

Asia Pacific College

School of Computing and Information Technology

Magallanes, Makati City

**IDENTIFYING FAKE NEWS ON SOCIAL MEDIA WEBSITES**

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

Ms. Rhea Luz

# Contents

[Contents 2](#_Toc497489356)

[Abstract 3](#_Toc497489357)

[I. Introduction 4](#_Toc497489358)

[Project Context 4](#_Toc497489359)

[Purpose and Description 4](#_Toc497489360)

[Objectives 4](#_Toc497489361)

[Scope and Limitations 5](#_Toc497489362)

[II. Review of Related Literature 6](#_Toc497489363)

[III. Technical Background 10](#_Toc497489364)

[IV. Design and Methodology 11](#_Toc497489365)

[VI. Conclusions and Recommendation 14](#_Toc497489366)

[VII. References 15](#_Toc497489367)

# Abstract

This document circles around fake news on the internet. In this research, the researchers will create a system that identifies if a link posted on social media sites leads to an article that is fake or not. The researchers will use an algorithm for classifying the link and 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 in an alarming level (Stecula, n.d.).

Mocha 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 photo was 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 application that 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 NBI there are no government agency that monitors fake news in the Philippines. This extension will also provide security against phishing sites. The proponents will create 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 create a system that determine whether a link leads to a fake news article or not.
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. The researchers will use web programming languages to create a web extension for identifying fake news. The gathered datasets will come from three online websites (kaggle.com, Politifact, Opensources) The web extension will only work on the latest and upcoming versions 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 an 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 aims to create a web extension that will scan the links for fake news.

**Fake News**

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. The semantic features contribute significantly to the 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).

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).

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 (Verstraete, 2016).

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.

**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.

**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.

**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.

# Design and Methodology

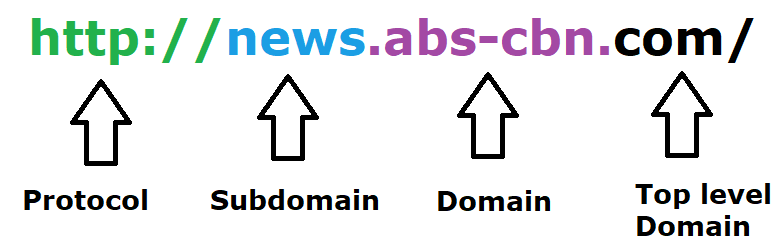
A web extension is a plug-in that extends the functionality of a web browser. The researchers will create a web extension that can identify fake news and will only run if the link is clicked. When the link is clicked, the system will compare the news link article from the list of fake news links and from the list of authentic news gathered by the team. The data for fake and authentic news list will be gathered from **(**kaggle.com, Politifact, Opensources**),** and it will be saved in a database. If the link is not in the list, the system then will check the domain of the URL.

Figure 1

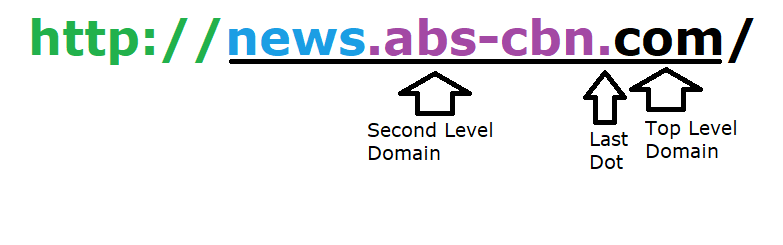
Everything that occurs before the last dot is the second level domain, everything that occurs after the last dot is the top-level domain. In this example, news and abs-cbn are part of the second level domain and “com” is the top-level domain. Well-known top-level domains are edu, gov, org and com.

Figure 2

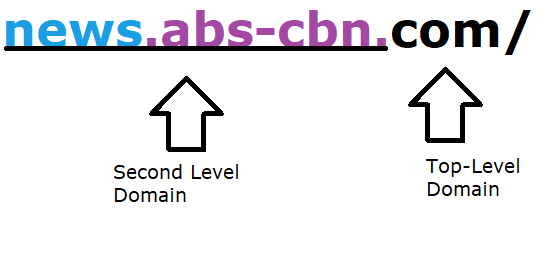


Figure 3

In this example (Figure 3) this is not a fake news site since .com is the top-level domain

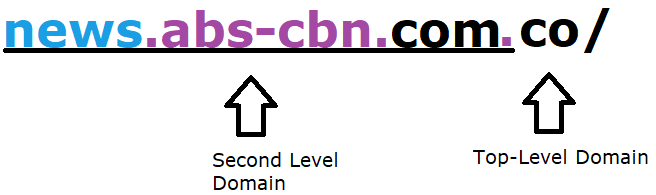


Figure 4

In this example (Figure 4) this news site is fake since .co is the top-level domain. The fake news site tries to be tricky with.com.co, but the .com is just part of the second level domain.

In the extension, the team will put a reporting system in which a user will have to choose an appropriate tag for the link.

* **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.

If the page has 5 or more reports the system will automatically flagged it as a fake website and it will show the user the reason why it is flagged as fake by getting the highest number of chosen tag (Fake News, Clickbait, Caution, Bias) . If the link is confirmed fake the system will notify the user with a dialog box popped up, but if the link article is authentic then the system will also notify the user.

The extension will be developed using web programming languages and it will be compatible to Google Chrome.

# 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 a 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 sites.

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

*Dezyre*. (2017, June 20). Retrieved from dezyre: https://www.dezyre.com/article/top-10-machine-learning-algorithms/202

Eugene Kiely, L. R. (2016, November 18). *FactCheck.org*. Retrieved from http://www.factcheck.org/2016/11/how-to-spot-fake-news/

Greg. (2017, January 24). *yhat*. Retrieved from yhat: http://blog.yhat.com/posts/why-support-vector-machine.html

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

*Machine Learning 101*. (n.d.). Retrieved from Medium: https://medium.com/machine-learning-101/chapter-2-svm-support-vector-machine-theory-f0812effc72

Mark Verstraete1, D. E. (n.d.). *Identifying and Countering fake news*. Retrieved from https://webcache.googleusercontent.com/search?q=cache:JZsN3vbXxE4J:https://law.arizona.edu/sites/default/files/asset/document/fakenewsfinal.pdf+&cd=2&hl=en&ct=clnk&gl=ph

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

Media Insight Project. (2016, March 17). *American Express Institute.* Retrieved from https://www.americanpressinstitute.org/publications/reports/survey-research/trust-news/

Novotny, E. (2017, August 1). *PennState*. Retrieved from libraries.

Palmisano, S. (2016, December 29). *EagleNews*. Retrieved from eaglenews: http://eaglenews.org/opinion/fake-news-leads-to-real-problems/

Patag, K. J. (2017, October 4). *Phil Star Global*. Retrieved from Phil Star Global: http://www.philstar.com/headlines/2017/10/04/1745491/no-govt-agency-monitoring-fake-news-social-media-nbi-exec-says

*Rappler*. (2017, April 10). Retrieved from http://www.rappler.com/technology/social-media/166326-how-to-spot-fake-news-facebook

Sathyanarayana, S. V. (n.d.). Data classification using Support vector. *Data classification using Support vector*.

*stackexchange*. (n.d.). Retrieved from crossvalidated: https://stats.stackexchange.com/questions/23391/how-does-a-support-vector-machine-svm-work

*Stackexchange*. (n.d.). Retrieved from ComputerScience: https://cs.stackexchange.com/questions/10304/which-classifier-is-more-accurate-for-a-svm-classification

Stecula, D. (n.d.). *The conversation*. Retrieved from theconversation: http://theconversation.com/the-real-consequences-of-fake-news-81179

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

Langford, J. (2004). *Telling Human and Computers Apart Automatically*. Retrieved from http:/www.captcha.net

Baljit, S. (2013). A Review of Bot Protection using CAPTCHA for Web Security. Retrieved from https://pdfs.semanticscholar.org/b67e/943b0c8feca317b4c44ca82d525e20132959.pdf

Jeremy, E. (2007). A captcha that exploits interest-aligned manual image categorization. Retrieved from https://www.researchgate.net/scientific-contributions/10074391\_Jon\_Howell