

Perbandingan Algoritma Naive Bayes Dan K-Nearest Neighbor Terhadap Analisis Sentimen Aplikasi Rosalia Indah Transport

Herbi Satrio¹, Eka Kusuma Pratama²

^{1,2}Informatika, Fakultas Teknik dan Informatika, Universitas Bina Sarana Informatika, Jl. Kramat Raya No.98, Kwitang, Kec. Senen Kota Jakarta Pusat DKI Jakarta, 10450

herbisatrio@gmail.com¹, eka.eem@bsi.ac.id²

Abstrak

Analisis sentimen yang diterapkan pada aplikasi Rosalia Indah Transport melibatkan pengumpulan ulasan atau umpan balik dari pengguna. Selanjutnya, algoritma analisis sentimen digunakan untuk mengklasifikasikan ulasan tersebut sebagai positif atau negatif. Analisis sentimen ini membantu perusahaan memahami pandangan pengguna tentang aplikasi Rosalia Indah Transport dan sejauh mana aplikasi ini mampu memenuhi kebutuhan pengguna. Proses analisis sentimen pada aplikasi Rosalia Indah Transport mencakup pengumpulan ulasan dari pengguna, yang kemudian diklasifikasikan menggunakan algoritma analisis sentimen.. Hasil penelitian menunjukkan bahwa KNN memiliki kinerja yang lebih unggul dibandingkan dengan Naive Bayes dalam semua metrik yang diuji. Akurasi KNN mencapai 81.54%, sementara Naive Bayes hanya mencapai 64.94%. Precision dan recall untuk KNN juga lebih tinggi, masing-masing sebesar 82.87%, 92.29%, dibandingkan dengan Precision dan recall Naive Bayes yang hanya mencapai 81%, 64.24%. Temuan ini menunjukkan bahwa KNN adalah metode yang lebih efektif untuk analisis sentimen ulasan aplikasi Rosalia Indah Transport, yang dapat membantu pengembang dan bisnis dalam memahami dan meningkatkan kualitas aplikasi mereka.

Kata kunci: Naive Bayes, K-Nearest Neighbor, Rosalia Indah Transport.

A. Pendahuluan

Transportasi adalah sarana yang digunakan seseorang untuk berpindah dari satu tempat ke tempat lain. Saat ini, terdapat berbagai jenis kendaraan yang umum dijumpai di jalan raya (Noor & Dwijayanti, 2021). Perkembangan teknologi di Indonesia mendorong digitalisasi di berbagai sektor, termasuk transportasi.

Aplikasi Rosalia Indah Transport yang dikembangkan oleh PO Rosalia Indah merupakan salah satu inovasi di bidang transportasi. Aplikasi ini mempermudah pengguna dalam memesan tiket bus Rosalia Indah secara daring. Pengguna sering kali memberikan ulasan mengenai pengalaman mereka, yang menjadi sumber data berharga bagi perusahaan dalam meningkatkan layanan. Ulasan tersebut mengandung sentimen positif atau negatif yang penting untuk dianalisis guna memahami kepuasan pelanggan.

Analisis sentimen adalah proses otomatis untuk memahami, mengekstrak, dan memproses data berbasis teks guna mengidentifikasi sentimen tersirat dalam opini pengguna(Fajar Sidik et al., 2022). Teknik ini dapat diterapkan menggunakan algoritma *Machine Learning* atau teknologi *Natural Language Processing (NLP)*, yang membantu memahami konteks teks dengan lebih mendalam (Pradana Rachman et al., 2021).

Dalam pemasaran, analisis sentimen digunakan untuk mengukur respons pelanggan terhadap produk atau merek, serta membantu pengembangan strategi pemasaran yang efektif (Rachman et al., 2021). Pada aplikasi Rosalia Indah Transport, analisis sentimen dapat dilakukan untuk memahami ulasan pengguna terkait aplikasi dan meningkatkan kualitas layanan berdasarkan ulasan tersebut.

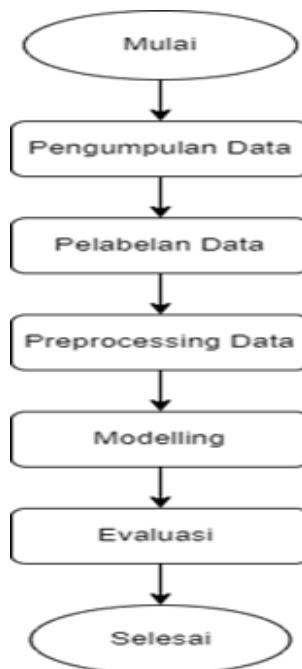
Beberapa studi terkait mendukung pentingnya analisis sentimen dalam sektor transportasi digital. Studi yang dilakukan oleh (Irfan & Erizal, 2024) membandingkan dua algoritma, yaitu Naive Bayes dan *K-Nearest Neighbor (KNN)*, dalam memprediksi sentimen ulasan aplikasi inDrive. Hasil penelitian menunjukkan bahwa algoritma Naive Bayes memberikan akurasi 97.50%, *precision* 92.71%, dan *recall* 100%, sedangkan KNN memberikan akurasi 83.21%, *precision* 85%, dan *recall* 57.30%. Dari hasil ini, dapat disimpulkan bahwa Naive Bayes lebih unggul dalam memprediksi sentimen ulasan dibandingkan KNN.

Penelitian lain yang dilakukan oleh(Wijaya et al., 2024) pada ulasan aplikasi Mitra Darat juga menunjukkan hasil serupa, di mana model Naive Bayes berhasil mencapai akurasi 99.28%, sementara KNN hanya memperoleh akurasi 80%. Evaluasi lainnya mendukung temuan ini, membuktikan bahwa model Naive Bayes efektif dalam analisis sentimen terhadap dataset ulasan aplikasi transportasi.

Penelitian ini membandingkan kinerja algoritma Naive Bayes (NB) dan *K-Nearest Neighbor (KNN)* dalam mengklasifikasikan sentimen ulasan pengguna aplikasi Rosalia Indah Transport. Algoritma NB menggunakan pendekatan statistik berdasarkan Teorema Bayes (Derajad Wijaya & Dwiasnati, 2020), sedangkan KNN bekerja dengan mencari titik data terdekat untuk memprediksi sentimen (Syahril Dwi Prasetyo et al., 2023). Indikator performa yang digunakan adalah akurasi, *precision*, *recall* dan nilai *AUC*. Hasil penelitian ini diharapkan memberikan wawasan mengenai algoritma yang lebih efektif dalam menganalisis sentimen di platform aplikasi transportasi digital.

B. Metode

Penelitian ini melakukan analisis sentimen pada aplikasi Rosalia Indah Transport dengan menggunakan algoritma Naïve Bayes dan *K-Nearest Neighbor (KNN)*. Setiap tahap penelitian disajikan melalui diagram alir untuk memastikan penelitian berjalan sesuai dengan tujuan yang ingin dicapai. Tahapan penelitian ditampilkan pada gambar 1.



Gambar 1. Tahapan Penelitian

1. Pengumpulan Data

Data yang diambil berupa ulasan pengguna berbahasa indonesia dari aplikasi Rosalia Indah Transport di *Google Play Store*, dengan pengumpulan data dilakukan melalui teknik *scraping* menggunakan bahasa pemrograman Python. Rentang waktu pengambilan data diambil secara acak mulai dari September 2018 hingga April 2024. Sebanyak 1392 data berhasil dikumpulkan dari hasil pengambilan data.

2. Pelabelan Data

Sebanyak 1.392 data ulasan yang telah terkumpul akan dibagi menjadi dua kategori, yaitu positif dan negatif. Kategori ini ditentukan berdasarkan bintang yang diberikan oleh pengguna dalam ulasan mereka. Ulasan dengan bintang 4 dan 5 akan dikelompokkan ke dalam kategori positif, sedangkan ulasan dengan bintang 3, 2, dan 1 akan dikategorikan sebagai negatif.

3. *Preprocessing Data*

Preprocessing dapat menjadi proses berulang, karena pilihan langkah pra-pemrosesan dapat bergantung pada karakteristik data dan persyaratan khusus dari algoritme pembelajaran mesin. Penting untuk mempertimbangkan dengan hati-hati langkah-langkah pra-pemrosesan yang sesuai untuk data dan masalah yang dihadapi, karena langkah-langkah ini dapat berdampak signifikan terhadap performa model. Tujuan dari tahap pengolahan data ini adalah untuk memudahkan analisis data pada tahap selanjutnya. Pada fase ini dilakukan pembersihan dan persiapan data agar data yang akan dianalisis lebih terorganisir dan siap dilakukan penelitian(Syahril Dwi Prasetyo et al., 2023).

Preprocessing merupakan teknik awal data mining yang bertujuan mengubah data mentah yang sudah terkumpul menjadi data bersih yang dapat digunakan pada proses berikutnya (Aziz & Fauziah, 2022). Teknik preprocessing yang diterapkan dalam penelitian ini yaitu:

a) *Cleaning*

Pada tahap ini dilakukan pembersihan data teks dari kesalahan, inkonsistensi, dan tidak relevan meliputi menghapus karakter khusus seperti tanda baca, simbol, dan angka (Syafrizal et al., 2023)

b) *Transform Cases*

Transform Cases adalah tahap mengubah semua huruf kapital menjadi huruf kecil (Ahmadi et al., 2021).

c) *Tokenize*

Tokenize adalah proses memisahkan teks atau kalimat menjadi unit-unit yang lebih kecil yang disebut token (Rifai et al., 2021).

d) *Stopword*

Stopword adalah proses penghapusan kata-kata yang dianggap tidak relevan atau terlalu umum dalam teks atau dokumen (Rahutomo et al., 2019).

e) *Stem*

Merupakan proses untuk menghilangkan awalan atau akhiran kata yang terdapat kata sambung, kata depan, kata ganti, menjadi kata dasar (Darwisi et al., 2020).

f) *Filter Token by Length*

Proses ini bertujuan untuk menghapus kata-kata yang memiliki jumlah karakter tertentu.

4. Modelling

Tahapan pemodelan dilakukan dengan menggunakan teknik pemodelan yang digunakan pada dataset yang telah disiapkan (Septi Rizqiah & Rahman Kadafi, 2022). Penelitian ini memanfaatkan algoritma klasifikasi Naïve Bayes dan *K-Nearest Neighbor* untuk teknik pemodelan. Proses pemodelan dilakukan menggunakan *tools RapidMiner*.

5. Evaluasi

Pada tahap evaluasi ini, dilakukan pengujian terhadap model yang telah dibangun sebelumnya untuk mengukur kinerjanya dalam mengklasifikasikan data ulasan. Evaluasi dilakukan dengan menggunakan *confusion matrix*, yang mampu memberikan gambaran mengenai performa model dalam mendeteksi kelas positif dan negatif secara akurat. Dalam hal ini, kedua algoritma yang digunakan, yaitu Naïve Bayes dan *K-Nearest Neighbor (KNN)*, dianalisis dan dibandingkan kinerjanya melalui metrik-metrik evaluasi seperti akurasi, *precision*, *recall*, serta nilai *AUC*. Proses evaluasi ini bertujuan untuk menentukan algoritma yang memberikan hasil klasifikasi terbaik untuk dataset yang digunakan.

C. Hasil dan Pembahasan

Bagian ketiga mencakup dokumentasi setiap langkah penelitian yang dilakukan. Pengambilan data akan dilakukan menggunakan Python, sedangkan analisis data akan dilakukan menggunakan *RapidMiner*.

1. Pengumpulan Data

Pengumpulan data dilakukan dengan memanfaatkan teknik *scraping* menggunakan bahasa pemrograman Python selama periode satu tahun, dimulai dari September 2018 hingga April 2024. Teknik ini memungkinkan penarikan data secara otomatis dari ulasan pengguna di aplikasi yang ditargetkan. Setelah proses scraping selesai, diperoleh total 1.392 baris data.

2. Pelabelan Data

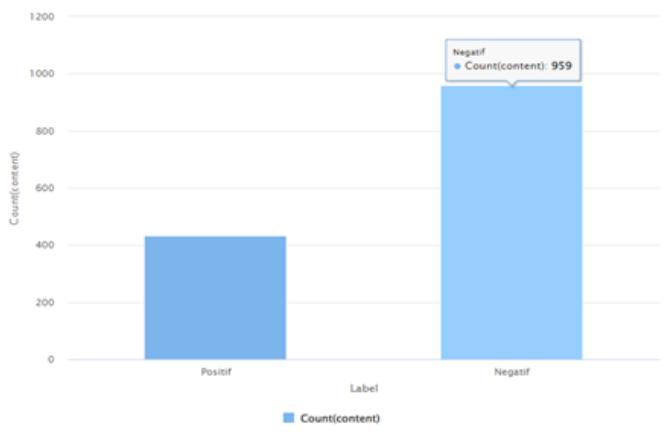
Sebanyak 1.392 data yang telah terkumpul akan dibagi menjadi dua kategori, yaitu kategori positif dan negatif. Kategori ini ditentukan berdasarkan rating atau "skor" pada ulasan. Ulasan dengan bintang 4 dan 5 masuk ke dalam kategori

positif, sedangkan ulasan dengan bintang 3, 2, dan 1 dikategorikan sebagai ulasan negatif. Hasil dari tahapan pelabelan data dapat dilihat pada Tabel 1.

Tabel 1. Hasil Pelabelan Data

| Ulasan | Label |
|--|---------|
| Pelayanan yg cukup baik gak pernah ngecewain, tapi kali ini kecewa tumben banget gak dapet Snack dan menu makanannya berbeda gak seperti biasa Tapi tetap suka naik transport rosin karna senyaman itu | Positif |
| Sering log out sendiri, pas log in ga dapet OTP, selang 2 hari baru dpt OTP. Kalo belum bisa munpuni bikin aplikasi, mending via web saja | Negatif |

Hasil pelabelan data yang telah dilakukan mendapatkan total 433 sentimen positif dan 959 sentimen negatif dari total 1392 baris data. Pembagian ini menunjukkan bahwa mayoritas ulasan dari dataset lebih cenderung memberikan evaluasi negatif terhadap aplikasi Rosalia Indah Transport. Visualisasi hasil dari proses pelabelan data dapat dilihat pada Gambar 2.



Gambar 2. Visualisasi Data Labeling

3. Preprocessing Data

Setelah proses pelabelan data ulasan selesai, langkah berikutnya adalah melakukan *preprocessing* teks pada kumpulan ulasan pengguna. Tahapan preprocessing teks meliputi:

- a) Cleaning
- b) Transform Cases
- c) Tokenize
- d) Stopword
- e) Stem

f) Filter Token by Length

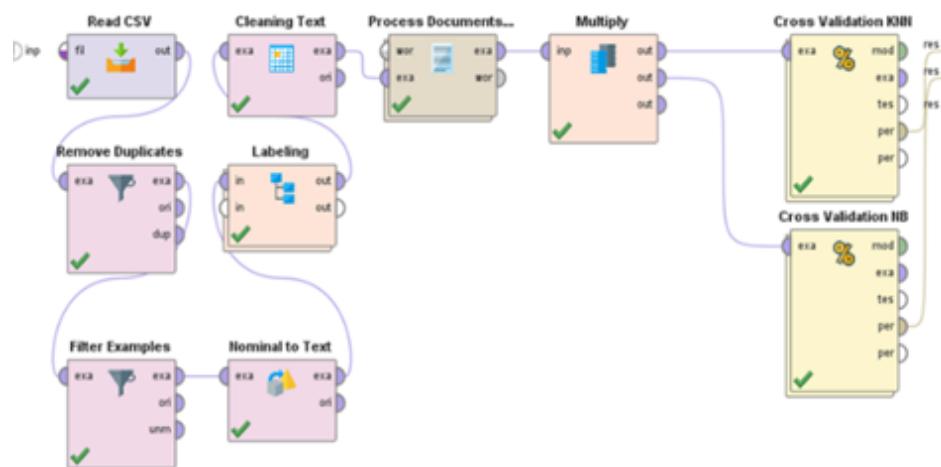
Hasil dari tahapan preprocessing dapat dilihat pada tabel 2.

Tabel 2. Hasil Preprocessing

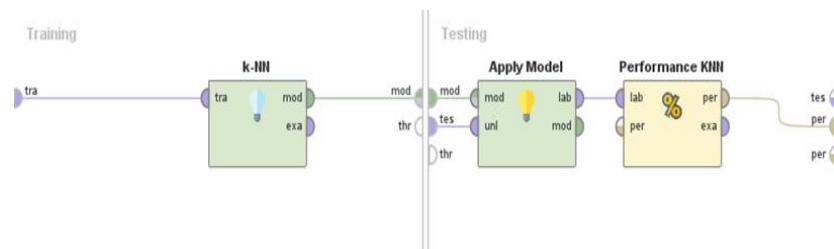
| Cleaning | Transform Cases | Tokenize | Stopword | Stem | Filter Token by Length |
|---|---|--|---|---|---|
| Pelayanan yg cukup baik gak pernah ngecewain Tapi kali ini kecewa tumben banget gak dapet Snack dan menu makanannya berbeda gak seperti biasa tapi tetap suka naik transport rosin karna senyaman itu | pelayanan yg cukup baik gak pernah ngecewain tapi kali ini kecewa tumben banget gak dapet snack dan menu makanannya berbeda gak seperti biasa tapi tetap suka naik transport rosin karna senyaman itu | [pelayanan, yg, cukup, baik, gak, pernah, ngecewain, Tapi, kali, ini, kecewa, tumben, banget, gak, dapet, snack, menu, makanannya, berbeda, gak, senyaman] | [pelayanan, ngecewain, kecewa, tumben, dapet, snack, menu, makanannya, berbeda, suka, transport, rosin, nyaman] | [pelayan, ngecewain, kecewa, tumben, dapat, snack, menu, makan, beda, suka, transport, rosin, nyaman] | [pelayan, ngecewain, kecewa, tumben, dapat, snack, menu, makan, beda, suka, transport, rosin, nyaman] |
| Sering log out sendiri, pas log in ga dapet otp selang hari baru dpt otp kalo belum bisa munpuni bikin aplikasi mending via web saja | sering log out sendiri pas log in ga dapet otp selang hari baru dpt otp kalo belum bisa munpuni bikin aplikasi mending via web saja | ['sering', 'log', 'out', 'sendiri', 'pas', 'log', 'in', 'ga', 'dapet', 'selang', 'hari', 'baru', 'dpt', 'otp', 'kalo', 'belum', 'bisa', 'munpuni', 'bikin', 'aplikasi', 'mending', 'via', 'web'] | ['log', 'out', 'pas', 'log', 'in', 'ga', 'dapet', 'otp', 'selang', 'dpt', 'otp', 'kalo', 'munpuni', 'bikin', 'aplikasi', 'mending', 'via', 'web'] | ['log', 'out', 'pas', 'log', 'in', 'ga', 'dapet', 'otp', 'selang', 'dpt', 'otp', 'kalo', 'munpuni', 'bikin', 'aplikasi', 'mending', 'via', 'web'] | ['dapet', 'selang', 'kalo', 'munpuni', 'bikin', 'aplikasi', 'mending'] |

4. Modelling

Proses pemodelan ini mengadopsi dua algoritma klasifikasi, yaitu Naïve Bayes dan *K-Nearest Neighbor (KNN)*, untuk mengategorikan sentimen data. Proses modellimg dapat dilihat pada gambar 3.

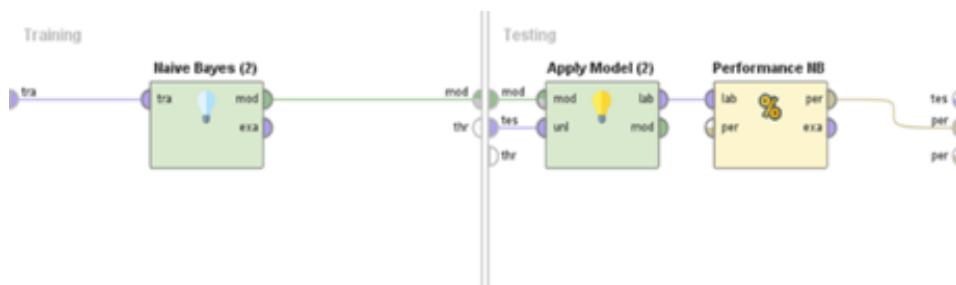


Gambar 3. Proses Modelling



Gambar 4. Pengujian Model K-Nearest Neighbor

Proses *Cross Validation* data dilakukan menggunakan algoritma klasifikasi *K-Nearest Neighbor*, yang mencakup tahap *training* dan *testing* menggunakan *apply model* dan *performance*. Setelah itu, pemodelan dilakukan dengan algoritma klasifikasi *K-Nearest Neighbor* untuk memperoleh nilai *accuracy*, *precision*, dan *recall* berdasarkan *performance vector*.



Gambar 5. Pengujian Model Naïve Bayes

Proses *Cross Validation* data dilakukan menggunakan algoritma klasifikasi Naïve Bayes, yang mencakup tahap *training* dan *testing* menggunakan *apply model* dan *performance*. Setelah itu, pemodelan dilakukan dengan algoritma klasifikasi Naïve Bayes untuk memperoleh nilai *accuracy*, *precision*, dan *recall* berdasarkan *performance vector*.

5. Evaluasi

Evaluasi dilakukan untuk menilai hasil model yang telah dibangun menggunakan *Confusion Matrix* serta *Cross Validation*. Metode *Cross Validation* yang digunakan adalah 10-fold *Cross Validation*, salah satu metode k-fold yang sering disarankan karena dapat memberikan perkiraan akurasi yang akurat. Berikut adalah hasil akurasi dari algoritma Naïve Bayes dan *K-Nearest Neighbor*.

accuracy: 64.94% +/- 3.34% (micro average: 64.94%)

| | true Positif | true Negatif | class precision |
|---------------|--------------|--------------|-----------------|
| pred. Positif | 288 | 343 | 45.64% |
| pred. Negatif | 145 | 616 | 80.95% |
| class recall | 66.51% | 64.23% | |

Gambar 6. Hasil Akurasi Naïve Bayes

Hasil yang diperoleh dari penerapan algoritma Naïve Bayes dalam mengklasifikasi 1392 data ulasan menunjukkan tingkat akurasi sebesar 64,94%. Nilai *precision* untuk kelas positif adalah 45,64%, sedangkan untuk kelas negatif adalah 80,95%.



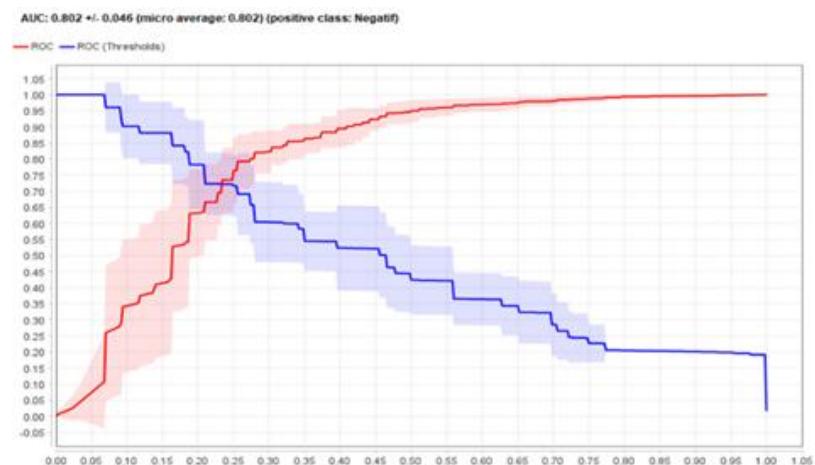
Gambar 7. Kurva Receiver Operating Characteristic Naïve Bayes

Grafik ROC yang diperoleh dari model Naïve Bayes memplot tingkat *true positive rate* terhadap *false positive rate*. Nilai AUC sebesar 0,530 menunjukkan bahwa luas area di bawah kurva ROC relatif kecil, yang mengindikasikan kinerja klasifikasi yang kurang baik.

| accuracy: 81.54% +/- 2.33% (micro average: 81.54%) | | | |
|--|--------------|--------------|-----------------|
| | true Positif | true Negatif | class precision |
| pred. Positif | 250 | 74 | 77,16% |
| pred. Negatif | 183 | 885 | 82,87% |
| class recall | 57,74% | 92,28% | |

Gambar 8. Hasil Akurasi K-Nearest Neighbor

Hasil yang diperoleh dari penerapan algoritma *K-Nearest Neighbor* dalam mengklasifikasi 1392 data ulasan menunjukkan tingkat akurasi sebesar 81,54%. Nilai *precision* untuk kelas positif adalah 77,16%, sedangkan untuk kelas negatif adalah 82,87%.



Gambar 9. Kurva Receiver Operating Characteristic K-Nearest Neighbor

Grafik ROC yang diperoleh dari model *K-Nearest Neighbor* memplot tingkat *true positive rate* terhadap *false positive rate*. Nilai AUC sebesar 0,802 menunjukkan bahwa luas area di bawah kurva ROC relatif besar, yang mengindikasikan kinerja klasifikasi yang sangat baik.

Confusion Matrix digunakan untuk menilai kinerja klasifikasi dengan melihat jumlah objek yang diklasifikasikan dengan benar maupun salah. Matriks ini menyajikan informasi mengenai data aktual dan hasil prediksi dalam sistem klasifikasi.

Tabel 3. Hasil Akurasi, Precision, Recall dan AUC

| | Akurasi | Precision | Recall | AUC |
|-----|---------|-----------|--------|-------|
| NB | 64,94% | 80,95% | 64,23% | 0,530 |
| KNN | 81,54% | 82,87% | 92,28% | 0,802 |

Berdasarkan Tabel 3, dapat disimpulkan bahwa algoritma *K-Nearest Neighbor* menunjukkan kinerja yang lebih unggul dibandingkan Naive Bayes dalam klasifikasi data. Algoritma *K-Nearest Neighbor* mencapai akurasi 81,54%, sementara Naive Bayes hanya memiliki akurasi 64,94%. Selain itu, *K-Nearest Neighbor* memiliki nilai recall dan precision yang lebih tinggi, yaitu 92,28% dan 80,95% secara berturut-turut. Nilai *AUC (Area Under Curve)* yang dihasilkan oleh *K-Nearest Neighbor* juga lebih besar, yaitu 0,802 dibandingkan dengan 0,530 dari Naive Bayes. Oleh karena itu, algoritma *K-Nearest Neighbor* dianggap lebih sesuai untuk klasifikasi data dalam penelitian ini.

D. Kesimpulan

Dalam penelitian ini, dilakukan analisis sentimen terhadap ulasan aplikasi Rosalia Indah Transport di *Google Play Store* menggunakan dua algoritma, yaitu Naive Bayes dan *K-Nearest Neighbor*. Berdasarkan evaluasi kinerja, dapat disimpulkan bahwa algoritma *K-Nearest Neighbor* menunjukkan performa yang lebih baik dibandingkan Naive Bayes dalam klasifikasi data. Algoritma *K-Nearest Neighbor* mencapai akurasi sebesar 81,54%, sementara Naive Bayes hanya mencapai 64,94%. Selain itu, *K-Nearest Neighbor* memiliki nilai *recall* dan *precision* yang lebih tinggi, yaitu masing-masing sebesar 92,28% dan 80,95%. Nilai *AUC (Area Under Curve)* dari *K-Nearest Neighbor* juga lebih besar, yakni 0,802 dibandingkan dengan 0,530 pada Naive Bayes. Oleh karena itu, algoritma *K-Nearest Neighbor* dinilai lebih cocok untuk klasifikasi data dalam penelitian ini, serta dapat memberikan kontribusi bagi pengembangan aplikasi Rosalia Indah Transport dan meningkatkan kualitas layanan bagi penggunanya.

Daftar Pustaka

- Ahmadi, M. I., Gustian, D., & Sembiring, F. (2021). Analisis Sentiment Masyarakat terhadap Kasus Covid-19 pada Media Sosial Youtube dengan Metode Naive bayes. In *Jurnal Sains Komputer & Informatika (J-SAKTI)* (Vol. 5, Issue 2).
- Aziz, A., & Fauziah. (2022). Analisis Sentimen Identifikasi Opini Terhadap Produk, Layanan dan Kebijakan Perusahaan Menggunakan Algoritma TF-IDF dan SentiStrength. In *Jurnal Sains Komputer & Informatika (J-SAKTI)* (Vol. 6, Issue 1). Abdul Aziz.
- Darwis, D., Shintya Pratiwi, E., Ferico, A., & Pasaribu, O. (2020). PENERAPAN ALGORITMA SVM UNTUK ANALISIS SENTIMEN PADA DATA TWITTER KOMISI PEMBERANTASAN KORUPSI REPUBLIK INDONESIA. In *Jurnal Ilmiah Edutic* (Vol. 7, Issue 1).
- Derajad Wijaya, H., & Dwiasnati, S. (2020). Implementasi Data Mining dengan Algoritma

- Naïve Bayes pada Penjualan Obat. *JURNAL INFORMATIKA*, 7(1).
<http://ejournal.bsi.ac.id/ejurnal/index.php/ji>
- Fajar Sidik, Ibnu Suhada, Azhar Haikal Anwar, & Firman Noor Hasan. (2022). *Analisis Sentimen Terhadap Pembelajaran Daring dengan Algoritma Naïve Bayes Classifier*.
- Irfan, M., & Erizal, E. (2024). Perbandingan Algoritma Naïve Bayes dengan K-Nearest Neighbor Untuk Analisis Sentimen Aplikasi InDrive di Playstore. *JURNAL MEDIA INFORMATIKA BUDIDARMA*, 8(3), 1535. <https://doi.org/10.30865/mib.v8i3.7780>
- Noor, F. H., & Dwijayanti, M. (2021). Analisis Sentimen Ulasan Pelanggan Terhadap Layanan Grab Indonesia Menggunakan Multinomial Naïve Bayes Classifier. In *JLK* (Vol. 4).
- Pradana Rachman, F., Santoso, H., & History, A. (2021). *Jurnal Teknologi dan Manajemen Informatika Perbandingan Model Deep Learning untuk Klasifikasi Sentiment Analysis dengan Teknik Natural Languange Processing Article Info ABSTRACT*. 7(2), 103–112.
<http://http://jurnal.unmer.ac.id/index.php/jtmi>
- Rachman, R., Handayani, R. N., & Artikel, I. (2021). Klasifikasi Algoritma Naive Bayes Dalam Memprediksi Tingkat Kelancaran Pembayaran Sewa Teras UMKM. *JURNAL INFORMATIKA*, 8(2). <http://ejournal.bsi.ac.id/ejurnal/index.php/ji>
- Rahutomo, F., Retno, A., & Ririd, T. H. (2019). *EVALUASI DAFTAR STOPWORD BAHASA INDONESIA*. 6(1), 41–48. <https://doi.org/10.25126/jtiik.201861226>
- Rifai, B., Danang Febryanto, B., Yulianto, F., Reflianah, N., & Tinggi Manajemen Informatika dan Komputer Nusa Mandiri, S. (2021). *Analisis Sentimen Opini Publik Terhadap Penerapan Kebijakan Social Distancing Dalam Pencegahan Covid-19*. 23(1).
<http://ejournal.bsi.ac.id/ejurnal/index.php/paradigma/issue/archive/>
- Septi Rizqiah, A., & Rahman Kadafi, A. (2022). Analisis Sentimen Pengguna Twitter Terhadap Vaksin Booster Menggunakan Metode Naïve Bayes Dan K-Nearst Neighbor. *Jurnal Sains Dan Teknologi*, 4(1), 1–5.
- Syafrizal, S., Afdal, M., & Novita, R. (2023). Analisis Sentimen Ulasan Aplikasi PLN Mobile Menggunakan Algoritma Naïve Bayes Classifier dan K-Nearest Neighbor. *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, 4(1), 10–19.
<https://doi.org/10.57152/malcom.v4i1.983>
- Syahril Dwi Prasetyo, Shofa Shofiah Hilabi, & Fitri Nurapriani. (2023). Analisis Sentimen Relokasi Ibukota Nusantara Menggunakan Algoritma Naïve Bayes dan KNN. *Jurnal KomtekInfo*, 1–7. <https://doi.org/10.35134/komtekinfo.v10i1.330>
- Wijaya, A., Rivaldo, M., & Rizky Pribadi, M. (2024). Analisis Sentimen Terhadap Aplikasi Mitra Darat Menggunakan Algoritma Naive Bayes Classifier dan K-Nearest Neighbor. In *Applied Information Technology and Computer Science* (Vol. 3, Issue 1).
<https://jurnal.politap.ac.id/index.php/aicoms>