

Analisis Sentimen Pada Data *Twitter* dengan Menggunakan *Text Mining* terhadap Suatu Produk

Eka Retnawiyati¹, Fatoni², Edi Surya Negara³

¹) Mahasiswa Informatika Universitas Bina Darma

²) Dosen Ilmu Komputer ³) Dosen Ilmu Komputer. Jl Jend. A. Yani No.12 Plaju,
Palembang 30264

Email: eka.retnawiyati@gmail.com¹, fatoni@binadarma.ac.id²),
edisuryanegara@binadarma.ac.id³)

Abstrak. Perkembangan teknologi informasi yang semakin meningkat memberi dampak pada pertukaran informasi dan komunikasi yang semakin mudah. Hal ini ditandai dengan munculnya media sosial seperti *Twitter*, *Facebook*, *Yahoo*, *Google*, *Youtube*, *Instagram*, *Path*. Pertumbuhan media sosial ini juga mendorong adanya informasi tekstual yang besar sehingga muncul kebutuhan penyajian data yang memudahkan pengguna mendapatkan informasi yang akurat. Pengguna *twitter* yang semakin meningkat ini terlihat dari jutaan *tweets* yang di *posting* setiap harinya dengan berbagai topik yang berbeda. Data *tweets* dapat berupa persepsi publik baik ekonomi, perilaku sosial, fenomena alam, perdagangan yang terjadi diseluruh dunia. Data yang besar tersebut bisa dijadikan sebagai objek penelitian. Penelitian ini mencoba melakukan analisis sentimen data dengan mengklasifikasi data *twitter* berbahasa Indonesia pada suatu produk. Data tersebut akan diproses dengan *text mining* untuk menghindari data yang kurang sempurna kemudian data *tweet* diklasifikasi menjadi klasifikasi positif, negatif, netral. Klasifikasi ini menggunakan algoritma *Naïve Bayes Classifier*.

1. Pendahuluan

1.1. Latar Belakang

Perkembangan teknologi informasi yang semakin meningkat memberi dampak pada pertukaran informasi dan komunikasi yang semakin mudah. Hal ini ditandai dengan munculnya media sosial seperti *Twitter*, *Facebook*, *Yahoo*, *Google*, *Youtube*, *Instagram*, *Path*. Pertumbuhan media sosial ini juga mendorong adanya informasi tekstual yang besar sehingga muncul kebutuhan penyajian data yang memudahkan pengguna mendapatkan informasi yang akurat.

Media sosial *twitter* merupakan salah satu media komunikasi populer saat ini. Hal ini terlihat dari peningkatan pengguna *twitter* yang tercatat di seluruh dunia. Berdasarkan data yang dirilis infografis (tahun 2014-2015) *Twitter* memiliki 302 juta pengguna aktif yang 80 persennya berasal dari perangkat mobile. Dari angka itu, 37 persen pengguna *Twitter* berusia 18-29 tahun, sedangkan 25 persen lainnya berada di rentan usia 30-49 tahun[2]. Dengan jumlah pengguna aktif sebanyak itu, *Twitter*

menerima kicauan sebanyak 500 juta setiap harinya. Sebanyak 68 persen berupa kicauan balasan, 26 persen berupa kicauan, dan 6 persen adalah kicauan ulang [2].

Pengguna *twitter* yang semakin meningkat ini terlihat dari jutaan *tweets* yang di *posting* setiap harinya dengan berbagai topik yang berbeda. Data *tweets* ini dapat berupa persepsi publik baik ekonomi, perilaku sosial, fenomena alam, perdagangan yang terjadi diseluruh dunia [1]. Hasil penelitian mengenai peristiwa jatuhnya pesawat AirAsia QZ8501 yang terjadi tanggal 28 Desember 2014 di Selat Karimata memperlihatkan dari negara mana saja yang ikut mengamati peristiwa jatuhnya pesawat tersebut berdasarkan bahasa yang digunakan pada saat mengunggah *tweets*, mengidentifikasi tipe pengguna *twitter*, rentang usia pengguna serta jenis kelamin [1].

Berdasarkan penelitian yang telah ada sebelumnya, penelitian ini mencoba melakukan analisis sentimen data dengan mengklasifikasi data *twitter* berbahasa Indonesia pada suatu produk. Data tersebut akan diproses dengan *text mining* untuk menghindari data yang kurang sempurna kemudian mengklasifikasi data *tweet* ke dalam tiga klasifikasi yaitu klasifikasi positif, negatif, netral. Klasifikasi ini menggunakan algoritma *Naïve Bayes Classifier*. adapun rumusan permasalahan dalam penelitian ini “Bagaimana Menganalisis dan mengklasifikasi Sentimen Pada Data *Twitter* Menggunakan *Text Mining* terhadap Suatu Produk.?”

2. Analisa dan Perancangan

2.1 Analisis Sistem

Analisis sistem yang digunakan untuk menyelesaikan masalah pada penelitian ini terdiri dari beberapa langkah – langkah yaitu:

1. Analisis masalah
2. Pengumpulan data
3. Proses *Text Mining*
4. Penerapan metode Klasifikasi *naïve Bayes Classifier*

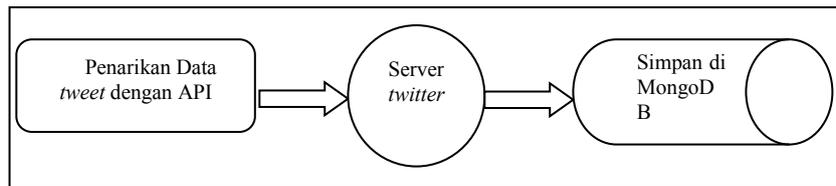
2.1.1. Analisis Masalah

Informasi yang terkandung dalam *tweets* dapat digunakan sebagai nilai jual bagi pelaku usaha. Salah satu contoh perusahaan yang ingin mengetahui nilai jual produknya laku dipasaran atau tidak. Kita sebut saja perusahaan yang memproduksi kosmetik seperti *lipstick* wardah, sabun cuci seperti Rinso. Perusahaan ini akan mencari opini dari pengguna produk tersebut apakah produknya di nilai positif atau negatif. Hal ini dilakukan untuk mengambil langkah strategis agar produknya tetap diminati masyarakat.

Penilaian produk pada suatu perusahaan disebut dengan analisis sentimen. Analisis sentiment yang ada sekarang dilakukan secara manual. Oleh karena itu untuk mendapatkan nilai sentiment secara otomatis dari suatu produk tersebut memiliki tantangan. Tantangan berupa teks yang tidak formal yang digunakan oleh pengguna *twitter* dalam mengeluarkan opini. Untuk itu dari penelitian ini menggunakan proses *text mining* yang di dalamnya terdapat tahapan *preprocessing* dan Ekstraksi fitur serta *naïve bayes classifier* sebagai metode klasifikasi nilai sentimen apakah opini bernilai positif atau negatif.

2.1.2. Pengumpulan Data

Pada penelitian ini pengumpulan data dimulai dengan penarikan data *tweet* dari *server twitter* yang kemudian disimpan kedalam MongoDB. Penarikan data *tweet* dilakukan dengan menggunakan fasilitas *Application Programming Interface (API)*. *API* adalah fasilitas yang telah disediakan oleh *twitter*. *Application Programming Interface (API)* ini mengambil data kotor *tweet* dari *server twitter*, yang selanjutnya akan dipilih menjadi data *tweet* bersih. Data *tweet* ini akan disimpan di dalam database MongoDB. Berikut adalah gambar proses pengumpulan data :



Gambar 1. Proses pengumpulan data *tweet*

2.1.3. Proses *Text Mining*

Pada proses *text mining* membutuhkan 2 (dua) tahapan yaitu :

1. *Preprocessing* ini dilakukan untuk menghindari data yang kurang sempurna, gangguan pada data, dan data-data yang tidak konsisten[3]

Tahapan pada *text preprocessing* yang dilakukan adalah:

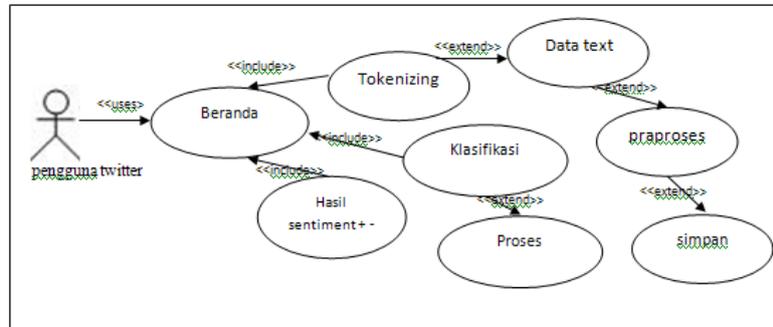
- a. Melakukan *Cleaning* menghilangkan kata yang tidak diperlukan agar dokumen bersih.
- b. *Case folding* dengan mengubah semua huruf dalam dokumen menjadi huruf kecil. Hanya huruf “a” sampai dengan “z” yang diterima. Karakter selain huruf dihilangkan dan dianggap delimiter[4].
- c. Tahap tokenizing / parsing adalah tahap pemotongan string input berdasarkan tiap kata yang menyusunnya.
- d. *Filtering* adalah tahap mengambil kata-kata penting dari hasil token. Bisa menggunakan algoritma *stoplist* (membuang kata yang kurang penting) atau *wordlist* (menyimpan kata penting). *Stoplist/stopword* adalah kata-kata yang tidak deskriptif yang dapat dibuang dalam pendekatan *bag-of-words*. Contoh *stopwords* adalah “yang”, “dan”, “di”, “dari”, dan seterusnya[4].
- e. Ekstraksi Fitur

2.1.4. Metode Klasifikasi

Klasifikasi pada penelitian ini menggunakan *Naïve Bayes Classifier*. Pada tahapan ini akan dilakukan klasifikasi dengan menggunakan algoritma *naïve bayes classifier*. Algoritma ini akan menghitung nilai sentimen positif, negatif atau netral.

2.2. Perancangan

Pada analisis sentimen ini terdiri dari 1 aktor yaitu pengguna *twitter*. Berikut gambaran diagram *use case* dari analisis sentiment data *twitter* :



Gambar 2. diagram *Use Case*

Gambar di atas menunjukkan alur kerja analisis sentiment penelitian ini dengan penjelasan sebagai berikut :

- Halaman Utama : merupakan proses untuk masuk ke tahapan selanjutnya.
- *Tokenizing* : merupakan proses yang terdiri dari praproses yang didalamnya ada *cleaning* dan *case folding* untuk membersihkan data *tweet*. Data *tweet* bersih akan dikategorikan secara manual sebagai data latih yang selanjutnya disimpan di MongoDB.
- Klasifikasi : merupakan proses pengklasifikasian *twitter* dengan data *tweet* yang belum diketahui kategorinya dimana proses ini melewati tahapan *preprocessing* dan proses *naïve bayes classifier* sehingga menghasilkan kategori sentimen positif, negatif ataupun netral.

3. Hasil dan Pembahasan

3.1. Hasil

3.1.1. Implementasi sistem

Setelah dilakukan analisa dan perancangan sistem maka akan dilakukan implementasi sistem analisa sentiment data *twitter* dengan menggunakan *text mining* pada suatu produk. Adapun spesifikasi perangkat keras dan lunak yang digunakan untuk mendukung penelitian ini adalah sebagai berikut :

1. Satu unit laptop dengan spesifikasi sebagai berikut :
 - Prosesor Core i5 3337U 1.8 GHz, 14" HD Color Shine, RAM 4 GB DDR3, Harddisk 500 GB, VGA Nvidia Geforce GT720M 2 GB, DVDRW, Wi-Fi, OS Windows 7.
2. Software yang digunakan adalah sebagai berikut :
 - NetBeans IDE 7.4, *Mongo DB*, *Java*.

Setelah Aplikasi dijalankan akan tampil Halaman beranda yang berisi menu pilihan untuk masuk ke form tokenizing dan form klasifikasi.

3.1.2. Pengujian

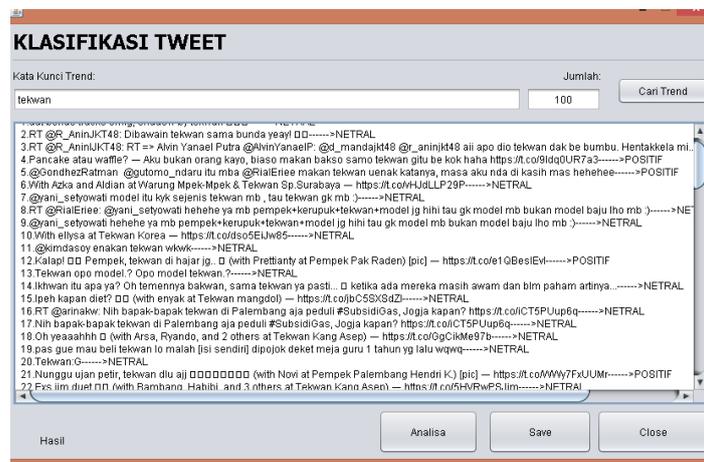
Pada tahap pengujian sistem akan mencari nilai klasifikasi terhadap data *tweet* secara otomatis. Nilai yang diperoleh berupa sentimen positif, sentiment negative atau netral pada setiap *tweet* yang ditampilkan. Pengujian ini menggunakan form klasifikasi dengan cara memasukkan trend dan jumlah data yang akan di analisa. Sebagai sampel data uji akan menggunakan produk makanan yaitu : tekwan dengan

jumlah 100 *tweet*. Berikut tampilan hasil cari trend mengenai tekwan sebelum dianalisa pada gambar 3 :



Gambar 3. Tampilan hasil mencari *tweet*

Setelah dapat 100 *tweet* kita klik tombol analisa untuk mencari nilai dari setiap *tweet* dengan tampilan hasil sebagai berikut :



Gambar 4. Tampilan hasil klasifikasi *tweet*

4. Kesimpulan

Dari hasil penelitian terhadap analisis sentimen data *twitter* dengan menggunakan *text mining* pada suatu produk dapat disimpulkan sebagai berikut :

1. Algoritma *Naïve Bayes* sebagai algoritma untuk mencari nilai probabilitas dari setiap *tweet* yang di analisa.

2. Dari hasil uji analisis sentimen pada data *twitter* didapat *tweet* dengan tiga kategori yaitu kategori positif, negatif dan netral.
3. Analisis sentiment dengan menggunakan *text mining* yang didalamnya terdapat tahapan *preprocessing* dan ekstraksi fitur untuk menghasilkan *tweet* kotor menjadi *tweet* bersih.
4. Semakin banyak data latih atau data yang ditoken saat *training* maka nilai sentimen yang didapat akan semakin akurat.
5. Penelitian ini bisa digunakan untuk mencari minat masyarakat terhadap suatu produk.

Referensi

1. Cahyanti, O.D., Saksono, P.H., Suryayusra, Negara, E.S., *Social Media Analytics Pemanfaatan Data Media Sosial Untuk Penelitian*, Palembang (2015).
2. Grafelly, Delvit Bagaimana perkembangan Twitter saat ini?. Diakses 10 Desember 2015, dari <http://www.techno.id/social/bagaimana-perkembangan-twitter-saat-ini-1509122.html>.
3. Hemalatha, I., Varma, P.G., dan Govardhan, A., 2012, Preprocessing the Informal Text for Efficient Sentiment Analysis, *International Journal of Emerging Trends & Technology in Computer Science (IJETTCS)*, Vol.1, July – August 2012, ISSN 2278-6856
4. Triawati, Chandra, *Metode Pembobotan Statistical Concept Based untuk Klustering dan Kategorisasi Dokumen Berbahasa Indonesia*, Institut Teknologi Telkom, Bandung (2009).