INSTITUTO
UNIVERSITÁRIO
DE LISBOA

Detecting Psychological Sentiments in Users from Social Networks

Patricia Sousa Santos

Master in Computer Engineering

Supervisor:

Doctor Ricardo Daniel Santos Faro Marques Ribeiro, Associate Professor,
Iscte – Instituto Universitário de Lisboa

Doctor Fernando Manuel Marques Batista, Associate Professor,
Iscte – Instituto Universitário de Lisboa

October, 2020

TECNOLOGIAS
E ARQUITETURA

Detecting Psychological Sentiments in Users from Social Networks

Patricia Sousa Santos

Master in Computer Engineering

Supervisor:

Doctor Ricardo Daniel Santos Faro Marques Ribeiro, Associate Professor,
Iscte – Instituto Universitário de Lisboa

Doctor Fernando Manuel Marques Batista, Associate Professor,
Iscte – Instituto Universitário de Lisboa

October, 2020

Resumo

Ao longo dos anos tem havido um aumento da população que sofre de depressão. Muitos utilizadores encontram um refúgio nas redes sociais em alternativa a profissionais de saúde, por se sentirem mais seguros e confortáveis a desabafar com outros utilizadores sem que isso os obrigue a expressar uma opinião. Apesar do impacto negativo que as redes sociais podem ter nos utilizadores, muitos estudos mostram que com base nos posts feitos, é possível fazer uma distinção entre utilizadores depressivos e não depressivos e uma análise comportamental e linguística. Desta maneira, a literatura reporta que foram criados e aperfeiçoados modelos de predição de utilizadores depressivos. Nesta dissertação, temos como objetivo testar e desenvolver um modelo que ajude na deteção de utilizadores depressivos em redes sociais com base nas palavras e emoções do posts, utilizando dois léxicos, NRC Emotion Lexicon e VAD Lexicon. Através do NRC Emotion Lexicon quisemos avaliar se a implementação de novas features num estudo apresentado no trabalho relacionado, permite uma melhoria no desempenho do modelo desenvolvido. Já com o VAD Lexicon, o objetivo é analisar se através dos valores obtidos de VAD é possível extrair as emoções presentes. Concluímos que o uso do VAD Lexicon não foi vantajoso visto que, não foi possível correlacionar os valores do VAD com os sentimentos. Para o NRC Emotion Lexicon, concluímos que, apesar da integração do léxico no modelo não ter resultado numa melhor performance do modelo, foi possível observar melhorias, demonstrando que o léxico contribuiu positivamente para a eficácia do modelo.

Palavras chave

Análise de sentimentos, Redes Sociais, Twitter, Reddit, Saúde Mental, Depressão

Abstract

Over the years there has been an increase in the population suffering from depression. Many users find refuge in social networks as an alternative to health professionals, because they feel safer and comfortable to open up with others without forcing them to express their opinion. Despite the negative impact of social networks on users, many studies show that based on the posts made, it is possible to make a distinction between depressed and non-depressed users and a behavioural and linguistic analysis. In this way, the literature reports that prediction models of depressive users have been created and improved. In this dissertation, we aim to test and develop a model that helps in the detection of depressive users in social networks based on the words and emotions of the posts, using two lexicons, NRC Emotion Lexicon and VAD Lexicon. Through the NRC Emotion Lexicon we wanted to evaluate if the implementation of new features in a study presented in the related work, allows an improvement in the performance of the model developed. With VAD Lexicon, the objective is to analyze if through VAD values it is possible to extract the emotions. We concluded that the VAD Lexicon was not advantageous since, it was not possible to correlate the VAD values with the emotions. With NRC Emotion Lexicon, we concluded that although the integration of the lexicon did not result in a better performance, it was possible to observe improvements, demonstrating that the lexicon contributed positively to the effectiveness of the model.

Keywords

Sentiment analysis, Social Networks, Twitter, Reddit, Mental Health, Depression

Acknowledgements

Quero agradecer ao Professor Ricardo Ribeiro e ao Professor Fernando Batista por me terem orientado e apoiado em todas as fases desta tese, mesmo com todos os contra-tempos que apareceram no caminho. Obrigada pela dedicação e por terem acreditado em mim e no meu tema para que esta tese fosse possível.

Aos meus pais, Inês e José, não existem palavras para agradecer todo o apoio, força e amor que me deram para que conseguisse atingir mais um dos meus objetivos. Se hoje sou como sou a vocês o devo.

Ao meu namorado, Bruno, quero agradecer por ter sido uma pessoa fundamental nesta fase, pelo apoio, força e, principalmente, por acreditar em mim.

Quero também agradecer aos meus amigos, em especial, ao João Pereira, Marta Pereira, Carolina Ladeira e Nuno Capela por me terem acompanhado nesta jornada, por todos os conselhos e pela motivação que me deram durante as múltiplas noitadas feitas para conseguir fechar este ciclo.

A todos os que não mencionei mas que fizeram parte deste percurso e me ajudaram a concluir esta etapa, muito obrigada.

Por último, quero dedicar esta tese à minha avó, Ana Serrano, por me ter ensinado a não desistir e a lutar sempre por aquilo que sonhamos, espero que esteja orgulhosa do meu percurso.

Contents

List of Figures

List of Tables

1 Introduction

This chapter presents, the history and initiatives to fight the prejudice present in society against people with mental illness, where depression is the main focus.

Section 1.1 shows one possible way to describe depression as well as, a statement of Solomon's [39] struggle against depression. We describe the continuous struggle of organizations, up to today, to end the stigma and how society shoud act when in the presence of a person suffering from any kind of mental illness. This section end with a list of pros and cons of social networks. Section 1.2 presents our objectives as well as our research questions. Section 1.3 explains the steps to taken into account when extracting data from the two social networks studied in this dissertation, Reddit and Twitter. Finally, we show how the dissertation is structured in Section 1.4.

1.1 Motivation

Mental health reflects the psychological, emotional and social well-being of an individual. This affects the way he thinks, feels and acts in different daily basis situations, being vital to maintain a mental balance in all stages of life, that is, childhood, adolescence and adulthood [46].

There are still issues that do not receive proper attention and therefore it is not possible to keep the population well informed on how to detect, act or even help if they come across any of these issues. For example, nowadays there is still a big stigma related to mental illness, mostly as a result of the negative connotation given by the media and social networks when mentioning mental health-related diseases, discouraging those who need to accept their condition and ask for help [8, 30].

To fight this and with the aim of alerting the society for the importance of mental health as well as possible causes and treatments in case of mental illness, in 1992 the World Federation of Mental Health (WFMH) chose the 10th of October as World Mental Health Support Day. In 1996 they started to approach a different theme each year to cover more subjects of mental health. Also, they aimed to include several common situations involving those who suffer from mental disorders and those who may or may not know how to interact

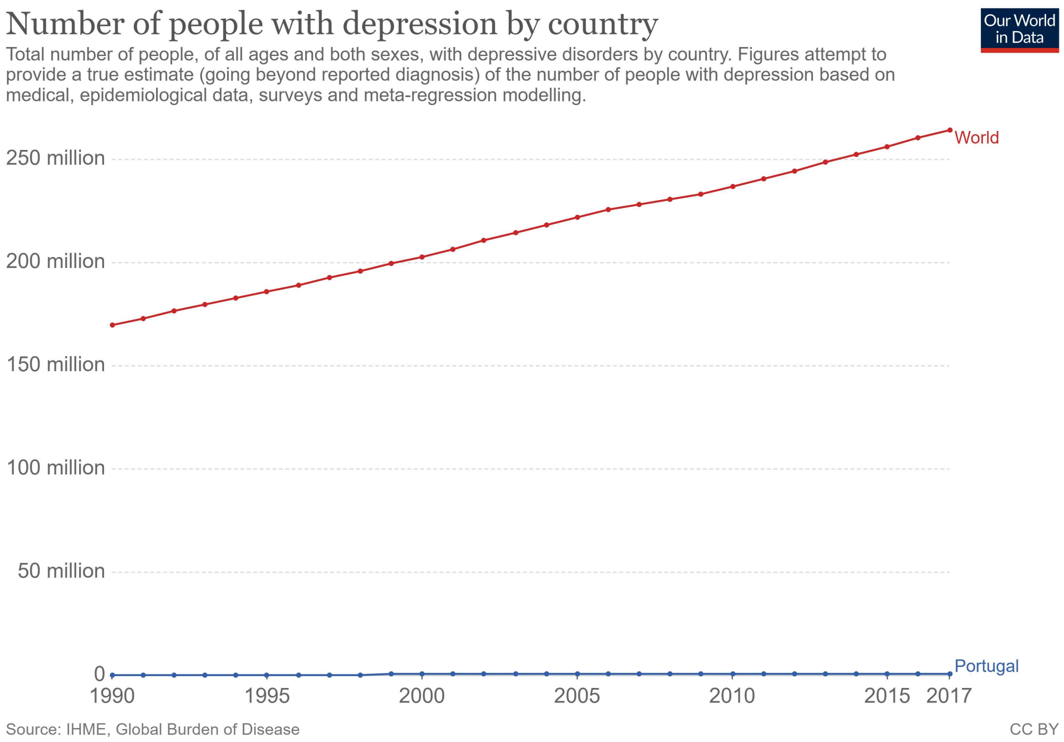

Figure 1.1: Number of people with depression in the world until 2017

with them. For example, "Mental health and work" or "Young People and mental health in a changing world" [4].

According to World Health Organization (WHO) studies, more than 264 million people in the world suffer from depression. This number has been growing over the last few years, as we can see in Figure 1.1 [15]. It is estimated that 1 in 15 adults suffers from depression each year and, 1 in 6 people will go through a depressive phase in their life [44]. Reinforcing these estimations, Figure 1.2 [16] shows that depression is the second mental illness affecting more people in the world.

Depression can manifest itself due to a significant event affecting, as has been said before, the ability to feel, think, sleep and perform daily tasks. However, it is not possible to give an exact definition of it because the symptoms can be experienced in different ways during different periods from person to person [20, 35].

The designation given by National Institute of Mental Health (NIMH) [45], for depression is that, depression is a severe depressive disorder that interferes with mental health, prejudicing daily activities. For someone to be diagnosed with depression, symptoms must be present for, at least, two weeks, depending on how advanced is the state of the disease.

Solomon [39], a writer of politics, culture and psychology, gave a speech at Ted Talks approaching his fight against depression. For him, the first warning signs were "losing interest in almost everything" and the feeling of not understanding why he felt that way.

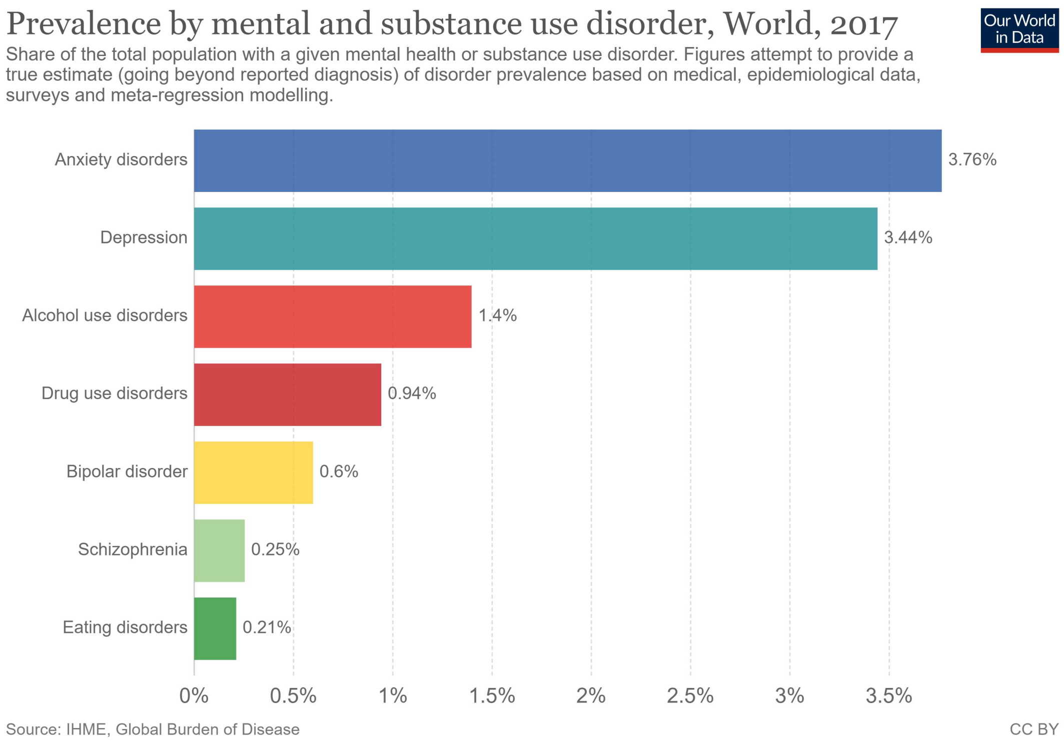

Figure 1.2: Percentage of population suffering from mental disorder

As an example, he said, making a simple meal to feed himself, left him sad for the time he would spent doing that task. During his speech, he highlights some points that he considers important. The first one is the use that is given today to the word depression: "we use this same word, depression, to describe how a kid feels when it rains on his birthday, and to describe how somebody feels the minute before they commit suicide". The word is banalized in multiple situations of our daily lives: it can be difficult to understand the importance and severity associated with it. The next point is related to the definition given to characterize depression, stating that "there are three things that people tend to confuse: depression, grief and sadness" and describing it as "much, much too much sadness, much too much grief" with a feeling of nullness compared to the rest of society. Finally, he clarifies that "one of the things that often gets lost in discussions of depression" is that the majority of people suffering from depression are aware of what they have.

Regardless of the importance and impact of mental health, there has never been a real concern at the level of health and community policies [8] with the intention of improving the quality of life of those suffering from a mental illness [30]. Portugal is with 30 years delay in the approach to mental health compared to other European countries. Only in 1909, with the presentation of Projeto de Lei de Proteção aos Alienados by deputy Miguel Bombarda, mental health began to be taken into account [30].

In 1964, with the Legislative Act 46102, mental health centers were created in Lisbon, Porto and Coimbra, previously approved by the Mental Health Law. However, despite these small advances, only in 2008 Mental Health was approved to be included in the National Health Plan [30].

Combining the information in *Relatório Primavera 2019* [30], Figure 1.3 [16] and Figure 1.4 [15], extracted from the website Our World in Data[1], it is possible to conclude that depression is the second mental illness affecting, in 2017, approximately 550 thousand people in Portugal.

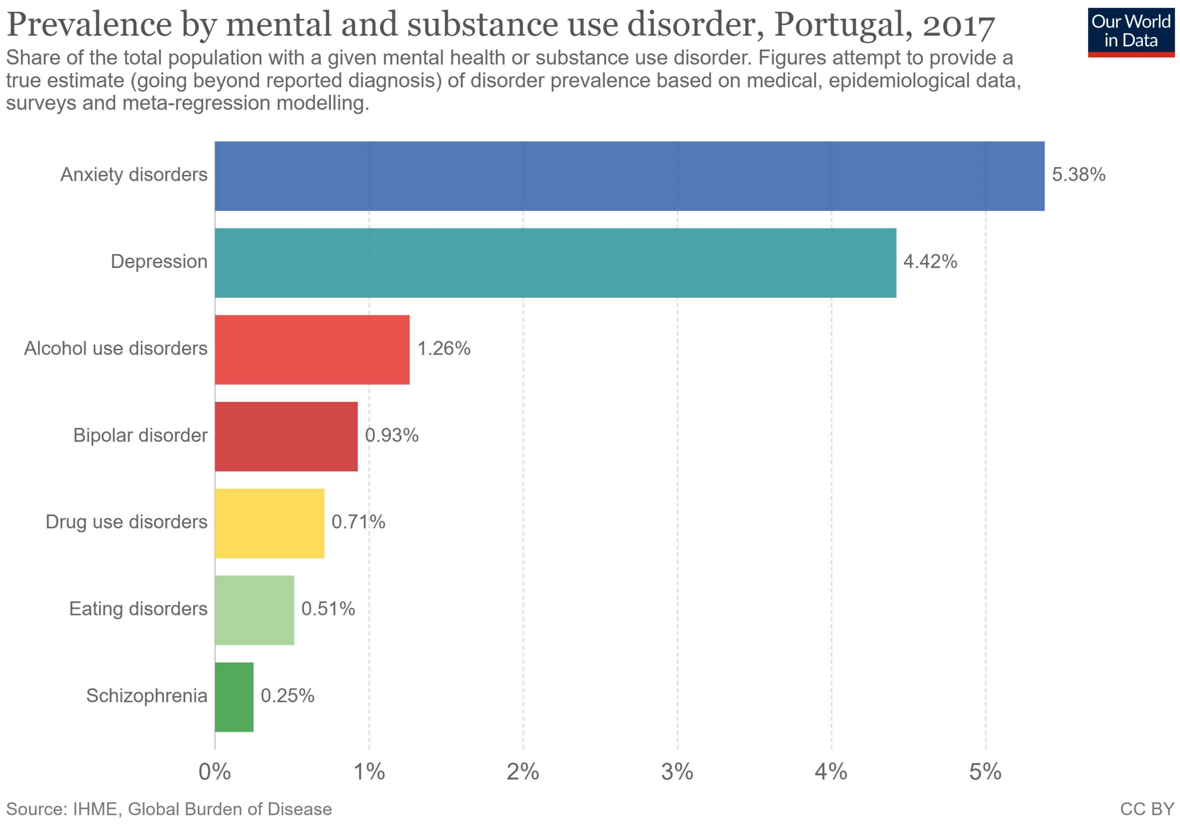

Figure 1.3: Percentage of population in Portugal suffering from mental disorder

[1] https://staging-owid.netlify.app/mental-health

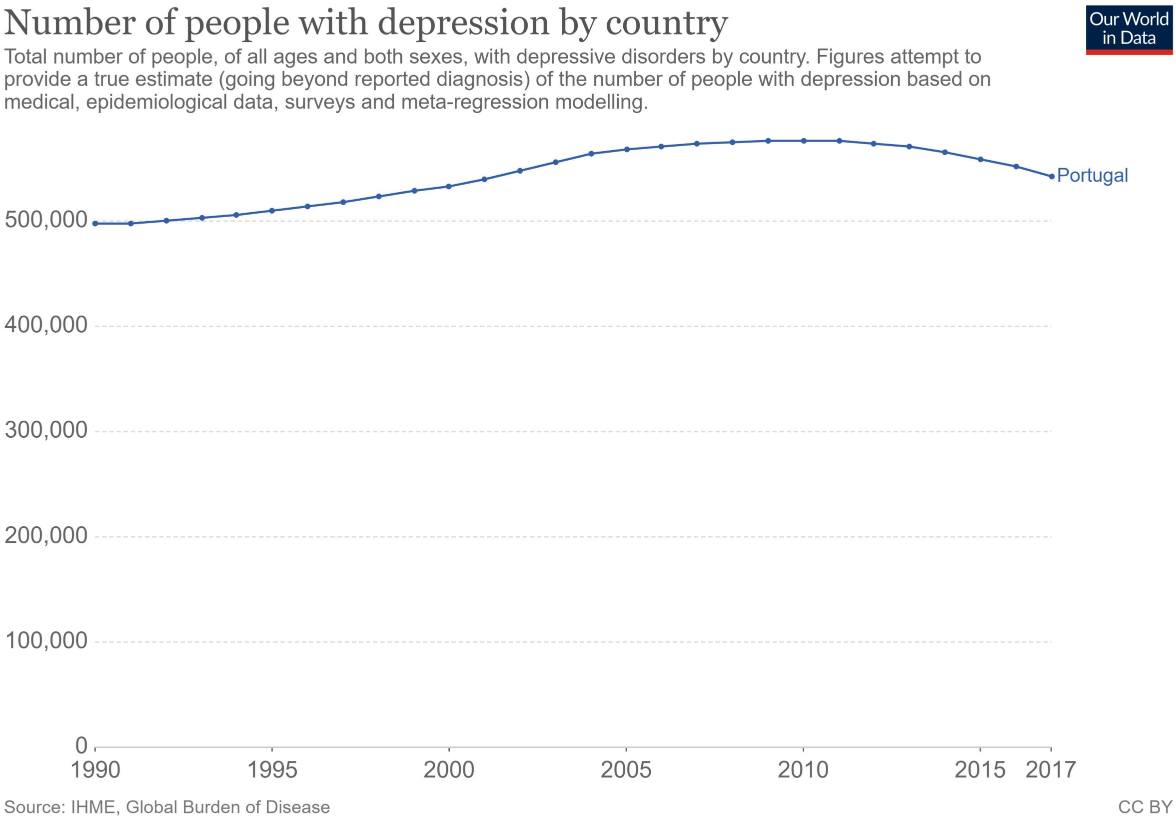

Figure 1.4: Number of people with depression in Portugal until 2017

According to WHO, more than 40% of the countries, have no mental health policy and, over 30% have no mental health program adding that, more than 33% of countries allocate less than 1% of their total health budgets to mental health, with another 33% spending just 1% of their budgets on mental health.

With the lack of resources provided by the state, those who suffer from mental disorders are forced to consult psychologists or psychiatrists who will eventually prescribe antidepressants, psychotherapy or even both as treatment [40]. Figure 1.5 shows that the consumption of antidepressants has increased over the last two decades, placing Portugal, according to OCDE statistics, in the fifth country with the highest consumption of antidepressants [8].

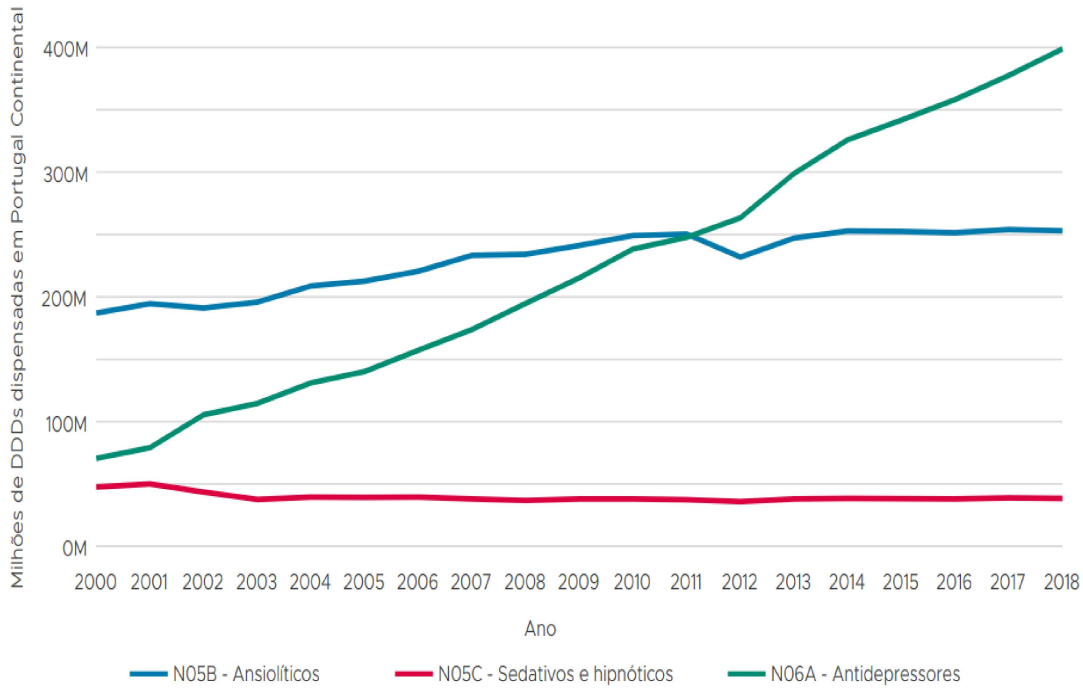

Figure 1.5: Daily Dose Defined dispensed in Portugal (except Açores and Madeira)

Taking these circumstances into account, Shen et al. [37] estimate that, more than 70% of the people that suffer from depression in initial stages do not look for an appropriate treatment. This contributes to a deterioration of their condition and many times, they seek refuge in social networks, for example, Facebook, Twitter, Snapchat, YouTube and Instagram where, most of the users share their musical taste, opinions about current issues in the world, histories from their life and feelings.

Figure 1.6 [14], presents the pros and cons that different social networks have. On one hand, social networks have a bad influence, due the fact that it leads to a social pressure and a fake image that someone's life is perfect and no mistakes are allowed [11]. Adding the fact that, excess of time spent connected can lead to loneliness and isolation, aggravating mental health problems such as anxiety and depression. On the other hand, social networks can be used as a tool to detect psychological states, based on the words used in the posts made by the user. Depressive people tend to use social networks very often, especially Twitter, Reddit and Facebook to express how they feel and how their lives goes. another reason for the use of social networks, given by one of the participants in the study of Park, McDonald, and Cha [32] is that their posts do not interfere directly with anyone's day, and it is not necessary for whoever is reading to stop the thask they are doing, and pay attention to what is being written. Reinforcing this idea, this participant adds that when expressing himself personally to his friends, he feels that they are worried and saddened by empathy, making him feel more frustrated. However, expressing the same in a social network in the

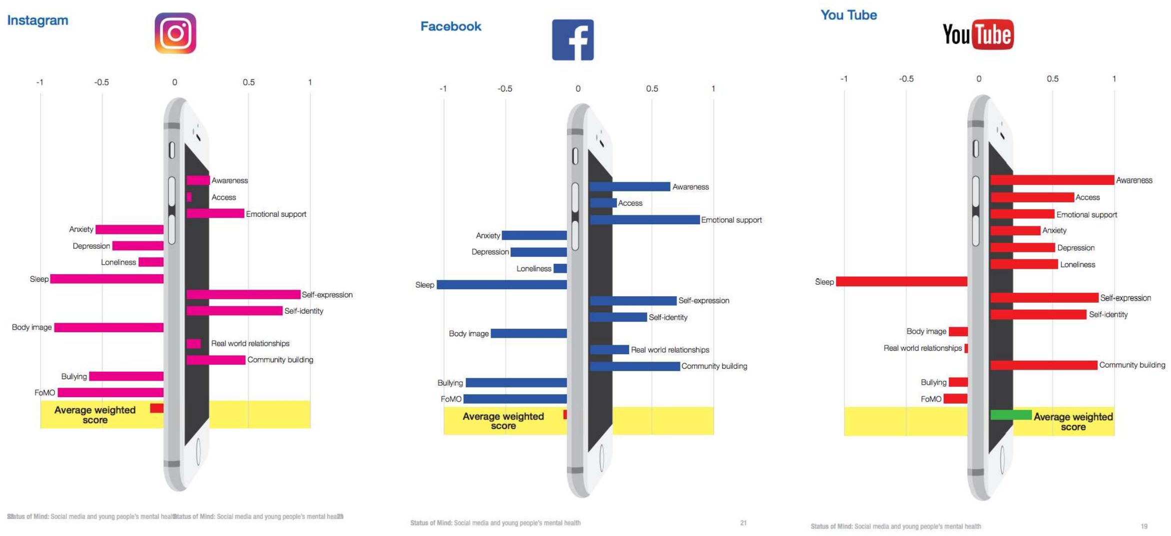

Figure 1.6: Relation between social networks and mental health

absence of emotional connection, the feedback received is more objective and useful.

1.2 Goals and Research Questions

The main objective of this dissertation is to detect signs of psychological suffering, especially depression, in posts made by users on their social networks. In order to achieve this goal, this study focus on two different social networks, Reddit and Twitter, based on posts in different languages, English and Portuguese, respectively.

Regarding Reddit, it is the eighth most used social network in the world, with a online forum which allows users to express their opinions on several topics.This social network is formed by several mini-forums, called, subreddits, which are associated to different topics [17].

Meanwhile, Twitter is considered a microblogging that allows users to send small posts, designated as tweets, with their thoughts/opinions on the topic they want to express their opinions [12].

Using Twitter, we intend to provide an analysis focused on the profile of users and respond to the question, can a chronological investigation help to identify signs of depression? If so, is it an isolated event or recursive?

Through Reddit, we expect to make a analysis focused on the classification of detection models using as a reference, the work developed by Losada and Crestani [24] and, respond to the question, does including new features lead to an improvement in the performance of the developed model?

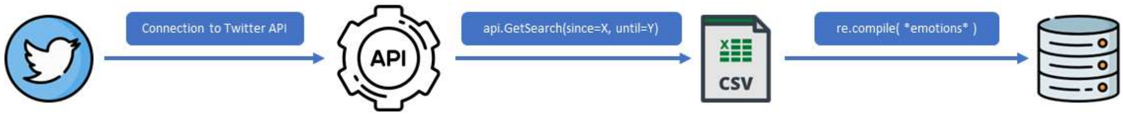

Figure 1.7: Extract Twitter data diagram

1.3 Methodolody

Depending on the desired social network and type of information, is necessary to use different tools to filter and extract the data that best fit the purpose of the study. This way, it is possible to obtain a dataset with fewer irrelevant data and more realistic as possible.

As mentioned above, in this work was necessary to extracted data from social networks with focus on data collected from Twitter and Reddit.

Figure 1.7, shows the steps need to extract data from Twitter [5], using Tweepy, a package provided by Python Twitter. Start by creating and authenticating a Twitter API object in order to get access to all the information available, for example, tweets, retweets, likes, direct messages, and followers. To extract the data, the `GetSearch()` method, available in Twitter API library, was used. Since our study required data within a specific period of time, we set the parameters *since* and *until* of the method were set with the desired dates. The data returned in this method was stored in a `.csv` file, where we applied a regular expression with the emotions associated to depression, obtaining the final dataset exported to a SQLite database.

Figure 1.8 shows the steps required to extract data from Reddit [43]. Using Python Reddit API Wrapper, PRAW, a python package that allows the access to Reddit. Following the authentication of the Reddit API object, which will give us access to all the subreddits available. To extract the data, the `subreddit()` method was used. This method receives as argument, the name of the subreddit where the desired information is. However, the return of this method will only give us, the id associated to the post and not the content. So, for that, the `dir()` method was used in the variable returned above, in order to understand which fields can be captured in the posts. After the analysis and choice of fields, we must invoke again the variable returned in the method `subreddit()` call with the name of the field from which we want to collect the content. The collected information was organized in one `.txt` file per user with their respective posts.

In our study, the criteria used to select depressed users was based on the confirmation of a diagnosis for depression in their posts, in other words, the user affirms being diagnosed with depression or confesses the use of depression medication.Which is why most of these users were in the subreddit dedicated to depression. Regarding the selection of non-depressed users, the criteria included different scenarios. Some users were picked up

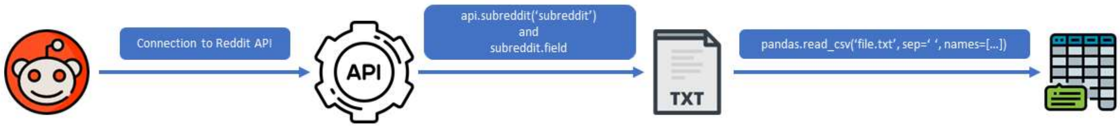

Figure 1.8: Extract Reddit data diagram

because they had a close relative with depression, other users were in the "depression" subreddit but did not suffer from depression and, finally, a random selection of users was made throughout the different subreddits.

After gathering the databases with the necessary data to the study, it is important to pre-process the data it in order to remove unnecessary information that will not benefit the learning of the prediction model. As shown in Chapters 3 and 4, the pre-processing done in our study involved the conversion of text to lower case, removal of url's, tags, punctuation through regular expressions and, finally, treatment of abbreviations and negation expressions.

1.4 Structure of the Document

This dissertation is organized in five chapters presenting all steps made in this study. Chapter 2 presents work previously studied regarding the differences between depressed and non-depressed users as well as, different implementations made to detect and predict depressed users in social networks. Chapter 3 discuss the obtained results using athe dataset from Twitter. Chapter 4 explains and present results for the experiences made using the dataset from Reddit. Chapter 5 draws conclusions obtained from the experiences made throughout this dissertation and, indicates some points to take into account for future work.

2 Related Work

This chapter reviews the literature through different research studies done for this topic. It starts by making an analysis of the main differences between depressed and non-depressed users, in Section 2.1, regarding the impact that Internet has on their lives, their behaviour on social networks and, the language used in their posts. Section 2.2 provides a overview of the different approaches for the prediction and detection of depressed users, describing which features and algorithms were implemented for user classification.

2.1 Depressed and Non-Depressed Users

Shaw and Gant [36] wanted to prove whether or not the use of the Internet had an impact on four areas, Depression, Loneliness, Self-Esteem and Perceived Social Support. In order to achieve this goal, they invited 40 participants, to communicate with each other through WebChat CT Software, during a period of 4 to 8 weeks. Throughout this trial, the participants answered three tests, in different periods, to evaluate the evolution in the areas of analysis referred previously. For depression, the Center for Epidemiological Studies Depression Scale (CES-D) was applied as a measure with a 20 questions quiz related to mood and behaviour. For Loneliness, the Revised University of California Los Angeles (UCLA) Loneliness Scale was used with 20 questions related to how well participants agreed with certain statements on a scale of 1 to 4. Self-Esteem was assessed using the Texas Social Behaviour Inventory (TSBI), a 32-question meter, where participants rated from 1 to 5 how much they would agree with a given statement. Finally, the Cohen-Hoberman Interpersonal Support Evaluation List (CHISEL) a 48-question true/false questionnaire related to social support aspects, helped measuring Social Support. After this trial, they concluded there is no negative impact of the Internet on a person’s life. However, if it had any negative influence, self-esteem would not be affected as quickly as the other three. They also emphasize that an online relationships significantly differ from face-to-face relationships because of the possible anonymity of the Internet, which might lead users to reveal problems and personal information more easily.

Using the analysis made by Park, McDonald, and Cha [32], we were able to understand the behaviour and opinion of depressed and non-depressed users on Twitter, learn about the mechanism used to follow someone, preferences to follow an account, reasons to unfollow, and finally, what they think about the limit imposed by Twitter for only seeing the 3200 most recent tweets. Regarding the first point, mechanism of following, both depressed and non-depressed users, agreed that they only added someone after analyzing the profile information of the person and decided whether those tweets are something they would want to read in the future. When asked what was the type of users they preferred to follow, the depressed group said they follow users who post life episodes and have emotional content, which help them, in some cases, realizing that their problem is not that bad and consequently, they gain more self-control of their issues. While non-depressed group follow those who can produce unique information, normally authoritative sources or institutions. Regarding the reasons why they unfollow someone, the non-depressed group claims that usually is because the person tweets too much, mastering their feed, this reason is also given by depressed group however, they add that being a user too "whining, gloomy" or depressed is a valid reason to unfollow because that state of mind might influence them. When it comes to the debate about Twitter or Facebook, both agreed that Twitter is a better platform despite the different reasons: on one hand, the depressed group sees Twitter like a medium, where everyone talks about their own stories and problems but yet, pays attention to different things that gives a feeling of community because no one it is obliged to listen their problems or give advice; on the other hand, the non-depressed group sees value in Twitter because that allows them to read the thoughts of public figures and collect information from authoritative and non-authoritative sources, such information might not be found in a close connected social media community like Facebook. As for the limit imposed by Twitter to only put available the 3200 most recent tweets, the depressed group agree with this approach because for them is like a disclose of their past emotions and daily situations. Otherwise, they would feel uncomfortable by knowing that some of their close friends find a post with their feeling exposed. On the other side, the non-depressed group do not approve this implementation because they lost information that they though it would be helpful in the future.

Rude, Gortner, and Pennebaker [35] compared the linguistic patterns between depressed and non-depressed users using an essay task as a study reference. To measure this task, which consisted of writing down the deepest thoughts and feelings that users had for returning to school, it was used the Beck Depression Inventory (BDI). BDI measures the level of depression in users and it has shown to obtain depression ratings quite similar to evaluations made by psychiatrists. As a complement to the previous measure, a 22-item self-report, The Inventory to Diagnose Depression (IDD-L), was used, which evaluates the extent and duration of the depressive symptoms felt in the week in which that feeling was more present. Finally, Linguistic Inquiry and Word Count (LIWC) was used to analyze the text, focusing on the use of the first person singular, first person plural, words with negative and positive emotions, as well as social references, for example, mentions made to friends or posts made. After the analysis, they concluded that depressed users have a more accentuated use of the first person singular and negative emotions. However, concerning social references there were no differences between depressed and non depressed users.

2.2 Automatic Detection of Signs of Depression in Social Networks Users

In 2013, De Choudhury, Counts, and Horvitz [10] wanted to create an index that would eventually serve as a characterizer of depression levels in the population. To reach this goal they used crowdsourcing to collect tweets from users after being selected through a combined Amazon's Mechanical Turk (AMT) questionnaire and Center for Epidemiologic Studies Depression Scale (CES's) questionnaire to determine which users could be classified as depressed and non-depressed and later, split into the respective groups. To help in the analysis of tweets, the extracted features are fed to a classification framework, in this case, a Support Vector Machine (SVM). The features were directed to both the tweet and the user, i.e., within the category of tweet, we can evaluate through emotions, time, language style and engagement, while for the user category we evaluate according to the number of followers, the number of people following, the number of links that are shared, among other aspects. After the experiments, the conclusions obtained related to the classifier were that the best result for average accuracy was around 70% and the highest precision rounded 74%.

In the same year, Choudhury et al. [7] wanted to evaluate the potential that social networks might have in the detection and diagnosis of severe depressed disorders in users. To collect and evaluate the users, they used the same approach mentioned above, namely the AMT and CES's questionnaire, adding the BDI questionnaire. In this way, they avoided noisy responses considering that if a user was truly depressed he would have high scores in both CES's and BDI questionnaire. As for the analysis of the posts, five groups of features were adopted: Engagement features, with those described in the previous study and insomnia index, which quantifies the pattern of posts published by the user throughout the day. For Egocentric Social Graph features, there are three types of measures, node properties, user-focused, dyadic properties, user interactions with others through @-reply and, finally, network properties that analyzed the user interaction with the whole network. The next group of features refers to emotion, where the positive affect (PA) and negative affect (NA) were analyzed, with the help of the LIWC lexicon while activation and dominance was based on the Affective Norms for English Words (ANEW) lexicon. LIWC lexicon was again applied, to help in the analysis of the 22 linguistic styles related to the Linguistic Style features group. Finally, for the depressive language features, a lexicon was created with terms used by depressed users in their posts as well as a measurement of the level of use of well-known antidepressant names for clinical depression treatments. Once again, the SVM was the chosen classifier to predict depression, using a Radial-basic Function (RBF) kernel. After the prediction was done and the results were collected, the conclusions reached were that linguistic style features improve the performance of the model, however, for the prediction task the features that obtained the best results were related to depressive language.

In 2017, the study made by Shen et al. [37] had the goal to detect depressed users in a timely way through data collected from social networks. As in previous studies, Twitter API was used to collect data using as criteria for depressed users, the confirmation of a diagnosis for depression in a post, while for non-depressed users the criteria used was those who had never mentioned the word "depress". Regarding the analysis of the posts they were based on the same features explained previously. In order to determine the best classifier to implement, they compared Naive Bayes (NB), Multiple Social Networking Learning (MSNL), Wasserstein Dictionary Learning (WDL) and, Multimodal Depressive Dictionary Learning (MDL). The obtained results showed that WDL achieved a better performance than NB, however, it was outperformed by MSNL and MDL. Ultimately, MDL surpassed MSNL by combining the dictionary learning strategy providing a better performance for depression detection. After applying the MDL classifier, the reached conclusion was that depressed users tend to tweet 41% more between 11 PM and 6 AM indicating they might suffer from insomnia. Concerning the ratio in a post, depressed user had 52% negative words to 37% positive words compared to non-depressed users that had 23% negative words to 17% positive words. These results showed that even though the bad mood is in both groups, depressed users have a tendency to show more their feelings.

Stephen and P. [41] tried to prove that it is possible to categorize a user as depressed or non-depressed, taking into account the calculation level of depression present in their posts. In order to achieve that, the Twitter API was used to collect posts from users who had hashtags with the word “abuse”, “bullying”, “anxiety”, “addict” and “addiction”. After the dataset was built, a preprocessing was done, starting with the conversion of JSON data from UTF-8 to ASCII, then the removal of exclamation marks, punctuations, digits and special characters at last. By joining a TM library and a list of positive words treated by Minqing Hu and Bing Liu the stops words present in the lexicon were removed. Afterwards, the analysis of the posts obtained was divided into three phases, the first phase consisted on calculating the emotion and polarity, positive or negative, present in each tweet through the help of syuzhet package [21] and a list of eight emotions, which includes anger, anticipation, disgust, fear, joy, sadness, surprise, and confidence. The second phase, had the help of three lexicons, AFINN, BING and NRC, to evaluate the feelings of the tweets per user, receiving a score from -1 to +1. Finally, the last stage consisted in the evaluation of the level of depression in a user’s tweets. Taking the above analysis into account, it was concluded that it was possible to determine whether a user is depressed or not, correlating the levels of depression with the content present in their posts.

These approaches are also being applied to other social networks. For example, Islam et al. [19] investigated which factors reflected depressive signs in Facebook users. The dataset was created based on public data available from the Facebook platform, captured by NCapture[1]. For this study, they decided to focus on four factors: emotion keywords composed of positive emotions, negative emotions, words related to sadness, anger and anxiety; another factor were temporal words in the past, present and future; for penultimate factor were chosen the standard linguistic dimensions extracted by LIWC; the last factor was a combination of all three factors mentioned above. For the prediction model four classifiers were explored, SVM, Decision Tree (DT), Ensemble and k-Nearest Neighbor (kNN), being the DT the one to obtain the best performance results as well as the highest values for precision, recall and F-measure when combined with the linguistic word factor. With an accuracy between 60 to 80%, they concluded that approximately 55% of the users present in the dataset communicated between midnight and noon, indicating signs of loneliness, stress or lack of energy and that signs of depression are more correlated with personal problems.

For Losada and Crestani [24], Reddit's platform was the best social network to use in order to achieve the proposed objective: evaluate the chronological evolution in the language present in posts of depressed people since, they consider time a fundamental factor. For the selection of depressed users they decided to adopt the approach proposed by Coppersmith, Dredze, and Harman [9] which was based on the use of an automatic method to identify people diagnosed with depression. Concerning the non-depressed group, a few users were randomly collected through the Reddit API and others were in the subreddit related to depression but did not suffer from it. After building a depression language classifier, they applied Logistic Regression with L1 regularization in four strategies, Random, Minority, First, and Dynamic. After the experimental phase, the results show that there has to be a “balance between correctness and time”, and that the *Dynamic* strategy is the most efficient considering the response time and, for this reason, obtained a better performance compared to the other strategies.

Tadesse et al. [42] tried to find the best solution to increase the performance of a prediction model through isolated features and through their combination. To achieve this, they used the dataset built by Pirina and Çöltekin [33], with depressed and non-depressed user posts, then divided the study into four points: the first focused on the relationship between depression and user language; the second in the grouping of three groups, them being, standard linguistic dimensions, psychological processes and personal concerns; for the penultimate point they evaluated the behaviour of N-grams, LIWC and Latent Dirichlet Allocation (LDA) with the isolated features; the last point showed the performance of the classifiers in the prediction of isolated and combined features. In the prediction process, the classifiers chosen were Linear Regression (LR), SVM, Random Forest (RF), Adaptive Boosting (AdaBoost) and Multilayer Perceptron (MLP). The conclusions obtained after applying the classifiers for individual features were that the best results were achieved by SVM and bigrams with 80% accuracy, followed by LIWC and RF, with 78% accuracy, and finally LDA and LR, with 77% accuracy. As for the combined features, the best result ob-

[1] https://www.kent.ac.uk/software/ncapture

tained was the junction of LIWC, LDA and bigrams with the MLP model with 91% accuracy.

Li, Lu, and Long [23] wanted to investigate the effectiveness of different lexicons using sentiment phrase analysis. To achieve this goal, they compared the behaviour of different lexicons, Valence Aware Dictionary and sEntiment Reasoner (VADER) sentiment lexicon, NRC Hashtag sentiment lexicon (HSenti), Evaluation Potency Activity (EPA) lexicon, Valence Arousal Dominance (VAD) lexicon and, finally, word embedding of 300 dimension. In the experimental phase they built three sentiment analysis tasks considering Stanford Sentiment Treebank (SST): the first regression task had to predict the punctuation of the sentiments present in sentences, called SST-R; the next task is of binary classification intended to convert the sentiments into labels, above 0.6 is considered positive feeling and below 0.4 negative feeling, this task is called SST-2c; finally, the last task, SST-3c, is ternary classification with the same objective as SST-2c however it focuses on the range between 0.4 to 0.6. For the SST-R task mean absolute error and Kendall rank correlation coefficient were used as evaluation metrics and SVM as classifier. For the SST-2c task they used accuracy and F-score and linear kernel as classifier. Finally, the SST-3c task used weighted accuracy and weighted F-score and, similarly to the SST-2c task applies the linear kernel as classifier. Based on the results obtained, they concluded that VADER lexicon had a better performance than EPA lexicon and, unexpectedly, found that multi-dimensional lexicons had worse results than VADER.

Since 2017 related shared tasks have been part of in the Conference and Labs of Evaluation Forum for Early Risk Prediction, more known as CLEF eRisk, this competition allows several researchers to create reusable references through early detection technologies and use them in different areas, mostly in health and safety [25, 42].

Almeida, Briand, and Meurs [2] study is based on the systems developed by the University of Quebec in Montreal (UQAM) team for CLEF eRisk Pilot Task 2017, which aims to predict as early as possible, the risk to mental health caused by posts made by the users on their social networks. Since they were based on the CLEF eRisk Pilot Task 2017, the datasets, training and testing, were already collected with data from Reddit. After the preprocessing where emojis were replaced by the related emotion, the authores applied an approach that has resulted from the combination of Information Retrieval (IR) and Supervised Learning (SL) predictions. In the IR it was given a score according to the probability that the analyzed document was produced by a depressed user. Subsequently the documents were sorted by score and users were classified as depressed or non-depressed in case the score was higher than a given threshold. For SL, the authors extracted four types of features such as, n-grams, dictionary words, selected Part-Of-Speech (POS), and user posting frequency and also elected three classifiers, Logistic Model Tree (LMT), Ensemble of Sequential Minimal Optimization (SMO), and Ensemble of RF. The conclusions achieved from these experiments were that, the team A (UQAMA) had the best performance and the best F1 and Recall values, while team D (UQAMD), had the best value for precision.

Benamara et al. [6] study was also based on the dataset made available by the CLEF eRisk 2017 task, with the purpose of detecting depression signs as early as possible through the analysis of posts made by users. To help the analysis of the posts, seven features were selected: Bag of words, Language Style, User behaviour in social media, Self-Preoccupation, Reminiscence, Symptoms and drugs, and finally Sentiment and emotion. Later, four models were created to be applied in two tasks. To detect if a user was depressed by their posts and, to detect signs of depression with the posts organized chronologically. The baseline will serve as a comparison to the models created and will use the feature bag of words, Model 1 uses the same bag of words, Language style and from User behaviour, Model 2 uses the same as Model 1 and Self-Preoccupation, following the same logic Model 3 uses the features used in Model 2 plus Reminiscence and Symptoms, finally Model 4 uses all the features. For the test phase, four classifiers were elected to apply to the described models, Sequential Minimal Optimization (SMO), NB, RF, and LR. After applying the classifiers to the models, it was concluded that the best results were achieved with RF, which detects the depressed users with a precision of 75%, a recall of 51.9%, and an accuracy of 91.5%.

Even though there are not articles related to CLEF eRisk 2020[2] available yet, since it happened very recently, depression was the target of one of the tasks, Task 2, aimed to explore the effectiveness of automatically evaluate the gravity of the symptoms associated to depression. To achieve this goal, it was considered the historical of the user's posts collected after answering the proposed surveys. Through the algorithms it was obtained a prediction of the supposed answer of the user to each one of the questions and, later, a comparison was made between the answer given by the user and the predicted one, evaluating the veracity and performance of the algorithm.

Taking into account the studies described, we conclude that the lexicons that best fit in our objective are, VAD Lexicon and NRC Emotion Lexicon. Regarding the classification method, we conclude that the Logistic Regression is the best classifier to be used in order to help us with the comparison of the results obtained by Losada and Crestani [24].

[2] https://erisk.irlab.org/

3 Experiments with Twitter

In this chapter, a database was created with tweets in Portuguese, to which we applied two different lexica. After a chronological investigation to detect signs of depression, we analyzed the emotions present on users tweets, in order to understand if it was an isolated event or recurrent. With that analysis we were able to classify the users present in our dataset, as depressed or non-depressed. Section 3.1 describes the extraction and the criteria to create the dataset. Section 3.2 explains the developed implementation using two different lexica, VAD Lexicon and NRC Emotion Lexicon. Finally, Section 3.3 presents the results obtained and the website created to simplify their interpretation when analyzed by a health professional.

3.1 Data

In order to make a better analysis focused on the profile of users, we decided to create a dataset based on Portuguese tweets, extracted through the Twitter API, imposing as criteria, tweets that contained the words ’medo’, ’assust’, ’morrer’, ’chor’, ’aguent’, ’sofr’, ’infeli’, ’vida’ and ’odeio’, during the period of 1st January 2018 to 31th December 2018, reaching a dataset composed of 259866 tweets and 22169 users.

Given the large number of users, we decided to analyse only the 40 most frequent users, meaning, those who had published over than 255 posts. Table 3.1 shows a small statistic of the data obtained for this dataset.

The data from our dataset, as we can see in the Figure 3.1 is composed by a unique tweet id, an id associated to each user present in the dataset, the date when the tweet was published, the text of the tweet and finally, information related to the user, name and location. It is extremely important to highlight that at no moment during this study the names

Number of	
Users	40
Total posts	13584
Posts per user (avg)	339,6

Table 3.1: Statistics from Twitter dataset

```
(xxxxxxx416, xxxxxx01, '2018-01-01 00:01:04', 'Estou a chorar e nem sei pq',
{"user": xxxxxxx, "place_name": "Felgueiras, Portugal")')
```

Figure 3.1: Example of a tweet

id	total tweets	max. words	min. words	avg. words	max. chars	min. chars	avg. chars
XXXXXX01	648	40	2	11.32	216	11	58.80

Table 3.2: Example of statistic based on a user

of the users have been disclosed or used, users are identified only by their id, maintaining their anonymity.

After we got the dataset with all 40 users we made an analysis and determined for each user, the total number of posts, the maximum, minimum and average number of words used in a post and the maximum, minimum and average number of characters present in a post. Table 3.2 shows how these data are saved.

3.2 Setup

Since our posts were in Portuguese and we wanted to test the polarity of the posts, one of the existing tools to obtain polarity, is TextBlob, however it only works for English texts. To get around this restriction, we tried to translate the tweets to English using different libraries. However, we found limitations that made this approach impossible, for example, due to the limit of HTTP Request or the maximum characters allowed to be submitted.

As a result of this limitation we chose to use lexicons. For our study, the lexicons chosen were, NRC Valence, Arousal, and Dominance (VAD) Lexicon and NRC Emotion Lexicon. These lexicons contained a list of words, in different languages, as well as the polarity and emotions associated with it.

VAD Lexicon [27] is a list that contains more than 20,000 English words as well as, translations in over 100 languages, with the respective scores of valence, arousal and dominance. For our study we used the columns 'Portuguese (Portugal, Brazil)-en', 'Valence', 'Arousal' and 'Dominance', with this analysis we expected that it would be possible to compare the VAD values obtained from the tweets with the VAD values characterized in the model spanned across the six basic emotions, as Figure 3.2 shows. However, since there is no range of VAD values associated with emotions related to depression, it was not possible to make a reliable interpretation of the results obtained, in Table 3.3 we demonstrate how the data was saved after this analysis.

Using NRC Emotion Lexicon we managed to analyze the tweets and the emotions associated, first we studied tweet by tweet for all users of our dataset and then, aggregated all

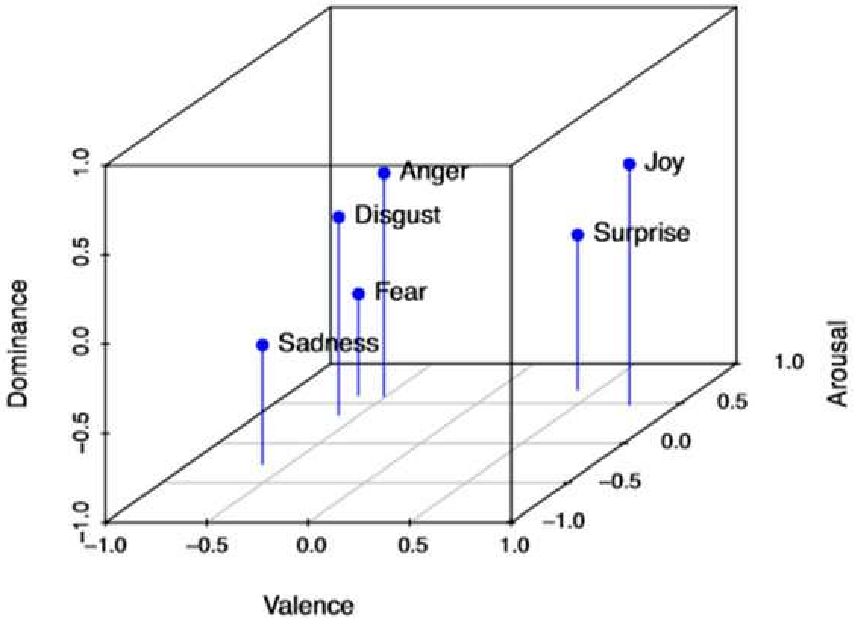

Figure 3.2: The VAD model spanned across the six basic emotions [3]

original tweet	analysed tweet	date / time	valence/ arousal/ dominance
Odeio que fiquem a inventar desculpas e assim, sejam logo sinceros	odeio que fiquem a inventar desculpas e assim sejam logo sinceros	03/01/2018 02:39:20	0.090 0.086 0.104
Odeio juro esses gajos q n sabem ver futebol, n saibam admitir as cenas e n saibam ser honestos, juro	odeio juro esses gajos que não neg_sabem neg_ver neg_futebol não neg_saibam neg_admitir neg_as neg_cenas neg_e neg_não neg_saibam neg_ser neg_honestos juro	04/01/2018 04:09:40	0.234 0.287 0.337
Dizem q sou diferente das redes sociais para a vida real	dizem que sou diferente das redes sociais para a vida real	17/03/2018 04:06:16	0.496 0.358 0.491
Odeio q n reparem em mim	odeio que não neg_reparem neg_em neg_mim	01/04/2018 11:15:43	0.129 0.091 0.140
Odeio n ser interessante	odeio não neg_ser neg_interessante	01/04/2018 18:30:51	0.180 0.295 0.358

Table 3.3: Example of VAD Lexicon analysis

the values obtained per user, calculated the mean and standard deviation of the VAD values obtained for the analyzed user, as well as the percentage and absolute value of positive, negative, neutral tweets. At last, the percentage and absolute value of tweets associated with each of the eight emotions evaluated, 'Anger', 'Anticipation', 'Disgust', 'Fear', 'Joy', 'Sadness', 'Surprise' and 'Trust'.

A manual analysis and distribution of all 40 users was made, into three different groups, since our dataset did not have this information. The depressed group, were represented by users who mentioned in their posts they had been diagnosed or showed signs of depression, for example, mentioning taking antidepressants, psychologist appointments, and, based on the works done by Shen et al. [37] and Choudhury et al. [7], the period of time where there is an higher amount of posts from depressed users. Non-Depressed refers to users who do not show signs of depression and had no diagnosis that proved the opposite. Finally, Potentially Depressed, refers to users who occasionaly demonstrate signs associated with depression, such as, anxiety. By the end of this analysis, we obtained 11 depressed users, 17 non-depressed users and 12 potential depressed users.

However, to continue with the development and implementation of the model, we would have to verify this diagnosis with a psychologist since, there was no information field extracted, which guaranteed us the veracity from the manual analysis done.

3.3 Results

In this Section we discuss the results obtained by the manual analysis explained in the Section 3.2.

We started by analysing the total number of tweets per user, shown in Figure 3.3. In contrast to the results presented in the studies mentioned in Section 2.1, with our data, we concluded that the group of depressed users has a smaller number of tweets compared to the other groups. However, this may be related to the fact that, in our study, we filtered the tweets collected based on the emotions found, as explained in Section 3.1, meaning that the tweets taken into account presented emotions related to depression.

The group of non-depressed users presents a higher number of tweets per user, due to the high frequency they describe their emotions as a result of the occurrences associated with social/leisure events, such as soccer games, concerts, watching a film/series or listening to music, as can be seen in the examples below.

> "porque é que eu choro em todos os filmes?"
>
> "estou a chorar pelo meu Porto"

For the group of potentially depressed users, it is possible to observe a similarity with the group of depressed users regarding the total number of tweets per user, however, as mentioned above there are occasional tweets that show signs of depression, such as anxiety. The

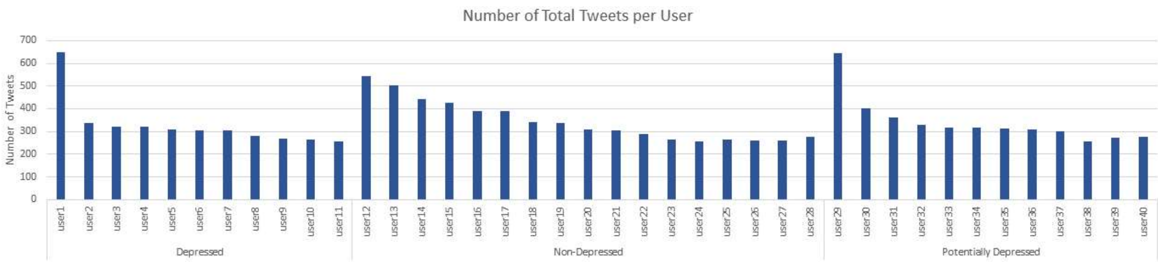

Figure 3.3: Total number of tweets per user

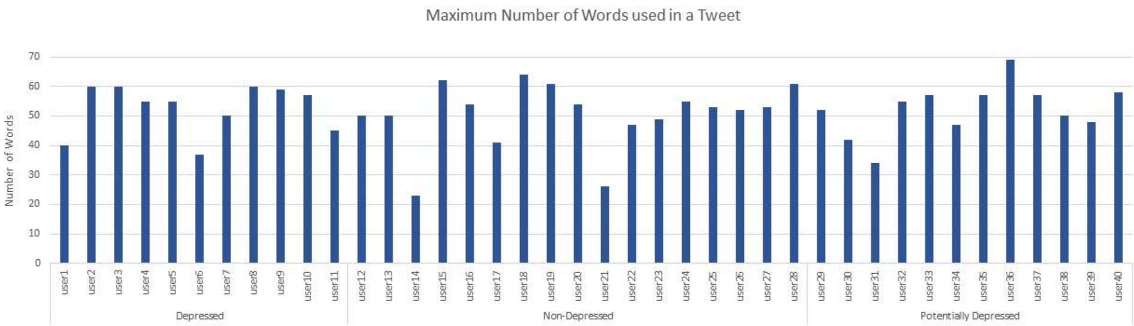

Figure 3.4: Maximum number of words

remaining tweets of this group report once again emotions associated with social/leisure events, as described in the group of non-depressed users.

We extracted the metrics related to the maximum and average number of words present in the users' tweets in order to analyze if they presented similar values within the same group and discrepancies between groups.

The results of Figure 3.4 and Figure 3.5 show that there is no significant difference between the groups, however, within the same group, we can observe some variation.

Another metric used to compare the different groups was the characters used in tweets. Again, the differences between the groups are not substantial, as can be seen in Figure 3.6 and Figure 3.7.

After the linguistic analysis, we decided to focus on polarity and emotions found in tweets of the users.

We started by observing how polarity was distributed throughout users from each group. Polarity is categorized in three different types: positive when the emotions in the tweets are mostly positive when compared to negative; neutral represents the balance between positive and negative emotions; and, finally, negative when the tweets contain more negative emotions than positive emotions.

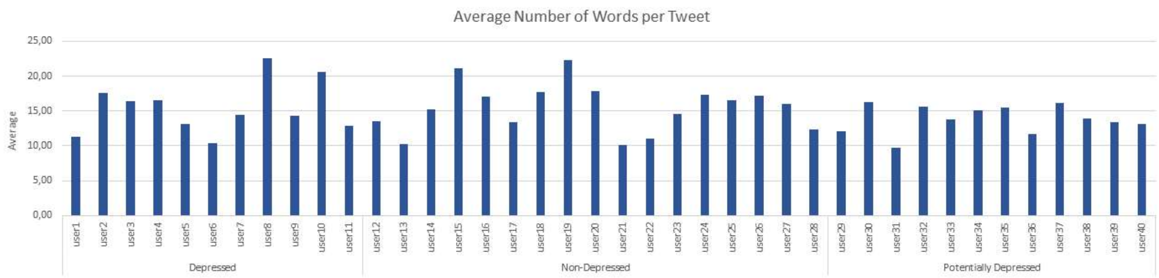

Figure 3.5: Average number of words

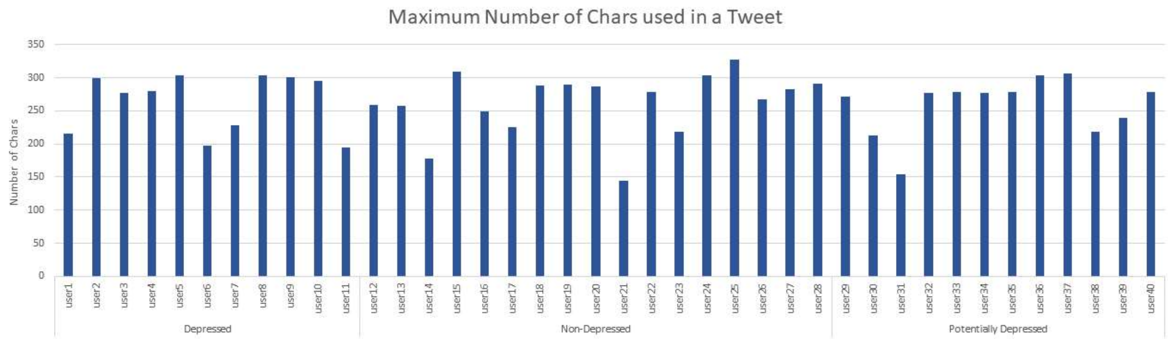

Figure 3.6: Maximum number of chars

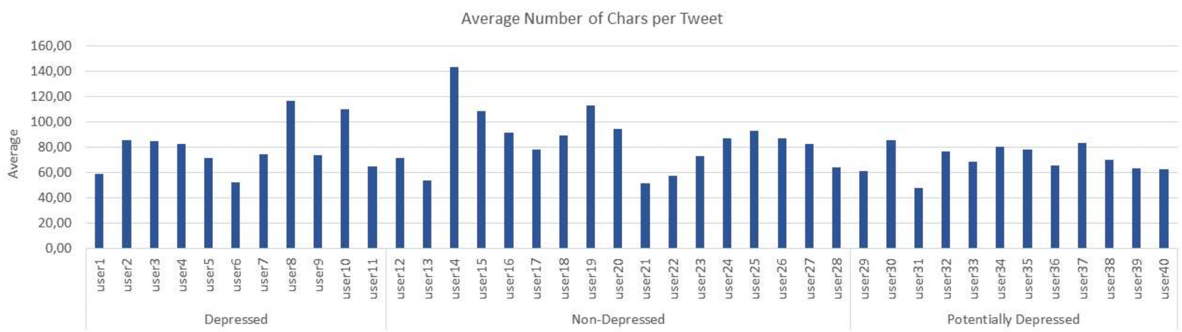

Figure 3.7: Average number of chars

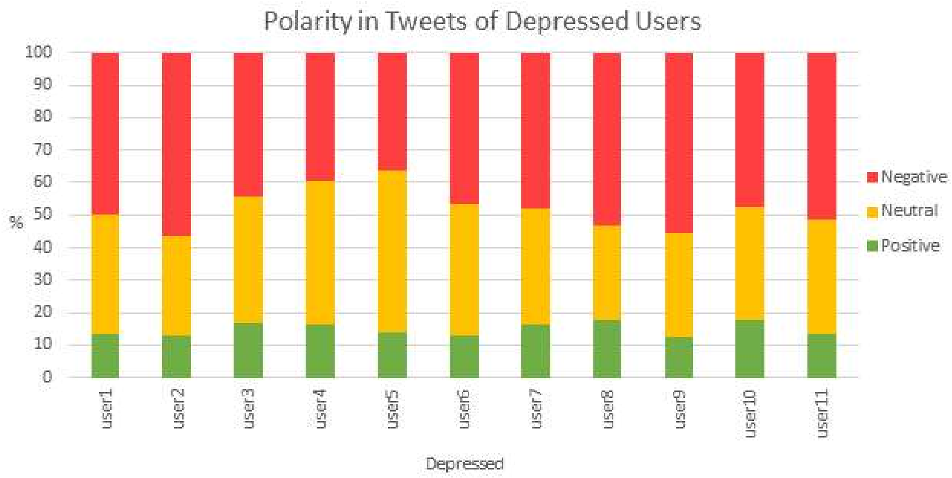

Figure 3.8: Polarity in depressed users

As expected and described in the studies developed by Shen et al. [37], the group of depressed users shows a higher percentage of negative tweets compared to the percentage of positive tweets, as can be seen in Figure 3.8. These values are combined with the fact that, this type of users, express their problems and emotions, most of them negatives, of daily situations they have faced.

For non-depressed users group, it is possible to observe by Figure 3.9 an increase in both percentage of positive and neutral tweets. This increase is due the fact that, users from this group do not expose their feelings but instead, as previously mentioned, reactions from activities that they have attended or participated in.

Once again, the group of potentially depressed users reflects a combination between results obtained in both groups described above. Based on the analysis provided by Figure 3.10, we notice an increase in the percentage of positive tweets compared to the group of depressed users.

The last analysis is related to the eight basic emotions illustrated in Figure 4.6 and, resulting from the words contained in the users' tweets. The colors used to represent the emotions in the graphics are based on the colors used in the Plutchik Wheel of Emotions [28].

As expected from the depressed user group, Figure 3.11 shows a strong presence of negative emotions such as Sadness, Disgust and Anger in approximately 35% of the tweets compared to a low percentage of positive emotions such as Joy, Trust and Surprise.

Figure 3.12, shows that, for the group of non-depressed users, there is a more balanced relationship between all emotions. As previously mentioned, many of these users reveal a

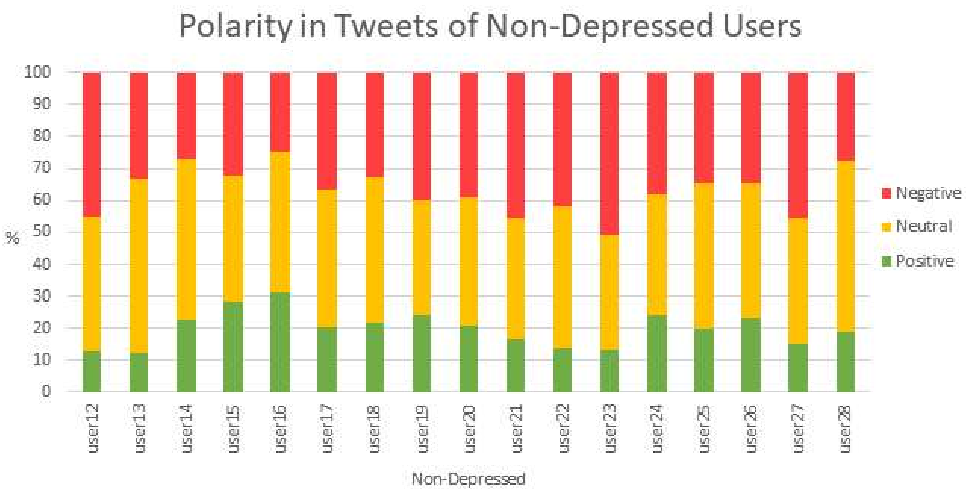

Figure 3.9: Polarity in non-depressed users

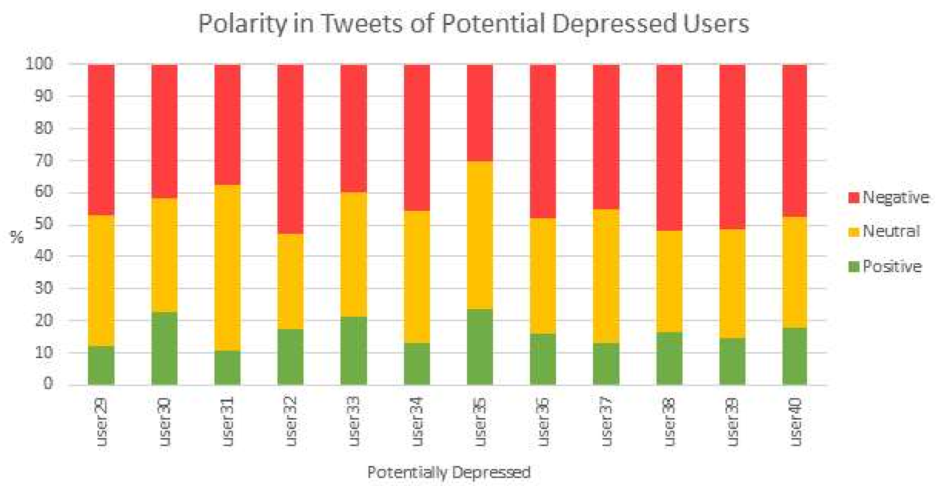

Figure 3.10: Polarity in potentially depressed users

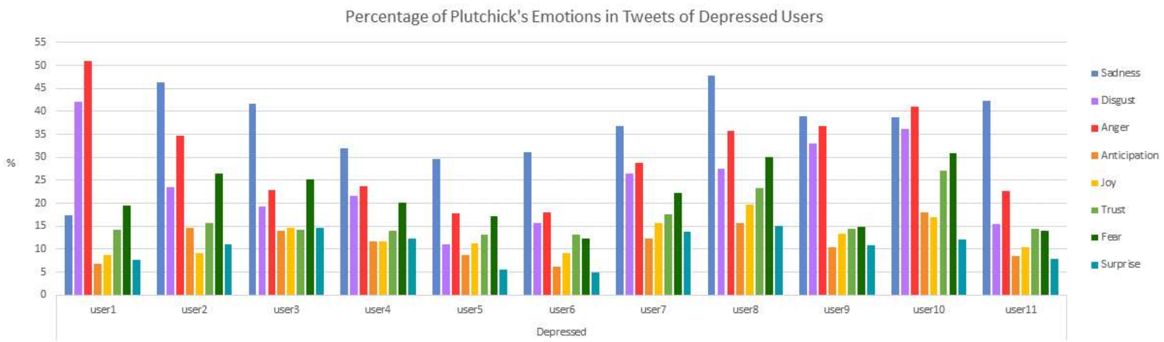

Figure 3.11: Percentage of Plutchik Emotions for depressed users

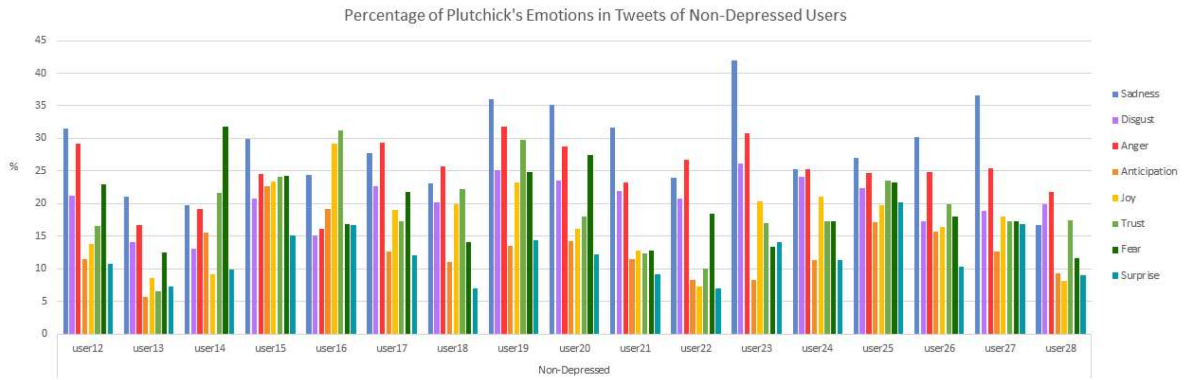

Figure 3.12: Percentage of Plutchik Emotions for non-depressed users

lot of emotions when describing activities that have been involved, for that reason, despite the negative emotions like Sadness and Anger being quite present they are not associated to the normal state of mind of the user, but to a specific event.

Finally, the group of potentially depressed users has once again, an oscillation between the different emotions, as shown in Figure 3.13. In addition to the usual high percentage for Sadness, Disgust and Anger, it is important to highlight that most users of this group, assumed to have episodes of anxiety which justifies the significant presence of Fear emotion, contrary to what happened in the other two groups.

After comparing the three groups, we decided to have the possibility to analyze more closely each user, therefore, we decided to create a website using Flask and Bootstrap. As mentioned before, the purpose of this website was to provide an easier interpretation of the information collected from users if the site is used by a health professional. This way, it is possible to analyze statistics and posts made by the user under analysis.

Once we enter the *Homepage* of the site we come across a field to insert the user Id

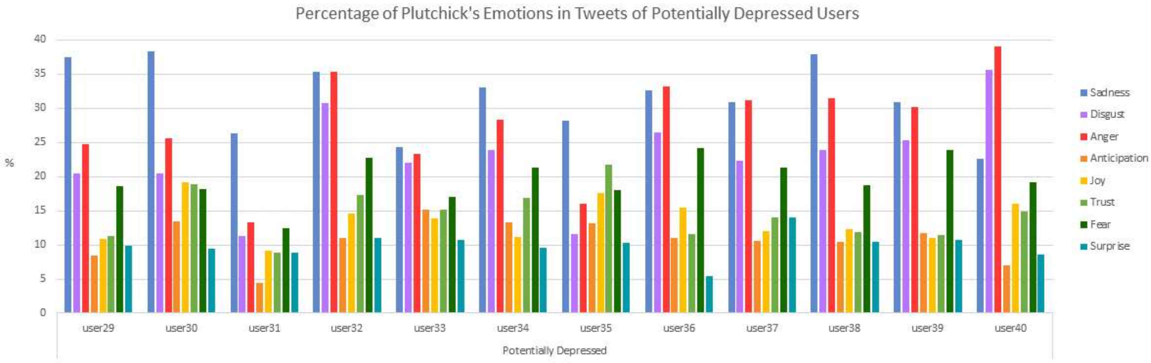

Figure 3.13: Percentage of Plutchik Emotions for potentially depressed users

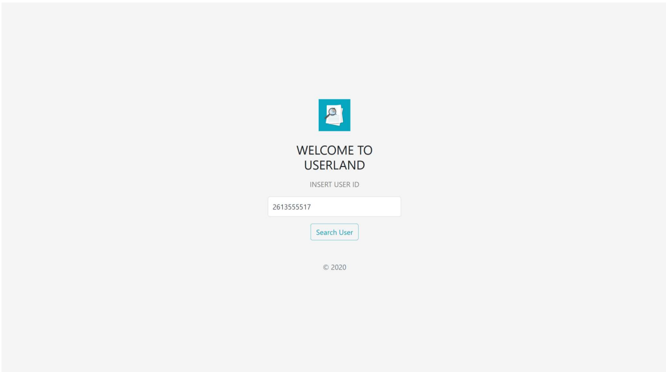

Figure 3.14: Homepage of the site

that we would like to examine. As shown in the Figure 3.14 we will base ourselves on user XXXXXXXX01 for the illustration of the different pages of our site.

After validating the user it is possible to proceed to the page represented on the Figure 3.15, *User Profile*. In this page, we have a brief presentation of the user, his gender, location, tweets collected and statistics obtained through tweets. Based on this example user, we can verify that it is a Female, from Felgueiras, with a total of 199 tweets in our dataset counting 54 words maximum in a post, a minimum of 3 and an average of 18 words per post.

In case we want to see the content of the tweets made by the user, we select the *Tweets* tab and, the website provids a table composed by the content of the post, date and time when they were posted, as shown in the Figure 3.16. It is important to emphasize that, at no time during the visualization of this data was revealed the identity of the user being analyzed or possible users who may have been mentioned in the posts made.

The *Graphics* tab allows us to have a clearer and more accurate perception of the user's state through the different graphs available, addressing different areas, for instance the percentage of tweets associated with an emotion and polarity, the number of tweets

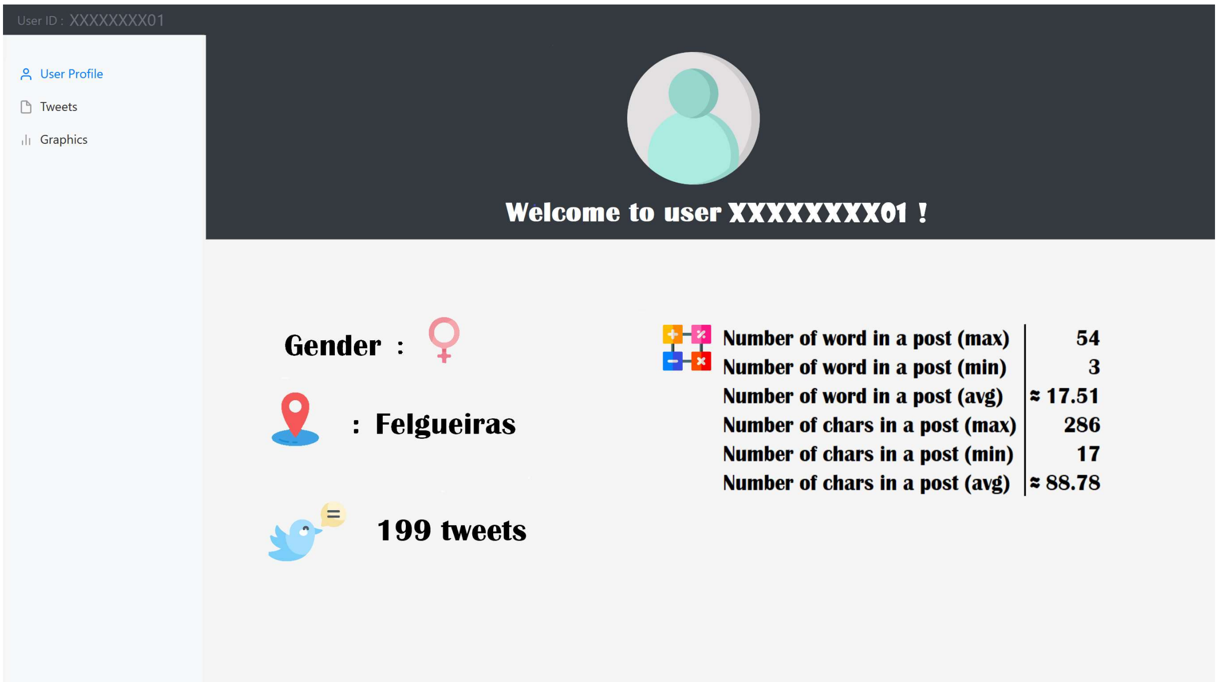

Figure 3.15: User Profile

Figure 3.16: Tweets from the user

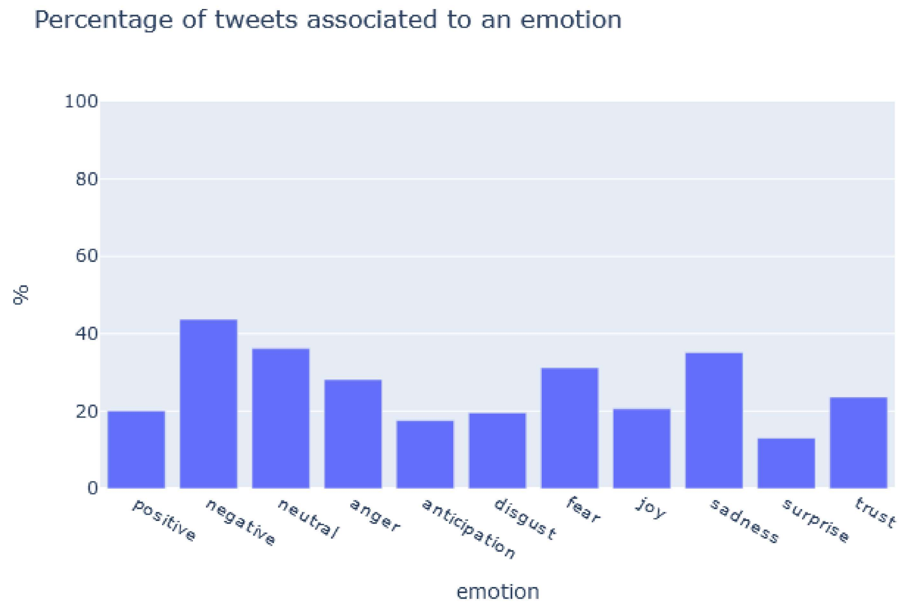

Figure 3.17: Percentage of tweets associated to a polarity and an emotion

made per day, the emotions present in the tweets made and the average of the VAD values given by the tweets made on the day.

As mentioned above, Figure 3.17 shows us the percentage of positive, negative and neutral tweets as well as the percentage of tweets associated to each one of the eight emotions presented. For the example user, we see that there is a higher number of negative tweets combined with anger, fear and sadness emotions.

The following graph gives us a perception of how many tweets the user made during the analysis period, in the case of the example user we can see in Figure 3.18, that the tweets reviewed cover the period from 1st January 2018 to 19th December 2018. We can also observe that during this time our user had a few days in which published more than usual, for example, on 02/14/2018. We highlight the fact that, even if the user had posted daily during this period of time, only those that contained the word depression or related to one of the emotions illustrated in the graphs can be found in our dataset since those are the ones that we are interested to analyze in detail and understand what our user could be feeling.

To complement the information above, Figure 3.19 shows which emotions are present in the tweets. Emotions with positive connotation as, joy, surprise and trust are represented above the x-axis while, emotions with negative connotation are represented below the x-axis, therefore we can have a quick reading of the possible tweet classification and the dominant condition of the user. We realize that in our example user, negative emotions are more predominant compared to positive emotions, giving us an idea that the user is in a phase marked by more negative moments.

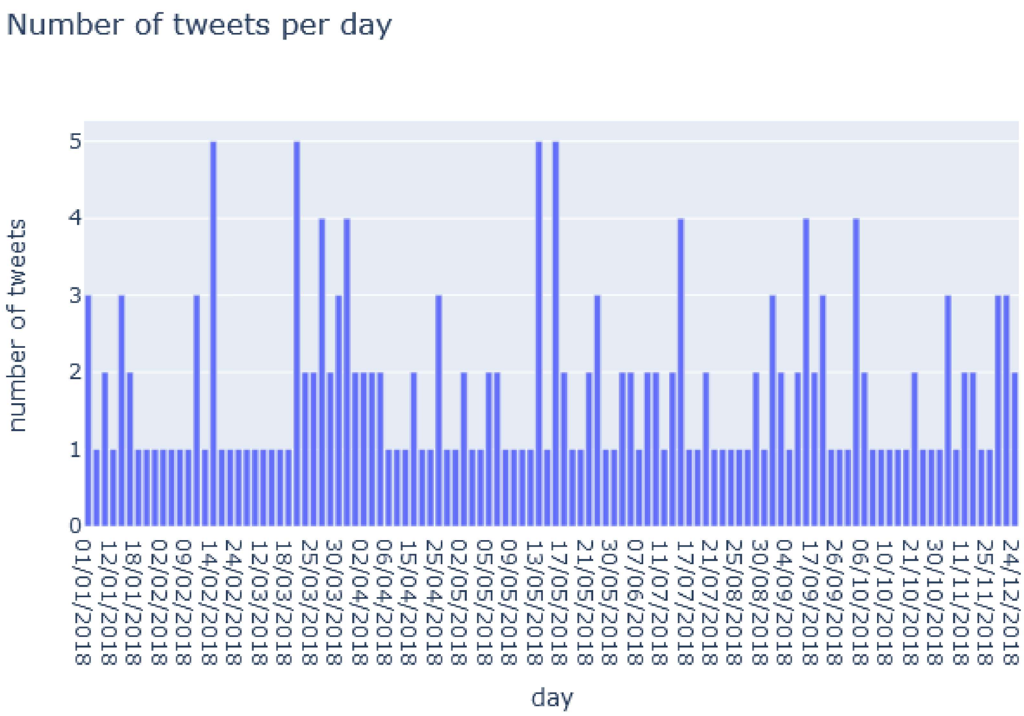

Figure 3.18: Number of tweets per day

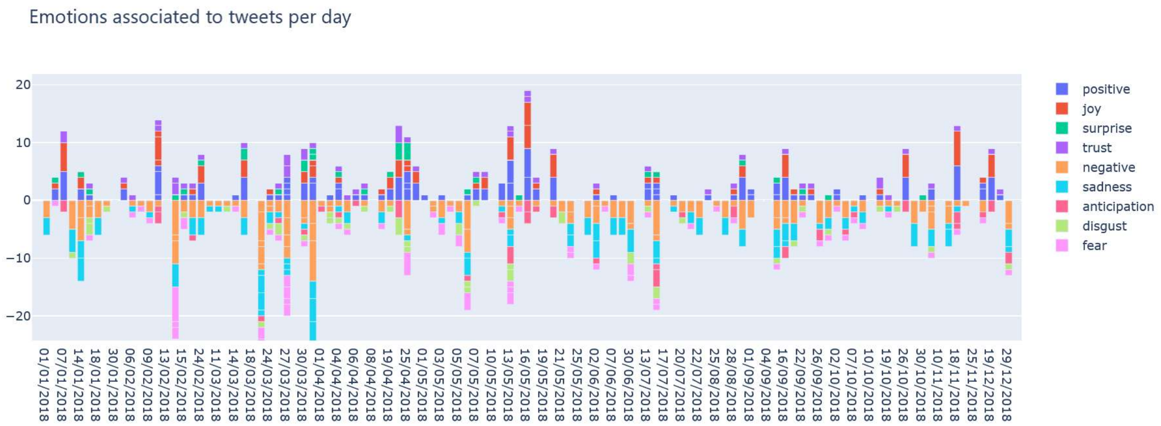

Figure 3.19: Emotions associated to tweets per day

4 Experiments with Reddit

This chapter presents experiments, using data extracted from Reddit, aimed at analyzing whether the implementation of new features in the model previously developed by Losada and Gamallo [26] leads to an improvement in its performance. Section 4.1 explains the criteria taken into account for the extraction and creation of the dataset used in this implementation, followed by the treatment and application of the new features before the training and testing phase in Section 4.2. Then, Section 4.3 introduces the explanation of the type of algorithm chosen and the reasons for this choice. To conclude this chapter, Section 4.4 lists the metrics used to evaluate the performance of the model, ending with the analysis and interpretation of the results obtained by it in Section 4.5.

4.1 Data

During the analysis made to understand the different methods to detect depression in the different social networks, we came across the work of Losada and Gamallo [26], that was applied on Reddit post.

The dataset is composed of posts in English, with 892 users and their respective posts. 137 of the total users has been diagnosed with depression (positive) and the remaining 755 users had no depression (negative). The data extraction of this dataset was done through Reddit's Python API and stored in XML files, as explained in Section 1.3. Each file represents a user, containing their posts and comments as depicted in Figure 4.1.

The file has the following structure :

- ID - anonymous identification of the user;
- Title - title of the post (if empty, it is considered a comment, otherwise, is a post);
- Date - when the post was made;
- Info - source of the post/comment;
- Text - content of the post/comment.

Before starting to pre-process the data we made statistics with the two group of data as it can be visualize in Table 4.1.

```
<INDIVIDUAL>
<ID>test_subject4534</ID>
<TITLE>   </TITLE>
<DATE> 2014-11-21 09:05:11 </DATE>
<INFO> reddit post </INFO>
<TEXT> I feel like my depression comes in waves. </TEXT>
</WRITING>
</INDIVIDUAL>
```

Figure 4.1: Example of a Reddit data

Number of	Depressed	Non-Depressed
Users	54	352
Total posts/comments	18,729	217,701
Posts per user (avg)	346.83	618.47
Words per post (min)	3	3
Words per post (max)	3523	7304
Words per post (avg)	≈ 48.38	≈ 37,73
Chars per post (min)	2	2
Chars per post (max)	19008	39474
Chars per post (avg)	≈ 250.72	≈ 200,92

Table 4.1: Statistics from Reddit dataset

4.2 Setup

Once we had access to the dataset and implementation made by Losada and Crestani [24], we decided to combine the implementation made by us with the one given by them. So, the first step was to ensure that the code and dataset lead to the results presented in the article. However, after some tests and improvement of the code, we realized that the values obtained with the files provided, were not the same as in the article. Therefore, we assume that after the publication of the article the code has been upgraded reaching better results.

To make the analysis of the dataset easier, we started by converting all the information in the files to two data frames, one for the depressed users and the other for non- depressed users, making it easier to read, search and change.

Starting to pre-process the posts, we removed the URLs, tags, punctuation, replaced the acronyms by the corresponding word, for example, we replaced "k" for "ok", and, finally, the treatment of negation, where we considered as negation words, "no", "never", "not". All words after a negation word will be marked by adding the prefix "not_", until one of the following phrase terminators is reached, ',!?:;", 'but', 'however', 'although', 'despite', 'even though', 'even if', 'though' and 'in spite of'. In Figure 4.2 we can see an example of the original post and what is obtained after applying a few of the pre-processing techniques listed above.

Classification is a technique used to categorize data into a certain number of classes so, a classification model is constructed in order to return a prediction based on the input values given for the trained dataset obtained in the training phase. When building a classi-

Original Post	Post with treatment
“ wealth doesn’t equal happiness. ”	“ wealth does not not_equal not_happiness ”

Figure 4.2: Example of Reddit pre-processing

fication model it is important to define which algorithm we intend to use in the experiment according to the type of output we are expecting to have.

In our dissertation, we will use Logistic Regression Algorithm, an algorithm that is a variance of Linear Regression Algorithm and has become the most used method of analysis over the years, according to Hosmer and Lemeshow [18]. Before explain the chosen algorithm is important to clarify the relation with Linear Regression Algorithm and why it cannot be used for our problem. Linear Regression Algorithm is a Supervised Regression Algorithm used to solve regression problems, the numerical prediction made for this type of problem takes into account a certain fixed number of parameters, that depend on the number of input features, with a number between 0 and 1 for the output. For example, predict the price of a trip to Barcelona depending on the season an duration of the trip [47].

However, the principal difference from Logistic Regression Algorithm consists of being a Supervised Classification Algorithm capable of resolving regression but mainly, classification problems. This algorithm calculate the probability of an event returning a categorical output which can be a probability, 1 or 0 , for this prediction is not required to have a direct relationship between dependent and independent variables. This ability of mapping the prediction value is given by the weighted sum of inputs through an activation function, more known as, sigmoid function, as we can see in Figure 4.3 [31], is represented with a S-Curve.

With the help of Figure 4.4, it will be easier to understand the differences mentioned above between these two algorithms.

In our implementation, we used sklearn Linear Model Logistic Regression [13] configured with L1 Regularization in order to avoid overfitting problems of the model, concerning the solver parameter. The ’ liblinear’ option was chosen, which is the most indicated for small datasets and for models with L1 regularization. Finally, the cost given to the two classes (depressed or non-depressed) was defined by the class_weight parameter, being the non-depressed class considered the majority class with a weight of 1/(1+w) and, for the minority class, depressed, a weight of w/(1+w) [26].

For the first experience we decided to apply the NRC Emotion Lexicon. Mohammad and Turney [29] created this lexicon with 14 200 English words using AMT. In 2017, they updated it to be multi-lingual, covering over one hundred languages[1]. Each word present in the lexicon is associated to ten affective categories: ’Anger’, ’Anticipation’, ’Disgust’,

[1] https://saifmohammad.com/WebPages/AccessResource.htm

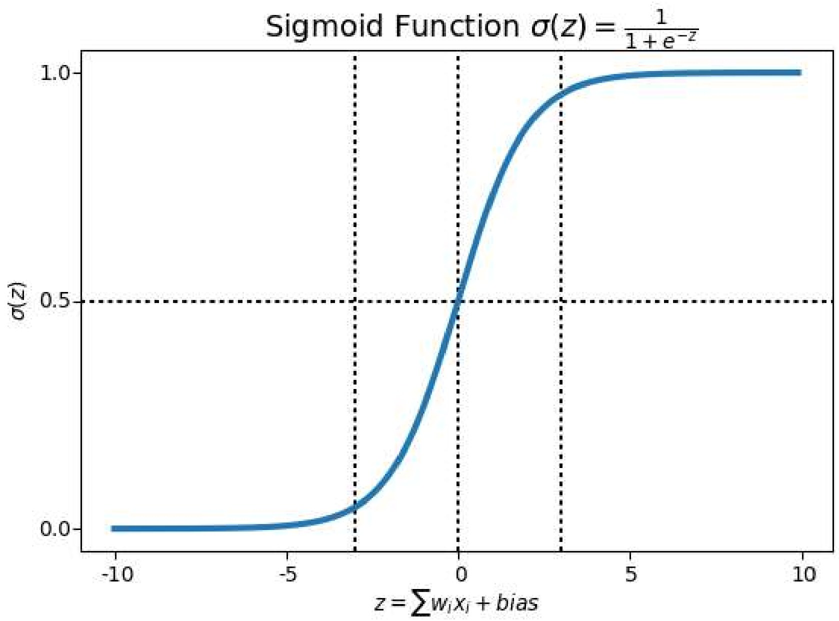

Figure 4.3: Sigmoid Function

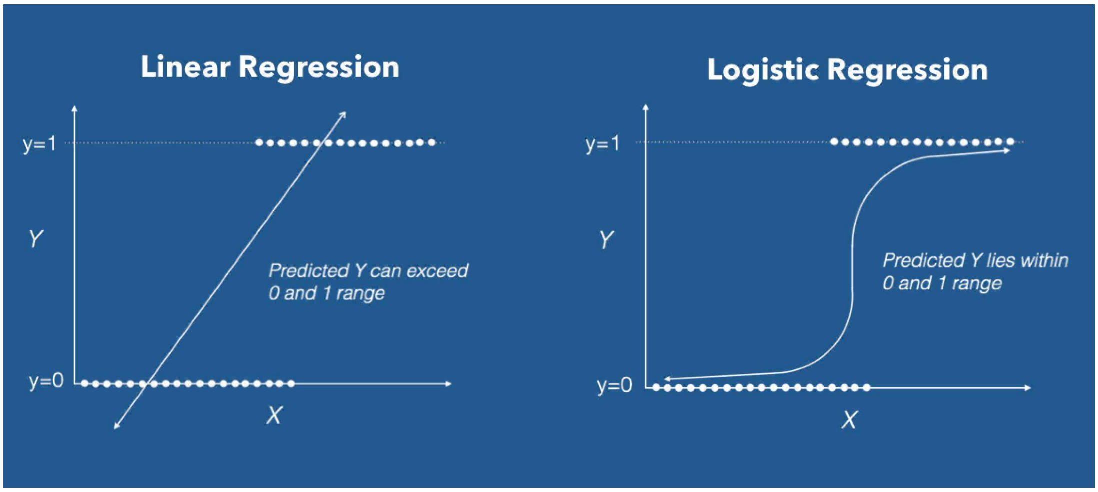

Figure 4.4: Linear Regression Algorithm VS Logistic Regression Algorithm

Structure	Example of word
TargetWord AffectCategory:AssociationFlag AffectCategory:AssociationFlag ...	aborto Positive:0 Negative:1 Anger:0 Anticipation:0 Disgust:1 Fear:1 Joy:0 Sadness:1 Suprise:0 Trust:0

Figure 4.5: Example of NRC Emotion Lexicon word representation

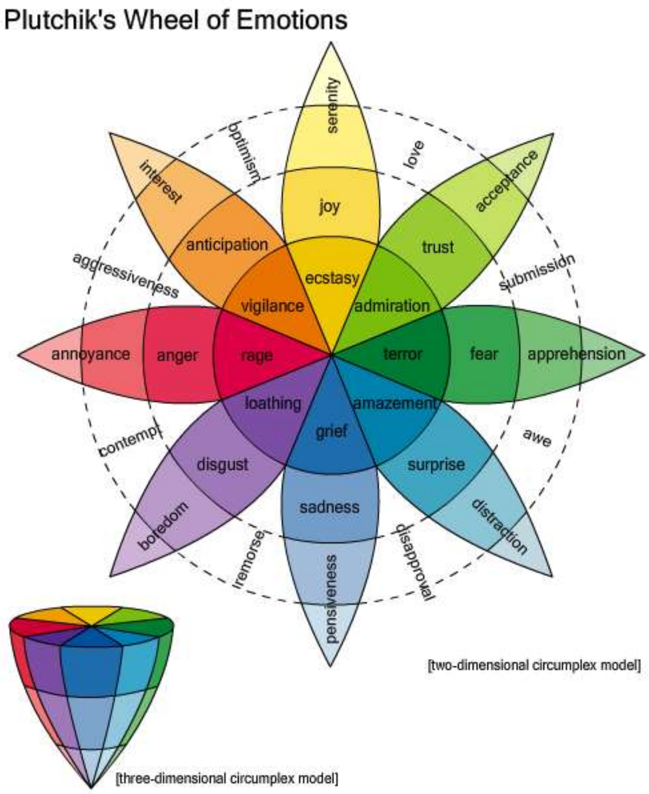

Figure 4.6: Plutchik's Wheel of Emotions

'Fear', 'Joy', 'Sadness', 'Surprise', 'Trust', 'Positive', 'Negative'. Eight of them correspond to Plutchik's basic emotions, illustrated in Figure 4.6 [28], and the two remaining represent polarity, negative or positive. As shown below, each line contains, the TargetWord, the AffectCategory that represents one of the ten possible categories and finally, AssociationFlag, which may have a value of 0 or 1, depending whether the word is associated with the emotion (1) or not (0).

For our approach we used the following information : 'English Word', 'Positive', 'Negative', 'Anger', 'Anticipation', 'Disgust', 'Fear', 'Joy', 'Sadness', 'Surprise', 'Trust'.

To process each post we create a dictionary with the emotions 'Anger', 'Anticipation', 'Disgust', 'Fear', 'Joy', 'Sadness', 'Surprise', 'Trust' and a counter for each emotion. After that, we went through our dataset to see if the words present in the posts were listed in the lexicon, if so, we incremented the counter of the emotion(s) associated to the word

Original Post	Post after applying NRC Emotion Lexicon
“ wealth doesn’t equal happiness. ”	“ wealth does not not_equal not_happiness __TRUST __NOT__ANTICIPATION ”

Figure 4.7: Example of a Reddit data after applying NRC Emotion Lexicon

however, if the word contained the prefix " not_" we would decrement the counter, this way we would try to represent the negation on the emotions.

Once the analysis of the post was finished, we sorted the dictionary according the values obtained on the counters, in descending order. Then, we added the emotion to the end of the post, to be later taken into account by the model, following the the next rules: if the counter value was greater than 0 we added the concatenation of " __" with the emotion; if the counter value was lower than 0 we added a negation by concatenating " __NOT__" to the emotion; finally, when the counter value was zero the emotion related was not added to the post. Figure 4.7 shows an example of the explained procedure.

In the training set, we started by creating a list with the writings of all the 406 users (depressed and non-depressed) that will be the final corpus to be analized.

After that, we applied Term Frequency Inverse Document Frequency, an algorithm used to “convert a collection of raw documents to a matrix of TF-IDF features” [22] that, in our case, was configured to remove the terms that are present in less than 20 documents. Before applying the LR classifier with a L1 regularization we used our corpus to learn the vocabulary and computed a document-term matrix.

For the test set we applied the same procedure applied in the training set, and then executed the four classifiers explained below .

The *Random* classifier, as the name indicates, provides a random decision on whether or not the user is depressed without any text analysis and it is only used as a baseline.

The *Minority* classifier, also does not do any text analysis and classifies all users as depressed.

The classifier *First n* predicts if the user belongs to the depressed or non-depressed groups based on the first *n* user posts. In our case we used 10, 100 and 500 posts. Since in Losada and Crestani [24] work they also classify the user based on all their posts, it is expected that the results using all posts will be better.

Finally, the Dyna*mic* classifier predicts according the thresholds rather than a fixed number of posts. For our study we executed this classifier three times changing the threshold value between 0.5, 0.75 and 0.9. For the prediction of the user group, it was done by concatenating the user posts and, after each post, the depression language classifier returns an output, that is the confidence value. If this value was higher than the threshold,

Original Post	Post after applying Emotion Vector
“ wealth does not equal happiness. ”	“ wealth does not not_equal not_happiness __NOT__HAPPINESS ”

Figure 4.8: Example of a Reddit data after applying emotion vector

the user was considered depressed, if not it keeps concatening user posts.

For the second experiment we repeated the whole process, however, instead of using the NRC Emotion Lexicon we simply decided to have a vector with the stemma following emotions, "anger", "anticipation", "disgust", "fear", "joy", "sad", "surprise", "trust", "happy". For this step we checked if each word in the post, or the Stemming of it, was present in the vector with the emotions. If it was true we added the emotion followed by the concatenation of the prefix " __" with the emotion in Upper Case, in cases of the word negation, we concatenated the negated emotion (i.e. "not_emotion") with " __NOT__ " and the emotion in Upper Case. This approach allowed us to process the emotion with a stronger connotation, contributing to better predict the type of user. In Figure 4.8 we can see an example of the outcome of this experience. In the training and test sets we will reproduce the steps described above.

4.3 Evaluation Model

The model evaluation includes two phases, as explained in Section 4.2, training and testing. In Figure 4.9, it is possible to have a better perception of the steps that need to be taken into account.

At the training phase, a model was created and trained using Logistic Regression Algorithm with an 80% of the dataset as input, which had already included the new features.

In the test phase, the trained model provided in the previous phase, was applied to the remaining 20% of the dataset, generating results for the classification of the user type, depressed or non-depressed.

4.4 Performance Metrics

After assigning the classification to the users from the model in the test phase, we take into account the three metrics listed, F1-Measure, Precision, Recall which help us analyze and understand the performance of the model created. However, it is important to clarify the variables involved before explaining the meaning and how these metrics are determined, using the Figure 4.10 [38] is possible to make a clearer and more perceptible explanation.

The *Precision* metric refers to the correct predictions made by the model over the total number of positive predictions performed, and is given by Equation 4.1.

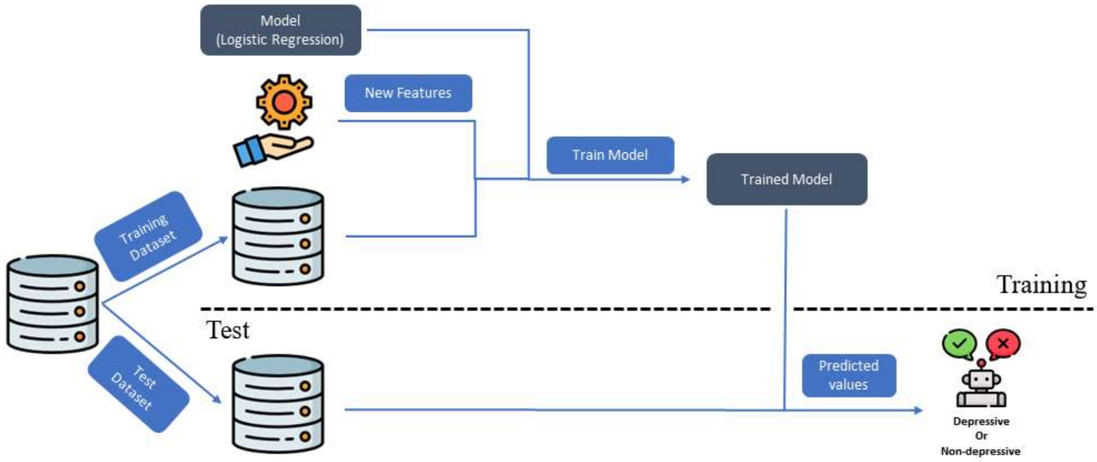

Figure 4.9: Model Evaluation

		Predicted	
		Negative	**Positive**
Actual	**Negative**	True Negative	False Positive
	Positive	False Negative	True Positive

Figure 4.10: Confusion matrix

$$precision = \frac{True\,Positives}{True\,Positives + False\,Positives} \tag{4.1}$$

The *Recall* metric, also known as Sensitivity, uses the TP and FN to determine the percentage of TP corrects compared to all positive cases as shown by Equation 4.2.

$$recall = \frac{True\,Positives}{True\,Positives + False\,Negatives} \tag{4.2}$$

The *F1-Measure* metric, also known as F1-score, is the weighted average between Precision and Recall as shown by Equation 4.3.

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{4.3}$$

Finally, ERDE metric, evaluates the speed of a true positive calculation, ERDE5 evaluated after 5 user posts and ERDE50 after 50 user posts. Contrary to the other metrics the lower the value the better.

4.5 Results

In this section we analyzed the results obtained by the three different approaches described above, the result in the article, first experience with the NRC Emotion Lexicon and the second experience with the vector of emotions.

In Table 4.2 we have the results obtained from the different approaches made throughout this study. The first two classifiers, Random and Minority, as expected, were the ones who obtained the worst results in metrics since they relied only in the first post to make their prediction.

As for the strategy that achieved the best performance, we concluded that the First 500 reached the highest value of F1 and the lowest of ERDE showing that it can take the right predictions in a shorter time. However, regarding the improvements obtained having as comparison the results obtained in the article [26], we concluded that both for the emotion vector and the NRC lexicon, the *Dynamic 0.9* strategy, had an increase in the number of correct predictions through the improvement of F1 and, at the same time, there was a decrease in the response time as can be seen in the reduction of the value of the percentage of ERDE.

	F1	P	R	ERDE5	ERDE50
RANDOM					
Losada and Crestani [24]	0.215	0.134	0.537	12.30%	12.28%
Vector Emotions	0.215	0.134	0.537	12.30%	12.28%
NRC Emotion Lexicon	0.215	0.134	0.537	12.30%	12.28%
MINORITY					
Losada and Crestani [24]	0.235	0.133	1.0	11.56%	11.53%
Vector Emotions	0.235	0.133	1.0	11.56%	11.53%
NRC Emotion Lexicon	0.235	0.133	1.0	11.56%	11.53%
FIRST 10					
Losada and Crestani [24]	0.333	0.667	0.222	10.93%	10.67%
Vector Emotions	0.265	0.643	0.167	11.54%	11.35%
NRC Emotion Lexicon	0.314	0.688	0.204	11.12%	10.89%
FIRST 100					
Losada and Crestani [24]	0.766	0.774	0.759	4.94%	4.55%
Vector Emotions	0.712	0.740	0.685	5.83%	5.43%
NRC Emotion Lexicon	0.704	0.704	0.704	5.71%	5.32%
FIRST 500					
Losada and Crestani [24]	**0.843**	**0.761**	**0.944**	**2.93%**	**2.54%**
Vector Emotions	0.824	0.754	0.907	3.36%	2.97%
NRC Emotion Lexicon	0.810	0.758	0.870	3.76%	3.36%
DYNAMIC 0.5					
Losada and Crestani [24]	0.741	0.596	0.982	3.16%	2.77%
Vector Emotions	0.706	0.546	1.0	3.24%	2.85%
NRC Emotion Lexicon	0.703	0.553	0.963	3.57%	3.18%
DYNAMIC 0.75					
Losada and Crestani [24]	0.748	0.705	0.7963	4.71%	4.35%
Vector Emotions	0.689	0.631	0.7593	5.33%	5.00%
NRC Emotion Lexicon	0.661	0.623	0.6852	6.05%	5.76%
DYNAMIC 0.9					
Losada and Crestani [24]	0.475	0.731	0.352	9.47%	9.31%
Vector Emotions	**0.533**	0.667	**0.444**	**8.57%**	**8.34%**
NRC Emotion Lexicon	**0.529**	0.697	**0.426**	**8.72%**	**8.49%**

Table 4.2: Results from Reddit analysis

5 Conclusions and Future Work

For this study we applied two lexicons and developed a model to detect and predict users on social networks that contain psychological signals of depression based on the historical of posts made with the aim of understanding if lexicons had some impact in that prediction.

To achieve this goal we made experiences using two different social networks, Twitter and Reddit. For both of them we had to obtain a good dataset with posts from users that showed signs of depression as well as posts from users that had never been diagnosed with depression or had episodes related to it.

Regarding Twitter social network, we obtained our data through Twitter API by imposing as collection criteria, the 40 users who contained more than 255 posts, whose tweets were in portuguese and contained the words 'medo', 'assust', 'morrer', 'chor', 'aguent', 'sofr', 'infeli', 'vida' or 'odeio'. After processing of the data, the experimental phase is based on the implementation of two lexicons, VAD Lexicon and NRC Emotion Lexicon. We could conclude that the use of the VAD lexicon was not advantageous for the prediction of depressed and non-depressed users, once it was not possible to correlate the VAD values with the sentiments present, even taking into account, the representation of the VAD values from the six most relevant emotions illustrated in Figure 3.2. However, using the website it was possible to analyze the user chronologically and understand whether or not he was a depressive user.

In the case of the Reddit platform, the dataset used was provided by the authors Losada and Gamallo [26], which already included posts by depressed and non-depressed users, as well as their classification, positive or negative. After the treatment of the data and considering the conclusions obtained in the approach described above, we excluded the VAD lexicon in the processing phase and implemented only the NRC Emotion Lexicon and added an experiment with a vector composed by eight emotions. In this approach we concluded that although the inclusion of the lexicon and the vector of emotions into the model has not resulted in the best prediction model, we were able to observe improvements in one of the strategies implemented, showing that the lexicon and the vector have positively influenced the effectiveness and efficiency of the model.

Throughout the development and test phases of this dissertation we identified some limitations and alternative algorithms that can be implemented in the future, with the purpose of improving our developments.

As already mentioned, the main constraint for the model's development when implemented on the Twitter social network, was not having the assessment by a health professional of the manual classification given to the 40 users of our dataset. Taking this into account, as a future work we should include a health professional to help with the interpretation of the posts, since this analysis can provide more reliable results and perceive whether the implementation made for Reddit would achieve good results here.

As for the development made for the Reddit social network, as a future work another types of lexicons should be added, such as ANEW, which is very popular and used in most of the studies already mentioned in Section 2.2. Other improvement would be to use different classifiers, such as SVM and Naive Bayes, which have a historical background of achieving good accuracy and performance values [34, 1].

Bibliography

[1] Maryam Mohammed Aldarwish and Hafiz Farooq Ahmad. “Predicting Depression Levels Using Social Media Posts”. In: *2017 IEEE 13th International Symposium on Autonomous Decentralized System (ISADS)*. 2017, pp. 277–280. DOI: 10.1109/ISADS.2017.41.

[2] Hayda Almeida, Antoine Briand, and Marie-Jean Meurs. “Detecting Early Risk of Depression from Social Media User-generated Content”. In: (2017), p. 12.

[3] Oana Bălan et al. “Emotion Classification Based on Biophysical Signals and Machine Learning Techniques”. In: *Symmetry* 12.1 (2019), p. 21. ISSN: 2073-8994. DOI: 10.3390/sym12010021. URL: https://www.mdpi.com/2073-8994/12/1/21.

[4] Dani Beck. *World Mental Health Day: facts, figures and the flaws of academia*. 2019. URL: https://www.neuro-central.com/world-mental-health-day-facts-figures-flaws-academia/.

[5] Martin Beck. *How to Scrape More Information From Tweets on Twitter*. Oct. 2020. URL: https://towardsdatascience.com/how-to-scrape-more-information-from-tweets-on-twitter-44fd540b8a1f.

[6] Farah Benamara et al. “Automatic Detection of Depressive Users in Social Media”. In: (2018), p. 16.

[7] Munmun De Choudhury et al. “Predicting Depression via Social Media”. In: *Seventh International AAAI Conference on Weblogs and Social Media*. 2013. URL: https://www.aaai.org/ocs/index.php/ICWSM/ICWSM13/paper/view/6124.

[8] - Conselho Nacional de Saúde. *Sem mais tempo a perder – Saúde mental em Portugal: um desafio para a próxima década*. 2019.

[9] Glen Coppersmith, Mark Dredze, and Craig Harman. “Quantifying Mental Health Signals in Twitter”. In: *Proceedings of the Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality*. Baltimore, Maryland, USA: Association for Computational Linguistics, 2014, pp. 51–60. DOI: 10.3115/v1/W14-3207. URL: http://aclweb.org/anthology/W14-3207.

[10] Munmun De Choudhury, Scott Counts, and Eric Horvitz. “Social Media As a Measurement Tool of Depression in Populations”. In: *Proceedings of the 5th Annual ACM Web Science Conference*. WebSci ’13. event-place: Paris, France. New York, NY, USA: ACM, 2013, pp. 47–56. ISBN: 978-1-4503-1889-1. DOI: 10.1145/2464464.2464480. URL: http://doi.acm.org/10.1145/2464464.2464480.

[11] Amandeep Dhir et al. “Online social media fatigue and psychological wellbeing—A study of compulsive use, fear of missing out, fatigue, anxiety and depression”. In: *International Journal of Information Management* 40 (2018), pp. 141–152. ISSN: 0268-4012. DOI: 10.1016/j.ijinfomgt.2018.01.012. URL: http://www.sciencedirect.com/science/article/pii/S0268401217310629 (visited on 10/23/2020).

[12] - Economic and Social Research Council. *What is Twitter and why should you use it?* URL: https://esrc.ukri.org/research/impact-toolkit/social-media/twitter/what-is-twitter/.

[13] Thomas J. Fan. *sklearn Logistic Regression — scikit-learn 0.23.2 documentation*. URL: https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LogisticRegression.html.

[14] Lena Firestone. *Which is Worst for Your Mental Health: Instagram, Facebook or YouTube?* 2017. URL: https://www.psychalive.org/worst-mental-health-instagram-facebook-youtube/.

[15] Global Burden of Disease Study 2017 (GBD 2017) Results. Seattle Global Burden of Disease Collaborative Network United States: Institute for Health Metrics and Evaluation (IHME). *Number of people with depression by country*. 2018. URL: https://ourworldindata.org/grapher/number-with-depression-by-country.

[16] Global Burden of Disease Study 2017 (GBD 2017) Results. Seattle Global Burden of Disease Collaborative Network United States: Institute for Health Metrics and Evaluation (IHME). *Prevalence by mental and substance use disorder*. 2018. URL: https://ourworldindata.org/grapher/prevalence-by-mental-and-substance-use-disorder.

[17] Grant H. *4 Social Media Platforms You Might Be (Wrongfully) Ignoring*. 2020. URL: https://www.lemonlight.com/blog/are-you-taking-full-advantage-of-social-media/.

[18] David W. Hosmer and Stanley Lemeshow. *Applied Logistic Regression*. 2º Edição. URL: http://resource.heartonline.cn/20150528/1_3kOQSTg.pdf.

[19] Md. Rafiqul Islam et al. “Depression detection from social network data using machine learning techniques”. In: *Health Inf Sci Syst* 6.1 (2018), p. 8. ISSN: 2047-2501. DOI: 10.1007/s13755-018-0046-0. URL: http://link.springer.com/10.1007/s13755-018-0046-0.

[20] Joshua A. Israel. “Remission in depression: definition and initial treatment approaches”. In: *J Psychopharmacol* 20.3_suppl (2006). Publisher: SAGE Publications Ltd STM, pp. 5–10. ISSN: 0269-8811. DOI: 10.1177/1359786806064306. URL: https://doi.org/10.1177/1359786806064306 (visited on 10/12/2020).

[21] Matthew Jockers. *Extracts Sentiment and Sentiment-Derived Plot Arcs from Text*. 2017. URL: https://cran.r-project.org/web/packages/syuzhet/vignettes/syuzhet-vignette.html.

[22] Guillaume Lemaitre. *sklearn TfidfVectorizer — scikit-learn 0.23.2 documentation*. URL: https://scikit-learn.org/stable/modules/generated/sklearn.feature_extraction.text.TfidfVectorizer.html.

[23] Minglei Li, Qin Lu, and Yunfei Long. "Are Manually Prepared Affective Lexicons Really Useful for Sentiment Analysis". In: (2017), p. 5.

[24] David E. Losada and Fabio Crestani. "A Test Collection for Research on Depression and Language Use". In: *Experimental IR Meets Multilinguality, Multimodality, and Interaction*. Ed. by Norbert Fuhr et al. Vol. 9822. Series Title: Lecture Notes in Computer Science. Cham: Springer International Publishing, 2016, pp. 28–39. ISBN: 978-3-319-44563-2 978-3-319-44564-9. DOI: 10.1007/978-3-319-44564-9_3. URL: http://link.springer.com/10.1007/978-3-319-44564-9_3.

[25] David E Losada, Fabio Crestani, and Javier Parapar. "CLEF 2017 eRisk Overview: Early Risk Prediction on the Internet: Experimental Foundations". In: (2017), p. 18.

[26] David E. Losada and Pablo Gamallo. "Evaluating and improving lexical resources for detecting signs of depression in text". In: *Lang Resources & Evaluation* 54.1 (2020), pp. 1–24. ISSN: 1574-020X, 1574-0218. DOI: 10.1007/s10579-018-9423-1. URL: http://link.springer.com/10.1007/s10579-018-9423-1.

[27] Saif Mohammad. "Obtaining Reliable Human Ratings of Valence, Arousal, and Dominance for 20,000 English Words". In: *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Melbourne, Australia: Association for Computational Linguistics, 2018, pp. 174–184. DOI: 10.18653/v1/P18-1017. URL: https://www.aclweb.org/anthology/P18-1017 (visited on 02/17/2020).

[28] Saif M. Mohammad and Peter D. Turney. "Crowdsourcing a Word-Emotion Association Lexicon". In: *arXiv:1308.6297 [cs]* (2013). arXiv: 1308.6297. URL: http://arxiv.org/abs/1308.6297 (visited on 10/26/2020).

[29] Saif M. Mohammad and Peter D. Turney. "National Research Council Canada Ottawa, Ontario, Canada, K1A 0R6". In: (2010), p. 234.

[30] - Observatório Português dos Sistemas de Saúde. *Relatório de Primavera 2019*. pt-PT. Library Catalog: opss.pt. 2019. URL: http://opss.pt/relatorios/relatorio-primavera-2019/.

[31] Ayush Pant. *Introduction to Logistic Regression*. 2019. URL: https://towardsdatascience.com/introduction-to-logistic-regression-66248243c148.

[32] Minsu Park, David W McDonald, and Meeyoung Cha. "Perception Differences between the Depressed and Non-Depressed Users in Twitter". In: (2013), p. 10.

[33] Inna Pirina and Çağrı Çöltekin. "Identifying Depression on Reddit: The Effect of Training Data". In: *Proceedings of the 2018 EMNLP Workshop SMM4H: The 3rd Social Media Mining for Health Applications Workshop & Shared Task*. Brussels, Belgium: Association for Computational Linguistics, 2018, pp. 9–12. DOI: 10.18653/v1/W18-5903. URL: http://aclweb.org/anthology/W18-5903.

[34] Bayu Yudha Pratama and Riyanarto Sarno. "Personality classification based on Twitter text using Naive Bayes, KNN and SVM". In: *2015 International Conference on Data and Software Engineering (ICoDSE)*. 2015, pp. 170–174. DOI: 10.1109/ICODSE.2015.7436992.

[35] Stephanie Rude, Eva-Maria Gortner, and James Pennebaker. "Language use of depressed and depression-vulnerable college students". In: *Cognition & Emotion* 18.8 (2004), pp. 1121–1133. ISSN: 0269-9931, 1464-0600. DOI: 10.1080/02699930441000030. URL: http://www.tandfonline.com/doi/abs/10.1080/02699930441000030.

[36] Lindsay H. Shaw and Larry M. Gant. "In Defense of the Internet: The Relationship between Internet Communication and Depression, Loneliness, Self-Esteem, and Perceived Social Support". In: *CyberPsychology & Behavior* 5.2 (2002), pp. 157–171. ISSN: 1094-9313, 1557-8364. DOI: 10.1089/109493102753770552. URL: http://www.liebertpub.com/doi/10.1089/109493102753770552.

[37] Guangyao Shen et al. "Depression Detection via Harvesting Social Media: A Multimodal Dictionary Learning Solution". In: *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence*. Melbourne, Australia: International Joint Conferences on Artificial Intelligence Organization, 2017, pp. 3838–3844. ISBN: 978-0-9992411-0-3. DOI: 10.24963/ijcai.2017/536. URL: https://www.ijcai.org/proceedings/2017/536.

[38] Koo Ping Shung. *Accuracy, Precision, Recall or F1?* 2020. URL: https://towardsdatascience.com/accuracy-precision-recall-or-f1-331fb37c5cb9.

[39] Andrew Solomon. *Depression, the secret we share*. 2013. URL: https://www.ted.com/talks/andrew_solomon_depression_the_secret_we_share.

[40] SPPSM. *Perturbações depressivas ou depressões*. URL: https://www.sppsm.org/informemente/guia-essencial-para-jornalistas/perturbacoes-depressivas-ou-depressoes/.

[41] Jini Jojo Stephen and Prabu P. "Detecting the magnitude of depression in Twitter users using sentiment analysis". In: *IJECE* 9.4 (2019), p. 3247. ISSN: 2088-8708, 2088-8708. DOI: 10.11591/ijece.v9i4.pp3247-3255. URL: http://ijece.iaescore.com/index.php/IJECE/article/view/16306 (visited on 09/27/2020).

[42] Michael M. Tadesse et al. "Detection of Depression-Related Posts in Reddit Social Media Forum". In: *IEEE Access* 7 (2019), pp. 44883–44893. ISSN: 2169-3536. DOI: 10.1109/ACCESS.2019.2909180. URL: https://ieeexplore.ieee.org/document/8681445/.

[43] Gilbert Tanner. *Scraping Reddit data*. URL: https://gilberttanner.com/.

[44] - The American Psychiatric Association. *What Is Depression?* URL: https://www.psychiatry.org/patients-families/depression/what-is-depression.

[45] - The National Institute of Mental Health. *Depression*. 2018. URL: https://www.nimh.nih.gov/health/topics/depression/index.shtml.

[46] - U.S. Department of Health & Human Services. *What Is Mental Health?* URL: https://www.mentalhealth.gov/basics/what-is-mental-health.

[47] Jaime Zornoza. *Linear Regression Explained*. 2020. URL: https://towardsdatascience.com/linear-regression-explained-d0a1068accb9.