

BAB I

PENDAHULUAN

1.1 Latar Belakang Masalah

Dalam perkembangannya, media penyaluran informasi masyarakat atau berita yang pada awalnya menyampaikan berita melalui media televisi, surat kabar, majalah atau radio sudah mulai menggunakan sistem berbasis komputer untuk menyampaikan beritanya secara *up to date*. Pada umumnya berita yang disampaikan dalam portal tersebut terdiri dari beberapa kategori seperti berita politik, olahraga, ekonomi, kesehatan, dan lain - lain sebagai contoh: *kompas*, *sindo*, dan *liputan6*. Masalahnya adalah dalam membagi berita ke dalam kategori-kategori tersebut untuk saat ini masih dilakukan secara manual, artinya dalam membagi berita harus terlebih dahulu membaca isi dari berita tersebut secara keseluruhan untuk selanjutnya dimasukkan ke dalam kategori yang tepat. Hal ini sangat merepotkan dalam mengkategorikan berita apabila jumlah berita yang ingin dikategorikan berjumlah banyak. Oleh karena itu, perlu adanya sistem dimana sistem tersebut dapat mengklasifikasikan berita secara otomatis sesuai dengan kategori-kategori berita yang ada sehingga mempermudah dalam proses mengklasifikasikan berita.

Text mining adalah salah satu teknik yang dapat digunakan untuk melakukan klasifikasi dimana, *text mining* merupakan variasi dari *data mining* yang berusaha menemukan pola yang menarik dari sekumpulan data tekstual yang berjumlah besar. Selain klasifikasi, *text mining* juga digunakan untuk menangani masalah *clustering*, *information extraction*, dan *information retrieval* [1]. Klasifikasi adalah proses untuk menemukan model atau fungsi yang menjelaskan atau membedakan konsep atau kelas data, dengan tujuan untuk dapat memperkirakan kelas dari suatu obyek [2]. Oleh karena itu, kelas yang ada tentulah lebih dari satu. Banyak ide telah muncul selama beberapa tahun belakangan tentang teknik *machine learning* untuk permasalahan klasifikasi. Misalnya yang dilakukan oleh Daniela Xhemali yang berkonsentrasi pada perbandingan tiga metode *Naïve Bayes*, *Pohon Keputusan* dan *Neural Networks*. Hasil

penelitian secara keseluruhan menunjukkan bahwa *Naïve Bayes* adalah pilihan terbaik untuk pelatihan domain [3].

Pada penelitian ini, proses yang akan dilakukan meliputi proses *preprocessing* data, proses seleksi fitur dan proses klasifikasi untuk mengetahui kelas dalam sebuah berita. Proses klasifikasi dilakukan dengan menerapkan metode klasifikasi *Naïve Bayes* dan *K-Nearest Neighbor*. *Naïve Bayes* merupakan suatu metode klasifikasi yang menggunakan perhitungan probabilitas. Penentuan kelas dari suatu dokumen dilakukan dengan cara membandingkan nilai probabilitas suatu sampel berada di kelas yang satu dengan nilai probabilitas suatu sampel berada di kelas yang lain [4]. Metode klasifikasi *Naïve Bayes* adalah metode pembelajaran Bayesian yang ditemukan sangat berguna dalam berbagai aplikasi. *Naïve Bayes* merupakan salah satu metode *supervised document classification*. Metode ini dikenal memiliki tingkat akurasi yang tinggi dengan perhitungan sederhana [5]. Algoritma *K-Nearest neighbor* (KNN) adalah sebuah metode untuk melakukan klasifikasi terhadap objek berdasarkan data pembelajaran yang jaraknya paling dekat dengan objek tersebut. KNN termasuk algoritma *supervised learning* dimana hasil dari *query instance* yang baru diklasifikasikan berdasarkan mayoritas dari kategori pada KNN. Nanti kelas yang paling banyak muncul yang akan menjadi kelas hasil klasifikasi [6].

Berdasarkan uraian diatas, maka disusun suatu laporan tugas akhir yang berjudul ***“Klasifikasi Konten Berita Menggunakan Metode Naïve Bayes dan K-Nearest Neighbor”***.

1.2 Rumusan Masalah

Berdasarkan uraian latar belakang yang ada maka dirumuskan masalah yaitu:

1. Bagaimana cara mengkasifikasikan topik berita secara otomatis menggunakan *Naïve Bayes* dan *K-Nearest Neighbor*?
2. Bagaimana mengukur tingkat presisi, *recall* dan *F1-Measure* dengan menggunakan metode *Naïve Bayes* dan *K-Nearest Neighbor* terhadap data uji?

1.3 Batasan Masalah

Agar penelitian yang dilakukan sesuai dengan tujuan yang diinginkan, maka penulis membatasi masalah yang akan dibahas sebagai berikut:

1. Kategori berita yang digunakan hanya 4 kategori yaitu berita autotekno, ekbis, lifestyle dan sport dimana data berita tersebut diambil dari media berita *online*.
2. *Dataset* yang digunakan didapat dari portal berita www.sindo.com pada bulan Juni - Juli 2015.
3. Berita yang digunakan dalam penelitian ini hanya berita berbahasa Indonesia.
4. Nilai K pada metode *K-Nearest Neighbor* yang diujikan adalah 3, 5, 7, 9.

1.4 Tujuan Penelitian

Adapun tujuan penelitian tugas akhir ialah sebagai berikut:

1. Melakukan pengklasifikasian terhadap data uji dengan menggunakan metode *Naïve Bayes* dan *K-Nearest Neighbor*.
2. Melakukan pengujian presisi, *recall* dan *F1-Measure* terhadap data uji dengan menggunakan metode *Naïve Bayes* dan *K-Nearest Neighbor*.

1.5 Sistematika Penelitian

Agar penulisan karya tulis ini dapat tersusun secara teratur, maka diperlukan sistematika. Berikut ini merupakan sistematika penulisan dari tugas akhir:

Bab I Pendahuluan, dalam bab ini menjelaskan mengenai latar belakang masalah, identifikasi masalah, perumusan masalah, batasan masalah, tujuan, metode penelitian dan sistematika penulisan.

Bab II Landasan Teori, dalam bab ini menjelaskan tentang teori-teori yang mendukung dalam mengerjakan tugas akhir.

Bab III Metodologi, dalam bab ini akan dibahas metodologi penelitian yang dilakukan.

Bab IV Analisis dan Perancangan, dalam bab ini akan dipaparkan kebutuhan-kebutuhan ketika melakukan penelitian. Untuk itu pembahasan ini akan terbagi atas analisis dan perancangan.

Bab V Implementasi dan Pengujian, dalam bab ini menjelaskan tentang tahapan implementasi. Pada bagian implementasi akan melakukan pengujian pada metode-metode yang telah dibuat pada perancangan sistem dan melakukan perhitungan *F1-Measure* metode. Bab ini juga membahas tentang hasil dari implementasi dan analisis.

Bab VI Penutup, dalam bab ini merupakan bab yang menguraikan tentang kesimpulan dan saran dari laporan tugas akhir.

