Towards the Identification of Fake News in Portuguese

João Rodrigues
Iscte – Instituto Universitário de Lisboa, Portugal
http://www.iscte-iul.pt
jfcrs@iscte-iul.pt

Ricardo Ribeiro
Iscte – Instituto Universitário de Lisboa, Portugal
INESC-ID, Lisboa, Portugal
http://www.inesc-id.pt
ricardo.ribeiro@iscte-iul.pt

Fernando Batista
Iscte – Instituto Universitário de Lisboa, Portugal
INESC-ID, Lisboa, Portugal
http://www.inesc-id.pt
fernando.batista@iscte-iul.pt

Abstract

All over the world, many initiatives have been taken to fight fake news. Governments (e.g., France, Germany, United Kingdom and Spain), on their own way, started to take action regarding legal accountability for those who manufacture or propagate fake news. Different media outlets have also taken a multitude of initiatives to deal with this phenomenon, such as the increase of discipline, accuracy and transparency of publications made internally. Some structural changes have lately been made in said companies and entities in order to better evaluate news in general. As such, many teams were built entirely to fight fake news – the so-called "fact-checkers". These have been adopting different techniques in order to do so: from the typical use of journalists to find out the true behind a controversial statement, to data-scientists that apply forefront techniques such as text mining and machine learning to support the journalist's decisions. Many of these entities, which aim to maintain or improve their reputation, started to focus on high standards for quality and reliable information, which led to the creation of official and dedicated departments for fact-checking. In this revision paper, not only will we highlight relevant contributions and efforts across the fake news identification and classification status quo, but we will also contextualize the Portuguese language state of affairs in the current state-of-the-art.

2012 ACM Subject Classification Computing methodologies → Natural language processing

Keywords and phrases Fake News, Portuguese Language, Fact-checking

Digital Object Identifier 10.4230/OASIcs.SLATE.2020.7

Funding This work was supported by national funds through FCT, Fundação para a Ciência e a Tecnologia, under project UIDB/50021/2020.

1 Introduction

The way in which each of us, regular consumers of contents in the environment that surround us, bridges ignorance and knowledge has been changing over time. This bridge, the channel responsible for making available and disseminating "common interest" content, has been suffering changes in its form, content and perception of reliability, from the consumer's perspective. Contrary to the period prior to the beginning of the Internet, these interventions, that moved according to the political, economic, social and scientific context of each society, are now at the mercy of a new context that has been gaining strength in recent years –

9th Symposium on Languages, Applications and Technologies (SLATE 2020).
Editors: Alberto Simões, Pedro Rangel Henriques, and Ricardo Queirós; Article No. 7; pp. 7:1–7:14
OpenAccess Series in Informatics
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

technology. Since the beginning of our existence, until the early 2003, humanity has generated 5 Exabytes of data [1]. Today, that same volume is produced in only two days. In parallel with this fact, the access points to every kind of information also grew, both for information and misinformation. Today, as we live in a world where the surging of the aforementioned data is dramatical, where the struggle for audiences on traditional media increases and new forms of information are now found in uncontrollable proportions, and where the thoroughness in the management and proliferation of information is declining, it is urgent to provide ourselves a critical and attentive eye to fight the avalanche of disinformation to which we are exposed every day. In the last decade, more traditional information channels such as newspapers and television have been forced to give space, and consequently power, to a new giant phenomenon that has been conquering the market – the social media. This migration of content consumption is essentially due to the popularity that certain social platforms, such as Facebook and Twitter, have started to gain in society [6]. With the emergence of social media, both positive and negative aspects have shown up in terms of impact for its users. On one hand, social networks have brought to life a tool that, due to its regular use in conjunction with its massive popularity, allowed not only an easy way to search for others, but also a huge ease in the almost instantaneous proliferation of news. Hence, news with transverse interest to the entire population, such as the reporting of events in times of crisis, can be obtained almost in real time, either through official news channels or by any user who uses social platforms. Despite this positive side, social media have also seemingly harmed our society in a variety of ways and fields of interest. In the traditional media field, the way that these large entities reached their listeners had to be rethought and reformulated, since there was a major shift in the interest in consumption of news by their target audience towards social media. In fact, nowadays it is quite easy for a user without any contractual affiliation to an audiovisual entity to achieve more views, in a specific content shared only by himself, than some contents presented on FOX News, CNN, and New York Times [3]. Media outlets, in order to avoid completely losing the race for the attention of their target audience, were forced to emphasize and focus on the number of views / clicks of their publications at the prejudice of their content [6]. Ethics, integrity and accountability have been transformed into sensationalism, views number and, ultimately, greed. With this, the perfect environment for the appearance of fake news is settled. News with a willful lack of attention to source confirmation, fake news, misleading news, rumors and especially click bait news, which the only goal is to attract the user to a content that seems to be relevant and interesting, but, after a quick glance at said content, it ends up being far below the user's expectations [2].

However, despite the fact that fake news are recently growing in the most traditional media outlets, it was not here that they have taken such proportions for the first time. In 2016, after the elections in the United States of America that resulted in the victory of Donald Trump, it was immediately possible to realize, in a clear way, that great consequences come from the proliferation of fake news on a large scale. Many were those who addressed this topic and concluded that most of the most debated fake news preceding the election favored Donald Trump over Hillary Clinton. Furthermore, it is unanimous that Donald Trump would never win the elections without the influence of these fake news [3].

The exploration of tasks such as the detection and classification of "Fake News" is quite recent. Due to the enormous media exposure in which the subject has been involved, essentially after the North American elections in 2016 where, allegedly, Russian influence jeopardized the outcome of the results [8], a boom of contributions and initiatives began to take form. Together, the global community started to develop a common sense of urgency to address the problem and started, as a whole, searching for different solutions and approaches [6].

Detection and classification tasks deal with unstructured textual information using NLP (Natural Language Processing) techniques. Different techniques and methodologies provide quite different results, and a large number of these tools can be applied within the scope of this paper.

2 Related Work

In this section, we will begin by conducting an analysis of the different definitions of "Fake News". Subsequently, a survey and its respective analysis will be presented, regarding the most used and recommended datasets by the academic and business community in an attempt to find a resolution to this type of problems and concerns. For each one of them, a summary is made with a brief introduction to its structure and composition, a mention to the authors who contributed most to its exploration and the most used features in each of the datasets. Finally, the last point of this section deals with the different methods/classifiers used in this type of cases, associating, to each one, the datasets in which they were used and the respective authors.

2.1 Fake News

The idea behind the "Fake News" concept is not a novelty. In fact, if we go back a few years, to the time when the Internet did not even exist, and despite the terms not being the exact same, the idea of misinformation and disinformation was already circulating in society and in the traditional media of these times. There are many historical examples that support the aforementioned statement. As early as 1835, for instance, the first major hoax manufactured by the New York Sun newspaper appeared, where countless articles reporting the discovery of life on the moon were published [3]. Already in 2006, another major fabrication of false news appeared in Belgium, where a Belgian television station reported that the Flemish parliament had declared independence from Belgium [3]. At this time, although the market was not as fragmented as it is today, and the fact there were very few media outlets, the power that these entities had at their disposal was already unanimously recognized. Information is power, and the greater the power, the greater the appetite for promiscuous interests and consequent corruption [7].

As for the term Fake News itself, its definition began to gain popularity in the period during and after the North American elections of 2016, which culminated in the election of Donald Trump as president of the United States [3]. Also, alongside this term, two other are now commonly used to describe this phenomenon: misinformation and disinformation. However, these words are often incorrectly used for the same purpose, while they mean two different things. Both are used to refer to any spread piece of content that is false or misleading, however, there is a key difference - the intent behind each content. "Disinformation", in contrast to "Misinformation", is a term used for describing any false or misleading content that is intentionally spread while knowing about the content's lack of truth.

Many papers have emerged after Donald Trump's election, and plenty suggested different definitions for the term "Fake News". The most consensual definition used by most scholars emphasizes the importance of intention and verification: fake news are any and all new news that are proven intentionally false [3]. Therefore, any news that, through different sources, can be disproved, proving to be categorically false, or any news through which the author has the clear intention of misleading, are considered as fake news.

Other authors look at the concept of fake news from different perspectives. In [7], a definition is given regarding three different aspects: publications based on manufactured content; publications inserted in the context of large fraudulent and defamatory campaigns;

and humorous publications. The work done by [6] also mentions three different aspects - humorous content, the need for verifiability, and a new perspective presented as "malicious content". According to these authors, it is necessary to consider the humorist content as misleading, since although it is directed to a public that recognizes the author's humorist intention at the outset, there is also a large portion of the community that fails to correctly identify it. Along with this definition comes the concept of "malicious content", which represents the definition of intention, as mentioned above (disinformation).

The European Commission, a political independent entity with numerous goals, such as the creation of legislation, policies and action programs that cross the interests of the whole European Union, suggested, in 2018, the creation of a group of highly specialized experts in the subject of fake news. This group had a mission - to ensure that the democratic process of the 2019 European elections ran without any kind of interference from both misinformation and disinformation contents. At the time, this group defined the operational concept of Fake News as "all information which is proven to be false or misleading and which is created, presented and disseminated in order to obtain economic advantage or to deliberately mislead the public, and which is likely to cause public harm" [8]. This was the first concept to address the importance of public harm and economic advantages. With these, public harm arises and it is likely the most blatant and alarming consequence of the fake news phenomenon. The authors define it as all the threats to democratic political processes and public goods such as health, environment and security. On the other hand, we also have one of the causes of the immense growth of Fake News contents and consequent public harm - economic interests. Economic lobbies, entities who usually place their interests above the common human being, are one of the most important reasons for the large investment in fake news. This, combined with political interests as well, causes a great deal of pressure and influence on traditional media, coming from multinational companies, the state itself and magnanimous entities.

Finally, in the National context, the definition of fake news by the ERC, the Regulatory Authority for the Media in Portugal, appears. In 2019, the ERC conducted a study entitled "Disinformation – European and National Context", which deals with this subject in great detail [8]. Fake News are defined by this entity in a very similar way to the one given by the European Commission, but it adds a new perspective which should be taken into account in this context. ERC mentions that a news story, in its definition, can never be false, but that the contents of the narratives that are inserted in it, can be false or misleading. This means, according to the author, that labelling any news as a "Fake New" might be a little abusive, semantically speaking, and misleading.

2.2 Datasets

This section is dedicated to the detailed analysis of the most used datasets in the literature in the context of fake news. For each one an overview is made, specifying when, by whom and for what purpose it was created. An analysis of each one's content is also made, such as the number of records, the different labels and the variables that define it. Finally, a comparative analysis of the different authors is presented, defining who made use of each of the datasets, the approach taken by each one regarding the basic textual processing performed, the NLP tasks performed, and the textual representations implemented, which will serve as input features to the machine learning algorithms presented in the next section.

2.2.1 Fake News Challenge (FNC-1)

The creation of this dataset took place in a challenge called "Fake News Challenge", in 2017, and counted on the joint effort of 100 volunteers from the academic and business fields [14]. The goal of this challenge was to find new methods and approaches that would be useful to present solutions to fight fake news. This dataset contains a set of news written in English. The dataset contains about 50,000 associations of statements with news. Each statement is associated with a particular news item and also with a label. The dataset is rather unbalanced in the sense that it has four different labels with a scatter sample distribution between them.

Table 1 Records by Label (FNC-1).

Label	**Records**	**Percentage**
Unrelated	36.545	73,1%
Disagree	8.909	17,8%
Agree	3.678	7,4%
Discuss	840	1,7%

This dataset has a different nature than the one expected in a typical Fake News problem. Often this problem is thought and addressed for the classification of titles and news bodies from a dichotomous perspective (usually called “truth labeling”), i.e., true or false. The purpose of this challenge is different: it is about trying to understand what is the “posture/relationship” of one statement before another, or a set of others (news/text body). Thus, the objective is to classify an affirmation using one of the previously mentioned labels: “Discuss” means that the body text neither confirms nor denies the statement; “Unrelated” means that there is no relationship between the body text and the statement; “Disagree” means that the body text does not agree with the statement; “Agree” means that the statement and the body text are related and agree with each other. According to the authors in [14], this approach is not a substitute for the truth labeling, but a means of supplementing it. The decision was made based on talks with journalists and fact-checkers where both parties mentioned that it is quite difficult to make a truth labeling classification. Both mentioned that they would rather have a semi-automatic solution to assist them in their work than a fully automatic solution whose performance would be far below expectations.

Many studies were based on this challenge and the respective dataset, and today this is one of the most used and explored datasets by all fake news researchers. Thus, each author has tried to approach it in his own way, using different tasks of text processing approaches and different types of representations of the text. These different representations form the features that will serve as input to the later machine learning and deep learning tasks.

Table 2 Pre-processing tasks and Textual Representations by Author (FNC-1).

Author	Pre-processing		Textual Representation
	Basic processing	NLP	
[9]	Regular expressions	Lemmatisation	Glove Embeddings
		Standford NER	
[5]	Label mapping	Sentiment Analysis	Word-Embeddings
			Google News CNN
	Tokenization		Number of n-grams
			TF-IDF
	Stemming		word2vec
			SVD
	N-grams generation		Number of positive and negative words

In [9], the authors started by using regular expressions to eliminate all unwanted links. Following that, they made use of lemmatization - the process of grouping together the inflected forms of a word as a single item. The Stanford Entity Name Recognizer was also used to replace entities such as names, organizations, and locations. Finally, pre-trained GloVe Embeddings models were used to represent all words in global semantic vectors [12]. In [5], the authors describe the work that led them to win first place in the competition. The "Solat In the Swan" team chose to use the combination of two classification approaches, and to do so they needed to create two sets of features. In order to get to both sets it was necessary, first of all, to perform a pre-processing step. In this phase, the tokenization of both titles and news were made and stemming was applied to each of the tokens. Finally, the various unigrams, bigrams and trigrams were generated from the list of tokens. Once the pre-processing was carried out, the first textual representation was created. For this, pre-trained Google News vectors were used, both on the titles and on the news themselves. More traditional features were applied in the second set: n-grams count; TF-IDF; SVD-based; word2vec and sentiment features.

2.2.2 BuzzFace

This dataset was developed from a set of news items published by the media company "BuzzFeed" after the 2016 North American elections and was manually tagged by journalists [16]. This dataset was also enriched with information from Facebook, such as comments, number of shares and reactions to the news published by the company on its website.

The dataset has a total of 2,282 entries. Each entry line corresponds to a share on Facebook from an official media outlet's page. Each post can have one of four possible classifications: "no factual content", "mostly true", "mostly false" and "mixture of true and false".

Table 3 Records by Label (BuzzFace).

Label	**Records**	**Percentage**
True	1.665	73%
Non Factual Content	274	12%
Mixture of True and False	251	11%
Mostly False	91	4%

This dataset is also of great importance since it has features related to the "context" of the news, i.e., number of shares, reactions and comments to the news post. This is the only dataset contemplated in this study that contains these types of features, which also means that it is the only one that does not depend directly on the intrinsic content of the news.

Table 4 Pre-processing and Textual Representations by Author (BuzzFace).

Author	Pre-processment		Textual Representation
	Basic processing	NLP	
[15]	-	-	Bag of words
	-	POS Tagging	Number of pronouns, verbs, adverbs, hashtags, punctuation.
	By “Linguistic Inquiry and Word Count”	By “Linguistic Inquiry and Word Count”	Psycholinguistic features (detection of biased and persuasive language)
	By “Google’s API”	By “Google’s API”	Semantic features (toxicity)
	By “Text Blop’s API”	By “Text Blop’s API”	Features subjectivity (subjectivity and feeling)

In [15], the authors start by ensuring a greater balance of the dataset, making the distribution of the records (for each label) more uniform. Thus, only two labels were considered: true and false, all cases with the label “mostly true” became “true”; the cases with the label “no factual content” were removed; and the records with the two remaining labels were converted to the label “false”. The authors then apply and detail a set of features appropriate to the problem in question, in which context-related features are also included. The features are thus divided into three main groups: (1) features extracted from the news content, (2) features extracted from the source of the news and (3) features extracted from the environment.

2.2.3 WSDM Cup

This dataset is one of the most recent in the literature and was developed with the purpose of being the object of study for those who participated in the challenge of the international conference WSDM (Web Search and Data Minning), WSDM Cup, which took place in 2019, organized by the ACM (Association for Computing Machinery).

The dataset in question was developed by ByteDance, a Chinese internet and technology company, which owns an online news platform. One of the greater challenges faced by ByteDance is the fight against fake news. To this end, the company has created a database to colect all kinds of fake news, so that all news can be properly verified regarding their veracity before being presented in the platform [10].

The dataset was originated from the database mentioned above, counting with a total of 360,767 records. Each record has in its constitution a title of a false news A, a title of a news B (news to be classified) and its label (Agreed, Disagreed or Unrelated). The objective here is to try to understand if the news item B addresses the same subject and agrees with the title A (agreed), which makes the news item B false; to try to understand if the title B does not agree with the news item A (disagreed), making the news item B true; or to identify that the news B has no relation whatsoever with news item A (unrelated). The dataset has news titles in two different languages, Chinese and English, its configuration follows a weight of 75% for training and 25% for testing, and presents the following label distribution:

Table 5 Records by Label (WSDM Cup).

Label	**Records**	**Percentage**
Unrelated	246,764	68.4%
Agreed	104,622	29.0%
Disagreed	9,379	2.6%

Table 6 shows some of the works using this dataset.

Table 6 Pre-processing and Textual Representations by Author (WSDM Cup).

Author	Pre-processing		Textual Representation
	Basic processing	NLP	
[10]	Dataset augmentation	-	Set of 25 pre-trained BERT's.
	Stopwords removal		
[13]	Dataset increase	N-grams	Text Based
	Text to lowercase		Statistics
	Add spaces between punctuations		Graph Based
	Tokenization		KNN (BERT's)

"Travel", the team that came second in the competition, developed an approach that achieved a "weighted accuracy score" of 0.88 [10]. Given the nature of the problem, the authors have chosen to attack the problem using NLI (Natural Language Inference) techniques. NLI is a subarea of NLP and its objective is to recognize textual implications - RTE (Recognizing Textual Entailment) - that is, in this case, from news B which is given as false, to be able to infer a hypothesis (which label characterizes news A) from a textual premise through the semantic similarities between them. Regarding the creation of the features that will feed the machine learning algorithms, the authors in question decided on three main steps. First, since the dataset was not properly balanced regarding the distribution of records per label, and in order to avoid an over-fit, an increase of quantity of the data was made through the transitivity of semantics. That is, if title A is related to title B and if title A is related to title C, then B and C are related. Subsequently, all stop words were removed, both in Chinese and English news. Finally, to address the problem of text representation, the authors chose to use BERT, a pre-trained linguistic model created by Google that has recently been gaining popularity.

In [13], the challenge's winning team ("IM"), suggests a pre-processing approach and textual representation similar to that used by the "Travel" team, but with some nuances. As far as pre-processing is concerned, the author suggests also separating text scores (by placing spaces) and the tokenisation of all titles. In the textual representation, an ensemble of features is suggested for input regarding future classification algorithms. The first set of features are "Text Based". Here all textual features are covered, such as the generation of n-grams of words and characters. After their generation, distance measurements are applied to pairs of titles, such as cosine, euclidean, city-block, jaccard, or simple addition and subtraction. The second set of features are "Statistics". This is where word counts are present, as well as stop words, tokens, characters, or a simple comparison of the textual size of the title pairs. The third set of features are the "Graph Based". In this case, the texts of the titles are represented as graph networks. The objective of these networks is to make a representation of each title (a node of the graph) and of each pair of titles. Through metrics, such as minimum or maximum distance between nodes and news pairs, assumptions can be made about their relation. Finally, the features "KNN" represent BERT embeddings with reduced dimensionality.

2.2.4 Fake.Br Corpus

For the Portuguese language, Fake.Br Corpus is the only dataset we found in this context. According to [11], this is the first fake news dataset with Brazilian Portuguese news and also the first dataset in the Portuguese language. The dataset resulted from a joint effort of researchers and analysts who gathered and classified news manually. The dataset contains 7200 news items and is divided in a balanced way regarding the number of records that are associated with the different labels (Table 7).

Table 7 Records by Label (Fake.Br Corpus).

Label	Records	Percentage
True	3,600	50%
False	3,600	50%

The authors defined a time span of two years (January 2016 to January 2018) and only collected news that were inserted in it. Some news, however, reference other news in previous time spaces. Other relevant information, such as the author of the news, the date of publication, the number of views and comments, were kept. The news were manually tagged and, for each false one, the authors used a semi-automatic process to find true news that could prove the tag on the corresponding "false" ones. The false news were manually extracted from four different newspapers. After this process, through web scrapping, 40,000 real news were taken from other sites, based on the most frequent words, verbs and names that were in each of the fake news previously taken. Then, a lexicon similarity measure (cosine) was applied to determine which were the true news that were closest to the fake news. After this process was completed, having already a smaller number of news items, they made the manual selection of the news items based on their actual degree of similarity.

Once again, the work on this dataset explored different pre-processing procedures and different types of text representations.

In [11], the authors use various approaches to the representation of the news text. Apart from the most common forms of textual representation (bag of words, term count, etc.), the authors also explore the use of linguistic features that may also be interesting, such as pausality, uncertainty, expressiveness, non-immediacy, and number of semantic classes.

In [4], the author presents two alternative approaches to the work of [11]. First, the author mentions the use of the chi-square method as a means of selecting the most relevant terms that resulted from his pre-processing task. According to his work, this method is an added value for textual processing tasks in the Portuguese language. Finally, the author states that textual representations based on word embeddings have shown better results than classical representations based on term-to-document matrices.

Table 8 Pre-processing and Textual Representations by Author (Fake.Br Corpus).

Author	Pre-processing		Textual Representation
	Basic processing	NLP	
[11]	Stopwords Removal	Stemming	Bag of Words
	Punctuation Removal		
	-	POS Tagging	Number of each Part of Speech takes place
	-	Enriched Lexicon	Number of semantic classes
	Words Removal	-	Pausality (number of punctuation characters)
	-	POS Tagging	Expressiveness (sum of adjetives and adverbs over the sum of nouns and verbs)
	Extraction of Specific Words (can, might)	-	Incertainty (number of modal verbs)
	-	POS Tagging	Non-Immediacy (number of first and second pronouns)
[4]	Stopwords Removal Tokenisaton	Stemming	Word-Embeddings
		Lemmatisation	
		Chi-square	

2.3 Methods

This section is dedicated to the analysis of the most commonly used methods in fake news detection. There are several ways to approach this topic according to the literature, however, most approaches cast this as a classification problem. In a classification problem the aim is to be able to associate a label, for example true or false, with small (in the case of a title, for example) or large (in the case of a news' body text, for example) textual portions. In order to respond to this task, most of the research body dedicated to this subject implements machine learning and especially deep learning techniques [6].

Within the classification problem, authors employ different methods depending on the features they have at their disposal. Typically, most of the approaches focus on using features extracted from the news content itself. However, other sets of features, such as information related to the source or context of the news [15], present another type of detail that may enrich the analysis. Regarding fake news classification strategies, a combination of methods is typically used, the so-called "ensemble models". Since in most works that ensemble methods are used the evaluation metrics refer to the set and not to each method itself, it is relevant to make a comparative analysis not only between methods, but also between standalone methods and ensemble methods.

There are several methods of machine learning that have been applied in the task of fake news detection. Of the most recent and best performing methods, there are three that typically stand out from the more traditional methods (such as KNN, Naive Bayes, Decision Trees, etc). The first method is the SVM. The SVM (Support Vector Machine) is a discriminative classifier formally defined by a hyper plane of separation [6]. This method has been used in several fake news tasks. In two of the four datasets mentioned in the previous section, there are authors who propose the use of the SVM in isolation [11, 15].

In [11], in a study using the dataset "Fake.Br Corpus", the authors used the SVM with only content features. Among several combinations of textual representations, the best performance obtained was a F1-score of 0.89. In [15], in a study using the dataset

"BuzzFace", the authors made use of all kinds of features and in all of them applied an SVM. In this work, the performance was a F1-score of 0.76, a performance that was below other methods also applied, such as Random Forest (0.81 F1-score) and Gradient Boosting (0.81 F1-score). An interesting approach used SVM [18]. The authors used a variation of SVM called "Graph-Kernel-Based SVM" to identify rumors using propagation structures and content features. In this study, the authors reported an accuracy of 0.91.

Another method that has been gaining prominence over more traditional methods is the "Gradient Boosting" approach. Gradient Boosting is a meta algorithm based on decision trees, and it is used to reduce biased predictions. Catboost and LightGBM, for instance, are versions of Gradient Boost that have been gaining popularity in recent times due to their advantages of fast processing and high prediction performance [13]. These types of algorithms typically appear in stance detection problems, but can also appear in truth labeling problems [9, 13, 15]. In [9], in a specific stance detection problem, an accuracy of 0.83 was achieved for this method alone. Although attaining a good performance, gradient boosting ended up behind two other classification methods also tested: LSTM and BiLSTM neural networks models. In [15], in a specific truth labeling problem, gradient boosting was also used achieving the best performance (0.81 F1-score) against other algorithms: SVM, Naive Bayes, and Random Forest.

Finally, deep learning models have brought great advances in several areas of Artificial Intelligence, such as image identification, speech recognition, and textual processing [13]. In this topic, the most commonly used neural networks are CNN (Convolutional Neural Networks) and RNN (Recurrent Neural Networks). In [4], a study on the dataset "Fake.Br" using neural networks, more specifically a CNN, achieved a performance of 0.91 accuracy which translated in a very successful result for a truth labeling problem. In [9], the authors using LSTM (Long Short-Term Memory) and BiLSTM (Bidirectional Long Short-Term Memory) neural networks managed to obtain a 0.92 and 0.93 accuracy, respectively, in this work regarding stance detection task.

As previously mentioned, ensemble methods have been a very successful approach to fake news detection. Typically combining deep learning and traditional machine learning techniques, these approaches achieved better results, in most cases. In the two public challenges that are described in this paper, WSDM Cup and Fake News Challenge, the winning teams made use of an ensemble method. In the Fake News Challenge 2016 (FNC-1), the authors who came first [5], after several attempts to apply methods individually, concluded that the best performance was achieved by combining the methods they were exploring. Thus, the best performing approach was a combination of CNNs and Gradient-Boosting Decision Trees methods. This approach had an average weighted score of 82.02%.

In [10, 13], works that finished first and second in the "WSDM Cup 2019" competition, the response to the problem also included a set of methods. The second placed team [10] used a total of 6 methods (3 SVM's, 1 Naive Bayes, 1 KNN and 1 Logistic Regression), obtaining an average accuracy score of 88.15%. The first team [13] chose to combine 28 methods (18 Neural Network Models, 9 Tree Based Models and 1 Logistic Regression), resulting in an average accuracy score of 88.28%.

3 Challenges for European Portuguese

Fake news detection, or fact-checking, has become a prominent facet of political, economic, sports and social news coverage. This task is defined by employing a variety of methodological practices, such as treating a statement containing multiple facts as if it were a single fact and categorizing it as accurate or inaccurate. These practices share the tacit presupposition that there cannot be genuine political debate about facts given that facts are unambiguous and not subject to interpretation. Therefore, when the black-and-white facts, as they appear to the fact checkers, conflict with the claims produced by politicians, these same fact-checkers are able to detect and expose lies [17].

In the past years, fact-checking has been a regular task that any journalist must practice in their work on a daily basis. It is not something they might do, but something they must do, since it is what their journalist deontological code implies. However, as mentioned, with the exponential growth of social media usage, and the urge of the traditional media to apply different methodologies to be able to keep up with the business and compete with those new forces, in many cases that deontological code has been put aside. Furthermore, since there does not seem to be, yet, legal consequences, at least not in Portugal, for the fabrication and propagation of fake news (mostly due to freedom of speech being a sensitive topic as we live in a democracy), they seem to have started appearing everywhere, from social media to most traditional media outlets.

With this high competition between media sectors alongside with this sense of impunity, sensationalism and made up news started to emerge. Today, not only are we living in the era of data – an era marked by the privilege of being able to access a massive amount of information (more than we ever could before) – which is already difficult to process, but we are also living in an era of disinformation. This is why it is imperative to have solutions to allow people to be better and well informed. Having this in mind, many journals around the world seem to have decided to adapt and build teams that are fully dedicated to fact checking duties. Although their focus, as any other private company, is also to make a profit, the strategy that they adopt is different - they focus on the quality and reputation of the group, and that is why, even knowing that this could cost more in the near-term, making sure that the information that they provide is reliable helps them build a sense of trust between the group and the community in the longer term.

In Portugal, as it was pointed out in the ERC study, there are two newspapers that have their own department of fact-checkers: **Polígrafo** and **Observador**. Both newspapers are certified by Poynter, the owner of the International Fact-Checking Network (IFCN), a unit fully dedicated to bringing together fact-checkers worldwide. This unit was launched in September 2015 in order to support a booming crop of fact-checking initiatives by promoting best practices and exchanges in this field. With this, both newspapers mentioned above should stand for Poynter principles.

Polígrafo is a recent digital newspaper launched by Sapo. It was announced for the first time in November 2018, during the Web Summit in Lisbon. Polígrafo is the first Portuguese newspaper dealing with fake news and its database has more than three thousand classified news. During the Covid-19 pandemic crisis, Polígrafo made a partnership with "Corona Verificado", a fact-check platform coordinated by the Brazilian "Agencia Lupa", which integrates information from 34 different fact-checkers entities from 18 Ibero-American countries. This platform alone has more than two thousand news classified.

Observador it is also a recent online newspaper, born in 2014. It defines itself as an independent and free online, daily newspaper that searches for the truth and submits to the facts. Observador stands for not being conditioned by partisan and economic interests or any group logic. They are accountable only to their readers. Over the years, Observador has been gaining a lot of popularity and respect by Portuguese population and even their peers. In 2015, the newspaper decided to create a section dedicated to fact-checking, and they became the first Portuguese newspaper having a department fully dedicated to fact-checking duties. Today, Observador has more than three hundred classified news.

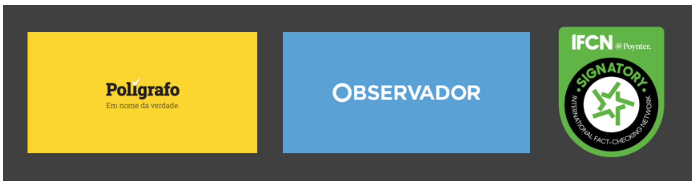

Figure 1 Polígrafo and Observador Recognition by Poynter.

These two sources can clearly constitute a fundamental resource for scientific research on automatic fake news detection in European Portuguese. Especially considering that specific aspects of language are cornerstone in this type of NLP task. The major challenges are constituting such resource and studying the suitability of the presented methods.

4 Conclusion

In this paper, we presented a detailed overview on automatic fake news detection. First, we introduced fake news regarding its context and definition according to the point of view of many different entities, from individual to national (ERC) and international ones (European Commission). Secondly, we presented the most used datasets in the context of fake news. For each dataset, we detailed its objectives, followed by how it was created, by how many records it has, and the distribution of entries by different target label. Also, for each dataset, an analysis was made on different works that used said dataset. The main papers that studied the datasets are here compared, not only by the type of pre-processing that they applied but also by their explored text representations. Thirdly, an overview of the different methods used in the last years was made, presenting their respective evolution and datasets in which they were applied. Finally, we discussed the importance of fact checking, where the need and the importance of those practices in our global society were shown. More specifically, in the case of the European Portuguese language, we presented the two major entities that are certified by Poynter as official fact-checkers, Polígrafo and Observador.

To conclude, most of the techniques necessary to successfully implement an automatic fake news detection system were addressed in this paper. From the most used preprocessing techniques, the most used classification methods (both for Portuguese and English datasets), to the resources available in Portuguese. This is a relevant asset for conducting future work on European Portuguese, where questions regarding the limitations of these resources and their impact on the Portuguese language itself can be, and should be, further analyzed and discussed.

References

1 Alberto Cairo. *The Functional Art: An Introduction to Information Graphics and Visualization*. New Riders, 2012.

2 Monther Aldwairi and Ali Alwahedi. Detecting fake news in social media networks. *Procedia Computer Science*, 141:215–222, 2018. doi:10.1016/j.procs.2018.10.171.

3 Hunt Allcott and Matthew Gentzkow. Social media and fake news in the 2016 election. *Journal of Economic Perspectives*, 31(2):211–236, 2017. doi:10.1257/jep.31.2.211.

4 Renan Rocha De Andrade. Utilização de técnicas de aprendizado de máquina supervisionado para detecção de Fake News. https://riuni.unisul.br/handle/12345/8649, 2019.

5 Sean Baird, Doug Sibley, and Yuxi Pan. Talos Targets Disinformation with Fake News Challenge Victory, 2017. URL: `http://blog.talosintelligence.com/2017/06/talos-fake-news-challenge`.
6 Alessandro Bondielli and Francesco Marcelloni. A survey on fake news and rumour detection techniques. *Information Sciences*, 497:38–55, 2019. `doi:10.1016/j.ins.2019.05.035`.
7 Yimin Chen, Niall J. Conroy, and Victoria L. Rubin. News in an online world: The need for an "automatic crap detector". *Proceedings of the Association for Information Science and Technology*, 52(1):1–4, 2015. `doi:10.1002/pra2.2015.145052010081`.
8 Entidade Reguladora para a Comunicação Social, editor. *A Desinformação - Contexto Europeu e Nacional*. Parlamento português, 2019.
9 Neema Kotonya and Francesca Toni. Gradual Argumentation Evaluation for Stance Aggregation in Automated Fake News Detection. In *Proceedings of the 6th Workshop on Argument Mining*, pages 156–166. ACL, 2019. `doi:10.18653/v1/w19-4518`.
10 Shuaipeng Liu, Shuo Liu, and Lei Ren. Trust or Suspect? An Empirical Ensemble Framework for Fake News Classification. *Proceedings of the 12th ACM International Conference on Web Search and Data Mining, Melbourne, Australia*, pages 1–4, 2019. URL: `http://www.wsdm-conference.org/2019/wsdm-cup-2019.php`.
11 Rafael A. Monteiro, Roney L.S. Santos, Thiago A.S. Pardo, Tiago A. de Almeida, Evandro E.S. Ruiz, and Oto A. Vale. Contributions to the Study of Fake News in Portuguese: New Corpus and Automatic Detection Results. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 11122 LNAI:324–334, 2018. `doi:10.1007/978-3-319-99722-3_33`.
12 Jeffrey Pennington, Richard Socher, and Christopher Manning. GloVe: Global Vectors for Word Representation. *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, page 4, 2014.
13 Lam Pham. Transferring, Transforming, Ensembling: The Novel Formula of Identifying Fake News. *Proceedings of the 12th ACM International Conference on Web Search and Data Mining, Melbourne, Australia*, 2019.
14 Dean Pomerleau and Delip Rao. The fake news challenge: Exploring how artificial intelligence technologies could be leveraged to combat fake news., 2017. URL: `http://www.fakenewschallenge.org/`.
15 Julio Reis, André Correia, Fabrício Murai, Adriano Veloso, and Fabrício Benevenuto. Supervised Learning for Fake News Detection. *IEEE Intelligent Systems*, 34(2):76–81, 2019. `doi:10.1109/MIS.2019.2899143`.
16 Giovanni C Santia and Jake Ryland Williams. BuzzFace : A News Veracity Dataset with Facebook User Commentary and Egos. In *Proceedings of the Twelfth International AAAI Conference on Web and Social Media (ICWSM 2018)*, pages 531–540. AAAI, 2018.
17 Joseph E. Uscinski and Ryden W. Butler. The Epistemology of Fact Checking. *Critical Review*, 25(2):162–180, 2013. `doi:10.1080/08913811.2013.843872`.
18 Ke Wu, Song Yang, and Kenny Q Zhu. False Rumors Detection on Sina Weibo by Propagation Structures. In *IEEE 31st International Conference on Data Engineering*, pages 651–662, 2015.