

ANALISIS SENTIMEN BERDASARKAN OPINI MASYARAKAT PADA TWITTER MENGGUNAKAN NAÏVE BAYES

Teguh Ansyor Lorosae¹⁾, Burhanudin Dwi Prakoso²⁾,
Saifudin³⁾, Kusrini⁴⁾

Magister Teknik Informatika, Universitas AMIKOM Yogyakarta
Jl. Ring Road Utar, Condong Catur, Depok, Sleman, Yogyakarta 55281
Email : ansyorlorosae95@gmail.com¹⁾, burhanudin29@hotmail.com²⁾, pex.syaifudin@gmail.com³⁾,
kusrini@amikom.ac.id⁴⁾

Abstrak

Analisis sentimen pada media sosial terkhusus twitter dalam pelayanan yang disediakan oleh penyedia jasa ekspedisi barang merupakan proses untuk mengelompokkan respon seseorang terhadap pelayanan pengiriman barang yang sedang atau telah dilakukan. Permasalahan yang terjadi dalam penelitian ini terdapat kesulitan untuk menentukan opini yang bersifat positif, negatif ataupun netral. Penelitian ini bertujuan untuk mengembangkan aplikasi yang mampu melakukan analisis sentimen konsumen, dimana Naive Bayes Classifier digunakan sebagai metode klasifikasi.

Terdapat penelitian terdahulu menghasilkan informasi sentimen masyarakat yang mengarah ke sentimen positif dan negatif mengenai asuransi dengan menggunakan metode Naive Bayes Classifier dengan menghasilkan akurasi hingga 95%. Sedangkan penelitian yang akan dilakukan yaitu membuat sistem yang mampu mengklasifikasikan sentimen pada twitter kedalam sentimen positif, netral atau negatif. Hasil yang didapatkan dari akurasi naive bayes pada data uji positif memperoleh ketepatan 84%.

Kata kunci: Analisis Sentimen, Naive Bayes, Opini, Twitter.

1. Pendahuluan

Analisis sentimen merupakan proses memahami, mengekstrak dan mengolah data tekstual secara otomatis untuk mendapatkan informasi sentimen yang terkandung dalam suatu kalimat opini. Analisis sentimen dilakukan untuk melihat pendapat atau kecenderungan opini terhadap sebuah masalah atau objek oleh seseorang, apakah cenderung berpandangan atau beropini negatif atau positif. Salah satu contoh penggunaan analisis sentimen dalam dunia nyata adalah identifikasi kecenderungan pasar dan opini pasar terhadap suatu objek barang. Besarnya pengaruh dan manfaat dari analisis sentimen menyebabkan penelitian dan aplikasi berbasis analisis sentimen berkembang pesat. Bahkan di Amerika terdapat sekitar 20-30 perusahaan yang memfokuskan pada layanan analisis sentiment.

Kemajuan teknologi informasi sudah semakin berkembang pesat disegala bidang kehidupan. Sehingga organisasi yang ada memiliki ketergantungan terhadap teknologi informasi, dengan fitur-fitur yang begitu lengkap, membuat organisasi tidak mungkin tanpa menggunakan teknologi informasi. Semua organisasi untuk mengakselerasi tugas-tugasnya untuk mencapai tujuannya, maka organisasi akan bergantung pada teknologi informasi.

Dengan demikian teknologi informasi sangat erat kaitannya dengan organisasi yang ada. Demikian pula dengan organisasi yang bergerak dalam bidang jasa ekspedisi. Sudah tentu teknologi informasi berperan penting, dimana setiap barang yang akan dikirim harus dikontrol berbasis teknologi informasi, mulai dari barang diserahkan oleh pengirim ke petugas ekspedisi, kemudian di catat, sampai barang sampai ke alamat tujuan. Setiap jasa pengiriman barang tentu memiliki berbagai macam cara dalam hal pelayanan terhadap konsumen. Salah satunya yaitu dengan memberikan pelayanan berupa konsultasi, tanya jawab seputar jasa pengiriman tersebut dengan memanfaatkan teknologi informasi berupa media sosial, diantaranya media sosial Twitter.

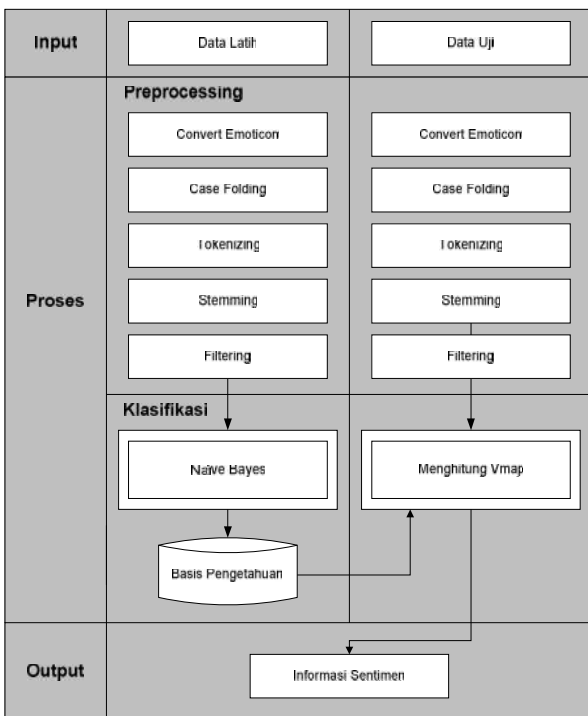
Terkait dengan kebutuhan, keinginan pengguna jasa ekspedisi, untuk memotret opini masyarakat bukan hal yang mudah untuk dilakukan. Hal itu akan melibatkan banyak aspek, salah satunya jejaring sosial twitter. Respon masyarakat sangat baik mengenai pelayanan yang diberikan pihak jasa pengiriman melalui media sosial twitter. Banyak opini-opini masyarakat yang disampaikan melalui tweet pada media sosial tersebut. Dari uraian opini masyarakat terkait dengan pelayanan yang diberikan, pihak jasa pengiriman belum dapat sepenuhnya melihat secara utuh terhadap pelayanan yang diberikan. Dibutuhkan sebuah sistem yang dapat mengambil informasi opini masyarakat menjadi suatu kesimpulan sentimen.

Metode naive bayes digunakan dalam hal klasifikasi. NBC bekerja dengan baik dalam mengklasifikasi, dan teknik yang sederhana hanya dengan menggunakan kemunculan & frekuensi fitur pada tiap-tiap opini[6]. Membandingkan beberapa metode seperti metode naive bayes, k-nearest neighbor, dan gabungan k-mean dan LVQ dalam

mengklasifikasikan kategori buku berbahasa Indonesia berdasarkan judul dan synopsis. Data yang digunakan berjumlah 200 buku, 150 data digunakan sebagai data latih, sedangkan 50 data digunakan sebagai data uji. Dari hasil penelitian yang dilakukan, menghasilkan akurasi pada setiap metode yang digunakan. Untuk metode KNN menghasilkan akurasi sebesar 96% , kemudian naïve bayes sebesar 98% , lalu untuk kombinasi k-Means dan LVQ menghasilkan akurasi sebesar 92,2%. Dihasilkan akurasi tertinggi menggunakan naïve bayes sebesar 95% [2]. Pada penelitian lainnya mengenai analisa sentiment pada opini produk dengan metode naïve bayes classifier. Data yang digunakan sebanyak 6000 data yang terdiri dari 2000 kalimat netral, 2000 kalimat positif dan 2000 kalimat negatif dengan menghasilkan akurasi antara 70,43% - 88,03%. Tingkat keakuratan dari pengklasifikasian sangat dipengaruhi oleh proses [10]. Kompleksitas data juga dapat mempengaruhi tingkat akurasi [11].

2. Pembahasan

Pada penelitian ini terdiri dari beberapa tahapan, diantaranya tahap identifikasi data masukan, pra-proses, klasifikasi dengan metode Naïve Bayes, dan keluaran. Data masukan berupa data latih dan data uji yang digunakan bersumber dari fanpage akun twitter perusahaan jasa ekspedisi TIKI berupa opini-opini masyarakat mengenai pelayanan jasa tersebut. Pada tahap pra-proses terdiri dari convert emoticon, case folding, tokenizing, stemming serta filtering. Data yang telah diproses pada pra-proses diproses kembali dengan menggunakan metode naïve bayes. Keluaran dari proses tersebut yaitu informasi sentimen. Penjelasan tersebut terdapat pada Gambar 1 di bawah ini :



Gambar 1. Rancangan Sistem

a. Convert Emoticon, Case Folding, Stemming, Filtering

Tahap convert emoticon ialah setiap emoticon akan dikonversikan ke dalam string yang bersesuaian, case folding merupakan tahapan untuk merubah seluruh huruf kapital yang terdapat pada dokumen menjadi huruf kecil. Stemming adalah tahapan untuk membuat kata yang berimbuhan kembali ke bentuk asalnya. Tahap filtering merupakan tahapan mengambil kata-kata yang penting (word list) dari hasil proses sebelumnya.

b. Metode Naïve Bayes

Naïve bayes classifier adalah penggolongan menggunakan statistik sederhana berdasarkan teorema bayes yang mengasumsikan bahwa keberadaan atau ketiadaan dari suatu kelas dengan fitur lainnya. Sifatnya yang sebagai model probabilitas, naïve bayes classifier dapat dilatih dengan efisien sebagai supervised learning . Pada metode ini, terdapat beberapa tahap yang dilakukan diantaranya pembentukan fitur (kata yang memiliki nilai sentimen), menghitung probabilitas setiap kelas sentimen, menghitung probabilitas setiap fitur, menghitung Vmap setiap kelas, menghitung Vmap maksimum. Model matematis naïve bayes classifier :

$$P(C | F_1, \dots, F_n) = (p(C)p(F_1, \dots, F_n|C)) / (p(F_1, \dots, F_n)) \dots (1)$$

Setelah dilakukan pembentukan fitur, dilanjutkan dengan menghitung probabilitas dari setiap kelas, dengan persamaan sebagai berikut :

$$p(c_i) = (fd(c_i) + 1) / (|D|) \dots (2)$$

Keterangan :

- fd(c_i) = jumlah dokumen yang termasuk kelas c_i
- |D| = jumlah data latih

Setelah probabilitas setiap kelas didapatkan, selanjutnya menghitung probabilitas setiap fitur pada kelas sentimen dengan cara sebagai berikut :

$$p(w_k|c_i) = (f(w_k, c_i) + 1) / (f(c_i) + |W|) \dots (3)$$

Keterangan :

- f(w_k, c_i) = nilai kemunculan kata w_k pada kelas c_i
- f(c_i) = jumlah keseluruhan kemunculan kata pada kelas c_i
- |W| = jumlah keseluruhan dari w_k

2.1 Hasil

Berdasarkan dari hasil analisis dan pengujian yang telah dilakukan dengan 100 data uji, penggunaan metode Naïve Bayes Classifier (NBC) dengan pre-processing yang digunakan menghasilkan rata-rata akurasi 84%. Dengan menggunakan Naïve Bayes Classifier (NBC) pada sistem ini dapat dihasilkan klasifikasi sentimen positif dan negatif. Contoh hasil yang didapatkan dari penelitian ini dapat dilihat pada Tabel 1 dibawah ini

Tabel 1. Hasil Klasifikasi Analisis Sentimen Tiki

No	Teks Ulasan	Hasil Klasifikasi Sentimen
1	Paket sudah sampai dengan baik, Terimakasih min	Positif
2	tiki telat lagi ? ke jogja belum nyampe juga, udah 3 hari	Negatif

Setelah melalui tahap pra-proses yang terdiri dari convert emoticon, case folding, tokenizing, stemming dan filtering, dilakukan perhitungan menggunakan metode Naïve Bayes Classifier dengan menghitung nilai probabilitas teks berdasarkan basis pengetahuan yang dimiliki. Terdapat dua contoh teks yang akan diklasifikasi seperti yang dapat dilihat pada Tabel 2 dibawah ini :

Tabel 2. Contoh Data Yang Digunakan

No	Komentar Opini
1	Paket sudah sampai dengan baik, Terimakasih min
2	tiki telat lagi ? ke jogja belum nyampe juga, udah 3 hari

Data latih yang digunakan akan dilakukan tahap preprocessing seperti case folding, tokenizing, stemming, filtering dengan hasil sebagai berikut ini:

Tabel 3. Hasil Preprocessing

Paket	Sudah	Baik	Terimakasih	Tiki
Telat	Belum	sudah		

Dari hasil preprocessing, dilakukan pmebentukan fitur, yaitu menentukan kata yang memiliki nilai sentimen, seperti berikut:

Tabel 4. Pembentukan Fitur

Data	Fitur (Kemunculan)	Kelas Sentimen
D1	Paket(1), sudah(1), baik(1), terimakasih(1)	Netral, positif, positif, positif
D2	Tiki(1), telat(1), belum(1), sudah(1)	Netral, negatif, negatif, positif

Tabel 5. Perhitungan Nilai Probabilitas Setiap Kelas

Kelas Sentimen	Data(j)		Fd(Cj)	P(Cj)
	D1	D2		
Positif	1	1	2	2/2
Negatif	0	1	1	1/2
Netral	1	1	2	2/2

Tabel 6. Perhitungan Nilai Probabilitas Setiap Fitur

Data f(wk, Ci)	Kelas Sentimen (c)		
	Positif	Negatif	Netral
Paket	$\frac{1}{4} + \frac{1}{8} = \frac{2}{12} = \frac{1}{6}$	$\frac{0}{2} + \frac{1}{8} = \frac{1}{10}$	$\frac{0}{2} + \frac{1}{8} = \frac{1}{10}$
Sudah	$\frac{1}{4} + \frac{1}{8} = \frac{2}{12} = \frac{1}{6}$	$\frac{0}{2} + \frac{1}{8} = \frac{1}{10}$	$\frac{0}{2} + \frac{1}{8} = \frac{1}{10}$
Baik	$\frac{1}{4} + \frac{1}{8} = \frac{2}{12} = \frac{1}{6}$	$\frac{0}{2} + \frac{1}{8} = \frac{1}{10}$	$\frac{0}{2} + \frac{1}{8} = \frac{1}{10}$
Terimakasih	$\frac{1}{4} + \frac{1}{8} = \frac{2}{12} = \frac{1}{6}$	$\frac{0}{2} + \frac{1}{8} = \frac{1}{10}$	$\frac{0}{2} + \frac{1}{8} = \frac{1}{10}$
Tiki	$\frac{0}{4} + \frac{1}{8} = \frac{1}{12}$	$\frac{0}{2} + \frac{1}{8} = \frac{1}{10}$	$\frac{1}{2} + \frac{1}{8} = \frac{5}{8}$
Telat	$\frac{0}{4} + \frac{1}{8} = \frac{1}{12}$	$\frac{1}{2} + \frac{1}{8} = \frac{5}{8}$	$\frac{0}{2} + \frac{1}{8} = \frac{1}{10}$
Belum	$\frac{0}{4} + \frac{1}{8} = \frac{1}{12}$	$\frac{1}{2} + \frac{1}{8} = \frac{5}{8}$	$\frac{0}{2} + \frac{1}{8} = \frac{1}{10}$
Sudah	$\frac{1}{4} + \frac{1}{8} = \frac{2}{12} = \frac{1}{6}$	$\frac{0}{2} + \frac{1}{8} = \frac{1}{10}$	$\frac{0}{2} + \frac{1}{8} = \frac{1}{10}$

Tabel 7. Data Uji

No	Data Uji
1	Paket saya ko lama belum sampai, udah telat 1 hari nih padahal pakai SDS

Tabel 8. Hasil Preprocessing Data Uji

Paket	Telat	Belum	Paket	Telat
Lama	Sds			

Menghitung Vmap pada data uji untuk menentukan probabilitas tertinggi dari masing-masing kelas.

$$V_{map} = p(wk | c) \times p(c) \dots \dots \dots (4)$$

a. Vmap untuk sentimen positif
 $V_{map}(\text{positif}) = P(\text{positif})P(\text{paket}|\text{positif})P(\text{belum}|\text{positif})P(\text{telat}|\text{positif})$

$$V_{map}(\text{positif}) = 2/2 \times 1/6 \times 1/12 \times 1/12 = 0,001157407$$

b. Vmap untuk sentimen pnegatif
 $V_{map}(\text{negatif}) = P(\text{negatif})P(\text{paket}|\text{negatif})P(\text{belum}|\text{negatif})P(\text{telat}|\text{negatif})$

$$V_{map}(\text{negatif}) = 1/2 \times 1/10 \times 1/5 \times 1/5 = 0,005$$

c. Vmap untuk sentimen netral
 $V_{map}(\text{netral}) = P(\text{netral})P(\text{paket}|\text{netral})P(\text{belum}|\text{netral})P(\text{telat}|\text{netral})$

$$V_{map}(\text{netral}) = 2/2 \times 1/10 \times 1/10 \times 1/10 = 0,001$$

Dari perhitungan tersebut dapat disimpulkan kategori dari data uji termasuk kedalam kategori sentimen negatif, karena memiliki probabilitas tertinggi yaitu 0,05.

Hasil pengujian tingkat akurasi yang dilakukan dengan menggunakan 100 data dengan metode K-fold cross validation, dapat dilihat pada table dibawah ini :

7	10	9	90%
8	10	9	90%
9	10	10	100%
10	10	7	70%
Rata-rata			84%

Setelah selesai tahap perhitungan probabilitas fitur, dilakukan hal yang sama pada data uji. Data uji dilakukan tahap preprocessing, lalu dihitung Vmap maksimum sebagai berikut :

2.2 Pengukuran Akurasi

Pengukuran akurasi dilakukan dengan menggunakan K-Flod Cross Validation. Sekenario pengujian yaitu sebanyak tiga kali dengan jumlah data uji yang berbeda beda, untuk pengujian pertama yaitu dengan menggunakan data sebanyak 50 data, pengujian kedua dengan 75 data, dan pengujian ketiga dengan 100 data, setiap pengujian dibagi menjadi beberapa blok dengan tiap blok menggunakan 10 data. Hasil dari pengujian tersebut ditunjukkan pada table dibawah ini :

Tabel 9. Pengujian 1

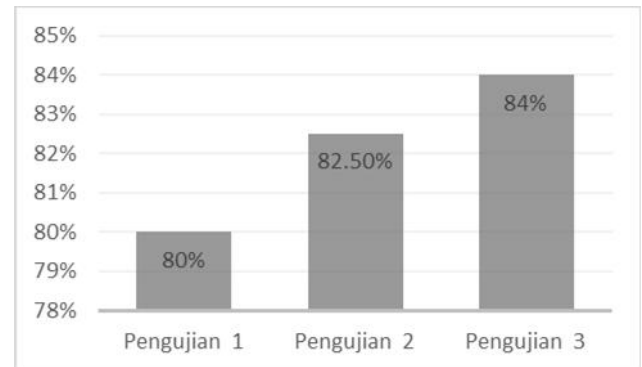
No	Jumlah Data Yang diuji	Data yang cocok	Hasil
1	10	9	90%
2	10	7	70%
3	10	8	80%
4	10	8	80%
5	10	8	80%
Rata-rata			80%

Tabel 10. Pengujian 2

No	Jumlah Data Yang diuji	Data yang cocok	Hasil
1	10	9	90%
2	10	7	70%
3	10	8	80%
4	10	8	80%
5	10	8	80%
6	10	9	90%
7	10	9	90%
8	5	4	80%
Rata-rata			82.50%

Tabel 11. Pengujian 3

No	Jumlah Data Yang diuji	Data yang cocok	Hasil
1	10	9	90%
2	10	7	70%
3	10	8	80%
4	10	8	80%
5	10	8	80%
6	10	9	90%



Gambar 2. Grafik Hasil Pengujian

3. Kesimpulan

Berdasarkan dari hasil penelitian yang telah dilakukan, dapat disimpulkan bahwa penelitian ini telah menghasilkan sebuah sistem analisis sentimen jasa pengiriman barang menggunakan naïve bayes. Sistem yang dibangun telah mampu mentransformasi sentimen yang berupa teks opini pada fanpage Tiki sehingga dapat menampilkan informasi sentimen masyarakat terhadap jasa pengiriman barang yang bersifat positif, negatif dan netral. Penggunaan metode naïve bayes classifier pada penelitian ini tidak menjamin ketepatan dalam proses klasifikasi. Akurasi pada sistem yang dibangun mencapai 84%. Tingkat akurasi akan semakin baik bila data latih yang digunakan lebih banyak, jumlah kategori pada setiap kelas sentimen pun akan berpengaruh.

Daftar Pustaka

- [1] Surya, H., Mohammad, A., Mochammad, H., 2017, Aplikasi Web Untuk Analisis Sentimen Pada Opini Produk Dengan Metode Naive Bayes Classifier, ITN Malang, 4 Februari.
- [2] Luthfia, O., Yulison, H.C., Rezki, Y., 2016, Text Mining Dalam Analisis Sentimen Asuransi Menggunakan Metode Naive Bayes Classifier, SNST.
- [3] Imam, F.R., Sholeh, H.P., Erfan, A.D, 2012, Implementasi Opinion Mining (Analisis Sentimen) untuk Ekstraksi Data Opini Publik pada Perguruan Tinggi, Jurnal EECCIS, No.1, Vol. 6.
- [4] Prananda, A., Rizal, S.P., Ali, F., 2017, Analisis Sentimen Tentang Opini Film Pada Dokumen Twitter Berbahasa Indonesia Menggunakan Naive Bayes Dengan Perbaikan Kata Tidak Baku, Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer, No. 12, Vol. 1, hlm. 1733-1741.
- [5] Amir, H., Naniek, W., 2015, Document Subjectivity and Target detection in Opinion Mining Using HMM POS-Tagger, ICTS.
- [6] Ling, J., Eka N, I.P., Bagus Oka, T., 2014, Analisis Sentimen Menggunakan Metode Naive Bayes Classifier Dengan Seleksi Fitur Chi Square, E-Jurnal Matematika, Agustus, Vol. 3, pp. 92-99.

- [7] Dinda, A.M, 2017, Analisis Sentimen Pada Review Restoran Dengan Teks Bahasa Indonesia Menggunakan Algoritma Naïve Bayes, E-ISSN, No. 2, Vol. 2.
- [8] Faisol, N., Sari, W.S., Afrizal, D., 2013, Analisis Sentimen Masyarakat terhadap Calon Presiden Indonesia 2014 berdasarkan Opini dari Twitter Menggunakan Metode Naïve Bayes Classifier, Jurnal ITSMART, No 2, Vol 2.
- [9] Falahah, Dyar, D.A.N., 2015, Pengembangan Aplikasi Sentimen Analysis Menggunakan Metode Naïve Bayes (Studi Kasus Sentiment Analysis dari media Twitter), SESINDO.
- [10] Rakhmat, S.A., Mukaharil, B.A., 2010, Analisis Sentimen Terhadap Acara Televisi Berdasarkan opini Publik, Jurnal Ilmiah Komputer dan Informatika.
- [11] Septian, N.Y., 2014, Data Mining Menggunakan Algoritma Naïve Bayes Untuk Klasifikasi Kelulusan Mahasiswa Universitas Dian Nuswantoro, Jurnal Ilmiah, September.

Biodata Penulis

Teguh Ansyor Lorosae, memperoleh gelar Sarjana Komputer (S.Kom), Jurusan Sistem Informasi STMIK Amikom Yogyakarta, lulus tahun 2016. Sedang menempuh pendidikan Magister Teknik Informatika di Universitas AMIKOM Yogyakarta.

Burhanudin Dwi Prakoso, memperoleh gelar Sarjana Komputer (S.Kom), Jurusan Teknik Informatika STMIK AMIKOM Yogyakarta, lulus tahun 2016.

Saifudin, memperoleh gelar Sarjana Komputer (S.Kom), Jurusan Sistem Informasi STMIK AMIKOM Yogyakarta, lulus tahun 2016.

Kusrini, memperoleh gelar Sarjana Komputer (S.Kom), Ilmu Komputer Universitas Gadjah Mada, lulus tahun 2002. memperoleh gelar Magister Komputer (M.Kom), Ilmu Komputer Universitas Gadjah Mada, lulus tahun 2006. memperoleh gelar Doktor (Dr), Ilmu Komputer Universitas Gadjah Mada, lulus tahun 2010.

