# A short review of some mathematical methods to detect fake news

Giuseppe Giordano[a], Serena Mottola[b], Beatrice Paternoster[a]

[a] Department of Mathematics, University of Salerno, Via Giovanni Paolo II, 132
84084 Fisciano, Italy
[b] Department of Economics and Law, University of Naples 'Parthenope', Via Amm. F. Acton, 38
80133 Napoli, Italy

**Abstract-** In this work we aim to illustrate some mathematical methods recently appeared in the scientific literature to detect fake news. The problem of fake news is an increasingly present topic in our society, from public debate to scientific research. The number of fake news produced is constantly increasing especially for the advantages of those who spread them. In fact, emotionally compelling news, in line with our thoughts, capture our attention, and lead to clicks and views, in the hope of attracting advertising. Understanding whether a news is false or not is not an easy problem to solve, given the large amount of data present on the internet. The detection mechanism should predict the information very quickly in order to stop the spread of fake news. This work is a review of four methods to detect fake news recently appeared in the literature [22, 33, 39, 47]. Different methodologies are observed among the various methods: statistical approach, artificial neural network, artificial intelligence and text approach. Furthermore, some results are shown.

**Keywords-** Fake News, Machine Learning, Naive Bayes Classifier

## I. INTRODUCTION

**F**AKE news is a term that is used to define articles that present invented, misleading information, created to misinform and make hoaxes viral through the internet.

Spreading fake news has always subjected the human brain to the need for a critical assessment but putting it into practice costs effort, although every individual grows up with defensive prejudices deriving from education, culture, environment.

Fake news cover all topics, political, sports, customs, science and medicine. For the last aspect, we can report the relationship between vaccines and autism. In fact, for a few years, a sharp drop in vaccinations has been observed, given by the diffusion of an article that correlated the appearance of autism with the use of vaccines. But very large studies, carried out subsequently, have shown that the incidence of this disorder is identical in vaccinated and unvaccinated children.

Although the topic of fake news has recently become of great interest, this problem has always existed. What seems to have clearly changed today compared to the past, is the amount of fake news circulating in information, especially on-line, and the weight that they are taking on. The most common mistake that is surely made today, in a society where the digital is increasingly integrated with our lifestyle, is to think that access to a large amount of data is automatically synonymous with freedom and democracy [20].

In today's society, fake news is spreading significantly faster, compared to real news, resulting from the birth of social media. In fact, their low cost, easy access, and rapid dissemination of information, lead people to seek out and consume news from social media. Various studies show how in the last decades the quantity of fake new on social media platforms has increased massively, with serious consequences in many decision-making processes, such as 2016 elections in USA [2, 26]. For this reason, the European Commission in 2018 reported the necessity of a deeper fact-checking activity from both journalists and social media platforms [18]. Moreover, to support a correct information on the topic, the Council of Europe in 2017 published a study about different types of "information disorder", which include – but is not limited to – fake news. While journalists are required to verify a story before sharing it, social media users are not. It's easy to understand how dangerous can be the use of social media platforms as primary source of knowledge and information.

The stronger is the content of the published post, the quicker it will circulate on social platforms. In order to understand the phenomena of fake news it is crucial to analyse it in relation to how social media work. When a false news become viral, there is no way to stop it. People who create fake news in most cases use powerful images, tough language and harsh words because it has been demonstrated that this kind of contents is most likely to be shared because of its emotional and persuasive power. [2, 7, 26, 29–31].

It has been demonstrated that fake news usually circulate faster and more broadly than the truth. A study published from the MIT in 2018 revealed that fake

news are 70% more likely to be shared on social media than real news, especially if the topic is politics [45]. Also, from an emotional point of view, fake news broad diffusion is related to the feelings they stimulate in the users, which are used to allow themselves to be bullied into a content that simply confirm them their thoughts and ideas, even if it's not the truth [1].

In times of emergency, such as that of COVID-19, fake news make its entrance markedly, especially on the web and social media, for their easy diffusion. During a pandemic, the fight against disinformation is crucial, especially for public health. Just like the virus, rumours and falsehoods flow beyond borders, causing, what is now called the World Health Organization (WHO), **universal infodemic** [9].

Some examples of fake news produced concern the relationship between 5G technology and the spread of COVID-19, as seen in the figure 1, stating that the radiation of these 5G towers are the real cause of virus deaths.



Fig. 1: Fake news articles about the relationship between 5G and COVID-19. Source: poynter.org
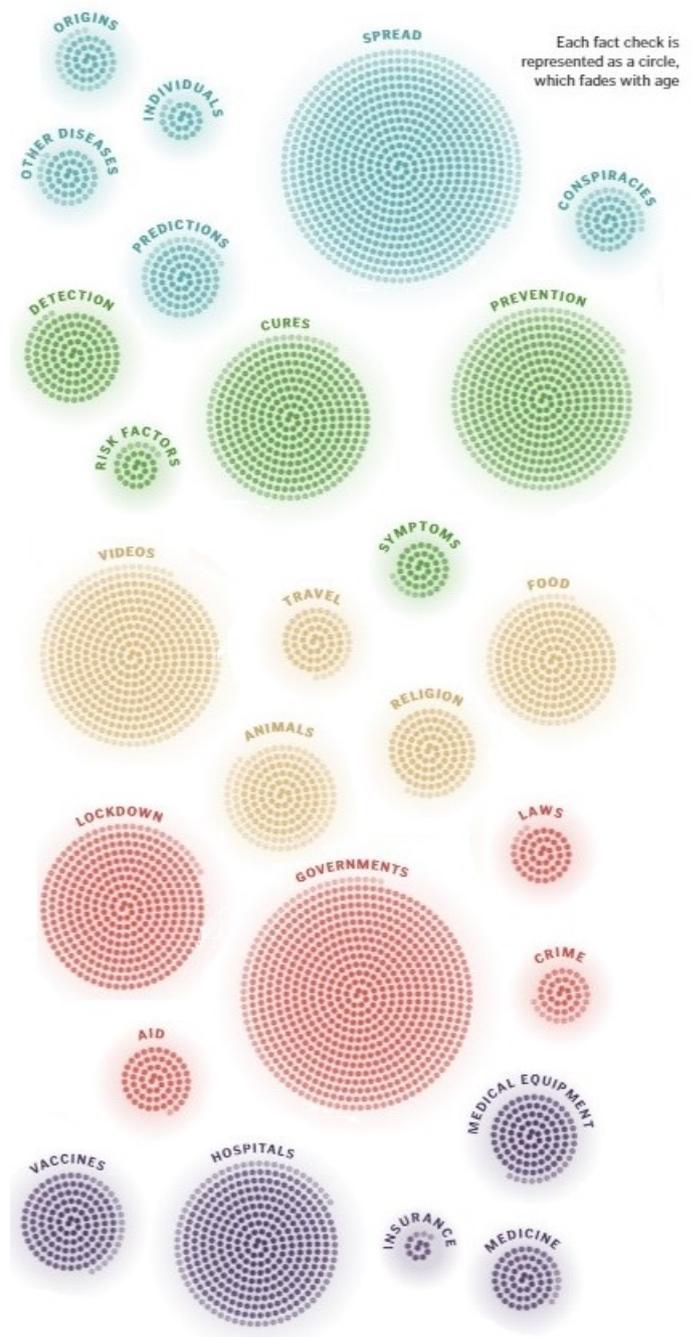


Fig. 2: Waves of hoaxes of COVID-19.Source: poynter.org

To try to counter the spread of fake news, a CoronaVirusFacts Alliance was created at the Poynter Institute, where the data database containing fake news is updated daily to insert new publications. For example, figure 2 shows the categories of fact-checks that help the alliance to identify successive waves of misinformation.

In the last few years the interest of the academia for the study of fake news is grown up, in particular, in the field of linguistics and communication sciences. Talking about fake news we talk about languages and discourses and about the use of certain types of terms and expressions to obtain a result and to share a particular message.

Even more thank texts, images play a fundamental role in fake news construction and have to be considered like texts, with their own grammar and syntax and the strong capacity of sending a message to the reader and persuade him to believe in it. For these reasons, we think that the contribution of linguists and communication experts is fundamental in order to reach a deep comprehension of the phenomena of disinformation and false news on social media platforms [30].

The papers is organized as it follows. In the second section we analysed a well-known statistic technique, the classification. In particular, we have treated the Naive Bayes classifier and how it is used for the treatment of fake news [22]. In the third section we have analysed a recent method to detection fake news [39]. We first reviewed the artificial neural networks and then we analysed the CSI method. Numerical tests are also reported on two very recent datasets such as Twitter and Weibo. In the fourth section, we analysed a method to detect fake news using text mining and intelligence artificial techniques [33]. In particular, we first described a text mining phase, in which the text is divided into smaller parts, and the stop-words are deleted. In the second phase, artificial intelligence algorithms are used to understand if the news is fake or not. In the fifth section, we reviewed a model to detect fake news using N-grams [47]. N-grams are a sequence of characters or numbers of n-dimensions. In particular, in this method there are two steps. In the first step, the size of the data is reduced, and in the second step two techniques are used to extract features from the text.

## II. Naive Bayes Classification

In this section we recall the statistic technique, named classification, that has the aim to detect a sub-population, with certain characteristics, in a larger set. *Naive Bayes classifiers* are a family of classifiers based on Bayes theorem. The Naive-Bayes algorithm is fairly simple but it has been found to be able to behave very well for problems of classification of textual documents. It assumes that all the features are independent of each other. This assumption implies that presence or absence of a feature does not influence the presence or absence of any feature. For sake of completeness, let us recall what conditional probability is, on which Bayes theorem is based. It works on conditional probability, which is the probability that an event will happen, given that a certain event has already occurred.
Bayes'theorem asserts that:

$$P(H|E) = \frac{P(E|H)P(H)}{P(E)} \qquad (1)$$

where,

- $P(H)$ is the prior probability of event $H$. Prior probability means that it does not consider any information about the occurrence of event $E$;

- $P(E)$ is the prior probability of event $E$;

- $P(H|E)$ is the conditional probability of event $H$ given that $E$ is true;

- $P(E|H)$ is the conditional probability of event $E$ given that $H$ is true.

For more details on the Naive Bayes Classifier see the paper [19].
To classify fake news, the following concept is used: fake news articles often use the same set of words, while true article have a particular set of words [22]. From (1) and in [22], the conditional probability that a news is a fake news is calculated and it is given by (2):

$$P(F|W) = \frac{P(W|F)P(F)}{P(W|F)P(F) + P(W|T)P(T)} \qquad (2)$$

where,

- $P(F|W)$ is the conditional probability that a news article is fake given that word $W$ appears in it;

- $P(W|F)$ is the conditional probability of finding word $W$ in fake news articles;

- $P(F)$ is the probability that given news article is fake news article;

- $P(W|T)$ is the conditional probability of finding word $W$ in true news article;

- $P(T)$ is the probability that given news article is true.

Let us suppose that the conditional probabilities $P(F|W)$ are known for each word of the article. In the next step, equation 3 combines these probabilities to calculate the probability that a news article is fake.

$$p_1 = P(F|W_1) \cdots P(F|W_n)$$
$$p_2 = (1 - P(F|W_1)) \cdots (1 - P(F|W_n)) \qquad (3)$$
$$p = \frac{p_1}{p_1 + p_2}$$

where,

- $n$ is the number of words in the news article;

- $p_1$ is the product of the probabilities that a news article is fake given that it contains a specific word for all of the words in the news article;

- $p_2$ is the same of $p_1$ but with the conditional probabilities;

- $p$ is the probability that a news article is fake.

The computation of the conditional probabilities to find that a specific word is a fake news article, is done by considering a *training set*, that contains a lot of news articles, labelled as true or false. Therefore this probability is calculated as the ratio of the fake news articles, that contain this word, to the total number of fake news articles.
In the same way, it defines the probabilities for true news articles.

**Rating by Category**

■ MOSTLY FALSE  ■ MIX. OF TRUE AND FALSE  ■ MOSTLY TRUE
□ NO FACTUAL CONTENT

**Mainstream**

| 94.8% |
0.0%  0.7%                                        4.5%

**Left**

| 14.4% | 56.3% | 24.6% |
4.7%

**Right**

| 12.3% | 25.4% | 47.9% | 14.4% |
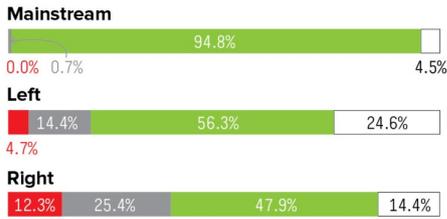
Fig. 3: Subdivision of the BuzzFeed dataset

### A. Dataset

In *BuzzFeed* [42] is present a dataset containing information about Facebook posts, during the American elections in 2016. The results are taken by [22]. They were collected in three large Facebook pages each from right and left political orientation, as well as three large mainstream political news pages. The team of journalists, during the period analysed, found that right-wing pages were more prone to sharing false information than left-wing pages. Mainstream pages didn't share any false information, but did publish a small number of posts that included unverified claims. Figure 3 shows this division in detail.

### B. Test of model

The implementation of the method is based on the following consideration [22]:

- News articles with label "mix of true and false" were not considered;

- The dataset was randomly shuffled and after was divided into three subset:

  1) *Training dataset*: it was used to train the Naive Bayes classifier;

  2) *Validation dataset*: it was used to tune some global parameter of the classifier;

  3) *Test dataset*: it was used to get an estimation of how well the classifier performs on new data

- If all of the words in the news article are unknown, the classifier reports that it can't classify the article.

To measure the accuracy of classifier, we define:

- **TP** (True Positive) is the number of news articles correctly classified as fake;

- **FP** (False Positive) is the number of news articles incorrectly classified as fake.

The precision of classifier is:

$$precision = \frac{TP}{TP + FP} \qquad (4)$$

The precision for the given classifier is 0.71 [22], as shown in the table 1.

Table 1: Results of Naive Bayes Classifier in terms of precision

| News article type | Total number of news in test dataset | Number of correctly classified news | Classification accuracy |
|---|---|---|---|
| True | 881 | 666 | 75.59% |
| Fake | 46 | 33 | 71.73% |
| Total | 927 | 699 | 75.40% |
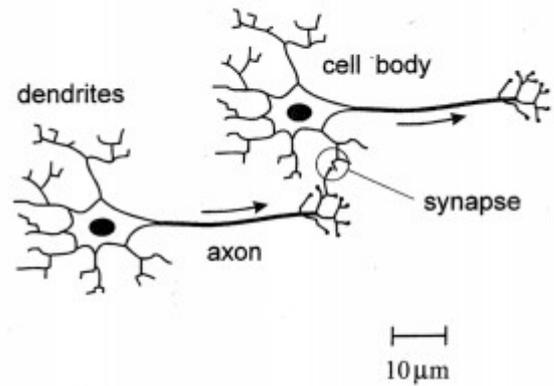


Fig. 4: Structure of a human neuron

## III. CSI

### A. Neural Network

The detection of fake news have been implemented also through the usage of Neural Networks (NN), i.e. a set of algorithms that are used to solve several different problems. [39].

Neural networks derive from studies on the information processing mechanism of the biological nervous system. A human neuron, as shown in the figure 4, is a special biological cell that processes information; it is composed of a cell body, or *soma*, and two types of out-reaching treelike branches: the *axon* and the *dendrite*. A neuron receives signals (impulses) from other neurons through its dendrites (receivers) and transmits signals generated by its cell body along the axon (transmitter) [5].

The simplest neural network is the single neuron [32]. It can be seen as a non-linear function which transforms a set of input variables $x_i$ into an output variable $z$. The signal $x_i$ at input $i$ is multiplied by a weight $w_i$, and it is added to all other weighted input signals to give a total input to the unit of the form [43]

$$a = \sum_{i=1}^{d} w_i x_i + w_0 \qquad (5)$$
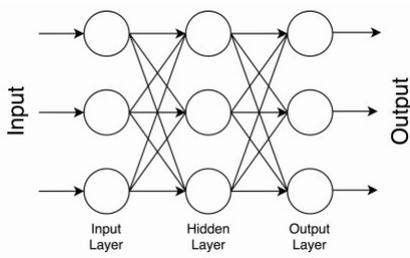
Fig. 5: Structure of a FeedForward Network



Fig. 6: Structure of a RNN

where the parameter $w_0$ is called *bias*. The output $z$ is given by operating on $a$ with a non-linear *activation function g*; so,

$$z = g(a). \tag{6}$$

The following functions are some activation functions:

- *Hyperbolic tangent*

$$tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \tag{7}$$

- *Sigmoid*, define by:

$$y = \frac{1}{1 + e^{-x}} \tag{8}$$

The connections of individual neurons to each other are done to build more complex neural networks. Neurons are organized in layer, such that the output of a given layer become the input of the next one. The most commonly used layers are the fully-connected layers, in which neurons have connections with all the other neurons. In order to solve more complex problems, in an optimal way, it is necessary that the last layer has the proper number of neurons.

There are two main types of neural networks [24]:

- **FeedForward Network** wherein there aren't cycles for the absence of connections between the nodes. It was the first and simplest type of artificial neuronal network devised. In this network, the information moves in only one direction, forward, from the input nodes, through the hidden layer, and to the output nodes. See figure 5.

- **Recurrent Network(RNN)**: where connections between nodes form a direct graph along a temporal sequence. This allows it to exhibit temporal dynamic behaviour. Unlike feed forward neural network, RNN can use their internal state (memory) to process sequences of inputs [23]. This makes them applicable to tasks such as detected fake news. See figure 6.

Within a neural network, the multiple connections between neurons take the form of a weighted oriented graph, where the nodes correspond to the individual neurons, the arcs to the weighted connections between them and the direction of the arcs to the direction of signal propagation.
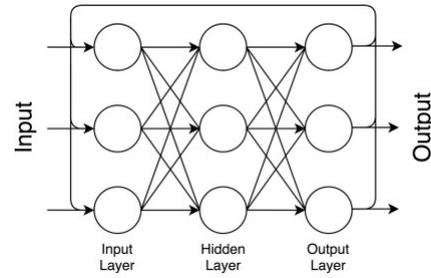
Since the purpose of the network is to minimize the error of predicting the output with respect to the expected value of the data, it is necessary to calculate this error and go backwards along the oriented graph to calibrate the weights based on how much they have affected the incorrect output. This technique is called BackPropagation. The principle behind error minimization is the gradient descent [40].

The technique of descending the gradient is based on the fact that the gradient of a function $f(\omega)$ indicates the direction of maximum growth (or decrease if considered in the opposite direction). Let $p = -\Delta f(\omega)$ denote the opposite of the gradient of the function $f$ calculated in $\omega$. To minimize the function, we start from a random $\omega$ and then proceed iteratively, calculating at each step $f(\omega)$ and updating $\omega = \omega - \Delta$, with $\Delta = \eta \cdot \Delta f(\omega)$, where $\eta$ represents a learning rate which defines at each iteration the length of the step that is used to move in the direction of decrease of the gradient.

The technique of descending the gradient, therefore, has two essential phases that iterate: the first, in which the output value of the function is calculated, the second in which the value of $\omega$ is calibrated in order to minimize the function. The algorithm ends if an output condition is verified or if the minimum is found, or if the function value does not decrease for N consecutive iterations.

In our case, $f(\omega)$ represents the cost function or loss, and is calculated as the average value of the errors computed for the set of inputs used. Since it represents the function to be minimized, its value during the gradient-descent execution tends to decrease.

### B. CSI-Caputure, Score, Integrate

Natali Ruchansky, Sungyong Seo, Yan Liu (University of Southern California) [39] developed CSI, a recent method to detect fake news, that uses artificial neuronal network and the Singular Value Decomposition (SVD) of matrix [21].

The following consideration is the base of the CSI method: there are three common features to fake news articles. In particular these features are: the text of an article, the response it receives and its source. CSI method uses all three feature to detect fake news. It is composed of three components:

- **Capture**, uses an RRN to obtain temporal repre-

sentation of articles;

- **Score**, assigns a score for each user that interacts with the article;

- **Integrate**, combines the information acquired from previous components to classify an article as true or false.

### C. Capture

In this component recurrent neural network is used to establish the number of users that engaged with an article $a_j$ and the engagements have been spaced over time.

The input of RNN is:

$$x_t = (\eta, \Delta t, x_u, x_\tau) \qquad (9)$$

where, $\eta$ is the number of engagements over time $\Delta t$. These parameters measure the frequency of the response an article received. $x_u$ represents the source of the article, and next the binary incident matrix is constructed to determinate the number of article that engage with a user, and the Singular Value Decomposition is applied to extract a lower-dimensional representation for each user $u_i$. $x_\tau$ is a vector which includes the text feature of an engagement with an article $a_j$.

The output is a vector $v_j = tanh(W_r h_T + b_r)$, where $X_r$ is a weight matrix, $h_T$ is the last hidden state and $b_r$ is the bias. This vector is a low dimensional representation of engagements for a given article $a_j$.

### D. Score

This component computes a score for each user. The following equation calculates it:

$$s_i = \sigma(w_s^T \cdot \tilde{y}_i + b_s) \qquad (10)$$

where $\sigma$ is the sigmoid function, $w_s$ is a weight vector, $b_s$ is the bias and $\tilde{y}_i$ is the vector representation of user and it is computed from

$$\tilde{y}_i = tanh(W_u y_i + b_u). \qquad (11)$$

### E. Integrate

To classify if an article is real or fake, the last component, Integrate, acquires the information of the previous components. In particular, it is necessary to introduce the vector $c_j$ obtained by concatenating vector $v_j$ to the vector obtained from vector $s$, by applying a mask $m_j$ to select only the components that have interacted with the article $a_j$.

The label $\tilde{L}_j$ is obtained as follows:

$$\tilde{L}_j = \sigma(w_c^T c_j + b_c) \qquad (12)$$

### F. Results

To show the results of CSI method, it has been tested on two dataset *Twitter* and *Weibo*, which are the only two free dataset that include all three characteristics: response, text and user information. In table 2, we observed some details. In particular, in Twitter the articles

are news stories, and in Weibo the articles are discussion topics [28].

Table 2: Details of the dataset for the CSI method

|  | Twitter | Weibo |
|---|---|---|
| *User* | 233719 | 2819338 |
| *Articles* | 992 | 4664 |
| *Engagements* | 592391 | 3752459 |
| *Fake articles* | 498 | 2313 |
| *True articles* | 494 | 2351 |

Table 3: Results in terms of accuracy and f-score of CSI method

|  | Twitter | | Weibo | |
|---|---|---|---|---|
|  | *Accuracy* | *F-score* | *Accuracy* | *F-score* |
| **DT-Rank** | 0.624 | 0.636 | 0.732 | 0.726 |
| **DTC** | 0.711 | 0.702 | 0.831 | 0.831 |
| **SVM-TS** | 0.767 | 0.773 | 0.857 | 0.861 |
| **LSTM-1** | 0.814 | 0.808 | 0.896 | 0.913 |
| **GRU-2** | 0.835 | 0.830 | 0.910 | 0.914 |
| **CI** | 0.847 | 0.846 | 0.928 | 0.927 |
| **CI-t** | 0.854 | 0.848 | 0.939 | 0.940 |
| **CSI** | 0.892 | 0.894 | 0.953 | 0.954 |

In particular CI and CI-$t$ are a variant of CSI model: CI uses only the text feature while CI-$t$ uses the text and temporal features of the articles. CSI shows very performing results, compared to other methods, present in the literature, both in terms of accuracy and F-score, as shown in table 3.

## IV. FAKE NEWS DETECTION USING SUPERVISED ARTIFICIAL INTELLIGENCE ALGORITHMS

Today, given the large amount of data circulating on the social media, it is unthinkable to suppose to solve the fake news problem by means of timely human control. In fact, many mathematical models make use of artificial techniques and text mining method, as in [25, 33, 36–38, 41, 44, 47]. In particular we want to analyse the model proposed by Ozabary et al. in [33].

### A. Text mining

The first step of the model is to apply the text mining method to structured data from an unstructured news article. The procedure begins with data pre-processing which consists of three phases:

- *Tokenization*: In this phase, the text is divided into smaller parts, which are called **token** and it removes all the numbers and punctuations from the text. Moreover, in this phase, all words that contain less than N characters are deleted.

- *Stop-words removal*: In this phase, the stop-words are completely detected from the given text. In fact, stop words are words which, given their high

frequency in a language, are usually considered insignificant, such as articles, conjunctions, generic words or widespread verbs.

- *Stemming*: In this phase, different grammatical forms have been transformed into its root forms, in such a way as to obtain the basic forms of the words whose meanings are the same, but the word forms are different from each other.

The second step of the text mining method consists in extraction and selection of features. In this phase, at each term in the data set is assigned a weight and each document has been converted into a vector or term weights. To compute these weights, the Term Frequency (TF) method is used.

The TF method calculates how many times a word is reported within the text, through the following equation:

$$TF = \frac{\eta_{ij}}{|d_i|} \tag{13}$$

where $d_i$ is the sum of all terms in the $i$th document and $\eta_{ij}$ is the number of $j$th word in the $i$th document.

After calculating these weights, the Document-Term Matrix (DTM) is built. The DTM is a matrix of dimension $m \times n$, where the rows represent the documents and the columns represent the terms.

$$DTM = \begin{pmatrix} w_{11} & w_{12} & \dots & w_{1n} \\ w_{21} & w_{22} & \dots & w_{2n} \\ \vdots & & & \vdots \\ w_{m1} & w_{m2} & \dots & w_{mn} \end{pmatrix} \tag{14}$$

### B. Supervised artificial intelligence algorithms

Different supervised artificial intelligence algorithms are applied to the structured news data sets. In particular, the following algorithms are reported.

- *Stochastic gradient descent (SGD)*: It is a variant of gradient descent technique that used an iterative method to optimize differential functions. The idea is to calculate the gradient no longer through the whole training set, but through a single sample randomly selected at each iteration [40].

- *Simple CART*: this method was developed by Leo Breiman in the 1980s [6]. CART is abbreviated as Classification and Regression Tree algorithm. Simple CART is a classification technique that generates the binary decision tree. The CART algorithm provides a foundation for important algorithms like bagged decision trees, random forest and boosted decision trees. In this method, the data is separated into two subgroups that have a different outcome. This process is terminated when the subgroup size is minimum [27].

For more details, see the list in [33].

Figure 7 shows a summary of the described scheme
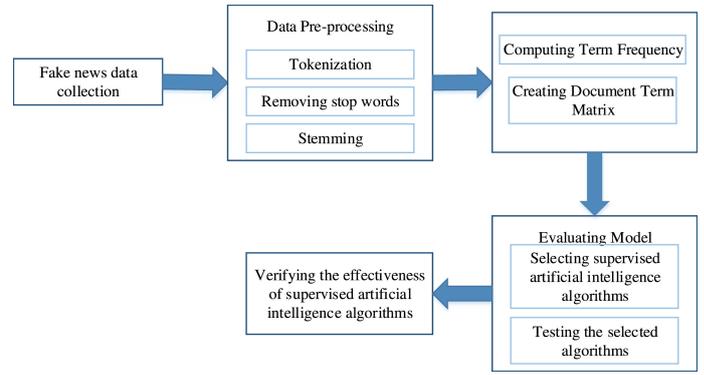


Fig. 7: Scheme of the model

### C. Results

This model has been tested on some available datasets, in particular in this paper we show the results obtained on BuzzFeed Political News data set, as mentioned in section II. In particular the dataset contains 1627 news Facebook articles about the 2016 American election, labelled as "Fake" or "Real". Some selected features from the BuzzFeed Political News dataset are: political, presidential, people, America, nation. All features are available in [33]. All these keywords have identified to label fake news in the BuzzFeed dataset. The news confirmed by reputed news platforms has been identified as real news.

Table 4: The performance of different supervised artificial intelligence algorithms for the BuzzFeed data set

|  | Accuracy | Precision | Recall | F-measure |
|---|---|---|---|---|
| BayesNet | 0.620 | 0.640 | 0.582 | 0.610 |
| JRip | 0.589 | 0.592 | 0.626 | 0.609 |
| OneR | 0.507 | 0.514 | 0.639 | 0.569 |
| Decision Stump | 0.532 | **0.747** | 0.534 | 0.534 |
| ZeroR | 0.510 | 0.509 | **1.000** | **0.675** |
| SGD | 0.605 | 0.619 | 0.590 | 0.604 |
| CVPS | 0.509 | 0.509 | **1.000** | **0.675** |
| RFC | 0.604 | 0.621 | 0.574 | 0.604 |
| LMT | 0.619 | 0.627 | 0.621 | 0.627 |
| LWL | 0.558 | 0.642 | 0.558 | 0.490 |
| CvC | 0.501 | 0.507 | 0.777 | 0.613 |
| WIHW | 0.509 | 0.509 | **1.000** | **0.675** |
| Ridor | 0.562 | 0.567 | 0.592 | 0.579 |
| MLP | 0.638 | 0.640 | 0.639 | 0.639 |
| OLM | 0.538 | 0.573 | 0.573 | 0.489 |
| SimpleCart | 0.646 | 0.654 | 0.649 | 0.652 |
| ASC | 0.563 | 0.616 | 0.563 | 0.516 |
| J48 | **0.655** | 0.655 | 0.681 | 0.668 |
| SMO | 0.619 | 0.629 | 0.616 | 0.622 |
| Bagging | 0.653 | 0.666 | 0.642 | 0.653 |
| Decision Tree | 0.634 | 0.626 | 0.707 | 0.664 |
| IBk | 0.513 | 0.441 | 0.480 | 0.460 |
| KLR | 0.521 | 0.481 | 0.583 | 0.527 |

To compare the performances of different supervised artificial intelligence algorithms, it is necessary to introduce the following quantities:

- True Positive (TP): If the method correctly classifies the fake news;

- False Positive (FP): If the method does not correctly classifies the fake news;

- True Negative (TN): If the method correctly classifies the real news;

- False Nagative (FN): If the method does not correctly classifies the real news.

The following measures are considered to evaluate the efficiency of the different algorithms:

$$Accuracy = \frac{|TN| + |TP|}{|FN| + |FP| + |TN| + |TP|}; \quad (15)$$

$$Precision = \frac{|TP|}{|FP| + |TP|}; \quad (16)$$

$$Recall = \frac{|TP|}{|FN| + |TP|}; \quad (17)$$

$$F - measure = 2 \times \frac{Recall \times Precision}{Recall + Precision}. \quad (18)$$

Table 4 shows the results, in terms of accuracy, precision, recall and F-score. In particular, it is observed that, in terms on $Accuracy$, the $J48$ algorithm is the best compared to the others. In terms of $Precision$, the highest value has been reached by $DecisionStump$. $ZeroR$, $CVPS$, $WIHV$ are the best in terms of $Recall$ and $F - measure$.
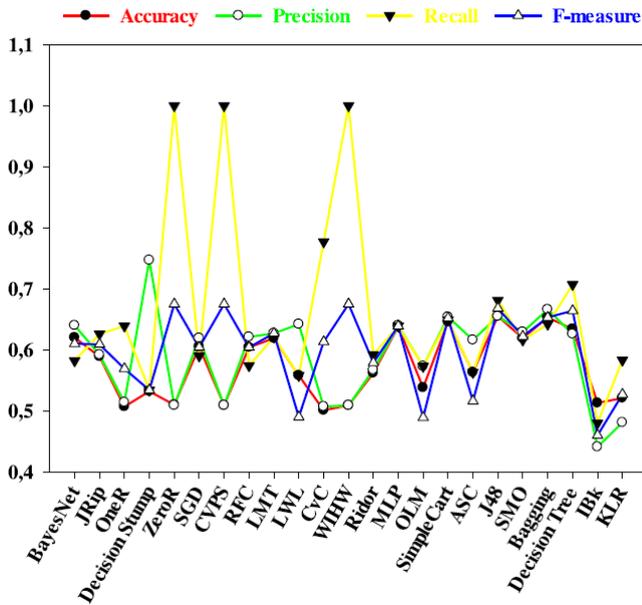


Fig. 8: Performances of supervised artificial intelligence algorithms [33]

Figure 8 shows the result obtained comparing the supervised artificial intelligence algorithms in terms os $accuracy, precion, recall$ and $F - measure$.

## V. Fake news detection using N-gram models

In this section we analyse the model proposed by Wynne at al. [47], who use the n-grams model to detect fake news. The method starts with a preprocessing phase, in which punctuation and all non-significant

words of the language are eliminated from the text, to reduce the dimension of the text. Next, N-grams features are extracted and a feature matrix is formed to represent the content. After that, TF and TFI techniques are used to extract features and calculate the feature values for the corresponding to all words in all contents in the training set. The last step is training the model.

### A. N-grams

An $N$-gram a sequence of numbers of items such as words, numbers, digits, characters or phonemes. An n-gram of size 1 is called $unigram$; a 2-gram is called $bigram$ and a 3-gram is called $trigram$. In general $N$-gram indicates a collection of $N$ words.

For example, we consider the sentence "we need to read this book soon", we can extract the word n-gram as follow:

- Unigrams: we, need, to, read, this, book, soon;

- Bigrams: we need, to read, this book, book soon;

- Trigrams: we need to, need to read, read this book, this book soon;

- 5-grams: we need to read this, to read this book soon;

- 7-grams: we need to read this book soon;

### B. Preprocessing

The preprocessing phase is important to reduce the size of the data. In the first phase, all punctuation is eliminated from the text and subsequently all the letters of the document are converted to lower case. After that, stop words are eliminated, as they are insignificant in a language. Then, the title and the body of each article are tokenized based on lengths of $N$-grams.

### C. Feature Extraction Techniques

There are two techniques for extracting features from the text:

- $TF$ (Term Frequency): it was described in the previous section, the matrix built with this technique will be indicated with TF.

- $IDF$ (Inverse Document Frequency): in the TF technique, all words are considered with the same importance. In a text, some words can appear very often, but they do not matter. For this reason, it is necessary to calculate the weight of the words considering all the words in the corpus. To calculate the IDF score, the following equation is used:

$$IDF(i) = \log\left(\frac{\text{Total Documents}}{\text{Document with term } i}\right) \quad (19)$$

Words that rarely occur have a higher IDF score.

- $TF$-$IDF$ (Term Frequency-Inverted Document Frequency) combines both techniques to generate a

score for each word within the text. To calculate the TF-IDF score, the following equation is used:

$$(TF - IDF)(i, j) = TF(i, j) \times IDF(i) \qquad (20)$$

where *TF-IDF* is a score for a term $i$ in document $j$.

*TfidfVectorizer* function of Python *sklearn.feature_extraction* library is used to generate TF-IDF n-gram features.

### D. Results

In order to analyse the method, a Kaggle dataset was used, containing 4009 articles, of which 2137 fake news and 1872 true news. For classification, two algorithms are used: **Random Forest** and **Gradient Boosting**. The size of the $N$-grams is between 1 and 4 and the number of features between 1000 and 10000. The results were derived in terms of accuracy, as in (15).

Table 5: Results, in terms of accuracy of model on dataset

| Feature extraction techniques | Classifiers | N-gram models | Lengths of n-gram | | | |
|---|---|---|---|---|---|---|
| | | | Uni-gram | Bi-gram | Tri-gram | Four-gram |
| TF (1000 features) | Random Forests | char | 0.92 | 0.92 | 0.89 | 0.89 |
| | | word | 0.91 | 0.90 | 0.91 | 0.89 |
| | Gradient Boosting | char | 0.92 | 0.94 | 0.93 | 0.92 |
| | | word | 0.91 | 0.92 | 0.91 | 0.89 |
| TF-IDF (1000 features) | Random Forests | char | 0.91 | 0.92 | 0.88 | 0.88 |
| | | word | 0.91 | 0.91 | 0.92 | 0.91 |
| | Gradient Boosting | char | 0.92 | 0.94 | 0.93 | 0.92 |
| | | word | 0.91 | 0.92 | 0.91 | 0.87 |
| TF (5000 features) | Random Forests | char | 0.91 | 0.92 | 0.92 | 0.91 |
| | | word | 0.91 | 0.92 | 0.92 | 0.91 |
| | Gradient Boosting | char | 0.92 | 0.93 | 0.93 | 0.94 |
| | | word | 0.94 | 0.92 | 0.91 | 0.87 |
| TF-IDF (5000 features) | Random Forests | char | 0.91 | 0.92 | 0.92 | 0.93 |
| | | word | 0.91 | 0.92 | 0.92 | 0.92 |
| | Gradient Boosting | char | 0.92 | 0.93 | 0.94 | 0.94 |
| | | word | 0.94 | 0.93 | 0.90 | 0.90 |
| TF (10000 features) | Random Forests | char | 0.92 | 0.93 | 0.93 | 0.93 |
| | | word | 0.92 | 0.94 | 0.94 | 0.92 |
| | Gradient Boosting | char | 0.93 | 0.94 | 0.94 | 0.95 |
| | | word | 0.94 | 0.94 | 0.93 | 0.88 |
| TF-IDF (10000 features) | Random Forests | char | 0.93 | 0.93 | 0.95 | 0.94 |
| | | word | 0.92 | 0.94 | 0.94 | 0.93 |
| | Gradient Boosting | char | 0.93 | 0.94 | **0.96** | **0.96** |
| | | word | 0.94 | 0.94 | 0.92 | 0.91 |

Table 5 shows the results of model. In particular, it is observed that the model achieves good results, with an accuracy of 96%, with both algorithms for classification, when using character tri-gram and character four-gram TF-IDF at 10,000 features. As expected, the worst result is obtained with an accuracy of 87%, with the Gradient Boosting algorithm, when using word four-gram.

## VI. CONCLUSION AND FUTURE WORK

The methods reported in this short review constitute only a small part of those present in the literature to detect and stop the fake news diffusion. In order to contrast fake news, two main approaches have been considered in literature: the first one tries to limit the diffusion of these news, e.g., by placing monitor software on some carefully chosen nodes [3], whereas the second one tries to limit the extent at which they may affect a public decision (e.g., by increasing the number of available alternatives [4]).

Developments in this research will concern new mathematical modelling to try to stop the fake news, but we will also try to anticipate the spread of fake news in advance, developing new numerical methods, based on differential equations [8, 10–17, 35], which are usually good for evolutionary problems.

Clearly, the detection of fake news is a relevant issue that can be handled through techniques of Artificial Intelligence, Neural Networks [24, 46], Metaheuristic Optimization Algorithms [34]. Future issues will concern the mathematical modelling of these aspects.

## REFERENCES

[1] Alba-Juez, L., Mackenzie J. Lachlan (2019). Emotion in Discourse, John Benjamins Publishing Company, Amsterdam, ISBN 9789027202390.

[2] Allcott Hunt, Matthew Gentzkow, Chuan Yu, 2019. Trends in the diffusion of misinformation on social media, "Research & Politics", vol 6(2).

[3] Amoruso, M., Anello, D., Auletta, V., Ferraioli, D., Contrasting the Spread of Misinformation in Online Social Networks, AAMAS 17, 1323–1331 (2017).

[4] Auletta, V., Ferraioli, D., Fionda, V., Greco, G., Maximizing the Spread of an Opinion when Tertium Datur Est, AAMAS 19, 1207–1215 (2019).

[5] Bishop Chris M. Neural networks and their applications, Review of scientific instruments, 65(6): 1803–1832, 1994.

[6] Breiman, L., Friedman, J., Stone, C. J., Olshen, R. A., Classification and regression trees. CRC press, 1984.

[7] Brody Dorje C., David M. Meier. How to model fake news, in arXiv:1809.00964.

[8] Butcher, J., D'Ambrosio, R. (2017). Partitioned general linear methods for separable Hamiltonian problems, Appl. Numer. Math. 117, 69-86.

[9] Cinelli, M., Quattrociocchi, W., Galeazzi, A., Valensise, C. M., Brugnoli, E., Schmidt, A. L., Zola, P., Zollo ,F., Scala, A., The covid-19 social media infodemic. arXiv preprint arXiv:2003.05004, 2020.

[10] Conte, C., D'Ambrosio, R., Jackiewicz, Z., Paternoster, B. (2013). Numerical search for algebraically stable two-step almost collocation methods, J. Comp. Appl. Math. 239(1), 304-321.

[11] Conte, C., D'Ambrosio, R., Paternoster, B. (2016). GPU acceleration of waveform relaxation methods for large differential systems, Numer. Algorithms 71(2), 293-310 (2016).

[12] D'Ambrosio, R., Paternoster, B., Numerical solution of a diffusion problem by exponentially fitted finite difference methods, Springer Plus 3(1), 425-431 (2014).

[13] D'Ambrosio, R., Moccaldi, M., Paternoster, B. (2018). Parameter estimation in IMEX-trigonometrically fitted methods for the numerical solution of reaction-diffusion problems, Comput. Phys. Commun. 226, 55-66.

[14] D'Ambrosio, R., Moccaldi, M., Paternoster, B. (2018). Numerical preservation of long-term dynamics by stochastic two-step methods, Discr. Cont. Dyn. Sys. - B 23(7), 2763-2773.

[15] D'Ambrosio, R., Paternoster, B. (2014). Exponentially fitted singly diagonally implicit Runge-Kutta methods, J. Comp. Appl. Math. 263, 277-287.

[16] D'Ambrosio, R., Paternoster, B. (2016). Numerical solution of reaction-diffusion systems of lambda-omega type by trigonometrically fitted methods, J. Comput. Appl. Math. 294 C, 436-445.

[17] D'Ambrosio, R., Paternoster, B. (2019). Multivalue collocation methods free from order reduction, J. Comput. Appl. Math., doi: 10.1016/j.cam.2019.112515.

[18] European Association for Viewers Interests EAVI (2017). Beyond Fake News - 10 Types of Misleading News. https://eavi.eu/beyond-fake-news-10-types-misleading-info/.

[19] Domingos P., Pazzini M., On the Optimality of the Simple Bayesian Classifier under Zero-One Loss, Machine Learning 29, 103–130 (1997);

[20] Giordano G. (2019). Metodi numerici per il rilevamento di fake news, Master thesis, University of Salerno.

[21] Golub Gene H, Van Loan Charles F, Matrix computations, Fouth Edition, The Johns Hopkins University Press, 2013.

[22] Granik, M., Volodymyr, M. Fake news detection using Naive Bayes Classifier, IEEE First Ukraine Conference on Electrical and Computer Engineering, pages 900–903, 2017.

[23] Hüsken Michael, Peter Stagge, Recurrent neural network for time series classification, Neurocomputing, 50: 223–235, 2003.

[24] Jain Anil K, Jianchang Mao, K Moidin Mohiuddin, Artificial neural networks: A tutorial, Computer, 20(3):31–44, 1996.

[25] Kaliyar, R.K., Goswami, A., Narang, P., Sinha, S., FNDNet – A deep convolutional neural network for fake news detection, Cognitive Systems Research, 61, 32-44, 2020.

[26] Lazer, D. et al. (2018). The science of fake news. "Science", 359(6380), 1094-1096.

[27] Loh, W. Y., Classification and regression tree methods. Wiley StatsRef: Statistics Reference Online, 2014.

[28] Ma J., Gao W., Mitra P., Kwon S., Jansen B.J., Wong K.F., Cha M., Detecting Rumors from Microblogs with Recurrent Neural Networks, Twenty-

[29] McGonagle, T. (2017). "Fake News": False fears or real concerns?, "Netherlands Quarterly of Human Rights", 35(4), 203-209. DOI: 10.1177/0924051917738685.

[30] Mottola S., Fake news as social phenomena. Linguistic analysis and persuasion in hoax in Italian and Spanish, to appear in Discurso & sociedad, Barcelona.

[31] Nicole R., "Title of paper with only first word capitalized," J. Name Stand. Abbrev., in press.

[32] Nielsen Michael A, Neural networks and deep learning, volume 25, Determination press San Francisco, CA, USA, 2015.

[33] Ozbay, F.A., Alatas, B., Fake news detection within online social media using supervised artificial intelligence algorithms, Physica A: Statistical Mechanics and its Applications, 540, 2020.

[34] Ozbay, F.A., Alatas, B., A Novel Approach for Detection of Fake News on Social Media Using Meta-heuristic Optimization Algorithms, ELEKTRONIKA IR ELEKTROTECHNIKA, 25, 62–67, 2019.

[35] Paternoster, B. (2002). Two step Runge-Kutta-Nyström methods for $y'' = f(x, y)$ and P-stability, Lecture Notes in Computer Science 2331, 459-466.

[36] Prabha, T.A., Aisuwariya, T., Kiran, M.V.K., Vasudevan, S.K., An innovative and implementable approach for online fake news detection through machine learning, Journal of Computational and Theoretical Nanoscience, 17, 130–135, 2020.

[37] Qawasmeh, E., Tawalbeh, M., Abdullah, M., Automatic Identification of Fake News Using Deep Learning, 2019 6th International Conference on Social Networks Analysis, Management and Security 2019, 383–388, 2019.

[38] Reddy, H., Raj, N., Gala, M., Basava, A., Text-mining-based Fake News Detection Using Ensemble Methods, International Journal of Automation and Computing, 2020.

[39] Ruchansky Natali, Sungyong Seo, Yan Liu, CSI: A hybrid deep model for fake news detection, in Proceedings of the 2017 ACM on Conference on Information and Knowledge Management, pages 797–806, ACM, 2017.

[40] Ruder S., An overview of gradient descent optimization algorithms, 2016, arXiv:1609.04747.

[41] Journal of Computational and Theoretical Nanoscience, Fagon: Fake news detection model using grammatical transformation on deep neural network, KSII Transactions on Internet and Information Systems, 13(10), 4958-4970, 2019.

[42] Silverman, S.S., Singer-Vine H., Hyperpartisan facebook pages are publishing false and misleading information at an alarming rate [Online].

[43] Simon S Haykin et al., Neural networks and learning machines, New York: Prentice Hall, 2009.

[44] Verma, A., Mittal, V., Dawn, S., FIND: Fake Information and News Detections using Deep Learn-

ing, 2019 12th International Conference on Contemporary Computing, IC3 2019.

[45] Vosoughi, S., Roy, D., Aral, S. (2018). "The spread of true and false news online", en Science, n. 359(6380), pp. 1146-1151. doi:10.1126/science.aap9559.

[46] Jang, S.Mo, Geng, T., Quennie Li, J.Y., Xia, R., Huang., C.T., Kim, H. Tang, J., A computational approach for examining the roots and spreading patterns of fake news: Evolution tree analysis,Computers in Human Behavior, 84, 103–113, 2018.

[47] Wynne H.E., Wint Z.Z., Content Based Fake News Detection Using N N-Gram Models, Association for Computing Machinery, 669–673, 2019.