

zenius

Kampus  
Merdeka  
INDONESIA AYU

# Final Project Presentation

Nomor Kelompok: 4

Nama Mentor: Diajeng Ciptaning Ayu

Nama Kelompok:

- Dwi Ahmad Dzulhijjah
- Rizal Fadia Al Fikri

Machine Learning Class

**Program Studi Independen Bersertifikat  
Zenius Bersama Kampus Merdeka**



1. Latar Belakang
2. Eksplorasi Data dan Visualisasi
3. Modelling
4. Kesimpulan



# Latar Belakang

# Latar Belakang Project

Sumber Data:

<https://www.kaggle.com/datasets/iabhishekofficial/mobile-price-classification>

Problem: **classification**

Tujuan:

- Mengklasifikasikan harga smartphone pada range tertentu
- Mengeksplorasi dan memperoleh insight dari data

# Latar Belakang Project

Berdasarkan dataset yang kami terima, bahwa client ingin mengetahui apa saja fitur yang mempengaruhi harga perangkat mobile.

Permasalahan bisnis ini bisa dipecahkan menggunakan metode klasifikasi pembelajaran mesin sehingga kami akan mencoba menjelaskan.

# Eksplorasi Data dan Visualisasi

# Business Understanding

Perangkat mobile adalah perangkat komputer yang dapat digenggam atau juga disebut sebagai ponsel cerdas (*smartphone*). Perangkat ini mempermudah pengguna untuk menggunakan komputer dengan mobilitas tinggi atau yang menginginkan perangkat komputer portabel. Perangkat mobile kini menjadi teknologi yang sering digunakan manusia sehari-hari.

Perangkat mobile selalu berkembang baik secara hardware maupun softwarenya. Fitur software misalnya sistem operasi dan aplikasi-aplikasi praktis. Hardware misalnya RAM, processor, kamera, dan layar. Keduanya setiap tahun selalu mengalami perkembangan mulai dari kapasitas maupun kemampuannya dalam operasi komputasi.

# Business Understanding

Fitur-fitur canggih mempengaruhi harga. Perusahaan akan menyesuaikan fitur-fitur dengan harga berdasarkan kecanggihannya. Teknologi kapasitas ram yang semakin besar dapat mempengaruhi harga menjadi semakin mahal. Tenaga baterai tertentu secara linier mempengaruhi harga.

Terdapat banyak fitur-fitur yang mempengaruhi harga dari sebuah perangkat mobile yakni tenaga baterai, ketersediaan bluetooth, dual sim, kamera depan dan kamera utama, jenis jaringan, clock speed, kapasitas penyimpanan internal, kapasitas RAM, dan lain-lain.



# Business Understanding

Mobile price atau harga perangkat mobile terdiri dari empat kategori yakni (harga murah, harga sedang, harga mahal, dan harga sangat mahal).

Dari dataset yang kami terima fitur-fitur ini terkategori ke dalam hardware. Maka fitur-fitur hardware ini akan dijadikan analisis klasifikasi harga.

# Data Cleansing

Dari dataset yang kami peroleh, kami mendapat informasi bahwa terdapat 21 kolom, 2000 baris, tipe data integer sebanyak 19 dan float sebanyak 2. Dengan 20 kolom digunakan untuk prediksi fitur yang mempengaruhi klasifikasi, dan 1 kolom target prediksi yakni “price\_range”.

Untuk rentang harga dalam kolom “price\_range” terdiri dari tiga kategori dengan kode numerik : [0] untuk harga murah (low cost), [1] untuk harga sedang (medium cost), [2] untuk harga mahal (high cost) dan [3] untuk harga sangat mahal (low very high).

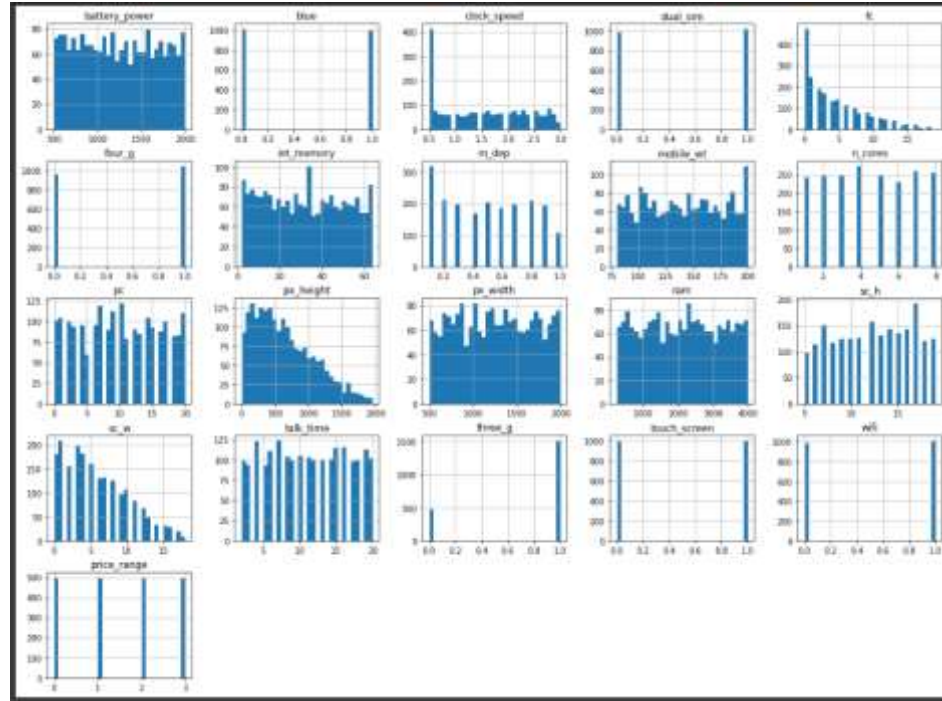
# Data Cleansing

Dari dataset yang kami peroleh kami tidak melakukan pembersihan data atau tindak lanjut data, karena berdasarkan hasil pemeriksaan:

- Tidak terdapat null values
- Data kategori sudah menjadi data numerik
- Tidak terdapat data yang duplikat
- Tidak terdapat outlier pada data
- Tidak terdapat kekeliruan data pada data kategorik (misal kasus biner)

# Exploratory Data Analysis

Cuplikan distribusi data menggunakan histogram.

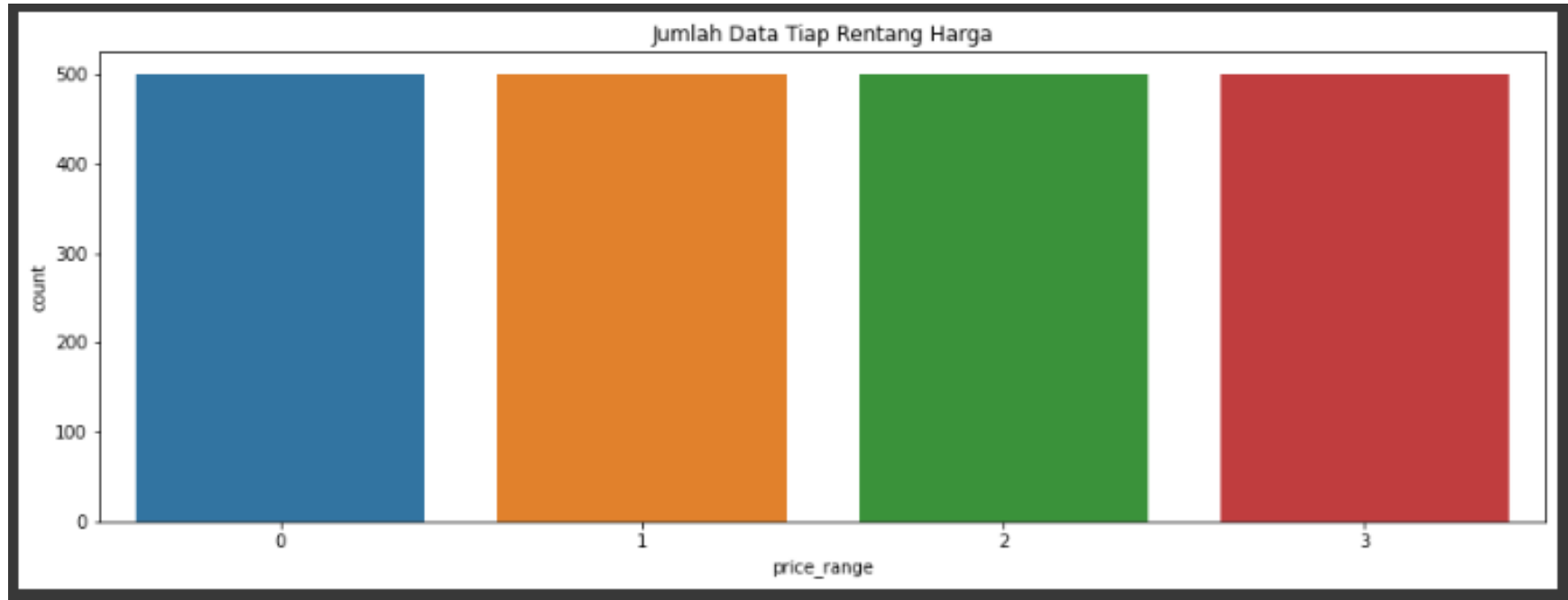


# Exploratory Data Analysis

Dari masing-masing plot visual histogram nampak bahwa tidak terdapat outlier dimana tidak ada bentuk yang berbeda dan berada di luar rentang normal.

Tidak terdapat kekeliruan data pada kolom numerik, misalnya pada kasus three\_g yang tidak terdapat nilai selain 0.0 (tidak memiliki three\_g) dan 1.0 (memiliki three\_g), begitupun untuk kolom four\_ge, touch\_screen, wifi, price\_range, blue, dan dual\_sim.

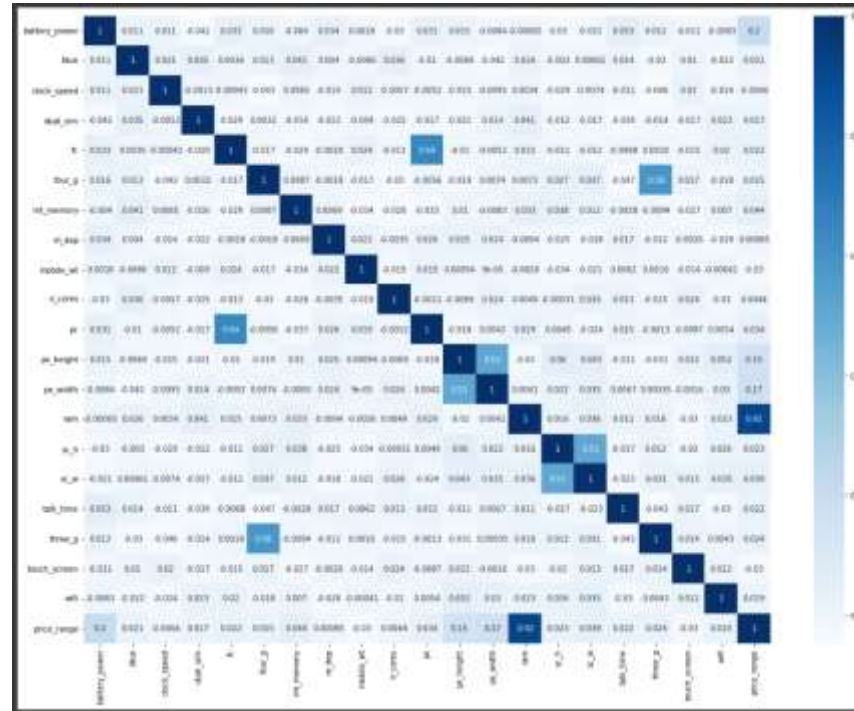
# Exploratory Data Analysis



# Exploratory Data Analysis

Jumlah data pada masing - masing kategori yang terdapat pada kolom price\_range sama banyaknya. Yakni harga dengan kategori [0] untuk harga murah (low cost) terdapat 500 data ,[1] untuk harga sedang (medium cost) terdapat 500 data, [2] untuk harga mahal (high cost) terdapat 500 data dan [3] untuk harga sangat mahal (low very high) terdapat 500 data .

# Exploratory Data Analysis



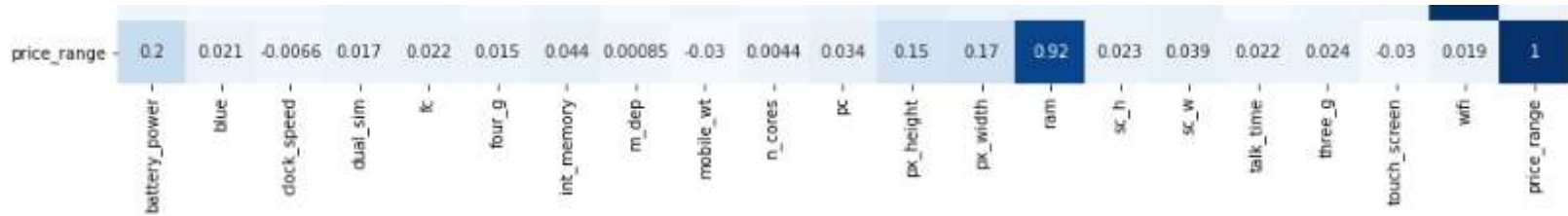


# Exploratory Data Analysis

Hasil:

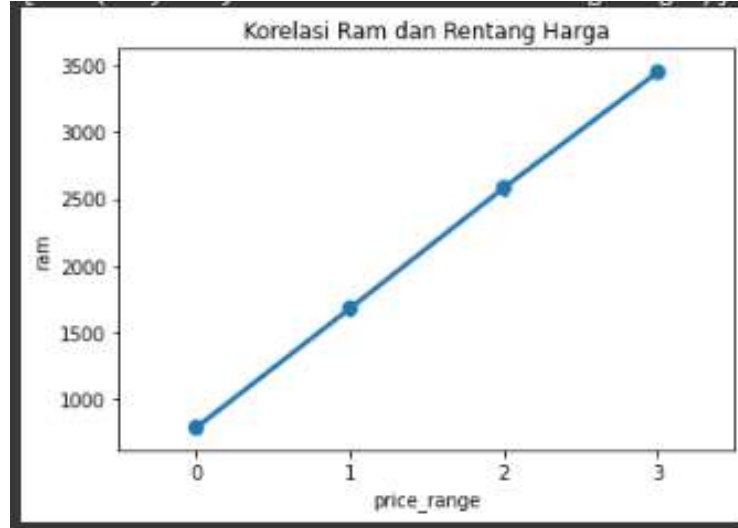
1. Ram memiliki korelasi positif yang tinggi terhadap kolom target yaitu price\_range
2. Kolom three\_g memiliki korelasi terhadap kolom four\_g, karena hampir semua smartphone yang sudah support four\_g pasti sudah support three\_g
3. Kolom pc memiliki korelasi terhadap kolom fc, karena seperti yang kita ketahui bahwa semakin tinggi megapixel yang di miliki front\_camera maka semakin tinggi pula megapixel primary\_camera nya
4. mobile\_wt dan touch\_screen memiliki korelasi negatif terbesar terhadap price\_range
5. Touchscreen tidak mempengaruhi price\_range. Kemungkinan data ini diambil dari data saat teknologi touchscreen belum mengalami trend

# Exploratory Data Analysis



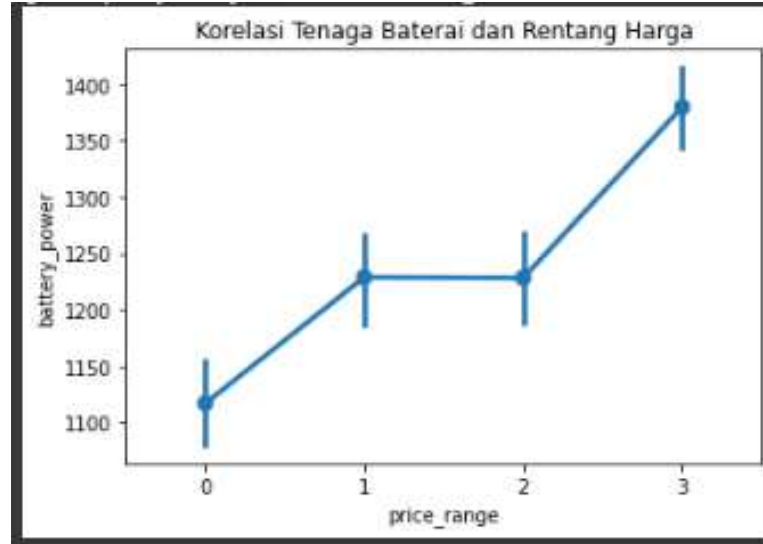
Koefisien korelasi menunjukkan fitur RAM memiliki korelasi besar dengan price\_range, dan fitur mobile\_wt dan touch\_screen tidak memiliki korelasi positif.

# Exploratory Data Analysis



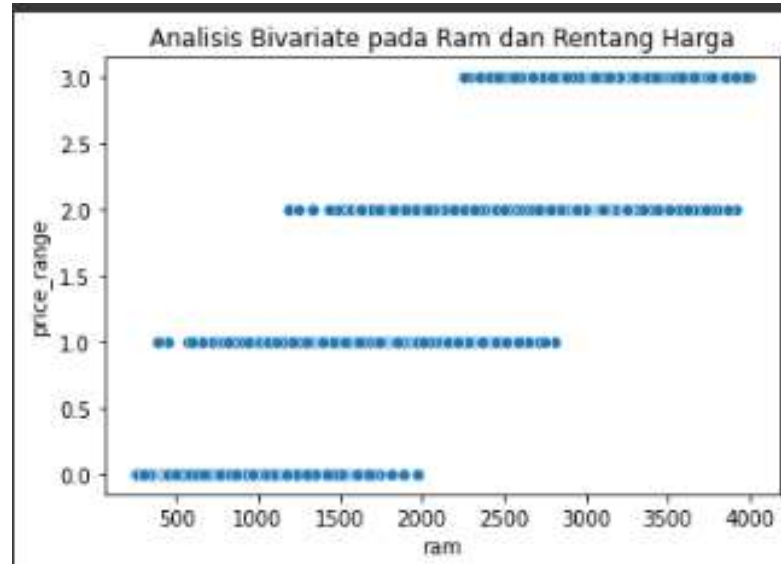
Semakin besar kapasitas RAM, semakin mahal rentang harga, saat RAM berkapasitas 3500 sudah menyentuh rentang harga berkategori 3, begitupun saat RAM berkapasitas 1000 menyentuh rentang harga berkisaran kisaran 0

# Exploratory Data Analysis

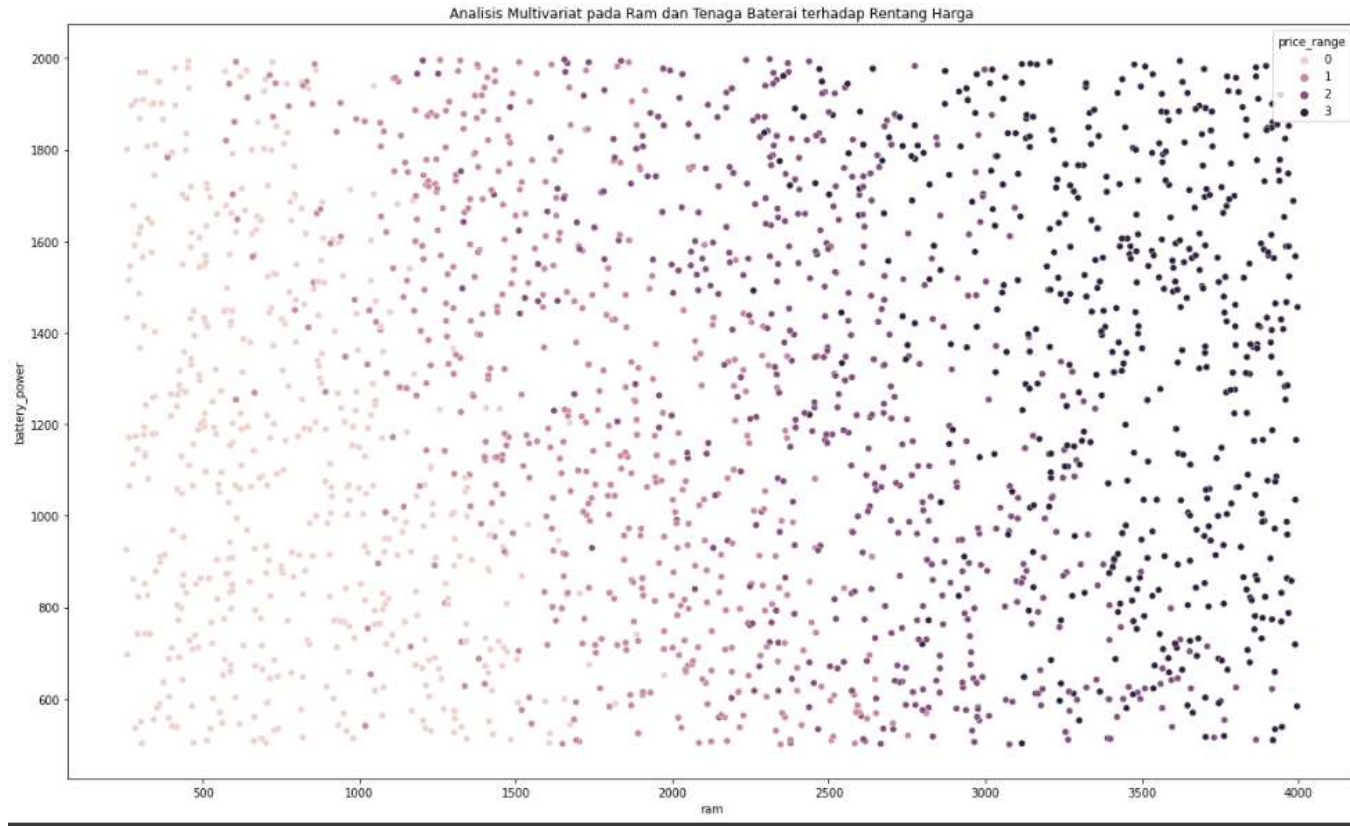


Semakin besar tenaga baterai maka cenderung mempengaruhi rentang harga. Namun pada tenaga baterai antara 1200 ke 1250 memiliki kecenderungan mempengaruhi rentang harga yang sama yakni [1] harga menengah, dan [2] harga mahal, sehingga untuk rentang tenaga baterai itu memiliki rentang harga menengah-mahal.

# Exploratory Data Analysis



# Exploratory Data Analysis

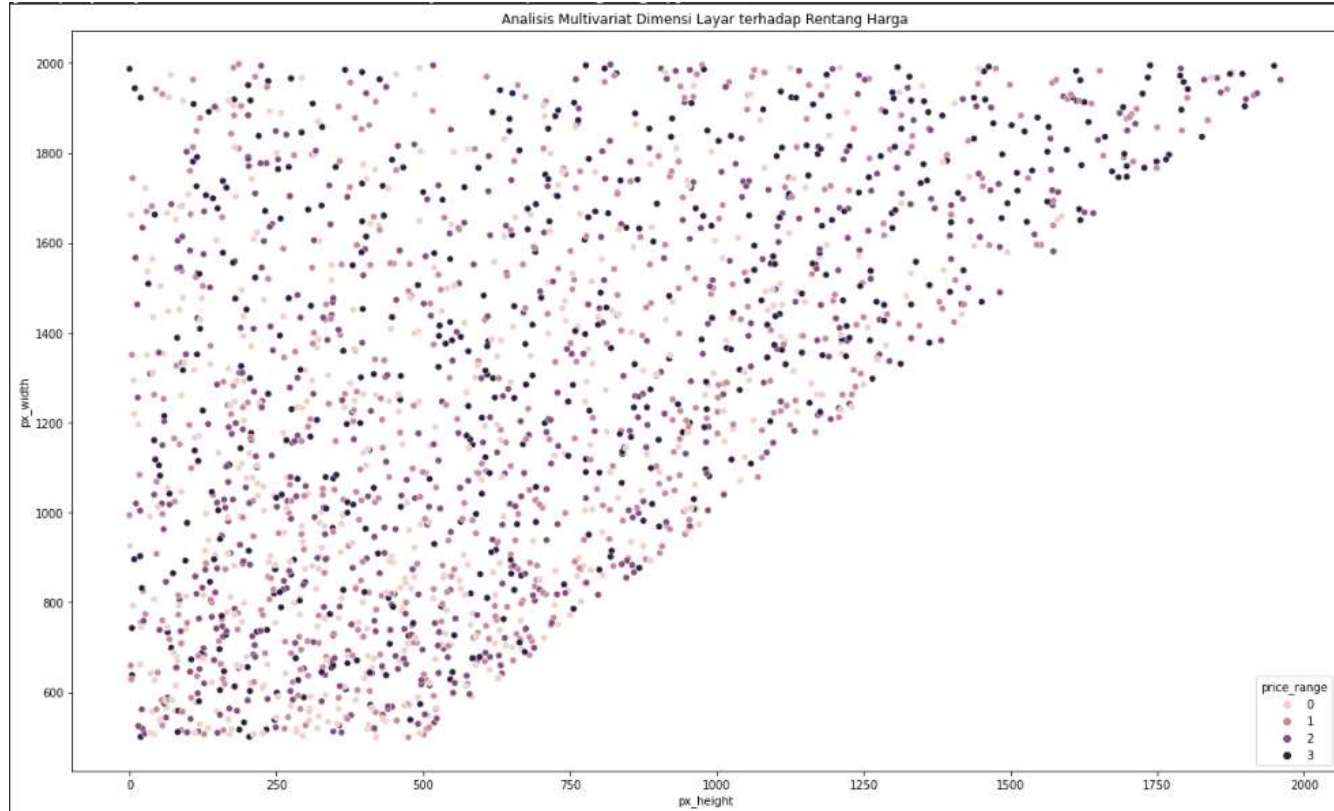


# Exploratory Data Analysis

Dari gambar tersebut nampak dengan dot-dot yang ada semakin besar kapasitas ram dan semakin besar tenaga baterai maka akan mempengaruhi ketebalan dari warna plotnya, artinya harganya semakin mahal.

Misal pada dot yang berada pada ram 3500 dan tenaga baterai 1800 maka dot akan tebal dengan kepekatan yang mengarah pada kategori rentang harga ke 3 yang menandakan harganya semakin mahal.

# Exploratory Data Analysis





# Exploratory Data Analysis

Dari gambar tersebut nampak dengan dot-dot yang tersebar begitu merata, sehingga secara visual nampak bahwa dimensi layar mobile tidak begitu mempengaruhi harga.

Pada gambar nampak seperti titik-titik terpotong, artinya dimensi layar memiliki rasio perbandingan, misal tinggi 700 dan lebar 800 pixel.

# Modelling



# Normalisasi Data

Normalisasi data menggunakan StandarScaler untuk fitur battery\_power, int\_memory, mobile\_wt, px\_height, px\_width, dan RAM.

# Train-Test Split

Train-test split kami menggunakan rasio 7:3 dimana 70% dari 2000 dataset untuk data latih, dan 30%nya untuk data test, yang apabila dijumlahkan 1.400 untuk data latih dan 600 untuk data test.

Pada train-test split ini kami akan menerapkannya kepada dua jenis data yakni data yang dinormalisasi dan data yang tidak dinormalisasi. Perlakuan sama untuk pembagian train-test split

## Dua Jenis Data

Data kami terbagi menjadi dua jenis berdasarkan normalisasinya, yakni data yang dinormalisasi menggunakan standar scaler dan data normalisasi yang tanpa standar scaler (data asal).

Pembagian ini digunakan untuk tiap-tiap model untuk melihat model mana dan jenis data apa yang optimal digunakan untuk klasifikasi.

# Matrix Evaluasi

Kami menggunakan matrix evaluasi yang berupa confusion matrix untuk masalah klasifikasi ini dengan laporan klasifikasi berupa presisi, recall, f1-score, support dan akurasi.

# Model-Model yang digunakan

Kami menggunakan model-model klasifikasi berikut :

- Logistic Regression
- Naive Bayes
- K-Nearest Neighbour (KNN)
- Decision Tree
- Random Forest
- Support Vector Machine (SVM)
- Adaboost Classifier
- Gradient Boosting Classifier
- XGBClassifier

# Upaya Menambah Akurasi Model

## Normalisasi vs Tanpa Normalisasi

Membuat perbandingan akurasi dengan normalisasi (standar scaler) dan tanpa normalisasi. Untuk beberapa model tidak ada perbedaan seperti naive bayes. Ada model yang drastis berubah akurasinya seperti KNN. Ada model yang meningkat akurasinya apabila dinormalisasi.

## KNN: Elbow Method

Dengan mencari jumlah tetangga terdekat. Menunjukkan nilai optimal pada grafik adalah pada jumlah tetangga terdekat sebanyak 8.



# Model Final untuk Prediksi

Berdasarkan model yang sudah diketahui akurasi :

- Model dengan data yang sudah dinormalisasi yakni *Support Vector Machine* (SVM) memiliki tingkat akurasi tertinggi sebesar 95.5%.
- Model dengan data yang sudah dinormalisasi yakni K-Nearest Neighbour memiliki akurasi terendah sebesar 31.5%.
- Model dengan data yang sudah dinormalisasi yakni XGB-Boost berada pada urutan ke 5, dengan akurasi sebesar 84.5%.

# Overfitting dan Underfitting Pada Model Terpilih (SVM)

- Untuk masalah overfitting dapat dilihat dari perbandingan nilai recallnya untuk class 0-3 tidak terlalu jauh, dan tidak ada nilai yang 1.00 , dan rentang akurasi data train sama data test tidak terlalu jauh.
- Untuk masalah underfitting model ini tidak termasuk underfitting kaena akurasi data test nya yang tinggi.

# Model Final Untuk Prediksi

Model final yang digunakan untuk prediksi adalah Support Vector Machine (SVM), karena model ini memiliki akurasi tertinggi pada data test sebesar 94.5%, selain itu nilai recall pada model Support Vector Machine (SVM) itu yang tertinggi dibandingkan dengan model lain dan tidak mengalami overfitting, yang artinya tingkat keberhasilan model dalam memprediksi sebuah data menjadi yang tertinggi.

# Kesimpulan

# Interesting Insights

- Faktor yang paling mempengaruhi price\_range dari sebuah smartphone adalah ram dan battery\_power, karena semakin besar ram dan battery\_power maka price\_range nya juga semakin tinggi, namun fitur yang paling besar berpengaruh terhadap price\_range adalah ram

# Rekomendasi

- Apabila hendak membeli smartphone yang memiliki harga murah, maka pilihlah smartphone yang memiliki kapasitas ram yang sedikit, begitu pula sebaliknya apabila hendak membeli smartphone yang memiliki harga mahal, maka pilihlah smartphone yang memiliki kapasitas ram yang besar.
- Dari model yang kami buat kami merekomendasikan untuk menggunakan SVM dengan normalisasi karena akurasi 94%.

# Semi Deployment

Proses Input Data :

```
Masukkan Tenaga Baterai : 1043
Apakah Mobilennya bluetooth? jawab 1 untuk Ya, dan 0 untuk Tidak : 1
Masukkan Clock Speednya : 3
Apakah Mobilennya mempunyai dual sim? jawab 1 untuk Ya, dan 0 untuk Tidak : 1
Berapa megapixel kamera depannya? : 
```

Semua data diinput :

```
Masukkan Tenaga Baterai : 1043
Apakah Mobilennya bluetooth? jawab 1 untuk Ya, dan 0 untuk Tidak : 1
Masukkan Clock Speednya : 0.5
Apakah Mobilennya mempunyai dual sim? jawab 1 untuk Ya, dan 0 untuk Tidak : 1
Berapa megapixel kamera depannya? : 2
Apakah Mobilennya mempunyai jaringan 4G? jawab 1 untuk Ya, dan 0 untuk Tidak : 1
Berapa memori internalnya? : 7
Berapa kedalaman (mobile deph)? : 0.7
Berapa berat mobilennya? : 145
Berapa Jumlah core processornya? : 3
Berapa megapixel kamera utamanya (kamera belakang)? : 6
Berapa tinggi pixelnya? : 1263
Berapa lebar pixelnya? : 1716
Berapa kapasitas RAM nya? : 2603
Berapa tinggi layar nya? : 11
Berapa lebar layarnya? : 2
Berapa lama waktu bicara (talktime)? : 9
Apakah Mobilennya mempunyai jaringan 3G? jawab 1 untuk Ya, dan 0 untuk Tidak : 1
Apakah Mobilennya sudah touchscreen? jawab 1 untuk Ya, dan 0 untuk Tidak : 1
Apakah Mobilennya mempunyai WiFi? jawab 1 untuk Ya, dan 0 untuk Tidak : 1
```

# Semi Deployment

Hasil prediksi

```
=====Hasil Prediksi=====
Mobile anda diprediksi tergolong ke dalam harga menengah
```

Menunjukkan bahwa dari data yang telah user input tergolong ke dalam harga menengah (kode : 2)



**Terima  
kasih!**  
Ada pertanyaan?

**zenius**



**Kampus  
Merdeka**  
INDONESIA JAYA

