Predicting Psychological Outcomes of Digital Life: A Machine Learning Analysis on European Data

Diogo Alexandre Rocha Moreira[1]

Master Thesis for the academic degree
Specialized Master of Science in Digital Neuroscience

Main Supervisor: Prof. Dr. Chantal Martin Sölch
Co-Supervisor: Dr. Simon Ruffieux
Co-Supervisor: MSc. Sebastien Chappuis

August 14, 2025

Department of Informatics - Master Thesis

Département d'Informatique - Departement für Informatik • Université de Fribourg - Universität Freiburg • Boulevard de Pérolles 90 • 1700 Fribourg • Switzerland

phone +41 (26) 300 84 65 fax +41 (26) 300 97 31 Diuf-secr-pe@unifr.ch http://diuf.unifr.ch

[1]diogo.rochamoreira@unifr.ch, Master Thesis with HUMAN-IST group, DIUF, University of Fribourg

Abstract

The last decades have seen the appearance and increase in digital technology use as well as the rise of concerns concerning their potential impact in psychological distress and time pressure. While the increasing amount of data allows a more extended analysis of such relationships, few researchers have attempted to use regression-based machine learning pipelines to analyze larger datasets. The present thesis investigates the use of machine learning algorithms to predict psychological distress (DASS) and chronic time pressure (CTPI) from psychometric and socio-demographic variables. This approach focuses on predictive accuracy and interpretability through a regression-based pipeline. Five regression algorithms, Random Forest, Extra Trees, XGBoost, LightGBM, and Linear Regression, were evaluated through extensive hyperparameter optimization and statistical testing. Results showed that boosting algorithms (XGBoost for DASS and LightGBM for CTPI) significantly outperformed the Linear Regression baseline. SHAP values revealed that for DASS, problematic internet use was a strongly positive predictor of distress. For CTPI, quality of digital experience had low positive predictive value, which went against initial suppositions. A cross-coutnry analysis revealed substantial variation in the importance of predictors which highlights potential cultural influences on the effects of digitalization. Overall, the study highlights the value of complex machine learning algorithms for larger psychometric datasets. However, careful tuning of the models are required to avoid overfitting and the interpretation of SHAP values depends greatly on the generalization capability of the model.

Keywords: Machine Learning, Regression, Psychological Distress, Time Pressure, SHAP

Acknowledgments

I would like to express my sincere gratitude to Prof. Dr. Chantal Martin Sölch for the possibility of working with an extremely interesting European project that is TIMED and for all the resources she provided me during this thesis such as psychology seminars to help me get up to date with the methodology and the TIMED Symposium where I was able to meet the authors of many references I use in my own work. I would also like to express my sincere gratitude to MSc Sebastien Chappuis and Dr. Karl Daher which helped me throughout the year with the psychology theory and the workings of the TIMED project. I also sincerely am grateful for Dr. Simon Ruffieux and Dr. Michael Papinutto for coordinating me as machine learning and data analysis experts across the entirety of the thesis. I would also thank Dr. Quentin Meteier for having proposed this theme as a master project and providing the initial pipeline that served as foundation for my improvement as a data analyst through machine learning algorithms. Finally, I would like to profoundly thank my parents and girlfriend for their unwavering support that allowed my mental health to be kept stable until the very end.

Contents

1 Introduction

Since the invention of the electronic chip by Jack Kilby in 1958, considered as the start of the digital revolution (Harper, 2003; Paige, 2024), digital technology (DT) has progressively entered all great spheres of society (Travkina, 2022). The evolution of DTs and introduction of the internet transformed the ways that individuals communicate. The new ways of communication impacted interactions both in work and life, a phenomenon further amplified by the COVID-19 pandemic (Amankwah-Amoah et al., 2021). The increase in DT use has raised concerns about well-being, with excessive use being tied to depression, anxiety, and stress in multiple studies (Cain and Gradisar, 2010; Dissing et al., 2022; Ho et al., 2014; Khalaf et al., 2023). Digital technology was also found to affect the way that individuals perceive time. Indeed, sociologists highlight that increased technology is increasing the pace of life and compressing time (Castells, 2011; Eriksen, 2001; Rosa, 2013). The worries relating to the effects of digitalization on wellbeing and time experience have led to the formation of the TIMED project. A multidisciplinary consortium grouping researchers and data from six European countries to explore how digitalization is shaping Europe.

Building upon a large cross-cultural dataset collected by the TIMED consortium, this thesis aims to investigate how digital behaviors and perceptions, such as problematic Internet use and the quality of digital experience, correlate with psychological distress and chronic time pressure. To achieve this, the study employs machine learning (ML) regression algorithms to predict psychological outcomes like DASS-21 (Lovibond and Lovibond, 1995), and CTPI (Denovan and Dagnall, 2019) scores. The choice of ML over traditional statistical methods comes from ML's ability to handle large, complex datasets and uncover intricate, non-linear relationships among variables without relying on rigorous assumptions (Orrù et al., 2020; Wang et al., 2023). The predictive focus and consideration of non-linearities of machine learning allows for greater predictive power and thus potentially greater accuracy (Wang et al., 2023). Thus the thesis explores two research questions: 1) which type of ML algorithm may attain the highest performance and surpass traditional methods such as linear regression; 2) how does digitalization affect psychological distress and time perception of European individuals. The first research question allows preparing an efficiency and accurate algorithm that can then be used to interpret interactions between digitalization, wellbeing, and time experience. More specifically, the aim of this study is to identify how problematic use of the Internet affects psychological distress, how perceived quality of digital life affects chronic time pressure, and whether these associations vary significantly across countries. To do so, the study starts with a theoretical overview of digitalization, psychological distress, time pressure, and machine learning methods. It follows by reviewing machine learning analyses of psychometric-based datasets and current gaps in the literature. The methodology of algorithm comparison and interpretation are then detailed, together with an exploration of the dataset and TIMED project. Finally, the results for each task are analyzed and discussed.

2 Psychological Theoretical Background

2.1 Digitalization

This section details the effects of digitalization on psychological wellbeing and time experience. The definition of digitalization used throughout this study is "the way many domains of social life are restructured around digital communication and media infrastructures" (Brennen and Kreiss, 2016, p. 1). It is understood here as the transformation of society through the integration of digital technology (DT).

Rapid advancements in digitalization over the last decades have changed communication patterns and global connectivity. In 2024, 68% of people worldwide had internet access, a great increase from the 10% in 2005 (International Telecommunication Union, 2024). In addition, 79% of young people between the ages of 15 and 24 are online, and over 95% of them live in high-income countries (International Telecommunication Union, 2024). This surge in connectivity has changed the way that individuals consume information. People now use digital devices such as computers and smartphones to entertain themselves and interact with others (Zhao and Wang, 2023). Many facets of life have been changed by digital technology: intimate relationships can now be started and managed online (Hobbs et al., 2017), social networks can be grown at all times and are less limited by geographic borders (Ellison et al., 2007), and employees can now be in constant connectivity with their workplace (Adisa et al., 2017).

Digital devices and platforms are extensively marketed for their ability to save time and increase efficiency (Wajcman, 2008), and there are many tools (e.g., social networking sites, productivity software, etc.) that should do so. However, some individuals report that time feels scarce (Wajcman, 2008), and scholars already in the late 1990s expressed concern about DT's effect on mental well-being (Kraut et al., 1998). Those concerns have only grown over years with multiple associations that will be developed in the following section (Lee and Zarnic, 2024; OECD, 2019).

2.1.1 Digitalization effect on psychological wellbeing

A review of the literature concerning the effects of digital technology on psychological wellbeing allowed for the identification of three problematic dimensions of digitalization: modes of use, platforms and environments, and the purposes of use. Some problematics concerning these dimensions are detailed in the following section.

Modes of Use An important factor observed when perusing the literature is the amount and manner in which technology is used. For instance, the "media displacement" theory posits that time spent on digital activities displaces essential offline activities like sleep, exercise, and face-to-face interaction, which are important for mental health (Jensen et al., 2019; Kraut et al., 1998) and cognitive functions (Cain and Gradisar, 2010; Khalaf et al., 2023). Studies also report how excessive screen time, particularly at night, is associated with poor sleep quality, increased stress, and symptoms of depression and anxiety (Cain and Gradisar, 2010; Dissing et al., 2022; Khalaf et al., 2023). Furthermore, excessive engagement may lead to problematic use patterns such as Problematic Smartphone Use (PSU) (Elhai, Dvorak, et al., 2017), Problematic Internet Use (PIU) (Spada, 2014), Problematic Social Network Site Use (PSNSU) (Franchina et al., 2018), and Internet Gaming Disorder (IGD) (Nasution et al., 2019). It is important to denote that only IGD is recognized as an actual disorder by the DSM-5 (American Psychiatric Association, 2013) while other types of problematic use are considered by authors to study interactions between excessive use and wellbeing. Problematic use is consistently associated with addictive behavior (Ko et al., 2008), negative impacts on overall psychological wellbeing (Odacı and Çikrici, 2017), depressive symptoms (Whang et al., 2003; Young and Rogers, 1998), social anxiety (Ioannidis et al., 2018). A meta-analysis of eight studies observed significant positive associations between PIU and alcohol abuse, ADHD, depression, and anxiety (Ho et al., 2014). Excessive use can be promoted through phenomenons such as "Fear of Missing Out (FoMO)" which is defined as the "pervasive apprehension that others might be having rewarding experiences from which one is absent" (Przybylski et al., 2013, p. 1). Individuals with high FoMO tend to engage more with social media to avoid missing events and feeling socially excluded which may lead to problematic use (Franchina et al., 2018).
In contrast, the "Goldilocks" hypothesis suggests that moderate digital engagement through digital limitations during weekdays may benefit wellbeing (Przybylski and Weinstein, 2017). Indeed, moderate Internet use was found to have beneficial effects such as emotional support and suicide prevention (Barak, 2007). It also facilitates information access (Anderson et al., 1995), and is associated with higher levels of social life (Suárez Álvarez and Vicente, 2024). Similarly, the "online and offline integration" hypothesis posits that healthier patterns of Internet use and superior psychological wellbeing can be attained when users find balance in their online and offline worlds (Lin et al., 2018). These observations suggest how the effect of digital technology (positive or negative) on wellbeing, is dependent on the balance and moderation of its use.

Platforms and environments Different platforms, such as mobile devices, social media, and digitalized workplace environments have been studied for their impacts on wellbeing. Smartphones, as explained above, can be detrimental to wellbeing with excessive and misplaced use. However, they were also found to decrease loneliness and depressive symptoms when used responsibly (Dissing et al., 2021). In addition, mobile devices can enhance employees performance while working (Leftheriotis and Giannakos, 2014; Wu, 2013) and provide flexibility which improves work-life balance (Adisa et al., 2017). At work, Information and Communication Technologies (ICTs) have indeed enabled remote work, increased productivity, and decreased commuting-related stress (Adisa et al., 2017; Amankwah-Amoah et al., 2021; Brennen and Kreiss, 2016). However, this new culture also increased obligations of constant connectivity and erasure of boundaries between family and work (Adisa et al., 2017; Thomée et al., 2011). This constant connectivity has been linked to higher stress levels, reduced mental wellbeing, and family conflict which counteract the promised benefits of digitalization with newer burdens (Dragano et al., 2021). Finally, social media platforms (e.g., facebook, instagram) facilitate the accumulation of social capital (i.e., benefits obtained through online or offline social connections) thus offering psychological benefits for many users (Ellison et al., 2007). Two competing hypotheses attempt to explain who benefits most: the "rich-get-richer" hypothesis posits that individuals with higher social capital reap most benefits, while the "poor-get-richer" hypothesis suggests that those with low social capital may gain more from social media engagement (Pouwels et al., 2021). These findings propose a variable impact of DT depending on the platform or environment that integrates it.

Purposes of Use The effects of digital technologies also appear to depend greatly on what they are used for. For example, some cases of social smartphone use at night were associated with lower loneliness (Dissing et al., 2022). Conversely, engagement in non-social activities on smartphones (e.g., entertainment, news browsing) during periods of psychological distress has been shown to worsen mental health outcomes (Elhai, Levine, et al., 2017). A study by Zhou et al., observed that compensatory use of social media (i.e., as a coping mechanism) was associated with an increase on depression levels of introverted college students. Later, in a meta-analysis of 141 studies, Godard and Holtzman, aimed to observe whether active (i.e., direct engagement with the platform and users) vs. passive (i.e., consumption of social media without direct contribution) led to different influences on wellbeing. However, little difference was found as both were associated with anxiety and also greater online support (Godard and Holtzman, 2024). These conclusions suggest how not only the motivation behind engagement but also the actual type of content that is engaged with, influence mental outcomes.

2.1.2 Digitalization effect on time pressure

Digitalization not only affects mental wellbeing but also how individuals perceive and experience time. Multiple sociologists have discussed the effects of technological advances over the last decades. Through this section, some of the main theories on how digital technology and technology advances overall affect time experience will be discussed.
A first concept was proposed by Anthropologist Thomas Hylland Eriksen where he characterizes our era as the "tyranny of the moment" (Eriksen, 2001, p 5). In his book, Eriksen explains how society is shaped by a mass of real-time information and constant communication. This excessive access to information is proposed as filling the gaps in life thus creating a series of overly saturated moments (Eriksen, 2001). The author points out, already in the early 2000s, the duality of time-saving devices that bombard individuals with information and leave little time for reflection. This lack of empty periods of time is proposed as reinforcing the urgency and acceleration of time perception individuals feel (Eriksen, 2001).
In line with the tyranny of the moment is Hartmut Rosa's "social acceleration theory". It suggests that modern communication networks, instantaneous data flows, and rapid production technologies allow for squeezing more activities in a same time period (Rosa, 2013). This leads to a condensation of tasks with individuals experiencing more in the same units of time. This is amplified by a culture of urgency that promotes the need for constant optimization of schedules and time. The result, as proposed by Rosa, is that individuals feel rushed and overwhelmed which leads to increased pressure (Rosa, 2013).
Expanding on the theme of altered temporality, Manuel Castells observes that the rise of a network society has redefined our experience of time. Castells describes how digital networks negate sequential clock time and overwrites the old linear scheduled time of the industrial era (Castells, 2011). The concept of multitasking, in line with the condensation of more tasks in the same units of time, leads to a "space of flows" where activities can happen in any order or all at once (Castells,

2011, p 406). The network society is thus proposed to intensify time pressure by creating an environment in which speed and immediacy are expected (Castells, 2011).
Judy Wajcman offers a more nuanced analysis of the situation. She notes that individuals often blame digital devices (e.g., smartphones) for the feeling of urgency. However, she proposes that the expectations and parameters one sets upon itself are the main driver of feeling rushed and pressured (Wajcman, 2015). The author extends this theory by explaining that although digitalization accelerates life, it also opens new opportunities to reshape time use depending in how individuals adapt and learn to use the available technologies. To conclude, her stance is that technology together with society work together to construct our experience of time (Wajcman, 2015).
More recently, Rutrecht et al., 2021 demonstrates how flow states influence the perception of time. Flow is a state of deep absorption in an activity and is often experienced in digital contexts like video games or programming (Rutrecht et al., 2021). In their recent experimental study, the authors observed that participants in a virtual reality game entered a state of flow where they stopped noticing the clock. The more flow players experienced, the less they thought about time and the faster they perceived time to pass (Rutrecht et al., 2021). This acceleration is not considered by the authors as innate time pressure but instead a compression of time during immersive enjoyable experiences. However, they do advise against the double-edged nature of this state where once the state of flow ends, individuals can come to feel an increased scarcity of time that may hinder other tasks (Rutrecht et al., 2021).

2.2 Psychological Distress

2.2.1 Conceptualizing Psychological Distress

Psychological distress is a term that refers to a broad state of emotional suffering. Multiple definitions of this concept are present throughout the literature. This study will focus on the definition used by Viertiö et al. where psychological distress refers to "non-specific symptoms of stress, anxiety and depression" (Viertiö et al., 2021, p. 2). Throughout this study we use the concept as a general indicator of mental strain and emotional discomfort and evaluate it through its components (stress, depression, and anxiety).

Rather than representing a specific psychiatric disorder, psychological distress reflects a continuum of mental health. It can range from mild discomfort to more severe levels of disturbance which may vary in duration and intensity (Dohrenwend et al., 1980; Viertiö et al., 2021). As evidenced by Massé is often used as an indicator of mental health in epidemiological studies and clinical trials which makes it particularly interesting for this study involving multi-cultural populations (Massé, 2000).

Given its close association with the domains of depression, anxiety, and stress (Drapeau et al., 2012; Viertiö et al., 2021), psychological distress is often discussed through these three components. In the following sections, each construct will be explored by detailing their prevalence, symptomatology and theoretical foundation.

2.2.2 Stress

The concept of psychological distress used within this study can be destructured into three components: stress, depression, and anxiety. This section will focus on presenting the concept of stress which is very tightly linked with the etiology of both depression and anxiety.

Early concepts of stress The concept of stress traces back to 1915 with the description of "fight-or-flight" response by the physiologist Walter Bradford Cannon. This theory posits that animals react to threats with an activation of the "Sympathetic Nervous System" (SNS), implicated in preparing the organism for either fighting or fleeing (Cannon, 1915). The activation of the SNS strains homeostasis and shifts the body towards a state of hyperarousal, also known as an acute stress response (S. Lu et al., 2021).

General Adaptation Syndrome The concept was extended by Hans Selye in 1936 with the "General Adaptation Syndrome" (GAS), a sum of systemic reactions of the body to long exposure to stress (Selye, 1946). Selye observed, through animal experimentation, that when an organism is severely damaged by nocuous agents, it produces a three stage adaptation stress response (alarm, resistance, and exhaustion) (Selye, 1946; Selye, 1936). The alarm reaction is the initial response to a stressor, it includes a shock phase (i.e., immediate systemic damage) and a counter-shock phase (i.e., defensive responses), these reactions prepare the organism to face the stressor (e.g., increased heart rate, adrenal cortex activation) (Selye, 1946). If the stressor persists, the organism enters the resistance stage where it acquires resistance to that particular stressor while decreasing resistance to others (Selye, 1946). The final stage of exhaustion appears when the stressor persists beyond the body's ability to adapt. During this stage, the adaptive energy is depleted, leading to reappearance of alert symptoms, decreased resistance, and possible severe consequences (e.g., irreversible damage, death) (Selye, 1946).

Four variations of stress Some years later, Selye moving beyond a purely negative view, described four variations of stress. On one axis, we find "eustress", considered as "good" stress (i.e., stressors that are challenging in a motivating way and can help an individual improve), and "distress", representing "bad" stress (i.e., the commonly known type of stressors arising from threatening and harmful events) (Selye, 1974). On the second axis, representing intensity, we locate "hyperstress" (i.e., excessive amount of stress from too many stressors) and "hypostress" (i.e., insufficient amount of stress associated with boredom or a monotonous environment) (Selye, 1974). This model links to the concept of "hormesis" (i.e., low doses of a stressor can induce beneficial adaptive responses in an organism, while high doses can be harmful or inhibitory) (Li and He, 2009). Through hormesis, eustress may induce mild, easily resolvable, stress responses that enhance the adaptability of the organism. In contrast, distress would cause strong responses which impair homeostasis, cannot be resolved, and thus last longer (S. Lu et al., 2021).

Transactional model of stress Exiting a purely physiological approach on stress, psychologists Richard Lazarus and Susan Folkman proposed, in 1984, a psychological interpretation of stress that does not consider specific events as being stressors by themselves. Their transactional model of stress and coping posits that stress arises as a transaction between the individual and their environment (Lazarus and Folkman, 1984). According to the authors, events are evaluated through a series of appraisals where an individual judges whether it is a threat and available coping resources to resolve the threat. The transaction represents the balance between the event's demands and the individual's ability to cope (i.e., dealing with adversity). A negative balance (i.e., situation demands are higher than the coping ability) induces a stress response. This model implies two cognitive appraisals, the "primary" appraisal (evaluating the demands and threat of an event) and the "secondary" appraisal (evaluation of coping resources). Additionally, previous events can be reappraised to threats or challenges later as the individual gets acquainted with it (Lazarus and Folkman, 1984). This theory highlights the importance of individual differences (e.g., perception, past experiences) in the appraisal of events as stressful or mild challenges. Indeed, the same event may be stressful for someone but mildly so for another.

Hypothalamic-pituitary-adrenal (HPA) Axis Stress responses are regulated through a neuroendocrinal system known as the "hypothalamic-pituitary-adrenal" (HPA) axis, figure 1 represents a simplified overview of the axis. Both physical and psychological stressors, when detected, leads to the secretion of corticotropin-releasing hormone (CRH) by the hypothalamus into the pituitary portal system. CRH then stimulates the anterior pituitary to release adrenocorticotropic hormone (ACTH). Finally, ACTH prompts the adrenal cortex to secrete cortisol (often called stress hormone). Cortisol helps the body switch to a "fight-or-flight" state (elevated blood glucose, raised blood pressure, modulated immune and metabolic functions) (Tsigos and Chrousos, 2002). Although acute stress responses are usually solved through a negative feedback where cortisol inhibits CRH secretion, chronic stress can become maladaptive. Chronic activation of the HPA axis may lead to a dysregulation of the negative feedback system, maintaining stress even after the initial stressor has been solved (Tsigos and Chrousos, 2002). A dysfunction of the HPA axis has been found to be associated with neuropsychiatric disorders such as depression and anxiety which will be detailed in the following sections (Mikulska et al., 2021).

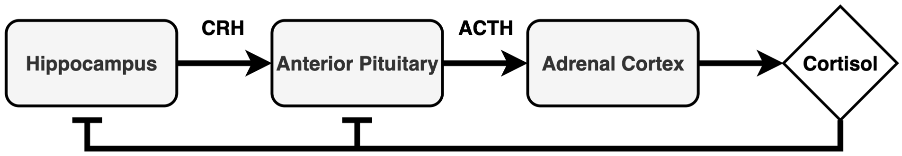


Figure 1: Simplified overview of the HPA Axis.

2.2.3 Depression

The term depression is often used colloquially to describe a range of mood disorders, that are, according to the DSM-5, characterized by a specific mood (sad, empty, or irritable) and functional changes (e.g., somatic, cognitive) (American Psychiatric Association, 2013). This section focuses on Major Depressive Disorder (MDD), also known as clinical depression, which is the classical condition in the group's disorders (American Psychiatric Association, 2013). This section will detail the prevalence, symptoms, and main theories involved in the etiology of MDD.

Prevalence of depression Recent research highlights the substantial global burden of depression, with prevalence rates varying greatly across populations, age groups, regions, and through the recent COVID-19 pandemic. An umbrella review of 83 meta-analyses estimated that approximately 30% of individuals experienced depression during the COVID-19 pandemic, with higher rates among medical students, teachers, and females (Mohseni et al., 2024). A global systematic review estimated pooled prevalence rates of 21.3% for mild-to-severe depression, 18.9% for moderate-to-severe depression, and 3.7% for major depression (B. Lu et al., 2024). In the United States, 2023 data show that 18.1% of adolescents aged 12-17 and 8.5% of adults aged 18 and older experienced a major depressive episode (MDE) in the past year, with the highest rates among multiracial individuals and young adults aged 18-25 (SAMSHA, 2024). In Europe, adult prevalence of clinically relevant depressive symptoms was reported as 6.54% based on the European Health Interview Survey (EHIS-3, 2018-2020) data. Wide variation was observed between countries which ranged from 1.85% in Greece to 10.72% in Sweden (Arias-de la Torre et al., 2023). The Global Burden of Disease (GBD) Study 2019 reported that depressive disorders affected approximately 280 million individuals worldwide, with a higher age-standardised prevalence among females (4158.4 per 100,000) compared to males (2713.3 per 100,000) (GBD 2019 Mental Disorders Collaborators, 2022). This is consistent with other findings showing women are nearly twice as likely to experience depression as men (Hagen and Rosenström, 2016). More recently, the GBD 2021 study indicated that depressive disorders accounted for 56.3 million years lived with disability (YLDs), making them the second leading cause of YLDs globally. The highest burden was observed among females aged 15-19 and 60-64 (The Lancet Psychiatry, 2024). These findings highlight the widespread and unequal distribution of depression, underscoring it as a leading and persistent public health challenge worldwide.

Symptoms of depression According to the DSM-5, Major Depressive Disorder is characterized by the presence of one or more major depressive episodes in the absence of a history of manic or hypomanic episodes. A major depressive episode involves a minimum of two weeks with a depressed mood or loss of interest in most activities (American Psychiatric Association, 2013). In addition to the MDE, at least four other symptoms must be present from the following: appetite or weight changes (increase or loss), sleep disturbances (hypersomnia or insomnia), psychomotor changes (agitation or retardation), general fatigue, feelings of worthlessness or guilt, impaired concentration or indecisiveness, recurrent thoughts of suicide (ideation or attempt) (American Psychiatric Association, 2013).

Biopsychosocial model The biopsychosocial model is a model that suggests mental health being affected by the interaction of biological, psychological, and social factors. It was propelled by Dr. George Engel in the late 20th century with the understanding that mind and body are closely associated and affect each other (Bolton, 2023). The model allows for a holistic approach on understanding the etiology of mental health disorders such as depression and anxiety. Akin to the unified model of depression by Beck and Bedemeier, it integrates findings from multiple levels of analysis, accounts for the entire symptomatology and provides a structure for understanding the antecedents, precipitation, and remission from the disorder (Beck and Bredemeier, 2016). Figure 2 presents an overview of the biopsychosocial model where biopsychosocial components serve as both moderators and antecedents in the development of mental disorders and their bidirectional conversion into physical illnesses. The following sections will thus briefly present current biological, psychological, and social models implicated in depression.

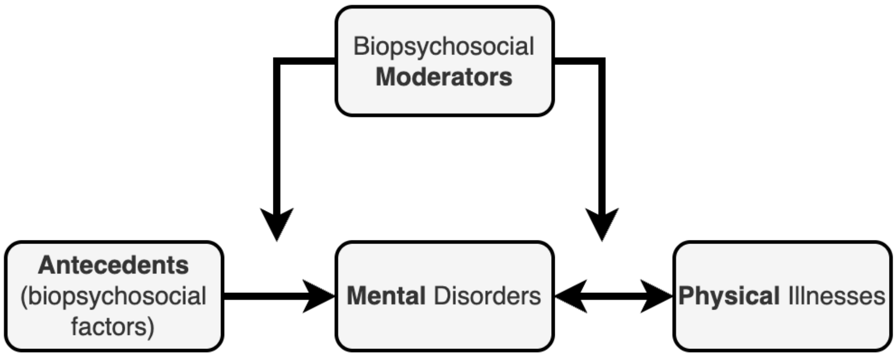


Figure 2: Overview of the biopsychosocial model of mental disorders. Illustrates interactions among antecedents and moderating biopsychosocial components with mental disorders and physical illnesses. Adapted from (L. H. Chen et al., 2023).

Biological theories Biological models propose that depression arises from physiological changes. This section will go through established biological theories such as the HPA axis, brain alterations, monoamine hypothesis, inflammation, and genetics.

HPA axis Dysregulation of the HPA axis has emerged as one of the most robust biological correlates of depression (Beck and Bredemeier, 2016). As detailed in 2.2.2, the HPA axis plays a central role in the body's adaptation to stressors. Du and Pang observed in their review that most depressed patients have high levels of cortisol and altered sensitivity to the HPA negative feedback system (Du and Pang, 2015). Other studies observed that although acute and severe subtypes of MDD have cortisol hypersecretion, this feature was absent from mild or atypical subtypes (Nandam et al., 2020). The role of the HPA axis and subsequently cortisol on depression is thought to be associated with an increase in stress sensitivity (i.e., being more susceptible to stressors) (Beck and Bredemeier, 2016). This sense is highlighted by the diathesis-stress model proposing that mental disorders arise from a vulnerability activated by stressful life events. A dysregulation of the HPA axis (making it more sensitive to stressors) fits the vulnerabilities of the models and may increase the risk of depression (Pruessner et al., 2017). Prolonged exposure to high cortisol levels may also lead to changes in brain structures such as hypertrophy of the amygdala and atrophy of the hippocampus (McEwen, 2003). These changes may further disturb the HPA feedback inhibition system and increase depression vulnerability (Mahar et al., 2014). Known associations between changes in these structures and depression will be elaborated in the next section.

Brain alterations Neuroimaging studies reveal consistent structural and functional alterations in the brain of patients with major depression compared to control subjects (Drevets, 2000). Key structures associated with depression are the prefrontal cortex, amygdala, and hippocampus (Beck and Bredemeier, 2016; Trifu et al., 2020). The prefrontal cortex (PFC) is anterior to the premotor and primary motor areas, and covers the front part of the frontal lobe of the brain. This structure is involved in executive functions such as decision making and working memory (Fuster, 2019). Impairment of this region may lead to disrupted emotional and cognitive processing and volume reduction of PFC areas was observed in patients with depression (Palazidou, 2012). These structural changes correlate with ruminations, impaired prosexual function, and irrationally negative cognition present in MDD (Trifu et al., 2020). The amygdala is a paired nuclear complex located medially in the temporal lobes. It is part of the limbic system and has a primary role in the processing of memory, decision-making, and emotional responses (e.g., fear and anxiety) (Pabba, 2013). Altered activation of the amygdala was found to correlate with the severity of depression (Palazidou, 2012). Depressed patients show hyperactivation of the amygdala when exposed to negative stimuli and hypoactivation when exposed to positive stimuli (Trifu et al., 2020). The hippocampi are a pair of structures located in both hemispheres of the brain in the medial temporal lobe. This structure plays an essential role in the consolidation of information from short-term to long-term memory (Kandel et al., 2021). In a recent study by Sun et al. (2025), patients with recurrent MDD were found to have progressive deterioration of their hippocampi with prolonged depression (Sun et al., 2025). In addition, lower hippocampus volume is a characteristic found in adolescents with higher depression risk (M. C. Chen et al., 2010), and those having experienced early life adversity (Rao et al., 2010). The role of the hippocampus in learning and memory as well as its role on the HPA axis (through the fornix that provides inhibitory feedback) makes the hippocampus a candidate for increasing the risk of depression when dysfunctional (Palazidou, 2012). Together the PFC, hippocampus, and amygdala form a core part of the limbic-cortical circuit which is responsible for integrating emotional regulation, memory processing, and executive control. Dysfunction within this circuit (e.g., atrophy, altered function) is proposed to disrupt emotional reactivity and increase vulnerability to mental disorders such as depression (Palazidou, 2012).

Monoamine hypothesis This theory posits that depression arises from deficits of monoamines such as serotonin, norepinephrine, and dopamine and is one of the most influential models across the past decades (Delgado, 2000). Antidepressant drugs such as Selective Serotonin Reuptake Inhibitors (SSRIs) or Serotonin-Norepinephrine Reuptake Inhibitors (SNRIs) work by blocking transporters (such as the serotonin transporter), thus increasing availability of the monoamines in the synaptic cleft. These drugs were found to increase neuroplasticity in certain brain regions and improve mood over weeks of treatment which led to the monoamine hypothesis (Delgado, 2000; Hillhouse and Porter, 2015). However, this theory was widely debated due to inconsistent evidence on the levels of monoamines in depressed patients. Furthermore, a recent umbrella review by Moncrieff et al. (2023) of major research areas related to the role of serotonin in depression found no consistent evidence supporting the hypothesis. Analyses of serotonin and its metabolites, receptor binding, transporter levels, tryptophan depletion, and genetic studies all failed to show reliable associations with depression (Moncrieff et al., 2023).

Inflammation Inflammation plays a significant role in the pathophysiology of depression. In fact, individuals suffering from severe depression often showing elevated levels of pro-inflammatory markers such as IL-6, TNF-α, CRP, and IL-1β (Miller and Raison, 2016). These markers are also associated with "sickness behavior" which is characterized by fatigue, reduced motivation, and social withdrawal, mirroring core depressive symptoms (Stieglitz et al., 2015). Beck's unified theory of depression proposes that such symptoms evolved as an adaptive response to the loss of vital resources (e.g., important relationships or a job), aiming to conserve energy (Beck and Bredemeier, 2016). However, in modern contexts, stress and early life adversity can amplify inflammatory responses, rendering this system maladaptive (Kaufmann et al., 2017; Miller and Raison, 2016). Inflammatory response may affect the brain by altering neurotransmitter metabolism, neuroplasticity, and neurocircuitry in pathways that regulate mood, motivation, and reward which may affect the development of depression (Miller and Raison, 2016). The relationship between depression and inflammation is bidirectional: inflammation can induce depressive symptoms, while depression can increase blood-brain barrier permeability thus amplifying neuroinflammatory effects (Kaufmann et al., 2017; Stieglitz et al., 2015). Elevated inflammation has been linked to increased risk of depression in population-based studies (Gimeno et al., 2009; Köhler et al., 2016), and high inflammatory markers in childhood predict adult depression (Goodwin, 2011; Khandaker et al., 2014). Conversely, childhood depression may sensitize individuals to inflammation later in life (Raison et al., 2010). These findings support the bidirectional relationship between depression and inflammatory responses. While anti-inflammatory treatments have shown benefits in high-inflammation profiles of depression (Köhler et al., 2016), their therapeutic value remains debated due to potential interference with SSRIs (Warner-Schmidt et al., 2011). Interestingly, SSRIs may reduce inflammation, which also supports a mechanistic link (Gałecki et al., 2018).

Genetics In his unified theory of depression, Beck highlighted the polygenic nature of depression vulnerability (Beck and Bredemeier, 2016). Specific gene variants, such as the 5-HTTLPR short allele (serotonin receptor, Karg et al., 2011), BDNF (brain-derived neurotrophic factor, Correia et al., 2023), and FKBP5 (glucocorticoid/cortisol receptor, Binder et al., 2004; Zimmermann et al., 2011) were found to be implicated on the moderation of the relationship between stress and depression. These genes affect biological systems related to stress reactivity and emotional regulation, including the HPA axis and brain regions such as the hippocampus and the amygdala. Genetic risk factors also interact with environmental stressors (e.g., parental loss or abuse) which influences the vulnerability. Furthermore, personal experiences can alter gene expression through epigenetic changes which may affect the vulnerability profile of individuals over their life (Beck and Bredemeier, 2016). A recent systematic review of cadidate gene studies by Norkeviciene identified 85 genes with potential associations to depression. Most implicated genes are involved in neurotransmission, synaptic plasticity, and neuronal survival (particularly within glutamatergic pathways and monoaminergic systems). However, few findings have been consistently replicated and due to lack of statistical power, the field still lacks definitive conclusions (Norkeviciene et al., 2022). These findings however highlight the polygenic nature of depression which gives rise to a grand variety of genetic profiles of vulnerability or resilience to depression.

Psychological theories This section will focus on two current psychological theories of depression: Seligman's learned helplessness theory and Beck's cognitive theory.

Cognitive Theory of Depression The cognitive theory of depression posits that the disorder arises from deeply rooted negative beliefs formed early in life. They influence the way that individuals interpret their life experiences and give rise to automatic negative thoughts about the self, the world, and the future. This is called the cognitive triad which is the core component of the cognitive theory. Beck argued that these distorted thought patterns play a central role in the development and maintenance of depressive symptoms (forming a self-reinforcing cycle) (Beck and Rush, 1979). More recently Beck and Bredemeier proposed a unified model of depression that places cognitive vulnerabilities (cognitive triad) within broader biological, evolutionary, and environmental contexts. This integrative framework conceptualizes depression as an adaptive response triggered by the perceived loss of vital resources (e.g., job, relationships) with the aim of conserving energy to enhance survival. Individuals have resistance profiles that depend on biological (e.g., genetic, neurological, endocrine) and social (e.g., socioeconomic status, life experiences) factors. These influence the information processing capabilities of individuals, further pushing them towards negative or positive interpretations of their experiences. Individuals with vulnerability predispositions will interpret events with a negative approach, reinforcing the cognitive triad. The cognitive appraisals then activate biological systems (autonomic, immune, neuroendocrine) which manifest symptoms like anergia, withdrawal, and anhedonia. Finally, the model proposes that depression can be decreased by restoring lost resources and with the help of exteranl factors (psychotherapy, social support, medication) to mitigate cognitive biases and replenish resilience (Beck and Bredemeier, 2016).

Learned Helplessness Seligman's learned helplessness provides a complementary perspective, focusing on the impact of uncontrollable experiences and attributional style. This theory originates from animal experimentation, where animals subjected to inescapable stress would later fail to escape even when possible. Applied to humans, the theory suggests that depression emerges from an individual observing that its actions cannot solve a situation. They learn to feel helpless, become depressed and stop trying to escape the situation (Seligman, 1972). Expanding on the learned helplessness model, the hopelessness theory of depression integrates attributional and cognitive processes to explain why some individuals are more susceptible to depression following adverse events. The theory identifies negative inferential styles (tendencies to attribute negative life events to internal, stable, and global causes) as key vulnerabilities. When individuals with such styles encounter negative life events, they are more likely to develop a sense of hopelessness which increases vulnerability to depression. This theory is grounded on the diathesis-stress model and bridges cognitive and attributional models, highlighting the crucial role of perceived causality and hopelessness in the onset and course of depression (R. T. Liu et al., 2015).

Focus on Beck's cognitive theory Abela and D'Alessandro, 2002; Beck, 2008; Beck and Bredemeier, 2016; Beck and Rush, 1979. Learned helplessness (potentially) (Seligman)

Social factors The biopsychosocial model of depression considers social determinants as a major driver of vulnerability to depression. Factors such as the socioeconomic status of an individual (i.e., economic and social standing) were found to be associated with mental and physical health. With higher social status correlating positively with better health, due to higher status facilitating access to vital resources (Bolton, 2023). Negative social events are another major factor as they are found to increase depression risk (Mazure, 1998) and can be directly tied to the loss of vital resources leading increased vulnerability (Beck and Bredemeier, 2016). In fact, early adversity was found to be associated with brain alterations that are often observed in depressed patients (Kim et al., 2019) and childhood maltreatment is consistently associated with depression (Nelson et al., 2017). This relationship was suggested to be bidirectional as individuals with depression tend to have worse social environments which elevate the symptoms of depression (Coyne, 1976). Depressed patients are also more likely to experience social rejection which amplifies the vulnerability further (Segrin and Dillard, 1992). Other factors such as higher reassurance seeking behavior (Joiner and Metalsky, 1995), high job strain (Seidler et al., 2022), and decrease in social support (Kendler and Gardner, 2014) can also increase the vulnerability of an individual towards developping depression when affected by stressors (Beck and Bredemeier, 2016).

2.2.4 Anxiety

The term anxiety is frequently used in everyday language to represent a range of emotional states involving worry and tension. According to the DSM-5, anxiety disorders are characterized by excessive or uncontrollable worry, fear, or apprehension. These are often accompanied by somatic, cognitive, or behavioral symptoms that cause clinically significant distress or impairment (American Psychiatric Association, 2013). This section offers a concise overview of the prevalence, symptoms, and current theories concerning anxiety.

Prevalence of Anxiety Anxiety disorders represent a significant global health burden. In 2021, they were the sixth leading cause of YLDs, accounting for 42.5 million worldwide (The Lancet Psychiatry, 2024). The GBD 2019 study estimated that approximately 301 million people (about 4.05% of the global population) were affected by anxiety disorders (GBD 2019 Mental Disorders Collaborators, 2022). Anxiety prevalence is generally higher in high-income, urban regions and among women, who are 1.66 times more likely to be affected (Javaid et al., 2023). During the COVID-19 pandemic, global anxiety prevalence rose sharply, with pooled estimates from an umbrella review showing a rate of 30.4% (Amini-Rarani et al., 2024). Anxiety disorders prevalence rivals that of depression which emphasizes them as another major public health issue (GBD 2019 Mental Disorders Collaborators, 2022).

Symptoms of Anxiety According to DSM-5, anxiety is characterized by excessive worry about future events, often accompanied by psychophysiological symptoms such as restleness, muscle tension, fatigue or irritability (American Psychiatric Association, 2013). Unlike typical anxiety (which may be experienced by most individuals), pathological anxiety causes significant distress and persists beyond a specific stressor that may have triggered it (American Psychiatric Association, 2013). Multiple anxiety disorders are detailed in the DSM-5, each having specific criteria that help distinguishing different forms of anxiety. One of such disorders closely embodies the generalized presentation of anxiety, it is called Generalized Anxiety Disorder (ah ma GAD). It is characterized by persistent and excessive anxiety or worry about a variety of events, occurring often during six months. This worry has to be difficult to control by the individual and associated with three or more symptoms such as restlessness, fatigue, difficulty concentrating, irritability, muscle tension, and sleep disturbances (American Psychiatric Association, 2013).

Biopsychosocial model As with depression, anxiety arises from the complex interplay of biological, psychological, and social factors. This supports the biopsychosocial model as one of the most comprehensive frameworks for understanding anxiety disorders (Jokinen and Hartshorne, 2022). Building on the same integrative framework outlined in section 2.2.3, the following sections will briefly present key theoretical models within each domain as they pertain to anxiety.

Biological theories Biological models have great importance in the study of anxiety disorders due to the strong association between anxiety and autonomic responses. The original adaptive role of anxiety is tied to the "fight-or-flight" response to threats, where caution is promoted to enhance survival. Multiple physiological systems are thought to have evolved over time to promote healthy doses of anxiety that would increase threat survival (e.g., predators, natural dangers) (Freeman and Freeman, 2012; Price, 2003). Some of these biological systems are detailed in the following section.

HPA axis Similar to depression, stress and the HPA axis were proposed as a key neuroendocrine component of anxiety disorders (Faravelli, Lo Sauro, Lelli, et al., 2012). A review by Faravelli et al. (2012) reported evidence to altered HPA axis function in anxiety disorders. HPA axis activity patterns vary by subtype: PTSD often shows high CRH with low basal cortisol; panic disorder findings are mixed; OCD tends toward HPA hyperactivity; social anxiety shows normal baseline levels but exaggerated reactivity to psychosocial stressors; and GAD is frequently linked to hypercortisolism. Some of the proposed causes to this dysregulation incluse: increase of stress responsivitiy through persistent anxiety levels; damage of hippocampal glucocorticoid receptors by persistent elevated cortisol levels (i.e., decrease negative feedback on CRH secretion, leading to higher CRH and cortisol concentrations); early stressors over-activating the HPA axis during developmental processes (Faravelli, Lo Sauro, Godini, et al., 2012). The disruption of the HPA axis is also proposed as leading to systemic inflammation and later neuroinflammation, increasing neurotoxic effects on limbic and pre-frontal brain structures and, by extent, vulnerability to anxiety disorders (Won and Kim, 2020).

Neurochemical factors In addition to the HPA axis, research with benzodiazepines (which enhance GABA, an inhibitory neurotransmitter) showed that replenishing GABA in emotion-processing brain regions allowed to reduce anxiety symptoms. This hints towards an association between decreased inhibition by GABA signaling and anxiety disorders (Nuss, 2015). Increased excitatory neurotransmission by glutamate in those same areas is equally thought to lead to anxiety (Martin et al., 2009). Likewise, the serotonergic system is implicated, with SSRIs being first-line treatments for chronic anxiety. Serotonin has been the main neurotransmitter associated with anxiety for the last decades, however its actual role is still highly debated. Indeed, an increase in serotonin was observed as beneficial in decreasing panic symptoms while facilitating more generalized forms of anxiety (Zangrossi et al., 2020). Other monoamines such as dopamine and norepinephrine are consistently implicated in the pathogenesis of anxiety disorders through their roles in monoaminergic systems targeted by anxyiolitic and antidepressant drugs (Martin et al., 2009).

Brain structures Multiple brain structures have been mapped as involved in anxiety disorders, among them, functional imaging converges on hyperactivation of the amygdala and insula regions. The amygdala has a central role in the evaluation of emotional and social stimuli and the perception of threats which is thought to be dysfunctional in anxiety disorders (Schmidt et al., 2018). The insula together with the anterior cingulate cortex constitute a "fear network" involved in conflict monitoring and fear learning. Hyperactivity of this network is consistently observed in anxiety (Schmidt et al., 2018). In contrast, the ventrolateral PFC was reported as helping regulate amygdala activity and controlling emotional responses. Examination of anxiety patients found hypoactivation of this region during a verbal fluency task and a negative correlation between its activation and social avoidance (Yokoyama et al., 2015).

Genetics Biological models also recognize the role of genetic factors in developing vulnerability to anxiety. Akin to depression, twin studies show substantial heritability for anxiety disorders, with estimates around 30 to 50% (Smoller, 2016). Of the various anxiety disorders, panic disorder has received particular attention and multiple candidate-genes exist (e.g., monoaminergic, GABA receptor, stress hormone, neurotrophic). However, although their revelevance is supported by imaging-genetic studies, association results are inconsistent (Smoller, 2016). Overall, no single anxiety gene exists, rather a polygenic model influencing vulnerability seems to be more appropriated (Gottschalk and Domschke, 2017). In addition to random mutations, epigenetic mechanisms are also of importance when considering the development of anxiety. Epigenetics refers to chemical modifications in the DNA or histones that modify gene expression without changing the actual genetic code. These are often caused in response to environmental factors, and some epigenetic associations with anxiety behaviors and disorders were reported (Nieto et al., 2016). Although genetic studies are inconclusive, heritability estimates do seem to point towards an important role of genetics in the development of anxiety.

These biological insights underscore that while the capacity for anxiety is deeply rooted in our biology, it is also dynamically shaped by our experiences.

Psychological theories Psychological theories focus on the mental processes and learning experiences that contribute to the onset and maintenance of anxiety disorders. They explore how thoughts, beliefs, and environmental influences shape anxious responses, and offer a complementary lens to biology through which to understand the complexity of this condition. Psychological vulnerabilities of anxiety lie in different levels which vary depending on the actual disorders. Some factors that are widely discussed are the learning and cognitive processes implicated in the development of such disorders as they underlie cognitive-behavioral therapy (CBT) (Craske and Stein, 2016). Learning theories such as Mowrer's two-factor model posit that anxiety can be acquired through conditioning and maintained through avoidance and safety behaviors. Neutral events can be associated with fear which elicit adaptive safety behaviors such as avoidance of the threat. However, such associations can be perpetuated over time and become maladaptive, leading to behaviors that reduce short-term anxiety but perpetuate it long-term. These theories do not account for individual differences in anxiety in individuals with similar negative experiences (Krypotos et al., 2015). This section will instead focus on relevant cognitive factors associated with anxiety.

Cognitive theory Founded by Aaron T. Beck, the cognitive theory of anxiety emphasizes that anxiety originates from how individuals interpret or appraise situations (rather than the situations themselves). It is based on the cognitive approach that shifted psychological focus from behavior to internal mental processes. The interpretation of stimuli can lead an individual to perceive it as a threat and anxiety arises when the individual believes they cannot handle it (Freeman and Freeman, 2012). These interpretations are shaped by schematic beliefs developed through life experiences (e.g., feeling vulnerable, assuming the worst). Anxious individuals also tend to have cognitive biases (e.g., overestimating threats, underestimating themselves) and to engage in safety behaviors (e.g., avoidance) which reinforce and maintain anxiety. Findings in cognitive complemented learning and behavioral theory to form the widely used "Cognitive Behavioral Therapy" that is used both in anxiety and psychological disorders. This therapy aims to change the negative appraisals and thought patterns formed by patients that increase vulnerability to anxiety (Bentley et al., 2021).

Social factors To complete the biopsychosocial model of anxiety, the consideration of social factors is essential. The diathesis-stress paradigm emphasizes not only the importance of biological/psychological predispositions (diathesis) that increase vulnerabilities, but also the environmental stressors that trigger a disorder (Pruessner et al., 2017). A systematic review by Gardner et al. (2019) reported that child maltreatment overall had a significant association with later anxiety disorders. In particular, the authors observed sexual abuse, physical abuse, and neglect as having particularly high associations compared to other types of early adversity (Gardner et al., 2019). Even later in life, quality of environmental relations was observed to have a consistent interaction with anxiety development. Notably, high workplace adversity (e.g., high job strain) (Seidler et al., 2022) and low social support (J. Liu et al., 2024) were observed as being associated with anxiety disorders. This overview is of course not exhaustive and many more social stressors can serve as triggers to anxiety disorders. However, these already demonstrate the high interaction between the social domain and anxiety disorders.

2.3 Time Pressure

Having outlined the key concepts of psychological distress as it is understood throughout this study, we next address time pressure. This section will conceptualize time pressure and clarify how the term is used throughout the study.

2.3.1 Conceptualizing Time Pressure

Complaints about shortage of time have become ubiquitous in modern society, with the experience of not having enough time becoming more popular in the past decades (Eriksen, 2001; Gleick, 1999). In the literature, multiple terms have been used to describe the experience of having little time (e.g., time pressure, scarcity, stress, famine, deficit, etc.), these have been used interchangeably and were found to commonly refer to some form of perceived "time pressure" (Denovan and Dagnall, 2019). Kleiner describes time pressure as "the sense that one's duties and responsibilities exceed one's ability to complete them in the time available" (Kleiner, 2014, page 1). The concept involves both the perception of time and role obligations necessary to accomplish within a specific time frame (Kleiner, 2014). Time pressure is composed of both an objective component and a subjective component. The objective aspect, sometimes called "time shortage", involves cognition about not having enough time for an activity. The subjective aspect represents the feeling of being rushed or harried, and focuses on a higher pace of life requiring to do tasks faster and constant vigilance about tight schedules (Szollos, 2009). These aspects were distinguished as the cognitive (i.e., objective) and affective (i.e., subjective) factors of time pressure (Denovan and Dagnall, 2019; Szollos, 2009).
Time pressure can negatively impact physical health (e.g., insomnia, fatigue, obesity, high blood pressure, hypertension) and mental health (e.g., depression, stress, emotional exhaustion, lower subjective well-being). Shortage of time has also been associated with harmful behaviors for the self (e.g., unhealthy eating, less exercising) and for others (e.g., dangerous driving, not providing help) (Rudd, 2019). Although time pressure is mainly considered as negative, some positive effects have been highlighted. Festini et al., for instance, assessed a positive correlation between busyness and cognitive performance (Festini et al., 2016). Increased time pressure can also be a sign of status and success in some cultures, with individuals aspiring to be busy (Bellezza et al., 2016).
In sociology, time pressure is viewed as a result of the accelerated pace of life from the contemporary era. The network society was described by Castells as being characterized by "timeless time", with new technologies promoting instantaneity, compression of time, and blurring of life's borders. Castells (Castells, 2011) theorized this evolution of the time experience of society as potentially increasing stress and pressure in multiple aspects of life. Later, sociologist Hartmut Rosa, described this phenomenon as "social acceleration" where technological acceleration induces an acceleration of social changes which, in turn, speeds up the pace of life and causes time pressure (Rosa, 2013). The theory was extended by Wajcman, acknowledging the impact of technological acceleration on perceived time shortage and highlighting that the actual effects of technology on time perception would likely vary depending on cultural factors (e.g., valuing busyness) (Wajcman, 2008, 2015). These sociological insights suggest that time pressure is structurally embedded in modern social systems, it is not merely a personal failing of time management, and appears to be a condition tied to cultural norms and institutional demands (e.g., workplace deadlines, digital communication expectations). Szollos (2009) proposed chronic time pressure as an umbrella for recurring experiences of time shortage, retaining the same cognitive (objective shortage) and affective (feeling harried) components (Szollos, 2009).

3 Machine Learning Model Theory

This section provides an overview of the machine learning algorithms used throughout the study. It begins with an introduction on the evaluation of regression models. Then follows with an explanation of individual algorithms such as Linear Regression and Decision Tree. Afterwards, it explains the ensemble algorithms as methods that combine multiple individual learners to produce better predictions. In particular, we discuss two ensemble techniques: bagging and boosting as extensions of decision trees. Under bagging, we explain Random Forests and Extra Trees, while the boosting section details XGBoost and Light GBM.

3.1 Regression Performance Evaluation

The nature of the dataset and targets requires the study to employ regression-based algorithms. Regression tasks can be evaluated through a multitude of metrics. The metrics chosen for this study are: Mean Squared Error, Mean Absolute Error, R^2.

3.2 Linear Regression

Linear regression is a bridge between traditional statistics and machine learning. Historically it was developed as a tool for statistical inference and later became one of the foundations of traditional machine learning. Simple linear regression considers only the interaction between one independent variable and the target. It can be represented as the following model $y = w_0 + w_1 x$. With w_0 as the intercept term, and w_1 as the slope or coefficient determining the interaction between the variables (Maulud and Abdulazeez, 2020). For training such a model we use the least squares method that will find the best weights (w_0, w_1) values that minimize the sum of squares of residuals. Residuals are the absolute difference between a predicted value by the model and the true value (Qu, 2024). The sum of residuals is the total error of the model and by minimizing it, we produce a model capable of predicting values close to truth (Maulud and Abdulazeez, 2020; Qu, 2024). For higher dimensional datasets, with more than one independent variable, multivariate linear regression is used $y = w_0 + w_1 x_1 + ... + w_n x_n$ with one weight/coefficient per feature (Maulud and Abdulazeez, 2020). After optimizing the parameters (intercept and weights) of a linear regression model we obtain a regression line that best fits the data. The main advantages of linear regression are the simplicity of interpretation and the computational efficiency. Interpretation is straightforward because we can easily observe the influence of each variable in prediction through their coefficients. It is however limited by its linear nature and simplicity which can cause underfitting in complex non-linear datasets.

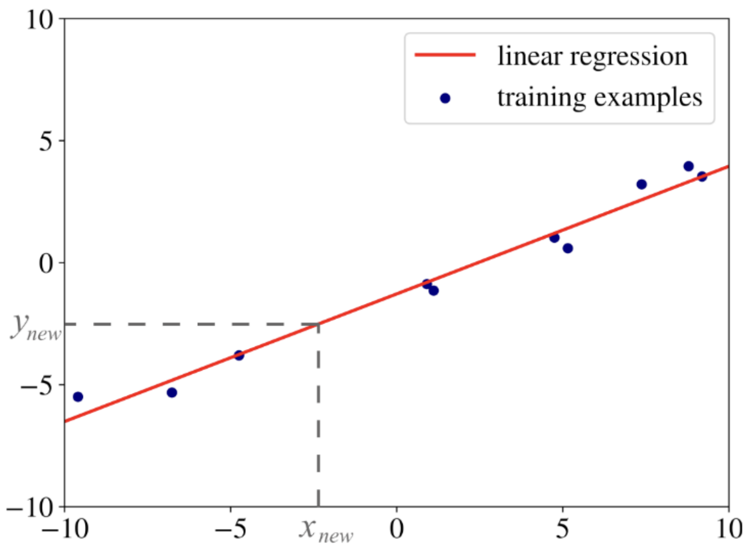


Figure 3: A fitted linear regression line over training data. Adapted from Burkov, 2019

3.3 Decision Tree

Decision Trees function by splitting the prediction through a series of decisions (deconstructing a problem into simpler problems) (Ahmad et al., 2018). The tree is composed of hierarchical conditions that are applied successively from the root (start of the problem) to one leaf (final prediction) of the tree (Breiman et al., 1984). The CART algorithm proposed by Breiman et al., 1984 supports both classification (Gini/entropy) and regression (variance reduction) tasks. The root node represents the entire dataset and first splitting decision. At each node, a splitting rule is applied to organize data into child nodes. Training a decision tree is about improving the decision rules according to a metric. For classification tasks, we optimize Gini impurity or information gain and want to have most of one class on each node after splitting (Breiman et al., 1984). When the target is continuous (regression task), the metric used is "variance reduction". When optimizing variance reduction, we want nodes to have low variance (i.e., values in the same node should be similar) after splitting (Breiman et al., 1984). During training, the decision tree tests many potential splits. The split that optimizes the chosen metric will be selected as it produced the most homogeneous groups. This homogeneity after splitting helps improve the final predictions. The leaf nodes of the tree are the final class labels or regression values used for prediction. The advantage of a decision tree is its clear decision-based structure that makes them a highly understandable and interpretable model. We can easily observe the criteria used for each split and understand the inner workings of the model. However, DTs are susceptible to high variance (i.e., small dataset changes can greatly modify tree structure) and tend to overfit without pruning branches (Ahmad et al., 2018).

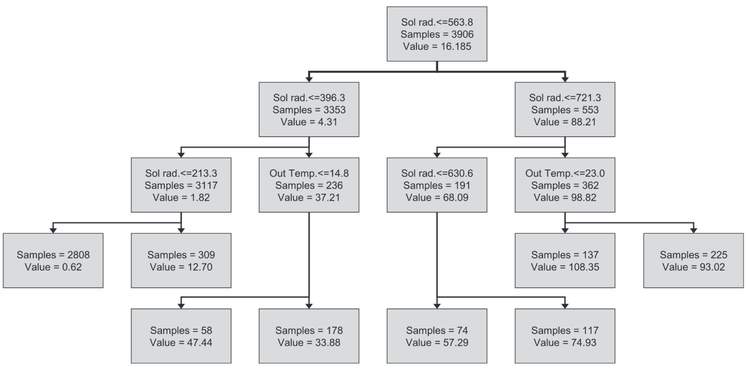


Figure 4: A regression decision tree that predicts energy gain from a solar collector. Adapted from Ahmad et al., 2018

3.4 Ensemble models

Ensemble models is a machine learning method focusing on the use of multiple individual learners (i.e., individual machine learning models) together on a task. Doing so generally improves performance and overcome limitations of the individual models (Dietterich, 2000). There are multiple ways of producing ensemble methods. Some of them are simple computations of the learner results such as voting or averaging the predictions. Voting is done for classification tasks, there each learner votes for their predicted label and the one getting most votes is the final prediction (Nti et al., 2020). For regression tasks a basic method is averaging the predictions of all learners (Nti et al., 2020). Then there are more advanced methods that merge learners. Stacking is an ensemble technique involving the training of a new model, a meta-learner trained on the predictions of all individual learners. The meta-model learns the optimal combination of the base predictions serving as meta-data (i.e., linear combination of weights) and produces a final prediction (Wolpert, 1992). In stacking, the base learners are trained with cross-validation and the meta-data is obtained through the out-of-fold predictions (Nti et al., 2020). Blending is conceptually similar to stacking but uses a holdout validation set instead of full cross-validation. This makes the method more computationally efficient but can affect performance since it uses less data (holdout set) to produce the meta-data thus risking information loss (Nti et al., 2020). In this study we will be focusing on two advanced ensemble techniques: Bagging and Boosting.

3.4.1 Bagging

Bagging, also known as bootstrap aggregating, works through the merging of multiple models predictions to decrease variance of the individual learners (Ahmad et al., 2018; Nti et al., 2020). The core idea is to train multiple versions of a same model type on different subsets of data. Each individual learner is produced with randomness and their predictions can be averaged or voted depending on the task. The two main algorithms that will be focused on from this method are Random Forest and Extra Trees. Random Forest uses actual bootstrap aggregating and trains each individual learner on a different subset of data and features (Breiman, 2001). Extra Trees, however, uses the full dataset for each learner (no bootstrap aggregating) while randomizing features still. It increases randomness by also randomizing the split rules for each learner (instead of optimizing each split) and is still usually considered as a bagging method due to the aggregating of individual learners (Baladram, 2024; Geurts et al., 2006).

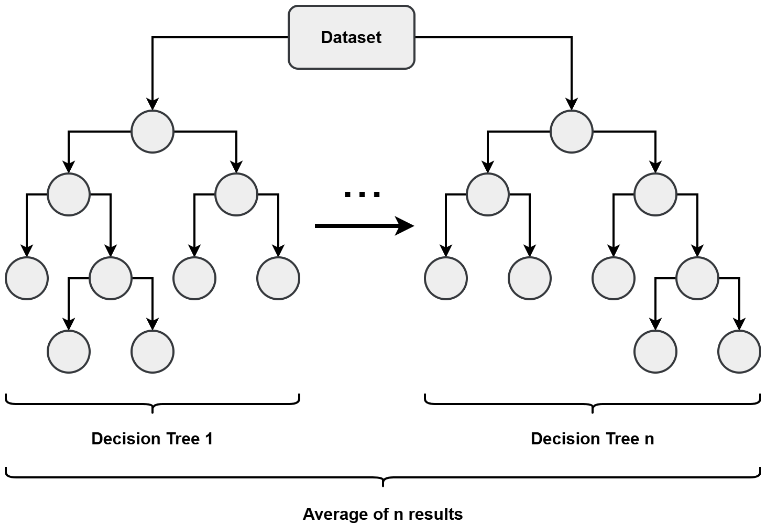


Figure 5: Diagram of a random forest as an example of bagging algorithms. N decision trees with random structures are produced in parallel.

Random Forest (RF) is a tree-based ensemble method, it was designed to improve upon individual Decision Trees. It consists of many weak decision tree learners that are trained in parallel to decrease the bias and variance of the overall model (Ahmad et al., 2018). A random forest is trained by drawing bootstrapped sample sets from the original dataset. Each sample is used to train a separate tree, then from the sample only a small and fixed number of predictors/features is used. This produces a specific chosen number of trees, each trained on different random data subsets and predictor subsets. By using samples and different predictor subsets instead of the full dataset, we ensure that each tree is considerably different from each other (this is bagging) which helps decrease the overall variance of the model (Breiman, 2001). In order to run inference on a random forest model, each individual learner gets the data and produces a prediction. The final result is the mean (for regression) or mode (for classification) of all predictions. For each individual learner of the random forest model, the rest of data not used for training is stored and used for computing the "out-of-bag score". This score is computed for each learner and averaged to obtain the performance of the algorithm as an unbiased estimation of generalization error (Breiman, 2001). Random Forest also provides model-specific interpretation through its internal coefficients. These are computed by averaging the difference between predictions of the individual learners that use a feature against those that do not use it. The difference reflects the contribution of the feature to predicting the target variable (thus its relationship to the target) (Breiman, 2001).

The advantages of RFs are i) They work well with high dimensional datasets. ii) Thanks to their accumulation of multiple learners with random structure they have better accuracy and less overfitting compared to simple decision trees. III) Ease of use due to their low dependence on parameter tuning (they achieve good results with minimal tuning). Their major disadvantage is that large RFs are computationally expensive and slow to train Choudhury et al., 2024.

Extra Trees (ET) Extra Trees (also called Extremely randomized trees) builds upon the foundation of Random Forests while adding additional randomness to the construction of their individual learners (Ahmad et al., 2018). Similar to RFs, they combine multiple individual trees that are trained with random subsets of features. However, instead of optimizing the split rules for each tree, for each selected feature, a split value is randomly selected from the feature's observed range (Geurts et al., 2006). This causes the model to have extreme randomization instead of the moderate one from random forests. Since there is no split optimization required, the computational requirements are lower (Geurts et al., 2006. Another difference with the Random Forest is that Extra Trees use the full dataset to train each learner (no bootstrap aggregating "bagging"). The model is still usually considered as being part of the bagging family since it creates multiple independent trees and aggregates their predictions through voting or averaging (Baladram, 2024). Extra Trees improve upon Random Forests with faster training and lower computational cost. These come from the random selection of decision rules instead of RFs' optimization. Their disadvantage lies in possible lower performance compared to RFs due to the random decision rules that might overlook optimal splits (Baladram, 2024).

3.4.2 Boosting

Boosting is a sequential ensemble technique that combines weak learners to form a stronger meta-model. Unlike bagging, where models are trained independently and in parallel on random subsets of the data, boosting builds models sequentially. Each new learner focuses on correcting the errors made by its predecessors (Mayr et al., 2014). This iterative process reduces bias by continuously refining the overall model, and can decrease variance through the ensemble effect (Nti et al., 2020). Although individual learners may perform poorly across the entire dataset, each is typically effective on different subsets of the data by contributing unique strengths to the general meta-model (Mayr et al., 2014).
The improvement of subsequent learners in boosting is guided by gradient descent techniques. This approach uses the gradient of the loss function to determine how to adjust the learner's parameters to minimize error (Friedman, 2001). This innovation allowed for more advanced and efficient boosting algorithms.
Two state-of-the-art gradient boosting algorithms will be used in this study: XGBoost, introduced in 2014, which enhances the original gradient boosting framework with regularization and system optimization for speed and scalability (T. Chen and Guestrin, 2016); and LightGBM, introduced by Microsoft in 2017, which builds on XGBoost's principles but employs two novel techniques that increase computational efficiency on large datasets by decreasing the amount of data points and features used to train the model (Ke et al., 2017).

XGBoost XGBoost for eXtreme Gradient Boosting, is an open-source machine learning system for tree boosting. It has become very popular in recent years with many machine learning competitions winning solutions using XGBoost either alone or in model ensembles (T. Chen and Guestrin, 2016). It is based on tree boosting, where decision trees are built sequentially and each improves upon previous learners mistakes. XGBoost further improves on tree boosting with several key innovations. First, a regularized objective function to improve generalization (T. Chen and Guestrin, 2016). Second, it employs second-order derivatives (Hessian) in addition to the first order derivatives usually employed in gradient descent for faster and more accurate loss minimization during tree construction (T. Chen and Guestrin, 2016). Third, it optimizes split-finding with a sparsity-aware algorithm able to deal with missing or zero values which improves efficiency in sparse datasets (T. Chen and Guestrin, 2016). Fourth, it uses algorithms that propose splits efficiently on large or distributed data (T. Chen and Guestrin, 2016). Together, these innovations

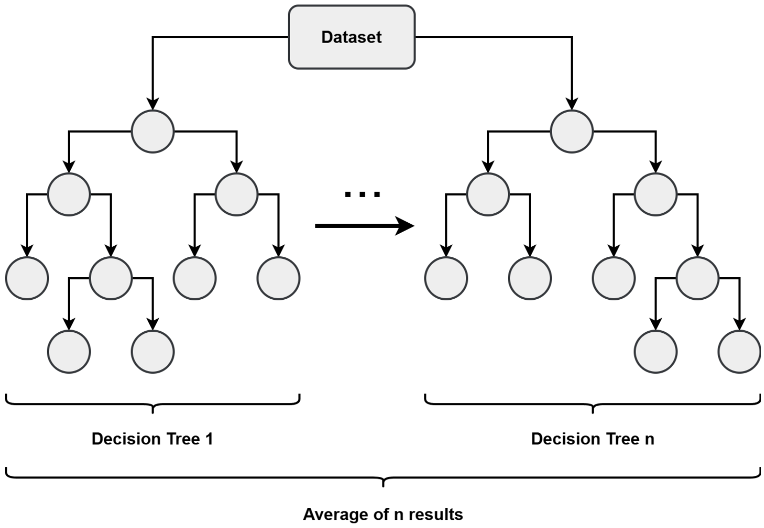


Figure 6: Diagram of gradient boosting algorithms. In gradient boosting, decision trees are produced sequentially. Each subsequent tree improves upon its predecessors.

make XGBoost faster, more scalable, and robust compared to traditional gradient boosting. The major disadvantage of XGBoost is its high dependence on parameter tuning. Attaining higher performances with this algorithm requires extensive knowledge over its multiple hyperparameters and their optimization for specific tasks (T. Chen and Guestrin, 2016; Choudhury et al., 2024).

LightGBM LightGBM is a modern gradient boosting framework that builds decision trees efficiently for large high dimensional datasets. Unlike traditional tree boosting methods that require scanning of all data points and features to find the best splits in each decision tree, LightGBM introduces two innovative techniques. First, Gradient-based One-Side Sampling (GOSS) keeps data with larger gradients that are expected to contribute more to information gain (Ke et al., 2017). This allows to drop instances with smaller gradients and focus on scanning only points that are most in need of correction. The method decreases the total amount of data processed while limiting accuracy loss thus increasing efficiency for large and high dimensional datasets (Ke et al., 2017). Second, it applies Exclusive Feature Bundling (EFB) that allows grouping together features that rarely take nonzero features at the same time (such as one-hot-encoded features) to decrease the number of features to consider (Ke et al., 2017). This algorithm's smart sampling and bundling strategies allow faster training and lower memory use while maintaining comparable performance to XGBoost (Ke et al., 2017). However, LightGBM's splitting strategy can produce highly complex decision trees thus resulting in overfitting (Choudhury et al., 2024). It is important to note that continuous improvements in the XGBoost and LightGBM libraries are leading to overlapping capabilities of both models. In fact, specific techniques used by LightGBM (such as the histogram-based method) are now also available in the XGBoost library.

3.5 Model Evaluation

Evaluating a machine learning model is crucial to determine its performance in predicting the target variable and if it is working correctly. There are two key sources of error in machine learning, "bias" and "variance", that impact the performance and generalization ability of a trained model. It is common, when using highly complex ML algorithms, for it to attain almost perfect performance on the data used for training. An algorithm with enough parameters, can simply learn all training data by heart and fit it perfectly, such a model has a high variance and is said to "overfit" to the data. The problem with such a model is that it also learns noise from the data and although it will perform well on the training data, performance will be very poor on unseen data (Emmert-Streib and Dehmer, 2019). The opposite situation exists where a model is too simple and cannot learn enough from the dataset for it to attain acceptable performance even in the data used to train it.

In this case, the model is said to have high "bias" and to "underfit" the data, since performance is low already in the training data, it will be low on unseen data as well (Emmert-Streib and Dehmer, 2019).
In this project, the models used are regression algorithms that predict continuous values and are evaluated mainly through the "coefficient of determination" also known as "R squared". R^2 is one of the most popular regression metrics (Chicco et al., 2021), it measures the proportion of variance of the target variable that is explained by the model after training. R^2 values usually range from 0 to 1, with 0 indicating that no variance is being explained by the model and it is not performing better than just predicting values randomly. The closest to 1, the better the regression model (e.g., 0.80 means that the model explains 80% of the variance). This metric was chosen due to its ease of interpretation and usefulness in comparing different models (Chicco et al., 2021).

3.6 Model Interpretation

Even if a model shows good performance and generalizes well to unseen data, out of the box we only get access to its predictions. This could be enough if the goal is purely to predict a value and we do not require to understand why the decision was made. However, there are many situations where pure predictions are not enough, such as when models are applied in sensitive domains (e.g., healthcare, finance) where transparency of the results is essential (Dwyer et al., 2018). Model interpretation methods allow to comprehend the decision-making process of machine learning models, specifically they allow to: identify the most influential features, understand the relationships learned by the models, and identify underlying patterns in the data. Moreover, if the explanations are in line with prior knowledge (e.g., known associations between variables), the model's predictions can be more easily trusted (Molnar, 2025). There are two types of interpretation methods: model-agnostic and model-specific.

3.6.1 Model-Agnostic Methods

Model agnostic methods are designed to work with any predictive model without relying on its internal structure or assumptions. They treat the model as a "black box" by only requiring inputs and outputs, which allows them to generate explanations (for instance, via local surrogate models like LIME or SHAP) for complex or opaque algorithms. This flexibility makes them broadly applicable across different types of models, although the explanations are often approximations based solely on observed behavior (Sim, 2022).
This work focuses on the use of SHAP values for interpreting the predictions of trained machine learning models. SHAP values are based on the concept of Shapley values from cooperative game theory, introduced by Lloyd Shapley in 1953 (Shapley, 1953). They were later popularized for machine learning in 2017 by Lundberg and Lee, with a rebranding from Shapley values to SHAP (SHapley Additive exPlanations), often termed SHAP values (Molnar, 2025). With this method, each feature in a model is considered a "player" in a game, and the prediction is the "payout" that has to be fairly distributed among the features. The SHAP value of a feature is computed by considering all possible combinations of features that exclude the feature in question. For each subset, we calculate the model's prediction with and without the feature, the difference between the predictions is the contribution of the feature. Finally, we average the contributions across all combinations of features. A positive SHAP value indicates that the feature increases the prediction compared to the average prediction. A negative SHAP value indicates that the feature decreases the prediction compared to the average one (Lundberg and Lee, 2017).

3.6.2 Model-Specific Methods

Model-specific methods, on the other hand, are tailored to a particular kind of model by leveraging its internal workings and structural characteristics. These methods use the inherent properties, such as coefficients in linear models or decision paths in tree-based models, to offer more precise and intrinsic explanations of the model's behavior. While this approach can yield more detailed insights, it is limited to models that expose such internal features and cannot be easily applied to models lacking them. Highly complex models such as neural networks cannot benefit from these methods, moreover the specificity of such methods to each model makes comparisons across models harder to provide which would make model-agnostic methods more desirable for comparison studies (Molnar, 2025).

4 Machine Learning in Psychological Studies

The following section will go through the multiple papers that were reviewed concerning the use of machine learning methods and algorithms in psychology studies. The aim is to observe what is usually done by researchers and identify the most promising algorithms for use in this study. Papers will be presented from oldest to most recent, with studies going from 2016 to 2025.

Ioannidis et al. (2016) used machine learning classification algorithms on a dataset composed of surveys from 2006 subjects. The authors used three types of algorithms (Logistic Regression, Random Forest, and Naïve Bayes) for the binary classification of problematic internet use. They used 50-replication cross-validation and observed that all three machine learning methods performed similarly (Ioannidis et al., 2016).

Kessler et al. (2016) tested a machine-learning algorithm to predict the persistence and severity of major depressive disorder. Their dataset consisted of psychometric data from 1056 respondents with DSM-III-R MDD. They used ensemble regression trees, penalized regression, and logistic regression in a classification task aiming at predicting MDD outcomes. They applied 10-fold cross-validation and found that machine learning models performed better compared to the logistic regression baseline (Kessler et al., 2016).

Dwyer et al. (2018) aimed to review the use of machine learning approaches for clinical psychology and psychiatry. Through their review article, where they evaluate algorithms across different types of datasets (MRI, EEG, clinical data, genetic, psychometric). The authors advocate for integrating machine learning into psychiatric research to overcome shortcomings of traditional methods (e.g., overfitting, limited generalizability, and small effect sizes). They highlight the need for the generalizability of machine learning algorithms to be assessed seriously while applying the method (Dwyer et al., 2018).

Srividya et al. (2018) classified mental health status (distressed, neutral, happy) on a psychometric dataset composed of 656 participants. They assessed a wide variety of machine learning algorithms (Support Vector Machine, Decision Tree, Naïve Bayes, K-Nearest Neighbors, Logistic Regression, Random Forest). For their splitting strategy, they separated the dataset into 80-20 train-test split, and used LIME for interpretation of the models. They reported the Random Forest, K-Nearest Neighbors, and Support Vector Machine as the highest performing classifiers (Srividya et al., 2018).

Watson et al. (2019) explored how ML can revolutionize clinical practice. They emphasize the advantage of ML on their complex pattern recognition and ability to learn complex datasets. However, the more complex ML models are, the harder it is to understand how they are making their decisions and this opacity challenges trust and informed consent. Finally, the authors highlight the existence of explainability methods that can be used to extract information on the decision making of complex ML models. These methods, such as SHAP and LIME allow to use highly accurate models, while maintaining a degree of understanding on their internal functioning (Watson et al., 2019).

Christodoulou et al. (2019) reported, through a systematic review, no performance benefit of machine learning over logistic regression for classification of clinical data. The authors compared Classification Trees, Random Forests, Neural Networks, Support Vector Machines, and Ensemble methods against logistic regression, however they failed to surpass baseline performance when validation was unbiased (Christodoulou et al., 2019).

Cho et al. (2019) reviewed machine learning algorithms in a classification task for the diagnosis of mental illness (PTSD, schizophrenia, depression). Multiple types of datasets were analyzed (psychometrics, EEG, MRI, fMRI, audio) through a variety of models (SVM, Random Forest, Gradient Boosting Machine, KNN, Naïve Bayes). They found that SVM and ensemble models performed best among the models (Cho et al., 2019).

Orrù et al. (2020) tested a great number of ML algorithms in a classification task to detect malingering (i.e., faking a survey's answers) on a psychometric dataset. They applied a stratified

10-fold cross-validation with a 20% holdout test set to ensure that the final results were representative of the generalization capability of the model. The authors reported that ML can effectively complement traditional statistics with ensemble models performing well with minimal overfitting. The authors suggest that Ml may be used to maximize accuracy and minimize replicability issues present with traditional statistics. They posit that the model agnostic (they do not rely on assumptions about data generation) nature of certain ML algorithms may help increase replicability (Orrù et al., 2020).

Conversely, Jacobucci and Grimm (2020), warn about the risk of measurement error on machine learning models. They reported through a simulated regression task, that linear regression may perform superiorly to higher complexity models such as a Gradient Boosting Machine, if the data has a high degree of measurement error. They conclude by highlighting that psychometric data is subject to participant errors (e.g., cognitive fatigue, malingering) and thus may be a better fit for simpler models that do not learn noise as easily (Jacobucci and Grimm, 2020).

Even with these risks, multiple psychological studies have been successful with complex ML algorithms in the classification of psychometric data. Tate et al. (2020) reported higher performance for Random Forest compared to a Logistic Regression baseline in the prediction of adolescent mental health problems (Tate et al., 2020). While Santana et al. (2020) observed higher performance of ensemble models (e.g., Extra Trees, XGBoost) against simpler models (e.g. Logistic Regression, SVC, KNN) (Santana et al., 2020). More recently, Ghorpade-Aher et al. (2023) observed high performance of AdaBoost and Support Vector Machine on the classification of severity levels of DASS-21 subscales (Depression, Anxiety, Stress) (Ghorpade-Aher et al., 2023). Later, Tian et al. (2024) reported that XGBoost attained high performance on the classification of psychological distress although slightly surpassed by a KNN model (Tian et al., 2024).

In summary, the reviewed literature highlights ensemble models, particularly those based on bagging (Random Forest, Extra Trees) and boosting (XGBoost, LightGBM) as having a high potential to surpass simpler models in psychometric data. When appropriately tuned, these models offer robustness against overfitting, handle complex non-linear relationships well, and are less reliant on strict statistical assumptions compared to traditional methods. While most studies reviewed focused on classification tasks, this study will instead produce a regression pipeline. This will allow better leverage over the continuous nature that some psychometric instruments have such as the DASS-21 (Lovibond and Lovibond, 1995) and the CTPI (Denovan and Dagnall, 2019). To allow the interpretation of these complex models and study the interaction between digitalization, wellbeing and time pressure, SHAP was selected as a model-agnostic method.

5 Research Questions and Hypotheses

The gaps in literature concerning the use of regression ML algorithms for prediction of continuous survey scores and the potential of ensemble methods led to the first research question.

Q1. How do tree-based machine learning regression models compare to linear regression on the prediction of psychological distress and time perception psychological scores in a psychometric multicultural dataset from the TIMED project ?

- H1.1: Regression models using bagging approaches, such as Random Forest, will surpass the performance of linear regression in predicting scores related to psychological distress and time perception.

- H1.2: Regression models based on boosting, such as XGBoost, will demonstrate superior performance compared to linear regression in predicting psychological distress and time perception scores.

The second research question was motivated by the great duality of digital technology use on both mental well-being and time pressure. Moreover, the proposition that cultural factors may influence how digitalization affects mental health and time experience.

Q2. How does digitalization affect psychological distress and time perception of European individuals ? Do these effects vary according to the cultural context ?

- H2.1: The problematic internet use score (PIUQ-9, Demetrovics et al., 2008) is a positive predictor of the sum of psychological distress subscales (DASS-21, Lovibond and Lovibond, 1995).
- H2.2: The mean quality of digital experience score (QDES, Witowska et al., 2025) is negatively correlated with the mean chronic time pressure score (CTPI, Denovan and Dagnall, 2019).
- H2.3: The predictors of the sum of psychological distress subscales (DASS-21, Lovibond and Lovibond, 1995) and mean chronic time pressure score (CTPI, Denovan and Dagnall, 2019) vary significantly depending on the country where the data were collected.

Due to the exploratory nature of applying multiple ML algorithms to a dataset (Tian et al., 2024), it is hard to determine accurate hypotheses. In this situation, it would have been interesting to just explore the interactions of different features with the target variables. However, with the aim of focusing this thesis, potential interactions were extracted from the literature between features present in the dataset and the target variables (DASS-21 Sum and CTPI Mean).
Since this project revolves around the utilization of machine learning algorithms on a psychology dataset, both research questions are highly complementary. Indeed, Q1 focuses on the comparison of multiple machine learning algorithms with the aim of finding one that is able to predict the target variables with sufficient accuracy which is required to justify the interpretations obtained from a trained model. These interpretations will then be used to answer the three hypotheses of the second research question.
In order to focus the hypotheses in a clear goal it was chosen to target scores that aggregate different subscales of a survey instead of predicting each individual subscale. Our two targets are I) "DASS Sum" which is the sum of the three different DASS-21 subscales "Depression", "Anxiety" and "Stress", and II) "CTPI Mean" which is the mean of the different CTPI subscales "Feeling harried" and "Cognitive awareness of time shortage". By predicting aggregated scores, we limit redundancy in the analysis of results and interpretations. Since the subscales are highly correlated with each other (being part of the same survey), predicting each individually would risk increasing redundancy of the results interpretation.

6 Methods

This section outlines the methodological approach used throughout the study to investigate the predictive relationships between digitalization and psychological outcomes using machine learning. It starts with a description of the dataset used in the study, including its origin, structure, and key variables it contains. It goes on to present the procedure (ethics and data collection), data cleaning and processing steps that were carried out to prepare the dataset for analysis. The following sections cover preparation of task-specific datasets, the configuration of prediction tasks, the machine learning pipeline for model training and optimization, and the evaluation framework used to compare model performance and model interpretation.

6.1 TIMe experience in Europe's Digital age (TIMED)

The TIMED project seeks to methodically evaluate how digital practices affect the perception, use, and allocation of time, and how it impacts the quality of life. The project is led by Dr. Ruth Ogden (Liverpool John Moores University) and involves an interdisciplinary team of psychologists, sociologists, neuroscientists, and computer scientists from six European countries: the United Kingdom, Germany, Switzerland, Spain, Poland, and the Czech Republic. This diverse collaboration allows for cross-cultural comparisons, making TIMED the first empirical mapping of the "Geography of European Time" in the digital age.
In terms of methodology, TIMED is organized into five work packages (WPs) that incorporate psychophysiological evaluations, quantitative surveys, qualitative interviews, and real-time behavioral monitoring through mobile technologies.
This thesis is based on the dataset collected in Work Package 2 (WP2), which is composed of

demographic information, measures related to time, measures of wellbeing, and measures of digital practice. WP2 gathered data from over 7,000 participants across the six participating countries using standardized psychometric instruments and measures of DT use developed in WP1. The instruments from WP1 are: "Immersion in Digital Life Scale (IDLS)" which measures the extent to which digital technology is integrated into everyday life (with higher values indicating a higher integration of DT into the respondent's life); and "Quality of Digital Experience Scale (QDES)" that assesses perceived quality of life as influenced by use of digital technology.

6.2 Participants

6.2.1 Demographics

The processed WP2 dataset contains the answers to the WP2 survey of 7,519 respondents, collected from six European countries. As reported in table 1, the age of participants ranges from 18 to 92 years (M = 45.73, SD = 15.70). The sample is geographically balanced, with each country contributing approximately one-sixth of the total responses: Poland (19.36%), Switzerland (16.27%), Germany (16.25%), Spain (16.12%), Czechia (16.04%), and the United Kingdom (15.96%). Gender distribution is nearly even, with 48.74% identifying as male and 50.72% as female, while a small proportion (0.53%) identified as another gender or declined to answer. The education level is varied, with 38.66% being at the university level, 33.67% having education below degree level, 25.44% possessing technical qualifications, and 2.22% reporting another form of education or withholding the information. Overall, the demographics are varied and well-balanced, reflecting the efforts made so that each country's sample represented was representative of its population.

6.2.2 Psychometric Profile

After pre-processing the overall survey contains 14 psychometric instruments that will be detailed in 6.4. These instruments can be separated into three types: digital technology use measures, time-related measures, and wellbeing measures. Descriptive statistics for each group are presented in tables 2, 3, and 4, respectively. Digital technology use measures revealed that participants exhibited a normal distribution for immersion in digital life, with most reporting a moderate level of DT pervasiveness. Problematic internet use scores were generally low, with only 38.28% of the sample exceeding the threshod of 22, as proposed by several authors (Koronczai et al., 2011; Laconi et al., 2018; Vally et al., 2021). Additionally, respondents tended to report a generally positive experience with digital technology across multiple dimensions. Time-related measures indicated that most participants leaned towards a faster experience of time and low estimated free time. In contrast, both time pressure and work-life balance scores were normally distributed among participants. Wellbeing measures reported overall low levels of depression, anxiety, and stress, with stress scoring slightly higher than the other two indicators of psychological distress. Regarding procrastination, 75% of participants scored below Steel's threshold of 3.5 for pure procrastination. This suggests that the majority of the sample does not engage in frequent procrastination (Steel, 2010)

Table 1: Demographics (N = 7,519)

Feature	Level / Statistic	Value	n	%
Age	Mean (SD)	45.73 (15.70)	7,519	–
Country	(1) United Kingdom	–	1,200	15.96%
	(2) Switzerland	–	1,223	16.27%
	(3) Spain	–	1,212	16.12%
	(4) Poland	–	1,456	19.36%
	(5) Czechia	–	1,206	16.04%
	(6) Germany	–	1,222	16.25%
Education	(1) Below degree	–	2,532	33.67%
	(2) Technical qualifications	–	1,913	25.44%
	(3) University	–	2,907	38.66%
	(4) Other	–	167	2.22%
Gender	(1) Male	–	3,665	48.74%
	(2) Female	–	3,814	50.72%
	(3) Other	–	40	0.53%

Percentages may not sum to 100% due to rounding.
Numeric codes in parentheses match original category coding.

Table 2: Digital Technology Use Measures (Mean [SD, Min–Max], Quartiles; N = 7,519)

Scale / Subscale	Mean (SD) [Min–Max]	Q1	Q2	Q3
Immersion in Digital Life (Witowska et al., 2025)				
Overall Score	49.47 (21.05) [0–100]	34.6	50.8	64.4
Quality of Digital Experience (Witowska et al., 2025)				
Overall Score	3.41 (0.70) [1–5]	3.00	3.46	3.88
Health & Wellbeing	3.36 (0.82) [1–5]	2.80	3.40	4.00
Social Connectedness	3.16 (0.85) [1–5]	2.58	3.25	3.83
Time & Efficiency	3.78 (0.72) [1–5]	3.44	3.89	4.22
Problematic Internet Use (Laconi et al., 2019)				
Overall Score	20.09 (7.19) [9–45]	14	19	25

Table 3: Time Measures (Mean [SD, Min–Max], Quartiles; N = 7,519)

Scale / Subscale	Mean (SD) [Min–Max]	Q1	Q2	Q3
Passage of Time (Ogden, 2020)				
Overall Score	70.19 (16.60) [2.6–100]	58.2	70.8	82.2
Time Expansion (Wittmann and Lehnhoff, 2005)				
Overall Score	1.49 (0.81) [0–4]	1.00	1.40	2.00
Chronic Time Pressure (Denovan and Dagnall, 2019)				
Overall Score	3.08 (0.69) [1–5]	2.62	3.08	3.62
Feeling Harried (affective)	3.05 (0.77) [1–5]	2.60	3.00	3.60
Cognitive awareness of time shortage (cognitive)	3.09 (0.74) [1–5]	2.50	3.12	3.62
Time Autonomy (Macan, 1994)				
Overall Score	3.26 (0.78) [1–5]	2.80	3.20	3.80
Work–Life Balance (Brough et al., 2014)				
Overall Score	3.28 (0.88) [1–5]	2.75	3.25	4.00
Temporal Focus (Shipp et al., 2009)				
Past	4.14 (1.22) [1–7]	3.00	4.00	5.00
Present	4.70 (0.94) [1–7]	4.00	4.75	5.25
Future	4.46 (1.15) [1–7]	3.75	4.50	5.25
Free Time Availability				
Overall Score	43.29 (22.73) [0–100]	28	40	60

Table 4: Wellbeing Measures (Mean [SD, Min–Max], Quartiles; N = 7,519)

Scale / Subscale	Mean (SD) [Min–Max]	Q1	Q2	Q3
WHO-5 Wellbeing (World Health Organization, 1998)				
Overall Score	13.02 (5.19) [0–25]	9	13	17
Satisfaction With Life (Diener et al., 1985)				
Overall Score	20.46 (6.76) [5–35]	16	21	25
Depression Anxiety Stress (Lovibond and Lovibond, 1995)				
Overall Score	17.73 (13.95) [0–63]	6	14	27
Depression	6.19 (5.45) [0–21]	2	5	10
Anxiety	4.63 (4.72) [0–21]	1	3	7
Stress	6.91 (4.93) [0–21]	3	6	10
Pure Procrastination (Steel, 2010)				
Overall Score	2.45 (0.89) [1–5]	1.75	2.42	3.08

6.3 Procedure

6.3.1 Ethics

The TIMED project is led by Dr. Ruth Ogden (Liverpool John Moores University) as project leader (PL) and coordinated by specific principal investigators for each of the participating countries. The WP2 of the TIMED project was approved by the Etics Committee of the Department of Psychology, University of Fribourg (Switzerland; 2022-795-R1) and the Internal Review Board of John Moore Liverpool University (United-Kingdom; UREC 23/PSY/071). The study did not present any physical, mental, or social risks to the participants. An online information sheet was provided in Qualtrics before start of the survey which informed of possible withdrawal at any time without negative consequences. Under the Swiss protocol, participants were able to request deletion of data within two weeks of survey completion. No compensation was provided by the research team for WP2 and investigators reported no conflicts of interest.

6.3.2 Data collection

WP2 was conducted as a fully anonymous online questionnaire hosted through Qualtrics across six countries (UK, Switzerland, Spain, Germany, Poland, Czechia). Criteria for participant eligibility were they were at least 18 years old, lived in the country, and spoke the language of the survey fluently. Samples were prepared to be representative of each country's general population in terms of age, gender, and education. The original dataset included 7,536 respondents and 46 variables. After data cleaning and processing, described in 6.6, 7,519 respondents remained and the feature set was reduced to 37 variables. The survey captured socio-demographic information and measures related to digital technology use, time perception, and wellbeing. The individual questionnaires used throughout the survey will be detailed in the following section.

6.4 Psychometric Instruments

6.4.1 Psychological Scores

The WP2 dataset is mostly composed of psychometric scores, which can be categorized into three domains: digital technology use, time perception, and well-being. A brief overview of these scores is provided in the following section to offer a general understanding of the dataset's structure and content.

Digital Technology Use Measures

Immersion in Digital Life Scale (IDLS): This score is obtained through a scale produced in TIMED's WP1. It measures the extent to which daily activities (socializing, leisure, time management) are conducted digitally (Witowska et al., 2025).

Quality of Digital Experience Scale (QDES): Also produced from TIMED's WP1, it assesses the respondent's perception of how digital technology improves quality of life. It specifically measures the impact of DT on three dimensions: mental health, relationships, and productivity. In addition to the overall score (mean of all item scores), three subscales are also present in the dataset (Health and well-being, time and efficiency, and social connectedness) (Witowska et al., 2025).

Problematic Internet Use (PIUQ): The score derives from the PIUQ-9 questionnaire Laconi et al., 2019. It assesses the severity of excessive or uncontrolled Internet use which may interfere with daily life, psychological well-being, and self regulation (Demetrovics et al., 2008; Laconi et al., 2019).

Time Measures

Passage of Time (PoT): The questionnaire is based on items from multiple sources aimed to measure the perceived speed of time in the present, past, future, and during digital technology use (Ogden, 2020; Wittmann and Lehnhoff, 2005).

Time Expansion (Texp): It corresponds to the time expansion subscale of the subjective time experience scale (Wittmann and Lehnhoff, 2005). Time expansion refers to a subjective experience where time feels as though it is moving slowly or dragging. It is often associated with boredom, lack of stimulation, and having too much unstructured or unused time (Wittmann and Lehnhoff, 2005).

Chronic Time Pressure (CTPI): This score derives from the Chronic Time Pressure Inventory developed by Denovan and Dagnall (Denovan and Dagnall, 2019). It measures the subjective experience of persistent time pressure (i.e., not having enough time to meet demands even in the absence of immediate stressors). It is complemented by two subscales, each representing one factor of time pressure: cognitive and affective (Denovan and Dagnall, 2019).

Time Autonomy (TA): Corresponds to the subscale from the Time Management Behavior scale called "perceived control over time". It measures perceived control over how one's time is spent, including self-organization and resistance to interruptions (Macan, 1994).

Work-Life Balance (WLB): The score originates from a questionnaire by Brough et al. in 2014. It assesses the respondent's subjective perceptions of their balance between work and non-work roles (Brough et al., 2014).

Temporal Focus (TFS): The TFS questionnaire was developed by Shipp et al. (2009). It measures the degree to which respondents allocate attention to the past, present, and future time frames (Shipp et al., 2009). This questionnaire does not include an overall score; instead, it provides three subscales that reflect the allocation of attention to the past, present, and future.

Free Time Availability (FreeTime): This score derives from a single-item visual analogue scale created by Ruth Ogden for the TIMED project. It represents the proportion of a respondent's time that they perceive as "free time".

Well-being Measures

WHO-5 Well-being (WHO): A widely used measure of subjective psychological well-being, published by the World Health Organization in 1998. Its measure encompasses overall emotional state and interest in everyday life (Topp et al., 2015).

Satisfaction with Life (SWLS): The SWLS was developed by Diener et al. (1985) and measures the satisfaction that individuals feel about their life as a whole (Diener et al., 1985).

Psychological Distress (DASS): The DASS was developed by Lovibond and Lovibond and measures the severity of three negative emotional states (depression, anxiety, and stress) (Lovibond and Lovibond, 1995). It can also be used and serves in this study as a measure of overall psychological distress.

Pure Procrastination (PPS): This score derives from the Pure Procrastination Scale developed by Piers Steel (2010). Assesses pure procrastination by using items that directly refer to the core behavior of voluntary delay despite the negative consequences (Steel, 2010; Svartdal and Steel, 2017).

6.5 Analysis Tasks

This section presents the two prediction tasks central to this study: DASS and CTPI. Both are regression problems which is motivated by the lack of validated cutoffs for CTPI and the literature gap in use of ML for regression in psychology. Moreover, as noted by Denovan and Dagnall, 2019, chronic time pressure is best evaluated using a dimensional rather than a categorical approach. While categorical classifications can help identify psychological disorders, not all constructs are suited for such binary distinctions (Kessler, 2002). Constructs that are inherently dimensional may be more accurately modeled using regression techniques.

6.5.1 DASS Task

The first regression task uses demographic features and psychological scores from the dataset to predict overall DASS-21 scores as a measure of psychological distress which are computed by summing the three DASS subscales (Depression, Anxiety, and Stress). Although all previously presented predictors (other than DASS subscales) are used in to train ML algorithms in this task (to improve performance), the main predictor that is observed (hypothesis 2.1) is problematic

Internet use (PIUQ). This task aims at observing the effects of digitalization, specifically problematic Internet use, on psychological distress and observe whether ML models can identify previously demonstrated positive associations.

Target - DASS: DASS-21 is the short-form of the DASS-42 questionnaire published by Lovibond and Lovibond (1995). The questionnaire is composed of 21 items, each subscale is assigned seven items and each item is scored from zero (did not apply to me at all) to three (applied to me very much or most of the time). Each subscale score ranges from 0 to 21, while the overall score ranges from 0 to 63. Additionally, the DASS measure is based on a dimensional rather than a categorical conception of psychological disorder, although standardized cutoffs were later developed (Lovibond and Lovibond, 1995). Each subscale measures the severity of a negative emotional state (depression, anxiety, and stress) and the overall score hints at a general measure of psychological distress.

Predictor - PIUQ: PIUQ-9 is a shortened form of the original 18-item PIUQ instrument. PIUQ-18 was published by Demetrovics et al. in 2008, while the shortened 9-item version was proposed by Laconi et al. in 2019 (Demetrovics et al., 2008; Laconi et al., 2019). The PIUQ-9 instrument consists of nine items that measure three key aspects of problematic Internet use: obsession (preoccupation with online activities when offline), neglect (neglecting daily tasks, relationships, and responsibilities due to Internet use), and control disorder (difficulty in controlling or decreasing time spent online). Each item is rated on a 5-point scale, ranging from one (never) to five (always or almost always). The overall score ranges from nine to 45 points and measures the severity and nature of problematic Internet use that interferes with daily life, psychological well-being, and self-regulation. A moderate positive correlation was previously identified with psychological distress measured through the Global Severity Index (GSI) (Laconi et al., 2019).

6.5.2 CTPI Task

The second regression task uses demographic and psychological features to predict the overall CTPI score, which is computed as the mean of all CTPI items. The main predictor of interest for this task (hypothesis 2.2) is the quality of digital experience (QDE), a novel measure from TIMED's Work Package 1 (Witowska et al., 2025). This task aims to observe how perceived quality of life provided by digital technology use affects sustained time pressure. Due to the novelty of QDE's development, there is a lack of literature on its associations which this study aims to address.

Target - CTPI: The Chronic Time Pressure inventory is a 13-item developed by Denovan and Dagnall (2019) that measures the subjective experience of persistent time pressure. Each item is scored in a 5-point scale ranging from one (strongly disagree) to five (strongly agree). It comprises two factors: feeling harried and cognitive awareness of time shortage which are reflected as the two subscales of the construct (Denovan and Dagnall, 2019). Feeling harried taps into the emotional and affective experience of being rushed (e.g., I feel rushed to do things I have to do). The cognitive awareness of time shortage focuses on the recognition that there is not enough time (e.g., I don't have enough hours in the day) (Denovan and Dagnall, 2019). These factors represent the affective and cognitive components of chronic time pressure (Denovan et al., 2023; Szollos, 2009).

Predictor - QDES: The Quality of Digital Experience Scale is a newly developed instrument by the TIMED consortium during its first Work Package. It is a 26-item questionnaire that measures the respondent's perceived quality of life as influenced by digital technology use. The measure consists of three dimensions: health and well-being (how DT supports psychological health and well-being), time and efficiency (how DT improves productivity and saves time), and social connectedness (how DT fosters social connection and relationships). Items are scored in a 5-point likert scale that ranges from one (strongly disagree) to five (strongly agree). The overall QDES score is the mean of all item scores and ranges from one to five. The subscale scores behave similarly as they are means of their respective items.

6.6 Data Cleaning and Processing

This section outlines the data cleaning and processing steps that were applied to the original WP2 dataset. Prior to this study, the TIMED team collected the dataset from Qualtrics and applied preliminary preprocessing to it. This included the removal of non-essential variables such as start date, end date, and survey language. Participants who failed attention check questions were excluded to limit measurement error from respondents that are not fully engaged. Questionnaire items were recoded and reverse-scored where necessary to reflect the meaning of the constructs. Total scores for each questionnaire and their subscales (either sums or means), and most individual items were discarded to emphasize overall scale effects. In addition, a z-scored version (ZSWLS) of one scale (SWLS) was produced to address the inconsistency of Spain using a five-item scale, while the other countries used a seven-item scale. The resulting dataset was exported as a SPSS file, and for the purposes of this study, was later converted into CSV format for analysis using Python.

6.6.1 General Processing

The initial dataset contained responses from 7,536 participants across 46 variables. As a first step, two metadata features, participant ID and survey completion time, were removed, as they did not carry informative content. This reduced the dataset to 44 features while retaining all 7,536 participants.

Subsequent analysis identified 17 duplicate entries, grouped into nine distinct sets. Although the duplicates had different participant IDs and response times, they showed identical values across all demographic and psychological variables. Given the proximity in submission times, these were interpreted as repeated submissions from nine distinct individuals. As they did not contribute new information, the duplicates were removed, reducing the number of participants to 7,519.

Next, the dataset was examined for missing values, as these cannot be used directly when training machine learning models. Features with more than 20% missing data were excluded, as imputing a large number of values can introduce noise and bias. Seven features had missing values: six exceeded the 20% threshold and were removed, while one had only a single missing value. The excluded features included two text description response variables: Gender_5_TEXT with 99.93% missing, and Education_9_TEXT with 99.55% missing. There were four psychological measures: CTQ_Sum, CTQ_Abuse_Sum, CTQ_Neglect_Sum (each with 35.4% missing), and RR_Sum (83.72%). The Childhood Trauma Questionnaire (CTQ) overall and subscale scores were entirely missing for participants from Spain and Poland. The Reward Responsiveness (RR_Sum) scores were only available for Swiss participants (83.72% missing). In both cases, the missing values exceeded the threshold and were removed to prevent cultural bias. One missing value in the fifth Passage of Time item (PoT_5) was imputed using the mean of the remaining items in the same scale. After these steps, the dataset consisted of 7,519 participants and 38 features.

The next stage involved the removal of the remaining individual item responses for two scales: the Immersion in Digital Life Scale (IDLS) and the Passage of Time scale (PoT). Since other scales did not provide individual items and aggregated scores were available, the five individual items from each scale were removed. This reduced the dataset to 28 features.

An inconsistency was identified in the Satisfaction With Life Scale (SWLS), where the Spanish scale consisted of only five items (range: 5-25), while other countries used seven items (range: 5-35). A z-scored version (ZSWLS) was provided to address this issue, however since it was calculated using the full dataset, it could result in data leakage during model training. Therefore, ZSWLS was discarded, and the existing SWLS scores were modified by rescaling the Spanish data to the [5, 35] range using the following linear transformation $x' = \frac{(x-5)}{20} * 30 + 5$. The formula allows for rescaling the maximum to 35 while keeping a minimum of five and is applied exclusively to Spanish respondents. After this transformation and the removal of ZSWLS, the dataset includes 27 features.

Three categorical features were identified in the dataset: Country, Gender, and Education. The Country feature consisted of six balanced groups (UK, Switzerland, Spain, Poland, Czechia, Germany). Gender originally included five categories: Male, Female, Non-binary/Transgender, Other, and Unknown. Male and Female account for 99.46% of all respondents. Although the last three categories only represent 0.54% of the participants, they were retained and merged into a single "Other" category to preserve potentially valuable psychological information while limiting the increase in sparsity. A similar method was applied to the Education feature. Among its six original categories (Below degree, technical qualifications, university, professional degree, other, unknown) the last three account for only 2.23% of all data and were similarly merged into an "Other" group. Finally, the categorical features have to be appropriately encoded to improve the training of ML algorithms. Initially, they were "label-encoded" with each group being assigned an unique number. This method works best when the categories are ordered, which is not the case for Country or Gender. The Education groups ordering is solid for the first three groups but confusing for the "Other" group. The features were instead encoded with "one-hot encoding" that replaces each categorical variable with a set of binary features easily understood by ML models. Following this encoding step, the final general dataset comprised 7,519 respondents and 37 features, which formed the foundation for producing task-specific datasets.

6.6.2 Task-Specific Processing

To enable the prediction of two distinct target variables, task-specific datasets were derived from the general processed WP2 dataset. This step required the removal of features that could lead to target leakage (i.e., features that are highly correlated with the target and would not be available in real-world prediction scenarios). For the psychological distress (DASS) prediction task, the three subscale scores (DASS_Depression, DASS_Anxiety, and DASS_Stress) were excluded. In their place, a total DASS score was computed by summing the three subscales, as this aggregated score was not originally included in the dataset and serves as the prediction target for the task. Similarly, for the chronic time pressure (CTPI) prediction task, the two subscales (CTPI_Harried and CTPI_Cognitive) were removed to prevent leakage as they were used to compute the overall score initially. The total DASS score was not used in the CTPI task, since it was not present in the original data and its subscales already contain most of its information. As a result of these adjustments, both the DASS and CTPI task-specific datasets consist of 7,519 respondents and 35 features each.

6.7 Analysis Structure

This section outlines the key methodological steps structuring the present study. It starts with an overview of the project, followed by the processes of model optimization and comparison. Afterwards, it details how models are interpreted to examine the influence of individual predictors. Finally, it describes the analysis of cross-country variations in predictor values.

6.7.1 Study Overview

This study investigates the prediction of psychological outcomes using machine learning models applied to a psychometric-based dataset in regression tasks. The key steps of the analysis are outlined in figure 7. After initial data cleaning and processing, which were described in section 6.6, the processed WP2 data was adapted into two task-specific datasets: one for predicting overall psychological distress (DASS and the other for predicting chronic time pressure (CTPI).

Both datasets were split into training and test sets using an 80-20 split with a fixed random state for reproducibility. The four previously selected models (Random Forest, Extra Trees, XGBoost, and Light GBM) are each tuned for 50 optimization trials using the Optuna algorithm. Each trial is evaluated on a 5-fold cross-validation of the training set, once all trials of a model have been completed, Optuna returns the model with the best average performance across the validation folds. The Linear Regression (LR) model does not require optimization and so it is directly evaluated on the validation folds without passing through Optuna.

Once all models have been optimized we compare, for each task dataset, the best performing model for each family of algorithms (Bagging and Boosting). The selected models are then tested against the Linear Regression baseline through a 5x2cv combined F-Test to determine if they significantly outperformed the traditional method. This step addresses the first set of hypotheses (1.1 and 1.2) regarding the superiority of ensemble models over Linear Regression.

Afterwards, the best overall model for each task is used to interpret how predictors of interest (PIUQ for DASS and QDE for CTPI) contribute to the predictions. SHAP values are computed on the test set to assess the each feature's contribution to the decision making process of the model. SHAP values allow for an evaluation of the direction and magnitude of the feature's influence which addresses hypotheses 2.1 and 2.2.
Finally, to assess cross-country variations (hypothesis 2.3), the task-specific datasets (DASS and CTPI) are further separated into subsets for each country. The subsets are further split into training (80%) and testing (20%) sets. The best model for each task is trained of the training subsets and SHAP values are computed in the test subsets. The variation in feature importance is quantified for each country using the mean absolute SHAP values and difference between countries is assessed through one-way ANOVA.

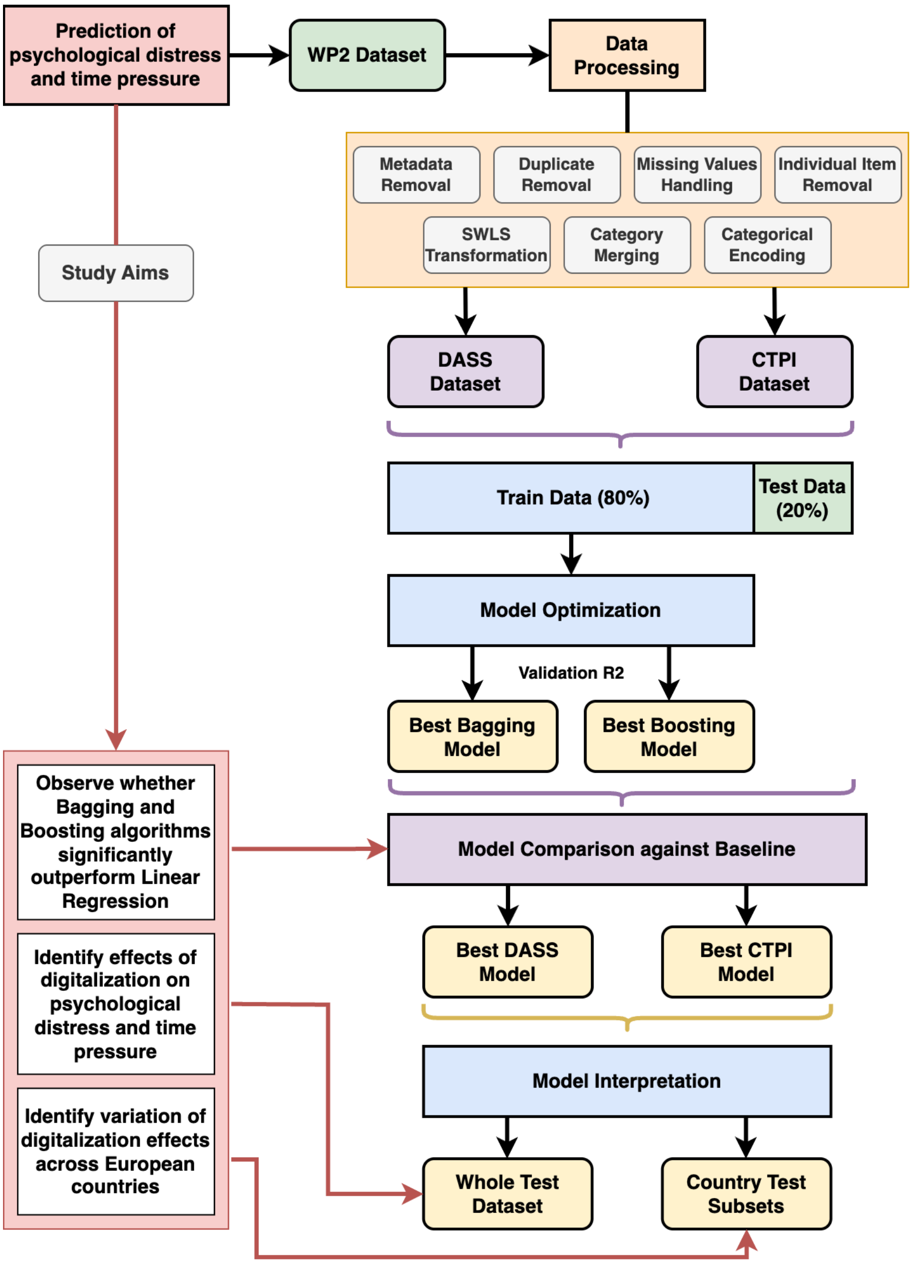


Figure 7: Study workflow diagram outlining the main steps of the machine learning pipeline used for comparison and interpretation of models. The key research goals are outlined and associated with their corresponding steps.

6.7.2 Data Splitting

Each task-specific dataset was divided into two sets: 80% for training and 20% for testing as presented in figure 7. The training set was utilized for performing 5-fold cross-validation during model optimization, allowing for robust estimation of performance and hyperparameter tuning. The test set was held out and used exclusively for significance testing of models against the baseline, and final evaluation and interpretation of the best-performing models. This gives an unbiased assessment of generalization performance. Data splitting was performed using scikit-learn's *train_test_split* function with a fixed *random_state* of 42 to ensure reproducibility across experiments.

6.7.3 Model Optimization

Machine Learning algorithms have hyperparameters, a property of the algorithm that influences how it works. These are not learned from data by the algorithm, instead having to be set by the data analyst before running it. They can greatly impact the overall performance, stability, and generalization ability of the final model (Burkov, 2019). Complex algorithms such as those chosen for this study (Random Forest, Extra Trees, XGBoost, Light GBM) have many hyperparameters and have to be optimized to attain their full potential. Linear Regression, however, does not have hyperparameters thus no optimization is required.

When optimizing hyperparameters, data scientists have to choose a hyperparameter search space defining the range of possible values that can be assigned to each hyperparameter. These spaces are often very large which requires the use of hyperparameter optimization algorithms that automatize the process. For this study, a hyperparameter optimization framework called Optuna was integrated inside the python machine learning pipeline. The framework allows users to specify the number of optimization trials, during which a machine learning model is repeatedly trained and evaluated. In each trial, Optuna selects a new set of hyperparameter values with the aim of improving model performance. It employs a smart search strategy where promising hyperparameter values are sampled using the Tree-structured Parzen Estimator (TPE) based on previous trial results. To reduce computation time, Optuna uses a pruning mechanism that stops trials that perform poorly early (Akiba et al., 2019).

The optimization steps are outlined in figure 8. Each trial runs the model through a 5-fold cross-validation where the training set is further divided in four training splits and one validation split, five times. The performance metric for optimization is the mean squared error of the model on the validation splits. The first trial selects hyperparameters randomly, however, subsequent trials choose promising hyperparameter values based on the performance of hyperparameters on previous trials. In this study, we used 50 trials of optimization for each algorithm which allows balancing performance and computational resources. Once all trials are completed, the trial having the best performance (i.e., lowest averaged mean squared error on the validation splits) is returned and its hyperparameters are used as the optimized model.

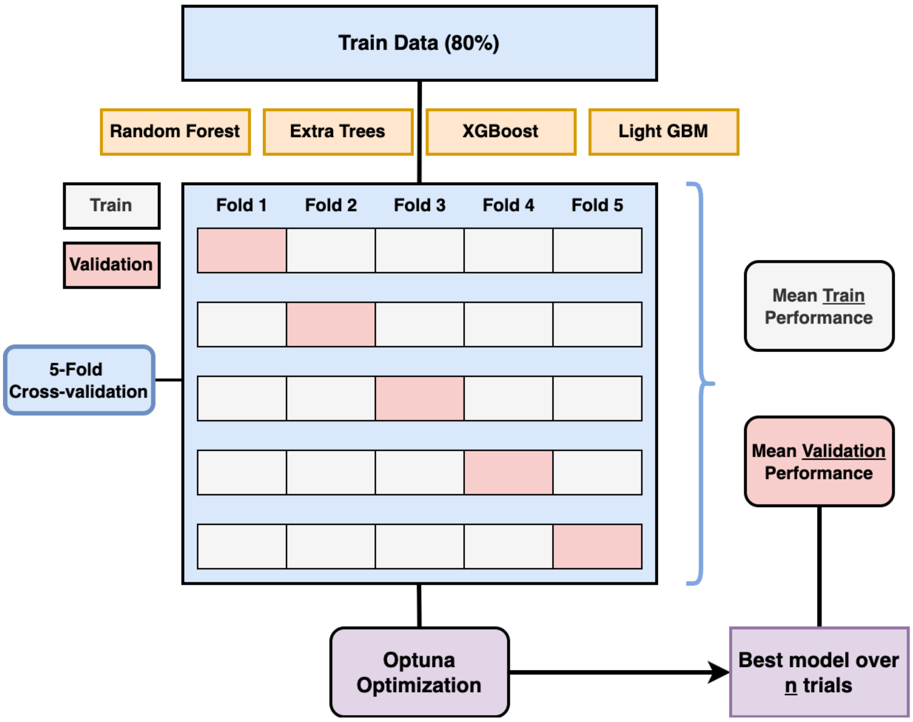


Figure 8: Diagram outlining the optimization of ML models using optuna. All machine learning algorithms (except linear regression) is optimized through this method. Each optimization trial runs the model through a 5-fold cross-validation on the training set. The first trial chooses hyperparameters randomly while subsequent trials select promising hyperparameter values based on the validation performance of previous trials. Once all trials have been completed, the model having attained the best validation performance is returned and its hyperparameters are used for the next steps of analysis.

6.7.4 Model Comparison

To determine whether ensemble models significantly outperform the Linear Regression baseline, each algorithm is run through 10 iterations of optimization. Since the initial choice of hyperparameters is random, it can lead to a poor choice of initial values. Considering that the starting point of optimization can greatly influence the final performance (i.e., getting stuck on a local optimum), we run multiple optimization iterations for each ML algorithm. Afterwards, for each algorithm type, we can select the iteration having the highest validation R^2 score. This iteration is the optimized version of the algorithm and can then be compared against the other optimized models.

Once all models (Random Forest, Extra Trees, XGBoost, and Light GBM) have been optimized, we identify for each family (bagging and boosting) the model having the highest validation performance on each task (DASS and CTPI). A statistical test is then used to validate the performance differences between bagging/boosting algorithms and the Linear Regression baseline. Two established methods were considered: the 5x2 cross-validation paired t-test by Dietterich (1998) (Dietterich, 1998) and the 5x2 cross-validation combined F-test by Alpaydin (1999) (Alpaydin, 1999). These address known statistical shortcomings of traditional tests, which often suffer from inflated Type I error rates due to overlapping training sets and non-independent samples (Dietterich, 1998). The 5x2cv paired t-test mitigates these issues by ensuring that each training and test set is independent. However it produce inconsistent results depending on the choice of fold in the t-statistic. The 5x2cv combined F-test addresses this problem by aggregating information from all folds which stabilizes the result and provides a more powerful and reliable hypothesis test (Alpaydin, 1999). As described in figure 9, the 5x2cv combined F-test uses the full task-specific dataset. The test has five replications where the data is randomly split in two halves (one for training each model and the other for testing the model). Once the model has been trained in one split and tested on the other, the splits are switched and the model is trained on the previous test split and tested on the previous train split. This allows the obtention of 10 performances for each model and the computation of 10 performance differences. The performance differences are then used to compute the F-statistic and the p-value to assess the significance of the performance difference.

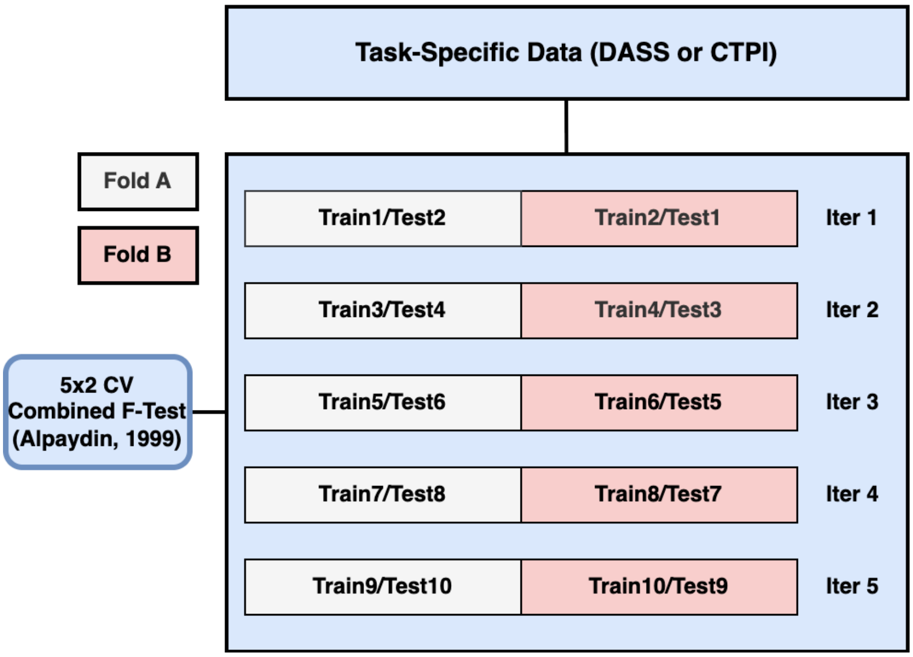


Figure 9: Diagram outlining the functioning of Alpaydin's 5x2cv combined F-test. Using the full task-specific dataset, five replications each with 2-fold cross-validation are performed. For each replication, the dataset is split randomly into two halves and each model is trained on one and evaluated on the other. Then, the same replication is switched, meaning that the fold previously used for testing is now being trained on the one used for training is tested on. This allows to obtain 10 performance differences between the two models which are used to compute the F-statistic (Alpaydin, 1999).

6.7.5 Model interpretation with SHAP values

The best-performing model from each task (DASS and CTPI) is further interpreted using SHAP (SHapley Additive exPlanations). These allow to assess the contribution of key predictors on the decision-making of the model. For this step, the best model for each task is trained on the full training dataset, and SHAP values are computed on the corresponding test set. This approach ensures interpretation based on unseen data while reducing computational cost. Predictor-specific SHAP values were analyzed to address the following hypotheses: H2.1. PIUQ positively contributes to DASS prediction; H2.2. QDE negatively contributes to CTPI prediction.

6.7.6 Cross-country predictor variation

To examine whether model interpretations varied across European countries (H2.3) the following procedure was applied: 1) Each task-specific dataset was further partitioned into country subsets (UK, Switzerland, Spain, Poland, Czechia, Germany). 2) For each country, a separate train-test split (80-20) was created. 3) The best model for each task (identified in previous steps) was trained independently on each country's training subset. 4) SHAP values were computed on each country-specific test set. 5) The mean absolute SHAP value of the predictor of interest (PIUQ for DASS; QDE for CTPI) was calculated for each country. Finally, to assess significant differences across countries, one-way ANOVA tests were conducted on the SHAP values for each task.

7 Results and Analysis

The results section of the project will be separated into three main parts: model comparison, model interpretation, and country variations. Each part is further divided into a subsection for the DASS (Lovibond and Lovibond, 1995) task and one for the CTPI (Denovan and Dagnall, 2019) task. First, we go through the results of the model comparison study, where we compare the performances of the four selected ML algorithms (Random Forest, Extra Trees, XGBoost, LightGBM) with linear regression as the baseline. Once the best model for each task is identified, we interpret its decision-making process with SHAP values through beeswarm plots (showing multiple feature contributions at once), and specific scatter plots (for the features of interest). Finally, we explore how input variables impact algorithmic predictions differently when applied to different countries, highlighting cross-cultural effects.

7.1 Model Comparison

In this section we go through the comparison of algorithms for both regression tasks, DASS (Lovibond and Lovibond, 1995) and CTPI (Denovan and Dagnall, 2019). Each of the five types of models follow 10 rounds of hyperparameter optimization, this allows not only comparison of types of ML models but also different parametrizations. The algorithm that performs the best overall is the one selected for interpretation.

7.1.1 Depression Anxiety Stress Scale

A preliminary comparison was conducted to evaluate the predictive performance of the five different ML algorithms on the overall DASS (Lovibond and Lovibond, 1995) score. Each model (except Linear Regression) was optimized 10 times with Optuna. This produced 41 total models to be compared. Before observing the results, it is important to know that the actual differences between models are very small, on the order of merely 3%. Although small, the differences will be tested for statistical significance later. Figure 10 provides an overview of the performances of the different algorithms on the DASS (Lovibond and Lovibond, 1995) task:

Random Forest achieved a mean R^2 of 0.613 with a narrow range between 0.611 and 0.614. This indicates that RF has a stable, but modest performance.

Extra Trees slightly outperformed its bagging counterpart, Random Forest, with a mean R^2 of 0.621 and performance ranging from 0.618 to 0.622.

XGBoost and LightGBM, both boosting algorithms, yielded higher performance, with mean R^2 scores of 0.626 and 0.625 respectively. Boosting models showed wider performance range (0.619 - 0.630 for XGBoost, and 0.618 - 0.630 for LightGBM). Both algorithms have very similar performances and higher sensitivity to hyperparameter configurations while proposing higher potential accuracy.

Linear Regression, as a baseline, demonstrated the lowest performance with a fixed R^2 of 0.599. With a 2% difference from bagging algorithms and a 3% difference from boosting algorithms, non-linear algorithms seem to have potential in the DASS (Lovibond and Lovibond, 1995) task.

To test the hypotheses related to the first research question (finding the best machine learning algorithm for the prediction tasks), we need first to select the best model for the boosting and bagging families. For bagging, Extra Trees shows a slight superior performance to the Random Forest (the best ET model has a score 0.8% superior to RF), and will thus be selected for testing against the baseline. The distinction between XGBoost and LightGBM is less clear, although XGB shows a slightly better stability, LightGBM has one iteration that achieves slightly better performance. This situation required an additional optimization round focused specifically on XGBoost and LightGBM.

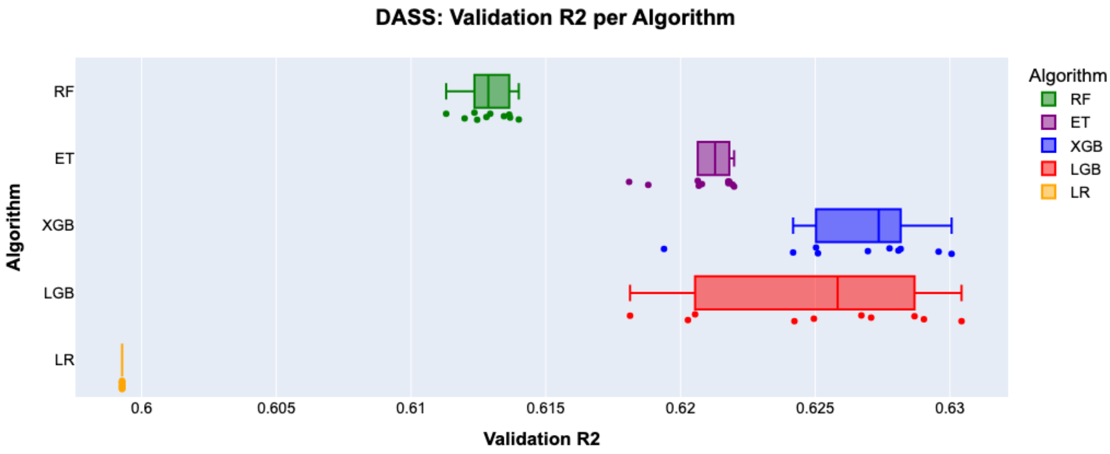


Figure 10: Boxplot comparison of five regression algorithms over 10 iterations of optuna optimization each for the prediction of overall (sum of all items) DASS (Lovibond and Lovibond, 1995) score. Random Forest (RF), Extra Trees (ET), XGBoost (XGB), LightGBM (LGB) were optimized 10 times on the training set and evaluated on the validation set; Linear Regression (LR) appears as a non-optimized baseline (no hyperparameters to optimize). The evaluation metric is the coefficient of determination (R^2) attained on the validation set.

This second round of optimization included 30 iterations for each of the algorithms (i.e., comparison of 30 versions of XGBoost against 30 versions of LightGBM). The LightGBM model had one of its iterations considered as an outlier due to its performance being very low ($R^2 < 0.52$), that outlier was removed from figure 11 so that it could be more easily interpreted (keeping the outlier caused the plot to be squished and made it hard to observe its trends). XGboost achieved higher performance in this optimization round, with a validation R^2 of 0.63.

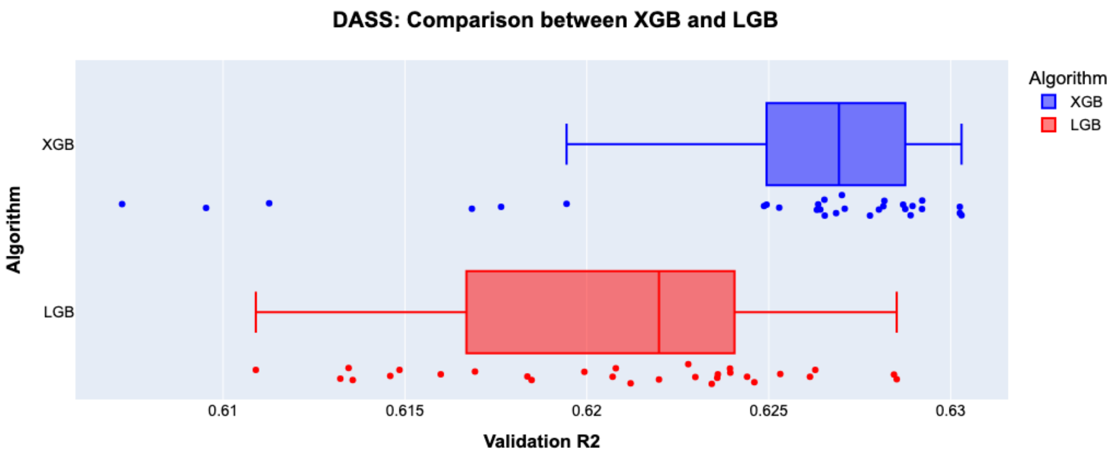


Figure 11: Boxplot focused comparison of XGBoost (XGB) and LightGBM (LGB) over 30 iterations of optimization each for the prediction of overall (sum of all items) DASS (Lovibond and Lovibond, 1995) score. Only 29 iterations are shown for the LightGBM as its worst performing iteration was considered an outlier ($R^2 < 0.52$) and removed.

To test whether the model with the best parametrization for each family (bagging and boosting) is significantly better in performance than the baseline linear regression model, we apply a statistical test efficient for comparing machine learning models called "combined 5x2 cross-validated F-Test" (Alpaydin, 1999) that was detailed in section 6. The LR model when tested on the DASS (Lovibond and Lovibond, 1995) task with the 5x2 cv F-Test scored an average of 0.597 (+- 0.007) R^2 on its 10 folds. The tenth iteration of Extra Trees was selected to represent bagging algorithms, with it achieving the highest validation score among all Random Forest and Extra Trees iterations during the first optimization round (0.622 validation R^2). When tested with 5x2 cross-validation, it achieved an average of 0.619 (+- 0.008) R^2 on the 10 folds, 2.23% above the baseline performance. The F-statistic was 20.12 with a p-value of 0.002 which is inferior to the threshold of 0.05 showing a significant superiority of ET over the baseline. XGBoost's ninth iteration was selected as the boosting algorithm for the DASS (Lovibond and Lovibond, 1995) task, having achieved the highest validation score among all XGBoost and Light GBM iterations in the second optimization round (0.63 validation R^2). During the 5x2 cv f-test, this model obtained an average of 0.629 (+- 0.008) R^2 across all 10 different folds which is 3.29% above the mean performance of the baseline. The F-value was 46.27 with a p-value of 0.0003 which is inferior to 0.05. The test demonstrates that XGBoost performs significantly better than the LR baseline for the prediction of the overall DASS (Lovibond and Lovibond, 1995) score. For the prediction of DASS (Lovibond and Lovibond, 1995) overall scores, both bagging and boosting algorithms perform significantly better than the Linear Regression baseline.

7.1.2 Chronic Time Pressure Inventory

The model selection for the CTPI (Denovan and Dagnall, 2019) task follows the same procedure as explained for the DASS (Lovibond and Lovibond, 1995) task. Again, the four different types of algorithms (RF, ET, XGB, LGB) each follow 10 rounds of optimization, producing models with different parametrizations. The Linear Regression algorithm is not included in the optimization due to the absence of optimizable parameters. The comparison for the prediction of CTPI (Denovan and Dagnall, 2019), shown in figure 12, shows a different pattern compared to the DASS (Lovibond and Lovibond, 1995) task. Overall, the performances are slightly higher for the CTPI (Denovan and Dagnall, 2019) task, with most results ranging between 0.64 and 0.675 R^2. This may hint towards the dataset having easier patterns and interactions to be learnt for the prediction of CTPI (Denovan and Dagnall, 2019). The results of the comparison were:

Random Forest and Extra Trees both bagging algorithms performed worst in the CTPI (Denovan and Dagnall, 2019) task. Random Forest was the overall worst performing algorithm with a mean R^2 of 0.647 and a range between 0.638 and 0.649. Extra Trees performed slightly better, attaining a mean R^2 of 0.652 with a range between 0.650 and 0.654.

XGBoost and LightGBM, again, yielded the highest performances, with mean R^2 scores of 0.668 and 0.669 respectively. In this task, XGBoost had was less stable than LightGBM, with a wider range of 0.656 - 0.673 compared to the range of 0.665 - 0.672 from LightGBM.

Linear Regression achieved a mean R^2 of 0.661, outperforming both bagging algorithms.

In the CTPI (Denovan and Dagnall, 2019) task, the bagging algorithms attained an inferior performance compared to the LR baseline. The best bagging model was Extra Trees, with its best parametrization attaining an average R^2 of 0.654. Again, the boosting algorithms show very close performances and lower stability (higher spread of performances) compared to the other models. In order to distinguish between them, another additional round of optimization, similar to the one for the DASS (Lovibond and Lovibond, 1995) task, with 30 optimization rounds for each type of algorithm is run.

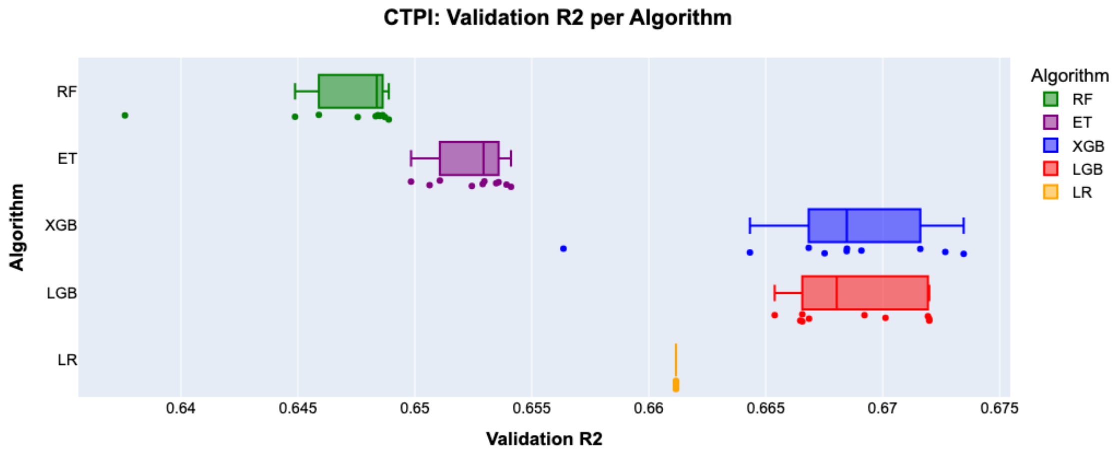


Figure 12: Boxplot comparison of five regression algorithms over 10 iterations of optuna optimization each for the prediction of overall (mean of all items) CTPI (Denovan and Dagnall, 2019) score. Random Forest (RF), Extra Trees (ET), XGBoost (XGB), LightGBM (LGB) were optimized 10 times on the training set and evaluated on the validation set; Linear Regression (LR) appears as a non-optimized baseline (no hyperparameters to optimize). The evaluation metric is the coefficient of determination (R^2) attained on the validation set.

When focusing on the comparison between XGBoost and LightGBM, we find that for the CTPI (Denovan and Dagnall, 2019) task, the LightGBM algorithm attains better performance overall. It achieved a mean R^2 of 0.667 on a range between 0.654 and 0.674, while XGBoost achieved a mean of 0.664 and a range between 0.653 and 0.670. For the boosting algorithms, the best parametrization of LightGBM is chosen, having obtained a R^2 of 0.674, 0.4% higher than its counterpart.

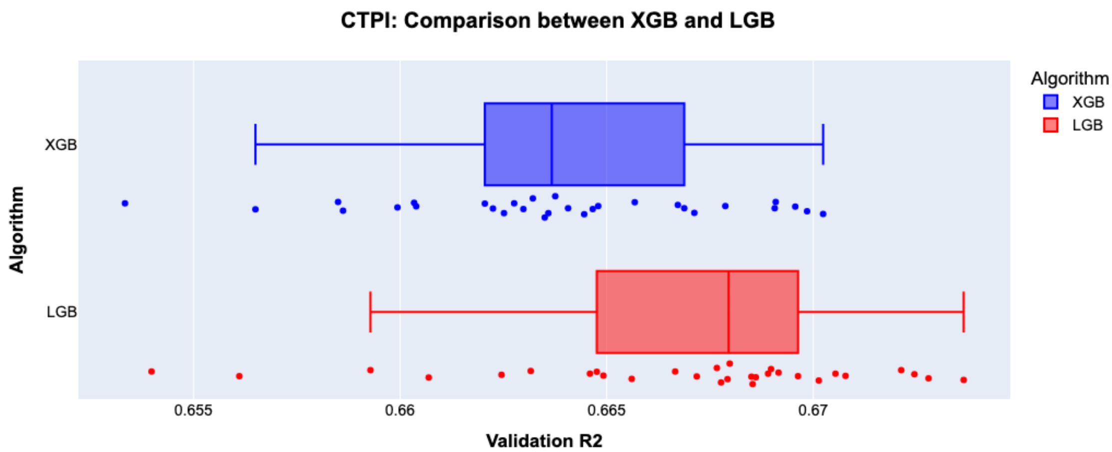


Figure 13: Boxplot focused comparison of XGBoost (XGB) and LightGBM (LGB) over 30 iterations of optimization each for the prediction of overall (mean of all items) CTPI (Denovan and Dagnall, 2019) score.

To test whether the model with the best parametrization for each family , we conducted again a F-Test. The algorithms selected for comparison against the baseline model (LR) are the Extra Trees (bagging), and the LightGBM (boosting). The Linear Regression (LR) model, when tested on the CTPI (Denovan and Dagnall, 2019) task with the 5x2 cv F-Test, scored an average of 0.656 (+- 0.008) R^2 on its 10 folds. The first iteration of Extra Trees was selected for the CTPI (Denovan and Dagnall, 2019) task, as it achieved a high validation score (R^2 of 0.654). When tested with 5x2 cross-validation against the LR baseline, it achieved a mean of 0.648 (+- 0.006) R^2 on the 10 folds, 0.8% lower compared to the baseline performance. The F-statistic was 4.98 with a p-value of 0.045 which is inferior to the threshold of 0.05 and indicates a significant inferiority of ET over the baseline for the CTPI (Denovan and Dagnall, 2019) task. LightGBM's 24th iteration was selected as the boosting algorithm for the CTPI task (validation R^2 of 0.674). During the 5x2 cv F-Test comparing it to the baseline, this model obtained an average of 0.666 (+- 0.007) R^2 across all 10 different folds, 1% higher compared to baseline. The F-value was 13.88 with a p-value of 0.005. The p-value being below 0.05, demonstrates that LightGBM performs significantly better than the LR baseline for the prediction of overall CTPI (Denovan and Dagnall, 2019) score.

In summary, for the prediction of CTPI (Denovan and Dagnall, 2019) overall scores, both the selected bagging (Extra Trees) and boosting (LightGBM) algorithms demonstrate statistically significant differences with the baseline. However, this time, bagging performed worse than the baseline, which shows that bagging models (e.g., Random Forest or Extra Trees) are less efficient for the prediction of CTPI (Denovan and Dagnall, 2019). The best boosting algorithm (24th iteration of LightGBM) achieved higher performance against the LR baseline, confirming the superiority of boosting algorithms for a psychometric dataset.

7.1.3 Hypotheses summary

In this section we will go through the results concerning the first two hypotheses H1.1 and H1.2. Both hypotheses aim to compare the performance of Decision Tree ensemble algorithms against a Linear Regression baseline. H1.1 supposes that bagging algorithms will have superior performance, and H1.2 supposes that boosting algorithms will have higher performance against the baseline. For H1.1, Extra Tree attained superior performance compared to Random Forest in both tasks. **H1.1 is only partially supported**, with bagging being significantly superior to the baseline in the DASS (Lovibond and Lovibond, 1995) task but not in the CTPI (Denovan and Dagnall, 2019) task (where the baseline achieved higher performance). For H1.2, the highest performing boosting model differed for the two tasks. A parametrization of XGBoost performed better when predicting DASS (Lovibond and Lovibond, 1995) and LightGBM performed better when predicting CTPI (Denovan and Dagnall, 2019). For both tasks, boosting algorithms achieved better performance compared to the baseline and also compared to the bagging algorithms, and so **hypothesis H1.2 is confirmed**. Seeing as boosting is the overall best family of machine learning algorithms for both tasks, these will be used for the next analysis tasks (model interpretation and cross-country variations). A summary of the model selection results for each task is presented in table 5

Table 5: Final model selection for each prediction task

Task	**Best Model**	**Iteration**	**Train** R^2	**Validation** R^2	**Test** R^2
DASS [a]	XGBoost	9	0.834	**0.630**	0.626
CTPI [b]	LightGBM	24	0.800	**0.674**	0.652

[a] Depression Anxiety Stress Scale (Lovibond and Lovibond, 1995)
[b] Chronic Time Pressure Inventory (Denovan and Dagnall, 2019)

7.2 Model Interpretation

Now that the best model for each task was identified, XGBoost for the prediction of DASS (Lovibond and Lovibond, 1995), and LightGBM for CTPI (Denovan and Dagnall, 2019), we can use a model interpretation method to understand how decisions are made. SHAP (Shapley Additive exPlanations) values are a model-agnostic method used for interpretation, they indicate how a feature contributes to the prediction. More precisely, these values are calculated for each data point of the test set (subset of data that was neither used for training nor for selecting the final models) and they represent how each feature influenced the final prediction for that specific data point.

7.2.1 Depression Anxiety Stress Scale

The selected parametrization of XGBoost was evaluated on the test set and achieved a R^2 of 0.626, slightly lower than the 0.63 validation R^2. The test and validation performances are close and show that we can expect this specific parametrization of XGBoost to explain around 0.625 to 0.63 of the variance when predicting DASS (Lovibond and Lovibond, 1995). Note that this model obtained a score of 0.834 on the training set which indicates that the model is overfitting to the training data. The model has learned patterns specific to the training data that do not generalize as well to unseen data, and we can expect the actual performance to be closer to the 0.63 R^2 mark.

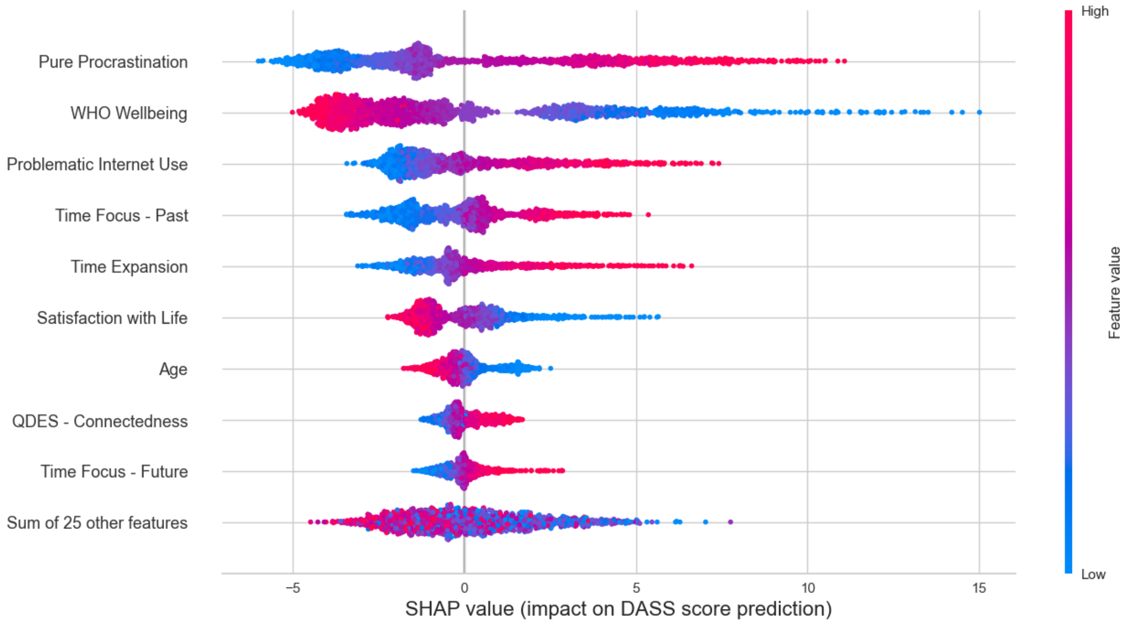


Figure 14: Beeswarm plot showing the SHAP values of the features that impact DASS (Lovibond and Lovibond, 1995) prediction the most. Blue points indicate a lower feature value, while red points indicate a higher feature value. In the x-axis, we observe the SHAP value which is the impact the feature has on the prediction. For example, the top feature "Pure Procrastination" is the one having the highest impact in the DASS prediction. This feature has most of its blue points in the negative side of the x-axis, while its red points are placed in the positive side of the x-axis. This indicates that a lower procrastination score (blue points) pushes the model to predict lower DASS values. Conversely, higher procrastination score (red points) pushes the model to predict higher DASS values.

Figure 14 presents the top 9 most important features and all their SHAP values calculated for each data point in the test set, these are described in table 6.

Table 6: Direction and approximate magnitude of SHAP associations with DASS (Lovibond and Lovibond, 1995)[a]

Feature	**Direction of Association**	**Magnitude**[b]
Pure Procrastination (Steel, 2010)	Positive	High
WHO Well-Being Index (WHO, 1998)	Negative	High
Problematic Internet Use (Laconi et al., 2019)	Positive	Moderate
Time Focus – Past (Shipp et al., 2009)	Positive	Moderate
Time Expansion (Wittmann and Lehnhoff, 2005)	Positive	Moderate
Satisfaction with Life (Diener et al., 1985)	Negative	Moderate
Age	Negative	Low
QDES – Connectedness (Witowska et al., 2025)	Positive	Low
Time Focus – Future (Shipp et al., 2009)	Positive	Low
Remaining 25 predictors[c]	Inconclusive	—

[a] Associations derived from the XGBoost SHAP beeswarm plot; positive (negative) means higher (lower) feature values tend to increase the predicted DASS score.
[b] "Magnitude" is based on relative SHAP value spread from 14: **High** $\simeq$ top three drivers, **Moderate** $\simeq$ mid-tier contributors, **Low** $\simeq$ small but interpretable effects.
[c] The aggregated SHAP signal of the residual feature set is diluted and does not reveal a clear, unified association.

We can already observe that our independent variable of interest (PIUQ, Laconi et al., 2019) comes in the third rank of importance in the prediction of psychological distress (DASS, Lovibond and Lovibond, 1995). When interpreting figure 14, the type of association (positive or negative) that is used by the model can be interpreted by checking at how a low (blue points) or high (red points) feature value affects the SHAP value (x-axis position). For the problematic internet use (PIUQ, Laconi et al., 2019) the blue points (low values of PIUQ) are concentrated on the negative side of the x-axis (negative SHAP values). This indicates that a low PIUQ (Laconi et al., 2019) value, pushes the model to predict a lower DASS (Lovibond and Lovibond, 1995) value. In contrast, the red points (high PIUQ score) are concentrated on the positive side of the x-axis (positive SHAP values), thus a high PIUQ (Laconi et al., 2019) score, pushes the model to predict a higher DASS (Lovibond and Lovibond, 1995) value. We can thus assume that the XGBoost model assumes a positive association between problematic internet use and its target variable (overall DASS score).

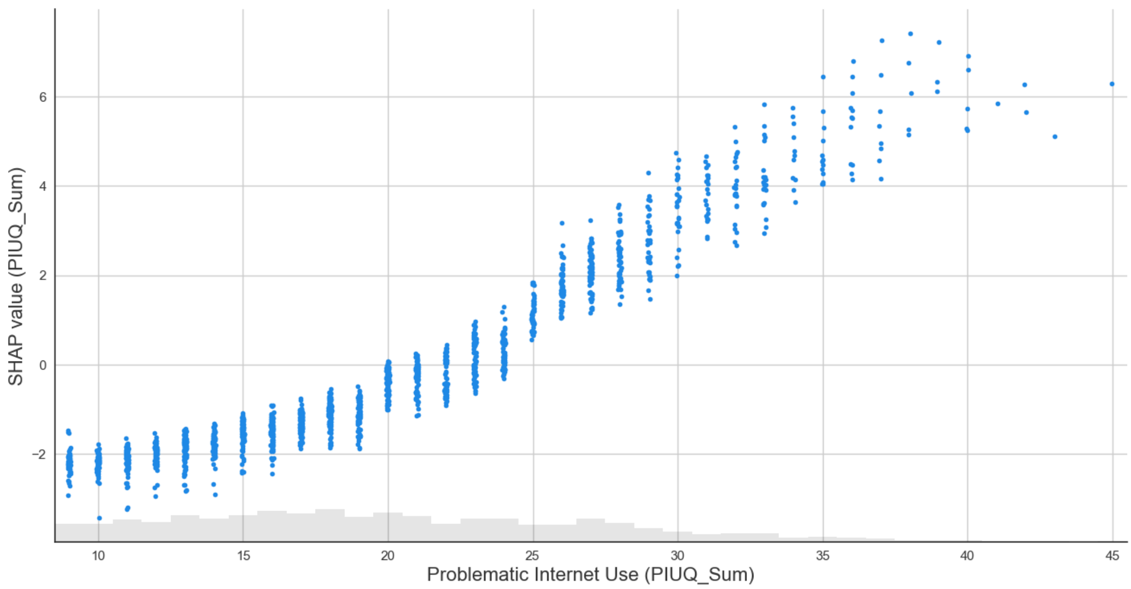


Figure 15: Scatterplot of the Problematic Internet Use (Laconi et al., 2019) SHAP values when predicting DASS (Lovibond and Lovibond, 1995).

7.2.2 Chronic Time Pressure Inventory

The model selected for the prediction of CTPI (Denovan and Dagnall, 2019) is the 24th iteration of LightGBM in the second optimization round. It achieved a train R^2 score of 0.80 and a validation score of 0.674. The evaluation on the test set returned a R^2 score of 0.652, the difference between validation and test for this model is greater than the XGBoost model for the DASS (Lovibond and Lovibond, 1995) task (0.022 vs. 0.004). This suggests that the effectiveness of the model on unseen data is less stable than for the DASS (Lovibond and Lovibond, 1995) task model. The overfitting problem is still present, with the model being much more performant on the training dataset compared to unseen data. However, overall, the CTPI (Denovan and Dagnall, 2019) task shows better performance compared to the DASS (Lovibond and Lovibond, 1995) task. Figure 16 presents the top 9 most important features and all their SHAP values calculated for each data point in the test set. To these, was added the independent variable of interest for hypothesis 2.2 (overall quality of digital life score) that was ranked 13th in terms of feature importance.

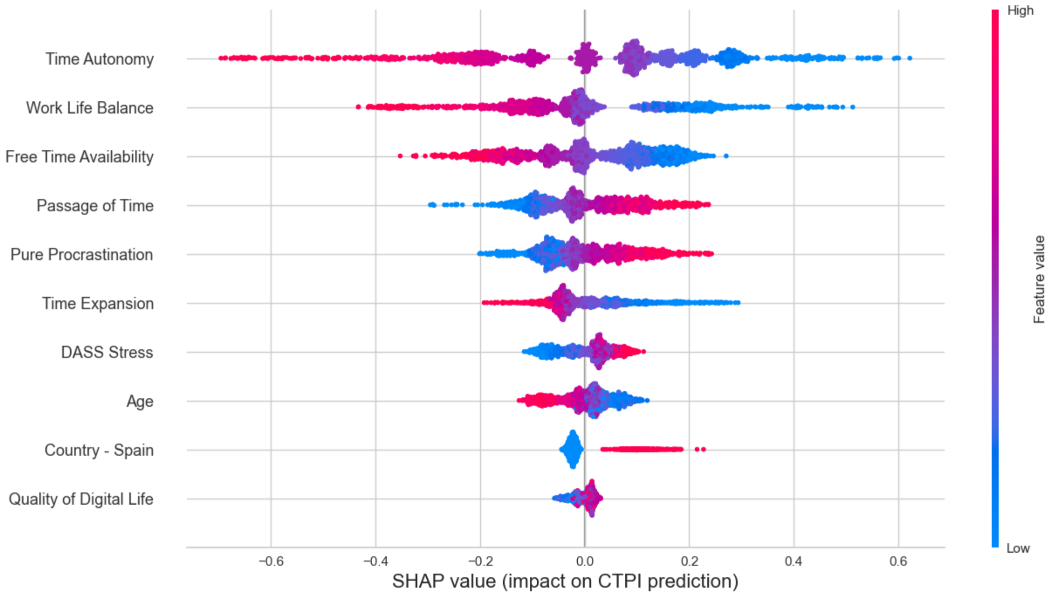


Figure 16: Beeswarm plot showing the SHAP values of the features that impact CTPI (Denovan and Dagnall, 2019) prediction the most. QDES (Witowska et al., 2025) was manually placed as the last value in the graph instead of the sum of all features as its value was not part of the top 9 due to its importance being very weak.

The interpretation of the SHAP values is described in table 7.

Table 7: Direction and approximate magnitude of SHAP associations with CTPI (Denovan and Dagnall, 2019)[a]

Feature	**Direction of Association**	**Magnitude[b]**
Time Autonomy (Macan et al., 1990)	Negative	High
Work-Life Balance (Brough et al., 2009)	Negative	High
Free Time Availability (Ruth Ogden, TIMED)	Negative	Moderate
Perceived Passage of Time [d]	Positive	Moderate
Pure Procrastination (Steel, 2010)	Positive	Moderate
Time Expansion (Wittmann and Lehnhoff, 2005)	Negative	Moderate
DASS - Stress (Lovibond and Lovibond, 1995)	Positive	Low
Age	Negative	Low
Country – Spain	Positive	Low
Quality of Digital Experience (Witowska et al., 2025)[c]	Positive	Very Low

[a] Associations derived from the LightGBM SHAP beeswarm plot; positive (negative) means higher (lower) feature values tend to increase the predicted CTPI score.
[b] "Magnitude" is based on relative SHAP value spread from 16: **High** $\simeq$ top three drivers, **Moderate** $\simeq$ mid-tier contributors, **Low** $\simeq$ small but interpretable effects.
[c] The overall quality of digital experience score was added as the tenth feature in the plot due to it being of particular interest for hypothesis 2.2. However, it is not the 10th most important feature from the dataset but the 13th in rank.
[d] Perceived Passage of Time (Ogden, 2020; Wittmann and Lehnhoff, 2005).

When interpreting figure 16 we see that the most important features for the model to predict chronic time pressure are: time autonomy, work-life balance, and free time availability. This is unsurprising as chronic time pressure is a measure of time and so are the top 3 features. This is further highlighted by five of the 9 top features being associated to time, of which the top 4 are all associated to time. Of course, it is sensible that the model focuses more on time-associated features to predict a time-associated target. When focusing on quality of digital experience, we observe that as the 13th ranked feature, it is of very small importance for the final decision. So the model did not seem to find very interesting patterns that would help predict time pressure from QDES (Witowska et al., 2025). However, it is still possible to distinguish a somewhat positive association between it and time pressure (as most blue points are grouped to the negative side of the x-axis and most red points are grouped to the positive side). A higher quality of digital experience does seem associated with higher time pressure to some extent as it slightly pushes the model to predict higher time pressure scores.
In figure 17 we can observe the SHAP values for the overall quality of digital experience score as well as its 3 different subcsales (Connectedness, Health and Wellbeing, and Time and Efficiency). When exploring the SHAP values specifically, we see that they are of very low amplitude and have non-linear associations.

Overall Score (QDES_Mean): Low to moderate (1-3) overall quality of digital life seem to motivate the model to predict lower chronic time pressure values (negative shap values). There is a shift around the overall QDES mean score of 3, where higher scores will push the predictions of time pressure to be higher. However, even after the shift, there are falls again into negative SHAP values at around 3.7 and 4.5 QDES (Witowska et al., 2025) values. Overall, for most of the test predictions, the mean QDES score seems to only push for higher time pressure predictions at 3-3.5 and 4-4.4 ranges.

Connectedness (QDES_Mean_C): Values of connectedness below 2 and around 2.5 seem to push the model to predict lower time pressure. Where values between 2 and 2.5, and values above 3 very slightly induce higher time pressure. Relatedness values below 2 seem to have the highest impact in decision-making, the rest of SHAP values being very close to 0.

Health and Wellbeing (QDES_Mean_HW): This subscale shows a somewhat negative association with chronic time pressure prediction. Low health and wellbeing scores seem to associate positively with time pressure (range 1 to 2.5), with a shift at 2.5 where the association becomes negative. From 2.5 to 5, higher health and wellbeing scores push the prediction of chronic time pressure to be smaller.

Time and Efficiency (QDES_Mean_TE): This subscale has values around 2-2.5 and 4.5 decreasing the prediction of time pressure. The rest is mostly oscillating close to 0, having little influence on the actual decision-making.

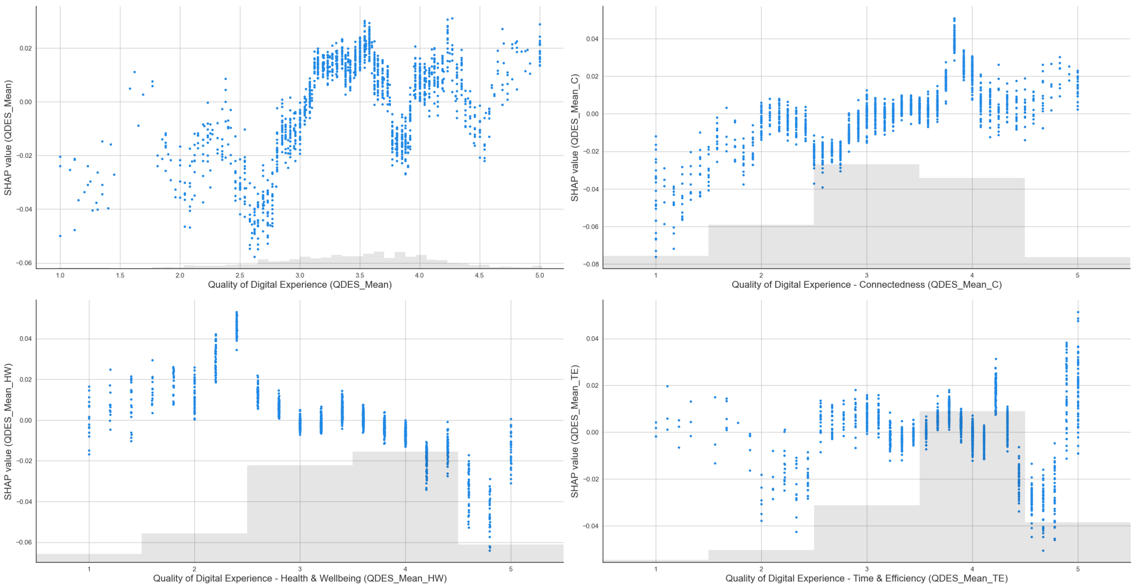


Figure 17: Scatterplot of QDES (Witowska et al., 2025) overall score and subscales SHAP values for the prediction of CTPI (Denovan and Dagnall, 2019). Only the health and wellbeing subscale has a negative association with CTPI. The overall associations are very weak.

Overall, quality of digital experience and its subscales have very weak effects on the actual decision-making processes of the machine learning model. We can identify some slight patterns, with the overall score and connectedness subscale having small positive associations, and the health and wellbeing subscale having a negative association. The time and efficiency subscale shows very weak SHAP values that gravitate around 0 with some falls into the negatives.

7.2.3 Hypotheses overview

Hypothesis 2.1 supposed that problematic Internet use would be positively associated with psychological distress. In the beeswarm plot 14, PIUQ (Laconi et al., 2019) is the third ranked feature in terms of importance when predicting DASS (Lovibond and Lovibond, 1995). When observing the SHAP values in figure 14 we see that when the value of overall PIUQ is low (blue points) the SHAP value is negative (pushes the model to predict a lower value of DASS). In contrast, when overall PIUQ (Laconi et al., 2019) is high (red points), the SHAP value is positive (pushes the model to increase the DASS value prediction). Likewise, figure 15 presents how SHAP values of PIUQ change in function of the overall PIUQ value. We can clearly see a positive association between overall PIUQ and its SHAP value (higher PIUQ score inducing higher SHAP values) with a y-intercept at approximately 22. This does confirm hypothesis H2.1 that problematic internet use is positively associated with psychological distress.
Hypothesis 2.2 supposed that quality of digital experience would be negatively associated with chronic time pressure. In the beeswarm plot 16, QDES (Witowska et al., 2025) has a very weak importance in the prediction of the overall score of the Chronic Time Pressure Inventory (CTPI_Mean). While interpreting the results, neither overall QDES nor the subscales were part of the top 10 variables in terms of importance, and so had to be added to figure 16 artificially to be compared with the others. We can also observe a small positive association between QDES (Witowska et al., 2025) and CTPI (Denovan and Dagnall, 2019) used by the model in its predictions. The hypothesis H2.2 cannot be affirmed with these results as the overall QDES score has a small positive association instead of the supposed negative association. However we can ob-

serve that the health and wellbeing (QDES_HW) subscale appears to be understood as a negative association by the model.

7.3 Cross-country Variations

In this section we go through the results that answer the final hypothesis (H2.3) that supposes the predictors, PIUQ (Laconi et al., 2019) and QDES (Witowska et al., 2025), of the two target variables, DASS (Lovibond and Lovibond, 1995) and CTPI (Denovan and Dagnall, 2019), vary depending on the country the data came from. To check how the features vary depending on the country, for each task, one model using the best parametrization for the task is trained on the data for each country. Since there are two tasks, and 6 countries, we end up with a total of 12 models (6 XGBoost and 6 LightGBM). Then we can compute the SHAP values for each of the models and calculate the importance of the features for each task on each of the 6 countries.

7.3.1 Variation of PIUQ's effect on DASS

The SHAP values of the problematic internet use feature for each country on the prediction of DASS (Lovibond and Lovibond, 1995) are presented on Figure 18. We can observe that there is a lot of variation on the importance of PIUQ (Laconi et al., 2019) depending on the country. Switzerland is the country where problematic internet use has the weakest importance in predicting psychological distress (average absolute SHAP of 0.832). In contrast, the United Kingdom is the country where PIUQ has the highest importance (average absolute SHAP of 2.688), with over 3 times the value of Switzerland. When analyzing the XGBoost model, the importance of problematic Internet use in predicting psychological distress is highest in the UK, followed by Spain. Germany and Poland show similar importance values, ranking next, with Czechia fifth and Switzerland last. A one-way ANOVA test on the SHAP values for PIUQ across countries reveals significant differences between all country pairs, except between Poland and Germany.

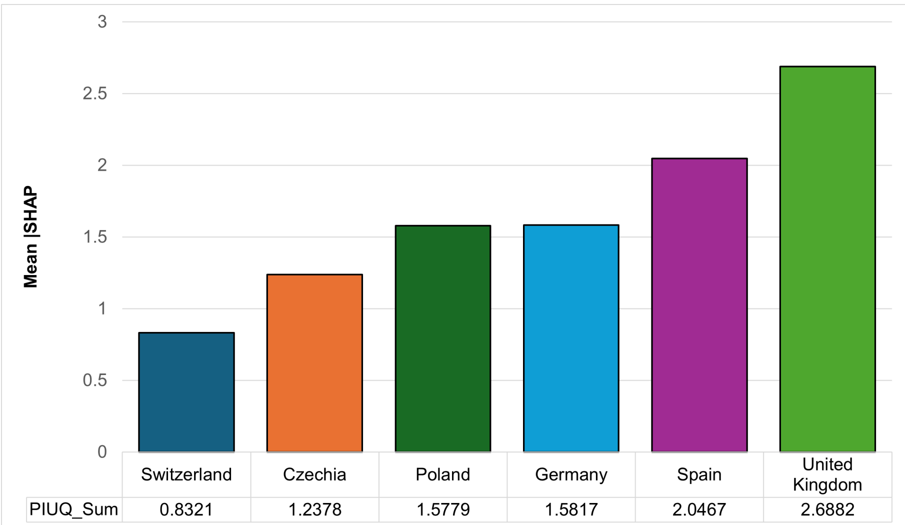
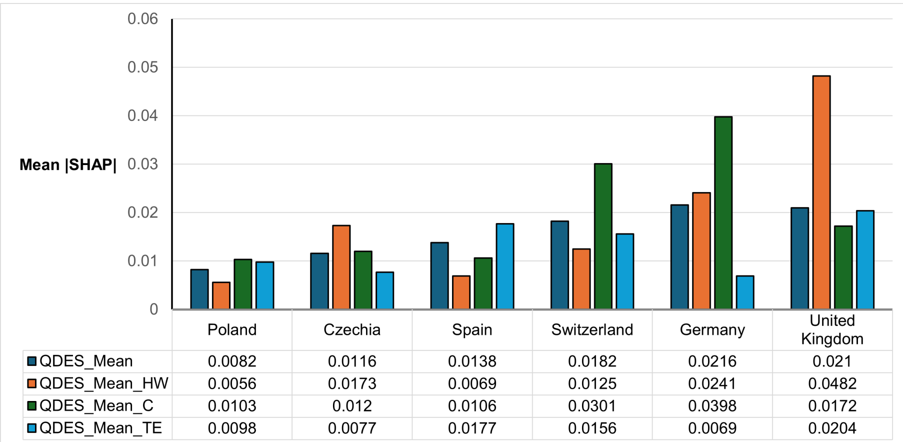


Figure 18: Variation of PIUQ (Laconi et al., 2019) overall score SHAP values for predicting DASS (Lovibond and Lovibond, 1995) depending on country subset. PIUQ_Sum indicates the overall score of the problematic internet use questionnaire which is the sum of all nine individual items.

7.3.2 Variation of QDES's effect on CTPI

The SHAP values for the mean QDES importance and its subscales for each country on the prediction of CTPI are presented on Figure 19. In terms of overall QDES (Witowska et al., 2025) importance, the United Kingdom (0.0210) and Germany (0.0216) showed the highest values, suggesting that, in these countries, overall quality of digital life contributed more to the prediction of chronic time pressure. Conversely, Poland (0.0082) and Czechia (0.0116) had the lowest SHAP

values for overall QDES, implying a weaker direct association.
When disaggregated by subscale, Health and Wellbeing (HW) showed particularly high importance in the UK (0.0482) and Germany (0.0241), while its relevance was markedly lower in Spain (0.0069) and Poland (0.0056). Connectedness (C) emerged as the most influential factor in Germany (0.0398) and Switzerland (0.0301), but was relatively less important in Spain (0.0106) and Poland (0.0103). The Time and Efficiency (TE) subscale showed its highest influence in the UK (0.0204) and Spain (0.0177), while having limited predictive value in Germany (0.0069) and Czechia (0.0077).
In terms of significance, the overall QDES (Witowska et al., 2025) score is significantly different for all pairs except three: Czechia and Spain, Germany and UK, and Switzerland and UK. For health and well-being only the Poland and Spain pair has no significant difference. Relatedness has three non significantly different pairs: Czechia-Poland, Czechia-Spain, and Spain-Poland. Finally, time and efficiency also had three pairs with no significant difference: Czechia-Germany, Czechia-Poland, and Spain-Switzerland.



	Poland	Czechia	Spain	Switzerland	Germany	United Kingdom
QDES_Mean	0.0082	0.0116	0.0138	0.0182	0.0216	0.021
QDES_Mean_HW	0.0056	0.0173	0.0069	0.0125	0.0241	0.0482
QDES_Mean_C	0.0103	0.012	0.0106	0.0301	0.0398	0.0172
QDES_Mean_TE	0.0098	0.0077	0.0177	0.0156	0.0069	0.0204

Figure 19: Variation of QDES (Witowska et al., 2025) overall score and subscales SHAP values for predicting CTPI (Denovan and Dagnall, 2019), depending on country subset. HW represents the Health and Wellbeing subscale, C represents the Connectedness subscale, and TE represents the Time and Efficiency subscale.

7.3.3 Hypothesis overview

The hypothesis H2.3 supposed that the predictors of interest for the target variables DASS (Lovibond and Lovibond, 1995) and CTPI (Denovan and Dagnall, 2019) would vary significantly depending on the country. Our analysis of the result using ANOVA tests mostly confirmed this hypothesis. Problematic Internet use was significantly different for all countries except Poland and Germany. Quality of digital experience and its subscales also showed much variation across countries although three pairs did not show significance for the overall score, connectedness, and time and efficiency, compared to just one pair not being significant for the health and wellbeing subscale. Overall, the hypothesis can be confirmed to some degree, as it is clear that there is great variation between countries on the importance of the features of interest. Each country does seem to have different patterns in how features are used for decision-making, which can highlight underlying social and cultural constructs.

8 Discussion

This master's thesis had two main objectives, first it aimed to apply Machine Learning algorithms for the prediction of psychological scores in a regression task. This was motivated by the capabilities of ML models to automatically extract patterns and non-linear information from larger datasets. The automatic nature of ML may allow the identification of unexpected interactions between variables and lead to more focused exploration on datasets where the total number of features is large. The dataset used throughout this study was collected during TIMED's Work Package

2 and contained responses of 7536 participants over 20 different questionnaires. These initially led to 46 features which after processing were converted to 49 features and separated into two datasets, one for each regression task. This study aimed to observe possible effects of increasing digital technology use on both mental health and time pressure, using the Depression Anxiety Stress Scale (Lovibond and Lovibond, 1995) and Chronic Time Pressure Inventory (Denovan and Dagnall, 2019) questionnaires as targets respectively. For this, promising algorithms were identified from the literature, optimized and compared against a Linear Regression baseline for each task. The best parametrization for each task was then interpreted using model interpretation methods to identify associations between the independent variables and the target used for decision-making. SHAP values allowed the identification of associations between problematic Internet use and psychological distress, and between quality of digital experience and chronic time pressure. The full dataset was later separated into subsets, one for each country of data collection, and the SHAP values of the predictors were compared across countries to observe whether significant variation was present that could highlight societal and cultural effects.

8.1 Overview of Main Findings

8.1.1 Research Question 1: Machine Learning Algorithm Comparison

The performance of ML algorithms was evaluated through the coefficient of determination (R^2) which evaluates how well a model fits to the provided data. This metric indicates the proportion of variance in the dependent variable that can be explained from the independent variables. In science fields dealing mostly with objective data (e.g., chemistry, physics) high R^2 scores over 0.8 are expected for a good ML model due to them dealing with molecules or materials with known properties and predictable behavior (Ozili, 2022). Note that obtaining high R^2 is often difficult in social sciences and psychology studies because they deal with highly subjective data (Ozili, 2022). The study by Smedslund et al. (Smedslund et al., 2022) analyzed 1,565 PsycINFO articles from 1956 to 2022 and found that the average explained variance reported in psychological research has remained consistently around 42.8%, corresponding to a R^2 of 0.428. In this thesis, performance for the psychological distress task was at 0.626 R^2 (62.6% explained variance), while performance for the chronic time pressure task attained 0.652 R^2 (65.2% explained variance). When comparing these performances to the average explained variance from psychological studies, Our approach, based on ML models, seems to have achieved acceptable performance levels.

H1.1. Bagging will perform better in a psychometric based dataset compared to a Linear Regression baseline This hypothesis had mixed results and was only partially confirmed. For both tasks, the algorithm representing bagging was the Extra Trees, the model performed significantly better than Linear Regression for the DASS task but significantly worse for the CTPI task. For the DASS task, since a linear model performs much worse compared to more complex models, prediction of DASS may benefit more from non-linear relationships. In contrast, the CTPI task has higher performance overall compared to DASS and also the Linear Regression baseline outperforms bagging models (even attaining close performance to boosting algorithms). CTPI prediction may benefit from strong linear relationships, this could be explained by the strongest predictors being time-related variables that contain similar information to the target. This can't be explained by the correlations between the target and predictors, since both tasks show similar patterns, with their top two predictors having absolute Pearson's r values above 0.6.

H1.2. Boosting will perform better in a psychometric based dataset compared to a Linear Regression baseline The second hypothesis that compared boosting algorithms to Linear Regression, was supported. Boosting models performed significantly better than Linear Regression for both the psychological distress and chronic time pressure tasks. XGBoost worked best for psychological distress, while LightGBM was best for chronic time pressure. Since both algorithms are closely related, their performance was expected to be similar, which was confirmed. Their hyperparameter spaces where also designed to be comparable. Although LightGBM had access to an additional boosting type (DART), the best-performing model still used the standard gradient boosting decision tree, just like XGBoost.

8.1.2 Research Question 2: Psychological Outcomes Prediction

H2.1. Problematic Internet use will be positively associated to psychological distress The interpretation of the XGBoost model when predicting DASS showed that a higher problematic internet use score pushed the model to predict higher psychological distress scores. This can indeed be interpreted as a positive association between PIU and psychological distress and confirms the our hypothesis. This association is also found in the literature, with multiple sources highlighting that problematic Internet use may induce psychological distress through symptoms such as depression, anxiety, and stress (Hatem and Ker, 2021; Ho et al., 2014; Ioannidis et al., 2016, 2018; Odacı and Çikrici, 2017; Spada, 2014; Whang et al., 2003; Young and Rogers, 1998). In Figure 15, we can observe the actual effect that problematic Internet use has on psychological distress prediction. The relationship is positive and mostly linear and we can observe a shift between negative and positive SHAP values around a value of 22. Interestingly, when using the problematic Internet use questionnaire for detecting problematic users, 22 is a commonly accepted cutoff between normal users and problematic ones (Koronczai et al., 2011; Laconi et al., 2018). The XGB model seems to have identified the cutoff and associates problematic users with higher psychological distress and normal users with lower distress. This further validates the reliability of the model's interpretation, highlighting that 0.626 test R^2 is indeed high enough for the model to use the predictors correctly for DASS prediction.

H2.2. Quality of digital life will be negatively associated to chronic time pressure The second hypothesis was disproved when interpreting the Light GBM model on CTPI prediction, where the overall quality of digital life score was found to have a weak positive association with chronic time pressure. The hypothesis supposed that higher perceived quality of life due to digital technology use would decrease chronic time pressure. Current sociological theories do point towards digital technology use increasing time pressure (Rosa, 2013; Wajcman, 2008, 2015), however there is little literature concerning the effect of perceived quality of digital experience (QDES) on chronic time pressure. The QDES questionnaire that produced the scores used as predictors for this hypothesis was developed during TIMED's Work Package 1 and was just recently published at the time of writing (Witowska et al., 2025). However, due to one of the QDES subscales being "time and efficiency", it is possible to suppose a negative association between the scale and time pressure. Indeed, a higher perception of DT as improving productivity should allow more time availability and less pressure. Moreover, as the type of digital technology use is a major driver of the association with psychological distress, with problematic use being associated with negative psychological outcomes (Dissing et al., 2021; Elhai, Dvorak, et al., 2017; Fioravanti et al., 2021; Franchina et al., 2018; Twenge and Campbell, 2018; Zhou et al., 2020) and balanced use being associated with positive or neutral outcomes (Granic et al., 2014; Przybylski and Weinstein, 2017). The same could be supposed for digital technology use and time pressure, with individuals perceiving DT as improving their quality of life suffering from lower time pressure. When interpreting the model (figure 17), both overall score (i.e., overall perceived quality of life through DT use) and the relatedness (i.e., how DT use facilitates social connections and interpersonal proximity) subscale of QDES had positive associations with CTPI. The time and efficiency subscale (i.e., how digital technology use improves productivity and helps save time) showed a mixed pattern with multiple peaks and troughs in its association. The health and wellbeing (i.e., how DT use contributes to mental health and well-being) subscale had a mostly negative association with time pressure. It does seem that the subscales are being used very differently by the model in its decision-making which could highlight heterogeneity of their effects on chronic time pressure. However, since the overall importance of these scores is very low for the prediction of CTPI, and the absence of literature, it is difficult to determine the reliability of these SHAP values.

H2.3. Predictors of psychological distress and chronic time pressure will vary in importance depending on the country The final hypothesis assumed that the observed predictors for both tasks (DASS and CTPI) would show significant variations in importance depending on the country where the data was collected. The hypothesis was mostly confirmed as, for both tasks (DASS and CTPI), the importance of the predictors (PIUQ and QDES) does vary significantly across most pairs of countries. For problematic Internet use (PIU), in figure 18, Switzerland is the country where PIU is the least important in predicting DASS and the United Kingdom is the country where it is most important. The UK's higher importance for problematic Internet use in predicting psychological distress may stem from its elevated exposure to digital risks: high DT usage inequality, high work-related stress, and youth overexposure with 37% of 15-year-olds being extreme Internet users in 2019 (OECD, 2019). In contrast, Switzerland has shown limited digital-work stress and lower youth risk exposure with rates of cyberbullying and extreme Internet use falling below OECD averages (OECD, 2019). For PIUQ's importance in DASS prediction, only Poland and Germany had non-significant differences, which highlights similarities between the countries in the experience of digitalization. Concerning the CTPI task and QDES's importance, the overall score and subscales vary among most pairs of countries although their impact, according to our results, are low. In figure 19 we can observe that the UK and Germany have the highest importances for the different scores, which would indicate that patterns of QDES's effect on CTPI are more easily identified in these countries. We can also observe that the health and wellbeing subscale has a much higher importance in the UK compared to the other QDES scores on other countries. In Germany and Switzerland, relatedness seems to be of particular importance for the prediction of chronic time pressure among the quality of digital life scales. However, Poland, Czechia, and Spain have very weak importances across all QDES scores for the prediction of chronic time pressure. This could indicate that the QDES instrument is not capturing the same level of information among all countries and its subscales are capturing different patterns for time pressure prediction. In its current state, the construct seems most effective for Switzerland, Germany, and the United Kingdom.

8.2 Key Challenges

8.2.1 Model selection

The first challenge was identifying relevant papers that used Machine Learning (ML) methods to predict psychological outcomes through psychometric-based datasets. Due to the absence of validated cutoffs for the Chronic Time Pressure Questionnaire at the beginning of the project, both prediction tasks were chosen to be regression tasks. While my research focused on regression tasks to predict continuous scores, I found that most existing studies in this domain emphasized classification tasks. The possibility of diagnosing mental health conditions or categorizing respondents in specific groups has been a greater focus when using ML in psychology, which is better suited for classification tasks. This mismatch imposed the selection of ML models and methods based mostly on classification studies.

8.2.2 Performance upper limitation

A performance plateau was reached early in the model comparisons and analysis of the WP2 dataset; models could hardly surpass 0.6 R^2 for DASS and 0.65 R^2 for CTPI. Multiple strategies were investigated to increase performance such as: the exclusion of subscales, the exclusion of aggregated scores, target transformations (log, box-cox, yeo-johnson) to improve their data distribution. However, only the hyperparameter optimization of the algorithms managed to slightly increase performance over the plateau. The evaluation of the algorithms was equally challenging as metrics used in regression tasks can be hard to interpret (e.g., mean squared error, mean absolute error, R^2) as their practical significance is not as clear as accuracy metrics and confusion matrices.

8.3 Limitations

One main limitation of the models, as demonstrated in the results section, is how they are overfitting to the training dataset. Despite the use of cross-validation and regularization during optimization, the models performed better on the training set than in the validation and test sets, for both tasks. This highlights that these models may be learning noise and/or specific patterns on the training data and do not generalize well on unseen data. Since the dataset used for the analysis is cross-sectional, the associations observed when interpreting the models cannot be used to establish cause-and-effect relationships. In addition, the patterns learned by the models in this dataset may be specific to the time of data collection. If associations evolve through time due to changes in societal and cultural norms, the models are not guaranteed to capture the changes and the prediction performance may decrease. As observed in the results section, boosting algorithms although outperforming other algorithms for both tasks, also had the highest instability (higher variation of results). This comes from the high stochasticity present in boosting algorithms (Friedman, 2002) and the hyperparameter optimization framework optuna (Akiba et al., 2019). This means the models can give very different results in each optimization round, so multiple rounds are needed to find the best parameters. Psychological constructs do not measure concepts perfectly, survey responses are highly subjective and can vary greatly depending on cognitive fatigue and response biases. Psychometric scores are also often computed as sums or averages of item responses and are not directly observed variables. This introduces additional noise in the data, which may be learned by highly complex non-linear ML algorithms and lead to overfitting training data as well as overlooking relevant patterns (Jacobucci and Grimm, 2020). When interpreting machine learning models, we examine how they make predictions. The reliability of these interpretations depends on how well the model performs and generalizes. Since psychometric datasets often contain a lot of noise and limit model performance, extra caution is needed. It's important to compare the model's findings with existing research to ensure meaningful interpretation. Another limitation lies in the internal structure of the models used for predicting psychological outcomes. Some scores, such as the DASS, contain many zero values which can pose problems when training certain algorithms. Regression models have trouble predicting exactly zero which will lead to performance issues when evaluating the model. If the target values are strictly positive (as is the case for DASS and CTPI) we have to ensure that the algorithm does not try to predict negative values. Both issues appeared when evaluating the DASS model, as it rarely predicted zero and even had a few negative predictions. In conclusion, the prediction of psychometric scores requires models that are tailored specifically for the type of score being predicted as these can vary greatly among constructs and their structure differs from usual continuous targets (e.g., discrete origins from summing ordinal, Likert-scale, items). Although psychometric scores can be used in regression tasks, they often behave like discrete variables, which can make model training more difficult. Using models or techniques that handle ordinal or categorical-like data could help address this issue. Finally, there is an inherent limitation to the dataset which is its high degree of multicollinearity. The dataset includes both overall scores and subscale scores, as well as measures that assess related constructs, leading to high intercorrelations among features. While excluding either the aggregate or subscale scores could reduce multicollinearity, doing may lead to inconsistency across measures as not all questionnaires provide aggregated scores or have subscales. The removal of subscales might strip away nuanced information about specific dimensions of a psychological construct, while removing aggregate scores could ignore the overall measure of the test.

8.4 Practical and Theoretical Applications

The model comparison component of the project showed how boosting decision tree algorithms attained high performance in a fully psychometric-based dataset. This can guide future regression ML studies on psychometric datasets to focus on optimizing and training boosting algorithms, such as XGBoost and Light GBM, skipping the time-consuming and computationally expensive model comparison processes for decision tree-based algorithms. This work can also serve for comparing performances of ML models in future regression studies with psychometric data as these are few in current literature. Additionally, this thesis highlights SHAP values as a sensible choice for interpreting the decision-making processes of complex algorithms when predicting psychological outcomes. The effects of digitalization in psychological distress and time pressure identified by the models motivate further investigation, with subscales of quality of digital life having different effects on time pressure prediction. The variations found in predictor importance across countries suggest an important cross-cultural effect of digitalization on both mental well-being and perceived time pressure.

8.5 Future Perspectives

The current study focused on the prediction of overall scores, however, as observed with QDES, different subscales can have different associations with other variables. Thus, it would be interesting to train models to predict the different subscales of DASS and CTPI constructs and observe how the predictor associations vary. In addition to the prediction of individual subscales, it would be interesting to predict the overall scores without including features that are closely associated to the target. The WP2 dataset is composed of four types of features: demographics, time measures, digitalization measures, and well-being measures. It is not surprising that when interpreting the models, the features having the highest importances are those of the same type as the target (e.g., DASS associates highly with PPS and WHO which are all well-being measures; CTPI associates highly with time autonomy and work-life balance which are both time measures). However, the strength of association between the target with similar features may be overestimating the performance these models could attain in real-life (where we are not guaranteed to have access to other well-being values when predicting psychological distress or other time measures when predicting time pressure) and hiding interesting association with less related features. Thus it would be interesting to extend this study and observe how performance and associations with digitalization predictors change when removing features of the same type as the target. The major advantage of Machine Learning lies in its capacity to automatically detect patterns in large datasets with high number of variables to predict a chosen target. This functioning would be better adapted for early exploration of a large dataset before applying more sophisticated statistical methods on interesting associations. For the purpose of the thesis, specific hypotheses were proposed focused on known variable associations. However, it would be interesting to run a large scale exploration of the WP2 dataset where a model could be trained to predict each feature based on the others, and then interpreted to identify unexpected variable associations in SHAP values. Although there may be multiple issues associated with doing so, such as the need to deal with target leakage for each task. Another interesting aspect to explore would be the use of multimodal data with machine learning algorithms to improve accuracy. While machine learning models can have bottlenecks induced by imprecise measures (coming from qualitative data), it would be interesting to leverage both subjective features (e.g., surveys) with objective measures (e.g., physiological data). The major advantage of complex ML algorithms is that they are able to be trained on complex inputs that merge different types of data (text, images, time series), that statistical methods and simple algorithms (e.g., LR) cannot. Of course the choice of the modalities requires consideration into ease of access to such modalities and usefulness in clinical settings. Ultimately, adding complementing psychometric data with objective data could lead to higher performances usable in clinical settings. The Quality of Digital Experience Scale showed considerable variation both across countries and among its subscales. This may hint towards potential cultural or contextual influences on how digital experiences are perceived. This variability highlights the need for further cross-cultural validation to ensure the scale's reliability and interpretability in diverse settings. Future studies should explore cultural differences in quality of digital experience across countries.

9 Conclusion

This study aimed to fill a gap of regression-based pipelines for the prediction of psychological outcomes using a dataset composed of socio-demographic and psychometric variables. Within this study, we investigated two families of machine learning algorithms, both based on the simpler decision tree. For each family, two algorithms were considered: Random Forest and Extra Trees for the "Bagging" family; XGBoost and LightGBM for the "Boosting" family. Each algorithm was trained and optimized on two tasks: the prediction of psychological distress, and the prediction of chronic time pressure. To ensure the value of using more complex models, they were compared against a Linear Regression baseline for both tasks. Our study demonstrated that boosting algorithms outperform both the baseline but also bagging algorithms for all tasks. XGBoost was the highest performing algorithm for the prediction of psychological distress, while LightGBM was the best performing model for prediction of chronic time pressure. The interpretation of models through SHAP values highlighted a positive association between problematic Internet use and psychological distress, affirming previous assumptions. Our analysis showed a shifting point between positive and negative effect of Internet use on psychological distress. The proximity of the shifting point to cutoffs present in the literature points towards good predictions from the XGBoost model. Analysis of the SHAP values of the LightGBM model highlighted weak positive association between overall quality of digital experience and time pressure. This contradicts the initial assumption that a high quality of digital experience would induce lower time pressure. The analysis of SHAP values for the subscales of quality of digital experience demonstrated different associations with time pressure. While connectedness has a weak positive association, health and wellbeing shows a weak negative association, and time and efficiency has mixed results. These findings suggest a complex interaction of the quality of digital experience scale, where the actual dimension of digital contribution influences time perception in different ways. Finally, the effects of digitalization on psychological distress were found to vary across countries. Switzerland had problematic Internet use as a weak predictor, where it is a strong predictor for the United Kingdom. This finding is in line with literature from the Organisation for Economic Co-operation and Development where digitalization affects wellbeing differently depending on the country. Our findings also point towards digitalization affecting time pressure differently among countries further strengthening our assumptions. In conclusion, this thesis shows the value of machine learning algorithms for the analysis of large psychological datasets. These can be useful to explore a large number of variable relationships before starting an in-depth statistical analysis. Our study also provides a complete methodology to identify non-linear relationships on psychometric datasets. These findings highlight the value of machine learning tools as exploratory instruments that can help uncover nuanced patterns in psychological data. Ultimately, these tools may help design targeted and culturally sensitive mental health interventions in an increasingly digital world.

References

Abela, J. R. Z., & D'Alessandro, D. U. (2002). Beck's cognitive theory of depression: A test of the diathesis-stress and causal mediation components [_eprint: https://bpspsychub.onlinelibrary.wiley.com/doi/pd *British Journal of Clinical Psychology*, *41*(2), 111–128. https://doi.org/10.1348/014466502163912

Adisa, T. A., Gbadamosi, G., & Osabutey, E. L. C. (2017). What happened to the border? The role of mobile information technology devices on employees' work-life balance. *Personnel Review*, *46*(8), 1651–1671. https://doi.org/10.1108/PR-08-2016-0222

Ahmad, M. W., Reynolds, J., & Rezgui, Y. (2018). Predictive modelling for solar thermal energy systems: A comparison of support vector regression, random forest, extra trees and regression trees. *Journal of Cleaner Production*, *203*, 810–821. https://doi.org/10.1016/j.jclepro.2018.08.207

Akiba, T., Sano, S., Yanase, T., Ohta, T., & Koyama, M. (2019). Optuna: A Next-generation Hyperparameter Optimization Framework. https://doi.org/10.48550/ARXIV.1907.10902

Alpaydin, E. (1999). Combined 5 × 2 cv F Test for Comparing Supervised Classification Learning Algorithms. *Neural Computation*, *11*(8), 1885–1892. https://doi.org/10.1162/089976699300016007

Amankwah-Amoah, J., Khan, Z., Wood, G., & Knight, G. (2021). COVID-19 and digitalization: The great acceleration. *Journal of Business Research*, *136*, 602–611. https://doi.org/10.1016/j.jbusres.2021.08.011

American Psychiatric Association (Ed.). (2013). *Diagnostic and statistical manual of mental disorders: DSM-5* (5th ed). American Psychiatric Association.

Amini-Rarani, M., Azami-Aghdash, S., Mousavi Isfahani, H., & Mohseni, M. (2024). Estimation of the prevalence of anxiety during the COVID-19 pandemic: A meta-analysis of meta-analyses. *BMC Public Health*, *24*(1), 2831. https://doi.org/10.1186/s12889-024-19729-7

Anderson, R., Bikson, T., Law, S., & Mitchell, B. (1995). *Universal access to e-mail: Feasibility and societal implications*. Rand.

Arias-de la Torre, J., Vilagut, G., Ronaldson, A., Bakolis, I., Dregan, A., Martín, V., Martinez-Alés, G., Molina, A. J., Serrano-Blanco, A., Valderas, J. M., Viana, M. C., & Alonso, J. (2023). Prevalence and variability of depressive symptoms in Europe: Update using representative data from the second and third waves of the European Health Interview Survey (EHIS-2 and EHIS-3). *The Lancet Public Health*, *8*(11), e889–e898. https://doi.org/10.1016/S2468-2667(23)00220-7

Baladram, S. (2024, November). Extra Trees, Explained: A Visual Guide with Code Examples.

Barak, A. (2007). Emotional support and suicide prevention through the Internet: A field project report. *Computers in Human Behavior*, *23*(2), 971–984. https://doi.org/10.1016/j.chb.2005.08.001

Beck, A. T. (2008). The Evolution of the Cognitive Model of Depression and Its Neurobiological Correlates [Publisher: American Psychiatric Publishing]. *American Journal of Psychiatry*, *165*(8), 969–977. https://doi.org/10.1176/appi.ajp.2008.08050721

Beck, A. T., & Bredemeier, K. (2016). A Unified Model of Depression: Integrating Clinical, Cognitive, Biological, and Evolutionary Perspectives [Publisher: SAGE Publications Inc]. *Clinical Psychological Science*, *4*(4), 596–619. https://doi.org/10.1177/2167702616628523

Beck, A. T., & Rush, A. J. (Eds.). (1979). *Cognitive therapy of depression* (13. print). Guilford Press.

Bellezza, S., Paharia, N., & Keinan, A. (2016). Conspicuous Consumption of Time: When Busyness and Lack of Leisure Time Become a Status Symbol. *Journal of Consumer Research*, *44*(1), 118–138. https://doi.org/10.1093/jcr/ucw076

Bentley, K. H., Bernstein, E. E., Wallace, B., & Mischoulon, D. (2021). Treatment for Anxiety and Comorbid Depressive Disorders: Transdiagnostic Cognitive-Behavioral Strategies. *Psychiatric annals*, *51*(5), 226–230. https://doi.org/10.3928/00485713-20210414-01

Binder, E. B., Salyakina, D., Lichtner, P., Wochnik, G. M., Ising, M., Pütz, B., Papiol, S., Seaman, S., Lucae, S., Kohli, M. A., Nickel, T., Künzel, H. E., Fuchs, B., Majer, M., Pfennig, A., Kern, N., Brunner, J., Modell, S., Baghai, T., . . . Muller-Myhsok, B. (2004). Polymorphisms in FKBP5 are associated with increased recurrence of depressive episodes and rapid response to antidepressant treatment [Publisher: Nature Publishing Group]. *Nature Genetics*, *36*(12), 1319–1325. https://doi.org/10.1038/ng1479

Bolton, D. (2023). A revitalized biopsychosocial model: Core theory, research paradigms, and clinical implications. *Psychological Medicine*, *53*(16), 7504–7511. https://doi.org/10.1017/S0033291723002660

Breiman, L. (2001). Random Forests. *Machine Learning*, *45*(1), 5–32. https://doi.org/10.1023/A:1010933404324

Breiman, L., Friedman, J., Olshen, R. A., & Stone, C. J. (1984). *Classification and Regression Trees*. Chapman and Hall/CRC. https://doi.org/10.1201/9781315139470

Brennen, J. S., & Kreiss, D. (2016, October). Digitalization. In K. B. Jensen, E. W. Rothenbuhler, J. D. Pooley, & R. T. Craig (Eds.), *The International Encyclopedia of Communication Theory and Philosophy* (1st ed., pp. 1–11). Wiley. https://doi.org/10.1002/9781118766804.wbiect111

Brough, P., Timms, C., O'Driscoll, M. P., Kalliath, T., Siu, O.-L., Sit, C., & Lo, D. (2014). Work–life balance: A longitudinal evaluation of a new measure across Australia and New Zealand workers. *The International Journal of Human Resource Management*, *25*(19), 2724–2744. https://doi.org/10.1080/09585192.2014.899262

Brough, P., Timms, C., & Shiels, R. (2009). Measuring work-life balance: Validation of a new measure across five Anglo and Asian samples. *8th Industrial and Organisational Psychology Conference (IOP)*
Accepted: 2017-05-03T14:09:50Z.

Burkov, A. (2019). *The hundred-page machine learning book*. Andriy Burkov.

Cain, N., & Gradisar, M. (2010). Electronic media use and sleep in school-aged children and adolescents: A review. *Sleep Medicine*, *11*(8), 735–742. https://doi.org/10.1016/j.sleep.2010.02.006

Cannon, W. B. (1915). *Bodily changes in pain, hunger, fear and rage: An account of recent researches into the function of emotional excitement*. D Appleton & Company. https://doi.org/10.1037/10013-000

Castells, M. (2011). *The rise of the network society* (2. ed. with a new preface, [reprint]). Wiley-Blackwell.

Chen, L. H., Law, W., Chang, D. H. F., & Sun, D. (2023). Editorial: The bio-psycho-social approach to understanding mental disorders [Publisher: Frontiers]. *Frontiers in Psychology*, *14*. https://doi.org/10.3389/fpsyg.2023.1225433

Chen, M. C., Hamilton, J. P., & Gotlib, I. H. (2010). Decreased hippocampal volume in healthy girls at risk of depression. *Archives of General Psychiatry*, *67*(3), 270–276. https://doi.org/10.1001/archgenpsychiatry.2009.202

Chen, T., & Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–794. https://doi.org/10.1145/2939672.2939785

Chicco, D., Warrens, M. J., & Jurman, G. (2021). The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. *PeerJ Computer Science*, *7*, e623. https://doi.org/10.7717/peerj-cs.623

Cho, G., Yim, J., Choi, Y., Ko, J., & Lee, S.-H. (2019). Review of Machine Learning Algorithms for Diagnosing Mental Illness. *Psychiatry Investigation*, *16*(4), 262–269. https://doi.org/10.30773/pi.2018.12.21.2

Choudhury, A., Mondal, A., & Sarkar, S. (2024, May). Searches for the BSM scenarios at the LHC using decision tree based machine learning algorithms: A comparative study and review of Random Forest, Adaboost, XGboost and LightGBM frameworks. https://doi.org/10.48550/arXiv.2405.06040

Christodoulou, E., Ma, J., Collins, G. S., Steyerberg, E. W., Verbakel, J. Y., & Van Calster, B. (2019). A systematic review shows no performance benefit of machine learning over logistic regression for clinical prediction models. *Journal of Clinical Epidemiology*, *110*, 12–22. https://doi.org/10.1016/j.jclinepi.2019.02.004

Correia, A. S., Cardoso, A., & Vale, N. (2023). BDNF Unveiled: Exploring Its Role in Major Depression Disorder Serotonergic Imbalance and Associated Stress Conditions. *Pharmaceutics*, *15*(8), 2081. https://doi.org/10.3390/pharmaceutics15082081

Coyne, J. C. (1976). Depression and the response of others [Place: US Publisher: American Psychological Association]. *Journal of Abnormal Psychology*, *85*(2), 186–193. https://doi.org/10.1037/0021-843X.85.2.186

Craske, M. G., & Stein, M. B. (2016). Anxiety. *The Lancet*, *388*(10063), 3048–3059. https://doi.org/10.1016/S0140-6736(16)30381-6

Delgado, P. L. (2000). Depression: The case for a monoamine deficiency. *The Journal of Clinical Psychiatry*, *61 Suppl 6*, 7–11.

Demetrovics, Z., Szeredi, B., & Rózsa, S. (2008). The three-factor model of Internet addiction: The development of the Problematic Internet Use Questionnaire. *Behavior Research Methods*, *40*(2), 563–574. https://doi.org/10.3758/BRM.40.2.563

Denovan, A., & Dagnall, N. (2019). Development and Evaluation of the Chronic Time Pressure Inventory. *Frontiers in Psychology*, *10*. https://doi.org/10.3389/fpsyg.2019.02717

Denovan, A., Dagnall, N., Drinkwater, K., & Escolà-Gascón, Á. (2023). Evaluating the psychometric properties of the Chronic Time Pressure Inventory using Rasch analysis. *PeerJ*, *11*, e15218. https://doi.org/10.7717/peerj.15218

Diener, E., Emmons, R. A., Larsen, R. J., & Griffin, S. (1985). The Satisfaction With Life Scale. *Journal of Personality Assessment*, *49*(1), 71–75. https://doi.org/10.1207/s15327752jpa4901_13

Dietterich, T. G. (1998). Approximate Statistical Tests for Comparing Supervised Classification Learning Algorithms. *Neural Computation*, *10*(7), 1895–1923. https://doi.org/10.1162/089976698300017197

Dietterich, T. G. (2000). Ensemble Methods in Machine Learning. *Multiple Classifier Systems*, 1–15. https://doi.org/10.1007/3-540-45014-9_1

Dissing, A. S., Andersen, T. O., Jensen, A. K., Lund, R., & Rod, N. H. (2022). Nighttime smartphone use and changes in mental health and wellbeing among young adults: A longitudinal study based on high-resolution tracking data. *Scientific Reports*, *12*, 8013. https://doi.org/10.1038/s41598-022-10116-z

Dissing, A. S., Hulvej Rod, N., Gerds, T. A., & Lund, R. (2021). Smartphone interactions and mental well-being in young adults: A longitudinal study based on objective high-resolution smartphone data. *Scandinavian Journal of Public Health*, *49*(3), 325–332. https://doi.org/10.1177/1403494820920418

Dohrenwend, B. P., Shrout, P. E., Egri, G., & Mendelsohn, F. S. (1980). Nonspecific Psychological Distress and Other Dimensions of Psychopathology: Measures for Use in the General Population. *Archives of General Psychiatry*, *37*(11), 1229–1236. https://doi.org/10.1001/archpsyc.1980.01780240027003

Dragano, N., Riedel-Heller, S. G., & Lunau, T. (2021). Haben digitale Technologien bei der Arbeit Einfluss auf die psychische Gesundheit? *Der Nervenarzt*, *92*(11), 1111–1120. https://doi.org/10.1007/s00115-021-01192-z

Drapeau, A., Marchand, A., Beaulieu-Prévost, D., Drapeau, A., Marchand, A., & Beaulieu-Prévost, D. (2012, January). Epidemiology of Psychological Distress. In *Mental Illnesses - Understanding, Prediction and Control*. IntechOpen. https://doi.org/10.5772/30872

Drevets, W. C. (2000). Neuroimaging studies of mood disorders. *Biological Psychiatry*, *48*(8), 813–829. https://doi.org/10.1016/S0006-3223(00)01020-9

Du, X., & Pang, T. Y. (2015). Is Dysregulation of the HPA-Axis a Core Pathophysiology Mediating Co-Morbid Depression in Neurodegenerative Diseases? [Publisher: Frontiers]. *Frontiers in Psychiatry*, *6*. https://doi.org/10.3389/fpsyt.2015.00032

Dwyer, D. B., Falkai, P., & Koutsouleris, N. (2018). Machine Learning Approaches for Clinical Psychology and Psychiatry. *Annual Review of Clinical Psychology*, *14*, 91–118. https://doi.org/10.1146/annurev-clinpsy-032816-045037

Elhai, J. D., Dvorak, R. D., Levine, J. C., & Hall, B. J. (2017). Problematic smartphone use: A conceptual overview and systematic review of relations with anxiety and depression psychopathology. *Journal of Affective Disorders*, *207*, 251–259. https://doi.org/10.1016/j.jad.2016.08.030

Elhai, J. D., Levine, J. C., Dvorak, R. D., & Hall, B. J. (2017). Non-social features of smartphone use are most related to depression, anxiety and problematic smartphone use. *Computers in Human Behavior*, *69*, 75–82. https://doi.org/10.1016/j.chb.2016.12.023

Ellison, N. B., Steinfield, C., & Lampe, C. (2007). The Benefits of Facebook "Friends:" Social Capital and College Students' Use of Online Social Network Sites. *Journal of Computer-Mediated Communication*, *12*(4), 1143–1168. https://doi.org/10.1111/j.1083-6101.2007.00367.x

Emmert-Streib, F., & Dehmer, M. (2019). Evaluation of Regression Models: Model Assessment, Model Selection and Generalization Error. *Machine Learning and Knowledge Extraction*, *1*(1), 521–551. https://doi.org/10.3390/make1010032

Eriksen, T. H. (2001). *Tyranny of the Moment: Fast and Slow Time in the Information Age*. Pluto Press. https://doi.org/10.2307/j.ctt183q4xv

Faravelli, C., Lo Sauro, C., Godini, L., Lelli, L., Benni, L., Pietrini, F., Lazzeretti, L., Talamba, G. A., Fioravanti, G., & Ricca, V. (2012). Childhood stressful events, HPA axis and anxiety disorders. *World Journal of Psychiatry*, *2*(1), 13–25. https://doi.org/10.5498/wjp.v2.i1.13

Faravelli, C., Lo Sauro, C., Lelli, L., Pietrini, F., Lazzeretti, L., Godini, L., Benni, L., Fioravanti, G., Talamba, G. A., Castellini, G., & Ricca, V. (2012). The role of life events and HPA axis in anxiety disorders: A review. *Current Pharmaceutical Design*, *18*(35), 5663–5674. https://doi.org/10.2174/138161212803530907

Festini, S. B., McDonough, I. M., & Park, D. C. (2016). "The busier the better: Greater busyness is associated with better cognition": Corrigendum. *Frontiers in Aging Neuroscience*, *8*.

Fioravanti, G., Casale, S., Benucci, S. B., Prostamo, A., Falone, A., Ricca, V., & Rotella, F. (2021). Fear of missing out and social networking sites use and abuse: A meta-analysis. *Computers in Human Behavior*, *122*, 106839. https://doi.org/10.1016/j.chb.2021.106839

Franchina, V., Vanden Abeele, M., Van Rooij, A. J., Lo Coco, G., & De Marez, L. (2018). Fear of Missing Out as a Predictor of Problematic Social Media Use and Phubbing Behavior among Flemish Adolescents. *International Journal of Environmental Research and Public Health*, *15*(10), 2319. https://doi.org/10.3390/ijerph15102319

Freeman, D., & Freeman, J. (2012). *Anxiety: A very short introduction*. Oxford Univ. Press.

Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *The Annals of Statistics*, *29*(5), 1189–1232. https://doi.org/10.1214/aos/1013203451

Friedman, J. H. (2002). Stochastic gradient boosting. *Computational Statistics & Data Analysis*, *38*(4), 367–378. https://doi.org/10.1016/S0167-9473(01)00065-2

Fuster, J. M. (2019, January). Chapter 1 - The prefrontal cortex in the neurology clinic. In M. D'Esposito & J. H. Grafman (Eds.), *Handbook of Clinical Neurology* (pp. 3–15, Vol. 163). Elsevier. https://doi.org/10.1016/B978-0-12-804281-6.00001-X

Gałecki, P., Mossakowska-Wójcik, J., & Talarowska, M. (2018). The anti-inflammatory mechanism of antidepressants – SSRIs, SNRIs. *Progress in Neuro-Psychopharmacology and Biological Psychiatry*, *80*, 291–294. https://doi.org/10.1016/j.pnpbp.2017.03.016

Gardner, M. J., Thomas, H. J., & Erskine, H. E. (2019). The association between five forms of child maltreatment and depressive and anxiety disorders: A systematic review and meta-analysis. *Child Abuse & Neglect*, *96*, 104082. https://doi.org/10.1016/j.chiabu.2019.104082

GBD 2019 Mental Disorders Collaborators. (2022). Global, regional, and national burden of 12 mental disorders in 204 countries and territories, 1990–2019: A systematic analysis for the Global Burden of Disease Study 2019. *The Lancet Psychiatry*, *9*(2), 137–150. https://doi.org/10.1016/S2215-0366(21)00395-3

Geurts, P., Ernst, D., & Wehenkel, L. (2006). Extremely randomized trees. *Machine Learning*, *63*(1), 3–42. https://doi.org/10.1007/s10994-006-6226-1

Ghorpade-Aher, J., Memon, A., Chugh, S., Chebolu, A., Chaudhari, P., & Chavan, J. (2023). DASS-21 Based Psychometric Prediction Using Advanced Machine Learning Techniques. *Journal of Advances in Information Technology*, *14*(3), 571–580. https://doi.org/10.12720/jait.14.3.571-580

Gimeno, D., Kivimäki, M., Brunner, E. J., Elovainio, M., De Vogli, R., Steptoe, A., Kumari, M., Lowe, G. D., Rumley, A., Marmot, M. G., & Ferrie, J. E. (2009). Associations of C-reactive protein and interleukin-6 with cognitive symptoms of depression: 12-year follow-up of the Whitehall II study. *Psychological medicine*, *39*(3), 413–423. https://doi.org/10.1017/S0033291708003723

Gleick, J. (1999). *Faster: The acceleration of just about everything* (1. ed). Pantheon Books.

Godard, R., & Holtzman, S. (2024). Are active and passive social media use related to mental health, wellbeing, and social support outcomes? A meta-analysis of 141 studies. *Journal of Computer-Mediated Communication*, *29*(1), zmad055. https://doi.org/10.1093/jcmc/zmad055

Goodwin, R. D. (2011). Association between infection early in life and mental disorders among youth in the community: A cross-sectional study. *BMC Public Health*, *11*, 878. https://doi.org/10.1186/1471-2458-11-878

Gottschalk, M. G., & Domschke, K. (2017). Genetics of generalized anxiety disorder and related traits. *Dialogues in Clinical Neuroscience*, *19*(2), 159–168. https://doi.org/10.31887/DCNS.2017.19.2/kdomschke

Granic, I., Lobel, A., & Engels, R. C. M. E. (2014). The benefits of playing video games. *American Psychologist*, *69*(1), 66–78. https://doi.org/10.1037/a0034857

Hagen, E. H., & Rosenström, T. (2016). Explaining the sex difference in depression with a unified bargaining model of anger and depression. *Evolution, Medicine, and Public Health*, *2016*(1), 117–132. https://doi.org/10.1093/emph/eow006

Harper, C. A. (2003, August). *Electronic Materials and Processes Handbook*. McGraw Hill Professional.

Hatem, L., & Ker, D. (2021, March). *Measuring well-being in the digital age* (Going Digital Toolkit Notes No. 6). https://doi.org/10.1787/1891bb63-en

Hillhouse, T. M., & Porter, J. H. (2015). A brief history of the development of antidepressant drugs: From monoamines to glutamate. *Experimental and clinical psychopharmacology*, *23*(1), 1–21. https://doi.org/10.1037/a0038550

Ho, R. C., Zhang, M. W., Tsang, T. Y., Toh, A. H., Pan, F., Lu, Y., Cheng, C., Yip, P. S., Lam, L. T., Lai, C.-M., Watanabe, H., & Mak, K.-K. (2014). The association between internet addiction and psychiatric co-morbidity: A meta-analysis. *BMC Psychiatry*, *14*(1), 183. https://doi.org/10.1186/1471-244X-14-183

Hobbs, M., Owen, S., & Gerber, L. (2017). Liquid love? Dating apps, sex, relationships and the digital transformation of intimacy. *Journal of Sociology*, *53*(2), 271–284. https://doi.org/10.1177/1440783316662718

International Telecommunication Union. (2024). Measuring digital development: Facts and Figures 2024. https://www.itu.int/hub/publication/D-IND-ICT_MDD-2024-4/

Ioannidis, K., Chamberlain, S. R., Treder, M. S., Kiraly, F., Leppink, E. W., Redden, S. A., Stein, D. J., Lochner, C., & Grant, J. E. (2016). Problematic internet use (PIU): Associations with the impulsive-compulsive spectrum. An application of machine learning in psychiatry. *Journal of Psychiatric Research*, *83*, 94–102. https://doi.org/10.1016/j.jpsychires.2016.08.010

Ioannidis, K., Treder, M. S., Chamberlain, S. R., Kiraly, F., Redden, S. A., Stein, D. J., Lochner, C., & Grant, J. E. (2018). Problematic internet use as an age-related multifaceted problem: Evidence from a two-site survey. *Addictive Behaviors*, *81*, 157–166. https://doi.org/10.1016/j.addbeh.2018.02.017

Jacobucci, R., & Grimm, K. (2020). Machine Learning and Psychological Research: The Unexplored Effect of Measurement. *Perspectives on Psychological Science*, *15*, 809–816. https://doi.org/10.1177/1745691620902467

Javaid, S. F., Hashim, I. J., Hashim, M. J., Stip, E., Samad, M. A., & Ahbabi, A. A. (2023). Epidemiology of anxiety disorders: Global burden and sociodemographic associations. *Middle East Current Psychiatry*, *30*(1), 44. https://doi.org/10.1186/s43045-023-00315-3

Jensen, M., George, M., Russell, M., & Odgers, C. (2019). Young Adolescents' Digital Technology Use and Mental Health Symptoms: Little Evidence of Longitudinal or Daily Linkages.

Clinical psychological science : a journal of the Association for Psychological Science, *7*(6), 1416–1433. https://doi.org/10.1177/2167702619859336

Joiner, T. E., & Metalsky, G. I. (1995). A prospective test of an integrative interpersonal theory of depression: A naturalistic study of college roommates [Place: US Publisher: American Psychological Association]. *Journal of Personality and Social Psychology*, *69*(4), 778–788. https://doi.org/10.1037/0022-3514.69.4.778

Jokinen, R. R., & Hartshorne, T. S. (2022). Anxiety Disorders: A Biopsychosocial Model and an Adlerian Approach for Conceptualization and Treatment [Publisher: University of Texas Press]. *The Journal of Individual Psychology*, *78*(2), 155–174. Retrieved August 7, 2025, from https://muse.jhu.edu/pub/15/article/855020

Kandel, E. R., Koester, J., Mack, S., & Siegelbaum, S. (Eds.). (2021). *Principles of neural science* (Sixth edition). McGraw Hill.

Karg, K., Burmeister, M., Shedden, K., & Sen, S. (2011). The Serotonin Transporter Promoter Variant (5-HTTLPR), Stress, and Depression Meta-Analysis Revisited: Evidence of Genetic Moderation. *Archives of general psychiatry*, *68*(5), 444–454. https://doi.org/10.1001/archgenpsychiatry.2010.189

Kaufmann, F. N., Costa, A. P., Ghisleni, G., Diaz, A. P., Rodrigues, A. L. S., Peluffo, H., & Kaster, M. P. (2017). NLRP3 inflammasome-driven pathways in depression: Clinical and preclinical findings. *Brain, Behavior, and Immunity*, *64*, 367–383. https://doi.org/10.1016/j.bbi.2017.03.002

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., & Liu, T.-Y. (2017). LightGBM: A Highly Efficient Gradient Boosting Decision Tree. *Advances in Neural Information Processing Systems*, *30*.

Kendler, K. S., & Gardner, C. O. (2014). Sex Differences in the Pathways to Major Depression: A Study of Opposite-Sex Twin Pairs [Publisher: American Psychiatric Publishing]. *American Journal of Psychiatry*, *171*(4), 426–435. https://doi.org/10.1176/appi.ajp.2013.13101375

Kessler, R. C., van Loo, H. M., Wardenaar, K. J., Bossarte, R. M., Brenner, L. A., Cai, T., Ebert, D. D., Hwang, I., Li, J., de Jonge, P., Nierenberg, A. A., Petukhova, M. V., Rosellini, A. J., Sampson, N. A., Schoevers, R. A., Wilcox, M. A., & Zaslavsky, A. M. (2016). Testing a machine-learning algorithm to predict the persistence and severity of major depressive disorder from baseline self-reports. *Molecular Psychiatry*, *21*(10), 1366–1371. https://doi.org/10.1038/mp.2015.198

Kessler, R. C. (2002). The categorical versus dimensional assessment controversy in the sociology of mental illness. *Journal of Health and Social Behavior*, *43*(2), 171–188.

Khalaf, A. M., Alubied, A. A., Khalaf, A. M., & Rifaey, A. A. (2023). The Impact of Social Media on the Mental Health of Adolescents and Young Adults: A Systematic Review. *Cureus*, *15*(8), e42990. https://doi.org/10.7759/cureus.42990

Khandaker, G. M., Pearson, R. M., Zammit, S., Lewis, G., & Jones, P. B. (2014). Association of Serum Interleukin 6 and C-Reactive Protein in Childhood With Depression and Psychosis in Young Adult Life. *JAMA psychiatry*, *71*(10), 1121–1128. https://doi.org/10.1001/jamapsychiatry.2014.1332

Kim, Y.-K., Ham, B.-J., & Han, K.-M. (2019). Interactive effects of genetic polymorphisms and childhood adversity on brain morphologic changes in depression. *Progress in Neuro-Psychopharmacology and Biological Psychiatry*, *91*, 4–13. https://doi.org/10.1016/j.pnpbp.2018.03.009

Kleiner, S. (2014). Subjective time pressure: General or domain specific? *Social Science Research*, *47*, 108–120. https://doi.org/10.1016/j.ssresearch.2014.03.013

Ko, C.-H., Yen, J.-Y., Yen, C.-F., Chen, C.-S., Weng, C.-C., & Chen, C.-C. (2008). The association between Internet addiction and problematic alcohol use in adolescents: The problem behavior model. *Cyberpsychology & Behavior: The Impact of the Internet, Multimedia and Virtual Reality on Behavior and Society*, *11*(5), 571–576. https://doi.org/10.1089/cpb.2007.0199

Köhler, O., Krogh, J., Mors, O., & Benros, M. E. (2016). Inflammation in Depression and the Potential for Anti-Inflammatory Treatment. *Current Neuropharmacology*, *14*(7), 732–742. https://doi.org/10.2174/1570159X14666151208113700

Koronczai, B., Urbán, R., Kökönyei, G., Paksi, B., Papp, K., Kun, B., Arnold, P., Kállai, J., & Demetrovics, Z. (2011). Confirmation of the Three-Factor Model of Problematic Internet Use on Off-Line Adolescent and Adult Samples. *Cyberpsychology, Behavior, and Social Networking*, *14*(11), 657–664. https://doi.org/10.1089/cyber.2010.0345

Kraut, R., Patterson, M., Lundmark, V., Kiesler, S., Mukopadhyay, T., & Scherlis, W. (1998). A Social Technology That Reduces Social Involvement and Psychological Well-Being ?

Krypotos, A.-M., Effting, M., Kindt, M., & Beckers, T. (2015). Avoidance learning: A review of theoretical models and recent developments. *Frontiers in Behavioral Neuroscience*, *9*, 189. https://doi.org/10.3389/fnbeh.2015.00189

Laconi, S., Kaliszewska-Czeremska, K., Gnisci, A., Sergi, I., Barke, A., Jeromin, F., Groth, J., Gamez-Guadix, M., Ozcan, N. K., Demetrovics, Z., Király, O., Siomos, K., Floros, G., & Kuss, D. J. (2018). Cross-cultural study of Problematic Internet Use in nine European countries. *Computers in Human Behavior*, *84*, 430–440. https://doi.org/10.1016/j.chb.2018.03.020

Laconi, S., Urbán, R., Kaliszewska-Czeremska, K., Kuss, D. J., Gnisci, A., Sergi, I., Barke, A., Jeromin, F., Groth, J., Gamez-Guadix, M., Ozcan, N. K., Siomos, K., Floros, G. D., Griffiths, M. D., Demetrovics, Z., & Király, O. (2019). Psychometric Evaluation of the Nine-Item Problematic Internet Use Questionnaire (PIUQ-9) in Nine European Samples of Internet Users. *Frontiers in Psychiatry*, *10*. https://doi.org/10.3389/fpsyt.2019.00136

Lazarus, R. S., & Folkman, S. (1984). *Stress, appraisal, and coping* (11. [print.]). Springer.

Lee, J., & Zarnic, Z. (2024, November). *The impact of digital technologies on well-being: Main insights from the literature* (OECD Papers on Well-being and Inequalities No. 29). https://doi.org/10.1787/cb173652-en

Leftheriotis, I., & Giannakos, M. N. (2014). Using social media for work: Losing your time or improving your work? *Computers in Human Behavior*, *31*, 134–142. https://doi.org/10.1016/j.chb.2013.10.016

Li, G., & He, H. (2009). Hormesis, allostatic buffering capacity and physiological mechanism of physical activity: A new theoretic framework. *Medical Hypotheses*, *72*(5), 527–532. https://doi.org/10.1016/j.mehy.2008.12.037

Lin, X., Su, W., & Potenza, M. N. (2018). Development of an Online and Offline Integration Hypothesis for Healthy Internet Use: Theory and Preliminary Evidence. *Frontiers in Psychology*, *9*. https://doi.org/10.3389/fpsyg.2018.00492

Liu, J., Chang, S., Wang, Z., & Raja, F. Z. (2024). Exploring the association between social support and anxiety during major public emergencies: A meta-analysis of the COVID-19 pandemic. *Frontiers in Public Health*, *12*, 1344932. https://doi.org/10.3389/fpubh.2024.1344932

Liu, R. T., Kleiman, E. M., Nestor, B. A., & Cheek, S. M. (2015). The Hopelessness Theory of Depression: A Quarter-Century in Review. *Clinical Psychology: Science and Practice*, *22*(4), 345–365. https://doi.org/10.1111/cpsp.12125

Lovibond, P. F., & Lovibond, S. H. (1995). The structure of negative emotional states: Comparison of the Depression Anxiety Stress Scales (DASS) with the Beck Depression and Anxiety Inventories. *Behaviour Research and Therapy*, *33*(3), 335–343. https://doi.org/10.1016/0005-7967(94)00075-U

Lu, B., Lin, L., & Su, X. (2024). Global burden of depression or depressive symptoms in children and adolescents: A systematic review and meta-analysis. *Journal of Affective Disorders*, *354*, 553–562. https://doi.org/10.1016/j.jad.2024.03.074

Lu, S., Wei, F., & Li, G. (2021). The evolution of the concept of stress and the framework of the stress system. *Cell Stress*, *5*(6), 76–85. https://doi.org/10.15698/cst2021.06.250

Lundberg, S., & Lee, S.-I. (2017, November). A Unified Approach to Interpreting Model Predictions. https://doi.org/10.48550/arXiv.1705.07874

Macan, T. H., Shahani, C., Dipboye, R. L., & Phillips, A. P. (1990). College students' time management: Correlations with academic performance and stress [Place: US Publisher: American Psychological Association]. *Journal of Educational Psychology*, *82*(4), 760–768. https://doi.org/10.1037/0022-0663.82.4.760

Macan, T. H. (1994). Time management: Test of a process model. *Journal of Applied Psychology*, *79*(3), 381–391. https://doi.org/10.1037/0021-9010.79.3.381

Mahar, I., Bambico, F. R., Mechawar, N., & Nobrega, J. N. (2014). Stress, serotonin, and hippocampal neurogenesis in relation to depression and antidepressant effects. *Neuroscience and Biobehavioral Reviews*, *38*, 173–192. https://doi.org/10.1016/j.neubiorev.2013.11.009

Martin, E. I., Ressler, K. J., Binder, E., & Nemeroff, C. B. (2009). The Neurobiology of Anxiety Disorders: Brain Imaging, Genetics, and Psychoneuroendocrinology. *The Psychiatric clinics of North America*, *32*(3), 549–575. https://doi.org/10.1016/j.psc.2009.05.004

Massé, R. (2000). Qualitative and Quantitative Analyses of Psychological Distress: Methodological Complementarity and Ontological Incommensurability. *Qualitative Health Research*, *10*(3), 411–423. https://doi.org/10.1177/104973200129118426

Maulud, D., & Abdulazeez, A. M. (2020). A Review on Linear Regression Comprehensive in Machine Learning. *Journal of Applied Science and Technology Trends*, *1*(2), 140–147. https://doi.org/10.38094/jastt1457

Mayr, A., Binder, H., Gefeller, O., & Schmid, M. (2014). The Evolution of Boosting Algorithms. *Methods of Information in Medicine*, *53*, 419–427. https://doi.org/10.3414/ME13-01-0122

Mazure, C. M. (1998). Life stressors as risk factors in depression [Place: United Kingdom Publisher: Blackwell Publishing]. *Clinical Psychology: Science and Practice*, *5*(3), 291–313. https://doi.org/10.1111/j.1468-2850.1998.tb00151.x

McEwen, B. S. (2003). Mood disorders and allostatic load. *Biological Psychiatry*, *54*(3), 200–207. https://doi.org/10.1016/s0006-3223(03)00177-x

Mikulska, J., Juszczyk, G., Gawrońska-Grzywacz, M., & Herbet, M. (2021). HPA Axis in the Pathomechanism of Depression and Schizophrenia: New Therapeutic Strategies Based on Its Participation [Number: 10 Publisher: Multidisciplinary Digital Publishing Institute]. *Brain Sciences*, *11*(10), 1298. https://doi.org/10.3390/brainsci11101298

Miller, A. H., & Raison, C. L. (2016). The role of inflammation in depression: From evolutionary imperative to modern treatment target [Publisher: Nature Publishing Group]. *Nature Reviews Immunology*, *16*(1), 22–34. https://doi.org/10.1038/nri.2015.5

Mohseni, M., Azami-Aghdash, S., Bashzar, S., Mousavi Isfahani, H., Parnian, E., & Amini-Rarani, M. (2024). An umbrella review of the prevalence of depression during the COVID-19 pandemic: Call to action for post-COVID-19 at the global level. *BMC Public Health*, *24*(1), 3562. https://doi.org/10.1186/s12889-024-21085-5

Molnar, C. (2025). *Interpretable machine learning: A guide for making black box models explainable* (3:e upplagan). Christoph Molnar.
OCLC: 1520125425.

Moncrieff, J., Cooper, R. E., Stockmann, T., Amendola, S., Hengartner, M. P., & Horowitz, M. A. (2023). The serotonin theory of depression: A systematic umbrella review of the evidence [Publisher: Nature Publishing Group]. *Molecular Psychiatry*, *28*(8), 3243–3256. https://doi.org/10.1038/s41380-022-01661-0

Nandam, L. S., Brazel, M., Zhou, M., & Jhaveri, D. J. (2020). Cortisol and Major Depressive Disorder—Translating Findings From Humans to Animal Models and Back [Publisher: Frontiers]. *Frontiers in Psychiatry*, *10*. https://doi.org/10.3389/fpsyt.2019.00974

Nasution, F. A., Effendy, E., & Amin, M. M. (2019). Internet Gaming Disorder (IGD): A Case Report of Social Anxiety. *Open Access Macedonian Journal of Medical Sciences*, *7*(16), 2664–2666. https://doi.org/10.3889/oamjms.2019.398

Nelson, J., Klumparendt, A., Doebler, P., & Ehring, T. (2017). Childhood maltreatment and characteristics of adult depression: Meta-analysis. *The British Journal of Psychiatry*, *210*(2), 96–104. https://doi.org/10.1192/bjp.bp.115.180752

Nieto, S. J., Patriquin, M. A., Nielsen, D. A., & Kosten, T. A. (2016). Don't worry; be informed about the epigenetics of anxiety. *Pharmacology, biochemistry, and behavior*, *146-147*, 60–72. https://doi.org/10.1016/j.pbb.2016.05.006

Norkeviciene, A., Gocentiene, R., Sestokaite, A., Sabaliauskaite, R., Dabkeviciene, D., Jarmalaite, S., & Bulotiene, G. (2022). A Systematic Review of Candidate Genes for Major Depression. *Medicina*, *58*(2), 285. https://doi.org/10.3390/medicina58020285

Nti, I. K., Adekoya, A. F., & Weyori, B. A. (2020). A comprehensive evaluation of ensemble learning for stock-market prediction. *Journal of Big Data*, *7*(1), 20. https://doi.org/10.1186/s40537-020-00299-5

Nuss, P. (2015). Anxiety disorders and GABA neurotransmission: A disturbance of modulation. *Neuropsychiatric Disease and Treatment*, *11*, 165–175. https://doi.org/10.2147/NDT.S58841

Odacı, H., & Çikrici, Ö. (2017). An exploration of the associations among internet use, depression, anxiety and stress among youths. *Mediterranean Journal of Clinical Psychology*, *5*(3). https://doi.org/10.6092/2282-1619/2017.5.1635

OECD. (2019, February). *How's Life in the Digital Age?: Opportunities and Risks of the Digital Transformation for People's Well-being*. https://doi.org/10.1787/9789264311800-en

Ogden, R. S. (2020). The passage of time during the UK Covid-19 lockdown (A. Nishi, Ed.). *PLOS ONE*, *15*(7), e0235871. https://doi.org/10.1371/journal.pone.0235871

Orrù, G., Monaro, M., Conversano, C., Gemignani, A., & Sartori, G. (2020). Machine Learning in Psychometrics and Psychological Research. *Frontiers in Psychology*, *10*. https://doi.org/10.3389/fpsyg.2019.02970

Ozili, P. K. (2022, June). The Acceptable R-Square in Empirical Modelling for Social Science Research. https://doi.org/10.2139/ssrn.4128165

Pabba, M. (2013). Evolutionary development of the amygdaloid complex [Publisher: Frontiers]. *Frontiers in Neuroanatomy*, *7*. https://doi.org/10.3389/fnana.2013.00027

Paige, M. (2024, May). The Evolution of Digital Transformation History: From Pre-Internet to Generative AI.

Palazidou, E. (2012). The neurobiology of depression. *British Medical Bulletin*, *101*(1), 127–145. https://doi.org/10.1093/bmb/lds004

Pouwels, J. L., Valkenburg, P. M., Beyens, I., van Driel, I. I., & Keijsers, L. (2021). Some socially poor but also some socially rich adolescents feel closer to their friends after using social media. *Scientific Reports*, *11*(1), 21176. https://doi.org/10.1038/s41598-021-99034-0

Price, J. S. (2003). Evolutionary aspects of anxiety disorders. *Dialogues in Clinical Neuroscience*, *5*(3), 223–236. https://doi.org/10.31887/DCNS.2003.5.3/jprice

Pruessner, M., Cullen, A. E., Aas, M., & Walker, E. F. (2017). The neural diathesis-stress model of schizophrenia revisited: An update on recent findings considering illness stage and neurobiological and methodological complexities. *Neuroscience & Biobehavioral Reviews*, *73*, 191–218. https://doi.org/10.1016/j.neubiorev.2016.12.013

Przybylski, A. K., Murayama, K., DeHaan, C. R., & Gladwell, V. (2013). Motivational, emotional, and behavioral correlates of fear of missing out. *Computers in Human Behavior*, *29*(4), 1841–1848. https://doi.org/10.1016/j.chb.2013.02.014

Przybylski, A. K., & Weinstein, N. (2017). A Large-Scale Test of the Goldilocks Hypothesis: Quantifying the Relations Between Digital-Screen Use and the Mental Well-Being of Adolescents. *Psychological Science*, *28*(2), 204–215. https://doi.org/10.1177/0956797616678438

Qu, K. (2024). Research on linear regression algorithm (J. Zhu, Ed.). *MATEC Web of Conferences*, *395*, 01046. https://doi.org/10.1051/matecconf/202439501046

Raison, C. L., Lowry, C. A., & Rook, G. A. W. (2010). Inflammation, Sanitation, and Consternation. *Archives of general psychiatry*, *67*(12), 1211–1224. https://doi.org/10.1001/archgenpsychiatry.2010.161

Rao, U., Chen, L.-A., Bidesi, A. S., Shad, M. U., Thomas, M. A., & Hammen, C. L. (2010). Hippocampal changes associated with early-life adversity and vulnerability to depression. *Biological Psychiatry*, *67*(4), 357–364. https://doi.org/10.1016/j.biopsych.2009.10.017

Rosa, H. (2013). *Social Acceleration: A New Theory of Modernity*. Columbia University Press.

Rudd, M. (2019). Feeling short on time: Trends, consequences, and possible remedies. *Current Opinion in Psychology*, *26*, 5–10. https://doi.org/10.1016/j.copsyc.2018.04.007

Rutrecht, H., Wittmann, M., Khoshnoud, S., & Igarzábal, F. A. (2021). Time Speeds Up During Flow States: A Study in Virtual Reality with the Video Game Thumper. https://doi.org/10.1163/22134468-bja10033

SAMSHA. (2024, August). *Key substance use and mental health indicators in the United States: Results from the 2023 national survey on drug use and health.* (Report) (Num Pages: 146 p.). SAMSHA. Rockville, MD. Retrieved July 21, 2025, from https://www.drugsandalcohol.ie/41611/

Santana, A. N., de Santana, C. N., & Montoya, P. (2020). Chronic Pain Diagnosis Using Machine Learning, Questionnaires, and QST: A Sensitivity Experiment. *Diagnostics*, *10*(11), 958. https://doi.org/10.3390/diagnostics10110958

Schmidt, C. K., Khalid, S., Loukas, M., & Tubbs, R. S. (2018). Neuroanatomy of Anxiety: A Brief Review. *Cureus*, *10*(1), e2055. https://doi.org/10.7759/cureus.2055

Segrin, C., & Dillard, J. P. (1992). The Interactional Theory of Depression: A Meta-Analysis of the Research Literature [Publisher: Guilford Publications Inc.]. *Journal of Social and Clinical Psychology*, *11*(1), 43–70. https://doi.org/10.1521/jscp.1992.11.1.43

Seidler, A., Schubert, M., Freiberg, A., Drössler, S., S. Hussenoeder, F., Conrad, I., Riedel-Heller, S., & Romero Starke, K. (2022). Psychosocial Occupational Exposures and Mental Illness. *Deutsches Ärzteblatt International*, *119*(42), 709–715. https://doi.org/10.3238/arztebl.m2022.0295

Seligman, M. E. P. (1972). Learned Helplessness. *Annual Review of Medicine*, *23*(Volume 23, 1972), 407–412. https://doi.org/10.1146/annurev.me.23.020172.002203

Selye, H. (1946). The general adaptation syndrome and the diseases of adaptation. *The Journal of Clinical Endocrinology and Metabolism*, *6*, 117–230. https://doi.org/10.1210/jcem-6-2-117

Selye, H. (1936). A Syndrome produced by Diverse Nocuous Agents. *Nature*, *138*(3479), 32–32. https://doi.org/10.1038/138032a0

Selye, H. (1974). *Stress without distress* (1st ed.). Lippincott.

Shapley, L. S. (1953, December). 17. A Value for n-Person Games. In H. W. Kuhn & A. W. Tucker (Eds.), *Contributions to the Theory of Games (AM-28), Volume II* (pp. 307–318). Princeton University Press. https://doi.org/10.1515/9781400881970-018

Shipp, A. J., Edwards, J. R., & Lambert, L. S. (2009). Conceptualization and measurement of temporal focus: The subjective experience of the past, present, and future. *Organizational Behavior and Human Decision Processes*, *110*(1), 1–22. https://doi.org/10.1016/j.obhdp.2009.05.001

Sim, M. K. (2022, January). Chapter 2 - Explanation using model-agnostic methods. In C. S. Nam, J.-Y. Jung, & S. Lee (Eds.), *Human-Centered Artificial Intelligence* (pp. 17–31). Academic Press. https://doi.org/10.1016/B978-0-323-85648-5.00008-6

Smedslund, G., Arnulf, J. K., & Smedslund, J. (2022). Smedslund et al. - 2022 - Is psychological science progressing Explained variance in PsycINFO articles during the period 1956. *Frontiers in Psychology*, *13*. https://doi.org/10.3389/fpsyg.2022.1089089

Smoller, J. W. (2016). The Genetics of Stress-Related Disorders: PTSD, Depression, and Anxiety Disorders. *Neuropsychopharmacology*, *41*(1), 297–319. https://doi.org/10.1038/npp.2015.266

Spada, M. M. (2014). An overview of problematic Internet use. *Addictive Behaviors*, *39*(1), 3–6. https://doi.org/10.1016/j.addbeh.2013.09.007

Srividya, M., Mohanavalli, S., & Bhalaji, N. (2018). Behavioral Modeling for Mental Health using Machine Learning Algorithms. *Journal of Medical Systems*, *42*(5), 88. https://doi.org/10.1007/s10916-018-0934-5

Steel, P. (2010). Arousal, avoidant and decisional procrastinators: Do they exist? *Personality and Individual Differences*, *48*(8), 926–934. https://doi.org/10.1016/j.paid.2010.02.025

Stieglitz, J., Trumble, B. C., Thompson, M. E., Blackwell, A. D., Kaplan, H., & Gurven, M. (2015). DEPRESSION AS SICKNESS BEHAVIOR? A TEST OF THE HOST DEFENSE HYPOTHESIS IN A HIGH PATHOGEN POPULATION. *Brain, behavior, and immunity*, *49*, 130–139. https://doi.org/10.1016/j.bbi.2015.05.008

Suárez Álvarez, A., & Vicente, M. R. (2024). Is Too Much Time on the Internet Making us Less Satisfied with Life? *Applied Research in Quality of Life*, *19*(5), 2245–2265. https://doi.org/10.1007/s11482-024-10326-9

Sun, F., Shuai, Y., Wang, J., Yan, J., Lin, B., Li, X., & Zhao, Z. (2025). Hippocampal gray matter volume alterations in patients with first-episode and recurrent major depressive disorder and their associations with gene profiles. *BMC Psychiatry*, *25*, 134. https://doi.org/10.1186/s12888-025-06562-4

Svartdal, F., & Steel, P. (2017). Irrational Delay Revisited: Examining Five Procrastination Scales in a Global Sample. *Frontiers in Psychology*, *8*. https://doi.org/10.3389/fpsyg.2017.01927

Szollos, A. (2009). Toward a psychology of chronic time pressure: Conceptual and methodological review. *Time & Society*, *18*(2-3), 332–350. https://doi.org/10.1177/0961463X09337847

Tate, A. E., McCabe, R. C., Larsson, H., Lundström, S., Lichtenstein, P., & Kuja-Halkola, R. (2020). Predicting mental health problems in adolescence using machine learning techniques. *PLOS ONE*, *15*(4), e0230389. https://doi.org/10.1371/journal.pone.0230389

The Lancet Psychiatry (Ed.). (2024). Global Burden of Disease 2021: Mental health messages [Publisher: Elsevier]. *The Lancet Psychiatry*, *11*(8), 573. https://doi.org/10.1016/S2215-0366(24)00222-0

Thomée, S., Härenstam, A., & Hagberg, M. (2011). Mobile phone use and stress, sleep disturbances, and symptoms of depression among young adults - a prospective cohort study. *BMC Public Health*, *11*(1), 66. https://doi.org/10.1186/1471-2458-11-66

Tian, X., Li, H., Li, F., Jiménez-Herrera, M. F., Ren, Y., & Shang, H. (2024). Development and validation of a web-based calculator for determining the risk of psychological distress based on machine learning algorithms: A cross-sectional study of 342 lung cancer patients. *Supportive Care in Cancer*, *33*(1), 63. https://doi.org/10.1007/s00520-024-09127-5

Topp, C. W., Østergaard, S. D., Søndergaard, S., & Bech, P. (2015). The WHO-5 Well-Being Index: A systematic review of the literature. *Psychotherapy and Psychosomatics*, *84*(3), 167–176. https://doi.org/10.1159/000376585

Travkina, N. M. (2022). Digitization of Society: Alternative Projections of the Future. *Herald of the Russian Academy of Sciences*, *92*(Suppl 6), S483–S491. https://doi.org/10.1134/S1019331622120115

Trifu, S. C., Trifu, A. C., Aluaş, E., Tătaru, M. A., & Costea, R. V. (2020). Brain changes in depression. *Romanian Journal of Morphology and Embryology*, *61*(2), 361–370. https://doi.org/10.47162/RJME.61.2.06

Tsigos, C., & Chrousos, G. P. (2002). Hypothalamic–pituitary–adrenal axis, neuroendocrine factors and stress. *Journal of Psychosomatic Research*, *53*(4), 865–871. https://doi.org/10.1016/S0022-3999(02)00429-4

Twenge, J. M., & Campbell, W. K. (2018). Associations between screen time and lower psychological well-being among children and adolescents: Evidence from a population-based study. *Preventive Medicine Reports*, *12*, 271–283. https://doi.org/10.1016/j.pmedr.2018.10.003

Vally, Z., Moussa, D., Khalil, E., Al Fahel, A., Al Azry, N., & Jafar, N. (2021). Celebrity worship in the United Arab Emirates: An examination of its association with problematic internet use, maladaptive daydreaming, and desire for fame. *Psychology of Popular Media*, *10*(1), 124–134. https://doi.org/10.1037/ppm0000276

Viertiö, S., Kiviruusu, O., Piirtola, M., Kaprio, J., Korhonen, T., Marttunen, M., & Suvisaari, J. (2021). Factors contributing to psychological distress in the working population, with a special reference to gender difference. *BMC Public Health*, *21*(1), 611. https://doi.org/10.1186/s12889-021-10560-y

Wajcman, J. (2008). Life in the fast lane? Towards a sociology of technology and time. *The British Journal of Sociology*, *59*(1), 59–77. https://doi.org/10.1111/j.1468-4446.2007.00182.x

Wajcman, J. (2015). *Pressed for Time: The Acceleration of Life in Digital Capitalism*. University of Chicago Press.

Wang, Y., King, R., & Leung, S. O. (2023). Understanding Chinese Students' Well-Being: A Machine Learning Study. *Child Indicators Research*, *16*(2), 581–616. https://doi.org/10.1007/s12187-022-09997-3

Warner-Schmidt, J. L., Vanover, K. E., Chen, E. Y., Marshall, J. J., & Greengard, P. (2011). Antidepressant effects of selective serotonin reuptake inhibitors (SSRIs) are attenuated by antiinflammatory drugs in mice and humans. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(22), 9262–9267. https://doi.org/10.1073/pnas.1104836108

Watson, D. S., Krutzinna, J., Bruce, I. N., Griffiths, C. E., McInnes, I. B., Barnes, M. R., & Floridi, L. (2019). Clinical applications of machine learning algorithms: Beyond the black box. *BMJ*, l886. https://doi.org/10.1136/bmj.l886

Whang, L. S.-M., Lee, S., & Chang, G. (2003). Internet Over-Users' Psychological Profiles: A Behavior Sampling Analysis on Internet Addiction. *CyberPsychology & Behavior*, *6*(2), 143–150. https://doi.org/10.1089/109493103321640338

Witowska, J., Ogden, R., Schoetensack, C., Goncikowska, K., Wittmann, M., Černohorská, V., Codina, N., Martin-Soelch, C., Fernández Boente, M., Giner, G., Pestana, J. V., & Papastamatelou, J. (2025). New measurements of digital technology use: The Immersion in Digital Life and Quality of Digital Experience scales [Publisher: Frontiers]. *Frontiers in Psychiatry*, *16*. https://doi.org/10.3389/fpsyt.2025.1595536

Wittmann, M., & Lehnhoff, S. (2005). Age effects in perception of time. *Psychological Reports*, *97*(3), 921–935. https://doi.org/10.2466/pr0.97.3.921-935

Wolpert, D. H. (1992). Stacked generalization. *Neural Networks*, *5*(2), 241–259. https://doi.org/10.1016/S0893-6080(05)80023-1

Won, E., & Kim, Y.-K. (2020). Neuroinflammation-Associated Alterations of the Brain as Potential Neural Biomarkers in Anxiety Disorders. *International Journal of Molecular Sciences*, *21*(18), 6546. https://doi.org/10.3390/ijms21186546

Wu, L. (2013). Social Network Effects on Productivity and Job Security: Evidence from the Adoption of a Social Networking Tool. *Information Systems Research*, *24*(1), 30–51. https://doi.org/10.1287/isre.1120.0465

Yokoyama, C., Kaiya, H., Kumano, H., Kinou, M., Umekage, T., Yasuda, S., Takei, K., Nishikawa, M., Sasaki, T., Nishimura, Y., Hara, N., Inoue, K., Kaneko, Y., Suzuki, S.-i., Tanii, H., Okada, M., & Okazaki, Y. (2015). Dysfunction of ventrolateral prefrontal cortex underlying social anxiety disorder: A multi-channel NIRS study. *NeuroImage: Clinical*, *8*, 455–461. https://doi.org/10.1016/j.nicl.2015.05.011

Young, K., & Rogers, R. C. (1998). The Relationship Between Depression and Internet Addiction. *CyberPsychology & Behavior*, *1*(1), 25–28. https://doi.org/10.1089/cpb.1998.1.25

Zangrossi, H., Del Ben, C. M., Graeff, F. G., & Guimarães, F. S. (2020, January). Chapter 36 - Serotonin in panic and anxiety disorders. In C. P. Müller & K. A. Cunningham (Eds.), *Handbook of Behavioral Neuroscience* (pp. 611–633, Vol. 31). Elsevier. https://doi.org/10.1016/B978-0-444-64125-0.00036-0

Zhao, Y., & Wang, M. (2023). Digital sociology: Origin, development, and prospects from a global perspective. *The Journal of Chinese Sociology*, *10*(1), 19. https://doi.org/10.1186/s40711-023-00198-1

Zhou, M., Li, F., Wang, Y., Chen, S., & Wang, K. (2020). Compensatory Social Networking Site Use, Family Support, and Depression Among College Freshman: Three-Wave Panel Study. *Journal of Medical Internet Research*, *22*(9), e18458. https://doi.org/10.2196/18458

Zimmermann, P., Brückl, T., Nocon, A., Pfister, H., Binder, E. B., Uhr, M., Lieb, R., Moffitt, T. E., Caspi, A., Holsboer, F., & Ising, M. (2011). Interaction of FKBP5 gene variants and adverse life events in predicting depression onset: Results from a 10-year prospective community study. *The American Journal of Psychiatry*, *168*(10), 1107–1116. https://doi.org/10.1176/appi.ajp.2011.10111577

AI Utilization

The following points describe the AI tools used during the project and their specific purposes. The author takes full responsibility for all content in this thesis.

ChatGPT (GPT-4o)

- Debugging python scripts or environments.
- Clarifying code documentation.
- Rephrasing and decluttering sentences.
- Summarizing research papers to assess for relevance.

GitHub Copilot

- Intelligent auto-completion.

Statement of Independence

I hereby certify that I have written this Master thesis independently without the help of third parties and without using any sources or aids other than those indicated.

14.08.2025, Diogo Alexandre Rocha Moreira

A Python version and libraries

All analyses were performed using Python 3.11.8 in Visual Studio Code. The major Python libraries used in this study are:

- jupyter 1.1.1 (https://jupyter.org/)
- matplotlib 3.7.5 (https://matplotlib.org/)
- pandas 2.1.4 (https://pandas.pydata.org/)
- plotly 5.24.1 (https://plotly.com/)
- numpy 1.26.4 (https://numpy.org/)
- scikit-learn 1.4.2 (https://scikit-learn.org/)
- shap 0.46.0 (https://shap.readthedocs.io)
- streamlit 1.40.2 (https://streamlit.io/)
- pycaret 3.3.2 (https://pycaret.org/)
- xgboost 2.1.3 (https://xgboost.readthedocs.io/)
- lightgbm 4.5.0 (https://lightgbm.readthedocs.io/)
- optuna 4.1.0 (https://optuna.org/)

B Hyperparameter Spaces

Random Forest (RF)

Parameter	Type	Range / Values
RF__n_estimators	Integer	100 to 2000 (step=100)
RF__max_depth	Integer	5 to 30
RF__min_samples_split	Integer	2 to 10
RF__min_samples_leaf	Integer	1 to 10
RF__bootstrap	Categorical	{True, False}

Extra Trees (ET)

Parameter	Type	Range / Values
ET__n_estimators	Integer	100 to 2000 (step=100)
ET__max_depth	Integer	5 to 30
ET__min_samples_split	Integer	2 to 10
ET__min_samples_leaf	Integer	1 to 10
ET__bootstrap	Categorical	{True, False}

XGBoost (XGB)

Parameter	Type	Range / Values
XGB__objective	Categorical	{'reg:squarederror'}
XGB__booster	Categorical	{'gbtree'}
XGB__tree_method	Categorical	{'approx', 'hist'}
XGB__n_estimators	Integer	100 to 2000 (step=100)
XGB__max_depth	Integer	5 to 30
XGB__learning_rate	Float (log)	10^{-3} to 0.3
XGB__min_child_weight	Integer	1 to 100
XGB__subsample	Float	0.5 to 1.0 (step=0.1)
XGB__colsample_bytree	Float	0.5 to 1.0 (step=0.1)
XGB__gamma	Float	0.0 to 5.0 (step=0.1)
XGB__lambda	Float (log)	10^{-8} to 10.0
XGB__alpha	Float (log)	10^{-8} to 10.0

LightGBM (LGB)

Parameter	Type	Range / Values
LGB__boosting_type	Categorical	{'gbdt', 'dart'}
LGB__n_estimators	Integer	100 to 2000 (step=100)
LGB__max_depth	Integer	5 to 30
LGB__learning_rate	Float (log)	10^{-3} to 0.3
LGB__num_leaves	Integer	31 to 512
LGB__min_child_samples	Integer	5 to 100
LGB__subsample	Float	0.5 to 1.0 (step=0.1)
LGB__colsample_bytree	Float	0.5 to 1.0 (step=0.1)
LGB__reg_alpha	Float (log)	10^{-8} to 10.0
LGB__reg_lambda	Float (log)	10^{-8} to 10.0
LGB__min_split_gain	Float	0.0 to 1.0
LGB__bagging_freq	Integer	1 to 10
LGB__max_bin	Integer	128 to 512

C Best Hyperparameters for the DASS Task

Extra Trees (ET)

Parameter	Best Value
ET__n_estimators	1800
ET__max_depth	21
ET__min_samples_split	4
ET__min_samples_leaf	1
ET__bootstrap	False

XGBoost (XGB)

Parameter	Best Value
XGB__objective	reg:squarederror
XGB__booster	gbtree
XGB__tree_method	approx
XGB__n_estimators	1700
XGB__max_depth	20
XGB__learning_rate	0.00330
XGB__min_child_weight	24
XGB__subsample	0.5
XGB__colsample_bytree	0.6
XGB__gamma	0.7
XGB__lambda	9.69e-07
XGB__alpha	1.92e-07

D Best Hyperparameters for the CTPI Task

Extra Trees (ET)

Parameter	Best Value
ET__n_estimators	1000
ET__max_depth	17
ET__min_samples_split	4
ET__min_samples_leaf	1
ET__bootstrap	True

LightGBM (LGB)

Parameter	Best Value
LGB__boosting_type	gbdt
LGB__n_estimators	2000
LGB__max_depth	30
LGB__learning_rate	0.01008
LGB__num_leaves	348
LGB__min_child_samples	38
LGB__subsample	0.6
LGB__colsample_bytree	0.6
LGB__reg_alpha	0.5938
LGB__reg_lambda	0.000118
LGB__min_split_gain	0.2267
LGB__bagging_freq	7
LGB__max_bin	384