

ANALISIS SENTIMEN KOMENTAR APLIKASI STOCKBIT DI GOOGLE PLAY STORE MENGGUNAKAN SUPPORT VECTOR MACHINE

Pandapotan KS, Fahmi A, Fergiano DR, Fadly D, Deni EP, Chaerur Rozikin

Informatika, Universitas Singaperbangsa Karawang

Jl.HS.Ronggo Waluyo, Puseurjaya, Telukjambe Timur, Karawang, Jawa Barat 41361, Telp.(0267) 641177

pandachristian27@gmail.com

ABSTRAK

Penelitian ini bertujuan menganalisis sentimen ulasan aplikasi investasi Stockbit di Google Play Store menggunakan algoritma Support Vector Machine (SVM). Data awal berjumlah 1000 ulasan, yang setelah tahap *dataset preparation* (penghapusan data duplikat, user ganda, dan data kosong) menyisakan 898 ulasan. Ulasan tersebut diberi label biner, yaitu positif (1) untuk rating 4–5 dan negatif (0) untuk rating 1–2, sementara rating 3 diabaikan. Proses penelitian mencakup *scraping*, *preprocessing* (cleaning, case folding, normalisasi, tokenizing, stopword removal, dan stemming), pembobotan TF-IDF, serta pembagian data ke dalam lima skenario (10:90, 20:80, 30:70, 40:60, dan 50:50). Hasil menunjukkan bahwa skenario 10% testing dan 90% training memberikan performa terbaik dengan F1-Score 0.846. Pengujian model dilakukan dengan kernel Linear dan RBF, di mana keduanya menghasilkan nilai F1-Score yang sama, namun kernel Linear lebih stabil dan disarankan karena kesederhanaannya. Klasifikasi akhir menunjukkan 607 sentimen positif dan 238 sentimen negatif, yang menggambarkan mayoritas pengguna memberikan ulasan positif terhadap aplikasi Stockbit. Penelitian ini dapat menjadi referensi untuk pengembangan sistem analisis sentimen pada aplikasi investasi online.

Kata kunci : Analisis Sentimen, Support Vector Machine, Stockbit, Google Play Store, TF-IDF, Kernel SVM

1. PENDAHULUAN

Dalam era digital, aplikasi investasi online telah menjadi sarana utama bagi masyarakat untuk berinvestasi, seiring dengan pesatnya perkembangan aplikasi digital lain seperti e-commerce dan ojek online yang kini menjadi bagian penting dalam kehidupan sehari-hari. Berdasarkan data dari Otoritas Jasa Keuangan (OJK), minat masyarakat terhadap investasi meningkat secara signifikan, yang terlihat dari semakin banyaknya aplikasi investasi online yang bermunculan. Untuk mengetahui tingkat keamanan dan kualitas suatu aplikasi investasi, informasi yang diperoleh dari ulasan pengguna sangatlah penting. Ulasan tersebut mencerminkan pengalaman langsung pengguna dan dapat memberikan wawasan berharga bagi pengembang dalam mengevaluasi kepuasan, fitur, serta kualitas layanan yang disediakan.

Analisis sentimen, yang merupakan salah satu bidang penting dalam pemrosesan bahasa alami (*Natural Language Processing*), digunakan untuk mengenali, mengidentifikasi, dan mengekstraksi opini, emosi, serta sentimen yang terkandung dalam teks. Berbagai penelitian terdahulu telah menerapkan metode ini untuk menganalisis sentimen pada ulasan aplikasi digital. Misalnya, penelitian yang menggunakan algoritma *Support Vector Machine* (SVM) untuk mengklasifikasi ulasan aplikasi e-commerce seperti Grab di Google Play Store [12]. Penelitian lainnya menerapkan algoritma *Random Forest* untuk menganalisis sentimen pada aplikasi Shopee, Tokopedia, Lazada, dan Blibli [10]. Selain itu, metode *Naïve Bayes Classifier* juga digunakan dalam penelitian terhadap ulasan aplikasi Shopee [3], Ojol The Game [9], dan SIGNAL [4]. Namun, beberapa penelitian menunjukkan bahwa metode *Naïve Bayes*

memiliki akurasi yang relatif lebih rendah, seperti pada penelitian terhadap aplikasi investasi Bibit yang hanya mencapai akurasi sebesar 75% [6].

Permasalahan utama yang muncul adalah bahwa penelitian mengenai analisis sentimen pada ulasan aplikasi investasi masih sangat terbatas dibandingkan dengan aplikasi e-commerce atau transportasi online. Belum banyak penelitian yang menerapkan algoritma *Support Vector Machine* (SVM) dengan kombinasi kernel *linear* dan *RBF* serta pengoptimalan parameter *C* dan *gamma* untuk meningkatkan akurasi klasifikasi sentimen. Selain itu, isu literasi keuangan dan keamanan data juga menjadi faktor penting dalam konteks penggunaan aplikasi investasi online. Kurangnya literasi keuangan dapat menyebabkan pengambilan keputusan investasi yang tidak tepat, sementara keamanan data pribadi menjadi perhatian utama dalam membangun kepercayaan pengguna terhadap aplikasi investasi.

Berdasarkan permasalahan tersebut, penelitian ini bertujuan untuk menganalisis sentimen pada ulasan pengguna aplikasi investasi Stockbit di Google Play Store dengan menggunakan algoritma *Support Vector Machine* (SVM). Penelitian ini dilakukan untuk mengidentifikasi opini pengguna terhadap kualitas layanan dan fitur aplikasi, sekaligus mencari model SVM terbaik berdasarkan variasi kernel dan parameter yang digunakan. Secara khusus, penelitian ini berfokus pada pengujian kernel *linear* dan *RBF* dengan tuning parameter *C* dan *gamma* guna memperoleh akurasi, *precision*, *recall*, dan *F1-score* tertinggi.

Untuk menyelesaikan permasalahan tersebut, penelitian ini dirancang melalui beberapa tahapan. Pertama, pengumpulan data dilakukan dengan teknik *web scraping* menggunakan bahasa pemrograman

Python untuk memperoleh 1.000 ulasan pengguna aplikasi Stockbit. Kedua, dilakukan *dataset preparation* dengan menghapus data duplikat dan tidak relevan, serta pelabelan semi-otomatis berdasarkan rating pengguna, di mana rating 1–2 dikategorikan sebagai sentimen negatif dan rating 4–5 sebagai sentimen positif, sedangkan rating 3 diabaikan. Selanjutnya dilakukan tahap *preprocessing* berupa *cleaning*, *case folding*, *tokenizing*, *normalization*, *stopword removal*, dan *stemming* untuk menghasilkan teks yang bersih. Data kemudian diberi bobot menggunakan metode *TF-IDF* dan dibagi menjadi data latih dan uji dengan beberapa perbandingan rasio. Model SVM kemudian diimplementasikan menggunakan kernel *linear* dan *RBF* dengan variasi parameter *C* dan *gamma*. Evaluasi dilakukan menggunakan *confusion matrix* untuk menentukan kombinasi parameter dan kernel terbaik yang menghasilkan akurasi tertinggi.

Melalui serangkaian tahapan tersebut, penelitian ini diharapkan dapat memberikan kontribusi dalam bidang analisis sentimen khususnya pada sektor investasi online, serta menjadi acuan bagi pengembang aplikasi dalam memahami persepsi pengguna. Selain itu, hasil penelitian ini diharapkan mampu memberikan wawasan tentang pengaruh parameter dan kernel terhadap performa SVM dalam klasifikasi teks, sehingga dapat dimanfaatkan dalam pengembangan sistem analisis sentimen yang lebih efektif dan akurat di masa mendatang.

2. TINJAUAN PUSTAKA

2.1. Penelitian Terdahulu

Beberapa penelitian sebelumnya telah menerapkan algoritma *Support Vector Machine* (SVM) dalam analisis sentimen pada berbagai aplikasi. Penelitian [11] menganalisis sentimen pengguna aplikasi *Tokopedia* dengan metode *TF-IDF* dan membandingkan tiga kernel, yaitu *linear*, *polynomial*, dan *RBF*. Hasilnya menunjukkan bahwa kernel *RBF* memberikan akurasi tertinggi sebesar 92,4%. Penelitian [15] menggunakan SVM untuk klasifikasi sentimen ulasan pengguna aplikasi *ShopeeFood* dengan tahapan *preprocessing* dan pembobotan *TF-IDF*, menghasilkan akurasi 90% serta *precision* 88% dan *recall* 91%. Sementara itu, penelitian [26] menerapkan SVM dengan kernel *linear* pada ulasan pengguna aplikasi *Grab* dan menunjukkan bahwa algoritma ini efektif dalam memisahkan kelas teks berdimensi tinggi. Secara umum, hasil dari ketiga penelitian tersebut membuktikan bahwa SVM memiliki performa tinggi dalam klasifikasi sentimen, dan hasil akurasi sangat dipengaruhi oleh pemilihan kernel, parameter, serta metode *preprocessing* yang digunakan.

2.2. Google Play Store

Google Play Store adalah platform yang memungkinkan pengguna untuk memberikan ulasan dan peringkat terhadap aplikasi yang mereka gunakan

[8] [10]. Fitur ini berfungsi sebagai sumber informasi berharga bagi perusahaan untuk memahami kepuasan dan persepsi pengguna, yang dapat menjadi dasar untuk perbaikan dan peningkatan layanan di masa depan [8] [10].

2.3. Aplikasi Investasi Stockbit

Stockbit adalah sebuah aplikasi investasi saham yang populer di Indonesia [7]. Aplikasi ini diawasi oleh Otoritas Jasa Keuangan (OJK) [7]. Stockbit menyediakan berbagai fitur untuk mendukung pengambilan keputusan investasi, termasuk analisis teknikal, komunitas investor, dan edukasi melalui *virtual trading* [7]. Penggunaan aplikasi investasi seperti Stockbit mengharuskan pengguna berbagi data pribadi, sehingga keamanan data dan privasi menjadi isu penting yang memengaruhi tingkat kepercayaan pengguna [7].

2.4. Scraping Data

Scraping adalah teknik pengumpulan data secara otomatis dari sumber lain, seperti halaman web, menggunakan program atau algoritma khusus [8]. Tujuannya adalah untuk mengekstrak data terstruktur yang dapat diakses melalui internet dan mengubahnya ke dalam format yang dapat digunakan untuk analisis lebih lanjut [8]. Dalam penelitian ini, teknik ini digunakan untuk mengumpulkan ulasan pengguna dari Google Play Store.

2.5. Dataset Preparation

Adalah tahap penting setelah pengumpulan data mentah untuk menganalisis sentimen [5] [10]. Tujuan dari tahap ini adalah untuk membersihkan dan menyaring data agar hanya ulasan yang relevan dan unik yang digunakan dalam penelitian [5] [10].

2.6. Labeling

Pelabelan data adalah proses pemberian label atau tag pada data mentah seperti teks, untuk membantu algoritma pembelajaran mesin memahami dan menginterpretasikan informasi yang terkandung di dalamnya [13]. Tujuannya adalah untuk mengklasifikasikan data ke dalam kategori yang relevan, yang nantinya akan digunakan untuk melatih model [14].

2.7. Pra-pemrosesan (Preprocessing)

Preprocessing adalah serangkaian tahapan yang dilakukan untuk membersihkan, mengubah, dan menyesuaikan data teks mentah menjadi format yang lebih terstruktur dan siap untuk analisis [2,16]. Tahapan ini meliputi:

- Pembersihan Data (*Data Cleaning*):** Menghapus elemen-elemen yang tidak relevan seperti simbol, angka, dan emotikon untuk meningkatkan kualitas data dan membersihkan data dari *noise* [2].
- Case Folding*:** Mengubah semua karakter huruf kapital menjadi huruf kecil untuk menyeragamkan data [2].

- c. Normalisasi (*Normalization*): Menyesuaikan data agar kompatibel dengan model yang akan diimplementasikan [1].
- d. *Tokenizing*: Memecah teks menjadi unit-unit individu, seperti kata-kata atau token [1].
- e. *Stopword Removal*: Menghapus kata-kata umum yang tidak memiliki makna signifikan dan tidak relevan untuk analisis [2].
- f. *Stemming*: Mengubah kata-kata menjadi bentuk dasarnya dengan menghilangkan imbuhan [2].
- g. Penggabungan Kalimat (*Sentence Merging*): Menggabungkan kalimat yang telah diproses kembali menjadi satu kesatuan teks [2].

2.8. Pembobotan TF-IDF

TF-IDF adalah algoritma pembobotan yang mengubah data teks menjadi representasi numerik [2] [4]. Metode ini berfungsi untuk memberikan bobot pada tiap kata ataupun fitur yang ada [2] [4]. Hasilnya adalah bentuk perkalian antara TF dengan IDF [2] [4]. Nilai ini dapat digunakan sebagai fitur untuk model klasifikasi [2] [4].

2.9. Splitting Data

Pembagian data menjadi training dan testing bertujuan melatih model sekaligus menguji kemampuannya pada data baru. Proporsi umum misalnya 80:20 atau 70:30, agar model tidak overfitting maupun underfitting [2] [4].

2.10. Support Vector Machine

SVM adalah algoritma pembelajaran mesin yang efektif untuk mengklasifikasikan data [12]. SVM bekerja dengan menemukan *hyperplane* terbaik yang memisahkan data ke dalam kelas-kelas berbeda dengan jarak margin terbesar. Keunggulannya adalah efektivitasnya pada ruang berdimensi tinggi dan kemampuannya menggunakan fungsi kernel yang berbeda untuk masalah non-linear.

2.11. Parameter

Parameter C dan gamma memiliki peran penting dalam mengoptimalkan kinerja model SVM. C adalah parameter regularisasi yang mengontrol *trade-off* antara kesalahan klasifikasi pada data latih dan kelancaran margin. Nilai C yang tinggi akan menghasilkan margin yang lebih sempit tetapi dapat meminimalkan kesalahan, sementara nilai C yang rendah akan menghasilkan margin yang lebih lebar. gamma menentukan jangkauan pengaruh satu sampel data latih, yang memengaruhi seberapa ketat model mengikuti data.

2.12. Evaluasi

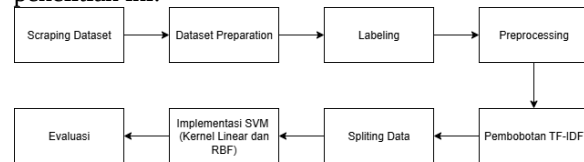
Evaluasi model dilakukan menggunakan Confusion Matrix. Matriks ini memberikan gambaran rinci tentang kinerja model dengan mengukur metrik seperti:

- a. Akurasi (*Accuracy*): Untuk mengetahui tingkat kemiripan nilai prediksi dengan nilai aktual [2].

- b. Presisi (*Precision*): Untuk mengetahui seberapa tepat atau akurat model dari yang diprediksi positif [2].
- c. *Recall*: Untuk menghitung berapa banyak dari nilai *Actual Positive* yang berhasil diidentifikasi oleh model melalui pelabelan sebagai *True Positive* [2].
- d. F1-Score adalah perbandingan rata-rata nilai presisi dan *recall* yang dibobotkan [2].

3. METODE PENELITIAN

Gambar 1 menunjukkan beberapa Langkah yang harus dilakukan untuk melakukan eksperimen penelitian ini.



Gambar 1. Alur Penelitian

3.1. Scaping Dataset

Pada tahap ini, data ulasan dikumpulkan secara otomatis dari Google Play Store. Data yang digunakan adalah ulasan aplikasi Stockbit. Proses ini menggunakan teknik *web scraping* dengan bantuan bahasa pemrograman Python. Penulis mengambil data sebanyak 1000 komentar pengguna. Data tersebut berasal dari ulasan pengguna aplikasi investasi online.

3.2. Dataset Preparation

Setelah data mentah terkumpul, tahap selanjutnya adalah mempersiapkannya untuk analisis. Proses ini meliputi:

- a. Menghapus komentar yang duplikat.
- b. Menghapus akun pengguna (*user*) yang duplikat.
- c. Menghapus data yang kosong atau tidak relevan.[15]

3.3. Labeling

Untuk mempercepat proses pelabelan data, penelitian ini akan mengimplementasikan metode semi-otomatis yang didasarkan pada peringkat (*rating*) yang diberikan oleh pengguna di Google Play Store [2, 4]. Peringkat ini akan berfungsi sebagai dasar awal untuk mengklasifikasikan sentimen:

- a. Sentimen Negatif: Ulasan dengan peringkat 1 atau 2 akan dilabeli sebagai sentimen negatif [2] [4].
- b. Sentimen Positif: Ulasan dengan peringkat 4 atau 5 akan dilabeli sebagai sentimen positif [2] [4].
- c. Sentimen Netral: Ulasan dengan peringkat 3 akan dianggap sebagai sentimen netral dan tidak digunakan dalam analisis lebih lanjut [2] [4]. Data dengan label netral ini akan dihapus dari dataset karena fokus penelitian ini adalah pada sentimen positif dan negatif [2] [4].

Dengan pendekatan ini, data ulasan akan dikategorikan secara efisien sebelum masuk ke tahap analisis lebih dalam.

3.4. Preprocessing

Data teks akan diproses melalui beberapa tahap:

- Cleaning: Menghapus karakter yang tidak relevan.
- Casfolding: Mengubah teks menjadi huruf kecil.
- Tokenizing: Memecah kalimat menjadi kata-kata.
- Normalization: Mengubah kata tidak baku menjadi kata baku.
- Stopword Removal: Menghilangkan kata-kata umum.
- Stemming: Mengubah kata berimbuhan menjadi kata dasar.

3.5. Pembobotan TF-IDF

Data teks yang telah melalui preprocessing akan diubah menjadi representasi numerik. TF-IDF adalah metode statistik yang memberikan bobot pada setiap kata berdasarkan seberapa sering kata tersebut muncul dalam sebuah cuitan, berbanding terbalik dengan seberapa sering kata itu muncul di seluruh dataset.

3.6. Splitting Data

Pada tahap ini akan dilakukan pembagian data dengan 5 perbandingan untuk menemukan split data terbaik berdasarkan nilai f1-score nya, Dimana perbandingannya sebagai berikut.

Tabel 1. Pembagian Data

Perbandingan	Jumlah Split
A	10% testing, 90% training
B	20% testing, 80% training
C	30% testing, 70% training
D	40% testing, 60% training
E	50% testing, 50% training

3.7. Implementasi SVM

Pada tahap ini, model klasifikasi dibangun menggunakan algoritma Support Vector Machine (SVM). Tujuan utamanya adalah menemukan nilai *F1-score* yang paling optimal dengan menguji berbagai kombinasi parameter dan kernel [5].

- Kernel Linear: Implementasi awal akan menggunakan kernel linear dengan nilai parameter *C* yang bervariasi untuk mengklasifikasikan data.
- Kernel RBF: Selanjutnya, model akan diimplementasikan dengan kernel RBF, yang menguji berbagai nilai parameter *C* dan *gamma* yang berbeda.

Proses ini bertujuan untuk mengoptimalkan performa model dan menemukan kombinasi parameter terbaik yang menghasilkan akurasi tertinggi dalam klasifikasi sentimen [5]. Evaluasi model akan dilakukan dengan membandingkan hasil dari setiap kombinasi parameter untuk mengidentifikasi konfigurasi yang paling efektif [5].

3.8. Evaluasi

Evaluasi menggunakan confusion matrix dimana akan menampilkan hasil dari nilai parameter dan evaluasi dinilai berdasarkan confusion matrix.

4. HASIL DAN PEMBAHASAN

4.1. Scraping Data

Dataset pada penelitian ini diperoleh dengan menggunakan proses *scraping* ulasan aplikasi Stockbit yang tersedia di platform Google Play Store. Proses *scraping* dilakukan dengan menggunakan library *google-play-scraper* pada bahasa pemrograman Python. Data yang diambil difokuskan pada ulasan pengguna dengan rentang tahun 2025. Dari proses tersebut berhasil dikumpulkan sebanyak 1.000 yang selanjutnya digunakan sebagai dataset penelitian. Hasil *scraping* dapat dilihat di gambar 2.

Jumlah data hasil scraping: 1000
Data berhasil diupload ke stockbit-reviews.csv

	reviewId	user	comment	rating	at
0	246e18d-896c-44b9-a953-e3dd39c8bdc	wahyu mulia	sumpah sangat kocwa udi 3hari gk di proses j...	1	2025-10-05 10:53:42
1	e874535-4b89-4128-a71a-94dc5299100	Asip Saepulah	sangat membantu, respon cepa...	5	2025-10-05 08:12:03
2	8db043e-e3e9-4b0c-b5cd-d8c7bbee51d	John Satra	Fitur semakin lengkap. Makin bagu...	5	2025-10-05 03:47:19
3	99b30dd-d7b8-42f2-e867-c91cd94ab64	Agus Sutarno	Aplikasi trading terbaik (fitur lengkap, mudah...	5	2025-10-05 03:17:57
4	6027af10-ad10-494d-a3a2-0a07694c2028	K Sanni	Tambahan fitur ini 🙌 1 untuk menambahkan f...	5	2025-10-05 00:52:55

Gambar 2. Scraping Dataset

4.2. Dataset Preparation

Setelah mendapatkan dataset, akan dilakukan dataset preparation dengan menghapus user, ulasan duplikat dan menghapus data kosong. Hasil dari dataset preparation dapat dilihat dari gambar 3 dimana sisa data setelah dihapus tersisa 898 data.

Jumlah data setelah preparation: 898

	reviewId	user	comment	rating	at
0	e462112-7003-4cfa-a896-45a271b0baa5	Zaid Avesta	Udah masuk ke Tahap T0 masih belum bisa diti...	1	2025-10-06 12:07:45
1	98c5a80b-7963-4698-88b-118aa1c48c	Grenly P Putra	kalo live support chat bisa di kasih fitur pane...	4	2025-10-06 09:33:18
2	b54118a9-7c7c-4a96-a8be-d1802a03d35	Sulton Azrial Firdaus	sangat bagus, dan banyak fitur nya, mudah di g...	5	2025-10-06 06:26:14
3	efba43db-93e9-4865-b1f1-826879a20427	Nur Adila	tolong wd nya jangan pakai tertunda min 🙌	4	2025-10-06 02:50:29
4	cc3c4c39-e49c-47d1-82f1-b12ba4ca578c	02 Abdi Narendra PP	👍👍👍	5	2025-10-05 15:22:41

Gambar 3. Dataset Preparation

4.3. Labeling Data

Selanjutnya melabeli data menjadi positif dan negative dimana positif = 1 jumlah datanya 607 sedangkan negative = 0 jumlah datanya 238, dapat dilihat hasilnya di gambar 4.

Distribusi label:

	count
label	
1	607
0	238

Gambar 4. Labeling Data

Kemudian pada gambar 5 ditampilkan data yang telah dilakukan labelling.

	reviewId	user	comment	rating	at	label
0	e462112-7003-4cfa-a896-45a271b0baa5	Zaid Avesta	Udah masuk ke Tahap T0 masih belum bisa diti...	1	2025-10-06 12:07:45	0
1	98c5a80b-7963-4698-88b-118aa1c48c	Grenly P Putra	kalo live support chat bisa di kasih fitur pane...	4	2025-10-06 09:33:18	1
2	b54118a9-7c7c-4a96-a8be-d1802a03d35	Sulton Azrial Firdaus	sangat bagus, dan banyak fitur nya, mudah di g...	5	2025-10-06 06:26:14	1
3	efba43db-93e9-4865-b1f1-826879a20427	Nur Adila	tolong wd nya jangan pakai tertunda min 🙌	4	2025-10-06 02:50:29	1
4	cc3c4c39-e49c-47d1-82f1-b12ba4ca578c	02 Abdi Narendra PP	👍👍👍	5	2025-10-05 15:22:41	1

Gambar 5. Hasil Labeling data

4.4. Preprocessing

a. Hasil Cleaning

Hasil cleaning ditunjukkan oleh gambar 6.

	comment	clean
0	Udah masuk ke Tahap T0 masih belum bisa ditari...	Udah masuk ke Tahap T0 masih belum bisa ditari...
1	kalo live support chat bisa dikasih fitur pane...	kalo live support chat bisa dikasih fitur pane...
2	sangat bagus, dan banyak fitur nya, mudah di g...	sangat bagus dan banyak fitur nya mudah di gun...
3	tolong wd nya jangan pakai tertunda min 🙏	tolong wd nya jangan pakai tertunda min

Gambar 6. Cleaning Data

b. Hasil CaseFolding

Proses case folding dilakukan dengan mengubah teks menjadi lowercase. Hasil casefolding perlihatkan oleh gambar 7.

	clean	lower
0	Udah masuk ke Tahap T0 masih belum bisa ditari...	udah masuk ke tahap t0 masih belum bisa ditari...
1	kalo live support chat bisa dikasih fitur pane...	kalo live support chat bisa dikasih fitur pane...
2	sangat bagus dan banyak fitur nya mudah di gun...	sangat bagus dan banyak fitur nya mudah di gun...
3	tolong wd nya jangan pakai tertunda min 🙏	tolong wd nya jangan pakai tertunda min

Gambar 7. Hasil CaseFolding

c. Hasil Normalisasi

kata-kata yang tidak baku akan dikonversi menjadi kata-kata yang baku seperti diperlihatkan oleh gambar 8.

	lower	normalized
0	udah masuk ke tahap t0 masih belum bisa ditari...	sudah masuk ke tahap t0 masih belum bisa ditar...
1	kalo live support chat bisa dikasih fitur pane...	kalo live support chat bisa dikasih fitur pan...
2	sangat bagus dan banyak fitur nya mudah di gun...	sangat bagus dan banyak fitur ya mudah di gunakan
3	tolong wd nya jangan pakai tertunda min 🙏	tolong wd ya jangan pakai tertunda min

Gambar 8. Hasil Normalisasi

d. Hasil Tokenizing

Pada proses tokenizing, setiap spasi dianggap sebagai pemisah dari sebuah token. Pengubahan kalimat menjadi token ditunjukkan oleh gambar 9.

	normalized	tokens
0	sudah masuk ke tahap t0 masih belum bisa ditar...	[sudah, masuk, ke, tahap, t0, masih, belum, bi...
1	kalo live support chat bisa dikasih fitur pane...	[kalo, live, support, chat, bisa, dikasih, fi...
2	sangat bagus dan banyak fitur ya mudah di gunakan	[sangat, bagus, dan, banyak, fitur, ya, mudah, ...
3	tolong wd ya jangan pakai tertunda min 🙏	[tolong, wd, ya, jangan, pakai, tertunda, min]

Gambar 9. Hasil Tokenizing

e. Hasil Stopword Removal

Pada tahap stopwords removal term yang terdiri dari 1 karakter huruf juga akan di hapus, seperti diperlihatkan oleh gambar 10.

	tokens	no_stopwords
0	[sudah, masuk, ke, tahap, t0, masih, belum, bi...	[masuk, tahap, t0, ditarik, cs, ngejawab, meng...
1	[kalo, live, support, chat, bisa, dikasih, fi...	[kalo, live, support, chat, dikasih, fitur, p...
2	[sangat, bagus, dan, banyak, fitur, ya, mudah, ...	[sangat, bagus, banyak, fitur, mudah, gunakan]
3	[tolong, wd, ya, jangan, pakai, tertunda, min]	[wd, jangan, pakai, tertunda, min]

Gambar 10. Hasil Stopword Removal

f. Hasil Stemming

Proses stemming dalam penelitian ini menggunakan algoritma nazief & adriani yang terdapat dalam library sastrawi. hasil stemming ditunjukkan oleh gambar 11.

	no_stopwords	stemmed
0	[masuk, tahap, t0, ditarik, cs, ngejawab, meng...	[masuk, tahap, t0, tarik, cs, ngejawab, asih, ...
1	[kalo, live, support, chat, bisa, dikasih, fitur, p...	[kalo, live, support, chat, kasih, fitur, pan...
2	[sangat, bagus, banyak, fitur, mudah, gunakan]	[sangat, bagus, banyak, fitur, mudah, guna]
3	[wd, jangan, pakai, tertunda, min]	[wd, jangan, pakai, tunda, min]

Gambar 11. Hasil Stemming

g. Penggabungan kalimat

Proses ini menggabungkan kalimat yang sipisah sebelumnya, hasil nya ditunjukkan oleh gambar 12.

	comment	preprocessed
0	Udah masuk ke Tahap T0 masih belum bisa ditari...	masuk tahap t0 tarik cs ngejawab asih solusi
1	kalo live support chat bisa dikasih fitur pane...	kalo live support chat kasih fitur panel noti...
2	sangat bagus, dan banyak fitur nya, mudah di g...	sangat bagus banyak fitur mudah guna
3	tolong wd nya jangan pakai tertunda min 🙏	wd jangan pakai tunda min

Gambar 12. Hasil Penggabungan Kalimat

4.5. Pembobotan TF-IDF

Hasil TF-IDF pada Gambar 13 menunjukkan matriks berukuran (845, 6895), yang berarti terdapat 845 dokumen/ulasan dan 6895 fitur kata unik. Setiap nilai dalam matriks merepresentasikan bobot kepentingan sebuah kata pada suatu ulasan relatif terhadap keseluruhan korpus.

Shape TF-IDF: (845, 6895)

Gambar 13. Hasil TF-IDF

4.6. Split data

Pada proses split ada dilakukan 5 perbandingan data training dan testing, setelah itu data akan dilatih menggunakan kernel svm linear untuk menemukan perbandingan daa terbaik untuk melatih kernel, dari gambar 14 dapat dilihat nilai F1-Macro atau biasa disebut F1-Score yaitu split data A yaitu 10% testing, 90% training.

	Split	Test_Size	F1_Macro	Accuracy
0	A	0.1	0.635714	0.788235
1	B	0.2	0.604557	0.775148
2	C	0.3	0.557839	0.755906
3	D	0.4	0.537167	0.751479
4	E	0.5	0.501508	0.742317

Gambar 14. Hasil Split Data

4.7. Implementasi SVM

Pada proses split data sudah didapat perbandingan data A yang terbaik, maka dari itu selanjutnya dilakukan pelaitan model menggunakan kernel Linear dan RBF.

a. Kernel Linear

Pengujian kernel linear dengan variasi parameter C menunjukkan bahwa nilai kecil (0.01 dan 0.1) menghasilkan F1-Score rendah (**0.417**). Performa meningkat tajam pada nilai C = 2 dengan F1-Macro tertinggi (**0.846**). Pada C = 5 dan 10, nilai sedikit menurun (**0.828** dan **0.823**). Dengan demikian, parameter optimal kernel linear terdapat nilai C = 2.

Hasil dapat dilihat di gambar 15.

	C	F1_Macro
0	0.01	0.417808
1	0.10	0.417808
2	1.00	0.823218
3	2.00	0.846681
4	5.00	0.828786
5	10.00	0.810268

Gambar 14. Hasil Kernel Linear

b. Kernel RBF

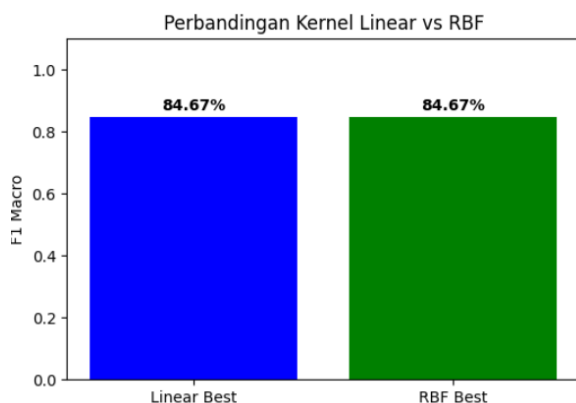
Pengujian kernel RBF dengan variasi parameter C dan gamma menunjukkan bahwa kombinasi C = 10, gamma = 0.1 memberikan performa terbaik dengan F1-Macro = 0.846. Kombinasi C = 5, gamma = 0.1 juga cukup baik (0.842). Sementara itu, nilai gamma “scale” maupun gamma lebih besar menghasilkan performa lebih rendah (0.775). Dengan demikian, parameter optimal kernel RBF adalah C = 10, gamma = 0.1.

	C	Gamma	F1_Macro
34	10.0	0.1	0.846681
28	5.0	0.1	0.823218
35	10.0	1	0.742647
18	2.0	scale	0.742647
24	5.0	scale	0.742647

Gambar 15. Hasil Kernel RBF

c. Perbandingan Kernel Linear vs RBF

Hasil pengujian menunjukkan bahwa kernel Linear dengan parameter C = 2 menghasilkan F1-Score = 0.846, sedangkan kernel RBF dengan parameter C = 10, gamma = 0.1 juga memberikan F1-Score = 0.846. Grafik pada Gambar di atas memperlihatkan bahwa kedua kernel memiliki performa yang seimbang, meskipun kernel linear cenderung lebih sederhana dan stabil, sedangkan kernel RBF memerlukan penyesuaian parameter gamma. Hasil perbandingan dapat dilihat dari gambar 16.



Gambar 16. Perbandingan Kernel

d. Workcloud Sentimen

Pada workcloud akan ditampilkan kata kata yang sering muncul baik sentimen positif dan negatif.

e. Workcloud Sentimen Positif

Pada gambar 17 ditampilkan workcloud dari proses awal sampai akhir dimana kata yang sering muncul adalah kata “mudah” sebanyak 131 jumlah muncul nya dapat dilihat di gambar 18.



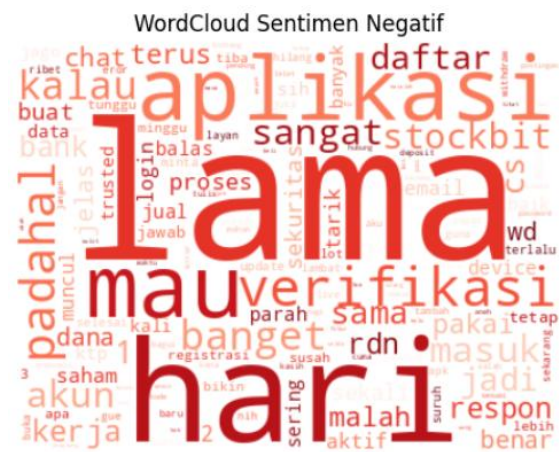
Gambar 17. Workcloud Positif

Kata	Jumlah
0	mudah 131
1	sangat 121
2	bagus 110
3	aplikasi 94
4	stockbit 85

Gambar 18. Jumlah kata

f. Workcloud Sentimen Positif

Pada gambar 19 ditampilkan workcloud dari proses awal sampai akhir dimana kata yang sering muncul adalah kata “lama” sebanyak 69 jumlah muncul nya dapat dilihat di gambar 20.



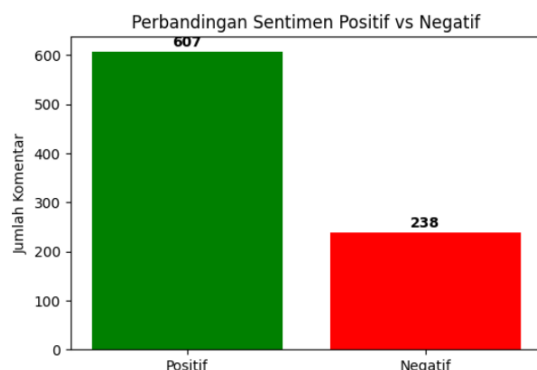
Gambar 19 Workcloud Negatif

	Kata	Jumlah
0	lama	69
1	hari	51
2	aplikasi	40
3	mau	40
4	verifikasi	39

Gambar 20. Jumlah kata

g. Perbandingan Sentimen

Setelah dilakukan proses dari awal sampai terakhir didapatkan hasil klasifikasi sentimen, dimana sentimen positif sebanyak 607 dan sentimen negatif sebanyak 238 dapat dilihat dari gambar 21 dan 22.



Gambar 21. Perbandingan Sentimen

	Sentimen	Jumlah
0	Positif	607
1	Negatif	238

Gambar 22. Jumlah sentimen

4.8. Evaluasi

Hasil evaluasi menggunakan confusion matrix menunjukkan bahwa baik kernel Linear ($C = 2.0$) maupun kernel RBF ($C = 10.0$, gamma

- True Negative (TN): 17 komentar negatif berhasil diprediksi dengan benar.
- False Positive (FP): 7 komentar negatif salah diprediksi sebagai positif.
- False Negative (FN): 3 komentar positif salah diprediksi sebagai negatif.
- True Positive (TP): 58 komentar positif berhasil diprediksi dengan benar.

Dari hasil ini terlihat bahwa kedua kernel mampu mengenali komentar positif dengan sangat baik (TP tinggi), meskipun masih terdapat kesalahan pada klasifikasi komentar negatif (FP relatif lebih besar). Secara keseluruhan, kinerja kernel Linear dan RBF sama baiknya pada dataset ini. Hasil Evaluasi dapat dilihat di gambar 17.

Confusion Matrix Linear ($C = 2.0$)

	Pred Neg	Pred Pos
True Neg	17	7
True Pos	3	58

Confusion Matrix RBF ($C = 10.0$ Gamma = 0.1)

	Pred Neg	Pred Pos
True Neg	17	7
True Pos	3	58

Gambar 17. Evaluasi

5. KESIMPULAN DAN SARAN

Penelitian ini menggunakan 1000 ulasan aplikasi Stockbit, yang setelah *dataset preparation* menjadi 898 ulasan dengan label positif (1) dan negatif (0). Setelah *preprocessing* dan pembobotan TF-IDF, data dibagi dalam lima skenario. Hasil terbaik diperoleh pada split 10:90 dengan F1-Score 0.846. Kernel Linear dan RBF sama-sama mencapai nilai tersebut, namun kernel Linear lebih disarankan karena sederhana dan stabil. Klasifikasi akhir menunjukkan 607 ulasan positif dan 238 ulasan negatif, menandakan mayoritas pengguna memberi ulasan positif terhadap Stockbit.

DAFTAR PUSTAKA

- [1] A. Nurian and B. N. Sari, "Analisis sentimen ulasan pengguna aplikasi DANA pada Google Play menggunakan Naïve Bayes," JITET (Jurnal Informatika dan Teknik Elektro Terapan), vol. 11, no. 3 S1, 2023.
- [2] A. N. Hasanah and B. N. Sari, "Analisis sentimen ulasan pengguna aplikasi jasa ojek online Maxim pada Google Play dengan metode Naïve Bayes Classifier," JITET (Jurnal Informatika dan Teknik Elektro Terapan), vol. 12, no. 1, 2024.
- [3] A. Nurian, M. S. Ma'arif, I. N. Amalia, and C. Rozikin, "Analisis sentimen pengguna aplikasi Shopee pada situs Google Play menggunakan Naïve Bayes Classifier," JITET (Jurnal Informatika dan Teknik Elektro Terapan), vol. 12, no. 1, 2024.
- [4] I. F. Rahman, A. N. Hasanah, and N. Heryana, "Analisis sentimen ulasan pengguna aplikasi Samsat Digital Nasional (SIGNAL) menggunakan metode Naïve Bayes Classifier," JITET (Jurnal Informatika dan Teknik Elektro Terapan), vol. 12, no. 2, 2024.
- [5] M. F. Y. Herjanto and C. Carudin, "Analisis sentimen ulasan pengguna aplikasi Sirekap pada Play Store menggunakan algoritma Random Forest Classifier," JITET (Jurnal Informatika dan Teknik Elektro Terapan), vol. 12, no. 2, 2024.
- [6] F. J. Sinaga, F. P. Putra, E. Pratama, and H. Irsyad, "Klasifikasi opini pengguna aplikasi Bibit pada Google Play Store menggunakan algoritma Naïve Bayes," 2nd MDP Student Conference (MSC), 2023.

- [7] A. S. D. Pratiwi, D. I. Kurniawan, F. Adestiani, L. Aulia, and Abdilah, "Analisis investasi online pada aplikasi Stockbit," *Journal Scientific of Mandalika (JSM)*, vol. 6, no. 4, 2025.
- [8] R. Zulfiqri, B. N. Sari, and T. N. Padilah, "Analisis sentimen ulasan pengguna aplikasi media sosial Instagram pada situs Google Play Store menggunakan Naïve Bayes Classifier," *JITET (Jurnal Informatika dan Teknik Elektro Terapan)*, vol. 12, no. 3, 2024.
- [9] S. Aulia and I. Ismail, "Analisis sentimen ulasan pengguna aplikasi Ojol The Game menggunakan algoritma Naïve Bayes," *JITET (Jurnal Informatika dan Teknik Elektro Terapan)*, vol. 12, no. 3, 2024.
- [10] A. Syah, F. Nurdiyansyah, and A. Y. Rahman, "Analisis sentimen aplikasi Shopee, Tokopedia, Lazada dan Blibli menggunakan leksikon dan Random Forest," *JITET (Jurnal Informatika dan Teknik Elektro Terapan)*, vol. 12, no. 3, 2024.
- [11] P. Aditiya, U. Enri, and I. Maulana, "Analisis sentimen ulasan pengguna aplikasi MyIM3 pada situs Google Play menggunakan Support Vector Machine," *JURIKOM (Jurnal Riset Komputer)*, vol. 9, no. 4, p. 1020, 2022.
- [12] R. Wahyudi and G. Kusumawardana, "Analisis sentimen pada aplikasi Grab di Google Play Store menggunakan Support Vector Machine," *J. Inform.*, vol. 8, no. 2, 2021.
- [13] M. I. Ahmadi et al., "Sentiment analysis online shop on the Play Store using method Support Vector Machine (SVM)," *Seminar Nasional Informatika*, 2020.
- [14] R. Al Anshari et al., "Komparasi payment digital untuk analisis sentimen berdasarkan ulasan di Google Play Store menggunakan metode Support Vector Machine," vol. 2, no. 3, pp. 118–128, 2023, doi: 10.55123.
- [15] R. Maulana, A. Voutama, and T. Ridwan, "NBC MyPertamina," *Jurnal Teknologi Terpadu*, vol. 9, pp. 42–48, 2023.
- [16] A. Salsabila, N. Sulistiyowati, and M. Jajuli, "Analisis sentimen pengguna Twitter terhadap fitur gratis ongkos kirim pada aplikasi Shopee Indonesia menggunakan algoritma K-Nearest Neighbor," 202