1. Introduction

Part-of-speech Tagging (POS Tagging) is an automatic word tagging in which is the part is considered appropriate from a tag will be given to a word based on words that have been tagged before [1]. POS Tagging is also called as grammatical tagging or disambiguation word categories [2]. The use of POS tagging is important because it is used in several Natural Processing Language (NLP) applications such as word disambiguation, sentence parsing, questions answering, and translation machine [3,4]. Tagging words manually is a task that can save a lot of time, because you need to have special accuracy and skills in its application [5]. It is hoped that with POS tagging giving a tag will save a lot of time. There are 2 types of labeling rules in the POS tagging method, namely Rule-Based-Tagging and Stochastic Tagging. Rule-Based Tagging is the process of labeling according to dictionaries or rules that have been determined from training data sets. This type is also commonly used to solve problems in cases of unknown words or ambiguity in words, morphological and semantic information [1]. Then the second type is Stochastic Tagging which is done by using the corpus dataset as a data train to later determine the probability of a class of words. The use of this rule is also to determine the best tag from the model which has predicted which word class is considered appropriate for the word to be tagged [1].

Indonesia is rich of ethnics settled in different regions. Each of them has different local language for communication in their regions, meanwhile the most common language for communication is Indonesian. Indonesian is the national language used to communicate officially throughout Indonesia [6]. Indonesian is also used as journalism language either in electronic or digital media. As other languages Indonesian language has several grammatical categories, such as verbs, nouns, adjectives, adverbs, and so on which is usually found in online news articles widely available on social networks [3].

According to KBBI News is an incident report about what the organization has recently learned about important or interesting things [7]. News articles have distinctive characteristics like a being actual, factual, and interesting. The use of Indonesian words in the news is formal. The role of POS Tagging in Indonesian news is considered to be very helpful because with POS tagging we can decipher words to get information quite easily and required shorter time than manual labeling [5].

The research that will be carried out is regarding to efficiency of POS taggers in Indonesian news. As the previous studies that have been carried out by several researchers with various methods and resulted the quite good methods accuracy. One of them was the research conducted by Ritu Banga and Pulkit Mehndiratta on a comparison of the five tagger methods including Perceptron Taggers, TnT, Conditional Random Fields Tagger, Brill Tagger, and Classifier BasedPOS Tagger (CPOS). Perceptron taggers have the highest accuracy of 88.7% [1]. However, the dataset used in this research comes from the Twitter API and is in English. There is a difference with the research that will be conducted by text from Twitter, where the text is much shorter than news articles. As for other research regarding POS Tagging in Indonesian with each POS Tagger method [2,8,9,10,11,12].

It has been described above that in this research, the authors plan to find the most efficient POS tagger for Indonesian news by using the Conditional Random Fields (CRF) and Hidden Markov Models (HMM) taggers. Conditional Random Fields (CRF) is a probabilistic calculation method for determining the order of labels to be assigned to the sequence of observations [13]. And the Hidden Markov Model (HMM) method is a process of providing tags that can classify one series or sequence of tags for each word in one sentence. HMM also uses a probabilistic technique, in which the resulting sequence of two stochastic coefficients, one of the processes, is unobservable (hidden state) [9,13].

The research conducted will use the HMM and CRF Tagger methods because both methods use probability techniques to determine a tag in a word. Then the dataset used is a dataset from online news in Indonesian. The corpus used is the corpus that comes from *Indonesia Manually Tagged* Corpus which contains the text of a news item that has been tagged manually with a total of 23 tag set [15]. The expected results in this research are that we can find out which POS taggers method between CRF and HMM taggers has the best efficiency. The method that is considered to have the best efficiency is the method that has a high accuracy value and a faster training time on the dataset.