

BAB I PENDAHULUAN

1.1 Latar Belakang

Pada teknologi tatap muka antara manusia dan komputer yang telah diteliti sejak lama. Namun belakangan pada pengenalan informasi lebih diperhatikan, yaitu pengenalan informasi nonverbal, khususnya ada pemrosesan emosi. Beberapa jenis ciri fisik yang umum digunakan dalam pemrosesan emosi diantaranya intonasi suara, ekspresi wajah, sikap badan, gerakan tubuh dan detak jantung. Diyakini oleh peneliti bahwa teknologi emosi merupakan bagian yang penting dalam kecerdasan buatan, khususnya komunikasi antar manusia[DAN-95].

Terdapat tiga media yang biasa digunakan dalam penelitian bidang emosi yaitu teks, suara dan gambar. Dalam penelitian ini lebih focus pada pengenalan emosi dari media teks yaitu berupa kalimat berbahasa Indonesia. Mengapa memilih media teks pada penelitian ini dikarenakan teks merupakan sebuah proses yang dapat berubah secara dinamis. Maka dari itu, penelitian emosi pada teks dilakukan berbentuk teks yang relative sederhana dibandingkan menggunakan media lainnya.

Teks adalah media yang paling sederhana akan tetapi teks juga memiliki peranan penting dalam komunikasi karena sebuah teks tidak hanya mempunyai pesan atau sebuah informasi belaka, namun juga menginformasikan tentang perilaku atau sifat manusia termasuk emosi. Media teks juga lebih mudah diperoleh dari bermacam-macam sumber misalnya pada buku, koran, majalah, situs, serta media elektronik lainnya seperti yang tersedia di internet lainnya. Dari penjelasan tersebut media teks mempunyai berbagai kelebihan, selain teks termasuk media yang populer juga merupakan media yang kaya akan emosi.

Jenis-jenis emosi sendiri ada 5 macam yang telah disepakati oleh para ahli sebagai emosi dasar, yaitu jenis emosi seperti senang, sedih, marah, takut. Jenis emosi tersebut telah dikenal sejak lama dalam media teks dan menjadi aspek penting dari sifat manusia[POW-97].

Namun dalam terapan emosi belum banyak digunakan ke dalam interaksi manusia dan computer. Tetapi emosi cenderung sangat berperan dalam komunikasi antar manusia di kehidupan sehari-hari. Sebagai contoh pada jejaring social yang semakin marak dan digemari para manusia untuk mengekspresikan emosi yang dirasakan ke dalam bentuk teks yaitu update status pada akun facebook atau pada akun twitter mereka. Oleh karena itu sistem interaksi manusia dan computer yang baik harus dapat mengenali, representasikan dan memproses emosi manusia.

Dalam perkembangan teknologi informasi yang semakin pesat, saat ini telah ditemui untuk melakukan pemrosesan teks yang bisa digunakan bermacam-macam metode. Diantaranya *Naïve bayes*, SVM, VCM, KNN, juga KNN-improve. Dari metode-metode tersebut digunakan untuk mengenali emosi pada teks.

Pada penelitian sebelumnya disimpulkan bahwa teknologi yang sudah ada mampu melakukan proses untuk mengenali emosi pada teks. Proses pengenalan emosi tidak dapat dipisahkan dari proses klasifikasi teks yang digunakan untuk menentukan jenis emosi teks tersebut. Setelah diketahui jenis emosi pada teks, kemudian dapat ditentukan proses selanjutnya [RAM-10].

Namun Pada penelitian ini akan dibahas proses klasifikasi emosi dari teks berbahasa Indonesia. Pengklasifikasian teks secara otomatis merupakan bagian dari *data mining* yang bertujuan untuk mendefinisikan kelas dari sebuah objek yang belum diketahui kelasnya. Pada klasifikasi teks, terlebih dahulu akan dilakukan proses *training* dan *testing* yang mana pada proses *training* akan digunakan dataset yang telah diketahui kelas objeknya [SAP-08].

Klasifikasi dan prediksi merupakan suatu teknik analisis data yang bisa digunakan untuk mengekstrak model dari data yang berisi kelas-kelas atau untuk memprediksi trend data yang akan datang. Klasifikasi memprediksi data dalam bentuk kategori, sedangkan prediksi memodelkan fungsi-fungsi dari nilai yang kontinyu. Misalnya model klasifikasi bisa dibuat untuk mengelompokkan beberapa aplikasi peminjaman pada bank apakah beresiko atau aman, sedangkan model prediksi dibuat untuk memprediksi pengeluaran untuk membeli peralatan berdasarkan pendapatan dan lokasi tinggalnya. Prediksi bisa dipandang sebagai pembetulan dan penggunaan model untuk menguji kelas dari sampel yang tidak

berlabel, atau menguji nilai atau rentang nilai dari suatu atribut. Klasifikasi salah satu teknik yang digunakan dalam penelitian ini yaitu untuk menentukan jenis emosi pada sebuah teks hingga diperoleh jenis emosi yang terdapat didalamnya.

Penelitian ini digunakan sebuah algoritma C5.0 untuk proses klasifikasi teks berbahasa Indonesia. Metode yang digunakan dalam penelitian ini merupakan salah satu metode klasifikasi teks yang digunakan untuk mengenali jenis emosi. Metode Algoritma C5.0 ini merupakan pengembangan dari metode sebelumnya yaitu ID3 dan juga metode C4.5. Maka dari itu saya menggunakan metode Algoritma C 5.0 karena merupakan metode yang masih jarang digunakan dalam penelitian klasifikasi teks. Kebanyakan metode ID3 dan juga metode terdahulu lainnya yang menjadi pilihan para peneliti sebelumnya.

Algoritma C5.0 adalah salah satu algoritma yang terdapat dalam klasifikasi data mining disamping algoritma CART, yang khususnya diterapkan pada teknik decision tree. C5.0 penyempurnaan dari algoritma sebelumnya yaitu ID3 dan C4.5. Dalam algoritma C5.0 pemilihan atribut yang diproses menggunakan information gain. Kalau dalam cabang suatu decision tree anggotanya berasal dari satu kelas maka cabang ini disebut pure. Kriteria yang digunakan adalah information gain. Jadi dalam memilih atribut untuk memecah obyek dalam beberapa kelas harus dipilih atribut yang menghasilkan *information gain* paling besar. Karena ukuran *information gain* tersebut digunakan untuk memilih atribut uji pada setiap node di dalam tree. Ukuran ini digunakan untuk memilih atribut atau node pada pohon. Atribut dengan nilai *information gain* tertinggi akan terpilih sebagai parent bagi *node* selanjutnya [KAZ-03].

Selain itu mengapa pada penelitian ini menggunakan Algoritma C5.0 sebagai metode klasifikasi teks berbahasa Indonesia ini karena, algoritma C5.0 tetap dianggap sebagai algoritma yang sangat membantu dalam melakukan klasifikasi data karena karakteristik data dari hasil yang diklasifikasi dapat diperoleh dengan jelas baik dalam bentuk struktur pohon keputusan maupun aturan *if-then*, dengan itu pengguna akan lebih mudah dalam melakukan penggalian informasi terhadap data yang bersangkutan lebih baik dari metode sebelumnya yaitu C4.5.

Berdasarkan latar belakang yang telah dipaparkan, maka judul yang diambil dalam skripsi ini adalah “ **Klasifikasi emosi untuk teks berbahasa Indonesia dengan menggunakan Algoritma C5.0**”.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah dijelaskan maka Rumusan masalah dalam Skripsi ini adalah :

1. Bagaimana penerapan metode C5.0 untuk pengklasifikasian emosi pada dokumen teks berita berbahasa Indonesia?
2. Bagaimana tingkat akurasi dari sistem pengklasifikasian dengan C5.0 pada sejumlah sampel emosi pada teks?

1.3 Tujuan

Beberapa tujuan yang ingin dicapai dalam Skripsi ini adalah:

1. Merancang sebuah sistem yang dapat mengklasifikasikan emosi pada teks bahasa Indonesia menggunakan Algoritma C5.0.
2. Mengetahui tingkat akurasi dalam system pengklasifikasian emosi menggunakan metode C5.0 pada teks berbahasa Indonesia

1.4 Batasan Masalah

Batasan masalah yang diangkat dalam skripsi ini adalah:

1. Klasifikasi hanya dalam teks bahasa Indonesia, tidak dalam bahasa yang lainnya.
2. Dokumen teks yang akan digunakan dalam skripsi ini berupa dokumen dalam format txt.
3. Suatu kata dianggap berdiri sendiri dan tidak terkait dari kata disekitarnya. Untuk hal ini ditujukan untuk penyederhanan proses pengolahan. Jika kata diproses secara berkelompok atau berpasangan tentu saja akan muncul berbagai banyak faktor yang harus diperhitungkan.
4. Untuk proses *stemming* tidak memperhitungkan adanya kata infiks (sisipan). Proses *stemming* yang dibangun hanya melakukan menghilangkan prefiks dan sufiks.

5. Kategori emosi yang digunakan terdiri dari 4 kategori, yaitu : sedih, senang, marah, dan takut

1.5 Manfaat

Manfaat hasil skripsi supaya nanti dapat digunakan untuk mengklasifikasi emosi pada teks berbahasa Indonesia ke dalam beberapa kategori emosi yang telah ditentukan. Sehingga pembaca bisa memilih berita apa yang ingin mereka baca.

1.6 Sistematika Penulisan

Sistematika penulisan skripsi ini adalah sebagai berikut :

BAB I : PENDAHULUAN

Pada bab ini membahas mengenai latar belakang, rumusan masalah, batasan masalah, tujuan, manfaat dan sistematika penulisan skripsi.

BAB II : TINJAUAN PUSTAKA

Bab ini menjelaskan menguraikan dasar teori tentang emosi, *text mining*, dan Algoritma C5.0 serta teori yang terkait dengan topik penulisan Skripsi ini yang diangkat menjadi acuan dasar dalam pembuatan sistem Klaisifikasi teks berbahasa Indonesia menggunakan Algoritma C5.0.

BAB III : METODOLOGI DAN PERANCANGAN

Bab ini menjelaskan mengenai metode-metode yang digunakan dalam pembuatan klasifikasian teks berita berbahasa Indonesia dengan menggunakan Algoritma C5.0

BAB IV : IMPLEMENTASI

Bab ini menjelaskan hasil implementasi yang diterapkan pada sebuah sistem dalam pengklasifikasian emosi pada teks bahasa Indonesia menggunakan C5.0.

BAB V : ANALISA HASIL DAN PEMBAHASAN

Bab ini penjelasan tentang analisa hasil pengujian data yang telah diperoleh.

BAB VI : PENUTUP

Bab ini berisi kesimpulan dari penelitian yang telah dilakukan, dan saran untuk pengembangan penelitian selanjutnya.

