

BAB I PENDAHULUAN

Pada bab ini akan dijelaskan latar belakang penelitian, permasalahan yang menyebabkan perlunya dilakukan penelitian, tujuan dari penelitian, ruang lingkup penelitian, metodologi yang digunakan dalam penelitian, serta sistematika dari penulisan laporan tugas akhir ini.

1.1. Latar Belakang

Kebutuhan akan informasi semakin meningkat. Keefisienan dan keefektifan pencarian informasi tersebut telah menjadi tolak ukur yang cukup penting untuk diperhitungkan. Kebutuhan untuk mendapatkan informasi bukan hanya menyangkut ketepatan saja, tetapi juga menyangkut kecepatan. Oleh karena itu, pengorganisasian informasi menjadi penting dalam mendukung tercapainya kebutuhan tersebut. Untuk jumlah data yang sedikit, kebutuhan ini memang dirasa tidak terlalu memberikan kontribusi yang besar. Namun, bila berbicara mengenai jumlah data yang besar, pengorganisasian yang baik dapat sangat membantu dalam proses pencarian informasi (*information retrieval*).

Seringkali, bila diberikan suatu koleksi dokumen, muncul kebutuhan untuk mengelompokkan dokumen tersebut ke dalam kluster-kluster berdasarkan tingkat kemiripan isi dari dokumen yang ada. Untuk jumlah dokumen yang sedikit, proses pengelompokan secara manual masih mungkin untuk dilakukan. Namun, untuk mengelompokkan dokumen dengan jumlah yang banyak, proses pengelompokan secara manual akan memakan waktu yang lama. Proses otomatisasi menjadi satu solusi yang bisa digunakan untuk mengurangi waktu yang diperlukan untuk melakukan pengelompokan.

Teknik pengelompokan dokumen atau *document clustering* telah berkembang dan menjadi suatu alat yang bisa digunakan untuk pengorganisasian dokumen dalam suatu koleksi yang besar. Dengan teknik pengelompokan dokumen yang baik, komputer dapat secara otomatis mengelompokkan dokumen ke dalam kluster yang tepat. Hal ini dapat mendukung pencarian informasi secara efisien.

Penelitian mengenai teknik pengelompokan dokumen dilakukan untuk menjawab kebutuhan yang ada. Begitu juga dengan penelitian teknik tersebut dalam domain bahasa Indonesia. Dengan adanya, penelitian ini diharapkan dapat membantu mewujudkan kebutuhan tersebut.

1.2. Permasalahan

Telah banyak teknik yang dikembangkan untuk mendukung proses pengelompokan dokumen. Penelitian pun sudah dilakukan dengan memakai beberapa teknik yang ada untuk melakukan pengelompokan dokumen. Dari percobaan tersebut dapat dipelajari hasil penerapan teknik untuk keperluan pengelompokan dokumen. Hanya saja, penelitian dengan teknik pengelompokan yang sudah ada umumnya mengolah dokumen dalam domain bahasa Inggris. Penerapan teknik untuk pengelompokan dokumen pada domain bahasa Indonesia, khususnya teknik Nonnegative Matrix Factorization dan Random Projection, belum pernah dilakukan sehingga tidak ada pengetahuan mengenai hasil penerapan teknik tersebut untuk keperluan pengelompokan dokumen.

1.3. Tujuan

Tujuan dari tugas akhir ini adalah menerapkan teknik Non-Negative Matrix Factorization (NMF) dan Random Projection untuk aplikasi pengelompokan dokumen bahasa Indonesia. Tantangan yang cukup besar dalam *information retrieval* adalah *high dimensionality* (dimensi tinggi) dimana perlu biaya komputasi yang tinggi. Oleh karena Non-Negative Matrix Factorization dan Random Projection akan mengurangi dimensi yang diperlukan, maka percobaan berusaha melihat kinerja dari masing-masing teknik tersebut dalam mengelompokkan dokumen ke dalam kluster dari sisi akurasi. Dengan demikian, bisa dilihat kemungkinan teknik tersebut digunakan sebagai teknik dalam aplikasi pengelompokan dokumen.

Penelitian ini juga bertujuan untuk melihat pengaruh beberapa variabel percobaan seperti jumlah kluster, jumlah dokumen terhadap akurasi dari masing-masing teknik. Dengan demikian, pola kerja dan kinerja dari masing-masing pun bisa diketahui secara lebih mendetil.

1.4. Ruang Lingkup

Ruang lingkup pengerjaan tugas akhir ini meliputi:

1. Teknik pengelompokan dokumen yang digunakan adalah Non-Negative Matrix Factorization dan Random Projection. Untuk teknik Random Projection, diperlukan tahap tambahan agar dokumen bisa dikelompokkan ke dalam kluster. Teknik tambahan yang digunakan adalah K-Means.
2. Pengelompokan dokumen dilakukan pada artikel media massa. Artikel media massa yang digunakan didapat dengan cara pengambilan langsung dari *website* Kompas dan Antara.
3. Fitur yang digunakan adalah fitur unigram dengan melakukan variasi pada jumlah fitur.
4. Pengerjaan tugas akhir dimulai dari persiapan data, pengolahan data, hingga penerapan teknik pengelompokan dokumen dan didapat akurasi. Biaya komputasi tidak menjadi fokus penelitian ini.

1.5. Metodologi Penelitian

Metodologi yang digunakan pada pengerjaan tugas akhir ini adalah metodologi eksperimental yang meliputi tahapan-tahapan sebagai berikut:

1. Studi literatur – Mencari dan mempelajari bahan-bahan pustaka terkait dengan teknik pengelompokan dokumen.
2. Perancangan – Melakukan perancangan terhadap implementasi *tools* dan perancangan percobaan yang akan dilakukan dengan mempersiapkan data dan menetapkan variabel percobaan yang akan digunakan.
3. Implementasi – Melakukan implementasi *tools* yang akan digunakan dalam percobaan dan melakukan percobaan yang telah dirancang dengan menerapkan teknik yang telah diimplementasi.
4. Analisis hasil – Mempelajari hasil percobaan dengan melihat akurasi untuk tiap teknik yang dilihat dari beberapa aspek seperti teknik, jumlah kluster, dan jumlah dokumen.

1.6. Sistematika Penulisan

Sistematika penulisan laporan tugas akhir ini adalah sebagai berikut:

- **BAB I PENDAHULUAN** - Pada bab ini akan dijelaskan latar belakang penelitian, permasalahan yang menyebabkan perlunya dilakukan penelitian, tujuan dari penelitian, ruang lingkup penelitian, metodologi yang digunakan dalam penelitian, serta sistematika dari penulisan laporan tugas akhir ini.
- **BAB II LANDASAN TEORI** – Pada bab ini akan dijelaskan landasan teori yang digunakan pada pengerjaan tugas akhir ini. Landasan teori meliputi penjelasan mengenai pengelompokan dokumen, teknik-teknik yang digunakan serta penelitian yang telah dilakukan terkait dengan pengelompokan dokumen.
- **BAB III PERANCANGAN** – Pada bab ini akan dijelaskan tahapan yang dilalui dalam melakukan perancangan penelitian yang akan dilakukan dalam tugas akhir ini. Tahapan tersebut meliputi perancangan implementasi *tools* yang akan digunakan dalam percobaan serta perancangan percobaan yang meliputi persiapan data dan penetapan variabel percobaan yang akan digunakan.
- **BAB IV IMPLEMENTASI** – Pada bab ini akan dijelaskan implementasi dari rancangan yang telah dibuat yang meliputi rancangan implementasi *tools* yang akan digunakan dalam percobaan serta rancangan percobaan yang meliputi variabel percobaan yang akan digunakan.
- **BAB V HASIL DAN PEMBAHASAN** – Pada bab ini akan dijelaskan hasil dari percobaan yang dilakukan mengenai pengelompokan dokumen dengan menggunakan teknik yang dijelaskan pada bab 2. Pembahasan meliputi perbandingan akurasi dari percobaan yang dilakukan dengan menggunakan variasi variabel percobaan.
- **BAB VI PENUTUP** – Pada bab ini dipaparkan kesimpulan dari seluruh percobaan yang dilakukan dan penjelasan mengenai kendala-kendala yang ditemui selama melakukan penelitian serta saran-saran terkait dengan pengerjaan penelitian.