

## ABSTRAK

Fenomena ujaran kebencian pada saat ini menjadi fenomena yang sering muncul di media sosial Indonesia. Fenomena ini muncul paling banyak pada twitter dan facebook. Namun, tidak menutup kemungkinan platform media sosial lain juga terdapat fenomena ini.

Ujaran kebencian pada awalnya muncul pada saat musim pemilu sekitar tahun 2014. Namun, seiring berkembangnya jaman ujaran kebencian kerap muncul dan dapat memecah belah persatuan bangsa. Pengguna media sosial seringkali tidak merasa bahwa apa yang dituliskan pada platform media sosial merupakan ujaran kebencian. Akibatnya, jika ada pelapor, penulis ujaran kebencian tersebut bisa saja terkena tindak pidana. Oleh sebab itu, penulis ingin meneliti mengenai pengenalan kalimat ujaran kebencian menggunakan metode *Convolutional Long Short Term Memory*. Harapannya dalam penelitian ini dapat dibedakan mana kalimat yang merupakan ujaran kebencian serta akurasinya.

Pada penelitian ini ada beberapa tahapan yang dilalui diantaranya adalah *preprocessing*, *feature extraction*, dan *training model*. Tahapan *preprocessing* terdiri dari beberapa proses seperti *case folding*, *remove punctuation*, *remove unnecessary char*, *remove emoticon*, *stemming*, *normalization*, *stopword removal*, hingga *tokenization*. Setelah itu pada tahap *feature extraction* dilakukan dengan metode *fasttext* untuk mendapatkan bobot masing – masing kata. Setelah mendapatkan bobot masing – masing kata, selanjutnya dilakukan pelatihan model dengan *Convolutional Long Short Term Memory*. Adapaun metode ini mendapatkan akurasi terbaik yaitu 83,43%

**Kata kunci** : Ujaran kebencian, *fasttext*, *convolutional long short term memory*.

## ABSTRACT

Hate speech phenomenon currently often appears in Indonesia social media. This phenomenon often appears in twitter and facebook. But, it is possible another social media also have this phenomenon.

Hate speech initially appear during of election season around 2014. However, along with the evolve of the era, hate speech more often appears and can divide national unity. Social media users, often not realize that what they write on social media is hate speech. As a result, if there is a reporter, the author of the hate speech may be subject of criminal act. Therefore, the author wants to examine the hate speech recognition using convolutional long short term memory method. It is hoped that in this study it can be distinguished which sentences are hate speech and their accuracy.

In this research, there were several stages that were passed, including : preprocessing, feature extraction, and model training. The preprocessing stage consists of several processes such as case folding, remove punctuation, remove unnecessary char, remove emoticon, stemming, normalization, stopword removal, and tokenization. After that, the feature extraction stage is carried out using the fasttext method to get the weight of each word. After getting the weight of each word, then the model training is carried out with Convolutional Long Short Term Memory. As for this method, the best accuracy is 83,43%

**Keywords:** hate speech, fasttext, convolutional long short term memory.