

PERBANDINGAN KINERJA ALGORITMA *DECISION TREE* DAN *RANDOM FOREST* UNTUK KLASIFIKASI PRODUK LARIS E-COMMERCE

FERDI B. M. ERARI



**PROGRAM SARJANA ILMU KOMPUTER
SEKOLAH SAINS DATA, MATEMATIKA, DAN INFORMATIKA
INSTITUT PERTANIAN BOGOR
BOGOR
2025**

PERNYATAAN MENGENAI SKRIPSI DAN SUMBER INFORMASI SERTA PELIMPAHAN HAK CIPTA

Dengan ini saya menyatakan bahwa skripsi dengan judul “Perbandingan Kinerja Algoritma *Decision Tree* dan *Random Forest* untuk Klasifikasi Produk Laris E-Commerce” adalah karya saya dengan arahan dari dosen pembimbing dan belum diajukan dalam bentuk apa pun kepada perguruan tinggi mana pun. Sumber informasi yang berasal atau dikutip dari karya yang diterbitkan maupun tidak diterbitkan dari penulis lain telah disebutkan dalam teks dan dicantumkan dalam Daftar Pustaka di bagian akhir skripsi ini.

Dengan ini saya melimpahkan hak cipta dari karya tulis saya kepada Institut Pertanian Bogor.

Bogor, Agustus 2025

Ferdi B. M. Erari
G64180112

ABSTRAK

FERDI B. M. ERARI. *Perbandingan Kinerja Algoritma Decision Tree dan Random Forest untuk Klasifikasi Produk Laris E-Commerce*. Dibimbing oleh SONY HARTONO WIJAYA dan FIRMAN ARDIANSYAH

Penelitian ini membandingkan kinerja algoritma Decision Tree dan Random Forest dalam mengklasifikasikan produk laris menggunakan data penjualan e-commerce periode November 2020–Oktober 2023. Data dianalisis melalui eksplorasi, praproses (imputasi nilai kosong, encoding fitur kategorikal, pembagian data 70:30), dan pelabelan produk laris berdasarkan skor gabungan frekuensi transaksi dan total unit terjual dengan ambang kuartil ketiga (Q3). Model awal dibangun menggunakan parameter bawaan, kemudian dioptimasi melalui hyperparameter tuning dengan GridSearchCV dan RandomizedSearchCV pada parameter kritis. Evaluasi menggunakan akurasi, presisi, recall, F1-score, dan confusion matrix. Hasil menunjukkan Random Forest default unggul dengan akurasi 92,56%, presisi 92,10%, recall 92,56%, dan F1-score 92,04%. Setelah tuning, akurasi menjadi 91,82%, presisi 92,59%, recall 98,39%, dan F1-score 95,40%. Peningkatan recall menandakan kemampuan deteksi produk laris yang lebih baik, meski akurasi sedikit menurun. Random Forest, khususnya setelah tuning, direkomendasikan untuk mendukung pengambilan keputusan berbasis data di e-commerce, seperti manajemen inventaris dan strategi pemasaran.

Kata kunci: *decision tree, e-commerce, hyperparameter tuning*, klasifikasi, *random forest*, produk laris.

ABSTRACT

FERDI B. M. ERARI. A Comparative Study of Decision Tree and Random Forest Algorithms for Best-Selling Product Classification in E-Commerce. Supervised by SONY HARTONO WIJAYA and FIRMAN ARDIANSYAH.

This study compares the performance of Decision Tree and Random Forest algorithms in classifying best-selling products using e-commerce sales data from November 2020 to October 2023. The data underwent exploratory analysis, preprocessing (missing value imputation, categorical feature encoding, 70:30 train-test split), and product labeling based on a combined score of transaction frequency and total units sold, with the third quartile (Q3) as the threshold. Baseline models were built using default parameters, then optimized through hyperparameter tuning with GridSearchCV and RandomizedSearchCV on key parameters. Evaluation metrics included accuracy, precision, recall, F1-score, and confusion matrix analysis. Results showed that the default Random Forest outperformed Decision Tree, achieving 92.56% accuracy, 92.10% precision, 92.56% recall, and a 92.04% F1-score. After tuning, Random Forest achieved 91.82% accuracy, 92.59% precision, 98.39% recall, and a 95.40% F1-score. The increase in recall indicates improved ability to detect best-selling products, although accuracy slightly decreased due to more false positives. Random Forest, particularly after tuning, is

recommended for data-driven decision-making in e-commerce, such as inventory management and marketing strategy.

Keywords: best-selling products, classification, decision tree, e-commerce, hyperparameter tuning, random forest,

@Hak cipta milik IPB University

IPB University



Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.

Judul Skripsi : Perbandingan Kinerja Algoritma *Decision Tree* dan *Random Forest* untuk Klasifikasi Produk Laris E-Commerce

Nama : Ferdi B. M. Erari

NIM : G64180112

Disetujui oleh

Pembimbing 1:

Dr. Sony Hartono Wijaya, S.Kom., M.Kom.



Pembimbing 2:

Firman Ardiansyah, S.Kom., M.Si.



Diketahui oleh

Ketua Program Sarjana Ilmu Komputer:

Dr. Sony Hartono Wijaya, S.Kom., M.Kom.

NIP. 19810809 200812 1 002



Tanggal Ujian:

11 Agustus 2025

Tanggal Lulus:



PRAKATA

Segala puji syukur hanya bagi Tuhan Yesus Kristus atas segala kasih karunia dan penyertaan-Nya sehingga karya ilmiah ini berhasil diselesaikan. Tema yang dipilih dalam penelitian yang dilaksanakan sejak bulan Mei 2025 sampai bulan Juli 2025 ini adalah hasil karya saya se, dengan judul “Perbandingan Kinerja Algoritma *Decision Tree* dan *Random Forest* untuk Klasifikasi Produk Laris E-Commerce”.

Terima kasih penulis ucapkan kepada para pembimbing, Dr. Sony Hartono Wijaya, S.Kom., M.Kom. dan Firman Ardiansyah, S.Kom, M.Si., yang telah membimbing dan banyak memberi saran. Ucapan terima kasih juga disampaikan kepada pembimbing akademik, moderator seminar, dan penguji luar komisi pembimbing. Ungkapan terima kasih juga disampaikan kepada ayah, ibu, serta seluruh keluarga yang telah memberikan dukungan, doa, dan kasih sayangnya hingga terselesaikannya penelitian ini. Penelitian ini bisa diselesaikan dengan baik tentunya dengan adanya bantuan beberapa pihak, yaitu kedua orang tua, Derek Kornelis Erari, dan Ibu Ferlin Tanatty, dan juga kakak sepupu saya Morgan Wayne Luther Sawaki dan adik saya Devilia Stin Erari yang telah memberikan dukungan serta doa bagi penulis.

Semoga karya ilmiah ini bermanfaat bagi pihak yang membutuhkan dan bagi kemajuan ilmu pengetahuan.

Bogor, Agustus 2025

Ferdi B. M. Erari



DAFTAR ISI

DAFTAR TABEL	vi
DAFTAR GAMBAR	vi
I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	1
1.3 Tujuan	1
1.4 Manfaat	2
1.5 Ruang Lingkup	2
II TINJAUAN PUSTAKA	3
2.1 E-Commerce dan Data Penjualan	3
2.2 <i>Machine Learning</i> dan Klasifikasi Produk	3
2.3 Algoritma <i>Decision Tree</i>	3
2.4 Algoritma <i>Random Forest</i>	3
2.5 <i>Hyperparameter Tuning</i>	4
2.6 Evaluasi	4
III METODE	5
3.1 <i>Exploratory Data Analysis(EDA)</i>	5
3.2 Praproses	6
3.3 Pemodelan	6
3.4 <i>Hyperparameter Tuning</i>	7
3.5 Evaluasi	7
IV HASIL DAN PEMBAHASAN	9
4.1 Pengumpulan Data	9
4.2 <i>Exploratory Data Analysis</i>	10
4.3 Praproses	13
4.4 Pemodelan <i>Decision Tree</i> dan <i>Random Forest</i>	14
4.5 <i>Hyperparameter Tuning</i>	19
V SIMPULAN DAN SARAN	21
5.1 Simpulan	21
5.2 Saran	21
DAFTAR PUSTAKA	22

DAFTAR TABEL

1	Detil atribut data	9
2	Statistik deskriptif kolom numerik terkait penjualan	10
3	Kuartil skor gabungan per label dan ambang Q3	11
4	Top-10 produk berdasarkan bkor gabungan ($\text{norm_freq} + \text{norm_qty}$)	12
5	Evaluasi <i>Decision Tree</i> dan <i>Random Forest</i>	18
6	Evaluasi <i>Random Forest (Tuned)</i>	19

DAFTAR GAMBAR

1	Metodologi penelitian	5
2	Boxplot produk Tidak Laris vs Laris	13
3	Skema pohon keputusan pada fitur terproses	15
4	<i>Confusion matrix decision tree (default)</i>	17
5	<i>Confusion matrix Random Forest (Default)</i>	18
6	<i>Confusion matrix Random Forest (Tuned)</i>	20