

BAB I

PENDAHULUAN

1.1 Latar Belakang

Google Scholar merupakan sebuah wadah untuk para peneliti mempublikasikan penelitian mereka. Jutaan artikel jurnal, prosiding dan lain-lain yang telah dipublikasikan oleh peneliti dapat ditemukan dalam *Google Scholar*, dengan sebuah kata kunci kita dapat menemukan beberapa dokumen yang mungkin sesuai dengan kata kunci yang kita masukan. Kata kunci ini bertujuan untuk memudahkan dalam mencari dokumen yang diinginkan, tanpa harus mencari satu persatu dokumen yang diinginkan di dalam jutaan dokumen yang ada di dalam *Google Scholar*.

Text mining bisa digunakan untuk membersihkan sebuah data untuk kemudian akan diolah dengan metode data *clustering*. Sehingga didapatkan sebuah kombinasi *cloudword* dari kata kunci sebuah dokumen yang akan dibandingkan dengan keseluruhan isi dokumen. Dengan demikian didapatkan lah sebuah kesimpulan keakuratan representasi dari *keyword* tersebut. *Clustering* sendiri telah digunakan dalam berbagai disiplin ilmu seperti rekayasa perangkat lunak, statistik, penambangan data, analisis gambar, *machine learning*, mesin *cluster web*, dan *text mining* untuk menyimpulkan kelompok dalam volume data yang besar. Sedangkan algoritma *K-Nearest Neighbor* adalah sebuah algoritma yang dapat membagi data kedalam cluster yang sejenis atau sama.

Dalam penelitian kali ini, kami akan menguji seberapa besar korelasi dan keakuratan sebuah kombinasi *keyword* yang didapat dari dokumen publikasi ilmiah dari hasil *crawling* yang nantinya *keyword-keyword* tersebut akan di eliminasi dengan menggunakan *text mining*, agar tidak terdapat *keyword* yang memiliki *duplicate* untuk ditampilkan kedalam *wordcloud*. Setelah terkumpul kombinasi kata pada *wordcloud* maka akan dibandingkan dengan keseluruhan isi dokumen dengan mengolah-nya menggunakan metode *K-Nearest Neighbor* dan akan dicari presentasi seringnya *keyword* tersebut keluar pada keseluruhan dokumen.

1.2 Rumusan Masalah

Adapun rumusan masalah dari pementaran proposal tugas akhir ini antara lain:

- a) Membangun sistem yang dapat menentukan *wordcloud* dari kata kunci yang di dapat dari seluruh isi dokumen.
- b) Membangun sistem yang dapat menampilkan *wordcloud* dengan *text mining*.
- c) Membuat sistem yang dapat menampilkan presentasi dengan metode *K-Nearest Neighbor*.
- d) Membangun sistem yang dapat melakukan *crawling* dokumen publikasi ilmiah dari *Google Scholar*.

1.3 Tujuan

Adapun tujuan dalam Tugas Akhir ini antara lain:

- a) Dapat melakukan *crawling* dokumen publikasi ilmiah dari *Google Scholar*.
- b) Dokumen yang didapat dari hasil *crawling* akan di ambil kata kunci-nya untuk menampilkan *wordcloud* dengan menggunakan *text mining*.
- c) Dapat merepresentasikan dengan menggunakan metode *K-Nearest Neighbor*.

1.4 Batasan Masalah

Adapun batasan masalah dalam Tugas Akhir ini antara lain:

- a) Data dokumen yang didapat hanya berasal dari *Google Scholar*.
- b) Kategori data *keyword* dokumen yang akan dihimpun kedalam metode *K-Nearest Neighbor* merupakan kumpulan *keyword* yang didapat dari dokumen hasil *crawling* berbahasa inggris.
- c) Dokumen publikasi yang digunakan memiliki keterangan kata kunci pada bagian abstrak
- d) Bahasa pemrograman yang digunakan adalah Bahasa R.

1.5 Metodologi Penelitian

Untuk menyelesaikan tugas akhir ini, penulis akan menerapkan beberapa metode untuk menyelesaikan masalah. Metode yang digunakan adalah sebagai berikut:

1. Studi Literatur

Metode ini dilaksanakan dengan melakukan studi kepustakaan yang relevan serta buku-buku maupun artikel-artikel atau *e-book* dan juga jurnal internasional yang didapatkan melalui internet.

2. Analisis

Pada tahap ini digunakan untuk mengolah data yang ada dan kemudian melakukan analisis terhadap hasil studi literatur yang diperoleh sehingga menjadi suatu informasi.

3. Perancangan Perangkat Lunak

Pada tahap ini, digunakan seluruh hasil analisa terhadap studi literatur yang dilakukan untuk merancang program yang akan dihasilkan. Dalam tahapan ini juga dilakukan perancangan visualisasi dari data yang di apat.

4. Implementasi dan Pengujian Sistem

Pada tahap ini dilakukan implementasi *crawling data*, *data mining* dan metode *K-Nearest Neighbor*.

5. Penyusunan Laporan Tugas Akhir

Pada tahap ini dilakukan penyusunan laporan tugas akhir serta pengumpulan dokumen yang terkait dengan sistem, format laporan yang digunakan mengikuti kaidah penulisan yang benar serta sesuai dengan ketentuan-ketentuan yang diterapkan oleh institusi.

1.6 Sistematika Penulisan

Penulisan Tugas Akhir ini akan dibagi menjadi lima bab, yaitu:

BAB I PENDAHULUAN

Bab ini berisi latar belakang pemilihan topik, rumusan dan batasan masalah, tujuan, metode penelitian dan sistematika penulisan.

BAB II LANDASAN TEORI

Bab ini membahas mengenai dasar-dasar teori yang akan digunakan pada penelitian dalam Tugas Akhir ini untuk memecahkan masalah yang diambil dari berbagai sumber.

BAB III PERANCANGAN SISTEM

Bab ini berisi tentang perancangan sistem yang akan dibangun pada Tugas Akhir ini, antara lain analisa kebutuhan sistem dan perancangan sistem.

BAB IV IMPLEMENTASI DAN PENGUJIAN SISTEM

Bab ini membahas mengenai pengujian dan analisis terhadap animasi yang telah dibangun.

BAB V KESIMPULAN DAN SARAN

Bab ini membahas mengenai kesimpulan yang bisa ditarik dari penelitian ini dan memberikan saran-saran yang bisa membangun penelitian selanjutnya.