

zenius

Kampus
Merdeka
INDONESIA JAYA

Final Project Presentation

Nomor Kelompok: Kelompok 6

Nama Mentor: Rauzan Sumara

Nama:

- <Feri Irawansyah>
- <Achmad Raihan>

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: [Mobile Price Classification | Kaggle](#)

Problem: **Classification** Tujuan:

- Mengetahui faktor yang memengaruhi kualitas *smartphone*.
- Menentukan *Machine Learning Model* terbaik dengan akurasi tinggi yang digunakan untuk Prediksi kualitas *smartphone*.

Explorasi Data dan Visualisasi

Business Understanding

- *Smartphone* merupakan salah satu perangkat elektronik dengan fungsi utama sebagai alat komunikasi.
- Perbedaan harga yang dirilis tiap perusahaan disebabkan karena perbedaan tujuan khusus dari spesifikasi *smartphone*.
- Berdasarkan spesifikasinya, *smartphone* juga dibuat dengan tujuan khusus, seperti untuk bermain *games*, fotografi, dll.

Data Cleansing

Pada kolom ***px_height*** dengan nilai minimum 0, index ke 1481 terdapat kejanggalan sebagai berikut:

- Terdapat *smartphone* berada di peringkat ke 3/kualitas terbaik
- Tidak memiliki *bluetooth*
- Dan memiliki RAM yang cukup besar
- Memiliki *battery power* yang besar
- Nilai pada kolom ***sc_width*** = 0

Tindakan terbaik yang dilakukan yaitu menghapus baris tersebut karena ambigunya informasi tersebut.

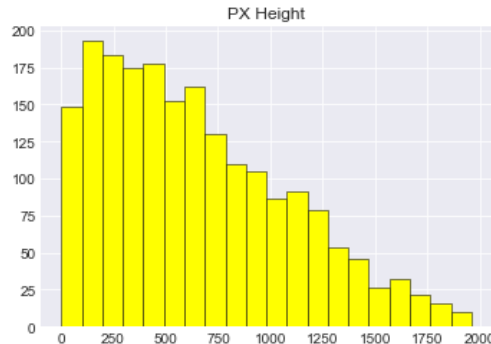
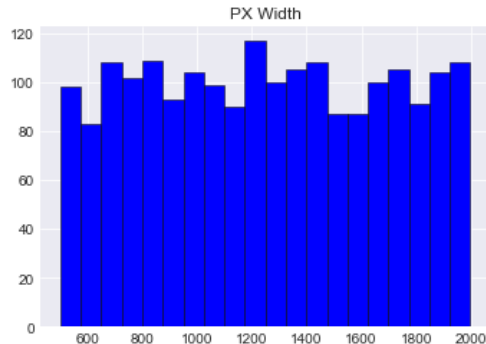
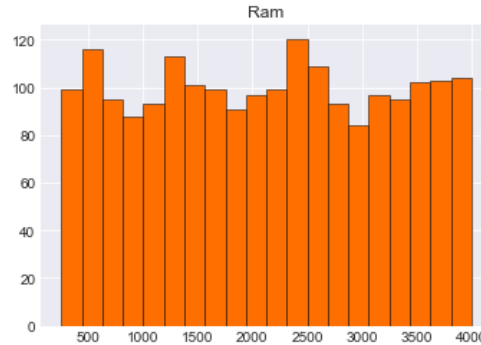
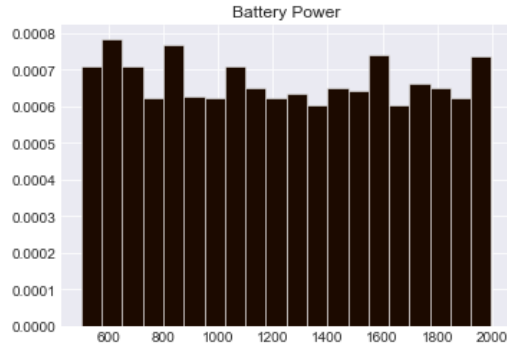
Exploratory Data Analysis

Terdapat 4 fitur yang memiliki korelasi diatas 10%, yaitu:

Fitur	Persentase
RAM	91%
Battery Power	20%
Pixel Width	16%
Pixel Height	14%

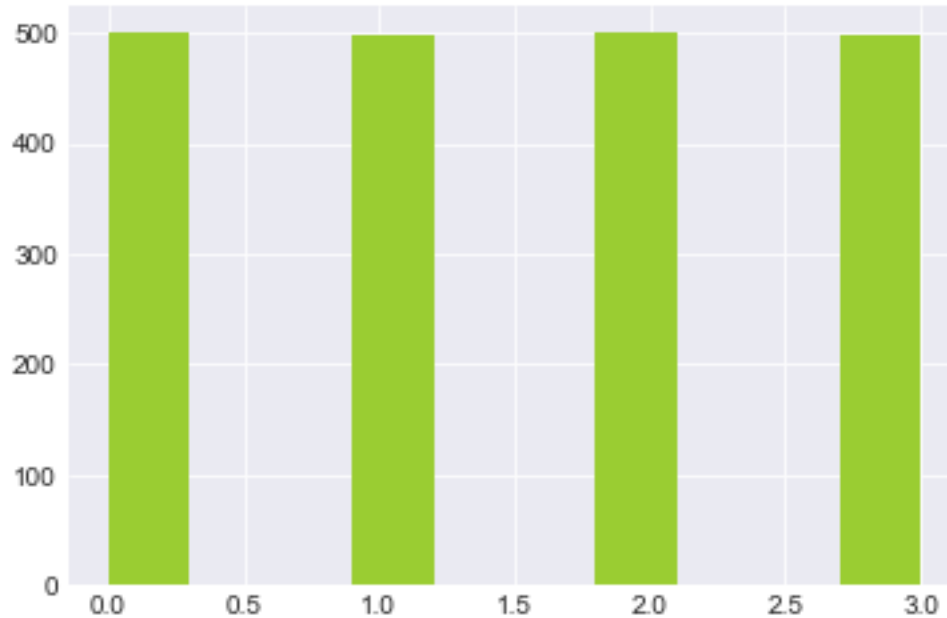
Berdasarkan pendapat dari Jonathan Sarwono Direktur Penjaminan Mutu di *International Women University* dalam artikel <http://www.databee.id/2020/12/jenis-uji-korelasi.html>, maka fitur tersebut akan dijadikan sebagai prediktor untuk prediksi kualitas *smartphone*.

Exploratory Data Analysis



Visualisasi untuk melihat adanya *outlier* berdasarkan distribusi dari 4 fitur tersebut.

Exploratory Data Analysis



Visualisasi untuk melihat keseimbangan jumlah *target variable* pada kolom *Price Range* pada tiap tingkat kualitasnya.

Modelling



Machine Learning Model

<pada bagian ini, silakan Anda menyusun sistematika slide sesuai dengan keinginan kelompok Anda>

hal-hal yang disarankan untuk dipresentasikan:

- Metode train test split / cross validation yang digunakan
- Metrik untuk melakukan evaluasi
- Jenis model awal yang dicoba
- Jenis model lain yang turut dicoba, serta tindakan-tindakan apa saja yang dilakukan untuk mencoba menambah akurasi model (hyperparameter tuning, dst)
- Model final
- Kolom-kolom apa saja yang menjadi prediktor dan target variable untuk model

final

Machine Learning Model

Akurasi *Machine Learning Model* yang digunakan untuk prediksi kualitas *smartphone* tanpa melakukan *Exploratory Data Analysis*

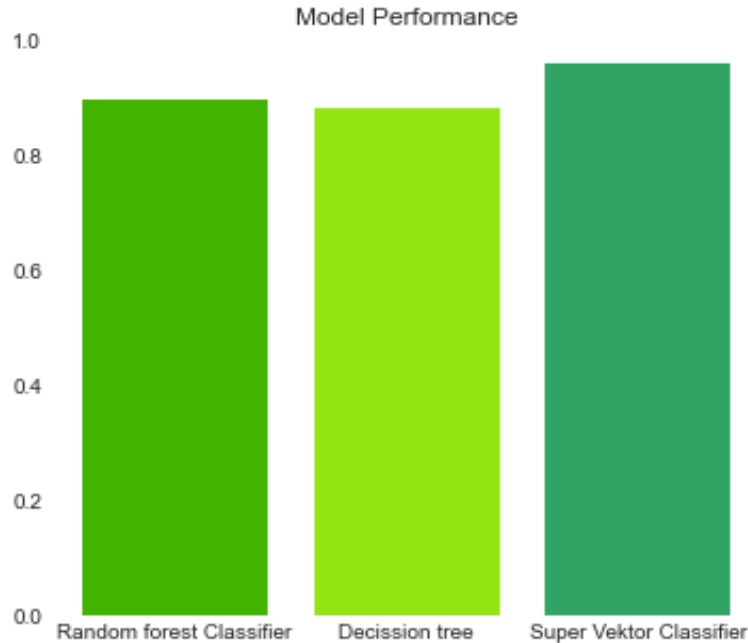
Machine Learning Model	Persentase Akurasi
Random Forest: Classifier	86%
Decision Tree	84%
Support Vector Machine: Classifier	87%

Machine Learning Model

Akurasi *Machine Learning Model* yang digunakan untuk prediksi kualitas *smartphone* dengan melakukan *Exploratory Data Analysis* dan *Hyperparameter Tuning*.

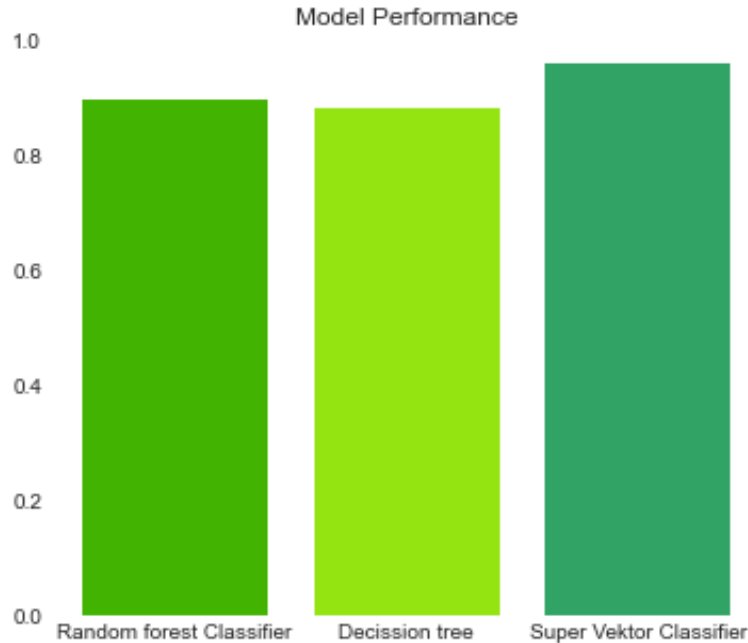
Machine Learning Model	Persentase Akurasi
Random Forest: Classifier	90%
Decision Tree	88%
Support Vector Machine: Classifier	97%

Machine Learning Model



Visualisasi akurasi *Machine Learning Model* dengan melakukan *Exploratory Data Analysis* dan *Hyperparameter Tuning*.

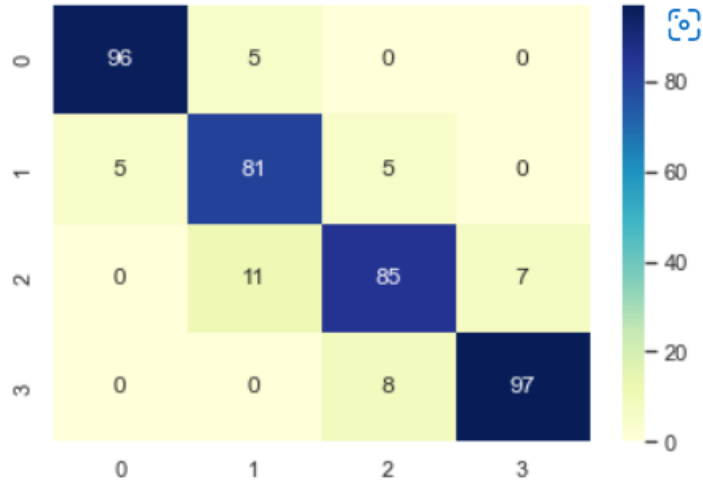
Machine Learning Model



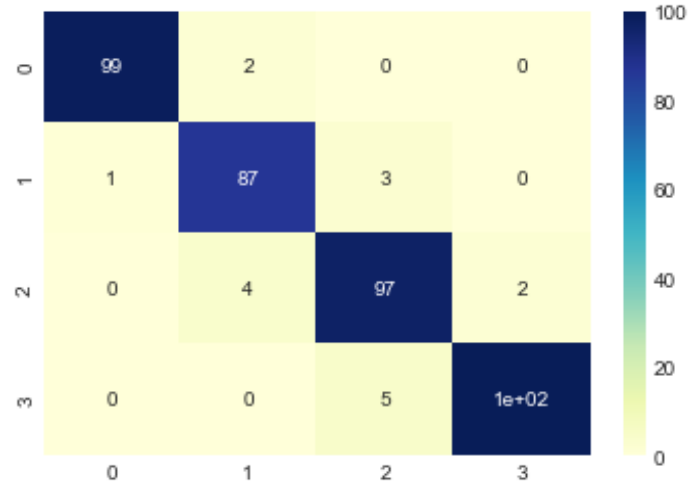
Visualisasi akurasi *Machine Learning Model* dengan melakukan *Exploratory Data Analysis* dan *Hyperparameter Tuning*.

Evaluation

Heatmap Confusion Matrix pada Random Forest: Classifier



Heatmap Confusion Matrix pada Support Vector Machine: Classifier



Evaluation

Pada *Heatmap Confusion Matrix Support Vector Machine: Classifier*, nilai prediksi melebihi kapasitas dari *data test* yang digunakan, dimana seharusnya berjumlah 400 smartphone, namun jika dilakukan perhitungan manual, hasilnya yaitu 402 dengan perhitungan berikut $99+1+2+87+4+3+97+5+2+102 = 402$. Sedangkan, *Heatmap Confusion Matrix pada Random Forest: Classifier* berada dalam kondisi normal.

Conclusion

<pada bagian ini, silakan paparkan kesimpulan Anda, apa saja insights/trend yang menarik, dan sertakan saran Anda kepada stakeholder>

<misal, proyek Anda adalah tentang properti, maka Anda bisa memberikan saran kepada calon pembeli properti, kira-kira properti yang seperti apa yang paling worth it>

<misal, proyek Anda adalah tentang churn, maka Anda bisa memberikan saran kepada perusahaan bagaimana untuk menurunkan churn, faktor-faktor apa yang harus diperhatikan, dst>

Berdasarkan hasil analisis yang telah dilakukan, terdapat 4 faktor yang harus diperhatikan kepada calon pembeli *smartphone*. Hal tersebut disebabkan karena akan memengaruhi kualitas *smartphone*, yaitu:

- RAM
- Battery Power
- Pixel Width
- Pixel Height

Berdasarkan evaluasi yang telah dilakukan, maka *Machine Learning Model* terbaik yang digunakan untuk prediksi kualitas *smartphone* yaitu *Random Forest: Classifier* dengan akurasi 90%.

**Terima
kasih!**
Ada pertanyaan?

zenius



**Kampus
Merdeka**
INDONESIA JAYA

