ANALISIS SENTIMEN TERHADAP PENGGUNAAN SISTEM PAYLATER PADA MEDIA SOSIAL X MENGGUNAKAN METODE SUPPORT VECTOR MACHINE

Reihan Almas Hediawan, Ahmad Faisol, Suryo Adi Wibowo
Teknik Informatika, Institut Teknologi Nasional Malang
Jalan Raya Karanglo km 2 Malang, Indonesia
Email: reihanalmas@gmail.com

ABSTRAK

Kemudahan dalam bertransaksi menggunakan layanan *paylater* telah memunculkan berbagai reaksi dari masyarakat, terutama di platform media sosial. Pandangan yang beragam, baik yang menekankan manfaat maupun potensi risiko dari layanan ini, menjadikan analisis sentimen penting untuk mendapatkan gambaran objektif atas opini publik. Penelitian ini bertujuan merancang sistem klasifikasi sentimen masyarakat terhadap layanan *paylater* ke dalam tiga kategori utama yaitu positif, negatif, dan netral, dengan memanfaatkan algoritma *Support Vector Machine* (SVM). Proses penelitian mencakup tahap pengumpulan data, pra-pemrosesan, pelabelan, serta pelatihan model. Data dikumpulkan dari unggahan *Twitter* yang menggunakan tagar #shopeepaylater, #gopaylater, dan #tiktokpaylater. Dari hasil evaluasi, pembagian data dengan proporsi 85% untuk pelatihan dan 15% untuk pengujian menghasilkan akurasi sebesar 83,5%. Capaian ini menunjukkan bahwa metode SVM cukup andal dalam mengelompokkan sentimen, dengan kecenderungan opini publik yang dominan bersifat negatif terhadap layanan tersebut.

Kata kunci : *Analisis Sentimen, Paylater, Support Vector Machine, Flask, Media Sosial X*

1. PENDAHULUAN

Perkembangan teknologi informasi dan komunikasi khususnya internet semakin pesat di dunia, hal tersebut sangat mempengaruhi berbagai bidang salah satunya yaitu pola konsumsi dan belanja masyarakat.[1] Pola konsumsi dan belanja masyarakat dimudahkan dengan kehadiran e-*commerce*. Kemajuan teknologi *fintech* atau *Financial Technology* memberikan kemudahan dalam bidang jasa keuangan yang beriringan dengan pola konsumsi dan belanja masyarakat. Salah satu bentuk inovasi dalam bidang *financial technology* (fintech) adalah hadirnya sistem *paylater*, yakni sebuah fitur yang memungkinkan pengguna untuk memperoleh barang atau jasa melalui platform tertentu dengan skema cicilan tanpa memerlukan kartu kredit.[2]

Perkembangan teknologi tersebut memberikan berbagai dampak diantaranya yaitu membuat seseorang menjadi konsumtif yang akan membentuk perilaku belanja *impulsif buying*.[3] Ditambah lagi dengan banyaknya *platform* untuk belanja secara *online* yang lebih disukai masyarakat karena begitu sederhana dengan harga yang lebih terjangkau. Banyak kemudahan yang didapat dari belanja *online* diantaranya adalah kemudahan transaksi yang bisa dilakukan dimana saja.[4]

Sebagai bagian dari perkembangan fintech, sistem *pay later* hadir sebagai solusi alternatif dalam transaksi keuangan yang menawarkan fleksibilitas pembayaran bagi konsumen. Dengan adanya sistem *pay later* memberikan dampak yang beraneka ragam dan beberapa opini yang bermacam-macam dari masyarakat [5]. Macam-macam data mengenai opini dari masyarakat bisa didapat melalui sosial media X. Hal tersebut dikarenakan sosial media X merupakan sebuah *platform* yang cukup banyak digunakan oleh masyarakat Indonesia, khususnya untuk memberikan suatu opini [6]. Opini mengenai sistem *pay later* dapat ditemukan dalam berbagai bentuk seperti penggunaan *hashtag* #shopeepaylater atau #gopaylater. Opini yang sudah terkumpul kemudian diklasifikasikan dan dianalisa dengan menggunakan metode untuk dapat mengelompokkan sentimen tersebut.

Untuk memahami dan mengelompokkan opini masyarakat, digunakan pendekatan yang dikenal sebagai analisis sentimen. Metode ini secara otomatis bekerja untuk mengenali, mengekstrak, serta memproses informasi dari teks guna mengungkap nuansa emosional atau sikap dalam suatu pernyataan [7]. Biasanya, analisis sentimen mengelompokkan data ke dalam tiga kategori utama, yaitu sentimen positif, negatif, dan netral. Dalam pelaksanaannya, sejumlah algoritma klasifikasi telah digunakan, seperti *Support Vector Machine* (SVM), *Naïve Bayes Classifier* (NBC), dan *K-Nearest Neighbor* (K-NN). Dari ketiganya, SVM sering kali dianggap unggul dalam pemrosesan teks dalam ranah *Natural Language Processing* (NLP) karena kemampuannya memberikan akurasi yang tinggi. Temuan dari penelitian Pamungkas mendukung keunggulan ini, dengan menunjukkan bahwa SVM mampu mencapai akurasi 90,1%, lebih tinggi dibandingkan *Naïve Bayes* yang memperoleh 79,20% dan K-NN dengan 62,10%.[8]

Berdasarkan hal tersebut, maka peneliti akan meneliti mengenai analisis sentimen untuk mengetahui opini masyarakat dengan adanya sistem *paylater* pada media sosial X dengan metode SVM. Metode SVM dipilih karena memiliki tingkat akurasi yang tinggi dalam klasifikasi teks berbasis sentimen, sebagaimana ditunjukkan dalam penelitian sebelumnya.

2. TINJAUAN PUSTAKA

2.1. Penelitian Terdahulu

Pamungkas melakukan penelitian yang berfokus pada analisis sentimen terhadap respons masyarakat Indonesia mengenai pandemi Covid-19 melalui media sosial *Twitter*. Dalam kajiannya, ia membandingkan efektivitas tiga algoritma klasifikas *Support Vector Machine* (SVM), *Naïve Bayes*, dan *K-Nearest Neighb*or (KNN) untuk menilai kemampuan masing-masing dalam mengelompokkan opini publik. Hasil penelitian menunjukkan bahwa algoritma SVM dengan kernel *linear* memberikan performa terbaik, mencatat akurasi hingga 90,01%, lebih tinggi dibandingkan dengan *Naïve Bayes* (79,20%) dan KNN (62,10%). Temuan ini menegaskan keunggulan SVM dalam menangani analisis teks dengan struktur yang kompleks dan berdimensi tinggi [8].

Raflizar melakukan penelitian mengenai analisis sentimen terhadap ujaran kebencian di platform media sosial dengan menggunakan algoritma klasifikasi *Support* Vector *Machine* (SVM) yang dikombinasikan dengan teknik pembobotan kata *Term Frequency-Inverse Document Frequency* (TF-IDF). Dalam studi ini, dirancang sebuah sistem berbasis *web* yang mampu membedakan dengan tepat antara teks yang mengandung unsur kebencian dan yang tidak. Berdasarkan hasil pengujian terhadap 100 kalimat, sistem menunjukkan performa yang cukup optimal dengan akurasi sebesar 81%, *precision* mencapai 96%, dan *recall* sebesar 76%. Hasil ini menunjukkan bahwa integrasi antara tahapan pra-pemrosesan teks, metode TF-IDF, dan algoritma SVM sangat efektif dalam mendeteksi ujaran kebencian berbahasa Indonesia, terutama dalam konteks Pemilu 2024 [9].

Cornelia melakukan studi yang menitikberatkan pada analisis sentimen terkait kampanye pengurangan plastik di media sosial, dengan menggunakan algoritma *Support Vector Machine* (SVM). Tujuan dari penelitian ini adalah untuk memahami tanggapan publik terhadap kampanye tersebut dengan mengklasifikasikan opini menjadi tiga jenis sentimen: positif, negatif, dan netral. Rangkaian proses dalam penelitian ini meliputi pengambilan data dari *Twitter*, tahap preprocessing untuk membersihkan dan menormalkan teks, pembobotan kata menggunakan metode *Term Frequency-Inverse Document Frequency* (TF-IDF), serta pelatihan dan evaluasi model klasifikasi dengan algoritma SVM. Dari total 1.555 data yang dianalisis, model menunjukkan tingkat akurasi sebesar 70,64%, dengan *recall* tertinggi terdapat pada sentimen positif. Hasil ini menunjukkan bahwa SVM merupakan metode yang cukup efektif untuk mendeteksi arah opini publik terkait isu lingkungan di media sosial[10].

Okta melakukan penelitian mengenai analisis sentimen pada *review* film dengan menerapkan teknik *Term Frequency-Inverse Document Frequency* (TF-IDF) sebagai metode untuk memberikan bobot pada kata, serta *algoritma Support Vector Machine* (SVM) dalam proses klasifikasi. Berdasarkan hasil evaluasi, algoritma SVM menunjukkan kinerja yang sangat baik dalam mengidentifikasi sentimen dalam teks ulasan, dengan akurasi sebesar 85%, *precision* mencapai 100%, *recall* sebesar 70%, dan F1-*Score* sebesar 82%. Hasil ini mengindikasikan bahwa penerapan TF-IDF bersama SVM mampu menangani teks berdimensi tinggi secara efektif dan menghasilkan akurasi klasifikasi yang memuaskan[7].

Ferra Junian Wahidna dan Paramitha Nerisafitra melakukan studi mengenai analisis sentimen terhadap pengguna layanan *paylater* dengan memanfaatkan algoritma *Support Vector Machine* (SVM) dan metode pembobotan berbasis *Lexicon*. Penelitian ini bertujuan untuk mengklasifikasikan opini publik yang diekspresikan melalui unggahan di *Twitter* mengenai penggunaan layanan *paylater*. Berdasarkan hasil penelitian, model klasifikasi yang dirancang mampu mendeteksi sentimen masyarakat terhadap Shopee Paylater dan Go Paylater secara akurat. Pengujian terhadap data Shopee Paylater menghasilkan akurasi sebesar 89,74%, sementara pada data Go Paylater, akurasinya mencapai 90,27%. Hasil tersebut menunjukkan bahwa penggunaan SVM cukup efektif dalam memahami persepsi pengguna terhadap layanan keuangan digital [2].

2.2. Text Preprocessing

Tahap pra-pemrosesan merupakan langkah awal yang krusial dalam pengolahan data, karena berfungsi untuk mengubah data mentah menjadi bentuk yang lebih terorganisir dan siap dianalisis. Tahapan ini umumnya mencakup proses pembersihan dari informasi yang tidak dibutuhkan serta penyesuaian struktur data agar dapat diproses secara optimal oleh sistem. Dalam konteks analisis sentimen, pra-pemrosesan memiliki peranan penting, terutama dalam menangani data dari media sosial yang kerap kali menggunakan bahasa tidak formal, istilah sehari-hari, dan gaya penulisan yang bervariasi sesuai karakteristik penggun [11].

2.3. Pembobotan TF-IDF

TF-IDF (*Term Frequency-Inverse Document Frequency*) merupakan salah satu metode populer dalam bidang Pemrosesan Bahasa Alami (NLP) yang digunakan untuk mengekstraksi fitur penting dari teks. Teknik ini memberikan bobot pada kata-kata dalam sebuah dokumen dengan memperhitungkan seberapa sering kata tersebut muncul di dalam dokumen tertentu, serta seberapa jarang kemunculannya di keseluruhan kumpulan dokumen. Pendekatan ini membantu dalam menyoroti istilah-istilah yang paling relevan dan informatif dalam konteks analisis teks [12].Pendekatan ini menentukan bobot suatu kata dalam dokumen dengan mempertimbangkan frekuensi kemunculannya di dokumen tersebut serta tingkat kelangkaannya di seluruh kumpulan dokumen (korpus). Nilai bobot yang dihasilkan dapat dimanfaatkan dalam berbagai aplikasi, seperti klasifikasi teks, pengelompokan dokumen

(clustering), dan pengembangan sistem pencarian informasi (information retrieval).

a. *Term Frequency* (TF)
TF (*Term Frequency*) mengacu pada jumlah kemunculan suatu kata dalam sebuah dokumen tertentu. Semakin sering kata tersebut muncul, semakin besar pula bobot atau tingkat kepentingannya terhadap isi dokumen, karena dianggap berkontribusi secara signifikan dalam merepresentasikan topik atau konteks keseluruhan dari dokumen yang dianalisis [9]. Persamaan (1) dapat digunakan untuk memformulasikan *Frequency Term* (TF).

$$F = 1 + \log(F_{t,d}), F_{t,d} > 0 \quad (1)$$

Keterangan :
TF = merupakan singkatan dari *Term Frequency*, yaitu ukuran yang menunjukkan seberapa sering suatu istilah muncul dalam sebuah dokumen.
$F_{t,d}$ = menyatakan jumlah kemunculan istilah tertentu di dalam dokumen yang sedang dianalisis.

b. *Inverse Document Frequency* (IDF)
Inverse Document Frequency (IDF) merupakan metrik yang digunakan untuk mengukur tingkat penyebaran suatu kata di seluruh koleksi dokumen. Jika sebuah istilah jarang ditemukan dalam berbagai dokumen, maka IDF-nya akan bernilai tinggi karena dianggap lebih spesifik dan memiliki nilai informasi yang lebih besar dibandingkan kata-kata umum yang sering muncul. Nilai IDF menunjukkan pentingnya suatu kata dalam membedakan antar dokumen, dan berkorelasi langsung dengan jumlah dokumen yang mengandung istilah tersebut. Perhitungan IDF biasanya dilakukan dengan menggunakan rumus yang ditampilkan pada Persamaan (2).

$$IDF = \log\left(\frac{D}{df}\right) \quad (2)$$

Keterangan :
IDF = merupakan singkatan dari *Inverse Document Frequency*, yaitu ukuran yang menggambarkan seberapa jarang sebuah kata muncul di seluruh kumpulan dokumen.
D = menyatakan total jumlah dokumen yang dianalisis dalam koleksi.
df = menunjukkan jumlah dokumen yang memuat istilah tertentu (*term*) yang sedang dianalisis.
Nilai TF-IDF dihitung dengan cara mengalikan *Term Frequency* (TF) dengan *Inverse Document Frequency* (IDF). Perhitungan ini dilakukan sesuai dengan rumus yang terdapat pada Persamaan (3), dengan tujuan untuk menilai sejauh mana suatu kata dianggap penting dalam sebuah dokumen dibandingkan dengan keseluruhan dokumen yang dianalisis.

$$W = tf \times idf \quad (3)$$

Keterangan:
W = hasil penghitungan bobot dengan menggunakan metode TF-IDF
tf = menyatakan frekuensi suatu kata dalam dokumen yang telah dihitung menggunakan rumus *Term Frequency*
idf = menunjukkan nilai IDF, yang diperoleh dari proses perhitungan terhadap distribusi kata dalam seluruh dokumen.

2.4. Support Vector Machine

Metode *Support Vector Machine* dapat digunakan untuk mengatasi persoalan klasifikasi linear melalui transformasi matematis terhadap ruang data dengan menggunakan fungsi kernel. Konsep dasar dari SVM adalah melakukan klasifikasi dengan mencari *hyperplane* optimal yang mampu memberikan jarak atau batas pemisah terbaik di antara dua kelas data yang berbeda.[13].

Hyperplane terbaik didapatkan dengan memilihi margin maksimum antara dua kelas. Dengan demikian, diasumsikan bahwa dua kelas data dapat dipisahkan secara sempurna oleh suatu *hyperplane* (*linear separable*). Namun, dalam praktiknya, kedua kelas dalam ruang *input* umumnya tidak dapat dipisahkan secara sempurna oleh *hyperplane* (*non-linear separable*). Untuk mengetahui *hyperplane* pada SVM dapat menggunakan Persamaan (4).

$$(W.X_i) + b = 0 \quad (4)$$

Untuk data x_i, yang diklasifikasikan ke dalam kelas -1, kondisi tersebut dapat ditulis sebagaimana terlihat pada Persamaan (5).

$$(W.X_i + b) \leq 1, y_i = -1 \quad (5)$$

Sementara itu, data yang termasuk dalam kelas 1 mengikuti Persamaan (6)

$$(W.X_i + b) \geq 1, y_i = 1 \quad (6)$$

2.5. Evaluasi Klasifikasi

Evaluasi bertujuan untuk mengukur sejauh mana model klasifikasi mampu bekerja dengan baik. Proses ini penting dilakukan agar dapat diketahui tingkat akurasi dan keandalan prediksi yang dihasilkan oleh sistem dalam mengelompokkan data ke dalam kelas-kelas yang tepat [14]. Untuk mendapatkan nilainya, dibutuhkan Matriks Evaluasi Prediksi seperti pada Tabel 1.

Tabel 1. Matriks Evaluasi Prediksi

Kelas Prediksi	**Kelas Sebenarnya**	
	Positif	**Netral**
Positif	Benar Positif(TP)	Salah Negatif (FN)
Negatif	Salah Positif (FP)	Benar Negatif (TN)

Berdasarkan tabel Matriks Evaluasi Prediksi yang disajikan dalam Tabel 1, dapat diidentifikasi berbagai metrik evaluasi kinerja model yang akan dijelaskan sebagai berikut [15]:

a. *Recall* adalah metrik yang digunakan untuk mengevaluasi sejauh mana suatu model berhasil mengenali hasil positif secara tepat. Metrik ini dikenal pula dengan istilah sensitivitas model, yang mencerminkan kemampuan model dalam mendeteksi keseluruhan *instance* positif yang ada.

Adapun rumus perhitungan *recall* dapat dilihat sebagai berikut[16]:

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

b. *Precision* mengukur sejauh mana model mampu memprediksi nilai positif dengan akurat. Dengan kata lain, *precision* menunjukkan tingkat ketepatan model dalam mengidentifikasi prediksi positif yang benar, yang juga dikenal sebagai *the positive predictive value* (nilai prediktif positif). Adapun rumus perhitungan *precision* dapat dilihat sebagai berikut[16]:

$$Precision = \frac{TP}{TP + FN} \times 100\% \quad (8)$$

c. Akurasi merupakan salah satu indikator evaluasi yang umum dipakai untuk mengukur kinerja sebuah model klasifikasi. Nilai akurasi diperoleh dari *confusion matrix* dengan menghitung perbandingan antara jumlah prediksi yang tepat dengan keseluruhan data yang dianalisis.[16]

$$Accuracy = \frac{TN+TP}{TN+TP+FN+FP} \times 100\% \quad (9)$$

Pada penelitian ini, penilaian terhadap performa model dilakukan menggunakan *Confusion Matrix* berukuran 3x3, yang dikembangkan sesuai dengan tiga kategori sentimen hasil klasifikasi, yakni positif, negatif, dan netral. Struktur matriks ini dirancang untuk merepresentasikan hubungan antara prediksi model dan kategori sentimen aktual dari data yang dianalisis.

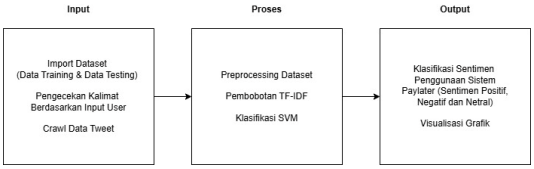
Tabel 2. *Multiclass Confusion Matrix*

		Prediksi		
		Positif	**Negatif**	**Netral**
Aktual	**Positif**	PP	PN	PT
	Negatif	NP	NN	NT
	Netral	TP	TN	TT

Keterangan:
P = Positif
N = Negatif
T = Netral

3. METODE PENELITIAN

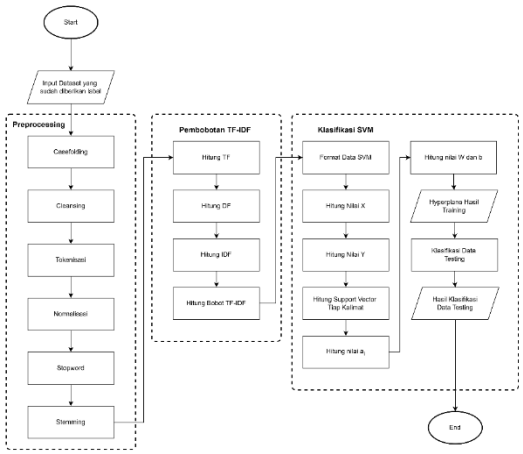
3.1 Diagram Blok Sistem



Gambar 1. Diagram Blok Sistem

Gambar 1 menyajikan representasi diagram dari aplikasi yang sudah dirancang, yang mana diantaranya adalah untuk mendukung proses *import* dataset (*training* dan *testing*), pengecekan kalimat, serta *crawling tweet* berdasarkan kata kunci tertentu. Sistem ini menjalankan tiga tahap utama yang meliputi proses *preprocessing*, pemberian bobot menggunakan metode TF-IDF, serta klasifikasi dengan menerapkan algoritma *Support Vector Machine*. Hasil akhirnya berupa label sentimen (positif, negatif, atau netral) disertai visualisasi dalam bentuk grafik.

3.2 Flowchart Metode



Gambar 2. *Flowchart* Metode

Gambar 2 menggambarkan alur kerja sistem klasifikasi dengan metode *Support Vector Machine* (SVM). Tahapan dimulai dari pemuatan *dataset* yang telah diberi label, kemudian dilanjutkan dengan proses pra-pemrosesan data, meliputi *casefolding*, pembersihan teks (*cleansing*), tokenisasi, normalisasi, penghapusan kata-kata umum (*stopword removal*), dan stemming. Setelah itu, data yang telah dibersihkan akan dibobot menggunakan teknik *Term Frequency-Inverse Document Frequency* (TF-IDF). Bobot tersebut dikonversi ke dalam bentuk vektor yang sesuai untuk digunakan dalam proses klasifikasi oleh SVM, termasuk identifikasi *support vector* serta perhitungan parameter a_i, w, dan b untuk membentuk *hyperplane*. *Hyperplane* inilah yang kemudian diuji menggunakan data validasi guna menghasilkan klasifikasi akhir terhadap sentimen dalam data.

3.3 Pengumpulan Data

Dalam tahap ini, sebanyak 1.617 data *tweet* berhasil dikumpulkan dari media sosial X, yang mencakup rentang waktu antara Januari 2021 hingga Mei 2025. Proses pengambilan data dilakukan dengan menggunakan kata kunci berbasis *hashtag*, seperti #shopeepaylater, #gopaylater, dan #tiktokpaylater, yang berkaitan langsung dengan objek analisis sentimen. Pengumpulan data dilakukan melalui teknik *crawling* menggunakan *library tweet-harvest*, dan hasilnya kemudian disimpan dalam format CSV untuk mendukung proses analisis selanjutnya.

3.4 Pelabelan Data

Sebanyak 1.617 data *tweet* yang telah dikumpulkan dari media sosial X kemudian diberi label sentimen oleh anotator, yaitu positif, negatif, atau netral, disertai dengan alasan yang mendasari pemberian label tersebut. Tabel 3 menyajikan salah satu contoh data yang telah melalui proses pelabelan oleh anotator.

Tabel 3. Pelabelan Data

Teks	Label	Alasan
cc @gopayindonesia min ini ko di blokir gopaylater dahal ga telat bayar ? ANEH	Negatif	Cuitan tersebut tergolong sentimen negatif karena pengguna Gopaylater mengeluhkan pemblokiran akun tanpa alasan, meskipun tidak ada keterlambatan pembayaran

3.5 Preprocessing Data

Langkah yang dilakukan sebelum proses klasifikasi adalah melakukan proses *preprocessing text*. *Preprocessing text* terdiri dari beberapa langkah, diantaranya adalah *Casefolding*, *Cleansing*, Tokenisasi, Normalisasi, *Stopword* serta *Stemming*. Detail tahap *preprocessing text* adalah sebagai berikut:

a. *Casefolding*

Tabel 4. Tahapan *Casefolding*

Input Proses	Output Proeses
Baru aktivin sPaylater menurutku gpp nyicil asal tau batas kemampuan yang bahaya kalo berani ambil cicilan apalagi yg besar tapi ga siap buat bayar. Btw aku pake spaylater buat nyari promo aja https://t.co/e8v0iM9SCF	baru aktivin spaylater menurutku gpp nyicil asal tau batas kemampuan yang bahaya kalo berani ambil cicilan apalagi yg besar tapi ga siap buat bayar. btw aku pake spaylater buat nyari promo aja https://t.co/e8v0im9scf

Tabel 4 menampilkan contoh *casefolding*, yakni konversi huruf kapital menjadi huruf kecil seperti pada kata "Baru", "sPaylater", dan "Btw" yang diubah menjadi "baru", "spaylater", dan "btw".

b. *Cleansing*

Tabel 5. Tahapan *Cleansing*

Input Proses	Output Proeses
baru aktivin spaylater menurutku gpp nyicil asal tau batas kemampuan yang bahaya kalo berani ambil cicilan apalagi yg besar tapi ga siap buat bayar. btw aku pake spaylater buat nyari promo aja https://t.co/e8v0im9scf	baru aktivin spaylater menurutku gpp nyicil asal tau batas kemampuan yang bahaya kalo berani ambil cicilan apalagi yg besar tapi ga siap buat bayar btw aku pake spaylater buat nyari promo aja

Tabel 5 menunjukkan contoh *cleansing*, yaitu pembersihan elemen seperti URL, emoji, dan simbol, ditandai dengan penghapusan URL pada data.

c. Tokenisasi

Tabel 6. Tahapan Tokenisasi

Input Proses	Output Proeses
baru aktivin spaylater menurutku gpp nyicil asal tau batas kemampuan yang bahaya kalo berani ambil cicilan apalagi yg besar tapi ga siap buat bayar btw aku pake spaylater buat nyari promo aja	'baru', 'aktivin', 'spaylater', 'menurutku', 'gpp', 'nyicil', 'asal', 'tau', 'batas', 'kemampuan', 'yang', 'bahaya', 'kalo', 'berani', 'ambil', 'cicilan', 'apalagi', 'yg', 'besar', 'tapi', 'ga', 'siap', 'buat', 'bayar', 'btw', 'aku', 'pake', 'spaylater', 'buat', 'nyari', 'promo', 'aja'

Tabel 6 menampilkan tokenisasi, yaitu pemisahan teks menjadi kata-kata terpisah berdasarkan spasi.

d. Normalisasi

Tabel 7. Tahapan Normalisasi

Input Proses	Output Proeses
baru aktivin spaylater menurutku gpp nyicil asal tau batas kemampuan yang bahaya kalo berani ambil cicilan apalagi yg besar tapi ga siap buat bayar btw aku pake spaylater buat nyari promo aja	Baru aktivin spaylater menurutku tidak apa-apa nyicil asal tahu batas kemampuan yang berisiko kalau berani ambil cicilan apalagi yang besar tapi tidak siap untuk membayar btw aku menggunakan spaylater untuk mencari promo aja

Tabel 7 menyajikan output dari proses normalisasi, yaitu langkah untuk mengonversi kata-kata tidak baku menjadi bentuk ejaan yang sesuai standar, dengan mencocokkannya terhadap entri yang terdapat dalam kamus normalisasi.

e. *Stopword*

Tabel 8. Tahapan *Stopword*

Input Proses	Output Proeses
Baru aktivin spaylater menurutku tidak apa-apa nyicil asal tahu batas kemampuan yang berisiko kalau berani ambil cicilan apalagi yang besar tapi tidak siap untuk membayar btw aku menggunakan spaylater untuk mencari promo aja	baru aktivin spaylater menurutku tidak apa-apa nyicil asal tahu batas kemampuan berisiko kalau berani ambil cicilan besar siap membayar aku menggunakan spaylater mencari promo

Tabel 8 memperlihatkan tahap penghapusan *stopword*, yaitu proses mengeliminasi kata-kata dengan bobot makna yang rendah atau kurang relevan, seperti "yang", "tapi", "tidak", "untuk", "btw", dan "aja", guna meningkatkan kualitas analisis teks.

f. *Stemming*

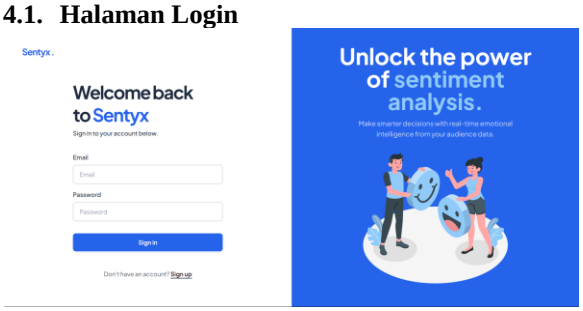
Tabel 9. Tahapan *Stemming*

Input Proses	Output Proeses
baru aktivin spaylater menurutku tidak apa-apa nyicil asal tahu batas kemampuan berisiko kalau berani ambil cicilan besar siap membayar aku menggunakan spaylater mencari promo	baru aktivin spaylater turut tidak apa nyicil asal tahu batas mampu risiko kalau berani ambil cicil besar siap bayar aku guna spaylater cari promo

Tabel 9 menunjukkan *stemming*, yaitu pengubahan kata ke bentuk dasar seperti “menurutku” menjadi “turut”, dan “membayar” menjadi “bayar”.

4. HASIL DAN PEMBAHASAN

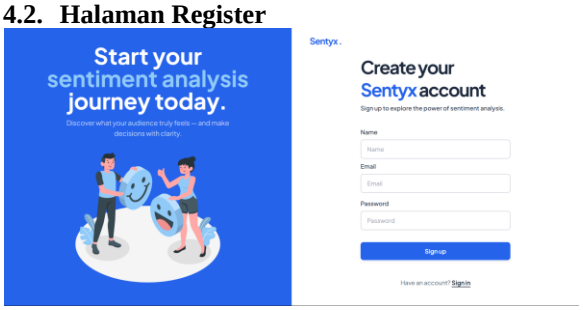
4.1. Halaman Login



Gambar 3. Halaman *Login*

Gambar 3 memperlihatkan antarmuka halaman login, yang meminta pengguna untuk memasukkan *email* dan kata sandi yang telah terdaftar di sistem guna mendapatkan akses ke dalam aplikasi.

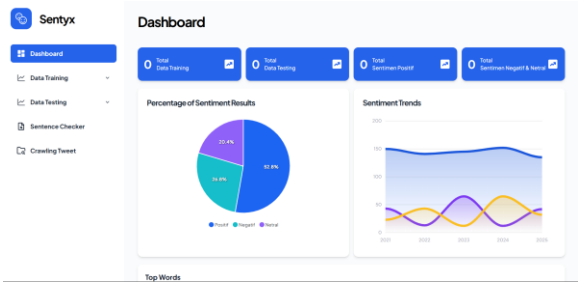
4.2. Halaman Register



Gambar 4. Halaman *Register*

Gambar 4 memperlihatkan tampilan halaman pendaftaran akun, di mana pengguna dapat membuat akun baru dengan mengisi nama, *email*, dan kata sandi guna memperoleh akses ke fitur *login* dan *dashboard*.

4.3. Halaman Dashboard

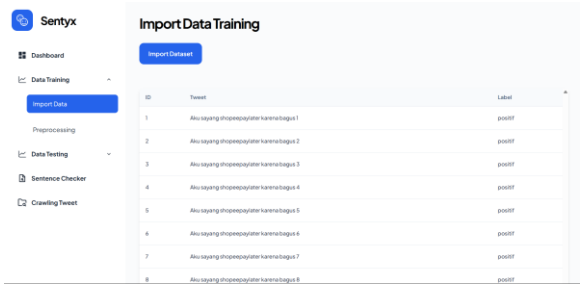


Gambar 5. Halaman *Dashboard*

Gambar 5 menampilkan tampilan *dashboard* yang berisi informasi mengenai jumlah data *training* dan *testing*, disertai grafik yang merepresentasikan distribusi sentimen positif, negatif, dan netral dalam bentuk persentase.

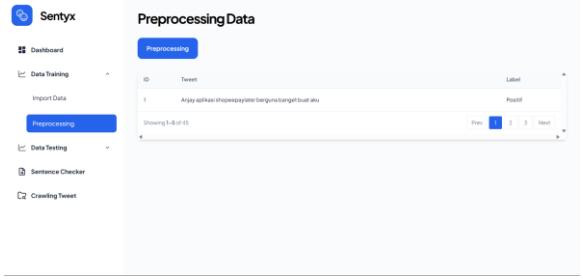
4.4. Halaman Import Data Training

Gambar 6 menampilkan antarmuka halaman unggah data pelatihan, di mana pengguna dapat memasukkan file berformat CSV dan melihat tampilan data yang berhasil diimpor ke dalam sistem.



Gambar 6. Halaman *Import* Data *Training*

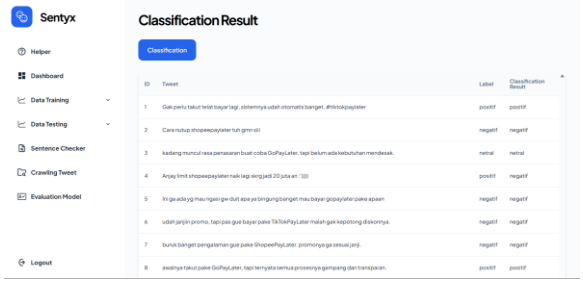
4.5. Halaman Preprocessing Data Training



Gambar 7. Halaman *Preprocessing* Data *Training*

Gambar 7 memperlihatkan tampilan halaman *preprocessing training* yang menyajikan hasil data setelah melalui tahap *preprocessing*.

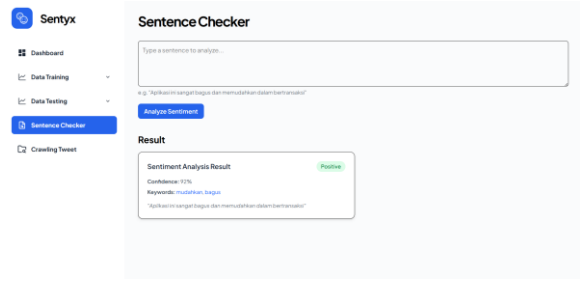
4.6. Halaman Hasil Klasifikasi Data Training



Gambar 8. Halaman Klasifikasi Data *Training*

Gambar 8 memperlihatkan tampilan halaman klasifikasi *training* yang menyajikan hasil klasifikasi menggunakan model SVM dari data yang telah dilatih.

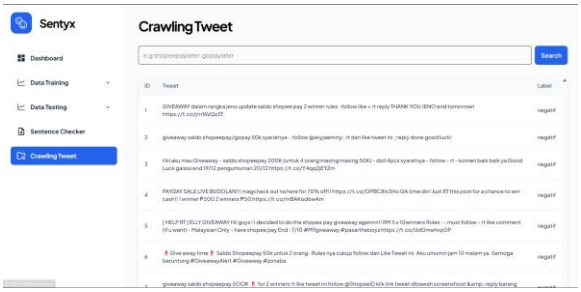
4.7. Halaman Cek Kalimat



Gambar 9. Halaman Cek Kalimat

Gambar 9 menampilkan halaman cek kalimat, di mana pengguna dapat memasukkan kalimat untuk dianalisis sentimennya, dan hasilnya ditampilkan pada kartu di bawah.

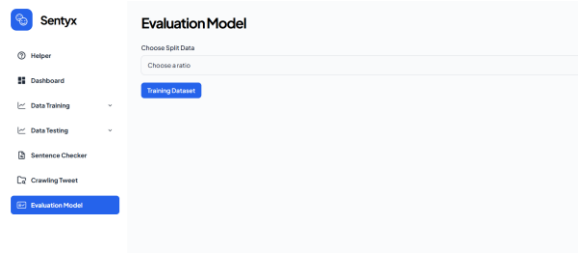
4.8. Halaman Crawl Tweet



Gambar 10. Halaman *Crawl Tweet*

Gambar 10 menampilkan halaman *crawling tweet*, yang memungkinkan pengguna mengambil *tweet* berdasarkan *keyword*, lalu memproses dan mengklasifikasikannya menggunakan SVM untuk menentukan sentimennya.

4.9. Halaman Evaluasi Model



Gambar 1.1 Halaman Evaluasi Model

Gambar 11 menampilkan halaman evaluasi model, yang menyajikan penilaian kinerja klasifikasi berdasarkan akurasi, presisi, *recall*, dan *confusion matrix* sesuai rasio data yang dipilih.

4.10. Pengujian Metode

Dalam penelitian untuk menguji kinerja model klasifikasi dengan memillih beberapa skema split data. Skema split data ini memiliki rasio diantaranya adalah 85:15 dan 90:10. Keakuratan sistem dievaluasi dengan membandingkan hasil klasifikasi pada data *training* dengan label yang diberikan anotator menggunakan *confusion matrix.*

Tabel 10. *Confusion Matrix* Rasio 85:15

True Class	Predicted Class			Total
	Negatif	Netral	Positif	
Negatif	87	4	2	93
Netral	9	55	9	73
Positif	12	4	61	77
Total	108	63	72	243

Tabel 11. *Confusion Matrix* Rasio 90:10

True Class	Predicted Class			Total
	Negatif	Netral	Positif	
Negatif	57	3	2	62
Netral	7	37	5	49
Positif	7	3	41	51
Total	71	43	46	162

Hasil dari *confusion matrix* digunakan untuk menentukan nilai *recall*, *precision* dan akurasi. Hasil perhitungan evaluasinya adalah sebagai berikut:

Tabel 12. Matriks Evaluasi Rasio 85:15

Label	Precision	Recall
Negatif	$\frac{87}{108}x100\% = 80.5\%$	$\frac{87}{93}x100\% = 93.5\%$
Netral	$\frac{55}{63}x100\% = 87.3\%$	$\mathbf{\frac{55}{73}x100\% = 75.3\%}$
Positif	$\frac{61}{72}x100\% = 84.7\%$	$\frac{61}{77}x100\% = 79.2\%$
Akurasi	$\frac{87+55+61}{243}x100\% = \frac{203}{243}x100\% = 83.5\%$	

Pada Tabel 12 menunjukkan kinerja terbaik pada kelas netral dengan *precision* tertinggi sebesar 87,3%, sedangkan kelas positif dan negatif memiliki *precision* 84,7% dan 78,7%. Dari sisi *recall*, model paling efektif mengenali data negatif yaitu 93,5%, sementara *recall* untuk kelas positif dan netral masing-masing 75,3% dan 79,2%. Secara keseluruhan, model berhasil mengklasifikasikan data dengan akurasi sebesar 83,5%.

Tabel 13. Matriks Evaluasi Rasio 90:10

Label	Precision	Recall
Negatif	$\frac{57}{71}x100\% = 80.2\%$	$\frac{57}{62}x100\% = 91.9\%$
Netral	$\frac{37}{43}x100\% = 86\%$	$\mathbf{\frac{37}{49}x100\% = 75.5\%}$
Positif	$\frac{41}{46}x100\% = 89.1\%$	$\frac{41}{51}x100\% = 80.3\%$
Akurasi	$\frac{57+37+41}{162}x100\% = \frac{135}{162}x100\% = 83.3\%$	

Berdasarkan data pada Tabel 13, model menghasilkan tingkat akurasi sebesar 83,3%. *Precision* tertinggi dicapai pada kategori positif sebesar 89,1%, disusul oleh kategori netral dengan 86%, dan negatif sebesar 80,2%. Sementara itu, dari sisi *recall*, kategori negatif menunjukkan performa paling optimal dengan 91,9%, diikuti oleh kategori positif dan netral masing-masing sebesar 80,3% dan 75,5%. Temuan tersebut menunjukkan bahwa model mampu mengklasifikasikan data dengan tingkat ketepatan dan keakuratan yang memadai.

Tabel 14. Evaluasi Akurasi

Split Data	Matrix Confussion	Akurasi
85:15	$\begin{bmatrix} 87 & 4 & 2 \\ 9 & 55 & 9 \\ 12 & 4 & 61 \end{bmatrix}$	83.5%
90:10	$\begin{bmatrix} 57 & 3 & 3 \\ 7 & 37 & 5 \\ 7 & 3 & 41 \end{bmatrix}$	83.3%

Berdasarkan hasil perhitungan *precision*, *recall* dan akurasi pada model dengan rasio data 85:15 menunjukkan akurasi sebesar 83.5%, dengan performa terbaik pada kelas netral dengan *precision* 87,3%. *Precision* untuk kelas positif dan negatif masing-masing 80,5% dan 84,7%, menandakan prediksi cukup

akurat. Dari sisi *recall*, model paling baik mengenali kelas negatif 93,5%, sementara kelas positif dan netral masing-masing 79,2% dan 75,3%.

5. KESIMPULAN DAN SARAN

Berdasarkan hasil penerapan dan evaluasi sistem klasifikasi sentimen terhadap layanan *paylater* di platform media sosial X dengan pendekatan *Support Vector Machine* (SVM), dapat disimpulkan bahwa seluruh skenario pengujian menunjukkan kinerja yang konsisten yaitu rasio split 85:15 menghasilkan akurasi 83,5%, sementara rasio 90:10 memberikan akurasi 83,3%. Kinerja terbaik dicapai pada skenario 85:15. Proses klasifikasi mengungkap bahwa sentimen negatif mendominasi, tercermin dari nilai *recall* tertinggi sebesar 93,5%, yang mengindikasikan kekhawatiran publik terkait bunga, keterlambatan pembayaran, serta risiko lainnya. Untuk pengembangan lebih lanjut, disarankan agar sistem dilengkapi dengan filter *crawling* yang lebih menyeluruh dan mendukung berbagai ragam bahasa guna mencakup opini publik secara lebih luas dan representatif.

DAFTAR PUSTAKA

[1] L. Rofiah and M. A. Graciafernandy, "Persepsi dan Minat Penggunaan Aplikasi Paylater," *J. Ilm. Aset*, vol. 25, no. 1, pp. 61–69, 2023, doi: 10.37470/1.25.1.216.

[2] F. J. Wahidna and P. Nerisafitra, "Analisis Sentimen Pengguna Sistem Pay Later Menggunakan Support Vector Machine Metode Pembobotan Lexicon," *J. Informatics Comput. Sci.*, vol. 04, pp. 334–343, 2023, doi: 10.26740/jinacs.v4n03.p334-343.

[3] A. T. Fadhilah Gina, Syafina Laylan, "Impact of Social Environment, Financial Education, and M-Banking Features on Impulsive Buying Behavior Among Indonesian University," vol. 5, no. 4, pp. 28–43, 2024.

[4] H. Purwanto, D. Yandri, and M. P. Yoga, "Perkembangan Dan Dampak Financial Technology (Fintech) Terhadap Perilaku Manajemen Keuangan Di Masyarakat," *Kompleks. J. Ilm. Manajemen, Organ. Dan Bisnis*, vol. 11, no. 1, pp. 80–91, 2022, doi: 10.56486/kompleksitas.vol11no1.220.

[5] Alfandi Safira and F. N. Hasan, "Analisis Sentimen Masyarakat Terhadap Paylater Menggunakan Metode Naive Bayes Classifier," *Zo. J. Sist. Inf.*, vol. 5, no. 1, pp. 59–70, 2023, doi: 10.31849/zn.v5i1.12856.

[6] A. K. Santoso, "Analisis Sentimen Twitter Bahasa Indonesia Menggunakan Pendekatan Machine Learning," *J. Inform. Kaputama*, vol. 6, no. 2, pp. 129–136, 2022, doi: 10.59697/jik.v6i2.111.

[7] O. I. Gifari, M. Adha, F. Freddy, and F. F. S. Durrand, "Analisis Sentimen Review Film Menggunakan TF-IDF dan Support Vector Machine," *J. Inf. Technol.*, vol. 2, no. 1, pp. 36–40, 2022, doi: 10.46229/jifotech.v2i1.330.

[8] F. S. Pamungkas and I. Kharisudin, "Analisis Sentimen dengan SVM, NAIVE BAYES dan KNNuntuk Studi Tanggapan Masyarakat Indonesia TerhadapPandemi Covid-19 pada Media Sosial Twitter," *Pros. Semin. Nas. Mat.*, vol. 4, pp. 1–7, 2021, [Online]. Available: https://journal.unnes.ac.id/sju/prisma/article/view/45038

[9] R. Deswandi Yahya, S. Adi Wibowo, and N. Vendyansyah, "Analisis Sentimen Untuk Deteksi Ujaran Kebencian Pada Media Sosial Terkait Pemilu 2024 Menggunakan Metode
Support Vector Machine," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 8, no. 2, pp. 1182–1189, 2024, doi: 10.36040/jati.v8i2.9076.

[10] Boro, C. L. T., Faisol, A., Rudhistiar D, "ANALISIS SENTIMEN TERHADAP KAMPANYE PENGURANGAN PLASTIK PADA MEDIA SOSIAL MENGGUNAKAN METODE SVM," vol. 7, no. 1, pp. 173–178, 2025.

[11] M. M. Mala Olhang, S. Achmadi, and F. . A. Wibisono, "Analisis Sentimen Pengguna Twitter Terhadap Covid-19 Di Indonesia Menggunakan Metode Naive Bayes Classifier (Nbc)," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 4, no. 2, pp. 214–221, 2020, doi: 10.36040/jati.v4i2.2695.

[12] M. F. Rizki, K. Auliasari, and R. Primaswara Prasetya, "Analisis Sentiment Cyberbullying Pada Sosial Media Twitter Menggunakan Metode Support Vector Machine," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 5, no. 2, pp. 548–556, 2021, doi: 10.36040/jati.v5i2.3808.

[13] R. Tineges, A. Triayudi, and I. D. Sholihati, "Analisis Sentimen Terhadap Layanan Indihome Berdasarkan Twitter Dengan Metode Klasifikasi Support Vector Machine (SVM)," *J. Media Inform. Budidarma*, vol. 4, no. 3, p. 650, 2020, doi: 10.30865/mib.v4i3.2181.

[14] N. Fitriyah, B. Warsito, and D. A. I. Maruddani, "Analisis Sentimen Gojek Pada Media Sosial Twitter Dengan Klasifikasi Support Vector Machine (Svm," *J. Gaussian*, vol. 9, no. 3, pp. 376–390, 2020, doi: 10.14710/j.gauss.v9i3.28932.

[15] A. N. Fauziah, "Analisis sentimen menggunakan naïve bayes classifier dan inset lexicon pada twitter (studi kasus: Mie Gacoan)," UNIVERSITAS ISLAM NEGERI SYARIF HIDAYATULLAH JAKARTA, 2023.

[16] A. Kulkarni, D. Chong, and F. A. Batarseh, "Foundations of data imbalance and solutions for a data democracy," *Data Democr. Nexus Artif. Intell. Softw. Dev. Knowl. Eng.*, pp. 83–106, 2020, doi: 10.1016/B978-0-12-818366-3.00005-8.