

## ABSTRACT

The news is a media that can spread the information to the audience. News media contain source texts that provide extensive information and extraction systems for rich text, one of which is the extraction of information. The extraction of information in the news can be performed to extract a sentence of a statement. Research on the extraction of sentences of quotations ever done by using linguistic methods, system rules (rule-based) as well as supervised learning. However, the research is limited to specific language and has not arrived to extract names of figures from the statement makers. A newsreader can easily find a response from a character regarding the statement raised on the news by extracting the statement sentence. The process of extracting sentence statement can be done by computer and by an expert. The process goes very slow if done by humans even though yields high accuracy. Humans have logic, reasoning, and feelings to differentiate the types of sentences. This research use approach to machine learning algorithm with Support Vector Machine (SVM), a Naive Bayes and Logistic Regression as the comparison algorithm. Machine learning approach used to identify the type of sentence. As for the entity, resource use recognizes the approach to Named Entity Recognition (NER). Of the proposed approach, the system introduction sentence quote gets value the accuracy of 0.843, 0.844 of precision value, and the recall of 0.839. The accuracy results can extract sentence quote from Indonesian Language that generates entity statements and the maker of the statement.

**Keywords :** Information Extraction, Named Entity Recognition, Machine Learning, Support Vector Machine, Naive Bayes, Logistic Regression

## INTISARI

Berita merupakan sebuah kabar yang dapat menginformasikan suatu peristiwa dan dapat dikabarkan kepada khalayak ramai melalui media. Media berita mengandung sumber teks yang menyediakan informasi yang luas dan kaya untuk sistem ekstraksi teks, salah satunya adalah ekstraksi informasi. Salah satu ekstraksi informasi dalam berita yaitu mengekstrak kalimat pernyataan tokoh dalam berita. Penelitian tentang ekstraksi kalimat pernyataan pernah dilakukan dengan menggunakan beberapa metode seperti linguistik, sistem aturan (*rule-based*) hingga *supervised learning*. Namun penelitian tersebut terbatas pada bahasa yang digunakan dalam penelitiannya dan belum sampai pada ekstraksi nama tokoh dari pembuat pernyataan. Dengan mengekstrak kalimat pernyataan tokoh dalam berita, pembaca berita dapat dengan mudah menemukan tanggapan dari seorang tokoh mengenai pernyataan yang diangkat pada berita. Proses ekstraksi kalimat pernyataan dapat dilakukan oleh komputer dan oleh seorang ahli. Namun proses ekstraksi kalimat pernyataan akan berjalan dengan sangat lambat jika dilakukan oleh manusia meskipun akan menghasilkan akurasi yang tinggi. Hal ini dikarenakan manusia memiliki logika, penalaran serta perasaan untuk membedakan jenis-jenis kalimat. Penelitian ini menggunakan pendekatan pembelajaran mesin dengan algoritme *Support Vector Machine* (SVM), *Naive Bayes* dan *Logistic Regression* sebagai algoritme pembandingan. Pendekatan pembelajaran mesin digunakan untuk mengenali jenis kalimat. Sedangkan untuk mengenali entitas narasumber menggunakan pendekatan Named Entity Recognition (NER). Dari pendekatan yang diusulkan, sistem pengenalan kalimat kutipan mendapatkan nilai akurasi sebesar 0,843, nilai presisi sebesar 0,844 dan nilai *recall* sebesar 0,839. Hasil akurasi tersebut dapat mengekstraksi kalimat kutipan dari Bahasa Indonesia yang menghasilkan entitas pembuat pernyataan dan pernyataannya.

**Kata kunci** – Information Extraction, Named Entity Recognition, Machine Learning, Support Vector Machine, Naive Bayes, Logistic Regression