

**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 main factors on 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 for clarifications and recommendations that we could follow on identifying fake news.

KEYWORDS: Facebook, Fake news, identifying

# 

# Introduction

## **Project Context**

Fake news is a type of yellow journalism that is consisting deliberate misinformation or hoaxes spread via traditional print and broadcast news media or online social media [1]. 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 [2].

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 [3]. It's giving harm to our community and our 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 our 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 our society.

## **Purpose and Description**

The main purpose of this project is to prevent

## **Objectives**

We now live in the age of fake news. People are so eager to accept anything put in front of them that they fail to recognize how false the information is [4]. Facebook users often share fake news or even death hoax that can mislead people. Eric Trump the son of Donald Trump tweeted an article about paid protestors from the domain “abcbews.com.co” that reinforced right wing conspiracy theories. The article was completely fabricated, but it resembled the real ABC News enough to fool those who weren’t paying attention [5]. The real-life consequences of fake news are unclear. In many social sites people are unaware on how to identify fake news that can lead to confusions. There are web plugins that can detect fake news like B.S Detector plugin that detects fake news base on users vote, but it cannot be guaranteed that it will work effectively since it is an open source people can add or remove websites links [6].

This researcher aims to create a web extension for Facebook that will scan the user’s news feed to identify fake news and bogus stories that give people wrong information. Using SMV(Support Vector Matching) algorithm this will classify the post whether it is fake or not. The researchers will base the credibility of the news base on the author of the source. The author then will fall under these criteria to verify its credibility: number of shares, number of likes and number of engagements in a post. Unknown authors will be automatically marked as fake. Confirmed fake news websites will be added to database for future preferences.

## **Scope and Limitations**

The study focuses on identifying fake news in Facebook using SMV algorithm. This will help Facebook users to identify unreliable sources of news that cause confusions to users. This will not cover any social media sites besides Facebook.

# 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. Furthermore, this chapter will also inform that the researchers aims to create a web extension for Facebook that will scan the user’s news feed for fake news.

Researchers will also be presenting various articles and studies related to the research that will link previous assumptions about fake news and its present picture. Lastly, the researchers will present and confirm given contents about fake news and what are the opinions of people about fake news as stated in the collected articles and statements.

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 [7].

According to American press institute survey research (2016), the factors drive people to trust news from different resources; Adults who say accuracy is a critical reason they trust a news source has a percentage of 85%, Adults who say having the latest details is a critical reason they trust a news source has a percentage of 76%, Adults who value news reporting that’s concise and gets to the point has a percentage of 72%, Political news consumers who highly value experts and data in reporting has a percentage of 79% and lastly, the Facebook news consumers with a lot of trust in the news they see there has a percentage of 12%. its low but still, this 12 percent of people had trusted a news from Facebook and these people did not know if it is real or fake news [8].

In Facebook, there is much fake news spreading because of engagements. Many people believe that because of its popularity. Rappler posted some tips on how to spot fake news. For example, b**e skeptical of headlines.**False news stories often have catchy headlines in all caps with exclamation points. If shocking claims in the headline sound unbelievable, they probably are fake [9].

# 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 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 face recognition method using PCA with neural network back error propagation learning algorithm. In this paper a feature is extracted using **principal component analysis** and then classification by creation of back propagation neural network. We run our algorithm for face recognition application using principal component analysis, neural network and also calculate its performance by using the photometric normalization technique: Histogram Equalization and comparing with Euclidean Distance, and Normalized correlation classifiers. The system produces promising results for face verification and face recognition.

**Population and Sample**

The population of interest for this study are the Facebook users.

**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

The paper has presented a face recognition system using PCA with neural networks in the context of face verification and face recognition using photometric normalization for comparison. The researchers will need to conduct experiments to show the Neural Network Euclidean distance rules using PCA for overall performance for verification. For recognition, Euclidean distance classifier must be studied in order the show the accuracy using the original face image. Also, the results of applying histogram equalization techniques on the face image must be studied to show its impact to the performance of the system in a controlled environment.

# Appendices

# References

|  |  |
| --- | --- |
| [1] | "Wikipidea," [Online]. Available: https://en.wikipedia.org/wiki/Fake\_news. |
| [2] | L. R. Eugene Kiely, "FactCheck.org," 18 November 2016. [Online]. Available: http://www.factcheck.org/2016/11/how-to-spot-fake-news/. |
| [3,9] | "Rappler," 10 April 2017. [Online]. Available:  http://www.rappler.com/technology/social-media/166326-how-to-spot-fake-news-facebook. |
| [4] | I. Sigal, "Fast Sigal," 23 March 2017. [Online]. Available: https://www.fastcompany.com/3069189/fake-fixes-fake-news-real-problems-social-media-journalism. |
| [5] | S. Palmisano, "Eagle News," 29 December 2016. [Online]. Available: http://eaglenews.org/opinion/fake-news-leads-to-real-problems/. |
| [6] | The Self Agency, [Online]. Available: http://bsdetector.tech/. |
| [7] | 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. |
| [8] | Media Insight Project, "American Express Institute," 17 March 2016. [Online]. Available: https://www.americanpressinstitute.org/publications/reports/survey-research/trust-news/. |
| [9] | S. V. Sathyanarayana, "Data classification using Support vector," *Data classification using Support vector.* |