

INTISARI

IDENTIFIKASI TWEET BANJIR DI JAKARTA PADA DATA TWITTER MENGGUNAKAN ALGORITMA MULTINOMIAL NAIVE BAYES DAN SUPPORT VECTOR MACHINE

Oleh

Muchammad Ersan Ramadhan
12/334666/PA/14899

Twitter merupakan salah satu media sosial yang umum digunakan oleh masyarakat. Masyarakat menuliskan informasi tentang banjir melalui media *twitter* dalam bentuk *tweet*. Beberapa *tweet* yang dituliskan tidak menunjukkan secara nyata bencana banjir yang sedang terjadi. Terdapat *tweet* tentang banjir menunjukkan *tweet* yang bukan kategori bencana, sebagai contoh banjir hadiah. Padahal informasi bencana banjir secara nyata dan cepat dibutuhkan untuk mengantisipasi banjir yang berkelanjutan. Oleh karena itu diperlukan sistem yang secara otomatis mengidentifikasi secara otomatis *tweet* banjir termasuk dalam kategori bencana atau banjir dalam arti lain.

Pada penelitian ini proses identifikasi *tweet* banjir menggunakan algoritma *Multinomial Naïve Bayes* dan *Support Vector Machine*. Data yang digunakan merupakan *tweet* yang mengandung kata banjir yang berlokasi di DKI Jakarta. Data *tweet* banjir diidentifikasi menjadi 2 kategori yaitu bencana dan bukan bencana. Jumlah *tweet* sebanyak 7789 sebagai data *training* dan 2327 sebagai data *testing* yang diambil pada 12 November 2014 – 30 Januari 2015.

Pengujian dilakukan dengan membandingkan algoritma *Multinomial Naïve Bayes* dan *Support Vector Machine*. Pengujian juga dilakukan dengan menggunakan metode *K-fold cross validation* dengan jumlah iterasi 10 kali. Dari hasil pengujian didapatkan nilai akurasi terbesar dengan menggunakan algoritma *Support Vector Machine* dengan rata-rata akurasi 78.9%

Kata Kunci : Banjir, *Multinomial Naïve Bayes*, *Support Vector Machine*.

ABSTRACT

FLOOD TWEETS IDENTIFICATION IN JAKARTA ON TWITTER DATA USING MULTINOMIAL NAÏVE BAYES AND SUPPORT VECTOR MACHINE ALGORITHMS

Oleh

Muchammad Ersan Ramadhan

12/334666/PA/14899

Twitter is one of the most common media sosial used by the society which is effective and fast in reporting the current situation, such as flood in DKI Jakarta. Flood tweets will be automatically identified so that it can provide information whether it is flooded or not flooded in certain area. The use of Twitter as a source of information regarding flood is potentially helpful to anticipate the occurrence of long-standing flood. That is why a system is needed to automatically categorized a flood containing tweets as flood as an event or other meaning

The identification process of flood tweets used in this research is Multinomial Naïve Bayes and Support Vector Machine algorithms. The data used are the tweets containing the word “banjir” located in DKI Jakarta. The data is identified into 3 categories, namely flooded, not flooded and unknown. Total tweets of 7789 is used as data training and 2327 tweets is used as data testing taken from November 12, 2014 - January 25, 2015.

The test is done by comparing Multinomial Naïve Bayes dan Support Vector Machine algorithms. The test is also done by using K-Fold Cross Validation method with 10 times iteration. The test result shows that the greatest accuracy obtained by using Support Vector Machine algorithms with average accuracy of 78.9%.

Keyword : Banjir, *Multinomial Naïve Bayes*, *Support Vector Machine*.