



# Klasifikasi Teks Komentar Pengguna Aplikasi Access By Kai di Google Play Store Menggunakan Metode Naïve Bayes

Fahrul Afandi\*, Rakhmat Kurniawan

Fakultas Sains dan Teknologi, Ilmu Komputer, Universitas Islam Negeri Sumatera Utara, Medan, Indonesia

Email: <sup>1,\*</sup>fahrulafandi886@gmail.com, <sup>2</sup>rakhmat.kr@uinsu.ac.id

Email Penulis Korespondensi: fahrulafandi886@gmail.com

**Abstrak**—Kemajuan teknologi informasi telah memengaruhi berbagai aspek kehidupan, termasuk transportasi. Aplikasi Access by KAI menyediakan layanan pemesanan tiket kereta api secara digital. Dengan jutaan pengguna, analisis tingkat kepuasan melalui ulasan di Google Play Store menjadi penting untuk meningkatkan kualitas layanan. Penelitian ini bertujuan mengklasifikasikan ulasan pengguna menggunakan algoritma Naïve Bayes untuk menentukan tingkat kepuasan, mengelompokkan ulasan berdasarkan kategori tertentu, serta mengevaluasi akurasi hasil klasifikasi. Penelitian menggunakan metode Naive bayes untuk mengklasifikasi teks dimana pengumpulan data ulasan dilakukan terlebih dahulu melalui proses scraping dari Google Play Store dengan total 1000 ulasan. Data dianalisis melalui tahap pre-processing seperti cleaning, case folding, tokenisasi, normalisasi, stopwords, stemming dan pemberian label sentimen menggunakan InSetLexicon. Selanjutnya, ulasan dikelompokkan berdasarkan kategori menggunakan metode clustering K-Means untuk mengelompokkan data menjadi tiga kategori, yaitu Fitur, Layanan, dan Sistem, guna meningkatkan akurasi klasifikasi diikuti klasifikasi menggunakan Naïve Bayes. Evaluasi dilakukan menggunakan confusion matrix untuk mengukur akurasi, presisi, recall, dan F1-score. Hasil klasifikasi menunjukkan bahwa pada kategori Fitur, precision adalah 79%, recall 99%, dan F1-score 88%. Pada kategori Layanan, precision mencapai 100%, recall 56%, dan F1-score 72%. Untuk kategori Sistem, precision adalah 94%, recall 68%, dan F1-score 79%. Secara keseluruhan, model mencapai akurasi sebesar 83%. Manfaat penelitian ini adalah memberikan pemahaman mendalam terhadap kebutuhan pengguna dan menjadi acuan pengembang untuk meningkatkan layanan aplikasi Access by KAI.

**Kata Kunci:** Access by KAI; Google Play Store; Klasifikasi Teks; K-Means; Naïve Bayes

**Abstract**—The advancement of information technology has influenced various aspects of life, including transportation. The Access by KAI application provides digital train ticket booking services. With millions of users, analyzing the level of satisfaction through reviews on the Google Play Store is important to improve service quality. This study aims to classify user reviews using the Naïve Bayes algorithm to determine the level of satisfaction, group reviews based on certain categories, and evaluate the accuracy of the classification results. The study uses the Naive Bayes method to classify text where review data collection is carried out first through a scraping process from the Google Play Store with a total of 1000 reviews. Data is analyzed through pre-processing stages such as cleaning, case folding, tokenization, normalization, stopwords, stemming and sentiment labeling using InSetLexicon. Furthermore, reviews are grouped by category using the K-Means clustering method to group data into three categories, namely Features, Services, and Systems, to improve classification accuracy followed by classification using Naïve Bayes. Evaluation is carried out using a confusion matrix to measure accuracy, precision, recall, and F1-score. The classification results show that in the Feature category, precision is 79%, recall 99%, and F1-score 88%. In the Service category, precision reaches 100%, recall 56%, and F1-score 72%. For the System category, precision is 94%, recall 68%, and F1-score 79%. Overall, the model achieves an accuracy of 83%. The benefits of this study are to provide a deep understanding of user needs and become a reference for developers to improve the Access by KAI application service.

**Keywords:** Access by KAI; Google Play Store; K-Means; Naïve Bayes; Text Classification

## 1. PENDAHULUAN

Dalam era digital kemajuan dan perkembangan Teknologi Informasi dan Komunikasi (TIK) sangat memengaruhi kehidupan manusia. Aplikasi ponsel pintar telah menjadi komponen penting dari kehidupan sehari-hari, termasuk menggunakan layanan transportasi. Transportasi merupakan usaha dan kegiatan mengangkut atau membawa barang atau penumpang dari satu tempat ke tempat lain [1]. Salah satu contoh nyata adalah aplikasi access by KAI yang merupakan platform resmi dari PT Kereta Api Indonesia (KAI) yang memudahkan pengguna dalam memesan tiket dan mendapatkan layanan informasi terkait perjalanan kereta api. Sebagai aplikasi yang digunakan oleh jutaan pengguna, evaluasi terhadap Tingkat kepuasan pengguna menjadi sangat penting untuk memastikan kualitas layanan dan meningkatkan performa aplikasi.

Sebagai aplikasi yang telah digunakan oleh jutaan pengguna, evaluasi terhadap tingkat kepuasan pengguna menjadi sangat penting guna memastikan kualitas layanan dan meningkatkan performa aplikasi. Aplikasi ini dapat dengan mudah diunduh melalui Google Play Store oleh pengguna perangkat berbasis Android. Google Play Store sendiri tidak hanya menyediakan sarana distribusi aplikasi, tetapi juga memungkinkan pengguna memberikan ulasan dan penilaian terhadap pengalaman mereka menggunakan suatu aplikasi [2]. Ulasan-ulasan pada aplikasi tersebut berperan sebagai sumber data penting dalam memahami kualitas kepuasan pengguna. Namun dengan banyaknya ulasan yang tersedia hal ini sering kali membuat proses analisis secara manual menjadi sulit dan memakan banyak waktu. Oleh karena itu, diperlukan sebuah metode otomatis untuk mengklasifikasikan ulasan-ulasan tersebut.

Untuk mengatasi permasalahan tersebut, klasifikasi teks dapat digunakan sebagai solusi. Klasifikasi teks merupakan teknik dalam Pemrosesan Bahasa Alami (Natural Language Processing/NLP) yang digunakan untuk mengelompokkan teks ke dalam kategori tertentu [3]. Tujuan dari klasifikasi teks adalah untuk menempatkan teks atau dokumen ke dalam kelas atau kategori yang sama berdasarkan pola atau fitur tertentu yang ada dalam teks tersebut. Proses



ini sangat bermanfaat untuk menganalisis data teks dalam jumlah besar, seperti ulasan pengguna, artikel, email, maupun posting media sosial. Pada aplikasi seperti Google Play Store, klasifikasi teks dapat diterapkan untuk menganalisis ulasan pengguna dan mengelompokkannya berdasarkan kategori tertentu, misalnya berdasarkan tingkat kepuasan, jenis keluhan, atau saran perbaikan. Teknik ini membantu perusahaan untuk memahami dengan lebih cepat dan efisien apa yang dirasakan oleh penggunanya, sehingga bisa mengambil tindakan yang tepat untuk memperbaiki layanan atau produk.

Salah satu algoritma klasifikasi adalah Naïve Bayes. Klasifikasi Naive Bayes adalah salah satu algoritma klasifikasi yang dapat digunakan untuk memprediksi kemungkinan keanggotaan kelas [4]. Klasifikasi ini didasarkan pada teorema Bayes dan dapat bekerja dengan sangat cepat dan akurat dalam database dengan dataset yang besar. Dalam proses ini, kata-kata kunci seperti "bagus", "kecewa", atau "cantik" dapat dianggap relevan. Dan untuk mengubah teks menjadi representasi numerik dari frekuensi kata, model klasifikasi dapat dipelajari dengan menghitung seberapa sering kata muncul dalam teks.

Berdasarkan penelitian sebelumnya yang dilakukan oleh Riyanah & Fatmawati, (2021) dimana dapat disimpulkan bahwa hasil akhir pengolahan data dengan menggunakan aplikasi Rapid Miner dari 35 data sample warga penerima bantuan SKTM dan 1 data testing dikategorikan layak dengan nilai akurasi keseluruhan 62.86% dan Class recall layak 78.57%, Class recall tidak layak 52.38%, Class precision layak 52.38% dan Class precision tidak layak 78.57% [5]. Dari hasil tersebut, terlihat bahwa meskipun akurasi belum terlalu tinggi, Naïve Bayes mampu memberikan hasil yang cukup signifikan dalam aspek precision, menjadikannya relevan untuk digunakan dalam klasifikasi opini pengguna.

Selanjutnya penelitian yang dilakukan oleh Adyatma Subagja, dkk (2021) dimana mengklasifikasikan ulasan aplikasi jenius pada google play store menggunakan algoritma naive bayes dan dapat diperoleh dengan pengujian model Naive Bayes Classifier menghasilkan akurasi sebesar 0.57, hasil precision dan recall terbesar terdapat pada class rating 1, sebesar 0,78 dan 0,87. Performa dari Naive Bayes Classifier dalam mengklasifikasikan ulasan pengguna pada aplikasi Jenius kurang baik, karena akurasi yang dihasilkan relatif kecil [6].

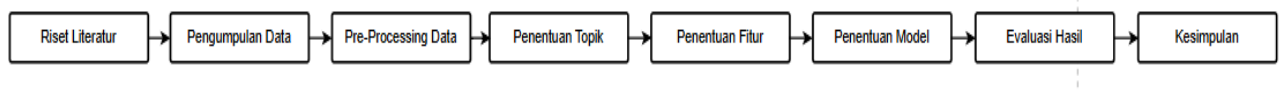
Dan penelitian yang dilakukan oleh Nurzaman, dkk (2024) dimana dengan implementasi Naïve Bayes, hasil evaluasi menunjukkan akurasi sebesar 63.61%, dengan 307 True Positif, 74 True Negatif, 26 False Positif, dan 192 False Negatif. Dengan tingkat precision sebesar 92.19% dan recall sebesar 61.52% [7].

Dari penelitian-penelitian terdahulu, ditemukan bahwa algoritma Naïve Bayes telah banyak diterapkan dalam berbagai konteks, namun belum secara khusus digunakan untuk mengklasifikasikan ulasan pengguna aplikasi Access by KAI. Selain itu, sebagian besar penelitian sebelumnya masih terbatas pada pengujian akurasi tanpa mengeksplorasi aspek spesifik dari ulasan pengguna transportasi kereta api di Indonesia. GAP ini menunjukkan adanya peluang untuk mengeksplorasi dan mengembangkan model klasifikasi sentimen berbasis Naïve Bayes secara lebih terfokus pada domain transportasi publik, khususnya aplikasi Access by KAI.

Berdasarkan uraian dan hasil penelitian terdahulu, maka penelitian ini dilakukan dengan tujuan untuk menganalisis sentimen ulasan pengguna terhadap aplikasi Access by KAI yang tersedia di Google Play Store dengan memanfaatkan algoritma Naïve Bayes. Algoritma ini dipilih karena mampu mengelola data teks dalam jumlah besar dengan efisiensi tinggi dan akurasi yang memadai, serta telah terbukti efektif dalam penelitian serupa. Penelitian ini diharapkan dapat memberikan manfaat bagi pengembang aplikasi dalam memahami kebutuhan dan pengalaman pengguna secara lebih mendalam, sekaligus menjadi masukan dalam peningkatan kualitas layanan. Novelty dari penelitian ini terletak pada penerapan algoritma Naïve Bayes secara spesifik pada ulasan aplikasi Access by KAI, yang sejauh ini masih jarang dibahas dalam literatur, serta penggunaan pendekatan klasifikasi sentimen untuk mengidentifikasi tingkat kepuasan dan masukan pengguna secara otomatis dan sistematis.

## 2. METODOLOGI PENELITIAN

Kerangka penelitian adalah suatu rancangan yang menjelaskan secara garis besar alur sebuah penelitian dengan menggunakan gambar sesuai dengan tahapan yang dilakukan dalam penelitian. Hal ini dilakukan agar tahapan penelitian lebih terstruktur. Penelitian ini menggunakan metode kuantitatif yaitu penelitian yang menguji teori dengan meneliti hubungan antara variabel. Berikut ini adalah kerangka penelitian yang diajukan oleh peneliti yaitu sebagai berikut.



**Gambar 1.** Kerangka Penelitian

Pada Gambar 1. Terdapat beberapa tahapan penelitian yang dilakukan yaitu dengan riset literatur, pengumpulan data, pre-processing data, penentuan topik, penentuan fitur, penentuan model dan penerapan algoritma naive bayes serta melakukan pengujian atau evaluasi dengan menghitung nilai akurasi, presisi, recall dan f-1 score.

### 2.1 Riset Literatur

Proses penelitian ini dimulai dengan riset literatur yaitu tahapan dalam memahami dari teori, metode dan pendekatan yang relevan dengan penelitian [8]. Pada tahap ini akan dilakukan studi terhadap berbagai sumber yang akurat seperti



buku, jurnal serta penelitian terdahulu yang berkaitan dengan text mining, analisis sentimen, Natural Language Processing (NLP) dan Model machine Learning yang akan digunakan.

KAI Access adalah aplikasi resmi milik PT Kereta Api Indonesia (PT. KAI) yang diluncurkan sejak 4 September 2014. Aplikasi ini dirancang untuk memenuhi kebutuhan penumpang kereta lokal, komuter, dan jarak jauh dalam hal pemesanan tiket, perubahan jadwal, serta pembatalan keberangkatan [9]. Meskipun terus mengalami peningkatan layanan dan dapat diakses melalui berbagai platform (iOS, Android, Windows), sebagian pengguna masih memilih pembelian tiket secara manual karena berbagai alasan [10].

Text mining adalah satu langkah dari analisis teks yang dilakukan secara otomatis oleh komputer untuk menggali informasi berkualitas dari suatu rangkaian teks yang terangkum dalam sebuah dokumen [11]. Tujuan utama text mining adalah menemukan pola pada data yang dapat digunakan oleh manusia untuk membantu pekerjaannya. Proses text mining menggunakan teks atau dokumen sebagai sumber data [12]. Salah satu langkah penting dalam text mining adalah klasifikasi teks, yaitu mengelompokkan dokumen ke dalam kategori tertentu menggunakan algoritma machine learning seperti Naive Bayes, SVM, KNN, Decision Tree, dan lainnya [13]. Representasi teks biasanya dikonversi ke bentuk numerik menggunakan teknik seperti Bag of Words (BoW), TF-IDF, atau Word2Vec [14]. Dalam penelitian ini, digunakan teknik TF-IDF untuk mengekstraksi fitur dari data ulasan pengguna.

Sementara itu, analisis sentimen merupakan metode dalam Natural Language Processing (NLP) yang digunakan untuk mengidentifikasi opini pengguna dan mengklasifikasikannya ke dalam sentimen positif maupun negatif terhadap produk atau layanan tertentu [15].

## 2.2 Pengumpulan Data

Pengumpulan data dilakukan secara langsung dari sumber utama, yaitu ulasan pengguna aplikasi Access by KAI yang dipublikasikan di Google Play Store. Peneliti mengakses platform tersebut untuk memperoleh data aktual dan relevan dari pengguna yang benar-benar menggunakan aplikasi. Sebanyak 1.000 ulasan dikumpulkan menggunakan teknik web scraping atau melalui API resmi Google Play Store (jika tersedia). Tahapan pengumpulan meliputi identifikasi sumber data, ekstraksi ulasan, serta pra-pemrosesan data (pre-processing) seperti normalisasi teks, penghapusan karakter khusus, dan pembersihan elemen yang tidak diperlukan untuk analisis lebih lanjut.

## 2.3 Pre-Processing Data

Pada Langkah selanjutnya melakukan proses pre-processing data yaitu diawali dengan proses cleaning pada dataset, proses ini dilakukan untuk menghapus elemen-elemen yang tidak relevan dalam teks, seperti simbol, angka dan karakter khusus yang tidak memberikan informasi berguna dalam analisis [16]. serta case folding untuk menyeragamkan huruf menjadi format kecil. Selanjutnya, dilakukan tokenisasi (tokenizing) guna memecah teks menjadi unit kata, normalisasi untuk menyamakan kata tidak baku ke bentuk standar, dan stopword removal untuk menghapus kata-kata umum yang tidak memiliki makna signifikan. Tahapan terakhir adalah stemming, yaitu mengubah kata ke bentuk dasar dengan menghilangkan imbuhan. Proses ini bertujuan untuk menghasilkan data teks yang bersih, konsisten, dan siap untuk dianalisis dalam klasifikasi maupun analisis sentimen.

## 2.4 Penentuan Topik

Setelah proses pra-pemrosesan, langkah selanjutnya adalah penentuan topik utama berupa kategori sentimen, yaitu positif dan negatif, serta aspek-aspek tertentu dalam ulasan. Pada penelitian ini, penentuan topik dilakukan dengan menganalisis kata kunci pada dataset ulasan pengguna aplikasi Access by KAI, dan menghitung probabilitas kemunculannya dalam kategori positif dan negatif menggunakan pendekatan klasifikasi Naïve Bayes. Hal ini bertujuan untuk menentukan kecenderungan sentimen pengguna terhadap aplikasi.

## 2.5 Penentuan Fitur

Tahap selanjutnya adalah penentuan fitur, yaitu proses mengubah teks mentah menjadi representasi numerik yang dapat diproses oleh model klasifikasi. Pada dataset yang telah dikumpulkan, peneliti menerapkan metode TF-IDF untuk mengekstraksi fitur dari 1.000 ulasan pengguna aplikasi. Proses ini menghasilkan representasi numerik dari teks yang digunakan sebagai input untuk algoritma klasifikasi [17]. TF-IDF menghitung bobot setiap kata berdasarkan frekuensi kemunculannya dalam suatu dokumen serta seberapa jarang kata tersebut muncul di seluruh dokumen [18]. Kata yang memiliki frekuensi tinggi pada satu dokumen namun jarang muncul di dokumen lain akan memiliki bobot lebih besar, sehingga lebih berpengaruh dalam proses klasifikasi.

## 2.6 Penentuan Model

Setelah penentuan fitur, langkah berikutnya adalah pemilihan model klasifikasi. Penelitian ini menggunakan algoritma Naive Bayes [19] pada data ulasan pengguna yang telah melalui proses pre-processing dan ekstraksi fitur TF-IDF. Model ini digunakan untuk mengklasifikasikan setiap ulasan ke dalam kategori positif atau negatif [20]. Model ini didasarkan pada Teorema Bayes dengan asumsi independensi antar fitur, sehingga memungkinkan proses klasifikasi yang efisien dan efektif pada data berukuran besar. Selain itu, K-Means juga diimplementasikan sebagai metode bantu untuk mengelompokkan ulasan berdasarkan pola kata yang serupa sebelum dilakukan klasifikasi. Sebelum melakukan klasifikasi sentimen, dilakukan proses TF-IDF (Term Frekuensi - Inverse Document Frekuensi) untuk mengekstraksi fitur

dari teks. TF-IDF merupakan metode yang mengukur seberapa penting suatu kata dalam dokumen dibandingkan dengan keseluruhan dataset.

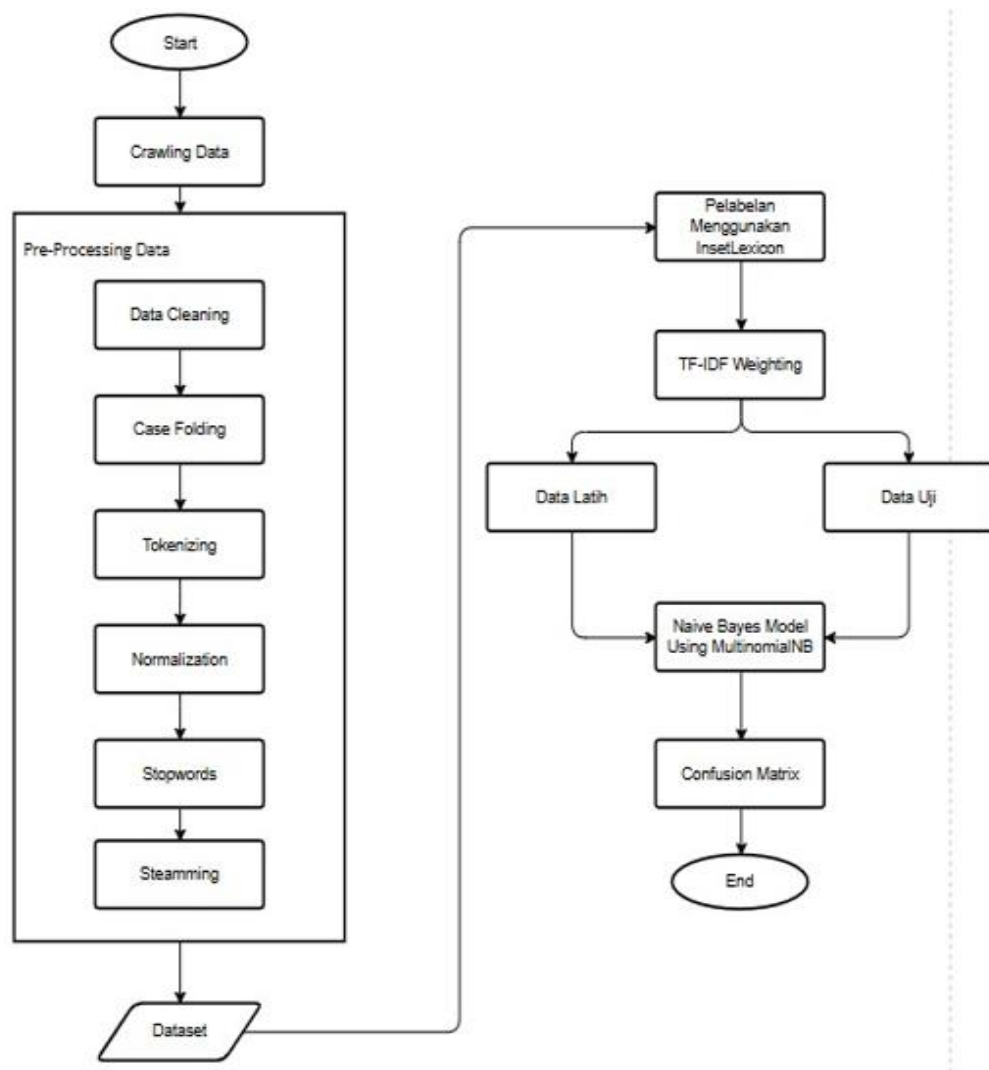
$$tf - idf = tf \times \log \frac{N}{df} \quad (1)$$

Dimana *Term Frequency* (TF) merupakan ukuran yang menunjukkan seberapa sering suatu kata muncul dalam sebuah dokumen. Sementara itu, *Inverse Document Frequency* (IDF) digunakan untuk menilai tingkat kepentingan kata tersebut di seluruh dokumen. Nilai IDF diperoleh dari perbandingan antara jumlah total dokumen ( $n$ ) dengan jumlah dokumen yang mengandung kata tersebut ( $df$ ). Semakin sedikit dokumen yang memuat kata tertentu, maka semakin tinggi nilai IDF-nya. Kombinasi TF dan IDF menghasilkan bobot TF-IDF yang merepresentasikan tingkat kepentingan kata dalam suatu dokumen secara proporsional terhadap keseluruhan korpus. Setelah itu dilakukan klasifikasi data dengan menggunakan metode Naïve Bayes Classifier, Dapat dirumuskan sebagai berikut:

$$P(H|X) = \frac{P(X|H) \times P(H)}{P(X)} \quad (2)$$

Dimana dalam konteks klasifikasi dengan metode *Naïve Bayes*, misalkan  $X$  merupakan data yang kelasnya belum diketahui, dan  $H$  adalah hipotesis bahwa  $X$  termasuk ke dalam kelas tertentu. Probabilitas  $P(H|X)$  disebut sebagai *posterior probability*, yaitu peluang hipotesis  $H$  benar diberikan data  $X$ . Nilai ini dihitung berdasarkan  $P(X|H)$ , yaitu peluang terjadinya data  $X$  jika diketahui hipotesis  $H$  benar (*likelihood*), dikalikan dengan  $P(H)$ , yaitu probabilitas awal hipotesis  $H$  sebelum data diamati (*prior probability*), dan dibagi dengan  $P(X)$  yang merupakan probabilitas data  $X$  secara keseluruhan (*prior probability* dari data).

Algoritma Naive Bayes dapat dipahami dengan menggunakan rumus conditional probabilities (probabilitas bersyarat). Flowchart dapat mempermudah dalam menemukan dan menyelesaikan suatu masalah khususnya pada masalah yang perlu dievaluasi dan diperelajari lebih dalam. Untuk alur prosesnya dapat dilihat pada Gambar 2 flowchart berikut.



Gambar 2. Flowchart Klasifikasi



## 2.7 Evaluasi Hasil

Pada tahap evaluasi, hasil prediksi model dianalisis menggunakan confusion matrix untuk menghitung metrik evaluasi seperti precision, recall, F1-score, dan accuracy. Precision mengukur proporsi prediksi positif yang benar, sementara recall menghitung proporsi data positif yang berhasil terdeteksi dengan benar oleh model. F1-score merupakan rata-rata harmonis antara precision dan recall, yang mencerminkan keseimbangan antara keduanya. Accuracy digunakan untuk mengukur tingkat keseluruhan prediksi yang benar. Evaluasi ini memberikan gambaran menyeluruh terhadap performa model dalam mengklasifikasikan sentimen ulasan pengguna.

## 3. HASIL DAN PEMBAHASAN

Dalam penelitian ini, data ulasan pengguna aplikasi Access by KAI diambil dari Google Play Store menggunakan teknik web scraping, dengan hasil sebanyak 1000 komentar yang disimpan dalam format file .csv. Ulasan ini mencerminkan opini masyarakat mengenai aplikasi tersebut dan selanjutnya diberi label positif dan negatif menggunakan kamus Insetlexicon serta algoritma Naïve Bayes untuk mengevaluasi accuracy, precision, recall, dan F1-Score. Sebelum dilakukan klasifikasi dengan Naïve Bayes, data terlebih dahulu dikelompokkan melalui clustering menggunakan model K-Means. Proses clustering ini bertujuan untuk mengelompokkan ulasan berdasarkan pola tertentu, seperti frekuensi kata dan kesamaan konten, menghasilkan tiga kluster: (1) Kluster 1 mengelompokkan ulasan terkait fitur aplikasi, (2) Kluster 2 mengelompokkan ulasan terkait layanan aplikasi, dan (3) Kluster 3 mengelompokkan ulasan terkait sistem aplikasi. Hasil clustering ini menjadi dasar untuk tahap analisis sentimen menggunakan Naïve Bayes, dengan menggunakan tools Jupyter Notebook. Contoh data komentar yang digunakan ditampilkan pada Tabel 1 berikut, yang terdiri dari 1000 data ulasan.

**Tabel 1.** Sampel Komentar

| No.  | Username  | score | at                     | content                                                                                                                                                                                                                                                                                                                                                                                                                                                        |
|------|-----------|-------|------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 1    | Xazandra  | 1     | 2024-09-23<br>21:01:10 | jelek banget aplikasinya, isi saldo tapi gak masuk saldo ke aplikasi KAI Access dan dari e-banking sudah sukses transfernya. perbaiki sistemnya buat orang rugi saja!                                                                                                                                                                                                                                                                                          |
| ...  | ...       | ...   | ...                    | ...                                                                                                                                                                                                                                                                                                                                                                                                                                                            |
| 1000 | Dhanu1243 | 2     | 2024-09-23<br>15:07:53 | Update aplikasi ke Access by KAI malah jadi gak bisa lihat boarding pass di HP ðŸŽŹ Gak bisa juga menampilkan keseluruhan stasiun di satu kota, misal milih Seluruh Stasiun di Jakarta, tidak kayak dulu. Masih bagus KAI Access yg lama, lebih simpel, enak dilihat, lengkap & mudah. Tolong bisa ditambahkan metode pembayaran : - Kartu kredit semua Bank (saat ini hanya debit Mandiri) - Gopay, Shopee Pay (saat ini hanya Ovo, itupun juga sering error) |

Langkah selanjutnya adalah drop data yang tidak dipakai seperti “username” dan “at” sehingga didapat seperti Tabel 2 berikut.

**Tabel 2.** Sampel Data

| No.  | Content                                                                                                                                                                                                                                                                                                                                                                                                                                                        |
|------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 1    | jelek banget aplikasinya, isi saldo tapi gak masuk saldo ke aplikasi KAI Access dan dari e-banking sudah sukses transfernya. perbaiki sistemnya buat orang rugi saja!                                                                                                                                                                                                                                                                                          |
| ...  | ...                                                                                                                                                                                                                                                                                                                                                                                                                                                            |
| 1000 | Update aplikasi ke Access by KAI malah jadi gak bisa lihat boarding pass di HP ðŸŽŹ Gak bisa juga menampilkan keseluruhan stasiun di satu kota, misal milih Seluruh Stasiun di Jakarta, tidak kayak dulu. Masih bagus KAI Access yg lama, lebih simpel, enak dilihat, lengkap & mudah. Tolong bisa ditambahkan metode pembayaran : - Kartu kredit semua Bank (saat ini hanya debit Mandiri) - Gopay, Shopee Pay (saat ini hanya Ovo, itupun juga sering error) |

Setelah melakukan proses penghapusan data yang tidak relevan (drop), langkah berikutnya dalam penelitian ini adalah pre-processing data. Adapun langkah pre-processing yang dilakukan adalah cleaning, case folding, Normalisasi, Tokenizing, Stopwords Removal, dan Stemming.

### 3.1 Pre-Processing

#### a. Teks Cleaning

Bertujuan untuk membersihkan data komentar yang sudah diperoleh. Komponen-komponen yang dibersihkan merupakan komponen yang tidak berarti atau tidak relevan untuk proses pengklasifikasian data. Menghilangkan link atau emotikon yang ada. Dalam kasus klasifikasi analisis sentimen, Contoh data sebelum dan sesudah proses cleaning dapat dilihat pada Tabel 3 berikut.

**Tabel 3.** Teks Cleaning

| No.  | Full text                                                                                                                                                                                                                                                                                                                                                                                                                                                  | Cleaning                                                                                                                                                                                                                                                                                                                                                                                                                                   |
|------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 1    | jelek banget aplikasinya, isi saldo tapi gak masuk saldo ke aplikasi KAI Access dan dari e-banking sudah sukses transfernya. perbaiki sistemnya buat orang rugi saja!                                                                                                                                                                                                                                                                                      | jelek banget aplikasinya isi saldo tapi gak masuk saldo ke aplikasi KAI Access dan dari ebanking sudah sukses transfernya perbaiki sistemnya buat orang rugi saja                                                                                                                                                                                                                                                                          |
| ...  | ...                                                                                                                                                                                                                                                                                                                                                                                                                                                        | ...                                                                                                                                                                                                                                                                                                                                                                                                                                        |
| 1000 | Update aplikasi ke Access by KAI malah jadi gak bisa lihat boarding pass di HP. Gak bisa juga menampilkan keseluruhan stasiun di satu kota, misal milih Seluruh Stasiun di Jakarta, tidak kayak dulu. Masih bagus KAI Access yg lama, lebih simpel, enak dilihat, lengkap & mudah. Tolong bisa ditambahkan metode pembayaran : - Kartu kredit semua Bank (saat ini hanya debit Mandiri) - Gopay, Shopee Pay (saat ini hanya Ovo, itupun juga sering error) | Update aplikasi ke Access by KAI malah jadi gak bisa lihat boarding pass di HP. Gak bisa juga menampilkan keseluruhan stasiun di satu kota misal milih Seluruh Stasiun di Jakarta tidak kayak dulu. Masih bagus KAI Access yg lama lebih simpel enak dilihat lengkap mudah. Tolong bisa ditambahkan metode pembayaran. Kartu kredit semua Bank saat ini hanya debit Mandiri. Gopay Shopee Pay saat ini hanya Ovo. itupun juga sering error |

## b. Teks Case Folding

Case folding adalah proses mengonversi semua huruf dalam teks menjadi huruf kecil untuk menyamakan bentuknya. Berikut penerapan case folding pada Tabel 4.

**Tabel 4.** Case Folding

| No.  | Cleaning                                                                                                                                                                                                                                                                                                                                                                                                                                   | Case Folding                                                                                                                                                                                                                                                                                                                                                                                                                               |
|------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 1    | jelek banget aplikasinya isi saldo tapi gak masuk saldo ke aplikasi KAI Access dan dari ebanking sudah sukses transfernya perbaiki sistemnya buat orang rugi saja                                                                                                                                                                                                                                                                          | jelek banget aplikasinya isi saldo tapi gak masuk saldo ke aplikasi kai access dan dari ebanking sudah sukses transfernya perbaiki sistemnya buat orang rugi saja                                                                                                                                                                                                                                                                          |
| ...  | ...                                                                                                                                                                                                                                                                                                                                                                                                                                        | ...                                                                                                                                                                                                                                                                                                                                                                                                                                        |
| 1000 | Update aplikasi ke Access by KAI malah jadi gak bisa lihat boarding pass di HP. Gak bisa juga menampilkan keseluruhan stasiun di satu kota misal milih Seluruh Stasiun di Jakarta tidak kayak dulu. Masih bagus KAI Access yg lama lebih simpel enak dilihat lengkap mudah. Tolong bisa ditambahkan metode pembayaran. Kartu kredit semua Bank saat ini hanya debit Mandiri. Gopay Shopee Pay saat ini hanya Ovo, itupun juga sering error | update aplikasi ke access by kai malah jadi gak bisa lihat boarding pass di hp. gak bisa juga menampilkan keseluruhan stasiun di satu kota misal milih seluruh stasiun di jakarta tidak kayak dulu. masih bagus kai access yg lama lebih simpel enak dilihat lengkap mudah. tolong bisa ditambahkan metode pembayaran. kartu kredit semua bank saat ini hanya debit mandiri. gopay shopee pay saat ini hanya ovo. itupun juga sering error |

## c. Tokenizing

Teks dapat dibagi menjadi kata-kata atau simbol yang terkait dengan aplikasi Access by KAI. Hasilnya seperti pada Tabel 5 berikut.

**Tabel 5.** Teks Tokenizing

| No.  | Case Folding                                                                                                                                                                                                                                                                                                                                                                                                                               | Tokenize                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                |
|------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 1    | jelek banget aplikasinya isi saldo tapi gak masuk saldo ke aplikasi kai access dan dari ebanking sudah sukses transfernya perbaiki sistemnya buat orang rugi saja                                                                                                                                                                                                                                                                          | ['jelek', 'banget', 'aplikasinya', 'isi', 'saldo', 'tapi', 'gak', 'masuk', 'saldo', 'ke', 'aplikasi', 'kai', 'access', 'dan', 'dari', 'ebanking', 'sudah', 'sukses', 'transfernya', 'perbaiki', 'sistemnya', 'buat', 'orang', 'rugi', 'saja']                                                                                                                                                                                                                                                                                                                                                                                                           |
| ...  | ...                                                                                                                                                                                                                                                                                                                                                                                                                                        | ...                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     |
| 1000 | update aplikasi ke access by kai malah jadi gak bisa lihat boarding pass di hp. gak bisa juga menampilkan keseluruhan stasiun di satu kota misal milih seluruh stasiun di jakarta tidak kayak dulu. masih bagus kai access yg lama lebih simpel enak dilihat lengkap mudah. tolong bisa ditambahkan metode pembayaran. kartu kredit semua bank saat ini hanya debit mandiri. gopay shopee pay saat ini hanya ovo. itupun juga sering error | ['update', 'aplikasi', 'ke', 'access', 'by', 'kai', 'malah', 'jadi', 'gak', 'bisa', 'lihat', 'boarding', 'pass', 'di', 'hp', 'gak', 'bisa', 'juga', 'menampilkan', 'keseluruhan', 'stasiun', 'di', 'satu', 'kota', 'misal', 'milih', 'seluruh', 'stasiun', 'di', 'jakarta', 'tidak', 'kayak', 'dulu', 'masih', 'bagus', 'kai', 'access', 'yg', 'lama', 'lebih', 'simpel', 'enak', 'dilihat', 'lengkap', 'mudah', 'tolong', 'bisa', 'ditambahkan', 'metode', 'pembayaran', 'kartu', 'kredit', 'semua', 'bank', 'saat', 'ini', 'hanya', 'debit', 'mandiri', 'gopay', 'shopee', 'pay', 'saat', 'ini', 'hanya', 'ovo', 'itupun', 'juga', 'sering', 'error'] |



#### d. Normalisasi Teks

Teks dapat dinormalisasi ke dalam Bahasa Indonesia yang baku sesuai KBBI, dengan mengganti kata-kata seperti "yg" menjadi "yang," "ga" menjadi "tidak," dan sebagainya. Dalam penelitian ini, proses normalisasi dilakukan menggunakan file \*kamusGaul.txt\* berbahasa Indonesia yang diperoleh dari GitHub Indonesian sebanyak  $\pm 10000$  kata seperti yang tercantum pada Tabel 6 dibawah.

**Tabel 6.** File Normalisasi

| No   | Sebelum    | Sesudah      |
|------|------------|--------------|
| 1    | woww       | wow          |
| 2    | hallo      | halo         |
| 3    | eeehhhh    | eh           |
| 4    | kata2nyaaa | kata-katanya |
| 5    | smga       | semoga       |
| ...  | ...        | ...          |
| 1000 | kau        | kamu         |

Setelah file normalisasi di uraikan, langkah selanjutnya sistem akan secara otomatis menyesuaikan penggunaan bahasa yang dimana sesuai dengan kamus KBBI. Hasilnya pada Tabel 7 berikut.

**Tabel 7.** Proses Normalisasi Data

| No.  | Case Folding                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               | Normalisasi                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         |
|------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 1    | ['jelek', 'banget', 'aplikasinya', 'isi', 'saldo', 'tapi', 'gak', 'masuk', 'saldo', 'ke', 'aplikasi', 'kai', 'access', 'dan', 'dari', 'ebanking', 'sudah', 'sukses', 'transferynya', 'perbaiki', 'sistemnya', 'buat', 'orang', 'rugi', 'saja']                                                                                                                                                                                                                                                                                                                                                                                                             | ['jelek', 'banget', 'aplikasinya', 'isi', 'saldo', 'tapi', 'enggak', 'masuk', 'saldo', 'ke', 'aplikasi', 'kai', 'access', 'dan', 'dari', 'ebanking', 'sudah', 'sukses', 'transferynya', 'perbaiki', 'sistemnya', 'buat', 'orang', 'rugi', 'saja']                                                                                                                                                                                                                                                                                                                                                                                                                   |
| ...  | ...                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        | ...                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 |
| 1000 | ['update', 'aplikasi', 'ke', 'access', 'by', 'kai', 'malah', 'jadi', 'gak', 'bisa', 'lihat', 'boarding', 'pass', 'di', 'hp', 'gak', 'bisa', 'juga', 'menampilkan', 'keseluruhan', 'stasiun', 'di', 'satu', 'kota', 'misal', 'milih', 'seluruh', 'stasiun', 'di', 'jakarta', 'tidak', 'kayak', 'dulu', 'masih', 'bagusan', 'kai', 'access', 'yg', 'lama', 'lebih', 'simpler', 'enak', 'dilihat', 'lengkap', 'mudah', 'tolong', 'bisa', 'ditambahkan', 'metode', 'pembayaran', 'kartu', 'kredit', 'semua', 'bank', 'saat', 'ini', 'hanya', 'debit', 'mandiri', 'gopay', 'shopee', 'pay', 'saat', 'ini', 'hanya', 'ovo', 'itupun', 'juga', 'sering', 'error'] | ['update', 'aplikasi', 'ke', 'access', 'by', 'kai', 'malah', 'jadi', 'enggak', 'bisa', 'lihat', 'boarding', 'pas', 'di', 'hp', 'enggak', 'bisa', 'juga', 'menampilkan', 'keseluruhan', 'stasiun', 'di', 'satu', 'kota', 'misal', 'memilih', 'seluruh', 'stasiun', 'di', 'jakarta', 'tidak', 'kayak', 'dulu', 'masih', 'bagusan', 'kai', 'access', 'yang', 'lama', 'lebih', 'simpler', 'enak', 'dilihat', 'lengkap', 'mudah', 'tolong', 'bisa', 'ditambahkan', 'metode', 'pembayaran', 'kartu', 'kredit', 'semua', 'bank', 'saat', 'ini', 'hanya', 'debit', 'mandiri', 'gopay', 'shopee', 'pay', 'saat', 'ini', 'hanya', 'ovo', 'itupun', 'juga', 'sering', 'error'] |

#### e. Filtering/Stopword Removal

Dalam teks, stopwords adalah kata-kata umum yang sering muncul, tetapi biasanya tidak memberikan makna yang signifikan untuk analisis, seperti "dan", "atau", "di", dan "yang". Menghilangkan stopwords membantu menyederhanakan data dan meningkatkan efisiensi serta akurasi dalam proses analisis teks. Berikut pada Tabel 8 merupakan implementasinya.

**Tabel 8.** Teks Stopwords

| No.  | Tokenize                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            | Stopwords                                                                                                                                                                                                                                           |
|------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 1    | ['jelek', 'banget', 'aplikasinya', 'isi', 'saldo', 'tapi', 'enggak', 'masuk', 'saldo', 'ke', 'aplikasi', 'kai', 'access', 'dan', 'dari', 'ebanking', 'sudah', 'sukses', 'transferynya', 'perbaiki', 'sistemnya', 'buat', 'orang', 'rugi', 'saja']                                                                                                                                                                                                                                                                                                                                                                                                                   | jelek banget aplikasinya isi saldo masuk saldo aplikasi kai access ebanking sukses transferynya perbaiki sistemnya orang rugi                                                                                                                       |
| ...  | ...                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 | ...                                                                                                                                                                                                                                                 |
| 1000 | ['update', 'aplikasi', 'ke', 'access', 'by', 'kai', 'malah', 'jadi', 'enggak', 'bisa', 'lihat', 'boarding', 'pas', 'di', 'hp', 'enggak', 'bisa', 'juga', 'menampilkan', 'keseluruhan', 'stasiun', 'di', 'satu', 'kota', 'misal', 'memilih', 'seluruh', 'stasiun', 'di', 'jakarta', 'tidak', 'kayak', 'dulu', 'masih', 'bagusan', 'kai', 'access', 'yang', 'lama', 'lebih', 'simpler', 'enak', 'dilihat', 'lengkap', 'mudah', 'tolong', 'bisa', 'ditambahkan', 'metode', 'pembayaran', 'kartu', 'kredit', 'semua', 'bank', 'saat', 'ini', 'hanya', 'debit', 'mandiri', 'gopay', 'shopee', 'pay', 'saat', 'ini', 'hanya', 'ovo', 'itupun', 'juga', 'sering', 'error'] | update aplikasi access by kai lihat boarding pas hp menampilkan stasiun kota memilih stasiun jakarta kayak bagusai kai access simpler enak lengkap mudah tolong metode pembayaran kartu kredit bank debit mandiri gopay shopee pay ovo itupun error |



## f. Stemming

Dalam NLTK (Natural Language Toolkit), stemming adalah salah satu langkah pre-processing yang digunakan untuk mengurangi kata-kata ke bentuk dasarnya, juga dikenal sebagai akar kata. Ini adalah bagian penting dari pipeline preprocessing teks dalam pengolahan bahasa alami (NLP). Proses ini menurunkan variasi kata dengan makna yang sama dan mempermudah analisis teks lainnya, seperti klasifikasi teks, analisis sentimen, dan pengelompokan dokumen seperti yang terlihat pada Tabel 9.

Tabel 9. Stemming Teks

| No.  | Stopwords                                                                                                                                                                                                                                        | Stemming                                                                                                                                                                                                                          |
|------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 1    | jelek banget aplikasinya isi saldo masuk saldo aplikasi kai access ebanking sukses transfernya perbaiki sistemnya orang rugi                                                                                                                     | jelek banget aplikasi isi saldo masuk saldo aplikasi kai access ebanking sukses transfer baik sistem orang rugi                                                                                                                   |
| 2    | nya aplikasi sjamohon bantuanyastlah aplikasi update akun emailno tlf masuk kbuka aplikasinya                                                                                                                                                    | nya aplikasi sjamohon bantuanyastlah aplikasi update akun emailno tlf masuk kbuka aplikasi                                                                                                                                        |
| ...  | ...                                                                                                                                                                                                                                              | ...                                                                                                                                                                                                                               |
| 1000 | update aplikasi access by kai lihat boarding pas hp menampilkan stasiun kota memilih stasiun jakarta kayak bagus kai access simpel enak lengkap mudah tolong metode pembayaran kartu kredit bank debit mandiri gopay shopee pay ovo itupun error | update aplikasi access by kai lihat boarding pas hp tampil stasiun kota pilih stasiun jakarta kayak bagus kai access simpel enak lengkap mudah tolong metode bayar kartu kredit bank debit mandiri gopay shopee pay ovo itu error |

## 3.1.1 Pemberian Label Sentimen Menggunakan InsetLexicon

Langkah selanjutnya adalah proses melabelkan data dimana dilakukan secara otomatis dengan mengimplementasikan kamus InSetLexicon. Nilai polaritas ditentukan dalam kamus InSetLexicon dengan kata-kata dan bobotnya. Untuk mengetahui nilai polaritas, Anda harus menghitung semua bobot kata dalam teks ulasan. Hasil perhitungannya adalah ulasan untuk setiap elemen dikategorikan sebagai sentimen positif atau negatif. Ulasan dianggap positif jika nilai polaritasnya lebih besar dari nol, dan negatif jika nilai polaritasnya lebih rendah dari nol.

Tabel 10. Positive Lexicon

| No.  | Word     | Weight |
|------|----------|--------|
| 1    | Hai      | 3      |
| 2    | Tetap    | 3      |
| 3    | Detail   | 2      |
| 4    | Bagus    | 2      |
| ...  | ...      | ...    |
| 3610 | Orisinal | 3      |

Tabel 11. Negative Lexicon

| No.  | Word      | Weight |
|------|-----------|--------|
| 1    | Isak      | -5     |
| 2    | Sakit     | -5     |
| 3    | Mulu      | -1     |
| 4    | Gamau     | -4     |
| ...  | ...       | ...    |
| 6610 | Mencoreng | -4     |

Di dalam Tabel 10 didapat beberapa komentar bernilai positif dan memiliki value atau bobot di tiap kata, sedangkan pada Tabel 11 menunjukkan tiap kata negatif juga memiliki bobot masing-masing bernilai negatif. Berikut merupakan contoh perhitungan untuk melabelkan sentiment positif dan negatif menggunakan InsetLexicon.

Tabel 12. Ilustrasi Perhitungan Dengan Kamus InSetLexicon

| Ulasan 1: suka versi lot ganggu seringkali error      |        |          |       |        |            |       |           |
|-------------------------------------------------------|--------|----------|-------|--------|------------|-------|-----------|
|                                                       | Suka   | Versi    | Lot   | Ganggu | Seringkali | error | Polaritas |
| Bobot Kata                                            | 3      | -        | -     | -4     | -          | -5    | -6        |
|                                                       | Label  |          |       |        |            |       | Negatif   |
| Ulasan 2: bangga aplikasi mudah cepat mengakses tiket |        |          |       |        |            |       |           |
|                                                       | Bangga | Aplikasi | Mudah | Cepat  | Mengakses  | tiket | Polaritas |
| Bobot Kata                                            | 3      | -4       | 4     | 3      | -          | -     | 6         |
|                                                       | Label  |          |       |        |            |       | Positif   |

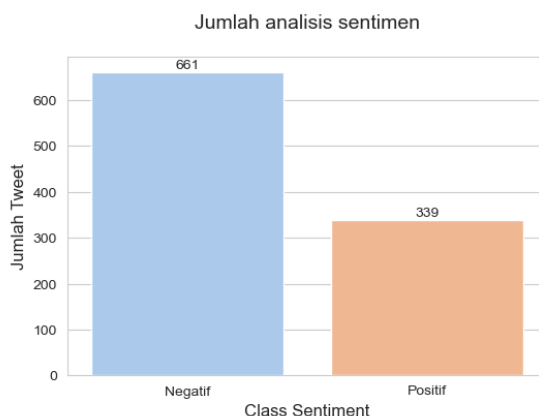


Di dalam Tabel 12 didapat dua komentar bernilai positif dan negatif. Proses ini dilakukan sampai semua dataset. Tabel 12 menunjukkan bahwa semua dataset memiliki labelnya masing-masing yaitu positif dan negatif. Berikut hasil pelabelannya pada Tabel 13 dibawah.

**Tabel 13.** Pelabelan Dataset

| No.  | steaming                                                                                                                                                                                                                          | Label   |
|------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------|
| 1    | jelek banget aplikasi isi saldo masuk saldo aplikasi kai access ebanking sukses transfer baik sistem orang rugi                                                                                                                   | Negatif |
| 2    | nya aplikasi sjamohon bantuanyastlah aplikasi update akun emailno tlf masuk kbuka aplikasi                                                                                                                                        | Negatif |
| 3    | booking trouble sesi bayar salah tulis email kai masuk setting hp aplikasi access by kai hapus cache voila back normal lancar bayar tiket moga bantu                                                                              | Negatif |
| 4    | saudara laki upgrade member basic enggak bisa mulu ya tulis hasil cek profil tampil daftar kali daftar ulang paksa beli tiket website kai ya sehatsehat bikin apk moga baik                                                       | Positif |
| 5    | saran kembali versi bug aplikasi sulit bayar guna qris bingung akses kode qris nya close                                                                                                                                          | Negatif |
| ...  | ...                                                                                                                                                                                                                               | ...     |
| 1000 | update aplikasi access by kai lihat boarding pas hp tampil stasiun kota pilih stasiun jakarta kayak bagus kai access simpel enak lengkap mudah tolong metode bayar kartu kredit bank debit mandiri gopay shopee pay ovo itu error | Negatif |

Jumlah komentar positif dan negatif terdapat pada Gambar 3 berikut, dimana proses dilakukan menggunakan pemrograman python di jupyter notebook.



**Gambar 3.** Tampilan Visualisasi

Dari hasil visualisasi tersebut terlihat bahwa mayoritas komentar memiliki sentimen Negatif sebanyak 661, sementara sentimen Positif hanya sebanyak 339. Hal ini menunjukkan bahwa lebih banyak pengguna memberikan komentar dengan nada tidak puas atau kritis dibandingkan dengan komentar positif. Analisis ini memberikan gambaran tentang persepsi atau tanggapan pengguna terhadap topik tertentu yang diulas dalam komentar aplikasi access by KAI.

### 3.1.2 Pembobotan Kata Menggunakan TF-IDF

Tahap awal yang di lakukan adalah menghitung nilai TF dari semua kata yang terdapat dalam teks (menghitung seberapa sering kata muncul). Seperti tabel 14 berikut yang menggunakan 3 kalimat yaitu :

- mantap aplikasi nya naik kereta online (Positif)
- ribet belit belit baik kasih metode bayar mbangking kak (Negatif)

**Tabel 14.** Term Frequency

| Kata     | TF        |           | DF |
|----------|-----------|-----------|----|
|          | Dokumen 1 | Dokumen 2 |    |
| mantap   | 1         | 0         | 1  |
| aplikasi | 1         | 0         | 1  |
| nya      | 1         | 0         | 1  |
| naik     | 1         | 0         | 1  |
| kereta   | 1         | 0         | 1  |
| online   | 1         | 0         | 1  |
| ribet    | 0         | 0         | 1  |
| belit    | 0         | 2         | 2  |
| baik     | 0         | 1         | 1  |
| kasih    | 0         | 1         | 1  |



| Kata     | TF        |           | DF |
|----------|-----------|-----------|----|
|          | Dokumen 1 | Dokumen 2 |    |
| metode   | 0         | 1         | 1  |
| bayar    | 0         | 1         | 1  |
| mbanking | 0         | 1         | 1  |
| kak      | 0         | 1         | 1  |

Tahap selanjutnya menghitung nilai IDF menggunakan persamaan:

$$idf_{(t,D)} = \log \frac{\text{Total Jumlah Dokumen}}{\text{Jumlah Dokumen dari TF}}$$

- $idf_{mantap} = \log \frac{2}{1} = 0.301$
- $idf_{aplikasi} = \log \frac{2}{1} = 0.301$
- $idf_{nya} = \log \frac{2}{1} = 0.301$

Proses perhitungan nilai IDF tidak dibahas secara keseluruhan Keseluruhan kata untuk hasil nilai IDF disajikan dalam Tabel 15 di bawah ini.

**Tabel 15.** Perhitungan IDF

| Kata     | DF | IDF   |
|----------|----|-------|
| mantap   | 1  | 0.301 |
| aplikasi | 1  | 0.301 |
| nya      | 1  | 0.301 |
| naik     | 1  | 0.301 |
| kereta   | 1  | 0.301 |
| online   | 1  | 0.301 |
| ribet    | 1  | 0.301 |
| belit    | 2  | 0     |
| baik     | 1  | 0.301 |
| kasih    | 1  | 0.301 |
| metode   | 1  | 0.301 |
| bayar    | 1  | 0.301 |
| mbanking | 1  | 0.301 |
| kak      | 1  | 0.301 |

Nilai TF dan IDF dihitung untuk setiap kata dalam komentar pengguna di aplikasi Google Play Store. Setelah memperoleh nilai TF untuk masing-masing kata, langkah selanjutnya adalah menghitung nilai TF-IDF dengan mengalikan nilai TF dan IDF dari setiap kata tersebut. Metode perhitungan TF dan IDF untuk setiap kata disajikan dalam bagian ini.

$$tf - idf_{t,d} = tf_{td} * idf_t$$

- $tf - idf_{(mantap,D1)} = tf_{mantap,D1} * idf_{mantap} = \frac{1}{6} \times 0.301 = 0.050$
- $tf - idf_{(aplikasi,D1)} = tf_{aplikasi,D1} * idf_{aplikasi} = \frac{1}{6} \times 0.301 = 0.050$

Proses perhitungan nilai TF-IDF menunjukkan hasil akhir perhitungan untuk seluruh kata yang telah dirangkul dan disajikan dalam bentuk tabel 16 di bawah ini, guna memberikan gambaran menyeluruh mengenai bobot masing-masing kata dalam analisis.

**Tabel 16.** TF-IDF

| Kata     | Dokumen | TF-IDF |
|----------|---------|--------|
| mantap   | 1       | 0.050  |
| aplikasi | 1       | 0.050  |
| nya      | 1       | 0.050  |
| naik     | 1       | 0.050  |
| kereta   | 1       | 0.050  |
| online   | 1       | 0.050  |
| ribet    | 2       | 0.0334 |
| belit    | 2       | 0      |
| baik     | 2       | 0.0334 |
| kasih    | 2       | 0.0334 |
| metode   | 2       | 0.0334 |

| Kata     | Dokumen | TF-IDF |
|----------|---------|--------|
| bayar    | 2       | 0.0334 |
| mbanking | 2       | 0.0334 |
| kak      | 2       | 0.0334 |

### 3.1.3 Klasifikasi Teks Menggunakan Naïve Bayes dan Menghitung Akurasi

#### a. Clustering K-Means

Proses clustering menggunakan algoritma K-Means dilakukan setelah data ulasan diubah menjadi vektor numerik dengan teknik TF-IDF Vectorizer. Algoritma ini bertujuan untuk mengelompokkan data berdasarkan kemiripan pola kata pada komentar pengguna. Berikut adalah langkah-langkah detail proses clustering:

Langkah pertama adalah dengan melakukan import library untuk model k-means dan memanggil data yang sudah di beri label positif dan negatif sebelumnya.

Kemudian melakukan transformasi atau vektorisasi data TF-IDF dimana Setiap ulasan dikonversi ke dalam representasi vektor dengan menghitung bobot TF-IDF untuk setiap kata sebanyak 1000 data. Hal ini dilakukan untuk menilai pentingnya suatu kata dalam ulasan tertentu dibandingkan dengan keseluruhan data seperti tabel 17 berikut.

**Tabel 17.** Hasil Vektorisasi

| abal | abis | acces | access   | account | aco | ada | adakan | adik | adil | aktif | ... | youtube |
|------|------|-------|----------|---------|-----|-----|--------|------|------|-------|-----|---------|
| 0    | 0    | 0     | 0.252927 | 0       | 0   | 0   | 0      | 0    | 0    | 0     | ... | 0       |
| 0    | 0    | 0     | 0        | 0       | 0   | 0   | 0      | 0    | 0    | 0     | ... | 0       |
| 0    | 0    | 0     | 0.187519 | 0       | 0   | 0   | 0      | 0    | 0    | 0     | ... | 0       |
| ...  | ...  | ...   | ...      | ...     | ... | ... | ...    | ...  | ...  | ...   | ... | ...     |
| 0    | 0    | 0     | 0.270    | 0       | 0   | 0   | 0      | 0    | 0    | 0     | ... | 0       |

Kemudian langkah selanjutnya adalah penerapan algoritma k-means. Dimana Model K-Means diterapkan dengan jumlah kluster yang telah ditentukan sebanyak tiga ( $k = 3$ ). Setiap kluster mewakili kategori berikut:

1. Klaster 1 (Fitur): Mengelompokkan ulasan terkait fitur aplikasi.
2. Klaster 2 (Layanan): Mengelompokkan ulasan terkait layanan pelanggan.
3. Klaster 3 (Sistem): Mengelompokkan ulasan terkait stabilitas sistem aplikasi.

Model ini bekerja dengan menginisialisasi *centroid* secara acak, lalu iteratif memindahkan titik data ke kluster dengan jarak terdekat hingga proses konvergensi tercapai. Berikut contoh perhitungan manual menggunakan K-Means *clustering* terhadap data yang digunakan pada vektorisasi data. Berikut centroid awal yang dipilih  $k = 3$ . Berikut pada tabel 18 merupakan centroid awal yang diperoleh.

**Tabel 18.** Centroid Awal

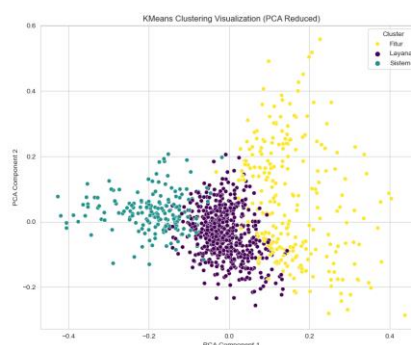
| abal | abis | acces | access | account | aco | ada | adakan | adik | adil | aktif |
|------|------|-------|--------|---------|-----|-----|--------|------|------|-------|
| 0.2  | 0    | 0.2   | 0      | 0       | 0   | 0   | 0      | 0    | 0    | 0     |
| 0    | 0    | 0     | 0      | 0.2     | 0   | 0   | 0      | 0    | 0    | 0     |
| 0    | 0    | 0     | 0      | 0       | 0   | 0.6 | 0      | 0    | 0    | 0     |

Menghitung jarak setiap data menggunakan *Euclidean distance*:

$$Data_{(1,1)} = \sqrt{(0 - 0.2)^2 + (0 - 0)^2 + (0.252927 - 0.2)^2 + \dots + (0 - 0)^2} = 0.2069$$

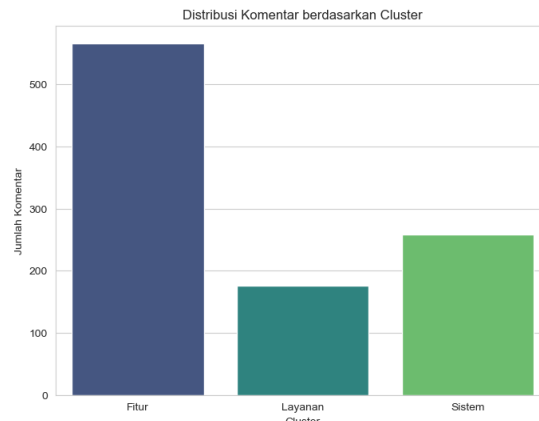
$$Data_{(1,2)} = \sqrt{(0 - 0)^2 + (0 - 0.1)^2 + (0.252927 - 0)^2 + \dots + (0 - 0)^2} = 0.2719$$

Kemudian menentukan *centroid* baru untuk iterasi selanjutnya dengan menghitung rata-rata (*mean*) pada masing-masing *cluster*. Dan terakhir adalah mengulagi langkah yang sama hingga iterasi tidak ada perpindahan. Berikut hasil klasterisasi pada Gambar 5 dengan K-Means berdasarkan fitur (C0), Layanan (C1) dan Sistem (C2).



**Gambar 4.** Visualisasi K-Means

Gambar 4 merupakan hasil visualisasi clustering menggunakan algoritma K-Means yang direduksi ke dalam dua dimensi dengan PCA (Principal Component Analysis). Data dikelompokkan menjadi tiga cluster yang dilabeli sebagai "Fitur" (ungu), "Layanan" (kuning), dan "Sistem" (hijau). Setiap titik merepresentasikan data yang diproyeksikan ke dalam komponen utama, dengan posisi menentukan hubungan antara fitur atau karakteristik data. Cluster terlihat cukup terpisah, terutama antara cluster kuning dan ungu, meskipun ada sedikit tumpang tindih antara cluster hijau dan ungu, menunjukkan beberapa data memiliki karakteristik yang mirip di antara kelompok tersebut.



**Gambar 5.** Distribusi Komentar

Gambar 6 ini adalah visualisasi distribusi komentar berdasarkan cluster yang telah dikelompokkan sebelumnya. Grafik batang menunjukkan bahwa cluster "Fitur" memiliki jumlah komentar terbanyak, yaitu lebih dari 566 komentar, diikuti oleh cluster "Sistem" dengan jumlah sekitar 258 komentar, dan cluster "Layanan" dengan jumlah yang sedikit lebih rendah, yaitu sekitar 176 komentar. Berikut pada Tabel 19 merupakan hasil klasterisasi yang diperoleh.

**Tabel 19.** Hasil Klasterisasi

| Steaming                                                                                                                                                                                      | Label   | cluster | pca x    | pca y    |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------|---------|----------|----------|
| jelek banget aplikasi isi saldo masuk saldo aplikasi kai access ebanking sukses transfer baik sistem orang rugi                                                                               | Negatif | 2       | 0.116913 | 0.303002 |
| nya aplikasi sjamohon bantuanyastlah aplikasi update akun emailno tlf masuk kbuka aplikasi                                                                                                    | Negatif | 0       | -0.14156 | 0.048958 |
| ...                                                                                                                                                                                           | ...     | ...     | ...      | ...      |
| jelass error sinyalpesen tiket error melulu jam ngeklik pesan tinggal bayar error mulu ayo kai nyaman kai enak mesan tiket online sering error mulu orang pilih transportasi simple mesan nya | Negatif | 0       | 0.091644 | -0.17506 |

b. Naive Bayes

Jumlah data yang digunakan dalam penelitian klasifikasi teks komentar pengguna aplikasi access by KAI di google play store dengan Algoritma Naïve Bayes adalah 1000 data. Data ini dibagi menjadi data latih dan data uji dengan perbandingan 80% untuk data latih dan 20% untuk data uji. Data latih berjumlah 800 data, sedangkan data uji terdiri dari 200 data. Dengan menggunakan 3 data sebelumnya sebagai data latih, 1 data uji dipilih sebagai berikut pada Tabel 20.

**Tabel 20.** Data Latih

| Steaming                                                                                                        | Label   | Cluster | Pca_x       | Pca_y       |
|-----------------------------------------------------------------------------------------------------------------|---------|---------|-------------|-------------|
| jelek banget aplikasi isi saldo masuk saldo aplikasi kai access ebanking sukses transfer baik sistem orang rugi | Negatif | 2       | 0.11691292  | 0.303002252 |
| aplikasi parah topup aktivasi kaipay aktivasi mana saldo topup kecewa duit mana                                 | Negatif | 2       | 0.038216422 | 0.186366933 |
| seluru aplikasi pakai ya kalo coba diskusi maksimalin luncur setengah setengah                                  | Positif | 0       | -0.03592816 | -0.00520466 |

Klasifikasi dilakukan secara otomatis menggunakan algoritma Naive Bayes, yang menggunakan fungsi MultinomialNB, yang membandingkan bobot setiap kata dalam dokumen. Ini dilakukan dengan menghitung kemungkinan dokumen data uji berdasarkan kemungkinan kata-kata yang ada dalam data latih. Akibatnya, setiap dokumen pelatihan akan memiliki jumlah probabilitas kata-kata positif dan negatif yang seimbang.

1. Perhitungan nilai probabilitas

$$P(\text{Kelas} | \text{Komentar}) = \frac{\text{Jumlah Kelas } X}{\text{Jumlah Komentar}}$$

Dengan menggunakan persamaan di atas, kita akan memperoleh probabilitas setiap kelas dalam sentimen.

a)  $P(\text{Positif} | \text{Komentar}) = \frac{1}{3} = 0.333$

$$\text{b) } P(\text{Negatif} | \text{Komentar}) = \frac{2}{3} = 0.667$$

## 2. Perhitungan Nilai Probabilitas Kondisional (Likelihood) :

Dari data latih :

- Label negatif (Data 1 dan Data 2), Dimana total kata pada data 1 sebanyak 17 kata dan Data 2 sebanyak 12 kata. Total = 29 kata.
- Label Positif (Data 3) sebanyak 11 kata.
- Fitur unik = 40 kata

$$P(F_i | \text{Kelas}) = \frac{\text{Frekuensi fitur pada label} + 1}{\text{Total Kata pada label} + \text{Jumlah fitur unik}}$$

Dengan menggunakan persamaan di atas, kita akan memperoleh probabilitas suku-suku di setiap kelas sentimen.

1. Probabilitas kata “saldo”

$$P(\text{Saldo} \mid \text{Positif}) = \frac{1+1}{11+40} = \frac{2}{51} = 0.392$$

$$P(\text{Saldo} \mid \text{Negatif}) = \frac{11+40}{29+40} = \frac{51}{69} = 0.043$$

## 2. Probabilitas kata “luar”

$$P(\text{Luar} \mid \text{Positif}) = \frac{0+1}{11+40} = \frac{1}{51} = 0.019$$

$$P(\text{Luar} \mid \text{Negatif}) = \frac{0+1}{29+40} = \frac{1}{69} = 0.019$$

3. Probabilitas kata “transportasi”

$$P(\text{Transportasi} \mid \text{Positif}) = \frac{0+1}{11+40} = \frac{1}{51} = 0.019$$

$$P(\text{Transportasi} \mid \text{Negatif}) = \frac{0+1}{29+40} = \frac{1}{69} = 0.019$$

Selanjutnya adalah mengambil data uji yaitu dengan klasifikasi data uji dengan mengalikan semua peluang. Nilai yang lebih tinggi merupakan kelas baru dari data tersebut. Data uji yang digunakan ada pada tabel 21 berikut,

**Tabel 21.** Data Uji Yang Digunakan

| Test Sentiment                                                                                                | Label | Cluster | Pca_x    | Pca_y    |
|---------------------------------------------------------------------------------------------------------------|-------|---------|----------|----------|
| login akun posisi luar negeri pengaruh loginn akun aplikasi transportasi indonesia kayak malu negara tetangga | ?     | 1       | -0.19019 | 0.024612 |

### 3. Perhitungan nilai probabilitas posterior

$$P(Komentar | Kelas) = P_{Term_1} x ... x P_{Term_n} x P(Kelas | Komentar)$$

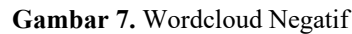
$$\begin{aligned} P(U_{ji} | \text{Positif}) &= P(\text{positif}) \times P(\text{saldo} | \text{positif}) \times P(\text{luar} | \text{positif}) \times \\ &\quad P(\text{transportasi} | \text{positif}) \\ &= 0.333 \times 0.392 \times 0.019 \times 0.019 \\ &= 0.000047123496 \end{aligned}$$

$$\begin{aligned} P(\text{Uji} | \text{Negatif}) &= P(\text{negatif}) \times P(\text{saldo} | \text{negatif}) \times P(\text{luar} | \text{negatif}) \\ &\quad P(\text{transportasi} | \text{negatif}) \\ &= 0.667 \times 0.043 \times 0.019 \times 0.019 \\ &= 0.000010353841 \end{aligned}$$

Kesimpulannya yaitu karena  $P(\text{Uji} \mid \text{Negatif})$  lebih kecil daripada  $P(\text{Uji} \mid \text{Positif})$ , maka komentar tersebut lebih mungkin diklasifikasikan sebagai “Positif”. Berikut pada gambar 7 dan 8 merupakan visualisasi menggunakan wordcloud positif dan negatif.



**Gambar 6.** Wordcloud Positif



Confusion Matrix

|                      | Actual<br>Futur | Actual<br>Layanan | Actual<br>Sistem |
|----------------------|-----------------|-------------------|------------------|
| Predicted<br>Futur   | 114             | 17                | 14               |
| Predicted<br>Layanan | 0               | 23                | 0                |
| Predicted<br>Sistem  | 1               | 1                 | 30               |

This Journal is licensed under a Creative Commons Attribution 4.0 International License



Fitur:

$$\text{precision} = \frac{TP}{TP+FP} = \frac{114}{114+31} = \frac{114}{145} = 0.786$$

$$\text{recall} = \frac{TP}{TP+FN} = \frac{114}{114+1} = \frac{114}{115} = 0.991$$

$$F1 - \text{Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} = 2 \times \frac{0.786 \times 0.991}{0.786 + 0.991} = 0.877$$

Layanan:

$$\text{precision} = \frac{TP}{TP+FP} = \frac{23}{23+0} = \frac{23}{23} = 1$$

$$\text{recall} = \frac{TP}{TP+FN} = \frac{23}{23+17} = \frac{23}{40} = 0.575$$

$$F1 - \text{Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} = 2 \times \frac{1 \times 0.575}{1 + 0.575} = 0.730$$

Sistem:

$$\text{precision} = \frac{TP}{TP+FP} = \frac{30}{30+1} = \frac{30}{31} = 0.968$$

$$\text{recall} = \frac{TP}{TP+FN} = \frac{30}{30+14} = \frac{30}{44} = 0.682$$

$$F1 - \text{Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} = 2 \times \frac{0.968 \times 0.682}{0.968 + 0.682} = 0.801$$

**Tabel 22.** Hasil Pengujian

| Aspek          | Recall | Precision | F1-Score | Support |
|----------------|--------|-----------|----------|---------|
| Fitur          | 0.786  | 0.991     | 0.887    | 115     |
| Layanan        | 1.0    | 0.575     | 0.730    | 41      |
| Sistem         | 0.968  | 0.682     | 0.801    | 44      |
| Total Data Uji |        |           |          | 200     |

Hasil confusion matrix yang ditampilkan dalam tabel 22 ini menunjukkan performa model klasifikasi sentimen berdasarkan tiga kategori yaitu Fitur, Layanan, dan Sistem. Pada kategori Fitur, model memiliki precision sebesar 0.79, recall sebesar 0.99, dan F1-score sebesar 0.88, menunjukkan kemampuan yang baik dalam mengenali data positif meskipun ada beberapa false positives. Untuk kategori Layanan, precision sangat tinggi (1.00), tetapi recall hanya 0.56, yang mengindikasikan model sering tidak mengenali data positif di kategori ini, menghasilkan F1-score sebesar 0.72. Sementara pada kategori Sistem, precision mencapai 0.94, recall 0.68, dan F1-score 0.79, yang menunjukkan keseimbangan yang lebih baik dibandingkan kategori Layanan. Secara keseluruhan, akurasi model adalah 83%, dengan rata-rata makro recall sebesar 0.74 dan weighted average recall sebesar 0.83. Dalam analisis ini, saya juga akan menampilkan seluruh hasil klasifikasi sentimen berdasarkan kategori Fitur (total 566 data), Layanan (176 data), dan Sistem (258 data), sebagai bagian dari keseluruhan evaluasi performa model. Berikut penjelasan klasifikasi fitur pada Tabel 23, layanan pada Tabel 24 dan sistem pada Tabel 25.

**Tabel 23.** Klasifikasi "Fitur"

| No  | Steamming                                                                                                                                                                                                                         |
|-----|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 1   | nya aplikasi sjamohon bantuanyastlah aplikasi update akun emailno tlf masuk kbuka aplikasi                                                                                                                                        |
| 2   | saran kembali versi bug aplikasi sulit bayar guna qris bingung akses kode qris nya close                                                                                                                                          |
| 3   | seluru aplikasi pakai ya kalo coba diskusi maksimalin luncur setengahsetengah                                                                                                                                                     |
| 4   | top up kaipay uang pin masuk salah tekan lupa pin email akses niat mudah sulit kereta bayar gratis                                                                                                                                |
| 5   | pesan tiket terang jadwal temu pas pakai apk jadwal hapus aplikasi nambahin beban negara                                                                                                                                          |
| ... | ...                                                                                                                                                                                                                               |
| 566 | update aplikasi access by kai lihat boarding pas hp tampil stasiun kota pilih stasiun jakarta kayak bagus kai access simpel enak lengkap mudah tolong metode bayar kartu kredit bank debit mandiri gopay shopee pay ovo itu error |

**Tabel 24.** Klasifikasi "Layanan"

| No | Steamming                                                                                                                                                                   |
|----|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 1  | saudara laki upgrade member basic enggak bisa mulu ya tulis hasil cek profil tampil daftar kali daftar ulang paksa beli tiket website kai ya sehatsehat bikin apk moga baik |
| 2  | member basic login daftar daftar mau pesan tiket aplikasi stupid                                                                                                            |
| 3  | akun login deh coba deh masuk login akun lupa sandi nya coba kali pas coba suruh tunggu jam kocak enggak jelas aplikasi                                                     |



| No  | Steamming                                                                                                                                                                                                                                                           |
|-----|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 4   | sorry ya bodoamat pokok ratting pantas haruanya minus stay tunggu jam pas jam segitu masuk kai pesan tiket masuk daftar tumpang enggak bisa dinext ku coba eh order kali order palalu yang pesan anjirpdahal wifi lancar ku coba pakai data lancar enggak bisa gila |
| 5   | login akun posisi luar negeri pengaruh loginn akun aplikasi transportasi indonesia kayak malu negara tetangga                                                                                                                                                       |
| ... | ...                                                                                                                                                                                                                                                                 |
| 176 | ayo aplikasi layanan senang beli kalo tuju apk kayak error daftar premium pilih pencet pas daftar basic error centang masuk baik aplikasi pergi                                                                                                                     |

Tabel 25. Klasifikasi "Sistem"

| No  | Steamming                                                                                                                                            |
|-----|------------------------------------------------------------------------------------------------------------------------------------------------------|
| 1   | jelek banget aplikasi isi saldo masuk saldo aplikasi kai access ebanking sukses transfer baik sistem orang rugi                                      |
| 2   | booking trouble sesi bayar salah tulis email kai masuk setting hp aplikasi access by kai hapus cache voila back normal lancar bayar tiket moga bantu |
| 3   | aplikasi parah topup aktivasi kaipay aktivasi mana saldo topup kecewa duit mana                                                                      |
| 4   | pakai qr kai pay cantum bayar hasil tiket bayar                                                                                                      |
| 5   | kai acces turun ganti nomer hangus nomer ambil sisa saldo kai pay hufftttt sulit                                                                     |
| ... | ...                                                                                                                                                  |
| 258 | top kai pay pakai shoepay ovo untung cs nya cepat respon coba pakai qris bayar nya pakai qris saudara laki nama usaha plat merah jelak               |

### 3.3 Pembahasan

Berdasarkan hasil klasifikasi teks komentar pengguna aplikasi Access by KAI menggunakan metode Naïve Bayes, diperoleh akurasi sebesar 83% dengan performa terbaik pada kategori Fitur (precision 79%, recall 99%, F1-score 88%), sedangkan kategori Layanan menunjukkan ketidakseimbangan dengan precision 100% namun recall hanya 56%. Hasil ini menunjukkan bahwa metode Naïve Bayes cukup efektif dalam mengelola data ulasan yang besar dan memberikan hasil yang kompetitif dibandingkan penelitian sebelumnya, seperti penelitian Riyanah & Fatmawati (2021) yang hanya mencapai akurasi 62,86%, dan penelitian Adyatma Subagja dkk. (2021) yang hanya mencapai akurasi 57%. Sementara itu, meskipun penelitian oleh Nurzaman dkk. (2024) mencatat precision tinggi (92,19%), akurasinya masih lebih rendah (63,61%) dibanding penelitian ini. Dengan demikian, pendekatan yang digunakan dalam penelitian ini terbukti lebih unggul, khususnya karena integrasi teknik clustering K-Means sebelum klasifikasi turut membantu meningkatkan akurasi model.

## 4. KESIMPULAN

Berdasarkan hasil penelitian klasifikasi teks komentar pengguna aplikasi Access by KAI di Google Play Store dengan menggunakan metode Naïve Bayes, diperoleh akurasi model yang cukup baik, yaitu sebesar 83%. Penelitian ini menggunakan total 1.000 komentar yang diperoleh melalui proses web scraping dari Google Play Store, kemudian dibagi menjadi 800 data latih dan 200 data uji untuk proses pembelajaran dan evaluasi model. Sebelum dilakukan klasifikasi sentimen, diterapkan algoritma K-Means untuk mengelompokkan komentar berdasarkan pola kemiripan kata ke dalam tiga kategori utama, yaitu Fitur, Layanan, dan Sistem. Pendekatan ini digunakan guna meningkatkan kualitas klasifikasi sentimen yang lebih terarah dan fokus terhadap aspek tertentu. Selanjutnya, pelabelan sentimen dilakukan secara otomatis menggunakan metode InSetLexicon berbahasa Indonesia, yang memberikan nilai polaritas positif atau negatif berdasarkan kamus sentimen yang telah disusun sebelumnya. Hasil evaluasi model menunjukkan bahwa kategori Fitur memiliki performa terbaik dengan nilai precision sebesar 79%, recall 99%, dan F1-score 88%. Untuk kategori Sistem, precision mencapai 94%, recall 68%, dan F1-score 79%. Sedangkan pada kategori Layanan, precision sangat tinggi sebesar 100%, namun recall-nya rendah, yaitu 56%, sehingga menghasilkan F1-score sebesar 72%. Secara keseluruhan, sentimen positif lebih dominan dibandingkan dengan sentimen negatif. Komentar terbagi ke dalam 566 untuk kategori Fitur, 258 untuk Layanan, dan 176 untuk Sistem. Hasil ini menunjukkan bahwa model bekerja paling stabil pada kategori Fitur dan Sistem, serta dapat diandalkan dalam membantu analisis opini pengguna terhadap aplikasi.

## REFERENCES

- [1] V. R. Modompit, "Analisis Permintaan Transportasi Gojek Online Di Kota Manado," *J. Berk. Ilm. Efisiensi*, vol. 20, no. 03, pp. 140–151, 2020.
- [2] Eskiyaturrofikoh and R. R. Suryono, "Analisis Sentimen Aplikasi X Pada Google Play Store Menggunakan Algoritma Naïve Bayes Dan Support Vector Machine (SVM)," *Jurnal Ilmiah Penelitian dan Pembelajaran Informatika (JIPI)*, vol. 8, no. 3, pp. 110–118, 2019, doi: 10.29100/jipi.v9i3.5392
- [3] Anwar Pauji, S. Aisyah, A. Surip, R. Saputra, and I. Ali, "Implementasi Algoritma K-Nearest Neighbor Dalam Menentukan Penerima Bantuan Langsung Tunai," *KOPERTIP J. Ilm. Manaj. Inform. dan Komput.*, vol. 4, no. 1, pp. 21–27, 2022, doi: 10.32485/kopertip.v4i1.114.
- [4] A. Surahman, U. Hayati, B. Sosial, and R. Miner, "Implementasi Algoritma Naïve Bayes," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 7, no. 1, pp. 347–352, 2023, doi: 10.36040/jati.v7i1.6302



- [5] N. Riyanah and F. Fatmawati, "Penerapan Algoritma Naive Bayes Untuk Klasifikasi Penerima Bantuan Surat Keterangan Tidak Mampu," *JTIM J. Teknol. Inf. dan Multimed.*, vol. 2, no. 4, pp. 206–213, 2021, doi: 10.35746/jtim.v2i4.117.
- [6] R. Adyatma Subagja, Y. Widiastiwi, and N. Chamidah, "Klasifikasi Ulasan Aplikasi Jenius pada Google Play Store Menggunakan Algoritma Naive Bayes," *Inform. J. Ilmu Komput.*, vol. 17, no. 3, p. 197, 2021, doi: 10.52958/iftk.v17i3.3652.
- [7] N. Nurzaman, N. Suarna, and W. Prihartono, "Analisis Sentimen Ulasan Aplikasi Threads Di Google Playstore Menggunakan Algoritma Naive Bayes," *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 8, no. 1, pp. 967–974, 2024, doi: 10.36040/jati.v8i1.8708.
- [8] P. Utomo, N. Asvio, and F. Prayogi, "Metode Penelitian Tindakan Kelas (PTK): Panduan Praktis untuk Guru dan Mahasiswa di Institusi Pendidikan," *Pubmedia J. Penelit. Tindakan Kelas Indones.*, vol. 1, no. 4, p. 19, 2024, doi: 10.47134/ptk.v1i4.821.
- [9] M. I. A. Ranius, T. Sutabri, and A. Y. Ranius, "Analisis Manajemen Pelayanan PT. KAI Sebagai Pengguna pada Aplikasi KAI ACCESS Berbasis Teknologi Informasi Menggunakan Framework ITIL Version 3," *Indones. J. Multidiscip. Soc. Technol.*, vol. 1, no. 2, pp. 135–140, 2023, doi: 10.31004/ijmst.v1i2.136.
- [10] S. D. Pratama and E. Syaodih, "Analisis Perilaku Konsumen dalam Memanfaatkan Aplikasi KAI Access," *Serv. Manag. Triangle J. ...*, vol. 3, no. 1, pp. 20–27, 2021.
- [11] M. P. R. Putra and K. R. N. Wardani, "Penerapan Text Mining Dalam Menganalisis Kepribadian Pengguna Media Sosial," *JUTIM (Jurnal Tek. Inform. Musirawas)*, vol. 5, no. 1, pp. 63–71, 2020, doi: 10.32767/jutim.v5i1.791.
- [12] M. Sholih 'afif, M. Muzakir, M. I. Al, and G. Al Awalaen, "Text Mining Untuk Mengklasifikasi Judul Berita Online Studi Kasus Radar Banjarmasin Menggunakan Metode Naïve Bayes," *Kumpul. J. Ilmu Komput.*, vol. 08, no. 2, pp. 199–208, 2021.
- [13] M. Ikhsan and R. R. Kurniawan, "Penerapan Text Mining pada Sistem Rekomendasi Pembimbing Skripsi Mahasiswa Menggunakan Algoritma Naïve Bayes Classifier," *Indones. J. Comput. Sci. Attrib.*, vol. 12, no. 6, pp. 2023–4196, 2023.
- [14] L. Efrizoni, S. Defit, M. Tajuddin, and A. Anggrawan, "Komparasi Ekstraksi Fitur dalam Klasifikasi Teks Multilabel Menggunakan Algoritma Machine Learning," *MATRIK J. Manajemen, Tek. Inform. dan Rekayasa Komput.*, vol. 21, no. 3, pp. 653–666, 2022, doi: 10.30812/matrik.v21i3.1851.
- [15] N. W. Utami and I. G. J. Eka Putra, "Text Minig Clustering Untuk Pengelompokan Topik Dokumen Penelitian Menggunakan Algoritma K-Means Dengan Cosine Similarity," *J. Inform. Teknol. dan Sains*, vol. 4, no. 3, pp. 255–259, 2022, doi: 10.51401/jinteks.v4i3.1907.
- [16] A. Z. Amrullah, A. Sofyan Anas, and M. A. J. Hidayat, "Analisis Sentimen Movie Review Menggunakan Naive Bayes Classifier Dengan Seleksi Fitur Chi Square," *Jurnal Bumigora*, vol. 2, no. 1, pp. 40–44, 2020, doi: 10.30812/bite.v2i1.804.
- [17] M. H. Mahendra, D. T. Murdiansyah, and K. M. Lhaksana, "Analisis Sentimen Tweet COVID-19 menggunakan K-Nearest Neighbors dengan TF-IDF dan Ekstraksi Fitur CountVectorizer," *DIKE J. Ilmu Multidisiplin*, vol. 1, no. 2, pp. 37–43, 2023, doi: 10.69688/dike.v1i2.35.
- [18] D. F. Zhafira, B. Rahayudi, and I. Indriati, "Analisis Sentimen Kebijakan Kampus Merdeka Menggunakan Naive Bayes dan Pembobotan TF-IDF Berdasarkan Komentar pada Youtube," *J. Sist. Informasi, Teknol. Informasi, dan Edukasi Sist. Inf.*, vol. 2, no. 1, pp. 55–63, 2021, doi: 10.25126/justsi.v2i1.24.
- [19] N. M. A. Mahar, Vihi Atina, and Nugroho Arif Sudibyo, "Pemodelan Prediksi Kelulusan Mahasiswa Dengan Metode Naïve Bayes Di Uniba," *J. Manaj. Inform. dan Sist. Inf.*, vol. 6, no. 2, pp. 148–158, 2023, doi: 10.36595/misi.v6i2.875.
- [20] L. Mayasari and D. Indarti, "Klasifikasi Topik Tweet Mengenai Covid Menggunakan Metode Multinomial Naïve Bayes Dengan Pembobotan Tf-Idf," *J. Ilm. Inform. Komput.*, vol. 27, no. 1, pp. 43–53, 2022, doi: 10.35760/ik.2022.v27i1.6184.