• U C •

FCTUC FACULDADE DE CIÊNCIAS
E TECNOLOGIA
UNIVERSIDADE DE COIMBRA

Mónica Catarina da Providência Wolters

Depressive States Identification on Social Networks Using Multimodal Models

Thesis submitted to the
University of Coimbra for the degree of
Master in Biomedical Engineering

Supervisors:
Prof. Dr. João Carvalho (Physics Department, Faculty of Science and Technology of the University of Coimbra)
Eng. Miguel Tavares (Whitesmith, Lda.)

Coimbra, 2018

This work was developed in collaboration with:

Whitesmith, Lda.

"I think everybody's weird. We should all celebrate our individuality and not be embarrassed or ashamed of it."

Johnny Depp

Acknowledgments

First of all, I would like to thank Whitesmith for making this adventure possible, for all the support and caring showed during the last year. A special thanks to my supervisor Miguel Tavares for all the patience to guide me through this project and teach me the topics I did not master.

I want to thank my University supervisor João Carvalho for going along with me on this project and for motivating me even when the outcome was not what I expected.

One of the biggest influencers of my academic journey was jeKnowledge, the Junior Enterprise of the Science and Technology Faculty of the University of Coimbra, which contributed to most of the soft skills I acquired during these five years and brought me a group of beloved people I will always carry with me. I'm thankful for all the incredible challenges it put me through, making possible for me to make a difference in the Junior Entrepreneurial European Movement.

To all my friends with whom I shared so many laughs, late hour projects, incredible trips, stressful exam weeks, unique academic traditions, and all the good and bad moments that came along these last years. I hope I can count on you for many more years and continue making incredible memories by your side in this new chapter of our lives. A special thanks to my awesome academic family, for all the support in these last months, for all the advice, counseling and all the kind words.

Finally to my mother, my father, and my sister, and to all my family for being my support for all these years, during the ups and downs, for always making me laugh and for always encouraging me to pursue my dreams and embark on new adventures. I want to thank you for all the great moments we spent together, for helping me grow in every single way and for making all of this possible, I would not have done it without you. I hope to return to you, as much as you give to me.

Resumo

Depressão é uma doença mental comum por todo o mundo. Esta condição pode causar muito sofrimento ao paciente, e também afetar o trabalho, escola e vida familiar. Em casos extremos, a depressão pode mesmo levar ao suicídio. Em Portugal, estima-se que cerca de 400,000 individuos sofram de depressão por ano. É também a maior causa de suicidio, sendo responsável por 70% dos suicídios.

Atualmente, os jovens dependem muito das redes sociais, principalmente os que sofrem com depressão. Estes jovens utilizam a internet para substituir as relações pessoais reais, procurando reconhecimento social e distanciamento da vida social habitual. A depressão tem tratamentos eficazes, mas menos de metade da população afetada acede a estes tratamento devido ao estigma social associado a doenças mentais e falta de recursos e profissionais de saúde treinados.

Neste projeto foi criado um modelo multimodal capaz de analisar textos de publicações, imagens partilhadas pelo utilizador e padrões de interação e utilização destes indivíduos. Este projeto foi dividido em quatro partes: distinção entre utilizadores depressivos e não depressivos, distinção entre os quatro estados de depressão (mínimo, suave, moderado e severo), previsão do score exato de depressão através de uma regressão e análise de depressão em imagens através de uma rede neuronal convolucional.

O nosso modelo atingiu uma exatidão de 90,5% aquando a previsão entre utilizadores depressivos e não depressivos e de 76,2% aquando a classificação entre os quatro estados de depressão, utilizando o Classificador Random Forest em ambos os casos. Na previsão dos scores exatos de depressão, foi obtido um modelo com um MSE de 79,37. Na previsão de depressão em imagens foi aplicada a técnica de transferência de aprendizagem com a rede pré-treinada VGG16 e obteve-se uma exatidão de 70,97%.

Palavras-chave: Aprendizagem Máquina, Depressão, Análise de Redes Sociais, Análise de Sentimento, Redes Neuronais Convolucionais. SVM

Abstract

Depression is a common mental illness all over the world. Especially when long lasting and with moderate to severe intensity, depression can cause a lot of suffering in the affected individual, having implications at work, school and family life. In extreme cases, depression can lead to suicide. In Portugal, depression is estimated to affect around 400,000 people per year. It is the leading cause of suicide, with a total of 70% when compared to other causes of death.

Nowadays, young people heavily rely on social networks, particularly those suffering from depression. These people use the internet to replace relationships in the real world, seeking social recognition and alienation from the ordinary social life. Depression has effective treatments, but less than half of the affected people receive them due to the social stigma associated with mental disorders and lack of resources as well as trained healthcare providers.

With the aim of predicting if a user is going through depression or not, we created a multimodal model, capable of analyzing the text of publications, images shared by the user and patterns of interaction of these individuals. This project was divided into four approaches: the distinction between depressive and non-depressive users, the distinction between the four states of depression (minimal, mild, moderate and severe), prediction of the exact depression scores using a regression and image depression analysis prediction using a convolutional neural network.

Our model got an accuracy of 90.5% when predicting between depressive and non-depressive users and 76.2% when classifying in one of the four depression states, both with the Random Forest Classifier. For exact depression score prediction we obtained a model with a mean squared error of 79.37. For the image depression analysis prediction we applied transfer learning using the trained VGG16, obtaining an accuracy of 70.97%.

Keywords: Machine Learning, Depression, Social Networks Analysis, Sentiment Analysis, Convolutional Neural Networks.

List of Acronyms

ANN Artificial Neural Network.
API Application Programming Interface.

BDI Beck Depression Inventory.
BDI-IA Beck Depression Inventory IA.
BDI-II Beck Depression Inventory II.

CNN Convulotional Neural Network.

KNN K-Nearest Neighbors.

MSE Mean Squared Error.

SVM Support Vector Machine.
SVR Support Vector Regression.

List of Figures

List of Tables

Contents

1

Introduction

1.1 Contextualization

Depression is a common mental illness all over the world. It is quite distinct from the usual mood fluctuations and short-term emotional responses to the day-to-day life since its symptoms normally last for at least two weeks. Especially when long lasting and with moderate to severe intensity, depression can become a serious health condition. This can cause a lot of suffering in the affected individual, having implications at work, school and family life. In extreme cases, depression can lead to suicide.[1, 2]

Depression has become more and more frequent in our society. According to the World Health Organization, more than 300 million people have this mental illness. This number increased more than 18% since 2005.[1]

Depression is now the worldwide leading cause of disability and the cause of about 800,000 suicides per year.[3] Amongst young adults, the number of people affected by depression has also increased from 8.8% in 2005 to 9.6% in 2014, and when analyzing statistics of young adults, suicide has become the second leading cause of death.[4]

In Portugal, depression is estimated to affect around 400,000 people per year, and is the leading cause of suicide, with a total of 70% when compared to other causes of death.[5]

1.2 Motivation

Nowadays, young people heavily rely on social networks, particularly those suffering from depression. These people use the internet to replace relationships in the real world, seeking social recognition and alienation from the ordinary social life.[6]

Depression has effective treatments, but less than half of the affected people receive them due to the social stigma associated with mental disorders and lack of resources as well as trained healthcare providers. It is also quite frequent people not being accurately diagnosed, which means the ones with the disease end up not receiving treatment and the ones who do not have the disease are prescribed antidepressants.[7]

With the development of technology, these days it is possible to track the publications of young adults on social networks and relate some of their behaviours to depression. This could help these members of society being self-aware of their condition allowing a more natural contact with organizations that would help getting the needed care.

1.3 Goals

This project aims to develop a rating system capable of collecting and processing social networking data and associate each user with a score that reflects the probability they are going through or into an episode of depression.

For this purpose, a multimodal Machine Learning system will be created, capable of analyzing the text of publications, images shared by the user and patterns of interaction of these individuals, and understand how this can be related to depression and help to improve the approaches used these days.

The target social networks will be Facebook, Instagram, and Twitter, allowing the analysis of text, images, shared content and behavior.

This project aims to develop a tool that could help young adults, since this is the age range that uses social networks the most, to get the necessary support from associations and, afterwards, from clinics and medical professionals, decreasing the number of individuals who have this mental disorder not diagnosed and preventing significant consequences.

1.4 Structure

Firstly, some general insights for this project are given, so that in this work we can go into detail about the problem that we aim to solve, the methods used and the different results obtained.

In the next chapter, State of the Art, it describes what depression is and what are some of the leading causes for this disease among young adults, followed by some information about the depression test used for scoring participants. Some of the most relevant researches in social network and sentiment analysis will also be reviewed there.

This will be followed by the Dataset and Methods chapter where participants' information is given, along with information on how the dataset was arranged, the different features extracted from each social network, the problem formulation and the methods used for its resolution, as well as the evaluation methods.

The following chapter is about Results and Discussion, where all the results of the different problems will be presented with some discussion.

Lastly, there is the Conclusion chapter where some future work is suggested, along with a discussion on some of the main challenges during this project.

2

State of the Art

2.1 Depression

Depression is a mental disorder, which is characterized by negative mood and thoughts. Joy and pleasure in life are lost, such as self-esteem and performance. The symptoms may sometimes occur in healthy individuals, but in the depression, they persist for at least two weeks, being more severe and visibly decreasing the quality of life of the individual.[2]

Some of its symptoms are permanent sadness, anxiety, pessimism, irritability, feelings of guilt, fatigue, difficulty of concentration, abnormal sleep and appetite patterns, and pain.[8]

There are several types of depression like major depressive disorder, persistent depressive disorder (dysthymia), postpartum depression, psychotic depression, seasonal affective disorder, bipolar disorder, treatment-resistant depression, subsyndromal depression, premenstrual dysphoric disorder, disruptive mood dysregulation disorder, substance-induced mood disorder or depression due to an illness.[2, 14]

On this project, we are focusing on the major depressive disorder. This condition is thought to be caused by a combination of three factors: genetic, environmental and psychological.[2] Risk factors such as family history of depression, significant changes in life, medications, chronic health problems and drug use may play an essential role in the development of the disease.[2, 8] It is estimated that about 40% of the risk is related to genetic factors.[8]

The disease is not diagnosed through specific laboratory tests[8], but other tests are done that may rule out other diseases with identical symptoms. Such as TSH (thyroid-stimulating hormone) and thyroxine blood tests to exclude hypothyroidism; basic electrolytes and serum calcium to shut off a metabolic disorder; a complete blood count, including ESR (erythrocyte sedimentation rate), to reject a systemic infection or chronic disease, and in older patients it is first necessary to reject dementia.[9]

The diagnosis is made by a psychologist or psychiatrist, who collects family cases, symptoms and current circumstances of the patient, as well as its history.[2] The mental health examination may include the use of a rating scale, such as the Hamilton Rating Scale for Depression[10], the Beck Depression Inventory[11] or the Suicidal Behavior Questionnaire-Revised[12]. However, these scales only serve as diagnostic aids and should not be used individually.[13]

2.1.1 Depression among Young Adults

A significant depression factor in the younger age group is stress. Young adults have experienced wars, economic crises, hear that an undergraduate degree is not enough. This stress can increase the risk of developing depression. Many young people nowadays also do not sleep the necessary resting hours, often staying up until dawn and then resorting to medicines and energy drinks to stay active the next day.[4]

The incidence of depression increases during adolescence and peaks early in adulthood. [17] This phase, defined as in between 18 and 30 years of age, is a time of significant change and can be associated with a higher risk of mental illness and high levels of stress.[15, 16]

The causes for this condition are complex, but unemployment or low-rewarding jobs, with few cognitive requirements and autonomy, can often be associated with depression among young adults.[18, 19] In the case of males, the higher the job status, the lower the signs of depression, while in females the signs of depression increase with the physical danger of employment.[20]

This age range is also a period when young people marry or start families. Although these moments can bring much happiness, new financial burdens, career demands, adjustment to married life and birth among young couples can influence negatively the mental health, especially women.[21]

Within this group of young individuals, women are twice as likely to have depression as men because they are primary caregivers for children and households, while at the same time have less labour and cultural control and power than their peers.[4]

On the other hand, this period also coincides with the legal age for alcohol consumption. Young people who frequently consume alcohol or tobacco are more likely to experience depression.[22, 23]

Depression in adolescence is associated with higher levels of depression and poorer health in the young adult age.[24, 25] An excellent parental relationship, especially close maternal relationship, is related to a lower likelihood of depression when entering the adult life.[26] A dangerous environment within the family, such as conflicts due to divorce or separation, can lead to much stress and frustration. Poverty, abuse, and violence can also be harmful to young people's mental health. Problems in relationships with other young people, whether loving or friendship, can additionally lead to loss of confidence and self-esteem problems.[14]

In addition to these factors, there are many other related to social pressure. In modern society, young people are fed with information that pressures them to meet often unrealistic needs. They are flooded with images of what they should look like, should have, and how to behave. If they feel they can not achieve what is expected, they may feel misplaced, not good enough or in disadvantage, and this can progress to depression.[14] A school is a definite place for learning, growing and developing, but it can additionally be very harmful. Performance pressures and stress on exams can cause frustration and trigger long-term depressed feelings. Bullying is becoming more common, both personally and through social networks, leading the victim to feelings of anxiety, low self-esteem, and poor concentration.[14]

Other events that may trigger depression in young adults are loss or illness of a loved one, own physical illness, cohabitation with a relative with depression, weight changes, embarrassing events, financial problems, or physical or emotional abuse.[14]

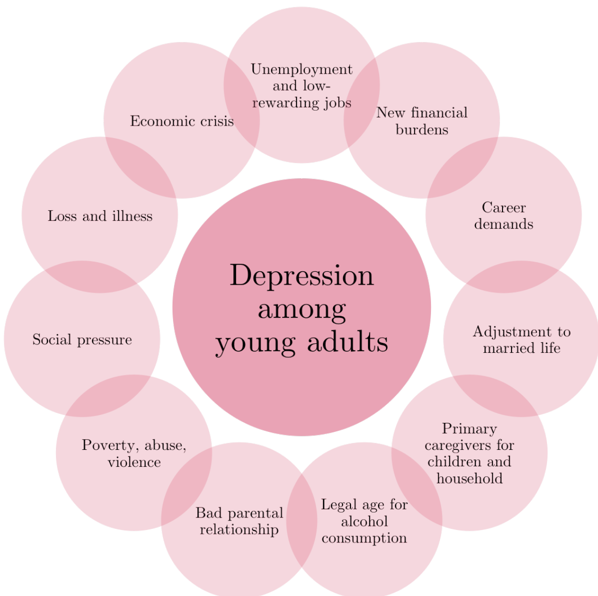


Figure 2.1: Causes for depression among young adults.

2.1.2 Beck Depression Inventory

The Beck Depression Inventory is a multiple-choice test composed of twenty-one questions created by Aaron T. Beck. It is one of the most commonly used psychometric tests worldwide in screening for severe depression.[28] This test allowed a different approach to the diagnosis of this disease.

There are three versions of this test, the first one published in 1961.[29] In the second version, BDI-IA, were removed options that caused some doubt and were worth the same score, and it was pointed out that the test was relative to what the patient felt during the previous two weeks.[30, 31]

The internal consistency for this version was good, with a Cronbach alpha coefficient of about 0.85, which means that inventory items are highly correlated with each other.[32] Cronbach's alpha coefficient is a way of estimating the reliability of a psychometric test. It measures the correlation between answers in this test by analyzing the answers given, presenting a mean correlation between the questions. The coefficient alpha is calculated from the variance of the individual items and from the variance of the sum of the items of each evaluator of all the items of a survey that uses the same measurement scale.[27]

Despite these improvements and good results, this version only referred to six of the nine depression criteria mentioned in the Diagnostic and Statistical Manual of Mental Disorders.

The last version of this test was published in 1996, developed especially after changing the diagnostic criteria of the same Manual mentioned above.[31] Items related to body image, hypochondria and work difficulties were replaced. Items such as loss of appetite and sleep were remodelled, having the possibility of increased appetite and number of sleeping hours. The final punctuation ranges of the test were also slightly modified.

The current version of this test, the BDI-II is used by patients from thirteen years of age and contain questions regarding symptoms such as irritability, guilt, abnormal eating and sleep patterns and changes in sexual life.

Each question has four response options, scored from 0 to 3, and at the end of the test, the higher the score, the more severe the depression. The score is distributed as follows:

- 0-13: minimal depression

- 14-19: mild depression
- 20-28: moderate depression
- 29-63: severe depression

The Cronbach alpha coefficient for internal consistency is 0.91, higher than the previous one.[31] Also when compared with other tests it obtained good results, an essential method of evaluating these new versions before being published. When correlated with the Hamilton Depression Rating Scale presented a Pearson r of 0.71.[33] The Pearson r or the Pearson correlation coefficient is used in statistics for measuring the linear correlation between two variables.

2.2 Social Network Service

A social networking service is an online application used to build social networks or social relationships between people.[36] It is often referred to as Web 2.0, since they are web-based people communities that interact with each other in the form of conversation, sharing of content (text, images, videos, news), forums, among others. The idea behind these networks is that its users feed its content.[34]

However, beyond these aspects, it is difficult to find a standard definition, since there are several social networks, very different from each other. Danah M. Boyd and Nicole B. Ellison give a possible definition *as web-based services that allow individuals to (1) construct a public or semi-public profile within a bounded system, (2) articulate a list of other users with whom they share a connection, and (3) view and traverse their list of connections and those made by others within the system.*[35] Social networks are operable on desktops, laptops, and mobile devices.

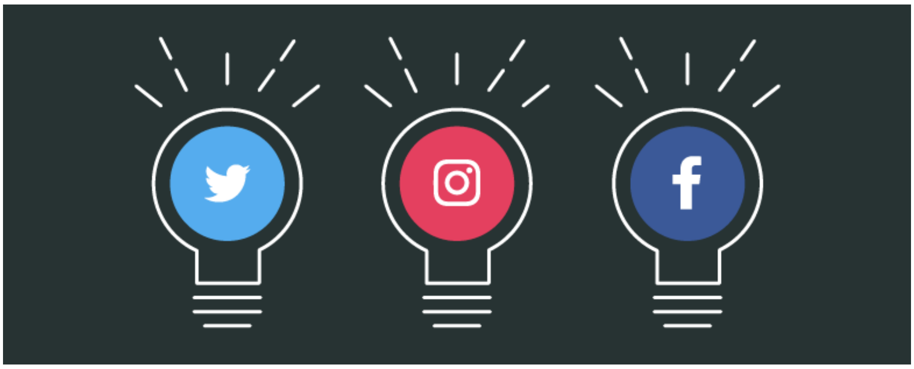


Figure 2.2: Some of the most used social networks, and the used in this project. Twitter, Instagram and Facebook. [83]

2.2.1 Social Networks' APIs

Application Programming Interface enables developers to write software that interacts with other applications, in this case, social networks. Some companies release public APIs so that their users will use them in their apps. There are also private APIs, for example when a company has multiple products and needs them to interact with each other.[68]

2.2.2 Text Sentiment Analysis on Social Networks

The words we use in daily life reflect who we are and the social relationships we are in.[60]

The most common way for people to express their thoughts and feelings is by using language. Language is also the way social psychologists try to understand us. Written language is no different from the spoken one, so with the development of technology and social networks, there are new possibilities to study human behavior and self-reported measures of personality, social behavior and cognitive styles.[60]

With the evolving of social networks, people increasingly generate content. Microblogs have every day more people sharing their thoughts because of its simplicity and quickness.[59]

Lamentably, there were no datasets available for sentiment analysis, making the development of this area very difficult since it is tough to get enough data for these researches. In 2013 a shared research called “Sentiment Analysis on Twitter” started, where more than forty teams participated, making it easier to have a suitable dataset.[58] Although this was a significant improvement for the text analysis on social networks, there is still little information from other social networks apart from Twitter, resulting in many publications regarding only on this microblog.

Text analysis is the most used analysis on social networks for the last 10-15 years. Several studies based on mood analysis have shown good results. Differentiating between negative and positive language, analyzing the frequency and time of posting, social engagement and linguistic style used, have been the most used methods in these studies.[55, 56]

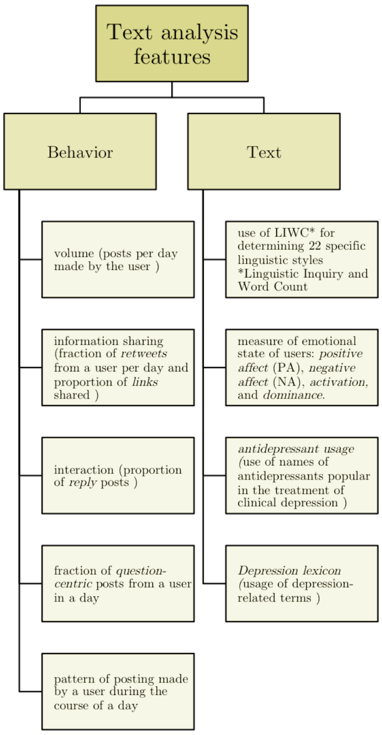


Figure 2.3: Text features extracted from the microblogs regarding M. Choudhury, S. Counts and E. Horvitz researches. [55]

Researchers M. Choudhury, S. Counts and E. Horvitz have several studies on this issue in which they conclude that people with depression had reduced social activity, a more negative dialogue, a greater focus on themselves, more significant concerns about medication and relationships and increased manifestation of religious thoughts. They were able to construct a SVM classifier that could predict depression in Twitter publications with 70% accuracy in one study and 73% in another.[55, 56]

S. Batra and D. Rao developed a sentiment cloud from labelled entities' Twitter posts, where words used in each post were classified as positive or negative, regarding its frequency in positive/negative posts.[57]

M. Krieck, J. Dreesman, L. Otrusina and K. Denecke tried to detect disease outbreaks using text analysis in the microblog Twitter. They conclude that it was difficult to have useful information from personal users about symptoms or other possible disease outbreaks, only news and information from public entities and individual subjects about the diseases were easily found. [61]

M. Park, C. Cha and M. Cha analyzed the behavior of a small group of Twitter users, 28 with moderate to severe depression and 41 with low depression (being considered as non-depressive). They concluded that users in the depression group when posting about their disease, they shared private and detailed information about the reasons that would make them or not depressed, medical diagnosis, medications changes and personal remedies. These users also posted more about themselves with a high prevalence of expressions related to anger, causation and friends. An interesting point they faced was that although several studies show a higher probability of depression in women than in men, they found an equal prevalence. A possible explanation they give was that female users are less likely to reveal their depressive thoughts on social media than men.[62]

2.2.3 Image and Multimodal Analysis on Social Networks

Contrary to the extensively studied text-based sentiment prediction, image-centred sentiment analysis receives much less attention. Recently, there has been the explosive popularity of image sharing social networks like Instagram. With these social networks, users express opinions and sentiments in a much more indirect way that in text-based networks. The understanding of these posts can have a significant benefit in a lot of real-world applications such as advertisement, recommendation, marketing and health-care, being this last one a crucial one for our society.[63]

Current work done in this area has approaches based on low-level visual features [64], mid-level visual features [65] and deep learning [66]. Low-level visual features are the characteristics of an image like the brightness, contrast and dominant colours. The mid-level visual features describe objects or characteristics seen in the image like an animal, the sky and a tree. For the deep learning, neural networks are used when predicting emotions in images.

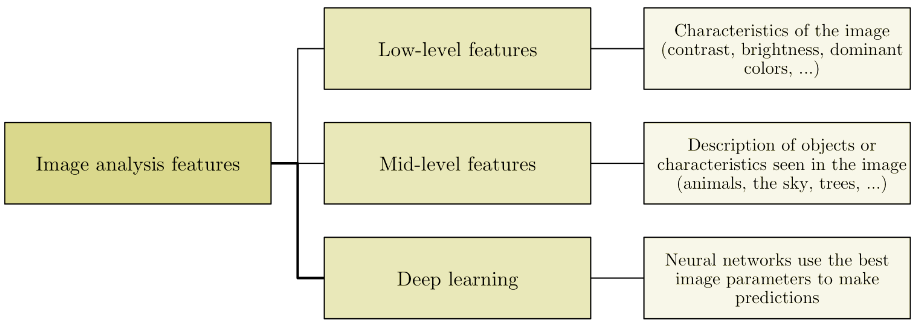


Figure 2.4: Types of image analysis features extracted.[64, 65, 66]

Most of image analysis work uses supervised learning, relying on labelled images to predict emotions, by training the classifiers with this information.[63] Low- and mid-level features do not have the semantic information to predict moods and emotions directly in images.[63]

Sentiment analysis of visual content can increase the extraction of user information about events, emotions or other topics, complementing the textual analysis.[66]

A picture is worth a thousand words. People with different backgrounds can easily understand the main content of an image or video.[66]

Today's computational infrastructure is much cheaper than some years ago and makes analyzing this vast amount of visual content available in social networks achievable.[66] During this significant data period, it has been concluded that integrating visual content in online social research has provided us with much better results and information.[67]

Few works attempted to predict emotions using features from images since it is incredibly challenging. Sentiment analysis is more laborious than object recognition since it is necessary a higher level of abstraction and subjectivity and a variety of visual tools to recognize life objects, scenes and actions. To be able to research this topic using supervised learning, it is essential to have a large and diverse labelled dataset, which requires a lot of work and time. These datasets also need to be generalized to cover all, or most of the situations.[67]

Because of these difficulties, much work using unsupervised learning has been done, using as input the several images and textual tags related to them so that the model could predict emotions in the images. So, in this case, images were extracted from social networks like Instagram and Flickr, as also tags that were part of these posts.[63]

Another type of research done on this topic is deep learning as mentioned previously. Here, the images are used to train a supervised convolutional neural network to predict and subsequently forecast sentiments in new images.[66]

A convolutional network takes the image pixels as input features. The layers are divided into two components. The hidden layers also called feature extraction part, is where the network performs a series of convolution and pooling operations and detects the features present in it. Afterward, comes the classification part, where the fully connected layers will work as a classifier on top of these extracted features.[82]

Several researches like You, Quanzeng, et al.[81] have used Flickr, a social network for sharing images, to extract labeled images (tags accompany them) to train convolutional neural networks and then be able to do sentiment analysis to new input data, for example from Twitter. These images were then classified into negative or positive, presenting results of the accuracy of 77.3% [81].

3

Dataset and Methods

3.1 Dataset

In order to construct a dataset, a project registration form was prepared (Appendix B). This form requested only some essential information from the users, such as e-mail to be able to make subsequent follow-ups, date of birth because only individuals from 18 to 30 years of age are being studied, the users accounts on social networks (Facebook, Instagram, and Twitter) and the filling of a survey of depression, in this case the BDI-II (Appendix A).

For the analysis of participants' depression, we wrote a script that covered participants' responses by adding a new column with the score obtained in the BDI-II survey.

Then, the social networks' profiles of each participant were analyzed, in order to understand if all had profiles in all social networks. The form with the BDI-II test was re-sent four months after the first one, in order to have another posts' period.

In total, we had 83 participants answering to our project, but we had to reject some because they were out of our age interval or did not have any of the three social networks, having in the end only 76 participants.

Since we asked for the users to answer a second time, we consider this as new users because their depression score was not exactly the same. In total we have 117 "users", having 72.6% with no depression (minimal depression) and 27.4% with depression (56.3% with mild depression, 15.6% with moderate depression and 28.1% with severe depression).

3.1.1 Instagram

From Instagram, we had 61 users with an account. From these, we collected all the posts from the individuals, including their photos, descriptions, likes and comments, through a package made available by a user in GitHub [37].

For this data collection, it was only necessary for the user to have a public profile or, in the case their profile was private, to accept being followed. Therefore, an account was created in this context.

After this collection, data that was not inside each user's study period was eliminated. The different features were extracted into three different datasets as outlined in the figure 3.1. These features were chosen by analyzing different publications in this topic; the text and usage metrics' features were chosen based on the researches done with twitter and other microblogs, the image features were chosen based on the works done with image analysis, using low-level visual features. The text features were chosen by analyzing some different posts, because there was not much information regarding features that were not linguistic, like size of the text, existence/number of tags and emojis.

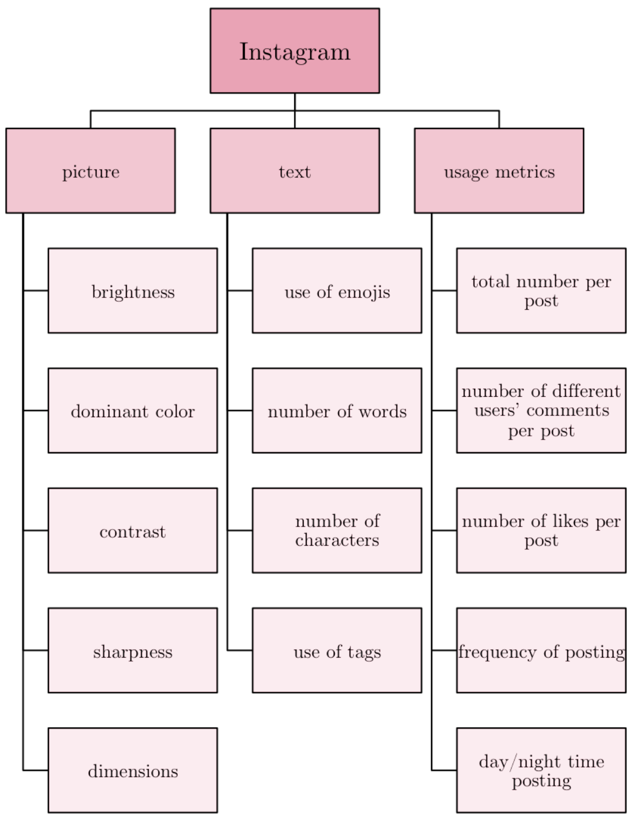


Figure 3.1: Features extracted from the social network Instagram.

We have a total of 85 post coming from users non depressive and 67 from users depressive (28 with mild depression, 16 with moderate depression and 23 with severe depression).

3.1.2 Facebook

From our 76 participants, 71 had a Facebook profile. To be able to collect data from Facebook users, a website (Appendix C) had to be constructed and a Facebook app. This website was only necessary in order to send it to the project participants and for them to log in to their Facebook account and permit us to collect their posts. This Facebook app was constructed and submitted but, because of the new regulations and unfortunate privacy scandals, Facebook restricted much of the information that could be extracted and our app was rejected for our post collecting features. Without this permission, we could not collect the data from Facebook.

During this accepting period, we analyzed what would be the best features to extract once the Facebook app was accepted. The features chosen were selected by the same criteria as the Instagram ones.

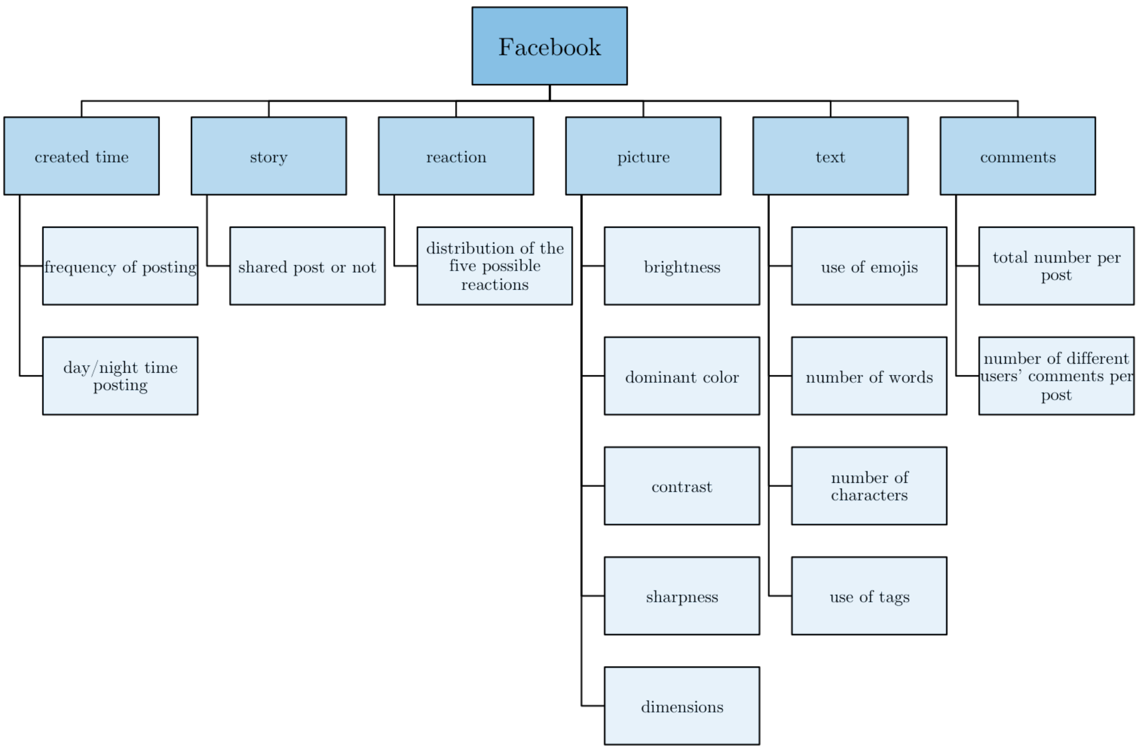


Figure 3.2: Features extracted from the social network Facebook.

3.1.3 Twitter

For Twitter, a script was written for accessing the Twitter API and collecting the tweets made by users. After analyzing the number of participants who had a twitter account, only 13, and from these the low regularity of its use, it was decided to leave this social network out, since it would not bring us much benefit in this study.

3.1.4 Features

The final features extracted from Instagram are specified in the tables 3.1, 3.2 and 3.3.

Table 3.1: Text features extracted for our dataset.

Text Features
Number of characters
Number of words
Existence of emojis
Number of emojis
Number of tags
Existence of tags

Table 3.2: Usage metrics' features extracted for our dataset.

Usage metrics' Features
Day/Night time
Number of likes
Number of comments
Number of different people comments
Days since last post
Average post frequency

Table 3.3: Image features extracted for our dataset.

Image Features
Brightness
Contrast
Sharpness
Red color quantity in the dominant color
Green color quantity in the dominant color
Blue color quantity in the dominant color
Height
Width

3.2 Methods

3.2.1 Problem Formulation

Now that we have all the data gathered, we have to decide how we want to handle the initial problem and what the best approaches are, what classifiers will we use and how will the results be interpreted.

Is it possible to predict if a post is coming from a user suffering from depression or a healthy user?

To answer this question, we are dividing our study into two parts: first, we will try to understand if the user behind the post is or not suffering from depression.

Is it also possible to predict between the different states of depression?

For the second part, we will not only try to predict if the user has this mental disorder but also in what stage he is right now (minimal, mild, moderate or severe depression).

So, we are facing a binary problem and a multiclass problem.

Would it be also possible to predict the exact depression score of the user?

After the first approach, we will also try to predict the exact score from the user behind the post. For this, we are using a Regression Model that will give us the exact depressive score, between 0-63. With this information, the user score could be tracked and recognize if there was being an improvement or rise of the depression. We would also be able to test users without needing them to fill a form.

And what if our features are not good for these predictions? Are there other ways for predicting between depressive and non-depressive users?

Apart from this two approaches that only uses features selected by us on the posts, usage metrics and pictures, we are also classifying our image data by using transfer learning with a convolutional neural network. The downside to this approach is that we do not know which features the CNN extracted from our pictures.

3.2.2 Supervised and Unsupervised Learning

In machine learning, we can divide the models into two main algorithm categories: supervised learning and unsupervised learning. Although machine learning algorithms keep evolving, they always fit in one of these two categories.[38]

Supervised Learning is the simplest model. Here, a function is trained with a training dataset and then applied to a new group of data, never seen by the model, to evaluate its prediction capacity. In this model, the dataset always comes with a labels vector that is used both to train and test the models. The goal of this models is to be able to generalize and predict correctly any data. The initial dataset is divided into two groups, one for training and the other for testing. The training dataset is used to train the function with the chosen algorithm until we manage to get a good model, then this model is tested with our test dataset. These data is a sample of data not used for training and is a useful evaluation method for our trained function.[41] These models use classification techniques, for the discrete responses, or regression techniques, for the continuous responses, to predict the output.[38] Some of the most widely used algorithms for supervised learning are support vector machines, linear regression, logistic regression, naive bayes, linear discriminant analysis, decision trees, k-nearest neighbor algorithm and neural networks.[39]

Unsupervised Learning does not require labels; the model classifies the data into different clusters based on hidden features. Firstly, the function segments the dataset into classes, so every input data becomes part of one of these classes, which are not labelled by the algorithm. Unlike the previous model, in this there are no correct answers, so the models cannot be evaluated.[41] They can be grouped in two different problems: clustering problems, where there are discovered different groups inside the input dataset or association groups, where specific rules are discovered that describe the dataset or significant portions of it.[40]

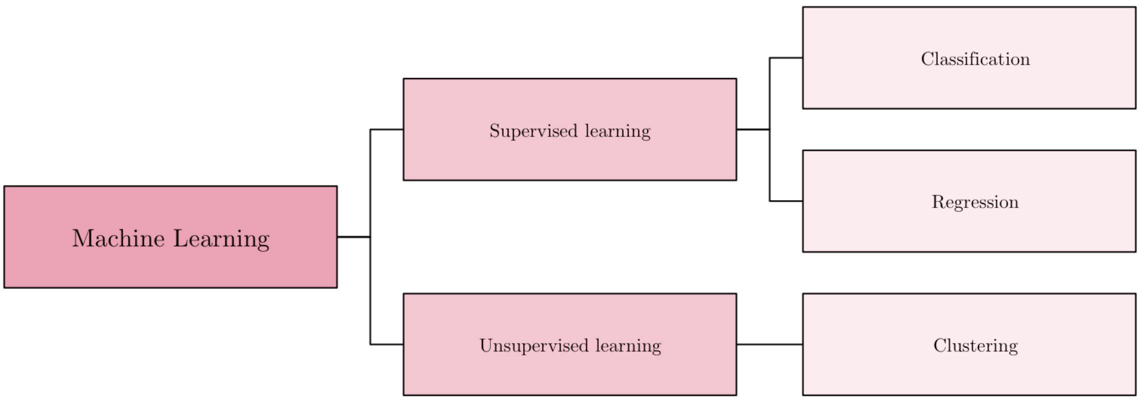


Figure 3.3: Algorithm categories for machine learning.

3.2.3 Feature Analysis and Feature Reduction

In this project, for feature analysis, we used two different tests. For the binary problem we used the Student's T-test, and for the multiclass problem, we used ANOVA (Analysis of Variance) F-test.

The Student's T-test is a method used for testing the null hypothesis that two groups have a similar mean score. Observing the p-value obtained from this test, if it is more significant than α, a threshold defined usually as 0.05 or 0.1, then we accept the null hypothesis, otherwise we can reject it and conclude that the groups have statistically different means and should be good to help distinguish these two groups, for example.[46]

The one-way analysis of variance (ANOVA) is used in statistics to measure if there are significant differences between the mean value of more than two groups. These groups have to be independent, and each dependent variable has to be normally distributed.[45]

In this test, the null hypothesis is that the means of these groups are identical. We will accept this hypothesis if we get a p-value larger than α. This α value is defined taking into consideration each set of data, so if the features we are analyzing have all high values of p-values, we set the α as 0.1. In the case that most of the p-values are lower than 0.1, we will set it as 0.05, so we have the most relevant features. If we get a p-value lower than α, we will take the alternative hypothesis, which at least two of the groups are statistically significantly different from each other, but we cannot know which groups with this test.[45]

3.2.4 Algorithms for the Classification and Regression Problems

For these approaches, we are going to use three different classifiers: Support Vector Machines (for the binary classification and regression problem), Random Forest (for the classification and regression approaches) and K-Nearest Neighbors (for the classification approaches).

3.2.4.1 Support Vector Machine/Support Vector Regression

A Support Vector Machine (SVM) is a supervised learning classifier that defines a hyperplane that separates two groups. For this, labelled training data is given, and the output is a hyperplane that tries to categorize new, unseen data.[47]

In the case we cannot define a line in a two-dimensional plane, a third dimension is added, the separation line defined and then transformed to the original plane, is possible to get a circle instead of a simple line as the output. These transformations are called kernels.[47]

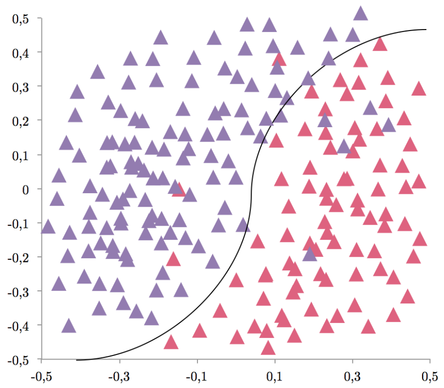


Figure 3.4: Example of a non-linear hyperplane dividing two classes with SVM.

To obtain better accuracy at predicting the right class for each point, the algorithm has some parameters that help increase accuracy in a reasonable amount of time.[47]

As mentioned previously, we have the kernel parameter. The kernel avoids the explicit mapping that is needed to get linear learning algorithms to learn a nonlinear function or decision boundary. There's also a regulation parameter, known as the C parameter, that defines how much we want to avoid misclassifying each training sample. For larger values of C, the hyperplane will present a smaller margin between the classes, if it can classify the training points correctly. On the other hand, choosing a smaller value for this parameter, the separating hyperplane will have a larger margin, but probably misclassifying more points. Another important one is the gamma parameter, used when we use a gaussian kernel, which defines which points influence the separation line between classes, meaning that for low values, points that are farther away from this line will be taken into consideration, and for higher values, only the closest ones are used.[47]

The Support Vector Regression uses the same algorithm as the Support Vector Machine Classifier but with some differences, like the parameter epsilon, a margin of tolerance that is set because the output prediction has infinite possibilities since the output is a real number and not a class. Another thing is to minimize the error; the margin should be maximized individualizing the hyperplane so that part of the error is tolerated.[54]

3.2.4.2 Random Forest Classifier and Regressor

The Random Forest algorithm is known as an ensemble algorithm, which is an algorithm that combines more than one for a better classification.[48] This model builds multiple decision trees and fuses them together to increase the accuracy of the prediction.[49] Decision trees are algorithms that divide the training data into different subgroups by identifying lines.[50] Despite that, decision trees and random forest are not the same thing: decision trees will formulate some rules when receiving a labelled training set, that will be after used for predictions with other data. On the other side, the random forest algorithm selects features and observations randomly and builds more than one decision tree and then averages the results of each of them. The Random Forest classifier also prevents overfitting by, before building the decision trees, creating smaller random subsets.[49]

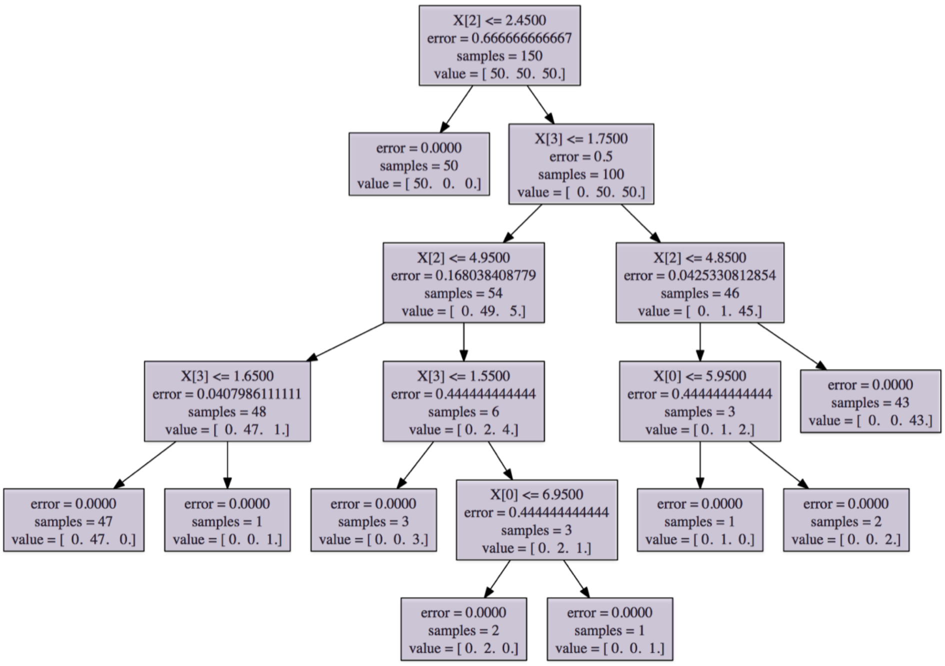


Figure 3.5: Example of a Decision Tree diagram for Random Forest.

Random Forest adds additional randomness to the model while growing the trees. Instead of searching for the most important feature while splitting a node, it searches for the best feature among a random subset of features. This results in a wide diversity that generally results in a better model.[49]

Like the previous models, this one also has some hyperparameters to increase the predictability of the data. We have the 'n_estimators' parameter, which is the number of trees built by the algorithm. The higher, the better the predictions, but the slower the computation. There's also the 'max_features' like the name says is the maximum number of features the algorithm should try. The 'min_sample_leaf' is the minimum number of leaves used to split an internal node. To define how many processors the algorithm is allowed to use we have the 'n_jobs' parameter, where for an unlimited number we type '-1'. We still have two parameters, the 'random_state' which makes the output replicable and the 'oob_score' which is a cross-validation method for this algorithm. Although there are no that many hyperparameters, the default values of them often do a pretty good job in the classifier. A significant disadvantaged of this algorithm is that a large number of trees in the training and prediction result in a slower process.[49]

The Random Forest Regressor uses the same algorithm as in the Random Forest Classifier but instead of predicting the new data into classes, creates a regression model to predict the new data.

3.2.4.3 K-Nearest Neighbors

The K-Nearest Neighbors algorithm has a particularity comparing to other classifiers; it does not make any assumptions on the underlying data distribution (non-parametric). It also doesn't do any generalization with the training dataset, making this first phase pretty fast, but slowing down the second one and taking much more memory to store all these training data that will be used in the testing phase. This algorithm uses the feature similarity to predict new data. The object is assigned to the most common class among its k (k is an integer) nearest neighbours.[51] This k number is defined in the parameter 'n_neighbors', is important to avoid k=1 because it frequently leads to overfitting of the model.[52] Some more parameters help to get better models like the distance metric used for the tree or the algorithm used to compute the nearest neighbours.[53]

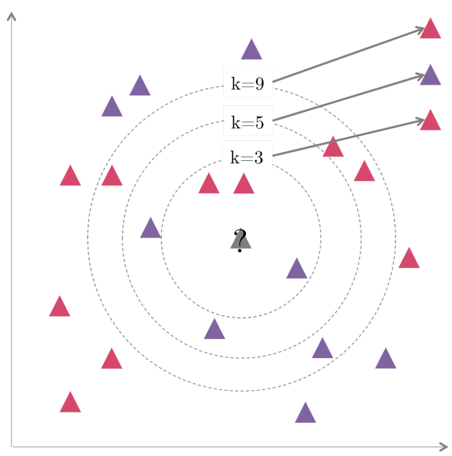


Figure 3.6: Example of KNN classification.

3.2.5 New Prediction Approach using a CNN

3.2.5.1 Neural Networks

A Neural Network is used to learn patterns and relationship in data, which do not require explicit coding of the problems.[73]

Artificial Neural Network are inspired by our brain's architecture. Despite that, both do not work the same way. ANNs have gradually become different from their biological relatives.[74]

ANNs result from academic investigations involving mathematical formulations to model nervous system operations.[72] These networks are the core of deep learning, and ideal to use in large and complex Machine Learning problems because they are very versatile, robust and scalable. There are a lot of different tasks where these are used, for example, speech recognition services, a significant number of images classification, recommendations for users on many platforms.[74]

ANNs were introduced in 1943 by the neurophysiologist Warren McCulloch and the mathematician Walter Pitts, where they presented an uncomplicated computational model of how biological neurons might act together in animal brains to perform complex computation using propositional logic.[74]

Starting with understanding how neurons function. A neuron is a cell found in animal cerebral cortexes, also known as the brain. It is composed of a cell body, which contains the nucleus and most complex components of the cell. Attached there are many branching extensions called dendrites and a longer one called the axon. The end of the axon splits into more branches called telodendria with synapses on the tips. These ends connect with other neurons and receive impulses (signals) from their synapses.[74]

Although this process may seem simple, neurons are organized in a vast network of billions of neurons. However, they can easily be compared to an ANN since they are an information processing system. The elements called neuron process the information and the connection links transmit the signal between them. These links have associated weights, which are multiplied along with the incoming signal (called network input). The output signal is achieved by applying activation to the network input. The neurons are grouped into layers. There is the input layer, which receives the information for the training of the networks; the output layer that communicates the output of the systems; and the in-between layers, called hidden layers.[73]

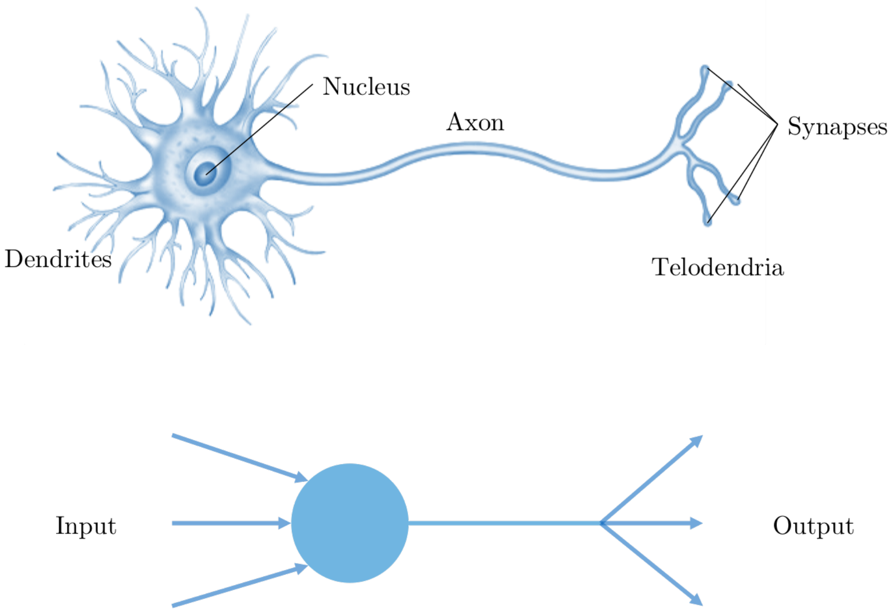


Figure 3.7: Comparison between a biological neuron and a artificial neuron.

3.2.5.2 Convolutional Neural Networks

A CNN takes arrays of pixel values as input to the network. The hidden layer consists of several distinct layers which carry out feature extraction. There is a fully connected layer that identifies the objects in the picture. Convolution operation forms the core of every convolution neural network. There are four layers in a CNN: Convolution layer, ReLU layer, Pooling layer, and Fully Connected Layer.[75, 76]

The Convolution layer applies a filter matrix over the array of pixels and executes a convolution operation to get a convolved feature map.[75, 76]

Next, there is a ReLU layer which introduces non-linearity to the network. It changes all negative pixels to zero and performs an element-wise operation. The original input image is scanned in multiple Convolution and ReLU layers for discovering hidden features and patterns in the image. The output is a Rectified Feature Map.[75, 76]

Then there is a Pooling layer that reduces the dimensionality of this feature map and outputs a Pooled feature map. Pooling layers use several filters to identify different parts of the image (edges, corners, body).[75, 76]

Lastly, we have Flattening, which is when this pooled feature map is converted into a long continuous linear vector. This flattened matrix goes through a Fully Connected Layer to classify the images.[75, 76]

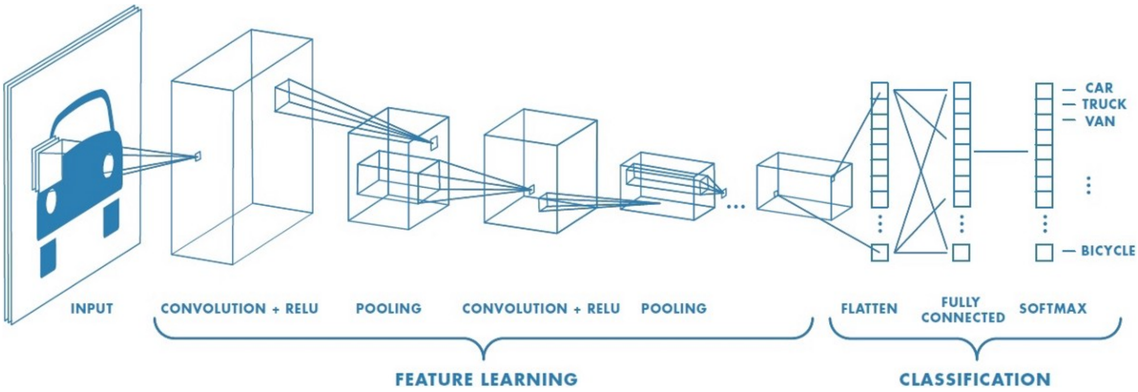


Figure 3.8: Example of a Convulotional Neural Network.[76]

3.2.5.3 Transfer Learning

Transfer Learning is used in Deep Learning when we have a small amount of data to train. Here we reuse a pre-trained model for our new problem. This is useful when we are dealing with real-world problems that generally don't have a significant amount of labeled data, and deep learning models are very complex to train.[78]

Summing up, in Transfer Learning, the knowledge of a previously trained Machine Learning model is applied to a different, nevertheless related problem. It tries to exploit what has been learned in one model to improve generalization in others, the weights that the network are transferred from a model to another.[78]

Transfer Learning is very commonly used in Computer Vision and Natural Language Processing Tasks like Sentiment Analysis, because of the complexity that is to train models for this.[78]

When using transfer learning, early and middle layers are used, and the last layers are re-train. For example, if we want to use the VGG16, that was trained to distinguish between 1000 different objects, to distinguish between dogs and cats. In the earlier layers, the model learned to recognize different objects, so now in the last layers, it can be re-trained to learn to distinguish between dogs and cats.

The main benefits of using this technique are that a lot of training time is saved, there is no need of a numerous amount of data, and in most cases, the neural network used performs better.[78]

3.2.5.4 VGG16

VGG is a convolutional neural network model introduced by K. Simonyan and A. Zisserman and achieves a top-5 test accuracy of 92.7% in ImageNet.[80]

The VGG16 model was trained to classify images of ImageNet competition. These images were divided into 1000 categories. The input shape for this model were images 224x224x3 [77] and this model has about 138 million parameters. This model is a really deep network, but the architecture can be easily understood. [79] Looking at the figure 3.9, we can observe the different layers of this model. For classifying dogs and cats, the last layer (Softmax) is removed and replaced by another with an output of 2 instead of 1000.

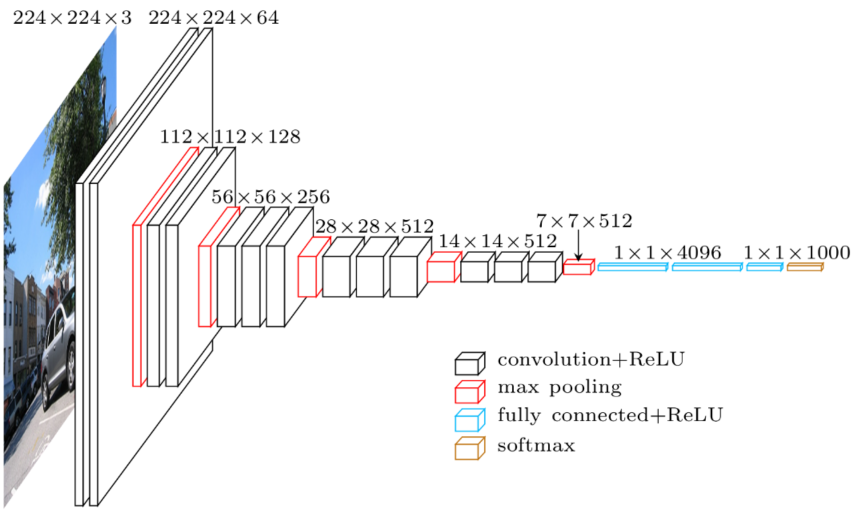


Figure 3.9: Layers of the VGG16 model.[80]

For our neural network problem, we are doing transfer learning using the trained CNN VGG16, where we are using our images as input. For our problem, we are removing the last layer and replace it with another layer that produces only two outputs so we could predict if the user was depressive or not. Although the features in the last layers are tailored to identify objects, since this network was trained to recognize different objects, and the mood of a picture might require more low-level features, we could only achieve this by choosing another pre-trained network, which is hard to find since training these networks is complex, or by cutting off more of the network and replacing it, but that would require a lot more data samples on our side.

3.2.6 Evaluation of the Models

3.2.6.1 Binary Problem

For evaluating our binary problem models, we are calculating the confusion matrix. A confusion matrix is a method for summarizing the performance of a classification algorithm. It gives us the number of true positives (number of well classified elements of the class 1), true negatives (number of well classified elements of the class 0), false positives (number of misclassified elements of the class 0) and false negatives (number of misclassified elements of the class 1).[69]

Predicted / Actual	1	0
1	TP	FP
0	FN	TN

Figure 3.10: Example of a confusion matrix.

With these numbers we can calculate four standard terms: accuracy, precision, specificity and sensitivity. The accuracy(3.1) is the ratio of correct predictions from both classes to total predictions made. Here we analyze the ability to evaluate the depressive user and non-depressive users correctly. The precision(3.2) is the ratio of correct predictions from the positive class to total of cases classified as positive. Here we analyze how many depressive users the algorithm correctly classifies. The specificity(3.3) is the ratio of correct predictions from the negative class to total of negative cases, both well or wrong classified. Here we analyze the ability of the method of classifying correctly the non-depressive users. The sensitivity(3.4) is the ratio of correct predictions from the positive class to total of positive cases, both well or wrong classified. Here we analyze the ability of the method of classifying correctly the depressive users.[70]

$$\text{accuracy} = \frac{TP + FP}{TP + TN + FP + FN} \tag{3.1}$$

$$\text{precision} = \frac{TP}{TP + FP} \tag{3.2}$$

$$\text{specificity} = \frac{TN}{TN + FP} \tag{3.3}$$

$$\text{sensitivity} = \frac{TP}{TP + FN} \tag{3.4}$$

3.2.6.2 Multiclass Problem

For the multiclass problem we only calculate the accuracy value, the ability of the algorithm to differentiate the classes, minimal depression, mild depression, moderate depression and severe depression, correctly. We also have in mind that not analyzing anything else may lead us to algorithms misclassifying an entire class and still present good accuracy values, so, we plot the real labels and the predicted ones to be sure that no class was entirely misclassified.

3.2.6.3 Regression Problem

In the regression problem the evaluation of the models is made calculating the mean squared error. Mean squared error (MSE) is an important criterion to measure the performance of an estimator, like a regression algorithm. For this measure, it is required a target of prediction (predicted data) and a predictor (given data). MSE is the average of squares of the "errors", the difference between these two attributes. It incorporates both the variance and bias of the estimator.[71]

3.2.6.4 Convolutional Neural Network Approach

For evaluating our last approach using a convolutional neural network, at the end of each epoch, we get the accuracy and loss from the network and from the validation set, to see how the classification evolves in the network. In the end, the CNN tries to predict a new group of data (test dataset), and we output the accuracy and loss.

The loss (3.5) is calculated for training and validation (and lastly for the testing dataset) using the categorical cross entropy, and its interpretation is how well the model is doing for these two sets. Unlike accuracy, the loss is not a percentage; it is a summation of the errors made for each example in training or validation sets.

$$H(p,q) = -\sum_i p_i \log q_i \tag{3.5}$$

4

Results and Discussion

4.1 Binary Problem: Non-depressive vs. Depressive

After having the final dataset, we altered the labels from scores (0-63) to the binary problem (0 or 1), so, all labels between 0 and 13 were changed into 0 and the rest of them into 1. We then wanted to analyze what the features could achieve in our classifiers without doing any feature reduction or feature extraction. Since we had just a few features (six text features, six usage metrics' features and eight image features), we tried to predict these two groups using all of them to train the classifiers so that we could have a global idea of their behavior and if it was a better option to analyze them one by one. In this case, the model parameters were tested manually. But further on in the project, to obtain the best results, it was used a function RandomizedSearchCV[42] for the Random Forest Classifier and GridSearchCV[43] for the SVM and K-Nearest Neighbors Classifiers, both from the Machine Learning package sklearn[44], that tests a lot of options in the parameters and chooses the ones that give better results to each model.

GridSearchCV performs a grid-search with cross-validation: splitting firstly multiple train and test sets, using a strategy defined by the parameter cv. Then, GridSearchCV will loop over each parameter configuration, fitting the model on one train set and evaluating it on the corresponding test set. In the end, given the best results considering the scoring parameter we choose, the best parameters will be returned. RandomizedSearchCV is similar to GridSearchCV, but not all parameter values are tried out, the number is given by n_iter.

After this first step, we could conclude that some models were not that good, and this could be for three main reasons. The features we have are not enough for the classifiers to predict depression; we have too few samples, and since each sample can be quite variable we need much more samples to have a proper dataset; or we are not using the features in the right way and should analyze them first. The only thing we could do at this point was to analyze the features and understand them which would be relevant for distinguish between the two classes.

With that said, we run a Student's T-Test between the two groups (non-depressive and depressive) in each feature. This test returns a t-statistic value and a p-value and observing the p-value if this is larger than 0.05 or 0.1, then we cannot reject the null hypothesis of identical average scores. Having identical average scores means the information in each of these groups is similar, so it will not be a good option for the classifier to divide into groups. We set the threshold of the p-value as 0.1 in most of the cases, unless the p-values were all really low, and in this case, the threshold was 0.05, so that we had the very best features. Analyzing the p-values, we kept the features with the value under our threshold, in some cases we only kept two features, in other we could keep some more.

So after all the feature extraction, it was time to normalize the data, so that the average score of the data was approximately zero and the standard deviation unitary. After this, the dataset was divided into a train and test datasets.

The parameters of the three classifiers used were analyzed in order to get the best performance but also to avoid overfitting the data. Considering that this dataset in the binary problem is not terrible unbalanced, the selection of the parameters is based on the prediction accuracy. Otherwise, it would not be the best option because the accuracy could be very high even if it misclassified one entire class.

Table 4.1: Parameters for the Random Forest Classifier that were tested to find the optimal one.

Parameters	Values
n_estimators	[10, 30, 50, 70, 100, 200, 400, 600, 800, 1000, 1200, 1400, 1600, 1800, 2000]
max_features	['auto', 'sqrt']
max_depth	[2, 4, 8, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 110, None]
min_samples_split	[2, 5, 10]
min_samples_leaf	[1, 2, 4]
bootstrap	[True, False]

Table 4.2: Parameters for the Support Vector Machine Classifier that were tested to find the optimal one.

Parameters	Values
kernel	['linear', 'rbf', 'poly']
C	[0.001, 0.01, 0.1, 1, 10]
gamma	[1e-7, 1e-4, 0.001, 0.01, 0.1, 1]

Table 4.3: Parameters for the K-Nearest Neighbors Classifier that were tested to find the optimal one.

Parameters	Values
n_neighbors	[1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29]
metric	['euclidean', 'cityblock']

After choosing the best parameters for each classifier with each group of features, we started training the classifiers dividing the training dataset into train and validation datasets.

Then, we trained each classifier 100 times, each time with a different training and validation dataset parted randomly, to observe the tendency of the results we could obtain and, after this, we chose the best models to afterward test them on our test dataset, which also undergo the same feature reduction and feature extraction.

When all models were tested, we chose the best model from each classifier and compared with each other to find the best overall model.

4.1.1 Text Features

Like explained previously, first we used the raw data directly to train our classifiers and analyze the results from these models and see if it would be a better option to analyze the features one by one.

Table 4.4: Parameters for the Random Forest, SVM and KNN Classifiers that were tested to find the optimal one to fit all text features in the binary problem.

Parameters	Values
n_estimators	800
max_features	'auto'
max_depth	50
min_samples_split	2
min_samples_leaf	2
bootstrap	False

Parameters	Values
kernel	'poly'
C	10
gamma	1

Parameters	Values
n_neighbors	9
metric	'euclidean'

Table 4.5: Best results on cross-validation for all text features in the binary problem with Random Forest Classifier using random train and validation datasets from the initial train dataset to evaluate the models.

	Accuracy	Specificity	Sensitivity	Precision
1	83.3%	87.5%	80.0%	88.9%
2	83.3%	83.3%	83.3%	71,4%
3	77.8%	100.0%	55.6%	100,0%
4	77.8%	80.0%	75.0%	75.0%

Table 4.6: Test results for the best models chosen from the cross-validation for all text features in the binary problem with Random Forest Classifier, using the test dataset to evaluate the models.

	Accuracy	Specificity	Sensitivity	Precision
1	63.6%	66.7%	61.5%	72.7%
2	59.1%	66.7%	53.8%	70.0%
3	59.1%	66.7%	53.8%	70.0%
4	54.5%	66.7%	46.2%	66.7%

Table 4.7: Best results on cross-validation for all text features in the binary problem with Support Vector Machine Classifier using random train and validation datasets from the initial train dataset to evaluate the models.

	Accuracy	Specificity	Sensitivity	Precision
1	83.3%	81.8%	85.7%	75.0%
2	72.2%	88.9%	55.6%	83.3%
3	72.2%	80.0%	62.5%	71.4%
4	66.7%	63.6%	71.4%	55.6%

Table 4.8: Test results for the best models chosen from the cross-validation for all text features in the binary problem with Support Vector Machine Classifier, using the test dataset to evaluate the models.

	Accuracy	Specificity	Sensitivity	Precision
1	59.1%	66.7%	53.8%	70.0%
2	50.0%	66.7%	38.5%	62.5%
3	54.5%	66.7%	46.2%	66.7%
4	59.1%	44.4%	69.2%	64.3%

Table 4.9: Best results on cross-validation for all text features in the binary problem with K-Nearest Neighbors using random train and validation datasets from the initial train dataset to evaluate the models.

	Accuracy	Specificity	Sensitivity	Precision
1	88.9%	90.9%	85.7%	85.7%
2	83.3%	90.0%	75.0%	85.7%
3	83.3%	75.0%	90.0%	81.8%
4	77.8%	90.0%	62.5%	83.3%

Table 4.10: Test results for the best models chosen from the cross-validation for all text features in the binary problem with K-Nearest Neighbors, using the test dataset to evaluate the models.

	Accuracy	Specificity	Sensitivity	Precision
1	40.9%	55.6%	30.8%	50.0%
2	63.6%	77.8%	53.8%	77.8%
3	54.5%	44.4%	61.5%	61.5%
4	54.5%	77.7%	38.5%	71.4%

Although the training models have interesting results, the test results from these models have much lower values for all the four standard terms, in all the classifiers.

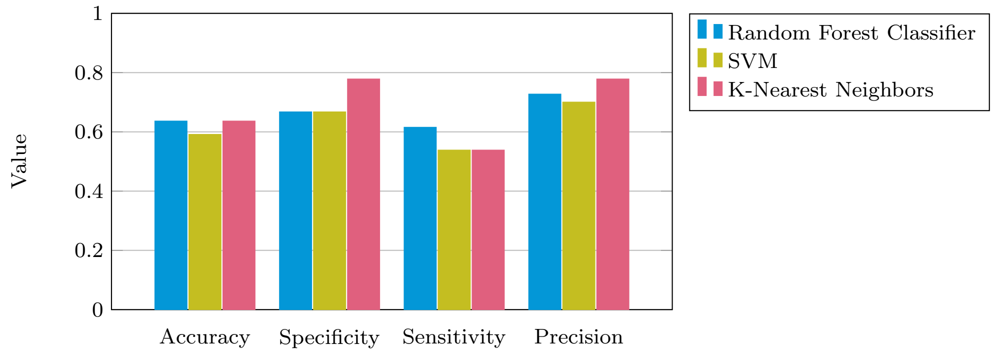


Figure 4.1: Best models from each classifier for all text features in the binary problem.

The previous tables show us the best results from the three classifiers used for text features. As we can observe, we did not have any good models for this data. Our next step will be to analyze each feature and understand which one has information that can be used to distinguish between the two groups and which ones we can neglect. With this feature analysis, we aim to improve the models.

Table 4.11: Values of the p-value from the Student's T-test on the text features in the binary problem.

	p-value
Number of characters	0.5601
Number of words	0.0799
Existence of emojis	0.3420
Number of emojis	0.2911
Number of tags	0.0778
Existence of tags	0.2835

Observing the p-values, we can conclude that some of these features are not useful for distinguishing these two groups, but the ones with the most different average scores are the number of words and number of tags. Our new dataset has now these two features. Our next step is to understand what parameters to use with this new dataset.

Table 4.12: Parameters for the Random Forest, SVM and KNN Classifiers that were tested to find the optimal one to fit the best text features in the binary problem.

Parameters	Values
n_estimators	1000
max_features	'auto'
max_depth	50
min_samples_split	2
min_samples_leaf	1
bootstrap	False

Parameters	Values
kernel	'rbf'
C	10
gamma	1

Parameters	Values
n_neighbors	21
metric	'cityblock'

After the parameters have been chosen, a cross-validation train was done, training 100 times each of the three classifiers with a train and a validation dataset split each time randomly out of the initial train dataset so the best models could be used with the test dataset. For analyzing these results, we plotted the four standard terms (accuracy, precision, specificity, and sensitivity) of each training.

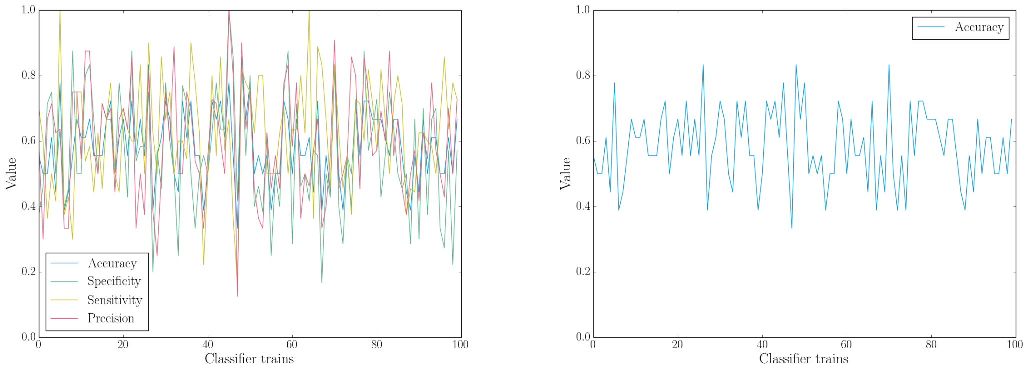


Figure 4.2: Cross-validation results of 100 trained Random Forest Classifiers with the train dataset of the best text features in the binary problem. On the left, the four standard terms and on the right, the accuracy.

Analyzing this picture, we can observe that the accuracy of these Random Forest models is between 45% and 85%. In tables 4.13 and 4.14, some of the best models were tested with our test dataset. Comparing the train and test results, we conclude that although some of the train models could be interesting, when testing with new data, they present poor overall results.

Table 4.13: Best results on cross-validation for the best text features in the binary problem with Random Forest Classifier, using random train and validation datasets from the initial train dataset to evaluate the models.

	Accuracy	Specificity	Sensitivity	Precision
1	83.3%	88.9%	77.8%	87.5%
2	77.8%	66.7%	83.3%	83.3%
3	77.8%	60.0%	84.6%	84.6%
4	72.2%	100.0%	50.0%	100.0%

Table 4.14: Test results for the best models chosen from the cross-validation for the best text features in the binary problem with the Random Forest Classifier, using the test dataset to evaluate the models.

	Accuracy	Specificity	Sensitivity	Precision
1	45.5%	36.4%	54.5%	46.2%
2	50.0%	36.4%	63.6%	50.0%
3	36.4%	36.4%	36.4%	36.4%
4	59.1%	54.5%	63.6%	58.3%

In this classifier, we could see that, although we did not have the best results, the models failed in both groups and not only one, so a possible reason for these low results is because we have very few samples to train the classifiers and our dataset has a considerable variation.

Next, we will use the Support Vector Machine Classifier for classifying our data, aiming better results than with the Random Forest Classifier. SVMs are helpful in text and hypertext categorization and have been proved to be a powerful and promising data classification and function estimation tool. We are first interpreting the results of 100 trained classifier, to understand what we could expect with these models. After this, we chose the best models to, posteriorly, test them with our test dataset.

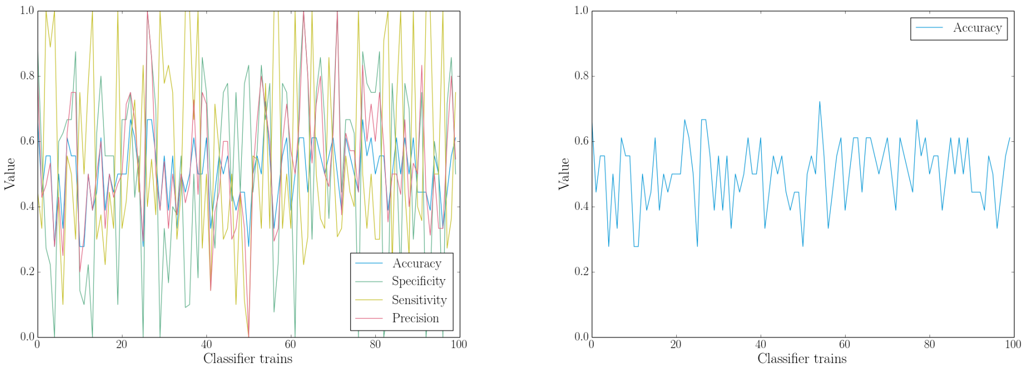


Figure 4.3: Cross-validation results of 100 trained Support Vector Machine Classifiers with the train dataset of the best text features in the binary problem. On the left, the four standard terms and on the right, the accuracy.

The first thing we notice is that the accuracy is much lower than in the previous classifier, rounding the 50% and hardly ever getting to the 80%. Despite this, we were able to find some reasonable results that we will test. We can also notice that the sensitivity is frequently 100%, which means that the positive class is always correctly classified. The reason for this to happen is that the model is overfitting the data. Overfitting implies that the model is too dependent on that data and it is likely to have a higher error rate on new unseen data, which we are going to conclude next.

Table 4.15: Best results on cross-validation for the best text features in the binary problem with the SVM Classifier, using random train and validation datasets from the initial train dataset to evaluate the models.

	Accuracy	Specificity	Sensitivity	Precision
1	83.3%	88.9%	77.8%	87.5%
2	83.3%	66.7%	100.0%	75.0%
3	83.3%	75.0%	90.0%	81.8%
4	77.8%	87.5%	70.0%	87.5%

Table 4.16: Test results for the best models chosen from the cross-validation for the best text features in the binary problem with the SVM Classifier, using the test dataset to evaluate the models.

	Accuracy	Specificity	Sensitivity	Precision
1	59.1%	63.6%	54.5%	60.0%
2	59.1%	63.6%	54.5%	60.0%
3	63.6%	72.7%	54.5%	66.7%
4	63.6%	72.7%	54.5%	66.7%

In the SVM classifiers, we observe the same as with the Random Forest Classifier, a significant decreasing of the results when testing the models. Besides this, we have better results with this classifier. Next, we will classify our data using the K-Nearest Neighbors Classifier. As previously, we started by understanding the behavior of the models, the results they could achieve. After this, we picked the best ones, and use them on our test dataset.

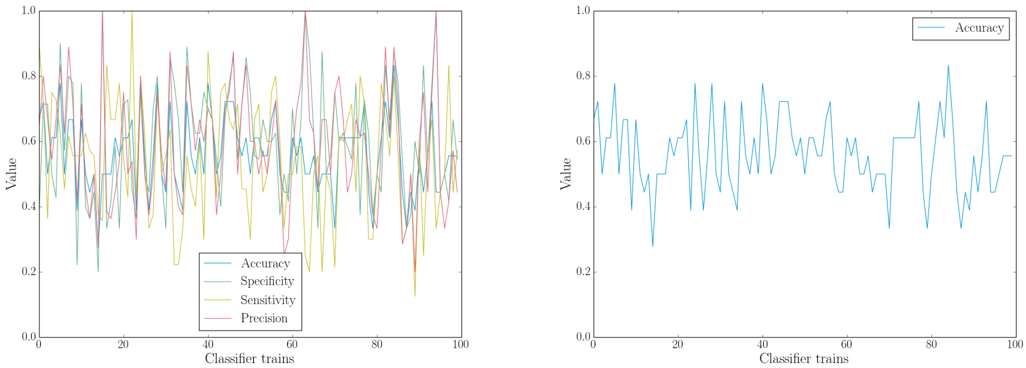


Figure 4.4: Cross-validation results of 100 trained K-Nearest Neighbors Classifiers with the train dataset of the best text features in the binary problem. On the left, the four standard terms and on the right, the accuracy.

Table 4.17: Best results on cross-validation for the best text features in the binary problem with K-Nearest Neighbors Classifier, using random train and validation datasets from the initial train dataset to evaluate the models.

	Accuracy	Specificity	Sensitivity	Precision
1	88.9%	88.9%	77.8%	87.5%
2	83.3%	66.7%	100.0%	75.0%
3	83.3%	75.0%	90.0%	81.8%
4	77.8%	87.5%	70.0%	87.5%

Table 4.18: Test results for the best models chosen from the cross-validation for the best text features in the binary problem with the K-Nearest Neighbors Classifier, using the test dataset to evaluate the models.

	Accuracy	Specificity	Sensitivity	Precision
1	68.2%	63.6%	54.5%	60.0%
2	72.7%	63.6%	54.5%	60.0%
3	63.6%	72.7%	54.5%	66.7%
4	54.5%	72.7%	54.5%	66.7%

In these models, the results from the test dataset also decreased significantly regarding the training, although it presented better results compared with the other two classifiers. The next step was to compare the best model of each classifier.

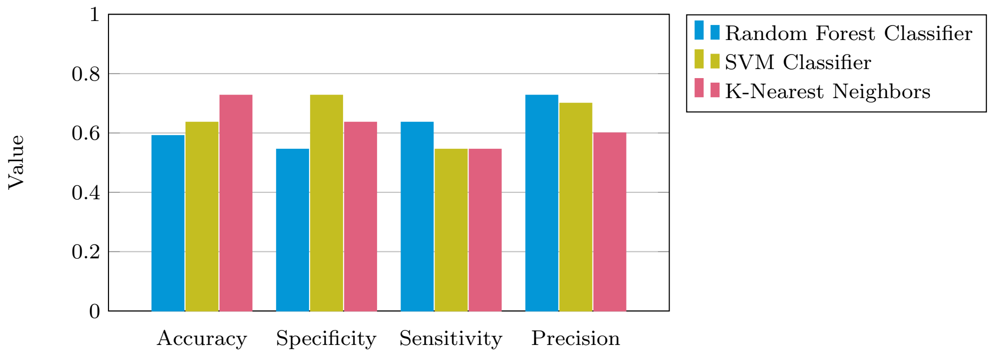


Figure 4.5: Best models from each classifier for the best text features in the binary problem .

In the overall prediction, we conclude that the KNN model was the best one, but had worse results when classifying the positive class (depressive users). This can be observed by analyzing the sensitivity of the model, which gives us the ratio between the well classified and all the positive class. This model also has lower precision than the others, meaning that from all the classified as being part of the positive class, it classified a lower percentage than the other two models (SVM and Random Forest) correctly.

It is important to notice that there weren't used many features for this classification, only the number of words and tags by the users showed relevance in distinguishing these two classes. It will be interesting to compare this result with the multiclass problem.

4.1.2 Usage Metrics' Features

Likewise the previous data analysis, we started using the raw data directly in the classifiers.

Table 4.19: Parameters for the Random Forest, SVM and KNN Classifiers that were tested to find the optimal one to fit all usage metrics' features in the binary problem.

Parameters	Values
n_estimators	30
max_features	'auto'
max_depth	10
min_samples_split	5
min_samples_leaf	2
bootstrap	False

Parameters	Values
kernel	'rbf'
C	1
gamma	1

Parameters	Values
n_neighbors	5
metric	'euclidean'

Table 4.20: Best results on cross-validation for all usage metrics' features in the binary problem with Random Forest Classifier, using random train and validation datasets from the initial train dataset to evaluate the models.

	Accuracy	Specificity	Sensitivity	Precision
1	94.4%	90.0%	100.0%	88.9%
2	94.4%	87.5%	100.0%	90.9%
3	88.9%	85.7%	90.9%	90.9%
4	77.8%	77.8%	77.8%	77.8%

Table 4.21: Test results for the best models chosen from the cross-validation for the all usage metrics' features in the binary problem with the Random Forest Classifier, using the test dataset to evaluate the models.

	Accuracy	Specificity	Sensitivity	Precision
1	72.7%	88.9%	61.5%	88.9%
2	72.7%	88.9%	61.5%	88.9%
3	72.7%	88.9%	61.5%	88.9%
4	81.8%	88.9%	76.9%	90.9%

Table 4.22: Best results on cross-validation for all usage metrics' features in the binary problem with SVM, using random train and validation datasets from the initial train dataset to evaluate the models.

	Accuracy	Specificity	Sensitivity	Precision
1	83.3%	80.0%	87.5%	77.8%
2	83.3%	75.0%	90.0%	81.8%
3	77.8%	87.5%	70.0%	87.5%
4	72.2%	80.0%	62.5%	71.4%

Table 4.23: Test results for the best models chosen from the cross-validation for the all usage metrics' features in the binary problem with the SVM, using the test dataset to evaluate the models.

	Accuracy	Specificity	Sensitivity	Precision
1	68.2%	88.9%	53.8%	87.5%
2	63.6%	88.9%	46.2%	85.7%
3	63.6%	77.8%	53.8%	77.8%
4	68.2%	88.9%	53.8%	87.5%

Table 4.24: Best results on cross-validation for all usage metrics' features in the binary problem with K-Nearest Neighbors, using random train and validation datasets from the initial train dataset to evaluate the models.

	Accuracy	Specificity	Sensitivity	Precision
1	83.3%	90.0%	75.0%	87.5%
2	77.8%	90.0%	62.5%	83.3%
3	77.8%	83.3%	66.7%	66.7%
4	77.8%	81.8%	71.4%	71.4%

Table 4.25: Test results for the best models chosen from the cross-validation for the all usage metrics' features in the binary problem with the KNN Classifier, using the test dataset to evaluate the models.

	Accuracy	Specificity	Sensitivity	Precision
1	54.5%	88.9%	30.8%	80.0%
2	54.5%	88.9%	30.8%	80.0%
3	59.1%	88.9%	38.5%	72.7%
4	54.5%	83.3%	30.8%	80.0%

Looking at these results, we can conclude that the results are much better than with the text features, although they are not satisfactory. For example, with the Random Forest Classifier, we have a model that when tested presents values of 80-90%, which makes it a good model regarding the amount of data we have for training, even though it was not one of the best models while training. This is why is important to choose more than one good model for testing afterward so that we do not end up with models that overfit the data and have poor results when tested with unseen data. The SVM and KNN Classifiers have both weak results when tested, having significant difficulty in classifying the positive class correctly (analyzing the sensitivity) and in the overall classification, resulting in low accuracy.

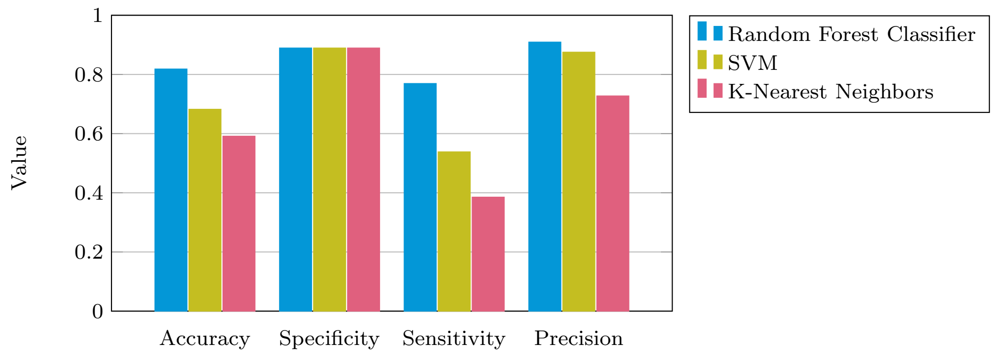


Figure 4.6: Best models from each classifier for all usage metrics' features in the binary problem.

Despite these results, we will also be going to analyze the features and try to improve the results. So for this group of features, we also run a Student's T-Test between the two groups (non-depressive and depressive) in each feature.

Table 4.26: Results from the Student's T-test on the usage metrics' features in the binary problem.

	p-value
Day/Night time	0.3286
Number of likes	0.0011
Number of comments	0.1967
Number of different people comments	0.0930
Days since last post	0.0206
Average post frequency	0.0026

Observing the p-values, we can conclude that most of these features are actually suitable for distinguishing these two groups, but we are only keeping the features with a p-value smaller than 0.1. We are excluding day/night time and the number of overall comments. With our new dataset, we are now choosing the optimal parameters so that we can find suitable models for our problem. After this, we are training 100 classifiers with different train and validation datasets, out of our train dataset, and analyze the results to choose the best models.

Table 4.27: Parameters for the Random Forest, SVM and KNN Classifiers that were tested to find the optimal one to fit the best usage metrics' features in the binary problem.

Parameters	Values
n_estimators	30
max_features	'auto'
max_depth	8
min_samples_split	2
min_samples_leaf	2
bootstrap	True

Parameters	Values
kernel	'rbf'
C	1
gamma	0.1

Parameters	Values
n_neighbors	3
metric	'euclidean'

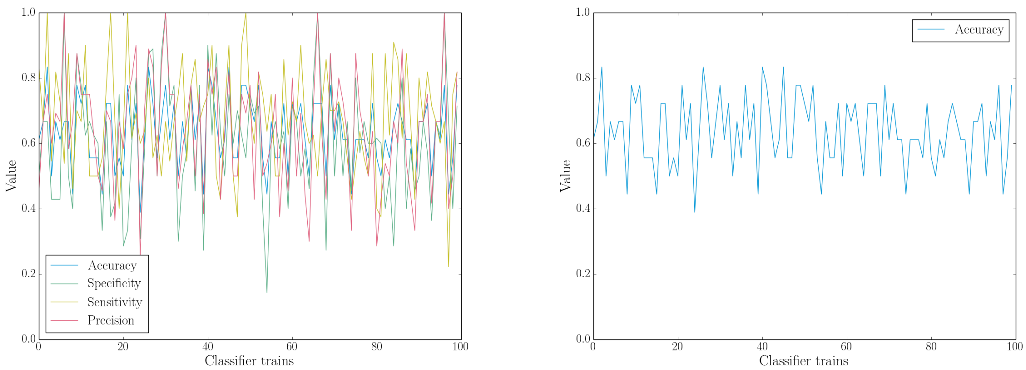


Figure 4.7: Cross-validation results of 100 trained Random Forest Classifiers with the train dataset of the best usage metrics' features in the binary problem. On the left, the four standard terms and on the right, the accuracy.

Taking a look at these two graphs in figure 4.7 we can conclude that there are good models. We will now choose the best ones to test them with new data.

Table 4.28: Best results on cross-validation for the best usage metrics' features in the binary problem with Random Forest Classifier, using random train and validation datasets from the initial train dataset to evaluate the models.

	Accuracy	Specificity	Sensitivity	Precision
1	83.3%	72.7%	100.0%	70.0%
2	77.8%	81.8%	71.4%	71.4%
3	77.8%	66.7%	83.3%	83.3%
4	72.2%	66.7%	75.0%	81.8%

Table 4.29: Test results for the best models chosen from the cross-validation for the best usage metrics' features in the binary problem with the Random Forest Classifier, using the test dataset to evaluate the models.

	Accuracy	Specificity	Sensitivity	Precision
1	86.4%	84.6%	88.9%	80.0%
2	81.8%	84.6%	77.8%	77.8%
3	77.3%	84.6%	66.7%	75.0%
4	72.7%	84.6%	55.6%	71.4%

With the Random Forest Classifier, the models are doing good on predicting the unseen data. They obtained even better results with this test dataset. When comparing with the results we had when using all features, and without optimizing the parameters, we can conclude that it is important not to skip these steps because we may be introducing much noise from the features that are not helping with the separation of the groups, make it difficult to have a good model. In the table 4.27 we can see the optimal parameters.

We will now train models with the Support Vector Machine Classifier.

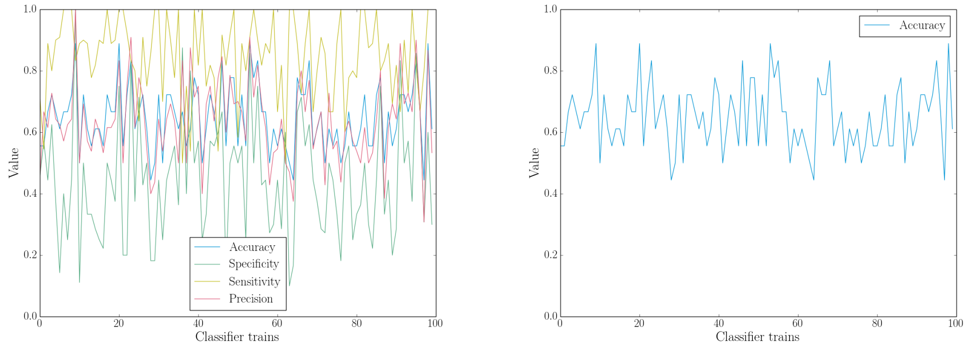


Figure 4.8: Cross-validation results of 100 trained SVM Classifiers with the train dataset of the best usage metrics' features in the binary problem. On the left, the four standard terms and on the right, the accuracy.

Table 4.30: Best results on cross-validation for the best usage metrics' features in the binary problem with SVM Classifier, using random train and validation datasets from the initial train dataset to evaluate the models.

	Accuracy	Specificity	Sensitivity	Precision
1	100.0%	100.0%	100.0%	100.0%
2	88.9%	100.0%	84.6%	100.0%
3	83.3%	60.0%	92.3%	85.7%
4	72.2%	75.0%	70.0%	77.8%

Table 4.31: Test results for the best models chosen from the cross-validation for the best usage metrics' features in the binary problem with the SVM Classifier, using the test dataset to evaluate the models.

	Accuracy	Specificity	Sensitivity	Precision
1	59.1%	38.5%	88.9%	50.0%
2	81.8%	84.6%	77.8%	77.8%
3	77.3%	69.3%	88.9%	66.7%
4	77.3%	69.2%	88.9%	66.7%

Although the Random Forest results were better, we also trained some good models with SVM. For example, the second model has results between 77% and 85% when tested. However, we also got a model that completely overfitted the data, the first model gave us 100% when trained in all of the four evaluation metrics but when tested with unseen data had low results.

Ultimately, we are going to classify the data with K-Nearest Neighbors models with the parameters showed in table 4.27.

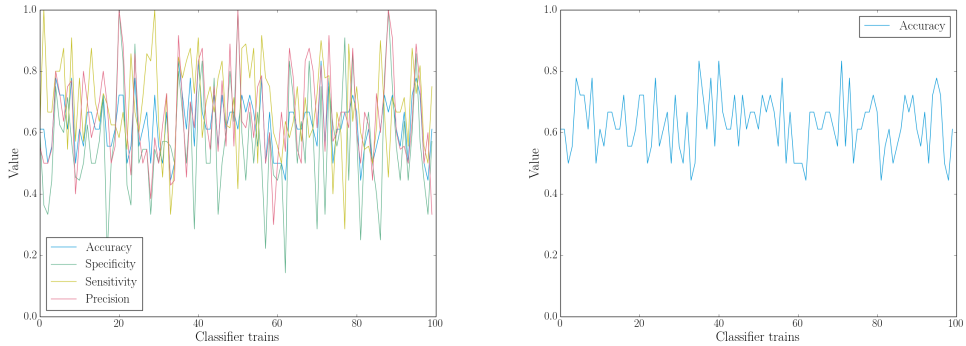


Figure 4.9: Cross-validation results of 100 trained K-Nearest Neighbors Classifiers with the train dataset of the best usage metrics' features in the binary problem. On the left, the four standard terms and on the right, the accuracy.

Table 4.32: Best results on cross-validation for the best usage metrics' features in the binary problem with the KNN Classifier, using random train and validation datasets from the initial train dataset to evaluate the models.

	Accuracy	Specificity	Sensitivity	Precision
1	88.9%	100.0%	83.3%	100.0%
2	83.3%	77.8%	88.9%	80.0%
3	77.8%	85.7%	72.7%	88.9%
4	77.8%	50.0%	100.0%	71.4%

Table 4.33: Test results for the best models chosen from the cross-validation for the best usage metrics' features in the binary problem with the KNN Classifier, using the test dataset to evaluate the models.

	Accuracy	Specificity	Sensitivity	Precision
1	77.3%	84.6%	66.7%	75.0%
2	72.7%	76.9%	66.7%	66.7%
3	81.8%	92.3%	66.7%	85.7%
4	77.3%	84.6%	66.7%	75.0%

For the KNN algorithm, we also got some reasonable models, although all of them when tested had more difficulty classifying the positive class (depressive users), concluded by analyzing the sensitivity of the model.

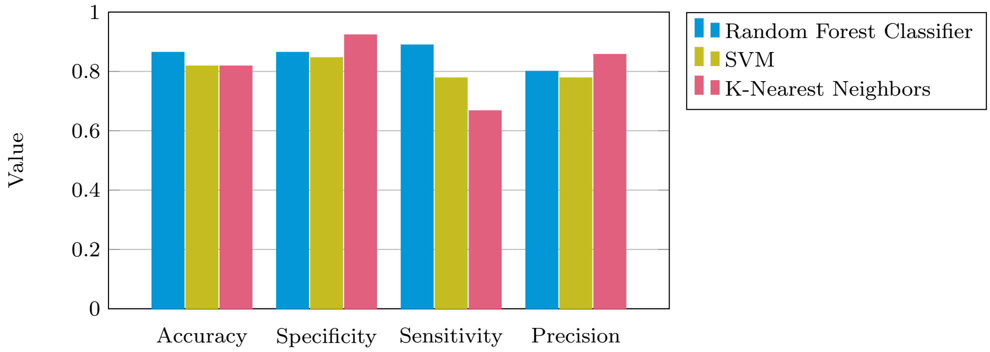


Figure 4.10: Best models from each classifier using the best usage metrics' features in the binary problem.

With this new dataset, we had much better results than with all the features. All classifiers have a model with accuracy over 80%, which means that of our test dataset that has 22 samples, at least 18 are correctly predicted.

Comparing to the text features, we got better models with these features with all of the three classifiers. We will next try to classify the depressive users using only image features.

4.1.3 Image Features

Like in the previous classifications, we will start by training the classifiers with the raw data.

Table 4.34: Parameters for the Random Forest, SVM and KNN Classifiers that were tested to find the optimal one to fit all text features in the binary problem.

Parameters	Values
n_estimators	10
max_features	'auto'
max_depth	90
min_samples_split	2
min_samples_leaf	1
bootstrap	True

Parameters	Values
kernel	'rbf'
C	1
gamma	1

Parameters	Values
n_neighbors	5
metric	'euclidean'

Table 4.35: Best results on cross-validation for all image features in the binary problem with Random Forest Classifier, using random train and validation datasets from the initial train dataset to evaluate the models.

	Accuracy	Specificity	Sensitivity	Precision
1	76.0%	91.7%	61.5%	88.9%
2	72.0%	84.6%	58.3%	77.8%
3	68.0%	92.3%	41.7%	83.3%
4	64.0%	66.7%	61.5%	66.7%

Table 4.36: Test results for the best models chosen from the cross-validation for all image features in the binary problem with the Random Forest Classifier, using the test dataset to evaluate the models.

	Accuracy	Specificity	Sensitivity	Precision
1	70.9%	78.3%	50.0%	44.4%
2	77.4%	78.3%	75.0%	54.5%
3	77.4%	78.3%	75.0%	54.5%
4	74.2%	78.3%	62.5%	50.0%

Table 4.37: Best results on cross-validation for all image features in the binary problem with SVM, using random train and validation datasets from the initial train dataset to evaluate the models.

	Accuracy	Specificity	Sensitivity	Precision
1	80.0%	73.3%	90.0%	69.2%
2	76.0%	90.9%	64.3%	90.0%
3	72.0%	90.9%	57.1%	88.9%
4	72.0%	83.3%	61.5%	80.0%

Table 4.38: Test results for the best models chosen from the cross-validation for all image features in the binary problem with the SVM Classifier, using the test dataset to evaluate the models.

	Accuracy	Specificity	Sensitivity	Precision
1	58.1%	60.9%	50.0%	30.8%
2	61.3%	69.6%	37.5%	30.0%
3	61.3%	73.9%	25.0%	25.0%
4	64.5%	69.6%	50.0%	36.4%

Table 4.39: Best results on cross-validation for all image features in the binary problem with K-Nearest Neighbors Classifier, using random train and validation datasets from the initial train dataset to evaluate the models.

	Accuracy	Specificity	Sensitivity	Precision
1	76.0%	81.3%	66.7%	66.7%
2	76.0%	78.6%	72.7%	72.7%
3	76.0%	75.0%	77.8%	63.6%
4	68.0%	83.3%	53.8%	77.8%

Table 4.40: Test results for the best models chosen from the cross-validation for all image features in the binary problem with the K-Nearest Neighbors Classifier, using the test dataset to evaluate the models.

	Accuracy	Specificity	Sensitivity	Precision
1	61.3%	69.6%	37.5%	30.0%
2	54.8%	69.6%	12.5%	12.5%
3	58.1%	65.2%	37.5%	27.3%
4	64.5%	73.9%	37.5%	33.3%

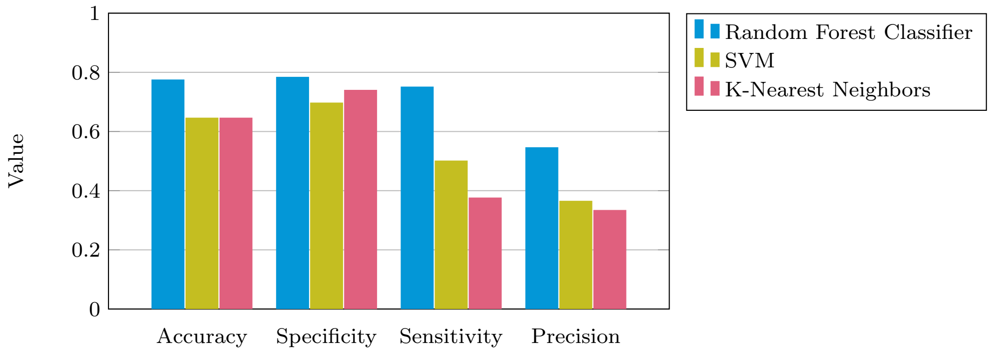


Figure 4.11: Best models from each classifier using all image features in the binary problem.

Analyzing these results, we can see that the precision of the three best models is low, which means that there were many cases predicted as depressive that were non-depressive. Moreover, the sensitivity was low in most of the models, also regarding the depressive class, a lower amount of depressive users than the non-depressive was correctly classified. On the other side, the specificity is high, meaning that the models correctly predicted most of the non-depressive users. What we can say about these models is that they are misclassifying the positive class. The next step is to analyze each feature and see if they are useful for dividing these two groups, to improve these results.

Table 4.41: Results from the Student's T-test on the image features in the binary problem.

Features	p-value
Brightness	0.4724
Contrast	0.0207
Sharpness	0.3264
Red color quantity in the dominant color	0.8876
Green color quantity in the dominant color	0.7542
Blue color quantity in the dominant color	0.9045
Height	0.2294
Width	0.0639

Observing the p-values, we can conclude that most of these features are not good enough for distinguishing these two groups, but we are keeping all features with a p-value less than 0.1. We are only using the contrast feature and the picture width.

After the final dataset is ready, we looked for the optimal parameters for the three classifiers.

Table 4.42: Parameters for the Random Forest, SVM and KNN Classifiers that were tested to find the optimal one to fit the best image features in the binary problem.

Parameters	Values
n_estimators	30
max_features	'auto'
max_depth	8
min_samples_split	2
min_samples_leaf	2
bootstrap	True

Parameters	Values
kernel	'rbf'
C	1
gamma	0.1

Parameters	Values
n_neighbors	3
metric	'euclidean'

With the parameters of the Random Forest Classifier optimized, we trained 100 models to inspect the range of results.

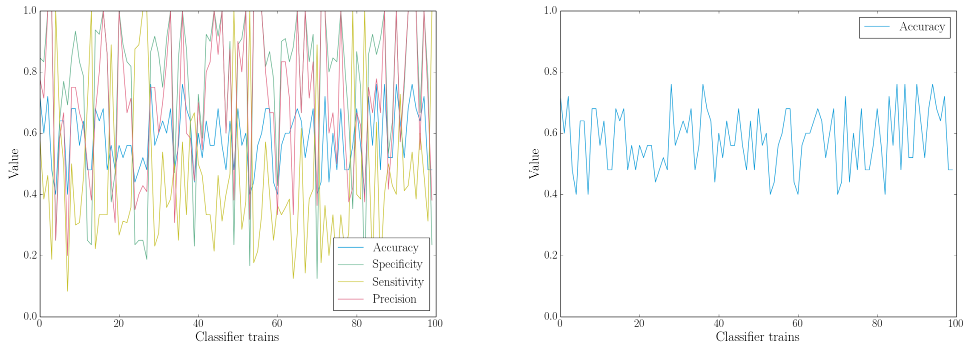


Figure 4.12: Cross-validation results of 100 trained Random Forest Classifiers with the train dataset with the best image features in the binary problem. On the left, the four standard terms and on the right, the accuracy.

After taking a look at the overall results, we chose the best ones to test with our test dataset.

Table 4.43: Best results on cross-validation for the best image features in the binary problem with Random Forest Classifier, using random train and validation datasets from the initial train dataset to evaluate the model.

	Accuracy	Specificity	Sensitivity	Precision
1	84.0%	100.0%	71.4%	100.0%
2	80.0%	91.7%	69.2%	90.0%
3	76.0%	92.3%	58.3%	87.5%
4	76.0%	100.0%	53.8%	100.0%

Table 4.44: Test results for the best models chosen from the cross-validation for the best image features in the binary problem with the Random Forest Classifier, using the test dataset to evaluate the models.

	Accuracy	Specificity	Sensitivity	Precision
1	58.1%	56.5%	62.5%	33.3%
2	51.6%	43.5%	75.0%	31.6%
3	45.2%	34.8%	75.0%	28.6%
4	54.8%	52.7%	62.5%	31.2%

Looking at the results from the Random Forest Classifier, we can observe that the models have poor overall results when tested, misclassifying most of the samples. The reason for this to happen may be that we are only using two features. We will train a SVM Classifier, hoping to obtain better models and to see if these results remain the same.

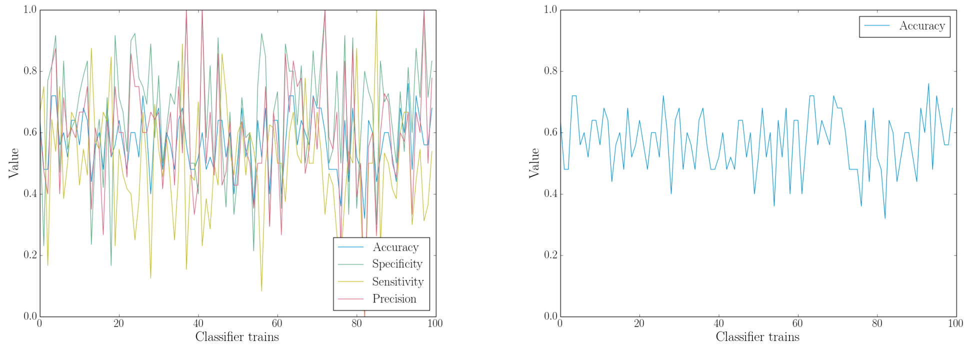


Figure 4.13: Cross-validation results of 100 trained SVM Classifiers with the train dataset with the best image features in the binary problem. On the left, the four standard terms and on the right, the accuracy.

Table 4.45: Best results on cross-validation for the best image features in the binary problem with SVM Classifier, using random train and validation datasets from the initial train dataset to evaluate the models.

	Accuracy	Specificity	Sensitivity	Precision
1	80.0%	92.3%	66.7%	88.9%
2	72.0%	73.3%	70.0%	63.6%
3	80.0%	81.2%	77.8%	70.0%
4	76.0%	92.3%	58.3%	87.5%

Table 4.46: Test results for the best models chosen from the cross-validation for the best image features in the binary problem with the SVM Classifier, using the test dataset to evaluate the models.

	Accuracy	Specificity	Sensitivity	Precision
1	71.0%	78.3%	50.0%	44.4%
2	74.2%	78.3%	62.5%	50.0%
3	67.7%	73.9%	50.0%	40.0%
4	61.3%	60.9%	62.5%	35.7%

For the SVM Classifier, the models are better than with the Random Forest Classifier, although they still have poor values of sensitivity and precision, meaning that the models are failing most in classifying the positive class.

Finally, we are using the K-Nearest Neighbors Classifier. First, we will choose the best parameters (Table 4.42).

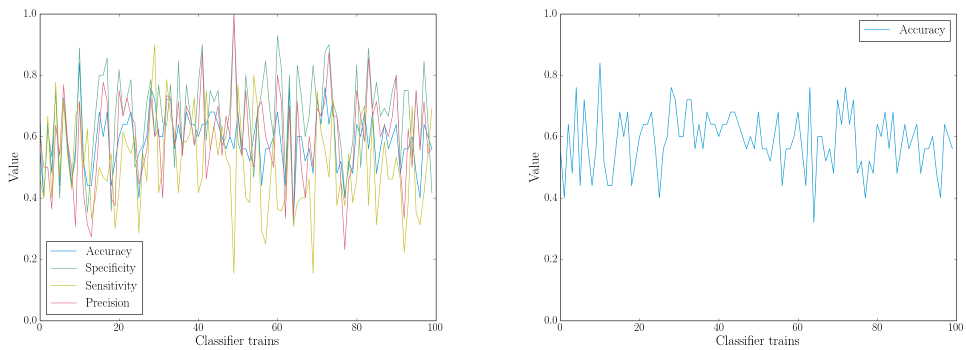


Figure 4.14: Cross-validation results of 100 trained K-Nearest Neighbors Classifiers with the train dataset of the best image features in the binary problem. On the left, the four standard terms and on the right, the accuracy.

Table 4.47: Best results on cross-validation for the best image features in the binary problem with KNN Classifier, using random train and validation datasets from the initial train dataset to evaluate the models.

	Accuracy	Specificity	Sensitivity	Precision
1	76.0%	84.6%	66.7%	80.0%
2	76.0%	86.7%	60.0%	75.0%
3	72.0%	72.7%	71.4%	76.9%
4	72.0%	90.9%	57.1%	88.9%

Table 4.48: Test results for the best models chosen from the cross-validation for the best image features in the binary problem with the Random Forest Classifier, using the test dataset to evaluate the models.

	Accuracy	Specificity	Sensitivity	Precision
1	77.4%	78.3%	75.0%	54.5%
2	77.4%	82.6%	62.5%	55.6%
3	77.4%	78.3%	75.0%	54.5%
4	77.4%	78,3%	75.0%	54.5%

The models using the KNN algorithm gave us the best results. In this case, only the precision has lower results. Although these results are better than the ones for the two previous classifiers, they are still not reasonable.

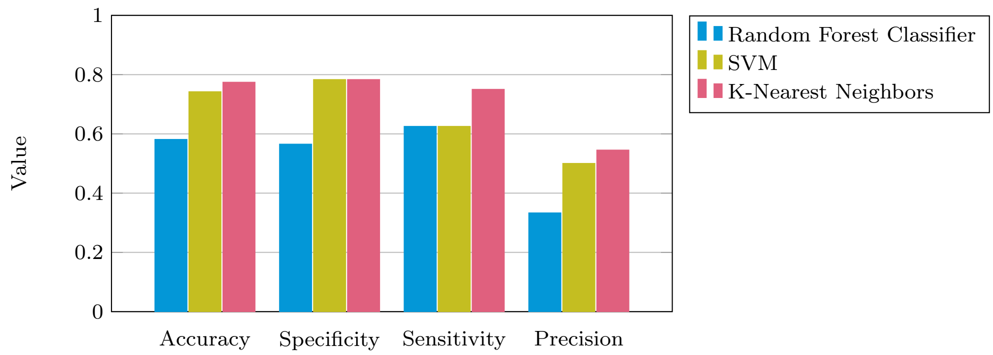


Figure 4.15: Best models from each classifier using the best image features in the binary problem.

Looking at the results of the three classifiers, we can conclude that the best results for image features were obtained with the K-Nearest Neighbors classifier, having all models in this classifier similar results. The results obtained are better than in the classification without taking into account the best features, and now, both groups are being correctly predicted.

4.1.4 Text, Usage Metrics' and Image Features combined

As the aim of this project is to create multimodal models for predicting depression among young adults on social networks, we are now training models with the combination of the text, usage metrics' and image features.

We started directly by analyzing this group of features and chose the best ones to distinguish between these two classes.

Table 4.49: P-value obtained from the Student's T-test on all features in the binary problem.

Features	p-value	Features	p-value
Number of characters	0.2268	Days since last post	0.0600
Number of words	0.0011	Average post frequency	0.2009
Existence of emojis	0.0700	Brightness	0.6947
Number of emojis	0.0280	Contrast	0.2683
Number of tags	0.0182	Sharpness	0.5789
Existence of tags	0.0051	Red color quantity in the dominant color	0.4112
Day/Night time	0.6838	Green color quantity in the dominant color	0.6586
Number of likes	0.1142	Blue color quantity in the dominant color	0.5890
Number of comments	0.4681	Height	0.3059
Number of different people comments	0.3215	Width	0.0063

After analyzing the p-value obtained we kept all features with this value lower than 0.1. Our new dataset will have the number of words, the existence and number of emojis, number and the existence of tags, the number of days since the user's last post and width of the image.

The next step was to choose the optimal parameters for the first classifier we are using, Random Forest Classifier.

Table 4.50: Parameters for the Random Forest, SVM and KNN Classifiers that were tested to find the optimal one to fit the best overall features in the binary problem.

Parameters	Values
n_estimators	200
max_features	'sqrt'
max_depth	60
min_samples_split	2
min_samples_leaf	2
bootstrap	True

Parameters	Values
kernel	'linear'
C	0.1
gamma	1e-7

Parameters	Values
n_neighbors	7
metric	'euclidean'

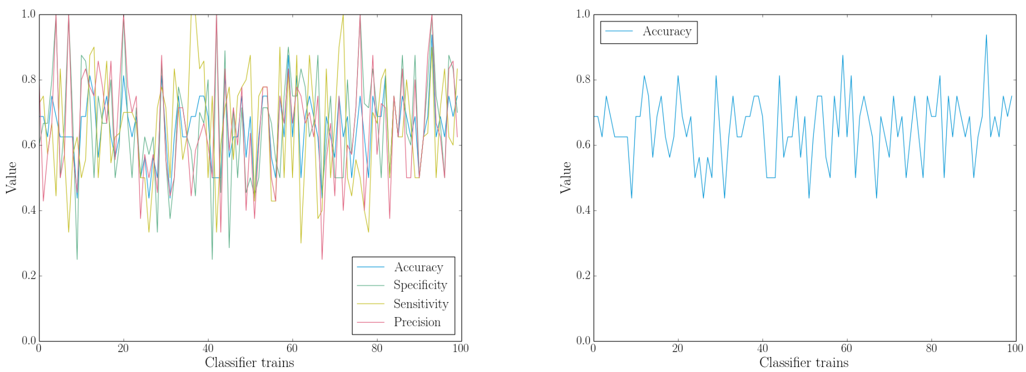


Figure 4.16: Cross-validation results of 100 trained Random Forest Classifiers with the train dataset of all features in the binary problem. On the left, the four standard terms and on the right, the accuracy.

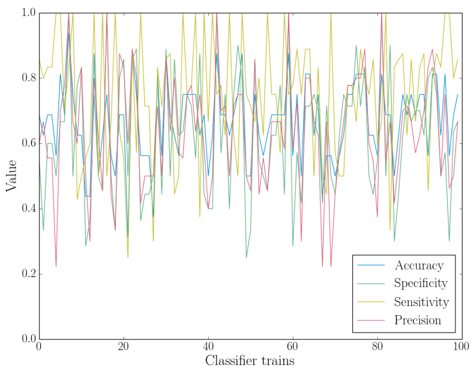
Table 4.51: Best results on cross-validation for the best overall features in the binary problem with Random Forest Classifier, using random train and validation datasets from the initial train dataset to evaluate the models.

	Accuracy	Specificity	Sensitivity	Precision
1	87.5%	100.0%	77.8%	100.0%
2	81.2%	88.9%	71.4%	83.3%
3	75.0%	77.8%	71.4%	71.4%
4	75.0%	70.0%	83.3%	62.5%

Table 4.52: Test results for the best models chosen from the cross-validation for the best overall features in the binary problem with the Random Forest Classifier, using the test dataset to evaluate the models.

	Accuracy	Specificity	Sensitivity	Precision
1	85.7%	83.3%	88.8%	80.0%
2	76.2%	66.7%	88.9%	66.7%
3	80.9%	75.0%	88.9%	72.7%
4	90.5%	91.7%	88.9%	88.9%

Looking at these results, we can find prominent models with values between 88% and 92%. An accuracy of 90.5% means that the model is only failing in predicting 2 in 22 samples.



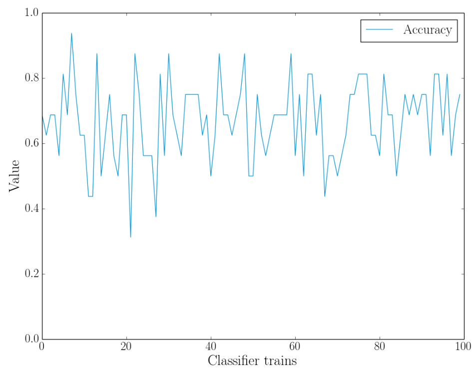


Figure 4.17: Cross-validation results of 100 trained SVM Classifiers for the train dataset of the best overall features in the binary problem. On the left, the four standard terms and on the right, the accuracy.

Table 4.53: Best results on cross-validation for the best overall features in the binary problem with SVM Classifier, using random train and validation datasets from the initial train dataset to evaluate the models.

	Accuracy	Specificity	Sensitivity	Precision
1	87.5%	80.0%	90.9%	90.9%
2	87.5%	71.4%	100.0%	81.8%
3	75.0%	88.9%	57.1%	80.0%
4	75.0%	83.3%	70.0%	87.5%

Table 4.54: Test results for the best models chosen from the cross-validation for the best overall features in the binary problem with the SVM Classifier, using the test dataset to evaluate the models.

	Accuracy	Specificity	Sensitivity	Precision
1	71.4%	75.0%	66.7%	66.7%
2	71.4%	66.7%	77.8%	63.6%
3	71.4%	75.0%	66.7%	66.7%
4	71.4%	75.0%	66.7%	66.7%

With this classifier, we did not obtain as good results as with the Random Forest Classifier. Here we have a similar problem as observed when using only one group of features; the positive class is being misclassified (depressive users). Finally, we are using the KNN algorithm to find the best model for these features.

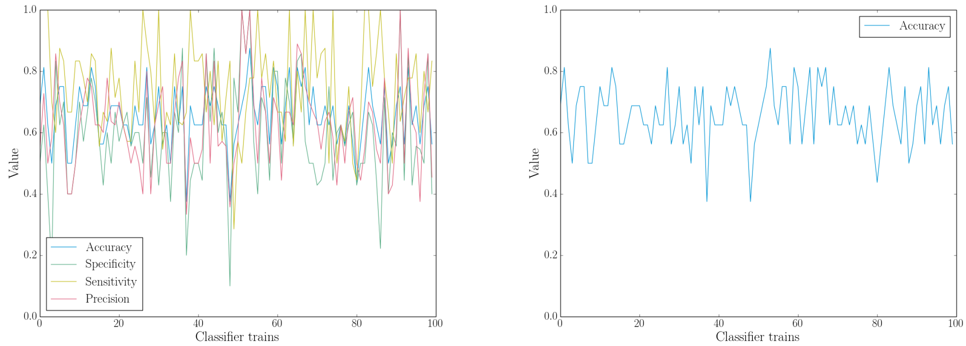


Figure 4.18: Cross-validation results of 100 trained K-Nearest Neighbors Classifiers for the image features in the binary problem. On the left, the four standard terms and on the right, the accuracy.

Table 4.55: Best results on cross-validation for the best overall features in the binary problem with KNN Classifier, using random train and validation datasets from the initial train dataset to evaluate the models.

	Accuracy	Specificity	Sensitivity	Precision
1	81.2%	77.8%	85.7%	75.0%
2	81.2%	100.0%	76.9%	100.0%
3	75.0%	71.4%	77.8%	77.8%
4	75.0%	77.8%	71.4%	71.4%

Table 4.56: Test results for the best models chosen from the cross-validation for the best overall features in the binary problem with the KNN Classifier, using the test dataset to evaluate the models.

	Accuracy	Specificity	Sensitivity	Precision
1	71.4%	66.7%	77.8%	63.6%
2	76.2%	83.3%	66.7%	75.0%
3	76.2%	75.0%	77.8%	70.0%
4	76.2%	83.3%	66.7%	75.0%

With the KNN algorithm, we also got good models, but not better than the ones with the Random Forest algorithm. In the next figure, we will compare the best from each algorithm.

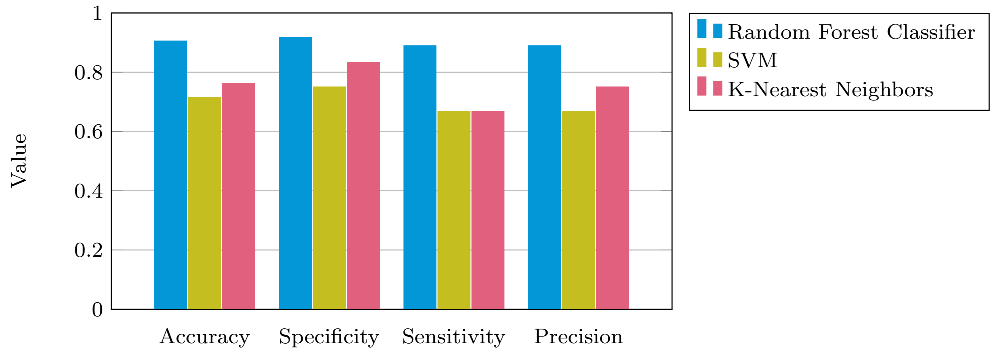


Figure 4.19: Best models from each classifier using the best overall features in the binary problem.

We can see that the best model uses the Random Forest Classifier. Using the combination of the best features of the three groups improved a lot our models and results.

Overall, we obtained acceptable results regarding the amount of data collected. For better results, it would be indispensable to have a significant number of users, divided into both classes. Social networks' posts are very variable from user to user, and from post to post, and it is essential to have as much data as possible to cover all of the differences that can occur between them.

Although the expectations are low, we will next try to predict the four groups (minimal depression, mild depression, moderate depression and severe depression).

4.2 Multiclass Problem: Minimal Depression, Mild Depression, Moderate Depression and Severe Depression

For this multiclass problem, we changed the labels from the original scores the four groups (minimal depression, mild depression, moderate depression and severe depression). Scores between 0 and 13 were changed to 0, between 14 and 19 were changed to 1, between 20 and 28 changed to 2 and between 29 and 63 changed to 3. Then, we analyzed the features to see if they were able to distinguish these four groups correctly. For this, we are doing an ANOVA between these four groups for each feature.

So after feature extraction and reduction, it was time to normalize the data, as done previously, so that the average score of the data was approximately zero and the standard deviation unitary.

After our dataset was ready, we tried to balance the dataset because there were more samples for the minimal depression group and severe depression group than the other two groups. After this, the dataset was divided into a train and test datasets.

Like in the previous problem, the parameters of the classifiers were chosen with the function RandomizedSearchCV for the Random Forest Classifier and GridSearchCV for the SVM and K-Nearest Neighbors Classifiers, both from the package sklearn.

After the best parameters were chosen, we trained each classifier 100 times and plotted the results to have an idea of the accuracy of these models. The next step was to analyze these results to see if we were able to obtain good models, and what we concluded was that these models were not good. The models were not good because some groups had only ten samples, and by balancing them we had very few samples in each group to properly train the classifiers. We tried to train these classifiers with the unbalanced dataset in order to see if it was possible to obtain better models, and we got better results. We also paid attention that the models were not wholly misclassifying classes, a common problem when working with unbalanced datasets.

4.2.1 Text Features

Unlike the previous problem, we are starting by analyzing the features, since we saw that the results were always better after doing this.

Table 4.57: p-values from the ANOVA F-test on the text features for distinguishing the four classes.

	p-value
Number of characters	0.0059
Number of words	0.0205
Existence of emojis	0.0948
Number of emojis	0.2749
Number of tags	1.394e-06
Existence of tags	0.0088

Most of the text features are good to distinguish these four groups, so we are keeping the features with p-values less than 0.05, keeping the number of characters, number of words, number of tags and existence of tags.

After having our balanced dataset ready with the features and the best parameters for each classifier, we search for the best parameters for our classifiers.

Table 4.58: Parameters for the Random Forest and KNN Classifiers that were tested to find the optimal one to fit the best text features in the multiclass problem.

Parameters	Values
n_estimators	800
max_features	'sqrt'
max_depth	2
min_samples_split	2
min_samples_leaf	4
bootstrap	True

Parameters	Values
n_neighbors	7
metric	'euclidean'

We then trained each classifier 100 times. Noticing that the obtained models were not good, we also trained all classifiers 100 times with the unbalanced dataset.

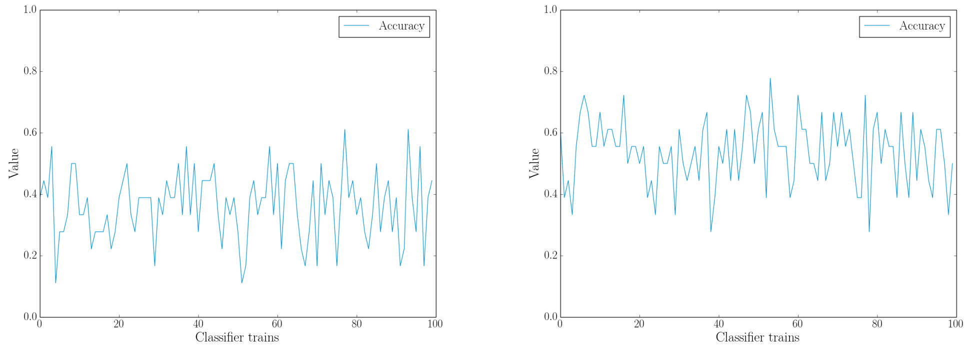


Figure 4.20: Cross-validation results of 100 trained Random Forest Classifiers with the train dataset of the best text features in the multiclass problem. On the left, the accuracy for the models using the balanced dataset and on the right, using the unbalanced dataset.

Comparing these two graphics in figure 4.20, we decided to use the unbalanced dataset, since the overall results are better. These poor values for the balanced dataset happen because the class with fewer samples has around 15, making us shorten our total dataset drastically.

Table 4.59: Best results on cross-validation for the best text features in the multiclass problem with Random Forest Classifier, using random train and validation datasets from the initial train dataset to evaluate the models.

	Accuracy
1	72.0%
2	66.7%
3	66.7%
4	61.1%

Table 4.60: Test results for the best models chosen from the cross-validation for the best text features in the multiclass problem with the Random Forest Classifier, using the test dataset to evaluate the models.

	Accuracy
1	40.9%
2	40.9%
3	40.9%
4	40.9%

The models we obtain have all poor results, which was expectable because of the little samples we have of each class. We will also use the K-Nearest Neighbors algorithm to obtain better results.

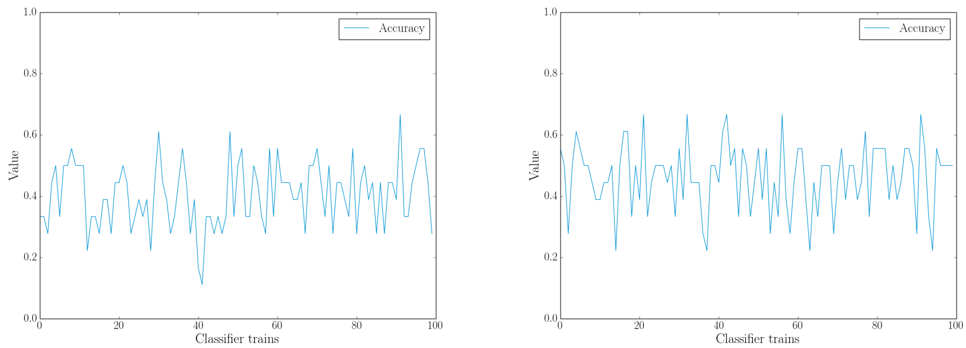


Figure 4.21: Cross-validation results of 100 trained K-Nearest Neighbors Classifiers with the train dataset of the best text features in the multiclass problem. On the left, the accuracy for the models using the balanced dataset and on the right, using the unbalanced dataset.

Table 4.61: Best results on cross-validation for the best text features in the multiclass problem with the KNN Classifier, using random train and validation datasets from the initial train dataset to evaluate the models.

	Accuracy
1	72.0%
2	66.7%
3	66.7%
4	66.7%

Table 4.62: Test results for the best models chosen from the cross-validation for the best text features in the multiclass problem with the KNN Classifier, using the test dataset to evaluate the models.

	Accuracy
1	40.9%
2	40.9%
3	40.9%
4	40.9%

Looking at these models, we can draw the same conclusion as with the previous one.

Analyzing the results obtained with all the models using the text features, we can conclude that with none of them we obtain good models. This was expectable as we had very few samples for some of the groups and could not balance the dataset properly without rejecting many samples that would be important to help train these models.

4.2.2 Usage Metrics' Features

Table 4.63: Results from the ANOVA F-test on the usage metrics' features in the multiclass problem.

	p-value
Day/Night time	0.3778
Number of likes	1.6270e-06
Number of comments	0.0014
Number of different people comments	0.0015
Days since last post	0.0772
Average post frequency	0.0090

The usage metrics' features are the best to distinguish these four groups. We are using only the ones with p-values less than 0.05, so we are keeping all except day-/night-time posting and days since the last post.

Table 4.64: Parameters for the Random Forest and KNN Classifiers that were tested to find the optimal one to fit the best usage metrics' features in the multiclass problem.

Parameters	Values
n_estimators	1400
max_features	'auto'
max_depth	30
min_samples_split	2
min_samples_leaf	2
bootstrap	True

Parameters	Values
n_neighbors	3
metric	'cityblock'

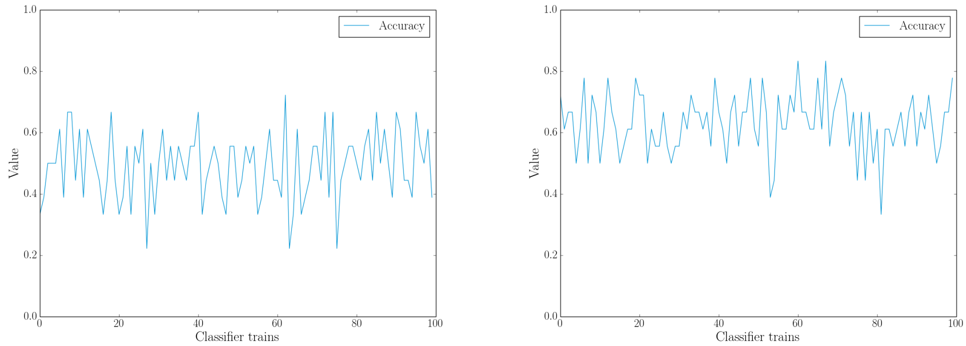


Figure 4.22: Cross-validation results of 100 trained Random Forest Classifiers for the usage metrics' features in the multiclass problem. On the left, the accuracy for the models using the balanced dataset and on the right, using the unbalanced dataset.

In this case, we also chose the dataset that is unbalanced for the Random Forest classifier, since it had better results than the balanced one.

Table 4.65: Best results on cross-validation for the best usage metrics' features in the multiclass problem with Random Forest Classifier, using random train and validation datasets from the initial train dataset to evaluate the models.

	Accuracy
1	83.3%
2	77.8%
3	77.8%
4	77.8%

Table 4.66: Test results for the best models chosen from the cross-validation for the best usage metrics' features in the multiclass problem with the Random Forest Classifier, using the test dataset to evaluate the models.

	Accuracy
1	45.5%
2	40.9%
3	45.5%
4	36.4%

Like with the previous group of features, when testing the models, the values for the accuracy obtained are weak, although when validating the models the results were good, above 77%.

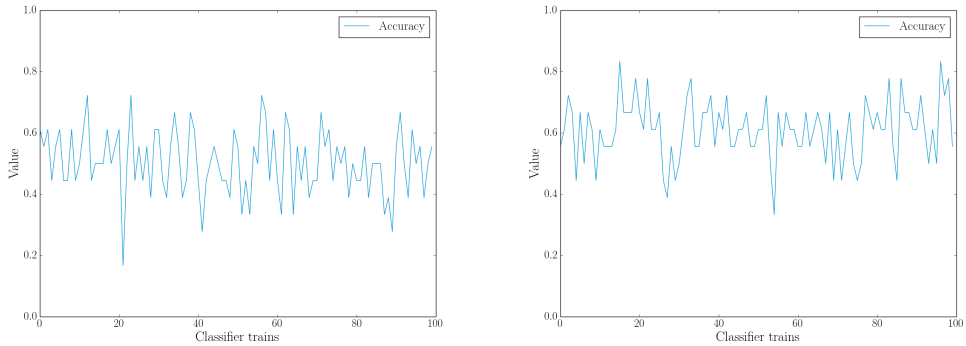


Figure 4.23: Cross-validation results of 100 trained K-Nearest Neighbors Classifiers for the usage metrics' features in the multiclass problem. On the left, the accuracy for the models using the balanced dataset and on the right, using the unbalanced dataset.

Table 4.67: Best results on cross-validation for the best usage metrics' features in the multiclass problem with K-Nearest Neighbors Classifier, using random train and validation datasets from the initial train dataset to evaluate the models.

	Accuracy
1	83.3%
2	83.3%
3	72.2%
4	72.2%

Table 4.68: Test results for the best models chosen from the cross-validation for the best usage metrics' features in the multiclass problem with the K-Nearest Neighbors Classifier, using the test dataset to evaluate the models.

	Accuracy
1	50.0%
2	40.9%
3	54.5%
4	45.5%

Comparing with the group of the text features, the results of these models are as weak as theirs. In both cases, the trained models presented reasonable values of accuracy, but when tested, were quickly neglected.

4.2.3 Image Features

We are now analyzing the features extracted from the images of the posts and look for better results than the two previous groups.

Table 4.69: P-values from the ANOVA F-test on the image features for the multiclass problem.

Features	p-value
Brightness	0.2724
Contrast	0.1034
Sharpness	0.3339
Red color quantity in the dominant color	0.3328
Green color quantity in the dominant color	0.5476
Blue color quantity in the dominant color	0.8828
Height	0.6053
Width	0.0044

Looking at these p-values, we conclude that only one feature is useful to distinguish these four groups. We are only using the width feature in this case. As in the previous groups, we are comparing the balanced dataset with the unbalanced, so we find the best results.

Table 4.70: Parameters for the Random Forest and KNN Classifiers that were tested to find the optimal one to fit the best image features in the multiclass problem.

Parameters	Values
n_estimators	2000
max_features	'sqrt'
max_depth	2
min_samples_split	2
min_samples_leaf	4
bootstrap	True

Parameters	Values
n_neighbors	25
metric	'euclidean'

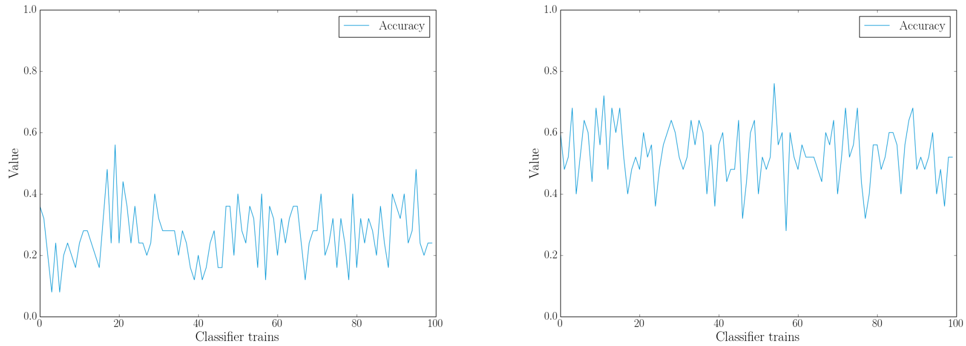


Figure 4.24: Cross-validation results of 100 trained Random Forest Classifiers with the train dataset of the best image features in the multiclass problem. On the left, the accuracy for the models using the balanced dataset and on the right, using the unbalanced dataset.

As previously, we also chose the unbalanced dataset in both classifiers, since it was the one with best results overall. We are now looking for the best models to test them with our unseen data.

Table 4.71: Best results on cross-validation for the best image features in the multiclass problem with Random Forest Classifier, using random train and validation datasets from the initial train dataset to evaluate the models.

	Accuracy
1	72.0%
2	68.0%
3	68.0%
4	52.0%

Table 4.72: Test results for the best models chosen from the cross-validation for the best image features in the multiclass problem with the Random Forest Classifier, using the test dataset to evaluate the models.

	Accuracy
1	51.6%
2	58.1%
3	51.6%
4	58.1%

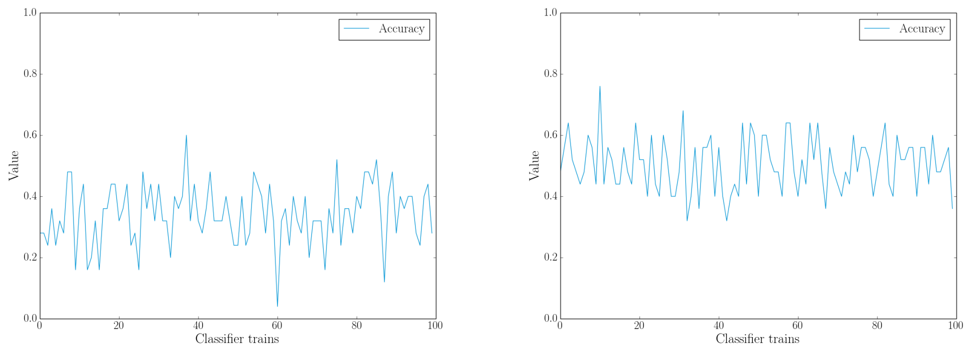


Figure 4.25: Cross-validation results of 100 trained K-Nearest Neighbors Classifiers with the train dataset of the image features in the multiclass problem. On the left, the accuracy for the models using the balanced dataset and on the right, using the unbalanced dataset.

The K-Nearest Neighbors Classifier could not predict any other class besides the minimal depression, so the models were not taken into account since the accuracy depended on the number of samples from this class present in the test dataset.

The image features gave us terrible models for this prediction, and the features we had from the images were not suitable for this separation.

4.2.4 Text, Usage Metrics' and Image Features combined

Table 4.73: p-value obtained from the Student's T-test on all features in the multiclass problem.

Features	p-value
Number of characters	0.4406
Number of words	9.283e-6
Existence of emojis	0.0018
Number of emojis	0.0009
Number of tags	0.0596
Existence of tags	0.0163
Day/Night time	0.0160
Number of likes	0.0442
Number of comments	0.2824
Number of different people comments	0.4908

Features	p-value
Days since last post	6.609e-6
Average post frequency	0.0190
Brightness	0.8719
Contrast	0.7058
Sharpness	0.5188
Red color quantity in the dominant color	0.3887
Green color quantity in the dominant color	0.6650
Blue color quantity in the dominant color	0.6683
Height	0.6099
Width	0.0581

Looking into the obtained p-values, we have some features good for distinguishing our four classes. We are adding to our new dataset the number of words, existence and number of emojis, existence and number of tags, day/night post time, number of likes, number of days since the user's last post, average post frequency and width of the posted image.

Table 4.74: Parameters for the Random Forest and KNN Classifiers that were tested to find the optimal one to fit the best overall features in the multiclass problem.

Parameters	Values
n_estimators	70
max_features	'auto'
max_depth	30
min_samples_split	10
min_samples_leaf	1
bootstrap	True

Parameters	Values
n_neighbors	7
metric	'cityblock'

As usually, we trained 100 times the classifiers both with balanced and unbalanced datasets so we could analyze the range of results the models could provide.

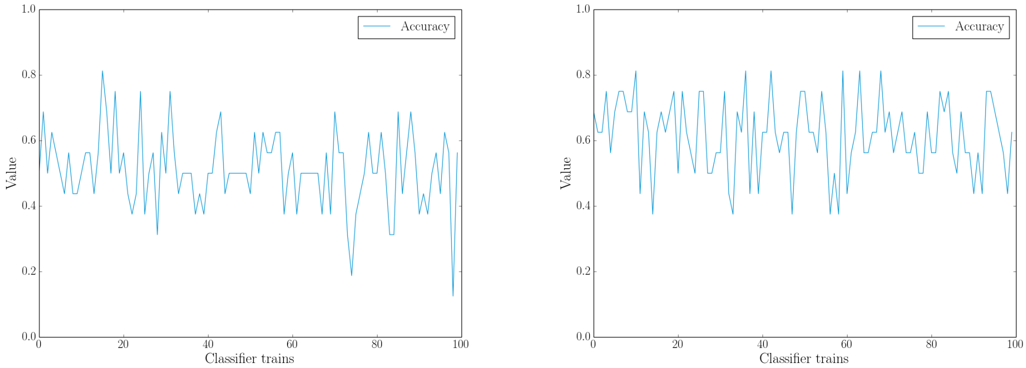


Figure 4.26: Cross-validation results of 100 trained Random Forest Classifiers with the train dataset of best overall features in the multiclass problem. On the left, the accuracy for the models using the balanced dataset and on the right, using the unbalanced dataset.

As previously, the unbalanced dataset gave us slightly better results, so this is the one we are using to train our classifiers.

Table 4.75: Best results on cross-validation for the best overall features in the multiclass problem with Random Forest Classifier, using random train and validation datasets from the initial train dataset to evaluate the models.

	Accuracy
1	75.0%
2	68.8%
3	68.8%
4	68.8%

Table 4.76: Test results for the best models chosen from the cross-validation for the best overall features in the multiclass problem with the Random Forest Classifier, using the test dataset to evaluate the models.

	Accuracy
1	71.4%
2	76.2%
3	76.2%
4	71.4%

As we can observe, the models are good for the amount of data we have for each class. The best ones present an accuracy over 76% when predicting new unseen data.

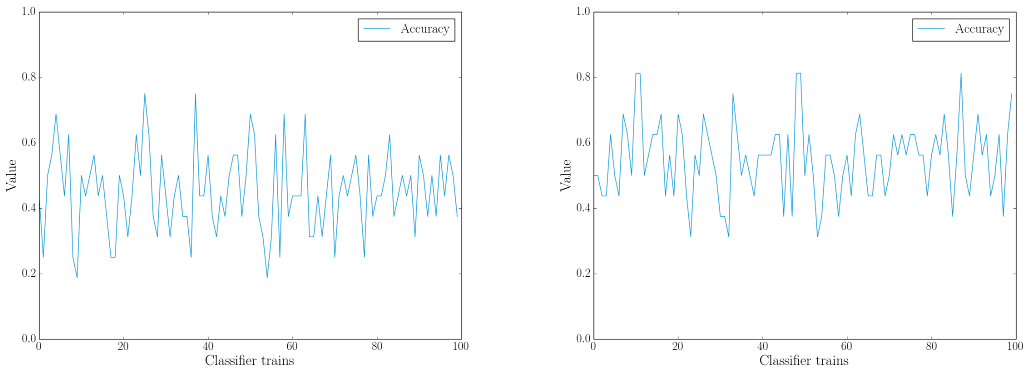


Figure 4.27: Cross-validation results of 100 trained K-Nearest Neighbors Classifiers for the best overall features in the multiclass problem. On the left, the accuracy for the models using the balanced dataset and on the right, using the unbalanced dataset.

As for the Random Forest Classifier, the K-Nearest Neighbors algorithm is presenting better results when using the unbalanced dataset.

Table 4.77: Best results on cross-validation for the best overall features in the multiclass problem with K-Nearest Neighbors, using random train and validation datasets from the initial train dataset to evaluate the models.

	Accuracy
1	81.3%
2	81.3%
3	68.8%
4	68.8%

Table 4.78: Test results for the best models chosen from the cross-validation for the best overall features in the multiclass problem with K-Nearest Neighbors, using the test dataset to evaluate the models.

	Accuracy
1	61.9%
2	57.1%
3	52.4%
4	61.9%

With the KNN algorithm, the models obtained worse results than with the Random Forest one.

Future work for this problem would be to get a much larger dataset, with samples from all of the groups, having enough samples to train the models and predicting these four different groups. A few more features could be chosen, at least for the image features group, because this group only has one feature used in the final dataset.

4.3 Regression Problem: distinguishing all Different Scores

For the regression problem, we started by normalizing our dataset so that all features would have a null average and a unit standard deviation. In this problem, we will try to predict the exact depression score from each post.

After the dataset processing, we divided our dataset into a train and test ones. We then trained each of the two regressors 100 times using a training and validation dataset randomly divided each time and plotted the mean square error value of each training sample. In this problem, we are using the Random Forest Regressor and the Support Vector Machine Regression.

The best models were then chosen and tested with the test dataset. We also plotted the scores from the validation and test datasets, and the scores predicted, to be able to analyze it visually.

Table 4.79: Parameters for the Random Forest Regressor that were tested to find the optimal one.

Parameters	Values
n_estimators	[10, 30, 50, 70, 100, 200, 400, 600, 800, 1000, 1200, 1400, 1600, 1800, 2000]
max_features	['auto', 'sqrt']
max_depth	[2, 4, 8, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 110, None]
min_samples_split	[2, 5, 10]
min_samples_leaf	[1, 2, 4]
bootstrap	[True, False]

Table 4.80: Parameters for the Support Vector Regression that were tested to find the optimal one.

Parameters	Values
epsilon	[0.1, 0.2, 0.5, 0.3]
kernel	['linear', 'rbf', 'poly']
C	[0.001, 0.01, 0.1, 1, 10]
gamma	[1e-7, 1e-4, 0.001, 0.01, 0.1, 1]

4.3.1 Text Features

We started by analyzing the Mean Squared Error of 100 trained Random Forest classifiers. The lower the MSE, the better the model. We used the following parameters on the regressors.

Table 4.81: Parameters for the Random Forest Regressor and SVR that were tested to find the optimal one to fit the text features in the regression problem.

Parameters	Values
n_estimators	1600
max_features	'sqrt'
max_depth	20
min_samples_split	10
min_samples_leaf	1
bootstrap	True

Parameters	Values
epsilon	0.05
C	1.5
gamma	1e-7
kernel	'linear'

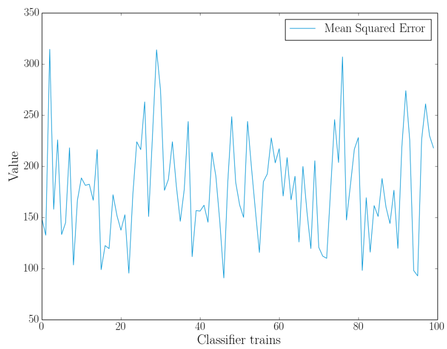


Figure 4.28: Cross-validation results of 100 trained Random Forest Classifiers with the train dataset of text features in the regression problem.

We can conclude that many models have a high value of the Mean Squared Error, we will try to work with the ones with the lowest.

Table 4.82: Best results on cross-validation for all text features in the regression problem with Random Forest Regressor, using random train and validation datasets from the initial train dataset to evaluate the models.

	Mean Squared Error
1	101.90
2	137.40
3	115.72
4	98.73

Table 4.83: Test results for the best models chosen from the cross-validation for the text features in the regression problem with the Random Forest Regressor, using the test dataset to evaluate the models.

	Mean Squared Error
1	153.83
2	132.52
3	138.16
4	152.52

For a better understanding of what these values mean, we plotted the actual labels and the predicted ones in figure 4.29.

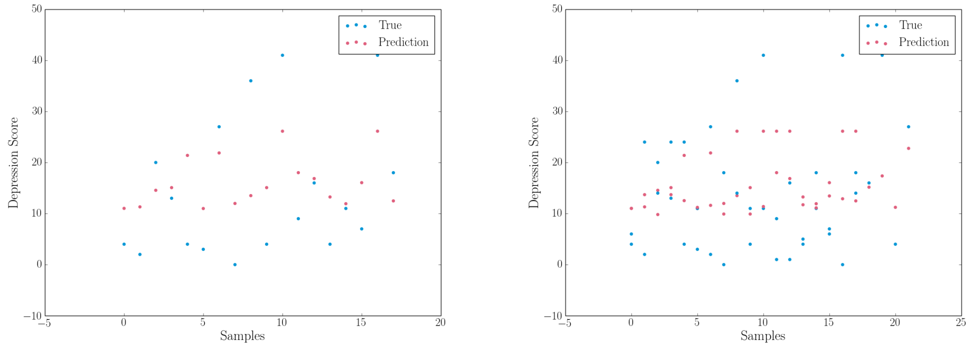


Figure 4.29: Best model (3) for the regression problem with the text features in the regression problem using the Random Forest Regression. On the left, the train results and on the right, the test results

Taking a look at this two graphics in figure 4.29 we can observe that most of the points of the predictions are less than ten scores away from their real value. This is not ideal, but comparing to what it was expected since we have a little amount of data, it shows us some impressive results.

We are now using the SVM Regression to predict the exact scores.

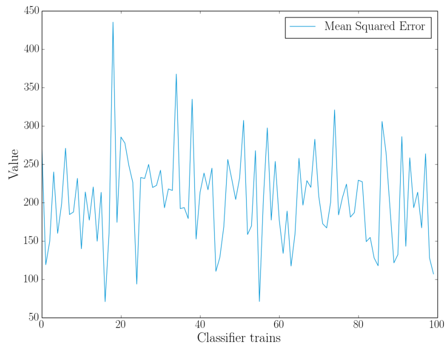


Figure 4.30: Cross-validation results of 100 trained Support Vector Machine Regression with the train dataset of text features in the regression problem.

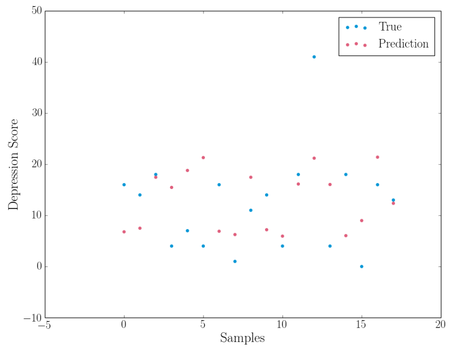
The values of the Mean Squared Error for the regressions with SVR are slightly higher than with the Random Forest Regressor.

Table 4.84: Best results on cross-validation for all text features in the regression problem with SVM Regression, using random train and validation datasets from the initial train dataset to evaluate the models.

	Mean Squared Error
1	123.47
2	153.34
3	117.87
4	94.11

Table 4.85: Test results for the best models chosen from the cross-validation for the text features in the regression problem with the SVM Regression, using the test dataset to evaluate the models.

	Mean Squared Error
1	118.67
2	132.02
3	95.12
4	92.83



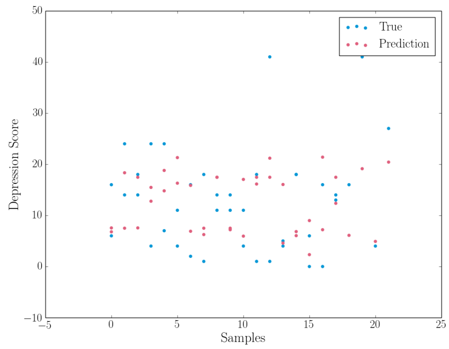


Figure 4.31: Best model (4) for the regression problem with the text features in the regression problem using the Support Vector Regression. On the left, the train results and on the right, the test results.

By analyzing the results and looking to the plot of the predicted and real scores, we can observe that the model with the SVR was better than the one with the Random Forest Regressor algorithm, since the scores were predicted with an average distance from their real value of eight scores, both for training and when testing.

4.3.2 Usage Metrics' Features

As previously, we started by training 100 regressors with the optimal parameters:

Table 4.86: Parameters for the Random Forest Regressor and SVR that were tested to find the optimal one to fit the usage metrics' features in the regression problem.

Parameters	Values
n_estimators	200
max_features	'auto'
max_depth	10
min_samples_split	5
min_samples_leaf	4
bootstrap	True

Parameters	Values
epsilon	0.5
C	1.5
gamma	1e-7
kernel	'linear'

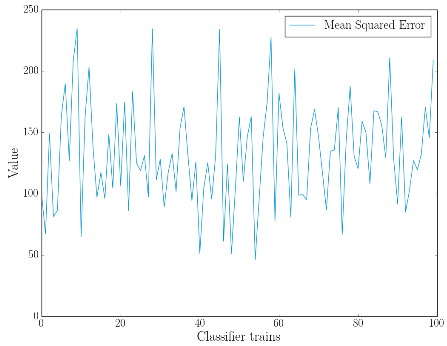


Figure 4.32: Cross-validation results of 100 trained Random Forest Regressor with the train dataset of text features in the regression problem.

Table 4.87: Best results for cross validation for usage metrics' features in the regression problem with Random Forest Regressor, using random train and validation datasets from the initial train dataset to evaluate the models.

	Mean Squared Error
1	76.17
2	77.42
3	73.44
4	73.45

Table 4.88: Test results for the best models for the usage metrics' features in the regression problem with Random Forest Regression,using the test dataset to evaluate the models.

	Mean Squared Error
1	96.68
2	97.51
3	99.88
4	98.01

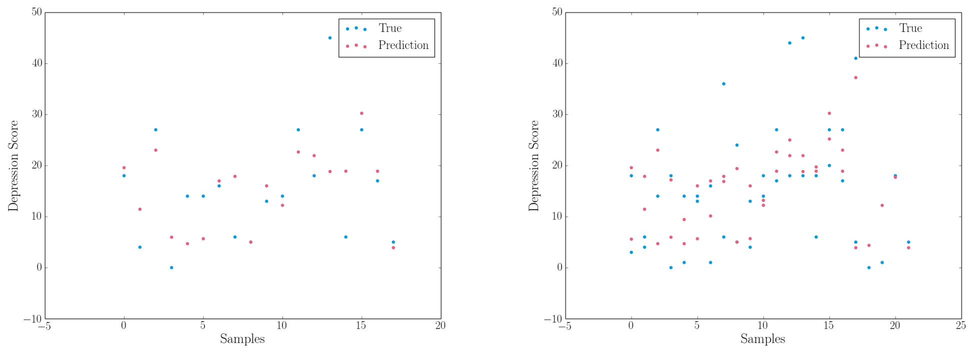


Figure 4.33: Best model (3) for the regression problem with the usage metrics' features in the regression problem using the Random Forest Regression. On the left, the train results and on the right, the test results.

The models obtained with this regressor are overall similar, so we chose one of them as the best model, which presented an average distance between the predicted and real scores of around six scores, which is better than the ones obtained with the text features. As we can also observe, the MSE in these models is smaller than the models with text features.

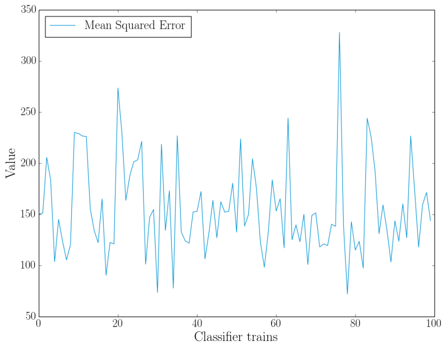


Figure 4.34: Cross-validation results of 100 trained Support Vector Regression with the train dataset of usage metrics' features in the regression problem.

With the SVR we obtained higher values when printing the MSE of the 100 trained regressors. We will now test the best models to analyze if this means the models are also worse than the obtained with the Random Forest Regressor.

Table 4.89: Best results for cross validation for usage metrics' features in the regression problem with Support Vector Regression, using random train and validation datasets from the initial train dataset to evaluate the models.

	Mean Squared Error
1	97.85
2	72.13
3	103.21
4	83.66

Table 4.90: Test results for the best models for the usage metrics' features in the regression problem with Support Vector Regression, using the test dataset to evaluate the models.

	Mean Squared Error
1	133.42
2	144.20
3	138.34
4	144.42

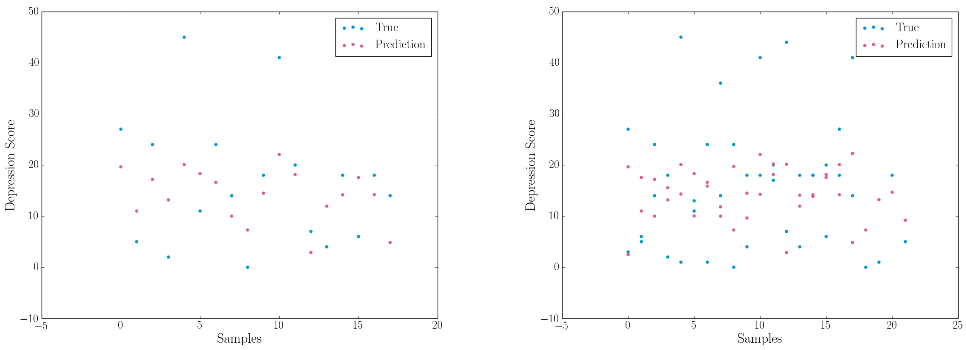


Figure 4.35: Best model (1) for the regression problem with the usage metrics' features in the regression problem using the Support Vector Regression. On the left, the train results and on the right, the test results.

Overall, we can see that the models are worse with this regressor, especially when tested with unseen data because the MSE when validating the models is similar to the other regressor in some models, but when tested, the MSE increases. The best model has an average distance between the predicted and real scores of around eight scores.

4.3.3 Image Features

First step, will be to search for the best parameters for our two regressors:

Table 4.91: Parameters for the Random Forest Regressor and SVR that were tested to find the optimal one to fit the image features in the regression problem.

Parameters	Values
n_estimators	1800
max_features	'auto'
max_depth	'None'
min_samples_split	2
min_samples_leaf	2
bootstrap	True

Parameters	Values
epsilon	0.1
C	10
gamma	1e-7
kernel	'linear'

After we have the optimal parameters, we are training 100 regressors with different random sets of training and validation dataset, and observe the MSE obtained in them by plotting the figure 4.36.

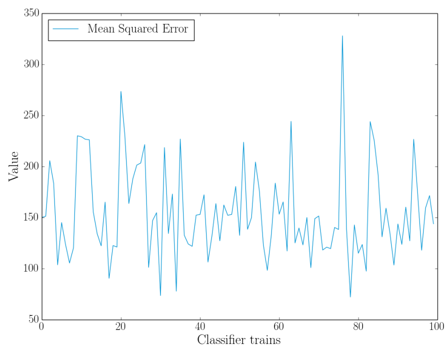


Figure 4.36: Cross-validation results of 100 trained Random Forest Regressor with the train dataset of image features in the regression problem.

Observing these MSEs, we can find some good models, although there are a lot of poor results. We are now testing the best ones.

Table 4.92: Best results for cross validation for image features in the regression problem with Random Forest Regressor, using random train and validation datasets from the initial train dataset to evaluate the models.

	Mean Squared Error
1	85.24
2	94.54
3	96.18
4	98.91

Table 4.93: Test results for the best models for the image features with in the regression problem Random Forest Regressor, using the test dataset to evaluate the models.

	Mean Squared Error
1	144.55
2	145.62
3	126.65
4	158.73

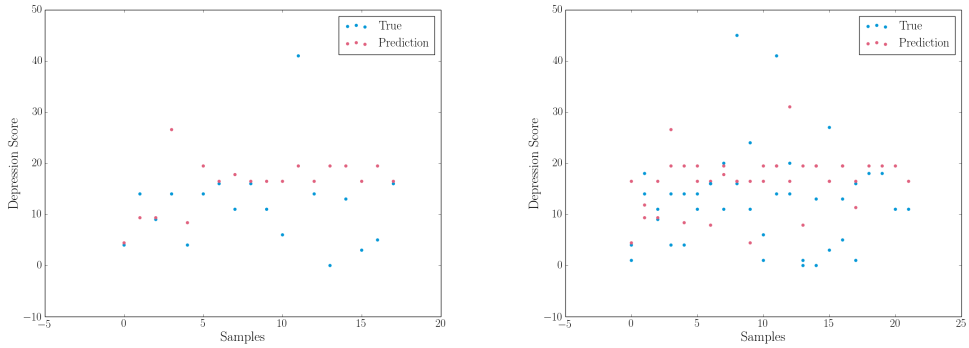


Figure 4.37: Best model (3) for the regression problem with the image features in the regression problem using the Random Forest Regression. On the left, the train results and on the right, the test results

When tested, these models increase their MSE compared with the validation values. Our best model (3) has an average of 8 scores difference between the predicted and the real values. The model is not optimal, but it predicts some of the scores correctly as seen in the figure 4.37.

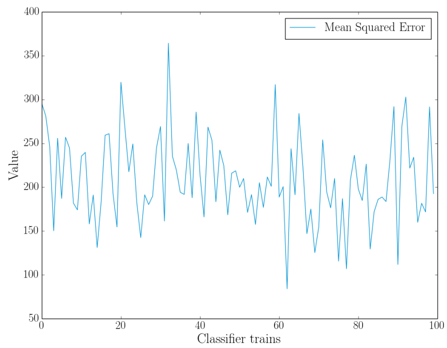


Figure 4.38: Cross-validation results of 100 trained Support Vector Regressions with the train dataset of image features in the regression problem.

The SVR, as in the previous groups, presents higher values of MSE overall for the 100 trained regressions, but we can find some with a much lower value, which we will test with our test dataset.

Table 4.94: Best results for cross validation for image features in the regression problem with Support Vector Regression, using random train and validation datasets from the initial train dataset to evaluate the models.

	Mean Squared Error
1	61.42
2	52.18
3	81.42
4	100.68

Table 4.95: Test results for the best models for the image features in the regression problem with Support Vector Regression, using the test dataset to evaluate the models.

	Mean Squared Error
1	116.56
2	105.48
3	114.17
4	105.84

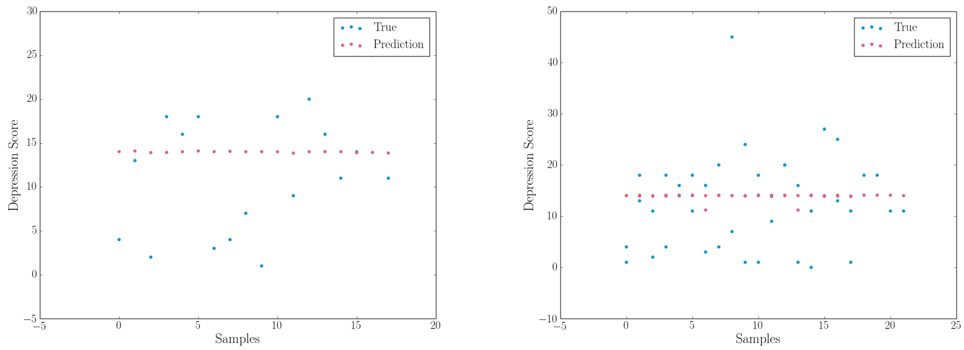


Figure 4.39: Best model (2) for the regression problem with the image features in the regression problem using the Support Vector Regression. On the left, the train results and on the right, the test results.

When analyzing the MSE with the tested data, we observe that the models present good results, but when analyzing the plotted predicted and real scores, we observe that the models fit a linear regression, so the results depend on the linearity of the scores in the datasets used to predict.

4.3.4 Text, Usage Metrics' and Image Features combined

Finally, we are combining the three datasets and attempt to improve the previous results from the groups separately.

Table 4.96: Parameters for the Random Forest Regressor and SVR that were tested to find the optimal one to fit the overall features in the regression problem.

Parameters	Values
n_estimators	800
max_features	'sqrt'
max_depth	50
min_samples_split	10
min_samples_leaf	4
bootstrap	True

Parameters	Values
epsilon	0.5
C	10
gamma	0.0001
kernel	'rbf'

Comparing to the previous groups, this one presents some lower values for the MSE of each of the trained regressor.

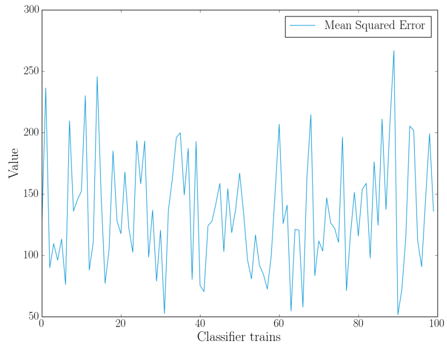


Figure 4.40: Cross-validation results of 100 trained Random Forest Regressors with the train dataset of the overall features in the regression problem.

Table 4.97: Best results for cross validation for the overall features in the regression problem with Random Forest Regressor, using random train and validation datasets from the initial train dataset to evaluate the models.

	Mean Squared Error
1	108.99
2	98.22
3	109.32
4	107.68

Table 4.98: Test results for the best models for the overall features in the regression problem with Random Forest Regressor, using the test dataset to evaluate the models.

	Mean Squared Error
1	79.37
2	85.14
3	84.66
4	79.91

For a better understanding of what these values mean, we plotted the actual labels and the predicted ones in figure 4.41.

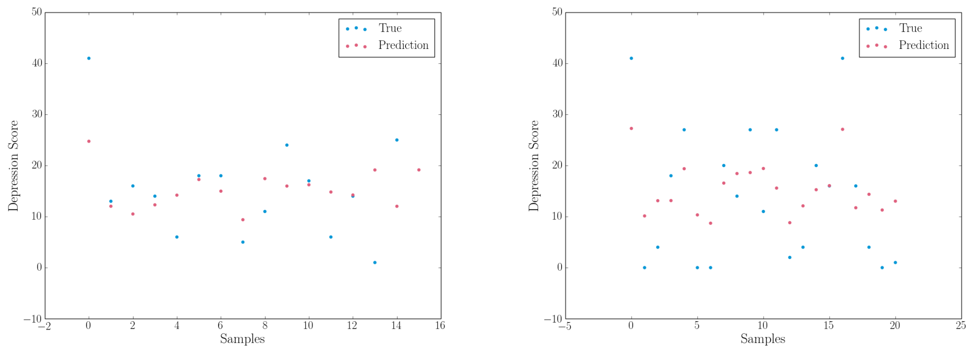


Figure 4.41: Best model (1) for the regression problem with the overall features in the regression problem using the Random Forest Regression. On the left, the train results and on the right, the test results

As we can observe, the predicted and real scores are close in a lot of the samples, which lead us to believe that if we got some more samples, the regressor could have much better results.

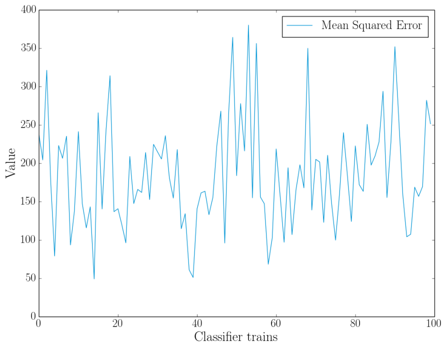


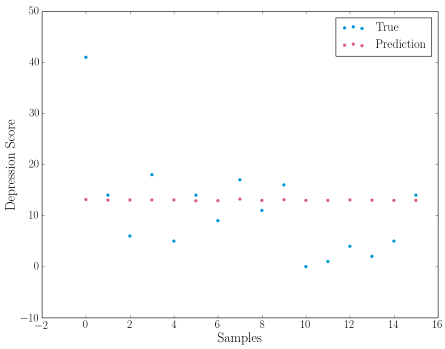
Figure 4.42: Cross-validation results of 100 trained Support Vector Regressions with the train dataset of the overall features in the regression problem.

Table 4.99: Best results for cross validation for the overall features in the regression problem with Support Vector Regressor, using random train and validation datasets from the initial train dataset to evaluate the models.

	Mean Squared Error
1	88.40
2	76.26
3	96.16
4	101.61

Table 4.100: Test results for the best models for the overall features in the regression problem with Support Vector Regressor, using the test dataset to evaluate the models.

	Mean Squared Error
1	165.54
2	166.62
3	164.47
4	166.46



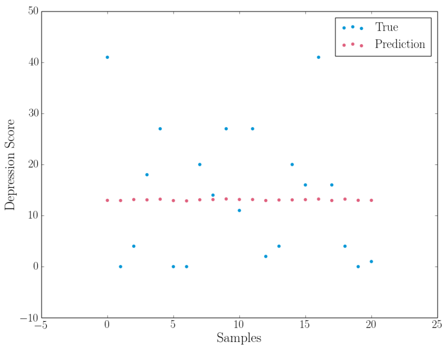


Figure 4.43: Best model (3) for the regression problem with the overall features in the regression problem using the Support Vector Regression. On the left, the train results and on the right, the test results.

When looking at the results from the SVR, we observe that we have the same problem as with the group of image features, the models do a linear regression that could fit some groups of data, but many times would not fit these type of data. So, these are not good models as we can also verify when testing them, obtaining much weaker results than in the validation.

Overall the regression problem presented some exciting results, giving us the desire to collect a more significant data sample to improve these models results.

4.4 Depression Prediction using a CNN

After analyzing the results obtained when only using the image features, we understood that the low-level features we were using were not suitable for predicting depression among user in social media. As we had some time before the delivery of this project, we decided to try to improve the results of image depression analysis by using convolutional neural networks. We chose to apply transfer learning to the pre-trained VGG16 network, since it is a well-known CNN for classifying images. We decided to only try to predict between depressive and non-depressive users, considering the amount of samples we had and the time we had.

For this problem, we started by processing all the images and creating the labels. After this, we shuffled the data and divided into train, validation and test dataset. We then loaded the VGG16 trained neural network, removed the last layer and added a layer suitable for our problem. The VGG16 has an output of 1000 different classes, and our problem has 2. The CNN was compiled, and we trained it with our training dataset, using the validation dataset for its validation. We trained the CNN with 20 and 100 epochs, but as we got similar results, we continue with only the 20 epochs so that it would take less time (2min/epoch). For evaluating the model, we test them with our unseen data (test dataset). In each epoch, we can see the values of the model training and validation loss, and training and validation accuracy, so at the end of the training, we plot two graphics, one with both accuracies during the different epochs and another one with both losses.

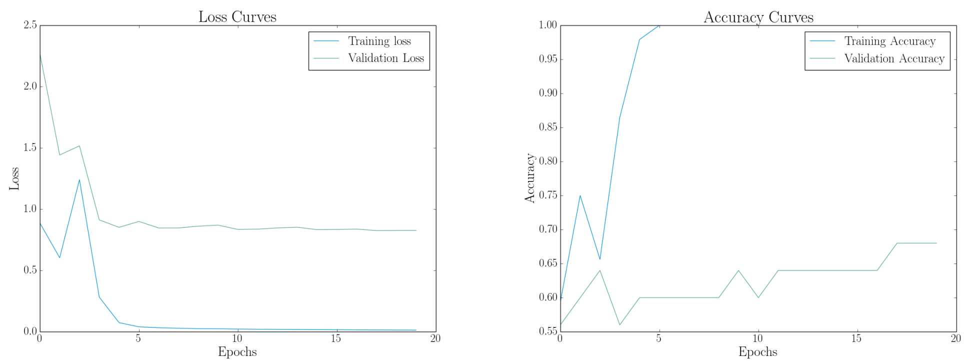


Figure 4.44: Loss and accuracy curves during the training of the Convulotional Neural Network.

We got an accuracy of 70.97% and a loss of 1.02%. It is essential to understand that we had few samples, only 152 images and we used 96 to train, 25 to validate and 31 to test.

To achieve better results, we tried to train the neural network during more epochs, but the results stayed similar. We also tried different optimizers and its parameters, being the sgd the one with best results (the parameters are described in the table 4.101).

Table 4.101: Parameters used in the sgd optimizer.

Parameter	Value
lr	0.01
decay	(default)
momentum	(default)
nesterov	(default)

5

Conclusion

In this project, we aimed to be able to predict if a social network user was going through depression. We divided it into four parts: prediction between a depressive and non-depressive user, prediction between the four states of depression (minimal, mild, moderate and severe), prediction of the exact depression scores and image depression analysis.

For this, we decided to collect user posts from three social networks (Facebook, Instagram, and Twitter) and evaluating these users with BDI-II, a depression tracking test. For Facebook, we needed to build an app that would be able to collect these posts. This app had to be approved by Facebook, and it was not accepted because of their Users' Privacy Policy. For Instagram we had a package that collected these data via their API, we only needed that the users allowed us to followed them via Instagram. For Twitter, it was similar to Facebook, but as only a few of our participants were using it, we did not use it.

Finally, we divided the post information from Instagram into three groups: image features, text features, and usage metrics' features.

In the three first approaches, we started by trying to predict with the separated features' groups. After analyzing the obtained results, we combined these three groups and make predictions using all the features. This led us to better results than in the first approach.

Our model got an accuracy of 90.5% when predicting between depressive and non-depressive users and 76.2% when classifying in one of the four depression states, both with the Random Forest Classifier. For exact depression score prediction we obtained a model with a mean squared error of 79.37. For the image depression analysis prediction we applied transfer learning to the pre-trained VGG16 and got an accuracy of 70.97%.

One of the biggest challenges in this project was the amount of data collected, which was too small considering that the posts can be very variable from person to person and the more samples, the easier it is to generalize. To obtain better results, it would also be important to test other features to be extracted, since some of them did not help us to distinguish the different classes, especially in the image features groups.

Regarding the image features group, we concluded that low-level features are not useful for predicting depression, as already concluded in research for sentiment analysis. Our second approach, deep learning, gave us much better results considering the number of samples we had.

The main conclusion we take from this project, is that a multimodal model has better results than unimodal models when predicting depression. Despite all the challenges that we came across during this project, there were always solutions found for the next step, making it possible to achieve its purpose.

5.1 Future Work

The most important future work with the most significant impact would be to collect a larger data sample, so the models could generalize more and achieve better results with any unseen data. It could also be essential to look for more features for the text and usage metrics' features, especially the usage metrics' since they were the best on predicting.

If better results could be achieved these models could be used by social networks to track users with depression and forward them to help institutions so they could get medical help. This could also be used to monitorize clinical patients by their psychologist and psychiatrist or help on depression diagnose.

As seen in the approach using a convolutional neural network, the results can be quite interesting. A good future work option would be to, instead of extracting features manually, to use the raw data directly in a neural network designed for the type of data we are classifying.

Finally, all age ranges could be covered, since there is also depression in younger and older users, than the ones studied.

Bibliography

[1] Depression and Other Common Mental Disorders: Global Health Estimates. Geneva: World Health Organization; 2017. Licence: CC BY-NC-SA 3.0 IGO.

[2] Depression (PDF). National Institute of Mental Health (NIMH). Archived from the original on 27 July 2011. Retrieved 15 May 2018.

[3] World Health Organization. 2018. *Depression*. [ONLINE] Available at: http://www.who.int/mediacentre/factsheets/fs369/en/. [Accessed 15 May 2018]

[4] National Trends in the Prevalence and Treatment of Depression in Adolescents and Young Adults Ramin Mojtabai, Mark Olfson, Beth Han Pediatrics Nov 2016, e20161878; DOI: 10.1542/peds.2016-1878

[5] Eu Luto Contra a Depressão. 2015. *Depressão*. [ONLINE] Available at: http://eulutocontraadepressao.eutimia.pt/depressao-em-portugal/. [Accessed 15 May 2018]

[6] Pedro Prata. 2017. *Uso excessivo das redes sociais pode ser indicativo de depressão*. [ONLINE] Available at: http://emais.estadao.com.br/noticias/comportamento,uso-excessivo-das-redes-sociais-pode-ser-indicativo-de-depressao,70001777894. [Accessed 15 May 2018]

[7] World Health Organization, Geneva. 2001. *Mental disorders affect one in four people*. [ONLINE] Available at: http://www.who.int/whr/2001/media_centre/press_release/en/. [Accessed 15 May 2018]

[8] American Psychiatric Association (2013), Diagnostic and Statistical Manual of Mental Disorders (5th ed.), Arlington: American Psychiatric Publishing, pp. 160–168, ISBN 978-0-89042-555-8, archived from the original on 31 July 2016, retrieved 15 May 2018

[9] Patton LL (2015). The ADA Practical Guide to Patients with Medical Conditions (2 ed.). John Wiley & Sons. p. 339. ISBN 9781118929285.

[10] Zimmerman M, Chelminski I, Posternak M (September 2004). "A review of studies of the Hamilton depression rating scale in healthy controls: implications for the definition of remission in treatment studies of depression". The Journal of Nervous and Mental Disease. 192 (9): 595–601. doi:10.1097/01.nmd.0000138226.22761.39. PMID 15348975.

[11] McPherson A, Martin CR (February 2010). "A narrative review of the Beck Depression Inventory (BDI) and implications for its use in an alcohol-dependent population". Journal of Psychiatric and Mental Health Nursing. 17 (1): 19–30. doi:10.1111/j.1365-2850.2009.01469.x. PMID 20100303.

[12] Osman A, Bagge CL, Gutierrez PM, Konick LC, Kopper BA, Barrios FX (December 2001). "The Suicidal Behaviors Questionnaire-Revised (SBQ-R): validation with clinical and nonclinical samples". Assessment. 8 (4): 443–54. doi:10.1177/107319110100800409. PMID 11785588.

[13] Sharp LK, Lipsky MS (September 2002). "Screening for depression across the lifespan: a review of measures for use in primary care settings". American Family Physician. 66 (6): 1001–8. PMID 12358212.

[14] Health. 2017. 12 Types of Depression, and What You Need to Know About Each. [ONLINE] Available at: http://www.health.com/depression/types-of-depression. [Accessed 6 June 2018].

[15] Turner, R. J. & Avison, W. R. (2003). Status variations in stress exposure among young adults: Implications for the interpretation of prior research. Journal of Health and Social Behavior, 44, 488-505.

[16] Graber, J. A. (2004). Internalizing problems during adolescence. In R. M. Lerner & L. Steinberg (Eds.), Handbook of Adolescent Psychology (Second ed.). Hoboken, NJ: John Wiley and Sons.

[17] Hankin, B., Abramson, L., Moffitt, T., Silva, P., McGee, R., & Angell, K. (1998). Development of depression from preadolescence to young adulthood: Emerging gender differences in a 10-year longitudinal study. Journal of Abnormal Psychology, 107(1) 128-140.

[18] Wiesner, M., Windle, M., & Freeman, A. (2005). Work stress, substance use, and depression among young adult workers: An examination of main and moderator effect model. Journal of Occupational Health Psychology, 10(2) 83-96.

[19] Zimmerman, F. J., Christakis, D.A., & Stoep, A.V. (2004). Tinker, tailor, soldier, patient: Work attributes and depression disparities among young adults. Social Science & Medicine, 58(10), 1889-1901.

[20] Child Trends Databank. (2015). Young adult depression. Available at: https://www.childtrends.org/?indicators=young-adult-depression[Accessed 6 June 2018]

[21] Horwitz, A. V., McLaughlin, J., & Raskin White, H. (1998). How the negative and positive aspects of partner relationships affect the mental health of young married people. Journal of Health and Social Behavior, 39(2), 124-136.

[22] Haarasilta, L. M., Marttunen, M. J., and Kaprio, J. (2004). Correlates of depression in a representative nationwide sample of adolescents (15-19 years) and young adults (20-24 years). European Journal of Public Health, 14(3), 280-285.

[23] Goodman, E., and Capitman, J. (2000). Depressive symptoms and cigarette smoking among teens. Pediatrics, 106, 748-755.

[24] Turner, H.A. & Butler, M.J. (2003). Direct and indirect effects of childhood adversity on depressive symptoms in young adults. Journal of Youth and Adolescence, 32(2), 89-103.

[25] Miller, D. K., Constance, H. L., & Brennan, P. A. (2007). Health outcomes related to early adolescent depression. Journal of Adolescent Health, 41(3), 256-262.

[26] Jones, D. J., Forehand, R., & Beach, S. R. H. (2000). Maternal and paternal parenting during adolescence: Forecasting early adult psychosocial adjustment. Adolescence, 35(139), 513-530.

[27] Nunnally, J. C. (1978). Assessment of Reliability. In: Psychometric Theory (2nd ed.). New York: McGraw-Hill.

[28] Richter, P., Werner, J., Heerlein, A., Kraus, A., Sauer, H. (1998). On the Validity of the Beck Depression Inventory - A Review. Psychopathology, 31,160-168.

[29] Beck AT, Ward CH, Mendelson M, Mock J, Erbaugh J (June 1961). "An inventory for measuring depression".

[30] Moran PW, Lambert MJ (1983). "A review of current assessment tools for monitoring changes in depression". In Lambert MS, Christensen ER, DeJulio S. The Assessment of Psychotherapy Outcomes. New York: Wiley.

[31] Beck AT, Steer RA, Ball R, Ranieri W (December 1996). "Comparison of Beck Depression Inventories -IA and -II in psychiatric outpatients". Journal of Personality Assessment.

[32] Ambrosini PJ, Metz C, Bianchi MD, Rabinovich H, Undie A (January 1991). "Concurrent validity and psychometric properties of the Beck Depression Inventory in outpatient adolescents". Journal of the American Academy of Child and Adolescent Psychiatry.

[33] Beck AT, Steer RA and Brown GK (1996) "Manual for the Beck Depression Inventory-II". San Antonio, TX: Psychological Corporation

[34] Stephanie Stiavetti. 2010. *What Is Electronic Social Networking?*. [ONLINE] Available at: https://socialnetworking.lovetoknow.com/What_is_Electronic_Social_Networking. [Accessed 13 June 2018].

[35] Boyd, D. M. and Ellison, N. B. (2007), Social Network Sites: Definition, History, and Scholarship. Journal of Computer-Mediated Communication, 13: 210-230. doi:10.1111/j.1083-6101.2007.00393.x

[36] Amichai-Hamburger, Y. and Hayat, T. (2017). Social Networking. In The International Encyclopedia of Media Effects (eds P. Rössler, C. A. Hoffner and L. Zoonen). doi:10.1002/9781118783764.wbieme0170

[37] Richard Arcega. 2016. Instagram-scraper. [ONLINE] Available at: https://github.com/rarcega/instagram-scraper. [Accessed 3 July 2018].

[38] MathWorks. 2018. Machine Learning. [ONLINE] Available at: https://www.mathworks.com/discovery/machine-learning.html. [Accessed 10 August 2018].

[39] Towards Data Science, James Le. 2018. A Tour of The Top 10 Algorithms for Machine Learning Newbies. [ONLINE] Available at: https://towardsdatascience.com/a-tour-of-the-top-10-algorithms-for-machine-learning-newbies-dde4edffae11. [Accessed 10 August 2018].

[40] Jason Brownlee. 2016. Supervised and Unsupervised Machine Learning Algorithms. [ONLINE] Available at: https://machinelearningmastery.com/supervised-and-unsupervised-machine-learning-algorithms/. [Accessed 10 August 2018].

[41] IBM, M. Tim Jones. 2017. Models for machine learning. [ONLINE] Available at: https://www.ibm.com/developerworks/library/cc-models-machine-learning/index.html. [Accessed 10 August 2018].

[42] scikit learn. 2017. sklearn.model_selection.RandomizedSearchCV. [ONLINE] Available at: http://scikit-learn.org/stable/modules/generated/sklearn.model_s election.RandomizedSearchCV.html. [Accessed 10 August 2018].

[43] scikit learn. 2017. sklearn.model_selection.GridSearchCV. [ONLINE] Available at: http://scikit-learn.org/stable/modules/generated/sklearn.model_selecti on.GridSearchCV.html. [Accessed 10 August 2018].

[44] scikit learn. 2017. sklearn.model_selection.GridSearchCV. [ONLINE] Available at: http://scikit-learn.org/stable/modules/generated/sklearn.model_selecti on.GridSearchCV.html. [Accessed 10 August 2018].

[45] Laerd statistics. 2018. One-way ANOVA. [ONLINE] Available at: https://statistics.laerd.com/statistical-guides/one-way-anova-statistical-guide.php. [Accessed 13 August 2018].

[46] The Editors of Encyclopaedia Britannica. 2018. Student's t-test. [ONLINE] Available at: https://www.britannica.com/science/Students-t-test. [Accessed 13 August 2018].

[47] Savan Patel. 2017. Chapter 2: SVM (Support Vector Machine)-Theory. [ONLINE] Available at: https://medium.com/machine-learning-101/chapter-2-svm-support-vector-machine-theory-f0812effc72. [Accessed 13 August 2018].

[48] Savan Patel. 2017. Chapter 5: Random Forest Classifier. [ONLINE] Available at: https://medium.com/machine-learning-101/chapter-5-random-forest-classifier-56dc7425c3e1. [Accessed 15 August 2018].

[49] Niklas Donges. 2018. The Random Forest Algorithm. [ONLINE] Available at: https://towardsdatascience.com/the-random-forest-algorithm-d457d499ffcd. [Accessed 15 August 2018].

[50] Savan Patel. 2017. Chapter 3 : Decision Tree Classifier-Theory. [ONLINE] Available at: https://medium.com/machine-learning-101/chapter-5-random-forest-classifier-56dc7425c3e1. [Accessed 15 August 2018].

[51] Adi Bronshtein. 2017. A Quick Introduction to K-Nearest Neighbors Algorithm. [ONLINE] Available at: https://medium.com/@adi.bronshtein/a-quick-

introduction-to-k-nearest-neighbors-algorithm-62214cea29c7. [Accessed 15 August 2018].

[52] Mohtadi Ben Fraj. 2017. In Depth: Parameter tuning for KNN. [ONLINE] Available at: https://medium.com/@mohtedibf/in-depth-parameter-tuning-for-knn-4c0de485baf6. [Accessed 15 August 2018].

[53] scikit learn. 2017. sklearn.neighbors.KNeighborsClassifier. [ONLINE] Available at: http://scikit-learn.org/stable/modules/generated/sklearn.neighbors.KN eighborsClassifier.html#. [Accessed 15 August 2018].

[54] Dr. Saed Sayad. 2018. Support Vector Machine - Regression (SVR). [ONLINE] Available at: http://www.saedsayad.com/support_vector_machine_reg.htm. [Accessed 15 August 2018].

[55] Gamon, Michael & de Choudhury, Munmun & Counts, Scott & Horvitz, Eric. (2013). Predicting Depression via Social Media. Association for the Advancement of Artificial Intelligence.

[56] De Choudhury, Munmun & Counts, Scott & Horvitz, Eric. (2013). Social media as a measurement tool of depression in populations. Proceedings of the 5th Annual ACM web science conference websci2013. 47-56. 10.1145/2464464.2464480.

[57] Batra, S., & Rao, D.S. (2010). Entity Based Sentiment Analysis on Twitter.

[58] Rosenthal, S., Farra, N., & Nakov, P. (2017). SemEval-2017 Task 4: Sentiment Analysis in Twitter. SemEval@ACL.

[59] Kumar, A & Sebastian, TM. (2012). Sentiment Analysis on Twitter.

[60] Tausczik, YR & Pennebaker, JW 2010, 'The psychological meaning of words: LIWC and computerized text analysis methods' Journal of Language and Social Psychology, vol. 29, no. 1, pp. 24-54. DOI: 10.1177/0261927X09351676

[61] Krieck, M., Dreesman, J., Otrusina, L., and Denecke,K. (2011). A new age of public health: Identifying disease outbreaks by analyzing tweets

[62] Park, Myoungouk & Cha, Chiyoung & Cha, M. (2012). Depressive moods of users portrayed in twitter. Proceedings of the ACM SIGKDD Workshop On Healthcare Informatics (HI-KDD)2012. 1-8.

[63] Yilin Wang, Suhang Wang, Jiliang Tang, Huan Liu, and Baoxin Li. 2015. Unsupervised sentiment analysis for social media images. In Proceedings of the 24th

International Conference on Artificial Intelligence (IJCAI'15), Qiang Yang and Michael Wooldridge (Eds.). AAAI Press 2378-2379.

[64] Jia, Jia & Wu, Sen & Wang, Xiaohui & Hu, Peiyun & Cai, Lianhong & Tang, Jie. (2012). Can We Understand van Gogh's Mood? Learning to Infer Affects from Images in Social Networks. ACM International Conference on Multimedia. 10.1145/2393347.2396330.

[65] Damian Borth, Rongrong Ji, Tao Chen, Thomas Breuel, and Shih-Fu Chang. 2013. Large-scale visual sentiment ontology and detectors using adjective noun pairs. In Proceedings of the 21st ACM international conference on Multimedia (MM '13). ACM, New York, NY, USA, 223-232. DOI: https://doi.org/10.1145/2502081.2502282

[66] Quanzeng You, Jiebo Luo, Hailin Jin, and Jianchao Yang. 2015. Robust image sentiment analysis using progressively trained and domain transferred deep networks. In Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence (AAAI'15). AAAI Press 381-388.

[67] Jin, X.; Gallagher, A.; Cao, L.; Luo, J.; and Han, J. 2010. The wisdom of social multimedia: using flickr for prediction and forecast. In ACM MM, 1235–1244. ACM.

[68] Jenn Chen. 2018. What Is an API & Why Does It Matter?. [ONLINE] Available at: https://sproutsocial.com/insights/what-is-an-api/. [Accessed 19 August 2018].

[69] Jason Brownlee. 2016. What is a Confusion Matrix in Machine Learning. [ONLINE] Available at: https://machinelearningmastery.com/confusion-matrix-machine-learning/. [Accessed 28 August 2018].

[70] Baratloo, A., Hosseini, M., Negida, A., & El Ashal, G. (2015). Part 1: Simple Definition and Calculation of Accuracy, Sensitivity and Specificity. Emergency, 3(2), 48–49.

[71] TutorVista. 2018. Mean Squared Error. [ONLINE] Available at: https://math.tutorvista.com/statistics/mean-squared-error.html. [Accessed 28 August 2018].

[72] University of Mississippi. Neural Networks. [ONLINE] Available at: http://faculty.bus.olemiss.edu/breithel/final%20backup%20of%20bus620%20summer%202000%20from%20mba%20server/Scott_Strech/week%202%20neural%20networks/neural_networks.htm. [Accessed 6 September 2018]

[73] Rameshkumar, G. P., & Samundeswari, S. (2014). Neural Network, Artificial Neural Network (ANN) and Biological Neural Network (BNN) in Soft Computing. Volume, 30(3), 3.

[74] Aurlien Gron. 2017. Hands-On Machine Learning with Scikit-Learn and Tensorflow: Concepts, Tools, and Techniques to Build Intelligent Systems (1st ed.). O'Reilly Media, Inc.

[75] Simplilearn. 2018. Convolutional Neural Network Tutorial (CNN) — How CNN Works — Deep Learning Tutorial — Simplilearn. [ONLINE] Retrieved from https://www.youtube.com/watch?time_continue=3833&v=Jy9-aGMB_TE

[76] Prabhu. 2018. Understanding of Convolutional Neural Network (CNN) — Deep Learning. [ONLINE] Available at: https://medium.com/@RaghavPrabhu/understanding-of-convolutional-neural-network-cnn-deep-learning-99760835f148. [Accessed 6 September 2018].

[77] Asheesh kumar singhalbhu. 2017. Transfer learning using VGG-16 model. [ONLINE] Available at: https://medium.com/@asheesh12feb/transfer-learning-using-vgg-16-model-2f9b5944b8b2. [Accessed 6 September 2018].

[78] Niklas Donges. 2018. Transfer Learning. [ONLINE] Available at: https://towardsdatascience.com/transfer-learning-946518f95666. [Accessed 12 September 2018].

[79] Shahariar Rabby. 2017. LeNet 5, AlexNet, VGG -16 from Deeplearning.ai. [ONLINE] Available at: https://medium.com/@shahariarrabby/lenet-5-alexnet-vgg-16-from-deeplearning-ai-2a4fa5f26344. [Accessed 12 September 2018].

[80] Davi Frossard. 2016. VGG in TensorFlow. [ONLINE] Available at: http://www.cs.toronto.edu/ frossard/post/vgg16/. [Accessed 12 September 2018].

[81] You, Q., Luo, J., Jin, H., & Yang, J. (2015, January). Robust Image Sentiment Analysis Using Progressively Trained and Domain Transferred Deep Networks. In AAAI (pp. 381-388).

[82] Daphne Cornelisse. 2018. An intuitive guide to Convolutional Neural Networks. [ONLINE] Available at: https://medium.freecodecamp.org/an-intuitive-guide-to-convolutional-neural-networks-260c2de0a050. [Accessed 13 September 2018].

[83] Brent Barnhart. 2017. 20 Social Media Ideas to Keep Your Brand's Feed Fresh. [ONLINE] Available at: https://sproutsocial.com/insights/social-media-ideas/. [Accessed 13 September 2018].

Appendices

A

BDI-II

O questionário seguinte é constituído por vários grupos de afirmações. Em cada grupo escolha UMA ÚNICA afirmação, a que melhor descreve a forma como se tem sentido NAS DUAS ÚLTIMAS SEMANAS.

1. Tristeza
   - Não me sinto triste
   - Ando triste muitas vezes
   - Sinto-me sempre triste
   - Estou tão triste ou infeliz que já não o suporto
2. Pessimismo
   - Não me sinto desencorajada em relação ao futuro
   - Sinto-me mais desencorajada em relação ao futuro do que costumava
   - Já não espero que os meus problemas se resolvam
   - Não tenho qualquer esperança no futuro e acho que tudo só pode piorar
3. Fracassos Passados
   - Não me considero uma falhada
   - Fracassei mais vezes do que deveria
   - Quando considero o meu passado, o que noto é uma quantidade de fracassos
   - Sinto-me completamente falhada como pessoa
4. Perda de Prazer
   - Tenho tanto prazer como costumava ter com as coisas que eu gosto
   - Eu não gosto tanto das coisas como costumava

- Tenho pouco prazer com as coisas que eu costumava gostar
- Não obtenho qualquer prazer das coisas que eu costumava gostar

5. Sentimentos de Culpa
    - Não me sinto particularmente culpada
    - Sinto-me culpada por muitas coisas que fiz ou deveria ter feito
    - Sinto-me bastante culpada a maioria das vezes
    - Sinto-me culpada durante o tempo todo
6. Sentimentos de Punição
    - Não sinto que estou a ser castigada
    - Sinto que posso ser castigada
    - Espero vir a ser castigada
    - Sinto que estou a ser castigada
7. Auto-depreciação
    - Aquilo que acho de mim é o que sempre achei
    - Perdi a confiança em mim própria
    - Estou desapontada comigo mesma
    - Não gosto de mim
8. Auto-criticismo
    - Não me culpo ou critico mais do que costumava
    - Critico-me mais do que costumava
    - Critico-me por todas as minhas falhas
    - Culpo-me por tudo o que de mal me acontece
9. Pensamentos ou Desejos Suicidas
    - Não tenho qualquer ideia de me matar
    - Tenho ideias de me matar mas não as levarei a cabo
    - Gostaria de me matar
    - Matar-me-ia se tivesse oportunidade

10. Choro
    - Não choro mais do que costumava
    - Choro mais do que costumava
    - Choro por tudo e por nada
    - Apetece-me chorar, mas já não consigo
11. Agitação
    - Não me sinto mais inquieta que o normal
    - Sinto-me mais inquieta que o habitual
    - Estou tão inquieta ou agitada que é difícil parar quieta
    - Estou tão inquieta ou agitada que tenho que me manter em movimento ou a fazer alguma coisa
12. Perda de interesse
    - Não perdi o interesse nas outras pessoas ou nas minhas actividades
    - Estou menos interessado pelas coisas e pelas outras pessoas do que antes
    - Perdi a maioria do meu interesse nas coisas e nas outras pessoas
    - É difícil interessar-me por qualquer coisa que seja
13. Indecisão
    - Tomo decisões como sempre fiz
    - Acho mais difícil tomar decisões do que o habitual
    - Tenho muitas mais dificuldades em tomar decisões do que antigamente
    - Sinto-me incapaz de tomar qualquer decisão
14. Sentimentos de inutilidade
    - Não me considero uma incapaz/inútil
    - Não me considero tão válida e útil como costumava
    - Sinto-me mais inútil, em relação às outras pessoas
    - Sinto-me completamente inútil
15. Perda de energia

- Tenho a mesma energia de sempre
- Sinto-me com menos energia do que o habitual
- Não me sinto com energia para muitas coisas
- Não me sinto com energia para nada

16. Alterações no Padrão de Sono nas duas últimas semanas
    - Não notei qualquer mudança no meu sono
    - Durmo um pouco mais do que o habitual
    - Durmo um pouco menos do que o habitual
    - Durmo muito mais do que o habitual
    - Durmo muito menos do que o habitual
    - Durmo a maioria do tempo durante o dia
    - Acordo cerca de 1-2 horas mais cedo que é costume e não consigo voltar a dormir
17. Irritabilidade
    - Não estou mais irritável que o normal
    - Estou mais irritável que o habitual
    - Estou mais irritável que o normal
    - Estou irritável o tempo todo
18. Alterações no Apetite
    - Não notei qualquer alteração no meu apetite
    - Tenho um pouco menos de apetite do que o habitual
    - Tenho um pouco mais de apetite do que o habitual
    - O meu apetite é muito menor que o normal
    - O meu apetite é muito maior que o normal
    - Perdi por completo o apetite
    - Anseio por comida o tempo todo
19. Dificuldades de Concentração

- Concentro-me tão bem como antes
- Não me consigo concentrar tão bem como antes
- É difícil manter as minhas ideias em qualquer coisa por muito tempo
- Acho que não consigo concentrar-me em nada

20. Cansaço ou Fadiga

- Não me sinto mais cansada/fatigada que o habitual
- Canso-me mais facilmente que o costume
- Estou demasiado cansada ou fatigada para fazer uma série de coisas que costumava fazer
- Estou demasiado cansada ou fatigada para fazer a maioria das coisas que costumava fazer

21. Perda de Interesse Sexual

- Não notei qualquer mudança recente no meu interesse pela vida sexual
- Encontro-me menos interessada pela vida sexual do que costumava estar
- Actualmente sinto-me menos interessada pela vida sexual
- Perdi completamente o interesse que tinha pela vida sexual

B

Users forms

B.1 First form

Projeto de Tese - Identificação de Estados Depressivos

Actualmente, as doenças mentais surgem como a principal causa de incapacidade a nível mundial. Estimativas da OMS indicam o número de casos de depressão na ordem das centenas de milhões, no entanto, os mecanismos de identificação, acompanhamento e suporte desta condição patológica ainda deixam a desejar, mesmo em países desenvolvidos, apoiando-se principalmente em diagnóstico por auto-avaliação do paciente ou avaliações subjectivas por terceiros.

Para melhorar estas condições, pretende-se desenvolver um sistema de classificação, end-to-end, capaz de recolher e processar dados de redes sociais e associar a cada utilizador um score que reflicta a probabilidade de este estar a passar ou ir entrar num episódio de depressão.

Para ser possível testar este sistema de classificação, será necessário um conjunto de dados de estudantes do ensino universitário ou jovens trabalhadores.

Caso aceites participar neste projeto:
Estarás a dar autorização de recolha e processamento de informação das suas redes sociais
Será garantida completa confidencialidade quanto à informação recolhida
Estarás a dar o teu consentimento informado para a utilização destes dados para fins de investigação

Para participares tens de ter mais de 18 anos. Necessitamos de grandes amostras e as respostas não serão analisadas individualmente.

*Obrigatório

Aceitas os termos e condições (acima descritos)? *

- ◯ Sim
- ◯ Não

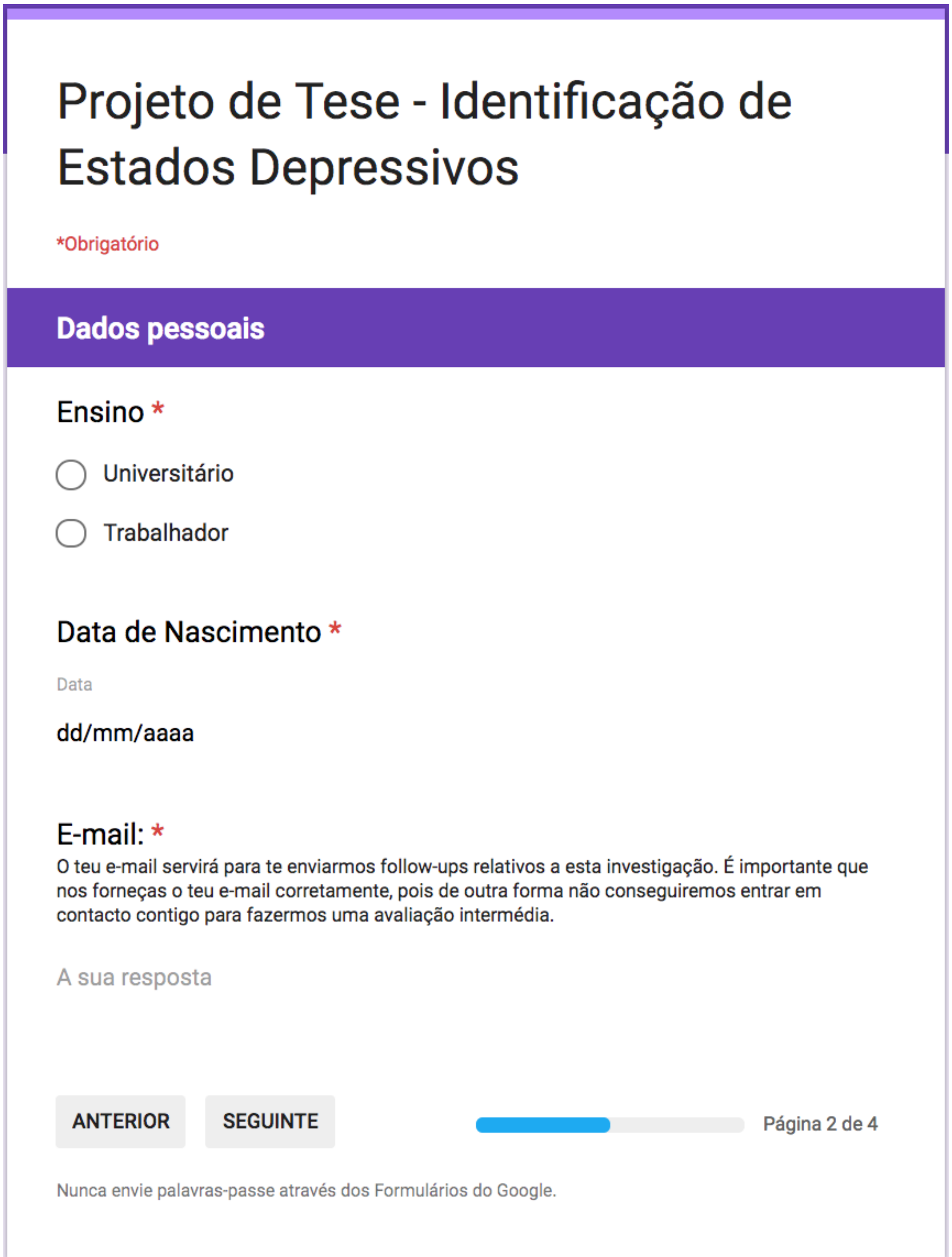
Aceitas responder honestamente às perguntas que te forem colocadas? *

- ◯ Sim
- ◯ Não

SEGUINTE

Página 1 de 4

Nunca envie palavras-passe através dos Formulários do Google.



Projeto de Tese - Identificação de Estados Depressivos

*Obrigatório

Dados pessoais

Ensino *

- Universitário
- Trabalhador

Data de Nascimento *

Data

dd/mm/aaaa

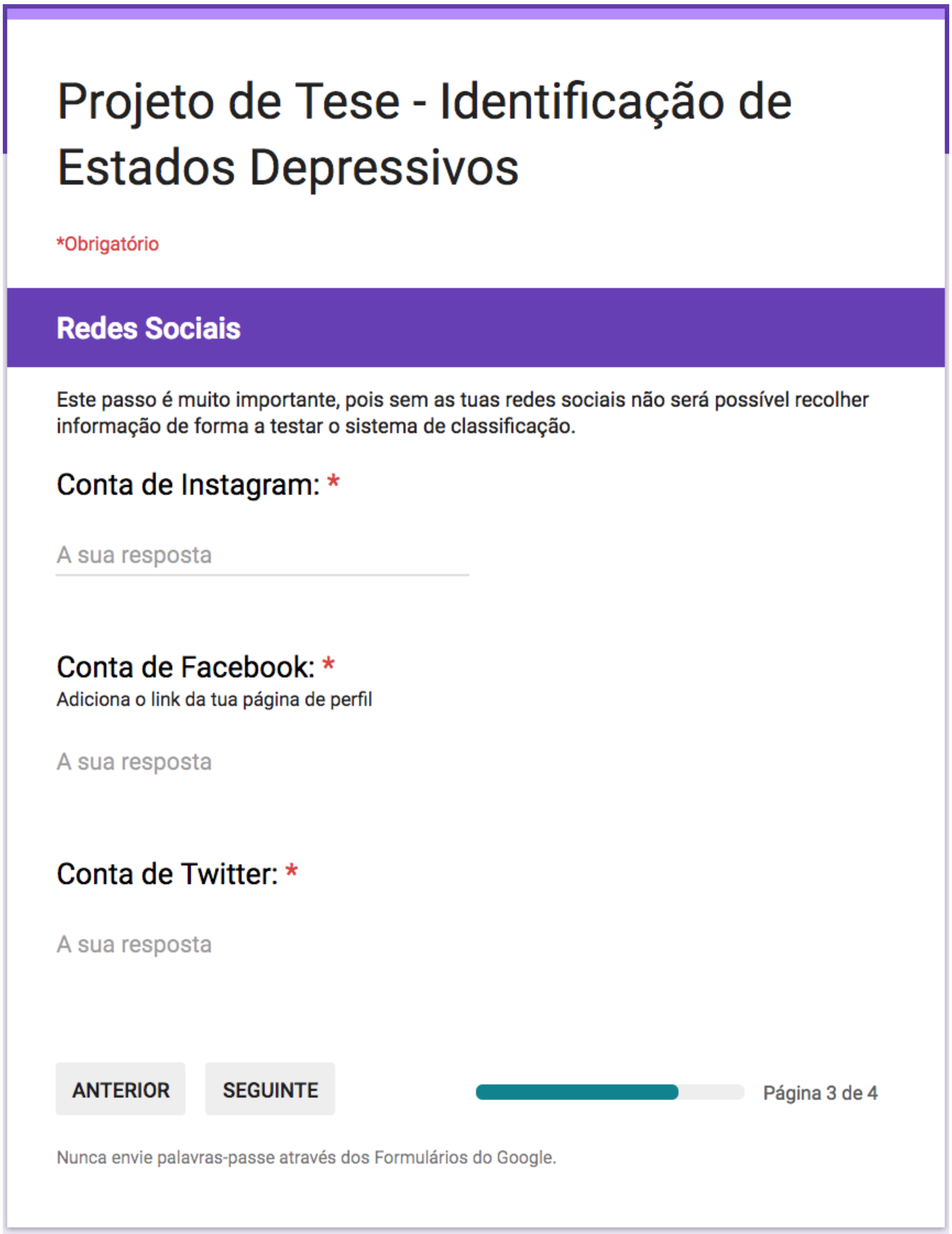
E-mail: *

O teu e-mail servirá para te enviarmos follow-ups relativos a esta investigação. É importante que nos forneças o teu e-mail corretamente, pois de outra forma não conseguiremos entrar em contacto contigo para fazermos uma avaliação intermédia.

A sua resposta

ANTERIOR SEGUINTE Página 2 de 4

Nunca envie palavras-passe através dos Formulários do Google.



Projeto de Tese - Identificação de Estados Depressivos

*Obrigatório

Redes Sociais

Este passo é muito importante, pois sem as tuas redes sociais não será possível recolher informação de forma a testar o sistema de classificação.

Conta de Instagram: *

A sua resposta

Conta de Facebook: *

Adiciona o link da tua página de perfil

A sua resposta

Conta de Twitter: *

A sua resposta

ANTERIOR SEGUINTE Página 3 de 4

Nunca envie palavras-passe através dos Formulários do Google.

Projeto de Tese - Identificação de Estados Depressivos

*Obrigatório

Questionário de avaliação do estado depressivo

O questionário seguinte é constituído por vários grupos de afirmações. Em cada grupo escolhe UMA ÚNICA afirmação, a que melhor descreve a forma como te tens sentido NAS DUAS ÚLTIMAS SEMANAS. São um total de 21 perguntas.

Lembra-te que sentires estas diversas emoções uma a duas vezes nas últimas duas semanas é normal no dia-a-dia de uma pessoa e não é relevante para o estudo de estado depressivo

.

1. Tristeza *

- ◯ Não me sinto triste
- ◯ Ando triste muitas vezes
- ◯ Sinto-me sempre triste
- ◯ Estou tão triste ou infeliz que já não o suporto

2. Pessimismo *

• • •

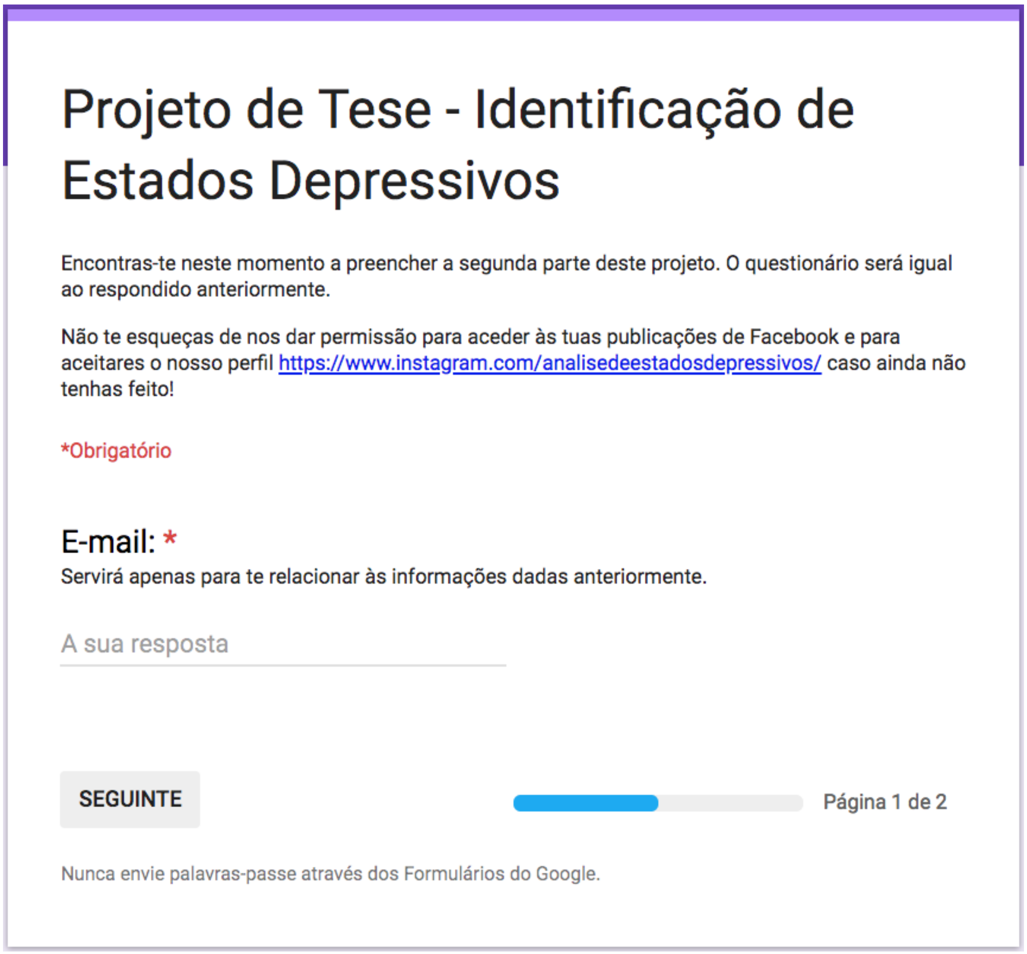
21. Perda de interesse sexual *

- ◯ Não notei qualquer mudança recente no meu interesse pela vida sexual
- ◯ Encontro-me menos interessada/o pela vida sexual do que costumava estar
- ◯ Actualmente sinto-me menos interessada/o pela vida sexual
- ◯ Perdi completamente o interesse que tinha pela vida sexual

ANTERIOR SUBMETER Página 4 de 4

Nunca envie palavras-passe através dos Formulários do Google.

B.2 Second form



Projeto de Tese - Identificação de Estados Depressivos

Encontras-te neste momento a preencher a segunda parte deste projeto. O questionário será igual ao respondido anteriormente.

Não te esqueças de nos dar permissão para aceder às tuas publicações de Facebook e para aceitares o nosso perfil https://www.instagram.com/analisedeestadosdepressivos/ caso ainda não tenhas feito!

*Obrigatório

E-mail: *
Servirá apenas para te relacionar às informações dadas anteriormente.

A sua resposta

SEGUINTE

Página 1 de 2

Nunca envie palavras-passe através dos Formulários do Google.

Projeto de Tese - Identificação de Estados Depressivos

*Obrigatório

Questionário de avaliação do estado depressivo

O questionário seguinte é constituído por vários grupos de afirmações. Em cada grupo escolhe UMA ÚNICA afirmação, a que melhor descreve a forma como te tens sentido NAS DUAS ÚLTIMAS SEMANAS. São um total de 21 perguntas.

Lembra-te que sentires estas diversas emoções uma a duas vezes nas últimas duas semanas é normal no dia-a-dia de uma pessoa e não é relevante para o estudo de estado depressivo
.

1. Tristeza *

- ◯ Não me sinto triste
- ◯ Ando triste muitas vezes
- ◯ Sinto-me sempre triste
- ◯ Estou tão triste ou infeliz que já não o suporto

2. Pessimismo *

• • •

21. Perda de interesse sexual *

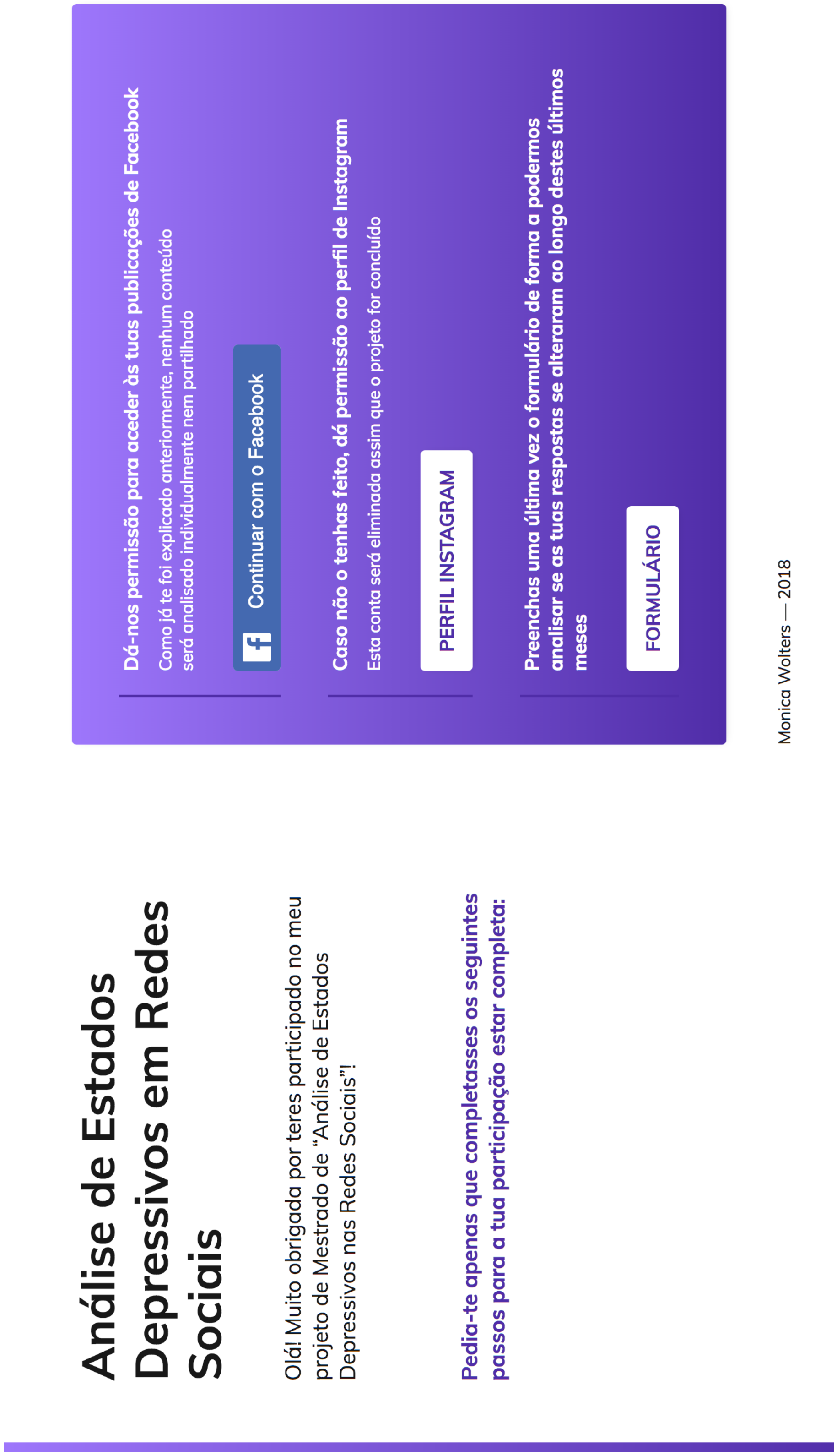
- ◯ Não notei qualquer mudança recente no meu interesse pela vida sexual
- ◯ Encontro-me menos interessada/o pela vida sexual do que costumava estar
- ◯ Actualmente sinto-me menos interessada/o pela vida sexual
- ◯ Perdi completamente o interesse que tinha pela vida sexual

ANTERIOR SUBMETER Página 2 de 2

Nunca envie palavras-passe através dos Formulários do Google.

C

Website for the Facebook APP


Análise de Estados Depressivos em Redes Sociais
Olá! Muito obrigada por teres participado no meu projeto de Mestrado de "Análise de Estados Depressivos nas Redes Sociais"!
Pedia-te apenas que completasses os seguintes passos para a tua participação estar completa:
Dá-nos permissão para aceder às tuas publicações de Facebook
Como já te foi explicado anteriormente, nenhum conteúdo será analisado individualmente nem partilhado
Continuar com o Facebook
Caso não o tenhas feito, dá permissão ao perfil de Instagram
Esta conta será eliminada assim que o projeto for concluído
PERFIL INSTAGRAM
Preenchas uma última vez o formulário de forma a podermos analisar se as tuas respostas se alteraram ao longo destes últimos meses
FORMULÁRIO
Monica Wolters — 2018