

ABSTRACT

Student Community Service-Community Empowerment Learning has been carried out for many years and has generated a lot of reports on the implementation of activities (Laporan Pelaksanaan Kegiatan, abbreviated as LPK). During this time, these report documents have not been fully utilized to be a source of other information, including the information about regional potency where the community service being held. The information mining about regional potency could be done by one method of text mining, namely document clustering. At the present time, clustering has only been used to group documents. Clustering has not been utilized to mine information. Mining of information using clustering could be done by bringing label up automatically on the cluster resulted.

This study aimed to perform cluster auto-labeling with term of regional potency using documents clustering. But, not every cluster has a label that related to regional potency. Thus, the process of clustering should be adjusted in order to bring up cluster labels that related to these term. Clustering algorithms used in this study were bisecting K-Means, STC, and LINGO. Clusters and cluster labels produced by the three algorithms would be analyzed to determine which one the best algorithm that perform cluster auto-labeling with term of regional potency.

In this study, cluster auto-labeling with term of regional potency could be done by removing the location name of the community service. This step was only good for Bisecting K-Means and LINGO, but it was not working for STC. Based on the evaluation, among these three algorithms, the best algorithm performing cluster auto-labeling with term of regional potency was bisecting k-means. From labels clusters yielded, it could be conclude that regional potency of research data set, which is Sleman Region, was beef, fish, lamb, bark, ginger, guava, cashew, tourism, sengon, craft of baskets (besek), and cocoa.

Keywords — document clustering, cluster auto-labeling, regional potency

INTISARI

Kuliah Kerja Nyata Pembelajaran Pemberdayaan Masyarakat (KKN-PPM) telah dilakukan bertahun-tahun dan menghasilkan banyak sekali laporan pelaksanaan kegiatan (LPK). Selama ini laporan tersebut belum dimanfaatkan secara maksimal untuk dijadikan sumber informasi lain, salah satunya informasi tentang potensi daerah lokasi KKN. Penggalan informasi tentang potensi daerah tersebut dapat menggunakan salah satu metode *text mining*, yaitu *document clustering*. Selama ini *clustering* hanya dimanfaatkan untuk mengelompokkan dokumen saja. *Clustering* belum dimanfaatkan untuk menambang informasi (*text mining*). Penggalan informasi dengan *clustering* dapat dilakukan dengan memunculkan label secara otomatis pada kluster yang terbentuk.

Tujuan penelitian ini adalah memunculkan istilah potensi daerah pada label kluster secara otomatis dengan clustering. Namun, tidak setiap kluster yang didapat dari proses *clustering* memiliki label yang berkaitan dengan istilah potensi daerah. Oleh karena itu, proses *clustering* ini perlu diarahkan agar *cluster auto-labelling* dapat memunculkan label yang memuat istilah potensi daerah. Algoritme *clustering* yang digunakan adalah Bisecting K-Means, STC, dan LINGO. Kluster dan label kluster yang dihasilkan oleh ketiga algoritme dianalisis untuk menentukan algoritme manakah yang paling baik dalam melakukan penamaan kluster otomatis dengan istilah potensi daerah.

Pada penelitian ini, penamaan kluster otomatis dengan istilah potensi daerah dapat dilakukan dengan cara menghilangkan nama lokasi KKN-PPM dari dokumen data set. Namun hal ini hanya berlaku untuk algoritme Bisecting K-Means dan LINGO, tetapi tidak pada STC. Berdasarkan hasil evaluasi, dari tiga algoritme yang digunakan, algoritme yang paling baik dalam melakukan penamaan kluster otomatis potensi daerah lokasi KKN-PPM UGM adalah Bisecting K-Means. Hasil analisis label kluster menyimpulkan bahwa potensi daerah sampel penelitian (Kabupaten Sleman) adalah sapi, ikan, kambing, ikan, salak, jahe, jambu, mete, wisata, sengan, kerajinan besek, kakao.

Kata Kunci— pengelompokan dokumen, penamaan kluster otomatis, potensi daerah