

COVID-19 FAKE NEWS DETECTION MODEL ON SOCIAL MEDIA DATA USING MACHINE LEARNING TECHNIQUES

Kelvin Liew Kai Xuan
Faculty of Computing
Universiti Malaysia Pahang
Al-Sultan Abdullah
Pahang, Malaysia
liewliew520@gmail.com

Mohaiminul Islam Bhuiyan
Faculty of Computing
Universiti Malaysia Pahang
Al-Sultan Abdullah
Pahang, Malaysia
rafe.aust@gmail.com

Nur Shazwani Kamarudin*
Faculty of Computing
Universiti Malaysia Pahang Al-
Sultan Abdullah
Pahang, Malaysia
nshazwani@ump.edu.my

Ahmad Fakhri Ab. Nasir
Faculty of Computing
Universiti Malaysia Pahang
Al-Sultan Abdullah
Pahang, Malaysia
afakhri@ump.edu.my

Muhammad Zulfahmi Toh Abdullah
Faculty of Computing
Universiti Malaysia Pahang
Al-Sultan Abdullah
Pahang, Malaysia
zulfahmi@ump.edu.my

Abstract—Social media sites like Instagram, Twitter, and Facebook have become indispensable parts of the daily routine. These social media sites are powerful instruments for spreading the news, photographs, and other sorts of information. However, since the emergence of the COVID-19 pandemic in December 2019, many articles and headlines concerning the COVID-19 epidemic have surfaced on social media. Social media is frequently used to disseminate fraudulent material or information. This disinformation may confuse consumers, perhaps causing worry. It is hard to counter the widespread dissemination of disinformation. As a result, it is critical to develop a model for recognizing fake news in the news stream. The dataset, which would be a synthesis of COVID-19-related news from numerous social media and news sources, is utilized for categorization in this work. Markers are retrieved from unstructured textual data gathered from a variety of sources. Then, to eliminate the computational burden of analyzing all of the features in the dataset, feature selection is done. Finally, to categorize the COVID -19 related dataset, multiple cutting-edge machine-learning algorithms were trained. Support Vector Machine (SVM), Naïve Bayes (NB), and Decision Tree (DT) are the machine learning models presented. Finally, numerous measures are used to evaluate these algorithms such as accuracy, precision, recall, and F1 score. The Decision Tress algorithm reported the highest accuracy of 100% compared to the Support Vector Machine 98.7% and Naïve Bayes 96.3%.

Keywords—fake news, social media, machine learning

I. INTRODUCTION

Social media platforms like Facebook, Twitter, Instagram, and others have risen in the twenty-first century, allowing information to travel swiftly. Users on social media can publish whatever they wish, regardless of the provenance and credibility of the published material, posing problems to information dependability assurance. Each social media user

could have as many accounts as they wish. With the COVID-19 epidemic, millions of posts or news are being sent out every day, with some detrimental repercussions for people and society. For example, the propagation of false information concerning COVID-19 patient data or symptoms that have yet to be confirmed. People might quickly become panicked as a result of social media's fake news and disinformation. The phrases fake news and disinformation are closely related and are sometimes used interchangeably. An automated false news detection system is required, which will rely on human annotation, machine learning, and natural language processing [1]. Since its emergence in December 2019, there have been several articles and media articles on the COVID-19 outbreak on the internet, traditional print, and digital media. These sources provide data from both credible and untrustworthy clinical sources. Furthermore, news from these outlets spreads swiftly. Spreading a piece of false information might prompt nervousness, undesirable openness to clinical cures, stunts for computerized showcasing, and may prompt destructive elements. As a result, a model for detecting fake news in the news stream is critical. Fake news identification is a new technique, and there is still more work to be done to attain specific levels of accuracy, particularly in news and information management [1].

In this day and age, machine learning (ML) algorithms and natural language processing (NLP) algorithms are critical tools for identifying fake news. For several advanced works in computing science, NLP has grown into a fundamental pillar for comprehending and applying the ideas of text summarization, text categorization, sentiment analysis, and opinion mining