

PERBANDINGAN METODE NAÏVE BAYES DAN SUPPORT VECTOR MACHINE UNTUK ANALISIS SENTIMEN PADA ULASAN PENGGUNA APLIKASI ALIBABA DI GOOGLE PLAY STORE

Rahmat, Abdul Rahim, Arbansyah

Teknik Informatika, Universitas Muhammadiyah Kalimantan Timur

Jl. Ir. H. Juanda, Samarinda, 75124 Indonesia

2111102441047@umkt.ac.id

ABSTRAK

Alibaba Group adalah perusahaan asal China yang telah memperluas jangkauannya ke pasar internasional, termasuk Indonesia. Dengan semakin banyaknya ulasan produk di Alibaba. Menghadapi masalah dengan semakin banyaknya ulasan produk yang sulit untuk dikelola dan dianalisis secara manual. Hal ini bisa menyebabkan penurunan kualitas layanan dan menyulitkan Alibaba dalam bersaing di pasar e-commerce yang ketat. Untuk itu, analisis sentimen menjadi penting agar Alibaba bisa lebih memahami apa yang diinginkan pelanggan, meningkatkan kualitas layanan, dan tetap kompetitif. Penelitian ini bertujuan untuk menganalisis dan membandingkan performa algoritma Naïve Bayes dan Support Vector Machine dalam analisis sentimen ulasan pengguna aplikasi Alibaba yang diambil dari Google Play Store pada periode Februari 2022 hingga Desember 2024. Sebanyak 10.000 ulasan diproses melalui teknik web scraping dan diklasifikasikan menjadi sentimen positif dan negatif. Proses ini mencakup tahapan preprocessing seperti Case Folding, Stopword Removal, Tokenizing, dan Stemming untuk memastikan data siap digunakan. Hasil pengujian menunjukkan bahwa algoritma Support Vector Machine memiliki keunggulan dalam analisis sentimen positif dengan akurasi sebesar 88,48%, precision 88,95%, recall 99,09%, dan f1-score 93,75%. Sebagai perbandingan, Naïve Bayes memiliki akurasi 87,46%, precision 87,47%, recall 99,93%, dan f1-score 93,29% pada sentimen positif. Namun, pada sentimen negatif, Naïve Bayes lebih unggul dalam precision sebesar 83,33%, sedangkan Support Vector Machine memiliki f1-score lebih tinggi yaitu 26,08%. Secara keseluruhan, Support Vector Machine lebih unggul dalam menganalisis sentiment ulasan pada aplikasi alibaba. Temuan ini memberikan wawasan bagi pengembang aplikasi Alibaba untuk memilih algoritma yang sesuai guna meningkatkan analisis sentimen dan mendukung pengambilan keputusan strategis.

Kata kunci : *Naïve Bayes, Support Vector Machine, Analisis Sentimen, Ulasan Pengguna, Alibaba.*

1. PENDAHULUAN

Alibaba Group adalah perusahaan asal China yang telah memperluas jangkauannya ke pasar internasional, termasuk Indonesia. Dengan masuknya Alibaba Group dan pertumbuhan pesat perusahaan E-commerce lokal, pemerintah Indonesia mulai menerapkan kebijakan untuk menyesuaikan dengan perkembangan ini. Salah satunya adalah pengenalan Paket Kebijakan Ekonomi 14 yang dikenal sebagai Roadmap E-commerce[1].

Orang-orang cenderung memilih belanja online karena dianggap lebih nyaman dan efisien dibandingkan dengan belanja di toko fisik. Selain itu, harga yang lebih murah dan berbagai promosi di platform online menjadi daya tarik utama. Hal ini membantu konsumen menghemat waktu, tenaga, dan uang. Akibatnya, semakin banyak orang beralih ke belanja online karena kemudahan dan kenyamanannya.[2]

Dengan semakin banyaknya ulasan produk di Alibaba. Menghadapi masalah dengan semakin banyaknya ulasan produk yang sulit untuk dikelola dan dianalisis secara manual. Hal ini bisa menyebabkan penurunan kualitas layanan dan menyulitkan Alibaba dalam bersaing di pasar e-commerce yang ketat. Untuk itu, analisis sentimen

menjadi penting agar Alibaba bisa lebih memahami apa yang diinginkan pelanggan, meningkatkan kualitas layanan, dan tetap kompetitif. Tanpa analisis sentimen, Alibaba mungkin akan kehilangan pandangan penting dari ulasan pelanggan, yang dapat mempengaruhi keputusan bisnis di masa depan.[3].

Tujuannya untuk memprediksi, menganalisis suasana publik, suasana hati, dan menggambarkan perasaan netizen secara otomatis terhadap suatu kasus.

Penelitian ini bertujuan untuk membandingkan algoritma klasifikasi Naïve Bayes (NB) dan Support Vector Machine (SVM) berdasarkan data dari google play store yang mengandung sentimen positif, negatif dan netral. Kedua algoritma ini Paling Relevan karena Pengalaman yang Sudah Teruji telah terbukti efektif di berbagai domain, termasuk e-commerce dan pengelolaan data besar, sehingga menawarkan fondasi yang kuat untuk penelitian.[4]

Tujuan dari penelitian ini adalah untuk mengetahui algoritma mana yang memiliki tingkat akurasi tertinggi dalam melakukan analisis sentimen terhadap penggunaan mobil listrik di Indonesia. Metode yang digunakan meliputi metrik akurasi, recall, precision, dan F1-score. Oleh karena itu, peneliti tertarik melakukan penelitian dengan

judul “Perbandingan Metode Naïve Bayes dan Support Vector Machine untuk Analisis Sentimen pada Ulasan Pengguna Aplikasi Alibaba.”

Penelitian yang dilakukan oleh Moh membandingkan dua algoritma dalam analisis sentimen terkait masalah kenaikan harga minyak goreng. Dengan menggunakan dataset yang terdiri dari 9.194 data, Vector Machine) mencapai akurasi sebesar 81,6%, sedangkan algoritma Naïve Bayes mencapai akurasi sebesar 74,06% [5].

Padapenelitian analisis sentiment ini, penulis Naïve Bayes dan Support Vector Machine, untuk menentukan metode mana yang memiliki tingkat akurasi lebih tinggi. Kedua metode ini, Naïve Bayes dan Support Vector Machine, memiliki persamaan karena sama-sama merupakan algoritma supervised learning (pembelajaran yang diawasi) dan dapat digunakan untuk masalah klasifikasi linier, sehingga memungkinkan untuk dibandingkan [6]

2. TINJAUAN PUSTAKA

Penelitian ini menggunakan algoritma seperti Naïve Bayes, Perceptron, dan Support Vector Machine (SVM) sering digunakan dalam analisis sentimen berbasis teks. Naïve Bayes adalah metode statistik yang sederhana namun efektif, dengan tingkat akurasi yang baik dalam klasifikasi Algoritma ini sangat efisien untuk data berskala besar seperti yang dihasilkan oleh aplikasi Alibaba. Karena pendekatannya yang sederhana, algoritma ini memungkinkan pemrosesan cepat tanpa memerlukan daya komputasi besar. [7].

Sedangkan SVM adalah alat klasifikasi yang kuat, yang bertujuan menemukan pemisah optimal antara kategori sentimen. SVM berfokus pada pemisahan maksimal antara kelompok data positif dan negatif SVM dikenal memiliki kinerja yang sangat baik dalam memberikan prediksi akurat untuk data yang tidak seimbang, seperti kasus penipuan yang langka dibandingkan transaksi normal. sehingga memungkinkan klasifikasi yang lebih akurat dan efisien berdasarkan sentimen dalam teks.[8]

Penelitian sebelumnya membahas perbandingan antara Naïve Bayes dan SVM dalam analisis sentimen terhadap 1007 dataset yang diambil dari ulasan aplikasi Zoom Cloud Meetings di Play Store Penelitian ini menemukan bahwa algoritma SVM mencapai akurasi sebesar 90,01 persen, sementara Naïve Bayes mencapai 79,20 persen. Berdasarkan penelitian sebelumnya Hasilnya, SVM memberikan nilai akurasi yang sedikit lebih baik dibandingkan dengan Naïve Bayes [9].

Salah satu perbedaan utama dari penelitian sebelumnya adalah bahwa mereka tidak menggunakan wordcloud, padahal wordcloud memiliki keunggulan dalam mengidentifikasi sentimen positif dan negatif yang sering disampaikan oleh pelanggan [10].

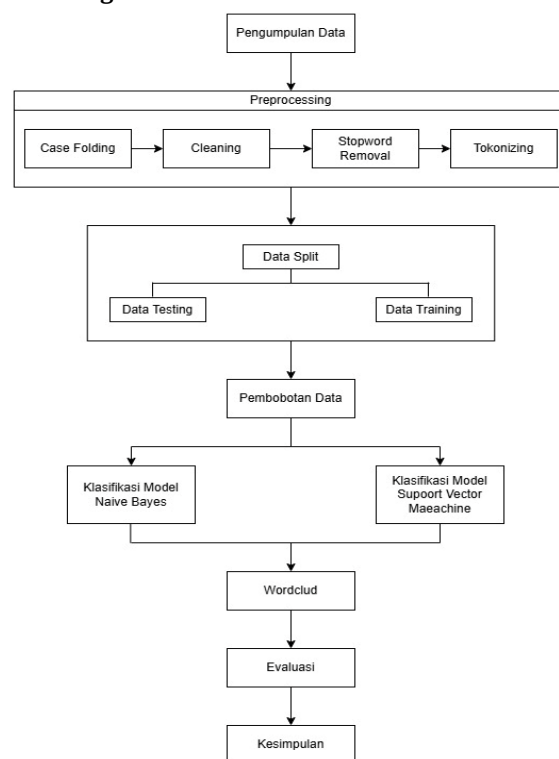
2.1. Objek Penelitian

Marketplace adalah sebuah platform atau situs web yang memungkinkan berbagai penjual untuk menawarkan produk atau layanan mereka kepada pembeli secara online. Berbelanja online atau melalui marketplace memudahkan kita karena bisa dilakukan dari rumah tanpa perlu bertemu langsung dengan penjual. Hal ini mengubah perilaku konsumen, yang kini lebih memilih belanja online dibandingkan belanja secara offline ke pasar [11].

Dalam penelitian ini, peneliti berfokus pada analisis ulasan positif dan negatif serta untuk mengetahui metode yang akurat diantara *Naïve Bayes* dan *Support Vector Machine* dalam menganalisis sentimen ulasan pengguna aplikasi Alibaba..

3. METODE PENELITIAN

3.1. Diagram Alur



Gambar 1. Diagram Alur

3.2. Data Penelitian

Penelitian ini mengumpulkan 10000 data dari ulasan aplikasi Alibaba di Google Play Store Ulasan ini mencakup tiga kategori sentimen positif, dan negatif melalui web scraping. Pengambilan data diambil pada bulan febuari 2022 sampai desember 2024 sebanyak 10000 data.

3.3. Data Split

Penelitian ini menggunakan rasio 80:20 dalam membagi dataset, yang memberikan keuntungan penting dalam pembangunan model. Pertama, alokasi 80% data untuk pelatihan. Kedua, 20% data yang digunakan untuk pengujian cukup memadai untuk

mengevaluasi efektivitas model pada data yang belum pernah dilihat sebelumnya.

3.4. Preprocessing

Preprocessing merujuk pada metode standar yang digunakan dalam analisis sentimen teks bahasa Indonesia. Proses ini terdiri dari empat langkah utama, yaitu penghapusan stop words, case folding, tokenizing, dan stemming.[12]

3.5. Case Folding

Case Folding atau tahap pembakuan, adalah proses mengubah seluruh teks menjadi huruf kecil (lowercase). Pada tahap ini, karakter selain huruf, seperti angka dan tanda baca, dihilangkan karena dianggap memiliki dampak yang minimal terhadap analisis teks.[13].

3.6. Tokenizing

Tokenizing Memecah teks menjadi kata kata atau perasa yang lebih kecil.[14]

3.7. Stopword Removal

Stopword Removal Tahap ini adalah proses menghapus kata-kata yang tidak memiliki makna penting dalam konteks analisis, seperti kata sambung, kata keterangan, conoth nya seperti “di”, “dan”, “yang”, dan.[15]

3.7.1. Stemming

Stemming adalah proses menghilangkan imbuhan pada awalan kata maupun akhiran kata sehingga Kembali menjadi kata dasar (Nursalim & Novita, 2023).

3.8. Pembobotan Data

Pembobotan data ini menggunakan TF-IDF (Term Frequency-Inverse Document Frequency) adalah salah satu metode yang digunakan untuk memberikan bobot pada kata-kata yang dihasilkan dari proses ekstraksi kata dengan menghitung frekuensi kata dalam sebuah dokumen.[16]

3.9. Word Cloud

WordCloud adalah teknik visualisasi yang menampilkan hubungan antara frekuensi kata dengan memperbesar kata-kata yang lebih sering muncul.[17]

3.10. Motode Naïve Bayes

Naïve Bayes adalah metode klasifikasi sederhana yang menghitung probabilitas berdasarkan teorema Bayes, dengan memanfaatkan kombinasi nilai frekuensi dari database [18]. Algoritma Naïve Bayes classification memiliki persamaan yang dapat digunakan sebagai referensi untuk menghitung probabilitas dalam pengambilan keputusan. Persamaan tersebut adalah sebagai berikut.[19] :

$$P(X | H) = \frac{P(X | H) \cdot P(x)}{P(H)} \quad (1)$$

Keterangan:

X : Data dengan *class* yang belum diketahui

H : Hipotesis data X merupakan suatu *class* spesifik

$P(H | X)$: Probabilitas hipotesis H berdasarkan kondisi X

$P(H)$: Probabilitas hipotesis H

$P(X | H)$: Probabilitas X berdasarkan kondisi pada hipotesis H

$P(X)$: Probabilitas X

3.11. Metode Support Vector Meachine

SVM (Support Vector Machine) adalah sebuah metode untuk menemukan hyperplane yang optimal dalam mengklasifikasikan kelas-kelas dalam suatu data. [20] SVM.

Secara umum, dalam klasifikasi biner, persamaan hyperplane pada Support Vector Machine dapat dinyatakan sebagai berikut.[21].

$$f(x) = \text{sign}(w \cdot x + b) \quad (2)$$

Keterangan:

$f(x)$: fungsi Keputusan

w : vector bobot

x : fitur vector data input

b : konstata bias

$\cdot$: operasi dot product

sign : fungsi tanda

3.12. Evaluasi

Evaluasi performa dilakukan untuk menguji hasil klasifikasi dengan mengukur tingkat keakuratan sistem. Parameter yang digunakan untuk mengukur keakuratan ini adalah nilai Accuracy, Recall, Precision, dan f1-score, yang semuanya akan ditampilkan dalam bentuk persentase [22].

3.13. Accuracy

Accuracy Akurasi menunjukkan kemampuan model untuk memprediksi nilai dengan benar, sehingga menggambarkan seberapa tepat dan akurat model tersebut. Dengan kata lain, akurasi adalah rasio antara jumlah prediksi yang benar (baik negatif maupun positif) dengan total data yang ada [23].

$$\text{Accuracy} = \frac{(TP+TN)}{(TP+TN+FP+FN)} \quad (2)$$

Keterangan:

TP : True Positive

TN : True Negative

FP : False Positive

FN : False Negative

3.14. Precision

Precision menggambarkan seberapa akurat model dalam menghasilkan temuan yang benar berdasarkan data yang dimilikinya, serta seberapa cepat sistem merespons permintaan informasi dari pengguna dan memberikan hasilnya [13] Rumus *Precision* yaitu:

$$\text{Precision} = \frac{TP}{TP+FP} \quad (3)$$

Keterangan:

TP : True Positive

FP : False Positive

3.15. Recall

Recall Score adalah tingkat keberhasilan sistem dalam menemukan kembali informasi. Ini mengukur seberapa banyak data positif yang benar-benar terdeteksi oleh model dari semua data yang sebenarnya positif. [24]

$$Recall = \frac{TP}{TP+FN} \times 100\% \quad (7)$$

Keterangan:

TP : True Positive

FN : False Negative

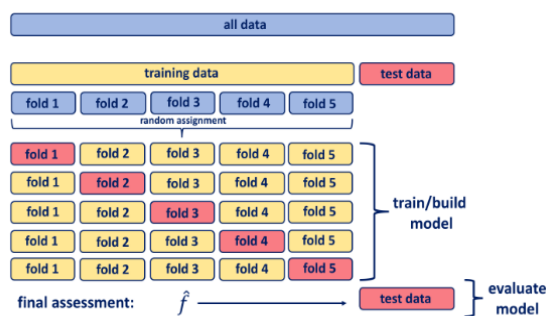
3.16. F1-Score

f1-score adalah rata-rata harmonis dari precision dan recall, yang digunakan untuk mengevaluasi model dengan mempertimbangkan keseimbangan antara precision dan recall yang tinggi [25]

$$F1 - score = 2 \times \frac{Presisi \times Recall}{Presisi + Recall}$$

3.17. K-Fold Cross Validation

K-Fold Cross Validation adalah teknik validasi model untuk mengevaluasi kemampuan model dalam menggeneralisasi data independen. Teknik ini digunakan untuk memprediksi performa model dan mengukur seberapa akurat model tersebut dalam praktik. [26].



Gambar 2. K-Fold Cros Validation

4. HASIL DAN PEMBAHASAN

4.1. Data Selection

Dengan menggunakan modul *Python google_play_scraper* dan *Visual Studio Code* untuk melaksanakan proses web scraping selama fase pemilihan data. Dengan menjalankan kode Python di *Visual Studio Code*, dan berhasil mengumpulkan data ulasan dari Google Play Store menggunakan modul tersebut. Total ulasan yang diperoleh berjumlah 10.000.

4.2. Transformation

Tahap ini menerapkan metodologi TF-IDF (*Term Frequency-Inverse Document Frequency*).

Term Frequency (TF) menghitung seberapa sering sebuah kata muncul dalam sebuah teks, Sementara itu, *Inverse Document Frequency* (IDF) mengukur kelangkaan atau keunikan kata dalam sebuah korpus teks. Kata yang jarang muncul di dokumen lain akan memiliki nilai IDF yang lebih tinggi.

Bentuk TF-IDF untuk data training: (7083, 5439)
Bentuk TF-IDF untuk data testing: (1771, 5439)

Kata-kata yang sering muncul:

bagus: 3.13
good: 3.28
ok: 3.42
mantap: 3.77
aplikasi: 3.84
barang: 3.97
alibaba: 4.11
coba: 4.19
belanja: 4.22
nya: 4.26

Kata-kata yang jarang muncul:

zkaqzxqocdd: 9.17
zkkkkkkkkk: 9.17
zm: 9.17
zpdddrzntsusn: 9.17
zpdf: 9.17
zqx: 9.17
zte: 9.17
ztrimskzdih: 9.17
zy: 9.17
zyukaa: 9.17

Gambar 3. Hasil Tranformation TF-IDF

4.3. Pengujian K-Fold Cross Validation

K-Fold Cross Validation diterapkan pada metode *Naïve Bayes* dan *Support Vector Machine* menggunakan parameter optimal dari evaluasi sebelumnya. Tabel 9 berikut, memberikan gambaran tentang konsistensi performa model dalam setiap lipatan uji.

Tabel 1. Pengujian K-Fold Cross Validation Naïve Bayes

Nilai K-Fold	Hasil
K-1	87.30
K-2	86.88
K-3	87.30
K-4	87.28
K-5	86.58
K-6	86.86
K-7	86.15
K-8	86.44
K-9	87.71
K-10	87.72
Rata-Rata	86.93

Hasil *K-Fold Cross Validation* untuk model *Naïve Bayes*, seperti pada Tabel 9 menunjukkan kinerja yang konsisten dan unggul. Nilai akurasi bervariasi antara 0,8730 hingga 0,8772, dengan rata-rata 0,8693. Fluktuasi minimal sebesar 0,42% menandakan kemampuan generalisasi model yang baik dan ketahanan terhadap variasi partisi data pada setiap fold.

Tabel 2. Pengujian K-Fold Cross Validation Support Vector Machine

Nilai K-Fold	Hasil
K-1	88.01
K-2	86.74
K-3	87.30
K-4	87.57
K-5	87.73
K-6	87.71
K-7	87.57
K-8	90.11
K-9	87.00
K-10	88.84
Rata-Rata	87.66

Hasil *K-Fold Cross Validation* untuk model *Support Vector Machine* pada Tabel 10 menunjukkan kinerja yang konsisten. Akurasi berkisar antara 0,8801 hingga 0,8884 dengan rata-rata 0,8766. Fluktuasi kecil sekitar 0,83% mencerminkan kemampuan generalisasi model yang sangat baik serta ketahanan terhadap variasi partisi data pada setiap fold.

4.4. Perbandingan Pengujian Naïve Bayes & Support Vector Machine

Pengujian perbandingan algoritma *Naïve Bayes* dan *Support Vector Machine* bertujuan untuk menilai kemampuan kedua model tersebut dalam pemrosesan dan prediksi data. Dalam pengujian, kedua model menunjukkan performa yang baik, tetapi masing-masing memiliki keunggulan dan kelemahan tergantung pada metrik evaluasi dan konteks data. Hasil perbandingan antara kedua algoritma berdasarkan hasil ulasan positif dapat dilihat pada table 11 berikut.

Tabel 3. Perbandingan Pengujian Hasil Ulasan Positif

Metrics	Naïve Bayes	Support Vector Machine
Accuracy	87.46%	88.48%
Precision	87.47%	88.95%
Recal	99.93%	99.09%
f1-score	93.29%	93.75%

Tabel 11 menyoroti perbandingan kinerja algoritma *Naïve Bayes* dan *Support Vector Machine* (SVM) dalam evaluasi ulasan positif berdasarkan berbagai metrik. Berikut adalah hasilnya. Akurasi *Naïve Bayes* 87,46% *Support Vector Machine* (SVM) 88,48% *Support Vector Machine* (SVM) memiliki akurasi yang lebih tinggi (1,02%), yang menunjukkan bahwa *Support Vector Machine* (SVM) lebih baik dalam memprediksi data positif dengan benar. Precision *Naïve Bayes* 87,47% *Support Vector Machine* (SVM) 88,95% *Support Vector Machine* (SVM) unggul dalam precision dengan nilai lebih tinggi. Ini berarti prediksi positif dari *Support Vector Machine* (SVM) lebih akurat, dengan lebih sedikit kesalahan memprediksi data positif yang sebenarnya negatif. Recall *Naïve Bayes* 99,93% *Support Vector Machine* (SVM) 99,09% *Naïve Bayes* unggul dalam

recall dengan selisih kecil. Ini menunjukkan bahwa *Naïve Bayes* mampu mendeteksi hampir semua data positif yang ada, meskipun perbedaannya dengan *Support Vector Machine* (SVM) sangat tipis.

F1-Score Naïve Bayes 93,29% *SVM* 93,75% *Support Vector Machine* (SVM) memiliki *F1-Score* yang sedikit lebih tinggi. Ini menunjukkan bahwa *Support Vector Machine* (SVM) memiliki keseimbangan yang lebih baik antara precision dan recall dibandingkan *Naïve Bayes*. Pada kelas positif, *Support Vector Machine* (SVM) memiliki performa lebih baik dalam hal akurasi, precision, dan *F1-Score*, sehingga prediksinya lebih akurat dan seimbang. Sementara itu, *Naïve Bayes* unggul dalam recall, menunjukkan kemampuannya mendeteksi hampir semua data positif dengan sangat baik.

Tabel 4. Perbandingan Pengujian Hasil Ulasan Negatif

Metrics	Naïve Bayes	Support Vector Machine
Accuracy	87.463%	88.48%
Precision	83.33%	72.00%
Recal	02.12%	15.92%
f1-score	03.10%	26.08%

Tabel 12 menunjukkan perbandingan performa *Naïve Bayes* dan *Support Vector Machine* (SVM) dalam evaluasi ulasan negatif berdasarkan beberapa metrik utama. negatif, terutama untuk aplikasi di mana identifikasi ulasan negatif menjadi prioritas utama. Akurasi *Naïve Bayes* 87,463% *Support Vector Machine* (SVM) 88,48% *Support Vector Machine* (SVM) memiliki akurasi sedikit lebih tinggi dibandingkan *Naïve Bayes*. Ini menunjukkan bahwa *Support Vector Machine* (SVM) sedikit lebih andal dalam memprediksi data dengan benar secara keseluruhan. Precision *Naïve Bayes* 83,33% *Support Vector Machine* (SVM) 72,00% *Naïve Bayes* memiliki precision yang lebih tinggi. Artinya, prediksi kelas negatif dari *Naïve Bayes* lebih akurat dibandingkan *Support Vector Machine* (SVM), dengan lebih sedikit data positif yang salah diklasifikasikan sebagai negatif. Recall *Naïve Bayes* 02,12% *Support Vector Machine* (SVM) 15,92% *Naïve Bayes* unggul signifikan dalam recall. Ini menunjukkan bahwa *Naïve Bayes* lebih baik dalam mendeteksi sebagian besar data negatif dibandingkan *Support Vector Machine* (SVM), meskipun performanya masih rendah.

F1-Score Naïve Bayes 43,10% *Support Vector Machine* (SVM) 26,08% *Naïve Bayes* memiliki *F1-Score* yang jauh lebih tinggi dibandingkan *Support Vector Machine* (SVM). Hal ini menunjukkan bahwa *Naïve Bayes* lebih seimbang dalam precision dan recall untuk kelas negatif. Kesimpulan Untuk kelas negatif, *Naïve Bayes* menunjukkan performa yang lebih baik dibandingkan *Support Vector Machine* (SVM). *Naïve Bayes* memiliki precision, recall, dan *F1-Score* yang lebih tinggi, yang menunjukkan kemampuannya lebih baik dalam mengenali data

negatif secara akurat, meskipun recall-nya masih rendah secara absolut. *Support Vector Machine* (SVM), di sisi lain, memiliki akurasi lebih tinggi, tetapi performanya lebih rendah dalam mendeteksi dan memprediksi kelas negatif.

4.5. Word Cloud

Data sentimen positif dari ulasan pengguna aplikasi alibaba menunjukkan bahwa terdapat sebanyak 50 ulasan yang mengandung sentimen positif. bertujuan untuk memberikan gambaran yang lebih jelas tentang kata-kata utama yang sering digunakan oleh pengguna dalam memberikan ulasan positif terhadap aplikasi. Hasil dari analisis ini ditampilkan secara lengkap pada Gambar 5 berikut.



Gambar 4. Wordclod Kata Sentimen Positif

Gambar WordCloud Kata Positif di atas menampilkan distribusi kata-kata yang paling sering muncul dalam ulasan pengguna yang bernada positif terhadap aplikasi alibaba. WordCloud di atas menggambarkan kata-kata yang sering muncul dalam ulasan positif pengguna. Kata-kata seperti "bagus", "mantap", "ok", dan "good" mencerminkan kepuasan pengguna terhadap produk atau layanan yang diterima. Secara keseluruhan, WordCloud ini menggambarkan bahwa sebagian besar ulasan positif berfokus pada kualitas, harga yang baik, dan pengalaman yang memuaskan dengan produk atau aplikasi tersebut.



Gambar 5. Wordclod Kata Sentimen Negatif

Gambar WordCloud Kata Negatif di atas memperlihatkan kumpulan kata yang paling sering muncul dalam ulasan negatif pengguna terhadap aplikasi Alibaba. Kata "harga" dan "aplikasi" tampak mendominasi, menunjukkan bahwa keluhan utama

pengguna berkaitan dengan biaya atau harga serta performa aplikasi yang kurang memuaskan. Secara keseluruhan, WordCloud ini memberikan gambaran mengenai keluhan utama yang disampaikan oleh pengguna, terutama terkait harga, kinerja aplikasi, kualitas layanan, serta biaya tambahan yang dirasakan tidak wajar atau berlebihan.

5. KESIMPULAN DAN SARAN

Hasil menunjukkan bahwa algoritma Support Vector Machine (SVM) unggul dalam analisis sentimen positif dengan akurasi 88,48%, precision 88,95%, recall 99,09%, dan f1-score 93,75%. Naïve Bayes memiliki akurasi 87,46%, precision 87,47%, recall 99,93%, dan f1-score 93,29% pada sentimen positif. Pada sentimen negatif, Naïve Bayes unggul dengan precision 83,33%, namun SVM memiliki f1-score lebih tinggi yaitu 26,08%. Hasil negatif kurang optimal karena lebih banyak data positif yang terkumpul. Secara keseluruhan, SVM lebih baik dalam analisis sentimen positif dan negatif, menjadikannya pilihan andal untuk analisis sentimen, meski perlu peningkatan dalam deteksi ulasan negatif.

Penelitian ini menunjukkan bahwa metode Support Vector Machine (SVM) lebih unggul dalam menganalisis sentimen positif, sedangkan Naïve Bayes lebih baik dalam mengenali sentimen negatif. Untuk meningkatkan hasil, disarankan untuk mencoba teknik pengolahan data yang lebih baik, seperti memperbaiki proses pembersihan teks dan menggunakan metode pembelajaran mesin lain seperti Random Forest atau deep learning (LSTM, BERT). Selain itu, dataset bisa diperluas dengan ulasan dari berbagai platform agar hasilnya lebih akurat.

DAFTAR PUSTAKA

- [1] 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.
- [2] M. Aulia and A. Hermawan, “Analisis Perbandingan Algoritma SVM, Naïve Bayes, dan Perceptron untuk Analisis Sentimen Ulasan Produk Tokopedia,” *J. Media Inform. Budidarma*, vol. 7, no. 4, pp. 1850–1859, 2023, doi: 10.30865/mib.v7i4.6839.
- [3] A. Safira and F. N. Hasan, “PAYLATER MENGGUNAKAN METODE NAIVE BAYES,” vol. 5, no. 1, pp. 59–70, 2023.
- [4] F. Adelia Irawan, A. Rialdy Atmadja, A. Wahana, T. Informatika, and U. Sunan Gunung Djati Bandung, “Analisis Sentimen Ulasan Aplikasi Bank Digital Menggunakan Algoritma Naive Bayes,” *J. Comput. Sci. Inf. Technol. E-ISSN*, vol. 4, no. 2, p. 60, 2024.
- [5] W. Ningsih, B. Alfiaanda, R. Rahmaddeni, and D. Wulandari, “Perbandingan Algoritma SVM

- dan Naïve Bayes dalam Analisis Sentimen Twitter pada Penggunaan Mobil Listrik di Indonesia,” *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 4, no. 2, pp. 556–562, 2024, doi: 10.57152/malcom.v4i2.1253.
- [6] Gishella Septania Al-Husna, Dian Asmarajati, Iman Ahmad Ihsannuddin, and Rina Mahmudati, “Perbandingan Metode Naïve Bayes Dan Support Vector Machine Untuk Analisis Sentimen Pada Ulasan Pengguna Aplikasi LinkedIn,” *STORAGE J. Ilm. Tek. dan Ilmu Komput.*, vol. 3, no. 2, pp. 139–144, 2024, doi: 10.55123/storage.v3i2.3602.
- [7] E. Suryati, Styawati, and A. A. Aldino, “Analisis Sentimen Transportasi Online Menggunakan Ekstraksi Fitur Model Word2vec Text Embedding Dan Algoritma Support Vector Machine (SVM),” *J. Teknol. dan Sist. Inf.*, vol. 4, no. 1, pp. 96–106, 2023.
- [8] G. Rachma, E. F. Ripanti, and H. Sujaini, “Implementasi Support Vector Machine untuk Analisis Sentimen Terhadap Pengaruh Program Promosi Event Belanja pada Marketplace,” vol. 8, no. 3, pp. 508–516, 2022.
- [9] B. A. Maulana, M. J. Fahmi, A. M. Imran, and N. Hidayati, “Analisis Sentimen Terhadap Aplikasi Pluang Menggunakan Algoritma Naive Bayes dan Support Vector Machine (SVM),” *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 4, no. 2, pp. 375–384, 2024, doi: 10.57152/malcom.v4i2.1206.
- [10] A. Hanafiah, A. H. Nasution, Y. Arta, R. Wandri, H. O. Nasution, and J. Mardafora, “Sentimen Analisis Terhadap Customer Review Produk Shopee Berbasis Wordcloud Dengan Algoritma Naïve Bayes Classifier,” *INTECOMS J. Inf. Technol. Comput. Sci.*, vol. 6, no. 1, pp. 230–236, 2023, doi: 10.31539/intecom.v6i1.5845.
- [11] M. A. Budiyanto, F. I. Prasetyo, and T. B. Pamungkas, “Analisis Pengaruh Harga , Citra Merk , dan Promosi Terhadap Keputusan Membeli Produk Online di Marketplace Shopee (Konsumen Shopee Sekaresidenan Pekalongan),” vol. 1, no. 3, 2022.
- [12] S. Khomsah and S. A. Ariwibowo, “Model Text-Preprocessing Komentar Youtube Dalam Bahasa Indonesia,” vol. 1, no. 10, pp. 648–654, 2021.
- [13] V. Fitriyana, L. Hakim, D. Candra, R. Novitasari, and A. Hanif, “Analisis Sentimen Ulasan Aplikasi Jamsostek Mobile Menggunakan Metode Support Vector Machine,” vol. 14, no. April, pp. 40–49, 2023.
- [14] Jasmarijal, Ramaddeni, Junadhi, and M. K. Anam, “Indonesian Journal of Computer Science,” vol. 13, no. 1, pp. 1438–1450, 2024.
- [15] M. F. Al-shufi and A. Erfina, “SENTIMEN ANALISIS MENGENAI APLIKASI STREAMING FILM MENGGUNAKAN ALGORITMA SUPPORT VECTOR MACHINE DI PLAY STORE,” pp. 156–162, 2021.
- [16] A. Sentimen, P. Maskapai, H. C. Husada, and A. S. Paramita, “Analisis Sentimen Pada Maskapai Penerbangan di Platform Twitter Menggunakan Algoritma Support Vector Machine (SVM) Sentiment Analysis of Airline on Twitter Platform Using Support Vector Machine (SVM) Algorithm,” vol. 10, no. 1, pp. 18–26, 2021, doi: 10.34148/teknika.v10i1.311.
- [17] E. Hasibuan and E. Allistair, “ANALISIS SENTIMEN PADA ULASAN APLIKASI AMAZON SHOPPING DI GOOGLE PLAY,” vol. 1, no. 3, pp. 13–24, 2022.
- [18] Syahril Dwi Prasetyo, Shofa Shofiah Hilabi, and Fitri Nurapriani, “Analisis Sentimen Relokasi Ibukota Nusantara Menggunakan Algoritma Naïve Bayes dan KNN,” *J. KomtekInfo*, vol. 10, pp. 1–7, 2023, doi: 10.35134/komtekinfo.v10i1.330.
- [19] K. Anwar, “Analisa sentimen Pengguna Instagram Di Indonesia Pada Review Smartphone Menggunakan Naive Bayes,” *KLIK Kaji. Ilm. Inform. dan Komput.*, vol. 2, no. 4, pp. 148–155, 2022, doi: 10.30865/klik.v2i4.315.
- [20] M. F. N. Fathoni, E. Y. Puspaningrum, and N. A. Sihananto, “Perbandingan Performa Labeling Lexicon InSet dan VADER pada Analisa Sentimen Rohingya di Aplikasi X dengan SVM,” *J. Inform. dan Sains Teknol.*, vol. 1, no. 3, pp. 62–76, 2024, [Online]. Available: <https://doi.org/10.62951/modem.v1i3.112>
- [21] M. Muchtar and R. A. Muchtar, “Perbandingan Metode Knn Dan Svm Dalam Klasifikasi Kematangan Buah Mangga Berdasarkan Citra Hsv Dan Fitur Statistik,” *J. Inform. dan Tek. Elektro Terap.*, vol. 12, no. 2, pp. 876–884, 2024, doi: 10.23960/jitet.v12i2.4010.
- [22] S. Rabbani, D. Safitri, N. Rahmadhani, A. A. F. Sani, and M. K. Anam, “Perbandingan Evaluasi Kernel SVM untuk Klasifikasi Sentimen dalam Analisis Kenaikan Harga BBM,” *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 3, no. 2, pp. 153–160, 2023, doi: 10.57152/malcom.v3i2.897.
- [23] R. Maulana, A. Voutama, and T. Ridwan, “Analisis Sentimen Ulasan Aplikasi MyPertamina pada Google Play Store menggunakan Algoritma NBC,” *J. Teknol. Terpadu*, vol. 9, no. 1, pp. 42–48, 2023, doi: 10.54914/jtt.v9i1.609.
- [24] D. Duei Putri, G. F. Nama, and W. E. Sulistiono, “Analisis Sentimen Kinerja Dewan Perwakilan Rakyat (DPR) Pada Twitter Menggunakan Metode Naive Bayes Classifier,” *J. Inform. dan Tek. Elektro Terap.*, vol. 10, no. 1, pp. 34–40, 2022, doi: 10.23960/jitet.v10i1.2262.
- [25] N. Fibriyanti Arminda, N. Sulistiyowati, and T.

- Nur Padilah, “Implementasi Algoritma Multinomial Naive Bayes Pada Analisis Sentimen Terhadap Ulasan Pengguna Aplikasi Brimo,” *JATI (Jurnal Mhs. Tek. Inform.,* vol. 7, no. 3, pp. 1817–1822, 2023, doi: 10.36040/jati.v7i3.7012.
- [26] M. A. Java, Mohammad Syafrullah, W. Windarto, and P. Painem, “Analisis Sentimen Ulasan Pengguna Aplikasi Threads pada Google Play Store Menggunakan Multinomial Naive Bayes dan Support Vector Machine,” *J. Ticom Technol. Inf. Commun.,* vol. 12, no. 2, pp. 75–80, 2024, doi: 10.70309/ticom.v12i2.112.