CHAPTER I
INTRODUCTION

As the years passed, information technology grew exponentially; therefore, humans are easier to obtain and even create information (Levitin, 2015). With just one click on a smartphone, humans can spread information quickly. In 2020, internet users in Indonesia reached 175.4 million people out of 272.1 million people in Indonesia. More interestingly, the number of connected smartphones reached 338.2 units or 124% compared to the population in Indonesia. Therefore, almost the average Indonesian population has more than one smartphone, and the possibility of information spread is also increasing (Kemp, 2020). The number of internet users contributing to spreading information will cause information overload. The abundance of this information results in a decrease in the quality of information (Menczer & Hills, 2020).

Irresponsible individuals utilize the ease of internet access to spread false information. KOMINFO found many sites containing false information in 2017, and about 800,000 sites circulated in the Indonesian language (Yuliani, 2017). Fake news spreads significantly further, faster, deeper, and more widely than true news across all categories of information. Its effects are more pronounced for fake political news than other fake news. It was found that people tend to share new information and that new information contains more fake news than true news. And the spread of fake news is undoubtedly the work of humans who cannot be controlled and result in fake stories that inspire fear, disgust, and surprise in reply, while the true story inspires anticipation, sadness, joy, and trust (Vosoughi et al., 2018).

Suppose false information is left to spread widely in the community. In that case, it will give rise to manipulation of public opinion, which leads to feelings of anger, suspicion, anxiety, and even depression by distorting our thinking (Erdelyi, 2020). In the economic field, false information can affect the behavior of investors. The existence of fake news can reduce investor confidence, resulting in a decrease in market value and losses (McElhaney, 2019).

Measuring the spread of false and true news in Indonesia can be done by modeling a cascade of information and looking at the mechanism using the Social Network Analysis and Susceptible-Infected models. Both models are suitable for investigating the spread of information from sources to larger audiences. The fake news can spread to 0.6414 more fractions and occurs 4.6 faster than true news. It is proven by discovering a cascade mechanism of false information by looking at the behavior of the Indonesian people (Alamsyah & Sonia, 2021).

Previous studies have used legacy methods that have been good at detecting hoax news. The Bidirectional Encoder Representation from Transformer (BERT) is a method for utilizing deep two-way representations of unlabeled text by jointly conditioning left and proper contexts in all layers (Devlin et al., 2019). BERT has widely applied to several studies, such as for classifying news articles as fake or real (Aggarwal et al., 2020), correcting spelling errors (Tan et al., 2020), and text-based emotion recognition (Adoma et al., 2020). IndoBERT is a model following the BERT Base configuration (Koto et al., 2020). The number of datasets used on IndoBERT is more than any other method limiting the number of datasets (Wilie et al., 2020). IndoBERT detects hoax news related to Covid-19 by using transfer learning from pre-trained transformer models such as Fine-tuned original pre-trained BERT, multilingual pre-trained mBERT, and monolingual pre-trained IndoBERT. The study results mentioned that fine-tuned mBERT base-cased has an accuracy of 97.93%; thus, the models perform better than others (Suadaa et al., 2012).

The implementation of IndoBERT for hoax detection is not much. For previous research, IndoBERT research in detecting hoax news related to Covid-19. Our contribution is to detect hoax news on topics such as economic and business, national news, stock market, politics, etc., using IndoBERT. Furthermore, we compare the IndoBERT result with other methods such as Support Vector Machine and Naïve Bayes to determine which methods best detect hoax news in terms of prediction accuracy.