

RANCANG BANGUN MODEL SENTIMEN ANALISIS *REVIEW* PRODUK PADA TOKO *ONLINE* MENGGUNAKAN *NAIVE BAYES*

Erawati Kartika¹, Jutono Gondohanindijo²

¹Program Studi Akuntansi, Universitas AKI Semarang

²Program Studi Teknik Informatika, Universitas AKI Semarang

Email: erawati.kartika@unaki.ac.id

ABSTRACT

The industrial revolution 4.0 encourages a lifestyle, especially shopping that was previously done offline, is now done online through online stores. In addition, various social businesses have emerged, such as selling on Facebook, which produce unlimited textual data, so that a presentation of information is needed without distorting the value of the information. One of the things that concerns consumers are reviews or comments on products. This study aims to make a system design or model capable of classifying sentiment analysis reviews or reviews of an online store product, namely positive or negative sentiments. Sentiment analysis stages consist of data collection, initial processing, data transformation, feature selection and classification stages using the naïve Bayes method. Data collection techniques using observation, interviews, and documentation. The initial processing applies the text mining stages, namely case folding, non alpha numeric removal, stop words removal, and stemming. The results of the sentiment analysis were measured using the Accuracy, Precision, Recall and F-measure criteria. The results of this study are a prototype product and a sentiment classification system model for online shop product analysis. The results of the test analysis with the Confusion Matrix showed an accuracy of 84.9%.

Keywords : *Sentiment Analysis; Product; Social Business; Online Shop; Naïve Bayes*

ABSTRAK

Revolusi industri 4.0 mendorong gaya hidup khususnya belanja yang sebelumnya dilakukan secara *offline* kini dilakukan secara *online* melalui toko *online*. Disamping itu muncul berbagai *social business* seperti jualan di *Facebook* yang memproduksi data tekstual tanpa batas, sehingga diperlukan sebuah penyajian informasi tanpa mendistorsi nilai dari informasi tersebut. Salah satu hal yang menjadi perhatian konsumen adalah *review* atau komentar terhadap produk. Penelitian ini bertujuan membuat sebuah rancang bangun sistem atau model yang mampu melakukan klasifikasi sentimen analisis *review* atau ulasan suatu produk toko *online*, yaitu sentimen positif atau negatif. Tahapan analisis sentimen terdiri dari pengumpulan data, pemrosesan awal, transformasi data, seleksi fitur dan tahapan klasifikasi menggunakan metode naïve bayes. Teknik pengumpulan data menggunakan observasi, *interview*, dan dokumentasi. Pemrosesan awal menerapkan tahapan *text mining* yakni *case folding*, *non alpha numeric removal*, *stop words removal*, dan *stemming*. Hasil analisis sentimen diukur menggunakan kriteria Akurasi, Presisi, *Recall* dan *F-measure*. Hasil dari penelitian ini berupa produk *prototype* dan model sistem klasifikasi sentimen analisis produk toko *online*. Hasil analisis pengujian dengan *Confussion Matrix* menunjukkan akurasi sebesar 84.9%.

Kata Kunci: *Sentimen Analisis; Produk; Social Business; Toko Online; Naïve Bayes*

PENDAHULUAN

Dijaman sekarang ini siapapun dapat membuka toko *online* baik melalui *website*, *e-commerce* maupun melalui media sosial seperti *Facebook*, *Instagram* maupun *Web Toko Online*. Pembeli dapat melakukan kunjungan ke toko *online* sambil duduk nyaman di depan komputer atau tiduran sambil pegang *smartphone*. Selanjutnya transaksi dapat dibayarkan melalui uang elektronik maupun melalui toko ritel yang bekerja sama sehingga efektif dan efisien.

Dari fakta bahwa sebanyak 75% konsumen *online* sebelum melakukan transaksi mereka melakukan investigasi ke toko bersangkutan, dan semuanya setuju bahwa 45% informasi di dapatkan dari media sosial untuk mendapatkan ulasan produk yang akan dibeli. Ulasan yang didapat di media sosial pun beragam mulai dari pujian maupun komplain terhadap produk yang dijual (Nielsen, 2014).

Dengan banyaknya komentar yang diberikan pengguna di internet, masyarakat harus jeli dan memilah-milah informasi tentang *review* atau ulasan sebuah produk toko *online*. Ini membutuhkan banyak waktu untuk pengguna membaca dan mengamati semua *review* / ulasan pengguna lain yang ada di internet terhadap produk yang akan dicarinya, sehingga dengan adanya sistem *review* / ulasan ini pengguna dapat dengan segera mengetahui ulasan baik atau tidaknya sebuah produk yang diberikan oleh sebuah toko *online*.

Penelitian analisis sentimen ini menganalisis sentimen untuk menentukan kecenderungan pandangan publik terhadap *review* produk atau barang pada toko *online* menggunakan metode klasifikasi Naïve Bayes, yaitu suatu algoritma yang digunakan untuk *classification*.

Preprocessing dilakukan untuk menghindari data yang kurang sempurna, gangguan pada data, dan data-data yang tidak konsisten (Han, 2012). Tahap preprocessing pada

penelitian ini diantaranya: *Cleansing*: tahap menghapus tanda baca dan angka yang terdapat pada teks komentar, *Case folding*: yaitu penyeragaman bentuk huruf menjadi *lower case* atau *upper case* serta penghapusan angka dan tanda baca, *Filtering*: *Filtering* dilakukan dengan menggunakan algoritma *stop word removal*. *Stopword removal* digunakan untuk membuang kata-kata yang sering muncul dan bersifat umum, kurang menunjukkan relevansinya dengan teks. *Stemming*, adalah proses mencari akar kata dan menghilangkan imbuhan pada kata. *Stemming* bertujuan mengurangi variasi kata yang memiliki kata dasar sama. *Convert negation*: proses mengganti negasi yang terdapat dalam komentar. Negasi adalah sesuatu yang dikenal dalam semua bahasa dan biasanya negasi digunakan untuk mengubah polaritas dari suatu pernyataan. Contoh kata-kata yang bersifat negasi adalah “kurang”, “tidak”, “enggak” dan “tak”.

Metode *Naïve Bayes Classifier* (NBC), merupakan suatu metode klasifikasi sederhana berdasarkan aplikasi teorema Bayes dengan asumsi antar variabel penjelas saling bebas (independen). Teorema bayes berfungsi untuk memprediksi peluang di masa depan berdasarkan pengalaman di masa sebelumnya (Zhang, 2016). Metode pengklasifikasian tersebut dikemukakan oleh ilmuwan Inggris Thomas Bayes. Berikut ini rumus yang digunakan:

$$P(C|X) = \frac{P(X|C)P(C)}{P(X)} \quad (1)$$

Keterangan:

X = peubah acak

C = kelas

$P(C | X)$ = probabilitas C didasarkan pada kondisi X

$P(X | C)$ = probabilitas X didasarkan pada kondisi C

Perhitungan yang dibutuhkan dalam NBC yaitu probabilitas tiap kelas, rumus mencari probabilitas tiap kelas dinotasikan sebagai berikut:

$$P(C) = \frac{N_c}{N} \quad (2)$$

Keterangan:

$P(C)$ = probabilitas C

N_c = jumlah total kelas C

N = jumlah total keseluruhan kelas

Evaluasi Performansi dilakukan untuk melihat performansi model klasifikasi dalam mencari hasil klasifikasi. Cara yang sering digunakan adalah menghitung akurasi, *precision*, *recall* dan *f-measure*. Untuk mempermudah perhitungan dibuatkan *Confussion Matrix* yang didapat dari hasil prediksi dan data aktual.

Tabel 1. *Confussion Matrix*

	Kelas Positif Prediksi	Kelas Negatif Prediksi
Kelas Positif Aktual	<i>True Positif (TP)</i>	<i>False Negatif (FN)</i>
Kelas Negatif Aktual	<i>False Positif (FP)</i>	<i>True Negatif (TN)</i>

Akurasi merupakan persentase dari suatu prediksi model yang sudah dibuat.

$$akurasi = \frac{(TP+TN)}{(TP+TN+FP+FN)} \quad (3)$$

Precision merupakan tingkat ketepatan antara informasi yang diminta oleh pengguna dengan jawaban yang diberikan oleh sistem.

$$precision = \frac{TP}{(FP+TP)} \quad (4)$$

Recall merupakan kualitas seberapa lengkap hasil relevan yang ditampilkan oleh sistem prediksi kelas.

$$recall = \frac{TP}{(FN+TP)} \quad (5)$$

F-measure merupakan salah satu perhitungan evaluasi dalam informasi temu kembali yang mengkombinasikan *recall* dan *precision*.

$$f - measure = \frac{(2 \cdot recall \cdot precision)}{(recall+precision)} \quad (6)$$

TP adalah hasil prediksi positif dan nilai aktual positif, FP adalah hasil prediksi positif dan nilai aktual negatif, FN adalah hasil prediksi negatif dan aktual positif, dan TN adalah hasil prediksi negatif dan nilai aktual negatif.

METODE PENELITIAN

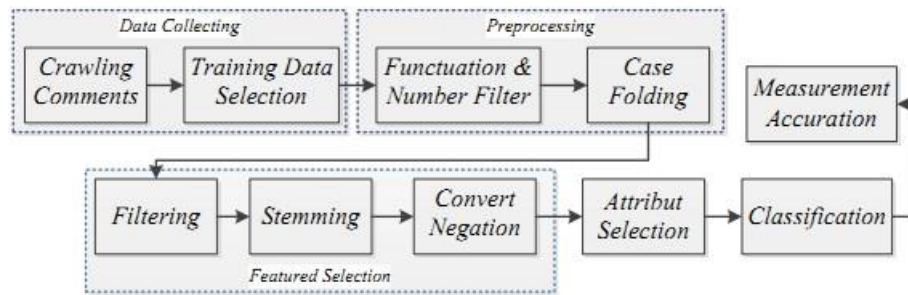
Metode penelitian berisi bahan-bahan utama yang digunakan dalam penelitian dan metode-metode yang digunakan dalam pemecahan permasalahan termasuk metode analisis.

Bahan penelitian ini berupa data yang diambil dari halaman *Web Toko Online SAMTHING* yang menjual produk-produk fashion. Data yang diperoleh berupa 210 komentar postingan yang terdiri dari 105 komentar positif dan 105 komentar negatif. Spesifikasi alat uji yang digunakan berupa satu set komputer.

Tabel 2. Spesifikasi komputer yang digunakan

Processor	Intel(R) Core (TM) i5 CPU
Memory	8.00 GB
Harddisk	1 TB
Sistem Operasi	Windows 7 64 bit
Bahasa Program	Python 3.8.3

Metode analisis dalam penelitian ini digambarkan dengan Diagram Alir proses klasifikasi sentimen analisis sebagai berikut:



Gambar 1. Diagram Alir Klasifikasi Sentimen Analisis

Alur Proses Klasifikasi Sentimen Analisis ini dimulai dari pengumpulan data komentar, preprocessing, featured selection, pemilihan atribut, klasifikasi, dan pengujian akurasi.

Data collecting merupakan tahapan pengumpulan *training* data menyaring *keywords* yang berhubungan dengan produk *fashion* dikombinasikan dengan kata-kata sentimen. Berikut ketentuan yang digunakan dalam memilih data training :

1. Menentukan target / entitas. Target entitas diperoleh setelah dilakukan frequent itemset terhadap data komentar yang masih belum melalui *preprocessing* :
 - Produk *fashion*, yaitu terkait kualitas produk, dan keyword terkait produk.
2. Menentukan *keyword* sentimen positif / negatif yang mengiringi suatu kalimat komentar :
 - a. Produk-sentimen-positif
 - b. Produk-sentimen-negatif
3. Melakukan *filtering* kata target / entitas dan sentimen ada dalam satu kalimat komentar: Jika salah satu tidak ada, maka tidak layak menjadi data training.

Tahap preprocessing diperlukan untuk membersihkan data dari noise, menyeragamkan bentuk kata dan mengurangi volume kata. Agar pada tahap masuk ke dalam metode klasifikasi lebih optimal dalam perhitungannya. Tahap feature selection merupakan tahap lanjut dari pengurangan dimensi pada proses transformasi teks.

Sebelum hasil *preprocessing* diproses dengan algoritma klasifikasi, data training akan dilakukan proses penentuan atribut (*attribut selection*). Tahap ini akan menyaring setiap kata yang muncul berdasarkan jumlah kemunculan kata.

Proses klasifikasi merupakan tahapan utama untuk mengetahui pola sentimen dari komentar-komentar yang telah *dipreprocessing*. Klasifikasi dilakukan dengan menggunakan metode Naïve Bayes. Evaluasi Performansi dilakukan dengan menghitung akurasi, *precision*, *recall* dan *f-measure*.

Dalam penyusunan penelitian ini, dibutuhkan data-data serta materi dan referensi yang mendukung. Beberapa metode yang digunakan adalah :

1. Studi Kepustakaan,

Studi kepustakaan yang dilakukan dengan mengumpulkan data melalui buku-buku, jurnal, internet, atau media informasi lain yang berhubungan dengan topik penelitian sebagai acuan dan landasan teori.

2. Pengumpulan Data,

- a. Wawancara pada pemilik toko dan pengguna.
- b. Analisa data statistic terkait *review* sebelum dan sesudah melakukan pengerjaan sistem.
- c. Analisa metode eksperimen.

3. Analisa data dan hasil eksperimen

HASIL DAN PEMBAHASAN

Sumber data teks yang akan diolah pada penelitian ini berasal dari halaman *web* toko *online* “SAMTHING” yang berlokasi di Salatiga. Data teks diambil dari komentar-komentar terkait produk dari pengunjung *web*. Jumlah data yang diambil sebanyak 210 data komentar selama tiga bulan, dari bulan September 2020 sampai dengan Nopember 2020.

A	
94	Barang bagus, mantap!
95	Desain bagus
96	Bahannya bagus, nggak gampang luntur
97	Memuaskan, kaosnya sesuai dengan pesanan
98	Kaosnya keren banget kak!
99	Jaketnya bagus banget!
100	Bahannya nyaman dipakeeee
101	Bajunya bagus
102	Tidak panas dipakai
103	Nyaman dipakai, sejuk
104	Bajunya enak dan nyaman dipakai
105	Sejuk dan tidak panas, bahannya bagus
106	Pesanan sesuai dengan yang diharapkan
107	ukuran LD nya ga sesuai deskripsi, biasanya pake LD 104-106 masih longgar. ini deskripsi uk LD nya 110 tp kekecilan, positif aja mungkin e
108	Hitamnya kayaknya luntur
109	Sayang sekali Warna salem lebih dominan ke transparan soalnya pas aku pake bajunya keliatan dalemnya
110	Kualitas produk kurang baik
111	Potongan bajunya aneh, gak kepake deh ??
112	Bajunya kaku bgt sumpah nyesell bgt
113	Bahannya panas ga nyerap keringat
114	Setelh di cuci & mau sy pake ternyt amat sangat mengecewakan. Jahitannya mengkerut (foto no 1), sambungan antra badan & lengan buk
115	Kaosnya ada yang luntur kak, jadi jelek nih.. :(
116	Benangnya ada yang lepas
117	Barang bagus, Cuma packing nya aja yang kurang rapi, ekspedisi juga terlalu lama, mungkin lain kali bisa bilang ke ekspedisinya agar lebih
118	Bajunya kegedean
119	Bajunya kekecilan woyy... gua pesen yang L yang datang malah S.. >:(

Gambar 2. Data Komentar Toko Online SAMTHING

Preprocessing dilakukan menggunakan Bahasa Program Python. *Preprocessing* diawali dari proses *cleansing* hingga *stopword removal/filtering*. Dilanjutkan dengan *stemming* dan *tokenizing*. Hasilnya adalah data komentar dengan tingkat *noise* lebih rendah.

```

[[('bagus', 2), ('aja', 2), ('beli', 2), ('harga', 1), ('gin', 1), ('tebel', 1), ('saran', 1), ('warna', 1), ('hitam', 1), ('kualitas', 1), ('ok', 1), ('produk', 1), ('original', 1), ('kirim', 1), ('cpt', 1)]
[[('harga', 3), ('produk', 2), ('sesuai', 2), ('barang', 2), ('bagus', 2), ('kualitas', 1), ('cepat', 1), ('kirim', 1), ('bagus', 1), ('bahan', 1), ('ukur', 1), ('jga', 1)]
[[('ajegileeee', 1), ('bagus', 1), ('banget', 1), ('bahan', 1), ('nyaa', 1), ('ademm', 1), ('suka', 1)]
[[('produk', 3), ('kualitas', 1), ('original', 1), ('harga', 1), ('cepat', 1), ('kirim', 1), ('ekspedisi', 1), ('tok', 1), ('kak', 3), ('donkk', 2), ('barang', 1), ('thanks', 1), ('kualitas', 1), ('produk', 1), ('original', 1), ('sesuai', 1), ('recommended', 1), ('muas', 1), ('customer', 1), ('puas', 1), ('sukaa', 1), ('moga', 1), ('kirim', 1), ('cepat', 1), ('banget', 2), ('bagus', 1), ('baju', 1), ('sukaasa', 1), ('kali', 1), ('order', 1), ('emang', 1), ('cucok', 1), ('bagus', 1), ('bahan', 1), ('terawang', 1), ('sesuai', 1), ('gambar', 1), ('pas', 1), ('bad', 1)]
[[('ld', 3), ('nya', 2), ('deskripsi', 2), ('ukur', 1), ('ga', 1), ('sesuai', 1), ('pake', 1), ('104106', 1), ('long', 1), ('hitam', 1), ('nya', 1), ('kya', 1), ('yg', 1), ('luntur', 1), ('kecewa', 1), ('kaya', 1), ('kemeja', 1), ('cowok', 1), ('sayang', 1), ('warna', 1), ('salem', 1), ('dominan', 1), ('transparan', 1), ('pas', 1), ('pake', 1), ('bajumya', 1), ('kualitas', 1), ('produk', 1), ('cepat', 1), ('kirim', 1)]
[[('potong', 1), ('baju', 1), ('aneh', 1), ('gk', 1), ('kepake', 1), ('deh', 1)]
[[('respon', 1), ('jual', 1), ('kacau', 1), ('parah', 1), ('w', 1), ('sen', 1), ('tanggal', 1), ('27', 1), ('mei', 1), ('g', 2), ('ld', 1), ('100', 1), ('barang', 1), ('nyampe', 1), ('muat', 1), ('alhasil', 1), ('kepake', 1)]
[[('bgt', 2), ('baju', 1), ('kaku', 1), ('sumpah', 1), ('nyesell', 1)]
[[('bahan', 1), ('panas', 1), ('ga', 1), ('nyerap', 1), ('keringat', 1)]
[[('no', 2), ('sambung', 2), ('lengan', 2), ('yg', 2), ('setelh', 1), ('cuci', 1), ('sy', 1), ('pake', 1), ('ternytt', 1)]
    
```

Gambar 3. Proses *Preprocessing*

Sebelum masuk ke proses klasifikasi, data atribut harus diberi label sentimen seperti pada Gambar 4.

	A	B
4	Warnanya bagus sesuai dengan iklan	1
5	Sesuai harapan saya, mantap desainnya	1
6	Lanjutkan! Pertahankan kualitasnya gan!	1
7	Lain kali mau pesen lagi lah, bagus kaosnya	1
8	Sesuai pesanan	1
9	Mantaaaaappppp.... Keren banget kaosnya!!	1
10	Jaketnya bagus!	1
11	Bahannya lembut bangeett	1
12	Overall bagus	1
13	Kaosnya bagus, pesen lagi ah	1
14	Kaosnya ada yang luntur kak, jadi jelek nih.. :(-1
15	Benangnya ada yang lepas	-1
16	Barang bagus, Cuma packing nya aja yang kurang rapi, ekspedisi juga terlalu lama	-1
17	Bajunya kegedean	-1
18	Bajunya kekecilan woyy... gua pesen yang L yang datang malah S.. >:(-1
19	Haduhhh... kaosnya lunturr... :(-1
20	Ini bajunya kekecilan atau aku yang gendutan yah? :(-1
21	Jahitannya kok gak rapi.	-1
22	Yang jual kurang bagus responnya	-1
23	Apaan nih, baju kegedean..	-1
24	Niat jualan gak sih?! Jahitannya jelek!	-1
25	Denial banyak alasan, hilang aja nggak ready kan udah selesai nake acara suruh	-1

Gambar 4. Label Klasifikasi

Pada tahap pengujian klasifikasi, pengujian dilakukan dengan *k-Fold Cross Validation* yang berfungsi menentukan data *training* dan data *testing*.

Setelah dilakukan proses klasifikasi menggunakan algoritma naïve bayes, hasil evaluasi adalah sebagai berikut :

Tabel 3. Accuracy

```
[31] from sklearn.metrics import accuracy_score
print(accuracy_score(predict, Y_test))

0.8490566037735849
```

Pada tabel 3 nampak Algoritma Naïve Bayes menghasilkan nilai akurasi sebesar 84,9%, dan parameter evaluasi yang lain nampak pada tabel 4 berupa Report *Classification* dengan nilai *precision* 83% (sentimen negatif) dan 87% (sentimen positif), *recall* 89% (

sentimen negatif) dan 80% (sentimen positif), F Measure (f1-score) 86% (sentimen negatif) dan 83% (sentimen positif) pada proses klasifikasi sentimen positif atau negatif.

Tabel 4. Report *Classification*

```
[32] print(classification_report(predict, Y_test))
```

	precision	recall	f1-score	support
-1	0.83	0.89	0.86	28
1	0.87	0.80	0.83	25
accuracy			0.85	53
macro avg	0.85	0.85	0.85	53
weighted avg	0.85	0.85	0.85	53

Nilai *Accuracy* dan Report *Classification* diambil dari Tabel Confusion Matrix (Tabel 5) dengan prediksi data testing sebanyak 25% dari data latih yaitu sebesar 53 data.

Tabel 5. Confusion Matrix

```
[30] confusion_matrix(predict, Y_test)
```

```
array([[25,  3],  
       [ 5, 20]])
```

SIMPULAN DAN SARAN

Dari analisis pengujian data yang dilakukan terkait sentimen terhadap Toko *Online* “SAMTHING” dengan menggunakan algoritma Naive Bayes dapat disimpulkan bahwa akurasi klasifikasi sentimen positif atau negatifnya adalah sebesar 84.9%.

Untuk meningkatkan akurasi kita bisa mencoba dengan membuat label data klasifikasi yang berimbang, sehingga bisa di analisa lebih jauh pengaruhnya terhadap akurasi. Analisis akurasi juga bisa dicoba dengan menambah data latih lebih banyak atau dengan memberikan algoritma pembobotan terhadap data hasil *preprocessing*.

Analisis sentimen ini sebaiknya diimplementasikan dengan lebih banyak data dalam jangka waktu beberapa bulan, sehingga akan memungkinkan melihat sentimen publik dari bulan kebulan.

DAFTAR PUSTAKA

- Firmansyah, Dimas Ricky. (2013). Implementasi Text Mining Klasifikasi Objek Wisata Dengan Metode Naive Bayes Classifier di Dinas Pariwisata dan Kebudayaan Jawa Barat. *Bandung : Universitas Komputer Indonesia*.
- Haddi, Emma., Liu, Xiaohui., & Yong Shi. (2013). The Role of Text Pre-processing in Sentiment Analysis. *Netherlands : Elsevier B.V.*
- Hamzah, Amir. (2014). Sentiment Analysis untuk Memanfaatkan Saran Kuesioner Dalam Evaluasi Pembelajaran dengan Menggunakan Naive Bayes Classifier. *Yogyakarta : IST AKPRIND*.
- Han, Jiawei., Kamber, Micheline., & Pei, Jian. (2012). *Data Mining Concepts and Techniques (3rded.)*. USA : Simon Frase University.
- Hayuningtyas, R., Y., Sari, R., (2019). Analisis Sentimen Opini Publik Bahasa Indonesia Terhadap Wisata TMII menggunakan Naive Bayes dan PSO. *Jurnal TECHNO Nusa Mandiri* Vol. 16, No. 1 Maret 2019.
- Hemalatha, I., Varma, G. P Saradhi., & Govardhan, A. (2012). Preprocessing the Informal Text of efficient Sentiment Analysis. India : *International Journal of Emerging Trends & Technology in Computer Science*.
- Ma, Jian dkk. (2012). An Ontology-Based Text-Mining Method to Cluster Proposals for Research Project Selection. *IEEE Trans. Syst., Man,Cybern. Part A, Syst & Humans* (vol. 42).
- Nielsen. (2014). *E-Commerce: Evolution or Revolution in the Fast Moving Consumer Goods World*. United States: New York
- Ramadhan, S., K., Muslim, L., K., (2018). Analisis Sentimen terhadap Toko Online menggunakan Naive Bayes pada Media Sosial Twitter. ISSN : 2355-9365 *e-Proceeding of Engineering* : Vol.5, No.3 Desember 2018.
- Syah, A., P., Adiwijaya, Faraby, S., A., (2017). Analisis Sentimen Pada Data Ulasan Produk Toko Online Dengan Metode Maximum Entropy. ISSN : 2355-9365 *e-Proceeding of Engineering* : Vol.4, No.3 Desember 2017.

Wati, R. (2016). Penerapan Algoritma Genetika Untuk Seleksi Fitur Pada Analisis Sentimen *Review* Jasa Maskapai Penerbangan menggunakan Naive Bayes. *Jurnal Evolusi*.

Zhang, Lei, and Bing Liu. 2016. "Sentiment Analysis and Opinion Mining." *Encyclopedia of Machine Learning and Data Mining*, no. May: 1–10. https://doi.org/10.1007/978-1-4899-7502-7_907-1.