

ABSTRAK

KLASTERISASI INFORMASI BANJIR DI TWITTER MENGGUNAKAN METODE *FREQUENT TERM-BASED CLUSTERING* (FTC) DAN EKSTRAKSI *NAMED ENTITY RECOGNITION* MENGGUNAKAN INNER PADA BADAN NASIONAL PENANGGULANGAN BENCANA (BNPB)

Oleh : Zahra Permata Jodea (1511502500)

BNPB merupakan lembaga pemerintah yang bertugas untuk menanggulangi bencana alam, dalam penanggulangnya BNPB belum menggunakan data Twitter sebagai sarana menanggulangi bencana alam secara maksimal di era globalisasi ini, pengguna media sosial memberikan beberapa informasi terkait terjadinya bencana alam seperti lokasi terjadinya bencana alam. Namun, informasi yang berasal dari media sosial memiliki struktur yang tidak baku serta kredibilitas yang rendah sebagai penyedia informasi. Data media sosial yang tidak terstruktur berdampak pada informasi terkait lokasi bencana alam menjadi sulit diidentifikasi. Oleh karena itu, harus digunakan suatu metode, salah satunya dengan pemanfaatan *text mining*, yang merupakan proses ataupun kegiatan untuk mengumpulkan data yang berukuran besar kemudian mengekstraksi data tersebut menjadi informasi yang nantinya dapat digunakan. Tujuan dari penelitian ini adalah untuk membantu BNPB dalam mencari data *tweet* bencana alam dengan konsep menggali informasi dan mengelompokkan data *tweet* yang terdapat pada *twitter*, pendekatan *text mining* diharapkan dapat menjawab permasalahan tersebut, dengan menggunakan metode *Clustering* dengan algoritme *Frequent Term Based Clustering*. *Frequent Term Based Clustering* merupakan sebuah algoritme klasterisasi menggunakan metode probabilitas dan statistik mentah untuk mengelompokkan suatu data dan mendapatkan informasi yang baik. Ditambah dengan *InNer*, pengguna dapat mengekstrak informasi lebih banyak dibanding *clustering* biasanya. Hasil identifikasi entitas Lokasi dapat membantu Badan Nasional Penanggulangan Bencana (BNPB) untuk mendapatkan informasi dini terkait bencana alam yang terjadi di Indonesia secara faktual dan aktual dalam melakukan *clustering tweet* atau mengelompokkan data yang berdasarkan kemiripannya. Penggunaan metode *clustering* dapat memudahkan pengguna dalam mendapatkan informasi dari sosial media *twitter*. Di sisi lain, informasi tersebut dapat digunakan sebagai langkah alternatif penyebaran informasi terkait bencana alam yang digunakan oleh masyarakat. Hal ini dibuktikan dengan validasi eksternal dengan pengukuran *purity* bencana alam memiliki tingkat homogen sebesar 0,60.

Kata Kunci: Data Mining, Text Mining , Bencana Alam, *Clustering*, *Frequent Term based Clustering*, *Named Entity Recognition*, INNER

xiii+60 halaman; 27 Gambar; 32 tabel; 1 lampiran