PREDIKSI KETERLAMBATAN PENGIRIMAN BARANG

Dibuat Untuk Memenuhi Tugas Besar Mata Kuliah Data Mining

Disusun Oleh :

Dewi Kresnawati	**(714220002)**
Rania Ayuni Kartini Fitri	**(714220032)**
Audyardha Nasywa Andini	**(714220020)**

PROGRAM STUDI DIV TEKNIK INFORMATIKA

SEKOLAH VOKASI

UNIVERSITAS LOGISTIK & BISNIS INTERNASIONAL

BANDUNG

TAHUN AJARAN 2025

HALAMAN PERNYATAAN 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 : .....

NIM : .....

Tanda Tangan : .....

Tanggal : .....

Mengetahui

Ketua :..... (.....tanda tangan.....)

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

KATA PENGANTAR

Puji syukur kami panjatkan ke hadirat Tuhan Yang Maha Esa atas rahmat dan karunia-Nya sehingga kami dapat menyelesaikan Tugas Besar Mata Kuliah Data Mining ini dengan baik. Penyusunan tugas besar ini merupakan bagian dari pemenuhan pembelajaran di Program Studi Sarjana Terapan Teknik Informatika, Universitas Logistik dan Bisnis Internasional (ULBI) Bandung.

Kami menyadari bahwa tersusunnya tugas besar ini tidak lepas dari dukungan, bimbingan, serta doa dari berbagai pihak sejak awal perkuliahan hingga proses penyelesaian tugas ini. Oleh karena itu, kami mengucapkan terima kasih sebesar-besarnya kepada:

1. Nisa Hanum Harani, S.Kom., M.T., CDSP, SFPC, selaku dosen pengampu mata kuliah Data Mining yang telah memberikan bimbingan dan arahan selama proses pengerjaan tugas besar ini.
2. Roni Andarsyah, S.T., M.Kom., SFPC, selaku Ketua Program Studi D4 Teknik Informatika.
3. Orang tua dan keluarga kami yang telah memberikan dukungan moral maupun material yang sangat berarti.
4. Teman-teman dan sahabat seperjuangan, yang telah memberikan semangat dan bantuan selama proses penyusunan tugas ini.

Akhir kata, kami berharap Tuhan Yang Maha Esa membalas segala kebaikan semua pihak yang telah membantu. Semoga tugas besar ini dapat memberikan kontribusi positif dalam pengembangan ilmu pengetahuan, khususnya di bidang data mining, serta bermanfaat bagi pihak-pihak yang berkepentingan.

Bandung, 10 Juli 2025

HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR UNTUK KEPENTINGAN AKADEMIS

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

Nama : Dewi Kresnawati

NIM : 714220002

Demi 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:

.....

.....

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 : 10 Juli 2025

Yang menyatakan

(Dewi Kresawati)

HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR UNTUK KEPENTINGAN AKADEMIS

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

Nama : Audyardha Nasywa Andini

NIM : 714220020

Demi 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:

.....

.....

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 : 10 Juli 2025

Yang menyatakan

(Audyardha Nasywa Andini)

HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR UNTUK KEPENTINGAN AKADEMIS

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

Nama : Rania Ayuni Kartini Fitri

NIM : 714220032

Demi 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:

.....

.....

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 : 10 Juli 2025

Yang menyatakan

(Rania Ayuni Kartini Fitri)

ABSTRAK

Keterlambatan pengiriman barang merupakan salah satu permasalahan yang sering terjadi dalam proses logistik dan distribusi, yang dapat berdampak pada kepuasan pelanggan, efisiensi operasional, serta reputasi perusahaan. Dalam era digital, banyak data dapat dikumpulkan dari jejak interaksi pengguna, seperti komentar konsumen di platform online, yang dapat dimanfaatkan untuk menggali pola-pola tertentu menggunakan metode data mining.

Penelitian ini dilakukan sebagai bagian dari Tugas Besar mata kuliah Data Mining, dengan tujuan untuk membangun model prediksi keterlambatan pengiriman barang berdasarkan data historis dan opini pelanggan. Proses penelitian melibatkan tahapan pengumpulan data primer, preprocessing data, analisis sentimen, serta penerapan algoritma klasifikasi seperti Decision Tree dan Random Forest. Hasil yang diperoleh menunjukkan bahwa algoritma data mining mampu memberikan prediksi yang cukup akurat terhadap kemungkinan keterlambatan pengiriman.

Model yang dikembangkan diharapkan dapat membantu perusahaan dalam mengidentifikasi potensi keterlambatan sejak dini dan mendukung pengambilan keputusan yang lebih proaktif dalam mengelola pengiriman barang secara tepat waktu dan efisien.

Kata kunci: prediksi keterlambatan, data mining, klasifikasi, pengiriman barang, analisis sentimen

ABSTRACT

Delivery delays are one of the common issues in logistics and distribution processes, often affecting customer satisfaction, operational efficiency, and the overall reputation of a company. In the digital era, a large amount of data can be collected from user interactions, such as customer comments on online platforms, which can be utilized to uncover meaningful patterns using data mining techniques.

This study was conducted as part of a final project for the Data Mining course, with the aim of developing a predictive model for delivery delays based on historical data and customer opinions. The research process involved collecting primary data, performing preprocessing, conducting sentiment analysis, and applying classification algorithms such as Decision Tree and Random Forest. The results indicate that data mining algorithms can provide accurate predictions regarding the likelihood of delivery delays.

The predictive model is expected to assist companies in early identification of potential delays and support more proactive decision-making to ensure timely and efficient delivery operations.

***Keywords**: delivery delay prediction, data mining, classification, shipment, sentiment analysis*

.

DAFTAR ISI

BAB I
PENDAHULUAN

1.1 Latar Belakang

Ketepatan waktu dalam pengiriman barang merupakan aspek fundamental dalam industri logistik modern [1]. Tingginya ekspektasi pelanggan terhadap kecepatan dan keakuratan layanan menuntut perusahaan untuk terus meningkatkan kualitas sistem distribusinya [2]. Sayangnya, realita di lapangan menunjukkan bahwa keterlambatan pengiriman masih menjadi kendala yang sering terjadi, baik pada sektor manufaktur, ritel, maupun penyedia jasa logistik pihak ketiga (3PL). Permasalahan ini tidak hanya berdampak pada konsumen secara langsung, tetapi juga memengaruhi reputasi perusahaan, meningkatkan biaya operasional, serta menurunkan loyalitas pelanggan secara signifikan [3] .

Faktor-faktor yang menyebabkan keterlambatan pengiriman sangat beragam dan kompleks, mencakup gangguan operasional internal seperti keterbatasan armada dan ketidakefisienan penjadwalan, hingga faktor eksternal seperti kondisi cuaca ekstrem, kemacetan lalu lintas, jarak tempuh, jenis layanan yang dipilih, serta karakteristik dari barang yang dikirim [4]. Kompleksitas tersebut menunjukkan bahwa keterlambatan bukanlah akibat dari satu penyebab tunggal, melainkan hasil dari interaksi berbagai elemen dalam rantai pasok [5]. Oleh karena itu, diperlukan pendekatan yang berbasis data dan analisis prediktif untuk membantu perusahaan dalam mengantisipasi potensi keterlambatan sebelum pengiriman terjadi.

Penelitian ini bertujuan untuk membangun sebuah model *machine learning* yang mampu memprediksi kemungkinan keterlambatan pengiriman barang berdasarkan atribut-atribut logistik yang relevan. Model prediktif ini diharapkan dapat memberikan wawasan kepada perusahaan logistik dalam pengambilan keputusan operasional secara proaktif, seperti menentukan rute alternatif, memilih jenis layanan yang lebih cepat, atau mengalokasikan armada dan kurir secara lebih strategis [6]. Untuk menunjang pengembangan model tersebut, data dikumpulkan dari berbagai sumber terkait aktivitas operasional logistik, termasuk data transaksi dari *platform e-commerce* dan sistem manajemen logistik internal perusahaan [7]. Dengan pendekatan ini, penelitian diharapkan mampu memberikan kontribusi nyata dalam meningkatkan efisiensi distribusi serta menciptakan pengalaman pelanggan yang lebih baik melalui ketepatan waktu pengiriman [8].

1.2 Rumusan Masalah

Berdasarkan uraian pada latar belakang, maka rumusan masalah yang dapat diidentifikasi dalam penelitian ini meliputi:

1. Mengapa keterlambatan pengiriman barang masih sering terjadi meskipun sistem distribusi logistik terus ditingkatkan?
2. Faktor-faktor logistik apa saja yang paling berpengaruh terhadap keterlambatan pengiriman barang dalam konteks operasional logistik modern?
3. Bagaimana pemanfaatan model machine learning dapat membantu memprediksi potensi keterlambatan pengiriman secara akurat dan proaktif?

1.3 Tujuan Penelitian

Tujuan dari penelitian ini sebagai berikut:

1. Menganalisis penyebab utama keterlambatan pengiriman barang berdasarkan data logistik dari berbagai sumber.
2. Membangun model prediktif berbasis *machine learning* untuk mengidentifikasi potensi keterlambatan pengiriman sebelum proses distribusi dilakukan.
3. Memberikan rekomendasi strategi operasional berbasis data yang dapat digunakan perusahaan logistik untuk mengurangi risiko keterlambatan.

1.4 Manfaat Penelitian

Manfaat dari penelitian ini sebagai berikut:

1. Membantu perusahaan logistik memahami faktor-faktor penyebab keterlambatan secara komprehensif, baik dari sisi internal maupun eksternal.
2. Menyediakan alat bantu berbasis machine learning untuk mendukung pengambilan keputusan operasional secara cepat dan tepat dalam menghadapi potensi keterlambatan.
3. Berkontribusi dalam peningkatan kepuasan pelanggan dan efisiensi distribusi melalui pemanfaatan teknologi prediktif dalam sistem logistik modern.

1.5 Ruang Lingkup

Penelitian ini dibatasi pada ruang lingkup sebagai berikut:

1. Fokus utama penelitian adalah membangun model prediksi keterlambatan pengiriman barang menggunakan algoritma *machine learning* berbasis data logistik historis.
2. Data yang digunakan mencakup atribut-atribut logistik seperti:
    a. Waktu pengambilan dan pengiriman
    b. Lokasi asal dan tujuan
    c. Jenis layanan
    d. Jenis barang yang dikirim
    e. Jarak tempuh, dan
    f. Variabel eksternal
3. Objek penelitian terbatas pada sektor logistik modern yang terlibat dalam *e-commerce* dan jasa pengiriman (termasuk 3PL/logistik pihak ketiga).
4. Penelitian tidak mencakup aspek non-logistik, seperti keputusan manajemen strategis, performa individu kurir, atau kondisi internal perusahaan secara keseluruhan.
5. Model prediktif yang dikembangkan hanya diuji secara eksperimental menggunakan data historis, tanpa implementasi langsung ke dalam sistem operasional *real-time* perusahaan.
6. Penelitian ini hanya sampai pada pengembangan dan evaluasi model prediksi, serta penyusunan rekomendasi operasional yang bersifat teoritis dan belum diuji langsung di lapangan.

BAB II
LANDASAN TEORI

2.1 *State Of The Art*

Permasalahan keterlambatan pengiriman barang telah menjadi salah satu isu utama dalam industri logistik modern. Sejumlah penelitian telah dilakukan untuk membangun model prediktif berbasis data guna mengidentifikasi potensi keterlambatan, terutama dengan pendekatan *machine learning* dan *data mining*. Berikut ini adalah beberapa penelitian terkini yang relevan:

a. Hardian Kokoh Pambudi et al. (2020) mengidentifikasi bahwa keterlambatan pengiriman dapat mempengaruhi kepuasan pelanggan dan biaya logistik, dengan rata-rata keterlambatan mencapai 27% dari total aktivitas bisnis. Penelitian ini menerapkan tiga metode *machine learning*, yaitu regresi logistik, *random forest*, dan *artificial neural network* (ANN), untuk memprediksi status pengiriman barang. Hasil penelitian menunjukkan bahwa metode *random forest* menghasilkan akurasi tertinggi sebesar 76,6% yang menunjukkan keunggulannya dalam mengatasi *overfitting* dan memberikan hasil yang lebih akurat dibandingkan dengan metode lainnya [9].

b. Pada penelitian yang dilakukan oleh Arif Rinaldi Dikananda et al. (2022) mengenai implementasi data mining pada ketepatan pengiriman barang mengungkapkan bahwa pengiriman barang yang tepat waktu sangat penting namun masih banyak perusahaan yang masih menggunakan metode manual yang mengakitkan kurangnya akurasi informasi. Penelitian ini menerapkan *Algoritma K-Nearest Neighbors* (KNN) untuk mengklasifikasikan dan mengelompokkan data keterlambatan pengiriman barang, dengan menggunakan aplikasi *machine learning RapidMiner*. Variabel yang dianalisis mencakup data pengiriman, keterlambaran, dan kategori ketepatan waktu pengiriman. Hasil penelitian menunjukkan bajwa metode KNN dapat mencapai akurasi 88,64%, yang menandakan kunggulannya dalam memberikan informasi akurat mengenai ketepatan pengiriman barang [10].

c. Moh. Alfian Firmansyah dan Mochammad Machlul Alamin (2024) mengembangkan sistem prediksi menggunakan metode *Naive Bayes* untuk mengklasifikasikan apakah pengiriman barang akan tepat waktu atau terlambar. Variabel yang dianalisis mencakup nama penerima, tujuan pengiriman, jenis kendaraan, hari, cuaca, volume paket, dan berat paket. Hasil penelitian menunjukkan bahwa sistem ini mencapai

akurasi 81,8% yang menandakan keunggulannya dalam memberikan prediksi akurat dengan data historis yang terbatas [11].

d. Dalam penelitian mengenai prediksi keterlambatan pengiriman di PT.X, Keshia Karina Mulia dan Iwan Halim Sahputra (2021) mengidentifikasi bahwa perusahaan mengalami masalah keterlambatan pengiriman kontainer kosong, dengan skor *On Time Performance* (OTP) yang hanya mencapai 71%, jauh di bawah target 95%. Penelitian ini menerapkan metode *Decision Tree Regression* untuk memprediksi keterlambatan berdasarkan data historis, dengan variabel yang mencakup waktu tempuh, status order, ukuran kontainer, grade kontainer, waktu pick up customer, dan lokasi depot. Hasil penelitian menunjukkan bahwa model dengan max depth 2 menghasilkan RMSE terkecil sebesar 7,021, meskipun masih jauh dari angka 0, yang menunjukkan bahwa model ini masih *underfitting* [12].
e. Muhammad Reza dan Suprayogi (2017) mengidentifikasi bahwa ketidakakuratan dalam prediksi waktu pengiriman di PT. Pos Indonesia dapat mengakibatkan ketidakpuasan pelanggan dan kerugian bagi perusahaan. Penelitian ini menerapkan algoritma *Backpropagation* dalam Jaringan Syaraf Tiruan untuk memprediksi waktu pengiriman barang, dengan variabel yang mencakup lokasi barang, kantor asal dan tujuan, jenis angkutan, serta lama pengiriman. Hasil penelitian menunjukkan bahwa algoritma ini dapat mencapai nilai *error* sebesar 2,1111%, yang menandakan tingkat akurasi yang tinggi dalam memprediksi waktu pengiriman [13].

Dalam beberapa tahun terakhir, penelitian di bidang *data mining* dan *machine learning* untuk prediksi keterlambatan pengiriman barang menunjukkan perkembangan yang pesat. Pergeseran signifikan terjadi dari penggunaan model-model tradisional seperti *Naïve Bayes* dan *K-Nearest Neighbor*, ke model yang lebih kompleks dan akurat seperti ensemble methods (contohnya *Random Forest* dan *XGBoost*) serta *neural networks*. Selain itu, pendekatan *AutoML* (*Automated Machine Learning***)** mulai banyak digunakan untuk mengotomatisasi proses pemilihan dan tuning model, sehingga mempercepat pengembangan sistem prediktif dalam skala besar.

Meski demikian, tantangan seperti ketidakseimbangan data masih menjadi isu umum dalam penelitian ini, karena data pengiriman yang terlambat umumnya jauh lebih sedikit dibandingkan yang tepat waktu. Untuk mengatasi masalah ini, mengandalkan pemilihan

fitur yang relevan serta pemodelan klasifikasi yang andal untuk meminimalkan bias terhadap kelas mayoritas.

Dengan menggunakan pendekatan klasifikasi berbasis algoritma seperti *Decision Tree* dan *Random Forest*, penelitian ini bertujuan untuk menghasilkan model prediksi keterlambatan yang akurat dan tetap dapat diinterpretasikan dengan baik oleh pihak operasional logistik. Pendekatan ini diharapkan mampu memberikan wawasan yang berguna dalam meningkatkan efisiensi proses pengiriman tanpa harus menambahkan kompleksitas dalam tahap *preprocessing data*.

2.2 Tinjauan Pustaka

Tinjauan pustaka bertujuan untuk memberikan landasan teoritis yang kuat terhadap penelitian ini, yang berfokus pada penerapan teknik data mining dalam memprediksi keterlambatan pengiriman barang. Subbab ini akan membahas secara sistematis teori-teori yang relevan, mulai dari pengertian data mining, algoritma yang digunakan, atribut-atribut penting dalam logistik, serta penelitian terdahulu yang mendasari pengembangan model prediktif.

2.2.1. *Data Mining*

Menurut Yuli Mardi, data mining adalah proses menemukan pola yang berguna dan tersembunyi dari kumpulan data yang besar [14]. Dalam konteks logistik, data mining digunakan untuk menggali pola keterlambatan pengiriman berdasarkan data historis seperti rute, waktu tempuh, cuaca, dan volume kiriman. Teknik ini dapat membantu mengidentifikasi faktor-faktor utama penyebab keterlambatan [15].

2.2.2. *Machine Learning*

Machine learning adalah cabang dari kecerdasan buatan yang memungkinkan sistem belajar dari data dan membuat prediksi tanpa diprogram secara eksplisit [16]. Dalam prediksi keterlambatan pengiriman, machine learning dapat digunakan untuk membuat model yang mempelajari pola dari data masa lalu dan memberikan estimasi kemungkinan keterlambatan pada pengiriman berikutnya [17].

2.2.3. Prediksi (Prediction)

Prediksi dalam data mining adalah salah satu metode klasifikasi atau regresi yang bertujuan untuk memperkirakan nilai atau status suatu objek berdasarkan data historis [18]. Dalam kasus keterlambatan pengiriman, prediksi digunakan untuk memperkirakan apakah suatu barang akan sampai tepat waktu atau mengalami

keterlambatan berdasarkan variabel seperti cuaca, jarak tempuh, rute, dan jenis kendaraan [19].

2.2.4. Keterlambatan Pengiriman Barang

Keterlambatan pengiriman merujuk pada kondisi di mana barang tidak sampai pada tujuan sesuai dengan waktu yang telah dijadwalkan. Faktor-faktor seperti cuaca, lalu lintas, kondisi kendaraan, human error, serta sistem logistik yang tidak efisien dapat menjadi penyebab keterlambatan [20]. Keterlambatan ini dapat berdampak pada kepuasan pelanggan dan biaya operasional.

2.2.5. *Supervised Learning*

Supervised learning adalah teknik *machine learning* di mana model dilatih menggunakan data *input* dan *output* yang telah diketahui. Teknik ini sangat umum dalam kasus klasifikasi (terlambat atau tidak terlambat) dan regresi (berapa lama keterlambatannya) [21]. Model seperti *Decision Tree*, *Random Forest*, dan *Logistic Regression* merupakan bagian dari *supervised learning*.

2.2.6. Klasifikasi

Klasifikasi merupakan sebuah teknik dalam supervised learning yang digunakan untuk memetakan input ke dalam kelas output yang telah ditentukan sebelumnya [sitasi]. Dalam konteks prediksi keterlambatan, klasifikasi bertujuan untuk mengidentifikasi apakah suatu pengiriman barang akan terlambat atau tidak terlambat, berdasarkan variabel-variabel historis seperti waktu tempuh, cuaca, jenis kendaraan, lokasi tujuan, dan volume barang [sitasi].

2.2.7. *Feature Selection* dan *Preprocessing*

Feature selection adalah proses memilih atribut (fitur) paling relevan dari dataset yang akan digunakan untuk pelatihan model, sehingga meningkatkan akurasi dan efisiensi [22]. Data preprocessing mencakup pengolahan data kosong, outlier, normalisasi, serta encoding data kategorikal agar model dapat memahami pola secara optimal [23].

2.2.8. Evaluasi Model

Evaluasi model dilakukan untuk mengukur seberapa baik model prediksi bekerja. Metode evaluasi umum meliputi akurasi, precision, recall, F1-score, RMSE (Root Mean Square Error), dan MAE (Mean Absolute Error) [24]. Dalam prediksi keterlambatan, pemilihan metrik tergantung pada jenis model dan jenis output (klasifikasi atau regresi).

2.2.9. Algoritma yang Digunakan

a. *Random Forest*

Random Forest adalah metode *ensemble learning* berbasis pohon keputusan yang bekerja dengan membuat banyak pohon keputusan dan menggabungkan hasilnya. Metode ini efektif dalam menghindari *overfitting* dan sering digunakan dalam prediksi klasifikasi maupun regresi [25]. Dalam logistik, *Random Forest* digunakan untuk mengklasifikasikan status pengiriman atau memperkirakan waktu kedatangan barang.

b. *HistGradient Boosting Classifier (HGBClassifier)*

HistGradient Boosting Classifier merupakan varian dari *algoritma Gradient Boosting* yang dioptimalkan untuk kecepatan dan efisiensi memori melalui teknik *histogram-based binning* pada fitur numerik [26]. Algoritma ini membangun model secara bertahap dengan meminimalkan *loss function* menggunakan pohon keputusan sebagai estimator lemah. Dalam konteks prediksi keterlambatan pengiriman barang, HGBC sangat efektif karena mampu menangani data berskala besar dan menghasilkan prediksi yang akurat, bahkan pada dataset yang kompleks dan tidak seimbang [27]. Selain itu, kemampuannya dalam mengatur regularisasi membuatnya cocok untuk mencegah overfitting pada sistem logistik yang dinamis [28].

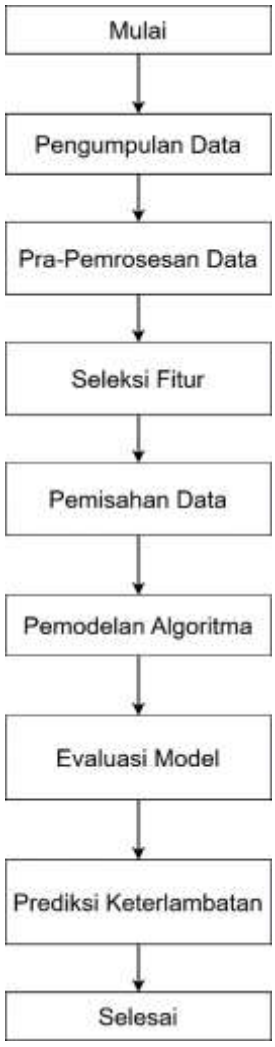
c. *Decision Tree*

Decision Tree adalah model prediktif berbasis struktur pohon, yang membagi data ke dalam cabang-cabang berdasarkan atribut tertentu. Model ini mudah diinterpretasikan dan digunakan secara luas dalam masalah klasifikasi dan regresi [10]. Dalam prediksi keterlambatan, *Decision Tree* memetakan kondisi yang memicu keterlambatan seperti waktu pengambilan, jenis barang, dan lokasi pengiriman [29].

2.2.10. Diagram Alir Konsep Metodologi

Diagram alur (flowchart) merupakan representasi visual dari proses atau tahapan penelitian yang dilakukan dalam membangun sistem prediksi keterlambatan pengiriman barang menggunakan teknik data mining. Diagram ini membantu untuk memahami alur kerja secara menyeluruh, mulai dari tahap awal hingga hasil akhir berupa model prediksi.

Adapun diagram alur dari sistem yang dikembangkan dapat dilihat pada Gambar 2.1 berikut:



Gambar 2. 1 Diagram Alir Konsep Metodologi

BAB III
METODOLOGI PENELITIAN

3.1 Tahapan Penelitian (CRISP-DM)

Penelitian ini menggunakan pendekatan CRISP-DM (*Cross Industry Standard Process for Data Mining*) sebagai kerangka metodologis yang umum diterapkan dalam proyek data mining. Tahapan-tahapan yang dilakukan meliputi:

a. Akuisisi Data

Tahap ini merupakan proses awal dalam penelitian, dimana data dikumpulkan sebagai dasar untuk dilakukan analisis lebih lanjut. Dalam penelitian ini, data diperoleh dari repository *GitHub* yang menyediakan dataset *open-source* mengenai atribut pengiriman barang dan status keterlambatannya. Dataset tersebut telah diunduh dalam format .csv, dan mencakup berbagai fitur seperti waktu pengiriman, jenis kendaraan, jarak, serta status ketepatan waktu pengiriman.

Pengambilan dataset dari *GitHub* dipilih karena:

- Ketersediaan data yang relevan dan realistis.
- Format file yang mudah diolah (CSV) karna data masih mentah.

b. Preprocessing

Data yang diperoleh dari sumber mentah mengalami proses pra-pemrosesan, yang mencakup:

- Penghapusan nilai kosong atau duplikat
- Penyesuaian format waktu dan tanggal
- Normalisasi data numerik
- Encoding untuk data kategorikal (misal: one-hot encoding)
- Tokenisasi dan pembersihan teks pada data opini pelanggan.

c. Feature Engineering

Dari data mentah yang telah dibersihkan, dilakukan pemilihan dan pembuatan fitur (atribut) yang relevan. Contoh fitur yang digunakan:

- Jarak tempuh
- Waktu pengiriman
- Volume paket
- Sentimen pelanggan (positif/negatif/netral)

- Cuaca, jenis kendaraan, dan jam pengiriman.

d. Pemodelan (*Modelling*)

Model dibangun menggunakan teknik klasifikasi berbasis *supervised learning*:

- *Decision Tree*
- *Random Forest*
- *HistGradient Boosting Classifier* untuk pengujian tambahan.

Dataset dibagi menjadi data latih dan data uji dengan perbandingan 80:20 untuk membangun dan mengevaluasi model.

e. Evaluasi

Hasil prediksi dari masing-masing model dievaluasi menggunakan metrik evaluasi standar:

- *Accuracy*: Persentase prediksi benar terhadap total data uji
- *Precision*: Ketepatan model dalam memprediksi keterlambatan
- *Recall*: Kemampuan model menangkap semua kasus keterlambatan
- *F1-score*: Harmonik dari precision dan recall
- (Jika regresi digunakan) RMSE untuk mengukur deviasi prediksi waktu keterlambatan.

3.2 Deskripsi Dataset

a. Sumber Dataset

Dataset diperoleh dari github https://github.com/Abhi1727/FedEx-Logistics-Performance-Analysis/blob/main/SCMS_Delivery_History_Dataset.csv

b. Ukuran Dataset

Jumlah Baris x Kolom sebelum pembersihan: 10324 baris × 33 kolom
Total jumlah data setelah pembersihan: ± 2000 baris
Terdiri dari data terstruktur dan hasil ekstraksi fitur sentimen dari teks

c. Atribut yang Digunakan

- waktu_pengiriman: waktu aktual pengiriman
- rute: asal dan tujuan
- volume_barang: dalam satuan m^3
- jenis_kendaraan: motor, mobil, truk
- cuaca: cerah, hujan, ekstrem

- sentimen: hasil analisis sentimen (positif/negatif/netral)
- label_keterlambatan: target (1 = terlambat, 0 = tepat waktu).

3.3 Algoritma/Data Mining *Tools*

Penelitian ini menggunakan beberapa algoritma klasifikasi dalam proses pemodelan untuk memprediksi kemungkinan keterlambatan pengiriman barang. Algoritma yang digunakan meliputi:

a. *Decision Tree*

Algoritma ini bekerja dengan membuat pemisahan data berdasarkan atribut-atribut tertentu dalam bentuk struktur pohon keputusan. Keunggulannya terletak pada interpretasi hasil yang mudah dipahami, sehingga cocok digunakan untuk menjelaskan alasan keterlambatan kepada pihak operasional.

b. *Random Forest*

Algoritma ensemble learning yang membentuk banyak decision tree secara acak dan menggabungkan hasilnya melalui voting (untuk klasifikasi). Algoritma ini efektif dalam mengurangi overfitting dan menghasilkan akurasi yang lebih stabil dibandingkan decision tree tunggal.

c. *HistGradientBoostingClassifier* (HGBC)

HistGradientBoostingClassifier adalah versi efisien dari algoritma *Gradient Boosting* yang memanfaatkan teknik histogram-based binning, sehingga lebih cepat dalam proses training dan hemat memori. Beberapa alasan dipilihnya algoritma ini:

- Mampu menangani data dalam jumlah besar
- Cocok untuk data yang tidak seimbang (misal: jumlah data keterlambatan jauh lebih sedikit)
- Memiliki pengaturan regularisasi yang baik untuk mencegah *overfitting*.
- Memberikan akurasi prediksi yang tinggi dalam berbagai kasus klasifikasi

  HGBC bekerja dengan membuat model secara iteratif, di mana model baru akan mempelajari kesalahan dari model sebelumnya, sehingga akurasi meningkat secara bertahap.

d. *Tools* yang Digunakan

a) Bahasa pemrograman**:** *Python*

b) Pustaka utama**:**

- *Scikit-learn*: untuk implementasi model klasifikasi dan evaluasi
- *Pandas, NumPy*: untuk manipulasi dan analisis data
- *Matplotlib, Seaborn*: untuk visualisasi
- NLTK / *TextBlob*: untuk analisis sentimen komentar pelanggan.

c) Platform coding**:** *Google Colab*.

3.4 Evaluasi Kinerja

Evaluasi kinerja model merupakan tahapan penting dalam proses data mining untuk mengetahui seberapa baik model dalam memprediksi keterlambatan pengiriman barang. Dalam penelitian ini, digunakan beberapa metrik evaluasi yang relevan untuk algoritma klasifikasi. Model dievaluasi menggunakan pendekatan *hold-out validation* dengan membagi data menjadi 80% data latih dan 20% data uji.

Adapun metrik evaluasi yang digunakan adalah sebagai berikut:

a. *Accuracy* (Akurasi)

Akurasi mengukur proporsi prediksi yang benar dari seluruh prediksi yang dilakukan. Metrik ini memberikan gambaran umum terhadap performa model secara keseluruhan.

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

Namun, akurasi tidak cukup jika data tidak seimbang, misalnya jumlah pengiriman yang tepat waktu jauh lebih banyak daripada yang terlambat.

b. *Precision*

Precision mengukur seberapa banyak prediksi "terlambat" yang benar-benar terlambat.

$$\text{Precision} = \frac{TP}{TP + FP}$$

Metrik ini penting untuk menghindari terlalu banyak prediksi salah terhadap pengiriman yang sebenarnya tepat waktu.

c. *Recall* (Sensitivity)

Recall mengukur seberapa banyak keterlambatan yang benar-benar berhasil diprediksi oleh model.

$$\text{Recall} = \frac{TP}{TP + FN}$$

Metrik ini berguna untuk mengetahui seberapa baik model dalam menangkap semua keterlambatan.

d. *F1-Score*

F1-Score merupakan rata-rata harmonik dari *precision* dan *recall*. Metrik ini cocok digunakan jika distribusi kelas tidak seimbang dan kita ingin keseimbangan antara ketepatan dan kelengkapan prediksi.

$$\text{F1-score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

e. RMSE (Root Mean Squared Error)

Jika model diperlakukan sebagai regresi untuk memprediksi estimasi berapa lama keterlambatan terjadi, maka digunakan metrik RMSE. RMSE mengukur rata-rata akar kuadrat dari selisih antara nilai prediksi dan nilai aktual.

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$$

Semakin kecil nilai RMSE, semakin baik model dalam memprediksi durasi keterlambatan.

f. Perbandingan Antar Model

Semua algoritma yang digunakan (*Decision Tree*, *Random Forest*, dan *HistGradientBoostingClassifier*) dibandingkan berdasarkan metrik-metrik di atas. Model dengan:

- *Accuracy* tertinggi
- *F1-score* yang seimbang
- dan *precision* tinggi pada kelas "terlambat"

DAFTAR PUSTAKA

[1] K. Karjono, E. D. Kusumawati, K. Karmanis, and D. Kusumaningrum, "Transformasi Pemasaran Industri Logistik Dalam Meningkatkan Efisiensi Dan Keunggulan Kompetitif," *Maj. Ilm. Bahari Jogja*, vol. 22, no. 2, pp. 125–136, 2024.

[2] M. Siti, "BUKU STRATEGI PENINGKATAN KUALITAS PELAYANAN PUBLIK DI ERA DIGITALISASI." Cv Mitra Ilmu, 2023.

[3] A. Nopriyanto, "Analisis pengaruh corporate social responsibility (CSR) terhadap nilai perusahaan," *Komitmen J. Ilm. Manaj.*, vol. 5, no. 2, pp. 1–12, 2024.

[4] N. S. Apriana, "ANALISIS RISIKO RANTAI PASOK MATERIAL TERHADAP KETERLAMBATAN PELAKSANAAN PROYEK KONSTRUKSI GEDUNG= RISK ANALYSIS OF MATERIAL SUPPLY CHAIN ON DELAYS IN BUILDING CONSTRUCTION PROJECT IMPLEMENTATION." Universitas Hasanuddin, 2024.

[5] U. S. Sulistyawati, "Decoding Big Data: Mengubah Data Menjadi Keunggulan Kompetitif dalam Pengambilan Keputusan Bisnis," *J. Manaj. Dan Teknol.*, vol. 1, no. 2, pp. 58–71, 2024.

[6] A. Maulana and S. Sy, *Manajemen Operasional: Inovasi Berbasis Teknologi pada Manajemen Logistik*. MEGA PRESS NUSANTARA, 2024.

[7] N. H. Maulida, "Studi Literatur Penerapan Metoda Prototype dan Waterfall dalam Pembuatan Sebuah Aplikasi atau Website," *Jur. Tek. Inform. Fak. Tek. …*, no. April, pp. 4–6, 2022, [Online]. Available: https://www.researchgate.net/profile/Nur-Maulida-2/publication/359814481_STUDI_LITERATUR_PENERAPAN_METODE_PROTOTAYPE_DAN_WATERFALL_DALAM_PEMBUATAN_SEBUAH_APLIKASI_ATAU_WEBSITE/links/624fcd304f88c3119ce8737e/STUDI-LITERATUR-PENERAPAN-METODE-PROTOTAYPE-DAN-

[8] A. R. Faatihah, "Inovasi, Teknologi dan Kepuasan Pelanggan: Kunci Keberhasilan UMKM di Pasar yang Kompetitif Diksi Metris1, Ahmad Rasyiddin2, Febri Sari Siahaan3, Khansa Rayyani Aulia4," 2025.

[9] H. K. Pambudi, P. G. A. Kusuma, F. Yulianti, and K. A. Julian, "Prediksi Status Pengiriman Barang Menggunakan Metode Machine Learning," *J. Ilm. Teknol. Infomasi Terap.*, vol. 6, no. 2, pp. 100–109, 2020, doi: 10.33197/jitter.vol6.iss2.2020.396.

[10] A. Rinaldi Dikananda, N. A. Wijaya, and A. Faqih, "Implementasi Data Mining Pada Ketepatan Pengiriman Barang Dengan Menggunakan Algoritma K-Nearest Neighbors," *J. Sist. Inf. dan Manaj.*, vol. 10, no. 2338–1523, pp. 259–265, 2022, [Online]. Available: https://ejournal.stmikgici.ac.id/

[11] 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.

[12] K. K. Mulia and I. H. Sahputra, "Rekomendari Solusi dan Pembangunan Model Prediksi Keterlambatan…," *J. Titra*, vol. 9, no. 2, pp. 247–254, 2021.

[13] M. Reza and - Suprayogi, “Prediksi Jangka Waktu Pengiriman Barang Pada PT. Pos Indonesia menggunakan Backpropagation,” *CogITo Smart J.*, vol. 3, no. 1, pp. 111–122, 2017, doi: 10.31154/cogito.v3i1.50.111-122.

[14] Y. Mardi, “Data Mining : Klasifikasi Menggunakan Algoritma C4.5,” *Edik Inform.*, vol. 2, no. 2, pp. 213–219, 2017, doi: 10.22202/ei.2016.v2i2.1465.

[15] C. Science, “Prediction of Top Trend Logistics Goods Delivery Services Using Linear Regression Algorithm at PT . XNH Prediksi Jasa Pengiriman Barang Top Trend Logistik Menggunakan Algoritma Regresi Linear pada PT . XNH,” vol. 4, no. October, pp. 1448–1455, 2024.

[16] L. N. Halimah, S. Riyadi, and A. F. Jurjani, “IMPLEMENTASI PENGGUNAAN MACHINE LEARNING DALAM PEMBELAJARAN : SUATU TELAAH DESKRIPTIF,” vol. 1, no. 1, pp. 1–10, 2025.

[17] T. G. Laksana and S. Mulyani, “Pengetahuan Dasar Identifikasi Dini Deteksi Serangan Kejahatan Siber Untuk Mencegah Pembobolan Data Perusahaan,” *J. Ilm. Multidisiplin*, vol. 3, no. 01, pp. 109–122, 2024, doi: 10.56127/jukim.v3i01.1143.

[18] M. Muharrom, “Analisis Komparasi Algoritma Data Mining Naive Bayes, K-Nearest Neighbors dan Regresi Linier Dalam Prediksi Harga Emas,” *Bull. Inf. Technol.*, vol. 4, no. 4, pp. 430–438, 2023, doi: 10.47065/bit.v4i4.986.

[19] F. R. Fitria and D. T. T. LAUT, “Analisis Alternatif Rute Angkutan Penyeberangan: Studi Kasus Lintas Jawa Timur, Bali, Dan Nusa Tenggara Barat,” *Inst. Teknol. Sepuluh Novemb.*, 2018, [Online]. Available: https://repository.its.ac.id/50532/1/04411340000019_UNDERGRADUATED_THESIS.pdf

[20] A. Utama. W, “Analisis Faktor-Faktor Yang Mempengaruhi Penundaan Pengiriman Barang Melalui Jalur Laut,” *Citra Widya Edukasi*, vol. X, no. 2, pp. 97–108, 2018, [Online]. Available: c:/Users/Hp/Downloads/FAKTOR PENUNDAAN PENGIRIMAN.pdf

[21] A. Nuzulia, *Memahami Data Mining dengan Python: Implementasi Praktis, Eureka Media Aksara*. 1967.

[22] A. Bengnga and R. Ishak, “Implementasi Seleksi Fitur Klasifikasi Waktu Kelulusan Mahasiswa Menggunakan Correlation Matrix with Heatmap,” *Jambura J. Electr. Electron. Eng.*, vol. 4, no. 2, pp. 169–174, 2022, doi: 10.37905/jjeee.v4i2.14403.

[23] G. Dwilestari *et al.*, “PREDIKSI PERSETUJUAN PINJAMAN MENGGUNAKAN DATASET LOAN APPROVAL MENGGUNAKAN ALGORITMA KLASIFIKASI,” vol. 9, no. 1, pp. 1342–1347, 2025.

[24] “KELAPA SAWIT MENGGUNAKAN CITRA UNMANNED AERIAL VEHICLE DENGAN MODEL RANDOM FOREST REGRESSION Disusun dan diajukan oleh : MUHAMMAD IJLAL NURHADI SAWIT MENGGUNAKAN CITRA UNMANNED AERIAL VEHICLE Disusun dan diajukan oleh,” 2024.

[25] B. Bayu Baskoro *et al.*, “Analisis Sentimen Pelanggan Hotel di Purwokerto

Menggunakan Metode Random Forest dan TF-IDF (Studi Kasus: Ulasan Pelanggan Pada Situs TRIPADVISOR)," *J. Informatics Inf. Syst. Softw. Eng. Appl. (INISTA)*, , vol. Volume 3 N, no. 2, pp. 21–029, 2021, doi: 10.20895/INISTA.V3.

[26] M. Maftoun, N. Shadkam, S. S. S. Komamardakhi, Z. Mansor, and J. H. Joloudari, "Malicious URL Detection using optimized Hist Gradient Boosting Classifier based on grid search method," 2024, [Online]. Available: http://arxiv.org/abs/2406.10286

[27] D. Mwiti, "Gradient Boosted Decision Trees [Guide]: a Conceptual Explanation," *neptune.ai*, 2025. https://neptune.ai/blog/gradient-boosted-decision-trees-guide

[28] GeeksforGeeks, "HistGradientBoostingClassifier in Sklearn," *06 Juni*, 2024. https://www.geeksforgeeks.org/machine-learning/histgradientboostingclassifier-in-sklearn/

[29] C. Sanjaya and Suhono Harso, "Predictive Analytics Menggunakan Machine Learning Untuk Memprediksi Waktu Keterlambatan Berdasarkan," *J. Sist. Cerdas*, vol. 03, no. 02, pp. 165–180, 2020.