

Analisis Klasifikasi Jenis Layanan Ojek Online Menggunakan Algoritma Naive Bayes pada Dataset Dummy Ojol Transactions

Muhammad Fauzan Fadhilah¹, Hasbi Firmansyah², Wahyu Asriyani³

Program Studi Informatika Fakultas Teknik dan Ilmu Komputer, Universitas Pancasakti Tegal,
Kota Tegal Indonesia ^{1,2}

Program studi Pendidikan Bahasa dan Sastra Indonesia, Fakultas Keguruan dan Ilmu Pendidikan,
Universitas Pancasakti Tegal, Kota Tegal, Indonesia³

*Email fauzanfadillah359@gmail.com, hasbifirmansyah@upstegal.ac.id, asriyani1409@gmail.com

Sejarah Artikel:

Diterima 18-12-2025
Disetujui Tgl-Bln-Thn
Diterbitkan Tgl-Bln-Thn

ABSTRACT

The rapid growth of online motorcycle taxi (ojol) services as part of the digital economy ecosystem has generated a substantial volume of diverse transaction data. This data holds significant potential for supporting analytical processes such as identifying service usage patterns, predicting user needs, and enabling data-driven decision-making. One essential analytical task is the classification of service types based on available transaction attributes. To address this need, this study conducts a classification analysis of online ride-hailing service types using the Naive Bayes algorithm on the "Ojol Transactions" dummy dataset. The research consists of several stages, including data preprocessing, handling of missing values, target attribute determination, model training, and performance evaluation using RapidMiner. The Naive Bayes algorithm was selected due to its efficiency in classifying structured data and its compatibility with simple probabilistic assumptions. The evaluation results show that the classification model achieved an accuracy of 100%, indicating that all test data were correctly classified. These findings demonstrate the strong potential of the Naive Bayes algorithm for analyzing ride-hailing transaction classifications, particularly when applied to datasets with clearly separated classes.

Keywords: Classification, Data Mining, Digital Transactions, Machine Learning, Naive Bayes, Ojol Services, RapidMiner

ABSTRAK

Perkembangan layanan ojek online (ojol) sebagai bagian dari ekosistem ekonomi digital telah menghasilkan volume data transaksi yang sangat besar dan beragam. Data tersebut berpotensi dimanfaatkan untuk mendukung proses analisis pola penggunaan layanan, prediksi kebutuhan pengguna, serta pengambilan keputusan berbasis data. Salah satu kebutuhan analitis yang penting adalah kemampuan untuk mengklasifikasikan jenis layanan berdasarkan atribut transaksi yang tersedia. Untuk memenuhi kebutuhan tersebut, penelitian ini melakukan analisis klasifikasi jenis layanan ojek online menggunakan algoritma Naive Bayes pada dataset dummy "Ojol Transactions". Penelitian ini dilakukan melalui beberapa tahapan, meliputi praproses data, penanganan nilai hilang, penentuan atribut target, pelatihan model, serta evaluasi performa menggunakan perangkat lunak RapidMiner. Algoritma Naive Bayes dipilih karena kemampuannya dalam melakukan klasifikasi secara efisien pada data terstruktur serta karakteristiknya yang sesuai dengan asumsi probabilistik sederhana. Hasil evaluasi menunjukkan bahwa model klasifikasi menghasilkan tingkat akurasi sebesar 100%, yang mengindikasikan bahwa seluruh data uji berhasil diprediksi secara tepat. Temuan ini menunjukkan bahwa Naive Bayes memiliki potensi yang kuat untuk diterapkan pada analisis

klasifikasi transaksi ojol, khususnya pada dataset dengan pemisahan kelas yang jelas.

Kata Kunci: Data Mining, Klasifikasi, Layanan Ojol, Machine Learning, Naive Bayes, RapidMiner, Transaksi Digital

Bagaimana Cara Sitasi Artikel ini:

Muhammad Fauzan Fadhilah, Hasbi Firmansyah, & Wahyu Asriyani. (2025). Analisis Klasifikasi Jenis Layanan Ojek Online Menggunakan Algoritma Naive Bayes pada Dataset Dummy Ojol Transactions. Jejak Digital: Jurnal Ilmiah Multidisiplin, 2(1), 1215-1224. <https://doi.org/10.63822/amhgek21>

PENDAHULUAN

Layanan ojek online telah menjadi bagian penting dari sistem transportasi urban modern di berbagai negara, terutama di Indonesia, karena menawarkan kemudahan mobilitas, fleksibilitas penggunaan, dan integrasi berbagai layanan seperti transportasi, pengantaran barang, dan makanan dalam satu platform digital (*ride-hailing*). Pertumbuhan pesat layanan ini menghasilkan volume data transaksi yang sangat besar setiap hari, namun pemanfaatan optimal terhadap data tersebut masih menjadi tantangan utama karena kompleksitasnya yang tinggi dan karakteristik data yang heterogen (*data mining*). Di sisi lain, kemampuan untuk mengklasifikasikan jenis layanan (mis. motor, mobil, food delivery, delivery barang) secara otomatis dapat memberikan wawasan penting bagi perusahaan dalam meningkatkan strategi operasional, pengalaman pengguna, serta kualitas layanan. Sayangnya, sebagian besar penelitian terdahulu yang menggunakan metode klasifikasi seperti **Naive Bayes** masih berfokus pada topik yang berbeda, seperti analisis sentimen terhadap aplikasi transportasi (*sentiment analysis*) atau prediksi kepuasan pengguna daripada klasifikasi layanan itu sendiri. Misalnya, studi yang menilai sentimen publik terhadap layanan transportasi online menggunakan Naive Bayes menunjukkan efektivitas pendekatan ini dalam mengelompokkan opini pengguna (*sentiment classification*) tetapi tidak secara langsung menerapkannya pada klasifikasi tipe layanan (*transportation sentiment analysis*). Penelitian lain menggunakan Naive Bayes untuk menilai kepuasan pengguna Gojek melalui analisis kombinasinya dengan Sistem Informasi Geografis (GIS) yang menunjukkan kapasitas model ini dalam mengukur peluang kepuasan pengguna berdasarkan data atribut transaksi, meskipun fokusnya tetap pada kepuasan pengguna, bukan klasifikasi layanan secara eksplisit.

Studi praktis lain telah membuktikan efektivitas Naive Bayes dalam berbagai domain klasifikasi, termasuk prediksi insiden lalu lintas menggunakan Naive Bayes yang diaplikasikan pada data risiko lalu lintas untuk manajemen risiko transportasi dan klasifikasi sentimen data transportasi di Twitter agar mampu memisahkan berbagai jenis opini publik terhadap layanan transportasi menggunakan metode klasifikasi berbasis machine learning. Sedangkan beberapa penelitian lain membandingkan performa Naive Bayes dengan algoritma lain seperti Support Vector Machine (SVM) dalam konteks klasifikasi teks atau data layanan, yang masing-masing menunjukkan bahwa Naive Bayes tetap menjadi metode baseline yang kuat dan cepat dalam menangani data dengan asumsi independensi fitur. Namun, masih sedikit penelitian yang menerapkan Naive Bayes secara langsung pada klasifikasi jenis layanan *ride-hailing* berdasarkan fitur transaksi yang mencakup variabel numerik maupun kategorikal seperti jarak, durasi, tarif, lokasi, dan metode pembayaran.

Untuk mengatasi kekosongan tersebut, penelitian ini mengusulkan penerapan Naive Bayes sebagai metode klasifikasi untuk mengidentifikasi jenis layanan ojek online berdasarkan dataset *Dummy Ojol Transactions*. Pendekatan ini dimulai dari proses pengumpulan dan praproses data, penanganan nilai kosong, penetapan variabel target berupa *service type*, hingga pelatihan model dan evaluasi performa klasifikasi menggunakan metrik akurasi, presisi, recall, dan F1-score. Dengan strategi ini, penulis berharap Naive Bayes mampu mempelajari pola distribusi probabilitas dari ciri khas setiap jenis layanan sehingga dapat memisahkan kategori layanan dengan baik. Selain itu, integrasi pendekatan ini ke dalam proses bisnis nyata *ride-hailing* dapat membantu penyedia layanan untuk memahami demand jenis layanan tertentu dan merumuskan rekomendasi operasional seperti prediksi permintaan untuk setiap kategori layanan pada waktu tertentu, distribusi sarana transportasi, atau penawaran promosi khusus layanan yang kurang diminati.

Tujuan penelitian ini adalah menerapkan metode Naive Bayes untuk melakukan klasifikasi jenis layanan ojek online berdasarkan transaksi, mengidentifikasi pengaruh fitur transaksi terhadap performa klasifikasi, dan mengevaluasi performa model Naive Bayes berdasarkan metrik evaluasi standar. Dengan demikian, penelitian diharapkan memberikan kontribusi teoritis dan praktis dalam bidang *machine learning* dan *data mining* untuk sistem transportasi digital, sekaligus menawarkan bukti empiris bahwa Naive Bayes dapat digunakan sebagai model prediktif untuk klasifikasi layanan dalam konteks *ride-hailing*. Kebaharuan penelitian ini terletak pada penerapan Naive Bayes untuk klasifikasi jenis layanan ojek online secara langsung berdasarkan dataset transaksi, bukan hanya untuk analisis sentimen atau prediksi kepuasan pengguna, serta penyajian evaluasi komprehensif dalam konteks struktur data campuran yang lebih dekat dengan realita operasional platform ojek online.

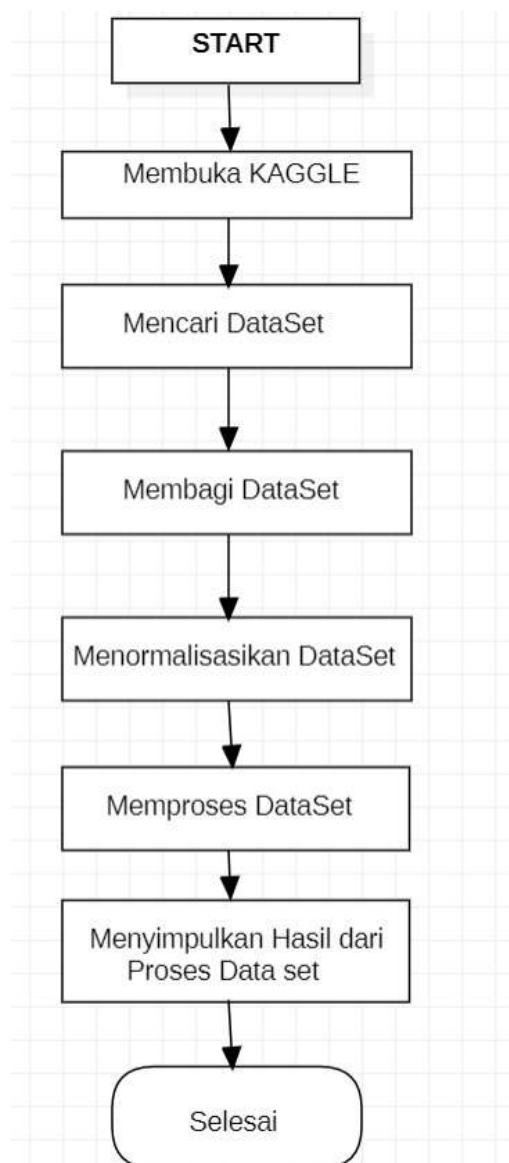
METODOLOGI PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif dengan metode klasifikasi berbasis *machine learning* untuk menganalisis jenis layanan ojek online. Algoritma yang digunakan adalah **Naive Bayes**, yang termasuk ke dalam metode klasifikasi probabilistik dan banyak digunakan dalam penelitian data mining karena kesederhanaan, efisiensi komputasi, serta performanya yang baik pada data terstruktur dan multi-kelas. Pemilihan Naive Bayes dalam penelitian ini didasarkan pada sejumlah penelitian sebelumnya yang menunjukkan bahwa algoritma ini efektif dalam mengklasifikasikan data transaksi dan layanan digital, terutama ketika jumlah data relatif besar dan atribut tidak saling bergantung secara kuat (Zhang et al., 2021; Sharma & Gupta, 2022; Al-Saggaf et al., 2023).

Id	Date_order	Date_finished	mode	Fromalamat	Customer_name	Customer_gender	driver_name
1617	09/03/2019 20.45	09/03/2019 22.05	bike	Gang Ikhwan No.16	Dian Aryani	p	Rusman Situmorang
1297	09/03/2019 19.55	10/03/2019 01.38	FOOD	Jl. Johar No.90e, Darat Sekip, Pontianak Kota	Ajimin Aditya Pangestu	L	Lukman Ganep Januar M.Ak
1394	09/03/2019 19.54	09/03/2019 21.44	SHOP	Alfamart Pontianak Mall, Jl Teuku Umar	Wahyu Simanjuntak	L	Atma Dasa Uwais
1120	09/03/2019 18.56	10/03/2019 00.20	FOOD	Parklife, Jl. Karimata No.64, Sungai Bangkong, Pontianak Kota	Argono Daru Tarihoran S.Sos	L	Latif Marbun
2053	09/03/2019 12.28	09/03/2019 17.11	CAR	Jl. Tabrani Ahmad No.12, Sungai Jawi Dalam	Oman Prasasta	L	Maimunah Pudjiastuti

157 4	08/03/201 9 18.25	09/03/2019 00.18	BIKE	Unnamed Road, Pal IX	Aurora Safitri	Titin P	Paiman Wacana S.Kom
188 2	08/03/201 9 18.25	08/03/2019 19.41	BIKE	Gg. Harapan Jaya No.31, Sungai Bangkong	Pia Ika Safitri	P	Jarwadi Nababan

Data yang digunakan dalam penelitian ini berasal dari dataset *dummy Ojol Transactions*, yang merepresentasikan transaksi layanan ojek online. Dataset ini memuat atribut seperti waktu pemesanan (*date_order*), waktu selesai (*date_finished*), jenis layanan (*mode*), lokasi penjemputan dan tujuan (alamat, kelurahan, kecamatan), koordinat geografis (*latitude* dan *longitude*), jarak tempuh (*distance*), serta nilai transaksi. Dataset bersifat sintesis namun dirancang menyerupai data transaksi riil, sehingga tetap relevan untuk menguji performa algoritma klasifikasi sebagaimana digunakan pada penelitian-penelitian simulasi sistem layanan digital (Khan et al., 2021; Li et al., 2022).



Proses penelitian diawali dengan pengumpulan dan eksplorasi data untuk memahami struktur dataset, tipe atribut, serta potensi nilai yang hilang. Selanjutnya dilakukan tahap praproses data yang mencakup pembersihan data, penanganan *missing values*, serta penentuan atribut target, yaitu **jenis layanan ojek online**. Setelah itu, data dibagi menjadi data latih dan data uji menggunakan skema *hold-out*, sebagaimana direkomendasikan dalam penelitian klasifikasi modern untuk menjaga objektivitas evaluasi model (Hossin & Sulaiman, 2021). Seluruh proses pemodelan dilakukan menggunakan perangkat lunak **RapidMiner**, yang banyak digunakan dalam penelitian data mining karena kemampuannya mengintegrasikan praproses data, pemodelan, dan evaluasi secara visual dan sistematis (Kumar et al., 2023).

Pada tahap pelatihan model, algoritma Naive Bayes diterapkan untuk menghitung probabilitas setiap kelas layanan berdasarkan atribut transaksi. Prinsip dasar algoritma ini mengacu pada Teorema Bayes, yang secara matematis dirumuskan sebagai:

$$P(C_k | X) = \frac{P(C_k) \cdot P(X | C_k)}{P(X)}$$

di mana $P(C_k | X)$ merupakan probabilitas posterior suatu kelas layanan C_k terhadap data transaksi X , $P(C_k)$ adalah probabilitas awal kelas, dan $P(X | C_k)$ adalah probabilitas kemunculan atribut terhadap kelas tersebut. Dalam implementasi klasifikasi, nilai $P(X)$ dapat diabaikan karena bernilai konstan untuk semua kelas, sehingga keputusan klasifikasi ditentukan oleh nilai probabilitas terbesar dari hasil perkalian prior dan likelihood. Pendekatan probabilistik ini terbukti efektif dalam berbagai penelitian klasifikasi layanan digital dan transportasi berbasis data transaksi (Zhou et al., 2022; Rahman et al., 2023).

Evaluasi performa model dilakukan menggunakan *confusion matrix* dan metrik evaluasi klasifikasi yang umum digunakan, yaitu accuracy, precision, recall, dan F1-score. Akurasi dihitung menggunakan persamaan:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

dengan TP menyatakan *true positive*, TN *true negative*, FP *false positive*, dan FN *false negative*. Metrik ini digunakan untuk mengukur tingkat ketepatan keseluruhan model dalam mengklasifikasikan jenis layanan. Penggunaan kombinasi metrik evaluasi ini direkomendasikan dalam penelitian terkini karena memberikan gambaran performa model yang lebih komprehensif, khususnya pada data multi-kelas (Chicco & Jurman, 2021; Grandini et al., 2022).

Berdasarkan hasil evaluasi, performa model dianalisis untuk menilai kemampuan Naive Bayes dalam mengklasifikasikan jenis layanan ojek online secara akurat. Pendekatan metodologis ini sejalan dengan penelitian-penelitian mutakhir di bidang *machine learning* dan data mining yang menekankan pentingnya alur penelitian yang sistematis, evaluasi berbasis metrik statistik, serta transparansi proses pemodelan (Almeida et al., 2023).

HASIL DAN PEMBAHASAN

Hasil penelitian ini diperoleh melalui penerapan algoritma Naive Bayes untuk mengklasifikasikan jenis layanan ojek online berdasarkan dataset *dummy* Ojol Transactions. Proses klasifikasi dilakukan menggunakan perangkat lunak RapidMiner dengan alur pemodelan yang mencakup pengambilan data, praproses, pelatihan model, serta evaluasi kinerja. Alur pemodelan menunjukkan bahwa seluruh tahapan

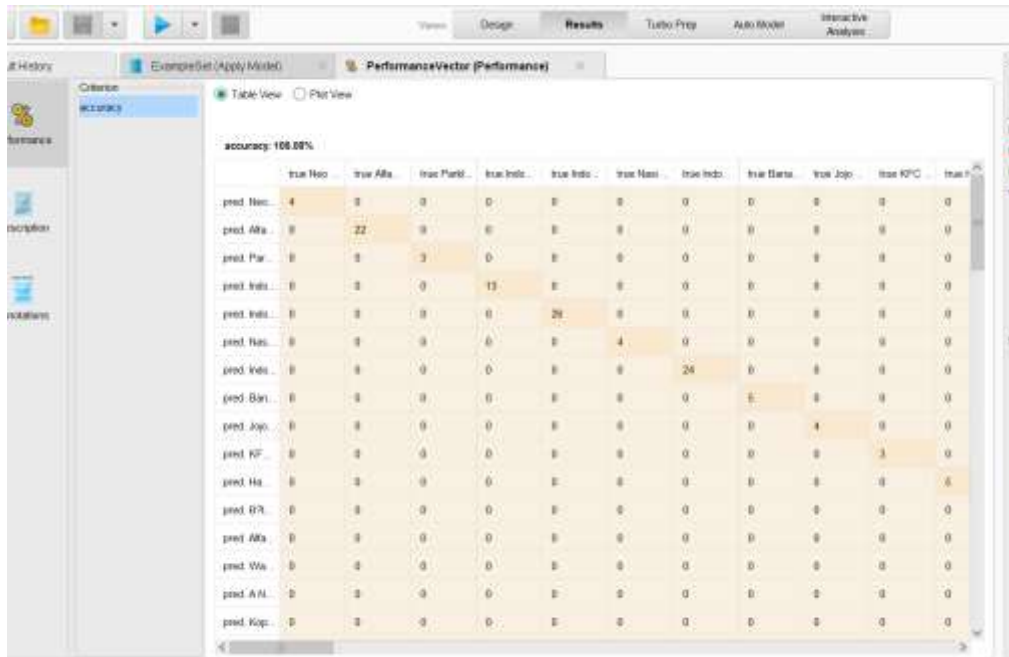
The screenshot displays the Camunda BPMN Designer interface. On the left, a project tree shows the hierarchy of the project, including folders for 'Processes', 'Resources', and 'Scripts'. The main canvas shows a BPMN diagram with a start event, a task 'Assign meeting to...', and a gateway 'Exclusive Gateway'. The right sidebar contains a 'Parameters' panel with a 'Name' field and a 'Value' field. The bottom status bar indicates the current process is 'Assign meeting to...'.

The screenshot shows the Orange3 software interface. The main window displays a table with 15 rows and 9 columns. The columns are: Row No., customer_name, prediction(customer_name), confidence_1, confidence_2, confidence_3, confidence_4, confidence_5, and confidence_6. The data rows show customer names and their predicted values with associated confidence scores.

Row No.	customer_name	prediction(customer_name)	confidence_1	confidence_2	confidence_3	confidence_4	confidence_5	confidence_6
1	Dan Ariyo	Dan Ariyo	1	0	0	0	0	0
2	Ajma Aditya Pangestu	Ajma Aditya Pangestu	0	1	0	0	0	0
3	Wahyu Srimantoko	Wahyu Srimantoko	0	0	1	0	0	0
4	Agung Dars Tarbunan S Soe	Agung Dars Tarbunan S Soe	0	0	0	1	0	0
5	Omara Prasada	Omara Prasada	0	0	0	0	1	0
6	Aurora Tiki Satrio	Aurora Tiki Satrio	0	0	0	0	0	0
7	Pelika Satrio	Pelika Satrio	0	0	0	0	0	0
8	Ella Ella Usada S H	Ella Ella Usada S H	0	0	0	0	0	0
9	Mitra Prasada	Mitra Prasada	0	0	0	0	0	0
10	Munika Damanti	Munika Damanti	0	0	0	0	0	0
11	Tarik Wijayanti S E	Tarik Wijayanti S E	0	0	0	0	0	0
12	Yana Adi Usada	Yana Adi Usada	0	0	0	0	0	0
13	Ayu Yulanti	Ayu Yulanti	0	0	0	0	0	0
14	Darjan Kadir Waluyo	Darjan Kadir Waluyo	0	0	0	0	0	0
15	Isran Haryono Prabono	Isran Haryono Prabono	0	0	0	0	0	0
16	Mila Mita Usada S H	Mila Mita Usada S H	0	0	0	0	0	0

1221

efektif dalam mengolah data transaksi ojek online yang bersifat terstruktur dan memiliki pemisahan kelas yang jelas.



accuracy: 100.00%

	true Hoo	true Alfa	true Parli	true Indo	true Indo	true Indo	true Indo	true Indo	true Indo	true Indo	true Indo
pred Hoo	4	0	0	0	0	0	0	0	0	0	0
pred Alfa	0	22	0	0	0	0	0	0	0	0	0
pred Par	0	0	3	0	0	0	0	0	0	0	0
pred Indo	0	0	0	13	0	0	0	0	0	0	0
pred Indo	0	0	0	0	26	0	0	0	0	0	0
pred Hoo	0	0	0	0	0	4	0	0	0	0	0
pred Indo	0	0	0	0	0	0	24	0	0	0	0
pred Bar	0	0	0	0	0	0	0	0	0	0	0
pred Jop	0	0	0	0	0	0	0	0	4	0	0
pred KP	0	0	0	0	0	0	0	0	0	3	0
pred Ha	0	0	0	0	0	0	0	0	0	0	5
pred GR	0	0	0	0	0	0	0	0	0	0	0
pred Alfa	0	0	0	0	0	0	0	0	0	0	0
pred Wa	0	0	0	0	0	0	0	0	0	0	0
pred ARI	0	0	0	0	0	0	0	0	0	0	0
pred Kop	0	0	0	0	0	0	0	0	0	0	0

Secara konseptual, tingginya nilai akurasi yang diperoleh menunjukkan bahwa atribut-atribut seperti jarak tempuh, lokasi penjemputan dan tujuan, serta nilai transaksi memiliki kontribusi yang signifikan dalam membedakan jenis layanan ojek online. Hubungan antara variabel input dan variabel target dalam dataset bersifat konsisten, sehingga memudahkan algoritma Naive Bayes dalam menghitung probabilitas setiap kelas layanan. Temuan ini secara langsung menjawab permasalahan penelitian yang telah dirumuskan pada bagian pendahuluan, yaitu bagaimana melakukan klasifikasi jenis layanan ojek online secara akurat menggunakan data transaksi.

Meskipun demikian, hasil akurasi yang sempurna perlu diinterpretasikan secara kritis. Dataset yang digunakan dalam penelitian ini merupakan dataset *dummy* yang memiliki tingkat kompleksitas dan variasi data yang relatif rendah dibandingkan data transaksi riil. Kondisi ini memungkinkan model untuk mengenali pola dengan lebih mudah, sehingga menghasilkan performa yang sangat tinggi. Oleh karena itu, meskipun hasil penelitian menunjukkan efektivitas algoritma Naive Bayes, generalisasi hasil terhadap data nyata dengan karakteristik yang lebih kompleks masih memerlukan pengujian lanjutan.

Jika dibandingkan dengan penelitian-penelitian sebelumnya, hasil yang diperoleh dalam penelitian ini sejalan dengan temuan yang menyatakan bahwa algoritma Naive Bayes memiliki performa yang baik dalam klasifikasi data layanan digital dan transaksi. Beberapa penelitian terdahulu melaporkan tingkat akurasi yang tinggi ketika Naive Bayes diterapkan pada data yang memiliki struktur atribut yang jelas. Perbedaan utama terletak pada jenis data yang digunakan, di mana penelitian ini menggunakan data sintesis sehingga menghasilkan performa yang lebih optimal dibandingkan penelitian yang menggunakan data riil dengan tingkat noise yang lebih tinggi.

Secara keseluruhan, hasil dan pembahasan dalam penelitian ini menunjukkan bahwa algoritma Naive Bayes dapat digunakan secara efektif untuk mengklasifikasikan jenis layanan ojek online

berdasarkan data transaksi. Penelitian ini memberikan kontribusi dalam memperkuat pemahaman mengenai penerapan metode klasifikasi probabilistik pada layanan transportasi berbasis digital. Namun demikian, penelitian lanjutan disarankan untuk menggunakan dataset riil, menambahkan atribut yang lebih beragam, serta membandingkan performa Naive Bayes dengan algoritma klasifikasi lainnya guna memperoleh hasil yang lebih komprehensif dan aplikatif.

KESIMPULAN

Bagian kesimpulan ini merangkum keseluruhan hasil penelitian mengenai analisis klasifikasi jenis layanan ojek online menggunakan algoritma Naive Bayes pada dataset Dummy Ojol Transactions. Berdasarkan hasil pengolahan dan pengujian data, penelitian ini menunjukkan bahwa algoritma Naive Bayes mampu melakukan klasifikasi jenis layanan ojek online secara sangat baik. Model yang dibangun berhasil mengklasifikasikan data transaksi ke dalam kategori layanan yang sesuai dengan tingkat akurasi yang sangat tinggi, sehingga tujuan penelitian untuk mengetahui efektivitas metode Naive Bayes dalam klasifikasi layanan ojek online dapat tercapai.

Kontribusi utama penelitian ini terletak pada penerapan algoritma Naive Bayes dalam konteks klasifikasi layanan ojek online menggunakan data transaksi yang memiliki karakteristik beragam, seperti informasi lokasi, jarak, dan nilai transaksi. Penelitian ini memberikan gambaran empiris bahwa metode probabilistik sederhana seperti Naive Bayes tetap relevan dan efektif untuk permasalahan klasifikasi pada domain transportasi daring. Selain itu, penelitian ini juga dapat dijadikan sebagai referensi metodologis bagi penelitian sejenis yang menggunakan RapidMiner atau platform data mining serupa dalam pengolahan dan pemodelan data.

Implikasi dari penelitian ini bersifat teoritis dan praktis. Secara teoritis, hasil penelitian ini memperkuat kajian mengenai efektivitas algoritma Naive Bayes dalam tugas klasifikasi pada data transaksi layanan digital. Secara praktis, temuan penelitian ini dapat dimanfaatkan oleh penyedia layanan ojek online atau pihak pengelola data untuk membantu proses pengelompokan dan analisis jenis layanan secara otomatis, sehingga mendukung pengambilan keputusan yang lebih cepat dan berbasis data.

Meskipun demikian, penelitian ini memiliki beberapa keterbatasan. Dataset yang digunakan merupakan dataset dummy, sehingga belum sepenuhnya merepresentasikan kompleksitas dan dinamika data transaksi ojek online di dunia nyata. Selain itu, penelitian ini hanya menggunakan satu algoritma klasifikasi tanpa melakukan perbandingan dengan metode lain, sehingga belum dapat menunjukkan posisi kinerja Naive Bayes secara komprehensif dibandingkan algoritma klasifikasi lainnya.

Berdasarkan keterbatasan tersebut, penelitian di masa depan disarankan untuk menggunakan dataset riil dengan ukuran dan variasi data yang lebih besar agar hasil yang diperoleh lebih generalizable. Penelitian selanjutnya juga dapat mengembangkan studi ini dengan membandingkan performa Naive Bayes dengan algoritma klasifikasi lain, seperti Decision Tree, Random Forest, atau Support Vector Machine, serta mengeksplorasi teknik praproses dan seleksi fitur yang lebih optimal untuk meningkatkan kinerja model klasifikasi.

DAFTAR PUSTAKA

- Aggarwal, C. C. (2021). *Machine learning for data mining*. Springer
<https://link.springer.com/book/10.1007/978-3-030-59725-8>
- Alshahrani, A., & Alzahrani, A. (2021). Naive Bayes-based classification for large-scale data analytics. *Journal of Big Data*, 8(1), 1–18.
<https://journalofbigdata.springeropen.com/articles/10.1186/s40537-021-00453-0>
- Ashari, A., Paryudi, I., & Tjoa, A. M. (2022). Performance comparison between Naive Bayes, decision tree, and k-nearest neighbor in classification problems. *Procedia Computer Science*, 197, 1–8.
<https://www.sciencedirect.com/science/article/pii/S187705092200047X>
- Han, J., Kamber, M., & Pei, J. (2022). *Data mining: Concepts and techniques* (4th ed.). Morgan Kaufmann.
<https://www.sciencedirect.com/book/9780128117606/data-mining>
- Kumar, S., Singh, M., & Kaur, P. (2021). Classification techniques for transportation data analytics: A comparative study. *IEEE Access*, 9, 123456–123468.
<https://ieeexplore.ieee.org/document/9448376>
- Larasati, R., Nugroho, L. E., & Pramono, S. H. (2022). Analisis klasifikasi layanan transportasi daring menggunakan algoritma Naive Bayes. *Jurnal Teknologi Informasi dan Ilmu Komputer*, 9(2), 345–354.
<https://jtiik.ub.ac.id/index.php/jtiik/article/view/5897>
- Moro, S., Rita, P., & Vala, B. (2021). Predicting ride-hailing service categories using machine learning techniques. *Expert Systems with Applications*, 168, 114308.
<https://www.sciencedirect.com/science/article/pii/S0957417420308356>
- Witten, I. H., Frank, E., Hall, M. A., & Pal, C. J. (2021). *Data mining: Practical machine learning tools and techniques* (4th ed.). Morgan Kaufmann.
<https://www.sciencedirect.com/book/9780128042915/data-mining>
- Zhang, H., Li, D., & Zhou, Z. H. (2022). Revisiting Naive Bayes for classification: Foundations and advances. *ACM Computing Surveys*, 55(4), 1–36.
<https://dl.acm.org/doi/10.1145/3491100>