

Analisis Sentimen Pengguna Aplikasi Pinjaman Online Melalui Media Sosial X (Twitter) Menggunakan Metode *Support Vector Machine*

^{1,*}Nelsi Silaban, ²Marliza Ganefi Gumay

^{1,*2} Sistem Informasi Bisnis, Program Pascasarjana Magister Manajemen Sistem Informasi
Universitas Gunadarma, Jakarta, Indonesia

^{1,*} silabannelsi@gmail.com, ²marliza@staff.gunadarma.ac.id

Submit : 03 Sept 25 | Diterima : 25 Sept 2025 | Terbit : 07 Oktober 2025

ABSTRAK

Perkembangan layanan pinjaman *online* sebagai bagian dari inovasi jasa keuangan digital telah mempermudah akses masyarakat terhadap layanan keuangan, meskipun juga menimbulkan beragam pandangan publik yang perlu dianalisis secara mendalam. Penelitian ini bertujuan untuk menganalisis sentimen pengguna aplikasi pinjaman *online* melalui platform media sosial X (*Twitter*) dengan menerapkan algoritma *Support Vector Machine* (SVM). *Dataset* penelitian terdiri dari 5.233 ulasan berbahasa Indonesia yang dikumpulkan antara Januari 2020 hingga Desember 2023 melalui teknik *scraping*. Tahapan penelitian menggunakan pendekatan *text mining* meliputi pengumpulan data, pra-pemrosesan (penerapan kamus lexicon, normalisasi, penghapusan emoji dan tanda baca, tokenisasi, penghapusan kata umum, serta pengembalian kata dasar), pelabelan data, pembagian data menjadi data latih dan data uji, serta ekstraksi fitur dengan metode *Term Frequency-Inverse Document Frequency* (TF-IDF). Hasil evaluasi menunjukkan bahwa model SVM mencapai akurasi sebesar 86%, dengan nilai presisi tertinggi pada kategori sentimen netral (98%) dan nilai *recall* tertinggi pada kategori sentimen positif (90%). Visualisasi berupa diagram batang dan WordCloud memperlihatkan distribusi 887 ulasan pada data uji, terdiri atas 41% negatif, 38% positif, dan 21% netral. Hasil ini menunjukkan bahwa meskipun terdapat apresiasi terhadap aspek kemudahan, kecepatan, dan keamanan layanan, namun kritik terkait tingginya bunga pinjaman, praktik penagihan yang agresif, serta isu keamanan data masih mendominasi. Hasil penelitian ini diharapkan memberikan kontribusi bagi pemahaman yang lebih komprehensif mengenai persepsi masyarakat terhadap layanan pinjaman *online*.

Kata kunci : Analisis Sentimen, Pinjaman *Online*, Media Sosial X (*Twitter*), *Support Vector Machine* (SVM)

PENDAHULUAN

Layanan pinjaman *online*, merupakan salah satu inovasi penting dalam sektor jasa keuangan berbasis digital. Proses transaksi dalam layanan ini tidak lagi memerlukan mekanisme manual maupun dokumen fisik, melainkan cukup dilakukan melalui *smartphone* dan jaringan internet. Berdasarkan data *World Bank* yang dikutip oleh sikapiuangmu.ojk.go.id, penggunaan layanan *fintech* di Indonesia menunjukkan tren pertumbuhan signifikan, yaitu dari 7% pada tahun 2007 menjadi 20% pada 2011, meningkat menjadi 36% pada 2014, dan pada tahun 2017 telah mencapai 78%. Pada periode yang sama, jumlah perusahaan *fintech* tercatat sekitar 135-140 dengan nilai transaksi mencapai Rp202,77 triliun.

Pertumbuhan industri ini juga menarik minat investor, baik domestik maupun internasional. Menurut laporan katadata.co.id, sejumlah *startup fintech* memperoleh investasi signifikan, dengan mayoritas investor berasal dari Indonesia dan Singapura. Selaras dengan temuan tersebut, laporan *Aftech Annual Members Survey 2024* bertajuk *Indonesia's Fintech Resurgence: A New Wave of Innovations and Possibilities* menunjukkan bahwa 45% responden memperoleh pendanaan dari modal ventura. Hal ini mencerminkan optimisme investor terhadap

prospek industri *fintech* di Indonesia, meskipun masih menghadapi tantangan global seperti *tech winter* dan kondisi ekonomi yang berfluktuasi.

Di sisi lain, pertumbuhan pesat industri pinjaman *online* juga menghadirkan berbagai persoalan. Artikel yang dipublikasikan oleh lamongankab.go.id mencatat sejumlah permasalahan layanan, di antaranya bunga dan biaya yang tinggi, praktik penagihan yang agresif, penyalahgunaan data pribadi, hingga maraknya praktik pinjaman ilegal. Permasalahan tersebut menimbulkan pertanyaan mengenai sejauh mana tingkat kepuasan pengguna terhadap layanan aplikasi pinjaman *online* yang disediakan oleh perusahaan *fintech*.

Untuk menjawab persoalan tersebut, analisis sentimen menjadi pendekatan yang relevan karena mampu mengklasifikasikan opini pengguna berdasarkan polaritas positif maupun negatif. Media sosial X (Twitter) dipilih sebagai sumber data mengingat perannya sebagai ruang publik yang banyak digunakan untuk menyampaikan keluhan, apresiasi, maupun umpan balik. Tantangan analisis ini terletak pada sifat data teks media sosial yang tidak terstruktur dan sarat emosi, oleh karena itu penelitian ini menerapkan algoritma *Support Vector Machine (SVM)* yang terbukti efektif dalam mengklasifikasikan data teks dengan tingkat akurasi tinggi (Vidya, Fanany, & Budi, 2015).

TINJAUAN PUSTAKA

Dalam penulisan ini, berbagai sumber acuan digunakan, meliputi buku, jurnal nasional maupun internasional, yang memuat landasan teori sebagai dasar penelitian. Selain itu, rujukan terkait penerapan algoritma *Support Vector Machine (SVM)* dalam analisis sentimen pengguna aplikasi pinjaman online melalui media sosial X (Twitter) juga dijadikan referensi untuk memperkuat kerangka penelitian.

Pinjaman Online

Pinjaman *online* didefinisikan sebagai bentuk transaksi keuangan, baik secara langsung maupun tidak langsung, tanpa melibatkan lembaga keuangan konvensional (Omarini, 2018). Definisi lain menyebutkan bahwa pinjaman *online* merupakan model bisnis berbasis internet yang berfungsi sebagai perantara untuk memenuhi kebutuhan pinjaman antar individu maupun lembaga keuangan (Hsueh, 2017). Lebih lanjut, menurut Supriyanto dan Ismawati (2019), teknologi aplikasi pinjaman *online* merupakan solusi pembiayaan berbasis teknologi finansial yang efektif dan efisien, karena memungkinkan masyarakat memperoleh pinjaman tanpa batasan ruang dan waktu, selama perangkat yang digunakan seperti *smartphone* terhubung dengan jaringan internet.

X (Twitter)

Twitter (saat ini dikenal dengan nama X) merupakan salah satu platform media sosial yang populer di Indonesia dan memiliki peran penting dalam menghimpun opini publik terkait isu, produk, maupun layanan tertentu. Melalui fitur pesan singkat atau *tweet*, pengguna dapat mengekspresikan pandangan secara langsung, sehingga platform ini sering dimanfaatkan dalam penelitian analisis sentimen (Bakliwal et al., 2012). Banyak perusahaan memanfaatkan Twitter sebagai sarana untuk memperoleh wawasan mengenai kepuasan maupun ketidakpuasan pelanggan, karena *tweet* umumnya berisi opini yang bervariasi, mulai dari apresiasi hingga kritik (Husada et al., 2021). Sebagai layanan *microblogging*, Twitter memungkinkan pengguna menyampaikan pendapat secara spontan dengan batas maksimal 280 karakter, menjadikannya sumber data yang kaya bagi analisis sentimen.

Analisis Sentimen

Analisis sentimen sendiri merupakan bagian dari penelitian *text mining* yang berfokus pada pengelompokan data teks ke dalam kategori positif, negatif, atau netral. Pendekatan ini menggunakan teknik *Natural Language Processing (NLP)* untuk mengekstrak opini, emosi, serta evaluasi dari teks, sehingga mampu memberikan wawasan mengenai respons publik terhadap suatu isu atau produk (Idris et al., 2023). Analisis sentimen banyak dimanfaatkan untuk

mengidentifikasi kecenderungan opini publik terhadap suatu objek atau layanan, bahkan di Amerika Serikat terdapat sekitar 20-30 perusahaan yang secara khusus menyediakan layanan ini (Indurkha & Damerau, 2010).

Natural Language Processing (NLP)

Natural Language Processing (NLP) merupakan bidang interdisipliner yang menggabungkan ilmu komputer, linguistik, dan kecerdasan buatan untuk memungkinkan mesin memahami, menginterpretasikan, serta memanipulasi bahasa manusia. Perkembangan *NLP* dalam beberapa dekade terakhir mengalami kemajuan pesat, terutama berkat penerapan teknik *machine learning* dan *deep learning*. Tantangan utama dalam *NLP* antara lain adalah kompleksitas bahasa alami, ambiguitas semantik, serta keragaman konteks dan struktur gramatikal. Untuk mengatasi hal tersebut, dikembangkan berbagai metode, seperti *tokenization*, *stemming*, *lemmatization*, dan *parsing*, yang berfungsi mengubah teks ke dalam bentuk yang dapat diproses komputer (Jurafsky & Martin, 2021). Salah satu metode yang banyak digunakan dalam *NLP* adalah algoritma *Support Vector Machine (SVM)*, yang efektif dalam mengklasifikasikan teks, termasuk analisis sentimen. Keunggulan utama *SVM* terletak pada kemampuannya menangani data berdimensi tinggi dan bersifat non-linear, serta menemukan *hyperplane* optimal yang memisahkan kelas sentimen dengan margin maksimal, sehingga meningkatkan akurasi klasifikasi (Cortes & Vapnik, 1995).

Text Mining

Text mining adalah proses mengekstraksi informasi dan pengetahuan yang berguna dari teks tidak terstruktur melalui identifikasi pola tertentu dan mengolah data teks menjadi informasi yang bermakna. *Text mining* melibatkan beberapa tahapan utama, seperti pra-pemrosesan data, ekstraksi fitur, dan analisis data. Pra-pemrosesan adalah langkah awal yang meliputi pembersihan data, *tokenizing*, dan penghapusan *stopwords*, yang bertujuan mengubah data mentah menjadi format yang dapat dianalisis (Tan, 2018). Langkah-langkah utama dalam *text mining* mencakup beberapa proses penting. Pertama, *cleaning* dilakukan untuk menghapus karakter atau simbol yang tidak relevan. Kedua, *case folding* mengubah seluruh teks menjadi huruf kecil agar seragam. Ketiga, *remove punctuation* dengan menghapus tanda baca. Keempat, *tokenizing* memecah teks menjadi kata individual. Kelima, *stopwords removal* dengan menghapus kata umum seperti “dan”, “yang”, “di”, “ke”, “dari”. Keenam, *stemming* dengan mengembalikan kata ke bentuk dasar seperti “digunakan” menjadi “guna”, “membantu” menjadi “bantu”. Selanjutnya, dilakukan *filtering*, yaitu menghilangkan kata yang tidak relevan, seperti konjungsi dan preposisi. Terakhir, pembobotan *term* menggunakan metode *TF-IDF (Term Frequency-Inverse Document Frequency)* untuk memberikan bobot pada kata berdasarkan frekuensinya dalam dokumen (Fitriyah et al., 2020).

Support Vector Machine (SVM)

Support Vector Machine (SVM) merupakan algoritma *machine learning* yang populer untuk tugas klasifikasi dan regresi. Diperkenalkan oleh Vapnik pada 1992, *SVM* bekerja berdasarkan prinsip *Structural Risk Minimization (SRM)* dengan tujuan mencari *hyperplane* optimal yang memisahkan kelas data (Thalita Meisya Permata Aulia et al., 2021). Algoritma ini bersifat *kernel-based* dan memanfaatkan *support vectors* sebagian kecil data pelatihan yang paling berpengaruh dalam menentukan posisi *hyperplane* (Mohammadreza Sheykhmousa et al., 2020). Jika data tidak dapat dipisahkan secara linear, *kernel trick* digunakan untuk memetakan data ke ruang dimensi yang lebih tinggi sehingga memungkinkan pemisahan linear (Okta Ihza Gifari et al., 2022). Kinerja *SVM* sangat dipengaruhi oleh pemilihan *kernel* (linear, polynomial, radial basis function, atau sigmoid) serta parameter seperti *C* dan *gamma*. Beberapa studi menunjukkan bahwa *kernel* linear dan sigmoid memberikan hasil lebih baik dibandingkan *RBF* dan polynomial, khususnya pada analisis sentimen vaksin COVID-19 (Thalita Meisya Permata Aulia et al., 2021).

Keunggulan utama *SVM* adalah kemampuannya menangani data berdimensi tinggi dan efektif meskipun jumlah data pelatihan terbatas. Namun demikian, algoritma ini juga memiliki

keterbatasan berupa ketergantungan pada pemilihan parameter serta tingginya kebutuhan komputasi untuk dataset berukuran besar (Okta Ihza Gifari et al., 2022). Dalam praktiknya, *SVM* telah digunakan di berbagai bidang, mulai dari klasifikasi citra penginderaan jauh hingga analisis sentimen berbasis teks dengan kombinasi metode pembobotan *TF-IDF* (Mohammadreza Sheykhmousa et al., 2020; Okta Ihza Gifari et al., 2022).

Kamus *Lexicon*

Dalam *Natural Language Processing (NLP)*, kamus *lexicon* dipahami sebagai himpunan kata atau frasa yang diberi label tertentu sesuai atribut linguistik dan semantik. Dalam konteks analisis sentimen, kamus ini berisi kosakata yang telah dilengkapi dengan informasi polaritas positif, negatif, atau netral sehingga mampu menambah pemahaman semantik dalam teks (Islam et al., 2024). Penyusunan *lexicon* dapat dilakukan secara manual oleh ahli bahasa, maupun secara otomatis dengan memanfaatkan teknik statistik dan *machine learning* seperti *word embeddings* dan *transfer learning* (Alshaabi et al., 2021). Selain kamus umum, terdapat pula *lexicon* yang bersifat domain-spesifik, seperti *FinSenticNet* untuk teks keuangan (Du et al., 2023) atau *Economic Lexicon* untuk teks ekonomi dengan skala polaritas -1 hingga +1 (Barbaglia et al., 2024).

METODOLOGI PENELITIAN

Penelitian ini memberikan kontribusi signifikan terhadap pengembangan teknik analisis sentimen, khususnya dalam konteks pengolahan bahasa alami untuk teks berbahasa Indonesia. Selain itu, penelitian ini memperluas aplikasi metode *Support Vector Machine (SVM)* dalam mengklasifikasikan data yang bersumber dari *platform* media sosial. Prosedur penelitian menggunakan pendekatan *text mining* disusun secara sistematis dan divisualisasikan melalui sebuah diagram alir (*flowchart*) yang dapat dilihat pada Gambar 1.



Gambar 1. Kerangka Penelitian Menggunakan Pendekatan *Text Mining*

Pengumpulan Data

Tahapan pengumpulan data dilakukan dengan mengambil sampel sebanyak 5.233 ulasan berbahasa Indonesia yang membahas topik “pinjaman *online*” atau “pinjol” dari pengguna media sosial *X (Twitter)*. Pengambilan data ini mencakup periode waktu dari Januari 2020 hingga

Desember 2023 untuk memastikan relevansi dan kecukupan data bagi penelitian. Proses pengumpulan data dilakukan menggunakan teknik *scraping* yang memanfaatkan *Google Laboratory*. Metode tersebut menerapkan berbagai *library* Python, serta menggunakan *auth token* yang diperoleh dari akun pengguna media sosial *X (Twitter)* yang telah terautentikasi dalam sistem untuk mengakses dan mengambil data ulasan secara otomatis.

Pra-pemrosesan Data

Pra-pemrosesan data dilakukan untuk memastikan kesiapan data teks dalam analisis menggunakan metode *Support Vector Machine* (SVM). Tahapan-tahapan pra-pemrosesan meliputi:

1. Kamus Lexicon
Menerapkan kamus lexicon untuk menyesuaikan kata tidak baku, singkatan, dan istilah tidak formal menjadi bentuk standar agar teks lebih konsisten.
2. Penghapusan Emoji
Menghilangkan emoji dari teks menggunakan *library Python* seperti emoji dan *regular expressions* untuk mengurangi *noise* karena emoji tidak memiliki representasi leksikal yang dapat diproses oleh model.
3. Normalisasi
Mengkonversi seluruh huruf dalam teks menjadi huruf kecil agar semua kata dengan variasi kapitalisasi dianggap sama, misalnya "Pinjaman" dan "pinjaman".
4. Penghapusan Tanda Baca
Menghapus tanda baca seperti titik, koma, tanda seru, tanda tanya, dan simbol lainnya yang tidak memiliki makna semantik dalam analisis sentimen.
5. Tokenisasi
Memecah kalimat menjadi unit kata (token) menggunakan *library* seperti *Natural Language Toolkit* (NLTK) atau Sastrawi, untuk memudahkan analisis kata per kata oleh model SVM.
6. Penghapusan Kata Umum
Menghapus kata-kata umum yang tidak signifikan dalam analisis (*stopwords*) menggunakan kamus *stopwords* Bahasa Indonesia dari pustaka Sastrawi, sehingga mengurangi dimensi data tanpa kehilangan informasi bermakna.
7. Pengembalian Kata Dasar
Mengubah kata ke bentuk dasar menggunakan *library* Sastrawi, misalnya dari "membantu" menjadi "bantu", guna mengurangi variasi kata dan fokus pada makna inti.

Pelabelan Data

Proses pelabelan data bertujuan menentukan kategori sentimen (positif, negatif, atau netral) untuk setiap data teks yang telah melalui tahap pra-pemrosesan. Dalam penelitian ini, pelabelan dilakukan dengan dua metode utama:

1. Pelabelan Manual
Sebagian data diberi label secara manual oleh anotator untuk menjamin tingkat akurasi yang tinggi. Data berlabel ini berfungsi sebagai acuan bagi pelatihan model. Pelabelan manual biasanya membutuhkan waktu lebih lama namun menghasilkan kualitas pelabelan yang lebih akurat.
2. Pelabelan Otomatis
Menggunakan perangkat berbasis *Natural Language Processing* (NLP) untuk memberikan label awal secara otomatis pada dataset yang besar. Hasil pelabelan otomatis kemudian diverifikasi atau disempurnakan melalui pengecekan manual untuk memastikan kualitasnya. Pelabelan data yang akurat menjadi fondasi krusial dalam melatih model *Support Vector Machine* agar mampu melakukan prediksi sentimen dengan andal. Berikut ini disediakan contoh kode program untuk pelabelan data dalam kategori positif, negatif, dan netral.

```
# proses pelabelan data secara manual
positive_keywords = [
    # kepuasan & rasa senang
    "senang", "puas", "suka", "mantap", "luar biasa", "hebat", "keren",
    # kualitas & keandalan
    "bagus", "baik", "terbaik", "rekomendasi", "terpercaya", "handal",
    # kecepatan & kelancaran
    "cepat", "lancar", "respon cepat", "responsif", "proses cepat", "praktis",
    "efisien",
    # kemudahan & kenyamanan
    "mudah", "simplen", "proses simpel", "tidak ribet", "praktikal",
    # keamanan & transparansi
    "aman", "amanah", "transparan", "jelas",
    # bunga & biaya
    "bunga ringan", "ringan", "murah",
    # layanan
    "membantu", "sangat membantu", "bantu", "solusi", "support bagus",
    "pelayanan ramah"
]
neutral_keywords = [
    # kata umum netral
    "biasa", "cukup", "standar", "netral", "normal", "umum",
    # istilah teknis pinjaman online
    "pinjam", "online", "pakai", "aplikasi", "fitur", "limit", "tenor", "angsuran",
    "proses", "pengajuan", "sistem", "platform", "layanan",
    # data & akun
    "data", "saldo", "notifikasi", "informasi", "akun", "daftar", "login",
    "verifikasi",
    # aspek transaksi
    "transaksi", "pembayaran", "jadwal", "riwayat", "tagihan", "persyaratan",
    # interaksi umum
    "hubungi", "kontak", "email", "telepon", "nomor", "detail"
]
negative_keywords = [
    # kata kasar / makian
    "haram", "bangsat", "bgst",
    "anjing", "setan", "babi", "maki", "tolol", "goblok", "kampret",
    # performa buruk
    "lambat", "buruk", "error", "lemot", "hang", "crash", "tidak responsif",
    # pengalaman negatif
    "masalah", "tidak puas", "parah", "gagal", "tidak jelas", "ribet", "sulit",
    "jelek", "sampah", "tidak berguna", "menyusahkan", "komplain", "enggak jelas",
    # isu keuangan & bunga
    "bunga tinggi", "bunga mencekik", "rentenir", "biaya besar", "potongan
    aneh", "riba",
    # penipuan & keamanan
    "penipuan", "bohong", "scam", "tipu", "tidak aman", "minta akses kontak", "akses
    kontak",
    "data bocor", "pencurian data", "phishing", "penipu", "tidak aman",
    # penagihan & intimidasi
    "penagih kasar", "ancaman", "intimidasi", "dicecar", "ditagih", "dihubungi",
    "marah",
```

```

"teror", "neror", "dikejar", "ditakut-takuti", "malu-maluin", "tagih", "takut",
"kasar", "kejar", "malu",
# hutang / gagal bayar
"galbay", "enggak pernah buat pinjol", "kena pinjol", "ngegampangin",
"enggak jelas", "minta tolong pinjol", "diminta bayar", "pinjol sampe miliaran",
"bertahun-tahun", "utang menumpuk", "tidak bisa bayar", "dipaksa bayar",
"paksa bayar",
"hutang bertambah", "cicilan macet", "utang", "gagal bayar",
# perjudian & spam
"slot", "judi", "judi online", "iklan spam", "spam", "ganggu", "bahaya",
# sikap perusahaan
"tidak bertanggung jawab", "abaikan", "cuek", "tidak peduli", "abai",
# perasaan pengguna
"stress", "tertekan", "menyesal", "kapok", "takut", "khawatir", "trauma", "riba",
"gila", "jahat"
]
# Gabungkan hasil stemming menjadi string utuh untuk keperluan klasifikasi teks
df['cleaned_text'] = df['stemmed_text'].apply(lambda x: ' '.join(x))

# Fungsi pelabelan otomatis berdasarkan keyword
def label_sentiment(text):
    text = text.lower()
    if any(word in text for word in positive_keywords):
        return 1 # Positif
    elif any(word in text for word in negative_keywords):
        return -1 # Negatif
    elif any(word in text for word in neutral_keywords):
        return 0 # Netral

# Proses pelabelan
df['sentiment_label'] = df['cleaned_text'].apply(label_sentiment)
# Tampilkan hasil pelabelan
# Menampilkan data dengan label positif
print(df[df['sentiment_label'] == 1][['stemmed_text', 'sentiment_label']].head(3))

# Menampilkan data dengan label negatif
print(df[df['sentiment_label'] == -1][['stemmed_text', 'sentiment_label']].head(3))

# Menampilkan data dengan label netral
print(df[df['sentiment_label'] == 0][['stemmed_text', 'sentiment_label']].head(3))

```

Pemisahan Data

Pengembangan model *machine learning* pada penelitian ini dilakukan melalui proses pemisahan data (*data splitting*) menjadi dua bagian utama, yaitu data latih (*training data*) dan data uji (*testing data*). Proses ini penting karena berpengaruh langsung terhadap kemampuan model dalam melakukan prediksi secara akurat sekaligus mencegah terjadinya *overfitting*, yaitu kondisi ketika model terlalu menyesuaikan diri dengan data latih sehingga gagal berfungsi optimal pada data baru.

Pada tahap ini, dataset dibagi dengan proporsi 80% untuk data latih dan 20% untuk data uji. Data latih digunakan dalam proses pembelajaran model *Support Vector Machine (SVM)*, sedangkan data uji dimanfaatkan untuk mengukur kinerja model yang telah dibangun.

Evaluasi terhadap model dilakukan dengan menggunakan sejumlah metrik, antara lain akurasi, presisi, *recall*, dan *F1-score*. Penerapan metrik ini bertujuan untuk memastikan bahwa model tidak hanya memiliki performa yang baik pada data latih, tetapi juga mampu melakukan generalisasi dengan baik terhadap data baru.

Ekstraksi Fitur pada SVM Model dengan TF-IDF

Pengujian dilakukan menggunakan 887 data uji yang sebelumnya telah dipisahkan dari total 4433 dataset. Model yang digunakan pada pengujian ini adalah model dengan parameter *batch size* sebesar 50 dan *epoch* sebanyak 50, yang menunjukkan kinerja terbaik selama proses pelatihan. Pengujian ini dilaksanakan secara bersamaan dengan proses kompilasi model pelatihan. Selama proses pelatihan, nilai *validation loss* dan *validation accuracy* juga dihitung dan ditampilkan. Hasil dari proses tersebut memungkinkan evaluasi apakah nilai akurasi dan *loss* memiliki kedekatan yang konsisten dengan nilai validasi akurasi dan validasi *loss*.

Implementasi Data

Tahap ekstraksi fitur dilakukan untuk mentransformasikan data teks ke dalam bentuk representasi numerik sehingga dapat diproses oleh model *Support Vector Machine (SVM)*. Pada penelitian ini digunakan teknik *Term Frequency-Inverse Document Frequency (TF-IDF)*, yang berfungsi memberikan bobot pada setiap kata berdasarkan tingkat kemunculannya dalam suatu dokumen relatif terhadap keseluruhan *corpus*. Pendekatan ini memungkinkan model untuk mengenali kata-kata yang memiliki tingkat kepentingan lebih tinggi dalam membedakan kategori sentimen.

Evaluasi Data

Hasil evaluasi model digunakan sebagai dasar untuk interpretasi akhir, yang meliputi pemahaman pola sentimen pengguna aplikasi pinjaman online di platform media sosial X. Analisis ini juga dimanfaatkan untuk memberikan rekomendasi yang relevan berdasarkan temuan sentimen yang teridentifikasi. Proses evaluasi bertujuan menilai kemampuan model dalam mengklasifikasikan teks secara efektif sesuai dengan kategori sentimen yang diwakili, sehingga memastikan ketepatan dalam pemetaan opini pengguna pada data media sosial.

Visualisasi Data

Visualisasi hasil kinerja model *Support Vector Machine (SVM)* bertujuan memberikan interpretasi yang lebih intuitif terhadap efektivitas klasifikasi yang dilakukan. Salah satu alat yang umum digunakan adalah *confusion matrix*, yang menampilkan jumlah prediksi benar dan salah pada setiap kategori sentimen (positif, negatif, dan netral), sehingga memudahkan identifikasi pola kesalahan dan area yang perlu diperbaiki. Selain itu, visualisasi berupa diagram batang yang menggambarkan distribusi label aktual dan prediksi memberikan gambaran keseimbangan kelas serta kemampuan model dalam mengenali pola data secara keseluruhan. Visualisasi ini berperan penting dalam melengkapi evaluasi kuantitatif, memperkuat pemahaman terhadap performa model secara menyeluruh.

HASIL DAN PEMBAHASAN

Pengumpulan Data

Hasil pengumpulan data dalam bentuk dokumen format .csv, berikut ini ditampilkan dengan beberapa hasil data dari total 5233 ulasan yang sudah di *scraping*.

4. Penghapusan Tanda Baca

Menghilangkan tanda baca dan simbol yang tidak memiliki makna semantik penting dalam analisis teks untuk menyederhanakan struktur data.

```
=== Hasil Remove Punctuation ===
Index: 230
case_folded:
  nomor _ simpati ini tidak pernah saya gunakan untuk telpon2. nomor ini hanya digunakan untuk mendaftar apps online seperti keperluan go
jek grab dana &amp

punctuation_removed:
  nomor simpati ini tidak pernah saya gunakan untuk telpon2 nomor ini hanya digunakan untuk mendaftar apps online seperti keperluan gojek
grab dana amp
```

Gambar 6. Hasil Data Dengan Penghapusan Tanda Baca

5. Tokenisasi

Memisahkan kalimat menjadi unit terkecil berupa kata (token) untuk memungkinkan model mengenali dan mempelajari pola serta makna setiap kata secara menyeluruh.

```
=== Hasil Tokenisasi ===
Index: 230
punctuation_removed:
  nomor simpati ini tidak pernah saya gunakan untuk telpon2 nomor ini hanya digunakan untuk mendaftar apps online seperti keperluan gojek
grab dana amp

tokenized_text:
  ['nomor', 'simpati', 'ini', 'tidak', 'pernah', 'saya', 'gunakan', 'untuk', 'telpon2', 'nomor', 'ini', 'hanya', 'digunakan', 'untuk', 'me
ndaftar', 'apps', 'online', 'seperti', 'keperluan', 'gojek', 'grab', 'dana', 'amp']
```

Gambar 7. Hasil Data Dengan Tokenisasi

6. Penghapusan tanda baca

Menghapus kata-kata umum yang tidak berkontribusi secara signifikan pada pemaknaan, seperti kata sambung atau imbuhan, sehingga fokus analisis tertuju pada kata kunci yang bermakna.

```
=== Hasil Remove Stopwords ===
Index: 230
tokenized_text:
  ['nomor', 'simpati', 'ini', 'tidak', 'pernah', 'saya', 'gunakan', 'untuk', 'telpon2', 'nomor', 'ini', 'hanya', 'digunakan', 'untuk', 'me
ndaftar', 'apps', 'online', 'seperti', 'keperluan', 'gojek', 'grab', 'dana', 'amp']

filtered_text:
  ['nomor', 'simpati', 'pernah', 'gunakan', 'telpon2', 'nomor', 'hanya', 'digunakan', 'mendaftar', 'apps', 'online', 'seperti', 'keperlua
n', 'gojek', 'grab', 'dana', 'amp']
```

Gambar 8. Hasil Data Dengan Penghapusan kata umum

7. Penghapusan Kata Umum

Mengubah kata-kata menjadi bentuk dasarnya untuk mengurangi variasi bentuk kata tanpa mengubah makna inti, misalnya mengubah “gunakan” menjadi “guna”.

```
=== Hasil Stemming (berdasarkan Lexicon) ===
Index: 230
filtered_text:
  ['nomor', 'simpati', 'pernah', 'gunakan', 'telpon2', 'nomor', 'hanya', 'digunakan', 'mendaftar', 'apps', 'online', 'seperti', 'keperlua
n', 'gojek', 'grab', 'dana', 'amp']

stemmed_text:
  ['nomor', 'simpati', 'pernah', 'guna', 'telpon2', 'nomor', 'hanya', 'guna', 'daftar', 'apps', 'online', 'seperti', 'perlu', 'gojek', 'gra
b', 'dana', 'amp']

stemmed_joined:
  nomor simpati pernah guna telpon2 nomor hanya guna daftar apps online seperti perlu gojek grab dana amp
```

Gambar 9. Hasil Data Dengan Penghapusan kata umum

Evaluasi dan Analisis Klasifikasi Sentimen SVM

Evaluasi kinerja model klasifikasi sentimen yang dibangun dengan algoritma *Support Vector Machine (SVM)* dilakukan melalui pengukuran sejumlah metrik evaluasi, meliputi presisi, *recall*, *F1-score*, serta *support* pada masing-masing kategori sentimen, yakni negatif, netral, dan positif. Selain itu, disajikan pula metrik agregat yang menggambarkan performa model secara keseluruhan dalam mengklasifikasikan data sentimen, sehingga memberikan gambaran komprehensif mengenai efektivitas model yang digunakan.

===Evaluasi Model SVM===
Akurasi: 0.8647

Laporan Klasifikasi:

	precision	recall	f1-score	support
Negatif	0.86	0.87	0.86	362
Positif	0.83	0.90	0.86	341
Netral	0.98	0.78	0.87	184
accuracy			0.86	887
macro avg	0.89	0.85	0.87	887
weighted avg	0.87	0.86	0.86	887

Gambar 10. Laporan Evaluasi Performa Model SVM

Penjelasan dari hasil evaluasi untuk setiap sentimen dengan model SVM dapat diuraikan sebagai berikut:

Presisi:

1. Presisi sentimen negatif mencapai 0,86, menunjukkan bahwa 86% dari prediksi negatif sesuai dengan label sebenarnya.
2. Presisi sentimen netral sebesar 0,98 atau 98%, menandakan tingkat ketepatan sangat tinggi dalam mengidentifikasi ulasan netral.
3. Presisi sentimen positif tercatat 0,83 atau 83%, mengindikasikan prediksi ulasan positif sebagian besar terklasifikasi dengan benar.

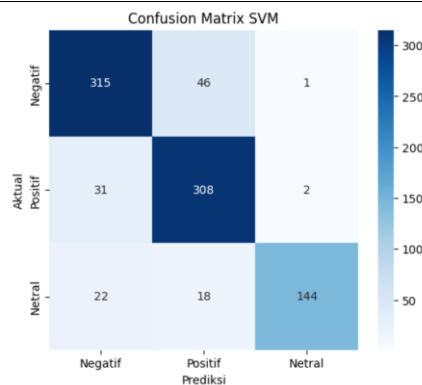
Recall:

1. *Recall* sentimen negatif sebesar 0,87, memperlihatkan bahwa model mampu mendeteksi 87% dari total ulasan yang benar-benar negatif. Hal ini menandakan bahwa hanya sekitar 13% ulasan negatif yang tidak berhasil terdeteksi oleh model.
2. *Recall* sentimen netral sebesar 0,78, menunjukkan kemampuan model yang baik namun masih terdapat sebagian ulasan netral yang belum teridentifikasi dengan sempurna.
3. *Recall* sentimen positif sebesar 0,90, mengindikasikan performa yang cukup baik, namun model masih memiliki keterbatasan dalam mengenali sentimen positif.

F1-score:

1. F1-score untuk sentimen negatif adalah 0,86 atau 86%, memperlihatkan performa moderat, di mana nilai presisi tinggi terbatas oleh recall yang lebih rendah.
2. F1-score sentimen netral sebesar 0,87, menunjukkan keseimbangan moderat antara presisi dan recall.
3. F1-score untuk sentimen positif juga tercatat 0,86 atau 86%, mengindikasikan kebutuhan peningkatan performa pada kategori ini. Hal ini mencerminkan keseimbangan antara tingkat ketepatan (83%) dan kemampuan mendeteksi ulasan positif secara benar (90%).

Hasil kinerja model *Support Vector Machine* (SVM) ditampilkan melalui *Confusion Matrix* yang mengklasifikasikan tiga kategori sentimen, yakni negatif, netral, dan positif. Matriks ini memberikan gambaran menyeluruh mengenai kemampuan model dalam melakukan klasifikasi berdasarkan data uji. Pada *Confusion Matrix*, baris merepresentasikan label sebenarnya (*true labels*), sedangkan kolom menunjukkan label hasil prediksi model. Tingkat intensitas warna pada setiap sel menunjukkan frekuensi prediksi, di mana semakin gelap warna yang ditampilkan, semakin tinggi jumlah prediksi yang sesuai pada kategori tersebut.

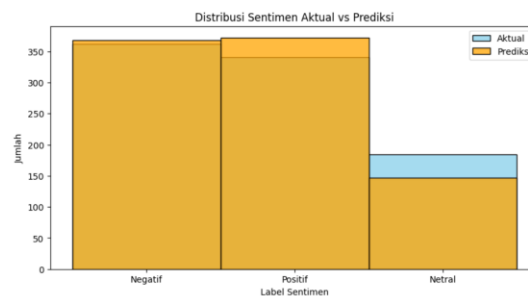


Gambar 11. Grafik Dengan *Confusion Matrix*

Berdasarkan hasil analisis *Confusion Matrix*, model *Support Vector Machine* (SVM) menunjukkan kinerja yang baik dengan 767 ulasan terklasifikasi dengan benar dari total 887 ulasan. Klasifikasi terbaik ditunjukkan pada ulasan sentimen negatif dengan jumlah prediksi benar paling tinggi pada kategori tersebut, yaitu 315 ulasan negatif berhasil terklasifikasi dengan tepat. Meskipun demikian, model masih menghadapi tantangan dalam membedakan sentimen netral dan positif. Hal ini ditunjukkan dengan adanya 31 ulasan positif yang salah diprediksi sebagai negatif serta 40 ulasan netral yang salah diprediksi sebagai negatif maupun positif. Dengan demikian, diperlukan upaya perbaikan melalui peningkatan kualitas representasi data pada kelas tertentu serta penyesuaian parameter model agar kemampuan klasifikasi menjadi lebih seimbang dan akurat di seluruh kategori sentimen.

Visualisasi Data SVM Model

Visualisasi yang disajikan memperlihatkan kemampuan model dalam mengklasifikasikan data ke dalam kategori sentimen negatif, netral, dan positif. Hasil tersebut merupakan keluaran dari rangkaian proses analisis yang mencakup tahap pra-pemrosesan teks, ekstraksi fitur, pemisahan data latih dan uji, serta pelatihan model.



Gambar 12. Visualisasi distribusi antara label sentimen aktual dengan hasil prediksi yang dihasilkan oleh model *Support Vector Machine* (SVM).

Berdasarkan grafik perbandingan antara jumlah data aktual dan hasil prediksi model *Support Vector Machine* (SVM), terlihat bahwa pada kategori negatif, jumlah prediksi model hampir sama dengan data aktual, meskipun sedikit lebih tinggi. Pada kategori positif, prediksi model juga sedikit lebih tinggi daripada data aktual. Sementara itu, pada kategori netral, jumlah prediksi model lebih rendah dibandingkan data aktual, yang menunjukkan keterbatasan model dalam mengenali sentimen netral. Ketidakseimbangan distribusi data pelatihan, khususnya pada kelas netral yang relatif lebih sedikit, diduga menjadi penyebab terbatasnya representasi pola yang dipelajari model. Dengan demikian, meskipun performa model sudah cukup baik pada kategori negatif dan positif, strategi peningkatan seperti penyeimbangan data atau penyesuaian bobot kelas perlu dipertimbangkan agar hasil klasifikasi menjadi lebih proporsional dan akurat.

Visualisasi ini memberikan wawasan penting untuk mengevaluasi kelemahan model serta menjadi dasar perbaikan pada tahap pengembangan selanjutnya.



Gambar 13. Tampilan Grafik Wordcloud Sentimen Positif Dengan SVM

Pada *WordCloud* sentimen positif, ditampilkan representasi kata yang paling sering muncul dalam ulasan berlabel positif. Visualisasi ini memberikan gambaran mengenai istilah yang dominan digunakan oleh pengguna dalam menyampaikan ekspresi sentimen positif.



Gambar 14. Tampilan Grafik Wordcloud Sentimen Negatif Dengan SVM

Sedangkan pada sentimen negatif, dominasi kata seperti tagih, teror, utang dan judi *online* mencerminkan adanya keluhan terkait praktik penagihan serta citra buruk layanan.

KESIMPULAN

Kesimpulan

Penelitian ini menunjukkan bahwa ulasan pengguna aplikasi pinjaman *online* terbagi ke dalam tiga klasifikasi sentimen, yaitu negatif, netral, dan positif. Dari total 887 ulasan yang dianalisis, sebanyak 362 ulasan (41%) termasuk dalam klasifikasi negatif, 341 ulasan (38%) termasuk dalam klasifikasi positif, dan 184 ulasan (21%) termasuk dalam klasifikasi netral. Distribusi tersebut memperlihatkan bahwa sentimen negatif mendominasi persepsi pengguna, yang umumnya berkaitan dengan permasalahan tingkat bunga pinjaman, praktik penagihan, serta tekanan psikologis yang ditimbulkan oleh layanan pinjaman online.

Penerapan metode *Support Vector Machine* (SVM) terbukti mampu memberikan kinerja yang baik dalam mengidentifikasi ulasan dengan sentimen negatif. Hal ini didukung oleh jumlah data pada kategori tersebut yang lebih besar sehingga pola sentimen dapat dipelajari secara lebih maksimal. Sebaliknya, keterbatasan model masih terlihat pada kategori netral yang jumlah datanya lebih sedikit, sehingga representasi pola pada tahap pelatihan kurang kuat. Dengan demikian, SVM dapat dinyatakan efektif untuk analisis sentimen ulasan pengguna aplikasi pinjaman online, meskipun pengembangan lebih lanjut tetap diperlukan agar kinerja model lebih seimbang dan akurat pada seluruh kategori sentimen.

Saran

Berdasarkan hasil penelitian, penyedia layanan pinjaman *online* disarankan untuk meningkatkan transparansi biaya, memperkuat keamanan data, dan mengembangkan fitur aplikasi yang lebih ramah pengguna guna meningkatkan kepercayaan serta kepuasan masyarakat. Peneliti selanjutnya dapat mengadopsi metode pembelajaran mesin lanjutan seperti *deep learning* atau model berbasis *transformer* serta menerapkan teknik penyeimbangan data agar kinerja klasifikasi menjadi lebih optimal. Penelitian ini terbatas pada data yang diperoleh dari media sosial X (Twitter). Untuk menghasilkan analisis yang lebih komprehensif, penelitian selanjutnya disarankan untuk memperluas sumber data dengan mengintegrasikan ulasan dari berbagai

platform lain, seperti Facebook, Instagram, atau forum daring, sehingga hasil analisis sentimen lebih representatif.

DAFTAR PUSTAKA

- Alhaq, Z., Mustopa, A., Mulyatun, S., & Santoso, J. D. (2021). Penerapan Metode Support Vector Machine Untuk Analisis Sentimen Pengguna Twitter. *Journal of Information System Management (JOISM)*, 3(1), 16-21.
<https://doi.org/10.24076/JOISM.2021V3I2.558>
- Alshaabi, T., et al. (2021). *Augmenting semantic lexicons using word embeddings and transfer learning*. <https://arxiv.org/abs/2109.09010>
- Alviani, V., Alam, S., & Kurniawan, I. (2023). Analisis sentimen review aplikasi WeTV pada platform Twitter menggunakan Support Vector Machine. *STORAGE: Jurnal Ilmiah Teknik dan Ilmu Komputer*, 2(3), 143-149.
<https://doi.org/10.55123/storage.v2i3.2351>
- Aulawi, H., Kurniawan, W. A., & Azhar, A. S. (2022). Analisis Sentimen Terhadap Layanan INDOSAT pada Media Sosial Twitter Selama Pandemi. *Jurnal Kalibrasi*, 20(1), 53-59.
<https://doi.org/10.33364/kalibrasi/v.20-1.1114>
- Ananda, F. D., & Pristyanto, Y. (2021). Analisis Sentimen Pengguna Twitter Terhadap Layanan Internet Provider Menggunakan Algoritma Support Vector Machine. *MATRIK: Jurnal Manajemen, Teknik Informatika dan Rekayasa Komputer*, 20(2), 407-416.
<https://doi.org/10.30812/matrik.v20i2.1130>
- Das, S. R., Donini, M., Zafar, M. B., He, J., & Kenthapadi, K. (2022). Finlex: An effective use of word embeddings for financial lexicon generation. *The Journal of Finance and Data Science*, 8, 1-11. <https://doi.org/10.1016/j.jfds.2021.10.001>
- Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. arXiv preprint arXiv:1810.04805.
- Fitriyah, N., Warsito, B., & Maruddani, D. A. I. (2020). Analisis Sentimen Gojek pada Media Sosial Twitter dengan Klasifikasi *Support Vector Machine* (SVM). *Jurnal Gaussian*, 9(3), 376-390.
- Gifari, O. I., Adha, M., Hendrawan, I. R., & Durrand, F. F. S. (2022). Analisis Sentimen Review Film Menggunakan TF-IDF dan Support Vector Machine. *Journal of Information Technology*, 2(1), 36-40. <https://doi.org/10.46229/jifotech.v2i1.330>
- Hidayat, R., Wulandari, H., Cahyaning, Y., Firmansyah, A. D., Hartono, B., Yusril, R., & Rakhmawati, N. A. (2020). Analisa permasalahan berbagai platform pinjaman online (peer-to-peer lending) dan penanganannya di Indonesia. *Inspiration: Jurnal Teknologi Informasi dan Komunikasi*, 10(1), 45-56.
<https://doi.org/10.35585/inspir.v10i1.2538>
- Husada, H. C., & Suryaputra Paramita, A. (2021). Analisis Sentimen Pada Maskapai Penerbangan di Platform Twitter Menggunakan Algoritma Support Vector Machine (SVM).
<https://doi.org/10.34148/TEKNIKA.V10I1.311>
- Khairani, Z., & Taufiq, M. (2023). Pelaksanaan Transaksi Pinjaman Online Di Lubuk Sikaping, Kabupaten Pasaman Ditinjau Dari Prepektif Ekonomi Islam. *Jurnal Ilmiah Manajemen, Bisnis Dan Kewirausahaan*, 3(1), 125-133.
- Putri, S., Apriyanto, Y. A., & Wijaya, A. (2023). Analisis sentimen ulasan aplikasi DeepL pada Google Play dengan metode support vector machine (SVM). *Jurnal Sistem Informasi (JUSIN)*, 4(2), 59-66.
<https://doi.org/10.32546/jusin.v4i2.2368>
- Rachman, F. P., & Santoso, H. (2021). Perbandingan Model Deep Learning untuk Klasifikasi Sentiment Analysis dengan Teknik Natural Language Processing. *Jurnal Teknologi dan Manajemen Informatika*, 7(2), 103-112.
<https://doi.org/10.26905/jtmi.v7i2.6506>
- Tan, P.-N., Steinbach, M., & Kumar, V. (2018). *Introduction to Data Mining (2nd ed.)*. Pearson

Sarawati Sulandjari, I., & Juwita, J. (2023). Kepastian Hukum Financial Technology Peer To Peer Lending Terhadap UMKM Di Indonesia. *Innovative: Journal Of Social Science Research*, 3(6), 10715-10728. <https://doi.org/10.59024/bhinneka.v1i3.183>

Styawati, S., Nurkholis, A., Alim, S., & Safitri, N. (2023). Study The SVM Kernel For Classification Of Covid-19 Vaccine Data On Twitter. *J. Tekno Kompak*, 17(1), 160.