

BAB I

PENDAHULUAN

1.1. Latar Belakang

Berita merupakan suatu informasi berdasarkan fakta atau laporan mengenai suatu kejadian yang sedang, atau telah terjadi. Berita dipublikasikan melalui media cetak, siaran, internet, maupun dari mulut ke mulut. kebutuhan informasi di Indonesia dalam bentuk digital baik berupa teks atau dokumen semakin meningkat seiring perkembangan teknologi. Salah satu sumber informasi yang banyak di akses oleh masyarakat adalah portal berita elektronik. Hal ini dikarenakan portal berita elektronik mengklasifikasikan artikel atau berita sesuai dengan kategorinya. Namun, pengklasifikasian berita masih dilakukan secara manual sehingga rentan terjadi kesalahan.

Banyaknya berita yang harus diklasifikasikan berdasarkan kategorinya dalam waktu yang singkat akan membuat editor mengalami kesulitan dan dapat menimbulkan kesalahan dalam pengkategorian. Hal ini dikarenakan editor harus membaca satu persatu berita-berita yang akan diklasifikasikan. Namun apabila pengklasifikasian berita dilakukan secara otomatis akan meminimalisir kesalahan pengkategorian dan lebih efisien. Pada alexa.com site berita detik.com masuk pada 10 top site di Indonesia[1]. Berdasarkan data tersebut kami mengambil berita dari situs detik.com dengan kategori berita olahraga yaitu sport.detik.com.

Penambangan data adalah tentang mencari pola dalam data. Demikian pula, penambangan teks adalah tentang mencari pola dalam teks, itu adalah proses menganalisis teks untuk mengekstrak informasi yang berguna untuk tujuan tertentu[2]. Praproses dalam penambangan data diantaranya adalah *tokenizing*, *case folding*, *stopwords*, dan *stemming*.

Klasifikasi adalah proses menemukan model (atau fungsi) yang menggambarkan dan membedakan kelas atau konsep data, untuk tujuan agar dapat menggunakan model untuk memprediksi kelas objek yang label kelasnya tidak diketahui. Model turunan didasarkan pada analisis satu set data pelatihan (yaitu, objek data yang label kelasnya diketahui)[3] Dari beberapa metode klasifikasi yang ada, salah satunya adalah metode *support vector machine*. *Support Vector Machine (SVM)* adalah sistem pembelajaran yang menggunakan ruang hipotesis berupa fungsi-fungsi linier dalam sebuah ruang fitur (*feature space*) berdimensi tinggi, dilatih

dengan algoritma pembelajaran yang didasarkan pada teori optimasi dengan mengimplementasikan learning bias yang berasal dari teori pembelajaran statistik[4]. SVM dipilih sebagai metode klasifikasi kategori berita karena pada penelitian sebelumnya metode ini berhasil memberikan akurasi tinggi dalam mengklasifikasi berita online berbahasa Indonesia secara otomatis[5].

Dengan uraian yang telah dijelaskan di atas, penelitian ini diharapkan dapat menyelesaikan permasalahan dalam pengklasifikasian berita olahraga berbahasa Indonesia sehingga didapatkan hasil klasifikasi kategori berita yang lebih baik dan akurat.

1.2. Rumusan Masalah

Berdasarkan latar belakang yang dijelaskan di atas, maka penulis merumuskan permasalahan yaitu “Seberapa akurat klasifikasi berita Olahraga berbahasa Indonesia menggunakan Metode *Support Vector Machine*?”

1.3 Tujuan Penelitian

Tujuan dilakukannya penelitian ini adalah untuk mengetahui keakuratan dari Metode *Support Vector Machine* pada klasifikasi berita olahraga bahasa Indonesia.

1.4 Manfaat Penelitian

Manfaat dari penelitian ini yaitu untuk memberikan kemudahan bagi pengguna untuk mengelompokkan berita olahraga sesuai dengan kategori isi berita.

1.5 Batasan Masalah

Dalam penelitian ini untuk membatasi permasalahan agar tetap terarah dan tidak menyimpang dari apa yang sudah direncanakan sebelumnya. Penulis membuat batasan-batasan masalah sebagai berikut:

- a. Berita yang diklasifikasikan adalah berita berbahasa Indonesia.
- b. Berita yang diklasifikasikan adalah berita dengan topik Olahraga.
- c. Berita yang diklasifikasikan hanya berita teks dan gambar tidak termasuk.
- d. Jumlah berita yang digunakan sebanyak 240, dimana terdapat 6 kategori berita yaitu Basket, Moto GP, Formula 1, Raket Sepakbola, Sport lain masing-masing kategori berjumlah 40 berita.

- e. Berita didapatkan dari website berita olahraga yaitu sport.detik.com.
- f. *Stemming* yang digunakan adalah dengan menggunakan Algoritma *Enhanced Confix Stripping Stemmer*.

1.6 Metodologi Penelitian

Dalam proposal penelitian ini terdapat beberapa tahapan yang akan dilakukan dalam perancangan sistem klasifikasi, di antaranya :

a. Studi Pustaka

Pada tahap ini dilakukan pengumpulan informasi terkait klasifikasi. Informasi diperoleh dari berbagai literatur seperti artikel, jurnal, buku dan lain-lain.

b. Analisa dan Perancangan

Analisa sistem diperoleh setelah mempelajari data-data yang telah terkumpul. Selanjutnya, sistem dirancang menurut urutan dari proses analisa data sehingga diperoleh hasil yang diinginkan.

c. Implementasi Sistem

Setelah rancangan sistem selesai dibuat, maka langkah selanjutnya yaitu mengimplementasikan rancangan tersebut menjadi sebuah program. Proses implementasi dalam bentuk aplikasi berbasis web.

d. Uji Coba Sistem

Pada tahap ini dilakukan pengujian sistem klasifikasi berita olahraga atau program yang telah dibuat. Pada klasifikasi berita olahraga terdapat 2 dokumen berita yaitu dokumen *training* dan dokumen *testing*. Dokumen *training* merupakan dokumen yang digunakan sebagai acuan dari dokumen *testing* untuk menghasilkan label kategori.

e. Analisa dan Evaluasi

Proses analisa dan evaluasi dari data pelatihan dilakukan dengan evaluasi *Recall*, *Presicion*, *F1-Score*, dan *Accuracy Score*. Evaluasi tersebut digunakan untuk mendapatkan nilai evaluasi klasifikasi sehingga tingkat akurasi hasil klasifikasi berita dapat diketahui.

f. Dokumentasi

Dokumentasi dilakukan mulai dari tahap studi pustaka hingga tahap evaluasi.

1.7 Sistematika Penulisan

Adapun sistematika pembahasan tugas akhir ini adalah sebagai berikut :

BAB I PENDAHULUAN

Pada bab ini berisikan latar belakang dari penelitian yang akan dilakukan, rumusan masalah, batasan masalah, tujuan dan manfaat penelitian, metodologi penelitian dan sistematika penulisan.

BAB II STUDI PUSTAKA

Pada bab ini dijelaskan mengenai materi-materi yang menjadi landasan dalam penyusunan penelitian ini. Materi-materi tersebut mengenai sistem klasifikasi, *Text Mining* dan *Processing*, *Term Frequency*, dan lain-lain.

BAB III PERANCANGAN SISTEM

Pada bab ini akan menjelaskan rancangan dari seluruh sistem yang akan dibuat, mulai kebutuhan *software* dan *hardware* yang dibutuhkan, rancangan sistem klasifikasi dengan model *Support Vector Machine*, model pengujian serta alur dari sistem yang akan dikembangkan.

BAB IV IMPLEMENTASI SISTEM

Pada bab ini akan dijelaskan implementasi sistem dari rancangan sistem klasifikasi yang telah dibuat pada bab sebelumnya. Kemudian sistem akan diuji coba berdasarkan model dan skenario pengujian yang telah ditentukan.

BAB V PENUTUP

Pada bab ini berisi kesimpulan dari hasil uji coba serta saran yang dapat digunakan untuk pengembangan sistem klasifikasi selanjutnya.