

ANALISIS SENTIMEN TOKO *ONLINE* MENGGUNAKAN ALGORITMA NAIVE BAYES CLASSIFIER

Mukti Qamal⁽¹⁾, Wahyu Fuadi⁽²⁾, Muslia⁽³⁾

Program Studi Teknik Informatika Universitas Malikussaleh
Jl. Batam, Blang Pulo, Muara Satu, Lhokseumawe, Aceh 24353

Abstrak

Belanja *online* merupakan salah satu kebiasaan masyarakat untuk mendapatkan barang atau jasa secara cepat dan mudah. Salah satu situs *online* yang paling banyak dikunjungi ialah Shopee yang mempunyai banyak fitur dan dapat digunakan oleh penggunanya untuk memberi ulasan (*review*) seperti memberikan bintang terhadap suatu produk atau juga dapat memberikan komentar terhadap suatu produk. Banyaknya komentar yang terdapat pada ulasan produk shopee membutuhkan pengklasifikasian untuk mengelompokkan suatu komentar termasuk positif atau sentimen negatif. Naive Bayes Classifier merupakan suatu algoritma yang dapat digunakan untuk mengklasifikasikan komentar – komentar tersebut, sehingga suatu produk dapat dinilai berdasarkan review sentimen positif. Pada penelitian ini digunakan 1000 data training, dimana 500 merupakan data training untuk sentimen positif dan 500 data training untuk sentimen negatif. Data uji dilakukan pada beberapa toko pakaian dengan mengambil 10 produk dan 10 komentar perproduknya, sehingga untuk data uji digunakan 100 data untuk sekali pengujian. Penelitian ini dapat di uji pada berbagai macam jenis toko pakaian, seperti toko pakaian wanita, toko pakaian pria dan toko pakaian anak –anak. Persentase sentimen positif yang paling tinggi diperoleh oleh toko pakaian anak – anak dari 100 komentar menghasilkan 90% sentimen positif. Persentase keakurasian pada toko *online* dengan algoritma naive bayes adalah 87 % akurat.

Kata kunci : e-commerce, review, analisis, naive bayes

Abstract

Online shopping is one of the habits of the people to get goods or services quickly and easily. One of the most visited online sites is Shopee which has many features and can be used by users to provide reviews such as giving stars to a product or also being able to provide comments on a product. The number of comments contained on shopee product reviews requires classification to classify a comment including positive sentiment or negative sentiment. Naive Bayes Classifier is an algorithm that can be used to classify these comments, so that a product can be assessed based on positive sentiment reviews. In this study, 1000 training data were used, of which 500 were training data for positive sentiment and 500 training data for negative sentiment. The

test data was carried out at several clothing stores by taking 10 products and 10 comments per product, so that the test data used 100 data for one test. This research can be tested on various types of clothing stores, such as women's clothing stores, men's clothing stores and children's clothing stores. Children's clothing store had the highest percentage of positive sentiment out of 100 comments resulting in 90% positive sentiment. The percentage of accuracy in online stores with the Naive Bayes algorithm is 87% accurate.

Keywords: e-commerce, review, analysis, naive bayes

1. PENDAHULUAN

Belanja online atau e-commerce merupakan sebuah transaksi yang dilakukan melalui media berupa situs – situs jual beli online ataupun jejaring sosial yang menyediakan barang atau jasa yang diperjual – belikan. Salah satu contoh e-commerce yang memiliki reputasi baik di Indonesia adalah Shopee. Shopee mempunyai fitur – fitur yang dapat digunakan oleh penggunanya untuk memberi ulasan (*review*) seperti memberikan bintang terhadap suatu produk atau juga dapat memberikan komentar terhadap suatu produk. *Review* dari pengguna tersebut biasanya dapat digunakan sebagai suatu tolak ukur yang efektif dan efisien untuk menemukan informasi terhadap suatu produk. *Review* atau ulasan dari pengguna aplikasi *e-commerce* umumnya berisi saran positif maupun keluhan yang bersifat negatif yang dituliskan secara tidak langsung akan berpengaruh pada calon pelanggan lainnya.

Analisis sentimen dapat dilakukan untuk mengklasifikasikan ulasan – ulasan menjadi sentimen positif ataupun sentimen negatif. Ada beberapa penelitian tentang analisis sentimen, yaitu Penelitian yang dilakukan oleh Mehdi Mursalat Ismail dan Kemas Muslim Lhaksamana dengan judul “Sentimen Analisis Pada Media Online Mengenai Pemilihan Presiden 2019 dengan Menggunakan Metode Naive Bayes”, *Output* dari *sentiment analysis* ini berupa *confusion matrix*. Penelitian ini membangun sistem yang dapat mendeteksi sebuah berita memiliki sentimen positif atau negatif pada salah satu pasangan calon presiden tahun 2019 dengan akurasi sebesar 79,5% untuk berita mengenai Jokowi-Ma’ruf dan 64% untuk berita mengenai Prabowo-Sandi. Penelitian yang dilakukan oleh Ni Putu Sri Merta Suryani, Linawati dan Komang Oka

Saputra dengan judul “Penggunaan Metode Naive Bayes Clasifier pada analisis sentimen Facebook Berbahasa Indonesia”, hasil dari penelitian tersebut adalah klasifikasi yang dihasilkan dari metode Naïve Bayes Classifier tergolong cukup baik. Untuk menguji tingkat keakuratan sistem dalam mengklasifikasikan opini, maka dilakukan pengujian hasil klasifikasi. Dari hasil pengujian diperoleh rata-rata akurasi sebesar 87,1 %. Semakin banyak data latih yang mirip dengan data testing maka hasil klasifikasi akan semakin bagus.

2. METODOLOGI PENELITIAN

2.1 Pengertian Data Mining

Data mining adalah suatu istilah yang digunakan untuk menguraikan penemuan pengetahuan di dalam *database*. *Data mining* adalah proses yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan *machine learning* untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai *database* besar. (Turban, dkk. 2005).

2.2 Text Mining

Text mining dapat diartikan sebagai penemuan informasi yang baru dan tidak diketahui sebelumnya oleh komputer, dengan secara otomatis mengekstrak informasi dari sumber-sumber yang berbeda. Kunci dari proses ini adalah menggabungkan informasi yang berhasil diekstraksi dari berbagai sumber. Proses text mining yang khas meliputi kategorisasi teks, text clustering, ekstraksi konsep/entitas, produksi taksonomi granular, sentiment analysis, penyimpulan dokumen, dan pemodelan relasi entitas (yaitu, pembelajaran hubungan antara entitas bernama).

2.3 Analisis Sentimen

Analisis sentimen adalah sebuah teknik untuk mengetahui opini masyarakat terhadap suatu subyek tertentu yang didapat dari sebuah kumpulan data. Pada jurnal yang ditulis oleh Liu menjelaskan bahwa analisis sentimen adalah bidang studi yang menganalisis pendapat, sentimen, evaluasi, sikap, dan emosi

seseorang dari bahasa tulisan. Analisis sentimen merupakan salah satu bidang penelitian yang paling aktif dalam pemrosesan bahasa alami dan juga banyak dipelajari dalam *data mining*, *web mining*, dan *text mining* (Liu, 2012).

Analisis sentimen juga menggunakan algoritma untuk mengolah dan melakukan klasifikasi terhadap data yang dibangun. Terdapat banyak algoritma yang dapat digunakan dalam penelitian analisis sentimen. Ada sepuluh algoritma terbaik yang biasa digunakan, di antaranya adalah *C4.5*, *The K-Means*, *Support Vector Machine*, *Apriori*, *Maximum Entropy PageRank*, *AdaBoost*, *k-nearest neighbor*, *Naive Bayes*, *CART*.

2.4 Algoritma Naïve Bayes

Bayesian classification adalah suatu metode pengklasifikasian data dengan model statistik yang dapat digunakan untuk menghitung probabilitas keanggotaan suatu kelas. Metode Bayesian classification digunakan menganalisis dalam membantu tercapainya pengambilan keputusan terbaik suatu permasalahan dari sejumlah alternatif. Pada pengklasifikasian menggunakan Naïve Bayes dibagi kedalam 2 proses, yaitu proses training dan testing. Proses training digunakan untuk menghasilkan model analisis sentimen yang nantinya akan digunakan sebagai acuan untuk mengklasifikasikan sentimen dengan data testing atau data mentah yang baru.

$$P(H | \mathbf{X}) = \frac{P(\mathbf{X} | H)P(H)}{P(\mathbf{X})} = P(\mathbf{X} | H) \times P(H) / P(\mathbf{X})$$

Keterangan :

- ❖ X → Data dengan class yang belum diketahui
- ❖ H → Hipotesis data X yang merupakan suatu class yang lebih spesifik
- ❖ $P(H|X)$ → Probabilitas hipotesis H berdasarkan kondisi X (*posteriori probability*)
- ❖ $P(H)$ → Probabilitas hipotesis H (*prior probability*)
- ❖ $P(X|H)$ → Probabilitas X berdasarkan kondisi pada hipotesis H

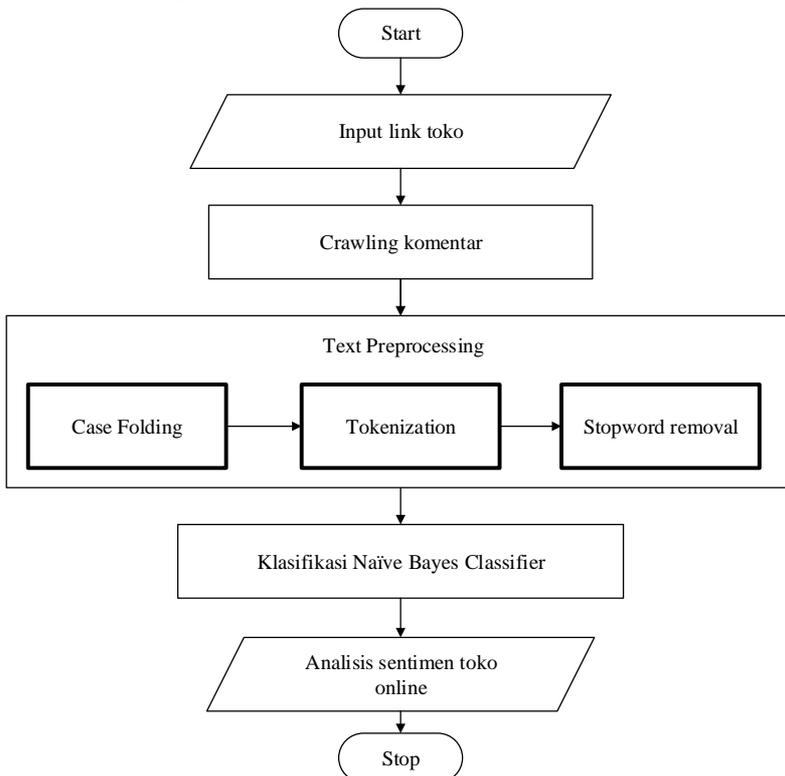
❖ $P(X) \rightarrow$ Probabilitas X

Tahapan algoritma naive bayes adalah sebagai berikut :

1. Baca Data Training
2. Hitung jumlah class
3. Hitung jumlah kasus yang sama dengan class yang sama
4. Kalikan semua nilai hasil sesuai dengan data X yang dicari class-nya

2.5 Skema Sistem

Skema aplikasi sentimen analisis toko *online* menggunakan algoritma *Naïve Bayes Classifier* dapat dilihat pada gambar 3.3 berikut.



Skema aplikasi di atas adalah skema yang menggambarkan mengenai proses perancangan keseluruhan sistem dari tahap *Start* hingga *Stop*.

a. *Start*

Merupakan proses inialisasi awal memulai aplikasi.

b. *Input link toko*

Merupakan proses memasukkan link toko untuk dilakukan crawling komentar yang digunakan sebagai data testing.

c. *Crawling komentar*

Merupakan proses untuk mengambil komentar pada produk yang ada dari link toko yang dimasukkan.

d. *Text Preprocessing*

Proses untuk proses text sebelum dilakukan pengecekan kelas, yang melalui tahap sebagai berikut.

1. *Case Folding*

Merupakan proses untuk mengubah semua huruf menjadi huruf kecil (lowercase).

2. *Tokenization*

Merupakan tahap pemotongan kata berdasarkan setiap kata penyusunnya menjadi potongan tunggal yang dipisahkan dengan spasi.

3. *Stopword Removal*

Merupakan tahap menghilangkan kata yang tidak diperlukan dalam kalimat.

e. *Klasifikasi Naïve Bayes Classifier*

Merupakan proses mengklasifikasikan komentar kedalam sentiment positif atau negatif.

f. *Analisis sentiment toko online*

Merupakan proses untuk menampilkan hasil klasifikasi yang telah dilakukan.

g. *Stop*

Merupakan proses perhentian dari aplikasi.

3. HASIL DAN PEMBAHASAN

3.1 Data Training dan Data Uji

Data training merupakan data yang digunakan untuk melatih algoritma. Jumlah data yang digunakan dalam pelatihan algoritma naïve bayes yang peneliti gunakan yaitu sebanyak 1000 data, dengan jumlah data training positif 500 kalimat dengan banyak kata dalam data training positif adalah 3628 kata,

data training negatif 500 kalimat dengan banyak kata dalam data training negatif adalah 3555 kata dan total keseluruhan kata sebanyak 7183 kata.

Data uji merupakan data yang digunakan untuk menguji algoritma naïve bayes dengan memanfaatkan data training atau data latih sebelumnya untuk analisis sentimen. Adapun data uji yang digunakan yaitu data 100 komentar yang diambil dari satu toko dengan 10 produk dan 10 komentar perproduknya.

3.2 Penerapan Naive Bayes

Perhitungan manual ini penulis ambil beberapa sampel untuk dihitung manual. Total data training 1000 data dengan 500 data training positif dan 500 data training negatif. Total kata pada training positif yaitu 3628 kata, sedangkan kata dalam training negatif yaitu 3555 kata. Berikut perhitungan manual naïve bayes classifier berdasarkan data uji di atas yang sudah dilakukan tahap text preprocessing. Misal kalimat “ barang bagus”.

kata	P(X)pos	P(X)neg	P(Ci)	P(X Ci)pos	P(X Ci)neg
barang	111	97	0.5	0.01	0.009
bagus	237	33	0.5	0.02	0.003

$$1. P(X|Ci) * P(Ci)$$

$$Positif = (0.01 * 0.02) * 0.5$$

$$Positif = 0.0001$$

$$Negatif = (0.009 * 0.003) * 0.5$$

$$Negatif = 1.35E - 5$$

Karena nilai posterior positif lebih tinggi daripada nilai posterior negatif, maka kalimat komentar “kualitas oke harga pengiriman cepat next time beli” diklasifikasikan sebagai analisis sentimen positif.

3.3 Pengujian sistem

Pengujian pada toko *online* dilakukan dengan menguji beberapa toko yang terdapat di Shopee. Pengujian ini dilakukan untuk melihat persentase pada setiap toko berdasarkan komentar. Detail dari beberapa pengujian pada toko di Shopee adalah sebagai berikut :

No	Username Toko	Nama Toko	Jumlah Komen	Sentimen Positif	Sentimen Negatif	Persentase
1.	Anekafashionstore	Anekafashionstore	100	80	20	80%
2.	alfargalery	Alfargalery	99	81	18	81,82%
3.	karaaa.id	Kara.id	99	81	18	81,82%
4.	bajukitaindonesia_	Sweatergrossir	100	68	32	68%
5.	batik_graha	Batik Graha	100	77	23	77%
6.	Agoenksu	Predz_id	100	86	14	86%

7.	priadi desain99	Project 99	99	76	23	76,77%
8.	erigostore	ERIGO Official Shop	100	82	18	82%
9.	kiosbalitafawa	Kios Balita Fawa	99	82	17	82,83%
10.	heybaby.store	HEYBABY STORE	100	90	10	90%

Dari pengujian beberapa toko online di Shopee dapat disimpulkan bahwa penelitian ini dapat di uji pada berbagai macam jenis toko pakaian, seperti toko pakaian wanita, toko pakaian pria dan toko pakaian anak –anak. Persentase sentimen positif yang paling tinggi diperoleh oleh toko pakaian anak – anak dari 100 komentar menghasilkan 90% sentimen positif.

3.4 Keakurasian Sistem

Pengujian akurasi dilakukan dengan beberapa pengujian data testing dengan menggunakan data training dengan jumlah yang berbeda. Akurasi merupakan persentase dari total sentimen yang benar dikenali. Perhitungan akurasi dilakukan dengan cara membagi jumlah data sentimen yang benar dengan total datadan data uji.

Pengujian	Data Set		Data Testing		Jumlah Sentimen		Persentase
	Pos	Neg	Pos	Neg	Benar	Salah	
I	100	100	82	18	86	14	86%
II	100	200	67	33	73	27	73%
III	200	100	89	11	95	5	95%
IV	200	200	77	23	85	15	85%
V	300	200	87	13	93	7	93%
VI	300	300	84	16	90	10	90%
VII	400	300	88	12	94	6	94%
VIII	400	500	78	22	84	16	84%
IX	500	400	87	13	93	7	93%
X	500	500	79	21	87	13	87%

Persentase keakurasian terbaik ketika data training positif lebih banyak daripada data training negatif. Sedangkan apabila data training positif sama dengan data training negatif maka keakurasiannya bersifat stabil.

4. Kesimpulan

Pada penelitian analisis sentimen toko online menggunakan algoritma naive bayes menggunakan 1000 data latih, dimana 500 merupakan data latih sentimen positif dan 500 sisanya merupakan data latih sentimen negatif. Data uji merupakan toko – toko pakaian dengan pengambilan 10 produk dan 10 komentar perproduknya, sehingga ada 100 komentar yang dilakukan untuk sekali pengujian. Penelitian ini dapat di uji pada berbagai macam jenis toko pakaian, seperti toko pakaian wanita, toko pakaian pria dan toko pakaian anak –anak. Persentase sentimen positif yang paling tinggi diperoleh oleh toko pakaian anak – anak dari 100 komentar menghasilkan 90% sentimen positif. Persentase keakurasian terbaik didapatkan ketika data training positif lebih banyak daripada data training negatif. Sedangkan apabila data training positif sama dengan data training negatif maka keakurasiannya bersifat stabil. Pada penelitian ini persentase keakurasian analisis sentimen toko online dengan menggunakan algoritma naive bayes adalah 87 %.

DAFTAR PUSTAKA

- Hemalatha, I., Varma, P.G., dan Govardhan, A., (2012), Preprocessing the Informal Text for Efficient Sentiment Analysis, *International Journal of Emerging Trends & Technology in Computer Science (IJETTCS)*, vol. 1, July-August 2012, ISSN 2278-6856.
- Ismail, M.M., dan Lhaksamana, K.M., (2019). Sentimen Analisis pada Media Online Mengenai Pemilihan Presiden 2019 dengan Menggunakan Metode Naive Bayes. *e-proceeding of Engineering*, vol.6, Agustus 2019, ISSN 2355-9305
- Liu, B, (2012), *Opinion Mining*. Chicago, United States of America.
- Ratnawati, F., (2018), Implementasi Algoritma Naive Bayes Terhadap Analisis Sentimen Opini Film pada Twitter, *Jurnal Inovtek Polbeng-Seri Informatika*, vol. 3, Juni 2018, ISSN 2527-9866.
- Rish, I., An Empirical Study of The *Naive Bayes Classifier*, In *Proceedings of IJCAI-01 workshop on Empirical Methods in AI*, New York, 2001
- Suryani, N.P., Linawati., dan Saputra, K.O., (2019). Penggunaan Metode Naive Bayes Clasifier pada Analisis Sentimen Facebook Berbahasa Indonesia. *Majalah Ilmiah Teknologi Elektro*, vol. 18, Januari-April 2019, ISSN 1693-2951
- Triawati, C, (2009), *Text Mining*. Bandung, JawaBarat, Indonesia.
- Turban, E.; et.al. *Decision Support and Business Intelligence Systems (edisi ke-9)*. New Jersey: Pearson Education, Inc. 2011.