

1. INTRODUCTION

Daily activities are inseparable from the two-way communication that has occurred in society from the past to the present. Of the various kinds of languages in the world, one of the most popular languages is Indonesian. Indonesian itself is the official language of the country of Indonesia. Derived from the Malay language, Indonesian as one of the regional languages in the archipelago, then developed into an intermediary language 'lingua franca' between communities [1]. Indonesian is one of the most widely spoken languages in the world, spoken by around 250 million speakers [2]. The use of the Indonesian language is spread through many media with formal and informal writing, one of which is the spread through the news media. Along with the development of the times, the use of slang or foreign words is increasingly being used in many mass media. News is also one of the important things to spread new information in the world.

In that sense, the news itself means information that can be presented through print, broadcast, internet, or even word of mouth, so the news is one of the important things for the wider community [3]. News that is developing at this time does not always have a formal language, but also informal. So many news articles are created and shared online every day, making it difficult for users to find the news they are interested in [4]. So in news writing, foreign languages and slang words are always a consideration for adding stories to the news to make it look more interesting. The choice of words used is also a consideration so that the article is well organized in terms of language. Along with the times, various kinds of languages can be processed using the help of computers for many purposes such as question-and-answer applications.

Even though Indonesian is one of the most widely used languages, it still has limited resources for Natural Language Processing (NLP) needs [5]. NLP itself means text processing which is processed into machine language. One of the basic techniques in NLP is Part-of-speech (POS) tagging because it is used in most NLP, such as applications for sentiment analysis, question-answering tools, word sense disambiguation, etc. [6]. In that sense, POS tagging is a process in word class labeling [7]. POS tagging works by automatically labeling word classes in a sentence [8]. POS tagging has several approaches commonly used including rule-based, probabilistic, and transformational-based. Rule-based POS tagger assigns a word class label to a word based on manually created linguistic rules. For a probabilistic approach by finding the highest probability of a class of words based on the context of the sentence [9]. Meanwhile, the transformational-based approach is a combination of rule-based and probabilistic approaches. In addition, one approach to POS tagging is using a neural network. Bidirectional Long-Short Term Memory (LSTM) is one of the types of Recurrent Neural Network (RNN) which has the ability to combine contextual information from opposite LSTM two-way input lengths and has been proven as an effective model for performing sequential labeling [10].

Lots of research on POS tagging has been carried out in various languages, one of which is Indonesian, which has been carried out by several researchers. One of them was done by Handrata et al. [11] by conducting research on POS tagging using the bidirectional LSTM method in Indonesian using a corpus from one of the universities, the University of Indonesia [11]. The corpus is used by Handrata et al. [11] as training data and test data. Other research on POS tagging for Indonesian was also carried out by other researchers using various methods. Cahyani et al. [7] did POS tagging in Indonesian using the Hidden Markov Model method. Pisceldo et al. [9] evaluated the two probabilistic models, Maximum Entropy and CRF (Conditional Random Fields). Rashel et al. [12] build an Indonesian language POS tagger using a rule-based approach. Local languages are also used, including Javanese [13] and Balinese [14].

Based on several considerations, the research conducted was to determine the performance of the POS tagging model built using the bidirectional LSTM model on Indonesian language news texts. By making several modifications to the Indonesian language corpus which is used to adapt the current news structure. Some of them are words that rarely have errors because the writing system has been adjusted along with the times. Because of that the research focuses on modifications to the word class "X" and "FW" as the main goal. In implementing the model for the tested news text, the selected text is a news text related to Telkom University. One of the reasons for choosing Telkom University's news is that it is one of the universities that are active in terms of reporting, one of which is the announcement related to scholarships.