

BAB 1

PENDAHULUAN

1.1. Latar Belakang Masalah

Twitter adalah jejaring sosial yang hingga kini masih digemari netizen untuk berbagi dalam mengomentari suatu kejadian bahkan produk yang dibeli maupun dilihatnya. Banyak produk yang akhirnya memiliki akun twitter yang terverifikasi (akun twitter yang telah resmi dimiliki perusahaan atau produk). Kebutuhan praktis sebagai pelapak di suatu toko online seharusnya memang harus selalu up to date dengan barang-barang terkini sesuai dengan jenis yang dijual pelapak. Beraneka ragam forum online yang ada di internet tentunya menjawab hal tersebut namun banyaknya forum online tentunya sangat sulit dipantau perubahan dan perkembangan review di dalam masing-masing forum.

Lalu penulis mengambil nilai praktis menjadikan data twitter sebagai data yang mewakili netizen dalam menilai suatu produk menggunakan text mining. *Text mining* adalah salah satu teknik yang dapat digunakan untuk melakukan klasifikasi dokumen dimana *text mining* merupakan variasi dari data mining yang berusaha menemukan pola yang menarik dari sekumpulan data tekstual yang berjumlah besar[1].

Sentiment analysis atau *opinion mining* adalah studi komputasional dari opini-opini orang, sentimen dan emosi melalui entitas dan atribut yang dimiliki yang diekspresikan dalam bentuk teks[2]. Lantas sentimen analisis ini akan mengetahui pendapat yang dikemukakan dalam kalimat atau dokumen tersebut

apakah bersifat positif, negatif atau netral [3]. Yang nantinya akan dikelompokkan polaritas dari teks yang ada dalam kalimat atau dokumen pada analisis sentimen terdapat 3 bagian yaitu crawling (pengambilan data), selanjutnya tahap training dan testing yang menggunakan algoritma untuk pengklasifikasiannya.

Data yang digunakan dalam penelitian ini adalah data real time yaitu data yang telah dituliskan oleh pengguna twitter itu sendiri yang ditentukan kurun waktunya.

Dari tweet banyak ditemukan kata-kata tidak baku misalnya "cuakep", "muantap", "kece" oleh karena itu preprocessing ini diperlukan.

Data twitter yang telah diambil akan dipreprocessing terlebih dahulu untuk menghilangkan eleme-elemen yang tidak diperlukan yang pada akhirnya akan menghasilkan nilai optimal saat dilakukan klasifikasi pada data-data twitter tersebut.

Oleh karena itu diperlukan sebuah proses untuk mengorganisasikan kata-kata yang tidak baku yang memiliki banyak variasi namun variasi tersebut merujuk kepada kata-kata yang baku maka pengkoreksian terhadap kata-kata tersebut dapat dilakukan dengan algoritma N-gram.

N gram merupakan sebuah metode yang diaplikasikan untuk pembangkitan kata atau karakter [4].

Berdasarkan pemaparan di atas maka dibuatlah sebuah aplikasi untuk mengatasi masalah-masalah tersebut menggunakan text mining yang bernama Sentimen Analisis Produk berbasis web menggunakan bahasa pemrograman php.

1.2. Rumusan Masalah

Dengan penjabaran dari latar belakang di atas dapat diumuskan ke dalam rumusan masalah yaitu:

1. Apakah metode n-gram untuk mengambil fitur pada tweet serta mengorganisasikan kata tidak baku dengan variasi kata yang termuat di dalam suatu tweet sudah bisa diterapkan?
2. Apakah kebutuhan praktis seorang penjual sudah bisa diatasi dengan review pada suatu produk?

1.3. Batasan Masalah

Agar penelitian ini optimal maka dirumuskan suatu batasan masalah antara lain:

1. Pengambilan data berupa twitter ini berbahasa Indonesia.
2. Tahap *preprocessing* yang dilakukan yaitu *convert emoticon, cleansing, case folding, tokenizing, convert negation, remove stopword* dan *stemming*
3. Proses *stopword* dan *stemming* hanya berlaku jika menggunakan bahasa Indonesia.
4. Klasifikasi *tweet* ke dalam kelas sentimen positif dan negatif.
5. Menggunakan metode *Naive Bayes* untuk klasifikasi *tweet*.
6. Menggunakan metode n-gram karakter untuk pengambilan fitur untuk seleksi kata.
7. Penelitian sentimen analisis ini tidak menggunakan thesaurus (similaritas) pada kata.
8. Penelitian ini menggunakan data Bahasa Indonesia.

9. Penelitian ini menggunakan metode Naive Bayes Classifier dan tidak ada membandingkan dengan algoritma lainnya.
10. Aplikasi web ini menggunakan bahasa pemrograman php dan database phpmyadmin

1.4. Maksud dan Tujuan

Maksud dari penelitian yang dilakukan adalah menerapkan metode n-gram untuk optimasi akurasi analisis sentimen suatu produk pada Twitter. Tujuan yang ingin dicapai adalah mengetahui pengaruh penerapan metode n-gram terhadap tingkat optimalnya akurasi yang akan digunakan untuk analisis sentimen pada Twitter.

1.5. Manfaat Penelitian

Manfaat penelitian ini antara lain:

1. Mengklasifikasikan sentimen pada twitter dengan data yang berjumlah besar secara otomatis.
2. Mencari informasi tentang penilaian netizen terhadap suatu produk.
3. Membantu melakukan survey baik dan buruknya suatu produk sebelum menjual ataupun membeli suatu produk.

1.6. Metode Penelitian

Metodologi penelitian yang akan dilakukan pada penelitian ini yaitu sebagai berikut:

1. Pengumpulan Data

Mengumpulkan data dengan cara mencari referensi jurnal dan diktat yang berkaitan dengan topik dari penelitian ini. Pengumpulan data juga yaitu mengumpulkan data *tweet* dengan cara mengambil langsung dari (*crawling data*) Twitter menggunakan API (*application interface*) pada Twitter.

2. Analisis dan Perancangan Model Aplikasi

Melakukan beberapa analisis dalam penelitian ini antara lain adalah analisis masalah yang diteliti, analisis aplikasi yang akan dibuat, analisis tahap *preprocessing* pada data yang telah diambil, analisis klasifikasi dengan melakukan penghitungan manual dari metode yang digunakan, serta analisis kebutuhan fungsional dari aplikasi yang akan dibangun. Lantas, peneliti membuat model aplikasi yang sebelumnya telah dianalisis kebutuhan fungsionalnya agar sesuai dengan metode yang diterapkan.

3. Pembangunan Aplikasi

Aplikasi yang sudah dijadikan modelnya, selanjutnya diimplementasikan. Tahap implementasi ini dilakukan bertahap, dilakukan per-modul dari mulai dari tahap *preprocessing*, tahap *training*, tahap *testing*, sampai ke tahap pengujian hasil akurasi *testing*. Dalam setiap modulnya akan dilakukan pengujian untuk mengetahui bila terjadi *error* pada aplikasi.

1.7. Sistematika Penulisan

Sistematika penulisan dibagi menjadi lima bab yaitu:

BAB I Pendahuluan

Bab ini berisikan konsep dasar untuk penyusunan skripsi

latar belakang masalah, rumusan masalah, batasan masalah, maksud dan tujuan, manfaat penelitian serta metode penelitian.

BAB II Dasar Teori

pada bab ini membahas beberapa teori yang mendukung pembahasan pada bab selanjutnya yaitu tinjauan pustaka, sentimen analisis, text mining, text preprocessing, cleansing, case folding, tokenizing.

BAB III Analisis dan Perancangan Perangkat Lunak

Pada bab ini membahas analisis pieces permasalahan dalam pembuatan aplikasi perangkat lunak dan menjelaskan tentang rancangan struktur menggunakan diagram flowchart program serta merancang interface dari perangkat lunak yang akan dibuat.

BAB IV Implementasi dan Pengujian Perangkat Lunak

Bab ini membahas implementasi dari perangkat lunak yang akan dibuat. Berisikan gambaran *user interface* yang akan dibuat. Selain itu, juga akan dilakukan pengujian untuk mengetahui apakah perangkat lunak yang dibuat berhasil dijalankan sesuai harapan atau tidak, serta untuk menemukan kesalahan-kesalahan (*error*) pada perangkat lunak.

BAB V Kesimpulan dan Saran

Bab terakhir ini berisi kesimpulan dan saran yang diharapkan dapat bermanfaat untuk penelitian selanjutnya.