1 2 9 0

UNIVERSIDADE Đ COIMBRA

Beatriz Pereira Patrício

CHARACTERIZING ACCEPTANCE OF INTERNET OF THINGS SOLUTIONS FOR ACTIVE AND HEALTHY AGEING

Dissertation in the context of the Master in Data Science and Engineering, supervised by Prof. Paulo de Carvalho and Prof. Jorge Henriques from the University of Coimbra and co-supervised by Prof. Giuseppe Fico and Dr. Patricia Abril from the Polytechnic University of Madrid and presented to the Department of Informatics Engineering of the Faculty of Sciences and Technology of the University of Coimbra.

September of 2022

FACULDADE DE
CIÊNCIAS E TECNOLOGIA
UNIVERSIDADE D
COIMBRA

DEPARTMENT OF INFORMATICS ENGINEERING

Beatriz Pereira Patrício

Characterizing acceptance of Internet of Things solutions for Active and Healthy Ageing

Dissertation in the context of the Master in Data Science and Engineering, supervised by Prof. Paulo de Carvalho and Prof. Jorge Henriques from the University of Coimbra and co-supervised by Prof. Giuseppe Fico and Dr. Patricia Abril from the Polytechnic University of Madrid and presented to the Department of Informatics Engineering of the Faculty of Sciences and Technology of the University of Coimbra.

September 2022

1 2 9 0

FACULDADE DE
CIÊNCIAS E TECNOLOGIA
UNIVERSIDADE D
COIMBRA

Departamento de Engenharia Informática

Beatriz Pereira Patrício

Characterizing acceptance of Internet of Things solutions for Active and Healthy Ageing

Dissertação no âmbito do Mestrado em Engenharia e Ciência de Dados, orientada pelo Professor Doutor Paulo de Carvalho e pelo Professor Doutor Jorge Henriques da Universidade de Coimbra e co-orientada pelo Professor Doutor Giuseppe Fico e pela Doutora Patricia Abril da Universidad Politécnica de Madrid e apresentada ao Departamento de Engenharia Informática da Faculdade de Ciências e Tecnologia da Universidade de Coimbra.

Setembro 2022

Agradecimentos

Em primeiro lugar gostaria de agradecer aos meus orientadores da Universidade de Coimbra, Professor Doutor Paulo de Carvalho e Professor Doutor Jorge Henriques, pela oportunidade de participar num projeto tão interessante, mas, acima de tudo, pelo apoio e pela paciência ao longo do último ano. Gostaria também de agradecer aos co-orientadores da Universidad Politécnica de Madrid, Doutora Patricia Abril e Professor Doutor Giuseppe Fico, pelo interesse e disponibilidade que demonstraram e por estarem sempre prontos a ajudar.

Aos meus amigos e colegas de mestrado, um obrigada pelo apoio nos últimos dois anos. O caminho foi mais fácil e mais agradável com vocês ao lado. Em especial, à Beatriz e ao Guilherme, parte do trio improvável que resultou melhor do que alguma vez poderia imaginar. Obrigada por terem tornado os projetos e as noitadas menos stressantes e mais divertidos.

Ao Coro Misto da Universidade de Coimbra, a melhor surpresa do meu 2º ano de mestrado. Com vocês descobri a magia de Coimbra e ganhei uma família nesta cidade. Espero que continuemos nas vidas uns dos outros por muitos e muitos anos.

As minhas últimas palavras vão para a minha família. Em especial para os meus pais, avós e tia-madrinha, a quem agradeço muito todo o carinho, ajuda e motivação que me deram ao longo da vida. Sem os vossos esforços e apoio este trabalho não seria possível.

Abstract

The digitalization of healthcare has the potential to alleviate the burden on national health systems due to the aging challenge. However, users may abandon a technological application or solution at any time due to a multitude of reasons, which could reduce the effectiveness of the intervention, and even increase the health problem or disease-associated risks. Therefore, it is of the utmost importance to identify users with higher risks of getting disengaged from a solution, even predicting when the disengagement could happen. This creates the opportunity of applying tailored personalized intervention strategies aiming at recovering from low adherence and prevention of dropout. In addition, this prediction could also reveal insight into which factors are causing attrition to adherence, which could allow for the implementation of more global strategies.

The main goal of this study is to research the factors, barriers and needs of aging users of healthcare technologies that may contribute to the early detection of lack of motivation and consequent disengagement. In particular, the goal is to research and develop models that 1) can detect early dropout patterns and 2) provide an explanation of this lack of engagement that allows the introduction of tailored interventions.

The study was supported by the ACTIVAGE Madrid Deployment Site database currently available in the Universidad Politécnica de Madrid (UPM), which includes the data of about 800 participants in a pilot project on the use of technology to promote healthy and active aging.

To achieve the proposed goal, a data science approach is used and three main models were created: a Regression Model for the prediction of an adherence percentage and two Binary Classification Models for the prediction of an adherence level (low or high), the second of which filtered for only when the period preceding the prediction had a high adherence level.

The final classifier achieved a F1-score of 0.81126 using a Random Forest Classifier and allowed for the inference of the most relevant features for the decrease of adherence, using SHAP values.

Keywords

Adherence, eHealth, Aging, Data Science, Prediction, Classifiers, Machine Learning, Interpretability, Shapley

Resumo

A digitalização dos cuidados de saúde tem o potencial de aliviar a carga sobre sistemas nacionais de saúde causado pelo envelhecimento da população. No entanto, os utilizadores podem abandonar uma aplicação ou solução tecnológica a qualquer momento por uma variado número de razões, o que pode reduzir a eficácia da intervenção e até mesmo aumentar o problema de saúde ou os riscos associados à doença. Neste sentido, é de extrema importância identificar os utilizadores com maiores riscos de deixarem de utilizar uma solução, e até mesmo prever quando a quebra da adesão poderá acontecer. Esta previsão gera a oportunidade de aplicar estratégias de intervenção personalizadas com o objetivo de recuperação da baixa adesão e prevenção do abandono da solução. Além disso, a previsão poderá ainda revelar insights sobre quais fatores que estão a criar atrito à adesão, o que pode permitir a implementação de estratégias mais globais.

O objetivo principal deste trabalho é pesquisar os fatores, barreiras e necessidades dos usuários de tecnologias de saúde que possam contribuir para a detecção precoce da falta de motivação e consequente abandono. Em particular, o objetivo é pesquisar e desenvolver modelos que 1) possam detectar padrões de abandono precoce e 2) fornecer uma explicação para essa falta de motivação para uso da solução que permita a introdução de intervenções personalizadas.

Este estudo foi apoiado pela base de dados ACTIVAGE Madrid Deployment Site atualmente disponível na Universidad Politécnica de Madrid (UPM), que inclui os dados de cerca de 800 participantes num projeto piloto sobre o uso da tecnologia para promover o envelhecimento saudável e ativo.

Para atingir o objetivo proposto, é utilizada uma abordagem de ciência de dados e foram criados três modelos principais: um Modelo de Regressão para a previsão de uma percentagem de adesão e dois Modelos de Classificação Binária para a previsão de um nível de adesão (baixo ou alto), o segundo dos quais filtrado para apenas quando o período anterior à previsão teve um nível de adesão alto.

O classificador final obteve um F1-score de 0,81126 usando um Random Forest Classifier e permitiu a inferência das características mais relevantes para a diminuição da adesão, usando valores SHAP.

Palavras-Chave

Adesão, eHealth, Envelhecimento, Ciência de Dados, Previsão, Classificadores, Inteligência Artificial, Interpretabilidade, Shapley

Contents

Acronyms

AHA Active and Healthy Ageing.

AI Artificial Intelligence.

DS Deployment Sites.

EDA Exploratory Data Analysis.

eHealth electronic health.

FFS Forward Feature Selection.

IoT Internet of Things.

KNN K-Neighbors Classifier.

LR Logistic Regression.

MLP Multi-Layer Perceptron.

NB Gaussian Naive Bayes.

RF Random Forest Regressor.

RMSE Root Mean Square Error.

SVM Support Vector Regressor.

UC Use Cases.

UPM Universidad Politécnica de Madrid.

WHO World Health Organization.

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Context and Motivations

The digitalization of healthcare has the potential to alleviate the burden on national health systems due to the aging challenge. According to the World Health Organization (WHO), more than a quarter of the world's countries have a critical healthcare workforce shortage [Hoque and Sorwar, 2017].

[Parra et al., 2014]

However, users may abandon a technological application or solution at any time due to a multitude of reasons, which could reduce the effectiveness of the intervention, and even increase the health problem or disease-associated risks. Therefore, it is of the utmost importance to identify users with higher risks of getting disengaged from a solution, even predicting when the disengagement could happen. This creates the opportunity of applying tailored personalized intervention strategies aiming at recovering from low adherence and prevention of dropout. In addition, this prediction could also reveal insight into which factors are causing attrition to adherence, which could allow for the implementation of more global strategies.

In this line, the main goal of this master thesis is to research the factors, barriers and needs of aging users of healthcare technologies that may contribute to the early detection of lack of motivation and consequent disengagement. In particular, the goal is to research and develop models that 1) can detect early dropout patterns and 2) provide a profile-based explanation of this lack of engagement that allows the introduction of tailored interventions.

The study will be supported by the ACTIVAGE Madrid Deployment Site database currently available in the Universidad Politécnica de Madrid (UPM), which includes the data of about 800 participants in a pilot project on the use of technology to promote healthy and active aging. [Fico et al., 2017]

1.2 Objectives

The main overarching goals and objectives for this thesis are the following:

- Research and develop a model for detection of the early dropout patterns.
- Support better understanding of the medium and long-term digital healthcare solutions acceptance and providing evidence of the early dropout causes.
- Develop a set of recommendations to apply intervention strategies aiming at recovering from disengagement.

1.3 Document Structure

The remainder of this document is structured as follows: Chapter 2 provides some general insight on previous research on the concepts of adherence and dropout and their use relating to technology, the current work on prediction of adherence of electronic health (eHealth) solutions as well as an overview of technical aspects, including the data science pipeline stages and interpretability on Machine Learning Models. Chapter 3 explains the context of the data acquisition, describes the dataset and describes the steps of the data science process that will be followed. Chapter 4 describes the data preparation steps, including data cleaning and feature engineering and presents the models that were built, along with the obtained results. Finally, Chapter 5 draws some conclusions from the project and presents suggestions of future work.

Chapter 2

State of the Art

2.1 Adherence and Dropout in Technology

2.1.1 Definitions of Adherence

The term adherence is strongly connected to the pharmaceutical industry (as in "adherence to medication") but can, nonetheless, be used in a broader sense. According to the World Health Organization's (WHO) definition, adherence to long-term therapy can be defined as "the extent to which a person's behaviour – taking medication, following a diet, and/or executing lifestyle changes, corresponds with agreed recommendations from a health care provider" [Sabaté et al., 2003].

When it comes to technology, however, specifically to electronic health (eHealth), there is not one single agreed upon definition, formula or way of measuring adherence.

Some studies choose to focus on "adherence metrics" and try to correlate them with the intended outcomes. [Donkin et al., 2011] captured the adherence data that was used across several studies. The metrics included "reporting the number of times the participant accessed or logged into the program, completed modules or activities, visits made to forums, posts made to the forum, and pages viewed and printed, as well as self-reported completion of activities away from the program or offline". However, only half of the studies presented the adherence data in relation to outcome measures. In [Horsch et al., 2015] some other metrics were suggested, like "the usage time of the technology and reports by a spouse or related others". On the other hand, [Evans et al., 2016] simply used "the percentages of data collected [which] were calculated by comparing the number of minutes, or days, of that particular type of data divided by the number of minutes, or days, that participants were involved in the study" to measure adherence.

[Sieverink et al., 2017] looked at several definitions of adherence relating to eHealth technology in the existing literature. [Christensen et al., 2009] defined adherence as "the degree to which individuals experience the content of the Internet intervention", which is missing the concept of "prescribed recommendations" from the WHO definition. [Donkin et al., 2011] referred to adherence as "the degree to which the user followed the program as it was designed”. This definition encompasses the concept of intended use. [Sieverink et al., 2017] concluded, therefore, that "the intended use is thus the minimum use to establish adherence". However, even using this concept, it can still be a challenge to operationalize the intended use for individual eHealth technologies.

Finally, and in conclusion, [Sieverink et al., 2017] suggest that the following three elements are necessary to determine adherence to eHealth technology:

1. the ability to measure the usage behavior of individuals,
2. an operationalization of intended use, and
3. an empirical, theoretical, or rational justification of the intended use.

In terms of adherence operationalizations, they divide the analyzed literature into three categories:

- **Category A**: When adherence was operationalized in terms of “the more usage, the better.” This does not include an operationalization of intended use, and therefore does not comply with Category B.
- **Category B**: Assigned when the intended use of a technology was provided but no justification was attached (eg, “a user is adherent when logging in at least once a week for three subsequent weeks”).
- **Category C**: Assigned when not only was the intended use of the technology provided, but a justification was given using theory, evidence, or rationale (eg, “we know from previous research that users benefit the most from the technology when finishing module 4, so a user is adherent once module 4 is completed”)."

About half of the analyzed operationalizations fall in Category A (under the assumption of “the more use, the better”) and do not include a threshold for the intended use. In some cases, especially when formulating a hypothesis, it is not known in advance what the intended use of a technology is, or defining it might not be a crucial part of the study. It is important to note that, according to the definitions used, these operationalizations should therefore not be referred to as adherence.

When the intended use for the technology was reported (Categories B and C), only a minority of the included studies featured justified Category C operationalizations. A reason for the lack of justifications for the intended use of eHealth technologies might be that there is a lack of knowledge regarding the working mechanisms of technology-based applications [Michie et al., 2017]. Moreover, the intended use has also been operationalized by linking the (positive) outcomes of individual users to their usage patterns to find the most effective patterns [Carolan et al., 2016], which can only be done *a posteriori*.

In conclusion, Category C operationalizations of adherence are preferable, but might not always be possible to achieve. A minimum of Category B should be used when operationalizing eHealth adherence, taking into account the concept of intended usage.

2.1.2 Dropout vs Adherence

[Christensen et al., 2009] uses the term "dropout" to "describe an individual who fails to complete the research trial protocol associated with an Internet intervention, and thus does not complete trial assessments".

If we use the definition of adherence as "the degree to which an individual follows the intended usage of a program" (adapted from [Donkin et al., 2011] with the considerations in [Sieverink et al., 2017]), we can conclude that both terms are interrelated but distinct (and not necessarily opposite) concepts. Individuals might have consistently low adherence rates, not reaching the full prescribed usage, but still complete the protocol.

Dropout is therefore a permanent final state that can be predicted with different rates of success from different points in time, whereas adherence is an evolving metric that refers to a specific moment or time frame.

2.1.3 Predictors of Adherence and Dropout

Several studies were analyzed to find common predictors of dropout and low adherence in technological health interventions. Many of the relevant predicting features were study-specific (such as reaching a module in a certain time frame). However, a few general predictors also emerged, such as:

- Longer time to complete earlier steps leads to higher dropout probability. [Bremer et al., 2020]
- Dropout is usually not abrupt and will be preceded by decreased adherence over time. [Pedersen et al., 2019]
- Dropout is most common at the beginning of the intervention. For example, in a program 65% of dropouts occurred within the first 2 weeks. [Coa and Patrick, 2016]
- In an intervention for depression, lower baseline rates of depression and younger age were found to be associated with increased adherence. In one for Post-traumatic stress disorder (PTSD), higher adherence was found in women, older persons, those who lived with a partner, and those less experienced with a computer. [Christensen et al., 2009]

As for the models used, [Pedersen et al., 2019] applied logistic regression, decision trees, and random forests, having found that the latter produced better results. [Ramos et al., 2021] included two models, logistic regression and random forests, having also obtained the best results using random forests. [Bremer et al., 2020] implemented logistic regression, linear regression, support vector machines, and boosted decision trees.

2.2 Technical Aspects

2.2.1 Data Processing Pipeline

There are several different tasks to be performed in order to solve a data-centric problem. These tasks can be organized into stages (such as acquisition, cleaning/curation, modeling, and so on). The collection of data science stages can be referred to as data science pipelines. [Biswas et al., 2022]

Even though following certain processing stages is essential, there is not one "golden" data science pipeline. Different authors and data scientists use different pipelines and might even adapt the pipeline according to the project at hand. Not only can the stages not be the same, but each stage can also consist of different tasks according to different sources. It is, therefore, important to study the common ground and variations when it comes to data science pipelines, in order to create a pipeline suitable for the desired purpose.

According to [Biswas et al., 2022], in literature about the Data Science Process, the following stages can be found among the various proposals of pipelines:

- Pre-Processing Layer:
  - **Data Acquisition:** Data are collected from the available sources, either manually or automatically. It is also in this stage that an effort must be made in order to understand the context of the problem and the project objectives and requirements from a business perspective. This will facilitate the design of a preliminary plan to achieve the objectives [Chapman et al., 2000]. This step is commonly referred to as "Business Understanding". Another important step is "Data Understanding", which entails exploring the data to become familiar with it and its nature, as well as verifying its quality.
  - **Data Preparation:** This stage involves further exploration of the data and the subsequent transformation of the raw format into usable material. Some common steps are Cleaning, Filtering, Organizing, and Formatting.
  - **Storage:** This stage, less often found in literature, encompasses the selection of hardware and software most appropriate to the storage and accessing of the data. For example, some types of databases might be more advantageous than others depending on the type of data.
- Model Building Layer:
  - **Feature Engineering:** Because not all features of the dataset will be relevant (or even beneficial) to the modeling stage, the appropriate

features must be identified and selected. Furthermore, some features which can come to be of considerable value might not be present in the dataset but can be constructed from the raw data, sometimes with the aid of external data sources.

- **Modeling:** The next stage, once data is prepared and ready, is the building of the model (or, most often, models). The model building stage requires model planning, model selection, mining, and deriving important properties of data [Biswas et al., 2022]. In order to create a suitable model, different algorithms and strategies are selected, according to the specific problem and domain (classification, clustering, etc).
- **Training:** In this stage, each model is trained with a subset of the labeled data (the training dataset). In order to improve the outcome, models and optimized and tuned with different parameters.
- **Evaluation:** After training each model, a subset of the data which was previously not used (the validation dataset) is used to determine the performance of the model when encountering new data. Appropriate metrics are selected depending on the problem and the balance of the dataset.
- **Prediction:** Finally, the best model is selected and used with another previously unused subset of the data: the test dataset. Final performance metrics are calculated.

* Post-Processing Layer:
  - **Interpretation:** Predicting a certain outcome is often not enough. Rather, it is frequently vital to explain why such a prediction was made and to translate it into knowledge that can be transformed into actionable guidance. In many cases, visualizations can be helpful in the decision-making process.
  - **Communication:** Because the data science process has a clear objective and does not exist in a vacuum, the following stage is naturally to share the acquired knowledge and information with the appropriate stakeholders or even with the scientific community.
  - **Deployment:** Lastly, the solution is deployed and the new knowledge is put into practice. If new data is acquired, the performance of the model is monitored and the model might even be altered using this data to adapt to new circumstances.

As previously mentioned, these stages (and their names) vary according to the source. It is interesting to note that the post-processing stages are less frequently found, as well as the stage referred to as "Storage". Looking at projects found in *Kaggle* (a platform where data scientists and students can publish and explore data sets, build models and participate in competitions), rather than to the literature, [Biswas et al., 2022] found the same lack of post-processing stages.

Finally, it is important to bear in mind that the data science pipeline is not linear. Rather, most stages have feedback loops to different parts of the pipeline.

Furthermore, stages do not have firm boundaries and it is possible to jump to another stage at several points in order to refine the project.

Certain tasks, such as hyperparameter choice and feature selection, naturally require the jump between multiple stages. Some of the most common feedback loops are evaluation to preparation, evaluation to modeling, and prediction to modeling. Feedback loops within the same layer are more common than feedback loops from one layer to another.

While these loops and jumps make the process rather complex, they are also crucial for the success of the process.

2.2.2 Explainable AI

More than simply being able to predict if and when dropout and low adherence occur, being able to understand the factors behind a prediction is of enormous value. Knowing the most probable causes for disengagement allows for the implementation of tailored interventions, more suited to each specific situation. Thus, being able to interpret and explain the reasons behind a model's prediction becomes crucial in this context.

Machine learning is a branch of Artificial Intelligence (AI) and computer science that uses methods and algorithms to make predictions based on data. In a way, it tries to imitate the way that humans learn when exposed to new information.

As machine learning and AI become more widely utilized and decisions that were previously made by humans are now made by machines, it becomes necessary for these mechanisms to explain themselves [Gilpin et al., 2018].

Interpretability can be defined as "the degree to which a human can understand the cause of a decision" [Miller, 2019] or "the degree to which a human can consistently predict the model's result" [Kim et al., 2016]. Therefore, "interpretable machine learning" is a term that signifies the “extraction of relevant knowledge from a machine-learning model concerning relationships either contained in data or learned by the model” [Murdoch et al., 2019].

Methods for machine learning interpretability can be classified as intrinsic or post hoc [Molnar, 2019].

Intrinsic methods achieve interpretability by restricting the complexity of the machine learning model. They refer to models that are considered interpretable due to their simple structure, such as short decision trees. The simplest way to achieve interpretability is therefore to restrict the chosen algorithms to only those that create interpretable models. These include:

- Linear Regression
- Logistic Regression
- Decision Tree

- Decision Rules
- RuleFit
- Naive Bayes Classifier
- K-Nearest Neighbors

Post hoc interpretability refers to the application of interpretation methods after model training. These methods can also be applied to intrinsically interpretable models. Model-agnostic interpretation methods (in contrast with model-specific ones) are those that can be applied to any model. This flexibility makes them easier to work with when testing several different models and comparing them in terms of interpretability. Model-agnostic methods can be either global or local.

Global methods describe the average behavior of a machine learning model and are often expressed as expected values based on the distribution of the data. These methods are particularly useful, for instance, when the modeler wants to understand the general mechanisms in the data. They can be used to make general decisions and interventions based on the data as a whole. Some model-agnostic global interpretation techniques are:

- The partial dependence plot (a feature effect method)
- Accumulated local effect plots (another feature effect method that works when features are dependent)
- Feature interaction (H-statistic) (quantifies to what extent the prediction is the result of joint effects of the features)
- Functional decomposition (a technique that decomposes the complex prediction function into smaller parts)
- Permutation feature importance (which measures the importance of a feature as an increase in loss when the feature is permuted)
- Global surrogate models (replaces the original model with a simpler model for interpretation)
- Prototypes and criticisms (representative data point of a distribution that can be used to enhance interpretability)

Local model-agnostic methods explain individual predictions. They can be used to make specific decisions and interventions based on a specific data point. Some model-agnostic local explanation methods are:

- Individual conditional expectation curves (the starting point for partial dependence plots, describe how changing a feature changes the prediction)
- Local surrogate models (LIME) (explain a prediction by replacing the complex model with a locally interpretable surrogate model)

- Scoped rules (anchors) (generate rules that describe which feature values anchor a prediction, i.e. lock the prediction)
- Counterfactual explanations (describe the smallest change to a feature value that changes the prediction to a predefined output)
- Shapley values (are a method from coalitional game theory that fairly assigns the prediction to individual features by calculating the average marginal contribution of a feature value across all possible coalitions)
- SHAP (another computation method for Shapley values that also proposes global interpretation methods based on combinations of Shapley values across the data)

Chapter 3

Methodology

This chapter describes the steps taken in order to build the machine learning models. There were three different approaches, in accordance with the project goals:

1. **Regression**: Predicting the adherence percentage of a user in a period with the information of the preceding window of similar length.
2. **Binary Classification**: Predicting the adherence level of a user in a period (high vs low), looking at a preceding period of similar length.
3. **Filtered Classification**: Similar to the previous approach but only when there is significant usage in the preceding period.

3.1 Dataset

3.1.1 Context

ACTIVAGE is a *Large Scale Pilot* project funded by the European Commission, which aims to demonstrate that Internet of Things (IoT) is essential in the use of Smart Living solutions that have a positive effect on Active and Healthy Ageing (AHA). This pilot was executed in several different cities across Europe, known as Deployment Sites (DS). One of them was the Madrid DS.

The Madrid DS had three main objectives: 1) to prevent the decrease in cognitive performance using brain training exercises and reminders, 2) to prevent falls by using physical training and exercises, and 3) to prevent social isolation by encouraging users to establish and maintain social interactions. [Fico et al., 2017]

In order to fulfill these objectives, participants were divided into four Use Cases (UC), according to their profiles and needs:

- UC 3 - Proactive users

- UC 5 - Fragile users
- UC 6 - Active users
- UC 7 - Isolated users

All participants were given access to the same mobile applications and, if necessary, were provided with devices on which to access these apps. However, participants from different UC were given different tasks regarding which apps to use and how often they should interact with them, in order to address specific problems of their aging profile (cognitive, physical, socialization, or depressive).

The data consists of the logs from the usage of these apps, as well as the participants' answers to several questionnaires and socio-demographic information.

Figure 3.1: App Home

3.1.2 Sociodemographic Information and Questionnaires

Sociodemographic Information Contains information on the sociodemographic characteristics of each participant, as well as basic information about the intervention they are receiving. This information gives a snapshot of the user at the moment of enrollment. Contains fields such as "Gender", "Year of birth", "Educational level", "Technology Level", and "Living Environment", among others.

UT-AUT Questionnaire The Unified Theory of Acceptance and Use of Technology (UT-AUT) is a technology acceptance model proposed by [Venkatesh et al., 2003]. The questionnaire aims to evaluate four constructs that play a significant role as determinants of user acceptance and usage behavior: performance expectancy, effort expectancy, social influence, and facilitating conditions.

EQ-5D-3L Questionnaire The EQ-5D-3L is a standard questionnaire for Quality of Life self-assessment provided by euroqol.org. The EQ-5D-3L descriptive system comprises the following five dimensions: mobility, self-care, usual activities, pain/discomfort, and anxiety/depression. Each dimension has 3 levels: no problems, some problems, and extreme problems. The patient is asked to indicate his/her health state by ticking the box next to the most appropriate statement in each of the five dimensions.

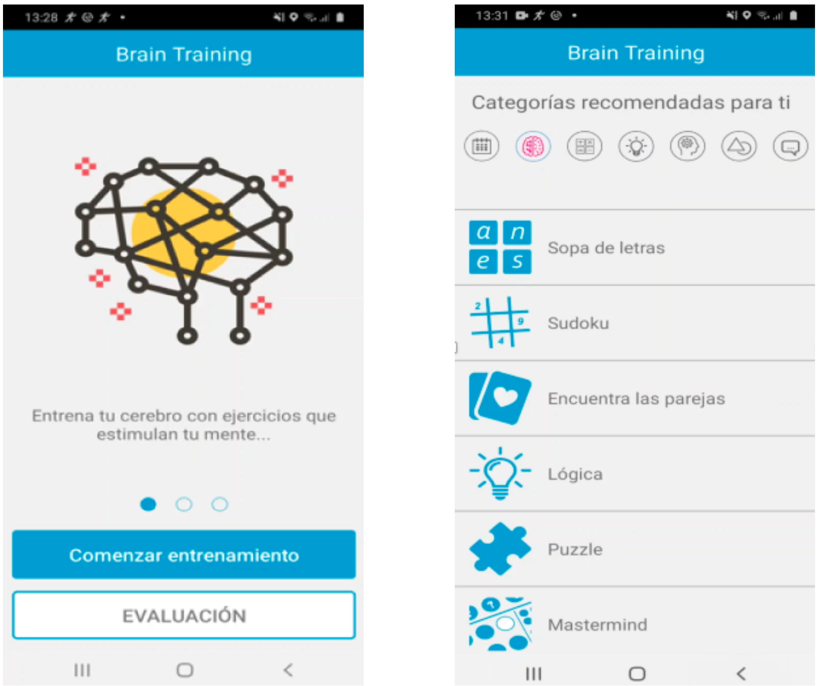
UCLA Loneliness Scale Questionnaire Developed by psychologist Daniel Russell (1996), the UCLA Loneliness Scale (Version 3) is a 20-item measure that assesses how often a person feels disconnected from others. Each possible answer ranges between 1 and 4, from "I never feel this way" to "I often feel this way".

Self Perception Questionnaire (SPQ) The Self Perception Questionnaire evaluates the overall self-perception of the participants on four instruments: quality of life, physical activity, social life and the provided IoT solution. Each answer ranges from 0 (Extreme problems/very negative) to 10 (Very good/no problems/very positive).

3.1.3 App Usage Data

Brain Games Brain games is an app composed of a set of different cognitive games oriented to train specific cognitive abilities (i.e memory, calculation, perception, reasoning, etc). The table contains an entry for every time a user attempted to solve a game, including a timestamp, the difficulty of the level, duration, if the game was completed and which specific cognitive game was played.

Figure 3.2: Brain Games App



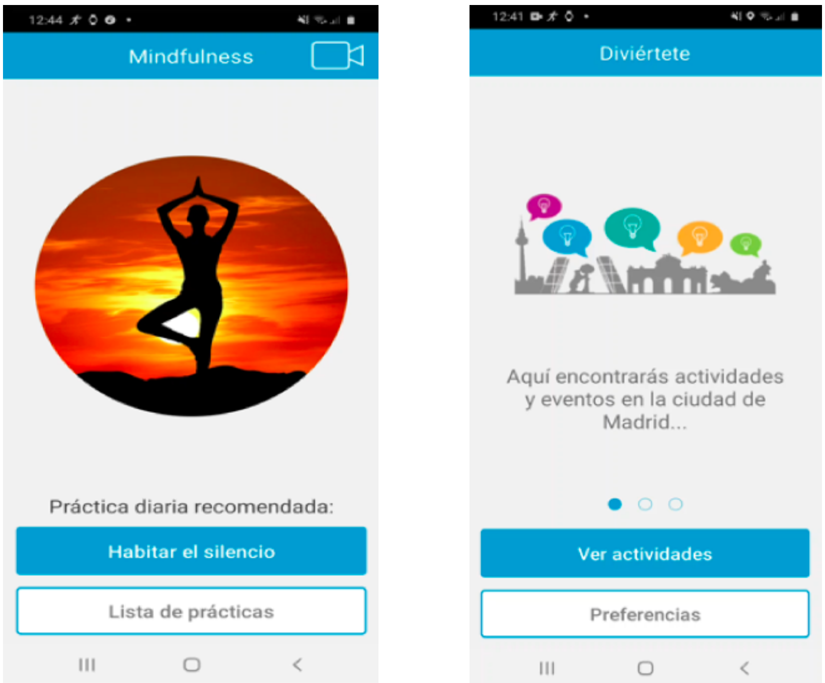
Physical Activity Similar game data as the Brain Games, but for the physical activity app, with balance and coordination exercises results.

Finger Tapping A subset of the physical activity games, tests coordination, rhythm and reaction time. Has specific fields, such as the number of taps, number of errors, and the minimum, maximum, mean and standard deviation for the reaction time.

Figure 3.3: Physical Activity and Finger Tapping Apps

Mindfulness Contains the results from logged sessions of the mindfulness app, with a timestamp, duration and whether the session was completed.

Figure 3.4: Mindfulness and Diviértete Apps



Digital Phenotyping This table contains all the logged information of actions taken by the user of the different apps from the Madrid DS. It includes all the pressed buttons and actions performed during the interaction with the app menus.

3.2 Pre-Processing

3.2.1 Data Acquisition & Data Preparation

The first step in the data processing pipeline is the transformation of the raw data from different sources into a unified dataset. In the case of data of user activity in an intervention (such as is the case) it might be necessary to aggregate observations, specifying the time window or interval for which points are aggregated. [Bremer et al., 2020]

In addition, it is necessary to handle missing data. This can be done in several different ways, such as:

- Ignoring and Discarding Data - either complete case analysis or discarding instances and/or attributes. Both should only be applied when the missing data is classified as Missing Completely At Random.
- Parameter Estimation - using maximum likelihood procedures to estimate the parameters of a model defined for the complete data.
- Imputation - filling in the missing values with estimated ones.

Imputation can be made through various methods, like replacing the missing data for a given attribute by the mean or mode of all known values of that attribute, or by developing statistical models that can predict the missing values based on other features. [Batista and Monard, 2003]

In order to clean this dataset, it was necessary to use two of the previous approaches: discarding data (both records and variables) and imputation.

3.2.2 Exploratory Data Analysis

Exploratory Data Analysis (EDA) refers to the process of performing initial investigations on data so as to discover patterns, to spot anomalies, to test hypothesis and to check assumptions with the help of summary statistics and graphical representations. [Patil, 2022]

This step, which can occur in conjunction with data preparation, is crucial in order to make sense of the data.

To explore the data of this project the following methods were used, among others: missing data density visualizations, feature correlation analysis, outliers detection using boxplots, and value distributions. The main takeaways from this stage of data processing can be found in the next chapter.

3.3 Model Building

3.3.1 Feature Engineering and Selection

Adding new features can very often lead to achieving better predictive performance and is a key step in the machine learning process. [Domingos, 2012] This can be done by two main approaches: handcrafted feature engineering, and automated or ad-hoc feature engineering.

Handcrafted feature engineering requires deep knowledge about the problem domain and can be a challenging (but important) task. Automated feature engineering is done by automatically generating a large number of candidate features and selecting the best by evaluating their predictive performance. One example, interaction features, can be the product of two other features and can lead to additional knowledge about their relationships. [Bremer et al., 2020]

Lastly, time window–based aggregation methods can be useful in the context of digital health interventions. [Bremer et al., 2020] In this case aggregations are performed on the original features for a specified time window (for example using the sum of the number of levels completed in a day or the average of the number of errors per level in that day).

A large number of features (which will be described in the next chapter) were created using a combination of the methods above. A large number required time window–based aggregation due to the nature of the data.

3.3.2 Target

The target for the models was the adherence level of the window in question. This measure was generated following the literature from the State of the Art and taking into account the concept of "intended use". Because users from distinct UCs were given different instructions, the formula for the adherence level had to take the users' UC into consideration.

Therefore, the adherence level was calculated as a percentage of the intended use as shown below:

- UC 3 - diviertete app once every 2 weeks
- UC 5 - physical activity or finger tapping app 5 times per week = 10 times over 2 week period
- UC 6 - brain games app 2 times per week = 4 times over 2 week period
- UC 7 - 4 complete mindfulness sessions per week = 8 complete sessions over 2 week period

The adherence percentage was capped at 100%, even if the actual use was superior than the intended used.

For percentages below 50% the adherence level was considered "low" and for percentages of 50% or above the adherence level was considered "high". The adherence percentage was used as the target in the first approach, while the binary adherence level was used as the target in the 2nd and 3rd approaches.

The adherence percentage was also calculated for the preceding window and used as a feature.

3.3.3 Modeling

For each of the 3 approaches, several models were built using a variety of algorithms. The chosen algorithms are diversified, which allows us to understand which are better suited to the problem, while using algorithms with different degrees of interpretability.

For the first approach (regression) the following algorithms were chosen:

- Logistic Regression (LR)
- LASSO Regression
- Random Forest Regressor (RF)
- Support Vector Regressor (SVM)
- Decision Tree Regressor

For the second and third approaches (classification) the following algorithms were chosen:

- K-Neighbors Classifier (KNN)
- Gaussian Naive Bayes (NB)
- Decision Tree
- Random Forest Classifier (RF)
- Support Vector Classifier (SVM)
- Multi-Layer Perceptron (MLP)

3.3.4 Training and Optimization

The models were initially trained using the Scikit-Learn default values provided for each one, in order to allow for a preliminary comparison. Each model was also trained using a combination of lengths for the prediction and provided time windows. Both windows were tested for 1, 2 and 4 weeks, creating a total of 9 combinations of window durations.

Based on the results obtained some models were selected to undergo further tuning and hyperparameterization.

Model Validation

In order to test the generated model's prediction capabilities data is usually split into training data and test data. Additionally, an extra split of the training data and the creation of a validation set can be used to compare different models and when experimenting with different features or tuning hyperparameters. This is called holdout validation. Another option for validation is to use cross-validation. In the most common scenario, k-fold cross-validation, the data is divided into k equally sized segments or folds. Subsequently, k iterations of training and validation are performed so that a different fold of the data is held out for validation in each iteration, while the remaining k—1 folds are used for training the model. [Refaeilzadeh et al., 2009]

For comparison of the trained models, 5-fold cross-validation was used on the training split of the dataset.

Scoring Function

The F-score or F-measure was chosen as the scoring function for the classification tasks because it is a combination of precision and recall into a single metric and our generated dataset is unbalanced, as there are many more low adherence windows than high adherence windows.

For the regression task, the scoring function was the Root Mean Square Error (RMSE).

Chapter 4

Results and Discussion

4.1 Data Preparation and Exploratory Data Analysis

While looking at the Sociodemographic data, a number of characteristics can be easily noted. There are 266 participants present in the data set (186 women and 80 men) who were on average born in 1944 (S.D.=7.4) (average 75 years old in 2019), with the birth year ranging from 1924 (95 years old in 2019) to 1961 (58 years old in 2019). Further sociodemographic information can be found in the table below:

Table 4.1: Sociodemographic information

	Experimental group
Gender (F)	186 (69.9%)
Year of Birth (m±SD)	1943.82±7.42
Educational level	
ISCED 0-2	124 (46.6%)
ISCED 3-5	80 (24.9%)
ISCED 6-8	62 (28.5%)
Technological Level	
Basic	141 (53.0%)
Intermediate	81 (30.5%)
Advanced	44 (16.5%)
Living Environment (Urban)	232 (87,3%)
Living Conditions (Home)	262 (98,5%)
Living Status (Alone)	91 (34.2%)

An important detail to note is that the distribution of the users among the Use Cases is not equal, as can be seen in Figure 4.1.

Further exploration reveals several instances of missing data, most of which in the SPQ questionnaire answers.

A number of outliers can also be found, particularly in the duration of the mindfulness and brain games activities, as can be seen in Figures 4.5 and 4.5.

Figure 4.1: Users by use case

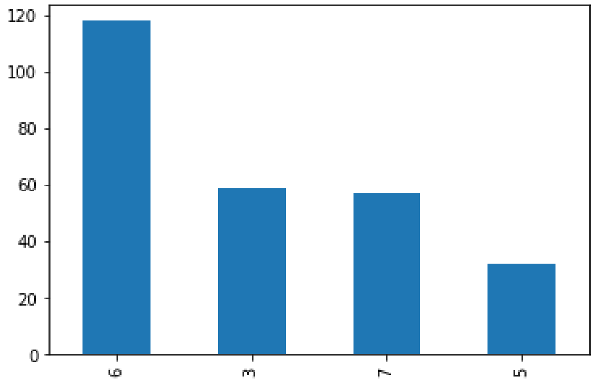


Figure 4.2: Missing values density analysis of the Sociodemo table features

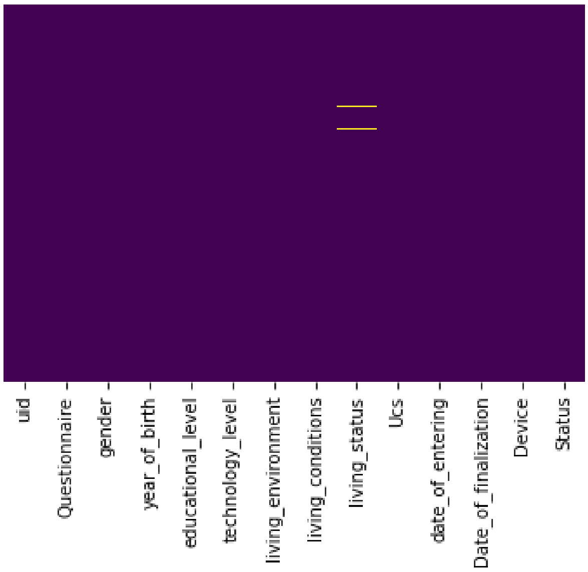


Figure 4.3: Missing values density analysis of the SPQ table features

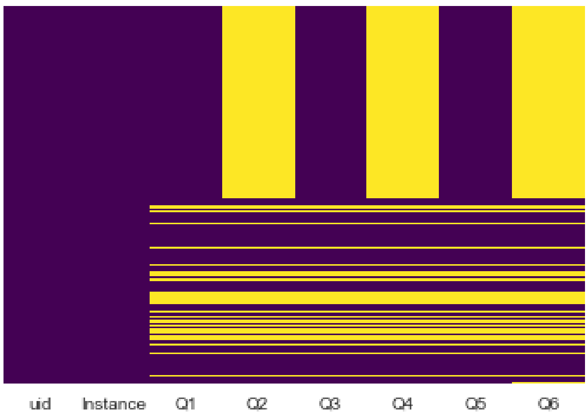


Figure 4.4: Missing values density analysis of the SPQ table features only with instance 1

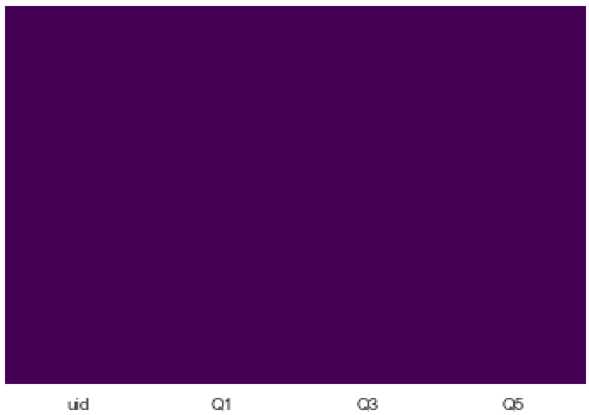


Figure 4.5: Boxplots and distribution analysis of the Mindfulness table features

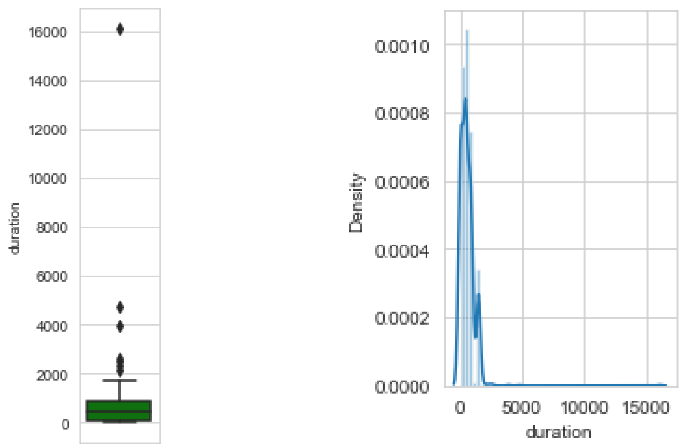
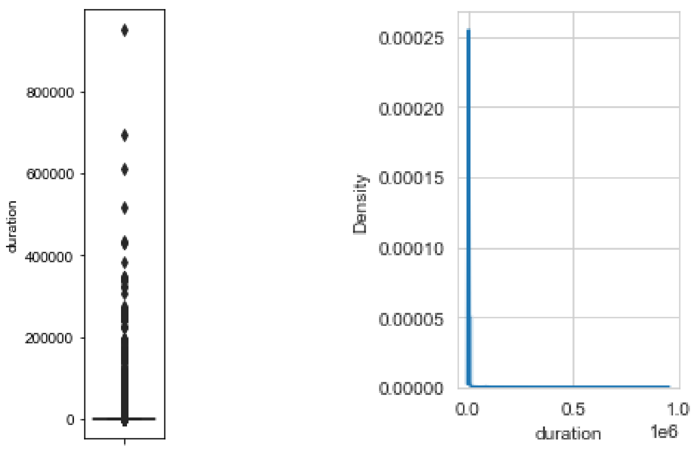


Figure 4.6: Boxplots and distribution analysis of the Brain Games table features



After merging the data from the sociodemographic and questionnaires tables, more missing data becomes visible. A large number of users did not respond to the UCLA questionnaire. It is also interesting to note that the biggest correlations between features occur within features of the same questionnaire.

Figure 4.7: Missing values density analysis of the Users table features

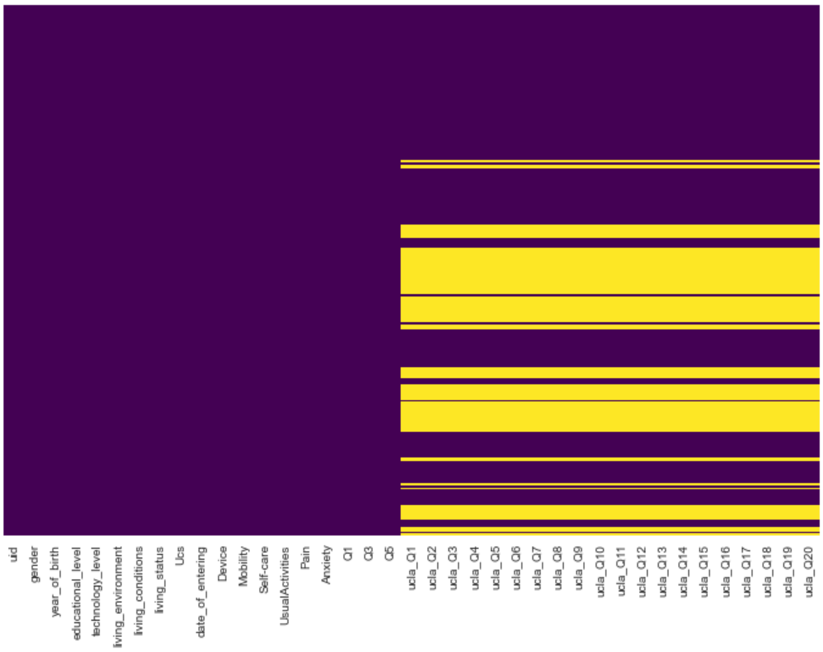


Figure 4.8: Correlation analysis between the Users table features

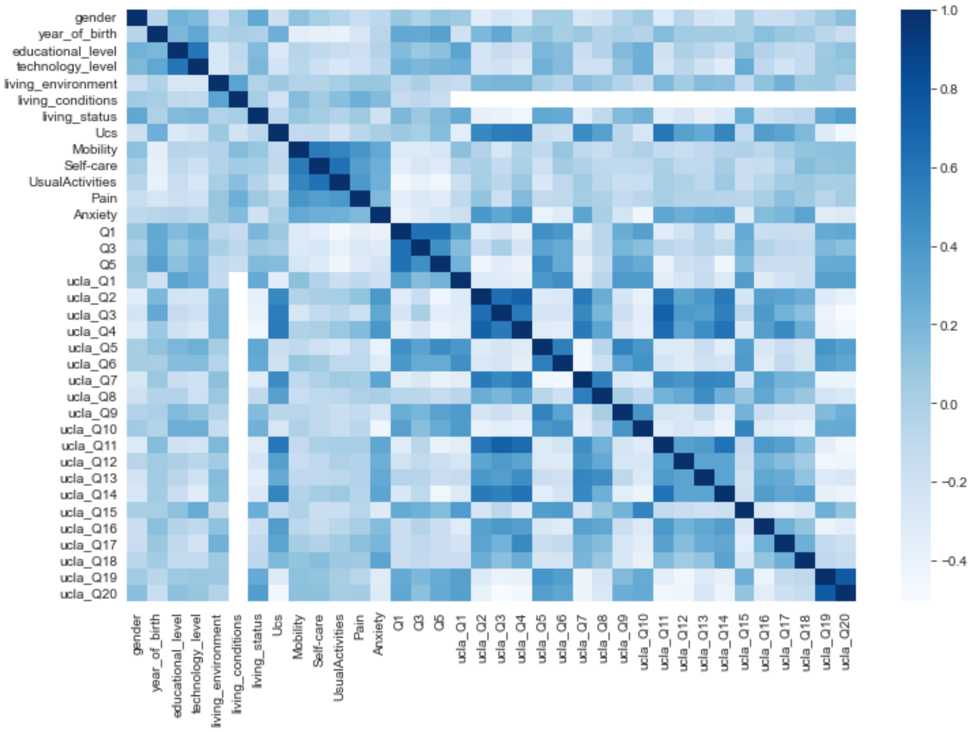


Table 4.2: Activities information

	brain games	**physical activity**	**finger tapping**	**mindful-ness**	**digital pheno-type**
Total number of records	61614	1421	805	1145	135089
Total number of users	174.000	86.000	94.000	157.000	198.000
N° of sessions per user (mean (SD))	354.103 (563.248)	16.523 (25.753)	8.564 (10.450)	7.293 (7.684)	682.267 (1227.898)
Duration per session (mean (SD))	1543.969 (12579.900)			578.878 (671.643)	

Figure 4.9: Number of sessions per user for each activity boxplots

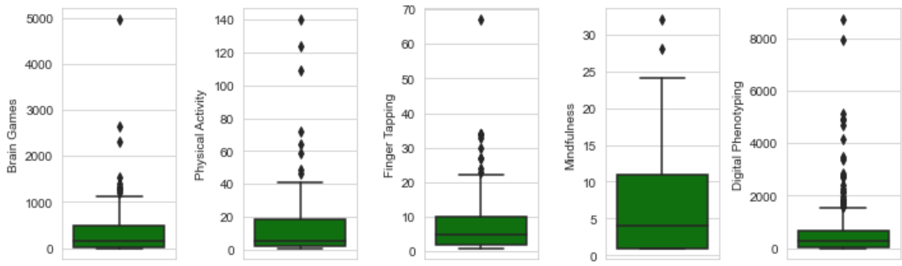
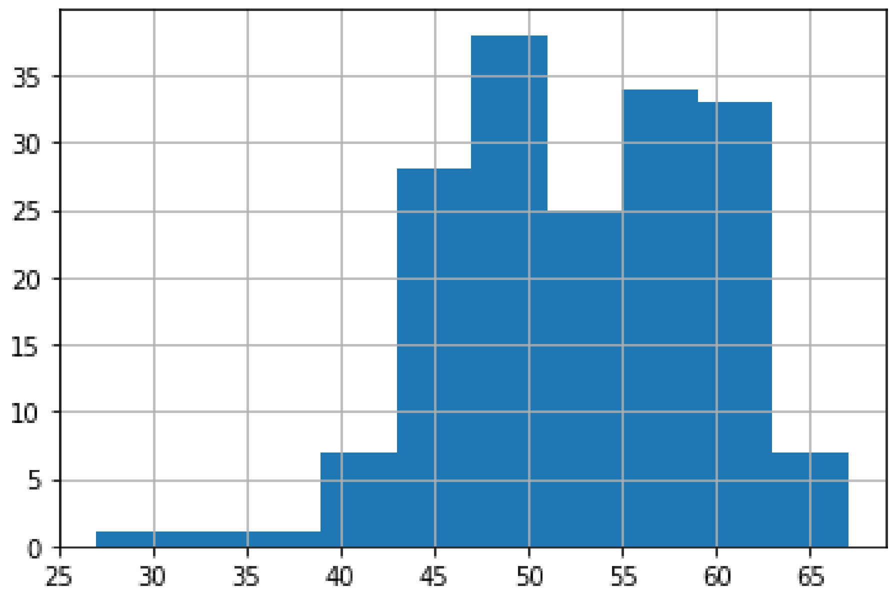


Figure 4.10: UCLA sum distribution



Data Preparation

After some initial data analysis of the 10 tables, several actions were performed. Firstly the dataset was divided into a train and a test sets, separating by user. All the basic following steps were repeated across both sets.

The next step was data cleaning. In the first place, the connecting field was renamed to match across all tables and its content was parsed into lowercase in order to allow for filtering (sometimes the field was called "uid" and sometimes "record_id" and the capitalization was not uniform). Then, all the dates were parsed into the same format and some numerical values were parsed from strings. Some fields that had one single value across all records were removed (for example, the "Questionnaire" column on the "Sociodemo" and "EQ5D3L" tables). The "log" field on the "digital phenotyping" table was fixed to correct a text parsing error that had resulted in foreign characters.

Because the objective of the model will be the prediction of adherence and/or dropout, all the data that was collected *a posteriori* was discarded. This included all the questionnaires that were administered at any point other than the initial interview (several instances of the "UCLA", "EQ5D3L" and "SPQ" questionnaires as well as the "UTAUT" questionnaire, which was only administered at the date of termination). The fields "status" and "date of finalization" from the "Sociodemo" table were discarded for the same reason.

Missing values were dealt with in several different ways. On the "digital phenotyping" and "finger tapping" tables the whole row was removed because missing values, when present, spanned across all fields. On the "Sociodemo" table, missing "living status" was replaced with the most common value for the same "living environment" and "living conditions". Missing answers on the "UCLA Loneliness Scale Questionnaire" were substituted by the mean answer for each question (calculated on the train set).

Feature Engineering

Using the data from the "digital phenotyping", a new table was created with the information of the usage of the "diviertete" app (which was prescribed to the UC 3).

In order to create the models capable of predicting the adherence percentage or level of a user in a 1, 2 or 4 week period with the information of the preceding 1, 2 or 4 week window, an aggregated dataset with several features was generated. For each user, a period of one year after entering the study was considered. The following features were calculated:

- days since entering the trial until the start of the window
- days since the beginning of the civil year (to account for time of year)
- working days in the preceding period
- working days in the target period
- number days with sessions on the physical activity app

- number of days with sessions on the finger tapping app
- number of days with sessions on the brain games app
- number of days with sessions on the mindfulness app
- number of days with sessions on the diviertete app
- total number of days with sessions across all apps
- number of sessions on the physical activity app
- number of sessions on the finger tapping app
- number of sessions on the brain games app
- number of sessions on the mindfulness app
- number of sessions on the diviertete app
- total number of sessions across all apps
- total duration of brain games app sessions
- average duration of brain games app sessions
- total duration of mindfulness app sessions
- average duration of mindfulness app sessions
- number of solved sessions on the physical activity app
- percentage of solved sessions on the physical activity app
- number of solved sessions on the finger tapping app
- percentage of solved sessions on the finger tapping app
- number of solved sessions on the brain games app
- percentage of solved sessions on the brain games app
- number of completed sessions on the mindfulness app
- percentage of completed sessions on the mindfulness app
- total number of completed/solved sessions across all apps
- total percentage of completed/solved sessions across all apps
- average number of fingertapping taps
- average number of fingertapping errors
- other fingertapping measures' averages
- number of unique physical activity games played

- number of unique fingertapping games played
- number of unique brain games types played
- number and percentage of easy, medium and hard brain games sessions played

In addition, the socio-demographic information and questionnaire answers were also used, including the total score of the UCLA questionnaire answers.

4.2 Approach 1

The first approach was a regression task to predict the adherence percentage of a user in a period with the information of the preceding window of similar length. To do this, 3 different windows for each period were used, as described before, for a total of 9 combinations and the 5 algorithms described in Section 3.3.3 were trained.

The results (as can be seen in Table C.1) appear very promising. However, it is important to note that, due to the imbalanced nature of the dataset, predicting an adherence percentage of 0% is a fairly easy task when the preceding period also has no engagement.

4.3 Approach 2

The second approach was the stepstone in trying to solve the data imbalance problem and the hope of obtaining some future actionable insights.

This approach was a binary classification task that tried to predict the adherence level of a user in a period (high vs low), looking at a preceding period of similar length. Once again, a plethora of algorithms were tested and the results were very good (see Table C.2). Random Forests reached better scores, even though all models achieved scores close to 0.9. However, the problem of the previous approach persists.

4.4 Approach 3

Lastly, the third approach tried to mitigate the imbalance created by permanent user dropout. This approach is similar to the previous one but the dataset was filtered to instances where the adherence level was high in the preceding period. The situations that require prediction for a possible intervention are exactly the unexpected ones, when the user is engaged with the solution right before diminishing interaction.

Figure 4.11: Adherence percentage distribution

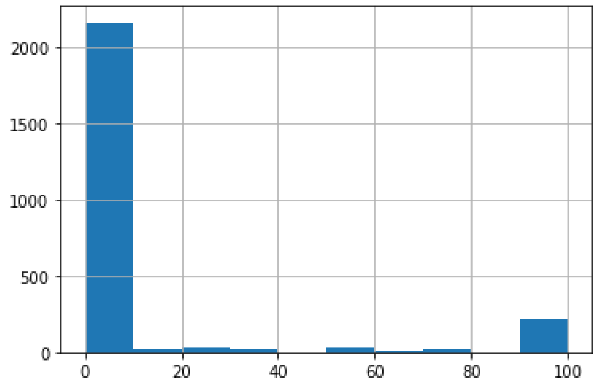
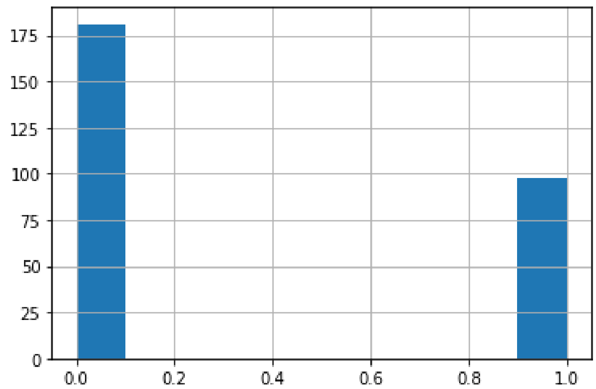


Figure 4.12: Adherence percentage distribution after filtering



Once again, the models were trained with the algorithms described previously and for the 9 combinations of window size. As expected, the score was inferior to the one from the previous approach.

The four algorithms that provided the best scores were then selected for fine-tuning using Grid Search of the model's hyperparameters, again for all window sizes. These algorithms were a Decision Tree, KNN and Random Forest and MLP. The results can be found in Tables C.3 to C.7. The parameters selected can be found in Table B.1.

Finally, two algorithms were selected to undergo Forward Feature Selection (FFS) with hyperparameter fine-tuning, using the same parameters as before. The selected algorithms were Random Forest (which provided the best results) and Decision Tree (due to its intrinsic interpretability, as well as good results).

This time, a fixed 2-week window length was selected both for the preceding period and the target period. These window lengths were selected for 2 reasons:

1. The best overall result was found with these lengths, using the Random Forest Classifier.

2. These lengths of time make the most sense to calculate the adherence metrics due to the intended use that the users were expected to follow.

Table 4.3: Approach 3 best results (Average 5-Fold F1-score)

	Decision Tree	Random Forest
Default Parameters	0.684	0.774
After Hyperparemeter Grid Search	0.738	0.810
After FFS with Hyperparemeter Grid Search	**0.793**	**0.811**

Decision Tree

The results after the last step were very satisfactory. In the Decision Tree, the best model returned an average F1-score of 0.793 (having selected 4 features and using the following hyperparameters: *{'criterion': 'gini', 'max depth': 4}*).

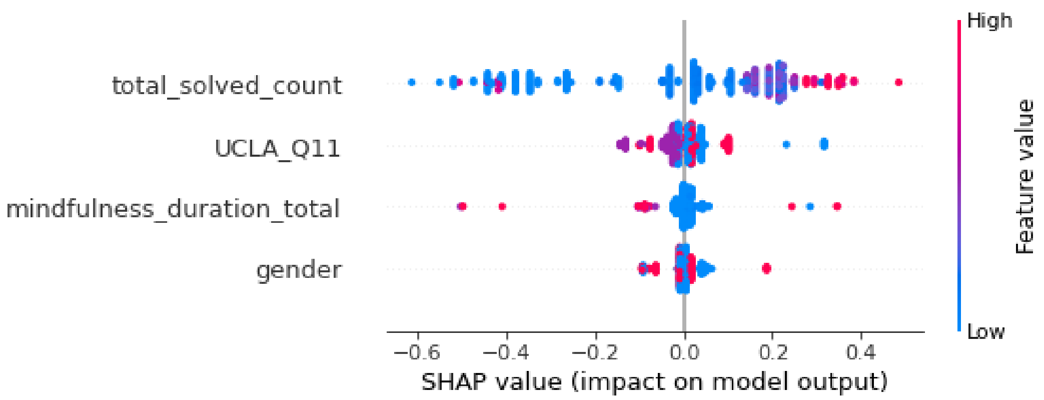
The selected features after FFS were

- The total number of solved or completed activities,
- The total duration of mindfulness sessions,
- Gender, and
- Question 11 of the UCLA questionnaire

The results from this step can be found in Appendix C, along with the generated Decision Tree.

Even though the Decision tree is an intrinsically interpretable model, we can use SHAP values to quicky visualize the importance of each feature:

Figure 4.13: Decision Tree SHAP values



Random Forest

In the Random Forest, the best model returned an average F1-score of 0.81126 (having selected 18 features and using the following hyperparameters: *{'criterion': 'entropy', 'max features': 'auto', 'min samples leaf': 2, 'min samples split': 2, 'n estimators': 25}*).

The selected features after FFS were

- number of days with brain games sessions,
- EQ5D3L questionnaire Anxiety score,
- percentage of solved physical activity sessions,
- number of solved physical activity sessions,
- living environment,
- living status,
- number of medium difficulty brain games sessions,
- average duration of brain games sessions,
- UCLA questionnaire Question 7,
- UCLA questionnaire Question 9,
- number of working days in target period,
- adherence trend within the preceding window,
- total percentage of solved or completed sessions,
- percentage of completed mindfulness sessions,
- percentage of solved fingertapping sessions,
- number of days with any sessions,
- UCLA questionnaire Question 20,
- average of mean rt in fingertapping sessions.

The results from this step can also be found in Appendix C.

In order to make sense of the results and the importance of each feature, SHAP values were used and produced the following results:

Figure 4.14: Random Forest SHAP values

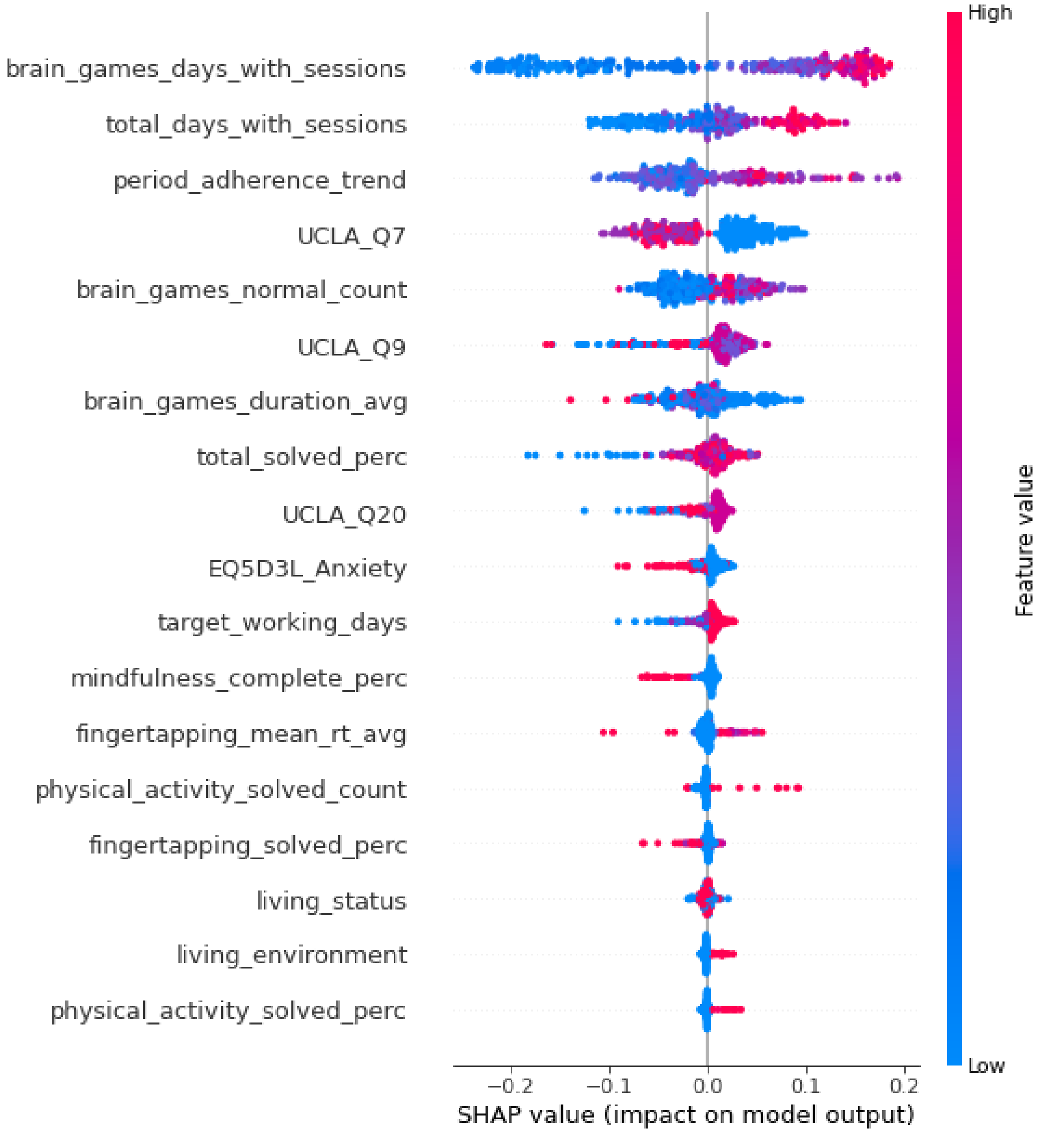
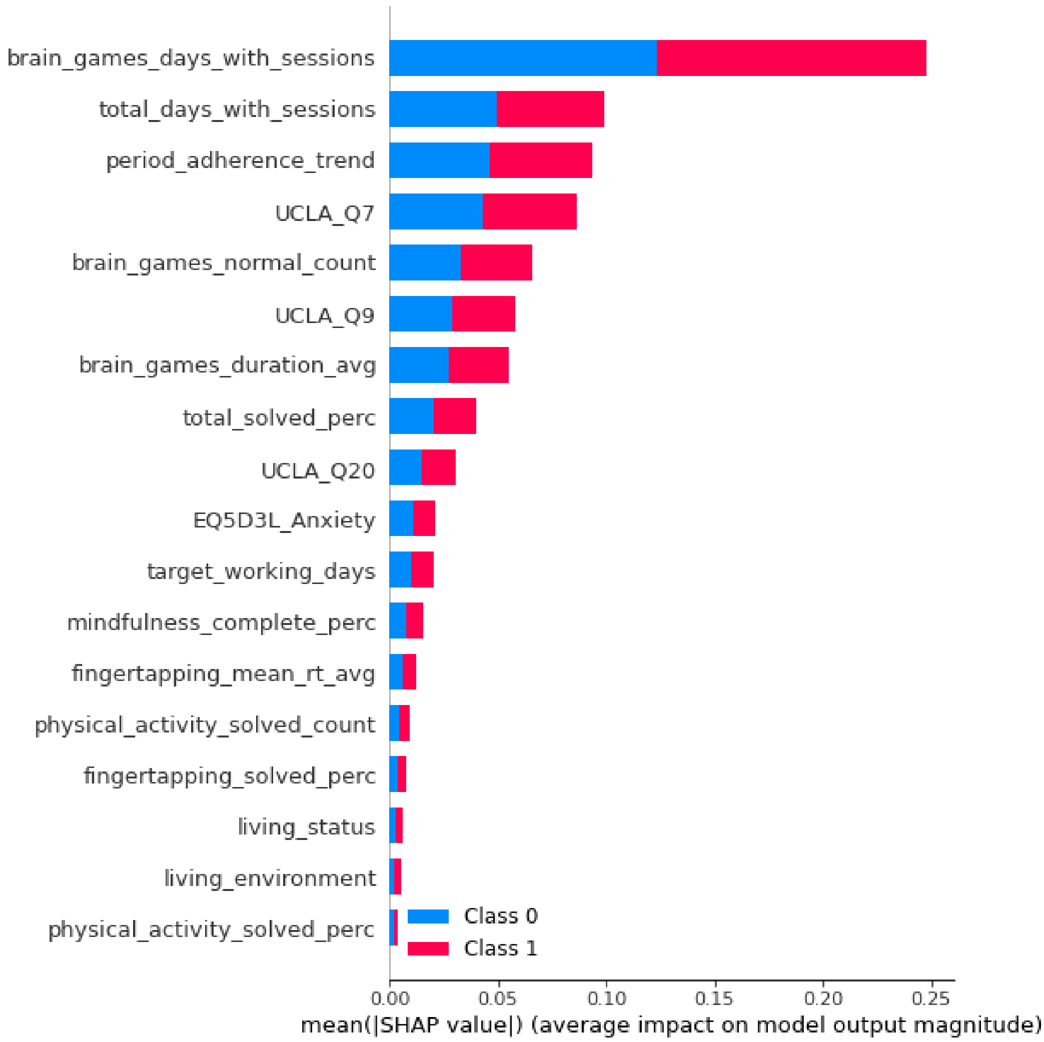


Figure 4.15: Random Forest SHAP values



4.5 Discussion: insights and recommendations

The final approach was capable of producing a model with good prediction capabilities of detecting a drop in user adherence. However, the most interesting results might be the insights gainned from the analysis of the impact of each feature on the model output. Specifically, we can infer the following:

- A good predictor of sustaining high adherence is the number of days in which the user interacted with the solution by completing a session. This is more relevant than the total number of sessions in the period. Therefore, the users should be encouraged to complete a smaller number of sessions consistently throughout the week rather than a larger number of sessions in a single day.
- Both the total number and the percentage of solved or completed sessions have a positive impact on high adherence. It is important to understand why a user is failing to solve the games or complete the sessions because this will lead to disengagement in the near future. It might be because the games are too hard or because the user is having difficulty with the technology.
- Similarly, a longer average duration of the brain games activities could lead to a drop in adherence. The longer duration could signal difficulty solving the activity as well and an effort should be made to understand the reason and assist the user.
- UCLA Loneliness questionnaire's questions 7 ("I am no longer close to anyone") and 9 ("I am an outgoing person") are also good predictors. Users with a better supporting system and more outgoing users are more likely to remain engaged with the solution. Perhaps an extra effort could be made to remain in more close contact with the users that score higher in the loneliness scale.
- Users with a lower anxiety score in the EQ5D3L questionnaire were more prone to staying adherent to the study. As with the above case, a more hands-on approach with more contact and regular check-ins might benefit these users.

Chapter 5

Conclusion

5.1 Reflection and Future Work

The main obstacles of the first part of the project were the definition of adherence and the operationalization of this measure, and the understanding of the problem, its context, and the data, with its features and acquisition process. The process of data cleaning was significant and represented a large chunk of the preparation phase, which will be very useful for the development of further models.

Creating a model that can identify and predict a decrease in adherence rather than an absolute measure of adherence, while leading to slightly worse results, can have a more practical effect and contribute to its usability. It would be interesting to implement the recommendations from the insights generated from the models and to study the results of these interventions in order to better access the usefulness of this project.

A next step could also be to test the usefulness of building different models for the different Use Cases (UC), due to how distinct they are in terms of the intended usage. However, for some of the UCs some more data could be needed in order to build satisfactory predictors, namely a larger number of in certain UCs.

Finally, in a real-time scenario it could be possible to understand the features behind a prediction of drop in adherence for a specific user in a certain date and foster tailored interventions even before this decrease in adherence occurs.

References

Gustavo EAPA Batista and Maria Carolina Monard. An analysis of four missing data treatment methods for supervised learning. *Applied artificial intelligence*, 17 (5-6):519–533, 2003.

Sumon Biswas, Mohammad Wardat, and Hridesh Rajan. The art and practice of data science pipelines. 2022.

Vincent Bremer, Philip I Chow, Burkhardt Funk, Frances P Thorndike, Lee M Ritterband, et al. Developing a process for the analysis of user journeys and the prediction of dropout in digital health interventions: Machine learning approach. *Journal of Medical Internet Research*, 22(10):e17738, 2020.

Stephany Carolan, Peter R Harris, Kathryn Greenwood, and Kate Cavanagh. Increasing engagement with, and effectiveness of, an online cbt-based stress management intervention for employees through the use of an online facilitated bulletin board: study protocol for a pilot randomised controlled trial. *Trials*, 17 (1):1–10, 2016.

Peter Chapman, Janet Clinton, Randy Kerber, Tom Khabaza, Thomas P. Reinartz, Colin Shearer, and Richard Wirth. Crisp-dm 1.0: Step-by-step data mining guide. 2000.

Helen Christensen, Kathleen M Griffiths, and Louise Farrer. Adherence in internet interventions for anxiety and depression: systematic review. *Journal of medical Internet research*, 11(2):e1194, 2009.

Kisha Coa and Heather Patrick. Baseline motivation type as a predictor of dropout in a healthy eating text messaging program. *JMIR mHealth and uHealth*, 4(3):e5992, 2016.

Pedro Domingos. A few useful things to know about machine learning. *Communications of the ACM*, 55(10):78–87, 2012.

Liesje Donkin, Helen Christensen, Sharon L Naismith, Bruce Neal, Ian B Hickie, and Nick Glozier. A systematic review of the impact of adherence on the effectiveness of e-therapies. *Journal of medical Internet research*, 13(3):e52, 2011.

Jarrett Evans, Amy Papadopoulos, Christine Tsien Silvers, Neil Charness, Walter R Boot, Loretta Schlachta-Fairchild, Cindy Crump, Michele Martinez, and Carrie Beth Ent. Remote health monitoring for older adults and those with heart failure: adherence and system usability. *Telemedicine and e-Health*, 22(6): 480–488, 2016.

Giuseppe Fico, Juan-Bautista Montalva, Alejandro Medrano, Nikos Liappas, Angeles Mata-Díaz, Gloria Cea, and Maria Teresa Arredondo. Co-creating with consumers and stakeholders to understand the benefit of internet of things in smart living environments for ageing well: the approach adopted in the madrid deployment site of the activage large scale pilot. In *EMBEC & NBC 2017*, pages 1089–1092. Springer, 2017.

Leilani H Gilpin, David Bau, Ben Z Yuan, Ayesha Bajwa, Michael Specter, and Lalana Kagal. Explaining explanations: An overview of interpretability of machine learning. In *2018 IEEE 5th International Conference on data science and advanced analytics (DSAA)*, pages 80–89. IEEE, 2018.

Rakibul Hoque and Golam Sorwar. Understanding factors influencing the adoption of mhealth by the elderly: An extension of the utaut model. *International journal of medical informatics*, 101:75–84, 2017.

Corine Horsch, Jaap Lancee, Robbert Jan Beun, Mark A Neerincx, and Willem-Paul Brinkman. Adherence to technology-mediated insomnia treatment: a meta-analysis, interviews, and focus groups. *Journal of medical Internet research*, 17(9):e214, 2015.

Been Kim, Rajiv Khanna, and Oluwasanmi O Koyejo. Examples are not enough, learn to criticize! criticism for interpretability. *Advances in neural information processing systems*, 29, 2016.

Susan Michie, Lucy Yardley, Robert West, Kevin Patrick, and Felix Greaves. Developing and evaluating digital interventions to promote behavior change in health and health care: recommendations resulting from an international workshop. *Journal of medical Internet research*, 19(6):e7126, 2017.

Tim Miller. Explanation in artificial intelligence: Insights from the social sciences. *Artificial intelligence*, 267:1–38, 2019.

Christoph Molnar. *Interpretable Machine Learning*. 2019.

W James Murdoch, Chandan Singh, Karl Kumbier, Reza Abbasi-Asl, and Bin Yu. Definitions, methods, and applications in interpretable machine learning. *Proceedings of the National Academy of Sciences*, 116(44):22071–22080, 2019.

Cristhian Parra, Patricia Silveira, Iman Khaghani Far, Florian Daniel, Eling D De Bruin, Luca Cernuzzi, Vincenzo D'Andrea, and Fabio Casati. Information technology for active ageing: A review of theory and practice. 2014.

Prasad Patil. What is exploratory data analysis?, May 2022. URL `https://towardsdatascience.com/exploratory-data-analysis-8fc1cb20fd15`.

Daniel Hansen Pedersen, Marjan Mansourvar, Camilla Sortsø, and Thomas Schmidt. Predicting dropouts from an electronic health platform for lifestyle interventions: analysis of methods and predictors. *Journal of medical Internet research*, 21(9):e13617, 2019.

Lucas A Ramos, Matthijs Blankers, Guido van Wingen, Tamara de Bruijn, Steffen C Pauws, and Anneke E Goudriaan. Predicting success of a digital self-help intervention for alcohol and substance use with machine learning. *Frontiers in psychology*, page 3931, 2021.

Payam Refaeilzadeh, Lei Tang, and Huan Liu. Cross-validation. *Encyclopedia of database systems*, 5:532–538, 2009.

Eduardo Sabaté, Eduardo Sabaté, et al. *Adherence to long-term therapies: evidence for action*. World Health Organization, 2003.

Floor Sieverink, Saskia M Kelders, and Julia EWC van Gemert-Pijnen. Clarifying the concept of adherence to ehealth technology: systematic review on when usage becomes adherence. *Journal of medical Internet research*, 19(12):e8578, 2017.

Viswanath Venkatesh, Michael G Morris, Gordon B Davis, and Fred D Davis. User acceptance of information technology: Toward a unified view. *MIS quarterly*, pages 425–478, 2003.

Appendices

Appendix A

Exploratory Data Analysis

Figure A.1: Missing values density analysis of the Sociodemo table features

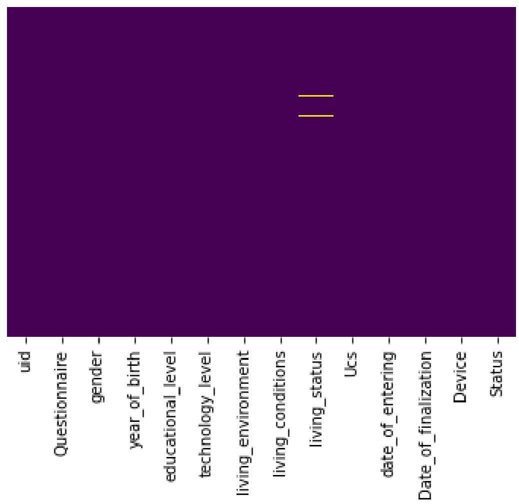


Figure A.2: Correlation analysis between the Sociodemo table features

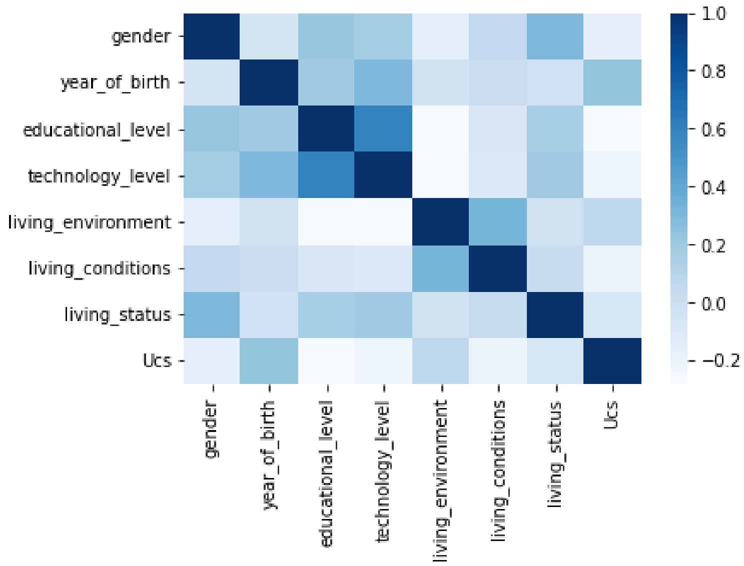


Figure A.3: Boxplots of the Sociodemo table features

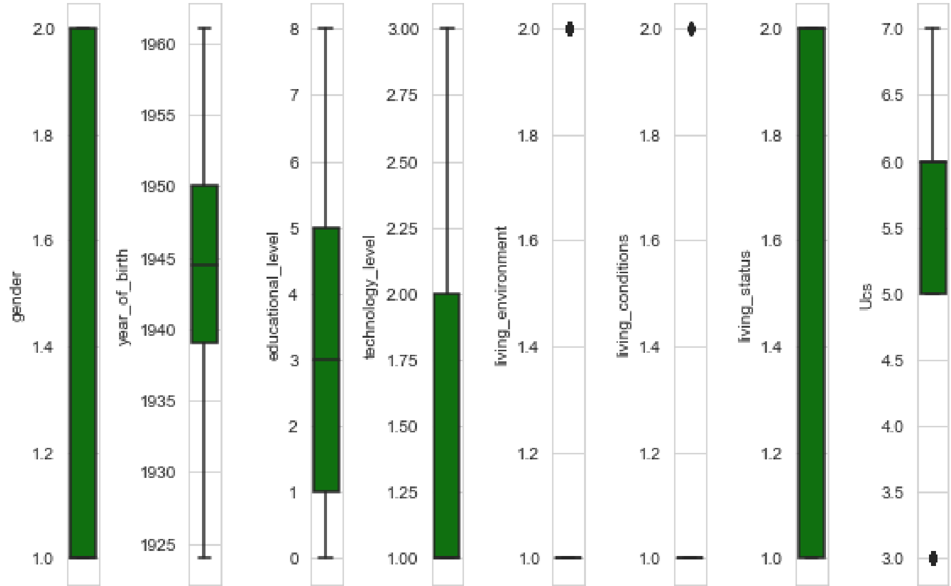


Figure A.4: Distribution analysis of the Sociodemo table features

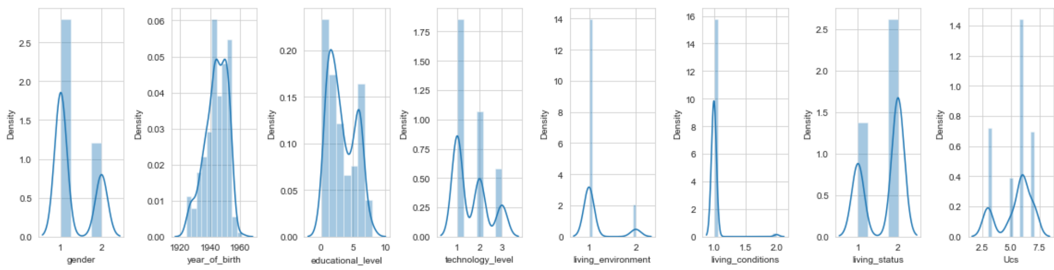


Figure A.5: Missing values density analysis of the UTAUT table features

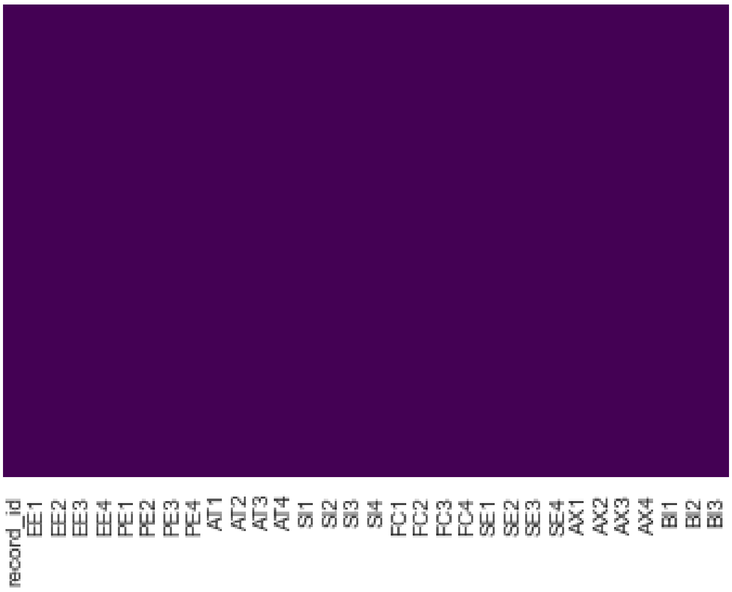


Figure A.6: Correlation analysis between the UTAUT table features

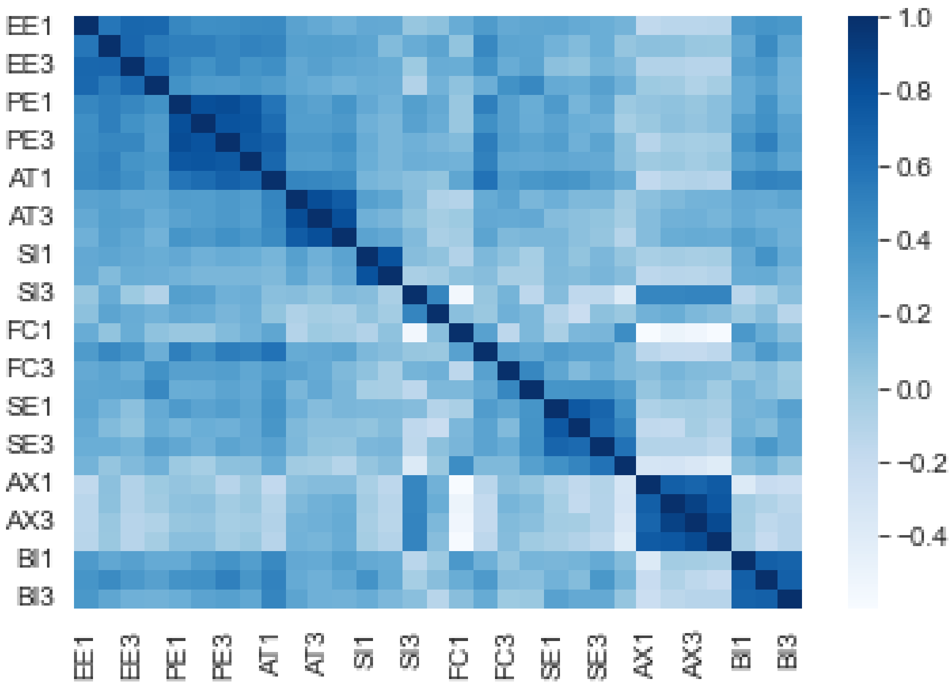


Figure A.7: Boxplots of the UTAUT table features

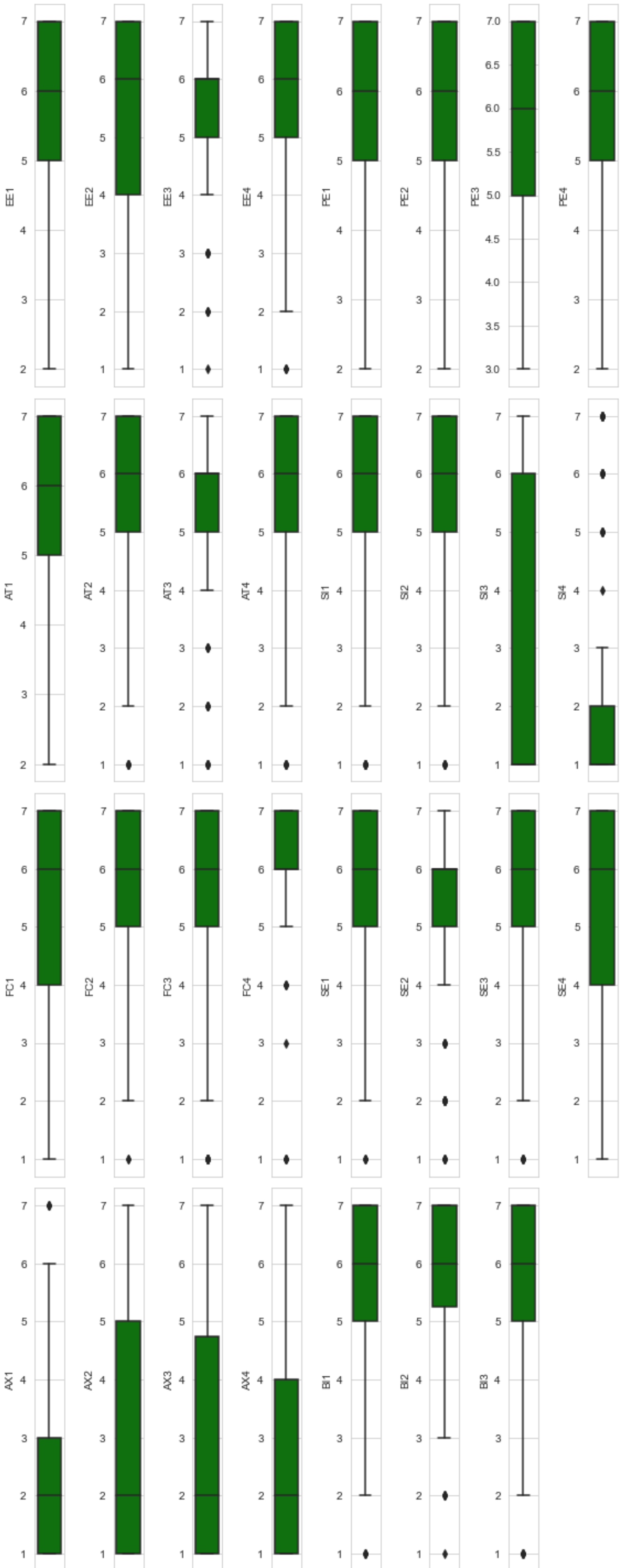


Figure A.8: Distribution analysis of the UTAUT table features

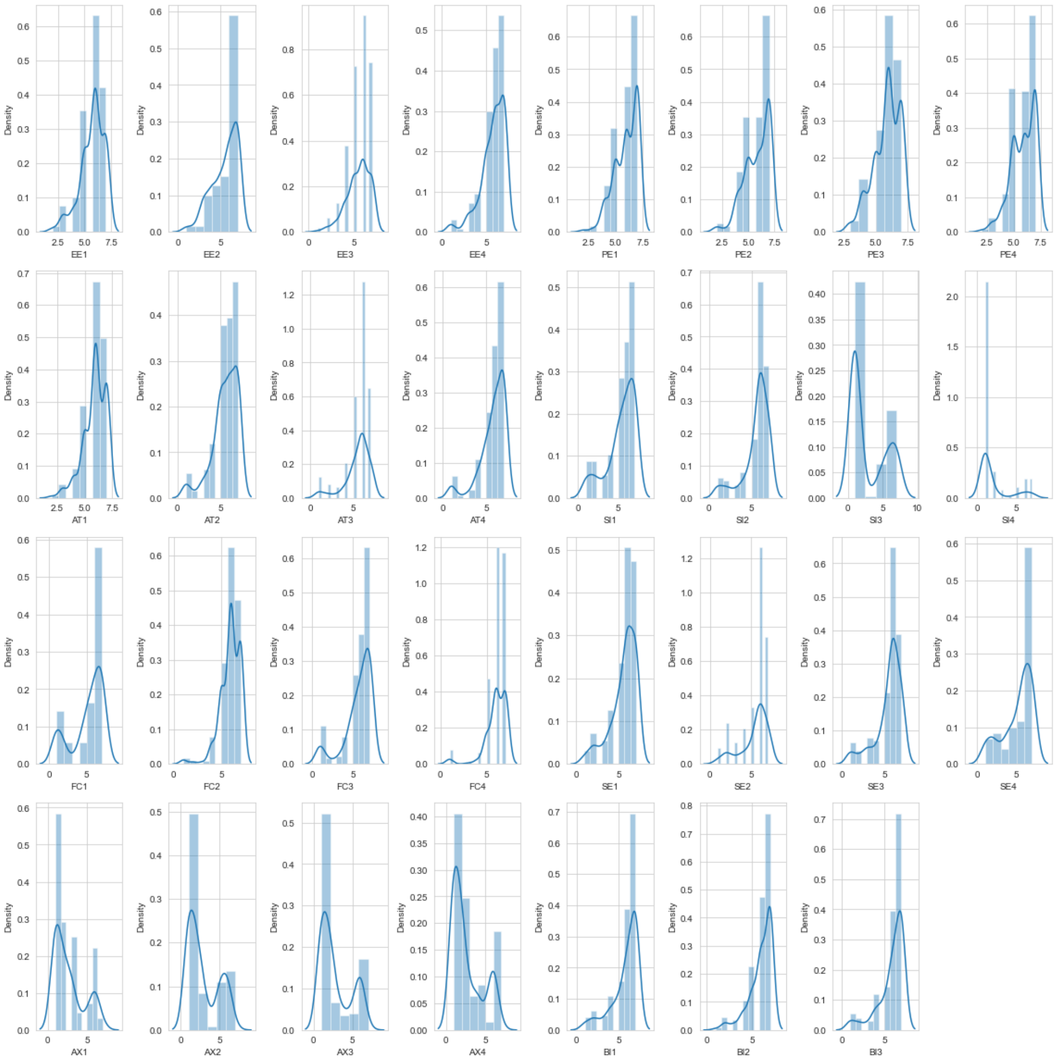


Figure A.9: Missing values density analysis of the EQ5D3L table features

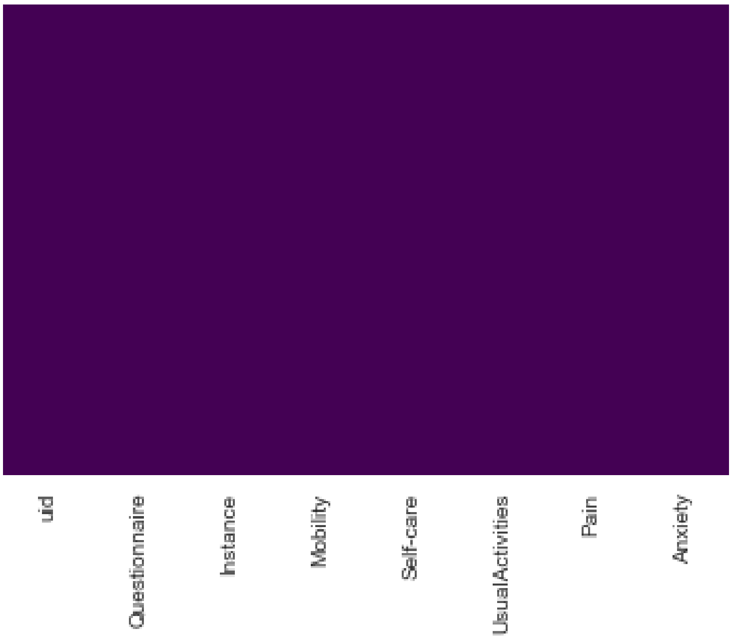


Figure A.10: Correlation analysis between the EQ5D3L table features

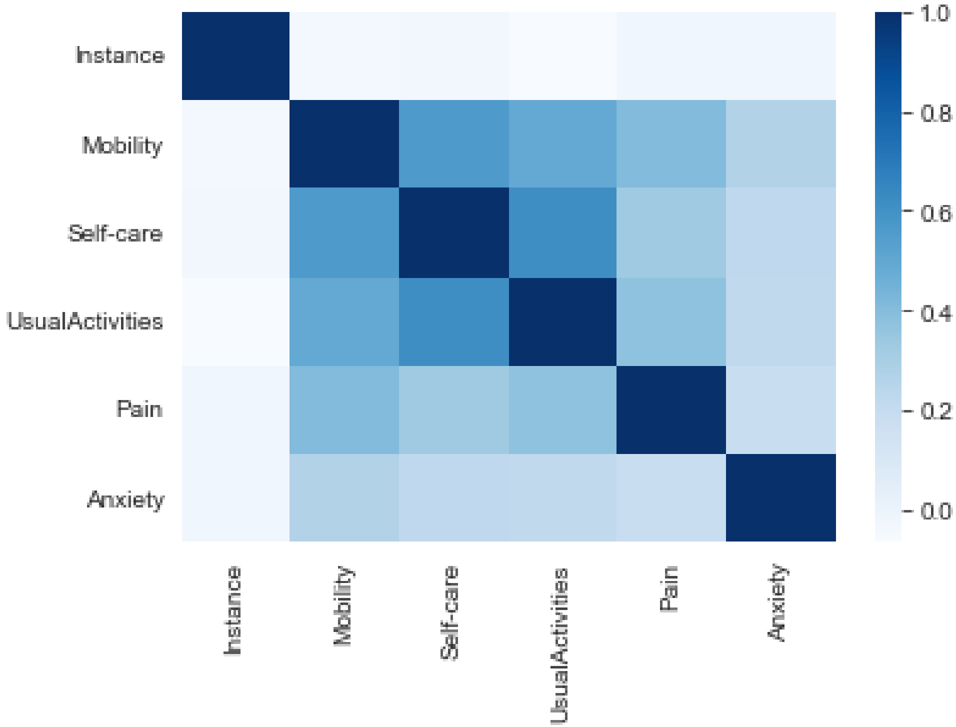


Figure A.11: Boxplots of the EQ5D3L table features

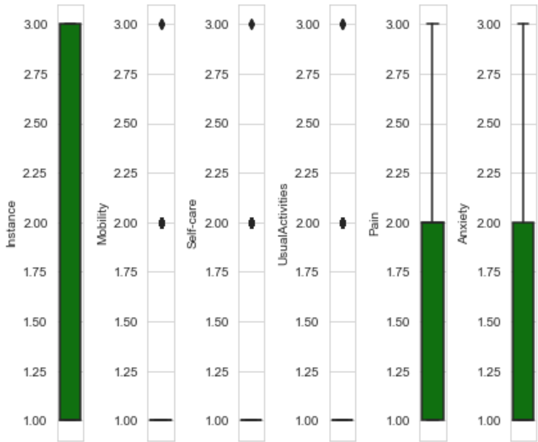


Figure A.12: Distribution analysis of the EQ5D3L table features

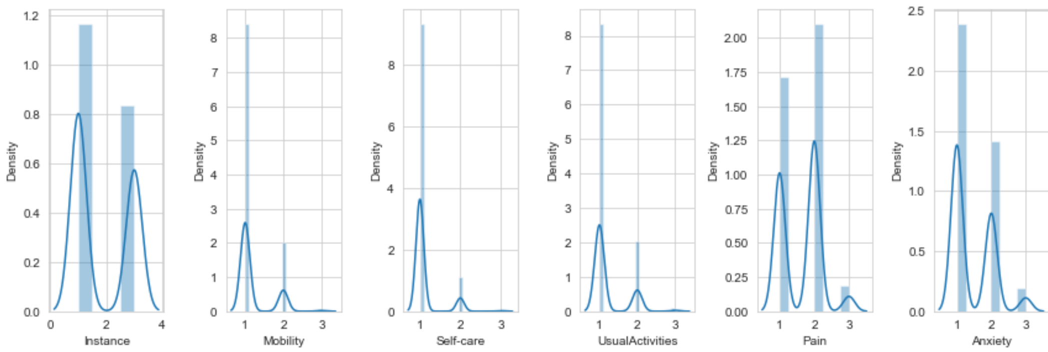


Figure A.13: Missing values density analysis of the SPQ table features

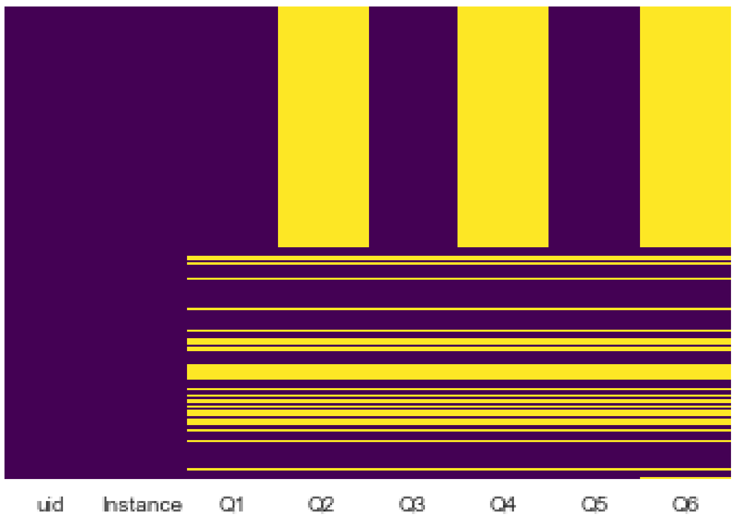


Figure A.14: Correlation analysis between the SPQ table features

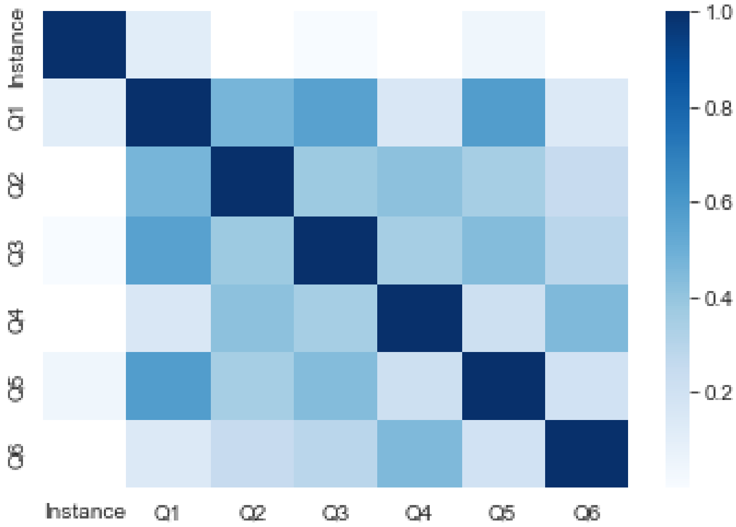


Figure A.15: Boxplots of the SPQ table features

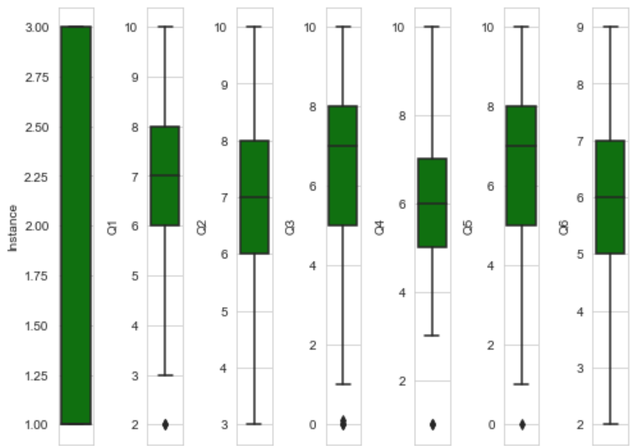


Figure A.16: Distribution analysis of the SPQ table features

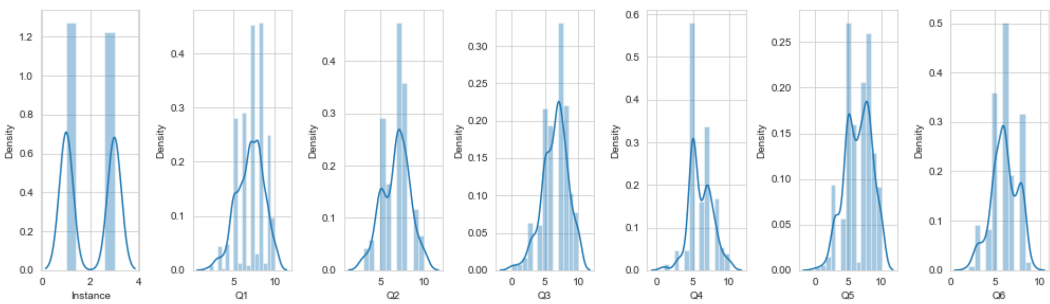


Figure A.17: Missing values density analysis of the UCLA table features

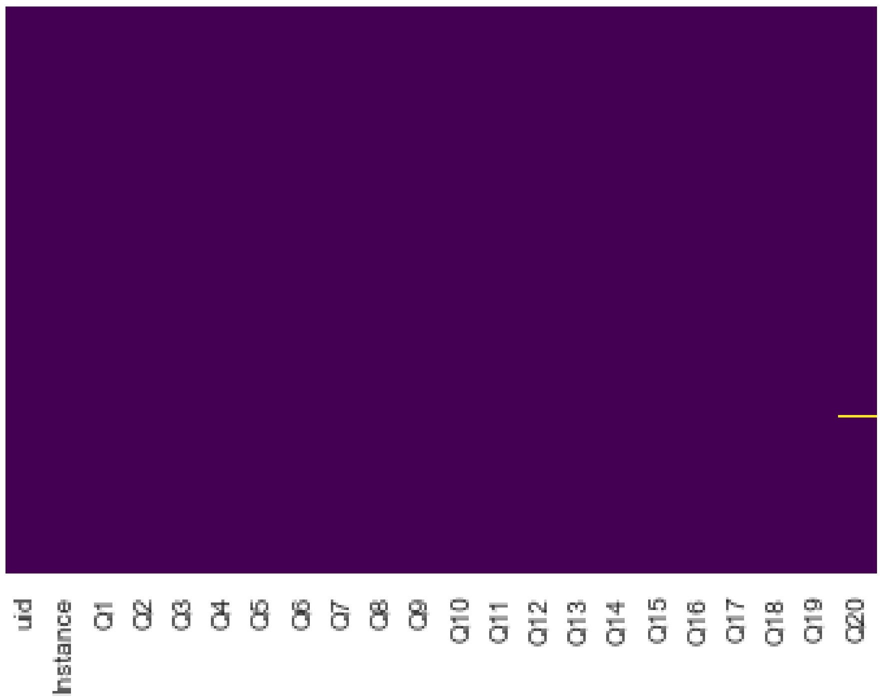


Figure A.18: Correlation analysis between the UCLA table features

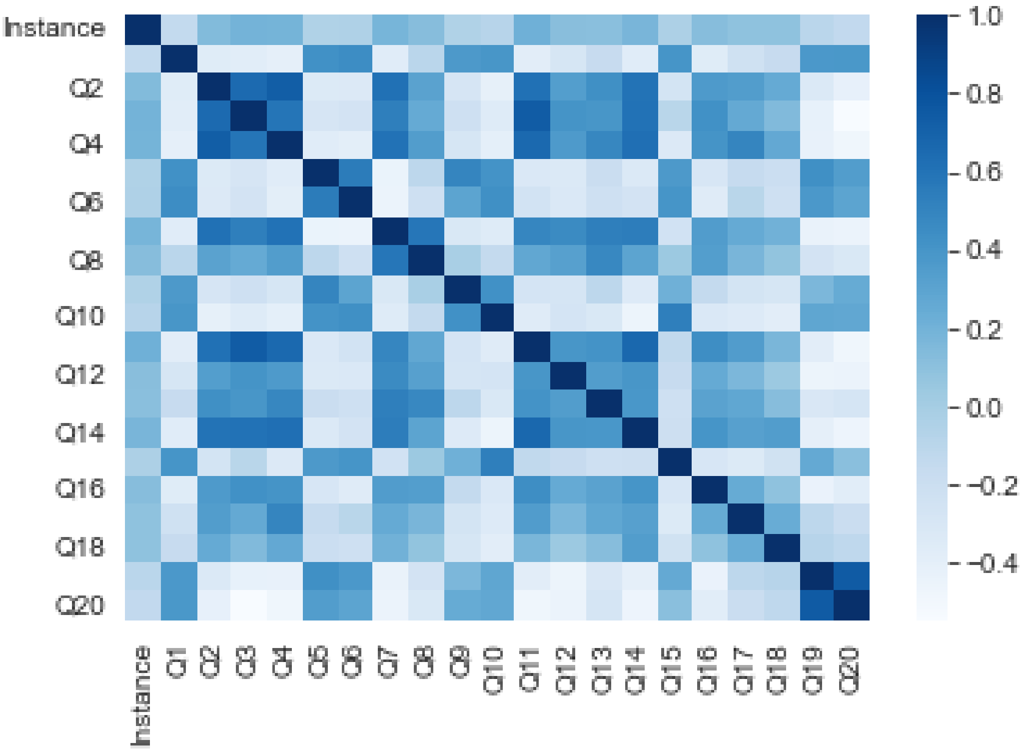


Figure A.19: Boxplots of the UCLA table features

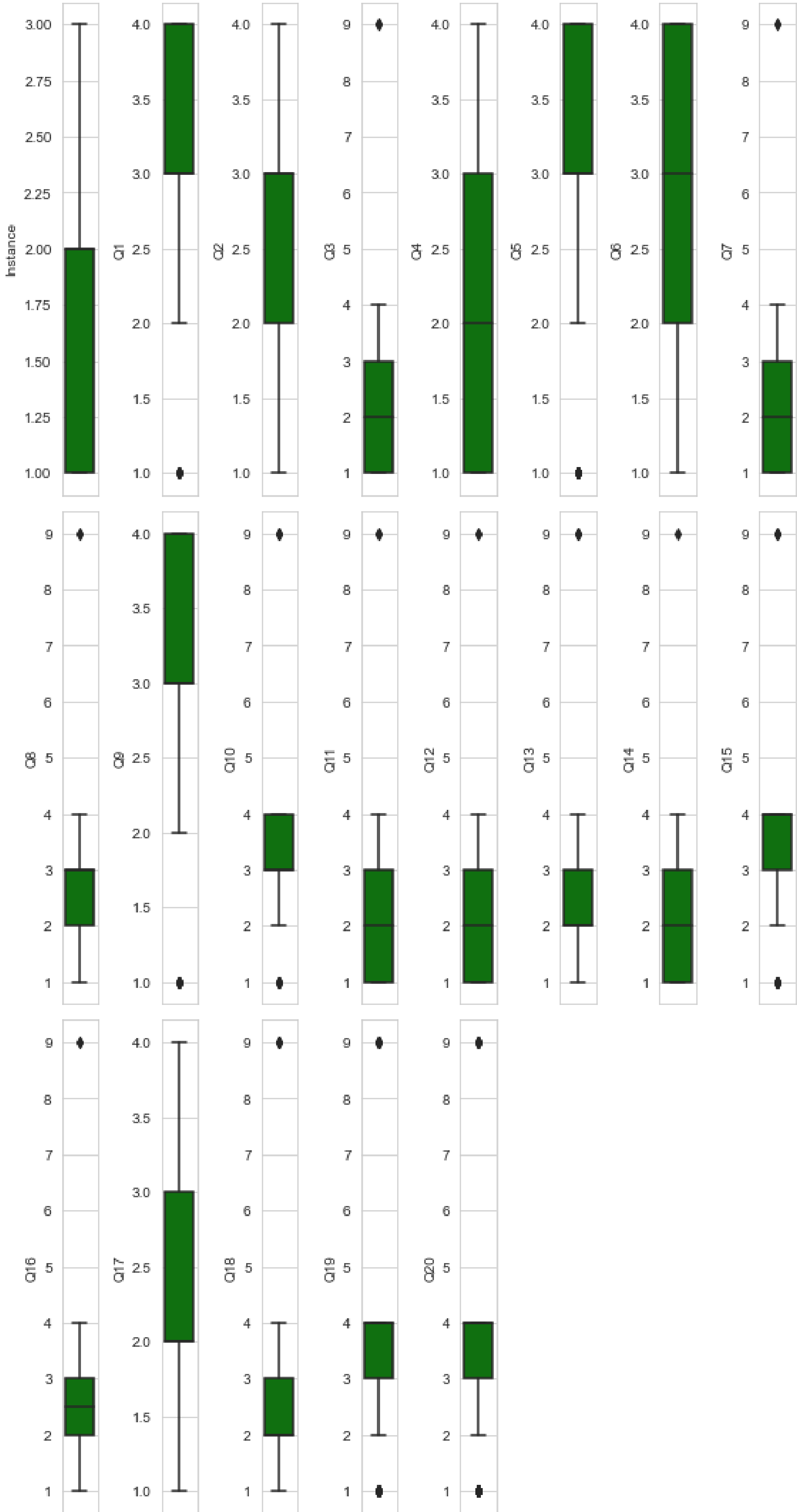


Figure A.20: Distribution analysis of the UCLA table features

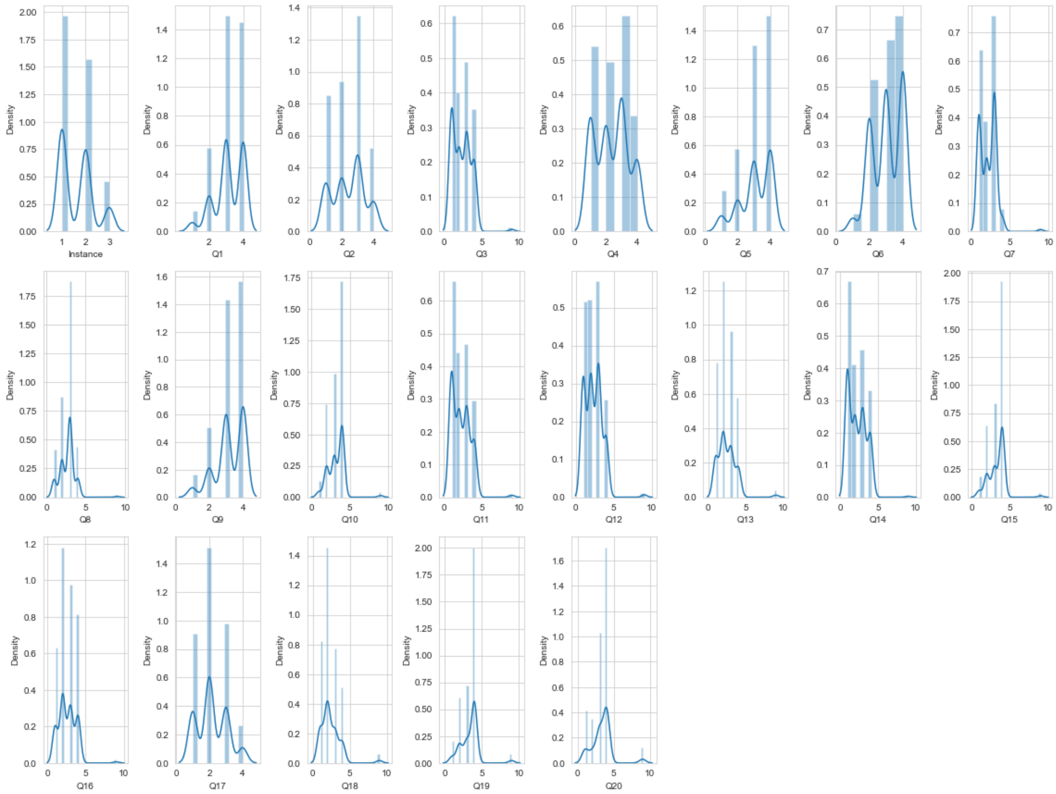


Figure A.21: Missing values density analysis of the Physical Activity table features

Figure A.22: Missing values density analysis of the Mindfulness table features

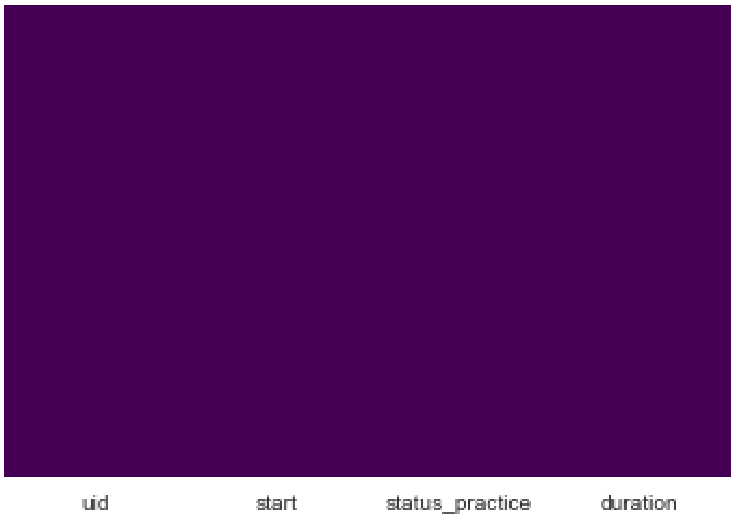


Figure A.23: Boxplots of the Mindfulness table features

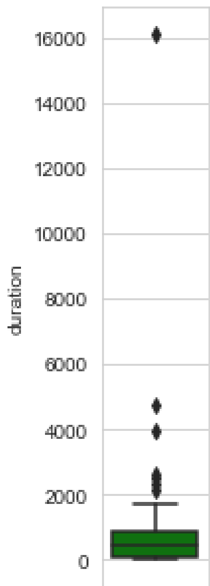


Figure A.24: Distribution analysis of the Mindfulness table features

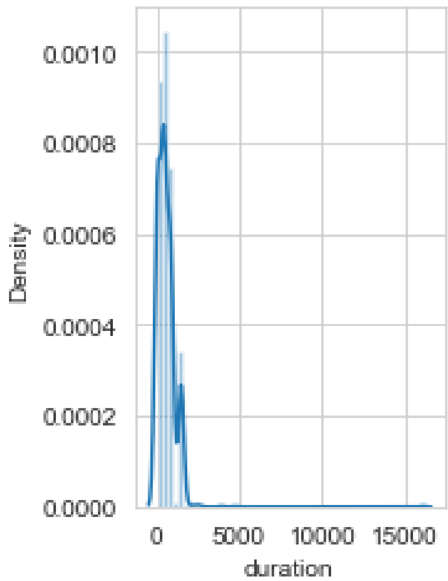


Figure A.25: Missing values density analysis of the Finger tapping table features

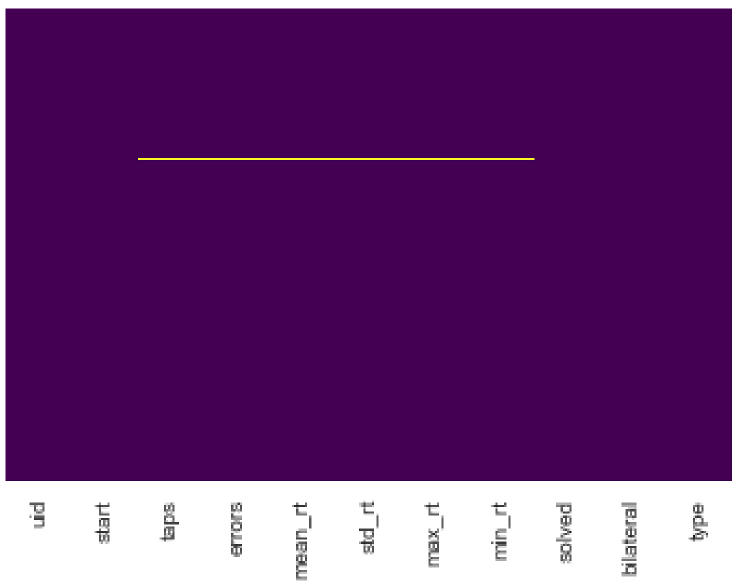


Figure A.26: Correlation analysis between the Finger tapping table features

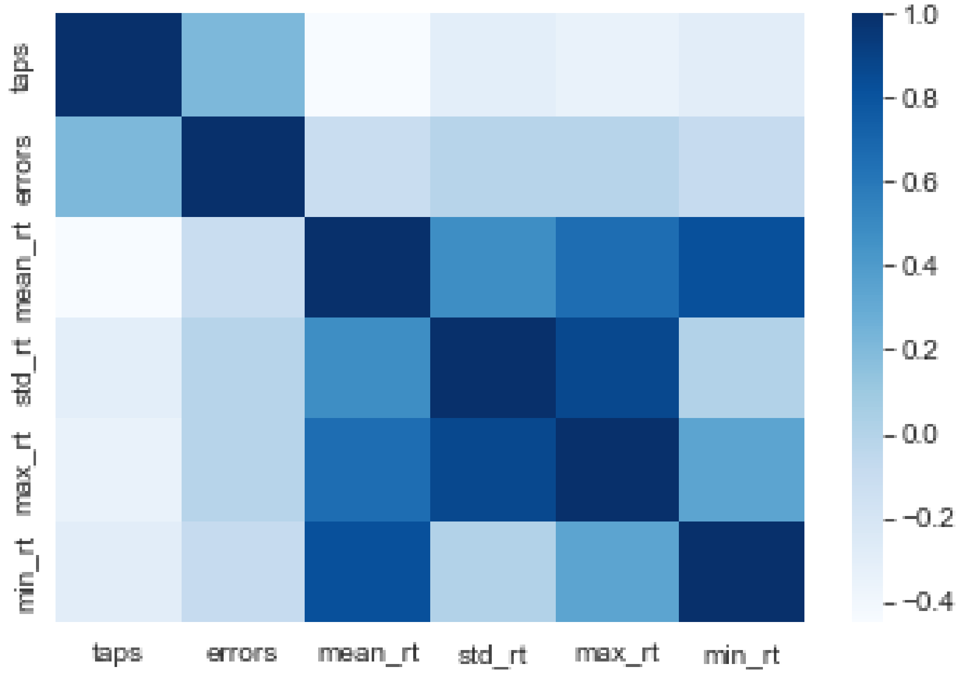


Figure A.27: Boxplots of the Finger tapping table features

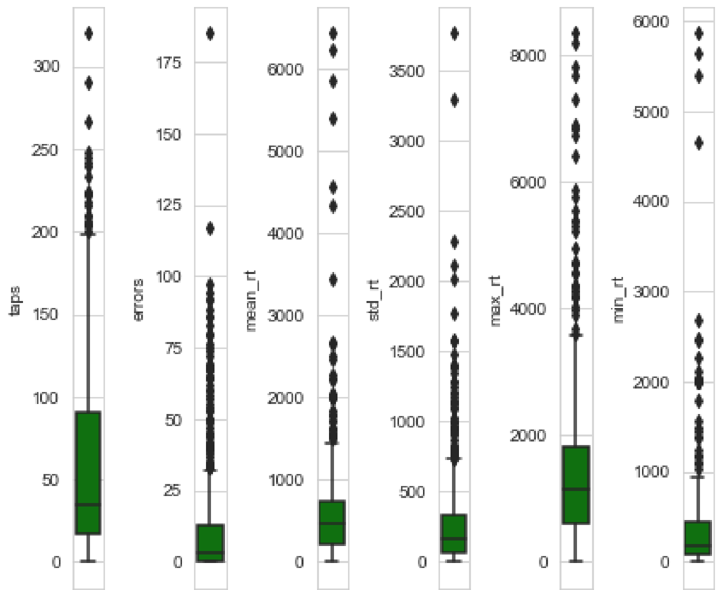


Figure A.28: Distribution analysis of the Finger tapping table features

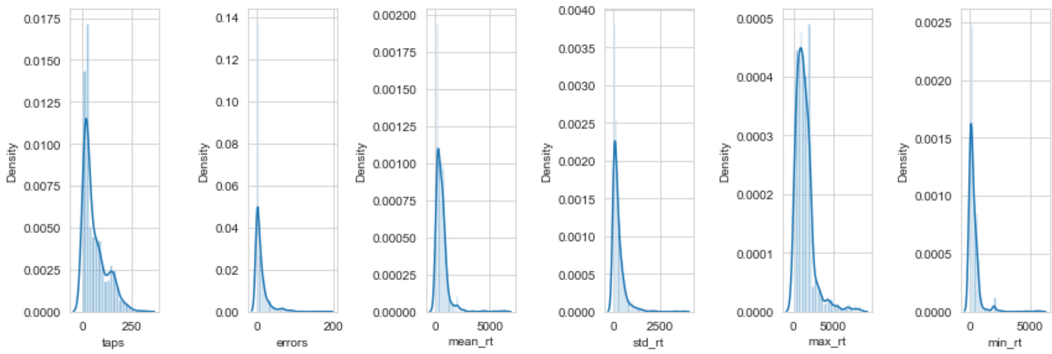


Figure A.29: Missing values density analysis of the Brain Games table features

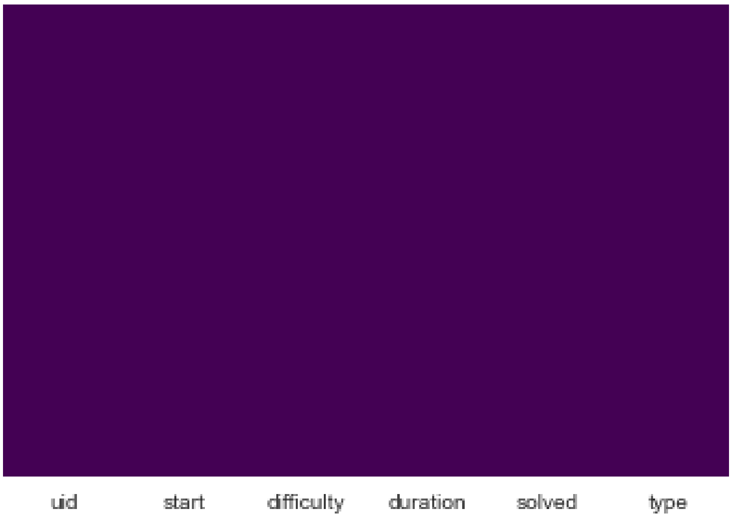


Figure A.30: Correlation analysis between the Brain Games table features

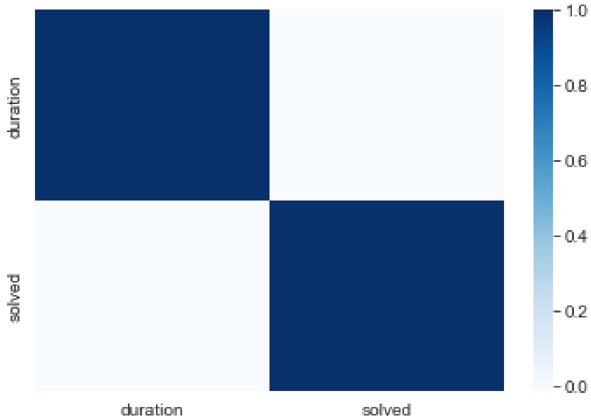


Figure A.31: Boxplots of the Brain Games table features

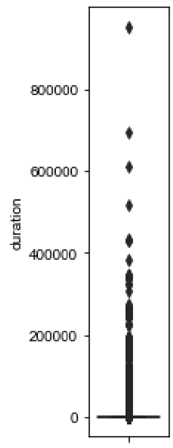


Figure A.32: Distribution analysis of the Brain Games table features

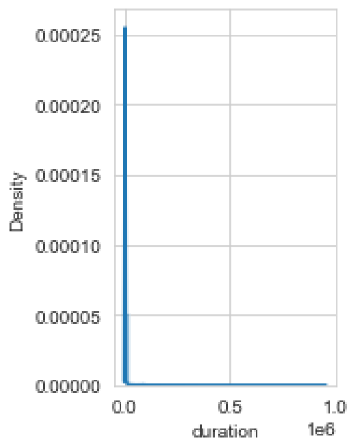


Figure A.33: Missing values density analysis of the Digital Phenotype table features

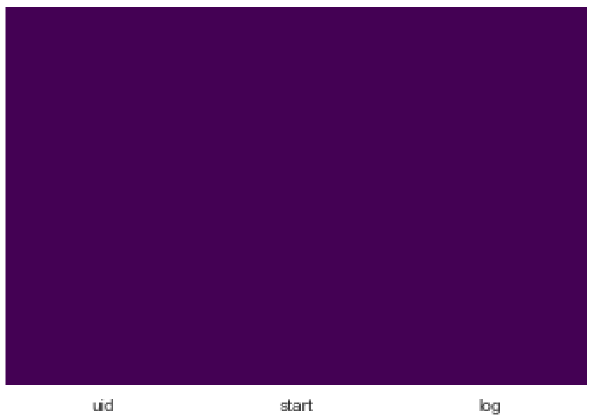


Figure A.34: Missing values density analysis of the Users table features

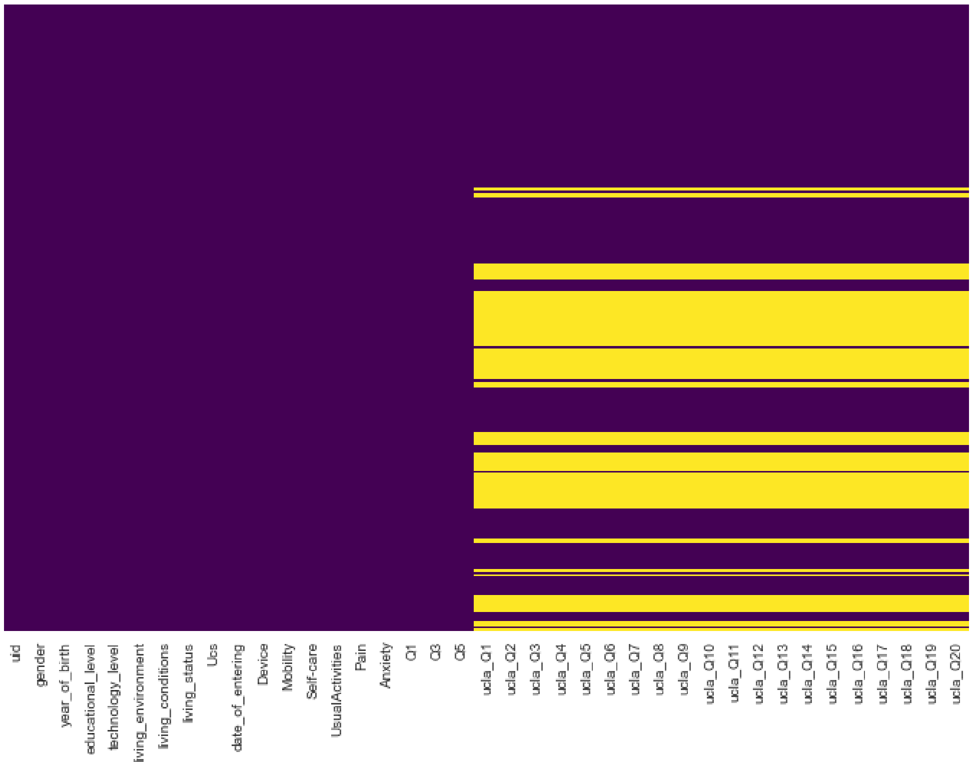


Figure A.35: Correlation analysis between the Users table features

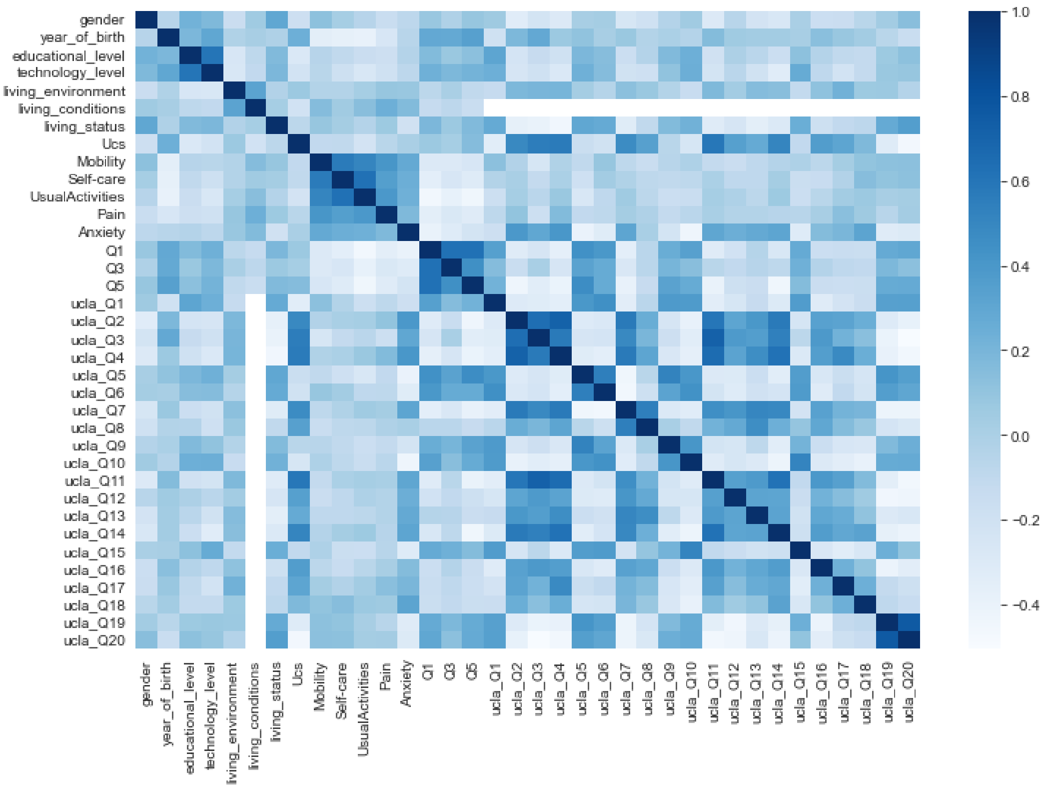


Figure A.36: Boxplots of the Users table features

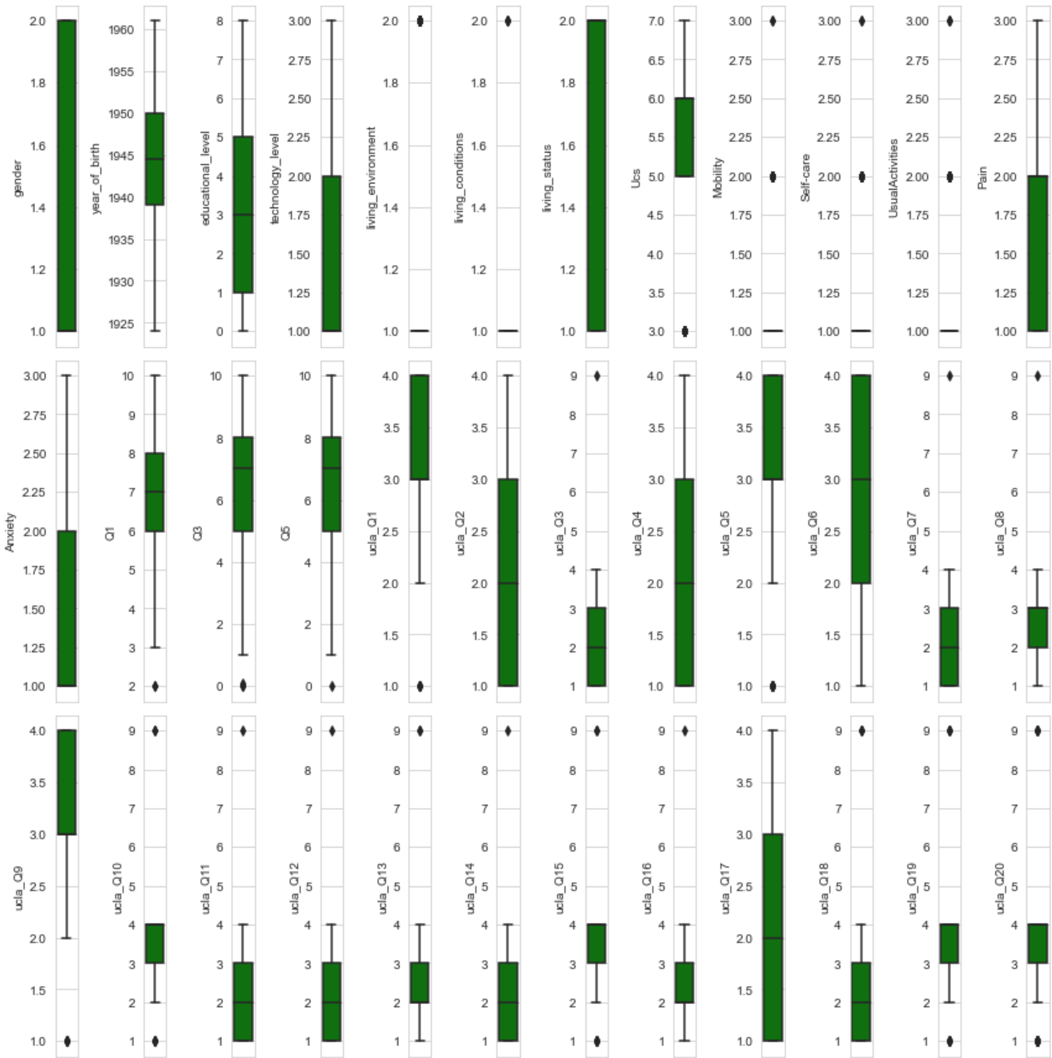
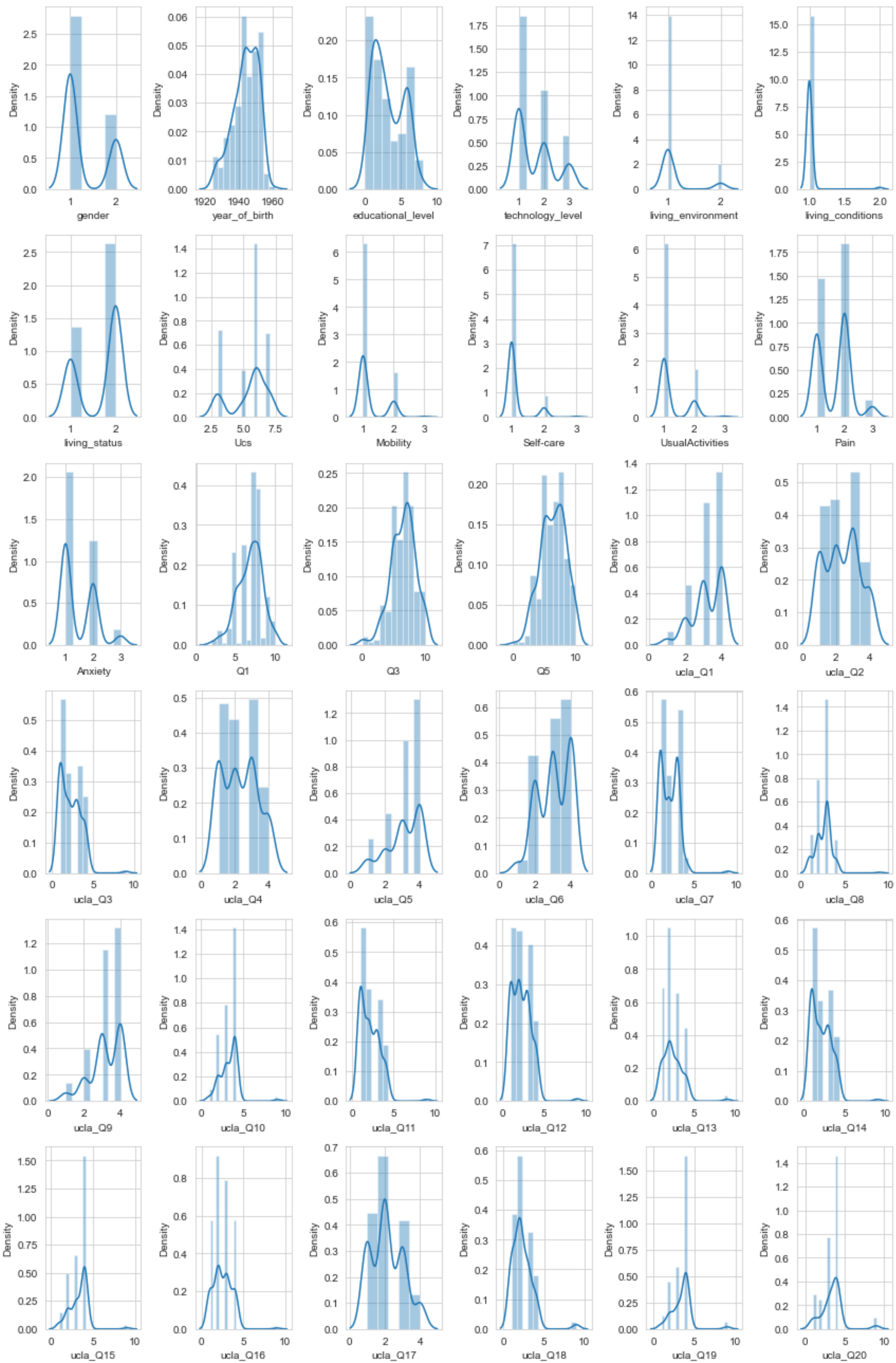


Figure A.37: Distribution analysis of the Users table features



Appendix B

Optimization Parameters

Table B.1: Parameters tested for each algorithm during Model Optimization.

Algorithm	**Parametres**	**Values**
	n_neighbors	list(range(1, 35, 2))
kNN	weights	['uniform', 'distance']
	leaf_size	list(range(1, 11, 2))
	metric	['euclidean', 'manhattan']
Decision Tree	criterion	['gini', 'entropy']
	max_depth	[2,3,4,5,10,15]
	n_estimators	list(range(5, 35, 5))
Random Forest	criterion	['gini', 'entropy']
	min_samples_leaf	[1,2,3]
	min_samples_split	[2,3,4,5,10,15]
	max_features	['auto', 'log2']
	max iter	[100, 200]
MLP	activation	['tanh', 'relu']
	solver	['sgd', 'adam']
	alpha	[0.0001, 0.01, 0.05]
	learning_rate	['constant','adaptive']

Appendix C

Results

Table C.1: Approach 1 results (Average RMSE ± SD)

	Linear Regression	Lasso Regression	Random Forest Regressor	Decision Tree Regressor	SVM
1 / 1	0.0 ± 0.0	0.038 ± 0.011	0.093 ± 0.093	0.0 ± 0.0	26.137 ± 7.215
1 / 2	0.0 ± 0.0	0.035 ± 0.011	0.256 ± 0.309	0.247 ± 0.315	29.918 ± 6.145
1 / 4	0.0 ± 0.0	0.034 ± 0.01	0.276 ± 0.217	0.411 ± 0.261	31.717 ± 7.916
2 / 1	0.0 ± 0.0	0.04 ± 0.021	0.168 ± 0.18	0.0 ± 0.0	28.273 ± 5.943
2 / 2	0.0 ± 0.0	0.037 ± 0.015	0.22 ± 0.195	0.233 ± 0.282	29.379 ± 9.516
2 / 4	0.0 ± 0.0	0.035 ± 0.02	0.291 ± 0.182	0.338 ± 0.271	32.026 ± 13.154
4 / 1	0.0 ± 0.0	0.039 ± 0.008	0.741 ± 0.688	0.751 ± 0.977	27.195 ± 5.92
4 / 2	0.0 ± 0.0	0.039 ± 0.026	0.382 ± 0.365	0.444 ± 0.453	30.673 ± 13.305
4 / 4	0.0 ± 0.0	0.038 ± 0.025	0.417 ± 0.295	0.51 ± 0.45	31.59 ± 10.945

Table C.2: Approach 2 results (Average F1-score ± SD)

	Decision Tree	**Gaussian Naive Bayes**	**kNN**	**Random Forest**	**SVM**	**MLP**
1 / 1	0.925 ± 0.009	0.921 ± 0.007	0.907 ± 0.009	**0.94 ± 0.005**	0.906 ± 0.01	0.915 ± 0.019
1 / 2	0.911 ± 0.002	0.901 ± 0.009	0.888 ± 0.007	**0.93 ± 0.007**	0.886 ± 0.011	0.894 ± 0.04
1 / 4	0.9 ± 0.011	0.885 ± 0.007	0.892 ± 0.013	**0.923 ± 0.014**	0.876 ± 0.008	0.875 ± 0.019
2 / 1	0.921 ± 0.009	0.915 ± 0.008	0.902 ± 0.005	**0.942 ± 0.005**	0.885 ± 0.006	0.899 ± 0.016
2 / 2	0.909 ± 0.014	0.911 ± 0.004	0.89 ± 0.009	**0.937 ± 0.011**	0.887 ± 0.018	0.895 ± 0.028
2 / 4	0.887 ± 0.028	0.893 ± 0.027	0.883 ± 0.031	**0.915 ± 0.024**	0.874 ± 0.037	0.869 ± 0.031
4 / 1	0.926 ± 0.01	0.918 ± 0.007	0.902 ± 0.011	**0.942 ± 0.014**	0.903 ± 0.016	0.898 ± 0.019
4 / 2	0.914 ± 0.018	0.9 ± 0.021	0.877 ± 0.03	**0.929 ± 0.015**	0.869 ± 0.035	0.865 ± 0.044
4 / 4	0.91 ± 0.011	0.893 ± 0.015	0.871 ± 0.008	**0.925 ± 0.016**	0.861 ± 0.016	0.867 ± 0.029

Table C.3: Approach 3 results (Average F1-score ± SD)

	Decision Tree	Gaussian Naive Bayes	kNN	Random Forest	SVM	MLP
1 / 1	0.682 ± 0.057	0.623 ± 0.07	**0.693 ± 0.025**	**0.694 ± 0.035**	0.552 ± 0.015	0.678 ± 0.073
1 / 2	**0.712 ± 0.036**	0.574 ± 0.067	0.695 ± 0.04	**0.73 ± 0.046**	0.632 ± 0.047	0.627 ± 0.078
1 / 4	0.687 ± 0.059	0.568 ± 0.068	**0.736 ± 0.046**	**0.759 ± 0.042**	0.553 ± 0.052	0.549 ± 0.172
2 / 1	0.66 ± 0.079	0.536 ± 0.039	0.595 ± 0.058	**0.723 ± 0.064**	0.468 ± 0.061	0.525 ± 0.129
2 / 2	0.684 ± 0.021	0.614 ± 0.06	0.673 ± 0.036	**0.774 ± 0.061**	0.519 ± 0.056	0.644 ± 0.079
2 / 4	**0.739 ± 0.028**	0.583 ± 0.039	0.71 ± 0.082	**0.723 ± 0.036**	0.522 ± 0.064	0.622 ± 0.029
4 / 1	0.72 ± 0.028	0.616 ± 0.078	0.686 ± 0.04	**0.749 ± 0.022**	0.642 ± 0.038	0.593 ± 0.108
4 / 2	**0.714 ± 0.04**	0.556 ± 0.098	0.613 ± 0.095	**0.728 ± 0.044**	0.43 ± 0.045	0.624 ± 0.09
4 / 4	0.645 ± 0.068	0.536 ± 0.074	0.595 ± 0.01	**0.697 ± 0.048**	0.438 ± 0.092	0.459 ± 0.181

Table C.4: Approach 3 MLP hyperparameter search results (Average F1-score)

Window Lengths	Score	Parametres	Values
1 / 1	0.7250	max iter	200
		activation	tanh
		solver	adam
		alpha	0.05
		learning_rate	adaptive
1 / 2	0.7337	max iter	200
		activation	tanh
		solver	adam
		alpha	0.01
		learning_rate	adaptive
1 / 4	**0.7564**	max iter	100
		activation	tanh
		solver	adam
		alpha	0.05
		learning_rate	adaptive
2 / 1	0.6693	max iter	100
		activation	tanh
		solver	adam
		alpha	0.05
		learning_rate	adaptive
2 / 2	**0.7380**	max iter	100
		activation	tanh
		solver	adam
		alpha	0.05
		learning_rate	adaptive
2 / 4	**0.7449**	max iter	200
		activation	tanh
		solver	adam
		alpha	0.0001
		learning_rate	adaptive
4 / 1	0.6893	max iter	200
		activation	tanh
		solver	adam
		alpha	0.05
		learning_rate	adaptive
4 / 2	0.6494	max iter	100
		activation	tanh
		solver	adam
		alpha	0.05
		learning_rate	adaptive
4 / 4	0.6565	max iter	200
		activation	tanh
		solver	sgd
		alpha	0.01
		learning_rate	adaptive

Table C.5: Approach 3 kNN hyperparameter search results (Average F1-score)

Window Lengths	Score	Parametres	Values
1 / 1	0.7321	n_neighbors	29
		weights	uniform
		leaf_size	1
		metric	manhattan
1 / 2	**0.7523**	n_neighbors	21
		weights	uniform
		leaf_size	1
		metric	euclidean
1 / 4	**0.7807**	n_neighbors	23
		weights	distance
		leaf_size	1
		metric	manhattan
2 / 1	0.6456	n_neighbors	13
		weights	uniform
		leaf_size	1
		metric	manhattan
2 / 2	0.7141	n_neighbors	17
		weights	uniform
		leaf_size	1
		metric	euclidean
2 / 4	0.7291	n_neighbors	9
		weights	distance
		leaf_size	1
		metric	manhattan
4 / 1	0.7027	n_neighbors	5
		weights	uniform
		leaf_size	1
		metric	manhattan
4 / 2	0.6396	n_neighbors	11
		weights	distance
		leaf_size	1
		metric	manhattan
4 / 4	0.6629	n_neighbors	17
		weights	uniform
		leaf_size	1
		metric	manhattan

Table C.6: Approach 3 Decision Tree hyperparameter search results (Average F1-score)

Window Lengths	Score	Parametres	Values
1 / 1	0.7066	criterion	gini
		max_depth	5
1 / 2	0.7669	criterion	gini
		max_depth	2
1 / 4	0.7254	criterion	gini
		max_depth	15
2 / 1	0.7160	criterion	entropy
		max_depth	20
2 / 2	0.7383	criterion	entropy
		max_depth	2
2 / 4	0.7490	criterion	gini
		max_depth	3
4 / 1	0.7266	criterion	entropy
		max_depth	2
4 / 2	0.7405	criterion	gini
		max_depth	2
4 / 4	0.6840	criterion	gini
		max_depth	20

Table C.7: Approach 3 Random Forest hyperparameter search results (Average F1-score)

Window Lengths	**Score**	**Parametres**	**Values**
1 / 1	0.7139	n_estimators	5
		criterion	entropy
		min_samples_leaf	1
		min_samples_split	15
		max_features	auto
1 / 2	**0.7806**	n_estimators	20
		criterion	gini
		min_samples_leaf	1
		min_samples_split	2
		max_features	auto
1 / 4	**0.7918**	n_estimators	25
		criterion	entropy
		min_samples_leaf	2
		min_samples_split	5
		max_features	log2
2 / 1	0.7542	n_estimators	30
		criterion	entropy
		min_samples_leaf	1
		min_samples_split	2
		max_features	auto
2 / 2	**0.8104**	n_estimators	25
		criterion	gini
		min_samples_leaf	1
		min_samples_split	5
		max_features	auto
2 / 4	0.7605	n_estimators	5
		criterion	gini
		min_samples_leaf	1
		min_samples_split	4
		max_features	log2
4 / 1	**0.7823**	n_estimators	20
		criterion	entropy
		min_samples_leaf	1
		min_samples_split	5
		max_features	log2
4 / 2	**0.7829**	n_estimators	30
		criterion	entropy
		min_samples_leaf	1
		min_samples_split	10
		max_features	auto
4 / 4	0.7632	n_estimators	25
		criterion	gini
		min_samples_leaf	2
		min_samples_split	15
		max_features	log2

Figure C.1: Decision Tree Feature Selection with Hyperparameterization

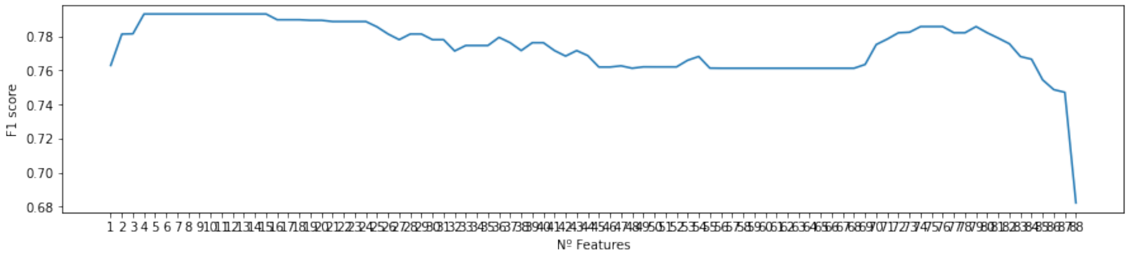


Figure C.2: Random Forest Feature Selection with Hyperparameterization

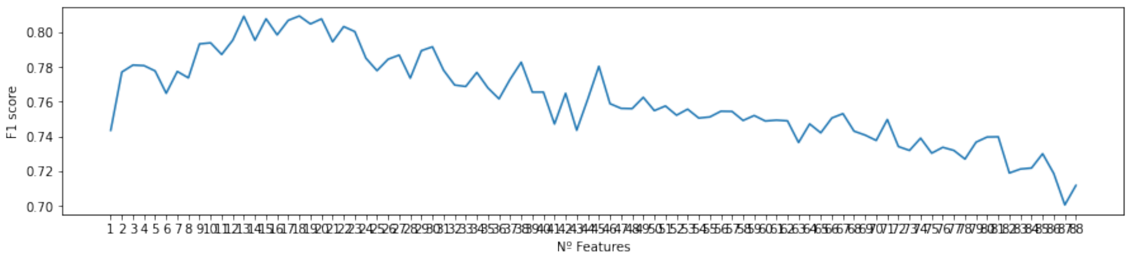


Figure C.3: Decision Tree