



PREDIKSI HARGA SMARTPHONE BERDASARKAN FITUR SMARTPHONE DENGAN *RANDOM FOREST REGRESSION*

Seftyan Wiguna¹, Syarifah Putri Agustini Alkadri², Istikoma³

^{1,2}Jurusan Teknik Informatika, Universitas Muhammadiyah Pontianak

³Jurusan Sistem Informasi, Fakultas Teknik dan Ilmu Komputer,
Universitas Muhammadiyah Pontianak

¹seftyan.wigunaa@gmail.com, ²agustini.putri@unmuhpnk.ac.id, ³istikoma@unmuhpnk.ac.id

Abstrak

Kemajuan teknologi telekomunikasi selama beberapa dekade terakhir telah secara drastis mengubah gaya hidup manusia di seluruh dunia. Penentuan harga yang tepat untuk ponsel pintar merupakan faktor kritis dalam strategi pemasaran dan penjualan yang sukses. Harga yang terlalu tinggi dapat mengurangi minat konsumen, sementara harga yang terlalu rendah dapat mengurangi potensi pendapatan perusahaan. Dalam konteks ini, banyak fitur penting yang perlu dipertimbangkan dalam memperkirakan harga ponsel, seperti prosesor, kapasitas baterai, ukuran layar, ketebalan, memori internal, resolusi kamera, dan kualitas video. Penelitian ini bertujuan untuk memprediksi harga smartphone berdasarkan fitur-fitur tersebut menggunakan metode Random Forest Regression, sebuah teknik machine learning yang dikenal mampu menangani data yang kompleks dan non-linear. Dalam penelitian ini sebuah aplikasi dikembangkan untuk memprediksi harga smartphone dengan menggunakan algoritma Random Forest. Hasil pengujian menunjukkan bahwa model prediktif ini memiliki performa yang cukup baik, dengan nilai MAPE sebesar 35,6%, MAE sebesar Rp. 760,969, MSE sebesar 3.023, dan RMSE sebesar Rp. 1,745,144. Hasil ini menunjukkan bahwa metode Random Forest Regression mampu memberikan prediksi yang mendekati nilai aktual. Aplikasi ini diharapkan dapat membantu produsen ponsel pintar dalam menentukan strategi harga yang lebih tepat dan mengoptimalkan keputusan bisnis. Dengan demikian, penelitian ini memberikan kontribusi dalam pengembangan alat prediksi harga berbasis fitur smartphone yang dapat digunakan secara praktis oleh pelaku industri teknologi.

Kata kunci: Prediksi, Harga Smartphone, *Random Forest Regression*

Article History

Received: January 2025
Reviewed: January 2025
Published: January 2025

Plagiarism Checker No 234
Prefix DOI :
10.8734/Koehesi.v1i2.365
Copyright : Author
Publish by : Koehesi



This work is licensed under
a [Creative Commons
Attribution-NonCommercial
4.0 International License](https://creativecommons.org/licenses/by-nc/4.0/)



1. PENDAHULUAN

Kemajuan teknologi telekomunikasi selama beberapa dekade terakhir telah secara drastis mengubah gaya hidup manusia di seluruh dunia. Salah satu inovasi utama di bidang ini adalah ponsel pintar yang telah menjadi bagian tak terpisahkan dari kehidupan sehari-hari masyarakat (Widodo & Qurniawati, 2017). Kemampuan ponsel pintar yang semakin maju tidak hanya memungkinkan akses mudah ke informasi dan komunikasi lintas batas geografis, tetapi juga telah membuka peluang baru di sektor bisnis.

Penentuan harga yang tepat untuk ponsel pintar menjadi faktor kunci dalam keberhasilan pemasaran dan penjualan (Suwarni et al., 2020). Harga yang terlalu tinggi dapat mengurangi minat konsumen, sedangkan harga yang terlalu rendah bisa mengurangi potensi pendapatan Perusahaan (Sari & Kusdiyanto, 2024). Oleh karena itu, diperlukan pendekatan yang tepat dan efisien untuk menentukan harga optimal dengan mempertimbangkan berbagai faktor yang mempengaruhi harga ponsel (Asim & Khan, 2018).

Banyak fitur penting yang harus dipertimbangkan dalam menetapkan harga ponsel, seperti prosesor, masa pakai baterai, ukuran dan ketebalan, memori internal, piksel kamera, dan kualitas video. Berdasarkan penjelasan mengenai fitur-fitur ini yang mendasari tujuan penelitian, dilakukan analisis terhadap sekumpulan fitur pada data set yang sangat mempengaruhi label 'harga'. Analisis ini menggunakan model Prediktif Random Forest Regression untuk menghitung pentingnya setiap fitur. *Random Forest Regression* (RFR) memiliki kemampuan bawaan untuk mengukur kepentingan relatif setiap fitur dalam membuat prediksi, yang membantu mengidentifikasi fitur-fitur dengan dampak signifikan pada harga ponsel pintar.

RFR adalah metode yang menggunakan konsep pohon keputusan secara ensemble, di mana sejumlah besar pohon keputusan dibangun untuk menghasilkan prediksi yang lebih akurat, serta mengurangi risiko overfitting seperti yang dapat terjadi pada *Decision Tree* (Yin et al., 2023). Selain itu, keunggulan lain dari metode RFR adalah kemampuannya menangani data yang kompleks dan non-linear, serta kemampuannya mengidentifikasi ketergantungan variabel yang kompleks dalam dataset (Wisyaladin et al., 2020). RFR juga memiliki toleransi yang baik terhadap outlier, yaitu data yang sangat berbeda dari data lainnya, yang penting dalam kasus prediksi harga, di mana perbedaan signifikan antara beberapa ponsel bisa menjadi hal yang umum (Tang & Asandimitra, 2023).

Penelitian ini bertujuan untuk mengembangkan model prediksi harga ponsel pintar menggunakan *Random Forest Regression* (RFR). Tujuan utama dari penelitian ini adalah untuk menguji kemampuan algoritma Random Forest Regression dalam memprediksi harga ponsel pintar dan mengidentifikasi fitur-fitur dalam dataset yang paling berpengaruh terhadap harga ponsel. Oleh karena itu, penelitian ini diharapkan dapat memberikan wawasan berharga bagi produsen ponsel pintar dan pelaku industri terkait dalam pengambilan keputusan bisnis yang lebih terinformasi dan akurat, misalnya melalui estimasi tren harga ponsel. Estimasi ini dapat meningkatkan efisiensi dalam penentuan harga ponsel.



Menggunakan data historis untuk memprediksi harga produk yang ada dan peluncuran produk baru merupakan latar belakang penelitian yang menarik bagi para peneliti Machine Learning. Penelitian sebelumnya oleh (Muhayat, 2022) telah membuktikan penggunaan Machine Learning dalam memprediksi harga ponsel, dan penggunaan algoritma *Random Forest* untuk memecahkan masalah regresi juga telah diuji oleh penelitian (Saadah & Salsabila, 2021). Setelah model dibangun, model ini diterapkan pada sebuah website sebagai antarmuka pengguna untuk memudahkan pengguna dalam memprediksi harga ponsel pintar.

2. LANDASAN TEORI

2.1 *Random Forest Regression*

Regresi adalah salah satu bentuk pembelajaran terawasi yang penting dalam pembelajaran mesin. Metode ini melibatkan penggunaan set pelatihan yang terdiri dari pengamatan atau fitur yang diikuti dengan nilai target kontinu yang sesuai (Hasibuan et al., 2022). Perbedaan utama antara regresi dan klasifikasi terletak pada sifat dari nilai yang dihasilkan: regresi menghasilkan nilai-nilai kontinu, sedangkan klasifikasi menghasilkan nilai-nilai diskrit. Oleh karena itu, regresi dan klasifikasi memiliki aplikasi yang berbeda dalam domain pembelajaran terawasi.

Random Forest adalah metode untuk klasifikasi dan regresi non-linier yang bergantung pada ansambel pohon keputusan (Syarovy et al., 2023). Dalam *Random Forest*, subset acak dari fitur dipilih untuk setiap pohon, dan fitur pemisahan terbaik dari subset tersebut digunakan untuk membagi setiap node. Ini menghasilkan sekumpulan pohon yang beragam dan tidak terlalu berkorelasi satu sama lain, yang meningkatkan kinerja model secara keseluruhan (Siregar et al., 2022). Keunggulan utama dari *Random Forest Regression* adalah kemampuannya menangani data berdimensi tinggi dan banyak fitur tanpa menghasilkan model yang overfitting. Selain itu, metode ini dapat mengukur pentingnya fitur untuk memahami peran relatif setiap fitur dalam dataset, serta digunakan untuk tugas klasifikasi dan regresi pada data kategorikal maupun numerikal.

Salah satu kelemahan utama dari *Random Forest Regression* adalah kebutuhan komputasi yang tinggi dan penggunaan memori yang besar. Meskipun tidak selalu menjadi algoritma paling akurat, *Random Forest* umumnya dianggap sebagai algoritma yang andal dan sering digunakan sebagai tolak ukur untuk algoritma lain. Kinerjanya mungkin tidak optimal pada data yang mengandung banyak noise atau hubungan yang sangat non-linear antara fitur dan variabel target (Amelia, 2023). Langkah-langkah utama dalam membangun model *Random Forest* meliputi pemilihan dataset, preprocessing data, pembagian data, pembentukan model dengan beberapa pohon keputusan, pelatihan model, pembuatan prediksi, dan evaluasi model menggunakan metrik yang sesuai.



2.2 Matrix Pengujian

Metrik pengujian digunakan untuk menilai kualitas model statistik atau machine learning. Dalam proyek machine learning, tahap evaluasi sama pentingnya dengan tahap pelatihan model. Ada berbagai jenis metrik evaluasi untuk masalah regresi. Penelitian ini menggunakan beberapa metrik evaluasi, termasuk *Mean Absolute Error* (MAE), yang mengukur rata-rata selisih absolut antara nilai aktual dan prediksi model. MAE dihitung dengan menjumlahkan semua selisih absolut antara nilai aktual dan prediksi, kemudian membaginya dengan total jumlah data.

$$MAE = \frac{\sum_{i=1}^n |y_i - \hat{y}_i|}{n} \quad (1)$$

Dalam rumus ini Y mewakili nilai aktual, \hat{Y} mewakili nilai prediksi, dan n adalah total jumlah data. MAE memberikan ukuran sederhana tentang akurasi prediksi, di mana nilai yang lebih rendah menunjukkan kinerja model yang lebih baik.

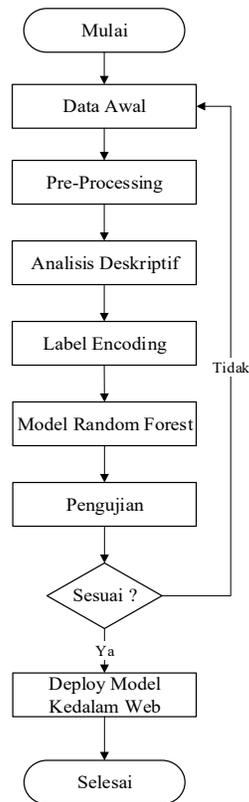
Mean Squared Error (MSE) adalah metrik yang mengukur rata-rata kuadrat selisih antara nilai aktual dan prediksi model. MSE dihitung dengan menjumlahkan semua kuadrat selisih antara nilai aktual dan prediksi, lalu membaginya dengan total jumlah data.

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (2)$$

Dalam rumus ini Y mewakili nilai aktual, \hat{Y} mewakili nilai prediksi, dan n adalah total jumlah data. MSE memberikan ukuran tentang seberapa besar kesalahan prediksi yang dihasilkan oleh model, dengan nilai yang lebih rendah menunjukkan akurasi prediksi yang lebih baik.

3. METODE PENELITIAN

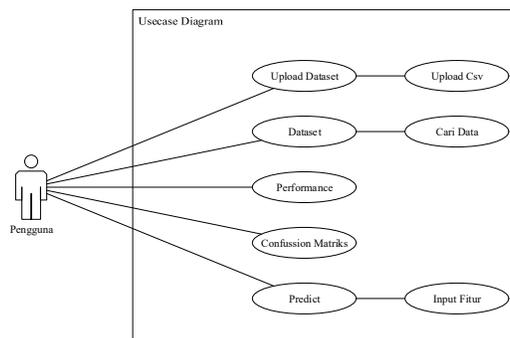
Penelitian ini menggunakan metodologi CRISP-DM, sebuah metodologi data mining yang terdiri dari empat tingkat abstraksi: phases, generic tasks, specialized tasks, dan process instances. Diagram alir data digunakan untuk memvisualisasikan langkah-langkah dan proses yang terlibat dalam data mining, dengan desain yang ditunjukkan pada Gambar berikut :



Gambar 1. Perancangan Diagram Air

Proses dimulai dengan pengambilan data awal berbentuk .csv dari *Kaggle*, yang kemudian melalui tahap *preprocessing* untuk menentukan variabel *y* dan *x* sebagai persiapan *machine learning*. Setelah itu, dilakukan analisis deskriptif untuk memahami data, diikuti dengan penggunaan label encoder untuk mengonversi data kualitatif menjadi angka. Model Random Forest digunakan untuk mendapatkan fitur penting dan memprediksi harga smartphone. Pengujian model dilakukan dengan RMSE, MAE, MAPE, dan confusion matrix; jika hasilnya tidak memuaskan, proses diulang dari awal. Jika memuaskan, model akan dideploy ke aplikasi web.

Perancangan diagram use case bertujuan untuk menggambarkan bagaimana sistem berinteraksi dengan pengguna. Rancangan diagram use case ini adalah sebagai berikut:



Gambar 2. Use Case Diagram



Gambar di atas menunjukkan use case diagram dari sistem yang dirancang, yang mencakup beberapa fitur utama: pengguna dapat mengunggah dataset dalam format .CSV, melihat data training dan jumlah data yang digunakan, serta melakukan pencarian data. Selain itu, pengguna dapat melihat hasil kinerja model, seperti scatter plot yang diperoleh dari 80% data training dan 20% data uji. Confusion matrix juga ditampilkan berdasarkan 20% data uji. Terakhir, pengguna dapat melakukan prediksi harga smartphone dengan memilih fitur-fitur tertentu yang akan diproses oleh sistem.

Pemilihan CRISP-DM dilakukan dengan mempertimbangkan metode lain seperti *Knowledge Discovery in Databases* (KDD) untuk keperluan deployment aplikasi dan integrasi bisnis. Pada tahap *Business Understanding*, analisis dilakukan terhadap industri Smartphone di Indonesia dan penelitian sebelumnya untuk menentukan kesesuaian dengan tujuan prediksi harga Smartphone. Pada tahap *Data Understanding*, data diambil dari Kaggle.com dengan struktur yang sesuai untuk permasalahan prediksi harga, dengan jumlah fitur 20 dan label numerik kontinu, serta jumlah sampel 1359 (<https://www.kaggle.com/>). Fitur-fitur tersebut dipilih berdasarkan studi literatur yang mengindikasikan bahwa fitur-fitur seperti RAM, kapasitas penyimpanan, prosesor, layar, dan baterai memiliki pengaruh yang signifikan terhadap harga Smartphone. Tahap *Data Preparation* melibatkan ekstraksi data, pemilihan fitur, dan *preprocessing* seperti mengubah label harga ke dalam mata uang rupiah, tanpa menghapus outlier atau mengubah data lainnya karena kualitas data yang sudah baik. Selanjutnya, dilakukan analisis deskriptif untuk menggambarkan karakteristik dasar dari data yang ada.

Pada tahap *Modelling* dilakukan pelatihan model menggunakan algoritma *Random Forest* dengan data yang telah dipreprocessing sebelumnya. Jumlah pohon yang dibuat untuk pelatihan ditentukan, dan setelah proses bagging dilakukan untuk membuat pohon-pohon tersebut, aturan keputusan dihasilkan sebagai model prediksi. Penggunaan jumlah pohon yang cukup besar seperti 42 membantu mengurangi risiko overfitting pada data pelatihan. Pada tahap *Evaluation*, model yang telah terpilih dievaluasi menggunakan beberapa parameter seperti *Root Mean Square Error* (RMSE), *Mean Absolute Error* (MAE), dan *Mean Absolute Percentage Error* (MAPE). Data dipisahkan menjadi data latih dan data uji untuk evaluasi, dan Confusion Matrix digunakan untuk evaluasi prediksi. Seluruh evaluasi divisualisasikan menggunakan *Google Collabs*. Tahap *Deployment* melibatkan penggabungan website yang dibangun dengan *framework Streamlite* dengan model prediksi yang telah dipilih untuk menciptakan sebuah alat yang dapat melakukan prediksi harga smartphone berdasarkan spesifikasi pengguna. Selain itu, penelitian ini juga mencakup variabel penelitian seperti variabel ordinal, variabel non-ordinal, dan variabel kuantitatif, serta teknik pengumpulan data yang melibatkan studi literatur dan pengambilan data training dari Kaggle.com.



4. HASIL DAN PENGUJIAN

Pada tahap ini data dalam format .CSV mengalami preprocessing sebagai persiapan untuk *machine learning*. Proses *preprocessing* meliputi penentuan fitur (X) dan label (y), di mana fitur berperan sebagai input dalam sistem dan label sebagai output atau hasil akhir. Dalam tabel penentuan fitur dan label, 18 variabel digunakan untuk fitur (X), sedangkan name dan model tidak termasuk dalam fitur. Selanjutnya, dilakukan analisis deskriptif untuk merangkum karakteristik dasar dari data. Analisis ini bertujuan untuk memberikan pemahaman yang lebih baik tentang data dengan menjelaskan, menggambarkan, dan menginterpretasikan data, serta memberikan informasi seperti rata-rata, deviasi standar, nilai minimum dan maksimum, serta persentil dari setiap kolom numerik dalam DataFrame. Tahap analisis deskriptif ini penting untuk memperoleh wawasan awal tentang distribusi dan statistik dasar dari data. Gambar 1 juga ditampilkan menggunakan kode program describe.

	Battery capacity (mAh)	Screen size (inches)	Resolution X	Resolution Y	Processor	RAM (MB)	Internal storage (GB)	Rear camera	Front camera	Number of SIMs	Price
count	1359.000000		1359.000000	1359.000000	1359.000000	1359.000000	1359.000000	1359.000000	1359.000000	1359.000000	1.359000e+03
mean	2638.489330		5.291210	811.549345	1490.777778	5.551194	2408.777778	30.654864	12.070199	7.037969	1.883701
std	873.514123		0.671257	270.707271	557.780120	2.189562	1664.442386	36.950241	8.948337	6.285448	0.374457
min	1010.000000		2.400000	240.000000	320.000000	1.000000	64.000000	0.064000	0.000000	0.000000	1.000000
25%	2300.000000		5.000000	720.000000	1280.000000	4.000000	1000.000000	8.000000	8.000000	2.000000	2.000000
50%	3000.000000		5.200000	720.000000	1280.000000	4.000000	2000.000000	16.000000	12.200000	5.000000	2.000000
75%	3500.000000		5.700000	1080.000000	1920.000000	8.000000	3000.000000	32.000000	13.000000	8.000000	2.000000
max	6000.000000		7.300000	2160.000000	3840.000000	10.000000	12000.000000	512.000000	108.000000	48.000000	3.000000

Gambar 3. Statistik Deskriptif Data Describe

Dari gambar 3 dalam analisis statistic di atas, kita memiliki data frame dengan 1359 entri atau data baris yang menggambarkan beberapa variabel penting terkait berbagai aspek seperti *battery capacity*, *screen size* dan yang lainnya.

4.1 Label *EnCoding*

Dikarenakan pemanfaatan pembelajaran mesin pada bahasa pemrograman Python tidak dapat dijalankan dalam skenario multi kelas atau lebih tepatnya ketika variabel input (X) dan output (y) memiliki tipe data yang tidak seragam, perlu dilakukan proses label enCoder. Label enCoder merujuk pada transformasi dari data dengan tipe string menjadi data bertipe integer. Untuk gambaran yang lebih komprehensif, silakan lihat ilustrasi dalam gambar 2 yang menunjukkan ragam tipe data asli dari rekaman yang diperoleh.



```
df.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1359 entries, 0 to 1358
Data columns (total 21 columns):
#   Column                                Non-Null Count  Dtype
---  ---                                -
0   Name                                  1359 non-null   object
1   Brand                                 1359 non-null   object
2   Model                                 1359 non-null   object
3   Battery capacity (mAh)                1359 non-null   int64
4   Screen size (inches)                  1359 non-null   float64
5   Touchscreen                           1359 non-null   object
6   Resolution x                           1359 non-null   int64
7   Resolution y                           1359 non-null   int64
8   Processor                             1359 non-null   int64
9   RAM (MB)                              1359 non-null   int64
10  Internal storage (GB)                  1359 non-null   float64
11  Rear camera                            1359 non-null   float64
12  Front camera                           1359 non-null   float64
13  Operating system                       1359 non-null   object
14  Wi-Fi                                  1359 non-null   object
15  Bluetooth                              1359 non-null   object
16  GPS                                     1359 non-null   object
17  Number of SIMs                         1359 non-null   int64
18  3G                                       1359 non-null   object
19  4G/ LTE                                1359 non-null   object
20  Price                                  1359 non-null   int64
dtypes: float64(4), int64(7), object(10)
memory usage: 223.1+ KB
```

Gambar 4. Bentuk Tipe Data Awal

Tipe data pada *dataset* ini terbagi menjadi numerik dan kategorik Ada 10 Fitur/Variabel dengan tipe data kategorik, Sisanya 11 Fitur lainnya memiliki tipe data numerik. Dari Informasi diatas juga dapat diketahui bahwa tidak ada data yang bernilai kosong/null.

```
Data columns (total 18 columns):
#   Column                                Non-Null Count  Dtype
---  ---                                -
0   Battery capacity (mAh)                1359 non-null   int64
1   Screen size (inches)                  1359 non-null   float64
2   Touchscreen                           1359 non-null   object
3   Resolution x                           1359 non-null   int64
4   Resolution y                           1359 non-null   int64
5   Processor                             1359 non-null   int64
6   RAM (MB)                              1359 non-null   int64
7   Internal storage (GB)                  1359 non-null   float64
8   Rear camera                            1359 non-null   float64
9   Front camera                           1359 non-null   float64
10  Operating system                       1359 non-null   object
11  Wi-Fi                                  1359 non-null   object
12  Bluetooth                              1359 non-null   object
13  GPS                                     1359 non-null   object
14  Number of SIMs                         1359 non-null   int64
15  3G                                       1359 non-null   object
16  4G/ LTE                                1359 non-null   object
17  Price                                  1359 non-null   int64
dtypes: float64(4), int64(7), object(7)
```

Gambar 5. Bentuk Tipe Data Akhir

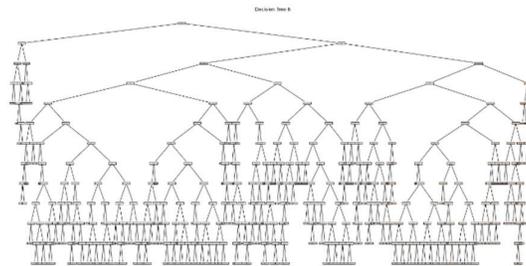
Pada gambar 5 merupakan bentuk tipe data akhir yang digunakan untuk melakukan prediksi nantinya.

4.2 Model *Random Forest*

Sebuah model yang tercipta melalui analisis *Random Forest* dibangun berdasarkan sampel data yang telah diambil. Model ini kemudian akan diaplikasikan kembali pada data baru dengan harapan dapat mencerminkan kondisi aktual. Dari 20 variabel yang digunakan dalam penelitian ini, sebanyak 10 fitur memiliki tipe data kategorikal, sedangkan 11 fitur lainnya memiliki tipe data numerik. Jumlah total data yang dimanfaatkan adalah sebanyak 1359. Sebelum pembentukan model *Random Forest*, tahap awal yang perlu dilakukan adalah mengonversi data kategorikal menjadi bilangan menggunakan skala nominal.



150	0.361 = 36.1%	MAE : 763968.4992942335 MSE : 2974743912428.204 RMSE : 1724744.5933900487 Cross Validation Score : 3449057830251.4517 MAPE : 0.36193607460267274
200	0.357 = 35.7%	MAE : 757086.6546362721 MSE : 2924952653638.8916 RMSE : 1710249.295757461 Cross Validation Score : 3450127261010.4595 MAPE : 0.3578354959619927

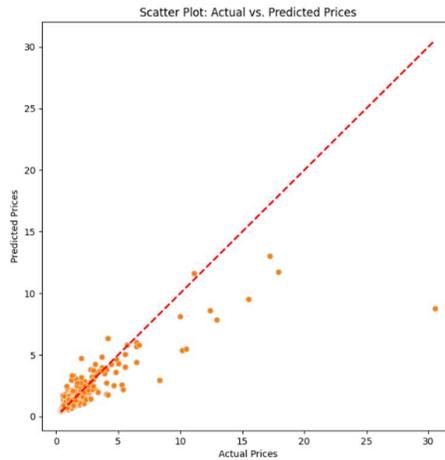


Gambar 7. Hasil Pohon Keputusan

Dalam penelitian ini, pengujian dilakukan untuk menentukan jumlah pohon yang menghasilkan nilai error terkecil. Berdasarkan data pada Tabel 1, dimulai dengan 50 pohon yang menghasilkan Mean *Absolute Percentage Error* (MAPE) sebesar 35.6%, menunjukkan nilai error terendah. Ketika jumlah pohon ditingkatkan menjadi 100, nilai MAPE sedikit meningkat menjadi 35.9%. Penambahan jumlah pohon hingga 150 menunjukkan peningkatan nilai error tertinggi, mencapai 36.1%. Sementara itu, pada 200 pohon, nilai MAPE sedikit menurun menjadi 35.7%. Dari hasil pengujian ini, dapat disimpulkan bahwa jumlah pohon terbaik yang digunakan dalam penelitian ini adalah 50 pohon, karena memiliki nilai error terendah sebesar 35.6% dibandingkan dengan jumlah pohon lainnya yang diuji. Selanjutnya, perbandingan antara harga aktual dengan harga prediksi menggunakan data yang ada menghasilkan hasil seperti yang terlihat pada Tabel 2.

Tabel 2. Pengujian Data Aktual dan Data Prediksi

No	Data Aktual	Hasil Prediksi
1	Rp.10.914.630	Rp. 10.684.726
2	Rp.5.179.815	Rp. 4.846.365
3	Rp.19.776.500	Rp. 11.042.634
4	Rp.6.462.050	Rp. 5.789.314
5	Rp.4.290.150	Rp. 4.484.183



Gambar 8. Perbandingan Data Prediksi dan Data Awal

Grafik pada Gambar 8 menampilkan perbandingan antara prediksi data testing dan data aktual. Meskipun prediksi model cenderung mendekati nilai aktual, terdapat kemungkinan nilai prediksi berada di atas atau di bawah nilai aktual. Hal ini menunjukkan bahwa penggunaan algoritma *Random Forest* dapat membantu dalam mengeksplorasi data untuk memahami pengaruh variabel independen terhadap variabel dependen. Dengan menggunakan algoritma yang sesuai, model yang dihasilkan memiliki tingkat akurasi yang cukup baik, sehingga nilai prediksi cenderung mendekati nilai aktual meskipun masih memiliki selisih.

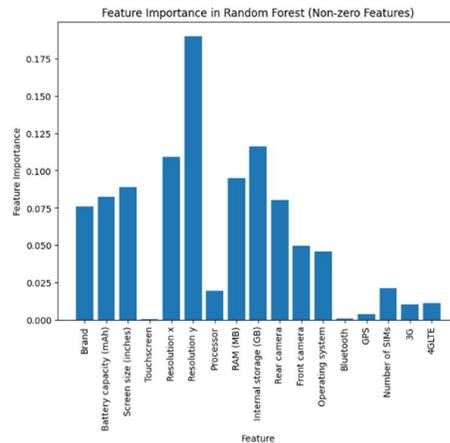
Tabel 3. Hasil Nilai Evaluasi

RMSE	MAE	MAPE	Akurasi
1745144	760969	35.6%	64.4%

Berikut adalah penjelasan tentang nilai-nilai evaluasi model yang dihasilkan dari model prediksi harga *smartphone*:

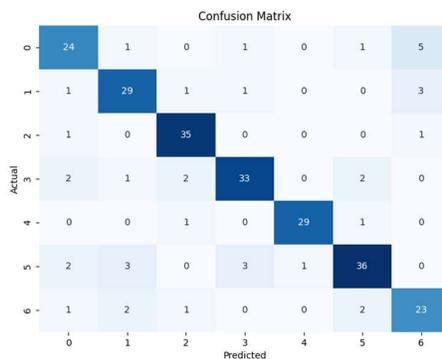
1. Dalam konteks ini, RMSE bernilai sekitar 1,745,144. Ini mengindikasikan seberapa jauh rata-rata nilai prediksi dari nilai sebenarnya dalam satuan mata uang yang digunakan.
2. Dalam kasus ini, MAE memiliki nilai sekitar 760,969. Ini mengindikasikan rata-rata kesalahan prediksi dalam satuan mata uang yang digunakan.
3. Dalam kasus ini, MAPE memiliki nilai sekitar 35,6%. Ini menunjukkan persentase rata-rata kesalahan prediksi sebagai bagian dari nilai sebenarnya.

Dalam konteks ini, nilai-nilai ini digunakan untuk mengevaluasi seberapa baik model dapat memprediksi harga *smartphone*. Semakin rendah nilai RMSE, MAE, serta semakin rendah nilai MAPE, semakin baik kinerja model tersebut. Sebaliknya, nilai yang lebih tinggi menunjukkan adanya kesalahan yang lebih besar dalam prediksi.



Gambar 9. Fitur Importance

Langkah berikutnya adalah mengevaluasi signifikansi variabel independen terhadap harga smartphone menggunakan nilai fitur penting (feature importance), sebagai alternatif karena kompleksitas pohon keputusan yang sulit diinterpretasi. Gambar 9 menampilkan nilai fitur penting yang membantu dalam memahami peran variabel dalam pembentukan model dan prediksi. Dari penelitian ini, ditemukan bahwa variabel resolution y memiliki signifikansi tertinggi dengan nilai 0.1903 dalam menentukan harga smartphone. Selanjutnya, internal storage (GB) memiliki nilai 0.1161, resolution x dengan 0.1094, RAM dengan 0.0951, screen size (inches) dengan 0.0888, dan seterusnya. Variabel dengan tingkat signifikansi tinggi menjadi faktor penentu dalam akurasi model prediksi karena prediksi akan lebih mendekati nilai aktual.



Gambar 10. Confusion Matrix

Hasil pengujian prediksi data uji dalam penelitian ini dievaluasi melalui penggunaan Confusion Matrix, yang memberikan gambaran menyeluruh tentang kemampuan model dalam mengklasifikasikan data. Sebagaimana ditunjukkan pada Gambar 10 di atas, terdapat 209 prediksi yang benar untuk data uji ini. Angka 209 ini didapatkan dengan menambahkan jumlah true positive dari setiap kelas, yaitu 24 untuk kelas A, 29 untuk kelas B, 35 untuk kelas C, 33 untuk kelas D, 29 untuk kelas E, 36 untuk kelas F, dan 23 untuk kelas G. Selain itu, terdapat 40 prediksi yang salah. Nilai 40 ini berasal dari jumlah false positive dan false negative untuk setiap



kelas, yang masing-masing dijelaskan secara rinci. Melalui Confusion Matrix, kinerja model klasifikasi dapat dianalisis secara mendalam, memberikan informasi penting untuk evaluasi dan perbaikan model. Dengan demikian, penggunaan Confusion Matrix menjadi langkah krusial dalam validasi dan peningkatan model klasifikasi.

5. KESIMPULAN DAN SARAN

Adapun kesimpulan yang didapat dalam penelitian ini adalah sebagai berikut:

1. Berdasarkan hasil penelitian ini, dapat disimpulkan bahwa prediksi harga smartphone menggunakan metode *Random forest* memberikan hasil yang cukup baik dalam memodelkan dan memprediksi harga smartphone, karena analisa yang dilakukan mampu menampilkan hasil prediksi harga dengan menggunakan metode random forest. Penggunaan teknik statistic deskriptif dan visualisasi data membantu memahami karakteristik dataset, sehingga fitur-fitur yang dihasilkan dari fitur engineering memperkaya informasi yang tersedia dan membantu dalam meningkatkan performa model.
2. Berdasarkan hasil pengujian menggunakan MAPE, MAE dan RMSE mendapatkan Tingkat error yang cukup rendah yaitu untuk MAPE mendapatkan nilai 35.6%, MAE mendapatkan selisih nilai Rp. 760,969, dan RMSE mendapatkan selisih nilai Rp.1,745,144. Meskipun terdapat perbedaan Tingkat error hasil ini menunjukkan bahwa metode Random forest memiliki performa yang cukup baik dalam memprediksi harga smartphone berdasarkan fitur yang ada.

Berdasarkan temuan dari penelitian ini, terdapat beberapa rekomendasi yang dapat diberikan untuk meningkatkan akurasi prediksi harga smartphone. Pertama, mempertimbangkan penambahan fitur atau variabel tambahan yang dapat berpotensi dalam prediksi, seperti rating marketplace, jenis prosesor, dan faktor lainnya yang relevan. Kedua, menggunakan dataset atau data training yang lebih besar dan beragam untuk meningkatkan akurasi model.



DAFTAR PUSTAKA:

- Amelia, Y. (2023). Perbandingan Metode Machine Learning Untuk Mendeteksi Penyakit Jantung. *IDEALIS%: InDonEsiA journal Information System*, 6(2), 220–225. <https://doi.org/10.36080/idealis.v6i2.3043>
- Asim, M., & Khan, Z. (2018). Mobile Price Class prediction using Machine Learning Techniques. *International Journal of Computer Applications*, 179(29), 6–11. <https://doi.org/10.5120/ijca2018916555>
- Hasibuan, E., Informasi, S., Ilmu, F., Informasi, T., Gunadarma, U., Margonda, J., No, R., Cina, P., & Jawa, D. (2022). Implementasi Machine Learning untuk Prediksi Harga Mobil Bekas dengan Algoritma Regresi Linear berbasis Web. *Jurnal Ilmiah Komputasi*, 21(4), 595–602. <https://doi.org/10.32409/jikstik.21.4.3327>
- Muhayat, T. (2022). Prediksi Harga Smartphone Menggunakan Algoritma Multiple Linear Regression. *Skripsi*.
- Saadah, S., & Salsabila, H. (2021). Prediksi Harga Bitcoin Menggunakan Metode Random Forest. *Jurnal Komputer Terapan*, 7(1), 24–32. <https://doi.org/10.35143/jkt.v7i1.4618>
- Sari, S. A. P., & Kusdiyanto, K. (2024). Analysis of the Effect of Marketing Mix (4p) on Purchase Decisions Fashion Products on Shopee's E-commerce (Case study on the Z Generation of the City of Surakarta). *Proceedings Series on Social Sciences & Humanities*, 15(Iconomics 2023), 127–141. <https://doi.org/10.30595/pssh.v15i.941>
- Siregar, B., Anthon Pangruruk, F., & Widjaja, P. A. (2022). Comparison of Various Forecasting Models for the Jakarta Stock Exchange Composite (JKSE) During the Covid-19 Pandemic. *Jurnal Multidisiplin Madani (MUDIMA)*, 2(2), 1035–1046.
- Suwarni, S., Aprika, L., & Indriasari, N. (2020). Analisis Bauran Pemasaran Pada Pt. Global Imoo Telekomunikasi Bengkulu. *EKOMBIS REVIEW: Jurnal Ilmiah Ekonomi dan Bisnis*, 8(1), 90–101. <https://doi.org/10.37676/ekombis.v8i1.935>
- Syarovy, M., Nugroho, A. P., & Sutiarmo, L. (2023). Pemanfaatan Model Neural Network Dalam Generasi Baru Pertanian Presisi Di Perkebunan Kelapa Sawit. *WARTA Pusat Penelitian Kelapa Sawit*, 28(1), 39–54. <https://doi.org/10.22302/iopri.war.warta.v28i1.97>
- Tang, I. M. H., & Asandimitra, N. (2023). Pengaruh Mental Accounting, Regret Aversion Bias, Herding Bias, Loss Aversion, Risk Perception, dan Financial Literacy Terhadap Keputusan Investasi Investor Generasi Z. *Jurnal Ilmu Manajemen*, 11(2), 458–473.
- Widodo, T., & Qurniawati, R. S. (2017). Pengaruh Ketergantungan Pada Smartphone Dan Sikap Terhadap Perilaku Pembelian Online Pada Mahasiswa. *Among Makarti*, 9(2), 99–122. <https://doi.org/10.52353/ama.v9i2.141>
- Wisyardin, M. K., Luciana, G. M., & Pariaman, H. (2020). Pendekatan LSTM untuk Memprediksi Kondisi Motor 10 kV pada PLTU Batubara. *Kilat*, 9(2), 311–318.
- Yin, L., Chen, K., Jiang, Z., & Xu, X. (2023). A Fast Parallel Random Forest Algorithm Based on Spark. *Applied Sciences (Switzerland)*, 13(10). <https://doi.org/10.3390/app13106121>
- Sumber Internet : <https://www.kaggle.com/>