

Analisis Sentimen Pinjaman Online: Studi Komparatif Algoritma Naïve Bayes, Decision Tree, dan KNN

Khyntia Miranda¹, Ryan Randy Suryono^{2,*}

¹ Program Studi Sistem Informasi, Universitas Teknokrat Indonesia, Indonesia

² Program Studi Ilmu Komputer, Universitas Teknokrat Indonesia, Indonesia

* Correspondence: ryan@teknokrat.ac.id

Copyright: © 2025 by the authors

Received: 3 Mei 2025 | Revised: 12 Mei 2025 | Accepted: 22 Mei 2025 | Published: 10 Agustus 2025

Abstrak

Perkembangan layanan pinjaman *online* di Indonesia telah memunculkan berbagai respon dari masyarakat di media sosial, termasuk keluhan mengenai metode penagihan dan kekhawatiran mengenai suku bunga yang tinggi. Penelitian ini bertujuan membandingkan kinerja algoritma *Naive Bayes*, *Decision Tree*, dan *K-Nearest Neighbors* (KNN). Jenis penelitian ini adalah kuantitatif dan data yang digunakan 5.941 *tweet* melalui teknik *crawling* dari media sosial X, dilanjutkan *preprocessing*, pelabelan data dengan *lexicon based*, ekstraksi fitur menggunakan TF-IDF, dan klasifikasi sentimen menggunakan ketiga algoritma tersebut. Tahapan evaluasi menggunakan *Confusion matrix* yang dapat menghitung seperti *accuracy*, *precision*, *recall* dan *F-1 score*. Hasil penelitian menunjukkan bahwa *Decision Tree* memberikan performa paling konsisten dengan akurasi 69% karena kemampuannya dalam mengenali pola data kompleks dan memahami hubungan antar fitur. *Naive Bayes* unggul dalam klasifikasi sentimen negatif dengan akurasi 68%, sementara KNN menunjukkan performa terendah dengan akurasi 44% karena tidak efektif dalam menangani data teks berdimensi tinggi. Hasil ini dapat dimanfaatkan oleh penyedia layanan pinjaman online dan regulator untuk membangun sistem pemantauan opini publik yang akurat, dalam rangka merespons isu-isu yang menjadi perhatian masyarakat dan meningkatkan kualitas layanan secara berkelanjutan.

Kata kunci: analilis sentiment; *decision tree*; *k-nearest neighbour*; *naïve bayes*; pinjaman *online*

Abstract

The development of online lending services in Indonesia has led to various responses from the public on social media, including complaints about billing methods and concerns about high interest rates. This study aims to compare the performance of Naive Bayes, Decision Tree, and K-Nearest Neighbors (KNN) algorithms. This type of research is quantitative, and the data used is 5,941 tweets through crawling techniques from X social media, followed by preprocessing, data labeling with a lexicon-based feature extraction using TF-IDF, and sentiment classification using the three algorithms. The evaluation stage uses a confusion matrix, which can calculate accuracy, precision, recall, and the F-1 score. The results show that the decision tree provides the most consistent performance with 69% accuracy due to its ability to recognize complex data patterns and understand relationships between features. Naive Bayes excels in negative sentiment classification with 68% accuracy, while KNN shows the lowest performance with 44% accuracy because it is not effective in handling high-dimensional text data. These results can be utilized by online loan service providers and regulators to build an accurate public opinion monitoring system in order to respond to issues of public concern and improve service quality on an ongoing basis.

Keywords: sentiment analysis; *decision tree*; *k-nearest neighbor*; *naïve bayes*; online loans



PENDAHULUAN

Kemajuan teknologi digital telah memberikan kemudahan bagi masyarakat dalam mengekspresikan opini serta respons terhadap berbagai isu sosial maupun ekonomi melalui platform media sosial (Ramli et al., 2023). Salah satu topik yang memperoleh perhatian luas dari publik adalah maraknya penggunaan layanan pinjaman online di Indonesia (Wati & Syahfitri, 2022). Melalui media sosial X, masyarakat menyampaikan berbagai reaksi, baik berupa keluhan, pengalaman pribadi, maupun dukungan, yang mencerminkan persepsi publik terhadap layanan pinjaman *online* tersebut (Wijayanti, 2022). Kondisi ini menciptakan dinamika opini publik yang kompleks dan memerlukan pemahaman yang mendalam. Namun, hingga saat ini, belum banyak kajian yang secara sistematis menganalisis sentimen masyarakat terhadap pinjaman *online* berbasis data media sosial, khususnya dalam konteks bahasa Indonesia yang informal (Arvante, 2022).

Pinjaman *online* telah menimbulkan beragam respons di kalangan masyarakat, Terutama sebagaimana terlihat di media sosial. Persepsi publik terhadap layanan pinjaman *online* perlu dianalisis secara sistematis karena dapat menjadi dasar penting dalam perlindungan konsumen dan pengambilan kebijakan finansial yang lebih responsif (Sugangga & Sentoso, 2020). Oleh karena itu, analisis sentimen diperlukan untuk menangkap opini masyarakat secara luas dan objektif, terutama karena data media sosial bersifat tidak terstruktur dan sangat dinamis (Pamungkas & Cahyono, 2024).

Penelitian ini menerapkan tiga algoritma klasifikasi, yaitu *naïve bayes*, *decision tree*, dan KNN. Ketiga metode tersebut dipilih karena dinilai efektif dalam menangani karakteristik data dari media sosial serta mampu memberikan hasil klasifikasi sentimen yang cukup akurat (Riyana & Fatmawati, 2021). Serta menerapkan analisis sentimen berbasis *Natural Language Processing* (NLP) yang didukung oleh algoritma pembelajaran *machine learning*. NLP memungkinkan sistem komputer untuk memahami serta memproses bahasa manusia secara otomatis, sehingga mampu mengekstraksi makna dan sentimen dari teks yang tidak terstruktur (Silalahi & Ginting, 2023).

Naive Bayes merupakan algoritma klasifikasi berbasis probabilistik yang efisien dan sederhana, serta cocok untuk memproses teks dalam jumlah besar (Agresia et al., 2025). *Decision Tree*, menyajikan proses klasifikasi dalam bentuk struktur pohon keputusan yang intuitif, sehingga memudahkan interpretasi terhadap proses pengambilan keputusan klasifikasi (Supriyadi, 2023). Sedangkan KNN bekerja berdasarkan prinsip kemiripan, yakni mengklasifikasikan data baru berdasarkan kemiripan dengan data-data yang sudah dilabeli sebelumnya. Pendekatan ini fleksibel dan tidak memerlukan proses pelatihan model yang kompleks (A'yuniyah & Reza, 2023). Dengan menerapkan kedua algoritma tersebut, diharapkan analisis sentimen mampu menghasilkan klasifikasi yang akurat dan reliabel terhadap opini publik mengenai layanan pinjaman online (Prayudi & Nasution, 2020).

Penelitian sebelumnya yang dilakukan oleh Safira & Hasan (2023) menggunakan algoritma *naive bayes* dengan total 405 tweet menunjukkan akurasi tertinggi sebesar 95%. Distribusi data dilakukan dengan 70% untuk data *training* dan 30% untuk data *testing*. Pada penelitian lainnya, menggunakan algoritma *decision tree* dalam menganalisis sentimen dengan total 477 *tweet* (Rahmianti, 2022). Hasil analisis mereka menunjukkan akurasi sebesar 94,27%, *precision* mencapai 100%, dan *recall* sebesar 89,29%. Sementara itu, penelitian yang dilakukan Siki et al. (2025) membahas tentang analisis sentimen pada pinjaman online menggunakan algoritma KNN. Dengan total 11.288 *tweet*, diperoleh akurasi sebesar 62,97%. Distribusi sentimen pada data menunjukkan sentimen negatif mendominasi sebesar 62,29%, diikuti sentimen netral sebesar 33,19%, dan sentimen positif hanya sebesar 5,52%.

Berdasarkan hasil beberapa temuan sebelumnya, algoritma *naive bayes* dan *decision tree* menunjukkan akurasi tinggi pada *dataset* terbatas, namun KNN kurang optimal pada data besar, dengan tantangan distribusi sentimen tidak seimbang. Oleh karena itu, penelitian ini dirancang

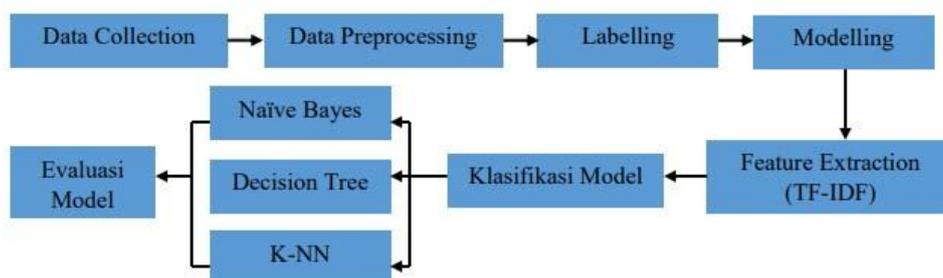
untuk melengkapi penelitian sebelumnya dengan mengevaluasi efektivitas pelabelan *lexicon-based* dalam klasifikasi sentimen dan membandingkan performa algoritma populer. Hasilnya diharapkan menjadi rujukan pengembangan metode yang lebih adaptif terhadap bahasa media sosial Indonesia.

Penelitian ini penting untuk dilakukan karena dapat memberikan gambaran komprehensif terkait efektivitas algoritma *machine learning* dalam menganalisis data teks tidak formal yang digunakan di media sosial. Temuan dari penelitian ini dapat digunakan untuk membangun sistem pemantauan opini publik yang lebih akurat bagi penyedia layanan pinjaman *online*, sehingga mereka dapat secara proaktif mengidentifikasi dan menangani aspek-aspek layanan yang menjadi sumber ketidakpuasan pengguna. Dengan menerapkan algoritma yang paling akurat, penyedia layanan pinjaman *online* dapat membangun sistem pemantauan opini publik yang mampu mengidentifikasi aspek layanan yang sering dikeluhkan, seperti kurangnya transparansi biaya, atau kualitas layanan pelanggan. Informasi ini dapat diintegrasikan dalam sistem manajemen mutu guna mendorong perbaikan berkelanjutan, meningkatkan kualitas layanan, serta meningkatkan kepuasan dan mengurangi ketidakpuasan pengguna (Ismail & Hakim, 2023).

Tujuan penelitian ini adalah untuk mengidentifikasi algoritma *machine learning* paling efektif dalam menganalisis sentimen data teks informal dan bervariasi berbahasa Indonesia, khususnya pada percakapan mengenai pinjaman *online* di *Twitter*. Selain mengidentifikasi, penelitian ini juga membandingkan kinerja algoritma *naive bayes*, *decision tree*, dan KNN untuk menentukan model yang paling akurat dalam menangani kompleksitas *linguistik* media sosial. Hasil penelitian ini diharapkan memberikan wawasan berharga regulator, penyedia layanan pinjaman *online*, serta berkontribusi pada pengembangan metode analisis sentimen yang lebih baik untuk data berbahasa Indonesia dalam bidang komputasi dan *linguistik* komputasional.

METODE

Penelitian ini menggunakan metode kuantitatif dengan data yang dikumpulkan dari *platform* media sosial X. Penelitian ini mengikuti alur yang terdiri dari 8 tahapan yaitu, *data collection*, *preprocessing*, *labelling*, *modeling*, *feature extraction*, *klasifikasi model*, dan *evaluasi model*. Setiap langkah dirancang untuk memastikan data yang digunakan berkualitas tinggi dan model yang dirancang dapat menghasilkan analisis sentimen yang optimal.



Gambar 1. Tahapan penelitian

Berdasarkan gambar 1, tahapan penelitian ini dimulai dari media sosial x dengan teknik *crawling* melalui pemanfaatan token otentikasi untuk memperoleh *Application Programming Interface* (API) dan *tweet harvest* yang didukung dengan bahasa pemrograman *python*. Proses otentikasi dalam pemanfaatan API untuk memperoleh data dari *platform* x dilakukan penggunaan API untuk memberi akses ke layanan yang disediakan oleh server. Data yang diperoleh berfokus pada penggunaan dengan *tweet* berbahasa Indonesia dengan Kata Kunci yang digunakan ialah pinjaman online, dan pinjaman cepat. Pengumpulan data dilakukan

dalam periode 11 november 2024 sampai dengan 19 februari 2025, yang menghasilkan data dengan total 6715 yang setelah itu disimpan dalam format file CSV. Tahapan berikutnya *Preprocessing* tahap ini meliputi *cleaning data*, *case folding*, normalisasi kata, *tokenizing*, *stopword* dan *stemming*. *Cleaning data* berfungsi untuk membersihkan teks yang dihapus seperti angka, simbol, tanda baca, *uniform resource locator* (URL), mention serta hastag. *Casefolding* dilakukan untuk mengubah seluruh kata menjadi huruf kecil, normalisasi teks bertujuan untuk mengubah kata yang tidak formal menjadi kata yang baku menggunakan kamus kaggle. Selanjutnya, *Tokenizing* digunakan untuk mencegah unit-unit kecil atau token. *Stopword* proses penghapusan kata yang umum yang sering muncul tetapi tidak berkontribusi dalam analisis “atau”, “yang”, “dan”. *Stemming* berfungsi untuk menghilangkan awalan dan akhiran kata dengan library NLP sastrawi (Setiawan & Suryono, 2024).

Proses pelabelan data menggunakan *lexicon-based*, yaitu metode yang memberikan label berdasarkan sentimen positif, negatif, dan netral. Hasil pelabelan menunjukkan 2.010 data berlabel sentimen positif, 2.067 data sentimen negatif, dan 1.864 data sentimen netral. Langkah selanjutnya adalah ekstraksi fitur menggunakan pendekatan *Term Frequency-Inverse Document Frequency* (TF-IDF) metode ini menekankan pada kata-kata yang sering muncul dalam satu dokumen namun jarang muncul di keseluruhan kumpulan dokumen. TF-IDF menetapkan nilai berdasarkan frekuensi penggunaan kata dalam sebuah dokumen dibandingkan dengan frekuensi kemunculannya di seluruh korpus. TF-IDF membantu membedakan kata-kata yang bermakna dari kata-kata umum yang kurang informatif, sehingga memudahkan model dalam mengenali informasi yang paling relevan (Al Rasyid & Ningsih, 2024).

Naïve Bayes menggunakan teknik klasifikasi yang mengkategorikan data berdasarkan fitur-fiturnya menggunakan prinsip probabilitas (Agustia et al., 2025). *Decision Tree* merupakan algoritma yang digunakan untuk membuat keputusan atau prediksi berdasarkan data. Algoritma ini bekerja dengan membagi data secara bertahap melalui serangkaian pertanyaan atau kondisi berdasarkan nilai fitur dalam data. *K-Nearest Neighbor* adalah algoritma pembelajaran mesin berbasis instans yang mengklasifikasikan data baru berdasarkan tingkat kesesuaiannya dengan data yang tersedia, dengan menghitung jarak terdekat.

Tahapan akhir evaluasi, evaluasi model adalah proses penting dalam analisis sentimen. Algoritma yang diterapkan dalam studi ini ialah *Naïve Bayes*, *Decision Tree*, dan *K-Nearest Neighbor*. Kinerja model ini dievaluasi menggunakan *Confusion matrix* yang dapat menghitung seperti *accuracy*, *precision*, *recall* dan *F-1 score*, yang membantu memahami lebih mendalam mengenai performa model dalam mengidentifikasi beberapa kelas secara bersamaan.

HASIL DAN PEMBAHASAN

Hasil

Data penelitian diperoleh dari media sosial x melalui proses *crawling* menggunakan *library tweet harvest* dengan kata kunci “pinjaman online” dan “pinjaman cepat”. Melalui proses ini data dikumpulkan sebanyak 6715 *tweet* kemudian disimpan dalam format CSV untuk memudahkan tahap selanjutnya. Pengelolaan data dilakukan menggunakan *Google Colab* dengan memanfaatkan kode autentikasi API dari media sosial x. dan didukung oleh Bahasa pemrograman *Phyton* dan *stemming* menghapus kata imbuhan menjadi kata dasar menggunakan bahasa NLP sastrawi. Tabel 1 menunjukkan bahwa tahap *preprocessing* mencakup beberapa langkah, yaitu *cleaning* dengan menghapus simbol, tanda baca, dan URL, *case folding* merubah huruf menjadi huruf kecil, normalisasi kata tidak baku, *tokenizing* pemisahan kalimat menjadi kata, *stopword removal* penghapusan kata-kata yang tidak memiliki makna penting, serta *stemming* perubahan kata ke bentuk dasarnya. Seluruh tahapan ini dilakukan untuk menghasilkan dataset yang bersih dan terstruktur. Hasil akhir menunjukkan

bahwa data menjadi lebih terorganisir dan relevan untuk proses analisis sentimen pada tahap selanjutnya.

Tabel 1. Hasil *preprocessing*

Tahapan	Hasil <i>Preprocessing</i>
<i>Cleaning data</i>	Yg mau jd dirtek harus terima klo jobdesk dirteknya dikerjain sama sang ketum raja rangkap jabatan Mungkin belum ada dirtek yg nganggur dan BU serta terbelit pinjol
<i>Case Folding</i>	yg mau jd dirtek harus terima klo jobdesk dirteknya dikerjain sama sang ketum raja rangkap jabatan mungkin belum ada dirtek yg nganggur dan bu serta terbelit pinjol
<i>Normalisasi kata</i>	yang mau jadi dirtek harus terima kalau jobdesk dirteknya dikerjain sama sang ketum raja rangkap jabatan mungkin belum ada dirtek yang nganggur dan bu serta terbelit pinjol
<i>Tokenizing</i>	[yang, mau, jadi, dirtek, harus, terima, kalau, jobdesk, dirteknya, dikerjain, sama, sang, ketum, raja, rangkap, jabatan, mungkin, belum, ada, dirtek, yang, nganggur, dan, bu, serta, terbelit, pinjol]
<i>StopWord</i>	[dirtek, terima, jobdesk, dirteknya, dikerjain, sang, ketum, raja, rangkap, jabatan, dirtek, nganggur, bu, terbelit, pinjol]
<i>Stemming</i>	dirtek terima jobdesk dirteknya dikerjain sang tum raja rangkap jabat dirtek nganggur bu belit pinjol

Setelah melalui tahapan *preprocessing*, data diberi label sentimen positif, negatif dan netral, dengan menggunakan dua library yang populer yaitu pelebelan *lexicon based*. Pelabelan *lexicon based* digunakan untuk secara otomatis menghasilkan label sentimen dengan memanfaatkan kamus kata-kata yang sudah memiliki nilai sentiment ketepatan *lexicon* merupakan gabungan dari Indonesia *sentiment lexicon (Inset)*. Penilaian sentiment dilakukan dengan penjumlahan nilai dari setiap kata skor > 0 dikategorikan positif, 0 dikategorikan sebagai netral dan < 0 dikategorikan negatif. Pendekatan ini praktis untuk proyek tahap awal, karena tidak memerlukan data yang diberi label secara manual, cepat diimplementasikan, serta mudah dipahami, hasil dari pelabelan ditampilkan dalam bentuk tabel 2.

Tabel 2. Hasil pelabelan

Tweet	Pelabelan Data
dirtek terima jobdesk dirteknya dikerjain sang tum raja rangkap jabat dirtek nganggur bu belit pinjol	Negatif
sih emaknya percaya minggu	Positif
harap	Netral

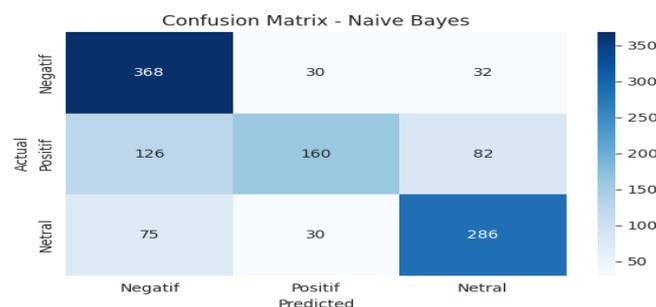
Setelah proses pelabelan selesai, jumlah data berkurang menjadi 5.941 *tweet* terjadi karena penghapusan *tweet* yang terdeteksi sebagai duplikat. Setelah pelabelan, 2.010 *tweet* diklasifikasikan sebagai positif, 2.067 negatif, dan 1.864 sebagai netral. Tahap pengujian dilakukan perbandingan untuk ketiga algoritma 80% untuk data latih dan 20% untuk data uji. Pengujian performa algoritma klasifikasi *naive bayes*, *decision tree*, dan KNN dilakukan menggunakan dataset *lexicon based*. Hasil klasifikasi dari tabel 3 menunjukkan bahwa *naives bayes* mencapai akurasi 68%, *decision tree* memberikan dengan jumlah 69% dan KNN memberikan akurasi 44%. *Decision tree* menunjukkan keunggulan dalam mengklasifikasikan sentimen karena kemampuannya menangani struktur data yang kompleks dan mempertimbangkan antar fitur, sehingga lebih efektif dalam memahami konteks dan variasi

bahasa informal pada media sosial. KNN memiliki performa terendah yang disebabkan oleh keterbatasan dalam memproses data berbasis TF-IDF yang berdimensi tinggi, rendahnya mengidentifikasi bahwa algoritma hanya efektif untuk data yang sangat mirip secara literal, namun gagal dalam konteks yang lebih luas dalam teks.

Tabel 3. Hasil klasifikasi

<i>Modelling</i>	<i>Accuracy</i>	<i>Sentimen</i>	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>
<i>Naives Bayes</i>	68%	Positif	71%	73%	72%
		Negatif	65%	86%	74%
		Netral	73%	43%	54%
<i>Decision Tree</i>	69%	Positif	72%	75%	73%
		Negatif	72%	66%	69%
		Netral	64%	67%	66%
<i>KNN</i>	44%	Positif	96%	31%	47%
		Negatif	76%	14%	23%
		Netral	35%	95%	52%

Penelitian ini dilakukan perbandingan *confusion matrix* dari ketiga algoritma untuk menilai efektivitas setiap model dalam klasifikasi sentimen. *Confusion matrix* memberikan informasi detail mengenai jumlah prediksi yang benar dan salah untuk masing-masing sentimen. Hasil perbandingan *confusion matrix* performa dari setiap algoritma bisa dilihat dalam gambar 2, 3 dan 4 yang menggambarkan perbedaan klasifikasi antara algoritma tersebut.

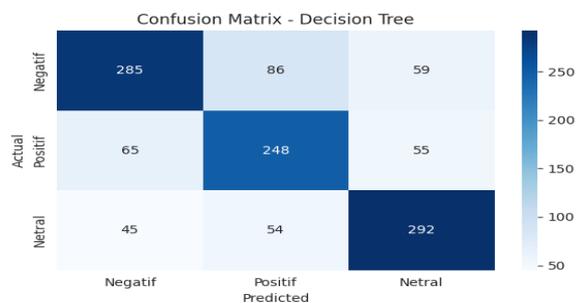


Gambar 2. *Confusion matrix naives bayes*

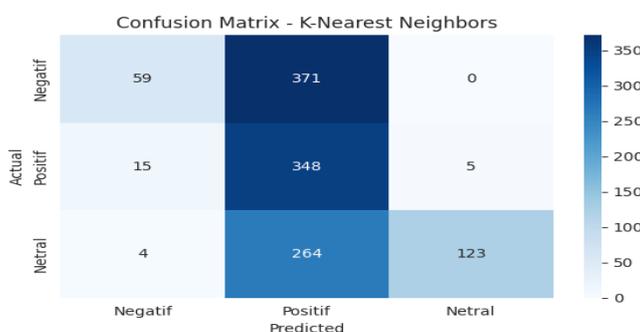
Confusion matrix model klasifikasi *Naive Bayes* pada gambar 2 menunjukkan bahwa berdasarkan semua data yang diuji, model menunjukkan keberhasilan dalam mengklasifikasikan 368 teks ke dalam kategori negatif secara tepat. Namun, terdapat 30 teks tersebut seharusnya masuk kategori positif dan 32 teks yang seharusnya masuk kategori netral, namun salah diklasifikasikan menjadi negatif. Dalam kategori positif, model mampu mengklasifikasikan 160 teks dengan tepat. Namun, sebanyak 126 teks tersebut sebenarnya masuk kategori negatif dan 82 teks dari kategori netral salah diklasifikasikan menjadi positif. Pada kategori netral, model mampu mengidentifikasi 286 teks dengan benar. Namun, terdapat 75 teks dari kategori negatif dan 30 teks dari kategori positif yang tidak tepat diklasifikasikan menjadi netral. Hasil dari *confusion matrix* menunjukkan bahwa model tersebut lebih optimal dalam mengidentifikasi teks netral, menunjukkan akurasi yang cukup baik, namun terdapat kendala dalam mengklasifikasikan tweet positif dan negatif.

Confusion matrix untuk model klasifikasi *decision tree* pada gambar 3 menunjukkan bahwa dari keseluruhan data uji yang digunakan, model menunjukkan keberhasilan mengklasifikasi 285 teks yang digunakan sebagai kategori negatif secara akurat. Namun terdapat 86 teks yang termasuk ke dalam kategori positif dan 59 teks termasuk kedalam

kategori netral yang salah diklasifikasikan menjadi negatif (*false positives*). Dalam kategori positif, model berhasil mengidentifikasi 248 teks dengan benar. Namun terdapat 65 teks yang termasuk ke dalam kategori negatif dan 55 teks termasuk ke dalam kategori netral yang salah diklasifikasikan sebagai (*false positives*). Pada kategori netral, model berhasil mengklasifikasikan 292 teks dengan benar. Namun terdapat 45 teks yang termasuk dalam kategori negatif dan 54 teks yang seharusnya termasuk ke dalam kategori positif yang salah diklasifikasikan sebagai netral.



Gambar 3. *Confusion matrix decision tree*



Gambar 4. *Confusion matrix knn*

Confusion matrix yang dihasilkan model klasifikasi KNN pada gambar 4 menyatakan bahwa model berhasil mengklasifikasikan 59 teks diidentifikasi sebagai negatif dengan benar. Namun 371 teks yang sebenarnya termasuk ke dalam kategori positif yang salah diklasifikasikan sebagai negatif dan 0 teks termasuk ke dalam kategori netral. Pada kategori positif, model berhasil mengklasifikasikan 348 teks dengan tepat. Namun ditemukan 15 teks yang seharusnya termasuk dalam kategori negatif dan 5 teks yang seharusnya masuk ke dalam kategori netral yang salah diklasifikasikan sebagai positif. Dalam kategori netral, menunjukkan keberhasilan dalam mengidentifikasi 123 teks dengan tepat. Namun 4 teks yang sebenarnya termasuk dalam kategori negatif dan 264 teks yang sebenarnya termasuk ke dalam kategori positif yang salah diklasifikasikan sebagai netral.

Kesalahan dalam kategorisasi sentimen positif, negatif, atau netral, dapat mempengaruhi persepsi publik tentang pinjaman online. Jika *tweet* negatif dianggap positif, hal ini dapat memberikan gambaran yang salah tentang penyedia layanan tentang mengenai tentang citra layanan, membuat penyedia lebih baik dari pada kenyataan. Sebaliknya, jika *tweet* positif salah dianggap negatif, hal tersebut dapat menyebabkan persepsi yang lebih buruk dari apa yang sebenarnya terjadi. Selain itu, kesalahan dalam menentukan sentimen netral juga dapat menyebabkan kekeliruan dalam opini public karena *tweet* netral tidak selalu dikategorikan sebagai positif atau negatif.

Berdasarkan Visualisasi *wordcloud* dan frekuensi kata pada gambar 5 menunjukkan bahwa istilah kata seperti “pinjol”, “proses cepat”, dan “modal usaha” Kata-kata tersebut menunjukkan bahwa kebanyakan pengguna membahas kemudahan proses pengajuan dan

Hasil *wordcloud* memperlihatkan kata-kata seperti “pinjol”, “proses cepat”, dan “modal usaha”, mencerminkan fokus pengguna pada kemudahan akses dan kebutuhan ekonomi. Istilah seperti “proses cepat” memiliki makna ganda, sehingga penting dikaitkan dengan hasil klasifikasi sentimen. Kata-kata seperti “bunga”, “galbay”, dan “teror” menunjukkan adanya persepsi negatif dari sebagian pengguna. Secara umum, persepsi masyarakat terhadap pinjaman online bervariasi. Sentimen positif muncul karena manfaat dan kemudahan, tetapi ada pula kekhawatiran terhadap risiko layanan. Hasil ini bernilai akademis dan praktis. Namun, pelabelan berbasis leksikon memiliki keterbatasan dalam menangkap makna ganda dalam bahasa. Jika dibandingkan dengan penelitian sebelumnya, hasilnya cukup berbeda.

Penelitian yang dilakukan Safira & Hasan (2023) menunjukkan bahwa algoritma *naive bayes* memperoleh akurasi 95%, sedangkan Rahmianti (2022) dengan *Decision Tree* memperoleh akurasi 94,27%. Sementara itu, penelitian (Siki et al., 2025) menggunakan KNN mencatat akurasi 62,97% dengan sentimen negatif sebesar 62,29%. Adapun dalam penelitian ini, algoritma KNN hanya mampu mencapai 44%. Perbedaan ini bisa disebabkan oleh variasi dalam metode pelabelan, karakteristik dataset, dan teknik *preprocessing* yang digunakan. Temuan ini menunjukkan bahwa efektivitas algoritma sangat bergantung pada sifat data dan pendekatan analisis yang diterapkan, sehingga pemilihan metode klasifikasi harus disesuaikan dengan karakteristik data agar memperoleh hasil yang optimal. Hasil ini memberikan kontribusi bagi pengembangan analisis sentimen dalam bahasa Indonesia, khususnya dalam meningkatkan akurasi klasifikasi pada data media sosial yang bersifat informal.

SIMPULAN

Penelitian ini menunjukkan bahwa algoritma *decision tree* memberikan kinerja terbaik dalam klasifikasi sentimen pinjaman online di media sosial X dibandingkan *naive bayes* dan KNN, berkat kemampuannya mengenali pola kompleks dan hubungan antar fitur pada data teks pendek, informal, dan kontekstual. Kelemahan KNN terletak pada penanganan data berdimensi tinggi. Analisis *wordcloud* mengungkap topik dominan seperti kemudahan proses, penggunaan modal usaha, serta isu negatif terkait bunga dan gagal bayar. Temuan ini menegaskan pentingnya analisis teks media sosial untuk memahami persepsi publik secara *real-time* dan memilih model klasifikasi yang tepat.

REFERENSI

- A'yuniyah, Q. A., & Reza, M. (2023). Penerapan Algoritma K-Nearest Neighbor Untuk Klasifikasi Jurusan Siswa Di Sma Negeri 15 Pekanbaru. *Indonesian Journal of Informatic Research and Software Engineering (IJIRSE)*, 3(1), 39–45. <https://doi.org/10.57152/ijirse.v3i1.484>
- Agresia, V., Suryono, R. R., Indonesia, U. T., Ratu, L., & Lampung, K. B. (2025). Comparison Of Svm , Naïve Bayes , And Logistic Regression Algorithms For Sentiment Analysis Of Fraud And Bots In Purchasing Concert Ticket. *INOVTEK Polbeng - Seri Informatika*, 10(2), 592-602.
- Agustia, D. N., Suryono, R. R., Indonesia, U. T., Ratu, L., & Lampung, K. B. (2025). Comparison Of Naïve Bayes , Random Forest , And Logistic Regression Algorithms For Sentiment Analysis Online Gambling. *INOVTEK Polbeng – Seri Informatika*. 10(1), 28 4-295. <https://Doi.Org/10.35314/Prk93630>
- Al Rasyid, R., & Ningsih, D. H. U. (2024). Penerapan Algoritma TF-IDF dan Cosine Similarity untuk Query Pencarian Pada Dataset Destinasi Wisata. *Jurnal JTik (Jurnal Teknologi Informasi Dan Komunikasi)*, 8(1), 170–178. <https://doi.org/10.35870/jtik.v8i1.1416>
- Arvante, J. Z. Y. (2022). Dampak Permasalahan Pinjaman Online dan Perlindungan Hukum Bagi Konsumen Pinjaman Online. *Ikatan Penulis Mahasiswa Hukum Indonesia Law Journal*, 2(1), 73–87. <https://doi.org/10.15294/ipmhi.v2i1.53736>

- Azmi, A. F. (2024). Prediksi Churn Nasabah Bank Menggunakan Klasifikasi Random Forest Dan Decision Tree Dengan Evaluasi Confusion Matrix. *Komputa : Jurnal Ilmiah Komputer Dan Informatika*, 13(1), 111-119. <https://doi.org/10.34010/komputa.v13i1.1263>
- Ismail, A. R., & Raden Bagus Fajriya Hakim. (2023). Implementasi Lexicon Based Untuk Analisis Sentimen Dalam Menentukan Rekomendasi Pantai Di DI Yogyakarta Berdasarkan Data Twitter. *Emerging Statistics and Data Science Journal*, 1(1), 37–46. <https://doi.org/10.20885/esds.vol1.iss.1.art5>
- Pamungkas, A. S., & Cahyono, N. (2024). Analisis Sentimen Review ChatGPT di Play Store menggunakan Support Vector Machine dan K-Nearest Neighbor. *Edumatic : Jurnal Pendidikan Informatika*, 8(1), 1–10. <https://doi.org/10.29408/edumatic.v8i1.24114>
- Prayudi, S., & Nasution, W. (2020). Ragam Bahasa Dalam Media Sosial Twitter: Kajian Sociolinguistik. *Jurnal Metamorfosa*. <https://doi.org/10.46244/metamorfosa.v8i2.1140>
- Rahmianti, I. (2022). Analisis Kelayakan Pemberian Kredit Koperasi Dengan Metode Data Mining Decision Tree. *Jurnal Informatika dan Rekayasa Elektronik*. 5(2). 153-161. <https://doi.org/10.36595/jire.v5i2.663>
- Ramli, M., Apriyanto, M., Azhar, A., Puspitasari, F., & KMS. Fikri, N. S. (2023). Dampak Konsumen Terhadap Pinjaman Online (PINJOL). *PaKMas: Jurnal Pengabdian Kepada Masyarakat*, 3(1), 52-58. <https://doi.org/10.54259/pakmas.v3i1.1638>
- Riyanah, N., & Fatmawati, F. (2021). Penerapan Algoritma Naive Bayes Untuk Klasifikasi Penerima Bantuan Surat Keterangan Tidak Mampu. *JTIM : Jurnal Teknologi Informasi Dan Multimedia*, 2(4), 206–213.
- Safira, A. & Hasan, N, Firman. (2023). Analisis Sentimen Masyarakat Terhadap Paylater Menggunakan Metode Naive Bayes Classifier. *Zonasi: Jurnal Sistem Informasi*. 5(1). 59-70.
- Setiawan, A., & Suryono, R. R. (2024). Analisis Sentimen Ibu Kota Nusantara menggunakan Algoritma Support Vector Machine dan Naïve Bayes. *Edumatic: Jurnal Pendidikan Informatika*, 8(1), 183– 192. <https://doi.org/10.29408/edumatic.v8i1.25667>
- Siki, Y.C.H., Tokan, T.B.A., Manehat, D.J., Ngaga, E. and Mau, S.D.B. 2025. Perbandingan Metode Naïve Bayes dan K-Nearest Neighbor Terhadap Sentimen Analisis Pinjaman Online. *Jurnal Media Informatika*. 6(3), 1614-1622. <https://doi.org/10.55338/jumin.v6i3.5687>
- Silalahi, N., & Guidio Leonarde Ginting. (2023). Rekomendasi Berita Berkaitan dengan Menerapkan Algoritma Text Mining dan TF-IDF. *Bulletin of Computer Science Research*, 3(4), 276–282. <https://doi.org/10.47065/bulletincsr.v3i4.266>
- Sinambela, D. P., Naparin, H., Zulfadhilah, M., & Hidayah, N. (2023). Implementasi Algoritma Decision Tree dan Random Forest dalam Prediksi Perdarahan Pascasalin. *Jurnal Informasi Dan Teknologi*, 5(3), 58–64. <https://doi.org/10.60083/jidt.v5i3.393>
- Sugangga, R., & Sentoso, E. H. (2020). Perlindungan Hukum Terhadap Pengguna Pinjaman Online (Pinjol) Ilegal. *Justice Journal Of Law*, 01, 47-61. <https://doi.org/10.33751/pajoul.v1i1.2050>
- Supriyadi, A. (2023). Perbandingan Algoritma Naive Bayes dan Decision Tree(C4.5) dalam Klasifikasi Dosen Berprestasi. *Generation Journal*, 7(1), 39-49. <https://doi.org/10.29407/gj.v7i1.19797>
- Wati, D., & Syahfitri, T. (2022). Dampak Pinjaman Online Bagi Masyarakat. *Community Development Journal : Jurnal Pengabdian Masyarakat*, 2(3), 1181- 1184 <https://doi.org/10.31004/cdj.v2i3.2950>
- Wijayanti, S. (2022). Dampak Aplikasi Pinjaman Online Terhadap Kebutuhan Dan Gaya Hidup Konsumtif Buruh Pabrik. *MIZANIA: Jurnal Ekonomi Dan Akuntansi*, 2(2), 230-235. <https://doi.org/10.47776/mizania.v2i2.592>