

# **BAB I**

## **Pendahuluan**

### **1.1 Latar Belakang**

Semakin berkembang pesatnya teknologi menyebabkan kebutuhan informasi yang besar dan juga tidak terbatas, salah satunya dalam bentuk jurnal, artikel, dan berita. Untuk berita yang panjang dan banyak pasti membutuhkan waktu yang lama juga untuk membacanya secara keseluruhan serta dapat memahami dan mengetahui isi dari dokumen berita tersebut. Berdasarkan hal tersebut maka akan dibutuhkan sistem untuk mengatasi hal tersebut dan salah satunya yaitu dengan membuat sistem peringkasan teks secara otomatis. Dengan pembuatan sistem peringkasan teks secara otomatis ini diharapkan pembaca akan lebih mudah dan cepat untuk memahami dan mengetahui informasi dari dokumen yang mereka baca.

Sistem peringkasan teks secara otomatis ini merupakan sistem yang dibangun untuk dapat meringkas sebuah teks atau dokumen berita secara otomatis dan dengan waktu yang singkat. Untuk dapat membangun sebuah sistem peringkasan teks otomatis ini diperlukan sebuah metode algoritma yang dapat membantu untuk pembuatan sistem. Ada banyak sekali metode yang dapat digunakan tapi disini penulis menggunakan metode yang cukup terkenal yaitu *K-Means Clustering*.

Ada banyak metode penelitian yang membahas tentang peringkasan teks otomatis ini salah satunya, yaitu penelitian oleh Riani, Amir Hamzah, Erna Kumalasari N pada tahun 2018. Riani dkk menggunakan metode *K-Means Clustering* untuk mengklasifikasi berita dalam bahasa Indonesia. Dalam penelitiannya disebutkan bahwa hasil akurasi dari metode *K-Means Clustering* mencapai nilai tertinggi sebesar 100% karena metode *K-Means Clustering* melakukan perhitungan secara random dan hasilnya bisa berubah. Artinya ketepatan data yang diperoleh sudah tepat karena data berita sudah sesuai, sehingga hasil akurasinya sudah sempurna dan diatas 75%. [12].

Kemudian ada penelitian dari Agus Riyanto pada tahun 2016. Agus Riyanto menggunakan metode *k-means clustering* untuk meringkas artikel berita bahasa

Indonesia. Dalam penelitiannya disebutkan bahwa jumlah kluster telah ditetapkan dan bernilai konstan sebab jumlah kluster itu sangat berpengaruh terhadap hasil ringkasan sistem. Dalam penelitian itu sistem memiliki nilai *precision* rata-rata sebesar 45.5%, rata-rata nilai *recall* sebesar 57.74% dan rata-rata nilai *f-measures* sebesar 46.13%[2].

Pada tahun 1976 J.B MacQueen pertama kali memperkenalkan metode *M-Means Clustering* ini. Kemudian metode ini menjadi salah satu algoritma *clustering* yang sangat umum digunakan untuk pengelompokan data yang sesuai dengan karakteristik yang serupa. Dalam metode *K-Means Clustering* ini data yang diolah merupakan data yang berbentuk angka. Kemudian setiap data yang diolah tadi akan dihitung kedekatannya dengan pusat kluster yang sudah ditentukan sebelumnya[5].

Berdasarkan pada penjelasan diatas, maka penulis dalam penelitian ini akan membuat sistem peringkasan otomatis untuk dokumen berita online pariwisata Madura dengan metode *K-Means Clustering*. Dengan metode *K-Means Clustering* ini didapat hasil rata-rata dari nilai akurasi yang dihasilkan kurang begitu bagus yaitu sebesar 69,95% dari 33 dokumen yang digunakan.

## **1.2 Rumusan Masalah**

Berdasarkan latar belakang diatas maka akan muncul masalah sebagai berikut:

### **1.2.1 Permasalahan**

Salah satu metode pengumpulan informasi berita pariwisata dengan membaca teks atau dokumen online. Untuk menggali intisari informasi dari sebuah dokumen online dibutuhkan metode peringkasan teks otomatis yang dapat menghasilkan ringkasan yang akurat.

### **1.2.2 Usulan Metode**

Dari permasalahan yang telah dijabarkan diatas, maka penulis mengusulkan metode *K-Means Clustering* untuk membantu mengukur tingkat akurasi dalam sistem peringkasan teks otomatis.

### **1.2.3 Pertanyaan Penelitian**

Berapa akurasi yang didapatkan oleh metode *K-Means Clustering* dalam meringkas dokumen berita pariwisata Madura pada sistem peringkasan teks otomatis?

## **1.3 Tujuan dan Manfaat**

Dari rumusan masalah diatas, maka tujuan dan manfaat yang ingin dicapai yaitu:

### **1.3.1 Tujuan penelitian**

Adapun tujuan yang ingin dicapai dalam penelitian ini yaitu untuk mengetahui seberapa akurat penerapan metode *K-Means Clustering* untuk peringkasan teks secara otomatis untuk dokumen berita online pariwisata Madura.

### **1.3.2 Manfaat penelitian**

Manfaat dari penelitian ini yaitu menghasilkan sistem peringkasan teks dokumen berita pariwisata Madura secara otomatis.

## **1.4 Batasan**

Batasan dari penelitian ini adalah:

1. Algoritma yang digunakan adalah *K-Means Clustering*.

2. Dokumen yang digunakan hanya berita pariwisata di daerah Madura.
3. Dokumen yang digunakan berupa teks berita online.
4. Proses *Preprocessing* yang dilakukan dalam penelitian ini yaitu Pemecahan kalimat, *Casefolding*, *Filtering*, *Tokenizing*, *Stemming* dan *Stopword Removal*.
5. Output yang dihasilkan oleh sistem peringkasan teks ini berupa ringkasan teks.
6. Sistem yang dihasilkan berupa aplikasi web.
7. Bahasa pemrograman yang dipakai yaitu bahasa pemrograman python.

### **1.5 Sistematika Skripsi**

Sistematika penyusunan skripsi ini dibuat untuk mengetahui deskripsi umum dari isi skripsi. Berikut merupakan sistematika penulisan skripsi ini.

#### **BAB I : PENDAHULUAN**

Pada bab ini berisi tentang latar belakang dari penelitian terkait, perumusan masalah, tujuan dan manfaat dari penelitian, batasan-batasan dan sistematika penulisan.

#### **BAB II : KAJIAN PUSTAKA**

Bab ini akan menjelaskan tentang dasar teori yang dijadikan landasan dalam penyusunan penelitian ini.

#### **BAB III : METODE USULAN**

Didalam bab ini berisi rancangan sistem yang akan dibuat, mulai dari kebutuhan *hardware* dan *software*, arsitektur sistem, algoritma.

#### **BAB IV HASIL DAN PEMBAHASAN**

Pada bab ini menjelaskan mengenai implementasi pembuatan sistem berdasarkan rancangan arsitektur yang dibuat sebelumnya.

#### **BAB V PENUTUP**

Dalam bab ini, berisikan kesimpulan penelitian yang disesuaikan dengan pertanyaan penelitian, dan saran mengenai metode dan hal-hal apa yang perlu dikembangkan kedepannya.