

Analisis Sentimen Ulasan Pengguna Aplikasi Shopee Indonesia Menggunakan Metode Naive Bayes

Muh.Novran Ramjani Rumkorem¹, Abigael Yarangga², Klemensia Dina Irma Indriani Solo³, Wedes F.A. Iwanggin⁴, Zwingly Isak Rumaseuw⁵, Tinglizhau⁶, Gracella Bakarbresy Paiki⁷, Riski Sulobua⁸, Heru Sutejo⁹

^{1,2,3,4,5,6,7,8,9} Universitas Sepuluh Nopember Papua

* Correspondence e-mail; novranrumkorem21@gmail.com

Article history

Submitted: 2024/12/08; Revised: 2024/12/15; Accepted: 2024/12/16:

Abstract

Penelitian ini bertujuan untuk menganalisis sentimen ulasan pengguna aplikasi Shopee Indonesia dengan menggunakan metode Naive Bayes. Data ulasan pengguna yang diperoleh kemudian diproses menggunakan RapidMiner untuk membangun model klasifikasi sentimen. Berdasarkan hasil pengujian menggunakan confusion matrix, model yang dihasilkan memperoleh tingkat akurasi sebesar 74.67%. Dalam analisis sentimen positif, model mencatatkan nilai recall sebesar 25.19% dan precision sebesar 82.93%, menunjukkan bahwa meskipun model memiliki ketepatan yang tinggi dalam memprediksi sentimen positif, kemampuannya dalam mendeteksi sentimen positif secara keseluruhan masih rendah. Sebaliknya, untuk sentimen negatif, model memperoleh recall sebesar 94.21% dan precision sebesar 53.02%, yang menunjukkan bahwa model lebih efektif dalam mendeteksi sentimen negatif, meskipun ketepatannya lebih rendah dibandingkan dengan sentimen positif. Hasil ini menggambarkan bahwa meskipun model memiliki kelebihan dalam mengidentifikasi sentimen negatif, masih terdapat kelemahan dalam memprediksi sebagian besar data sentimen positif. Penelitian ini memberikan wawasan penting terkait potensi dan tantangan dalam menerapkan metode Naive Bayes untuk analisis sentimen ulasan pengguna aplikasi e-commerce.

Keywords

analisis sentimen, Shopee, Naive Bayes, RapidMiner, e-commerce



© 2024 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution 4.0 International (CC BY SA) license, <https://creativecommons.org/licenses/by-sa/4.0/>.

PENDAHULUAN

Kemajuan teknologi informasi dan komunikasi telah membawa perubahan signifikan dalam berbagai bidang, termasuk sektor perdagangan elektronik atau e-commerce. Salah satu platform e-commerce yang paling menonjol di Indonesia adalah Shopee. Shopee telah menjadi platform pilihan bagi masyarakat Indonesia untuk berbelanja online karena menyediakan berbagai produk, layanan, dan promosi yang menarik. Dengan jumlah pengguna aktif yang terus meningkat, Shopee juga

menerima banyak ulasan dari pengguna yang mencerminkan pengalaman mereka dalam menggunakan aplikasi tersebut. Ulasan ini berisi informasi penting yang dapat dimanfaatkan untuk mengevaluasi tingkat kepuasan dan preferensi pelanggan.

Menganalisis sentimen dari ulasan pengguna menjadi langkah strategis untuk memahami persepsi pelanggan terhadap aplikasi Shopee. Sentimen yang terkandung dalam ulasan, baik itu bersifat positif, negatif, maupun netral, mampu memberikan gambaran yang lebih mendalam mengenai kekuatan dan kelemahan layanan yang ditawarkan. Informasi ini sangat penting bagi pengembang aplikasi dan tim manajemen Shopee dalam merumuskan strategi peningkatan layanan, sehingga dapat menjaga kepuasan pelanggan sekaligus menarik lebih banyak pengguna baru.

Metode Naive Bayes adalah salah satu algoritma yang sering digunakan untuk analisis sentimen. Metode ini memiliki kemampuan untuk mengklasifikasikan teks dengan efisien menggunakan prinsip probabilitas. Naive Bayes bekerja dengan menghitung kemungkinan suatu teks termasuk dalam kategori tertentu berdasarkan pola kata yang ada di dalamnya. Sederhana namun efektif, algoritma ini telah menjadi pilihan utama dalam banyak penelitian di bidang pengolahan bahasa alami (Natural Language Processing/NLP).

Proses analisis sentimen dalam penelitian ini dilakukan dengan menggunakan RapidMiner, sebuah perangkat lunak yang dirancang khusus untuk keperluan data mining. RapidMiner menyediakan antarmuka yang intuitif, memungkinkan pengguna untuk membangun model analisis data tanpa perlu keterampilan pemrograman yang rumit. Dengan RapidMiner, setiap tahap dalam proses analisis, mulai dari pembersihan data (data preprocessing), pembuatan model, hingga evaluasi hasil, dapat dilakukan dengan mudah dan terstruktur. Keunggulan ini membuat RapidMiner menjadi alat yang sangat membantu, terutama bagi peneliti yang ingin fokus pada analisis tanpa terbebani oleh kompleksitas teknis.

Penelitian ini bertujuan untuk mengeksplorasi pola sentimen yang terkandung dalam ulasan pengguna aplikasi Shopee Indonesia. Dengan memanfaatkan metode Naive Bayes dan fitur-fitur yang disediakan RapidMiner, penelitian ini berupaya memberikan wawasan baru mengenai persepsi pengguna terhadap aplikasi Shopee. Selain itu, penelitian ini juga dapat menjadi referensi bagi pengembang aplikasi atau peneliti lain yang tertarik untuk mempelajari penerapan analisis sentimen dalam e-commerce.

Penelitian yang dilakukan oleh (Samsir et al., 2021) dengan judul Analisis Sentimen Pembelajaran Daring Pada Twitter di Masa Pandemi COVID-19 Menggunakan Metode Naïve Bayes menunjukkan bahwa metode Naïve Bayes efektif

dalam mengklasifikasikan sentimen. Penelitian ini mengungkap bahwa pembelajaran daring di masa pandemi memiliki 30% sentimen positif, 69% sentimen negatif, dan 1% sentimen netral selama periode penelitian tersebut.

Penelitian lainnya oleh (Rustiana & Rahayu, 2017) dengan judul Analisis Sentimen Pasar Otomotif Mobil: Tweet Twitter Menggunakan Naive Bayes menunjukkan performa akurasi yang tinggi, yaitu sebesar 93%. Precision positif dan negatif masing-masing mencapai 90%, sedangkan sentimen netral memiliki tingkat akurasi hingga 100%. Hasil ini membuktikan bahwa algoritma Naive Bayes dapat menghasilkan klasifikasi yang sangat akurat dalam konteks sentimen pasar otomotif.

Penelitian oleh (Indriastuti et al., 2018) mengkaji perbandingan metode K-Nearest Neighbor (K-NN) dan Naive Bayes dalam pengklasifikasian sentimen positif dan negatif pada ulasan film berbahasa Inggris. Hasilnya menunjukkan bahwa metode Naive Bayes memiliki akurasi lebih tinggi (81%) dibandingkan K-NN (71%). Namun, penelitian ini juga mencatat bahwa pengolahan kata tidak baku belum dilakukan, yang membatasi akurasi Naive Bayes.

Selain itu, penelitian oleh (Gustian, 2019) berjudul Sistem Pakar Dengan Algoritma Naive Bayes Untuk Prediksi Hasil Produksi Ayam Broiler Plasma menunjukkan keberhasilan algoritma Naive Bayes dalam memprediksi keuntungan dan kerugian pada produksi ayam broiler. Dengan menggunakan perangkat lunak Weka, rata-rata akurasi algoritma mencapai 96,36%. Selain itu, nilai Receiver Operating Curve (ROC) sebesar 0,9995 mendekati angka sempurna, menandakan kualitas prediksi yang sangat baik.

Melalui penelitian ini, diharapkan dapat dihasilkan temuan-temuan yang relevan dalam rangka meningkatkan kualitas layanan Shopee berdasarkan masukan dari pengguna. Selain itu, penelitian ini juga memberikan kontribusi terhadap pengembangan studi analisis sentimen di Indonesia, khususnya dalam konteks penggunaan perangkat lunak seperti RapidMiner dan algoritma Naive Bayes untuk mengolah data ulasan pengguna secara lebih terstruktur dan efektif.

METODE

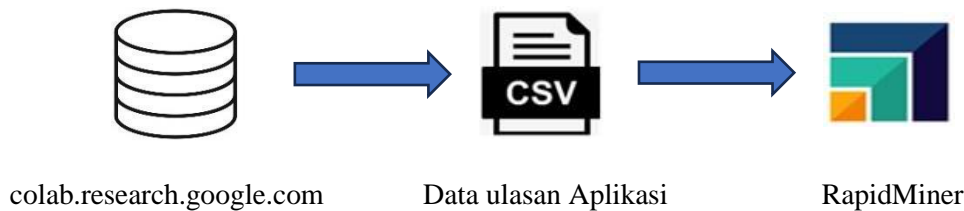
Penelitian ini menggunakan pendekatan analisis sentimen dengan algoritma Naive Bayes pada ulasan pengguna Shopee Indonesia yang diambil dari Google Playstore melalui scraping. Data ulasan diproses menggunakan RapidMiner, mencakup pelabelan sentimen (positif dan negatif), preprocessing data (cleansing, tokenisasi, dan penghapusan stopword), hingga evaluasi performa algoritma menggunakan confusion matrix untuk menghitung accuracy, precision, dan recall.

Metode ini bertujuan mengubah data tidak terstruktur menjadi informasi terstruktur yang bermanfaat untuk memahami sentimen pengguna.

HASIL DAN PEMBAHASAN

Pengumpulan Data

Pada penelitian ini, data ulasan aplikasi Shopee Indonesia dikumpulkan dengan menggunakan teknik web scraping. Data yang diambil meliputi ulasan pengguna dari Desember 2020 hingga Agustus 2023, dengan total sebanyak 895 ulasan. Ulasan tersebut diekstraksi dari halaman Google Play Store dan disimpan dalam format CSV. Data ini akan dijadikan dataset untuk analisis sentimen. Berikut adalah Gambar 1, yang menggambarkan proses pengumpulan data.



Gambar 2. Ilustrasi pada proses pengumpulan data

Pelabelan Dataset

Setelah data terkumpul, proses berikutnya adalah pelabelan dataset. Setiap ulasan dikategorikan sebagai sentimen positif atau negatif. Proses pelabelan dilakukan secara manual menggunakan Microsoft Excel, di mana sebanyak 300 ulasan digunakan sebagai data latih, sementara sisanya sebagai data uji. Berikut adalah Gambar 2, yang menunjukkan hasil pelabelan data ulasan aplikasi Shopee Indonesia.

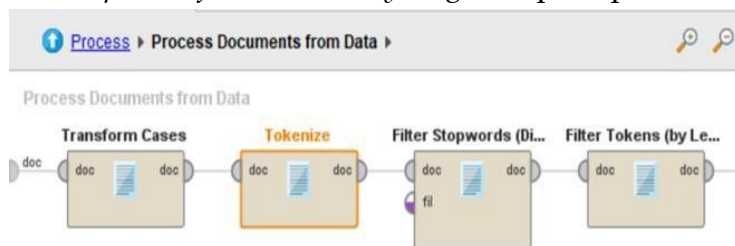
	userName	score	at	content
0	Aida EnFebriant	4	2023-09-13 13:47:37	Sangat baik. Sebagai saran. Untuk list produk ...
1	Aizzan RaQila	2	2023-09-04 15:04:57	Aplikasi lemot, terlalu banyak iklan, sekarang...
2	Arif Fathul Andalus	1	2023-08-29 03:13:23	Aplikasi pengganggu, neipon berkali-kali menaw...
3	Francisca	1	2023-09-12 06:20:07	Fitur beri tip untuk kurir shopee express regu...
4	Feby Wulandari	1	2023-09-09 08:16:30	Shopee makin kesini makin bad pelayanannya, ap...

Gambar 3. Ilustrasi pada proses pengumpulan data

Preprocessing Data

Proses selanjutnya peneliti melakukan preprocessing pada dataset ulasan aplikasi Shopee dengan menggunakan aplikasi *RapidMiner*. Pada proses

preprocessing, data yang berisi teks ulasan aplikasi Shopee yang tidak beraturan akan diubah menjadi teks yang beraturan sehingga tidak terjadi kesalahan. Pada tahap proses preprocessing ini ada tahap beberapa tahap yaitu *cleaning*, *tokenize*, *transform cases*, *stopword*, *filter tokens (by length)*. Selain itu, pada penelitian ini peneliti menggunakan operator di aplikasi RapidMiner yaitu operator *process document* untuk melakukan penyesuaian terhadap operator yang akan digunakan pada tahapan *tokenize*, *transform cases*, *stopword*, *filter tokens (by length)* seperti pada Gambar 4 berikut.



Gambar 4. Tahap Preprocessing

Cleansing merupakan proses yang dilakukan untuk melakukan proses pembersihan dengan pada teks ulasan dengan menghapus berbagai simbol atau karakter seperti : ,!/ dan seterusnya . Berikut contoh ulasan sebelum dan sesudah melewati proses *cleansing* pada Tabel 1.

Tabel 1. Proses *cleansing*

Sebelum <i>cleansing</i>	Sesudah <i>cleansing</i>
Barangnya bagus banget, sesuai deskripsi! Terima kasih 😊	Barangnya bagus banget sesuai deskripsi terima kasih
Saya pesan warna biru, tapi yang datang warna merah. Kenapa ya??? 😡	Saya pesan warna biru tapi yang datang warna merah kenapa ya

Transform cases merupakan proses untuk merubah setiap huruf besar dataset untuk diubah menjadi huruf kecil . Hal tersebut dilakukan untuk mencegah kesalahan pada saat *tokenize*. Berikut merupakan contoh dari beberapa data sebelum dan sesudah melalui proses *transform cases* pada Tabel 2.

Tabel 2. Proses *transform cases*

Sebelum <i>transform cases</i>	Sesudah <i>transform cases</i>
Barangnya bagus banget sesuai deskripsi terima kasih	Barangnya bagus banget sesuai deskripsi terima kasih
Saya pesan warna biru tapi yang datang warna merah kenapa ya	Pesan warna biru kenapa yang datang merah

Tokenizing merupakan tahapan di mana teks dipisah menjadi bagian-bagian yang dinamakan sebagai token untuk dilakukannya analisis nantinya [29]. Berikut merupakan contoh ulasan sebelum dan sesudah melewati proses tokenize pada Tabel 3.

Tabel 3. Proses *tokenizing*

Sebelum <i>tokenizing</i>	Sesudah <i>tokenizing</i>
Barangnya bagus banget sesuai deskripsi	['barangnya', 'bagus', 'banget', 'sesuai', 'deskripsi']

Filter stopwords adalah tahap menghilangkan kata yang tidak memiliki arti [30]. Peneliti menggunakan *stopwords* Bahasa Indonesia yang didapat dari www.kaggle.com. Berikut merupakan contoh ulasan yang telah melalui proses filter *stopwords* pada Tabel 4.

Tabel 4. Proses *stopword*

Sebelum <i>stopword</i>	Sesudah <i>stopword</i>
Barangnya bagus banget sesuai deskripsi dan pengiriman cepat	barangnya bagus deskripsi pengiriman cepat
Saya pesan warna biru, tetapi yang datang warna merah. Kenapa ya?	pesan warna biru datang warna merah kenapa

Filter tokens by length merupakan tahapan untuk menghapus kata yang telah ditetapkan jumlah hurufnya oleh peneliti. Berikut merupakan contoh ulasan yang telah melalui proses *filter tokens by length* pada Tabel 5.

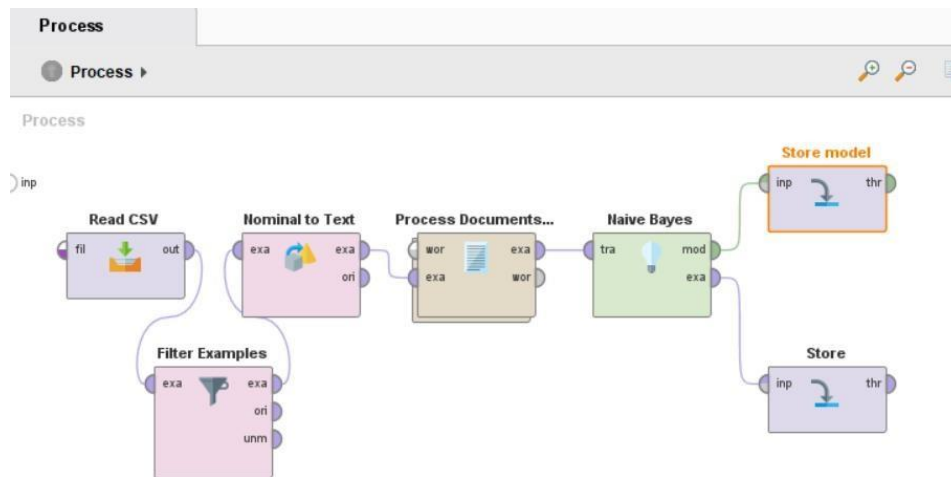
Tabel 5. Proses *filter by length*

Sebelum <i>filter tokens by length</i>	Sesudah <i>filter tokens by length</i>
Barangnya bagus banget sesuai deskripsi	barangnya bagus banget sesuai deskripsi
Pesan warna biru datang warna merah kenapa ya	pesan warna biru datang warna merah kenapa

Implementasi Naïve Bayes

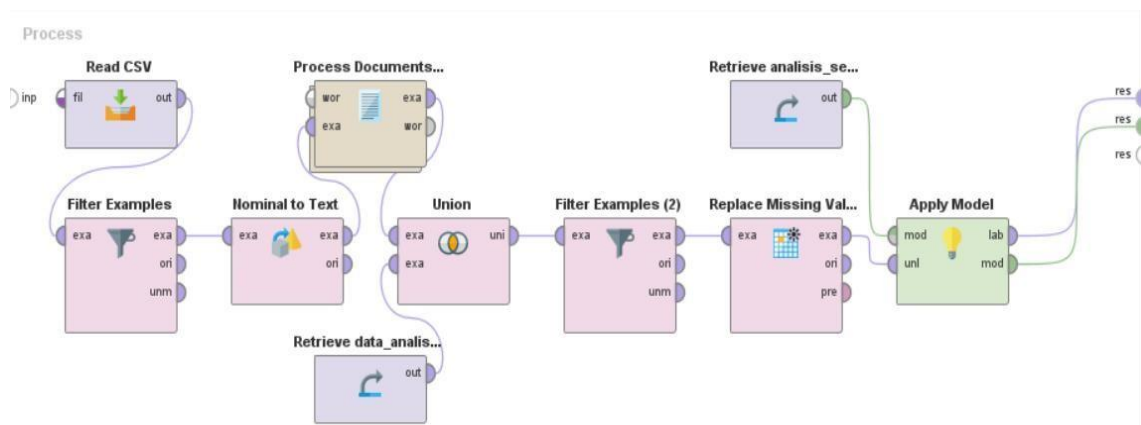
Setelah tahap *preprocessing*, langkah selanjutnya adalah pengimplementasian algoritma Naïve Bayes. Proses ini dilakukan menggunakan aplikasi RapidMiner. Peneliti terlebih dahulu membuat model algoritma Naïve Bayes beserta data latih dengan menghubungkan operator *Process Document* dari tahap *preprocessing* ke operator Naïve Bayes. Data yang digunakan merupakan ulasan aplikasi Shopee yang sudah diberi label sentimen.

Pembagian data dilakukan dengan rasio 60:40, di mana 60% untuk data latih dan 40% untuk data uji. Gambar 5 memperlihatkan alur pemrosesan model Naive Bayes dan pembuatan data latih



Gambar 5. Proses pembuatan model Naive Bayes dan data latih.

Setelah model Naive Bayes dan data latih tersedia, langkah selanjutnya adalah proses klasifikasi. Data uji yang berisi ulasan tanpa label sentimen difilter untuk memilih data dengan sentimen kosong, lalu diproses melalui *preprocessing* menggunakan operator *Process Document*. Hasil *preprocessing* digabungkan dengan data latih menggunakan operator *Union* untuk menyelaraskan atribut antara data latih dan data uji. Setelah itu, data uji yang telah siap diklasifikasikan menggunakan model Naive Bayes dengan operator *Apply Model*. Proses ini menghasilkan data yang secara otomatis dilabeli sentimennya berdasarkan pola yang telah dipelajari oleh model Naive Bayes. Tahapan pemberian label sentimen ini ditunjukkan pada Gambar 6.



Gambar 6. Proses pelabelan secara otomatis dengan Naive Bayes di aplikasi

RapidMiner

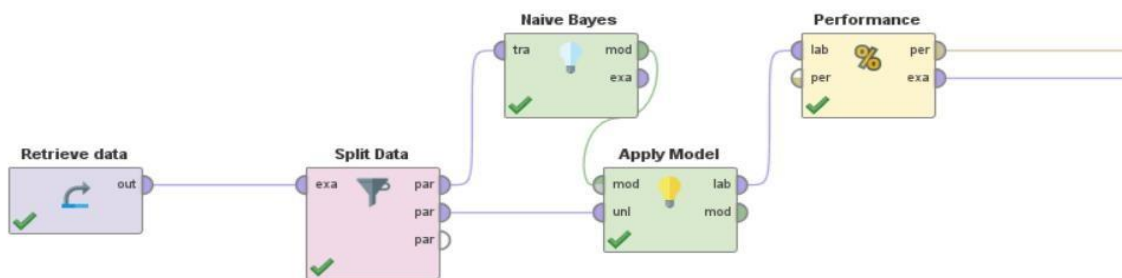
Setelah proses pemberian label sentimen secara otomatis menggunakan metode algoritma Naïve Bayes pada ulasan pengguna aplikasi Shopee Indonesia, hasil prediksi sentimen akan muncul. Algoritma Naïve Bayes telah dilatih sebelumnya menggunakan data latih yang sudah diberi label sentimen secara manual. Dari hasil pelatihan tersebut, algoritma dapat mengenali pola dalam data latih sehingga mampu memprediksi sentimen pada data uji menjadi dua kategori, yaitu positif dan negatif. Hasil akhir dari proses ini ditampilkan pada **Gambar 7**.

2	Promonya bagus. Tapi tolong jasa kirimnya kebi...	5	Positif
3	Kalau bisa toko online harus bisa bersaing seh...	1	Negatif
4	Akun belanja kesayangan, karena banyak promo d...	5	Positif
5	shopee memang gk jelas,padahal cuman liat aj,t...	1	Negatif
6	Maaf aku kasih bintang 1 soalnya makin lama ko...	1	Negatif
7	Ada beberapa toko yg lama mengantar barang pad...	1	Negatif
8	Sekarang cek barang dan toko lebih mudah. Pili...	5	Positif
9	tadinya males banget belanja via shopee, selal...	5	Positif
10	Aplikasi Shopee mmbuat belanjaku semakin mudah...	5	Positif
11	Makin kesini makin kesana lemot nya gak ketulu...	2	Negatif

Gambar 7. Hasil dari proses pemberian label secara otomatis dengan Naïve Bayes

Pengujian Evaluasi

Tahapan selanjutnya setelah klasifikasi dengan algoritma Naïve Bayes yaitu melakukan proses pengujian/evaluasi dari algoritma Naïve Bayes. Pada pengujian evaluasi ini *confusion matrix* digunakan untuk melakukan pengujian evaluasi. Berikut Gambar 8 merupakan alur dilakukannya pengujian algoritma Naïve Bayes.



Gambar 8. alur pengujian algoritma Naïve Bayes dengan confusion matrix

Pada operator split data peneliti akan menggunakan split data dengan perbandingan yaitu sebesar 60:40. Hasil dari proses pengujian algoritma Naïve bayes ini berupa nilai *Precision*, *recall*, dan *accuracy*. Berikut Gambar 9 merupakan *confusion matrix* hasil dari proses pengujian algoritma Naïve Bayes.

Confusion Matrix

	true NEGATIF	true POSITIF	class precision
pred. NEGATIF	114	101	53.02%
pred. POSITIF	7	34	82.93%
class recall	94.21%	25.19%	

Gambar 9. Confusion matrix dari hasil pengujian algoritma Naive Bayes

Berdasarkan hasil *confusion matrix* dari algoritma yang digunakan, model mendapatkan nilai *accuracy* sebesar **74.67%**. Sentimen **positif** pada proses pengujian memperoleh *recall* sebesar **25.19%** dan *precision* sebesar **82.93%**. Sementara itu, sentimen **negatif** memperoleh *recall* sebesar **94.21%** dan *precision* sebesar **53.02%**. Hal ini menunjukkan bahwa model lebih baik dalam mendeteksi sentimen negatif dibandingkan sentimen positif, namun masih memiliki kelemahan dalam memprediksi sebagian besar data positif.

KESIMPULAN

Penelitian ini berhasil menganalisis sentimen ulasan pengguna aplikasi Shopee Indonesia menggunakan algoritma Naive Bayes. Proses analisis melibatkan tahapan preprocessing data, pelabelan sentimen, dan implementasi algoritma Naive Bayes dengan menggunakan RapidMiner. Berdasarkan hasil evaluasi menggunakan confusion matrix, algoritma ini menunjukkan akurasi sebesar 76,67%. Sentimen positif memiliki nilai recall 25,19% dan precision 82,93%, sedangkan sentimen negatif memperoleh nilai recall 94,21% dan precision 53,02%.

Temuan ini menegaskan bahwa Naive Bayes merupakan metode yang efektif dan sederhana untuk menganalisis sentimen pada data teks, khususnya dalam konteks ulasan pengguna aplikasi e-commerce. Penelitian ini juga memberikan kontribusi signifikan dalam membantu pengembang aplikasi memahami pola sentimen pengguna, sehingga dapat digunakan untuk meningkatkan kualitas layanan di masa mendatang. Penelitian lebih lanjut dapat dilakukan dengan memanfaatkan data yang lebih besar atau mengombinasikan algoritma ini dengan metode pembelajaran mesin lainnya untuk meningkatkan akurasi.

REFERENSI

Gustian, D. (2019). Sistem Pakar SISTEM PAKAR DENGAN ALGORITMA NAIVE BAYES UNTUK PREDIKSI HASIL PRODUKSI AYAM BROILER PLASMA (STUDI KASUS : PT.SEKAWAN SINAR SURYA). *Jurnal Riset Sistem Informasi*

dan Teknologi Informasi (JURSISTEKNI), 1(3).

<https://doi.org/10.52005/jursistekni.v1i3.12>

Indriastuti, M., Hani'ah, & Sukmono, A. (2018). Analisis Kepadatan Bangunan Menggunakan Interpretasi Hibrida Citra Satelit Landsat Di Kecamatan Ungaran Timur Dan Ungaran Barat Kabupaten Semarang Tahun 2009-2018. *Jurnal Geodesi Undip, 5(4).*

Rustiana, D., & Rahayu, N. (2017). ANALISIS SENTIMEN PASAR OTOMOTIF MOBIL: TWEET TWITTER MENGGUNAKAN NAÏVE BAYES. *Simetris: Jurnal Teknik Mesin, Elektro dan Ilmu Komputer, 8(1).*

<https://doi.org/10.24176/simet.v8i1.841>

Samsir, S., Ambiyar, A., Verawardina, U., Edi, F., & Watrianthos, R. (2021). Analisis Sentimen Pembelajaran Daring Pada Twitter di Masa Pandemi COVID-19 Menggunakan Metode Naïve Bayes. *JURNAL MEDIA INFORMATIKA BUDIDARMA, 5(1).* <https://doi.org/10.30865/mib.v5i1.2580>