

PERBANDINGAN ALGORITMA LOGISTIC REGRESSION DAN XGBOOST DALAM PREDIKSI KETERLAMBATAN PENGIRIMAN E-COMMERCE

Hilmi Nurpadilah
Program Studi Teknik Informatika
Fakultas Teknologi Informasi
Universitas Bale Bandung

ABSTRAK: Perkembangan e-commerce yang semakin pesat membuat layanan pengiriman menjadi salah satu faktor penting dalam menjaga kepuasan pelanggan. Namun, dalam praktiknya, keterlambatan pengiriman masih sering terjadi dan dapat menimbulkan keluhan dari pelanggan. Oleh karena itu, diperlukan solusi yang mampu membantu memprediksi kemungkinan keterlambatan pengiriman sejak awal. Capstone project ini membahas penerapan machine learning untuk memprediksi keterlambatan pengiriman pada layanan e-commerce. Dua algoritma yang digunakan dalam penelitian ini adalah Logistic Regression dan XGBoost. Dataset yang digunakan berasal dari data historis pengiriman e-commerce yang diperoleh dari Kaggle dan telah melalui proses pembersihan serta pengolahan data. Kinerja kedua model dievaluasi menggunakan beberapa metrik, yaitu accuracy, F1-score, dan ROC-AUC. Hasil pengujian menunjukkan bahwa algoritma XGBoost memiliki performa yang lebih baik dibandingkan Logistic Regression dalam memprediksi keterlambatan pengiriman. Model terbaik kemudian diimplementasikan ke dalam aplikasi berbasis web menggunakan Streamlit sehingga dapat digunakan secara interaktif. Diharapkan hasil dari capstone project ini dapat memberikan gambaran pemanfaatan machine learning dalam bidang logistik e-commerce serta membantu meningkatkan kualitas layanan pengiriman.

Kata kunci: Prediksi keterlambatan pengiriman, E-commerce, Machine learning, Logistic Regression, XGBoost

ABSTRACT: The rapid growth of e-commerce has made delivery services an important factor in maintaining customer satisfaction. However, delivery delays still frequently occur and may lead to customer complaints. Therefore, an approach that can help predict potential delivery delays is needed. This capstone project explores the use of machine learning to predict delivery delays in e-commerce services. Two algorithms are applied in this study, namely Logistic Regression and XGBoost. The dataset used is an e-commerce delivery dataset obtained from Kaggle, which was processed through data cleaning and preprocessing stages. Model performance was evaluated using accuracy, F1-score, and ROC-AUC metrics. The results show that XGBoost performs better than Logistic Regression in predicting delivery delays. The best-performing model was then implemented into a web-based application using Streamlit to allow interactive predictions. This capstone project is expected to demonstrate the practical use of machine learning in e-commerce logistics and support improvements in delivery service quality.

Keywords: Delivery delay prediction, E-commerce, Machine learning, Logistic Regression, XGBoost

PENDAHULUAN

1.1 Latar Belakang

Kemajuan e-commerce yang semakin berkembang menjadikan proses jual beli lebih mudah dan cepat. Saat ini, banyak masyarakat memilih berbelanja secara online karena dinilai praktis dan efisien. Seiring dengan meningkatnya aktivitas tersebut, layanan pengiriman barang memegang peranan penting dalam menentukan kualitas layanan e-commerce secara keseluruhan.

Namun, Meskipun sistem e-commerce terus berkembang, permasalahan keterlambatan pengiriman masih sering terjadi. Keterlambatan ini dapat disebabkan oleh berbagai faktor, seperti metode pengiriman yang digunakan, proses penanganan barang di gudang, berat produk, hingga jumlah interaksi pelanggan dengan layanan *customer service* [1]. Kondisi tersebut dapat menurunkan tingkat kepuasan pelanggan dan berpotensi memengaruhi kepercayaan terhadap layanan e-commerce.

Untuk membantu mengatasi permasalahan tersebut, dibutuhkan suatu pendekatan yang mampu menganalisis data pengiriman secara lebih efektif. Salah satu pendekatan yang dapat digunakan adalah *machine learning*. Dengan memanfaatkan data historis pengiriman, machine learning memungkinkan sistem untuk mempelajari pola tertentu dan menghasilkan prediksi mengenai kemungkinan terjadinya keterlambatan pengiriman.

Dalam bidang *machine learning*, banyak algoritma yang dapat digunakan untuk permasalahan klasifikasi. Dua algoritma yang cukup sering digunakan adalah *Logistic Regression* dan *XGBoost*. *Logistic Regression* dikenal sebagai algoritma yang sederhana dan mudah dipahami, sedangkan *XGBoost* merupakan algoritma berbasis ensemble yang mampu menangani data dengan tingkat kompleksitas yang lebih tinggi [2]. Maka dari itu, pada *capstone project* ini kedua algoritma tersebut digunakan untuk memprediksi keterlambatan pengiriman e-commerce serta dibandingkan kinerjanya.

1.2 Rumusan Masalah

Berdasarkan latar belakang tersebut, *capstone project* ini mengangkat sejumlah masalah, antara lain:

1. Faktor apa saja yang memengaruhi keterlambatan pengiriman pada layanan e-commerce?
2. Bagaimana kinerja algoritma Logistic Regression dalam memprediksi keterlambatan pengiriman?
3. Bagaimana kinerja algoritma XGBoost dalam memprediksi keterlambatan pengiriman?
4. Algoritma manakah yang memiliki performa terbaik berdasarkan metrik evaluasi yang digunakan?

1.3 Batasan Masalah

Supaya pembahasan dalam *capstone project* ini tetap fokus pada masalah utama, penelitian ini dibatasi oleh hal-hal berikut:

1. Dataset berasal dari dataset e-commerce publik dan telah melalui proses pembersihan data.
2. Variabel yang digunakan dibatasi pada fitur-fitur yang tersedia dalam dataset, seperti warehouse block, mode of shipment, customer rating, berat barang, dan diskon.
3. Algoritma yang digunakan dibatasi pada Logistic Regression dan XGBoost.
4. Kinerja model dievaluasi menggunakan beberapa metrik, yaitu *accuracy*, *precision*, *recall*, *F1-score*, dan *ROC-AUC*.

1.4 Tujuan Penelitian

Tujuan dari penelitian *capstone project* ini adalah sebagai berikut:

1. Menerapkan machine learning untuk memprediksi keterlambatan pengiriman pada layanan e-commerce.
2. Membangun model prediksi keterlambatan pengiriman menggunakan algoritma Logistic Regression.
3. Membangun model prediksi keterlambatan pengiriman menggunakan algoritma XGBoost.
4. Membandingkan performa kedua algoritma untuk menentukan model terbaik dalam memprediksi keterlambatan pengiriman.

Hasil dari *capstone project* ini diharapkan dapat dijadikan sebagai bahan referensi bagi pihak pengelola e-commerce dalam meningkatkan kualitas layanan logistik serta sebagai bahan kajian akademis terkait penerapan *machine learning* pada bidang logistik dan distribusi.

TINJAUAN PUSTAKA

2.1 Penelitian Terdahulu

Penelitian mengenai penerapan machine learning dalam berbagai permasalahan klasifikasi telah banyak dilakukan, termasuk pada bidang logistik, distribusi, dan layanan e-commerce. Penelitian-penelitian terdahulu tersebut menjadi dasar dalam pemilihan metode serta pendekatan yang digunakan pada penelitian ini.

Penelitian oleh [3] membahas penerapan algoritma *machine learning* untuk prediksi keterlambatan pengiriman barang. Hasil penelitian menunjukkan bahwa pendekatan berbasis data historis pengiriman mampu membantu perusahaan logistik dalam mengidentifikasi potensi keterlambatan secara lebih dini.

Penelitian lain yang dilakukan oleh [4], [5], menggunakan algoritma *Logistic Regression* untuk permasalahan klasifikasi pada data layanan pelanggan. Penelitian tersebut menyimpulkan bahwa *Logistic Regression* memiliki keunggulan dalam hal interpretabilitas model dan dapat dijadikan sebagai model *baseline* yang baik.

Selanjutnya, penelitian oleh [4], [6], menerapkan algoritma *XGBoost* untuk prediksi berbasis data tabular dan menunjukkan bahwa *XGBoost* menghasilkan hasil atau performa yang lebih baik dibandingkan metode konvensional, terutama pada data dengan kompleksitas tinggi.

Penelitian oleh [7], [8], membandingkan beberapa algoritma klasifikasi, yaitu *Logistic Regression* dan metode *ensemble*. Hasil penelitian tersebut menunjukkan bahwa algoritma ensemble cenderung menghasilkan tingkat akurasi dan stabilitas yang lebih baik.

Penelitian lain yang relevan dilakukan oleh [3], [9], yang memanfaatkan algoritma *XGBoost* dalam prediksi pada sektor layanan digital. Penelitian tersebut menyatakan bahwa *XGBoost* efektif dalam menangani hubungan non-linear antar variabel.

Berdasarkan beberapa penelitian terdahulu tersebut, dapat disimpulkan bahwa *Logistic Regression* dan *XGBoost* merupakan algoritma yang relevan dan layak digunakan dalam permasalahan klasifikasi, termasuk prediksi keterlambatan pengiriman e-commerce. Namun, perbandingan kinerja kedua algoritma tersebut pada konteks pengiriman e-commerce masih perlu dikaji lebih lanjut.

2.2 Landasan Teori

2.2.1 Machine Learning

Machine learning merupakan cabang dari kecerdasan buatan yang memungkinkan sistem belajar langsung dari data untuk menghasilkan prediksi tanpa perlu aturan yang ditentukan secara manual [9]. Pendekatan ini banyak dimanfaatkan dalam permasalahan klasifikasi dan prediksi.

Pada capstone project ini, machine learning digunakan untuk mengenali pola pada data pengiriman dan memperkirakan kemungkinan terjadinya keterlambatan pengiriman.

2.2.2 Logistic Regression

Logistic Regression adalah algoritma klasifikasi yang menghasilkan prediksi berdasarkan hubungan antara variabel input dan target dalam bentuk probabilitas [8].

Pada capstone project ini, algoritma tersebut digunakan sebagai model banding awal untuk melihat pengaruh hubungan linear antar fitur terhadap prediksi keterlambatan pengiriman.

2.2.3 XGBoost

XGBoost adalah algoritma *machine learning* yang cara kerjanya adalah menggabungkan banyak pohon keputusan agar hasil prediksi yang dihasilkan menjadi lebih baik [10].

Algoritma ini dikenal mampu menangani data yang kompleks dan memiliki performa yang baik pada permasalahan klasifikasi. Oleh karena itu, *XGBoost* digunakan sebagai model utama dalam capstone project ini untuk dibandingkan dengan *Logistic Regression*.

Beberapa penelitian menunjukkan bahwa *XGBoost* sering memberikan hasil yang lebih unggul dibandingkan algoritma klasifikasi lainnya pada data tabular. Oleh karena itu, *XGBoost* digunakan dalam penelitian ini sebagai model

utama untuk dibandingkan dengan *Logistic Regression*.

2.2.4 Evaluasi Kinerja Model

Evaluasi model bertujuan untuk mengukur sejauh mana model *machine learning* mampu menghasilkan prediksi yang baik. Dalam klasifikasi, penilaian model tidak cukup jika hanya menggunakan satu metrik, sehingga diperlukan beberapa metrik seperti *accuracy*, *precision*, *recall*, *F1-score*, dan *ROC-AUC* agar performa model dapat terlihat dengan lebih jelas, terutama pada data yang tidak seimbang.

Pada *capstone project* ini, evaluasi performa model difokuskan pada penggunaan metrik *accuracy*, *F1-score*, dan *ROC-AUC*. Ketiga metrik tersebut digunakan untuk membandingkan kinerja algoritma *Logistic Regression* dan *XGBoost* dalam memprediksi keterlambatan pengiriman pada layanan *e-commerce*.

METODOLOGI PENELITIAN

3.1 Waktu dan Tempat Penelitian

Dilaksanakan pada semester berjalan tahun akademik 2025/2026. Proses penelitian dilakukan secara daring (online), meliputi pengumpulan data, pengolahan data, pemodelan *machine learning*, hingga evaluasi model. Seluruh proses pengolahan serta analisis data dilakukan dengan memanfaatkan perangkat laptop pribadi yang digunakan oleh peneliti.

3.2 Alat dan Bahan Penelitian

3.2.1 Alat Penelitian

Alat yang digunakan adalah sebagai berikut:

1. Perangkat keras berupa laptop dengan spesifikasi yang memadai untuk pengolahan data.
2. Sistem operasi Windows.
3. Bahasa pemrograman Python.
4. Perangkat lunak dan library pendukung dari Python

3.2.2 Bahan Penelitian

Data yang digunakan dalam *capstone project* ini adalah dataset pengiriman *e-commerce* yang diperoleh dari platform *Kaggle*. Dataset tersebut memuat data historis pengiriman yang mencakup informasi terkait pelanggan, produk, serta proses pengiriman. Data ini kemudian dimanfaatkan

untuk data latih serta data uji dalam proses pemodelan *machine learning*.

3.3 Metode Penelitian

Penelitian ini menggunakan pendekatan eksperimen dengan metode kuantitatif. Pendekatan tersebut diterapkan untuk menguji serta membandingkan kinerja algoritma *Logistic Regression* dan *XGBoost* dalam memprediksi keterlambatan pengiriman pada layanan *e-commerce*. Proses penelitian dilakukan secara bertahap dan terstruktur, dimulai dari pengumpulan data, pengolahan data, hingga evaluasi hasil dari model yang dibangun.

3.4 Metode Pengumpulan Data

Pengumpulan data dilakukan dengan cara mengunduh dataset tersebut untuk digunakan dalam tahap pengolahan dan pemodelan.

3.5 Tahapan Penelitian

3.5.1 Data Preprocessing

Tujuan dari tahap preprocessing data adalah untuk mempersiapkan data agar siap digunakan dalam proses pemodelan. Tahapan preprocessing yang dilakukan meliputi:

1. Pembersihan data untuk memastikan tidak terdapat nilai kosong atau data tidak valid.
2. Transformasi data kategorikal menjadi bentuk numerik menggunakan teknik encoding.
3. Normalisasi atau standarisasi data numerik untuk meningkatkan performa model.

3.5.2 Pembagian Data

Dataset yang digunakan dalam penelitian ini dibagi menjadi dua bagian, yaitu data latih dan data uji dengan persentase 80% data latih dan 20% data uji. Data pelatihan untuk membangun dan melatih model *machine learning*, sedangkan data pengujian dimanfaatkan untuk mengevaluasi kinerja model setelah proses pelatihan selesai.

3.5.3 Pembuatan dan Pelatihan Model

Pada tahap pemodelan, dibuat model prediksi dengan menggunakan dua algoritma, yaitu *Logistic Regression* dan *XGBoost*. *Logistic Regression* dimanfaatkan sebagai model awal untuk keperluan pembandingan, sementara

XGBoost digunakan sebagai model dengan pendekatan *ensemble*. Proses pelatihan kedua model tersebut dilakukan menggunakan data latih yang telah disiapkan sebelumnya.

3.5.4 Evaluasi Model

Evaluasi model dilakukan untuk mengetahui kinerja masing-masing algoritma dalam memprediksi keterlambatan pengiriman. Metrik evaluasi yang digunakan dalam penelitian ini yaitu *accuracy*, *precision*, *recall*, *F1-score*, dan *ROC-AUC*. Hasil evaluasi digunakan sebagai dasar dalam menentukan model terbaik.

3.5.5 Implementasi Model

Model dengan kinerja terbaik berdasarkan hasil evaluasi kemudian diterapkan ke dalam sebuah aplikasi berbasis web berbasis *Streamlit*. Dari aplikasi tersebut, pengguna bisa data pengiriman dan langsung memperoleh hasil prediksi keterlambatan secara interaktif.

Dengan tahapan metodologi yang telah dilakukan, *capstone project* ini diharapkan mampu menghasilkan model prediksi keterlambatan pengiriman e-commerce yang memiliki tingkat akurasi yang baik serta dapat dimanfaatkan sebagai sistem pendukung keputusan.

HASIL DAN PEMBAHASAN

4.1 Deskripsi Data

Dataset yang digunakan berasal dari platform Kaggle dan berisi data historis pengiriman e-commerce dengan berbagai atribut yang berkaitan dengan pelanggan, produk, serta proses pengiriman. Sebanyak 11.000 data digunakan dan dibagi ke dalam data pelatihan serta data pengujian.

Beberapa atribut yang dimanfaatkan dalam penelitian ini meliputi *warehouse block*, *mode of shipment*, jumlah panggilan *customer care*, *customer rating*, harga produk, jumlah pembelian sebelumnya, tingkat kepentingan produk, jenis kelamin pelanggan, diskon yang diberikan, serta berat barang dalam gram. Sementara itu, target klasifikasi pada penelitian ini adalah status keterlambatan pengiriman yang dibagi ke dalam dua kelas, yaitu pengiriman tepat waktu dan pengiriman terlambat.

4.2 Hasil Preprocessing Data

Pada tahap *preprocessing*, data terlebih dahulu dipastikan bersih dari nilai kosong maupun data yang tidak valid. Setelah itu, atribut yang bersifat kategorikal dikonversi ke bentuk numerik melalui proses encoding agar dapat digunakan dalam pemodelan *machine learning*. Data numerik kemudian dinormalisasi untuk meningkatkan performa model, khususnya pada algoritma *Logistic Regression*.

Hasil *preprocessing* menunjukkan bahwa dataset telah siap digunakan untuk proses pemodelan dan tidak ditemukan anomali data yang signifikan.

4.3 Hasil Pemodelan Logistic Regression

Logistic Regression digunakan sebagai model awal pembanding pada penelitian ini. Model dilatih menggunakan data latih dan selanjutnya diuji menggunakan data uji untuk melihat hasil performanya.

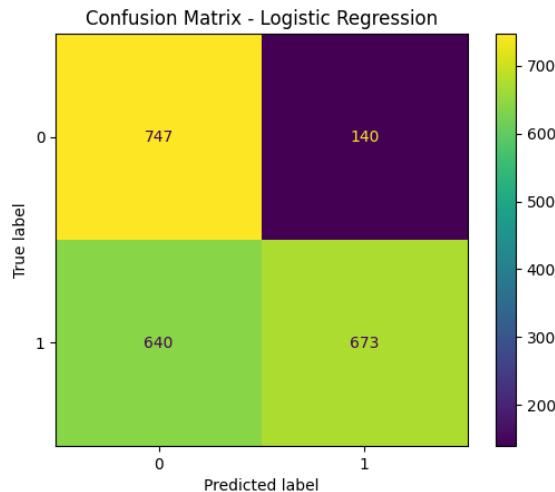
Hasil evaluasi menunjukkan bahwa *Logistic Regression* mampu memprediksi keterlambatan pengiriman dengan tingkat akurasi yang tergolong cukup baik.

	Model	Accuracy	F1-Score	ROC-AUC
1	XGBoost	0.675909	0.653379	0.734613
0	Logistic Regression	0.645455	0.633114	0.716983

Tabel 1. Hasil Evaluasi Model Logistic Regression dan XGBoost

Berdasarkan hasil pengujian yang dilakukan, model *Logistic Regression* memperoleh nilai akurasi sebesar 64,55%, *F1-score* sebesar 63,31%, dan *ROC-AUC* sebesar 71,69%. Hasil yang didapatkan menunjukkan bahwa model

dapat melakukan prediksi dengan kinerja yang cukup baik.

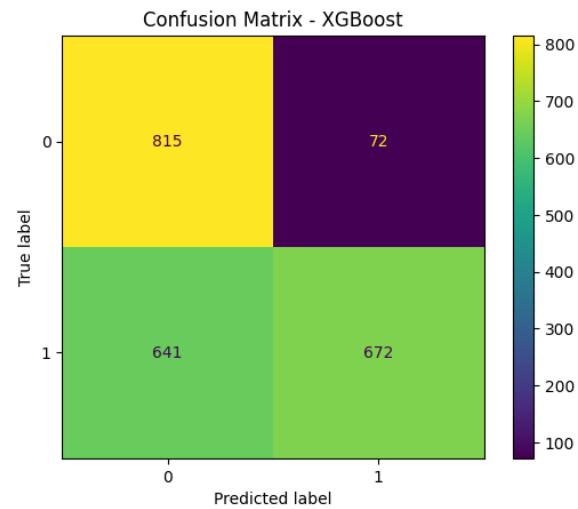


Gambar 1. Confusion Matrix - Logistic Regression

Dari hasil *confusion matrix* dapat dilihat bahwa model cukup baik dalam mengenali pengiriman yang tepat waktu. Namun, masih ada beberapa pengiriman yang mengalami keterlambatan dan belum dapat diprediksi dengan baik oleh model.

4.4 Hasil Pemodelan XGBoost

Model *XGBoost* digunakan sebagai model pembanding dalam penelitian ini. Proses pelatihan dan pengujian dilakukan dengan skema yang sama seperti pada model *Logistic Regression*. Hasil pengujian menunjukkan bahwa kemampuan *XGBoost* lebih baik dibandingkan *Logistic Regression*. Berdasarkan evaluasi yang dilakukan, model *XGBoost* memperoleh nilai akurasi sebesar 67,59%, F1-score sebesar 65,34%, serta ROC-AUC sebesar 73,46%.



Gambar 2. Confusion Matrix XGBoost

Berdasarkan *confusion matrix*, model *XGBoost* mampu mengurangi kesalahan prediksi dibandingkan model *Logistic Regression*, khususnya dalam memprediksi data pengiriman tepat waktu.

4.5 Perbandingan Kinerja Model

Hasil perbandingan menunjukkan bahwa algoritma *XGBoost* memberikan kinerja yang lebih baik dibandingkan *Logistic Regression* pada seluruh metrik evaluasi yang digunakan. Hal tersebut mengindikasikan bahwa *XGBoost* lebih mampu menangani karakteristik data pengiriman e-commerce yang memiliki pola kompleks dan tidak bersifat linear.

Peningkatan performa *XGBoost* dibandingkan *Logistic Regression* sejalan dengan penelitian sebelumnya yang menyatakan bahwa algoritma berbasis *boosting*, khususnya *XGBoost*, memiliki kinerja yang lebih unggul dibandingkan metode klasifikasi konvensional pada data tabular yang kompleks [11].

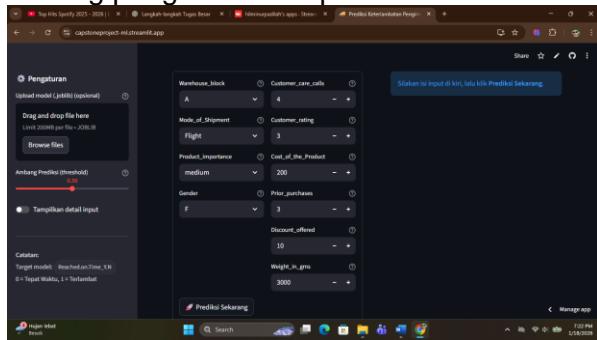
4.6 Pembahasan

Berdasarkan hasil yang diperoleh, *machine learning* dapat membantu dalam memprediksi keterlambatan pengiriman pada layanan e-commerce dengan cukup baik. *Logistic Regression* berperan sebagai model awal yang memberikan gambaran hubungan antar variabel, sedangkan *XGBoost* mampu mempelajari pola data yang lebih kompleks.

Beberapa fitur, seperti besarnya diskon, berat barang, serta metode pengiriman, terbukti

berpengaruh terhadap hasil prediksi keterlambatan pengiriman. Temuan tersebut sejalan dengan penelitian terdahulu yang menyebutkan bahwa faktor produk dan proses distribusi memegang peranan penting dalam menentukan ketepatan waktu pengiriman.

Penerapan model dengan performa terbaik ke dalam aplikasi berbasis *Streamlit* menunjukkan bahwa hasil *capstone project* ini tidak hanya terbatas pada kajian teoritis, tetapi juga dapat digunakan secara praktis sebagai sistem pendukung pengambilan keputusan.



Gambar 3. Antarmuka Web

Aplikasi ini memungkinkan pengguna untuk melakukan prediksi keterlambatan pengiriman secara interaktif berdasarkan data yang dimasukkan.

Berdasarkan hasil yang diperoleh, *capstone project* ini menunjukkan bahwa pendekatan machine learning, terutama algoritma *XGBoost*, memiliki potensi yang kuat untuk membantu meningkatkan efisiensi serta kualitas layanan logistik pada sistem e-commerce.

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Berdasarkan proses implementasi serta hasil evaluasi yang telah dilakukan pada *capstone project* ini, beberapa kesimpulan dapat disusun sebagai berikut:

1. Penerapan *machine learning* dapat digunakan untuk memprediksi keterlambatan pengiriman pada sistem e-commerce dengan memanfaatkan data historis pengiriman.
2. Algoritma *Logistic Regression* mampu digunakan sebagai model baseline dalam memprediksi keterlambatan pengiriman,

dengan performa yang cukup baik dan mudah diinterpretasikan.

3. Algoritma *XGBoost* menunjukkan performa yang lebih baik dibandingkan *Logistic Regression* berdasarkan metrik evaluasi *accuracy*, *F1-score*, dan *ROC-AUC*.
4. Faktor-faktor seperti diskon yang diberikan, berat barang, metode pengiriman, dan karakteristik produk memiliki pengaruh terhadap hasil prediksi keterlambatan pengiriman.
5. Model terbaik hasil evaluasi berhasil diimplementasikan ke dalam aplikasi berbasis web *Streamlit*, sehingga hasil *capstone project* ini dapat digunakan sebagai sistem pendukung keputusan secara interaktif.

Secara keseluruhan, *capstone project* ini membuktikan bahwa algoritma *machine learning*, khususnya *XGBoost*, memiliki potensi yang baik dalam membantu meningkatkan efisiensi dan ketepatan waktu pengiriman pada layanan e-commerce.

5.2 Saran

Berdasarkan hasil *capstone project* ini, beberapa saran untuk pengembangan selanjutnya adalah sebagai berikut:

1. Performa model bisa lebih baik apabila penelitian selanjutnya menggunakan data dengan jumlah dan variasi yang lebih besar.
2. Pengembangan selanjutnya juga dapat mencoba menerapkan algoritma *machine learning* lain, seperti *Random Forest*, *LightGBM*, atau *Neural Network*, guna memperoleh performa prediksi yang lebih optimal.
3. Penyesuaian *threshold* klasifikasi dapat diteliti lebih lanjut untuk menyesuaikan kebutuhan bisnis, misalnya untuk meminimalkan risiko keterlambatan pengiriman.
4. Pengembangan aplikasi dapat ditingkatkan dengan menambahkan fitur visualisasi, integrasi database, atau penggunaan data *real-time* dari sistem e-commerce.

Saran-saran tersebut diharapkan dapat menjadi acuan untuk pengembangan *capstone*

project atau penelitian selanjutnya di bidang prediksi dan analisis logistik e-commerce.

DAFTAR PUSTAKA

- [1] H. Lintang, F. Hafidz, M. Rafi, A. Naufal, and M. A. Muthalib, "Penerapan AI untuk Optimasi Rute Secara Real-time dan Meningkatkan Efisiensi Pengiriman," *Jurnal Sains Masyarakat*, vol. 01, doi: 10.1016/j.ijlm.2022.04.003.
- [2] D. T. Murdiansyah, "Prediksi Stroke Menggunakan Extreme Gradient Boosting," *JIKO (Jurnal Informatika dan Komputer)*, vol. 8, no. 2, p. 419, Sep. 2024, doi: 10.26798/jiko.v8i2.1295.
- [3] E. Mustika Sari, C. Sabila, R. Fakhrizal Adam, and R. Kurniawan, "Analisis dan Prediksi Indeks Kualitas Udara Jakarta: Penerapan Algoritma XGBoost," *Jurnal Nasional Teknologi dan Sistem Informasi*, vol. 11, no. 2, pp. 161–169, Sep. 2025, doi: 10.25077/teknosi.v11i2.2025.161-169.
- [4] I. Azis *et al.*, "Penentuan Faktor Kemiskinan Indonesia Menggunakan Regresi Logistik," 2023. [Online]. Available: <http://www.ojs.unm.ac.id/jmathcos>
- [5] F. Bukhari, S. Nurdianti, M. Khoirun Najib, and R. Nurul Amalia, "Deteksi Penyakit Jantung Menggunakan Metode Klasifikasi Decision Tree dan Regresi Logistik," 2023.
- [6] A. Haris, M. Sholeh, L. Muflikhah, and N. Yudistira, "Peningkatan Akurasi Prediksi Harga Barang Impor Menggunakan XGBoost dan Particle Swarm Optimization," *Jurnal Teknologi Informasi dan Ilmu Komputer*, vol. 12, no. 2, pp. 351–358, Apr. 2025, doi: 10.25126/jtiik.2025129419.
- [7] R. Siringoringo, R. Perangin Angin, and B. Rumahorbo, "MODEL KLASIFIKASI GENETIC-XGBOOST DENGAN T-DISTRIBUTED STOCHASTIC NEIGHBOR EMBEDDING PADA PERAMALAN PASAR," 2022. [Online]. Available: <https://archive.ics.uci.edu/ml/datasets/onlin e+retail>
- [8] S. Wahyuni Kalumbang, "PERBANDINGAN REGRESI LOGISTIK, KLASIFIKASI NAIVE BAYES, DAN RANDOM FOREST (COMPARISON THE LOGISTIC REGRESSION, NAIVE BAYES CLASSIFICATION, AND RANDOM FOREST)," vol. 03, no. 02, p. 2021.
- [9] 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," *Journal of Information Technology and Computer Science (INTECOMS)*, vol. 8, no. 6, 2025.
- [10] R. Yunanto and U. Budiyanto, "Implementasi XGBoost dan SMOTE untuk Meningkatkan Deteksi Transaksi Fraud di Industri Jasa Keuangan," *Jurnal Pendidikan dan Teknologi Indonesia*, vol. 4, no. 11, pp. 525–535, Dec. 2024, doi: 10.52436/1.jpti.518.
- [11] M. Ravly Andryan *et al.*, "KOMPARASI KINERJA ALGORITMA XGBOOST DAN ALGORITMA SUPPORT VECTOR MACHINE (SVM) UNTUK DIAGNOSA PENYAKIT KANKER PAYUDARA," *Jurnal Informatika dan Komputer*, vol. 6, no. 1, pp. 1–5, 2022.