

**ANALISIS PERBANDINGAN ALGORITMA LOGISTIC
REGRESSION DAN RANDOM FOREST PADA
PREDIKSI PEMBELIAN ONLINE**

SKRIPSI

Diajukan untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Informatika



disusun oleh

WIJAYANTO AGUNG WIBOWO

22.11.4552

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2026

**ANALISIS PERBANDINGAN ALGORITMA LOGISTIC
REGRESSION DAN RANDOM FOREST PADA
PREDIKSI PEMBELIAN ONLINE**

SKRIPSI

Diajukan untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Informatika



disusun oleh

WIJAYANTO AGUNG WIBOWO

22.11.4552

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2026

HALAMAN PERSETUJUAN

SKRIPSI

ANALISIS PERBANDINGAN ALGORITMA LOGISTIC
REGRESSION DAN RANDOM FOREST PADA
PREDIKSI PEMBELIAN ONLINE

yang disusun dan diajukan oleh

Wijayanta Agung Wibowo

22.11.4552

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 11 Mei 2026

Dosen Pembimbing,



Ike Verawati, S.Kom., M.Kom.
NIK. 190302237

HALAMAN PENGESAHAN

SKRIPSI

ANALISIS PERBANDINGAN ALGORITMA LOGISTIC
REGRESSION DAN RANDOM FOREST PADA
PREDIKSI PEMBELIAN ONLINE

yang disusun dan diajukan oleh

Wijayanto Agung Wibowo

22.11.4552

Telah dipertahankan di depan Dewan Penguji
pada tanggal 11 Mei 2026

Susunan Dewan Penguji

Nama Penguji

Tanda Tangan

Dr. Andi Sunyoto, S.Kom., M.Kom.,
NIK. 190302052

Majid Rahardi, S.Kom., M.Eng.
NIK. 190302393

Ike Verawati, S.Kom., M.Kom.
NIK. 190302237



Skrripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 11 Mei 2026

DEKAN FAKULTAS ILMU KOMPUTER



Prof. Dr. Kusriani, M.Kom.
NIK. 190302106

HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

Nama mahasiswa : Wijayanto Agung Wibowo
NIM : 22.11.4552

Menyatakan bahwa Skripsi dengan judul berikut:

Analisis Perbandingan Algoritma Logistic Regression dan Random Forest pada Prediksi Pembelian Online

Dosen Pembimbing : Ike Vrawati, S.Kom., M.Kom.

1. Karya tulis ini adalah benar-benar ASLI dan BELUM PERNAH diajukan untuk mendapatkan gelar akademik, baik di Universitas AMIKOM Yogyakarta maupun di Perguruan Tinggi lainnya.
2. Karya tulis ini merupakan gagasan, rumusan dan penelitian SAYA sendiri, tanpa bantuan pihak lain kecuali arahan dari Dosen Pembimbing.
3. Dalam karya tulis ini tidak terdapat karya atau pendapat orang lain, kecuali secara tertulis dengan jelas dicantumkan sebagai acuan dalam naskah dengan disebutkan nama pengarang dan disebutkan dalam Daftar Pustaka pada karya tulis ini.
4. Perangkat lunak yang digunakan dalam penelitian ini sepenuhnya menjadi tanggung jawab SAYA, bukan tanggung jawab Universitas AMIKOM Yogyakarta.
5. Pernyataan ini SAYA buat dengan sesungguhnya, apabila di kemudian hari terdapat penyimpangan dan ketidakbenaran dalam pernyataan ini, maka SAYA bersedia menerima SANKSI AKADEMIK dengan pencabutan gelar yang sudah diperoleh, serta sanksi lainnya sesuai dengan norma yang berlaku di Perguruan Tinggi.

Yogyakarta, 11 Mei 2026

Yang Menyatakan,



Wijayanto Agung Wibowo

HALAMAN PERSEMBAHAN

Segala puji dan syukur kepada hadirat Tuhan Yang Maha Esa atas segala rahmat dan karunia-Nya atas kelancaran dalam pembuatan skripsi ini dengan baik. Oleh karena itu tidak mengurangi rasa hormat saya. Saya persembahkan skripsi ini sebagai ungkapan rasa hormat saya kepada :

1. Kedua orang tua tercinta yaitu Ibu Siti Rahayu dan Bapak Rokhidin yang senantiasa menjadi sumber kekuatan, doa, dan kasih sayang yang tiada henti. Ucapan terima kasih yang sebesar-besarnya saya sampaikan atas segala bentuk pengorbanan, dukungan, serta kepercayaan yang diberikan selama ini.
2. Ibu Ike Verawati, S.Kom., M.Kom., selaku dosen pembimbing, yang dengan penuh kesabaran membimbing, memberikan arahan, serta ilmu yang sangat berharga sepanjang proses penyusunan skripsi ini. Atas segala bimbingan dan perhatian yang telah diberikan, saya mengucapkan terima kasih yang setulus-tulusnya.
3. Kakak dan adik saya yang senantiasa memberikan dukungan, semangat, serta bantuan selama proses penyusunan skripsi ini. Terima kasih atas perhatian, doa, dan kebersamaan yang selalu diberikan sehingga penulis dapat menyelesaikan skripsi ini dengan baik.
4. Kepada diri saya pribadi, apresiasi yang setulusnya saya sampaikan atas keteguhan hati dan usaha yang telah dilakukan dalam menjalani setiap proses, mulai dari perkuliahan hingga penyusunan skripsi ini. Meski tidak selalu mudah, setiap tantangan yang dihadapi telah memberikan pelajaran yang bermakna untuk membentuk pribadi yang kuat dan matang. Semoga pencapaian ini menjadi awal dari langkah menuju masa depan yang lebih cerah dan bermakna.

KATA PENGANTAR

Segala puji dan syukur penulis panjatkan ke hadirat Tuhan Yang Maha Esa atas segala rahmat dan karunia-Nya, sehingga penulis dapat menyelesaikan penyusunan skripsi ini dengan baik dan lancar. Skripsi yang berjudul "Analisis Perbandingan Algoritma Logistic Regression dan Random Forest pada Prediksi Pembelian Online" ini disusun sebagai salah satu syarat untuk memperoleh gelar Sarjana pada Program Studi Teknik Komputer, Fakultas Ilmu Komputer, Universitas Amikom Yogyakarta.

Terselesaikannya skripsi ini tidak lepas dari bantuan dan dukungan berbagai pihak, untuk itu penulis menyampaikan terima kasih yang sebesar-besarnya kepada:

1. Ibu Ike Verawati, S.Kom., M.Kom., selaku dosen pembimbing, yang telah dengan sabar membimbing, memberikan masukan, serta arahan yang sangat berharga sepanjang proses penyusunan skripsi ini.
2. Kedua orang tua, serta kakak dan adik tercinta yang senantiasa memberikan doa, dukungan, dan kasih sayang yang tulus serta tak ternilai selama masa studi hingga terselesaikannya skripsi ini.
3. Seluruh dosen dan staf Program Studi Teknik Komputer, Fakultas Ilmu Komputer, Universitas Amikom Yogyakarta, yang telah membagikan ilmu, wawasan, dan pengalaman selama masa perkuliahan.

Sebagai penutup, penulis menyadari bahwa skripsi ini masih memiliki berbagai keterbatasan dan belum sepenuhnya sempurna. Oleh karena itu, penulis sangat mengharapkan kritik dan saran yang bersifat membangun sebagai bahan perbaikan di masa yang akan datang. Penulis juga berharap karya ini dapat memberikan manfaat serta menjadi kontribusi yang positif bagi pengembangan ilmu pengetahuan, khususnya pada bidang terkait.

Yogyakarta, 27 April 2026

Penulis

DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN PERSETUJUAN	ii
HALAMAN PENGESAHAN	iii
HALAMAN PERNYATAAN KEASLIAN SKRIPSI	iv
HALAMAN PERSEMBAHAN	v
KATA PENGANTAR	vi
DAFTAR ISI	vii
DAFTAR TABEL	xi
DAFTAR GAMBAR	xii
DAFTAR LAMPIRAN	xiii
DAFTAR LAMBANG DAN SINGKATAN	xiv
DAFTAR ISTILAH	xv
INTISARI	xvi
<i>ABSTRACT</i>	xvii
BAB I PENDAHULUAN	18
1.1 Latar Belakang	18
1.2 Rumusan Masalah	20
1.3 Batasan Masalah	20
1.4 Tujuan Penelitian	20
1.5 Manfaat Penelitian	21
1.6 Sistematika Penulisan	22

BAB II TINJAUAN PUSTAKA	24
2.1 Studi Literatur	24
2.2 Dasar Teori	30
2.2.1 <i>E-commerce</i>	30
2.2.2 <i>Data Mining</i>	31
2.2.3 Klasifikasi	32
2.2.4 Algoritma Klasifikasi	32
2.2.4.1 Logistic Regression	32
2.2.4.2 Random Forest	34
2.2.5 <i>Preprocessing Data</i>	35
2.2.5.1 <i>Cleaning data</i>	35
2.2.5.2 <i>Missing Values</i>	35
2.2.5.3 <i>Label Encoding</i>	36
2.2.5.4 Normalisasi Data	36
2.2.5.5 SMOTE	36
2.2.6 <i>Stratified Shuffle Split</i>	37
2.2.7 <i>Recursive Feature Elimination</i>	37
2.2.7 <i>Confusion matrix</i>	37
2.2.8 <i>Hyperparameter Tuning</i>	38
BAB III METODE PENELITIAN	39
3.1 Objek Penelitian	39
3.2 Alur Penelitian	39
3.3 Alat dan Bahan	42
3.3.1 Data Penelitian	42

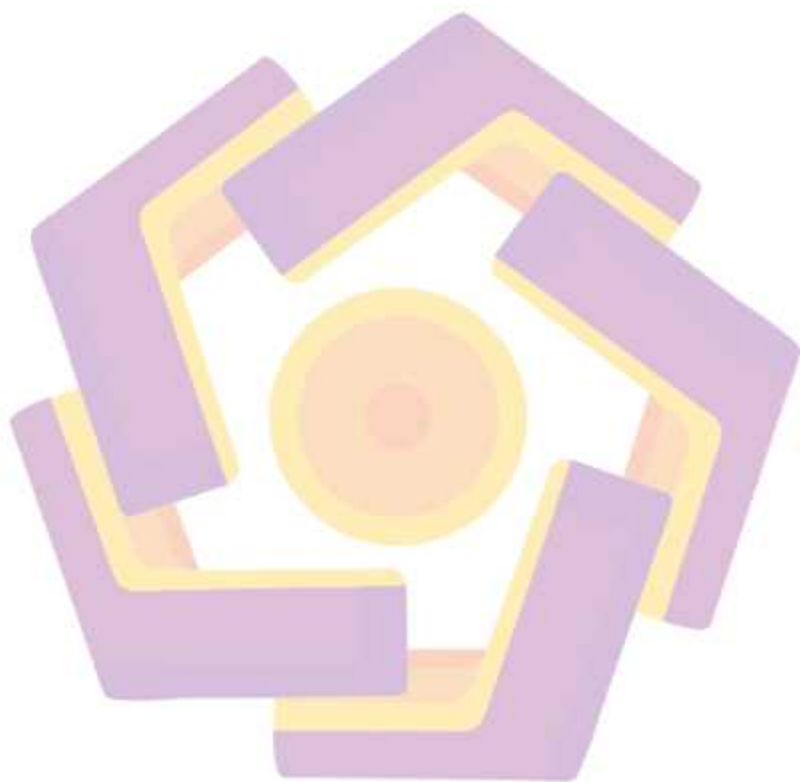
3.3.2	Alat/instrumen Penelitian	43
BAB IV HASIL DAN PEMBAHASAN		45
4.1	Akuisisi Data	45
4.2	EDA	47
4.3	Data cleaning	50
4.4	Encoding	52
4.5	<i>Split Dataset</i>	53
4.6	Normalisasi Data	54
4.7	SMOTE	54
4.8	Pemilihan Fitur	57
4.8.1	RFE Logistic Regression	57
4.8.2	RFE Random Forest	59
4.9	Modeling	60
4.9.1	Modeling Logistic Regression	61
4.9.2	Modelling Random Forest	62
4.10	<i>Hyperparameter Tuning</i>	62
4.10.1	<i>Hyperparameter</i> Logistic Regression	63
4.10.2	<i>Hyperparameter</i> Random Forest	63
4.11	Hasil Evaluasi	65
4.12	Deployment	72
BAB V PENUTUP		74
5.1	Kesimpulan	74
5.2	Saran	74

REFERENSI

76

LAMPIRAN

80



DAFTAR TABEL

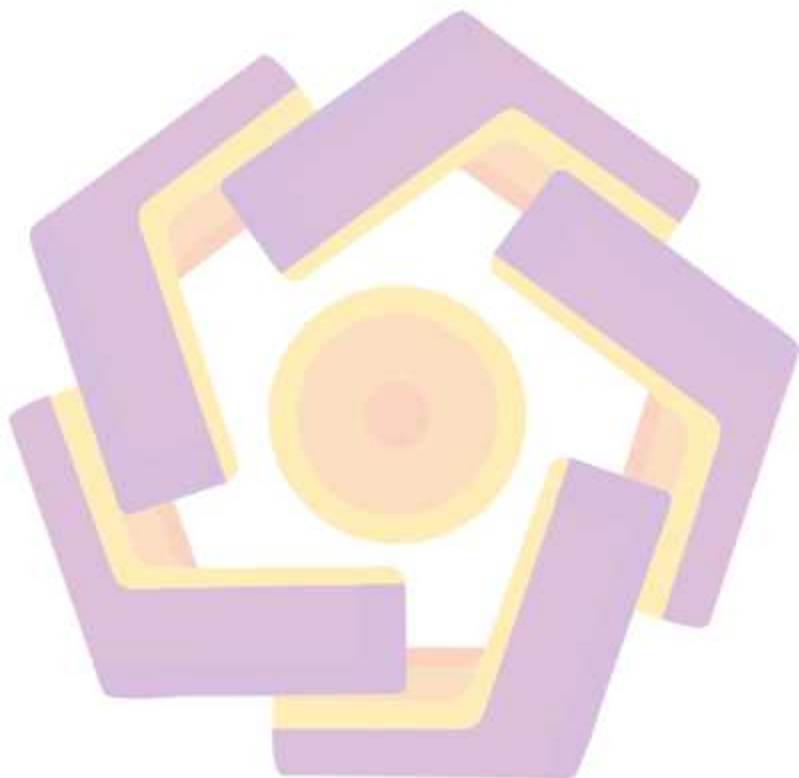
Tabel 2.1 Keaslian Penelitian	26
Tabel 3.1 Sampel data	42
Tabel 3.2 Perangkat keras	43
Tabel 3.3 Perangkat Lunak	44
Tabel 4.1 Sampel data	45
Tabel 4.2 Deskripsi Keterangan Fitur	47
Tabel 4.3 Hasil Perbandingan Data Setelah SMOTE	56
Tabel 4.4 Rincian Performa RFE Logistic Regression	58
Tabel 4.5 Rincian Performa RFE Random Forest	60
Tabel 4.6 <i>Hyperparameter</i> Terbaik Logistic Regression	63
Tabel 4.7 <i>Hyperparameter</i> Terbaik Random Forest	64
Tabel 4.8 Perbandingan Logistic Regression setelah <i>Hyperparameter Tuning</i>	70
Tabel 4.9 Perbandingan Random Forest setelah <i>Hyperparameter Tuning</i>	70
Tabel 4.10 Perbandingan Model Logistic Regression dan Random Forest	71

DAFTAR GAMBAR

Gambar 2. 1 Proses Data Mining	31
Gambar 2.2 Hasil Fungsi Logistic Regression	33
Gambar 2.3 Tahapan Random Forest	34
Gambar 3.1 Alur Penelitian	39
Gambar 4.1 Alur Proses EDA	48
Gambar 4.2 Output Data Information	48
Gambar 4.3 Output Deskripsi Data	49
Gambar 4.4 Korelasi Antar Fitur	50
Gambar 4.5 Boxplot Awal Sebelum Ditangani	51
Gambar 4.6 Hasil Deteksi Outlier Pada Data	51
Gambar 4.7 Hasil Visualisasi Boxplot Setelah Ditangani	52
Gambar 4.8 Output Split Dataset	53
Gambar 4.9 Hasil Komposisi Variable Target	55
Gambar 4.10 Distribusi Variabel Target	55
Gambar 4.11 Distribusi Setelah SMOTE	56
Gambar 4.12 Visualisasi Grafik Performa RFE Logistic Regression	58
Gambar 4.13 Visualisasi Grafik Performa RFE Random Forest	59
Gambar 4.14 Pengukuran Performa Model Awal Logistic Regression	61
Gambar 4.15 Pengukuran Performa Model Awal Random Forest	62
Gambar 4.16 Tahapan Evaluasi Model	65
Gambar 4.17 Matrix Evaluasi <i>Hyperparameter</i> Logistic Regression	66
Gambar 4.18 Matrix Evaluasi <i>Hyperparameter</i> Random Forest	66
Gambar 4.19 Confussion Matrix <i>Hyperparameter</i> Logistic Regression	67
Gambar 4.20 Confussion Matrix <i>Hyperparameter</i> Random Forest	68
Gambar 4.21 Grafik AUC-ROC dari <i>Hyperparameter</i> Logistic Regression	69
Gambar 4.22 Grafik AUC-ROC dari <i>Hyperparameter</i> Random Forest	69
Gambar 4.23 Grafik AUC-ROC Model	72
Gambar 4.24 Halaman Web Prediksi Pembelian	73

DAFTAR LAMPIRAN

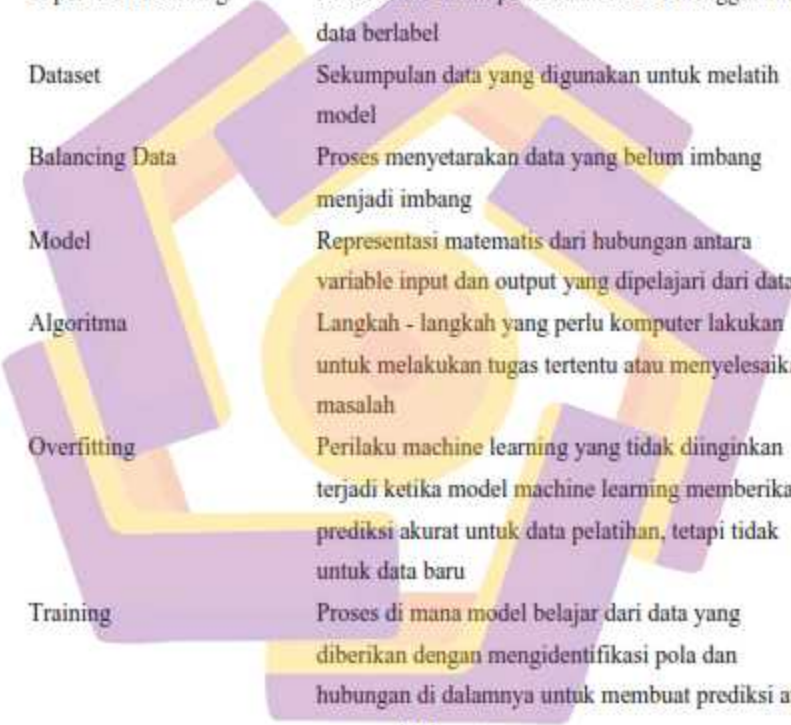
Lampiran 1. Dataset penelitian	80
Lampiran 2. Source code penelitian	80
Lampiran 3. Link deployment	80



DAFTAR LAMBANG DAN SINGKATAN

LR	Logistic Regression
RF	Random Forest
SMOTE	Synthetic Minority Over-sampling Technique
IQR	Interquartile Range
RFE	Recursive Feature Elimination
ROC	Receiver Operating Characteristic
AUC	Area Under the Curve
TP	True Positive
TN	True Negative
FP	False Positive
FN	False Negative

DAFTAR ISTILAH



Data Mining	Ilmu pengembangan algoritma dan model secara statistik
Machine Learning	Ilmu pengembangan algoritma dan model secara statistik
Supervised Learning	Pendekatan dalam pembuatan model menggunakan data berlabel
Dataset	Sekumpulan data yang digunakan untuk melatih model
Balancing Data	Proses menyetarakan data yang belumimbang menjadiimbang
Model	Representasi matematis dari hubungan antara variable input dan output yang dipelajari dari data
Algoritma	Langkah - langkah yang perlu komputer lakukan untuk melakukan tugas tertentu atau menyelesaikan masalah
Overfitting	Perilaku machine learning yang tidak diinginkan terjadi ketika model machine learning memberikan prediksi akurat untuk data pelatihan, tetapi tidak untuk data baru
Training	Proses di mana model belajar dari data yang diberikan dengan mengidentifikasi pola dan hubungan di dalamnya untuk membuat prediksi atau mengambil keputusan
Testing	Proses menguji kinerja model pada data yang tidak digunakan selama pelatihan untuk mengevaluasi seberapa baik model tersebut dapat melakukan prediksi pada data baru
<i>Confusion matrix</i>	Tabel yang digunakan dalam machine learning untuk mengevaluasi kinerja model klasifikasi

INTISARI

Niat pembelian online menjadi faktor penentu dalam kesuksesan strategi pemasaran digital pada *E-commerce*, khususnya di tengah persaingan pasar yang semakin kompetitif. Ketidakmampuan memprediksi perilaku pengunjung ini dapat berdampak pada inefisiensi biaya promosi dan hilangnya potensi retensi pelanggan. Oleh karena itu, penelitian ini bertujuan untuk membandingkan performa dua algoritma klasifikasi, yaitu Logistic Regression dan Random Forest, dalam memprediksi pembelian online menggunakan dataset Online Shoppers Intention dari UCI Machine Learning Repository. Proses penelitian meliputi eksplorasi data, *preprocessing* menggunakan teknik normalisasi dan Synthetic Minority Over-sampling Technique (SMOTE), kemudian kinerja ditingkatkan melalui proses tuning dengan GridSearchCV untuk menemukan kombinasi *hyperparameter* terbaik. Model diuji berdasarkan performa dalam memprediksi pembelian melalui berbagai metrik evaluasi. Hasil analisis menunjukkan bahwa Random Forest memiliki performa lebih unggul dengan akurasi 0.88, presisi 0.59, *Recall* 0.73, *F1-Score* 0.65, dan ROC-AUC 0.92. Sementara Logistic Regression mencatat akurasi 0.86, presisi 0.55, *Recall* 0.74, *F1-Score* 0.63, ROC-AUC 0.89. Temuan ini mengindikasikan keunggulan Random Forest dalam mengenali pola non-linear perilaku pengguna, sehingga lebih akurat dalam prediksi pembelian. Hasil penelitian diharapkan dapat menjadi referensi bagi praktisi dan perusahaan dalam pemilihan model Machine Learning yang optimal dan pengembangan strategi pemasaran digital berbasis data.

Kata kunci: E-commerce, Prediksi Pembelian Online, Logistic Regression, Random Forest, Machine Learning.

ABSTRACT

Online purchase intent is a key factor in the effectiveness of digital marketing strategies in e-commerce, particularly in the face of increasingly competitive markets. The lack of ability to predict visitor behavior can lead to inefficient promotional costs and the loss of potential customer retention. Therefore, this study aims to compare the performance of two classification algorithms, Logistic Regression and Random Forest, in predicting online purchases using the Online Shoppers Intention dataset from the UCI Machine Learning Repository. The research process includes data exploration, preprocessing using normalization techniques and the Synthetic Minority Over-sampling Technique (SMOTE), followed by performance tuning using GridSearchCV to identify the optimal hyperparameter combination. The models were evaluated based on their performance in predicting online purchases using various evaluation metrics. The analysis results show that Random Forest outperforms Logistic Regression with an accuracy of 0.88, precision of 0.59, recall of 0.73, F1-Score of 0.65, and ROC-AUC of 0.92. Meanwhile, Logistic Regression had the following results: accuracy 0.86, precision 0.55, recall 0.74, F1-Score 0.63, and ROC-AUC 0.89. These findings indicate the superiority of Random Forest in recognizing non-linear patterns in user behavior, making it more accurate in predicting purchases. The research results are expected to serve as a reference for practices and companies in selecting the optimal Machine Learning model and developing data-driven digital marketing strategies.

Keyword: E-commerce, Online Purchase Prediction, Logistic Regression, Random Forest, Machine Learning.