

BAB I PENDAHULUAN

1.1 Latar Belakang

Di era digital saat ini, pertumbuhan informasi yang cepat dan luas memunculkan tantangan dalam mengelola dan mengakses informasi yang relevan. Situs berita online seperti Detik.com (Sely Karmila, 2022) menyediakan platform yang penting bagi masyarakat untuk mendapatkan informasi terkini dan beragam. Dengan ratusan judul berita yang dipublikasikan setiap harinya, penting untuk dapat mengelompokkan berita-berita tersebut ke dalam kategori yang jelas dan mudah diakses. Klusterisasi judul berita adalah pendekatan yang dapat digunakan untuk mengelompokkan berita-berita tersebut berdasarkan kesamaan topik atau tema.

Clustering text (Wisna et al., 2023) melibatkan data dalam bentuk teks tidak terstruktur, di mana teknik dalam text mining digunakan untuk menemukan kata atau pola yang penting dalam proses clustering. Penambangan teks melibatkan identifikasi informasi dan pengetahuan yang berguna dengan mengekstraksi pola dari berbagai sumber data teks. Proses penambangan teks seperti pelipatan kasus, tokenisasi, pemfilteran, dan stemming berupaya mengurangi atau mengekstraksi data dan meminimalkan gangguan dalam data. Diperlukan suatu metode untuk merepresentasikan data yang diproses dalam proses clustering sebagai vektor kata. Memanfaatkan pembobotan kata (term Weighting) untuk menentukan seberapa sering setiap istilah muncul. Dalam situasi ini, teknik yang biasanya digunakan untuk pembobotan kata adalah Term Frequency (Tf) dan Inverse Document Frequency (Idf) (Hanami, 2023).

Klusterisasi judul berita (Sely Karmila, 2022) merupakan teknik yang penting dalam analisis konten digital untuk mengelompokkan judul-judul berita yang serupa ke dalam kelompok-kelompok yang lebih terorganisir dan bermakna. Dalam konteks media digital, website berita seperti Detik merupakan sumber informasi yang menyajikan berbagai topik berita dari berbagai kategori, mulai dari politik, ekonomi, hingga hiburan, dengan frekuensi update yang tinggi.

Algoritma K-Means dipilih untuk penelitian ini karena data input yang akan diproses relatif sederhana, sehingga K-Means dianggap lebih cocok untuk melakukan pengelompokan. Selain itu, penelitian sebelumnya menunjukkan bahwa Algoritma K-Means efektif dalam mengelompokkan dokumen teks. Misalnya, (Rachman et al., 2020) berhasil Dilakukan clustering skripsi mahasiswa Universitas Ma Chung dengan algoritma K-Means, menghasilkan nilai kemurnian sebesar 76%. Hal ini menunjukkan bahwa sistem telah mengkategorikan sekitar 76% dokumen secara akurat.

Dalam konteks ini, sistem diharapkan mampu mengelompokkan berita sesuai dengan kesamaan yang dimiliki. Sehingga dari permasalahan tersebut penelitian ini akan menawarkan solusi yaitu melakukan Klusterisasi Judul Berita Pada Website Detik Menggunakan Kmeans untuk mengetahui pola kesamaan yang terdapat di judul berita.

1.2 Rumusan Masalah

Berdasarkan dari uraian latar belakang diatas maka rumusan masalah pada penelitian ini yaitu

1. Bagaimana menerapkan metode K-means untuk klusterisasi judul berita pada website Detik?
2. Seberapa efektif metode K-means Klusterisasi terhadap judul berita pada situs Detik?

1.3 Tujuan Masalah

Dari berbagai uraian diatas maka adapun tujuan penelitian ini yaitu klusterisasi Judul Berita pada Website Detik yang diharapkan mampu mengelompokkan berita sesuai dengan kesamaan

1.4 Manfaat Penelitian

Adapun manfaat yang diharapkan dalam penelitian ini adalah sebagai berikut:

1. Manfaat bagi Penulis

Penulis dapat memanfaatkan pembelajaran dan kemampuan yang diperoleh

di perkuliahan, dan memenuhi syarat kelulusan Sarjana (S1) Teknik Informatika Universitas Muhammadiyah Jember, sekaligus memperluas ilmu dan pengalaman di bidang IT.

2. Manfaat bagi Universitas Muhammadiyah Jember

Memahami tingkat keterampilan yang dimiliki mahasiswa dalam menggunakan bidang keahliannya, khususnya di bidang IT. Meningkatkan kualitas lulusan unggul melalui penelitian yang dilakukan mahasiswa Universitas Muhammadiyah Jember. Berfungsi sebagai sumber tambahan untuk meningkatkan penelitian akademis oleh siswa dalam studi mendatang.

1.5 Batasan Penelitian

Agar pembahasan tidak menyimpang dari yang sudah dijelaskan, maka penulis membuat sebuah Batasan masalah antara lain :

1. Penelitian ini hanya mengambil data dari situs Detik.com.
2. Adapun data yang diambil yaitu 500 data text.

