Evangelia Grigoria ZVE

Master's degree (2nd year) Statistics and Econometrics (Distance Learning)

Academic Year: 2018/2019

MASTER THESIS

"Fake News Detection using social context and textual data"

Master's Thesis Supervisor : Yoann Pitarch

Associate Professor (Maître de Conférences)

UNIVERSITE TOULOUSE 1 CAPITOLE

2 rue du Doyen-Gabriel-Marty - 31042 Toulouse cedex 9 - France - Tél. : 05 61 63 35 00 - Fax : 05 61 63 37 98 - www.ut-capitole.fr

ACKNOWLEDGEMENTS

I would like to thank my supervisor Mr. Yoann Pitarch, for his guidance throughout the research project. His feedback and support, whether it was good or critical, helped me to structure and improve this thesis and to continue the project at tough points.

My gratitude also goes to my professors for their collaboration, support and understanding all over these years that I had to accomplish the difficult task to work in parallel of my studies. Next to our projects, I am grateful for their feedback, discussions, encouragement.

Finally, I would like to express my gratitude to the people dearest to me. I would like to thank my parents and my sister for their love, support. Thank you for being part of this journey, even when being at a distance.

ABSTRACT

Social Media, due to their easiness of use and accessibility, are becoming a more and more important source of information for the big public over the years. Their contribution to the convenience of the information and freedom of speech is precious for society. They nonetheless enable the broad propagation of Fake News, i.e., news with intentionally false information. Detecting Fake News is an important and challenging mission. Not only ensuring that users receive accurate information, but also trying to maintain a reliable news ecosystem. The most commonly used detection algorithms concentrate on clues extraction from news contents. This information is not enough as Fake News is very often intentionally written to mislead users by imitating true stories. Therefore, we need to explore additional information to improve detection. According to researches, considering the news dissemination process on Social Media and extracting valuable information from that, can contribute to enhancing the quality of the Fake News detection. This combination of data sources is likely the key to improve Fake News detection algorithms, and thus guarantee the reliability of the information disseminated on Social Media.

TABLE OF CONTENTS

1. Introduction

The objective of my Master thesis is to give an overview of the Fake News detection landscape, more specifically in Social Media and give a concrete example of an algorithm that detects Fake News based on the combination of two different types of information (Textual & Social Context). Looking across different research publications, papers, thesis, we can understand the importance to combine different data types in order to be able to efficiently and accurately predict Fake News on Social Media. I will be concentrating on specific methodologies and hypotheses intending to be as concrete as possible. I will not focus and cover every aspect of the Fake News detection but only the ones that in my opinion have the most critical impact.

Moreover, the different hypotheses I can define depend on the information I can extract from the dataset I possess. Therefore, the data restrict the options for feature extraction.

In the first part, you will find some introduction elements about the Fake News, their definition, use, and types. In addition to that, I will present you with some examples, Fake News used to provide political propaganda and massive manipulation. I will use the example of the USA presidential elections. Another important part of my thesis will be the presentation of the legal framework and the obligation of Social Media to put in place measures regarding this problem.

In the next part, I will present to you the different research approaches and methodologies used in our days to detect Fake News on Social Media with the help of algorithms.

In the third and most significant part, I will be concentrating on specific approaches, methodologies, and define different hypotheses for significant features. Those hypotheses will be evaluated along with the model construction and selection using statistical tests and evaluation metrics. A final model with the selected features will be the result of this part. In order to achieve the previous, I will develop an algorithm in Python 3. The algorithm development requires preliminary work on data preparation and feature extraction. To conclude, I will compare two model types (Naïve Bayes classifier and Random forests) but I will not go in more in-depth mathematical specifications given that my thesis focuses on feature extraction. As a conclusion, I will state the advantages to combine both social/network context and textual data for better and more accurate Fake News predictions in Social Media.

1.1 Definition of Fake News

Before defining what is Fake News and being able to intuitively recognize them, it is essential to have a clear conception of what is truth[1] and who decides what is true. In Wikipedia we find the following definition "*Truth is most often used to mean being in accord with fact or reality, or fidelity to an original or standard. Truth is also sometimes defined in modern contexts as an idea of "truth to self", or authenticity*"[2].

In the case of Fake News disseminated on Social Media, the questions that are raising are less abstract, especially when it comes to political propaganda. We can be in a situation where we can objectively judge if an article is Fake News or not (e.g., bots and fake profiles identification, expert's analysis).

A general definition could be that Fake News or information, Fake information, or Fake News is false information that is delivered to manipulate or mislead an audience.

These last years, the phenomenon of Fake News has become very present on the web to the detriment of internet users and caused massive manipulation notably related to political matters.

This false information propagates for different purposes. Some are intended to mislead the reader or to influence his opinion on a subject. Others are from scratch, so they are entirely invented with an attractive and catchy title to boost traffic and increase the number of visitors to a website to monetize this activity.

"Fake News" becomes a buzzword, notably after the U.S. presidential election in 2016. It is a practice characterized by a lot of misinformation and Fake News spreads. Mainstream media have widely reported on Fake News, and political organisations around the world have discussed and put in place solutions to control the situation. We will be discussing the legal framework of Fake News lately in this chapter.

Below there are stated a couple of definitions concerning the "Fake News" term by institutions, renowned media, and websites around the web.

The European Commission uses the term "Disinformation" to designate the Fake News neologism:

> Disinformation is understood as verifiably false or misleading information that is created, presented and disseminated for economic gain or to intentionally deceive the public, and may cause public harm. Public harm includes threats to democratic processes as well as to public

> goods such as Union citizens' health, environment or security. Disinformation does not include inadvertent errors, satire and parody, or clearly identified partisan news and commentary. The actions contained in this Action Plan only target disinformation content that is legal under Union or national law. They are without prejudice to the laws of the Union or of any of the Member States that may be applicable, including rules on illegal content.[3]

According to Webopedia:

> Fake News, or hoax news, refers to false information or propaganda published under the guise of being authentic news. Fake News websites and channels push their Fake News content in an attempt to mislead consumers of the content and spread misinformation via social networks and word-of-mouth.[4]

Another definition comes from an article found on PolitiFact website: "*Fake News is made-up stuff, masterfully manipulated to look like credible journalistic reports that are easily spread online to large audiences willing to believe the fictions and spread the word"* [5].

Wikipedia gives its definition "*Fake News is a neologism often used to refer to fabricated news. This type of news, found in traditional news, Social Media or Fake News websites, has no basis in fact, but is presented as being factually accurate"* [6].

In Cambridge dictionary "Fake News" is defined as "*false stories that appear to be news, spread on the internet or using other media, usually created to influence political views or as a joke: There is concern about the power of Fake News to affect election results".*

To conclude, a narrower definition of Fake News involves only intentionally and verifiably Fake News articles. Satire, hoax, rumors, conspiracy theories, and unintentionally generated misinformation are not included in this definition of Fake News.

A more general word used for the detection of Fake News and misinformation is called detection of deception.

For automatic Fake News detection analysis, keeping the Fake News definition as narrow as possible has more benefits to maintain the focus down to the key components and eliminate ambiguity. Therefore, in my master thesis I will be assuming this definition "*Fake News articles are intentionally fabricated to be deceptive and can be proven that they are false."*

Nevertheless, we should also take into consideration that the questions are complex, and it is crucial to make the distinction between Fake News and the expression of ideas and opinions. Also, other tricky elements should be taken into consideration. For example, the satire and humorous news articles are sometimes categorized as Fake News by algorithms because they do not contain real events and can mislead readers. It is also challenging to detect humour in Real News. Therefore, it can be a false cue/signal to detect Fake News automatically.

Moreover, a news article may sometimes contain partially false aspects. These elements do not make the entire news false, but they can trigger deception. Those fake components are usually more difficult to identify than identifying the entire Fake News article, and they make it difficult to classify the article as fake or real by an algorithm.[7]

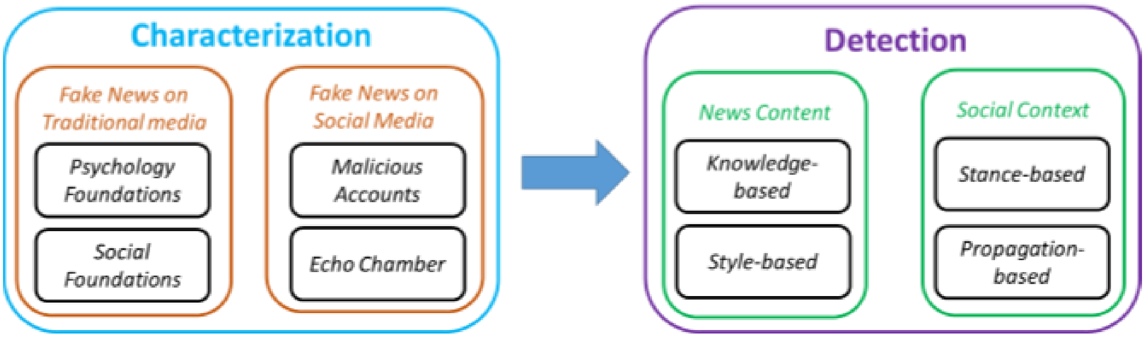

Figure 1: Fake News on Social Media: from characterization to detection
Source: Shu, Kai, Amy Sliva, Suhang Wang, Jiliang Tang and Huan Liu, "Fake News Detection on Social Media: A Data Mining Perspective," Figure 1.

In the next section, there is a presentation of the different types of Fake News, not with the narrow perspective of Fake News.

1.2 Categories of Fake News

The commonly called “Fake News” brings together several types of erroneous information, published with different intentions. They could be split into two main categories. Those that are under the “misinformation”, when people share content without knowing it is wrong, because they trust the author, or they did not realize they were on a satirical site. Also, those who are in the “disinformation” category, where the internet users share articles while knowing very well that the news is fake, but they serve their cause or their ideal. The first case is about an error and it is innocent. In the second case, the news is a lie, and it is intentional. It is a lie intended to propagate a political idea, support a candidate for an election or earn money by broadcasting viral content that creates an audience and therefore, advertising revenue.[8]

Searching across several articles and researches around the web, I can identify six types of Fake News that are sub-categories of the above mentioned categories: (1) news satire, (2) news parody, (3) fabrication, (4) manipulation, (5) advertising, and (6) propaganda[9]:

- **News Satire**:

The most common use of Fake News in articles is satire, which refers to Fake News, which usually use humor or exaggeration to present news to the public.

- **News Parody:**

News Parody has many common characteristics with Satire. They both uses humor as there mean to attract the audience interest. They often imitate mainstream media on the way the present the news. The main element that makes parody different from satire is that parody doesn’t reply on real information to inject humor, so most of the time the context is invented contrary to satire.

- **Fabrication:**

Those are articles that have no factual basis but are published in a way that allows them to look like regular press articles to manipulate the public. Contrary to parody, there is no implicit understanding between the author and the reader that the element is fake. The intention of the author is most of the time, the opposite, so, to misinform the audience.

- **Manipulation:**

This type of Fake News refers to photo and image manipulation. The manipulation of images became very common in the digital era and represents a significant part of the Fake News spreads. Digital photos, dynamic image manipulation, and knowledge of advanced techniques can make completely fake images look like real ones. Some examples of photo manipulation are the color saturation, removing or inserting a person or a minor element from the photo.

- **Advertising:**

Advertising is Fake News when malicious content is created to generate revenue through increased web traffic. Native Advertising is a specific case of online Advertising, which becomes more and more popular. Those advertisements are not easy to recognize, because they adapt to the content, inserts paid messaging right into the mix alongside news articles. Therefore, it hard to recognize which online content is news and which is a paid promotion. Another case of Advertising Fake News could be the clickbait news articles which use appalling or appealing headlines to attract attention and generate clicks on the website of the editor, usually at the expense of the truth or accuracy.

- **Propaganda:**

Propaganda refers to news created by a political entity with the purpose to influence the public opinion. The explicit aim is for a public person, organization, or government to profit. It is interesting to note that there is a gray zone between advertising and propaganda as there may share similar motives.

Claire Wardle of First Draft News[10] explains briefly seven types of Fake News that she distinguished due to her research:

- satire or parody ("no intention to cause harm but has potential to fool")
- false connection ("when headlines, visuals or captions don't support the content")
- misleading content ("misleading use of information to frame an issue or an individual")
- false context ("when genuine content is shared with false contextual information")
- impostor content ("when genuine sources are impersonated" with false, made-up sources)
- manipulated content ("when genuine information or imagery is manipulated to deceive", as with a "doctored" photo)
- fabricated content ("new content is 100% false, designed to deceive and do harm")[11]

First Draft News is a project created in 2015 by nine organisations brought together by the Google News Lab to tackle Misinformation and Disinformation online. It involves Facebook, Twitter, Open Society foundations, and various philanthropies[12]

The seven types of Misinformation, according to M. Mitchell Waldrop from the National Academy of Sciences of the United States of America.

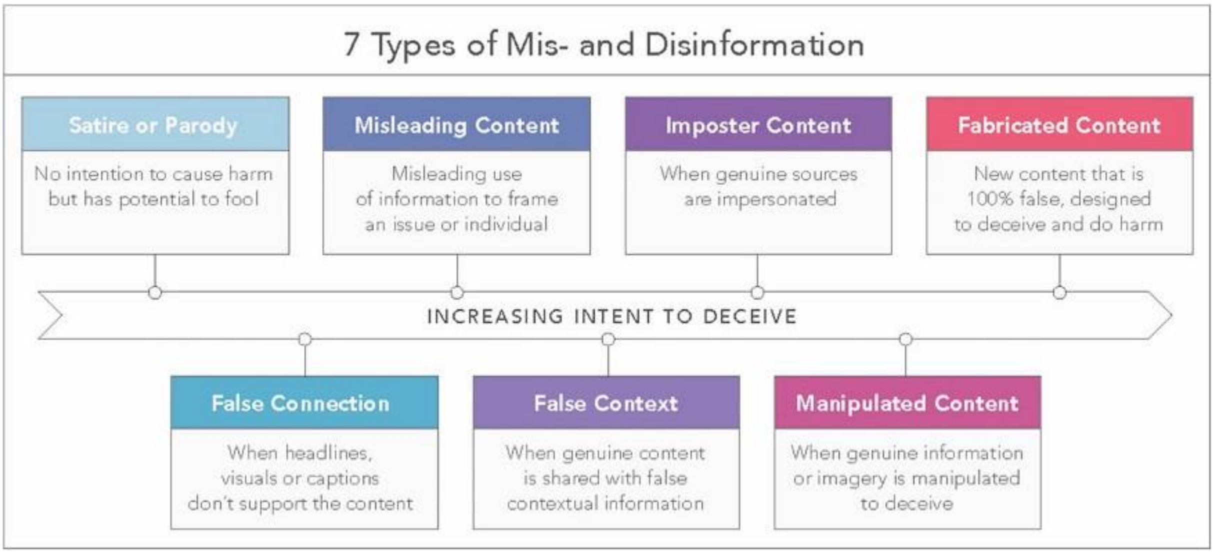

Figure 2: 7 Types of Mis- and Disinformation

Source: Waldrop, M. Mitchell. "News Feature: The genuine problem of Fake News." *Proceedings of the National Academy of Sciences of the United States of America 114 48* (2017): 12631-12634, Figure 2.

Among the above types of Fake News, I will be focusing on the identification of Fake News that are intentional lies. In Chapter 1.1 I gave the definition of Fake News that it will be considered for my analysis.

1.3 The diffusion of Fake News

Fake News can propagate through various media, but Social Media has the most significant proportion of Fake News spreads. Social Media's powerful nature and the fact that each person can be a publisher, make Social Media ideal for the propagation of any information. In addition, we notice that false, misleading, and unilateral news in Social Media spread faster than many serious, Real News stories.[13] It is crucial to note that Social Media is widely used as a news source for multiple news web pages and has become a priority news source, notably for the younger generation.[14]

To better understand the Fake News propagation and the dimension that can take, social bots is another significant fact to discuss. A social bot relates to a Social Media account operated by a computer algorithm that generates content automatically and interacts on Social Media with people (or other bot users). Researchers[15] concluded that bots have a significant responsibility in spreading misinformation and Fake News on social networks. They are mostly active in the early phases of spreading, targeting influential users who are more likely to spread false information.

Furthermore, there are significant personal biases when it comes to the recognition of the news character. Individuals are more likely to classify news articles that they disagree with as Fake News.[16] Recent research[17], shows that humans can distinguish only 70% of the total number of Fake News. In another research[18], we can find that 75% of individuals categorized the Fake News as True News. These significant biases influence what people share on Social Media and the spread of Fake News.

In addition to the categorization presented in Chapter 1.2, social science scientists have explored Fake News from different angles and provided in recent articles a general categorization of the different types of Fake News[19]. That categorisation of Fake News is the base of different methodologies and techniques used around the web for Fake News detection.

Below is a short description of the different categories[20]:

- **Visual-based:**

Fake News uses visual representation objects, such as images modified with photoshop, videos, or a mixture of both.

- **User-based:**

Fake News is related to fake accounts, and they target specific audiences grouped by age, gender, culture, political orientation.

- **Post-based:**

Fake News focuses primarily on appearing on Social Media platforms. For example, a Facebook post with a fake picture or video, a tweet, a meme.

- **Network-based:**

Fake News is targeted at members of a specific organisation. The members are related in one way or the other. This practice could concern a group of friends on Facebook or a group of people linked to each other on LinkedIn.

- **Knowledge-based:**

Fake News that pretend to give scientific or reasonable explanations for unsolved problems, such as, Fake News article that gives the solution or medicine to cure a specific type of cancer.

- **Style-based:**

Fake News focuses on the way an article is written and presented to its public. In this case, the authors are, most of the time, people who are not journalists, and thus, the different style of writing may help to recognize Fake News.

- **Stance-based:**

Stance-based Fake News type is related to the style-based type that is described above. This type of Fake News focuses on how statements are presented. True News articles provide the reader with enough information about the topic/subject matter. It is on readers to understand and perceive the meaning of the story. Generally, Stance-based Fake News articles do not provide the reader with enough information about the subject matter and use arguments that are not true.

Below we can find an illustration of the Fake News life Cyrcle as explained by Xinyi Zhou, and Reza Zafarani[21]. They consider four different perspectives of Fake News that are very similar to the ones mentioned above. We can also say that summarises them.

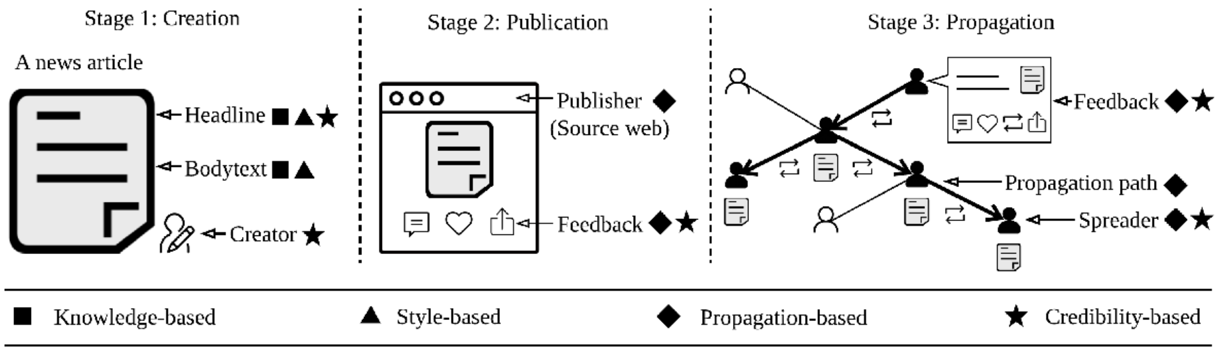

Figure 3: Fake News Life Cycle and Connections to the Four Fake News Perspectives
Source: Zhou, Xinyi and Reza Zafarani, "Fake News: A Survey of Research, Detection Methods, and Opportunities," Figure 1.

1.4 Socio-political context

Fake News have always existed in our history, they nonetheless gained attention in the aftermath of the unexpected outcomes of electoral contests in Western Europe and North America in 2016. In the same time, their influence is today more important because of the virality of their dissemination on the Social Media networks. Although conventional news media remain predominant world-wide, the media developed on the internet constitute one of the primary sources of information for the public and it becomes little by little the most important source of information. Among these media, there are digital versions of the print media, web sites dedicated to news or social networks (e.g., Facebook, Twitter, Instagram).

Social networks were not initially designed to become a news media. However, the ease of sharing information has made social networks attractive to the public who seeks for easily accessible and quick information. It also gives the freedom to the user sharing a news article that for him is important, and the freedom of the expressing his opinion openly and widely. Thus, social media becomes a medium for all people express their opinion about the news and Social networks are characterized by the viral dissemination of the information. Nevertheless, that liberty came without putting in place measures that verify the content published on those networks. Those specificities have made social networks a medium where it is relatively easy to share Fake News.

Another important information is that Social Media bots can become malicious entities designed for harmful purposes, such as manipulating the public by spreading Fake News.

Studies show that social bots have distorted online conversations on a big scale during the 2016 U.S. presidential election[22], and that about 19 million bot accounts tweeted in support of either Trump or Clinton during the week before the elections[23].

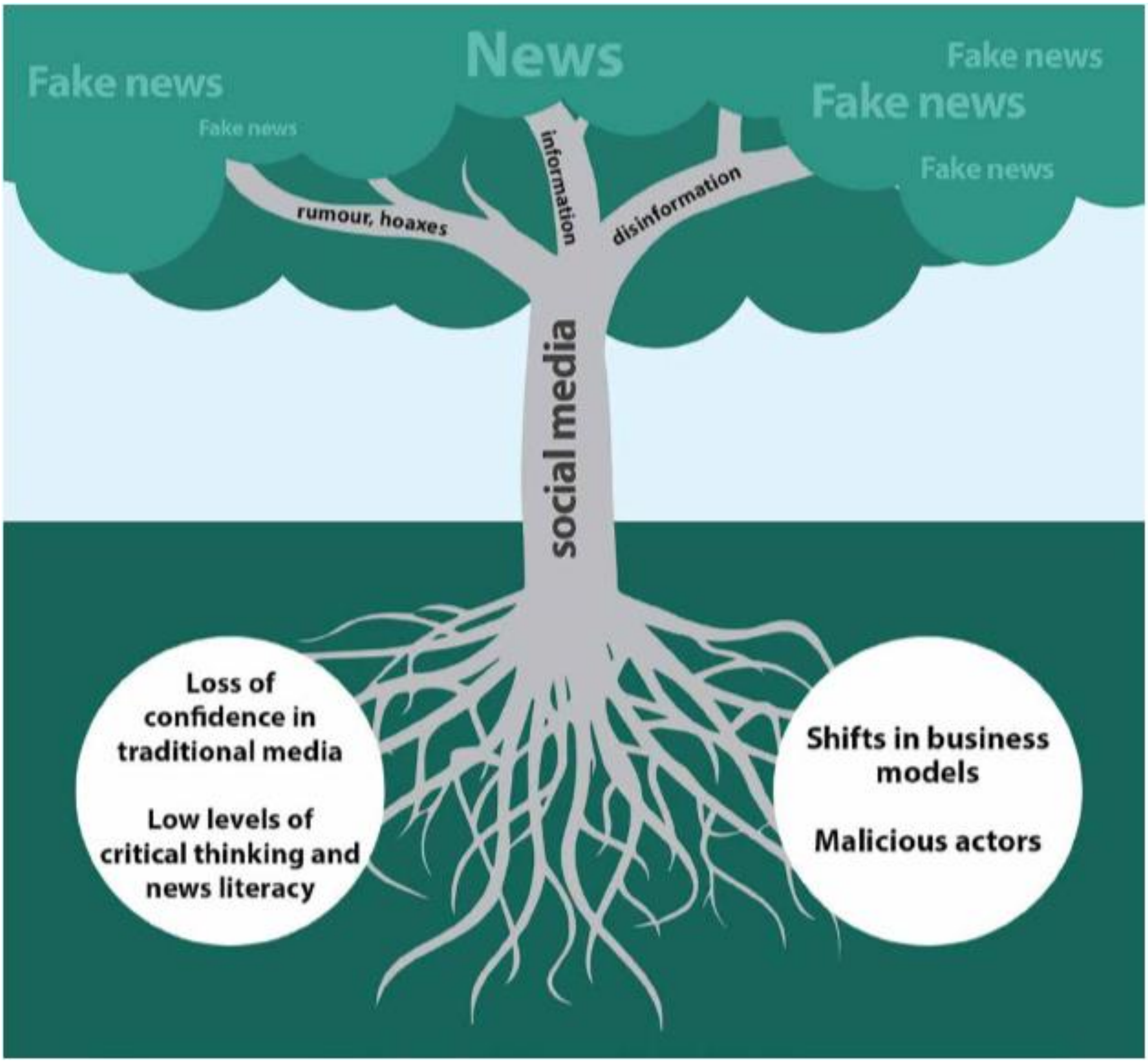

Figure 4: The roots of 'Fake News'

Source: Stremlau, Nicole, Iginio Gagliardone, and Monroe Price. "World Trends in Freedom of Expression and Media Development 2018." *UNESCO - World Trends in Freedom of Expression and Media Development* (2018), Figure 2-8.

1.4.1 Fake News cases

According to Hunt Allcott and Matthew Gentzkow (Journal of Economic Perspectives) [24], the impact of the "Fake News" in the U.S. 2016 presidential election, was very important. More specifically, recent researches and surveys show that:

- 62% of US adults get informed on Social Media [25],
- on Facebook, the most famous Fake News stories were shared more commonly than the most common news stories[26] ,
- many people who see Fake News stories report that they believe them[27]
- the most discussed Fake News stories tended to favor Donald Trump over Hillary Clinton[26].

"Fake News" term was used during the 2016 election campaign to describe Fake News stories from websites that post hoaxes, as well as, hyper-partisan websites that they are supposed to provide Real News. The last months before the election, some of those stories, such as, Fake News about Pope Francis supporting Trump and Hillary Clinton selling arms to ISIS, were widely spread on Facebook and could influence the decision of the voters. On that matter, several renowned journalists claimed that Donald Trump would not have been elected president if there was not that political propaganda using Fake News (see articles from Parkinson 2016[28], Read 2016[29], Dewey 2016[30] for examples).

In a more recent news case, Donald Trump seems to intent weaponizing the term, turning its meaning on its head for his political ends. He claimed that this word usage had provoked global distrust concerning the American press, especially in Russia. His claims gave legitimacy to the stories diffused in the Russian media, labeling American news, in particular the news of the crimes committed by the Syrian regime against its own individuals. More specifically, it was stated by Trump that "munitions at the air base had as much to do with chemical weapons as the test tube in the hands of Colin Powell[31] had to do with weapons of mass destruction in Iraq" to respond to Russian media[32] and gain the public trust.

That huge difference in the position that media from different countries take around the world can be an element that indicates a high probability of the Fake News nature of the news disseminated in one country or another.

It is worthy of mentioning that in May 2018, Donald Trump declared on Twitter his distrust against the accuracy of the news spread for him across the online Social Media:

"*The Fake News is working overtime. Just reported that, despite the tremendous success we are having with the economy & all things else, 91% of the Network News about me is negative (Fake). Why do we work so hard in working with the media when it is corrupt? Take away credentials?"* [33]

A more recent example of Fake News concerning Donald Trump raised at the beginning of August 2019. In an assault at a Walmart in El Paso, a gunman murdered 20 individuals and injured dozens more. Before the assault, he supposedly wrote a racist, anti-immigrant manifesto that outlined Latino "invasion" concerns and had "Fake News" references.[34]

President Trump said "Fake News" has contributed to provoking "anger and rage" in the country in the context of two mass shootings that happened at the beginning of May 2019, marking the second time he has criticized media coverage in the aftermath of a tragedy. More specifically, to respond to these events Trump tweeted,

"The Media has a big responsibility to life and safety in our Country. Fake News has contributed greatly to the anger and rage that has built up over many years. News coverage has got to start being fair, balanced and unbiased, or these terrible problems will only get worse!"[35]

Fake news is not only a matter in the United States but also in Europe. Below there are two short examples of Fake News that went viral on Social Media for political propaganda purposes.

- **Brexit:**

The leaders of the pro-Brexit campaign have deliberately fabricated a lie: "*Leaving the European Union, Britain could save 350 million per week and allocate them to social security*." False information calibrated to attract votes from the citizens and disseminated massively via the new digital media. [36]

- **Presidential elections in France:**

During the 2017 presidential campaign, according to Richard Ferrand, the campaign manager for Mr Macron's En March party, the candidate Emmanuel Macron and his team faced massive computer attacks and the publication of Fake News information against him on social media. The government claimed that the attacks have their sources in Russia and all these maneuvers are all attempts at foreign interference that could have affected the results of the elections. [37]Emmanuel Macron said, "Today, we have to look at the facts: two major media platforms Russia Today and Sputnik, that belong to the Russian state, broadcast false news daily. Then these news are recuperated and quoted and come and weight on our democratic life."[38]

1.4.2 The legal Framework of the Fake News

In this section, we will find some interesting introduction elements about the legal framework fighting against Fake News. It concerns geographical areas that are considered relevant to this topic due to numerous examples of Fake News dissemination. The objective is to give an overview of the Fake News legal context at the international level.

Measures have been taken by western governments and institutions to cope with this problem by introducing specific legislations and regulations. Notably in Europe, their objective is to put in place measures that force the internet platforms to use new technologies in order to enable automatic Fake News post removal on Social Media and deactivate individual accounts that spread Fake News (i.e., social bots).

Many organisations interpret this as an effort to suppress freedom of expression. The fine line is drawn differently depending on the country. Major Social Media networks like Facebook and Twitter in Europe can face penalties and harsher regulations. The previous can happen if they ignore or fail to remove "harmful and/or illegal" articles and accounts.

That being said, the legal basis for any policy measures to fight misinformation finds a fundamental restriction on the existence of the fundamental right of the freedom of the speech, which must get balanced with the right of the public to be duly informed.

Below I citate two European articles established to protect and preserve the freedom of the speech:

Article 11.1 of the Charter of Fundamental Rights of the European Union (2000/C 364/01) recognizes the freedom of expression and information:

> Everyone has the right to freedom of expression. This right shall include freedom to hold opinions and to receive and impart information and ideas without interference by public authority and regardless of frontiers. [39]

Article 10 of the European Convention on Human Rights (hereinafter, ECHR) says:

1. Everyone has the right to freedom of expression. This right shall include freedom to hold opinions and to receive and impart information and ideas without interference by public authority and regardless of frontiers. This Article shall not prevent States from requiring the licensing of broadcasting, television or cinema enterprises.

2. The exercise of these freedoms, since it carries with it duties and responsibilities, may be subject to such formalities, conditions, restrictions or penalties as are prescribed by law and are necessary in a democratic society, in the interests of national security, territorial integrity or public safety, for the prevention of disorder or crime, for the protection of health or morals, for the protection of the reputation or rights of others, for preventing the disclosure of information received in confidence, or for maintaining the authority and impartiality of the judiciary.[39]

The above Articles have been considered by European Union in the process of creating a legal framework that fight against misinformation, and Fake News by respecting the balance discussed above.

1.4.2.1 Actions in Europe

Always at the heart of the news, the Fakes News has been these last years, the subject of a new institutional initiative to fight against their spread.

The European Commission has worked to introduce a transparent, thorough and extensive set of measures to combat the spread and impact of online misinformation and Fake News in Europe and to guarantee that European and democratic values are protected.

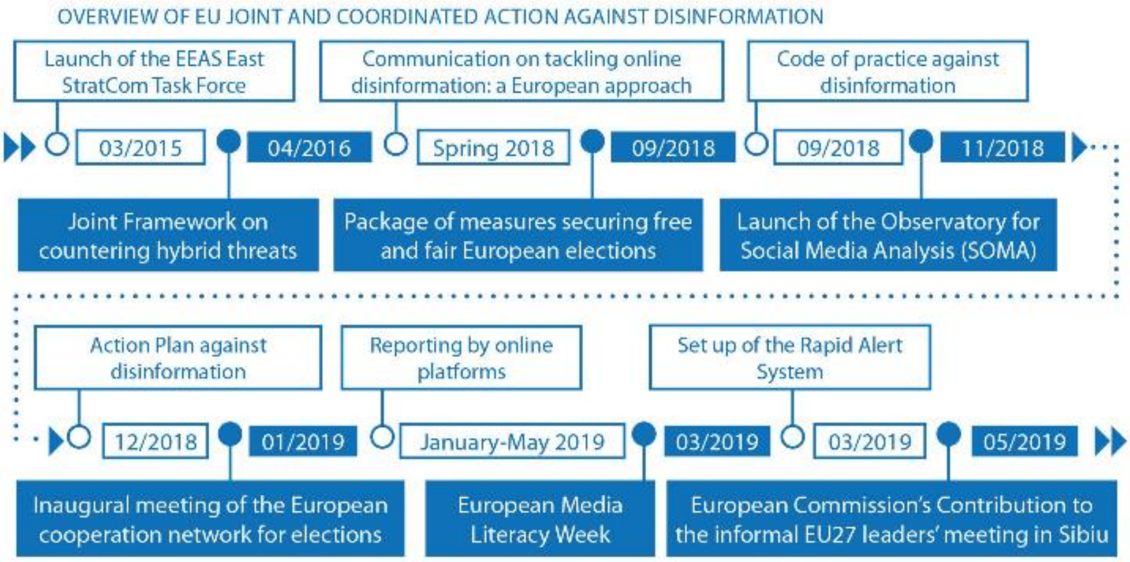

Figure 5: Overview of EU joint and coordinated action against disinformation.

Source: “Tackling Online Disinformation.” *Digital Single Market - European Commission. The Directorate-General for Communications Networks, Content and Technology*, June 17, 2019. https://ec.europa.eu/digital-single-market/en/tackling-online-disinformation.

The Action plan against Disinformation[3]

The European Union established an Action Plan to boost efforts for fighting disinformation, misinformation and Fake News in Europe. The plan focuses on four key points (cf. below list). This plan serves to build the EU's capabilities and strengthen cooperation between the Member States by

1. improving disinformation detection, analysis, and exposure
2. enhancing cooperation and joint reactions to threats
3. enhancing collaboration with internet platforms and business to combat disinformation
4. increasing awareness and social resilience

The measures were based on the April 2018 Communication on tackling internet disinformation[40], highlighting the role of civil society and the private industry in fighting disinformation spread. The attempts aim at avoiding misinformation and Fake News dissemination before the 2019 European elections.

Code of Practice on Disinformation

Representatives of online platforms, leading social networks, advertisers, and the advertising industry had already agreed on a code of practical self-regulation to tackle the spread of online misinformation and Fake News.

Digital Economy and Society Commissioner Mariya Gabriel considers this as a shift in the desired direction but invites the online platforms to reinforce their initiatives to combat the spread of online disinformation:

> The Code of Practice submitted today by the industry is the first tangible outcome of the Communication that the Commission adopted last April.
>
> It is an important step in tackling a problem which has become increasingly pervasive and threatens Europeans' trust in democratic processes and institutions. This is the first time that the industry has agreed on a set of self-regulatory standards to fight disinformation worldwide, on a voluntary basis. The industry is committing to a wide range of actions, from

> transparency in political advertising to the closure of fake accounts and demonetisation of purveyors of disinformation, and we welcome this. These actions should contribute to a fast and measurable reduction of online disinformation. To this end, the Commission will pay particular attention to its effective implementation.
>
> The Code of Practice should contribute to a transparent, fair and trustworthy online campaign ahead of the European elections in spring 2019, while fully respecting Europe's fundamental principles of freedom of expression, a free press and pluralism. This step complements the Commission's Recommendation presented by President Juncker in his 2018 State of the Union Address on election cooperation networks, online transparency, protection against cybersecurity incidents and fighting disinformation campaigns. Online platforms need to act as responsible social players especially in this crucial period ahead of elections. They must do their utmost to stop the spread of disinformation.
>
> I urge online platforms and the advertising industry to immediately start implementing the actions agreed in the Code of Practice to achieve significant progress and measurable results in the coming months. I also expect more and more online platforms, advertising companies and advertisers to adhere to the Code of Practice, and I encourage everyone to make their utmost to put their commitments into practice to fight disinformation.
>
> I will meet the signatories of the Code of Practice in the coming weeks to discuss the specific procedures and policies that they are adopting to make the Code a reality.
>
> As foreseen in the Communication, the Commission will closely follow the progress made and analyse the first results of the Code of Practice by the end of 2018. Should the results prove unsatisfactory, the Commission may propose further actions, including actions of a regulatory nature.[41]

Also, as specific list of best practices to fight against Fake News has been officially established.[42,43]

1.4.2.2 Actions in France

French law had repressed the publication of Fake News since a long time ago[44]. However, this regulation did not allow to fight against the spread of Fake News effectively. Other texts have recently been adopted to try to involve other actors than only Fake News spreaders.

For example, a decree put in place in 2017[45] forces digital advertising vendors to provide detailed information on the characteristics of their services, allowing notably advertisers to check the advertising context and to ensure that advertisements do not appear on websites and pages that contain Fake News content.

Moreover, the "Fake News law"[46] allows the sanctioning of the online platforms (whose activity exceeds a specific limit of connections in France) which, during an election period, fail to provide visitors with a certain amount of information on the content related to the debate of general interest.

This law enables judges during election campaigns to order the immediate removal of "Fake News."

The law also says that "information that is fair, clear and transparent" on how their data are exploited must be given to users.

The law was rejected by the Senate twice before it was adopted by the parliament formed by President Emmanuel Macron. On the one side, those who support the law claim that the objective is to combat against Fake News spread, not censor opinions. On the other hand, the French government has been accused by both right and left politicians of trying to create a form of "thought police".[47]

1.4.2.3 Actions in the US

The United States Constitution's first amendment protects the freedom of the expression and the press. It protects the right to exchange ideas and opinions openly, whether controversial or incorrect. Censorship, as well as a prior restraint, is generally unconstitutional, which is a public action that prohibits speech or other expressions before it can occur. That means that it is not possible to ban Fake News in the U.S.

Nevertheless, there are several legal remedies for people who are the topic of Fake News. In other words, they may take action for defamation or other speech-related torture (e.g., false light privacy invasion), intentional infliction of emotional distress[48] or tortious interference[49].

That being said, before the Fake News article is published, no legal action can be taken, and remedies can only be sought after the publication if the latter has caused damage to the individual.[50]

Facebook boss Mark Zuckerberg said on Wednesday, June 26, that "*a lack of action by US authorities on fake political content on the platform after the 2016 US election helped pave the way for a subsequent avalanche of online disinformation.*"[51]

<u>How Fake News can be identified on Social Media</u>

Thanks to machine learning algorithms, there are a couple of measures that can be taken to detect if a news article disseminated on a Social Media is True or Fake. For example, the URL of the page may revendicate malicious content or discrepancies with real media patterns or outlets.

In addition, there are several websites working on the manual and automatic identification of Fake News across the internet (e.g., Snopes.com, FactCheck.org, PolitiFact.com). Their role is crucial for the effectiveness of the Fake News detection as many information included in datasets used to train machine learning algorithms have their source from the manual fact-checking done by experts.

Table 1: A Comparison among Expert-based Fact-checking Websites

	Topics Covered	**Content Analyzed**	**Assessment Labels**
PolitiFact[3]	American politics	Statements	True; Mostly true; Half true; Mostly false; False; Pants on fire
The Washington Post Fact Checker[4]	American politics	Statements and claims	One pinocchio; Two pinocchio; Three pinocchio; Four pinocchio; The Geppetto checkmark; An upside-down Pinocchio; Verdict pending
FactCheck[5]	American politics	TV ads, debates, speeches, interviews and news	True; No evidence; False
Snopes[6]	Politics and other social and topical issues	News articles and videos	True; Mostly true; Mixture; Mostly false; False; Unproven; Outdated; Miscaptioned; Correct attribution; Misattributed; Scam; Legend
TruthOrFiction[7]	Politics, religion, nature, aviation, food, medical, etc.	Email rumors	Truth; Fiction; etc.
FullFact[8]	Economy, health, education, crime, immigration, law	Articles	Ambiguity (no clear labels)
HoaxSlayer[9]	Ambiguity	Articles and messages	Hoaxes, scams, malware, bogus warning, fake news, misleading, true, humour, spams, etc.

Source: Zhou, Xinyi and Reza Zafarani, "Fake News: A Survey of Research, Detection Methods, and Opportunities," Table 3.

We will be discussing the practices that can put in place for Fake News detection on Social Media, throughout the next chapters of my master's thesis.

2. Research Approaches and Methodologies for Fake News detection

Due to the increased dissemination of Fake News, the amount of research in this subject and the development of new models and methodologies for Fake News detection is also rapidly growing. The frequently used automatic detection methodologies and research directions for Fake News are summarized in this chapter.

2.1 Research directions

This part is dedicated to the introduction of the main research directions, as presented by Sue Kai et al., in the following paper "Fake News Detection on Social Media: A Data Mining Perspective"[52]. The Fake News detection algorithms on social media is a rapidly evolving field of research, so, it is worthy of highlighting the promising research directions from a data mining point of view.

As presented in the below figure, there are four main research directions: Data-oriented, Feature-oriented, Model-oriented and Application-oriented.

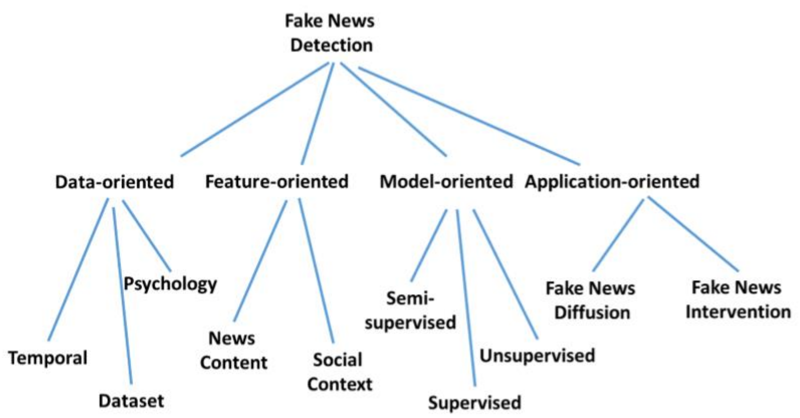

Figure 6: Future directions and open issues for Fake News detection on social media.
Source: Shu, Kai, Amy Sliva, Suhang Wang, Jiliang Tang and Huan Liu, "Fake News Detection on Social Media: A Data Mining Perspective," 11, Figure2.

In the 3rd chapter of my master dissertation, I will be focusing on the Feature-oriented perspective, giving less importance to the other aspects of Fake News detection. However, essential elements related to the other aspects are vital to be considered in order to develop a quality algorithm that detects Fake News.

2.1.1 Data-oriented

From a dataset point of view, researchers demonstrated that currently, no dataset possesses all necessary resources to be able to extract all necessary features for the Fake News prediction. Another interesting point is the early detection of Fake News, which aims to give early notifications of Fake News existence during the dissemination process, before it becomes virally propagated. The research on this topic is getting more and more important. A promising approach is to develop a large news benchmark dataset that scientists can use to enable further research in this field. However, it is important to note that obtaining a reliable and complete Dataset could be time-consuming, demanding intensive work, and to implicate many parties.

The literature on social psychology has studied various aspects of Fake News qualitatively[53]. However, quantitative studies are limited to confirm these psychological hypotheses. In fact, capture various specific data is still very difficult for the researchers of that field.

Also, the intention detection from news articles represents a very sophisticated information, but researchers have not been able until now to find how to extract such information from a News dataset. Most Fake News studies focus on authenticity detection and ignore the intentional aspect of Fake News. The detection of intention is complicated because the intention is often explicitly unavailable. Thus, exploring how to use data mining techniques to validate and capture psychological intentions is very challenging and would be a big step forward for the research in this field.

2.1.2 Model-oriented

Model oriented Fake News research aims to create effective, practical, and accurate models for Fake News detection. The natural process of model-oriented research is incorporating those features extracted into various types of models, such as Support Vector Machines, Random Forest, Naive Bayes, Decision Tree, Logistic Regression and then select the classifier that performs the best based on some specific criteria.

More research can be done to develop more complex and sophisticated models and to better use extracted features, such as "*aggregation methods*", "*probabilistic methods*", "*ensemble methods*", or "*projection methods*".[54]

Another topic of research from the Model oriented perspective would be to explore the possibility to build efficient and accurate semi-supervised and unsupervised models as actual Fake News detection algorithms rely on supervised approaches.

2.1.3 Application-oriented

The application-oriented research contributes indirectly to Fake News detection. There are two primary approaches concerning the Application-oriented research: *Fake News diffusion* and *Fake News intervention*[55].

Fake News intervention methods intend to decrease the impact of Fake News by minimizing the spread scope proactively or after Fake News goes viral by putting in place reactive intervention methods.

Fake News diffusion describes the Fake News propagation paths and patterns on social media. Research studies have demonstrated[56,57] that True and Fake information follow different patterns while disseminating on social media. Exploring the different propagation patterns and extract meaningful information from that, it is in the heart of the Application-oriented research.

2.1.4 Feature-oriented

The purpose of feature-oriented Fake News research is to identify key characteristics to detect Fake News from various sources of information. As it has already been explained in the introduction, but also it will be further analysed in Chapter 2.3 "Methodologies for Feature Extraction", there are two primary sources of information. The one is the news content, and the other is social context information.

From the news content point of view, we implement linguistic (e.g., NLP tasks: text classification and clustering, author identification[58], deception detection[59]) and visual-based methods to extract features.

Although all these evolutions on the research in this field, the underlying characteristics of Fake News have not been fully understood, and further research is needed. For instance, embedding techniques such as word embedding[60], and deep neural networks[61] have the potential to be further developed.[62] Also, visual characteristics obtained from images are shown to be significant indicators for Fake News detection[63]. However, there are very few studies that focus on how to exploit the visual characteristics effectively. That is a challenge for the researchers to reveal within the next years.

Regarding the social context features extraction, there is already a lot of quality research on this field. The new challenge is to extend the current researches on how other networks can be constructed in terms of different aspects of relationships among relevant users and posts. Advanced methods of network representations, such as network embedding are very promising.[64,65]

The feature-oriented methodology is a supervised binary classification process. Supervised binary classification is a machine learning method where the output categories are predefined and are used to categorize new probabilistic observations (News Articles), into said categories (Fake News or True News). That being said, to realize a binary classification, we need labelled data in our Training set.

Unlike supervised machine learning, unsupervised machine learning cannot adapt to the Fake News prediction problem. That is because the model takes X number of features as input and no corresponding output. The objective of the unsupervised machine learning is to model the underlying structure or find unknown patterns in the data. The two categories of Unsupervised machine learning are Clustering and Association problems. It is possible to use unsupervised Machine Learning techniques to extract meaningful features that can improve the quality of the Fake News prediction. For example, we can use unsupervised methods for community detection in social networks by dividing a graph into several groups such that they share similarities.

Feature Selection is one of the core concepts in machine learning classification process that has a significant impact on the performance of a model. The data features that we use to train a machine learning classification model have a significant influence on the quality and accuracy of the prediction. Irrelevant or partially relevant features can negatively impact model performance. My Master’s thesis focus on the advantage of combining social context and news content features to construct a model that optimizes the quality of the prediction.

2.2 Model construction Methods

The traditional approach to detecting Fake News is to use a knowledge-based fact-checking system. This system compares relational knowledge extracted from to-be-verified news content and stored in a knowledge graph, with a truth dataset collected from the web.[66]

However, the most critical disadvantage of using that system is that it can only detect False News and not Fake News (i.e., intentional false news)[66]. Their difference has been explained in the first Chapter, more in detail. Below you can find a table that summarises the different Model construction methods.

Table 2: Summary and Comparison of Perspectives to study Fake News

	Knowledge-based	**Style-based**	**Propagation-based**	**Credibility-based**
Potential Research Task(s)	Fake news analysis and detection	Fake news analysis and detection	Fake news analysis, detection, and intervention	Fake news analysis, detection, and intervention
Fake News Stage(s) Studied	Creation, publication and propagation	Creation, publication and propagation	Propagation	Creation, publication and propagation
Information Utilized	News-related	News-related	Primarily social-related	News-related and social-related
Objective(s)	News Authenticity Evaluation	News Intention Evaluation	News Authenticity and Intention Evaluation	News Authenticity and Intention Evaluation
Techniques	Relation-based	Feature-based	Primarily Relation-based	Relation-based and Feature-based
Resources	Knowledge graphs, e.g., Knowledge Vault	Theories, e.g., reality monitoring; however, not many theories focus on fake news	Theories in Table 2	Theories in Table 2.
Related Topic(s)	Fact-checking	Deception analysis and detection	Epidemic modeling, rumor analysis and detection.	Clickbait analysis and detection, (review and Web) spam detection.

Source: Zhou, Xinyi and Reza Zafarani, "Fake News: A Survey of Research, Detection Methods, and Opportunities," Table 10.

2.2.1 Knowledge-based

Studying Fake News from a knowledge-based point view[21] is aimed at detecting Fake News using a method called fact-checking. Fact-checking that originates from journalism seeks to evaluate the authenticity of news by comparing the information obtained from the news content (e.g., its claims or statements) with known facts (i.e., real knowledge).

In this part, we will be introducing the two methodologies used for fact-checking, which is the traditional one, also known as manual fact-checking, as well as, the automatic fact-checking. Here is an interesting presentation of the fact-checking technique:

> Fact checking and analysis (FCA, for short), viewed as the process of analyzing a piece of information, crossing it with existing knowledge, verifying its accuracy and possibly enriching it with nuances, comments and connections to reputable sources, has an inherent part of human effort, thus it is unlikely to ever be completely automatized. This is because FCA requires not only modus ponens-style manipulation of facts (i.e., automated reasoning, for which mature software tools exist by now), but also involves information extraction tasks at which humans are still better than software, as well as judging (through the context of one's experience) the cultural environment of the target audience of the analysis result.[67]

Manual fact-checking

The manual fact-checking can be split into two main categories: expert-based checks, crowd-sourced checks.

Expert-based manual fact-checking relies on the opinion of experts for a given subject. They are also called fact-checkers. Commonly there is a group of experts formed by highly credible fact-checkers. The results of this fact-checking are exceptionally performing and bring very high accuracy in the prediction of Fake News. It is nonetheless very time consuming and costs a lot. That being said, this methodology is not appropriate when the volume of news to checking increases. Many websites have appeared these last years that have the objective to better serve the public with knowledgeable fact-checking. In Table 1 of chapter 1, there is a list of well-known websites doing this work.

Crowd-sourced Manual Fact-checking is based on a big population of ordinary people acting as fact-checkers (i.e., collective intelligence). Crowd-sourced fact-checking is hard to manage compared to expert-based fact-checking, less credible due to the fact-checkers' political prejudice and their inconsistent annotations. Crowd-sourced fact-checking websites are still in early development, unlike expert-based fact-checking. An example is Fiskkit, where users can upload articles, ratings for articles expressions and sentences, and tags what better describe them. The sources provided in the articles serve to identify the type of content (e.g., if it is news or something else) and evaluate their credibility.

<u>Automatic Fact-checking</u>

As also explained before, the manual fact-checking is not appropriate for the massive quantity of newly generated news notably on social media. For that reason, there were developed Automated fact-checking methods relying mainly on information retrieval (IR) and natural language processing (NLP), as well as, Network/graph theory[68] that contributes to tackle the scalability of news propagation.

Automatic Fact-checking requires a standardized representation of knowledge that can be processed automatically by machines.[69]

- **Fact Extraction:**

Knowledge (Facts) is often extracted from the web as “raw facts” also known as relation extraction[70]. Before to create a knowledge base or a knowledge graph, the raw facts are further processed and cleaned up by knowledge processing functions.

Four principal forms of Web content can be extracted. These are text, tabular data, structured webpages, and human annotations.

- **Fact-checking:**

Applying fact-checking for authenticity detection requires to compare the knowledge extracted from the news content to be verified, and the facts stored in the knowledge base(s) or knowledge graph(s) where the real knowledge is.

The fact-checking strategy for a triple SPO (Subject, Predicate, Object) is generally to assess the possibility that the edge labeled “Predicate” exists from the node labeled “Subject” to the node representing “Object” in a knowledge graph. Where “Subject” and “Object” are entities and “Predicate” is the relation between them.[71]

For example, below is the knowledge graph elements created for the following phrase, “Leonard Nimoy was an actor who played the character Spock in the science-fiction movie Star Trek.”

subject	*predicate*	*object*
(LeonardNimoy,	*profession,*	*Actor)*
(LeonardNimoy,	*starredIn,*	*StarTrek)*
(LeonardNimoy,	*played,*	*Spock)*
(Spock,	*characterIn,*	*StarTrek)*
(StarTrek,	*genre,*	*ScienceFiction)*

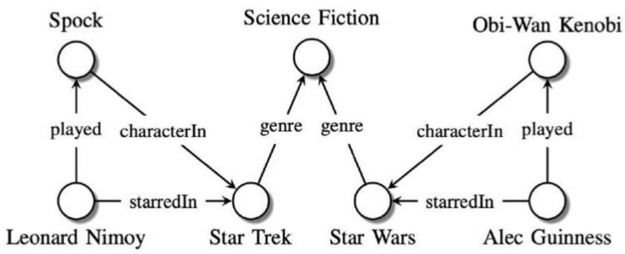

Figure 7: Sample knowledge graph. Nodes represent entities, edge labels represent types of relations, edges represent existing relationships.
Source: Nickel, Maximilian, Kevin Murphy, Volker Tresp and Evgeniy Gabrilovich. "A Review of Relational Machine Learning for Knowledge Graphs.", Figure 1.

2.2.2 Style-Based

Likewise knowledge-based, studying Fake News from a style-based perspective also emphasizes on investigating the news content. Although knowledge-based methods assess the news article authenticity, style-based methods try to determine news intention, i.e., is there an intention to mislead the public opinion or not?

Fake News style is "*A set of quantifiable characteristics (e.g., machine learning features) that can well represent Fake News and differentiate Fake News from the truth*."[21] Style-based methods attempt to identify Fake News by catching news content styles. The intention of the Fake News is generally to mislead the public and, on this purpose, their authors use specific writing styles designed to attract and convince a wide range of people not seeing in the Real News articles.[72] There are two style-based methods, the Deception-oriented and the Objectivity-oriented.

Objectivity-oriented

This methodology captures style signals in news articles characterized by low objectivity. Low objectivity can indicate the intention of an article to mislead the readers. Hyper partisan styles and yellow journalism are two examples of signals. Hyper partisan styles concern radical behaviour in favour of a political party that often correlates with a strong motivation to create Fake News. Yellow journalism concerns articles with attention-catching headlines (i.e., clickbait) that have the objective to dramatize, sensationalize, and mislead the public. Clickbait titles can be a very reliable indicator for Fake News detection[73].

Linguistic based features are commonly used to detect hyper partisan articles and describe content style from four primary language levels: lexicon, syntax, semantic and discourse[21]. Their quantification mostly relies on Natural Language Processing (NLP) techniques.

I will go in more in-depth details regarding style-based feature extraction in the third part of this chapter, as well as, in the third Chapter.

Deception-oriented

Deception-oriented style-based techniques rely on forensic psychology[74] and capture deceptive statements or claims in news articles. Several forensic tools have been developed until now, including criteria-based Content Analysis[75] and Scientific-based Content Analysis[76].

There are two main perspectives of the natural language processing (NLP) models to detect deception in news articles: deep syntax and rhetoric structure. Deep syntax models use Probabilistic context-free grammers (PCFG) to transform sentences into rules that describe the syntax structure.[52] The rhetorical structure can be used to capture the differences between Fake and True sentences in a news article[77]

2.2.3 Propagation-Based

Unlike knowledge-based and style-based approaches that study Fake News through their content, when studying Fake News from a propagation-based approach, we base our analysis on data related to Fake News dissemination. For example, we try to identify how news propagates and how users spread it. Propagation-based methods may be the most interesting and promising research direction on the Fake News detection topic, especially when the studies focus on the news propagation over time.

It has also been proven that the mechanism of Fake News dissemination is like the diffusion of an infectious disease[78] and can be explained with models of network epidemics[79]. In addition, there is considerable empirical evidence that Fake News proliferates differently from True News [80] forming patterns that enable automatic detection of Fake News. Moreover, it is worth considering that propagation-based features can be generalized across different languages, locations, and geographies, contrary to content-based features that need to be created exclusively for each specific language. Two main Fake News detection techniques based on propagation models exist[81]:

- Cascade-based Fake News detection techniques, which take direct advantage of news propagation paths and news cascades to identify Fake News. This technique enables Fake News detection by capturing the similarity of its cascade to that of other Fake News or standardizing the representation of its cascade to facilitate distinguishing Fake News from True News. [82]

- Network-based Fake News detection methods which construct flexible networks to detect Fake News propagation indirectly. The networks built may be homogeneous, heterogeneous, or hierarchical. Homogeneous networks consist of a single type of entities, such as post or event. Heterogeneous networks are related to various types of entities, such as posts, sub-events, and events. Hierarchical network includes news aspects (i.e., latent sub-events), multiple layers and utilize a graph optimization framework to infer event credibility.[66]

2.2.4 Credibility-based

Studying Fake News from a credibility-based perspective, it means that we focus on both news content and social context information. For example, intuitively, we can understand that a news article posted on unreliable websites and forwarded by unreliable users is more likely to be Fake News compared to the news posted by credible users to reliable websites.

Fake News study from a credibility perspective overlaps with a propagation-based study of Fake News, where we investigate the interactions between news articles and elements such as publishers[83], users[84], and posts[85].

Content-based Models:

Content-based models evaluate the credibility of posts using a sequence of language characteristics obtained from user comments and adopt a strategy comparable to style-based detection of Fake News.

Behaviour-based Models:

Behaviour-based models analyse key characteristics of unreliable comments obtained from user-behaviour related metadata. We organize these related behavioural features into five categories: burstiness, activity, timeliness, similarity, and extremity. Mukherjee[86] propose the Author Spamicity Model (ASM), including behavioural features from all five categories. It is worthy of mentioning that studies have shown that the arrival times of ordinary reviewers are stable and uncorrelated to their temporal rating patterns, while spam attacks are generally bursting and either positively or negatively correlated to the rating.[87]

2.3 Methodologies for Feature Extraction

Many of the current Fake News detection methods introduced in the previous section rely on the feature extraction, as so, various authors oriented their research related to Fake News detection on feature extraction methodologies. In this section, we find a brief presentation of the main features categories and we analyse their advantages and limitations.

It is essential to mention that feature extraction is a large topic with an extensive range of sub-categories of methodologies. In this section, we will be focusing on the most commonly used ones that I identified by reading articles and papers on this topic. It is also worthy to point out that the objective of my master thesis is to highlight the importance to combine different methods of feature extraction in order to achieve better accuracy models. More specifically, feature extraction though News content information and Social context information. These are the two-parent categories concerning almost all feature detection methods.

The third and last chapter of my master thesis, it will be based and inspired by the information presented in this chapter.

2.3.1 News content features

News content features aim to analyse a news article through its associated meta-information. Here are the main parts of news articles that consist the source of the feature extraction[52,88:]

- Source: The author or the publisher of the news article
- Headline: Short title text intended to attract readers ‘attention and describe the article’s main topic
- Body Text: Main text elaborating the news article details
- Image/Video: Part of a news article body text that gives visual elements to support and accompany the news article.

Thanks to the above content attributes, different kinds of features can be extracted and evaluated. The extraction of content-based features is the traditional method that it is also used in traditional news media and, more generally, in all cases where no social information can be retrieved. Historically, this method was used to detect spam email messages[89] and web pages[90].

One inconvenient of the content-based methods is that there is a significant number Fake News that are intentionally written to mislead the users, which makes it very difficult to detect Fake News only based on news content information[83].

2.3.1.1 Linguistic Features based Methods

Linguistic based features can be broken out into three main subcategories:

Lexical Features

Lexical features are character-oriented (character-level) and word-oriented (word-level) features. Some examples of lexical features are total words, characters per word, frequency of large words, unique words, average word length, informal or slang words. Researchers[91] calculate word complexity with the help of readability indexes, which attribute to a news article a grade-level score based on the number of syllables per word. Also, they use a metric called "Type Token Ratio" ("TTR") to capture the lexical diversity of a document.

Syntactic Features

Syntactic features concern sentence oriented (sentence-level) features. Examples of syntactic features are the frequency of function words and phrases (i.e., "n-grams" and bag-of-words features), the parts-of-speech (POS) tagging, sentence-level complexity (sentence's syntax tree depth), punctuation count, average sentence length. Syntactic feature extraction methodologies can also use domain-specific approaches, such as quoted words, external links, number of graphs, and the average length of graphs[13].

Psycho-linguistic Features

Emotionality and inflammatory language are two of the most common characteristics of Fake News articles. Examples of Psycho-linguistic features are positive, neutral, or negative sentiment (sentiment analysis metrics[92]), high occurrences of personal pronouns, and sensationalist (click-bait) language.

An example of a sentiment analysis tool is SentiStrength which allows measuring the level of positive or negative sentiment intensity of each news article. Researchers[93] also extract psycho-linguistic features metrics with the use of Linguistic Inquiry and Word Count (LIWC) program.

2.3.1.2 Visual Features based Methods

A Fake News image is a Fake News image attached in a Fake News post. In propaganda, they are often used to cause anger or other emotional response. According to recent research[94], Fake News images concern not only maliciously manipulated (digitally modified) images, but also Real Images that represent unrelated events (non-manipulated images, but the content is misleading).

Visual characteristics of news articles play a vital role in Fake News detection. That is because their use is widespread in Social Media for propaganda purposes.

The question is how the visual content can be used to determine a news article as True or Fake, which is equivalent to classifying a specified image as a Fake News image or a Real News image.

On that purpose, various visual-based and statistical-based features enable Fake News detection[95]. Examples of visual features are scores related to clarity, coherence, similarity distribution histogram, diversity, and clustering. Some examples of statistical features are count, image ratio, aspect ratio, multi-image ratio, hot image ratio, long image ratio.

2.3.2 Social Context Features

Recent research shows that there are three main methods for the social context feature extraction. These are User-based features, generated Posts-based features, and Network-based features. Below there are details on how to work with social context features using these three methods.

2.3.2.1 Post-based Features Methods

Post-based features are valuable information retrieved from multiple elements that characterise a social media publication. These features help to evaluate the truthfulness of a news article posted on social media. They are divided into three levels: post, group, temporal features.

Post-level features

Post-based features characteristics are represented by the users' reaction in the social media post concerning stances, topics, or credibility. This reaction can also be considered as the expression of an emotion (e.g., sensational reactions) or opinion (e.g., sceptical opinion) of the user towards a Fake News post on social media. Thus, extracting quality post-based characteristics is important in strengthening Fake News detection through social media reactions as reflected by users in posts.

- Stance features show the reaction of the users to the news, such as encouragement or rejection[96].
- Topic features that can be extracted using topic models like latent Dirichlet allocation (LDA)[97].
- Credibility features assess the level of credibility of content posted on social media[98].

Group-level features

Group-level features are the aggregated values of the post-level feature values concerning posts of a given news article. Two examples of group-level features are the average Credibility score of a social media post and the number of posts.

Temporal-level features

Temporal level features take into account the temporal variability of the post-level feature values[97]. Depending on the shape of the time-series that the feature values generate (e.g., number of posts), we can calculate mathematical features, like SpikeM parameter.[99]

2.3.2.2 Network-based Features Methods

Network-based feature extraction uses social context information retrieving through news propagation patterns. Researches[66] have shown that Fake News dissemination in networks demonstrates patterns that are very different from Real News dissemination patterns. Below there is a presentation of the different Fake News patterns and how they can express them as a range of quantifiable and valuable features across network levels (i.e., node, triad, community, network).

It is important to note there are two different types of networks: the *homogeneous* and the *heterogeneous*[100]. Here we examine a Homogeneous Network which has a unique type of nodes and edges.

1) More Spreader Pattern – Features list

a) General-Non-Attributed Spreaders:

- The number of users that spread each Fake or True News article.

b) <u>Specific-Attributed-Spreaders:</u>

- User Susceptibility features: the number of times the user participated in the propagation of different Fake News and the frequency of this action.
- User Influence features: an approximation of a user influence can be calculated with the help of centrality scores within the network. Below are examples of the different types of centrality metrics, all of which use node-to-node links to define their network positions:
  (i) Degree centrality
  (ii) Closeness centrality
  (iii) Betweenness centrality
  (iv) PageRank score
  (v) Eigenvector centrality

2) Farther Distance Pattern – Features

- Geodesic Distance: shortest path length between the two most distant spreaders within the network.

3) Stronger Engagement Pattern – Features

a) <u>Group Engagements features:</u>

- Number or proportion of times that susceptible users have spread the news story.
- Number or proportion of times that normal users have spread the news story.

b) <u>Individual Engagements features:</u>

- Average spreading frequencies of (susceptible, normal, all) users who have participated in the news article dissemination.

4) Denser Network Pattern – Features

a) <u>Ego Level features</u>

At the ego level, to evaluate the density of networks created by users-spreaders that have interacted with a certain news article, we look at the numbers and proportions of connections that these users have generally created with other users-spreaders, and specifically with other normal or susceptible users-spreaders.

b) Triad Level features

A triad is a set of three linked users, which is the most basic subgraphs of networks.

- General Triads feature: total number of triads within a graph.
- Specific Triads (triads formed between susceptible and normal) feature: the number and proportion of every type of triads within each Fake and True News

c) Community Level features:

In networks, a community structure refers to the occurrence of groups of nodes in a network that is more densely connected internally than with the rest of the network.

- the number of communities within each Fakes News and True News

- the proportion of the communities within each Fakes News and True News

2.3.2.3 User-based Features Methods

As also explained previously, Fake News is likely to be propagated by social bots or cyborgs. Therefore, including user-based features in the Fake News detection process can provide very useful information.

In addition, various studies on real datasets[101] have shown that there are social media users who trust Fake News more easily than Real News and generally these users have different characteristics from those who are more likely to trust Real News. These observations can facilitate Fake News detection by including related user-based features in Fake News detection algorithms.

These features can form two big categories of features: the individual-oriented features and group-oriented features.

Individual-level characteristics are useful to determine the reliability of the social media users by using various elements concerning user demographics (e.g., age, gender, hometown), number of followers/followees, number of friends, number of status, but also user-personality (Five Factor Model or "Big Five"[102]) and number of original tweets posted by the user, etc.

Group level user features represent general characteristics of the users' groups. The hypothesis is that spreaders of Fake News and True News belong to different groups that have distinguishing characteristics that can be represented by group-level features. Examples of group-level features are the average number of spreaders or followers, as well as, percentage of verified users that spread or tweeted a piece of news[103].

3. Fake News Prediction Algorithm

The objective in this 3rd chapter is to present the different steps, aspects, and process of a Fake News detection algorithm by developing and explaining an example of an algorithm that uses Social-context and Textual features as input and predicts whether a News article is Fake or True.

This 3rd chapter of my master thesis comes along with the Python 3 code I have written on that purpose. You can find it in the .zip file sent along with my manuscript. Thus, every diagram, plot, or result you can see in this chapter is based on that code. In the .zip file, you can also find the data used for the analysis and a Readme.txt file that gives instructions on how to execute the code. Moreover, the code is commented to facilitate reading it and understand to which chapter is associated.

After having done researches on the different methodologies of Fake News identification, I decided to concentrate my analysis on the feature's extraction methodologies. Feature-oriented research direction (cf., chapter 2.1.4) refers to various aspects and techniques in its application. I give you some introduction elements about existing methodologies on that topic and I focus on the two main categories of features, which are the Textual features retrieved from the News article Body Text and Social based features.

Concerning the textual features, I will be focusing on style-based features and patterns, more specifically to lexical and syntactic features but also psycho-linguistic features. Concerning Social context analysis, I will be concentrating on propagation-based and network-based approaches for feature extraction.

By creating that hybrid model that combines two different types of data, I expect to have better prediction results and provide a better model accuracy compared to other techniques that use only content or social context features each time.

My work will be based on the tri-relationship ("TriFN") embedding framework described in recent researches[104], which has five different aspects that you can find listed below:

- news contents embedding
- user embedding
- user-news interaction embedding
- publisher-news relation embedding
- news classification.

The publisher-news relation embedding, and news classification will not be considered fin my analysis. Therefore, I will be focusing on the first three components listed above, which are the news content embedding, the user embedding, the user-news interaction.

It's worthy of mentioning that a couple of data preparation tasks must be done before to be able to extract features from a dataset. These preparation tasks concerns library import, loading datasets and data cleaning. In Wikipedia we find the following definition of the Data Preparation. "*Data preparation is the first step in data analytics projects and can include many discrete tasks such as loading data or data ingestion, data fusion, data cleaning, data augmentation, and data delivery*."[105] According to Steve Lohr of The New York Times "*Data scientists, according to interviews and expert estimates, spend 50 percent to 80 percent of their time mired in the mundane labor of collecting and preparing unruly digital data, before it can be explored for useful nuggets*."[106] You will find more in-depth details on the data preparation tasks by looking at the Python 3 code that accompanies this manuscript.

3.1 The Objective

The objective of my model will be to classify a News article into Fake or True News category by following a classification process that combines features extracted from the article body text and features extracted from Social-context analysis.

We define Fake News detection as a binary classification problem. I will be using Naïve Bayes classifier and Random Forest classifier to realize the classification task.

The output will be binary classes:

Fake News (1)

True News (0)

3.2 Social Context Features Extraction - Network Analysis

Social context characteristics can be produced on social networks through user-driven news involvement. Researchers less investigate social Network features compared to content features. It is nonetheless a very challenging field for further research as the first studies have already shown that news propagation patterns have a significant impact on Fake News detection. Social engagement reflects the news propagation process and provides valuable information that allow us to evaluate the integrity of the news articles. As also explained in Chapter 2.3.3 “Social Context Features”, the social media context features are based on three key components: The Users, the created Posts, and the Networks. In the data I possess for my analysis, apart from the News articles divided into Training and Test set and their labels (Fake or True News), I own the following information: the news-user relationship (posts/spreads), the user-user relationship (followers and followees) and from those I can extract additional and auxiliary meaningful information.

In this chapter, I give you some context elements about the different approaches that are used to extract social context information, I present the different hypothesis I constructed, and the different variables that I selected to test each hypothesis. Finally, I give some descriptive statistics about the variables by using diagrams and I realize statistical tests to have a first impression of whether there is a dependence between the response variable, which is in our case the variable “Target” and each explanatory variable.

Creating a graph

In order to be able to realize the above-mentioned tasks in Python 3, the first thing I must do is to create a Network diagram from the data I own. Network diagrams that are also called Network Graphs, they show interconnections between entities. Each entity is represented by a node and the connections between nodes are represented by edges. Here the nodes are the social network users, and the edges are the relationships between the nodes (users). The relation between the nodes here is related to the fact that a user follows another user and thus we have a Follower and Followee relationship. Facebook or LinkedIn create an undirected graph as friendships apply in both directions. However, on Twitter or other social media networks where the type of relationship is the followers and non-followers, the concept of "followers" rather than "friendships" makes these graphs directed. To realize the latter, I will use from_pandas_edgelist() function of the NetworkX package for Python 3, which returns a graph from a Pandas DataFrame containing an edge list.

3.2.1 Community Detection

The community detection has the objective to identify highly connected groups of individuals that in our case are social media users, inside the networks. These groups are called communities. Various methods have been developed[107] over the years that enable us to efficiently identify communities in complex networks. A common practice aims at maximizing a measure called "modularity" in the network, which means maximizing the number of edges (relationships) inside the communities and minimizing the number of edges between the communities.

To realize the community detection, I will be using the community.best_partition() community detection algorithm from the community API package for Python 3. This algorithm computes the partition of the graph nodes which maximises the modularity using the Louvain Heuristics. This is called the partition of highest modularity.

The creators of the community detection algorithm describe the algorithm as follows:

> We propose a simple method to extract the community structure of large networks. Our method is a heuristic method that is based on modularity optimization. It is shown to outperform all other known community detection method in terms of computation time. Moreover, the quality of the communities detected is very good, as measured by the so-called modularity. This is shown first by identifying language communities in a Belgian mobile phone network of 2.6 million customers and by analysing a web graph of 118 million nodes and more than one billion links. The accuracy of our algorithm is also verified on ad-hoc modular networks. [108]

We have in total 408 824 edges and 23 865 nodes. By applying the Louvain method for community detection, we obtain four communities (groups). That means the users (nodes) in our data set are assigned to one of those four communities. In reality, the algorithm identified in total 192 different communities, but if we look closer to the groups formed, we can easily understand that starting from the 5th group the communities have less than four nodes/users associated. Thus, I assume that those users are not part of any community. Below you can see a pie chart that shows the percentage of the total nodes in each community. You can observe that the community no.3 contains the highest number of users (12 442), followed by community no.0 (4 537 users), no.1 (3 447 users), no.2 (2 530 users). Finally, 3.8% of the total number of users are part of small communities of a maximum of 4 users each. These users are not part of any community, and thus I will attribute a community with value “-1” to be able to distinguish them from the rest of the observations.

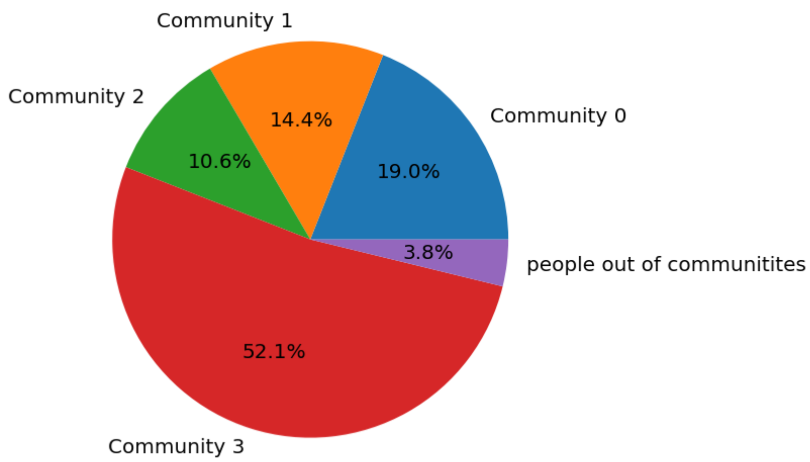

Figure 8: percentage of users per Community group.

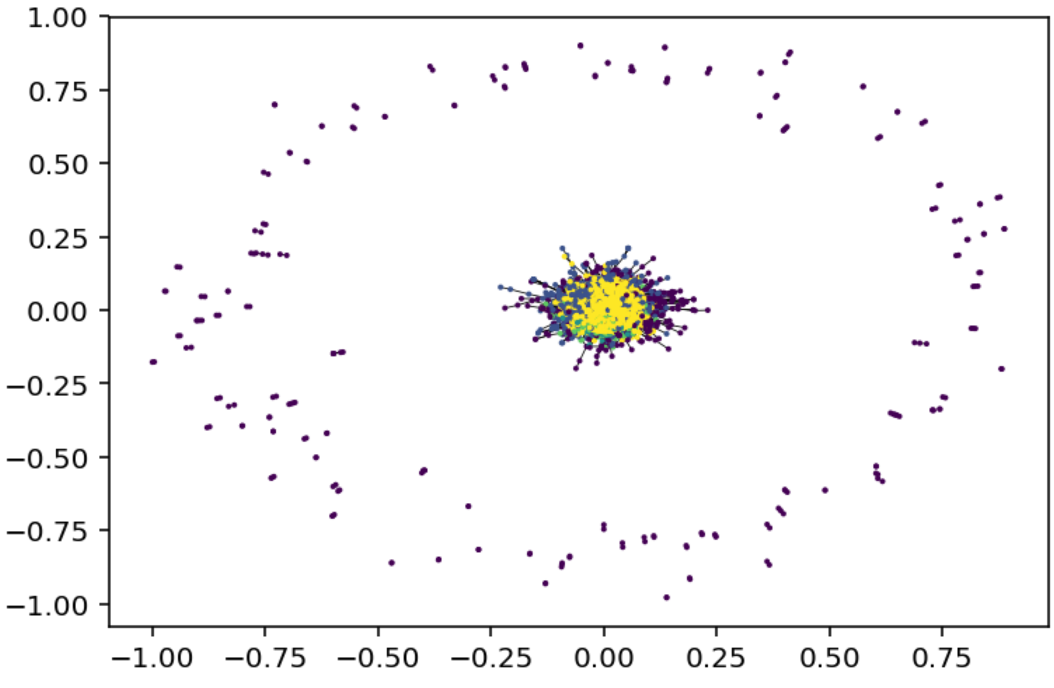

Figure 9: Network Graph Visualisation that includes users out of communities

The Nodes of the community "-1" (nodes that are not part of communities) form a circle around the rest of the node's communities. If we delete those nodes and we visualise again a sub-graph, we can observe more clearly the 4 communities formed (i.e., Figure 10).

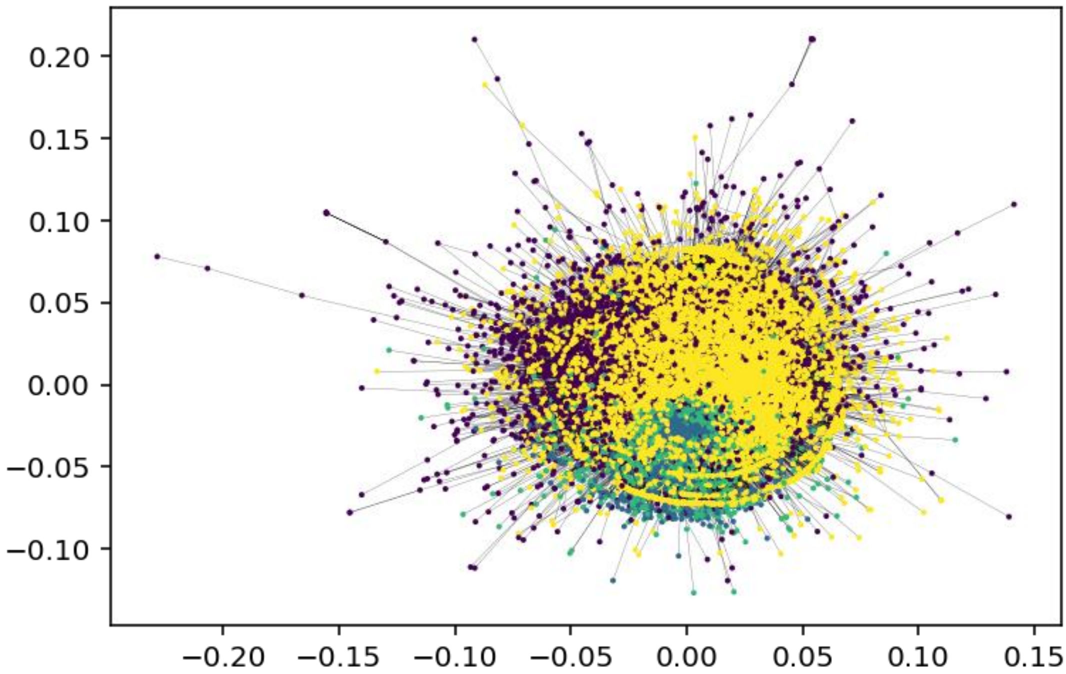

Figure 10: Network Graph Visualisation that included users out of communities

Based on the above analyses and the result I have got; I am now able to make hypotheses related to the communities formed by the social network users and test those hypotheses.

List of Hypotheses:

1) Fake News spreaders are part of a specific community. In other words, Fake News users form specific communities.
2) The most frequent community per News can indicate if a News article is Fake or not.

List of features:

I chose the below two features to test my hypotheses.

1) **Most frequent community of users per News article (hypothesis 2)**

Descriptive analysis and statistical tests

Below you can find a plot that visualize the volume Fake News spreads occurred within each community group and a second graph that shows the volume of communities occurred within Fake or True News.

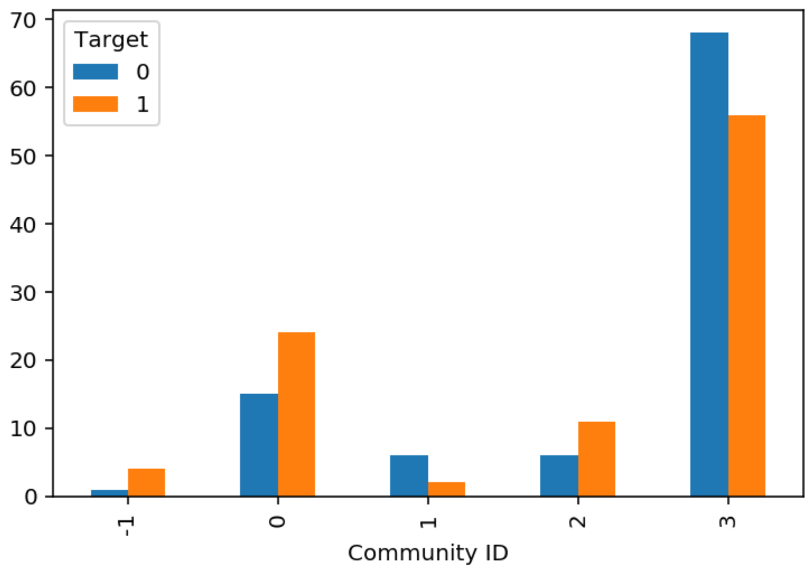

Figure 11: Most frequent community per News and volume of True and Fake news within each community.

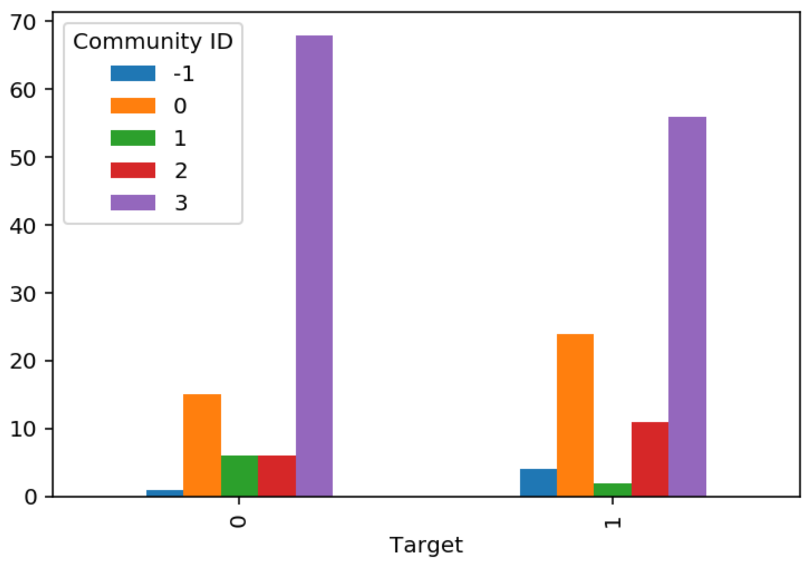

Figure 12: Most frequent community per News and volume of each community within Fake and True News categories.

From the above diagrams we can observe that users that are part of the Community 2 or Community 0 are more likely to spread Fake News compared to Users that are part of other communities. I realized a Chi-Squared statistical test to validate or not my hypothesis. The Chi-Squared test is a statistical hypothesis test that assumes (the null hypothesis) that the observed frequencies for a categorical variable match the expected frequencies for the categorical variable. The test calculates a statistic that has a chi-squared distribution.

Contingency Table

	Target	
Community	True News (0)	Fake News (1)
-1	1	4
0	15	24
1	6	2
2	6	11
3	68	56

Chi-squared test results

The test result gives a p-value = 0.07 which is superior to 5% significant level calculated by inverting the 95% probability used in the critical value interpretation. Therefore, we fail to reject the null Hypothesis, so the most frequent community per news is independent to “Target” value which is the response variable for Fake and True News detection, and the opposite.

Below are the results given by using Python 3 chi2() and chi2_contingency() function of the Scipy.stats package :

```
probability=0.950, critical=9.488, stat=8.504
Independent (fail to reject H0)
significance=0.050, p=0.075
Independent (fail to reject H0)
```

2) Number of different communities occurred per News Article (hypothesis 1)

<u>Descriptive analysis and statistical test</u>

Even though the number of communities that occurred per News Article is basically a numeric variable, I will be considering this as Categorical given that the only possible values are 1, 2, 3, 4. Below you can find a bar plot that shows the number of communities occurred within Fake and True News and the number True and Fake News against the number of communities.

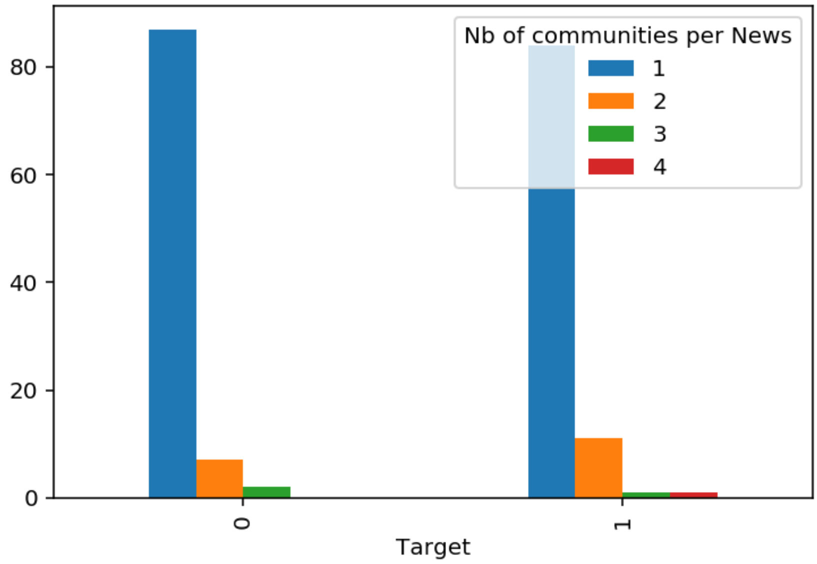

Figure 13: Number of different communities occurred within Fake and True News categories.

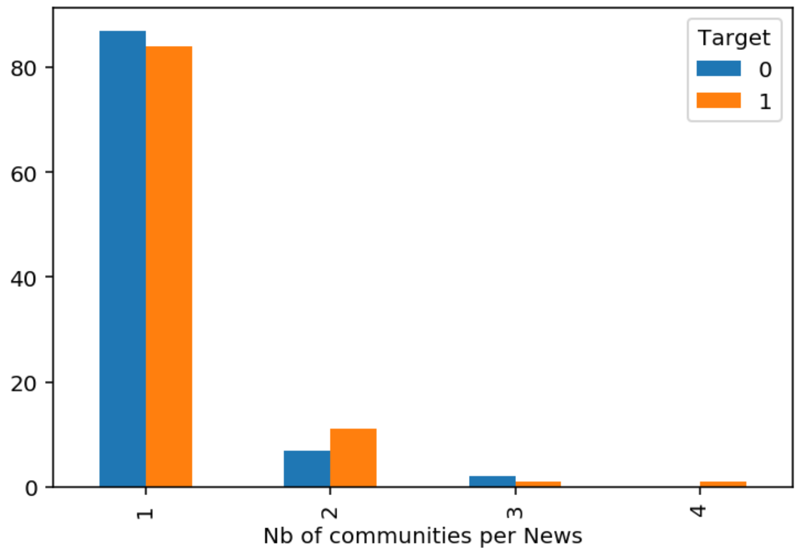

Figure 14: Number of True and Fake News for the different group's occurrences.

By looking at the above graphs we can easily understand that most of the News (Fake or True) are spread from one community of users. In the same time, it looks like there is not such a difference in the number between True and Fake News. Below I will realize a Chi-squared test to test my hypothesis.

Contingency Table

	Target	
Nb. Of communities per News	True News (0)	Fake News (1)
1	87	84
2	7	11
3	2	1
4	0	1

Chi-squared test results

The test result gives a p-value = 0.518 which is superior to 5% significant level calculated by inverting the 95% probability used in the critical value interpretation. Therefore, we fail to reject the null Hypothesis, so the variable Nb. of communities per News is independent to "Target" variable which is the response variables for Fake and True News.

It looks like the communities of users in our dataset don't have very distinguishing characteristics in order to form communities of users with significantly different behaviour. In addition, by looking at the Network graph we could observe that the communities formed they were very "closed" the one to the other.

Results using Python 3 chi2() and chi2_contingency() function of the Scipy.stats package :

```
probability=0.950, critical=7.815, stat=2.270
Independent (fail to reject H0)
significance=0.050, p=0.518
Independent (fail to reject H0)
```

3.2.2 Network Centrality and User Importance

After attributing to every user to his associated community and extract information from that, another important information we can retrieve is to find which nodes are the most influential in our network. On that purpose we can calculate Centrality measures. There are a couple of different ways for answering the following question "*Which nodes are the most important*?" and there are various ways of calculating user centrality. Some of the most famous are the degree centrality, the betweenness centrality, and the eigenvector centrality scores. Page Rank algorithm is another metric that determines the importance of a node within a Network. Unlike other centrality measures, Page Rank is based on the importance of the neighbourhood but not the distance. Below you can find two hypotheses I have done based on the user importance and centrality within a Network.

Hypothesis:

1) Fake News spreaders have a higher Eigenvector centrality score compared to True News spreaders.
2) Fake News spreaders have a lower Page Rank centrality score compared to True News spreaders.

List of features:

Below you can find the two variables selected to test the above two hypotheses.

1) **Average Eigenvector Centrality score per News (hypothesis 1)**

Eigenvector Centrality

I chose the Eigenvector centrality score, given that it is the one for which the calculation is taking less time and allows me to calculate it with my machine efficiently. Eigenvector centrality is measuring the influence of a node/user in a network, in our case social network. A score is assigned to each node, based on the principle that "*connections to high-scoring nodes contribute more to the score of the node in question than equal connections to low-scoring nodes*"[109]. When a node has a high eigenvector centrality score, that means it is connected to many nodes/users who on their side have high eigenvector centrality scores.

Descriptive analysis and statistical test

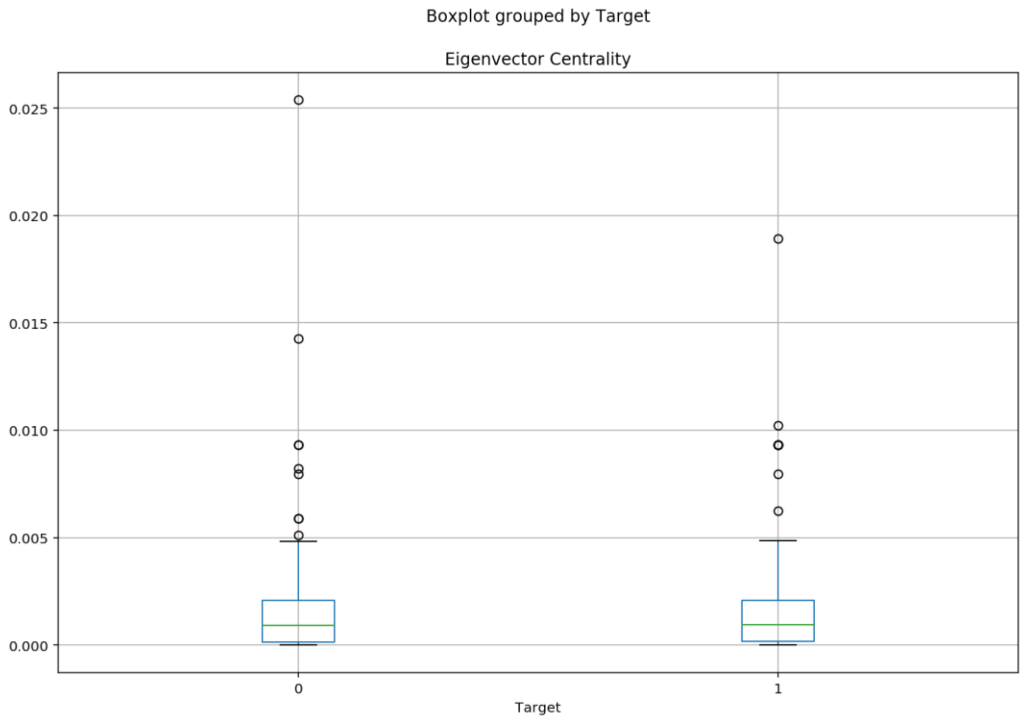

Figure 15: Boxplot Eigenvector Centrality Score for Fake News (1) and True News (0).

By looking at the above boxplot, we cannot easily understand if there is a significant difference of the average Eigenvector Centrality score between Fake News spreaders and True News spreaders. We will be evaluating the dependence based on a statistical test. In addition, we can observe that there are outlier values. I decided to keep them in the dataset as these outliers could concern very influential users that spread a specific class of News.

Statistical tests-Student's t-test

The null hypothesis here is that there is no significant difference in the mean test scores of the two sample groups (Fake News and True News) and that any difference down to chance. As the p-value is > 5%, I reject that hypothesis. Therefore, Eigenvector Centrality score doesn't have a significant relationship with the variable Target.

```
Ttest_indResult(statistic=0.023676849802760615,
pvalue=0.9811350895931088)
```

2) Average PageRank score per News (hypothesis 2)

<u>PageRank score</u>

PageRank (or PR) is a technique for detecting influential nodes in a graph. It is different from other centrality algorithms because the influence of a node depends on the influence of its neighbours. PageRank can be considered as a variant of Eigen Centrality. The primary differentiation between those metrics is that PageRank takes into consideration the link direction between the nodes and the weight[110].

It was developed by Google founders Larry Page and Sergei Brin, and it was initially destinated to be used in Google Search to rank web pages in their search engine results. More specifically, it gives to each page a relative score of importance and authority by evaluating the quality and quantity of its links. According to Cambridge Intelligence *"Each webpage is treated as a node in a network and is assigned a score based upon its number of in-coming links (its 'indegree'). These links are also weighted depending on the relative score of its originating node."* [110]

There are three distinct factors that determine the PageRank of a node (in our case Social Media User): (i) the number of links it receives, (ii) the link propensity of the linkers, and (iii) the centrality of the linkers.

Descriptive analysis and statistical test

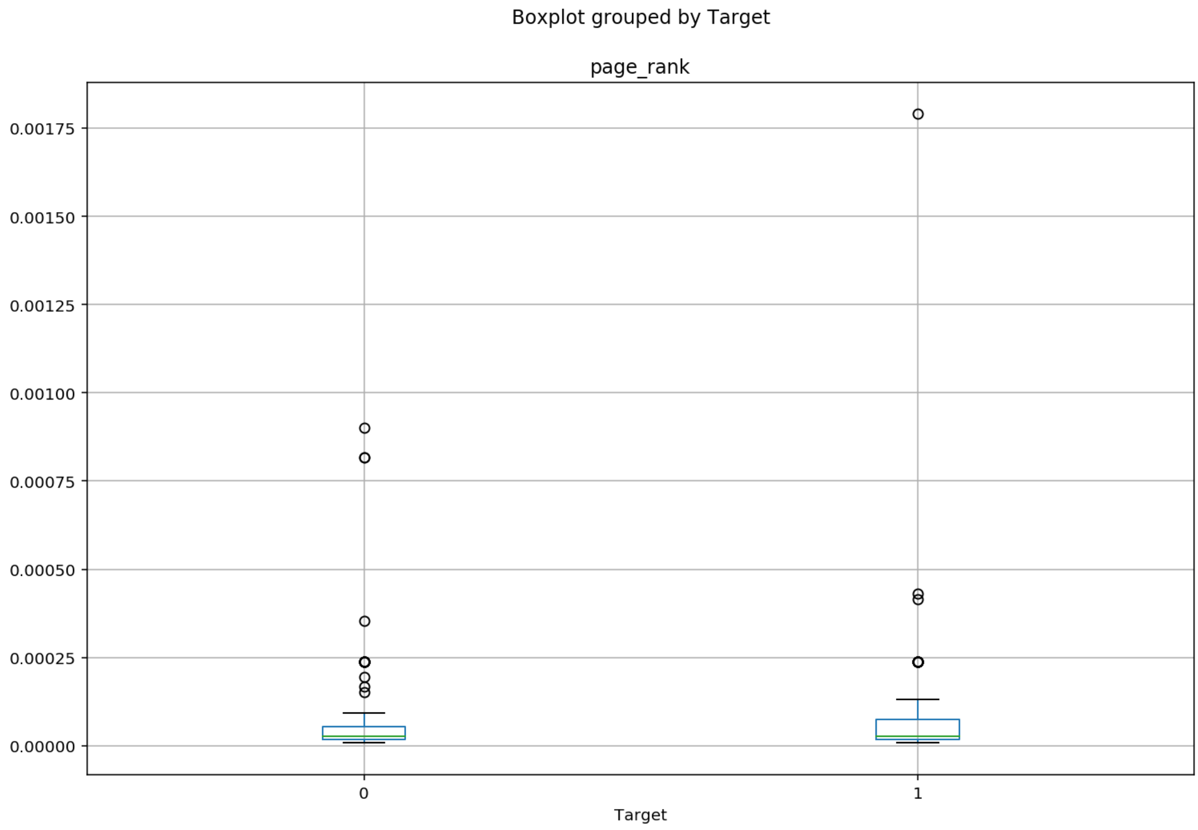

Figure 16: Boxplot Page Rank score for Fake News (1) and True News (0)

Like for Eigenvector Centrality, by looking at the page_rank /Target boxplot, we cannot easily understand if there is a significant difference of the average Page Rank score between Fake News spreaders and True News spreaders, even though the mean of the Page Rank score for the Fake News users seems slightly higher. We will be evaluating the dependence with the help of a statistical test. In addition, we can observe that there are outlier values. As I have done previously, I will keep the outliers observations in the dataset, as they could concern very influential users that spread a specific category of News.

Statistical test-Student's t-test

The null hypothesis here is that there is no significant difference in the mean test scores of the two sample groups (Fake News and True News) and that any difference down to chance. As the p-value is > 5%, I reject that hypothesis. Therefore, Page Rank score doesn't have a significant relationship with the variable Target.

```
Ttest_indResult(statistic=-0.15413461777311413, pvalue=0.8776663333427704)
```

3.2.3 News Propagation - General measures

In this part, I base my hypothesis to the theory introduced in chapter 2.3.2.2. I will be considering the different Fake News patterns and representation in Networks, as explained by Xinyi Zhou and Reza Zafarani[66]

List of Hypothesis:

1) More users spread Fake News than True News (More-spreader hypothesis 1)

2) Spreaders engage more actively with Fake News than True News (Stronger-Engagement hypothesis 2)

3) A Fake News is shared more times that the True News (Stronger-Engagement hypothesis 3)

List of features:

1) **Number of unique users involved in spreading each Fake or True News (related to Hypothesis 1)**

<u>Descriptive analysis and statistical tests</u>

By looking at the below bar plot we can observe that more unique users spread Fake News (1) compared to True News (0).

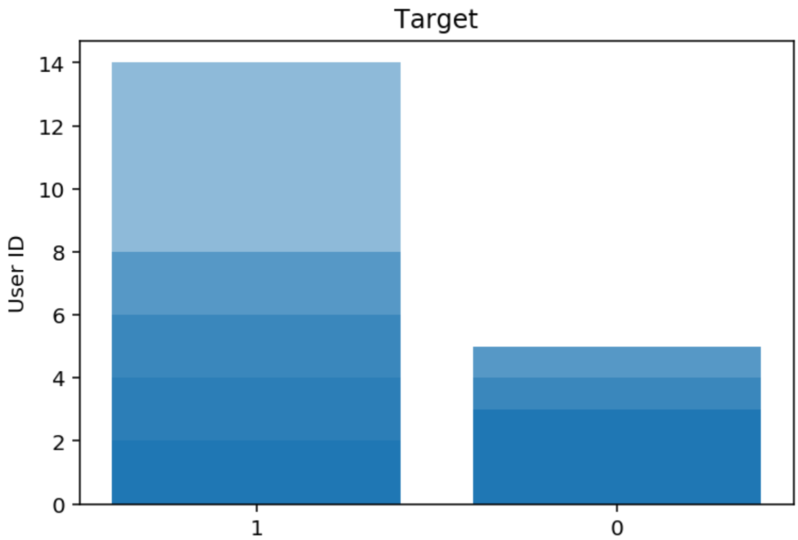

Figure 17: Bar plot Unique spreaders per News for Fake News (1) and True News (0)

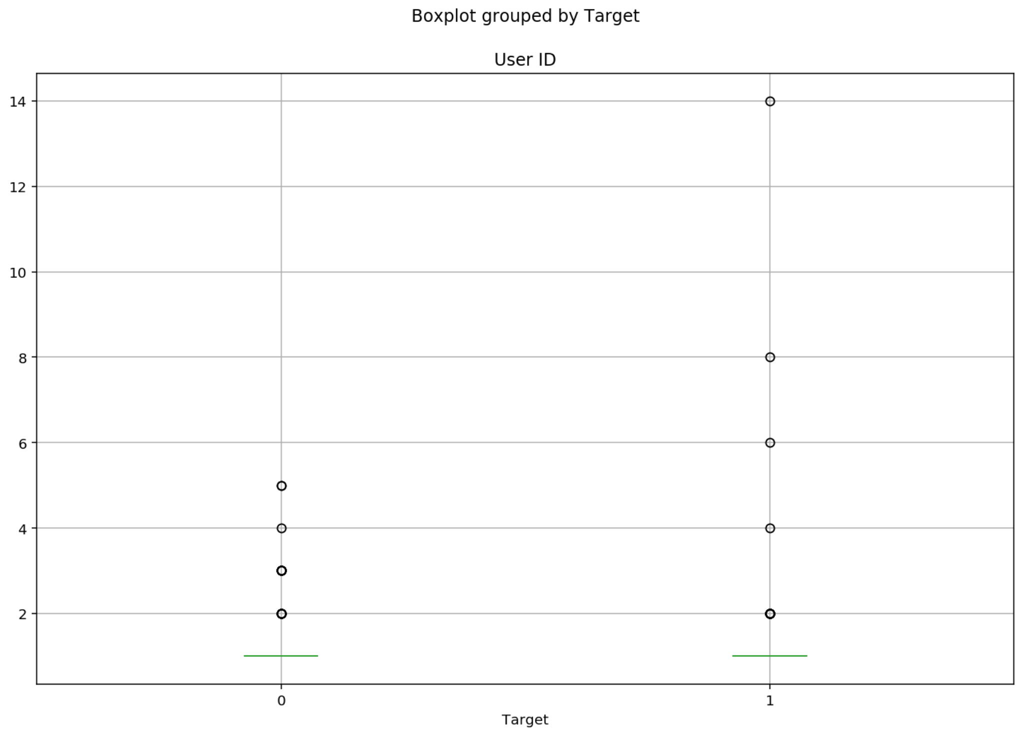

Figure 18: Boxplot Unique spreaders per News for Fake News (1) and True News (0)

Statistical test – Student's t-test

The null hypothesis here is that there is no significant difference in the mean test scores of the two sample groups (Fake News and True News) and that any difference down to chance. As the p-value is > 5%, I don't reject that hypothesis. Therefore, the number of unique users involved in spreading News doesn't have a significant relationship with the variable Target.

```
Ttest_indResult(statistic=-0.25587551464690245, pvalue=0.7983223213995143)
```

2) Individual Engagements (average spreading frequencies of all users who have participated in the news propagation) (related to Hypothesis 2)

Descriptive analysis and statistical tests

By looking at the below plot we cannot see clearly whether the average spreads per User is different for Fake News (1) and True News (0). A statistical test below will show whether there is a significant relationship between the two variables.

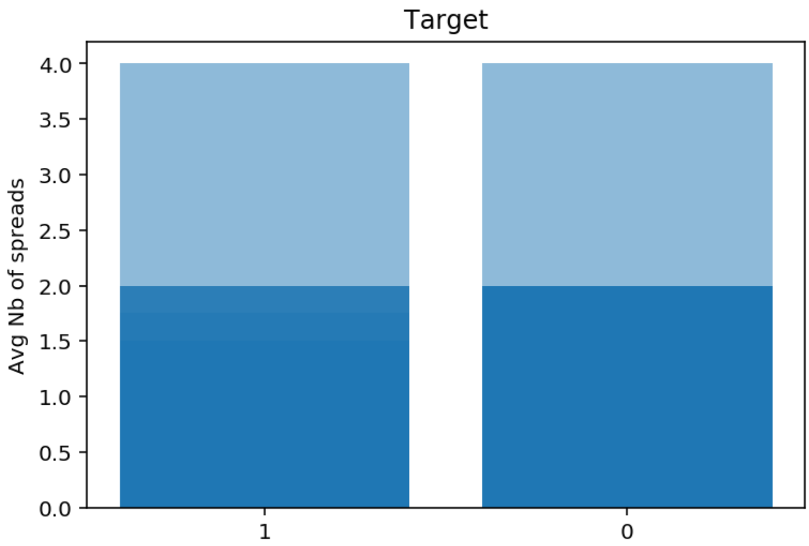

Figure 19: Bar plot Average number of spreads per User for Fake News (1) and True News (0)

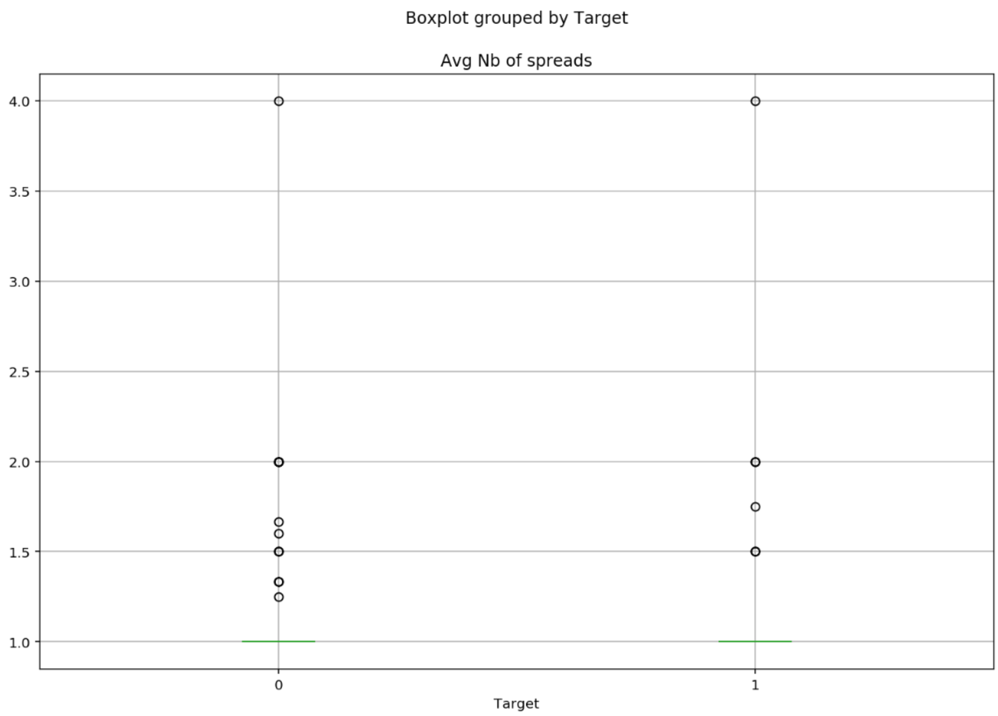

Figure 20: Boxplot Average number of spreads per User for Fake News (1) and True News (0)

<u>Statistical test – Student's t-test</u>

The null hypothesis here is that there is no significant difference in the mean test scores of the two sample groups (Fake News and True News) and that any difference down to chance. As the p-value is > 5%, I don't reject that hypothesis. Therefore, the average number of spreads per user doesn't have a significant relationship with the variable Target.

```
Ttest_indResult(statistic=0.9202193826474958, pvalue=0.35861888513543827)
```

3) Group Engagements (total number of times that the news story has been spread) (related to Hypothesis 3)

<u>Descriptive analysis and statistical tests</u>

By looking at the below plot we can see that Fake News (1) propagates more compared to True News (0).

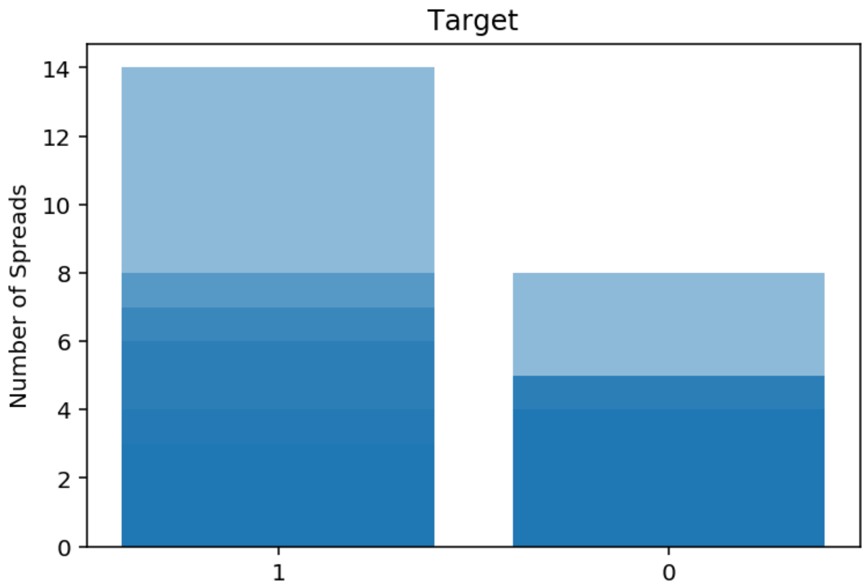

Figure 21: Bar plot Total number of spreads for Fake News (1) and True News (0)

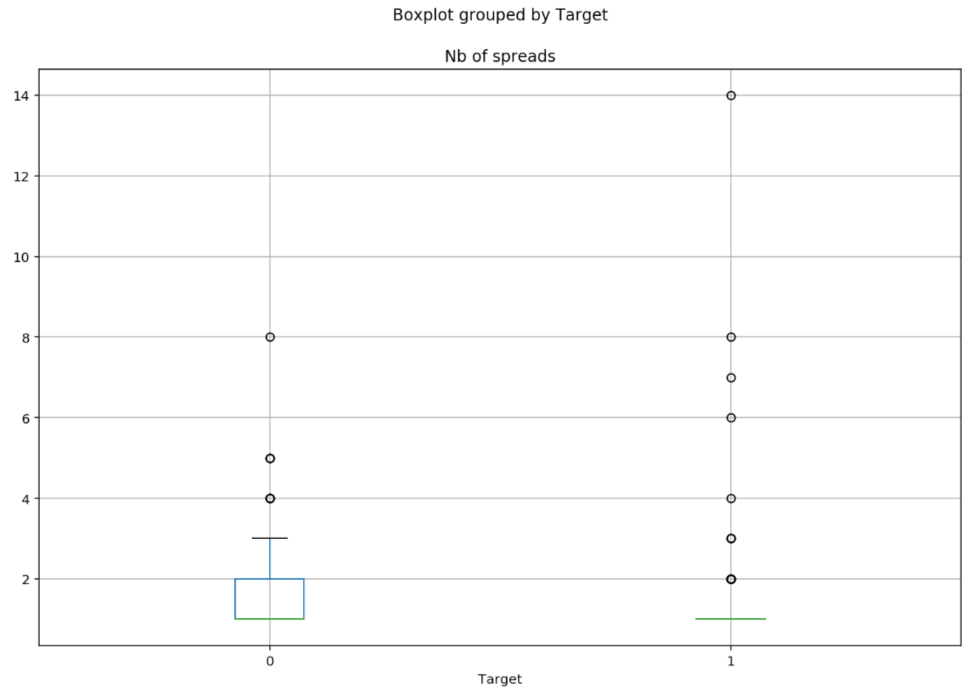

Figure 22: Boxplot Total number of spreads for Fake News (1) and True News (0)

<u>Statistical test – Student's t-test</u>

The null hypothesis here is that there is no significant difference in the mean test scores of the two sample groups (Fake News and True News) and that any difference down to chance. As the p-value is > 5%, I don't reject that hypothesis. Therefore, the total number of times a News is spread doesn't have a significant relationship with the variable Target.

```
Ttest_indResult(statistic=0.22105342557882543, pvalue=0.8252869668659539)
```

3.3 News Content features

News content features describe a piece of news with meta information. As also explained in chapter two, there are some major categories of these features: source information (author or publisher) of the news, body text (Main text) of the News, Image/Video (part of the body content of a news story that has visual cues to frame the story).

I consider in my analysis the body text as the most important part of the News article that should be the core of the Fake News detection analysis. According to recent research[52] the body text part of the news article is, in reality, the one that contains the most valuable information that can contribute to predict if a News article is Fake or True.

To find linguistic variations between Fake and True News articles, I extract features on a linguistic level. More specifically, to the following sub-categories: lexical, syntactic, and psycholinguistic (sentiment analysis). These categories have been presented in the second chapter of my master thesis (Linguistic Features based Methods). Linguistic based-analysis is part of a more significant group, which is called Style-based methods.

The performance of machine learning models depends on an excellent deal on features design. Hence, I have extracted a broad range of features for which I will evaluate their importance in predicting Fake News.

First, I will present you the data preparation tasks that are necessary before extracting features from a text document. You can find the associated Python 3 code file in the .zip file that accompanies my manuscript.

3.3.2 Data Preparation

Removing useless punctuation

I remove all punctuation besides "?" and "!" that I will keep them at the beginning to calculate their frequency as a feature and evaluate if they contribute the Fake News detection.

Word Tokenization

Tokenization is the process of cutting the Body text into separate words and sentences. Word tokenization is the technique of separating into words a sample of text. That is a prerequisite for the natural language processing tasks where each word must be captured and further analysed to extract extra information. In addition to that, it is also provided as input for further text cleaning steps, such as Stemming. I use the function word_tokenize() of the NLTK package to split a sentence into words.

Sentence tokenization

Like the word tokenization, we can implement sentence tokenization. This information will be used to create features, such as average sentence length per News or the total number of sentences per News. To accomplish the above, we use PunktTrainer () and PunktSentenceTokenizer () of the nltk package for Python 3.

POS Tagging preparation

We apply the POS tagging technique to determine the grammatical category of each "word". To do so, I apply the pos_tag () function of the nltk package on the result of the tokenization.

I perform a POS tagging to determine the grammatical category of each "word" of the already created tokens. To do so, I apply the pos_tag() function of the nltk package on the result of the tokenization. These data will be used afterwards to extract further information and create new features, such as the frequency of each grammatical category within each News.

Below there is the explanation of some abbreviations:

- CC coordinating conjunction
- CD cardinal digit
- DT determiner
- EX existential there (like: “there is”… think of it like “there exists”)
- FW foreign word
- IN preposition/subordinating conjunction
- JJ adjective ‘big’
- JJR adjective, comparative ‘bigger’
- JJS adjective, superlative ‘biggest’
- LS list marker 1)
- MD modal could, will
- NN noun, singular ‘desk’

- NNS noun plural ‘desks’
- NNP proper noun, singular ‘Harrison’
- NNPS proper noun, plural ‘Americans’

The full list of the abbreviations and their explanation is here[111].

Stopwords removal

I remove frequent words that are not informative. These words are named “stop-words”. I use the stopwords.words('english') instruction of the nltk package to store the stop words into a list and then I proceed to their deletion in the tokenized text.

Stemming

The stemming[112] is a crucial step in pre-processing text data. Indeed, it avoids considering words of the same root as two different words by considering the root of each word. Here I use the Porter algorithm by implementing the PorterStemmer() function of the nltk package for Python 3.

The Stemming technique aims at converting the words into their base word or stem word. For example, “tastefully”, “tasty”, these words are converted to stem word called “tasti”). That reduces the vector dimension because we do not consider all similar words.

Another technique that can replace Stemming is the Lemmatization technique. Both techniques have the same purpose, which is to reduce the inflectional forms of each word into a common base or root.[113] For my analysis I implement the Stemming technique.

3.3.3 Style-Based Textual Features Engineering

In this chapter I explore some popular and effective methods for handling text data and extracting meaningful features from them.

As also explained within the second chapter, the Style-based perspective focuses on investigating the news content. This approach aims to assess news intention. Fake News style that derives from the Text Style will be outlined as a group of quantifiable features that may well represent Fake News and differentiate Fake News from True News.[21]

3.3.3.1 Lexical analysis

Lexical features include character-level and word-level features. In this section I state the different hypothesis I have done regarding the lexical characteristics of Fake News and True News articles. I will the provide a list of variables/features that enables me to test my hypotheses. Moreover, I give some descriptive statistics using charts and I realize statistical tests to have a first insight on the significant effect of the variables chosen for my analysis.

List of Hypothesis

1) The number of words in a News article body text varies significantly between Fake and True News.
2) The number of unique words in a News article body text varies significantly between Fake and True News.
3) The appearance of words in a News article body text varies significantly between Fake and True News.
4) The appearance of the combination of words in a News article body text varies significantly between Fake and True News.
5) The importance of words in a News article body text varies significantly between Fake and True News.

List of features

1) **Number of words per News (related to hypothesis 1)**

<u>Descriptive analysis</u>

By looking at the below chart I can see that True News contains more words that Fake News.

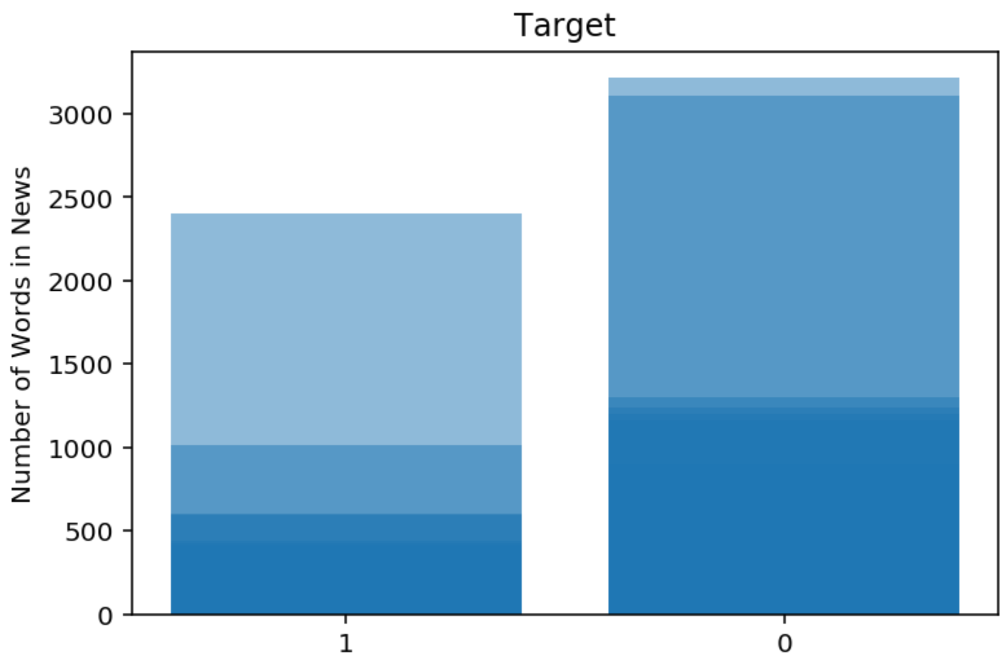

Figure 23: Bar plot Number of Words for Fake News (1) and True News (0)

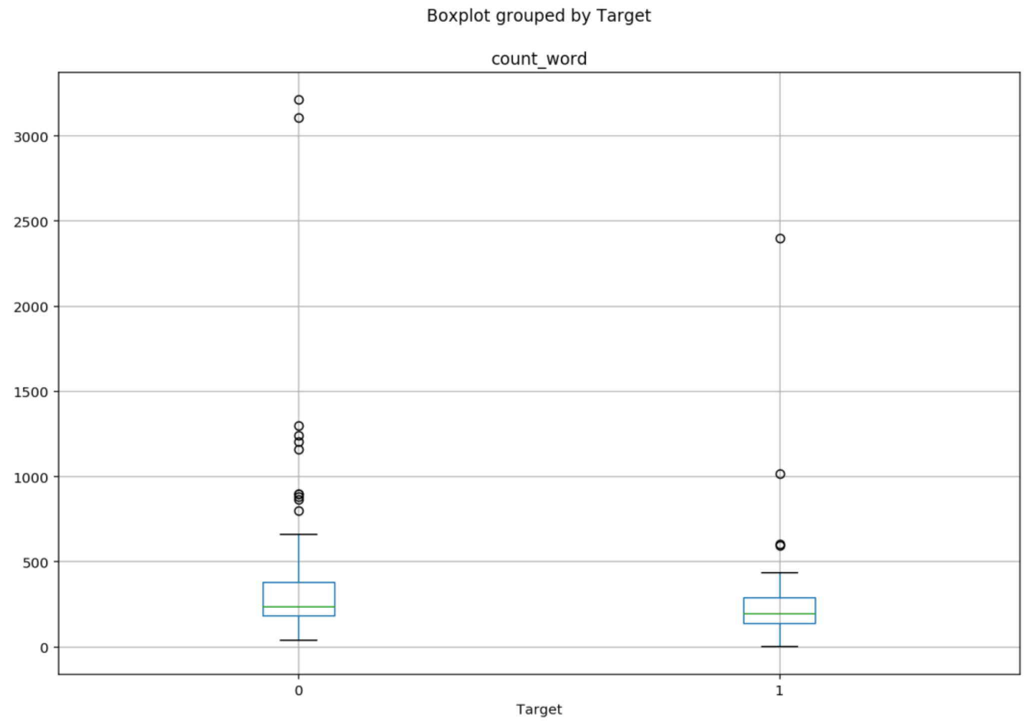

Figure 24: Boxplot Number of Words for Fake News (1) and True News (0)

Statistical test - Student's t-test

The null hypothesis here is that there is no significant difference in the mean test scores of the two sample groups (Fake News and True News) and that any difference down to chance. As the p-value is < 5%, I reject that hypothesis. Therefore, the total number of the words in a News article have a significant relationship with the variable Target which is the response variable for Fake (1) or True (0) News.

```
Ttest_indResult(statistic=2.5463499339509976, pvalue=0.011674465802643806)
```

2) Number of unique words per News (related to hypothesis 2)

By looking at the below diagrams I can observe that True News articles include more unique words than the Fake News.

Descriptive analysis

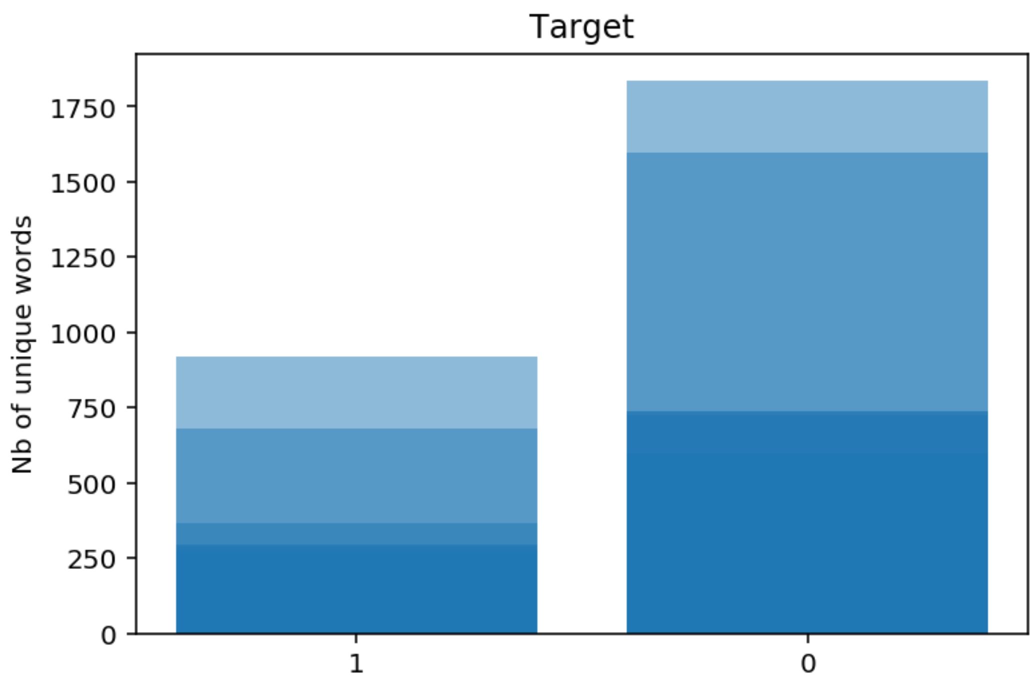

Figure 25: Bar plot Number of Unique Words for Fake News (1) and True News (0)

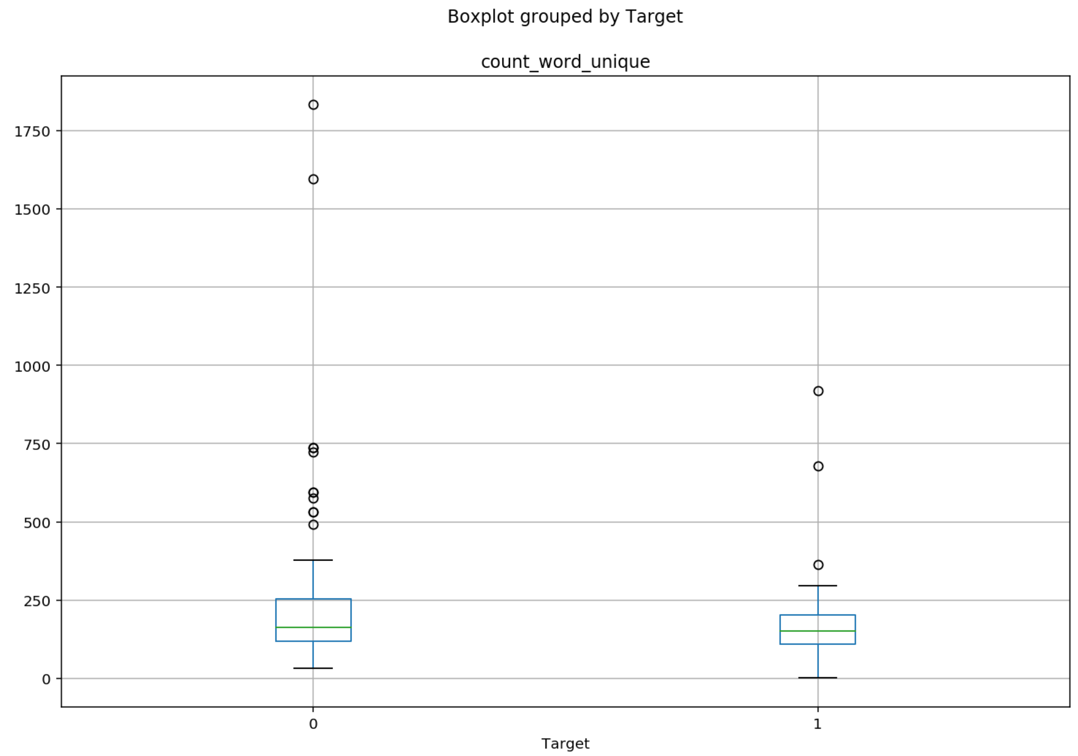

Figure 26: Boxplot Number of Unique Words for Fake News (1) and True News (0)

<u>Statistical test - Student’s t-test</u>

The null hypothesis here is that there is no significant difference in the mean test scores of the two sample groups (Fake News and True News) and that any difference down to chance. As the p-value is < 5%, I reject that hypothesis. Therefore, the number of unique words in a News article have a significant relationship with the variable Target which is the response variable for Fake (1) or True (0) News.

```
Ttest_indResult(statistic=2.6037560687530115, pvalue=0.009945496724163807)
```

Bag of words approaches for features extraction:

3) CountVectorizer features extraction including unigrams and bigrams (related to hypothesis 3,4)

In this section I evaluate the need to use both CountVectorizer and TfidfTransformer features, as well as, selecting the number ngrams that I will consider for my analysis. The result will be different for Naïve Bayes classifier and Random Forest classifier. Moreover, I select the fifty most significant features of CountVectorizer and the the fifty most significant features of TfidfTransformer.

CountVectorizer function of the sklearn package offers an easy way of both tokenizing a collection of texts and building a vocabulary of known words, but also encoding new data using that vocabulary.

I use it as follows:

- I create a CountVectorizer class instance.
- To learn/train a vocabulary from one or more articles, I use the fit() function.
- I call the transform () function to encode each as a vector on one or more documents.

<u>Naive Bayes classifier</u>

In this part, I define parameters for which I will perform tuning. The parameter names must begin with the classifier name. With “vect__ngram_range” I am telling to use unigram and bigrams and choose the optimal one. With “tfidf__use_idf” I test if it is worth proceeding with tfidf features extraction.

The results showed that the accuracy is ~ 74.0% for the Naïve Bayes classifier, and the corresponding parameters are {'tfidfuse_idf': True, 'vectngram_range': (1, 1)}. Hence, that means for Naïve Bayes model, we will use only unigrams, so the fourth hypothesis stated previously “*The appearance of the combination of words in a text varies significantly between Fake and True News*” is rejected when we use the Naive Bayes model. Also, according to the above results, we can conclude that it is meaningful to use the TfidfVectorizer method for feature extraction along with Naïve Bayes classifier, as it will bring more accuracy to the model.

Below you can find a chart that shows the frequency of the fifty most frequent words (unigrams) in our dataset, followed by charts that show the most frequent words in Fake News and True News.

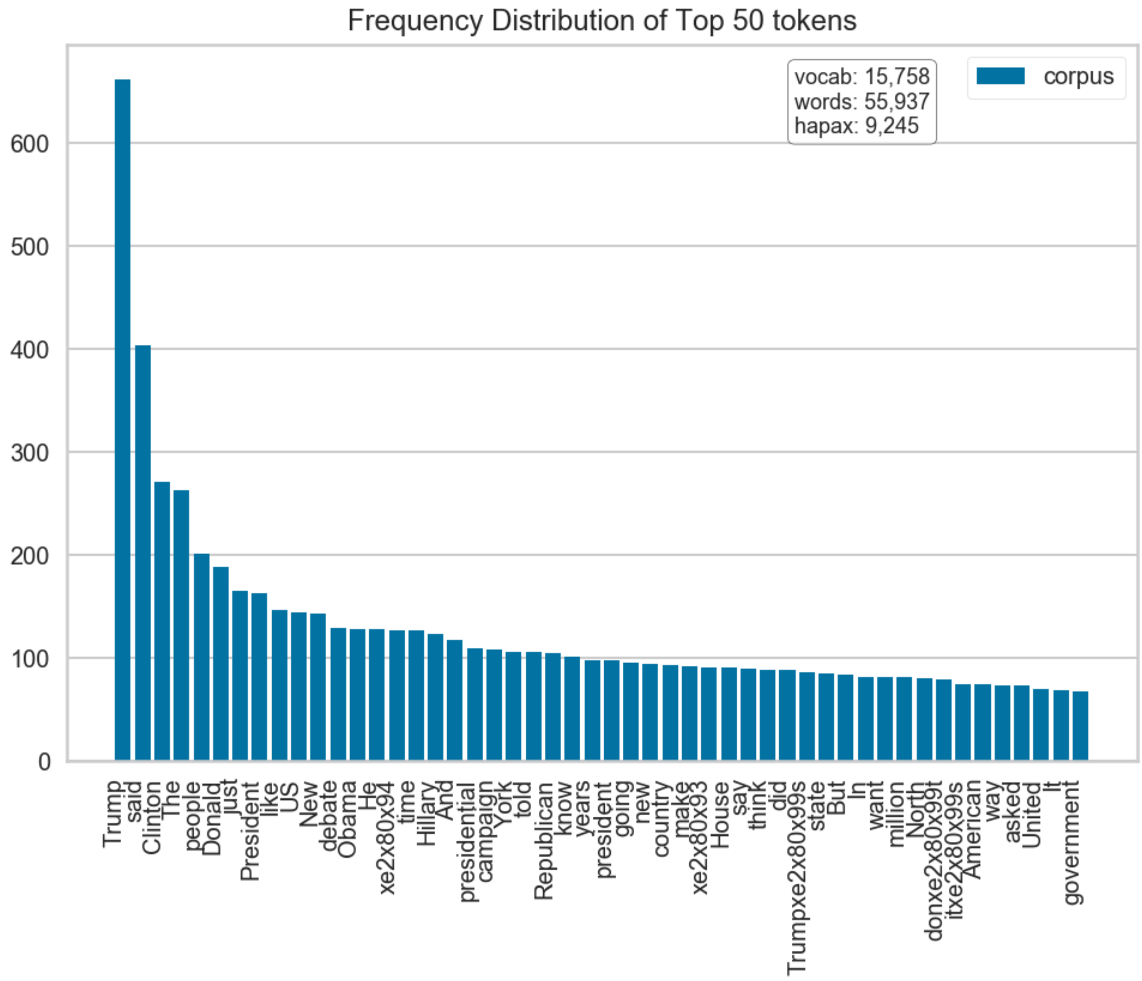

Figure 27: Bar plot – frequency distribution of top 50 words – only unigrams

True News:

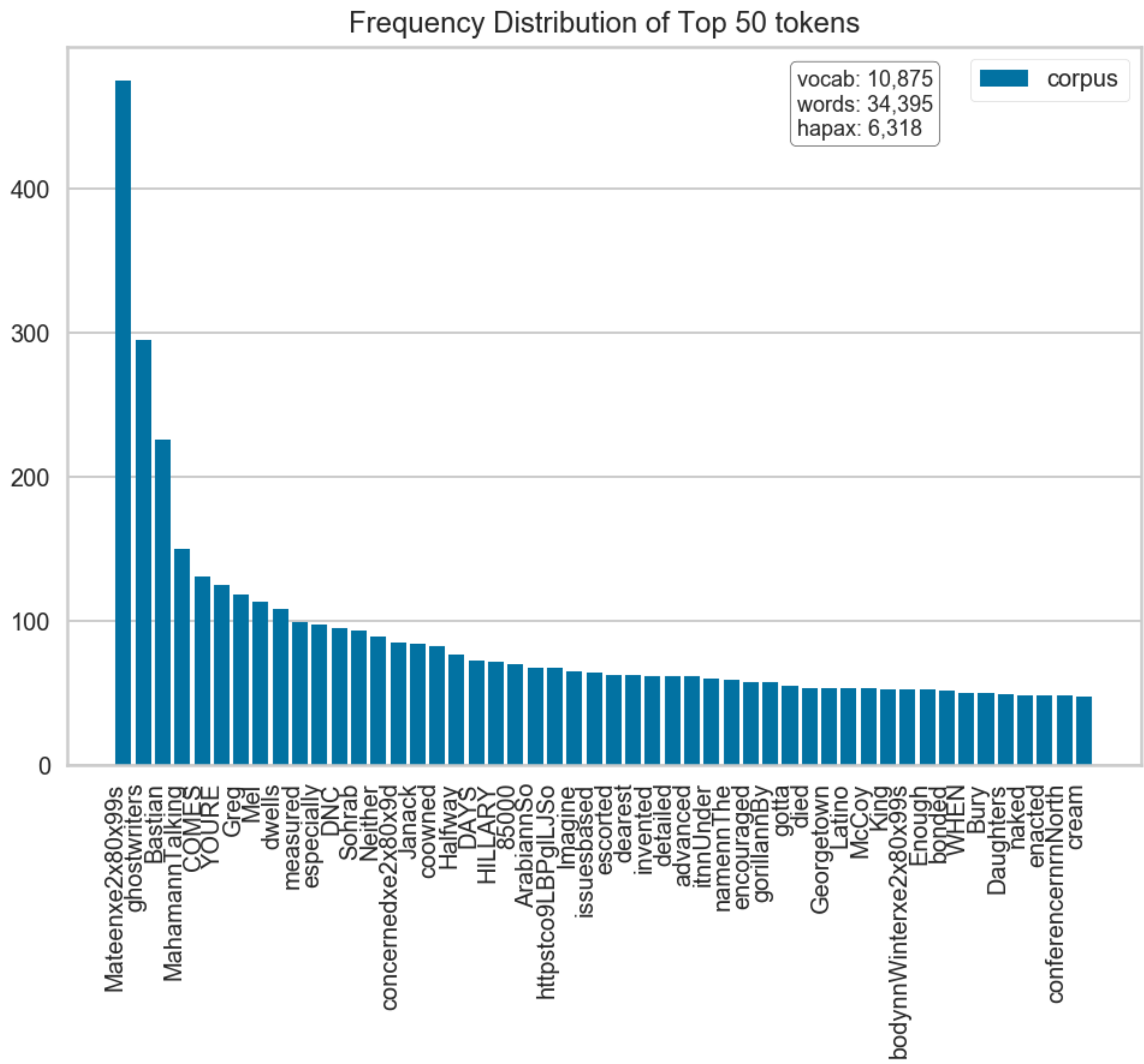

Figure 28: Bar plot – frequency distribution of top 50 words in True News – only unigrams

Fake News:

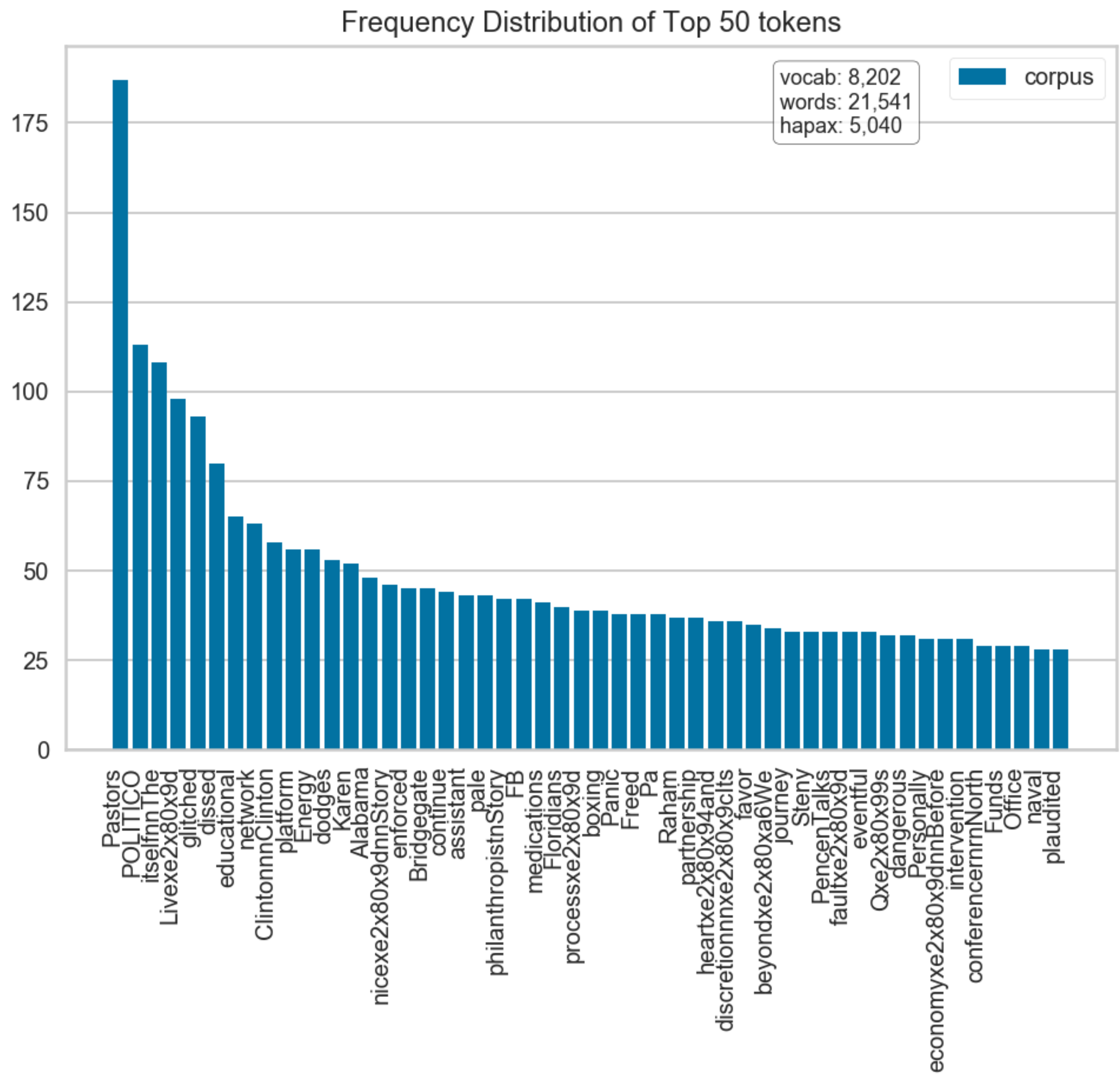

Figure 29: Bar plot – frequency distribution of top 50 words in Fake News – only unigrams

<u>Random Forest classifier</u>

Concerning Random Forest classifier, I am proceeding the same way as I did for Naive Bayes in order to evaluate the need for using the TfidfVectorizer method and decide the number of ngrams I will be using for my analysis. The results show us that the accuracy percentage is approximately 64.2% for the Random Forest classifier, and the corresponding parameters are {'tfidfuse_idf': True, 'vectngram_range': (1, 2)}. Hence, that means for the Random Forest classifier, we will use both unigrams and bigrams, so the fourth hypothesis stated previously "*the appearance of the combination of words in a text varies significantly between Fake and True News*" is accepted when we use the Random Forest classifier. In addition, according to the result, we should be using tfidf vectoriser.

Below you can find a chart that shows the frequency of the fifty most frequent words (unigrams & Bigrams) in our dataset, followed by charts that show the most frequent words in Fake News and True News.

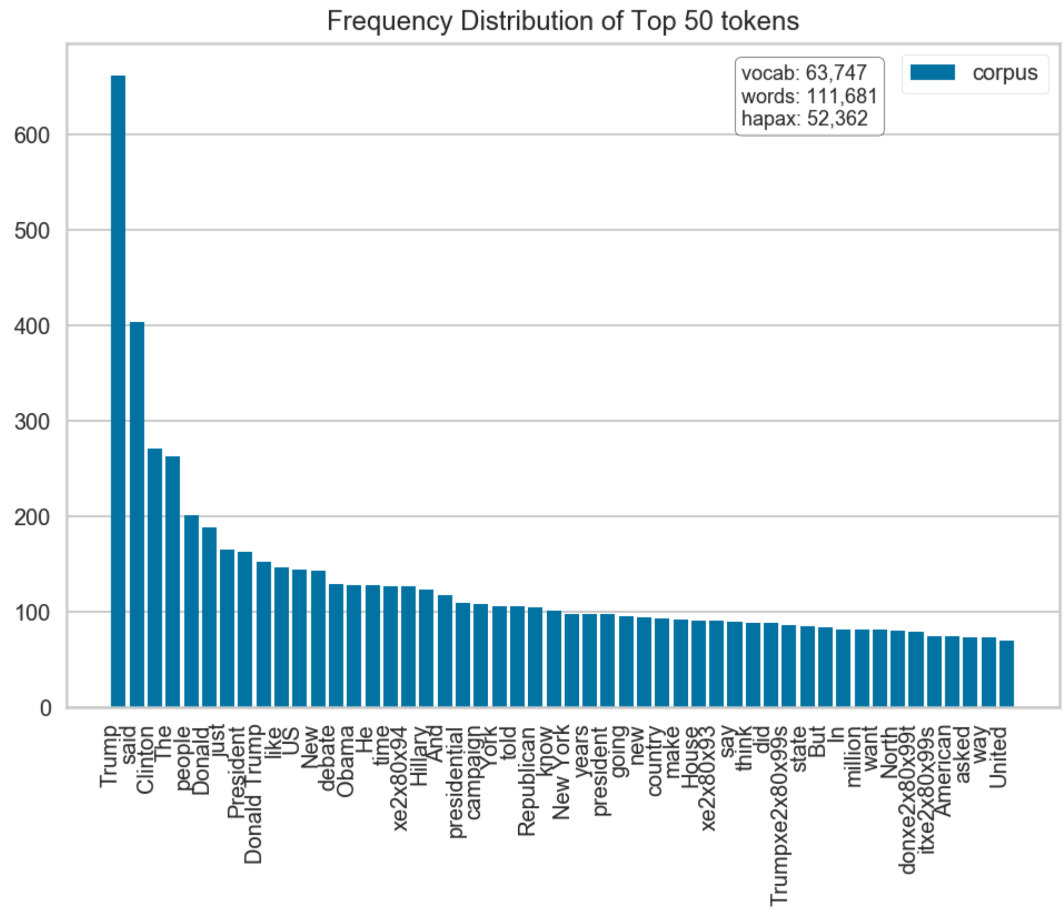

Figure 30: Bar plot – frequency distribution of top 50 words – unigrams & bigrams

True News:

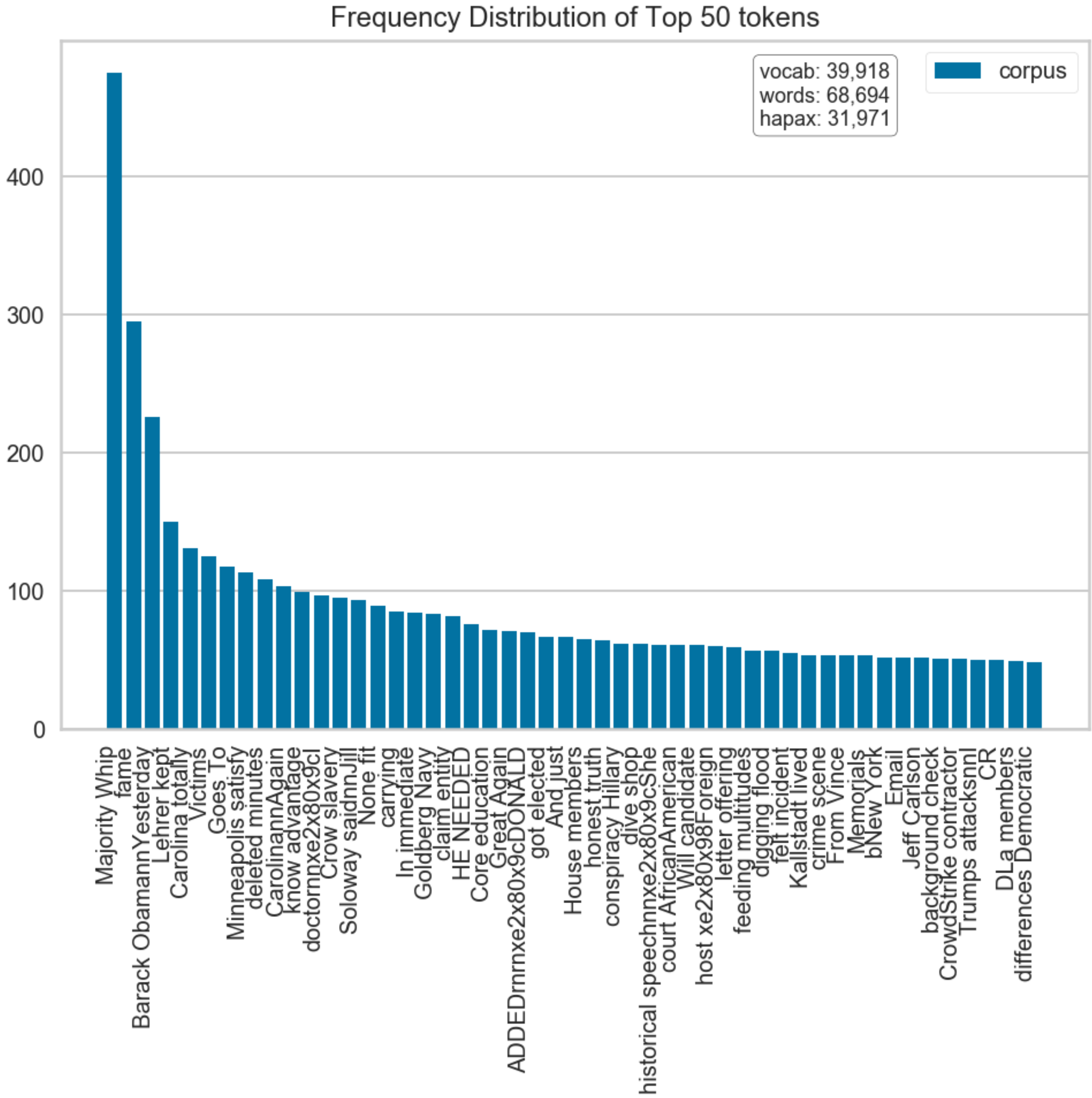

Figure 31: Bar plot – frequency distribution of top 50 words for True News – unigrams & bigrams

Fake News:

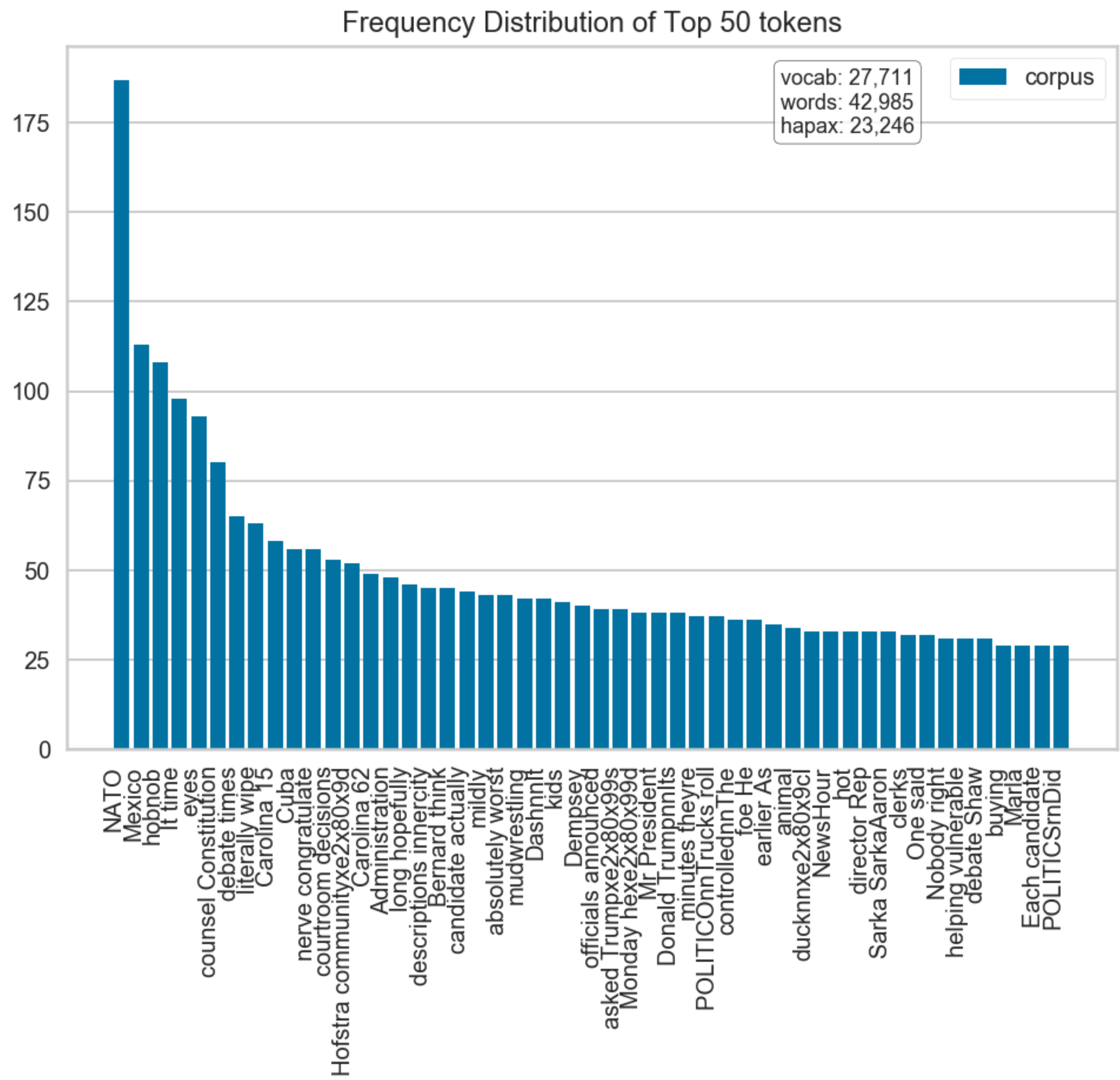

Figure 32: Bar plot – frequency distribution of top 50 words for Fake News – unigrams & bigrams

4) TfidfVectorizer features extraction (related to hypothesis 5)

According to the results and decisions taken in the previous section, we should be using Tfidf features in both Naïve Bayes and Random Forest classifiers. Thus, the fifth hypothesis is validated "*The importance of words in a News article body text varies significantly between Fake and True News*."

TF-IDF is one of the most broadly and commonly used techniques to process textual data in order to retrieve valuable information. According to Wikipedia, "*In information retrieval, tf–idf or TFIDF, short for term frequency–inverse document frequency, is a numerical statistic that is intended to reflect how important a word is to a document in a collection or corpus*" [114].

In more detail, term Frequency (tf), gives the frequency of the word in each news article. It represents the ratio of number of times the word occurs in the text compared to the total number of words. Thus, it becomes greater while the number of occurrences of the word within the text getting higher. The Inverse Data Frequency (idf) aims to measure the weight of words across all news articles/texts/documents in the corpus. The words that do not occur often (they occur rarely) in the corpus have a high IDF score. [115] Combining the above two metrics, we obtain the TF-IDF score for each word in the text.

I consider that combining CountVectorizer results with Tf-idfVectorizer function results, I can bring more accuracy and precision to the predictions of my model. On that purpose I will be using the TfidfVectorizer() function of the sklearn package that allows first to tokenize, then count tokens, then transform the raw counts to TF-IDF values.

Naive Bayes classifier

Below you can find a chart that shows the frequency of the fifty most important words (unigrams) in our dataset, followed by charts that show the most important words in Fake News and True News.

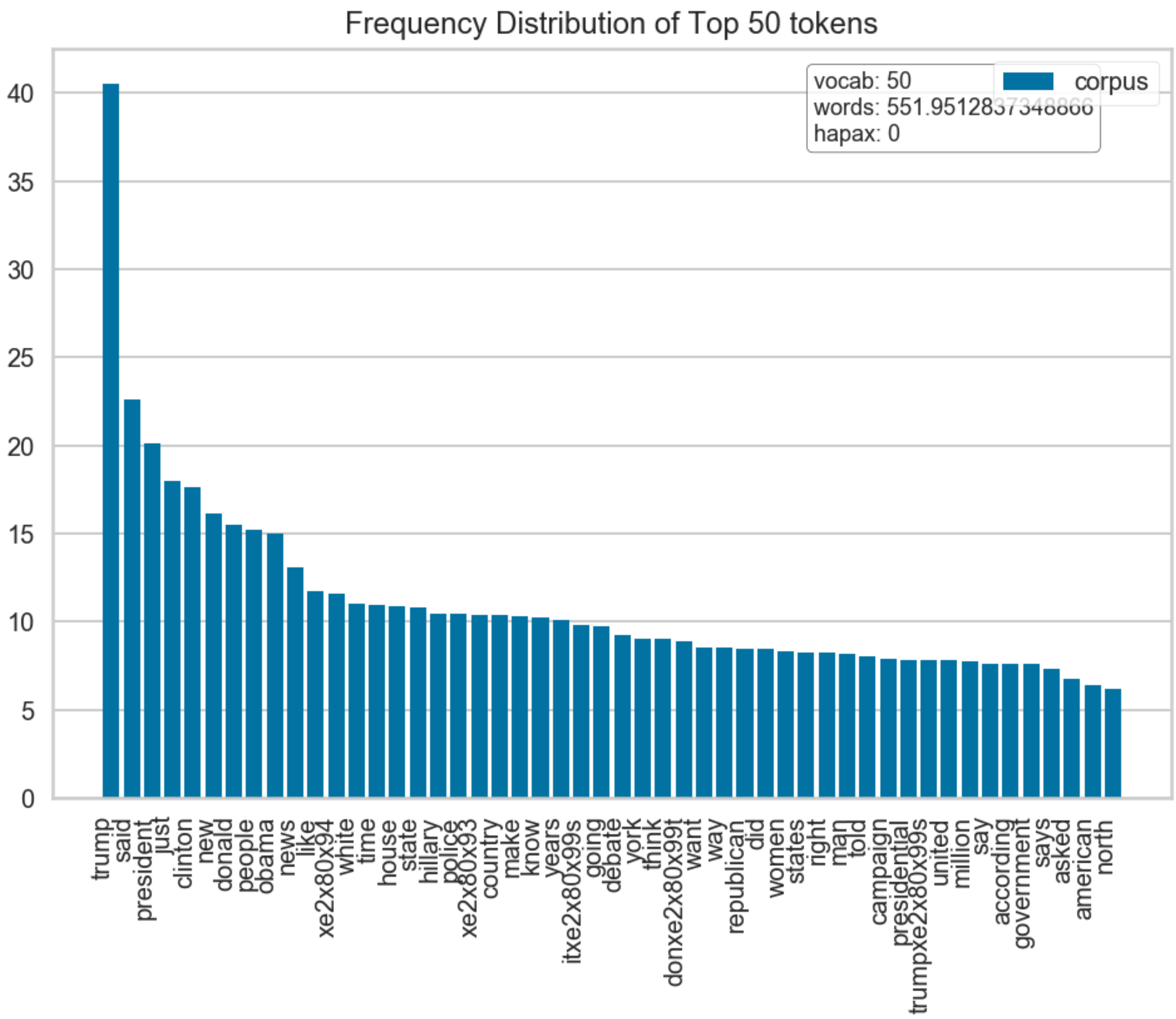

Figure 33: Bar plot – frequency distribution of top 50 words – unigrams - tfidf

True News:

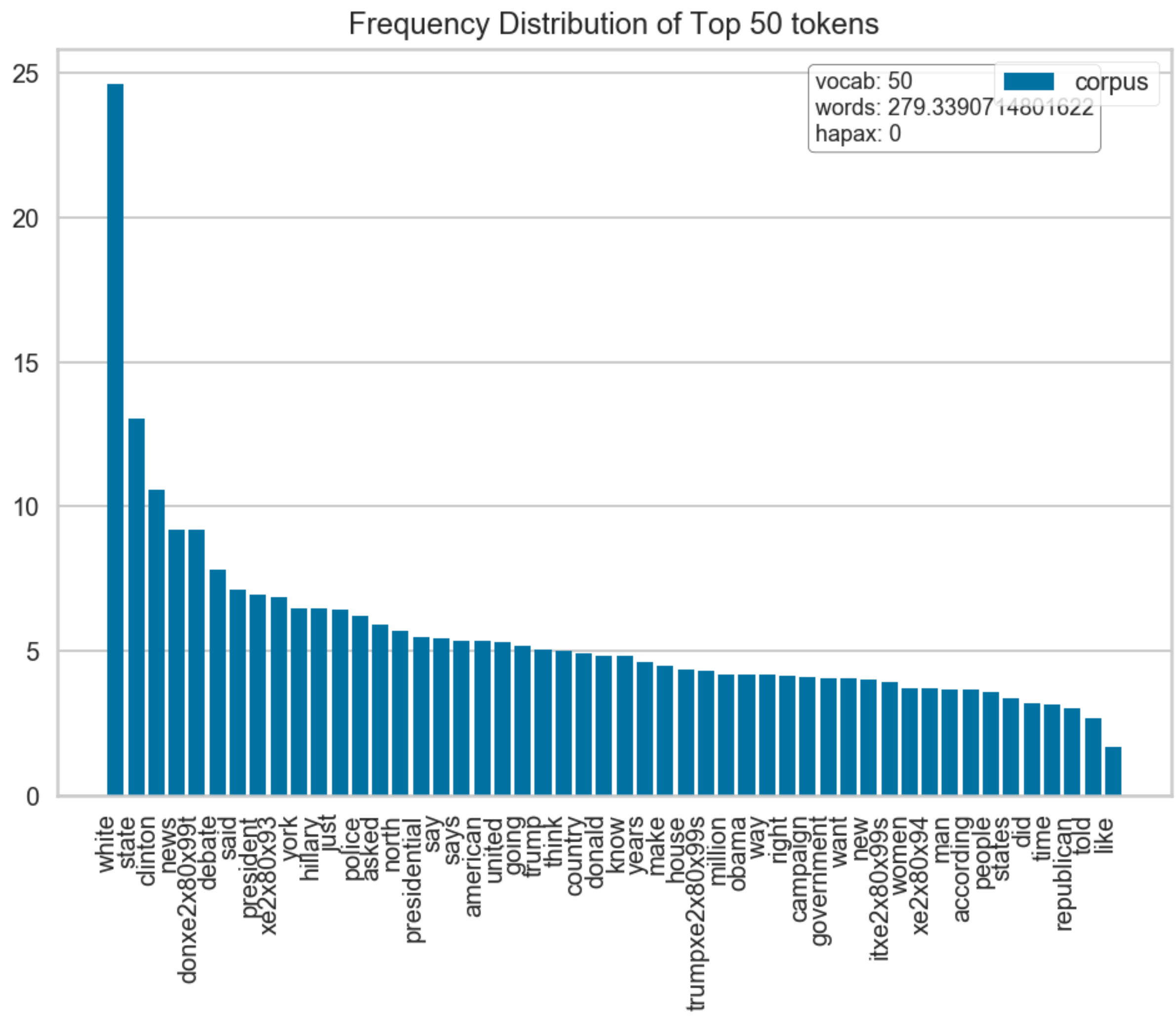

Figure 34: Bar plot – frequency distribution of top 50 words in True News – unigrams - tfidf

Fake News:

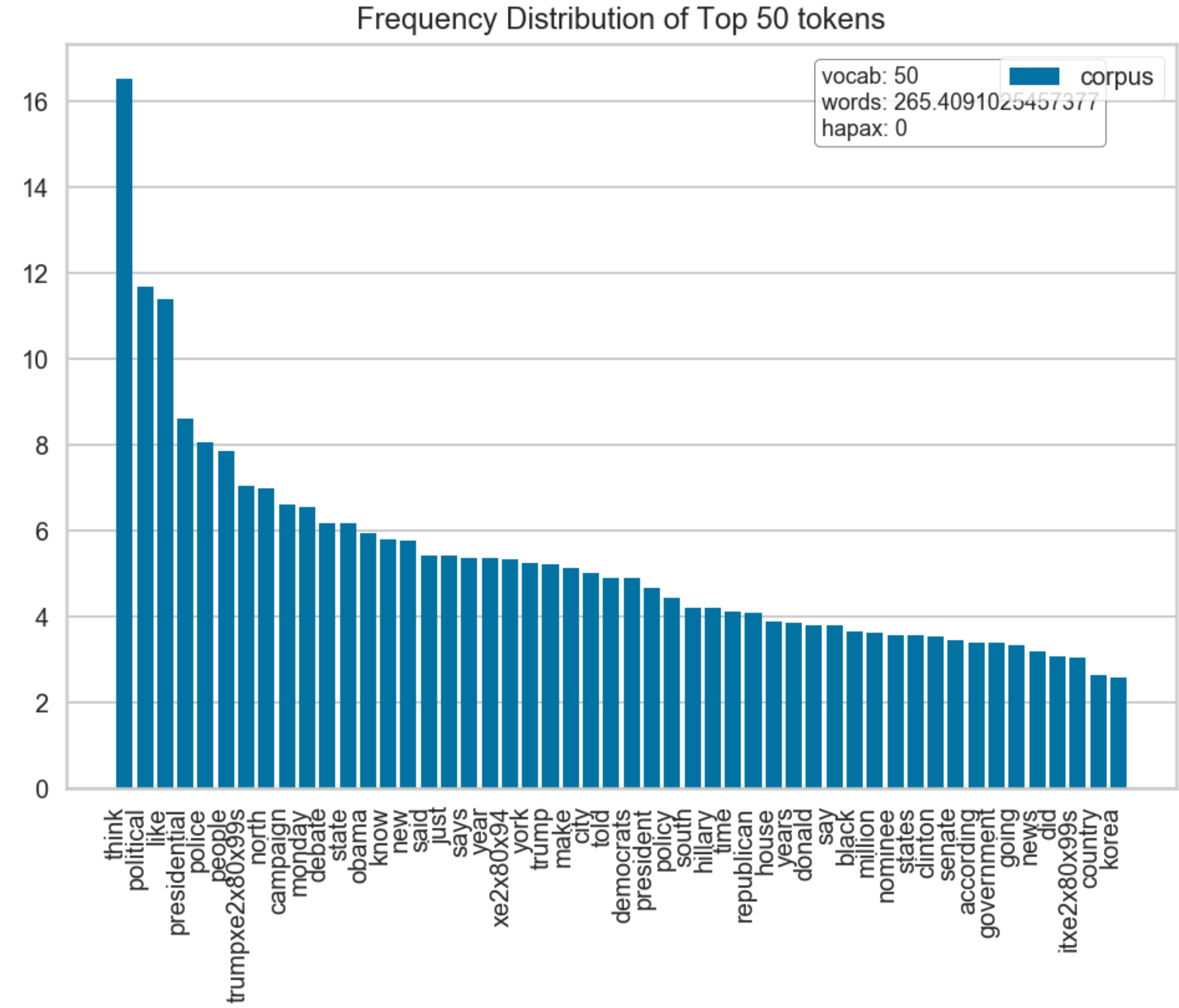

Figure 35: Bar plot – frequency distribution of top 50 words in Fake News – unigrams - tfidf

Random Forest classifier

Below you can find a chart that shows the frequency of the fifty most important words (unigrams & bigrams) in our dataset, followed by charts that show the most important words in Fake News and True News.

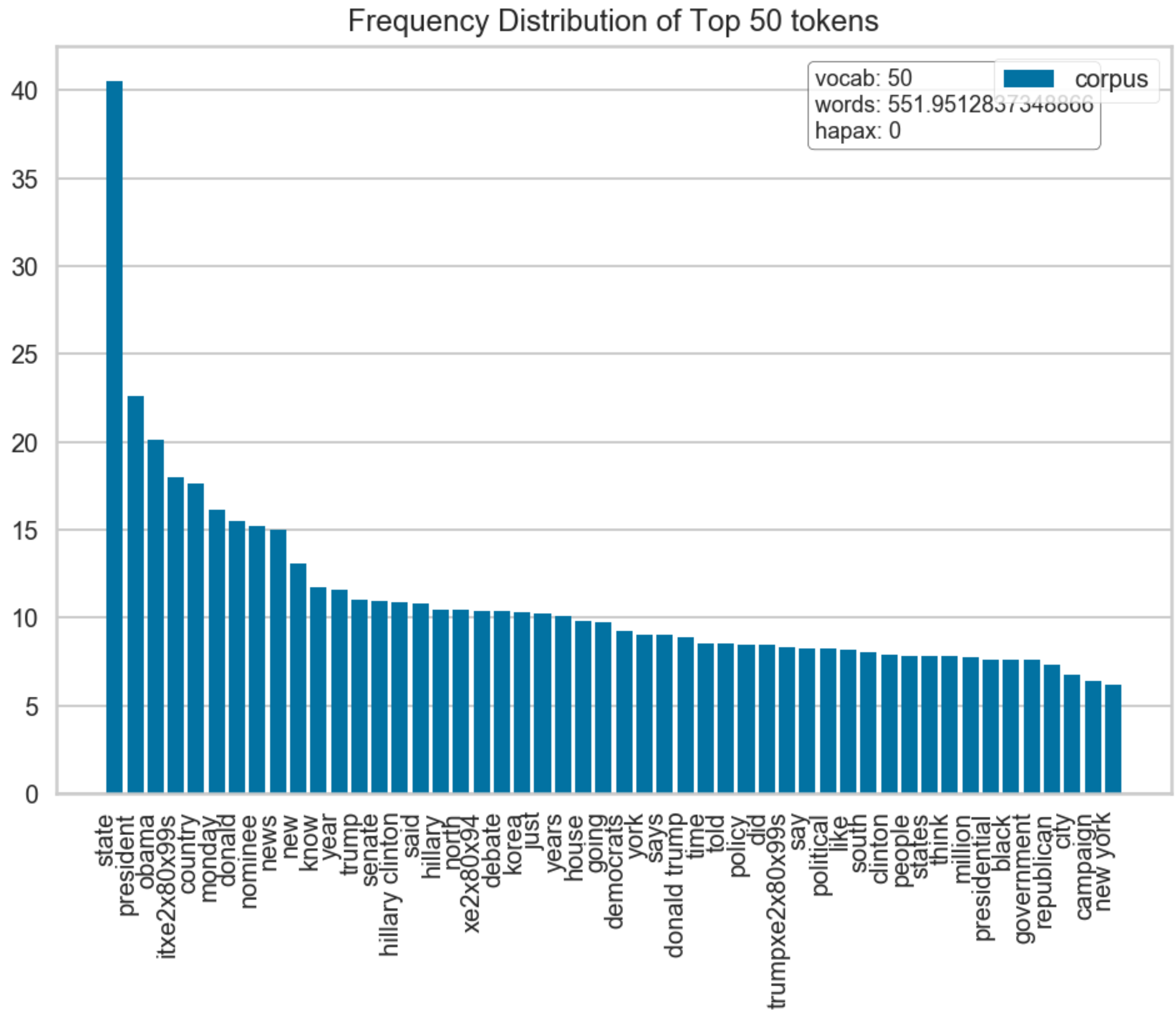

Figure 36: Bar plot – frequency distribution of top 50 words – unigrams & bigrams - tfidf

True News:

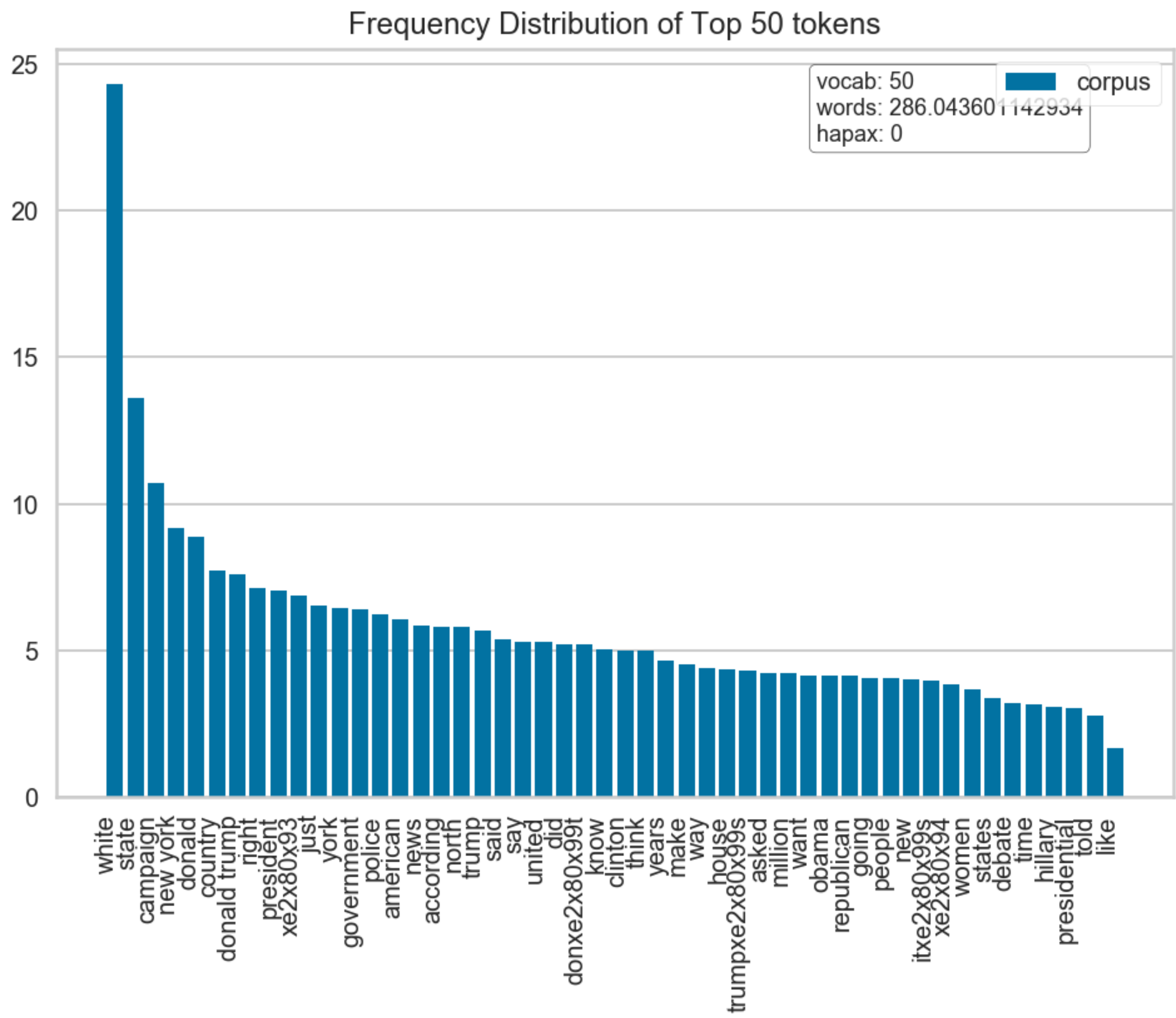

Figure 37: Bar plot – frequency distribution of top 50 words in True News – unigrams & bigrams - tfidf

Fake News:

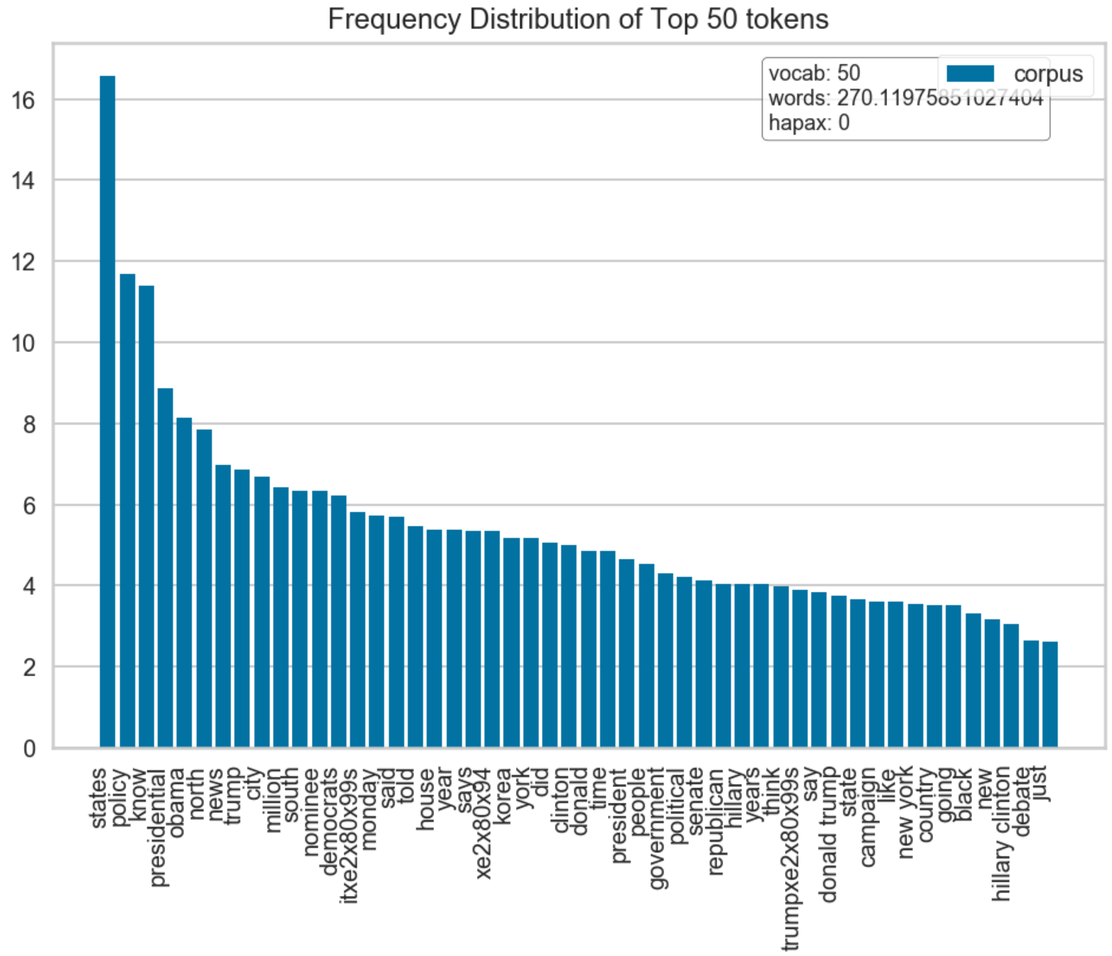

Figure 38: Bar plot – frequency distribution of top 50 words in Fake News – unigrams & bigrams - tfidf

3.3.3.2 Syntactic analysis

As explained in the second chapter, in order to understand the syntax, text style, and grammatical components of each article content, I need to extract stylistic/syntactic features with the use of Natural Language Processing techniques.

Before proceeding to the analysis concerning the importance of the syntactic features, as first step I have to realize several pre-processing tasks. I will be using the Python Natural Language Toolkit (Bird 2006) and Part of Speech (POS) tagger in order to keep a count of how many times each tag appears in the article. In addition to that, I will keep track of the number of punctuations ("?" & "!"), the lexical diversity, the number of sentences, the average length of the sentences in each news article text.

Below you can find the list of the hypothesis I have done and inspired me to decide the syntactic features that I will include in my analysis.

List of hypotheses

1) The number of sentences per news varies significantly between Fake and True News
2) The average length of a sentence per news varies significantly between Fake and True News
3) The frequencies of the different grammatical categories in a News varies significantly between Fake & True News
4) The presence of "?" and "!" punctuations varies significantly between Fake and True News. More punctuations likely show that there is a higher probability that a News is Fake.
5) Lexical diversity varies significantly between Fake and True News.

List of features

1) **Number of sentences per news (related to hypothesis 1):** I am counting the number of sentences per news article by using the already tokenized text in sentences.

By looking at the below charts we can see that there slightly more sentences in Fake News than True News.

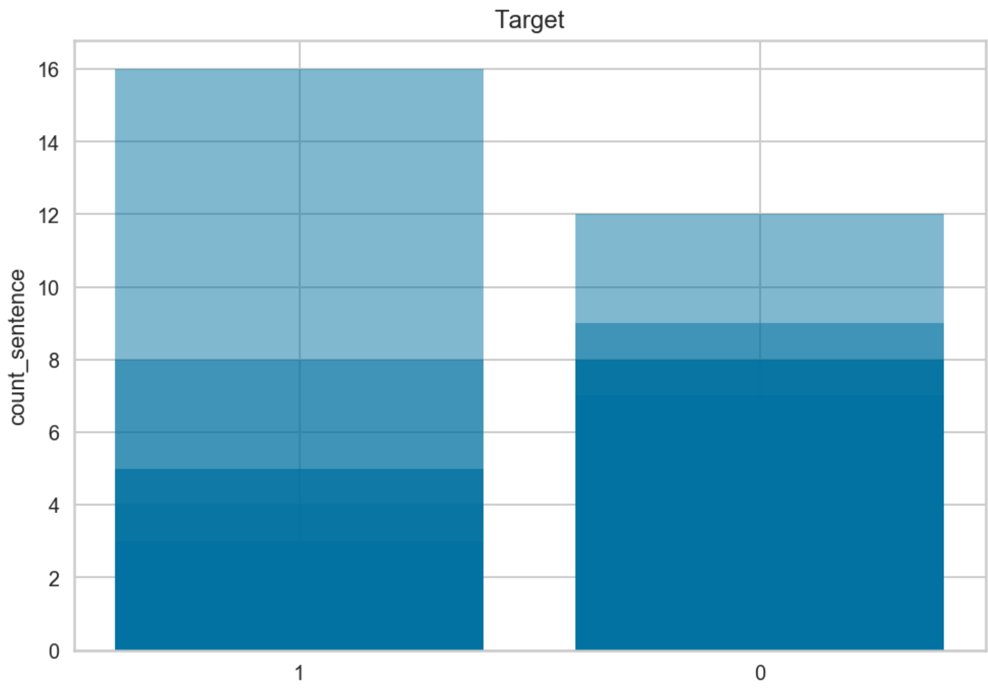

Figure 39: Bar plot – Number of sentences per Fake (1) and True (0) News

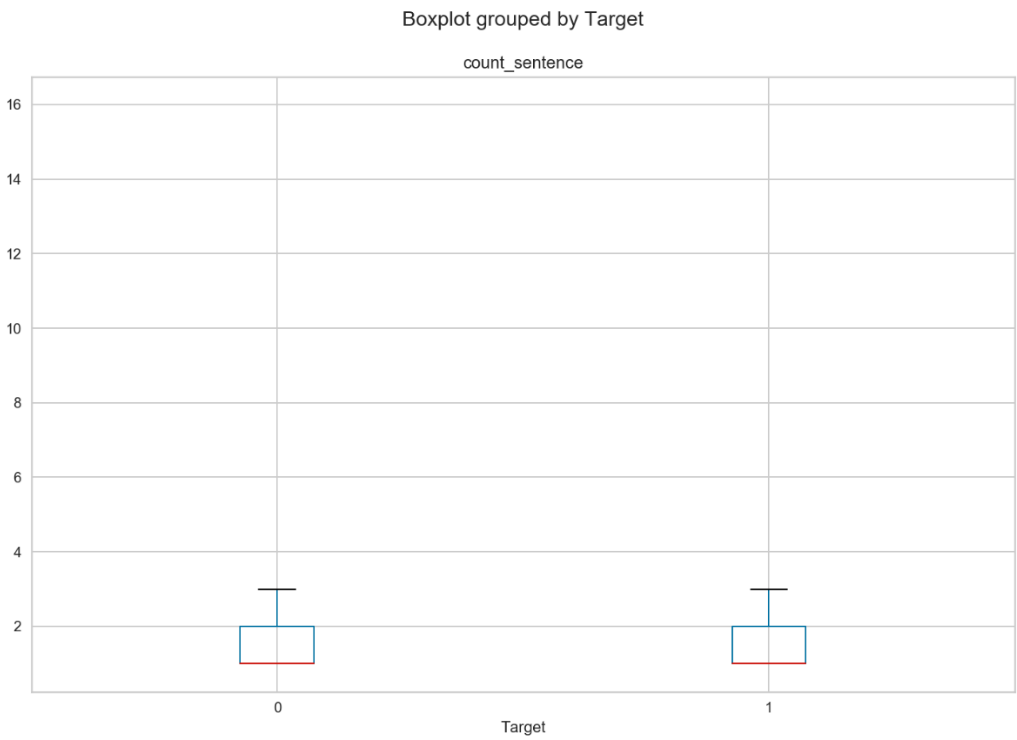

Figure 40: Boxplot – Number of sentences per Fake (1) and True (0) News

<u>Statistical Test – Student's t-test</u>

The null hypothesis here is that there is no significant difference in the mean test scores of the two sample groups (Fake News and True News) and that any difference down to chance. As the p-value is > 5%, I don't reject that hypothesis. Therefore, the total number of sentences in a News article doesn't have a significant relationship with the variable Target which is the response variable for Fake (1) or True (0) News.

```
Ttest_indResult(statistic=1.2035528803490185, pvalue=0.23025200803791124)
```

2) **Average length of sentence per news (related to hypothesis 2):** I am calculating the average length of sentences per news article by using the already tokenized text in sentences. The length is calculated based on the number of letters.

By looking at the below charts, I can see that True News have a higher average length of sentence compared to Fake News.

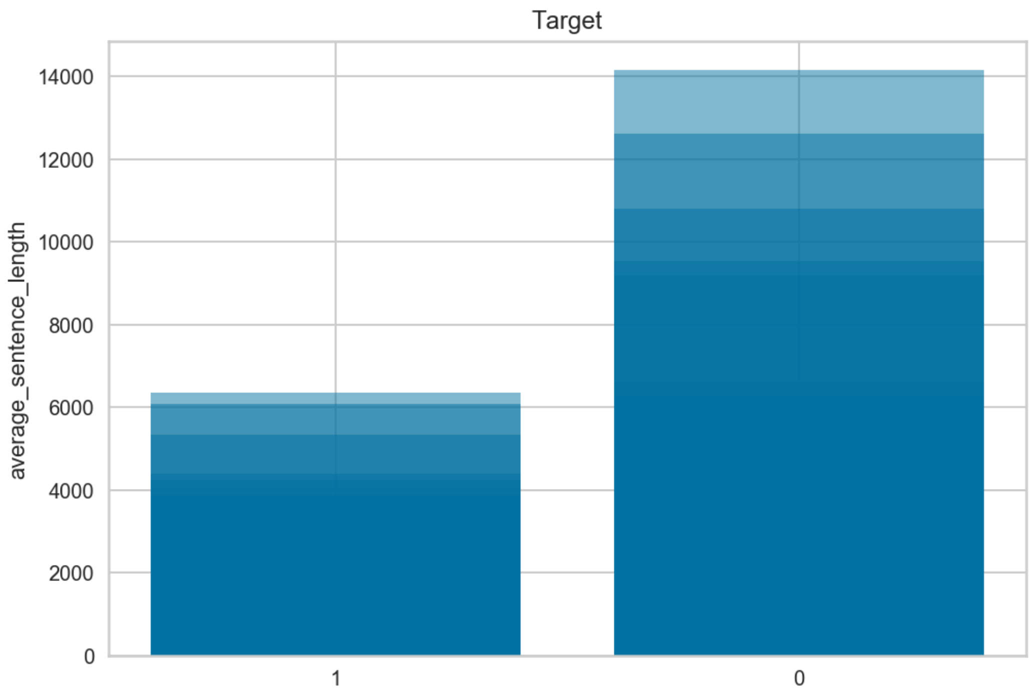

Figure 41: Bar plot – Average sentence length for Fake (1) and True (0) News

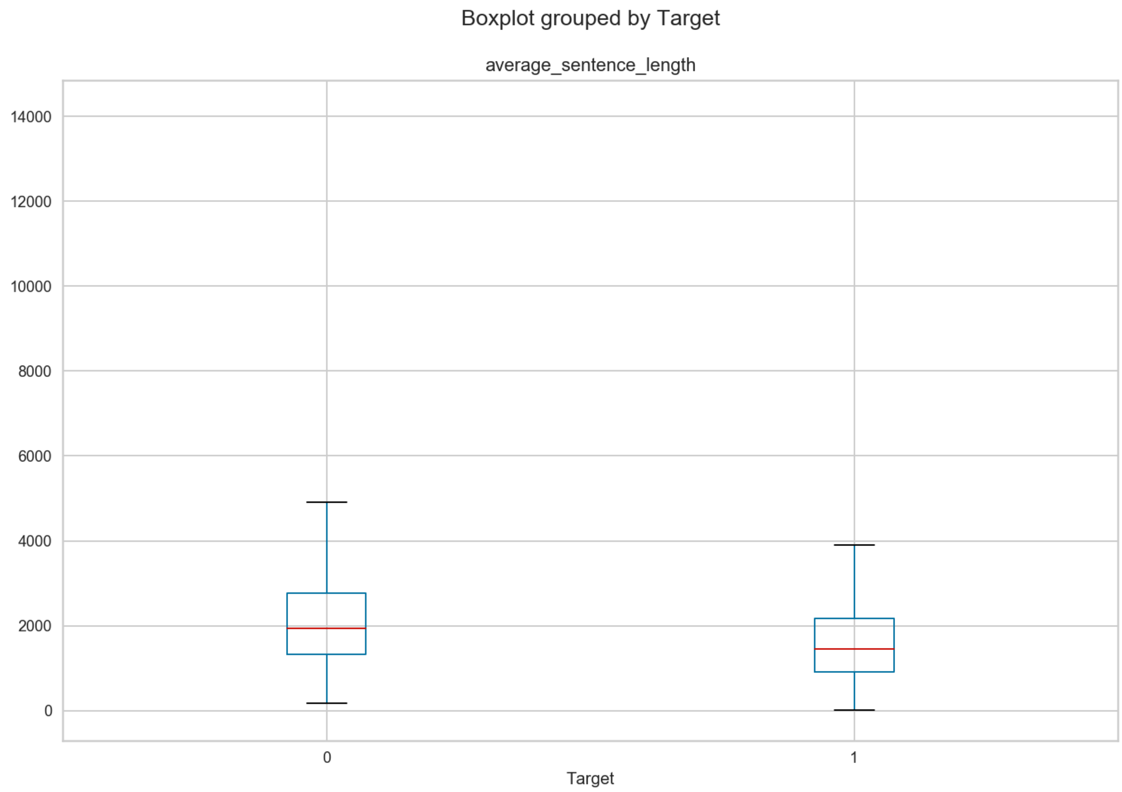

Figure 42: Boxplot – Average sentence length for Fake (1) and True (0) News

<u>Statistical test – Student's t-test</u>

The null hypothesis here is that there is no significant difference in the mean test scores of the two sample groups (Fake News and True News) and that any difference down to chance. As the p-value is < 5%, I reject that hypothesis. Therefore, the average sentence length in a News article have a significant relationship with the variable Target which is the response variable for Fake (1) or True (0) News.

```
Ttest_indResult(statistic=2.901616318204166, pvalue=0.004148537695175065)
```

3) **POS Tagging Features extraction (related to hypothesis 3):** To test the hypotheses mentioned previously, I create a feature for each grammatical category, and I count the number of occurrences per news for these grammatical categories[116].

<u>Cardinal Digit – CD</u>

By looking at the below chart we can see that the number of Cardinal Digits is higher in True News articles than Fake News articles.

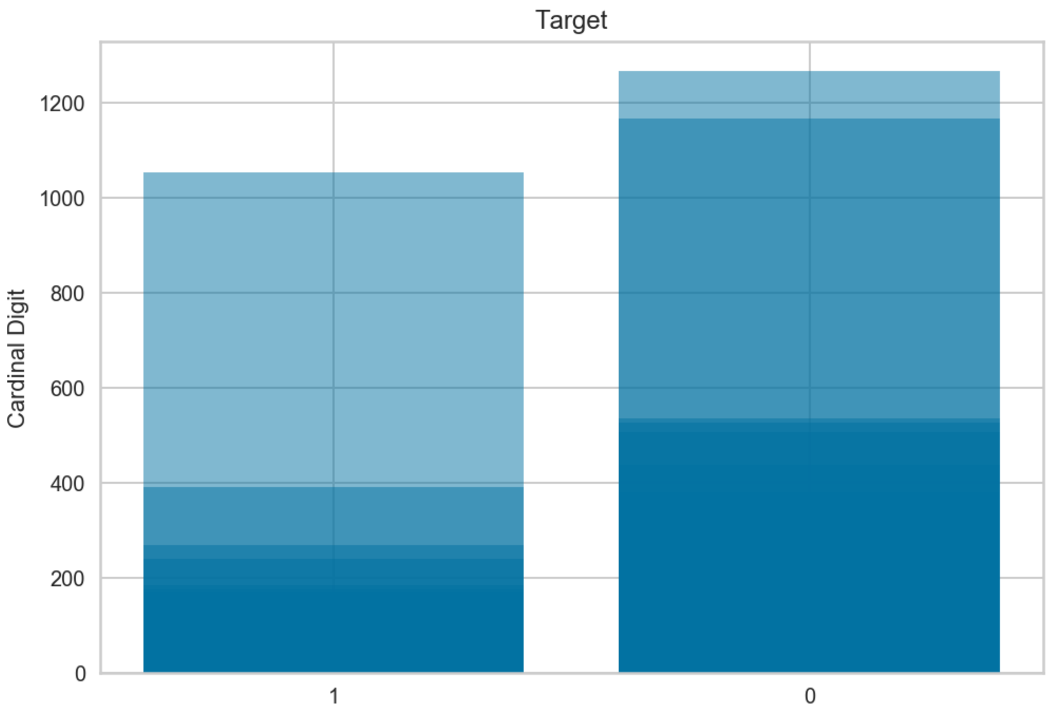

Figure 43: Bar plot – Number of Cardinal Digits in Fake (1) and True (0) News

Student's t-test result:

The null hypothesis here is that there is no significant difference in the mean test scores of the two sample groups (Fake News and True News) and that any difference down to chance. As the p-value is < 5%, I reject that hypothesis. Therefore, the total number of Cardinal Digits (CD) in a News article have a significant relationship with the variable Target which is the response variable for Fake (1) or True (0) News.

```
Ttest_indResult(statistic=2.4256843123235075,pvalue=0.016209346769572697)
```

<u>Adjective – JJ</u>

By looking at the below chart we cannot really understand if there is a difference in the number adjectives presented in Fake and True News.

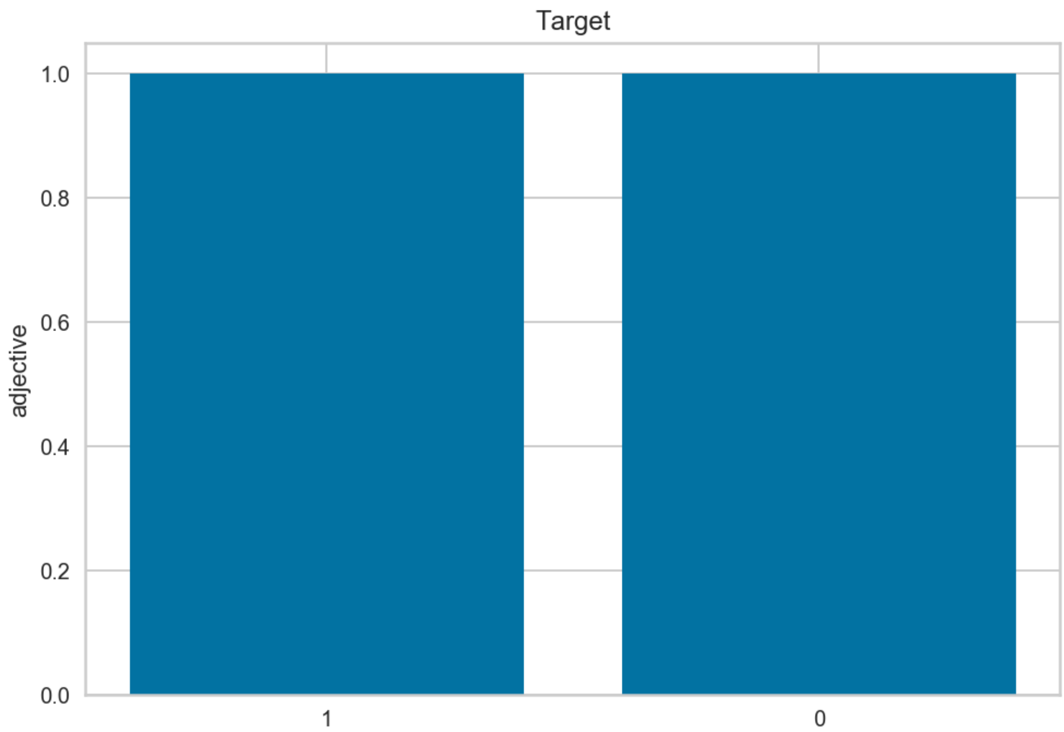

Figure 44: Bar plot – Number of adjectives in Fake (1) and True (0) News

Student's t-test result:

The null hypothesis here is that there is no significant difference in the mean test scores of the two sample groups (Fake News and True News) and that any difference down to chance. As the p-value is > 5%, I don't reject that hypothesis. Therefore, the total number of Adjectives (JJ) in a News article doesn't have a significant relationship with the variable Target which is the response variable for Fake (1) or True (0) News.

```
Ttest_indResult(statistic=0.9948051596666935, pvalue=0.3210896553395598)
```

<u>Noun Singular – NN</u>

By looking at the below chart we see that there aren't enough occurrences to judge if there is a difference in the number of singular nouns presented in Fake and True News.

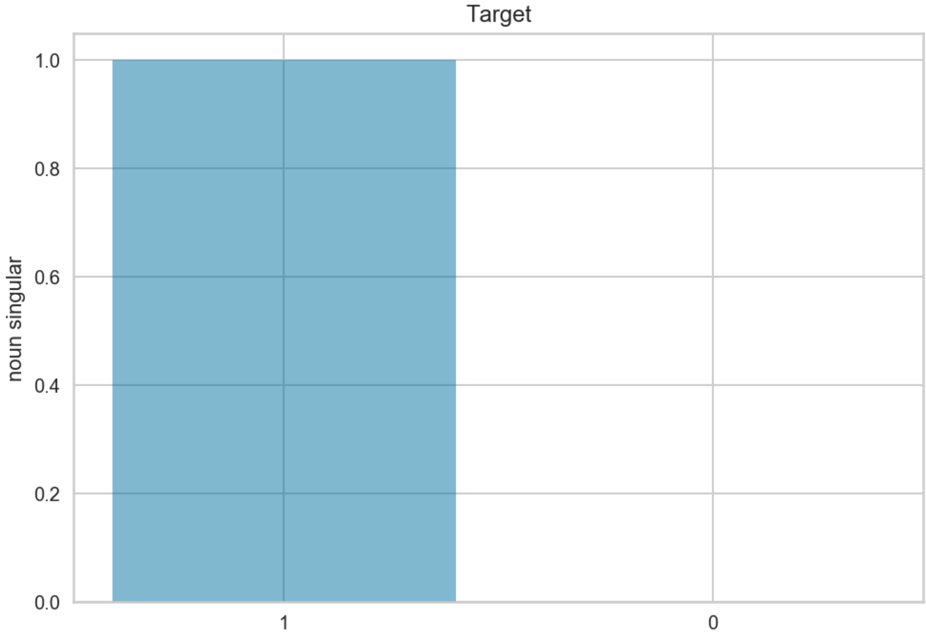

Figure 45: Bar plot – Number of Singular Nouns in Fake (1) and True (0) News

Student's t-test result:

The null hypothesis here is that there is no significant difference in the mean test scores of the two sample groups (Fake News and True News) and that any difference down to chance. As the p-value is > 5%, I don't reject that hypothesis. Therefore, the total number of Singular Nouns (NN) in a News article doesn't have a significant relationship with the variable Target which is the response variable for Fake (1) or True (0) News.

```
Ttest_indResult(statistic=-0.9948051596666969, pvalue=0.32108965533955813)
```

<u>Noun plural – NNS</u>

By looking at the below chart we can see that there isn't any remarkable difference in the number of nouns in plural between Fake and True News articles.

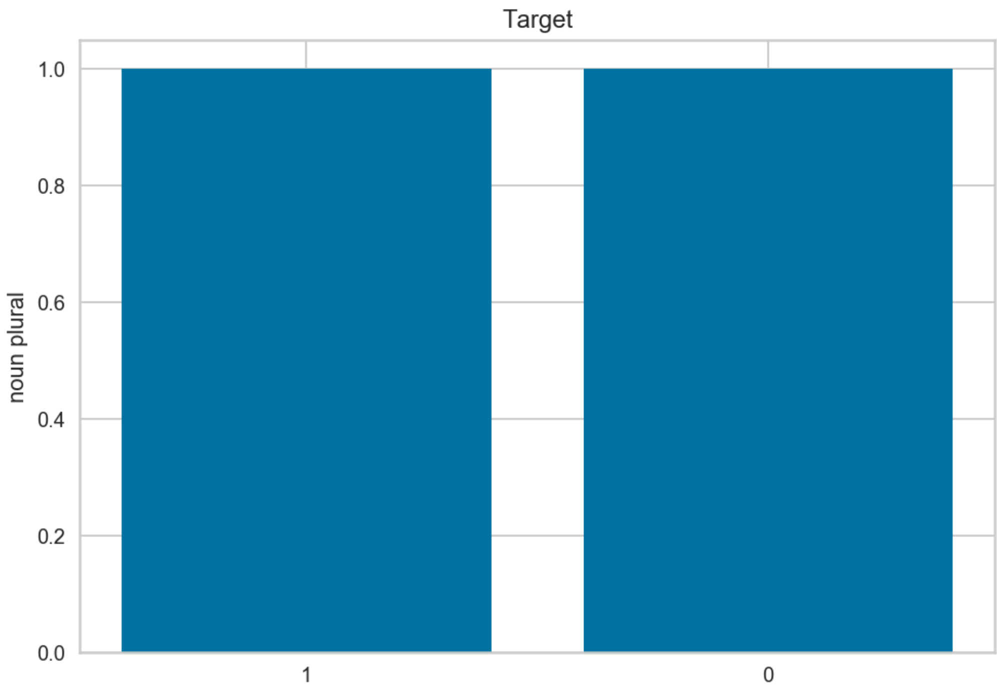

Figure 46: Bar plot – Number of Nouns in plural in Fake (1) and True (0) News

Student's t-test result:

The null hypothesis here is that there is no significant difference in the mean test scores of the two sample groups (Fake News and True News) and that any difference down to chance. As the p-value is < 5%, I reject that hypothesis. Therefore, the total number of Nouns in plural (NNS) in a News article have a significant relationship with the variable Target which is the response variable for Fake (1) or True (0) News.

```
Ttest_indResult(statistic=0.9948051596666935, pvalue=0.3210896553395598)
```

<u>Possessive ending parent's – POS</u>

By looking at the below chart we can observe that the number of possessive ending parents are more in True News than Fake News.

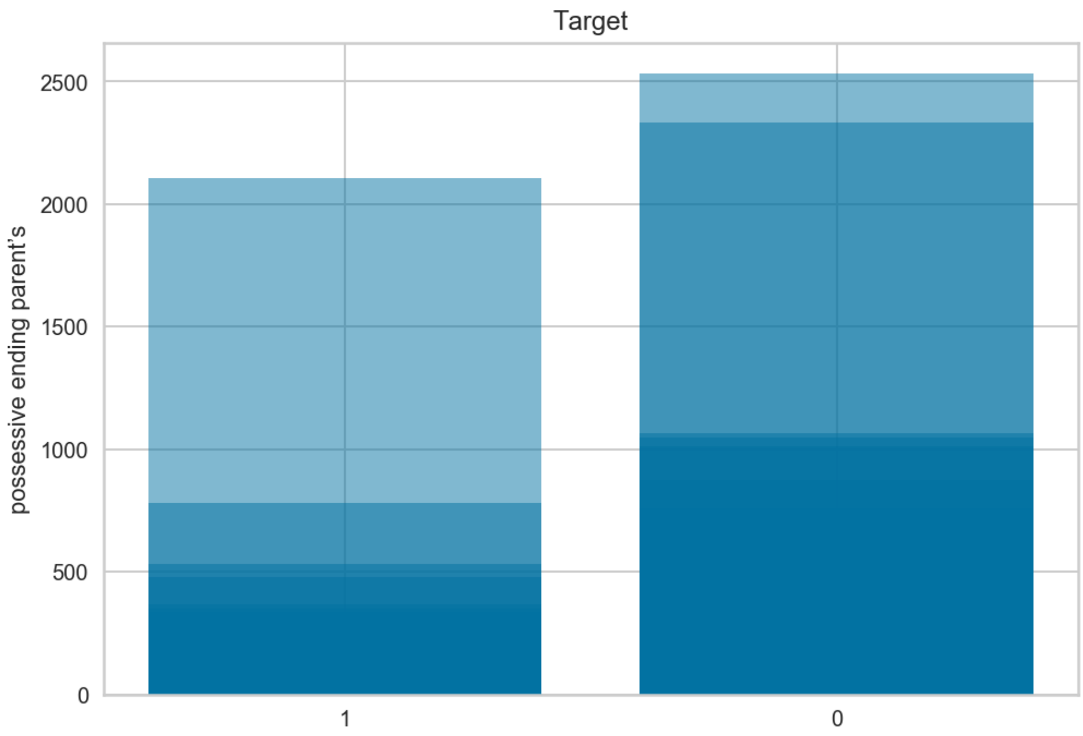

Figure 47: Bar plot – Number of Possessive ending parent's in Fake (1) and True (0) News

Student's t-test result:

The null hypothesis here is that there is no significant difference in the mean test scores of the two sample groups (Fake News and True News) and that any difference down to chance. As the p-value is < 5%, I reject that hypothesis. Therefore, the total number of Possessive ending parent's (POS) in a News article have a significant relationship with the variable Target which is the response variable for Fake (1) or True (0) News.

```
Ttest_indResult(statistic=2.4258652693558806, pvalue=0.016201514845892124)
```

4) **Punctuations counts (related to hypothesis 4):** Previous research of Victoria Rubin[117] on Fake News detection, as well as, research of Daniel W. Otter[118] on spam detection say that using punctuation could help distinguish truthful texts from misleading Fake News. To test my hypotheses, I only consider "?" and "!" as punctuations as I consider those as the most important.

By looking at the below chart, we can observe that there are slightly more punctuations in True News than in Fake News and that is an unexpected result. However, we don't know if that difference is significant. On that purpose we realize a statistical test.

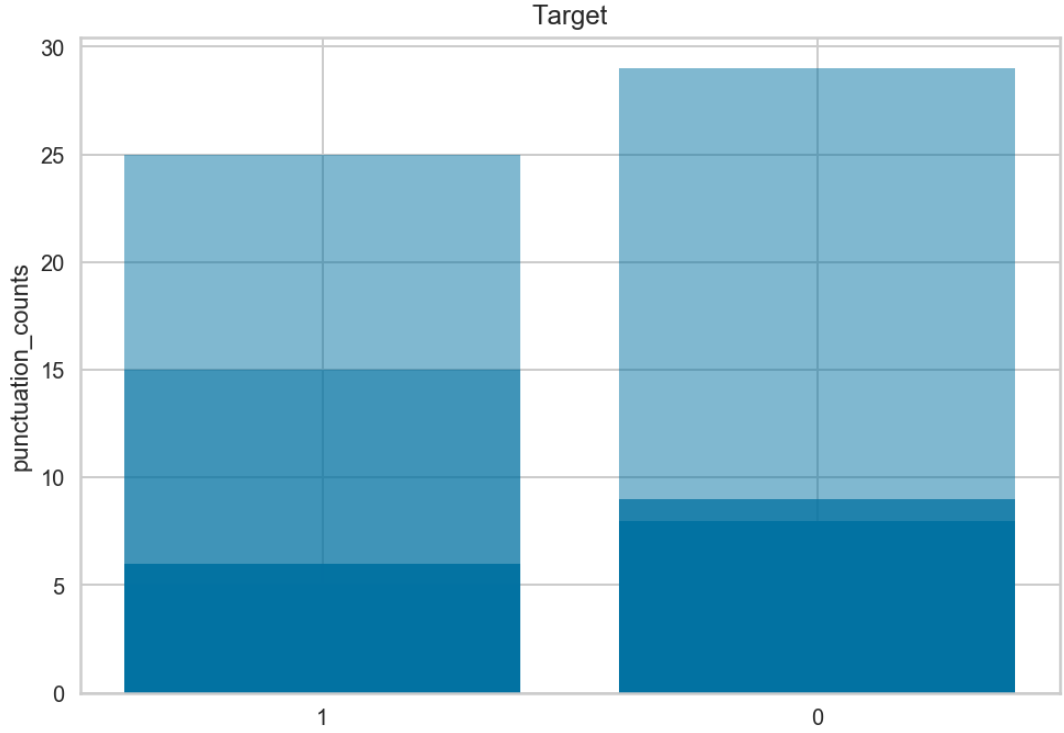

Figure 48: Bar plot – Number of Punctuations in Fake (1) and True (0) News

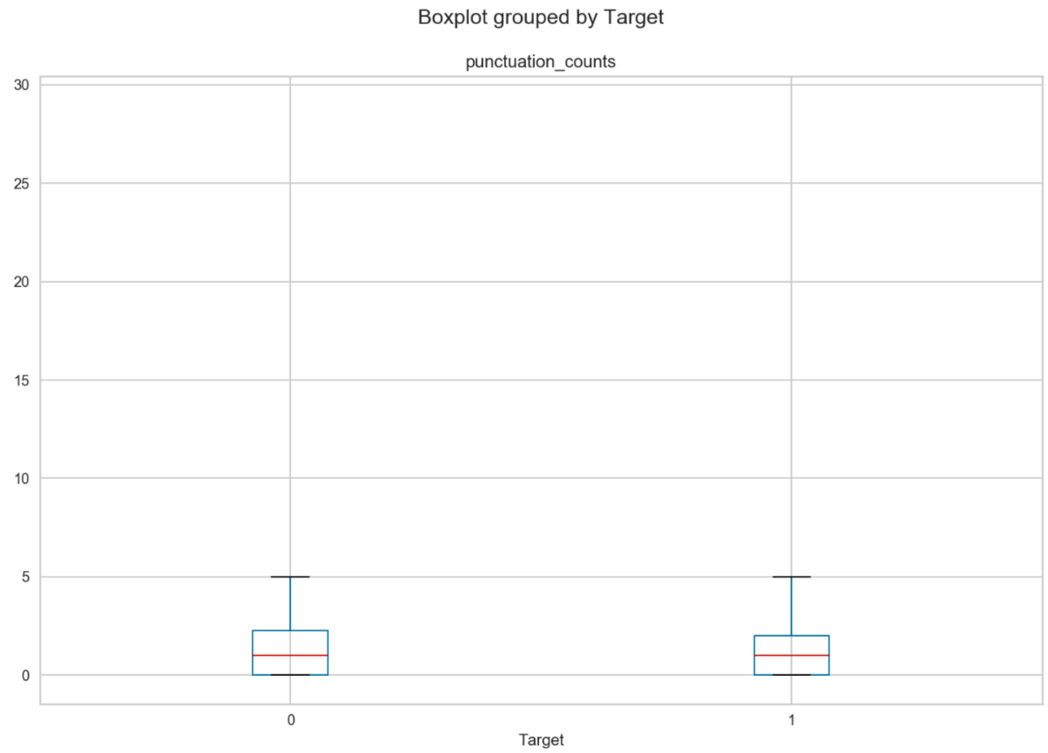

Figure 49: Boxplot – Number of Punctuations in Fake (1) and True (0) News

Statistical test – Student's t-test

The null hypothesis here is that there is no significant difference in the mean test scores of the two sample groups (Fake News and True News) and that any difference down to chance. As the p-value is > 5%, I don't reject that hypothesis. Therefore, the total number of Punctuations in a News article doesn't have a significant relationship with the variable Target which is the response variable for Fake (1) or True (0) News.

```
Ttest_indResult(statistic=0.597213018303081, pvalue=0.5510724811833383)
```

5) Lexical Diversity level using Text-Type Ratio (TTR) (related to hypothesis 5):

Recent researches[21] have shown that Lexical Diversity can be an important indicator for Fake News detection. In Wikipedia, the lexical diversity is defined as below:

> Lexical diversity is one aspect of "lexical richness" and refers to the ratio of different unique word stems (types) to the total number of words (tokens). The term is used in applied linguistics and is quantitatively calculated using numerous different measures including Text-Type Ratio (TTR), vocd, and the measure of textual lexical diversity (MTLD).
>
> A common problem with lexical diversity measures, especially TTR, is that text samples containing large number of tokens give lower values for TTR since it is often necessary for the writer or speaker to re-use several function words. One consequence of this is that lexical diversity is better used for comparing texts of equal length. Newer measures of lexical diversity attempt to account for sensitivity to text length. [119]

In my analysis, I will be using the Text-Type Ratio (TTR) as the measure of lexical diversity in the news body text. On that purpose, I will be using the lex_div() function of the Lexical-Diversity package. To be more precise, what this the function does is to look at the news text and say there are x many unique word types and then divide that by the number of tokens.

By looking at the below chart we can see that there is slightly more lexical diversity in Fake News compared to True News.

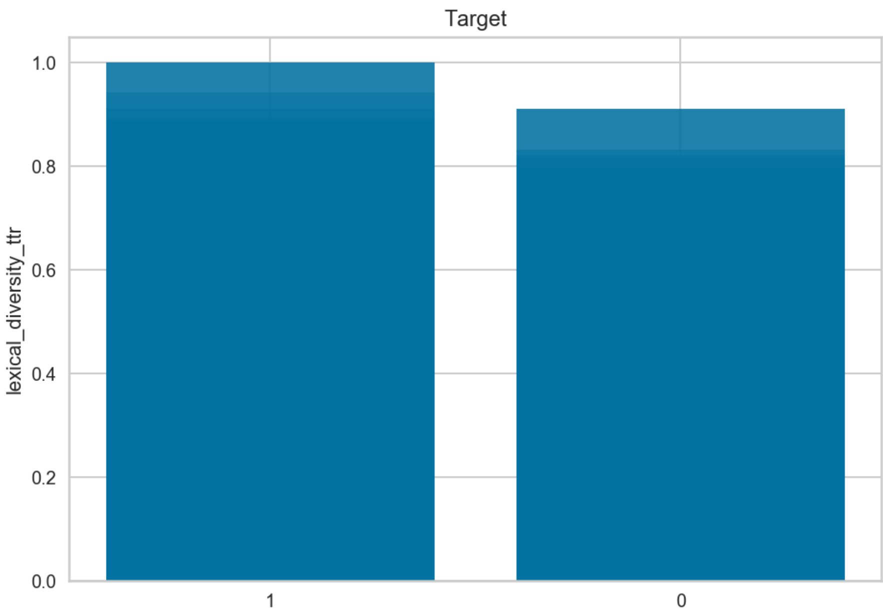

Figure 50: Bar plot – Lexical Diversity in Fake (1) and True (0) News

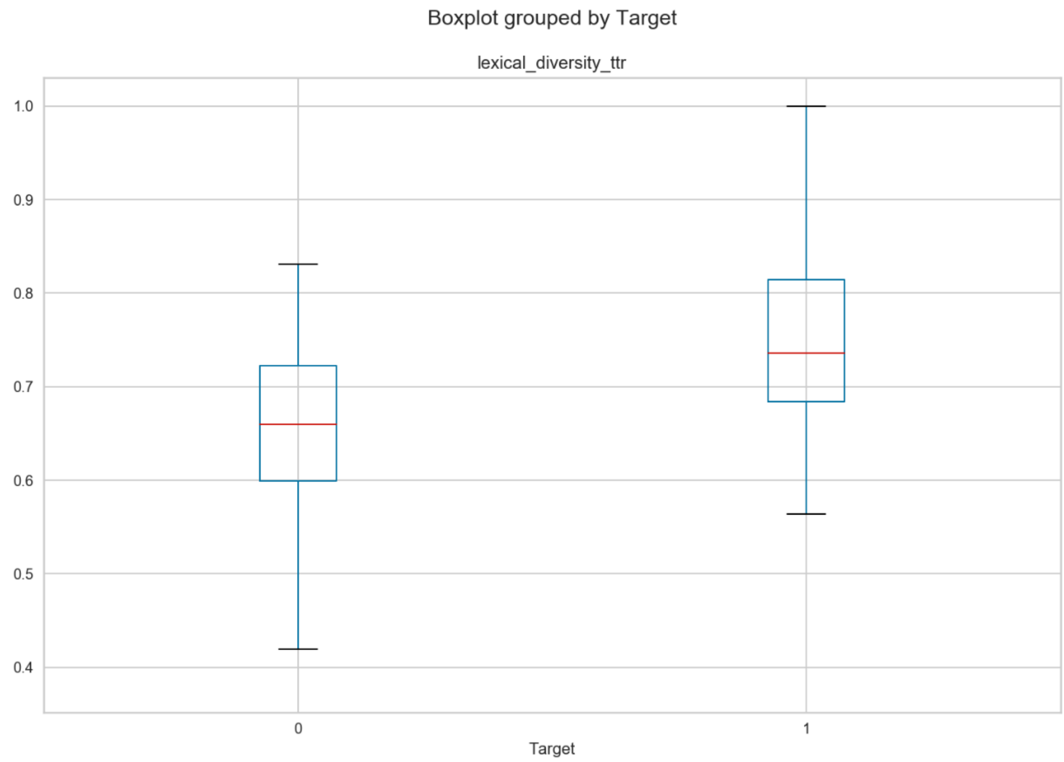

Figure 51: Boxplot – Lexical Diversity in Fake (1) and True (0) News

<u>Statistical test – Student's t-test</u>

The null hypothesis here is that there is no significant difference in the mean test scores of the two sample groups (Fake News and True News) and that any difference down to chance. As the p-value is < 5%, I reject that hypothesis. Therefore, the Lexical Diversity in a News article have a significant relationship with the variable Target which is the response variable for Fake (1) or True (0) News.

```
Ttest_indResult(statistic=-5.553222552186553, pvalue=9.301566551574583e-08)
```

3.3.3.3 Psycho-linguistic analysis

Hypothesis:

1) I suppose that Fake News articles have mostly positive or negative sentiment but not neutral.

List of Features:

1) Polarity Based Sentiment scores

As also explained in the second chapter, sentiment analysis can be described as the method of statistically determine whether a text is positive, negative, or neutral. It is sometimes called "opinion mining"[120]. There are two main methods to realise sentiment analysis. The one is polarity-based, where texts are categorized as either positive or negative, and the other one is the valence-based, where the sentiment strength is considered. I the context of my analysis, I will use the polarity-based method by calculating the polarity_scores using SentimentIntensityAnalyzer () function of the vaderSentiment package.

The tool developers describe Vader as follows:

> VADER (Valence Aware Dictionary and sEntiment Reasoner) is a lexicon and rule-based sentiment analysis tool that is specifically attuned to sentiments expressed in social media.
>
> VADER uses a combination of a sentiment lexicon is a list of lexical features (e.g., words) which are generally labelled according to their semantic orientation as either positive or negative. [121]

The human raters of Vader used 5 heuristics to analyse the sentiment:

- Punctuation — I love pizza vs I love pizza!!
- Capitalization — I'm hungry!! vs I'M HUNGRY!!
- Degree modifiers (use of intensifiers)— I WANT TO EAT!! VS I REALLY WANT TO EAT!!
- Conjunctions (shift in sentiment polarity, with later dictating polarity) — I love pizza, but I really hate Pizza Hut (bad review)
- Preceding Tri-gram (identifying reverse polarity by examining the tri-gram before the lexical feature— Canadian Pizza is not really all that great. [122]

Vader is generally used through the comments published on posts by social media users. However, this technique can be applied to any other type of text. Therefore, it can be applicable in the case of a news article.

We classify a text as positive when the rate is higher than 0.05, as negative when the rate is less than - 0.05 and neutral when it is in between.

I will define the different degrees as follows:

- Positive = 3
- Neutral = 2
- Negative = 1

Descriptive analysis

Contingency Table

	Target	
Community	True News (0)	Fake News (1)
1 - Negative	53	49
2 - Neutral	1	4
3 - Positive	42	44

By looking at the below charts I can observe that very few articles are neutral either Fake News or True News. Moreover, it doesn’t look like there is a big difference in the sentiment intensity between Fake News and True News. I will realize a statistical test to be able to see if there is a significant relationship between the sentiment score and the response variable “Target”.

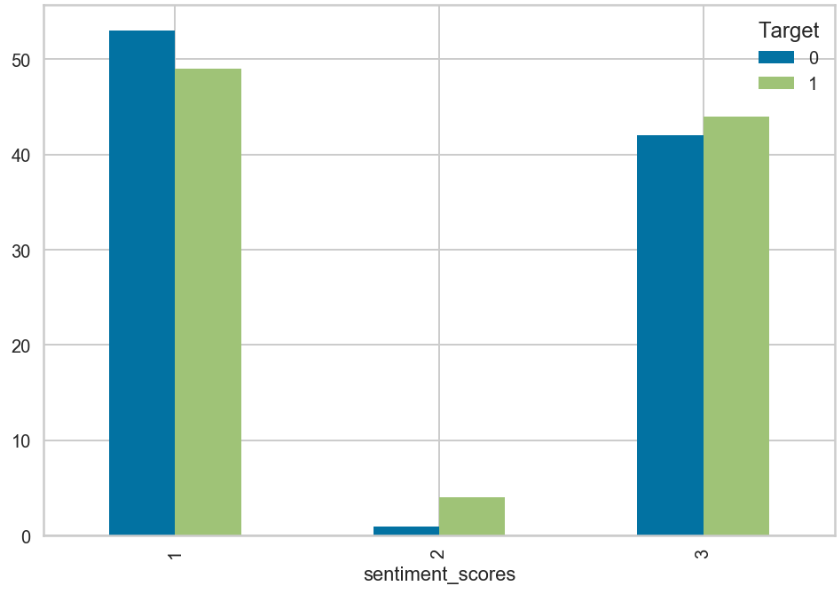

Figure 52: Bar plot – Fake (1) and True (0) News by Sentiment category

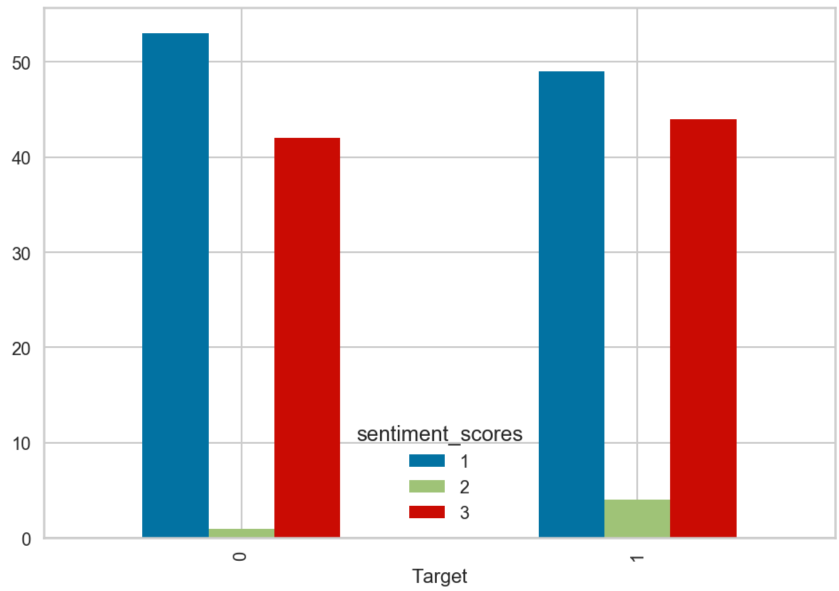

Figure 53: Bar plot – Sentiment category by Fake (1) and True (0) News articles

Chi-squared test results

The test result gives a p-value = 0.368 which is superior to 5% significant level calculated by inverting the 95% probability used in the critical value interpretation. Therefore, we fail to reject the null Hypothesis, so the Variables “Most frequent community per news” is independent to “Target” which is the Response variables for Fake and True News.

Results using Python 3 chi2() and chi2_contingency() function of the Scipy.stats package :

```
probability=0.950, critical=5.991, stat=1.998
Independent (fail to reject H0)
significance=0.050, p=0.368
Independent (fail to reject H0)
```

3.4 Results

3.4.1 Dataset description

Online news articles can have various sources. The most important source are news agencies websites, search engines, and social media websites. Evaluating the truthfulness of news manually is a very demanding task that involves annotators with domain expertise who performs an in-depth analysis of claims and additional evidence, context, and reports from authoritative sources[52].

News data that are already annotated as Fake or True can be collected with the contribution of expert journalists, fact-checking websites, industry detectors, and crowdsourced workers. An introduction on those techniques has given in the second chapter which focuses on the knowledge-based Fake News detection models.

For my analysis, I will be working on the Fake News detection dataset, and its name is FakeNewsNet[123]. The dataset is collected from two platforms doing fact-checking for political context articles: BuzzFeed and PolitiFact. Both datasets possess news article text content with labels and social context information.

News content includes the meta attributes of the news such as the body text, and social context includes the associated user social engagements of the news items, such as user posts/shares of news in Twitter and their frequencies.

In our dataset we have in total 240 documents (articles) that correspond to 193 documents in the training set and 47 documents in the test set. Below you can find two graphs, one with the number of documents per training and test set, as well as, the percentage for each category.

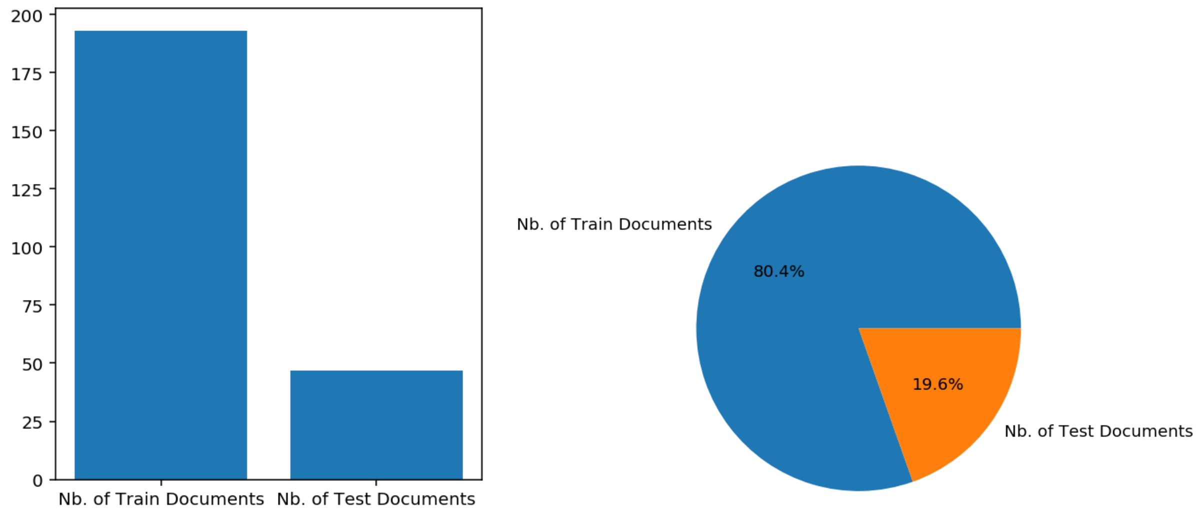

Figure 69: Number and Percentage of News on the Training and Test sets

The total number of unique users that shared at least once, one article is 240. Moreover, the number of Fake News in the Training set is 97 and the number of True News is 96. Therefore, our dataset is well equilibrated.

As we can see in the below plot, the number of Fake and True News is almost the same (97 Fake New against 96 True News) in our dataset. This can be perceived as a positive element as we do not have an overrepresented category. Having an imbalanced (unequal distribution of classes within a dataset) dataset when we want to apply classification methods can bring bias to the results. In those cases, there are methods developed to tackle imbalanced classes (e.g., under sampling, oversampling and generating synthetic data)

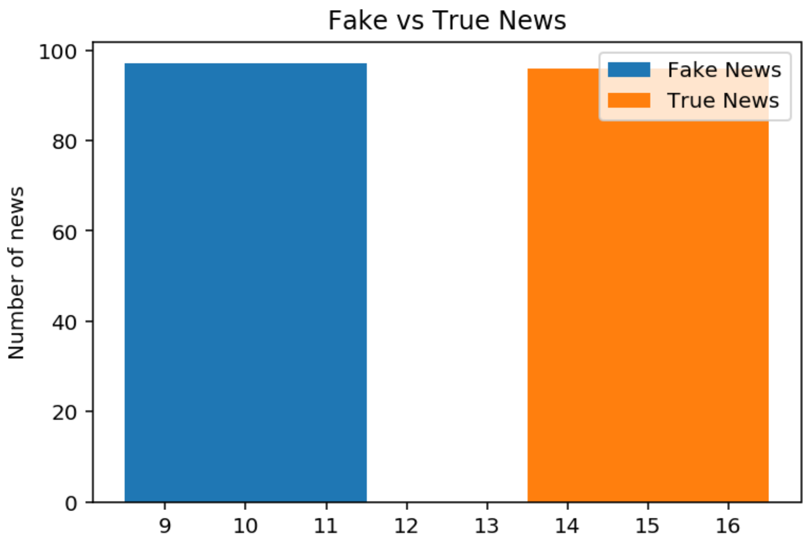

Figure 54: Number of Fake News and True News in our dataset

3.4.2 Study Design

In this chapter, we will find details on the different aspects of the classification process. We will start by introducing two classification machine learning algorithms that are commonly used for Fake News detection (Naïve Bayes & Random forest classifier). Furthermore, I will present you the Features Selection method that I decided to apply for each model type, and I will present you the results. Then, I will train and compare models based on the different scenarios listed below. I will realize prediction on the test set and compare models based on evaluation metrics.

List of scenarios:

- **Scenario 1:** Social features with Naïve Bayes classifier (model 1) vs Text features with Naïve Bayes classifier (model 2)
- **Scenario 2:** Social features with Random Forest classifier (model 3) vs Text features with Random Forest classifier (model 4)
- **Scenario 3:** Social + Text features with Naïve Bayes Classifier (model 5) vs Text features with Naïve Bayes classifier
- **Scenario 4:** Social + Text features with Random Forest classifier (model 6) vs Text features with Random Forest classifier

3.4.2.1 Classification algorithms

Naive Bayes (NB) classifier

Naïve Bayes derives from Bayes Theorem that is used for calculating conditional probability, the "*probability that something will happen, given that something else has already occurred*"[124]. Thus, we can compute the likelihood of a certain outcome by using past knowledge of it. Naïve Bayes is indeed a type of classifier which is regarded to be a supervised learning algorithm that belongs to the Machine Language category and works by predicting "member probabilities" for each class (e.g., binary response), such as the likelihood that the given evidence or record belongs to a class.

The advantages of the Naïve Bayes classifier are that it is relatively fast and highly accessible. The biggest inconvenience of this method is that it assumes all the features to be independent, which is not always the case. Hence, there is no relationship learned among the features[124].

Random Forest classifier

Random forest classifier generates a set of decision trees from a training set's randomly chosen subset. In order to decide the output of the test object, it aggregates the votes from distinct decision trees. Random forests are very commonly used by machine learning experts and that is due to their prediction performance and accuracy, low overfitting and their easy interpretation.

Tony Yiu described the Random Forest classification process in Toward Data Science as follows,

> The Random Forest Classifier. Random forest, like its name implies, consists of a large number of individual decision trees that operate as an ensemble. Each individual tree in the random forest spits out a class prediction and the class with the most votes becomes our model's prediction[125]

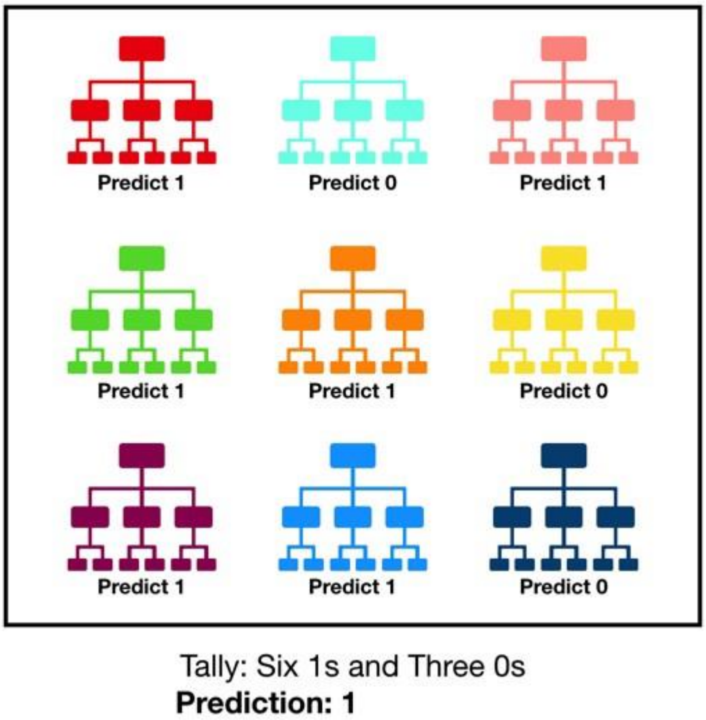

Figure 55: Random Forest Classifier

Source: Yiu, Tony, "Understanding Random Forest," *Medium, Towards Data Science*.

It is important to note that Random forests can automatically perform the task of feature selection: deciding which features of the data are most relevant in answering a given question, in our case if a news articles is Fake or True. On that matter, I will give you more details in the next chapter "Features Selection".

3.4.2.2 Features Selection Methodology

In the previous chapters 3.2 and 3.3, we realized bivariate analysis to evaluate the dependence between the features extracted (explanatory variables) and the response variable “Target.” We implement Student’s t-test for the numeric explanatory variables and Chi-2 test for categorical variables. That is a simple technique that allows the understanding of the relationship between an explanatory variable and the predicted variable. However, those tests are reliable only in the case that all explanatory variables are independent.

Student’s t-test (or independent t-test, or the two-sample t-test), is an inferential statistical test that determines whether there is a statistically significant difference between the means in two unrelated groups. I have implemented this test by making the assumptions about the scale of measurement, random sampling, normality of data distribution, adequacy of sample size (at least 50 observations) and equality of variance in standard deviation in the two groups. As we will see later in this chapter, some of the numeric variables which they had not significant difference in their mean in the two groups based on the t-test, they will be selected by the methods presented in this chapter (e.g., variable Page Rank score, Eigenvector Centrality score). That is likely because one or more of the assumptions of the Student’s t-test were not actually satisfied. Although we know that the assumption of the sample size is satisfied.

The chi-square statistic is commonly used for testing relationships between categorical variables. The null hypothesis of the chi-square test is that no relationship exists on the categorical variables in the population; they are independent. The chi-square is appropriate when the sampling method is simple random sampling, the variables under study are categorical and the expected number observations in each level of the variable is at least five. In chi-square case the assumptions seem to be satisfied, as a result we are expecting more coherence in the results between the bivariate tests and the results presented in this chapter.

Naïve Bayes classifier

In this part I select the features that they will be used for the Naive Bayes classifier. Features selection is a technique where I automatically select those features in my dataset that contribute significantly to the prediction of the response variable which is in our case the variable “Target” that takes the value “1” for Fake News and “0” for True News. The features selection is a critical task of the classification process, as model performance will be reduced by having too many irrelevant features in our dataset. The main advantages of evaluating the importance of the features and realize features selection before modelling are the following:

- Reduces overfitting
- Improves accuracy
- Less data implies that algorithms train more quickly.

I apply the Recursive Feature Elimination with cross-validation (RFECV) method by using the RFECV () function of the Sklearn package for Python 3 that basically realize feature ranking with recursive feature elimination and cross-validated selection of the best number of features.

More specifically, Recursive Feature Elimination (RFE) eliminates features iteratively. The algorithm begins with the full regression model containing all features and then removes the least useful predictor (worst feature importance) in each iteration.

Textual & Social – Features selection

The Recursive Feature Elimination algorithm selected the best Naïve Bayes classifier with 77 features. As we can see in the below chart, the cross-validated accuracy for classification start to decrease when the features are more than 77. The original features were 119.

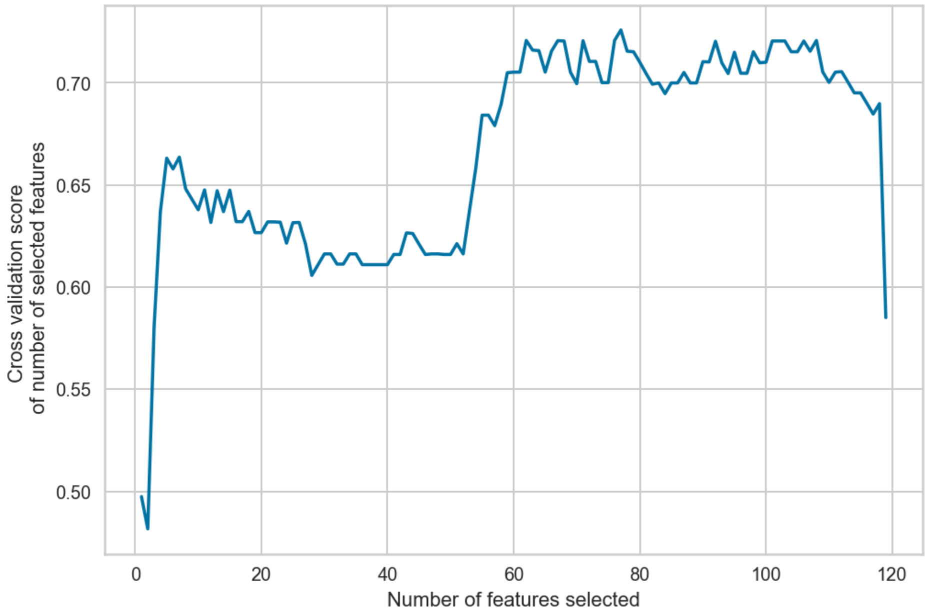

Figure 56: Accuracy score depending on the number of features selected (model with both Textual and Social Based features).

List of Selected Features:

```
'Eigenvector Centrality', 'page_rank', 'NN', 'American_count_vect',
'But_count_vect', 'Hillary_count_vect', 'New_count_vect',
'North_count_vect', 'Republican_count_vect',
'Trumpxe2x80x99s_count_vect', 'US_count_vect', 'United_count_vect',
'York_count_vect', 'asked_count_vect', 'campaign_count_vect',
'debate_count_vect', 'did_count_vect', 'government_count_vect',
'itxe2x80x99s_count_vect', 'million_count_vect', 'new_count_vect',
'president_count_vect', 'presidential_count_vect', 'say_count_vect',
'state_count_vect', 'think_count_vect', 'xe2x80x94_count_vect',
'according', 'american', 'asked', 'campaign', 'clinton', 'country',
'debate', 'did', 'donald', 'donxe2x80x99t', 'going', 'government',
'hillary', 'house', 'itxe2x80x99s', 'just', 'know', 'like', 'make',
'man', 'million', 'new', 'news', 'north', 'obama', 'people', 'police',
'president', 'presidential', 'republican', 'right', 'said', 'say',
'says', 'state', 'states', 'think', 'time', 'told', 'trump',
'trumpxe2x80x99s', 'united', 'want', 'way', 'white', 'women',
'xe2x80x93', 'xe2x80x94', 'years', 'york'
```

We can observe that both Social context and Textual features were selected. Count vectorizer features are having "_count_vect" in their names. Any other word variable in the results without that mention is a TFIDF feature.

Textual only – Features selection

The Recursive Feature Elimination algorithm selected the best Naïve Bayes classifier with 61 features. As we can see in the below, the cross-validated accuracy for classification stays the same or decreases when the features are more than 61. The original features were 112.

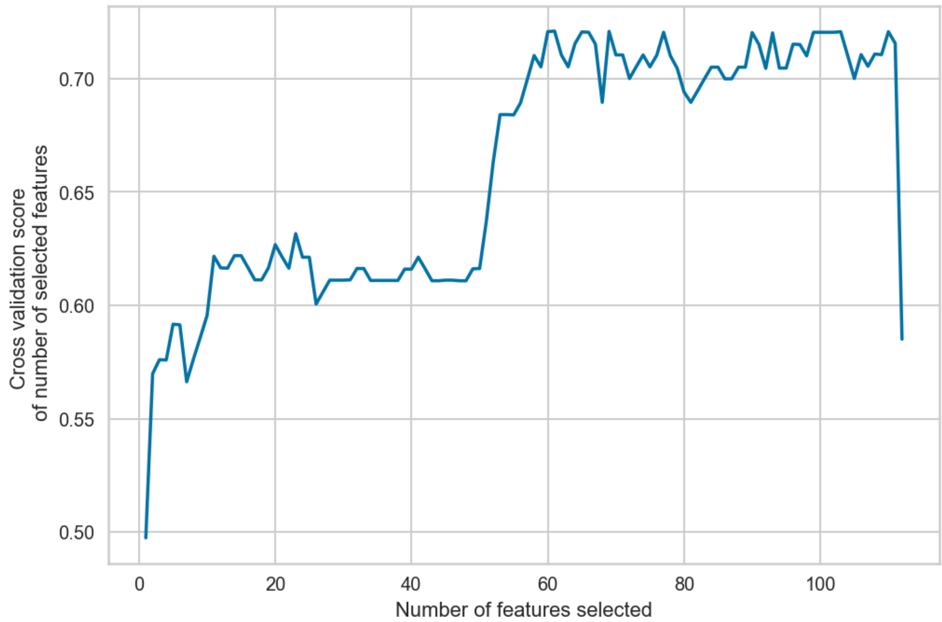

Figure 57: Accuracy score depending on the number of features selected (model with only Textual features).

<u>List of Selected Features:</u>

```
'NN', 'But_count_vect', 'North_count_vect', 'Republican_count_vect',
'York_count_vect', 'campaign_count_vect', 'debate_count_vect',
'government_count_vect', 'million_count_vect',
'presidential_count_vect', 'state_count_vect', 'according', 'american',
'asked', 'campaign', 'clinton', 'country', 'debate', 'did', 'donald',
'donxe2x80x99t', 'going', 'government', 'hillary', 'house',
'itxe2x80x99s', 'just', 'know', 'like', 'make', 'man', 'million', 'new',
'news', 'north', 'obama', 'people', 'police', 'president',
'presidential', 'republican', 'right', 'said', 'say', 'says', 'state',
'states', 'think', 'time', 'told', 'trump', 'trumpxe2x80x99s', 'united',
'want', 'way', 'white', 'women', 'xe2x80x93', 'xe2x80x94', 'years',
'york'
```

We can observe that count vectorizer features, POS tagging and tfidf features were selected.

Social context only – Features selection

The Recursive Feature Elimination algorithm selected the best Naïve Bayes classifier with 6 features (7 in total). As we can see in the below, the cross-validated accuracy for classification decreases when the features are more than 6.

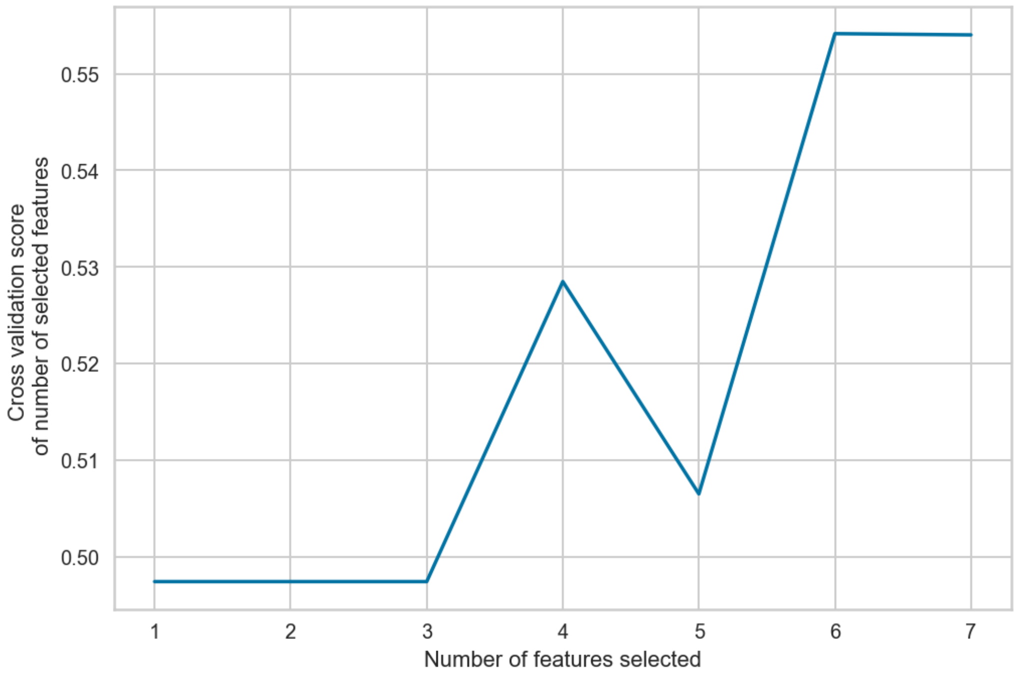

Figure 58: Accuracy score depending on the number of features selected (model with only social context features).

Features Selected

```
'page_rank', 'Community ID', 'User ID', 'Nb of spreads',
'Eigenvector Centrality', 'Avg Nb of spreads',
'Nb of communities per News'
```

Random Forest classifier

As explained also previously, Random Forests classifier allows us to derive the importance of the contribution of the features on the tree decision in the classification process thanks to built-in feature selection methods. Features selection with Random Forest classifier falls under the Embedded methods for features selection.[126]

Chris Albon explained that process in more details,

> Random Forests are often used for feature selection in a data science workflow. The reason is because the tree-based strategies used by random forests naturally ranks by how well they improve the purity of the node. This mean decrease in impurity over all trees (called gini impurity). Nodes with the greatest decrease in impurity happen at the start of the trees, while notes with the least decrease in impurity occur at the end of trees. Thus, by pruning trees below a particular node, we can create a subset of the most important features. [127]

Akash Dubey explained the features selection process as follows,

> In the Random Forest classification process each tree is also a sequence of yes-no questions based on a single or combination of features. At each node (this is at each question), the three divides the dataset into 2 buckets, each of them hosting observations that are more similar among themselves and different from the ones in the other bucket. Therefore, the importance of each feature is derived from how "pure" each of the buckets is. To give a better intuition, features that are selected at the top of the trees are in general more important than features that are selected at the end nodes of the trees, as generally the top splits lead to bigger information gains.[126]

To realize features selection in Python 3, I will be using SelectFromModel () function of the Sklearn.feature_selection package. I will select those features which importance is greater than the mean importance of all the features by default, but we can alter this threshold if we want.

Below you can find the results of the selected features for the different models (only textual features, only social-based features and combination of the two).

Textual & Social – Features Selection

43 features have been selected among the 119 original features. Below is the list of the selected features.

```
'Eigenvector Centrality', 'page_rank', 'count_word',
'count_word_unique', 'average_sentence_length', 'CD', 'POS',
'lexical_diversity_ttr', 'Donald_count_vect', 'New_count_vect',
'New York_count_vect', 'President_count_vect', 'Republican_count_vect',
'Trump_count_vect', 'York_count_vect', 'debate_count_vect',
'presidential_count_vect', 'said_count_vect', 'xe2x80x94_count_vect',
'clinton', 'country', 'debate', 'did', 'donald', 'hillary',
'itxe2x80x99s', 'just', 'know', 'like', 'make', 'new', 'news',
'president', 'presidential', 'republican', 'said', 'states', 'time',
'trump', 'white', 'xe2x80x93', 'xe2x80x94', 'york'
```

We can observe that both Social context and Textual features were selected. We can also see that only unigrams have been selected and not bigrams.

Textual only – Features Selection

37 features have been selected among the 112 original features. Below is the list of the selected features.

```
'count_word', 'count_word_unique', 'average_sentence_length', 'CD',
'POS', 'lexical_diversity_ttr', 'New_count_vect', 'New York_count_vect',
'President_count_vect', 'Republican_count_vect', 'Trump_count_vect',
'York_count_vect', 'debate_count_vect', 'presidential_count_vect',
'said_count_vect', 'xe2x80x94_count_vect', 'campaign', 'debate',
'donald', 'donxe2x80x99t', 'going', 'hillary', 'itxe2x80x99s', 'just',
'like', 'new', 'news', 'president', 'presidential', 'republican',
'said', 'time', 'trump', 'united', 'xe2x80x93', 'xe2x80x94', 'york'
```

We can observe that lexical Tfidf, count vectorizer and syntactic features have been selected. We can also see that only unigrams have been selected and not bigrams.

Social only – Features Selection

2 features have been selected among the 7 original features. Below is the list of the selected features.

```
'page_rank', 'Eigenvector Centrality'
```

3.4.2.3 Test Scenarios

In this section, I analyse and compare the general performance of the proposed models in the Fake News prediction. I expect to obtain results demonstrating that the hybrid detection model that considers both content-based features and social-based features can perform comparatively well against the model that considers only content-based ones. For fairness of comparison, I report the performances of the methods that rely on two different supervised classifiers, the Naïve Bayes classifier, and the Random Forest classifier.

I will proceed as follows: I will be training the six models defined at the beginning of this chapter. I will evaluate their performance and the role of the training dataset on that performance, based on their Learning curve. I will then compare the four Scenarios listed at the beginning of the chapter by implementing prediction on the test set.

Naïve Bayes with only social context features (Model 5)

In this section, I train the models with the variables selected in the previous section. In addition, I implement cross validation methods to have insights on how the training score evaluates.

By looking at the below Learning curve (Sklearn learning_curve() function), we can observe that the training score is high at the beginning and decreases and the cross-validation score is low at the beginning and decreases more. Both the cross-validation score and the training score converge to a value that is generally lower with increasing size of the training set, so we will not benefit much from more training data.

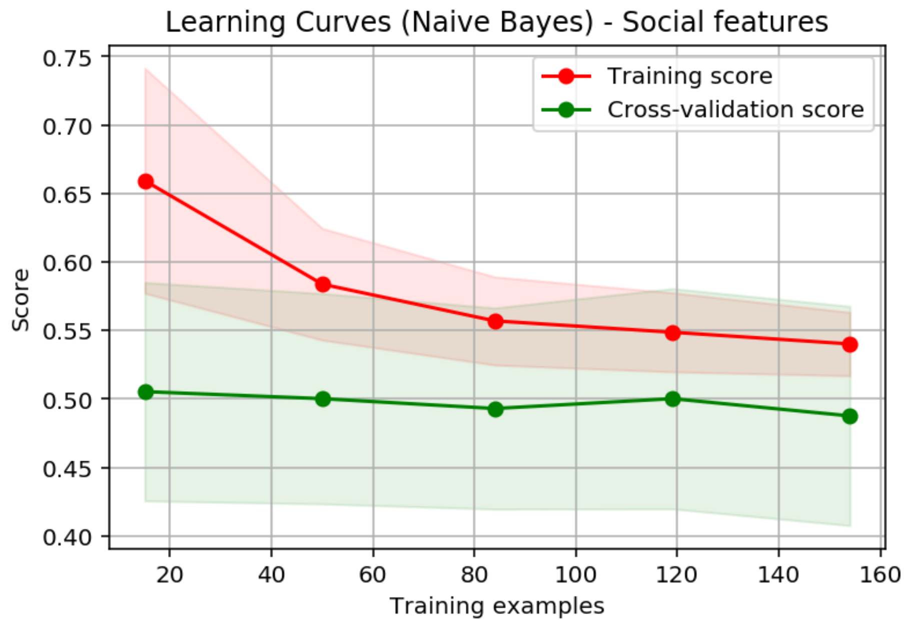

Figure 59: Learning Curve for Naïve Bayes classifier with Social context features

Naïve Bayes with only Textual features (Model 2)

By looking at the below Learning curve, we can observe that the training score is very high at the beginning and decreases and the cross-validation score is very low at the beginning and increases. Both the cross-validation score and the training score converge to a value that is generally lower with increasing size of the training set, so we will not benefit much from more training data.

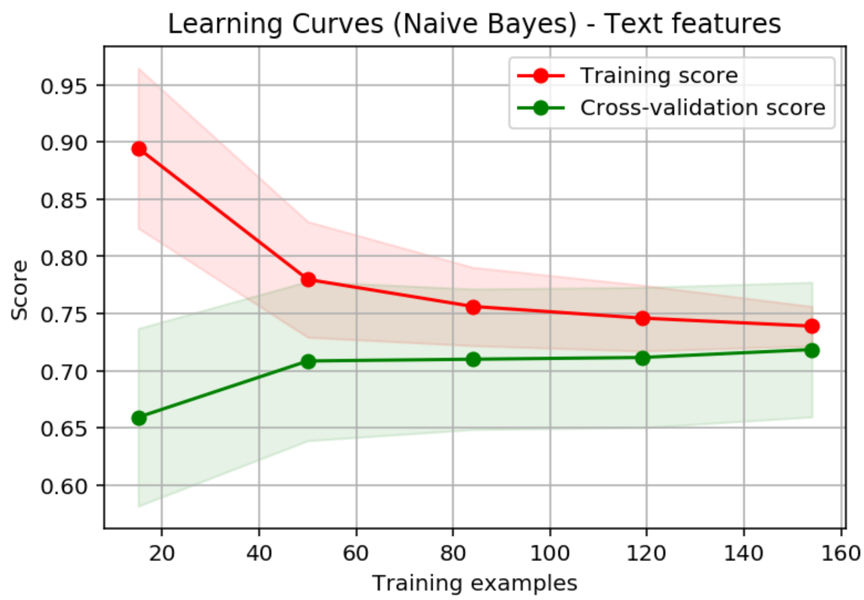

Figure 60: Learning Curve for Naïve Bayes classifier with Textual features

Naïve Bayes with both Social and Text Features (model 5)

Like the Naïve Bayes models with only textual data, the model that contains both textual and social based features, the cross-validation score and the training score converge to a value that is generally lower with increasing size of the training set, so we will not benefit much from more training data.

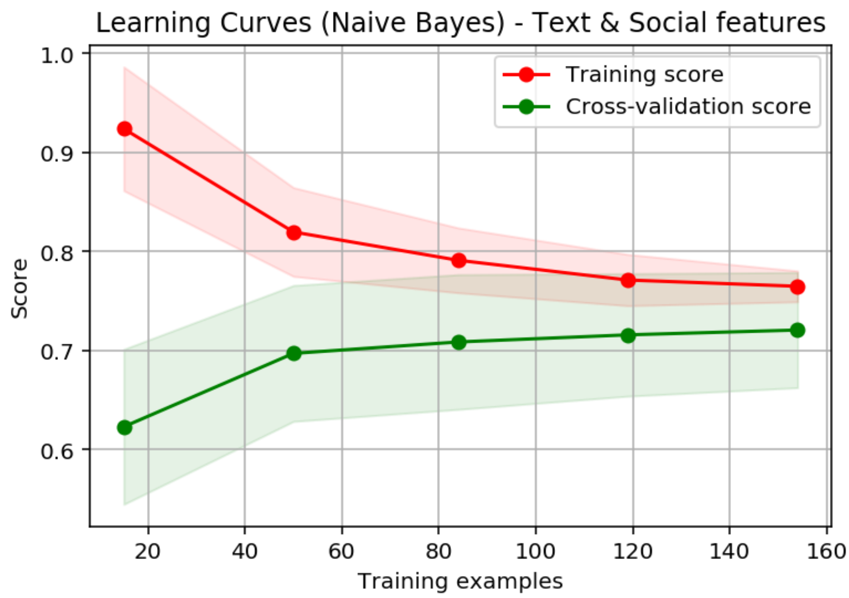

Figure 61: Learning Curve for Naïve Bayes classifier with Textual and Social context features

Random Forest Classifier with only social context features (Model 3)

By looking at the below Learning curve, we can observe that the training score is very high at the beginning and decreases and the cross-validation score is very low at the beginning and becomes lower. Both the cross-validation score and the training score converge to a value that is generally lower with increasing size of the training set, so we will not benefit much from more training data.

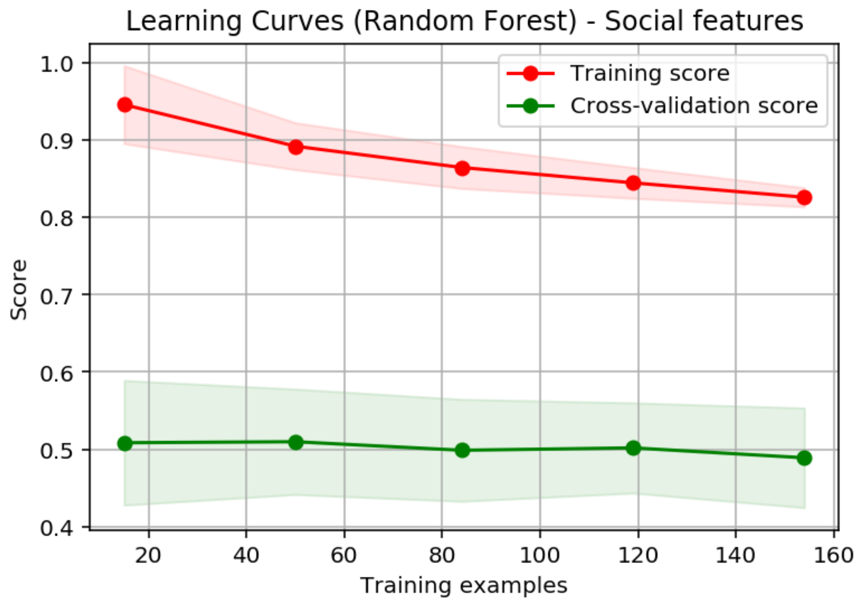

Figure 62: Learning Curve for Random Forest classifier with Textual & Social context features

Random Forest Classifier with only textual features (Model 4)

By looking at the below Learning curve, we can observe that the training score is very high at the beginning and stays the same, likely because of the overfitting on the training data. The cross-validation score is low at the beginning and increases. Cross-validation score which is a more reliable metric compared to training score converge to a value that is generally higher with increasing size of the training set, so in this case we will benefit much from more training data.

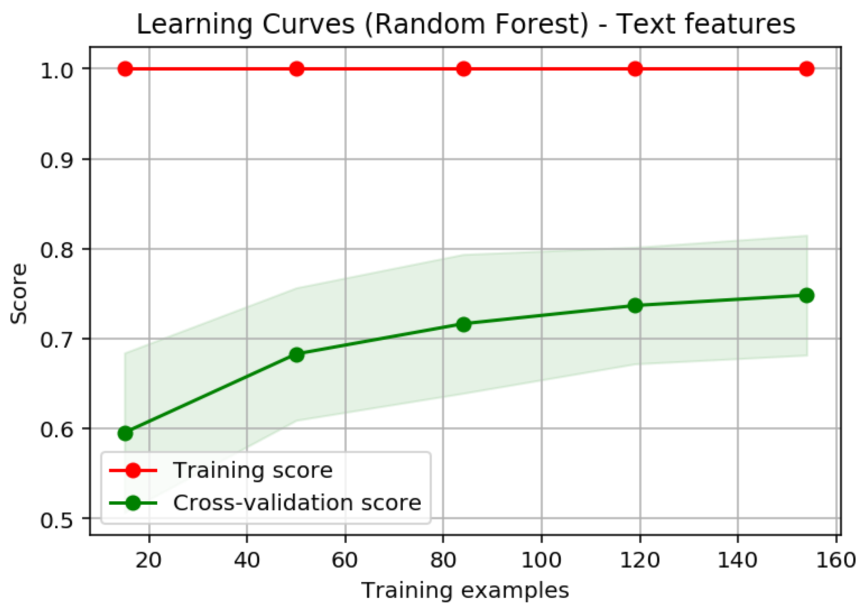

Figure 63: Learning Curve for Random Forest classifier with Textual features

Random Forest Classifier with both textual & social context features (Model 6)

Like the previous model, by looking at the below Learning curve, we can observe that the training score is very high at the beginning and stays the same, likely because of the overfitting on the training data. The cross-validation score is low at the beginning and increases. Cross-validation score which is a more reliable metric compared to training score converge to a value that is generally higher with increasing size of the training set, so in this case we will benefit much from more training data.

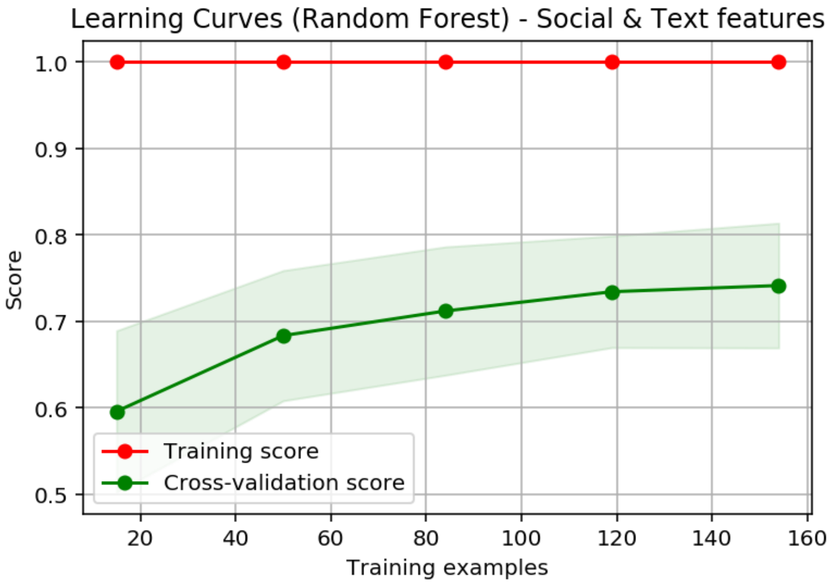

Figure 64: Learning Curve for Random Forest classifier with Textual and Social context features

- Testing Scenarios

To compare the below models, I will mainly consider the accuracy score.

Test Scenario 1: Social features with Naïve Bayes classifier (model 1) vs Text features with Naïve Bayes classifier (model 2)

<u>Model 1 results:</u>

	precision	recall	f1-score	support
True News	0.50	0.75	0.60	24
Fake News	0.45	0.22	0.29	23
accuracy			0.49	47
macro avg	0.48	0.48	0.45	47
weighted avg	0.48	0.49	0.45	47

<u>Model 2 results</u>:

	precision	recall	f1-score	support
True News	0.92	0.46	0.61	24
Fake News	0.63	0.96	0.76	23
accuracy			0.70	47
macro avg	0.77	0.71	0.68	47
weighted avg	0.78	0.70	0.68	47

We can observe that Model 2 performs better than Model 1.

Test Scenario 2: Social features with Random Forest classifier (model 3) vs Text features with Random Forest classifier (model 4)

<u>Model 3 results</u>:

	precision	recall	f1-score	support
True News	0.46	0.46	0.46	24
Fake News	0.43	0.43	0.43	23
accuracy			0.45	47
macro avg	0.45	0.45	0.45	47
weighted avg	0.45	0.45	0.45	47

<u>Model 4 results</u>:

	precision	recall	f1-score	support
True News	0.78	0.58	0.67	24
Fake News	0.66	0.83	0.73	23
accuracy			0.70	47
macro avg	0.72	0.70	0.70	47
weighted avg	0.72	0.70	0.70	47

Model 4 performs better than the Model 3.

Test Scenario 3: Social & Text features with Naïve Bayes Classifier (model 5) vs Text features with Naïve Bayes classifier

Model 5 results:

	precision	recall	f1-score	support
True News	0.73	0.46	0.56	24
Fake News	0.59	0.83	0.69	23
accuracy			0.64	47
macro avg	0.66	0.64	0.63	47
weighted avg	0.67	0.64	0.63	47

Model 2 results:

	precision	recall	f1-score	support
True News	0.92	0.46	0.61	24
Fake News	0.63	0.96	0.76	23
accuracy			0.70	47
macro avg	0.77	0.71	0.68	47
weighted avg	0.78	0.70	0.68	47

Model 5 performs better than the Model 2.

Test Scenario 4: Social & Text features with Random Forest classifier (model 6) vs Text features with Random Forest classifier

Model 6 results**:**

	precision	recall	f1-score	support
True News	0.78	0.58	0.67	24
Fake News	0.66	0.83	0.73	23
accuracy			0.70	47
macro avg	0.72	0.70	0.70	47
weighted avg	0.72	0.70	0.70	47

Model 4 results:

	precision	recall	f1-score	support
True News	0.78	0.58	0.67	24
Fake News	0.66	0.83	0.73	23
accuracy			0.70	47
macro avg	0.72	0.70	0.70	47
weighted avg	0.72	0.70	0.70	47

The two models perform equally.

Based on the above results, the most performant models are the Model 4 (Text features with Random Forest classifier), Model 6 (Text and Social Features with Random Forest classifier) and Model 2 (Text features with Naïve Bayes classifier).

I will compare those dominant models using the AUC score for which you will find more details in the next chapter where the evaluation metrics are presented.

- **Model 6**:

AUC score: 0.7164750957854406

ROC curve:

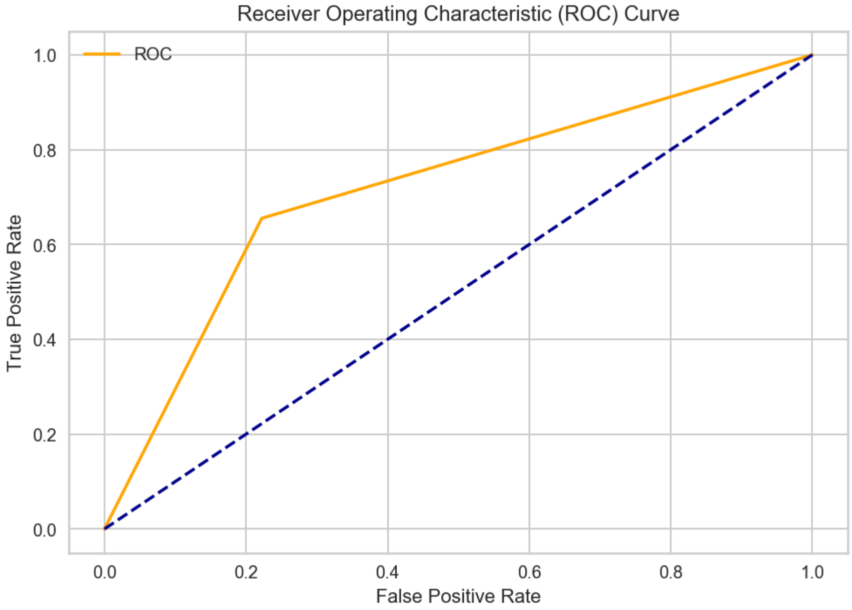

Figure 65: ROC Curve: Random Forest classifier with Textual and Social context Features

- **Model 4:**

AUC Score: 0.7164750957854406

ROC Curve:

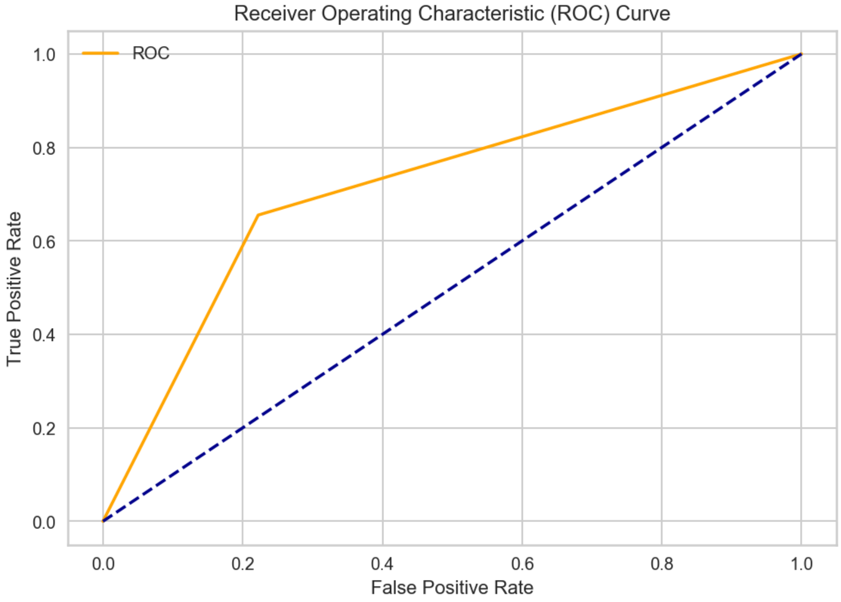

Figure 66: ROC Curve: Random Forest classifier with Textual Features

- **Model 2:**

AUC score: 0.7726190476190476
ROC curve:

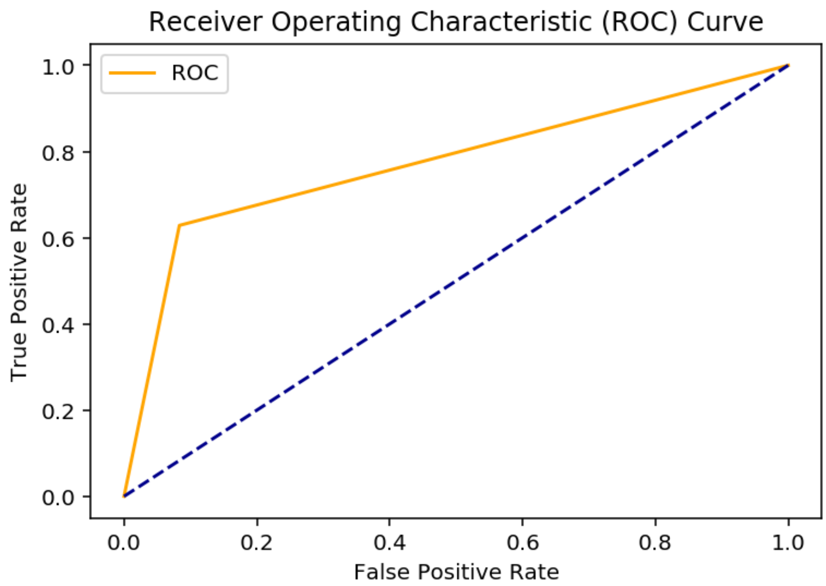

Figure 67: ROC Curve: Naïve Bayes classifier with Textual Features

The Naïve Bayes classifier with Text only features is more performant in terms of AUC score compared to Random Forest Classifier with only text data, thus I suppose that Naïve Bayes classifier performs better for models with only textual features.

By looking at the results of Random Forest classifier on the training set at the beginning of this section, we have observed that the more data we have in our dataset, the better. Therefore, even though the performance score is quite good but not very different from the other models, we have an indicator showing that with a bigger training set, the model performance can be higher. In addition to that, we have seen that the Random Forest Classifier has a higher accuracy score compared to Naïve Bayes classifier when the prediction is on the training set.

In the next section you will find some introduction elements regarding the metrics that have been used in this chapter to compare the models.

3.4.2.4 Evaluation Metrics

In this section, I will introduce you the different metrics that are necessary to evaluate model performance and take the good decisions based on them.

There is something called the "No Free Lunch" theorem said by Wolpert concerning supervised machine learning. The theorem implies that no algorithm performs well for any problem. This is commonly valid in Prediction Models where we train our dataset on an algorithm, and then we use the trained model for predictions on new data.

Performance indicators - Evaluation Metrics

In my analysis, in order to assess the performance of the Fake News detection classifier, I have been looking mainly to two evaluation metrics for which it has been demonstrated that they provide us with valuable information about the quality of the prediction. Those two are the Accuracy score[128] and the AUC score.[128]

Given that we are in a classification case, the results of our predictions can be categorized as True Positive, True Negative, False Negative, and False Positive. Those metrics are the components of the confusion matrix[129] which is the base of the calculation of the most commonly used performance metrics (e.g., Precision[128], Accuracy, Recall[128], AUC score)

AUC stands for "Area under the ROC Curve" and measures the entire two-dimensional area underneath the entire ROC curve[130] from (0,0) to (1,1). That curve plots the True Positive Rate and the False Positive Rate:

$$TPR = \frac{TP}{TP + FN}$$

$$FPR = \frac{FP}{FP + TN}$$

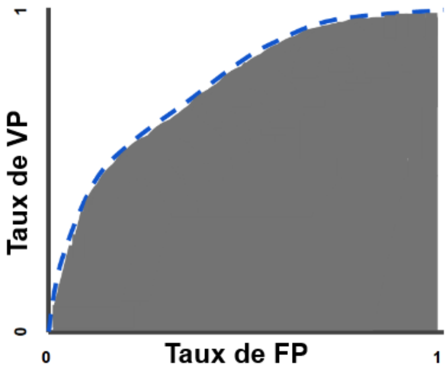

Figure 68 : AUC (aire sous la courbe ROC Curve)

Source: "Machine Learning Crash Course." *Google Developers*. Accessed August 31, 2019. Figure 5. https://developers.google.com/machine-learning/crash-course/classification/roc-and-auc.

Google Developers describe AUC, "*AUC provides an aggregate measure of performance across all possible classification thresholds. One way of interpreting AUC is as the probability that the model ranks a random positive example more highly than a random negative example…*"[128].

Accuracy score for binary classification is calculated in terms of positives and negatives as the number of correct predictions divided by the total number of predictions:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

It is worth also mentioning that according to researches AUC is more statistically consistent and more discriminating than accuracy[131].

3.4.3 Results and Discussion

The features selected across my analysis which could be the base for further research on the Fake News detection topic are the Textual and Social-based features with the use of the Random Forest Classifier. You can find below the list of the selected features (their names and the names of the variables in the dataset) and their category.

Social Context Features:

User Influence

- Eigenvector Centrality (“Eigenvector Centrality”)
- Page Rank (“page_rank”)

Textual Features:

Lexical Features

- Number of Words in the text (count_word)
- Number of Unique Words in the text (count_word_unique)

Syntactic Features

- Average sentence length (“average_sentence_length”)
- Cardinal Digit - Pos Tagging (“CD”)
- Possessive ending parent’s - Pos Tagging (“POS”)
- Lexical Diversity (“lexical_diversity_ttr”)

Lexical Features (Count Vetorizer and Tfidf features)

- The word “Donald” (“Donald_count_vect”)
- The word “New” (“New_count_vect”)
- The word “New York” (“New York_count_vect”)
- The word “President” (“President_count_vect”)
- The word “Republican” (“Republican_count_vect”)
- The word “Trump” (“Trump_count_vect”)
- The word “York” (“York_count_vect”)
- The word “debate” (“debate_count_vect”)
- The word “presidential” (“presidential_count_vect”)
- The word “said” (“said_count_vect”)
- The word “clinton” (tfidf feature)

- The word “country” (tfidf feature)
- The word “debate” (tfidf feature)
- The word “did” (tfidf feature)
- The word “donald” (tfidf feature)
- The word “hillary” (tfidf feature)
- The word “just” (tfidf feature)
- The word “know” (tfidf feature)
- The word “like” (tfidf feature)
- The word “make” (tfidf feature)
- The word “new” (tfidf feature)
- The word “news” (tfidf feature)
- The word “president” (tfidf feature)
- The word “presidential” (tfidf feature)
- The word “republican” (tfidf feature)
- The word “said” (tfidf feature)
- The word “states” (tfidf feature)
- The word “time” (tfidf feature)
- The word “trump” (tfidf feature)
- The word “white” (tfidf feature)
- The word “York” (tfidf feature)
- The word “xe2x80x94” ('xe2x80x94_count_vect')
- The word “itxe2x80x99s” (tfidf feature)
- The word “xe2x80x93” (tfidf feature)

The performance of the Random Forest classifier model that combines Textual and Social-based features gave promising results. The accuracy of the prediction on the training data was ~98% and ~70% on test data, with an AUC ROC score of ~98% and ~71% on training and test sets respectively.

The performance of the model that includes only Textual features gave as good results as the hybrid model.

The model that contains only Social features did not give the results we were expecting as the prediction accuracy on the test set was only ~50%. Improving the dataset by adding data related to News propagation on social media could be an efficient way to improve the latter results. As also mentioned in the theoretical part of my master thesis, researches have shown that Fake News propagates more rapidly than True News articles. Thus, it would be interesting to have the necessary data to extract features that allow analysing News propagation over time.

In addition to that, having more News sources variety on our dataset, it could improve the accuracy of the model. Different sources mean also people (users) that read articles from these sources and may have interesting behaviours to analyse. Having a large variety of sources may help to identify distinguishing communities of users, and from that extract significant features. Moreover, including demographic or interest's information about the Users that shared the articles (cf. chapter 2.3.2.3 "User-based features methods") could also be a way improve the model.

Also, it is worthy of mentioning again that the Random Forest classifier gave significantly better results when we look notably at the predictions on the training data (cf. "3.9.2.2 Test Scenarios"). Although the very good results obtained with the hybrid model, the application of this method on the present dataset does not improve significantly the performance of the model comparing to the one that contains Textual only features. That can be the object of further study and analysis. This topic has already been considered as very promising by the research community.

My future work can focus on the improvement of the feature extraction and selection but also on the modelling part by implementing tuning for parameter selection and cross-validation techniques. Those two methodologies can improve the quality of the prediction significantly and give more objective evaluation results. Another measure of improvement that I mentioned previously, could be to improve the dataset. That stands for answering the following questions: how do we measure our dataset's quality and improve it? And how much data do we need to get useful results? For example, we have seen that for Random Forest Classifier the dataset size plays a crucial role in the prediction performance.

4. Conclusion

The research on machine learning techniques aiming to assess the news integrity is still in the early phases, although there is a growing body of research and bibliography on this topic. As more and more individuals choose social media as a primary source of information, it becomes crucial to have reliable detection systems in place to support preventing disinformation spreads.

Many researchers concentrate their work on evaluating the different existing methodologies to realize Fake News detection automatically using classification machine learning techniques, and others create their techniques and approaches.

As stated, and discussed in previous chapters, algorithms that include a combination of content and social-based features outperformed Fake News detection in several pieces of research. It is about experimental results on Fake News datasets of real articles that have proven the effectiveness of the proposed methodology.

In this context, each data type has its one advantage and contributes with its way to improve the prediction of the news veracity. On the one hand, for predicting Fake News before it has been propagated on social media (at an early stage), the textual data plays a critical role. On the other hand, social-based features contribute to avoiding bias that can occur when Fake News is intentionally written to mislead the reader.

Thus, both techniques have their advantages and disadvantages, and their combination makes the balance.

5. References

[1] “Truth.” *The dictionary by Merriam-Webster is America's most trusted online dictionary for English word definitions, meanings, and pronunciation*. Retrieved from https://www.merriam-webster.com/dictionary/truth

[2] Wikipedia contributors. “Truth.” *Wikipedia The Free Encyclopaedia*. https://en.wikipedia.org/wiki/Truth

[3] High Representative of the Union for Foreign Affairs and Security Policy. “Action Plan against Disinformation.” Brussels: European Commision, 2018. https://ec.europa.eu/digital-single-market/en/news/action-plan-against-disinformation

[4] Forrest Stroud. “What Is Fake News?” *Webopedia - Online Tech Dictonnary for Tech Professionals*. Accessed August 27, 2019. https://www.webopedia.com/TERM/F/fake-news.html.

[5] Angie Drobnic Holan. “2016 Lie of the Year: Fake News.” *PolitiFact*, December 13, 2016. https://www.politifact.com/truth-o-meter/article/2016/dec/13/2016-lie-year-fake-news/.

[6] Wikipedia contributors. “Fake News.” In *Wikipedia The Free Encyclopaedia*. https://en.wikipedia.org/wiki/Fake_news.

[7] Özgöbek, Özlem and Jon Atle Gulla. “Towards an Understanding of Fake News.” (2018)

[8] Lemaistre, Clémence. “Les Fake News, Une Multitude D'intentions.” *Les Echos*, February 26, 2018. https://www.lesechos.fr/2018/02/les-fake-news-une-multitude-dintentions-985291.

[9] Tandoc, Edson & Wei Lim, Zheng & Ling, Rich. “Defining “Fake News”: A typology of scholarly definitions.” *Digital Journalism*. (2017): 1-17.

[10] “About.” First Draft. Accessed August 17, https://firstdraftnews.org/about/.

[11] Wikipedia contributors. “Fake News.” *Wikipedia The Free Encyclopaedia*. https://en.wikipedia.org/wiki/Fake_news.

[12] Wikipedia contributors. “First Draft News.” *Wikipedia The Free Encyclopaedia*. https://en.wikipedia.org/wiki/First_Draft_News

[13] Potthast, Martin, Johannes Kiesel, Kevin Reinartz, Janek Bevendorff and Benno Stein. “A Stylometric Inquiry into Hyperpartisan and Fake News.” *ACL* (2017).

[14] Marchi, Regina. “With Facebook, Blogs, and Fake News, Teens Reject Journalistic ‘Objectivity.’” Journal of *Communication Inquiry 36* (2012): 246–262.

[15] Shao, Chengcheng, Giovanni Luca Ciampaglia, Onur Varol, Kai-Cheng Yang, Alessandro Flammini and Filippo Menczer. “The spread of low-credibility content by social bots.” *Nature Communications* (2018).

[16] Ribeiro, Manoel Horta, Pedro H. Calais, Virgílio A. F. Almeida and Wagner Meira. “Everything I Disagree With is #FakeNews”: Correlating Political Polarization and Spread of Misinformation.” *ArXiv abs/1706.05924 (2017)*: 1-3.

[17] Pérez-Rosas, Verónica, Bennett Kleinberg, Alexandra Lefevre and Rada Mihalcea. “Automatic Detection of Fake News.” *COLING* (2017).

[18] Ruchansky, Natali, Sungyong Seo and Yan Liu. “CSI: A Hybrid Deep Model for Fake News Detection.” *CIKM* (2017).

[19] Conroy, Niall, Victoria L. Rubin and Yimin Chen. “Automatic deception detection: Methods for finding Fake News.” *ASIST* (2015).

[20] Parikh, Shivam B. and Praveen Kumar Atrey. “Media-Rich Fake News Detection: A Survey.” *2018 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR)* (2018): 436-441.

[21] Zhou, Xinyi and Reza Zafarani. “Fake News: A Survey of Research, Detection Methods, and Opportunities.” *ArXiv abs/1812.00315* (2018).

[22] Bessi, Alessandro, and Emilio Ferrara. "Social bots distort the 2016 US Presidential election online discussion." *First Monday* 21, no. 11-7 (2016).

[23] “Resource for Understanding Political Bots.” *The Computational Propaganda Project*, November 18, 2016. https://comprop.oii.ox.ac.uk/research/public-scholarship/resource-for-understanding-political-bots/.

[24] Allcott, Hunt, and Matthew Gentzkow. "Social media and Fake News in the 2016 election." *Journal of economic perspectives* 31, no. 2 (2017): 211-36.

[25] Shearer, Elisa, and Jeffrey Gottfried. “News Use Across Social Media Platforms 2016.” *Pew Research Center's Journalism Project*, December 27, 2017. https://www.journalism.org/2016/05/26/news-use-across-social-media-platforms-2016/.

[26] Silverman, Craig. “This Analysis Shows How Viral Fake Election News Stories Outperformed Real News On Facebook.” *BuzzFeed News*, November 16, 2016. https://www.buzzfeednews.com/article/craigsilverman/viral-fake-election-news-outperformed-real-news-on-facebook.

[27] Silverman, Craig. “The True Story Behind The Biggest Fake News Hit Of The Election.” *BuzzFeed News*, December 16, 2016. https://www.buzzfeednews.com/article/craigsilverman/the-strangest-fake-news-empire.

[28] Parkinson, Hannah Jane. "Click and Elect: How Fake News Helped Donald Trump Win a Real Election | Hannah Jane Parkinson." *The Guardian*, November 14, 2016. https://www.theguardian.com/commentisfree/2016/nov/14/fake-news-donald-trump-election-alt-right-social-media-tech-companies.

[29] Read, Max. "Donald Trump Won Because of Facebook." *Intelligencer,* November 9, 2016. http://nymag.com/intelligencer/2016/11/donald-trump-won-because-of-facebook.html.

[30] Dewey, Caitlin. "Facebook Fake-News Writer: 'I Think Donald Trump Is in the White House Because of Me.'" *The Washington Post*, November 17, 2016. https://www.washingtonpost.com/gdpr-consent/?destination=/news/the-intersect/wp/2016/11/17/facebook-fake-news-writer-i-think-donald-trump-is-in-the-white-house-because-of-me/

[31] Wikipedia contributors. "Colin Powell." *Wikipedia, the free encyclopaedia*. https://en.wikipedia.org/wiki/Colin_Powell.

[32] Rutenberg, Jim. "A Lesson in Moscow About Trump-Style 'Alternative Truth'." *The New York Times,* April 17, 2017. https://www.nytimes.com/2017/04/16/business/media/vladimir-putin-moscow-press-trump.html.

[33]Donald Trump. Twitter Post. May 9, 2018, 04 :38 a.m. //twitter.com/realDonaldTrump/status/994179864436596736

[34] Samuels, Brett. "Trump Says 'Fake News Has Contributed Greatly to the Anger and Rage' after Shootings." *The Hill*, August 5, 2019. https://thehill.com/homenews/administration/456142-trump-says-fake-news-has-contributed-greatly-to-the-anger-and-rage.

[35]Donald Trump. Twitter Post. August 5, 2019, 04:32 a.m. https://twitter.com/realdonaldtrump/status/1158340095608610816

[36] "La Fabrique Du Mensonge Brexit Et Micro-Ciblage Publicitaire." *France.tv - Programmes et replay des chaînes France Télévisions*. Video File. March 19, 2019. https://www.france.tv/france-5/la-fabrique-du-mensonge/la-fabrique-du-mensonge-saison-1/943623-extrait-brexit.html.

[37] "La fabrique du mensonge Emmanuel Macron : sa campagne entachée par les Russes ?" *France.tv - Programmes et replay des chaînes France Télévisions*. Video File. April 5, 2019. https://www.france.tv/france-5/la-fabrique-du-mensonge/956303-emmanuel-macron-sa-campagne-entachee-par-les-russes-la-fabrique-du-mensonge.html

[38] Farand, Chloe. "French Presidential Frontrunner Emmanuel Macron 'Targeted by Russian-Controlled State Media.'" *The Independent,* February 13, 2017. https://www.independent.co.uk/news/world/europe/emmanuel-macron-russia-le-pen-en-march-french-elections-cyber-attacks-fake-news-a7578011.html.

[39] Renda, Andrea. “The Legal Framework to Address ‘Fake News.’” *CEPS (Centre for European Policy Studies)*. European Parliament, (2018): 22-23. https://www.ceps.eu/ceps-publications/legal-framework-address-fake-news-possible-policy-actions-eu-level/.

[40] “Communication - Tackling Online Disinformation: a European Approach.” *Digital Single Market - European Commission*. *The Directorate-General for Communications Networks, Content and Technology (DG CONNECT)*, April 26, 2018. https://ec.europa.eu/digital-single-market/en/news/communication-tackling-online-disinformation-european-approach.

[41] “Statement by Commissioner Gabriel on the Code of Practice on Online Disinformation.” *European Commission - Press Releases*, September 26, 2018. https://europa.eu/rapid/press-release_STATEMENT-18-5914_en.htm.

[42] “Best Practices - Code of Practice on Disinformation.” *Digital Single Market - European Commission*. *The Directorate-General for Communications Networks, Content and Technology*, (2018): 1-14.

[43] “EU Code of Practice on Disinformation.” *Digital Single Market - European Commission*. *The Directorate-General for Communications Networks, Content and Technology*, (2018): 1-12.

[44] “Fake News: a Bill to Combat the Manipulation of Information.” *Gouvernement.fr*, June 7, 2018. https://www.gouvernement.fr/en/fake-news-a-bill-to-combat-the-manipulation-of-information.

[45] Bouchenard, Julia Darcel Claire. “New Reporting Duties on Seller of Digital Ad Spaces in France.” *Osborne Clark*, April 27, 2017. https://marketinglaw.osborneclarke.com/advertising-regulation/new-reporting-duties-seller-digital-ad-spaces-france/.

[46] General Watin-Augouard, Marc. “Law No. 2018-1202 of 22 December 2018 on Combating the Manipulation of Information.” *FIC Observatory*, January 11, 2019. https://observatoire-fic.com/en/law-no-2018-1202-of-22-december-2018-on-combating-the-manipulation-of-information-by-general-marc-watin-augouard-founder-of-the-fic/.

[47] Fiorentino, Michael-Ross. “France Passes Controversial 'Fake News' Law.” *Euronews*, November 22, 2018. https://www.euronews.com/2018/11/22/france-passes-controversial-fake-news-law.

[48] Wikipedia contributors. “Intentional Infliction of Emotional Distress.” *Wikipedia. the free encyclopaedia*. https://en.wikipedia.org/wiki/Intentional_infliction_of_emotional_distress.

[49] Wikipedia contributors. “Tortious Interference.” *Wikipedia. the free encyclopaedia*. https://en.wikipedia.org/wiki/Tortious_interference.

[50] Garance, Mathias. “What Is Fake News? Can It Be Regulated?” *Mathias Avocats*, January 18, 2018. https://www.avocats-mathias.com/actualites/fake-news-regulation.

[51] France-Presse, Agence. “Zuckerberg Says U.S. Government Inaction Allowed Fake News to Spread.” *Rappler*, June 27, 2019. https://www.rappler.com/technology/news/234053-zuckerberg-us-government-inaction-let-fake-news-spread.

[52] Shu, Kai, Amy Sliva, Suhang Wang, Jiliang Tang, and Huan Liu. "Fake news detection on social media: A data mining perspective." *ACM SIGKDD Explorations Newsletter* 19, no. 1 (2017): 22-36.

[53] Reed, Edward S., Elliot Turiel and Terrance Brown. “Naive Realism in Everyday Life: Implications for Social Conflict and Misunderstanding.” *The Jean Piaget symposium series. Values and knowledge* (1997): 103-135.

[54] Shu, Kai, Suhang Wang, Jiliang Tang, Reza Zafarani and Huan Liu. “User Identity Linkage across Online Social Networks: A Review.” *SIGKDD Explorations 18* (2016): 5-17.

[55] Farajtabar, Mehrdad, Jiachen Yang, Xiangzhou Ye, Huan Xu, Rakshit Trivedi, Elias Boutros Khalil, Shuang Li, Le Song and Hongyuan Zha. “Fake News Mitigation via Point Process Based Intervention.” *ArXiv abs/1703.07823* (2017): 1-18.

[56] Vicario, Michela Del, Alessandro Bessi, Fabiana Zollo, Fabio Petroni, Antonio Scala, Guido Caldarelli, Harry Eugene Stanley and Walter Quattrociocchi. “The spreading of misinformation online.” *Proceedings of the National Academy of Sciences of the United States of America 113 (3)* (2016): 554-559.

[57] Menczer, Filippo. “The spread of misinformation in social media.” *WWW* (2016): 717-717.

[58] Houvardas, John and Efstathios Stamatatos. “N-Gram Feature Selection for Authorship Identification.” *AIMSA* (2006): 77–86.

[59] Feng, Song, Ritwik Banerjee, and Yejin Choi. "Syntactic stylometry for deception detection." *Association for Computational Linguistics* (2012): 171-175

[60] Wikipedia contributors. “Word Embedding.” *Wikipedia. the free encyclopedia*. https://fr.wikipedia.org/wiki/Word_embedding.

[61] “Deep Neural Network”. *Technopedia Dictionary*. Retrieved from: https://www.techopedia.com/definition/32902/deep-neural-network

[62] Wang, Suhang, Jiliang Tang, Charu C. Aggarwal and Huan Liu. “Linked Document Embedding for Classification.” *CIKM* (2016).

[63] Jin, Zhiwei, Juan Cao, Yongdong Zhang, Jianshe Zhou and Qi Tian. “Novel Visual and Statistical Image Features for Microblogs News Verification.” *IEEE Transactions on Multimedia 19* (2017): 598-608.

[64] Tang, Jian, Meng Qu, Mingzhe Wang, Mingjie Zhang, Jun Yan and Qiaozhu Mei. "LINE: Large-scale Information Network Embedding." *WWW* (2015): 1067-1077.

[65] Wang, Suhang, Jiliang Tang, Charu C. Aggarwal, Yi Chang and Huan Liu. "Signed Network Embedding in Social Media." *SDM* (2017).

[66] Zhou, Xinyi and Reza Zafarani. "Network-based Fake News Detection: A Pattern-driven Approach." ArXiv abs/1906.04210 (2019)

[67] Goasdoué, François, Konstantinos Karanasos, Yannis Katsis, Julien Leblay, Ioana Manolescu and Stamatis Zampetakis. "Fact checking and analyzing the web." *SIGMOD Conference* (2013): 997-1000.

[68] Cohen, Sarah, James T. Hamilton and Fred Turner. "Computational journalism." Commun. *ACM 54* (2011): 66-71.

[69] Nickel, Maximilian, Kevin Murphy, Volker Tresp and Evgeniy Gabrilovich. "A Review of Relational Machine Learning for Knowledge Graphs." *Proceedings of the IEEE 104* (2015): 11-33.

[70] Pawar, Sachin, Girish Keshav Palshikar and Pushpak Bhattacharyya. "Relation Extraction: A Survey." *ArXiv abs/1712.05191* (2017).

[71] Nickel, Maximilian, Kevin Murphy, Volker Tresp, and Evgeniy Gabrilovich. "A review of relational machine learning for knowledge graphs." *Proceedings of the IEEE* 104, no. 1 (2015): 11-33.

[73] Chen, Yimin, Niall Conroy and Victoria L. Rubin. "Misleading Online Content: Recognizing Clickbait as "False News"." *In Proceedings of the 2015 ACM on Workshop on Multimodal Deception Detection* (2015): 15-19.

[74] Amado, Bárbara G., Ramón Arce, and Francisca Fariña. "Undeutsch hypothesis and Criteria Based Content Analysis: A meta-analytic review." *The European Journal of Psychology Applied to Legal Context* 7, no. 1 (2015): 3-12.

[75] Vrij, Aldert. "Criteria-Based Content Analysis: A Qualitative Review of the First 37 Studies." *Psychology, Public Policy, and Law* 11, no. 1 (2005): 3.

[76] Lesce, Tony. "SCAN: Deception detection by scientific content analysis." *Law and Order* 38, no. 8 (1990): 3-6.

[77] Rubin, Victoria L., and Tatiana Lukoianova. "Truth and deception at the rhetorical structure level." *Journal of the Association for Information Science and Technology* 66, no. 5 (2015): 905-917.

[78] Kucharski, Adam. "Post-truth: Study epidemiology of Fake News." *Nature* 540, no. 7634 (2016): 525.

[79] Keeling, Matt J., and Ken TD Eames. "Networks and epidemic models." *Journal of the Royal Society Interface* 2, no. 4 (2005): 295-307.

[80] Vosoughi, Soroush, Deb Roy, and Sinan Aral. "The spread of true and false news online." *Science* 359, no. 6380 (2018): 1146-1151.

[81] Jin, Zhiwei, Juan Cao, Yongdong Zhang and Jiebo Luo. “News Verification by Exploiting Conflicting Social Viewpoints in Microblogs.” *AAAI* (2016).

[82] Sharma, Karishma, Feng Qian, He Jiang, Natali Ruchansky, Ming Zhang, and Yan Liu. "Combating fake news: A survey on identification and mitigation techniques." *ACM Transactions on Intelligent Systems and Technology (TIST)* 10, no. 3 (2019): 21.

[83] Shu, Kai, Suhang Wang, and Huan Liu. "Exploiting tri-relationship for Fake News detection." *arXiv preprint arXiv:1712.07709* (2017).

[84] Zhang, Jiawei, Limeng Cui, Yanjie Fu and Fisher B. Gouza. “Fake News Detection with Deep Diffusive Network Model.” *ArXiv abs/1805.08751* (2018).

[85] Ma, Jing, Wei Gao and Kam-Fai Wong. “Rumor Detection on Twitter with Tree-structured Recursive Neural Networks.” *ACL* (2018): 1980-1989.

[86] Mukherjee, Arjun, Abhinav Kumar, Bing Liu, Junhui Wang, Meichun Hsu, Malú Castellanos and Riddhiman Ghosh. “Spotting opinion spammers using behavioural footprints.” *KDD* (2013): 632-640.

[87] Xie, Sihong, Guan Wang, Shuyang Lin and Philip S. Yu. “Review spam detection via temporal pattern discovery.” *KDD* (2012): 823-831

[88] Singh, Vivek, Rupanjal Dasgupta, Darshan Sonagra, Karthik Raman and, Isha Ghosh. “Automated Fake News detection using linguistic analysis and machine learning.” I*n International Conference on Social Computing, Behavioral-Cultural Modeling, & Prediction and Behavior Representation in Modeling and Simulation (SBPBRiMS)* (2017): 1–3.

[89] Vukovic, Marin, Kresimir Pripuzic and Hrvoje Belani. “An Intelligent Automatic Hoax Detection System.” *KES* (2009): 318-325.

[90] Sharifi, Mehrbod, Eugene Fink and Jaime G. Carbonell. “Detection of Internet scam using logistic regression.” *2011 IEEE International Conference on Systems, Man, and Cybernetics* (2011): 2168-2172.

[91] Horne, Benjamin D. and Sibel Adali. “This Just In: Fake News Packs a Lot in Title, Uses Simpler, Repetitive Content in Text Body, More Similar to Satire than Real News.” *ArXiv abs/1703.09398* (2017).

[92] Elmurngi, Elshrif, and Abdelouahed Gherbi. "Detecting fake reviews through sentiment analysis using machine learning techniques." *IARIA/data analytics* (2017): 65-72.

[93] Volkova, Svitlana and Jin Yea Jang. "Misleading or Falsification: Inferring Deceptive Strategies and Types in Online News and Social Media." *WWW* (2018): 575-583.

[94] Boididou, Christina, Katerina Andreadou, Symeon Papadopoulos, Duc-Tien Dang-Nguyen, Giulia Boato, Michael Riegler and Yiannis Kompatsiaris. "Verifying Multimedia Use at MediaEval 2016." *MediaEval* (2015).

[95] Jin, Zhiwei, Juan Cao, Yongdong Zhang, Jianshe Zhou and Qi Tian. "Novel Visual and Statistical Image Features for Microblogs News Verification." *IEEE Transactions on Multimedia 19* (2017): 598-608.

[96] Jin, Zhiwei, Juan Cao, Yongdong Zhang, and Jiebo Luo. "News verification by exploiting conflicting social viewpoints in microblogs." In *Thirtieth AAAI Conference on Artificial Intelligence* (2016).

[97] Ma, Jing, Wei Gao, Zhongyu Wei, Yueming Lu and Kam-Fai Wong. "Detect Rumors Using Time Series of Social Context Information on Microblogging Websites." *CIKM* (2015): 1751-1754.

[98] Castillo, Carlos, Marcelo Mendoza and Barbara Poblete. "Information credibility on twitter." *WWW* (2011): 675-676.

[99] Kwon, Sejeong, Meeyoung Cha, Kyomin Jung, Wei James Chen and Yajun Wang. "Prominent Features of Rumor Propagation in Online Social Media." *2013 IEEE 13th International Conference on Data Mining* (2013): 1103-1108.

[100] Hu, Xia, Qiaoyu Tan, and Ninghao Liu. "Deep Representation Learning for Social Network Analysis." *Frontiers in Big Data* 2 (2019): 2

[101] Shu, Kai, Suhang Wang, and Huan Liu. "Understanding user profiles on social media for Fake News detection." In *2018 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR)*, IEEE (2018): 430-435.

[102] Gumelar, Gumgum, Zarina Akbar, and Erik Erik. "Engagement and the spread of Fake News: Personality Trait as moderator." In *International Conference of Communication Science Research (ICCSR 2018)*. Atlantis Press (2018).

[103] Kwon, Sejeong, Meeyoung Cha, Kyomin Jung, Wei Chen, and Yajun Wang. "Prominent features of rumor propagation in online social media." In *2013 IEEE 13th International Conference on Data Mining*, IEEE (2013): 1103-1108.

[104] Shu, Kai, Suhang Wang, and Huan Liu. "Beyond news contents: The role of social context for Fake News detection." In *Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining*, ACM (2019): 312-320.

[105] Wikipedia contributors. “Data Preparation.” *Wikipedia the free encyclopaedia*. https://en.wikipedia.org/wiki/Data_preparation.

[106] Lohr, Steve. “For Big-Data Scientists, 'Janitor Work' Is Key Hurdle to Insights.” *The New York Times*, August 18, 2014. https://www.nytimes.com/2014/08/18/technology/for-big-data-scientists-hurdle-to-insights-is-janitor-work.html.

[107] “Community Detection in Social Networks.” *Kernix. Digital factory and Data Lab*, November 7, 2016. https://www.kernix.com/blog/community-detection-in-social-networks_p13.

[108] Blondel, Vincent D., Jean-Loup Guillaume, Renaud Lambiotte, and Etienne Lefebvre. "Fast unfolding of communities in large networks." *Journal of statistical mechanics: theory and experiment* 2008, no. 10 (2008).

[109] Srebrenik, Brian. “Ego Network Analysis for the Detection of Fake News.” Medium. Towards Data Science, February 20, 2019. https://towardsdatascience.com/ego-network-analysis-for-the-detection-of-fake-news-da6b2dfc7c7e.

[110] “EigenCentrality & PageRank.” *Cambridge Intelligence*, May 2, 2019. https://cambridge-intelligence.com/eigencentrality-pagerank/.

[111] Rachiele, Gianpaul. “Tokenization and Parts of Speech (POS) Tagging in Python's NLTK Library.” *Medium*, February 19, 2018. https://medium.com/@gianpaul.r/tokenization-and-parts-of-speech-pos-tagging-in-pythons-nltk-library-2d30f70af13b.

[112] Hafsa, Jabeen. “Stemming and Lemmatization in Python.” DataCamp Community, October 13, 2018. https://www.datacamp.com/community/tutorials/stemming-lemmatization-python.

[113]García, Clara, Juan Pedro Cabanilles, and Benjamín Ramirez. “What Is the Difference between Stemming and Lemmatization?” *Bitext. We help AI understand humans. Artificial Training for Chatbots*, February 28, 2018. https://blog.bitext.com/what-is-the-difference-between-stemming-and-lemmatization/.

[114] Wikipedia contributors. “Tf–Idf.” Wikipedia. the free encyclopaedia. https://en.wikipedia.org/wiki/Tf–idf.

[115] Tripathi, Mayank. “How to Process Textual Data Using TF-IDF in Python.” *Free Code Camp*, June 24, 2018. https://www.freecodecamp.org/news/how-to-process-textual-data-using-tf-idf-in-python-cd2bbc0a94a3/.

[117] Rubin, Victoria L., Yimin Chen, and Niall J. Conroy. "Deception detection for news: three types of fakes." In *Proceedings of the 78th ASIS&T Annual Meeting: Information Science with Impact: Research in and for the Community*. American Society for Information Science (2015): 83

[118] Otter, Daniel W., Julian R. Medina, and Jugal K. Kalita. "A survey of the usages of deep learning in natural language processing." *arXiv preprint arXiv:1807.10854* (2018).

[119] Wikipedia contributors. "Lexical Diversity." *Wikipedia the free encyclopaedia*. https://en.wikipedia.org/wiki/Lexical_diversity.

[120] Rai, Ankit. "Python: Sentiment Analysis Using VADER." *GeeksforGeeks*, January 23, 2019. https://www.geeksforgeeks.org/python-sentiment-analysis-using-vader/.

[121] Pandey, Parul. "Simplifying Sentiment Analysis Using VADER in Python (on Social Media Text)." *Medium*. *Analytics Vidhya*, August 13, 2019. https://medium.com/analytics-vidhya/simplifying-social-media-sentiment-analysis-using-vader-in-python-f9e6ec6fc52f

[122] Yip, Jason. "Algorithmic Trading Using Sentiment Analysis on News Articles." *Medium. Towards Data Science*, November 25, 2018. https://towardsdatascience.com/https-towardsdatascience-com-algorithmic-trading-using-sentiment-analysis-on-news-articles-83db77966704.

[123] Shu, Kai, Deepak Mahudeswaran, Suhang Wang, Dongwon Lee, and Huan Liu. "Fakenewsnet: A data repository with news content, social context and dynamic information for studying Fake News on social media." *arXiv preprint arXiv:1809.01286* (2018).

[124] Saxena, Rahul. "How the Naive Bayes Classifier works in Machine Learning." *Data Science, Machine Learning* (2017).

[125] Yiu, Tony. "Understanding Random Forest." *Medium. Towards Data Science*, June 12, 2019. https://towardsdatascience.com/understanding-random-forest-58381e0602d2.

[126] Dubey, Akash. "Feature Selection Using Random Forest." Medium. Towards Data Science, December 15, 2018. https://towardsdatascience.com/feature-selection-using-random-forest-26d7b747597f.

[127] Genuer, Robin, Jean-Michel Poggi, and Christine Tuleau-Malot. "Variable selection using random forests." *Pattern Recognition Letters* 31, no. 14 (2010): 2225-2236.

[128] "Machine Learning Crash Course." *Google Developers*. Accessed August 31, 2019. https://developers.google.com/machine-learning.

[129] Bendersky, Eli. "The Confusion Matrix in Statistical Tests." *DZone Big Data*, March 28, 2018. https://dzone.com/articles/the-confusion-matrix-in-statistical-tests.

[130] Brownlee, Jason. "How and When to Use ROC Curves and Precision-Recall Curves for Classification in Python." *Machine Learning Mastery*, August 31, 2018. https://machinelearningmastery.com/roc-curves-and-precision-recall-curves-for-classification-in-python/.

[131] Ling, Charles X., Jin Huang, and Harry Zhang. "AUC: a statistically consistent and more discriminating measure than accuracy." *Ijcai*, vol. 3 (2003): 519-524