

BAB I

PENDAHULUAN

Pada bab ini akan dibahas mengenai latar belakang masalah, rumusan masalah dan pokok bahasan, tujuan penelitian, manfaat penelitian dan sistematika penulisan skripsi.

1.1 Latar Belakang

Twitter merupakan aplikasi jejaring sosial yang memungkinkan bagi pengguna untuk berbagi informasi dalam bentuk teks-teks pendek (140 karakter) [PHU-10:120]. Informasi-informasi berupa teks pendek yang dikirim oleh pengguna disebut sebagai *tweet* dapat dilihat oleh pengguna lain, sebutan bagi pengguna lain yang membaca *tweet* dari seorang pengguna adalah *follower*. Informasi yang datang secara acak pada pengguna dengan jumlah yang banyak dalam waktu yang singkat membuat pengguna mengalami kewalahan dalam mendapatkan informasi yang dibutuhkannya.

Permasalahan informasi yang datang secara acak pada *tweet* dengan jumlah banyak dapat diminimalisasi dengan melakukan pengkategorian pesan singkat pada kategori-kategori tertentu. Tujuan dari penelitian adalah untuk mengkategorikan *tweet* secara otomatis ke dalam kategori yang berbeda sehingga pengguna tidak kewalahan dengan membaca informasi *tweet* dalam kondisi yang masih acak [SRI-10:841]. Berbagai metode telah diterapkan dan masih terus dikembangkan oleh para peneliti di seluruh dunia. Salah satu metode yang banyak digunakan untuk melakukan pengkategorian dokumen dengan metode klasifikasi adalah *Naive Bayes*. Algoritma *Naive Bayes* tidak rumit dan efektif dalam klasifikasi teks [GUO-10:699].

Naive Bayes Classifier adalah sebuah metode *information retrieval* yang menggunakan pendekatan probabilistik dalam melakukan inferensi, yakni berbasis teorema Bayes secara umum. Metode ini paling banyak digunakan dalam proses klasifikasi teks. Pada algoritma *Naive Bayes* mengasumsikan bahwa probabilitas

kemunculan sebuah kata independen terhadap posisinya di dalam teks. Metode *Naive Bayes* secara teoritis terbagi ke dalam dua proses utama, yaitu tahap pelatihan dokumen dan tahap klasifikasi dokumen. Selama tahap pelatihan program perlu mengekstrak kata-kata yang ada pada dokumen pelatihan untuk setiap kategori. Kemudian frekuensi dan nilai-nilai probabilitasnya dihitung dengan algoritma pelatihan *Naive Bayes classifier*. Pada tahap klasifikasi data uji akan diekstrak kata-kata dari dokumen pengujian lalu diproses dengan algoritma klasifikasi *Naive Bayes classifier* untuk menentukan kategori bagi dokumen tersebut.

Pada pengkategorian teks singkat otomatis pada jejaring social Twitter diberikan beberapa kategori yaitu berita umum, keuangan, olahraga, hiburan, teknologi, dan otomotif yang berbahasa Indonesia. Bahasa Indonesia merupakan bahasa yang mempunyai morfologi yang berbeda dengan bahasa Inggris sehingga dalam proses ekstraksi kata menggunakan metode *stemming* untuk bahasa Indonesia. Pengkombinasian antar metode yang diterapkan pada pengkategorian teks pendek secara otomatis ini diharapkan bisa memberikan kategori yang sesuai dan akurat.

1.2 Rumusan Masalah

Berdasarkan paparan latar belakang yang telah diuraikan, sehingga masalah-masalah yang akan dibahas sebagai berikut:

1. Bagaimana merancang dan mengimplementasikan perangkat lunak yang dapat mengkategorikan pesan teks singkat pada jejaring sosial Twitter secara otomatis menggunakan metode klasifikasi *Naive Bayes*?
2. Bagaimana pengaruh metode *Stemming* bahasa Indonesia dalam pengkategorian pesan teks singkat pada jejaring sosial Twitter?
3. Bagaimana cara kerja metode klasifikasi *Naive Bayes* dalam melakukan pengkategorian pesan teks singkat pada jejaring sosial Twitter?
4. Bagaimana akurasi hasil pengkategorian dari aplikasi apabila dibandingkan dengan pengkategorian secara manual?

1.3 Batasan Masalah

Berdasarkan paparan latar belakang dan rumusan yang telah diuraikan maka penelitian ini mempunyai batasan masalah sebagai berikut:

1. Dokumen yang diproses adalah dokumen pesan singkat pada jejaring sosial *Twitter* berbahasa Indonesia.
2. Kategori yang dipakai adalah berita, keuangan, olahraga, hiburan, teknologi, dan otomotif.
3. Data pelatihan yang digunakan berasal dari dokumen RSS.
4. Metode yang digunakan untuk proses stemming adalah Metode Arifin.
5. Metode yang digunakan untuk pengklasifikasian dokumen adalah metode Naive Bayes.
6. Sistem tidak memperhatikan persamaan kata (sinonim) dalam proses *stemming*.
7. Sistem tidak memperhatikan kesalahan ejaan kata.
8. Pada proses *stemming* tidak terlalu memperhatikan imbuhan sisipan.
9. Pengimplementasian aplikasi pengkategorian pesan singkat *tweet* berbahasa Indonesia menggunakan bahasa pemrograman PHP.
10. Perancangan dan Implementasi pembuatan aplikasi pengkategorian pesan singkat *tweet* berbahasa Indonesia.

1.4 Tujuan Penelitian

Tujuan yang ingin dicapai dalam melakukan penelitian ini yaitu membuat aplikasi pengkategorian pesan singkat *Tweet* berbahasa Indonesia menggunakan metode *Naive Bayes* untuk menghasilkan klasifikasi secara otomatis dan memberikan kategori yang sesuai dan akurat.

1.5 Manfaat Penelitian

Penelitian ini diharapkan mempunyai manfaat yang baik dan berguna bagi pembaca dan penulis, adapun manfaat yang diharapkan adalah sebagai berikut:

a) Bagi Penulis

1. Sebagai media untuk pengimplementasian ilmu pengetahuan teknologi khususnya di bidang *Information Retrieval*.
2. Mendapatkan pengetahuan dan wawasan tentang pengklasifikasian text.

b) Bagi pembaca

1. Mendapatkan wawasan akan pengimplementasian dari *Information Retrieval*.
2. Mendapatkan wawasan akan penerapan metode *Stemming* Arifin.
3. Mendapatkan wawasan akan bagaimana cara kerja metode *Naive Bayes* dalam mengklasifikasikan dokumen.

1.6 Sistematika Pembahasan

Sistematika isi dan penulisan proposal skripsi ini antara lain :

Bab I : Pendahuluan

Berisi tentang latar belakang masalah, perumusan masalah dan pokok-pokok bahasan, tujuan dan manfaat dari penelitian serta sistematika skripsi

Bab II : Tinjauan Pustaka

Berisi tentang karakteristik khusus pesan singkat pada jejaring sosial Twitter, tahapan *text mining*, metode *stemming*, klasifikasi pada *naive bayes*, dan penjelasan mengenai metode untuk pengujian hasil dari ringkasan (*Precision Recall*) .

Bab III : Metodologi Penelitian

Bab ini berisi tentang gambaran umum perancangan sistem pakar, meliputi: *study literatur*, penggunaan dasar teori, analisa sistem perancangan, *engine* yang digunakan, serta bentuk pengujian yang tepat.

Bab IV : Perancangan

Pada bab ini akan dibahas mengenai perancangan sistem yang akan dibuat.

Bab V : Implementasi

Pada bab ini akan dibahas implementasi berdasarkan metodologi dan perancangan

Bab VI : Pengujian

Pada bab ini akan dipaparkan tentang pengujian terhadap hasil ringkasan dan aplikasi.

Bab VII : Kesimpulan dan Saran

Pada bab ini akan dipaparkan kesimpulan dan saran terhadap penelitian ini.

