

ANALISIS SENTIMEN ISU PENYALAHGUNAAN DATA PADA LAYANAN PINJAMAN ONLINE MENGGUNAKAN SUPPORT VECTOR MACHINE DI PLATFORM X

Ari Fullah^{1*}, Muhammad Arif Septian², Muhammad Rizky Haritama Putra³, Muhammad Iqbal Fadiatama⁴, Vashih Al Farizi⁵, Novita Puspitasari⁶, Ummul Hairah⁷

^{1,2,3,4,5,6,7}Universitas Mulawarman; Jalan Sambaliung No.9, Kota Samarinda 75119 Kalimantan Timur; telp. +(62) 811 555 1962, +(0541) 541 749 159

Keywords:

Analisis Sentimen;
Penyalahgunaan Data;
Pinjaman Online;
Support Vector Machine.

Correspondent Email:

arifullahtgr@gmail.com

Layanan pinjaman online di Indonesia menunjukkan peningkatan drastis dalam beberapa tahun terakhir, memberikan kemudahan dalam akses keuangan namun juga memunculkan permasalahan serius berupa penyalahgunaan data pribadi pengguna. Kondisi ini menimbulkan keresahan masyarakat dan menekankan pentingnya perlindungan data digital Platform media sosial, Khususnya X, menjadi ruang utama bagi masyarakat dalam menyampaikan opini dan keluhan terkait isu tersebut. Penelitian ini bertujuan menganalisis sentimen publik terhadap kasus penyalahgunaan data pada layanan pinjaman online untuk memahami persepsi masyarakat. Data dikumpulkan dari unggahan berbahasa Indonesia di Platform X dan diproses melalui tahap preprocessing meliputi pembersihan teks, tokenisasi, penghapusan stopword, normalisasi, stemming, serta pembobotan kata menggunakan TF-IDF. Dataset dibagi dengan metode train-test split rasio 80:20 menggunakan stratified sampling. Algoritma Support Vector Machine (SVM) digunakan sebagai model klasifikasi, sedangkan evaluasi dilakukan dengan akurasi, presisi, recall, dan F1-score. Hasil pengujian menunjukkan akurasi 93% dengan macro average F1-score 0,81 serta dominasi sentimen negatif.



Copyright © [JITET](#) (Jurnal Informatika dan Teknik Elektro Terapan). This article is an open access article distributed under terms and conditions of the Creative Commons Attribution (CC BY NC)

Online lending services in Indonesia have shown a significant increase in recent years, providing easier access to financial services but also raising serious issues in the form of misuse of users' personal data. This condition has caused public concern and emphasized the importance of digital data protection. Social media platforms, particularly X, have become the main space for people to express their opinions and complaints regarding this issue. This study aims to analyze public sentiment toward cases of personal data misuse in online lending services to better understand public perceptions. Data were collected from Indonesian-language posts on Platform X and processed through several preprocessing stages, including text cleaning, tokenization, stopword removal, normalization, stemming, and word weighting using TF-IDF. The dataset was split using an 80:20 train-test ratio with stratified sampling. The Support Vector Machine (SVM) algorithm was applied as the classification model, while evaluation was conducted using accuracy, precision, recall, and F1-score. The testing results showed an accuracy of 93% with a macro average F1-score of 0.81 and a dominance of negative sentiment.

1. PENDAHULUAN

Munculnya berbagai layanan berbasis teknologi, seperti pinjaman online, yang memungkinkan masyarakat mendapatkan dana dengan mudah, didorong oleh transformasi digital sektor keuangan [1]. Pertumbuhan ini didasarkan pada data Otoritas Jasa Keuangan (OJK), total *Outstanding* biaya pinjaman online (pinjol) pada Juni 2024, mencapai Rp 66,79 triliun, yang merupakan peningkatan tahunan sebesar 26,73% [2]. Namun pertumbuhan ini diikuti dengan peningkatan jumlah penyelenggara baik legal maupun ilegal. Laporan resmi OJK menerima 1.081 aduan baru terkait pinjol dari Januari hingga Maret 2025; sebagian besar aduan berasal dari layanan pinjol ilegal yang tidak memiliki izin resmi dari OJK [3].

Salah satu permasalahan utama dalam pelaksanaan layanan pinjaman online, baik yang terdaftar secara resmi maupun yang tidak, adalah penyalahgunaan atau penyebaran data pribadi pengguna [4]. Berbagai laporan mengungkap praktik penyebaran informasi sensitif, seperti nomor telepon identitas, maupun kontak darurat, yang sering kali berujung pada teror dari penagih utang hingga pencemaran nama baik. Fenomena ini menimbulkan dampak yang tidak hanya bersifat material, tetapi juga sosial dan psikologis bagi korban. Secara regulasi, praktik ini melanggar Undang-undang (UU) Nomor 27 Tahun 2022 tentang Perlindungan Data Pribadi [5], yang bertujuan menjaga kerahasiaan dan keamanan data konsumen. Tidak heran jika kemudian keresahan masyarakat kerap muncul dan mencari ruang untuk diekspresikan, salah satunya melalui media sosial yang menjadi kanal utama diskusi publik.

Perkembangan media sosial memberikan ruang bagi masyarakat untuk mengekspresikan opini masyarakat terkait isu pinjol. Platform X menjadi salah satu kanal yang banyak dimanfaatkan, mengingat karakteristiknya yang cepat, terbuka, dan mencerminkan dinamika opini publik. Fenomena ini menjadikan analisis sentimen publik di media sosial sangat penting untuk memahami cara masyarakat berpikir tentang bagaimana hal ini dapat mempengaruhi persepsi masyarakat secara keseluruhan [6].

Dalam konteks ini, metode klasifikasi berbasis algoritma pembelajaran mesin, seperti *Support Vector Machine* (SVM), menjadi

relevan karena kemampuannya menangani data teks pendek dengan tingkat akurasi tinggi. Penelitian sebelumnya menggunakan SVM untuk analisis sentimen aplikasi pinjol di Google Play Store, menunjukkan mayoritas ulasan negatif, dengan Kredivo paling banyak sentimen positif dan Easycash paling banyak negatif. Model SVM mencapai akurasi 72%, presisi 76%, dan *recall* 85%, bahkan 83% pada aplikasi Kredit Pintar [7]. Penelitian lain mengenai Analisis Sentimen Terhadap Penggunaan Aplikasi Shopee Menggunakan Algoritma Support Vector Machine (SVM) juga membuktikan kinerja yang cukup baik, dengan hasil akurasi sebesar 98% dan *f1-score* sebesar 0.98 atau 98% [8]. Namun, penelitian-penelitian tersebut masih fokus pada kualitas aplikasi secara umum, sedangkan isu penyalahgunaan data pribadi belum banyak diteliti, padahal relevansinya meningkat seiring banyaknya pengaduan publik.

Kondisi ini menimbulkan pertanyaan mengenai bagaimana masyarakat benar-benar memandang kasus penyalahgunaan data pada pinjol, terutama di ruang diskusi publik seperti media sosial. Pemetaan opini semacam itu dapat memberikan gambaran lebih luas mengenai persepsi dan keresahan masyarakat, sekaligus menjadi refleksi sejauh mana kebijakan dan regulasi yang ada telah menjawab persoalan yang timbul.

Berdasarkan uraian tersebut, penelitian ini bertujuan untuk menganalisis sentimen warganet di Platform X terhadap isu penyalahgunaan data dalam layanan pinjaman online di Indonesia, sekaligus menerapkan algoritma SVM untuk mengklasifikasikan sentimen menjadi kategori positif dan negatif. Penelitian ini juga mencakup pembuatan visualisasi *Word Cloud* untuk menampilkan kata-kata yang paling sering muncul dalam komentar masyarakat. Selain itu, kinerja SVM akan dievaluasi menggunakan akurasi, presisi, *recall*, dan *F1-score*. Hasil penelitian diharapkan dapat memberikan masukan bagi pihak-pihak terkait dalam memahami opini publik serta memperkuat kebijakan perlindungan data di sektor pinjol.

2. TINJAUAN PUSTAKA

2.1. Analisis Sentimen

Analisis sentimen, atau yang kerap disebut penambangan opini, merupakan proses untuk memahami, mengekstraksi, serta mengolah data teks secara otomatis guna memperoleh informasi mengenai perasaan atau sikap yang terkandung dalam suatu pendapat[9]. Metode ini merupakan metode untuk mengenali serta mengelompokkan opini yang terdapat dalam suatu teks, apakah bersifat positif, negatif, maupun netral [10]. Di sisi lain, analisis sentimen juga dapat dimanfaatkan untuk menilai kinerja, layanan, dan aspek lainnya.

2.2. Platform X

Twitter, sekarang dikenal sebagai X, adalah platform sosial media dengan banyak pengguna di seluruh dunia yang dapat berbagi informasi dalam bentuk teks, gambar, dan pesan suara dengan orang lain dan dengan publik [11].

2.3. Support Vector Machine

SVM, juga dikenal sebagai prinsip minimisasi risiko struktural, adalah algoritma umum yang didasarkan pada batasan risiko terjamin dari teori pembelajaran statistik [10]. SVM berfungsi untuk menemukan fungsi pemisah terbaik antara kelas. Fungsi pemisah terbaik adalah yang menghasilkan nilai margin terbesar antara dua vektor dari dua kelas yang berbeda dan berada di tengah-tengah kedua vektor tersebut [12].

2.4. TF-IDF

TF-IDF merupakan metode untuk memperoleh informasi yang mengutamakan frekuensi istilah dan kebalikannya dalam dokumen (IDF). Setiap kata atau frasa memiliki nilai TF dan IDF, di mana bobot dari istilah tersebut disebut sebagai *TF-IDF*. Algoritma *TF-IDF* berfungsi untuk memberikan nilai pada kata kunci dalam setiap dokumen serta menghitung frekuensi kemunculannya di dalam dokumen tersebut [13].

2.5. Word Cloud

Metode *text mining* yang dikenal sebagai *word cloud* menampilkan grafik frekuensi kata yang menonjolkan kata-kata yang lebih sering ditemukan dalam teks sumber [14]. Kata yang paling sering muncul akan ditampilkan dengan ukuran lebih besar, sedangkan kata dengan frekuensi rendah akan ditampilkan lebih kecil.

2.6. Preprocessing Data

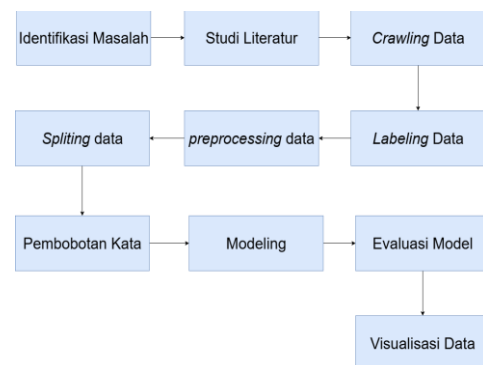
Tujuan dari tahap *preprocessing* data adalah untuk menyederhanakan pengolahan data sehingga hasil penilaian yang diperoleh dapat tersusun dengan lebih sistematis [15].

2.7. Confusion Matrix

Confusion matrix merupakan tabel yang menunjukkan jumlah data uji yang terklasifikasi dengan benar serta jumlah data uji yang terklasifikasi dengan salah [16]. *Confusion matrix* menjadi acuan dalam perhitungan metrik evaluasi seperti akurasi, *precision*, *recall*, dan *F1-score*.

3. METODE PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif dengan metode eksperimen komputasional. Metode ini dipilih untuk melakukan analisis sentimen terhadap isu penyalahgunaan data pada layanan pinjaman online dengan memanfaatkan algoritma SVM sebagai metode klasifikasi utama. Data penelitian diperoleh dari Platform X dalam bentuk opini publik yang dikumpulkan selama periode 1 Januari 2023 hingga 30 Agustus 2025. Tahapan penelitian ini meliputi identifikasi masalah, studi literatur, *crawling* data, *labeling* data, *preprocessing* data, *splitting* data, pembobotan kata, *modeling*, evaluasi model, dan visualisasi data. Alur tahapan penelitian tersebut dapat dilihat pada Gambar 1.



Gambar 1. Tahapan Penelitian

3.1. Identifikasi Masalah

Tahap ini bertujuan untuk merumuskan permasalahan utama penelitian, yaitu adanya isu penyalahgunaan data pada layanan pinjol yang menimbulkan berbagai reaksi pada masyarakat. Dari permasalahan tersebut, ditetapkan tujuan penelitian untuk menganalisis sentimen publik menggunakan pendekatan komputasional.

3.2. Studi Literatur

Tahap ini dilakukan untuk mengumpulkan teori, konsep, serta penelitian

terdahulu yang relevan. Literatur yang digunakan mencakup kajian mengenai analisis sentimen, algoritma SVM, serta tren penelitian terkait isu penyalahgunaan data. Hasil studi literatur menjadi dasar dalam merancang alur penelitian.

3.3. *Crawling Data*

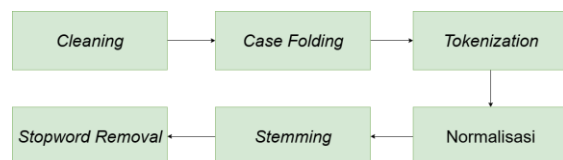
Data diperoleh dari platform X dengan teknik web scraping dengan pustaka Python Tweet Harvest. Unggahan yang diambil adalah teks berbahasa Indonesia yang mengandung kata kunci seperti “pinjaman online”, “penyalahgunaan data”, dan istilah sejenis. Data yang dikumpulkan mencakup isi *tweet*, tanggal unggahan, serta metadata yang relevan.

3.4. *Labeling Data*

Data yang diperoleh kemudian dilakukan pemberian label sentimen menjadi dua kategori, yaitu positif dan negatif. Proses pelabelan dilakukan secara manual oleh peneliti berdasarkan isi teks, dengan memperhatikan konteks kalimat agar sesuai dengan makna sebenarnya.

3.5. *Preprocessing Data*

Proses ini meliputi *cleaning*, *case folding*, *tokenization*, normalisasi, *Stemming*, dan *Stopword removal*. Alur dari tahapan *preprocessing* dapat dilihat pada Gambar 2.



Gambar 2. Tahapan *Preprocessing*

3.5.1. *Cleaning*

Pada tahap *cleaning*, dilakukan penghapusan elemen yang tidak relevan, seperti mention, URL, dan karakter khusus, sehingga data menjadi lebih bersih dan siap untuk diproses.

3.5.2. *Case Folding*

Pada tahap *case folding*, seluruh teks dikonversi menjadi huruf kecil untuk menjaga konsistensi serta memudahkan proses analisis. Dengan langkah ini, perbedaan antara huruf kapital dan nonkapital tidak lagi memengaruhi hasil pengolahan data.

3.5.3. *Tokenization*

Pada tahap *tokenization*, suatu kalimat diuraikan menjadi potongan kata atau token

agar lebih mudah dianalisis. Proses ini memungkinkan setiap kata diperlakukan sebagai unit data yang berdiri sendiri.

3.5.4. *Normalisasi*

Pada tahap normalisasi, kata yang tidak baku diubah menjadi bentuk baku berdasarkan kamus yang digunakan. Tahap ini bertujuan untuk menyeragamkan variasi penulisan sehingga data lebih konsisten.

3.5.5. *Stemming*

Pada tahap *stemming*, setiap kata dikembalikan ke bentuk dasarnya dengan cara menghapus imbuhan, baik awalan, sisipan, maupun akhiran. Tahap ini bertujuan agar kata dengan makna yang sama dapat diperlakukan sebagai satu kesatuan. Proses *stemming* dilakukan dengan bantuan pustaka NLTK.

3.5.6. *Stopword Removal*

Pada tahap *stopword removal*, kata-kata umum yang tidak memiliki makna signifikan, seperti “dan”, “di”, atau “yang”, dihapus agar tidak mengganggu proses analisis. Tahap ini dilakukan dengan menggunakan daftar *stopword* dari pustaka NLTK.

3.6. *Splitting Data*

Splitting data adalah proses membagi dataset menjadi dua atau lebih bagian terpisah [17]. Pada penelitian ini, dataset yang telah melalui tahap *preprocessing* dibagi menjadi data latih dan data uji. Data latih digunakan untuk membangun model klasifikasi sentimen sedangkan data uji digunakan untuk mengukur performa model terhadap data yang belum pernah dilihat sebelumnya.

3.7. *Pembobotan Kata*

Data teks yang telah diproses kemudian diubah ke dalam bentuk numerik menggunakan metode *TF-IDF*. Representasi ini memungkinkan model untuk mengukur tingkat kepentingan suatu kata dalam sebuah dokumen.

3.8. *Modeling*

Tahap ini menggunakan SVM sebagai metode klasifikasi utama untuk membedakan sentimen menjadi kategori positif dan negatif. Proses pemodelan dilakukan pada data latih.

3.9. *Evaluasi Model*

Evaluasi kinerja model dilakukan menggunakan *confusion matrix* yang memuat metrik akurasi, *precision*, *recall*, dan *F1-Score*. Metrik ini dipilih untuk memberikan gambaran komprehensif mengenai kemampuan model dalam mengklasifikasikan data sentimen.

3.10. *Visualisasi Data*

Hasil analisis ditampilkan dalam bentuk *word cloud* untuk memberikan representasi visual terhadap kata-kata yang paling sering muncul pada dataset.

4. HASIL DAN PEMBAHASAN

Hasil dan pembahasan pada penelitian ini disajikan melalui beberapa tahapan mulai dari proses pengumpulan hingga pengolahan data. Untuk mempermudah pemahaman, tabel yang ditampilkan hanya memuat sampel dari keseluruhan data. Dengan cara ini, pembaca tetap dapat melihat gambaran nyata dari bentuk data pada setiap proses tanpa harus menampilkan seluruh dataset yang jumlahnya sangat besar.

4.1. Crawling Data

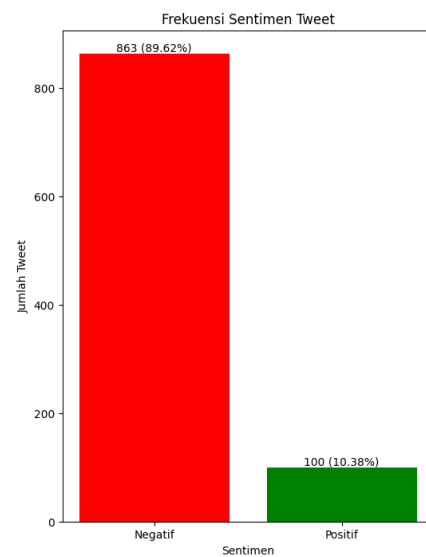
Pada tahap pertama, sebanyak 969 unggahan publik di Platform X dikumpulkan untuk pengecekan data, hasilnya menunjukkan bahwa ada data duplikat dan data kosong pada atribut `full_text`, sehingga total data menjadi 963. Sampel hasil *crawling* data ditampilkan pada Tabel 1.

Tabel 1. Sampel hasil *crawling* data

full_text
@250312052208we @tanyakanrl btw kalo ambil pinjol dan udah lunas. tapi nomer masih kesebar itu namanya penyalahgunaan data bkn sih? jadi ada tmn yg nomernya disebar sekalian sama nik foto dan alamat juga. ga sekali dua kali.. tapi berkali² disebar. kenapa ya..
Modus penipuan baru unlocked Kali ini modusnya adalah penyalahgunaan data seseorang untuk mengajukan pinjaman online/pinjol ilegal.

4.2. Labeling Data

Hasil labeling manual menunjukkan bahwa 100 data berlabel sentimen positif dan 863 data berlabel sentimen negatif, sebagaimana ditunjukkan pada Gambar 3.



Gambar 3. Hasil Labeling

4.3. Preprocessing Data

4.3.1. Cleaning

Sampel hasil *cleaning* dapat dilihat pada Tabel 2.

Tabel 2. Sampel hasil *cleaning*

full_text	cleaned
@grok @sugarmaker__ @Pendekar969 @kompascom Aaah ribet amat hidup baru ngurusin nomor aja ribetnya minta ampun ngurusin kebocoran data masyarakat bertahun² data disalahgunakan utk teror pinjol dll sampai ada yg bunuh diri ga selesai² sampai skrg ... Ngemeng aja selangit mau ditertibkan hingga detik ini mana??	Aaah ribet amat hidup baru ngurusin nomor aja ribetnya minta ampun ngurusin kebocoran data masyarakat bertahun² data disalahgunakan utk teror pinjol dll sampai ada yg bunuh diri ga selesai² sampai skrg Ngemeng aja selangit mau ditertibkan hingga detik ini mana
@_nuyyyy @worksfess Bener kak apalagi kemaren ada kasus	Bener kak apalagi kemaren ada kasus penyalahgunaan data buat pinjol Makin

penyalahgunaan data buat pinjol. Makin was was	was was
--	---------

4.3.2. Case Folding

Sampel hasil case folding ditampilkan pada Tabel 3.

Tabel 3. Sampel hasil *case folding*

cleaned	casefolded
Abis baca tweet yang dia korban penyalahgunaan data pribadi yang dipake pinjol Sumpah ngeri bangeett pinjol gini tolong dibasmi semua	abis baca tweet yang dia korban penyalahgunaan data pribadi yang dipake pinjol sumpah ngeri bangeett pinjol gini tolong dibasmi semua
Bahaya Pinjaman Online Bunga tinggi penyalahgunaan data stres mental dan siklus utang Jangan sampai terjebak	bahaya pinjaman online bunga tinggi penyalahgunaan data stres mental dan siklus utang jangan sampai terjebak

4.3.3. Tokenization

Sampel hasil *tokenization* ditampilkan pada Tabel 4.

Tabel 4. Sampel hasil *tokenization*

casefolded	tokens
waspada penyalahgunaan data pribadi untuk pinjaman online	['waspada', 'penyalahgunaan', 'data', 'pribadi', 'untuk', 'pinjaman', 'online']
mengapa pinjol akan diperketat atau ditutup maraknya pinjol ilegal banyak pinjol ilegal beroperasi tanpa izin menjerat masyarakat dengan bunga tinggi	['mengapa', 'pinjol', 'akan', 'diperketat', 'atau', 'ditutup', 'maraknya', 'pinjol', 'ilegal', 'banyak', 'pinjol', 'ilegal', 'beroperasi', 'tanpa', 'izin', 'menjerat',

penagihan tidak beretika dan penyalahgunaan data pribadi langkah penutupan diarahkan untuk menghentikan aktivitas ilegal	'masyarakat', 'dengan', 'bunga', 'tinggi', 'penagihan', 'tidak', 'beretika', 'dan', 'penyalahgunaan', 'data', 'pribadi', 'langkah', 'penutupan', 'diarahkan', 'untuk', 'menghentikan', 'aktivitas', 'ilegal']
--	---

4.3.4. Normalisasi

Sampel hasil normalisasi ditampilkan pada Tabel 5.

Tabel 5. Sampel hasil normalisasi

tokens	normalized
['efek', 'samping', 'pinjol', 'data', 'kalian', 'mudah', 'diketahui', 'sudah', 'tidak', 'privacy', 'rawan', 'scam', 'dan', 'penyalahgunaan', 'data', 'sebar', 'data', 'yg', 'dilakukan', 'oknum', 'dc', 'pinjol', 'masih', 'mau', 'pakai', 'pinjol', 'sayangilah', 'data', 'kalian', 'kalo', 'gw', 'nih', 'lebih', 'baik', 'miskin', 'daripada', 'pakai', 'pinjol', 'hidup', 'lebih', 'tenang']	efek samping pinjaman online data kalian mudah diketahui sudah tidak privasi rawan scam dan penyalahgunaan data sebar data yang dilakukan oknum dc pinjaman online masih mau pakai pinjaman online sayangilah data kalian kalau gue nih lebih baik miskin daripada pakai pinjaman online hidup lebih tenang
['kalau', 'buat', 'pinjol', 'bukannya', 'penyalahgunaan', 'data', 'pribadi', 'yak', 'main', 'aman', 'ajaaa']	kalau buat pinjaman online bukannya penyalahgunaan data pribadi ya main aman saja

4.3.5. Stemming

Sampel hasil *stemming* ditampilkan pada Tabel 6.

Tabel 6. Sampel hasil *stemming*

normalized	stemmed
pinjaman online ilegal bisa dengan mudah mengeksploitasi data pribadi kita mulai dari penyalahgunaan data hingga teror pinjaman	pinjam online ilegal bisa dengan mudah eksploitasi data pribadi kita mulai dari penyalahgunaan data hingga teror pinjam
kita perlu ketahui bahwa joki pinjaman online mendatangkan banyak bahaya jika digunakan yaitu penyalahgunaan data pribadi risiko penipuan dan tarif joki yang mahal bahkan sebenarnya joki pinjaman online tidak memiliki kontrol atas pinjaman yang kamu miliki	kita perlu tahu bahwa joki pinjam online datang banyak bahaya jika guna yaitu penyalahgunaan data pribadi risiko tipu dan tarif joki yang mahal bahkan benar joki pinjam online tidak milik kontrol atas pinjam yang kamu milik

4.3.6. Stopword Removal

Sampel hasil *stopword removal* ditampilkan pada Tabel 7.

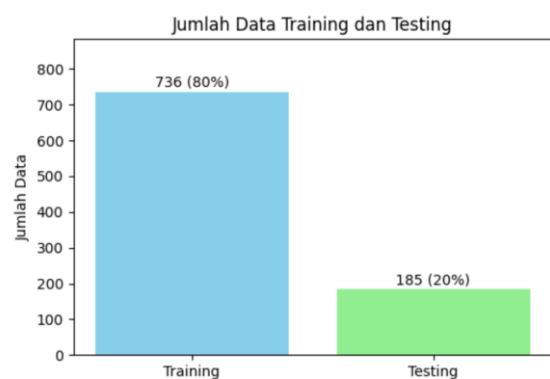
Tabel 7. Sampel hasil *stopword removal*

stemmed	stopword_removal
aksi jahat penyalahgunaan data lalu platform pinjam online ilegal di internet makin marak jadi nih	aksi jahat penyalahgunaan data platform pinjam online ilegal internet marak nih
manis rayu pinjam online ilegal dan jerat penyalahgunaan data pribadi	manis rayu pinjam online ilegal jerat penyalahgunaan data pribadi

4.4. Splitting Data

Setelah dilakukan tahap preprocessing terdapat data yang duplikat sehingga total data dari 963 menjadi 921. Tahap berikutnya setelah proses preprocessing adalah membagi dataset menjadi dua bagian, yaitu data latih dan data uji. Pembagian ini dilakukan menggunakan metode *train-test split* dengan rasio 80 persen untuk data latih dan 20 persen untuk data uji. Teknik *stratified sampling* digunakan agar distribusi jumlah data pada kedua kelas, yaitu *Negatif* dan *Positif*, tetap seimbang baik pada data latih maupun data uji. Selain itu, nilai *random state* ditetapkan sebesar 42 agar hasil pembagian data bersifat konsisten dan dapat direproduksi pada saat penelitian diulang.

Dari total 921 data yang digunakan dalam penelitian ini, sebanyak 736 data (80%) ditetapkan sebagai data latih untuk melatih model, sementara 185 data (20%) dijadikan data uji untuk mengevaluasi performa model. Meskipun target pembagian adalah 80:20, hasil aktual menunjukkan persentase 80% untuk training dan 20% untuk testing, yang masih sesuai dengan proporsi yang diharapkan. Untuk memperjelas pembagian tersebut, dibuat visualisasi dalam bentuk diagram batang yang menggambarkan jumlah serta persentase data latih dan data uji. Diagram batang jumlah data *training* dan *testing* dapat dilihat pada Gambar 4.

**Gambar 4.** Persentase data Training dan data Testing.**4.5. Pembobotan Kata**

Proses ini bertujuan untuk membagi setiap kata berdasarkan tingkat kepentingannya

dalam dokumen. 20 kata dengan nilai *TF-IDF* tertinggi, berikut adalah hasil pembobotan, seperti yang ditunjukkan pada Gambar 5.

	fitur	rata_rata_bobot
191	data	0.074747
698	pakai	0.051128
764	pribadi	0.043294
121	bocor	0.042848
223	data pribadi	0.040460
609	ojk	0.039503
709	pakai pinjam	0.036186
348	ilegal	0.035915
639	online ilegal	0.031752
123	bocor data	0.030579
994	ya	0.029305
729	penyalahgunaan	0.028930
677	orang	0.028520
730	penyalahgunaan data	0.026806
115	blokir	0.022773
220	data pakai	0.022607
472	ktp	0.021668
70	banget	0.020621
955	tutup	0.019405
443	kerja	0.018597

Gambar 5. 20 kata nilai *TF-IDF* tertinggi

4.6. Modeling

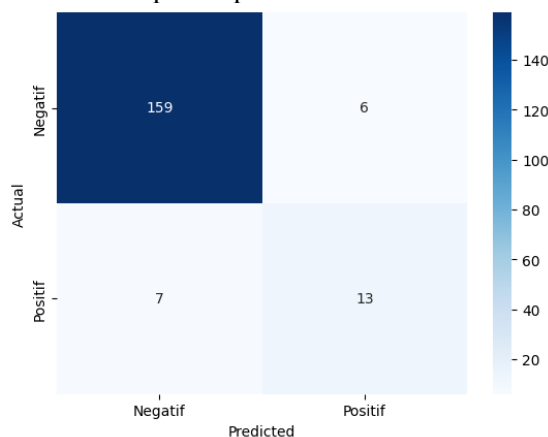
Implementasi kode program untuk tahap pemodelan dengan algoritma SVM ditampilkan pada Gambar 6.

```
# === 2. Model ===
model = LinearSVC(C=1.0, class_weight="balanced", random_state=42)
```

Gambar 6. Implementasi SVM

4.7. Evaluasi Model

Hasil dari proses pengujian menunjukkan bahwa penelitian ini menghasilkan *confusion matrix* ditampilkan pada Gambar 7.



Gambar 7. *Confusion matrix*

Hasil perhitungan *confusion matrix* menghasilkan metrik akurasi, *recall*, *precision*, serta *f1-score* seperti terlihat pada Gambar 8.

	precision	recall	f1-score	support
Negatif	0.96	0.96	0.96	165
Positif	0.68	0.65	0.67	20
accuracy			0.93	185
macro avg	0.82	0.81	0.81	185
weighted avg	0.93	0.93	0.93	185

Gambar 8. *classification report*

Berdasarkan *classification report* pada Gambar 7, model mampu mengklasifikasikan sentimen negatif dengan nilai *precision*, *recall*, dan *f1-score* sebesar 0,96. Namun, pada sentimen positif, kinerja model menurun dengan nilai *precision* sebesar 0,68, *recall* 0,65, dan *f1-score* 0,67. Secara keseluruhan, model menghasilkan akurasi sebesar 93% dengan nilai *macro average f1-score* sebesar 0,81. Tingginya kinerja pada kelas negatif dipengaruhi oleh dominasi data berlabel negatif dalam *dataset*, sedangkan kinerja pada kelas positif relatif lebih rendah karena jumlah data positif yang lebih sedikit. Ketidakseimbangan ini muncul karena di Platform X mayoritas unggahan pengguna cenderung mengekspresikan sentimen negatif, sehingga model lebih terlatih pada pola teks negatif dibandingkan positif.

4.8. Visualisasi Data

Hasil *word cloud* ditampilkan pada Gambar 9.



Gambar 9. Hasil *word cloud*

Berdasarkan visualisasi *word cloud* pada Gambar 8, kata yang paling dominan adalah “pinjam”, “data”, “online”, “pakai”, dan “pribadi”. Temuan ini menunjukkan bahwa ulasan pengguna banyak berfokus pada isu peminjaman, data pribadi, serta penggunaan layanan secara daring. Selain itu, kata “bocor”, “penyalahgunaan”, dan “ilegal” juga muncul dengan ukuran cukup besar, yang mengindikasikan adanya kekhawatiran masyarakat terhadap risiko penyalahgunaan data dan praktik pinjaman online ilegal.

5. KESIMPULAN

- a. Penelitian ini berhasil menganalisis sentimen publik terhadap kasus penyalahgunaan data pada layanan pinjaman online dengan menggunakan algoritma Support Vector Machine (SVM), menghasilkan akurasi 93% dan macro average F1-score sebesar 0,81.
- b. Hasil analisis menunjukkan dominasi sentimen negatif, yang mengindikasikan keresahan masyarakat terhadap isu keamanan data pribadi pada layanan pinjaman online.
- c. Kelebihan penelitian ini terletak pada proses preprocessing teks yang komprehensif (pembersihan teks, tokenisasi, stopword removal, normalisasi, stemming, serta TF-IDF) dan pemilihan algoritma SVM yang terbukti efektif dalam klasifikasi teks.
- d. Kekurangan penelitian ini adalah keterbatasan data yang hanya bersumber dari Platform X, sehingga belum sepenuhnya merepresentasikan opini masyarakat secara luas di berbagai media sosial.
- e. Untuk pengembangan selanjutnya, penelitian dapat diperluas dengan menggunakan lebih banyak sumber data, membandingkan kinerja dengan algoritma lain seperti Random Forest atau LSTM, serta menambahkan analisis temporal untuk melihat perubahan sentimen dari waktu ke waktu.

UCAPAN TERIMA KASIH

Ucapan terima kasih disampaikan kepada semua pihak yang telah memberikan dukungan dalam pelaksanaan penelitian ini.

DAFTAR PUSTAKA

- [1] R. Sungangga and E. H. Sentoso, "Perlindungan hukum terhadap pengguna pinjaman online (pinjol) ilegal," PAJOUL (Pakuan Justice Journal of Law), vol. 1, no. 1, pp. 306–312, 2024, doi: 10.33751/pajoul.v1i1.2050.
- [2] F. Saputra, "Outstanding pembiayaan fintech P2P lending capai Rp 66,79 triliun pada Juni 2024." [Online]. Available: <https://keuangan.kontan.co.id/news/outstanding-pembiayaan-fintech-p2p-lending-capai-rp-6679-triliun-pada-juni-2024>. Accessed: Aug. 25, 2025.
- [3] Badan Pembinaan Hukum Nasional, "Pinjaman online ilegal kian marak, berikut proses hukum hingga pencegahannya." [Online]. Available: <https://bphn.go.id/berita-utama/pinjaman-online-ilegal-kian-marak-berikut-proses-hukum-hingga-pencegahannya>. Accessed: Aug. 25, 2025.
- [4] E. S. Darmayanti and S. A. Wiraguna, "Tanggung jawab hukum pinjaman online terhadap penyebaran data nasabah secara ilegal," ALADALAH: Jurnal Politik, Sosial, Hukum dan Humaniora, vol. 3, no. 2, pp. 233–251, May 2025, doi: 10.59246/aladalah.v3i2.1313.
- [5] "Undang-undang (UU) Nomor 27 Tahun 2022 tentang Pelindungan Data Pribadi." [Online]. Available: <https://peraturan.bpk.go.id/Details/229798/uu-no-27-tahun-2022>. Accessed: Aug. 27, 2025.
- [6] R. N. Muhammad, L. W. S. Wulandari, and B. Tanggahma, "Pengaruh media sosial pada persepsi publik terhadap sistem peradilan: analisis sentimen di Twitter," UNES Law Review, vol. 7, no. 1, pp. 507–516, 2024, doi: 10.31933/unesrev.v7i1.2327.
- [7] M. Iqbal, M. Afdal, and R. Novita, "Implementasi algoritma support vector machine untuk analisa sentimen data ulasan aplikasi pinjaman online di Google Play Store," MALCOM: Indonesian Journal of Machine Learning and Computer Science, vol. 4, no. 4, pp. 1244–1252, Jul. 2024, doi: 10.57152/malcom.v4i4.1435.
- [8] I. S. K. Idris, Y. A. Mustofa, and I. A. Salihi, "Analisis sentimen terhadap penggunaan aplikasi Shopee menggunakan algoritma support vector machine (SVM)," Jambura Journal of Electrical and Electronics Engineering, vol. 5, no. 1, pp. 32–35, Jan. 2023.
- [9] A. S. Rusydiana and L. Marlina, "Analisis sentimen terkait sertifikasi halal," JEBA (Journal of Economics and Business

- Aseanomics), vol. 5, no. 1, pp. 69–85, Jul. 2020, doi: 10.33476/j.e.b.a.v5i1.1405.
- [10] R. A. P. Wahyu and A. I. Mahendra, "Penggunaan media sosial X pada akun @Txdrrberseragam untuk mengetahui tingkat kepercayaan masyarakat Yogyakarta terhadap Kepolisian RI," MUKASI: Jurnal Ilmu Komunikasi, vol. 3, no. 3, pp. 236–248, Aug. 2024, doi: 10.54259/mukasi.v3i3.3005.
- [11] H. S. W. Hovi, A. I. Hadiana, and F. R. Umbara, "Prediksi penyakit diabetes menggunakan algoritma support vector machine (SVM)," Informatics and Digital Expert (INDEX), vol. 4, no. 1, pp. 40–45, Jan. 2023, doi: 10.36423/index.v4i1.895.
- [12] D. Irawan, E. B. Perkasa, Y. Yurindra, D. Wahyuningsih, and E. Helmud, "Perbandingan klasifikasi SMS berbasis support vector machine, naive Bayes classifier, random forest dan bagging classifier," Jurnal Sisfokom (Sistem Informasi dan Komputer), vol. 10, no. 3, pp. 432–437, 2021, doi: 10.32736/sisfokom.v10i3.1302.
- [13] J. Supriyanto, D. Alita, and A. R. Isnain, "Penerapan algoritma k-nearest neighbor (K-NN) untuk analisis sentimen publik terhadap pembelajaran daring," Jurnal Informatika dan Rekayasa Perangkat Lunak, vol. 4, no. 1, pp. 74–80, Mar. 2023, doi: 10.33365/jatika.v4i1.2468.
- [14] J. J. A. Limbong, I. Sembiring, and K. D. Hartomo, "Analisis klasifikasi sentimen ulasan pada e-commerce Shopee berbasis word cloud dengan metode naive Bayes dan k-nearest neighbor," Jurnal Teknologi Informasi dan Ilmu Komputer, vol. 9, no. 2, p. 347, Feb. 2022, doi: 10.25126/jtiik.2022924960.
- [15] Z. Rani and B. K. Khotimah, "Analisis sentimen publik terhadap karapan sapi di Twitter menggunakan kombinasi metode k-means dan support vector machine," Jurnal Informatika dan Teknik Elektro Terapan, vol. 13, no. 1, Jan. 2025, doi: 10.23960/jitet.v13i1.5685.
- [16] D. Normawati and S. A. Prayogi, "Implementasi Naïve Bayes Classifier dan Confusion Matrix pada Analisis Sentimen Berbasis Teks pada Twitter," J. Sains Komput. Inform. (J-SAKTI), vol. 5, no. 2, pp. 697–711, 2021.
- [17] R. Merdiansah and A. A. Ridha, "Analisis sentimen pengguna X Indonesia terkait kendaraan listrik menggunakan IndoBERT," Jurnal Ilmu Komputer dan Sistem Informasi (JIKOMSI), vol. 7, no. 1, pp. 221–228, 2024.