Walisongo Journal of Information Technology
Vol 8, No 1 (2026): 01-12. DOI: 10.21580/wjit.2026.8.1.27311
ISSN 2715-0143 (online) | 2714-9048 (printed)

Analisis Komparatif Algoritma Naïve Bayes dan Support Vector Machine untuk Klasifikasi Sentimen Ulasan Pengguna RedNote di Google Play Store

Ananda Akbar Maulana[1], Maya Rini Handayani[2], Masy Ari Ulinuha[3], Khothibul Umam[4]
[123] Universitas Islam Negeri Walisongo Semarang
Corresponding Author: 2208096057@student.walisongo.ac.id

Abstract: *This study compares the performance of Naïve Bayes and Support Vector Machine (SVM) algorithms in classifying user sentiment toward the RedNote (Xiaohongshu) application based on reviews collected from the Google Play Store. A total of 2,364 reviews were collected through web scraping, comprising 1,373 positive reviews and 991 negative reviews. The research process involved text preprocessing, including case folding, non-alphabetic character removal, tokenization, stopword removal, and stemming. Text features were represented using Term Frequency–Inverse Document Frequency (TF-IDF), and the dataset was divided into training and testing sets using an 80:20 ratio. Model performance was evaluated using accuracy, precision, recall, and F1-score. The evaluation results showed that Naïve Bayes achieved an accuracy of 86%, outperforming SVM, which achieved 84% accuracy on the test set. Naïve Bayes also demonstrated a smaller gap between training and testing accuracy, whereas SVM exhibited a larger performance gap, suggesting a potential overfitting tendency. These findings indicate that Naïve Bayes is a more suitable model for the dataset and experimental setting examined in this study. This research provides a comparative evaluation of two machine learning algorithms for classifying RedNote user reviews and highlights the importance of assessing model performance beyond training accuracy.*

Keyword: *Naïve Bayes, RedNote, Sentiment Classification, Support Vector Machine, TF-IDF.*

Copyright © 2026 Walisongo Journal of Information Technology

Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License.

1. Pendahuluan

Perkembangan teknologi digital telah mendorong aplikasi seluler menjadi salah satu sarana utama dalam komunikasi, pertukaran informasi, dan interaksi sosial. Selain menyediakan berbagai layanan digital, aplikasi seluler memungkinkan pengguna menyampaikan pengalaman, penilaian, dan keluhan melalui fitur ulasan. Ulasan tersebut merupakan sumber informasi yang dapat dimanfaatkan untuk memahami persepsi pengguna, mengidentifikasi permasalahan layanan, serta mengevaluasi aspek aplikasi yang perlu ditingkatkan. Namun, jumlah ulasan yang besar dan karakteristik teks yang tidak terstruktur menyebabkan analisis secara manual membutuhkan waktu dan berpotensi menghasilkan interpretasi yang tidak konsisten. Oleh karena itu, pendekatan Natural Language Processing

(NLP) dan pembelajaran mesin diperlukan untuk mengolah ulasan pengguna secara sistematis serta mengidentifikasi kecenderungan opini yang terkandung di dalamnya. Pemanfaatan klasifikasi sentimen untuk menganalisis ulasan aplikasi telah diterapkan pada berbagai layanan digital, termasuk BSI Mobile (Arifiyanti et al., 2023) , Google Play Store secara umum (Nurian, 2023) , dan aplikasi media sosial X (Eskiyaturrofikoh & Suryono, 2024) .

Salah satu platform digital yang relevan untuk dikaji adalah RedNote atau Xiaohongshu, yang mengintegrasikan fitur jejaring sosial, berbagi konten gaya hidup, dan interaksi antarpengguna. Pada Januari 2025, platform ini memperoleh perhatian internasional seiring dengan migrasi sejumlah pengguna TikTok ke Xiaohongshu di tengah ketidakpastian mengenai keberlanjutan akses terhadap TikTok di Amerika Serikat. Fenomena tersebut menunjukkan bahwa migrasi ant platform digital tidak hanya berkaitan dengan perubahan pilihan layanan, tetapi juga melibatkan dinamika interaksi sosial dan adaptasi lintas budaya di lingkungan digital (Xiao & Zhang, 2025; Zhao et al., 2026) .Perkembangan tersebut memperlihatkan pentingnya memahami pengalaman pengguna terhadap platform digital yang mengalami perubahan dinamika pengguna. Dalam konteks ini, ulasan RedNote pada Google Play Store dapat dimanfaatkan sebagai sumber data untuk mengidentifikasi opini positif dan negatif yang disampaikan pengguna. Analisis terhadap ulasan tersebut dapat memberikan gambaran mengenai pengalaman pengguna, termasuk apresiasi terhadap fitur aplikasi maupun keluhan mengenai kendala teknis dan kualitas layanan.

Klasifikasi sentimen merupakan salah satu pendekatan dalam pengolahan teks yang bertujuan mengelompokkan opini ke dalam kategori sentimen tertentu. Naïve Bayes dan Support Vector Machine (SVM) merupakan dua algoritma pembelajaran mesin yang telah digunakan dalam berbagai penelitian klasifikasi sentimen. Keduanya memiliki karakteristik yang berbeda: Naïve Bayes menggunakan pendekatan probabilistik dengan asumsi independensi bersyarat antarfitur, sedangkan SVM membangun batas keputusan berdasarkan prinsip pemaksimalan margin. Pada pengolahan teks, kedua algoritma dapat dikombinasikan dengan representasi fitur Term Frequency–Inverse Document Frequency (TF-IDF) untuk mengubah dokumen tekstual menjadi representasi numerik. Penerapan kombinasi tersebut dapat ditemukan dalam penelitian ulasan aplikasi Teman Bus oleh (Soraya & Indriyanti, 2024) , SMILE Indonesia oleh (Rhomaningtias et al., 2025) , serta Grab Indonesia oleh (Nurfauzi et al., 2025) .

Penelitian terdahulu memperlihatkan bahwa algoritma dengan kinerja terbaik dapat berbeda menurut karakteristik dataset dan konfigurasi eksperimen. (Soraya & Indriyanti, 2024) melaporkan bahwa SVM dengan kernel Radial Basis Function memperoleh akurasi 85%, lebih tinggi daripada Naïve Bayes sebesar 82% pada ulasan aplikasi Teman Bus. Sebaliknya, (Rhomaningtias et al., 2025) memperoleh akurasi 79% untuk Naïve Bayes dan 75% untuk SVM pada ulasan aplikasi SMILE Indonesia. Hasil yang berbeda juga ditemukan oleh (Eskiyaturrofikoh & Suryono, 2024) dalam klasifikasi sentimen aplikasi X, ketika akurasi SVM dan Naïve Bayes sebelum penerapan SMOTE masing-masing mencapai 75,5% dan 75%. Temuan tersebut menunjukkan bahwa performa algoritma tidak dapat ditentukan hanya berdasarkan karakteristik teoretisnya, tetapi perlu dievaluasi berdasarkan data dan rancangan eksperimen yang digunakan.

Perbandingan pada aplikasi lain turut memperlihatkan variasi hasil tersebut. (Mola et al., 2024) mengkaji ulasan aplikasi Halo BCA dengan membandingkan Naïve Bayes, SVM, dan Random Forest, sedangkan (Nurfauzi et al., 2025) membandingkan Naïve Bayes dan SVM pada ulasan Grab Indonesia. Penelitian (Setyani et al., 2026) menggunakan 6.000 ulasan Jamsostek Mobile dan menggabungkan tahapan prapemrosesan, pembobotan TF-IDF, serta evaluasi klasifikasi. Sementara itu, (Kurniawan et al., 2025) mengevaluasi Naïve Bayes, SVM, dan Random Forest pada ulasan aplikasi Ajaib. Penelitian-penelitian tersebut memperlihatkan

bahwa klasifikasi sentimen ulasan aplikasi telah diterapkan pada berbagai konteks layanan digital dengan variasi algoritma dan prosedur pengujian. Oleh karena itu, hasil dari satu aplikasi tidak dapat langsung digeneralisasikan untuk aplikasi lainnya.

Selain itu, (Arifiyanti et al., 2023) menunjukkan bahwa pendekatan supervised learning dapat digunakan untuk menganalisis ulasan pengguna BSI Mobile, tetapi model klasifikasi yang dihasilkan masih memiliki keterbatasan dalam memprediksi sentimen negatif. (Nurian, 2023) juga mengkaji klasifikasi sentimen ulasan pengguna aplikasi Google Play menggunakan Naïve Bayes. Secara keseluruhan, penelitian-penelitian tersebut menunjukkan pentingnya mengevaluasi kinerja model melalui lebih dari satu metrik agar hasil pengujian tidak hanya bergantung pada akurasi keseluruhan. Penggunaan presisi, recall, dan F1-score juga diperlukan untuk memberikan gambaran mengenai kemampuan model dalam mengenali masing-masing kelas sentimen.

Meskipun penelitian mengenai perbandingan Naïve Bayes dan SVM pada ulasan aplikasi telah dilakukan dalam berbagai konteks, penerapan kedua algoritma pada ulasan RedNote tetap perlu dikaji secara khusus. Berdasarkan penelitian terdahulu yang ditelaah dalam studi ini, kajian komparatif yang telah ditemukan berfokus pada aplikasi dan konteks yang berbeda, sehingga hasilnya belum dapat digunakan secara langsung untuk menentukan algoritma terbaik bagi dataset RedNote yang digunakan dalam penelitian ini. Karakteristik bahasa, distribusi kelas sentimen, prosedur prapemrosesan, representasi fitur, serta konfigurasi model merupakan aspek yang perlu diperhatikan ketika menafsirkan hasil perbandingan. Dengan demikian, penelitian ini berfokus pada evaluasi empiris kedua algoritma menggunakan korpus ulasan RedNote yang dikumpulkan dari Google Play Store, bukan pada klaim bahwa salah satu algoritma selalu lebih unggul dalam seluruh permasalahan klasifikasi sentimen.

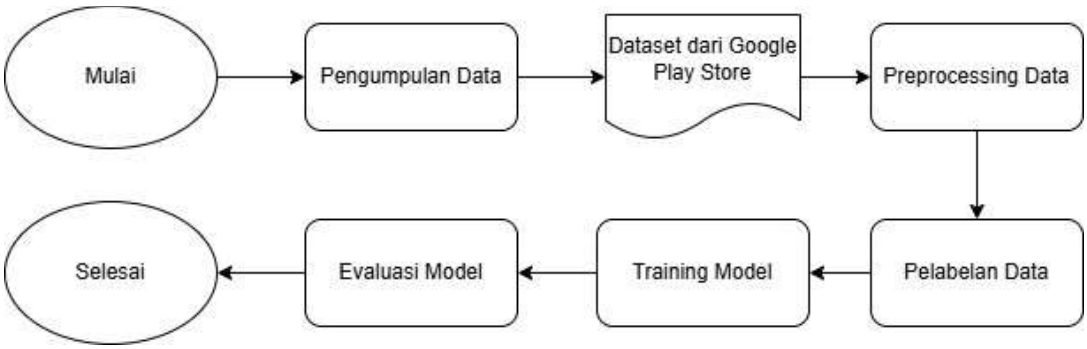
Penelitian ini bertujuan membandingkan kinerja algoritma Naïve Bayes dan SVM dalam mengklasifikasikan ulasan pengguna RedNote ke dalam dua kategori sentimen, yaitu positif dan negatif. Sebanyak 2.364 ulasan dikumpulkan melalui teknik web scraping, yang terdiri atas 1.373 ulasan positif dan 991 ulasan negatif. Pelabelan dilakukan berdasarkan rating pengguna, dengan rating empat dan lima bintang dikategorikan sebagai positif, sedangkan rating satu hingga tiga bintang dikategorikan sebagai negatif. Data selanjutnya melalui tahapan prapemrosesan teks dan ekstraksi fitur menggunakan TF-IDF sebelum digunakan untuk melatih serta menguji kedua model. Kinerja model dievaluasi menggunakan akurasi, presisi, recall, dan F1-score. Melalui perbandingan tersebut, penelitian ini diharapkan dapat memberikan bukti empiris mengenai kinerja kedua algoritma pada dataset yang digunakan serta menjadi pertimbangan dalam pemilihan metode klasifikasi sentimen untuk analisis ulasan aplikasi digital.

2. Metode

Penelitian ini menggunakan pendekatan kuantitatif eksperimental untuk membandingkan kinerja algoritma Naïve Bayes dan Support Vector Machine (SVM) dalam mengklasifikasikan opini pengguna aplikasi RedNote berdasarkan ulasan pada Google Play Store. Pendekatan komparatif ini dilakukan dengan menerapkan kedua algoritma pada dataset yang sama, kemudian mengevaluasi hasil klasifikasinya menggunakan metrik pengukuran yang telah ditentukan. Perbandingan algoritma pada data ulasan aplikasi juga diterapkan dalam penelitian (Salassa et al., 2026; Setyani et al., 2026)

Tahapan penelitian meliputi pengumpulan data, pelabelan sentimen, prapemrosesan teks, pembagian dataset, klasifikasi menggunakan Naïve Bayes dan SVM, serta evaluasi model. Setiap tahapan dilakukan secara berurutan untuk menghasilkan data yang siap digunakan dalam proses klasifikasi dan memungkinkan perbandingan kinerja kedua algoritma.

Gambar 1. Diagram alir penelitian



2.1 Pengumpulan Data

Data penelitian diperoleh dari ulasan pengguna aplikasi RedNote (Xiaohongshu) pada Google Play Store melalui teknik web scraping. Pengumpulan data dilakukan menggunakan Google Play Scraper API dengan bantuan pustaka Python google-play-scraper. Data yang dikumpulkan mencakup nama pengguna (username), teks ulasan, dan rating aplikasi dalam skala satu hingga lima bintang. Informasi lain yang tidak digunakan dalam analisis tidak disertakan dalam dataset penelitian.

Seluruh data yang terkumpul disimpan dalam format CSV untuk mempermudah pengelolaan dan pemrosesan pada tahap berikutnya. Dataset penelitian terdiri atas 2.364 ulasan pengguna. Penggunaan teknik web scraping untuk memperoleh ulasan Google Play Store juga diterapkan dalam penelitian (Salassa et al., 2026; Setyani et al., 2026).

2.2 Pelabelan Data

Pelabelan dilakukan dengan mengacu pada rating bintang yang diberikan pengguna terhadap aplikasi. Ulasan dengan rating empat dan lima bintang dikategorikan sebagai sentimen positif, sedangkan ulasan dengan rating satu hingga tiga bintang dikategorikan sebagai sentimen negatif. Berdasarkan aturan tersebut, dataset terdiri atas 1.373 ulasan positif dan 991 ulasan negative.

Pelabelan berbasis rating digunakan untuk mengubah penilaian numerik pengguna menjadi kategori sentimen yang dapat digunakan dalam proses klasifikasi. Pendekatan pelabelan berdasarkan rating juga digunakan dalam penelitian. Meskipun demikian, label yang diperoleh melalui pendekatan ini merepresentasikan kategori yang ditetapkan berdasarkan rating, sehingga tidak selalu identik dengan interpretasi sentimen yang diperoleh melalui anotasi manual terhadap teks ulasan.

2.4 Pra-Pemrosesan Data

Prapemrosesan dilakukan untuk membersihkan teks ulasan dan menyeragamkan format data sebelum proses klasifikasi. Tahapan yang diterapkan meliputi *case folding*, penghapusan karakter nonalfabetik, tokenisasi, penghapusan *stopwords*, dan *stemming*. Tahapan prapemrosesan tersebut juga digunakan dalam penelitian klasifikasi sentimen ulasan aplikasi oleh (Aji Putra Wicaksana & Andryana, 2026; Mandaya & Fajri, 2026).

Tahap *case folding* mengubah seluruh huruf dalam teks menjadi huruf kecil agar perbedaan kapitalisasi tidak menghasilkan representasi kata yang berbeda. Selanjutnya, karakter nonalfabetik, seperti tanda baca, angka, dan simbol khusus, dihapus untuk mengurangi

unsur yang tidak digunakan dalam analisis. Proses tokenisasi dilakukan untuk memecah teks menjadi unit-unit kata menggunakan fungsi word_tokenize dari pustaka NLTK (*Natural Language Toolkit*).

Tahap berikutnya adalah *stopword removal*, yaitu menghapus kata-kata umum yang dianggap kurang memberikan kontribusi terhadap proses klasifikasi berdasarkan daftar *stopwords* yang digunakan dalam implementasi. Setelah itu, dilakukan *stemming* untuk mengubah kata ke bentuk dasarnya dengan menghilangkan imbuhan sesuai aturan algoritma yang diterapkan. Dalam naskah penelitian ini, metode *Porter Stemmer* disebut digunakan untuk proses *stemming*. Hasil prapemrosesan selanjutnya digunakan pada tahap ekstraksi fitur dan klasifikasi.

2.5 Pembagian Data

Dataset yang telah melalui proses pelabelan dan prapemrosesan dibagi menjadi data pelatihan (training set) dan data pengujian (testing set) dengan rasio 80:20. Dari total 2.364 ulasan, sebanyak 1.891 data digunakan untuk pelatihan, sedangkan 473 data dialokasikan untuk pengujian. Pembagian dilakukan menggunakan fungsi train_test_split dari pustaka scikit-learn dengan stratifikasi berdasarkan label sentimen, sebagaimana pendekatan pembagian data yang digunakan dalam penelitian (Salassa et al., 2026).

Data pelatihan digunakan untuk membangun model klasifikasi, sedangkan data pengujian digunakan untuk menilai kemampuan model dalam mengklasifikasikan ulasan yang tidak digunakan selama pelatihan. Penggunaan pemisahan data pelatihan dan pengujian juga diterapkan dalam penelitian komparatif klasifikasi sentimen oleh (Jatmiko & Dometian, 2026) .

2.6 Klasifikasi Opini

Klasifikasi opini dilakukan dengan menerapkan dua algoritma, yaitu Naïve Bayes dan Support Vector Machine (SVM). Sebelum klasifikasi, teks hasil prapemrosesan direpresentasikan dalam bentuk numerik menggunakan metode *Term Frequency–Inverse Document Frequency* (TF-IDF), sebagaimana disebutkan dalam tahapan pengolahan data pada naskah penelitian ini. TF-IDF digunakan untuk memberikan bobot pada kata berdasarkan frekuensi kemunculannya dalam dokumen dan kelangkaannya pada keseluruhan kumpulan dokumen.

Algoritma Naïve Bayes digunakan sebagai metode klasifikasi berbasis probabilitas dengan asumsi independensi bersyarat ant fitur. Sementara itu, SVM digunakan untuk menentukan batas pemisah antarkelas berdasarkan karakteristik fitur yang direpresentasikan secara numerik. Kedua algoritma diterapkan pada data pelatihan yang sama agar kinerja keduanya dapat dibandingkan secara konsisten. Perbandingan Naïve Bayes dan SVM dalam klasifikasi sentimen ulasan Google Play Store juga dilakukan oleh Setyani et al. (2026), Salassa et al. (2026), dan Jatmiko dan Dometian (2026).

2.7 Evaluasi Model

Evaluasi dilakukan untuk mengukur kemampuan Naïve Bayes dan SVM dalam mengklasifikasikan ulasan ke dalam kategori positif dan negatif. Pengukuran kinerja menggunakan empat metrik, yaitu *accuracy*, *precision*, *recall*, dan *F1-score*. Metrik tersebut digunakan untuk memberikan gambaran mengenai ketepatan prediksi model serta kemampuannya dalam mengenali setiap kelas sentimen. Penggunaan metrik evaluasi tersebut juga ditemukan dalam penelitian klasifikasi sentimen ulasan aplikasi oleh (Aji Putra Wicaksana & Andryana, 2026; Jatmiko & Dometian, 2026; Sari et al., 2026) .

Accuracy mengukur proporsi seluruh prediksi yang benar terhadap jumlah data yang dievaluasi. *Precision* mengukur proporsi prediksi suatu kelas yang benar, sedangkan *recall*

mengukur proporsi data aktual pada suatu kelas yang berhasil dikenali oleh model. Adapun *F1-score* merupakan rata-rata harmonis antara *precision* dan *recall*, sehingga dapat digunakan untuk menilai keseimbangan kedua metrik tersebut.

Selain keempat metrik tersebut, *confusion matrix* digunakan untuk melihat jumlah prediksi benar dan kesalahan klasifikasi pada masing-masing kelas. Hasil evaluasi kedua algoritma kemudian dibandingkan untuk menentukan model yang menunjukkan kinerja lebih baik pada dataset ulasan RedNote yang digunakan dalam penelitian ini.

Evaluasi model dalam penelitian ini menggunakan Confusion Matrix untuk mengukur kinerja klasifikasi, dengan menghitung empat metrik utama: accuracy, precision, recall, dan F1-score.

- Accuracy menunjukkan proporsi prediksi yang benar dari seluruh data yang diuji, dihitung dengan rumus:

  $$\text{Accuracy} = \frac{TP+TN}{TP+FP+FN+TN} \times 100\% \quad (1)$$

- Precision mengukur ketepatan prediksi positif, dihitung dengan:

  $$\text{Precision} = \frac{TP}{TP+FP} \times 100\% \quad (2)$$

- Recall menunjukkan kemampuan model menemukan semua data positif, dengan rumus:

  $$\text{Recall} = \frac{TP}{TP+FN} \times 100\% \quad (3)$$

- F1-Score adalah rata-rata harmonis antara precision dan recall, dihitung menggunakan:

  $$\text{F1-Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \times 100\% \ (4)$$

Pendekatan evaluasi ini memastikan bahwa analisis tidak hanya mempertimbangkan akurasi keseluruhan, tetapi juga faktor lain seperti ketidakseimbangan kelas dan variasi ekspresi dalam ulasan pengguna. Hasil dari evaluasi ini akan digunakan sebagai dasar untuk menentukan algoritma terbaik yang paling sesuai untuk tugas klasifikasi sentimen pada ulasan aplikasi RedNote.

3. Hasil dan Pembahasan

3.1 Pengumpulan Data

Pengumpulan data dilakukan terhadap ulasan pengguna aplikasi Xiaohongshu (RedNote) yang tersedia di Google Play Store. Proses pengambilan data menggunakan teknik *web scraping* berbasis Python dengan memanfaatkan pustaka google-play-scraper. Pengumpulan data difokuskan pada ulasan berbahasa Inggris dari pengguna di Amerika Serikat untuk memperoleh data yang sesuai dengan cakupan penelitian.

Gambar 2: Distribusi rating pengguna



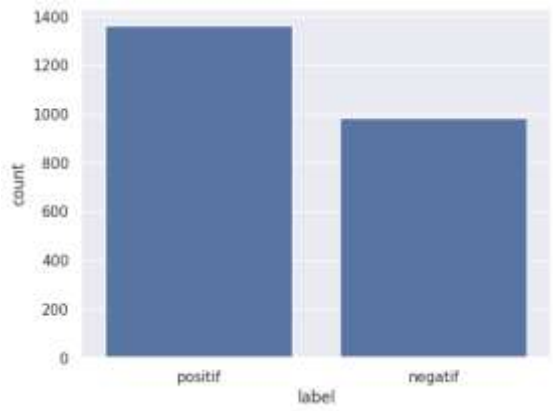
Hasil pengumpulan data awal menghasilkan 2.364 ulasan. Setiap data mencakup tiga komponen utama, yaitu *username* pengguna, rating aplikasi pada skala 1-5 bintang, dan teks ulasan. Rating dan teks ulasan selanjutnya digunakan dalam proses pelabelan sentimen serta tahapan klasifikasi. Distribusi rating pengguna ditampilkan pada gambar distribusi rating.

Data yang terkumpul menjadi dasar untuk mengidentifikasi kecenderungan opini pengguna terhadap RedNote. Keberadaan rating dan teks ulasan dalam satu data memungkinkan rating digunakan sebagai dasar penentuan label, sedangkan teks ulasan menjadi masukan bagi proses klasifikasi. Dengan demikian, tahap pengumpulan data menghasilkan korpus awal yang akan diproses lebih lanjut melalui pelabelan, pra-pemrosesan teks, ekstraksi fitur, dan pengujian algoritma.

3.2 Pelabelan Data

Pelabelan sentimen dilakukan berdasarkan rating yang diberikan pengguna pada Google Play Store. Ulasan dengan rating 4–5 bintang dikategorikan sebagai sentimen positif, sedangkan ulasan dengan rating 1–3 bintang dikategorikan sebagai sentimen negatif. Berdasarkan aturan tersebut, dari total 2.364 ulasan yang terkumpul, sebanyak 1.373 ulasan (58,08%) termasuk dalam kategori positif dan 991 ulasan (41,92%) termasuk dalam kategori negatif.

Gambar 3: Hasil pelabelan



Distribusi tersebut menunjukkan bahwa jumlah ulasan positif lebih banyak dibandingkan ulasan negatif, dengan selisih 382 ulasan. Meskipun demikian, kedua kategori tetap terwakili dalam dataset sehingga dapat digunakan untuk membangun model klasifikasi biner. Hasil pelabelan selanjutnya disimpan dalam format CSV untuk mendukung pengolahan data pada tahap berikutnya.

Secara umum, ulasan positif dalam dataset antara lain memuat kata-kata seperti *helpful*, *recommended*, dan *creative*. Sementara itu, ulasan negatif antara lain memuat kata-kata seperti *error*, *slow*, dan *crash*. Kata-kata tersebut merupakan contoh ekspresi yang ditemukan dalam ulasan, bukan indikator tunggal untuk menentukan sentimen. Penetapan label dalam penelitian ini tetap didasarkan pada rating pengguna.

Gambar 4: Review dengan label

	review	label
0	fantastic tiktok alternative. i love this app ...	positif
1	Absolutely love the welcoming community.	positif
2	Xiaohongshu has an incredible community.	positif
3	I have never been in a space with this many pe...	positif
4	the best social media app	positif

Ulasan positif cenderung menggunakan kata seperti "helpful", "recommended", dan "creative", sementara ulasan negatif sering mengandung kata "error", "slow", dan "crash". ini menghasilkan 1.373 data positif dan 991 data negatif yang siap untuk tahap analisis lebih lanjut. Dataset akhir disimpan dalam format CSV, penyimpanan dalam format ini memungkinkan kemudahan akses dan pengolahan data.

3.3 Pra-Pemrosesan Data

Pra-pemrosesan dilakukan untuk menyiapkan teks ulasan sebelum proses ekstraksi fitur dan klasifikasi. Tahapan ini bertujuan menyeragamkan bentuk teks serta mengurangi karakter yang tidak diperlukan dalam representasi data. Proses yang dijelaskan dalam penelitian meliputi *case folding*, penghapusan karakter nonalfabetik, tokenisasi, penghapusan *stopwords*, dan *stemming*.

Pada tahap *case folding*, seluruh teks ulasan diubah menjadi huruf kecil agar perbedaan kapitalisasi tidak menghasilkan representasi kata yang berbeda. Selanjutnya, karakter nonalfabetik, seperti tanda baca, angka, dan simbol khusus, dihapus untuk mengurangi unsur yang tidak digunakan dalam analisis. Tahap tokenisasi dilakukan dengan memecah teks menjadi unit kata menggunakan fungsi word_tokenize dari pustaka NLTK.

Tahap berikutnya adalah *stopword removal*, yaitu menghapus kata-kata umum yang dianggap kurang informatif bagi proses klasifikasi. Setelah itu, *stemming* digunakan untuk mengubah kata ke bentuk dasarnya sehingga variasi bentuk kata dapat dikurangi. Teks hasil pra-pemrosesan kemudian digunakan sebagai masukan untuk ekstraksi fitur menggunakan TF-IDF.

Gambar 5: Hasil Pra-pemrosesan data

	review
0	fantast tiktok altern love app enjoy person es...
1	absolut love welcom commun
2	xiaohongshu incred commun
3	never space mani peopl internet experi noth jo...
4	best social media app

Hasil pra-pemrosesan ditampilkan pada gambar yang memperlihatkan perubahan teks sebelum dan sesudah pengolahan. Penyajian contoh tersebut penting untuk menunjukkan bagaimana tahapan pemrosesan diterapkan terhadap data ulasan yang menjadi objek penelitian.

3.4 Pembagian Data

Dataset yang telah melalui proses pelabelan dibagi menjadi data pelatihan (*training set*) dan data pengujian (*testing set*) dengan rasio 80:20. Berdasarkan uraian metode, sebanyak 1.891 data digunakan untuk melatih model, sedangkan 473 data dialokasikan untuk menguji kinerja model. Pembagian dilakukan menggunakan fungsi train_test_split dari pustaka scikit-learn dengan parameter stratifikasi berdasarkan label sentimen.

Stratifikasi digunakan untuk mempertahankan proporsi kelas positif dan negatif pada kedua subset agar distribusi label tidak berubah secara drastis setelah pembagian data. Data pelatihan digunakan untuk membangun model Naïve Bayes dan Support Vector Machine (SVM), sedangkan data pengujian digunakan untuk mengevaluasi kemampuan kedua model dalam memprediksi data yang tidak digunakan selama pelatihan.

3.5 Klasifikasi Opini

Klasifikasi opini dilakukan dengan membandingkan dua algoritma pembelajaran mesin, yaitu Naïve Bayes dan Support Vector Machine (SVM). Kedua algoritma digunakan untuk mengelompokkan ulasan ke dalam dua kelas sentimen, yaitu positif dan negatif. Sebelum klasifikasi, teks hasil pra-pemrosesan direpresentasikan dalam bentuk numerik menggunakan metode *Term Frequency–Inverse Document Frequency* (TF-IDF).

TF-IDF memberikan bobot pada kata berdasarkan frekuensi kemunculannya dalam suatu dokumen dan tingkat kelangkaannya pada kumpulan dokumen. Representasi ini memungkinkan teks ulasan diubah menjadi fitur numerik yang dapat diproses oleh algoritma klasifikasi. Selanjutnya, model Naïve Bayes melakukan klasifikasi berdasarkan pendekatan probabilistik, sedangkan SVM menentukan pemisah antarkelas dengan mencari *hyperplane* yang optimal.

Kedua model dilatih menggunakan subset data pelatihan dan kemudian diuji menggunakan subset data pengujian. Performa masing-masing model dibandingkan berdasarkan akurasi, presisi, *recall*, dan F1-score. Perbandingan tersebut digunakan untuk menilai kemampuan model dalam mengklasifikasikan ulasan sekaligus mengidentifikasi perbedaan kinerja pada kelas positif dan negatif.

3.6 Evaluasi Model

Evaluasi dilakukan untuk membandingkan kinerja algoritma Naïve Bayes dan Support Vector Machine (SVM) dalam mengklasifikasikan sentimen ulasan aplikasi RedNote. Metrik yang digunakan meliputi akurasi, presisi, *recall*, dan F1-score. Akurasi menunjukkan proporsi prediksi yang benar dari seluruh data pengujian, sedangkan presisi mengukur ketepatan prediksi pada suatu kelas. *Recall* mengukur kemampuan model dalam mengenali data yang benar-benar termasuk ke dalam kelas tersebut, sementara F1-score merangkum keseimbangan antara presisi dan *recall*.

Tabel 1: Evaluasi algoritma Naive Bayes

Label	Precission	Recall	F1-Score	Support
Negative	0.81	0.86	0.83	197
Positive	0.89	0.85	0.87	272
Accuracy			0.86	469
Macro avg	0.85	0.86	0.85	469
Weighted avg	0.86	0.86	0.86	469

Hasil evaluasi Naïve Bayes. Berdasarkan tabel evaluasi, Naïve Bayes memperoleh akurasi sebesar 0,86 atau 86%. Pada kelas negatif, model menghasilkan presisi 0,81, recall 0,86, dan F1-score 0,83. Pada kelas positif, nilai presisi mencapai 0,89, recall 0,85, dan F1-score 0,87. Nilai tersebut menunjukkan bahwa presisi model lebih tinggi pada kelas positif, sedangkan recall sedikit lebih tinggi pada kelas negatif.

Akurasi data pelatihan yang dilaporkan sebesar 93%, dibandingkan dengan akurasi pengujian sebesar 86%. Perbedaan tersebut menunjukkan adanya penurunan kinerja ketika model diterapkan pada data pengujian. Namun, besarnya perbedaan akurasi saja belum cukup untuk memastikan bahwa model bebas dari *overfitting*. Penilaian yang lebih kuat memerlukan pemeriksaan hasil pengujian, proses validasi, dan konfigurasi model.

Tabel 2: Evaluasi algoritma SVM

Label	Precission	Recall	F1-Score	Support
Negative	0.82	0.80	0.81	197
Positive	0.86	0.88	0.87	272
Accuracy			0.84	469
Macro avg	0.84	0.84	0.84	469
Weighted avg	0.84	0.84	0.84	469

Hasil evaluasi Support Vector Machine. Berdasarkan tabel evaluasi, SVM memperoleh akurasi sebesar 0,84 atau 84%. Pada kelas negatif, model menghasilkan presisi 0,82, recall 0,80, dan F1-score 0,81. Sementara itu, pada kelas positif, nilai presisi mencapai 0,86, recall 0,88, dan F1-score 0,87. Hasil tersebut menunjukkan bahwa model memiliki recall lebih tinggi pada kelas positif daripada kelas negatif, sedangkan F1-score kelas positif juga lebih tinggi daripada F1-score kelas negatif.

Akurasi data pelatihan SVM dilaporkan sebesar 97%, sedangkan akurasi pengujiannya sebesar 84%. Selisih sebesar 13 poin persentase tersebut mengindikasikan adanya kesenjangan kinerja antara data pelatihan dan data pengujian. Kondisi ini dapat menjadi indikasi awal overfitting, tetapi perlu dikonfirmasi melalui validasi tambahan atau pengujian dengan konfigurasi yang berbeda.

Perbandingan kedua algoritma. Berdasarkan akurasi pengujian yang dilaporkan, Naïve Bayes mencapai 86%, sedangkan SVM mencapai 84%. Dengan demikian, Naïve Bayes unggul sebesar 2 poin persentase pada metrik akurasi. Nilai F1-score pada kelas positif untuk kedua model sama-sama sebesar 0,87, sedangkan F1-score kelas negatif pada Naïve Bayes sebesar 0,83 dan pada SVM sebesar 0,81.

Berdasarkan hasil tersebut, Naïve Bayes menunjukkan kinerja keseluruhan yang lebih baik menurut akurasi pengujian, serta nilai F1-score kelas negatif yang lebih tinggi. Meskipun demikian, selisih akurasi kedua model relatif kecil. Oleh karena itu, kesimpulan mengenai algoritma yang paling sesuai perlu mempertimbangkan konsistensi hasil evaluasi dan memastikan bahwa jumlah sampel yang digunakan pada tabel benar-benar sesuai dengan jumlah data pengujian.

4. Kesimpulan

Penelitian ini membandingkan kinerja algoritma Naïve Bayes dan Support Vector Machine (SVM) dalam mengklasifikasikan opini pengguna aplikasi RedNote berdasarkan ulasan di Google Play Store. Dataset penelitian terdiri atas 2.364 ulasan, yang mencakup 1.373 ulasan positif dan 991 ulasan negatif. Tahapan penelitian meliputi pelabelan berdasarkan rating, pra-pemrosesan teks, ekstraksi fitur menggunakan TF-IDF, serta evaluasi model menggunakan akurasi, presisi, recall, dan F1-score.

Berdasarkan hasil pengujian yang dilaporkan, Naïve Bayes memperoleh akurasi sebesar 86%, sedangkan SVM memperoleh akurasi sebesar 84%. Naïve Bayes juga menghasilkan F1-score kelas negatif sebesar 0,83, lebih tinggi dibandingkan SVM sebesar 0,81. Sementara itu, kedua algoritma memperoleh F1-score kelas positif sebesar 0,87. Hasil tersebut menunjukkan bahwa Naïve Bayes memberikan kinerja yang lebih baik berdasarkan akurasi pengujian dan F1-score pada kelas negatif dalam eksperimen ini. Dengan demikian, Naïve Bayes dapat dipertimbangkan sebagai model yang lebih sesuai untuk klasifikasi sentimen pada dataset yang digunakan.

Meskipun demikian, hasil penelitian ini perlu ditafsirkan dengan mempertimbangkan beberapa aspek. Selisih akurasi kedua algoritma relatif kecil, sehingga keunggulan Naïve Bayes belum dapat digeneralisasikan untuk seluruh karakteristik ulasan aplikasi. Selain itu, perbedaan antara akurasi pelatihan dan pengujian pada SVM mengindikasikan kemungkinan overfitting yang masih perlu diperiksa melalui validasi lebih lanjut. Penelitian selanjutnya dapat menguji konfigurasi parameter yang berbeda, menerapkan validasi silang, serta membandingkan algoritma tambahan untuk memperoleh evaluasi yang lebih komprehensif.

5. References

Aji Putra Wicaksana, & Andryana, S. (2026). Sentiment Analysis of Indonesian Mobile Banking Reviews on Google Play Store in 2025: A Comparison of Machine Learning Algorithms. *Jurnal Informatika*, *13*(2), 113–128. https://doi.org/10.31294/informatika.v13i2.13512

Arifiyanti, A. A., Shantika, N. R., & Syafira, A. O. (2023). ANALISIS SENTIMEN ULASAN PENGGUNA BSI MOBILE PADA GOOGLE PLAY DENGAN PENDEKATAN SUPERVISED LEARNING. *Jurnal Informatika Polinema*, *9*(3). https://doi.org/10.33795/jip.v9i3.1003

Eskiyaturrofikoh, E., & Suryono, R. R. (2024). ANALISIS SENTIMEN APLIKASI X PADA GOOGLE PLAY STORE MENGGUNAKAN ALGORITMA NAÏVE BAYES DAN SUPPORT VECTOR MACHINE (SVM). *JIPI (Jurnal Ilmiah Penelitian Dan Pembelajaran Informatika)*, *9*(3). https://doi.org/10.29100/jipi.v9i3.5392

Jatmiko, S., & Dometian, C. (2026). Analisis Sentimen Ulasan Google Play Store: Studi Komparatif Algoritma SVM, Naïve Bayes, dan Logistic Regression. *JURNAL FASILKOM*, *15*(3). https://doi.org/10.37859/jf.v15i3.10016

Kurniawan, N. D., Ferdian, P. R., & Hidayati, N. (2025). Analisis Sentimen Algoritma Naïve Bayes, Support Vector Machine, dan Random Forest Pada Ulasan Aplikasi Ajaib. *Jurnal Nasional Teknologi Dan Sistem Informasi*, *11*(1). https://doi.org/10.25077/teknosi.v11i1.2025.87-97

Mandaya, Y. W., & Fajri, L. R. H. A. (2026). Comparative Analysis of Naive Bayes and Support Vector Machine for Sentiment Analysis of Mobile Banking Application Reviews in Indonesia. *Journal of Advances in Information Systems and Technology*, *8*(1). https://doi.org/10.15294/jaist.v8i1.48426

Mola, S. A. S., Baun, D. L. B., Nunes, I. O., & Sani, M. M. A. R. (2024). ANALISIS SENTIMEN APLIKASI HALO BCA DI GOOGLE PLAY STORE MENGGUNAKAN METODE NAIVE BAYES, SUPPORT VECTOR MACHINE DAN RANDOM FOREST. *HOAQ (High Education of Organization Archive Quality) : Jurnal Teknologi Informasi*, *15*(2). https://doi.org/10.52972/hoaq.vol15no2.p69-79

Nurfauzi, O. M., Hilabi, S. S., Nurapriani, F., & Huda, B. (2025). Analisis Sentimen, Grab Indonesi Analisis Sentimen Grab Indonesia Pada Ulasan Google Play Store Menggunakan Algoritma Naïve Bayes Dan SVM. *SMARTICS Journal*, *11*(1). https://doi.org/10.21067/smartics.v11i1.11789

Nurian, A. (2023). ANALISIS SENTIMEN ULASAN PENGGUNA APLIKASI GOOGLE PLAY MENGGUNAKAN NAÏVE BAYES. *Jurnal Informatika Dan Teknik Elektro Terapan*, *11*(3s1). https://doi.org/10.23960/jitet.v11i3s1.3348

Rhomaningtias, L., Khairunisa, A., Shella May Wara, S., & Maulida Hindrayani, K. (2025). ANALISIS SENTIMEN ULASAN APLIKASI SMILE INDONESIA MENGGUNAKAN METODE NAIVE BAYES DAN SUPPORT VECTOR MACHINE (SVM). *HOAQ (High Education of Organization Archive Quality) : Jurnal Teknologi Informasi*, *16*(1). https://doi.org/10.52972/hoaq.vol16no1.p79-91

Salassa, N. E., Patanduk, A., Yusupa, A., & Rindengan, Y. D. Y. (2026). Comparative Analysis of Naive Bayes and Support Vector Machine for Sentiment Classification of Indonesian-Language Mobile Application Reviews on Google Play Store. *Jurnal Ilmiah Informatika Dan Komputer*, *3*(1). https://doi.org/10.69533/informatech.volume3number1.515

Sari, M. M., Purba, E. H., Supriyanti, D., Firdaus, R., & Astriyani, E. (2026). Pengukuran Kepuasan Pengguna Mobile JKN pada Play Store Menggunakan Naïve Bayes Dan SVM. *Journal CERITA*, *12*(2), 248–261. https://doi.org/10.33050/cerita.v12i2.4405

Setyani, T., Sari, K., Heidy, H. N., Suryono, R. R., Komputer, M. I., Teknik, F., & Komputer, I. (2026). *Sentiment Analysis of Jamsostek Mobile Application Reviews on Google Play Store Using Support Vector Machine and Naive Bayes Algorithms*. *6*(1), 373–384. https://doi.org/https://doi.org/10.57152/malcom.v6i1.2526

Soraya, F. A., & Indriyanti, A. D. (2024). Perbandingan Algoritma Naive Bayes dan Support Vector Machine dalam Analisis Sentimen Aplikasi teman Bus. *Journal of Emerging Information Systems and Business Intelligence (JEISBI)*, *5*(2).

Xiao, Y., & Zhang, J. (2025). "No longer our place": TikTok refugees and the politics of digital migration to Xiaohongshu. *New Media and Society*. https://doi.org/10.1177/14614448251368890

Zhao, H., Ru, P., & Jia, H. (2026). TikTok Users Migration to Xiaohongshu (Rednote): Emotional Dynamics, Platform Governance, and an NCA-SEM Analysis in Cross-Cultural Adaptation. *International Journal of Human-Computer Interaction*, *42*(4). https://doi.org/10.1080/10447318.2025.2530088