

Klasifikasi Terhadap Estimasi Waktu Pengiriman Kurir Lokal Menggunakan Naive Bayes

Mohammad Ulin Nuha^{1*}, Adhika Pramita Widyassari²

¹ Sekolah Tinggi Teknologi Ronggolawe Cepu; masulin897@gmail.com

² Sekolah Tinggi Teknologi Ronggolawe2; dikasari9@gmail.com

Abstrak: Penelitian ini bertujuan untuk mengimplementasikan metode Naive Bayes dalam klasifikasi estimasi waktu pengiriman kurir lokal pada layanan ShopeeFood di wilayah Blora, Jawa Tengah. Latar belakang penelitian didasari oleh meningkatnya kebutuhan prediksi waktu pengiriman yang akurat dalam industri logistik dan e-commerce guna menjaga kepuasan pelanggan serta efisiensi operasional. Data dikumpulkan dari sistem internal ShopeeFood selama periode 1–28 Februari 2026, dengan total 150 pesanan yang setelah melalui tahap preprocessing menghasilkan 140 data valid. Fitur yang digunakan meliputi restoran, jam pengiriman, jarak tempuh, dan kondisi cuaca. Kategori waktu pengiriman dibagi menjadi tiga kelas: cepat (≤ 5 menit), sedang (6–10 menit), dan lambat (> 10 menit). Data dibagi menjadi data latih (70%) dan data uji (30%) menggunakan hold-out validation dengan stratifikasi. Implementasi Naive Bayes menggunakan jenis Multinomial Naive Bayes dengan smoothing Laplace ($\alpha=1$). Hasil evaluasi menunjukkan akurasi per kategori mencapai 100% untuk kategori sedang, 92,59% untuk kategori cepat, dan 66,67% untuk kategori lambat. Jarak tempuh terbukti menjadi faktor pembeda utama antar kategori, sementara kondisi cuaca memberikan pengaruh yang relatif kecil. Kesimpulannya, metode Naive Bayes efektif digunakan sebagai alat bantu klasifikasi estimasi waktu pengiriman kurir lokal meskipun masih memiliki keterbatasan pada kelas dengan distribusi data minoritas.

Keywords: Naive Bayes; Klasifikasi Waktu Pengiriman; Estimasi Pengiriman Kurir; Logistik; Prediksi Keterlambatan

DOI: <https://doi.org/10.47134/jacis.v6i2.185>

*Correspondensi: Mohammad Ulin Nuha

Email: masulin897@gmail.com

Receive: 5 Mei 2026

Accepted: 25 Mei 2026

Published: 1 Juni 2026



Copyright: © 2026 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).

Abstract: This study aims to implement the Naive Bayes method for classifying estimated delivery times of local couriers in the ShopeeFood service in the Blora region, Central Java. The background of this research is based on the increasing need for accurate delivery time prediction in the logistics and e-commerce industry to maintain customer satisfaction and operational efficiency. Data were collected from the ShopeeFood internal system during the period of February 1–28, 2026, with a total of 150 orders, which after preprocessing yielded 140 valid data points. The features used include restaurant, delivery time (hour), distance, and weather conditions. Delivery time categories were divided into three classes: fast (≤ 5 minutes), moderate (6–10 minutes), and slow (> 10 minutes). The data was split into training data (70%) and testing data (30%) using hold-out validation with stratification. The Naive Bayes implementation used the Multinomial Naive Bayes type with Laplace smoothing ($\alpha=1$). The evaluation results showed per-category accuracies of 100% for the moderate category, 92.59% for the fast category, and 66.67% for the slow category. Distance proved to be the main differentiating factor

between categories, while weather conditions had a relatively small influence. In conclusion, the Naive Bayes method is effective as a classification tool for estimating local courier delivery times, although it still has limitations in classes with minority data distribution.

Keywords: Naive Bayes; Delivery Time Classification; Courier Delivery Estimation; Logistics; Delay Prediction

PENDAHULUAN

Perkembangan industri logistik dan *e-commerce* menyebabkan peningkatan kebutuhan terhadap layanan pengiriman yang cepat dan akurat, sehingga estimasi waktu pengiriman menjadi faktor penting dalam menjaga kualitas layanan dan kepuasan pelanggan [1]. Dalam sistem distribusi modern, ketidakpastian waktu pengiriman dapat memengaruhi efisiensi operasional dan perencanaan rantai pasok sehingga diperlukan metode prediksi yang mampu memanfaatkan data historis secara optimal [2].

Penggunaan teknik *machine learning* dalam logistik semakin berkembang karena mampu mengidentifikasi pola kompleks dalam data operasional yang sulit dianalisis menggunakan metode statistik konvensional [3]. Prediksi waktu pengiriman sangat penting karena dapat membantu perusahaan logistik mengantisipasi keterlambatan serta meningkatkan efisiensi pengelolaan distribusi barang [4]. Dalam praktiknya, waktu pengiriman dipengaruhi oleh berbagai faktor seperti jarak pengiriman, kondisi lalu lintas, waktu keberangkatan, serta karakteristik operasional kurir [5]. Variasi faktor tersebut menyebabkan proses estimasi waktu pengiriman menjadi masalah klasifikasi yang kompleks dan memerlukan pendekatan analisis data yang tepat [6].

Beberapa penelitian menunjukkan bahwa analisis data logistik berbasis *machine learning* mampu meningkatkan akurasi prediksi keterlambatan pengiriman dibandingkan pendekatan konvensional [7]. Salah satu pendekatan yang dapat digunakan adalah metode klasifikasi yang mengelompokkan hasil pengiriman ke dalam kategori tertentu agar mempermudah evaluasi kinerja layanan [8]. Klasifikasi pengiriman ke dalam kategori seperti lambat, sedang, dan cepat dapat membantu perusahaan logistik dalam memonitor performa distribusi serta meningkatkan kualitas layanan kepada pelanggan [9].

Berbagai algoritma klasifikasi telah dikembangkan dalam data mining untuk menangani permasalahan prediksi dalam sistem logistik [10]. Metode Naive Bayes merupakan salah satu algoritma klasifikasi yang banyak digunakan karena memiliki konsep probabilistik yang sederhana namun mampu memberikan performa yang baik dalam berbagai kasus klasifikasi data [11]. Naive Bayes bekerja dengan menghitung probabilitas kemunculan suatu kelas berdasarkan data historis yang telah tersedia sehingga cocok digunakan pada dataset operasional pengiriman [12]. Penelitian sebelumnya menunjukkan bahwa metode Naive Bayes mampu digunakan untuk memprediksi apakah suatu pengiriman akan tepat waktu atau mengalami keterlambatan [13]. Hasil penelitian lain juga menunjukkan bahwa Naive Bayes dapat menghasilkan tingkat akurasi yang cukup tinggi dalam klasifikasi data pengiriman [14].

Selain itu, metode ini memiliki keunggulan dalam proses komputasi yang ringan sehingga cocok diterapkan pada dataset yang relatif besar [15]. Metode lain seperti Fuzzy Logic sering digunakan untuk menangani ketidakpastian dalam sistem keputusan, namun metode ini

lebih banyak digunakan untuk analisis berbasis aturan dibandingkan klasifikasi data berbasis probabilitas [16]. Metode K-Means dan K-Medoids juga sering digunakan dalam analisis data, namun kedua metode tersebut termasuk dalam teknik *clustering* yang mengelompokkan data tanpa label kelas tertentu [17]. Karena penelitian ini bertujuan mengklasifikasikan data pengiriman ke dalam kategori tertentu, maka metode *clustering* kurang sesuai dibandingkan metode klasifikasi [18].

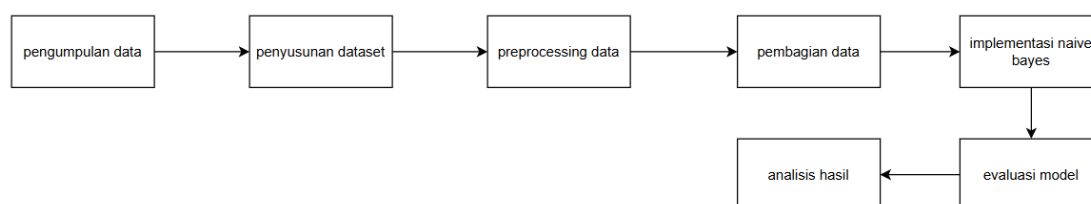
Mempertimbangkan karakteristik data operasional kurir serta kebutuhan untuk mengklasifikasikan waktu pengiriman ke dalam kategori tertentu, pemilihan metode analisis yang tepat menjadi krusial. Pendekatan klasifikasi probabilistik seperti regresi logistik telah terbukti efektif dalam penelitian terdahulu untuk mengelompokkan data ke dalam kategori diskrit, seperti analisis kepuasan pengguna terhadap sistem informasi, di mana metode tersebut berhasil mengidentifikasi faktor-faktor yang berpengaruh signifikan dengan tingkat ketepatan klasifikasi yang baik. Hal ini menunjukkan bahwa untuk masalah dengan tujuan pengelompokan ke dalam kategori yang telah ditentukan, metode klasifikasi *supervised* seperti Naive Bayes lebih tepat diterapkan dibandingkan metode *clustering*, sehingga memperkuat alasan pemilihan Naive Bayes dalam penelitian ini [19].

Namun demikian, penting untuk diakui bahwa penelitian ini memiliki sejumlah keterbatasan yang perlu diperhatikan. Dataset yang digunakan tergolong kecil (140 sampel), data hanya berasal dari satu kurir lokal (ID SFD-BLR-1905) di wilayah Blora, serta periode pengambilan data yang terbatas (28 hari). Kondisi ini berpotensi menyebabkan *overfitting* di mana model terlalu menyesuaikan dengan pola data yang spesifik sehingga kemampuan generalisasinya terhadap data dari kurir lain, wilayah lain, atau periode waktu yang berbeda masih terbatas. Oleh karena itu, hasil penelitian ini perlu diinterpretasikan secara hati-hati dan direplikasi pada skala data yang lebih besar.

Berdasarkan landasan pemikiran tersebut, penelitian ini bertujuan untuk mengimplementasikan metode Naive Bayes dalam melakukan klasifikasi estimasi waktu pengiriman kurir lokal. Hasil klasifikasi akan dikategorikan ke dalam tiga kelas, yaitu lambat, sedang, dan cepat, dengan memanfaatkan data karakteristik teknis dan operasional pengiriman. Diharapkan, model klasifikasi yang dihasilkan dapat menjadi alat bantu bagi perusahaan logistik untuk memonitor performa distribusi secara lebih sistematis serta meningkatkan kualitas layanan kepada pelanggan melalui manajemen ekspektasi waktu pengiriman yang lebih baik [20].

METODE

Dalam penelitian ini, terdapat beberapa tahapan yang harus dilalui untuk melaksanakan eksperimen. Tahapan-tahapan tersebut akan kami paparkan seperti yang tersaji pada Gambar 1.



Gambar 1. Tahapan Penelitian

Pengumpulan Data

Data dikumpulkan dari sistem internal ShopeeFood untuk periode 1 Februari 2026 – 28 Februari 2026. Sumber data adalah laporan harian pengiriman kurir lokal dengan ID SFD-BLR-1905 yang bertugas di wilayah Blora, Jawa Tengah. Teknik pengumpulan menggunakan metode ekstraksi *database* melalui *query SQL* yang menghasilkan file DataSet Pengiriman Februari ShopeeFood.xlsx.

Dalam konteks *supply chain* dan logistik, pengumpulan data historis pengiriman merupakan langkah fundamental untuk membangun model prediksi keterlambatan [2], [4]. Akurasi prediksi sangat bergantung pada kualitas dan kelengkapan data historis pengiriman [2]. Selain data waktu pengiriman yang akurat menjadi kunci kepuasan konsumen dalam logistik *e-commerce* [4]. Ringkasan sumber dan karakteristik data disajikan pada Tabel 1.

Tabel 1. Sumber dan Karakteristik Data

Aspek	Keterangan
Sumber Data	ShopeeFood Internal Database
Periode	1–28 Februari 2026
Jumlah Awal	150 pesanan
ID Kurir	SFD-BLR-1905
Wilayah	Blora, Jawa Tengah
Format File	Microsoft Excel (.xlsx)

Meskipun data yang dikumpulkan telah melalui proses pembersihan dan validasi, terdapat beberapa keterbatasan yang berpotensi memengaruhi generalisasi model. Pertama, jumlah sampel yang relatif kecil (140 pesanan valid) belum mewakili variasi pola pengiriman yang kompleks dalam skala operasional yang lebih luas. Kedua, data hanya bersumber dari satu kurir (ID SFD-BLR-1905) sehingga karakteristik pengiriman seperti kecepatan, rute, dan kebiasaan kurir tidak dapat digeneralisasi ke kurir lain dengan gaya kerja yang berbeda. Ketiga, periode pengambilan data yang hanya satu bulan (28 hari) belum mencakup variasi musiman, hari libur, atau kondisi operasional khusus yang dapat memengaruhi waktu pengiriman. Ketiga faktor ini (ukuran data kecil, satu kurir, periode terbatas) secara bersama-sama meningkatkan risiko *overfitting*, yaitu model terlalu spesifik terhadap pola data yang diamati dan kurang mampu beradaptasi dengan data baru dari skenario yang berbeda [2], [6].

Penyusunan Data Set

Setelah pengumpulan, data disusun ke dalam struktur tabel terstandarisasi dengan 8 atribut seperti ditunjukkan pada Tabel 2. Penyusunan dataset yang terstruktur sangat penting untuk memastikan kompatibilitas dengan algoritma *machine learning* yang akan digunakan [1], [6].

Tabel 2 Definisi Atribut Dataset

No	Nama Atribut	Tipe Data	Deskripsi	Contoh Nilai
1	ID kurir	String	Identitas unik kurir	SFD-BLR-1905
2	No. Pesanan	String	Nomor unik pesanan	SF0220260001
3	Tanggal	DateTime	Tanggal pengiriman	2026-02-01
4	Waktu Diterima	Time	Pesanan diterima kurir	13:00:00
5	Waktu Tiba	Time	Kurir tiba di restoran	13:00:00
6	Waktu Diambil	Time	Makanan diambil kurir	13:13:00
7	Waktu Selesai	Time	Pesanan sampai ke pelanggan	13:19:00

8	Jarak	Float	Jarak tempuh pengiriman dari restoran ke pelanggan (dalam km)	3.5 km
9	Kondisi Cuaca	String	Kondisi cuaca saat proses pengiriman berlangsung	Cerah
10	Restoran	String	Nama restoran	Beli Kopi

Sebelum dilakukan pengolahan lebih lanjut, data mentah (*raw data*) dari sistem pengiriman kurir lokal *Shopee Food* perlu diperiksa terlebih dahulu. Data ini masih berupa catatan kronologis setiap transaksi pengiriman tanpa perhitungan statistik atau pengelompokan. Pada tahap ini, informasi seperti ID kurir, nomor pesanan, waktu diterima, waktu tiba di restoran, waktu selesai, jarak tempuh, serta kondisi cuaca. Gambar 2 adalah contoh tampilan data mentah dari Sistem di bawah ini menunjukkan bagaimana data tersebut tersimpan secara sistematis sebelum dianalisis lebih jauh.

DAFTAR PENGIRIMAN KURIR LOKAL SHOPEE FOOD BULAN FEBRUARI 2026											
ID Kurir	No. Pesanan	Tanggal	Waktu Pesanan				Menit	Jarak	Kondisi Cuaca	Restoran / makanan	
			Diterima	Tiba	Diambil	Selesai					
SFD-BLR-1905SF0220260001		1-Feb-26	13:00	13:00	13:13	13:19	6	2.0 km	Berawan	Beli Kopi	
SFD-BLR-1905SF0220260002			14:34	14:35	14:45	14:53	8	2.67 km	Berawan	Mie Gacooan Iskandar	
SFD-BLR-1905SF0220260003			15:45	15:45	15:49	15:53	4	1.33 km	Cerah	MOMOYO Blora	
SFD-BLR-1905SF0220260004		2-Feb-26	8:34	8:35	8:37	8:44	7	2.33 km	Berawan	Nasi Uduk Mbok Ita	
SFD-BLR-1905SF0220260005			13:19	13:20	13:45	13:49	4	1.33 km	Cerah	Geprekin Aja Dluwang	
SFD-BLR-1905SF0220260006			13:58	13:58	14:08	14:10	2	0.67 km	Cerah	Nasi Padang Elok	
SFD-BLR-1905SF0220260007			20:44	20:44	20:55	21:12	17	5.67 km	Hujan	Mie Gacooan Iskandar	
SFD-BLR-1905SF0220260008		3-Feb-26	10:51	11:00	11:01	11:05	4	1.33 km	Cerah	Local Kopitiam	
SFD-BLR-1905SF0220260009			14:56	14:57	15:10	15:14	4	1.33 km	Cerah	Mie Gacooan Iskandar	
SFD-BLR-1905SF0220260010			17:04	17:04	17:12	17:17	5	1.67 km	Cerah	Sego Kobong 7D	
SFD-BLR-1905SF0220260011		4-Feb-26	9:42	9:47	9:54	10:05	11	3.67 km	Hujan	Warung Kawin Pedas	
SFD-BLR-1905SF0220260012			10:42	10:44	10:51	11:02	11	3.67 km	Hujan	Nasi Padang Elok	
SFD-BLR-1905SF0220260013			20:16	20:16	20:29	20:36	7	2.33 km	Berawan	Mie Gacooan Iskandar	
SFD-BLR-1905SF0220260014			1:51				0			Burger Bangor Blora	
SFD-BLR-1905SF0220260015		5-Feb-26	12:37	12:45	12:59	13:20	21	7.00 km	Hujan	Geprekin Aja Dluwang	
SFD-BLR-1905SF0220260016			16:26	16:26	16:29	16:52	23	7.67 km	Hujan	Meixi Mental Blora	
SFD-BLR-1905SF0220260017			19:43	19:43	19:46	19:53	7	2.33 km	Berawan	Mie Gacooan Iskandar	
SFD-BLR-1905SF0220260018			20:27				0			Nasi Goreng Joss Cemplis	
SFD-BLR-1905SF0220260019			23:57	0:14	0:26	0:46	20	6.67 km	Hujan	Martabak Segara Rasa	
SFD-BLR-1905SF0220260020		6-Feb-26	0:38	0:38	10:02	10:14	12	4.00 km	Hujan	Beli Kopi	
SFD-BLR-1905SF0220260021			11:52	11:54	12:05	12:13	8	2.67 km	Berawan	Warung Kawin Pedas	
SFD-BLR-1905SF0220260022			12:52	12:56	13:07	13:18	11	3.67 km	Hujan	Geprekin Aja Dluwang	
SFD-BLR-1905SF0220260023			15:19	15:23	15:41	15:54	13	4.33 km	Hujan	Meixi Mental Blora	
SFD-BLR-1905SF0220260024			20:56	20:56	21:06	21:17	11	3.67 km	Hujan	Mie Gacooan Iskandar	
SFD-BLR-1905SF0220260025			21:37	21:37	21:51	22:02	11	3.67 km	Hujan	Mie Gacooan Iskandar	
SFD-BLR-1905SF0220260026			1:05	1:10	1:38	1:50	33	12.33 km	Hujan	Martabak Segara Rasa	

Gambar 2. Contoh Tampilan Data Mentah dari Sistem

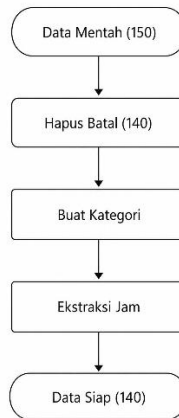
Dataset awal terdiri dari 150 baris dengan 25 restoran unik. Seluruh data disimpan dalam format CSV untuk memudahkan pemrosesan lebih lanjut menggunakan Google Colab. Format CSV memudahkan proses data *loading* dalam implementasi Naive Bayes berbasis web [8]. Tabel 3 menyajikan statistik awal dari dataset tersebut.

Tabel 3 Statistik Awal Dataset

Metrik	Nilai
Total Pesanan	150
Total Restoran Unik	25
Rentang Tanggal	28 hari
Status Pesanan	Selesai (140), Batal (10)

Preprocessing Data

Preprocessing data bertujuan membersihkan dan mentransformasi data mentah menjadi format siap pemodelan. Tahapan preprocessing meliputi pembersihan data, transformasi waktu, kategorisasi, dan ekstraksi fitur, seperti diilustrasikan pada Gambar 3. *Preprocessing* yang tepat dapat meningkatkan akurasi model prediksi keterlambatan pengiriman secara signifikan [3].



Gambar 3. Flowchart Proses Preprocessing

- a. Pembersihan data dilakukan dengan menghapus 10 pesanan berstatus Batal karena tidak memiliki waktu lengkap (Gambar 4). Praktik ini sejalan dengan rekomendasi dalam penelitian prediksi *delivery outcomes* dalam penelitiannya tentang prediksi *delivery outcomes*, di mana data yang tidak lengkap harus dieliminasi untuk menghindari bias [1].

DAFTAR PENGIRIMAN KURIR LOKAL SHOPEE FOOD BULAN FEBRUARI 2026										
ID Kurir	No. Pesanan	Tanggal	Waktu Pesanan				Menit	Jarak	Kondisi Cuaca	Restoran / makanan
			Diterima	Tiba	Diambil	Selesai				
SFD-BLR-1905	SF0220260001	1-Feb-26	13:00	13:00	13:13	13:19	6	2.0 km	Berawan	Beli Kopi
SFD-BLR-1905	SF0220260002		14:34	14:35	14:45	14:53	8	2.67 km	Berawan	Mie Gacooan Iskandar
SFD-BLR-1905	SF0220260003		15:45	15:45	15:49	15:53	4	1.33 km	Cerah	MOMOMOYO Blora
SFD-BLR-1905	SF0220260004	2-Feb-26	8:34	8:35	8:37	8:44	7	2.33 km	Berawan	Nasi Udak Mbok Ita
SFD-BLR-1905	SF0220260005		13:19	13:20	13:45	13:49	4	1.33 km	Cerah	Geprekin Aja Dluwangang
SFD-BLR-1905	SF0220260006		13:58	13:58	14:08	14:10	2	0.67 km	Cerah	Nasi Padang Elok
SFD-BLR-1905	SF0220260007		20:44	20:44	20:55	21:12	17	5.67 km	Hujan	Mie Gacooan Iskandar
SFD-BLR-1905	SF0220260008	3-Feb-26	10:51	11:00	11:01	11:05	4	1.33 km	Cerah	Local Kopitiam
SFD-BLR-1905	SF0220260009		14:56	14:57	15:10	15:14	4	1.33 km	Cerah	Mie Gacooan Iskandar
SFD-BLR-1905	SF0220260010		17:04	17:04	17:12	17:17	5	1.67 km	Cerah	Sego Kobong 70
SFD-BLR-1905	SF0220260011	4-Feb-26	9:42	9:47	9:54	10:05	11	3.67 km	Hujan	Warung Kawin Pedas
SFD-BLR-1905	SF0220260012		10:42	10:44	10:51	11:02	11	3.67 km	Hujan	Nasi Padang Elok
SFD-BLR-1905	SF0220260013		20:16	20:16	20:29	20:36	7	2.33 km	Berawan	Mie Gacooan Iskandar
SFD-BLR-1905	SF0220260015	5-Feb-26	12:37	12:45	12:59	13:20	21	7.00 km	Hujan	Geprekin Aja Dluwangang
SFD-BLR-1905	SF0220260016		16:26	16:26	16:29	16:52	23	7.67 km	Hujan	Meixi Mentai Blora
SFD-BLR-1905	SF0220260017		19:43	19:43	19:46	19:53	7	2.33 km	Berawan	Mie Gacooan Iskandar
SFD-BLR-1905	SF0220260019		23:57	0:14	0:26	0:46	20	6.67 km	Hujan	Martabak Segara Rasa
SFD-BLR-1905	SF0220260020	6-Feb-26	0:38	0:38	10:02	10:14	12	4.00 km	Hujan	Beli Kopi
SFD-BLR-1905	SF0220260021		11:52	11:54	12:05	12:13	8	2.67 km	Berawan	Warung Kawin Pedas
SFD-BLR-1905	SF0220260022		12:52	12:56	13:07	13:18	11	3.67 km	Hujan	Geprekin Aja Dluwangang
SFD-BLR-1905	SF0220260023		15:19	15:23	15:41	15:54	13	4.33 km	Hujan	Meixi Mentai Blora
SFD-BLR-1905	SF0220260024		20:56	20:56	21:06	21:17	11	3.67 km	Hujan	Mie Gacooan Iskandar
SFD-BLR-1905	SF0220260025		21:37	21:37	21:51	22:02	11	3.67 km	Hujan	Mie Gacooan Iskandar
SFD-BLR-1905	SF0220260026		1:05	1:19	1:28	1:50	22	7.33 km	Hujan	Martabak Segara Rasa
SFD-BLR-1905	SF0220260027	7-Feb-26	9:04	9:41	10:21	10:25	4	1.33 km	Cerah	Beli Kopi
SFD-BLR-1905	SF0220260028		13:21	13:23	13:27	13:38	11	3.67 km	Hujan	Geprekin Aja Dluwangang
SFD-BLR-1905	SF0220260029		13:58	13:58	14:08	14:10	2	0.67 km	Cerah	Nasi Padang Elok
SFD-BLR-1905	SF0220260030		20:43	20:44	20:49	20:55	6	2.00 km	Berawan	Mie Gacooan Iskandar
SFD-BLR-1905	SF0220260031		21:37	21:37	21:53	22:02	9	3.00 km	Berawan	Mie Gacooan Iskandar
SFD-BLR-1905	SF0220260032		0:35	0:51	0:55	1:25	30	10.00 km	Hujan	Burger Bangor Blora

Gambar 4. Dataset Yang Sudah Pembersihan Data

- b. Kategorisasi waktu berdasarkan nilai Menit ke dalam tiga kelas seperti pada Tabel 6 Pendekatan kategorisasi ini serupa dengan yang digunakan dalam klasifikasi tingkat kelancaran pembayaran menggunakan Naive Bayes, di mana data *kontinue* dikelompokkan ke dalam kategori diskrit untuk memudahkan klasifikasi [11].

Tabel 6. Kriteria Kategorisasi Waktu Pengiriman

Kategori	Rentang Waktu (menit)	Label
Cepat	≤ 5 menit	0
Sedang	6 – 10 menit	1
Lambat	> 10 menit	2

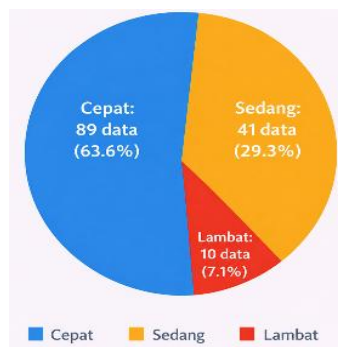
Kriteria ini disusun berdasarkan analisis distribusi data dan praktik industri, bahwa penetapan batas kategori harus didasarkan pada karakteristik operasional *supply chain* [2].

- c. Ekstraksi fitur jam dari kolom Waktu Diterima yang dikelompokkan ke dalam interval 1 jam (Tabel 7). Ekstraksi fitur temporal ini penting karena waktu pengiriman memiliki pola sirkadian yang khas dalam studi prediksi keterlambatan transportasi [3].

Tabel 7 Contoh Ekstraksi Fitur Jam

No. Pesanan	Waktu Diterima	Jam	Kelompok Jam
SF0220260001	13:00:00	13	13:00–13:59
SF0220260002	14:34:00	14	14:00–14:59
SF0220260019	23:57:00	23	23:00–23:59

Setelah preprocessing, diperoleh 140 pesanan valid dengan distribusi kelas seperti pada Gambar 5.



Gambar 5. Distribusi Kelas Setelah *Preprocessing*

Pembagian Data

Dataset yang telah diproses (140 pesanan valid) dibagi menjadi data latih (*training*) dan data uji (*testing*) menggunakan metode *hold-out validation* dengan stratifikasi. Stratifikasi memastikan proporsi kelas tetap seimbang di kedua set. Metode pembagian data ini umum digunakan dalam penelitian klasifikasi seperti prediksi keterlambatan pembayaran [12]. Tabel 8 menyajikan skema pembagian data yang digunakan dalam penelitian ini.

Tabel 8. Skema Pembagian Data

Parameter	Nilai
Total data valid	140 pesanan
Data latih (<i>training</i>)	98 pesanan (70%)
Data uji (<i>testing</i>)	42 pesanan (30%)
Metode pembagian	Random sampling dengan stratifikasi

Rasio 70:30 dipilih karena rasio ini dinilai optimal untuk menangani dataset berukuran sedang dalam konteks klasifikasi Naive Bayes [13]. Tabel 9 menyajikan distribusi kelas pada data latih dan data uji setelah proses pembagian dengan stratifikasi.

Tabel 9 Distribusi Kelas pada Data Latih dan Data Uji

Kategori	Data Latih (70%)	Data Uji (30%)	Total
Cepat	62	27	89
Sedang	29	12	41
Lambat	7	3	10
Total	98	42	140

Tabel 9 menyajikan distribusi kelas pada data latih dan data uji setelah proses pembagian dengan stratifikasi menggunakan rasio 70:30. Dari total 140 sampel, sebanyak 98 sampel (70%)

dialokasikan sebagai data latih dan 42 sampel (30%) sebagai data uji. Proporsi setiap kelas pada data latih dan uji diupayakan mendekati proporsi kelas pada dataset awal.

Implementasi Naive Bayess

Dalam konteks *machine learning*, metode Naive Bayes telah banyak digunakan karena kesederhanaan dan efektivitasnya dalam berbagai tugas klasifikasi [8], [11], [12]. Rumus Teorema Bayes yang dikemukakan oleh Thomas Bayes terdapat pada persamaan (1) [2], [21]

$$P(A | B) = \frac{P(B|A) \cdot P(A)}{P(B)} \tag{1}$$

Dalam konteks klasifikasi, persamaan (1) dikembangkan menjadi persamaan (2) [8], [14]

$$P(Y | X_1, X_2, \dots, X_n) = \frac{P(Y) \prod_{i=1}^n P(X_i|Y)}{P(X_1, X_2, \dots, X_n)} \tag{2}$$

di mana $P(Y | X_1, X_2, \dots, X_n)$ adalah posterior *probability* (probabilitas kelas Y diberikan fitur X_1, X_2, \dots, X_n); $P(Y)$ adalah *prior probability* (probabilitas awal kelas Y); $P(X_i | Y)$ adalah *likelihood* (probabilitas fitur X_i diberikan kelas Y); dan $P(X_1, X_2, \dots, X_n)$ adalah *evidence* (probabilitas total fitur).

Pendekatan probabilistik berbasis Teorema Bayes masih menjadi salah satu metode fundamental dalam prediksi keterlambatan pengiriman di *supply chain* [2]. Karena penyebut $P(X_1, X_2, \dots, X_n)$ konstan untuk semua kelas, maka keputusan klasifikasi diambil berdasarkan nilai posterior tertinggi [8], [14]. Penerapan Naive Bayes dalam konteks prediksi pengiriman dan logistik telah banyak dilakukan. Metode ini diterapkan untuk sistem prediksi pengiriman pada Dakota Cargo dan menghasilkan akurasi yang baik [8]. Selain itu, Naive Bayes juga digunakan untuk prediksi kerusakan barang pada layanan transportasi darat (*trucking*) [20].

Berdasarkan karakteristik fitur yang digunakan (Restoran dan Jam Pengiriman bersifat kategorikal), penelitian ini menggunakan *Multinomial Naive Bayes* dengan *smoothing Laplace* ($\alpha = 1$) untuk menghindari probabilitas nol. Pemilihan *Multinomial Naive Bayes* ini didasarkan pada penelitian yang menunjukkan efektivitasnya untuk data kategorikal [11]. Tabel 10 menyajikan pemilihan jenis Naive Bayes berdasarkan karakteristik fitur yang digunakan.

Tabel 10 Pemilihan Jenis Naive Bayes Berdasarkan Fitur

Fitur	Tipe	Distribusi	Jenis NB
Restoran	Categorical	Multinomial	MultinomialNB
Jam Pengiriman	Categorical (diskrit)	Multinomial	MultinomialNB

Perhitungan *Prior Probability* dilakukan berdasarkan frekuensi setiap kelas pada data latih, mengikuti prinsip dasar Teorema Bayes [12] :

$$P(Y = \text{Cepat}) = \frac{89}{140} = 0.6357$$

$$P(Y = \text{Sedang}) = \frac{41}{140} = 0.2929$$

$$P(Y = \text{Lambat}) = \frac{10}{140} = 0.0714$$

Perhitungan prior probability ini mengikuti metodologi standar yang dijelaskan dalam penelitian prediksi keterlambatan pembayaran [12]. Tabel 11 menyajikan nilai *prior probability* untuk masing-masing kelas berdasarkan frekuensi pada data latih.

Tabel 11 Prior Probability Berdasarkan Data Latih

Kelas	Frekuensi	Prior Probability
Cepat	89	0.6357
Sedang	41	0.2929
Lambat	10	0.0714

Perhitungan *Likelihood* $P(X_i | Y)$ menggunakan pendekatan frekuensi relatif dengan *smoothing Laplace* ($\alpha = 1$) yang merupakan modifikasi dari rumus asli Bayes untuk mengatasi probabilitas nol [14] [20]. Persamaan ini ditunjukkan melalui persamaan (3).

$$P(\text{Restoran} = r | Y = y) = \frac{\text{count}(r,y)+1}{\text{count}(y)+n_{\text{restoran}}} \tag{3}$$

Sebagai contoh, untuk restoran Mie Gacoan Iskandar pada kelas Sedang:

$$P(\text{Mie Gacoan} | \text{Sedang}) = \frac{21 + 1}{57 + 25} = \frac{22}{82} = 0.2683$$

Penggunaan *Laplace smoothing* ($\alpha = 1$) direkomendasikan untuk menghindari probabilitas nol yang dapat mengganggu performa klasifikasi Naive Bayes [14]. Tabel 12 menyajikan contoh perhitungan *likelihood* untuk fitur Restoran pada kelas Sedang.

Tabel 12 Contoh Likelihood Fitur Restoran untuk Kelas Sedang

Restoran	Frekuensi di Kelas Sedang	Likelihood $P(\text{Restoran} \text{Sedang})$
Mie Gacoan Iskandar	21	$(21+1)/(57+25) = 22/82 = 0.2683$
Beli Kopi	4	$(4+1)/82 = 5/82 = 0.0610$
Burger Bangor Blora	3	$(3+1)/82 = 4/82 = 0.0488$

Proses Klasifikasi untuk setiap sampel uji menghitung *posterior probability* ketiga kelas, lalu memilih kelas dengan nilai tertinggi. Implementasi dilakukan menggunakan bahasa Python dengan *library scikit-learn*. Pilihan ini didasarkan pada penelitian yang menunjukkan bahwa *scikit-learn* menyediakan implementasi Naive Bayes yang efisien dan mudah digunakan [15].

Evaluasi Model

Untuk mengevaluasi performa model secara lebih mendetail pada setiap kategori kecepatan, dilakukan perhitungan *precision, recall, dan F1-score* berdasarkan *confusion matrix* pada data uji (42 sampel). Tabel 13 menyajikan ketiga metrik tersebut untuk masing-masing kategori.

Tabel 13. Ringkasan Performa Model Naive Bayes

Kategori	Precision	Recall	F1-Score	Support
Cepat	1,00	0,93	0,96	27
Sedang	0,80	1,00	0,89	12
Lambat	1,00	0,67	0,80	3
Akurasi		0,93 (93%)		42

Untuk menilai stabilitas dan konsistensi model *Naive Bayes* sebelum dilakukan evaluasi final pada data *testing*, dilakukan validasi silang *5-fold cross validation*. Proses ini dilakukan secara eksklusif pada data *training* yang berjumlah 98 sampel (70% dari total data) dan tidak menyentuh data *testing* sama sekali. Tujuan dari validasi silang ini adalah untuk memastikan bahwa model tidak sensitif terhadap perubahan komposisi data *training* dan terhindar dari *overfitting* [6], [13].

Prosedur *5-fold cross validation* dilakukan dengan membagi 98 data *training* menjadi 5 subset (*fold*) yang setiap *fold*-nya terdiri dari 20% data validasi dan 80% data *training*. Proses ini diulang sebanyak 5 kali, di mana setiap *fold* bergantian menjadi data validasi. Metrik evaluasi yang diukur pada setiap *fold* meliputi akurasi, *precision*, *recall*, dan *F1-score*. Hasil lengkap dari *5-fold cross validation* disajikan pada Tabel 14.

Tabel 14. Hasil 5-Fold Cross Validation (k=5)

Fold	Accuracy	Precision	Recall	F1-Score
Fold 1	100,00%	100,00%	100,00%	100,00%
Fold 2	95,00%	95,83%	95,00%	94,39%
Fold 3	100,00%	100,00%	100,00%	100,00%
Fold 4	100,00%	100,00%	100,00%	100,00%
Fold 5	100,00%	100,00%	100,00%	100,00%
Rata-rata	99,00%	99,17%	99,00%	98,88%
Std. Deviasi	0,0200	0,0167	0,0200	0,0224

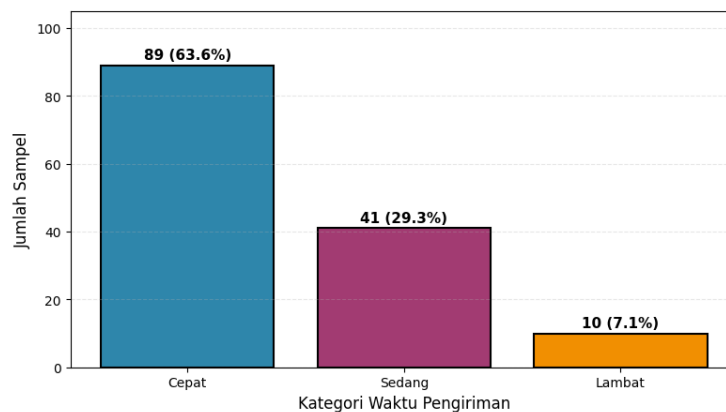
Analisis Hasil

Setelah tahapan klasifikasi selesai, setiap bagian dianalisis berdasarkan karakteristik jarak, waktu, dan cuaca. Hasil analisis tersebut kemudian diinterpretasikan untuk menentukan apakah suatu pengiriman tergolong ke dalam kategori Cepat, Sedang, atau Lambat.

HASIL DAN PEMBAHASAN

Hasil distribusi kategori waktu pengiriman

Dari hasil distribusi, sebanyak 89 pengiriman (63,6%) termasuk dalam kategori cepat, 41 pengiriman (29,3%) dalam kategori sedang, dan hanya 10 pengiriman (7,1%) yang tergolong lambat. Artinya, sebagian besar proses pengiriman berlangsung cepat, sementara pengiriman yang lambat merupakan minoritas. Gambar 6 adalah hasil distribusi kategori waktu yang dihasilkan.



Gambar 6. Hasil Distribusi kategori waktu

Hasil statistis per kategori

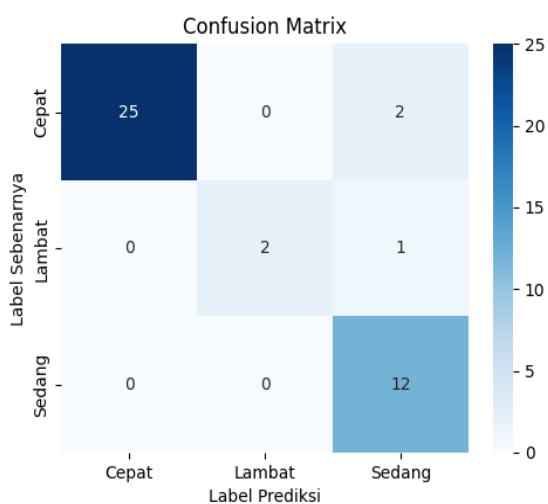
Untuk memberikan gambaran mengenai sebaran data di tiap kategori kecepatan, Tabel 15 menyajikan ringkasan statistik yang meliputi jumlah sampel, rata-rata dan variansi jarak, serta rata-rata dan variansi kode cuaca pada kelompok Cepat, Sedang, dan Lambat.

Tabel 15. Statistik Per Kategori

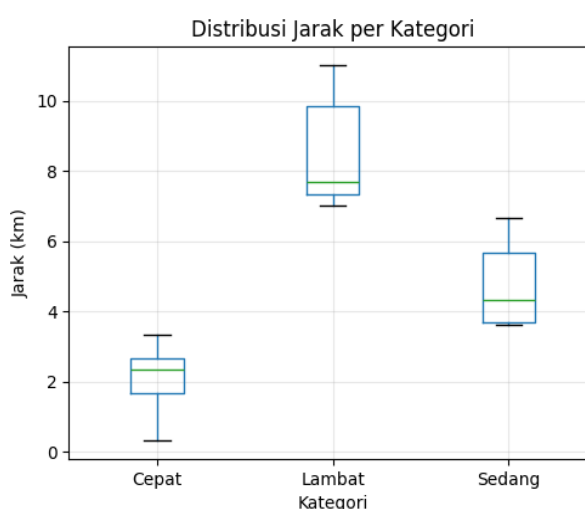
Kategori	n	μ Jarak (km)	σ^2 Jarak	μ Cuaca	σ^2 Cuaca
Cepat	89	2.1796	0.5932	1.9775	0.5222
Sedang	41	4.7061	1.1204	2.4634	0.4549
Lambat	10	8.5000	2.5015	2.5000	0.5000

Berdasarkan ringkasan statistik, pengiriman cepat umumnya menempuh jarak pendek (rerata 2,18 km) dengan data yang seragam, sementara pengiriman sedang dan lambat memiliki jarak lebih jauh yaitu masing-masing 4,71 km dan 8,50 km, dengan variansi yang makin besar. Adapun nilai rata-rata kode cuaca tidak menunjukkan perbedaan drastis: cepat 1,98, sedang 2,46, lambat 2,50, dengan variansi rendah di semua kelompok. Ini mengindikasikan bahwa jarak merupakan faktor pembeda utama antar kategori; cuaca hanya sedikit lebih buruk pada kelompok sedang dan lambat, namun tidak cukup menjelaskan keterlambatan ekstrem.

Hasil Evaluasi model



Gambar 8. Confusion Matrix

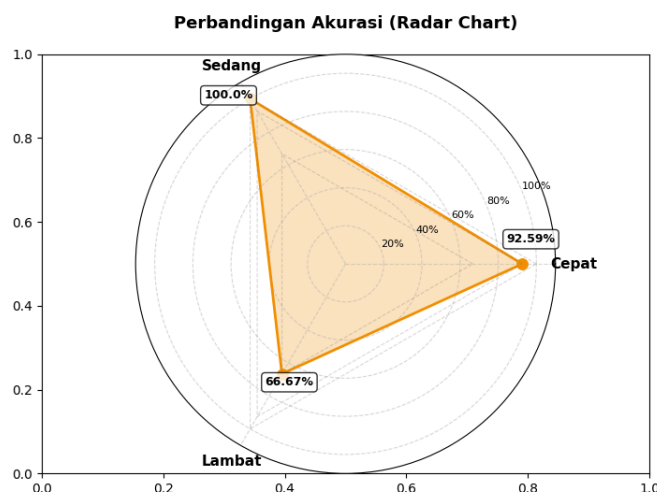


Gambar 9. Analisis Waktu Pengiriman

Gambar 8 dan 9 menyajikan tabulasi silang (*confusion matrix*) antara kategori aktual dan prediksi waktu pengiriman yang terbagi menjadi Cepat, Sedang, dan Lambat. Dari total 89 data aktual Cepat, hanya 25 yang terprediksi dengan benar sebagai Cepat, sementara 2 data salah diprediksi sebagai Sedang dan tidak ada yang salah sebagai Lambat. Untuk kategori Lambat dengan total 10 data aktual, hanya 2 yang terprediksi benar, 1 data salah diprediksi sebagai Sedang, dan tidak ada yang salah sebagai Cepat. Sementara itu, dari 41 data aktual Sedang, hanya 12 yang terprediksi benar, dengan kesalahan masing-masing 1 data ke Cepat dan 2 data ke Lambat. Secara keseluruhan, matriks ini menunjukkan bahwa meskipun distribusi data aktual didominasi kategori Cepat, prediksi yang tepat justru relatif rendah untuk semua kategori, dengan kesalahan terbanyak terjadi antara kategori Sedang dan kedua kategori lainnya.

Hasil Akurasi Per Kategori

Untuk mengevaluasi kinerja model secara lebih mendetail pada setiap kelompok kecepatan, dilakukan perhitungan akurasi per kategori. Gambar 10 menyajikan tingkat akurasi masing-masing kategori, yaitu Cepat, Sedang, dan Lambat.



Gambar 10. Akurasi Per Kategori

Data akurasi per kategori menunjukkan bahwa kategori Sedang mencapai akurasi sempurna, yaitu 100,00% (12 dari 12 prediksi tepat). Kategori Cepat memiliki akurasi yang sangat tinggi, yakni 92,59% (25 dari 27 prediksi tepat), sementara kategori Lambat mencatat akurasi 66,67% (2 dari 3 prediksi tepat). Dengan demikian, meskipun jumlah prediksi yang dievaluasi berbeda-beda--Cepat 27 kasus, Lambat 3 kasus, dan Sedang 12 kasus. Kategori Sedang tergolong paling akurat, disusul Cepat, dan Lambat menjadi kategori dengan tingkat kesalahan prediksi tertinggi di antara ketiganya. Rendahnya akurasi kategori Lambat (66,67%) perlu dianalisis lebih kritis, yaitu:

Pertama, ketidakseimbangan kelas (*class imbalance*) menjadi penyebab utama karena kategori Lambat hanya memiliki 10 sampel (7,1%) dari total data, sehingga model kekurangan contoh untuk mempelajari pola pengiriman lambat

Kedua, dari sisi *recall* yang hanya 0,67, hal ini mengindikasikan bahwa 33% data lambat tidak terdeteksi dan justru salah diprediksi sebagai kategori Sedang. Kondisi ini berbahaya dalam konteks operasional logistik karena keterlambatan yang tidak terdeteksi dapat mengakibatkan ketidakpuasan pelanggan.

Ketiga, dengan hanya 3 sampel uji untuk kategori Lambat, satu saja kesalahan prediksi sudah menurunkan akurasi sebesar 33,33%, sehingga hasil ini belum reliabel secara statistik. Dengan demikian, rendahnya akurasi kategori Lambat bukan disebabkan oleh kelemahan algoritma Naive Bayes, melainkan lebih karena keterbatasan data pada kelas minoritas.

Keterbatasan Penelitian dan Potensi Overfitting

Meskipun model Naive Bayes menunjukkan performa yang baik (akurasi 93% pada data testing, rata-rata akurasi 5-fold CV 99% dengan standar deviasi rendah), terdapat keterbatasan dataset yang perlu dicermati sebelum hasil penelitian digeneralisasi secara luas. Pertama, ketidakseimbangan kelas (*class imbalance*) cukup mencolok di mana kategori Cepat mendominasi dengan 89 sampel (63,6%), kategori Sedang 41 sampel (29,3%), sementara kategori Lambat hanya 10 sampel (7,1%) yang berdampak langsung pada performa model dengan *recall* Lambat hanya 0,67 (33% data lambat tidak terdeteksi), yang mengindikasikan model cenderung bias memprediksi kelas mayoritas (Cepat) karena memiliki lebih banyak contoh pembelajaran. Kedua, periode pengambilan data yang terbatas pada 1-28 Februari

2026 menyebabkan model belum mampu menangkap pola musiman seperti bulan Ramadhan, hari raya Idul Fitri, musim liburan sekolah, akhir pekan, atau musim hujan yang memiliki karakteristik lalu lintas dan volume pesanan berbeda. Ketiga, ukuran dataset yang relatif kecil (140 sampel) membuat model rentan *overfitting*, sementara data yang hanya berasal dari satu kurir (ID SFD-BLR-1905) menyebabkan karakteristik pengiriman tidak dapat digeneralisasi ke kurir lain. Dengan demikian, hasil penelitian ini perlu diinterpretasikan secara hati-hati dan direplikasi pada data yang lebih besar, lebih seimbang, serta mencakup variasi musiman sebelum dapat digeneralisasi secara luas.

SIMPULAN

Metode *Naive Bayes* terbukti efektif untuk mengklasifikasikan estimasi waktu pengiriman kurir lokal ke dalam kategori cepat, sedang, dan lambat. Istilah "efektif" di sini didasarkan pada capaian akurasi keseluruhan sebesar 93% pada data uji, serta performa *precision* dan *F1-score* yang tergolong baik (*macro average F1-Score* 0,88). Namun demikian, efektivitas ini masih bersifat terbatas karena model memiliki kelemahan signifikan pada prediksi kategori lambat (akurasi hanya 66,67%) akibat ketidakseimbangan data ke dalam kategori cepat, sedang, dan lambat, dengan akurasi tertinggi pada kategori sedang (100%), diikuti cepat (92,59%), dan lambat (66,67%). Jarak tempuh menjadi faktor pembeda utama sementara kondisi cuaca berpengaruh kecil, meskipun model masih memiliki kelemahan pada prediksi kategori lambat akibat ketidakseimbangan data. Oleh karena itu, disarankan bagi penelitian selanjutnya untuk memperkaya fitur yang digunakan guna meningkatkan kualitas model, antara lain: kepadatan lalu lintas *real-time* (berdasarkan data Google Maps atau API lalu lintas) karena waktu perjalanan sangat dipengaruhi oleh kondisi kemacetan, jenis kendaraan (motor matic, motor bebek, atau mobil) yang memiliki kecepatan dan kelincahan berbeda di jalanan perkotaan, riwayat performa kurir (rata-rata waktu pengiriman 10 pesanan terakhir) untuk menangkap karakteristik individu kurir serta zona waktu pengiriman (pagi, siang, sore, malam) untuk menangkap pola kemacetan jam sibuk. Penambahan fitur-fitur ini diharapkan dapat meningkatkan akurasi model, terutama pada kategori lambat yang saat ini masih rendah (66,67%). s

DAFTAR PUSTAKA

- [1] P. Jain, "Research in Engineering Management and Science (Ijprems) Predicting Delivery Outcomes in Supply Chain Management Using Machine Learning : a Random Forest Classifier Approach," pp. 1107–1115, 2024.
- [2] N. Müller, P. Burggräf, F. Steinberg, C. R. Sauer, and M. Schütz, "An analytical review of predictive methods for delivery delays in supply chains," *Supply Chain Anal.*, vol. 11, no. May, 2025, doi: 10.1016/j.sca.2025.100130.
- [3] C. Boonma, "Thonburi University , Bangkok , Thailand Thonburi University , Bangkok , Thailand," no. 2022, pp. 72–78, 2025.
- [4] A. A. Ramadhan, "Waktu Untuk Meningkatkan Kepuasan Konsumen Dalam Logistik E-Commerce Indonesia," 2023.
- [5] A. Purnomo, N. Gia Ginasta, S. Syafrianita, and S. F. Pane, "Predictive Modeling of Delivery Delays in Transportation Using Machine Learning: A Comparative Study of

- Service Types," *Dinasti Int. J. Educ. Manag. Soc. Sci.*, vol. 7, no. 2, pp. 1765–1779, 2025, doi: 10.38035/dijemss.v7i2.5736.
- [6] Noraini Abu Talib, Rafiq Ahmad, and Siti Norbaya Noor, "A Comparative Analysis of Machine Learning Models for Time Series Forecasting in Finance," *Int. J. Appl. Math. Comput.*, vol. 1, no. 2, pp. 20–26, 2024, doi: 10.62951/ijamc.v1i2.71.
- [7] K. et al 2023, "OPTIMASI MODEL CVRP DAN MACHINE LEARNING UNTUK EFISIENSI DAN KETAHANAN RANTAI PASOK PANGAN," vol. 32, no. 3, pp. 167–186, 2021.
- [8] M. A. Firmansyah and M. M. Alamin, "Sistem Prediksi Pengiriman Pada Dakota Cargo Menggunakan Metode Naive Bayes Berbasis Web," *Jutisi J. Ilm. Tek. Inform. dan Sist. Inf.*, vol. 13, no. 1, p. 324, 2024, doi: 10.35889/jutisi.v13i1.1802.
- [9] A. D. Cahyani *et al.*, "Analisis Optimalisasi Distribusi Logistik untuk Meningkatkan Efisiensi dan Keefektifan Operasional pada PT. XX," *Ind. J. Ilm. Tek. Ind.*, vol. 8, no. 3, pp. 654–665, 2024, doi: 10.37090/indstrk.v8i3.1580.
- [10] N. D. Gunawan, J. Tji Beng, N. J. Perdana, D. Francesco, and D. Matthew, "Klasifikasi Kinerja Pengiriman Logistik Berbasis Prediksi Eta Menggunakan Machine Learning Di Pt. De Besta Trans Logistics Delivery Performance Classification Based on Eta Prediction Using Machine Learning At Pt. De Besta Trans," *J. Inf. Technol. Comput. Sci.*, vol. 8, no. 6, 2025.
- [11] R. Rachman and R. N. Handayani, "Klasifikasi Algoritma Naive Bayes Dalam Memprediksi Tingkat Kelancaran Pembayaran Sewa Teras UMKM," *J. Inform.*, vol. 8, no. 2, pp. 111–122, 2021, doi: 10.31294/ji.v8i2.10494.
- [12] D. M. Alfiansyah and H. Soetanto, "Prediksi Keterlambatan Pembayaran SPP Siswa dengan Pendekatan Metode Naive Bayes dan K-Nearest Neighbors," *Build. Informatics, Technol. Sci.*, vol. 5, no. 4, pp. 706–719, 2024, doi: 10.47065/bits.v5i4.4643.
- [13] K. Kartarina, N. K. Sriwinarti, and N. luh P. Juniarti, "Analisis Metode K-Nearest Neighbors (K-NN) Dan Naive Bayes Dalam Memprediksi Kelulusan Mahasiswa," *JTIM J. Teknol. Inf. dan Multimed.*, vol. 3, no. 2, pp. 106–112, 2021, doi: 10.35746/jtim.v3i2.159.
- [14] M. Ikhsan, L. N. Hayati, and D. Widyawati, "Analisis Sentimen Masyarakat Terhadap System Perbelanjaan Di Alfragift Dengan Metode Naive Bayes," *LINIER Lit. Inform. dan Komput.*, vol. 2, no. 2, pp. 255–263, 2025, doi: 10.33096/linier.v2i2.3134.
- [15] F. Panjaitan *et al.*, "Studi Komparatif Algoritma Machine Learning Pada Analisis Sentimen Media Sosial," *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 9, no. 2, pp. 3145–3152, 2025, doi: 10.36040/jati.v9i2.13277.
- [16] A. Wijaya and R. A. Putri, "Penerapan Metode Fuzzy Sugeno Untuk Optimalisasi Persediaan Pakaian," *Bull. Comput. Sci. Res.*, vol. 5, no. 5, pp. 1224–1234, 2025, doi: 10.47065/bulletincsr.v5i5.770.
- [17] M. D. Ananda, K. N. Malik, A. F. N. Masruriyah, and M. Mardiah, "Studi Komparatif Algoritma K-Means dan K-Medoids untuk Segmentasi Informasi Kesehatan," *Comput. Sci.*, vol. 5, no. 2, pp. 103–112, 2025, doi: 10.31294/coscience.v5i2.9207.
- [18] C. Chairunnita, A. Handayanto, and F. M. Dewanto, "Penerapan Algoritma K-Means Clustering dalam Analisis Pengelompokan Produk Toko Oleh-Oleh Berdasarkan Data Penjualan," *J. Inf. Syst. Res.*, vol. 6, no. 4, pp. 1816–1825, 2025, doi: 10.47065/josh.v6i4.7832.
- [19] R. A. Subektianto, M. Ratna, and I. Zain, "Analisis Kepuasan Pasien dan Pelaksana Terhadap Sistem Informasi Manajemen Puskesmas Elektronik (Simpustronik) di

- Kabupaten Blitar Menggunakan Regresi Logistik Biner," *Inferensi*, vol. 1, no. 1, p. 29, 2018, doi: 10.12962/j27213862.v1i1.6718.
- [20] M. Fatihul, F. Dzulfaqqor, I. Kadek, and D. Nuryana, "Journal of Emerging Information System and Business Intelligence Prediction of Goods Damage in Land Transportation Services (Trucking) Using Naïve Bayes," *J. Emerg. Inf. Syst. Bus. Intell.*, vol. 6, no. 3, pp. 399–410, 2025.
- [21] Ahmad Bahar, Tri Astuti, and Primandani Arsi, "Performance Comparison of Svm, Naive Bayes, and Logistic Regression Classification Algorithms in Analyzing Noice App User Reviews," *J. Tek. Inform.*, vol. 5, no. 4, pp. 469–477, 2024, doi: 10.52436/1.jutif.2024.5.4.2061.