

BAB I

PENDAHULUAN

1.1 Latar Belakang Masalah

Melesatnya perkembangan teknologi informasi saat ini membuat apapun menjadi mungkin dilakukan. Apalagi saat ini kebutuhan internet seakan sudah menjadi kebutuhan utama bagi banyak kalangan. Dunia perdagangan pun mengalami perkembangan yang serupa. Tak heran bila kini banyak masyarakat lebih memilih berbelanja *online* dibandingkan dengan cara konvensional. Keuntungan yang memudahkan pembeli seperti keuntungan waktu, jarak, dan pembayaran juga turut mendorong pembeli semakin menyukai belanja *online*.

Saat ini jumlah *e-commerce* pun semakin banyak, dan konsumen dapat memilih *website* belanja yang terpercaya dan menjadi favorit mereka. Karena sifatnya *online* maka barang yang akan dibeli pun tak bisa dilihat langsung ataupun dicoba terlebih dahulu. Pembeli *online* lebih banyak menggunakan *review* produk tersebut untuk lebih yakin ketika akan melakukan pembelian [5]. Hal ini dikarenakan suatu *review* merupakan tulisan berupa opini tentang pengalaman pemakaian terhadap produk tersebut. Pembeli tersebut tentu akan membeli produk tersebut jika *review* tentang produk tersebut lebih banyak berkomentar positif, dan sebaliknya, pembeli akan ragu untuk melakukan pembelian jika *review* tersebut lebih banyak berkomentar negatif [6]. Akan tetapi, ternyata *review* yang beredar di *e-commerce* ini tidak semuanya dapat dipercaya [7]. *Review* yang tidak dapat dipercaya ini disebut *opinion spam* [6]. *Review spammer* akan memberi komentar positif untuk membantu penjualan produk mereka atau memberi komentar negatif untuk menjatuhkan pesaing produk mereka [6]. Tentu hal ini akan sangat merugikan pembeli yang benar-benar percaya terhadap *review* tersebut. Karena hal itulah banyak peneliti saat ini meneliti tentang bagaimana mendeteksi *fake review* dan karakteristiknya.

Pendeteksian *fake review* ini berbeda dengan pendeteksian *opinion spam* lainnya, *reviews on brand only* dan *non-review*. Karena dua tipe *spam* ini dapat dikenali dengan kasat mata dan dapat dilakukan pelabelan dengan mudah. Berbeda dengan *fake review*, *spam* tipe ini sulit dikenali dan pelabelannya pun harus membutuhkan usaha yang lebih keras.

Pada penelitian oleh [6], pelabelan yang digunakan adalah dengan menggunakan teknik *duplicate and near duplicate*. Namun teknik ini dinilai tidak tepat sasaran [7,16], karena *review* yang duplikat belum tentu *spam* dan sebaliknya. Pada tugas akhir ini, penulis akan mencoba melakukan pelabelan secara manual seperti yang dilakukan oleh [7] karena pelabelan dengan cara ini akan melabelkan *review* dengan lebih tepat karena dibantu dengan informasi membedakan *review spam*. Selain itu penelitian yang dilakukan oleh [5,6,7] tidak menggunakan tidak menggunakan NLP dan sumber lexicon dalam memproses *feature*, dalam Tugas Akhir ini akan dicoba menggunakan kedua hal tersebut untuk menghasilkan hasil *feature* yang lebih tepat. Penggunaan kamus juga akan membantu menghasilkan *sentiment feature* yang lebih baik, hal ini tidak dilakukan oleh peneliti sebelumnya [5,6,7]. Kemudian penulis akan melakukan

proses klasifikasi dengan menggunakan metode *supervised* yaitu *naïve Bayes*. Metode *naïve Bayes* dipilih karena penelitian yang dilakukan oleh Fangtao Li mendapatkan hasil terbaik dibandingkan dengan metode lain dalam proses identifikasi *review spam* [7]. Pada penelitian ini penulis akan memilih *library tagger* yang digunakan dalam sistem, dan akan dianalisis hasilnya apakah penggunaan *tagger* akan mempengaruhi sistem. Selain itu penggunaan kamus untuk membantu dalam proses ekstraksi *sentiment feature* juga akan dianalisis hasilnya dan dilihat seberapa pengaruhnya kamus dalam akurasi sistem. Kemudian pemilihan kelas *feature* juga akan dianalisis pada penelitian ini, dan akan dilihat kelas *feature* yang paling berpengaruh dalam sistem.

1.2 Perumusan Masalah

Berdasarkan latar belakang sebelumnya, maka terdapat beberapa rumusan permasalahan yang dapat diangkat dalam tugas akhir ini antara lain :

1. *Feature* apa yang paling berpengaruh pada proses pengklasifikasian manipulasi *review*?
2. Bagaimana menganalisis dan mengklasifikasikan *review* yang dimanipulasi dengan menggunakan *Naïve Bayes*?
3. Bagaimana tingkat akurasi pengklasifikasian manipulasi *review* menggunakan *Naïve Bayes*?

1.3 Batasan Masalah

Batasan masalah yang digunakan dalam penelitian tugas akhir ini adalah sebagai berikut :

1. Dataset yang digunakan merupakan data set berbahasa Inggris.
2. Dataset tidak mengandung link dan data gambar.
3. Dataset merupakan kalimat sempurna yang tulisannya tidak disingkat.
4. Tidak mengatasi kata yang memiliki arti ganda dan berbeda pada suatu kalimat.
5. Pelabelan data *review* yang ditandai sebagai spam dilakukan secara manual.
6. Tidak menangani data *review* yang panjangnya lebih dari 1500 karakter.

1.4 Tujuan

Tujuan dari tugas akhir ini adalah sebagai berikut :

1. Mengetahui *feature* yang berpengaruh pada proses pengklasifikasian manipulasi *review*.
2. Menganalisis dan mengklasifikasikan *review* yang dimanipulasi dengan menggunakan *Naïve Bayes*.
3. Mengetahui tingkat akurasi pengklasifikasian manipulasi *review* menggunakan *Naïve Bayes*.

1.5 Metodologi Penelitian

Metodologi penelitian yang digunakan dalam penyelesaian tugas akhir ini adalah sebagai berikut :

1. Studi Literatur
Pada tahap ini akan dilakukan pencarian referensi teori yang relevan atau sesuai dengan permasalahan yang ditemukan seperti teori tentang Manipulasi *Review*, ekstraksi *feature*, metode klasifikasi Naïve Bayes.
2. Perancangan Sistem
Pada tahap ini akan dilakukan analisis perancangan sistem yang diperlukan dalam membangun sistem, seperti sistem *preprocessing* data, proses ekstraksi *feature*, proses klasifikasi *Naïve Bayes*, dan perhitungan akurasi.
3. Implementasi
Pada tahap ini dilakukan penerapan implementasi sistem berdasarkan rancangan yang telah diuraikan sebelumnya.
4. Pengujian sistem dan Analisis hasil pengujian
Pada tahap ini akan dilakukan pengujian sistem yang telah dibuat dengan beberapa skenario pengujian sistem untuk mengetahui pengaruh penggunaan *library tagger* yang digunakan, mengetahui pengaruh penggunaan kamus, mengetahui *feature* apa saja yang berpengaruh pada pengklasifikasian manipulasi *review*, dan mengetahui pengaruh kombinasi *feature* yang digunakan. Kemudian hasil yang diperoleh akan dianalisis berdasarkan nilai *precision*, *recall*, dan *F-score* serta akurasinya.
5. Kesimpulan dan penyusunan laporan
Pada tahap ini akan dilakukan penarikan kesimpulan terhadap hasil pengujian dan analisis hasil pengujian. Setelah itu penulis akan melakukan proses penyusunan laporan.