

## INTISARI

# KLASIFIKASI TWEET DAN PENGENALAN ENTITAS BERNAMA PADA TWEET BENCANA DENGAN SUPPORT VECTOR MACHINE

Oleh

Rizki Dermawan

11/316702/PA/13831

Indonesia termasuk negara yang sering terkena bencana. Ketika terjadi sebuah bencana pada suatu daerah, banyak pengguna media sosial, khususnya pengguna Twitter, memberikan informasi terkait bencana. Informasi yang diberikan dapat berupa lokasi terjadinya suatu bencana, kondisi dari suatu daerah, ataupun kebutuhan dari masyarakat pada daerah bencana. Informasi tersebut dapat dimanfaatkan untuk pemetaan kejadian bencana atau pemetaan kebutuhan masyarakat ketika terjadi suatu bencana. Akan tetapi, media sosial memiliki kredibilitas yang rendah sebagai penyedia informasi. Selain itu, tidak terstrukturanya data pada media sosial membuat pendataan akan lokasi, kondisi, dan kebutuhan masyarakat menjadi sulit.

Pada penelitian ini, dibangun sistem yang dapat mengklasifikasi apakah suatu tweet terkait bencana atau tidak. Jika tweet tersebut terkait dengan bencana, dilakukan pengenalan entitas terhadap tweet tersebut sehingga dapat diketahui lokasi yang dibicarakan, kondisi yang terjadi, dan kebutuhan masyarakat pada daerah bencana. Sistem pengklasifikasian tweet dan sistem pengenalan entitas bernama dibangun dengan *supervised learning*. Algoritma *supervised learning* yang digunakan adalah *support vector machine*.

Pada pengklasifikasian tweet, dicoba dua metode pembobotan yakni  $tf$  dan  $tf - idf$ . Pada penelitian ini,  $tf - idf$  unggul dari  $tf$  dengan rata-rata akurasi 78%. Sedangkan pada sistem pengenalan entitas bernama, dibandingkan dua metode segmentasi entitas yakni BIO dan BILOU. Pada penelitian ini metode segmentasi BIO lebih unggul dibanding BILOU dengan  $F_1Score$  86%.

Kata-kata kunci : support vector machine, pengenalan entitas bernama, sosial media, bencana.

## **ABSTRACT**

# **TWEET CLASSIFICATION AND NAMED ENTITY RECOGNITION IN DISASTER TWEET WITH SUPPORT VECTOR MACHINE**

By

Rizki Dermawan

11/316702/PA/13831

Indonesia is a country that is often affected by disaster. When there is a disaster in an area, many users of social media, especially Twitter users, provide information related to the disaster. The information provided can be the location of a disaster, the conditions of an area, or the needs of the people in the disaster area. Such information can be used for mapping the event of disaster or mapping needs of the community in the event of a disaster. However, social media has low credibility as an information provider. Moreover, unstructured data on social media will make data collection of location, condition, and needs of the community becomes difficult.

In this research, system that can classify whether a tweet related to the disaster or not has been built. If the tweet is related to the disaster, the named entity is recognized so the location of disaster, a condition of the area, and the need of community is known. Tweet classification system and named entity recognition system built by supervised learning. Supervised learning algorithm that is used is a support vector machine.

In the classification of the tweet, two methods of weighting  $tf$  and  $tf - idf$  has been tried. In this study,  $tf - idf$  is better than  $tf$  with an average accuracy of 78%. In the named entity recognition system, two methods of segmentation entities BIO and BILOU is compared. In this study, BIO segmentation method is better than BILOU with  $F_1 Score$  86%.

**Keywords :** support vector machine, named entity recognition, social media, disaster.