

Peningkatan kinerja metode k-means dengan pendekatan mini batch untuk masalah pendeteksian topik = Performance improvement of k means method with mini batch approach for topics detection issues

Siti Rofiqoh Fitriyani, author

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=20423373&lokasi=lokal>

Abstrak

Media online menjadi suatu media penting dalam mengakses berbagai macam informasi yang dibutuhkan, misal berita-berita aktual pada portal berita ataupun media sosial. Dinamika perkembangan berita dan informasi pada media online ini semakin tinggi seiring dengan kebutuhan berita teraktual dari setiap peristiwa. Isi informasi yang terkandung dalam berita pun sangat beragam. Oleh karena itu, dibutuhkan metode pendeteksian topik secara otomatis agar mempermudah pengguna mengakses informasi. Salah satu metode otomatis untuk pendeteksian topik adalah metode K-means clustering. Namun, dimensi data yang cukup besar menjadi kendala dalam pendeteksian topik secara otomatis, sehingga dibutuhkan pendekatan lain dalam proses pengolahan data. Pada penelitian ini pendekatan yang digunakan adalah metode mini batch yang hanya menggunakan sekelompok kecil data dalam proses clustering. Akurasi dari metode ini menunjukkan hasil yang cukup baik untuk waktu komputasi yang jauh lebih singkat.

Online media became an important media access to various kinds of information needed, e.g. actual news on news portals or social media. The dynamics of the news and information on online media is higher in line with the needs of most actual news of each event. The information contained in the news are very diverse. Therefore, it takes automatic topic detection methods in order to facilitate users to access the information. One of which method for the automatic topic detection is the K-means clustering. However, the large dimensions of data become obstacles, so it needs a different approach in data processing. In this research, the approximation that used is a mini-batch method that uses only a small group of data in the clustering process. The accuracy of this method showed good results for the computing time is much shorter.