

BAB I

PENDAHULUAN

1.1. Latar Belakang Masalah

Internet telah menjadi bagian yang tidak bisa di lepaskan dari hidup manusia pada jaman sekarang. Dengan perkembangan pesat dan semakin popular nya internet, membuat internet telah merambah ke berbagai aspek masyarakat [1]. Google scholar merupakan sebuah wadah untuk para peneliti mempublikasikan penelitian mereka. Jutaan artikel jurnal, prosiding dll yang telah di publikasikan 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 di inginkan, tanpa harus mencari satu persatu dokumen yang di inginkan di dalam jutaan dokumen yang ada di dalam google scholar.

Text mining bisa digunakan untuk membersihkan sebuah data untuk kemudian akan di olah dengan metode *data clustering*. Sehingga didapatkan sebuah kombinasi *Cloudword* dari kata kunci sebuah dokumen yang akan di bandingkan dengan keseluruhan isi dokumen. Dengan demikian di dapatkan 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-Means* adalah sebuah alogaritma 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 di dapat dari dokumen publikasi ilmiah. Yang nantinya *Keyword-keyword* tersebut akan di *Filter*, agar tidak terdapat *keyword* yang memiliki *duplicate* untuk di tampilkan kedalam *cloudword*. Setelah terkumpul kombinasi kata pada *cloudword* maka akan di bandingkan dengan keseluruhan isi dokumen dengan mengolah nya menggunakan metode *K-Means Clustering*. Kemudian akan di cari presentasi seberapa seringnya *keyword* tersebut keluar pada keseluruhan dokumen.

1.2. Rumusan Masalah

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

- a) Membangun sistem yang dapat menentukan *Cloudword* dari kata kunci yang di inputkan secara manual.
- b) Membangun sistem yang dapat menampilkan presentasi dari *keyword* yang didapat dari Dokumen publikasi ilmiah.
- c) Membuat system yang dapat menampilkan presentasi *plot* dengan metode *K-Means Clustering*.
- d) Membangun system yang dapat melakukan *crawling* dokumen publikasi ilmiah dari *Google Scholar*.

1.3. Tujuan dan Manfaat

Adapun tujuan dalam Tugas Akhir ini antara lain:

- a) Mendapatkan *Cloudword* sebagai hasil dari *filtering* kata kunci yang di inputkan secara manual.
- b) Menampilkan *Cloudword 2* sebagai hasil dari metode *K-means clustering*.
- c) Menampilkan *Clusterplot* dari perhitungan jarak *centroid* menggunakan *K-means Clustering*.
- d) Mendapatkan file publikasi ilmiah dengan melakukan *crawling* dari *Google Scholar*.

1.4. Batasan Masalah

Adapun batasan masalah dalam Tugas Akhir ini antara lain:

- a) Data dokumen yang dikumpulkan hanya berasal dari *google scholar*.
- b) Kategori data *keyword* dokumen yang akan di himpun kedalam metode *K-Means Clustering* merupakan kumpulan *keyword* yang didapat dari dokumen hasil *crawling* berbahasa inggris.
- c) Dokumen publikasi yang di gunakan memiliki keterangan kata kunci pada bagian abstrak.
- d) Bahasa pemrograman yang digunakan adalah Bahasa R.

1.5. Metode Penelitian

Dalam penelitian ini, tahapan-tahapan yang akan dilalui adalah sebagai berikut:

- a) 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.

b) Analisis

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

c) 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 dapat.

d) Implementasi dan Pengujian Sistem

Pada tahap ini dilakukan implementasi *crawling data*, *data mining* dan metode *K-Means Clustering*

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 masalah, tujuan dan manfaat, batasan masalah, 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.