

ABSTRACT

CLUSTERING IMPLEMENTATION USING DBSCAN ALGORITHM TO DETECT NATURAL DISASTER ON MICROBLOGGING TWITTER

By

M. JUAN SHAPUTERA

Social media is one of media or outlet which uses computer that mainly used by people for giving, sharing, or writing information in any form such as document,image, or video to others using internet connection. Nowadays, many social media offered to people for use, with Twitter being one of them. Twitter is a microblogging social media where its user are capable of writing maximum 140 character of short message which can be embedded with image or video that we called as tweet. In this paper, we focused on detecting tweet that contains word relating to natural disaster, and using clustering method supported by DBSCAN algorithm, we try to cluster them and find the overall similarity value of the cluster, while also seeing the difference of value between cluster which uses classification and cluster which doesn't use classification. The testing afterwards giving good result, where mainly cluster that uses classification giving bigger overall similarity value.

Keywords : DBSCAN, cluster, natural disaster, classification, Twitter.

ABSTRAK

IMPLEMENTASI CLUSTERING PADA ALGORITMA DBSCAN UNTUK DETEKSI KEJADIAN BENCANA ALAM PADA MICROBLOGGING TWITTER

Oleh

M. JUAN SHAPUTERA

Media sosial adalah salah satu sarana atau *outlet* menggunakan komputer yang banyak digunakan masyarakat untuk mengirim, men-share, atau menulis informasi dalam berbagai bentuk seperti dokumen, gambar, atau video untuk orang lain. Di era modern ini, banyak media social bermunculan dan digunakan oleh masyarakat umum, salah satunya adalah Twitter. Twitter adalah salah satu media sosial *microblogging* dimana penggunanya dapat menulis *tweet*, yaitu pesan pendek dengan panjang maksimum 140 karakter dan dapat disertai dengan video atau gambar. Dalam penelitian ini, dilakukan pendekripsi *tweet* yang mengandung kata yang berhubungan dengan bencana alam, dan dengan menggunakan metode *clustering* dengan algoritma DBSCAN, dilakukan perbandingan antara overall similarity *cluster tweet* yang menggunakan klasifikasi dan *cluster tweet* yang tidak menggunakan klasifikasi. Pengujian yang dilakukan memberikan hasil yang baik, dimana kebanyakan *cluster* yang menggunakan klasifikasi memiliki *overall similarity value* yang lebih besar dari *cluster* yang tidak menggunakan klasifikasi.

Kata kunci : DBSCAN, cluster, bencana alam, klasifikasi, Twitter.