

**Synopsis
On
Fake news detection**

By

Depandra yadav (2000290140041)
Ashwani Chaudhary (2000290140023)
Session:2021-2022 (4th Semester)

Under the supervision of

Ms. Vidushi

KIET Group of Institutions, Delhi-NCR, Ghaziabad



**DEPARTMENT OF COMPUTER APPLICATIONS
KIET GROUP OF INSTITUTIONS, DELHI-NCR,
GHAZIABAD-201206
(MARCH- 2022)**

ABSTRACT

In our modern era where the internet is ubiquitous, everyone relies on various online resources for news. Along with the increase in the use of social media platforms like Facebook, Twitter, etc. news spread rapidly among millions of users within a very short span of time. The spread of fake news has far-reaching consequences like the creation of biased opinions to swaying election outcomes for the benefit of certain candidates. Moreover, spammers use appealing news headlines to generate revenue using advertisements via click-baits. In this paper, we aim to perform binary classification of various news articles available online with the help of concepts pertaining to Artificial Intelligence, Natural Language Processing and Machine Learning. We aim to provide the user with the ability to classify the news as fake or real and also check the authenticity of the website publishing the news.

TABLE OF CONTENTS

1. Introduction	2
2. Literature Review	3
3. Project / Research Objective	4
4. Research Methodology	5
5. Project / Research Outcome	6
6. Proposed Time Duration	7
References	8

INTRODUCTION

As an increasing amount of our lives is spent interacting online through social media platforms, more and more people tend to hunt out and consume news from social media instead of traditional news organizations.[1] The explanations for this alteration in consumption behaviors are inherent within the nature of those social media platforms: (i) it's often more timely and fewer expensive to consume news on social media compared with traditional journalism , like newspapers or television; and (ii) it's easier to further share, discuss , and discuss the news with friends or other readers on social media. For instance, 62 percent of U.S. adults get news on social media in 2016, while in 2012; only 49 percent reported seeing news on social media [1].

It had been also found that social media now outperforms television because the major news source. Despite the benefits provided by social media, the standard of stories on social media is less than traditional news organizations. However, because it's inexpensive to supply news online and far faster and easier to propagate through social media, large volumes of faux news, i.e., those news articles with intentionally false information, are produced online for a spread of purposes, like financial and political gain. it had been estimated that over 1 million tweets are associated with fake news "Pizzagate" by the top of the presidential election. Given the prevalence of this new phenomenon, "Fake news" was even named the word of the year by the Macquarie dictionary in 2016

2. Literature Review

2.1 What is fake news detection :

A type of yellow journalism, fake news encapsulates pieces of news that may be hoaxes and is generally spread through social media and other online media. This is often done to further or impose certain ideas and is often achieved with political agendas. Such news items may contain false and/or exaggerated claims, and may end up being viralized by algorithms, and users may end up in a filter bubble.

2.2 How does fake news detection works :

Essentially, **Linguistic Cue approaches detect fake news by catching the information manipulators in the writing style of the news content.** The main methods that have been implemented under the Linguistic Cue approaches are Data Representation, Deep Syntax, Semantic Analysis, and Sentiment Analysis

Project Objective

Our project is an web application which gives you the guidance of the day to day routine of fake news, spam message in daily news chanel , Facebook, Twitter, Instagram and other social media. We have shown some data analysis from our dataset which have retrieve from many online social media and display the main source till now fake news and true news are engaged.

Our project is tangled with multiple model trained by our own and also some pretrained model extracted from Felipe Adachi. The accuracy of the model is around 95% for all the selfmade model and 97% for this pretrained model. This model can detect all news and message which are related to covid-19, political news, geology ,etc.

4- Research Methodology

3.1 Methodology used

- **Logistic Regression:**

The logistic function, also called the sigmoid function was developed by statisticians to describe properties of population growth in ecology, rising quickly and maxing out at the carrying capacity of the environment. It's an S-shaped curve that can take any real-valued number and map it into a value between 0 and 1, but never exactly at those limits.

$$\text{sigmoid}(Z) = 1 / (1 + e^{-Z})$$

$$\text{Hypothesis} \Rightarrow Z = WX + B$$

$$h_{\Theta}(x) = \text{sigmoid}(Z)$$

- **Decision Tree Classification :**

Decision Tree is a Supervised learning technique that can be used for both classification and Regression problems, but mostly it is preferred for solving Classification problems. It is a tree-structured classifier, where internal nodes represent the features of a dataset, branches represent the decision rules and each leaf node represents the outcome.

- **Gradient Boosting Classifier :**

Gradient Boosting is a popular boosting algorithm. In gradient boosting, each predictor corrects its predecessor's error. In contrast to Adaboost, the weights of the training instances are not tweaked, instead, each predictor is trained using the residual errors of predecessor as label.

Passive Aggressive Classifier Algorithm :

Passive-Aggressive algorithms are generally used for large-scale learning. It is one of the few '**online-learning algorithms**'. In online machine learning algorithms, the input data comes in sequential order and the machine learning model is updated step-by-step, as opposed to batch learning, where the entire training dataset is used at once. This is very useful in situations where there is a huge amount of data and it is computationally infeasible to train the entire dataset because of the sheer size of the data. We can simply say that an online-learning algorithm will get a training example, update the classifier, and then throw away the example.

5.1 References

- [1] Kai Shu, Amy Sliva, Suhang Wang, Jiliang Tang, and Huan Liu, "Fake News Detection on Social Media: A Data Mining Perspective" arXiv:1708.01967v3 [cs.SI], 3 Sep 2017
- [3] M. Granik and V. Mesyura, "Fake news detection using naive Bayes classifier," 2017 IEEE First Ukraine Conference on Electrical and Computer Engineering (UKRCON), Kiev, 2017, pp. 900-903.
- [4] Fake news websites. (n.d.) Wikipedia. [Online]. Available: https://en.wikipedia.org/wiki/Fake_news_website. Accessed Feb. 6, 2017
- [5] Cade Metz. (2016, Dec. 16). The bittersweet sweepstakes to build an AI that destroys fake news.
- [6] Conroy, N., Rubin, V. and Chen, Y. (2015). "Automatic deception detection: Methods for finding fake news" at Proceedings of the Association for Information Science and Technology, 52(1), pp.1-4.

1.