I. Introduction

In today's digital era, the spread of information has become increasingly rapid and easy to reach. With the internet, people can access various information through various sources obtained through news portals, online articles, and social media [1]. In January 2024, it was recorded that internet users in Indonesia reached 183.5 million with a percentage of 66.5% and there were 353.3 million active cellular connections, this figure is equivalent to 126.8% of the total population. Based on these data, it is proven that there are groups of people who have more than one smartphone and this can lead to an increasing spread of information [2]. Abundant information can result in a decrease in the quality of information [3]. The public is increasingly vulnerable to fake news with the widespread use of information accompanied by decreasing quality [4].

Fake news can be defined as false or misleading content presented in the form of news in oral, written, printed, electronic, and digital communication formats [5]. In Indonesia itself, based on a survey by the Indonesian Telematics Society (Mastel) in 2019, 21.80% of respondents admitted to having difficulty checking the truth of the news and only 16.20% could immediately identify that the news they received was fake news [6]. Added to the results of a survey by the Indonesian Internet Service Providers Association (APJII) in the period December 2023 - January 2024, it was stated that the content category that was most vulnerable to fake news was political content with a percentage of 24.7% [7]. Unwittingly, this can cause social unrest because it is often used as a weapon by irresponsible parties to commit fraud, defamation, cyberbullying, and others [8]. There are several websites that are well-known for routinely identifying and publishing fake Indonesian news such as TurnBackHoax, CekFakta, and Hoax or Not. However, the identification process on these platforms is generally done manually from individual reports. This makes the system less effective in handling the large number of fake news spread across various media every day. Thus, system automation is needed to detect fake news and control its very massive spread.

Leaving this problem, we conducted a research to detect fake and real Indonesian news scraped from the TurnBackHoax and CekFakta sites. Various methods have been used to create a fake news detection system, in this study we used the Natural Language Inference (NLI) approach. The basic concept of NLI is a system that imitates the methods used by news experts to identify fake news. New news is called fake if it contradicts verified facts. Conversely, the news is classified as true if the news matches the confirmed news. This method is considered innovative because the inference method has been applied for the first time in identifying fake news and data sources outside the content and context of the news have been used [9]. NLI concludes the relationship between two text narratives, namely premises and hypotheses. Premises and hypotheses have a relationship called true (entailment), neutral, and false (contradiction). This field is a developing field in natural language processing (NLP) [10].

In the development of NLP in Indonesia, language and cultural issues are one of the biggest challenges faced. Indonesian language has a distinct morphological, syntactic, and semantic structure that calls for a particular method of building NLP technologies. In addition, the diversity of languages and dialects in various regions in Indonesia also adds to the level of complexity in the NLP research and development process [11]. To address this issue, the NLI approach is collaborated with the deep learning model Indonesian Bidirectional Encoder Representations from Transformers (IndoBERT), which is a monolingual model specifically trained for Indonesian [12]. So the combination of NLI and IndoBERT is expected to be an innovation that can understand the context in fake news in Indonesian. The news used is also specific in political news topics because based on data [7], political content is the content that is most prone to fake news in Indonesia. Thus, compared to previous studies [9], [13], this project aims to improve the use of the NLI approach to detect fake news in Indonesian language, as most Indonesians still have difficulty identifying real and fake news based on previously presented data.