

BAB I

PENDAHULUAN

1.1 Latar Belakang Masalah

Kritik dan saran merupakan sesuatu yang penting bagi suatu lembaga, baik itu lembaga didalam suatu universitas maupun perusahaan. Berbagai macam upaya yang dilakukan oleh lembaga tersebut untuk mendapatkan sebuah saran atau komentar terhadap lembaga dengan berharap dapat membantu lembaga tersebut. Upaya tersebut sering dilakukan baik secara langsung maupun di media sosial mengenai kritik dan saran. Masing-masing kritik dan saran pun bermacam-macam ada yang berupa pujian, gagasan, dan juga hujatan.

Di Dalam Lembaga Universitas AMIKOM Yogyakarta, setiap alumni diberikan kesempatan untuk memberi kritik dan saran kepada pihak lembaga. Penyampaian kritik dan saran dilakukan ketika mahasiswa sudah menjadi alumni melalui pihak tracer study alumni. Penyampaian kritik dan saran ini sudah dilakukan sejak tahun 2011 hingga sekarang. Bentuk kritik dan saran bisa mengenai proses pembelajaran, fasilitas, dan bentuk lainnya.

Seiring berjalannya waktu, angka alumni dari Universitas AMIKOM Yogyakarta terus bertambah maka data kritik dan saran dari alumni pun semakin mengalami peningkatan. Hal itu menyebabkan proses pengevaluasian oleh pihak Lembaga Universitas AMIKOM Yogyakarta relatif semakin sulit.

Dengan adanya masalah tersebut, perlu adanya pengelompokkan data sesuai dengan topik kritik dan saran dari alumni Universitas AMIKOM Yogyakarta. Dari kedua data opini kritik dan saran, penelitian ini hanya akan berfokus pada data kritik sebagai acuannya. Pemilihan data kritik dilakukan karena melalui kritik dari alumni, pihak lembaga dapat mengetahui apa saja bentuk kritik yang perlu dievaluasi sehingga dapat membantu pihak lembaga dalam meningkatkan kualitas dan layanan dari universitas amikom yogyakarta.

Text mining merupakan solusi untuk mengatasi permasalahan diatas. *Text mining* merupakan suatu metode yang digunakan untuk mencari informasi didalam suatu data teks. *Text Mining* adalah sebuah proses dalam menemukan informasi dari kumpulan koleksi teks dalam jumlah besar, serta mengidentifikasi pola-pola menarik serta keterkaitannya dalam data tekstual [1].

Penggunaan tugas khusus dari *text mining* yaitu *text clustering* yang merupakan metode yang digunakan untuk mengelompokkan data kata ke dalam kelompok-kelompok tertentu yang memiliki kesamaan karakteristik. Data teks yang tidak terstruktur akan dikelompokkan berdasarkan kesamaan topik. Metode yang digunakan untuk melakukan klasterisasi data kritik adalah K-Means dengan pembobotan kata yang digunakan yaitu TF-IDF.

Metode tersebut digunakan karena jumlah dari data kritik alumni Universitas AMIKOM Yogyakarta cenderung banyak, maka tingkat akurasi dari metode K-means akan semakin baik. Setelah data dikelompokkan, setiap kelompok data diberi label keyword sesuai topiknya menggunakan TF-IDF. Dengan klasterisasi, data dapat dikelompokkan berdasarkan kesamaan topik kritik untuk mempermudah pencarian opini alumni sehingga dokumen lebih mudah dicari dan dievaluasi oleh pihak lembaga Universitas AMIKOM Yogyakarta.

Tujuan dari penelitian ini adalah untuk membantu Universitas AMIKOM Yogyakarta untuk melakukan proses pengevaluasian kinerja yang didapat melalui data hasil tracer study berdasarkan kritik atau komentar dari alumni mengenai Lembaga Universitas AMIKOM Yogyakarta dengan menggunakan metode K-Means dengan pembobotan kata TF-IDF. Selain itu penelitian ini bertujuan untuk mengetahui efisiensi kinerja dari kombinasi metode K-Means dan TF-IDF.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah disampaikan, maka permasalahan yang dapat dirumuskan adalah sebagai berikut.

1. Bagaimana klasterisasi data teks kritik alumni Universitas AMIKOM Yogyakarta dengan menggunakan algoritma K-Means?
2. Bagaimana kinerja dari K-Means dan TF-IDF terhadap data kritik ?

1.3 Batasan Masalah

Batasan masalah pada penelitian ini digunakan untuk pembahasan penelitian yang lebih spesifik dan tidak melebar dari topik. Adapun mengenai batasan masalah pada penelitian ini sebagai berikut.

1. Data yang dianalisis adalah data dari hasil tracer study alumni yang dilakukan oleh pihak lembaga Universitas AMIKOM Yogyakarta.
2. Metode yang digunakan adalah metode K-Means yang dimaksudkan untuk melakukan pengklasterisasian data dan TF-IDF sebagai pembobotan kata.
3. Data yang digunakan adalah data kritik hasil tracer study periode tahun 2019 - 2020.
4. Jumlah data yang digunakan sebanyak 3000 data.

1.4 Maksud dan Tujuan Penelitian

Adapun maksud dan tujuan dari penelitian ini adalah sebagai berikut.

1. Untuk melakukan klasterisasi data kritik Alumni Universitas AMIKOM Yogyakarta menggunakan algoritma K-Means pada tiap-tiap kritik dari alumni kepada Universitas AMIKOM Yogyakarta.
2. Mengukur kinerja dan efisiensi dari kombinasi metode K-Means dan TF-IDF.

1.5 Manfaat Penelitian

Hasil dari penelitian ini diharapkan dapat membantu Lembaga Universitas AMIKOM Yogyakarta dalam melakukan pengolahan data hasil tracer study dan pengevaluasian kinerja dan layanan lembaga terhadap mahasiswanya. Manfaat lain penelitian ini dapat mengetahui kinerja dari kombinasi antara K-Means dengan TF-IDF pada bidang klasterisasi data.

1.6 Metodologi Penelitian

Metodologi penelitian yang dilakukan untuk mendapatkan berbagai informasi yang terkait dengan objek permasalahan dalam penelitian. Untuk mengumpulkan data, peneliti menggunakan langkah-langkah sebagai berikut.

1.6.1 Metode Pengumpulan data

Metode pengumpulan data yang dilakukan antara lain :

1. Studi Pustaka

Metode studi pustaka yaitu metode yang dilakukan untuk mengumpulkan data dari berbagai literatur seperti buku-buku pustaka, artikel, jurnal-jurnal ilmiah dan berbagai macam sumber lainnya yang terkait dengan topik penelitian.

2. Dataset Internal

Pengumpulan data kritik yang tersedia dari hasil kegiatan tracer study Universitas AMIKOM Yogyakarta yang dilakukan selama tahun 2020.

1.6.2 Metode Analisis

Merupakan tahapan dalam menganalisis atau mendefinisikan permasalahan - permasalahan yang akan dibangun. Adapun metode analisis yang digunakan adalah sebagai berikut.

1. Analisis masalah
2. Pengolahan data dimulai dengan melakukan kegiatan persiapan data mentah untuk pemrosesan lebih lanjut. *Preprocessing* data terdiri dari *case folding*, *tokenizing*, *stopwords removal* dan *stemming*. Proses ini menghasilkan data teks yang telah dikonversikan menjadi huruf kecil, menghilangkan karakter selain huruf, segmentasi kata pada kalimat dan paragraf menjadi individu teks atau *tokenizing*, dan pengembalian kata yang kemudian disusun pada sebuah daftar kalimat.
3. Data diolah dengan pemilihan fitur kata yang biasa disebut *feature selection*.
4. Data teks ditransformasikan ke dalam bentuk numerik dengan memberikan pembobotan nilai TF-IDF pada setiap kalimat.
5. Pembentukan kluster yang dilakukan dengan algoritma K-Means dimana perhitungan jarak antar objek yang dihitung.
6. Penentuan jumlah kluster yang optimal dilakukan dengan menggunakan parameter kalkulasi *Sum of Square Error (SSE)*.

1.6.3 Metode Perancangan

Pada penelitian ini, program dibuat dengan menggunakan bahasa pemrograman *Python*.

1.6.4 Metode Testing (pengujian)

Metode testing yang digunakan dalam penelitian ini adalah metode *Sum of Square Error (SSE)* untuk mendapatkan nilai kluster data kritik yang lebih optimal.

1.7 Sistematika Penulisan

Pada penyusunan sistematika penulisan bertujuan untuk memudahkan pembaca dalam mengikuti apa yang dipaparkan dalam laporan skripsi ini. Adapun sistematika penulisan skripsi ini adalah sebagai berikut.

BAB I PENDAHULUAN

Dalam bab ini, berisikan gambaran umum penelitian yang dilakukan meliputi latar belakang masalah, identifikasi masalah, tujuan penelitian, batasan masalah, rumusan masalah, metode penelitian dan sistematika penulisan yang merupakan panduan dalam penyusunan landasan teori.

BAB II LANDASAN TEORI

Bab ini membahas tentang dasar-dasar teori yang dapat mendukung dalam melakukan proses klusterisasi teks dengan menggunakan algoritma K-means dan TF-IDF sebagai pembobotan kata. Dalam bab ini berfungsi sebagai referensi dalam melakukan penelitian.

BAB III ANALISIS DAN PERANCANGAN

Dalam bab ini membahas tentang metodologi yang digunakan dalam penelitian. Mencakup pada metode klusterisasi teks dengan menggunakan

algoritma K-Means dan juga metode-metode yang lainnya seperti pengumpulan dataset, *preprocessing* data, *feature selection*, pembobotan kata, klasterisasi teks dan evaluasi.

BAB IV HASIL DAN PEMBAHASAN

Bab ini berisikan penjelasan hasil dari pembahasan mengenai klasterisasi data kritik Alumni universitas AMIKOM Yogyakarta. Hasil klasterisasi yang didapatkan dengan menggunakan algoritma K-Means meliputi tiga tahapan proses klasterisasi yaitu *preprocessing* data, *feature selection* dan klasterisasi teks. Pembahasan dilakukan secara detail dengan melihat pada teori sebelumnya dengan memperhatikan dan menimbang berdasarkan eksperimen yang sudah dilakukan.

BAB V PENUTUP

Bab ini merupakan penutup dari penulisan skripsi ini yang berisi kesimpulan dari apa yang telah dipaparkan pada bab sebelumnya dan dan saran untuk pengembangan selanjutnya.

DAFTAR PUSTAKA

Daftar pustaka berisi tentang pustaka yang digunakan penulis sebagai bahan literature dalam pembuatan skripsi.