

**Asia Pacific College**

**School of Computing and Information Technology**

**Magallanes, Makati City**

**IDENTIFYING FAKE NEWS IN FACEBOOK**

**Project Documentation Submitted**

**To the Faculty of School of**

**Computing and Information Technologies**

**Of**

**Asia Pacific College**

**by**

**Marc Anthony Nares**

**Aleo De Leon**

**Joshua Cruz**

**Wyatt Holgado**

**BSCS-SS / SS152**

**To**

**Sir. Manuel Sebastian Sanchez**

**TABLE OF CONTENTS**

Abstract …………………………………………………………………………………………………………………………. 3

Introduction …………………………………………………………………………………………………………………… 4

Related Literature ………………………………………………………………………………………………………….. 6

Technical Background ……………………………………………………………………………………………………. 7

Design / Methodology ………………………………………………………………………………………………….. 13

Conclusions ………………………………………………………………………………………………………………….. 16

Appendices ………………………………………………………………………………………………………………….. 17

# Abstract

This document circles around fake news on the internet. Defining what fake news really is and how us people should deal with it whenever we get to encounter them on the internet, most likely in the social media. In this research the researchers will identify fake news and differentiate it from authentic news in Facebook. The researchers will also discuss the effects of it in our daily lives and in our society. Considering that there’s a lot of various ways on identifying fake news, this document will be providing conclusions and recommendations that we could follow on identifying fake news.

KEYWORDS: Facebook, Fake news, identifying

# 

# Introduction

## **Project Context**

Fake news is a deliberate misinformation or hoaxes that spread via traditional print, broadcast news media or online social media (Wikipidea, n.d.). A lot of these viral claims aren’t “news” at all, it is just stated as fake news, all but fiction and efforts to fool readers into thinking they’re for real. Fake news is nothing new, it is as same as bogus stories that are spreading and reaching more people and more quickly on the internet specially in social media (Eugene Kiely, 2016).

It is spreading and keeps popping up everywhere in the society, misleading people and makes the world less informed or, giving people wrong information that makes people to give wrong assumptions to what's really happening in the society. People don't want to see false news on social media or let's say on the internet and neither do we (Rappler, 2017). It's giving harm to the community and to the industry into an alarming level. This isn't a new problem. Each new technology requires new and creative solutions. It's up to all of us in technology, media, academia to fight it.   
 For the current state of the society, and the way how fake news is spreading, it would be ideal for people to have the tools and to have ideas on how to identify fake news and bogus stories. It is crucial for us to know how to identify fake news having the fact that it really is alarming how badly it affects the society.

## **Purpose and Description**

The main purpose of this project is to identify fake news in social media, since almost everybody goes through social media; and there are still a number of users whom believe whatever they see or read on these websites. A majority of the ones affected by the fake news are believers of “click bait”, which is a false or exaggerated title in which people see and believe instantly. A big reason why this happens is to make people interested in the site, but most users don’t or can’t open these links and articles as the “free fb” promo is famous in the Philippines.

The proponents will be using Support vector machine algorithm for classifying. Support Vector Machine is a supervised machine learning algorithm for classification or regression problems where the dataset teaches SVM about the classes so that SVM can classify any new data. It works by classifying the data into different classes by finding a line (hyperplane) which separates the training data set into classes. As there are many such linear hyperplanes, SVM algorithm tries to maximize the distance between the various classes that are involved and this is referred as margin maximization. If the line that maximizes the distance between the classes is identified, the probability to generalize well to unseen data is increased.

## **Objectives**

This research aims to help the Facebook society to prevent the spreading of fake news stories

Thus, the objectives are:

1. To create a web extension for Facebook that can Identify fake news.
2. To use SMV (Support Vector Machine) algorithm to classify the post whether it is fake or not.

## **Scope and Limitations**

The study focuses on identifying fake news in Facebook using SMV algorithm. This study will cover Facebook and fake news links.

# Review of Related Literature

In this chapter, the researchers will be presenting an articles and related readings about this study. Moreover, this chapter proposes about fake news and how fake news affects people that give them false information.

**Fake News**

Fake news, a term initially coined by the news media to describe stories on the internet posted by websites of questionable integrity. The term has since been turned back on the media, some people in America including Donald Trump he charges accusations of the fake news against the likes of CNN, the New York Times, and others. We now have a constant back and forth in the public sphere: the media accuses online outlets of being fake, and politicians, in turn, accuse the media of being fake. Fake news had affected many people (Ordway, 2017).

Fake News is circulated largely through social media. Based on web browsing data, archives of fact-checking websites, and results from a new online survey, in 2016 election, there are many fake news appeared in the three months before the election, those favoring Trump were shared a total of 30 million times on Facebook, while those favoring Clinton were shared 8 million times; the average American adult saw on the order of one or maybe several fake news stories in the months around the election, with just over half of those who remembered seeing them believing them; and people are much more likely to believe stories that favor their preferred candidate, especially if they have social media networks (Alcott, 2017). These reviews are one of the examples of how fake news affects people.

**Identifying Fake News using classifier**

According to Conroy et al. (2015), classifiers are useful for automated numerical analysis. One common example of classifiers is Support Vector Machines (SVM). It shown high accuracy results in classification. It provides a basis for the design of a comprehensive fake news detection tool. Using classifier the accuracy which measures the number of articles correctly classified as real or fake.

# Technical Background

**Theoretical Framework**

**Support Vector Machine**

Support vector machine are good at solving problems. It is used computational biology due to their high accuracy, the ability to deal with large datasets, and the flexibility in modeling different sources of data. These are controlled learning models used for analyzing data and for its classification and regression analysis using sorting algorithms. Support vector machines use its ability to create forecasts based of the given set of data (Ben-Hur et al., 2008). It is also used for specific tasks such as classification of data and analysis. SVM used for sentiment analysis of opinion mining is directly focused on two things: classifying and forecasting. The support vector machine will be able to identify new inputs to the model and then designate each input to its rightful category just as how it processes previous ones. The application of SVM ranges from text categorization, image segmentation, hand-written character recognition (Jadav et al., 2016). The researchers will use SVM to identify fake news and real news in Facebook and it will prove its accuracy.

**Text Mining**

Text mining is the extraction of information from different resources. The extracted information used to form new facts or new hypothesis to be explored further by means of experimentation (Hearst, 2009). Text mining also referred as text data mining or knowledge discovery from textual databases, states to the process of extracting non-trivial patterns or knowledge from text. It can be observed as an extension of data mining from structured databases. it involves dealing with text data that are naturally unstructured. Text mining involves information retrieval, text analysis, information extraction, clustering, categorization, visualization, database technology, machine learning, and data mining (Tan, 1998). Text mining is the process of deriving [information](https://en.wikipedia.org/wiki/Information) from [text](https://en.wikipedia.org/wiki/Plain_text). High-quality information is typically derived through the devising of patterns and trends. The goal of Text data mining is to discover or derive new information from data, finding patterns through datasets (Hearst, 1998).

**Conceptual Framework**



.

# Design and Methodology

  This chapter includes the research design of the study, the target respondents, research tools, data gathering procedure, and statistical treatment that will be used to achieve accurate data interpretation

**Nature of the Study**

The researchers propose a fake news detector as a web extension for internet browsers. We run our algorithm for identifying fake news using Support Vector Machine algorithm. The system produces promising results for identifying fake news.

**Population and Sample**

The population of interest for this study are the Facebook users. The researcher will gather the fake news links dataset from Kaggle and CBCP (Catholic Bishops Conference of the Philippines).

**Research Instruments**

The researchers used literature related to the study with topics about Identifying fake news, text mining, support vector machine.

To further support the statements of this research, other materials will be used such as EBSCO Host, and Google Scholar for factual references.

**Data Analysis**

The steps to perform data classification using Support Vector Machine algorithm are as

follows:

Step1: Define a set of n data points in an array say

X= array([[x11, x21], [x12, x22], ....... [x1n, x2n]])

Step2: Define class of each data point in a vector of list type say Y = [-1, -1, -1 .....1, 1, 1]

Step3: **F**it the SVM model using the statements

clf = svm.SVC(kernel='linear') and clf.fit(X, Y)

Step4: Get the separating hyperplane xx as x1 coordinates anf yy as x2 coordinates

w = clf.coef\_[0]

a = -w[0]/w[1]

xx = np.linspace(-1, 8, 10, 1)

yy = a\*xx - (clf.intercept\_[0])/w[1]

Step5: Get the parallels to the separating hyperplane that pass through the support vectors

b = clf.support\_vectors\_[0]

yy\_down = a\*xx + (b[1] - a\*b[0]) (positive support plane)

b = clf.support\_vectors\_[-1]

yy\_up = a\*xx + (b[1] - a\*b[0]) (negative support plane)

Step6: Plot the line, the points, and the nearest vectors to the plane using appropriate python commands

# Conclusions

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. This program will be a plugin for the users’ internet browsers, but will only work for Facebook, 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.

# Appendices

# References

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

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

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

*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*.

*Wikipidea*. (n.d.). Retrieved from https://en.wikipedia.org/wiki/Fake\_news

Hunt Allcott (2017). Retrieved from

http://www.nber.org/papers/w23089.pdf

D. M. Ordway, "Journalist's Resource," [Online]. Available:

https://journalistsresource.org/studies/society/internet/fake-news-conspiracy-theories-journalism-research?utm\_source=feedburner&utm\_medium=feed&utm\_campaign=Feed%3A+journalistsresource+

%28Journalist%27s+Resource%29.

http://onlinelibrary.wiley.com/doi/10.1002/pra2.2015.145052010082/full