

IMPLEMENTASI LLM GEMINI 2.0 FLASH DAN NAÏVE BAYES UNTUK KLASIFIKASI RATING ULASAN SIREKAP 2024

Muhamad Ridwan, Nauval Azmi Verdikha, Kuncoro Hadi, Novariyani Nur Ramadhani

Teknik Informatika, Universitas Muhammadiyah Kalimantan Timur
Jl. Ir. H. Juanda No. 15 Kota Samarinda 75124, Kalimantan Timur, Indonesia
mr624@umkt.ac.id

ABSTRAK

Aplikasi Sistem Informasi Rekapitulasi (SIREKAP) 2024 memegang peranan krusial dalam mendukung transparansi, efisiensi, dan akuntabilitas penyelenggaraan pemilihan umum di Indonesia melalui ulasan digital pengguna. Namun, tingginya volume ulasan yang tidak terstruktur serta masalah ketidakseimbangan data yang ekstrem membuat proses analisis sentimen maupun evaluasi kualitas aplikasi secara manual menjadi sangat tidak efisien. Oleh karena itu, penelitian ini bertujuan untuk mengotomatisasi klasifikasi ulasan pengguna SIREKAP 2024 ke dalam lima kelas rating menggunakan pendekatan pembelajaran mesin. Metodologi yang diterapkan mencakup pra-proses teks berbasis Large Language Model (LLM) Gemini 2.0 Flash untuk merapikan ejaan dan tata bahasa, ekstraksi fitur Term Frequency-Inverse Document Frequency (TF-IDF), serta pengklasifikasian menggunakan algoritma Multinomial Naïve Bayes yang dioptimasi parameternya dengan Optuna. Evaluasi keandalan model dilakukan melalui teknik 10-Fold Cross Validation dan Confusion Matrix terhadap dataset sebanyak 8.358 ulasan. Hasil pengujian menunjukkan rata-rata F1-score sebesar 31,00%. Kinerja model yang terbatas ini utamanya disebabkan oleh dominasi ulasan pada kelas Rating 1 yang mencapai 64,1%, sehingga model mengalami bias terhadap kelas mayoritas dan kesulitan mengklasifikasikan kelas minoritas. Temuan ini menyoroti pentingnya penanganan ketidakseimbangan data untuk penelitian selanjutnya.

Kata kunci : *TF-IDF, Naïve Bayes, Klasifikasi, MultinomialNB, SIREKAP, F1-Score.*

1. PENDAHULUAN

Bagi Indonesia sebagai negara kepulauan digitalisasi pemilu menjadi solusi efektif untuk menjangkau daerah terpencil, menekan biaya operasional, dan meningkatkan partisipasi publik melalui akses informasi real-time [1]. Salah satu inovasi utama yang mencerminkan transformasi ini adalah aplikasi Sistem Informasi Rekapitulasi Elektronik (SIREKAP). SIREKAP 2024 muncul sebagai aplikasi digital yang dirancang oleh Komisi Pemilihan Umum (KPU) untuk merekam, memproses, dan mengirimkan hasil suara secara langsung dari Tempat Pemungutan Suara (TPS) aplikasi ini dibuat sebagai langkah untuk meningkatkan transparansi, efisiensi, dan akurasi dalam penyelenggaraan pemilu di Indonesia [2].

Dalam konteks aplikasi SIREKAP 2024, rating memiliki peran krusial sebagai cerminan kepercayaan publik terhadap kualitas, keandalan, dan akuntabilitas sistem digital dalam mendukung transparansi pemilu [3]. Aplikasi SIREKAP 2024 menghadapi tantangan dalam hal keakuratan data, kendala sistem, serta responsivitas antarmuka sebagaimana aplikasi e-government lain yang juga mengalami masalah serupa dalam memenuhi ekspektasi pengguna di platform digital [4].

Pasca penyelenggaraan Pemilu 2024, aplikasi SIREKAP menuai berbagai respons dari masyarakat yang diekspresikan melalui ribuan ulasan di platform distribusi digital seperti Google Play Store. Volume data ulasan yang besar dan tidak terstruktur ini menjadi tantangan untuk dianalisis padahal di dalamnya terkandung wawasan mengenai persepsi publik,

keluhan teknis, dan apresiasi pengguna [5]. Oleh karena itu diperlukan pendekatan analisis teks berbasis Natural Language Processing (NLP) yang mampu mengolah data ulasan dalam jumlah besar secara otomatis dan sistematis.

Perkembangan Natural Language Processing (NLP) ditandai dengan hadirnya Large Language Model (LLM) yaitu model berskala besar yang mampu memahami konteks dan pola bahasa dari data teks dalam jumlah besar. LLM banyak digunakan dalam pemrosesan teks tidak terstruktur karena kemampuannya dalam melakukan normalisasi bahasa dan klasifikasi teks [6]. Dalam penelitian ini implementasi Gemini 2.0 Flash digunakan pada tahapan pra-proses untuk analisis sentimen dan klasifikasi teks mengingat efektivitasnya yang telah terbukti dalam klasifikasi teks seperti yang ditunjukkan dalam deteksi berita palsu [7].

Dalam analisis teks sebuah algoritma klasifikasi tidak dapat memproses data dalam bentuk kata-kata mentah melainkan memerlukan input berupa format numerik. Sebab itu metode Term Frequency-Inverse Document Frequency (TF-IDF) diterapkan sebagai sebuah teknik untuk mengubah teks menjadi angka sekaligus mengukur tingkat kepentingan dari setiap kata [8]. Sebuah kata dianggap penting jika sering muncul dalam satu dokumen spesifik (Term Frequency), namun pada saat yang sama kata tersebut juga jarang ditemukan pada dokumen-dokumen lainnya (Inverse Document Frequency). Dalam penerapannya metode ini berhasil mencapai akurasi sebesar 86.48% pada sebuah studi kasus klasifikasi teks [9].

Penelitian pada [10] mengenai analisis sentimen ulasan aplikasi MyPertamina didapatkan tingkat akurasi 75% menggunakan metode Naïve Bayes dibandingkan dengan metode K-Nearest Neighbor yang memiliki tingkat akurasi berkisar antara 56% sampai 73%. Penelitian lain pada [11] yang membandingkan tiga metode didapatkan hasil lebih tinggi metode Naïve Bayes dengan akurasi sebesar 79,8% dibandingkan algoritma K-Nearest Neighbor sebesar 50,23% dan Support Vector Machine sebesar 75,29%. Berdasarkan kedua penelitian tersebut, terlihat pola yang konsisten bahwa metode Naïve Bayes menunjukkan performa akurasi yang lebih unggul dibandingkan algoritma klasifikasi lain dalam studi kasus klasifikasi teks.

Penelitian ini menguji efektivitas penerapan pra-proses menggunakan model LLM Gemini 2.0 Flash, ekstraksi fitur TF-IDF, dan algoritma Naïve Bayes untuk klasifikasi ulasan aplikasi SIREKAP 2024, data ulasan yang digunakan dalam penelitian ini diambil dari penelitian terdahulu yang dilakukan oleh [12]. Pendekatan ini bertujuan untuk menguji apakah pra-proses dengan LLM menghasilkan kinerja yang berbeda secara signifikan dibandingkan metode pra-proses manual. Keberhasilan model akan dievaluasi berdasarkan metrik f1-score..

2. TINJAUAN PUSTAKA

Proses pengklasifikasian ulasan pengguna aplikasi SIREKAP 2024, akan dibagi ke dalam beberapa pembahasan utama. Pembahasan tersebut dimulai dari pemahaman mengenai klasifikasi, LLM Gemini hingga konsep dasar pemrosesan bahasa alami (Natural Language Processing) yang mencakup tahapan pra-pemrosesan, ekstraksi fitur Term Frequency-Inverse Document Frequency (TF-IDF), dan algoritma Naïve Bayes.

2.1. Klasifikasi

Klasifikasi adalah proses menemukan model atau fungsi yang menjelaskan dan membedakan kelas data serta konsepnya. Proses ini bertujuan untuk mengelompokkan data ke dalam kategori tertentu berdasarkan pola atau karakteristik yang ada. Model yang terbentuk dari data berlabel kemudian digunakan untuk memprediksi kelas dari data baru yang belum diketahui [17].

2.2. Natural Language Processing (NLP)

Natural Language Processing adalah cabang kecerdasan buatan yang memungkinkan komputer memahami dan mengolah bahasa manusia. NLP berperan dalam memproses teks ulasan melalui tahapan – tahapan untuk mempersiapkan data yang akan dianalisis oleh model klasifikasi.

2.3. LLM Gemini

Gemini adalah Large Language Model yang dikembangkan oleh Google dan mampu memahami konteks bahasa secara mendalam. Dalam penelitian ini

gemini digunakan pada tahapan pra – proses untuk membantu normalisasi teks dan perbaikan struktur kalimat sehingga data ulasan menjadi lebih terstruktur sebelum dilakukan klasifikasi.

2.4. Term Frequency-Inverse Document Frequency (TF-IDF)

TF-IDF adalah metode ekstraksi fitur yang digunakan untuk mengubah data teks menjadi representasi numerik yang dapat digunakan oleh algoritma machine learning dengan cara mengukur seberapa penting sebuah kata dalam dokumen terhadap keseluruhan data, sehingga kata kata yang bersifat lebih informatif memiliki kontribusi yang lebih besar dalam proses klasifikasi. Pada penelitian yang dilakukan oleh [13] menunjukkan bahwa TF-IDF mampu merepresentasikan karakteristik teks secara lebih efektif dibandingkan pendekatan Bag-Of-Words dalam klasifikasi teks.

TF-IDF sering dipadukan dengan algoritma Naïve Bayes karena kemampuannya dalam merepresentasikan data teks berdimensi tinggi dan bersifat sparse. Pada penelitian [14] membuktikan bahwa penggunaan TF-IDF sebagai metode ekstraksi fitur dapat meningkatkan performa klasifikasi ketika dipadukan dengan algoritma Naïve Bayes. Hasil serupa juga ditunjukkan oleh [15] yang menerapkan TF-IDF dan Naïve Bayes pada ulasan layanan teknologi di mana kombinasi ini terbukti efektif dalam proses klasifikasi.

Implementasi TF-IDF dalam penelitian ini dilakukan menggunakan Python dengan bantuan library scikit-learn melalui TfidfVectorizer. Prosesnya diawali dengan tahap pra-pemrosesan teks, kemudian teks diubah menjadi matriks TF-IDF yang merepresentasikan tingkat kepentingan setiap kata dalam dokumen. Representasi numerik ini selanjutnya digunakan sebagai input bagi algoritma Naïve Bayes untuk melakukan klasifikasi maupun analisis teks lainnya. Representasi numerik berupa matriks TF-IDF ini kemudian digunakan sebagai input pada tahap klasifikasi teks seperti yang juga dilakukan pada penelitian terdahulu [16].

2.5. Naïve Bayes

Metode Naïve Bayes adalah algoritma klasifikasi yang berbasis probabilitas dan menganggap bahwa fitur – fitur saling independen satu terhadap yang lainnya. Metode ini menggunakan prinsip Teorema Bayes untuk menghitung probabilitas bahwa sebuah ulasan masuk ke dalam kategori tertentu berdasarkan distribusi kata dalam teks, keunggulan dari pendekatan ini adalah kemampuannya untuk memproses data teks dengan cepat dan tepat khususnya pada data yang berukuran besar.

3. METODE PENELITIAN

Prosedur penelitian ini terdiri dari beberapa tahapan sistematis yang dilakukan untuk memperoleh hasil klasifikasi rating aplikasi SIREKAP 2024

berdasarkan analisis ulasan pengguna. Adapun alur metode penelitian dapat dilihat pada Gambar 1.



Gambar 1. Tahapan Penelitian

Metodologi yang diterapkan dalam penelitian untuk mengklasifikasikan ulasan pengguna aplikasi SIREKAP 2024 mencakup jenis penelitian, sumber data, teknik pengumpulan data, hingga tahapan pemrosesan data yang meliputi pra-pemrosesan, ekstraksi fitur, klasifikasi, dan evaluasi model.

3.1. Input Data

Input data diperoleh dari ulasan pengguna terhadap aplikasi SIREKAP 2024 yang bersumber dari hasil penelitian sebelumnya. Ulasan tersebut dikumpulkan dari Google Play Store menggunakan teknik web scraping kemudian di ekstraksi dan disimpan dalam format file CSV melalui prompt pada Gemini 2.0 Flash. Data yang digunakan dalam penelitian ini sebanyak 8.358 ulasan aplikasi.

3.2. Tahapan Pra-Proses

Tahapan pra – proses dilakukan menggunakan Gemini 2.0 Flash karena kemampuannya memahami konteks bahasa indonesia tidak baku pada ulasan aplikasi SIREKAP. Model digunakan untuk memperbaiki kesalahan ejaan, menormalkan bahasa, menghilangkan noise seperti simbol dan emoji, serta merapikan struktur kalimat. Proses diawali dengan konfigurasi API dan penetapan model dilengkapi dengan manajemen cache berbasis JSON untuk efisiensi pemanggilan layanan serta protokol prompt guna menjaga output. Tahapan ini meliputi lowercase, remove unnecessary character, spellchecker, dan stemming agar data menjadi lebih bersih dan siap digunakan pada tahapan selanjutnya.

3.3. Ekstraksi Fitur TF-IDF

TF-IDF merupakan metode ekstraksi fitur yang mengubah data teks menjadi representasi numerik dengan mengukur tingkat kepentingan kata dalam dokumen terhadap keseluruhan data sehingga kata

yang lebih informatif memiliki bobot lebih besar dalam proses klasifikasi. Metode ini sering dipadukan dengan Naïve Bayes karena dapat merepresentasikan data teks berdimensi tinggi dan bersifat sparse, serta terbukti meningkatkan performa klasifikasi pada berbagai penelitian sebelumnya. Dalam penelitian ini, TF-IDF diimplementasikan menggunakan TfidfVectorizer dari library Scikit-learn untuk menghasilkan matriks bobot kata yang selanjutnya digunakan sebagai input pada proses klasifikasi Naïve Bayes.

Rumus yang akan digunakan untuk ekstraksi fitur TF-IDF ada pada persamaan (1) dan (2)

$$B_{xy} = tf(x, y) \times idf(x) \quad (1)$$

Keterangan:

B_{xy} = bobot pada *term* x pada dokumen y

tf_{xy} = frekuensi kemunculan *term* x pada dokumen y

Rumus matematis untuk menghitung IDF :

$$idf(x) = \log \frac{1+n}{1+df_x} + 1 \quad (2)$$

Keterangan:

idf_x = nilai Inverse Document Frequency dari kata x

df_x = jumlah dokumen yang mengandung kata x

n = jumlah total dokumen

log = fungsi logaritma

Rumus matematis untuk TF-IDF :

$$v_{norm} = \frac{v}{\|v\|_2} = \frac{v}{\sqrt{v_1^2 + v_2^2 + v_3^2 + \dots + v_n^2}} \quad (3)$$

Keterangan:

v_{norm} v_{norm} = Vektor hasil normalisasi.

v v = Vektor TF-IDF asli dari dokumen

$\|v\|_2$ $\|v\|_2$ = Jumlah kuadrat nilai TF-IDF dari semua dokumen.

3.4. Klasifikasi Naïve Bayes

Klasifikasi data adalah proses menemukan model atau fungsi yang menjelaskan dan membedakan kelas data serta konsepnya. Proses ini bertujuan untuk mengelompokkan data ke dalam kategori tertentu berdasarkan pola atau karakteristik yang ada. Model yang terbentuk dari data berlabel kemudian digunakan untuk memprediksi kelas dari data baru yang belum diketahui [17].

Dalam penelitian ini algoritma yang digunakan adalah Naïve Bayes Multinomial karena metode ini merupakan algoritma klasifikasi berbasis probabilitas sederhana yang banyak digunakan dalam klasifikasi teks karena bekerja berdasarkan frekuensi kemunculan kata dalam setiap data [18].

Rumus untuk klasifikasi Naïve Bayes terdapat pada persamaan (4):

$$y^{\wedge} = \underset{y}{\operatorname{argmax}} \left(\log P(y) + \sum_i x_i \log \frac{N_{iy+\alpha}}{\sum_j N_{jy+\alpha|V|}} \right) \quad (4)$$

Keterangan:

$y^{\wedge}y^{\wedge}$ = Kelas hasil prediksi yang memiliki nilai probabilitas terbesar

$argmax_y argmax_y$ = Proses pemilihan kelas y yang memaksimalkan nilai fungsi probabilitas

$P(y)P(y)$ = Probabilitas awal(prior) dari kelas y

$logP(y)logP(y)$ = Logaritma dari probabilitas prior kelas

$x_i x_i$ = Bobot fitur ke – i (berasal dari TF-IDF)

$N_{iy} N_{iy}$ = Jumlah kemunculan fitur ke – i pada kelas y

$\sum_j N_{jy} \sum_j N_{jy}$ = Total kemunculan seluruh fitur pada kelas y

$|V||V|$ = Ukuran kosakata atau jumlah seluruh fitur

$\alpha \alpha$ = Parameter Laplace smoothing yang digunakan untuk menghindari probabilitas 0

3.5. Klasifikasi Naïve Bayes

Evaluasi dalam penelitian ini bertujuan untuk mengukur kinerja model Naïve Bayes dalam mengklasifikasikan ulasan aplikasi SIREKAP 2024, salah satu metode yang digunakan adalah Confusion Matrix yang berguna untuk membandingkan hasil prediksi dengan data aktual serta K-Fold Cross Validation untuk menilai akurasi model. K-Fold Cross Validation merupakan teknik pengambilan sampel ulang data yang membagi dataset menjadi K bagian dengan ukuran yang sama kemudian dilakukan proses latih dan uji secara berulang sebanyak K kali sehingga setiap bagian data berperan sebagai data uji secara bergantian [20].

4. HASIL DAN PEMBAHASAN

Pembahasan secara mendalam berdasarkan hasil pengujian dan evaluasi model yang telah dipaparkan. Pembahasan difokuskan pada interpretasi kinerja algoritma Naïve Bayes yang dikombinasikan dengan ekstraksi fitur TF-IDF dalam mengklasifikasikan ulasan pengguna aplikasi SIREKAP 2024 ke dalam lima kelas rating.

4.1. Hasil Analisis Data

Evaluasi dalam penelitian ini bertujuan untuk mengukur kinerja model Naïve Bayes dalam mengklasifikasikan ulasan aplikasi SIREKAP 2024, salah satu metode yang digunakan adalah Confusion Matrix yang berguna untuk membandingkan hasil prediksi dengan data aktual serta K-Fold Cross Validation untuk menilai akurasi model. K-Fold Cross Validation merupakan teknik pengambilan sampel ulang data yang membagi dataset menjadi K bagian dengan ukuran yang sama kemudian dilakukan proses latih dan uji secara berulang sebanyak K kali sehingga setiap bagian data berperan sebagai data uji secara bergantian [20].

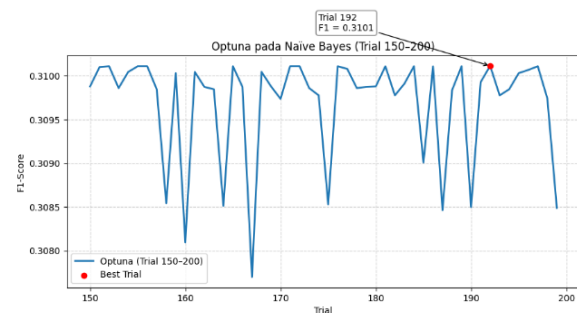
Tabel 1. Distribusi Ulasan SIREKAP 2024 Berdasarkan Rating

Rating	Jumlah Ulasan	Presentase
1	5.361	64,1%
2	736	8,8%
3	478	5,7%
4	255	3,1%
5	1.528	18,3%
Total	8.358	100%

Data Tabel 1 didominasi secara ekstrem oleh pengguna yang memberikan Rating 1, yakni sebanyak 5.361 ulasan atau 64,1% dari total populasi data. Jika digabungkan dengan Rating 2 (8,8%), maka total sentimen negatif mencapai 72,9%. Selain tingginya Rating 1, pada Rating 5 yang menempati posisi terbanyak dengan 1.528 ulasan (18,3%). Di sisi lain, penilaian menengah yang cenderung netral atau moderat seperti Rating 3 (5,7%) dan Rating 4 (3,1%) justru sangat minim.

4.2. Optimasi Parameter

Dilakukan proses optimasi hyperparameter dengan metode optuna untuk menguji beberapa nilai α dalam rentang tertentu dan memilih nilai α yang menghasilkan performa terbaik berdasarkan metrik evaluasi yang digunakan [19]. Setelah dilakukan optimasi model menunjukkan peningkatan meskipun relative kecil. Trial terbaik terdapat pada trial ke 192 dengan nilai α sebesar 0.0009 dan menghasilkan rata – rata f1-score tertinggi 0.3101. Grafik dari optimasi hyperparameter dapat dilihat pada gambar 2.



Gambar 2. Grafik Optimasi Hyperparameter

Berdasarkan Gambar 2 grafik pengujian Optuna pada rentang trial 150 hingga 200, pencarian kombinasi hyperparameter menghasilkan performa tertinggi pada Trial 192 dengan nilai F1-Score sebesar 0,3101 (31,01%). Nilai ini menandakan titik puncak (global optimum) dari pencarian parameter yang dieksplorasi oleh algoritma berbasis Bayesian Optimization.

F1-Score sepanjang 50 trial terakhir ini bergerak dalam rentang yang sangat terbatas, yaitu dari titik terendah 0,3077 (pada Trial 167) hingga titik tertinggi 0,3101 (pada Trial 192). Selisih puncak dan lembah grafik hanya sebesar 0,0024 (0,24%). Kurva yang cenderung mendatar dan berulang (plateauing) ini

mengindikasikan bahwa pemodelan telah mengalami kejenuhan (saturation).

Penyesuaian nilai hyperparameter seperti smoothing parameter pada Naïve Bayes maupun parameter ekstraksi TF-IDF tidak lagi mampu memberikan dorongan peningkatan performa.

4.3. Hasil Evaluasi

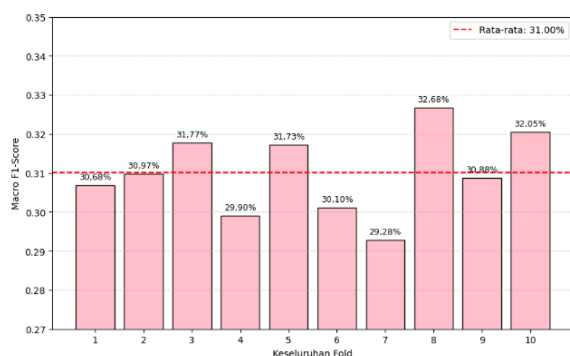
Untuk menguji konsistensi dan stabilitas kinerja model Naïve Bayes dengan nilai alpha terpilih dilakukan evaluasi menggunakan metode k-fold cross validation. Hasil pengujian pada setiap fold yang menunjukkan nilai f1- score pada masing – masing fold dapat dilihat pada tabel 2.

Berdasarkan grafik pengujian Optuna pada rentang trial 150 hingga 200, pencarian kombinasi hyperparameter menghasilkan performa tertinggi pada Trial 192 dengan nilai F1-Score sebesar 0,3101 (31,01%).

Tabel 2. Hasil Evaluasi Setiap Fold

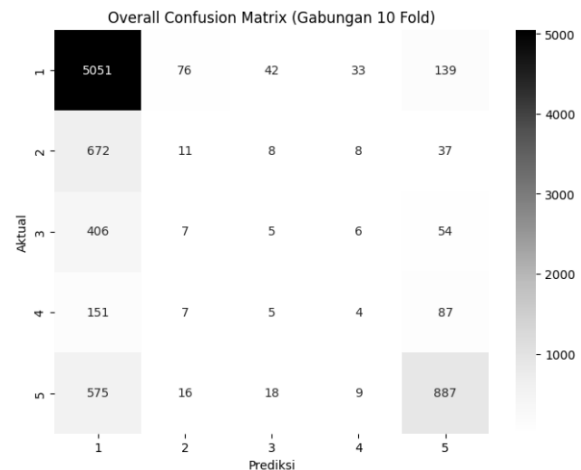
Fold	F1-Score
1	30,66%
2	30,97%
3	31,77%
4	29,90%
5	31,73%
6	30,10%
7	29,28%
8	32,68%
9	30,88%
10	32,05%
Rata-rata	31,00%

Berdasarkan hasil Tabel 2 pengujian nilai evaluasi tertinggi terdapat pada fold ke-8 dengan nilai 32,68% sedangkan nilai terendah pada fold ke-7 dengan nilai 29,28% perbedaan nilai antar fold ini menunjukkan adanya variasi karakteristik data pada setiap pembagian namun secara keseluruhan rentang nilai yang dihasilkan antar fold relatif kecil sehingga performa model Naïve Bayes Multinomial dapat dikatakan cukup stabil meskipun distribusi data dan pola kata antar fold berbeda – beda. Berikut pada gambar 3 juga dipaparkan hasil dalam bentuk diagram.



Gambar 3. Diagram Rata – Rata Hasil Fold

Berdasarkan grafik Gambar 3 model memiliki rata – rata f1-score sebesar 31.00% dengan variasi performa yang tidak terlalu ekstrem antar fold. Hal ini menunjukkan bahwa hasil evaluasi model cukup stabil. Untuk melihat kinerja model secara lebih rinci dilakukan analisis menggunakan confusion matrix untuk menunjukkan perbandingan antara label aktual dan label hasil prediksi yang dihasilkan. Setiap baris pada confusion matrix merepresentasikan kelas aktual sedangkan setiap kolom merepresentasikan kelas hasil prediksi.



Gambar 4. Hasil Confusion Matrix

Dapat dilihat pada gambar 4 menunjukkan bahwa model memiliki performa yang tidak seimbang antar kelas di mana kelas 1 mendominasi data dan juga hasil prediksi terlihat dari jumlah prediksi benar yang sangat tinggi yaitu 5051 data. Kondisi ini menyebabkan model cenderung mempelajari pola kata yang paling sering muncul pada kelas tersebut sehingga mengakibatkan sebagian besar data dari kelas lain khususnya kelas 2,3, dan 4 yang memiliki jumlah data lebih sedikit cenderung diklasifikasikan sebagai kelas 1.

5. KESIMPULAN DAN SARAN

Berdasarkan hasil penelitian yang telah dilakukan dapat disimpulkan bahwa metode Naïve Bayes Multinomial berhasil diterapkan dalam proses klasifikasi ulasan pengguna aplikasi SIREKAP 2024. Proses klasifikasi dilakukan melalui tahapan pra-proses teks berbasis LLM Gemini 2.0 Flash untuk membersihkan data kemudian menggunakan ekstraksi fitur TF-IDF agar data teks dapat direpresentasikan dalam bentuk numerik dan siap untuk diklasifikasi.

Evaluasi menggunakan 10-fold cross validation model menghasilkan rata – rata nilai f1-score sebesar 31,00%. Nilai tersebut menunjukkan bahwa model memiliki kinerja yang relatif konsisten namun masih menghadapi tantangan dalam menangani variasi konteks bahasa dan ketidakseimbangan distribusi kelas pada data ulasan.

DAFTAR PUSTAKA

- [1] N. Fransco et al., "Pemanfaatan Teknologi dalam melakukan Pengawasan Pemilu : Meningkatkan Transparansi dan Kepercayaan Publik," pp. 616–628, 2024.
- [2] M. Degly, A. Wea, and A. Muljanto, "Implementasi Aplikasi Sirekap Pada Pilkada (Pemilihan Kepala Daerah) 2024 Di Kabupaten Sidoarjo," vol. 4, no. 2, pp. 3362–3370, 2025.
- [3] E. Cintanesa et al., "Analisis Penggunaan Sistem Informasi Rekapitulasi (SIREKAP) pada Pemilihan Umum di Kota Tanjungpinang Tahun 2024," no. 1, 2025.
- [4] A. I. Tanggraeni and M. N. N. Sitokdana, "Analisis Sentimen Aplikasi E-Government pada Google Play Menggunakan Algoritma Naïve Bayes," JATISI (Jurnal Tek. Inform. dan Sist. Informasi), vol. 9, no. 2, pp. 785–795, 2022, doi: 10.35957/jatisi.v9i2.1835.
- [5] M. R. Hanafi and R. K. R., "Sentiment Analysis on Sirekap App Reviews on Google Play Using Naive Bayes Algorithm Analisis Sentimen pada Ulasan Aplikasi Sirekap di Google Play Menggunakan Algoritma Naive Bayes," vol. 4, no. October, pp. 1578–1586, 2024.
- [6] A. T. U. B. Lubis, "QUESTION ANSWERING SYSTEM MENGGUNAKAN LARGE," no. Llm, 2024.
- [7] N. Agustina, A. Adrian, and M. Hermawati, "Implementasi Algoritma Naïve Bayes Classifier untuk Mendeteksi Berita Palsu pada Sosial Media," Fakt. Exacta, vol. 14, no. 4, p. 206, 2022, doi: 10.30998/faktorexacta.v14i4.11259.
- [8] D. Septiani, I. Isabela, F. Teknik, U. N. Jakarta, F. Teknik, and U. N. Jakarta, "Analisis term frequency inverse document frequency (tf-idf) dalam temu kembali informasi pada dokumen teks," vol. 25, pp. 81–88, 2023.
- [9] R. Alfawas, T. I., Rahim, A., & Rudiman, "Naive-Bayes," IJCAI 2001 Work. Empir. methods Artif. Intell., vol. 4, no. January, pp. 41–46, 2024, [Online]. Available: <http://www.cc.gatech.edu/home/isbell/classes/reading/papers/Rish.pdf>
- [10] H. Taufiqurrahman, F. Tri Anggraeny, and M. Muharrom Al Haromainy, "Perbandingan Algoritma Naïve Bayes Dan K-Nearest Neighbor Pada Analisis Sentimen Ulasan Aplikasi Mypertamina," JATI (Jurnal Mhs. Tek. Inform., vol. 7, no. 6, pp. 3934–3939, 2024, doi: 10.36040/jati.v7i6.7801.
- [11] M. H. Asnawi, I. Firmansyah, R. Novian, and R. S. Pontoh, "Perbandingan algoritma naïve bayes, k-nn, dan svm dalam pengklasifikasian sentimen media sosial," Pros. Semin. Nas. Stat., vol. 10, no. June 2023, pp. 20–20, 2021, doi: 10.1234/pns.v10i.85.
- [12] R. Ridhoi, N. Azmi Verdikha, and F. Yulianto, "Analisis Klasifikasi Ulasan Aplikasi Sirekap 2024 menggunakan Ekstraksi Fitur DistilBert Dan Metode Support Vector Machine," J. Ilm. Inform., no. 15, 2025.
- [13] A. Davina, M. Putri, N. Sulistianingsih, and R. Rismayati, "JTIM : Jurnal Teknologi Informasi dan Multimedia Pengaruh Teknik Representasi Teks Bag-of-Words dan TF-IDF," vol. 7, no. 4, pp. 675–688, 2025.
- [14] Ardiansyah and Kurniawan, "Optimasi Metode Naïve Bayes Classifier Menggunakan Pendekatan Term Frequency-Inverse Document Frequency (TF-IDF) Pada Analisis Sentimen," vol. 7, no. 3, pp. 458–464, 2024.
- [15] S. Khoerunnisa, D. F. Shiddiq, and D. Nurhayati, "Application of the Naive Bayes Algorithm with TF-IDF and Cross Validation Techniques for Sentiment Analysis Towards Starlink Penerapan Algoritma Naive Bayes dengan Teknik TF-IDF dan Cross Validation untuk Analisis Sentimen Terhadap Starlink," vol. 5, no. April, pp. 566–577, 2025.
- [16] H. Purnomo et al., "Klasifikasi Sentimen Ulasan Pengguna MyPertamina Menggunakan Algoritma Naive Bayes dan TF-IDF 1, 2," pp. 801–810, 2025.
- [17] N. B. Putri and A. W. Wijayanto, "Analisis Komparasi Algoritma Klasifikasi Data Mining Dalam Klasifikasi Website Phishing," Komputika J. Sist. Komput., vol. 11, no. 1, pp. 59–66, 2022, doi: 10.34010/komputika.v11i1.4350.
- [18] A. Gerliandeva, Y. H. Chrisnanto, and H. Ashaury, "Optimasi Klasifikasi Sentimen pada Komentar Online menggunakan Multinomial Naïve Bayes dan Ekstraksi Fitur TF-IDF serta N-grams Optimization of Sentiment Classification on Online Comments using Multinomial Naïve Bayes and TF-IDF Feature Extraction and N-grams," no. X, pp. 259–272, 2024, doi: 10.56873/jpkm.v9i2.5585.
- [19] M. Ridwan and E. Utami, "Optimized Hyperparameter Tuning for Improved Hate Speech Detection," vol. 5, no. 158, pp. 525–534, 2026.
- [20] A. I. Pradana and V. Atina, "Teknik K-Fold Cross Validation untuk Mengevaluasi Kinerja Mahasiswa," pp. 239–248, 2024, doi: 10.33364/algoritma/v.21-1.1618.