Mobile Sensing for Emotion Recognition

Piskioulis Orestis

UNIVERSITY CENTER OF INTERNATIONAL PROGRAMMES OF STUDIES

SCHOOL OF SCIENCE AND TECHNOLOGY

A thesis submitted for the degree of

Master of Science (MSc) in E-Business & Digital Marketing

January 2021

Thessaloniki-Greece

Student Name:	Orestis Piskioulis
SID:	254316824873
Supervisor:	Prof. Katerina Tzafilkou

I hereby declare that the work submitted is mine and that where I have made use of another's work, I have attributed the source(s) according to the Regulations set in the Student's Handbook.

January 2021

Thessaloniki-Greece

Abstract

This work concerns the classification of emotions based on data from smartphones. The data are gathered from people who play our game parallel to the sensor recording app on their phones. Although other emotion detection methods are generally more intrusive or estimate for longer durations, accelerometer data are an opportunity to detect emotions both in real time and non-intrusive. An Android application was developed to track the accelerometers of the people who attended the study and ask them to present their emotionality. This collects the data from a natural instead of a laboratory environment. The recorded information is then processed and used to compare various classifiers. The machine learning algorithms decision tree, support vector machine, neural networks and linear and logistics regression are used for this purpose.

The prediction of two basic emotions, enjoyment, and frustration was investigated. While the recognition rate of predicting these two emotions is high in the current bibliography, we succeeded to achieve interesting prediction rates, combining our tons of data and machine learning evaluation models. There are several ways to predict an individual's state of activation. Accelerometer and gyroscope are high predictors of user emitters during gameplay activities. In addition, good predictions have been made using camera, microphone, but also camera and accelerometer combinations. Such projections can be used in combination with other emotional methods to adjust services to the customer's emotional state.

Without constant supervision by Dr Katerina Tzafilkou, the supervisor who have made the greatest contribution in this study, this effort could not have been accomplished.

Keywords: Mobile sensing; Mobile emotion sensing Emotion Detection; Smartphone Sensing; Mobile learning emotion recognition;

Piskioulis Orestis

04/01/2021

Table of Contents

List of Tables

1. Introduction

In recent years, mood sensing has received substantial attention from social psychology, neuroscience, and computer science (Fowler & Christakis, 2009). The state of mood or emotion plays a significant role in everyday human life, and has a great impact on the communication, perception, social behavior, and decision making of people (Likamwa et al., 2013; Y. Zhang et al., 2010).

A difficult task of our era is the automated mood identification, which envisions a wide variety of new mood-conscious application scenarios. For example, individuals enjoy different music and film styles, which rely not only on their taste, but also on their mood and personality.

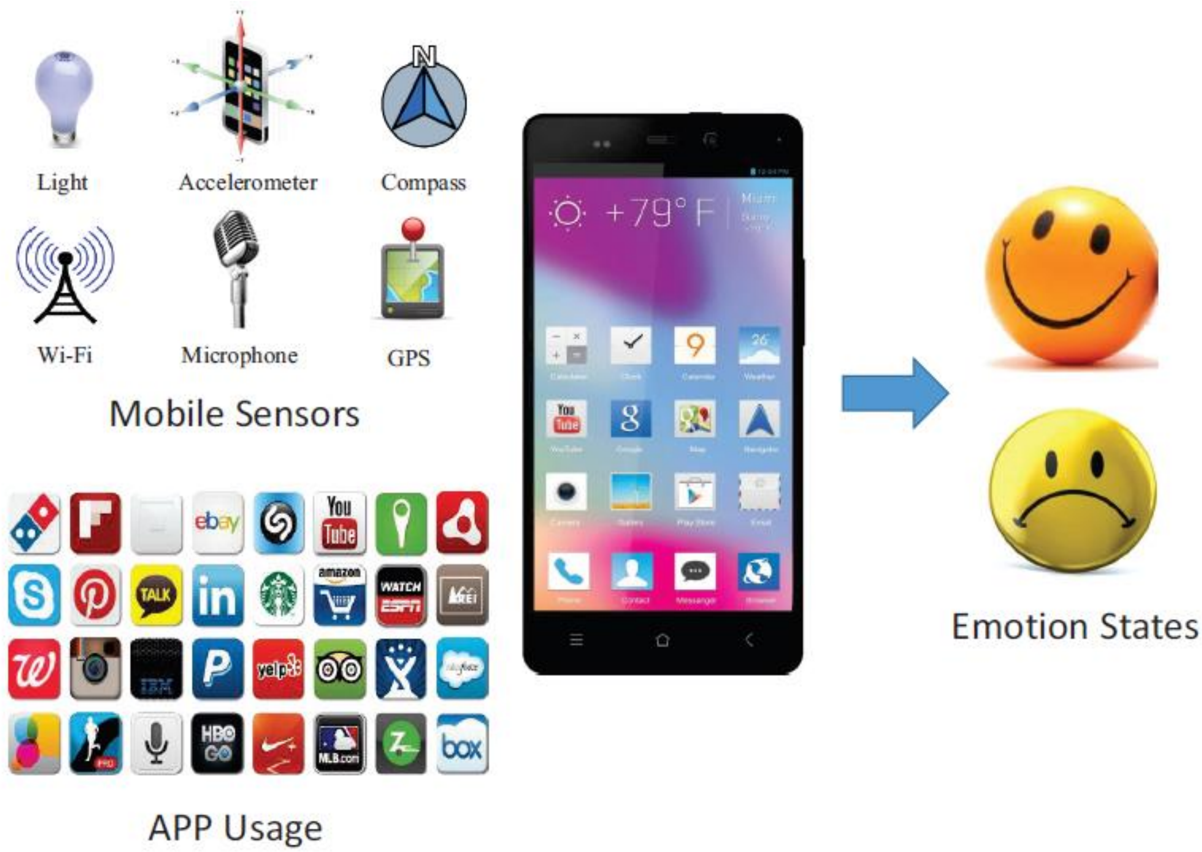


Figure 1: Mobile Sensors, App Usage and Emotion States

The mood to improve user engagement should also be considered by a smart recommendation framework. Another great example is advertising, which is fascinating or irritating for different individuals with different states of emotion. Advertising will also be more effective and customized if it is possible to understand the emotions of people. Sharing the mood between family and close friends will allow individuals to reinforce their connections and boost the way of social communication with the prevalence of social network apps.

Furthermore, robots will be commonly used in different areas of our lives soon. If they can ''read the mood'' of the person for whom they work, the robots can be more intelligent and humanized.

Moreover, to assess the psychological health and mental well-being of individuals, it is crucial to understand the emotion condition and its evolution (Diener et al., 2003; Gross & John, 2003).

With the use of body physiological signals such as heart rhythm, blood pressure, breath rate, mood detection was recorded (Gluhak et al., 2007; Leng et al., 2007; M. Zhao et al., 2018). By monitoring such signals, however, relies on costly dedicated computers, which are infeasible for device-free omnipresent detection. Some current studies have tried to classify emotions through audio and video signals (Ang et al., 2002).

For example, Ang explored speech-based recognition of two negative emotions, annoyance, and frustration (Zeng et al., 2009). Ashraf detected the expression of pain by facial signal recognition. The MoodMeter suggested recognizing the smiling face through video cameras on campus (Hernandez et al., 2012). Visual feature recognition, however, only represents the gestures of people in a snapshot. In addition, audio and video data collection and analysis is a highly computational task, and without large deployment of cameras, such signals cannot be captured everywhere.

Smartphones are now commonly used in industry for business, social and entertainment purposes in people’s everyday life (Bao et al., 2013; Likamwa et al., 2013). As we can see in the image below, there are many sensors embedded:

Microphone, accelerometer, electronic compass, GPS, proximity, etc. The information obtained from the sensors has profound information and can be used to evaluate the social habits of the user, such as physical activity, social communications, location, etc. Intuitively, mobile use and background knowledge are associated with the emotions of users. For instance, when they are happy, people may play mobile phone games, or they may feel depressed if they live in a noisy environment. Several works were inspired by this, using smartphone sensing data for emotion detection.

A multifactorial statistical model to understand everyday stress by comprehensively analyzing cell phone data and weather conditions has been proposed by Bogomolov (Bogomolov et al., 2014) .For cold-start emotion prediction using transfer learning, Sun used sensor data, APP usage information, and SMS content.(Sun et al., 2017)

Their approach, however, is content-based, which highly involves privacy-sensitive data from users, such as SMS content. LiKamWa suggested a mood detection system using the period of email, SMS, venue, and app use as features (Du et al., 2014). What needs to be emphasized is that only one emotion identification is considered in the literature, which means that people in a time are in only one emotion state. We weaken those assumptions, unlike existing works, and research the co-existence of multiple emotions called compound emotions.

According to the theory of Plutchik, (PLUTCHIK, 1980) emotion is not required in a pure state and may be a mixture of fundamental emotions. Compound facial feelings in human facial expressions were also seen in the analysis of (Likamwa et al., 2013; Zhou et al., 2015). For example, a compound emotional expression that incorporates basic feelings of happiness and surprise is "happily surprised".

1.1 Research objectives

The following research aims to enrich the existing literature on the prediction of human emotions in mobile users. In this context, an attempt was made to study the provocation of two basic human emotions, enjoyment, and frustration. Based on an improvised psychometric game, we tried to provoke some of these emotions to study them in more detail. In fact, with the help of appropriate technology, we measured through the senses of mobile devices the general behavior of use during the game to quantify this data and process it to obtain a result that will answer our research questions.

Furthermore, we should clarify that in this research with the use of an app that is called Sensor Record, we track and measure the whole behavior our users when they play with our game. To this, we make use of two accelerometers, the classic and the linear and we also use the gyroscope to give us a clear and safe image about the whole interaction of our users. With the accelerometer we practically measure the linear acceleration based on vibration. In comparison, with gyroscope we determine an angular position based on the principle of rigidity of space.

In this research, it is very crucial to categorize users based on some of their characteristics, e.g., based on how much time they spend playing games per week, how familiar they are with technology, what kind of game they play, what emotions they are provoked by the use of a new application or technology. This kind of categorization will lead us to hit a user profile and be able to study them in more depth. Besides, it has been observed that there are some timeless correlations when we talk about game users and the behaviors that they tend to risk. Thus, we should spend time to tailor our data so that is compatible with that.

The present research process is aimed at the average mobile phone user. Through the rapid use of mobile phones, we have concluded that all mobile phone users are capable of being a sample of our research. That is why we did not focus on a specific segment of the population. We tried to simply target people who were willing to contribute to this whole research process and we tried to get them interested so that they would be motivated to play and fill out the questionnaire with due care. This can be said to have been achieved since we did not receive a wrong answer to the concentration question, which leads us to believe that the users of our sample received due attention. After all, familiarity or not with psychometric games is not a factor of our research.

To the final processing of the results of our research, we try to exact knowledge from our data. To this, we process our sensor data into ML algorithm and try to build a model which it can predict the two basic emotions, enjoyment, and frustration, that we wanted to investigate on that research effort. We ended up defining some benchmarks in our research because of comparing specific user features.

2. Bibliography

2.1 Affective Computing

Affective computing is the research and development of technologies and systems capable of detecting, reading, processing, and simulating human impacts. It is an interdisciplinary field spanning computer science, cognitive science, and psychology. Although the origins of the field can be traced back to early metaphysical studies into emotion ("affect" is simply a synonym for "emotion.") (Jaques et al., 2015).

Affective computing technology senses the user's emotional state (via sensors, microphones, cameras, software) and reacts with real, pre-defined features of products / services, including changing a contest or suggesting a collection of videos to match the learner 's mood. The more computers we have in our lives, the more politically and socially wise we want them to be. We don't want unimportant information to bother us. That kind of common-sense reasoning requires an understanding of the person's emotional state (Jaques et al., 2015).

Emotional intelligence envisions a more 'human-oriented' technology world by providing computers with emotion-awareness and enabling organizations and applications to highly personalize to highly personalize their services. To this, affective computing has introduced a range of instruments and methodologies from behavioral and physiological data to diagnose the affective states of the user. Stepping on the relationship between cell phone sensor knowledge and emotional status (Wang et al., 2020).

This 'privilege' is introduced to smart mobile devices by the field of Mobile Sensing by enabling them to identify 'hidden' affective states within the contact behaviors of users. For the compilation of emotional responses without interrupting users, mobile devices are considered acceptable platforms (Shu et al., 2019).

To date, much of Affective Computing 's research work has focused primarily on unimodal or multimodal emotion recognition using text, picture, and speech data(Poria et al., 2017). The implementation of machine learning and deep learning algorithms for supervised visual, textual, and aural information learning has shown interesting results, while Convolutionary Neural Networks (CNNs) have been shown to be very successful (showing the highest accuracy ratings) for visual inputs, e.g., in face-tracking studies. In addition, according to recent studies (Kanjo et al., 2019), when large numbers of sensor inputs are used, the adoption of deep learning methods in the classification of human emotions is more successful than conventional machine learning.

Unfortunately, in customized e-learning systems emotions are the most overlooked aspect and most of those emotionally observable E-learning systems use overbearing methods to detect them (Wahid & Rasheed, 2019). Student and instructor research have shown clear connections between learning and emotion that are also expressed in formal and informal mobile learning environments. To provide affective recommendation, adaptation, or personalization services, it is useful to predict or diagnose emotions associated with learning. (Ashwin & Guddeti, 2020) have recently developed and suggested a method for investigating intervention using the affective states of students. In its framework, the authors predict the affective states of the unobtrusive and multi-modal students using facial expressions, hand movements and body postures.

2.2 Models for emotion and mood measurement

Mobile Sensing is a common technology that in recent years has been attracting research. The productivity plateau is anticipated to be reached in around two or five years, according to Gartner's hype period of ''Emotion Detection/Recognition''(Goertz, 2018) .

There has been detailed research of emotion and mood in psychology, sociology, and neuroscience (Posner et al., 2005). In general, "emotion" refers to the present instantaneous feeling, and "mood" refers to the longer-term normal feelings (X. Zhang et al., 2018).

Such models can be used to quantify both instantaneous and long-term sensations, and they can be used to quantify emotion and mood without limiting differentiation. The Positive and Negative Affect Schedule (PANAS) model is a widely used indicator for the general affective states (Crawford & Henry, 2004).

Participants completing the PANAS are asked to rate the degree to which each of the 20 emotions they encountered on a 5-point Likert scale ranging from "very slightly" to "very strongly". Participants can be asked how they feel right now or over longer periods of time (e.g., over the past month), depending on the intent of emotion or mood assessment. The discrete model of the category represented emotion by a collection of categories.

Figure 2: The 6 basic emotions of Ekman

The discrete model of the category represented emotion by a collection of categories. Ekman's six basic categories are among the most common models (Ekman et al., 1987): happiness, sadness, anger, surprise, fear, disgust (Ekman, 1992). The Ekman's model is intuitive and understandable to average users, and it allows multiple emotions to coexist with various intense levels. The

Ekman’s emotion model has been widely adopted by many studies with its high applicability (Reisenzein et al., 2013; Zhou et al., 2015).

Tomkins compiled a further list of emotions and later added them to the list of 15 emotions that culminated in 9 fundamental emotions of interest, joy, surprise, sorrow, fear, guilt, disdain, rage, and disgust. The List of Tomkins provides a history to more recent studies on the relationship between emotions and Lövheim neurotransmitters (Lövheim, 2012). Although these choices are relatively small, the emotional states of an individual contain a reasonably high number.

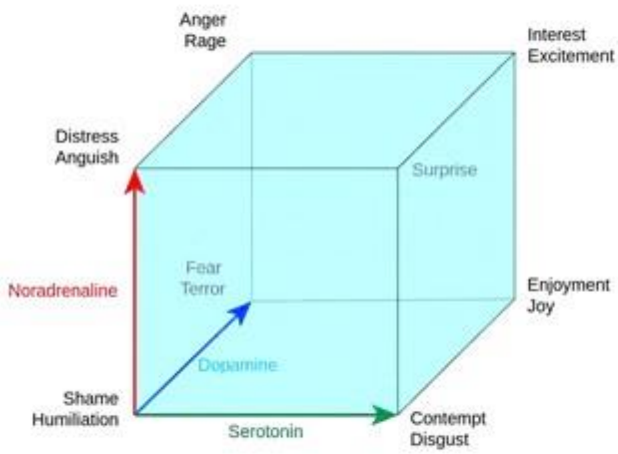


Figure 3: The Lövheim model of the connection between neurotransmitters and emotions

The Circumplex mood model (Reisenzein et al., 2013) employed a circumplex of two dimensions to represent the emotional state of participants: the factor of satisfaction tests the degree of positive and negative emotions, and the behavior factor tests the user’s probability of taking action under mood condition. Each dimension is quantified by means of ranking, so the Circumplex mood model provides continuous measurement of the mood and has adopted in many studies (Likamwa et al., 2013; Sun et al., 2017).

2.3 Emotion recognition

Emotion recognition is a fine-grained analysis of sentiment that seeks to recognize emotions from different sources of information, such as video, picture, text, and so on. A significant number of works have been conducted in this field in recent years. The researchers detected smile faces through the camera in the campus in the work of (Hernandez et al., 2012).

The research discovered that the emotional states of the users had a cyclical pattern and were strongly associated with external events. The emotional state is normally categorized into one distinct group in the conventional emotion classification. Some works followed the continuous distribution of probability to represent an image’s emotional state and suggested using the Gaussian mixture model for emotion recognition (S. Zhao et al., 2015).

As for text-based emotion analysis the authors categorized sentence-level emotion about label context dependence and formalized the problem as a problem of multi-label classification, allowing multiple emotions to be detected in a single sentence (Li et al., 2015).The EQ-Radio recently suggested the use of wireless signals to monitor the heartbeats of individuals, which have also been used as emotion detection features (M. Zhao et al., 2018).Their function, however, was based on dedicated Wi-Fi devices and on-body sensors such as ECG monitors.

2.4 Detection of human emotions via smartphones

Nowadays, researchers have shown with the rapid adoption of smartphones that mobile sensing data can be adopted to infer and detect human mental well-being such as stress, anxiety, depression, emotion and mood (Cao et al., 2017; MacKerron & Mourato, 2013; Servia-Rodríguez et al., 2017; Stütz et al., 2015; Suhara et al., 2017; Zhou et al., 2015).

DeepMood (Cao et al., 2017) detected bipolar affective disorder on a multi-view neural network-based smartphone using typing dynamics and accelerometer sensor data. Herdem (Herdem, 2012) aimed to assist mobile individuals when they need emotional help to connect offline with friends. A multifactorial statistical model was proposed by Bogomolov to understand everyday stress by comprehensively analyzing cell phone data and weather conditions (Bogomolov et al., 2014).

Via mobile mobility trace study, Canzian (Canzian & Musolesi, 2015) tracked user depression states. The identification of positive and negative effects using wild mobile product sensors was suggested by Mottelson (Mottelson & Hornbæk, 2016) Sensor data, APP usage information and SMS content were used by iSelf system (Sun et al., 2017) for emotion prediction using transfer learning (Likamwa et al., 2013).

However, only single emotion detection is considered in the current works, which ignored the fact that multiple emotions can co-exist in a time (Gao et al., 2012). In our paper, we are trying to predict two basic emotions, enjoyment and frustration, by making use of smartphones data from a collection of sensing information (Shah et al., 2015). To this point, we should make comprehensive that touch data, like finger-stroke, tapping, drag and drop, scaling, typing and swyping, will create data for us and help us in our research (Ghosh, Hiware, et al., 2019; Ghosh, Sahu, et al., 2019; Shu et al., 2019).

2.5 Sensors

Modern smartphones have a high number of sensors available, in addition to software-based recordings of how the phone is used. For Android phones, the following sensors might be relevant for the task of identifying emotions. In the present research and using the appropriate technology we chose to record data from the following sensors:

- Accelerometer (with gravity, motion and position detection)
- Accelerometer Linear (without gravity, motion and position detection)
- Gyroscope (affected by other vibrations, rotation detection)

Both the accelerometer and the gyroscope provide measurements for each axis in a 3-dimensional coordinate system, and for this project we will focus on accelerometer readings. The output is a vector consisting of the acceleration, in each of three directions x,y and z: when information from the accelerometer is sought:

$$\vec{A} = \{a_x, a_y, a_z\}$$

Equation 1: A vector expressed in X,Y,Z axes

Ideally, the accelerometer is built into a lying phone the values A = {0, 0, ±g} should be output still on a flat surface. However, the accelerometer data is not fully reliable. You can quickly see the noise by leaving the phone when displaying the accelerometer data, still. The values are determined not be constant but shift from one reading to another. That's it. The consistency of the data recorded is threatened by inaccuracy. Accelerometers are usually not used for monitoring movement and orientation to detect emotions or affective data generally. Although affective data cannot be accessed directly through the accelerometer, the forces exerted on the system are clearly generated. This enables us to determine the device's orientation and movement (Dzedzickis et al., 2020).

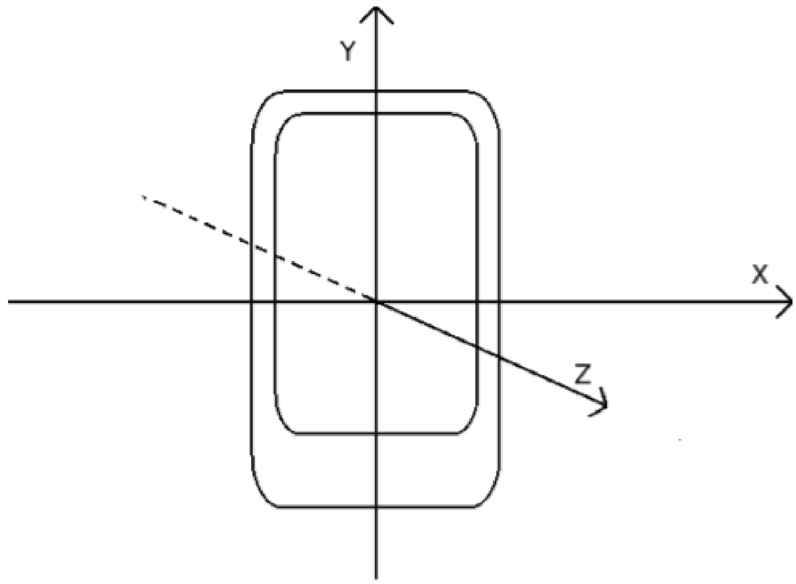


Figure 4: Image explanation X,Y,Z

A Huawei P10 with HiSilikon Kirin960 processor, with 4GB RAM and internal storage of 64GB was used for initial testing. It has a combined accelerometer, accelerometer linear and gyroscope. With a potential sampling rate of up to 4000Hz. It has a combined accelerometer, accelerometer linear and gyroscope, with a potential sampling rate of up to 4000Hz. It is reasonable to assume that the sensors can provide a sufficiently fine-grained set of data. In practice however the sampling rate does not reach such high frequencies, but it should not be necessary anyway. The people who were recruited to participate in the data collection for this research had many different smartphones, and they all managed to output the decided sampling frequency.

While the accuracy of such methods is always strong, do not make an instant assessment of the emotions of a person without utilization of every external machinery. This implies that there is some restricted practical utility, and approaches that the public cannot immediately use. In this study, I will answer these matters by implementing a framework for the prediction of the two basic emotions of a person based solely on information from the accelerometer, accelerometer linear and gyroscope sensors while user is playing.

The accelerometer enables us to record some details linking movement with emotions. In addition to enabling us to perform simple tasks such as telephone calls and sending text messages, smartphones are becoming more and more computationally efficient. This helps us to record and analyze the data of the accelerometer without interrupting phone use.

While emotion recognition is a fascinating term itself, but it has many Applications potential. Due to the intrinsic vulnerability, the findings would possibly not be used alone, but as part of this work. A series of various ways of emotions identification. Below you can find some Interesting application areas where an accelerometer could be used Be inclusive fairly.

A study conducted by Jaques et al (Jaques, Taylor, Azaria, et al., 2015; Jaques, Taylor, Sano, et al., 2015) amongst students at MIT aimed to identify which students were at higher risk of depression. The researchers tried to identify the everyday mood of the students as self-reported emotional by using different sensor and smartphone data (including the accelerometer, but not specifically to detect emotion states).

The gaming industry is continuously advancing technological advances. Such developments that make the games more realistic are more informative diagrams, improved physical simulations and greater control precision. It is easy to imagine that it will be fun for both developers and players to throw emotions into a game. In a 2011 novel entitled Ready Player One the author Ernest Cline describes a game that has become a substitute for normal human interactions (Cline, 2011): Todd13 scowled, and his face turned red – an indication that he hadn't bothered to disabuse the emotive function of his account in real time, thus making his avatar represent your face and body language. Pokémon Go (htt), a mobile game where players walk around collecting what are termed Pokémon's, is the most successful game in summer 2016 because it is a strong candidate for emotion detection based on readings of accelerometers, because players have to walk around to play the game around. The player's emotions could in some way impact how the game works.

3. Prototype App

3.1. Overview

Our game is a prototype app which aims to investigate the basic psychometric characteristics of its users. For this reason, this app has been build based on same other commercial games that are used massively by big companies in our days to give them a detailed overview of the psychometric profile of its employees.

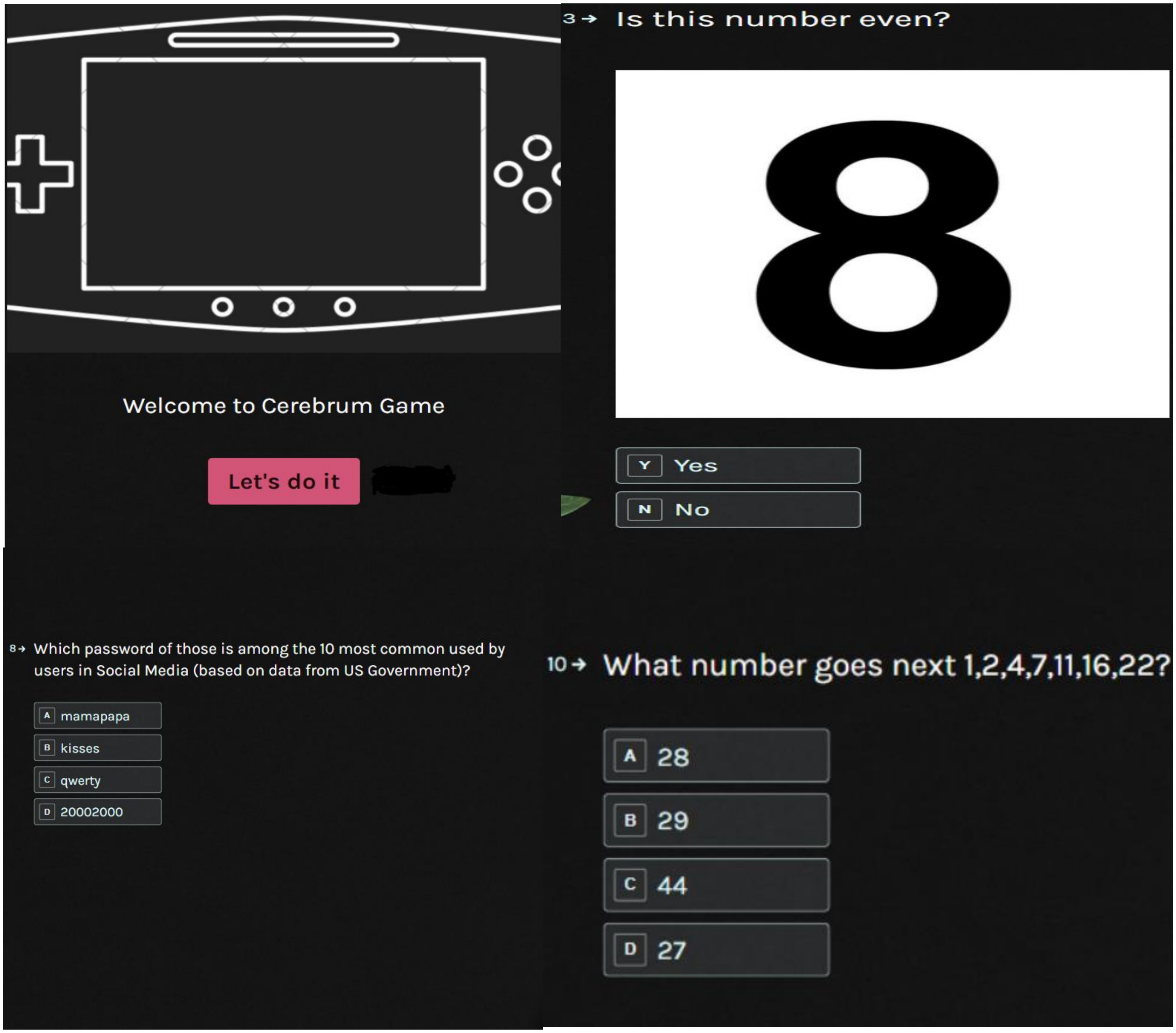


Figure 5: Preview of App 1

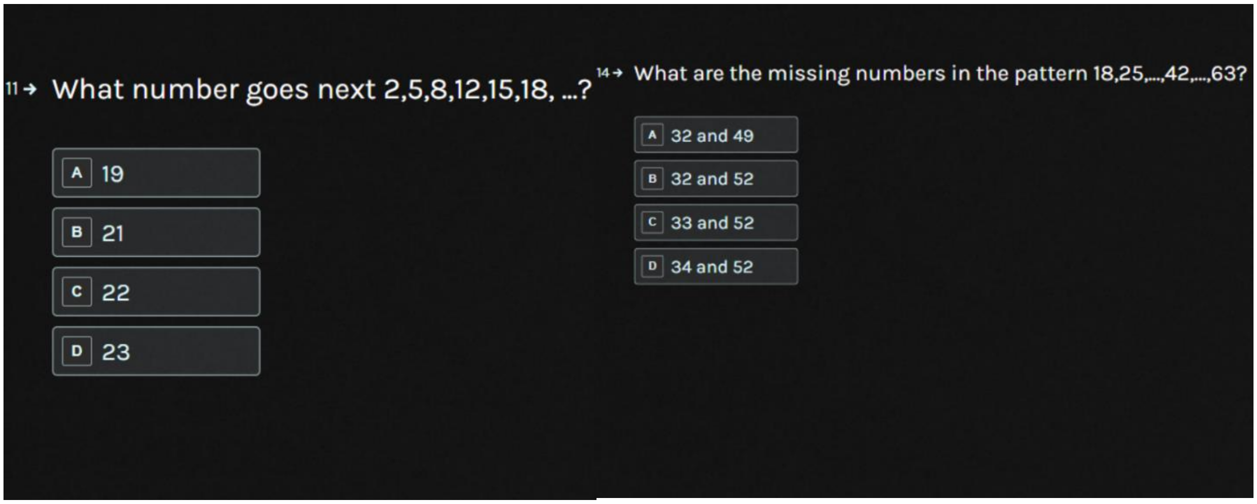


Figure 6: Preview of App 2

3.2. Overall architecture and technologies used

For the creation and successful execution of this game we looked for a platform that could give us a fully compatible game with the application of recording the sensors we chose for data collection. It was finally decided to be the open-source game development platform, MIT App Inventor (htt1). Through it we managed to set up a game in the standards of Skyrise City (htt2), capable of giving us a simple and short psychometric game that will arouse the interest of users.

Those kinds of game-based assessments are grounded in validated personality and cognitive neuroscience frameworks and use your real behaviors to create a psychometric profile. This type of games is set up so that through the challenges faced by the user can sketch a user profile which is wrapped around six psychological characteristics:

- Aptitude: the abilities that affect how you process and use information to perform mental operations.
- Personal Style: explain your personal preferences in dealing with situations.
- Cognition: the abilities that affect how you process and use information to perform mental operations. In practice, we perceive how you tend to process larger amount of information, how you tend to process information quickly and how the information processed at a consistent speed over time.
- Drive: reflects how you deliver results in the workplace. There are four indicators on this section: how you tend to bounce back more quickly after setbacks, how you focus more

influenced by values & goals, how you are motivated by reward and how you show initiative in organizing and completing longer projects with a consistent approach.

- Interpersonal Style: explain how you interact with other people. More certainly, it reflects how focused you are on needs of others, how you tend to adjust behavior to content or you prefer you own style, how you tend to feel more energized by social situations, how you tend to be more dominant and assertive and how willing you are to create conflict to achieve aims.
- Thinking Style: explain how you tend to approach and appraise problems and make decision. There are nine indicators for that category:

→shows how comfortable you feel with uncertainty & unpredictability

→shows how you prefer new and experimental ways to solve problems

→reflects how you tend to think more precisely or "outside of the box"

→reflects how you tend to be realistic or optimistic

→shows how you prefer familiarity and consistency or variety and regular change

→shows how you prefer or not to focus on immediate outcomes when making decisions

→reflects the tendency to make fast decisions involving risk that are based on emotional and physiological cues.

→shows the tendency to make decision based on your intuition or on your rational thought process

→shows how you tend to be deliberative or impulsive by carefully considering your action before reacting

3.2.1 Implementation of Sensor Record

The implementation of the data recordings took place with the simultaneous operation of the game and the application of the Sensor Record. More specifically, in the instructions sent to the users who were our sample, a link was sent that via emulator started at the same time our game and the application for recording the sensors of the mobile phone. A prerequisite is that the MIT App Inventor application and the Sensor Record application are pre-installed. Thus, we simplified the data collection process and automated the entire research process.

Additionally, with the completion of the game, the users who stopped the application of Sensor Record showed a frame to send their sensor data to our e-mail. In this way and after continuous communications and data shipments we managed to collect this large volume of data from the sensors of mobile users.

3.3. Pilot Testing

During the application development several different kinds of questions were checked. At first, we included psychometric problems that had to do with the perception of subordinate emotions. There was also a stage where the user had to base the memory to unlock a lock by saving patterns. While we regarded the types of challenges as necessary, it turned out at the end of the day that they triggered some minor errors, which make the study difficult. We would now say, because we collected data on mobile sensors, and even such minor delays will lead to a future leverage, that is necessary to avoid these delays. During the review of the survey data this will be observed. Since we have omitted any questions which have caused such problems, we may assume that the application we have produced is highly suitable for the research purpose.

Moreover, the literature notes that when we research user behavior using cell phones, the implementation of short methods is recommended. In this way, we adapted the basic process of study and sought to be motivated by the same pieces as our literature could understand. After this test, the steps of other research processes were considered essential to follow and not attempt and generate anything from scratch entirely. Finally, the application we created gave our research an interactive application that helped collect a large amount of data in conjunction with the sensor recording system.

4. Survey Instrument

Initially, in this research we wanted to build a more general profile of users regarding the familiarity in games as well as the availability of new technologies. So, after the necessary demographic questions, some followed that aimed at capturing this feature by the users. In fact, this is very important in our research since we must show a willingness to investigate the familiarity of the user so that then we can understand in depth his behavior.

With this in mind, we set up a questionnaire that examines and records users' predisposition to a new game or new technology, which tends to overwhelm users most of the time before they begin their interaction with each game. In this, we also sought to find out how many hours they spend each week playing games and what kind of games they prefer. So, in addition to the familiarization part we tried to collect data on the preferences and time they spend playing games.

Then, the second part of our questionnaire and research was set up so that we could record the user's feelings based on the Likert scale after playing our psychometric game. In this part, the questions were aimed at capturing the user's emotional transitions and the most successful recording of them in the questionnaire. But beyond this change of emotions, an attempt was made to capture the user's impressions of the overall experience he had with the game. That is why there are questions that study whether users are satisfied with the level, the design and whether they encountered difficulties. Now, it should be noted that there were some possible failures because users also had the Sensor Record application, which collected valuable data from mobile sensors for research, opened up together with the application.

It is worth mentioning that in most cases we ran our application designed with the MIT App Inventor via emulator and with the simultaneous execution of the data recording application of the mobile sensors. In essence, there was constant oversight of the research process with the aim of optimizing user performance and avoiding problems that may have resulted from the simultaneous execution of these two applications on a mobile phone.

We must mention that the whole survey consists of twenty-two questions. The first four were basic demographics of the respondents. The next six looked at users' familiarity with games and the use of new technologies. Then there was a question to know the mood of the user before starting to play the game. Then, and after the completion of the game on the part of the user, he was asked to answer the next 10 questions about the emotions caused by this experience as well as for the whole experience in general (difficulties, level, application design, etc.). Within them there was also a concentration question to filter and ultimately exclude those who answered by chance and gave hasty and contradictory answers.

4.1 Descriptive Statistics

We chose our sample to consist of 40 people since we wanted to collect a lot of data from the sensors of the users' mobiles to use them in the final part of the work. Thus, we judged as satisfactory a sample of 40 people who in the part of the questionnaires gives us 40 answers but in the part of the sensor data from their mobile phones, they sent us a lot of useful data.

Relying on the part of the answers we received, we realize that most users were men at a rate of 55% compared to 45% of women and a 5% who preferred not to disclose their gender.

Regarding the age part of the users, we observe that an equal percentage, ie 30%, belongs to the age groups 25-34 and 35-44. Also, the remaining 20% of the users of our sample belong to the age group 45-54, while the remaining group of users 18-24 with 17.5% and finally a 2.5% for people belonging to the age group 55-64.

When asked about the maximum level of education of the respondents, we received the equally impressive 52.5% from postgraduate level graduates. Next, we received a 27.5% of users who have completed university level education. The remaining percentage corresponds to a 15% for holders of a doctoral degree and a 5% for high school graduates.

In the next question which examined whether the people who completed our questionnaire were married or not, we received an overwhelming for the data of the question 67.5% of unmarried people and the remaining 32.5% stated that they are married.

We continue with the questions of preferences regarding games and user familiarity regarding the use of new technologies / applications. In the first question, which concerned the hours they spend weekly on mobile and computer games, we received the percentage that 42.5% of users spend from 4 to 10 hours in this type of activity. Also, 30% stated that they dedicate 1 to 3 hours, while 10% received the answers that stated up to one hour per week and not at all. Finally, 7.5% stated that they spend more than 10 hours in this activity.

Then we sought to find out the users' preferences about the games. Specifically, we studied the type of games they are said to prefer. There was a great deal of disagreement on this question but also many choices. Indicatively, we mention that 27.5% prefer sports games, 12.5% strategy games and 7.5% of users spend time on action, adventure, war, and car games.

We move on to questions about evoking emotions towards exposure to an application or technology. When asked if they feel nervous when they are about to learn / use a new application or technology, 30% of users answered that they disagree with it and 22.5% that they completely disagree. This condemns that 52.5% are in the opposite direction and do not seem to feel nervous when exposed to a new application / technology. Of course, the percentage of 27.5% is equally important, stating that they really agree that they feel nervous in the face of this situation. In fact, the remaining 12.5% declared neutrality towards this situation and only 7.5% expressed their confidence that they fully agree with this question.

The next question was whether users generally choose to avoid a new application or technology. In this, there was an overall percentage of 55%, who stated that 30% strongly disagree with this case and the remaining 25% that they disagree with it. However, 40% agree with this assumption, either to a large or lesser extent, with only 5% declaring neutral on the use of new applications and technologies.

In addition, users' views on whether they enjoy using new applications and technologies were sought. To this question, 37.5% reportedly agreed, 27.5% declared neutral in the use of new technologies. The users who strongly agreed and those who disagreed gathered 15%. In fact, 5% said they completely disagreed.

Many young people have noticed that they feel confident when using a new technology, this case has been made available to users. In this case a percentage of 50% agreed. Specifically, 32.5% agreed and 17.5% stated that they fully agree. 25% did not support either side. Finally, those who disagreed and those who declared complete disagreement got 12.5%.

The next question, closing the circle of previous familiarity questions, attempted to capture the current mood of users just before they started playing the game. In it we were asked to choose a mental mood from the six set by Ekman (Ekman, 1992). A percentage of 42.5% said they were surprised, 25% scared, 17.5% happy. A percentage of 10% said they were angry and the remaining 5% said they were sad. None of the detainees said they were disgusted.

At this point, we come to the part where we talk about the middle ground. This has to do with completing it after users have played the game and trying to reflect the change in users' emotions due to our psychometric game.

Descriptives

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
Howmanyhoursperweekdoyouspendonaverageplayingcomputer	40	1	5	3.28	1.086
Ifeelanxiouseverytimethatlmustlearnuseanewapplicatio	40	1	5	2.68	1.309
Igenerallyavoidusingnewapplicationsortechnologies	40	1	5	2.68	1.474
Ienjoylearningusingnewapplicationandtechnologies	40	1	5	3.43	1.083
Ifeelconfidentonusingtechnology	40	1	5	3.30	1.265
Ifeelhappy	40	2	5	3.63	.979
Ithoughtitwasfun	40	2	5	3.53	.933
Ienjoyedit	40	2	5	3.60	.841
Playingthisgamemakesmemoreintelligent	40	2	5	4.00	.877
Theactionsinthisgamearerespectable	40	2	5	3.65	.700
Itgavemeabadmood	40	1	5	2.65	1.075
Ifounditiresome	40	1	4	2.02	.947
Thegamedesignwaspoororcomplicated	40	1	5	2.82	.984
Iwouldprefereasiertasks	40	1	5	3.13	1.067
Ifacedseveraldifficultiesbecauseofthedesignorothertec	40	1	4	2.28	.960
Valid N (listwise)	40				

Figure 7: Descriptive Statistics

The first question after completing the game examines whether users felt happy during the game. 52.5% said they were happy, 15% said they fully agreed with it and 20% disagreed. The remaining 12.5% said they were indifferent to this claim. It is worth noting that the average of these answers is equal to 3.63, which shows that most people tend to embrace this case. In fact, this is aided by the fact that the prevailing price is 4 and coincides with the average.

Continuing with the study of positive emotions. In this question, we investigate whether users found this game fun. Most users answered 57.5% that they found it fun or a lot of fun. Still, 25% stated neutrality about this feeling while 17.5% stated opposition to this question. In this question, the average is 3.53, especially high while the mean value coincides with the normal and is equal to 4.

We wanted to ask our users if they enjoyed this overall experience. In this question, the overwhelming 52.5% said they agree and a 10% strongly agree. Then, 25% stated that they do not take a position on this question and 12.5% expressed their opposition to this case. The average of the total answers was calculated equal to 3.6 and the average value was calculated and coincides with the most common value and are both equal to 4.

To understand how competitive and lucrative our game queries were, we asked users if they thought that game made them more capable. In this, 77.5% agreed to a greater or lesser degree. The remaining 15% said they were neutral and 7.5% disagreed with the case. The average of the answers was calculated equal to 4 and is the only case that coincides with the average as well as with the most common value.

We wanted to raise the question of whether users consider the actions of this game to be worthy of respect. Of these, 55% said they agreed, while 32.5% said they were neutral on the question. Still, 7.5% said they strongly agree and only 5% disagreed. In these, we must mention that the average is equal to 3.65, the standard value and the average were calculated equal to 4.

Unraveling the tangle of negative emotions, we begin by exploring the general question of causing a negative mood. In this question, a percentage of 37.5% expressed their disagreement, a 27.5% stated neutrality towards the question. Also, 17.5% stated that they were provoked negatively during the game, while 12.5% stated that they were completely against it and 5% completely agree with this question. The average of the answers was calculated equal to 2.65, with the most common value being the second option, i.e., the disagreement on the specific answer and a median value equal to 2.5.

The next question was posed to filter if and to what extent the user is with us and does not answer by chance. Mention that if someone answered this simple question incorrectly, they would be excluded from our sample as it would give us the impression that he chooses the answers randomly and is not reliable to include in our sample. Fortunately, we did not receive an answer beyond the obvious and this makes us optimistic about the progress of the investigations. This method is a frequent filtering that aims to reduce the biased participation of individuals in research.

We continue with our questions, with this one about whether users found this game exhausting. A fairly large percentage of 75% disagree with this claim to a greater or lesser degree. The remaining

15% said they were neutral on the question and only 10% found it tedious. On the answers, an average of 2.03 was calculated, with an average and standard value equal to 2.

The next question asked the users' opinion on the design of the application. In it, 40% stated that they do not take a position on this question with 25% disagreeing, 22.5% agreeing, 10% completely disagreeing and the remaining 2.5% strongly agreeing. Also, the average was calculated equal to 2.83 with an average and usual value of 3.

In addition, we asked if users would prefer easier tests. 40% are said to agree with this, while 22.5% expressed their opposition. In fact, 25% said they did not share this question, 7.5% strongly disagreed and 5% identified with the question. On the statistical analysis of the results, the average is equal to 3.13 with a median equal to 3 and a more common value of 4.

Finally, we wanted to investigate whether users encountered any technical difficulties they might have with the application design and other technical compatibility issues with the application. To the following question, 47.5% of users answered that they were neutral about what we are studying. Beyond that, 30% said they strongly disagreed, 17.5% disagreed and only 5% said they had difficulties. The average was calculated as 2.28 on the answers, with an average and standard value equal to 3.

5. Data Collection and Processing

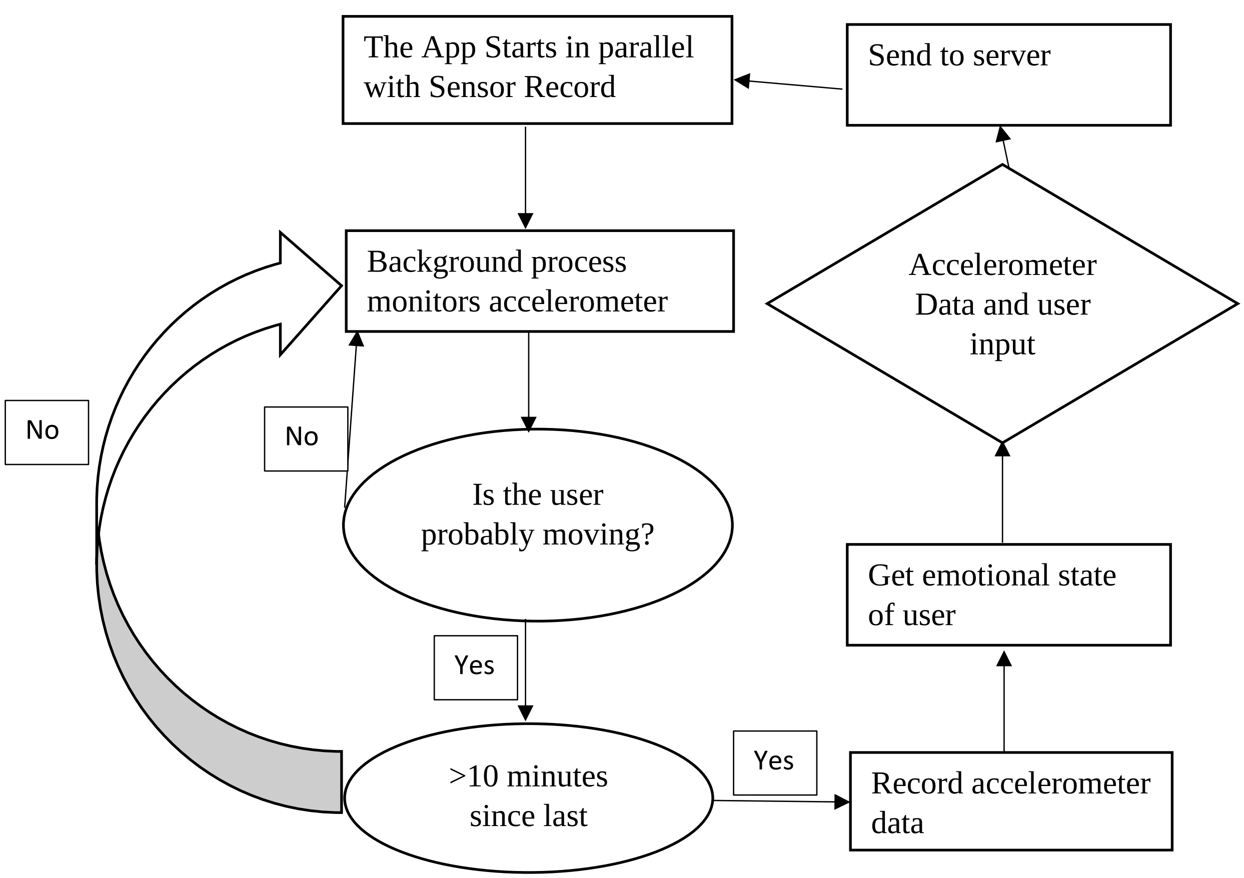
5.1 Field Study (Participants and Procedure)

During the research process we had completed the initial part of the questionnaire to investigate the demographic characteristics. Then followed the process of the psychometric game we created called "Cerebrum Game" and we had programmed it to run in parallel with the data collection application through the sensors of the mobile "Sensor Record". Finally, the sampling was completed by completing the second part of the questionnaire, which was set up to evaluate the user's previous experience.

Conducting this research, we set up our questionnaire on the special site for creating research questionnaires, questionpro.com. Also, during the sampling procedure, which was carried out with basic supervision throughout the process, the following mobile phone device was used:

Mobile Phone	HuaWei P10
Android	8.0.0
RAM	3G
ROM	64GB
CPU	4x2.1GHz ARM Cortex-A53, 4x1.7GHz ARM Cortex-A53, Cores 8
GPU	ARM Mali-T830 MP2, 900MHz, Cores 2
Dimensions	72x146.5x7.2 mm

Figure 8: Mobile Phone Characteristics & Mobile Sensors data tracking method diagram



During the data collection process, the phone can be in many different states of movement and usage:

1. Lying still and not being used
2. Lying still and used
3. Held in hand and being used while the user is stationary
4. Held in hand and being used while the user is moving
5. Carried and not being used while the user is moving

5.2 Data Overview

During our research, a large amount of data was collected and processed. Starting with the self-reported emotional data, the participants have five options to choose from in each dimension of emotions. While such a data resolution might be interesting, to increase the classification accuracy, the two highest options and the two lowest options will be combined in order to get three different classes on each scale: *low, medium, and high* as we observe from the table below.

User-reported data	Class
1-2	Low
3	Medium
4-5	High

Figure 9: User-reported data & classes

From the table below we have an overview of the participants that contributed to this research:

#asked: 82
#accepted: 42
#participated: 40
#recordings: 40

Figure 10: Participants Overview

Given that for research purposes we entered the process of recording the sensors of the mobile, below we have a picture of the volume of data collected by the three sensors we chose to collect the data.

Sensor	Amount of data (expressed in excel rows)
Accelerometer	450.886
Accelerometer Linear	455.549
Gyroscope	460.614
Total	1.367.049

Figure 11: Amount of sensor data

From the next table we have an overview of various features that have been calculated based on the data of our research:

Averages	Accelerometer	Accelerometer Linear	Gyroscope
X	0.27	0.48	0.0044
Y	0.23	0.08	-0.0010
Z	9.83	4.87	0.00067
Enjoyment	3.9	3.9	3.9
Frustration	2.0	2.0	2.0
Variations	Accelerometer	Accelerometer Linear	Gyroscope
X	0.000886	0.001689	0.000806
Y	0.000326	0.001435	0.000200
Z	0.003332	0.024508	8.43855E-06

Figure 12: Calculation of X,Y,Z data

5.3 Data (pre)processing

The accelerometer data presents some difficulties when it comes to data quality:

- The accelerometer (linear or not) is not perfectly accurate, adding a little noise.
- When you move and play with the phone, you cause confusion to the sensors and the readings from individual axes is difficult to compare.
- The non-placement of the mobile in a stable place can make the telephone feel more jerky
- The participant will influence the recordings with gait features.
- Whether the participant moved for the entire recording period is very uncertain.

Firstly, we have cleaned out data by isolating:

- Standard Types which the pandas recognized as out of the box
- Non-standard Types, that pandas have not automatically recognize as missing values
- Unexpected Types when a column of names contains unexpected data types
- Summarizing
- Replacing, when making simple replacements

5.4 Feature Selection

In order to determine the features correctly, we had to start with the following definitions first. Given that we had to export features from a huge amount of data, we set up a specific computing database. This was introduced as the calculation of the average of every twenty observations so that we can proceed to the following calculations. Practically, these 20 observations are equal to 200 milliseconds, or 0.2 seconds. In this way, we've been able to extract accurate results for a variety of features as you have the opportunity to read below.

For the success rate of a classifier, the choice of good features is essential. There are a series of features in this section and the most promising features are chosen in the next chapter. Three different axes are measured by the sensors. In this work, we do not intend to consider the position of your telephone. Therefore, when calculating the features, it is reasonable to consider the total movement instead of the movement on the individual axes. The readings $a_{x,i}$,$a_{y,i}$, $a_{z,i}$ start with time t_0=0 and end with time t_{M-1} for each recording of the M accelerometer are as follows:

$$
\begin{aligned}
\vec{a_x} &= \{a_{x,0}, a_{x,1}, \ldots, a_{x,M-1}\} \\
\vec{a_y} &= \{a_{y,0}, a_{y,1}, \ldots, a_{y,M-1}\} \\
\vec{a_z} &= \{a_{z,0}, a_{z,1}, \ldots, a_{z,M-1}\}
\end{aligned}
$$

Figure 13: Definition of Ax, Ay, Az

Those values form the basis for calculating the characteristics and we assume that the data is segmented into steps in this section. This means that in acceleration vector values a_x, a_y, a_z consist only of the parts of the original recording where we have found that the participant has been moving, i.e., the location of the detected steps.

The choice is mainly based on what the related projects have tried, (Adibuzzaman et al., 2013; Bernhardt, 2010; Bonomi et al., 2009; Coutrix & Mandran, 2012; Crane & Gross, 2007; Gong et al., 2010; Gyllensten & Bonomi, 2011; Preece et al., 2009; Z. Zhang et al., 2016).

The features indicated below are gray marked.

Acceleration:

$$
\text{For } i = 0 \ldots M-1: \quad a_i = \sqrt{a_{x,i}^2 + a_{y,i}^2 + a_{z,i}^2}
$$

giving the acceleration vector

$$
\vec{a} = \{a_0, a_1, \ldots, a_{M-1}\}
$$

Figure 14: Acceleration Vector Type

Mean Acceleration (5.1):

$$\bar{a} = \frac{1}{M}\sum_{i=0}^{M-1} a_i$$

Figure 15: Mean Acceleration Equation

The standard deviation of a set of values tells us something about how these values are distributed and may be of interest as a feature.

Standard Deviation of acceleration (5.2):

$$\sigma_a = \sqrt{\frac{1}{M}\sum_{i=0}^{M-1}(a_i - \bar{a})^2}$$

Figure 16: Standard Deviation Equation

The peak jerk can also be calculated for each step, giving a measure of the strongest movement in a step. With s steps and the peak acceleration for step k denoted as $a_{top,k}$ a measure for the mean peak acceleration a_{top} can be calculated as:

$$\bar{a}_{top} = \frac{1}{s}\sum_{k=0}^{s-1} a_{top,k} \quad \textbf{(5.3)}$$

Figure 17: Mean Peak Acceleration Equation

And the **standard deviation (5.4)**:

$$\sigma_{a_{top}} = \sqrt{\frac{1}{s}\sum_{k=0}^{s-1}(a_{top,k} - \bar{a}_{top})^2}$$

Figure 18: Standard Deviation Equation

Another measure of the energy in the acceleration is the root of the sum of the squared acceleration values. Compared to the mean acceleration, this feature lets the highest values give a larger contribution.

Squared Acceleration Values:

$$\bar{a}_{rss} = \sqrt{\frac{1}{M}\sum_{i=0}^{M-1} a_i^2}$$

Figure 19: Squared Acceleration Values

Now, we will focus on jerk. The definition of jerk is a quick suddenly arrested push, pull or twist. Basically, is refers to when we make a sudden spasmodic motion. Jerk is the first derivative of acceleration.

$$\begin{aligned}
&\text{For } i = 0 \ldots M-2: \\
&\quad j_{x,i} = a_x'(t) \approx \frac{a_{x,i+1} - a_{x,i}}{\Delta t} \\
&\quad j_{y,i} = a_y'(t) \approx \frac{a_{y,i+1} - a_{y,i}}{\Delta t} \\
&\quad j_{z,i} = a_z'(t) \approx \frac{a_{z,i+1} - a_{z,i}}{\Delta t} \\
&\quad j_i = \sqrt{j_{x,i}^2 + j_{y,i}^2 + j_{z,i}^2}
\end{aligned}$$

giving the jerk vector

$$\vec{j} = \{j_0, j_1, \ldots, j_{M-2}\}$$

Figure 20: Jerk Vector

Mean jerk (5.5):

$$\bar{j} = \frac{1}{M-1}\sum_{i=0}^{M-2} j_i$$

Figure 21: Mean Jerk Type

Standard Deviation of jerk (5.6):

$$\sigma_j = \sqrt{\frac{1}{M-1}\sum_{i=0}^{M-2}(j_i - \bar{j}_i)^2}$$

Figure 22: Standard Deviation of jerk

On the same way of thinking, the peak jerk for each step can be calculated. With s steps and the peak jerk for step k denoted as $j_{top,k}$, a measure for the mean peak acceleration j_{top} can be calculated as follows:

Mean peak jerk (5.7):

$$\bar{j}_{top} = \frac{1}{s}\sum_{k=0}^{s-1} j_{top,k}$$

Figure 23: Mean Peak Jerk

And the **standard deviation (5.8):**

$$\sigma_{a_{top}} = \sqrt{\frac{1}{s}\sum_{k=0}^{s-1}(a_{top,k} - \bar{a}_{top})^2}$$

Figure 24: Standard Deviation of Peak Jerk

Step Duration: Step limits are defined in the calculation of steps. These data points can be used to calculate the duration of each step because the total time for individual recordings and the number of data points are known in each recording. With steps consisting of p_k data points, a total time of T and M data points, each step is averaged by: The mean step time:

$$\bar{t}_s = \frac{1}{s}\sum_{k=0}^{s-1}\frac{p_k T}{M} \quad \textbf{(5.9)}$$

Figure 25: Mean Step Type

and the standard deviation:

$$\sigma_{t_s} = \sqrt{\frac{1}{s}\sum_{k=0}^{s-1}\left(\frac{p_k T}{M} - \bar{t}_s\right)^2} \quad \textbf{(5.10)}$$

Figure 26: Standard Deviation of mean step

Fourier Transform:

A Timing Signal is transformed into a frequency domain with Fourier Transform. (htt3) This applies especially to data with a certain periodicity, which applies to accelerometer data recorded in this project. The acceleration transformation of Fourier is calculated using the Fast Fourier algorithm and is based on the following equation:

$$A_k = \sum_{m=0}^{M-1} a_m e^{-i2\pi km/M}$$

Figure 27: Fourier Transform

Where k to 0, …, M – 1

The resulting Fourier coefficients A_k, or a subset thereof, can be used as features.

Skewness and Kurtosis

The statistical measurements of the asymmetry of the form of the distribution are Skewness (R.J. Larsen, 2014) and Kurtosis (htt4). For a specific set of values, skewness is calculated as follows:

$$Skew(A) = E\left[\left(\frac{A-\mu}{\sigma}\right)^3\right]$$

Figure 28: Skew Type

This function is calculated in accordance with the alternative calculation of the mean acceleration and jerk characteristics above, for every step. The stage measurement of the skewness can be calculated as each step involving the p_k data points and the start index for each step of start.

$$Skew(A)_{step} = \frac{1}{s}\sum_{k=0}^{s-1}\left(\frac{\frac{1}{p_k}\sum_{i=start_k}^{start_k+p_k-1}(a_i-\bar{a})^3}{\left(\frac{1}{p_k}\sum_{i=start_k}^{start_k+p_k-1}(a_i-\bar{a})^2\right)^{\frac{3}{2}}}\right) \quad \textbf{(5.11)}$$

Figure 29: Skew step

Similarly, kurtosis is calculated as:

$$Kurt(A)_{step} = \frac{1}{s}\sum_{k=0}^{s-1}\left(\frac{\frac{1}{p_k}\sum_{i=start_k}^{start_k+p_k-1}(a_i-\bar{a})^4}{\left(\frac{1}{p_k}\sum_{i=start_k}^{start_k+p_k-1}(a_i-\bar{a})^2\right)^{2}}\right) \quad \textbf{(5.12)}$$

Figure 30: Kurtosis step

Power Spectral Density (PSD): PSD is the measurement of the frequency power distribution of a time series signal in which power is measured by the time series square. Utilizing the Fourier Transform, the discrete version of PSD is calculated as follows:

$$\tilde{S}_{aa}(\omega) = \frac{(\Delta t)^2}{T}\left|\sum_{j=1}^{M} x_j e^{-i\omega j}\right|^2$$

Figure 31: PSD Equation

From the resulting values we can calculate the mean:

$$\bar{S}_{aa} = \frac{1}{l}\sum_{k=0}^{l-1}\tilde{S}_{aa,k} \quad \textbf{(5.13)}$$

Figure 32: Mean (PSD)

$$\sigma_{S_{aa}} = \sqrt{\frac{1}{l}\sum_{k=0}^{l-1}(\tilde{S}_{aa,k} - \bar{S}_{aa})^2} \quad \textbf{(5.14)}$$

Figure 33: S_{aa} array type

An overview of the various features:

Name	Equation	Enjoyment	Frustration
Mean Acceleration	5.1	3.77	2.61
Std Acceleration	5.2	0.034	0.031
Mean Peak Acceleration	5.3	0.12	0.23
Std peak Acceleration	5.4	0.0009	0.0012
Mean Jerk	5.5	0.002	0.005
Std Jerk	5.6	0.0001	0.0003
Mean Peak Jerk	5.7	0.07	0.13
Std peak Jerk	5.8	0.0004	0.0012
Mean step duration	5.9	0.01	0.024
Std step duration	5.10	0.0002	0.0005
Skewness	5.11	-1.18	0.32
Kurtosis	5.12	1.46	2.00
Mean PSD	5.13	0.02	0.04
Std PSD	5.14	0.001	0.003

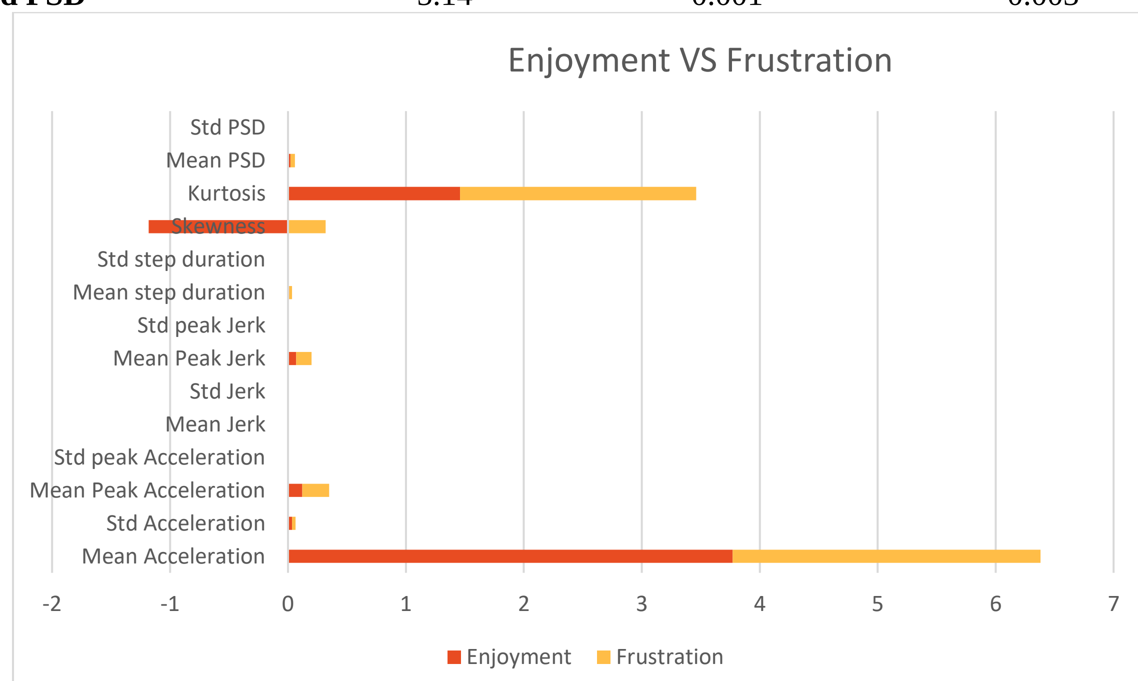


Figure 34: Enjoyment vs Frustration Values

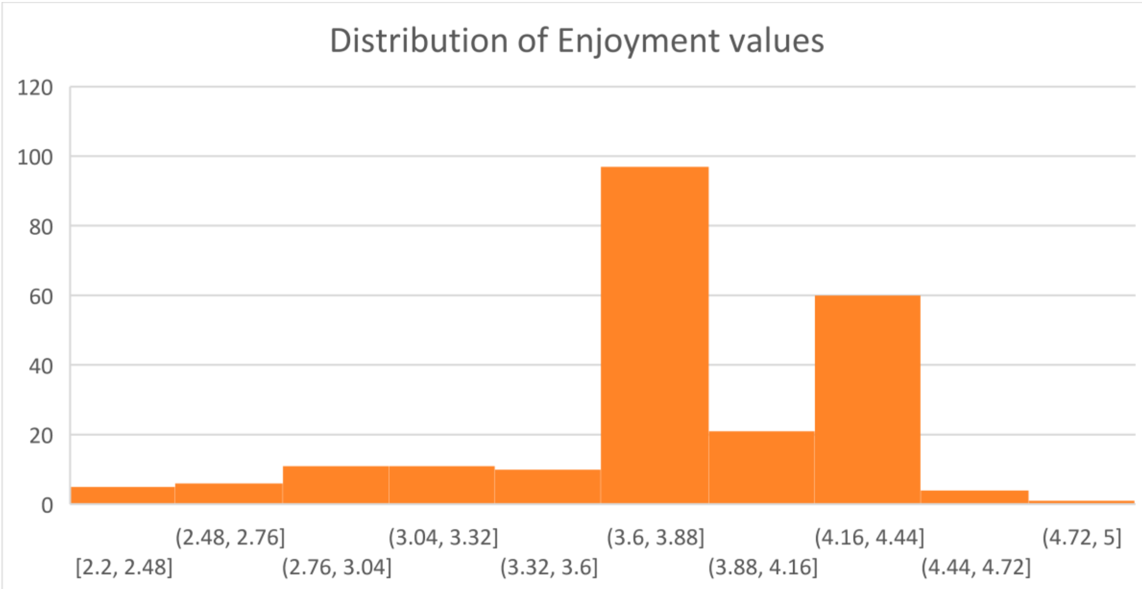


Figure 35: Distribution of Enjoyment

From the analysis of the data and their distribution we understand that regarding the term of ‘’Enjoyment’’ the most common value we receive is between the range 3.6 and 3.88. Also, the second most common price range is that of 4.16 to 4.44. To all this, if we add the third most common which is the intermediate range, i.e., 3.88 to 4.16 we can conclude that **78% of users gave us an average value that is between 3.6 and 4.44**. So, in conclusion, we realize that this particular game tends to cause a small to large enjoyment value to its users.

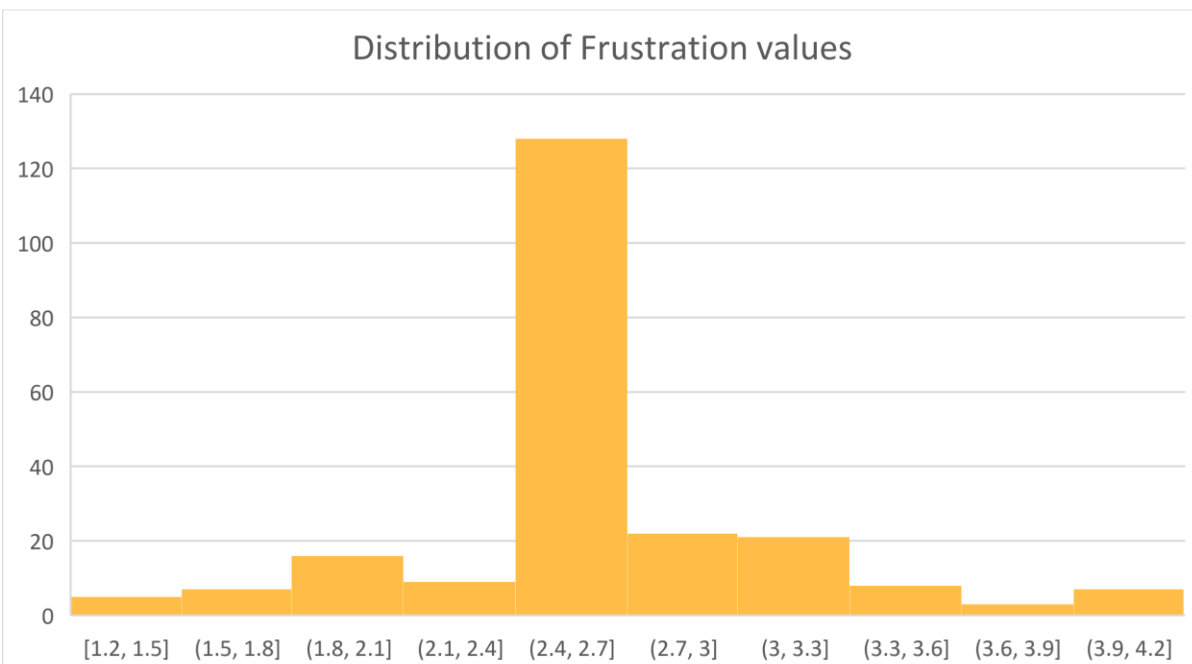


Figure 36: Distribution of Frustration

Regarding the term ‘’Frustration’’, the values we received by 56% are between the range 2.4 to 2.7. In this regard we observe a less broad distribution since most users seem to be limited between this price range. Finally, the other 2 most common options are the ranges 2.7 to 3 and 3 to 3.3 in which we meet a percentage of users equal to 9.2% and 8.8%.

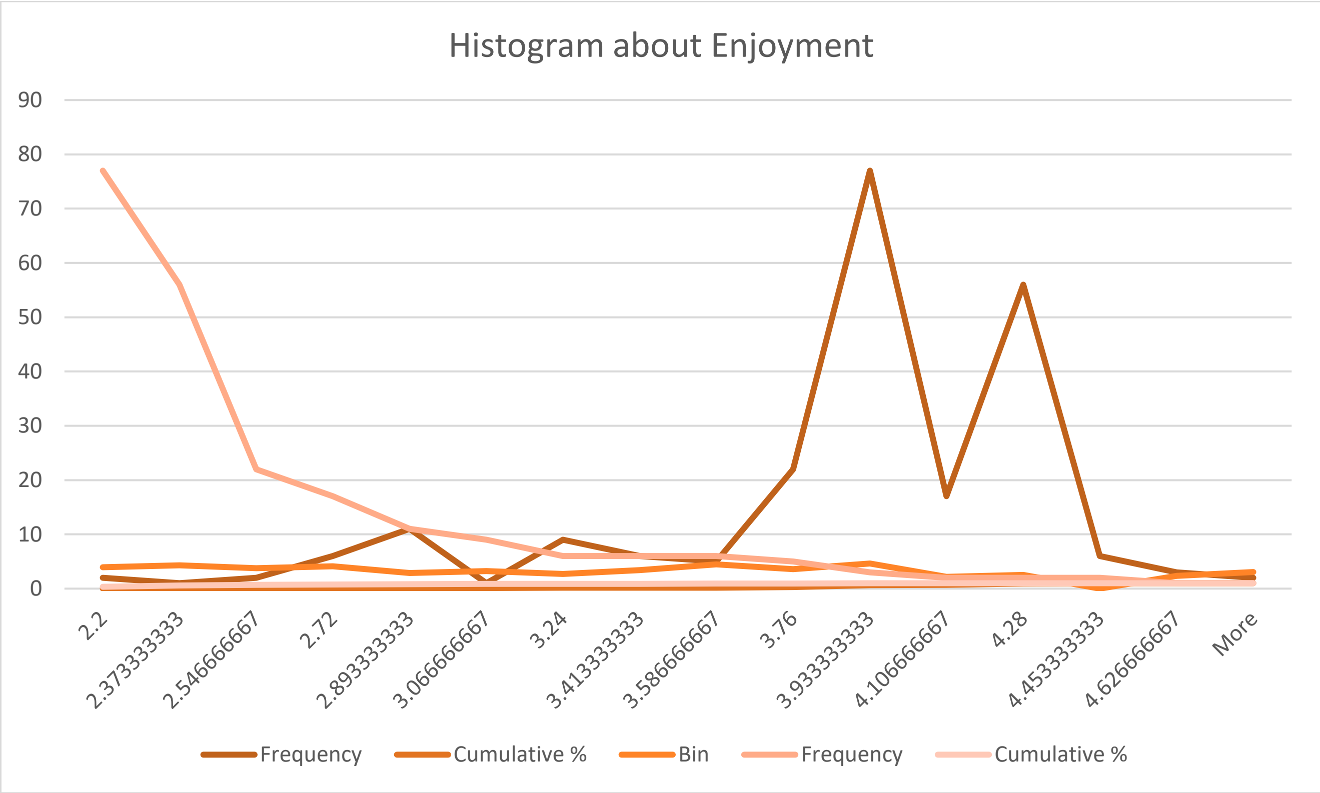


Figure 37: Frequencies of Enjoyment

Observing the frequency histogram of ‘’Enjoyment’’, we realize that most values are distributed at 3.93 and 4.28.

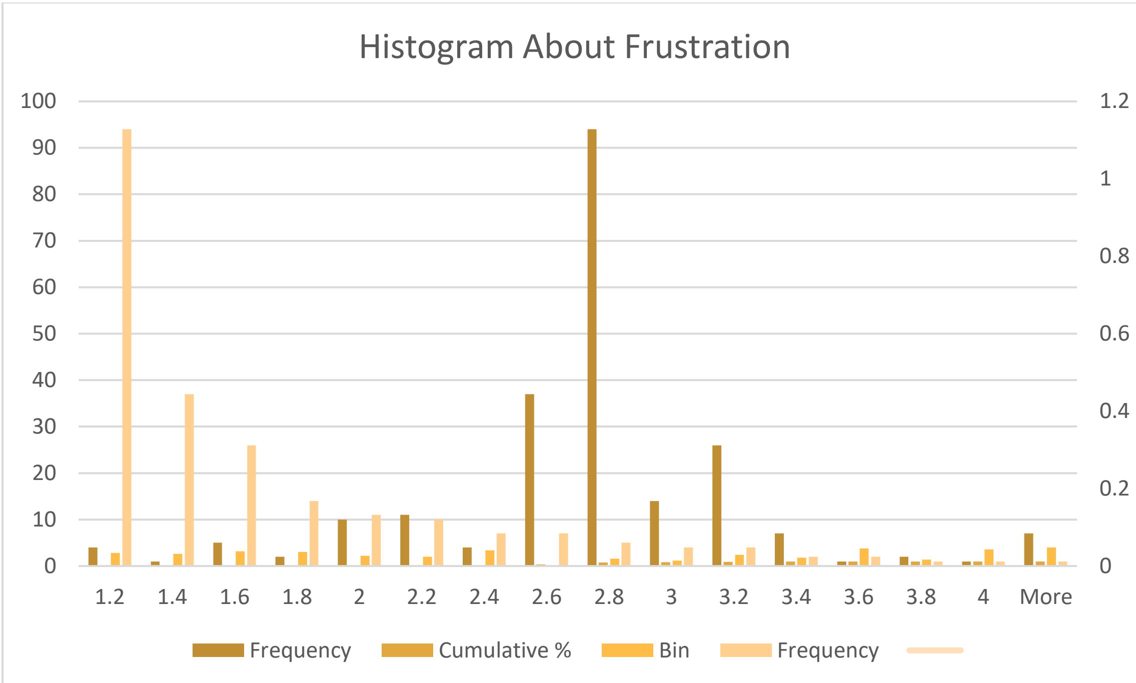


Figure 38: Frequencies of Frustration

From the creation of the frequency histogram for the term Frustration we understand that we have the largest distribution in the values 2.8 and 2.6 which reflect 57.70% of user responses.

While the theoretically infinite range of features is possible, only a subset of features will have positive discrimination. The selection process tries to identify the good characteristics and eliminate the bad ones. Only the best functions can successfully improve the results of classification and at the same time reduce the time.

This may also be part of the extraction process but is a separate procedure in this work. Improvements are particularly likely if the data set is small and because the risk of overpowering exceeds many features, the algorithm is adapted to the training set too much to reduce the capacity to generalize. Some selection of features should therefore be carried out. The risk of overfitting and thus reduction of general generalization is high, especially since the data set collected for this project consists of recordings collected from very small numbers of people.

One way to perform feature selection is through a method called recursive feature elimination (RFE). As the name suggests, the idea is to begin with all the functions and remedially remove the feature which contributes the least to the classification (i.e., has the lowest absolute weight). This is repeated until you have the desired number of functions. The following criterion will be applied to this project, instead of deciding on a specific number of features: remove functions if classification accuracy is not declined. This means that the training set is classified as low as possible with the most precise features that are still to be classified. RFE is available in the Python Framework for Scikit-Learn. (Elimination, n.d.)

Every participant has created big rows of data by playing this game and tracking their actions with the three sensors that we have seen. To facilitate the use of machine learning on this data, every participant's data needed to be represented by several features. The strategy was to include features found in related work, features derived from emotion theory, in addition to conceivable features computable using gathered data.

Some features are applicable for all three tasks (such as distance between fingers in the scaling task). Each feature is represented by several sub-features: the minimum, maximum, average, and median value, with some variation based on the applicability of the specific feature. Distances were normalized over screen sizes. We extracted a total amount of almost 550.000 to 600.000 raw data sensors (for every sensor).

5.5 Reducing class imbalance in emotion recognition

To this step, we want to measure the reliability and balance between our survey data we will use the Cronbach Alpha statistical test.

Cronbach's Alpha measures internal consistency between items in a scale. We should make sure not to mix positively and negatively worded questions-Alpha will be negative. We should use reverse code variables if needed.

A "high" value for alpha does not imply that the measure is unidimensional. If, in addition to measuring internal consistency, you wish to provide evidence that the scale in question is unidimensional, additional analyses can be performed. Exploratory factor analysis is one method of checking dimensionality. Technically speaking, Cronbach's alpha is not a statistical test – it is a coefficient of reliability (or consistency).

To this research, we try to recognize two basic human emotions: enjoyment and frustration.

For the first one, by running the statistical test, we received Cronbach's Alpha rate based on standardized items equal to 0.755 or 75.5%, a percentage higher than 0.7. Therefore, the data of our research about enjoyment are considered reliable and far from leverage.

Furthermore, we will run the above statistical test for the second human emotion we study in our research. Therefore, we received the price 0.709 or 70.9%, a percentage that is also considered reliable since it exceeds the reliability base of 0.7. Finally, we can conclude that none of our questions predispose the user to select or direct in one direction of answers and therefore the questions are considered inviolable and impartial, not guiding the user to certain pairs of answers.

Reliability Statistics: CA	Cronbach's Alpha	CA Standardized Items	N of items
Enjoyment	.755	.738	5
Frustration	.704	.709	5

Figure 39: Cronbach's Alpha reliability tests

At this point, remember that in both cases, the number of items is five as we have done our research by using five questions to calculate enjoyment and five about the user frustration.

6. Evaluation/Prediction Models

In this step we will try to predict two basic human feelings (Enjoyment, Frustration) based on the mobile sensor data from accelerometer, accelerometer linear and gyroscope that we have tracked in this research. To this effort we have constructed the following detecting emotion recognition models for the two basic emotions that we have targeted.

Before doing that, in this research we will use the term emotion mean value. We will define this term as the average value we received from the users of our research regarding the emotional effect that they had by playing our game. Next, we examine to what extent it has occur changes in the two basic emotional states that we try to predict. To this we should remember that we have used Likert scale from 1 to 5 from our research.

To give an example, when we receive a mean value of 3.8 about enjoyment, we concluded that our game had a positive impact in this user because it makes them feel entertainment. In simple terms, his emotional state about this particular variable has changed from 3 (which we assume all users start in our research purposes) to 3.8. Finally, this whole process has caused a positive change to the term of enjoyment which is equal to 0.8.

Conversely, calculating frustration mean values we received 2.6. In the same way of thinking, we understand that from the general emotion value which we have set equal to 3, we have moved to 2.6. This is a negative change and practically reflects that this user did not feel frustration after the whole procedure and even at a rate equal to 0.4.

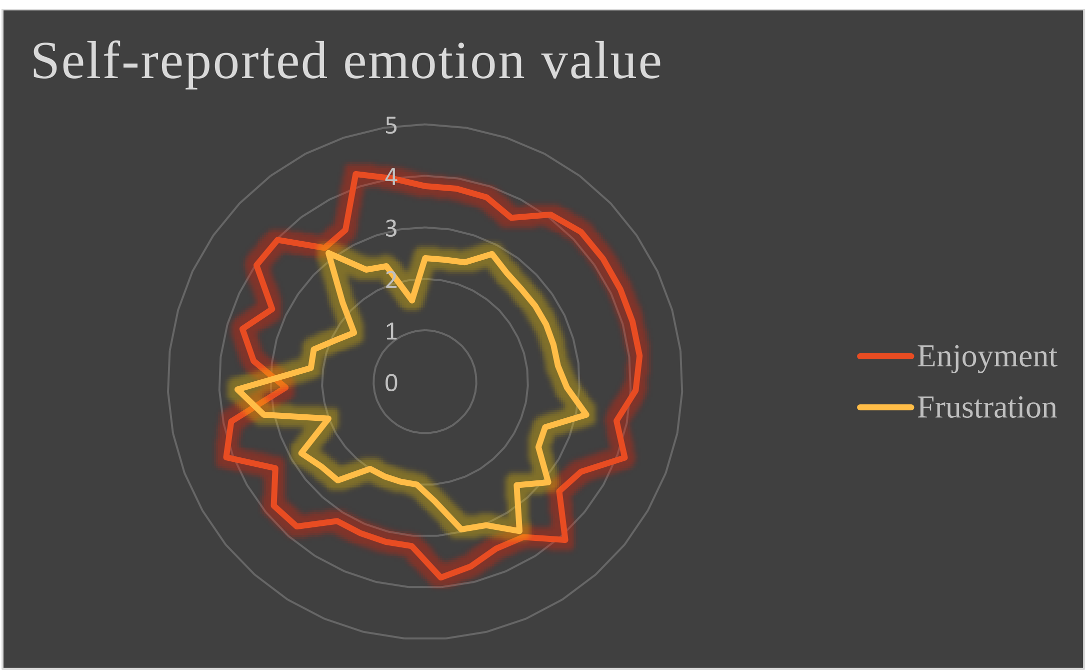


Figure 40: Self-reported emotion value

6.1 KNIME Analytics Platform

In the following research we try to discover the prediction model that will give us the greatest accuracy as soon as we feed it with the data we have captured from the users' sensors. The examination is conducted with the use of the latest updated KNIME 4.2 Analytics Platform. On this subsection, a reference on the KNIME's tools used for the current research is held.

KNIME, the Constanz Information Miner, is an open-source data analysis, documentation, and integration tool. It combines different components for machine learning and data mining with its modular pipelining concept. The graphical user interface allows the assembly of nodes that combine different data sources, including pre-processing, without or with minimal programming, for modeling, data analysis and visualization.

KNIME Analytics Platform is an open access data science program. Intuitive, open, and constantly incorporating new technologies, KNIME produces data comprehension and data science workflows. KNIME was used to predict one dependent variable for others. In addition, as the abbreviation reads, the Konstanz Information Miner provides a wide range of building blocks and tools from third parties. Files are stored in a workspace that includes a workflow. Workflows may be programs or procedures that define the steps used for loading, simulating, or converting data (Melcher & Silipo, n.d.).

The architecture of KNIME is based on three key concepts. The visual interactive platform allows data to be controlled by drag-and-drop from several processing units. Customized systems can also be modeled by individual data pipelines. The modularity of the different processing units and data containers does not require reciprocal dependence, making it easier to deliver computing and enabling the autonomous development of different algorithms. Data types are already incarnated, i.e., categories are not predefined, while new types are easily added along with type-specific renderers and comparators. The above is accomplished by a pipeline of nodes, connected by edges that transport data or templates, in order to process the data processing. Each node processes the incoming data and produces output results (Bakos, 2013).

6.2 Linear Regression

Linear regression is a linear model i.e., a model that assumes a linear relationship (straight-line relationship) between the input variables (x) and the single output variable(y).

Linear Regression process:

- Add Partitioning note to File Reader output

-top port should have 70% of the rows

-draw randomly such rows

- Add Linear Regression Learner to top output port of Partitioning node

-select feeling column to be learned

-execute the node and open its scatter plot views. Which column is most correlated to each feeling (column selection tab)?

- Set up the Regression Predictor

-predicts test set (remaining 30% rows) by simply connecting the remaining unconnected output ports

- Remove rows with missing prediction
- Add Numeric Scorer to Regression Prediction Output

-reference column: the column you learned

-predicted column: the new column created by the predictor node

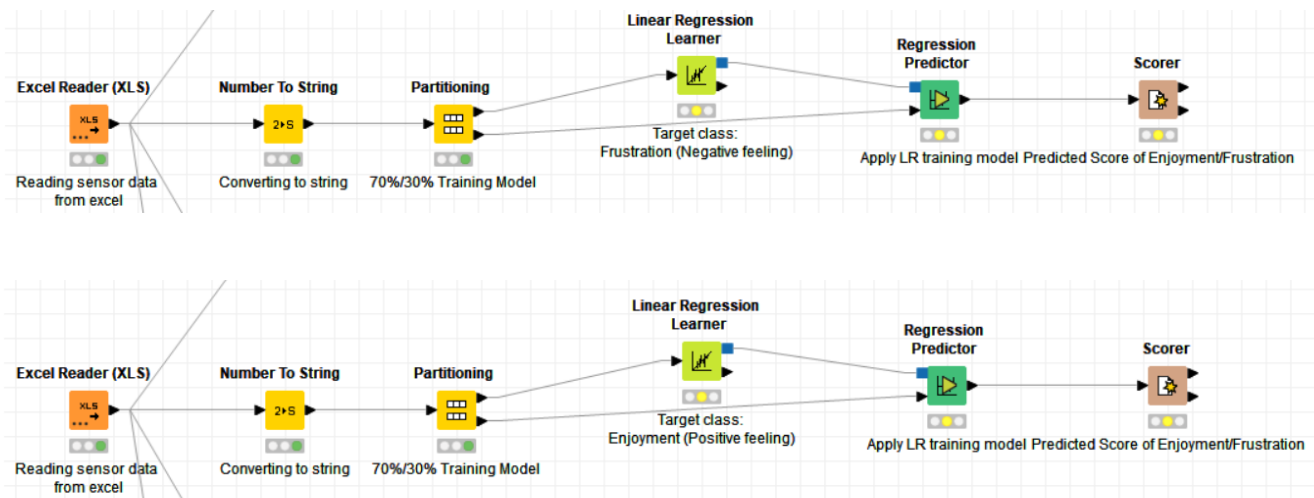


Figure 41: Linear Regression Model

This workflow evaluates the performance of a regression model and checks some of the assumptions of a linear regression model. Extensively, this model firstly reads the data that we have collected from our survey users and then it removes all the missing values in the columns of:

1. Accelerometer
2. Accelerometer Linear
3. Gyroscope

Next, we have partitioned our dataset into two, the 30% that we have used for testing and the following 70% that we have used to train our Linear Regression Model. With this 70% of our data, we have finally trained our model. Important to mention that we have also added the linear correlation node to check the multicollinearity. This term refers to a situation in which two or more explanatory variables in a multiple regression model are highly linear related.

6.3 Decision trees

A decision tree is a flowchart-like structure in which each internal node represents a test on a feature, in this research the two basic features are (the prediction of) enjoyment and frustration based on the data that we have collected from the sensors of mobile phones. Each leaf node represents a class label (decision taken after computing all features) and branches represent conjunctions of features that lead to those class labels. The paths from root to leaf represent classification rules. In the KNIME workflows below, we try to predict to what extent enjoyment and frustration were caused to users when using the psychometric game (Gyllensten & Bonomi, 2011).

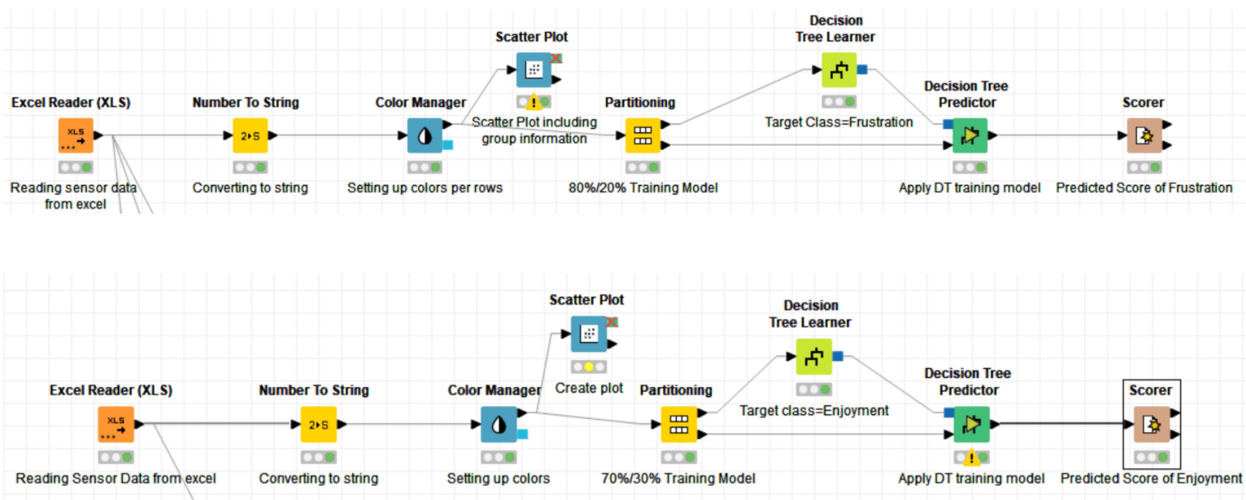


Figure 42: Decision Trees Model

Decision trees are designed using an algorithmic method to divide a data package into various conditions. It is one of the most used and practical learning processes. Decisions are a tool used for both classification and regression tasks which is non-parameter driven learning. Classification trees are called tree models where the destination variable can take a discreet collection of values. Decision trees where constant values (typically real numbers) can be taken from the target variable are called regression trees. The general term for this is Classification and Regression Tree (CART) (Quinlan, 1986).

6.4 Logistic Regression

Logistic Regression is a Machine Learning algorithm which is used for the classification problems, it is a predictive analysis algorithm and based on the concept of probability.

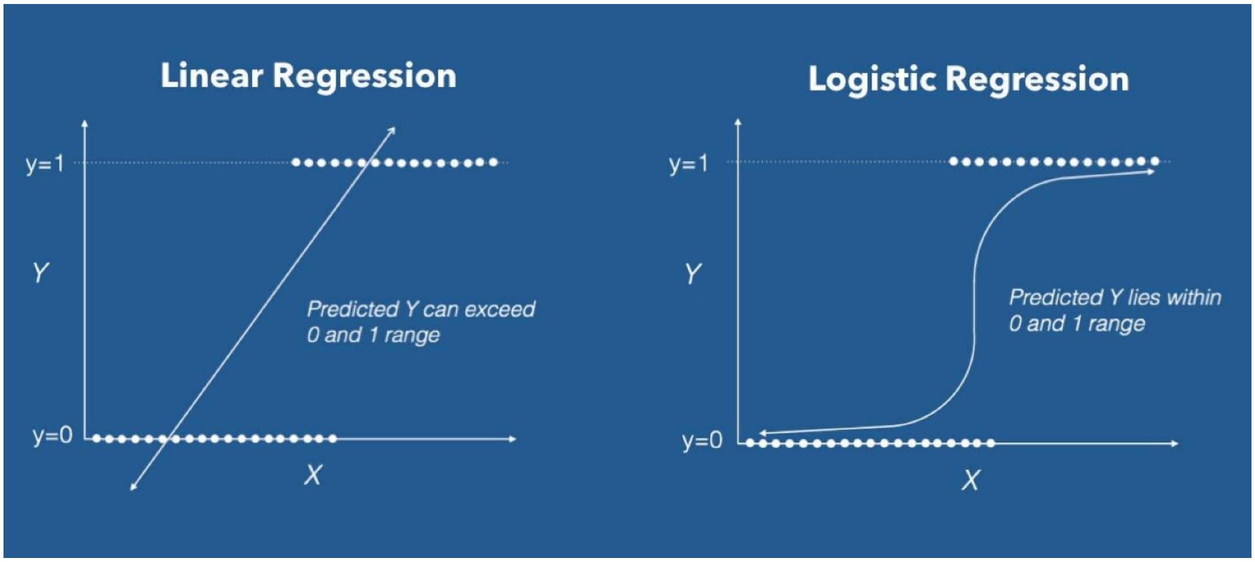


Figure 43: Linear Regression VS Logistic Regression Graph| Image: Data Camp

We can call a Logistic Regression a Linear Regression model, but the Logistic Regression uses a more complex cost function, this cost function can be defined as the 'Sigmoid function' or also known as the 'logistic function' instead of a linear function.

We might call a "Logistic Regression," but the "Logistic Regression" uses a more complex costing function, which can be described as the "Sigmoid Function" or as the "Logical Function" instead of a linear Function. The logistic regression hypothesis tends to reduce the cost function from 0 to 1. Linear functions thus fail to represent it, as a value greater than 1 or less than 0 is not feasible in accordance with the logistic regression hypothesis.

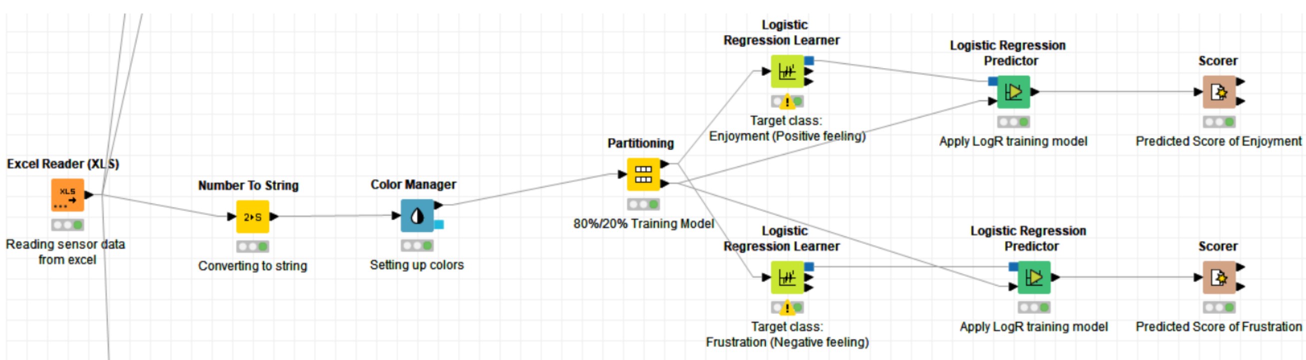


Figure 44: Logistic Regression Model

In the following workflow, we will see how we construct our flow on the platform KNIME and how it predicts value for our two big variables, enjoyment, and frustration.

6.5 Neural Network

Below we used the prediction model of Neural Networks based on which we checked if and to what extent the two variables of our research can be predicted. Below we have the workflow of the process based on which we received the results that you will see below.

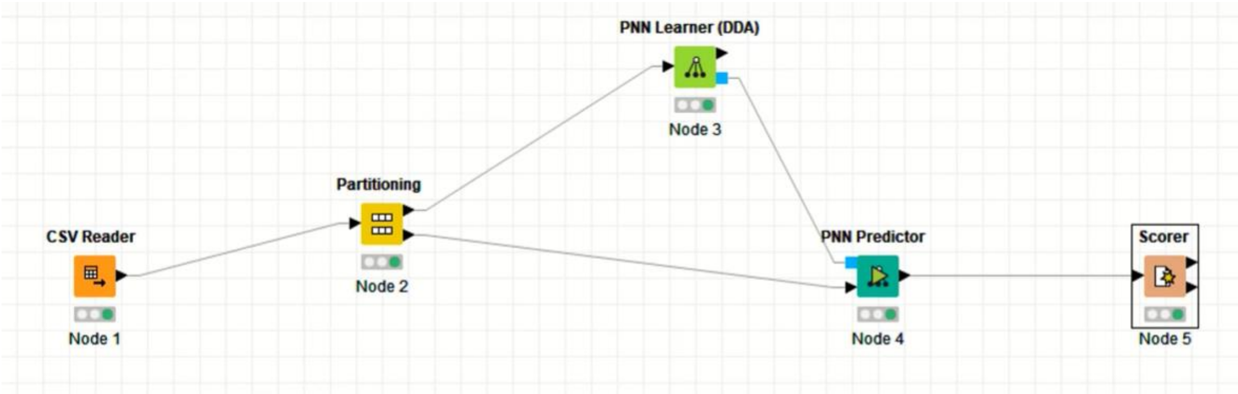


Figure 45: Neural Network Model

6.6 Support Vector Machine

The objective of the support vector machine algorithm is to find a hyperplane in an N-dimensional space (N — the number of features) that distinctly classifies the data points.

To separate the two classes of data points, there are many possible hyperplanes that could be chosen. Our objective is to find a plane that has the maximum margin, i.e., the maximum distance between data points of both classes. Maximizing the margin distance provides some reinforcement so that future data points can be classified with more confidence.

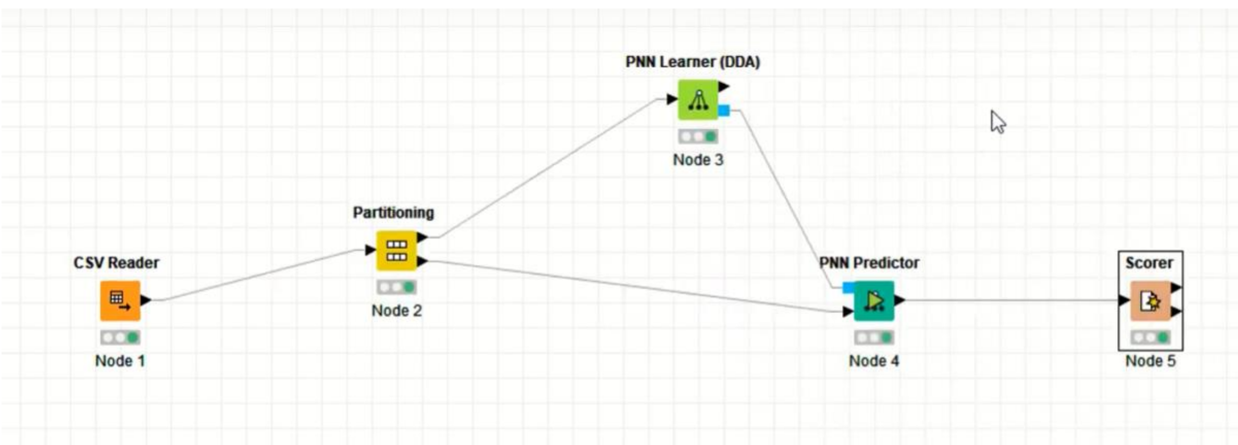


Figure 46: Support Vector Machine Model

6.7 Evaluation/Prediction Model Conclusions

From the application of the following algorithms, we obtained the following results regarding the prediction success of the two main variables that we are trying to predict. More specifically, it was observed that by running the Decision Tree at KNIME, we obtained the optimal prediction rates for the two basic human emotions we are trying to predict. In fact, with this method we exceeded the average percentage of accuracy for the detection of emotions using the sensors of the mobile and which is 77% for this method. Also, we must mention that the use of decision trees is specially to predict negative emotions, especially to detect stress, irritability, sadness, boredom, and positive emotions associated with enjoyment and frustration.

ML Algorithm	Enjoyment	Frustration
Linear Regression	65.13%	57.88%
Decision Trees	**87.90%**	**89.45%**
Logistic Regression	61.20%	42.12%
Support Vector Machine	72.08%	74.22%
Neural Network	77.16%	79.10%

Figure 47: ML Algorithm percentages

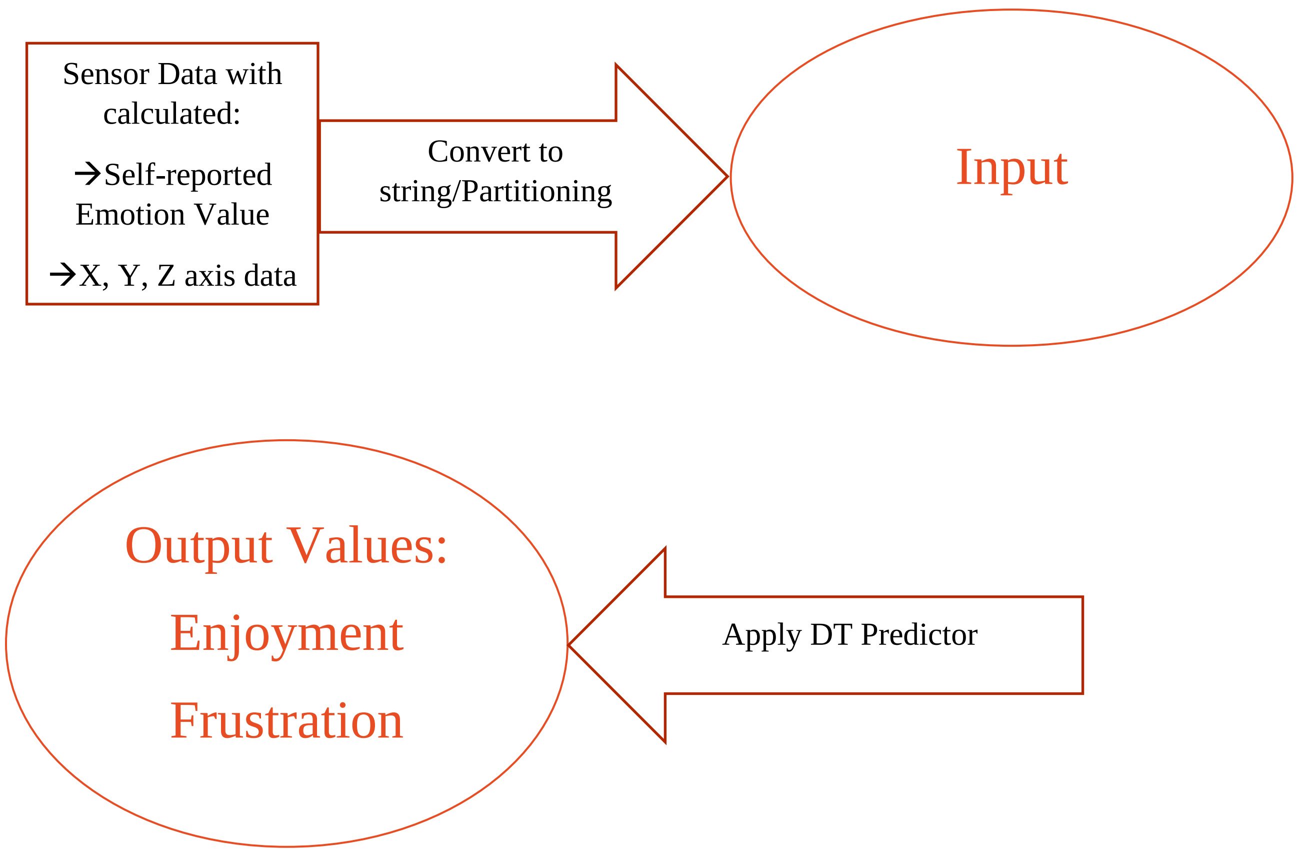
We realize that the decision tree has exceeded our expectations and is considered to be the most accurate model we have set up. It is important to note that the accuracy of the models found in the literature was influenced by two factors. The use of camera and accelerometer data to feed the model. In this research we have used only the accelerometer and gyroscope data in our research. This is also the reason why some forecasting models did not provide us with the expected accuracy that we thought we were starting the whole process and studying similar models (Tzafilkou, 2021).

Finally, in our research process to evaluate the model ultimately, we can choose the decision trees method based on the total performance percentages of the following prediction models. It seems we've achieved by studying the general constants when you study the insight of predicting human emotions using data from user mobile phone sensors in literature. We can therefore be sure about the way in which we will develop our final model and how we can evaluate this prediction models effectively.

7. Emotion Recognition Model

7.1 Model Selection

From the above analysis and presentation of the evaluation percentages of the models, we have selected the Decision Tree Algorithm for the prediction of both basic emotions. We have chosen this method because it gives us the most accurate and continuous training of the model that we have managed to achieve satisfactory accuracy in the data of this research. In the following diagram, we present in a simple way the mode of operation that feeds back the specific model. In fact, with this feed, we hope for a higher percentage of accuracy in the future, since this machine learning model can perform better as it is trained and practiced with real data.



This model needs two types of data to work:

→X, Y, Z axis data

→Self-reported emotion values

Feeding the model with these values begins to work, and the more data we give it, the more accuracy it returns to us. This input is vital for this model as it pushes it to work and improves its accuracy.

Secondly, this model uses the Decision Tree Predictor to try to predict the emotion value with the higher accuracy that it cans. This value is about the change in the mental state of the user in relation to the variables of enjoyment and frustration that we study. To calculate this value, all the data with which this model has been powered are utilized.

We conclude that this model can predict the self-reported emotion value for enjoyment with 87.90% accuracy and the self-reported emotion value for frustration with 89.45% accuracy.

This model was created with the ability to be continuously fed back and to try to increase its accuracy on a regular basis, if it is equipped with an appropriate form of data related to what we have seen above. Thus, one realizes that it can be used in similar research and gives an idea of how a model for predicting human emotions should work and be nurtured. Finally, one realizes that such models need a lot of data to be successful, something that we need to consider when dealing with the emotional change of users.

On the duration and process model training, it must be clarified that we have chosen 70 per cent of the sample to be used in the model and 30 per cent to be used for the needs of its preparation. Also, this process, as we are referring to a large amount of data, lasted ten days. This is because we have chosen to give time to this model and gradually monitor how it works to avoid providing the model with biased data. In conclusion, the model was developed gradually and with constant supervision of the entire feedback process.

Bibliographically, it is noted that in situations where there is a large amount of data, it is important to allow time for each new model to mature and gradually increase its accuracy.

8. Discussion

Some potential applications have been mentioned in the introduction and will not be further explored. There are, however, many opportunities to improve the results of classification and to adapt the methods to practical use.

Data from 40 users have been collected during the current research process. This is a satisfactory example based on what we study but a larger sample could be chosen in the future, as the data adds value to the model. If the volume of data were over doubled, we would have achieved a higher level of accuracy in the prediction models. Those who want to conduct future research in these fields should take that into consideration.

In retrospect, it is easy to see that additional incentives for participants to collect more data have been sufficient. However, it is difficult to recompense high levels of participation in this context, because the participants would not want to actively try to produce the application request information. This would counteract the idea of data collection and would possibly produce less data of a natural environment. If the project were to be resumed, the use of statistics on how their responses had progressed with the passage of time would have been interesting to implement. That is no reward per capita, and it does not promote high activity or competition, but it allows users to see outcome and adds a more autonomous reason to participate. More functionality may be added, taking account of user statistics, but it is also necessary to carry out these functions.

As a starting point for a later study, the predictor that has already been taught in this project. The model would initially adapt to the participant over time using the existing parameters and predict from the beginning. This can be used to show the user how the participation affects the workings of the algorithm.

The performance of the model depends heavily on training data of individual users. However, because of the wild nature of the study, a balanced dataset may not be possible. While some filters and some techniques have been used to overcome the imbalance problem, this should be carefully considered to obtain balanced data during such application deployment. Machine learning algorithms can be used specifically for unbalanced datasets (Chen et al., 2004; Tang et al., 2009).

One limitation of this work is that we assume that in a session the emotion of the user does not change. While we observe that 80% of the sessions are less than 4 minutes in our dataset, this assumption is probably untrue, but this may not always be true. The design of advanced ESM schedules to trace the emotional changes in the session can mitigate this problem (Vachon et al., 2019).

Furthermore, mobile user experience can be improved for better and smoother data collection. An additional possible limitation of these works.

In addition, people often do not have a pure kind of emotion. An individual can be excited and nervous when a new process is applied, for instance. For a more accurate and comprehensive emotional inference service it is therefore essential to describe and infer complex emotional statements containing multiple basic emotions at one time. Moreover, people are subjectively experiencing and expressing emotions. The intensity of the emotions expressed externally by people, such as facial expressions, is different. For behavioral responses such as device use, the difference is more obvious which suggests a more personalized deference model that best interprets an individual's emotional changes. It is time-consuming to collect enough personal data with it. As a result, we need techniques which can transmit knowledge from a variety of large collective information to a personalized model of inference, which can deduce user emotion in a cold start.

It is crucial to say that the deployment of emotion-sensitive applications on personal mobile devices challenges privacy issues. For training models, applications must collect information from their users. Because of their limited computational capability or battery power, devices can discharge the inferring task into cloud or another device. This means that user data may be revealed. Location, device use, data from the activity, calls and communications contain all the information to reveal the identity of a user. For this reason, the use of personal data by emotion sensing functionality is required by a uniform Privacy Protection Standard.

The model can be used in several ways. Once we can detect the emotional state of the user from his device interaction, we can change the look and feel at the interface in our effort to generate to the emotional state. Depending on the current state of user emotion we can also make changes in the way work is done. This can result in 'political' interfaces that are 'empathic' or 'friendly.' These qualities will in turn improve usability and improve user experience.

Finally, we should also mention a number of technical issues related to the model creation process. In particular, the creation of such a model of prediction requires the sacrifice of large volumes of data in order to train and begin to perform at a satisfactory pace. In fact, during this process, we had to modify some of the settings in our system and in knime platform so that we could use all the capacity of the computer's RAM to train and finally create the model.

9. Conclusions

Emotion recognition is a powerful and very useful technique for assessing human emotional states and predicting their behavior to provide the most appropriate advertising material in the marketing or educational field. In addition, the recognition and evaluation of emotions is very useful in the development of various human machine interaction systems. Relationships between particular emotions and human body reactions have been known for a long time, but many uncertainties remain in the selection of measurement and data analysis methods. There are some of the most widely used methods in this field, which are based on measurements of different parameters and an innumerable majority of methods of data analysis and practical applications.

This chapter sums up the results of this research and shows improvements to improve the results and to make the methods more practical.

There are several elements that have an impact on the resulting predicting rates:

- Calculation of the emotional mean value for both of our variables has an impact on the results.
- Individual movement bias appears to be less important than initially expected. However, this needs further exploration before any conclusion can be reached.
- Selection of features is generally important for the accuracy of our model, although a lot of the information needed to extract the expected emotion value score based on the data that we have collected from mobile sensors.
- The definition of emotion value helped in the course of the research and in the wider analysis of the data and the fluctuations of the change of the emotional state of the users
- The decision tree algorithm performs on the highest prediction model accuracies for our both variables. Our prediction model accuracy is 87.90% for enjoyment and 89.45% for frustration. Over time, there is a greater degree of accuracy when trying to predict negative emotions than positive ones.

While there are many improvements to be made, the results indicate that the emotional state of the person is reflected to some extent on the movements that the phone is subjected to while lying in the pocket of the person as he/she is walking. This is a highly non-obtrusive way of detecting emotions and does not require any additional equipment or interaction with the user.

The accelerometer-based method of emotional detection described in this thesis could therefore be useful in practical applications. It can either be part of a more complex system for detecting emotions, with potential high accuracy, or can be used on its own as a low-precision indicator of a person's emotional state.

Finally, the combination of these methods and the implementation of machine learning for data analysis appears to be an extremely powerful combination that will create breakthroughs in practical application in all fields, starting with advertising and marketing and finishing with industrial engineering applications.

References

A. Bogomolov, B. L. (2014). Daily stress recognition from mobile phone data, weather conditions and individual traits. *22nd ACM international conference on Multimedia*, 477–486.

A. Dzedzickis, A. K. (2020). *Human Emotion Recognition: Review of Sensors and Methods.* Vilnius: MDPI.

A. Gluhak, M. P. (2007). Towards mood based mobile services and applications. *In European Conference on Smart Sensing and Context*, 159-174.

Adibuzzaman, M. (2013). *In situ effect detection in mobile devices: a multimodal approach for advertising using social network.* ACM SIGAPP Applied Computing Review, p.67-77.

Ashwin T.S., G. R. (2020). *Impact of inquiry interventions on students in e-learning and classroom environments using affective computing framework.* Springer.

B. Cao, L. Z. (2017). *Deepmood: Modeling mobile phone typing dynamics for mood detection.*

B. Sun, Q. M. (2015). *Towards cold-start emotion labeling using transfer learning with smartphones, IEEE Conference on Computer Communications (INFOCOM)*, 1203–1211.

B. Sun, Q. M. (2015). iSelf system: Towards cold-start emotion labeling using transfer learning with smartphones. *IEEE Conference on Computer Communications (INFOCOM)*, 1203–1211.

Bakos, G. (2013). *KNIME Essentials-Packt Publishing.* Birmingham: Packt Publishing.

Bonomi, A. G. (2009). *Improving assessment of daily energy expenditure by identifying types of physical activity with a single accelerometer.* Journal of Applied Physiology.

Bonomi, I. C. (2011). *Identifying types of physical activity with a single accelerometer: evaluating laboratory- trained algorithms in daily life.* IEEE Transactions, p.2656–2663.

Breiman L., C. C. (2004). *Using random forest to learn imbalanced data .* J. Machine Learning Res. .

Céline Coutrix, N. M. (2012). *'Identifying emotions ex- pressed by mobile users through 2D surface and 3D motion gestures'.* Proceedings of the 2012 ACM Conference on Ubiquitous Computing, p. 311–320.

Cline, E. (2011). *Ready Pleayer One.* Crown Publishers.

Daniel Bernhardt, P. R. (2007). *Detecting Emotions from everybody body movaments.* Barcelona: Presencia PhD Sym. .

E. Diener, S. O. (2003). Personality, culture, and subjective well-being: Emotional and cognitive evaluations of life. *Annual review of psychology*, 403-425.

Ekman, P. (1992). An argument for basic emotions. *Cogn. Emotion*, 169-200.

Elimination, R. F. (n.d.). Retrieved from https://scikit-learn.org/stable/modules/generated/sklearn.feature_selection.RFE.html

Elizabeth Crane, M. G. (2007). *Motion capture and emotion: Affect detection in whole body movement.* International Conference on Affective Computing and Intelligent Interaction, p. 95–101.

Goertz, W. (2018). *Hype Cycle for Human-Machine Interface, 2018.* Gartner Research.

Gong, L. (2010). *Recognizing Affect from Non-stylized Body Motion Using Shape of Gaussian Descriptors.* Sierre, Switzerland: Proceedings of the 2010 ACM Symposium on Applied Computing. SAC '10, p. 1203–1206.

H. Fowler, N. A. (2008). Dynamic spread of happiness in a large social network: longitudinal analysis over 20 years in the framingham heart study. *British Medical Journal*, 337:a2338.

H. Leng, Y. L. (2007). An experimental study on physiological parameters toward driver emotion recognition. *In International Conference on Ergonomics and Health Aspects of Work with Computers*, 237-246.

Herdem, K. C. (2012). Reactions: Twitter based mobile application for awareness of friends' emotions. *2012 ACM Conference on Ubiquitous Computing,*, 796–797.

Hornbæk, A. M. (2016). An affect detection technique using mobile commodity sensors in the wild. *ACM International Joint Conference on Pervasive and Ubiquitous Computing*, 781–792.

Hugo Vachon, W. V.-G. (2019). *Compliance and Retention With the Experience Sampling Method Over the Continuum of Severe Mental Disorders: Meta-Analysis and Recommendations.* Leuven, Belgium: Journal of Medical Internet Research.

Hui, S. Z. (2014). A survey on mobile affective computing. *arXiv, Cornell University*.

Illapha Cuba Gyllensten, A. G. (2011). *Identifying types of physical activity with a single accelerometer: evaluating laboratory- trained algorithms in daily life.* Biomedical Engineering, IEEE Transactions, p. 2656–2663.

J. Ang, R. D. (2002). Prosody-based automatic detection of annoyance and frustration in human-computer dialog. *Interspeech*.

J. Hernandez, M. E. (2012). Mood meter: counting smiles in the wild. *ACM Conference on Ubiquitous Computing*, 301-310.

J. R. Crawford, J. D. (2004). The positive and negative affect schedule (panas): Construct validity, measurement properties and normative data in a large non-clinical sample. *British Journal of Clinical Phychology*, 245-265.

Jaques, N. (2015). Predicting students happiness from physiology, phone, mobility, behavioral data, Affective Computing and Intelligent Interaction, IEEE.

Jaques, N. (2015). *Multitask, Multi-Kernel Learning for Estimating Individual Wellbeing.* Montreal: Workshop on Multimodal Machine Learning.

Jiayu Shu, M. C. (2019). *Emotion Sensing for Mobile Computing.* IEEE Communications Magazine.

John, J. J. (2003). Individual differences in two emotion regulation processes: implications for affect, relationships, and well-being. *Journal of personality and social psychology*, 85(2):348.

Jonathan Posner, J. A. (2005). The circumplex model of affect: An integrative approach to affective neuroscience, cognitive development, and psychopathology. *Development and psychopathology*, 715-734.

Kanjo E., Y. E. (2019). *Deep learning analysis of mobile physiological, environmental and location sensor data for emotion detection.* ScienceDirect.

Kathrin Melcher, R. S. (2020). *Codeless Deep Learning with KNIME.* PACKT Publishing LTD. .

Lövheim, H. (2012). *'A new three-dimensional model for emotions and monoamine neurotransmitters'.* p. 341–348.: Medical hypotheses.

M. Zhao, F. A. (2016). Emotion recognition using wireless signals. *Annual International Conference on Mobile Computing and Networking* , 95-108.

Mourato, G. M. (2013). Happiness is greater in natural environments. *Global Environmental Change*, 23(5):992–1000.

Musolesi, L. C. (2015). Trajectories of depression: unobtrusive monitoring of depressive states by means of smartphone mobility traces analysis. . *2015 ACM international joint conference on pervasive and ubiquitous computing*, 1293–1304.

P. Ekman, W. V.-T.-B. (1987). Universals and cultural differences in the judgments of facial expressions of emotion. *Journal of personality and social psychology*, 53(4):712.

Picard, R. W. (1995). *Affective Computing.* M.I.T Media Laboratory Perceptual Computing Section Technical Report.

Ping Wang, L. D. (2020). *Clustering-Based Emotion Recognition Micro-Service Cloud Framework for Mobile Computing.* IEEE Communications Magazine.

Plutchik, R. (1980). A general psychoevolutionary theory of emotion. *Elsevier*, 1(3-31):4.

Poria S., C. E. (2017). *A review of affective computing.* ScienceDirect.

Preece, S. J. (2009). *Activity identification using body-mounted sensors—a review of classification techniques.* Physiological meas.

Quinlan, J. R. (1986). *Induction of decision trees.* Machine Learning, p.81-106.

R. LiKamWa, Y. L. (2013). Moodscope: building a mood sensor from smartphone usage patterns. *Proceeding of the 11th annual international conference on Mobile systems, applications, and services*, (pp. 389-402).

R. Reisenzein, M. S. (2013). Coherence between emotion and facial expression: Evidence from laboratory experiments. *Emotion Review*, 5(1):16–23.

R.J. Larsen, D. B. (2014). *Personality Phychology: Domains of Knowledge about Human Nature.* 5th edition McGraw-Hill.

S. Du, Y. T. (2014). Compound facial expressions of emotion. *Proceedings of the National Academy of Sciences*, (pp. 111(15):E1454–E1462).

S. Ghosh, S. S. (2019). An Emotion-aware Smartphone Keyboard for Mental Health Monitoring. *11th International Conference on Communication Systems & Networks (COMSNETS)*, 496-499.

S. Li, L. H. (2013). Sentence-level emotion classification with label and context dependence. *ACL*, 1045–1053.

S. Servia-Rodriguez, K. K. (2017). Mobile sensing at the service of mental well-being: a large-scale longitudinal study. *26th International Conference on World Wide Web*, 103–112.

S. Zhao, H. Y. (2015). Predicting continuous probability distribution of image emotions in valence-arousal space. *ACM international conference*, 879–882.

Sachin Shah, J. N. (2015). Towards affective touch interaction: predicting mobile user emotion from finger strokes. *Journal of Interaction Science*, 3-6.

Surjya Ghosh, K. H. (2019). Emotion detection from touch interactions during text entry on smartphones. *Intenational Journal of Human-Computer Studies*, 47-57.

T. Stutz, T. K. (2015). Smartphone based stress prediction. *In International Conference on User Modeling Adaptation, and Personalization*, 240–251.

Tang Y., Z. Y. (2009). *SVMs modeling for highly imbalanced classification.* IEEE Trans. Syst. Man. Cybernet p.281-288.

Tzafilkou, K. (2021). *Mobile sensing for emotion recognition: A literature review on non-instructive methodologies and m-learning contexts.* Thessaloniki: Unplublished work/under review.

Wahid A., R. F. (2019). *Emotion Detection Using Unobtrusive methods: An Integrated Approach.* 2nd International Conference on Intelligent Computing, Instrumentation and Control Technologies.

X. Bao, S. F. (2013). Your reactions suggest you liked the movie: Automatic content rating via reaction sensing. *ACM international joint conference on Pervasive and ubiquitous computing*, 197-206.

Xiao Zhang, W. L. (2017). MoodExplorer: Towards Compound Emotion Detection via Smartphone Sensing. *Association for Computing Machinery*, 176.

Y. Gao, N. B.-B. (2012). What does touch tell us about emotions in touchscreen-based game-paly? *ACM Trans. Computer-Human Interaction*, 31.

Y. Suhara, Y. X. (2017). Deepmood: Forecasting depressed mood based on self-reported histories via recurrent neural networks. . *26th International Conference on World Wide Web*, 715–724.

Y. Zhang, J. T. (2010). Emotion prediction via dynamic continuous factor graph model. *International Conference on Data Mining*, 1193-1198.

Y. Zhou, H. X. (2015). Emotion distribution recognition from facial expressions. *23rd ACM international conference on Multimedia*, 1247–1250.

Z. Zeng, M. P. (2009). A survey of affect recognition methods: Audio, visual, and spontaneous expressions. *IEEE transactions on pattern analysis and machine intelligence*, 31(1):39–58.

Zhang, Z. (2016). *Emotion recognition based on customized smart bracelet with built-in accelerometer.* PeerJ PrePrints 4.

(n.d.). Retrieved from https://www.pokemongo.com/en-us/

(n.d.). Retrieved from http://www.appinventor.org/Architecture2

(n.d.). Retrieved from https://play.google.com/store/apps/details?id=com.arcticshores.skyrisecity&hl=en&gl=US

(n.d.). Retrieved from https://en.wikipedia.org/wiki/Fourier_transform

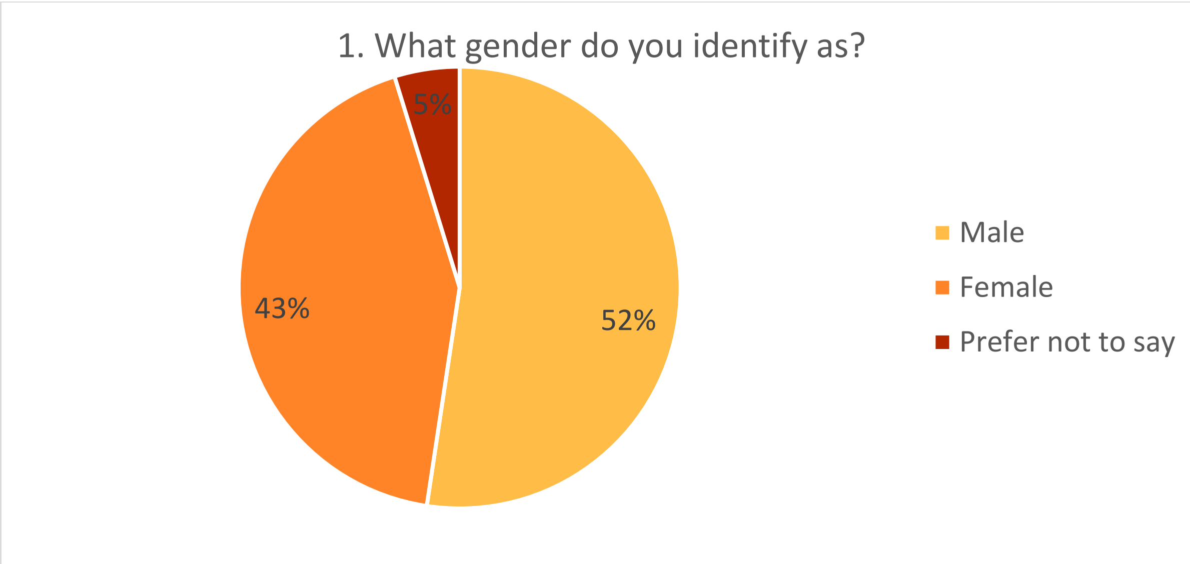
(n.d.). Retrieved from https://en.wikipedia.org/wiki/Kurtosis

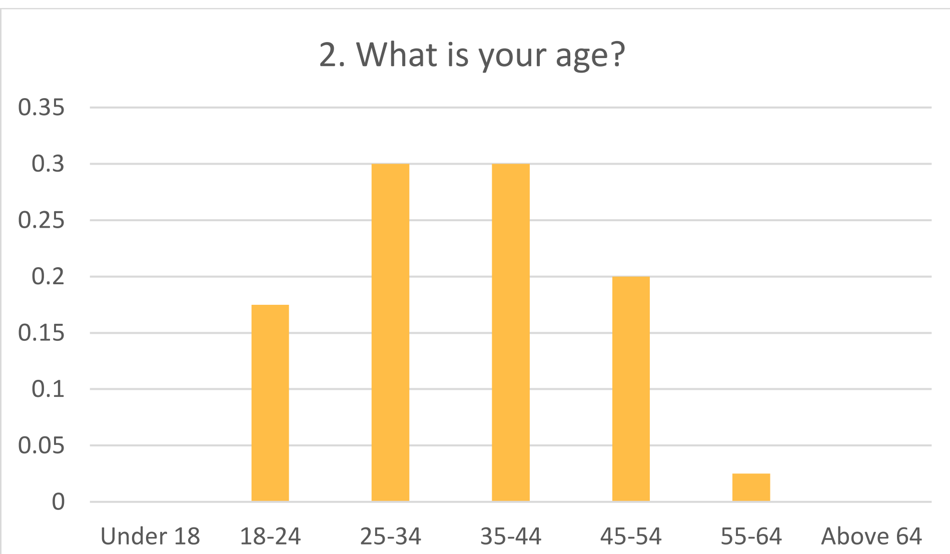
Appendix: Statistics

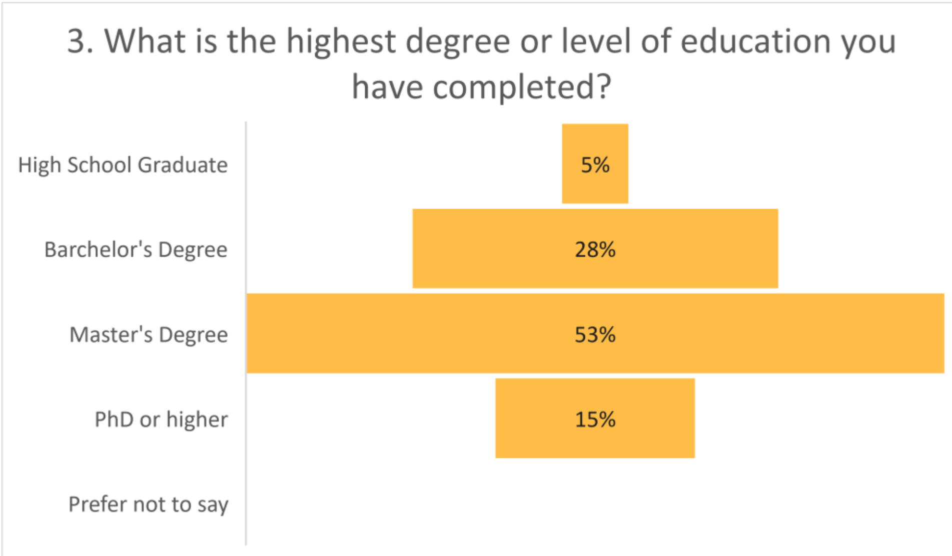
In the following appendix there have been the results of

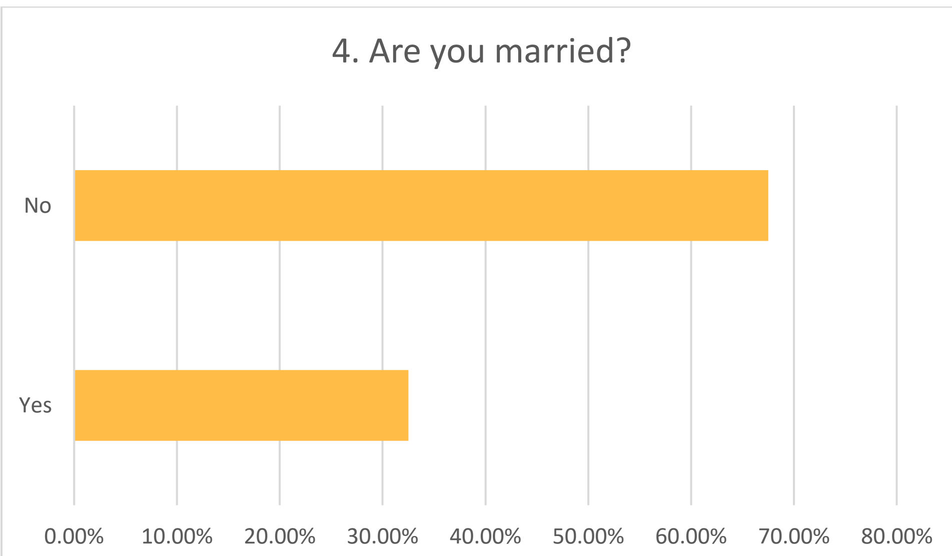
- Pre-Survey Questions: 1 to 11
- Post-Game Questions: 12 to 22

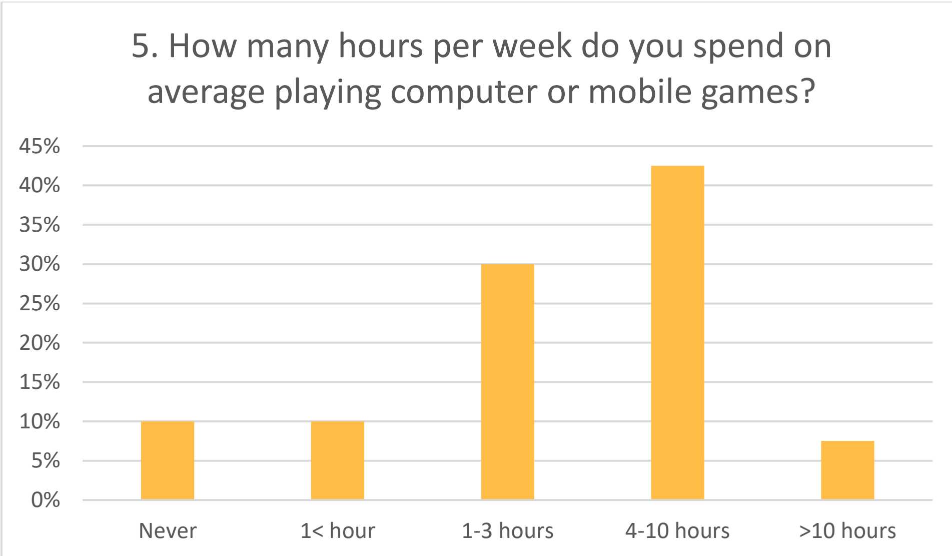
with their graphical representations.

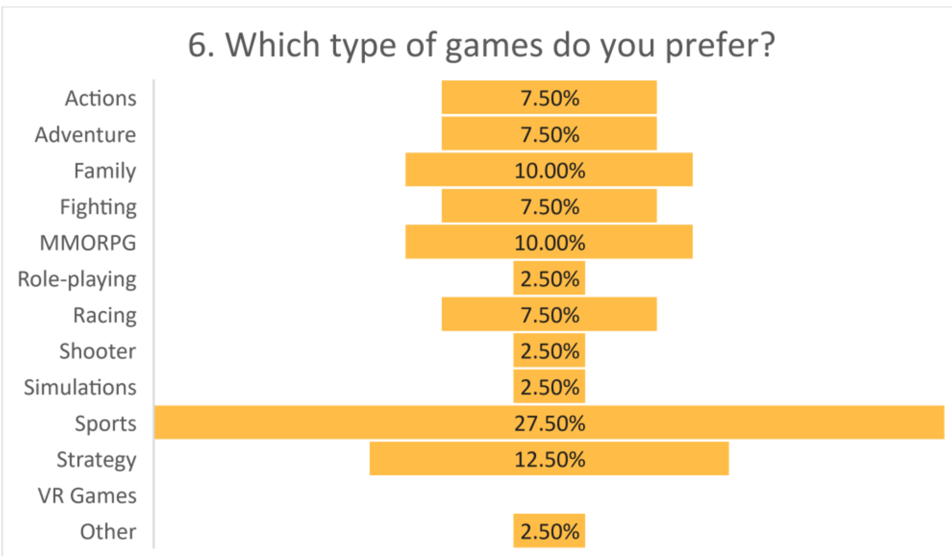


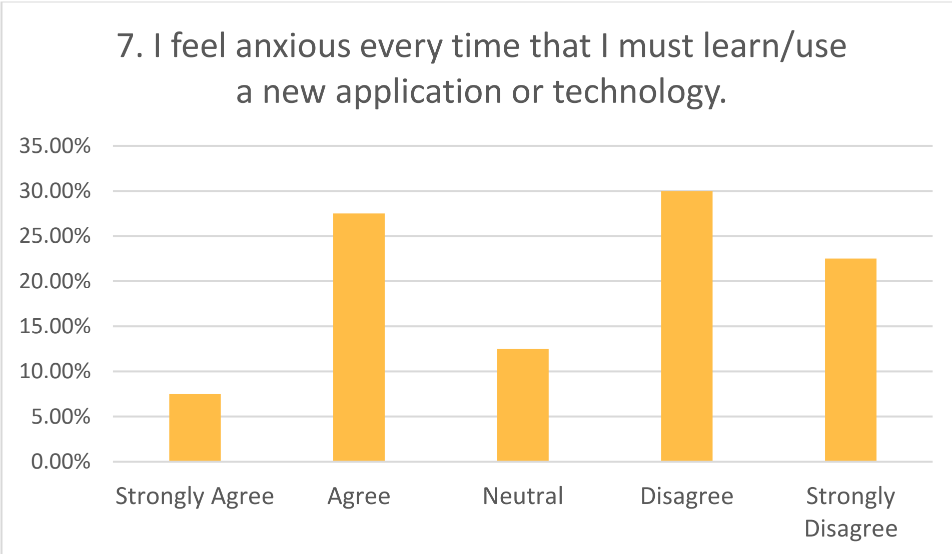


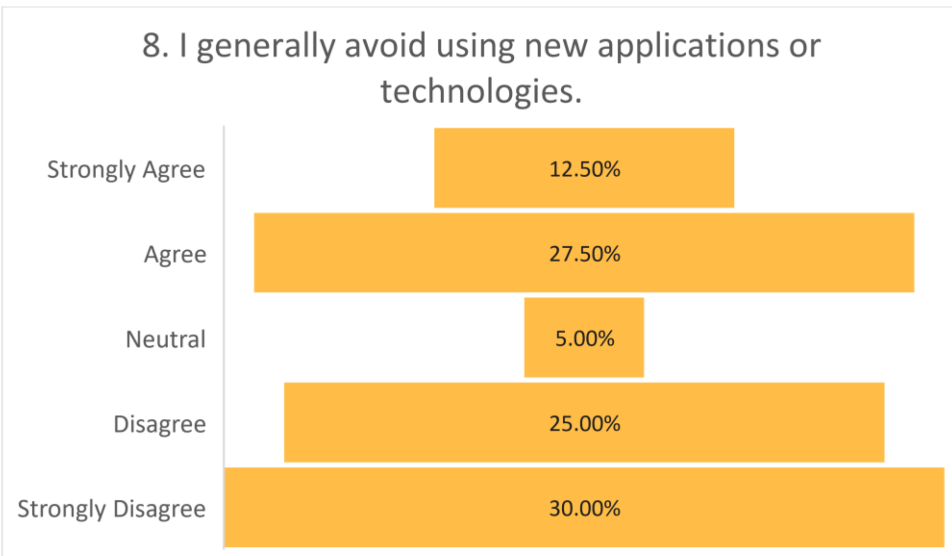

3. What is the highest degree or level of education you have completed?
High School Graduate
Barchelor's Degree
Master's Degree
PhD or higher
Prefer not to say
5%
28%
53%
15%

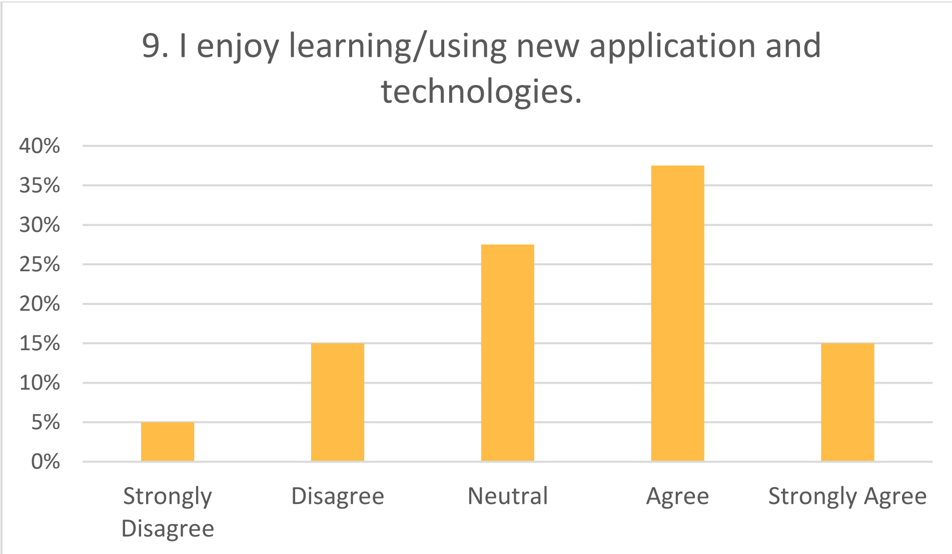

4. Are you married?
No
Yes
0.00%
10.00%
20.00%
30.00%
40.00%
50.00%
60.00%
70.00%
80.00%


5. How many hours per week do you spend on average playing computer or mobile games?
45%
40%
35%
30%
25%
20%
15%
10%
5%
0%
Never
1< hour
1-3 hours
4-10 hours
>10 hours

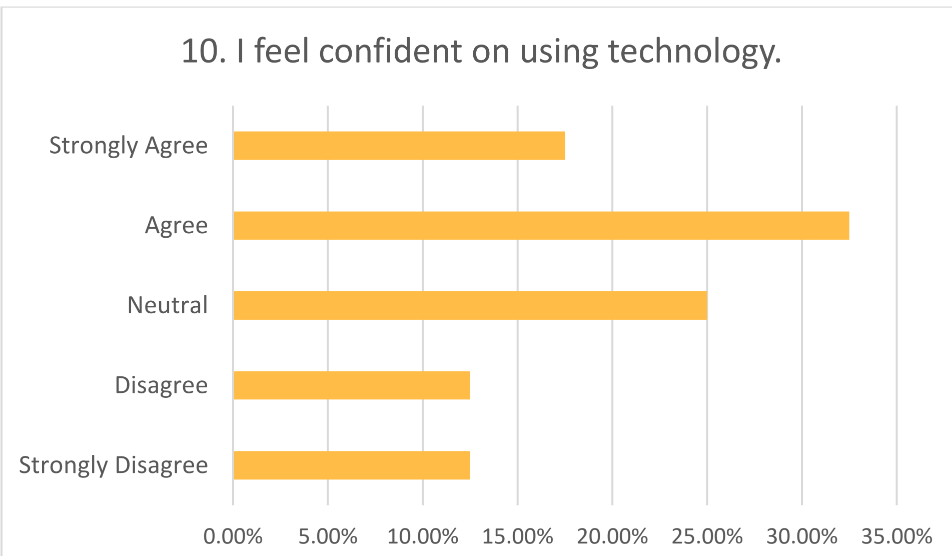

6. Which type of games do you prefer?
Actions
Adventure
Family
Fighting
MMORPG
Role-playing
Racing
Shooter
Simulations
Sports
Strategy
VR Games
Other
7.50%
7.50%
10.00%
7.50%
10.00%
2.50%
7.50%
2.50%
2.50%
27.50%
12.50%
2.50%


7. I feel anxious every time that I must learn/use a new application or technology.
35.00%
30.00%
25.00%
20.00%
15.00%
10.00%
5.00%
0.00%
Strongly Agree
Agree
Neutral
Disagree
Strongly Disagree

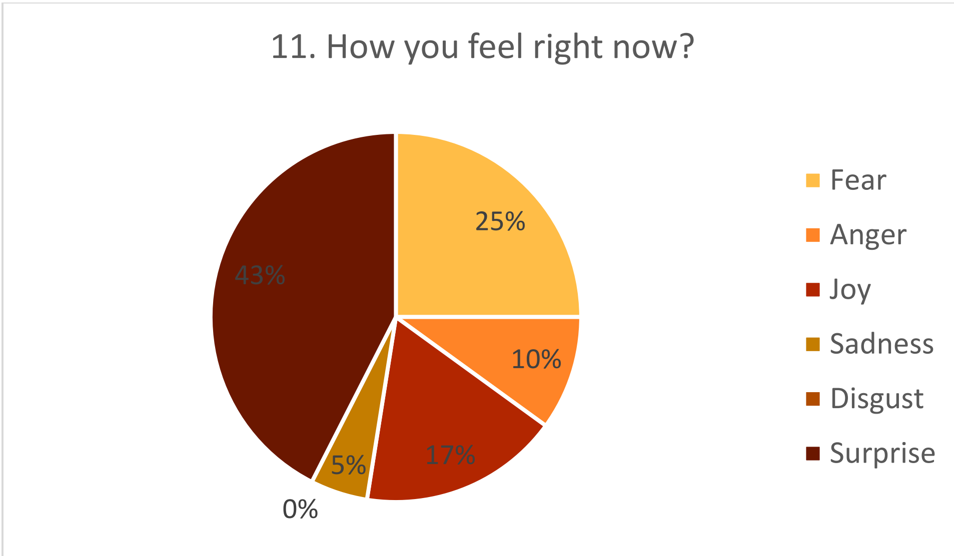

8. I generally avoid using new applications or technologies.
Strongly Agree
12.50%
Agree
27.50%
Neutral
5.00%
Disagree
25.00%
Strongly Disagree
30.00%

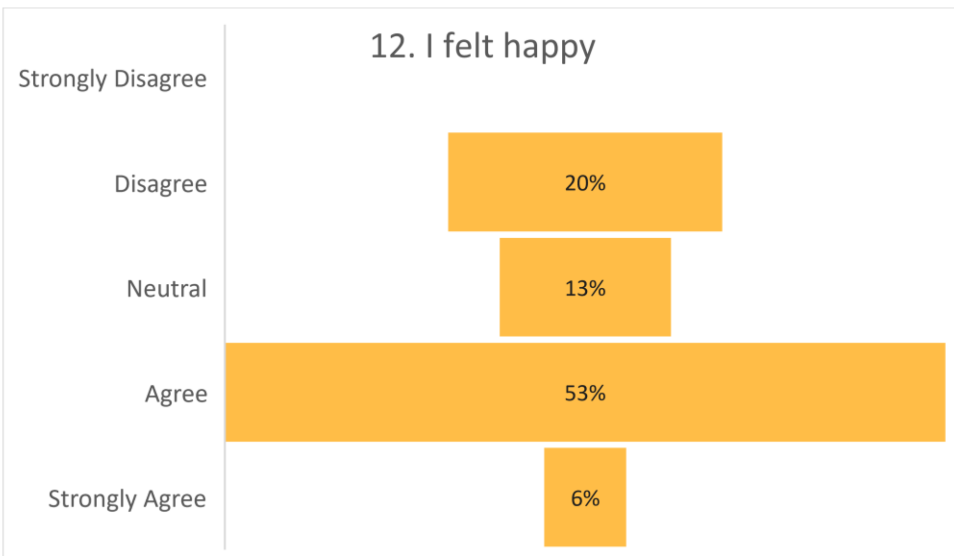


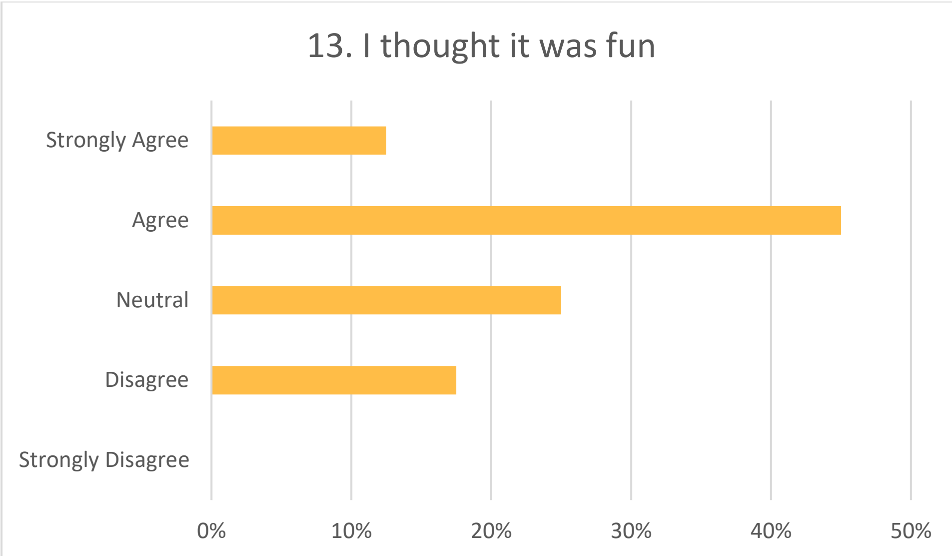
9. I enjoy learning/using new application and technologies.
40%
35%
30%
25%
20%
15%
10%
5%
0%
Strongly Disagree
Disagree
Neutral
Agree
Strongly Agree

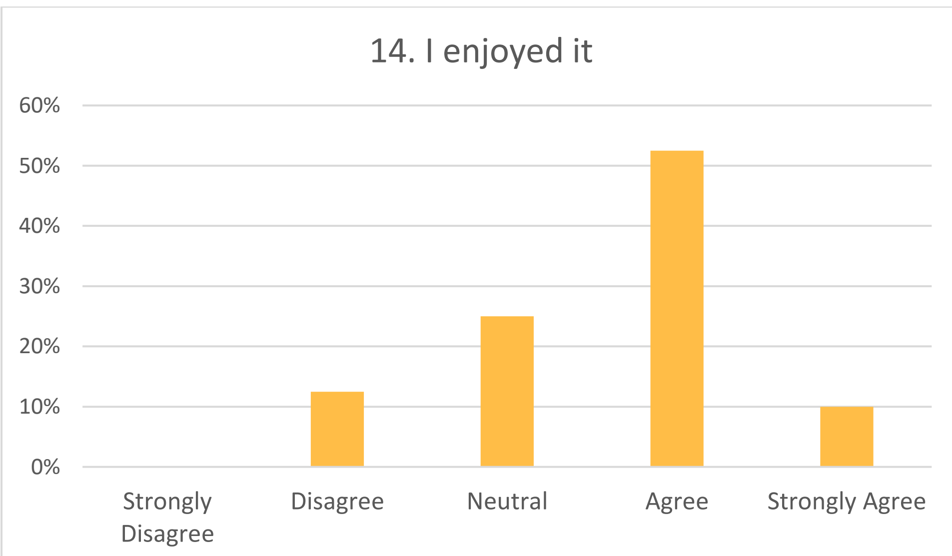


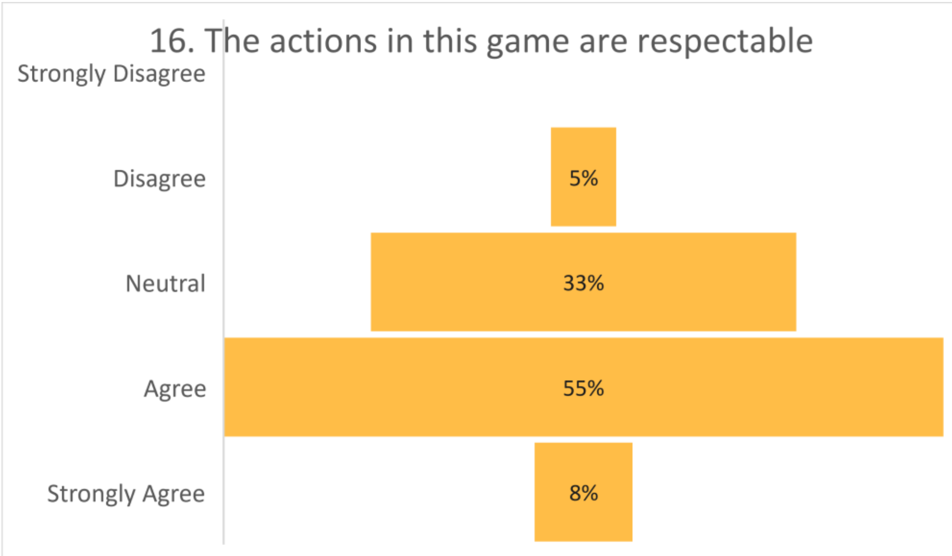
10. I feel confident on using technology.
Strongly Agree
Agree
Neutral
Disagree
Strongly Disagree
0.00%
5.00%
10.00%
15.00%
20.00%
25.00%
30.00%
35.00%

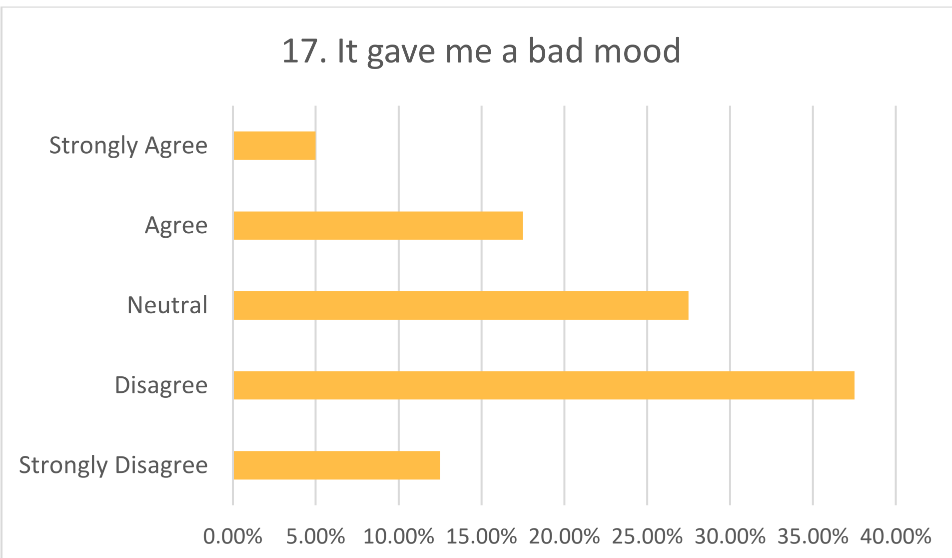

11. How you feel right now?
25%
43%
10%
17%
5%
0%
Fear
Anger
Joy
Sadness
Disgust
Surprise

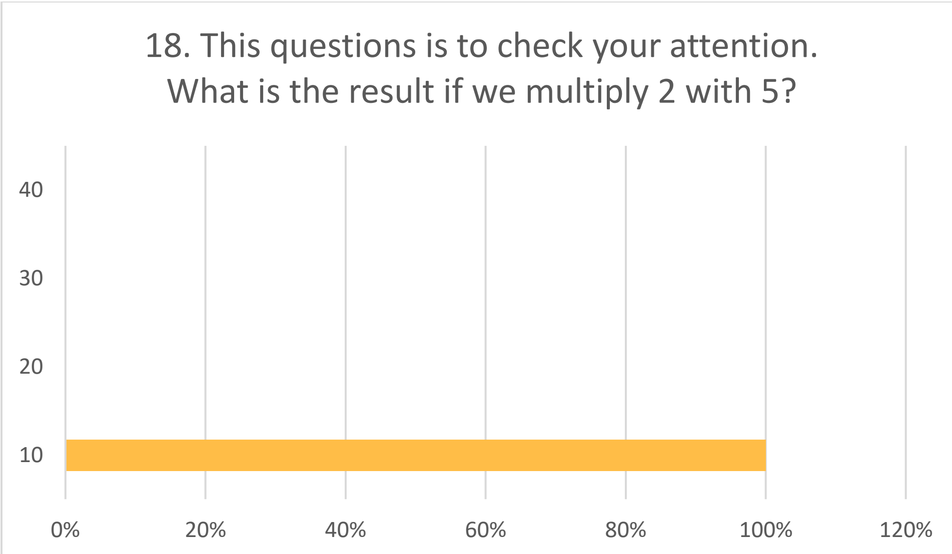

12. I felt happy
Strongly Disagree
Disagree
Neutral
Agree
Strongly Agree
20%
13%
53%
6%


13. I thought it was fun
Strongly Agree
Agree
Neutral
Disagree
Strongly Disagree
0%
10%
20%
30%
40%
50%

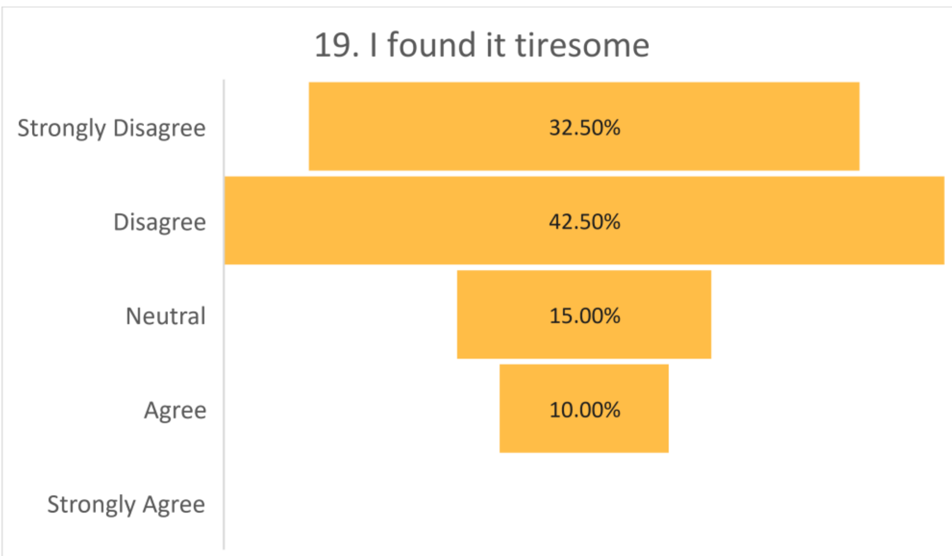

14. I enjoyed it
60%
50%
40%
30%
20%
10%
0%
Strongly Disagree
Disagree
Neutral
Agree
Strongly Agree


16. The actions in this game are respectable
Strongly Disagree
Disagree
5%
Neutral
33%
Agree
55%
Strongly Agree
8%

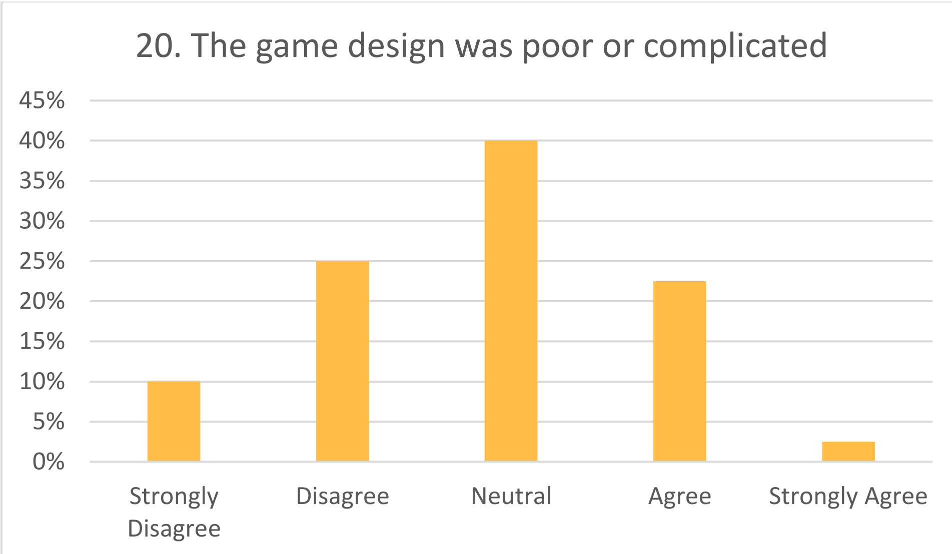

17. It gave me a bad mood
Strongly Agree
Agree
Neutral
Disagree
Strongly Disagree
0.00%
5.00%
10.00%
15.00%
20.00%
25.00%
30.00%
35.00%
40.00%

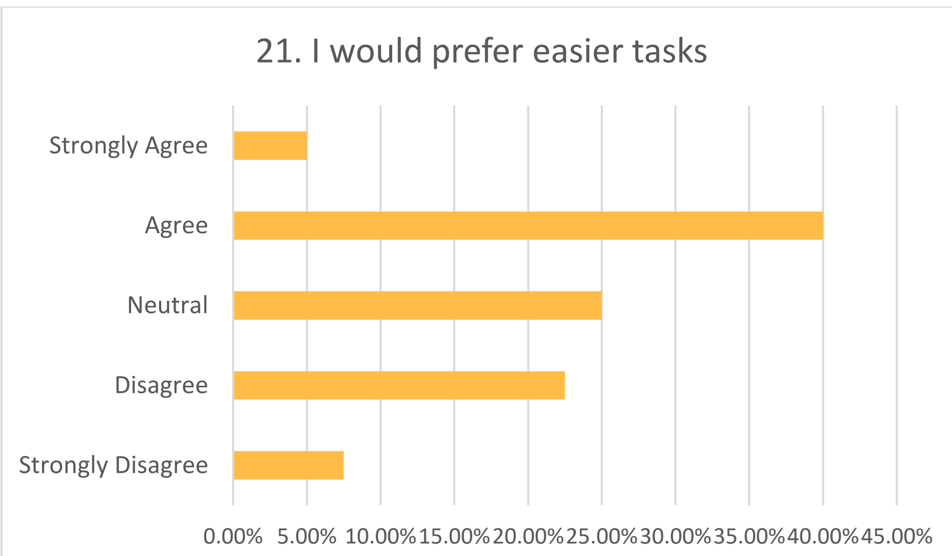


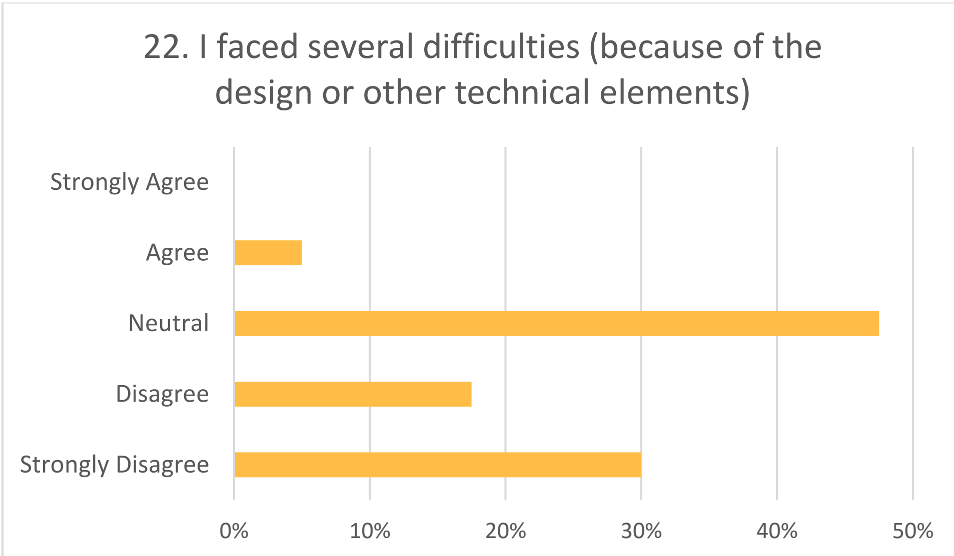
18. This questions is to check your attention.
What is the result if we multiply 2 with 5?
40
30
20
10
0%
20%
40%
60%
80%
100%
120%



19. I found it tiresome
Strongly Disagree
32.50%
Disagree
42.50%
Neutral
15.00%
Agree
10.00%
Strongly Agree


20. The game design was poor or complicated
45%
40%
35%
30%
25%
20%
15%
10%
5%
0%
Strongly Disagree
Disagree
Neutral
Agree
Strongly Agree


21. I would prefer easier tasks
Strongly Agree
Agree
Neutral
Disagree
Strongly Disagree
0.00%
5.00%
10.00%
15.00%
20.00%
25.00%
30.00%
35.00%
40.00%
45.00%


22. I faced several difficulties (because of the design or other technical elements)
Strongly Agree
Agree
Neutral
Disagree
Strongly Disagree
0%
10%
20%
30%
40%
50%

Cronbach's Alpha Statistical Test for Enjoyment

Reliability

Scale: ALL VARIABLES

Case Processing Summary

		N	%
Cases	Valid	40	100.0
	Excluded[a]	0	.0
	Total	40	100.0

a. Listwise deletion based on all variables in the procedure.

Reliability Statistics

Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
.755	.738	5

Item Statistics

	Mean	Std. Deviation	N
Ifeelhappy	3.63	.979	40
Ithoughtitwasfun	3.53	.933	40
Ienjoyedit	3.60	.841	40
Playingthisgamemakesme moreintelligent	4.00	.877	40
Theactionsinthisgamearere spectable	3.65	.700	40

Inter-Item Correlation Matrix

	Ifeelhappy	Ithoughtitwasfun	Ienjoyedit	Playingthisgamemakesmoreintelligent	Theactionsinthisgamearerespectable
Ifeelhappy	1.000	.558	.529	.478	.028
Ithoughtitwasfun	.558	1.000	.666	.501	.092
Ienjoyedit	.529	.666	1.000	.556	.148
Playingthisgamemakesmemoreintelligent	.478	.501	.556	1.000	.042
Theactionsinthisgamearerespectable	.028	.092	.148	.042	1.000

Summary Item Statistics

	Mean	Minimum	Maximum	Range	Maximum / Minimum	Variance
Item Means	3.680	3.525	4.000	.475	1.135	.034
Inter-Item Correlations	.360	.028	.666	.638	23.732	.059

Summary Item Statistics

	N of Items
Item Means	5
Inter-Item Correlations	5

Item-Total Statistics

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Squared Multiple Correlation
Ifeelhappy	14.77	5.871	.580	.385
Ithoughtitwasfun	14.87	5.702	.674	.512
Ienjoyedit	14.80	5.959	.709	.534
Playingthisgamemakesmemoreintelligent	14.40	6.297	.571	.369
Theactionsinthisgamearerespectable	14.75	8.705	.093	.027

Item-Total Statistics

	Cronbach's Alpha if Item Deleted
Ifeelhappy	.689
Ithoughtitwasfun	.649
Ienjoyedit	.642
Playingthisgamemakesme moreintelligent	.692
Theactionsinthisgamearere spectable	.827

Scale Statistics

Mean	Variance	Std. Deviation	N of Items
18.40	9.579	3.095	5

Cronbach's Alpha Statistical Test for Frustration

Reliability

Scale: ALL VARIABLES

Case Processing Summary

		N	%
Cases	Valid	40	100.0
	Excluded[a]	0	.0
	Total	40	100.0

a. Listwise deletion based on all variables in the procedure.

Reliability Statistics

Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
.704	.709	5

Item Statistics

	Mean	Std. Deviation	N
Itgavemeabadmood	2.65	1.075	40
Ifoundittiresome	2.03	.947	40
Thegamedesignwaspoororc omplicated	2.83	.984	40
Iwouldprefereasiertasks	3.13	1.067	40
Ifacedseveraldifficultiesbec auseofthedesignorothertec	2.28	.960	40

Inter-Item Correlation Matrix

	Itgavemeabadmood	Ifounditiresome	Thegamedesignwaspoororcomplicated	Iwouldprefereasiertasks
Itgavemeabadmood	1.000	.563	.304	.196
Ifounditiresome	.563	1.000	.445	.276
Thegamedesignwaspoororcomplicated	.304	.445	1.000	.266
Iwouldprefereasiertasks	.196	.276	.266	1.000
Ifacedseveraldifficultiesbecauseofthedesignorothertec	.021	.359	.486	.366

Inter-Item Correlation Matrix

	Ifacedseveraldifficultiesbecauseofthedesignorothertec
Itgavemeabadmood	.021
Ifounditiresome	.359
Thegamedesignwaspoororcomplicated	.486
Iwouldprefereasiertasks	.366
Ifacedseveraldifficultiesbecauseofthedesignorothertec	1.000

Summary Item Statistics

	Mean	Minimum	Maximum	Range	Maximum / Minimum	Variance
Item Means	2.580	2.025	3.125	1.100	1.543	.191
Inter-Item Correlations	.328	.021	.563	.542	26.667	.023

Summary Item Statistics

	N of Items
Item Means	5
Inter-Item Correlations	5

Item-Total Statistics

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Squared Multiple Correlation
Itgavemeabadmood	10.25	8.192	.371	.385
Ifoundittiresome	10.88	7.548	.612	.456
Thegamedesignwaspoororcomplicated	10.08	7.712	.540	.344
Iwouldprefereasiertasks	9.78	8.179	.379	.173
Ifacedseveraldifficultiesbecauseofthedesignorothertec	10.63	8.343	.426	.375

Item-Total Statistics

	Cronbach's Alpha if Item Deleted
Itgavemeabadmood	.694
Ifoundittiresome	.594
Thegamedesignwaspoororcomplicated	.622
Iwouldprefereasiertasks	.690
Ifacedseveraldifficultiesbecauseofthedesignorothertec	.669

Scale Statistics

Mean	Variance	Std. Deviation	N of Items
12.90	11.631	3.410	5