

BAB I. PENDAHULUAN

1.1. Latar Belakang

Kemajuan teknologi merupakan fenomena yang kita rasakan setiap hari di era digital ini. Tidak hanya terus memunculkan teknologi baru, kemajuan teknologi juga membawa pengetahuan, metode, sampai perspektif baru yang dapat membantu manusia dalam menyelesaikan masalah sederhana maupun kompleks. Bahkan saat ini teknologi sudah menjadi kebutuhan primer manusia yang digunakan di semua segi kehidupan manusia (Cholik, 2021). Hal ini tentu berpengaruh pada berbagai bidang, salah satunya adalah bidang pendidikan. Pada bidang ini penemuan pengetahuan, metode, dan perspektif baru ini tentu tidak bisa ditelan mentah begitu saja, namun diperlukan latihan dan pengarahan yang baik khususnya kepada mahasiswa tujuannya agar pengetahuan tersebut dapat dipahami mahasiswa dengan baik sehingga tidak disalahgunakan.

Salah satu cara untuk memastikan mahasiswa memahami betul suatu pengetahuan adalah dengan cara diberi soal uraian, karena soal uraian memberikan evaluasi yang lebih akurat dibandingkan dengan jenis soal lainnya. Oleh karena itu, disarankan agar mahasiswa dievaluasi menggunakan soal uraian (Aalaei et al., 2021). Soal uraian dianggap sebagai cara yang efektif karena melibatkan kemampuan siswa dalam pemecahan masalah dan penalaran. Akan tetapi, dengan banyaknya jawaban dari soal uraian, sulit bagi guru untuk mengetahui kemampuan dan variasi jawaban siswa secara keseluruhan (Aisyiyah, 2018).

Namun hal ini tentu memunculkan masalah baru dimana dosen seringkali masih belum menemukan cara yang efektif dalam menilai jawaban soal uraian. Adapun tantangan tersendiri bagi dosen saat menilai jawaban mahasiswa, seperti memberikan umpan balik untuk setiap jawaban, menentukan standar dari jawaban uraian untuk suatu pertanyaan, transparansi dalam melakukan penilaian, dan masalah efisiensi waktu dalam melakukan penilaian (Zulkifli et al., 2019).

Peneliti sendiri telah melakukan wawancara kepada salah satu dosen Jurusan Teknologi Informasi Politeknik Negeri Malang dan mendapat hasil berupa pada tahun ajaran 2023, dosen tersebut mengajar 2 mata kuliah dengan jumlah kelas

yang diajar sebanyak 5 kelas, dimana pada setiap kelas terdapat 20 sampai 30 mahasiswa. Jika dihitung rata-rata maka dosen tersebut mengajar sebanyak 125 mahasiswa. Hasil ini menunjukkan rasio jumlah mahasiswa terhadap dosen di Jurusan Teknologi Informasi Politeknik Negeri Malang memiliki perbandingan yang sangat tinggi, dan tentu jika satu dosen dapat mengajar 125 mahasiswa lalu pada satu semester dilakukan 4 kali evaluasi maka dosen tersebut harus menilai jawaban mahasiswa sebanyak 500 kali, sehingga proses akademik yang dilakukan seperti menilai jawaban uraian akan sangat melelahkan dan menghabiskan banyak waktu.

Clustering merupakan salah satu cara yang dapat digunakan untuk mengetahui variasi jawaban siswa secara keseluruhan. Dengan mengetahui variasi jawaban siswa, guru dapat mengetahui kemampuan siswa secara keseluruhan dan mengetahui kesalahan konsep yang mungkin terjadi sehingga dapat dilakukan evaluasi pada proses pembelajaran (Aisyiyah, 2018). *Clustering* sendiri digunakan untuk mengelompokkan dokumen ke dalam kelompok berdasarkan kesamaan dokumen tersebut dengan dokumen lainnya (Onan, 2019). Proses ini cocok untuk mengolah data jawaban karena bukan hanya dapat mendeteksi topik, *clustering* juga dapat mendeteksi pola tersembunyi dan kesalahan atau anomali yang memisahkan sebuah dokumen dengan dokumen lainnya.

Kmeans merupakan algoritma *clustering* yang dinilai paling kuat dan paling populer. Hal ini terjadi karena kmeans sangat simpel dan sangat mudah digunakan (M. Ahmed et al., 2020). Selain itu kmeans juga dapat diandalkan jika dibandingkan dengan metode lainnya, seperti *K-Nearest Neighbors* pada penelitian yang berkesimpulan metode yang diajukan yaitu kmeans menghasilkan tingkat akurasi lebih tinggi daripada metode *K-Nearest Neighbors* dalam mengolah informasi (N, Al-Obaydy, Hashim, Najm, & Jalal, 2022). Hierarchical Clustering pada penelitian yang berkesimpulan pada dataset yang digunakan Kmeans lebih cocok digunakan dibandingkan Hierarchical Clustering (Abdulhafedh, 2021). Average Linkage yang berkesimpulan berdasarkan nilai centroidnya, metode kmeans lebih cocok dibandingkan metode average linkage (Zikir et al., 2022).

Oleh karena itu penelitian yang berjudul *Clustering terhadap Jawaban Uraian Mahasiswa menggunakan TF-IDF dan KMeans* ini dilakukan dengan

harapan bahwa kemajuan teknologi dapat membantu dosen untuk menentukan metode dan teknologi yang tepat sehingga akhirnya dapat mempermudah dosen mengelompokkan jawaban dari mahasiswa yang akan membuat proses penilaian setiap jawaban uraian dari mahasiswa jauh lebih cepat.

Dengan Clustering ini setiap kata pada jawaban mahasiswa akan dibobot berdasarkan seberapa seringnya kata tersebut muncul di setiap jawaban yang ada dan jawaban mahasiswa yang sudah dibobot tersebut akan dikelompokkan sesuai dengan kemiripan inti satu jawaban dengan jawaban lainnya dimana kelompok-kelompok inilah yang akan menjadi acuan dosen dalam memberikan nilai.

1.2. Rumusan Masalah

Rumusan masalah yang diangkat dari skripsi dengan judul “Clustering Jawaban Uraian Mahasiswa Menggunakan TF-IDF dan KMeans”, adalah sebagai berikut:

1. Bagaimana cara mengolah data menggunakan TF-IDF dan Kmeans untuk mengelompokkan jawaban uraian mahasiswa?
2. Bagaimana pengaruh menggunakan TF-IDF dan KMeans dalam membantu dosen mengelompokkan jawaban uraian mahasiswa?
3. Bagaimana kualitas penggunaan TF-IDF dan KMeans dalam mengelompokkan jawaban uraian mahasiswa?

1.3. Tujuan

Tujuan dari dilakukannya skripsi dengan judul “Clustering Jawaban Ujian Mahasiswa Menggunakan TF-IDF dan KMeans”, adalah sebagai berikut:

1. Mengetahui bagaimana cara mengolah data menggunakan TF-IDF dan Kmeans untuk mengelompokkan jawaban uraian mahasiswa.
2. Mengetahui bagaimana pengaruh menggunakan TF-IDF dan KMeans dalam membantu dosen mengelompokkan jawaban uraian mahasiswa.
3. Mengetahui bagaimana kualitas penggunaan TF-IDF dan KMeans dalam mengelompokkan jawaban uraian mahasiswa.

1.4. Batasan Masalah

Batasan masalah yang diatur dalam pembuatan skripsi dengan judul “Clustering Jawaban Uraian Mahasiswa Menggunakan TF-IDF dan KMeans”, adalah sebagai berikut:

1. Penelitian hanya dilakukan terhadap jawaban mahasiswa pada Jurusan Teknologi Informasi Politeknik Negeri Malang.
2. Penelitian tidak dilakukan menilai jawaban uraian mahasiswa, namun untuk mempermudah dosen menilai jawaban uraian mahasiswa dengan cara mengelompokkan jawaban uraian mahasiswa tersebut.
3. Sistem yang dibuat hanya untuk *platform website*
4. Data jawaban uraian yang dapat diproses hanya berupa teks bahasa Indonesia.
5. Untuk jawaban dari soal bertipe kebahasaan tidak dapat langsung diproses.
6. Metode pembobotan jawaban menggunakan *TF-IDF* vectorizer.
7. Metode *clustering* yang digunakan adalah *Kmeans*.
8. Perhitungan jarak pada *Kmeans* menggunakan *euclidean distance*.

1.5. Sistematika Penelitian

Sistematika penulisan dari laporan ini berguna untuk mempermudah melihat dan mengetahui pembahasan yang ada pada setiap bab pada laporan. Sistematika ini juga merupakan kerangka dan pedoman penelitian skripsi. Adapun sistematika penelitiannya adalah sebagai berikut:

1. BAB I PENDAHULUAN
Pada bab ini berisikan tentang latar belakang, rumusan masalah, batasan masalah, tujuan, manfaat, dan sistematika penelitian
2. BAB II LANDASAN TEORI
Pada bab ini berisikan tentang studi literatur penelitian terdahulu serta beberapa dasar teori yang mendasari penelitian dan pembuatan sistem
3. BAB III METODOLOGI PENELITIAN
Pada bab ini berisikan tentang tahapan yang dilakukan untuk pembuatan sistem, diantaranya berisi waktu dan tempat penelitian, dan metode penelitian.
4. BAB IV ANALISIS DAN PERANCANGAN

Pada bab ini berisikan tentang analisis dan perancangan dalam pembuatan sistem. Analisis terdiri dari analisis permasalahan dan kebutuhan sistem, sedangkan perancangan terdiri dari perancangan sistem.

5. **BAB V IMPLEMENTASI DAN PENGUJIAN**

Pada bab ini berisikan tentang proses implementasi dan pengujian pada sistem seperti implementasi data, implemetasi algoritma, implementasi sistem, dan pengujian

6. **BAB VI HASIL DAN PEMBAHASAN**

Pada bab ini berisikan hasil dan pembahasan dari hasil implementasi dan data uji terhadap sistem

7. **BAB VII KESIMPULAN DAN SARAN**

Pada bab ini berisikan kesimpulan yang didapatkan dari hasil pengujian dan saran-saran yang dapat menyempurnakan sistem pada penelitian selanjutnya