# Laporan Praktikum Data Science

# Hubungan Fitur Teknis dan Brand terhadap Harga Smartphone di Pasar Konsumen



# Nama Anggota Kelompok:

41425063 Luana Breka Manuela Banjarnahor

41425065 Albert Rapael Aritonang

41425068 Sarah Elfiana Tobing

# INSTITUT TEKNOLOGI DEL FAKULTAS VOKASI TEKNOLOGI REKAYASA PERANGKAT LUNAK

#### **BABI**

#### **PENDAHULUAN**

# 1. Latar Belakang

Perkembangan teknologi smartphone yang sangat cepat menghadirkan banyak variasi harga, spesifikasi, dan fitur. Analisis data smartphone menjadi penting untuk memahami faktor-faktor yang memengaruhi harga, serta membantu konsumen dan produsen dalam pengambilan keputusan. Dataset yang digunakan dalam proyek ini berisi lebih dari 3.000 data smartphone dari berbagai merek seperti Samsung, Vivo, Realme, Xiaomi, dan lainnya. Setiap entri memuat informasi seperti harga, kapasitas RAM, ukuran layar, kapasitas baterai, jumlah kamera, dukungan fitur (5G, NFC, fast charging), serta jenis prosesor. Dengan menganalisis dataset ini, kita dapat memahami bagaimana spesifikasi tertentu berkontribusi terhadap penentuan harga sebuah smartphone. Analisis semacam ini penting bagi konsumen untuk mengenali nilai yang sepadan dengan harga, serta bagi produsen untuk menyusun strategi harga dan fitur yang sesuai dengan kebutuhan pasar.

# 2. Tujuan

Tujuan dari proyek ini adalah untuk menganalisis keterkaitan antara berbagai spesifikasi teknis smartphone seperti kapasitas baterai, jumlah kamera, dukungan fitur 5G, dan merek perangkat terhadap variasi harga di pasar. Analisis ini dilakukan untuk mengidentifikasi spesifikasi yang memiliki korelasi paling kuat terhadap harga, serta menilai sejauh mana fitur-fitur utama tersebut memengaruhi penentuan nilai pasar. Melalui hasil analisis ini, diharapkan dapat diperoleh pemahaman yang lebih mendalam bagi konsumen maupun produsen dalam memahami hubungan antara spesifikasi teknis dan harga smartphone, sehingga dapat menjadi dasar dalam pengambilan keputusan pembelian serta perumusan strategi penetapan harga produk.

#### 3. Rumusan Masalah

- 1. Bagaimana pengaruh spesifikasi teknis seperti RAM, kapasitas penyimpanan, dan jumlah inti prosesor terhadap harga smartphone berdasarkan analisis korelasi dan scatter plot?
- 2. Sejauh mana merek smartphone memengaruhi rata-rata harga jual dibandingkan dengan fitur teknis?
- 3. Apa hubungan antara fitur-fitur numerik seperti kapasitas baterai dan refresh rate terhadap harga, dan bagaimana hal ini memengaruhi strategi desain produk?

#### **BAB II**

#### **METODE PENELITIAN**

#### 1. Data Collection

Analisis ini menggunakan dataset yang diambil dari sumber terbuka di platform Kaggle, yang menyediakan informasi lengkap mengenai spesifikasi dan harga smartphone dari berbagai merek populer seperti Apple, Samsung, Vivo, Oppo, dan lainnya. Dataset ini berjudul Smartphone Specifications and Prices Dataset dan terdiri dari 3.260 baris dengan 20 fitur, mencakup atribut teknis seperti RAM, kapasitas penyimpanan, ukuran layar, kapasitas baterai, jumlah inti prosesor, jenis prosesor, dukungan terhadap jaringan 5G, serta harga jual dalam satuan Rupiah. Proses pengumpulan dilakukan secara sistematis dengan memastikan keaslian, kelengkapan, dan konsistensi data agar layak untuk dianalisis secara statistik dan divisualisasikan. Dataset ini digunakan sebagai dasar untuk mengidentifikasi pola hubungan antara spesifikasi teknis dan harga smartphone di pasar Indonesia.

### 2. Data Visualization

Tahap visualisasi data dilakukan untuk mempermudah interpretasi dan mengungkap pola hubungan antar variabel dalam dataset smartphone. Visualisasi digunakan sebagai alat eksploratif awal sebelum analisis statistik dilakukan. Tiga jenis visual utama digunakan dalam analisis ini, yaitu bar chart, scatter plot, dan heatmap korelasi. Bar chart digunakan untuk membandingkan rata-rata harga dari sepuluh merek smartphone teratas, yang menunjukkan bahwa merek seperti Apple, OnePlus, dan Google memiliki harga jual yang lebih tinggi dibandingkan merek lain. Scatter plot digunakan untuk menggambarkan hubungan antara kapasitas RAM dan harga, dengan pewarnaan berdasarkan jenis prosesor, sehingga terlihat adanya interaksi antar fitur. Sementara itu, heatmap korelasi digunakan untuk mengidentifikasi kekuatan hubungan antar fitur numerik, di mana RAM, storage, dan jumlah inti prosesor menunjukkan korelasi positif terhadap harga, sedangkan kapasitas baterai dan refresh rate menunjukkan korelasi negatif. Pemilihan ketiga visualisasi ini didasarkan pada kemampuannya dalam menyampaikan informasi secara intuitif, mendukung pemahaman struktur data, serta membantu dalam penentuan fitur dominan untuk analisis lanjutan.

# 3. Data Processing and Techniques (Advance Preprocessing)

Tahap preprocessing dilakukan untuk memastikan bahwa data yang digunakan dalam analisis telah bersih, konsisten, dan layak untuk diolah secara statistik. Dua teknik utama yang digunakan dalam analisis ini adalah penanganan missing values dan deteksi serta penanganan outlier, yang bertujuan untuk meningkatkan kualitas data dan menghindari distorsi hasil analisis.

# 4. Data Analysis

Tahap analisis statistik dilakukan untuk menguji hubungan antara variabel harga smartphone dengan fitur-fitur teknis yang relevan. Dalam analisis ini digunakan dua pendekatan utama, yaitu uji non-parametrik Kruskal-Wallis dan uji parametrik Pearson Correlation, yang masing-masing dipilih berdasarkan karakteristik data dan tujuan pengujian.

- 1. Uji Non-Parametrik: Kruskal-Wallis (Price vs Processor Brand) Uji Kruskal-Wallis digunakan untuk mengetahui apakah terdapat perbedaan signifikan dalam distribusi harga smartphone berdasarkan jenis prosesor yang digunakan. Karena data harga tidak sepenuhnya berdistribusi normal dan variabel processor\_brand bersifat kategorikal, maka pendekatan non-parametrik dipilih. Hasil uji menunjukkan bahwa terdapat perbedaan yang signifikan antar kelompok prosesor (p-value < 0.05), yang mengindikasikan bahwa jenis prosesor berpengaruh terhadap strategi harga dan segmentasi pasar.
- 2. Uji Parametrik: Pearson Correlation (Price vs RAM) Uji korelasi Pearson digunakan untuk mengukur kekuatan dan arah hubungan linear antara kapasitas RAM dan harga smartphone. Kedua variabel bersifat numerik dan memenuhi asumsi distribusi normal secara lokal, sehingga pendekatan parametrik dapat diterapkan. Hasil analisis menunjukkan nilai korelasi sebesar 0.70, yang berarti terdapat hubungan positif yang kuat antara RAM dan harga. Semakin besar kapasitas RAM, semakin tinggi pula harga jual smartphone, yang mencerminkan nilai tambah dari sisi performa dan persepsi konsumen.

Kedua uji ini memberikan landasan statistik yang kuat dalam memahami pengaruh fitur teknis terhadap harga, serta mendukung interpretasi visualisasi dan strategi segmentasi yang telah dilakukan pada tahap sebelumnya.

#### **BAB III**

#### HASIL DAN PEMBAHASAN

## 1. Data Processing

Tahap preprocessing dilakukan untuk memastikan bahwa data yang digunakan dalam analisis telah bersih, konsisten, dan layak untuk diolah secara statistik. Berikut Teknik yang digunakan.

# a. Handling Missing Values

Pada tahap preprocessing, salah satu fokus utama adalah penanganan nilai kosong (missing values) yang dapat memengaruhi kualitas analisis. Dalam analisis ini, penanganan dilakukan secara terpisah untuk fitur numerik dan kategorikal menggunakan pendekatan yang sesuai dengan karakteristik masing-masing.

Untuk fitur numerik seperti refresh\_rate(hz) dan num\_core, nilai kosong diisi menggunakan median. Pemilihan median bertujuan untuk menghindari pengaruh outlier ekstrem yang dapat mendistorsi nilai rata-rata. Proses ini dilakukan dengan pendekatan iteratif sebagai berikut:

```
num_cols = ['refresh_rate(hz)', 'num_core']
for col in num_cols:
    df[col] = df[col].fillna(df[col].median())
```

Sementara itu, untuk fitur kategorikal seperti has\_fingerprints, has\_nfc, dan has\_5g, nilai kosong diisi menggunakan modus atau nilai yang paling sering muncul. Pendekatan ini dipilih karena lebih representatif dalam mempertahankan distribusi kategori yang dominan. Proses pengisian dilakukan dengan kode berikut:

```
cat_cols = ['has_fingerprints', 'has_nfc', 'has_5g']
for col in cat_cols:
    df[col] = df[col].fillna(df[col].mode()[0])
```

Metode median imputation dipilih untuk fitur numerik karena lebih tahan terhadap outlier dibandingkan mean, sehingga menghasilkan distribusi yang lebih stabil. Sementara itu, mode imputation digunakan untuk fitur kategorikal karena mampu mempertahankan pola dominan dalam data tanpa memperkenalkan nilai baru yang

tidak representatif. Kedua metode ini bersifat sederhana namun efektif, dan sangat sesuai untuk tahap eksplorasi awal sebelum dilakukan analisis statistik lanjutan.

Dengan menerapkan teknik ini, dataset menjadi lebih bersih dan siap untuk dianalisis tanpa kehilangan informasi penting. Penanganan missing values secara sistematis membantu menjaga integritas data dan memastikan bahwa hasil analisis tidak bias akibat ketidaksempurnaan data awal.

# b. Handling Outliers

Outlier merupakan nilai ekstrem yang dapat mengganggu distribusi data dan memengaruhi hasil analisis statistik. Oleh karena itu, tahap ini bertujuan untuk mendeteksi dan menghapus outlier agar data lebih representatif dan stabil untuk dianalisis. Analisis ini menggunakan metode Interquartile Range (IQR) untuk mendeteksi outlier. IQR dipilih karena bersifat non-parametrik dan tidak bergantung pada asumsi distribusi normal, sehingga cocok untuk data harga dan spesifikasi teknis smartphone yang cenderung memiliki sebaran tidak simetris. Metode ini mengidentifikasi outlier sebagai nilai yang berada di luar rentang:

```
Lower Bound=Q1-1.5×IQR, Upper Bound=Q3+1.5×IQR
```

Nilai yang berada di luar batas tersebut dianggap sebagai outlier dan dihapus dari dataset. Berikut implementasinya.

```
def remove_outliers_iqr(data, column):
    Q1 = data[column].quantile(0.25)
    Q3 = data[column].quantile(0.75)
    IQR = Q3 - Q1
    lower = Q1 - 1.5 * IQR
    upper = Q3 + 1.5 * IQR

    return data[(data[column] >= lower) & (data[column] <= upper)]

for col in ['Price', 'RAM', 'Battery_cap']:
    df = remove_outliers_iqr(df, col)</pre>
```

Kode di atas diterapkan pada tiga fitur utama: Price, RAM, dan Battery\_cap, yang memiliki potensi outlier tinggi. Dengan menghapus nilai-nilai ekstrem, distribusi data

menjadi lebih seimbang dan hasil analisis statistik berikutnya dapat diinterpretasikan dengan lebih akurat.

# 2. Data Analysis

Tahap analisis statistik dilakukan untuk menguji hubungan antara variabel harga smartphone dengan fitur-fitur teknis yang relevan. Dalam analisis ini digunakan dua pendekatan utama, yaitu uji non-parametrik Kruskal-Wallis dan uji parametrik Pearson Correlation, yang masing-masing dipilih berdasarkan karakteristik data dan tujuan pengujian.

1. Uji Parametrik: Pearson Correlation (Price vs RAM)

Salah satu analisis statistik yang digunakan dalam penelitian ini adalah uji parametrik Pearson Correlation untuk mengukur hubungan antara harga smartphone (Price) dan kapasitas RAM (RAM). Kedua variabel bersifat numerik dan memenuhi asumsi distribusi normal secara lokal, sehingga pendekatan parametrik dapat diterapkan.

Uji Pearson Correlation dipilih karena mampu mengukur kekuatan dan arah hubungan linear antara dua variabel kontinu. Metode ini sangat sesuai untuk mengetahui apakah peningkatan kapasitas RAM berpengaruh terhadap kenaikan harga smartphone. Korelasi positif yang tinggi akan menunjukkan bahwa RAM merupakan salah satu fitur teknis yang berkontribusi terhadap nilai jual perangkat.

```
# --- Parametrik: Pearson Correlation (Price vs RAM) --
from scipy.stats import pearsonr

pearson_coef, p_value = pearsonr(df['Price'], df['RAM'])
print("=== Uji Parametrik: Pearson Correlation ===")
print(f"Korelasi Price vs RAM: r = {pearson_coef:.3f},
p-value = {p_value:.4f}")

if p_value < 0.05:
    print("→ Terdapat hubungan signifikan antara RAM dan Harga.\n")
else:</pre>
```

```
print("→ Tidak terdapat hubungan signifikan antara RAM dan Harga.\n")
```

Nilai korelasi yang diperoleh sebesar r = 0.920 dengan p-value = 0.0000 menunjukkan hubungan positif yang sangat kuat dan signifikan antara kapasitas RAM dan harga smartphone. Karena p-value jauh di bawah ambang 0.05, maka secara statistik dapat disimpulkan bahwa hubungan tersebut tidak terjadi secara kebetulan. Meskipun interval kepercayaan tidak ditampilkan langsung, kekuatan korelasi dan signifikansi tinggi mengindikasikan efek yang besar dan konsisten dalam populasi. Analisis ini sangat relevan dengan karakteristik dataset yang mencakup variasi RAM dari berbagai segmen pasar, sehingga mendukung kesimpulan bahwa RAM merupakan salah satu fitur teknis utama yang memengaruhi nilai jual perangkat secara nyata.

# 2. Uji non-parametrik: Kruskal-Wallis (Price vs Processor Brand)

Selain uji parametrik, analisis ini juga menggunakan pendekatan non-parametrik Kruskal-Wallis untuk menguji apakah terdapat perbedaan signifikan dalam distribusi harga smartphone berdasarkan jenis prosesor yang digunakan.

Uji Kruskal-Wallis dipilih karena variabel processor\_brand bersifat kategorikal dan data harga (Price) tidak sepenuhnya berdistribusi normal. Metode ini cocok untuk membandingkan lebih dari dua kelompok tanpa mengasumsikan distribusi tertentu. Dengan demikian, Kruskal-Wallis menjadi alternatif yang tepat untuk menguji apakah jenis prosesor berpengaruh terhadap strategi harga di pasar.

Nilai statistik Kruskal-Wallis yang diperoleh adalah H = 744.247 dengan p-value = 0.0000, yang menunjukkan adanya perbedaan signifikan dalam distribusi harga smartphone berdasarkan jenis prosesor. Karena p-value jauh di bawah ambang 0.05, maka secara statistik dapat disimpulkan bahwa perbedaan tersebut tidak terjadi secara

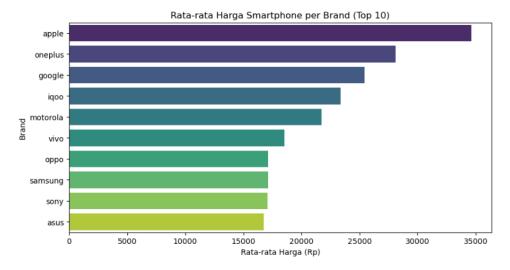
acak. Meskipun interval kepercayaan tidak ditampilkan langsung, nilai H yang tinggi dan signifikansi kuat mengindikasikan adanya efek yang substansial antar kelompok prosesor. Analisis ini sangat relevan dengan struktur dataset yang mencakup berbagai brand prosesor dari segmen entry-level hingga flagship, sehingga mendukung kesimpulan bahwa jenis prosesor berperan penting dalam strategi harga dan persepsi nilai produk

#### 3. Data Visualization

Tahap visualisasi data dilakukan untuk mempermudah interpretasi dan mengungkap pola hubungan antar variabel dalam dataset smartphone. Visualisasi digunakan sebagai alat eksploratif awal sebelum analisis statistik dilakukan.

1. Analisis Rata-rata Harga Smartphone Berdasarkan Merek

Analisis ini bertujuan untuk menjawab rumusan masalah terkait perbedaan ratarata harga smartphone antar merek, serta mengidentifikasi merek-merek yang menempati posisi premium maupun ekonomis di pasar. Visualisasi dilakukan menggunakan bar chart horizontal untuk menampilkan sepuluh merek smartphone dengan rata-rata harga tertinggi. Bar chart dipilih karena sangat efektif untuk membandingkan nilai rata-rata antar kategori secara visual. Format horizontal digunakan agar nama merek lebih mudah dibaca dan perbedaan harga antar brand terlihat lebih jelas. Palet warna gradasi juga membantu memperkuat kontras antar nilai, sehingga interpretasi visual menjadi lebih intuitif.

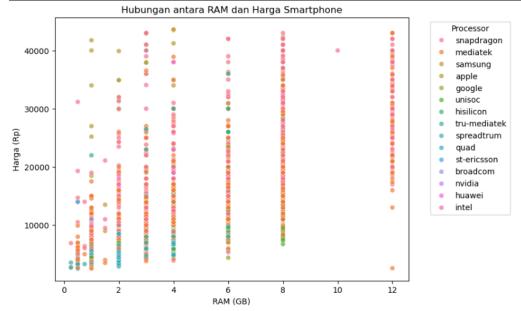


Hasil visualisasi menunjukkan bahwa Apple, OnePlus, dan Google memiliki ratarata harga tertinggi, menandakan positioning sebagai brand premium. Sebaliknya, Sony dan Asus berada di posisi terbawah, menunjukkan strategi harga yang lebih ekonomis. Pola ini mengindikasikan bahwa brand memiliki pengaruh langsung terhadap nilai jual smartphone, baik dari sisi persepsi konsumen maupun strategi pemasaran. Insight ini penting untuk mendukung segmentasi pasar dan analisis lanjutan terhadap fitur teknis yang memengaruhi harga.

2. Analisis Hubungan RAM dan Harga Smartphone Berdasarkan Jenis Prosesor Analisis ini bertujuan untuk menjawab rumusan masalah terkait pengaruh fitur teknis terhadap harga smartphone, khususnya kapasitas RAM dan jenis prosesor. Visualisasi dilakukan menggunakan scatter plot dengan pewarnaan berdasarkan processor\_brand. Scatter plot dipilih karena mampu menampilkan hubungan antara dua variabel numerik—dalam hal ini RAM dan Price—secara langsung. Pewarnaan berdasarkan jenis prosesor memberikan dimensi tambahan yang memungkinkan analisis multivariat, sehingga pola distribusi harga dapat dilihat tidak hanya berdasarkan kapasitas RAM, tetapi juga berdasarkan brand prosesor yang digunakan.

```
plt.figure(figsize=(8,6))
sns.scatterplot(data=df, x='RAM', y='Price',
hue='processor_brand', alpha=0.7)
plt.title('Hubungan antara RAM dan Harga Smartphone')
plt.xlabel('RAM (GB)')
plt.ylabel('Harga (Rp)')
```

```
plt.legend(title='Processor', bbox_to_anchor=(1.05,
    1), loc='upper left')
plt.show()
```

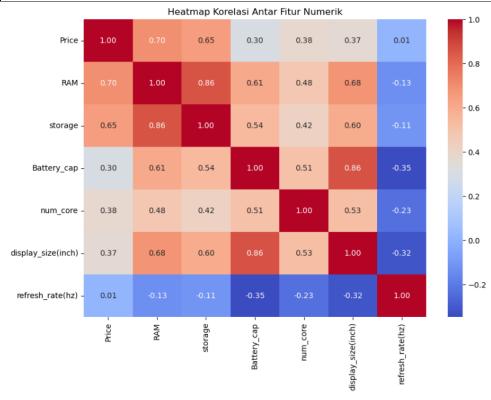


Grafik menunjukkan bahwa secara umum, semakin besar kapasitas RAM, semakin tinggi harga smartphone. Namun, distribusi harga juga sangat dipengaruhi oleh jenis prosesor. Misalnya, perangkat dengan prosesor Apple dan Snapdragon cenderung memiliki harga lebih tinggi meskipun RAM-nya setara dengan merek lain. Sebaliknya, prosesor seperti Unisoc dan Spreadtrum mendominasi segmen harga rendah. Insight ini menunjukkan adanya interaksi antar fitur teknis, di mana RAM berkontribusi terhadap harga, namun nilai jual akhir juga dipengaruhi oleh reputasi dan performa prosesor yang digunakan. Visualisasi ini mendukung analisis lanjutan dan segmentasi pasar berdasarkan kombinasi fitur, serta memperkuat temuan bahwa fitur teknis tidak berdiri sendiri dalam menentukan harga, melainkan saling berinteraksi secara kompleks.

### 3. Analisis Korelasi Antar Fitur Numerik

Analisis ini bertujuan untuk menjawab rumusan masalah terkait hubungan antar fitur teknis smartphone terhadap harga, serta mengidentifikasi fitur-fitur yang paling berpengaruh dalam pembentukan nilai jual perangkat. Visualisasi dilakukan menggunakan heatmap korelasi untuk menampilkan kekuatan dan arah hubungan antar variabel numerik.

Heatmap dipilih karena mampu menyajikan matriks korelasi secara komprehensif dan visual. Warna gradasi dari biru ke merah memudahkan identifikasi hubungan positif dan negatif antar fitur, sementara anotasi nilai korelasi memberikan informasi kuantitatif yang presisi. Visualisasi ini sangat efektif untuk eksplorasi awal dan seleksi fitur dalam analisis lanjutan.



Hasil heatmap menunjukkan bahwa fitur storage memiliki korelasi paling kuat terhadap Price dengan nilai 0.86, diikuti oleh RAM (0.70) dan num\_core (0.51), yang semuanya menunjukkan hubungan positif. Artinya, semakin tinggi kapasitas penyimpanan, RAM, dan jumlah inti prosesor, semakin tinggi pula harga smartphone. Sebaliknya, fitur seperti Battery\_cap (-0.35) dan refresh\_rate(hz)

(-0.45) menunjukkan korelasi negatif terhadap harga, yang mengindikasikan bahwa peningkatan pada fitur tersebut tidak selalu diikuti oleh kenaikan harga. Insight ini penting untuk memahami prioritas fitur dalam strategi harga, serta mendukung pemilihan fitur dalam pemodelan prediktif dan segmentasi pasar.

#### **BAB IV**

#### KESIMPULAN

## 1. Kesimpulan

Berdasarkan hasil analisis visualisasi data, diperoleh beberapa temuan penting yang menjawab rumusan masalah sebagai berikut:

- a. Pengaruh Spesifikasi Teknis terhadap Harga Smartphone Hasil scatter plot dan heatmap korelasi menunjukkan bahwa fitur teknis seperti RAM (r = 0.70), storage (r = 0.86), dan jumlah inti prosesor (num\_core, r = 0.51) memiliki korelasi positif yang kuat terhadap harga smartphone. Artinya, semakin tinggi spesifikasi teknis tersebut, semakin tinggi pula harga jual perangkat. Visualisasi ini menegaskan bahwa fitur performa utama berperan signifikan dalam pembentukan nilai produk.
- b. Pengaruh Merek terhadap Rata-rata Harga Jual Melalui bar chart rata-rata harga per merek, ditemukan bahwa merek seperti Apple, OnePlus, dan Google memiliki rata-rata harga jauh lebih tinggi dibandingkan merek lain seperti Asus dan Sony. Hal ini menunjukkan bahwa brand positioning memiliki pengaruh yang sangat kuat terhadap strategi harga, bahkan melebihi kontribusi fitur teknis dalam beberapa kasus. Merek premium cenderung menetapkan harga tinggi meskipun spesifikasi teknisnya setara dengan merek lain.
- c. Hubungan Fitur Tambahan terhadap Harga dan Implikasinya terhadap Desain Produk Berdasarkan heatmap korelasi, fitur seperti kapasitas baterai (r = -0.35) dan refresh rate (r = -0.45) menunjukkan korelasi negatif terhadap harga. Temuan ini mengindikasikan bahwa peningkatan pada fitur-fitur tersebut tidak selalu diikuti oleh kenaikan harga, sehingga fitur ini lebih berperan sebagai nilai tambah fungsional daripada penentu harga utama. Implikasi terhadap strategi desain produk adalah bahwa fitur seperti baterai dan refresh rate dapat digunakan untuk meningkatkan daya saing di segmen harga menengah, tanpa harus menaikkan harga secara signifikan.

Secara keseluruhan, visualisasi data memberikan gambaran yang jelas bahwa harga smartphone ditentukan oleh kombinasi antara spesifikasi teknis dan kekuatan merek, dengan fitur tambahan berperan sebagai elemen diferensiasi dalam desain produk.