

BAB I

PENDAHULUAN

1.1 LATAR BELAKANG

Karya ilmiah yaitu karya tulis yang telah diakui dalam bidang ilmu pengetahuan, teknologi atau seni. Pada dasarnya karya ilmiah merupakan perwujudan kegiatan ilmiah yang dikomunikasikan lewat bahasa tulisan. Penulisan karya ilmiah juga menjadi salah satu kegiatan pokok di perguruan tinggi. Karya ilmiah ditulis sesuai dengan tata cara ilmiah dan mengikuti pedoman atau konvensi ilmiah yang telah disepakati atau ditetapkan oleh suatu lembaga pendidikan tinggi.

Seiring dengan kemajuan di bidang teknologi informasi yang semakin berkembang, kemudahan para akademisi ataupun para peneliti dalam membuat suatu karya ilmiah juga semakin mudah. Hanya menunggu kurang dari 5 detik, berbagai macam dokumen yang ingin dicari dan berhubungan dengan karya ilmiahnya di internet akan muncul. Hal tersebut tentu tidak lepas dari baik dan buruknya, baiknya yaitu memudahkan para peneliti tersebut untuk membantu menyelesaikan karya ilmiahnya dan buruknya ialah terjadinya plagiat yaitu penjiplakan atau pengambilan karangan, pendapat orang lain dan menjadikannya seolah-olah karangan sendiri (KBBI, 1997: 775).

Tindakan penjiplakan atau plagiat ini sering dilakukan oleh banyak orang yang tujuannya untuk memudahkannya mendapatkan gelar akademik dalam waktu singkat. Perilaku tersebut tentu menjadi perhatian pemerintah khususnya negara Indonesia untuk melakukan pencegahan agar kualitas para pelajar maupun pendidik dalam menghasilkan karya tulis ilmiahnya dapat lebih berkualitas tanpa harus merugikan orang lain.

Berbagai metode telah diteliti untuk mencari cara terbaik, waktu tercepat maupun tingkat kemiripan yang paling tepat pada dokumen teks seperti karya ilmiah. Maka dari permasalahan tersebut penulis akan meneliti salah satu metode yang dapat diterapkan dalam menghitung tingkat kemiripan untuk mengetahui

kemiripan dokumen karya ilmiah dengan metode *K-Nearest Neighbor* yaitu sebuah metode untuk melakukan klasifikasi terhadap objek berdasarkan data pembelajaran yang jaraknya paling dekat dengan objek tersebut.

1.2 RUMUSAN MASALAH

Adapun rumusan masalah dalam penelitian ini adalah :

1. Bagaimana menghitung pembobotan kata pada dokumen teks karya ilmiah menggunakan algoritma *tf/idf (term frequency/inverse document frequency)*.
2. Menghitung tingkat kemiripan dokumen teks berupa karya ilmiah menggunakan metode *k-nearest neighbor*.
3. Membandingkan hasil kemiripan dokumen karya ilmiah menggunakan perhitungan rumus *cosine similarity* dengan *euclidean distance*.

1.3 BATASAN MASALAH

Adapun batasan masalah dalam penelitian ini adalah :

1. Dokumen teks yang dibandingkan berupa karya ilmiah.
2. Karya ilmiah yang akan diidentifikasi kemiripannya menggunakan bahasa indonesia.
3. Karya ilmiah yang digunakan pada peneltian ini berupa jurnal maupun skripsi yang diambil dari beberapa situs kepustakaan *online* ilmiah.
4. File dokumen teks yang diproses berekstensi *.pdf*.
5. Pola yang diambil pada dokumen untuk diproses adalah teks.
6. Menggunakan algoritma *stemming porter* untuk bahasa indonesia
7. Dokumen yang diuji berjumlah 10 dokumen yang akan diidentifikasi kemiripannya dengan 100 dokumen yang telah di *preprocessing* dan disimpan ke dalam database.

1.4 TUJUAN PENELITIAN

1. Untuk merancang aplikasi yang dapat mengidentifikasi kemiripan suatu karya ilmiah yang satu dengan yang lainnya.

2. Untuk melakukan pengklasifikasian dokumen teks pada perangkat lunak menggunakan algoritma *k-nearest neighbor*.

1.5 RELEVANSI

Pengidentifikasian kemiripan karya ilmiah dengan metode *k-nearest neighbor* ini diharapkan dapat membantu para peneliti khususnya di bidang *text mining* agar dapat mengembangkannya untuk lebih baik lagi sehingga tindakan penjiplakan karya ilmiah dapat lebih mudah diketahui.