Development of a Smartphone Application and Chrome Extension to Detect Fake News in English and European Portuguese

Ricardo Afonso, and João Rosas

***Abstract*—In a digital society, the truth portrayed by information is crucial in promoting education, security, and evolution. However, fake news raises a significant concern in that regard. Although there has been a continuous effort in the fight against fake news, it is still a multifaceted challenge in constant change as the menace renovates itself. Thus, in our approach, several machine learning and deep learning models were developed to obtain models that can detect fake content that appears online. The models can then be interfaced with users' devices, namely in the form of browser extensions and smartphone applications. The classification models run on a cloud server and are accessible via web services. These models can detect fake news in English and European Portuguese, with a stronger focus on the latter, given the reduced number of projects in this specific field and language. Besides developing the first public dataset for fake news detection in European Portuguese through web scraping, the models achieved better performance than previous work while being trained with a significantly higher amount of data from a wider variety of sources.**

Link to graphical and video abstracts, and to code: https://latamt.ieeer9.org/index.php/transactions/article/view/8547

***Index Terms*—Machine Learning, Deep Learning, Web Scraping, Natural Language Processing, Extra Gradient Boosting**

I. Introduction

Our world is constantly changing, with millions of events happening every time and everywhere. To not only share information with the rest of the world but also keep up with everything happening, we resort to media in the form of radio, television, press, and the internet, with the latter being one of the most prominent nowadays.

In this context, the integrity of information is essential, as it determines our opinions, attitudes, feelings, and decisions regarding ourselves and others. When we base our decisions on unreliable information, their effects can be harmful, even to our security and health. When lies spread through social media, they can cause social, political, and economic disruptive, damaging effects, although they can provide unfair benefits to the perpetrators [1]. When deception and lies spread through social networks, this phenomenon is commonly known as fake news. In these cases, the consequences are even more severe and can manipulate public opinion, polarize people, cause social tensions, and influence elections. Mitigating this phenomenon requires rational and balanced oversight.

Ricardo Afonso and João Rosas are with NOVA School of Science and Technology, Caparica, Portugal (e-mails: ro.afonso@campus.fct.unl.pt and p187@fct.unl.pt).

However, the ease of access to information and its almost worldwide availability come at the price of complex monitoring, regulation, and credibility checks of not only the information shared but also the responsible entity since anyone can create their website or social media account for a wide variety of purposes, including more nefarious ones such as spreading false information of all kinds.

The amount of fake news has increased in recent years, as has the sophistication of the strategies used to deceive people [2], [3]. It is, therefore, necessary to make a continuous effort to develop new mechanisms to combat fake news and update existing mechanisms. This necessity means applying more research to obtain more sophisticated and complex detection models and processes.

Unlike trusted sources like journalists and media outlets, the internet has little to no regulation over shareable content and conduct principles, so one must filter both content and sources available. However, this is usually hard to do, as many publishers and individuals who spread fake news try to look trustworthy by impersonating well-known sources [4].

Fake news poses no significant threat when used for more fun-related purposes, such as parodies or other forms of humor. The real threat emerges when fake news is used for more malicious purposes, such as manipulating one's principles, ideals, perceptions, and behavior, which has been found within political, social, and economic affairs.

Such is the case of the 2016 presidential election in the United States of America, during which President Donald Trump resorted to false statements aimed at his political opponents, questioning their rights to govern the country [5].

The ongoing COVID-19 pandemic is another example, with many fake videos and photos regarding their origin and impacts being shared over social media, where panic spread faster than the actual virus [6].

Many fake news articles revolving around the ongoing war in Ukraine have also been spread online by both pro-Russian and pro-Ukrainian groups, which end up raising even more unnecessary conflicts and fear across the world [7].

The sheer amount of information alongside the different types of content makes it impossible for humans to review every news article and effectively identify what is fake and what is real. This is where approaches based on Machine Learning (ML) and Deep Learning (DL) must be used.

This article proposes a system based on ML and DL models, alongside cloud and web services, to protect users from fake news in English and European Portuguese, with a focus on the latter given the lack of projects in this field and language.

The methods used to detect fake news are explored in Section II and Section III describes the proposed approaches in English and European Portuguese. The results of each approach are discussed in Section IV and final considerations are mentioned in Section V, alongside future work.

II. Fake News Detection Methods and Techniques

Detecting fake news involves various methods and techniques to identify fake or misleading content inside, e.g., news and social media. Since most fake news articles are generated by bots and then spread by people unaware of their false nature, most share specific characteristics within the data, which can be observed during content analysis. As such, data must be analyzed in such a way as to discover these peculiarities within the questionable content so that AI classifiers can learn from them and effectively manage to spot them afterward. This ability requires using Natural Language Processing (NLP) and Machine Learning to achieve this feature, usually known as Text Mining.

Text mining consists in extracting valuable and meaningful information from unstructured text. It involves several techniques to transform large volumes of unstructured text into structured and analyzable data. Examples of unstructured text are the contents contained in social media, product reviews, videos, and audio files. Other content can be obtained from semi-structured text, which includes Extensible Markup Language (XML), Javascript Object Notation (JSON), and HyperText Markup Language (HTML) files.

Overall, the processes found within fake news detection can be divided into different phases, resorting to a wide variety of datasets, data processing techniques, and classification models. Fig. 1 illustrates a typical sequence of processes behind text classification.

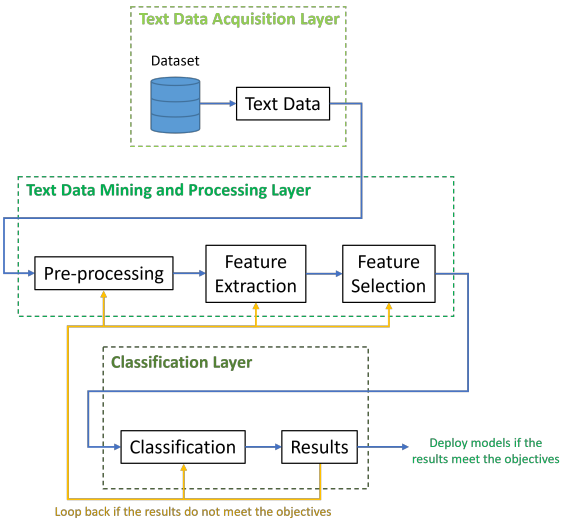


Fig. 1. Text classification processes

A. Data Acquisition

The first step is related to data acquisition, typically involving the use of datasets. Datasets play an important role in any Artificial Intelligence (AI) task, and detecting fake news through ML and DL methods is no exception. After all, the data being processed will define the performance and behaviour of the models, so using appropriate data translates to better future predictions and outcomes.

With these concepts in mind, many organizations and individuals have created several publicly available datasets for this kind of task. For example, Awf Abdulrahman and Muhammet Baykara [8] resorted to a social media fake news dataset from Kaggle, Antoun et al. [9] developed two models based on a dataset provided by QICC made of 384 articles, Khanam et al. [10] resorted to the LIAR-PLUS Master, an extended version of the LIAR dataset from Kaggle, and Thota et al. [11] resorted to the Fake News Challenge (FNC-1) dataset, which is from a specific digital journalism project for rumour debunking with 1684 articles.

It is also common to combine different datasets into a single one, in order to train models with a wider variety of data patterns. For example, Ahmad et al. [12] and Mishra et al. [13] resorted to datasets from Kaggle and the ISOT Fake News Dataset to obtain news articles from many different domains, including politics, sports, technology, and entertainment, while Verma et al. [14] created the WELFake dataset by merging popular datasets from various sources, including Kaggle, McIntire, Reuters, and BuzzFeed Political, among others.

Many authors have also created their own datasets after gathering unstructured data from different sources and transforming them into semi-structured data, which researchers and developers then end up accessing during the development of their projects.

Social media is often considered when extracting fake news. For instance, Sahoo and Gupta [15] resorted to a web crawler and the Facebook Application Programming Interface (API) to create a dataset comprised of 15000 news from a total of 5000 different profiles, Khan et al. [6] resorted to diverse fake and real COVID-19-related news articles manually gathered from various sources, such as Facebook, Twitter, The New York Times and Harvard Health Publishing, and Patwa et al. [16] collected 10700 social media posts and news related to COVID-19 in English.

One can also resort to news websites and blogs. For example, Silva et al. [17] created the Fake.Br Corpus, a dataset made of 7200 news articles in Brazilian Portuguese, half real, half fake, gathered from different online media sources and journals, João Rodrigues [18] created FakePT, a dataset of 3764 news articles in European Portuguese gathered from several European Portuguese fact-checking websites through web-scraping techniques, and Márcia Teixeira [19] also created a dataset of 708 news articles in European Portuguese collected from different sources through web crawling and News Feeds.

B. Pre-processing

Pre-processing in Text Mining is a fundamental stage that involves cleaning and preparing unstructured text data before

carrying out more advanced analysis, such as text classification, topic analysis, and association mining. This process consists of several stages described below.

Stopword filtering is a pre-processing technique focused on removing stop words, such as "a", "an", "in, "the" and more, which are common words that are not very relevant, as a way to allow algorithms to focus on more important ones that help in the posterior machine learning steps.

The **tokenization** of texts is another crucial technique. It involves splitting paragraphs and sentences into smaller units, often single words, and removing spaces and punctuation elements so that algorithms can better understand the data [10].

Stemming involves reducing words into their base forms by deleting prefixes and suffixes (e.g., changing the word "jumps" into "jump"). A process similar to stemming is **lemmatization**, in which inflectional endings from words are removed to obtain the base or canonical form found in dictionaries (e.g., turning "running" into "run"). However, unlike in lemmatization, removing prefixes and suffixes in stemming can cause overstemming or understemming of certain words that may become misleading or not be found in dictionaries (e.g., turning "better" into "bet") [11], [17].

Part of Speech (POS) is another text pre-processing technique that categorizes words as parts of speech, such as verbs, nouns, and adjectives, with tags according to their definition and corresponding context, making it possible to perform a semantic analysis on unstructured text [18].

Sentiment Analysis consists of detecting the polarity (positive or negative) and intensity of sentiments or feelings found in the text. This technique has been increasingly used for the past few years due to increased stylistic techniques found in fake news that affect readers' emotions [6].

C. Feature Extraction

Feature extraction is concerned with converting text into structured format, such as tables composed of words and their frequency, suitable for analysis and machine learning. Some methods for feature extraction are described below.

Term Frequency-Inverse Document Frequency (TF-IDF) is a vectorization model that encodes text using a Vector Space Model (VSM) format. It determines the importance of each term within a single document and a document collection [16].

TF-IDF is often used alongside a **Bag of Words (BoW)**, also known as **Count Vectorizer (CV)**, which is used to determine the frequency of each word found within a document. A list is generated, with each key being the word and each value its corresponding number of occurrences. BoW is similar to **Word Embedding (WE)** like Word2Vec, which creates a vector per word [8].

The **n-gram** model offers the ability to store spatial information, which allows the calculation of not only the term frequency as described before but also the frequency of several terms found after each other. A bi-gram corresponds to a sequence of 2 words (n=2), a tri-gram to a sequence of 3 words (n=3), and so on [13].

D. Feature Selection

Feature selection is a technique used to select the subset of features for training a classification model. There are three main types of methods: Wrapper, Filter, and Embedded.

Wrapper methods iterate through different subsets of features, evaluating the importance of each feature for each iteration as a way to determine the most optimal model with the best combination. **Forward Selection** and **Backward Selection** are two examples of wrapper methods.

Filter methods resort to a ranking procedure to select features based on a useful descriptive measure other than error. Many filter methods, including **Pearson Correlation**, measure the linear correlation between two variables, indicating how relevant the data is. **Information Gain**, in turn, measures the contribution of a term to a classification task based on its presence or absence within a document. There is also the **Chi-square Statistic** method, in which the independence between terms and classes is measured, and more [20].

Embedded methods can be described as the middle term between the previous ones, since the selection and tuning of feature subsets are done during the model creation process.

L2 Regularization is an embedded method that reduces the impact of multicollinearity by decreasing the correlation strength found within less relevant variables.

L1 Regularization is similar to L2 Regularization, with the most significant difference being the ability to remove features from a model, which also contributes to reducing the model's complexity [21].

E. Machine Learning

The text data mining and processing phase is followed by the classification phase, in which the data is fed as input to classification algorithms so that machines can learn and make predictions based on said input data.

Machine Learning falls under the umbrella of Computer Science as a subset of Artificial Intelligence. The algorithms and techniques involved in ML aid computers when analysing problems and making decisions to solve them. Some of them are described below.

Decision Trees is a supervised learning model used in classification and regression. The input data is divided recursively into subsets and each tree node represents the possible outcomes [16].

Random Forest combines a collection of several decision trees to make better predictions. The accuracy of this classifier increases with the number of trees and its predictions are not affected by overfitting or omitted values [12].

Extremely Randomised Trees, also known as Extra Trees, are algorithms similar to Random Forests but with a random subset of features for training to estimate and classify the importance of each characteristic [22].

Adaptive Boosting (AdaBoost) is a boosting algorithm that identifies the misclassified cases and penalises them by assigning more weightage to them to improve the following learning classifiers' performance so that the final model can avoid said "mistakes" [23].

Gradient Boosting (GBoost) is another boosting algorithm. Instead of penalising misclassified cases as found in AdaBoost, the different classifiers are gradually and sequentially trained in GBoost according to a loss function. GBoost typically performs better than AdaBoost, but it is more susceptible to overfitting and longer computational time [24].

Extreme Gradient Boosting (XGBoost) represents an enhancement over the Gradient Boosting technique. The algorithm resorts to regularisation to reduce overfitting and parallel running to improve runtime speed and tree pruning [25].

K-Nearest Neighbours (KNN) is another supervised learning algorithm used to identify patterns and trends by considering k cases when analysing the whole dataset [26].

Logistic Regression is a supervised learning algorithm and is one of the most used algorithms for classification problems. The weighted sum of inputs is mapped between 0 and 1 through a Sigmoid curve (S-curve), which contains a threshold value used for predictions [8].

Linear Regression is also a supervised learning algorithm mainly used for regression problems to find connections between its predictions and respective variables [27].

Support Vector Machine (SVM) is another supervised learning algorithm that is used to solve linear classification and regression problems. The model is created according to a set of previously trained data. Its creation is determined by selecting the best hyperplane, a margin separating the input data into their respective classes [28].

Linear Support Vector Machine (LSVM) is one of the most used algorithms for binary classification problems. The boundary that separates both classes is represented by a straight line, unlike SVM which has different representations.

Naïve Bayes (NB) encompasses a family of classification algorithms that leverage Bayes' Theorem and is used when many independent characteristics influence a decision of a particular class. Moreover, **Multinomial Naïve Bayes (MNB)** is one of the many Naïve Bayes that resort to multinomial distribution. It can identify an article's topic, such as politics, entertainment, and others [19].

F. Deep Learning

Deep Learning is a subfield of Machine Learning that attempts to simulate the human brain's behaviour through neural networks with multiple layers that allow a progressive extraction of higher-level characteristics and features from input data. Some methods are described below.

An **Artificial Neural Network (ANN)** comprises three main layers: an input layer, a hidden layer and, an output layer. These layers are connected with associated weight and threshold levels. The nodes receive inputs, process them, and produce an output, according to a specific threshold, which then translates to the data transmission to the subsequent layer of the network. A **Deep Neural Network (DNN)** is an ANN with more than one hidden layer, with different weights applied to the inputs [11].

A **Convolutional Neural Network (CNN)** is a class of DNNs with a Convolutional Layer in which a map with in-depth convoluted features is created, a Pooling Layer in which the average mean of nearby outputs represents the output of the network, and a Fully Connected Layer, which defines the outputs representation with interconnected nodes [29].

A **Recurrent Neural Network (RNN)** considers information from preceding inputs to determine the current input and output of the sequence, while also sharing parameters across each layer of the network. These aspects allow RNNs to capture relevant contextual information by finding patterns associated with training data dependence [30].

Increasing layers using activation functions makes it harder to train due to the decreasing gradient value, corresponding to a loss of past information. **Long Short-term Memory (LSTM)** is a type of RNN capable of avoiding the decreasing gradient value related to information loss by memorising prior information and interpreting data in different ways [8].

Bi-directional Long Short-term Memory (LSTM) is similar to LSTM, but its input flow is bi-directional, which means both future and past information can be preserved. This feature is accomplished using two separate LSTMs that register information from all input and output nodes [15].

There are also other deep learning models commonly used for NLP tasks, including **Bidirectional Encoder Representations from Transformers (BERT)**, which resorts to surrounding text as a way to grasp the context of it, as we humans do when reading polysemous words, **Extreme Language Understanding Network (XLNET)**, which allows learning bidirectional contexts and overcomes certain limitations of BERT, and **Robustly Optimized BERT Pretraining Approach (RoBERTa)**, which has an improved mask language modeling when compared to BERT [9], [31].

G. Results Evaluation

To measure the performance of the models, one typically resorts to well-known evaluation methods, including **k-fold cross-validation** and respective confusion matrices, alongside metrics such as **Accuracy**, **Precision**, **Recall** and **F1-score**. F1-score is often the best metric in the context of fake news since it also considers false positives and negatives [32].

III. Proposed Approaches

By relying on the several methods described in the previous section, a project pipeline was created with both English and European Portuguese approaches, as illustrated in Fig. 2.

The main difference between the approaches is related to the data acquisition phase. The data in the English approach can be obtained by merging different publicly available datasets. In contrast, a new dataset had to be created from scratch in the European Portuguese approach by resorting to the implementation and use of Web Scrapers to gather real and fake news articles from trustworthy and untrustworthy European Portuguese websites.

Regarding the English Approach, three datasets were considered to create the final English dataset:

1) The "WELFAKE dataset", which contains 72,134 news articles from Kaggle, McIntire, Reuters, and BuzzFeed Political, as already described before.

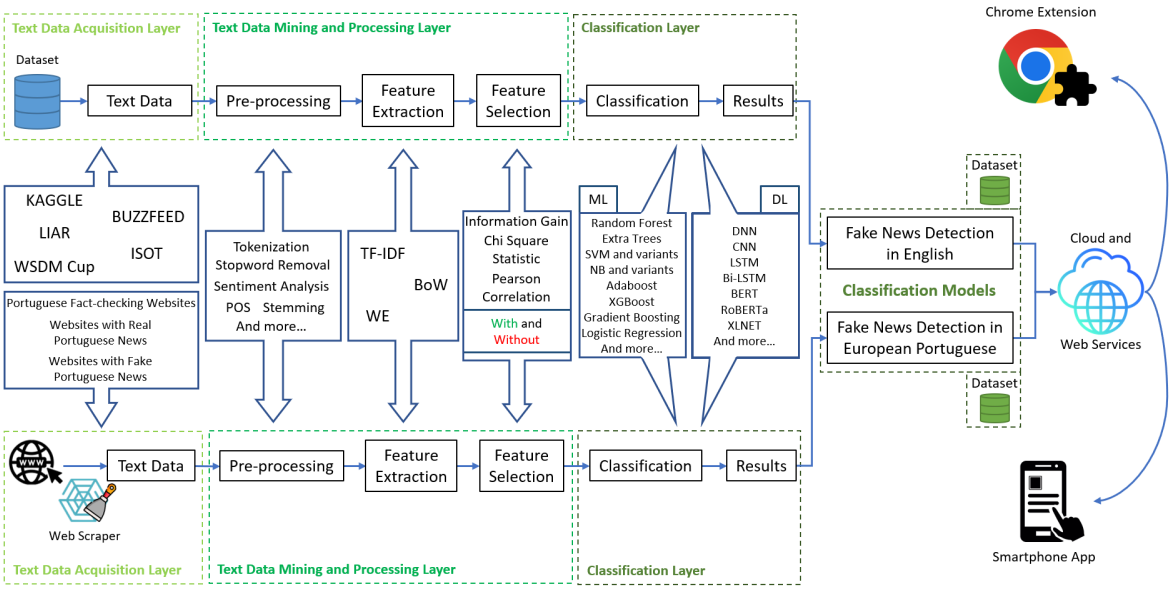
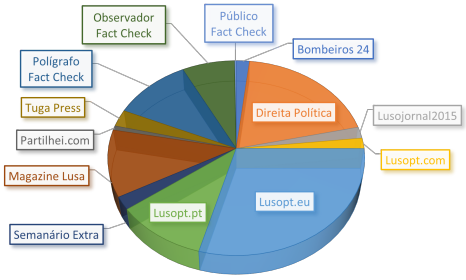


Fig. 2. Pipeline of the project

2) The "Misinformation & Fake News text dataset" created by Steven from Kaggle, which contains 79,000 misinformation, fake news, and propaganda articles.
3) The "COVID-19 Fake News dataset" created by Patwa et all. as previously mentioned, with a total of 10,479 statements from several social media posts, comments, and news articles.

Once cleaned, the English dataset contained over 140,000 articles, totalling 72,925 fake and 68,167 real news articles.

Regarding the European Portuguese approach, a total of 31,716 real and 31,520 fake news were considered from many different websites. Figs. 3 and 4 show their distribution, along with the number of articles and statements scraped.



(a) Fake news distribution chart

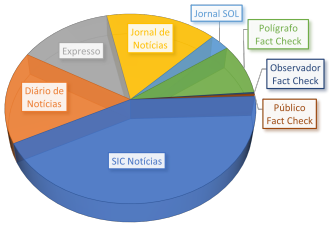
Fake News	
Website	Count
Bombeiros 24	602
Direita Política	6425
Lusojornal2015	623
Lusopt.com	574
Lusopt.eu	9620
Lusopt.pt	3486
Semanário Extra	710
Magazine Lusa	3652
Partilhei.com	238
Tuga Press	927
Polígrafo Fact Check	3559
Observador Fact Check	2385
Público Fact Check	80

(b) Fake news websites

Fig. 3. Portuguese fake news distribution chart and table

Fake news articles were gathered from a wide variety of websites flagged as unreliable news sources with the intent to manipulate and deceive readers, based on thorough studies in this field [33], [34], [35]. Real news articles were extracted from trustworthy news websites with a strong presence in the daily lives of millions of Portuguese people.

The extraction of data from websites implies understanding how information is distributed according to HTML format,



(a) Real news distribution chart

Real News	
Website	Count
SIC Notícias	14718
Diário de Notícias	5369
Expresso	4537
Jornal de Notícias	5369
Jornal SOL	950
Polígrafo Fact Check	2925
Observador Fact Check	90
Público Fact Check	143

(b) Real news websites

Fig. 4. Portuguese real news distribution chart and table

which describes the structure of web documents. The content on a page is structured by creating an architecture of text documentation in the form of elements identified by tags.

These elements are used by browsers such as Chrome, Edge, Firefox, and Safari to determine what content or information to display. This includes titles, headings, subheadings, paragraphs, hyperlinks, tables, lists, basic text, images, interactable components such as buttons and text boxes, and much more.

The Beautiful Soup package allowed the extraction of news data from each website by sending several requests to specify the HTML tags associated with the desired information. Selenium was also used given that some websites required user interaction to load and display the desired content. This included the need to click buttons, scroll up and down, close tabs or pop-ups, and more before actually retrieving the data with Beautiful Soup.

Besides fake and real news websites, many fact-checks were also considered from reputable sources such as Polígrafo, Público, and Observador. Fig. 5 shows the rating categories used by each fact-checking website and the corresponding label set when creating the European Portuguese dataset.

Unlike Rodrigues [18] who resorted to each fact-check's overview in the main webpage to gather data, the data from

Polígrafo Rating	Label
Verdadeiro	Real
Verdadeiro, mas...	Real
Descontextualizado	Fake
Impreciso	Fake
Manipulado	Fake
Falso	Fake
Pimenta na Língua	Fake

(a) Polígrafo's conversion

Observador Rating	Label
Certo	Real
Praticamente certo	Real
Esticado	Fake
Inconclusivo	Fake
Enganador	Fake
Errado	Fake

(b) Observador's conversion

Público Rating	Label
Verdadeiro	Real
Parcialmente verdadeiro	Real
Parcialmente falso	Fake
Falso	Fake
Inconclusivo	Fake

(c) Público's conversion

Fig. 5. Fact-checking websites' categories conversion

fact-checks in this project was extracted from the actual articles instead of their overviews, more specifically the first 4 paragraphs in which the context of the fact-checked statement was described, thus improving the performance of the models.

The topics and fields discussed by the extracted articles were often divided into different categories on each website such as politics, science, sports, fame, justice, society, and more. There were also both real and fake news articles about COVID-19 and the ongoing war between Russia and Ukraine. Some fake news websites also had articles tagged as jokes or satire, which were ignored during the web scraping step.

It is also important to mention that many fake news websites had news from credible sources as well, including ones that were used to extract real news data. This was probably done in an attempt to fool readers into thinking that each of these websites was trustworthy, which made their misinformative nature not so obvious and easily spotted. To avoid possible incorrect labels, all news articles from misinformation websites with credible sources were ignored by the web scrapers, while those with an untrustworthy source were labelled as fake.

The steps that followed the data acquisition phase were the same for both English and European Portuguese approaches. To transform the data in multiple ways, several combinations of pre-processing, feature extraction, and feature selection techniques were considered in different orders. This approach exposes the ML and DL models to a wide variety of inputs, which help maximise their performance by resorting to k-fold cross-validation and respective metrics, alongside the customization of hyperparameters.

The resulting datasets and classification models are then available to users via cloud and web-based services, as depicted in Fig. 9. The cloud platform comprises a Flask RESTful application run on a docker container inside an AWS EC2 instance. A Chrome extension and an Android application were also developed. Through POST and GET requests, they communicate with the web application, allowing users to check whether a given news article is real or fake.

Users can also report fake or real news articles, which are then processed in a script running on a separate computer with a dedicated Graphical Processing Unit (GPU). The script resorts to a DistilBERT model capable of capturing context and sentence embeddings, which alongside Cosine Similarity allows the program to measure the agreement among users for a particular text or statement.

A representative text is then selected for each similarity group that is generated. To overcome the challenges raised by similar but contradicting texts or more complex or subtle negations, the Law of Large Numbers was considered. In the context of fake news detection, this Law states that between two contradicting labels, the number of users that reported the correct one will in the end be higher than its counterpart [36].

The models are then fine-tuned with the filtered feedback data by resorting to Transfer Learning. The data patterns initially learned are transferred to a new model and a top layer is added to be specifically trained on feedback data. This method ensures previous knowledge is retained and the received feedback is able to either teach the model about new topics or unfolding events or improve already existing ones within the scope of its knowledge, as depicted in Fig. 6.

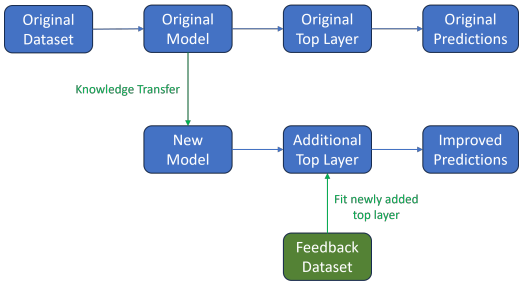
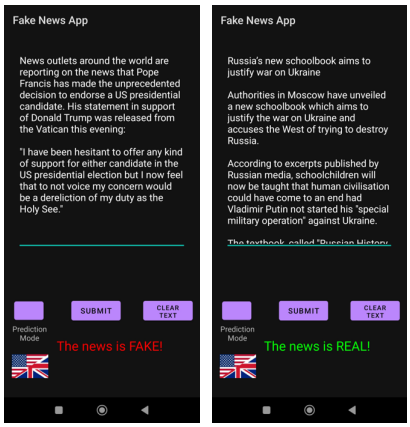


Fig. 6. Model improvement through Transfer Learning

Finally, the models are sent over to the cloud instance through Secure Shell (SSH) and Secure File Transfer Protocol (SFTP) commands, as well as a POST request to update them in Flask. Figs. 7 and 8 show different examples of fake and real news being correctly predicted in the Android app and Chrome Extension, respectively.



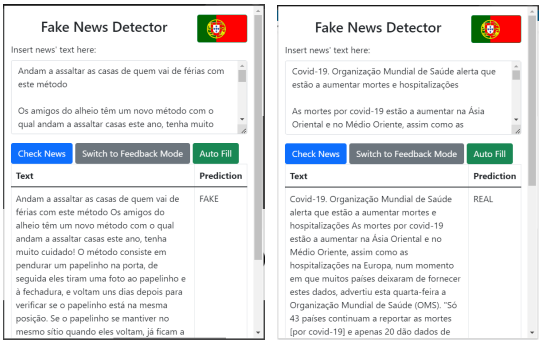
(a) Fake News (b) Real News

Fig. 7. Example of fake and real predictions of English news on the Android app

IV. Results

Despite both ML and DL models being considered during the content classification phase, the resulting predictions of news credibility were significantly different in each approach. The best results in the English approach corresponded to DL models, as shown in Table I.

The English models struggled when working with the different pre-processing, feature extraction, and feature selection



(a) Fake prediction on Portuguese news

(b) Real prediction on Portuguese news

Fig. 8. Example of fake and real predictions of European Portuguese news on the Chrome extension

TABLE I
BEST ENGLISH MODELS

	Accuracy	Precision	Recall	F1-score
DNN	0.55	0.59	0.54	0.48
CNN	0.58	0.69	0.57	0.49
LSTM	0.63	0.7	0.62	0.58
Bi-LSTM	0.66	0.68	0.65	0.64
BERT	**0.96**	**0.96**	**0.96**	**0.96**
RoBERTa	0.77	0.77	0.77	0.77
XLNET	0.83	0.83	0.83	0.83

techniques, which translated to a much worse performance when compared to previous projects. This was a surprising outcome since the models are commonly used for NLP tasks.

Nevertheless, the best performance with the final dataset was observed when tokenizing the text data, with BERT achieving the best results with a macro average F1-score of 0.96 by tuning the following layers and hyper-parameters:

- Input shape of 128
- A single dense layer with "relu" as the activation function
- Dropout layer set to 0.2
- Output layer with "sigmoid" as the activation function
- Learning rate of 2e-5
- Train/test and train/validation split set to 0.2
- Batch size of 64
- 2 epochs

The best results in the European Portuguese approach corresponded to ML models, as shown in Table II.

Unlike the English ones, the European Portuguese models showed better results with many different pre-processing and feature extraction techniques. Besides using stopword removal and lemmatization to pre-process the text data, POS tagging and Sentiment Analysis were also considered.

The source of each news article also helped improve the results since many fake news websites promoted news articles from other fake news websites, thus helping classifiers establish similar patterns shared between the news from fake news websites while also differentiating them from real ones.

Out of all the models, XGBoost achieved the best results with a macro average F1-score of 0.957 after tuning the following hyper-parameters in a Grid Search with 5 folds:

- Colsample by tree set to 0.8

TABLE II
BEST EUROPEAN PORTUGUESE MODELS

	Accuracy	Precision	Recall	F1-score
Decision Trees	0.95	0.95	0.95	0.948
Random Forest	0.92	0.92	0.92	0.92
Extra Trees	0.92	0.92	0.92	0.92
XGBoost	**0.96**	**0.96**	**0.96**	**0.957**
AdaBoost	0.94	0.94	0.94	0.94
Gradient Boosting	0.95	0.95	0.95	0.95
Bagging	0.95	0.95	0.95	0.95
Logistic Regression	0.95	0.95	0.95	0.952
SVC	0.94	0.94	0.94	0.94
Linear SVC	0.95	0.95	0.95	0.947
Multinomial NB	0.91	0.91	0.91	0.91
Bernoulli NB	0.87	0.87	0.87	0.87
KNN (k=20)	0.85	0.85	0.85	0.85

- Learning rate set to 0.2
- Max depth set to 5
- N estimators set to 300
- Subsample set to 1

The results showed that DL models performed much better in the English approach than ML models, especially pre-trained language models such as XLNET and BERT, which are known for their ability to capture the context of sentences and generate better word embeddings.

Since the European Portuguese dataset is less than half the size of the English one, this behaviour could be related to DL models typically performing much better in bigger datasets than ML models [37].

The fact that ML models are simpler and yet achieved better results in the European Portuguese approach than DL models also corroborates this idea. This aspect could also raise the possibility of a loss in text context and patterns by oversimplifying the English data with pre-processing, feature extraction, and feature selection techniques.

On the other hand, the combination of the three English datasets into a single one may have significantly increased the complexity of the patterns as well, to the point where most models struggled during the training phase. This is a very plausible cause, especially because the same models achieved much better results when trained in each dataset separately.

Although different projects and datasets were explored and considered regarding the detection of fake news in English, these only represent a fraction of the total number of approaches that have been developed in this context, thus making it hard to compare the BERT model of the English approach with all the remaining projects.

On the other hand, the Portuguese approach had its data extraction phase much more limited due to the scarcity of public datasets. This aspect also applies to the other projects explored in European Portuguese that also resorted to similar sources, which translates to a much more coherent comparison. The main aspects of said comparison are shown in Table III.

The proposed approach can be summarized as a mix of the other approaches with extra features and characteristics. Both fact-checking and news websites were considered when extracting data, and the resulting dataset is much bigger than the others. Besides extracting much more data than the

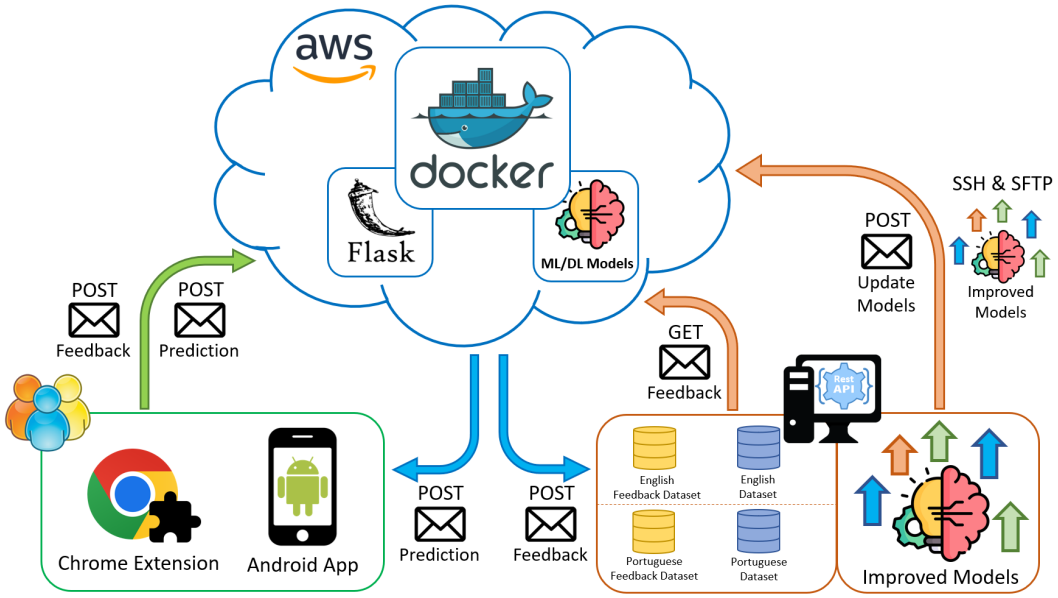


Fig. 9. Processes behind the application development and deployment phases

TABLE III
PORTUGUESE APPROACHES COMPARISON

	Rodrigues [18]	Teixeira [19]	**Our Approach**
Sources	Fact-checks	News Websites	**Fact-checks, News Websites**
Dataset Size	3764 (872 real and 2889 fake)	718 (543 real and 175 fake)	**63,236 (31,716 real, 31,520 fake)**
Best Pre-process	Punctuation Tokenization Stopwords Lowercase	Stemming Tokenization Stopwords Lowercase	**Lemmatization Stopwords Sentiment Score POS Tagging**
Best Feature Extraction	BoW, TF-IDF and One Hot	TF-IDF (Uni and Bi-grams)	**TF-IDF (Uni, Bi, Tri-grams)**
Best Model	Extra Trees (0.74 F1-score)	MNB (0.95 Accuracy)	**XGBoost (0.957 F1-score)**

previous European Portuguese projects, by resorting to web scraping, a much wider variety of sources was also considered.

These aspects play an essential role in acquiring a more representative sample of fake and real news found in real life, which consequently translated into developing the first publicly available fake news dataset in European Portuguese.

Sentiment Analysis had not been considered before in this language, and it proved very useful, given its impact on model performance, alongside the previously mentioned techniques. XGBoost achieved the best F1-score, which was also higher than previous work, thus proving the impact of boosting algorithms in ML tasks, specifically in fake news detection.

This achievement is even more remarkable given the fact that much more data was considered when training the models than the previous projects, which can be challenging due to the increase of data patterns, thus leading to the development of much more robust and complex models designed to detect fake news in this language.

A Chrome extension and Android application were developed alongside the Flask application ran on Docker and deployed on the AWS EC2 cloud instance. This allowed the creation of a system that can be easily scalable and in which the developed models can be put to practice, proving their usefulness and possible impact when protecting their users from the dangers of fake news.

Furthermore, developing a RESTful API capable of improving the models by filtering user feedback is a valuable step, as it contributes to enhancing the ML and DL models and, ultimately, the system as a whole. This feature was not considered in the previous projects.

The results of this research work are available in our GitHub repository with all the datasets, the cloud-based web application source code, the Flask and Android applications, and the ML and DL models [38].

V. CONCLUSION

Fake news threatens our society with emerging forms of manifestation and the harmful effects they cause. It requires an appropriate response by developing more sophisticated tools to combat them. This research work contributed to this effort, focusing on obtaining models for detecting fake news in English and European Portuguese, with a more significant contribution in the latter.

The developed models achieved remarkable performance. The wider variety of data and techniques used in the European Portuguese approach allowed the creation of more robust and complex models with better performance when compared to previously developed projects.

The developed work also led to the creation of the first public dataset for fake news detection in European Portuguese through web scraping due to the lack of public datasets. This is

a commendable step in addressing the challenge of this field, as it can pave the way for more research and innovation in this specific language.

The cloud platform with the Flask application, the Chrome extension, and the Android application are available in a public repository. The project can be used to develop a platform that can be scaled and benefit a broader range of users, protecting them from the influence of fake news. Furthermore, the platform can receive user feedback, which allows the continuous improvement of the classification models.

Despite its proven impact and contribution to the fight against the threat of fake news, the developed system also has its limitations, including the need to consider more data and more news categories over a longer period.

Furthermore, only a well-informed human with access to trustworthy sources at all times can effectively determine whether the information is real or fake, which is why the predictions generated by models should be seen as auxiliary indicators to warn and protect users from misinformation.

It is crucial to develop models that are able to successfully filter online content specifically in European Portuguese given the differences found between Portuguese dialects. These are not limited to sentence-level aspects, but also the news content itself, which is bound to differ taking into account the diverse realities found in the different countries.

As a consequence, additional ML and DL models and larger datasets from more sources can be considered for future work, alongside their deployment on a cloud instance with more scalable resources. It would also be interesting to develop more methods to filter user feedback data, as this would have a direct impact when fine-tuning the models with even more relevant data.

Moreover, testing the models with other variants of Portuguese could allow the development of a system capable of detecting online misinformation in different dialects. However, this would naturally require much more data from a significantly wider variety of sources and fields in all the Portuguese variants taken into account.

As newer manifestations of fake news emerge, so does the need to adapt and create better tools, which is probably a never-ending effort.

ACKNOWLEDGEMENTS

This research was supported in part by the Portuguese FCT program, Center of Technology and Systems (CTS) UIDB/00066/2020 / UIDP/00066/2020.

REFERENCES

[1] U. Gneezy, "Deception: The role of consequences," *American Economic Review*, vol. 95, pp. 384–394, March 2005.

[2] X. Zhang and A. A. Ghorbani, "An overview of online fake news: Characterization, detection, and discussion," *Information Processing & Management*, vol. 57, p. 102025, Mar. 2020.

[3] C. Shao, G. L. Ciampaglia, O. Varol, K. Yang, A. Flammini, and F. Menczer, "The spread of low-credibility content by social bots," *Nature Communications*, vol. 9, p. 4787, Nov. 2018.

[4] J. McGarrigle, "Explained: What is Fake news? | Social Media and Filter Bubbles," Available at https://www.webwise.ie/teachers/what-is-fake-news/, 2018.

[5] S. Maheshwari, "10 Times Trump Spread Fake News," Available at https://nyti.ms/3Qb9kA6, 2017.

[6] S. Khan, S. Hakak, N. Deepa, B. Prabadevi, K. Dev, and S. Trelova, "Detecting COVID-19-Related Fake News Using Feature Extraction," *Frontiers in Public Health*, vol. 9, p. 788074, Jan. 2022.

[7] M. Holroyd, "Five of the most viral misinformation posts since Ukraine war began," Available at https://www.euronews.com/my-europe/2022/08/24/ukraine-war-five-of-the-most-viral-misinformation-posts-and-false-claims-since-the-conflic, 2022.

[8] A. Abdulrahman and M. Baykara, "Fake News Detection Using Machine Learning and Deep Learning Algorithms," in *2020 International Conference on Advanced Science and Engineering (ICOASE)*, pp. 18–23, IEEE, Dec. 2020.

[9] W. Antoun, F. Baly, R. Achour, A. Hussein, and H. Hajj, "State of the Art Models for Fake News Detection Tasks," in *2020 IEEE International Conference on Informatics, IoT, and Enabling Technologies (ICIoT)*, pp. 519–524, IEEE, Feb. 2020.

[10] Z. Khanam, B. N. Alwasel, H. Sirafi, and M. Rashid, "Fake News Detection Using Machine Learning Approaches," *IOP Conference Series: Materials Science and Engineering*, vol. 1099, p. 012040, Mar. 2021.

[11] A. Thota, P. Tilak, S. Ahluwalia, and N. Lohia, "Fake News Detection: A Deep Learning Approach," *SMU Data Science Review*, vol. 1, no. 3, 2018.

[12] I. Ahmad, M. Yousaf, S. Yousaf, and M. O. Ahmad, "Fake News Detection Using Machine Learning Ensemble Methods," *Complexity*, vol. 2020, pp. 1–11, Oct. 2020.

[13] S. Mishra, P. Shukla, and R. Agarwal, "Analyzing Machine Learning Enabled Fake News Detection Techniques for Diversified Datasets," *Wireless Communications and Mobile Computing*, vol. 2022, pp. 1–18, Mar. 2022.

[14] P. K. Verma, P. Agrawal, I. Amorim, and R. Prodan, "Welfake: Word embedding over linguistic features for fake news detection," *IEEE Transactions on Computational Social Systems*, vol. 8, no. 4, pp. 881–893, 2021.

[15] S. R. Sahoo and B. Gupta, "Multiple features based approach for automatic fake news detection on social networks using deep learning," *Applied Soft Computing*, vol. 100, p. 106983, Mar. 2021.

[16] P. Patwa, S. Sharma, S. Pykl, V. Guptha, G. Kumari, M. S. Akhtar, A. Ekbal, A. Das, and T. Chakraborty, "Fighting an infodemic: COVID-19 fake news dataset," in *Combating Online Hostile Posts in Regional Languages during Emergency Situation*, pp. 21–29, Springer International Publishing, 2021.

[17] R. M. Silva, R. L. S. Santos, T. A. Almeida, and T. A. S. Pardo, "Towards automatically filtering fake news in Portuguese," *Expert Systems with Applications*, vol. 146, p. 113199, 2020.

[18] J. F. C. Rodrigues, "Fake News Classification in European Portuguese Language," Available at http://hdl.handle.net/10071/22194, 2020.

[19] M. R. P. Teixeira, "Índice de Credibilidade de Conteúdos Noticiosos em Língua Portuguesa para Uso em Ambiente Escolar," Available at http://hdl.handle.net/10400.22/18330, 2021.

[20] F. P. Shah and V. Patel, "A review on feature selection and feature extraction for text classification," in *2016 International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET)*, pp. 2264–2268, IEEE, Mar. 2016.

[21] B. Venkatesh and J. Anuradha, "A review of feature selection and its methods," *Cybernetics and Information Technologies*, vol. 19, no. 1, pp. 3–26, 2019.

[22] P. Geurts, D. Ernst, and L. Wehenkel, "Extremely randomized trees," *Machine Learning*, vol. 63, pp. 3–42, Apr. 2006.

[23] R. Wang, "AdaBoost for Feature Selection, Classification and Its Relation with SVM, A Review," *Physics Procedia*, vol. 25, pp. 800–807, Jan. 2012.

[24] A. Natekin and A. Knoll, "Gradient boosting machines, a tutorial," *Frontiers in Neurorobotics*, vol. 7, p. 21, Dec. 2013.

[25] T. Chen and C. Guestrin, "XGBoost: A Scalable Tree Boosting System," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 785–794, Aug. 2016.

[26] Z. Zhang, "Introduction to machine learning: k-nearest neighbors," *Annals of Translational Medicine*, vol. 4, p. 218, June 2016.

[27] D. Maulud and A. Mohsin Abdulazeez, "A Review on Linear Regression Comprehensive in Machine Learning," *Journal of Applied Science and Technology Trends*, vol. 1, pp. 140–147, Dec. 2020.

[28] T. Evgeniou and M. Pontil, "Support Vector Machines: Theory and Applications," in *Machine Learning and Its Applications, Advanced Lectures*, vol. 2049, pp. 249–257, Sept. 2001.

[29] S. Indolia, A. K. Goswami, S. P. Mishra, and P. Asopa, "Conceptual Understanding of Convolutional Neural Network- A Deep Learning Approach," *Procedia Computer Science*, vol. 132, pp. 679–688, Jan. 2018.
[30] N. M. Rezk, M. Purnaprajna, T. Nordström, and Z. Ul-Abdin, "Recurrent Neural Networks: An Embedded Computing Perspective," *IEEE Access*, vol. 8, pp. 57967–57996, 2020.
[31] L. S. Moreira, G. M. Lunardi, M. d. O. Ribeiro, W. Silva, and F. P. Basso, "A Study of Algorithm-Based Detection of Fake News in Brazilian Election: Is BERT the Best?," *IEEE Latin America Transactions*, vol. 21, pp. 897–903, Sept. 2023.
[32] H. Dalianis, "Evaluation Metrics and Evaluation," in *Clinical Text Mining: Secondary Use of Electronic Patient Records*, pp. 45–53, Springer International Publishing, 2018.
[33] M. Sintra, "Fake News e a Desinformação: Perspetivar comportamentos e estratégias informacionais," Available at http://hdl.handle.net/10362/79564, 2019.
[34] Observador, "Como é o mundo clandestino dos "sites" em português associados às 'fake news'," Available at https://observador.pt/2019/02/20/como-e-o-mundo-clandestino-dos-sites-em-portugues-associados-as-fake-news/, 2019.
[35] D. de Notícias, "Diário de Notícias: "Fake news: sites portugueses com mais de dois milhões de seguidores"," Available at https://apav.pt/apav_v3/index.php/pt/1866-diario-de-noticias-fake-news-sites-portugueses-com-mais-de-dois-milhoes-de-seguidores, 2018.
[36] K. Sedor, "The Law of Large Numbers and its Applications," Available at https://www.lakeheadu.ca/sites/default/files/uploads/77/images/Sedor%20Kelly.pdf, 2015.
[37] "Deep Learning vs Machine Learning: The Ultimate Battle," *Turing*, Available at https://www.turing.com/kb/ultimate-battle-between-deep-learning-and-machine-learning, 2022.
[38] R. Afonso, "fake-news-pt-eu," Available at https://github.com/ro-afonso/fake-news-pt-eu, 2024.

Ricardo Afonso has a Master's degree in Electrical and Computer Engineering from NOVA School of Science and Technology, Portugal. His research interests include the world of data and Artificial Intelligence.

João Rosas received his PhD in Electrical Engineering in 2010 from NOVA School of Science and Technology, Portugal, where he is currently a Professor in the Department of Electrical and Computer Engineering. His research interests are in the Internet of Things, Digital Twins, Information Systems, Digital Games, and Machine Learning. He has several publications in international journals, conference proceedings, and book chapters. He has participated in several research projects funded by the European Commission.