

BAB I PENDAHULUAN

1.1 Latar Belakang

Wisata adalah kegiatan yang menjadi pilihan kebutuhan pokok manusia. Objek-objek wisata tersebar di seluruh permukaan bumi, mulai dari wisata alam, buatan, kuliner, religi dan budaya. penyebaran wisata di Indonesia menjadi salah satu sumber devisa bagi negara. Namun semakin luas dan banyaknya sebaran semakin sulit dalam pemilihan objek yang sesuai dengan kebutuhan. Oleh karena itu dibutuhkan cara pengelompokan secara otomatis[1].

Information Retrieval merupakan ilmu komputer yang berkaitan dengan pengambilan informasi dari dokumen-dokumen berdasarkan pada isi dan konteks dari dokumen-dokumen tersebut. Proses Information Retrieval dideskripsikan sebagai suatu proses untuk mendapatkan keluaran yang relevan dari suatu koleksi melalui pencarian query yang diberikan user [2].

Klasifikasi menjadi salah satu cara untuk penyusunan data secara rapi untuk mengelompokkan objek yang memiliki karakteristik atau ciri yang sama kedalam beberapa kategori [4]. Secara umum klasifikasi merupakan usaha yang menata pengetahuan atau juga benda yang pada umumnya sistematis. Algoritme klasifikasi yang banyak digunakan secara luas, yaitu *Decision / Classification Trees*, *Bayesian Classifiers / Naive Bayes Classifiers*, *Neural Networks*, *Analisis Statistik*, *Algoritme Genetika*, *Rough Sets*, *K-Nearest Neighbor*, *Metode Rule Based*, *Memory Based Reasoning* dan *Support Vector Machines (SVM)*[2].

K-Nearest Neighbor adalah suatu metode yang menggunakan algoritme *supervised*, dengan tujuan untuk menemukan pola baru dalam data dengan menghubungkan pola data yang sudah ada dengan data yang baru. Sedangkan pada *unsupervised learning*, data belum memiliki pola apapun, dengan tujuan untuk menemukan pola dalam sebuah data[3]. Metode *K-Nearest Neighbor* melakukan klasifikasi berdasarkan jarak terdekat antara *data testing* dengan *data training* berdasarkan jumlah nilai k (k = banyaknya *data training*). Proses menghitung dekat atau jauhnya jarak dapat dihitung dengan menggunakan *Dice Similarity*[2]. *Dice similarity* merupakan metode pengukuran jarak kesamaan antar dua objek dengan

cara mengalikan keterkaitan setiap pasangan vektor (*inner product*) dalam ruang kata [3].

Pada penelitian sebelumnya[5] menggunakan algoritme *K-Nearest Neighbor* dengan menekankan perbandingan metode pengukuran jarak dari *cosine similarity*, *ecludian distance*, *improved scoring scheme*, *manhattan distance* untuk mengklasifikasi dokumen yang terbagi menjadi 2700 dokumen pelatihan dan 90 dokumen testing pada dokumen Bengali. Berdasarkan hasil pengujian ke-4 perbandingan tersebut, didapatkan pengukuran jarak *cosine similarity* dengan hasil yang terbaik yaitu 93,33%.

Pada penelitian lain[6] menggunakan metode *Vector Space Model* dan algoritme *Web Query Classification Algorithm (WQCA)* dengan menggunakan jarak *Dice Similarity* untuk mengklasifikasi *query* yang diinputkan dengan 220 dokumen pelatihan dengan 22 kategori. Berdasarkan hasil pengujian tersebut, didapatkan rata-rata pengujian *F-measure* yaitu 85%.

Dengan demikian, dibuatlah sistem untuk mengklasifikasi Dokumen Pariwisata menggunakan metode algoritme *K-Nearest Neighbor* menggunakan jarak *dice similarity* yang mampu meningkatkan hasil klasifikasi dokumen pariwisata.

1.2 Rumusan Masalah

1.2.1 Permasalahan

Banyaknya sumber informasi pariwisata saat ini membuat penyedia situs masih belum melakukan pengelompokan objek pariwisata indonesia yang sesuai dengan kategori dengan kategori Wisata alam, buatan, kuliner, religi, dan budaya. Dampaknya dokumen pariwisata yang tersebar tidak terklasifikasi dengan benar. Berdasarkan studi kasus dari Dinas Pariwisata Bangkalan ada beberapa pengelompokan kategori wisata, diantaranya alam, buatan, kuliner, religi, dan budaya.

1.2.2 Usulan Solusi

Diusulkan Metode *K-Nearest Neighbor* untuk mengklasifikasi Dokumen Pariwisata dengan jarak menggunakan *dice similarity*.

1.2.3 Pertanyaan Penelitian

1. Berapa waktu dan akurasi yang diperlukan untuk mengklasifikasi dokumen menggunakan *Dice Similarity* ?

1.3 Tujuan dan Manfaat

1.3.1 Tujuan

Penerapan *K-Nearest Neighbor* pada klasifikasi dokumen pariwisata dengan membandingkan metode pengukuran jarak antara *Cosine Similarity* dan *Dice Similarity*.

1.3.2 Manfaat

Dihasilkan aplikasi yang mampu mengklasifikasi dokumen berita pariwisata sesuai dengan kategori sebenarnya.

1.3.3 Batasan Masalah

1. Dokumen yang diuji merupakan dokumen teks sebanyak 200 Dokumen Pariwisata dari situs <http://www.pesona.travel/>.
2. Teks yang digunakan sebagai dokumen uji merupakan dokumen yang menggunakan bahasa Indonesia yang baku.

1.4 Sistematika Proposal

Sistematika penulisan Proposal ini terdiri dari beberapa bab, yaitu :

BAB I PENDAHULUAN

Pada bab ini membahas tentang latar belakang, rumusan masalah, tujuan, manfaat, batasan masalah, sistematika penulisan laporan.

BAB II KAJIAN PUSTAKA

Pada bab ini membahas tentang landasan teori dan penelitian sebelumnya.

BAB III METODE USULAN

Pada bab ini membahas tentang metode yang diusulkan (algoritme), arsitektur(*Input, Proses, Output*), DATA, tahapan penelitian, serta rencana pengujian.

BAB IV HASIL DAN PEMBAHASAN

Pada bab ini membahas mengenai implementasi sistem yang telah dikerjakan berdasarkan hasil dari perancangan pada BAB III.

BAB V PENUTUP

Pada bab ini berisi kesimpulan dari tugas akhir dan saran untuk pengembangan sistem.