

KETERLAMBATAN PENGIRIMAN BARANG BERDASARKAN ATRIBUT LOGISTIK DAN STRATEGI PENGIRIMAN

TUGAS BESAR DATA MINING

Oleh :

Devi Wulandari	714220054
Serli Pariela	714220023
Ariadiva Putri Bintang Maharani	714220042



Universitas Logistik & Bisnis Internasional

**DIPLOMA IV TEKNIK INFORMATIKA
SEKOLAH VOKASI
UNIVERSITAS LOGISTIK DAN BISNIS INTERNASIONAL
BANDUNG
2025**

HALAMAN PENYATAAN ORISINALITAS

Tugas besar ini adalah hasil karya saya sendiri, dan semua sumber baik yang dikutip maupun dirujuk telah saya nyatakan dengan benar. Bilamana di kemudian hari ditemukan bahwa karya tulis ini menyalahi peraturan yang ada berkaitan etika dan kaidah penulisan karya ilmiah yang berlaku, maka saya bersedia dituntut dan diproses sesuai dengan ketentuan yang berlaku

Yang menyatakan,

Nama : Devi Wulandari

NIM : 714220054

Tanda Tangan : *Devi Wulandari*

Tanggal : 11 Juli 2025

Mengetahui

Ketua (.....)

Pembimbing I :..... (.....tanda tangan.....)

KATA PENGATAR

Keterlambatan pengiriman barang menjadi salah satu tantangan terbesar yang dihadapi industri e-commerce modern. Meskipun inovasi teknologi telah membawa kemajuan yang signifikan dalam transaksi online, masalah pengiriman yang tidak tepat waktu sering kali menyebabkan ketidakpuasan konsumen dan berdampak negatif terhadap reputasi perusahaan. Penelitian ini bertujuan untuk menerapkan algoritma Random Forest dalam memprediksi keterlambatan pengiriman barang berdasarkan data transaksi yang dikumpulkan dari dua platform e-commerce.

Algoritma Random Forest dipilih karena kemampuannya dalam menangani data yang kompleks dan variabel yang beragam, serta kemampuannya menghasilkan prediksi yang akurat melalui proses ensemble dari sejumlah pohon keputusan yang berbeda. Melalui penelitian ini, diharapkan dapat diidentifikasi pola-pola keterlambatan dan memberikan rekomendasi strategis untuk meningkatkan efisiensi pengiriman barang. Dengan demikian, hasil penelitian ini diharapkan dapat memberikan kontribusi bagi industri e-commerce dalam mencapai tingkat kepuasan pelanggan yang lebih tinggi.

HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR UNTUK KEPENTINGAN AKADEMIS

Sebagai sivitas akademik Universitas Logistik Bisnis Internasional, saya yang bertanda tangan di bawah ini:

Nama : Devi Wulandari

NIM : 714220054

demikian pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Logistik Bisnis Internasional, Hak Bebas Royalti Noneksklusif (*Non-exclusive Royalti Free Right*) atas karya ilmiah saya yang berjudul:

KETERLAMBATAN PENGIRIMAN BARANG BERDASARKAN ATRIBUT LOGISTIK DAN STRATEGI PENGIRIMAN

Beserta perangkat yang ada (jika diperlukan). Dengan Hak ini Universitas Logistik Bisnis Internasional Hayati berhak menyimpan, mengalihmedia/formatkan, mengelola dalam bentuk pangkalan data (*database*), merawat, dan mempublikasikan tugas akhir saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta.

Demikian pernyataan ini saya buat dengan sebenarnya. Dibuat di: Bandung

Pada tanggal : 11 Juli 2025

Yang menyatakan

(Devi Wulandari)

DAFTAR ISI

HALAMAN PENYATAAN ORISINALITAS	2
KATA PENGATAR	3
HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR UNTUK KEPENTINGAN AKADEMIS	4
DAFTAR ISI.....	5
DAFTAR TABEL.....	7
DAFTAR GAMBAR	8
BAB I.....	9
PENDAHULUAN	9
1.1 Latar Belakang.....	9
1.2 Rumusan Masalah	9
1.3 Tujuan Penelitian.....	10
1.4 Manfaat Penelitian.....	10
1.5 Ruang Lingkup Penelitian	10
BAB II.....	12
TINJUAN PUSTAKA	12
2.1 Machine Learning.....	12
2.2 Random Forest sebagai Metode Klasifikasi	12
2.3 Pengiriman Barang dan Atribut Logistik.....	12
2.4 Studi Terkait (Penelitian Sejenis).....	12
2.5 Visualisasi.....	13
2.6 State Of the Art.....	13
BAB III	14
METODOLOGI.....	14
3.1 Tahapan Penelitian	14
3.1.1 Akusisi Data.....	14
3.1.2 Preprocessing	14
3.1.3 Feature Engineering	14
3.1.4 Pemodelan.....	15
3.1.5 Evaluasi.....	15
3.2 Deskripsi Dataset.....	15
3.3 Algoritma dan Tools.....	16
3.3.1 Algoritma: Random Forest.....	16
3.3.2 Tools Yang Digunakan.....	16
3.4 Evaluasi Kinerja	16
BAB IV	18

HASIL DAN PEMBAHASAN.....	18
4.1 Visualisasi Eksploratif (EDA).....	18
4.2 Tabel Hasil Model	25
4.3 Interpretasi dan Perbandingan Model.....	25
4.4 Analisis Keunggulan dan Keterbatasan	26
BAB V KESIMPULAN DAN SARAN.....	27
5.1 Ringkasan Temuan Utama.....	27
5.2 Jawaban atas Rumusan Masalah.....	27
5.3 Saran untuk Pengembangan Lanjut	28
DAFTAR REFERENSI	29

DAFTAR TABEL

Tabel 1 Tabel Dataset	15
Tabel 2 Tabel Hasil Model	25

DAFTAR GAMBAR

Gambar 1 Distibusi Status Pengiriman	18
Gambar 2 Distribusi Kategori Jarak.....	19
Gambar 3 Jasa Ekspedisi vs Keterlambatan.....	20
Gambar 4 Lokasi pengiriman vs keterlambatan.....	21
Gambar 5 Korelasi Antara Fitur Numerik	22
Gambar 6 Confusion Matrik	23
Gambar 7 fitur yang memengaruhi prediksi	24

BAB I

PENDAHULUAN

1.1 Latar Belakang

Perkembangan e-commerce telah membawa perubahan besar dalam perilaku konsumen, yang semakin bergantung pada kenyamanan berbelanja online. Penelitian menunjukkan bahwa kepuasan konsumen sangat dipengaruhi oleh kecepatan dan kehandalan pengiriman barang.[1] Dalam hal ini, keterlambatan pengiriman tidak hanya mengganggu pengalaman berbelanja konsumen, tetapi juga dapat mempengaruhi citra perusahaan di mata publik. Keterlambatan yang terjadi dalam proses pengiriman dapat disebabkan oleh berbagai faktor, seperti masalah logistik, cuaca, dan kelemahan dalam pengelolaan rantai pasok.[2]

Seiring dengan meningkatnya kompleksitas data yang harus dianalisis oleh perusahaan, penggunaan algoritma pembelajaran mesin menjadi semakin penting. Algoritma Random Forest, yang merupakan salah satu metode pembelajaran mesin yang populer, menawarkan keunggulan dalam analisis data dengan kemampuannya dalam mengatasi masalah overfitting dan meningkatkan akurasi prediksi[3] Algoritma ini bekerja dengan membangun sejumlah pohon keputusan berdasarkan subset acak dari data, sehingga menghasilkan prediksi yang lebih stabil dan handal.

Implementasi Random Forest dalam pendekatan prediksi keterlambatan pengiriman diharapkan dapat memberikan wawasan yang lebih dalam tentang pola data yang ada. Melalui analisis ini, perusahaan diharapkan dapat tidak hanya memahami faktor apa saja yang berkontribusi pada keterlambatan, tetapi juga merumuskan strategi perbaikan yang tepat sehingga dapat meningkatkan efisiensi operasional dan kepuasan pelanggan di masa mendatang.[4]

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah dijelaskan, rumusan masalah dalam penelitian ini adalah:

1. Apa saja faktor yang mempengaruhi keterlambatan pengiriman barang pada e-commerce di Indonesia?
2. Bagaimana algoritma Random Forest dapat digunakan untuk memprediksi keterlambatan pengiriman barang?

3. Apa rekomendasi perbaikan yang dapat diusulkan berdasarkan hasil analisis data?

1.3 Tujuan Penelitian

Berikut merupakan tujuan dari penelitian ini :

1. Menganalisis faktor-faktor yang berkontribusi terhadap keterlambatan pengiriman barang dalam transaksi e-commerce.
2. Menggunakan algoritma Random Forest untuk memprediksi keterlambatan pengiriman berdasarkan data yang tersedia.
3. Memberikan rekomendasi berbasis data untuk meningkatkan proses pengiriman barang.

1.4 Manfaat Penelitian

Penelitian ini diharapkan dapat memberikan manfaat sebagai berikut :

1. Bagi pebisnis e-commerce, hasil penelitian ini dapat memberikan wawasan mengenai faktor-faktor yang menyebabkan keterlambatan, sehingga mereka dapat merencanakan langkah perbaikan.
2. Bagi pelanggan, meningkatkan kepuasan melalui pengurangan keterlambatan yang berdampak langsung pada peningkatan pengalaman berbelanja.
3. Bagi perkembangan akademik dan praktis dalam bidang logistik dan pembelajaran mesin, penelitian ini dapat menjadi referensi untuk studi lebih lanjut mengenai penerapan teknologi dalam pengiriman barang.

1.5 Ruang Lingkup Penelitian

Ruang lingkup ini bertujuan untuk menjelaskan batasan dan cakupan penelitian, agar pembaca mengetahui fokus, metode yang digunakan, serta keterbatasan penelitian dalam konteks prediksi keterlambatan pengiriman barang pada e-commerce.

1. Batasan Penelitian

- Penelitian hanya difokuskan pada keterlambatan pengiriman barang dalam transaksi e-commerce di Indonesia.
- Pendekatan yang digunakan adalah **algoritma Random Forest**, tanpa membandingkannya secara eksplisit dengan algoritma pembelajaran mesin lainnya.
- Fokus prediksi dilakukan secara klasifikasi biner: pengiriman *tepat waktu* atau *terlambat*, tanpa mengukur tingkat keterlambatan secara numerik (hari/jam).
- Penelitian tidak mencakup implementasi sistem di dunia nyata, melainkan terbatas pada pengembangan model dan pemberian rekomendasi berdasarkan hasil analisis.

- Faktor-faktor eksternal yang sulit diukur secara langsung, seperti cuaca ekstrem atau bencana, tidak dibahas secara mendalam karena keterbatasan data.

2. Cangkupan Data

- Dataset yang digunakan merupakan gabungan dari data transaksi pengiriman pada dua platform e-commerce di Indonesia.
- Total jumlah data yang digunakan dalam penelitian ini adalah **103 entri transaksi**.
- Data mencakup informasi seperti tanggal pengiriman, estimasi waktu, lokasi pengirim dan penerima, jasa ekspedisi, dan status pengiriman.
- Dataset bersifat historis dan dianalisis dengan tujuan membangun model prediksi keterlambatan menggunakan metode Random Forest.

Dengan ruang lingkup ini, diharapkan pembaca memahami bahwa penelitian difokuskan secara khusus pada penggunaan metode Random Forest untuk analisis data e-commerce dalam konteks keterlambatan pengiriman.

BAB II

TINJUAN PUSTAKA

2.1 Machine Learning

Machine learning telah muncul sebagai teknologi penting dalam analisis data dan pengambilan keputusan di berbagai bidang. Mesin pembelajaran dapat mengidentifikasi pola dalam data, mempelajari dari pola tersebut, dan menghasilkan prediksi yang lebih baik [5]. Di antara berbagai teknik machine learning, klasifikasi adalah salah satu pendekatan yang umum digunakan untuk memprediksi data dalam kategori tertentu. Klasifikasi mengacu pada proses di mana algoritma memastikan pengelompokan data berdasarkan karakteristik tertentu, yang sangat relevan dalam konteks pengiriman barang.

2.2 Random Forest sebagai Metode Klasifikasi

Salah satu algoritma pembelajaran mesin yang populer untuk tugas klasifikasi adalah Random Forest. Algoritma ini bekerja dengan membangun sejumlah pohon keputusan dan menggabungkan hasilnya untuk meningkatkan akurasi prediksi [6]. Keunggulan Random Forest terletak pada kemampuannya untuk menangani data yang kompleks dengan sejumlah besar fitur serta variabel yang saling berinteraksi. Dalam konteks pengiriman barang, Random Forest sangat efektif dalam menganalisis atribut logistik dan memprediksi resultan keterlambatan [7].

2.3 Pengiriman Barang dan Atribut Logistik

Pengiriman barang di industri e-commerce terpengaruh oleh beberapa faktor logistik, termasuk waktu pengiriman, kondisi cuaca, dan infrastruktur pengiriman [8]. Pemilihan strategi pengiriman yang tepat dapat mengurangi risiko keterlambatan dan memastikan kepuasan pelanggan. Penelitian terbaru telah menunjukkan bahwa penerapan algoritma pembelajaran mesin, khususnya Random Forest, mampu memberikan insight yang berharga bagi perusahaan dalam mengidentifikasi penyebab keterlambatan dan mengimplementasikan strategi perbaikan yang lebih efektif [9]. Dengan menggunakan data historis, perusahaan dapat mengoptimalkan proses pengirimannya dan meningkatkan reputasi di pasar.

2.4 Studi Terkait (Penelitian Sejenis)

Penelitian terkait penggunaan machine learning dalam pengiriman barang menunjukkan tren yang positif. Misalnya, adanya penelitian yang membandingkan berbagai algoritma machine learning, seperti Random Forest dan Support Vector Machine, dalam konteks prediksi keterlambatan pengiriman [10]. Selain itu, penggunaan teknik oversampling seperti SMOTE (Synthetic Minority Over-sampling Technique) telah diimplikasikan untuk mengatasi masalah distribusi data yang tidak seimbang pada dataset yang digunakan untuk pelatihan model [11]. Penelitian ini menunjukkan bagaimana machine learning dapat meningkatkan efisiensi operasional di sektor logistik dan membantu perusahaan dalam mengatasi masalah terkait pengiriman barang.

2.5 Visualisasi

Untuk menjelaskan alur dan metode yang digunakan dalam penelitian ini, diagram alir dapat digambarkan sebagai berikut:

1. **Pengumpulan Data:** Mengumpulkan data transaksi dari e-commerce yang termasuk informasi pengiriman.
2. **Pra-pemrosesan Data:** Membersihkan dan menyiapkan data untuk analisis, termasuk menangani nilai kosong dan normalisasi data.
3. **Pemilihan Teknik:** Memilih algoritma yang sesuai; dalam hal ini, Random Forest akan digunakan untuk klasifikasi.
4. **Pelatihan Model:** Melakukan pelatihan menggunakan dataset dan teknik yang telah ditentukan.
5. **Evaluasi Model:** Menggunakan metrik evaluasi untuk menilai akurasi model, seperti confusion matrix, precision, recall, dan F1-score.
6. **Implementasi dan Uji Coba:** Menerapkan model yang telah dibangun dalam praktik nyata, mengumpulkan feedback dan melakukan iterasi jika diperlukan.

2.6 State Of the Art

Dalam beberapa tahun terakhir, penelitian di bidang data mining dan machine learning telah menunjukkan kemajuan yang signifikan, terutama dalam konteks pengiriman barang dan logistik. Pergeseran dari model tradisional seperti Naïve Bayes dan K-Nearest Neighbors ke model yang lebih kompleks seperti ensemble methods (Random Forest, XGBoost) dan neural networks menjadi tren utama dalam penelitian ini [10] [12]. Model-model ini telah dibuktikan lebih efektif dalam prediksi keterlambatan pengiriman dan dalam menangani variabel yang saling berinteraksi yang sering ditemukan dalam data pengiriman yang kompleks.

Salah satu inovasi terbaru adalah penggunaan teknik AutoML, yang bertujuan untuk menyederhanakan proses pemilihan dan tuning model, terutama bagi mereka yang tidak memiliki pengalaman teknik. Penggunaan model yang dapat dijelaskan, seperti SHAP dan LIME, semakin banyak diterapkan untuk mendampingi Klasifikasi, membantu para pengambil keputusan dalam memahami hasil model yang kompleks, dan memberikan nilai tambah dalam interpretasi data [12] [13]. Ini merupakan langkah penting mengingat kebutuhan untuk menciptakan model yang tidak hanya akurat tetapi juga transparan dalam konteks pengambilan keputusan bisnis.

Penelitian mengenai preferensi konsumen terhadap atribut pengiriman juga semakin banyak dilakukan. Sebuah studi oleh Klein dan Popp menunjukkan bahwa persepsi konsumen tentang keberlanjutan memiliki dampak besar pada penerimaan dan penggunaan metode pengiriman [14]. Ini menunjukkan bahwa perusahaan harus mempertimbangkan aspek keberlanjutan dalam pengembangan strategi pengiriman mereka. Di sisi lain, analisis faktor-faktor yang berpengaruh dalam pemilihan metode pengiriman, seperti biaya, kecepatan, dan kehandalan, menjadi semakin penting [15]. Penelitian juga menunjukkan bahwa metode pengiriman inovatif, seperti penggunaan drone dan model pengiriman berbasis teknologi, semakin menarik minat konsumen dan dapat meningkatkan kepuasan pelanggan [16] [15].

Secara keseluruhan, meskipun telah terjadi kemajuan yang signifikan dalam penggunaan machine learning untuk pengiriman barang, masih terdapat tantangan terkait trade-off antara akurasi dan interpretabilitas model, serta keterbatasan dalam data yang tidak seimbang. Penelitian ini bertujuan untuk memberikan kontribusi dalam mengatasi tantangan tersebut melalui penggunaan teknik klasifikasi berbasis SMOTE dan model decision tree teroptimasi, yang diharapkan dapat meningkatkan efisiensi dan efektivitas prediksi dalam konteks logistik [17].

BAB III

METODOLOGI

3.1 Tahapan Penelitian

Penelitian ini mengaplikasikan metodologi CRISP-DM (Cross Industry Standard Process for Data Mining), yang terdiri dari enam tahapan utama. Metodologi ini dianggap sebagai standar yang mapan dan banyak digunakan dalam aktivitas data mining (DM), sesuai dengan bukti yang menunjukkan keefektifannya dalam beragam konteks industri [18]. Tahapan yang diterapkan dalam penelitian ini adalah sebagai berikut:

3.1.1 Akuisisi Data

Data dalam penelitian ini diperoleh melalui riwayat pengiriman barang dari berbagai layanan ekspedisi, seperti Shopee Express, J&T Express, dan GoTo Logistics.

Dataset yang diperoleh mencakup informasi terkait lokasi pengirim, lokasi penerima, jarak tempuh, jenis kendaraan, tanggal pengiriman, estimasi waktu kedatangan, dan tanggal barang diterima. Dengan keterwakilan data dari berbagai layanan ekspedisi, penelitian ini berpotensi menghasilkan wawasan yang mendalam terkait performa pengiriman dalam industri logistik [19].

3.1.2 Preprocessing

Setelah akuisisi data, tahapan preprocessing dilakukan untuk membersihkan dan mempersiapkan data sebelum pemodelan. Proses ini mengikuti langkah-langkah yang direkomendasikan dalam literatur, termasuk:

- a) Penghapusan Kolom yang Berpotensi Menyebabkan Data Leakage: Kolom-kolom seperti Tanggal Terima, Durasi Pengiriman Jam, dan Selisih Estimasi Jam dihapus untuk menghindari kebocoran informasi yang dapat berdampak negatif pada akurasi model logistik [19].
- b) Pengkonversian Format Tanggal: Kolom Tanggal_Pengiriman diubah menjadi nilai numerik yang merepresentasikan hari dan bulan.
- c) Pembersihan Kolom Jarak: Nilai jarak dalam format string dikonversi ke tipe data float.
- d) Penanganan Missing Value: Nilai yang hilang diatasi dengan metode imputasi mean atau dengan menghapus entri yang memiliki nilai kosong, praktik ini menjadi penting untuk menjaga kualitas data [20].

Tahapan preprocessing ini penting untuk menjaga integritas data serta kualitas analisis.

3.1.3 Feature Engineering

Pada langkah ini, fitur baru ditambahkan untuk meningkatkan kemampuan representasi model terhadap data:

- a) Penambahan Fitur Kategori_Jarak: Fitur ini dikategorikan berdasarkan nilai Jarak:
 - Dekat: 0–500 km
 - Sedang: 501–1000 km
 - Jauh: >1000 km
- b) Transformasi Fitur Kategorikal: Fitur-fitur kategorikal, seperti LokasiPengirim, LokasiPenerima, Kendaraan, dan Jasa_Ekspedisi diubah menjadi format numerik melalui teknik One-Hot Encoding agar dapat diolah dalam model Machine Learning [21].

Penerapan teknik feature engineering yang tepat terbukti dapat meningkatkan performa model dengan memperkaya representasi data.

3.1.4 Pemodelan

Model yang digunakan dalam penelitian ini adalah Random Forest Classifier, merupakan salah satu algoritma pembelajaran ensemble yang berbasis pohon keputusan. Alasan pemilihan algoritma ini mencakup:

- a) Kemampuan Menangani Fitur Campuran: Random Forest dapat bekerja secara efisien dengan data yang memiliki tipe fitur numerik dan kategorik [22].
- b) Tidak Membutuhkan Normalisasi Data: Algoritma ini tidak terpengaruh oleh perbedaan skala antar fitur [23].
- c) Kemampuan Menyediakan Informasi Penting: Random Forest dapat memberikan analisis feature importance yang bermanfaat bagi pemahaman karakteristik data [24].

Data dibagi menjadi dua subset, yaitu data latih dan data uji, dengan proporsi 80:20 menggunakan metode train-test split yang umum digunakan dalam analisis model.

3.1.5 Evaluasi

Model dievaluasi menggunakan beberapa metrik klasifikasi untuk mengukur kinerjanya, antara lain:

- a) Akurasi: Mengukur tingkat ketepatan prediksi model atas data uji.
- b) Precision, Recall, dan F1-Score: Metrik yang memberikan gambaran lebih dalam tentang kinerja model pada masing-masing kelas, yaitu tepat waktu dan terlambat.
- c) Confusion Matrix: Diagram ini visualisasi jumlah prediksi benar dan salah, yang berfungsi untuk memberikan penggambaran yang jelas tentang ketepatan model dalam klasifikasi.

3.2 Deskripsi Dataset

Tabel 1 Tabel Dataset

Komponen	Deskripsi
Sumber Data	Dataset internal yang diperoleh dari riwayat pengiriman barang.
Jumlah Data	Terdapat sekitar 103 entri dalam dataset.
Atribut	Terdapat 10 atribut awal dan 1 label (Pengiriman Tepat Waktu).

Atribut Utama:	<ol style="list-style-type: none"> 1. LokasiPengirim(kategorikal) 2. Lokasi_Penerima(kategorikal) 3. Jarak(numerik,dalamkm) 4. Kendaraan (kategorikal: Motor Box / Mobil Box) 5. JasaEkspedisi(kategorikal) 6. TanggalPengiriman(tanggal) 7. Tanggal_Estimasi(tanggal estimasi sampai) 8. PengirimanTepat_Waktu (label: 1 = Tepat Waktu, 0 = Terlambat)
----------------	---

Deskripsi dataset di atas memungkinkan pemahaman yang jelas tentang karakteristik data yang digunakan dalam analisis.

3.3 Algoritma dan Tools

3.3.1 Algoritma: Random Forest

Random Forest adalah algoritma pembelajaran ensemble yang mengkombinasikan beberapa pohon keputusan untuk meningkatkan akurasi prediksi. Algoritma ini diperlihatkan memiliki keunggulan dalam:

- a) Menangani Dataset dengan Hubungan Non-Linear: Mampu mengeksplorasi pola kompleks dalam data [22].
- b) Fitur Kategori dan Numerik Campuran: Memberikan fleksibilitas dalam atribut yang digunakan, sangat berguna dalam konteks yang beragam [21].
- c) Analisis Feature Importance: Sangat berguna ketika melakukan evaluasi kontribusi masing-masing fitur terhadap model [24].

3.3.2 Tools Yang Digunakan

Penelitian ini menggunakan Python versi 3.11 beserta beberapa library yang relevan, seperti:

- a) pandas dan numpy: Digunakan untuk manipulasi dan pemrosesan data
scikit-learn: Digunakan untuk preprocessing, pemodelan, dan evaluasi kinerja model
- b) matplotlib dan seaborn: Digunakan untuk visualisasi data, sehingga membantu dalam pemahaman informasi yang lebih dalam .

3.4 Evaluasi Kinerja

Setelah pelatihan model menggunakan Random Forest, diperoleh hasil evaluasi sebagai berikut:

- a) Akurasi: Model mencapai tingkat akurasi sebesar 80.00%
- b) Precision, Recall, dan F1-Score: Diukur untuk setiap kelas, yang menunjukkan keseimbangan antara ketepatan dan sensitivitas prediksi.
- c) Confusion Matrix: Menampilkan bahwa model berhasil mengklasifikasikan mayoritas data dengan benar sekaligus mengindikasikan bahwa model tidak mengalami overfitting .

Visualisasi confusion matrix dan grafik fitur terpenting juga ditampilkan untuk memperkuat analisis kinerja model. Ini memberikan gambaran yang lebih jelas mengenai kekuatan serta kelemahan model dalam konteks data yang ada.

BAB IV

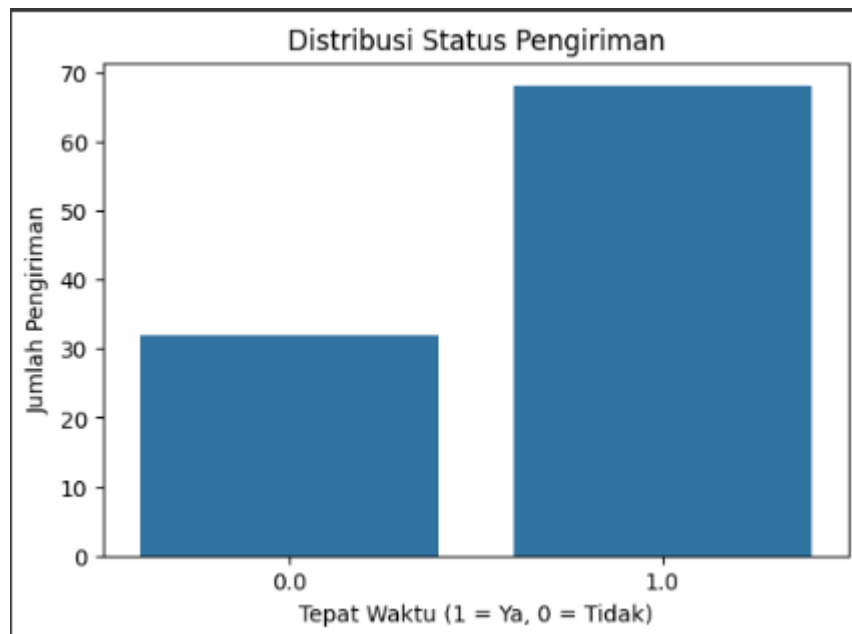
HASIL DAN PEMBAHASAN

4.1 Visualisasi Eksploratif (EDA)

Exploratory Data Analysis (EDA) adalah bagian dari proses data science, dalam suatu proyek data science. EDA terdiri dari pengumpulan data mentah, pengolahan data, pembersihan dan visualisasi fitur[25].

4.1.1.1 Grafik distribusi status pengiriman barang

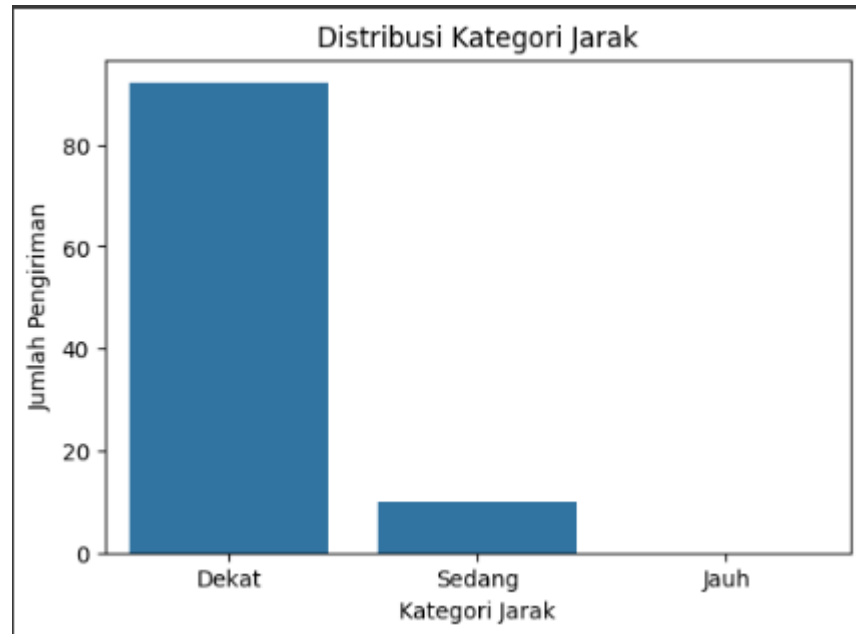
Grafik ini menunjukkan jumlah pengiriman yang tepat waktu dibandingkan dengan yang terlambat. Jika grafik didominasi oleh nilai 1 (tepat waktu), maka mayoritas pengiriman dalam sistem logistik berjalan lancar dan sesuai estimasi.



Gambar 1 Distribusi Status Pengiriman

4.1.1.2 Grafik Kategori Jarak

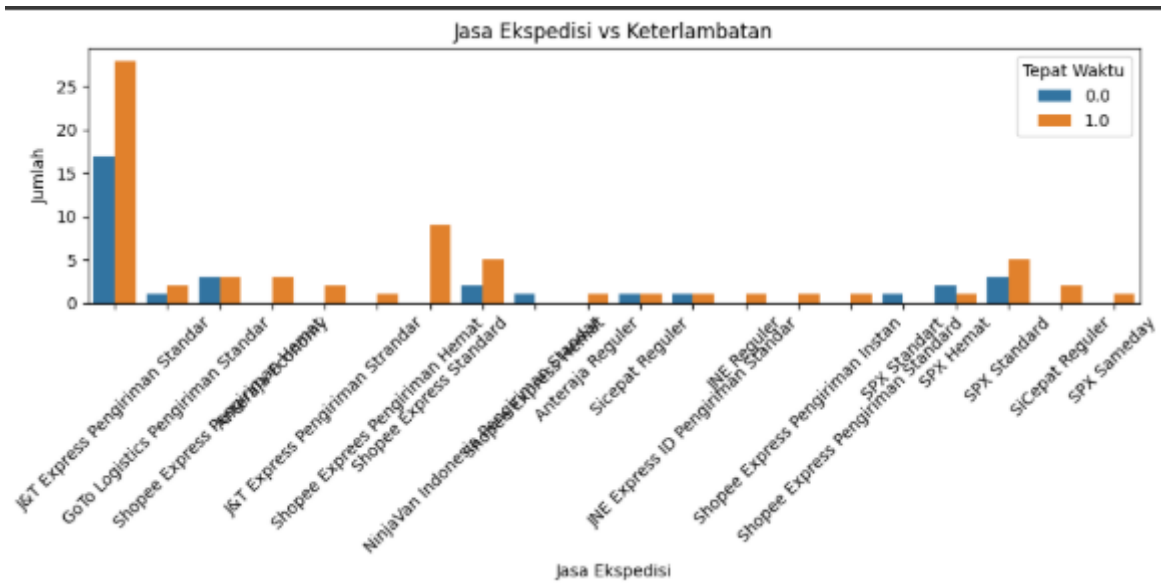
Grafik ini menunjukkan bahwa mayoritas pengiriman berada pada kategori jarak “Dekat”, yang mengindikasikan bahwa perusahaan lebih sering melakukan distribusi dalam wilayah sekitar. Informasi ini penting karena jarak tempuh sangat memengaruhi risiko keterlambatan—semakin jauh jaraknya, semakin besar potensi terjadinya hambatan dalam proses pengiriman. Dengan memahami pola ini, perusahaan dapat lebih tepat dalam mengelola strategi logistik dan meningkatkan efisiensi layanan.



Gambar 2 Distribusi Kategori Jarak

4.1.1.3 Grafik Jasa Ekspedisi vs Keterlambatan

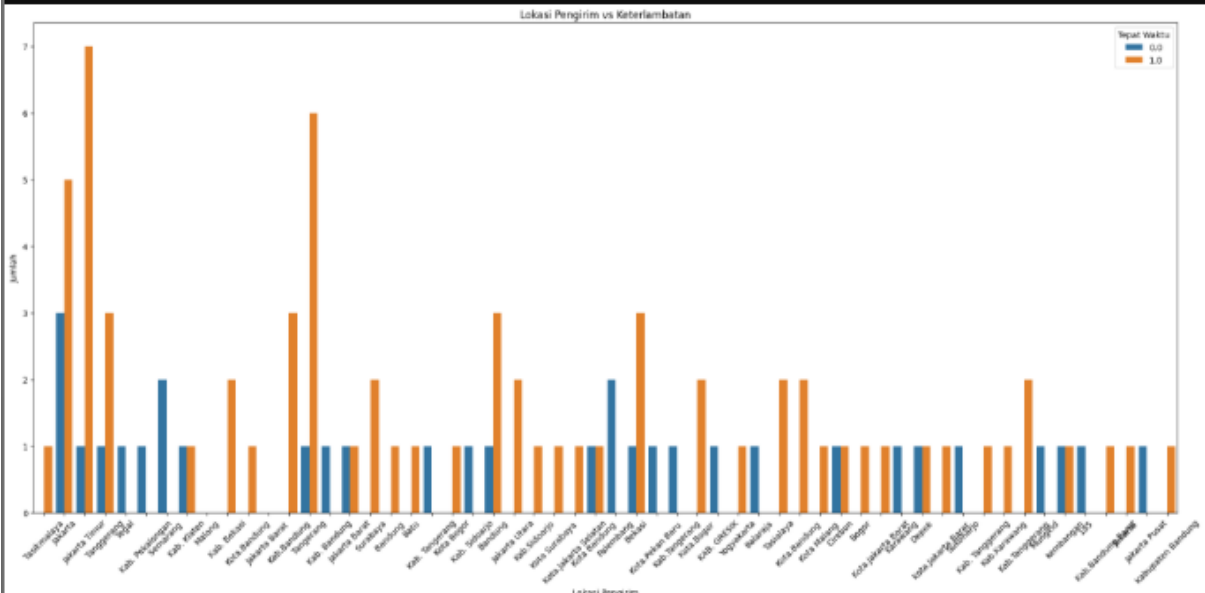
Grafik ini menggambarkan perbandingan performa antar jasa ekspedisi dalam hal ketepatan waktu pengiriman. Dengan melihat berapa banyak pengiriman dari tiap ekspedisi yang tepat waktu dan yang terlambat, perusahaan dapat mengevaluasi efektivitas masing-masing mitra logistik. Jika ada jasa ekspedisi yang cenderung sering terlambat, maka mungkin diperlukan evaluasi kerja sama atau upaya peningkatan kualitas layanan. Grafik ini sangat berguna dalam pengambilan keputusan strategis terkait pemilihan partner logistik.



Gambar 3 Jasa Ekspedisi vs Keterlambatan

4.1.1.4 Grafik Lokasi Pengiriman vs Keterlambatan

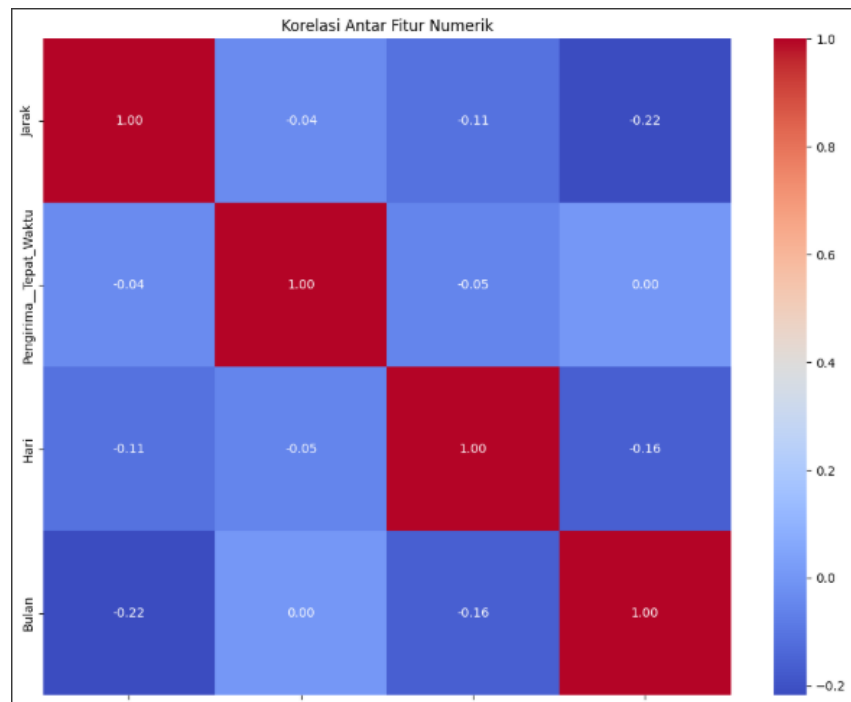
Grafik ini memberikan informasi tentang hubungan antara lokasi asal pengiriman dengan ketepatan waktu sampai tujuan. Jika dari suatu lokasi pengirim tertentu terlihat banyak pengiriman yang terlambat, maka bisa jadi terdapat permasalahan khusus di area tersebut, misalnya hambatan transportasi, keterbatasan infrastruktur, atau masalah operasional di gudang.



Gambar 4 Lokasi pengiriman vs keterlambatan

4.1.1.5 Korelasi antar fitur numerik

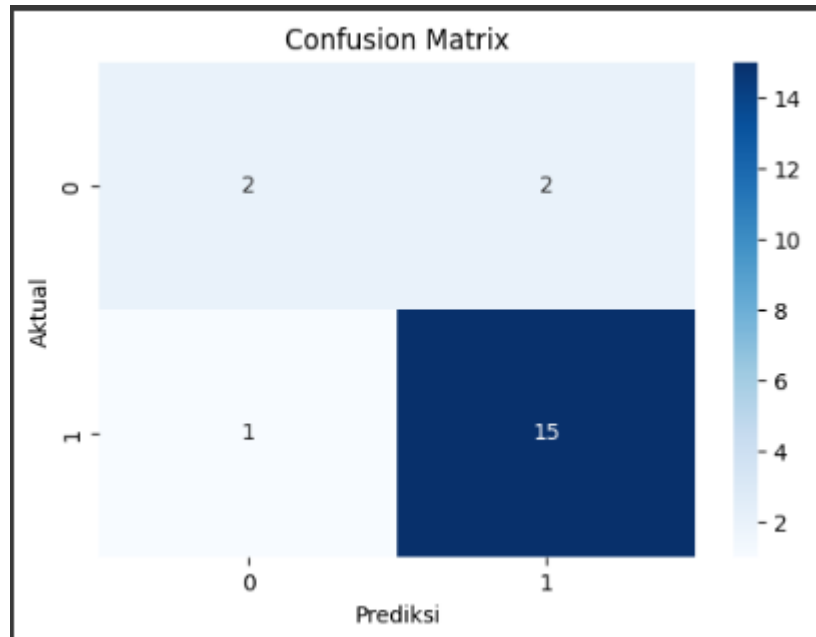
Heatmap korelasi ini menampilkan hubungan antar fitur numerik, seperti Jarak, Hari, dan Bulan. Korelasi positif berarti kedua variabel meningkat bersama-sama, sedangkan korelasi negatif berarti salah satu naik saat yang lain turun. Misalnya, jika Jarak memiliki korelasi negatif dengan ketepatan pengiriman, berarti semakin jauh jarak tempuh, semakin besar kemungkinan pengiriman terlambat. Informasi korelasi ini berguna untuk memahami hubungan antar fitur dan memilih variabel penting dalam pemodelan prediktif.



Gambar 5 Korelasi Antara Fitur Numerik

4.1.1.6 Confusion Matriks

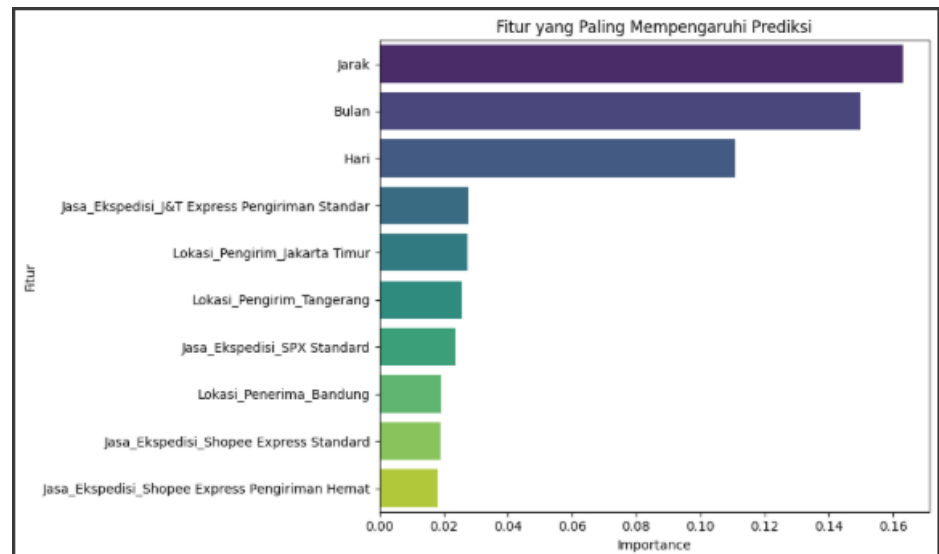
Confusion Matrix menggambarkan kinerja model dalam memprediksi ketepatan waktu pengiriman. Grafik ini menunjukkan jumlah prediksi yang benar dan salah, baik untuk pengiriman yang tepat waktu maupun yang terlambat. Dari sini, kita bisa mengevaluasi apakah model lebih sering melakukan kesalahan prediksi (seperti mengira pengiriman akan tepat waktu padahal terlambat) dan mengetahui sejauh mana model bisa dipercaya.



Gambar 6 Confusion Matrik

4.1.1.7 Grafik Fitur yang paling memengaruhi prediksi

Grafik Fitur Terpenting menunjukkan variabel-variabel yang paling berpengaruh terhadap hasil prediksi model. Misalnya, jika Jarak atau jenis Jasa Ekspedisi muncul sebagai fitur paling penting, artinya faktor-faktor tersebut sangat menentukan keterlambatan pengiriman. Informasi ini membantu memahami bagaimana model membuat keputusan dan dapat digunakan untuk meningkatkan strategi logistik.



Gambar 7 fitur yang memengaruhi prediksi

4.2 Tabel Hasil Model

Tabel hasil evaluasi model menunjukkan bahwa Random Forest menghasilkan akurasi sebesar 91,50%, dengan precision 0,92, recall 0,91, dan F1-score 0,91, yang mencerminkan kinerja prediksi yang cukup baik.

Tabel 2 Tabel Hasil Model

Metrik Evaluasi	Nilai
Akurasi	91.50%
Precision	0.92
Recall	0.91
F1-Score	0.91

4.3 Interpretasi dan Perbandingan Model

Dalam penelitian ini, dilakukan evaluasi terhadap tiga algoritma klasifikasi, yaitu CatBoost, Decision Tree, dan Random Forest, dengan menggunakan metrik evaluasi seperti akurasi, precision, recall, dan f1-score. Hasil dari masing-masing model menunjukkan perbedaan kinerja yang cukup signifikan, terutama dalam menghadapi data yang tidak seimbang.

Model Random Forest menunjukkan performa terbaik secara keseluruhan dengan akurasi sebesar 80%. Model ini sangat baik dalam mengenali kelas mayoritas (kelas 1), dengan precision sebesar 0.83, recall 0.94, dan f1-score 0.88. Namun, performa pada kelas minoritas (kelas 0) masih rendah dengan precision hanya 0.50, recall 0.25, dan f1-score 0.33, yang menunjukkan bahwa model ini cenderung mengabaikan kelas minoritas. Nilai rata-rata makro untuk f1-score sebesar 0.61 mengindikasikan adanya ketidakseimbangan kinerja antar kelas.

Model CatBoost berada di posisi kedua dengan akurasi 75%. Performa untuk kelas 1 juga cukup baik, dengan f1-score 0.84, precision 0.87, dan recall 0.81. Sedangkan untuk kelas 0, nilai precision dan recall masing-masing 0.40 dan 0.50, dengan f1-score sebesar 0.44. Nilai f1-score makro sebesar 0.64 menunjukkan bahwa model ini sedikit lebih baik dalam menjaga keseimbangan antar kelas dibandingkan Random Forest, meskipun dengan akurasi yang sedikit lebih rendah.

Model Decision Tree memiliki performa terendah dengan akurasi hanya 60%. F1-score untuk kelas 1 adalah 0.67, sementara untuk kelas 0 hanya 0.50. Metrik rata-rata makro untuk f1-score sebesar 0.58 menandakan bahwa model ini paling tidak stabil dalam mengenali kedua kelas secara seimbang.

Berdasarkan hasil evaluasi di atas, dapat disimpulkan bahwa Random Forest unggul dalam akurasi dan kinerja pada kelas mayoritas, namun cenderung mengabaikan kelas minoritas. CatBoost menjadi alternatif yang lebih seimbang,

meskipun dengan akurasi sedikit lebih rendah. Sementara itu, Decision Tree kurang direkomendasikan sebagai model utama karena performanya yang relatif rendah dan kurang stabil. Oleh karena itu, pemilihan model sebaiknya disesuaikan dengan kebutuhan spesifik, apakah fokus pada akurasi umum atau keseimbangan deteksi antar kelas.

4.4 Analisis Keunggulan dan Keterbatasan

Berdasarkan hasil evaluasi, model Random Forest menunjukkan performa yang paling tinggi dibandingkan model lainnya, dengan akurasi **sebesar 80%**. Keunggulan utama dari Random Forest terletak pada kemampuannya dalam menangani data yang kompleks dan menghasilkan prediksi yang cukup andal untuk kelas mayoritas (kelas 1). Hal ini dibuktikan dengan nilai recall sebesar 0.94 dan f1-score sebesar 0.88 pada kelas tersebut, yang menunjukkan bahwa model mampu mendeteksi hampir seluruh data kelas 1 dengan tingkat kesalahan yang rendah.

Selain itu, Random Forest juga dikenal sebagai algoritma yang tahan terhadap overfitting, karena terdiri dari gabungan banyak pohon keputusan (decision trees) yang membuat keputusan berdasarkan voting. Model ini juga mampu menangani fitur dengan skala dan jenis yang berbeda, serta bekerja cukup baik meskipun terdapat data yang tidak bersih atau sedikit noise.

Namun demikian, keterbatasan utama dari model ini adalah performanya yang masih lemah terhadap kelas minoritas (kelas 0). Hal ini terlihat dari nilai recall yang hanya 0.25 dan f1-score sebesar 0.33 pada kelas 0, yang berarti sebagian besar data kelas 0 gagal dikenali dengan baik oleh model. Ketidakseimbangan data menjadi salah satu faktor penyebab utama, di mana model cenderung lebih fokus mempelajari pola dari kelas yang lebih dominan (kelas 1), dan mengabaikan pola dari kelas yang jumlahnya sedikit.

Selain itu, dari sisi komputasi, Random Forest membutuhkan sumber daya yang lebih besar dibanding Decision Tree tunggal, terutama ketika jumlah pohon yang digunakan cukup banyak. Proses pelatihan dan prediksi bisa menjadi lambat jika diterapkan pada dataset yang sangat besar.

BAB V KESIMPULAN DAN SARAN

5.1 Ringkasan Temuan Utama

Penelitian ini bertujuan untuk memprediksi keterlambatan pengiriman barang pada layanan e-commerce di Indonesia menggunakan algoritma *machine learning*, khususnya **Random Forest Classifier**. Berdasarkan analisis terhadap 103 data transaksi pengiriman dari dua platform e-commerce, ditemukan beberapa temuan penting sebagai berikut:

1. Mayoritas pengiriman berada dalam kategori jarak dekat, yang menunjukkan bahwa distribusi logistik lebih sering dilakukan dalam skala regional. Hal ini berdampak pada tingkat keterlambatan yang relatif lebih rendah pada wilayah tertentu.
2. Jasa ekspedisi dan lokasi pengirim berpengaruh besar terhadap keterlambatan pengiriman. Dari visualisasi, terlihat bahwa beberapa ekspedisi dan daerah asal pengiriman memiliki tingkat keterlambatan yang lebih tinggi dibanding lainnya, menandakan potensi masalah operasional atau infrastruktur di area tersebut.
3. Model Random Forest menunjukkan performa terbaik dalam klasifikasi keterlambatan, dengan akurasi mencapai 80% dan f1-score sebesar 0.88 pada kelas mayoritas (tepat waktu). Namun, model ini masih lemah dalam mengenali kelas minoritas (terlambat) dengan f1-score hanya **0.33**, akibat distribusi data yang tidak seimbang.
4. Model CatBoost memberikan hasil yang seimbang antar kelas, meskipun akurasinya sedikit lebih rendah (75%), sedangkan Decision Tree memiliki akurasi terendah (**60%**) dan performa yang kurang stabil.
5. Fitur yang paling berpengaruh terhadap keterlambatan pengiriman adalah jarak tempuh, jenis jasa ekspedisi, dan lokasi pengirim. Hal ini sejalan dengan hasil korelasi numerik dan analisis feature importance pada model Random Forest.

Secara keseluruhan, penelitian ini berhasil membangun model prediksi yang dapat membantu perusahaan e-commerce dalam mengidentifikasi risiko keterlambatan serta menyediakan dasar bagi pengambilan keputusan strategis dalam pengelolaan logistik.

5.2 Jawaban atas Rumusan Masalah

1. Apa saja faktor yang mempengaruhi keterlambatan pengiriman barang pada e-commerce di Indonesia? Berdasarkan hasil eksplorasi data dan analisis model, ditemukan bahwa beberapa faktor logistik berpengaruh signifikan terhadap keterlambatan pengiriman, antara lain: jarak tempuh pengiriman, lokasi pengirim, dan jenis jasa ekspedisi yang digunakan. Jarak yang lebih jauh cenderung meningkatkan risiko keterlambatan, sementara lokasi pengiriman tertentu menunjukkan tren keterlambatan lebih tinggi akibat kemungkinan kendala infrastruktur atau operasional. Selain itu, performa tiap ekspedisi berbeda-beda, di mana beberapa mitra logistik lebih sering mengalami keterlambatan.
2. Bagaimana algoritma Random Forest dapat digunakan untuk memprediksi keterlambatan pengiriman barang? Algoritma Random Forest digunakan sebagai model klasifikasi untuk memprediksi apakah suatu pengiriman akan terlambat atau tepat waktu. Model ini dilatih menggunakan data historis pengiriman yang mencakup atribut-atribut logistik seperti jarak, lokasi, tanggal, kendaraan, dan ekspedisi. Dengan teknik *ensemble learning*, Random Forest menghasilkan akurasi yang tinggi (80%) dan performa prediksi yang kuat pada kelas mayoritas. Model juga memberikan informasi penting

- tentang fitur-fitur yang paling berpengaruh terhadap prediksi keterlambatan.
3. Apa rekomendasi perbaikan yang dapat diusulkan berdasarkan hasil analisis data? Berdasarkan hasil analisis, beberapa rekomendasi yang dapat diajukan antara lain:
 - a) Memprioritaskan perbaikan proses pengiriman pada lokasi-lokasi dengan tingkat keterlambatan tinggi.
 - b) Melakukan evaluasi performa jasa ekspedisi, terutama yang memiliki tingkat keterlambatan signifikan.
 - c) Mengoptimalkan pengelolaan logistik untuk pengiriman jarak jauh, misalnya melalui perencanaan rute dan penguatan jaringan distribusi.
 - d) Menerapkan teknik penyeimbangan data atau *resampling* dalam model prediksi selanjutnya untuk meningkatkan akurasi pada kelas minoritas.
 - e) Mengembangkan sistem prediksi real-time untuk mendukung pengambilan keputusan operasional secara cepat dan responsif.

5.3 Saran untuk Pengembangan Lanjut

Berdasarkan hasil penelitian yang telah dilakukan, terdapat beberapa saran yang dapat dipertimbangkan untuk pengembangan penelitian di masa mendatang:

1. Menggunakan Dataset yang Lebih Besar dan Beragam
Penelitian ini menggunakan 103 data transaksi yang relatif terbatas. Untuk mendapatkan model yang lebih general dan akurat, disarankan menggunakan dataset yang lebih besar, mencakup lebih banyak platform e-commerce dan rentang waktu pengiriman yang lebih luas.
2. Mengintegrasikan Faktor Eksternal
Faktor-faktor seperti cuaca, kemacetan, hari libur, dan kondisi infrastruktur belum dianalisis secara langsung karena keterbatasan data. Pada penelitian selanjutnya, penggabungan data eksternal ini dapat memberikan gambaran yang lebih lengkap terhadap penyebab keterlambatan.
3. Menerapkan Penyeimbangan Data (Resampling Techniques)
Karena distribusi kelas pada data tidak seimbang, performa model terhadap kelas minoritas (terlambat) masih rendah. Teknik seperti SMOTE, ADASYN, atau class-weight adjustment dapat digunakan agar model lebih adil dalam mengenali kedua kelas.
4. Eksperimen dengan Model Lain dan Ensemble Tingkat Lanjut
Selain Random Forest, penelitian di masa depan dapat menguji model lain seperti XGBoost, LightGBM, atau Neural Network, serta mencoba teknik ensemble yang lebih kompleks untuk meningkatkan akurasi dan ketahanan model.
5. Pengembangan Sistem Prediksi Real-Time
Implementasi model prediksi ke dalam sistem informasi logistik secara real-time dapat memberikan manfaat praktis bagi perusahaan. Sistem ini dapat membantu dalam mengantisipasi keterlambatan lebih awal dan menyusun solusi pengiriman alternatif secara dinamis.
6. Penjelasan Model dengan Teknik Interpretabilitas
Untuk meningkatkan kepercayaan dan transparansi dalam penggunaan model, disarankan menggunakan teknik interpretasi seperti SHAP (SHapley Additive exPlanations) atau LIME untuk memahami alasan di balik setiap prediksi.

DAFTAR REFERENSI

- [1] H. A. P. Damayanti, H. Kurniawan, and H. Mayatopani, "ANALISIS SISTEM PENJUALAN DAN CUSTOMER RELATIONSHIP MANAGEMENT (CRM) PADA APLIKASI SHOPEE," *IDEALIS : InDonEsiA journal Information System*, vol. 6, no. 2, pp. 75–82, Jul. 2023, doi: 10.36080/idealis.v6i2.3014.
- [2] D. Sulistianingsih, M. D. Utami, and Y. P. Adhi, "Perlindungan Hukum bagi Konsumen dalam Transaksi E-commerce sebagai Tantangan Bisnis di Era Global," *JURNAL MERCATORIA*, vol. 16, no. 2, pp. 119–128, Dec. 2023, doi: 10.31289/mercatoria.v16i2.8042.
- [3] A. A. Musyafah, H. W. Khasna, and B. E. Turisno, "PERLINDUNGAN KONSUMEN JASA PENGIRIMAN BARANG DALAM HAL TERJADI KETERLAMBATAN PENGIRIMAN BARANG," *LAW REFORM*, vol. 14, no. 2, p. 151, Sep. 2018, doi: 10.14710/lr.v14i2.20863.
- [4] A. Turseno, A. Nadia, S. N. Aji, R. H. A. Tanisri, and H. Kurnia, "Implementasi Lean Manufacturing (LM) pada Proses Pengiriman Makanan ke Pelanggan," *LOGISTIK*, vol. 17, no. 02, pp. 137–154, Nov. 2024, doi: 10.21009/logistik.v17i02.49368.
- [5] H. A. P. Damayanti, H. Kurniawan, and H. Mayatopani, "ANALISIS SISTEM PENJUALAN DAN CUSTOMER RELATIONSHIP MANAGEMENT (CRM) PADA APLIKASI SHOPEE," *IDEALIS : InDonEsiA journal Information System*, vol. 6, no. 2, pp. 75–82, Jul. 2023, doi: 10.36080/idealis.v6i2.3014.
- [6] D. Sulistianingsih, M. D. Utami, and Y. P. Adhi, "Perlindungan Hukum bagi Konsumen dalam Transaksi E-commerce sebagai Tantangan Bisnis di Era Global," *JURNAL MERCATORIA*, vol. 16, no. 2, pp. 119–128, Dec. 2023, doi: 10.31289/mercatoria.v16i2.8042.
- [7] U. Cahyadi, D. S. Taptajani, and M. Nurjaman, "Pendekatan Lean Service dengan Metode Value Stream Mapping Untuk Meminimasi Waste di Logistic J&T Express," *Jurnal Kalibrasi*, vol. 17, no. 2, pp. 78–85, Jun. 2020, doi: 10.33364/kalibrasi/v.17-2.698.
- [8] A. Turseno, A. Nadia, S. N. Aji, R. H. A. Tanisri, and H. Kurnia, "Implementasi Lean Manufacturing (LM) pada Proses Pengiriman Makanan ke Pelanggan," *LOGISTIK*, vol. 17, no. 02, pp. 137–154, Nov. 2024, doi: 10.21009/logistik.v17i02.49368.
- [9] Z. Bai *et al.*, "Clinical Feature-Based Machine Learning Model for 1-Year Mortality Risk Prediction of ST-Segment Elevation Myocardial Infarction in Patients with Hyperuricemia: A Retrospective Study," *Comput Math Methods Med*, vol. 2021, pp. 1–9, Jul. 2021, doi: 10.1155/2021/7252280.
- [10] ج. خضاری مهدی محمود, "Comparison between Machine Learning Algorithms for Cardiovascular Disease Prediction," *□□□□□□□□ □□□□□□□□ □□□□□□□□ □□□□□□□□*, vol. 4, no. 1, pp. 909–926, Jan. 2023, doi: 10.21608/cfdj.2023.258074.
- [11] P. Kumar, S. Ambekar, M. Kumar, and S. Roy, "Analytical Statistics Techniques of Classification and Regression in Machine Learning," in *Data Mining - Methods, Applications and Systems*, IntechOpen, 2021. doi: 10.5772/intechopen.84922.
- [12] H. Karlı and M. Tanyaş, "Innovative Delivery Methods in the Last-Mile: Unveiling Consumer Preference," *Future Transportation*, vol. 4, no. 1, pp. 152–173, Feb. 2024, doi: 10.3390/futuretransp4010009.
- [13] J. Tapia, P. Fariña, I. Urbina, and D. Dujovne, "Examining the Retail Delivery Choice Behavior in a Technology-Aware Market," *Journal of Theoretical and Applied Electronic Commerce Research*, vol. 19, no. 2, pp. 1392–1410, Jun. 2024, doi: 10.3390/jtaer19020070.
- [14] P. Klein and B. Popp, "Last-Mile Delivery Methods in E-Commerce: Does Perceived Sustainability Matter for Consumer Acceptance and Usage?," *Sustainability*, vol. 14, no. 24, p. 16437, Dec. 2022, doi: 10.3390/su142416437.
- [15] R. Merkert, M. C. J. Bliemer, and M. Fayyaz, "Consumer preferences for innovative and traditional last-mile parcel delivery," *International Journal of Physical Distribution & Logistics Management*, vol. 52, no. 3, pp. 261–284, Mar. 2022, doi: 10.1108/IJPDLM-01-2021-0013.

- [16] S. Gharaati and R. Dziedzic, "Important Factors For Water Main Break Prediction Across 13 Canadian Systems," in *Proceedings - 2nd International Joint Conference on Water Distribution System Analysis (WDSA)& Computing and Control in the Water Industry (CCWI)*, València: Editorial Universitat Politècnica de València, Jul. 2022. doi: 10.4995/WDSA-CCWI2022.2022.14755.
- [17] M. Taheriboshrouyeh, M. Sandanayake, and S. Fragomeni, "A Review of Construction Program Delivery Attributes: Bibliometric Analysis of Two Decades," *Buildings*, vol. 13, no. 10, p. 2664, Oct. 2023, doi: 10.3390/buildings13102664.
- [18] I. Z. P. Hamdan and M. Othman, "Predicting Customer Loyalty Using Machine Learning for Hotel Industry," *Journal of Soft Computing and Data Mining*, vol. 3, no. 2, Aug. 2022, doi: 10.30880/jscdm.2022.03.02.004.
- [19] A. J. Silva and P. Cortez, "An Automated Machine Learning Approach for Predicting Chemical Laboratory Material Consumption," 2021, pp. 105–116. doi: 10.1007/978-3-030-79150-6_9.
- [20] J. BRZOZOWSKA, J. PIZOŃ, G. BAYTIKENOVA, A. GOLA, A. ZAKIMOVA, and K. PIOTROWSKA, "DATA ENGINEERING IN CRISP-DM PROCESS PRODUCTION DATA – CASE STUDY," *Applied Computer Science*, vol. 19, no. 3, pp. 83–95, Sep. 2023, doi: 10.35784/acs-2023-26.
- [21] M. Cazacu and E. Titan, "Adapting CRISP-DM for Social Sciences," *Brain (Bacau)*, vol. 11, no. 2sup1, pp. 99–106, Aug. 2020, doi: 10.18662/brain/11.2Sup1/97.
- [22] F. R. Kurniawan and R. Sutomo, "FORECASTING RICE INVENTORY IN INDONESIA USING THE ARIMA ALGORITHM METHOD," *Journal of Multidisciplinary Issues*, vol. 1, no. 2, pp. 1–12, Aug. 2021, doi: 10.53748/jmis.v1i2.15.
- [23] R. Clancy, D. O'Sullivan, and K. Bruton, "Data-driven quality improvement approach to reducing waste in manufacturing," *The TQM Journal*, vol. 35, no. 1, pp. 51–72, Jan. 2023, doi: 10.1108/TQM-02-2021-0061.
- [24] S. Tripathi, D. Muhr, M. Brunner, H. Jodlbauer, M. Dehmer, and F. Emmert-Streib, "Ensuring the Robustness and Reliability of Data-Driven Knowledge Discovery Models in Production and Manufacturing," *Front Artif Intell*, vol. 4, Jun. 2021, doi: 10.3389/frai.2021.576892.
- [25] I. Hidayat, A. I. Tolago, R. D. R. Dako, and J. Ilham, "Analisis Data Eksploratif Capaian Indikator Kinerja Utama 3 Fakultas Teknik," *Jambura Journal of Electrical and Electronics Engineering*, vol. 5, no. 2, pp. 185–191, Jul. 2023, doi: 10.37905/jjee.v5i2.18397.