

CapstoneProject_301230004_Hi Imi Nurpadilah.pdf

by chujukisoa@gmail.com chujukisoa@gmail.com

Submission date: 18-Jan-2026 10:23PM (UTC+0900)

Submission ID: 2857792925

File name: CapstoneProject_301230004_Hilmi_Nurpadilah.pdf (299K)

Word count: 3193

Character count: 21964

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

H³⁵ Nurpadilah
Prodi Teknik Informatika
Fakultas Teknologi Informatasi
Universitas Bale Bandung

ABSTRAK: Keterlambatan pengiriman merupakan salah satu permasalahan yang sering terjadi pada sistem e-commerce dan dapat berdampak pada kepuasan pelanggan. Oleh karena itu, diperlukan suatu pendekatan yang mampu memprediksi potensi keterlambatan pengiriman secara lebih awal. Penelitian ini bertujuan untuk membandingkan kinerja algoritma Logistic Regression dan XGBoost dalam memprediksi keterlambatan pengiriman e-commerce. Dataset yang digunakan merupakan data pengiriman e-commerce yang diperoleh dari Kaggle dan mencakup informasi terkait pelanggan, produk, serta proses pengiriman. Metode penelitian yang digunakan meliputi tahap preprocessing data, pembagian data latih dan data uji, pelatihan model, se²⁵ evaluasi kinerja model menggunakan metrik accuracy, F1-score, dan ROC-AUC. Hasil penelitian menunjukkan bahwa kedua algoritma mampu melakukan prediksi keterlambatan pengiriman dengan baik, namun algoritma XGBoost menunjukkan kerja yang lebih unggul dibandingkan Logistic Regression berdasarkan hasil evaluasi yang diperoleh. Penelitian ini diharapkan dapat memberikan kontribusi dalam pengembangan sistem pendukung keputusan untuk meningkatkan efisiensi dan ketepatan pengiriman pada layanan e-commerce.

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

ABSTRACT: Delivery delays are one of the common problems in e-commerce systems and can negatively affect customer satisfaction. Therefore, an approach is needed to predict potential delivery delays at an early stage. This study aims to compare the performance of Logistic Regression and XGBoost algorithms in predicting e-commerce delivery delays. The dataset used in this study consists of e-commerce shipment data obtained from Kaggle, which includes information related to customers, products, and delivery processes. The research methodology includes data preprocessing, data splitting into training and testing sets, model training, and performance evaluation using accuracy, F1-score, and ROC-AUC metrics. The results show that both algorithms are capable of predicting delivery delays effectively; however, the XGBoost algorithm demonstrates better performance compared to Logistic Regression based on the evaluation results. This study is expected to contribute to the development of decision support systems to improve delivery efficiency and timeliness in e-commerce services.

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

24
PENDAHULUAN**1.1 Latar Belakang**

Perkembangan e-commerce di Indonesia [4] mengalami pertumbuhan yang sangat pesat seiring dengan meningkatnya peningkatan internet dan perubahan perilaku konsumen. Aktivitas jual beli yang sebelumnya dilakukan secara konvensional kini beralih ke platform digital, sehingga menuntut perusahaan untuk menyediakan layanan logistik yang cepat, tepat, dan andal. Salah satu indikator utama kualitas layanan logistik adalah ketepatan waktu pengiriman barang kepada pelanggan.

Namun, dalam praktiknya, keterlambatan pengiriman masih menjadi permasalahan yang sering terjadi pada sistem distribusi e-commerce. Keterlambatan pengiriman dapat disebabkan oleh berbagai faktor, seperti metode pengiriman, kepadatan gudang, berat barang, diskon yang diberikan, hingga intensitas interaksi pelanggan dengan layanan customer care. Permasalahan ini tidak hanya berdampak pada menurunnya kepuasan pelanggan, tetapi juga dapat memengaruhi reputasi perusahaan dan tingkat loyalitas konsumen.

Seiring berkembangnya teknologi, pemanfaatan machine learning menjadi salah satu solusi yang banyak digunakan untuk memprediksi dan meminimalkan risiko keterlambatan pengiriman. Beberapa penelitian menunjukkan bahwa algoritma machine learning mampu menangani data yang kompleks dan bersifat non-linear, sehingga dapat menghasilkan prediksi yang lebih akurat dibandingkan metode konvensional. Algoritma seperti Logistic Regression banyak digunakan karena interpretabilitasnya yang baik, sedangkan XGBoost dikenal memiliki performa tinggi dalam menangani data tabular dan tidak seimbang.

Berdasarkan penelitian-penelitian sebelumnya, XGBoost terbukti unggul dalam berbagai kasus prediksi, seperti prediksi harga, kualitas udara, dan peramalan berbasis data kompleks, sedangkan regresi logistik tetap relevan sebagai baseline model yang sederhana dan mudah diinterpretasikan [1], [2], [3]. Oleh karena itu, diperlukan penelitian yang mengkaji dan membandingkan performa kedua algoritma tersebut dalam konteks prediksi keterlambatan pengiriman e-commerce.

Penelitian ini memanfaatkan dataset e-commerce yang tersedia secara publik dan mengimplementasikan algoritma Logistic Regression dan XGBoost untuk memprediksi status keterlambatan pengiriman. Diharapkan hasil penelitian [5] dapat memberikan gambaran model terbaik yang dapat digunakan sebagai sistem pendukung keputusan dalam manajemen logistik e-commerce.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah diuraikan, rumusan masalah dalam penelitian ini adalah sebagai berikut:

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

Agar penelitian lebih terarah dan fokus, maka batasan masalah dalam penelitian ini adalah:

1. Dataset yang digunakan 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 hanya Logistic Regression dan XGBoost.
4. Evaluasi model dilakukan menggunakan metrik akurasi, precision, recall, F1-score, dan ROC-AUC.

1.4 Tujuan Penelitian

Tujuan dari penelitian ini adalah:

1. Mengidentifikasi faktor-faktor yang berpengaruh terhadap keterlambatan pengiriman 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 penelitian ini diharapkan dapat menjadi referensi bagi pengelola e-commerce dalam meningkatkan kualitas layanan logistik serta sebagai bahan kajian akademis terkait penerapan machine learning pada bidang logistik dan distribusi.

41

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 [2], 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 [1], 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 [3], menerapkan algoritma XGBoost untuk prediksi berbasis data modular dan menunjukkan bahwa XGBoost mampu menghasilkan performa yang lebih baik dibandingkan metode konvensional, terutama pada data dengan kompleksitas tinggi.

65 Penelitian oleh [4], membandingkan beberapa algoritma klasifikasi, termasuk Logistic Regression dan metode ensemble. Hasil

penelitian menunjukkan bahwa algoritma ensemble cenderung memberikan tingkat akurasi dan stabilitas yang lebih baik.

Penelitian lain yang relevan dilakukan oleh [5], 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 berfokus pada pengembangan sistem yang mampu belajar dari data dan meningkatkan performanya tanpa harus diprogram secara eksplisit. Machine learning banyak digunakan untuk menyelesaikan permasalahan klasifikasi, regresi, dan clustering. Dalam penelitian ini, machine learning digunakan untuk memprediksi keterlambatan pengiriman berdasarkan data historis e-commerce.

2.2.2 Logistic Regression

Logistic Regression merupakan algoritma klasifikasi yang digunakan untuk memprediksi probabilitas suatu kejadian dengan keluaran bersifat biner. Algoritma ini bekerja dengan memodelkan hubungan antara variabel independen dan probabilitas kelas target menggunakan fungsi logistik. Logistic Regression banyak digunakan karena modelnya sederhana, mudah diinterpretasikan, serta efisien dalam komputasi.

Dalam konteks prediksi keterlambatan pengiriman, Logistic Regression digunakan sebagai model baseline untuk mengetahui sejauh mana hubungan linear antar fitur dapat menjelaskan status keterlambatan pengiriman.

2.2.3 XGBoost

Extreme Gradient Boosting (XGBoost) merupakan algoritma ensemble berbasis gradient boosting yang mengombinasikan beberapa model pohon keputusan untuk menghasilkan prediksi yang lebih akurat. XGBoost dikenal memiliki performa tinggi, mampu menangani data berskala besar, serta efektif dalam menangani data tidak seimbang.

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 kinerja model merupakan tahapan penting dalam penelitian machine learning untuk mengetahui tingkat keberhasilan model dalam melakukan prediksi. Metrik evaluasi yang umum digunakan pada permasalahan klasifikasi antara lain akurasi, precision, recall, F1-score, dan ROC-AUC. Penggunaan beberapa metrik evaluasi bertujuan untuk memberikan gambaran kinerja model secara lebih menyeluruh, terutama pada dataset yang memiliki distribusi ¹⁷ yang tidak seimbang.

Dalam penelitian ini, evaluasi kinerja model dilakukan menggunakan metrik accuracy, F1-score, dan ROC-AUC untuk membandingkan performa Logistic Regression dan XGBoost dalam memprediksi keterlambatan pengiriman e-commerce.

10 METODOLOGI PENELITIAN

3.1 Waktu dan Tempat Penelitian

Penelitian ini 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 dan analisis data dilakukan menggunakan perangkat komputer pribadi peneliti dengan bantuan perangkat lunak pendukung.

12 3.2 Alat dan Bahan Penelitian

3.2.1 Alat Penelitian

Alat yang digunakan dalam penelitian ini adalah sebagai berikut:

1. Perangkat keras berupa komputer atau laptop dengan spesifikasi yang memadai untuk pengolahan data.
2. Sistem operasi Windows.
3. Bahasa pemrograman Python.
4. Perangkat lunak dan library pendukung, antara lain:
 - Pandas dan NumPy untuk pengolahan data
 - Scikit-learn untuk pemodelan Logistic Regression dan evaluasi model
 - XGBoost untuk pemodelan Extreme Gradient Boosting
 - Matplotlib dan Seaborn untuk visualisasi data
 - Streamlit untuk pembuatan antarmuka aplikasi prediksi

10 3.2.2 Bahan Penelitian

Bahan penelitian yang digunakan adalah dataset pengiriman e-commerce yang diperoleh dari platform Kaggle. Dataset ini berisi data historis pengiriman dengan berbagai atribut yang berkaitan dengan pelanggan, produk, dan proses pengiriman. Dataset tersebut digunakan sebagai data latih dan data uji dalam proses pemodelan machine learning.

4 3.3 Metode Penelitian

Metode penelitian yang digunakan dalam penelitian ini adalah metode eksperimen dengan pendekatan kuantitatif. Metode ini digunakan untuk menguji dan membandingkan kinerja algoritma Logistic Regression dan XGBoost dalam memprediksi keterlambatan pengiriman e-commerce. Tahapan penelitian dilakukan secara sistematis mulai dari pengumpulan data hingga evaluasi hasil model.

18 3.4 Metode Pengumpulan Data

Data yang digunakan dalam penelitian ini diperoleh dari dataset publik yang tersedia di Kaggle. Dataset tersebut dikumpulkan oleh pihak ketiga dan telah digunakan dalam berbagai penelitian sebelumnya. Proses pengumpulan data dilakukan dengan mengunduh dataset, kemudian melakukan seleksi atribut yang relevan dengan tujuan penelitian.

3.5 Tahapan Penelitian

3.5.1 Data Preprocessing

39 Tahap preprocessing data bertujuan 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. Penghapusan atribut yang tidak relevan dengan proses prediksi.
3. Transformasi data kategorikal menjadi bentuk numerik menggunakan teknik encoding.
4. Normalisasi atau standarisasi data numerik untuk meningkatkan performa model.

3.5.2 Pembagian Data 6

Dataset dibagi menjadi data latih (training data) dan data uji (testing data). Data latih digunakan untuk melatih model machine learning, sedangkan data uji digunakan untuk mengukur kinerja model yang telah dilatih. Pembagian data dilakukan dengan proporsi tertentu agar model dapat dievaluasi secara objektif.

3.5.3 Pembuatan dan Pelatihan Model 63

Pada tahap ini dilakukan pembuatan model prediksi menggunakan dua algoritma, yaitu Logistic Regression dan XGBoost. Model Logistic Regression digunakan sebagai baseline model, sedangkan XGBoost digunakan sebagai model 53 banding dengan pendekatan ensemble. Proses pelatihan dilakukan menggunakan data latih yang telah dipersiapkan.

3.5.4 Evaluasi Model 49

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

3.5.5 Implementasi Model

Model terbaik hasil evaluasi kemudian diimplementasikan dalam bentuk aplikasi berbasis web menggunakan Streamlit. Aplikasi ini memungkinkan pengguna untuk memasukkan data pengiriman dan memperoleh hasil prediksi keterlambatan secara interaktif.

Dengan tahapan metodologi tersebut, penelitian ini diharapkan dapat menghasilkan model prediksi keterlambatan pengiriman e-commerce yang akurat dan dapat digunakan sebagai sistem pendukung keputusan.

22 HASIL DAN PEMBAHASAN

4.1 Deskripsi Data

Dataset yang digunakan dalam penelitian ini merupakan dataset pengiriman e-commerce yang diperoleh dari platform Kaggle. Dataset tersebut berisi data historis pengiriman barang dengan berbagai atribut yang berkaitan dengan pengiriman, produk, serta proses pengiriman. Jumlah data yang digunakan dalam penelitian ini adalah sebanyak 11.000 data, yang kemudian dibagi menjadi data latih dan data uji.

Atribut yang digunakan dalam penelitian ini antara lain warehouse block, mode of shipment, customer care calls, customer rating, cost of the product, prior purchases, product importance, gender, discount offered, dan weight in grams. Target klasifikasi pada penelitian ini adalah status keterlambatan pengiriman yang direpresentasikan dalam bentuk label biner, yaitu tepat waktu dan terlambat.

4.2 Hasil Preprocessing Data

Pada tahap preprocessing, dilakukan pembersihan data untuk memastikan tidak terdapat nilai kosong (missing value) dan data yang tidak valid. Selain itu, dilakukan transformasi data kategorikal menjadi data numerik menggunakan teknik encoding agar dapat diproses oleh algoritma 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

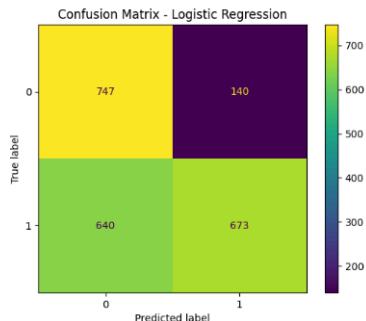
Model Logistic Regression digunakan sebagai baseline model dalam penelitian ini. Setelah proses pelatihan menggunakan data latih, model diuji menggunakan data uji untuk mengetahui kinerjanya.

30 Hasil evaluasi model Logistic Regression menunjukkan bahwa model mampu melakukan prediksi keterlambatan pengiriman dengan tingkat akurasi yang 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

8 Berdasarkan hasil evaluasi, model Logistic Regression menghasilkan nilai akurasi sebesar 64,55%, nilai F1-score sebesar 63,31%, dan nilai ROC-AUC sebesar 71,69%.



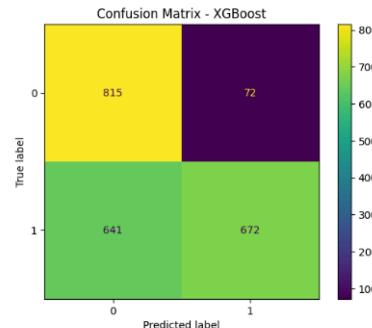
Gambar 1. Confusion Matrix Logistic Regression

Hasil confusion matrix menunjukkan bahwa model mampu mengenali data pengiriman tepat waktu dengan baik, namun masih terdapat sejumlah data keterlambatan yang belum terprediksi secara optimal.

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 evaluasi menunjukkan bahwa model XGBoost mampu menghasilkan performa yang lebih baik dibandingkan Logistic Regression.

Model XGBoost menghasilkan nilai akurasi sebesar 67,59%, nilai F1-score sebesar 65,34%, dan nilai 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

Perbandingan kinerja antara Logistic Regression dan XGBoost menunjukkan bahwa XGBoost memiliki performa yang lebih unggul pada seluruh metrik evaluasi yang digunakan. Hal ini menunjukkan bahwa algoritma XGBoost lebih efektif dalam menangani kompleksitas data pengiriman e-commerce yang bersifat non-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 [6].

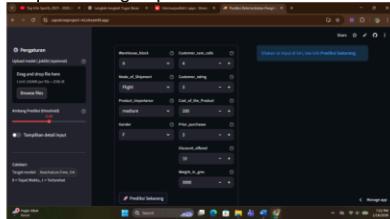
4.6 Pembahasan

Hasil penelitian ini menunjukkan bahwa penerapan machine learning dapat membantu dalam memprediksi keterlambatan pengiriman e-commerce secara efektif. Logistic Regression sebagai model baseline memberikan gambaran awal hubungan antar variabel, sedangkan XGBoost mampu menangkap pola yang lebih kompleks dalam data.

Fitur-fitur seperti diskon yang diberikan, berat barang, dan metode pengiriman memiliki pengaruh yang signifikan terhadap hasil prediksi

44 keterlambatan pengiriman. Hal ini sejalan dengan hasil penelitian terdahulu yang menyebutkan bahwa karakteristik produk dan proses distribusi merupakan faktor utama dalam menentukan ketepatan waktu pengiriman.

Implementasi model terbaik ke dalam aplikasi berbasis Streamlit menunjukkan bahwa hasil penelitian ini tidak hanya bersifat teoritis, tetapi juga dapat diaplikasikan secara praktis sebagai sistem pendukung keputusan.



Gambar 3. Antarmuka Web

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

Dengan demikian, hasil dari capstone project ini membuktikan bahwa pendekatan machine learning, khususnya XGBoost, memiliki potensi yang besar dalam meningkatkan efisiensi dan kualitas layanan logistik pada sistem e-commerce.

4 KESIMPULAN DAN SARAN

5.1 Kesimpulan

Berdasarkan hasil implementasi dan evaluasi yang telah dilakukan pada capstone project ini, dapat diambil beberapa kesimpulan 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 unggul 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 menggunakan 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.

19 5.2 Saran

Berdasarkan hasil capstone project yang telah dilakukan, terdapat beberapa saran yang dapat dijadikan bahan pengembangan selanjutnya, antara :

1. Penelitian selanjutnya dapat menggunakan dataset yang lebih besar dan beragam agar model memiliki kemampuan generalisasi yang lebih baik.
2. Dapat dilakukan eksperimen dengan algoritma machine learning lainnya, seperti Random Forest, LightGBM, atau Neural Network, untuk memperoleh performa 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] A. Haris, M. Sholeh, L. Muflihah, 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.
- [2] 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.
- [3] I. Azis *et al.*, "Penentuan Faktor Kemiskinan Indonesia Menggunakan Regresi Logistik," 2023. [Online]. Available: <http://www.ojs.unm.ac.id/mathcos>
- [4] 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>
- [5] 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.
- [6] 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.



PRIMARY SOURCES

1	ejournal.unibba.ac.id Internet Source	4%
2	Submitted to Universitas Budi Luhur Student Paper	1%
3	Asrul Abdullah, Della Udy Khairah, Menur Wahyu Pangestika. "COMPARISON OF RANDOM FOREST AND XGBOOST ALGORITHMS IN CREDIT CARD FRAUD CLASSIFICATION", Jurnal CoSciTech (Computer Science and Information Technology), 2025 Publication	1%
4	docplayer.info Internet Source	1%
5	Submitted to Forum Perpustakaan Perguruan Tinggi Indonesia Jawa Timur II Student Paper	1%
6	Submitted to STT PLN Student Paper	1%
7	jurnal.itg.ac.id Internet Source	1%
8	Submitted to Universitas Muhammadiyah Sumatera Utara Student Paper	1%
9	Submitted to University of Derby Student Paper	1%
10	digilib.unila.ac.id Internet Source	1%
eprints.utdi.ac.id		

11	Internet Source	1 %
12	adoc.pub Internet Source	1 %
13	Panji Pangestu Saputra, Hasbi Firmansyah, Rizki Prasetyo Tulodo, Priyo Haryoko, Wahyu Asriyani. "Penerapan Algoritma k-Nearest Neighbor untuk Klasifikasi Kondisi Lingkungan Pertanian Berbasis IoT", Jurnal Pengabdian Masyarakat dan Riset Pendidikan, 2025 Publication	1 %
14	repository.pnj.ac.id Internet Source	1 %
15	journal.mediapublikasi.id Internet Source	1 %
16	Submitted to Universitas Muhammadiyah Kotabumi Student Paper	1 %
17	publikasi.uylindo.ac.id Internet Source	1 %
18	lembagakita.org Internet Source	1 %
19	text-id.123dok.com Internet Source	1 %
20	Submitted to Academic Library Consortium Student Paper	1 %
21	www.ipm2kpe.or.id Internet Source	1 %
22	Submitted to Universitas Amikom Student Paper	<1 %
23	repo.bunghatta.ac.id Internet Source	<1 %
24	repository.its.ac.id Internet Source	

		<1 %
25	Christine Eirene, Dian Syafitri, Neny Sulistianingsih, Khasnur Hidjah, Hairani Hairani. "Classification of Learning Styles of Junior High School Students Using Random Forest & XGBoost Algorithm", Jurnal Bumigora Information Technology (BITe), 2025	<1 %
	Publication	
26	eprints.unram.ac.id	<1 %
	Internet Source	
27	Submitted to Universitas Sumatera Utara	<1 %
	Student Paper	
28	jpti.journals.id	<1 %
	Internet Source	
29	repositori.usu.ac.id	<1 %
	Internet Source	
30	Indri Nurfiani, Jumadi Jumadi, Muhammad Deden Firdaus. "PEMANFAATAN STFT DAN CNN DALAM PENGOLAHAN DATA SUARA UNTUK MENKLASIFIKASIKA SUARA BATUK", Rabit : Jurnal Teknologi dan Sistem Informasi Univrab, 2024	<1 %
	Publication	
31	ejurnal.seminar-id.com	<1 %
	Internet Source	
32	eprints.umpo.ac.id	<1 %
	Internet Source	
33	jurnal.stmikroyal.ac.id	<1 %
	Internet Source	
34	kc.umn.ac.id	<1 %
	Internet Source	
35	library.unibba.ac.id	<1 %
	Internet Source	

36	Submitted to Universitas Muhammadiyah Purwokerto	<1 %
37	artikelpendidikan.id	<1 %
38	repository.uin-suska.ac.id	<1 %
39	securityphresh.com	<1 %
40	www.journal.irpi.or.id	<1 %
41	www.researchgate.net	<1 %
42	Muh. Faried Muchtar, Rahma Laila, Dwi Shinta, H. M. Yazdi Pusadan. "Perbandingan Algoritma Naïve bayes Dan Support Vektor Machine Untuk Klasifikasi Status Stunting Pada Balita", The Indonesian Journal of Computer Science, 2024	<1 %
43	Suriyanti Suriyanti, Ramlawati Ramlawati, Nur Amaliah, Dwi Putri Asis. "Analisis Beban Kerja Kurir Terhadap Ketepatan Pengiriman Barang Pada PT. Indah Logistik Cargo Cabang Makassar", RIGGS: Journal of Artificial Intelligence and Digital Business, 2026	<1 %
44	digilib.uinsby.ac.id	<1 %
45	id.123dok.com	<1 %
46	manueluels418418.post-blogs.com	<1 %

47	Internet Source	<1 %
48	Ferry Khusnil Arief, Nurul Kholidiah, Rifki Mistahul Munir, Rahma Yunita. "Comparative Analysis of CNN Architectures (ResNet152V2, InceptionV3, and Xception) in Animal Species Classification", Indonesian Journal of Interdisciplinary Research in Science and Technology, 2025 Publication	<1 %
49	Ichsani Mursidah, Remi Sanjaya, Bambang Yulianto, Dhian Sweetania, Puji Sularsih. "Klasifikasi Sentimen Google Play Store Aplikasi ChatGPT Berbahasa Indonesia Berbasis IndoBERT", Jurnal Minfo Polgan, 2025 Publication	<1 %
50	Maria Anjelina Domu Tukan, Alfian Nara Weking, Dominikus Boli Watomakin. "Implementasi Algoritma K-Nearest Neighbors untuk Rekomendasi Diet Tinggi Serat dalam Mencegah Penyakit Jantung", RIGGS: Journal of Artificial Intelligence and Digital Business, 2025 Publication	<1 %
51	Syahrul Sobari, Ade Irma Purnamasari, Agus Bahtiar, Kaslani Kaslani. "MENINGKATKAN MODEL PREDIKSI KELULUSAN SANTRI TAHFIDZ DI PONDOK PESANTREN AL-KAUTSAR MENGGUNAKAN ALGORITMA RANDOM FOREST", Jurnal Informatika dan Teknik Elektro Terapan, 2025 Publication	<1 %
52	asescongress.com Internet Source	<1 %
53	docshare.tips Internet Source	<1 %

54	ejournal.uksw.edu Internet Source	<1 %
55	ejournal3.undip.ac.id Internet Source	<1 %
56	ichaimucu.wordpress.com Internet Source	<1 %
57	journal.unpacti.ac.id Internet Source	<1 %
58	jurnal.kominfo.go.id Internet Source	<1 %
59	jurnal.polibatam.ac.id Internet Source	<1 %
60	repository.usni.ac.id Internet Source	<1 %
61	www.fitb.itb.ac.id Internet Source	<1 %
62	Muhammad Nasim, Alda Cendekia Siregar, Rachmat Wahid Saleh Insani. "KLASIFIKASI OBESITAS DENGAN ALGORITMA C5.0 BERDASARKAN POLA MAKAN DAN KONDISI FISIK", Jurnal Khatulistiwa Informatika, 2024 Publication	<1 %
63	Rahmat Satria Buana, Windu Gata, Alda Zevana Putri Widodo, Hendra Setiawan, Khairunisa Hilyati. "Analisis Sentimen pada Komen Twitter Pawang Hujan Mandalika dengan Support Vector Machine (SVM) dan Naïve Bayes", Jurnal JTAK (Jurnal Teknologi Informasi dan Komunikasi), 2023 Publication	<1 %
64	Edo Ridho Lidinillah, Tatang Rohana, Ayu Ratna Juwita. "Analisis sentimen twitter terhadap steam menggunakan algoritma logistic regression dan support vector	<1 %

machine", TEKNOSAINS : Jurnal Sains, Teknologi dan Informatika, 2023

Publication

-
- 65 Mhd Arief Hasan, Novia Bimby. "Analisis Sentimen Publik Terhadap Kenaikan Pajak PPN di Indonesia Tahun 2024 Menggunakan Algoritma Machine Learning", JURNAL FASILKOM, 2025 <1 %
- Publication
-
- 66 Muhammad Ikhsan Anugrah, Junta Zeniarja, Dicky Setiawan Setiawan. "Peningkatan Performa Model Hard Voting Classifier dengan Teknik Oversampling ADASYN pada Penyakit Diabetes", Edumatic: Jurnal Pendidikan Informatika, 2024 <1 %
- Publication
-

Exclude quotes Off
Exclude bibliography On

Exclude matches Off