

zenius

Kampus  
Merdeka  
INDONESIA JAYA

# Final Project Presentation

Nomor Kelompok: 5

Nama Mentor: Diajeng Ciptaning Ayu

Nama:

- Arnold Julian
- Kayliev Nabil Auza

Machine Learning Class

Program Studi Independen Bersertifikat  
Zenius Bersama Kampus Merdeka



- 1. Latar Belakang**
- 2. Explorasi 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:

- Memprediksi pengaruh harga smartphone terhadap spesifikasi
- Mencari tahu model algoritma untuk menentukan akurasi terhadap harga smartphone
- Menganalisis Range penjualan dari harga smartphone

# Explorasi Data dan Visualisasi

# Business Understanding

Smartphone sudah menjadi bagian penting dalam hidup, selain sebagai alat komunikasi bisa juga membantu kegiatan sehari-hari seperti mengatur jadwal kegiatan, mencari informasi, atau sebagai media hiburan

Harga smartphone juga dipengaruhi oleh spesifikasinya, semakin bagus spesifikasi smartphone maka akan semakin mahal harga smartphone



# Data Cleansing

```
battery_power    0  
blue             0  
clock_speed     0  
dual_sim        0  
fc              0  
four_g          0  
int_memory      0  
m_dep           0  
mobile_wt       0  
n_cores         0  
pc              0  
px_height       0  
px_width        0  
ram             0  
sc_h            0  
sc_w            0  
talk_time       0  
three_g         0  
touch_screen    0  
wifi            0  
price_range     0  
dtype: int64
```

Data tersebut tidak perlu dibersihkan karena dari awal data sudah bersih, sehingga tidak ada baris atau kolom yang missing/null

# Exploratory Data Analysis

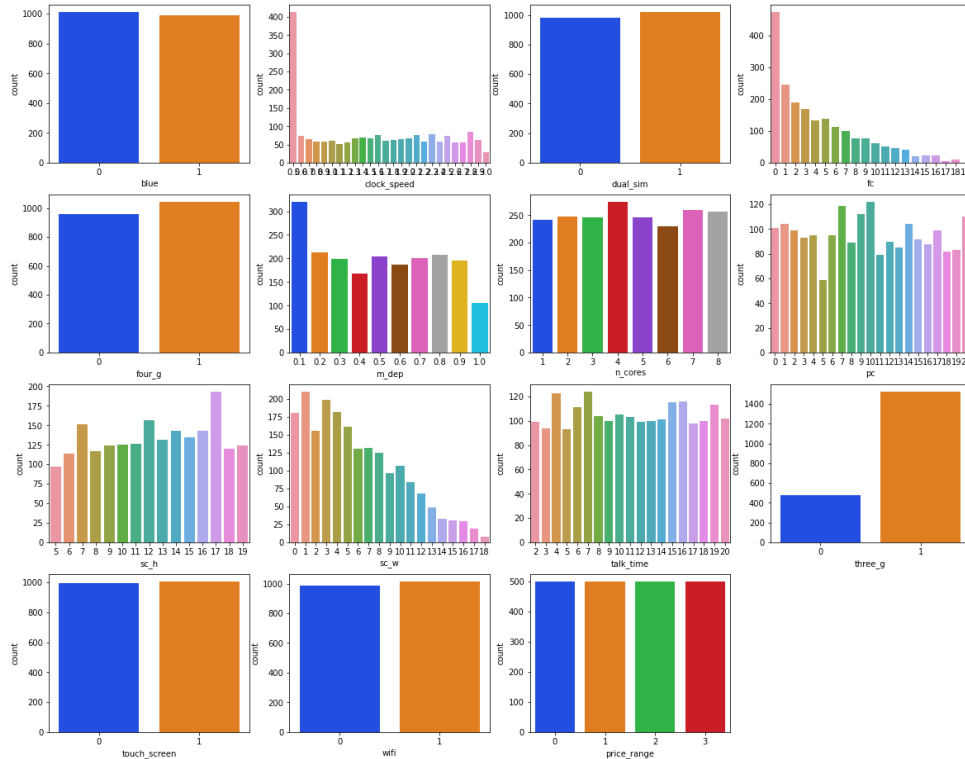
	battery_power	blue	clock_speed	dual_sim	fc	four_g	int_memory	m_dep	mobile_wt	n_cores	...	px_height	px_width	ram	sc_h	sc_w	talk_time	three_g	touch_screen	wifi	price_range
0	842	0	2.2	0	1	0	7	0.6	188	2	...	20	756	2549	9	7	19	0	0	1	1
1	1021	1	0.5	1	0	1	53	0.7	136	3	...	905	1988	2631	17	3	7	1	1	0	2
2	563	1	0.5	1	2	1	41	0.9	145	5	...	1263	1716	2603	11	2	9	1	1	0	2
3	615	1	2.5	0	0	0	10	0.8	131	6	...	1216	1786	2769	16	8	11	1	0	0	2
4	1821	1	1.2	0	13	1	44	0.6	141	2	...	1208	1212	1411	8	2	15	1	1	0	1

5 rows x 21 columns

```
Index(['battery_power', 'blue', 'clock_speed', 'dual_sim', 'fc', 'four_g',  
      'int_memory', 'm_dep', 'mobile_wt', 'n_cores', 'pc', 'px_height',  
      'px_width', 'ram', 'sc_h', 'sc_w', 'talk_time', 'three_g',  
      'touch_screen', 'wifi', 'price_range'],  
      dtype='object')  
(2000, 21)
```



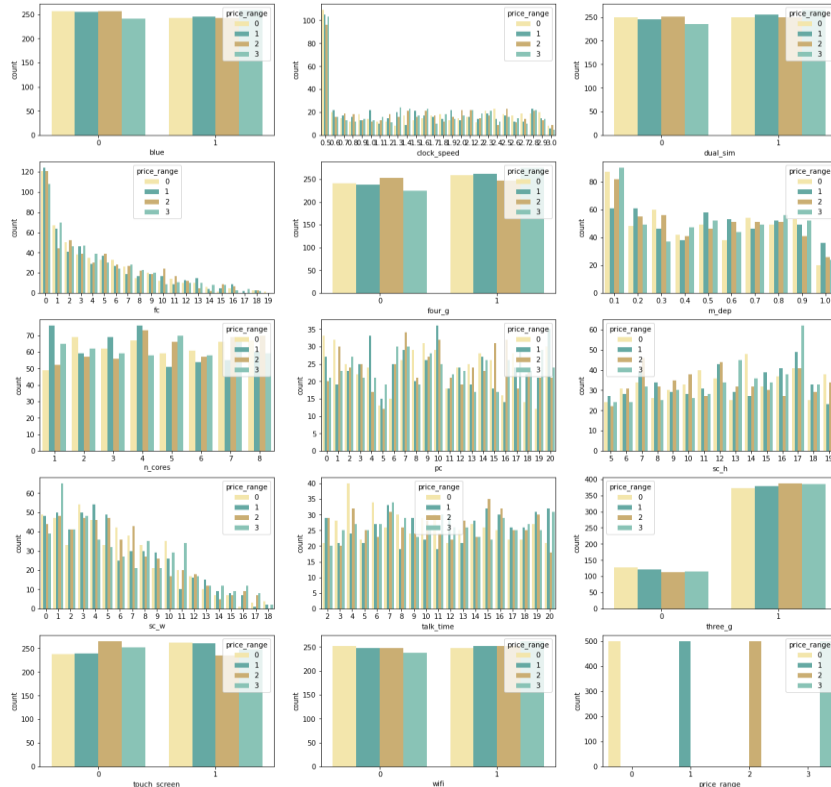
# Exploratory Data Analysis



## Categorical Features

```
[ 'blue',
  'clock_speed',
  'dual_sim',
  'fc',
  'four_g',
  'm_dep',
  'n_cores',
  'pc',
  'sc_h',
  'sc_w',
  'talk_time',
  'three_g',
  'touch_screen',
  'wifi',
  'price_range']
```

# Exploratory Data Analysis

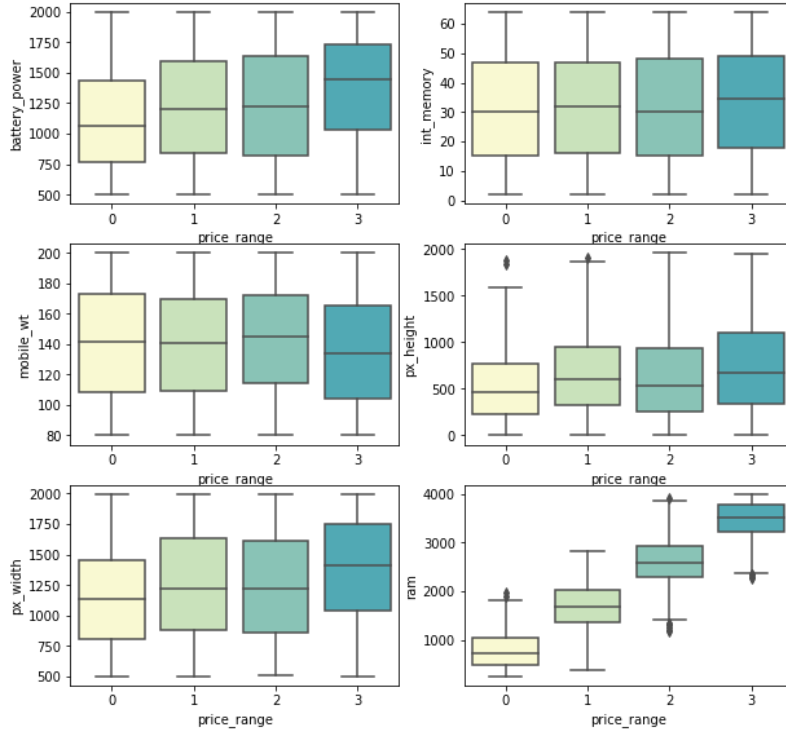


## Categorical Features

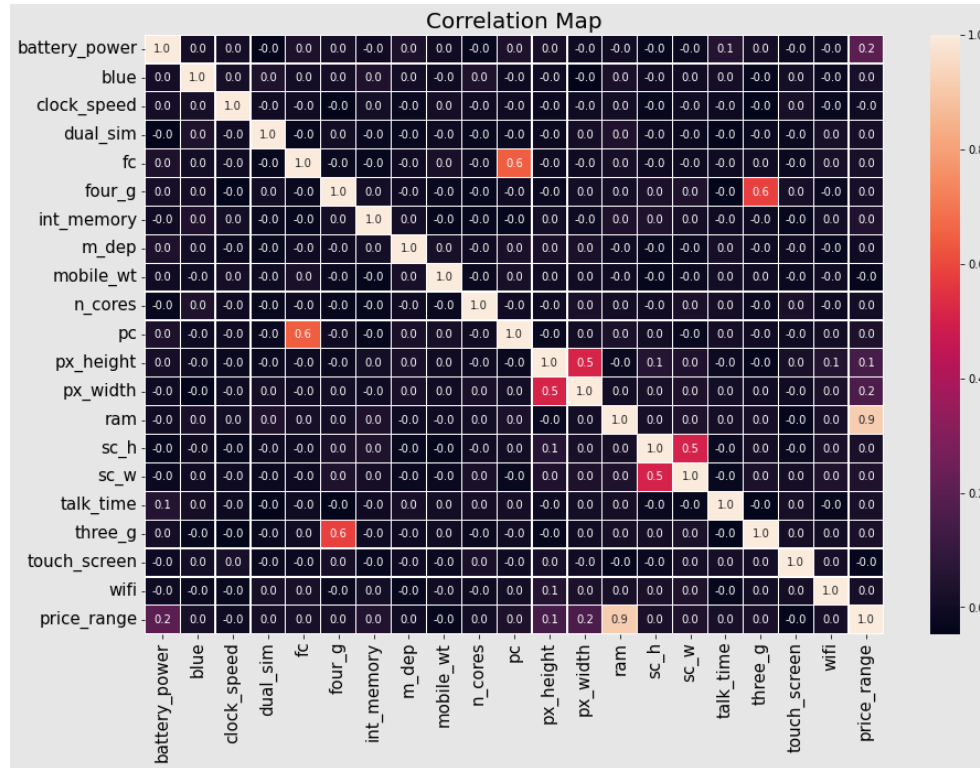
```
[ 'blue',
  'clock_speed',
  'dual_sim',
  'fc',
  'four_g',
  'm_dep',
  'n_cores',
  'pc',
  'sc_h',
  'sc_w',
  'talk_time',
  'three_g',
  'touch_screen',
  'wifi',
  'price_range']
```

# Exploratory Data Analysis

## Numerical Features using Boxplot



# Exploratory Data Analysis



## Heatmap

# Modelling



# Preparation

- Membuat Variable x dan y dimana x = semua kolom kecuali price range dan y = price range
- Membagi variabel x dan y menjadi 4 ( x train, x test, y train, y test)
- Mengecek variabel dari x dan y

# Logistic Regression Model

- Membuat variabel  $lr$  sebagai logistic regression
- Melakukan fit untuk variabel  $x$  train dan  $y$  train
- Melakukan tes pada model dan didapatkan hasil 0.6283

# Decision Tree Model

- Membuat variabel dtree untuk decision tree classifier
- Melakukan fit untuk variabel x train dan y train
- Melakukan tes pada model dan didapatkan hasil 0.8433



# Random Forest Model

- Membuat variabel rfc sebagai Random Forest Classifier
- Melakukan fit untuk variabel x train dan y train
- Melakukan tes pada model dan didapatkan hasil 0.8716

# Evaluation

- Accuracy Score

```
Akurasi untuk Logistic Regression: 0.63  
Akurasi untuk Decision Tree 0.84  
Akurasi untuk Random Forest 0.87
```

- Precision Score

```
Precision score untuk Logistic Regression: 0.63  
Precision score untuk Decision Tree: 0.85  
Precision score untuk Random Forest: 0.87
```

- Recall Score

```
Recall score untuk Logistic Regression: 0.64  
Recall score untuk Decision Tree: 0.85  
Recall score untuk Random Forest: 0.87
```

Dari ketiga model dapat disimpulkan bahwa akurasi untuk random forest paling tinggi yaitu 0.87

# Conclusion

# Kesimpulan

Dari Analisis tersebut, dapat disimpulkan bahwa ketika kita ingin membeli suatu smartphone, hal yang perlu kita lihat yaitu ram nya

Tapi kita juga perlu melihat aspek dari baterai dan juga pixel kamera dari visualisasi heatmap masih ada korelasi dengan price range

**Terima  
kasih!**  
**Ada pertanyaan?**

**zenius**



**Kampus  
Merdeka**  
INDONESIA JAYA

