



Hak cipta dan penggunaan kembali:

Lisensi ini mengizinkan setiap orang untuk menggubah, memperbaiki, dan membuat ciptaan turunan bukan untuk kepentingan komersial, selama anda mencantumkan nama penulis dan melisensikan ciptaan turunan dengan syarat yang serupa dengan ciptaan asli.

Copyright and reuse:

This license lets you remix, tweak, and build upon work non-commercially, as long as you credit the origin creator and license it on your new creations under the identical terms.

BAB I

PENDAHULUAN

1.1 Latar Belakang

Revolusi teknologi informasi ikut mengubah perilaku masyarakat modern dengan mencoba masuk dan menghirup atmosfer kebudayaan baru, menjadikan masyarakat tanpa kertas (*paperless society*) (Hernawati, 2005). Berdasarkan buku pedoman Perpustakaan Perguruan Tinggi (Depdiknas, 2004), salah satu tugas perpustakaan perguruan tinggi adalah menghasilkan karya-karya penelitian yang dapat dipublikasikan untuk kepentingan pembangunan masyarakat dalam berbagai bidang. Menurut Yuadi (2006), perpustakaan tradisional memiliki keterbatasan dalam penyimpanan dan akses informasi, maka banyak perpustakaan yang mulai beralih ke arah penyimpanan dokumen digital (tanpa kertas) untuk akses yang lebih luas. Besarnya dan berkembangnya jumlah penelitian yang diterbitkan, termasuk peningkatan publikasi membuat tugas untuk identifikasi studi menjadi kompleks dan memakan waktu yang lama (O'Mara-Eves dkk., 2015).

Mengakses data dalam jumlah besar dan tidak terstruktur sangat menyulitkan untuk menemukan intisari atau kesimpulan data. Teknik *text mining* sangat tepat digunakan untuk melakukan ekstraksi informasi dari data tekstual tidak terstruktur yang kemudian diidentifikasi dan dieksplorasi untuk mencari pola informasi penting (Feldman dan Sanger, 2007). Menurut Gupta dan Lehal (2009) *text mining* berbeda dengan pencarian *web* karena pada pencarian *web* ditujukan untuk pencarian data yang telah diketahui dan telah ditulis oleh orang lain, sedangkan *text mining* digunakan untuk mendapatkan informasi yang belum

diketahui dan belum ditulis oleh orang lain.

Terkadang mahasiswa menghabiskan banyak waktu di perpustakaan hanya untuk mencari referensi skripsi dan hasil pencarian yang diharapkan seringkali tidak sesuai dengan kenyataan yang ada (Prabowo dkk., 2013). Berdasarkan pernyataan di atas, waktu yang banyak dihabiskan mahasiswa disebabkan karena banyaknya jumlah dan isi dokumen yang panjang. Membaca dokumen dengan isi yang panjang memerlukan waktu yang lama untuk menemukan intisari dari dokumen (Ahda, 2015). Pengetahuan tentang tren topik skripsi mahasiswa di sebuah universitas pada umumnya maupun di program studi tertentu pada khususnya dapat membawa manfaat yang sangat positif bagi pengembangan kurikulum maupun perencanaan roadmap penelitian skala institusi (Prilianti dan Wijaya, 2014). Diperlukan metode-metode analisis untuk mendapatkan informasi mengenai tren topik skripsi. Metode *clustering* dapat digunakan untuk analisis dan mengelompokkan dokumen skripsi berdasarkan tingkat kemiripan satu sama lain. Analisis *cluster* merupakan suatu teknik yang dipergunakan untuk mengklasifikasikan objek ke dalam kelompok yang relatif homogen tanpa ditentukan terlebih dahulu di awal (Sitepu dkk., 2011).

Terdapat 2 pembagian metode analisis *cluster*, yaitu metode analisis *hierarchical* dan *non-hierarchical*. Menurut Johnson dan Wichern (2002) metode *non-hierarchical* dapat diterapkan pada kasus dengan jumlah objek yang sangat besar. Metode K-Medoids atau disebut *Partitioning Around Medoids* (PAM) adalah metode analisis *cluster non-hierarchical* dan merupakan pengembangan dari metode K-Means. Keduanya memiliki kesamaan, yaitu memecah *dataset* menjadi kelompok-kelompok. Perbedaan dari keduanya adalah PAM bekerja dengan

medoids yang merupakan entitas dari *dataset* yang mewakili kelompok tempat dimasukkannya entitas tersebut, sedangkan K-Means berkerja dengan *centroid* yang merupakan artifisial hasil entitas untuk mewakili kelompok (Kaufman dan Rousseeuw, 2005). Ada penelitian yang menyatakan bahwa metode K-Medoids lebih baik performanya dibandingkan dengan metode K-Means, seperti pada penelitian *A K-Mean-like Algorithm for K-medoids Clustering and Its Performance* (Park dkk., 2006) dan penelitian Perbandingan Partition Around Medoids (PAM) dan K-Means Clustering untuk Tweets (Wibisono, 2011). Sebenarnya sudah ada penelitian sebelumnya untuk menentukan tren topik skripsi, tetapi penelitian tersebut menggunakan metode K-Means, berbasis *desktop*, studi kasus yang berbeda dan hasil akhirnya hanya berupa grafik frekuensi kemunculan kata (Prilianti dan Wijaya, 2014). Hasil dari penelitian sebelumnya, menyatakan bahwa hasil analisis meminimalkan keterlibatan pengguna dan hasil akhir yang didapatkan positif terhadap aspek kinerja dari algoritma K-Means *clustering*, penggunaan aplikasi, dan akurasi konklusi.

Hal tersebutlah yang menjadi dasar penelitian dalam menerapkan metode K-Medoids pada rancang bangun aplikasi *text mining* penentu tren topik skripsi.

1.2 Rumusan Masalah

Rumusan masalah dalam penelitian ini adalah sebagai berikut.

1. Bagaimana cara merancang dan membangun aplikasi *text mining* untuk penentuan tren topik skripsi dengan metode K-Medoids *clustering*?
2. Bagaimana cara mengetahui ketepatan pengelompokan objek dalam sebuah *cluster* dalam penelitian ini?

1.3 Batasan Masalah

Berikut adalah batasan masalah dalam penelitian ini.

1. Data yang digunakan sebagai masukan adalah *file* abstrak dari skripsi mahasiswa fakultas Teknologi Informasi dan Komunikasi (ICT) di Universitas Multimedia Nusantara (Teknik Informatika, Sistem Informasi, dan Sistem Komputer) dari tahun 2011 hingga 2015.
2. Proses *stemming* pada tahap *preprocessing* data hanya untuk bahasa Indonesia saja dengan menggunakan algoritma Nazief Adriani.
3. Jumlah maksimal *cluster* yang digunakan dalam penelitian ini sebesar 3 *cluster* dan jumlah kata yang masuk dalam *clustering* dibatasi hanya untuk 100 kata dengan jumlah frekuensi teratas.
4. Proses *clustering* dilakukan secara manual, yaitu dilakukan masing-masing sesuai dengan *filter* tahun dan jurusan yang dipilih.
5. Tipe ekstensi *file* abstrak yang digunakan hanya PDF (.pdf).

1.4 Tujuan Penelitian

Berdasarkan rumusan masalah, tujuan penelitian ini adalah sebagai berikut.

1. Merancang dan membangun sebuah aplikasi *text mining* berbasis *web* untuk penentuan tren topik skripsi dengan metode K-Medoids *clustering*.
2. Mengetahui ketepatan pengelompokan objek dari sebuah *cluster* pada penelitian ini dengan menggunakan metode *purity*.

1.5 Manfaat Penelitian

Penelitian ini memiliki manfaat untuk membantu para pengguna aplikasi mendapatkan rekomendasi topik skripsi hasil dari pengelompokan *cluster* berdasarkan frekuensi kemunculan setiap kata pada *file* abstrak skripsi.

1.6 Sistematika Penulisan

Sistematika penulisan yang digunakan dalam penyajian laporan skripsi ini adalah sebagai berikut.

BAB I PENDAHULUAN

Bab ini berisi latar belakang masalah, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, dan sistematika penulisan.

BAB II LANDASAN TEORI

Bab ini menjelaskan teori-teori dan konsep dasar yang mendukung penelitian terkait, yaitu teori dasar mengenai *text mining*, tahapan *preprocessing* data, algoritma Nazief Adriani, analisis *cluster*, K-Medoids, Manhattan Distance, *purity*, Likert Scale, dan *framework* CodeIgniter.

BAB III METODOLOGI DAN PERANCANGAN SISTEM

Bab ini menjelaskan metode penelitian dan rancangan dari aplikasi yang dibuat.

BAB IV IMPLEMENTASI DAN UJI COBA

Bab ini berisi implementasi sistem, hasil penelitian beserta dengan hasil

analisis dari data tersebut, dan uji coba aplikasi kepada beberapa mahasiswa Fakultas Teknologi Informasi dan Komunikasi.

BAB V SIMPULAN DAN SARAN

Bab ini berisi simpulan hasil penelitian terhadap tujuan yang ingin dicapai dalam penelitian ini dan saran pengembangan penelitian lebih lanjut.

