

ANALISIS SENTIMEN PADA JASA OJEK ONLINE MENGUNAKAN METODE NAÏVE BAYES

Didik Garbian Nugroho*, Yulison Herry Chrisnanto, Agung Wahana

Jurusan Informatika, Fakultas Matematika dan Ilmu Pengetahuan Alam

Universitas Jenderal Achmad Yani

Jalan Terusan Jenderal Sudirman No. 148, Cimahi 40513, Telp./Fax: (022) 6656190

*E-mail: garbian1.didik@gmail.com

Abstrak

Ojek online adalah sebuah terobosan baru di bidang transportasi yang semakin diminati masyarakat. Analisis sentimen pada jasa ojek online merupakan proses mengekstraksi pendapat, sentimen, evaluasi, dan emosi orang tentang pelayanan ojek online yang tertulis. Di media sosial masyarakat mengeluarkan beragam opini tentang pelayanan dari transportasi ini dengan jumlah yang banyak, sehingga terdapat kesulitan untuk menentukan opini yang bersifat positif, negatif ataupun netral. Penelitian terdahulu mengenai analisis sentimen pada Twitter menggunakan metode Naïve Bayes dengan data sebanyak 3000 data yang terdiri dari 1000 kalimat netral, 1000 kalimat positif dan 1000 kalimat negatif dengan hasil akurasi hingga 88%. Sedangkan penelitian yang akan dilakukan yaitu membuat sistem yang mampu mengklasifikasikan sentimen ke dalam sentimen positif, netral atau negatif serta menyalurkan opini tersebut ke setiap jasa yang bersangkutan dengan opini yang muncul. Hasil yang didapatkan dari akurasi naïve bayes memperoleh ketepatan 80%.

Kata kunci: Naïve Bayes, Ojek Online, Opini, Twitter

1. PENDAHULUAN

Transportasi adalah hal yang dekat dengan masyarakat. Transportasi ada berbagai macam jenisnya, baik roda dua maupun roda empat. Transportasi umum masih digunakan masyarakat sebagai sarana untuk menuju suatu tempat walaupun untuk memiliki kendaraan pribadi di zaman sekarang sangatlah mudah. Saat ini, ojek *online* menjadi tren transportasi umum terbaru di kalangan masyarakat. Ojek *online* sendiri merupakan transformasi dari ojek konvensional yang bertempat di suatu pangkalan untuk menunggu pelanggan, atau biasa disebut ojek pangkalan. Kemudahan dalam pemesanannya dengan menggunakan teknologi *smartphone* membuat ojek *online* semakin diminati. Hal ini dapat menjadi solusi transportasi di Indonesia, terutama di kota besar yang seringkali mengalami kemacetan. Ojek yang merupakan kendaraan menggunakan sepeda motor, sangat efektif untuk mobilitas di kemacetan kota.

Mayoritas perusahaan ojek *online* bermitra dengan para pengemudi ojek berpengalaman dan terdapat di area kota-kota besar, seperti Jabodetabek, Bandung, Bali, Yogyakarta & Surabaya. Layanan utama ojek *online* adalah mengantar penumpang ke tempat tujuan. Namun selain mengantar penumpang juga melayani antar jemput barang, pemesanan makanan, belanja, tiket, dan sebagainya (Marikxon, 2015). Cara kerja ojek *online* yaitu dengan membuka aplikasi yang terpasang di *smartphone*, lalu memilih layanan yang diinginkan dan mengkonfirmasi pesanan. Di dalam aplikasi terdapat layanan untuk mengarahkan pengemudi/*driver* ke tempat pemesan berada. Setelah mengkonfirmasi pesanan, *driver* akan segera mengantar pesanan menuju pelanggan. Ini menjadi keunggulan tersendiri bila dibandingkan dengan ojek konvensional.

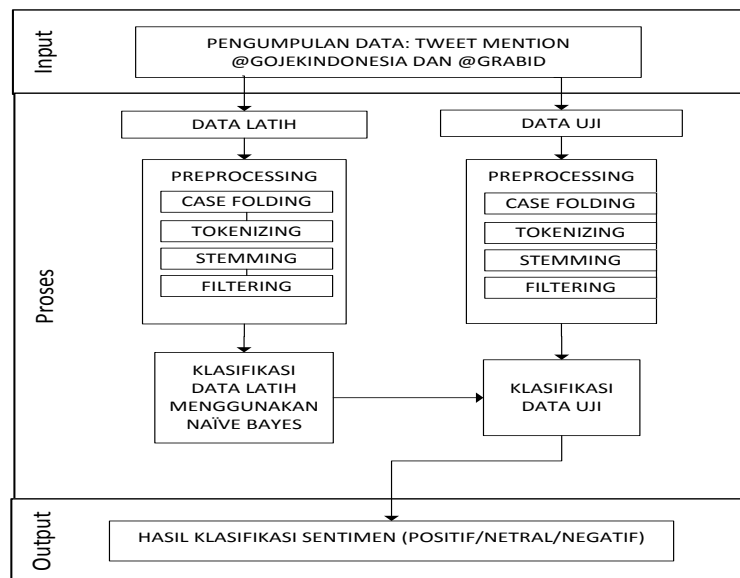
Seiring berkembangnya jasa ojek *online*, masyarakat menjadi sering membicarakannya. Masyarakat memberikan pendapat dan opininya melalui berbagai media, salah satunya media sosial *Twitter*. Opini yang diberikan masyarakat terhadap jasa ojek *online* beragam, seperti pelayanan, aplikasi dan pengemudi. Di *Twitter*, perusahaan ojek *online* memiliki akun resmi untuk memberikan informasi *up-to-date* tentang layanan maupun menampung *tweet* komentar-komentar dari masyarakat dan pelanggan. Berdasarkan penelitian PeerReach, Indonesia merupakan pengguna *Twitter* ketiga teraktif di dunia, yang berarti pengguna *Twitter* di Indonesia termasuk paling aktif di dunia (Tarigan, 2015). Apabila diteliti lebih lanjut terhadap *tweet* tersebut maka akan didapatkan sebuah sentimen yang apabila dikumpulkan akan dapat dibuat kesimpulan jasa ojek manakah yang memiliki sentimen baik di *Twitter*. Analisis sentimen adalah studi komputasi dari opini-opini, sentimen, serta emosi yang diekspresikan dalam teks (Liu, 2011). Tugas dasar dalam analisis

sentimen adalah mengelompokkan polaritas dari teks yang ada dalam dokumen, kalimat, atau pendapat. Polaritas mempunyai arti apakah teks yang ada dalam dokumen, kalimat, atau pendapat memiliki aspek positif atau negatif.

Salah satu teknik pembelajaran mesin untuk analisis sentimen adalah *Naïve Bayes Classifier* (NBC). NBC merupakan teknik pembelajaran mesin yang berbasis probabilistik. NBC adalah metode sederhana tetapi memiliki akurasi dan performansi yang tinggi dalam pengklasifikasian teks (Routray, Swain & Prava, 2013). Analisis sentimen yang dibangun menggunakan NBC memperoleh akurasi sebesar 83% dan rata-rata harmonik sebesar 90,713% (Ling, Eka & Bagus, 2014). Perbandingan beberapa metode seperti metode *Naïve Bayes*, *K-nearest Neighbor*, dan gabungan *K-means* dan *LVQ* dalam mengklasifikasikan kategori buku berbahasa Indonesia dengan data yang digunakan berjumlah 200 buku, 150 buku digunakan sebagai data latih, sedangkan 50 buku digunakan sebagai data uji. Dari hasil penelitian yang dilakukan, metode KNN memperoleh akurasi sebesar 96%, kemudian *Naïve Bayes* sebesar 98%, lalu kombinasi *K-Means* dan LVQ menghasilkan akurasi sebesar 92,2%. Metode *Naïve Bayes* mendapatkan hasil akurasi yang tertinggi (Darujati, 2010). Tingkat keakurasian pengklasifikasian sangat dipengaruhi oleh proses *training* (Sentiaji & Bachtiar, 2014). Kompleksitas data juga dapat mempengaruhi hasil akurasi (Septian, 2014).

2. METODOLOGI

Berikut ini adalah gambaran umum sistem pada penelitian analisis sentimen jasa ojek *online*.



Gambar 1. Gambaran Umum Sistem.

2.1 Pengumpulan Data

Data masukan yang akan digunakan bersumber dari *tweet* mention akun @gojekindonesia dan @GrabID. Dari data tersebut, sebagian menjadi data latih dan sebagian menjadi data uji. Kedua data tersebut dilakukan *pre-processing* terlebih dahulu untuk membuat teks menjadi ‘bersih’.

2.2 Tahap *Pre-processing*

Pada tahap *pre-processing* data, data *tweet* mentah terlebih dahulu dilakukan proses *case folding*, *tokenizing*, *stemming*, serta *filtering*. Hasil dari tahapan ini menghasilkan fitur yang digunakan sebagai data pembelajaran mesin oleh NBC. Dalam tahapan *preprocessing*, terdiri dari proses *case folding*, *tokenizing*, *stemming* dan *filtering*. Berikut ini merupakan penjelasan dari tahapan *preprocessing*:

- 1) Pada tahapan *case folding*, teks dilakukan proses perubahan dari huruf besar menjadi huruf kecil dan menghilangkan seluruh tanda baca pada kalimat.

- 2) Pada tahapan *tokenizing*, setiap kata akan dipisahkan berdasarkan spasi yang ditemukan.
- 3) Pada tahapan *stemming*, yaitu perubahan kata berimbuhan menjadi kata dasar.
- 4) Pada tahapan *filtering*, yaitu pembuangan kata-kata tidak penting dari hasil token.

2.3 Tahap Klasifikasi Naïve Bayes Classifier

Pada tahap ini akan dilakukan pengklasifikasian teks bersih yang sudah melalui tahap *pre-processing*, untuk mengetahui kategori dari teks tersebut berdasarkan data latih. Metode Naïve Bayes merupakan proses pengklasifikasian atau pengujian suatu teks baru yang belum pernah diketahui kategorinya berdasarkan basis pengetahuan yang dimiliki. Pada tahapan ini terdapat empat tahapan yang dilakukan.

1) Pembentukan Fitur Data Latih

Tahap ini merupakan tahapan awal melakukan pelatihan data latih. Dalam penelitian ini yang dimaksud dengan fitur adalah kata kunci yang akan menjadi parameter satuan data latih, yaitu dokumen (*tweet*) untuk diklasifikasikan ke dalam kelas yang telah ditentukan (positif, negatif dan netral). Dalam kata lain, fitur adalah kata yang memiliki nilai sentimen. Probabilitas dihitung dengan persamaan sebagai berikut.

$$p(wk|vj) = (nk + 1) / (\text{jumlah frekuensi} + \text{jumlah kata}) \tag{1}$$

2) Perhitungan Nilai Probabilitas Data Latih

Setelah dibentuk fitur dengan kemunculannya dari data latih selanjutnya menghitung probabilitas dari setiap kelas dengan cara sebagai berikut.

3) Menentukan Probabilitas Kelas

Setelah didapat probabilitas dari setiap kelas, selanjutnya menghitung probabilitas setiap fitur pada setiap kelas menggunakan persamaan berikut.

$$p(wk | vj) = (f(w_{ki} | c_i) + 1) / (\sum f()_{ci} + |W|) \tag{2}$$

4) Penghitungan Vmap

Di tahap ini dilakukan penghitungan nilai probabilitas teks berdasarkan data latih. Teks yang akan diuji dilakukan tahap *pre-processing* dan penghitungan Vmap. Vmap adalah perhitungan yang digunakan NBC untuk menentukan probabilitas tertinggi data uji dari masing-masing kelas berdasarkan dari proses *learning*. Persamaan yang digunakan adalah sebagai berikut.

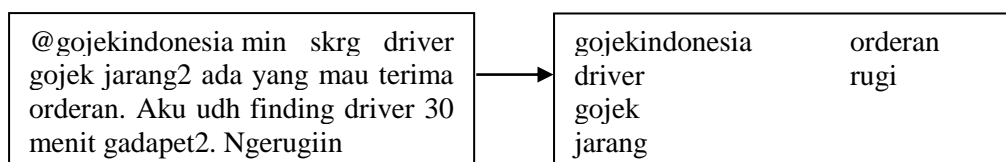
$$Vmap = \underset{\{positif, negatif\}}{argmax} p(wk|c) \times p(c) \tag{3}$$

3. HASIL DAN PEMBAHASAN

3.1 Tahapan Pre-Processing

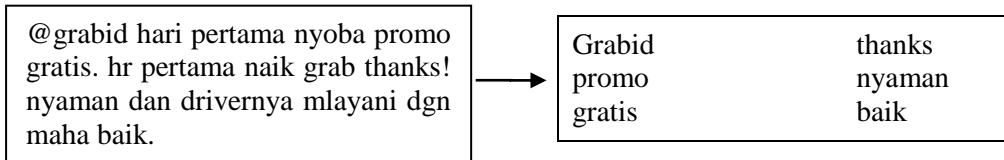
Sebelum data *tweet* digunakan, tahap *pre-processing* dilakukan untuk mendapatkan data bersih. Tahapan yang dilakukan antara lain *case folding*, *tokenizing*, *stemming*, dan *filtering*.

Berikut ini adalah data latih 1 sebelum dan setelah dilakukan *pre-processing*.



Gambar 2. Data Latih 1 Sebelum dan Sesudah Pre-processing.

Dan berikut ini adalah data latih 2 sebelum dan setelah dilakukan *pre-processing*.



Gambar 3. Data Latih 2 Sebelum dan Setelah *Pre-processing*.

3.2 Tahapan Naïve Bayes Classifier

Pada data latih dilakukan proses pembentukan fitur, perhitungan nilai probabilitas, dan menentukan probabilitas kelas. Berikut ini adalah hasil pembentukan fitur pada masing-masing data latih.

Tabel 1. Hasil Pembentukan Fitur Data Latih.

Data	Fitur	Kemunculan	Kelas Sentimen
D1	gojekindonesia	1	netral
	driver	1	netral
	gojek	1	netral
	jarang	1	negatif
	orderan	1	netral
	rugi	1	negatif
D2	grabid	1	netral
	promo	1	positif
	gratis	1	positif
	thanks	1	positif
	nyaman	1	positif
	baik	1	positif

Selanjutnya data latih dihitung probabilitas kemunculan kata pada setiap kelas. Hasilnya adalah sebagai berikut.

Tabel 2. Frekuensi Kemunculan Kata.

Kelas sentimen	Data(j)		Fd(C _j)	P(C _j)
	D1	D2		
Positif	0	1	1	1/2
negatif	1	0	2	1/2
netral	1	1	2	2/2

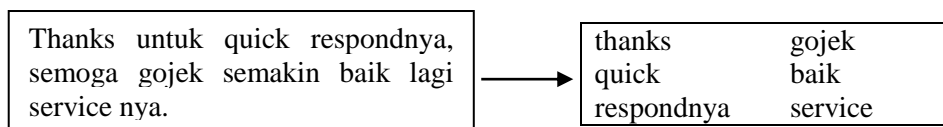
Setelah didapat probabilitas dari setiap kelas, selanjutnya menghitung probabilitas setiap fitur pada setiap kelas dengan hasil sebagai berikut.

Tabel 3. Probabilitas Tiap Fitur pada Kelas Sentimen.

Data $f(w_{ki}, c_i)$	Kelas sentimen (c)		
	Positif	Negatif	Netral
gojekindonesia	$\frac{0+1}{5+12} = \frac{1}{17}$	$\frac{0+1}{2+12} = \frac{1}{14}$	$\frac{1+1}{5+12} = \frac{2}{17}$
driver	$\frac{0+1}{5+12} = \frac{1}{17}$	$\frac{0+1}{2+12} = \frac{1}{14}$	$\frac{1+1}{5+12} = \frac{2}{17}$
gojek	$\frac{0+1}{5+12} = \frac{1}{17}$	$\frac{0+1}{2+12} = \frac{1}{14}$	$\frac{1+1}{5+12} = \frac{2}{17}$
jarang	$\frac{0+1}{5+12} = \frac{1}{17}$	$\frac{1+1}{2+12} = \frac{2}{14}$	$\frac{0+1}{5+12} = \frac{1}{17}$
pesan	$\frac{0+1}{5+12} = \frac{1}{17}$	$\frac{0+1}{2+12} = \frac{1}{14}$	$\frac{1+1}{5+12} = \frac{2}{17}$
rugi	$\frac{0+1}{5+12} = \frac{1}{17}$	$\frac{1+1}{2+12} = \frac{2}{14}$	$\frac{0+1}{5+12} = \frac{1}{17}$
grabid	$\frac{0+1}{5+12} = \frac{1}{12}$	$\frac{0+1}{2+12} = \frac{1}{14}$	$\frac{1+1}{5+12} = \frac{1}{17}$
promo	$\frac{1+1}{5+12} = \frac{2}{17}$	$\frac{0+1}{2+12} = \frac{1}{14}$	$\frac{0+1}{5+12} = \frac{1}{17}$
gratis	$\frac{1+1}{5+12} = \frac{2}{17}$	$\frac{0+1}{2+12} = \frac{1}{14}$	$\frac{0+1}{5+12} = \frac{1}{17}$
thanks	$\frac{1+1}{5+12} = \frac{2}{17}$	$\frac{0+1}{2+12} = \frac{1}{14}$	$\frac{0+1}{5+12} = \frac{1}{17}$
nyaman	$\frac{1+1}{5+12} = \frac{2}{17}$	$\frac{0+1}{2+12} = \frac{1}{14}$	$\frac{0+1}{5+12} = \frac{1}{17}$
baik	$\frac{1+1}{5+12} = \frac{2}{17}$	$\frac{0+1}{2+12} = \frac{1}{14}$	$\frac{0+1}{5+12} = \frac{1}{17}$

Pada data uji, dilakukan penghitungan V_{map} untuk menentukan probabilitas tertinggi dari masing-masing kelas berdasarkan dari proses pelatihan. Nilai probabilitas tertinggi merupakan sentimen dari data *tweet* tersebut.

- 1) Hasil sebelum dan sesudah *pre-processing* data uji.



Gambar 2. Data Latih 1 Sebelum dan Setelah *Pre-processing*.

- 2) V_{map} untuk sentimen positif

$$\begin{aligned}
 V_{map}(\text{positif}) &= P(\text{"positif"}) P(\text{"thanks"}|\text{positif}) P(\text{"gojek"}|\text{positif}) P(\text{"baik"}|\text{positif}) \\
 &= \frac{1}{2} \times \frac{2}{17} \times \frac{1}{17} \times \frac{2}{17} = 0.000396981
 \end{aligned}$$

- 3) V_{map} untuk sentimen negatif

$$\begin{aligned}
 V_{map}(\text{negatif}) &= P(\text{"negatif"}) P(\text{"thanks"}|\text{negatif}) P(\text{"gojek"}|\text{negatif}) P(\text{"baik"}|\text{negatif}) \\
 &= \frac{1}{2} \times \frac{1}{14} \times \frac{1}{14} \times \frac{1}{14} = 0.0001789555
 \end{aligned}$$

- 4) Vmap untuk sentimen netral

$$\begin{aligned} V_{\text{map}}(\text{netral}) &= P(\text{"netral"}) P(\text{"thanks"}|\text{netral}) P(\text{"gojek"}|\text{netral}) P(\text{"baik"}|\text{netral}) \\ &= \frac{2}{2} \times \frac{1}{17} \times \frac{2}{17} \times \frac{1}{17} = 0.000393588 \end{aligned}$$

3.3 Tahapan Pengujian

Pada penelitian ini, pengujian dilakukan dengan menggunakan data sebanyak 800 *tweet* yang dibagi menjadi 300 data latih dan 500 data uji. Kemudian, dari 300 data latih dibagi pada setiap klasifikasinya menjadi 100 data latih positif, 100 data latih negatif dan 100 data latih netral. Hasil dari pengujian metode Naïve Bayes akurasi yang didapat sebesar 80%.

4. KESIMPULAN

Penelitian ini menghasilkan sebuah sistem analisis sentimen pada jasa ojek *online* menggunakan metode *Naïve Bayes*. Proses penelitian ini menggunakan masukan berupa *tweet mention* jasa ojek *online* di Twitter. Hasil dari pengujian, sistem mampu mengklasifikasi sentimen menggunakan *Naïve Bayes* dengan akurasi yang dihasilkan sebesar 80% berdasarkan 800 data *tweet* yang terdiri atas 300 data latih dan 500 data uji. Terdapat kesalahan pada data uji fitur yang muncul tidak sesuai dengan klasifikasinya. Akurasi klasifikasi dapat ditingkatkan dengan menambah jumlah data latih.

DAFTAR PUSTAKA

- Darujati, C. (2010). Perbandingan Klasifikasi Dokumen Teks Menggunakan Metode Naive Bayes dengan K-Nearest Neighbor. *Jurnal Informatika*, 13.
- Ling, J., Eka N, I., & Bagus Oka, T. (2014). Analisis Sentimen Menggunakan Metode Naive Bayes Classifier Dengan Seleksi Fitur Chi Square. *E-Jurnal Matematika*, 3, pp 92-99.
- Liu, B. (2012). *Sentiment Analysis and Opinion Mining*. Morgan & Claypool.
- Marikxon. (2015). *Go-Jek, Startup Panggilan Ojek Yang Kini Bisa Dipesan Dari Smartphone*. Retrieved from <https://www.maxmanroe.com/go-jek-startup-panggilan-ojek-yang-kini-bisa-dipesan-dari-smartphone.html>
- 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*, 21 (2).
- Tarigan, I. A. (2013). *Pengguna Twitter Indonesia Teraktif Ketiga di Dunia*. Retrieved 2016, from CHIP Online Indonesia http://chip.co.id/news/apps-social_media/9030/pengguna_twitter_indonesia_teraktif_ketiga_di_dunia
- Routray, P., Swain, C. K., & Prava, S. (2013). A Survey on Sentiment Analysis. *International Journal of Computer Application*. pp. 1-8.