Universidade do Minho
Escola de Engenharia
Departamento de Informática

Nuno Ricardo Araújo Silva

Age prediction through the influence of fatigue levels in human-computer interaction

May 2022

Universidade do Minho
Escola de Engenharia
Departamento de Informática

Nuno Ricardo Araújo Silva

Age prediction through the influence of fatigue levels in human-computer interaction

Master dissertation
Integrated Master's in Informatics Engineering

Dissertation supervised by
Paulo Jorge Freitas de Oliveira Novais
André Pimenta Ribeiro

May 2022

STATEMENT OF INTEGRITY

I hereby declare having conducted this academic work with integrity. I confirm that I have not used plagiarism or any form of undue use of information or falsification of results along the process leading to its elaboration.

I further declare that I have fully acknowledged the Code of Ethical Conduct of the University of Minho.

ACKNOWLEDGEMENTS

The completion of this dissertation was an enriching experience on all levels. From the study of new areas and investigation of new Machine Learning techniques, to the possibility of correlating different areas and aspects to achieve the intended purpose. However, this whole experience presented itself as an obstacle course which was only possible to complete with the support of several people, to whom I am deeply grateful.

To my advisors, Professor Paulo Novais and André Pimenta, PhD, CEO & Co-Founder at Anybrain, for their patience, experience, encouragement and critical spirit while investigating this work. Furthermore, to Professor Davide Carneiro and Head of Data Science at Anybrain, Pedro Duarte, for their advice and suggestions.

To my family for their support throughout my years as a student. To the twenty-six for accompanying me in this five-year course and also to my closest friends for their suggestions and support that enriched this dissertation.

Finally to Telma Torres for all her support, encouragement and patience, which allowed me to achieve the intended results as well as to keep my focus throughout this journey.

ABSTRACT

The evolution of current times and the available technology is making it easier to access potentially inappropriate content. Therefore, the ability to detect the age of the human being, by non-invasive methods, is increasingly necessary to reduce possible false claims. All of these claims arise through interactions with the device, so, and taking into account the demands and the fast pace of everyday life, the intent is to develop a system capable of detecting age groups, taking into account the presence of human factors like fatigue or stress that can change the interaction patterns. This system will use biometric features created by keyboard and mouse events, describing typing velocity, mouse acceleration, and so on in the period of five minutes. However, keeping in mind the everyday pace and the growth in mobile phone use, a similar system is created for this case study.

Keywords: Age, Detection System, Fatigue, Interactions

RESUMO

A evolução dos tempos modernos e das tecnologias existentes está a tornar mais fácil o acesso a conteúdos potencialmente impróprios. Assim, a capacidade para detetar a idade de um ser humano, por métodos não invasivos, é cada vez mais necessário de forma a reduzir potenciais falsas alegações. Todas estas alegações provêm através de interações com um dispositivo, dessa forma, e tendo em conta as exigências e o ritmo acelerado do quotidiano, o objetivo passa pelo desenvolvimento de um sistema capaz de detetar idades, considerando a presença de fatores humanos que poderão influenciar os padrões de interação, como fadiga ou stress. Este sistema irá utilizar biometrias criadas a partir de eventos de teclado e rato, descrevendo velocidade de escrita, aceleração do rato, entre outras no período de cinco minutos. Contudo, tendo em conta o ritmo acelerado do quotidiano e crescimento do uso de telemóveis, um sistema similar é criado para este caso estudo.

Palavras-Chave: Fadiga, Idade, Interações, Sistemas Deteção

CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACRONYMS

A

AC Affective Computing.

AI Artificial Intelligence.

API Application Programming Interface.

C

CRISP-DM CRoss Industry Standard Process for Data Mining.

G

GBM Gradient Boosting Machine.

GDPR General Data Protection Regulation.

H

HCI Human-Computer Interaction.

M

MAE Mean Absolute Error.

ML Machine Learning.

MRI Magnetic Resonances Imaging.

MSE Mean Squared Error.

R

RFECV Recursive Feature Elimination Cross-Validation.

RMSE Root Mean Squared Error.

S

SDK Software Development Kit.

SMOTE Synthetic Minority Oversampling Technique.

STE Short Term Energy.

X

XGBOOST Extreme Gradient Boosting.

Z

ZCR Zero-Crossing Rate.

1

INTRODUCTION

Since the 1900s, mental fatigue and its implications have been studied by several researchers (e.g. Thorndike (1900); Muscio (1921)). As such, several proposals appear to try to explain its effects:

- Frederic Barlett, claims that the effects of mental fatigue are correlated with a decreased cognitive control. In other words, a lower ability to coordinate more complex activities, may lead to an increase in inevitability and alertness to physical discomfort (Bartlett, 1941);
- Ruth Kanfer and Phillip Ackerman, indicate that the effects of mental fatigue arise due to the decrease in cognitive resources and the decline in motivation, compromising the allocation of resources for the execution of the current task (Kanfer and Ackerman, 1989);

However, new proposals have emerged, claiming that fatigue can arise as a result of a cost-benefit assessment and that it can serve as an alert system for the performance decline (Boksem and Tops, 2008). The effects of this state are presented as real consequences, there's not only a decrease in productivity but it can also be the main cause of accidents at work (de Jong et al., 2018). However, mental fatigue depends on factors such as the person or the environment in which they are inserted, and its effects can be overcome through motivation, for example through a monetary prize, or the ingestion of a cup of coffee (de Jong et al., 2020). There are still other factors that serve as an important example of how they can influence how mental fatigue is expressed, this is the case with age; yet, age is also one of the main factors in measuring mental status. It's very important to understand the effects of mental fatigue in older people, due to the average age of the working population, which has increased in the past century. Although, little is known among the interactions of age with mental fatigue in performance, as a rule, older people are slower to respond in addition to being more susceptible to making mistakes (Verhaeghen and Salthouse, 1997), on the other hand, younger populations are faster but commit more mistakes (Starns and Ratcliff, 2010).

Finally, and taking these into account, this dissertation aims to predict a users age group through its interactions with a computer. However, these interactions may be influenced by high levels of fatigue, which in turn may be intensified due to the task being performed.

1.1 MOTIVATION

Being tired is something that appears in a person's daily life as something natural and related to their life style. This work is very motivating due to the possibility of analyzing the impact of fatigue with the progression of age and how they are related to each other. Keeping in mind the possibility of creating a robust system capable of reducing the number of false claims, avoiding access to inappropriate content, among others.

The verification methods that exist today are based only on pop-ups that appear when accessing a website or any other application, asking the user if he is of legal age for the content, or in the form of questioning the date of birth. These checks are unreliable because they accept any input you provide that meets their age requirements, without raising any other questions or age confirmation methods. This type of verification can be seen on the *Jack Daniels*[1] and *Super Bock*[2] web pages, as we can see in the following images.

(a) Jack Daniels

(b) Super Bock

Figure 1: Unreliable verification methods

1 https://www.jackdaniels.com/
2 https://www.superbock.pt/pt/pt/

Stronger methods in terms of age verification, request the availability of data very susceptible to theft, such as a photo of the citizen card, tax identification number, and others, making it a very intrusive way that not every user feels comfortable to provide. This data may come in different forms, yet, it may not reveal racial/ethnic origin, political opinions, religious or philosophical beliefs, biometric data for unambiguous identification of a user, data relating to a person's health, sexual orientation and sex life.

Many of these methods may not comply with the *General Data Protection Regulation* (*GDPR*), as the data subject's consent is required for the use of their information for a particular legal purpose. Or an authorization of a parent or guardian to collect data from minors under 16 years of age. Nonetheless, as soon as the collected data is not necessary for processing needs, they must be generalized through the replacement of direct or indirect identifiers by values that make less sense or even be deleted (Union, 2016).

One of the most notorious cases in recent years regarding this matter is the scandal involving Cambridge Analytica and Facebook. The case dates back to 2018 when the company Facebook admitted the incorrect use of data from around 50 million users, which in turn were improperly obtained by Cambridge Analytica, to create a psychological profile. These same profiles were used for the United States elections, by candidate Donald Trump, in addition to having been used for the 'Brexit' referendum[3][4].

Therefore, the vision involves creating a model, using techniques related to *Artificial Intelligence* (*AI*), which can predict a person's age based on their interactions. As a challenge, it would be interesting to create benchmarks, to evaluate the performance of the created model, as well as different optimization techniques for hyperparameters and frameworks already existing and consolidated in the market.

1.2 OBJECTIVES

This work aims to create a system capable of predicting a user's age, taking into account mental states, such as fatigue, through his interaction with a computer. Therefore, the

3 https://www.businessinsider.com/cambridge-analytica-a-guide-to-the-trump-linked-data-firm-that-harvested-50-million-facebook-profiles-2018-3

4 https://www.nytimes.com/2018/04/04/us/politics/cambridge-analytica-scandal-fallout.html

objective is to create a resource capable of preventing false claims through age verification, using keystroke dynamics and mouse movements.

Taking into account the case study of this project, an analysis will be carried out that aims at the correlation between fatigue and age progression, to contemplate the possible causes for the occurrence of tiredness. As well as studying the impact that this fatigue may have on a user's performance, in addition to an analysis on the task that may present a greater tendency to cause tiredness.

The proposed objectives of this dissertation are:

- Learning about the foundations for this case study;
- Learning the patterns of fatigue occurrence according to different tasks at the time of data capture;
- Analysis of the impact of fatigue on a user's interactions;
- Analysis of the correlation of fatigue levels with age progression;
- Specification of a classification system.

1.3 DOCUMENT STRUCTURE

This document will follow a structure based on seven chapters to present the project in an organized way. The chapter in which we find ourselves intends to make an introduction to the problem that will be addressed, explaining the context, the motivation of the development, the proposed objectives and how it will proceed for its conclusion. The next chapter will present the state of art addressing relevant topics such as Human-Computer Interaction, Affective Computing, Fatigue and Age Detection.

In the third chapter, the architecture of the system is presented along with the proposal for its integration into the market in addition to which tools will be used for the system's design. In the fourth and fifth chapters, the objective is to explain the necessary procedures to prepare each data set and the application of Machine Learning techniques to achieve the main goal proposed in this thesis. Following these chapters, the results obtained are presented in the Results Discussion chapter.

Finally, the seventh and last chapter aims at the conclusions drawn about the project, application areas and the presentation of future works.

1.4 WORK METHODOLOGY

The development of this project will follow an action-research methodology. This type of methodologies, unlike others, follows a line of identification of adversity, formulation and development (Kyriacou, 2007). Throughout the investigation, all data and information relevant to the cause are compiled to devise a solution. Finally, in a final phase, conclusions are formulated that will allow the evaluation of the obtained results.

Thus, the project will proceed as follows:

1. Specification of the functional and non-functional requirements, as well as their characteristics;
2. System modelling and implementation;
3. System deployment;
4. Validation of the implemented system;
5. Result analysis;

Using the following Gantt diagram we can ilustrate, in a more detailed way, the task schedule.

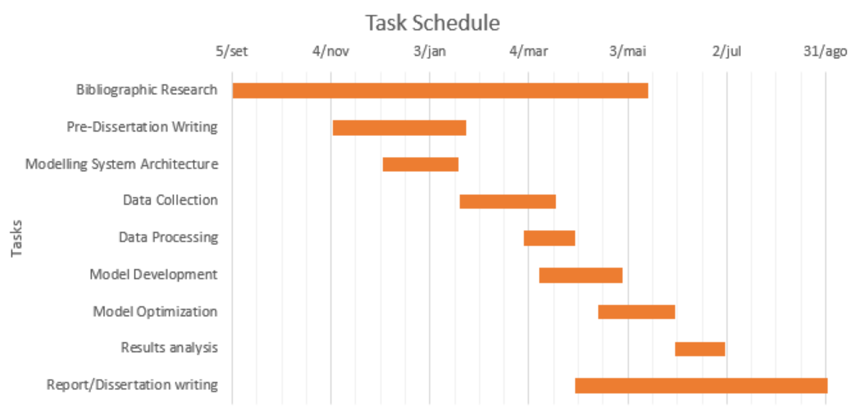

Figure 2: Scheduled tasks

As can be seen the expected duration for this master's thesis is approximately one year. And throughout this period, around 9 tasks will be carried out:

- **Bibliographic Research:** At this stage, all types of research material related to the project will be analyzed;
- **Pre-Dissertation Writing:** At this point, and based on the material collected in the previous phase, a report is made in which themes related to the project will be discussed, an explanation of its areas of application and similar methods for age detection;
- **Modelling System Architecture:** For the modelling of the system architecture, an analysis is made of how the system can be applied, as well as the design to solve this project;
- **Data Collection:** For this phase, data is collected from volunteers;
- **Data Processing:** In this task, the collected data is processed, from correlation analysis, outliers, treatment of missing data, among others;
- **Model Development:** Creation of a machine learning model to detect the age of a user;
- **Model Optimization:** In this phase, an optimization is made of the hyperparameters of the model created in the previous phase;
- **Results Analysis:** From the data collected from the already optimized model, graphs will be generated that will help to conclude this project;
- **Report/Dissertation Writing:** In this final phase, it's the culmination of the entire project developed, where all relevant material will be exposed, as well as all conclusions are drawn;

2

STATE OF ART

2.1 HUMAN COMPUTER INTERACTION

In the early 1980s, *Human-Computer Interaction* (*HCI*) emerged, initially aimed at areas in computer science by combining cognitive sciences and human factors engineering (Nielsen, 2013). These factors are defined as a science that explores the capabilities, behaviour, and how this is incorporated in the design, evaluation, operation and maintenance of products and systems (Boring, 2002). In general, it's defined as a discipline concerned with all aspects of the development of interactive computer systems for the human being and with the study of the phenomena around them (Gulliksen, 2017). It allows the description of physiological limitations, performance and relevant knowledge about a given task, reducing the barrier between a human mental model about an objective and computational support (Wei Shen and Xiaolei Zhou, 2015; Panda, 2017).

This type of interactions, as previously indicated, is highly influenced by the cognitive sciences, which outline cognitive processes in terms of rules and objectives, and methods that will isolate mental phenomena from the context of the outside world (Boring, 2002). Besides the influence originated by cognitive science, the *HCI* is shown to be equally interfered by aspects like physiological state, experience and stress, compromising the usability of a system (Saxena et al., 2013). Given the application of this type of interaction, it's often found in areas such as Speech Recognition, Behavioural Biometrics, Security, Human Performance, etc.

2.1.1 *Speech Recognition*

In the mid-1950s, the first speech recognizer, developed by Bell Telephone Laboratories, appears. This recognizer was able to identify English numbers, however, it was limited to the number of times it accepted to be able to carry out its operations successfully.

Such a system is defined as an interdisciplinary area that combines phonetics and digital signal processing and is also characterized by flexibility, multi-channel coordination and temporality. However, as it is a system that uses speech as its main source of input, it may present limitations with the presence of background noise, whatever its origin, speech rate and frequency (Li et al., 2019).

By nature, the speech recognition is the prefered way to interact with the computer. This choice is corroborated by entities linked to the speech recognition area and marketing area, as it is the most natural way to establish communication (Abu Shariah et al., 2007). Thanks to this recognition capacity, the creation of tools geared towards the areas of education and entertainment have increasingly improved, making it possible to make use of this type of *HCI* for different purposes such as (Panda, 2017):

- Writing through voice;
- Data input;
- Preparation of structured documents;
- Speech-to-text processing;
- Aircraft;

Through these possibilities, it's possible to name existing products with these capabilities such as:

- **Dictation:**[1] Has the ability to recognize and transcribe different languages and dialects. Enables the writing of structured documents and emails using simple commands like "New line", "Delete all text", "Copy to Clipboard", and others. However, it's limited, as it's only possible to use it in Google Chrome.

1 https://dictation.io/

- **Twilio:**[2] Develops communication applications, such as scheduled messages, appointment reminders, among other offers. It's worth noting the possibility of payments by telephone call, in which payment details are captured and the order is subsequently processed.

- **Otter.ai:**[3] It allows the capture of meeting notes, with the possibility of highlighting keywords, audio transcriptions and attaching images. While also presenting integration in several tools for meetings, such as Google Meet, Zoom, Microsoft Teams, among others.

- **Microsoft Word:**[4] it's a product of the Microsoft company, with resources for processing and formatting text. In addition to these basic capabilities, it presents others such as spell checking, version control, and text dictation, with the possibility of using voice commands such as "New Line", etc.

In this way, it allows the chance of assistance to physically disabled people, to document writing, web browsing and overcoming writing adversities.

2.1.2 *Behavioural Biometrics*

With the development of new technologies and in turn new designs, today's devices undergo large amounts of processing in such a small space, yet, they are unable to understand and interpret human behaviour, for all its potential to be explored. With this increase and evolution of technologies, there have been more and more cognitive demands on the part of users and, in turn, a greater workload, which in turn can be reflected on the performance (Roth et al., 2000).

The ability to measure performance effectively is not always easy, due to the influence of factors such as type and time between the start and finish of a task and its requirements for the mental workload. With the existence of all these influences, the possibility of the occurrence of fatigue is increased, for its detection the most common means are in the form of self-report or through the use of electroencephalography (EEG). However, new ways for

2 https://www.twilio.com/docs/voice/tutorials/how-capture-your-first-payment-using-pay
3 https://get.otter.ai/
4 https://www.microsoft.com/en/microsoft-365/word

its detection have been studied, among which behavioural biometrics present themselves as a good alternative because they are manifested naturally in the course of a task.

Behavioural Biometrics is a science whose objective is to identify individuals through their behaviour during the course of daily tasks. Since all people are unique in their way of being, these biometrics present themselves as an efficient form of identification. Similar to other types of biometrics such as iris, fingerprint, palm, which have the same capacity, behavioural biometrics, however, target other aspects such as walking and looking. Just like performance, these biometrics are influenced by factors such as mood, fatigue or even stress (Jain et al., 2004). The possibility of analyzing this type of behavioural patterns can be developed in non-intrusive or non-invasive ways through the observation of a user and his interactions.

Through the use of behavioural biometrics, especially the dynamics of keys and the mouse, it's possible to obtain the type of task, its duration, as well as the detection of different levels of mental fatigue. These detection capabilities come from a set of variability resources extracted from the interaction between a human and a computer (McClernon and Miller, 2011). In this way, the data collected describe in what way an individual behaves while using a computer peripheral, through features like velocity, acceleration and distance travelled by the mouse, as well as typing rhythms, and others. Using this information and others provided by the context, like type and duration of the task, may be used to define the user performance, and to detect and monitor mental fatigue (Ribeiro et al., 2013, 2014).

2.1.3 *Security*

Security is a vital process for any organization or computer system, as they may be victims of illicit access through physical means or access by electronic means. The alternatives to prevent these accesses may be unaffordable for small organizations as they require, for example, the use of specialized hardware in facial recognition, magnetic card reader (Haider et al., 2000).

With this, authentication presents itself as a first protection barrier, defining how a process of verification of digital or physical entities is authentic. These entities can be verified through (Karnan et al., 2011):

- **Passwords:** Very susceptible to be guessed and even the most complex ones can be deciphered using methods like Brute Force;
- **PIN**(Personal Identification Number): They can easily be forgotten;
- **Key/Smart Card:** Susceptible to theft or being lost;
- **Biometric Authentication** (e.g. fingerprint): More convenient and more accurate;

Each of these alternatives have their advantages and disadvantages, implementation price and convenience for a user. However, more and more studies have emerged that safely add the potential of keystroke dynamics.

Keystroke dynamics are manifested through interactions with any interface as long as they have a keyboard, and are based on the user's rhythmic and writing patterns at the time of interaction. According to the National Science Foundation (NSF) and National Institute of Standards and Technology (NIST), United States of America, typing patterns present themselves as unique to the typist (Karnan et al., 2011; Haider et al., 2000).

In the current market, we can already find security systems with this type of implementation, such as:

- **TypingDNA:**[5] It presents itself as a product that allows its implementation in situations where security is a priority, such as login or even entering data susceptible to theft such as a pin or credit card number. This technology is based on recording the writing patterns of a user using its Keystroke Dynamics, and later, when an access attempt is made, a comparison is done between the new pattern and what was already recorded.
- **KeyTrac:**[6] This product is very similar to TypingDNA, following the same writing pattern recognition technology, however, it has the ability to guarantee academic integrity such as preventing identity fraud and cheating in online tests, as well as preventing extortion with hacked credentials.

Finally, there are still cases where content is not suitable for all types of people or ages. Content that varies from images, videos or information, from pornographic sites, chat rooms without moderation, which encourage acts of vandalism, crime or even terrorism. For these

5 https://www.typingdna.com/
6 https://www.keytrac.net/en/

cases, access control methods are easily circumvented or even non-existent, showing a great need for safe and economical methods for the end consumer that do not require the use of external tools such as those previously indicated.

2.2 AFFECTIVE COMPUTING

The emotions of a human being present themselves as a complex psychological state that can be expressed in different ways, such as speech, writing, facial expressions and body language, thus marking a strong presence in everyone's daily life (Zucco et al., 2017; Yang et al., 2019). By being a psychological state, emotions can influence a variety of factors, such as cognitive perception, decision making, learning or even communication (Gowda et al., 2017).

In 1995, in the proposal of Professor Rosalind Picard, *Affective Computing* (*AC*) emerged. This is designated as the study and development of systems and devices that can recognize, interpret, process and simulate human affections (Zucco et al., 2017). Thanks to emotions in both cognition and human perception, affective computers will be able to improve human assistance capacities, as well as decision-making capacities (Picard, 1997), enabling its application in areas such as E-learning, Call Centers, Medicine, Legal Assistance, etc (Gowda et al., 2017).

An investigation in *AC* mainly includes three parts (Yang et al., 2019):

- **Emotion Model:** The hidden Markov model presents itself as the best, as it tells us that people will have an emotional state at any time and that the emotional state at different times can be converted with a certain probability;

- **Sentiment Classification:** Taking into account different proposals for categories of emotions, several investigators consider that the most representative are those proposed by Ekman (Happiness, Surprise, Boredom, Sadness, Fear, Anger) and Plutchik (Fear, Anger, Sadness, Happiness, Disgust, Trust, Expectation, Surprise), when combining them we arrive at the following categories: Happiness, Surprise, Neutrality, Anger, Fatigue and Confusion;

- **Emotional Recognition:** According to Professor Picard, the emotional calculation process can be summarized in the following graphic:

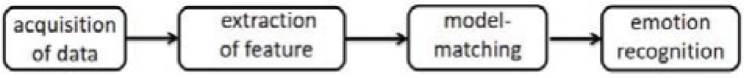

Figure 3: Emotional Recognition Model, proposed by Professor Picard

Source: Yang et al. (2019)

Given its enormous applicability, *AC* can also be used together with Sentimental Analysis, which refers to the processing of natural language and textual mining methodologies, to identify an opinion or feeling (Zucco et al., 2017). This type of approach is based on 3 categories (Cambria, 2016):

- **Knowledge-based techniques**, this type of technique has several limitations. For instance, it has a weak ability to recognize affection when the linguistic rules are evolved and doesn't allow dealing with different concepts;
  Example:
  - **WordNet-Affect:**[7] Developed by WordNet Domains, which presents a multi-language extension of Word-Net, through the selection of synsets (groups of synonyms) that can represent affective concepts. WordNet-Affect aims to add a hierarchy of affective tags, independent of the domain, in which the synsets that represent the affective concepts are noted;
  - **SentiWordNet:**[8] Is a lexical resource for the classification of feelings. This resource applies to each synset of WordNet, three possibilities of feeling, "positive", "negative" and "objective";
  - **SenticNet:**[9] Offers a set of semantics and polarities associated with natural language concepts, thus allowing the detection of polarity and sentimental recognition present through connotative and denotative information in a text.

7 https://wndomains.fbk.eu/wnaffect.html
8 https://github.com/aesuli/SentiWordNet
9 https://sentic.net/

- **Statistical methods**, although, they are semantically weak and require huge amounts of data to achieve a good recognition capacity;
  Example: SVM, Deep Learning;

- **Hybrid methods**, use knowledge-based techniques and statistical methods to detect emotions and detect polarity from the text;
  Example:
  - **iFeel**[10] It's a mobile platform, which connects a patient with a specialized psychologist for the area of need. It provides total confidentiality, personalized treatment plan and daily communication.

To understand more clearly about the affective capacity of a computer, it's possible to perform a test, called the Turing test to which an assessment is made if in a conversation between humans and computers, by textual means, the human cannot detect whether the responses you get are generated by a machine or another human. This type of test was conducted several times by Professor Clifford Nass, who concluded that human-computer interactions are inherently natural and social, given that communication is something natural between people and in turn is more natural with a computer that detects and express affections (Picard, 1997).

Through dialogues between humans and computers, it's possible to capture signs of emotions such as anxiety, still this capture and recognition must be constant due to the path that a dialogue can take (Han and Li, 2014). These interactions can arise through biometric attributes, which may be manifested as physiological (face, palm, fingerprint, iris) or behavioural (voice, writing, keystroke dynamics, mouse movements), also showing wide applicability, for example in security systems (Kołakowska, 2013).

2.3 FATIGUE

Occasionally, after a day, a person may feel a certain physical and mental exhaustion, this phenomenon is called fatigue. Usually, this sensation can be used to describe different states, from nausea to lack of concentration (van der Linden and Eling, 2006). Therefore,

10 https://ifeelonline.com/en/

the concept of fatigue manifests itself as something very broad and subjective. Because it can present itself as a combination of symptoms, including poor performance (e.g. loss of attention, very slow reaction times, poor performance on usual tasks that reflect good performance, etc.) and other symptoms, such as drowsiness and tiredness (Perelli, 1980; Williamson et al., 2005).

With the initial objective of its formalization, fatigue can be defined as a symptom to which physical and cognitive functions are limited. On the other hand, Frederic Bartlett presents a more ambiguous definition that encompasses all the states mentioned above (Bartlett, 1953):

> *"Fatigue is a term used to cover all those determinable changes in the expression of an activity which can be traced to the continuing exercise of that activity under its normal operational conditions, and which can be shown to lead, either immediately or after delay, to deterioration in the expression of that activity, or, more simply, to results within the activity that are not wanted."*

So, the fatigue present in everyday life can be considered as a physical or mental insufficiency, which in turn can result in discomfort or reduced task performance. It can manifest itself in the form of tiredness, psychological disorders, lack of energy and concentration (Faber et al., 2012; Miller, 2008).

Mental fatigue appears as a state that involves a set of effects that in turn influence cognitive, emotional and motivational skills, which in turn results in a state of discomfort and poor performance (van der Linden et al., 2003). These limitations may lead to a lesser willingness to perform or complete a task. But, despite several studies already carried out on mental fatigue, monitoring it's still difficult to achieve (Boksem et al., 2006; Muscio, 1921; J. and Hockey, 1997). We can characterize it as a lack of mental energy, and people who are affected by this type of fatigue feel that they have less energy than normal. It can lead to mental exhaustion if the workload is excessive (Gartner and Murphy, 1977).

One of the main contributors to the occurrence of fatigue is impaired sleep, which causes a reduction in brain energy and hormone levels (Åkerstedt et al., 2018), its origins can refer to prolonged mental or physical work, long periods of stress or anxiety and even performance of boring or repetitive tasks. This state has an influence on activities that a person can perform, and can manifest itself in the form of (Caldwell et al., 2019):

- Reduced decision-making ability;
- Reduced communication skills;
- Reduced productivity or performance;
- Reduced ability to handle stress on the job;
- Increased errors in judgement;
- Increased rates of adverse incidents;

A person's profile provides valuable information about his condition, through physical and psychological characteristics and states, as well as information about his daily life. This information includes:

- **Age**: It indicates an individual's mental age and it's important to understand the expected cognitive abilities, as they are likely to degrade with advancing age;
- **Gender**: Mental states differ between men and woman;
- **Consumption of alcohol or drugs**: The use of certain substances may lead to a change in physical and cognitive abilities
- **Occupation**: Some jobs are more tiring than others;
- **Mood**: The person's mood can influence their mental state, through motivation the person can overcome the effects of fatigue;
- **Stress**: It can be defined as demands or concerns placed on a person by external stimuli;
- **Sleepiness**: Symptom strongly related to mental fatigue and can be characterized as a form of an alert, on the part of the brain, to indicate the scarcity of resources.

2.4 HUMAN PERFORMANCE

Human performance is a factor considered critical in any activity or work. This is correlated with the execution of tasks by an individual, and it can be defined as an interpretation competence to the associated stimuli. In this way, efficiency can be associated with the answers given to these interpretations. Regarding its measurement, all of them must be taken into account, which in turn is useful for detecting fatigue states or even the possibility of failures.

Excessive workloads, whether physical or mental, are presented as one of the main sources for the decrease in performance, in addition to cost-benefit assessments that may lead to a deterioration in performance if the execution of a task exceeds the benefits of its completion (de Jong et al., 2018).

Cognitive performance is greatly influenced by the way the domain characteristics are processed, however, with the presence of mental fatigue, this processing is adulterated and can lead to slower reaction times, also affecting speed and accuracy, which are characteristic measures that make it possible to measure performance (de Jong et al., 2020, 2018).

With advances in technology, and in turn, the emergence of autonomous systems, human performance has been influenced due to the lack of need for a human hand in decision-making and its information collection capabilities are not stimulated. Thus, the effectiveness of a Human-Computer system depends on the ability to monitor the performance of an autonomous system (Roth et al., 2000).

For example, on February 19, 1985, there was an accident with a Boeing 747 plane from China Airlines, on a flight originating in Taipei, Taiwan and bound for Los Angeles, California, United States of America. The airplane in question suffered a loss of power from one of its engines, with this event the autopilot compensated and maintained the flight level without alerting the pilots. Upon reaching its limit, the plane starts to roll until it enters a vertical dive (National Transportation Safety Board Bureau of Accident Investigation Washington, 1986).

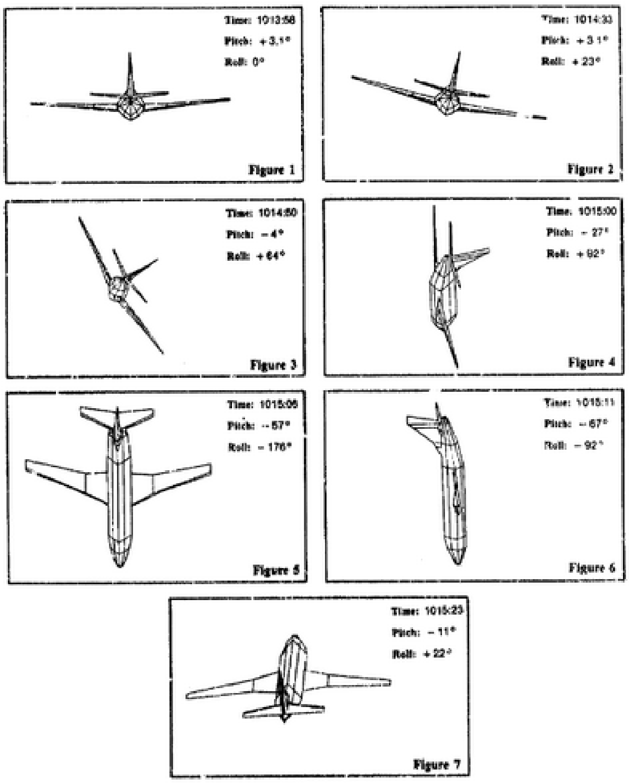

Figure 4: Representation of the plane's trajectory

Source: National Transportation Safety Board Bureau of Accident Investigation Washington (1986)

Donald Norman, former professor at the University of California, considers that the failure of automated systems to provide feedback comes from the designers of these systems. For any situation or system, the possibility of unforeseen circumstances must be taken into account, as in the same way that a human is capable of dealing with these situations, so a system should be, through information and control mechanisms (Norman, 1990).

2.5 AGE DETECTION

The age of a person is an attribute that represents their lifetime, providing with their progression, access to the most diverse experiences. However, it carries responsibilities, and it's often falsely alleged for malicious practices or improper access. Therefore, the estimation of a person's age has been increasingly investigated, enabling its application in several areas such as forensic medicine, helping criminal investigations, security, limiting access and preventing the purchase of alcohol or tobacco (Erbilek et al., 2014). The sources that make this estimation possible come from the most varied ways, such as the use of forms, biometrics, writing, interactions with devices (e.g. Tablets, Mobile Phones, Computers), facial recognition, and even magnetic resonances.

Online chats are presented as a way to connect people from the most remote parts of the planet but have very little security regarding content. A study by the Crimes Against Children Research Center, in 2005, indicated that young people receive less and less unsolicited sexual content, one of the reasons could be education or even discouragement by the law. To catch predators, the authorities pretend to be minors, however, the results are scarce due to the existing number of employees and volunteers, arising the need for an autonomous system for age detection, using text messages. Systems of this type are based on the use of natural language processing (NLP) techniques, such as n-grams (e.g. uni, bi, tri, etc.) as they show the style as lexical and contextual information. As well as the analysis in the punctuation and slang as they help in the detection of age groups (Tam and Martell, 2009).

Facial recognition presents itself as a technology already well established in our society and its properties and aspects enable the creation of technological solutions for different purposes, such as people counting, age/gender detection, mask use detection, face tracking, etc (Lobo and Kwon, 1998). Several companies such as Microsoft[11], Shikino[12] and SightCorp[13], have already presented their solutions in the current market.

- **Microsoft Azure Face Recognition:** Its implementation allows the recognition of people based on a private repository, emotional perception and the ability to detect multiple face;
- **Shikino's Age and Gender Detection Camera:** Enables the detection of a person's age and gender with a stand-alone camera;

11 https://www.microsoft.com/pt-pt/
12 https://www.shikino.co.jp/eng/
13 https://sightcorp.com/

Figure 5: Stand-Alone Camera

Source: Shikino's website

- **Sightcorp's Face Detection** Enables the user to choose between a *Software Development Kit* (*SDK*) or a Toolkit, allowing the system adaptation as the user intends;

All of the facial recognition technology have the same basis. These technologies use facial landmarks, as they allow the creation of properties that describe in detail characteristics of facial attributes such as the nose, eyes, eyebrows and others (Farley et al., 2019).

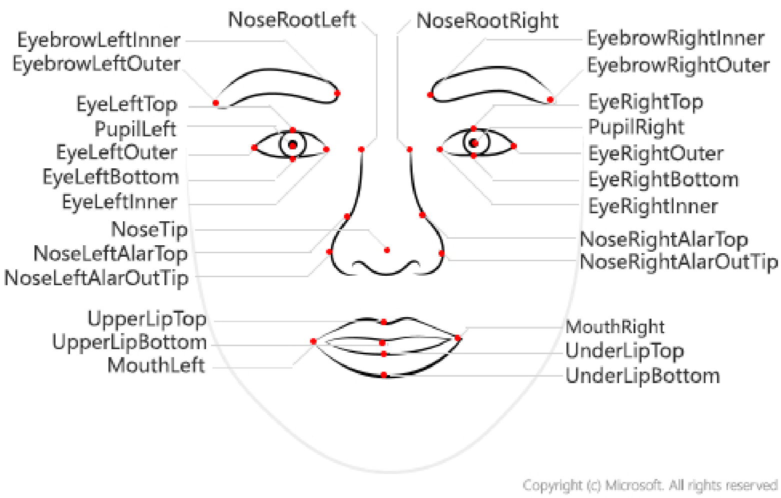

Figure 6: Facial Landmarks identified by Microsoft

Source: Farley et al. (2019)

Some systems calculate and analyze the wrinkles present on the face of a user to help in the detection, or the use of point-and-shoot cameras to eliminate the possibility of error due to adjustments by those who took the photo. For the age or gender detection, a frame is made within the limits of each age and the output comes from the highest probability of the frame (Lobo and Kwon, 1998). Due to its nature, a system of this type is very susceptible to several aspects such as clarity, lighting, misinterpretation by the algorithm, etc., leading to a wrong estimation, as we can see in Figure 7 (Clark Howard, 2015).

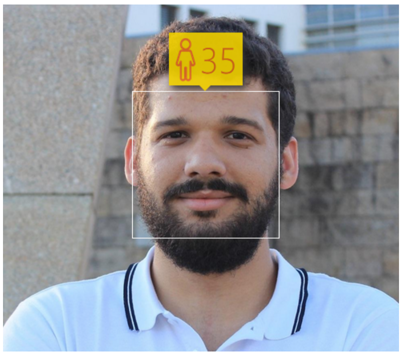

Figure 7: Wrong prediction of the author's age

Source: Prediction made with Microsofts How Old[14]

Human-human interactions, as a rule, are established by voice, gestures or by a textual way, taking into account that with the progression of age, human organs such as the glottis or vocal tract are subject to change. The possibility of identifying age groups based on speech signal arises and this signal can be broken down into segments, representing vocal and non-vocal regions. These segments can then be classified using *Zero-Crossing Rate* (*ZCR*) and *Short Term Energy* (*STE*) methods to distinguish between sound and non-sound parts. A speech is produced by the excitation of the vocal tract by the airflow in the glottis, generally showing a low zero-crossing count, while non-sonic parts are produced by the tract constriction, enough to cause turbulent airflow, resulting in noise with the high zero-crossing count. That is, a high *ZCR* indicates segments without a voice, and in turn, a low value indicates voice segments. In 2014, Dipen Nath and Sanjib Kalita proposed the use of these methods to identify age groups, in which they concluded that older age groups, 50 years old, are easily distinguishable from others, people aged 30 and 10 years. On the other hand, age groups of 30 and 10 years, present more similar results (Nath and Kalita, 2014).

New forms of detection have appeared, and of the most varied forms, such as the use of *Magnetic Resonances Imaging* (*MRI*). This technology allows the obtaining of better images than other techniques, such as computed tomography, due to its high contrast in soft tissue.

14 https://www.how-old.net/

Knowing the age of a brain allows the prevention of premature deaths and the prediction of diseases like Alzheimer's. To detect the age of a brain, the authors, suggested the application of segmentation in the images from an *MRI*, to facilitate its analysis, and later the use of machine learning models, previously trained, such as a convolutional neural network and a deep neural network. With these types of algorithms and data processing, they accomplished an accuracy of 79% (Siar and Teshnehlab, 2019).

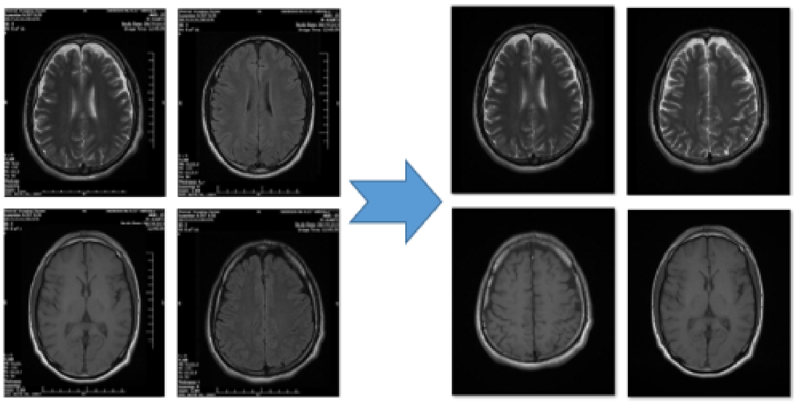

Figure 8: Example processing of an MRI of the brain

Source: Siar and Teshnehlab (2019)

There are other ways to validate operations, these can come in the form of forms or other simple operations like a checkbox. One of the most well-known and well-established ways in the financial market is Know Your Customer, which is based on the addition of confidential information or questions by a user, to make authentication more robust; based on the information provided, they are assigned a ranking, this information is then subsequently required based on the risk/likelihood of identity theft. That is, if there is a greater risk/probability, then higher ranking information will be required. This type of operations can be used, for example, after a successful login or even moments before completing a higher risk operation, such as bank transfers (Mondal et al., 2016).

With the increased usage of mobile devices, the use of technologies that make use of the touch-screen emerges as one of the most used forms among *HCI*, and the gesture that characterizes them, is more intuitive. Using these devices allows the collection of information that enables the characterization of the user or his status. Through this collection, it's possible to obtain mechanisms, variations in intensity, duration, touch area, among others. These interactions make it possible to estimate age (e.g. an older user has longer and more intense touches). With these possibilities in mind, the authors of *"The influence of age and gender in the interaction with touch screens"* proposed to use this type of interactions, using machine learning algorithms, like *Gradient Boosting Machine* (*GBM*) to make the prediction. Thanks to this estimation, the interaction patterns may emerge as relevant indicators for health monitoring and the appearance of cognitive or physical problems (Rocha et al., 2019).

2.6 SUMMARY

With the end of the exposition of the topics of great relevance to the theme of this dissertation, it was noticeable the importance of the areas concerned to achieve the intended end. *HCI*'s present themselves as the discipline aimed at an entire interactive system and all phenomena, which may reside in it, allowing a detailed description of possible physiological limitations on the part of humans. This type of interactions are then influenced by harmful physical and emotional states. Since the state of a user is always susceptible to the phenomena that surround him, such as the task he performs, the work environment or even other factors such as anxiety, stress or even fatigue, his interactions are also influenced. Thus, there is a need for systems capable of detecting, interpreting and processing human emotions. *HCI* are inherently natural and social, given that communications are natural, social and convey emotional states. This area is called *AC*, and it's used in areas related to psychology, medicine, among others.

From these interactions, it's possible to identify adverse states, which can be described by a mouse's writing patterns, speed and acceleration, among other dynamics made possible by mouse and keyboards. These dynamics affect the area of Behavioural Biometrics, which manifests itself as the science whose objective is the identification of individuals through their interactions in daily tasks, and since each person is unique, these biometrics present themselves as an alternative for use in identification.

The term fatigue is a very broad and subjective term, as it can represent a combination of symptoms, including poor performance, nausea, dizziness, tiredness. Given the type of interactions being addressed, physical fatigue is negligible given that hand contractions only occur between 6 to 20% of cases and thus, are considered to be low. So, mental fatigue can be defined as a state that involves a set of effects that influence cognitive, emotional and motivational functions, which in turn results in discomfort and reduced performance. Fatigue can arise through impaired sleep, long periods of work, stress, anxiety or even boring tasks.

A person's performance is related to the execution of tasks and can be associated with the answers given to interpretations of problems. As previously indicated, this is influenced by emotional states, fatigue, also, it can be impaired if, in a cost-benefit assessment, the cost of performing a task exceeds its benefits. However, these limitations can be overcome through breaks, motivation (e.g. monetary awards based on performance), or even a cup of coffee.

With all these factors in sync, the possibility for successful detection arises, and there is already software for age detection but they come from different sources. These sources arise in the most varied ways, from the face (in which reference points are used and fit the values obtained in an age group), voice (where they calculate methods that identify vocal and non-vocal regions and from their intensity fit a respective age group). Or even through brain *MRI* (where segmentation of an *MRI* of a brain is made and later convolutional and deep neural networks are applied to detect the age of the brain). The use of these type of technologies can also be used in areas of security (e.g. online chats, age validations, validation access, etc.), prediction of cognitive or physical disabilities, among others.

3

MAIN COMPONENTS ARCHITECTURE

This chapter will address the architectures proposed for the work in question, as well as the technologies for their design. Regarding architectures, we can identify two types, monolithic and based on micro-services. Monolithic architectures can be considered as the whole service in a single system, presenting themselves with a good performance, simple deployment and low complexity in their implementation and integration. However, due to everything being in the same service, a failure in the system can mean that it stops and if the functionality is overloaded, it's not possible to replicate it in isolation. An update to the system implies a new deployment, difficult to maintain and not able to scale.

On the other hand, architectures based on micro-services, are better organized (each service has its specific work) and allows an independent deployment for each service. They are technologically independent and make it possible to isolate. In addition to its development being faster, it has greater resilience and is easy to maintain. However, this type of architecture presents latency in the communication between its micro-services, its development is more expensive, the testing of the entire system is complex and it's difficult to create a layer capable of encapsulating common transversal concerns of the various modules.

An *Application Programming Interface* (*API*) is a set of definitions and protocols for creating and integrating applications. Emphasizing its flexibility, design and use, enabling products and services to communicate with others without prior knowledge of their implementation. They are also presented as a simplified way to connect an infrastructure with the development of cloud-based applications.[1]

1 https://www.redhat.com/en/topics/api/what-are-application-programming-interfaces

3.1 MULTI PLATFORM APPROACH

The proposed system, aims at its integration in multi-platforms, such as websites, through libraries, or applications through a *SDK*. This tool provides intelligence, enabling the prevention of access to inappropriate content, and avoiding intrusive methods that allow confirmation of the alleged age.

The following figure describes the intended system, in a very succinct way.

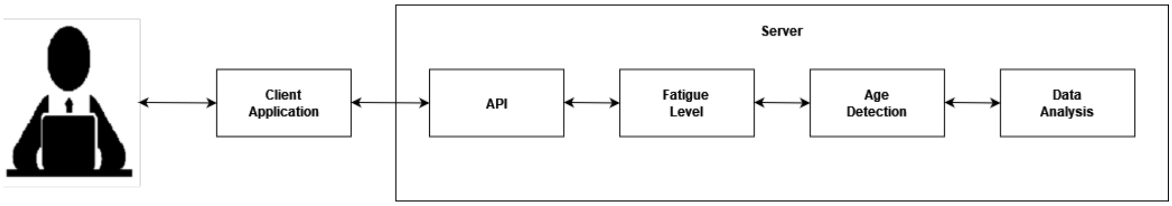

Figure 9: Main system architecture

In Figure 9, it's described the steps to achieve the age prediction and the analysis of the claim are described.

- **Client Application**: The system starts with the interaction between a user and a device (e.g. computer or mobile), in which events related to these interactions are captured, using sensors that already existed in these devices;
- **API**: After the events captured in the client application, this data is used as an input for the *API*, where it will be processed into biometric features;
- **Fatigue Level**: With the processed biometrics, the Fatigue Levels are predicted through machine learning algorithms;
- **Age Detection**: Concluded the Fatigue Level prediction, now it's possible to obtain the user's age using the biometrics and the levels as an input to *ML* algorithms;
- **Data Analysis**: Finally, with the predicted age it's performed an analysis, which will define if the claim made is true, and so granting or not access to the content requested.

This classification is achieved thanks to data mining techniques, as we be seen below. These aim to remove as much of the obtained data as well as to understand the relationships and characteristics that reside in them. Enabling the application of appropriate algorithms to classify the target, where they themselves, in a previous phase, were optimized reducing the error capacity in the predictions that are requested.

As mentioned, to get the most of the available data, the methodology *CRoss Industry Standard Process for Data Mining* (*CRISP-DM*) was adopted. In this methodology, it's followed a few steps, such as:

1. **Business Understanding**: Initial study of the problem in question and the areas related to it;
2. **Data Collection**: Collect data and evaluate their quality;
3. **Data Preparation**: Selection of features and data cleaning;
4. **Modeling**: Model creation to solve the problem in question;
5. **Evaluation**: Compare the predicted results with ones already available;
6. **Deployment**: Model deployment for production purposes.

In the following figure, the pipeline followed to achieve age prediction is described in more detail.

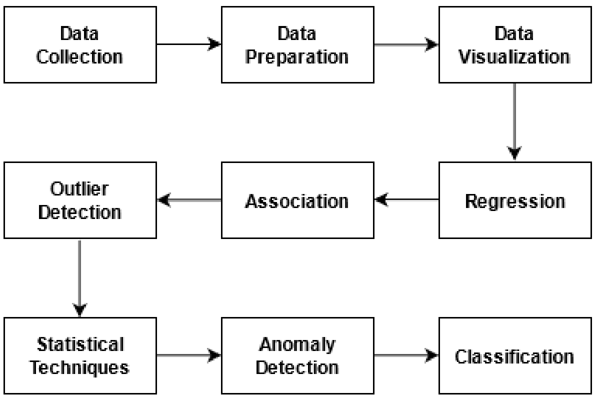

Figure 10: Age group classification

Looking at Figure 10, it's possible to see that before achieving the age group classification, the data will be subjected to 9 phases. These phases are described as data mining techniques, as mentioned previously, and are intended to achieve a better understanding of the dataset and possibly filter features that may harm the age group prediction [2].

2 https://www.talend.com/resources/data-mining-techniques/

- **Data Collection**: Raw data capture through mouse and keyboard events, as well as the task being performed;
- **Data Preparation**: The process where raw data is cleaned and formatted to convenient and more insightful data. These include different methods, such as data modelling, data migration, transformation, and so on;
- **Data Visualization**: Creation of graphical forms for the characterization of the data, and its relationships with other featues, allowing a deeper knowledge about possible patterns and, in turn, about the dataset;
- **Outlier Detection**: Detection of possible outliers through boxplots, and in turn, application of the "Winsorizing" technique for their treatment;
- **Association**: Analysis of the correlation between different features through a Heatmap. Filtering of features that have the highest correlation with each other;
- **Regression**: Use of regressive *ML* models for missing data prediction;
- **Statistical Techniques**: Filtration of less significant features for the final objective through the P-Value test;
- **Anomaly Detection**: Detection of anomalous, global or age group records, using the Isolation Forest model;
- **Classification**: Prediction of a user's age group through *ML* models, provided by the H2o.ai framework.

3.2 API AS A SERVICE

Software as a Service (SaaS) is a delivery model, commonly known as "software on-demand". This type of model can drastically reduce the total cost of ownership for adopting software, has eliminated the burden of problems with local hardware, as well as possible scalability challenges. Thanks to this, it allowed companies to focus on the business itself, reducing the burden of IT operations.

With this growth, around 73% of organizations have at least one application or a part of their computing infrastructure already in the cloud. Emerging a need for product integration in applications, and hence it's necessary to use API, which tend to offer upfront billing. So it's noticeable that APIs are a distinct subset of SaaS, and in turn, more extensible and customer-friendly (Levine, 2019).

API as a Service, thusly, introduces itself as a product stage that permits the client to cooperate with outsider APIs, os just deal with their own. As a stage, it presents two principal components (Fanchi, 2019):

- Capacity to fabricate, test and actualize with your administrations;
- Capability to connect with third parties;

An incredible illustration of the utilization of this kind of innovation is the organization's Amazon Web Administrations (AWS), Stripe and Twilio, which present incomes in the order of millions or even billions (Levine, 2019).

3.3 MACHINE LEARNING

Before approaching how the model will be created, first it's necessary to understand the techniques, and what is meant by machine learning. *Machine Learning* (*ML*) is a subarea of *Artificial Intelligence* (*AI*) and the goal is to build applications that learn from data to improve its accuracy over time without the need to be programmed (Education, 2020).

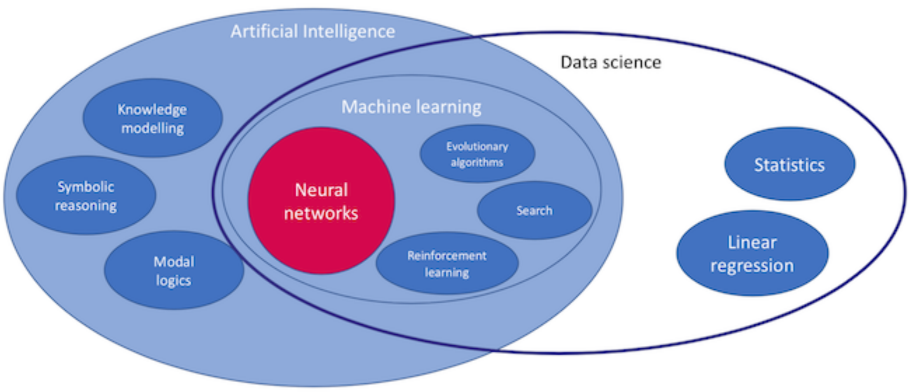

Figure 11: The relation between *ML*, *AI* and Data Science

Source: ICT Institute[3]

ML allows the use of three types of techniques: supervised, unsupervised and reinforcement learning.

- **Supervised Learning**: In this type of learning, the final result is known, allowing the establishment of a relation between the pretended data and values produced by the system. Can be divided into two categories:
  - **Classification**: The results are discrete (e.g. "Blue", "White", ...);
  - **Regression**: The results are continuous (e.g. variation of temperature, atmosphere pressure, ...);
- **Unsupervised Learning**: The results are unknown, only the data is known. The objective is to establish a data distribution or modulate a structure. It can also be divided into two categories:
  - **Segmentation**: When the intention is to organize the data into coherent groups;
  - **Association**: When the objective is to know the rules that establish a relation between them;
- **Reinforcement Learning**: Doesn't have information about the results, however it has the capabilities of knowing if the obtained results are good or bad.

3 https://ictinstitute.nl/ai-machine-learning-and-neural-networks-explained/

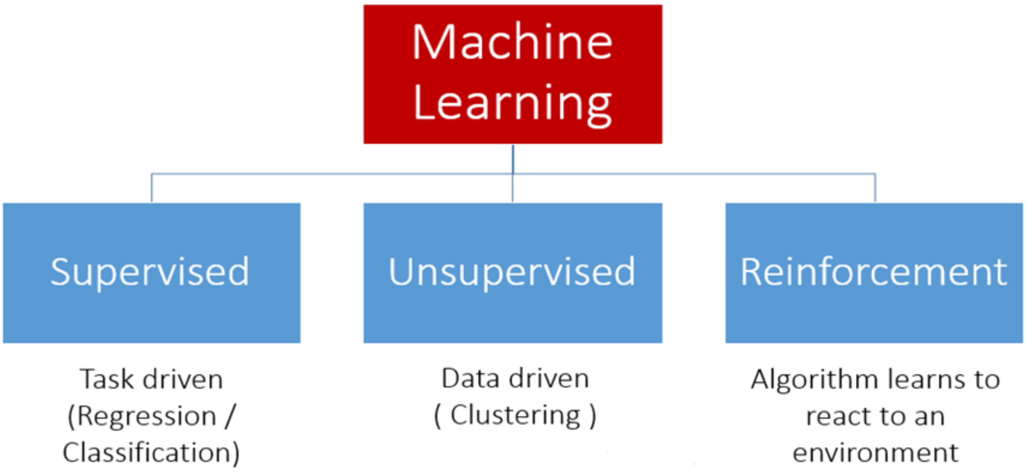

Figure 12: Types of techniques available

Source: Medium[4]

With this in mind, the goal is to create the model with the Anaconda[5] platform and using the Python[6] programming language. The Anaconda platform is a popular tool in the data science and machine learning community, it encompasses a wide variety of tools, such as virtual environments, making it possible to use different versions of libraries in isolation. The Python language is a programming language with a simple syntax, allowing its integration easily in other system components. Both this platform and programming language, are compatible with frameworks such as Tensorflow[7], Keras[8] and H2O.ai[9]. These libraries are well-optimized and stable, making them a good option for model creation.

Regarding the creation and analysis of graphics that will collaborate in the decision-making process, the Matplotlib[10] and the Seaborn[11] packages will be used.

4 https://medium.com/deep-math-machine-learning-ai/different-types-of-machine-learning-and-their-types-34760b9128a2
5 https://www.anaconda.com/
6 https://www.python.org/
7 https://www.tensorflow.org/
8 https://keras.io/
9 https://www.h2o.ai/
10 https://matplotlib.org/
11 https://seaborn.pydata.org/

3.4 MODEL ENDPOINT

After designing the model, the next step to be developed is an endpoint to allow its integration across multiple platforms. For the case study of this dissertation, RESTFul Web Service developed using the Spring Boot framework, the Java programming language, the Maven tool and MOJOs of the developed models.

Since this service is destined for the prediction of a user AgeGroup, the data that serves as input will need pre-processing. The following picture display all the phases and technologies used to achieve the desired objective.

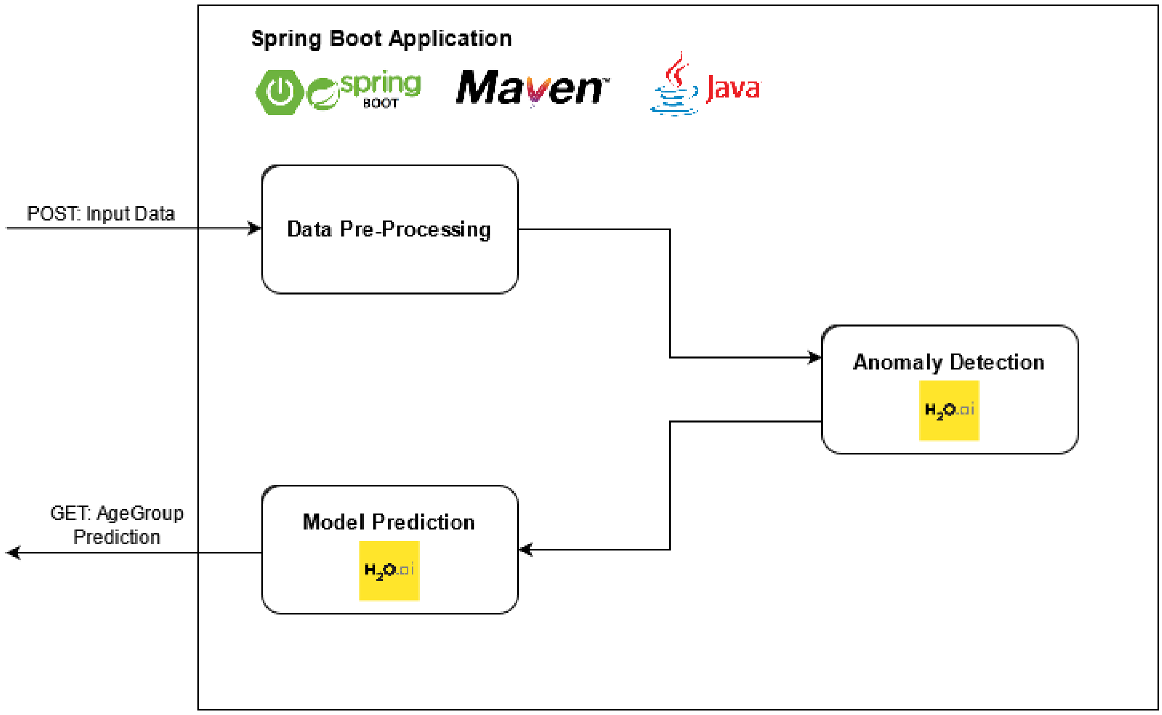

Figure 13: Endpoint representation

The diagram above intends to show the entire processing line to which the data will be subjected. To arrive at a prediction about the age group that a user is in, the data received as input will go through 3 phases:

1. **Data Pre-Processing:** This first phase's goal is to extract the registration time from the timestamp, and in turn, determine what part of the day it is;

2. **Anomaly Detection:** After pre-processing, the data is used to predict whether it is an anomalous record as a whole or for some age group in question. With these predictions made, two new features are added that will be used in turn for the final prediction;

3. **Model Prediction:** Similarly to what happened in the 2nd phase, the new data is used to predict a user's age group, where finally this prediction is sent back to a user's device through an HTTP GET method;

The use of a Web Service, shown above, has several advantages, such as:[12][13]

- The execution of each order is done independently;
- They feature great compatibility with HTTP protocols;
- Facilitates data exchange between different applications and platforms;

In addition to the advantages presented above, it's possible to notice that since the existing relationship is a "Client-Server" relationship, it allows the existence of independent evolution of each of the intervening parties.[14] At the same time, it makes it possible to simplify the entire processing line to achieve the prediction of the user's age group remotely, i.e. the forecast of the intended group does not induce any additional processing load to the already existing in the user's device. However, it also presents certain counterparts, such as the usage of HTTP protocols which enables the existence of failures in the delivery and response of data.[13] Also, given the purpose in question, several variables were used, to satisfy the entire process, enabling the achievement of the intended objective, yet it may cost too much time to conclude the whole process.

12 https://www.tutorialspoint.com/restful/restful_statelessness.htm
13 https://blog.idwall.co/qual-a-diferenca-entre-api-e-web-service/
14 https://www.mulesoft.com/resources/api/restful-api

4

AGE DETECTION THROUGH MOUSE AND KEYBOARD EVENTS

The following chapter looks at all the processes needed to predict a user's age group using *HCI*. These interactions are characterized by biometric data collected from mouse and keyboard events. In addition to these biometric data, some variables describe fatigue levels, from 1 to 7, where 1 corresponds to a state without any existence of tiredness and 7 to a state of total exhaustion (Samn and Perelli, 1982). However, there are some records with fatigue level being 0, meaning that no calculation was made to estimate the fatigue status of this person.

The data to be analyzed and subsequently applied by *ML* algorithms was made available by Anybrain, S.A.[1]

4.1 DATA SET CHARACTERIZATION

The data set provided presents about one million records, to be more precise 1051505 records, which represent the average of 5 minutes of interaction, equivalent to about ten years of data collected, and 76 variables. In these, it's possible to highlight those that represent fatigue levels, *FatigueLevel*; the task to be performed, *Task*; *birthday*; record timestamp, *TimeStamp*; user ID, *User*; user gender, *gender*; and the rest relating to the events referred to above.

As mentioned before, two types of events are recorded, generated from a mouse and a keyboard. These aim to describe various biometrics, such as the speed of the mouse's movement, its acceleration, or even the time that takes place between pressing two consecutive

1 https://anybrain.gg/

keys. Some of these biometrics are presented below, by event, and a description of their functionalities.

- Mouse Events
  - MMAVar (Mean Mouse Accelaration Variance): Mouse acceleration for a certain period is calculated from the following formula;

$$\bar{a} = \frac{\Delta v}{\Delta t} \tag{1}$$

  - AEDVar (Average Excess Distance Variance): Average of the distance excess travelled by the pointer, considering two consecutive clicks. It's calculated as the difference between the travelled distance and the distance in a straight line between the two clicks;
  - CDMean (Mean Click Duration): Describe how long a left-click took to complete in milliseconds. It's obtained through the difference between the timestamp of a MOUSE_LEFT_DOWN and the following MOUSE_LEFT_UP event;
- Keyboard Events
  - KIVar (Key Interval Variance): Time between every two keys pressed, the calculation follows the same line of thought as 'CDMean';
  - KDTMean (Mean Key Down Time): How long was a given key pressed in milliseconds;
  - WVMean (Mean Writing Velocity): Average of the number of keys pressed per minute;

Since the data set in question has records from several years, new variables were added throughout the time, thus leading to certain previous records having missing values in these same new features. Therefore, 32 features present more than 60% of records as missing values and about 6773 records with more than 50% of missing values.

4.2 FEATURE ENGINEERING

With the existence of a variable relating to the timestamp of each record, it is also possible to calculate the age of each player as well as the time of registration, and the part of the day it is on. The last two features were created to introduce a relation between them with FatigueLevel feature, influencing the collected interactions.

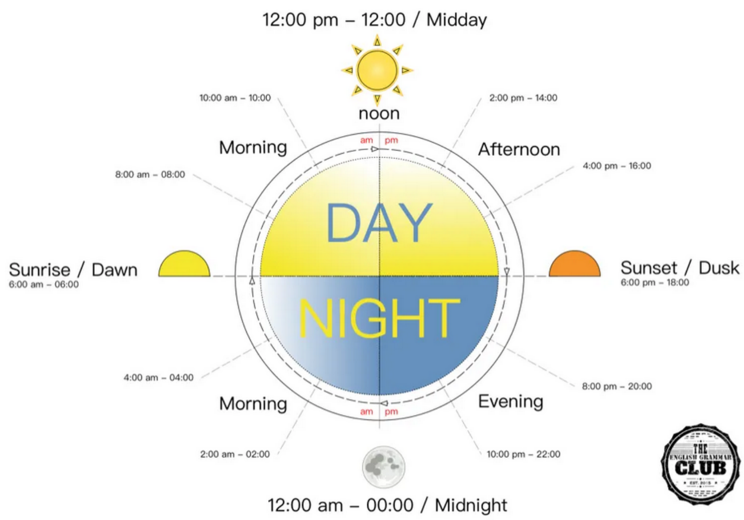

Figure 14: Representation of the parts of day

Source: The English Grammar Club[2]

The parts of the day, introduced to the data set, are based on the previous image. Although with a few modifications, "Morning" occurs between 6 am to midday, followed with "Afternoon" until 6 pm, "Evening" begins at 6 pm and ends at 10 pm, the remaining hours of the day, not mentioned, are considered to be part of "Night".

As mentioned before, the prediction made in this section will be of the classification type, more specifically, the forecasting of the age group to which the user belongs. The groups created were based on an article made available by StatCan[3], a government agency whose objective is to provide statistical information about the Canadian population (e.g. population, resources, economy, society and culture). However, the suggested groups weren't exactly as described in the article, more specifically, the group refering to "Youth" that contains ages

2 https://grammar.tips/grammar-tips/parts-of-the-day/
3 https://www.statcan.gc.ca/eng/concepts/definitions/age2

between 15 and 24 years old which were divided into two groups, one for the ages between 15 and 17 years old, and a second one for 18 years old up to 24. This enabled the separation between minors and adults, allowing, in turn, the implementation of this type of forecast in systems that require this type of knowledge.

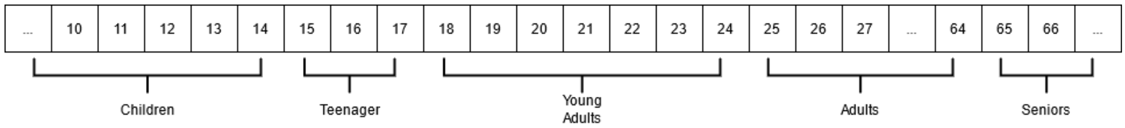

Figure 15: Age provision by their groups

In Figure 15, the implemented age groups are displayed and as can be seen, 5 age groups are presented. However, only 3 are present in the data set in question, being 'Teenager', 'Young Adult' and 'Adult'.

After the generated transformations, 3 new features were obtained (e.g. 'AgeGroup', 'CaptureDate_Hour' and 'PartsOfDay'). In this way, the objective for the *ML* models is defined, more specifically the age group, and the definition of two new variables, which in turn enable the establishment of a link between fatigue levels and the time of recording, which will be discussed in the following section.

4.3 CASE STUDY

As previously mentioned, the data set that will be subject to a case study is composed of biometrics that describes interactions with mouse and keyboard as well as other features, that in turn, describe the type of task to be performed, timestamp and the part of the day in which these interactions were registered.

In order to obtain the age of an user, unreliable data was removed. In this way, around 119 records which are referring to the birth date 7 August 1005, were removed. This specific date is used to highlight registration errors or simply an invalid input on the birthday parameter. Then, 661 records that have missing Timestamp were removed, due to it being a high importance variable for age calculation; as well as possible duplicates and ages that are considered fallacious, with only ages between 6 and 80 years old being studied, making a total of 1047791 records in the data set.

With the removal of these data, it is now possible to analyze the age distribution found for the case study. The case in question was created from the interactions of approximately 319 people, aged between 16 and 50 years, as can be seen in the following graph.

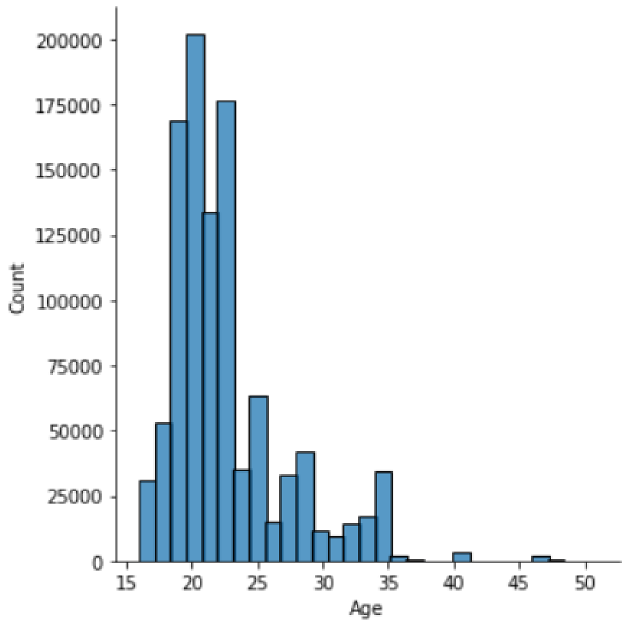

Figure 16: Occurrences per age

	Occurrences	Relative Frequency
[15,17]	31021	0.029606
[18,24]	769489	0.734391
[25,64]	247281	0.236003

Table 1: User age statistics

From Figure 16 and Table 1, it is possible to verify the distribution of the registered ages, and in turn, their age group. At first sight, it's possible to see that almost three-quarters of users are between 18 and 24 years old, consequently creating an unbalanced distribution, which in turn will influence the predictions made by the *ML* models.

It's also possible to see that most of the records refer to males, as can be observed in Figure 17.

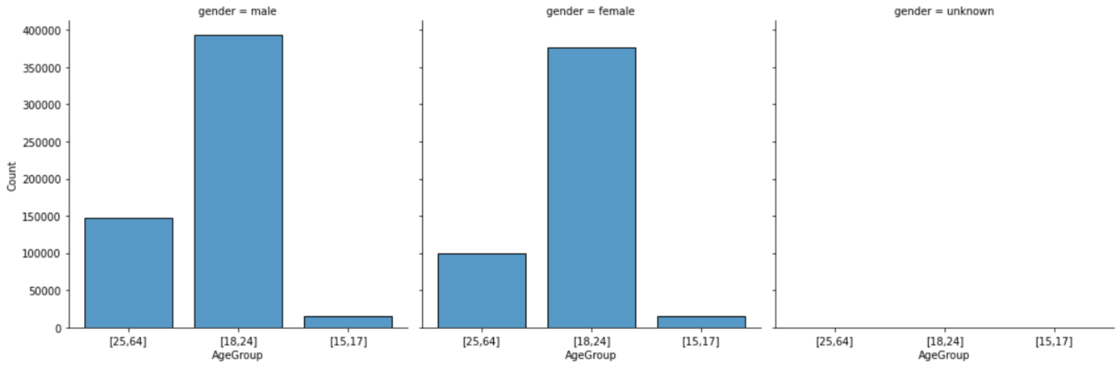

Figure 17: Gender occurrences through age group

As already mentioned, each of these records is associated with a task describing the type of action that took place at the time of capture. Some of these may induce higher levels of fatigue in a user, thus influencing the collected biometrics, as it's possible to see in the following figure.

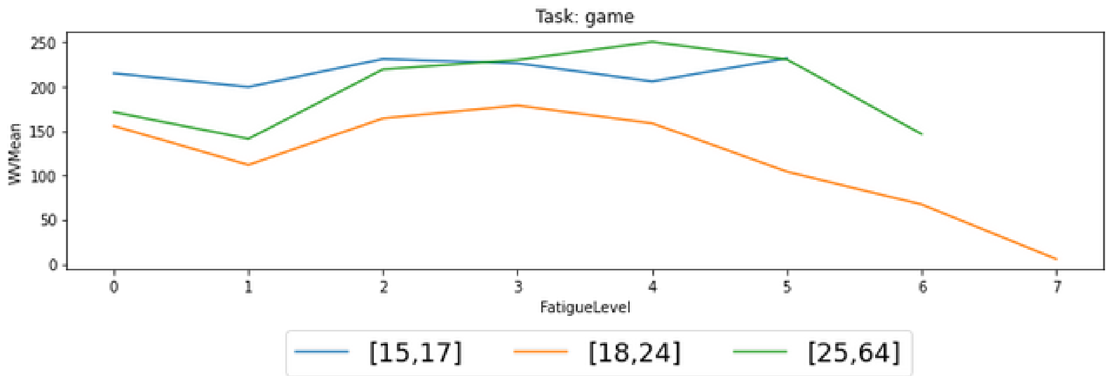

Figure 18: Impact of fatigue levels on the 'WVMean' feature

From Figure 18, it can be seen that as fatigue levels increase, the variable 'WVMean' decreases. Meaning that fewer keys are pressed over the course of a fatigued state, which can lead to a worse performance given the nature of the task at hand. This scenario checks for all age groups, however the '[15,17]' group doesn´t have many records, so there is the possibility of not capturing all possible scenarios, yet it's possible to see that it follows the same trend as the others groups.

The following figure aims to present which tasks show the highest levels of fatigue and the period of the day in which they occur. Therefore, it is possible to identify 10 tasks, which in turn characterize what type of interaction occurred at the time of the recording.

- net: All interactions with a web browser (e.g. 'Chrome', 'Firefox' and 'Opera');
- game: Playing games such as 'League of Legends', 'Fortnite' and so on;
- programming: Usage of code editors (e.g. 'Sublime Text' and 'Notepad++') or IDE (e.g. 'Visual Studio Code' and 'BlueJ');
- chat: Exchange messages from applications such as 'WhatsApp' or 'Facebook Messenger';
- office: Usage of tools such as 'Word', 'PowerPoint' and 'Excel';
- reading: Use of programs such as 'Adobe Acrobat Reader' for reading pdf.;
- drawing: Interactions with programs, like 'Paint' and similar;
- video: Watching videos through programs like 'VLC';
- music: Listening to music through 'Spotify', for example;
- others: Any other interaction that has not been identified by the other tasks;

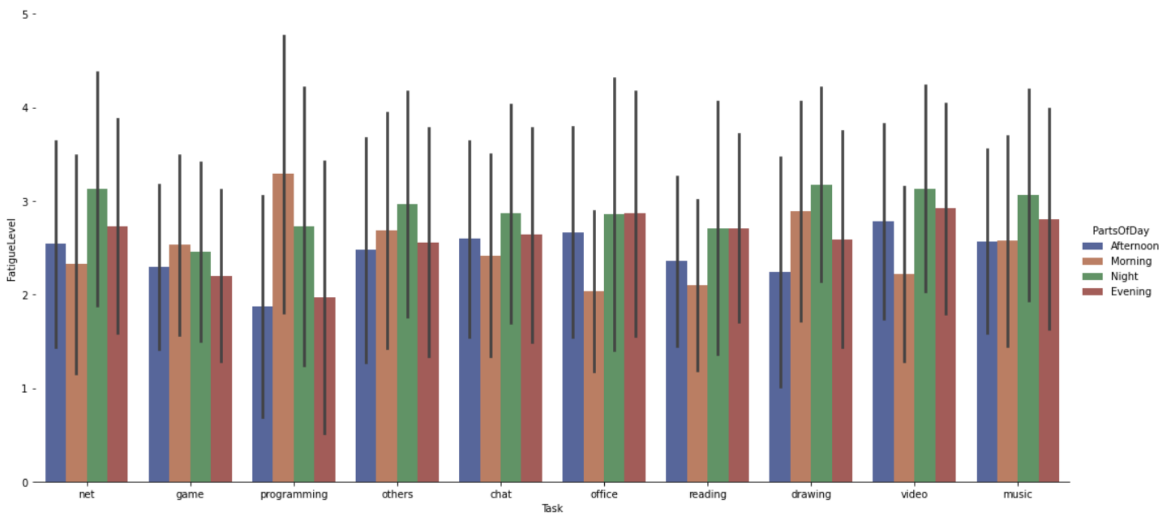

Figure 19: Fatigue levels occurrence through task and parts of the day

As can be seen in Figure 19, many of these tasks have similar fatigue levels, yet it's clear that some particular tasks tend to have higher fatigue levels for any part of the day, as is the case of 'music'. While others present a small impact throughout the day, such as 'game' and there are still others that have higher levels throughout the day, such as 'video', 'chat' and 'net'.

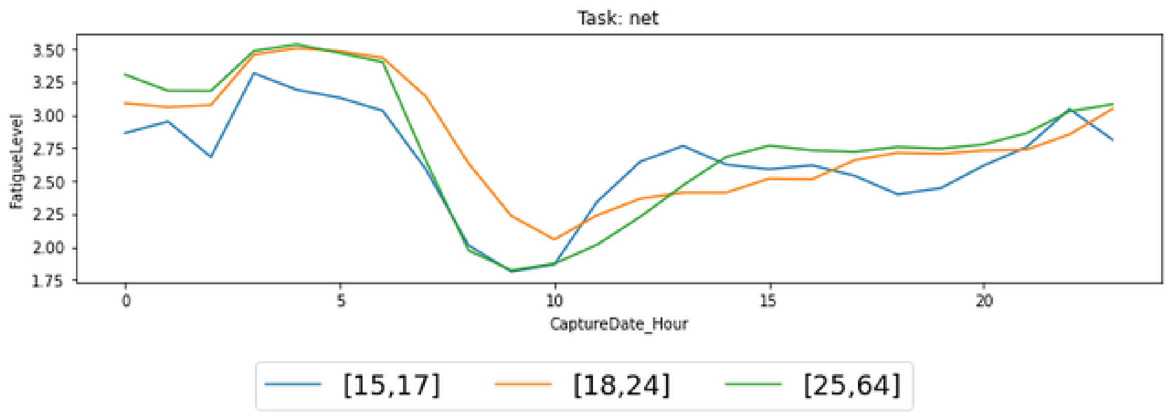

Figure 20: Evolution of fatigue levels of the 'net' task over a day

As mentioned before, there are cases wherein a task it's noticeable that fatigue levels gradually increase as the day progresses. This scenario is confirmed by Figure 20, in which in the morning periods, fatigue levels appear at their minimum, advance over time and reach their peak in more advanced periods of the night, approximately at three in the morning.

4.3.1 *Missing Values*

The applied treatment may be in the form of insertion and forecast of values. However, if it's not possible to administer any treatment described above, the solution may be the removal of that same record.

Since certain tasks may require more keyboard or mouse related biometrics, it's necessary to maintain the integrity of this variable. So, to avoid introducing possible errors that can be misleading, for example, the substitution of missing values for the most frequent ones, 28 records related to the task being performed were also removed.

With the removal of features that have more than 60% of missing values, there are still other variables to deal with (e.g. 'KIVar', 'MMAVar', 'CDMean', ...) and these cases have much lower numbers, from 1 up to 14357 of records with missing values. Thus, an analysis was made on these same features, finding out that some variables have missing values, but their "homologous" does not, such as the variable 'TBKVar' which has missing values although the 'TBKMean' doesn't. Therefore, for each of these cases, some graphics were created to illustrate the relationship between these two variables and, in turn, an analysis was made for each of these cases, followed by a treatment for the missing values.

Regarding the case mentioned above, for the variable 'TBKVar', a graphic was created intending to illustrate its relationship with its counterpart.

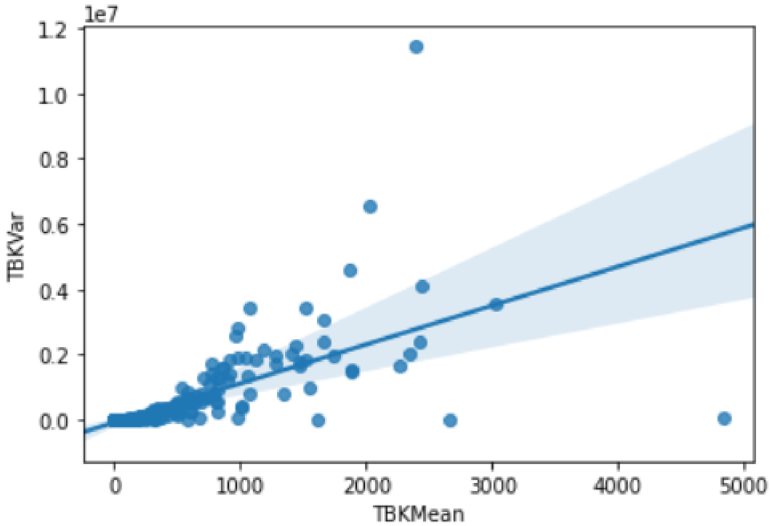

Figure 21: Relation between TBKMean and TBKVar

As we can see, these two variables present a positive correlation, that is, when one of them increases the other does as well. Therefore, for this particular case, a specific ML algorithm was used, called Random Forest, to predict the missing values in the 'TBKVar' feature from the 'TBKMean' values. Then it was applied the "Standardization" technique to transform the scale into a normal distribution. After this, the same data was learned by the mentioned model, presenting an *Mean Squared Error* (*MSE*) of 0,15 and an R^2 of 0,71. With the trained model, it is used to predict the missing values from the 'TBKMean' values of those same records. For the feature 'AEDVar', initially, the same line of thought was followed to the variable described above. However, from the generated graphic the only conclusion drawn is that, for any value of 'AEDMean', 'AEDVar' will be 0. This detail can be easily verified when taking into account the formula for calculating the variance.

$$\sigma^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1} \tag{2}$$

Once again, for the variables 'MouseExcessDistance' and 'MouseDistance', graphics were created to understand the relationship between these two variables.

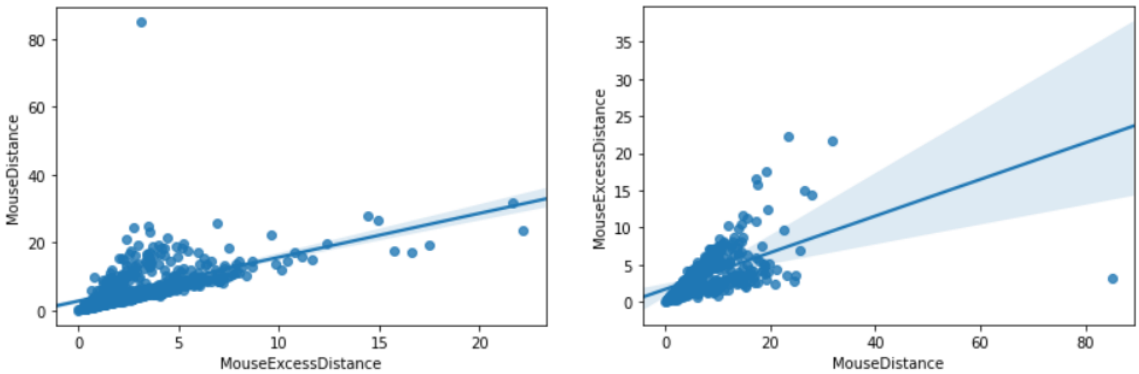

Figure 22: Relation between MouseDistance and MouseExcessDistance

It's possible to notice that these same features have a positive correlation, yet there are several scattered points along with the existing trend. Considering there are a total of 3 missing values between these two variables, these will be replaced by the median of the feature values in the group where they are found. These groups are made up of variables that allow for a better distinction, namely 'FatigueLevel', 'Task', 'PartsOfDay', 'AgeGroup' and 'gender'. That is, a subsection is created from the variables mentioned above, through this it's calculated the median of the values in the feature containing missing values, where this result will replace these gaps.

Mouse Distance	FatigueLevel	Task	PartsOfDay	AgeGroup	gender
55.5	2.0	game	Evening	[18,24]	female
?	2.0	game	Evening	[18,24]	female
92.3	2.0	game	Evening	[18,24]	female

Mouse Distance	FatigueLevel	Task	PartsOfDay	AgeGroup	gender
55.5	2.0	game	Evening	[18,24]	female
73.9	2.0	game	Evening	[18,24]	female
92.3	2.0	game	Evening	[18,24]	female

Figure 23: Example of the proposed treatment

After the previously mentioned treatments, there are still about 12 variables, one of them being the 'CDMean' feature, that have about 6112 missing values each. Taking this into account, these same columns were analyzed, concluding that these missing values appear in equal amounts on the same 'Task' values. Therefore, to reduce the processing required to apply the technique previously mentioned, it was investigated in which parameters these missing values arise, having found that all these values appear for males and a level of fatigue of 0. After applying all these techniques, all missing values were treated or removed.

4.3.2 *Correlation*

With the resolution of the missing values, the next step followed was the analysis of the existing correlation between the current features. Correlation is a statistical method that expresses the extent to which two variables are linearly related, i.e. both evolve at the same pace. In other words, the higher the value, the more similar two variables are, making it possible to remove one of them.

Considering there are several relationships between all the variables present in the data set, a Heat map was created to represent these relationships and their intensity. This type of graphic differs from others created, by the usage of a color scale representing the intensity of the relationship, allowing a better knowledge of the data set itself.

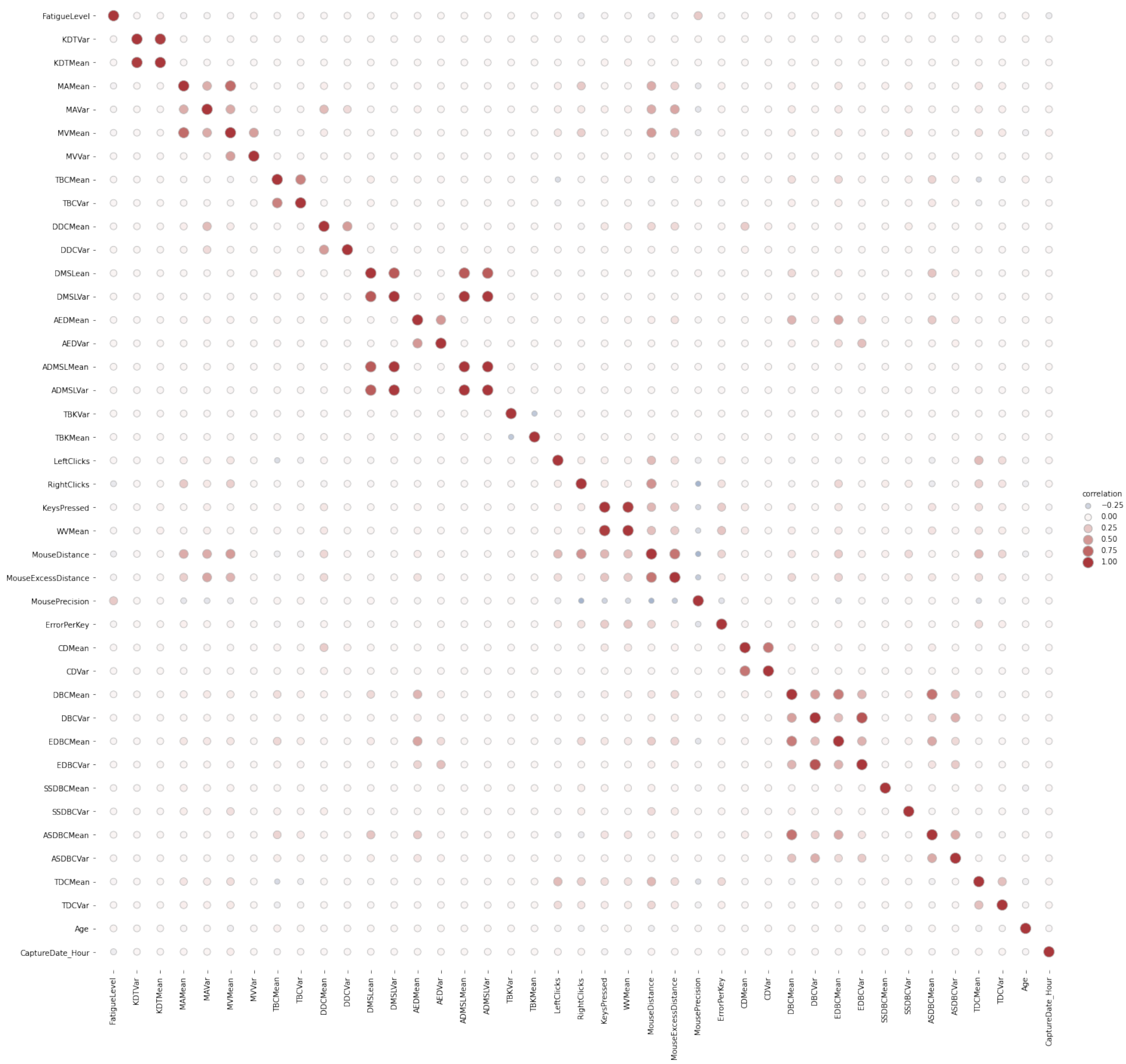

Figure 24: Correlation heat map

As expected, the biometrics described through their mean and variance show correlations of at least 0.75. Besides, other relationships can be mentioned, such as 'KeysPressed' and 'WVMean', which are related to interactions with a keyboard. On the other hand, there are values below 0, as in the relationship between 'RightClicks' and 'MousePrecision', despite both being interactions via mouse.

From the information collected from Figure 24, it is then possible to remove features that present high correlations. The value that stood out best in the tests, which will be detailed in the following section, is a correlation of 0.8. Therefore, features that have at least this value for correlation were removed, thus eliminating the variables 'EDBCVar', 'KDTMean', 'ADMSLVar', 'DMSLVar', 'WVMean' and 'ADMSLMean' from the data set.

4.3.3 *Outliers*

With the treatment of missing data and the elimination of unnecessary features, an analysis of the existing outliers in this data set was carried out. Therefore, an outlier can be defined as one that significantly differentiates from other observations[4].

As such, a boxplot is a good way to describe data at the middle and the end of a distribution. This type of chart uses the median, the upper and lower quantiles, from which it allows the calculation of fences making it possible to identify different types of outliers, as can be seen in the following image.

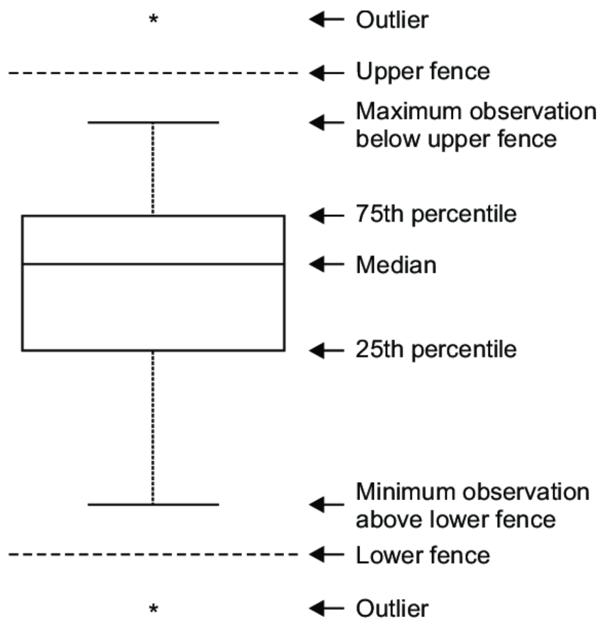

Figure 25: Description of a boxplot and its fences

Source: (Kwak and Kim, 2017)

Figure 25 shows the different details present in a boxplot, from the quartiles, median and fences. From these fences, it's possible to categorize the outliers, where a point present after the inner fence is called a mild outlier, and after the outer fence is considered an extreme

4 https://www.itl.nist.gov/div898/handbook/prc/section1/prc16.htm

outlier. With this in mind, boxplots were generated for each of the variables, enabling the discovery of outliers in a more graphic way.

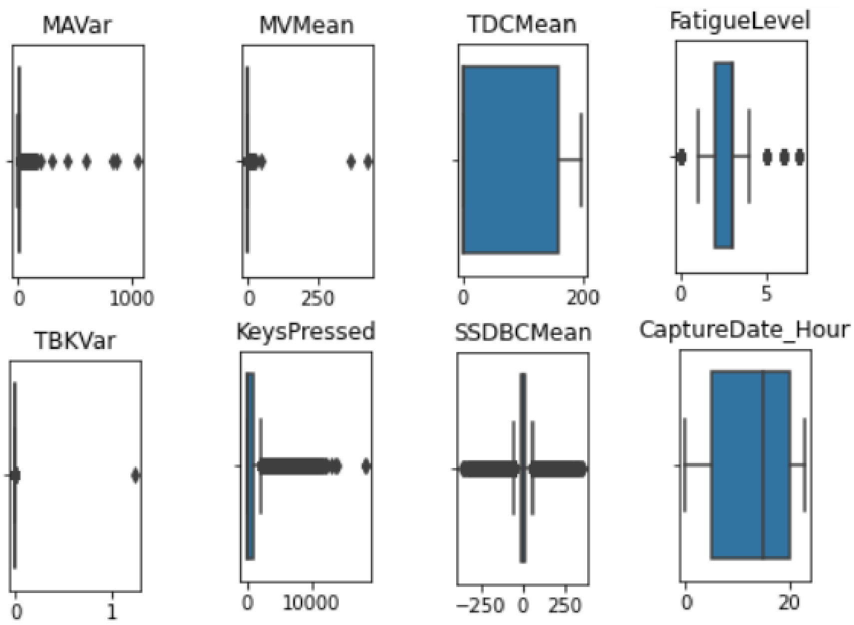

Figure 26: Display of a boxplot and its fences before the application of Winsorizing

At first glance, it's possible to see that, within the presented variables, some have outliers, except 'TDCMean' and 'CaptureDate_Hour'. A possible solution discovered for the treatment of these outliers is the Winsorizing technique, which is described by the limitation of extreme values. This limitation is described by replacing values below and/or above a determined percentile with that same percentile[5].

Before proceeding with the application of this technique, several tests were performed to detect the best percentiles to be used, with the 0 and 95th limits standing out. This is due to present, in general, a lower standard deviation than the others, with this, the technique was applied to the existing variables, except for the 'FatigueLevel', 'SSDBCMean', 'TDCMean' and 'CaptureDate_Hour'. These variables were ignored due to the distribution presented, more specifically 'SSDBCMean' and another due to its purpose for this data set.

5 https://blogs.sas.com/content/iml/2017/02/08/winsorization-good-bad-and-ugly.html

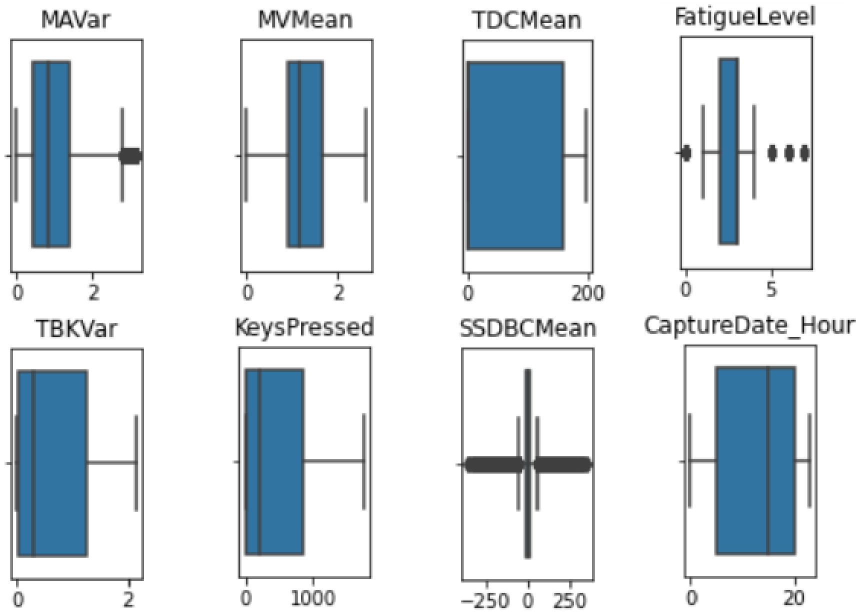

Figure 27: Effect of the winsorizing technique

As can be observed, almost all outliers were dealt with, except for some situations which were resolved manually.

4.3.4 *P-Value*

Once the analysis of the existing outliers was completed, a statistical method was applied to obtain only the variables that present the highest significance for the problem. The chosen method was the p-value, which can be defined as the probability of obtaining results, at least as extreme as the observed results of a statistical hypothesis test, assuming that the null hypothesis is correct. This method, as mentioned, is used to provide the lowest significance level at which the null hypothesis will be rejected, i.e. the smaller the p-value, the greater its statistical significance[6].

Considering this, and to take advantage of this statistical method, a scenario was created in which various combinations of variables were tested, to obtain the features that have greater significance for the objective. After the execution of the previous scenario, three features were removed, 'CDMean', 'DDCVar' and 'AEDMean', i.e. these variables have a p-value

6 https://www.investopedia.com/terms/p/p-value.asp

higher than 0.05 threshold. After removing these variables, graphics were generated for each of these characteristics to analyze the distributions presented by the variables concerning the age group.

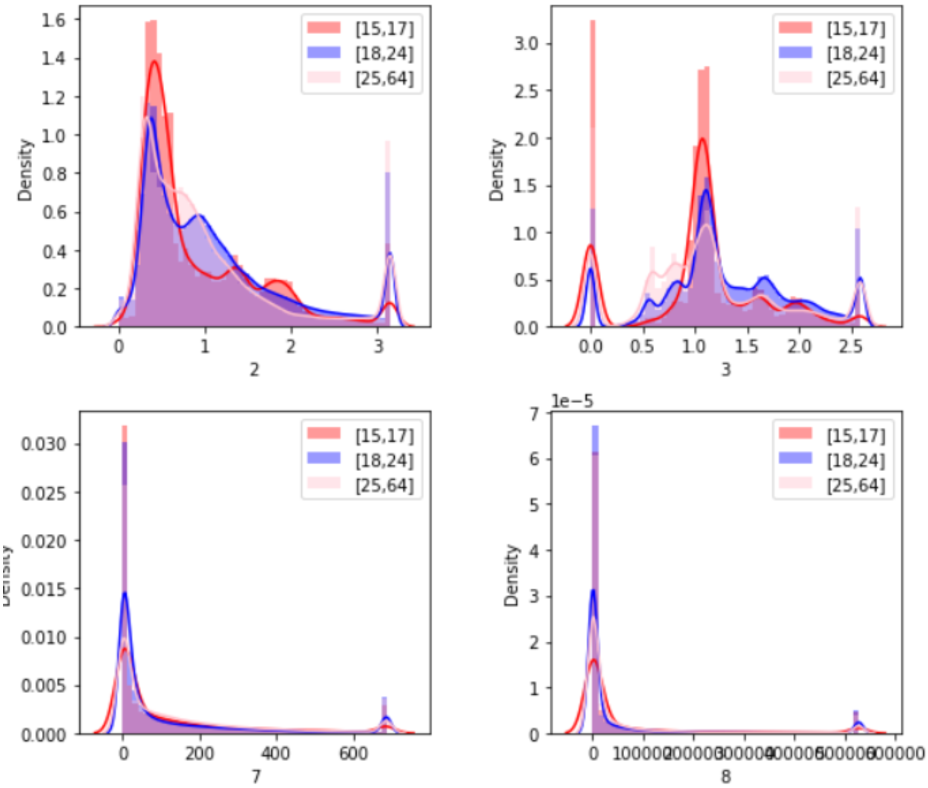

Figure 28: Features distribution amongst age groups

Figure 28 intends to present the distributions of four features across different age groups, as expected, the variables show similar trends, differing only in the values presented by each of the age groups.

4.3.5 *Encodings*

Once the proposed and previously applied treatments are finished, the application of encodings for categorical values remains. The framework that will be used, H2O.ai, allows these same categorical values, by assuming, in turn, continuous values during the training of the model. However, the purpose of applying these encodings is to transmit notions of order or even equality. Therefore, the label coding technique was applied to the 'PartsOfDay' resource, i.e. each of its values were replaced by a numerical value, more specifically, the value corresponding to an earlier part of the day will have a lower value, and by turn the later part of the day, a higher value. In addition to this applied encoding, one-hot encoding was also used in the 'gender' and 'Task' features. This technique is based on creating a new variable for each value of the mentioned features where in these new variables, the value of 1 where the registration occurred and 0 where not, will be taken.

KeysPressed	MousePrecision	Task
0.0	12.7	reading
208.0	98.857	game

KeysPressed	MousePrecision	Task	reading	game
0.0	12.7	reading	1	0
208.0	98.857	game	0	1

Figure 29: Application of the one hot encoding technique

From Figure 29, we can see how the one-hot encoding technique is processed, yet after its application, the feature to which this method was applied remains, so its removal will be necessary. Therefore, the 'gender' and 'Task' variables will be removed, together with the 'Age' variable as it would influence the model results, and consecutively its knowledge would require a user to provide it.

4.4 ANOMALY DETECTION

With the treatment completed, detection of possible anomalies in the data set is now performed. The objective is to identify suspicious observations, events and items, which may indicate some problems during data collection (e.g. damaged sensors, typos on forms, ...) or unexpected events, such as security breaches, server failures and so on.

This type of detection will be made with an Isolation Forest model. In this type of model, the trees are divided randomly, as well as each leaf that isolates exactly one observation from the data set. Thus, assuming that one register is similar to others, more arbitrary divisions will be needed to isolate that observation perfectly, rather than isolating an outlier. Therefore, the smaller the number of divisions required for the register, the more likely it is to be anomalous[7].

So, to obtain these predictions, the data set resulting from the treatments mentioned above was used. This data set will be subject to four forecasts, one aimed at detecting global anomalies and another three for local anomalies, specifically each age group.

The first anomaly to be detected will be the global one, so the entire data set will be used. The model output shows the average number of splits needed to isolate a record across all generated trees.

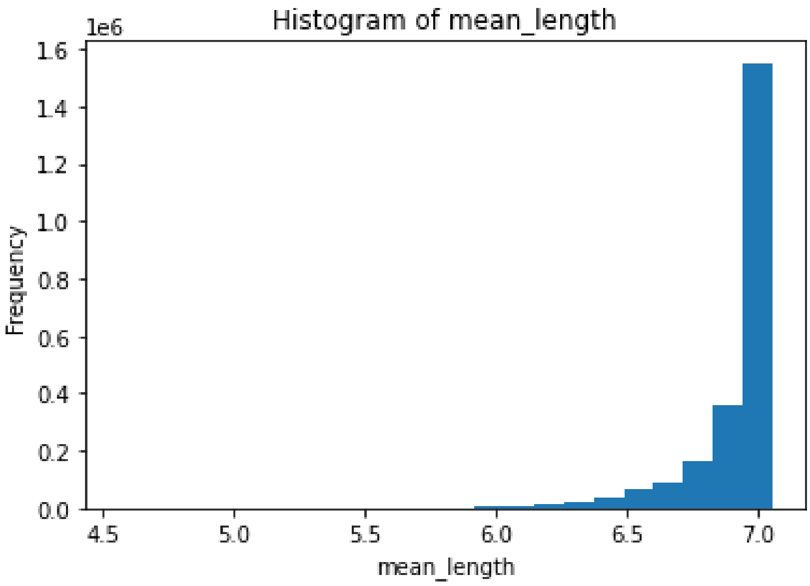

Figure 30: Global mean splits

Figure 30 intends to demonstrate the frequency of the main average numbers of splits, showing that the most frequent minimum number is around 6.0. Therefore, all records having a mean_length less than 6.0 will be considered anomalous values, so around 15508 records were identified. To better understand why a value is considered anomalous, a test scenario was created using an anomaly as an objective. For this new test scenario, a Random

7 https://www.h2o.ai/blog/anomaly-detection-with-isolation-forests-using-h2o/

Forest model was created, which allows the creation of a diagram describing why this new objective is considered an anomaly.

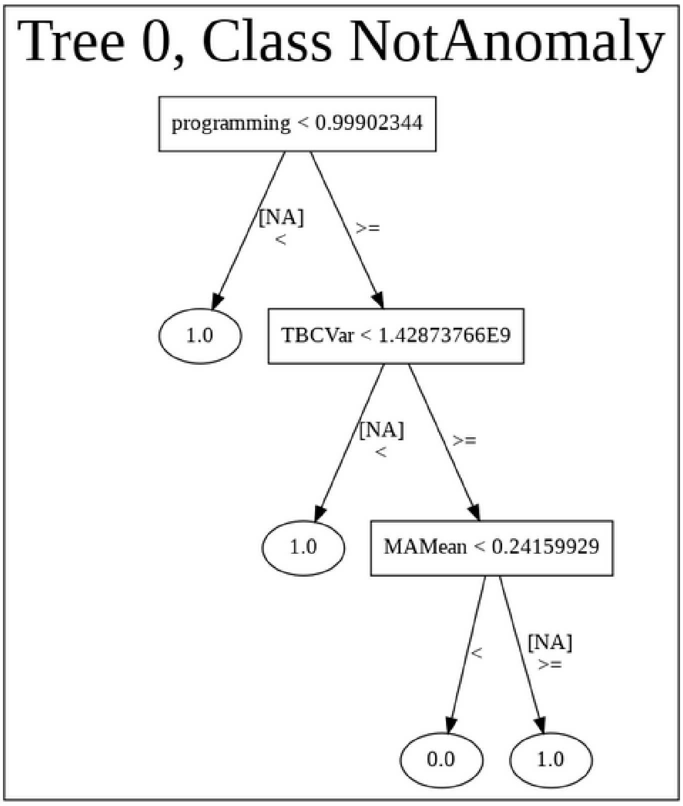

Figure 31: Splits necessary to detect anomaly

The diagram showed above displays which features were used to predict the anomaly in question and which values were considered for decision making. The values represented on the leaves indicate the probability that a register is not anomalous, so the interest goes through the lower values. As can be seen, for this test case, an anomalous record has a 'programming' value greater than or equal to approximately 0.99902344, a 'TBCVar' value greater or equal to 1.48737, and lastly a 'MAMean' less than 0.241599[8].

8 https://github.com/h2oai/h2o-tutorials/blob/master/tutorials/isolation-forest/interpreting_isolation-forest.ipynb

4.5 TEST SCENARIOS

The following section aims to describe the tests performed for each of the algorithms used, to achieve an effective and efficient prediction. The algorithms used, *GBM* and *Extreme Gradient Boosting* (*XGBoost*) proved themselves as being multifaceted, i.e. both models can be used for either type of prediction, among other applications. However, they also have certain disadvantages, the main one being the computation needed for large data sets. In addition to these models, it's also used the Stacked Ensemble algorithm which will be created from the previously generated *GBM* and *XGBoost* models. This type of algorithm has the advantage of combining the knowledge of other models into one, enabling a performance superior to that obtained by a single model.

In these tests, several experiments are performed, such as, filtering through different correlation data, removing certain variables (e.g. 'FatigueLevel', 'gender', 'Task' or 'PartsOfDay'), or variables that indicate the presence of anomalous data or not. Along with using other approaches to balance the data set through the inclusion of synthetic data. In addition to all these tests mentioned, no data transformation in terms of scale was used, this was due to avoiding any "scale" overshoots, leading to unreliable results.

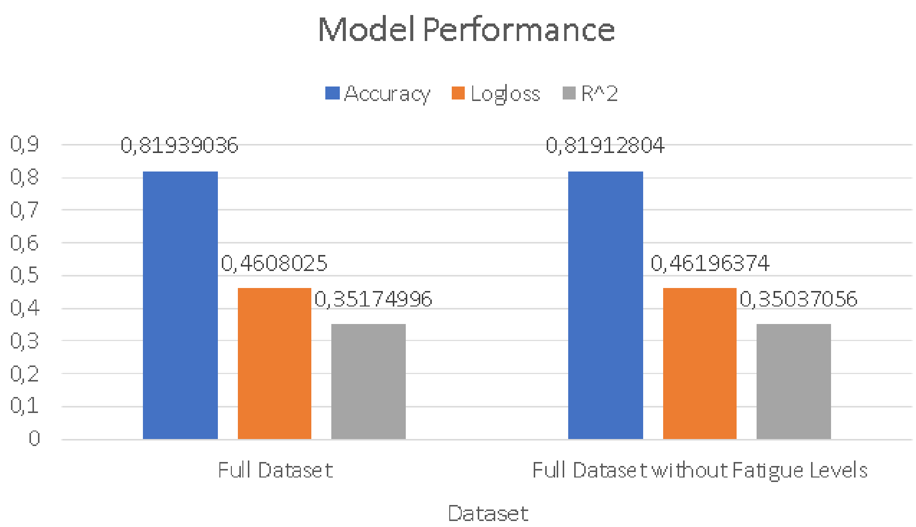

Figure 32: Performance differences with existence of Fatigue Level feature

In Figure 32, it's noticeable that the presence of the variable 'FatigueLevel' leads to better results from the models generated with the *GBM* algorithm. This same scenario can also be verified for the *XGBoost* algorithm.

From the results obtained in the previous figure, it's also possible to answer one of the main questions posed at the beginning of this project. In other words, the presence of variables that describe the fatigue state of an individual helps to obtain better results in terms of age prediction through human-computer interactions.

4.5.1 *GBM*

As mentioned previously in Section 4.5, several tests were performed to obtain the best combination of features for the algorithm in question. The test that stood out was obtained through the application of the *Synthetic Minority Oversampling Technique* (*SMOTE*) and the removal of the same features through 0.8 correlation and P-Value.

The following graph shows the performance differences between the best data set and a "base" data set, that is, containing all the features having been dealt with the missing values and having applied feature engineering as in Section 4.2.

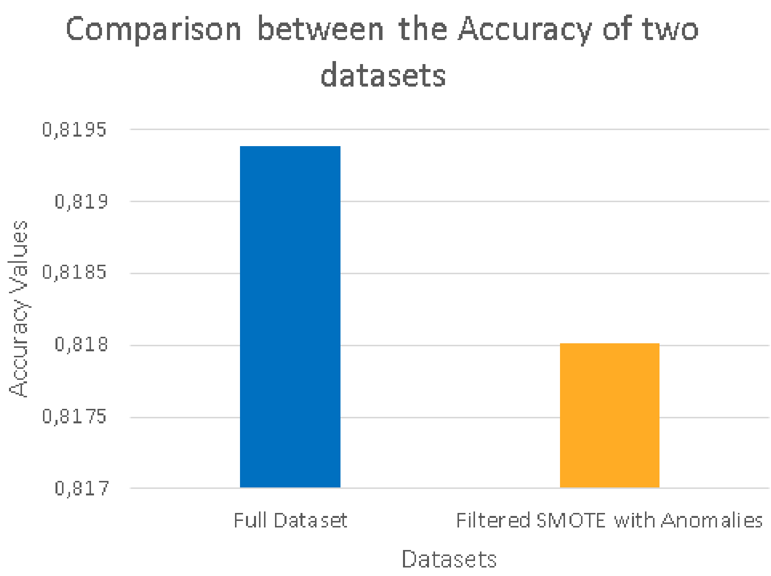

Figure 33: Accuracy results across data set

Through Figure 33, it's possible to see some performance differences between the two data sets. However, as it can be seen, the chosen data set does not have the highest accuracy value among the two, so two more metrics were taken into account (e.g. Logloss and R^2). The Logloss is an indication of how close the predictions are to the actual values, while the R^2 indicates how well the input data can understand present variations.

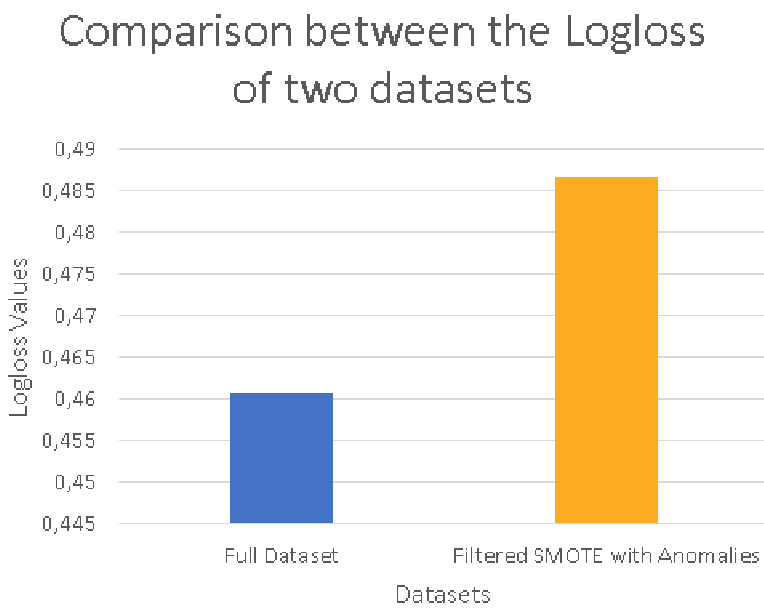

Figure 34: Logloss results across data set

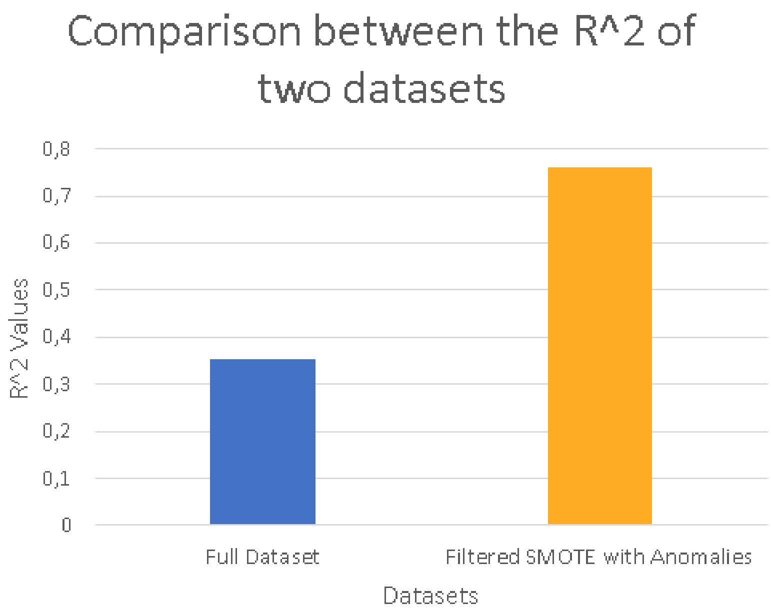

Figure 35: R^2 results across data set

As can be seen, although the chosen data set presents a worse performance regarding the accuracy and logloss, however, it displays better results in the R^2 metric. The difference in performance between these two data sets can be explained by the amount of data that exists for each label. That is, as the original data set is highly unbalanced, a model for this data set ends up just "learning " the patterns for a particular age group.

After obtaining the best data set, the optimization of the hyperparameters was performed, for which the Random Search optimization method was used. Through this method, four optimization phases were created, where in each one a search is conducted to obtain the best combination of hyperparameters in that phase.

1. Hystogram type & Fold Assignment;
2. Number of trees & Learning Rate;
3. Depth of the tree & Min Rows & Min Split Improvement;
4. Sample Rate & Column Sample Rate, per Tree or Level;

Therefore, the following graphics aim to illustrate the evolution of accuracy and logloss, along with these optimization phases.

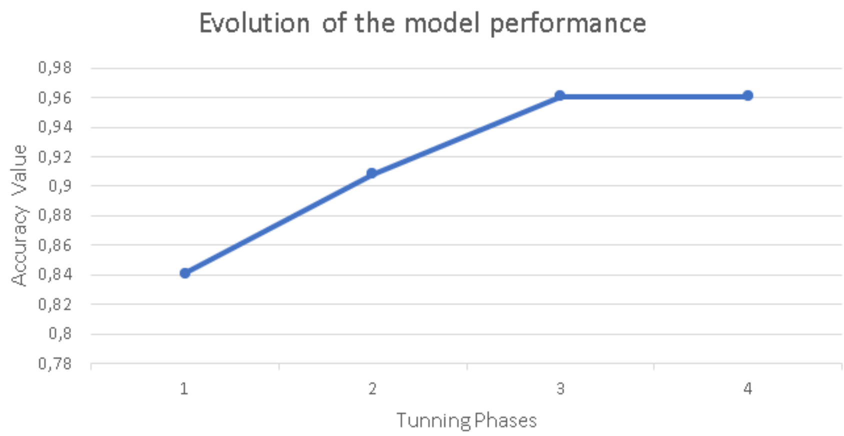

Figure 36: Accuracy evolution along the tuning phases

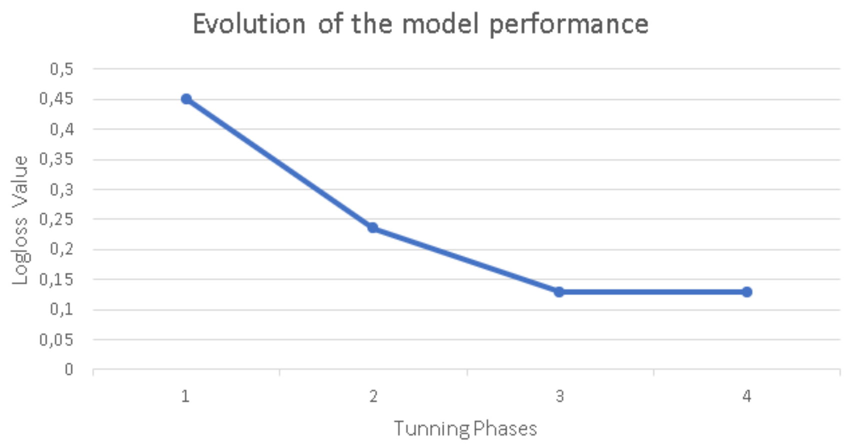

Figure 37: Logloss evolution along the tuning phases

In the previous figures, it is possible to visualize the performance of the generated model, along with its optimization phases. These, as indicated, were optimized through the Random Search method, where ten random models were created from a search space to find the best hyperparameters. The parameters selected for each phase are based not only on the performance of the 'Accuracy' metrics but also on the 'Logloss' and 'R^2' metrics. In other words, the parameters that present the best performance in the globality of these three metrics are chosen. The following table aims to show the disposition of the evolution of the generated model's performance.

		Tuning Phases			
	Base Model	1	2	3	4
Accuracy	0.8180197	0.8407494	0.90844	0.96105	0.96105
Logloss	0.48663276	0.44965398	0.23447	0.12883	0.12883
R^2	0.7612363	0.78321123	0.89403	0.95314	0.95314

Table 2: GBM performance summary

4.5.2 *XGBoost*

Based on the line of thinking described in Section 4.5.1, the same number of tests and their purpose was repeated to obtain the best data set for the *XGBoost* algorithm. After performing these tests, it was found that the data set that presents the best result is achieved by applying the *SMOTE* technique and feature elimination through 0.8 correlation and P-Value. This technique creates synthetic values that will balance the data set in question, thus avoiding less biased and, in turn, more credible predictions; however, the same features were removed, as in Section 4.5.1. Once again, the following graphs aim to present the performance obtained from the best data set and the base data set.

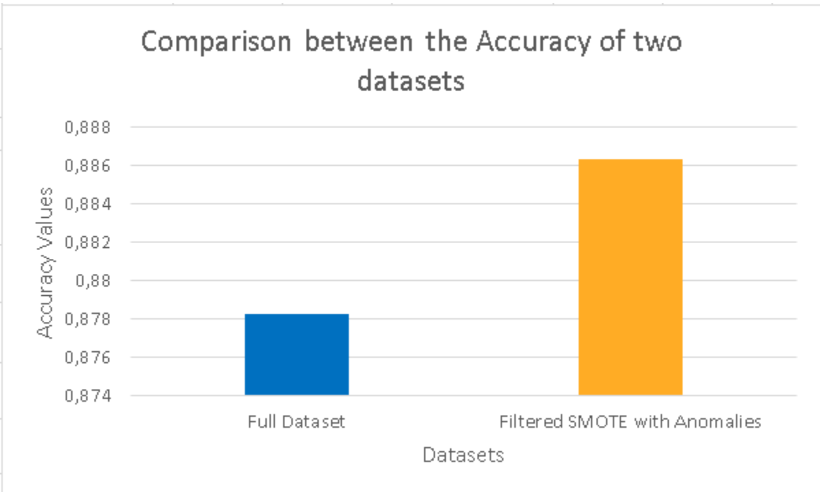

Figure 38: XGBoost accuracy results across data set

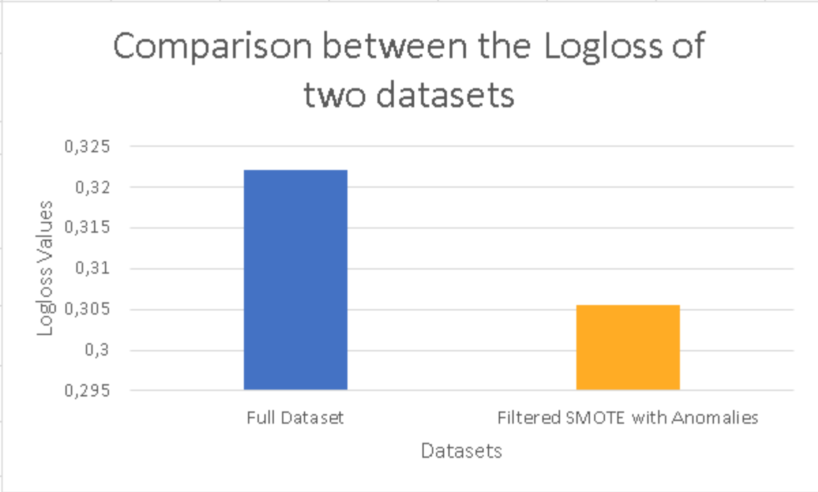

Figure 39: XGBoost logloss results across data set

As can be seen, and despite the scale, the difference in performance between these two data sets is not very large, so a third metric was taken into account, once again R^2 was chosen.

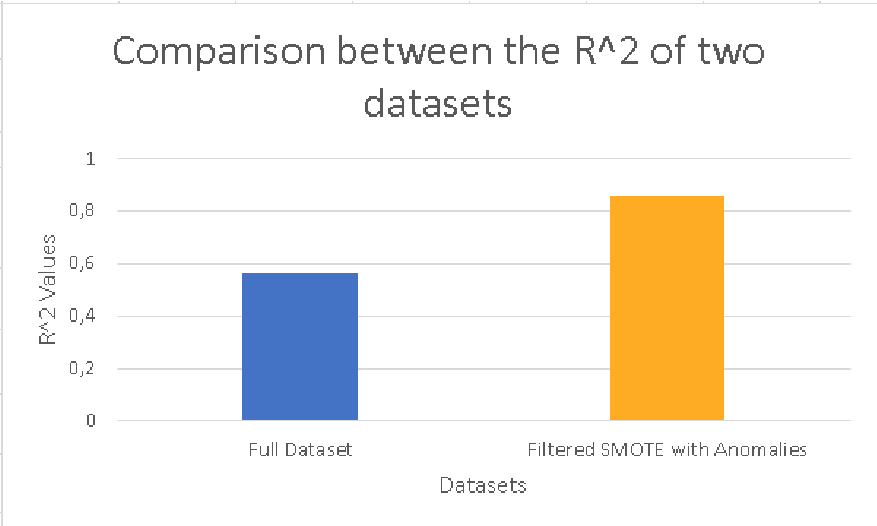

Figure 40: XGBoost R^2 results across data set

From Figures 38 to 40, it's possible to see that the data set generated by the *SMOTE* technique presents the best results for the three selected metrics. Identical to what was done in Section 4.5.1, the best data set was subjected to the same optimization method. However, the phases diverge, given the hyperparameters available in this new algorithm.

1. Fold Assignment & Booster & Tree Method & Grow Policy;

2. Sample type & Normalize type;
3. Number of trees & Learning Rate;
4. Depth of tree & Min Rows & Min Split Improvement;
5. Sample Rate & Column Sample Rate, per Tree;
6. Rate Drop, Skip & One Drop;
7. Reg Lambda, Alpha;

The following graphs tend to demonstrate the evolution of the model's performance through the phases described above.

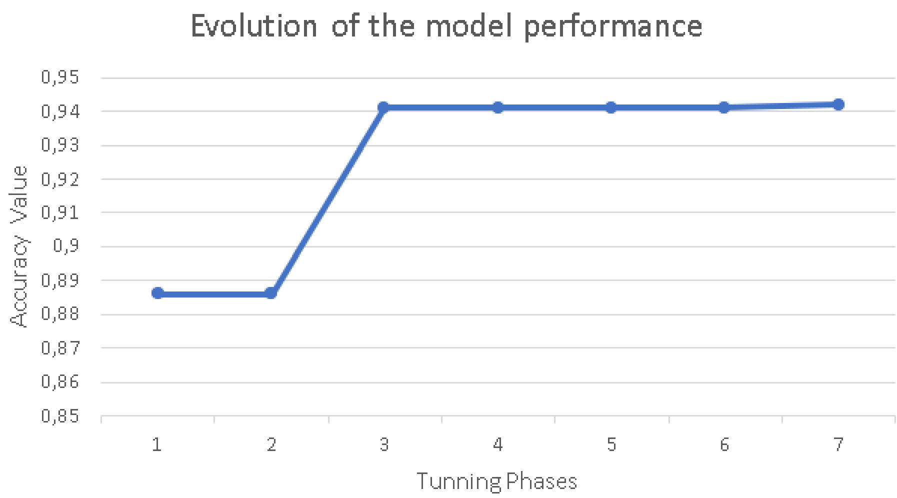

Figure 41: Accuracy evolution along the tuning phases

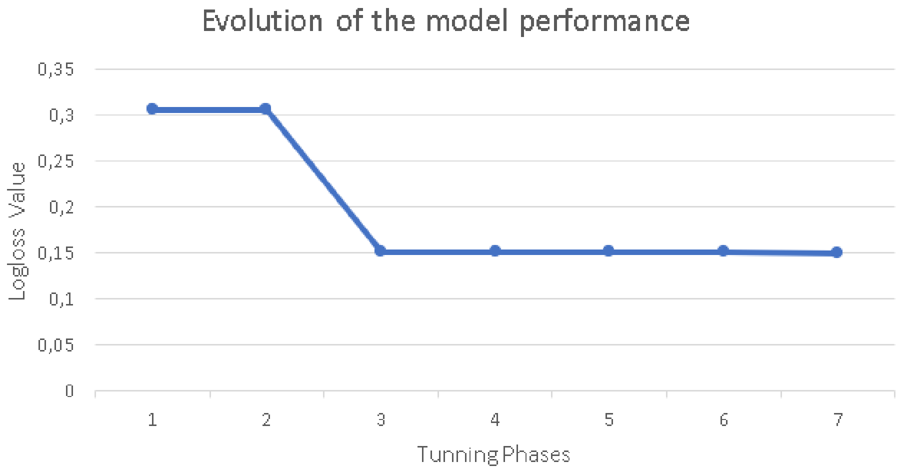

Figure 42: Logloss evolution along the tuning phases

		Tuning Phases						
	Base Model	**1**	**2**	**3**	**4**	**5**	**6**	**7**
Accuracy	0.8863446	0.8860619	0.8860619	0.9412085	0.9412085	0.9412085	0.9412085	0.9419889
Logloss	0.30546758	0.30600512	0.30600512	0.15216033	0.15216033	0.15216033	0.15216033	0.15015613
R^2	0.8603539	0.8601301	0.8601301	0.9325447	0.9325447	0.9325447	0.9325447	0.9334495

Table 3: XGBoost performance summary

4.5.3 *Stacked Ensemble*

After the previously presented models have been tuned, there is still the possibility of obtaining higher performance based on their knowledge. The algorithm that will be presented below is called Stacked Ensemble. Also known as "Super Learning", Stacking is a class of algorithms whose objective is to train a "meta learner" to optimize the knowledge of base models[9].

From the use of this type of algorithms, the objective is to find the best combination of base models, through the models presented in the previous sections, and the best "meta learner" for that combination. Thus, the combination that stood out in its generality in all available "meta learners" is composed of the *XGBoost* tuned models with Bayesian Optimization method and the *GBM* model obtained via a Random Search optimization method.

9 https://docs.h2o.ai/h2o/latest-stable/h2o-docs/data-science/stacked-ensembles.html

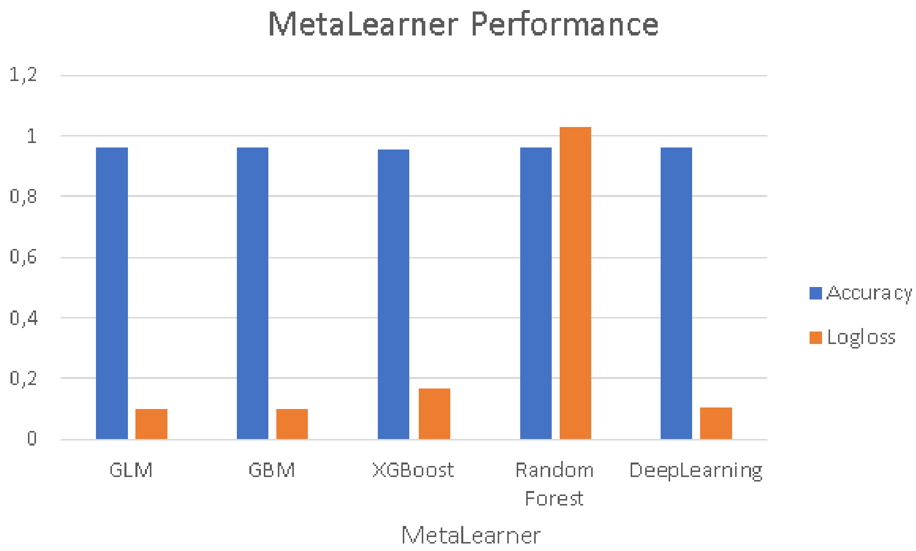

Figure 43: MetaLearner performance

Through Figure 43, it is possible to verify the different performances for each one of the "meta learners" for the combination mentioned above. Here, the *GBM* algorithm stands out, as it presents the best results for the Accuracy and Logloss metrics.

4.6 RESULT ANALYSIS

With the end of the creation and optimization of *ML* models, it's possible to conclude that all the generated algorithms presented a better performance for a data set generated with the aid of the *SMOTE* technique and with the removal of certain features. This data set displayed about 2 million records and 46 variables that, in turn, characterize biometrics, such as typing speed, mouse speed, the time between keys, among others, or the description of the user's state of fatigue. The following graphic aims to present the performance of each of the models created when submitted to a data set unknown to it.

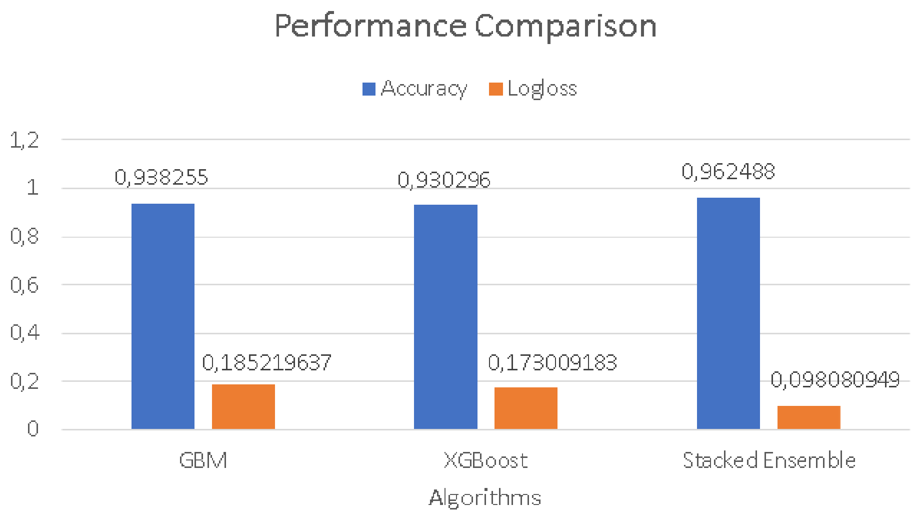

Figure 44: Performance comparison between three algorithms

From the previous figure, it's noticeable that the results obtained are very satisfactory for all algorithms. Regarding the initial algorithms, more specifically the *GBM* and *XGBoost* models, this is due to the random search optimization which produced a performance gain of approximately 12.81% and 4.96%, respectively. However, the next step implemented, assimilates the knowledge of these two models and in turn, returns a more corroborated and improved results. This fact was proven due to the improvement in metrics such as Accuracy, which gained around 2.4% and reduced the value of Logloss to about half compared to the best result of this metric.

5

AGE DETECTION IN MOBILE ENVIRONMENTS

As shown in the previous chapters, *HCI* presents itself as a reliable source of information for the successful prediction of a user's age. However, the data used comes from input peripherals (e.g. mouse and keyboard), and in turn, the generated models cannot be applied in different environments. Thus, the following section aims to apply this idea but on other platforms, more specifically on mobile devices.

The data collected by these platforms comes from a publicly available game. To which controls were added to acquire interactions that in turn characterize the dynamics created between the game and the user.

5.1 CHARACTERIZATION OF THE GAME

The game in question is based on a simple set of moves (e.g. up, down, left and right), in which the objective is to avoid the obstacles that may occur in the path of the user's avatar, and go as far as possible.

Throughout the game and as the users score increases, the game velocity grows gradually, diminishing when the avatar loses one of its lives by colliding into any obstacle. However, there are some items, such as extra life, magnets and multiplicative score, which make it possible to reach higher values in less time.

Figure 45: Possible game movements

5.2 DATA TREATMENT

This section will present the treatment applied to all the data collected, from the form and the game. The form retrieves crucial information such as the username used in the game, birthday and other relevant information. Whereas the game collects all type of interactions during playing time, as well as the given username, timestamp and operative system.

5.2.1 *Missing Usernames*

The merge of the game records with the users' data, collected by the form, is made through the provided usernames. Some of these did not coincide so it was necessary to treat them, that is, to replace a username with its corresponding one in the form records. However, this method alone was not sufficient, as there's game records associated with the default username, "Trash Cat" and in this situation, it was used the "machineID" variable. This variable represents a unique identifier(id) associated with a particular mobile phone model, ids that have two usernames associated to it, being one of them the default, it's assumed that these records belong to the other username of that id. However, there are ids with more than two usernames associated, for this case, game records with the default username will be assigned the most frequent username for that "machineID".

Despite all these methods, there are still 106 records to which it was not possible to associate a user, these were removed from the data set to avoid the insertion of possible errors, allowing the merge of the two data sets.

5.3 CASE STUDY

After treating the data set, it's already possible to verify some characteristics about the universe that we will come across. Thanks to the answers obtained through the form, it is possible to characterize the data through graphics. So, first of all, a profile of the newly formed data set is necessary, which in turn has 1545 records and 70 features.

The case study is faced with users aged between 11 and 36, in a population of 59 people, as can be seen in the following graph:

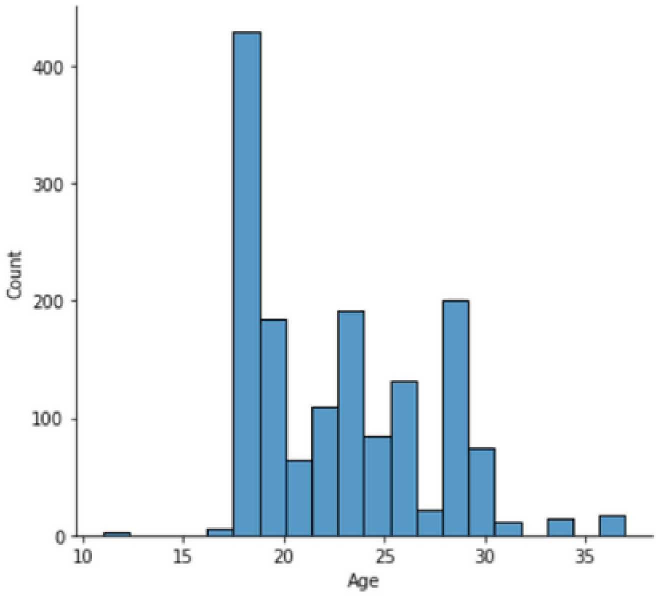

Figure 46: Age occurrences

	Occurrences	Relative Frequency
11	3	0.001942
17	6	0.003883
18	429	0.277670
19	70	0.045307
20	115	0.074434
21	64	0.041424
22	110	0.071197
23	182	0.117799
24	10	0.006472
25	85	0.055016
26	131	0.084790
27	21	0.013592
28	178	0.115210
29	23	0.014887
30	75	0.048544
31	12	0.007767
34	14	0.009061
36	17	0.011003

Table 4: Age distribution

As observed in Figure 46 and Table 4, around 27% of the records are related to users aged 18 years. There is also evidence of a disparity between records, more specifically in terms of quantity, which in turn will influence future results.

Considering that this case study focuses on interactions between humans and mobile devices, the available biometrics will not be the same as those presented in Chapter 4. One of the main differences will be the inexistence of a feature that represents the presence of fatigue in these interactions, yet this may be present given the difference in values throughout the day.

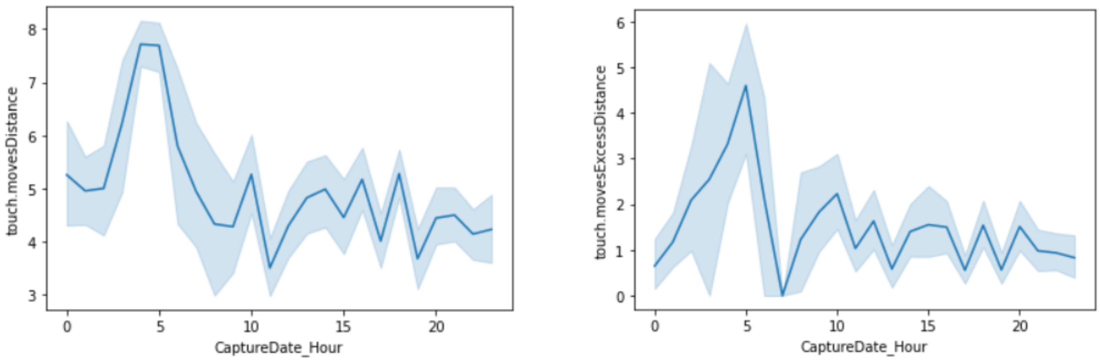

Figure 47: Evolution of distance and its excess over the course of a day

Through Figure 47, it's possible to see that the types of distance increase throughout the day, reaching their peak in later hours, this phenomenon may indicate the presence of fatigue in these interactions.

5.4 TEST SCENARIOS

Following the same line of thought shown in Section 4.5, once again, the *GBM* and *XGBoost* algorithms were used, as well as the Stacked Ensemble method formed from the best-obtained algorithms. To start the prediction process in the best way, several performance tests were conducted for different data set combinations, from filtering resources through different correlation values, identifying anomalous records, *Recursive Feature Elimination Cross-Validation* (*RFECV*), etc. These same tests had a second objective, the detection of the best distribution for the context in question.

5.4.1 *GBM*

As indicated, several tests were made to encounter the best combination, in this way, the best performance was achieved by removing features such as occupation, predominant hand, timestamp, and applying the *RFECV* method as well as an oversampling technique called *SMOTE*. With this technique, it's possible to balance the occurrences of the label so in this way, the model will not be biased. While using the *RFECV* method, several combinations of existing features are tested to reduce the amount of data used without neglecting the performance. In this case the feature universe is reduced to 41.

While carrying out this test, the data set was not subjected to any data normalization, that is, to reduce each feature to a scale between 0 and 1. This combination shows itself to be a relatively better option, an option with all the variables and application, as it presents less complexity and is not subject to "scales" that could in turn be broken, leading to questionable results.

The following graphics allow us to demonstrate the differences in terms of performance, comparing the best data set and its counterpart (same techniques applied except *SMOTE*).

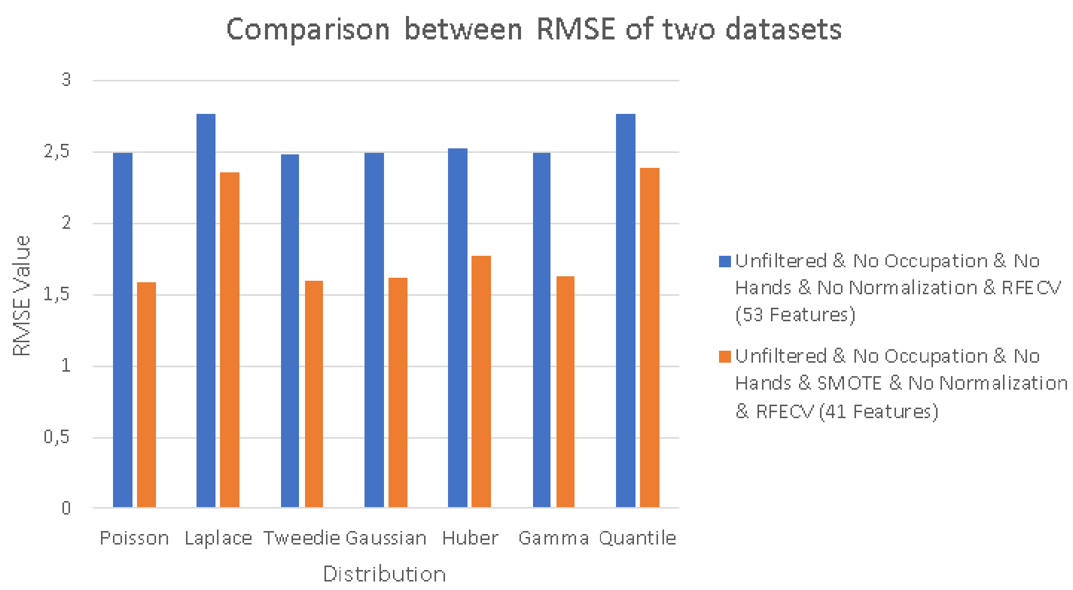

Figure 48: *RMSE* results across distribution

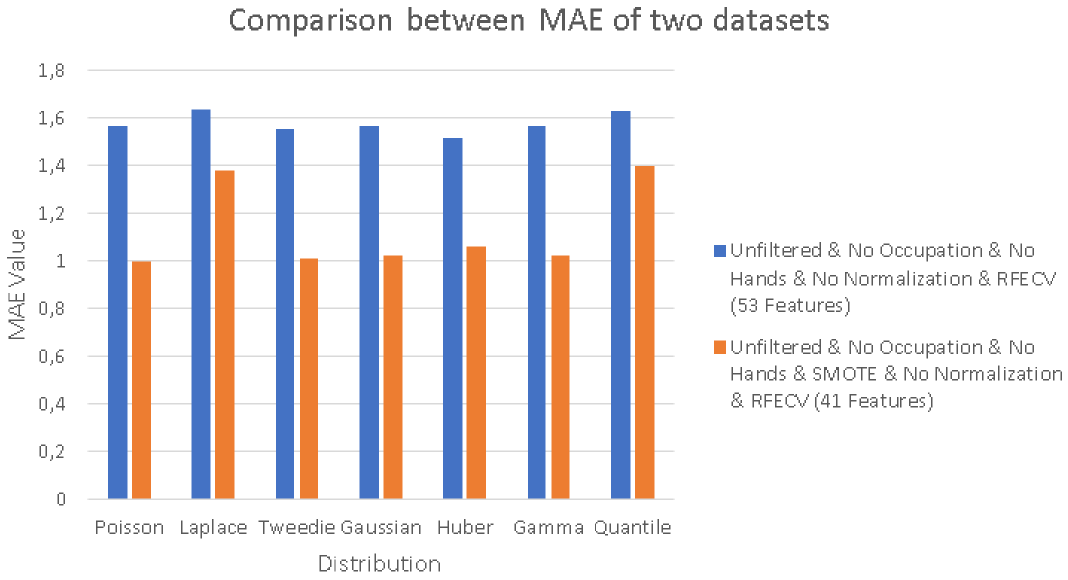

Figure 49: *MAE* results across distribution

As can be seen, the small differences that distinguish these two data sets have significant impacts on the performance of the machine learning model, *GBM*. It is notable that in all cases, the data set with *SMOTE* applied presents the lowest results in both metrics, *Root Mean Squared Error* (*RMSE*) and *Mean Absolute Error* (*MAE*). Additionally, through these graphics, it's possible to retrieve the distribution that outperformed the others, in this case, Tweedie reveals himself as a good option due to presenting the lowest value for the *RMSE*, 1.6002, and the second-lowest for the *MAE*.

With the selection of the best test data set and model distribution, it's now time for a tuning phase, in which two scenarios were created. One of the scenarios is the Grid Search method and the other is the Bayesian Optimization method. These two scenarios were subjected to four optimization phases, referring to the search for the best combination of highly correlated hyperparameters:

1. Hystogram type & Fold Assignment;
2. Number of trees & Learning Rate;
3. Depth of the tree & Min Rows & Min Split Improvement;
4. Sample Rate & Column Sample Rate, per Tree or Level;

Therefore, the following graphics aim to illustrate the impact of each of these phases on the used model.

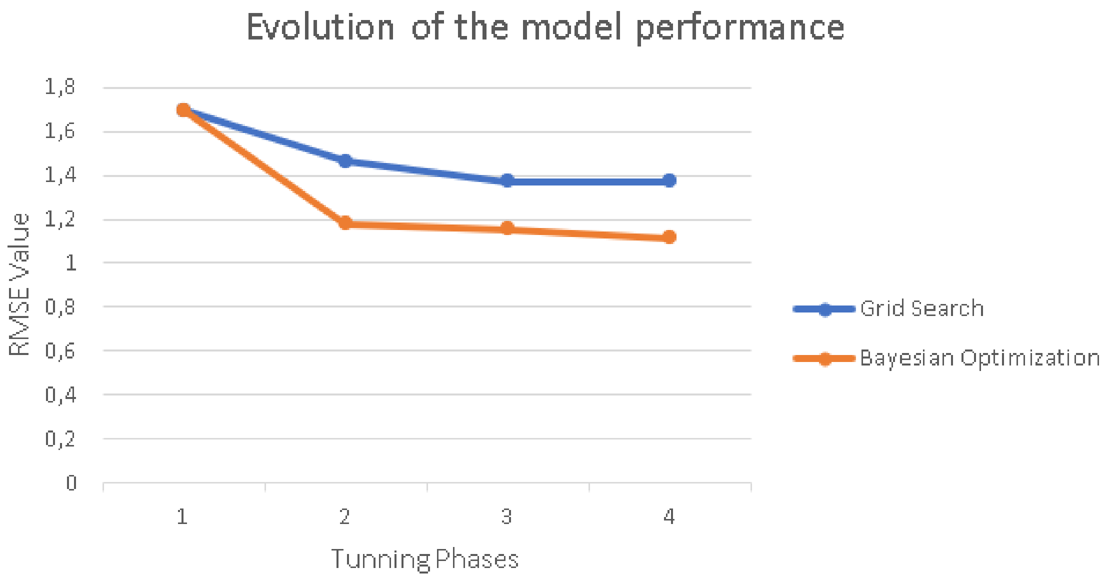

Figure 50: *RMSE* evolution along tuning phases

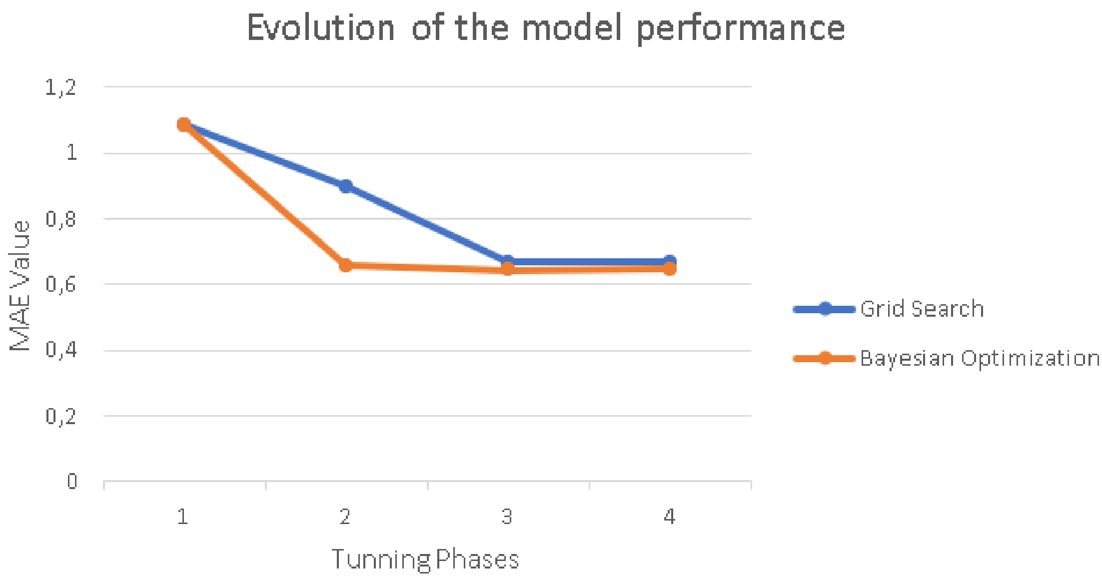

Figure 51: *MAE* evolution along tuning phases

From the graphics illustrated above, it's possible to consider that the evolution in the two metrics is positive, clearly showing the "speed" with which the Bayesian Optimization method needs to reach increasingly smaller *RMSE* values. This is due to the ample space for research, enabling a range of more hyperparameter alternatives. On the other hand, when using the Grid Search method, the initial search space is selected based on assumptions, and depending on its performance, a new search is made in a narrower space, taking into account the best values for both metrics mentioned earlier, as we can verify those differences in the evolution of *MAE* values between the two tuning methods.

5.4.2 *XGBoost*

Following the line of thought demonstrated previously, the same number of tests will be executed, to find the best "base" data set for this type of algorithm. Similarly to the results obtained from the *GBM* algorithm, the best data set presents 43 features, retrieved with *RFECV* method, and the application of the *SMOTE* technique.

As done previously, when performing these tests, a second objective emerges, which is the search for the best distribution to use for this case study and algorithm. In the following graphics, the way these distributions behave can be ascertained.

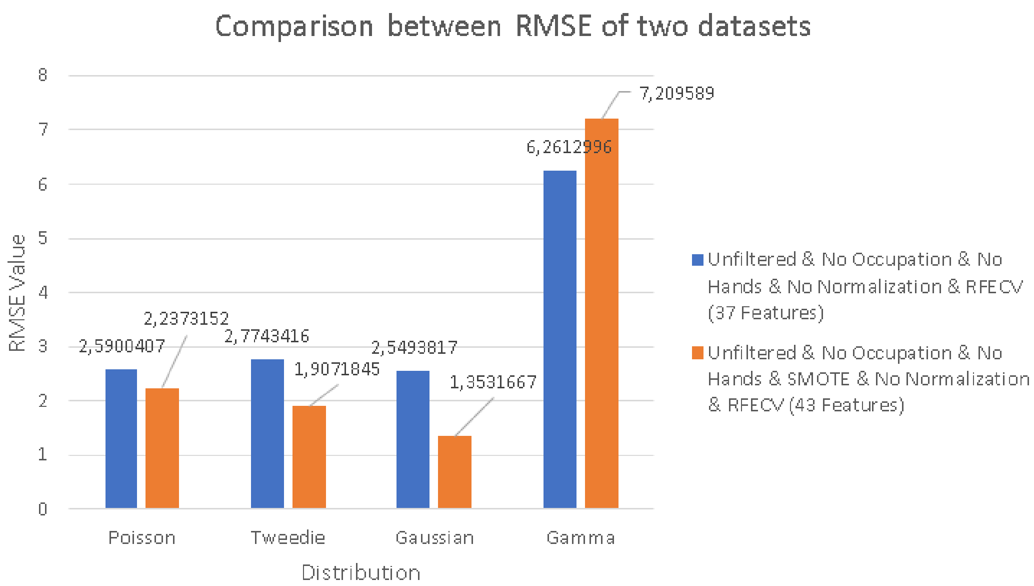

Figure 52: XGBoost *RMSE* results across distribution

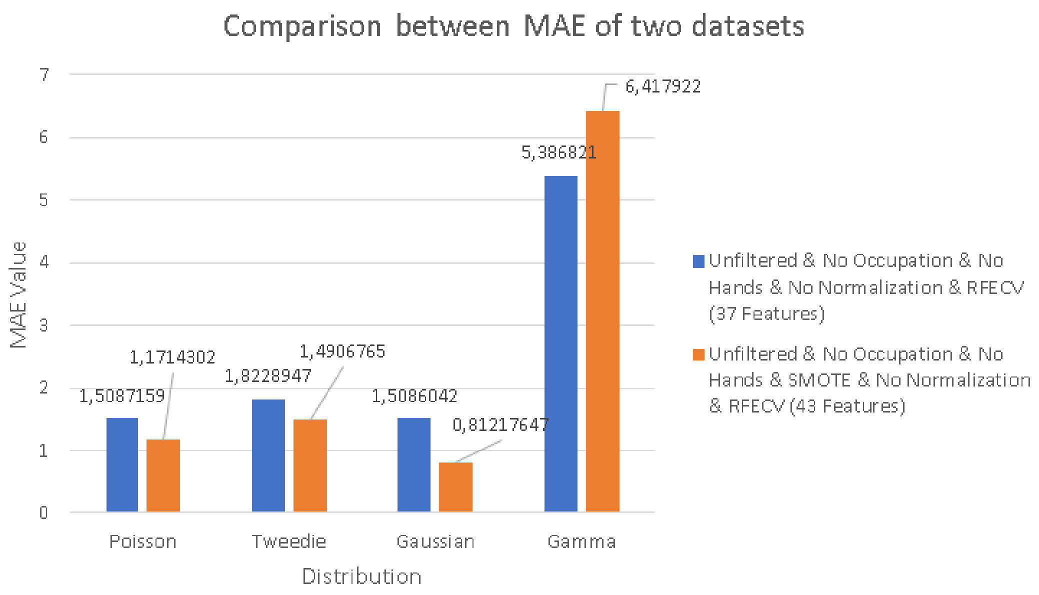

Figure 53: XGBoost *MAE* results across distribution

In relation to this specific case, both *RMSE* and *MAE* reach the lowest value of these tests for a Gaussian distribution. A phenomenon that also happens to the homologous data set, where the *SMOTE* technique was not applied.

As was done for the *GBM* algorithm, the best data set was subjected to the same two tuning scenarios, Grid Search and Bayesian Optimization methods. However, the phases differ compared to those that occurred for the *GBM*, due to the number of hyperparameters available.

1. Fold Assignment & Booster & Tree Method & Grow Policy;
2. Sample type & Normalize type;
3. Number of trees & Learning Rate;
4. Depth of tree & Min Rows & Min Split Improvement;
5. Sample Rate & Column Sample Rate, per Tree;
6. Rate Drop, Skip & One Drop;
7. Reg Lambda, Alpha;

The following graphs aim to show the evolution of the models throughout these phases. Note that for the Bayesian Optimization method, phases six and seven are performed simultaneously.

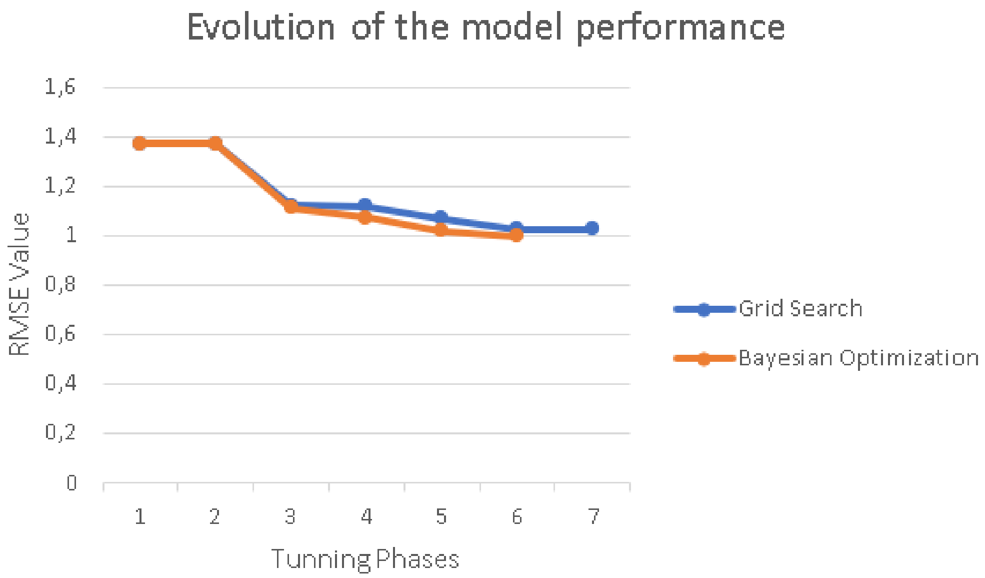

Figure 54: XGBoost *RMSE* evolution along tuning phases

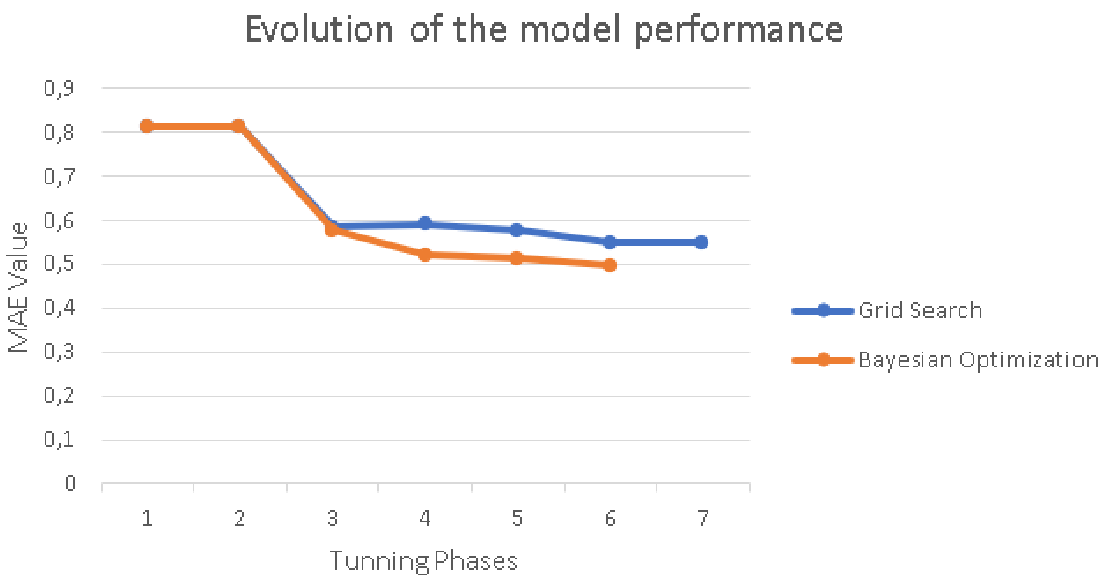

Figure 55: XGBoost *MAE* evolution along tuning phases

From the figure above, it's noticeable a positive evolution concerning the *RMSE*, for both tuning methods. However, the same is not true for the *MAE* values, specifically, for the model optimized with Grid Search, which presents an increase between the third and fourth tuning phase. This may be due to the assumptions used to build the search space. On the other hand, the Bayesian Optimization method presents a positive evolution, achieving better results in both metrics.

5.4.3 *Stacked Ensemble*

After obtaining the best hyperparameters for the different tuning methods and *ML* models, a new scenario was created where the objective is the convergence of the knowledge of these tuned models to improve the results obtained. For this, a Stacked Ensemble was used, which permits to join models in a way to create a better one through already trained models, to which several combinations of the models, built previously, are used as "base models" and tests different combinations for the "meta learner".

Consequently, the following figure comes to demonstrate the impact of each possible "meta learner" option on the best "base models" combination encountered, containing both *XGBoost* models and a Grid Search tuned *GBM* model.

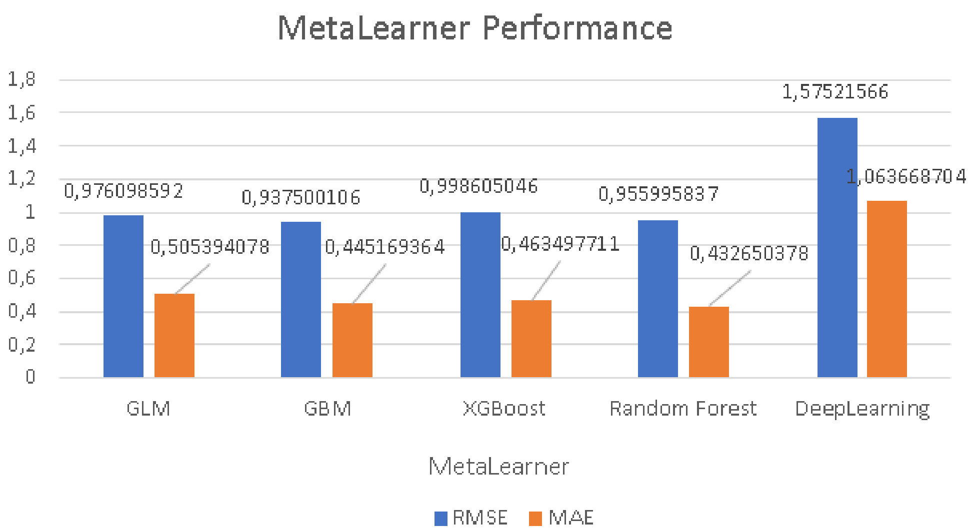

Figure 56: Metalearner performance

Through Figure 56, it's possible to notice that each one of the different possibilities for the "meta learner" presents different impacts, ones more similar to others, however, the *GBM* algorithm has the best results for both metrics, thus choosing this learner.

5.5 RESULT ANALYSIS

With all the models tuned and trained, the focus is now on understanding their actual performance. In this way, it was used a test data set containing 703 records, with ages between 24 and 36. With the models training concluded, all five models were subjected to this new data, standing out the Stacked Ensemble model. The following figure intends to show all the results obtained with this new data for all the developed models.

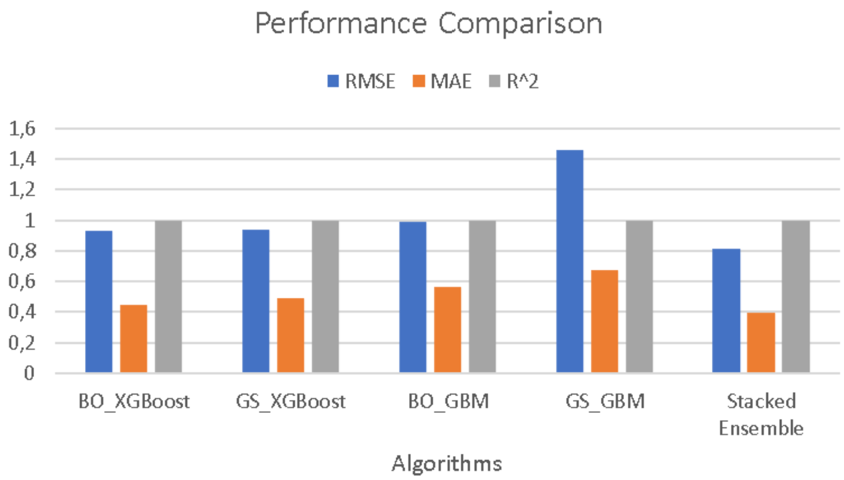

Figure 57: Comparison of results

In Figure 57, it's possible to see the different performances of each of the generated models where it's clear that the *GBM* model optimized through the Grid Search method, defined as 'GS_GBM', presents the worst results compared with the others. Regarding the other models, tuned by the previously mentioned methods, they display a similar performance to each other. Finally, the Stacked Ensemble model, as expected, exhibits the best results in all metrics, achieving this way the objective proposed in Section 5.4.3.

6

RESULTS DISCUSSION

With the end of tests and optimizations proposed for both scenarios mentioned, it's now possible to present a retrospective on the obtained results.

First of all, it is possible to verify that human-computer interactions present themselves as a reliable source of data for the main objective of this dissertation. Then, it was possible to verify that variables that are presented as descriptions of a user's states of fatigue positively influence the *ML* models and, in turn, their results.

Regarding the model, relative to the data captured from a computer, the initial data set showed a clear disparity in terms of age groups along with very large dimensions. Thus, several tests and techniques were carried out to reduce these dimensions without harming the quality of the data set, later an oversampling technique was applied. These aim to balance the data set via synthetic data generated from statistical and clustering methods.

After all the optimizations, the created models were submitted to an unknown data set collected from an Overwolf[1] application. The obtained results (Figure 43), follows the same line of results presented in training (Tables 2 and 3), proving the veracity of the results as well as their ability to capture the different patterns across the different age groups identified. On the other hand, the objective of this dissertation presents itself as something very broad, therefore, the same line of thought was followed, but for mobile environments. The results for this one were truly satisfactory, as can be seen in Figure 57. However, for a better analysis of the results, the following graph was generated.

1 https://www.overwolf.com/

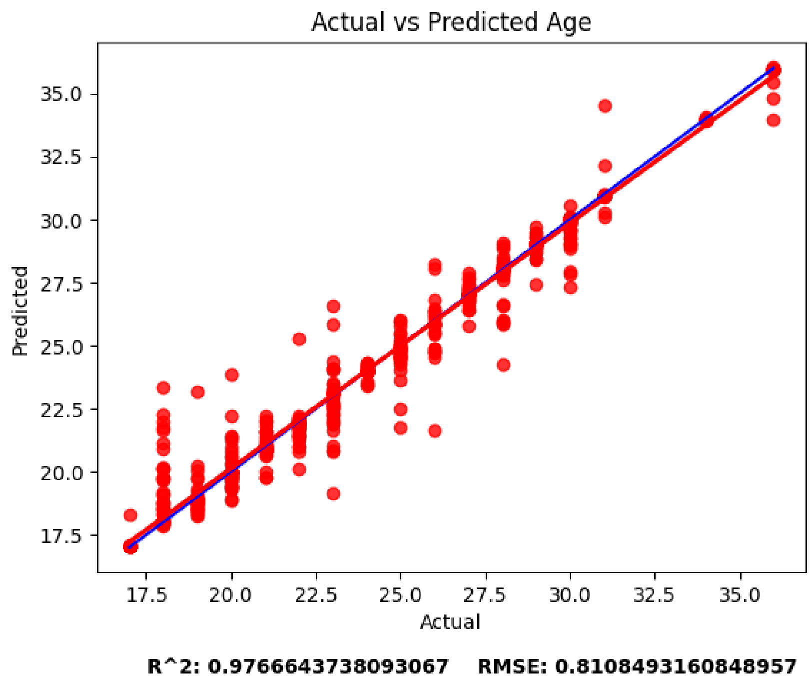

Figure 58: Disposition of prediction errors along each age

From Figure 58, it can be seen, first of all, that the *RMSE* value is around 0.8108. For this case in question, this metric represents the prediction error, meaning that if a model indicates that a given user has an age of 25 years, he can have an age between 24.189 and 25.8108.

Despite the balance obtained from the *SMOTE* technique, the model has certain disparities, for example, there are many records of an 18-year-old user, in which it was indicated that he was over 20 years old as well as other disparities, at other ages. The reason behind these values may come from the amount of original data that existed for that age. In other words, as there is little data, when applying the oversampling technique, the cluster for this age will be very small, thus leading to a greater concentration on the same patterns, which may lead to errors at closer ages.

7

CONCLUSION

Throughout this project, several aspects were addressed, from affective computing, human performance, fatigue and human-computer interactions. From these aspects, it was acquired the knowledge that the physical and psychological state of a person significantly influences their social and motor interactions, as well as their performance in executing any task. So, to obtain a description of these interactions, biometric data from mouse and keyboard events was used, as well as other controllers to capture the type of task to be performed.

When analyzing some of the tasks, it was found that they have different impacts from each other. Some of these tasks have different influences throughout the day, such as the 'net' task, in which the occurrence of fatigue increases over time. Understandably, this type of task may have an evolution in the impact of a user's mental state due to its nature. Since it describes all possible interactions with a web browser, existing the possibility of this type of interaction being more monotonous leads to a greater reluctance to perform the task.

Knowing that age alters the effects of mental fatigue on the dynamics of typewriting, but that in turn, based on these interactions, it´s possible to monitor and analyze mental fatigue (de Jong et al., 2018, 2020; Ribeiro et al., 2014, 2013), the possibility arises of predicting the user age from them. Therefore, it was possible to create *ML* models to effectively predict the user age group based on their interactions with the computer. The use of this type of data proves to be a less intrusive form of age validation, allowing its use for various purposes, from gambling to online shopping. However, the idea behind these models still allows its application to, for example, mobile devices as previously proven (Rocha et al., 2019). Thus, a mobile phone game is used to collect interactions with this type of device, and in turn, to train *ML* models with these data for age prediction.

Finally, it was possible to develop an endpoint that consecutively uses the *ML* model generated from interactions with the computer, enabling its integration on multiple platforms.

The results obtained indicate that the knowledge of fatigue levels present in interactions with a computer is an asset, given that these appear as a factor with a high impact on several metrics and, in turn, human performance. Furthermore, the results also indicate that *AI* can be used to predict this range in future cases.

7.1 APPLICATION AREAS

In an analysis of the existing forms of age verification, it can be seen that they are either easily circumventable or very invasive and so, an area of an application would be to guarantee the age requirements that does not cause discomfort to a user.

Therefore, the proposed system can be applied to web pages that contain inappropriate content (e.g. pornographic sites), gambling sites (e.g. Bet[1], Betano[2], etc.) or purchase sites (e.g. alcohol, tobacco, etc.). These pages will be able to guarantee these validations through the use of a library but the possible applications are not restricted to web pages. It can also be used in-game distribution applications such as Steam[3], Ubisoft Connect[4], Epic Games Launcher[5], where the validations would be provided by an *SDK*.

Finally, there is still the possibility of its use in systems designed for detecting cognitive decline, which follows the same line of thought and can be ensured either by using a library or *SDK*.

7.2 FUTURE WORK

The proposed objectives for this dissertation were completed. However, there is the possibility of carrying out new tests/experiments and thus, take better advantage of the obtained data set.

1 https://www.bet.pt/apostas-desportivas/
2 https://www.betano.pt/
3 https://store.steampowered.com/
4 https://ubisoftconnect.com/pt-PT/
5 https://www.epicgames.com/store/pt-BR/

Regarding the prediction of an age group, a possible example for optimization would be, in the first phase, the design of a binomial model that predicts whether a given user is of legal age or not. This result would be taken into account in the main model. The idealization of this type of optimization emerged as a means of reducing erroneous predictions, especially in age groups that contain people of near legal age and people of legal age. Another example could be a more intensive search regarding the features and, from these, generate new ones, enabling the characterization of certain states.

As the last suggestion for future work, concerning forecasting via interactions with mobile devices, it would be interesting to obtain a more comprehensive dataset in terms of ages, i.e. a dataset with several records of different ages, both minors and of legal age.

BIBLIOGRAPHY

M. A. M. Abu Shariah, R. N. Ainon, R. Zainuddin, and O. O. Khalifa. Human computer interaction using isolated-words speech recognition technology. In *2007 International Conference on Intelligent and Advanced Systems*, pages 1173–1178, 2007. doi: 10.1109/ICIAS.2007.4658569.

F. C. Bartlett. Fatigue Following Highly Skilled Work. *Nature*, 147(3736):717–718, 06 1941. ISSN 0028-0836, 1476-4687. doi: 10.1038/147717a0.

Frederic Bartlett. *Psychological criteria of fatigue*, volume 119. 1953. doi: 10.1126/science.119.3095.564.b. URL https://www.science.org/doi/10.1126/science.119.3095.564.b.

Maarten A.S. Boksem and Mattie Tops. Mental fatigue: Costs and benefits. *Brain Research Reviews*, 59(1):125–139, 11 2008. ISSN 01650173. doi: 10.1016/j.brainresrev.2008.07.001.

Maarten A.S. Boksem, Theo F. Meijman, and Monicque M. Lorist. Mental fatigue, motivation and action monitoring. *Biological Psychology*, 72(2):123–132, 05 2006. ISSN 03010511. doi: 10.1016/j.biopsycho.2005.08.007.

Ronald Boring. Human-computer interaction as cognitive science. *Proceedings of the Human Factors and Ergonomics Society*, 46:1767–1771, 09 2002. doi: 10.1177/154193120204602103. URL https://www.researchgate.net/publication/229068297_Human-Computer_Interaction_as_Cognitive_Science.

John A. Caldwell, J. Lynn Caldwell, Lauren A. Thompson, and Harris R. Lieberman. Fatigue and its management in the workplace. *Neuroscience & Biobehavioral Reviews*, 96:272–289, 01 2019. ISSN 01497634. doi: 10.1016/j.neubiorev.2018.10.024. URL https://linkinghub.elsevier.com/retrieve/pii/S0149763418305220.

E. Cambria. Affective computing and sentiment analysis. *IEEE Intelligent Systems*, 31(2):102–107, 2016. doi: 10.1109/MIS.2016.31.

Brian Clark Howard. How microsoft's new age detection software works (and doesn't), 2015. URL https://www.nationalgeographic.com/news/2015/05/150501-how-old-microsoft-face-aging-software-science/.

Marlon de Jong, Jacob Jolij, André Pimenta, and Monicque M. Lorist. Age modulates the effects of mental fatigue on typewriting. *Frontiers in Psychology*, 9:1113, 2018. ISSN 1664-1078. doi: 10.3389/fpsyg.2018.01113.

Marlon de Jong, Anne M. Bonvanie, Jacob Jolij, and Monicque M. Lorist. Dynamics in typewriting performance reflect mental fatigue during real-life office work. *PLOS ONE*, 15(10):1–15, 10 2020. doi: 10.1371/journal.pone.0239984.

IBM Cloud Education, 2020. URL https://www.ibm.com/cloud/learn/machine-learning.

M. Erbilek, M. Fairhurst, and M. Da Costa-Abreu. Improved age prediction from biometric data using multimodal configurations. In *2014 International Conference of the Biometrics Special Interest Group (BIOSIG)*, pages 1–7, 2014. URL https://ieeexplore.ieee.org/document/7029422.

Léon G. Faber, Natasha M. Maurits, and Monicque M. Lorist. Mental Fatigue Affects Visual Selective Attention. *PLoS ONE*, 7(10):e48073, 10 2012. ISSN 1932-6203. doi: 10.1371/journal.pone.0048073.

Christopher Fanchi. What is api as a service?, 2019. URL https://backendless.com/what-is-api-as-a-service/.

P. Farley, J. Wells, T. Christiani, and P. Thurman, 2019. URL https://docs.microsoft.com/en-us/azure/cognitive-services/face/concepts/face-detection.

W.B. Gartner and M.R. Murphy. Pilot workload and fatigue: A critical survey of concepts and assessment techniques. *Applied Ergonomics*, 8(4):238, 1977. ISSN 0003-6870. doi: https://doi.org/10.1016/0003-6870(77)90183-1.

R. K. Gowda, V. Nimbalker, R. Lavanya, S. Lalitha, and S. Tripathi. Affective computing using speech processing for call centre applications. In *2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, pages 766–771, 2017. doi: 10.1109/ICACCI.2017.8125934.

Jan Gulliksen. Institutionalizing human-computer interaction for global health. *Global Health Action*, 10(sup3):1344003, 2017. doi: 10.1080/16549716.2017.1344003. URL https://doi.org/10.1080/16549716.2017.1344003. PMID: 28838309.

S. Haider, A. Abbas, and A. K. Zaidi. A multi-technique approach for user identification through keystroke dynamics. In *Smc 2000 conference proceedings. 2000 ieee international conference on systems, man and cybernetics. 'cybernetics evolving to systems, humans, organizations, and their complex interactions' (cat. no.0)*, volume 2, pages 1336–1341 vol.2, 2000. doi: 10.1109/ICSMC.2000.886039.

L. Han and X. Li. The appliance of affective computing in man-machine dialogue: Assisting therapy of having autism. In *2014 Fourth International Conference on Communication Systems and Network Technologies*, pages 1093–1096, 2014. doi: 10.1109/CSNT.2014.223.

Robert J. and G. Hockey. Compensatory control in the regulation of human performance under stress and high workload: A cognitive-energetical framework. *Biological Psychology*, 45(1-3):73–93, 03 1997. ISSN 03010511. doi: 10.1016/S0301-0511(96)05223-4.

A. K. Jain, A. Ross, and S. Prabhakar. An introduction to biometric recognition. *IEEE Transactions on Circuits and Systems for Video Technology*, 14(1):4–20, 2004. doi: 10.1109/TCSVT.2003.818349.

Ruth Kanfer and Phillip L. Ackerman. Motivation and cognitive abilities: An integrative/aptitude-treatment interaction approach to skill acquisition. *Journal of Applied Psychology*, 74(4):657–690, 1989. ISSN 1939-1854, 0021-9010. doi: 10.1037/0021-9010.74.4.657.

M. Karnan, M. Akila, and N. Krishnaraj. Biometric personal authentication using keystroke dynamics: A review. *Applied Soft Computing*, 11(2):1565–1573, 03 2011. ISSN 15684946. doi: 10.1016/j.asoc.2010.08.003. URL https://linkinghub.elsevier.com/retrieve/pii/S156849461000205X.

A. Kołakowska. A review of emotion recognition methods based on keystroke dynamics and mouse movements. In *2013 6th International Conference on Human System Interactions (HSI)*, pages 548–555, 2013. doi: 10.1109/HSI.2013.6577879.

Sang Kyu Kwak and Jong Hae Kim. Statistical data preparation: management of missing values and outliers. *Korean Journal of Anesthesiology*, 70(4):407, 2017. ISSN 2005-6419, 2005-7563. doi: 10.4097/kjae.2017.70.4.407. URL http://ekja.org/journal/view.php?doi=10.4097/kjae.2017.70.4.407.

Chris Kyriacou. Action research: a methodology for change and development - by bridget somekh. *British Journal of Educational Studies*, 55(4):468–469, 2007. doi: 10.1111/j.1467-8527.2007.00388_3.x.

Daniel Levine. Apis are the next saas wave, 2019. URL https://techcrunch.com/2019/09/06/apis-are-the-next-big-saas-wave/.

N. Li, T. Zhou, Y. Zhou, C. Guo, D. Fu, X. Li, and Z. Guo. Research on human-computer interaction mode of speech recognition based on environment elements of command and control system. In *2019 5th International Conference on Big Data and Information Analytics (BigDIA)*, pages 170–175, 2019. doi: 10.1109/BigDIA.2019.8802812.

Niels Lobo and Young Kwon. Automatic feature detection and age classification of human faces in digital images, 07 1998. URL https://patents.google.com/patent/US5781650A/en.

Christopher McClernon and James Miller. Variance as a measure of performance in an aviation context. *The International Journal of Aviation Psychology*, 21:397–412, 10 2011. doi: 10.1080/10508414.2011.606765.

James C. Miller. Fundamentals of Shiftwork Scheduling. *Ergonomics in Design: The Quarterly of Human Factors Applications*, 16(3):13–17, 07 2008. ISSN 1064-8046. doi: 10.1518/106480408X333109.

P. C. Mondal, R. Deb, and M. N. Huda. Know your customer (kyc) based authentication method for financial services through the internet. In *2016 19th International Conference on Computer and Information Technology (ICCIT)*, pages 535–540, 2016. doi: 10.1109/ICCITECHN.2016.7860255.

B. Muscio. IS A FATIGUE TEST POSSIBLE? *British Journal of Psychology. General Section*, 12 (1):31–46, 06 1921. ISSN 03732460. doi: 10.1111/j.2044-8295.1921.tb00036.x.

D. Nath and S. K. Kalita. An effective age detection method based on short time energy and zero crossing rate. In *2014 2nd International Conference on Business and Information Management (ICBIM)*, pages 99–103, 2014. doi: 10.1109/ICBIM.2014.6970942.

D.C. 20594 National Transportation Safety Board Bureau of Accident Investigation Washington. *China Airlines Boeing 747-SP, N4522V, 300 Nautical Miles Northwest of San Francisco, California, February 19, 1985*. 1986. URL https://www.ntsb.gov/investigations/AccidentReports/Reports/AAR8603.pdf.

Lene Nielsen. *Personas In "The Encyclopedia of Human-Computer Interaction, 2nd Ed."*. 2013. URL http://www.interaction-design.org/encyclopedia/personas.html.

Donald Norman. The 'problem' with automation: Inappropriate feedback and interaction, not 'over-automation'. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 327:585–93, 05 1990. doi: 10.1098/rstb.1990.0101.

Soumya Priyadarsini Panda. Automated speech recognition system in advancement of human-computer interaction. In *2017 International Conference on Computing Methodologies and Communication (ICCMC)*, pages 302–306, 2017. doi: 10.1109/ICCMC.2017.8282696.

L.P. Perelli. *Fatigue stressors in simulated long-duration flight. Effects on performance, information processing, subjective fatigue, and physiological cost*, volume 186. 1980. URL http://onlinelibrary.wiley.com/o/cochrane/clcentral/articles/734/CN-00179734/frame.html.

Rosalind W. Picard. *Affective computing*. MIT Press, 1997. ISBN 9780262161701. URL https://affect.media.mit.edu/pdfs/95.picard.pdf.

André Ribeiro, Davide Carneiro, Paulo Novais, and Jose Neves. Monitoring mental fatigue through the analysis of keyboard and mouse interaction patterns. In *Hybrid Artificial Intelligent Systems*, pages 222–231, 09 2013. ISBN 9783642408458. doi: 10.1007/978-3-642-40846-5_23.

André Ribeiro, Davide Carneiro, Paulo Novais, and Jose Neves. Analysis of human performance as a measure of mental fatigue. In *Hybrid Artificial Intelligence Systems*, pages 389–401. Springer International Publishing, 2014. ISBN 978-3-319-07617-1. doi: 10.1007/978-3-319-07617-1_35.

Rodrigo Rocha, Davide Carneiro, and Paulo Novais. The influence of age and gender in the interaction with touch screens. In Paulo Moura Oliveira, Paulo Novais, and Luís Paulo Reis, editors, *Progress in Artificial Intelligence*, pages 3–12. Springer International Publishing, 2019. ISBN 978-3-030-30244-3. doi: 10.1007/978-3-030-30244-3_1.

Emilie Roth, Emily Patterson, Randall Mumaw, E. Mumaw, and John Sons. Cognitive engineering: Issues in user-centered system design. 09 2000. URL https://www.researchgate.net/publication/2801690_Cognitive_Engineering_Issues_in_User-Centered_System_Design.

Sherwood Samn and Layne Perelli. Estimating aircrew fatigue: A technique with application to airlift operations. page 29, 12 1982. URL https://www.researchgate.net/publication/235150215_Estimating_Aircrew_Fatigue_A_Technique_with_Application_to_Airlift_Operations.

S. Saxena, P. Saxena, and S. K. Dubey. Various levels of human stress their impact on human computer interaction. In *2013 International Conference on Human Computer Interactions (ICHCI)*, pages 1–6, 2013. doi: 10.1109/ICHCI-IEEE.2013.6887808.

M. Siar and M. Teshnehlab. Age detection from brain mri images using the deep learning. In *2019 9th International Conference on Computer and Knowledge Engineering (ICCKE)*, pages 369–374, 2019. doi: 10.1109/ICCKE48569.2019.8964911.

Jeffrey J. Starns and Roger Ratcliff. The effects of aging on the speed–accuracy compromise: Boundary optimality in the diffusion model. *Psychology and Aging*, 25(2):377–390, 2010. ISSN 1939-1498, 0882-7974. doi: 10.1037/a0018022.

J. Tam and C. H. Martell. Age detection in chat. In *2009 IEEE International Conference on Semantic Computing*, pages 33–39, 2009. doi: 10.1109/ICSC.2009.37.

Edward Thorndike. Mental fatigue. I. *Psychological Review*, 7(6):547–579, 1900. ISSN 0033-295X. doi: 10.1037/h0069511. URL http://content.apa.org/journals/rev/7/6/547.

European Union. Regulation (eu) 2016/679, 2016. URL https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016R0679&from=BG.

Dimitri van der Linden and Paul Eling. Mental fatigue disturbs local processing more than global processing. *Psychological Research Psychologische Forschung*, 70(5):395–402, 09 2006. ISSN 0340-0727, 1430-2772. doi: 10.1007/s00426-005-0228-7.

Dimitri van der Linden, Michael Frese, and Theo F Meijman. Mental fatigue and the control of cognitive processes: effects on perseveration and planning. *Acta Psychologica*, 113(1): 45 – 65, 2003. ISSN 0001-6918. doi: https://doi.org/10.1016/S0001-6918(02)00150-6.

Paul Verhaeghen and Timothy A. Salthouse. Meta-analyses of age–cognition relations in adulthood: Estimates of linear and nonlinear age effects and structural models. *Psychological Bulletin*, 122(3):231–249, 11 1997. ISSN 1939-1455, 0033-2909. doi: 10.1037/0033-2909.122.3.231.

Wei Shen and Xiaolei Zhou. Research on the human-computer interaction mode designed for elderly users. In *2015 4th International Conference on Computer Science and Network Technology (ICCSNT)*, volume 01, pages 374–377, 2015. doi: 10.1109/ICCSNT.2015.7490772.

Richard J. Williamson, Shaun Purcell, Abram Sterne, Simon Wessely, Matthew Hotopf, Anne Farmer, and Pak C. Sham. The relationship of fatigue to mental and physical health in a community sample. *Social Psychiatry and Psychiatric Epidemiology*, 40(2):126–132, 02 2005. ISSN 0933-7954, 1433-9285. doi: 10.1007/s00127-005-0858-5.

J. Yang, Y. Xue, Z. Zeng, and W. Guo. Research on multimodal affective computing oriented to online collaborative learning. In *2019 IEEE 19th International Conference on Advanced Learning Technologies (ICALT)*, volume 2161-377X, pages 137–139, 2019. doi: 10.1109/ICALT.2019.00045.

C. Zucco, B. Calabrese, and M. Cannataro. Sentiment analysis and affective computing for depression monitoring. In *2017 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pages 1988–1995, 2017. doi: 10.1109/BIBM.2017.8217966.

Torbjörn Åkerstedt, Andrea Discacciati, Anna Miley-Åkerstedt, and Hugo Westerlund. Aging and the Change in Fatigue and Sleep – A Longitudinal Study Across 8 Years in Three Age Groups. *Frontiers in Psychology*, 9:234, 03 2018. ISSN 1664-1078. doi: 10.3389/fpsyg.2018.00234. URL http://journal.frontiersin.org/article/10.3389/fpsyg.2018.00234/full.