

Asia Pacific College

School of Computing and Information Technologies

Magallanes, Makati City

**IDENTIFYING FAKE NEWS**

Project Documentation Submitted

to the Faculty of School of

Computing and Information Technologies

of

Asia Pacific College

By

Marc Anthony Nares

Aleo De Leon

Wyatt Holgado

BSCS-SS / SS152

To

Mr. Manuel Sebastian Sanchez

# Table of Contents

[Table of Contents 2](#_Toc509989818)

[Abstract 3](#_Toc509989819)

[I. Introduction 4](#_Toc509989820)

[Project Context 4](#_Toc509989821)

[Purpose and Description 4](#_Toc509989822)

[Objectives 4](#_Toc509989823)

[Scope and Limitations 4](#_Toc509989824)

[II. Review of Related Literature 5](#_Toc509989825)

[III. Technical Background 10](#_Toc509989826)

[IV. Design and Methodology 12](#_Toc509989827)

[VI. Conclusions and Recommendation 14](#_Toc509989828)

[VII. References 15](#_Toc509989829)

# Abstract

This document circles around fake news on the internet. In this research, the researchers will create a web extension in google chrome that identifies if a link posted on social media sites leads to an article that is fake or not. It will be implemented through a web extension that will only run on social media sites. Considering that there are various ways in identifying fake news, this document will be providing solutions on how to identify fake news.

KEYWORDS: Fake news, classification

# Introduction

## **Project Context**

Fake news is a deliberate misinformation or hoax that spreads via traditional print, broadcast news media or online social media (Novotny, 2017). It misleads people and make the world less informed. It harms the community and the industry at an alarming level (Stecula, n.d.).

Mocha Uson a secretary of the Presidential Communications Operations Office(PCOO) shared a post of a Duterte supporter which claimed that a young girl was murdered due to the drug problem in the Philippines. She was reportedly outraged as to why the Commission on Human Rights didn’t focus on the incident. However, the link showed a photo of a nine-year-old Brazilian girl who was raped and murdered in 2014. BBC called her on it and she later took down the post (ARIAS, 2017). With the way on how fake news is spreading, it would be ideal to have an application that will help identify fake news online.

## **Purpose and Description**

The main purpose of this project is to give awareness to online users from being tricked by fake news since according to the National Bureau of Investigation(NBI) there is no government agency that monitors fake news in the Philippines. The proponents will use an algorithm that can identify fake news in social media sites. It will be implemented through an extension that will work only on Google Chrome and it will only identify fake news in social media sites.

## **Objectives**

The objectives of this study are:

1. To identify fake news in the internet.
2. To create a web extension that can identify whether a link clicked by a user leads to a fake news article or not.

## **Scope and Limitations**

The study is about identifying fake news in the internet and to create a web extension that can help spot fake news. The gathered datasets will come from two online websites (kaggle.com, Opensources). The web extension will only work on the latest and upcoming versions of Google Chrome and it will not work on any other browsers. This study is only limited in identifying fake news on social media sites and it will not identify fake news on other sites.

# Review of Related Literature

In this chapter, the researchers will be presenting articles and related readings about this study. Moreover, this chapter suggests about fake news, web extension or plugin and image-based captcha. Furthermore, this chapter will also inform that the researchers to create a web extension that will scan the links for fake news.

**Fake News**

According to the article, (Japantimes 2017) Facebook and Google are now testing apps to spot fake news. A research scientist (Ciampaglia, G.2017) at Indiana University, created a tool that tracks unconfirmed claims that spread through the internet. They started creating this tool after seeing a lot of fake content spreading affecting many. Their studies have yielded tools that help track how alternative facts spread, and others that let you identify fake news. Based on this article, the researchers proved that fake news really gives us a problem, not just a problem, but a huge problem in the social media.

In addition to Japantimes (2017) they also tackle about the rumors on Twitter. Mitra, M. (2017), a doctoral student at the Georgia Institute of Technology, started a project three years ago that sees how misinformation and fake news spread through Twitter. Another project led by Georgia Tech PhD candidate Tanushree Mitra scanned 66 million tweets linked to nearly 1,400 real-world events to build a language model that identified words and phrases that lead to strong or weak perceived levels of credibility on Twitter. An Indian-origin researcher put together a system to analyse whether a tweets is credible or not. Twitter opened up its verified service to public in order to prove one’s real identity to the followers of one’s handle, but there is a need for something to value one’s tweet. According to (Mitra 2017), studies are being conducted on social media credibility in recent years but very little is known about what types of words or phrases create credibility perceptions during rapidly unfolding events. These articles will help the researchers build a language model and a system that identifies and analyzes words and phrases that determines the levels of credibility.

Another tool (Mencer, F.2016) was created at Indiana University that visualizes claims in the news and fact checks those claims. That tool called Hoaxy, Ciampaglia, G. (2017) also coordinated on this project. They said a user can enter a claim into the service’s website and see results that show both incidents of the claim in the media and attempts to fact-check it by independent organizations such as snopes.com, politifact.com and factcheck.org. These results can then be selected to generate a visualization of how the articles are shared across social media. The site's search results display headlines that appeared on sites known to publish inaccurate, unverified or satirical claims based upon lists compiled and published by reputable news and fact-checking organizations. Based on this article, it will help the researchers on adding this feature to improve the web extension.

Badaskar et al. (n.d.) focused on the topic about fake news. The researcher measures the accuracy of number of articles that are classified correctly as real or fake and the log-probability that measures the classification decision, the researchers used a classification-task based formalism for evaluating various features with the objective of improving conventional language models. Features that perform well in the task for classifying real and fake articles. These features are syntactic, semantic and empirical. This article contributes significantly to the semantic features on classification task accuracy.

McIntire (2017) conducted a study on a model that fake news articles could classify accurately. He tested his model in his fake news datasets. Remarkably, it was able to correctly identify them as fake. But It turns out that his hypothesis predicting model would struggle at classifying news articles was quite wrong. McIntire thought that it would be excellent and managed to surpass that by a significant margin. McIntire created a model that convinced that it is as good as it appears. To understand why this might have happened, McIntire takes a look at the fake and real words in the data. The researcher borrowed a technique from the author named Kevin Markham which he started off with table 2 columns wide and 10558 rows long. The first column represented how many times a given word appeared in articles classified as fake and the second column was how many times a word appeared in a real article.  Then he divided the fake column by the total number of fake articles the model classified and so on for the real column. Next, added the number one to every value in the data because he created a new column of Fake:Real ratios and didn’t want to get an error by dividing zero. This Fake:Real is a pretty good but by no means perfect metric of just how fake or real a certain word. The logic is pretty simple, if a word shows up a lot of fake articles and rarely in real articles then its fake to real ratio will score be pretty high. In addition, McIntire (2017) think that a standard Naive Bayes text classification model can help and provide insight into addressing the issue, a more powerful tool should be employed to fight fake news (McIntire, 2017). This article will also help on analyzing the fake and real texts and words in the data.

According to Ruchansky et al. (2017), they study about the problems of fake news detection. The researchers read several articles that addressed the problem by focusing on the text, the response an article receives, or the users who source it. They claim that it is important to combine all three. They propose the CSI model which is composed of three modules. First module, Capture, captures the abstract temporal behavior of user encounters with articles, as well as temporal textual and user features, to measure response as well as the text. The second is score; it estimates a source suspiciousness score for every user, which is then combined with the first module to produce a predicted label for each article. The separation into modules allows CSI to output a prediction separately on users and articles, incorporating each of the three characteristics, resulting into combining the information for classification. The model demonstrated the accuracy of CSI in classifying fake news articles.

The researchers read an article about a new database of fake news sites details how much fakery has spread from Trump v. Clinton to local news. The database published by PolitiFact, one of Facebook’s partners in its hoax-combatting program, showed a list of 156 sites that contains fake stories. The sites are divided into four categories; parody or joke sites, which contain some disclaimer somewhere that are meant to be satire; news imposter sites, these sites attempt to trick readers into thinking they are newspapers or radio or television stations; fake news sites, most of these sites join services that allow them to post a collection of provocative ads to make money off clicks; and lastly, sites that contain some fake news (Owen & Gillin, 2017). Based on this article, the researchers will need a database of fake news sites that divided into many categories.

Verstraete (2016) conducted a study entitled ‘Identifying and Countering fake news’. The researchers presented a study that addresses a useful classification of fake news based on their creators’ intent to deceive and motivation. In particular, it identifies four key categories: satire, hoax, propaganda, and trolling. This analytical framework will help policymakers and commentators alike by providing accuracy to debates over the issue. The study identifies key structural problems that make it difficult to design interventions that can address fake news effectively. These include the ease with which authors can produce user-generated content online. Authors often have a mixture of motives in creating content, making it less likely that a single solution will be effective. Consumers of fake news have limited incentives to invest in challenging or verifying its content, particularly when the material supports their existing beliefs and perspectives. Finally, the researchers said that fake news rarely appears alone, it is frequently mingled with more accurate stories, such that it becomes harder to categorically reject a source. Then, the report classifies existing and proposed interventions based upon the four regulatory modalities catalogued by Larry Lessig; law, code, social norms and markets. It assesses the potential and shortcomings of extant solutions. The most important is the study offers a set of model interventions, classified under the four regulatory modalities, to generate discussion and to provide a starting point for policymakers who want to reduce the effects of fake news. The goal of researchers is to create a foundation to help advance dialogue about fake news and to suggest tools that might mitigate it

According to Goel et.al (2016) the team developed a technology project that tries to build an algorithm that authenticates what is real and what is fake on Facebook. The chrome-extension goes through user’s Facebook feed in real time as the user browse it and verifies the authenticity of posts. Posts can be status updates, images or links. The backend AI checks the facts within these posts and verifies them using image recognition, keyword extraction, and source verification and a twitter search to verify if a screenshot of a twitter update posted is authentic. The posts then are visually identified on the top right corner in accordance with their trust score. If a post is found to be false, the AI tries to find the truth and shows it to users. Each time the user posts or shares content, chat bot uses a web hook to get a call. This chat bot then uses the same backend AI as content consumption to determine if the new post by the user contains any unverified information. If so, the user is notified and can choose to either take it down or let it exist. The chrome-extension is built using javascript that uses advanced web scraping techniques to extract links, posts, and images. This is then sent to an AI. The AI is a collection of API calls that we collectively process to produce a single trust factor. The APIs include Microsoft's cognitive services such as image analysis, text analysis, Bing web search, Twitter's search API and Google's Safe Browsing API. The backend is written in Python and hosted on Heroku. The chat bot was built using Facebook's wit.ai.

According to Shu et.al (2017) in the article, Fake News Detection on Social Media: A Data Mining Perspective, they discovered the fake news problem by reviewing existing literature in two phases: characterization and detection. But the researchers focused mainly in the detection phase. In the detection phase, the researchers studied existing fake news detection approaches from a data mining perspective, including feature extraction and model construction. They also discussed the datasets, evaluation metrics, and promising future directions in fake news detection research and expand the field to other applications. That application named BS Detector. The dataset is gathered from a browser extension called BS detector developed for checking news authenticity. The extension searches all links on a given webpage for references to unreliable sources by checking against a manually complied list of domains. The labels are the outputs of BS detector, rather than human annotators. In addition to BS Detector, the extension is powered by the OpenSources. This research gives the idea of using an extension to identify fake news.

Conroy et.al (2015) conducted a study about linguistic and network‐based approaches it showed high accuracy results in classification tasks within limited domains. The study presented methods that needs more evaluation, and provides a basis for the design of a comprehensive fake news detection tool. Techniques based from different approaches may be developed together in a system. For example, linguistic processing used for analyzing words on multiple layers. Linguistic approach extracted and analyzed messages to associate language patterns with deception while network approach, the network information, like message metadata or structured knowledge network queries can be attached to provide cumulative deception measures. These machine learning techniques are used for training classifiers to suit the analysis. The two approaches utilize the most effective deception detection methods for the implementation of a fake news detection tool. According to Kinmont, K, (2017) they conducted a study on detection on fake news in social media site, twitter. The researcher also used the method linguistic approach for credibility checking.

In addition to fake news detection (Singh el.al) , proposed a project about text analysis based on computational approach to automatically identify or detect fake news. Based on the researcher’s dataset it comes from the kaggle fake news dataset provided by SBP-BRIMS. The dataset contains a news reports that marked as fake but the data needs to be compared with other dataset that contains valid news reports. For research purposes, they created another dataset of valid news reports that contains selected news agencies. The methods that researchers used are linguistic analysis and word count and created multiple machine learning models that gave a best prediction results. Based on the results, text-processing based machine learning correctly detected that the news report is fake with 87% of accuracy. The team can use Singh’s method for detecting fake news.

**Image-based CAPTCHA**

According Saini et.al (2013) CAPTCHA was invented by Luis Von Ahn, Manuel Blum, Nicholas J. Hooper and John Langford in 2000. CAPTCHA stands for Completely Automated Public Turing Test to tell Computers and Humans Apart. CAPTCHA is based on identifying the distorted text, the color of image, object or the background. The researchers will focus mainly in image. Image-based CAPTCHAs are challenge tests in which the users have to guess those images that have some similarity. CAPTCHA follows a reverse Turing test in which CAPTCHA program acts like an evaluator and participant acts like a user. If the test is passed by the user, then he is considered as human otherwise it is a bot.

In addition to Saini et.al (2013) CAPTCHA is a defensive system that acts as a tool to prevent bots from abusing online services on the internet. The system also gives the user a challenge, when the user gives correct answer then he is considered as human otherwise, a bot. Elson et al (2007) developed an image-based CAPTCHA but according to the researchers, it is originally called Asirra that relies on a database of images of pets from several animal shelters. In order to pass the CAPTCHA, the user must select all images depicting either cats or dogs from a set of random images from both categories. The system takes advantage of the fact that users can easily distinguish between semantically different visual content, while the problem is difficult for computers.

The researchers will conclude that these articles and journals will help them to learn more about web browser extensions and fake news that will apply to the study about identifying fake news using web extension. Also, with the help of journals that provides a description about image-based CAPTCHA, the researchers will use these for security purposes. The researchers also read about the open source web extension to flag fake news. The project named projectFiB.

# Technical Background

**Web Extension**

Extensions are bits of code that modify the functionality of a web browser. They are written using standard Web technologies - JavaScript, HTML, and CSS - plus some dedicated JavaScript APIs. Among other things, extensions can add new features to the browser or change the appearance or content of particular websites. Basically, the researchers will create an extension that can identify fake news in social media sites.

**JavaScript**

JavaScript is a programming language that enables you to create dynamically updating content, control multimedia, animate images, and pretty much everything else. The researchers will use this language to create an extension.

**IndexedDb Database**

A collection of data is called a Database. In order to manage a data in a database, it will need to have a Database Management System (DBMS). DBMS can help the user interact with the database itself and it can also help them to handle and capture the data. The researchers will use IndexedDb. IndexedDB is a low-level API for client-side storage of significant amounts of structured data, including files/blobs. This API uses indexes to enable high performance searches of this data.

**Image-based CAPTCHA**

CAPTCHA stands for Completely Automated Public Turing Test to tell Computers and Humans Apart. Using this feature, it will provide security through challenge tests based on images to verify users if it is a human or a bot.

**Machine Learning**

Machine Learning is an application of artificial intelligence (AI) that performs tasks intelligently without being programmed for those activities. It is also the science of getting computers to learn and act like humans do and improve their learning by feeding them data and information in the form of observation.

**Linguistic Approaches**

Most fake news sites use their language strategically to avoid being caught. Linguistic approaches base their assessment on the content of the news specifically the usage of language. This approach aims to identify the language leak such as the patterns of pronoun, conjunction, and negative emotion word usage found in the content of a news in a webpage.

**Fakebox**

Fakebox contains a Machine Learning model that uses Natural Language Processing to understand news articles and blog posts. It analyzes the content of an article and predicts whether its fake or not. Facebox checks for:

* Title - Titles can be clickbait or biased
* Content - The textual content of an article can be analyzed to determine if its written like real news or not
* Domain name - Some domains are known for hosting certain types of content, Fakebox knows about the most popular sites

# Design and Methodology

A web extension is a plug-in that extends the functionality of a web browser. The researchers will create a web extension that can identify fake news and will only run if the link is clicked. The extension will be uploaded to Google Web Store and it will be accessible by Chrome users.

1. **Gathering of Data**

The data for fake news list will be gathered from (kaggle.com, Opensources.com). Kaggle.com is a data science website. Opensources.com provide a continuously updated database of information sources for researchers to leverage in the fight against fake, false, and misleading news. It is also maintained by professionals who have analyzed each source, looking for overall inaccuracy extreme biases, lack of transparency, and other kinds of misinformation (Opensources, n.d.). The researchers will download the file from the source manually.

1. **Analyzing the link**

There are two phases for analyzing the link. First the web extension would compare the link from the list of fake news sites. If the link is in the list of fake news sites the extension will warn the user that the page may contain false information. Second if the link was not in the list of fake news sites then the web extension will analyze the news articles using an API from Fakebox to assess whether they are fake or not by looking at a range of available aspects of an article (title, content and URL).

1. **Updating**

*Figure SEQ Figure \\* ARABIC 1*

*Figure 2*

*Figure 3*

The extension is automatically updated without user intervention through Google developer’s dashboard which automatically uploads the file through Google’s server. The researchers will update the data of fake new list monthly.

1. **Reporting System**

In the extension, the team will put a reporting system in which a user will have an option to report a page by choose an appropriate tag for the link. The researchers will be using indexedDB for the client-side storage. The data from the clients will be synced to MySQL. It will show the user the reason why it is flagged as fake (Fake News, Clickbait, Caution, Bias) by getting the highest number of votes. If the link is confirmed fake the system will notify the user with a dialog box popped up. A captcha would also be used to prevent and spammers and bots.

**Below are the tags.**

* **Fake News-** Sources that entirely fabricate information, disseminate deceptive content, or grossly distort actual news reports
* **Clickbait-** Sources that provide generally credible content, but use exaggerated, misleading, or questionable headlines, social media descriptions, and/or images.
* **Caution -** Sources that may be reliable but whose contents require further verification.
* **Bias** - Sources that come from a particular point of view and may rely on propaganda, decontextualized information, and opinions distorted as facts.

# Conclusions and Recommendation

Fake news can be lessened or better yet eliminated, since people keep falling prey for such posts; it needs to be stopped. This paper is meant to do just that; the program we are proposing is meant to detect if the news posted in your social media websites, which would literally get marked as a fake if it is one. To identify fake news, one must check URL. This program will be a plugin for the users’ internet browsers, but will only work for social media sites, and will not detect if posted news are fake from other websites. If this program could be implemented, there will definitely be a big change in how news would be spread, since users will no longer fall for faulty news articles and posts. This extension will also help users from falling to a phishing site.

To help increase the accuracy of the extension, the researchers recommend adding more features for classifying data and to work not only in Chrome but also on other browsers.

# References

Arias, J. (2017, October 4). *Inquirer.net*. Retrieved from Preen: http://preen.inquirer.net/58185/a-list-of-mocha-usons-fake-news-posts

Bisson, D. (2015, September 13). *Most Suspicious TLDs Revealed by Blue Coat Systems*. Retrieved from The State of Security: https://www.tripwire.com/state-of-security/security-data-protection/cyber-security/most-suspicious-tlds-revealed-by-blue-coat-systems/

Conroy, N. J., Rubin, V. L., & Chen, Y. (n.d.). Automatic Deception Detection: Methods for Finding Fake. 1-4.

*DeZyre.* (2016, January 29). Retrieved from https://www.dezyre.com/article/top-10-machine-learning-algorithms/202

*Facebook, Google testing apps to spot fake news but experts say key is being educated*. (2017, February 7). Retrieved from Japan Times: https://www.japantimes.co.jp/news/2017/02/07/business/facebook-google-testing-apps-spot-fake-news-experts-say-key-educated/#.WlM6JlWWZpg

Google. (n.d.). *Google Recaptcha*. Retrieved from Google: https://www.google.com/recaptcha/intro/android.html

IBM. (n.d.). *How SVM works*. Retrieved from IBM: https://www.ibm.com/support/knowledgecenter/en/SS3RA7\_15.0.0/com.ibm.spss.modeler.help/svm\_howwork.htm

*Indiana University researchers launch tool to understand spread of fake news*. (2016, December 21). Retrieved from IU BLOOMINGTON NEWSROOM: http://archive.news.indiana.edu/releases/iu/2016/12/iu-hoaxy.shtml

Inquirer. (2017, February 7). *Fighting fake news isn’t just up to Facebook and Google*. Retrieved from Inquirer.net: http://technology.inquirer.net/58483/fighting-fake-news-isnt-just-facebook-google

Kiely, E., & Robertson, L. (2016, November 18). *FactCheck.org*. Retrieved from How to Spot Fake News: https://www.factcheck.org/2016/11/how-to-spot-fake-news/

Kinmont, K. (n.d.). Fake News Detection in Twitter. 4-4.

McIntire, G. (n.d.). *Machine Learning Finds “Fake News” with 88% Accuracy*. Retrieved from KDnuggets: http://www.kdnuggets.com/2017/04/machine-learning-fake-news-accuracy.html

*Opensources*. (n.d.). Retrieved from http://www.opensources.co/

Palmisano, S. (2016, December 29). *FAKE NEWS LEADS TO REAL PROBLEMS*. Retrieved from Eagle News: http://eaglenews.org/opinion/fake-news-leads-to-real-problems/

Patel, S. (2017, May 3). *Chapter 2 : SVM (Support Vector Machine) — Theory.* Retrieved from Machine Learning 101: https://medium.com/machine-learning-101/chapter-2-svm-support-vector-machine-theory-f0812effc72

Philstar. (2017, October 4). Retrieved from Philstar: https://www.philstar.com/headlines/2017/10/04/1745491/no-govt-agency-monitoring-fake-news-social-media-nbi-exec-says

Project, M. I. (2016, April 17). *A new understanding: What makes people trust and rely on news*. Retrieved from AmericanPress Institute: https://www.americanpressinstitute.org/publications/reports/survey-research/trust-news/

Saini, B. S., & Bala, A. (n.d.). A Review of Bot Protection using CAPTCHA for Web Security. *Journal of Computer Engineering*, 1-7.

Singh, V., Dasgupta, R., Sonagra, D., Raman, K., & Ghosh, I. (n.d.). Automated Fake News Detection Using Linguistic Analysis. 1-3.

Stecula, D. (2017, July 27). *The real consequences of fake news*. Retrieved from The Conversation: http://theconversation.com/the-real-consequences-of-fake-news-81179

vvv. (n.d.).

*Why use SVM?* (n.d.). Retrieved from alteryx community: http://blog.yhat.com/posts/why-support-vector-machine.html