

BAB 1 PENDAHULUAN

1.1. Latar Belakang

Perkembangan teknologi di dunia saat ini sudah sangat merubah hidup manusia menjadi lebih mudah dan praktis. Teknologi dapat dikatakan menjadi kebutuhan dalam keseharian manusia untuk melakukan berbagai hal termasuk membaca berita. Perkembangan media berita sangat berkembang pesat seiring dengan berkembangnya teknologi, banyak media berita cetak yang bertransformasi menjadi portal berita online dengan jumlahnya yang tersebar di internet semakin bertambah banyak. Selain itu, banyaknya situs berita yang terdaftar pada *World Wide Web* (WWW) memungkinkan hasil pencarian berita menjadi sangat luas.

Pariwisata di Indonesia merupakan industri yang berkembang pesat sehingga berpotensi meningkatkan perekonomian daerah dan menciptakan lapangan kerja [1]. Pariwisata merupakan sebuah kebutuhan dalam kehidupan masyarakat saat ini dengan tujuan untuk menghibur dan menghilangkan rasa bosan setelah melakukan aktivitas sehari-hari namun, tujuan dari pariwisata semakin berkembang seiring dengan berkembangnya teknologi. Tujuan dari pariwisata saat ini tidak sekedar untuk menghibur dan menghilangkan rasa bosan melainkan juga untuk mengikuti sebuah kegiatan yang sedang populer atau sering dikunjungi banyak orang hanya untuk menambah koleksi foto pada media sosial [2].

Mesin pencarian atau *search engine* di era modern ini sering kali menjadi tempat pertama yang dikunjungi ketika mencari informasi. *Search engine* bekerja dengan menghitung kemiripan *query* dengan dokumen berita pariwisata yang terdapat dalam *database* dan menampilkan berita yang paling relevan, dengan begitu pengguna dapat memilih hasil pencarian yang disajikan oleh *search engine* dan membaca berita yang paling bernilai bagi mereka [3]. SEBI (Search Engine Bahasa Indonesia) merupakan salah satu *search engine* yang mampu mengumpulkan halaman web dari internet, melakukan *preprocessing*, membangun indeks, menangani *query* dengan sempurna serta mengkategorikan dokumen [4].

Terdapat beberapa metode yang dapat digunakan dalam menghitung kemiripan antara *query* dengan data dalam *database* seperti *Cosine*, *Dice*, *Vector Support*

Machine (VSM), dan lainnya. Dari beberapa metode tersebut, *Cosine* merupakan salah satu metode yang terbaik untuk menghitung kemiripan. Namun, dalam penelitian yang dilakukan oleh Lokesh Sahu menunjukkan bahwa *Cosine* memiliki hasil yang kurang optimal dalam memperkecil jarak *intracluster* untuk mengelompokkan dokumen sehingga peneliti tersebut menggunakan *Modified Cosine* untuk memperkecil jarak *intracluster* [5].

Banyak penelitian yang telah dilakukan untuk membuat *search engine* serta mencari kemiripan antara dua dokumen, seperti yang dilakukan oleh Rizki Tri Wahyuni [6] untuk mengklasifikasi dokumen skripsi dengan menerapkan algoritma *Cosine Similarity* dan pembobotan *TF-IDF* menunjukkan kelayakan sistem dengan persentase pengujian 88,3% ketika diuji oleh pakar kearsipan. Penelitian lainnya dilakukan oleh Lokesh Sahu untuk pengelompokan dokumen menggunakan algoritma *K-means* yang ditingkatkan menggunakan *Modified Cosine Distance Measure* menggunakan *Mahout* dengan *Hadoop* menunjukkan bahwa dengan *Modified Cosine* hasil intra-cluster yang diperoleh lebih baik dibandingkan metode *Cosine* yaitu 93% [5]. Karter D. Putung, Arie Lumenta, Agustin Jacobus melakukan penelitian untuk menerapkan metode *cosine* dalam sistem temu kembali informasi pada kumpulan dokumen skripsi dan menunjukkan hasil *recall* bernilai 1 pada setiap pengujian yang menunjukkan semua dokumen yang relevan dapat ditemukan oleh sistem [7].

Dari uraian latar belakang ini, sistem pencari berita pariwisata diperlukan agar memudahkan pembaca untuk mencari berita tentang pariwisata yang paling relevan dengan *query* menggunakan metode *Modified Cosine*.

1.2. Perumusan Masalah

1.2.1. Permasalahan

Dalam mencari berita pariwisata, tidak semua berita yang ditampilkan relevan dengan *query* sehingga perlu dilakukan pengecekan ulang yang memakan banyak waktu sehingga diperlukan sebuah sistem pencarian berita pariwisata yang dapat menampilkan berita pariwisata yang paling relevan dengan *query*.

1.2.2. Metode Usulan

Metode yang diusulkan dalam penelitian ini yaitu *Modified Cosine* karena metode *Modified Cosine* mampu memperkecil jarak antar dokumen dalam *clustering* dokumen dimana dalam mencari kemiripan antara dokumen berita

pariwisata dengan *query*, jika jarak antar dokumen diperkecil maka nilai *precision*, *recall*, dan *f-measure* akan semakin baik.

1.2.3. Pertanyaan Penelitian

- 1) Berapa nilai *precision*, *recall*, dan *f-measure* pada metode *Modified Cosine* dalam menghitung kemiripan dokumen berita pariwisata dengan *query*?

1.3. Tujuan Dan Manfaat Penelitian

1.3.1. Tujuan Penelitian

Mengetahui nilai *precision*, *recall*, dan *f-measure* pada metode *Modified Cosine* dalam menghitung kemiripan *query* dengan dokumen berita pariwisata.

1.3.2. Manfaat Penelitian

Dengan *Software* yang dibuat dalam penelitian ini diharapkan *precision*, *recall*, dan *f-measure* hasil pencarian berita pariwisata menjadi lebih baik sehingga dapat digunakan pada portal berita online.

1.4. Batasan Masalah

1. Dokumen berita pariwisata diambil dari beberapa situs berita online

1.5. Sistematika Proposal

BAB I PENDAHULUAN

Bab ini menjelaskan latar belakang masalah, rumusan masalah, batasan masalah, tujuan dan manfaat penelitian, metode penelitian dan sistematika penulisan laporan.

BAB II KAJIAN PUSTAKA

Pada bab ini menjelaskan mengenai materi yang menjadi dasar dalam penelitian diantaranya, penelitian terkait menghitung kemiripan dokumen, serta algoritma *Modified Cosine*.

BAB III METODE USULAN

Pada bab ini menjelaskan mengenai rancangan sistem yang akan dibuat, tahapan analisa permasalahan, perancangan sistem, dan pengujian sistem.

BAB IV HASIL DAN PEMBAHASAN

Bab ini menjelaskan implementasi dari aplikasi yang dibuat berdasarkan hasil rancangan yang di buat sebelumnya.

BAB V PENUTUP

Bab berikut berisi kesimpulan serta saran terhadap penelitian yang telah dilakukan