KOMPARASI NAÏVE BAYES, SUPPORT VECTOR MACHINE, DAN RANDOM FOREST DALAM ANALISIS SENTIMEN APLIKASI SHOPEE DI GOOGLE PLAY STORE

Putri Anggraini, Winarsih
Sistem Informasi, Universitas Nasional
Jalan Sawo Manila, Pasar Minggu, Jakarta Selatan, Indonesia
winarsih@civitas.unas.ac.id

ABSTRAK

Dalam era digital, analisis sentimen penting untuk memahami opini pengguna terhadap layanan *e-commerce*. Shopee sebagai salah satu platform belanja daring terbesar memiliki ribuan ulasan yang dapat dianalisis untuk meningkatkan layanan. Tantangan utama adalah memilih algoritma *machine learning* yang paling efektif dalam klasifikasi sentimen. Penelitian ini membandingkan performa *Naïve Bayes*, *Support Vector Machine* (SVM), dan *Random Forest* dalam analisis sentimen ulasan pengguna Shopee di Google Play Store. Data dikumpulkan melalui *web scraping* dengan Google Play Scraper, kemudian diproses melalui tahap *pre-processing* (*case folding*, tokenisasi, *stop word removal*, normalisasi, *stemming*, dan rekonstruksi kalimat) dan dikonversi ke representasi numerik menggunakan TF-IDF. Model diuji dengan skema pembagian data (90:10 hingga 10:90) dan dievaluasi menggunakan akurasi, *precision*, *recall*, dan F1-*score*. Hasil menunjukkan bahwa *Naïve Bayes* memiliki akurasi tertinggi 77,13% pada skema 90:10, kemungkinan karena asumsi independensi fitur yang kurang sesuai untuk analisis teks, sementara *Random Forest* dan SVM lebih unggul dengan akurasi masing-masing 87,11% dan 86,90% pada skema 80:20. Visualisasi *confusion matrix*, *word cloud*, dan *line chart* digunakan untuk mengidentifikasi pola sentimen dan tren model. Penelitian ini menyimpulkan bahwa *Random Forest* dan SVM lebih efektif dalam klasifikasi sentimen ulasan *e-commerce* dibandingkan *Naïve Bayes*, meskipun tantangan seperti ketidakseimbangan data dan absennya model *deep learning* masih dapat dieksplorasi untuk peningkatan akurasi di masa depan.

Kata kunci : *analisis sentimen, naïve bayes, random forest, support vector machine, visualisasi data*

1. PENDAHULUAN

Berbagai aspek kehidupan manusia telah diubah oleh kemajuan pesat teknologi informasi dan internet, termasuk industri perdagangan. *E-commerce* atau belanja daring menjadi salah satu sektor bisnis yang berkembang pesat dan terus mengalami peningkatan jumlah pengguna. Shopee adalah salah satu platform *e-commerce* terkemuka di Indonesia dengan banyak fitur yang membuat transaksi jual beli lebih mudah [1].

Popularitas Shopee dapat dilihat dari jumlah unduhan aplikasinya di Google Play Store yang telah melebihi 100 juta dengan lebih dari 15 juta ulasan pengguna.

Analisis sentimen ulasan pengguna di Google Play Store dapat memberikan informasi penting tentang tingkat kepuasan pengguna terhadap suatu aplikasi. Oleh karena itu, ini adalah cara yang berguna untuk menilai kualitas layanan dan pengalaman pengguna. Dengan menggunakan analisis sentimen, pendapat dalam teks dapat dikategorikan menjadi positif, negatif, atau netral [2].

Dengan memanfaatkan analisis sentimen, perusahaan dapat memperoleh informasi penting untuk meningkatkan layanan dan memahami kebutuhan pelanggan. Tujuan analisis sentimen adalah untuk menganalisis opini dan emosi dalam teks secara otomatis dan mengidentifikasi sentimen positif atau negatif pada berbagai tingkatan, seperti dokumen, kalimat, dan aspek [3].

Tujuan utama analisis sentimen adalah untuk mengidentifikasi polaritas data dalam bentuk teks dan menghasilkan informasi dengan menampilkan tiga kategori, yaitu positif, negatif, dan netral [4].

Selain itu, penelitian ini juga bertujuan untuk membandingkan metode mana yang memiliki akurasi terbaik dalam mengklasifikasikan ulasan [5].

Berbagai metode klasifikasi telah dikembangkan untuk meningkatkan akurasi dalam analisis sentimen, di antaranya adalah *Naïve Bayes*, SVM, dan *Random Forest*. *Naïve Bayes* adalah metode berbasis probabilitas yang cepat dan efisien untuk klasifikasi teks [6].

SVM merupakan algoritma berbasis vektor yang bekerja dengan mencari *hyperplane* optimal untuk memisahkan data ke dalam kategori tertentu [7].

Sementara itu, *Random Forest* adalah algoritma berbasis pohon keputusan yang dapat meningkatkan akurasi dengan menggabungkan beberapa pohon keputusan dalam suatu proses klasifikasi [2].

Setelah melakukan perbandingan antara ketiga metode tersebut, penelitian ini diharapkan dapat memberikan rekomendasi algoritma terbaik untuk mengklasifikasikan sentimen pengguna setelah melakukan analisis terhadap performa algoritma *Naïve Bayes*, SVM, dan *Random Forest* dalam melakukan analisis sentimen pada ulasan pengguna terhadap aplikasi Shopee yang tersedia di Google Play Store.

2. TINJAUAN PUSTAKA

Pada bagian ini akan dibahas landasan teori yang mendukung penelitian, termasuk konsep analisis sentimen, penerapan *machine learning*, serta penggunaan Google Colab dalam proses pelatihan model.

2.1. Analisis Sentimen

Analisis sentimen menempatkan pendapat dalam teks ke dalam klasifikasi positif, negatif, atau netral. Proses ini memanfaatkan teknik data *mining* untuk mengumpulkan dan menganalisis informasi dari ulasan pengguna [5].

Hal ini juga bertujuan untuk melihat pendapat teks dan memberikan tolak ukur baik atau tidaknya suatu layanan atau produk menurut pelanggan [8].

2.2. Machine Learning

Machine learning digunakan untuk proses klasifikasi dalam pembelajaran terawasi. *Support Vector Machine* (SVM) adalah algoritma yang bertujuan untuk menemukan *hyperplane* terbaik untuk memisahkan dua kelas dalam ruang *input* [9].

Dengan beberapa metode, sistem dapat belajar dari data dan membuat prediksi dan keputusan berdasarkan informasi tersebut [2].

2.3. Google Colab

Google Colab merupakan platform berbasis *cloud* yang mendukung eksekusi kode Python, yang sering digunakan untuk tugas-tugas seperti data *mining* dan *machine learning*. Google Colab memiliki banyak keuntungan yang membuatnya menjadi favorit untuk analisis data dan *machine learning* [10].

Pengambilan data dilakukan di Google Colab menggunakan bahasa Python dengan bantuan *library* tertentu [11].

2.4. Naïve Bayes

Dengan asumsi bahwa fitur independen, metode klasifikasi probabilistik *Naive Bayes* menggunakan *Teorema Bayesian*. Karena kesederhanaannya, kecepatan, dan akurasi yang tinggi dalam memproses *database* yang besar, algoritma ini sering digunakan dalam analisis teks dan pengklasifikasian dokumen [7].

Di bawah ini merupakan persamaan dasar metode berdasarkan *teorema Bayes*, dengan rumus adalah sebagai berikut:

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

2.5. Support Vector Machine

Metode *machine learning* (SVM) untuk pengklasifikasian teks sangat disukai dan berhasil dalam banyak domain. Untuk memaksimalkan hasil klasifikasi, algoritma ini dapat menemukan *hyperplane* yang membedakan dua kelas yang berbeda [12]. Persamaan dasar metode untuk SVM ini adalah sebagai berikut:

$$y(x) = wT\varphi(x) + b \quad (2)$$

2.6. Random Forest

Random Forest adalah metode untuk regresi dan klasifikasi yang termasuk dalam kelompok algoritma *ensemble* [9].

Metode ini membangun beberapa pohon keputusan dan menggabungkan hasil prediksi dari setiap pohon untuk mencapai prediksi yang lebih stabil dan akurat. Membangun pohon prediksi dengan menggunakan teknik *sampling bootstrap* dilakukan. Setiap pohon keputusan diprediksi dengan prediktor acak.

2.7. Evaluasi Performa

Evaluasi performa adalah proses mengukur sejauh mana suatu sistem atau model berhasil atau gagal dalam mencapai tujuan tertentu. Dalam konteks *machine learning*, Metode untuk mengevaluasi kemampuan model untuk menyelesaikan tugas klasifikasi tertentu dikenal sebagai evaluasi performa. Evaluasi hasil bisa ditampilkan menggunakan *confusion matrix* yang berisi informasi asli dan yang diprediksi, dari mana dapat diketahui akurasi, presisi, dan *recall* [2].

Akurasi adalah metrik evaluasi yang mengukur persentase prediksi benar dari total prediksi yang dilakukan. Metrik ini memberikan gambaran umum tentang performa model dalam klasifikasi. Akurasi yang tinggi menunjukkan model mengenali pola dengan baik, sedangkan akurasi rendah menandakan banyak kesalahan prediksi. Namun, jika data tidak seimbang, akurasi bisa kurang representatif, sehingga perlu didukung metrik lain. Akurasi dihitung dengan rumus berikut:

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

Presisi adalah metrik evaluasi yang mengukur seberapa akurat model dalam memprediksi kelas positif. Metrik ini menunjukkan proporsi prediksi positif yang benar dibandingkan dengan total prediksi positif yang dibuat oleh model. Nilai presisi yang tinggi menunjukkan bahwa model jarang memberikan prediksi positif yang salah, sehingga lebih dapat diandalkan dalam menghindari *false positive*. Namun, presisi saja tidak cukup untuk menilai performa model secara keseluruhan, terutama jika *dataset* tidak seimbang. Oleh karena itu, presisi sering digunakan bersama metrik lain seperti *recall* dan *F1-score*. Presisi dihitung dengan rumus berikut:

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

Recall adalah metrik evaluasi yang mengukur seberapa baik model dalam mendeteksi semua kasus

positif yang sebenarnya ada dalam data. Nilai *recall* yang tinggi menunjukkan bahwa model mampu mengidentifikasi hampir semua sampel positif, sehingga sangat berguna dalam situasi di mana mendeteksi positif lebih penting daripada menghindari *false positive*, seperti dalam diagnosa penyakit atau deteksi penipuan. Namun, *recall* yang tinggi sering kali berbanding terbalik dengan presisi, sehingga diperlukan keseimbangan antara keduanya. *Recall* dihitung dengan rumus berikut:

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

F1-score adalah metrik evaluasi yang menggabungkan *precision* dan *recall* dalam satu nilai harmonis untuk memberikan gambaran yang lebih seimbang tentang performa model, terutama ketika terdapat ketidakseimbangan antara keduanya. Metrik ini sangat berguna dalam situasi di mana penting untuk mempertimbangkan baik kemampuan model dalam mengidentifikasi kelas positif (*recall*) maupun keakuratan prediksi positifnya (*precision*). *F1-score* dihitung sebagai rata-rata harmonik dari *precision* dan *recall*, sehingga nilai tinggi menunjukkan bahwa model memiliki keseimbangan yang baik antara keduanya. Perhitungan *F1-score* dilakukan dengan rumus berikut:

$$\text{F1} = \frac{(2 \times \text{precision} \times \text{recal})}{(\text{precision} + \text{recall})} \quad (6)$$

2.8. Penelitian Terdahulu

Berbagai penelitian telah dilakukan untuk menganalisis sentimen ulasan pengguna terhadap aplikasi di Google Play Store menggunakan berbagai metode klasifikasi. [2] melakukan analisis sentimen terhadap aplikasi Mobile Legend dengan algoritma *Naïve Bayes*, SVM, *Random Forest*, *Decision Tree*, dan *Logistic Regression*. Hasil penelitian menunjukkan bahwa algoritma SVM memiliki performa terbaik dalam mengklasifikasikan sentimen ulasan pengguna. Penelitian lain oleh [13] menganalisis sentimen ulasan pengguna aplikasi MySAPK BKN menggunakan *Naïve Bayes* dan SVM, dengan hasil akurasi masing-masing 92,47% dan 94,14%.

Sementara itu, [12] meneliti sentimen terhadap aplikasi Ruangguru menggunakan tiga algoritma, yaitu *Naïve Bayes*, *Random Forest*, dan SVM, dengan hasil terbaik diperoleh dari *Random Forest* dengan akurasi 97,16% dan AUC sebesar 0,996.

Selain itu, [10] mengombinasikan metode TF-IDF dan *Long Short-Term Memory* (LSTM) untuk menganalisis sentimen aplikasi Shopee, yang menghasilkan akurasi sebesar 83%. [11] meneliti ulasan aplikasi GBWhatsApp menggunakan *Naïve Bayes Classifier* (NBC) dan *Random Forest Classifier* (RFC), dengan hasil menunjukkan bahwa NBC memiliki performa lebih baik dengan akurasi 71,43% dibandingkan RFC yang memiliki akurasi 64,94%.

Penelitian lainnya oleh [4] membandingkan tiga algoritma, yaitu *Naïve Bayes*, SVM, dan *K-Nearest Neighbor* (KNN), dalam menganalisis reaksi masyarakat terhadap aplikasi Peduli Lindungi. Hasil penelitian menunjukkan bahwa SVM memiliki akurasi tertinggi sebesar 76,5%, sedangkan *Naïve Bayes* dan KNN masing-masing memiliki akurasi 72,3% dan 59,1%.

Lebih lanjut, [14] meneliti pengaruh komposisi pembagian data terhadap performa akurasi analisis sentimen menggunakan *Naïve Bayes* dan SVM. Hasil penelitian menunjukkan bahwa model SVM dengan rasio 80:20 memberikan akurasi tertinggi sebesar 76,66% dengan F1-*score* 77%. [7] membandingkan algoritma data *mining* dalam analisis sentimen aplikasi PeduliLindungi dan menemukan bahwa *Naïve Bayes* dengan TF-IDF *Vectorizer* memiliki akurasi tertinggi sebesar 89,05%. Sementara itu, [6] melakukan perbandingan enam algoritma dalam analisis sentimen ulasan Shopee dan menemukan bahwa SVM memiliki akurasi tertinggi sebesar 88%, diikuti oleh *Extra Trees Classifier* dan *Logistic Regression* dengan akurasi 86% dan 85%.

Terakhir, [5] membandingkan metode *Logistic Regression*, *Multinomial Naïve Bayes*, SVM, dan KNN dalam analisis sentimen ulasan Gojek, dengan hasil menunjukkan bahwa *Logistic Regression* memiliki performa terbaik dengan akurasi 82,45%.

Berdasarkan tinjauan pustaka di atas, dapat disimpulkan bahwa algoritma yang sering digunakan dalam analisis sentimen adalah *Naïve Bayes*, SVM, dan *Random Forest*. SVM sering kali menunjukkan performa yang lebih baik dibandingkan dengan metode lainnya, meskipun *Random Forest* juga memiliki tingkat akurasi yang tinggi. Selain itu, pemilihan teknik *feature extraction* seperti TF-IDF dan *Count Vectorizer* berpengaruh terhadap akurasi model. Oleh karena itu, penelitian ini akan melakukan analisis sentimen terhadap ulasan pengguna Shopee di Google Play Store dengan membandingkan algoritma *Naïve Bayes*, SVM, dan *Random Forest*, serta mengeksplorasi pengaruh pembagian data terhadap performa model.

3. METODE PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif dengan teknik pengumpulan data melalui *web scraping* menggunakan *library* Google Play Scraper dalam bahasa pemrograman Python. Teknik ini memungkinkan pengambilan data ulasan pengguna dari Google Play Store secara otomatis, termasuk informasi seperti nama pengguna, rating, tanggal ulasan, serta teks ulasan. Setelah data diperoleh, tahap selanjutnya adalah persiapan data, yang mencakup pembuatan *dataset*, penambahan atau penghapusan kolom yang tidak diperlukan, serta pengecekan kelengkapan data untuk memastikan tidak ada nilai yang hilang (*missing values*).

Setelah data dikumpulkan, dilakukan serangkaian proses *pre-processing* untuk

meningkatkan kualitas data sebelum digunakan dalam analisis sentimen. Tahapan *pre-processing* yang diterapkan meliputi *case folding*, yaitu mengubah seluruh teks menjadi huruf kecil untuk menghindari perbedaan pengenalan kata akibat huruf kapital, serta tokenisasi, yang memecah kalimat ke dalam bentuk kata-kata atau token individu. Selanjutnya, dilakukan *stopword removal* dengan menghapus kata-kata yang tidak memiliki makna signifikan dalam analisis sentimen, seperti "dan", "yang", "di", "ke", dan sebagainya. Proses normalisasi juga diterapkan untuk mengubah kata-kata tidak baku menjadi bentuk baku, misalnya "gk" menjadi "tidak" atau "trs" menjadi "terus". Setelah itu, dilakukan stemming untuk mengembalikan kata-kata ke dalam bentuk dasar menggunakan algoritma Porter Stemmer atau Sastrawi dalam bahasa Indonesia. Contohnya, kata "membeli" dan "membelikan" akan dikembalikan ke bentuk dasar "beli". Setelah melalui tahap-tahap ini, kata-kata yang telah dibersihkan disusun kembali menjadi kalimat yang dapat digunakan untuk analisis lebih lanjut.

Tahapan berikutnya adalah pelabelan sentimen terhadap data ulasan. Dalam penelitian ini, sentimen dikategorikan menjadi tiga kelas utama, yaitu sentimen positif, yang menunjukkan kepuasan pengguna terhadap aplikasi, sentimen negatif, yang menyatakan ketidakpuasan pengguna terhadap aplikasi, serta sentimen netral, yang tidak memiliki kecenderungan emosional yang jelas. Pelabelan dapat dilakukan secara otomatis berdasarkan skor rating, misalnya rating 4-5 dianggap positif, rating 1-2 negatif, dan rating 3 netral, atau menggunakan pendekatan berbasis leksikon yang mencocokkan kata-kata dengan kamus sentimen. Setelah pelabelan dilakukan, jumlah masing-masing kategori sentimen divisualisasikan menggunakan histogram untuk melihat distribusi data ulasan yang diperoleh.

Untuk mengubah data teks menjadi bentuk numerik yang dapat digunakan dalam algoritma *machine learning*, digunakan metode *Term Frequency - Inverse Document Frequency* (TF-IDF). Teknik ini memberikan bobot lebih tinggi pada kata-kata yang sering muncul dalam satu dokumen tetapi jarang muncul dalam keseluruhan *dataset*, sehingga meningkatkan efektivitas pemodelan. Setelah data dikonversi menjadi representasi numerik, dilakukan pemodelan menggunakan beberapa algoritma *machine learning*, yaitu *Naïve Bayes*, algoritma probabilistik yang sering digunakan untuk klasifikasi teks, *Support Vector Machine* (SVM), algoritma yang bekerja dengan mencari *hyperplane* terbaik untuk memisahkan kelas sentimen, serta *Random Forest*, algoritma berbasis pohon keputusan yang bekerja dengan membentuk banyak pohon dan menggabungkan hasilnya untuk meningkatkan akurasi klasifikasi. Model dilatih menggunakan data latih (*training data*) dan diuji dengan data uji (*test data*) dengan komposisi tertentu, misalnya 80% data latih dan 20% data uji.

Setelah model dilatih, dilakukan evaluasi performa menggunakan beberapa metrik, yaitu akurasi, yang menunjukkan persentase prediksi yang benar dibandingkan total data uji, presisi, yang mengukur proporsi prediksi positif yang benar terhadap keseluruhan prediksi positif, serta *recall*, yang mengukur kemampuan model dalam mendeteksi seluruh data dari suatu kelas tertentu. Selain itu, digunakan pula F1-*score*, yang merupakan *harmonic mean* dari presisi dan *recall* untuk memberikan gambaran seimbang tentang performa model, serta *confusion matrix*, yaitu matriks yang menunjukkan jumlah prediksi benar dan salah untuk setiap kategori sentimen.

Hasil analisis sentimen divisualisasikan menggunakan beberapa metode untuk mendapatkan wawasan lebih lanjut. *WordCloud* digunakan untuk menampilkan kata-kata yang paling sering muncul dalam ulasan pengguna dengan ukuran yang proporsional terhadap frekuensinya, sedangkan *Line Chart* digunakan untuk membandingkan performa berbagai model berdasarkan metrik evaluasi guna menentukan model terbaik.

Berdasarkan hasil analisis dan evaluasi, penelitian ini akan menyimpulkan model mana yang memberikan performa terbaik untuk analisis sentimen ulasan pengguna aplikasi serta memberikan rekomendasi berdasarkan temuan penelitian.

4. HASIL DAN PEMBAHASAN

Pada bagian ini untuk menganalisis dan membandingkan performa model, hasil pengujian divisualisasikan melalui *confusion matrix*, tabel, dan *line chart*, guna memberikan pemahaman lebih jelas terhadap efektivitas masing-masing algoritma dalam analisis sentimen.

4.1. Performa Algoritma

Penelitian ini mengevaluasi akurasi model klasifikasi pada berbagai pembagian data latih dan uji (90:10 hingga 10:90). Akurasi digunakan sebagai metrik utama untuk menilai kemampuan model dalam memprediksi data baru, yang penting untuk mengukur generalisasi model.

Hasil perbandingan performa ketiga algoritma dapat dilihat pada Tabel 1, yaitu membandingkan akurasi tiga algoritma *machine learning* berdasarkan pembagian data. Kolom "*test size*" menunjukkan persentase data uji, sementara sisanya digunakan untuk pelatihan. Misalnya, *test size* 0.1 berarti 90% data untuk pelatihan dan 10% untuk pengujian. Kolom "*Naive Bayes*", "*Random Forest*", dan "SVM" menampilkan akurasi masing-masing model dalam memprediksi data uji.

Tabel 1. Perbandingan Performa

Test Size	Naïve Bayes	Random Forest	SVM
90:10	0.771370	0.864993	0.869064
80:20	0.763908	0.871099	0.866689
70:30	0.753505	0.858209	0.860697
60:40	0.739484	0.839722	0.847015
50:50	0.727815	0.819539	0.833379

Test Size	Naïve Bayes	Random Forest	SVM
40:60	0.715400	0.810267	0.818182
30:70	0.689281	0.785036	0.803353
20:80	0.669267	0.746523	0.779003
10:90	0.656490	0.702397	0.756144

Tabel 1 menunjukkan akurasi tiga model pada berbagai ukuran data uji (10%–90%). Akurasi mengukur proporsi prediksi benar, yang mencerminkan kemampuan model dalam generalisasi.

Naive Bayes memiliki akurasi 77.137% (*test size* 10%) hingga 65.649% (*test size* 90%). Model ini cepat dan efisien, tetapi kinerjanya menurun dengan meningkatnya data uji karena asumsi independensi fitur yang kurang sesuai untuk dataset kompleks.

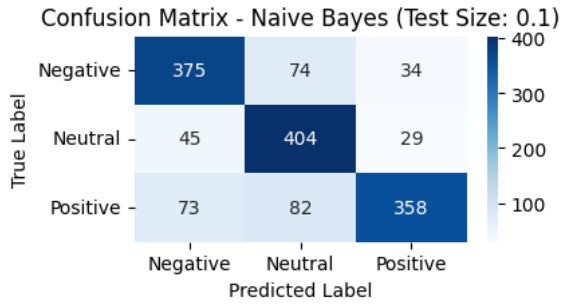
Random Forest lebih stabil dengan akurasi 86.499%–70.239%. Sebagai model *ensemble*, ia lebih akurat dan tahan terhadap *overfitting*, meskipun tetap terpengaruh oleh peningkatan data uji.

SVM menunjukkan performa tinggi, dari 86.906% hingga 75.614%. Model ini efektif dalam memisahkan kelas dengan *hyperplane* optimal, tetapi akurasinya menurun pada dataset besar.

Kesimpulannya, *Random Forest* dan SVM lebih unggul dalam akurasi dibandingkan *Naive Bayes*, terutama untuk dataset besar. Namun, pemilihan model bergantung pada *trade-off* antara akurasi, kompleksitas, dan efisiensi komputasi.

4.2. Confusion Matrix

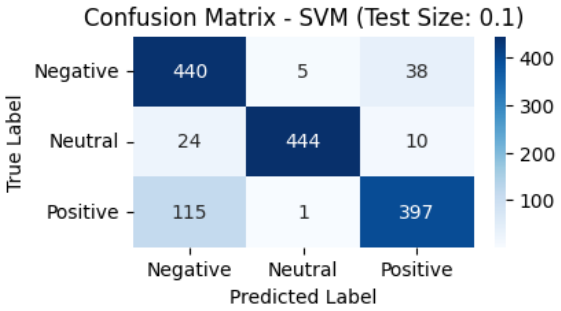
Berikut hasil performa algoritma *Naïve Bayes* dengan pembagian data *training* dan data *test* sebesar 90:10 menggunakan *confusion matrix*.



Gambar 1. Confusion Matrix Naïve Bayes

Confusion matrix pada Gambar 1 menunjukkan bahwa model berhasil mengidentifikasi 375 data negatif, 404 data netral, dan 358 data positif. Namun, terdapat kesalahan klasifikasi, khususnya antara kelas netral dan positif, dengan 82 data netral salah dikategorikan sebagai positif. Hal ini menunjukkan tantangan dalam membedakan dua kelas tersebut. Meskipun akurasinya cukup baik, performa dapat ditingkatkan dengan optimasi fitur atau eksplorasi algoritma lain.

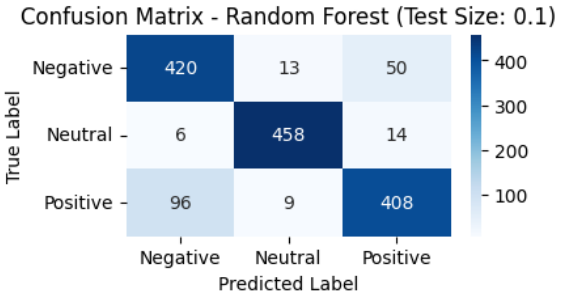
Sedangkan berikut Gambar 2 merupakan hasil performa algoritma SVM dalam melakukan analisis dengan pembagian data *training* dan data *test* sebesar 90:10 menggunakan *confusion matrix*.



Gambar 2. Confusion Matrix SVM

Confusion matrix pada Gambar 2 untuk model SVM menunjukkan performa yang lebih baik dalam klasifikasi tiga kelas. Model berhasil mengidentifikasi 440 data negatif, 444 data netral, dan 397 data positif, dengan sedikit kesalahan klasifikasi. Kesalahan utama terjadi dalam membedakan data positif dan negatif. Dengan *tuning* parameter atau penambahan data, model ini berpotensi meningkatkan akurasi lebih lanjut.

Selanjutnya, Gambar 3 menunjukkan hasil performa algoritma *Random Forest* dalam melakukan analisis dengan pembagian data *training* dan data *test* sebesar 90:10 menggunakan *confusion matrix*.



Gambar 3. Confusion Matrix Random Forest

Confusion Matrix untuk model *Random Forest* (*Test Size*: 90:10) menunjukkan bahwa model berhasil mengklasifikasikan 420 sampel negatif, 458 netral, dan 408 positif dengan benar. Namun, terdapat kesalahan klasifikasi, seperti 50 sampel negatif diprediksi sebagai positif, 14 netral sebagai positif, dan 96 positif sebagai negatif. Model cukup baik dalam mengenali kelas netral, tetapi masih memiliki kesalahan tinggi pada kelas positif yang diprediksi sebagai negatif.

4.3. Wordcloud

Lalu, visualisasi berikutnya menggunakan *wordcloud*, yang menampilkan kata-kata yang paling sering digunakan dalam ulasan pengguna Shopee. Semakin banyak kata yang ditampilkan, semakin sering kata tersebut muncul dalam ulasan.

Gambar 4. *Wordcloud* Sentimen Positif

Visualisasi *wordcloud* Gambar 4 menunjukkan bahwa istilah seperti "belanja", "bagus", "kirim", dan "aplikasi" mendominasi, menandakan aspek utama yang diperhatikan pengguna dalam ulasan positif.

Selanjutnya Gambar 5, adalah *wordcloud* untuk sentimen negatif, yang menunjukkan kata-kata yang sering muncul dalam ulasan negatif pengguna Shopee.

Gambar 5. *Wordcloud* Sentimen Negatif

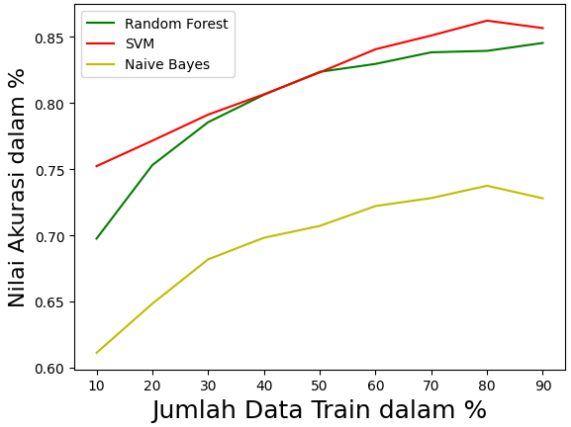
Pada Gambar 5, kata-kata seperti "tolong", "sudah", dan "tidak" sering muncul, menunjukkan adanya ketidakpuasan pengguna, terutama terkait kendala pengiriman atau pengalaman penggunaan aplikasi. Selain visualisasi *wordcloud* untuk sentimen positif dan negatif, berikut adalah *wordcloud* untuk sentimen netral.

Gambar 6. *Wordcloud* Sentimen Negatif

Pada Gambar 6, kata-kata seperti “barang”, “paket”, dan “kirim” mendominasi, mencerminkan ulasan yang lebih bersifat deskriptif atau berisi kritik dan saran.

4.4. Line Chart

Selanjutnya, dilakukan visualisasi perbandingan ketiga algoritma menggunakan diagram *line chart*, yang ditampilkan pada Gambar 7.

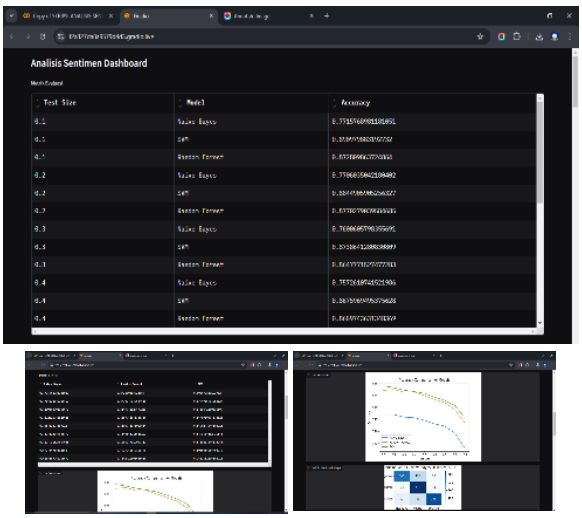


Gambar 7. Perbandingan Algoritma dengan Line Chart

Grafik ini membandingkan kinerja *Naïve Bayes*, SVM, dan *Random Forest* dalam melakukan prediksi ketika jumlah data pelatihan meningkat. Hasil menunjukkan bahwa akurasi ketiga algoritma meningkat seiring bertambahnya data pelatihan. *Random Forest* dan SVM memiliki performa yang lebih stabil dibandingkan *Naïve Bayes*, mengindikasikan keunggulan keduanya dalam analisis sentimen.

4.5. Visualisasi Website

Sebagai langkah akhir, ditampilkan visualisasi dari *website* yang merangkum kesimpulan utama secara interaktif dan mudah diakses, membantu pengguna memahami hasil penelitian dengan lebih jelas. Dapat dilihat tampilan *website* menggunakan gradio yang disajikan pada Gambar 8



Gambar 8. Tampilan *Website*

Tampilan pada Gambar 8 menunjukkan *website* interaktif yang menyajikan hasil perbandingan performa algoritma yang telah dianalisis menggunakan Gradio. *Website* ini dirancang untuk menampilkan kesimpulan utama dalam bentuk visualisasi yang informatif dan mudah dipahami, seperti tabel perbandingan, grafik performa, *confusion matrix*, dan *wordcloud* untuk tiap kategori sentimen. Dengan tampilan yang interaktif, pengguna dapat dengan mudah mengeksplorasi keunggulan dan kelemahan masing-masing algoritma tanpa perlu mengunduh *file* secara manual, sehingga mempermudah proses analisis dan pengambilan keputusan.

5. KESIMPULAN DAN SARAN

SVM dan *Random Forest* menunjukkan akurasi lebih baik dibandingkan *Naïve Bayes* dalam analisis sentimen. *Naïve Bayes* mencapai 77.137% pada data 90:10 tetapi turun hingga 65.649% seiring meningkatnya data uji, menunjukkan keterbatasannya pada data kompleks. *Random Forest* memiliki akurasi awal 86.499% dan menurun ke 70.239%, unggul dalam menangani data besar dengan metode *ensemble*. SVM paling konsisten dengan akurasi 86.906% hingga 75.614%, serta tingkat kesalahan klasifikasi lebih rendah. Disarankan menggunakan SVM untuk akurasi tinggi atau *Random Forest* sebagai alternatif fleksibel, sementara *Naïve Bayes* tetap berguna untuk efisiensi. Peningkatan performa dapat dilakukan dengan *tuning hyperparameter*, pemrosesan teks lanjutan, atau eksplorasi metode *ensemble* dan *deep learning* pada *dataset* besar.

DAFTAR PUSTAKA

[1] D. Pratmanto, R. Rousyati, F. F. Wati, A. E. Widodo, S. Suleman, and R. Wijianto, "App Review Sentiment Analysis Shopee Application in Google Play Store Using Naive Bayes Algorithm," *J. Phys. Conf. Ser.*, vol. 1641, pp. 1–7, 2020, doi: 10.1088/1742-6596/1641/1/012043.

[2] N. C. Ramadani, "Analisis Sentimen Untuk Mengukur Ulasan Pengguna Aplikasi Mobile Legend Menggunakan Algoritma Naive Bayes, SVM, Random Fores, Decision Tree, dan Logistic Regression," *JSI J. Sist. Inf.*, vol. 16, no. 1, pp. 123–138, 2024, [Online]. Available: https://jsi.ejournal.unsri.ac.id/index.php/jsi/index

[3] A. H. Nurdy, A. Rahim, and Arbansyah, "Analisis Sentimen Ulasan Game Stumble Guys Pada Playstore Menggunakan Algoritma Naïve Bayes," *TEKNIKA*, vol. 13, no. 3, pp. 388–395, 2024, doi: 10.34148/teknika.v13i3.993.

[4] A. Salma and W. Silfianti, "Sentiment Analysis of User Review on COVID-19 Information Applications Using Naïve Bayes Classifier, Support Vector Machine, and K-Nearest Neighbors," *Int. Res. J. Adv. Eng. Sci.*, vol. 6, no. 4, pp. 158–162, 2021.

[5] A. Maulana, I. K. Afifah, A. Mubarrak, K. R. Fauzan, A. Dwintara, and B. P. Zen, "Perbandingan Metode Logistic Regression, MultinomialNB, SVM, Dan K-NN Pada Analisis Sentimen Ulasan Aplikasi Gojek di Google Play Store," *J. Tek. Inform.*, vol. 4, no. 6, pp. 1487–1494, 2023, doi: https://doi.org/10.52436/1.jutif.2023.4.6.863.

[6] K. Hasanah, "Comparison of Sentiment Analysis Model for Shopee Comments on Google Play Store," *J. Sisfokom (Sistem Inf. dan Komputer)*, vol. 13, no. 1, pp. 21–30, 2024, doi: 10.32736/sisfokom.v13i1.1916.

[7] H. P. Doloksaribu and Y. T. Samuel, "Komparasi Algoritma Data Mining Untuk Analisis Sentimen Aplikasi Pedulilindungi," *J. Teknol. Inf. J. Keilmuan dan Apl. Bid. Tek. Inform.*, vol. 16, no. 1, pp. 1–11, 2022, doi: 10.47111/jti.v16i1.3747.

[8] R. Wahyudi and G. Kusumawardhana, "Analisis Sentimen pada review Aplikasi Grab di Google Play Store Menggunakan Support Vector Machine," *J. Inform.*, vol. 8, no. 2, pp. 200–207, 2021, doi: 10.35889/progresif.v20i1.1614.

[9] M. R. Rahman, A. F. Diansyah, and Hanafi, "Sentiment Analysis on the Shopee Application on Playstore Using the Random Forest Classification Method," *J. Ilm. Bid. Teknol. Inf. dan Komun.*, vol. 9, no. 1, pp. 20–24, 2023, doi: 10.25139/inform.v9i1.5465.

[10] Musfiroh, A. Tholib, and Z. Arifin, "Analisis Sentimen Terhadap Ulasan Aplikasi Shopee di Google Play Store Menggunakan Metode TF-IDF dan Long Short-Term Memory (LSTM)," *J. Electr. Eng. Comput.*, vol. 6, no. 2, pp. 371–381, 2024, doi: 10.33650/jeecom.v4i2.

[11] A. Prasetyo, T. Ridwan, and A. Voutama, "Analisis Sentimen Terhadap Aplikasi Gbwhatsapp Menggunakan Naive Bayes Classifier Dan Random Forest Classifier," *JSiI (Jurnal Sist. Informasi)*, vol. 11, no. 1, pp. 1–9, 2024, doi: 10.30656/jsii.v11i1.6936.

[12] E. Fitri, Y. Yuliani, S. Rosyida, and W. Gata, "Analisis Sentimen Terhadap Aplikasi Ruangguru Menggunakan Algoritma Naive Bayes, Random Forest Dan Support Vector Machine," *J. Transform.*, vol. 18, no. 1, pp. 71–80, 2020.

[13] R. I. Alhaqq, I. M. K. Putra, and Y. Ruldeviyani, "Analisis Sentimen terhadap Penggunaan Aplikasi MySAPK BKN di Google Play Store," *J. Nas. Tek. Elektro dan Teknol. Inf.*, vol. 11, no. 2, pp. 105–113, 2022, doi: 10.22146/jnteti.v11i2.3528.

[14] Y. A. Prasetyo, E. Utami, and A. Yaqin, "Pengaruh Komposisi Split Data Terhadap Performa Akurasi Analisis Sentimen Algoritma Naïve Bayes dan SVM," *J. Electr. Eng. Comput.*, vol. 6, no. 2, pp. 382–390, 2024, doi: 10.33650/jeecom.v4i2.