# A sentiment analysis approach to increase authorship identification

Ricardo Martins, José João Almeida, Pedro Henriques, Paulo Novais
*Algoritmi Centre / Department of Informatics*
*University of Minho, Braga - Portugal*
*ricardo.martins@algoritmi.uminho.pt, {jj, prh, pjon}@di.uminho.pt*

Keywords: Sentiment Analysis, Machine Learning, Natural Language Processing

Abstract: Writing style is considered the manner in which an author expresses his thoughts, influenced by language characteristics, period, school, or nation. Often, this writing style can identify the author. One of the most famous examples comes from 1914 in Portuguese literature. With Fernando Pessoa and his heteronyms Alberto Caeiro, Álvaro de Campos and Ricardo Reis, who had completely different writing styles, led people to believe that they were different individuals. Currently, the discussion of authorship identification is more relevant because of the considerable amount of widespread fake news in social media, in which it is hard to identify who authored a text and even a simple quote can impact the public image of an author, especially if these texts or quotes are from politicians. This paper presents a process to analyse the emotion contained in social media messages such as Facebook to identify the author's emotional profile and use it to improve the ability to predict the author of the message. Using preprocessing techniques, lexicon-based approaches and machine learning, we achieved an authorship identification improvement of approximately 5% in the whole dataset and more than 50% in specific authors when considering the emotional profile on the writing style, thus increasing the ability to identify the author of a text by considering only the author's emotional profile, previously detected from prior texts.

## 1 Introduction

Since Barack Obama's election, politicians have been using social media to maintain direct contact with the voters, using it to increase their credibility through this direct channel that includes photos, posts, and comments. Social media enables a correct perception by the voters about politics, creating opinions about the subjects they consider essential. This phenomenon is increasingly turning politicians into digital influencers. Thus, the way politicians communicate on social media can be considered their "personal brand"; thus, their concerns about how they are interpreted are crucial.

With extensive information from social media, digital influencers and their followers validate, reinforce, and amplify news, which is often faked. As the primary objective of these individuals is to be "liked, loved, and shared," it is essential to correctly choose the words contained in their texts to maximize the sentiment raised in the readers. Thus, the emotional characteristics contained in the messages make up an "emotional profile" about the author, which, along with the words used in the text, helps to determine the message's author profile while writing. For example, the following posts are from different authors, but the theme is the same: the Paris Climate Agreement; however, the writing styles are different and arouse different emotions. While the first uses positive and negative words in the text ("fight", "force", "progress"), the second uses mostly words with negative emotions ("hurt", "stop", "needs"):

> "Today marks a crucial step forward in the fight against climate change, as the historic Paris Climate Agreement officially enters into force. Let's keep pushing for progress" (Barack Obama);
>
> "I'm optimistic we can stop climate change and help those who are being hurt the most by it—all while meeting the world's energy needs" (Bill Gates).

### 1.1 Plagiarism

When we think about author identification, the first idea for using this tool is to detect plagiarism. [19] define plagiarism as "theft of intellectual property". The definition expanded by [14] to include different types of plagiarism, such as copy-paste plagiarism, paraphrasing and translated plagiarism, among others.

However, detecting plagiarism is not easy to perform automatically, and there are several works regarding this issue, such as the approach presented by [26], that identifies plagiarism through a framework designed for this purpose. [30] presented a plagiarism detector that uses the Levenshtein distance to identify plagiarism.

The intention of this work is not to detect plagiarism but to create an alternative to increasing authorship identification of text in a non-contextual comparison with painting in which the artist who painted a painting that was unknown by the experts around the world is identified through the techniques and artistic characteristics.

### 1.2 Fake News

Another critical issue that authorship identification raises is about fake news. [1] defined fake news as "to be news articles that are intentionally and verifiably false, and could mislead readers." Considering this definition, a necessary implication of fake news is the massive dissemination of misattributed texts or quotes. For example, Figure 1 presents fake quotes; in some situations, the identification of a fake text is trivial because we know the author's profile; however, there are situations in which it is difficult to identify whether the quote is fake.

### 1.3 Objectives

In this paper, we extended the work presented by [17], which consists of an approach using the author emotional profile, aiming to improve the authorship identification. The expansion includes a new syntactical analysis section that analyses the syntactical writing style for each author and provides more in-depth explanations about the processes applied.

To achieve this purpose, we adopted Plutchik's model to represent emotions because we consider it to be more realistic, easy to use, and this model allows us to represent several different emotions through dyad emotions. Moreover, there are some libraries and lexicons used in this work that represent and process emotions according to this model.

The remainder of this paper is organized as follows. In Section 2, we introduce the concept of emotion and present some theories about emotion representation and analysis, and introduce the concept of the emotional lexicon in Section 3. Section 4 presents some work in this area to detect emotion from social media. Section 5 describes our proposal explaining the steps used in our analysis and discusses the results obtained from a set of tests performed. The paper concludes in Section 6 with the conclusion and future work.

## 2 Emotional theories

Historically, several models have been created to systematize the emergence of emotions and their associated behaviours and discuss how our cognitive system elicits emotions. The primary research theories are discrete, dimensional and appraisal theories.

**Discrete emotional theories** propose the existence of basic emotions (*happiness*, *anger*, *sadness*, *surprise*, *disgust*, and *fear*, for instance) that are universally displayed and recognized. Discrete models group emotions into categories and assume that they are independent. In the literature, among the discrete models, a well-known model is basic emotions, proposed by [8]. This model proposes the existence of six basic emotions: happiness, sadness, fear, anger, surprise, and disgust. One of the main advantages of discrete models is that, through psychophysical experiments, the perception of emotions by human beings is discrete. [24], claims that any sentiment is composed of a set of 8 basic emotions: *Anger*, *Anticipation*, *Disgust*, *Fear*, *Joy*, *Sadness*, *Surprise* and *Trust*, and can be represented as a "wheel of emotions". Additionally, the combination of basic emotions results in *dyad* emotions, enabling the representation of any emotion as a combination of basic emotions.

In contrast to this view, **dimensional theories** attempt to explain emotions regarding two or three dimensions. The most common dimensional characterization of emotions uses two dimensions: arousal and valence. Valence is related to positive or negative evaluation and is associated with the feeling state of pleasure (vs. displeasure). Arousal reflects the general degree of intensity felt. Low arousal is associated with less energy and high arousal with more energy. However, using this two-dimensional approach is challenging for differentiating emotions that share the same values of valence and arousal, such as *anger* and *fear*.

For this reason, it is common to adopt a third dimension to support this differentiation. According to [12] "the third view emphasises the distinct component of emotions, and is often named the componential view". For [27], "the components are appraisals, subjective feelings, physiological changes, motor expressions, and action tendencies. Appraisals are regarded as driving changes in the other emotion components leading to full-blown emotions when the different components are synchronised."

Emotional-cognitive psychologists focus their studies mainly on the **appraisal process**. According to [27], the central idea is that emotions are triggered and differentiated by a subjective analysis of an event, situation or object. This cognitive assessment performed personally is called an appraisal. For instance, Paul and John are watching a basketball game where their favourite teams

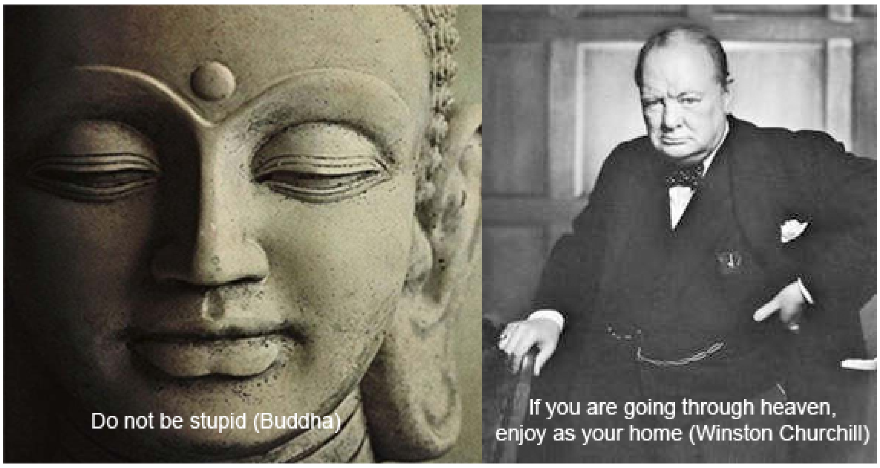


Figure 1: Fake quotes examples

are playing. John's team wins (event). Paul's appraisal is that an undesirable event happened: Paul's team lost, and he is sad. For John, the situation's appraisal is that the event is desirable, and he is happy. Therefore, emotion and reason are not disconnected. In fact, emotions require cognitive processes to generate or retrieve preferences and meanings. Emotions are triggered by the personal interpretation of the annoying or cheerful aspects of an event, the appraisal. Moreover, the appraisal is a cognitive process that triggers emotions.

Despite different theories, they have in common the sense of positive and negative emotions. According to [6], "polarity detection is a popular Natural Language Processing (NLP) task focused on the binary classification of snippets of text into either positive or negative". In other words, polarities are the scores associated with positive, negative and neutral parts of a sentence.

## 3 Lexicons & Emotional Lexicons

For [10], in NLP-context, "a lexicon is a component of a system that contains information (semantic and/or grammatical) about words or expressions, whereas the term dictionary usually refers to objects (printed books or electronic) intended for human readers, but also accessible by computers". For example, when searching in the Wordnet lexicon [20], the word "kill", there are three different meanings for the noun "kill" and fifteen different meanings for the verb "kill", with examples of sentences using the word, in addition to their synonyms.

The standard construction of a lexicon is an arduous task due to the large volume of information and the amount of time that is spent to carry out the steps. For this task, there are efforts in the creation of lexicons through computational techniques, for instance, as presented by [23]. Another method for the construction of computational lexicons is the analysis and improvement of existing lexicons.

The initial point of any approach to study emotion in a text is the use of specific affective lexicons. [7] presented one of the first studies targeting the problem of the referential structure of the affective lexicon. Additionally, [22] argued that affective lexicons should not only contain terms related to emotion but must also contain other terms and affective conditions (affection, mood and sentiment). Terms such as "affection" and "emotion" are sometimes used as synonyms. The distinction occurs when the term affection refers to anything whose valence value is positive or negative. Affection has a broader category when compared to emotion. Types of affective conditions cause emotions, but not all affective conditions are emotions. For example, children prefer to eat fries as opposed to another kind of food (sometimes even more tasty food). This preference is affective and is not an emotion, even when it is the cause of a heated emotional interchange. According to [22], "affect is a very general category of which emotion is a relatively small part. Emotions are particular kinds of affective conditions; so that all emotions are affective conditions, but not all affective conditions are emotions".

At the beginning of studies in affective lexicons, [3] analysed the data selected and considered having affective connotations from [2]. The objective was to develop a method called "semantics", which would map a universe of words with affective characteristics. However, not all words that have affectivity were included in the study that "justify that any division between affective concepts is necessarily vague and arbitrary" [3].

Despite the existence of several well-known emotional lexicon, as WordNet Affect [29], SentiWordNet [9] and ANEW [5], the EmoLex lexicon [21] was used in this work. In addition to being the most recent lexicon, the choice for this lexicon is justified by its structure that links each word to the existence or nonexistence of each Plutchik's basic emotion, creating a referential to analyse the sentences computing the sum of emotions for each word individually. According to the author, each word was analysed using a Mechanic Turk to classify whether the word contains some of Plutchik's basic emotions and its polarities. It is important to emphasize that any word can contain more than one basic emotion. Table 1 is a short fragment of EmoLex lexicon that illustrates the words and their associated emotions.

# 4 Related work

Despite the vast number of works using sentiment analysis, none of them considers the author's emotional profile as a component of the writing style for authorship identification. Thus, each work cited below has partially inspired our work, as will be mentioned.

The work of [28] inspired the usage of emotions in social media, which predicts the individual happiness, as measured by a life satisfaction scale, through the language people used on social media. This prediction is made using randomly selected posts from Facebook and a lexicon-based approach to identify the text words polarity. Moreover, [4] have presented another exciting work involving lexicons and ontologies to extract emotions including sadness, happiness, surprise, fear and anger, which contributed to the emotional profile creation.

The framework developed for authorship identifying based on online messages presented by [31] considers features as syntactic, lexical, structural and content-specific that contribute to the use of machine learning techniques to predict authorship.

# 5 Methods

To predict the authors of a post based on the emotion contained in the text, we collected 2,100 Facebook posts from 8 different authors from different areas, such as politics, business, entertainment, and sports. All data were collected during the same period, reducing temporal situations interference in the text emotions. To compare all the information, we manually labelled the posts into two categories: politicians and non-politicians.

The task of predicting the author of a text is composed of several intermediate steps. First, some preprocessing tasks were needed to reduce the data size by removing unnecessary text from the original message.

## 5.1 Preprocessing

Preprocessing is a significant step in text mining processes and applications. It is the first step not only for text mining approaches but also in data mining. Several preprocessing techniques are used to extract information from text, and their usage is based on the characteristics of the desired information. Although some techniques were created in data mining, they are used in text mining approaches since the same technique can be applied for both information extraction, information retrieval, or combined.

In this work, preprocessing after tokenization was divided into three parallel tasks, as presented in Figure 2: part-of-speech tagging (POS-T), named entity recognition (NER) and stopwords removal. We chose this strategy because both POS-T and NER need the text in the original format to return the correct data from the analysis. Later, the intersection of three task outcomes is stemmed, creating the preprocessed file used to analyse the emotions.

An example of text preprocessing using a real post from Barack Obama is presented in Figure 3.

In the sequel, a more in-depth view of each task in the preprocessing pipeline.

### 5.1.1 Part of speech tagging (POS-T)

The POS-T process identifies the textual grammatical structure of a sentence. Through a grammatical analysis, each word in the sentence is labelled according to its respective grammatical category. For example, using the Stanford Core NLP [15] to analyse the sentence grammatically "Four little monkeys jump on the bed", the result is "Four/**CD** little/**JJ** monkeys/**NNS** jump/**VBP** on/**IN** the/**DT** bed/**NN**", where:

$CD = cardinal$
$JJ = adjective$
$NNS = nounplural$
$VBP = verbinthethirdpersonofsingularpresent$
$IN = prepositionorsubordinatingconjunction$
$DT = determiner$
$NN = noun$

In our tests, concerning text cleaning, **the POS-T process removes all grammatical categories different than nouns, verbs, adverbs, and adjectives**, which is important because only these grammatical categories provide emotional information. More formally, the tok-

Table 1: EmoLex lexicon words examples

| **Word** | **Positive** | **Negative** | **Anger** | **Anticipation** | **Disgust** | **Fear** | **Joy** | **Sadness** | **Surprise** | **Trust** |
|---|---|---|---|---|---|---|---|---|---|---|
| aback | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| abandon | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 0 |
| abandoned | 0 | 1 | 1 | 0 | 0 | 1 | 0 | 1 | 0 | 0 |
| hate | 0 | 1 | 1 | 0 | 1 | 1 | 0 | 1 | 0 | 0 |
| love | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 |
| majesty | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| prepared | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 |
| punch | 0 | 1 | 1 | 0 | 0 | 1 | 0 | 1 | 1 | 0 |
| wonderful | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 1 |

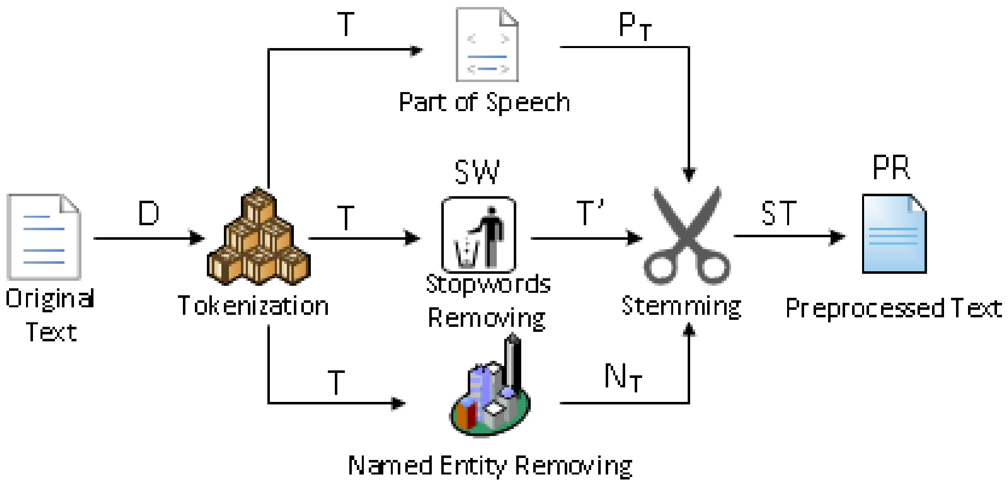


Figure 2: Preprocessing tasks

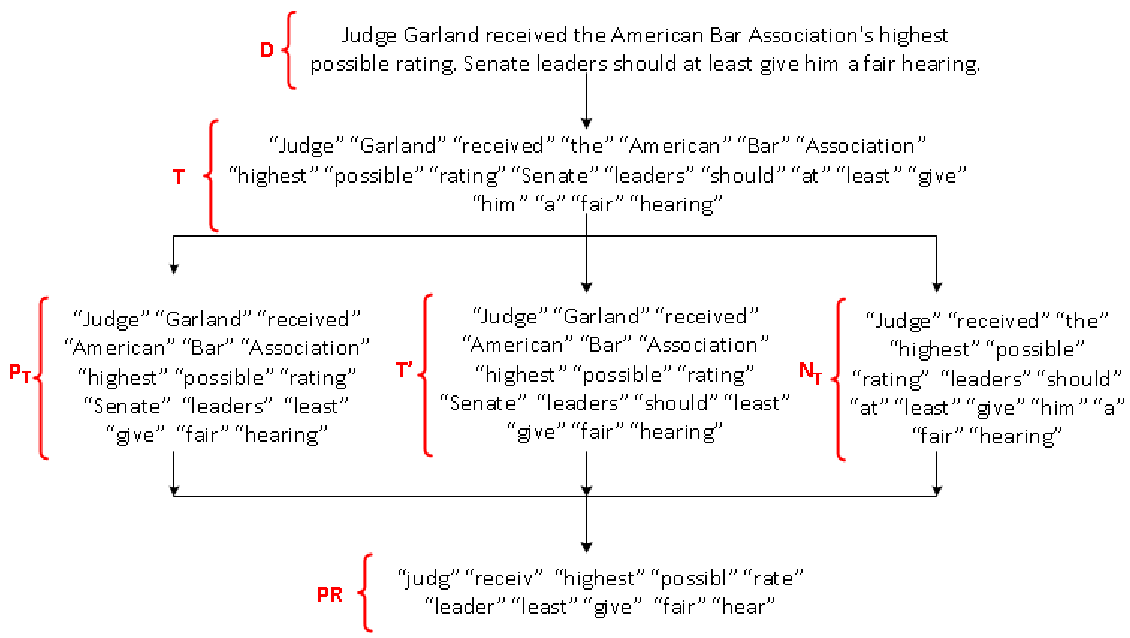


Figure 3: Preprocessing text example

enization process converts the original text $D$ into a set of tokens $T = \{t_1, t_2, ..., t_n\}$ where each element contained in $T$ is part of the original document $D$. POS-T process labels each token with semantic information. Later, a process collects all nouns, verbs, adverbs and adjectives in a set $P_T$, where $P_T = \{p_{T_1}, p_{T_2}, ..., p_{T_k}\}$ and $0 \leq k \leq n$ and $P_T \subset T$.

#### 5.1.2 Named Entity Recognition (NER)

Named entity recognition (NER) is a process that considers a string of text (sentence or paragraph) as input to identify relevant names (people, places, and organizations) mentioned in that text. This process is essential to avoid misunderstandings between names and nouns. For example, during text preprocessing without an NER step, the USA city named *Riverside* can be confused with the noun *riverside*, leading to a different understanding.

In our tests, to avoid emotional bias for places, names, and organizations, once identified tokens in one of these categories, they are removed from the text. Using the Stanford Core NLP [15], all sentences were analysed and, similar to POS-T, NER labels were added to the text. For example, the sentence "Trump will make America great again" produced as a result "Trump/PERSON will/O make/O America/LOCATION great/O again/O". Therefore, in this case, the tokens "Trump" and "America" were discarded because they correspond to people and place, respectively, as presented in Figure 4, while the tokens labelled with the tag /O remain because they were not identified as a name, person or organization and are kept to the next step.

In a formal definition, a set $N_T = \{n_{(T_1)}, n_{(T_2)}, ..., n_{(T_j)}\}$ is constructed based on identified word category and where $\forall j, cat(N_j) = "O"$. A mandatory requirement for NER step is that it must be done in parallel with POS-T because some locations can be confused with nouns (as Long Beach).

#### 5.1.3 Stopwords Removal

The stopwords removal process is a task that checks the existence of predefined (and not allowed) words in the text. In the case of existence, the process removes these words from the texts.

In a formal definition, this task is based on a personal predefined set $SW = \{sw_1, sw_2, ...sw_y\}$ of stop words [1], manually created according to several similar lists available on the internet. This step will return a set $T' = t'_1, t'_2, ..., t'_n$, where $T' \cap SW = \emptyset$.

#### 5.1.4 Stemming

After the 3 preprocessing tasks finish, the outcoming set $ST$ is defined as $ST = T' \cap P_T \cap N_T$.

Later, a stemming algorithm is responsible for obtaining the stem of a word, which is its morphological root, through clearing the parts of the word that are grammatical or lexical information, considering all inflected words as only one, and producing the preprocessed text. For this task, an implementation of the Lovins stemmer [13], resulting in a set of stemmed words $PR = \{ST_1, ST_2, ..., ST_z\}$ ready to be analysed.

[1] stop words are words that are filtered out before or after processing of natural language data (text)

### 5.2 Syntactical analysis

To know the writing style of each author, an approach used identifies how the author expresses their texts syntactically, i.e., how is each sentence from the author syntactically composed. To achieve this objective, the authors had their non-processed texts labelled according to the Part of Speech Penn Treebank [16] tags using Stanford Core NLP. It is important to emphasize that the text must be analysed before the preprocessing due to the words being deleted in each process creating an impact on the grammatical analysis. Moreover, we considered only nouns, verbs, adverbs and adjectives for this analysis, and all subcategories were identified as their "chunk" (for example, "NNS" - noun, plural -, "NNP" - proper noun singular -, "NNPS" - proper noun, plural - were identified as "NN" - noun).

After all text analyses, it was possible to determine the grammatical style of each author, according to Table 2.

When applying the Pearson's correlation coefficient ($r^2$) between each author, it is possible to verify that they are strongly correlated, as presented in Table 3.

After analysing these data, it is possible to notice that in general, non-politicians use nouns more frequently than politicians. When correlating the percentages of nouns, adjectives, verbs, and adverbs between all authors, the values obtained represent a very strong correlation between the authors. This means that, statistically, the writing styles of all authors are very similar; however, despite the strong correlation, in general, non-politicians have higher correlation values between non-politicians, while politicians have higher correlation values between politicians, i.e., the grammatical writing style inside the two groups (politicians and non-politicians) is higher than outside the groups.

### 5.3 Polarity analysis

The first analysis was aimed at determining the posts polarities. To achieve this objective, after the preprocessing, all sentences contained in $PR$ were compared against the EmoLex lexicon [21] to identify the positive and negative words contained in the text. This analysis did not consider the intensity of the polarities or the emotions.

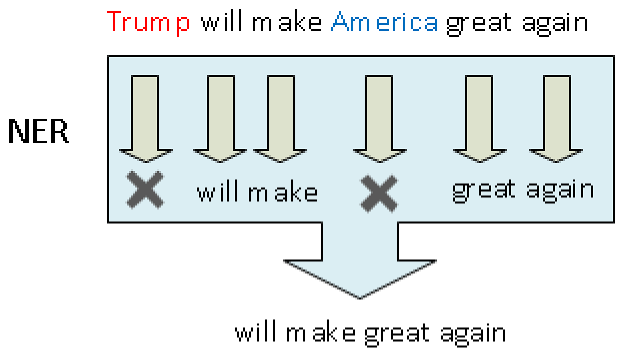


Figure 4: NER process

Table 2: Grammatical frequencies

| Author | Category | Nouns | Adjectives | Verbs | Adverbs |
|---|---|---|---|---|---|
| Barack Obama | Politician | 35% | 8% | 15% | 4% |
| Bill Gates | Non Politician | 32% | 8% | 16% | 4% |
| Donald Trump | Non Politician | 44% | 7% | 11% | 3% |
| Hillary Clinton | Politician | 26% | 6% | 18% | 6% |
| Jeremy Corbyn | Politician | 33% | 7% | 16% | 3% |
| Leonardo Di Caprio | Non Politician | 39% | 8% | 14% | 3% |
| Magic Johnson | Non Politician | 41% | 6% | 13% | 3% |
| Thereza May | Politician | 27% | 8% | 17% | 4% |

When comparing the posts' polarities according to their author's category (politicians and non-politicians), the data did not reveal relevant differences between politicians and non-politicians, as shown in Figure 5. The same analysis was confirmed using the chi-squared test, where a value of $\chi^2 = 1$ was obtained, indicating that both polarities data (politicians and non-politicians) are not independent.

However, this interpretation may lead to an incorrect understanding of the scenario. According to Figure 6, when comparing the authors' polarities, it is possible to conclude that while politicians tend to have posted in the same area in a normal distribution, non-politicians tend to be in the extremes - as shown in Figure 6 where the non-politicians Bill Gates and Magic Johnson are represented by the extremities -, i.e., they are blunter than politicians when expressing through Facebook and indicating that each author has its own "emotional signature" in his posts.

This information is confirmed in Table 4, which presents the positive and negative polarities by authors.

## 5.4 Lexicon-based emotion analysis

To analyse the emotions contained in the text, we employed a lexicon-based approach, which consists of comparing the labelled emotion contained in the EmoLex lexicon with the preprocessed texts described earlier. Using the emotions model proposed by [25], where all sentiment is composed of a set of 8 basic emotions (*anger*, *anticipation*, *disgust*, *fear*, *joy*, *sadness*, *surprise* and *trust*), all posts were analysed according to this model, and a list of emotions in each post was generated, according to Table 5.

Hence, when applying the Person's correlation coefficient ($r^2$) between polarities and basic emotions, as presented in Table 6, it is possible to note which emotions are related with polarities.

In a scale ranging from -1 to 1, emotions related to a high r2 value indicate a strong relationship with polarity (as anger and negative polarity), while high negative $r^2$ values indicate a strong inverse relationship (as fear and positive polarity). In our approach, ambiguous emotions are classified when the standard deviation for r2 polarity's value is less than 10% (i.e., 0.2).

Many authors consider that positive and negative emotions are essential to describing the author's emotional pattern, while neutral emotions do not have a significant contribution to achieving this objective, which is relevant because once the emotions contained in the text are identified, it will enable the identification of specific emotions and how they contribute to positive and negative polarity. Moreover, it helps to justify why a sentence or author is more negative or positive than others.

The emotions classified in the text according to polarities are as follows:

- Positive polarity - Joy;

Table 3: Correlations between authors

| | Barack Obama | Bill Gates | Donald Trump | Hillary Clinton | Jeremy Corbyn | Leonardo Di Caprio | Magic Johnson | Thereza May |
|---|---|---|---|---|---|---|---|---|
| **Barack Obama** | 1.000 | 0.998 | 0.985 | 0.955 | 0.998 | 0.997 | 0.994 | 0.979 |
| **Bill Gates** | 0.998 | 1.000 | 0.974 | 0.971 | 1.000 | 0.991 | 0.986 | 0.988 |
| **Donald Trump** | 0.985 | 0.974 | 1.000 | 0.897 | 0.972 | 0.994 | 0.998 | 0.931 |
| **Hillary Clinton** | 0.955 | 0.971 | 0.897 | 1.000 | 0.969 | 0.931 | 0.924 | 0.980 |
| **Jeremy Corbyn** | 0.998 | 1.000 | 0.972 | 0.969 | 1.000 | 0.991 | 0.985 | 0.990 |
| **Leonardo Di Caprio** | 0.997 | 0.991 | 0.994 | 0.931 | 0.991 | 1.000 | 0.998 | 0.964 |
| **Magic Johnson** | 0.994 | 0.986 | 0.998 | 0.924 | 0.985 | 0.998 | 1.000 | 0.951 |
| **Thereza May** | 0.979 | 0.988 | 0.931 | 0.980 | 0.990 | 0.964 | 0.951 | 1.000 |

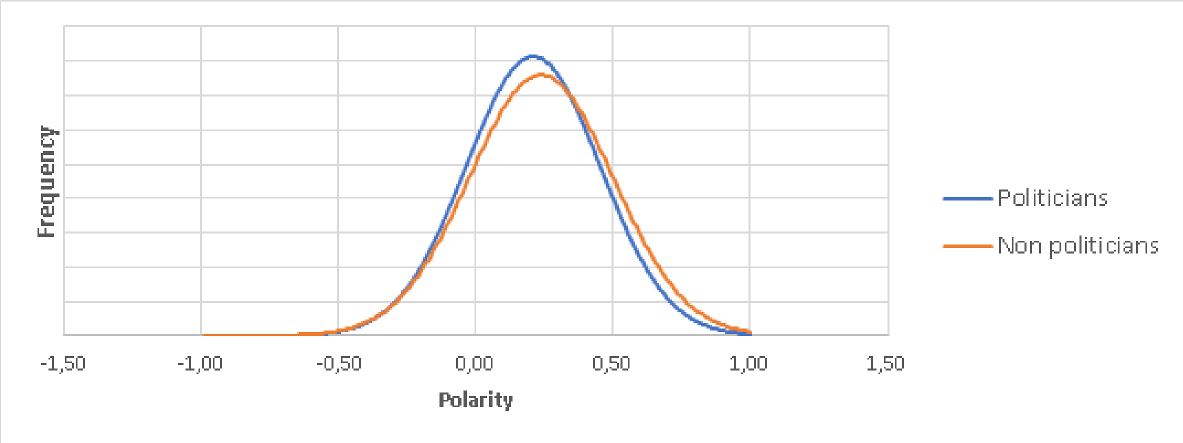


Figure 5: Polarities distribution by category

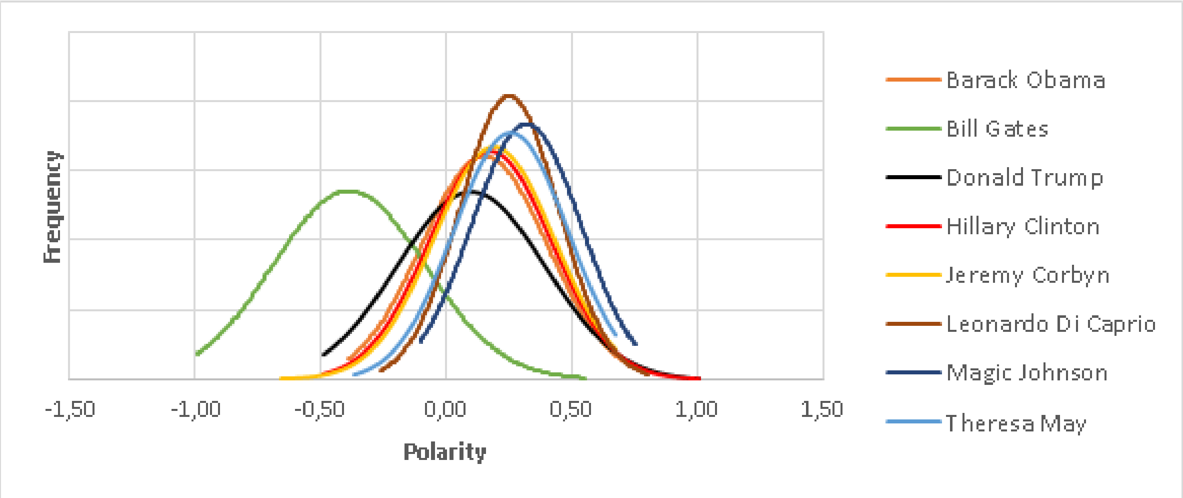


Figure 6: Polarities distribution by author

Table 4: Polarities by author

| Author | Positive | Negative | Category |
|---|---|---|---|
| Barack Obama | 0.28 | 0.13 | Politician |
| Bill Gates | 0.30 | 0.11 | Non Politician |
| Donald Trump | 0.25 | 0.16 | Non Politician |
| Hillary Clinton | 0.35 | 0.17 | Politician |
| Jeremy Corbyn | 0.30 | 0,13 | Politician |
| Leonardo Di Caprio | 0.34 | 0.09 | Non Politician |
| Magic Johnson | 0.37 | 0.06 | Non Politician |
| Theresa May | 0.36 | 0.10 | Politician |

Table 5: Basic emotions average per author

| Author | Anger | Anticipation | Disgust | Fear | Joy | Sadness | Surprise | Trust |
|---|---|---|---|---|---|---|---|---|
| Barack Obama | 0.08 | 0.15 | 0.03 | 0.10 | 0.13 | 0.05 | 0.05 | 0.21 |
| Bill Gates | 0.06 | 0.14 | 0.04 | 0.08 | 0.15 | 0.06 | 0.06 | 0.14 |
| Donald Trump | 0.06 | 0.12 | 0.02 | 0.09 | 0.12 | 0.10 | 0.04 | 0.16 |
| Hillary Clinton | 0.14 | 0.26 | 0.02 | 0.07 | 0.22 | 0.15 | 0.12 | 0.30 |
| Jeremy Corbyn | 0.08 | 0.16 | 0.03 | 0.08 | 0.09 | 0.08 | 0.06 | 0.23 |
| Leonardo Di Caprio | 0.04 | 0.11 | 0.01 | 0.07 | 0.09 | 0.03 | 0.03 | 0.16 |
| Magic Johnson | 0.03 | 0.19 | 0.03 | 0.05 | 0.21 | 0.04 | 0.07 | 0.21 |
| Theresa May | 0.06 | 0.17 | 0.02 | 0.06 | 0.14 | 0.07 | 0.07 | 0.22 |

Table 6: Correlation between polarities and emotions

| Polarity | Anger | Anticipation | Disgust | Fear | Joy | Sadness | Surprise | Trust |
|---|---|---|---|---|---|---|---|---|
| Positive | -0.10 | 0.49 | -0,26 | -0.90 | 0.48 | -0.22 | 0.44 | 0.40 |
| Negative | 0.83 | 0.27 | -0,08 | 0.60 | 0.01 | 0.89 | 0.34 | 0.37 |

- Negative polarity - Anger, Fear, Sadness;
- Ambiguous polarity - Anticipation, Disgust, Surprise, Trust.

When transposing these emotions to polarities, it is possible to determine the polarity profile for each author, as presented in Table 7.

### 5.5 Machine learning-based emotion analysis

Once the average of each emotion was identified from the author, the next analysis was to identify the emotional pattern of the author. To identify this emotional pattern, an approach based on machine learning (ML) techniques was applied. The first attempt tested the same messages in its original state, i.e., with no preprocessing and only the authors' identification in an ML approach. Once the result was obtained only by original texts with no preprocessing, it was considered the lowest acceptable precision rate, and in cases in which this rate decreases, it may be interpreted as a negative influence of preprocessed texts in the authors' prediction. In our initial tests, the best precision rate was presented by an SVM implementation using String2WordVector as word embeddings through Weka [11] and 10-fold cross-validation in the whole dataset, with a correct prediction precision of 85% when predicting authors.

When the lowest prediction rate was identified, the next step was to classify using the preprocessed information. By using the previous preprocessed texts, polarity values, and each basic emotion rate, a new dataset was generated for the ML process. The most relevant algorithms for text classification, such as SVM, naive Bayes, and random forests, were used; however, using a naive Bayes multinomial implementation through Weka and 10-fold cross-validation in the whole dataset, returned a precision of 88% of correct predictions when predicting authors. Both results (non-preprocessed and preprocessed) are presented in Table 8.

## 6 Conclusion

This paper presents a combination of lexicon-based and machine learning approaches to explore the emotions contained in a text through the best practices in sentiment analysis to increase the results' accuracy in authorship identification. Everyone has particular characteristics of expressing his or her thoughts and feelings about the surrounding events and behaviours, and these personal characteristics are naturally reflected or transmitted to his or her texts. Today, internet users are bombarded with intrusive digital content such as advertisements, quotes and news - and many of them are fake news - so ensuring the origin of the information assures the confidence that they are consuming information about whom we want, not whom the author claims to be. For this reason, knowing the author's emotional writing style profile is important, and using this emotional information contained in a given text helps to increase the accuracy of authorship identification. We base this claim on the successful prediction rate increasing from 82% to 87.41% in our tests in addition to the values of precision, recall and f-measure, which increased in the majority of the cases when using emotionally labelled data. This improvement can be interpreted as a promising outcome of our proposal.

Additionally, knowing the emotional profile for different groups enables the identification of the "emotional profile pattern", which can lead to the identifica-

Table 7: Polarities profiles

| Author | Positive | Negative | Ambiguous |
|---|---|---|---|
| Barack Obama | 17% | 24% | 59% |
| Bill Gates | 22% | 21% | 57% |
| Donald Trump | 20% | 25% | 56% |
| Hillary Clinton | 19% | 19% | 62% |
| Jeremy Corbyn | 12% | 22% | 66% |
| Leonardo Caprio | 18% | 22% | 61% |
| Magic Johnson | 27% | 10% | 63% |
| Theresa May | 19% | 16% | 65% |

Table 8: Detailed accuracy results for non-preprocessed and preprocessed texts

| | Non-preprocessed texts | | | Preprocessed texts | | |
|---|---|---|---|---|---|---|
| **Author** | **Precision** | **Recall** | **F-Measure** | **Precision** | **Recall** | **F-Measure** |
| Barack Obama | 0,906 | 0,877 | 0,84 | 0,847 | 0,968 | 0,903 |
| Bill Gates | 0,413 | 0,553 | 0,58 | 0,872 | 0,54 | 0,667 |
| Donald Trump | 0,587 | 0,705 | 0,713 | 0,62 | 0,778 | 0,69 |
| Bill Clinton | 0,82 | 0,876 | 0,867 | 0,899 | 0,852 | 0,875 |
| Jeremy Corbyn | 0,837 | 0,811 | 0,77 | 0,861 | 0,87 | 0,865 |
| Leonardo Di Caprio | 0,876 | 0,889 | 0,862 | 0,949 | 0,905 | 0,927 |
| Magic Johnson | 0,931 | 0,868 | 0,835 | 0,914 | 0,829 | 0,869 |
| Thereza May | 0,558 | 0,646 | 0,641 | 0,74 | 0,779 | 0,759 |
| **Overall** | **85%** | **85%** | **84%** | **88%** | **87%** | **87%** |

tion of different information, such as the domain where the conversation occurs [18].

As future work, determining the author's emotional intensity profile is planned. This is an improvement that will enable different emotional sentences (for example, "this is a silly game" and "this is a stupid game", to have the same emotions, but the word "stupid" is more intense to describe the emotion anger than "silly") in the same emotional proportion by combining with other text analysis metrics to increase authorship identification.

## Acknowledgements

FCT has supported this work – Fundação para a Ciência e Tecnologia within the Project Scope: UID/CEC/00319/2019.

# REFERENCES

[1] Hunt Allcott and Matthew Gentzkow. Social media and fake news in the 2016 election. *Journal of Economic Perspectives*, 31(2):211–36, 2017.

[2] Gordon W Allport and Henry S Odbert. Trait-names: A psycho-lexical study. *Psychological monographs*, 47(1):i, 1936.

[3] James R Averill. *A semantic atlas of emotional concepts*. American Psycholog. Ass., Journal Suppl. Abstract Service, 1975.

[4] Matteo Baldoni, Cristina Baroglio, Viviana Patti, and Paolo Rena. From tags to emotions: Ontology-driven sentiment analysis in the social semantic web. *Intelligenza Artificiale*, 6(1):41–54, 2012.

[5] Margaret M Bradley and Peter J Lang. Affective norms for english words (anew): Instruction manual and affective ratings. Technical report, Technical report C-1, the center for research in psychophysiology, University of Florida, 1999.

[6] Erik Cambria, Amir Hussain, Catherine Havasi, and Chris Eckl. Senticspace: Visualizing opinions and sentiments in a multi-dimensional vector space. In *KES (4)*, pages 385–393. Springer, 2010.

[7] Gerald L Clore, Andrew Ortony, and Mark A Foss. The psychological foundations of the affective lexicon. *Journal of personality and social psychology*, 53(4):751, 1987.

[8] Paul Ekman. An argument for basic emotions. *Cognition & emotion*, 6(3-4):169–200, 1992.

[9] Andrea Esuli. Automatic generation of lexical resources for opinion mining: models, algorithms and applications. In *ACM SIGIR Forum*, volume 42, pages 105–106. ACM, 2008.

[10] Louise Guthrie, James Pustejovsky, Yorick Wilks, and Brian M Slator. The role of lexicons in natural language processing. *Communications of the ACM*, 39(1):63–72, 1996.

[11] Mark Hall, Eibe Frank, Geoffrey Holmes, Bernhard Pfahringer, Peter Reutemann, and Ian H Witten. The weka data mining software: an update.

*ACM SIGKDD explorations newsletter*, 11(1):10–18, 2009.

[12] Howard Leventhal and Klaus Scherer. The relationship of emotion to cognition: A functional approach to a semantic controversy. *Cognition and emotion*, 1(1):3–28, 1987.

[13] Julie Beth Lovins. Development of a stemming algorithm. *Mech. Translat. & Comp. Linguistics*, 11(1-2):22–31, 1968.

[14] Romans Lukashenko, Vita Graudina, and Janis Grundspenkis. Computer-based plagiarism detection methods and tools: an overview. In *Proceedings of the 2007 international conference on Computer systems and technologies*, page 40. ACM, 2007.

[15] Christopher D. Manning, Mihai Surdeanu, John Bauer, Jenny Finkel, Steven J. Bethard, and David McClosky. The stanford corenlp natural language processing toolkit. In *Association for Computational Linguistics (ACL) System Demonstrations*, pages 55–60, 2014.

[16] Mitchell P Marcus, Mary Ann Marcinkiewicz, and Beatrice Santorini. Building a large annotated corpus of english: The penn treebank. *Computational linguistics*, 19(2):313–330, 1993.

[17] Ricardo Martins, José Almeida, Pedro Henriques, and Paulo Novais. Increasing authorship identification through emotional analysis. In Álvaro Rocha, Hojjat Adeli, Luís Paulo Reis, and Sandra Costanzo, editors, *Trends and Advances in Information Systems and Technologies*, pages 763–772, Cham, 2018. Springer International Publishing.

[18] Ricardo Martins, José João Almeida, Pedro Henriques, and Paulo Novais. Domain identification through sentiment analysis. In Fernando De La Prieta, Sigeru Omatu, and Antonio Fernández-Caballero, editors, *Distributed Computing and Artificial Intelligence, 15th International Conference*, pages 276–283, Cham, 2019. Springer International Publishing.

[19] Hermann A Maurer, Frank Kappe, and Bilal Zaka. Plagiarism-a survey. *J. UCS*, 12(8):1050–1084, 2006.

[20] George A Miller. Wordnet: a lexical database for english. *Communications of the ACM*, 38(11):39–41, 1995.

[21] Saif Mohammad and Peter D. Turney. Crowdsourcing a word-emotion association lexicon. *Computational Intelligence*, 29(3):436–465, 2013.

[22] Andrew Ortony, Gerald L Clore, and Mark A Foss. The referential structure of the affective lexicon. *Cognitive science*, 11(3):341–364, 1987.

[23] Martha Palmer, Daniel Gildea, and Paul Kingsbury. The proposition bank: An annotated corpus of semantic roles. *Computational linguistics*, 31(1):71–106, 2005.

[24] Robert Plutchik. *Emotion: A psychoevolutionary synthesis*. Harpercollins College Division, 1980.

[25] Robert Plutchik. Emotions: A general psychoevolutionary theory. *Approaches to emotion*, 1984:197–219, 1984.

[26] Martin Potthast, Benno Stein, Alberto Barrón-Cedeño, and Paolo Rosso. An evaluation framework for plagiarism detection. In *Proceedings of the 23rd international conference on computational linguistics: Posters*, pages 997–1005. Association for Computational Linguistics, 2010.

[27] Klaus R Scherer, Tim Dalgleish, and Mick Power. Handbook of cognition and emotion. *Handbook of cognition and emotion*, 1999.

[28] H Andrew Schwartz, Maarten Sap, Margaret L Kern, Johannes C Eichstaedt, Adam Kapelner, Megha Agrawal, Eduardo Blanco, Lukasz Dziurzynski, Gregory Park, David Stillwell, et al. Predicting individual well-being through the language of social media. In *Biocomputing 2016: Proceedings of the Pacific Symposium*, pages 516–527, 2016.

[29] Carlo Strapparava, Alessandro Valitutti, et al. Wordnet affect: an affective extension of wordnet. In *LREC*, volume 4, pages 1083–1086, 2004.

[30] Zhan Su, Byung-Ryul Ahn, Ki-Yol Eom, Min-Koo Kang, Jin-Pyung Kim, and Moon-Kyun Kim. Plagiarism detection using the levenshtein distance and smith-waterman algorithm. In *Innovative Computing Information and Control, 2008. ICICIC'08. 3rd International Conference on*, pages 569–569. IEEE, 2008.

[31] Rong Zheng, Jiexun Li, Hsinchun Chen, and Zan Huang. A framework for authorship identification of online messages: Writing-style features and classification techniques. *Journal of the Association for Information Science and Technology*, 57(3):378–393, 2006.