

ANALISIS SENTIMEN PELANGGAN TOKO ONLINE JD.ID MENGUNAKAN METODE NAÏVE BAYES CLASSIFIER BERBASIS KONVERSI IKON EMOSI

Fransiska Vina Sari

Magister Ilmu Komputer Fakultas Teknologi Informasi
Universitas Budi Luhur, Jakarta
Email: fransiskavinasari17@gmail.com

Arief Wibowo

Magister Ilmu Komputer Fakultas Teknologi Informasi
Universitas Budi Luhur, Jakarta
Email: arief.wibowo@budiluhur.ac.id

ABSTRAK

Analisis Sentimen adalah suatu teknik mengekstrak data teks untuk mendapatkan informasi tentang sentimen bernilai positif, netral maupun negatif. Analisis sentimen diberikan oleh pengguna internet pada media sosial untuk memberikan suatu penilaian atau opini pribadi. Salah satu toko online Indonesia yang sering mendapatkan sentimen pengguna melalui media sosial adalah JD.id. Adanya sentimen opini dari konsumen tentang JD.id dapat dianalisis dan dimanfaatkan untuk mendapatkan informasi yang berguna bagi pelanggan lain maupun pihak toko. Dengan menggunakan teknik *Text Mining* metode klasifikasi, akan diketahui suatu sentimen bernilai positif, netral atau negatif. Salah satu algoritme yang banyak digunakan dalam analisis sentimen adalah metode klasifikasi *Naïve Bayes*. Penelitian ini menggunakan metode *Naïve Bayes Classifier (NBC)* dengan pembobotan *tf-idf* disertai penambahan fitur konversi ikon emosi (*emoticon*) untuk mengetahui kelas sentimen yang ada dari *tweet* tentang toko JD.id. Hasil penelitian menunjukkan bahwa metode *Naïve Bayes* tanpa penambahan fitur mampu mengklasifikasi sentimen dengan nilai akurasi sebesar 96,44%, sementara jika ditambahkan fitur pembobotan *tf-idf* disertai konversi ikon emosi mampu meningkatkan nilai akurasi menjadi 98%.

Kata kunci: analisis sentiment; *naïve bayes classifier*; konversi ikon emosi.

ABSTRACT

Sentiment Analysis is a technique of extracting text data to get information about positive, neutral or negative sentiments. Sentiment analysis is provided by internet users on social media to provide an assessment or personal opinion. One of the Indonesian online stores that often get user sentiment through social media is JD.id. The existence of sentiment from consumers' opinions about JD.id can be analyzed and utilized to obtain useful information for other customers and the store. By using the Text Mining technique classification method, we will find a positive, neutral or negative sentiment. One algorithm that is widely used in sentiment analysis is the Naïve Bayes classification method. This study uses the Naïve Bayes Classifier (NBC) method with tf-idf weighting along with the addition of the emotion icon conversion feature (emoticon) to determine the existing sentiment class from a tweet about the JD.id store. The results showed that the Naïve Bayes method without the addition of features was able to classify sentiments with an accuracy value of 96.44%, while if the tf-idf weighting feature was added along with the conversion of emotional icons it could increase the accuracy value to 98%.

Keyword: *sentiment analysis; naïve bayes classifier; convert emoticon.*

1. PENDAHULUAN

Perkembangan teknologi dan informasi yang semakin pesat termasuk salah satunya adalah internet. Pengguna internet mulai memberikan opini mereka terhadap berbagai topik di media sosial salah satunya adalah media sosial *twitter*. Selain media sosial industri-industri juga sudah mulai merambah ke internet, industri ritel salah satunya sudah mulai memasarkan produknya secara *online*. Dengan adanya toko *online* masyarakat menjadi lebih mudah dalam melakukan transaksi berbelanja tanpa harus datang langsung ke toko penjual, salah satunya adalah toko *online* yang ada di Indonesia yaitu JD.id. Adanya keluhan dari masyarakat terkait toko *online* JD.id menunjukkan bahwa pelanggan merasa kurang puas dalam

berbelanja. Toko *online JD.id* kesulitan dalam memantau sentimen masyarakat terhadap kinerja *marketplace* nya di media sosial *Twitter*.

Banyaknya opini masyarakat pada *twitter* yang membahas mengenai toko *online JD.id*. Opini tersebut bersifat subyektif. Opini tersebut merupakan *textual* data yang dapat dianalisa dan dimanfaatkan untuk mendapatkan informasi guna memantau sentimen masyarakat terhadap kinerja *JD.id* di media sosial *Twitter* menggunakan menggunakan teknik text mining yaitu sentimen analisis. Sentimen tersebut dikategorikan menjadi sentimen positif, netral dan negatif. Penelitian ini memodelkan proses klasifikasi sentimen pengguna *twitter* terhadap toko *online JD.id* untuk mengetahui layanan maupun produk dari *marketplace* yang bermanfaat bagi pelanggan maupun penjual menggunakan metode *Naïve Bayes Classifier* (NBC) yang dikembangkan pembobotan *tf-idf* dan fitur konversi ikon emosi

Penelitian terdahulu yang pernah dilakukan tentang sentimen analisis, di antaranya adalah penggunaan *Naïve Bayes* dalam klasifikasi sentimen analisis dari toko *online Indonesia* [1] [2], studi tersebut menyatakan bahwa metode *Naïve Bayes* mampu memiliki nilai akurasi klasifikasi sentimen di atas 90%. Studi lain dilakukan dengan *Naïve Bayes* dan menambahkan fitur analisis negasi [3] dan mencapai nilai akurasi 95%. Metode *Naïve Bayes* juga digunakan untuk melakukan sentimen analisis sentimen terhadap jasa maskapai penerbangan [4] maupun produk telekomunikasi [5]. Studi yang menekankan aspek seleksi fitur tersebut mampu melakukan klasifikasi sentimen dengan nilai akurasi sebesar 89,50%.

Studi lain yang menggunakan *Naïve Bayes* untuk mendeteksi komentar spam, dilakukan tanpa penambahan fitur, dan hanya mendapatkan akurasi sebesar 77,25% [6]. Tantangan melakukan analisis sentimen pada data media sosial, telah diselesaikan pada studi lain yang masih menggunakan metode *Naïve Bayes* pada *Twitter* [7] dan *Instagram* [8] dengan nilai akurasi di kisaran 90%. Seluruh studi terdahulu tersebut menjadi referensi utama untuk dilakukannya penelitian pada studi ini yang juga menggunakan *Naïve Bayes*. Studi yang dilakukan saat ini akan menambahkan fitur konversi ikon emosi atau emoticon karena diduga akan meningkatkan nilai akurasi jika dilakukan pengklasifikasian tanpa penambahan fitur apapun.

1.1 Text Mining

Text mining merupakan proses penambangan yang dilakukan oleh komputer untuk mendapatkan sesuatu yang baru, dan tidak diketahui sebelumnya, atau menemukan kembali informasi yang tersirat secara implisit. Temuan tersebut berasal dari informasi yang diekstrak secara otomatis dari sumber-sumber data teks yang berbeda. *Text Preprocessing* adalah bagian dari *Text Mining* yang dilakukan untuk menghapus *noise* pada kalimat. *Text Preprocessing* bertujuan untuk menghindari data yang kurang sempurna, gangguan pada data dan data yang tidak konsisten.

1.2 Sentimen Analisis

Analisis sentimen merupakan sebuah metode yang digunakan untuk mengekstrak data opini, memahami serta mengolah tekstual data secara otomatis untuk melihat sentimen yang terkandung dalam sebuah opini.

1.3 Naïve Bayes Classifier (NBC)

Naïve Bayes merupakan sebuah metode pengklasifikasian dengan menggunakan probabilitas sederhana yang berakar pada *Teorema Bayes* dan memiliki asumsi ketidaktergantungan (*independent*) yang tinggi dari masing – masing kondisi atau kejadian, Bentuk umum *teorema bayes* adalah sebagai berikut:

$$P(H|X) = \quad (1)$$

Keterangan:

X = Data dengan kelas yang belum diketahui

H = Hipotesa data X merupakan suatu kelas spesifik

P(H|X) = Probabilitas hipotesis H berdasarkan kondisi X (*posterior probability*)

P(H) = Probabilitas hipotesis H (*prior probability*)

2. METODOLOGI PENELITIAN

Metode *sampling* yang digunakan dalam penelitian ini menggunakan metode *Non Random Sampling* atau Sampel Tidak Acak. *Non Random Sampling* adalah metode dimana setiap elemen atau

anggota populasi tidak mempunyai peluang yang sama untuk dijadikan sampel dikarenakan pengambilan sampel mempunyai kriteria tertentu.

Non Random Sampling memiliki beberapa teknik, salah satunya adalah *sampling kuota (quota sampling)*. *Quota Sampling* merupakan teknik *sampling* yang menentukan jumlah sampel dari populasi yang memiliki ciri tertentu hingga jumlah kuota yang diinginkan tercapai. Pengambilan sampel berdasarkan pertimbangan peneliti saja, besar dan kriteria sampel telah ditentukan. Sampel data yang digunakan adalah 900 tweet, 300 *tweet* positif, 300 *tweet* netral dan 300 *tweet* negatif. Pengumpulan data *tweet* pada *twitter* menggunakan *RStudio* dengan memasukkan settingan *twitter API* dan mengetikkan kata kunci *JD.id*. Langkah-langkah dalam penelitian ini dibuat untuk dapat memenuhi prosedur agar hasil yang didapatkan sesuai dengan tujuan. Berikut merupakan langkah-langkah dalam penelitian ini:

- Identifikasi Masalah
- Tujuan Penelitian
- Studi Literatur
- Pengumpulan Data Tweet pada Twitter
- Pengolahan Data Menggunakan Text Preprocessing
- Pembobotan TF-IDF
- Klasifikasi menggunakan *Naïve Bayes Classifier* (NBC) dan penambahan fitur konversi ikon emosi
- Evaluasi dan Validasi
- Implementasi Aplikasi

2.1 Pemrosesan Teks

Setelah data dikumpulkan, dilakukan pengolahan data berupa *text preprocessing* (*Case Folding*, *Convert Emoticon*, *Cleansing*, *Text Transformation*, *Stopword Removal*, *Tokenizing* dan *Stemming*), pembobotan *tf-idf*, kemudian dilakukan pengklasifikasian menggunakan metode *Naïve Bayes Classifier* (NBC). Eksperimen dilakukan dengan menambahkan fitur konversi *emoticon* pada data teks yang diduga akan meningkatkan nilai akurasi.

2.2 Evaluasi dan Validasi

Evaluasi menggunakan *confused matrix* untuk mengukur tingkat akurasi dalam pengujian metode dan validasi menggunakan *k-fold cross validation*.

3. HASIL DAN PEMBAHASAN

3.1 Text Preprocessing

Rangkaian kegiatan pra-pemrosesan teks diawali dengan pengubahan bentuk teks menjadi huruf kecil semua. Hal ini telah terbukti mempercepat waktu pemrosesan secara keseluruhan. Tahap berikutnya adalah konversi ikon emosi yang menggunakan lima belas simbol ikon untuk diterjemahkan menjadi kata *ikonpositif*, *ikonnegatif*, *ikonnetral*. Berikut ini adalah tabel konversi ikon emosi yang digunakan seperti terlihat pada Tabel 1.

Tabel 1. Convert emoticon

No.	Emoticon	Hasil Convert Emoticon	Kelas
1.	:)	Senang	Positif
2.	:(Sedih	Negatif
3.	:D	Senang	Positif
4.	:s	Kesal	Negatif
5.	:')	Bahagia	Positif
6.	:'))	Bahagia	Positif
7.	:))	Senang	Positif
8.	:")	Senang	Positif
9.	:)	Senang	Positif
10.	:((Sedih	Negatif
11.	^^	Senang	Positif
12.	;(Sedih	Negatif
13.	:'(Sedih	Negatif
14.	: '((Sedih	Negatif
15.	: "(Sedih	Negatif

Proses *cleansing* dilakukan pada dataset dan mereduksi sekitar 13% bagian *corpus* yang tidak digunakan. Teks transformasi dilakukan dengan berimbas perubahan pada 61,7% *corpus* dengan penekanan perubahan singkatan dan tata bahasa yang tidak sesuai kaidah Bahasa Indonesia. Proses tokenisasi dan *stemming* juga dilakukan dengan memberikan dampak berupa terbentuknya kata dasar yang berasal dari 77,8% *corpus* yang dianalisis.

Pada tahap selanjutnya, dilakukan pembobotan pada *term-term* yang telah melalui proses *stemming* kemudian dihitung bobotnya dengan menggunakan *TF-IDF*. Bobot ditujukan untuk memberikan skor terhadap frekuensi kemunculan sebuah kata. *Term Frequency* adalah konsep pembobotan dengan mencari seberapa sering (frekuensi) munculnya sebuah *term* dalam satu dokumen. Hasil pembobotan dapat terlihat pada Tabel 2.

Tabel 2. Hasil Pembobotan *TF-IDF*

<i>Term</i>	<i>Dokumen ke-</i>															<i>DF</i>	<i>N/DF</i>	<i>IDF</i>
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15			
Aman	1	0	2	0	0	0	0	1	0	0	0	0	0	0	0	3	5	0.698
Buruk	0	0	0	0	0	0	0	0	1	1	1	0	0	1	1	5	3	0.477
Cepat	0	1	0	0	0	0	1	0	0	0	0	0	1	0	0	3	5	0.698
Cinta	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	2	7.5	0.875
Senang	0	1	0	1	0	0	0	0	0	0	0	1	0	0	0	3	5	0.698

3.2 *Pemodelan dengan Naïve Bayes Classifier*

Berdasarkan persamaan (1) maka dilakukan perhitungan pada dataset, adapun sampel data yang dilakukan perhitungan terlihat pada Tabel 3.

Tabel 3. Perhitungan *NBC*

	<i>Sudah</i>	<i>Belanja</i>	<i>Harga</i>	<i>Bayar</i>	<i>Mau</i>	<i>Kelas</i>
Text001	1	0	0	0	1	positif
Text002	0	1	0	0	1	netral
Text078	0	2	0	0	0	netral
Text218	0	0	0	0	1	negatif
Text316	1	1	0	1	0	negatif
Text474	0	1	2	0	0	?

Perhitungan probabilitas pada *Naïve Bayes* berdasarkan dokumen Text474 adalah sebagai berikut:

- Hitung probabilitas bersyarat (*likelihood*) dokumen Text474 pada kelas positif, netral dan negatif.

Kelas Positif:

$$P(\text{Text474}|\text{Positif}) = 0$$

Kelas Netral:

$$P(\text{Text474}|\text{Netral}) = 0$$

Kelas Negatif:

$$P(\text{Text474}|\text{Negatif}) = 0,0164$$

- Probabilitas *prior* dari kelas positif, netral dan negatif dihitung dengan proporsi dokumen pada tiap kelas:

$$P(\text{positif}) = 1/6 = 0.17$$

$$P(\text{netral}) = 2/6 = 0.33$$

$$P(\text{negatif}) = 3/6 = 0.50$$

- Hitung Probabilitas *posterior* dengan persamaan dari rumus *Naïve Bayes Classifier* adalah sebagai berikut:

$$P(\text{Text474}|\text{Positif}) = 0$$

$$P(\text{Text474}|\text{Netral}) = 0$$

$$P(\text{Text474}|\text{Negatif}) = 0,0013$$

Dari hasil perhitungan probabilitas tersebut, dapat disimpulkan bahwa dokumen Text474 termasuk dalam kelas negatif dikarenakan $P(\text{Negatif}|\text{Text474})$ lebih besar daripada $P(\text{Positif}|\text{Text474})$ dan $P(\text{Netral}|\text{Text474})$.

3.3 Evaluasi Model

Tahap evaluasi model dilakukan dengan menggunakan instrumen *Confusion Matrix* dengan hasil sebagaimana terlihat di Tabel 4.

	Tabel 4. <i>Confusion matrix</i>		
	<i>Pred Positif</i>	<i>Pred Netral</i>	<i>Pred Negatif</i>
<i>True Positif</i>	300	0	0
<i>True Netral</i>	11	288	1
<i>True Negatif</i>	6	0	294

Dari bentuk matriks yang diperoleh maka didapatkan nilai evaluasi model sebagai berikut:

Accuracy : $((300+288+294) / 900) * 100\% = 98\%$

Precision : $((300) / 300+11+6) * 100\% = 94,64\%$

Recall : $((300) / 300) * 100\% = 100\%$

Penelitian ini menggunakan metode *Naïve Bayes Classifier (NBC)* disertai perbandingan jika menambahkan pembobotan *tf-idf* serta fitur konversi ikon emosi yang hasilnya terlihat pada Tabel 5.

	Tabel 5. Perbandingan hasil akurasi			
<i>Klasifikasi</i>	<i>Tanpa Convert Emoti-con</i>	<i>Dengan Convert Emoti-con</i>	<i>Tanpa Convert Emoti-con</i>	<i>Dengan Convert Emoti-con</i>
<i>Naïve Bayes Classifier</i>	96,44%	97,11%	97,67%	98,00%

Dari penelitian tersebut dapat disimpulkan bahwa metode *Naïve Bayes Classifier* dengan pembobotan *tf-idf* serta dengan fitur konversi ikon emosi memiliki nilai akurasi tertinggi. Dengan menggunakan pembobotan *tf-idf* dan fitur konversi ikon emosi dapat meningkatkan akurasi. Dari hasil tersebut dapat dinyatakan bahwa hipotesis terpenuhi.

4. KESIMPULAN

Kesimpulan yang didapatkan dari hasil penelitian yang telah dilakukan, sebagai berikut:

- NBC* tanpa pembobotan *tf-idf* dan konversi ikon emosi (*convert emoticon*) memiliki nilai akurasi 96,44% sedangkan *NBC* dengan pembobotan *tf-idf* dan konversi ikon emosi (*convert emoticon*) memiliki nilai akurasi 98%, dengan demikian terjadi peningkatan akurasi sebesar 1,56%.
- Bahasa pemrograman yang digunakan dalam pembuatan aplikasi adalah *PHP* dan *HTML*. *Tools* yang digunakan dalam mengambil data *tweet* adalah *RStudio*, *tools* yang digunakan untuk melakukan eksperimen adalah *RapidMiner* dengan *text preprocessing* yang dilakukan yaitu *case folding*, *convert emoticon*, *cleansing*, *text transformation*, *stopword removal*, *tokenizing* dan *stemming*.
- Sentimen masyarakat terhadap toko online *JD.id* bersifat positif, netral dan negatif.

Saran-saran yang diusulkan agar penelitian ini dapat menjadi lebih baik di masa mendatang yaitu melakukan analisis pada bentuk frasa asing yang ditemui, kata-kata negasi serta kata-kata ambigu agar nilai akurasi dari proses klasifikasi yang didapatkan menjadi lebih baik.

DAFTAR PUSTAKA

- [1] C. Fiarni, H. Maharani, and R. Pratama, "Sentiment Analysis System for Indonesia Online Retail Shop Review Using Hierarchy Naive Bayes Technique," in 16 Fourth International Conference on Information and Communication Technologies (ICoICT), 2016, vol. 4.

- [2] Y. M. Febrianti, I. Indriati, and A. W. Widodo, "Analisis Sentimen Pada Ulasan 'Lazada ' Berbahasa Indonesia Menggunakan K-Nearest Neighbor (K-NN) Dengan Perbaikan Kata Menggunakan Jaro Winkler Distance," Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer, vol. 2, no. 10, pp. 3689–3698, 2018.
- [3] L. Oktasari, Y. H. Chrisnanto, and R. Yuniarti, "Text Mining Dalam Analisis Sentimen Asuransi Menggunakan Metode Naïve Bayes Classifier," in Prosiding Seminar Nasional Sains dan Teknologi, 2016, pp. 37–42.
- [4] R. Wati, "Penerapan Algoritma Genetika Untuk Seleksi Fitur Pada Analisis Sentimen Review Jasa Maskapai Penerbangan Menggunakan Naive Bayes," Jurnal Evolusi, vol. 4, no. BSI, pp. 25–31, 2016.
- [5] F. N. Zuhri and A. Alamsyah, "Analisis Sentimen Masyarakat Terhadap Brand Smartfren Menggunakan Naive Bayes Classifier Di Forum Kaskus," in e-Proceeding of Management, 2017, vol. 4, no. 1, pp. 242–251.
- [6] A. Rachmat C. and Y. Lukito, "Deteksi Komentar Spam Bahasa Indonesia Pada Instagram Menggunakan Naive Bayes," Jurnal Ultimatics, vol. IX, no. June, pp. 50–60, 2017.
- [7] A. F. Hidayatullah and A. Azhari SN, "Analisis Sentimen Dan Klasifikasi Kategori Terhadap Tokoh Publik Pada Twitter," in Prosiding Seminar Nasional Informatika 2014 (semnasIF 2014), 2014, vol. 1, no. UPN Veteran Yogyakarta, pp. 115–122.
- [8] W. A. Luqyana, I. Cholissodin, and R. S. Perdana, "Analisis Sentimen Cyberbullying pada Komentar Instagram dengan Metode Klasifikasi Support Vector Machine," Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer, vol. 2, no. 11, pp. 4704–4713, 2018.