

BAB I

PENDAHULUAN

1.1 Latar Belakang

Artikel merupakan sebuah karya tulis yang berisi informasi dari suatu fakta maupun opini yang dipublikasi dan bertujuan untuk menarik minat pembaca sehingga membuat orang-orang datang untuk mencarinya. Pariwisata adalah salah satu topik yang hangat di kalangan remaja dan orang dewasa. Dengan seiring bertambahnya jumlah sumber artikel yang ada diperlukan waktu yang banyak untuk membacanya. Salah satu cara tercepat membaca suatu artikel yaitu membaca intisarnya saja. Dan dengan hal tersebut dibutuhkan peringkasan teks otomatis yang menghasilkan intisari dari artikel yang ada [1].

Automatic Text Summarization atau peringkasan teks otomatis adalah proses mengurangi isi dokumen menggunakan program komputer untuk menghasilkan sebuah ringkasan yang berisi intisari dari dokumen asli. Secara umum terdapat 2 metode dasar yang digunakan dalam peringkasan teks. Yaitu ekstraksi dan abstraksi, dalam metode ekstraksi ringkasan yang dihasilkan tersusun dari kata-kata dari dokumen asli. Sedangkan metode abstraksi ringkasan yang dihasilkan berisi kata-kata yang tidak termasuk termasuk dari dokumen asli [2].

Dalam penelitian yang dilakukan oleh Nikita M dan Sharvari S. Govilkar membagi dua pendekatan yang dapat digunakan dalam peringkasan teks yaitu pendekatan statistik dan linguistik. Dalam pendekatan statistik metode yaitu *Title method*, *Location method*, *tf-idf method*, *Cue word method*. Sedangkan dalam pendekatan linguistik metode yang dapat digunakan yaitu *Lexical chain*, *Word Net*, *Graph theory*, *Clustering* [3]. Dalam penelitian ini metode yang akan digunakan menggunakan metode *ekstraktif*. *TF-IDF* sering digunakan dalam peringkasan teks untuk menghitung bobot dalam suatu kata dalam dokumen, yang dikatakan dapat menghasilkan hasil relatif yang baik [4].

K-Medoid merupakan metode *clustering* yang memiliki efisiensi lebih baik dalam menangani data yang besar. Dalam penelitian purnamasari K K yaitu membandingkan peringkasan teks dengan menggunakan pengelompokan *K-Mean* dengan *K-Medoid* untuk mencari metode terbaik diantara keduanya. Berdasarkan dari 50 dokumen teks, hasil akurasi dari *K-Mean* sebesar 51.16% sedangkan akurasi *K-Medoid* sebesar 63.35%, perhitungan akurasi *K-Mean* berdasarkan ukuran dan pusat *cluster* awal yang dipilih. Hal tersebut membuktikan bahwa peringkasan menggunakan *K-Medoid* menghasilkan peringkasan yang lebih akurat dari pada *K-Mean*[5].

Pada penelitian ini akan mengimplementasikan metode *TF-IDF* untuk menghasilkan bobot dalam suatu kalimat dan *K-medoid clustering* untuk mengelompokkan kalimat-kalimat yang dianggap memiliki makna yang sama.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang sebelumnya sudah dijabarkan, munculah sebuah masalah antara lain sebagai berikut:

1.2.1 Permasalahan

Salah satu metode untuk mengumpulkan informasi dari artikel online pariwisata yaitu dengan membaca teks atau dokumen online. Untuk menggali intisari informasi dari sebuah dokumen online dibutuhkannya metode peringkasan teks otomatis yang dapat menghasilkan ringkasan yang akurat.

1.2.2 Usulan Metode

Menggunakan Metode *ekstraktif* dengan *TF-IDF* untuk menghasilkan bobot dalam kalimat dan *K-medoid* untuk mengelompokkan hasil bobot kalimat sehingga diharapkan menghasilkan ringkasan yang akurat.

1.2.3 Pertanyaan Penelitian

Berapakah nilai akurasi yang dihasilkan dari penerapan *TF-IDF* untuk menghasilkan bobot kalimat dan *K-Medoid Clustering* untuk mengelompokkan hasil bobot kalimat dalam peringkasan teks otomatis pada artikel online pariwisata madura?

1.3 Tujuan & Manfaat

Berdasarkan perumusan masalah yang dijabarkan di atas, maka tujuan dan manfaat yang ingin dicapai dalam penelitian ini adalah sebagai berikut:

1.3.1 Tujuan Penelitian

Adapun tujuan dari penelitian ini untuk mengetahui berapa nilai akurasi yang akan dihasilkan dari penerapan metode *TF-IDF* dan *K-Medoid Clustering* untuk peringkasan teks otomatis artikel online pariwisata madura.

1.3.2 Manfaat

1. Dihasilkannya sistem peringkasan teks otomatis yang menghasilkan ringkasan teks.
2. Menjadi rujukan untuk penelitian selanjutnya agar sistem yang dibangun menjadi lebih baik lagi.

1.4 Batasan – Batasan

Dalam penelitian ini terdapat beberapa batasan-batasan masalah diantaranya sebagai berikut:

1. Metode yang digunakan merupakan metode ekstraktif dalam peringkasan teks otomatis.
2. Dokumen artikel yang digunakan adalah dokumen berbahasa Indonesia.
3. Topik artikel yang digunakan yaitu pariwisata di madura.
4. Masukkan berupa teks artikel online pariwisata madura.
5. Library yang digunakan dalam pemrograman python meliputi Flask, nltk dan sastrawi.

1.5 Sistematika Penulisan

Adapun sistematika penulisan dalam penelitian ini antara lain sebagai berikut:

BAB I PENDAHULUAN

Pada bab ini menjelaskan tentang latar belakang, rumusan masalah, tujuan dan manfaat, batasan – batasan dalam penelitian dan sistematika penulisan.

BAB II KAJIAN PUSTAKA

Bab ini menjelaskan tentang landasan teori meliputi Dokumen, *Extractive Automatic Text Summarization*, *Term Frequency-Inverse Document Frequency* (TF-IDF), *K-Medoid Clustering*, dan penelitian-penelitian yang terkait dengan *TF-IDF*, *Elbow* dan *K-Medoid clustering*.

BAB III METODE PENELITIAN

Bab ini menjelaskan tentang tahapan dalam penelitian, algoritma yang akan digunakan, perancangan sistem, dokumen yang akan digunakan, arsitektur sistem, dan Skenario pengujian.

BAB IV HASIL DAN PEMBAHASAN

Pada bab ini menjelaskan tentang hasil dan implementasi pembuatan sistem berdasarkan rancangan arsitektur yang dibuat sebelumnya dan penerapan metode yang diusulkan ke dalam sistem.

BAB V PENUTUP

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