PEMODELAN SUPPORT VECTOR MACHINE (SVM) DAN RANDOM FOREST DALAM ANALISIS SENTIMEN ULASAN APLIKASI SOCO BY SOCIOLLA

Zannuba Fatichatul Rizqi[1*], Ria Faulina[2]
[1]*Program Studi Statistika, Fakultas Sains dan Tenologi, Universitas Terbuka, Indonesia*

**Penulis korespondensi: zannubarizqi@gmail.com*

ABSTRAK

Sociolla merupakan salah satu aplikasi kecantikan paling popular di Indonesia. Aplikasi SOCO by Sociolla adalah platform yang memadukan media online dengan e-commerce untuk memberikan ulasan atau umpan balik dari para konsumen terhadap barang yang telah digunakan. Tujuan dari penelitian ini yaitu melakukan analisis sentimen pada ulasan pengguna aplikasi SOCO by Sociolla di Google Play Store yang diklasifikasikan menjadi tiga kategori yaitu positif, netral, dan negatif. Data yang digunakan sebanyak 2.815 ulasan setelah dilakukan praproses tesk. *Support Vector Machine* (SVM) dan *Random Forest* merupakan algoritma yang diterapkan pada penelitian ini. Pembagian *data training* dan *data testing* dengan presentase 80:20 serta 90:10 sebagai objek penelitian. Dari pembagian data tersebut memperlihatkan pembagian data 90:10 mampu memodelkan data lebih baik dibandingkan 80:20 untuk algoritma *Support Vector Machine* (SVM) dan *Random Forest*. Algoritma *Support Vector Machine* mendapatkan nilai akurasi sebesar 75% dan *Random Forest* 73%. Hal ini membuktikan bahwa algoritma *Support Vector Machine* lebih baik digunakan dalam klasifikasi ulasan aplikasi SOCO by Sociolla.

Kata kunci: Analisis sentimen, SOCO by Sociolla, *support vector machine*, *random forest*

1 PENDAHULUAN

Seiring dengan meningkatnya kesadaran masyarakat pada sangat pentingnya untuk menjaga kesehatan dan perawatan pada kulit, dalam beberapa tahun terakhir bisnis kecantikan di Indonesia berkembang secara drastis. Menurut data dari Kementerian Koordinator Bidang Perekonomian Republik Indonesia (2024) ada 1.010 perusahaan kosmetik di Indonesia di pertengahan tahun 2023, meningkat 21,9% dari 913 perusahaan pada tahun 2022. Pada tahun 2023 potensi ukuran pasar mencapai 467.919 produk yang menunjukkan lebih dari sepuluh kali lipat dari 5 tahun sebelumnya. Dengan rata – rata pertumbuhan tahunan sebesar 5,5%, maka nilai pasar diperkirakan akan mencapai USD 473,21 miliar secara global pada tahun 2028. Meningkatnya pengguna media sosial dan platform e-commerce, yang memudahkan konsumen untuk membandingkan produk, mengakses informasi, dan menyelesaikan transaksi secara online, merupakan faktor lain yang berkontribusi terhadap peningkatan ini. Konsumen kini semakin mengandalkan platform digital untuk mendapatkan informasi, dan evolusi teknologi digital telah mengubah cara orang berinteraksi dan berbelanja. Selain membeli barang, konsumen juga dapat membaca ulasan dan berbicara dengan konsumen lain tentang pengalaman mereka dengan produk.

Sociolla merupakan salah satu situs e-commerce kecantikan yang sangat andal dan komprehensif di Indonesia, bekerja sama dengan distributor berlisensi dan setiap produknya terjamin keasliannya, karena telah disertifikasi Badan Pengawasan Obat dan Makanan (BPOM) (Sociolla, 2025). Sociolla mempunyai beberapa fitur andalan, seperti Lilla by Sociolla, Beauty Journal, dan SOCO. SOCO merupakan platform yang memadukan media online dengan e-

commerce untuk memberikan ulasan atau umpan balik dari para konsumen terhadap barang yang telah digunakan.

Aplikasi SOCO by Sociolla tersedia pada Google Play Store dan dapat diakses pada desktop dan perangkat selulur. Pengguna yang telah menggunakan apikasi dapat membagikan pengalaman mereka dengan memberikan ulasan atau komentar dan memberikan rating dalam bentuk bintang pada Google Play Store (Alhaqq dkk., 2022). Selain membantu perusahaan memahami kualitas layanan, ulasan pengguna juga dapat memberikan gambaran langsung tentang pengalaman mereka selama menggunakan aplikasi. Hal ini sangat membantu calon pengguna dalam mengevaluasi kualitas aplikasi sebelum menggunakannya. Analisis sentimen dapat digunakan untuk mengklasifikasikan ulasan – ulasan ini dengan menggunakan tekstual. Karena apabila manual akan menjadi tidak efisien. Oleh karena itu, untuk menentukan apakah suatu sudut pandang memiliki kecenderungan ke arah positif atau negatif diperlukan analisis sentimen (Muttaqin & Kharisudin, 2021).

Analisis sentimen adalah subbidang dari *text mining* yang terfokus pada pemeriksaan sudut pandang yang ditemukan dalam materi tertulis. Analisis sentimen yaitu cara untuk menentukan apakah sebuah teks memiliki makna yang positif, netral, ataupun negatif (Asri, 2024). Metode komputasi digunakan dalam proses ini untuk mengumpulkan ulasan pengguna. Hal ini dapat membantu peneliti dalam menentukan tingkat kepuasan pengguna. Analisis sentiment menggunakan pendekatan komperatif yang menggunakan *Support Vector Machine* (SVM) dan *Random Forest*.

Penelitian analisis sentimen telah banyak algoritma dari *machine learning* yang digunakan, salah satunya penelitian oleh Novianti dkk., (2024) yang melakukan analisis sentimen pada metaverse dengan membandingkan algoritma *Support Vector Machine* dan *Naïve Bayes* diperoleh hasil bahwa SVM memiliki akurasi tertinggi sebesar 90,32%, *precision* 0,90, dan *recall* 0,86, sementara *Naïve Bayes* memperoleh akurasi sebesar 84,23%, *precision* 0,87, dan *recall* 0,84. Kemudian penelitian oleh Pratama dkk., (2023) yang membandingkan *Naïve Bayes* dan *Random Forest* untuk analisis sentimen pada kenaikan harga BBM di Indonesia menghasilkan metode *Naïve Bayes* memiliki akurasi sebesar 79,74%, sedangkan metode *Random Forest* memiliki akurasi 85.15%. Berdasarkan hasil penelitian, algoritma *Support Vector Machine* (SVM) dan *Random Forest* mempunyai kinerja lebih baik dibandingkan algoritma *Naïve Bayes*. Selain itu, Berdasarkan penelitian Firdaus dkk., (2024) menjelaskan tentang penggunaan data latih dan data uji 60:40, 70:30, 80:20, dan 90:10 dalam analisis sentimen dengan membandingkan algoritma *Random Forest* dan *Support Vector Machine*. Pengujian tersebut menghasilkan akurasi terbaik pada pembagian data 90:10 dengan algoritma SVM memperoleh akurasi 96% dan *Random Forest* 94%.

Berdasarkan latar belakang di atas, tujuan pada penelitian ini untuk menganalisis sentimen terhadap ulasan SOCO by Sociolla yang diberikan oleh pengguna dan juga membandingkan tingkat akurasi dari pendekatan *Support Vector Machine* (SVM) dan *Random Forest*. Dengan melakukan penelitian ini akan memberikan gambaran umum tentang perspektif masyarakat pada aplikasi SOCO by Sociolla, termasuk apakah evaluasi secara umum menguntungkan, netral, atau tidak menguntungkan.

2 METODE

2.1 Data dan Sumber Data

Pengambilan data ulasan pengguna aplikasi SOCO by Sociolla didapatkan pada bagian ulasan Google Play Store melalui proses *web scraping* dengan memanfaatkan Google

Colaboratory dengan Bahasa python. Populasi dari sampel data adalah semua data ulasan pengguna aplikasi SOCO by Sociolla. Nama pengguna, ulasan, rating, dan tanggal ulasan yang disampaikan adalah data yang di ambil. Data yang digunakan sebanyak 3.388 dengan periode mulai dari 06 April 2019 hingga 14 Oktober 2025.

Variabel penelitian yang digunakan meliputi variabel *review* dan rating. Klasifikasi teks menggunakan variabel *review* yang berisi ulasan pengguna aplikasi SOCO by Sociolla dan variabel rating berisi penilaian berupa bintang (1 hingga 5) pada pengguna aplikasi SOCO by Sociolla, yang akan digunakan juga sebagai analisis deskriptif. Ulasan rating dikategorikan dalam kelas sentimen sebagai berikut:

Table 1. Kategori Sentimen Berdasarkan Rating

Rating	Kategori Sentimen
1 dan 2	Negatif
3	Netral
4 dan 5	Positif

Sumber: (Zuhdi & Prasetiyo., 2025)

2.2 Analisis Sentimen

Menurut Mola dkk., (2024) analisis sentimen sering diistilahkan *opinion mining* yaitu cabang penelitian untuk meneliti pemikiran, perasaan, penilaian, dan emosi orang tentang berbagai topik yang meliputi organisasi, pariwisata, isu, barang jasa, dan atributnya. Fitri dkk., (2019) menyatakan bahwa analisis sentimen merupakan subbidang dari *text mining* yang terfokus pada pemeriksaan sudut pandang yang ditemukan dalam materi tertulis. Metode untuk menetapkan apakah sebuah teks memiliki makna yang negatif, netral, atau positif dikenal sebagai analisis sentimen (Asri, 2024). Saat ini, analisis sentimen sangat penting karena memungkinkan berbagai pihak yang berkepentingan untuk memahami opini yang diungkapkan dalam data tekstual. Maka, untuk memungkinkan pengambilan keputusan yang lebih strategis dan efektif, hasil analisis sentimen sering kali digunakan untuk membantu bisnis, pemerintah, dan organisasi lain dalam pengumpulan informasi mengenai tren pasar, kepuasan pelanggan, citra merek, dan respon kebijakan.

2.3 *Support Vector Machine* (SVM)

Prinsip *Structural Risk Minimazation* (SRM) dioperasikan pada *Support Vector Machine* (SVM) bertujuan untuk menentukan *hyperplane* optimal yang dapat membedakan kelas – kelas dalam *input space*.

Support Vector Machine (SVM) biasanya digunakan untuk tugas regresi dan klasifikasi dengan konsep memaksimalkan jarak antar kelas. Untuk membagi kelas data dengan jarak terbesar dari vektor pendukung, SVM menentukan *hyperplane* yang optimal (Amalia & Yustanti, 2021). Dalam SVM, istilah *hyperplane* merupakan representasi geometris dari keputusan yang diambil oleh SVM untuk memisahkan data dengan label kelas yang berbeda. *Hyperplane* yaitu garis pemisah terbaik antara dua kelas data dengan mengukur *margin hyperplane* serta menemukan titik maksimalnya. Menurut Supian dkk., (2024) pencarian *hyperplane* terbaik bertujuan untuk memisahkan 2 kelas data, yakni kelas positif (+1) dan kelas negatif (-1). Jarak antara *hyperplane* dan pola data terdekat dari masing – masing kelas disebut sebagai *margin*.

Beberapa fungsi kernel yang biasanya digunakan pada umumnya sebagai berikut:

a. Kernel Linear

$$K(x, y) = x.y$$

b. Kernel *Polynomial* dengan Variabel Bebas q

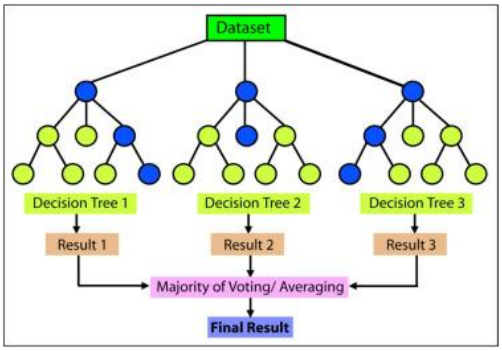
$$K(x_i, y_j) = ((x_i, y_j) + c)^{d}$$

c. Kernel Gaussian atau RBF

$$K(x_i, y_j) = \exp\left(\frac{|x_i - y_j|^2}{2\sigma^2}\right)$$

2.4 *Random Forest*

Random Forest merupakan teknik pembelajatan *ensembel* untuk aplikasi regresi serta klasifikasi. Pendekatan ini dipilih karena keunggulannya yang mecakup kapasitas untuk mengalisis data set yang besar dengan berbagai karakteristik, menghindari *overfitting*, dan mempertahankan stabilitas kinerja yang tinggi dan kuat (Siregar dkk., 2023). Selain itu, *Random Forest* dapat mengevalusi signifikansi variabel dalam proses klasifikasi dan efisien dalam mengelola data yang hilang. Jalal dkk., (2022) menyatakan bahwa *Random Forest* membuat keputusan dengan menggunakan banyak pohon keputusan. Alih – alih bergantung pada satu pohon keputusan *Random Forest* memprediksi hasil berdasarkan suatu mayoritas dari perkiraan yang dibuat oleh setiap pohon keputusan.



Gambar 1. Struktur pohon pada model *random forest* (Chowdhury, 2024)

Menurut Aljedaani dkk., (2022) dalam hal matematika, *Random Forest* dapat didefinisikan sebagai.

$$rf = mode\,\{T_1, T_2, T_3, \dots, T_n\}$$

Atau

$$rf = mode\left\{\sum_{i=1}^{n} T_1\right\}$$

Keterangan:

T_1, T_2, T_3 : pohon – pohon dalam *Random Forest*

n : jumlah pohon

Setelah kumpulan pohon dalam *Random Forest* terbentuk, proses prediksi dilakukan dengan menentukan hasil berdasarkan mayoritas suara (*voting*) untuk klasifikasi. Tingkat keyakinan model terhadap hasil prediksinya diukur menggunakan *Margin Function*. Fungsi margin menghitung sejauh mana jumlah rata – rata suara pada (X, Y) bagi kelas yang tepat melebihi jumlah rata – rata suara bagi kelas lainnya. Disini X adalah *vector predictor* dan Y adalah klasifikasi. Fungsi margin diberikan sebagai berikut:

$$mg(X,Y) = av_k I(h_k(X) = Y) - max_{j \neq Y} av_k I(h_k(X) = j)$$

di sini I(.) adalah fungsi indikator.

Margin berbanding lurus dengan keyakinan dalam klasifikasi. Kekuatan dari *Random Forest* diberikan dalam bentuk nilai yang diharapkan dari fungsi margin sebagai berikut.

$$S = E_{X,Y}(mg(X,Y)$$

Akurasi yang dapat dicapai dan kemungkinan masalah *overfitting* dapat dihindari meningkat seiring dengan bertambahnya jumlah pohon dalam *decision forest* (Habibi dkk., 2023). Kesalahan generalisasi diperlukan dalam mencegah *overfitting* serta menghasilkan prediksi yang akurat dengan data baru yang belum digunakan pada pelatihan. Kesalahan (PE*) dapat dinyatakan sebagai berikut.

$$PE^* = P_{x,y}\big(mg(X,Y)\big) < 0$$

di mana $mg(X,Y)$ adalah fungsi margin.

Kesalahan generalisasi dalam metode *ensamble* dipengaruhi oleh rata – rata korelasi antara pengklasifikasi dasar serta kekuatan rata – rata model. Jika ρ mewakili nilai rata – rata korelasi, maka batas atas maksimum dari kesalahan generalisasi dapat dinyatakan sebagai berikut.

$$PE^* = \rho(1 - s^2)/s^2$$

2.5 Langkah Analisis

Analisis data yang digunakan dalam penelitian ini yaitu sebagai berikut:

a) *Scraping data*, yaitu tahap pengumpulan data dengan Google Colaboratory untuk mendapatkan ulasan pengguna aplikasi SOCO by Sociolla.
b) Praproses data, yaitu melakukan tahapan *cleaning data* untuk membersihkan data agar lebih mudah dipahami. Selanjutnya melakukan *case folding*, *tokenizing*, *normalisasi*, *stopword*, dan *stemming*.
c) Pelabelan data, membagi rating jadi 3 kategori yaitu 1 hingga 2 negatif, 3 netral dan 4 hingga 5 positif (Zuhdi & Prasetiyo, 2025).
d) Split data menggunakan 80:20 dan 90:10.
e) *Feature extraction*, tahap mengubah data menjadi fitur numerik agar dapat diproses oleh model menggunakan fitur TF-IDF.
f) Klasifikasi, melakukan analisis dengan mengguakan *Support Vector Machine* (SVM) dan *Random Forest*.
g) Evalusi model, untuk membandingkan metode berdasarkan nilai tingkat akurasi. Rumus akurasi sebagai berikut.

$$Accuracy = \frac{A + E + I}{A + B + C + D + E + F + G + H + I}$$

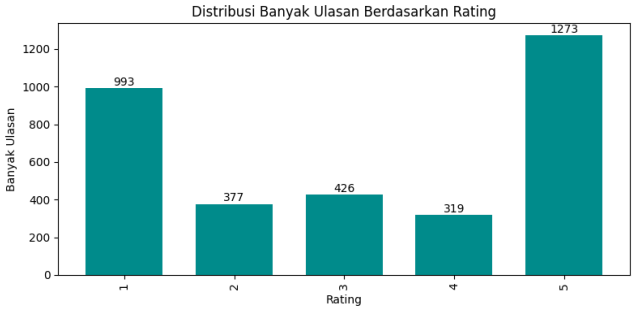
h) Visualisasi *word cloud*, menganalisis kata atau topik yang banyak dibicarakan dalam ulasan pengguna aplikasi SOCO by Sociolla di Google Play Store.

3 HASIL DAN PEMBAHASAN

3.1 Analisis Diskriptif

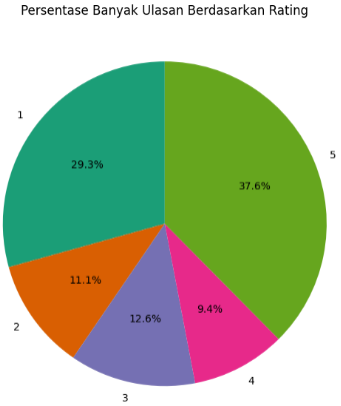
Sebelum melakukan praproses data teks dilakukan, data yang telah dikumpulkan dieksplorasi. Sangat penting untuk menggunakan data yang dieksplorasi dalam penelitian ini untuk memperoleh pemahaman yang lebih baik tentang karakteristik dan karakteristik umum yang diberikan oleh pengguna aplikasi SOCO by Sociolla yang berada di Google Play Store. Dalam ulasan tersebut,

penilaian dibagi menjadi kategori 1 – 5, seperti yang ditunjukkan pada diagram batang sebagai berikut.



Gambar 2. Distribusi ulasan berdasarkan rating

Gambar 2 menunjukkan frekuensi sebaran penilaian/ rating pengguna aplikasi SOCO by Sociolla pada sampel data yang telah diambil. Dari data tersebut diperoleh informasi bahwa rating yang diberikan bervariatif, seperti ulasan paling tinggi pada rating 5 sebesar 1.273 dan paling rendah pada rating 4 sebesar 319. Ulasan tersebut memiliki rentang dari 1 hingga 5, yang memiliki arti secara berurutan yaitu "sangat buruk", "buruk", "rata – rata", "baik", dan "sangat baik". Untuk mengetahui perbandingan frekuensi antar kategori penilian disajian *pie chart* sebagai berikut.



Gambar 3. Proporsi ulasan berdasarkan rating

Berdasarkan Gambar 3 dapat diketahui bahwa sebagian pengguna aplikasi SOCO by Sociolla memberikan penilaian yang positif. Dapat dilihat untuk bintang 5 sebesar 37,6% memberikan rating "sangat baik" dan 9,4% memberikan rating "baik" pada bintang 4, sehingga secara keseluruhan 47% pengguna mempunyai pemahaman positif tehadap aplikasi tersebut. Sedangkan, dapat dilihat pada bintang 1 – 3 mempunyai penilaian "sangat buruk" sebesar 29,3%, "buruk" sebesar 11,1%, dan "rata – rata" sebesar 12,6%, sehingga secara keseluruan 53%. Hal ini menunjukkan bahwa walaupun ulasan positif mendominasi, namun masih banyak pemahaman pengguna yang cukup besar memberikan rating sangat rendah. Sehingga, membuat analisis sentimen diperlukan untuk mengevaluasi ulasan para pengguna aplikasi SOCO by Sociolla. Agar dapat dilakukan evaluasi untuk memahami faktor penyebab ketidakpuasan pengguna.

3.2 Praproses Data

Pada praproses data dilakukan beberapa tahapan mengolah data mentah yang akan digunakan dalam proses klafisikasi. Berikut tahapan yang dilakukan pada penelitian ini (Mola dkk., 2024):

a) *Cleaning Data* dan *Case Folding*
Melakukan pembersihan data, di mana elemen seperti tanda baca, angka, simbol, URL, karakter khusus, dan spasi yang tidak perlu dihilangkan dari data teks. Kemudian, *case folding* merupakan tahap mengubah semua huruf dalam teks menjadi huruf kecil (*lowercase*) untuk menghindari ketidaksuaian antara pengolahan huruf besar dan huruf kecil. Tahap ini membuat data teks lebih sederhana, sehingga membuat lebih seragam serta memudahkan analisis pada tahap berikutnya.

b) *Tokenizing*
Tokenizing merupakan tahap untuk memecahkan teks menjadi bagian terkecil, biasanya kata atau frasa. Dengan melakukan ini, maka setiap kata dapat dianalisis secara terpisah.

c) Normalisasi
Normalisasi merupakan proses penyesuaian teks agar lebih standar dengan memperbaiki ejaan, menggantikan singkatan, dan menyelaraskan kata – kata gaul atau informal ke bentuk standar. Pada tahapan ini mencakup perbaikan ejaan dan penggantian kata – kata baku menjadi bentuk yang sesuai dengan norma Bahasa.

d) *Stopword*
Stopword merupakan proses penyaringan kata – kata seperti “dan”, “di”, “ke”, dan “yang” merupakan contoh kata – kata yang dianggap tidak penting atau tidak relevan untuk analisis karena tidak memiliki makna spesifik, sehingga dapat diabaikan untuk meningkatan efisiensi pemrosesan data.

e) *Stemming*
Stemming merupakan proses mengubah kata – kata berdasarkan konteks gramatikal mereka. Proses ini membantu menyatukan variansi bentuk kata, sehingga analisis dapat lebih konsisten. Terutama untuk mengukur kemunculan kata – kata dengan makna yang sama.

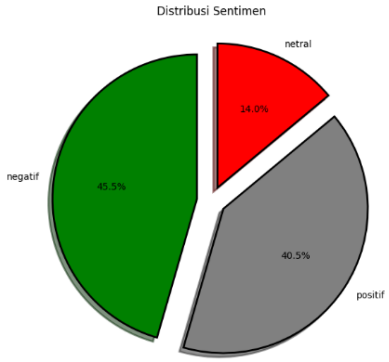
Tabel 2. Praproses data

Teks awal	Skin care original,tidak ada biaya admin..pengiriman cepat .Pokoknya top??
Cleaning Data* dan *Case Folding	Skin care original tidak ada biaya admin pengiriman cepat pokoknya top
Tokenizing	[‘skin’, ‘care’, ‘original’, ‘tidak’, ‘ada’, ‘biaya’, ‘admin’, ‘pengiriman’, ‘cepat’, ‘pokoknya’, ‘top’]
Normalisasi	[‘skin’, ‘care’, ‘original’, ‘tidak’, ‘ada’, ‘biaya’, ‘admin’, ‘pengiriman’, ‘cepat’, ‘pokoknya’, ‘top’]
Stopword	[‘skin’, ‘care’, ‘original’, ‘biaya’, ‘admin’, ‘pengiriman’, ‘cepat’, ‘pokoknya’, ‘top’]
Stemming	Skin care original biaya admin kirim cepat pokok top

Setelah dilakukan *cleaning data* dan praproses data teks, jumlah data menjadi 2.815 data, yang sebelumnya berjumlah 3.388 data. Langkah selanjutnya yaitu melakukan palabelan berdasarkan rating diperoleh perbandingan jumlah data sebagai berikut.

Tabel 3. Hasil pelabelan

Kelas	Jumlah
Negatif	1282
Positif	1140
Netral	393



Gambar 4. Distribusi kelas sentimen

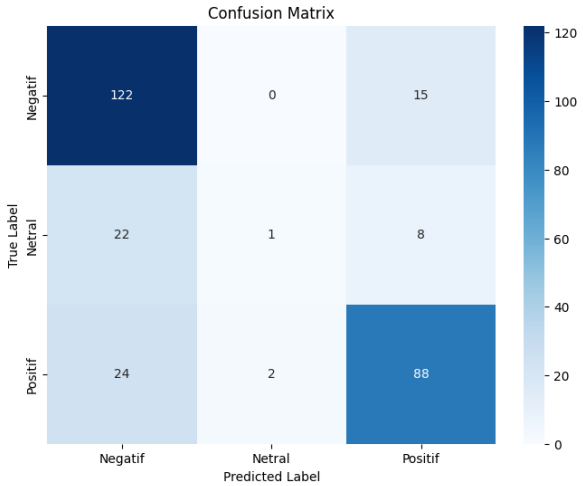
Berdasarkan Tabel 3 serta Gambar 4 dapat memberitahukan bahwa ulasan pengguna aplikasi SOCO by Sociolla cenderung mengandung sentimen negatif yaitu 1.282 atau 45.5%. Kemudian, pada sentimen positif sebesar 1.140 atau 40,5% dan pada sentimen netral hanya 393 atau 14%. Hal ini menandakan bahwa mayoritas pengguna menyampaikan ulasan yang cenderung mengungkapkan ketidakpuasan terhadap aplikasi SOCO by Sociolla. Selanjutnya teks dikonversi menjadi bentuk numerik menggunakan TF-IDF untuk menentukan bobot masing – masing kata berdasarkan tingkat kemunculannya dalam dokumen dan seluruh data.

3.3 *Support Vector Machine* (SVM)

Klasifikasi pertama yaitu *Support Vector Machine* (SVM), data yang digunakan merupakan data yang sudah dilatih serta diuji. 80:20 dan 90:10 merupakan pembagian data yang digunakan dalam klasifikasi. Berikut merupakan hasil pemodelan klasifikasi menggunakan *Support Vector Machine*.

Tabel 4. Hasil Evaluasi Metode SVM

Split Data	Akurasi
80:20	73%
90:10	75%



Gambar 5. *Confusion matrix* metode SVM

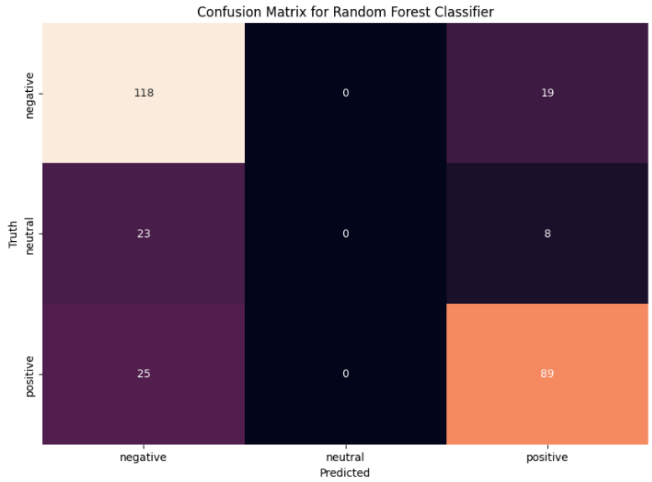
Pada Tabel 4 dapat diketahui bahwa pemodelan *Support Vector Machine* (SVM) dengan split data 90:10 menghasilkan performa terbaik dengan akurasi 75%. Pada Gambar 5 menunjukkan hasil evaluasi performa dari model SVM. Hasil *confusion matrix* masih menunjukkan bahwa hasil klasifikasi model terbaik masih terdapat kesalahan prediksi. Terdapat sebanyak 137 ulasan berlebel negatif dengan 122 yang terprediksi benar, tidak ada yang terprediksi kelas netral, dan 15 terprediksi kelas positif. Kemudian terdapat 31 ulasan berlabel netral dengan 1 terprediksi benar, 22 terprediksi kelas negatif, dan 8 terprediksi kelas positif. Sedangkan, terdapat sebanyak 114 ulasan berlabel positif dengan 88 ulasan terprediksi benar, 24 terprediksi kelas negatif, dan 2 terprediksi kelas netral.

3.4 *Random Forest*

Klasifikasi kedua yaitu *Random Forest*, menggunakan data yang sudah dilatih dan diuji dalam proses sebelumnya. 80:20 dan 90:10 merupakan data latih dan data uji yang digunakan dalam klasifikasi. Berikut hasil pengujian dari *Random Forest*.

Tabel 5. Hasil Evaluasi Metode *Random Forest*

Split Data	Akurasi
80:20	72%
90:10	73%



Gambar 6. *Confusion matrix* metode *Random Forest*

Pada Tabel 5 dapat diketahui bahwa pemodelan *Random Forest* dengan pembagian data 90:10 menghasilkan performa terbaik dengan akurasi 73%. Pada Gambar 6 menunjukkan hasil evaluasi performa dari model *Random Forest*. Hasil *confusion matrix* masih menunjukkan bahwa hasil klasifikasi model terbaik masih terdapat kesalahan prediksi. Terdapat sebanyak 137 ulasan berlebel negatif dengan 118 yang terprediksi benar, tidak ada yang terprediksi kelas netral, dan 19 terprediksi kelas positif. Kemudian terdapat 31 ulasan berlabel netral tidak ada yang terprediksi benar, 23 terprediksi kelas negatif, dan 8 terprediksi kelas positif. Sedangkan, terdapat sebanyak 114 ulasan berlabel positif dengan 89 ulasan terprediksi benar, 25 terprediksi kelas negatif, dan tidak ada yang terprediksi kelas netral.

3.5 Perbandingan Metode

Setelah mendapatkan hasil dari kedua metode, berikut merupakan perbandingan hasil dari metode *Support Vector Machine* (SVM) dan *Random Forest*. Berikut tabel perbandingan kedua metode tersebut.

Tabel 6. Perbandingan Hasil Evaluasi

Metode	Split Data	Akurasi
***Support Vector Machine* (SVM)**	80:20	73%
	90:10	**75%**
Random Forest	80:20	72%
	90:10	73%

Menurut Tabel 6 menunjukkan metode *Support Vector Machine* (SVM) memiliki tingkat akurasi terbaik yaitu nilai akurasi sebesar 75% menggunakan pembagian data 90% untuk data pelatihan dan 10% untuk data pengujian.

3.6 Visualisasi Data

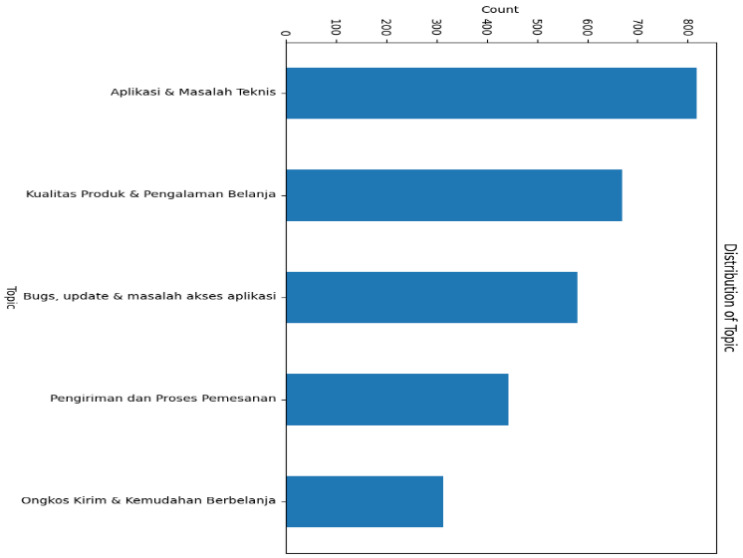
Word cloud untuk menampilkan visualisasi ulasan pengguna aplikasi SOCO by Sociolla pada Google Play Store. *Word cloud* merupakan metode visualisasi yang menyajikan sekumpulan kata yang selalu muncul dari teks atau dokumen. Kata tersebut ditampilkan dalam ukuran yang bervariasi, semakin sering muncul sebuah kata, maka ukurannya akan semakin besar (Fahrudin dkk., 2022). Kata – kata dalam *word cloud* dikelompokkan berdasarkan ketegori sentimen,

sehingga dapat terlihat istilah yang selalu muncul di setiap katerogi sentimen. Berikut hasil visualisasi *word cloud* dalam analisis sentimen ulasan aplikasi SOCO by Sociolla.

Gambar 7. *Word cloud* sentimen positif pada ulasan sociolla

Gambar 8. *Word cloud* sentimen negatif pada ulasan sociolla

Dalam Gambar 7, lima kata yang sering digunakan untuk menunjukkan sentimen positif adalah aplikasi, sociolla, bagus, belanja, dan produk. Pada Gambar 8, lima kata yang paling sering digunakan untuk menunjukkan sentimen negatif adalah aplikasi, barang, buka, update, dan belanja.



Gambar 9. Distribusi ulasan berdasarkan aspek pada ulasan

Berdasarkan Gambar 9 diketahui bahwa sebagian besar ulasan pengguna fokus pada aplikasi dan masalah teknis yang menunjukkan pada keluhan tentang error dalam aplikasi dan proses loading lambat. Aspek kualitas produk dan pengalam belanja menunjukkan kepuasan pelanggan terhadap produk dan pengalaman berbelanja di Sociolla. Bugs, update dan masalah akses aplikasi menunjukkan keluhan tentang aplikasi yang sering error dan kesulitan login. Sementara itu, pengiriman dan proses pemesanan serta ongkos kirim dan kemudahan berbelanja menunjukkan keluhan tentang proses pengiriman barang dan masalah transaksi pembayaran.

4 KESIMPULAN

Berdasarkan 3.388 data ulasan pengguna aplikasi SOCO by Sociolla setelah melakukan tahapan *cleaning data* menjadi 2.815 data. Hasil pelabelan analisis sentimen menunjukkan bahwa lebih banyak ulasan sentimen negatif 45,5%, ulasan sentimen positif 40,5%, dan ulasan sentimen netral 14%. Hal ini menunjukkan bahwa mayoritas pengguna menyampaikan ketidakpuasan terhadap aplikasi. Setelah melakukan tahapan analisis menggunakan *Support Vector Machine* (SVM) dan *Random Forest* yang dilakukan pada pembagian data 80:20 dan 90:10 menghasilkan *Support Vector Machine* (SVM) pada pembagian data 90:10 memiliki performa terbaik dengan nilai akurasi 75%. Keluhan utama pengguna adalah pada masalah teknis aplikasi yang error, bug, lambat, dan kendala login. Hal ini menunjukkan stabilitas dan performa aplikasi, serta perbaikan proses pengiriman dan transaksi pembayaran, agar dapat meningkatkan kepuasan pengguna.

DAFTAR PUSTAKA

Aljedaani, W., Rustam, F., Mkaouer, M. W., Ghallab, A., Rupapara, V., Washington, P. B., Lee, E., & Ashraf, I. (2022). Sentiment analysis on Twitter data integrating TextBlob and deep learning models: The case of US airline industry. *Knowledge-Based Systems*, *255*. https://doi.org/10.1016/j.knosys.2022.109780

Amalia, D. H., & Yustanti, W. (2021). Klasifikasi Buku Menggunakan Metode *Support Vector Machine*pada Digital Library. *Journal of Informatics and Computer Science (JINACS)*, *3*(01). https://doi.org/10.26740/jinacs.v3n01.p55-61

Ary Prandika Siregar, Dwi Priyadi Purba, Jojor Putri Pasaribu, & Khairul Reza Bakara. (2023). Implementasi Algoritma *Random Forest* Dalam Klasifikasi Diagnosis Penyakit Stroke. *Jurnal Penelitian Rumpun Ilmu Teknik*, *2*(4). https://doi.org/10.55606/juprit.v2i4.3039

Asri, Y. K. D. (2024). *Machine Learning & Deep Learning: Analisis Sentimen Menggunakan Ulasan Pengguna Aplikasi*. Uwais Inspirasi Indonesia.

Chowdhury, M. S. (2024). Comparison of accuracy and reliability of random forest, support vector machine, artificial neural network and maximum likelihood method in land use/cover classification of urban setting. *Environmental Challenges*, *14*. https://doi.org/10.1016/j.envc.2023.100800

Fahrudin, T. M., Sari, A. R. F., Lisanthoni, A., & Lestari, A. A. D. (2022). Analisis Speech-To-Text Pada Video Mengandung Kata Kasar Dan Ujaran Kebencian Dalam Ceramah Agama Islam Menggunakan Interpretasi Audiens Dan Visualisasi Word Cloud. *SKANIKA*, *5*(2). https://doi.org/10.36080/skanika.v5i2.2942

Firdaus, R., Al Hariri, R., & Amran, H. F. (2024). Sentimen Analisis Masyarakat Tentang Penetapan Hari Raya Idul Adha Tahun 2023 Pada Video Youtube Menggunakan Algoritma *Random Forest* dan Support Vector Machine. *JURNAL FASILKOM*, *14*(1), 278–285. https://doi.org/10.37859/jf.v14i1.7012

Fitri, E. (2020). Analisis Sentimen Terhadap Aplikasi Ruangguru Menggunakan Algoritma Naive Bayes, *Random Forest* Dan Support Vector Machine. *Jurnal Transformatika*, *18*(1). https://doi.org/10.26623/transformatika.v18i1.2317

Fitri, V. A., Andreswari, R., & Hasibuan, M. A. (2019). Sentiment analysis of social media Twitter with case of Anti-LGBT campaign in Indonesia using Naïve Bayes, decision tree, and *Random Forest* algorithm. *Procedia Computer Science*, *161*. https://doi.org/10.1016/j.procs.2019.11.181

Habibi, R., Dwiyanti, Z. A., Prianto, C., & Fahira. (2023). *Prediksi Diagnosa Kehamilan Menggunakan Algoritma C4.5 dan Random Forest*. Penerbit Buku Pedia. https://books.google.co.id/books?id=-GvWEAAAQBAJ

Jalal, N., Mehmood, A., Choi, G. S., & Ashraf, I. (2022). A novel improved *Random Forest* for text classification using feature ranking and optimal number of trees. *Journal of King Saud University - Computer and Information Sciences*, *34*(6). https://doi.org/10.1016/j.jksuci.2022.03.012

Kementerian Koordinator Bidang Perekonomian Republik Indonesia. (2024, February 3). *Hasilkan Produk Berdaya Saing Global, Industri Kosmetik Nasional Mampu Tembus Pasar Ekspor dan Turut Mendukung Penguatan Blue Economy*. Kementerian Koordinator Bidang Perekonomian Republik Indonesia.

Mola, S. A. S., Rumlaklak, N. D., Polly, D. P. N., & Widiastuti, T. (2024). *Analisis Sentimen dengan Metode Random Forest*. Kaizen Media Publishing. https://books.google.co.id/books?id=vuIWEQAAQBAJ

Muttaqin, M. N., & Kharisudin, I. (2021). Analisis Sentimen Pada Ulasan Aplikasi Gojek Menggunakan Metode *Support Vector Machine*dan K Nearest Neighbor. *UNNES Journal of Mathematics*, *10*(2).

Naufal Zuhdi, H., & Prasetiyo, B. (2025). Analisis Sentimen pada Ulasan Aplikasi iPusnas di Google Play Store Menggunakan Naive Bayes Classifier. *Indonesian Journal of Informatic Research and Software Engineering (IJIRSE)*, *5*(1), 12–19. https://doi.org/10.57152/ijirse.v5i1.1846

Novianti, D. N., Shiddieq, D. F., Roji, F. F., & Susilawati, W. (2024). Komparasi Algoritma *Support Vector Machine*dan Naïve Bayes untuk Analisis Sentimen pada Metaverse. *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, *4*(1), 231–239. https://doi.org/10.57152/malcom.v4i1.1061

Pratama, A. R. I., Latipah, S. A., & Sari, B. N. (2022). Optimasi Klasifikasi Curah Hujan Menggunakan *Support Vector Machine*(Svm) dan Recursive Feature Elimination (Rfe). *JIPI (Jurnal Ilmiah Penelitian Dan Pembelajaran Informatika)*, *7*(2). https://doi.org/10.29100/jipi.v7i2.2675

Pratama, M. L., Via, Y. V., & Mandyartha, E. P. (2023). Analisis Performansi Naive Bayes Dan *Random Forest* Terhadap Sentimen Kenaikan Harga BBM di Indonesia. *Scan: Jurnal Teknologi Informasi Dan Komunikasi*, *18*(1). https://doi.org/10.33005/scan.v18i1.3837

Raksaka Indra Alhaqq, I Made Kurniawan Putra, & Yova Ruldeviyani. (2022). Analisis Sentimen terhadap Penggunaan Aplikasi MySAPK BKN di Google Play Store. *Jurnal Nasional Teknik Elektro Dan Teknologi Informasi*, *11*(2). https://doi.org/10.22146/jnteti.v11i2.3528

Sociolla. (2025). *About Sociolla*. Sociolla.

Supian, A., Tri Revaldo, B., Marhadi, N., Efrizoni, L., & Rahmaddeni, R. (2024). Perbandingan Kinerja Naïve Bayes Dan Svm Pada Analisis Sentimen Twitter Ibukota Nusantara. *JURNAL ILMIAH INFORMATIKA*, *12*(01), 15–21. https://doi.org/10.33884/jif.v12i01.8721