EXAMENSARBETE INOM TEKNIK,
GRUNDNIVÅ, 15 HP

STOCKHOLM, *SVERIGE 2017*

COMBATING DISINFORMATION

Detecting fake news with linguistic models and classification algorithms

PHILIP RUMMAN

MIKAEL SVÄRD

KTH
SKOLAN FÖR DATAVETENSKAP OCH KOMMUNIKATION

Abstract

The purpose of this study is to examine the possibility of accurately distinguishing fabricated news from authentic news stories using Naive Bayes classification algorithms, this involves a comparative study of two different machine learning classification algorithms. The work also contains an overview of how linguistic text analytics can be utilized for detection purposes and an attempt to extract interesting information was made using Word Frequencies. A discussion of how different actors and parties in businesses and governments are affected by, and how they handle deception caused by fake news articles is also included in this work. The study further tries to ascertain what collective steps could be made towards introducing a functioning solution to combat fake news. The results were inconclusive and the simple Naive Bayes algorithms used did not yield fully satisfactory results. Word frequencies alone did not give enough information for detection. They were however found to be potentially useful as part of a larger set of algorithms and strategies as part of a solution to handling of misinformation.

Contents

1 Introduction **1**
- 1.1 Problem Definition . . . 1
- 1.2 Scope and Constraints . . . 2

2 Background **2**
- 2.1 Definition: Fake News . . . 3
- 2.2 More than one type of fake . . . 3
- 2.3 History . . . 5
- 2.4 Impact . . . 6
  - 2.4.1 Businesses . . . 7
  - 2.4.2 Government and National security . . . 13

3 Theoretical Framework **14**
- 3.1 Machine Learning . . . 14
  - 3.1.1 Broad Categories . . . 14
  - 3.1.2 Training and test partitions . . . 15
  - 3.1.3 Classification and Regression . . . 16
  - 3.1.4 Support vector machines . . . 17
  - 3.1.5 Naive Bayes . . . 17
  - 3.1.6 Evaluation Methods . . . 18
- 3.2 Text Analytics . . . 19
  - 3.2.1 Text analytics and machine learning . . . 19
  - 3.2.2 Requirements for a fake news detection corpus . . . 19
  - 3.2.3 N-gram . . . 20
  - 3.2.4 Bag-of-words . . . 20
  - 3.2.5 Deep Syntax . . . 20
- 3.3 Previous studies . . . 21

4 Methodology **21**
- 4.1 Overview . . . 21
- 4.2 Data Gathering . . . 22
- 4.3 Selection of classification algorithms . . . 22
- 4.4 Evaluation . . . 22
- 4.5 Present and compare results . . . 23

5 Data and Results **23**

6 Discussion and conclusion **25**
- 6.1 Error sources . . . 25
- 6.2 Conclusion . . . 26
- 6.3 Future studies . . . 27

1 Introduction

With social media networks playing a part in the daily life of millions of people around the world, fake news stories spreading on these sites are quickly becoming a major concern for the corporate world as well as for the national security of nations. The practice of fabricating and spreading fake news and false content on the web has been commercialized by organizations and political blocks. Producers of fake content has enjoyed financial gain and the ability to influence the public on large scales. As the interconnection of people and the flow of information in society increases, so does the reach that information has. Individuals can connect to large audiences without the need of much resources and this creates an environment where the effect misinformation has can have serious complications. Many actors are affected and involved in the flow of information and as such it has become of interest for them to differentiate between what is true and what is false. What exactly the best way of achieving this is remains unclear. Many different solutions are being discussed and tested by the public, the research community, companies and defence organizations. One possible route is with the use of machine learning.

Machine learning is, in general, about learning to do better in the future based on what was experienced in the past. A practical implementation of a machine learning system or algorithm based software to detect fake news and stopping disinformation from spreading could mitigate the effect misinformation has on public opinion. Such a solution would in fact be interesting to companies like Facebook and Google who are both becoming platforms where falsified stories tends to spread. The demand and pressure for a solution has never been higher. This study will attempt and address the problem by using classification methods that have so famously been used in, among other things, spam filters. It will do this using an algorithm based on Bayes theorem, a method that will naively attempt to classify a text based on how a pre-prepared set of texts have been classified before it. An attempt will also be made to give some broader background on the subject and inquire how this rapidly developing field has gone about dealing with the issue of fake news.

1.1 Problem Definition

The research in this paper contains an analysis of how classification algorithms compare against each other in a test containing false and authentic news stories. An attempt will be made to isolate and locate comparable advantages and disadvantages of the algorithms. The paper also introduces and discusses different linguistic models and methods for detection. The study will also try to present how actors affected have gone about dealing with the problem of fake news and misinformation. The problem statement the paper attempts to answer are:

How would one accurately and to what extent is it possible to differentiate between fake news articles using natural language processing and classification

algorithms?

How are different actors and parties in businesses and governments affected by deception caused by fake news articles and what collective steps could be made towards introducing a solution?

1.2 Scope and Constraints

The practical part of this study, as mentioned, will focus on a linguistic text analysis and machine learning approach to automatically detect deception in news articles. There are several digital methods to detect deception and many different classification algorithms and text analysis approaches. In order to evaluate and motivate what method will best suit the problem and data concerned by this study, a presentation of a few selected techniques will have to be made. Thus the linguistic models and machine learning algorithms that are involved in this project are described in detail in section 3. The parameter that will be used to construct the linguistic models and train the classification algorithms is the raw text content of a training set of news articles. This implies an assumption that the syntactic patterns of the fake news can be distinguished from other news. The idea is to train the models on a dataset and from that analyze the results and effectiveness they have on discovering deception. In reality, a wide range of other parameters can be used that are not taken into consideration in tests this work performs.

The field of news and media is huge and because of that it crosses over to many different fields and research opportunities. Because of the complexity and given the many recent events on the particular topic of fake news, it lies within the scope of this study to present a walk-through of the market of fake news and the mediums on which they propagate. The market is however huge and this work will mainly consider a few players affected by the misinformation problem, namely Google and Facebook as well as the governments role. The work will also go through other deception detection methods used by these companies in the field today. The work limits itself by only presenting these methods and chooses not to go into technical details and present evaluations of how they perform. That is only the case of the machine learning algorithms relevant to the actual tests performed.

2 Background

This section is meant to explain the term "Fake News" and introduce and paint an overview of the history, recent events and the impact large scale deception in news has had on society.The section also covers some solutions implemented to handle fake content. With many of these solutions being fairly recently implemented on the platforms discussed, actual impact and performance however is not yet always clear.

2.1 Definition: Fake News

The term fake news is easily thrown around. A recent example is how the tweets and statements from just elected President Donald Trump, calling out and accusing the American news channels CNN, ABC and NBC for being spreaders of fake news which caught the eyes of the public[Alderman, 2017]. Another interesting example of how the term is being used is the way the Russian Foreign Ministry has begun labeling news stories as fake on their web page claiming that they contain false information about Russia [The Ministry of Foreign Affairs of the Russian Federation, 2017]. In order to carry out analysis and research on the topic, the term needs to be properly defined. This is not entirely obvious since fake news is a term that can mean different things depending on the context. Three closely related terms that ties into each other are explained in this section.

This study will utilize the following definition of the term: *Fake news* , or *Hoax news*, refers to intentionally false information or propaganda published under the guise of being authentic [Stroud, 2017]. These news are then pushed by fake news channels and websites on social networks in an attempt to mislead the consumers and challenge public opinion. The goal is to spread the fake news via social media shares or word-of-mouth. Closely related is another definition of deception in media: *Journalistic deception*. Journalistic deception is an act of communicating messages verbally (a lie) or non-verbally through the withholding or twisting of information with the intention to initiate or sustain a false belief [Lee, 2004]. *Yellow journalism* or *Yellow press* should also be brought up in the context. This is the type of reporting that through exaggerations, exploitation's and unethical and unprofessional manners, often publishes stories without evidence or with biased opinion, sometimes also bending truths. With the use of colorful and inciting headlines and pictures yellow press tries to either in traditional media; sell more newspapers, or on social media get more clicks for increased ad revenue[Klockars and Williams, 2012]. Depending on who is asked the definition differ and it is difficult drawing a line between these three definitions. Observers agree that fake news cannot be straightforwardly defined. "Fake News" comes in many different shades. This need not be taken as proof of the futility of investigating this phenomenon. On the contrary: accepting that there is no easy way to demarcate between "fake" from "non-fake" across all cases opens up interesting research opportunities[The Public Data Lab, 2017].

2.2 More than one type of fake

Clearly not all types of news are the same, and because they are not the same they can be expected to have different properties and distinguishing attributes. An article or news story can be fabricated in different ways and therefore there exists different kinds of fakes with different degrees and gravity. The world is

not black-and-white and it becomes a problem if the issue is tackled with a black-and-white thinking mindset. For identification to be made by a machine learning algorithm a separation can be made into three subgroups A) Serious fabrications B) Hoaxes C) Humourous fakes [L.Rubin et al., 2015]

The definition of fake news in the previous section defined fake news as being intentional. This leads to a number of related cousins of fake news being ruled out in our study. Thus, the machine learning algorithms tested described in chapter 4.3 will be trained with deceptive news in the subgroup A and B and will ignore the following related cousins.

- unintentional reporting mistakes
- rumors that knows it is a rumour
- humorous satire or humorous fakes

Figure 1 shows examples of falsified stories that the study perceives as typical serious fabrications and hoaxes. Such news are example targets to identify as fake by the trained models. Both stories emerged in the recent American election. To the left is the "Pope Francis endorses Trump" story that gained widespread attention claiming that Pope Francis supported Trump in his campaign for presidency. To the right is the fake news story about the alleged paedophilia ring involving people at the highest levels of the Democratic Party operating a paedophilia ring out of a Washington pizza restaurant. On the picture is democrat John Podesta and the involved pizzeria, Comet Pizza. The story grew its initial following after the leak of Clinton's emails that was later picked up on the internet forum 4chan [Wendling, 2016].

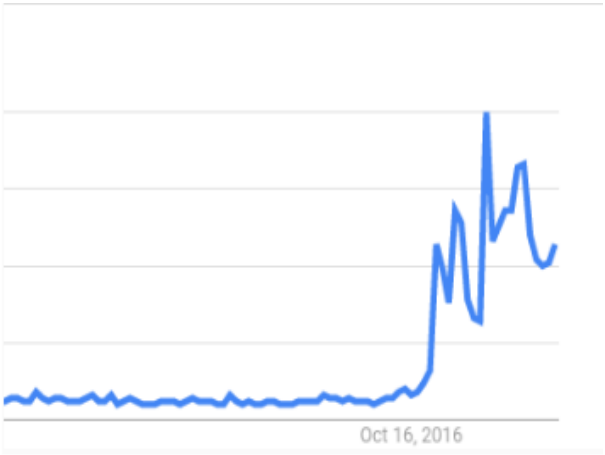
Figure 1: Serious Fabrications/Hoaxes:

Pope endorses Trump and Pizzagate paedophile ring

2.3 History

It is important to realize that hoaxes and deceptive news have existed for a long time [Soll, 2016]. It is not a new phenomenon that appeared out of nowhere. Fabricated news have existed ever since humans felt the need to influence the thinking of others, to perhaps overthrow a rule or incite revolts that would cause problems in order to gain their own cause. Bending truths for ideological or political reasons is hardly a new phenomenon, so why the recent rise of interest and urgency? It is a fact that a clear lift-off has seen a in public interest on the subject demonstrated by this chart collected from Google Trends.

Figure 2: Chart from Google Trends showing the IoT or Interest over time for the search term "Fake News". A peak can be seen around the time of the recent US election on November 8, 2016.



A paradigm shift has been taking place in the mass media industry during the recent decade. The popularity of and advancements in technology has resulted in new media significantly increasing its market share over traditional media [James and Shearer, 2016]. Traditional or conventional media can be defined as the medium, which attains one-way communication with the audience, for instance, newspapers, magazines, broadband, and motion pictures; implying that conventional media is made up of four media sectors including publishing, radio, film, and television [Zhang and Authors, 2012]. New media is defined as the communication that through the medium of the Internet promotes the interaction between media and audience. New media is based on computing technologies, use digital information, and usually involve connection to an open digital communication network [Kung, 2008]. New media is by that definition very tied to new technology, making the term very vague. A characteristic of new media is that it is extremely decentralized, and requires very low investment, provides greater interactivity and public participation and is much more

difficult to control [Dizard, 1999]. The result is a double-edged sword, less control means increased freedom of press, but also a lot of noise and malicious fake content. Today's society and use of the Internet has created a platform for fake news to spread like virus. This was not always the case and journalism has in the past been more controlled and small scaled in comparison to today's media, leaving little room for malicious parties for reasons demonstrated by the following three observations.

Spreading disinformation used to require resources.

It would cost money and time to build up and reach a large number of consumers. Conventional media used to reach audience, both locally and nationally, but with a higher distribution cost than today. Even though the globalization strategy from big media conglomerates made media content available internationally, the high distribution costs was still a limitation to that. There is less limitation, however, for distribution through the Internet.

Conventional media is regulated.

There are laws that media outlets and publishers need to follow lest they end up being sued. Most people got their daily update of the world from network news departments and newspapers that normally followed broadly accepted editorial rules of fairness and objectivity.

Connected World

The social media boom changed the game for malicious players. Resources are barely no longer needed. Huge proportions of consumers can be reached instantly. The Internet broadens the way media content is reaching the audience, with the characteristics of multiple, anytime, and anywhere. Since media content have been digitized, it easily reaches the audience through multi-platforms - computer, laptop, smartphone, and game console - besides the traditional ways.

[Zhang and Authors, 2012]

2.4 Impact

This section will discuss the impact fake news has on some affected parties. These will for this study be grouped into *Businesses* and *Government and National security*.

2.4.1 Businesses

In order to discuss the impact fake news has had on businesses and consumers of their services, we must first define a set of relevant actors. The business areas and organizations selected and involved here are Social Media Networks: *Facebook, Twitter*, Search engines: *Google* as well as *Advertising networks* and *Fake news producers and consumers*. Considering the nature of this fast moving area and given the recent turn of events and lack of new scientific research on the involved companies newly released counteractions and solutions. The study contains a large amount of unscientific sources to explain and illustrate, such as articles and blog posts. This study deems it necessary to use up-to-date and new posts about the companies discussed.

The Social Networks

Social media platforms has faced pressure from consumers and civil society to reduce the appearance of fake news on their systems. An important and crucial step to achieving that is to understand the motivations of fake news producers and to stop them at the source. One motive for fake content producers that has been identified is the financial motive. Articles that go viral on social media can draw significant revenue from advertising when users click and get redirected to the page. A way to measure the importance of social media networks for fake news suppliers is to measure the source of their web traffic. Each time a user visits a web-page, that user has either navigated directly to the server or has been referred from some other site. These referral sources are largely consisting of and including social media and search engines [Allcott and Gentzkow, 2017]. Because of the sheer size of Facebook and Google they are identified as the major referral sources and have been receiving criticism for not doing enough in terms of limiting fake content. There are examples where fake news producers have made thousands of dollars from ad-revenue [Smith and Banic, 2016]. Following the criticism, Facebook has been focusing on three main areas in its attempt to combat fake news from spreading.

One of the areas focused on is the *disruption of the economic incentives* of producing and spreading fake stories. According to Facebook the steps taken to accomplish that include making it as difficult as possible to buy ads on the platform for people posting fake content by reinforcing their policies. As well as better identifying false news with the help of the community and third-party fact-checking organizations so they can limit the spread of identified articles, making it uneconomical [Mosseri, 2017].

The second area focused on is *building new products*. One specific function Facebook has introduced to receive help from the community and users of the platform is the flagging of fake content [Facebook Help Center, 2017]. To flag

a fake news article, users click on the upper right hand corner of a post. The more times a post is flagged as false by users, the less often it will show up in the News Feeds. Facebook won't delete heavily flagged posts, but they end up with a disclaimer: "Many people on Facebook have reported that this story contains false information [Wohlsen, 2015]. However, there is a flipside to this solution. It is possible to launch a "mass flagging" - coup to flag true stories as false with malicious intent. In fact, it is possible to take advantage of structural, temporal, content, and user features to create social bots [Ferrara et al., 2016]. These social bots has proven to be influential in the spread of fake news due to the fact that they are designed to increase the reach of fake news and exploit the vulnerabilities that stem from our cognitive and social biases [Shaoi et al., 2016]. For example, they create the appearance of popular topics and campaigns to manipulate attention, and target influential users to induce them to reshare misinformation [D.Conover et al., 2011]. Single users may also flag the story simply because one does not agree with the article content. Critics have argued that regular users often are not in a position to assess the validity of the links they see, making the new flagging function unreliable [Stanford History Education Group, 2016]. This could be the reason as to why Facebook sends heavily flagged posts to third-party fact-checking organizations. If at least two fact-checking organizations marks a story as disputed, users will begin seeing a banner under the article if it appears in their News Feed. The banner reads: "Disputed by 3rd Party Fact Checkers." Facebook will provide links to articles debunking the posted item that appear below it, explaining why. Stories that have been disputed will also get pushed down in News Feed. Users who try to share a disputed article are asked if they are sure they want to share it [Guynn, 2017]. But relying solely on the users of social media is not a foolproof solution. The whole idea builds upon enough people realizing a story is fake and flagging it in order for a red alarm to turn on on the side of Facebook. However if this was always the case then fake news would not be as big of a problem in the first place.

It is actually very difficult for many people assessing the level of truth in articles, as pointed out by the critics of the flagging function. Therefore, the third and last key area of Facebook is to *educate the public* and helping people make more informed decisions when they encounter false news to achieve an informed community [Dillet, 2017]. Facebook says it's algorithms are also helping by rooting out fake articles. The algorithms decide what news or information is included in an individual's Facebook feed considering, among other factors, its source, when it originated, the type of content, and the number of interactions with the content.

On Twitter , fake news shared by real people is concentrated to a small set of websites and highly active "cyborg" users [Lazer et al., 2017]. These cyborg users have evolved to be industrialized by states and terrorist organizations and these professional groups are sometimes called Troll farms. According to a study conducted by students in computer science at KTH, a potential algorithm to

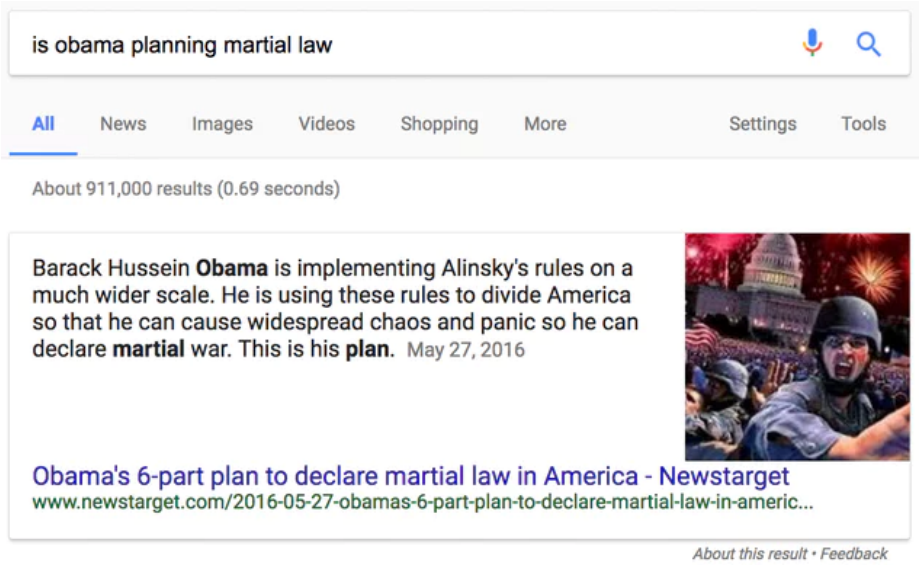
track down Troll farms in Twitter are cluster analysis algorithms [Da Silva and Englind, 2016]. Detection is key in order to shut down these accounts to stop the spread.

Search Engines

In a similar attempt to Facebook, Google has also rolled out a fact checking tool as a means towards decreasing the impact of fake news. The search giant's new tool also relies on users and contains similar functionalities to the one introduced by Facebook. This tool mainly contains functions that are aimed at *Auto-complete suggestions* and *Featured Snippets* [Kosslyn and Yu, 2017].

A Featured Snippet is the box that appears at the top of searches in an attempt to answer user questions without them having to be redirected by clicking links. These answers are visually set apart, encased in a virtual box with a slight drop shadow. The Featured Snippets are provided by an automatic and algorithmic match to a given search query. When the snippets work, the user is directly given relevant answers to the questions asked. But when they do not work, they can be pulling information from fake and propaganda sources. Google has been struggling with the results returned by these Featured Snippets and there are numerous examples of questionable results(see Figure 3) for an example. The new tool mitigate harmful effects caused by results like this appearing on top by accompanying the Featured Snippets with a credibility score given by external fact checking organizations. Users have also been given the option to report offensive and/or misleading Featured Snippets [Williams, 2017].

Figure 3: An example of fake news appearing on Featured Snippets



[Jeffries, 2017]

Auto-Complete suggestions are the options Google presents to the user at the moment of typing in the search query. The autocomplete suggestions are based on the user's previous searches, the actual terms typed and other user searches and trends [Google Search Help, 2017]. Predictions have been known to return inappropriate queries, for example, the suggestions have been recorded to promote violence and sexism. Incidents involving both the Featured Snippets and Autocomplete are an embarrassment to the company and hurts the image they are trying to build as a reliable service [Liedtke, 2017].

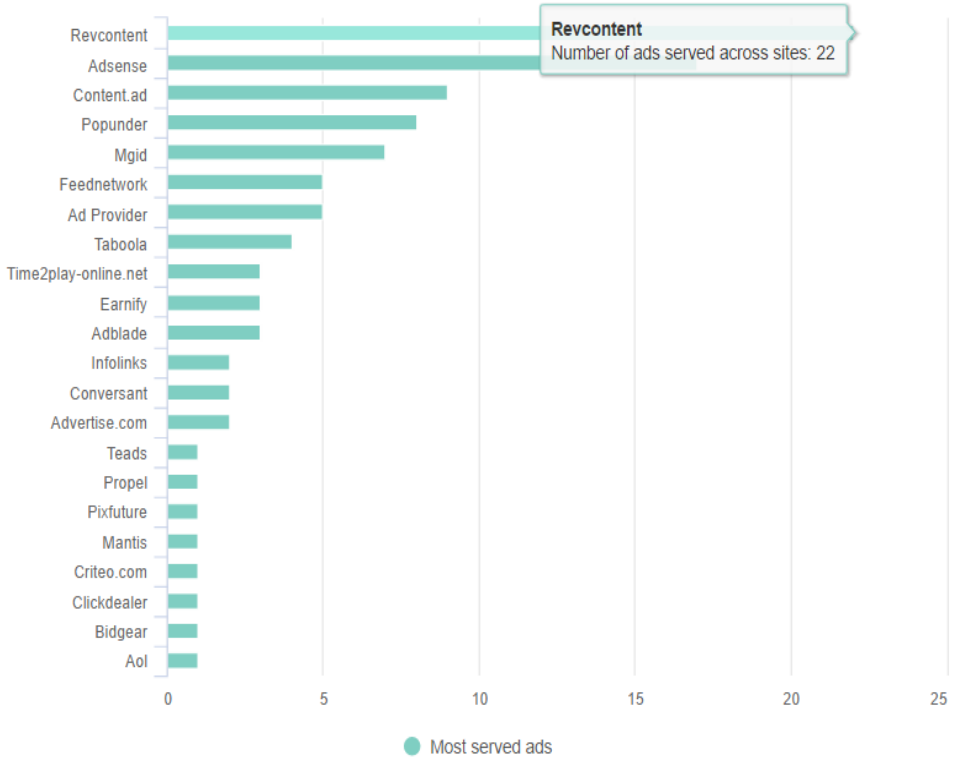
Digital Advertising Networks

Both previously covered giants Facebook and Google fall under this section since advertisement solutions are a part of their huge product portfolio. Given the explosion of bogus sites seen, top-executives and management at brands, agencies and companies involved in advertising technology/digital advertising have become concerned about the safety of digital advertising and being associated with those websites. Customers of the ad tech companies could have their banners and commercials placed on these sites without knowing it until after the fact. This is due to the fact that digital advertising systems goes through many intermediates between the publisher of the ad and the advertiser. An advertiser of a campaign can partner up with an Ad Network that in turn buys a certain volume of, what in the sense of online marketing, is called "impressions" from different publishers. An impression is counted as whenever an ad is fetched from its source and loaded onto the page of the publisher. After the purchase the ad publishers guarantees them the agreed upon amount of impressions. But the outsourcing may cause communication issues and an advertiser may not always know which of the publishers the ad networks purchases its impressions from. Publishers can even outsource further by acting as an ad network if they do not have capacity to fill out the agreed amount on their own. Leading to even more uninformed advertisers. There's also a risk of publishers URLs being withheld or spoofed [Timmers, 2016].

The signum has been fake news and real ads, both working together to create an environment where fake news articles are thriving and being a big reason as to why it has been so profitable for fake news producers to push their content.

Figure 4: In an investigating study conducted by Buzzfeed news they mapped which digital advertising networks had their ads showing on fake news stories. Interesting is that Google adsense almost tops the list.

This graphic shows the most common ad networks found on fake news sites reviewed by BuzzFeed News:



[Silverman et al., 2017]

Effect on business reputation and stock prices

One aspect that has not yet been accounted for and touched upon is the actual content of the fake news articles regarding businesses. What impact can the text content have on corporate reputation and how much does made-up stories actually affect the reputation of a company and could it in fact go so far as affecting stock prices? An example can be found in Sweden's first conviction of its kind where two medical students was convicted for market manipulation. It was concluded that they illegally affected the stock market by systematically producing fake recommendations of small stocks under false aliases and blogs. The medical students then amplified the spread of these falsified recommendations by posting them on forums and social media. After the publishing of the blogs-entries the price of the involved stocks saw an increase between 4,89-53,76 percent. Earning them millions in revenue from their long positions [Stockholms Tingsrätt, 2016]. An indication that what is read affects opinions and perceptions of companies, and even causes savers to act on these stories by purchasing stock.

This is related to fake news because one factor driving the value of a company's stock price is market sentiment, and fake news affect sentiments [Harper, 2017]. Hedge funds and large investors are getting ready for the next spike in sentiment changes on companies and making trades based on sentiment analysis is now noticeable. Hedge funds sit patiently on the sidelines waiting for something causing a big swing in sentiment. One thing that could cause such a swing is for example a new tweet from President Trump. A previous one aimed at the company Lockheed lowered the company's valuation by roughly 4 billion dollars and Hedge funds were quick to capitalize [Reilly, 2017]. This is only one example of many. By searching the web it is possible to easily find many more. The consensus is that companies can experience significant damage to their reputation and experience serious financial damage from fake news. An example of a company that uses a sentiment tracking algorithm in its product portfolio is Alva. The product can, among other things, notify costumers when a change in sentiment occurs, making it possible for the costumer to react quickly in such an event [Alva, 2017].

2.4.2 Government and National security

As touched upon earlier in the study, the idea of deception, disinformation and fake news is nothing new. Propaganda has long been used to affect people, create support for a cause and strike fear in the enemy. Propaganda is often separated into three groups.

- White propaganda where the sender is visible and clearly defined and the content is objectively true.
- Black propaganda, where the source is hidden and the content often is false or slanted.

- Grey propaganda which is somewhere on the borderline between the black and white. An example would be the US radio broadcasts to the Eastern Bloc during the cold war. It attempts to sway opinions with half hidden and distorted facts and unclear senders.

Another example of grey propaganda happend 2002-2008 when the US military headquarters recruited and trained 75 pensioned officers to comment and feature in tv-news on for example Iraq's possible ownership of weapons of mass destruction. NATO has special PSYOPs, psychological operations as a part of their strategies. These are meant to weaken the public of the opponent and strengthen the own support as well as gain support from the undecided factions and people. This is done by, for example, starting newspapers, radio and tv-channels under names that hide any connections [Nygren and Hök, 2016].

What has changed from earlier variants of propaganda is the way today's information society works. It is possible to, as an individual, reach large audiences. Not only that however but in this information society people make their own decisions, and fast. We listen to what is currently going on and act on that very quickly. This means that we are more reliant on information and because of this you can reach greater effects. A state, unlike perhaps some other actors mentioned, naturally has many tools to achieve it's goal such as diplomatic, military economic ones and information is one of these. A totalitarian state has an easier time organizing this since the state controls everything. An example of disinformation is the Ukraine crisis 2014, where a state invaded and annexed another state's territory and lied about it. This delayed the response from the rest of the world. This was not only done by false information but also by doing and going through activities that eased the spreading of lies [Folk och Försvar, 2017]

3 Theoretical Framework

3.1 Machine Learning

It was mentioned that Facebooks and Googles algorithms was clearing inappropriate content on their services. Not only concerning the area of fake news and disinformation but also other content, like pornoghraphy, violence, spam and more. But what does that mean practically and how are they able to achieve this. In order to understand that an understanding of machine learning is needed.

3.1.1 Broad Categories

Supervised learning

Supervised learning is a machine learning technique used for building predictive models from a known dataset called the training and testing set. The concepts

and differences of the training and testing are explained in section 3.1.2. In supervised learning, the training set contains input data and appropriate response values. In other words the data is organized as N input-output pairs. Given the training set, the aim of a supervised learning algorithm is to create a model that makes predictions on the response values (Output) of new unseen data when introduced to a presence of uncertainty. The presence of uncertainty is the untested input the algorithm has never seen before. The adaptive algorithm searches for patterns in the data when given observations determining the relationship between input and output. More observations will therefore generally improve the algorithm's predictive performance.

Unsupervised learning

Unlike supervised learning, unsupervised learning algorithms does not use a known training set consisting of input and response values. Instead the dataset contains data without labeled responses. The aim of an unsupervised learning algorithm is to understand the input data and draw conclusion from it by searching for a deeper understanding and hidden patterns in the attributes and features of the data. This characteristic makes the unsupervised algorithms useful in cases where little or nothing at all is known about the outputs of the data.

3.1.2 Training and test partitions

An important context in machine learning is the separation of data into two divisions or sets, the training set and the testing set. Each of these sets of data has a distinct role, the roles of fitting and evaluating an algorithm. Although it could appear intuitive, the training set or learning division has one single and crucial role, to provide the raw material needed to train the model. A novice in the field might have the notion that more training of the model using more data would give a more fitted model, why even bother splitting the data into two partitions and instead use all data as raw material in the training?

The objective of many machine learning implementations is to develop predictive models that guides decisions on real world problems, and to assess the quality of the model one must simulate real world encounters. To simulate encounters with the future, a part of the accumulated data is reserved for testing, this is the role of the very important test partition. In the most simple partitioning the test partition is not used and is not needed for building the model. In the

simplest method of partitioning the data nothing is to be learned from the test set. It is used uniquely for evaluation of the model. This is the partition often made in cases when a large dataset is available. When not sufficient data has been gathered one can use methods of rotation on the data, see Cross validation below. In the event that a model becomes overly-complicated and grows in complexity it will start to show decreasing accuracy in its predictions on unseen data, this is known as overfitting [Steinberg, 2014]. Without a test partition it is harder to find a sweet spot of ideal model size and complexity.

Cross validation

In the simplest cases, testing sets are constructed just by splitting some original dataset into more than one part. But when doing this the evaluations received tend to reflect the particular way the data was divided up. The solution is to use cross-validation strategies to get more accurate predictions and measurements when the data set is too small for simple partitioning. This alternative way of training and testing has the goal to make sure that every object from the dataset has the same chance of appearing in the training and testing set. One basic protocol of cross-validation is k-fold cross validation.

In k-fold cross-validation, the original dataset sample is partitioned into k equal size data subsets. These k partitions are partitioned randomly. One of the subsets are selected as the testing set. The remaining k-1 as training sets. This process is repeated k times and a different subset is used as the testing set each time. The results are then averaged or combined to produce a single error estimation. This method has the advantage of ensuring the observations are used for both training and testing, and each observation is used for testing exactly once.

3.1.3 Classification and Regression

In the subgroup of classification algorithms, the goal is to assign a categorical response value or label value to the output. These categories can either be *Binary Splits* ('Yes' or 'No') translated by the computer as 1 or 0. Or more than two categories, called *multi-label problems*. The classification algorithms are used where the predicted output of the data can be separated into specific "classes". In mathematical terms a classifier is a function f that maps input feature vectors $x \in X$ to output class labels $Y \in 1, ..., C$, where X is the feature space. Feature space refers to the N-dimensions where your variables live. In the category of regression algorithms the goal is to predict continuous-response values and measurement for a given observation. Here the output are therefore real numbers and common uses can be seen in stock price prediction, consumer goods price prediction etc.

3.1.4 Support vector machines

A Support Vector Machine (SVM) is a supervised machine learning algorithm that can be employed for both classification and regression purposes. There are several Support Vector Machines used for different purposes. Although the goal of all of them is to find the optimal separating hyperplane which maximizes the margin of the training data. Meaning an algorithm in the SVM family takes an input vector X_1 of the dataset, we can say that X_1 is a *P-dimensional vector* if it has P dimensions. It was previously stated is section 3.1.1 that a dataset contained input and output class labels in Supervised learning implementations. A more mathematical and formal defintion of a dataset D is the set of N couples of elements (X_i, Y_i).

The SVM takes input vector X_i, transforms it to a higher dimension space and then goes on finding the linear hyperplane which separates the data into two classes with least error and largest distance margin between the respective classified points. In other words data is represented as points in space divided by a gap that is drawn as wide as possible. Finding the optimal hyperplane requires the algorithm to solve an optimization problem. Optimization problems are themselves somewhat tricky to solve.

3.1.5 Naive Bayes

Naive Bayes methods are a set of algorithms that are based on Bayes' theorem with a naive assumption of independence between every pair of features. Given a variable y and a dependent vector x_1 through x_n this theorem states:

$$P(\, y \mid x_1, ..., x_n) \, = \frac{P(\, y) \, P(\, x_1, ..., x_n \mid y)}{P(\, x_1, ..., x_n)}$$

with a naive assumption that:

$$P(\, x_i \mid x_1, ..., x_{i-1}, x_{i+1}, ..., x_n) \, = P(\, x_i \mid y)$$

an estimation can then be used to get $P(\, y)$ and $P(\, x_i \mid y)$ where the former is the relative frequency of y in the training set. The different classifiers differ mainly on the assumptions made regarding $P(\, x_i \mid y)$ Even with the naive simplification of feature independence Naive Bayes work in many real-world situations like document classification and spam filtering. They all require a small amount of training data to estimate necessary parameters. These classifiers can be very fast compared to more advanced methods.

Multinomial Naive Bayes is one of the classic variants of Naive Bayes used in text classification. Data is often represented as word vector counts with distributions parameterized as vectors

$$\theta_y = (\, \theta_{i1}, ..., \theta_{yn})$$

for each y where n is the number of features, which in text classifications means the size of the vocabulary. θ_{yi} is the probability

$$P(\, x_i \mid y)$$

of feature i appearing in a sample of y θ_y is estimated by smoothed maximum likelihood.

Bernoulli Naive Bayes is another variant and has a decision rule based on

$$P(\, x_i \mid y) \; = P(\, i \mid y)\, x_i + (\, 1 - P(\, i \mid y)\,)\,(\, 1 - x_i\,)$$

which is different from multinomial Naive Bayes rule by explicitly penalizing non-occurrence of a feature i that is an indicator for y where the multinomial one just ignores non-occurring features.

3.1.6 Evaluation Methods

Accuracy/Error

Evaluating an algorithm can be done with the use of the test data, where the predictions from testing are divided into four sets. In classification purposes those sets are defined as *True Positives*(T_P) observation is positive, and is predicted to be positive. *True Negatives*(T_N) observation is negative, and is predicted to be negative. *False Positives*(F_P) observation is negative, but is predicted positive. *False Negatives*(F_N) observation is positive, but is predicted negative. A *classification accuracy rate* is obtained when correct predictions are divided by the total N number of predictions made.

$$Accuracy = \frac{True\ Positives + True\ Negatives}{N}$$

classification error rate can then be defined as

$$Error = 1 - Accuracy$$

Precision/Recall

Two measures *Recall* and *Precision* are often used for capturing the effectiveness of an algorithm. Recall(R) is a measure of the rate of documents successfully classified by the algorithm (i.e. its effectiveness), while Precision(P) measures the percentage of how many of the returned documents are correct.

$$Precision = \frac{True\ Positives}{True\ Positives + False\ Positives}$$

$$Recall = \frac{True\ Positives}{True\ Positives + False\ Negatives}$$

F1-score

A system with high recall but low precision returns many results. However most of its predicted labels will be incorrect when compared to the training labels. A system with high precision and low recall is just the opposite. It will return very few results, but most of its predicted labels will be correct when compared to the training labels. When measuring how well a system performs it can be useful to have a single number that describes performance. One can achieve this by calculating a combined metric *F1-score* of the algorithm defined as the harmonic mean of the ratios Precision and Recall. The F1-score can be seen like an 'average' between the two that also takes into account how similar the two values are.

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

3.2 Text Analytics

3.2.1 Text analytics and machine learning

Machine Learning (in the context of text analytics) is a set of statistical techniques for identifying some aspect of text (parts of speech, entities, sentiment, etc). The techniques can be expressed as a model that is then applied to other text (supervised), or could be a set of algorithms that work across large sets of data to extract meaning (unsupervised).

3.2.2 Requirements for a fake news detection corpus

The data used in a fake news corpus should preferably meet the following conditions:

1. Availability of both truthful and deceptive instances. To predict based on patterns we need both positive and negative data points.

2. Digital textual format accessibility. The prefered medium is text, videos audio and images will have to be transcribed or converted.

3. Verifiability of ground truth. The question of how one knows what is genuine or fabricated. Credible news sources with reputations based on a system of "checks and balances" are preferred.

4. Homogeneity in lengths. Articles of comparable length should be used. That is, a short tweet, a Facebook one-paragraph summary and a longform co-ed does not constitute a fitting homogeneous dataset.

5. Homogeneity in writing matter. The news should be aligned along genres and topics and use similar types of authors (e.g. serious or humorous) and be compared across news outlets.

6. Predefined timeframe. It should have a thought-out timeframe. A snapshot of breaking news might have greater variations than all news on a topic over a 2-3 year interval.

7. The manner of news delivery (humor, newsworthiness, believability absurdity, sensationalism). How the news are provided creates context for interpretations. Truth-biased readers might shift to a lie-biased perspective when reading satire.

8. Pragmatic concerns include copy-right costs, public availability etc.

9. Language and culture. English is the predominant language in deception detection and only a handful of other languages are explored and reported in research with little consideration given to language and cultural differences.

[L.Rubin et al., 2015]

3.2.3 N-gram

An N-gram is a sequence of N words. For example a 2-gram (more commonly called bigram) is a sequence of two words like "welcome back". A 3-gram would be a three word sequence such as "please come again". Using N-grams you can estimate probabilities of words given previous words as well as assigning probabilities to entire sequences.

3.2.4 Bag-of-words

This is one of the simpler ways of representing text. It regards each word as a single, equally significant unit. In a bag-of-words individual words or n-gram frequencies are aggregated and analyzed to, in this case, allow for detection of cues of deception. The simplicity does lead to some problems though. Besides relying only on language it often relies on isolated n-grams separated from useful context information. It has still found its use by researchers as a complementary tool of analysis [L.Rubin et al., 2015].

3.2.5 Deep Syntax

Analysis of words is not always enough for accurate predictions however. One example of deeper analysis is by using the syntax. Deep syntax analysis can be implemented with Probability Context Free Grammars (PCFG). Here sentences are transformed after a set of rewrite rules (parse tree) to describe the syntax structure which in turn is rewritten by their syntactic constituent parts. The final set of rewrites produces a parse tree with probabilities assigned. This method can be used to distinguish rule categories (lexicalized, unlexicalized, parent nodes etc.) for deception detection with 85-91% accuracy depending on which rule is used. This kind of syntax analysis might not be sufficient alone but can again be combined with other methods.

3.3 Previous studies

The general concept of deception detection is wide and there are many studies that stretches over many areas. In a paper (Machine Learning for E-mail Spam Filtering: Review, Techniques and Trends , A. Bhomik, S. M. Hazarika, 2016) machine learning in spam filtering, among other things, is discussed. More specifically it goes into detail about Naive Bayes and touches on topics like textual fingerprints and authorship identification techniques. While in the context of spam filtering it is still related in the sense of attempting to classify texts.

In an article from Language and Information Technology Research Lab (LIT.RL) Faculty of Information and Media Studies from the University of Ontario named Deception Detection for News: Three Types of Fakes by Victoria L. Rubin, Yimin Chen and Niall J. Conroy the researchers discusses three types of fake news and weighs their pros and cons as a corpus for text analytics and predictive modeling. This with the purpose of reviewing what a detection system would need for properly filtering, vetting, and verifying online information. This particular study is relevant for the problem this paper tries to answer because it highlights the attributes and demands a reliable corpus for deception detection would need to abide. The study also divides the group of fake news into different sub-sets that all have different properties.

A Field Guide To Fake News compiled by Liliana Bounegru, Jonathan Gray, Tommaso Venturini and Michele Mauri is an unfinished huge project with many contributors that is only available as a sample on the website (fakenews.publicdatalab.org). It contains a detailed walkthrough of how to map and investigate fake news, online misinformation, how to track the spread and lots of other information on fake associated material on social media and the web in general.

4 Methodology

4.1 Overview

Detecting fake news is a difficult task for many humans, let alone machines. This is because many factors come into play when trying to determine if a story is fabricated or not. To make this task more manageable this study has restricted and limited the scope to only take the raw text of the news stories into consideration. In order to implement and analyze an algorithm a number of things have to be done. The first step is to build a corpus by scraping the web for data or find an existing one that fits the task. The corpus should contain news that are predefined and classified as true or false. The next step is to implement and fit the algorithms which will handle the classifying itself which is done using the majority of the corpus. Following this a known set of data using a small part of the corpus will be used for testing, analyzing and evaluating.

4.2 Data Gathering

For this test a total of 201 different mostly American news articles were gathered. Of those 201 articles 120 were fake articles and 81 were real articles. All of them was tagged appropriately. After tagging 10 fakes and 10 real articles were selected and used for testing while the remaining were used for fitting. The 201 fake news articles were gathered from an online corpus. This corpus was in turn based on data gathered from 244 websites tagged by the Google Chrome-extension "BS detector" [Risdal, 2016]. The real news articles were all gathered by hand. An article was deemed to be true if a trustworthy source had reported on the same news-story and that article did not contradict facts stated in the original article. Trustworthy sources in this study were BBC, Reuters, Washington Post, New York Times and CNN. This does not however necessarily mean that all chosen articles are from these sources, but all real news articles collected have been "backed-up" by one or more of these sources. The articles collected were mainly on world events and U.S. politics, the same goes for the fake articles inside the corpus. This in an attempt to keep the subjects of the data as close as possible. The size of the data-sample collected by hand was mainly limited by the time and work effort of choosing the real news stories.

4.3 Selection of classification algorithms

In this study two different algorithms were used. Bernoulli Naive Bayes and Multinomial Naive Bayes. These are both examples of Naive Bayes methods and were chosen for being both simple and easy to use and able to perform well with smaller sample sizes while still hopefully being different enough from each other to produce contrasting results.

4.4 Evaluation

After the trained model has been tested the result should be evaluated. A test sample with known labels is compared to the models prediction, while there is no hard rule for how big such a test sample should be, common recommendations often range between 10-30% of the data set. Evaluation methods on the algorithm described in chapter 3.1.6 will be calculated and demonstrated. 10 fake and 10 real articles as well as their classification were separated for the test. They were chosen as the 10 last articles in the file for the fake news and the 10 first for the real news. This was purely for ease of coding. The articles were stored in one array and the classifications in another, order intact. Using the fitted models and the predict method of the Naive Bayes functions an attempt to predict whether the articles of the test-set were fake or not was made. A larger test based on 10 real and 100 fake news was also made with methodology and size of training sample remaining the same. Complementary to this a brief word frequency analysis was made based on the entire training set with stop list words removed.

4.5 Present and compare results

When generating the result from the different algorithm the result needs to be shown in a way so that they can easily be compared to each other. Using the known separated classifications of the test-set a comparison was made with the predicted results. If both were true then is was classified as a True Positive, if both were false then it was a True Negative. If the real value was true but the prediction was false then it was a False Negative. If the real value was false but the prediction was true then it was a False Positive. The word frequencies of the 20 most commonly used words in the real and fake news respectively can also be seen in a diagram form.

5 Data and Results

Table 1: Smaller test of 10 real and 10 fake news articles

	True Positive	True Negative	False Positive	False Negative	Total News (Real/Fake)
Multinomial NB	10	2	0	8	20 (10/10)
Bernoulli NB	4	8	6	2	20 (10/10)

Table 2: Bigger test of 10 real and 100 fake news articles

	True Positive	True Negative	False Positive	False Negative	Total News (Real/Fake)
Multinomial NB	10	39	0	61	110 (10/100)
Bernoulli NB	4	94	6	6	110 (10/100)

Table 3: Evaluation scores for Table 1

	Recall(%)	Precision(%)	F1-Score(%)	Accuracy(%)
Multinomial NB	55.56	100.00	71.47	60.00
Bernoulli NB	66.67	40.00	32.00	60.00

Table 4: Evaluation scores for Table 2

	Recall(%)	Precision(%)	F1-Score(%)	Accuracy(%)
Multinomial NB	14.48	100.00	24.07	44.54
Bernoulli NB	40.00	40.00	40.00	89.09

Figure 5: Word frequencies for the 20 most commonly used words in the fake news articles of our study. Stop words from the classification stop list are excluded.

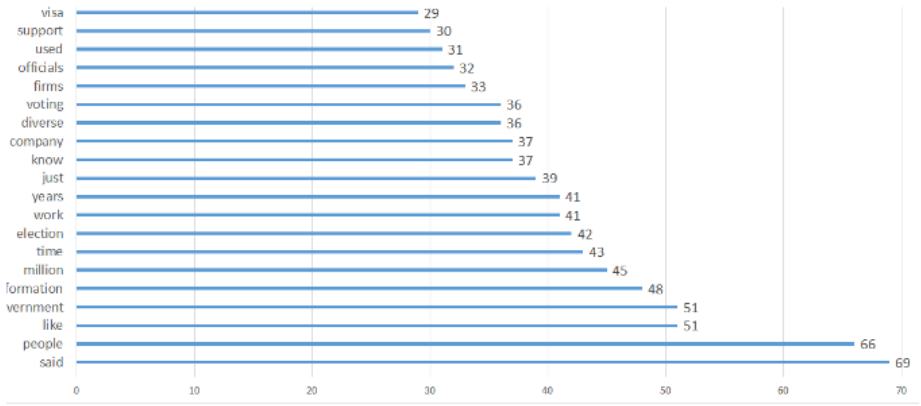
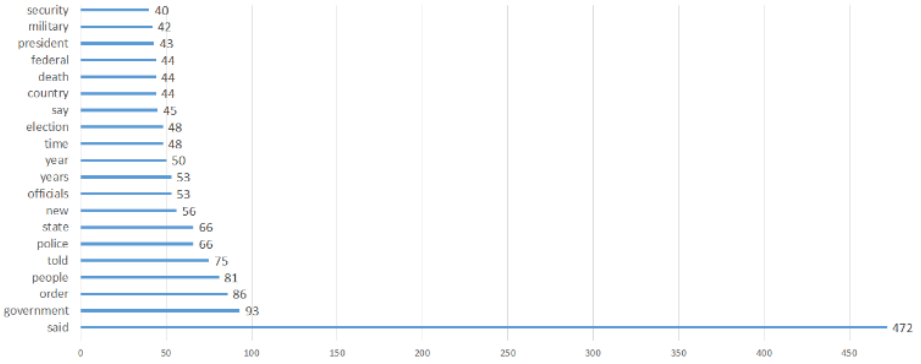


Figure 6: Below are the word frequencies for the 20 most commonly used words in the real news articles of our study. Stop words from the classification stop list are excluded.



6 Discussion and conclusion

6.1 Error sources

The result are mediocre which is hardly surprising considering the simplicity of the method. The number of articles are also on the low side, especially the true news articles. This is because of the time and work selecting the news took. When it comes to the fake news corpus used in the study, the very nature of how the corpus was created using the tagging of a browser extension is unreliable. The extension might present either news that are not fake or create a bias towards certain kinds of news, perhaps poorly written ones, that are not entirely representative. There is also the possibility that what is assumed to be real news might not be real news. This hints at the problem that arises if credible sources pick up news from less credible sources and rewrite them. This would presumably create an article with similar syntax as other real news but containing false information. By looking at text syntax there is a potential difference in subject that occurs naturally considering the nature of fake news. If a real article covered the same news story as a fake one then it would not be false. Another source of error to be accounted for is that a few of the articles are from British sources, like the BBC, which might create a difference in language. The number of these are however fairly low. The stop list used was a default one and might need some more tweaking. This can be noticed in the word frequencies seen. In retrospect, for example the word "said" could have been removed considering the strange over-representation of it among the real articles. On the other hand determining the relevance of words too heavily can be difficult and may run the risk of becoming arbitrary.

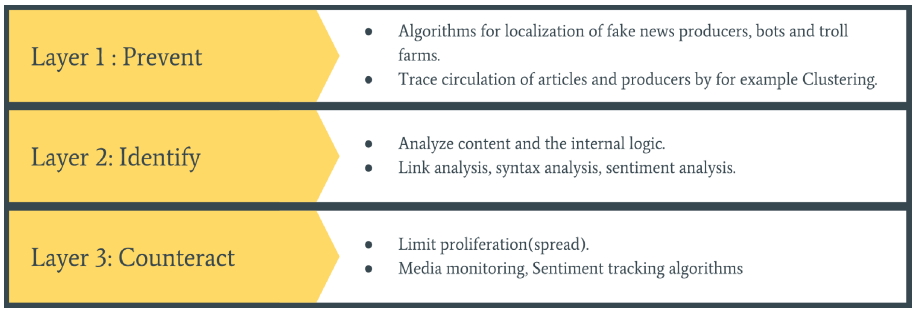
It was previously stated in section 2.2 that unintentional reporting mistakes and self-aware rumors was not included in the study's definition of fake news and therefor should not be included in our data and testing. I reality, there is no easy way for this study to actually correctly separate all unintentional mistakes and rumours from intentional deception and this affects the credibility of the test results. Another cause for concern over the result is that some aspects of the corpus required for a qualified sentiment analysis described in chapter 3.2.2 was not met. This is also a trade-off made because of the time it would take to collect a corpus that abide to those rules. For example the study has taken no significant regards to homogeneity in lengths and predefined timeframes. Leading to the final result being questionable.

6.2 Conclusion

The presented methodology turns out to be no silver bullet for fake news. The two classification methods, on inspection, seem to be fairly accurate within specific areas (true positive and true negative respectively) and Bernoulli NB presents decent accuracy rates in both tests. The Multinomial NB however drops to very poor levels and the recall rates are poor in the smaller test and perform noticeably worse in the larger test. Possibly this reflects the difference in vocabulary that a larger sample size creates. The true news stories remains a limitation due to the small size of material available. Any increase in testing would mean a decrease in training and fitting data. Interestingly the Precision-rate does not change at all. It is also interesting to see how the F1-score changes between the tests. The smaller test has the Multinomial in lead by far but it drops sharply in the larger test. To contrast this the Bernoullis score actually goes up slightly. Why exactly this is remains uncertain.

The word frequencies does show some difference between the real and fake news but in the end it does not appear to be any words sticking out and characterizing distinctive features of the two categories. It does not seem to be that easy to differentiate between authentic and fake. This perhaps is reasonable considering the classification results. However, there still is value in looking at the syntactic patterns of the fake news stories as a sub-area in a larger system. The larger system of defences could involve algorithms for localization of producers, bots and troll farms. This would be an example of defence by offence and targets the source. Here visualizing and mapping the spread of fake news is the important aspect, thoroughly tracing their circulation and clustering fake news sites into geographical groups [The Public Data Lab, 2017]. Another layer of defence could be algorithms that examine the story itself, by analyzing its internal content and logic. Although a rather unsuccessful attempt in this study, there is much more that can be done with the content. With a better suited and larger dataset the study forecasts a better result. However, very little points to the textual content being separable in false and fake news stories. A smarter solution perhaps would be to have an algorithm identify facts and claims stated in an article. Then let that algorithm search for confirmations on these facts. This method also has pitfalls however, because the very nature of news is to add to the human corpus of knowledge, a new turn of events and new articles may not have other sources backing it up. That does not mean the article is not true. As a final layer of defence sentiment tracking algorithms and media monitoring can be used by companies to quickly be able to counteract the effects of the fake news. A quickly published refute of the misinformation can limit the damage received.

Figure 7: An example of how the different layers could look.



6.3 Future studies

Much remains to be done within the area and no doubt research will continue. Indeed much has happened just during the writing of this paper as important elections are staged. An example is how Facebook for the, at time of writing, upcoming snap election in the UK has announced it will be running ads to raise awareness about fake news [BBC News, 2017]. More extensive studies have to be made regarding the possibility of detection by language-based classification models, based on bigger samples and perhaps more algorithms. More studies and testing also have to be made on the possibilities and viability of other methods of detection. Future research on the subject could for example examine parameters beside the raw text of news articles. Furthermore, more extensive in-depth research can be made on the market of fake news and how businesses and stocks are affected, or perhaps analysis of the results of the techniques that has been introduced in trying to deal with fake stories on social media. There is also room for further inquiry on the role and importance of individual words in both real and fake news. The possibilities for further research are in other words many.

References

J. Alderman. Trump has called dozens of things fake news. none of them are, Feb. 2017. URL https://www.mediamatters.org/research/2017/02/13/trump-has-called-dozens-things-fake-news-none-them-are/215326.

H. Allcott and M. Gentzkow. Social media and fake news in the 2016 election. *National Bureau of Economic Research*, 2017. URL http://www.nber.org/papers/w23089.

Alva. Alva, know where you stand, 2017. URL http://www.alva-group.com/us/.

BBC News. Facebook publishes fake news ads in uk papers, May 2017. URL http://www.bbc.com/news/technology-39840803.

F. Da Silva and M. Englind. Troll detection: A comparative study in detecting troll farms on twitter using cluster analysis. *DD151X Examensarbete i Datateknik, grundnivå*, 2016. URL http://www.diva-portal.org/smash/get/diva2:927209/FULLTEXT02.

M. D.Conover, J. Ratkiewicz, M. Francisco, B. Goncalves, F. Menczer, and A. Flammini. Political polarization on twitter. *AAAI Publications, Fifth International AAAI Conference on Weblogs and Social Media*, 2011. URL https://www.aaai.org/ocs/index.php/ICWSM/ICWSM11/paper/view/2847.

R. Dillet. Facebook runs full page newspaper ads against fake news in france ahead of the election, Apr. 2017. URL https://techcrunch.com/2017/04/14/facebook-runs-full-page-newspaper-ads-against-fake-news-in-france-ahead-of-the-election/.

W. Dizard. *https://www.amazon.com/Old-Media-New-Communications-Information/dp/080133277X*. Pearson, 1999. ISBN 080133277X.

Facebook Help Center. How do i mark a news story as false?, 2017. URL https://www.facebook.com/help/572838089565953.

E. Ferrara, O. Varol, C. Davis, F. Menczer, and A. Flammini. The rise of social bots. *Communications of the ACM*, pages 96–104, 2016. URL https://cacm.acm.org/magazines/2016/7/204021-the-rise-of-social-bots/fulltext.

Folk och Försvar. Rikskonferensen 2017, 2017. URL https://www.youtube.com/watch?v=h7_fTVDYzNM.

Google Search Help. Search using autocomplete, 2017. URL https://support.google.com/websearch/answer/106230?hl=en.

J. Guynn. Facebook begins flagging 'disputed' (fake) news, Mar. 2017. URL https://www.usatoday.com/story/tech/news/2017/03/06/facebook-begins-flagging-disputed-fake-news/98804948/.

D. Harper. Forces that move stock prices, 2017. URL http://www.investopedia.com/articles/basics/04/100804.asp.

G. James and E. Shearer. News use across social media platforms 2016, May 2016. URL http://www.journalism.org/2016/05/26/news-use-across-social-media-platforms-2016/.

A. Jeffries. Google's' featured snippets are worse than fake news, Mar. 2017. URL https://theoutline.com/post/1192/google-s-featured-snippets-are-worse-than-fake-news.

A. Klockars and M. Williams. Scientific yellow journalism. *Small GTPases*, 2012. URL http://dx.doi.org/10.4161/sgtp.22289.

J. Kosslyn and C. Yu. Fact check now available in google search and news around the world, Apr. 2017. URL https://blog.google/products/search/fact-check-now-available-google-search-and-news-around-world/.

L. Kung. *Strategic Management in the Media: Theory to Practice*. Sage Publications (CA), 2008. ISBN 1412903130.

D. Lazer, M. Baum, N. Grinberg, L. Friedland, K. Joseph, W. Hobbs, and C. Mattsson. Combating fake news: An agenda for research and action. May 2017. URL https://shorensteincenter.org/combating-fake-news-agenda-for-research/.

S. T. Lee. Lying to tell the truth: Journalists and the social context of deception. *Small GTPases*, 2004. URL http://dlib.nyu.edu/undercover/sites/dlib.nyu.edu.undercover/files/documents/uploads/editors/Lee_Lying.pdf.

M. Liedtke. Google targets 'fake news,' offensive search suggestions, Apr. 2017. URL https://phys.org/news/2017-04-google-fake-news-offensive.html.

V. L.Rubin, Y. Chen, and N. J.Conroy. Deception detection for news: Three types of fakes. *Proc. Assoc. Info. Sci. Tech.*, 2015. URL http://dx.doi.org/10.1002/pra2.2015.145052010083.

A. Mosseri. Working to stop misinformation and false news, Apr. 2017. URL https://newsroom.fb.com/news/2017/04/working-to-stop-misinformation-and-false-news/.

G. Nygren and G. Hök. Ukraina och informationskriget – journalistik mellan ideal och självcensurs. *Myndigheten För Samhällsskydd och beredskap*, 2016. URL https://www.msb.se/sv/Produkter--tjanster/Publikationer/Publikationer-fran-MSB/Ukraina-och-informationskriget--journalistik-mellan-ideal-och-sjalvcensur/.

E. Reilly. How bussiness must communicate in the age of twitter and fake news, Mar. 2017. URL http://www.prweek.com/article/1426241/business-communicate-age-twitter-fake-news.

M. Risdal. Getting real about fake newstext metadata from fake biased news sources around the web, 2016. URL https://www.kaggle.com/mrisdal/fake-news.

C. Shaoi, G. L. Ciampaglia, A. Flammini, and F. Menczer. Hoaxy: A platform for tracking online misinformation. *Proceedings of the 25th International Conference Companion on World Wide Web*, pages 745–750, 2016. URL http://dl.acm.org/citation.cfm?id=2872518.2890098.

C. Silverman, J. Singer-Vine, and L. T. Vo. Fake news, real ads: Fake news publishers are still earning money from big ad networks, Apr. 2017. URL https://www.buzzfeed.com/craigsilverman/fake-news-real-ads?utm_term=.eaYoNLBKR.

A. Smith and V. Banic. Fake news: How a partying macedonian teen earns thousands publishing lies, Dec. 2016. URL http://www.nbcnews.com/news/world/fake-news-how-partying-macedonian-teen-earns-thousands-publishing-lies-n692451.

J. Soll. The long and brutal history of fake news, Dec. 2016. URL http://www.politico.com/magazine/story/2016/12/fake-news-history-long-violent-214535.

Stanford History Education Group. Evaluating information: The ccornerstone of civic online reasoning, Nov. 2016. URL https://sheg.stanford.edu/upload/V3LessonPlans/Executive%20Summary%2011.21.16.pdf.

D. Steinberg. Why data scientists split data into train and test, Mar. 2014. URL http://info.salford-systems.com/blog/bid/337783/Why-Data-Scientists-Split-Data-into-Train-and-Test.

Stockholms Tingsrätt. Aktiebloggande läkarstudenter döms för grov otillbörlig marknadspåverkan, Dec. 2016. URL http://www.stockholmstingsratt.se/Om-tingsratten/Nyheter-och-pressmeddelanden/Aktiebloggande-lakarstudenter-doms-for-grov-/.

F. Stroud. fake news, 2017. URL http://www.webopedia.com/TERM/F/fake-news.html.

The Ministry of Foreign Affairs of the Russian Federation. Published materials that contain false information about russia, 2017. URL http://www.mid.ru/en/nedostovernie-publikacii.

The Public Data Lab. A field guide to fake news, 2017. URL http://fakenews.publicdatalab.org/.

B. Timmers. Everything you wanted to know about fake news, Dec. 2016. URL https://integralads.com/resources/everything-wanted-know-fake-news.

M. Wendling. The saga of 'pizzagate': The fake story that shows how conspiracy theories spread, Dec. 2016. URL http://www.bbc.com/news/blogs-trending-38156985.

W. Williams. Google rolls out new 'fact check' tool worldwide to combat fake news, Apr. 2017. URL http://www.csmonitor.com/Technology/2017/0407/Google-rolls-out-new-Fact-Check-tool-worldwide-to-combat-fake-news.

M. Wohlsen. Stop the lies: Facebook will soon let you flag hoax news stories, May 2015. URL https://www.wired.com/2015/01/facebook-wants-stop-lies-letting-users-flag-news-hoaxes/.

S. Zhang and Authors. The convergence of conventional media and new technology in the cases of the new york times and nbc. 2012. URL http://opensiuc.lib.siu.edu/gs_rp/293/.