

Abstrak

Perkembangan teknologi sekarang ini dapat memudahkan kegiatan yang biasanya harus dilakukan oleh manusia. Dengan banyaknya informasi di internet maka manusia akan dengan mudah mengakses dan mengolah informasi dari manapun. Seperti melakukan pengecekan tingkat kemiripan dokumen yang membutuhkan waktu lama dan tenaga yang besar, maka dapat digantikan oleh aplikasi yang menerapkan *text mining*. Menurut Kamus Besar Bahasa Indonesia, kemiripan adalah hampir sama atau serupa. Pada tugas akhir ini akan dilakukan perhitungan kemiripan dokumen yaitu data abstrak tugas akhir mahasiswa Fakultas Informatika Telkom University.

Text Mining adalah suatu bahasan yang menganalisis dan mengelola teks menjadi sebuah informasi yang dapat diolah untuk tujuan tertentu. Pada *Text Mining* dikenal istilah *preprocessing* yang terdiri dari *case folding*, *tokenizing*, *filtering*, dan *stemming*. Hal ini yang dilakukan sebelum proses pembobotan *tf-idf* dan perhitungan *cosine similarity*. Metode yang digunakan adalah Cosine Similarity. Cosine Similarity merupakan metode perhitungan jarak antara vektor A dan B yang menghasilkan sudut kosinus x diantara kedua vektor tersebut. Nilai sudut kosinus antara dua vektor menentukan kesamaan dua buah objek yang dibandingkan dimana nilai terkecil adalah 0 dan nilai terbesar adalah 1. Nilai 0 menandakan bahwa kedua abstrak yang dibandingkan tidak mirip sama sekali dan semakin mendekati nilai 1 berarti tingkat kemiripannya semakin besar.

Dengan didapatkannya hasil sebesar 0.5729 pada perhitungan koefisien korelasi Pearson, maka dapat dinyatakan bahwa korelasi antara perhitungan menggunakan TF-IDF dan metode Cosine Similarity dengan penilaian secara manual memiliki nilai korelasi linear positif karena berada diantara 0 sampai 1.

Kata kunci : *text mining, kemiripan, tugas akhir, Cosine Similarity, Telkom University*