

BAB I

PENDAHULUAN

1.1. Latar Belakang

Dalam penelitian membutuhkan sebuah proses pencarian sumber pustaka untuk dapat menyempurnakan penelitian yang akan dibuat. Peneliti dalam hal ini mahasiswa Program Studi Teknik Informatika kesulitan dalam melakukan pencarian sumber pustaka untuk mengetahui topik pembahasan yang telah dilakukan oleh mahasiswa sebelumnya. Program Studi Teknik Informatika Unissula, merupakan objek yang dilakukan pada penelitian ini dimana memiliki data tugas akhir digital yang pemanfaatannya masih belum maksimal. Mahasiswa pada program studi teknik informatika ini juga tidak atau belum memiliki akses digital terhadap tugas akhir yang pernah dilakukan serta mahasiswa harus mencari tugas akhir secara fisik pada perpustakaan. Hal ini tentu membuat kesulitan mahasiswa dalam pencarian sumber pustaka yang tepat untuk dilakukannya.

Sistem pencarian merupakan salah satu teknologi yang paling sering digunakan pada saat ini. Perannya dalam mencari informasi sangat memudahkan pengguna. Ada banyak metode yang dapat digunakan seperti, Algoritma K-Nearest Neighbor (KNN) adalah algoritma yang digunakan untuk melakukan klasifikasi terhadap suatu objek, berdasarkan k buah data latih yang jaraknya paling dekat dengan objek tersebut. Syarat nilai k adalah tidak boleh lebih besar dari jumlah data latih, dan nilai k harus ganjil dan lebih dari satu (Rivki & bachtiar, 2017).

Adapun kelebihan k-NN sendiri adalah sangat *non-linear*. k-NN merupakan salah satu algoritma (model) pembelajaran mesin yang bersifat *non-parametrik*, yaitu model yang tidak mengasumsikan apa-apa mengenai distribusi *instance* di dalam *dataset*. Model *non-parametrik* biasanya lebih sulit diinterpretasikan, namun salah satu kelebihanannya adalah garis keputusan kelas yang dihasilkan model tersebut bisa jadi sangat fleksibel dan *non-linear*. Kemudian Mudah dipahami dan diimplementasikan. Dari paparan yang diberikan dan penjelasan cara menghitung jarak dalam artikel ini, cukup jelas bahwa algoritma k-NN mudah dipahami dan juga mudah diimplementasikan. Untuk mengklasifikasi

instance x menggunakan k -NN, kita cukup mendefinisikan fungsi untuk menghitung jarak antar-*instance*, menghitung jarak x dengan semua *instance* lainnya berdasarkan fungsi tersebut, dan menentukan kelas x sebagai kelas yang paling banyak muncul dalam k *instance* terdekat.

Adapun kekurangan dari metode k -NN adalah perlu menentukan parameter k (Jumlah Tetangga Terdekat). Lalu tidak menangani nilai hilang (*Missing Value*) secara implisit artinya, jika terdapat nilai hilang pada satu atau lebih variabel dari suatu *instance*, perhitungan jarak *instance* tersebut dengan *instance* lainnya menjadi tidak terdefinisi. Bagaimana coba, menghitung jarak dalam ruang 3-dimensi jika salah satu dimensi hilang? Karenanya, sebelum menerapkan k NN kerap dilakukan imputasi untuk mengisi nilai-nilai hilang yang ada pada *dataset*. Contoh teknik imputasi yang paling umum adalah mengisi nilai hilang pada suatu variabel dengan nilai rata-rata variabel tersebut (*mean imputation*). Kemudian rentan terhadap variabel yang non-informatif, Meskipun kita telah menstandarisasi rentang variabel, k -NN tetap tidak dapat mengetahui variabel mana yang signifikan dalam klasifikasi dan mana yang tidak (Cahya, 2018).

Metode Cosine Similarity merupakan metode yang digunakan untuk menghitung similarity (tingkat kesamaan) antar dua buah objek. Secara umum penghitungan metode ini didasarkan pada vector space similarity *measure*. Metode cosine similarity ini menghitung similarity antara dua buah objek (misalkan $D1$ dan $D2$) yang dinyatakan dalam dua buah *vector* dengan menggunakan *keywords* (kata kunci) dari sebuah dokumen sebagai ukuran. Kelebihan dari metode cosine similarity yaitu metode cosine similarity mempunyai konsep normalisasi panjang vektor data dengan membandingkan *N-gram* yang sejajar satu sama lain dari 2 pembandingan (Ogie, dkk, 2016).

Kelemahan dari metode Cosine Similarity sendiri yaitu memerlukan waktu yang lama dalam tahap perhitungan nilai kemiripan, sebelum melakukan perhitungan klasifikasi terlebih dahulu dilakukan perhitungan untuk mencari jarak *query*, jarak dokumen dan juga perhitungan *inner product* (Komalasari, dkk, 2018).

Salah satu metode dalam melakukan pencarian yaitu dengan menggunakan *information retrieval* yang merupakan sebuah sistem dimana dapat membantu pengguna dalam menemukan hasil pencarian yang dia inginkan atau biasa disebut informasi temu kembali. Cara kerja dari sebuah *Information Retrieval* yaitu dengan melakukan perankingan terhadap dokumen yang ada terhadap *query* yang diberikan oleh pengguna, dimana dari *query* tersebut akan mendapatkan dokumen yang memiliki nilai kemiripan atau relevansi paling tinggi sesuai keinginan pengguna (Brata dan Hetami, 2015). Sistem temu kembali bertujuan menemukan kembali dokumen-dokumen pada sebuah kumpulan data, yang dimunculkan berdasarkan *query* yang diberikan pengguna apakah sesuai atau tidak. Sehingga dapat membantu pengguna dalam menyelesaikan masalahnya (Putung dkk, 2016). Pada penelitian ini objek yang digunakan adalah laporan tugas akhir pada program studi Teknik Informatika Unissula Semarang.

Penggunaan sistem *Information Retrieval* (IR) dalam penelitian ini guna mendapatkan suatu mesin pencarian referensi data digital tugas akhir dari kata kunci yang telah dituliskan atau diketikkan oleh pengguna. Dengan kelebihan pada metode cosine similarity tersebut maka digunakanlah metode ini penelitian rancang bangun sistem temu kembali dokumen laporan tugas akhir.

1.2. Rumusan Masalah

Rumusan masalah pada penelitian ini yaitu pengembangan sistem *information retrieval* atau informasi temu kembali di antaranya:

1. Tidak atau belum adanya akses digital terhadap laporan Tugas Akhir yang telah dilakukan sebelumnya, sehingga mahasiswa mengalami kesulitan dalam mencari judul atau referensi untuk melengkapi laporan Tugas Akhir yang dibuat.
2. Mahasiswa mengalami kesulitan dalam mencari referensi laporan Tugas Akhir yang pernah dilakukan di Program Studi Teknik Informatika sesuai tema penelitian yang diinginkan guna mengolah dan mempertimbangkan laporan yang akan dibuat.

1.3. Batasan Masalah

Batasan masalah pada penelitian ini yaitu:

1. Lingkup pencarian dokumen hanya pada judul dan abstrak laporan tugas akhir mahasiswa di program studi Teknik Informatika Unissula.
2. Dokumen yang digunakan berjumlah 30 laporan tugas akhir dengan seluruh tema yang ada.
3. Keluaran yang dimunculkan berupa dokumen yang paling relevan dari kata kunci.

1.4. Tujuan Penelitian

Tujuan dari penelitian ini yaitu:

1. Merancang dan membangun suatu sistem temu kembali yang dapat membantu mahasiswa dalam melakukan pencarian dokumen laporan tugas akhir sesuai dengan kata kunci yang dituliskan oleh mahasiswa.
2. Mengimplementasikan metode *cosine similarity* untuk mendapatkan nilai dalam perhitungan kemiripan sehingga dokumen tugas akhir yang berupa judul dan abstrak sebagai perhitungan dapat di ranking dari nilai yang tinggi ke nilai yang paling rendah.

1.5. Sistematika Penulisan

Sistematika dari penulisan tugas akhir sebagai berikut :

BAB 1 : PENDAHULUAN, berisikan latar belakang, rumusan masalah batasan masalah, tujuan penelitian, manfaat penelitian, dan sistematika penulisan.

BAB 2 : TINJAUAN PUSTAKA DAN DASAR TEORI, pada tahap ini menampilkan tentang konsep dan prinsip dasar guna memecahkan permasalahan pada tugas akhir dengan bersumber dari berbagai referensi yang ada. Menguraikan hal-hal yang ada pada tugas akhir secara relevan sesuai penelitian yang dilakukan.

BAB 3 : METODE PENELITIAN, pada bab ini di sampaikan metode yang digunakan untuk melakukan perancangan sistem serta langkah-langkah yang dilakukan untuk memperoleh hasil yang tepat. Dapat berupa perhitungan, simulasi dalam komputer dan desain sistem.

BAB 4 : HASIL DAN ANALISIS PENELITIAN, memuat hasil penelitian dan pengujian alat yang dibuat maupun data yang dibuat.

BAB 5 : KESIMPULAN DAN SARAN, memuat kesimpulan dari seluruh bab yang ada dan saran-saran dari hasil yang diperoleh dan harapan pemanfaatan dari pengembangan yang dilakukan.