

BAB VI PENUTUP

Bab ini berisi rangkuman singkat dari apa yang telah dilakukan pada tugas akhir ini, kesimpulan dari penelitian yang telah dilakukan, kendala yang dihadapi selama penelitian dilakukan. Pada subbab terakhir diberikan saran-saran yang terkait dengan pengembangan lebih lanjut dari penelitian ini.

6.1. Rangkuman

Penelitian ini bertujuan untuk menerapkan teknik reduksi dimensi Nonnegative Matrix Factorization dan Random Projection untuk aplikasi pengelompokan dokumen. Dari penerapan tersebut kemudian dapat dipelajari dan dianalisa kinerja dari masing-masing teknik. Penerapan kedua teknik ini telah dilakukan pada dokumen bahasa Inggris oleh Berry & Shahnaz (2004) dan Lin & Gunopulos (2003). Namun, penerapan pada dokumen bahasa Indonesia belum ada. Oleh karena itu, penelitian ini menerapkan kedua teknik tersebut pada dokumen bahasa Indonesia. Penelitian dilakukan dengan menerapkan teknik Nonnegative Matrix Factorization dan Random Projection pada artikel media massa yang diambil dari Kompas dan Antara.

Konsep yang digunakan pada pengelompokan dokumen yang dilakukan pada penelitian ini adalah konsep *unigram*. Setiap dokumen dibedakan dengan cara melihat kemunculan kata (fitur) pada dokumen yang bersangkutan. Namun, sebelumnya dimensi dari representasi dokumen direduksi terlebih dahulu.

Beberapa percobaan dilakukan dengan menggunakan variasi dari beberapa variabel percobaan yang digunakan untuk menganalisa kinerja dari masing-masing teknik. Variabel percobaan yang digunakan adalah penggunaan *stopwords*, jumlah fitur yang digunakan, jenis informasi fitur, parameter-parameter khusus masing-masing teknik, ukuran kluster, jumlah kluster, keseragaman ukuran kluster, teknik pengelompokan dokumen, aspek kemiripan kluster, pengelompokan dokumen dari sumber berbeda, dan pengelompokan dokumen ke dalam kluster yang jumlahnya melebihi kategori yang digunakan.

6.2. Kesimpulan

Dari percobaan yang telah dilakukan dapat disimpulkan bahwa teknik reduksi dimensi Nonnegative Matrix Factorization dan Random Projection dapat diterapkan dalam aplikasi pengelompokan dokumen bahasa Indonesia. Hal ini dikarenakan akurasi yang dihasilkan masing-masing teknik sudah cukup baik (untuk beberapa konfigurasi telah mencapai akurasi hingga 97%). Hasil ini hampir sama dengan penerapannya pada dokumen bahasa Inggris yang telah dilakukan oleh Berry & Shahnaz (2004) yang juga mencapai akurasi sekitar 97%. Selain itu, hal ini juga menunjukkan bahwa penggunaan fitur berupa kata dalam membentuk informasi untuk membedakan satu dokumen dengan dokumen lain cukup efektif.

Secara umum dari aspek teknik yang digunakan, teknik Nonnegative Matrix Factorization menghasilkan akurasi yang lebih baik dibandingkan dengan teknik Random Projection. Disamping itu, penggunaan jenis informasi fitur berupa *presence* memberikan hasil terbaik dibandingkan dengan jenis informasi fitur yang lain yaitu *frequency*, *frequency normalized term frequency*, dan *term frequency-inverse document frequency*. Jumlah fitur dan penggunaan *stopwords* juga dapat mempengaruhi akurasi pengelompokan.

Dari pola yang muncul pada hasil percobaan, aspek dokumen atau data yang digunakan juga dapat mempengaruhi akurasi pengelompokan. Aspek dokumen yang dimaksud adalah ukuran kluster, jumlah kluster, kemiripan kluster, dan sumber dokumen. Akurasi pengelompokan bisa mencapai nilai yang lebih tinggi lagi dari apa yang didapat dari percobaan ini jika data yang digunakan lebih besar. Hal ini sesuai dengan pola yang muncul pada hasil percobaan.

6.3. Kendala

Kendala utama yang dihadapi selama melakukan penelitian ini adalah kurangnya sumber daya data yaitu dokumen bahasa Indonesia. Dalam penelitian ini, jumlah dokumen yang digunakan dirasa masih kurang banyak. Walaupun dokumen ini diambil secara otomatis oleh *crawler*, terdapat beberapa spesifikasi *website* yang menyebabkan jumlah dokumen yang diambil tidak banyak. Artikel yang hanya ditampilkan per hari pada *website* Kompas misalnya, menyebabkan pengambilan

harus dilakukan tiap hari dan karena keterbatasan waktu, jumlah dokumen yang diambil tidak banyak.

6.4. Saran

Beberapa saran untuk pengerjaan topik penelitian ini lebih lanjut diantaranya:

1. Dalam penelitian ini jumlah dokumen yang digunakan sangat terbatas. Oleh karena itu, penelitian dengan jumlah data yang lebih besar dengan variasi (kategori) yang lebih banyak memungkinkan untuk mendapatkan hasil dan kesimpulan yang lebih baik.
2. Percobaan dengan menggunakan fitur berupa kata dasar melalui proses *stemming* masih belum dilakukan. Diperkirakan, penggunaan fitur berupa kata dasar bisa memberikan hasil yang lebih baik.
3. Dalam penelitian ini, teknik pengelompokan yang digunakan setelah dimensi direduksi dengan teknik Random Projection adalah K-Means. Masih terdapat banyak teknik pengelompokan yang bisa digunakan.
4. Pada penerapan teknik Nonnegative Matrix Factorization dalam penelitian ini, iterasi perubahan (*update*) *term-document matrix* dikendalikan oleh parameter jumlah iterasi. *Stopping criteria* lain yang bisa digunakan adalah selisih besaran matriks antar 2 iterasi yang berurutan. Jika selisih besaran matriks tersebut lebih kecil dari nilai yang telah ditentukan, maka matriks dianggap telah konvergen dan iterasi pun berhenti.