

Abstract—News is a tool used to disseminate information through various media, one of which is the internet. Various kinds of news articles have words that are not recognized in the dictionary such as slang words and have foreign words that do not exist in the corpus. How can a POS tagging model built on the corpus be able to handle word class labeling in Indonesian news. The research was conducted to check the results of POS tagging on a collection of news about Telkom University which was selected manually. By using the bidirectional LSTM model, three test scenarios were attempted to improve the performance of the built model, the test scenarios were applying the best padding for the corpus, comparing the performance results of the modified corpus model with the original corpus model, and determining the dimensions of the Word2vec vector. Then the selected model from each corpus is implemented on the news that has been labeled manually. One of the best scenario tests is obtained by modifying the corpus by removing double words in the word class "X" and changing some of the word classes "X" which are more likely to be foreign words so that they are changed to the word class "FW". The best performance results in the implementation of news about Telkom University using the bidirectional LSTM model which was built based on the modified corpus get accuracy values of 92.74%, precision of 92.85%, recall of 92.74%, and F1-score 92.48%.

Keywords: POS tagging; bidirectional LSTM; Indonesian; news