1. Introduction

On March 11, 2020, the WHO (World Health Organization) determined that a new pandemic had occurred in Wuhan, China. This virus is caused by infection with the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), which can affect or infect the respiratory system in humans [1]. The characteristics of people infected with the virus are increasingly diverse, ranging from high fever over 39°C, dry cough, difficulty breathing, shortness of breath, and loss of smell and taste [2]. However, many are unaware that they have been exposed to the coronavirus, so it is very easy to transmit it to each other. The coronavirus has spread in various parts of the world, including Indonesia. The Indonesian government has made multiple efforts to deal with the pandemic, such as limiting community activities, using masks when leaving the house, washing hands, maintaining distance, and self-quarantining if the COVID virus is detected [3]. The following solution by the government is to accelerate vaccination for all Indonesians. The information is disseminated to all news channels and social media, making it easier for the general public to get the news.

According to Silverman, hoax news is a series of pieces of information that are misled but are 'sold' as accurate news [4]. The amount of information from news sites and social media that spread hoax news about the side effects of the COVID-19 vaccine has made people nervous about vaccination. The spread of hoax news is very dangerous for the community, especially now that it is easy to find information on social media platforms such as Facebook, Instagram, Twitter, YouTube, TikTok, and WhatsApp. Hoax news is usually made to influence views in the political field or just be news [5]. This makes the government unable to contain the COVID virus in Indonesia because many people do not believe in spreading fake news. One of the efforts that can be made is to detect fake news to see whether the news is true or false.

Several studies related to the detection of hoax news have been carried out previously, such as research conducted by Awalina et al. [6], which detected hoax news using Deep Learning Network Transformers with the BERT method and Transformer Network as the basis for reference and comparison between other methods using CNN, Hybrid CNN-BiLSTM, and BiLSTM. The results obtained by the BERT method show that the accuracy value reaches 90%, supported by the f1-score, which is 90% greater than the other methods. Santoso et al. [7], conducted research to classify fake and genuine Twitter accounts using the Naïve Bayes method has an accuracy value of 80%, using a 5% training set. Another study was conducted by Panjaitan et al. [8], using Random Forest, SVM, Gradient Boosting, Naïve Bayes, and Logistic Regression methods. The result with a good accuracy value is Random Forest, with an accuracy value of 96.05%. Das et al. [9] researched the hoax news detection system using the ensemble method, which yielded an f1-score accuracy value of 0.98%. Research by Gowthami et al. [10], has identified fake news by using a comparison of the SVM and Random Forest algorithm methods; the results obtained show that the Random Forest algorithm can predict by producing a value of 98% compared to the SVM algorithm, which produces an accuracy value of 70%. Therefore, the Random Forest algorithm has better and more efficient performance.

From several series of studies that have been carried out, we have not found any topics regarding comparisons regarding the detection of hoax news related to the news of the COVID-19 vaccine. In addition, the study conducted a classification using only the title. This has prompted us to conduct research related to the classification of hoax news regarding the COVID-19 vaccine by using data in the form of titles and news content to find out whether the news is hoax news or not.

In this study, we classify those related to the spread of hoax news about vaccination in the community. The dataset used comes from a government website specifically related to COVID-19 news, namely https://covid19.go.id/. The data is taken from the period 01 January 2020 to 28 February 2022 [11]. In this case, we use two test scenarios, namely a data scenario and a classification scenario. In the data scenario, the data is used using three data types, namely data that is different only in the "Title", the "News Content" section, and the combination of "Title + News Content". Then, in the classification scenario, we conducted an algorithm comparison experiment using the basic ensemble approach, namely Random Forest and AdaBoost. Forest Random Selection is known for its simplicity and effectiveness in generating good classifications [12], and the choice of the AdaBoost algorithm is because it is suitable for unbalanced data [13]. The purpose of this experiment is to find out which ensemble method effective in detecting is good to use in detecting hoaxes related to COVID-19 vaccine data.