

zenius

Kampus
Merdeka
INDONESIA JAYA

Final Project Presentation

Nomor Kelompok: 2

Nama Mentor: Rauzan Sumara

Nama:

- Randi Julian Saputra
- Wahyu Berkhe Bestari
- David

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:

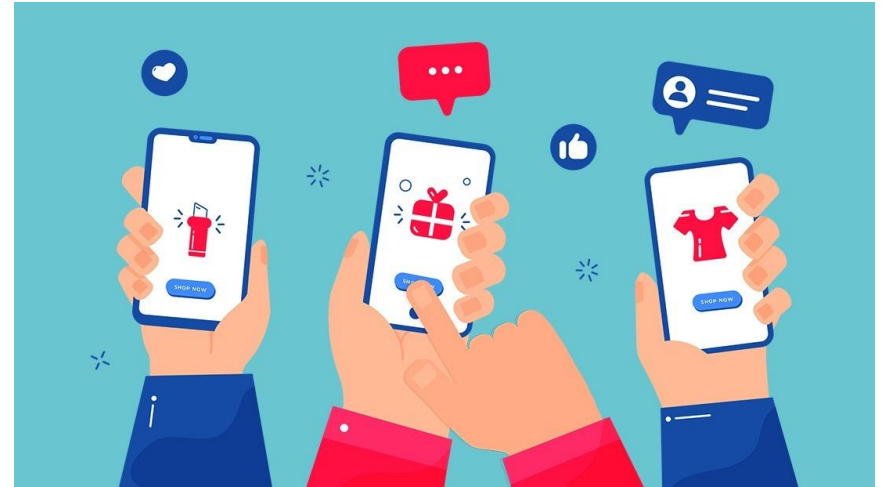
- Memprediksi pengaruh spesifikasi terhadap harga smartphone.
- Mencari tahu model algoritma untuk menentukan akurasi terhadap harga smartphone.
- Menganalisis range penjualan dari harga smartphone.

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 × 21 columns

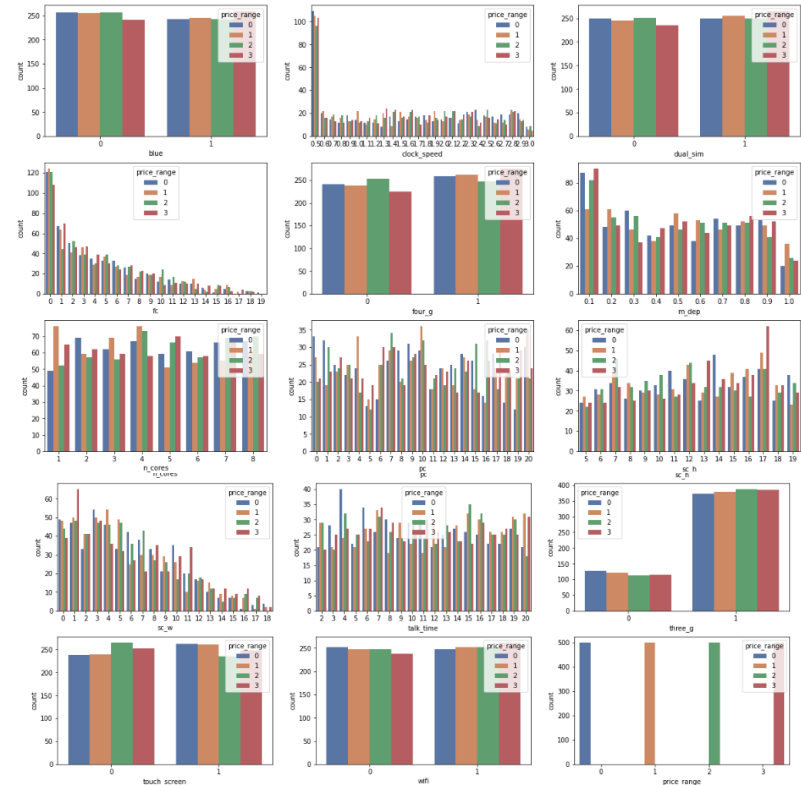
```
Index(['battery_power', 'blue', 'clock_speed', 'dual_sim', 'fc', 'four_g',  
      'int_memory', 'm_dep', 'mobile_wt', 'n_cores', 'pc', 'px_height',  
      'px_width', 'ram', 'sc_h', 'sc_w', 'talk_time', 'three_g',  
      'touch_screen', 'wifi', 'price_range'],  
      dtype='object')
```

(2000, 21)

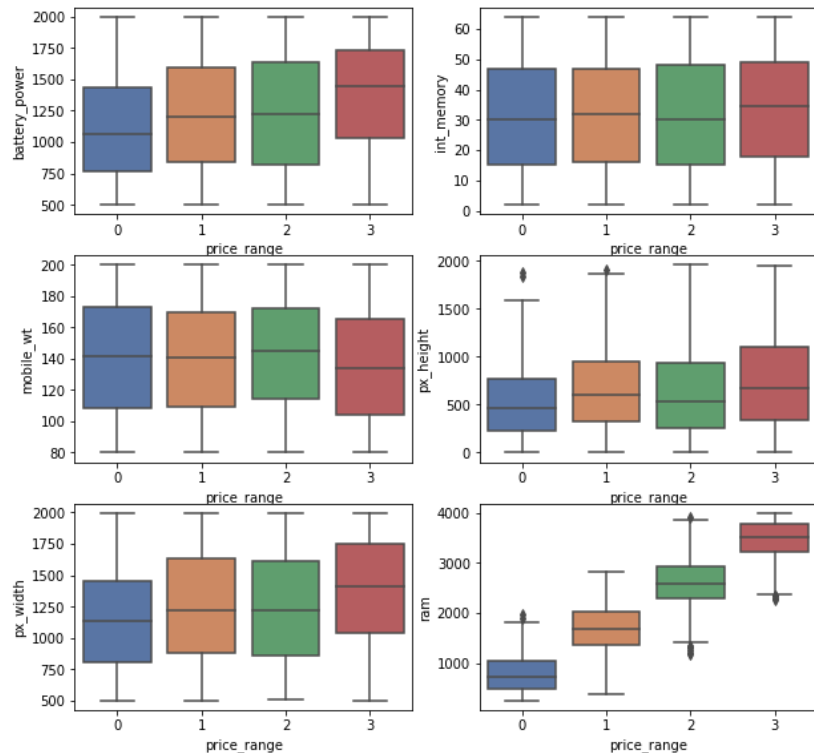
Exploratory Data Analysis

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' ]
```



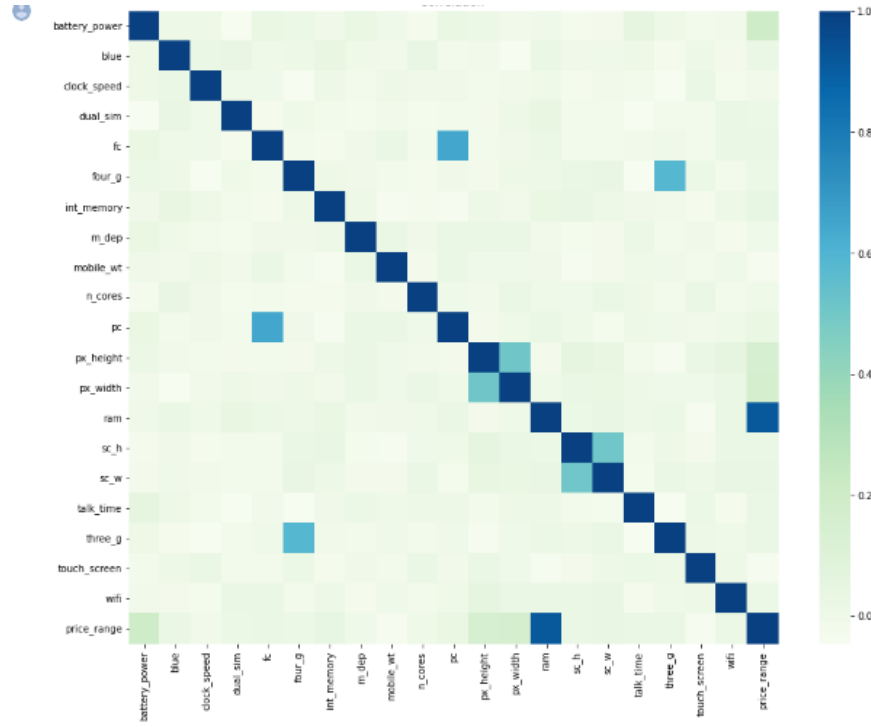
Exploratory Data Analysis



Numerical Features

```
['battery_power', 'int_memory', 'mobile_wt', 'px_height', 'px_width', 'ram']
```

Exploratory Data Analysis



Heatmap

price_range memiliki korelasi terhadap ram, px_width, px_height, dan battery power.

Modelling



Splitting Train and Test

- Membagi menjadi 4 variable yaitu : X_train, X_test, y_train, y_test.
- x merupakan variable yang berisikan data kolom selain price range, sementara y berisikan kolom price range.
- Mengambil data train sebesar 80% dan data test sebesar 20%

Naïve Bayes Classifier

- Melakukan import sklearn untuk library Gaussian Naïve Bayes
- Membuat variable GNB sebagai Gaussian Naïve Bayes
- Dilakukan fit dengan model GNB pada x_train dan y_train dan didapatkan test accuracy sebesar 0.835

Random Forest Classifier

- Melakukan import sklearn untuk library Random Forest Classifier
- Membuat variable rfc sebagai Random Forest Classifier
- Dilakukan fit dengan model rfc pada x_train dan y_train dan didapatkan test accuracy sebesar 0.8675

Decision Tree Classifier

- Melakukan import sklearn untuk library Decision Tree Classifier
- Membuat variable dtree sebagai Decision Tree Classifier
- Dilakukan fit dengan model dtree pada x_train dan y_train dan didapatkan test accuracy sebesar 0.8275

Evaluation based on model

```
Test accuracy with Naive Bayes : 0.835
Test accuracy with Random Forest : 0.8675
Test accuracy with Decision Tree : 0.8275
```

Evaluation based on model

- Decision Tree

	precision	recall	f1-score
0	0.90	0.93	0.91
1	0.79	0.74	0.76
2	0.73	0.75	0.74
3	0.89	0.89	0.89

Evaluation based on model

- Random Forest

	precision	recall	f1-score
0	0.94	0.98	0.96
1	0.81	0.79	0.80
2	0.79	0.75	0.77
3	0.91	0.94	0.93

Evaluation based on model

- Naïve Bayes

	precision	recall	f1-score	s
0	0.96	0.94	0.95	
1	0.76	0.75	0.75	
2	0.72	0.71	0.71	
3	0.89	0.93	0.91	

Conclusion

Dari analisis yang sudah dilakukan dapat diambil kesimpulan bahwa ram, px_width, px_height, dan battery power dapat mempengaruhi price range dari smartphone terutama ram yang memiliki perbedaan value terbanyak.

**Terima
kasih!**
Ada pertanyaan?

zenius



**Kampus
Merdeka**
INDONESIA JAYA

