

Analisis Perbandingan Algoritma Linear Regression dan Gradient Boosting dalam Memprediksi Harga Smartphone Berdasarkan Spesifikasi Perangkat

Raihan^{1*}, Deni Risdiansyah², Erni³

^{1,2,3}Fakultas Teknik dan Informatika, Program Studi Informatika Kampus Kota Pontianak, Universitas Bina Sarana Informatika, Pontianak, Indonesia

Email: ^{1*}15220459@bsi.ac.id, ²deni.drx@bsi.ac.id, ³erx@bsi.ac.id

(*Email Corresponding Author: 15220459@bsi.ac.id)

Received: 30 Juni 2026 | Revision: 20 Juli 2026 | Accepted: 6 Agustus 2026

Abstrak

Pesatnya perkembangan teknologi informasi mentransformasikan *smartphone* menjadi komoditas ekonomi digital dengan fluktuasi harga yang tinggi akibat kompleksitas hubungan antarspesifikasi secara non-linear. Penelitian ini bertujuan untuk menguji efektivitas metode statistik tradisional dibandingkan dengan pendekatan *ensemble learning* dalam memprediksi harga *smartphone*. Analisis dilakukan melalui perbandingan kinerja antara algoritma *Linear Regression* dan *Gradient Boosting* berdasarkan lima komponen utama, yaitu RAM, Storage, Baterai, Brand, dan Resolusi Display menggunakan perangkat lunak *Orange Data Mining*. Penelitian ini memanfaatkan dataset sekunder sebesar 1.715 baris data dengan proporsi pembagian sampel 70% sebagai data latih (1.201 baris) dan 30% sebagai data uji (514 baris). Akurasi model diukur secara kuantitatif melalui metrik *R-Squared* (R^2), *Root Mean Square Error* (RMSE), dan *Mean Absolute Percentage Error* (MAPE). Hasil pengujian membuktikan bahwa algoritma *Gradient Boosting* memiliki kinerja lebih baik dengan mencatatkan nilai R^2 sebesar 0.998 (99.8%), RMSE sebesar 16.766, serta MAPE sebesar 0.014 (1.4%). Sementara itu, *Linear Regression* menghasilkan nilai R^2 sebesar 0.990 (99.0%), RMSE sebesar 34.933, dan MAPE sebesar 0.087 (8.7%). Hasil ini membuktikan hipotesis bahwa kelompok algoritma *ensemble learning* terbukti sangat adaptif dan presisi dalam mengekstrak interaksi non-linear yang rumit dari spesifikasi fisik perangkat tanpa menimbulkan gangguan *noise*.

Kata Kunci: Smartphone, *Linear Regression*, *Gradient Boosting*, *Orange Data Mining*, Prediksi Harga.

Abstract

The rapid development of information technology has transformed smartphones into a digital economy commodity with high price fluctuations due to the complexity of the relationship between specifications in a non-linear manner. This study aims to test the effectiveness of traditional statistical methods compared to the ensemble learning approach in predicting smartphone prices. The analysis was carried out through a performance comparison between the Linear Regression and Gradient Boosting algorithms based on five main components, namely RAM, Storage, Battery, Brand, and Display Resolution using Orange Data Mining software. This study utilizes a secondary dataset of 1,715 lines of data with a sample distribution proportion of 70% as training data (1,201 lines) and 30% as test data (514 lines). Model accuracy was measured quantitatively through the *R-Squared* (R^2), *Root Mean Square Error* (RMSE), and *Mean Absolute Percentage Error* (MAPE) metrics. The test results proved that the Gradient Boosting algorithm had better performance by recording an R^2 value of 0.998 (99.8%), RMSE of 16,766, and MAPE of 0.014 (1.4%). Meanwhile, Linear Regression resulted in an R^2 value of 0.990 (99.0%), RMSE of 34.933, and MAPE of 0.087 (8.7%). These results prove the hypothesis that ensemble learning algorithm groups are shown to be highly adaptive and precise in extracting complex non-linear interactions from the physical specifications of the device without causing noise disturbance.

Keywords: Smartphone, *Linear Regression*, *Gradient Boosting*, *Orange Data Mining*, Price Prediction

1. PENDAHULUAN

Berdasarkan laporan dari *International Data Corporation (IDC) Worldwide Quarterly Mobile Phone Tracker*, pengiriman *smartphone* secara global mengalami peningkatan sebesar 2,4% dibandingkan periode yang sama tahun sebelumnya, dengan total mencapai 331,7 juta unit di kuartal keempat tahun 2024 (4Q24). Tren ini menandai enam kuartal berturut-turut adanya pertumbuhan dalam pengiriman, yang sekaligus menutup tahun tersebut dengan tingkat kenaikan sebesar 6,4% dan total pengiriman mencapai 1,24 miliar unit. Pertumbuhan ini menunjukkan adanya pemulihan yang cukup signifikan setelah dua tahun sebelumnya mengalami penurunan yang cukup menantang. Prediksi pasar menunjukkan bahwa tren pertumbuhan akan berlanjut di tahun 2025, meskipun dengan laju yang lebih lambat, dikarenakan siklus pembaruan perangkat yang semakin meningkat dan terpenuhinya permintaan yang tertahan.

Globalisasi telah menjadi faktor utama yang memacu kemajuan pesat serta transformasi teknologi informasi dan komunikasi (TIK) di seluruh dunia [1]. Di era digital yang terus berkembang, internet telah menjadi fondasi utama dari perubahan dalam komunikasi modern, memungkinkan interaksi manusia tanpa terbatas oleh ruang dan waktu sebelumnya

[2]. Kemajuan teknologi digital ini turut memengaruhi berbagai aspek kehidupan masyarakat perkotaan, termasuk dalam akses informasi, cara berkomunikasi, dan aktivitas konsumsi [3]. Dengan perkembangan teknologi yang pesat, kebutuhan masyarakat terhadap perangkat teknologi, khususnya smartphone, meningkat secara signifikan [4].

Penggunaan smartphone di seluruh dunia telah menjadi bagian tak terpisahkan dari kehidupan sehari-hari. Saat ini, sulit sekali menemukan orang yang tidak menggunakan smartphone pintar dibandingkan mereka yang menggunakannya [5]. Smartphone tidak hanya berfungsi sebagai alat komunikasi, tetapi juga mengalami perkembangan fitur baru seperti kamera canggih, aplikasi beragam, dan fitur lainnya yang memenuhi kebutuhan dan keinginan pengguna, sehingga permintaan terhadap perangkat ini terus meningkat dari tahun ke tahun [6].

Dalam konteks penetapan harga, terdapat variabel-variabel yang saling berkaitan erat dalam menentukan nilai ekonomi sebuah perangkat. Komponen utama seperti kapasitas *RAM*, *Storage* (*penyimpanan internal*), dan *baterai* merupakan variabel independen yang berpengaruh terhadap performa sekaligus harga jual di pasar. Hubungan antar variabel ini sering kali bersifat kompleks; misalnya, peningkatan kapasitas *RAM* dengan kapasitas *storage* dan *baterai* yang lebih besar akan menyebabkan kenaikan harga yang jauh lebih signifikan dibandingkan hanya peningkatan kapasitas *baterai* saja [7].

Untuk memodelkan hubungan tersebut, penelitian ini memanfaatkan perangkat lunak *Orange Data Mining*. Pemilihan platform *Orange* didasarkan pada kemampuannya menyediakan alur kerja data yang bersifat visual dan transparan, mulai dari tahap pra-pemrosesan hingga evaluasi model [8]. Dalam studi ini, penulis membandingkan dua algoritma machine learning, yaitu *Linear Regression* dan *Gradient Boosting*. Algoritma pembelajaran mesin dapat memberikan hasil prediksi yang lebih objektif dalam menangani variabel numerik yang beragam. *Linear Regression* digunakan sebagai metode dasar untuk mengidentifikasi hubungan linear, sedangkan *Gradient Boosting* dipilih karena kemampuannya menangkap pola non-linear yang lebih kompleks secara iteratif [9].

Keakuratan kedua model tersebut diukur menggunakan tiga metrik evaluasi utama, yaitu *Root Mean Square Error* (*RMSE*), *Mean Absolute Percentage Error* (*MAPE*), dan *R-Squared* (*R²*). *RMSE* sangat penting untuk mengukur seberapa besar penyimpangan prediksi dalam satuan harga asli, sedangkan *MAPE* memberikan gambaran persentase rata-rata dari kesalahan prediksi. Sementara itu, *R²* digunakan untuk menilai sejauh mana variabel-variabel spesifikasi seperti *RAM*, *Storage*, dan *Baterai* mampu menjelaskan variasi harga smartphone [10].

Penelitian dibatasi pada lima spesifikasi utama (*RAM*, *Storage*, *Baterai*, *Brand*, dan *Resolusi Display*) yang dipilih berdasarkan penelitian terdahulu karena memiliki pengaruh paling signifikan terhadap harga smartphone.

Machine Learning adalah salah satu cabang dari kecerdasan buatan yang berfokus pada penciptaan algoritma yang memungkinkan komputer untuk belajar dari data tanpa harus diprogram secara eksplisit [11]. Pendekatan ini dipandang sebagai metode ilmiah berbasis data (*data-driven framework*) yang menggunakan algoritma komputer untuk mendeteksi pola, memahami hubungan kompleks antar variabel, serta menghasilkan prediksi dan pengambilan keputusan secara otomatis tanpa perlu penulisan aturan bisnis secara manual [12].

Linear Regression adalah model prediktif dasar yang mengasumsikan bahwa perubahan pada variabel target (harga) dapat dijelaskan melalui kombinasi linier dari fitur-fitur input. Tujuan utama algoritma ini adalah menentukan bobot koefisien (β) sedemikian rupa sehingga fungsi kerugian (*loss function*), biasanya berupa *Sum of Squared Errors* (*SSE*), diminimalkan [13]. Regresi linear untuk menjelaskan seberapa besar perubahan nilai rata-rata variabel dependen (*Y*) apabila variabel independennya (*X*) mengalami kenaikan atau penurunan [14].

Gradient Boosting merupakan metode pembelajaran ansambel (*ensemble*) yang membangun pohon-pohon keputusan (*decision trees*) secara berurutan atau sekuensial (*sequentially*). Algoritma ini mengombinasikan konsep *Gradient Boosting* dengan teknik regularisasi. Tujuan dari regularisasi ini adalah meningkatkan kemampuan generalisasi model dan mengoptimalkan performa keseluruhan, sehingga sangat efisien dan memiliki tingkat akurasi tinggi untuk menangani tugas prediksi data terstruktur (*structured data*) [15]. *Gradient Boosting* dipahami sebagai paradigma model pembelajaran mesin non-linear berbasis pohon (*tree-based non-linear models*) yang sangat kuat untuk memetakan hubungan dan interaksi rumit antar-fitur data yang bersifat tidak linear [16].

Orange adalah sebuah perangkat lunak *open-source* yang berbasis komponen, digunakan untuk penambangan data (*data mining*), machine learning, dan visualisasi data. Keunggulan utama dari *Orange* terletak pada antarmuka visualnya yang menerapkan konsep pemrograman visual, sehingga memudahkan peneliti dalam melakukan proses *preprocessing*, pemodelan, hingga evaluasi data tanpa perlu menulis kode yang kompleks [8]. Secara operasional, platform ini memanfaatkan elemen visual mandiri yang disebut sebagai *Widgets* untuk merepresentasikan setiap fungsionalitas analisis data. Pengguna dapat membangun sebuah alur kerja eksperimental (*workflow*) yang transparan secara sekuensial dengan cara menghubungkan satu *Widget* ke *Widget* lainnya di atas kanvas kerja [9]. Fleksibilitas interaktif ini sangat mendukung efisiensi pengujian, karena penyesuaian parameter pada tahap *preprocessing* dapat langsung dialirkan ke dalam blok algoritma prediktif, seperti *Linear Regression* dan *Gradient Boosting*, serta dievaluasi performanya secara bersamaan dalam satu lingkungan kerja yang terintegrasi [12].

Smartphone adalah perangkat seluler yang menggabungkan fitur komunikasi dengan pemrosesan data tingkat tinggi, sistem operasi canggih, dan koneksi internet. Dalam ekonomi digital, nilai jual perangkat ini tidak ditentukan oleh satu faktor saja, melainkan dibentuk oleh kombinasi dari berbagai spesifikasi perangkat kerasnya.

Fenomena fluktuasi pangsa pasar yang terjadi secara konstan menegaskan bahwa kesesuaian antara kualitas fisik perangkat dengan nilai harga yang ditawarkan merupakan determinan utama yang paling sensitif bagi konsumen dalam ekosistem pasar digital tersebut [17].

Fenomena ini dipertegas oleh temuan [18] yang menyatakan bahwa *smartphone* telah bertransformasi menjadi elemen krusial yang tidak dapat dipisahkan dari kehidupan modern, sekaligus bertindak sebagai simbol gaya hidup dan representasi identitas pribadi dari penggunaanya.

Oleh karena itu, penelitian ini memfokuskan analisis pada komponen-komponen arsitektur utama perangkat yang secara empiris terbukti menjadi variabel prediktor paling signifikan terhadap fluktuasi harga *smartphone*. Sebagaimana dieksplorasi oleh [19], komponen arsitektur fisik *smartphone* merupakan perwujudan dari konfigurasi spesifikasi perangkat keras fisik *smartphone*, salah satunya ditunjukkan melalui integrasi lensa kamera beresolusi 12 megapiksel dengan ukuran bukaan (*aperture*) $f/1.5$.

Berdasarkan eksperimen regresi yang dilakukan oleh [20], spesifikasi fisik perangkat keras internal terbukti memberikan pengaruh yang signifikan dalam menentukan nilai estimasi harga *smartphone*. Di sisi lain, tinjauan literatur sistematis secara global oleh [9] juga memperkuat bahwa fitur arsitektur dasar perangkat merupakan komponen penentu utama (*key determining features*) yang paling konsisten dalam memetakan nilai pasar sebuah *smartphone*. Untuk menjaga objektivitas dan efisiensi model pembelajaran mesin selama tahapan pra-pemrosesan hingga evaluasi, karakteristik fisik tersebut ditransformasikan ke dalam dataset terstruktur berbasis nilai numerik standar [12].

2. METODOLOGI PENELITIAN

Kerangka penelitian merupakan struktur logis dan sistematis yang menggambarkan tahapan-tahapan eksperimen dalam menyelesaikan penelitian ini. Alur kerja ini dirancang untuk mempermudah pemahaman proses komparasi algoritma *Linear Regression* dan *Gradient Boosting* menggunakan perangkat lunak *Orange Data Mining*. Penjelasan detail dari setiap tahapan pada *flowchart* kerangka penelitian adalah sebagai berikut:



Gambar 1. Flowchart Alur Penelitian

2.1 Sumber Data

Data yang digunakan dalam penelitian ini sepenuhnya menggunakan data sekunder. Data sekunder merupakan data yang diperoleh secara tidak langsung atau telah dikumpulkan oleh pihak lain terlebih dahulu. Dataset sekunder dalam penelitian ini didapatkan dari platform repositori data *Kaggle* dengan nama dataset "*phone-information-2024*" url <https://www.kaggle.com/datasets/willianoliveiragibin/phone-information-2024>. Dataset ini disimpan dalam format .csv (*Comma-Separated Values*) dan memiliki ukuran awal sebanyak 1.715 baris data (*records*) serta 31 atribut/kolom. Data ini memuat spesifikasi teknis riil serta harga pasar dari 22 merek smartphone pintar dengan rentang tahun peluncuran dari tahun 2017 hingga 2024.

2.2 Tahapan Pra-pemrosesan Data (*Data Preprocessing*)

Seluruh alur pengolahan data dilaksanakan menggunakan perangkat lunak Orange Data Mining. Tahap pra-pemrosesan diawali dengan proses *load data* menggunakan *Widget File* untuk memuat sebanyak 1.715 baris data mentah. Selanjutnya, dilakukan pengecekan *missing value* melalui *Widget Feature Statistics*, di mana hasil analisis menunjukkan persentase data hilang sebesar 0% (*missing value* = 0) sehingga tidak diperlukan adanya prosedur imputasi data. Proses dilanjutkan dengan pengecekan data duplikat (*duplicate data*) menggunakan skrip *Python (Pandas)*, yang mengidentifikasi 1.006 baris data dengan kombinasi *brand*, *model*, *RAM*, dan *storage* yang sama. Namun, karena nilai variabel target (*Price*) pada baris-baris tersebut bervariasi akibat perbedaan wilayah distribusi, waktu pencatatan, atau penjual, data tersebut tetap dipertahankan untuk mencerminkan variabilitas pasar riil. Tahapan pra-pemrosesan diakhiri dengan pengecekan penciran (*outlier*) menggunakan *Widget Outliers* dengan metode *Local Outlier Factor (LOF)* pada tingkat kontaminasi 10%. Pengujian ini mengidentifikasi 171 data *outliers* dan 1.544 data *inliers*. Guna menjaga representasi penuh dari populasi data, seluruh 1.715 data tetap digunakan secara utuh dalam proses pelatihan maupun pengujian.

2.3 Seleksi Fitur (*Feature Selection & Role Mapping*)

Tahapan seleksi fitur dan pemetaan peran variabel (*role mapping*) dilakukan memanfaatkan *Widget Select Columns* pada perangkat lunak *Orange Data Mining*. Pada tahapan ini, atribut-atribut *dataset* dikelompokkan ke dalam tiga kategori utama berdasarkan fungsi dan karakteristiknya. Variabel independen (*features*) mencakup atribut numerik yang meliputi *RAM*, *Storage*, dan *Baterai*, serta atribut kategorikal yang terdiri dari *Brand* dan *Resolusi Display*. Sementara itu, variabel dependen (*target*) yang digunakan adalah *Price* yang bersifat numerik kontinu. Adapun atribut *Model*, *Chipset*, dan *CPU* dialokasikan sebagai variabel meta (*metas / ignored*) sehingga diisolasi secara penuh dari kalkulasi matematis algoritma regresi. Pengisolasian atribut deskriptif tersebut bertujuan untuk mencegah dampak tingginya tingkat kardinalitas (mencapai 758 nilai unik), mengantisipasi permasalahan *curse of dimensionality*, serta mengeliminasi potensi redundansi maupun multikolinearitas antarvariabel.

2.4 Pembagian Data (*Data Splitting*)

Proses pembagian data (*data splitting*) dilaksanakan dengan menggunakan *Widget Data Sampler* melalui metode *Fixed Proportion*. Seluruh dataset yang berjumlah 1.715 baris data dibagi secara acak (*random sampling*) menggunakan rasio proporsi 70% berbanding 30%. Pembagian ini menghasilkan data latih (*training set*) sebanyak 1.201 baris data (70%) yang dialokasikan khusus untuk melatih dan membangun model prediksi. Di sisi lain, sebanyak 514 baris data sisanya (30%) dialokasikan sebagai data uji (*testing set*) yang diisolasi secara ketat dan tidak dilibatkan dalam proses pelatihan guna menguji serta mengevaluasi performa prediktif model secara objektif.

2.5 Pemodelan & Konfigurasi Hyperparameter

- Linear Regression*: Menggunakan *Ordinary Least Squares (OLS)* yang dilengkapi regularisasi *Elastic Net Regression* ($\text{Alpha} = 0.02$, $\text{Elastic Net Mixing} = 0.50$) Penerapan regularisasi ini bertujuan untuk mencegah terjadinya penyesuaian berlebih (*overfitting*) sekaligus menyeimbangkan penalti L1 (*Lasso*) dan L2 (*Ridge*).
- Gradient Boosting*: Dikonfigurasi menggunakan parameter utama yang meliputi jumlah pohon keputusan sebanyak 100 ($\text{number of trees} = 100$), laju pembelajaran (*learning rate*) sebesar 0,1, batas kedalaman maksimum pohon ($\text{limit depth of individual trees}$) sebesar 4, serta pembatasan sampel minimum untuk pemisahan simpul (min. split subset) sebesar 10.

2.6 Evaluasi Metrik dan Uji Hipotesis Statistik

Evaluasi performa model regresi diukur secara kuantitatif menggunakan tiga metrik utama, yaitu *Root Mean Square Error (RMSE)* untuk mengukur akar kuadrat dari rata-rata kuadrat kesalahan prediksi, *Mean Absolute Percentage Error (MAPE)* untuk mengukur persentase rata-rata kesalahan absolut, serta *Koefisien Determinasi (R²)* untuk mengukur proporsi variansi variabel target yang dapat dijelaskan oleh fitur-fitur independen. Untuk membuktikan signifikansi

perbedaan performa antar-algoritma secara objektif, dilakukan pengujian hipotesis statistik. Tahap pengujian dimulai dengan Uji Normalitas Residual menggunakan metode *Kolmogorov-Smirnov* terhadap selisih *absolute error* untuk menganalisis pola distribusi data residual. Berdasarkan karakteristik distribusi data tersebut, pengujian beda signifikansi dilanjutkan menggunakan *Wilcoxon Signed-Rank Test* sebagai uji non-parametrik utama, serta *Paired Sample t-Test* sebagai penguji pelengkap (*robustness check*) guna mengonfirmasi konsistensi hasil pengujian.

3. HASIL DAN PEMBAHASAN

3.1 Hasil Pra-pemrosesan Data (*Data Preprocessing*)

Berdasarkan serangkaian operasional pra-pemrosesan data (*data preprocessing*) yang telah dilakukan menggunakan perangkat lunak *Orange Data Mining*, diperoleh hasil transformasi data yang signifikan dari kondisi mentah (*raw data*) hingga menjadi dataset yang bersih (*clean data*). Ringkasan komprehensif dari hasil pemrosesan tersebut disajikan dalam tabel 1 berikut:

Tabel 1. Hasil Pra-pemrosesan Data

No	Tahapan Preprocessing	Instrumen / Widget Orange	Hasil Pemeriksaan / Temuan	Tindakan Metodologis	Dampak terhadap Jumlah Data
1	Pemuatan Data (Load)	Widget File	Terdeteksi 1.715 baris data mentah awal	Mengunci format pemisah koma (comma-separated)	1.715 baris data tetap utuh
2	Pengecekan Missing Value	Feature Statistics	0 baris data kosong (0% Missing)	Tidak dilakukan proses imputasi nilai rata-rata	1.715 baris data tetap utuh
3	Pengecekan Duplicate Data	Python Script (Pandas)	1.006 baris data memiliki spesifikasi kembar	Mengidentifikasi tingkat redundansi spesifikasi fisik smartphone	Pembatasan parameter fitur untuk analisis
4	Pengecekan Outlier	Outliers Widget (Local Outlier Factor)	171 baris data pencilan (outliers) dan 1.544 baris data normal (inliers)	Tidak dilakukan proses eliminasi data pencilan	1.715 baris data tetap utuh

Pada tabel 1 dijelaskan bahwa tahapan pra-pemrosesan diawali dengan memuat 1.715 baris data mentah *phone-information-2024.csv* yang terisi penuh tanpa *missing value* (0%). Analisis duplikasi menemukan 1.006 data redundan akibat perbedaan wilayah distribusi/retail. Selanjutnya, deteksi pencilan menggunakan *Local Outlier Factor (LOF)* pada tingkat kontaminasi 10% berhasil mengidentifikasi 171 baris data sebagai *outliers* dan 1.544 baris data sebagai *inliers* (normal). Seluruh 1.715 baris data tetap dipertahankan dan digunakan untuk pemodelan *Linear Regression* dan *Gradient Boosting*

3.2 Hasil Analisis Korelasi Spearman

Melalui pengekseskuan pada *Widget Python Script* di platform *Orange Data Mining* (divalidasi melalui eksekusi skrip statistik *Python*), keseluruhan 1.715 baris data (*records*) dianalisis secara komprehensif untuk mengukur tingkat kekuatan, arah hubungan monotonik, serta tingkat signifikansi statistik antara variabel independen spesifikasi perangkat terhadap variabel target (*Price*). Hasil analisis korelasi Rank Spearman dari instrumen pengujian ditabulasikan pada tabel 2 sebagai berikut:

Tabel 2. Hasil Uji Korelasi Spearman

No	Variabel Spesifikasi	Koefisien Korelasi	p-value	Kekuatan dan Arah Hubungan	Signifikansi ($\alpha = 0.05$)
1	RAM	-0.118	0.0000	Sangat Lemah (Negatif)	Signifikan
2	Storage	0.242	0.0000	Lemah / Cukup (Positif)	Signifikan
3	Baterai	-0.134	0.0000	Sangat Lemah (Negatif)	Signifikan

4	Brand	-0.167	0.0000	Sangat Lemah (Negatif)	Signifikan
5	Resolusi Display	0.080	0.0009	Sangat Lemah (Positif)	Signifikan

Pada tabel 2 dijelaskan bahwa hasil uji korelasi Spearman menunjukkan bahwa masing-masing variabel spesifikasi memiliki koefisien korelasi yang rendah atau sangat lemah terhadap variabel target (dengan nilai absolut di bawah 0.3), namun seluruhnya bernilai signifikan ($p\text{-value} < 0.05$). Meskipun hubungan secara individual tergolong lemah, dalam konteks pemodelan prediktif berbasis machine learning, kombinasi seluruh fitur ini mampu menangkap pola interaksi non-linear yang kompleks secara simultan. Hal ini menjelaskan mengapa fitur-fitur tersebut tetap efektif dan esensial sebagai prediktor utama dalam membangun model dengan tingkat akurasi yang tinggi.

3.3 Uji Petik Kasus Konkret Hasil Estimasi di Predictions

Untuk meninjau akurasi pada tingkatan baris data individual secara objektif, *Predictions* mengekstrak beberapa sampel spesifikasi *smartphone* secara acak untuk membandingkan nilai harga asli dengan nilai numerik hasil prediksi secara langsung pada tabel 3 berikut:

Tabel 3. Hasil Predictions

Model	RMSE	MAPE	R2
Linear Regression	34.933	0.087	0.990
Gradient Boosting	16.766	0.014	0.998

Pada tabel 3 dijelaskan bahwa algoritma *Gradient Boosting* menunjukkan performa prediksi yang lebih tinggi dibandingkan dengan *Linear Regression*, yang ditandai dengan nilai error dari *RMSE* dan *MAPE* yang lebih rendah, serta *R2* yang lebih tinggi. Keunggulan ini terjadi karena *Gradient Boosting* merupakan algoritma *ensemble* berbasis pohon keputusan (*decision trees*) yang mampu menangkap hubungan non-linear yang kompleks dan efek interaksi antarfitur *smartphone*. Sebaliknya, *Linear Regression* hanya mampu memetakan hubungan searah yang bersifat linear (garis lurus) dan sangat kaku terhadap variasi harga di pasar *smartphone*. Melalui proses iterasi sekuensial yang terus-menerus memperbaiki kesalahan prediksi dari pohon keputusan sebelumnya, *Gradient Boosting* berhasil meminimalkan deviasi prediksi secara signifikan, sehingga menghasilkan estimasi harga yang jauh lebih akurat, stabil, dan mendekati nilai riil di pasar.

Tabel 4. Hasil Prediksi berdasarkan Variabel Independen

Variabel Independen	Linear Regression			Gradient Boosting		
	RMSE	MAPE	R2	RMSE	MAPE	R2
RAM	289.061	0.728	0.326	190.175	0.411	0.708
Storage	265.463	0.594	0.431	201.33	0.434	0.673
Baterai	309.635	0.904	0.226	200.639	0.548	0.675
Brand	285.362	0.805	0.343	216.754	0.579	0.621
Resolusi Display	245.117	0.647	0.515	202.477	0.483	0.669

Pada tabel 4 dijelaskan bahwa dilakukan pengujian prediksi secara individual terhadap variabel yang akan diteliti menggunakan *Widget Predictions* untuk membandingkan nilai harga aktual (*ground truth*) dengan hasil estimasi dari kedua algoritma secara langsung. Berdasarkan data perbandingan tersebut, terlihat bahwa Algoritma *Gradient Boosting* secara konsisten mengungguli *Linear Regression* pada setiap variabel independen karena karakteristik dasarnya sebagai metode *ensemble* berbasis pohon keputusan (*tree-based*). Teori *machine learning* menjelaskan bahwa *Linear Regression* memaksa hubungan garis lurus konstan di sepanjang data. Realitasnya, variabel seperti *RAM*, *Storage*, *Baterai*, *Brand*, dan *Resolusi Display* memiliki pola non-linear dan batas jenuh di pasar nyata (misalnya, harga tidak naik secara linear di setiap penambahan 1 GB *RAM*). *Gradient Boosting* mampu memecah data secara fleksibel melalui *node splitting* (percabangan) untuk menangkap pola non-linear dan interaksi kompleks tersebut tanpa memerlukan asumsi linearitas, sehingga menghasilkan estimasi yang jauh lebih mendekati kondisi riil.

3.4 Evaluasi Kuantitatif Performa Model di Test and Score

Kalkulasi akurasi model diukur secara otomatis berdasarkan tiga metrik evaluasi standar, yaitu *Root Mean Square Error (RMSE)*, *Mean Absolute Percentage Error (MAPE)*, dan koefisien determinasi *R-Squared (R2)*. Hasil ekstraksi kuantitatif dari *Widget Test and Score* dirangkum secara valid pada tabel 5 berikut:

Tabel 5. Hasil Test and Score

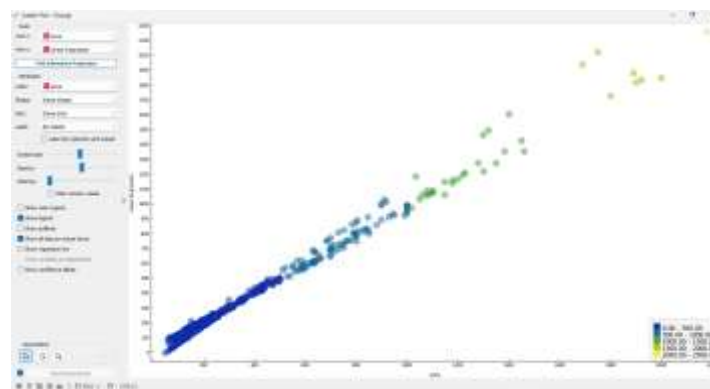
Model	RMSE	MAPE	R2
Linear Regression	33.216	0.093	0.990

Gradient Boosting	16.923	0.017	0.997
-------------------	--------	-------	-------

Pada tabel 5 dijelaskan bahwa *Gradient Boosting* memiliki performa lebih baik dibandingkan *Linear Regression* yang dicerminkan oleh penurunan nilai *RMSE* dan *MAPE* serta *R2* yang lebih tinggi berakar pada karakteristik arsitekturnya. Sebagai algoritma *ensemble* berbasis pohon keputusan, *Gradient Boosting* sangat adaptif dalam menangkap pola hubungan non-linear yang kompleks serta efek interaksi multivariat antarspesifikasi fisik smartphone. Sebaliknya, *Linear Regression* terbatas pada pemetaan linear (garis lurus) yang terlalu kaku untuk mengakomodasi fluktuasi harga pasar. Melalui mekanisme iterasi sekuensial yang mengoreksi kesalahan prediksi dari tahapan pohon sebelumnya secara bertahap, *Gradient Boosting* mampu memangkas deviasi prediksi secara signifikan guna menghasilkan estimasi harga final yang jauh lebih presisi, stabil, dan realistis.

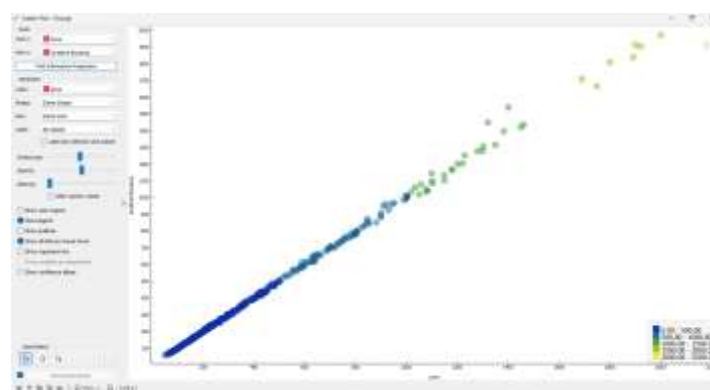
3.5 Analisis Visual Sebaran Prediksi di *Scatter Plot*

Guna memvalidasi akurasi kuantitatif di atas, *Scatter Plot* digunakan untuk memproyeksikan visualisasi koordinat dua dimensi yang mempertemukan nilai harga aktual smartphone (Sumbu X) dengan nilai harga hasil estimasi algoritma (Sumbu Y). Proyeksi visual grafik tersebut disajikan pada gambar 2 berikut:



Gambar 2. Hasil visual grafik pada harga pada sumbu x dan Linear Regression pada sumbu y

Pada gambar 2 dijelaskan bahwa sebaran koordinat pada model *Linear Regression* cenderung melebar dan membentuk jarak deviasi yang renggang dari garis lurus diagonal utamanya, terutama pada klaster harga tinggi (*high-end segment*). Semakin jauh posisi titik dari garis diagonal, semakin besar nilai kesalahan prediksi yang dihasilkan model. Penyebaran titik yang semakin lebar menunjukkan bahwa model kurang mampu mempelajari pola hubungan antara variabel independen dengan variabel dependen sehingga akurasi prediksinya menurun. Hal ini mencerminkan tingginya nilai kesalahan prediksi ketika dihadapkan pada variabilitas harga yang ekstrem.



Gambar 3. Hasil visual grafik pada harga pada sumbu x dan Gradient Boosting pada sumbu y

Pada gambar 3 dijelaskan bahwa sebaran koordinat pada model *Gradient Boosting* memperlihatkan kerapatan titik data yang sangat padat, konsisten, dan berhimpitan secara presisi mengikuti arah garis lurus diagonal. Semakin dekat posisi titik terhadap garis diagonal, semakin kecil selisih antara nilai aktual dan nilai prediksi. Hal ini menunjukkan bahwa model memiliki tingkat akurasi yang semakin baik karena mampu menghasilkan prediksi yang mendekati nilai sebenarnya. Visualisasi ini membuktikan secara empiris keakuratan model dalam memetakan target harga di seluruh klaster spesifikasi perangkat.

3.6 Hasil Uji Signifikansi Statistik Beda Performa Model

Untuk memastikan bahwa perbedaan performa antara algoritma *Linear Regression* dan *Gradient Boosting* signifikan secara statistik dan bukan terjadi karena variasi acak (*random chance*), dilakukan pengujian hipotesis statistik terhadap nilai *error* mutlak (*absolute error*) dari 514 sampel data uji.

a. Uji Normalitas Residual

Pengujian normalitas dilakukan terhadap selisih *error* mutlak (*absolute error*) kedua model menggunakan metode *Kolmogorov-Smirnov* dengan tingkat signifikansi $\alpha = 0.05$. Hasil pengujian disajikan pada tabel 6 berikut:

Tabel 6. Hasil Uji Normalitas Residual (Kolmogorov-Smirnov)

Indikator Tingkat Kesalahan	Nilai KS- Statistic	p-value	Signifikansi α	Keputusan H_0	Kesimpulan Sebaran
Selisih Absolute Error (LR - GB)	0.0651	0.0245	0.05	Ditolak	Tidak Berdistribusi Normal

Pada tabel 6 dijelaskan bahwa hasil pengujian normalitas terhadap selisih *error* mutlak (*absolute error*) antara algoritma *Linear Regression* dan *Gradient Boosting* menggunakan metode *Kolmogorov-Smirnov* menghasilkan nilai statistik *KS-Statistics* sebesar 0.0651 dengan probabilitas signifikansi *p-value* sebesar 0.0245. Nilai *p-value* tersebut kemudian dibandingkan dengan tingkat signifikansi kritis yang ditetapkan dalam penelitian, yaitu $\alpha = 0.05$. Karena nilai $p\text{-value} \leq 0.05$, maka secara statistik inferensial H_0 ditolak dan H_1 diterima. Penolakan H_0 ini membuktikan secara ilmiah bahwa sebaran data selisih *absolute error* kedua model tidak berdistribusi normal. Secara metodologis, tidak terpenuhinya asumsi normalitas ini berimplikasi langsung pada pemilihan uji inferensial selanjutnya. Penggunaan pengujian parametrik seperti *Paired Sample t-Test* tidak dapat dijadikan acuan utama karena berpotensi menghasilkan bias akibat pelanggaran asumsi distribusi. Oleh karena itu, untuk menjaga keabsahan (*validity*) dan ketepatan analisis, pengujian beda signifikansi statistik secara sah dialihkan menggunakan metode non-parametrik *Wilcoxon Signed-Rank Test* sebagai uji utama.

b. Uji Beda Signifikansi Statistik

Pengujian beda performa dilakukan menggunakan *Wilcoxon Signed-Rank Test* sebagai uji utama (non-parametrik). Selain itu, pengujian *Paired Sample t-Test* (parametrik) juga ditampilkan sebagai analisis pelengkap (*robustness check*) untuk menguji konsistensi hasil. Hasil disajikan pada tabel 7 berikut

Tabel 7. Hasil Uji Signifikansi Statistik Beda Performa Model

Metode Uji	Sifat Uji	Statistik Uji	p-value	Keputusan	Kesimpulan Hipotesis
Wilcoxon Signed-Rank Test	Utama (Non- Parametrik)	$Z = 22929.0000$	0.0000	H_0 Ditolak	Perbedaan performa signifikan secara nyata (H_1 diterima)
Paired Sample t-Test	Pelengkap (Parametrik)	$t = 15.0044$	0.0000	H_0 Ditolak	Perbedaan performa signifikan secara nyata (H_1 diterima)

Pada tabel 7 dijelaskan bahwa pengujian beda performa menggunakan uji utama *Wilcoxon Signed-Rank Test* menghasilkan nilai statistik $Z = 22929.0000$ dengan $p\text{-value} = 0.0000$. Karena nilai $p\text{-value} < 0.05$, maka H_0 ditolak dan H_1 diterima. Konsistensi hasil ini juga didukung oleh uji pelengkap (*robustness check*) menggunakan *Paired Sample t-Test* yang menghasilkan nilai statistik $t = 15.0044$ dan $p\text{-value} = 0.0000$ ($p < 0.05$). Konfirmasi dari kedua pendekatan statistik ini (baik parametrik maupun non-parametrik) menyimpulkan bahwa keunggulan akurasi algoritma *Gradient Boosting* atas *Linear Regression* bersifat signifikan secara nyata. Penurunan nilai kesalahan mutlak (*absolute error*) yang dihasilkan oleh *Gradient Boosting* terbukti secara empiris bukan merupakan hasil dari variasi acak (*random chance*) pada data uji, melainkan karena kemampuan model dalam mempelajari pola non-linier data secara lebih optimal.

3.7 Pembahasan Teoretis

Keunggulan performa algoritma *Gradient Boosting* dibandingkan *Linear Regression* berdasarkan metrik evaluasi kuantitatif tidak terjadi secara kebetulan (*random chance*), melainkan dapat dijelaskan secara teoretis melalui perbedaan fundamental pada arsitektur matematika, mekanisme pembelajaran, serta kapabilitas kedua algoritma dalam memodelkan karakteristik data harga smartphone.

a. Keterbatasan Asumsi Parametrik dan Hyperplane pada Linear Regression

Linear Regression mencoba membangun bidang lurus global (*global hyperplane*). Model ini mengasumsikan efek penambahan spesifikasi bersifat aditif konstan. Padahal, dalam ekonomi pasar smartphone, penambahan 4 GB RAM pada smartphone murah memberikan pengaruh kenaikan harga yang berbeda jika dibandingkan dengan penambahan RAM pada smartphone *flagship*. Selain itu, regresi linier standar tidak secara otomatis mampu mengekstrak interaksi antar-fitur (*feature interaction*), sehingga menghasilkan *unexplained variance* yang cukup tinggi.

- b. Keunggulan Arsitektur Ensemble Non-Parametrik Gradient Boosting
- Gradient Boosting* bekerja secara non-parametrik melalui dua mekanisme utama:
1. Partisi Ruang Fitur (*Node Splitting*): Membagi ruang data multidimensi menjadi wilayah keputusan (*decision regions*) lokal berdasarkan ambang batas nilai (*thresholds*), sehingga mampu mendekati fungsi non-linear yang rumit.
 2. Pembelajaran Sekuensial Aditif (*Gradient Descent in Function Space*): Setiap pohon keputusan baru dilatih secara khusus untuk mengoreksi *pseudo-residuals* (sisa kesalahan) dari kumpulan pohon sebelumnya. Mekanisme koreksi teriterasi ini secara sistematis menekan akumulasi *error* kuadrat (*RMSE*) dan persentase *error* (*MAPE*) hingga tingkat terendah.

4. KESIMPULAN

Berdasarkan hasil penelitian terhadap 1.715 data smartphone menggunakan perangkat lunak *Orange Data Mining*, dapat disimpulkan bahwa algoritma *Linear Regression* dan *Gradient Boosting* berhasil diimplementasikan melalui tahapan *preprocessing*, pembagian data menggunakan metode *hold-out* dengan proporsi 70% data latih dan 30% data uji, serta proses pemodelan untuk memprediksi harga smartphone. Variabel *RAM*, *Storage*, *Baterai*, *Brand*, dan *Resolusi Display* digunakan sebagai prediktor utama dan berkontribusi terhadap pembentukan model prediksi harga. Hasil evaluasi menunjukkan bahwa *Gradient Boosting* memberikan performa yang lebih baik dibandingkan *Linear Regression*, dengan nilai R^2 sebesar 0.998 atau 99.8%, *RMSE* sebesar 16.766, dan *MAPE* sebesar 0.014 atau 1.4%. Sementara itu, *Linear Regression* memperoleh nilai R^2 sebesar 0.990 atau 99.0%, *RMSE* sebesar 34.933, dan *MAPE* sebesar 0.087 atau 8.7%. Perbedaan tingkat kesalahan prediksi kedua model juga terbukti signifikan berdasarkan *Wilcoxon Signed-Rank Test* dengan nilai $p < 0.05$. Dengan demikian, H_0 ditolak dan H_1 diterima. Kontribusi dan kebaruan penelitian ini terletak pada perbandingan langsung algoritma *Linear Regression* dan *Gradient Boosting* menggunakan dataset “*phone-information-2024*” untuk prediksi harga smartphone melalui pemrograman visual pada platform *Orange Data Mining*, serta pemanfaatan kombinasi lima variabel utama sebagai prediktor harga. Hasil penelitian menunjukkan bahwa *Gradient Boosting* lebih efektif dalam menangkap hubungan non-linear dan interaksi kompleks antar spesifikasi smartphone sehingga menghasilkan prediksi harga yang lebih akurat dibandingkan *Linear Regression*.

REFERENCES

- [1] Tiara and Ismawati, “Analisis Pengaruh Kualitas Produk, Harga, dan Promosi Terhadap Keputusan Pembelian Handphone Samsung pada Mahasiswa Kampus Negeri di Surakarta,” *Analisis Pengaruh Kualitas Produk, Harga, dan Promosi Terhadap Keputusan Pembelian Handphone Samsung pada Mahasiswa Kampus Negeri di Surakarta*, vol. 2, pp. 1–7, Nov. 2025, doi: <https://doi.org/10.36600/rma.v16i2.497>.
- [2] P. A. Candra and S. M. Sagga, “ANALYSIS OF THE INFLUENCE OF BRAND EQUITY ON CUSTOMER LOYALTY THROUGH CUSTOMER SATISFACTION FOR SAMSUNG SMARTPHONES: A CASE STUDY OF SMARTPHONE USERS AT TMT BUILDING, CILANDAK, SOUTH JAKARTA ANALISIS PENGARUH EKUITAS MEREK TERHADAP LOYALITAS PELANGGAN MELALUI KEPUASAN PELANGGAN SMARTPHONE MEREK SAMSUNG: STUDI KASUS PADA PELANGGAN SMARTPHONE DI TMT BUILDING CILANDAK, JAKARTA SELATAN ARTICLE INFO ABSTRACT Correspondent,” *Journal of Social and Economics Research*, vol. 7, no. 1, 2025, doi: <https://doi.org/10.54783/jsr.v7i1.917>.
- [3] D. S. Dewi et al., “PENGARUH GAYA HIDUP SERTA KUALITAS PRODUK TERHADAP KEPUTUSAN PEMBELIAN SMARTPHONE IPHONE DI WILAYAH KOTA DKI JAKARTA,” 2026. doi: <https://doi.org/10.53825/jmbjayakarta.v7i2.430>.
- [4] Felia and Nurul, “PENGARUH KUALITAS PRODUK, HARGA PRODUK DAN KEPUASAN PELANGGAN TERHADAP KEPUTUSAN PEMBELIAN BRAND HP OPPO YANG DIGUNAKAN OLEH MASYARAKAT JEPARA,” *PENGARUH KUALITAS PRODUK, HARGA PRODUK DAN KEPUASAN PELANGGAN TERHADAP KEPUTUSAN PEMBELIAN BRAND HP OPPO YANG DIGUNAKAN OLEH MASYARAKAT JEPARA*, vol. 15, pp. 1–27, Dec. 2025, doi: <https://doi.org/10.52643/jam.v15i4.6914>.
- [5] T. S. Muliadi, B. Romadhoni, S. Indrayani, F. Ekonomi, D. Bisnis, and U. Muhammadiyah Makassar, “PENGARUH KUALITAS PRODUK, HARGA DAN WORD OF MOUTH TERHADAP MINAT BELI PRODUK SMARTPHONE OPPO PADA TOKO ITC CELL KOTA MAKASSAR,” 2025. doi: <https://doi.org/10.37476/jbk.v14i2.5101>.
- [6] G. Graceallah, “PENGARUH KOMPETITIF PRICING DAN CUSTOMER PERCEIVED VALUE TERHADAP KEPUTUSAN MENJUAL SMARTPHONE INFINIX PADA GERAI PENJUALAN SMARTPHONE DI PONTIANAK DAN KUBU RAYA DENGAN BRAND IMAGE SEBAGAI VARIABEL PEMEDIASI,” 2025. doi: <https://doi.org/10.30738/md.v9i3.20738>.
- [7] J. Cui, X. Wang, and Y. Zheng, “Analysis of Factors Influencing Smartphone Prices,” *INSTICC*, Sep. 2025, pp. 65–69. doi: <https://doi.org/10.5220/0013814100004708>.

- [8] E. Tri, A. A'fena, A. Ristyawan, and A. Nugroho, "Pengaruh Normalisasi Fitur Pada Perbandingan Algoritma Random Forest Regressor dan KNN Regressor dalam Menentukan Harga Smartphone Effect of Feature Normalization on Comparison of Random Forest Regressor and KNN Regressor Algorithms in Determining Smartphone Prices," *KRESNA: Jurnal Riset dan Pengabdian Masyarakat*, vol. 5, pp. 74–83, 2025, doi: <https://doi.org/10.36080/kresna.v5i1.226>.
- [9] D. W. Naaman, B. Tahir Ahmed, and O. S. Kareem, "Comparative Review of Machine Learning Models for Mobile Price Prediction Based on Specifications: A Systematic Literature Analysis," vol. 16, no. 2, pp. 84–109, 2025, doi: <https://doi.org/10.31849/digitalzone.v16i2>.
- [10] Y. Ruizhe, "Mobile Phone Price Prediction Model Based on Deep Learning," *ITM Web of Conferences*, vol. 78, p. 03030, 2025, doi: <https://doi.org/10.1051/itmconf/20257803030>.
- [11] Edi Ismanto, Ahmad Gunawan Dalimunthe, Muhammad Iqbal, and Fauza Addinunnisa, "Analisis Perbandingan Model Machine Learning dan Deep Learning untuk Peramalan Harga Saham," *Jurnal Ilmu Komputer dan Teknik Informatika*, vol. 2, no. 1, pp. 131–139, Jan. 2026, doi: <https://doi.org/10.64803/juikti.v2i1.114>.
- [12] B. B. Özen, M. F. Alaeddinoğlu, and T. Aydın, "Analysing the Refurbished Smart Phone Market with Machine Learning," *Trends in Business and Economics*, vol. 40, no. 1, pp. 42–59, Jan. 2026, doi: <https://doi.org/10.16951/trendbusecon.1607949>.
- [13] -PhD, Ahad Rezayan, A. Mansouri -PhD Candidate, and A. Rezayan, "Applying Neural Network Modeling to Mobile Pricing: A New Approach to Smartphone Price Prediction," Jan. 2026. doi: 10.2139/ssrn.5419155.
- [14] R. S. Rahmah, "Pengaruh Harga dan Promosi terhadap Keputusan Pembelian Iphone ada Kalangan Followers X," 2025. doi: <https://doi.org/10.35972/kindai.v2i1i.2109>.
- [15] P. J. C. Rivera, C. Bunn, and M. Ryo, "Smartphone-Based Monitoring Identifies the Importance of Farm Size and Soil Type for Coffee Tree Productivity at a Large Geographic Scale," *Journal of Sustainable Agriculture and Environment*, vol. 4, no. 4, Dec. 2025, doi: <https://doi.org/10.1002/sae2.70111>.
- [16] S. Monir and S. Ahmed, "Prediction of Smartphone Sales Revenue Through Machine Learning Approach," Dec. 2025. doi: <https://doi.org/10.46254/ba08.20250230>.
- [17] A. K. Suaib, Z. Ruma, N. fadilah Aswar, C. imran Musa, and I. wardhana Haeruddin, "Pengaruh Kualitas Produk, Harga dan Promosi Terhadap Keputusan Pembelian Smartphone Oppo di Makassar," *RIGGS: Journal of Artificial Intelligence and Digital Business*, vol. 4, no. 3, pp. 2871–2883, Aug. 2025, doi: <https://doi.org/10.31004/riggs.v4i3.2388>.
- [18] M. A. Maulana, A. Agung, and K. Diatmika, "Pengaruh Pemasaran Media Sosial Terhadap Ekuitas Merek Berbasis Konsumen, Pengalaman Merek Dan Niat Pembelian Produk Smartphone," 2026. doi: <https://doi.org/10.56799/peshum.v5i2.12864>.
- [19] P. H. J. Fernandes, S. L. Davies-Jenkins, W. Z. Liang, M. E. Drewnoski, and Y. Xiong, "Assessing forage mass using smartphone imagery for cereal rye under grazing conditions in Nebraska," *Crop, Forage and Turfgrass Management*, vol. 12, no. 1, Jun. 2026, doi: <https://doi.org/10.1002/cft2.70101>.
- [20] M. Irsyad, Silvy Amelia, and Y. M. Ardi, "PREDIKSI HARGA PONSEL BERDASARKAN SPESIFIKASINYA MENGGUNAKAN ALGORITMA LINEAR REGRESSION," *INTI Nusa Mandiri*, vol. 19, no. 2, pp. 251–258, Feb. 2025, doi: <https://doi.org/10.33480/inti.v19i2.6292>.