SVM Based Learning System for the Detection of Depression in Social Networks

Juan Arturo Pérez Cebreros, Eduardo Vázquez Fernández, Alma Partida Herrera, Geovani Peña Ramírez
Instituto Politécnico Nacional, Escuela Superior de Ingeniería Mecánica y Eléctrica Culhuacán, Departamento de Ingeniería en Computación, México
{perezcebreros, partidaherreraalma, geovpe}@gmail.com, eduardovf@hotmail.com

Abstract. Depression represents a problem of public concern that is now prioritized in many health care agendas with the intention of preventing future suicides, which have devastating impact not only because of tragic loss of life, but also for the grieving family and friends. Investigations in each country reveal a reduction in physical and mental well-being; for this reason, the proposal presented in this article comprises an attempt to detect the feelings expressed in text sentences presented in social networks.

Keywords. Social networks, depression, machine learning.

1 Introduction

Today, the world is going through a time of transformation; daily life has made a 360 degree turn, where the protagonist, a viral strain known as SARS-CoV-2 has caused somewhat more than four million deaths. Apart from the economic consequences, social confinement is usually an unpleasant experience, which can lead to different stress factors that generate mental health outcomes [1]. Because this situation is new and very much expanding, it is still premature to estimate the emotional consequences of the epidemic outbreak. However, research carried out in [2,3] suggests that fear of the unknown and uncertainty can lead to the evolution of different mental health diseases such as: stress disorders, anxiety, depression, somatization and degenerative behaviors, resulting in an increase in the consumption of alcohol, tobacco and other substances harmful to health [4]. In particular, people with chronic illnesses are expected to have higher levels of psychological symptoms [5]. Older people are also predicted to be more psychologically vulnerable than young people in this crisis [6].

This project has emerged because of the great problem of suicide cases in our country among young people [7,8,9]. For this reason, we have decided to develop a tool that is capable of alerting possible cases so that they can be contained.

1.1 Depression and Artificial Intelligence

Notably in Mexico, it was found that young adults (that is, between 15 and 25 years old) present suicidal ideas and show greater depressive states, meaning depression is evident in 67.3% of those who have attempted suicide and in 81.1% of those who manifest suicidal ideas [10]. Likewise, people with mental illness tend to disclose their mental condition on social media, as a way of seeking relief [11].

However, research on employing social media, as a means to understanding behavioral health disorders, is still in its infancy.

In [12], web activity patterns of university students were analyzed, as they may indicate depression.

Similarly, in [13] they showed that Facebook status updates may reveal symptoms of depressive episodes. Some differences have been noted, such as the fact that depressed users more frequently use first person pronouns, [14] as well as words indicating negative emotions and anger.
For this reason, depression has been associated with the use of linguistic markers such as greater use of first person pronouns. Many other studies of language and depression have been limited to clinical settings, and therefore to analyzing spontaneous speech or written essays.

Following this lead, some research [15, 16] has proposed innovative methodologies to amass textual content shared by people diagnosed with depression. However, there are no publicly available sources. This is because the text is often taken from social networking sites such as Twitter or Facebook that do not permit redistribution [17]. Hence, these previous studies direct us towards detecting depression in social networks as the first step against suicide. The main procedure in mental health studies using social networks has traditionally been carried out by applying surveys, where the number of users is limited by those who manage to complete the survey.

For example, in [18] Twitter users were requested to take a Center for Epidemiological Studies Depression Scale (CES-D) and share their profile with the public. This type of study has produced high quality data; however it is limited in size and scope. Therefore, in this research we will examine depression by considering automatically obtained samples from large amounts of Twitter data. The Internet has allowed us to follow the evolution of language and is providing us with a very accessible medium for people to express their feelings anonymously.

Hence, we have adapted the method in [15] for the construction of this data set in Spanish, we will proceed to identify self-expressions of diagnoses of mental illnesses and we take advantage of these messages to build our data set.

1.2 Analysis of Sentiments

Generally, the word feeling refers to a way of thinking (opinion) or sentiment (emotion) about something [20]. One of the best known tasks involved in sentiment analysis (SA) is the detection of polarity that can be synthesized in the following classification problem: “Given a text T as input, depending on its content, determine whether T contains a positive/negative/neutral opinion (and eventually determine a force parameter that indicates how much of the content is positive/negative). This sentiment analysis task is widely used in contexts such as the review of products or services, political predictions, among others [19].

Recently, the main focus of another line of research is medical and psychological, where the task of emotion recognition is carried out on forums, chats, and social networks. Data on social networks such as Twitter and Facebook, where users post their reactions and comments in real time, pose new and different challenges.

First of all, it seems that studies can be divided into supervised and lexical-based methods. The supervised methods are based on training classifiers, such as Naive Bayes, Vector Support Machines, Random Forest, whilst, lexicon-based methods determine the tendency of the sentiment in a text based on the use of pre-established lexicons of previously weighted words, in terms of the feelings expressed [20].

Similarly, the analysis of real-time data from social networks (SMA), has received considerable attention in recent years in the context of analysis for the detection of abnormal events/activities. Today, we can say that there are millions of Twitter posts, millions of Facebook posts and billions of forums on web pages and various documents that can be reviewed in order to determine the opinions expressed behind the words.

2 Related work

Originally, SA was associated with business intelligence [21, 23], but it has spread to other areas such as politics [23, 24], medicine [25], education [26], recommendations [27, 28], screening for plagiarism [29], news influences [30], deception detection [31, 32], irony detection [33] and account classification [34], among others.

In particular, SA is a prominent research topic in the field of computational linguistics. Tasks include classifying the polarity of sentiment expressed in text (e.g. positive, negative, and neutral), identifying the target/theme of sentiment, and identifying the sentiment in terms of various aspects of a theme.

The sentiment polarity classification problem is often modeled as two-way (positive / negative) or three-way (positive / negative / neutral) [35]. It is
important to note that this task of detection and classification is not easy, firstly because tweets are short messages where the indicators of depression tend to manifest themselves in a very subtle way.

Due to the widespread adoption of social media and the availability of large-scale data from social media, approaches to using this data for screening for depression are receiving increased attention from researchers.

In [22], it is shown that college students show symptoms that indicate depression on Facebook.

In [27], the differences between Twitter users with and without depression are analyzed by analyzing their activities.

In [28], a similar analysis is performed by analyzing Facebook data, using multiple regression analysis.

Some recent studies have submitted encouraging results for the detection of users suffering from depression, but more studies are still required [36,37].

In our research, we have formulated models in a bidirectional way (positive/negative); leaving to the future, the task of evaluating greater intensities of depressive sentiment: strong positives, strong negatives, mild positives, and mild negatives.

Detecting sentiments from the phrases is a complicated task, for example: "life is like jazz, better if it is improvised"; the sentiment of the opinion is positive because the word "life" implies something good.

However, the same word in another context, as shown in the following statement: 'my life is meaningless', implies a negative feeling - it is bad because negation reduces the positive in the word 'life'. Thus the problem involves the use of language, which is a very complex and huge problem.

### 3 Methodology

To solve this problem, a three-phase model was proposed:

#### 3.1 Collection phase

During this phase, we took advantage of the large amount of data provided by Twitter. The collection method is based on two main stages: first, the tweets are filtered out from regular expressions and then these are classified into: negative and positive.

To acquire the tweets for this study, we developed an application that uses the Twitter search API [21].

To filter tweets that are not written in Spanish, we used the free language detection library [22]. This library is based on Bayesian filters and has a precision of 0.99 in the detection of the 53 languages to which it permits entry. The tweets were acquired over 180 days (from December 1, 2020 to June 1, 2021), producing data sets with approximately 3800 tweets for Spanish.

To generate data from a set of tweets with depressive traits, we considered tweets from people who declared to have been diagnosed with the disease of depression.

Table 1 shows the regular expressions used to detect people who refer to depression in their tweets; however the main intention is to identify people who make a direct and open statement that they were diagnosed with the disease of depression.

Subsequently, the tweets are extracted from the list of people who asserted through a tweet that they suffer from this disease.

#### 3.2 Preprocessing phase

Data preprocessing is an often neglected but important stage in the process. It involves techniques to transform the raw data into a more understandable format. The main ones are data cleansing, data integration, data transformation and data reduction.

As apparent in Figure 1, our preprocessing mechanism includes:

a. Extraction,

<table>
<thead>
<tr>
<th>Table 1. Regular expressions for tweet detection</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Word</strong></td>
</tr>
<tr>
<td>Depression</td>
</tr>
<tr>
<td>Associated phrases</td>
</tr>
</tbody>
</table>

ISSN 2007-9737
b. Elimination of numbers and URLs that may have an effect on our analysis but do reduce noise and our efficiency [23].

c. Elimination of stop words such as articles, pronouns, and prepositions [24].

d. Word derivation, which is used to transform different word forms into a standard root form [25].

In this phase, in addition to these techniques, we incorporate a weighting step using the Term Frequency-Inverse Document (TF-IDF) algorithm. The TF-IDF reflects the importance of a word in a document; and this level of importance increases when the word appears many times, to the point that we can determine the themes that are trending [26].

Term Frequency (TF) is the frequency with which words appear in a document. The term $t_i$ in a document can be formulated as follows:

$$Tf_{i,j} = n_{i,j}.$$  \(1\)

In (1), we have that $n_{i,j}$ is the number of the word $t_i$ occurs in the document $d_j$. In contrast, Inverted Document Frequency (IDF) measures the overall importance of a word in a document. We can formulate this in the following way:

$$idf_{i,j} = \log \frac{D}{df_{i,j}}.$$  \(2\)

In (2), we have that $D$ is the total number of text documents and $df_{i,j}$ is the number of documents $d_j$ which contain the term $t_i$.

Finally, we have that TF-IDF is a combination of TF and IDF; the formula would look like this:

$$Tf-idf_{i,j} = Tf_{i,j} \times idf_{i,j}.$$  \(3\)

### 3.3 Identification/classification phase

The classification algorithm based on vector support machines (SVM) is a supervised learning machine, which requires training data and test data. This consists in finding an optimal hyperplane as the function that separates two types of data. The classification with the lowest error is the one obtained from the hyperplane that maximizes the margin, that is, whose distance between the plane and the support vectors is as large as possible. Despite its simplicity, this has proven to be a robust...
algorithm that generalizes well to real-life problems [44-48].

4 Results

The proposed method involves the classification and identification of tweets that allows us to have an accurate and direct visualization, it can determine whether the phrase that was extracted from Twitter is indicative of depression or not, making it possible to help the person who is in need.

In Figure 2, we can visualize the problem of analyzing the messages posted on Twitter in terms of the sentiments these messages express. Here, our first task was to tag a set of tweets in Spanish, obtained using the methodology described. Likewise, when we label, it is important to consider the presence of negation, because negation plays a very important role when detecting the polarity of a message (positives become negative and vice versa).

This classification is not a trivial task and one of the characteristics of Twitter is that it is a type of informal communication, with limitations in length. This makes it different from previous research on sentiment analysis, using conventional texts.

Table 2 shows the ten words with the highest positive and negative frequency, respectively. Notably, the word ‘life’ appears both on the positive and negative sides. Later, in Table 3, we explain this change in polarity.

Because of the limitations of this work, it will be necessary to do more studies to reduce dispersion, in order to apply semantic smoothing techniques, among others [49].

The results obtained using the Bayesian classifier and vector support machines were compared using the following metrics: accuracy, precision and sensitivity where,

\[
\text{Accuracy} = \frac{TP + TN}{TP + TN + FN + FP} \quad (4)
\]

Positive sensitivity (5) and negative sensitivity (6) is the sensitivity ratio and is calculated as follows:

\[
\text{Sensitivity p} = \frac{TP}{(TP + FN)}, \quad (5)
\]

\[
\text{Sensitivity n} = \frac{TN}{(FP + TN)} \quad (6)
\]

Positive precision (7) and negative precision (8) is the precision ratio and is calculated as follows:

\[
\text{Precision p} = \frac{TP}{(TP + FP)}, \quad (7)
\]

\[
\text{Precision n} = \frac{TN}{(FN + TN)} \quad (8)
\]
In Table 3, four tweets extracted from the data set are presented; we can see that negation plays a very important role in detecting the polarity of a sentence (positives become negative and vice versa), in addition to negation, adjectives that accompany the noun and that change their quality must be considered.

Table 4 shows the performance comparison between the Bayesian classifier and vector support machines, respectively, in terms of precision and sensitivity. Similarly, Table 5 shows the performance of the classifiers in terms of accuracy.

5 Conclusions

The increasing trend for depression and suicide represents a serious public health problem. Undoubtedly, this is a problem that the Mexican health system must face with urgency, firstly considering that the country is in a stage of economic uncertainty (derived from the current pandemic), and secondly, because there are evident mental health care needs.

Our method can provide the basis for more social computing studies and opens the doors to future research on AI algorithms that make use of other training data of the multifactorial and multilevel type, such as social, economic and political variables. In order to explore mental health, the central idea of this research starts from the principle of classifying a text as positive, or negative, using AI algorithms. As a first step, we describe a methodology from which to generate a data set in Spanish and using this, some essential steps are established for the classification of depressive traits.

We have applied the Bayesian classifier and the vector support machine classifier to classify texts with depressive features, obtaining very good results.

In future works, we will proceed to increase the size of the data set using the methodology described. We will also analyze different techniques for representing texts, for example we will incorporate a dimensionality reduction through a bag of words model (BOW). We could also combine our algorithms with multimodal information so as to offer a new dimension to traditional text analysis, where we could take into account different modalities such as visual and audio data, among others [50, 52]. Likewise, we can incorporate deep learning techniques into our method, using hierarchical architectures to increase scalability and precision [53, 54].

References


53, No. 8, pp. 5559–5588. DOI: 10.1007/s10462-020-09828-3.


Article received on 29/07/2021; accepted on 30/09/2021.
Corresponding author is Eduardo Vázquez Fernández.