

BAB I

PENDAHULUAN

1.1. Latar Belakang

Berita merupakan suatu informasi berdasarkan fakta yang ada atau laporan mengenai suatu kejadian yang sedang atau telah terjadi yang dipublikasikan melalui media cetak, siaran, internet maupun dari mulut ke mulut. Berita menjadi kebutuhan sehari-hari dalam masyarakat. Dengan adanya berita, masyarakat menjadi lebih tahu mengenai kejadian terkini. Berita olahraga merupakan salah satu berita yang paling banyak diakses oleh masyarakat. Pada alexa.com situs berita olahraga seperti sport.detik.com, bola.net masuk pada *25 top site* di Indonesia. Di era sekarang ini, kebanyakan masyarakat lebih memilih mengakses berita olahraga melalui internet. Berbeda dengan media cetak ataupun siaran, berita melalui internet dapat kapan saja diakses tanpa adanya batas waktu. Akan tetapi seiring dengan berjalannya waktu, jumlah berita olahraga pada internet akan semakin besar. Besarnya jumlah berita membuat berita tersebut harus dilakukan pengelompokan atau kategorisasi sesuai dengan isi dari berita, dengan tujuan yaitu agar berita tersebut terorganisir. Kategorisasi dapat dilakukan secara manual oleh manusia, akan tetapi perlu waktu yang tidak singkat. Sehingga harus menerapkan teknik data mining yaitu klasifikasi dengan tujuan untuk kategorisasi dokumen berita olahraga secara otomatis, yang dapat mempercepat dan mempermudah dalam melakukan kategorisasi sehingga dapat meningkatkan efisiensi waktu.

Data mining adalah suatu proses ekstraksi atau penggalian data yang belum diketahui sebelumnya, namun dapat dipahami dan berguna dari database besar serta digunakan untuk membuat suatu keputusan bisnis yang sangat penting [1]. Sedangkan klasifikasi adalah pengolahan untuk menemukan kumpulan model (atau fungsi) yang menggambarkan dan membedakan kelas data atau konsep, dengan tujuan mampu menggunakan model untuk memprediksi kelas dari objek yang label kelasnya tidak diketahui. Model didasarkan pada analisa kumpulan data latih [2].

Sebelum berita dapat diklasifikasi, berita tersebut harus dilakukan tahap *preprocessing*. *Preprocessing* terdiri dari beberapa tahapan. Adapun tahapan *preprocessing* berdasarkan [3] , yaitu: *case folding*, *tokenizing/ parsing*, *filtering* dan *stemming*. Pada tahapan *preprocessing*



ini, proses *stemming* merupakan proses yang terpenting karena pada *stemming* terjadi penghilangan kata imbuhan sehingga menghasilkan kata dasar. Algoritma yang digunakan pada proses *stemming* ini adalah Algoritma *Enhanced Confix Stripping Stemmer*. Algoritma *Enhanced Confix Stripping Stemmer* merupakan pengembangan dari Algoritma *Confix Stripping* dengan melakukan beberapa perbaikan aturan sehingga memiliki tingkat kesalahan paling sedikit dibandingkan algoritma sebelumnya [4].

Terdapat beberapa metode klasifikasi, salah satunya yaitu *Naïve Bayes Classifier*. *Naïve Bayes Classifier* adalah model probalistik yang simpel untuk klasifikasi data kedalam kelas yang spesifik berdasarkan fitur data yang berbeda [5]. *Naïve Bayes* telah menjadi metode utama yang digunakan pada klasifikasi untuk variasi pada data mulai dari medis, jaringan komputer dan pengakuan teks karena kesederhanaan, keefektifan, dan kapabilitas menangkap penalaran data dalam model grafis [6][7][8]. Kelebihan *Naïve Bayes Classifier* yaitu konsep yang mudah dimengerti, tidak sensitif terhadap fitur yang relevan, dapat menangani data real maupun diskrit [9]. *Naïve Bayes Classifier* adalah metode klasifikasi yang berasumsi bahwa setiap variable X bersifat bebas (*independence*). Dengan kata lain, NBC mengasumsikan bahwa keberadaan sebuah atribut (*variable*) tidak ada kaitannya dengan keberadaan atribut (*variable*) yang lain [10]. Meskipun asumsi bahwa fitur-fitur dataset bersifat independent adalah sebuah asumsi yang kurang baik tetapi kenyataannya hasil klasifikasi berbasis *Naïve Bayes* memiliki kinerja yang mampu berkompetisi dengan metode-metode lain yang lebih kompleks dalam aspek komputasinya [11]. *Naïve Bayes Classifier* memiliki keunggulan dalam hal waktu komputasi dibandingkan dengan algoritma klasifikasi lainnya [12].

Pada penelitian ini akan melakukan klasifikasi berita olahraga menggunakan metode *Naïve Bayes* dengan *Enhanced Confix Stripping Stemmer* sebagai algoritma *stemming*. Dengan harapan yaitu dapat menghasilkan hasil klasifikasi yang lebih akurat.

1.2. Perumusan Masalah

Berdasarkan latar belakang masalah yang telah dipaparkan diatas, maka penulis merumuskan permasalahan yaitu “Seberapa akurat klasifikasi berita olahraga menggunakan Metode *Naïve Bayes* dengan *Enhanced Confix Stripping Stemmer*?”



1.3. Tujuan dan Manfaat

Berdasarkan rumusan masalah di atas maka tujuan dan manfaat yang dapat diambil dari penelitian ini adalah sebagai berikut:

1.3.1. Tujuan

Tujuan dilakukannya penelitian ini adalah untuk mengetahui keakuratan dari Metode *Naïve Bayes* dengan *Enhanced Confix Stripping Stemmer* pada klasifikasi berita olahraga.

1.3.2. Manfaat

Manfaat yang dihasilkan dalam penelitian ini adalah menghasilkan *software* yang dapat melakukan kategorisasi/pengkategorian berita olahraga.

1.4. 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:

1. Berita yang diklasifikasi adalah berita berbahasa Indonesia.
2. Jumlah berita latih yang digunakan sebanyak 151 berita, dimana terdapat 6 kategori berita yaitu Basket, MotoGP, Formula 1, Raket, Sepak Bola, masing-masing kategori berjumlah 25 berita dan berita olahraga lainnya yang berjumlah 26 berita.
3. Berita didapatkan dari *website* berita olahraga yaitu *sport.detik.com*.
4. Berita uji yang digunakan adalah berita olahraga, kecuali berita yang mengandung nama pemain atau nama klub, akan tetapi isi dari berita bukan berita berkategori olahraga.

1.5. Metodologi Penelitian

Secara garis besar tahapan metodologi penelitian yang digunakan dalam penelitian Tugas Akhir ini adalah sebagai berikut :

1. Studi Pustaka
Pada tahap ini dilakukan pengumpulan informasi terkait klasifikasi. Informasi diperoleh dari berbagai literatur seperti artikel, jurnal, buku dan lain-lain.
2. Perancangan Sistem
Pada tahap ini dilakukan perancangan sistem klasifikasi berita olahraga yang dibangun dengan mengacu pada hasil pemahaman dan studi pustaka.



3. Implementasi Sistem
Pada tahap ini dilakukan implementasi dari rancangan sistem klasifikasi berita olahraga yang telah dibuat ke dalam suatu program dalam bentuk aplikasi berbasis web.
4. 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 latih dan dokumen uji. Dokumen latih merupakan dokumen yang digunakan sebagai acuan dari dokumen uji untuk menghasilkan label kategori.
5. Analisa dan Evaluasi
Pada tahap ini dilakukan analisa untuk mengetahui keakuratan klasifikasi berita olahraga yang dihasilkan.
6. Dokumentasi
Pada tahap ini dilakukan penyusunan laporan dan dokumentasi dari pelaksanaan secara keseluruhan, mulai dari tahap awal sampai tahap analisa dan evaluasi.

1.6. Sistematika Penulisan

Secara garis besar, Sistematika penulisan buku dalam Tugas Akhir ini adalah sebagai berikut :

BAB I : PENDAHULUAN

Pada bab ini menjelaskan tentang latar belakang penelitian, perumusan masalah, tujuan, manfaat, batasan masalah, metode penelitian, serta sistematika penulisan dalam laporan tugas akhir ini.

BAB II : STUDI PUSTAKA

Bab ini membahas mengenai landasan teori-teori pokok yang digunakan dalam menyusun penelitian ini. Landasan teori yang dibahas meliputi teori umum dan teori khusus.

BAB III : PERANCANGAN SISTEM

Pada bab ini menjelaskan tentang rancangan sistem dan keseluruhan sistem yang dibuat beserta rancangan alur *flowchart* pada tiap-tiap proses pada sistem.

BAB IV : IMPLEMENTASI SISTEM



Pada bab ini menjelaskan tentang implementasi sistem berdasarkan rancangan pada BAB III. Dan menjelaskan hasil analisa dari sistem yang telah dibuat.

BAB V : PENUTUP

Pada bab penutup ini menjelaskan tentang kesimpulan dari hasil uji coba yang telah dilakukan serta saran yang menjadi masukan dalam pengembangan penelitian berikutnya.



[HALAMAN INI SENGAJA DIKOSONGKAN]

