

Analisis Tingkat Kepopuleran E-Commerce Di Indonesia Berdasarkan Sentimen Sosial Media Menggunakan Metode Naïve Bayes

Adika Sri Widagdo¹⁾, Bambang Soedijono W. A²⁾, Asro Nasiri³⁾

^{1,2,3)} Magister Teknik Informatika Universitas AMIKOM Yogyakarta

Jalan Ring Road Utara, Condongcatur, Sleman, Yogyakarta 55281

¹adika.sriwidagdo@gmail.com, ²bambang.s@amikom.ac.id, ³asro@amikom.ac.id

Abstrak

Saat ini semakin banyak pelaku usaha, baik perusahaan besar maupun ritel, beralih atau mengembangkan usaha ke arah digital yaitu berupa e-commerce. tidak bisa dipungkiri lagi bahwa e-commerce dapat mempermudah transaksi antara penjual dan pembeli, selain dalam bentuk website ditambah dengan fasilitas e-commerce berupa aplikasi smartphone sehingga masyarakat bisa dengan mudah mengakses dimanapun dan kapanpun. Selain itu menurut dari data APJII pada setiap tahunnya masyarakat indonesia yang bertransaksi menggunakan e-commerce semakin bertambah, sehingga banyak bermunculan e-commerce baru yang membuat masyarakat bingung dalam memilih e-commerce yang dipercaya, untuk itu diperlukan sebuah penilaian yang dapat mengetahui tingkat kepopuleran e-commerce di indonesia yang berdasarkan sentimen pada sosial media. Pada tulisan ini akan menggunakan Naïve Bayes untuk melakukan analisis sentimen dengan hasil akhir prosentase kepopuleran tiap e-commerce di Indonesia yaitu E-commerce 1 = 91,2%, E-commerce 2 = 76,4, E-commerce 3 = 37,8% E-commerce 4 = 33,8% dimana prosentase tersebut dihasilkan dari selisih antara sentimen positif dan negatif.

Kata Kunci: *naïve bayes*, sentimen *e-commerce*.

1. PENDAHULUAN

a. Latar Belakang

Saat ini perdagangan menjadi salah satu daya tarik para pelaku usaha. Banyak pelaku usaha, baik perusahaan besar maupun ritel, beralih atau mengembangkan usaha ke arah digital yaitu berupa e-commerce. Tidak bisa dipungkiri lagi bahwa e-commerce dapat mempermudah transaksi antara penjual dan pembeli, selain dalam bentuk website ditambah dengan fasilitas *e-commerce* berupa aplikasi smartphone sehingga masyarakat bisa dengan mudah mengakses dimanapun dan kapanpun. Selain itu menurut dari data APJII pada setiap tahunnya masyarakat indonesia yang bertransaksi menggunakan *e-commerce* semakin bertambah, sehingga banyak bermunculan *e-commerce* baru yang membuat masyarakat bingung dalam memilih *e-commerce* yang dipercaya, untuk itu diperlukan sebuah penilaian yang dapat mengetahui tingkat kepopuleran e-commerce di indonesia berdasarkan analisis sentimen pada sosial media yang diungkapkan dalam teks yang berupa opini positif atau negatif (Ling et al, 2014).

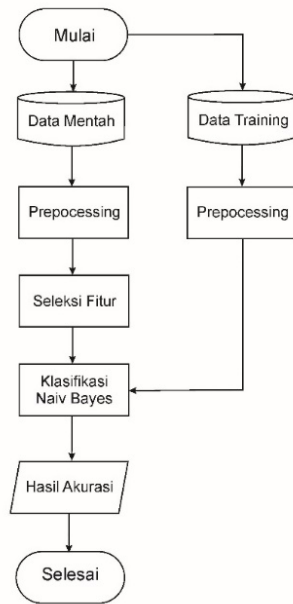
Pada penelitian ini media sosial yang digunakan adalah *twitter*, sebab banyak masyarakat yang menggunakan *twitter* dalam menyampaikan opininya daripada menyampaikan secara langsung. *Twitter* memungkinkan penggunanya saling berinteraksi satu sama lain sehingga memungkinkan untuk mendapatkan informasi yang dibutuhkan dalam pengambilan keputusan (Afshoh, 2017).

Dalam menganalisis sentimen terdapat sebuah metode Naïve Bayes yaitu teknik *machine learning* yang dibuat untuk mengklasifikasikan data berupa teks (Pratama et al, 2018). Klasifikasi digunakan untuk pengelompokan kata berdasarkan kelas pada opini. Naïve Bayes merupakan teknologi *pre-processing* dalam klasifikasi fitur yang menambah skalabilitas, akurasi dan efisiensi pada proses klasifikasi teks.

Berdasarkan permasalahan tersebut penelitian ini akan menganalisa tingkat kepopuleran *e-commerce* di Indonesia berdasarkan analisis sentimen pada media sosial *twitter* menggunakan metode *naïve bayes*.

2. METODE PENELITIAN

Berikut tahapan alur penelitian :

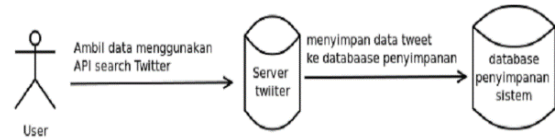


Gambar 1. Flowchart Penelitian.

Penelitian ini memiliki beberapa tahap antara lain yaitu pengumpulan data data mentah didapatkan dari proses *crawling twitter* menggunakan RStudio, pre-processing, pelabelan dan pembangun sistem klasifikasi menggunakan algoritma naïve bayes.

a. Pengumpulan Data

Data yang digunakan pada penelitian ini adalah data *tweet* maupun *retweet* pengguna *twitter* (Nugroho, 2018). Data terbatas pada 4 *e-commerce* di Indonesia yaitu Bukalapak, Lazada, Tokopedia dan Shopee. Pengumpulan data dilakukan secara terpisah antara *e-commerce* satu dengan yang lain, karena tujuan penelitian ini adalah mencari *e-commerce* terpopuler berdasarkan jumlah komentar positif yang terdapat pada setiap *e-commerce*. Jika pengambilan data secara keseluruhan, maka bisa dianggap tidak adil karena hasil dari *crawl* belum tentu memberikan jumlah data yang sama pada setiap *e-commerce*. Jumlah data yang digunakan pada penelitian ini adalah 500 data *tweet* pada setiap *e-commerce* sehingga keseluruhan data terdapat 2000 *tweet*.



Gambar 2. Skema Pengumpulan Data dari Twitter

b. Preprocessing

Tahap *preprocessing* merupakan tahapan pengolahan data yang bertujuan untuk menghilangkan data-data yang tidak digunakan untuk proses selanjutnya. Penelitian ini menggunakan 4 tahap *preprocessing*, yaitu :

- 1) *Cleansing data* : Proses ini melakukan pembersihan terhadap data *tweet* (kata-kata) yang telah didapat atau sering disebut dengan *noise*. Kata-kata yang tidak penting yaitu seperti URL, hashtag (#), *username* (@username), email, emoticon (:@, :*, :D), tanda baca seperti koma (,), titik (.) dan juga tanda baca lainnya. Contoh *cleansing data*, input kalimat : "@LazadaIDCare padahal sudah satu minggu pembayaran, kok belum dikirim sih ? #Kesel #Kecewa.", output kalimat : "padahal sudah satu minggu pembayaran kok belum dikirim sih Kesel Kecewa".
- 2) *Case Folding* : Data *tweet* yang didapat terdapat banyak perbedaan bentuk huruf, tahapan ini dilakukan untuk merubah bentuk semua huruf pada kalimat menjadi huruf kecil (lower case) atau dapat disebut juga penyeragaman bentuk huruf. Contoh *case folding*, input kalimat : padahal sudah satu minggu pembayaran kok belum dikirim sih Kesel Kecewa", output kalimat : "padahal sudah satu minggu pembayaran kok belum dikirim sih kesel kecewa".
- 3) *Tokenizing* : Proses ini disebut juga proses parsing yaitu tahapan pemotongan atau memisahkan per-kata atau bisa disebut pemenggalan kalimat menjadi kata (Artanti et al, 2018). Contoh *tokenizing*, input kalimat: padahal sudah satu minggu pembayaran kok belum dikirim sih kesel kecewa", output kalimat : "padahal, sudah, satu, minggu, pembayaran, kok, belum, dikirim, sih, kesel, kecewa ".

3. TINJAUAN PUSTAKA

a. Metode Naïve Bayes

Naive Bayes dikembangkan oleh *Reverend Thomas Bayes* pada abad ke 18. Klasifikasi dengan metode Naïve Bayes secara umum dilakukan dengan pendekatan peluang atau probabilitas (*Lamurias et al*, 2014). Metode Naïve Bayes adalah suatu metode yang digunakan untuk proses klasifikasi data yang menggunakan perhitungan probabilitas dan statistic (*Feldman et al*, 2007). Konsep dasar yang digunakan oleh Bayes adalah Teorema peluang bersyarat Bayes berikut:

$$P(A|B) = P(A) P(B|A)$$

Persamaan 1. Teorema peluang bersyarat
Peluang kejadian A bersyarat B ditentukan dari peluang A dan peluang B bersyarat A. Persamaan ini dikembangkan menjadi persamaan berikut:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

Persamaan 2. Peluang A Bersyarat B
Pada pengaplikasiannya persamaan ini dapat digunakan untuk klasifikasi dokumen dengan melakukan perhitungan nilai probabilitas. Klasifikasi dilakukan untuk menentukan kategori dari suatu dokumen. Persamaan ini dapat diubah kedalam bentuk persamaan berikut:

$$P(C_i|D) = \frac{P(D|C_i)P(C_i)}{P(D)}$$

Persamaan 3.

Metode Naive Bayes menganggap setiap variabel berdiri bebas satu sama lain dan tidak ada keterkaitan dengan variabel lain, sehingga sebuah dokumen akan dianggap sebagai kumpulan dari kata-kata yang menyusun dokumen tersebut, dan tidak memperhatikan urutan kemunculan kata pada dokumen. Sehingga perhitungan probabilitas dapat dianggap sebagai hasil perkalian dari probabilitas kemunculan kata-kata pada dokumen. Persamaan Naive Bayes Classifier dapat dilihat pada persamaan-persamaan berikut:

$$P(C_i) = \frac{fD(C_i)}{|D|}$$

Persamaan 4. Probabilitas

$P(C_i)$ adalah Probabilitas dari suatu kategori dokumen, $fD(C_i)$ adalah Frekuensi dokumen yang memiliki kategori C_i dan $|D|$ adalah Jumlah seluruh dokumen latih.

$$P(W_{kj}|C_i) = \frac{f(W_{kj}|C_i) + 1}{f(C_i) + |W|}$$

Persamaan 5. Probabilitas kemunculan W_{kj}

$P(W_{kj}|C_i)$ adalah Probabilitas kemunculan W_{kj} pada suatu dokumen dengan kategori kelas C_i . W_{kj} adalah frekuensi kata ke-k pada setiap kategori. W adalah jumlah kata pada dokumen test. $f(C_i)$ adalah frekuensi dokumen berkategori C_i . (1)

Pada persamaan ini terdapat suatu penambahan angka 1 pada pembilang untuk mengantisipasi jika terdapat suatu kata pada dokumen uji yang bernilai nol (0) karena tidak terdapat pada dokumen latih.

4. HASIL DAN PEMBAHASAN

a. Pengumpulan Data

Sumber data teks yang akan diolah pada penelitian ini berasal dari data *tweet* ataupun *retweet* oleh para pelanggan *e-commerce* yaitu Lazada, Tokopedia, Bukalapak dan Shopee. Setiap *e-commerce* diambil 500 data *tweet* dimana data tersebut hanya terbatas satu minggu kebelakang dari tanggal *crawling* sesuai dengan peraturan yang diberikan *twitter*.

text	(3)
1	Ada banyak alasan kenapa kamu harus coba Tokopedia: Belanja mud.
2	Tokopedia. Satu Aplikasi untuk Semua Kebutuhan https://t.co/FG3Vxxn
3	jual sweater hitam. 100 ribu. WA 089683148184. Kirim via JNE https://t.co/LKY7y3lojL
4	jual cover dakimakura sexy https://t.co/LKY7y3lojL
5	@imbaguss Hi Kak Bagus, Coki bisa informasikan untuk saat ini pe..
6	0838-6760-8070 (WA) obat kolesterol alami ampuh https://t.co/mpphF
7	AfifaAlra Selamat siang Ibu Almira, untuk merchant yang sudah terda.
8	@hajjoe_ Hai Kak, apakah yang dimaksud adalah transaksi INV/2019.
9	MERCI Propolis, obat herbal, ekstrak sarang lebah, anti virus, bakteri,
10	@TokopediaCare @tokopedia sy sudah bayar dan nominal yang sy ba
11	ayo serbu Tay Pin San di Tokopedia, dapatkan cashbacknya!!!! https://t.co/STORY OF SETH Wulanfadi
12	STORY OF SETH Wulanfadi . Diskon 20% jadi Rp 70.400 (normalnya 88
13	RT @ridwankamil: BAYAR PAJAK KENDARAAN BERMOTOR sekarang bisa
14	Grosir Gelang Sawan Sambet TolakBala New Model Terbaru Termurah

Gambar 3. Data Hasil Crawling dari Twitter

b. Preprocessing

Preprocessing yang dilakukan adalah membersihkan karakter-karakter yang tidak digunakan seperti URL, @, #, dan simbol yang tidak digunakan dalam proses selanjutnya. Setelah menghilangkan karakter yang tidak digunakan selanjutnya melakukan

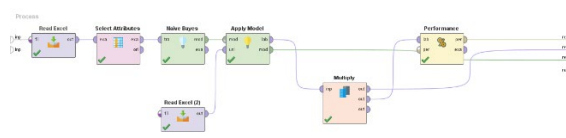
case folding untuk membuat karakter menjadi *lower case* semua. Hasilnya adalah sebagai berikut :

Value
List of length 1000
'jual' 'ambejos' 'salep' 'salwa' 'obat' 'wasir' ...
'tokopediacare' 'tokopedia' 'search' 'merah' 'hati' 'yg' ...
'obat' 'herbal' 'multi' 'khasiat' 'apollo' 'cody' ...
'run' 'denature' 'indonesia' 'majenang' 'tokopedia' 'https' ...
'rt' 'diva' 'fiction' 'po' 'masih' 'dibuka' ...
'rlthingy' 'tokopedia' 'banyak' 'cari' 'ajaaa' 'jenny' ...
'tokopediacare' 'halo' 'min' 'saya' 'pesan' 'barang' ...
'hannakartika' 'hai' 'kak' 'hanna' 'selengkapnya' 'mengenai' ...
'tokopediacare' 'min' 'mau' 'tanya' 'kenapa' 'no' ...
'cupcake' 'diva' 'tulip' 'papercup' 'cup' 'cake' ...
'jenis' 'emblem' 'i' 'vtec' 'red' 'bahan' ...
'tokopediacare' 'bagaimana' 'ini' 'kelanjutannya' 'katanya' 'setelah' ...
'respatiadityo' 'terima' 'kasih' 'atas' 'infonya' 'kak' ...
'jual' 'jersey' 'jersey' 'klub' 'dan' 'produk' ...

Gambar 4. Data Hasil Proses *Preprocessing*

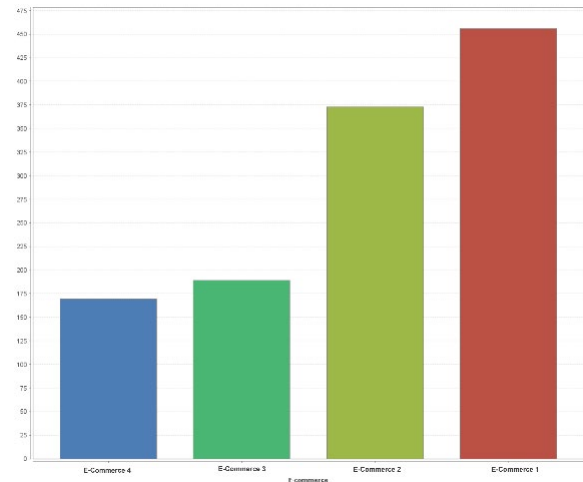
c. Klasifikasi Naïve Bayes

Dalam proses klasifikasi Naive Bayes, penelitian ini menggunakan aplikasi Rapidminer. Data yang telah di *pre-processing* dimasukkan kedalam aplikasi tersebut dengan langkah-langkah yaitu memasukkan dataset (data latih) yang telah diberi label, kemudian memilih atribut sebagai *primary key*, kemudian memasukkan data training lalu dilakukan klasifikasi Naïve Bayes. Berikut adalah tampilan proses klasifikasi Naive Bayes menggunakan Rapidminer :



Gambar 5. Proses Klasifikasi Naive Bayes

Sesuai tujuan dari penelitian ini yaitu mengetahui tingkat kepopuleran *E-Commerce* di Indonesia, maka proses klasifikasi Naive Bayes diatas didapatkan hasil seperti pada diagram dibawah ini :



Gambar 6. Diagram *E-Commerce*

d. Pengujian

Proses pengujian ini bertujuan untuk mengetahui akurasi yang didapatkan dari klasifikasi Naïve Bayes. Metode pengujian yang digunakan pada penelitian ini adalah Pengujian Performance. Pengujian ini merupakan *library* yang dimiliki aplikasi RapidMiner. Setelah dilakukan pengujian pada klasifikasi dengan menggunakan Metode Naïve Bayes antara dataset dengan data training didapatkan hasil presentase sentimen Negatif dan Positif *E-Commerce* sebagai berikut :

Accuracy : 97.14%

Tabel 1. Presentase sentimen.

	True Negatif	True Positif	Dass Precisio n
pred. Negatif	412	4	99.04%
pred. Positif	30	741	96.11%
class recall	93.21%	99.46%	

5. KESIMPULAN DAN SARAN

a. Kesimpulan

Berdasarkan penelitian yang telah dilakukan dengan menggunakan Algoritma Naive Bayes dengan menggunakan 500 data sentiment pada tiap *E-Commerce*, didapatkan hasil presentase tingkat kepopuleran *E-Commerce* di Indonesia dengan hasil sebagai berikut :

Tabel 2. Prosentase Sentimen

N o	E- Commere e	Positi f	Negati f	Prosentase
1.	E 1	456	44	91.2%
2.	E 2	373	127	74.6%
3.	E 3	189	311	37.8%
4.	E 4	169	331	33.8%

Untuk mencari presentase di atas peneliti menggunakan rumus sebagai berikut :

$$\text{Presentase} = \frac{\text{Total Sentimen Positif}}{\text{Total Sentimen}} \times 100\%$$

b. Saran

Dari hasil penelitian Analisis Tingkat Kepopuleran *E-Commerce* Di Indonesia Berdasarkan Sentimen Sosial Media adapun saran untuk bahan penelitian selanjutnya antara lain :

- 1) Tidak ada hasil pembandingan karena hanya menggunakan 1 metode.
- 2) Untuk penelitian selanjutnya diharapkan menggunakan data lebih dari 10000
- 3) Sumber data disarankan tidak hanya berasal dari *twitter*
- 4) Pengujian hanya menggunakan salah satu fasilitas dari Rapidminer yaitu *Library Performance*, diharapkan penelitian selanjutnya menggunakan pengujian *Split Validation* dan RSME

6. REFERENSI

- Pratama, N. D., Sari, Y. A., & Adikara, P. P. Analisis Sentimen Pada Review Konsumen Menggunakan Metode Naive Bayes Dengan Seleksi Fitur Chi Square Untuk Rekomendasi Lokasi Makanan Tradisional. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer e-ISSN, 2548, 964X*.
- Nugroho, A. (2018). Analisis Sentimen Pada Media Sosial *Twitter* Menggunakan Naive Bayes Classifier Dengan Ekstraksi Fitur N-Gram. *J-SAKTI (Jurnal Sains Komputer dan Informatika)*, 2(2), 200-209.
- Ling, J., Kencana, I. P. E. N., & Oka, T. B. (2014). Analisis Sentimen Menggunakan Metode Naive Bayes Classifier Dengan

Seleksi Fitur Chi Square. *E-Jurnal Matematika*, 3(3), 92-99.

Afshoh, F., Pamungkas, E. W., Kom, S., & Kom, M. (2017). *Analisa Sentimen Menggunakan Naive Bayes Untuk Melihat Persepsi Masyarakat Terhadap Kenaikan Harga Jual Rokok Pada Media Sosial Twitter* (Doctoral dissertation, Universitas Muhammadiyah Surakarta).

Artanti, D. P., Syukur, A., & Prihandono, A. (2018). Analisis Sentimen Untuk Penilaian Pelayanan Situs Belanja Online Menggunakan Algoritma Naive Bayes. *Konferensi Nasional Sistem Informasi (KNSI) 2018*.

Feldman, R., & Sanger, J. (2007). *The text mining handbook: advanced approaches in analyzing unstructured data*. Cambridge university press.

Lamúrias, A. F. M. (2014). *Identifying interactions between chemical entities in text* (Doctoral dissertat)