

Analisis Sentimen Terhadap Aplikasi Mitra Darat Menggunakan Algoritma Naive Bayes Classifier dan K-Nearest Neighbor

Ananda Wijaya¹, Mario Rivaldo², Muhammad Rizky Pribadi³,
Program Studi Informatika, Fakultas Ilmu Komputer dan Rekayasa

Universitas Multi Data Palembang, Palembang
Jl. Rajawali No.14, 9 Ilir, Kec. Ilir Tim. II, Kota Palembang, Sumatera Selatan, 30113, Indonesia
¹wijayaanda28@mhs.mdp.ac.id, ²mariorivaldo@mhs.mdp.ac.id, ³rizky@mhs.mdp.ac.id

Abstrak. Industri transportasi sekarang menjadi elemen penting seiring dengan berkembangnya jaman terutama untuk generasi muda sekarang. Mitra Darat sendiri juga salah satu dari industri tersebut. Aplikasi yang memungkinkan untuk pengguna dengan mudah mengetahui jadwal keberangkatan bus yang akan mereka tumpangi dimana pun dan kapan pun di perangkat seluler mereka. Ulasan pasti diberikan untuk setiap aplikasi yang tersedia baik positif dan negatif. Dengan ini, kami mencoba melakukan penelitian analisis sentimen untuk aplikasi Mitra Darat melalui ulasan komentar dari google play store agar kami dapat mengidentifikasi sentimen yang terkait dengan penggunaan aplikasi Mitra Darat, serta memberikan wawasan beharga kepada penyedia layanan transportasi darat untuk memahami pandangan pengguna dan meningkatkan pelayanan pengguna dari hasil analisis sentimen kami. Algoritma yang digunakan kami ialah KNN dan NBC. Kedua algoritma ini sudah umum digunakan oleh banyak orang karena keahlian dalam mengklasifikasi data analisis sentimen dan juga popular di kalangan peneliti. Berdasarkan hasil pengujian kami bisa disimpulkan untuk model analisis sentimen kami yang dirancang menggunakan algoritma NB menampilkan performa akurasi lebih tinggi dibandingkan KNN. Akurasi model NB mencapai 99,28%, sedangkan KNN mendapatkan akurasi sebesar 80%. Ini menunjukkan bahwa algoritma naïve bayes lebih cocok digunakan untuk mendapatkan keakuratan yang maksimal dibandingkan menggunakan k-nearest neighbor.

Kata Kunci: *Sentimen Analisis, Mitra Darat, NB, KNN, Klasifikasi, Google Play Store, Data Mining.*

1 Pendahuluan

Dalam era digital yang terus berkembang, aplikasi mitra darat telah menjadi elemen penting dalam industri transportasi khususnya bus. Aplikasi mitra darat ini memungkinkan pengguna untuk dengan mudah mengetahui jadwal keberangkatan bus melalui perangkat seluler mereka. Seiring dengan meningkatnya penggunaan aplikasi mitra darat, sangat penting untuk memahami sentimen pengguna terhadap aplikasi tersebut. Analisis sentimen adalah metode yang luas digunakan untuk mendapatkan wawasan tentang pandangan dan perasaan pengguna mengenai suatu produk atau layanan.

Dalam studi ini, kami melaksanakan melakukan analisis sentimen terhadap aplikasi mitra darat dengan memanfaatkan algoritma *Naive Bayes Classifier* dan *K-Nearest Neighbor*. Algoritma *Naive Bayes Classifier* merupakan teknik klasifikasi yang populer dalam pengolahan bahasa alami dan analisis sentimen. Algoritma ini berbasis pada teorema *Bayes* yang melibatkan perhitungan peluang untuk mengkategorikan data ke dalam kategori yang sesuai. Dalam konteks analisis sentimen, algoritma *Naive Bayes Classifier* dapat mengklasifikasikan ulasan pengguna aplikasi mitra darat ke dalam klasifikasi sentimen positif, negatif, atau netral.

Kami juga menggunakan algoritma *k*-NN dalam analisis ini. *K*-NN merupakan teknik klasifikasi sederhana namun efektif, yang bekerja dengan prinsip bahwa objek sama cenderung terletak di satu sama lain dalam ruang fitur. Algoritma ini mengklasifikasikan sebuah data baru dari mayoritas label *k* terdekatnya dalam dataset pelatihan. Pada analisis sentimen, *K*-NN dapat digunakan untuk menentukan kategori sentimen dari sebuah ulasan berdasarkan kemiripannya dengan ulasan-ulasan lain yang telah dikategorikan sebelumnya. *K*-NN dikenal karena kemampuannya menangani data yang non-linear dan tidak memerlukan asumsi mengenai distribusi data, menjadikannya alat yang fleksibel dalam berbagai tugas klasifikasi, termasuk sentimen analisis.

Tujuan penelitian ini ialah untuk mengidentifikasi sentimen yang terkait dengan penggunaan aplikasi mitra darat. Dengan demikian, hasil analisis sentimen dapat memberikan wawasan berharga kepada penyedia layanan transportasi darat untuk memahami pandangan pengguna dan meningkatkan pengalaman pengguna secara keseluruhan.

Penelitian sebelumnya telah menggunakan berbagai metode untuk melakukan analisis sentimen terhadap berbagai jenis data, termasuk ulasan produk dan layanan. Salah satu studi terkait yang relevan adalah penelitian Dzul Asfi Warraihan[1] yang menerapkan algoritma NBC dan KNN untuk analisis sentimen terkait penggunaan transportasi online Maxim di Instagram. Studi ini menghasilkan hasil yang menjanjikan untuk *Naïve Bayes Classifier* dengan tingkat akurasi yang tinggi yakni 94% dalam mengklasifikasikan sentimen pengguna, sedangkan penggunaan *KNN* hanya menghasilkan akurasi sebesar 84% yang Dimana *Naïve Bayes Classifier* lebih unggul dalam mengklasifikasikan sentimen pengguna.

2 Studi Terkait

Beberapa studi terkait berkaitan dengan penelitian ini, yang pertama yang ditulis oleh Sartika Mandasari[2] menyediakan pandangan berharga dalam bidang analisis sentimen transportasi online. Penelitian ini dimaksudkan untuk mengotomatisasi pengelompokan sentimen terkait ulasan aplikasi, memungkinkan pengembang dan pengguna untuk memahami pandangan pengguna tentang aplikasi yang digunakan.

Hasil penelitian mengindikasi bahwa klasifikasi yang diambil dari tweet sentimen netral menunjukkan tingkat kualitas layanan Grab Indonesia dengan layanan terbaik pada GrabFood sebesar 42,29%, GrabBike sebesar 32,57%, dan GrabCar sebesar 25,14%. Pengujian analisis sentimen pada layanan Grab Indonesia dengan metode Multinomial Naive Bayes Classifier ini menghasilkan akurasi sebesar 86,57%.

Selanjutnya, penelitian oleh Syahril Dwi Prasetyo[3]. Data yang diambil peneliti berasal dari ulasan komentar twiiter terkait pemindahan ibu kota, dengan total 2000 data komentar pengguna di platform Twitter. Tujuan penelitian ini ialah untuk menganalisis sentimen pengguna terkait topik pemindahan ibu kota.

Berdasarkan hasil penelitian, penggunaan algoritma k-NN pada penelitian ini memiliki tingkat keakuratan tertinggi untuk sentimen komentar twitter dengan nilai 88,12%, sedangkan untuk dengan NB memiliki tingkat keakuratan sebesar 82,27. Penelitian ini berasumsi bahwa menggunakan algoritma k-NN terdapat akurasi paling tinggi karena rata-rata yang dihasilkan lebih unggul dibandingkan dengan algoritma lain.,

Dalam penelitian sebelumnya yang dilakukan oleh Eka Wahyu Sholeha[4], penelitian ini menggunakan dataset yang berasal dari 3 fanpage Facebook agen penjualan online untuk menganalisis 1500 komentar di bulan september, oktober, dan november 2020.

Hasil evaluasi menunjukkan akurasi sebesar 85,11% untuk penggunaan algoritma k-NN dan 83,02% untuk penggunaan algoritma NB, dapat disimpulkan bahwa rata-rata penelitian menyimpulkan penggunaan algoritma k-NN jauh lebih akurat dibandingkan penggunaan Naïve Bayes Classifier.

3 Metode Penelitian

3.1 Dataset

Proses pada mendapatkan dataset, kami menggunakan Jupyter Notebook. Jupyter Notebook adalah alat yang sangat populer di kalangan data scientist dan peneliti karena kemampuannya untuk mengintegrasikan kode, visualisasi, dan teks penjelasan dalam satu dokumen interaktif. Jupyter Notebook menyediakan platform interaktif yang memungkinkan penulisan dan eksekusi kode Python dalam blok-blok sel.

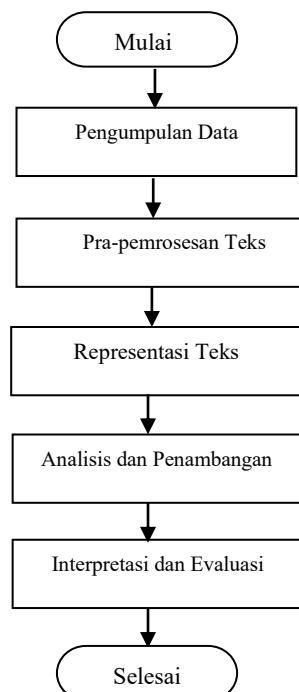
Dataset digunakan untuk penelitian ini didapat dari ulasan pengguna pada halaman Play Store yang termasuk bagian dari Google untuk aplikasi mitra darat. Dataset yang dikumpulkan berisi 200 komentar dengan atribut yang meliputi teks komentar, rating yang diberikan pengguna, dan tanggal komentar tersebut diposting. Rating

tersebut membantu dalam melakukan analisis awal terkait sentimen pengguna, meskipun fokus utama tetap pada teks komentar.

3.2 Text Mining

Text mining dan data mining dianggap satu ilmu karena algoritma yang sama dapat digunakan pada kedua konsep data mining. Namun, keduanya berbeda karena mining data berkaitan dengan data terstruktur, sedangkan teks berkaitan dengan fitur-fitur tertentu dan relative tidak terstruktur serta memerlukan pra-pemrosesan. Text mining adalah bidang yang berkaitan erat dengan atau pemrosesan bahasa alami pemrosesan Bahasa alami atau Natural Language Processing (NLP)[5].

Sedangkan menurut Fira Fathonah & Asti Herliana[6], text mining melibatkan proses eksplorasi dan analisis data teks jumlah besar yang tidak terstruktur. Proses ini dibantu oleh perangkat lunak yang mampu mengenali konsep, pola, topik, kata kunci, dan atribut lainnya dalam data tersebut. Agar megetahui struktur proses Text Mining dapat dilihat pada gambar 1.



Gambar. 1. Flowchart Text Mining

3.3 Opinion Mining

Opinion mining atau analisis sentimen adalah cabang ilmu dari data mining yang berfungsi untuk menganalisis, memproses, dan mengekstrak data tekstual terkait entitas, seperti layanan, produk, individu, organisasi, peristiwa, atau topik tertentu. Tujuan dari analisis ini adalah untuk memperoleh informasi dari sekumpulan data yang tersedia. Analisis sentimen merupakan bidang baru dalam Natural Language Processing yang bertujuan mengidentifikasi subjektivitas dalam teks serta melakukan klasifikasi sentimen pada opini[7].

Analisis sentimen adalah salah satu bidang penting dalam machine learning atau pembelajaran mesin. Penelitian yang menganalisis opini, evaluasi, sentimen, sikap, penilaian, dan emosi terhadap entitas seperti layanan, individu, isu, topik, dan atribut terkait juga didasarkan pada analisis sentimen[8].

3.4 KNN

K-Nearest Neighbor (KNN) adalah bagian dari kelompok instance-based learning dan merupakan salah satu teknik lazy learning. Algoritma ini bekerja dengan mencari sekelompok k objek dalam data pelatihan yang

paling mirip atau paling dekat dengan objek dalam data baru atau data pengujian. Dibutuhkan sistem klasifikasi yang mampu mencari informasi secara efektif[9].

Untuk menentukan hasil klasifikasi, algoritma K-Nearest Neighbor mengevaluasi jarak terdekat antara objek dengan masing-masing kelompok[10]. Algoritma KNN sangat sederhana tetapi cukup efektif untuk mengkategorikan teks. Namun, KNN memiliki kelemahan karena sangat bergantung pada parameter k dan membutuhkan komputasi yang tinggi[11]. Untuk persamaan atau rumus dari algoritma KNN dapat dihitung dengan rumus pada persamaan 1.

$$d(x_i, x_j) = \sqrt{\sum_{r=1}^n ((a_r(x_i) - a_r(x_j))^2)} \quad (1)$$

3.4 Naive Bayes Classifier

Naive Bayes Classifier atau dengan istilah lain multinomial naïve bayes ialah metode yang dipakai untuk mengklasifikasikan sekumpulan dokumen. Algoritma ini memanfaatkan prinsip probability dan statistic yang ditemukan oleh ilmuwan Inggris Thomas Bayes. Metode NB melibatkan dua langkah dalam proses klasifikasi teks, yaitu langkah pelatihan dan langkah pengujian (klasifikasi)[12].

Menurut Mohamad Efendi Lasulika[13], Naive Bayes merupakan sebuah teknik prediksi probabilitas yang menggunakan Teorema Bayes. Metode ini didasarkan pada asumsi bahwa setiap fitur dalam data adalah independen satu sama lain. Teknik ini merupakan salah satu bentuk sederhana dari klasifikasi, dengan persamaan yang terkait terlihat pada persamaan 2.

$$P(x|y) = \frac{P(y|x)P(x)}{P(y)} \quad (2)$$

4 Hasil dan Pembahasan

4.1 Informasi Dataset

Data yang dianalisis untuk penelitian ini adalah Kumpulan ulasan aplikasi Mitra Darat yang didapatkan dari Google Play Store melalui teknik web scraping menggunakan software Jupyter Notebook. Dataset terdiri dari 200 entri dengan tiga kolom, yaitu Score, Review dan Sentimen.

4.2 Prapemrosesan Data

Pada tahap ini, proses prapemrosesan atau praprocessing sebuah data dilakukan dengan mengubah semua teks menjadi huruf kecil, menghilangkan tanda baca, dan menghapus karakter yang tidak diperlukan. Selanjutnya, dilakukan tokenisasi, yaitu memecah kalimat menjadi token atau kata, satu per satu. Selain itu, diterapkan penghapusan kata umum (stopword removal) dan stemming untuk mengubah kata-kata menjadi bentuk dasarnya yang relevan.

4.3 Model Analisis Sentimen

Analisis sentimen dalam penelitian ini menggunakan algoritma NB dan KNN. Penelitian ini menggunakan RapidMiner Studio untuk menjalankan klasifikasi model. Proses pelatihan dibagi jadi dua bagian, yakni data training dan data testing atau klasifikasi dengan proporsi 70:30. Di tahap pelatihan, algoritma KNN dan NB dipakai untuk klasifikasi tiap dokumen ke dalam kategori sentimen positif atau negatif. Bedasarkan hasil studi, dinyatakan bahwa model ini dapat melakukan analisis sentimen untuk dataset ulasan pengguna aplikasi Mitra Darat dengan total akurasi 84% untuk algoritma KNN dan 99% untuk algoritma NB. Ini membuktikan bahwa algoritma NB dapat memberikan hasil analisis sentimen yang sangat baik.

4.4 Evaluasi Model

Untuk mengevaluasi model yang telah dibuat, digunakan metode confusion matrix. Confusion matrix digunakan untuk menampilkan total hasil klasifikasi yang tepat dan tidak tepat dari model. Terdapat empat nilai penting pada matrix confusion yakni, Nilai *True Positive*(TP), Nilai *True Negative*(TN), Nilai *False Positive*(FP), dan Nilai *False Negative*(FN).

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \times 100\% \quad (3)$$

$$Precision = \frac{TP}{TP + FP} \times 100\% \quad (4)$$

$$Recall = \frac{TP}{FN + TP} \times 100\% \quad (5)$$

Berdasarkan struktur tabel confusion matrix, maka confusion matrix terdiri dari 4 nilai: TP adalah jumlah prediksi yang benar Ketika model memprediksi positif dan data tersebut memang positif. TN adalah jumlah prediksi yang benar ketika model memprediksi negatif dan data tersebut memang negatif. FP adalah jumlah prediksi yang salah ketika model memprediksi positif, tetapi data tersebut sebenarnya negatif. FN ialah jumlah prediksi yang salah Ketika model memprediksi negatif, tetapi data tersebut sebenarnya positif[14].

Untuk mengukur performance matrixada beberapa paremeter yang dijadikan acuan dalam menilai kinrja algoritma diantaranya adalah akurasi, presisi dan recall. Akurasi adalah rasio antara jumlah prediksi yang benar dengan total jumlah data. Nilai presisi adalah perbandingan antara jumlah prediksi benar dengan total jumlah prediksi yang diklasifikasikan sebagai benar[15]. Untuk hasil confusion matrix dan evaluasi matrix untuk percobaan dengan algoritma KNN di Rapid Miner Studio dapat dilihat pada tabel 1 dan 2.

Tabel 1. Confusion Matrix

Confusion Matrix	Hasil Klasifikasi <i>k</i> -NN
True Positive	70
False Positive	3
False Neutral	9
False Negative	1
True Neutral	3
True Negative	45

Di tabel 1 terdapat tambahan neutral karena labeling di dataset kami terdapat 3 sentimen yakni positive, neutral, dan negative.

Tabel 2. Evaluasi Matrix

Recall	98,59%
Precision	85,37%
Accuracy	84,89%

Sementara itu, kami mencoba algoritma lain yakni NB dengan dataset yang sama, untuk hasil confusion matrix dan evaluasi matrix untuk percobaan dengan algoritma Naïve Bayes Classifier bisa dilihat di tabel 3 dan 4.

Tabel 3. Confusion Matrix

Confusion Matrix	Hasil Klasifikasi <i>k</i> -NN
True Positive	70
False Positive	0
False Neutral	0
False Negative	1
True Neutral	20
True Negative	48

Tabel 2. Evaluasi Matrix

Recall	98,59%
Precision	100%
Accuracy	99,28%

5 Kesimpulan

Dari hasil penelitian, dapat disimpulkan bahwa model Analisis Sentimen yang dikembangkan menggunakan algoritma NB menunjukkan kinerja akurasi yang cukup tinggi dibandingkan KNN. Akurasi model NB mencapai 99,28%, sedangkan KNN mendapatkan akurasi dengan nilai 80%. Evaluasi lainnya juga menghasilkan hasil yang sama, membuktikan bahwa model dengan NB ini efektif dalam melakukan analisis sentimen pada dataset ulasan pengguna aplikasi Mitra Darat. Hasil penelitian ini memberikan kepastian bahwa algoritma KNN bisa diharapkan untuk tugas analisis sentimen pada dataset yang serupa.

Tetapi, ada beberapa faktor yang dapat mempengaruhi akurasi model, seperti kualitas dataset yang digunakan, pemilihan fitur pada Rapid Miner Studio, dan parameter pada algoritma. Oleh karena itu, untuk menambah keakuratan model, diharapkan penelitian berikutnya bisa focus pada meningkatkan elemen-elemen yang dijelaskan sebelumnya.

Referensi

- [1] D. A. Warraihan, I. Permana, M. Mustakim, and R. Novita, "Analisis Sentimen Pengguna Transportasi Online Maxim Pada Instagram Menggunakan Naïve Bayes Classifier dan K-Nearest Neighbor," *J. Media Inform. Budiadarma*, vol. 7, no. 3, pp. 1134–1143, 2023, doi: 10.30865/mib.v7i3.6336.
- [2] S. Mandasari, B. H. Hayadi, and R. Gunawan, "Sentiment Analysis of Online Transportation Users Towards Grab Indonesia Services Using Multinomial Naïve Bayes Classifier," *J-SISKO TECH (Jurnal Teknol. Sist. Inf. dan Sist. Komput. TGD)*, vol. 5, no. 2, p. 118, 2022, [Online]. Available: <https://ojs.trigunadharma.ac.id/index.php/jsk/article/view/5635>
- [3] Syahrial Dwi Prasetyo, Shofa Shofiah Hilabi, and Fitri Nurapriani, "Analisis Sentimen Relokasi Ibukota Nusantara Menggunakan Algoritma Naïve Bayes dan KNN," *J. KomtekInfo*, vol. 10, pp. 1–7, 2023, doi: 10.35134/komtekinfo.v10i1.330.
- [4] E. W. Sholeha, S. Yunita, R. Hammad, V. C. Hardita, and K. Kaharuddin, "Analisis Sentimen Pada Agen Perjalanan Online Menggunakan Naïve Bayes dan K-Nearest Neighbor," *JTIM J. Teknol. Inf. dan Multimed.*, vol. 3, no. 4, pp. 203–208, 2022, doi: 10.35746/jtim.v3i4.178.
- [5] A. K. Fauziyyah, "Analisis Sentimen Pandemi Covid19 Pada Streaming Twitter Dengan Text Mining Python," *J. Ilm. SINUS*, vol. 18, no. 2, p. 31, 2020, doi: 10.30646/sinus.v18i2.491.
- [6] F. Fathonah and A. Herliana, "Penerapan Text Mining Analisis Sentimen Mengenai Vaksin Covid - 19 Menggunakan Metode Naïve Bayes," *J. Sains dan Inform.*, vol. 7, no. 2, pp. 155–164, 2021, doi: 10.34128/jsi.v7i2.331.
- [7] V. K. S. Que, A. Iriani, and H. D. Purnomo, "Analisis Sentimen Transportasi Online Menggunakan Support Vector Machine Berbasis Particle Swarm Optimization," *J. Nas. Tek. Elektro dan Teknol. Inf.*, vol. 9, no. 2, pp. 162–170, 2020, doi: 10.22146/jnteti.v9i2.102.
- [8] N. P. G. Naraswati, R. Nooraeni, D. C. Rosmilda, D. Desinta, F. Khairi, and R. Damaiyanti, "Analisis Sentimen Publik dari Twitter Tentang Kebijakan Penanganan Covid-19 di Indonesia dengan Naïve Bayes Classification," *Sistemasi*, vol. 10, no. 1, p. 222, 2021, doi: 10.32520/stmsi.v10i1.1179.
- [9] A. M. Argina, "Penerapan Metode Klasifikasi K-Nearest Neigbor pada Dataset Penderita Penyakit Diabetes," *Indones. J. Data Sci.*, vol. 1, no. 2, pp. 29–33, 2020, doi: 10.33096/ijodas.v1i2.11.
- [10] D. Pajri, Y. Umaidah, and T. N. Padilah, "K-Nearest Neighbor Berbasis Particle Swarm Optimization untuk Analisis Sentimen Terhadap Tokopedia," *J. Tek. Inform. dan Sist. Inf.*, vol. 6, no. 2, pp. 242–253, 2020, doi: 10.28932/jutisi.v6i2.2658.
- [11] S. Syafrizal, M. Afdal, and R. Novita, "Analisis Sentimen Ulasan Aplikasi PLN Mobile Menggunakan Algoritma Naïve Bayes Classifier dan K-Nearest Neighbor," *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 4, no. 1, pp. 10–19, 2023, doi: 10.57152/malcom.v4i1.983.
- [12] A. D. Wibisono, S. Dadi Rizkiono, and A. Wantoro, "Filtering Spam Email Menggunakan Metode Naïve Bayes," *TELEFORTECH J. Telemat. Inf. Technol.*, vol. 1, no. 1, 2020, doi: 10.33365/tft.v1i1.685.
- [13] M. E. Lasulika, "Komparasi Naïve Bayes, Support Vector Machine Dan K-Nearest Neighbor Untuk Mengetahui Akurasi Tertinggi Pada Prediksi Kelancaran Pembayaran Tv Kabel," *Ilk. J. Ilm.*, vol. 11, no. 1, pp. 11–16, 2019, doi: 10.33096/ilkom.v11i1.408.11-16.
- [14] M. Luthfi Bangun Permadi and R. Gumilang, "Penerapan Algoritma CNN (Convolutional Neural Network) Untuk Deteksi Dan Klasifikasi Target Militer Berdasarkan Citra Satelit," *J. Sos. Teknol.*, vol. 4, no. 2, pp. 134–143, 2024, doi: 10.59188/jurnalsoftech.v4i2.1138.
- [15] A. Nugroho and Y. Religia, "Analisis Optimasi Algoritma Klasifikasi Naïve Bayes menggunakan Genetic Algorithm dan Bagging," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 5, no. 3, pp. 504–510, 2021, doi: 10.29207/resti.v5i3.3067.