

## **Analisis Sentimen Ulasan Pengguna Aplikasi Alibaba.Com Pada Google Playstore Menggunakan Naïve Bayes**

**Rafi Fadhlur Rahman, Faldy Irwiensyah\***

Fakultas Teknologi Industri dan Informatika, Informatika, Universitas Muhammadiyah Prof. Dr. Hamka, Jakarta, Indonesia

Email: <sup>1</sup>fadhlurrahman.2002@gmail.com, <sup>2</sup>\*faldy@uhamka.ac.id

Email Penulis Korespondensi: faldy@uhamka.ac.id

Submitted: 25/10/2024; Accepted: 15/11/2024; Published: 15/11/2024

**Abstrak**—Abstrak *Alibaba.com*, sebagai salah satu platform terkemuka, terus berupaya meningkatkan layanannya berdasarkan umpan balik dari pengguna. Salah satu pendekatan yang digunakan adalah melalui pengumpulan ulasan pengguna di *Google Play Store*. Untuk meningkatkan kualitas layanan dan pengalaman pengguna, analisis sentimen terhadap ulasan tersebut menjadi sangat penting. Pada penelitian ini, algoritma *Naïve Bayes* diterapkan untuk menganalisis sentimen ulasan dengan tujuan menentukan apakah sentimen tersebut positif atau negatif. Data berupa ulasan diambil melalui *web scraping*, Menghasilkan 998 ulasan yang diproses melalui tahap *preprocessing*. Selanjutnya, dataset dibagi menjadi data uji dan data latih melalui perbandingan 60:40, di mana 599 ulasan dilabeli secara manual untuk pelatihan, dan 399 ulasan digunakan sebagai data uji. Algoritma *Naïve Bayes* kemudian mengelompokkan ulasan menjadi sentimen positif atau negatif. Berdasarkan evaluasi menggunakan *confusion matrix*, model ini menunjukkan akurasi sebesar 77,44%, *precision* sebesar 83,39%, dan *recall* sebesar 85,16%. Sebanyak 721 ulasan diklasifikasikan sebagai sentimen positif, sementara 277 ulasan dikategorikan sebagai sentimen negatif. Permasalahan utama yang diidentifikasi dalam ulasan negatif ini mencakup kendala terkait bahasa dan pembayaran. Selain itu, terdapat juga keluhan mengenai penipuan jual beli barang online, yang merupakan isu signifikan di platform ini. Banyak pengguna melaporkan pengalaman negatif yang terkait dengan transaksi yang tidak sesuai, barang yang tidak diterima, atau produk yang tidak sesuai dengan deskripsi. Hal ini menunjukkan pentingnya sistem *verifikasi* dan keamanan yang lebih baik untuk melindungi pengguna dari penipuan. Pada penelitian ini membuktikan kalau Algoritma *Naïve Bayes* cukup efektif dalam menilai persepsi pengguna dari ulasan aplikasi *Alibaba.com*.

**Kata Kunci:** Analisis; Aplikasi Alibaba.com; Google Playstore; Naïve Bayes, RapidMiner

**Abstract**—Abstract *Alibaba.com*, as one of the leading platforms, continues to strive to improve its services based on user feedback. One approach used is the collection of user reviews on the *Google Play Store*. To enhance service quality and user experience, sentiment analysis of these reviews becomes crucial. In this study, the *Naïve Bayes* algorithm is applied to analyze the sentiment of the reviews with the aim of determining whether the sentiment is positive or negative. The data, consisting of reviews, was obtained through *web scraping*, resulting in 998 reviews that were processed through *preprocessing* stages. The dataset was then divided into training and testing data with a 60:40 ratio, where 599 reviews were manually labeled for training, and 399 reviews were used as test data. The *Naïve Bayes* algorithm subsequently categorized the reviews as either positive or negative sentiment. An evaluation with a *confusion matrix* was then used to assess performance, this model showed an accuracy of 77.44%, precision of 83.39%, and recall of 85.16%. A total of 721 reviews were categorized as positive sentiment, while 277 reviews were categorized as negative sentiment. The main issues identified in the negative reviews included challenges related to language and payment. Additionally, there were complaints regarding online buying and selling fraud, which is a significant issue on this platform. Many users reported negative experiences related to transactions that did not match expectations, items that were not received, or products that did not match their descriptions. This highlights the importance of better verification and security systems to protect users from fraud. This study demonstrates that the *Naïve Bayes* algorithm is quite efficient in analyzing user review sentiments on the *Alibaba.com* application.

**Keywords:** Analysis; Alibaba.com Application; Google Playstore; Naïve Bayes; RapidMiner

## **1. PENDAHULUAN**

Pada zaman yang semakin maju memasuki era digital, penggunaan teknologi digital telah membawa dampak signifikan dalam berbagai unsur kehidupan, termasuk di bidang perdagangan telah mendorong pengembangan model perdagangan elektronik (*e-commerce*) yang diharapkan dapat meningkatkan permintaan konsumen serta menekan harga barang dan jasa [1]. Kemajuan pesat dalam teknologi dan infrastruktur internet telah mengubah perdagangan tradisional menjadi perdagangan elektronik, atau yang lebih Dikenal sebagai *e-commerce*, ini merupakan aktivitas transaksi penjualan yang dilakukan secara online menggunakan sistem komputer dan perangkat digital lainnya, dengan berbagai model dan jenis *e-commerce* yang beragam [2].

Salah satu alat utama yang digunakan dalam era digital ini adalah media internet, yang memungkinkan bisnis dan restoran untuk memanfaatkan *e-commerce* sebagai sarana penjualan dan pembelian secara online [3]. *E-commerce* memberikan kemudahan bagi penjual untuk memasarkan barang mereka ke publik tanpa perlu menyewa toko fisik [4]. Selain itu, masyarakat Indonesia sudah semakin terbiasa dengan penggunaan *e-commerce* dan bisnis elektronik. *E-commerce*, yang juga dikenal sebagai perdagangan elektronik, Meliputi aktivitas jual beli atau pertukaran barang, jasa, dan informasi melalui jaringan informasi, termasuk internet [5].

*Alibaba.com* adalah salah satu situs *e-commerce* terbesar di Indonesia, yang mulai berdiri pada tahun 1998, Platform ini memungkinkan pengguna untuk mengunggah, menjual, dan membeli barang secara online. Pada tahun 1999, Jack Ma, pendiri *Alibaba*, mempekerjakan 18 karyawan untuk mengembangkan bisnis yang

memanfaatkan internet sebagai alat untuk membantu pengusaha kecil bersaing dengan perusahaan besar di pasar domestik maupun internasional [6].

Banyak platform *e-commerce*, termasuk Alibaba.com, menghadapi masalah keamanan transaksi dan kepercayaan pelanggan, terutama dalam hal penipuan jual beli online. Beberapa pengguna di Indonesia mengklaim telah menjadi korban penipuan saat mereka membeli barang di *Alibaba.com*. Modus penipuan ini termasuk penjual yang mengirimkan produk yang berbeda dari yang tercantum dalam deskripsi, produk palsu, atau bahkan tidak mengirimkan produk sama sekali setelah pembayaran diterima. Melacak dan mengajukan klaim pengembalian dana juga sulit karena penjual menutup akun setelah transaksi penipuan. Kasus-kasus seperti ini menantang kepercayaan pengguna terhadap *e-commerce* internasional dan menunjukkan bahwa regulasi dan pengawasan yang lebih ketat diperlukan untuk transaksi yang dilakukan di platform global.

Selain *e-commerce*, *Google Play Store* juga menjadi platform penting dalam era digital, menyediakan berbagai aplikasi yang berkaitan dengan game, musik, film, dan buku [7]. *Play Store* memiliki fitur rating dan ulasan, yang memungkinkan pengguna menyampaikan pendapat mereka tentang aplikasi yang telah digunakan [8]. Ulasan ini dapat berupa tanggapan positif maupun negatif, sesuai dengan pengalaman pengguna.

Analisis sentimen adalah metode yang populer digunakan untuk menempatkan ulasan ke dalam kategori positif, negatif, atau netral. Metode ini sering diterapkan dalam menganalisis ulasan pengguna terkait aplikasi *e-commerce* di *Google Play Store* [9]. Salah satu algoritma yang sering digunakan dalam analisis sentimen adalah Naive Bayes, yang terkenal karena kesederhanaan dan kecepatan dalam pemrosesannya [10]. Algoritma ini telah terbukti efektif dalam berbagai penelitian, dengan tingkat akurasi yang sangat tinggi. Sebagai contoh, dalam analisis sentimen ulasan aplikasi Primaku, algoritma *SVM* dengan kernel linear menunjukkan akurasi terbaik hingga 97,5% [11].

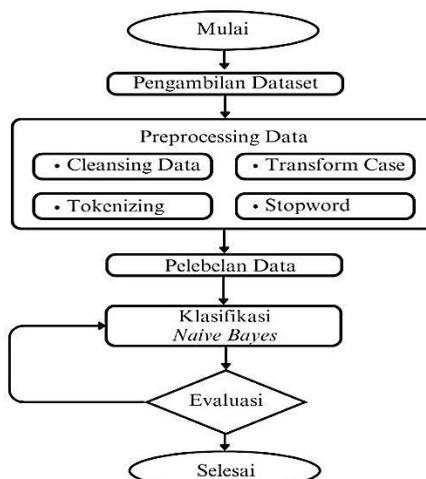
Penelitian terdahulu yang mana menerapkan algoritma *Naive Bayes* untuk menganalisis sentimen ulasan di platform Shopee juga menghasilkan hasil yang mengesankan, dengan akurasi mencapai 99,5%. Studi tambahan memanfaatkan kombinasi Metode *Naive Bayes* dan *K-Nearest Neighbor* untuk menganalisis sentimen komentar di Tokopedia, juga menghasilkan data yang signifikan, dengan tingkat akurasi sebesar 75,30% [12]. Selanjutnya penelitian yang dilakukan oleh Dendy Aprilianto Nugroho dengan judul “Analisis Sentimen Kegiatan Pembersihan Sampah Pada Media Sosial X Menggunakan *SVM* dan *Naïve Bayes*” mendapatkan akurasi sebesar 91,67% [13]. Selanjutnya penelitian yang dilakukan oleh Khairul Fadli dengan judul “Analisis Sentimen Tanggapan Masyarakat Terhadap Penutupan TikTok Shop Menggunakan Metode *Naïve Bayes*” mendapatkan tingkat akurasi sebesar 90,24% [14]. Dari penelitian sebelumnya dapat dibuktikan bahwa metode *Naive Bayes* dapat memberikan akurasi yang tinggi pada evaluasi sentimen.

Berdasarkan latar belakang tersebut, penulis menggunakan analisis sentimen pada aplikasi *Alibaba.com* dengan memanfaatkan Data ulasan diambil dari *Google Play Store*. Studi ini mengeksplorasi sentimen positif, negatif, dan netral dalam ulasan pengguna dengan menggunakan algoritma klasifikasi *Naive Bayes*.

## 2. METODOLOGI PENELITIAN

### 2.1 Tahapan Penelitian

Dalam penelitian ini fokus pada penerapan metode *Naive Bayes* pada proses klasifikasi analisis sentimen dari data ulasan yang telah dikumpulkan. Dataset ulasan dikumpulkan menggunakan layanan *Google Colab* untuk pengolahan data. Lalu proses klasifikasi dilakukan dengan bantuan dari *RapidMiner*. Gambar 1 menunjukkan proses tahapan metode penelitian.



**Gambar 1.** Tahapan Penelitian

Penelitian ini menerapkan berbagai tahap pada analisis sentimen terhadap komentar pengguna jasa aplikasi Alibaba.com yang ada Google Play Store. Dimulai Setelah data diambil melalui *scrapping*, proses *preprocessing* termasuk pembersihan (*Cleansing Data*), tokenisasi (*Tokenizing*), transformasi case (*Transform Case*), dan penghapusan *stopword*. Data yang telah diproses kemudian diberi label untuk kategori sentimen positif dan negatif, dan algoritma *Naive Bayes* digunakan untuk mengklasifikasikan data. Proses evaluasi dilakukan untuk mengukur kinerja model dalam hal akurasi, presisi, dan recall. Hasilnya menunjukkan betapa efektifnya algoritma dalam mengklasifikasikan sentimen pada kumpulan data ini. Penelitian ini diharapkan untuk memberikan lebih banyak informasi tentang cara bisnis *e-commerce* dapat menggunakan analisis sentimen.

## 2.2 Pengumpulan Data

Pada tahapan ini, penulis menggunakan teknik *crawling* dataset pada aplikasi *Google Play Store*. Tujuan dari teknik pengambilan dataset ini adalah untuk mendapatkan informasi tentang ulasan dari platform *Play Store* dan mengekstrak dataset menjadi bentuk dokumen. [15]. Langkah pertama adalah menggunakan *aplikasi Google Collab* untuk mengambil data ulasan dari aplikasi *Alibaba.com di google play store*. Setelah data diambil dengan sukses, file tersebut diubah ke format *file.csv*.

## 2.3 Preprocessing Data

Tahap pengolahan data merupakan langkah penting dalam analisis data, karena memastikan data mentah diformat dengan tepat untuk keperluan analisis lanjutan. [16]. Dengan melakukan pengolahan data yang efektif, pengguna dapat memastikan bahwa data yang dianalisis berkualitas tinggi dan siap memberikan wawasan yang akurat dan bernilai. Proses preprocessing untuk memproses dataset penelitian ini dijelaskan di bawah ini:

- a. *Cleansing* adalah proses untuk menghilangkan tanda baca, angka, simbol, *URL*, dan nama pengguna dari teks. Tujuannya adalah menghasilkan data cuitan yang bersih dan bebas dari elemen-elemen yang tidak relevan, sehingga analisis dapat dilakukan dengan lebih mudah dan keaslian informasi dalam setiap cuitan tetap terjaga [17].
- b. *Tokenizing* adalah proses memisahkan setiap kata dalam dokumen sambil mengubah huruf besar menjadi huruf kecil. Proses ini hanya berfokus pada pengolahan huruf, sehingga karakter atau tanda baca lainnya dihapus. Selama tahap ini, hanya huruf dari 'a' hingga 'z' yang dipertahankan. *Tokenizing* memungkinkan teks diurai menjadi unit terkecil, sehingga lebih mudah diolah dalam analisis teks atau pemrosesan bahasa alami[18].
- c. *Transform Cases* adalah proses standarisasi atau normalisasi teks dengan mengonversi semua huruf menjadi huruf kecil atau huruf besar. Sebagai contoh, kata "BARANG" dan "barang" akan dianggap sama setelah proses ini [19]. Langkah ini membantu meningkatkan akurasi dalam klasifikasi sentimen dengan mengurangi variasi pada data yang mungkin dianggap berbeda oleh model.
- d. *Filter Stopwords* adalah tahap di mana kata-kata penting diambil dari hasil *tokenizing*. Pada tahap ini, digunakan algoritma *stoplist*, yaitu daftar kata untuk menghilangkan kata-kata yang dianggap kurang relevan atau tidak penting [20]. Selain itu, ada daftar khusus yang berfungsi menyimpan kata-kata yang dinilai penting atau relevan untuk analisis atau pemrosesan lebih lanjut.

## 2.4 Pelabelan Data

Setelah tahap preprocessing selesai, pelabelan data dilakukan. Tujuan pelabelan dataset adalah untuk membantu memahami cara mengidentifikasi sentimen data dengan menggunakan aplikasi *RapidMiner* baik secara manual maupun otomatis dan membuat representasi objek data yang terdapat dalam ulasan. [21]. Pelabelan data ini harus disesuaikan dengan perbandingan antara data uji dan data latih.

## 2.5 Klasifikasi Naïve Bayes

*Naive Bayes* adalah algoritma *klasifikasi* berbasis probabilitas yang sangat efektif, terutama untuk data yang berukuran besar. Di *RapidMiner*, berbagai operator mendukung setiap tahap proses klasifikasi, membuat proses membangun dan menerapkan model *Naive Bayes* sederhana dan mudah dipahami. Dengan *RapidMiner* menggunakan *Naive Bayes* [22], Model yang akurat dan prediksi yang dibuat berdasarkan data baru membantu pengguna membuat keputusan yang lebih baik.

## 2.6 Evaluasi

Siklus analisis data mencakup langkah penting yang dikenal sebagai evaluasi data, yang bertujuan untuk menilai kinerja model berdasarkan metrik tertentu. Ini dilakukan untuk memastikan bahwa model yang dibangun sangat akurat serta dapat menghasilkan prediksi yang akurat ketika diaplikasikan pada data baru. Asli *Positif* (*TP*), Asli *Negatif* (*TN*), *False Positif* (*FP*), dan *False Negatif* (*FN*) adalah empat komponen *confusion matrix* yang digunakan untuk melakukan evaluasi untuk memastikan hasil yang didapat [23]. *Precision*, *Recall*, dan *Accuracy* adalah tiga rumus yang digunakan dalam penelitian ini, masing-masing dengan rumus berikut.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

$$Precision = \frac{TP}{TP+FP} \quad (2)$$

$$Recall = \frac{TP}{TP+FN} \quad (3)$$

### 3. HASIL DAN PEMBAHASAN

#### 3.1 Pengambilan Dataset (Scrapping)

Pada awal analisis ini dimulai dari pengumpulan dataset menggunakan metode *scrapping* dataset pada 22 Januari 2024. Dataset yang diambil dari aplikasi *Alibaba.com* pada *playstore* berupa ulasan yang di *scrapping* menggunakan *Google Collab* kemudian disimpan dalam format *CSV*, seperti yang terlihat pada Gambar 2

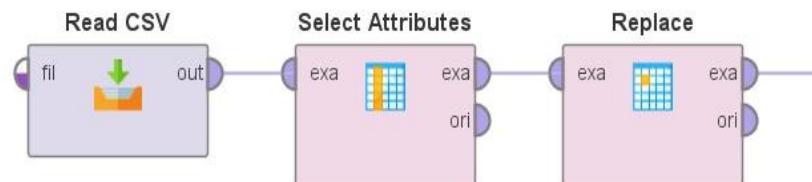


**Gambar 2.** Proses Pengambilan Data

Pengumpulan data ulasan aplikasi *Alibaba.com* dijalankan dengan teknik *scrapping* di Google Colab menggunakan tautan aplikasi 'com.alibaba.intl'. Proses ini melibatkan dua library, yaitu NumPy dan Pandas. Dengan jumlah 998 ulasan pengguna berhasil dikumpulkan melalui metode *scrapping* dengan kategori ulasan terbaru (*newest*). Data yang diperoleh mencakup beberapa atribut, seperti hal nya nama user (*username*), penilaian (*score*), tanggal ulasan (*at*), serta isi ulasan (*content*).

#### 3.2 Preprocessing

Tahapan *preprocessing* meliputi *cleansing*, *tokenizing*, *transform cases*, *filter stopwords*. Alat yang digunakan untuk proses *preprocessing* adalah *RapidMiner*. Gambar 3 menggambarkan proses *cleansing*, di mana tahap ini dimulai dengan membaca seluruh data ulasan menggunakan operator *readCSV*. Setelah itu, operator *replace* digunakan untuk menghilangkan karakter-kharakter khusus dan tanda baca dari setiap ulasan yang muncul.



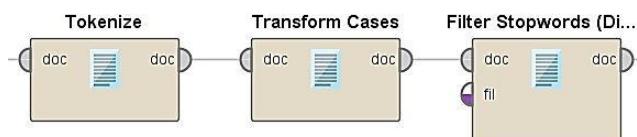
**Gambar 3.** Proses *Cleansing*

Langkah awal yang dilakukan adalah proses *Cleansing*, yang bertujuan untuk menghapus tanda baca serta karakter-karakter khusus dalam data ulasan, Seperti pada tabel 1.

**Tabel 1.** Hasil *Cleansing*

Sebelum <i>Cleansing</i>	Sesudah <i>Cleansing</i>
Baju bagus2 tp Onkirnya, mahal..	Baju bagus2 tp Onkirnya mahal
Alhamdulillah keren sih, makin nyaman pake aplikasi ini	Alhamdulillah keren sih makin nyaman pake aplikasi ini
application sangat Bagus....bahasa tidak auto Terjemahkan indonesia..	application sangat Bagus bahasa tidak auto Terjemahkan indonesia
Mo belanja orang di bikin ribet.... aneh...	Mo belanja orang di bikin ribet aneh
Produknya bagus, harga terjangkau	Produknya bagus harga terjangkau

Menunjukkan bagian operator *preprocessing*. Setelah menyelesaikan tahap awal pembersihan, proses dilanjutkan dengan beberapa langkah berikutnya, yaitu tokenisasi, transformasi huruf, dan penyaringan *stopwords*.



**Gambar 4.** Isi operator Preprocessing

Langkah kedua yaitu *tokenizing*. *Tokenizing* diperlukan supaya pemecahan teks menjadi kata-kata yang memiliki makna pada sebuah kalimat. Kata pada ulasan diubah menjadi potongan-potongan menggunakan operator *tokenizing*, Seperti pada tabel 2.

**Tabel 2.** Hasil *Tokenizing*

Sebelum <i>Tokenizing</i>	Sesudah <i>Tokenizing</i>
Baju bagus2 tp Onkirnya mahal	Baju, bagus2, tp, Onkirnya, mahal
Alhamdulillah keren sih makin nyaman pake aplikasi ini	Alhamdulillah, keren, sih, makin, nyaman pake, aplikasi, ini
application sangat Bagus bahasa tidak auto	Application, sangat, Bagus, bahasa, tidak, auto,
Terjemahkan indonesia	Terjemahkan, indonesia
Mo belanja orang di bikin ribet aneh	Mo, belanja, orang, di, bikin, ribet, aneh
Produknya bagus harga terjangkau	Produknya, bagus, harga, terjangkau

Langkah ketiga adalah *transform cases*. *Transform cases* adalah mengubah huruf kapital atau besar dalam suatu kalimat menjadi huruf kecil atau sebaliknya, Seperti pada tabel 3.

**Tabel 3.** Hasil *Transform cases*

Sebelum <i>Transform cases</i>	Sesudah <i>Transform cases</i>
Baju, bagus2, tp, Onkirnya, mahal	baju, bagus2, tp, onkirnya, mahal
Alhamdulillah, keren, sih, makin, nyaman pake, aplikasi, ini	alhamdulillah, keren, sih, makin, nyaman pake, aplikasi, ini
Application, sangat, Bagus, bahasa, tidak, auto, Terjemahkan, indonesia	application, sangat, bagus, bahasa, tidak, auto, terjemahkan, indonesia
Mo, belanja, orang, di, bikin, ribet, aneh	mo, belanja, orang, di, bikin, ribet, aneh
Produknya, bagus, harga, terjangkau	produknya, bagus, harga, terjangkau

Langkah keempat adalah pemfilteran *stopword*. Pemfilteran *stopword* adalah proses di mana kata-kata yang tidak memiliki arti yang signifikan, seperti kata sambung, kata keterangan, dan sejenisnya, dihilangkan, Seperti pada tabel 4.

**Tabel 4.** Hasil *Filter Stopword*

Sebelum <i>Transform cases</i>	Sesudah <i>Transform cases</i>
Baju, bagus2, tp, Onkirnya, mahal	baju, bagus2, onkirnya, mahal
Alhamdulillah, keren, sih, makin, nyaman pake, aplikasi, ini	keren, makin, nyaman pake, aplikasi, ini
Application, sangat, Bagus, bahasa, tidak, auto, Terjemahkan, indonesia	application, sangat, bagus, bahasa, tidak, auto, terjemahkan
Mo, belanja, orang, di, bikin, ribet, aneh	belanja, di, bikin, ribet
Produknya, bagus, harga, terjangkau	produknya, bagus, harga, terjangkau

### 3.2 Labelling

Setelah menyelesaikan tahap pembersihan dalam proses *preprocessing*, tahap selanjutnya adalah pelabelan. Pelabelan untuk menentukan sentimen positif atau negatif dilakukan dengan mengikuti acuan dari kamus sentimen (*Lexicon-Based Approach*) dengan melihat jumlah kata yang lebih banyak dipakai antara kata positif atau negatif pada suatu ulasan. Dari total 1000 data yang telah dibersihkan, jumlah data yang tersisa menjadi 998. Dataset lalu terbagi menjadi dua bagian, yaitu berupa dataset pelatihan dan dataset pengujian dengan skala 60:40. Pemberian label data dikerjakan dengan cara manual, menghasilkan 599 data yang dilabeli dan 399 data yang tidak dilabeli. Dataset yang telah dilabeli nantinya akan dimanfaatkan sebagai dataset pelatihan, sedangkan dataset yang tidak dilabeli akan berfungsi sebagai data uji untuk model penerapan *Naive Bayes*.

**Tabel 5.** Labelling Data

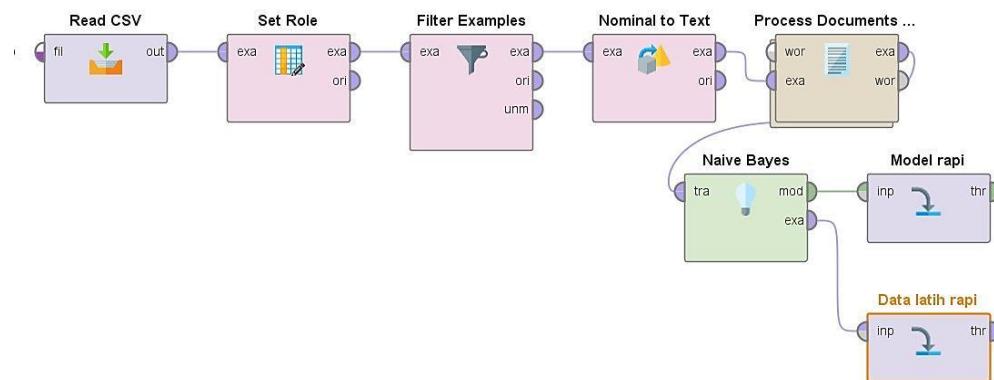
Ulasan	Sentimen
Layanan cepat dan mudah	Positif
Kemudahan mendapat info barang	Positif
Sangat Meribetkan bagi pengguna	Negatif

Ulasan	Sentimen
Bagus & bermanfaat	Positif
sangat dipersulit mengisi alamat	Negatif

Data yang sudah dilabeli berdasarkan acuan dari kamus sentimen (*Lexicon-Based Approach*), ada 3 ulasan yang termasuk kedalam ulasan positif dan 2 ulasan yang masuk kedalam ulasan negatif, Seperti yang terlihat pada tabel 5.

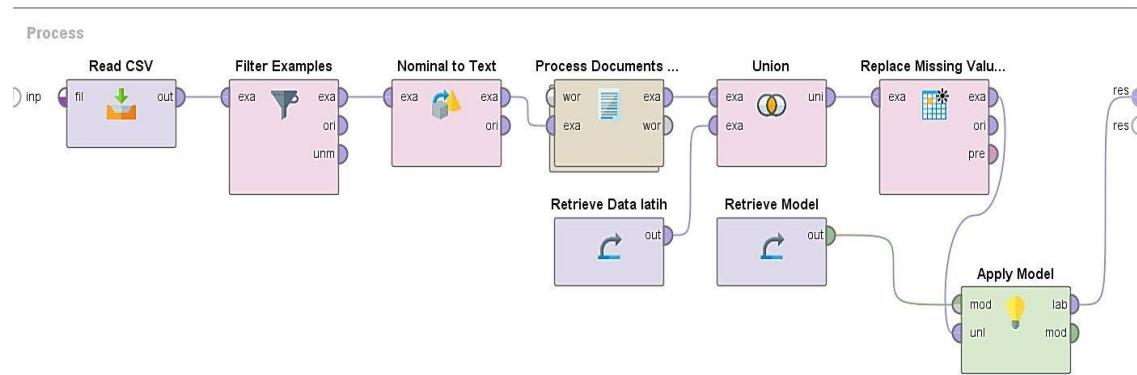
### 3.3 Implementasi Naive Bayes

Berikut nya adalah tahapan implementasi *Naive Bayes* pada gambar 5. dengan tahap awal menggunakan operator *read csv* untuk membaca file yang sudah dilabeli dan tidak dilabeli secara manual. Untuk menggunakan kolom sentimen sebagai label, operator *set role* dan operator *read csv* dihubungkan. Kemudian, operator *filter example* digunakan untuk menyaring data yang tidak memiliki nilai kosong atau yang telah dilabeli secara manual. Selanjutnya, operator *nominal to text* digunakan agar merubah data nominal berubah jadi bentuk format teks. Proses berikutnya melibatkan *process documents* yang mencakup pengisian tokenisasi, transformasi huruf, dan pemfilteran *stopword*. Selanjutnya, data dilatih menggunakan *Naive Bayes*. Dengan menghubungkan data latih dan model, kemudian hasil pelatihan disimpan.



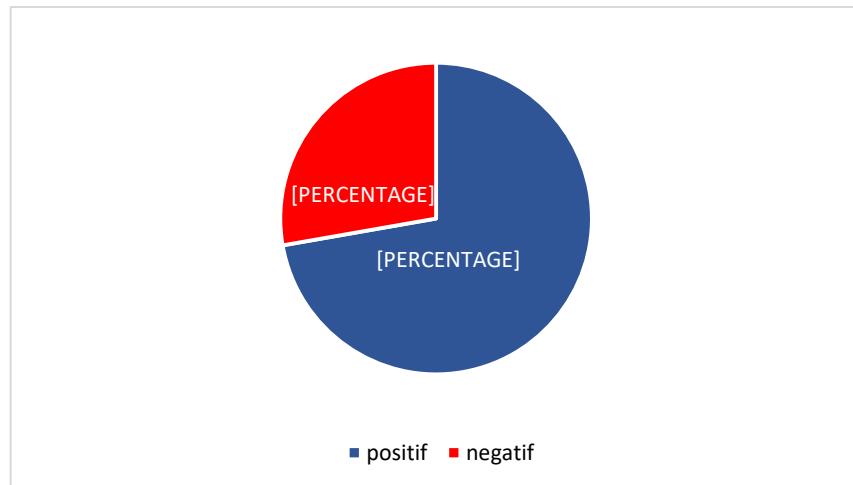
Gambar 5. Proses Pembuatan Data Latih

Setelah dataset pelatihan dibuat, penerapan *Naive Bayes* adalah langkah berikutnya, ini ditunjukkan pada gambar 6. Proses ini dimulai dengan data yang dibaca menggunakan operator *readCSV*, diikuti dengan penyaringan data yang hilang atau belum dilabeli menggunakan *filter examples*. Penjelasan mengenai operator *nominal to text* dan *process document* sama seperti pada tahap sebelumnya. Data dari hasil analisis *store data* latih digabungkan dengan hasil dari *process document* dan disambungkan dengan *filter examples*, lalu dilanjutkan dengan *replace missing values*. Sementara itu, keluaran dari *storemodel* dihubungkan dengan *applymodel* untuk menghasilkan model prediksi.



Gambar 6. Proses Pengimplementasian *Naive Bayes*

Setelah melalui tahap *preprocessing*, pelabelan, dan penerapan *Naive Bayes*, diperoleh hasil prediksi. Dari 998 data ulasan, 721 termasuk sentimen positif dan 277 termasuk sentimen negatif. Hasil ini menunjukkan bahwa sebagian besar pengguna memberikan ulasan positif lebih banyak, dengan selisih 444 ulasan lebih banyak dibandingkan dengan ulasan negatif. Penyebab utama dari sentimen negatif ini adalah ketidakpuasan pengguna terhadap kesulitan dalam segi bahasa yang digunakan dan pembayaran pada saat berbelanja pada aplikasi *alibaba.com*. Hasil ini divisualisasikan dalam bentuk diagram pie dan *wordcloud* untuk melihat kata-kata yang paling sering tampil. Seperti pada gambar 7.



**Gambar 7.** Diagram Pie

Untuk menemukan kata-kata yang paling sering muncul dari 880 ulasan pengguna, hasil penerapan *Naïve Bayes* juga divisualisasikan dalam bentuk *wordcloud*, selain menggunakan diagram pie. Gambar 8 menunjukkan kata-kata yang diurutkan dari frekuensi kemunculan tertinggi hingga terendah. Kata ‘barang’ muncul paling banyak dengan 399 kali, diikuti oleh kata ‘yang’ sebanyak 304 kali, dan ‘nya’ yang muncul 277 kali. Visualisasi ini mempermudah pemahaman pola kata yang kerap muncul dalam ulasan pengguna di aplikasi *Alibaba.com*.



**Gambar 8.** Visualisasi Wordcloud

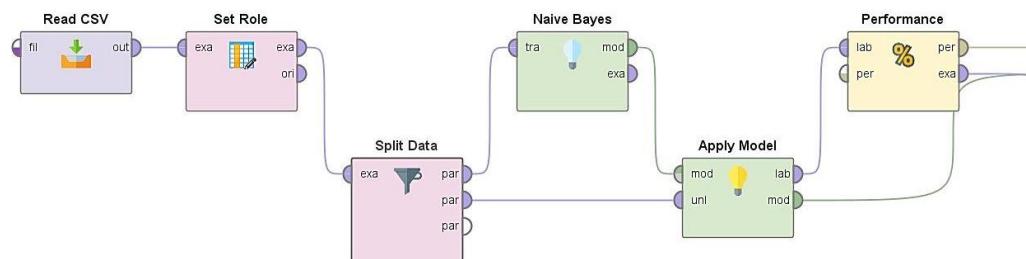
### 3.4 Evaluasi

Performa model dinilai melalui penghitungan *akurasi*, *presisi*, dan *recall*, seperti pada tabel 6. Evaluasi ini bertujuan untuk mengukur seberapa efektif model dalam memprediksi data secara akurat serta memberikan gambaran keseluruhan tentang kinerja model tersebut.

**Tabel 6. Confusion Matrix**

Evaluasi	Prediction class	
	Positif	Negatif
Actual Class	Positif	True Positif (TP)
	Negatif	False Negatif (FN)
		True Negatif (TN)

Tahapan selanjutnya pembuatan confusion matrix pada gambar 9. Dimulai dengan menginput data dari gambar ke dalam operator *read CSV*, diikuti oleh *set role* guna menetapkan *field* prediksi sebagai label. Data kemudian dibagi menggunakan *split data* (60% pelatihan, 40% pengujian) dan diterapkan melalui *apply model*. Hasil model dievaluasi dengan *performance*, yang menampilkan *akurasi*, *presisi*, dan *recall*.



**Gambar 9.** Proses Confusion Matrix

Dari hasil proses confusion matrix menunjukkan bahwa model memiliki akurasi sebesar 77,44%, dengan 241 *True Positive (TP)*, 42 *False Negative (FN)*, 48 *False Positive (FP)*, dan 68 *True Negative (TN)*. Tabel 7 menampilkan hasil *accuracy* dari proses klasifikasi tersebut.

**Tabel 7.** Hasil Dari Klasifikasi

	<i>True Positif</i>	<i>True Negatif</i>	<i>Class Precision</i>
<i>Pred. Positif</i>	241	48	83,39%
<i>Pred. Negatif</i>	42	68	61,82%
<i>Class Recall</i>	85,16%	58,62%	

Hasil dari perhitungan akurasi, presisi, dan recall yang dilakukan menggunakan persamaan confusion matrix berikut ini.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} = \frac{241+68}{241+68+48+42} = \frac{309}{399} = 0,7744 = 77,44\%$$

$$Precision = \frac{TP}{TP+FP} = \frac{241}{241+48} = \frac{241}{289} = 0,8339 = 83,39\%$$

$$Recall = \frac{TP}{TP+FN} = \frac{241}{241+42} = \frac{241}{283} = 0,8516 = 85,16\%$$

Berdasarkan hasil dari perhitungan diatas, mendapatkan evaluasi dari klasifikasi pengguna aplikasi *Alibaba.com* didapatkan *Accuracy* sebesar 77,44%, *Precision* sebesar 83,39% dan *Recall* sebesar 85,16%.

## 4. KESIMPULAN

Berdasarkan hasil penelitian menunjukkan bahwa, implementasi algoritma *Naive Bayes* yang digunakan dalam analisis sentimen komentar pada aplikasi *Alibaba.com* menunjukkan kinerja yang memuaskan. Proses ini melibatkan tahapan pengumpulan data melalui metode *scrapping*, *preprocessing data*, serta pelabelan data. Setelah data dilatih dan diuji menggunakan *Naive Bayes*, hasil evaluasi menggunakan *confusion matrix* menunjukkan bahwa model memiliki akurasi sebesar 77,44%, dengan nilai *precision* sebesar 83,39% dan *recall* sebesar 85,16%. Dengan jumlah dataset yang dilabeli secara manual sebanyak 599 dan 399 yang tidak dilabeli sebagai data uji. Mayoritas ulasan dari pengguna bersifat positif dengan jumlah sebanyak 721 ulasan, meskipun terdapat sejumlah ulasan negatif sebanyak 277 ulasan yang terkait dengan masalah bahasa, pembayaran dan kasus penipuan. Kasus penipuan yang sering dilaporkan mencakup produk yang tidak sesuai deskripsi, barang palsu, atau produk yang tidak dikirimkan setelah pembayaran dilakukan pada aplikasi *Alibaba.com*. Dapat disimpulkan dengan tingkat akurasi yang mencapai lebih dari 77,44%, model ini dapat dikatakan efektif dalam mengklasifikasikan sentimen ulasan pengguna. Untuk penelitian berikutnya, disarankan memanfaatkan fitur-fitur atau metode lain yang tersedia di *RapidMiner*, seperti *term frequency*, *term occurrences*, dan *binary term occurrences*, juga dengan menggunakan dataset yang lebih besar guna meningkatkan akurasi dan kinerja evaluasi model, serta mengevaluasi lebih lanjut kendala-kendala terkait keamanan dan kepercayaan pengguna dalam transaksi online di platform global.

## REFERENCES

- [1] H. N. Utami and S. N. Wiyono, "Manfaat Yang Diharapkan Dari Model Perdagangan Digital Produk Pangan: Perspektif Pelaku Agribisnis," *J. Agristan*, vol. 5, no. 1, pp. 61–73, 2023, doi: 10.37058/agristan.v5i1.6999.
- [2] J. Riyanto, "Analisa Sistem Aplikasi Marketplace Facebook Dalam Pengembangan Dunia Bisnis," *J. Media Inform. Budidarma*, 2020, doi: 10.30865/mib.v4i4.2346.
- [3] Nursania Dasopang, "Jiemas E – Commerce Bisnis Dan Internet," *J. Ilm. Ekon. Manaj. dan Syariah JIEMAS*, vol. 2, pp. 129–135, 2023, [Online]. Available: <https://jiemas.stai-dq.org/index.php/home>
- [4] N. Laila Wardani and J. Susyanti, "Analisis Pemanfaatan E-commerce Dalam Pengembangan Bisnis Pelaku Usaha Mikro, Kecil, dan Menengah (UMKM) di Era Revolusi 4.0 (Studi Kasus Pada UMKM di Kabupaten Malang)," *J. Ilm. Manaj. Ekon. Dan Akunt.*, vol. 1, no. 4, pp. 268–275, 2024, [Online]. Available: <https://doi.org/10.62017/jimea>
- [5] Novan, "Pemanfaatan E-commerce dalam Meningkatkan Daya Saing UMKM," *J. Pengabdi. Mandiri*, vol. 2, no. 2, pp. 639–648, 2023.
- [6] M. Iqbal and M. Ashry Sallatu, "Dampak Ekspansi Alibaba Group Terhadap Perkembangan E-Commerce Di Indonesia," *Hasanuddin J. Int. Aff.*, vol. 2, no. 1, pp. 2775–3336, 2022.
- [7] N. Agustina, D. H. Citra, W. Purnama, C. Nisa, and A. R. Kurnia, "Implementasi Algoritma Naive Bayes untuk Analisis Sentimen Ulasan Shopee pada Google Play Store," *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 2, no. 1, pp. 47–54, 2022, doi: 10.57152/malcom.v2i1.195.
- [8] N. Herlinawati, Y. Yuliani, S. Faizah, W. Gata, and S. Samudi, "Analisis Sentimen Zoom Cloud Meetings di Play Store Menggunakan Naïve Bayes dan Support Vector Machine," *CESS (Journal Comput. Eng. Syst.*

- Sci.*, vol. 5, no. 2, p. 293, 2020, doi: 10.24114/cess.v5i2.18186.
- [9] T. P. Lestari, “Analisis Text Mining pada Sosial Media Twitter Menggunakan Metode Support Vector Machine (SVM) dan Social Network Analysis (SNA),” *J. Inform. Ekon. Bisnis*, vol. 4, no. 3, pp. 65–71, 2022, doi: 10.37034/infeb.v4i3.146.
  - [10] A. H. Hasugian, M. Fakhriza, and D. Zukhoiriyah, “Analisis Sentimen Pada Review Pengguna E-Commerce Menggunakan Algoritma Naïve Bayes,” *J-SISKO TECH (Jurnal Teknol. Sist. Inf. dan Sist. Komput. TGD)*, vol. 6, no. 1, p. 98, 2023, doi: 10.53513/jsk.v6i1.7400.
  - [11] A. M. Tanniewa, “Implementasi Algoritma Support Vector Learning Terhadap Analisis Sentimen Penggunaan Aplikasi Tiktok Shop Seller Center,” *Pros. SISFOTEK*, vol. 7, no. 1, pp. 165–170, 2023.
  - [12] M. F. El Firdaus, N. Nurfaizah, and S. Sarmini, “Analisis Sentimen Tokopedia Pada Ulasan di Google Playstore Menggunakan Algoritma Naïve Bayes Classifier dan K-Nearest Neighbor,” *JURIKOM (Jurnal Ris. Komputer)*, vol. 9, no. 5, p. 1329, 2022, doi: 10.30865/jurikom.v9i5.4774.
  - [13] D. A. Nugroho and F. N. Hasan, “Analisis Sentimen Kegiatan Pembersihan Sampah Pada Media Sosial X Menggunakan SVM dan Naïve Bayes,” *J. Media Inform. Budidarma*, vol. 8, no. 2, p. 843, 2024, doi: 10.30865/mib.v8i2.7562.
  - [14] F. N. H. Khairul Fadli, “Analisis Sentimen Terhadap Penutupan Tiktok Shop Dengan Metode Naïve Bayes,” *J. Media Inform. Budidarma*, vol. 6, no. 3, pp. 407–416, 2024, doi: 10.47065/josh.v6i1.6060.
  - [15] R. A. Syakura and Ade Davy Wiranata, “Analisis Sentimen Ulasan Kepuasan Pengguna Aplikasi Bsi Mobile Dengan Menggunakan Naïve Bayes Rais,” *Indones. J. Comput. Sci.*, vol. 12, no. 2, pp. 284–301, 2024, [Online]. Available: <http://ijcs.stmikindonesia.ac.id/ijcs/index.php/ijcs/article/view/3135>
  - [16] A. Gaizka, A. R. Dzikrillah, and E. Sinduningrum, “Analisis Sentimen Masyarakat Sebelum Dan Sesudah Terpilihnya Gibran Sebagai Cawapres Prabowo Menggunakan Naïve Bayes,” *KLIK Kaji. Ilm. Inform. dan Komput.*, vol. 4, no. 6, pp. 2830–2841, 2024, doi: 10.30865/klik.v4i6.1876.
  - [17] D. Al Ghozi and F. N. Hasan, “Implementation of User Sentiment with Naïve Bayes Algorithm to Analyze LinkedIn Application Regarding Job Vacancies in the Play Store,” *J. Media Inform. Budidarma*, vol. 8, no. 3, p. 1647, 2024, doi: 10.30865/mib.v8i3.7879.
  - [18] N. L. P. C. Savitri, R. A. Rahman, R. Venyutzky, and N. A. Rakhmawati, “Analisis Klasifikasi Sentimen Terhadap Sekolah Daring pada Twitter Menggunakan Supervised Machine Learning,” *J. Tek. Inform. dan Sist. Inf.*, vol. 7, no. 1, pp. 47–58, 2021, doi: 10.28932/jutisi.v7i1.3216.
  - [19] D. Duei Putri, G. F. Nama, and W. E. Sulistiono, “Analisis Sentimen Kinerja Dewan Perwakilan Rakyat (DPR) Pada Twitter Menggunakan Metode Naïve Bayes Classifier,” *J. Inform. dan Tek. Elektro Terap.*, vol. 10, no. 1, pp. 34–40, 2022, doi: 10.23960/jitet.v10i1.2262.
  - [20] M. I. Ahmadi, F. Apriani, M. Kurniasari, S. Handayani, and D. Gustian, “Sentiment Analysis Online Shop on the Play Store Using Method Support Vector Machine (Svm),” *Semin. Nas. ...*, vol. 2020, no. Semnasif, pp. 196–203, 2020, [Online]. Available: <http://jurnal.upnyk.ac.id/index.php/semnasif/article/view/4101>
  - [21] M. K. Amelia Ariska, “Deteksi Hate Speech pada Kolom Komentar Tiktok dengan menggunakan SVM,” *Indones. J. Comput. Sci.*, vol. 12, no. 2, pp. 284–301, 2024, [Online]. Available: <http://ijcs.stmikindonesia.ac.id/ijcs/index.php/ijcs/article/view/3135>
  - [22] A. D. W. Akbar Cleary Syafi'i, “Analisis Sentimen Terhadap Rangka E-SAF Honda Pada Media Sosial X Dengan Algoritma Naïve Bayes Akbar,” *Scientica*, vol. 2, no. 1, pp. 126–138, 2024, doi: 10.30865/klik.v5i1.1993.
  - [23] R. Sulistiawati and M. Kamayani, “Analisis Sentimen Aplikasi Maskapai Penerbangan Lion Air Menggunakan Metode SVM dan Naïve Bayes,” *Indones. J. Comput. Sci.*, vol. 12, no. 2, pp. 284–301, 2024, [Online]. Available: <http://ijcs.stmikindonesia.ac.id/ijcs/index.php/ijcs/article/view/3135>