

ABSTRAK

Pengenalan Entitas Bernama pada Tweet tentang Bencana Menggunakan *Conditional Random Field*

Oleh

TIO RAHADITYA LUTHFITAMA

15/378076/PA/16551

Sosial media merupakan salah satu media yang dapat digunakan untuk mencari informasi yang sedang terjadi di dunia. Salah satu jenis informasi yang dapat didapatkan adalah bencana yang sedang terjadi. Sayangnya, media sosial tidak memiliki struktur kalimat yang baku dalam penulisannya. Sehingga, akan dibutuhkan waktu yang cukup lama untuk mengenali entitas lokasi, bencana, dan tanggal terjadinya.

Dengan berkembangnya pendekatan *Natural Language Processing (NLP)* dan didukung teknologi yang ada, maka dibuat sebuah model yang mampu melakukan pengenalan entitas pada kumpulan *tweet* yang terindikasi sebagai *tweet* tentang bencana. Sedangkan *tweet* yang tidak termasuk dalam *tweet* tentang bencana, tidak digunakan dalam *dataset*.

Model ini menggunakan algoritma *Conditional Random Field (CRF)* dan dibandingkan dengan algoritma *Multinomial Naïve Bayes (MNB)*. Tiga entitas yang digunakan sebagai komponennya yaitu DIS (bencana apa yang sedang terjadi), LOC (dimana lokasi bencana itu terjadi), dan DATE (tanggal kejadian bencana tersebut). Pada penelitian ini, hasil dari evaluasi model CRF lebih baik dibandingkan dengan MNB. Nilai Evaluasi CRF yang dihasilkan adalah 0.93 untuk *precision*, 0.89 untuk *recall* dan 0.91 untuk *f1-score*.

Kata kunci: Pemrosesan Bahasa Alami (NLP), Pengenalan Entitas Bernama (NER), *Conditional Random Field (CRF)*.

ABSTRACT

NAMED ENTITY RECOGNITION FROM TWEET ABOUT DISASTER USING CONDITIONAL RANDOM FIELD

By

TIO RAHADITYA LUTHFITAMA

15/378076/PA/16551

Social media is one of the media that can be used to find information that is happening in the world. One type of information that can be obtained is a disaster that is happening. Unfortunately, social media does not have a standard sentence structure in writing. So, it will take a long time to recognize the location, disaster, and date of occurrence.

With the development of the Natural Language Processing (NLP) and supported by existing technology, a model is able to recognize entities to a tweets that indicated as tweets about disaster. Tweets that not about disasters are not used in the dataset.

This model uses the Conditional Random Field (CRF) algorithm and compared with the Multinomial Naïve Bayes (MNB) algorithm. Three entities that are used as components are DIS (what disaster is happened), LOC (where the disaster happened), and DATE (the date of the disaster happened). In this study, the results of the evaluation of the CRF model were better than MNB. The Evaluation Value of CRF produced is 0.93 for precision, 0.89 for recall and 0.91 for f1-score.

Keywords: Natural Language Processing (NLP), Named Entities Recognition (NER), Conditional Random Field (CRF).