

ANALISIS SENTIMEN JASA EKSPEDISI BARANG MENGUNAKAN METODE NAÏVE BAYES

Raflialdy Raksanagara*, Yulison Herry Chrisnanto, Asep Id Hadiana

Jurusan Informatika, Fakultas Matematika dan Ilmu Pengetahuan Alam

Universitas Jenderal Achmad Yani

Jalan Terusan Jenderal Sudirman PO BOX 148 Cimahi - Bandung, Telp./Fax: (022) 6656190

*e-mail: raflialdy18@gmail.com

Abstrak

Analisis sentimen pada twitter dalam jasa ekspedisi barang merupakan proses mengekstraksi pendapat, sentimen, evaluasi, dan emosi orang tentang pelayanan pengiriman barang yang tertulis. Permasalahan yang terjadi dalam penelitian ini terdapat kesulitan untuk menentukan opini yang bersifat positif, negatif ataupun netral, serta hasil analisis sentimen disalurkan ke setiap pelayanan yang tersedia. Terdapat penelitian terdahulu mengenai analisis sentimen pada facebook dan twitter menggunakan metode naïve bayes yang menggunakan data sebanyak 6000 data yang terdiri dari 2000 kalimat netral, 2000 kalimat positif dan 2000 kalimat negatif dengan menghasilkan akurasi hingga 85,44%. Sedangkan penelitian yang akan dilakukan yaitu membuat sistem yang mampu mengklasifikasikan sentimen pada twitter kedalam sentimen positif, netral atau negatif serta menyalurkan opini tersebut ke setiap bagian pelayanan yang bersangkutan dengan opini yang muncul. Hasil yang didapatkan dari akurasi naïve bayes pada data uji negative memperoleh ketepatan 76% sedangkan untuk data uji positif memperoleh 83%.

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

1. PENDAHULUAN

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.

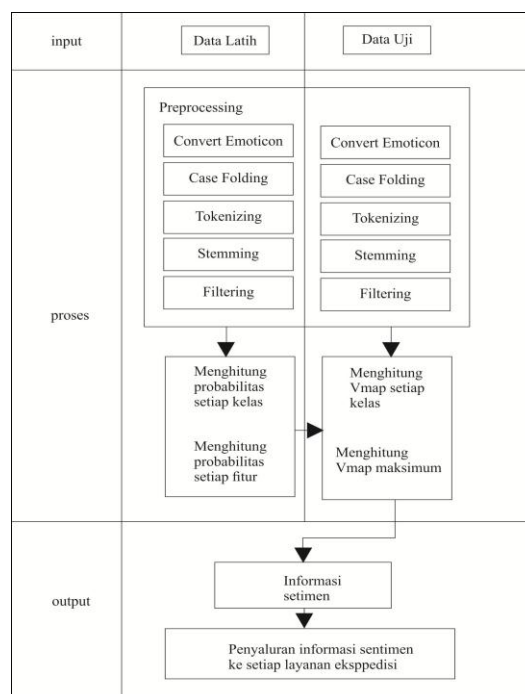
Terkait dengan kebutuhan, keinginan pengguna jasa ekspedisi, untuk memotret opini masyarakat bukan hal yang mudah untuk dilakukan. Hal itu akan melibatkan banyak aspek, salah satunya jejaring sosial twitter. Respon masyarakat sangat baik mengenai pelayanan yang diberikan pihak jasa pengiriman melalui 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 naïve 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 (Ling, Eka & Bagus Oka, 2014). Membandingkan beberapa metode seperti metode naïve 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 98% (Darujati, 2010). Pada penelitian lainnya mengenai analisa sentiment menggunakan mengenai masyarakat pengguna telepon genggam menggunakan metode naïve bayes. Data yang digunakan sebanyak 200 opini, terdiri dari 100 opini positif dan 100 opini negative. 100 opini digunakan sebagai data latih dan 100 opini digunakan sebagai data uji. Hasil yang didapatkan akurasi naïve bayes sebesar 83% dan rata-rata harmonic sebesar 90,713% (Ganisaputra & Tan, 2014). Tingkat keakurasian dari pengklasifikasian sangat dipengaruhi oleh proses training (Rakhmat & Mukaharil, 2010). Kompleksitas data juga dapat mempengaruhi tingkat akurasi (Septian, 2014).

2. METODOLOGI

Metodologi pada penelitian ini terdiri dari beberapa tahap, diawali dengan identifikasi data input atau masukan kemudian dilakukan tahap preprocessing , dilakukan klasifikasi data masukan menggunakan metode *Naïve bayes* lalu hasil klaksifikasi di sortir berdasarkan tipe yang sudah ditentukan. Data masukan yang digunakan didapatkan dari akun twitter perusahaan jasa ekspedisi JNE berupa opini-opini masyarakat mengenai pelayanan jasa tersebut. Untuk mendapatkan hasil sentimen dari data masukan, dapat dilihat pada gambar 1.



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 (*wordlist*) 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* :

$$P(C|F_1, \dots, F_n) = \frac{p(C)p(F_1, \dots, F_n|C)}{p(F_1, \dots, F_n)} \quad (1)$$

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

$$p(c_i) = \frac{fd(c_i)}{|D|} \quad (2)$$

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 :

$$p(wk|c_i) = \frac{f(wk, c_i) + 1}{f(c_i) + |W|} \quad (3)$$

Keterangan :

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

3. HASIL DAN PEMBAHASAN

3.1. Hasil

Berdasarkan hasil dari pengujian yang dilakukan dengan 1500 data uji menggunakan metode *naïve bayes*, terbukti mampu menghasilkan klasifikasi sentimen positif, negatif dan netral. Contoh hasil dari penelitian dapat dilihat pada tabel 1.

Tabel 1. Contoh Data yang digunakan

No	Opini Masyarakat pada Twitter
1	Paket sudah sampai dengan baik, Terimakasih min
2	jne 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 2. Hasil preprocessing

Paket	Sudah	Baik	Terimakasih	Jne
Telat	Belum	sudah		

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

Tabel 3. Pembentukan Fitur

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

Dilanjutkan dengan menghitung probabilitas setiap kelas, dengan cara sebagai berikut:

Tabel 4. 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

Menentukan probabilitas setiap fitur pada kelas sentimen, dengan cara sebagai berikut :

Tabel 5. Perhitungan Nilai Probabilitas setiap fitur

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

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

Tabel 6. Data uji

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

Tabel 7. Hasil preprocessing data uji

Paket	Telat	Belum
Lama	yes	

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

$$Vmap = p(wk|c)xp(c) \tag{4}$$

- a. Vmap untuk sentimen positif

$$V_{\text{map}}(\text{positif}) = P(\text{positif})P(\text{paket}|\text{positif})P(\text{belum}|\text{positif})P(\text{telat}|\text{positif})$$

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

b. V_{map} untuk sentimen negatif

$$V_{\text{map}}(\text{negatif}) = P(\text{negatif})P(\text{paket}|\text{negatif})P(\text{belum}|\text{negatif})P(\text{telat}|\text{negatif})$$

$$V_{\text{map}}(\text{negatif}) = \frac{1}{2} \times \frac{1}{10} \times \frac{1}{5} \times \frac{1}{5} = 0,05$$

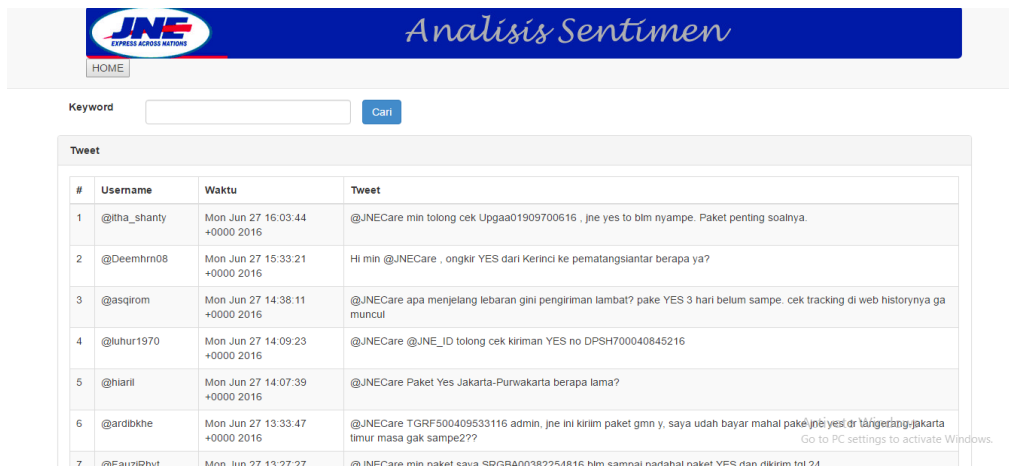
c. V_{map} untuk sentimen netral

$$V_{\text{map}}(\text{netral}) = P(\text{netral})P(\text{paket}|\text{netral})P(\text{belum}|\text{netral})P(\text{telat}|\text{netral})$$

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

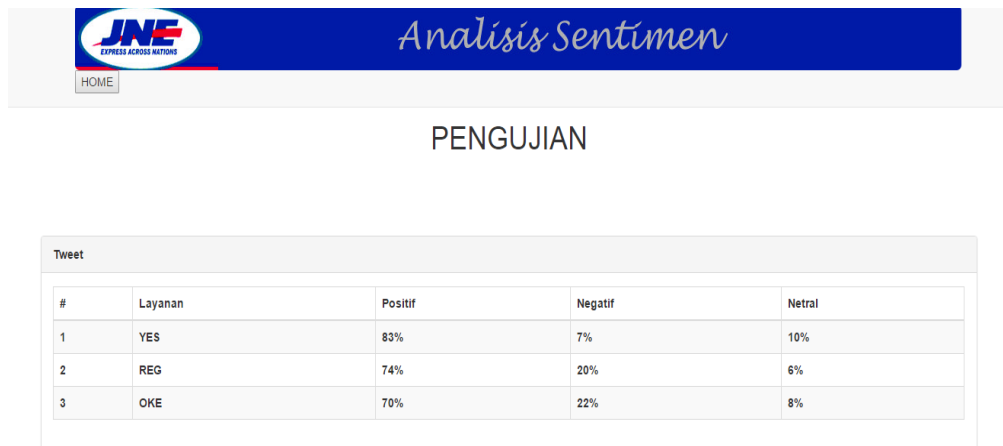
Dari perhitungan tersebut dapat disimpulkan kategori dari data uji termasuk kedalam kategori sentimen negatif, karena memiliki probabilitas tertinggi yaitu 0,05.

Hasil program dari proses pengujian yang dilakukan, dapat dilihat pada gambar 2 dan gambar 3.



#	Username	Waktu	Tweet
1	@ltha_shanty	Mon Jun 27 16:03:44 +0000 2016	@JNECare min tolong cek Upgaa01909700616 , jne yes to blm nyampe. Paket penting soalnya.
2	@Deemtrn08	Mon Jun 27 15:33:21 +0000 2016	Hi min @JNECare , ongkir YES dari Kerinci ke pematangsiantar berapa ya?
3	@asqiom	Mon Jun 27 14:38:11 +0000 2016	@JNECare apa menjelang lebaran gini pengiriman lambat? pake YES 3 hari belum sampe. cek tracking di web historynya ga muncul
4	@luhur1970	Mon Jun 27 14:09:23 +0000 2016	@JNECare @JNE_ID tolong cek kiriman YES no DPSH700040845216
5	@hiaril	Mon Jun 27 14:07:39 +0000 2016	@JNECare Paket Yes Jakarta-Purwakarta berapa lama?
6	@ardibkhe	Mon Jun 27 13:33:47 +0000 2016	@JNECare TGRF500409533116 admin. jne ini kirim paket gmn y. saya udah bayar mahal pake jne yes dr Tangerang-Jakarta timur masa gak sampe???
7	@FauziBhat	Mon Jun 27 13:27:27	@JNECare min naket epus SBGRAN0382254816 hlm camnal naktah naket YES dan dlkrim tml 24

Gambar 2. Pengambilan Data Latih



Gambar 3. Hasil Pengujian

4. KESIMPULAN

Berdasarkan dari hasil penelitian yang telah dilakukan, penggunaan metode *naïve bayes classifier* terhadap opini berupa tweet pada akun twitter jasa ekspedisi barang JNE dapat mengklasifikasi ke dalam 3 kelas sentimen yaitu, positif, negatif dan netral. Seluruh data latih dan data uji dilakukan tapan preprocessing. Akurasi pada sistem ini dapat mencapai 83%. Tingkat akurasi akan semakin baik bila data latih yang digunakan lebih banyak, jumlah kategori pada setiap kelas sentimen pun akan berpengaruh.

DAFTAR PUSTAKA

- Darujati, C., 2010. Perbandingan Klasifikasi Dokumen Teks Menggunakan Metode Naive Bayes dengan K-Nearest Neighbor. *Jurnal Informatika*, Januari. Volume 13.
- Ganisaputra, y. & Tan, R., 2014. Pembuatan Aplikasi Data mining Facebook dan Twitter dengan Naive Bayes Classifier. *Jurnal Teknologi Informasi*, 12 Agustus.
- 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, Volume 3, pp. 92-99.
- Rakhmat Sentiaji, A. & Mukaharil Bachtiar, A., 2010. Analisis Sentimen Terhadap Acara Televisi Berdasarkan opini Publik. *Jurnal Ilmiah Komputer dan Informatika*.
- Septian Nugroho, Y., 2014. Data Mining Menggunakan Algoritma Naive Bayes Untuk Klasifikasi Kelulusan Mahasiswa Universitas Dian Nuswantoro. *Jurnal Ilmiah*, September.21(2).