KAI: An AI-powered Chatbot To Support Therapy

Monitoring Report

Mariama C. Djalo D.

Bachelor Thesis in Computer Science
Specialization in Computing

Director: Javier Béjar Alonso
Fall, 2022

Contents

1 Introduction **5**
- 1.1 Context . . . 5
- 1.2 Concepts . . . 6
  - 1.2.1 Chatbot . . . 6
  - 1.2.2 Cognitive Behavioral Therapy . . . 6
  - 1.2.3 Artificial Intelligence . . . 8
  - 1.2.4 NLP . . . 8
  - 1.2.5 Machine Learning . . . 9
  - 1.2.6 Text Classification . . . 9
  - 1.2.7 Bias . . . 9
  - 1.2.8 Data Crowdsourcing . . . 9
- 1.3 Justification . . . 10
- 1.4 Scope . . . 10
  - 1.4.1 Objectives . . . 10
- 1.5 Requirements . . . 11
- 1.6 Risks . . . 11
- 1.7 Methodology . . . 11
- 1.8 Stakeholders . . . 12

2 Initial Project Planning **13**
- 2.1 Description of tasks . . . 13
  - 2.1.1 Project Management . . . 13
  - 2.1.2 Project Research . . . 14
  - 2.1.3 Project Theory . . . 14
  - 2.1.4 Data Generation . . . 15
  - 2.1.5 Project Experimentation . . . 15
  - 2.1.6 Project Development . . . 15
  - 2.1.7 Project Documentation . . . 16
  - 2.1.8 Thesis Defense Preparation . . . 16
- 2.2 Risk management: alternative plans . . . 16

3 Initial Budget **21**
- 3.1 Costs Per Activity . . . 21
- 3.2 Generic Costs . . . 22
- 3.3 Contingency . . . 22
- 3.4 Incidental Costs . . . 22
- 3.5 Management control . . . 22

4 Sustainability report **23**
- 4.1 Self assessment . . . 23
- 4.2 Economic dimension . . . 23
- 4.3 Environmental dimension . . . 23
- 4.4 Social dimension . . . 24

5 Final Project Planning **26**
- 5.1 Deviations in the project development . . . 27
- 5.2 Deviations in the project documentation . . . 27

6 Final Methodology **29**

7 Technical Competences **29**

8 Identification of Laws and Regulations **30**
8.1 Academic Regulations for the Degree Final Project . . . 30
8.2 BotUI License . . . 30
8.3 GDPR Privacy Policy . . . 30
8.3.1 Company's contact details . . . 30
8.3.2 The Purposes and Legal Basis for Processing . . . 30
8.3.3 Sharing of user's personal data . . . 31
8.3.4 Sharing of user's personal data to a third country . . . 31
8.3.5 Period of time storage of user's personal data . . . 31
8.3.6 User's Rights . . . 31
8.4 The EU Regulatory Environment of Medical Device Software Development . . . 31

9 Dialogue Flow **35**
9.1 Identifying Potential Cognitive Distortions . . . 37
9.2 Challenging Potential Cognitive Distortions . . . 37
9.3 Relaxation Exercises . . . 37

10 Supervised Machine Learning Algorithms **37**
10.1 Multinomial Naive Bayes . . . 37
10.2 Multinomial Logistic Regression . . . 38
10.3 Support Vector Machine . . . 38
10.4 XGBoost (XBGClassifier) . . . 39
10.5 AdaBoost (AdaBoostClassifier) . . . 39
10.5.1 AdaBoost vs XGBoost . . . 40
10.6 Random Forest . . . 40
10.7 LightGBM (LGBMClassifier) . . . 41
10.8 K-nearest neighbor . . . 41
10.8.1 Hyperparameters . . . 42
10.9 Performance . . . 42
10.9.1 Bias . . . 42

11 Dataset **42**
11.1 Data Generation . . . 42

12 Implementation **42**

13 Adversities **42**

List of Figures

1 A drawing that depicts how a situation causes feelings, from the book *A Therapist's Guide to Brief Cognitive Behavioral Therapy* [1]. 7
2 Diagram illustrating the structure of the Cognitive Model, from the book *A Therapist's Guide to Brief Cognitive Behavioral Therapy* [1]. 7
3 A diagram showing the different subfields of Artificial Intelligence, from *EuropeanValley* [2], a website. 8
4 ML model workflow overview, from *Google Machine Learning Education* [3], a website. 9
5 Gantt Chart illustrating the project's schedule following a Waterfall model. [Own Creation] . 19
6 15 major cognitive distortions by PositivePyschology, a website [4] 20
7 EU Regulatory Environment for MDSW Development. [5] 33
8 Caption 36

List of Tables

1 Task Table containing a summary of all task information. T and GEPT means Tutor and GEP Tutor respectively. [Own Creation] 18
2 Salary of the different roles extracted from PayScale, a compensation software company [6] multiplied by 1.35 to include the cost of social security [Own creation]. 21
3 Budget Structure of the project [Own creation] 25
4 Task Table showing the progress of the project [Own creation] 26
5 Final version of the task table [Own creation] 27
6 Final version of the budget structure [Own creation] 28
7 Standard and guidance documents useful to demonstrate MDSW compliance with MDR. [5] . 34

1 Introduction

This is a Bachelor Project Thesis done at the Barcelona School of Informatics (FIB) under the supervision of Javier Béjar Alonso for the Computer Science Degree with a specialization in Computing.

This project aims to address the problem of a significant gap that exists between people who are struggling with mental health issues and need assistance and those who have access to mental health care. The design and implementation of an AI-powered chatbot that supports therapy by automatically detecting cognitive distortions is planned to achieve this goal. The goal of this project is to train various supervised machine learning techniques for text classification in order to identify cognitive distortions from chatbot conversations and use them to aid in self-directed CBT.

The project is aimed at people who meet all of the following conditions:

- People who have mild to moderate symptoms of mental health issues and can function normally. [1]
- People who do not have access to or prefer not to speak with a therapist due to privacy concerns and/or a fear of being judged as a result of the stigma associated with mental health issues.
- People who prefer to be autonomous and prefer to learn and treat themselves with self-directed therapy.

1.1 Context

According to WHO [7], depression is a leading cause of disability and suicide is the fourth leading cause of death among people aged 15-29. Many mental health conditions can be effectively treated at low cost, but there is still a significant gap between those who need care and those who have access to care.

Therapeutic and mental health Chatbots are an intriguing way to bridge this gap. Chatbots, also known as conversational user interfaces, are tools that simulate human communication via voice or text using machine learning and artificial intelligence.

On the other hand, it is widely known that a few sessions of cognitive-behavioral therapy (CBT) can be extremely beneficial in the treatment of anxiety and depression. However, many people do not have access to a CBT therapist because they cannot afford it, it is not covered by their insurance, or there are none nearby. It can also be difficult to go to therapy due to lack of time because of work or having to take care of kids for instance. However, in some cases, a therapist may not be required. There are numerous options for doing CBT without a therapist, such as self-help books and internet-based treatment, or by conducting your own research on the material. Self-directed CBT has been shown in numerous studies to be very effective [8]. This particular type of therapy promotes independence and self-therapy. The methodology of CBT is based on **identifying cognitive distortions**, known as negative thoughts and **challenging them** by replacing them for alternative and more positive thoughts in order to **improve mood and the overall mental health**.

The goal of this project is to combine the two previously mentioned methodologies to help people dealing with mental health issues. KAI, a Chatbot that provides therapy support in self-directed CBT, will be designed and implemented as part of the project. KAI will detect Cognitive Distortions in text using Text Classification, an NLP technique that will assist in the application of CBT.

Bear in mind that **the chatbot is not a therapist and thus does not provide mental health diagnosis**; rather, the chatbot is a tool that assists with the automatic detection of potential cognitive distortions and guides you through the process of identifying and challenging them. In other words, KAI is a self-directed AI-based CBT tool that assists people in learning how to deal with cognitive distortions.

[1] Those who are severely depressed or have severe mental health problems will most likely require one-on-one therapy with a professional

1.2 Concepts

Before proceeding with the document, a few key concepts may require clarification. Thus, all of the key concepts required to comprehend the project are covered in this section.

1.2.1 Chatbot

1.2.2 Cognitive Behavioral Therapy

The goal of cognitive-behavioral therapy (CBT), a type of psychotherapy, is to help patients understand how their thoughts, beliefs, and attitudes affect how they feel and behave. The foundation of cognitive behavioral therapy (CBT) is the premise that our thoughts, feelings, and behaviors are connected, and that by altering our thoughts and beliefs, we can alter our behaviors and emotions.

With the aim of recognizing and altering negative thought and behavior patterns, CBT often entails systematic, goal-oriented therapy with a therapist. Together, the client and therapist determine the issues facing the client and establish clear, quantifiable therapy objectives. The therapist next works with the client to put these techniques into practice in daily life after helping the client build strategies to accomplish these goals.

The fact that CBT is an evidence-based treatment, which means it has been proven successful in numerous scientific research, is one of its distinguishing characteristics. The efficacy of CBT in treating a variety of mental health issues, such as depression, anxiety, eating disorders, substance misuse, and personality disorders, has been repeatedly demonstrated by research. Additionally, it has been proven to be successful in addressing physical health issues like irritable bowel syndrome and chronic pain.

The fact that CBT is a relatively short-term therapy, with most courses lasting between 10 and 20 sessions, is one of its benefits. This enables many people to afford it as a treatment choice. Additionally, CBT is a highly adaptable therapeutic strategy that can be customized to each person's unique needs. It can be carried out in a variety of situations, including face-to-face, over the phone or video, or using self-help resources like books and websites.

CBT is a successful treatment, but it is not a one-size-fits-all approach, and not everyone is a candidate for it. The structured, goal-oriented aspect of CBT may be challenging for some people, while it may be challenging for others to recognize and question their negative ideas and actions. Alternative therapy techniques might be more suitable in these circumstances.

Self-Directed Therapy

Self-directed cognitive-behavioral therapy (CBT) enables patients to work alone, with little to no supervision from a therapist, to recognize and modify the unhelpful thoughts and behaviors that contribute to their emotional and behavioral issues. If a patient prefers to work alone or does not have access to a therapist, self-directed CBT may be an useful treatment alternative.

Using self-help tools like books, websites, or apps, self-directed CBT often entails the individual being guided through the process of recognizing and altering harmful thinking and behavior patterns. These resources frequently provide details on CBT principles and methods in addition to exercises and other activities that can be used to assist the person apply these ideas to their personal situation.

The ability to work at one's own pace and in one's own setting is one of the benefits of self-directed CBT, which can be especially beneficial for those who may not have the time or means to attend regular therapy sessions. It can also be a more affordable alternative to regular therapy because self-help books and other resources are usually less expensive.

However, not everyone should use self-directed CBT. Some people might find it challenging to seek treatment on their own, especially if they are dealing with serious mental health issues or if they find it challenging to recognize and resist their negative thoughts and behaviors. In these situations, working with a therapist or looking into alternate therapy alternatives might be more effective.

In general, self-directed CBT can be a useful therapeutic option for patients who favor working alone or who might not have access to a therapist. Before deciding on self-directed CBT as a kind of treatment, it is crucial for people to thoroughly assess their own requirements and preferences.

Cognitive Model

The use of cognitive models is crucial in CBT. These models offer a framework for comprehending how psychological issues arise and how they might be handled and serve to explain how a person's thoughts and beliefs influence their emotions and behaviour.

The cognitive model is a theoretical paradigm for explaining how thoughts, feelings, and behaviors are associated. Most individuals believe that their emotions are caused by situations as shown in Figure 1.

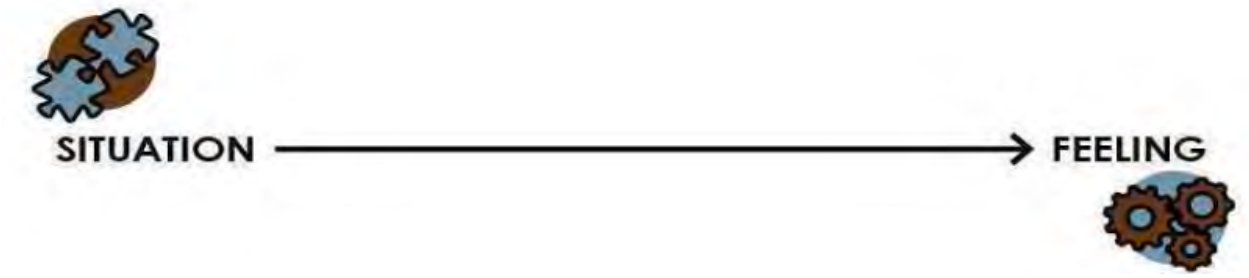


Figure 1: A drawing that depicts how a situation causes feelings, from the book *A Therapist's Guide to Brief Cognitive Behavioral Therapy* [1].

The cognitive model challenges this subjective experience and suggests, instead, that it is the thoughts we have about situations that give rise to emotions as illustrated in Figure 2. The cognitive model of depression, created by Aaron T. Beck, is one well-known cognitive model that is frequently applied in CBT. This paradigm claims that depressed individuals have cognitive biases that lead them to view the world negatively and pessimistically. These unfavorable cognitive biases can produce unfavorable thoughts, which in turn might produce unfavorable behaviors and emotions.

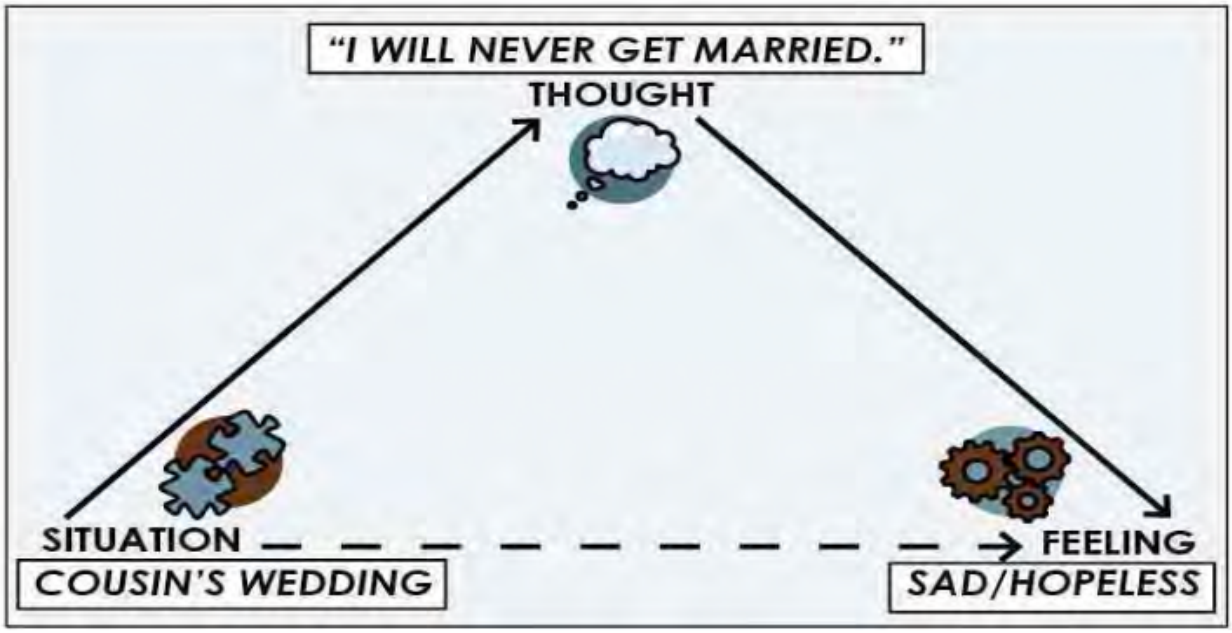


Figure 2: Diagram illustrating the structure of the Cognitive Model, from the book *A Therapist's Guide to Brief Cognitive Behavioral Therapy* [1].

Cognitive Distortions

Cognitive distortions are distorted, biased, or unreasonable ways of thinking that can influence negative feelings and actions. These distortions affect how we view and understand events, which might drive us to make unreasonable or unhealthy judgments that have harmful consequences.

Typical examples of cognitive distortions are as follows:

- **All-or-nothing thinking**: Seeing things in black-and-white terms, with no shades of gray. For example, *"I am a complete failure because I didn't pass the exam on my first try."*
- **Jumping to conclusions**: Drawing judgments without enough evidence. For example, *"She's not responding so she must be ignoring me or even worse, she must be mad at me"*
- **Mental filter**: Focusing only on the negative part of everything. For example, *"I did well on most of the exam, but I got one question wrong, so I must be a complete failure."*
- **Emotional reasoning**: Thinking that because you feel something it must be real. For example, *"I feel like a failure, so I must be a failure."*

Individuals can learn to identify and confront their negative thoughts and actions by becoming aware of and addressing cognitive distortions, which will improve their emotional and behavioral results. CBT is a type of psychotherapy that is especially created to assist people in recognizing and altering cognitive distortions and other harmful thought and behavior patterns.

1.2.3 Artificial Intelligence

Artificial intelligence (AI) refers to the ability of a machine or computer system to perform tasks that typically require human intelligence, such as learning, problem-solving, and decision-making. In other words, computer systems that mimic human intelligence. AI applications include face recognition, chatbots, and self-driving cars.

1.2.4 NLP

Natural language processing (NLP), as illustrated in Figure 3, is a subfield of artificial intelligence. NLP focuses on teaching computers to comprehend text and spoken language in the same way that humans do.

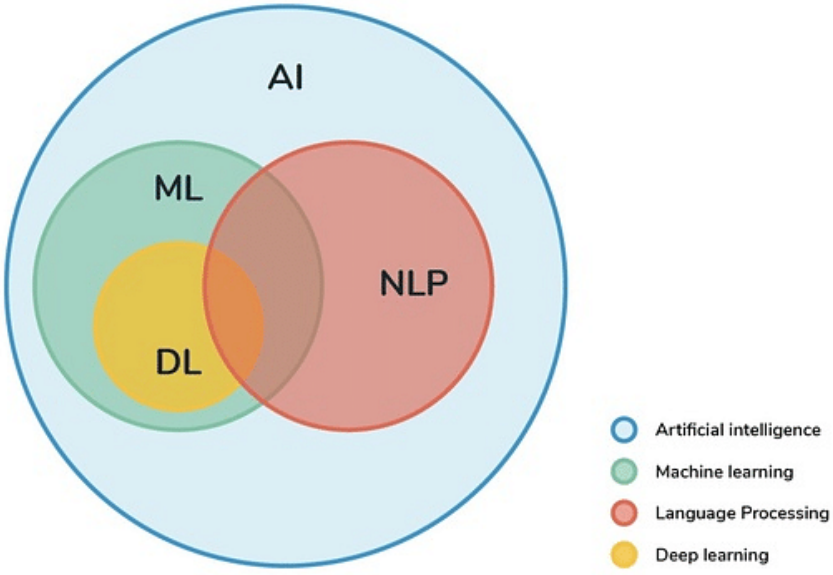


Figure 3: A diagram showing the different subfields of Artificial Intelligence, from *EuropeanValley* [2], a website.

1.2.5 Machine Learning

Machine Learning, as shown in Figure 3, is another subfield of AI that focuses on the development of algorithms and models that can learn from data and improve their performance over time.

A dataset, which consists of input data and matching labels or output data, is used to train a machine learning model. The objective is to learn a function that can translate data from the input to the output. The performance of the model is then assessed on unseen data.

There are two main types of machine learning: supervised learning and unsupervised learning. In supervised learning, the model is trained on labeled data, where the correct output is provided for each example in the training set. The goal is to learn a function that can predict the output for a new input. In unsupervised learning, the model is not provided with labeled training examples, but must discover the underlying structure of the data through techniques such as clustering.

1.2.6 Text Classification

Text Classification is a supervised machine learning technique in NLP used to classify text into different categories. Common types of Text Classification are Sentiment Analysis and Spam Email detector.

As shown in Figure 4, the Text Classification Workflow begins with gathering data (obtaining a dataset) and continues with data exploration using descriptive analytics. The next step is to prepare the data. Since I'll be working with texts and ML models don't understand them, I'll need to do Text Vectorisation, which is the process of converting text into numerical representation.

Following that, the model is built, trained with the dataset, and evaluated to see how well it performs. In my case, I'll train several models before selecting the best one.

Finally, after Hyperparameters Tuning for improved performance, the model is deployed. Hyperparameters are the parameters that define the model architecture, and Hyperparameter Tuning is the process of searching for the best model architecture. The maximum depth is an example of a Hperparamter in decision trees (a ML Model).

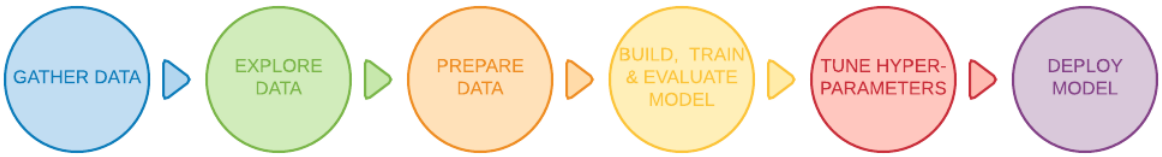


Figure 4: ML model workflow overview, from *Google Machine Learning Education* [3], a website.

1.2.7 Bias

A bias is a strong inclination or prejudice for or against someone or something. Humans are naturally biased; it's a natural response, and thus **humans are the ultimate source of bias in machine learning**. Because all models are created by humans, they all have human biases. Even the most reliable model will have many biases.

Despite the fact that this problem exists, we should avoid bias in machine learning models to avoid inaccurate predictions or predictions that may violate the law or be unethical. The COMPAS model, which is widely used in the United States, is the most well-known example of a heavily biased ml model. A regression model is used in this system to predict whether or not a perpetrator will commit another crime again. The model predicted that African Americans would have twice as many false positives for recividism as Caucasians.

1.2.8 Data Crowdsourcing

Crowdsourced data collection is a collaborative method of creating a dataset with the assistance of a large group of people. This method is extremely useful when there is no available dataset for the problem you

wish to solve or when you require assistance labeling data.

1.3 Justification

The use of AI in mental healthcare and psychiatry, particularly therapeutic chatbots, is still in its early stages, with limited research data and datasets available to fully explore the field's true potential.

Shickel et al. [9] published the most relevant research in 2019. The authors used supervised machine learning techniques such as SVM, XGBoost, and RNN to find cognitive distortions in text automatically. The authors obtained a weighted F1 scores of 0.68.

Since the detection of cognitive distortions is a novel machine learning task, there are currently no publicly accessible datasets with labelled text sections with distortions. In order to solve this challenging issue, they gathered data through crowdsourcing and a real-world online therapy program.

Toledo et al. [10], on the other hand, took a different and very innovative approach: they developed cognitive distortion responses to support CBT interactions. Which is really relevant because one of the central concepts of Cognitive Behavioural Therapy (CBT) is the ability to convert negative or distorted thoughts (cognitive distortions) into more realistic alternatives (positive) ones. The authors used Transformers learners to generate the responses.

It's also worth mentioning that there are self-help apps and chatbots for mental health, many of which use CBT, on the market, including in Spain[2]. Wysa and Woebot are the most well-known.

As mentioned before, there are currently no publicly available datasets. Crouwdsourcing and data generation are two options for resolving this challenging issue. Since Crowdsourcing is a paid service, I will opt for the second option, in other words: I will create the dataset myself. The process will be completed by providing sufficient examples of the 15 major cognitive distortions. To ensure that the data is accurate, examples will be drawn or be inspired from psychological books, websites, and articles specializing in psychology.

Unfortunately there is no public information available on how to design a CBT chatbot. Therefore, the design will be done from scratch with the aid of a very helpful and introductory manual book to CBT [1] that will help have a better understanding about the structure of cognitive behavioural therapy session.

1.4 Scope

In this section, the objectives, the functional requirements, and the potential obstacles and risks that may arise during the project are explained.

1.4.1 Objectives

The main objective of this project is to design and implement an AI-powered Chatbot that supports therapy to assist people suffering from mental health issues such as anxiety or depression. Kai will use CBT technique and with the help of AI he will automatically detect cognitive distortions. To achieve this goal, the project has been divided into several sub-objective:

Theoretical part

1. Investigate CBT to gain a thorough understanding of how it works.
2. Design the chatbot's conversation structure and dialogue flow.
3. Investigate the best supervised machine learning techniques for text classification.

Practical Part

1. Generate a suitable dataset to train the machines.
2. Train different supervised machine learning models.
3. With the help of metrics, choose the best model.
4. Implement the chatbot.

[2]This fact confirms and clarifies that it is perfectly legal to make **support therapy chatbots**

1.5 Requirements

There are some requirements that must be met in order to ensure the final project's quality:

- **Data pre-processing.** Data pre-processing is essential for ensuring that the models perform optimally.
- **Hyperparameter Tuning.** Understanding how different algorithms work is essential for knowing how to make good Hyperparameter Tuning decisions to achieve the best results.
- **Optimization of the code.** Optimize the code for all imputation methods as well to improve efficiency.
- **Avoid bias.** Avoiding bias is key to ensure that the models are accurate.
- **Good programming practices.** Using good programming practices such as a readable style, good comments, and as little complexity as possible.

1.6 Risks

There may be some risks that prevent the project from progressing smoothly:

- **Not being able to generate an appropriate or a representative dataset to train models.** The main risk would be not being able to find/generate a suitable (without bias) dataset to train the models, which would prevent the project from progressing.
- **Deadline of the project.** The project also has a deadline for completion that must be met. This forces to make difficult decisions. As a result, strong organizational skills and the ability to meet deadlines are essential for finishing the project on time. Some libraries, on the other hand, have flaws.
- **Bugs in libraries.** Some libraries may have bugs in certain functions, resulting in incorrect code.

1.7 Methodology

In order to have more flexibility, I will use a hybrid of Waterfall and Agile workflow methodology for this project. Waterfall is a project management methodology that is based on a sequential design process. Agile is a methodology that prioritizes development through evolution. This method enables sprint work and the resolution of issues that arise during iterations. The Agile-Waterfall hybrid method combines the best features of both methods: Agile allows you to check for bugs, test the code, and correct it in a progressive way without having to wait until the entire implementation is completed. Waterfall, on the other hand, allows you to keep track of all the dependencies between tasks to better organize the project.

The Kanban framework which falls under the Agile methodology will be used. Since the 1950s, the Japanese phrase "kanban," which means "visual board" or a "sign," has been used to refer to a process definition. Toyota invented it and used it as the first just-in-time factory scheduling system. The "Kanban Method," which was initially defined in 2007, is known and connected with the capitalized term "Kanban," on the other hand.

Kanban boards are used to efficiently show and control the workflows. The essential elements are:

- **Kanban Cards** are used to express tasks visually. Each card contains details on the work and its progress, including the due date, the person assigned to it, the description, etc.
- **Kanban Columns**: On the board, each column corresponds to a distinct step of your operation. The workflow is applied to the cards till they are finished completely.

In my case, I'll have four columns.

1. **To Do:** composed by all of the tasks that haven't been started yet.
2. **In progress:** composed by all of the tasks that are still in progress.

3. **Tested:** composed by all of the tasks that have already been completed but need to be tested to ensure they work properly.

4. **Completed:** composed by all completed and tested tasks.

There are many project management tools that follow the kanban methodology, but I'm going to use Trello because I believe it's the better option for small teams or one-person teams, such as freelancers (which is my case).

On the other hand, a Gantt chart with a Waterfall workflow will also be used to keep track of the dependencies and the required time for each task.

I'll use a Github repository as a version control tool to make sure I can restore earlier versions in the event of serious errors because it's securely kept in the cloud.

The GitHub Flow is the Git branch strategy that I'm going to employ. Its structure is made up of a main branch where the production-ready code is placed and other branches, called feature branches, which contains work on new features and bug fixes and will be merged back into the main branch once the work is complete and properly reviewed. Smaller teams, which, as I previously stated, is my case, benefit the most from this method.

For the practical part, the **cross-validation** method will be used to choose the optimum **Hyper-parameters** for the models and also to check their **performance** and if there is **bias**. I'll also set the random state to an integer to prevent having different outcomes each time I run the model.

Last but not least, whenever I have queries or run into problems, I shall email or meet with my tutor online. Extraordinary in-person meetings will be scheduled if I encounter any serious issues with my project or if I believe that communicating with my tutor in person will be more convenient.

1.8 Stakeholders

In this section the stakeholders who will benefits from the completion of this project will be enumerated.

The project's completion is important not only for people suffering from mental illnesses, but also for hospitals, clinics, educational institutions, and communities/groups in general. They could use the chatbot to treat and improve the mental health of patients/community members, thereby improving people's overall well-being.

- The project's completion is specially important for people struggling with mental health.

- Hospitals, clinics, educational institutions, and communities/groups in general also benefit from this project. They could use the chatbot to treat and improve the mental health of patients/community members, thereby improving people's overall well-being

2 Initial Project Planning

The project will last approximately 579 hours spread over 126 days, beginning on September 20th, 2022 and ending on January 23rd, 2022.[3]. Since the date for the project defense has not yet been determined, the previous deadline is the earliest we can have. It is planned to work an average of 5 hours per day, but some flexibility may be required due to exams or personal issues.

This section begins with a task description, followed by the resources required for the project's development, and finally by an explanation of risk management. Furthermore, Table 1 summarizes all of the defined tasks, as well as their dependencies and required resources, and Figure 5 captures the project schedule.

2.1 Description of tasks

The identification and description of the tasks that will be completed during the course of the project are presented in this section. It is given an estimate of the time needed for each task in hours, as well as a description of the logical sequence and dependencies between them. The tasks are divided into the following groups: **Project Management, Project Research, Project Theory, Data Generation, Project Experimentation, Project Development, Project Documentation and Thesis Defense Preparation.**

2.1.1 Project Management

Project management is most likely one of the project's pillars. It defines the scope of the project, the tasks and their planning, as well as the budget and sustainability.

- **PM1 - ICT tools for project and team management.**
  - **Description:** To support the project's development, the best technology, devices, and concepts that fit the nature of our project are required. To accomplish this, it is necessary to conduct research on various types of software for various tasks.
  - **Resources:** PC with internet acces.
  - **Approximate duration:** 1 hour.
- **PM2 - Context and Scope.**
  - **Description:** The project's scope, as well as its contextualization, are defined. The general goal of the project, its justification, developments, and tools are all discussed in this section.
  - **Resources:** This part requires a PC with internet connection, Overleaf to document and both The GEP [4] Tutor and the Tutor of the project for the feedback.
  - **Approximate duration** 35 hours.
- **PM3 - Time planning.**
  - **Description:** It is critical to plan ahead of time to ensure that the deadline is met. A good plan can help us identify which tasks require more attention than others and which are critical.
  - **Resources:** Making the Time Planning requires a PC, Overleaf, The GEP Tutor and the Tutor and TeamGantt to make the Gantt chart.
  - **Approximate duration:** 30 hours.
- **PM4 - Budget and sustainability.**
  - **Description:** This task focuses on creating a budget and analyzing the project's sustainability. This is critical in order to determine the total cost of the project and the impact it will have on the development.

[3]Because of the project's complexity, I worked "unofficially" on it over the summer, so the hours estimate is actually higher.

[4]GEP is a course that must be passed before submitting the final version of the thesis. It stands for "Gestió de Projectes" which translates to Project Management.

- **Resources:** This part needs a PC, Overleaf, The GEP Tutor and the Tutor as resources.
- **Approximate duration:** 30 hours.

- **PM5 - Meetings.**
  - **Description:** Meetings with the project's tutor will be scheduled as needed (E.g. when doubts or critical problems that impede the proper development of the project arise. I have added a "reserved time" for the meetings to make a better estimation of the total hours required for the project.
  - **Resources:** Tutor.
  - **Approximate duration:** 1 hour a week (in total, 18 hours).

2.1.2 Project Research

This part has been divided into the following tasks :

- **PR1 - Psychology research.**
  - **Description:** This project's methodology is based on psychology, specifically Cognitive Behavioral Therapy. Before embarking on the implementation phase, thorough research in CBT is essential.
  - **Resources:** PC, Books, Research Papers, Articles.
  - **Approximate duration:** 10 hours.
- **PR2 - ML research.**
  - **Description:** Documentation on various types of supervised machine learning models, as well as the statistics behind them, is also highly needed.
  - **Resources:** PC, Books, Research Papers, Articles.
  - **Approximate duration:** 10 hours.

2.1.3 Project Theory

In the theoretical part, I will study the structure of a CBT Therapy session in order to properly design the dialogue flow. Furthermore, I will consider what are the best ML models for both text classification and small datasets [5].

This part is divided into the following tasks:

- **PT1 - Design.**
  - **Description:** Design of the structure and dialogue flow of the chatbot.
  - **Resources:** PC, Books, Research Papers, Articles.
  - **Approximate duration:** 10 hours.
- **PT2 - Choose.**
  - **Description:** Choose between the top 5-7 supervised ML models for text classification and small datasets.
  - **Resources:** PC, Books, Research Papers, Articles.
  - **Approximate duration:** 10 hours.
- **PT3 - Select the Hyperparameters.**
  - **Description:** Select the Hyperparameters of every model to optimize.
  - **Resources:** PC, Books, Research Papers, Articles.
  - **Approximate duration:** 10 hours.

[5] Since I will be creating the data, the dataset will most likely be quite small (less than 1000 examples of cognitive distortions).

2.1.4 Data Generation

- **Description:** It is necessary to **generate the dataset** before beginning the experimentation with the models to determine which is the best. As previously stated, the generation will be created by using examples or finding inspiration on the internet, books, and articles specialized in psychology. The dataset will include examples of the 15 major cognitive distortions shown in Figure 6.
- **Resources:** PC, Books, Research Papers, Articles.
- **Approximate duration:** 14 hours.

2.1.5 Project Experimentation

In the experimentation section, the ml model workflow is completed as described in the previous assignment, and the best model is selected after analyzing the metrics.

In summary, the tasks are divided into the following:

- **PE1 - Apply the workflow.** This part will require 16 hours.
- **PE2 - Hyperparameters tuning.** Experiment with every model optimizing them with the aid Hyperparameters tuning selected in the Project Theory part (thus there is a dependency). This part will require 8 hours.
- **PE3 - Choose the best ml model.** Analyse the performance of every model and choose the best one for the text classification. This part will require 2 hours.

All these tasks will be done in Colaboratory, best known as "Colab". This tool is a product from **Google Research**. **Colab** allows anyone to write and run arbitrary Python code in the browser, making it ideal for machine learning, data analysis, and education. Colab notebooks are stored in Google Drive so it's safely stored in the Cloud. Furthermore, a PC, books, research papers, articles, programming languages and Github will be needed.

2.1.6 Project Development

The Project Development takes place after designing the structure and dialogue flow of the chatbot and consists on the implementation of a telegram chatbot.

The tasks are organized as follows:

- **PDEV1 - Implementation.**
  - **Description:** Implement the telegram chatbot.
  - **Resources:** PC, Github, Colab and programming languages.
  - **Approximate duration:** 200 hours.
- **PDEV2 - Testing.**
  - **Description:** Test the correct functioning of the telegram bot. This will be done during and after the implementation of the apps in order to make sure all parts work correctly on time.
  - **Resources:** PC, Github, Colab and programming languages.
  - **Approximate duration:** 60 hours.

2.1.7 Project Documentation

To avoid having to do everything at the end, the project documentation will be completed concurrently with the project's development (after the research part is done). For these task we need a PC, Overleaf/Texifier and Trello to keep track of the progress. The Project Documentation has been broken down into the following tasks:

- **PDOC1 - Annotation of events:** Annotation of all the events that are done during the project development. This task will be done intermittently and will aproximately require 10 hours.
- **PDOC2 - Revision of the events:** Once the project development is done is time to check all the documentation done during the project to better organize the ideas, correct changes and structure the final document correctly. This task will require 20 hours.
- **PDOC3 - Write Final Document:** After PDOC2 is done the writing of the final documentation begins. This task will require approximately 70 hours.

2.1.8 Thesis Defense Preparation

Finally, after the project documentation is completed, the oral defense preparation begins. To do so, it is necessary to practice and prepare for potential tribunal questions. This will require 25 hours.

2.2 Risk management: alternative plans

During the course of the project, difficulties that could jeopardize the project's proper progress may arise. All of the previously introduced potential problems in section 1.6 will be addressed in this section by introducing new tasks and accommodating the planning. A level of risk is also added to each of them.

- **Not being able to generate an appropriate or a representative dataset to train the models** **[Extreme Risk]**. Bias may be introduced into the data by providing examples by myself because I am forcing the examples to fit the criteria for cognitive distortions, which may or may not reflect how cognitive distortions occur in the real world. CD *"in the wild"* may be more subtle or more than one can appear in a thought/phrase. If that were to occur these are the steps that would be followed:
  - **Scrapping the internet or obtaining a dataset from a public mental health/therapy forum.** In my case, I would use a public dataset obtained from Reddit (specifically, from mental health subreddits) [11]. This method would yield "real-world" examples. Unfortunately, Manual labeling of the dataset into the 15 major cognitive distortions would be required.
  - **Crowdsourcing.** People will give examples that fit the CD criteria, so there may still be some bias. However, it would be significantly more varied than giving examples by myself in this case because people from different parts of the world would give examples (that may or may not apply to their real-life situation) that would most likely differ from mine due to differences in backgrounds, for example.
  - **Resources to reuse:** PC, programming languages, and TeamGantt.
  - **Estimated delay:** between 1-2 weeks.
- **Deadline of the project** **[High Risk]**. It could be caused by an accurate preliminary estimation of the tasks and their duration, which is completely normal because we do it before we begin. It is important to plan ahead of time, but it is also important to be flexible and adapt to changing circumstances. If that were to occur these are the steps that would be followed:
  - **Replanning in a more advanced of the project.** We can easily solve this problem by replanning in a more advanced part of the project, allowing us to do it more accurately.
  - **Increasing the hours dedicated to the project.** If, despite planning, it is still difficult to meet the deadline, it may be resolved by increasing the number of hours dedicated to the project as a last resort.

  - **Resources to reuse:** PC and TeamGantt.
  - **Estimated delay:** 1-2 weeks.

- **Bugs in libraries** [Medium Risk]. Third-party libraries will be used during project development, and they may contain bugs. Waiting until the library is updated, which should hopefully fix the bug, is one possible solution. However, due to the tight deadline, this option is out of the question. As a result, coding the function from scratch and testing its correct operation would be required, increasing the overall duration of the project.
  - **Resorces to reuse:** PC, TeamGantt and programming languages.
  - **Estimated delay:** 1-2 weeks.

ID	Name	Time (h)	Dependencies	Resources
PM	**Project Management**	**114**		
PM1	ICT tools for project and team management	1		PC
PM2	Context and Scope	35		PC, Overleaf, GEPT, T
PM3	Time planning	30	PM2	PC, Overleaf, GEPT, T, TeamGantt
PM4	Budget and sustainability	30	PM3	PC, Overleaf, GEPT, T
PM5	Meetings	18		T
PR	**Project Research**	**20**		PC, Books, Research Papers, Articles
PR1	Psychology Research	10		
PR2	ML Research	10		
PT	**Project Theory**	**30**		
PT1	Design Chatbot	10	PR1	PC, Books, Research Papers, Articles
PT2	Choose supervised ML models	10	PR2	PC, Books, Research Papers, Articles
PT3	Select the Hyperparameters to optimize	10	PT2	PC, Books, Research Papers, Articles
DG	**Data Generation**	**14**		PC, Books, Research Papers, Articles
PE	**Project Experimentation**	**26**		
PE1	Apply the Workflow	16	DG	PC, Books, Research Papers, Articles, Programming languages, Github, Colab
PE2	Hyperparameters Tuning	8	PT2, PE1	PC, Books, Research Papers, Articles, Programming languages, Github, Colab
PE3	Performance analysis of every model	2	PE2	PC, Books, Research Papers, Articles, Programming languages, Github, Colab
PDEV	**Project Development**	**260**		
PDEV1	Telegram bot Implementation	200	PT1	PC, Programming languages, Github, Colab
PDEV2	Testing	60	PT1	PC, Programming languages, Github, Colab
PDOC	**Project Documentation**	**90**		
PDOC1	Annotation of events	10	PR	PC, Textit, Trello
PDOC2	Revision of the events	20	PR	PC, Textit, Trello
PDOC3	Write final documentation	60	PR	PC, Textit, Trello
TDP	**Thesis Defense Preparation**	**25**	**PDOC**	PC, Textit, results
Total		**579**		

Table 1: Task Table containing a summary of all task information. T and GEPT means Tutor and GEP Tutor respectively. [Own Creation]

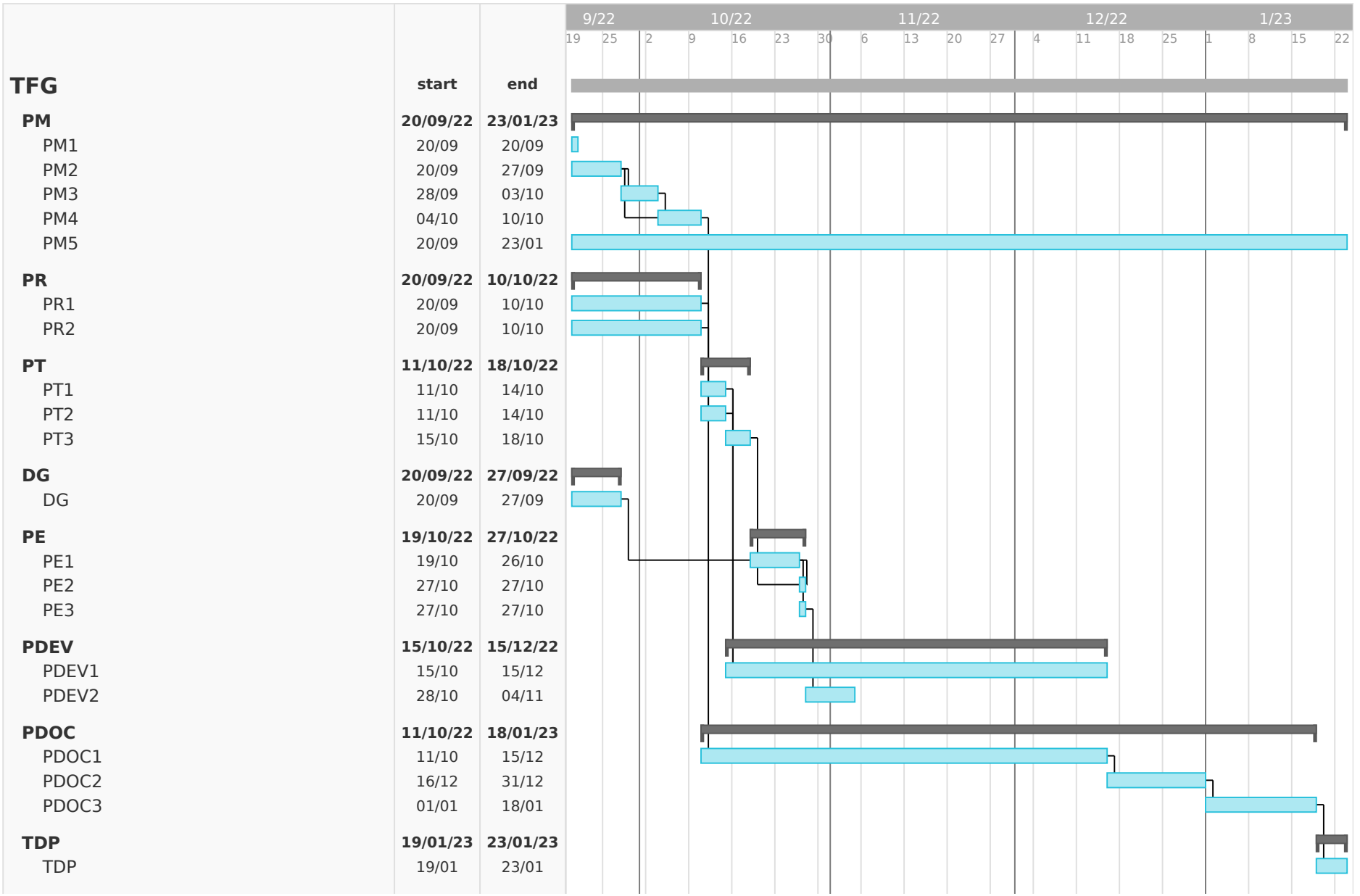


Figure 5: Gantt Chart illustrating the project's schedule following a Waterfall model. [Own Creation]

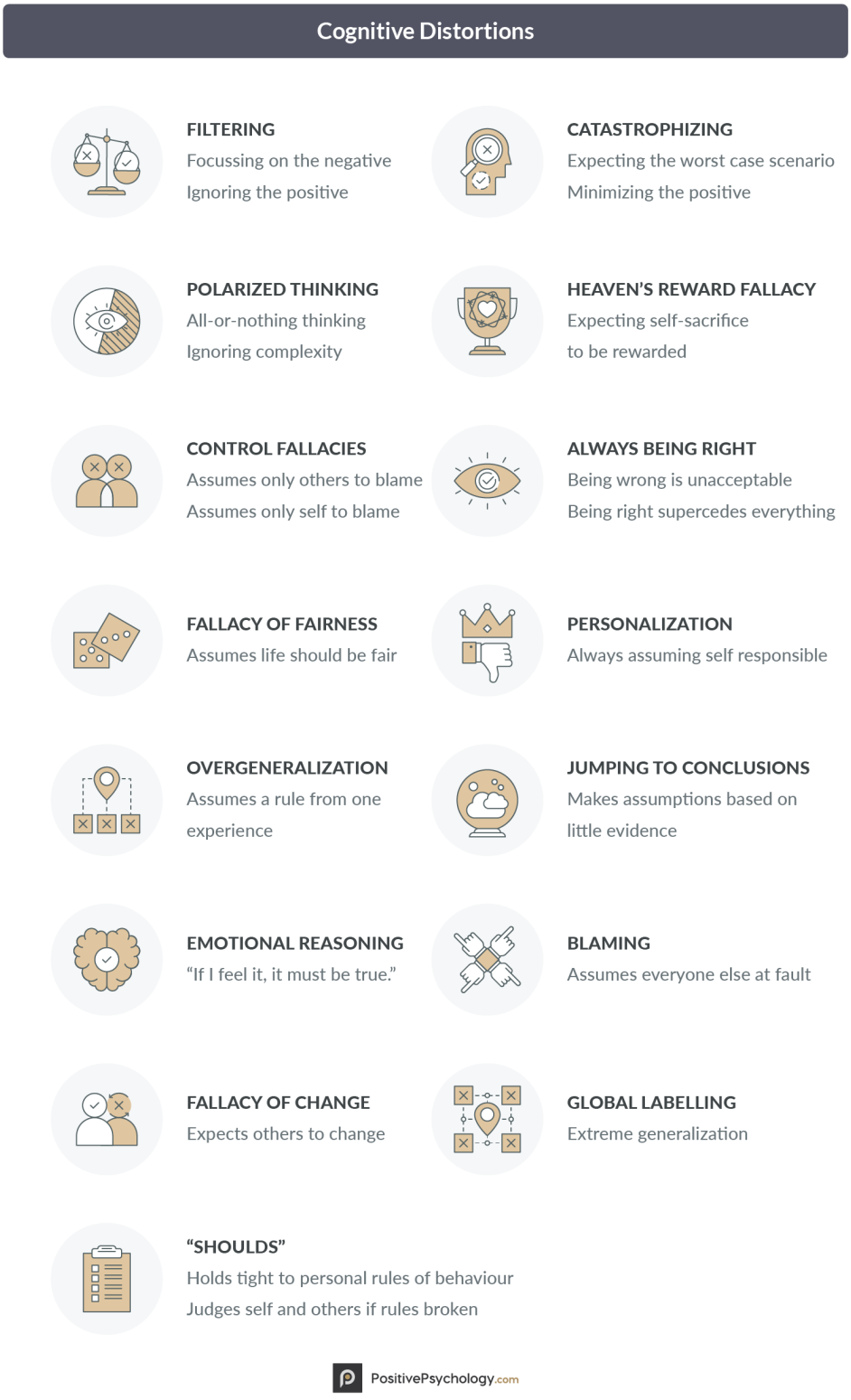


Figure 6: 15 major cognitive distortions by PositivePyschology, a website [4]

3 Initial Budget

In this section the economic cost of the project is discussed. First, the staff cost is described and analysed, then the generic and indirect costs are also calculated. Furthermore, the mechanism for controlling potential budget deviations is also explained. Finally, in Figure 3 we can see that the budget estimation is **17066,08€**.

3.1 Costs Per Activity

To accurately estimate the project's costs and create a budget, we must consider all of the resources required. Human resources is one of them. Even though the project is going to be done only by my with the guidance of my tutors, 7 roles are created to better simulate the required human resources to develop the project in order to better estimate the cost by task. The following is a description of the various responsibilities of each role:

- **Project Manager.** The project manager is in charge of the project's planning and development; in other words, the project manager oversees the project's progress.
- **Software engineer.** The software engineers implements the chatbot.
- **Tester.** The tester is in charge of verifying that the implementation is correct.
- **Research ML.** The researcher is responsible for investigating the best supervised machine learning models for the project and selecting the best hyperparameters for optimization.
- **Research psychologist** is responsible of designing the proper dialogue flow structure following the cbt method.
- **Technical writer** is in charge of documenting the project and presenting.
- **ML engineer** is responsible of implementing, tuning and analysing the models to choose the best one.

The Project Managements roles are going to be played by the tutors and the rest of roles by me.

In this section it is computed the Total Personnel Cost Per Activity (CPA). Each task or activity (previously defined in 2.1) is associated with the staff cost who are involved in that task. In this project there are 7 roles, each one with a different hourly salary which translates into **cost per hour** shown in Figure 2.

Role	Gross Annual Salary (€)	Price per hour (€)
Project Manager	52899,6	25,425
Software Engineer	46556,9	22,375
Tester	39533	19
Research ML	47239,4	22,7
Research psychologist	71436,3	34,35
Technical Writer	41600	20
ML engineer	47239,4	22,7

Table 2: Salary of the different roles extracted from PayScale, a compensation software company [6] multiplied by 1.35 to include the cost of social security [Own creation].

The computation of CPA is done by multiplying the hours required per task/activity with the cost per hour of the role that is involved in the activity. The total CPA is the sum of the CPA of every task of the Gantt Chart. As shown in Figure 3, the total cost of recruitment (CPA) is **13366,05€**.

3.2 Generic Costs

There are many resources that aren't directly tied to a task: the generic costs. To calculate the generic costs we need to take into account the **amortisation** of the resources used. In this project, all the software products are free so we are going to focus on the calculation of the hardware costs. I will be working 5 hours a day on average during 126 days. The computation of the amortisation is done with formula 1.

$$Amortisation(€) = Resource\ Price \cdot \frac{1}{Years\ of\ Use} \cdot \frac{1}{Days\ of\ Work} \cdot \frac{1}{Hours\ per\ Day} \cdot Hours\ Used \quad (1)$$

The indirect costs are identified to make the budget more realistic. Since I'll be working from home for the project (unless an extraordinary in-person meeting with the tutor is required), the transportation cost is zero. On the other hand, internet costs around 70€ per month, and electricity costs 100€ per month. The total Generic Cost, as shown in Figure 3, is **1223,11€**.

3.3 Contingency

Unexpected events are common during the development of a project, and one must plan ahead to account for them. As a result, a contingency plan is created in order to avoid potential delays during the planning process. Since contingency margins in the IT sector typically range from 10% to 20%, I decided to have a 15% contingency margin for this project, which amounts to **2188,37€**.

3.4 Incidental Costs

Incidental costs define all potential risks that could cause project delays. The most extreme risk of the project in this case is detecting bias in the machine learning models and thus having to generate more data alternatively, as explained in previous sections. The project is delayed as a result of this risk. Total Incidental Costs are **288,54€**.

3.5 Management control

In this section it is presented the mechanisms to control the budget. Also, the control indicators that aid to supervise cost deviations during the development of the project are defined.

While doing each planned task, the deviation from the estimated cost is calculated with the Formula 2.

$$Deviation(€) = Cost_{Estimated} - Cost_{Real} \quad (2)$$

If the deviation is negative, part of the contingency fund must reallocated in order to cover the deviation. In the positive case, it means there has been a overestimation of cost and reallocation of the extra money to incidents would be more productive.

4 Sustainability report

It is well known about climate change and the consequences that we are going to suffer or that we are even suffering now. Thus is really important for humans and companies to stop being selfish and think about the future of the world and cooperate in order to reduce pollution urgently. Thus, it is important to check the footprint of a project to see how does it impact in the environment. Assessing for the economical impact is also important, it helps us to optimize cost and savings. Finally keeping track of the social impact of a company/project is also very important. New technologies in particular have changed millions of people life including minorities and in developing countries. These three elements compose a sustainability report.

4.1 Self assessment

Students were asked to complete a survey for their bachelor thesis. This survey asks respondents about their knowledge of sustainability in various fields. Some of these fields include economic, environmental, and social sustainability. After doing the survey, I realize that the Environmental field is my weak spot. In particular, I don't know which indicators to use to measure the impact in this aspect. Regarding the economical field, I have some intuition on how to measure and control the economical impact since I previously did a a budget. Finally in the social field, I think I know how new technologies and in particular my project impacts society. So in conclusion, I have a below average level of knowledge about sustainability, especially in the environmental field.

4.2 Economic dimension

Regarding the PPP, have you estimated the cost of undertaking the project (human and material resources)?

In section 3 the estimated costs of the project are identified and calculated and the budget is also shown.

Regarding the exploitation, how is the problem that you wish to address resolved currently (state of the art)? In what ways will your solution economically improve existing solutions?

Nowadays most people go to in-person therapy with a professional therapist which is really expensive: in Spain the average price for one session is 50€and taking into account that on average a person needs between 8-20 sessions, the final costs amounts to 400-1000€. Online therapy sessions are getting popular giving people the flexibility to receive support and help with the need to transport. This option is usually more economical than traditional therapy (in person). Self-help therapy chatbots currently available in the market work on a free basis but if you want access to more content you need to subscribe, furthermore if you want to have access to a therapist: rates vary depending on the therapist or works on a subscription basis. Since KAI is free, it will help people embark in their self-help journey in therapy with the guide of the chatbot in a more affordable way.

4.3 Environmental dimension

Regarding the PPP, have you estimated the environmental impact of undertaking the project? Have you considered how to minimise the impact, for example by reusing resources?

Due to the project's nature, estimating its true impact is difficult. For example, it will be developed using only one laptop and thus will have little environmental impact. As far as I'm concerned, the only point that could possibly have an impact on the environment would be the energy consumption due to the training and evaluation of different machine learning models. One approach would be to execute in parallel to reduce training and evaluation time. This can be accomplished by increasing the njobs parameters above 0: the sklearn python library includes a parameter for determining the number of jobs to run in parallel for cross-validation.

Regarding the exploitation, how is the problem that you wish to address resolved currently (state of the art)? In what ways will your solution environmentally improve existing solutions?

As mentioned before, most people go to therapy in the traditional way. Since people can have access to KAI without the need to move from home, the project helps in reducing the pollution because they

no longer need to take a mains of transport to go to a session. So we can conclude that KAI is more environmentally-friendly.

4.4 Social dimension

Regarding the PPP, what do you think undertaking the project has contributed to you personally?

It has aided me in learning more about psychology, a field in which I have always been interested, and how technology can help people with their psychological needs. Furthermore, it has helped me in becoming more familiar with the machine learning field, particularly in the healthcare/medical sector, in which I am very interested. It also made me realize that Python is a powerful programming language, particularly for AI, due to the existence of an extensive library. What's more, the project experience has assisted me in determining whether I truly want to pursue a career in AI in healthcare or medicine. Finally, the most significant and special contribution of the project to me has been the opportunity to provide a mental health support tool like KAI to my sister, who unfortunately suffers from a severe mental illness.

Regarding the exploitation, how is the problem that you wish to address resolved currently (state of the art)? In what ways will your solution socially improve (quality of life) existing

The project will help to close the gap between those who need and those who receive mental health care. They will also have 24/7 access to support. More importantly, **it will allow people to have access to mental health support in a more affordable, autonomous, and time-efficient way**.

Activity	Amount (€)	Observations
PM1 - ICT tools for project and team management	25,43	Project Manager , 1 hour
PM2 - Context and Scope	889,88	Project Manager , 35 hours
PM3 - Time planning	762,75	Project Manager , 30 hours
PM4 - Budget and sustainability	762,75	Project Manager , 30 hours
PM5 - Meetings	457,65	Project Manager , 18 hours
PR1 - Psychology Research	343,50	Research psychologist, 10 hours
PR2 - ML Research	340,50	Research ML, 10 hours
PT1 - Design Chatbot	343,50	Research psychologist, 10 hours
PT2 - Choose supervised ML models	227,00	Research ML, 10 hours
PT3 - Select the Hyperparameters to optimize	227,00	Research ML, 10 hours
DG - Data Generation	480,90	Research psychologist, 14 hours
PE1 - Apply the Workflow	363,20	ML engineer, 16 hours
PE2 - Hyperparameters Tuning	181,60	ML engineer, 8 hours
PE3 - Performance analysis of every model	45,40	ML engineer, 2 hours
PDEV1 - Telegram bot Implementation	4475,00	Software Engineer, 200 hours
PDEV2 - Testing	1140,00	Tester, 60 hours
PDOC1 - Annotation of events	200,00	Technical Writer, 10 hours
PDOC2 - Revision of the events	400,00	Technical Writer, 20 hours
PDOC3 - Write final documentation	1200,00	Technical Writer, 60 hours
TDP - Thesis Defense Preparation	500,00	Technical Writer, 25 hours
Total CPA (Cost Per Activity)	**13366,05**	Total personnel costs by activty (Gantt actvities)
Hardware		
Laptop	220,57	Mackbook Air 2017, Purchase Price: 1200e
Peripheral devices	122,54	Display + mouse + keyboard, Purchase Price: 400e
Software		
Overleaf	0,00	Free to use
Google sheets	0,00	Free to use
TeamGantt	0,00	Free to use
Colab	0,00	Free to use
GitHub	0,00	Free to use
Space		
Electricity	400,00	100€/month x 4 months (duration of project)
Furniture	200,00	Table + Chair
Internet	280,00	70€/month x 4 months (duration of project)
Transport	0,00	Work from home
Total GC (Cost computed Generically)	**1223,11**	
Total Cost (Total CPA + Total GC)	**14589,16**	
Contingency	2188,37	Contingency margin = 15%
Total DC (direct cost) + IC (indirect cost) + Contingency	**16777,54**	
Data Generation Delay (1 Week)	240,45	Cost: Research psychologist, 14 hours. Risk: 50%
Data Generation Delay (2 Week)	48,09	Cost: Research psychologist, 14 hours. Risk: 10%
Total incidentals (or unforseen costs)	**288,54**	
TOTAL	**17066,08**	

Table 3: Budget Structure of the project [Own creation]

5 Final Project Planning

The steps of project planning may be seen in Table 4, and the project is currently in the development phase (as it was planned in the initial project planning). As can be seen in Table 5, there were two significant changes: one that affected the project's development technique and the other that affected the documentation of the monitoring report. The changes will be described in the section that follows.

Project Planning Phase	Description
Project Management	*Definition of the scope of the project, the tasks and their planning, as well as the budget and sustainability.*
Project Research	*Research the best machine learning models for topic classification and CBT (Cognitive Behavioural Therapy).*
Project Theory	*Design chatbot, choose supervised ML models and select the Hyperprameters to optimize.*
Data Generation	*Data Generation by giving examples based on books, books and research papers of CBT.*
Project Experimentation	*Data training and analysis of every model previously chose in Project Theory.*
Project Development	*Implementation of the chatbot with Jupyter Notebook and Javascript (BotUI framework)*
~~Telegram bot~~ Chatbot Implementation	*Use JS (BoutUI) instead of Telegram API to develop the chatbot*
Testing	
Project Documentation	
Monitoring Report	
Annotation of events	
Revision of the events	
Write final documentation	
Thesis Defense Preparation	

Legend
Done
In Progress
To Do
Change Added

Table 4: Task Table showing the progress of the project [Own creation]

ID	Name	Time (h)	Dependencies	Resources
PM	**Project Management**	**114**		
PM1	ICT tools for project and team management	1		PC
PM2	Context and Scope	35		PC, Overleaf, GEPT, T
PM3	Time planning	30	PM2	PC, Overleaf, GEPT, T, TeamGantt
PM4	Budget and sustainability	30	PM3	PC, Overleaf, GEPT, T
PM5	Meetings	18		T
PR	**Project Research**	**20**		PC, Books, Research Papers, Articles
PR1	Psychology Research	10		
PR2	ML Research	10		
PT	**Project Theory**	**30**		
PT1	Design Chatbot	10	PR1	PC, Books, Research Papers, Articles
PT2	Choose supervised ML models	10	PR2	PC, Books, Research Papers, Articles
PT3	Select the Hyperparameters to optimize	10	PT2	PC, Books, Research Papers, Articles
DG	**Data Generation**	**14**		PC, Books, Research Papers, Articles
PE	**Project Experimentation**	**26**		
PE1	Apply the Workflow	16	DG	PC, Books, Research Papers, Articles, Programming languages, Github, Colab
PE2	Hyperparameters Tuning	8	PT2, PE1	PC, Books, Research Papers, Articles, Programming languages, Github, Colab
PE3	Performance analysis of every model	2	PE2	PC, Books, Research Papers, Articles, Programming languages, Github, Colab
PDEV	**Project Development**	**260**		
PDEV1	~~Telegram bot~~ Chatbot Implementation	200	PT1	PC, Programming languages, Github, Jupyter Notebook, VS Code
PDEV2	Testing	60	PT1	PC, Programming languages, Github, Jupyter Notebook, VS Code
PDOC	**Project Documentation**	**105**		
PDOC1	Monitoring Report	15		PC, Overleaf, Trello, T
PDOC2	Annotation of events	10	PR	PC, Textit, Trello
PDOC3	Revision of the events	20	PDOC2	PC, Textit, Trello
PDOC4	Write final documentation	60	PDOC3	PC, Textit, Trello
TDP	**Thesis Defense Preparation**	**25**	**PDOC**	PC, Textit, results
Total		**594**		

Table 5: Final version of the task table [Own creation]

5.1 Deviations in the project development

After revising the documentation of the telegram app API to develop the chatbot, I opted to use another alternative: BotUI a javascript framework. The reason to this was that I found the Telegram API documentation rather unclear and not easy to use for the functionalities needed for the chatbot.

5.2 Deviations in the project documentation

Writing the monitoring report wasn't factored into the initial plans. As a result, the monitoring report has been included in the project's final planning. This modification affected both the computation of the budget, as shown in Table 6, and the projected number of hours required to complete the project.

Activity	Amount (€)	Observations
PM1 - ICT tools for project and team management	25,43	Project Manager , 1 hour
PM2 - Context and Scope	889,88	Project Manager , 35 hours
PM3 - Time planning	762,75	Project Manager , 30 hours
PM4 - Budget and sustainability	762,75	Project Manager , 30 hours
PM5 - Meetings	457,65	Project Manager , 18 hours
PR1 - Psychology Research	343,50	Research psychologist, 10 hours
PR2 - ML Research	340,50	Research ML, 10 hours
PT1 - Design Chatbot	343,50	Research psychologist, 10 hours
PT2 - Choose supervised ML models	227,00	Research ML, 10 hours
PT3 - Select the Hyperparameters to optimize	227,00	Research ML, 10 hours
DG - Data Generation	480,90	Research psychologist, 14 hours
PE1 - Apply the Workflow	363,20	ML engineer, 16 hours
PE2 - Hyperparameters Tuning	181,60	ML engineer, 8 hours
PE3 - Performance analysis of every model	45,40	ML engineer, 2 hours
PDEV1 - Telegram bot Chatbot Implementation	4475,00	Software Engineer, 200 hours
PDEV2 - Testing	1140,00	Tester, 60 hours
PDOC1 - Monitoring Report	300,00	Technical Writer, 15 hours
PDOC2 - Annotation of events	200,00	Technical Writer, 10 hours
PDOC3 - Revision of the events	400,00	Technical Writer, 20 hours
PDOC4 - Write final documentation	1200,00	Technical Writer, 60 hours
TDP - Thesis Defense Preparation	500,00	Technical Writer, 25 hours
Total CPA (Cost Per Activity)	**13666,05**	Total personnel costs by activty (Gantt actvities)
Hardware		
Laptop	226,29	Mackbook Air 2017, Purchase Price: 1200e
Peripheral devices	125,71	Display + mouse + keyboard, Purchase Price: 400e
Software		
Overleaf	0,00	Free to use
Google sheets	0,00	Free to use
TeamGantt	0,00	Free to use
Colab	0,00	Free to use
GitHub	0,00	Free to use
Space		
Electricity	400,00	100€/month x 4 months (duration of project)
Furniture	200,00	Table + Chair
Internet	280,00	70€/month x 4 months (duration of project)
Transport	0,00	Work from home
Total GC (Cost computed Generically)	**1232,00**	
Total Cost (Total CPA + Total GC)	**14898,05**	
Contingency	2234,71	Contingency margin = 15%
Total DC (direct cost) + IC (indirect cost) + Contingency	**17132,76**	
Data Generation Delay (1 Week)	240,45	Cost: Research psychologist, 14 hours. Risk: 50%
Data Generation Delay (2 Week)	48,09	Cost: Research psychologist, 14 hours. Risk: 10%
Total incidentals (or unforseen costs)	**288,54**	
TOTAL	**17421,30**	

Table 6: Final version of the budget structure [Own creation]

6 Final Methodology

The project's methodology hasn't changed; the hybrid mode between waterfall and agile technique is well suited for the project, and it's because of this that the previously mentioned deviations haven't had a significant impact on the project's proper development.

The gantt chart, which employs a waterfall methodology, was used throughout the project's development to determine the dependencies and the order in which to strategize the tasks. On the other hand, the project's coding and testing phases used an agile methodology with kanban boards to keep track of all the tasks.

7 Technical Competences

During the development of the project thesis, the following technical competences from computing specialization were followed:

CCO2.1
To demonstrate knowledge about the fundamentals, paradigms and the own techniques of intelligent systems, and analyse, design and build computer systems, services and applications which use these techniques in any applicable field. [Quite]

During the project design and development of an application (a chatbot) using machine learning has been done.

CCO2.2
Capacity to acquire, obtain, formalize and represent human knowledge in a computable way to solve problems through a computer system in any applicable field, in particular in the fields related to computation, perception and operation in intelligent environments. [Quite]

This competence was achieved with the acquisition of human knowledge and the representation through machine learning to detect possible cognitive distortions.

CCO2.3
To develop and evaluate interactive systems and systems that show complex information, and its application to solve person-computer interaction problems. [A little]

This competence was achieved with the design and development of the chatbot that extracts and shows complex information (possible cognitive distortions) from user input, a form of human-computer interaction (conversational user interface).

CCO2.4
To demonstrate knowledge and develop techniques about computational learning; to design and implement applications and system that use them, including these ones dedicated to the automatic extraction of information and knowledge from large data volumes. [In depth]

The study included extensive research on the top machine learning methods for topic classification. In addition, the design and development of the chatbot that automatically detects potential cognitive distortions using a machine learning technique.

8 Identification of Laws and Regulations

Understanding the laws and regulations that have an impact on the design and development of the chatbot is one of the most crucial components of the project thesis.

8.1 Academic Regulations for the Degree Final Project

The UPC has a documentation of the regulations for the Degree Final Project available online [12]. This documents defines and describes the characteristics of the final project. The document also explains all the process needed to do the project. This document is of course very important and must be followed to ensure the correct development of the project.

8.2 BotUI License

BotUI is a JavaScript framework to develop conversational chatbots. The MIT License [13] states that the framework is free to use with the only requirement of preserving the copyrights. To comply with this requirement, the copyright notice and the permission notice will be included in the project.

8.3 GDPR Privacy Policy

A first step toward granting EU people and residents more control over how their data are used in organizations is the EU General Data Protection Regulation (GDPR). No matter where they are in the globe, businesses must adhere to the GDPR if they handle the personal data of people who reside in the EU.

A key requirement for businesses subject to the GDPR is that they make transparent and easily accessible information about the personal data they are processing available to the public. A clear and thorough privacy policy will help one to achieve this.

A privacy notice is a public statement from a company outlining how it manages customer information and adheres to data protection laws. A GDPR privacy notice is a crucial tool for assisting customers and users in making informed choices regarding the data you gather and use.

According to the GDPR [14], organizations are required to give customers a privacy disclosure that is:

- In a clear, visible, understandable, and readily available format
- Written in a straightforward manner, especially for any information aimed exclusively towards children
- Delivered on schedule
- Provided free of charge

In the following sections, all the information that must be included in a privacy notice is explained.

8.3.1 Company's contact details

Article 13(1)(a) [15] of the GDPR requires providing to users with: **"the identity and the contact details of the controller and, where applicable, of the controller's representative"**. An individual who determines how and why personal data is handled is referred to as "the controller" or a "data controller."

Article 13(1)(b) [15] of the GDPR also requires providing: **"the contact details of the data protection officer, where applicable"**. A data protection officer is required for some firms of a specific size or those that consistently handle sensitive personal data (DPO).

8.3.2 The Purposes and Legal Basis for Processing

Article 13 (1)(c) [15] of the GDPR requires providing information about: **"the purposes of the processing for which the personal data are intended as well as the legal basis for the processing"**. To put it another way, is not allowed to process personal data unless there is a purpose for doing so. Additionally, there must be a legal justification for every form of data processing is carried out.

The GDPR sets out six legal bases at Article 6.

A person's personal data may only be processed if at least one of the following conditions is met:

- You have their consent.
- To carry out or enter into a contract with them, you must process their personal data.
- It's required by law that you handle their personal information.
- Failure to process their personal data could endanger their lives or the life of another person.
- Processing their personal data is something you're doing in the public interest.
- You have a legitimate interest in processing their personal data.

The app falls under the category "You have a legitimate interest in processing their personal data" since it collects user data in order to identify potential cognitive distortions based on user input.

8.3.3 Sharing of user's personal data

Article 13 (1)(e) [15] requires to provide information about: **"the recipients or categories of recipients of the personal data, if any"**. In the app's case the data is never shared with third party companies.

8.3.4 Sharing of user's personal data to a third country

Article 13(1)(f) [15] of the GDPR requires providing information about: **"the fact that the controller intends to transfer personal data to a third country or international organization and the existence or absence of an adequacy decision by the commission"**. A "third country" refers to a country outside of the EU.

The list of nations with "sufficient" data protection rules is maintained by the European Commission. You must indicate if a country is on the list if you are sending data to a third country. In the app's case the data is never shared with third countries.

8.3.5 Period of time storage of user's personal data

Article 13(2)(a) [15] of the GDPR requires informing users: **"the period for which the personal data will be stored, or if that is not possible, the criteria used to determine that period"**. It's crucial to comply with the GDPR's prohibition on keeping personal data longer than necessary. In the case of the app, user data is never stored.

8.3.6 User's Rights

Chapter 3 of the GDPR [16] sets out the rights that people have over their data. The GDPR not only requires you to not only make it easier for your users to access these rights, **but also to inform them of those rights in your Privacy Policy**. Additionally, you must let the user know how to file a complaint with their local data protection authority.

8.4 The EU Regulatory Environment of Medical Device Software Development

Software that is designed to be used alone for one or more medical purposes without being part of a hardware medical device is referred to as "Software as a Medical Device" (SaMD). This terminology is largely used by the IMDRF (International Medical Device Regulation Forum). The chatbot falls into this category.

The relevant General Safety and Performance Requirements (GSPRs) will have to be complied with by all software that falls within the Medical Device category. MDSW lawful manufacturers must put out a dossier or technical document (TD) for their product in order to prove conformity with GSPR. The details and explanations of the documents' contents are provided in the section that follows.

Applicable GSPRs mainly refer to one of the following general fields:

- **Quality Management System (QMS) requirements**. MDSW developers must work following a QMS methdology.

- **Risk Management System (RMS) requirements**. The basic objective of an RMS is to make sure that any potential dangers are recognized, categorised, and minimized without negatively influencing the device's risk-benefit ratio. To do this, the manufacturer must develop and implement a risk management strategy that accurately identifies all hazards related to the devices, establishes the necessary risk mitigation measures, and evaluates the effectiveness of each strategy.
- **Clinical Evaluation and Post-market surveillance requirements**. The clinical evaluation of a MDSW must be carried out following *MDCG 2020-1, guidance on Clinical Evaluation of MDSW*, and a Clinical Evaluation Report drafted providing the following information:
  - A valid clinical association of the software with the targeted clinical condition or physiological state, usually by means of literature references.
  - An analytical evaluation of the software to show that it is capable of processing data appropriately.
  - The software's output is then validated clinically to guarantee that it is accurate and dependable in the context of the clinical setting.
- **Usability requirements**. SW developers must make sure that as many user errors as possible are prevented via the user interface. *IEC 62366 Medical devices — Part 1: Application of usability engineering to medical devices* must be followed when planning and conducting usability tests for this. From a cybersecurity and safety standpoint, each and every one of the discovered user errors must be taken into account in the risk analysis and contributed to the risk management strategy and report. As with any other risk, preventive steps must be taken if the possibility of user errors cannot be entirely removed. Some of them include increasing training or adding particular warnings to the user handbook.

In addition, theses requirements are specifically applicable to MDSW:

- **Software lifecycle requirements**. The software lifecyle is described in *IEC 62304 Medical device software — Software life cycle processes*. It provides a series of steps that should be taken by SW developers. In Figure 7 we can see the Software Lifecycle.
- **Cybersecurity requirements**. This mainly regard patient data protection and protection from other cyber threats.

As with any other MD, it is advised that MDSW developers use approved techniques and standardized procedures like the ones listed below in order to adhere to the relevant GSPRs:

- **International standards** (mainly ISO [6], IEC [7] and ANSI/AAMI [8] standards)
- **MDCG** or **IMDRF** guidance documents

Table 7 provides an exhaustive list of the standards and guidelines now in existence that are advised to be followed by MDSW developers in order to accomplish with applicable GSPRs, together with their most recent updates.

[6] International Organization for Standardization (ISO)
[7] International Electrotechnical Commission (IEC)
[8] American National Standards Institute (ANSI)/ Association for the Advancement of Medical Instrumentation (AAMI)

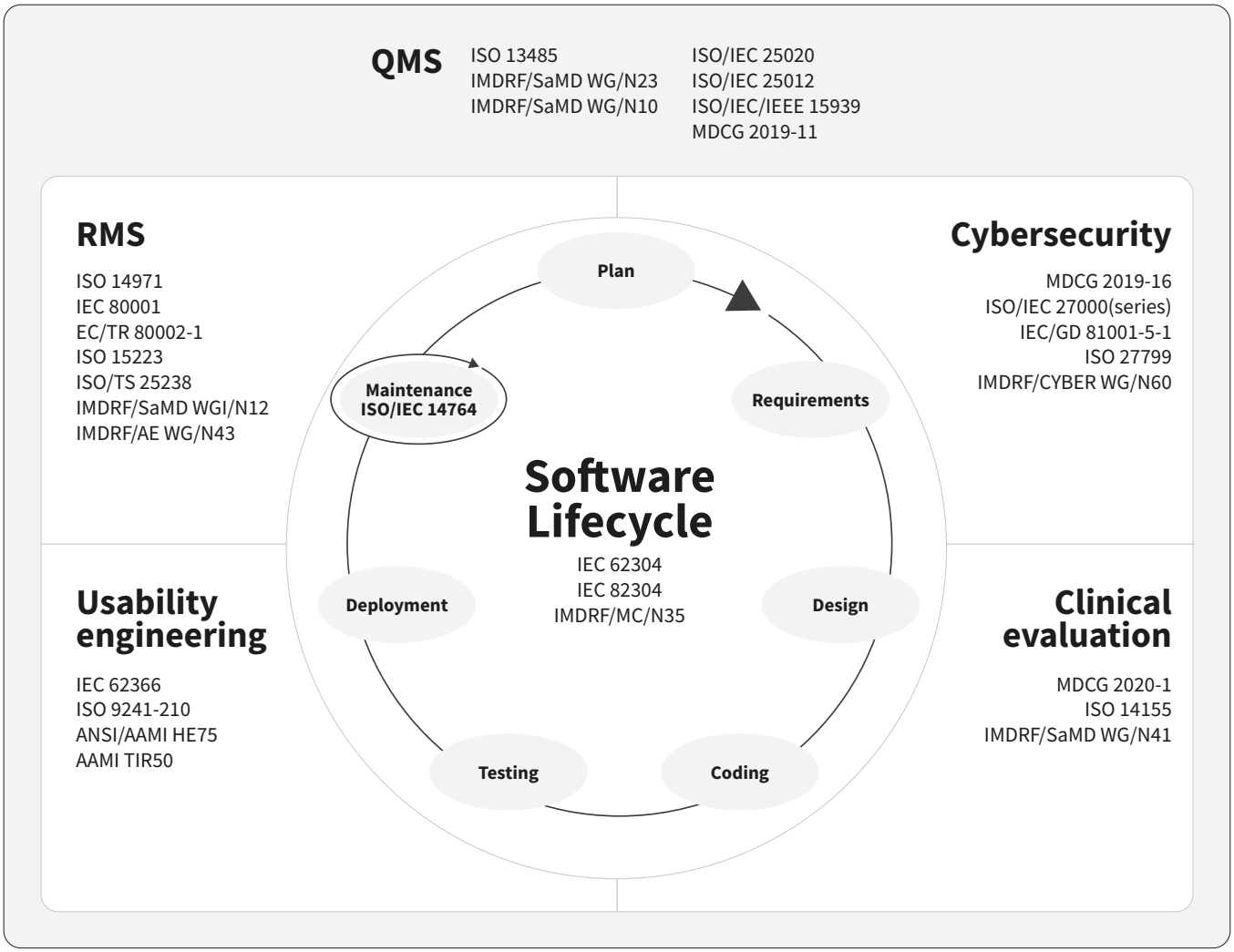


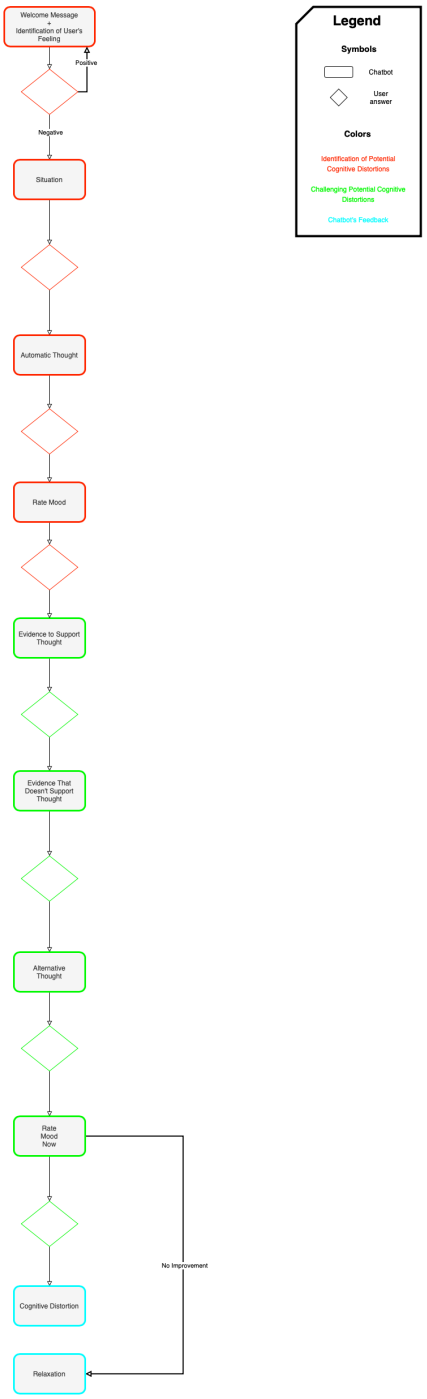
Figure 7: EU Regulatory Environment for MDSW Development. [5]

Requirement	Standard or Guidance	Title
Quality Management System (QMS)	**EN ISO 13485:2016/AC:2018 (*)**	Medical devices - Quality management systems - Requirements for regulatory purposes
	ISO/IEC 25020:2019	Systems and software engineering — Systems and software Quality Requirements and Evaluation (SQuaRE) — Quality measurement framework
	ISO/IEC 25012:2008	Software engineering -- Software product Quality Requirements and Evaluation (SQuaRE) -- Data quality model
	ISO/IEC/IEEE 15939:2017	Systems and software engineering — Measurement process
	MDCG 2019-11	Qualification and classification of software - Regulation (EU) 2017/745 and Regulation (EU) 2017/746
	IMDRF/SaMD WG/N23 FINAL: 2015	Software as a Medical Device (SaMD): Application of Quality Management System
	IMDRF/SaMD WG/N10 FINAL:2013	Software as a Medical Device (SaMD): Key Definitions
Risk Management System (RMS)	**EN ISO 14971:2019 (*)**	Medical devices - Application of risk management to medical devices
	IEC 80001-1:2010 (series)	Application of risk management for IT-networks incorporating medical devices — Part 1: Roles, responsibilities and activities
	EC/TR 80002-1:2009	Medical device software — Part 1: Guidance on the application of ISO 14971 to medical device software
	ISO/TS 25238:2007	Health informatics — Classification of safety risks from health software
	EN ISO 15223-1:2016 (*)	Medical devices - Symbols to be used with medical device labels, labelling and information to be supplied - Part 1: General requirements
	IMDRF/SaMD WG/N12 FINAL:2014	"Software as a Medical Device": Possible Framework for Risk Categorization and Corresponding Considerations
	IMDRF/AE WG/N43 FINAL:2020 & Annexes	IMDRF terminologies for categorized Adverse Event Reporting (AER): terms, terminology structure and codes
Clinical Evaluation	**EN ISO 14155:2020 (*)**	Clinical investigation of medical devices for human subjects - Good clinical practice
	MDCG 2020-1	Guidance on clinical evaluation (MDR) / Performance evaluation (IVDR) of medical device software
	IMDRF/SaMD WG/N41 FINAL:2017	Software as a Medical Device (SaMD): Clinical Evaluation
Cybersecurity	**ISO/IEC 27000:2018(en) (series)**	Information technology — Security techniques — Information security management systems — Overview and vocabulary
	ISO 27799:2016	Health informatics — Information security management in health using ISO/IEC 27002
	IEC/CD 81001-5-1 (draft 2021)	Health software and health IT systems safety, effectiveness and security — Part 5-1: Security — Activities in the product lifecycle
	MDCG 2019-16 rev.1	Guidance on cybersecurity for medical devices
	IMDRF/CYBER WG/N60FINAL:2020	Principles and Practices for Medical Device Cybersecurity
Usability	**IEC 62366-1:2015 (*)**	Medical devices - Application of usability engineering to medical devices
	ISO 9241-210:2010	Ergonomics of human-system interaction - Human-centered design for interactive systems
	ANSI/AAMI HE75:2009/(R)2018 (*)	Human factors engineering- Design of medical devices
	AAMI TIR50:2014 (*)	Post-market surveillance of use error management
Software lifecycle	**EN 62304:2006/AC:2008(*)**	Medical device software - Software life-cycle processes
	IEC 82304-1:2016	Health software — Part 1: General requirements for product safety
	ISO/IEC 14764:2006	Software Engineering — Software Life Cycle Processes — Maintenance
	IMDRF/MC/N35 FINAL:2015	Statement regarding Use of IEC 62304:2006 "Medical device software -- Software life cycle processes"
(*) Although they are not software-specific, these standards are highly relevant for the development of MDSW.		

Table 7: Standard and guidance documents useful to demonstrate MDSW compliance with MDR. [5]

9 Dialogue Flow

This section explains how the design of the chatbot's dialogue flow. The dialogue flow is divided in three main parts as shown in figure which are the following: identifying potential cognitive distortions, challenging potential cognitive distortions, and relaxation exercises.


Welcome Message
+
Identification of User's Feeling
Positive
Negative
Situation
Automatic Thought
Rate Mood
Evidence to Support Thought
Evidence That Doesn't Support Thought
Alternative Thought
Rate Mood Now
No Improvement
Cognitive Distortion
Relaxation
Legend
Symbols
Chatbot
User answer
Colors
Identification of Potential Cognitive Distortions
Challenging Potential Cognitive Distortions
Chatbot's Feedback

9.1 Identifying Potential Cognitive Distortions

The first step in the cognitive part of the therapy is to identify cognitive distortions. By first identifying the user's automatic thoughts, potential cognitive distortions can be identified. A idea that arises automatically or spontaneously, without conscious effort, is referred to as an automatic thought. Automatic thoughts can be brought on by certain events or circumstances and are frequently connected to negative emotions.

Automatic thoughts are frequently founded on our beliefs, values, and expectations and are immediate, automatic responses to occurrences. Our present emotional state as well as our prior experiences can have an impact on them. Automatic thoughts, often referred to as cognitive distortions, can be both positive and negative, although they are typically negative and, if unrecognized and unchallenged, can result in bad feelings and behaviors.

9.2 Challenging Potential Cognitive Distortions

Cognitive distortions can be fought with a variety of methods. The majority of times, a thought record is used. Patients are able to challenge their own beliefs and thoughts in a cooperative, Socratic manner, which helps them to utilize their own arguments to refute dysfunctional thinking. The impact of the counter-thoughts and beliefs is increased by creating counter-statements based on the information patients bring to the session.

Socratic questioning is a therapeutic approach for challenging cognitive distortions. In this procedure, a patient is guided to identify his or her thoughts, feelings, or actions related to a certain scenario by way of a series of open-ended, brief questions. Socratic questioning is nonjudgmental, but it is founded on the therapist's knowledge that the underlying foundation of a thought or belief might not be accurate; as a result, the questions are made to reveal the dysfunctional idea or belief so that it can be refuted.

In cognitive-behavioral therapy (CBT), a Dysfunctional Thought Record (DTR) is a tool used to identify and address negative ideas and beliefs that contribute to emotional and behavioral problems. The DTR is a systematic form that assists people in recognizing their negative attitudes and beliefs, weighing the evidence supporting and refuting them, and coming up with more reasonable and adaptable alternatives.

9.3 Relaxation Exercises

A variety of psychotherapy strategies known as relaxation techniques are used to ease tension, stress, worry, and/or anxiety.

There are various reasons why relaxation techniques are crucial in brief therapy. They first concentrate on techniques for reducing the frequently crippling stress, anxiety, worry, and tension that obstruct patient performance. Second, patients generally find stress, anxiety, concern, and tension to be extremely uncomfortable. By helping them to feel better, you can greatly improve their treatment expectations and rapport.

Deep breathing is a physiologically grounded approach for relaxation. The goal of deep breathing is to slow down the shallow, irregular breathing that frequently takes place when people are stressed, worried, or anxious. Rapid and shallow breathing can result in the patient's blood oxygen levels dropping, which can impair their ability to focus and cause symptoms including hyperventilation, lightheadedness, and dizziness. As an alternative, boosting the amount of oxygen-rich blood flow with a deep, complete breath might create a sense of serenity or slowing.

10 Supervised Machine Learning Algorithms

10.1 Multinomial Naive Bayes

Multinomial Naive Bayes is a classification algorithm that is based on the Naive Bayes algorithm and is specifically designed for classification tasks with multiple classes. It is commonly used for text classification tasks, such as spam filtering or sentiment analysis, because it can handle the large number of features that are typically present in text data.

The Multinomial Naive Bayes algorithm works by estimating the probability of each class label occurring based on the frequency of the label in the training data. It also estimates the probability of each feature occurring given a specific class label, which is known as the likelihood. In contrast to the standard Naive Bayes algorithm, which assumes that all features are independent of each other, the Multinomial Naive Bayes algorithm takes into account the frequency of each feature in the training data. This is useful for text classification tasks because the presence or absence of a particular word may be indicative of the class label.

To classify a new data point, the Multinomial Naive Bayes algorithm uses the Bayes theorem to calculate the probability of each class label given the features of the data point. The class label with the highest probability is chosen as the prediction for the data point.

One advantage of the Multinomial Naive Bayes algorithm is that it is able to handle large numbers of features and can accurately classify text data. It is also relatively simple and efficient to implement. However, like the standard Naive Bayes algorithm, the assumption of independence between features may not always hold in real-world data, which can affect the accuracy of the classifier.

In summary, the Multinomial Naive Bayes algorithm is a classification method that is based on the Naive Bayes algorithm and is specifically designed for text classification tasks with multiple classes. It estimates the probability of each class label occurring and the likelihood of each feature occurring given a specific class label, and uses the Bayes theorem to classify new data points. While it is effective for text classification tasks, the assumption of independence between features may not always hold in real-world data, which can affect the accuracy of the classifier.

BAYES THEOREM (IMAGE??)

10.2 Multinomial Logistic Regression

Multinomial logistic regression is a classification method that is used to predict a categorical dependent variable, with multiple categories, from one or more independent variables. It is an extension of the binary logistic regression model, which is used to predict a binary outcome.

In multinomial logistic regression, the dependent variable is assumed to have multiple categories, each of which can be predicted based on the values of the independent variables. The model estimates the probability of each category, and the category with the highest probability is chosen as the predicted outcome.

The model is based on the assumption that the log-odds of the dependent variable are a linear combination of the independent variables. This is expressed as:

log(p(y=k)/(1-p(y=k))) = b0 + b1x1 + b2x2 + ... + bn*xn

where y is the dependent variable, k is a category of y, x1, x2, ..., xn are the independent variables, and b0, b1, b2, ..., bn are the coefficients that are estimated by the model.

To fit the model, the maximum likelihood estimation method is used to find the values of the coefficients that maximize the likelihood of the observed data. This is done by minimizing the negative log-likelihood of the data, which is expressed as:

-ln(L) = -[y=k]ln(p(y=k))

where L is the likelihood of the data, y is the dependent variable, and k is a category of y.

There are several advantages to using multinomial logistic regression, including its ability to handle multiple categories and its ability to model the relationships between the independent variables and the dependent variable. However, it is important to note that the model assumes that the independent variables are independent of each other, which may not always be the case in practice.

In conclusion, multinomial logistic regression is a useful classification method for predicting a categorical dependent variable with multiple categories from one or more independent variables. It is based on the assumption that the log-odds of the dependent variable are a linear combination of the independent variables, and is fitted using the maximum likelihood estimation method.

10.3 Support Vector Machine

Support Vector Machines (SVMs) are a class of supervised learning algorithms that can be used for classification, regression, and outlier detection. In this paper, we will focus on the classification aspect of SVMs and how they can be used to construct a hyperplane to separate different classes of data.

SVMs work by finding a hyperplane in a high-dimensional space that maximally separates different classes of data. Given a set of training data with input features (x1, x2, ..., xn) and corresponding labels (y1, y2, ..., ym), an SVM algorithm will find the hyperplane that maximally separates the data points of different classes.

The hyperplane is defined by a weight vector (w) and a bias term (b). The equation of the hyperplane is given by:

w*x + b = 0

where x is a feature vector and w and b are the parameters of the hyperplane. The distance of a point x from the hyperplane is given by:

distance = (w*x + b) / ——w——

where ——w—— is the norm of the weight vector.

The goal of an SVM is to find the hyperplane that maximally separates the classes and also has the largest margin, i.e., the distance between the hyperplane and the closest data points from either class. The margin is important because it determines how well the hyperplane generalizes to unseen data. A larger margin leads to better generalization.

To find the hyperplane, the SVM algorithm uses a technique called the "kernel trick". The kernel trick maps the input data into a higher-dimensional space, where it becomes easier to find the hyperplane. The kernel function used depends on the type of data and the complexity of the problem. Commonly used kernel functions include linear, polynomial, and radial basis functions.

Once the hyperplane is found, the SVM can be used to classify new data points by calculating the distance of the point from the hyperplane. If the distance is positive, the point is classified as belonging to one class, and if it is negative, it is classified as belonging to the other class.

SVMs have several advantages over other classification algorithms. They are robust to noise and can handle high-dimensional data. They also have a solid mathematical foundation and have been widely studied and used in a variety of applications.

In summary, Support Vector Machines are a powerful and widely used classification algorithm that can be used to find a hyperplane to separate different classes of data. They work by mapping the input data into a higher-dimensional space and finding the hyperplane that maximally separates the classes with the largest margin. They are robust to noise and can handle high-dimensional data, making them a good choice for many classification problems.

10.4 XGBoost (XBGClassifier)

XGBoost (eXtreme Gradient Boosting) is a popular and efficient implementation of gradient boosting for machine learning. It is a powerful tool for classifying data and has been widely used in a variety of applications, including natural language processing, computer vision, and recommendation systems.

The XGBClassifier is a class in the XGBoost library that implements the gradient boosting algorithm for classification problems. It is a flexible and efficient model that can handle a wide range of data types and can be easily fine-tuned to improve performance.

Gradient boosting is a machine learning technique that combines a set of weak learners (e.g., decision trees) to form a strong learner that can make accurate predictions. The basic idea behind gradient boosting is to train a sequence of decision trees, with each tree attempting to correct the errors made by the previous tree. This is done by fitting the residual errors of the previous tree as a target for the next tree in the sequence.

The XGBClassifier uses a variant of gradient boosting called extreme gradient boosting, which is optimized for speed and performance. It uses a number of techniques to achieve this, such as using approximate second-order gradient information, weighting the contributions of different trees in the ensemble differently, and using sparsity-aware splits.

10.5 AdaBoost (AdaBoostClassifier)

AdaBoost, short for Adaptive Boosting, is a machine learning algorithm used to improve the accuracy of a classifier by combining the predictions of multiple weak classifiers. AdaBoost is an iterative algorithm that trains a sequence of weak classifiers, with each classifier in the sequence being trained on the errors made by

the previous classifier. The final prediction is made by taking a weighted majority vote of the predictions made by each of the classifiers in the sequence.

One of the key advantages of AdaBoost is that it is able to incorporate a wide variety of weak classifiers, allowing it to be used in a variety of contexts. Additionally, AdaBoost is resistant to overfitting and is able to handle large amounts of data, making it a popular choice for use in real-world applications.

To implement AdaBoost, we start by training a base classifier on the training data. The base classifier is typically a simple, weak classifier such as a decision tree with a shallow depth. We then assign weights to each of the training examples, with higher weights being assigned to examples that are misclassified by the base classifier.

Next, we train a second classifier on the weighted training examples. This classifier is also a weak classifier, and its predictions are combined with those of the base classifier to make a final prediction. The weights of the training examples are adjusted again based on the performance of this second classifier, with higher weights being assigned to examples that are misclassified.

This process is repeated for a fixed number of iterations, or until the performance of the classifier reaches a desired level. At each iteration, the classifier is updated to focus more on the examples that it has struggled to classify correctly in the past.

In summary, AdaBoost is a powerful and flexible machine learning algorithm that can be used to improve the accuracy of a classifier by combining the predictions of multiple weak classifiers. Its ability to handle large amounts of data and resist overfitting make it a popular choice for use in real-world applications.

10.5.1 AdaBoost vs XGBoost

AdaBoost and XGBoost are both ensemble learning algorithms that can be used to improve the accuracy of a classifier by combining the predictions of multiple weak learners. However, there are some key differences between the two algorithms:

Boosting method: AdaBoost uses the "Adaptive Boosting" method, while XGBoost uses the "Extreme Gradient Boosting" method. Weak learner: AdaBoost uses decision trees as its weak learner, while XGBoost can use a variety of weak learners including decision trees, linear models, and neural networks. Regularization: XGBoost includes a regularization term to prevent overfitting, which is not present in AdaBoost. Speed: XGBoost is generally faster than AdaBoost due to its use of more efficient data structures and algorithms. Overall, both AdaBoost and XGBoost are powerful ensemble learning algorithms that can be used to improve the accuracy of a classifier. The choice between the two will depend on the specific requirements of the problem at hand and the resources available.

10.6 Random Forest

Random forests are a type of ensemble learning method for classification and regression that operate by constructing a multitude of decision trees at training time and outputting the class that is the mode of the classes (classification) or mean prediction (regression) of the individual trees. The method combines both bias and variance reduction techniques by constructing a large number of decision trees and then aggregating their predictions.

The decision trees that make up a random forest are trained using bootstrapped samples of the training data, with a random subset of the features chosen for each split at each step in the tree-building process. This process is repeated a specified number of times, with the final model being the average or majority vote of all the individual trees.

One of the main benefits of random forests is that they can be trained using very little preprocessing of the data, and are able to handle categorical and numerical features without requiring any additional transformation or encoding. They also have the ability to handle missing values and can be parallelized, making them efficient to train on large datasets.

In addition to their versatility and ease of use, random forests have also been shown to perform well in a variety of settings and have achieved state-of-the-art results on many tasks. They have been applied to a wide range of problems, including credit risk prediction, protein structure prediction, and gene expression analysis.

Despite their many advantages, there are also some limitations to random forests. One disadvantage is that they can be computationally expensive to train, particularly when dealing with large datasets with many features. Additionally, random forests are not as interpretable as some other models, as it can be difficult to understand the specific decisions made by individual trees in the ensemble.

Overall, random forests are a powerful and widely-used tool for both classification and regression tasks, and have been shown to perform well in many different settings.

10.7 LightGBM (LGBMClassifier)

LightGBM is an open-source, gradient boosting framework that is widely used for efficient and effective training of machine learning models, particularly in the field of predictive modeling and classification. It was developed by Microsoft in 2017 and has since gained significant popularity due to its fast training speed, high performance, and low memory usage.

In a classification problem, the goal is to predict the class label of an input instance based on a set of features. LightGBM classifier is a variant of the LightGBM model specifically designed for classification tasks. It utilizes the gradient boosting algorithm to learn a highly predictive function from the input data.

The gradient boosting algorithm works by iteratively training weak models on the input data and combining their predictions to form a stronger model. Each weak model is trained to correct the mistakes made by the previous model in the sequence. This process is repeated until the desired level of performance is achieved or a pre-defined stopping criteria is reached.

LightGBM classifier uses decision trees as the weak models. It constructs the trees in a greedy manner, choosing the split that results in the highest reduction in loss at each step. To prevent overfitting, LightGBM introduces a regularization term in the objective function and uses leaf-wise tree growth, which has been shown to perform better than the traditional level-wise tree growth used in other gradient boosting frameworks.

In addition to these techniques, LightGBM classifier also includes a number of other optimizations such as histogram-based split finding, categorical feature support, and efficient handling of missing values. These features make it particularly suitable for large-scale classification tasks with high-dimensional data.

Overall, LightGBM classifier is a powerful tool for classification tasks that offers fast training speed, high performance, and low memory usage. Its adoption has been widespread in the machine learning community, and it has been used to achieve state-of-the-art results on a number of classification benchmarks.

10.8 K-nearest neighbor

K-Nearest Neighbors (KNN) is a simple and effective machine learning algorithm for classification and regression problems. It is a non-parametric method, which means that it does not make any assumptions about the underlying data distribution. Instead, it relies on the underlying structure of the data to make predictions.

In the KNN algorithm, a new data point is classified based on the majority class of its "nearest neighbors". The number of neighbors, "K", is a hyperparameter that is specified by the user.

To classify a new data point, the KNN algorithm follows these steps:

Calculate the distance between the new data point and all the training data points. Select the K training data points that are closest to the new data point. Determine the majority class of the K nearest neighbors. Assign the new data point to the majority class. One of the main advantages of the KNN algorithm is its simplicity. It requires very little training data and is easy to implement. Additionally, the KNN algorithm can be used for both classification and regression problems, making it a versatile tool in the machine learning toolkit.

However, the KNN algorithm also has some limitations. It can be computationally expensive to calculate the distance between the new data point and all the training data points, especially for large datasets. Additionally, the KNN algorithm can be sensitive to the choice of K and may not always perform well when the underlying data distribution is highly non-linear.

Despite these limitations, the KNN algorithm remains a popular choice for many practical applications, including image classification, natural language processing, and recommendation systems.

In summary, the K-Nearest Neighbors (KNN) algorithm is a simple and effective machine learning method for classification and regression problems. It is easy to implement and can be used for a wide range of applications, but it can be computationally expensive and may not always perform well with highly non-linear data distributions.

10.8.1 Hyperparameters

10.9 Performance

TABLE

10.9.1 Bias

11 Dataset

11.1 Data Generation

12 Implementation

13 Adversities

References

[1] J. A. Cully and A. L. Teten. *A Therapist's Guide to Brief Cognitive Behavioral Therapy*. Department of Veterans Affairs South Central MIRECC, Houston, 2008.

[2] J. Calvo. 2022. *Aprendizaje por transferencia: NLP Blog Europeanvalley. [online] Europeanvalley. es. Available at: https://www.europeanvalley.es/noticias/aprendizaje-por-transferencia-nlp/ [Accessed*, 26, September 2022.

[3] Introduction — Machine Learning — Google Developers — developers.google.com. `https://developers.google.com/machine-learning/guides/text-classification`. [Accessed 26-Sep-2022].

[4] MA. Courtney E. Ackerman. CBT Techniques: 25 Cognitive Behavioral Therapy Worksheets — positivepsychology.com. `https://positivepsychology.com/cbt-cognitive-behavioral-therapy-techniques-worksheets/`. [Accessed 27-Sep-2022].

[5] Asphalion - Scientific and Regulatory Affairs consultancy — asphalion.com. `https://www.asphalion.com`. [Accessed 06-Dec-2022].

[6] Payscale - Salary Comparison, Salary Survey, Search Wages — payscale.com. `https://www.payscale.com`. [Accessed 06-Oct-2022].

[7] Mental health — who.int. `https://www.who.int/health-topics/mental-health#tab=tab_1`. [Accessed 25-Sep-2022].

[8] Michiyo Hirai and George A. Clum. A meta-analytic study of self-help interventions for anxiety problems. *Behavior Therapy*, 37(2):99–111, June 2006.

[9] Benjamin Shickel, Scott Siegel, Martin Heesacker, Sherry Benton, and Parisa Rashidi. Automatic detection and classification of cognitive distortions in mental health text. *Automatic Detection and Classification of Cognitive Distortions in Mental Health Text*, 9 2019.

[10] Ignacio de Toledo Rodriguez, Giancarlo Salton, and Robert Ross. Formulating automated responses to cognitive distortions for cbt interactions. 2021.

[11] Daniel M Low, Laurie Rumker, John Torous, Guillermo Cecchi, Satrajit S Ghosh, and Tanya Talkar. Natural language processing reveals vulnerable mental health support groups and heightened health anxiety on reddit during covid-19: Observational study. *Journal of medical Internet research*, 22(10):e22635, 2020.

[12] Degree final project — FIB - Barcelona School of Informatics — fib.upc.edu. `https://www.fib.upc.edu/en/studies/bachelors-degrees/bachelor-degree-informatics-engineering/degree-final-project`. [Accessed 29-Nov-2022].

[13] botui/LICENSE at master · botui/botui — github.com. `https://github.com/botui/botui/blob/master/LICENSE`. [Accessed 03-Dec-2022].

[14] Writing a GDPR-compliant privacy notice (template included) - GDPR.eu — gdpr.eu. `https://gdpr.eu/privacy-notice/`. [Accessed 05-Dec-2022].

[15] Art. 13 GDPR - Information to be provided where personal data are collected from the data subject - GDPR.eu — gdpr.eu. `https://gdpr.eu/article-13-personal-data-collected/`. [Accessed 05-Dec-2022].

[16] Chapter 3 (Art. 12-23) Archives - GDPR.eu — gdpr.eu. `https://gdpr.eu/tag/chapter-3/`. [Accessed 05-Dec-2022].