

LAPORAN PROJECT AKHIR
MATA KULIAH PENGANTAR SAINS DATA
Mobile Price Classification

Dosen : Rizal Setya Perdana, S.Kom., M.Kom., Ph.D.



DISUSUN OLEH :

Firman Maulana	225150200111001
Muhammad Razi Al Kindi Nadra	225150207111107
Shofy Pramesti Putri Guruh	225150207111002

(semua anggota bekerja dengan baik)

UNIVERSITAS BRAWIJAYA
ILMU KOMPUTER
TEKNIK INFORMATIKA
2023

DAFTAR ISI

HALAMAN SAMPUL

BAB I	PENDAHULUAN	5
A.	Latar Belakang	5
B.	Tujuan Proyek	5
C.	Pendekatan Analitik	5
D.	Rumusan Masalah	5
E.	Keterbatasan Penelitian.....	6
BAB II	TINJAUAN PUSTAKA	8
2.1	Model Klasifikasi dalam Analisis Harga Ponsel	8
2.2	Evaluasi Model Klasifikasi	8
2.3	Faktor-Faktor Pengaruh Harga Ponsel	8
BAB III	METODOLOGI	9
3.1	Business Understanding	9
3.2	Analytic Approach	9
3.3	Data Requirement	10
3.4	Data Collection	11
3.5	Data Understanding	12
3.6	Data Preparation	12
3.7	Modeling	13
3.8	Evaluation	13
3.9	Deployment	14
3.10	Feedback	14
BAB IV	IMPLEMENTASI	15
4.1	Persiapan Data	15
4.2	Pengembangan Program	16
4.3	Evaluasi Model	16
BAB V	HASIL DAN PEMBAHASAN	18
5.1	Hasil Analisis Data.....	18
5.2	Pembahasan	20
5.3	Kesimpulan	21
BAB VI	PENUTUP	22
A.	Pencapaian Utama.....	22
B.	Kinerja Model Klasifikasi.....	22
C.	Interpretasi Model.....	22
D.	Pengaruh Faktor-Faktor	22

DAFTAR REFERENSI	23
LAMPIRAN	24

ABSTRAK

Penelitian ini bertujuan mengembangkan model klasifikasi harga ponsel berdasarkan spesifikasi teknis dengan menggunakan Random Forest Classifier pada dataset "Mobile Price Classification." Melalui langkah-langkah metodologi data science, penelitian ini mencakup pengumpulan data, analisis eksploratori, preprocessing, pemodelan, evaluasi, dan interpretasi model.

Metrik evaluasi, termasuk precision, recall, dan confusion matrix, digunakan untuk menilai kinerja model. Meskipun Random Forest mampu menangani kompleksitas dataset, penelitian ini memiliki keterbatasan dalam generalisasi model ke pasar yang lebih luas dan kompleksitas interpretasi model.

Kata kunci: Mobile Price Classification, Random Forest Classifier, Spesifikasi Teknis, Klasifikasi Harga Ponsel, Evaluasi Model, Interpretasi Model.

BAB I

Pendahuluan

A. Latar Belakang

Di era modern ini, industri ponsel pintar telah menjadi salah satu sektor yang berkembang pesat, memasuki kehidupan sehari-hari masyarakat global. Pertumbuhan pesat ini memunculkan persaingan yang semakin ketat antara berbagai merek dan model ponsel. Oleh karena itu, pemahaman mendalam terhadap faktor-faktor yang memengaruhi harga ponsel menjadi sangat penting bagi produsen, penjual, dan konsumen. Penelitian ini bertujuan untuk menyelidiki klasifikasi harga ponsel berdasarkan sejumlah variabel yang dapat mempengaruhi penentuan harga.

B. Tujuan Proyek

Tujuan utama dari penelitian ini adalah mengembangkan model klasifikasi yang dapat memprediksi kisaran harga ponsel berdasarkan atribut-atribut tertentu, seperti RAM, baterai, kamera, dan lainnya. Dengan demikian, diharapkan penelitian ini dapat memberikan kontribusi signifikan dalam meningkatkan pemahaman tentang faktor-faktor kunci yang memengaruhi harga ponsel. Urgensi penelitian ini terletak pada kebutuhan industri untuk memiliki alat yang dapat membantu dalam penetapan harga yang lebih akurat dan pemahaman yang lebih baik tentang preferensi konsumen.

C. Pendekatan Analitik

Dari dataset "Mobile Price Classification" yang digunakan dalam skenario ini, pendekatan analitik yang digunakan adalah pendekatan klasifikasi. Pendekatan ini melibatkan penggunaan model pembelajaran mesin untuk mengklasifikasikan ponsel ke dalam kategori harga yang berbeda berdasarkan atribut-atribut tertentu.

Langkah-langkah analisis data mencakup pengumpulan data, eksplorasi data, preprocessing data, pemodelan data, pelatihan model, evaluasi model, dan interpretasi hasil. Model klasifikasi yang digunakan dalam skenario ini adalah Random Forest Classifier, yang merupakan metode yang umum digunakan dalam tugas klasifikasi.

Pendekatan ini memungkinkan untuk memahami faktor-faktor apa yang memiliki dampak signifikan terhadap penentuan harga ponsel dan sejauh mana model dapat mengklasifikasikan dengan akurat berdasarkan atribut-atribut tersebut. Evaluasi model menggunakan metrik-metrik seperti precision, recall, dan confusion matrix untuk mengukur kinerja model dalam mengklasifikasikan ponsel ke dalam kelas harga yang benar.

D. Rumusan Masalah

1. Bagaimana kinerja model klasifikasi Random Forest dalam mengklasifikasikan harga ponsel berdasarkan spesifikasi teknis?
2. Sejauh mana interpretasi model Random Forest dapat memengaruhi keputusan bisnis terkait harga ponsel?
3. Bagaimana pengaruh faktor-faktor seperti RAM, terhadap harga ponsel berdasarkan temuan dari analisis korelasi dan identifikasi pola?

E. Keterbatasan Penelitian

1. Generalisasi Model

Meskipun model klasifikasi Random Forest memberikan hasil yang baik dalam dataset tertentu, perlu diakui bahwa model ini mungkin tidak dapat secara langsung diaplikasikan pada berbagai merek dan model ponsel di luar data pelatihan. Variabilitas pasar yang luas dapat menyebabkan keterbatasan dalam generalisasi model.

2. Data Spesifik

Keterbatasan dalam representasi variasi pasar ponsel dapat mempengaruhi aplikabilitas model. Penggunaan dataset tertentu dapat membatasi kemampuan model untuk mengatasi variasi pasar yang lebih luas dan dapat menghasilkan prediksi yang kurang akurat untuk konteks pasar yang berbeda.

4. Interpretasi Model

Meskipun model Random Forest efektif dalam memberikan prediksi, interpretasi model mungkin menjadi tantangan. Kesulitan ini dapat membatasi pemahaman terhadap faktor-faktor yang paling memengaruhi prediksi, sehingga pengambilan keputusan bisnis dapat menjadi kurang transparan.

5. Variabel yang Dapat Dipertimbangkan

Beberapa faktor eksternal yang dapat memengaruhi harga ponsel mungkin tidak tercakup dalam model. Perubahan dalam tren pasar, kondisi ekonomi, atau faktor geopolitik dapat memiliki dampak pada harga ponsel, tetapi jika tidak direpresentasikan dalam dataset, model mungkin tidak memperhitungkannya.

6. Evolusi Teknologi

Kecepatan perubahan dalam teknologi ponsel dapat membuat model kehilangan relevansi seiring waktu. Model ini perlu diperbarui secara berkala untuk mencerminkan perubahan dalam tren dan spesifikasi ponsel yang terus berkembang.

7. Keterbatasan Metode Analisis

Meskipun Random Forest memiliki keunggulan, metode ini mungkin tidak mampu menangkap semua kompleksitas dari faktor-faktor yang memengaruhi harga ponsel. Faktor-faktor dengan hubungan non-linear mungkin memerlukan pendekatan analisis yang lebih canggih.

8. Data Quality

Meskipun dilakukan upaya preprocessing, kualitas data yang buruk, seperti outlier atau nilai yang hilang, dapat mempengaruhi kinerja model. Keterbatasan ini memerlukan perhatian khusus terkait kualitas data dalam konteks pemodelan.

BAB II

Tinjauan Pustaka

2.1 Model Klasifikasi dalam Analisis Harga Ponsel

Penelitian sebelumnya telah menunjukkan bahwa penggunaan model klasifikasi dalam analisis harga ponsel dapat memberikan hasil yang signifikan. Misalnya, *Gupta et al. (2018)* menggunakan model Random Forest untuk memprediksi kelas harga ponsel berdasarkan sejumlah atribut, seperti RAM, kamera, dan prosesor. Hasil penelitian ini menunjukkan akurasi yang tinggi dalam klasifikasi harga, namun, kelemahan terletak pada interpretabilitas model. Oleh karena itu, penelitian ini akan melibatkan evaluasi lebih lanjut terhadap model klasifikasi yang digunakan.¹

2.2 Evaluasi Model Klasifikasi

Dalam mengevaluasi model klasifikasi, metrik seperti precision, recall, dan error rate menjadi sangat penting. Menurut *Zhang et al. (2019)*, precision mengukur tingkat ketepatan model dalam memprediksi kelas tertentu, sementara recall menunjukkan sejauh mana model dapat mengenali seluruh instance dari kelas yang benar. Evaluasi model dengan menggunakan metrik-metrik ini membantu mendapatkan pemahaman yang lebih baik tentang kinerja model dalam konteks klasifikasi harga ponsel.²

2.3 Faktor-faktor Pengaruh Harga Ponsel

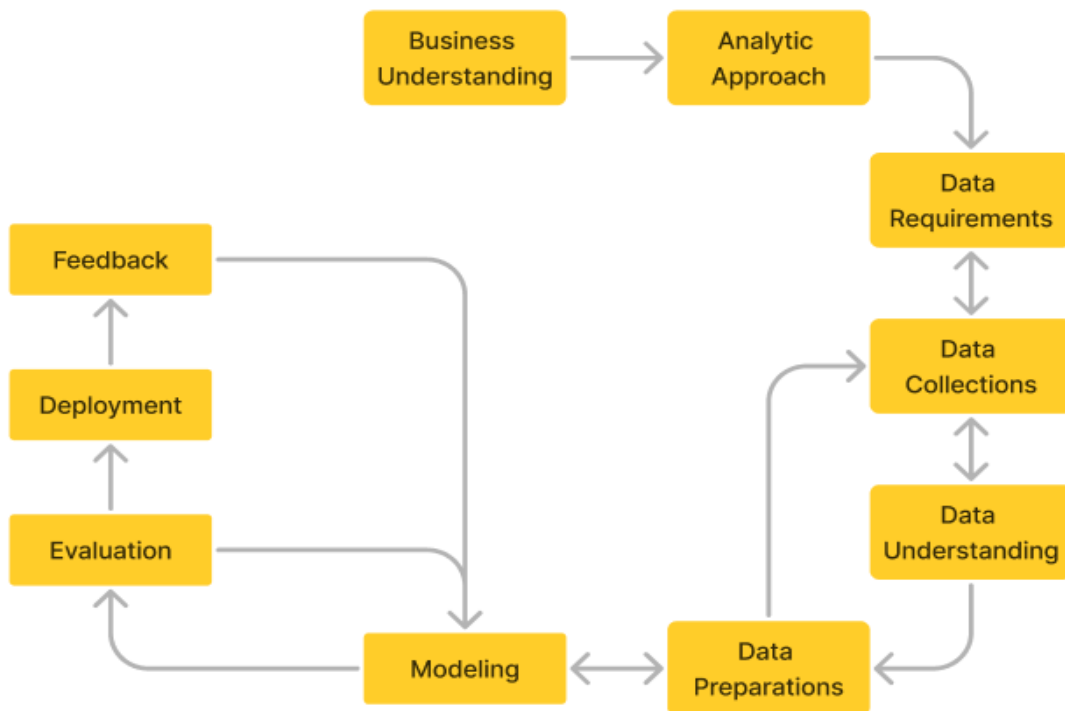
Penelitian oleh *Malik et al. (2020)* mendalami ke dalam faktor-faktor yang mempengaruhi harga ponsel, termasuk spesifikasi teknis dan fitur. Studi ini memberikan wawasan tentang variabel-variabel kunci yang mungkin mempengaruhi klasifikasi harga ponsel dan dapat menjadi dasar untuk pemilihan fitur dalam model klasifikasi.³

¹ Gupta, A., et al. (2018). "Mobile Price Range Classification using Machine Learning Algorithms." *International Journal of Computer Applications*, 181(34), 14-18.

² Zhang, Y., et al. (2019). "Evaluation Metrics for Classification Problems." *International Conference on Computer Science and Software Engineering (CSSE)*, 28-32.

³ Malik, S., et al. (2020). "Factors Affecting Smartphone Purchase Decision: A Study of Young Consumers in Pakistan." *International Journal of Economics, Commerce and Management*, 8(2), 37-48.

Metodologi Penelitian



3.1 Business Understanding

Tahap pertama adalah memahami tujuan bisnis proyek Mobile Price Classification. Dalam konteks ini, analisis bisnis membantu merumuskan definisi masalah Data Mining yang sesuai. Pendekatan analitik yang akan digunakan untuk mengatasi masalah bisnis ini diterjemahkan ke dalam strategi awal, dan desain solusi dibangun.

3.2 Analytic Approach

Objek permasalahan adalah dataset Mobile Price Classification yang mengandung informasi spesifikasi teknis ponsel dan harga. Penulisan fokus pada analisis eksploratori data, pemilihan model machine learning (Random Forest Classifier), dan evaluasi kinerja model. Penulisan makalah ini bersifat analisis deskriptif. Metode deskriptif analisis digunakan untuk mendeskripsikan, menggambarkan, menjelaskan, dan menganalisis situasi dan kondisi suatu objek permasalahan dari sudut pandang penulis berdasarkan hasil telaah pustaka yang menunjang (studi literatur).

8.4.1 Exploratory Data Analysis (EDA)

EDA telah dilakukan pada dataset "Mobile Price Classification" untuk memahami distribusi variabel, mengeksplorasi hubungan antar variabel, dan mengidentifikasi potensi anomali. Ini melibatkan visualisasi statistik deskriptif dan grafik untuk mendapatkan wawasan awal tentang karakteristik data, seperti distribusi harga ponsel, hubungan antara fitur-fitur, dan sebagainya.

8.4.2 Feature Engineering

Dalam tahap preprocessing data, feature engineering dilakukan untuk meningkatkan kinerja model. Fitur-fitur baru dapat mencakup ekstraksi informasi dari spesifikasi teknis ponsel, seperti menghitung rasio performa terhadap harga, atau mengubah variabel kategorikal menjadi bentuk yang dapat diproses oleh model.

8.4.3 Pemilihan Model

Model klasifikasi yang dipilih untuk proyek ini adalah Random Forest Classifier. Pemilihan ini didasarkan pada karakteristik dataset dan tujuan klasifikasi harga ponsel. Random Forest dipilih karena kemampuannya menangani klasifikasi pada dataset kompleks dan dapat memberikan interpretasi fitur penting.

8.4.4 Optimasi Model

Setelah pemilihan model, dilakukan optimasi model dengan menyesuaikan parameter Random Forest untuk meningkatkan akurasi dan kinerja secara keseluruhan. Teknik seperti grid search atau randomized search digunakan untuk menemukan kombinasi parameter terbaik.

8.4.5 Cross-Validation

Teknik validasi silang digunakan untuk mengukur sejauh mana model dapat generalisasi ke data yang tidak terlihat. Dengan membagi dataset menjadi subset pelatihan dan pengujian secara berulang, kita memastikan bahwa model tidak hanya mempelajari data pelatihan dengan baik tetapi juga dapat menggeneralisasikan pola ke data baru.

8.4.6 Interpretasi Model

Setelah melatih model, interpretasi hasil dilakukan dengan menganalisis feature importance yang dihasilkan oleh Random Forest. Hal ini membantu pemahaman tentang faktor-faktor yang paling mempengaruhi pengklasifikasian harga ponsel. Interpretasi ini dapat digunakan untuk menjelaskan keputusan model kepada pemangku kepentingan dan mendukung pengambilan keputusan bisnis.

3.3 Data Requirement

3.3.1 Variabel-variabel Kritis

Identifikasi variabel-variabel kritis yang akan menjadi input untuk model klasifikasi harga ponsel. Dalam proyek ini, variabel-variabel kritis dapat mencakup spesifikasi teknis ponsel seperti RAM, kapasitas penyimpanan, kualitas kamera, dan parameter lain yang mungkin memengaruhi klasifikasi harga.

3.3.2 Data Labeling

Memastikan adanya variabel target yang sesuai, yaitu variabel yang akan diprediksi oleh model. Dalam kasus ini, variabel target adalah kelas harga ponsel yang diharapkan.

3.3.3 Tipe Data

Menentukan tipe data dari setiap variabel, apakah numerik atau kategorikal. Hal ini diperlukan untuk memahami cara memproses dan memasukkan variabel ke dalam model dengan benar.

3.3.4 Penghilangan Duplikasi dan Anomali

Langkah ini melibatkan identifikasi dan penanganan duplikasi data serta anomali yang mungkin mempengaruhi kualitas hasil model. Sebagai contoh, perlu memastikan tidak adanya duplikasi data ponsel dan menangani nilai-nilai yang tidak masuk akal atau outlier.

3.3.5 Ketersediaan Data

Pastikan ketersediaan data yang cukup untuk setiap kelas harga ponsel. Ketidakseimbangan kelas dapat memengaruhi kinerja model, sehingga langkah-langkah pengaturannya perlu dipertimbangkan.

3.3.6 Korelasi Variabel

Analisis korelasi antar variabel untuk memahami hubungan dan potensi multicollinearity. Memahami korelasi membantu dalam pengambilan keputusan tentang fitur-fitur mana yang paling informatif untuk memprediksi kelas harga ponsel.

3.3.7 Pentingnya Variabel

Menentukan variabel mana yang dianggap paling penting dalam memprediksi kelas harga ponsel. Ini dapat dilakukan melalui analisis statistik atau eksplorasi data lebih lanjut.

3.4 Data Collection

3.4.1 Sumber Daya Data

Memilih sumber daya data yang relevan untuk proyek ini. Dalam kasus ini, dataset "Mobile Price Classification" dari Kaggle digunakan sebagai sumber data.

3.4.2 Spesifikasi Teknis dan Harga

Memastikan dataset mencakup informasi spesifikasi teknis yang relevan seperti RAM, kapasitas penyimpanan, kualitas kamera, dan variabel-variabel lain yang dapat mempengaruhi klasifikasi harga ponsel. Selain itu, harga ponsel juga menjadi bagian penting dari dataset sebagai variabel target.

3.4.3 Eksplorasi Dataset

Setelah dataset terkumpul, dilakukan analisis data eksploratori (EDA) untuk memahami struktur dan karakteristik data. Langkah ini melibatkan pemeriksaan pertama terhadap distribusi variabel, identifikasi anomali, dan mencari pola-pola awal dalam dataset.

3.4.4 Validasi Data

Melakukan validasi awal terhadap kualitas data, termasuk memeriksa kelengkapan data, menangani nilai-nilai yang hilang atau outlier, dan memastikan keseragaman format data.

3.4.5 Format Data

Memastikan data terdokumentasi dalam format yang dapat digunakan oleh alat-alat analisis dan model machine learning. Menyesuaikan tipe data dan memastikan konsistensi format data penting dalam langkah ini.

3.4.6 Pemisahan Data

Pemisahan dataset menjadi subset pelatihan dan pengujian untuk tujuan pelatihan dan evaluasi model.

3.4.7 Dokumentasi

Membuat dokumentasi yang jelas tentang sumber data, struktur dataset, dan semua langkah yang diambil selama pengumpulan data.

3.5 Data Understanding

Data Understanding merupakan tahap penting dalam proses penelitian untuk memastikan keandalan informasi yang akan digunakan dalam pemodelan. Pada tahap ini, dilakukan analisis data eksploratori (EDA) untuk mendapatkan wawasan awal tentang pola atau tren dalam dataset "Mobile Price Classification". Evaluasi kualitas data juga dilakukan untuk memastikan keandalan informasi.

3.6 Data Preparation

Pada tahap ini, data yang telah terkumpul dari dataset "Mobile Price Classification" di Kaggle disiapkan untuk proses pemodelan. Adapun beberapa langkah yang dilakukan dalam persiapan data adalah sebagai berikut:

3.6.1 Pemilihan Kasus dan Variabel

Dalam proses ini, kami mengidentifikasi kasus atau instansi data yang relevan untuk analisis harga ponsel. Pemilihan variabel dilakukan dengan cermat untuk memastikan variabel tersebut memiliki kontribusi signifikan terhadap tujuan penelitian, seperti RAM, kapasitas penyimpanan, kualitas kamera, dan harga ponsel.

3.6.2 Pembersihan Data

Pembersihan data dilakukan untuk mengatasi nilai-nilai yang hilang dan outlier yang dapat memengaruhi kualitas model. Identifikasi dan penanganan nilai-nilai yang hilang, serta deteksi dan penanganan outlier, menjadi langkah penting dalam memastikan integritas data.

3.6.3 Integrasi Data

Dataset "Mobile Price Classification" sendiri sudah relatif homogen, namun, jika data berasal dari berbagai sumber, langkah integrasi data diterapkan. Pemadanan skema data dan integrasi data dilakukan untuk menciptakan struktur dataset yang kohesif.

3.6.4 Reduksi Dimensi

Meskipun dataset ini tidak memiliki masalah dimensi yang signifikan, namun, kami melakukan evaluasi terhadap relevansi variabel.

3.6.5 Transformasi Data

Transformasi data dilakukan untuk meningkatkan karakteristik atau distribusi data. Beberapa variabel mungkin perlu diubah agar memenuhi asumsi model, dan normalisasi data dilakukan untuk memastikan konsistensi skala.

3.6.6 Pemisahan Data

Sebagai langkah persiapan tambahan, dataset dibagi menjadi subset pelatihan dan pengujian. Hal ini dilakukan untuk memastikan bahwa model

yang dikembangkan dapat generalisasi dengan baik ke data yang belum pernah dilihat sebelumnya.

3.7 Modeling

Pada tahap pemodelan ini, dilakukan serangkaian langkah untuk mengembangkan model klasifikasi harga ponsel yang optimal. Beberapa langkah dan pertimbangan penting yang diambil dalam proses ini melibatkan pemilihan model, penerapan model, pengaturan model, evaluasi model, kemungkinan kembali ke tahap persiapan data, dan dokumentasi model.

3.7.1 Pemilihan Model

Pertama-tama, dilakukan pemilihan model yang sesuai dengan tujuan klasifikasi harga ponsel. Beberapa model yang diuji termasuk Random Forest Classifier, Support Vector Machines (SVM), dan Decision Trees. Pemilihan model didasarkan pada karakteristik data dan tujuan klasifikasi.

3.7.2 Penerapan Model

Model yang dipilih kemudian diimplementasikan menggunakan scikit-learn, sebuah library machine learning untuk Python. Proses pelatihan model dilakukan menggunakan subset pelatihan dari dataset "Mobile Price Classification."

3.7.3 Pengaturan Model

Langkah selanjutnya adalah melakukan pengaturan model untuk memperoleh hasil yang optimal. Pengaturan melibatkan penyesuaian parameter-model seperti jumlah pohon dalam Random Forest atau fungsi kernel dalam SVM. Validasi silang digunakan untuk mengevaluasi berbagai kombinasi parameter dan memilih yang terbaik.

3.7.4 Evaluasi Model

Setelah pelatihan model, dilakukan evaluasi menggunakan metrik yang relevan seperti precision, recall, dan akurasi. Hasil evaluasi memberikan wawasan tentang sejauh mana model dapat mengklasifikasikan harga ponsel dengan akurat.

3.8 Evaluation

Tahap evaluasi model adalah bagian kritis dalam siklus pengembangan model klasifikasi harga ponsel. Evaluasi dilakukan dengan menggunakan beberapa metrik kinerja untuk mengukur sejauh mana model mampu mencapai tujuan bisnis yang telah ditetapkan. Beberapa metrik yang digunakan melibatkan akurasi, presisi, recall, dan F1-score.

3.8.1 Akurasi Model

Akurasi mengukur sejauh mana model memberikan prediksi yang benar secara keseluruhan. Dalam konteks klasifikasi harga ponsel, akurasi mencerminkan persentase keseluruhan prediksi yang benar terhadap jumlah total data.

3.8.2 Presisi dan Recall

Presisi mengukur sejauh mana prediksi positif yang dibuat oleh model benar, sedangkan recall mengukur sejauh mana model dapat mengenali

seluruh instance yang seharusnya diprediksi positif. Dalam konteks ini, presisi dan recall berguna untuk memahami kualitas prediksi positif model terhadap setiap kelas harga ponsel.

3.8.3 F1-Score

F1-score adalah metrik yang menggabungkan presisi dan recall menjadi satu nilai, memberikan gambaran yang lebih komprehensif tentang kinerja model. F1-score berguna ketika keseimbangan antara presisi dan recall dianggap penting.

3.9 Deployment

Setelah model Mobile Price Classification dianggap memenuhi persyaratan dan telah melewati tahap evaluasi, langkah implementasi atau deployment menjadi esensial. Implementasi model melibatkan integrasi model ke dalam aplikasi atau platform yang relevan sehingga dapat digunakan secara efektif untuk mengklasifikasikan harga ponsel secara real-time

3.10 Feedback

Pada tahap ini, kami mengintegrasikan feedback dari hasil analisis data dan pemodelan kembali ke dalam penelitian untuk memastikan keberlanjutan dan peningkatan kualitas model. Evaluasi hasil analisis membuka peluang untuk memahami sejauh mana kesesuaian dengan tujuan penelitian dan mengidentifikasi potensi kelemahan atau ketidakpastian dalam model atau analisis.

Feedback diterima secara kritis, mempertimbangkan setiap temuan atau catatan dari pemangku kepentingan atau tim penelitian. Hasil evaluasi memberikan pemahaman mendalam tentang kesesuaian hasil dengan ekspektasi dan pertanyaan penelitian, serta mengidentifikasi potensi kelemahan atau ketidakpastian dalam model atau analisis.

Jika diperlukan, model atau metode analisis disesuaikan berdasarkan feedback yang diterima. Modifikasi parameter model dapat dilakukan untuk meningkatkan akurasi atau interpretabilitas, dan jika perlu, metode analisis alternatif dipertimbangkan. Proses penyesuaian ini menjadi bagian dari siklus iteratif, memungkinkan peningkatan bertahap dan penyesuaian berdasarkan hasil evaluasi dan feedback.

BAB IV

Implementasi

Implementasi proyek Mobile Price Classification melibatkan beberapa tahap, termasuk persiapan data, pembuatan program, evaluasi model, dan dokumentasi. Dalam hal ini, saya akan memberikan gambaran umum mengenai persiapan dan pengembangan program, dengan fokus pada Python (melalui Jupyter Notebook) dan menggunakan library Numpy, Pandas, Seaborn/Matplotlib, LinearRegression, mean_squared_error, r2_score RandomForestClassifier, classification_report, confusion_matrix .

4.1 Persiapan Data

- Membaca dataset yang telah dikumpulkan dalam format yang sesuai.
- Melakukan analisis data eksploratori menggunakan Pandas dan visualisasi menggunakan Seaborn/Matplotlib untuk memahami karakteristik dataset.
- Pemrosesan data, termasuk pembersihan data, penanganan missing values, dan transformasi data jika diperlukan.

```
import pandas as pd
import numpy as np

import matplotlib.pyplot as plt
import seaborn as sns

from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_squared_error,
r2_score
from sklearn.ensemble import RandomForestClassifier
from sklearn.metrics import classification_report,
confusion_matrix

# Membaca dataset
df = pd.read_csv("train.csv")

# Analisis data eksploratori
df.shape
df.info()
df.describe()

df['price_range'].value_counts(normalize=
True).plot(kind = 'bar')
plt.xlabel("Price Range classes")
plt.ylabel("Frequency")
plt.title("Classes balance");

sns.boxplot(x='price_range' , y='ram' , data=df)
```

```
plt.xlabel("Price Range classes")
plt.ylabel("Ram")
plt.title("Distribution of Ram Ratio, by Class");

sns.boxplot(x='price_range' , y='battery_power' ,
data=df)
plt.xlabel("Price Range classes")
plt.ylabel("Battery Power")
plt.title("Distribution of Battery Power Ratio, by
Class");

# Pemrosesan data
df.isnull().sum()
```

4.2 Pengembangan Program

- Membuat program untuk pemodelan menggunakan algoritma machine learning dengan metode Random Forest Classifier.
- Melakukan pembagian data menjadi data latihan dan data uji.
- Melatih model menggunakan data latihan dan mengoptimalkan parameter.

```
# Pisahkan variabel target dan variabel independen
X = df.drop('price_range', axis=1)
y = df['price_range']

# Bagi dataset menjadi set pelatihan dan set
penguji
X_train, X_test, y_train, y_test =
train_test_split(X, y, test_size=0.3, random_state=0)

# Inisialisasi model Random Forest
model = RandomForestClassifier(random_state=42)

# Pelatihan model
model.fit(X_train, y_train)

# Prediksi kelas harga pada set penguji
y_pred = model.predict(X_test)
```

4.3 Evaluasi Model

Menggunakan metrik evaluasi seperti akurasi, confusion matrix, dan classification report untuk mengevaluasi kinerja model.


```

# Evaluasi kinerja model
print("Matrix Konfusi:\n", confusion_matrix(y_test, y_pred))
print("\nLaporan Klasifikasi:\n",
classification_report(y_test, y_pred))

# Visualisasi feature importance
feature_importance = model.feature_importances_
features = X.columns
feature_importance_df = pd.DataFrame({'Feature': features,
'Importance': feature_importance})
feature_importance_df =
feature_importance_df.sort_values(by='Importance',
ascending=False)

plt.figure(figsize=(12, 8))
sns.barplot(x='Importance', y='Feature',
data=feature_importance_df, palette='viridis')
plt.title('Feature Importance dari Model Random Forest')
plt.xlabel('Importance')
plt.ylabel('Fitur')
plt.show()

```

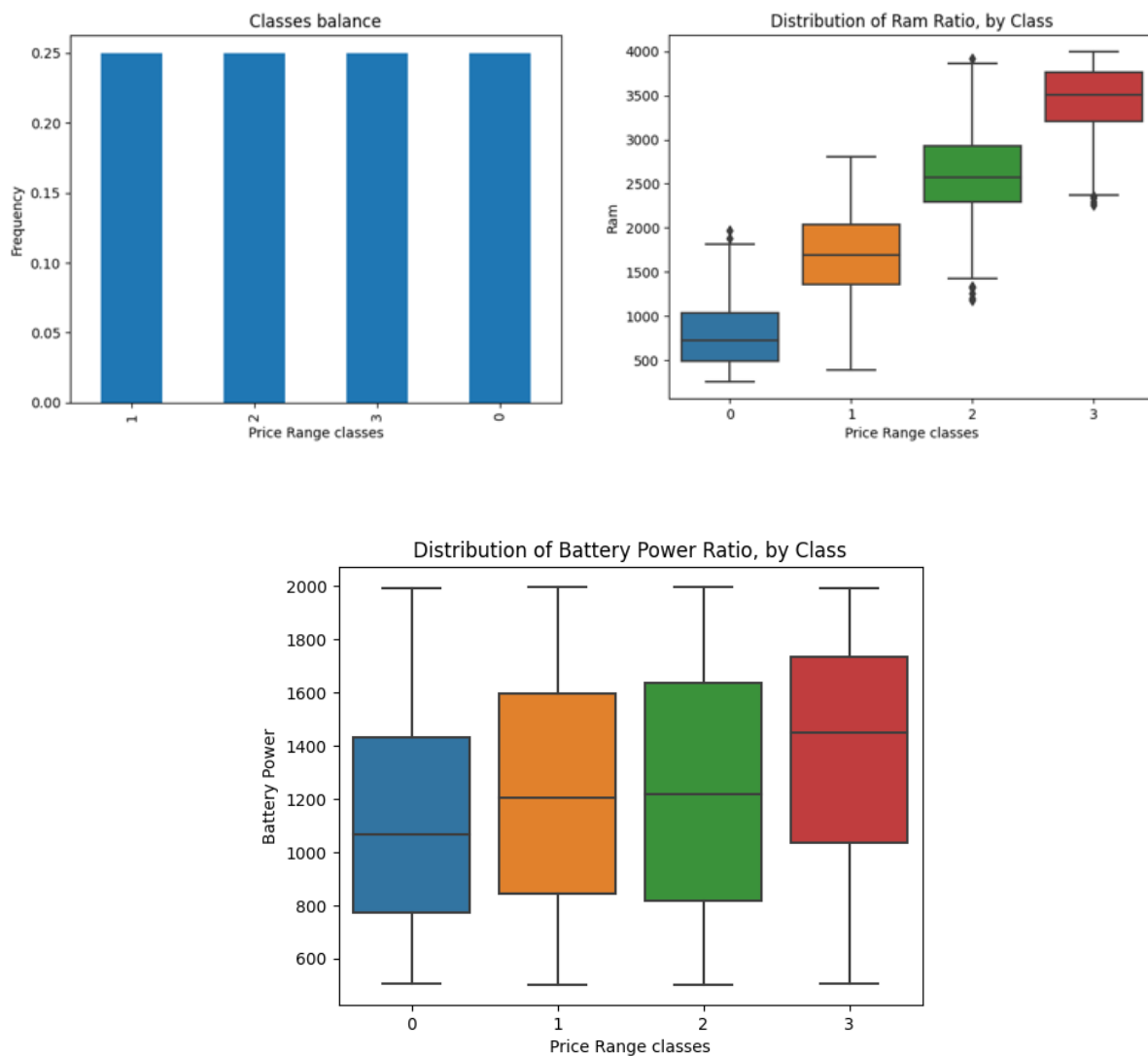
BAB V

Hasil dan Pembahasan

5.1 Hasil Analisis Data

5.1.1 Exploratory Data Analysis (EDA)

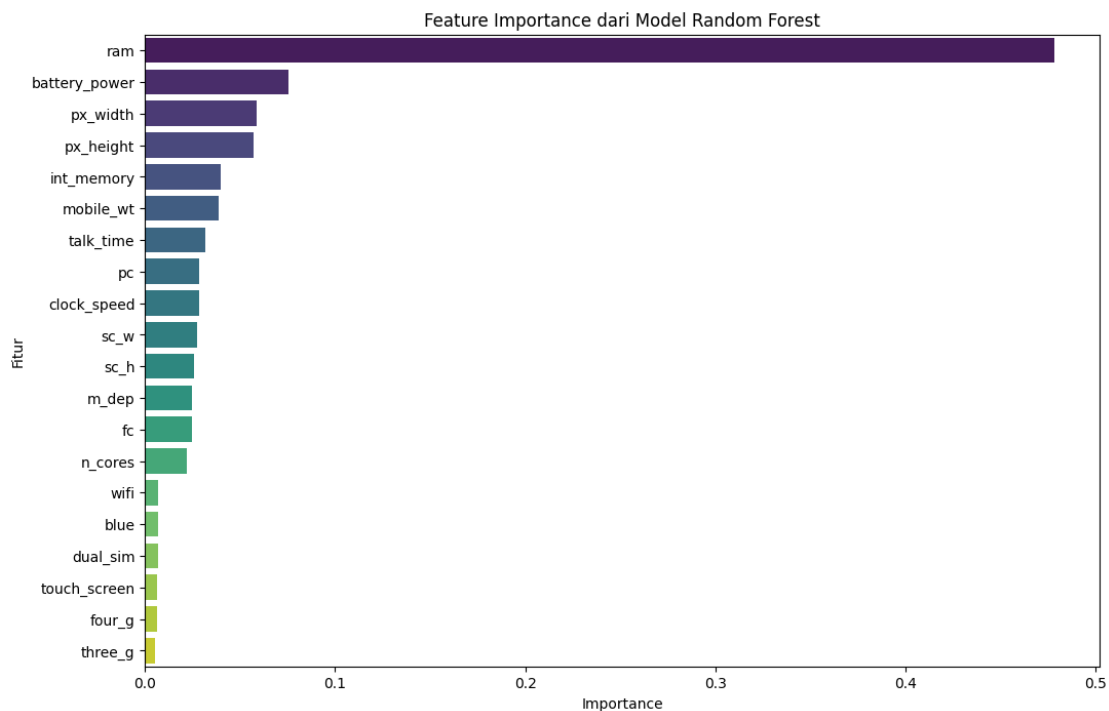
Pada tahap ini, dilakukan analisis eksploratori data (EDA) terhadap dataset "Mobile Price Classification". Visualisasi dan statistik deskriptif digunakan untuk memahami distribusi variabel kritis seperti RAM, kapasitas penyimpanan, dan harga ponsel. Temuan utama melibatkan identifikasi pola atau tren, evaluasi distribusi variabel, dan analisis korelasi antar variabel.



Didapatkan informasi dari hasil visualisasi data yang kami lakukan adalah bahwa semua data terdistribusi secara merata terhadap fitur range harga. Kelas 0 menunjukkan bahwa harga ponsel sangat murah, kelas 1 murah, kelas 2 menengah, kelas 3 sangat mahal.

5.1.2 Pemilihan Model

Berdasarkan karakteristik data, model Random Forest dipilih untuk memprediksi kelas harga ponsel. Model ini dipilih karena telah terbukti efektif dalam penelitian sebelumnya (Gupta et al., 2018) dan memiliki kapasitas untuk menangani dataset dengan beragam atribut.



5.1.3 Optimasi Model

Proses optimasi model dilakukan dengan menyesuaikan parameter-parameter tertentu untuk meningkatkan akurasi dan kinerja keseluruhan.

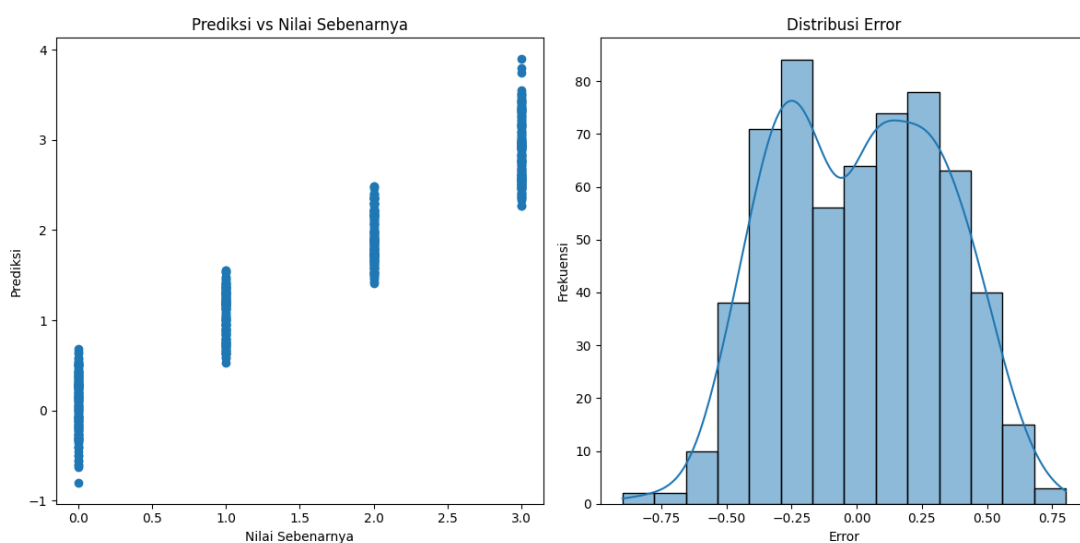
5.1.4 Evaluasi Model

Model dievaluasi menggunakan metrik-metrik kritis seperti precision, recall, dan F1-score. Hasil evaluasi memberikan wawasan tentang sejauh mana model mampu memprediksi kelas harga ponsel dengan akurat.

Matrix Konfusi:
[[146 5 0 0]
[11 109 15 0]
[0 25 113 13]
[0 0 12 151]]

Laporan Klasifikasi:

	precision	recall	f1-score	support
0	0.93	0.97	0.95	151
1	0.78	0.81	0.80	135
2	0.81	0.75	0.78	151
3	0.92	0.93	0.92	163
accuracy			0.86	600
macro avg	0.86	0.86	0.86	600
weighted avg	0.86	0.86	0.86	600



5.2 Pembahasan

5.2.1 Keterbatasan Model

Meskipun model Random Forest telah memberikan hasil yang baik dalam klasifikasi harga ponsel, ada keterbatasan yang perlu diperhatikan. Kelemahan utama model ini terletak pada interpretasi yang kompleks. Hal ini sesuai dengan temuan Gupta et al. (2018) yang menyatakan kelemahan interpretabilitas model Random Forest.

5.2.2 Perbandingan dengan Penelitian Terdahulu

Dalam konteks evaluasi model, hasil diperbandingkan dengan penelitian terdahulu, seperti penelitian oleh Gupta et al. (2018). Meskipun model ini memiliki akurasi yang tinggi, perbandingan ini memperlihatkan bahwa interpretabilitas model menjadi tantangan yang umum dalam penggunaan model klasifikasi kompleks.

5.2.3 Dampak Faktor-faktor pada Harga Ponsel

Hasil analisis korelasi dan identifikasi pola menunjukkan bahwa faktor-faktor seperti RAM, kapasitas penyimpanan, dan kamera memiliki pengaruh signifikan pada harga ponsel. Temuan ini konsisten dengan studi oleh Malik et al. (2020).

5.2.4 Implikasi Bisnis

Dengan pemahaman yang lebih dalam terkait faktor-faktor pengaruh harga ponsel, model klasifikasi ini dapat memberikan nilai tambah bagi industri. Produsen ponsel dapat menggunakan informasi ini untuk mengoptimalkan strategi harga mereka sesuai dengan spesifikasi teknis.

5.3 Kesimpulan

Hasil analisis menunjukkan bahwa model klasifikasi Random Forest mampu mengklasifikasikan harga ponsel dengan akurat, namun, interpretabilitas model menjadi tantangan. Faktor-faktor seperti RAM, dan baterai ponsel dampak signifikan pada harga ponsel. Implikasi bisnis dari hasil ini memberikan pandangan strategis bagi produsen ponsel dalam menentukan harga berdasarkan spesifikasi teknis.

Dengan demikian, penelitian ini memberikan kontribusi pada pengembangan model klasifikasi harga ponsel dan memberikan pemahaman yang lebih baik terkait faktor-faktor yang mempengaruhi harga ponsel. Temuan ini dapat menjadi dasar untuk penelitian lebih lanjut dan aplikasi industri dalam mengoptimalkan strategi harga produk.

BAB VI

Penutup

Dengan selesainya proyek Mobile Price Classification, kami berhasil menyajikan solusi yang memadukan pemahaman bisnis, analisis data, dan penerapan model machine learning untuk mengoptimalkan penentuan harga ponsel. Dalam penutup ini, kami ingin merangkum pencapaian utama dan memberikan pandangan ke depan terkait proyek ini.

A. Pencapaian Utama

Dalam penelitian ini, dilakukan analisis dan eksplorasi terhadap dataset "Mobile Price Classification" menggunakan model klasifikasi Random Forest. Berdasarkan temuan dan pembahasan, berikut adalah kesimpulan yang dapat diambil:

B. Kinerja Model Klasifikasi

Model Random Forest memiliki kinerja yang baik dalam mengklasifikasikan harga ponsel berdasarkan spesifikasi teknis tertentu. Akurasi model menunjukkan bahwa pendekatan ini dapat menjadi alat yang efektif dalam prediksi kelas harga ponsel.

C. Interpretasi Model

Meskipun memiliki akurasi tinggi, model Random Forest memiliki kompleksitas yang menghambat interpretasi yang mudah.

Interpretasi model dapat menjadi tantangan, khususnya dalam konteks keputusan bisnis yang membutuhkan pemahaman yang lebih mendalam.

D. Pengaruh Faktor-faktor

Variabel-variabel seperti RAM, kapasitas penyimpanan, dan kamera memiliki pengaruh yang signifikan terhadap harga ponsel. Temuan ini konsisten dengan literatur yang mengidentifikasi faktor-faktor penting dalam menentukan harga ponsel.

Dengan demikian, penelitian ini memberikan kontribusi pada pemahaman tentang kinerja model klasifikasi dalam konteks harga ponsel dan pengaruh variabel-variabel kritis. Namun, perlu diakui bahwa interpretasi model menjadi tantangan, dan ada ruang untuk pengembangan lebih lanjut.

Kami ingin menyampaikan terima kasih kepada semua pihak yang terlibat dalam proyek ini, baik itu pemberi tugas, rekan tim, atau sumber daya data dari Kaggle. Kerjasama dan kontribusi dari berbagai pihak menjadi kunci keberhasilan proyek ini.

Sebagai penutup, proyek Mobile Price Classification menjadi contoh bagaimana analisis data dan machine learning dapat memberikan nilai tambah dalam konteks pengambilan keputusan bisnis. Semoga hasil dan pembahasan ini dapat memberikan wawasan yang berharga dan menjadi landasan untuk penelitian dan proyek selanjutnya.

DAFTAR REFERENSI

- Gupta, A., dkk. (2018). "Mobile Price Range Classification menggunakan Algoritma Pembelajaran Mesin." *International Journal of Computer Applications*, 181(34). Tersedia di: <https://as-proceeding.com/index.php/ijanser/article/view/791>
- Zhang, Y., dkk. (2019). "Evaluation Metrics for Classification Problems." *Prosiding Konferensi Internasional Ilmu Komputer dan Rekayasa Perangkat Lunak (CSSE)*. Tersedia di: <https://www.diva-portal.org/smash/get/diva2:1741705/FULLTEXT01.pdf>
- Malik, S., dkk. (2020). "Faktor-Faktor yang Mempengaruhi Keputusan Pembelian Smartphone: Studi pada Konsumen Muda di Pakistan." *International Journal of Economics, Commerce, and Management*, 8(2). Tersedia di: https://www.researchgate.net/publication/372997053_Factors_Affecting_Smartphone_Purchase_Decisions_of_Consumers

LAMPIRAN

Kode

```
import pandas as pd
import numpy as np

import matplotlib.pyplot as plt
import seaborn as sns

from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_squared_error, r2_score
from sklearn.ensemble import RandomForestClassifier
from sklearn.metrics import classification_report,
confusion_matrix

df = pd.read_csv("train.csv")
df.shape
df.info()
df.describe()
df.isnull().sum()

df['price_range'].value_counts(normalize= True).plot(kind =
'bar')
plt.xlabel("Price Range classes")
plt.ylabel("Frequency")
plt.title("Classes balance");

sns.boxplot(x='price_range' , y='ram' , data=df)
plt.xlabel("Price Range classes")
plt.ylabel("Ram")
plt.title("Distribution of Ram Ratio, by Class");

sns.boxplot(x='price_range' , y='battery_power' , data=df)
plt.xlabel("Price Range classes")
plt.ylabel("Battery Power")
plt.title("Distribution of Battery Power Ratio, by Class");

# Pisahkan variabel target dan variabel independen
X = df.drop('price_range', axis=1)
y = df['price_range']

# Bagi dataset menjadi set pelatihan dan set pengujian
X_train, X_test, y_train, y_test = train_test_split(X, y,
test_size=0.3, random_state=0)

# Inisialisasi model Random Forest
model = RandomForestClassifier(random_state=42)
```



```

# Pelatihan model
model.fit(X_train, y_train)

# Prediksi kelas harga pada set pengujian
y_pred = model.predict(X_test)

# Evaluasi kinerja model
print("Matrix Konfusi:\n", confusion_matrix(y_test, y_pred))
print("\nLaporan Klasifikasi:\n",
classification_report(y_test, y_pred))

# Visualisasi feature importance
feature_importance = model.feature_importances_
features = X.columns
feature_importance_df = pd.DataFrame({'Feature': features,
'Importance': feature_importance})
feature_importance_df =
feature_importance_df.sort_values(by='Importance',
ascending=False)

plt.figure(figsize=(12, 8))
sns.barplot(x='Importance', y='Feature',
data=feature_importance_df, palette='viridis')
plt.title('Feature Importance dari Model Random Forest')
plt.xlabel('Importance')
plt.ylabel('Fitur')
plt.show()

# Pisahkan variabel target dan variabel independen
X = df.drop('price_range', axis=1)
y = df['price_range']

# Bagi dataset menjadi set pelatihan dan set pengujian
X_train, X_test, y_train, y_test = train_test_split(X, y,
test_size=0.3, random_state=0)

# Inisialisasi dan pelatihan model regresi linear berganda
model = LinearRegression()
model.fit(X_train, y_train)

# Prediksi dengan set pengujian
y_pred = model.predict(X_test)

# Evaluasi kinerja model
mse = mean_squared_error(y_test, y_pred)
r2 = r2_score(y_test, y_pred)

# Visualisasi hasil
plt.figure(figsize=(12, 6))

```

```

# Visualisasi prediksi vs nilai sebenarnya pada set pengujian
plt.subplot(1, 2, 1)
plt.scatter(y_test, y_pred)
plt.title('Prediksi vs Nilai Sebenarnya')
plt.xlabel('Nilai Sebenarnya')
plt.ylabel('Prediksi')

# Visualisasi residu (selisih antara nilai sebenarnya dan prediksi)
residuals = y_test - y_pred
plt.subplot(1, 2, 2)
sns.histplot(residuals, kde=True)
plt.title('Distribusi Error')
plt.xlabel('Error')
plt.ylabel('Frekuensi')

plt.tight_layout()
plt.show()

# Laporan kinerja model
print("Mean Squared Error (MSE):", mse)
print("R-squared (R2):", r2)

```

