

# Machine Learning Evaluation & Supervised Learning

Ensemble Methods



# Rezki Trianto

Data guy who spent the last 7 years to work around data analytics and machine learning. Working in a unicorn company with various domain expertise to work with, from an e-commerce, OTA, and ride-hailing company.



Rezki Trianto

[linkedin.com/in/rezkitrianto](https://www.linkedin.com/in/rezkitrianto)

## Education Background



**2010-2014**  
Bachelor Degree  
Computer Science



**2015-2017**  
Master Degree  
Computer Science

### Rezki Trianto

5+ years experience in Data  
Science & Analytics

# Hands On Required

**Hands - On : 4. Ensemble Method**

Klik disini untuk  
mengakses folder Hands  
On

# Topik Supervised Learning: Ensemble Methods

- ☐ Introduction: Ensemble Method
- ☐ Bagging
- ☐ Hands On - Bagging
- ☐ Boosting
- ☐ Hands On - Boosting
- ☐ Stacking
- ☐ Hands On - Stacking
- ☐ Explainable AI
- ☐ Hands On - Shap Values

# Topik Supervised Learning: Ensemble Methods



Introduction: Ensemble Method



Bagging



Hands On - Bagging



Boosting



Hands On - Boosting



Stacking



Hands On - Stacking



Explainable AI



Hands On - Shap Values



# Apa itu Ensemble Methods

Ensemble learning membantu meningkatkan performa model machine learning dengan melakukan kombinasi dari beberapa model menjadi 1 predictive model saja dengan tujuan:

- Mengurangi bias (mencegah underfitting) - boosting
- Mengurangi variance(mencegah overfitting) - (bagging)
- Mengurangi bias/variance dengan kombinasi algoritma yang berbeda (stacking)

# Advantages & Disadvantages

## Keuntungan:

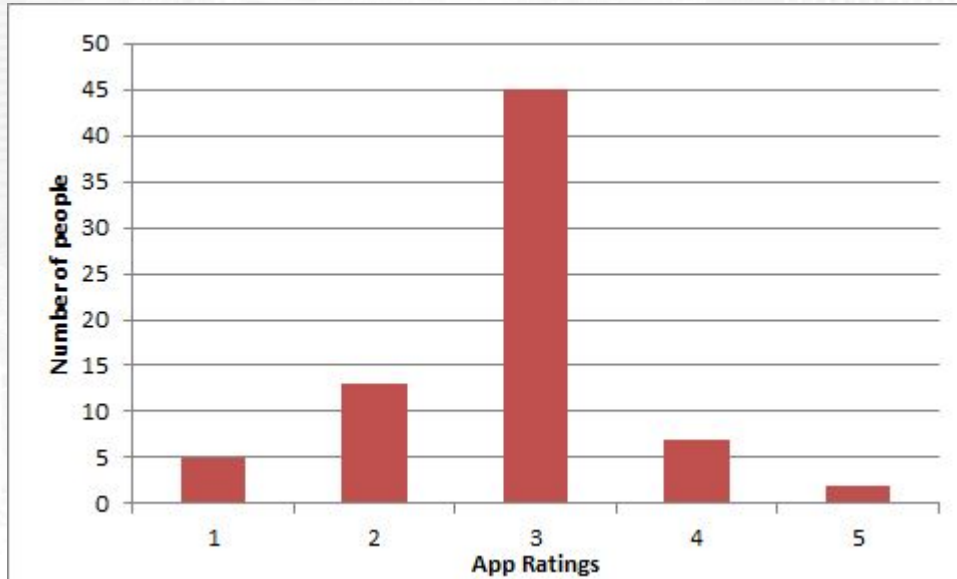
- Prediksi model lebih akurat
- Model lebih stabil dan robust

## Kerugian:

- Semakin sulit untuk interpretasi model untuk ditarik business insightnya
- Proses komputasi yang cukup tinggi
- Pemilihan model yang tepat untuk dilakukan ensemble cukup sulit

# Simple Ensemble Method

Terdapat chart dari hasil review di apps store sebagai berikut  
Sebagai seorang DS, kita mau summary hasil reviewnya dengan cara:



1. Majority Voting
  - a. Major Rating = 3
2. Average
  - a. Avg. Rating = 2.8
3. Weighted Average



# Simple Ensemble Method

Contoh pada algoritma:

	Logistic Regression	kNN	Decision Tree
Label	0	1	1
Probabilitas	0.4	0.5	0.6
Majority Voting	1 (Karena mayoritas prediksi label 1)		
Average	$(0.4 + 0.5 + 0.6)/3 = 0.5 \rightarrow$ label 1 (karena melebihi threshold $\geq 0.5$ )		
Weighted Average	$(0.4 + 2*0.5 + 2*0.6)/5 = 0.52$ $\rightarrow$ label 1 (karena melebihi threshold $\geq 0.5$ )		

# Topik Supervised Learning: Ensemble Methods



Introduction: Ensemble Method



Bagging



Hands On - Bagging



Boosting



Hands On - Boosting



Stacking



Hands On - Stacking



Explainable AI



Hands On - Shap Values

# Topik Supervised Learning: Ensemble Methods



Introduction: Ensemble Method



Bagging



Hands On - Bagging



Boosting



Hands On - Boosting



Stacking



Hands On - Stacking



Explainable AI



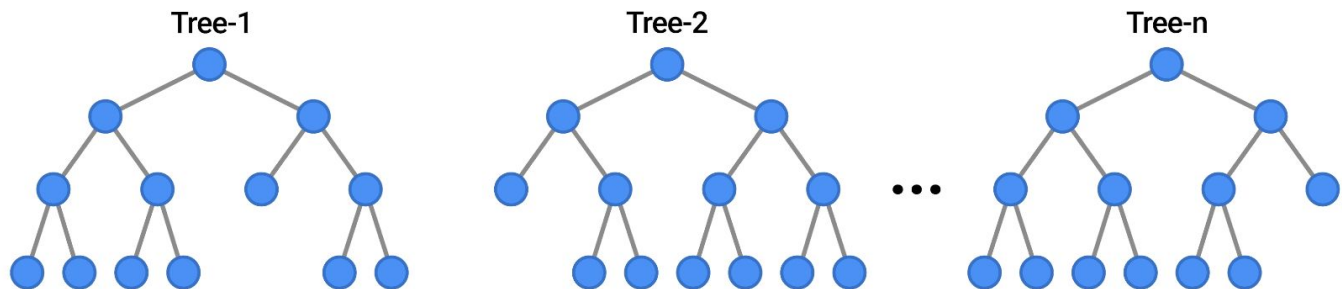
Hands On - Shap Values

# Bagging

Bootstrap AGGREGatING

# Contoh Algoritma: Random Forest

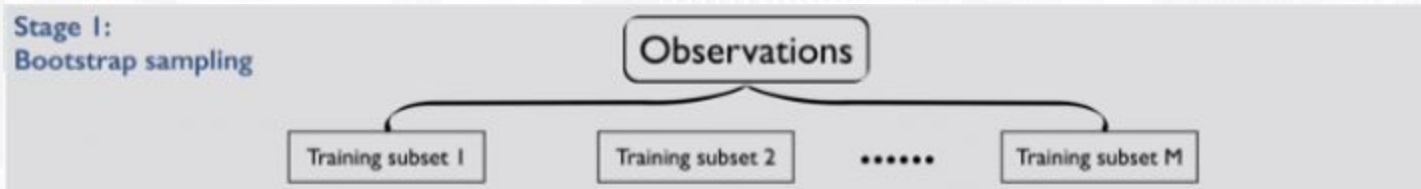
- Random Forest terdiri dari beberapa subset decision tree yang memiliki training set berbeda pada masing-masing tree (yang di generate secara random)





# Step by Step Bagging

1. Langkah awalnya adalah dengan melakukan bootstrap dan membuat random sample dari training set.

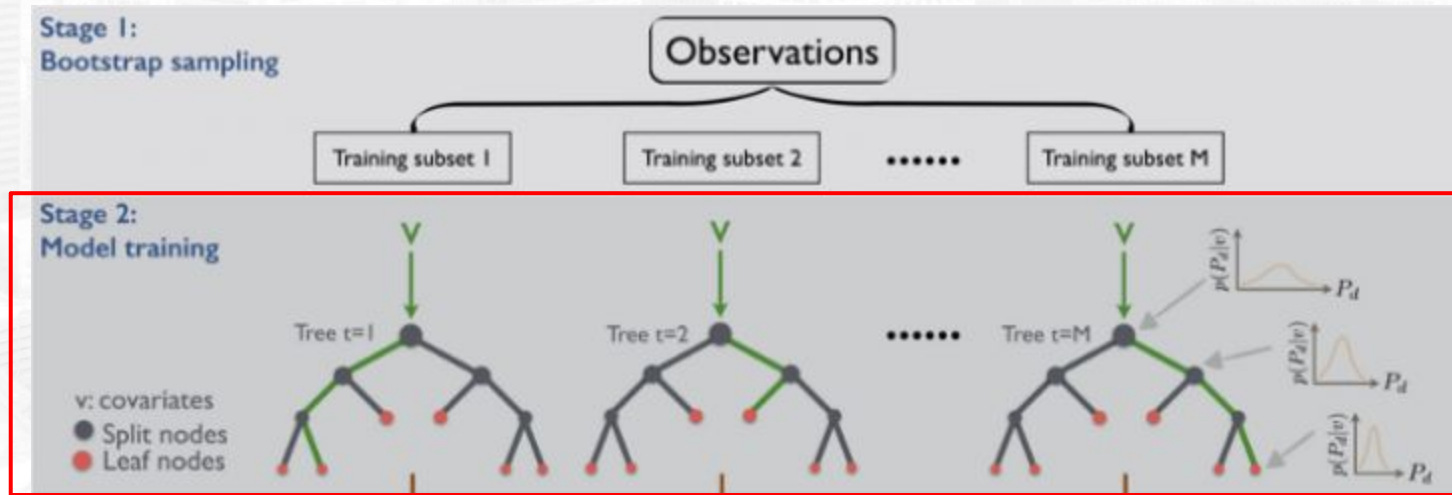


Contoh bootstrap:



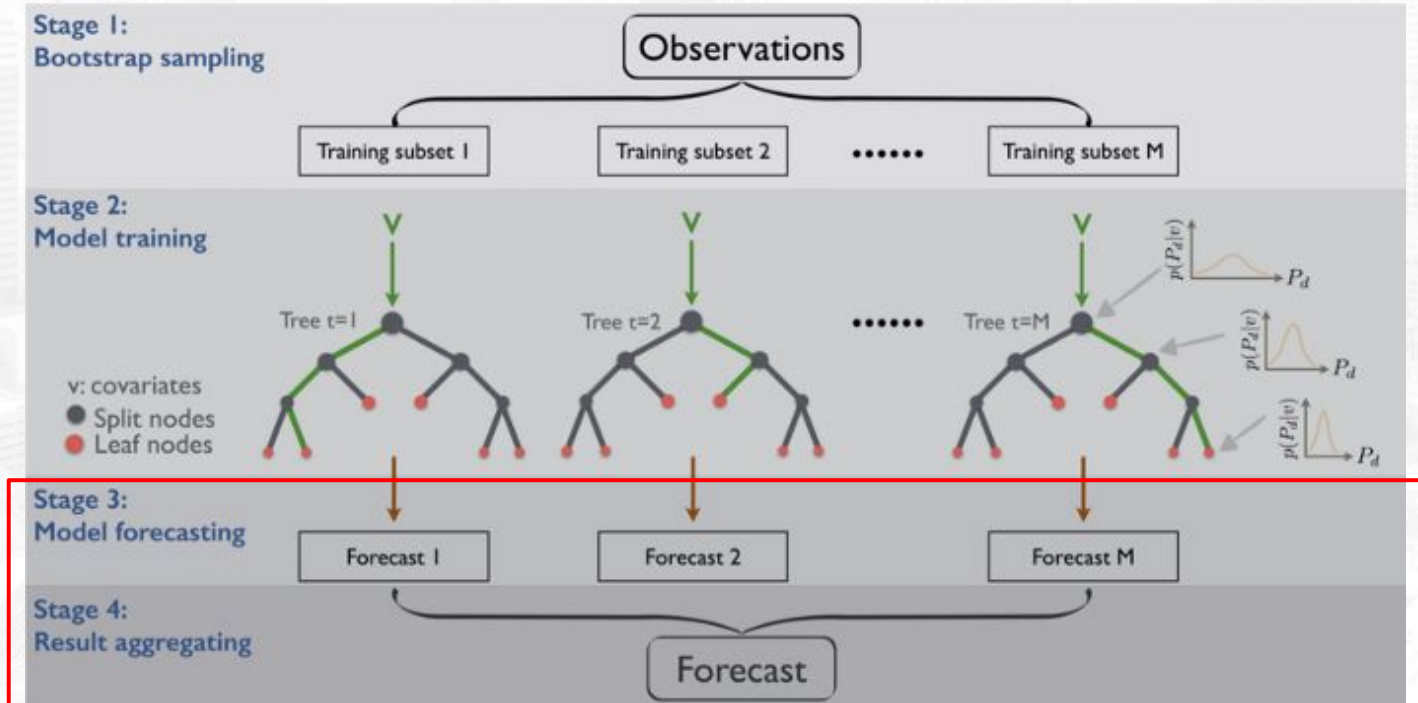
# Step by Step Bagging

2. Dari masing-masing sample tersebut akan dibentuk sebuah model.



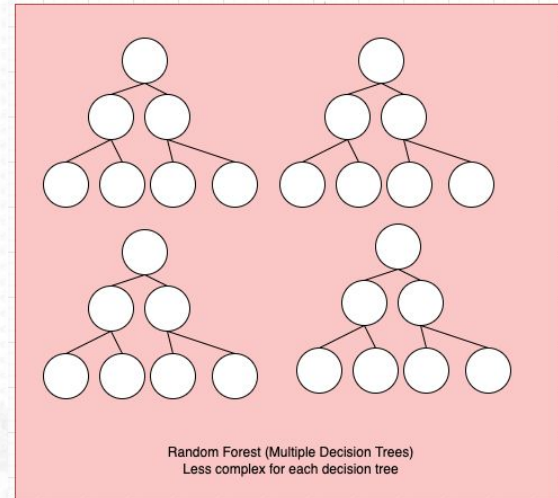
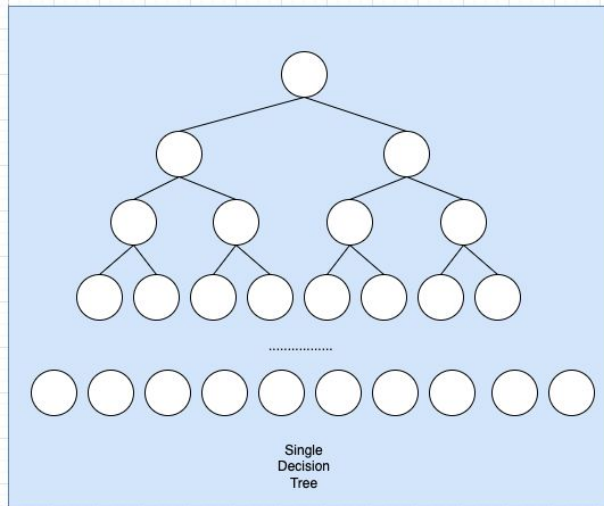
# Step by Step Bagging

3. Hasil prediksi dari model-model tersebut akan dikombinasikan dengan majority voting



# Bagging

- Bagging dapat membantu untuk mengurangi variance error (mencegah overfitting), karena tidak terlalu terpusat menggunakan data training lengkap yang sama pada sebuah model.



# Topik Supervised Learning: Ensemble Methods

- ☒ Introduction: Ensemble Method
- ☒ Bagging
- ☐ Hands On - Bagging
- ☐ Boosting
- ☐ Hands On - Boosting
- ☐ Stacking
- ☐ Hands On - Stacking
- ☐ Explainable AI
- ☐ Hands On - Shap Values



# Topik Supervised Learning: Ensemble Methods



Introduction: Ensemble Method



Bagging



Hands On - Bagging



Boosting



Hands On - Boosting



Stacking



Hands On - Stacking



Explainable AI



Hands On - Shap Values

# Contoh Implementasi Random Forest

```
from sklearn.ensemble import RandomForestClassifier

clf = RandomForestClassifier()
clf.fit(X_train, y_train)
```



*Buat model ML  
sesuai task*

Dokumentasi sklearn Random Forest:

<https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html>

# Hands On

# Topik Supervised Learning: Ensemble Methods

- |                                     |                               |                          |                        |
|-------------------------------------|-------------------------------|--------------------------|------------------------|
| <input checked="" type="checkbox"/> | Introduction: Ensemble Method | <input type="checkbox"/> | Stacking               |
| <input checked="" type="checkbox"/> | Bagging                       | <input type="checkbox"/> | Hands On - Stacking    |
| <input checked="" type="checkbox"/> | Hands On - Bagging            | <input type="checkbox"/> | Explainable AI         |
| <input type="checkbox"/>            | Boosting                      | <input type="checkbox"/> | Hands On - Shap Values |
| <input type="checkbox"/>            | Hands On - Boosting           |                          |                        |

# Topik Supervised Learning: Ensemble Methods



Introduction: Ensemble Method



Bagging



Hands On - Bagging



Boosting



Hands On - Boosting



Stacking



Hands On - Stacking



Explainable AI

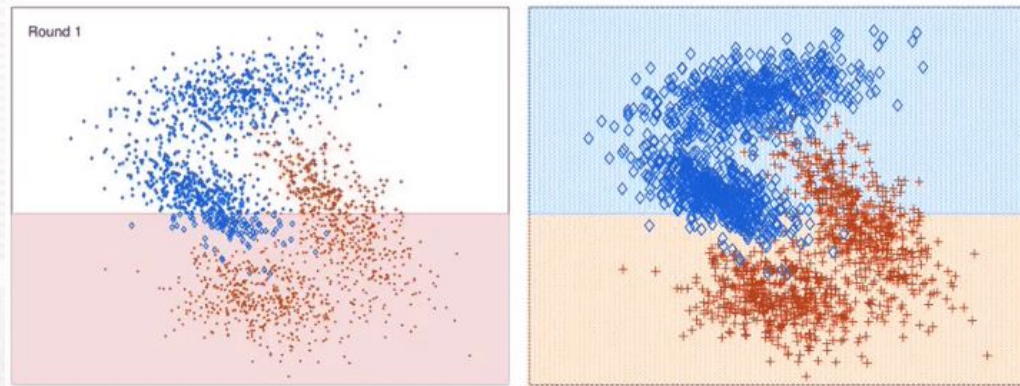


Hands On - Shap Values



# Boosting

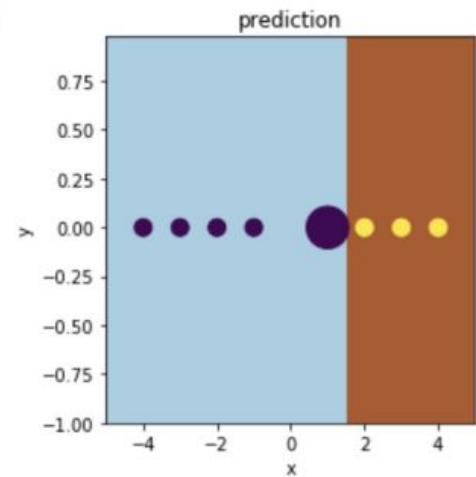
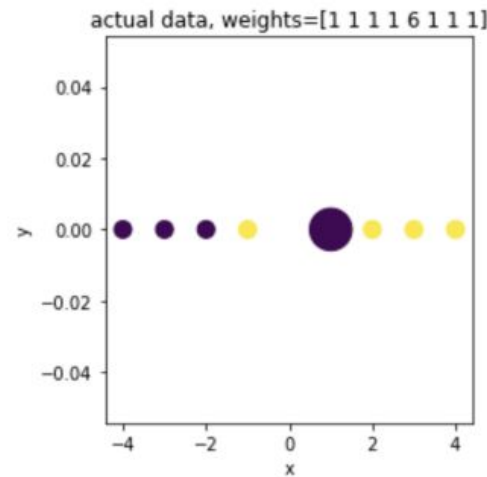
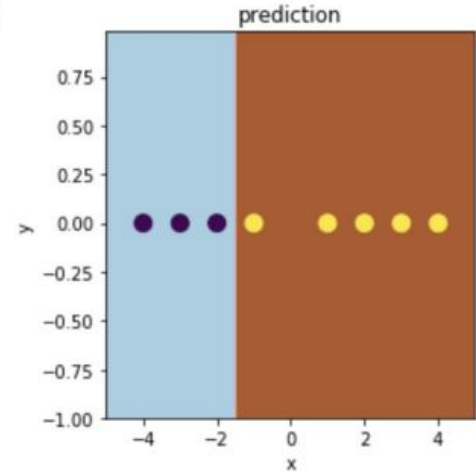
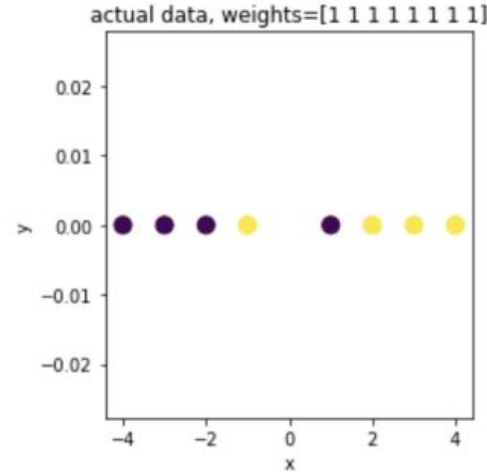
# Apa itu Boosting?



- Boosting menurunkan bias dengan membuat model prediksi yang lebih kuat dari sebelumnya.
- Boosting melakukan perbaikan weight dari hasil klasifikasi sebelumnya. Jika hasil prediksi tidak tepat, proses boosting mencoba meningkatkan weight dari data tersebut
- Hasil prediksi cenderung dapat lebih baik daripada bagging, namun ada kecenderungan overfitting
- Penting untuk dilakukan parameter tuning untuk menghindari overfitting

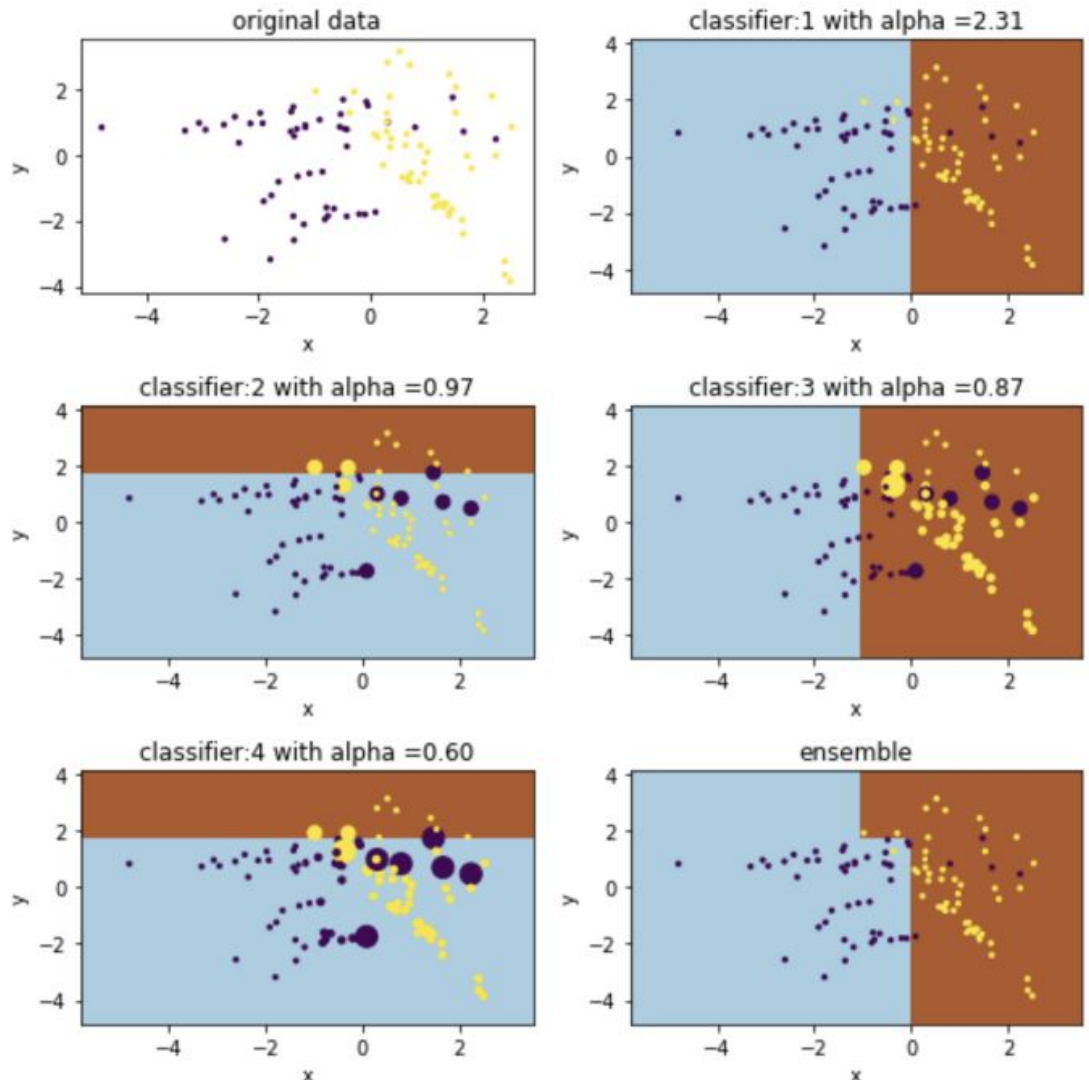
# Step by step Boosting

- Dibentuk model dengan menggunakan keseluruhan data training
- Dari hasil prediksi model pertama, boosting akan memberikan weight lebih besar pada data yang diprediksi kurang tepat pada model sebelumnya
- Proses yang digunakan adalah proses iteratif dengan menambahkan algoritma baru pada data yang sudah diberi bobot hingga limit yang telah ditentukan



# Contoh

- Hasil prediksi dari classifier 1 memberikan bobot yang lebih pada data yang terprediksi salah pada classifier 2, dst.
- Pada akhir proses ensemble, dilakukan majority voting



# Contoh Implementasi: Adaboost

```
from sklearn.ensemble import AdaBoostClassifier

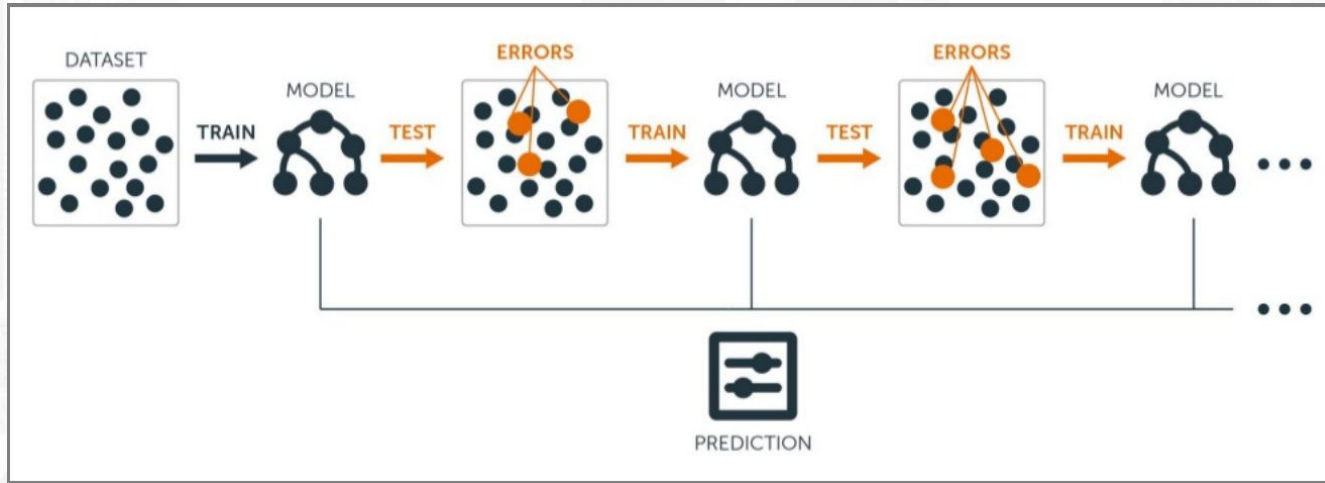
clf = AdaBoostClassifier()
clf.fit(X_train, y_train)
```

Dokumentasi sklearn Adaboost:

<https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.AdaBoostClassifier.html>



# XGBoost



- XGBoost melakukan perbaikan terhadap model tree yang dibangun pada iterasi sebelumnya
- Memanfaatkan konsep regularization untuk mencegah overfitting

# Contoh Implementasi: XGBoost

```
from xgboost import XGBClassifier  
  
xg = XGBClassifier()  
xg.fit(X_train,y_train)
```

Dokumentasi XGBoost:

<https://xgboost.readthedocs.io/en/latest/parameter.html>

# Algoritma lainnya

Jenis algoritma lain yang dapat kamu explore

[Naive Bayes](#)

[Support Vector Machine](#)

[LightGBM](#)

[Neural Network](#)

dll.

# Topik Supervised Learning: Ensemble Methods

- |                                     |                               |                          |                        |
|-------------------------------------|-------------------------------|--------------------------|------------------------|
| <input checked="" type="checkbox"/> | Introduction: Ensemble Method | <input type="checkbox"/> | Stacking               |
| <input checked="" type="checkbox"/> | Bagging                       | <input type="checkbox"/> | Hands On - Stacking    |
| <input checked="" type="checkbox"/> | Hands On - Bagging            | <input type="checkbox"/> | Explainable AI         |
| <input checked="" type="checkbox"/> | Boosting                      | <input type="checkbox"/> | Hands On - Shap Values |
| <input type="checkbox"/>            | Hands On - Boosting           |                          |                        |

# Topik Supervised Learning: Ensemble Methods



Introduction: Ensemble Method



Bagging



Hands On - Bagging



Boosting



Hands On - Boosting



Stacking



Hands On - Stacking



Explainable AI




Hands On - Shap Values

# Hands On



# Topik Supervised Learning: Ensemble Methods

- |   |                               |   |                        |
|---|-------------------------------|---|------------------------|
|  | Introduction: Ensemble Method |  | Stacking               |
|  | Bagging                       |  | Hands On - Stacking    |
|  | Hands On - Bagging            |  | Explainable AI         |
|  | Boosting                      |  | Hands On - Shap Values |
|  | Hands On - Boosting           |   |                        |

# Topik Supervised Learning: Ensemble Methods



Introduction: Ensemble Method



Bagging



Hands On - Bagging



Boosting



Hands On - Boosting



Stacking



Hands On - Stacking



Explainable AI

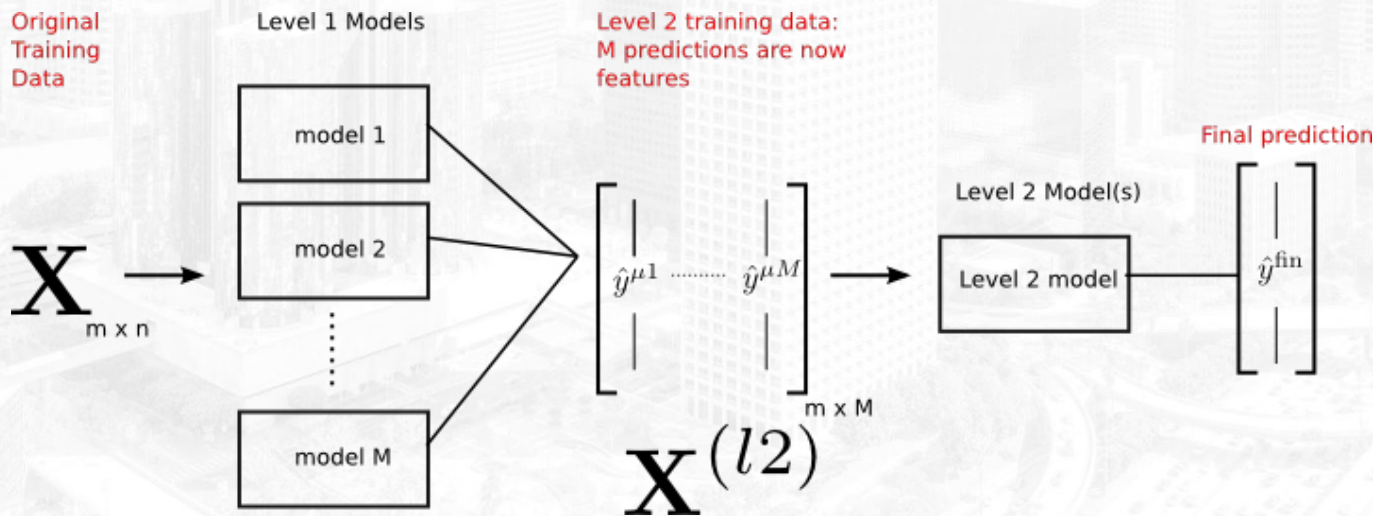


Hands On - Shap Values

# Stacking

# Apa itu Stacking?



- Stacking melakukan kombinasi dari banyak model klasifikasi/regresi.
- Model (level-1) awalnya dilatih pada keseluruhan train set, lalu meta-model selanjutnya (level-2) akan dilatih dari hasil prediksi model sebelumnya.



# Bagaimana memilih model pada Stacking?

- There's no free lunch. Coba kombinasi model yang berbeda dengan parameter yang berbeda juga. Hingga akhirnya kita paham bahwa beberapa jenis algoritma baik dalam beberapa kasus dan bisa kita lakukan stacking.

# Topik Supervised Learning: Ensemble Methods

- |   |                               |   |                        |
|---|-------------------------------|---|------------------------|
|  | Introduction: Ensemble Method |  | Stacking               |
|  | Bagging                       |  | Hands On - Stacking    |
|  | Hands On - Bagging            |  | Explainable AI         |
|  | Boosting                      |  | Hands On - Shap Values |
|  | Hands On - Boosting           |   |                        |



# Topik Supervised Learning: Ensemble Methods



Introduction: Ensemble Method



Bagging



Hands On - Bagging



Boosting



Hands On - Boosting



Stacking



Hands On - Stacking



Explainable AI



Hands On - Shap Values

# Hands On

## Contoh Implementasi

```
from sklearn.ensemble import StackingClassifier

level1 = list()
level1.append(('lr', LogisticRegression()))
level1.append(('knn', KNeighborsClassifier()))
level1.append(('cart', DecisionTreeClassifier()))
level1.append(('svm', SVC()))
level1.append(('bayes', GaussianNB()))

# definisikan model level 2
level2 = LogisticRegression()

# define the stacking ensemble
model = StackingClassifier(estimators=level1, final_estimator=level2,
cv=5)











# fit the model on all available data
model.fit(X_train, y_train)
```

# Hands On

# Topik Supervised Learning: Ensemble Methods

- ☒ Introduction: Ensemble Method
- ☒ Bagging
- ☒ Hands On - Bagging
- ☒ Boosting
- ☒ Hands On - Boosting
- ☒ Stacking
- ☒ Hands On - Stacking
- ☐ Explainable AI
- ☐ Hands On - Shap Values

# Topik Supervised Learning: Ensemble Methods

-  Introduction: Ensemble Method
-  Bagging
-  Hands On - Bagging
-  Boosting
-  Hands On - Boosting
-  Stacking
-  Hands On - Stacking
-   Explainable AI
-  Hands On - Shap Values

# Explainable AI

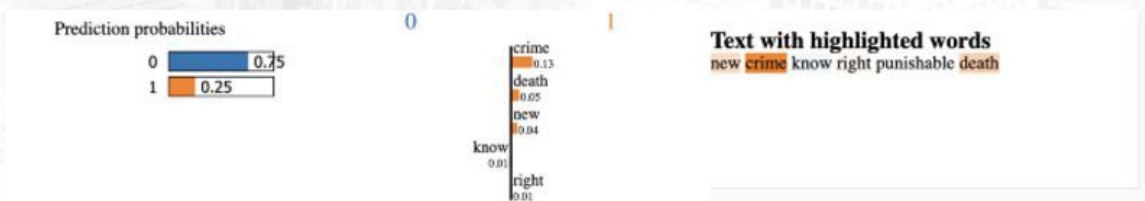


# Explainable AI (XAI)

Cara bagaimana manusia memahami bagaimana hasil prediksi dari sebuah model machine learning melakukan prediksi.



Grad-CAM gives you a class-discriminative visual explanation for the predictions of your CNN model.



LIME for interpret your model

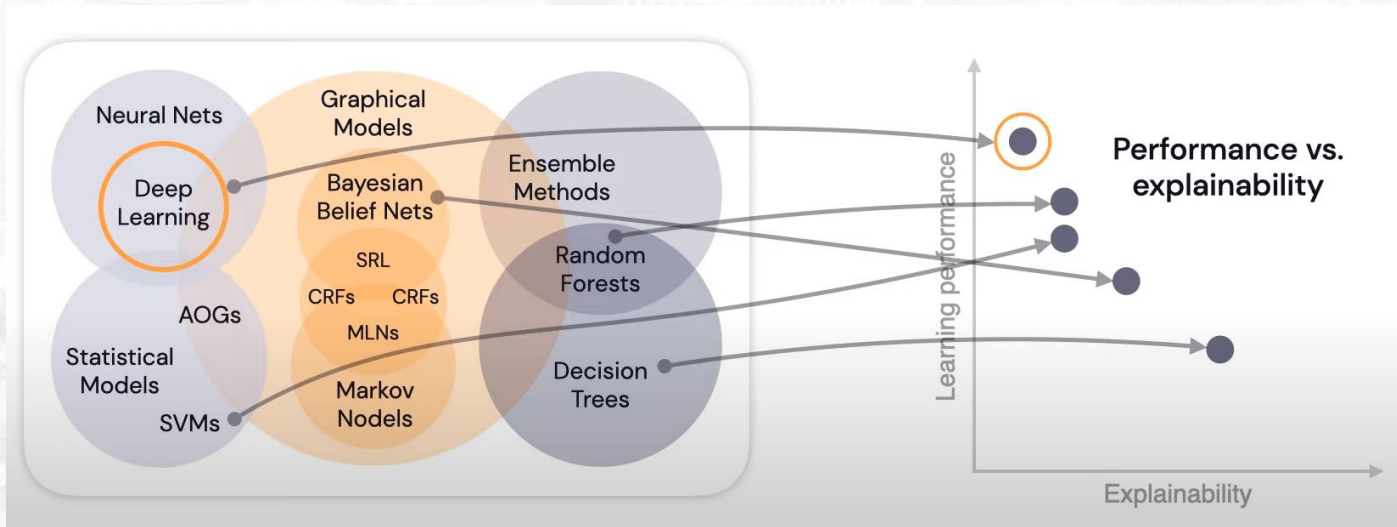
# Mengapa penting untuk masing-masing stakeholder?

XAI digunakan oleh masing-masing stakeholder sesuai dengan kebutuhannya: untuk memahami model lebih jelas, serta meningkatkan transparansi dan trust.

Engineers	Consumers	Regulators
Increase Understanding	Increase Trust	Increase Trust
Improve Performance	Bias & Transparency	Bias & Transparency
Create Better Algorithms	Understand Impact	Compliance
Produce Models	Reports & Analyses	Reports

# Apakah model butuh di interpretasi?

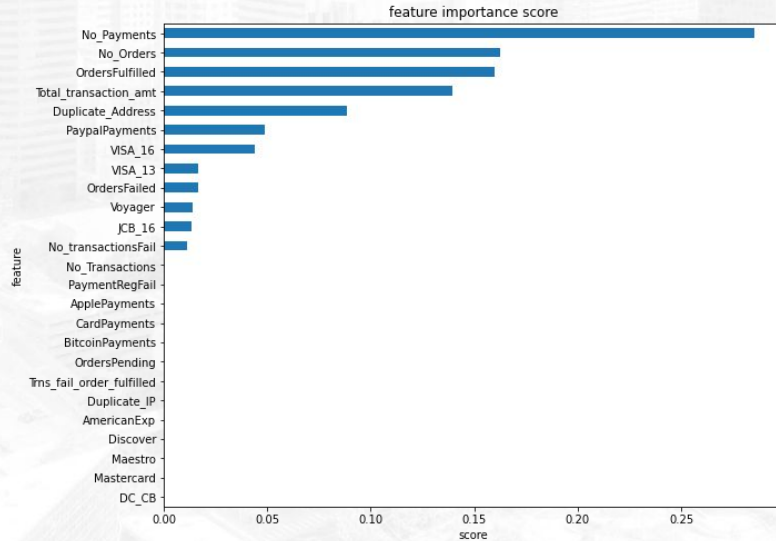
Komunikasikan dengan stakeholder, jika butuh di interpretasi dan diambil keputusan dari feature importance, atau model yang black box sudah cukup?



Complexity vs. Explainability

# Feature Importance salah satu XAI sederhana

Beberapa algoritma sederhana menyediakan Feature Importance untuk dipahami pengaruh dari masing-masing feature namun secara global atau menyeluruh. Bagaimana cara interpretasi per-data point?



# Metode Populer untuk XAI

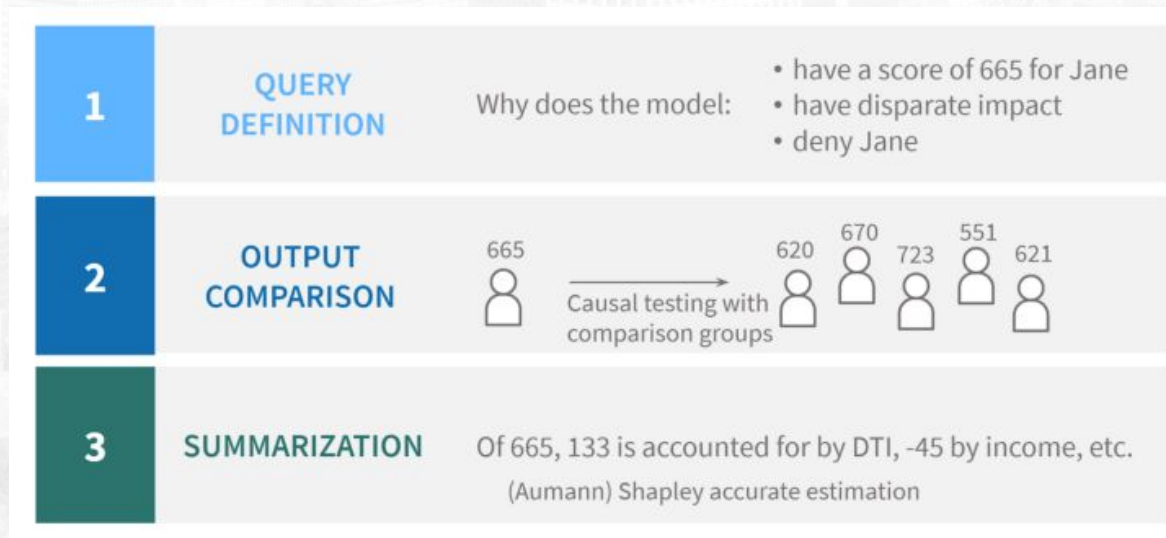
Beberapa metode XAI populer:

- SHAP (SHapley Additive exPlanations)
- Local interpretable model-agnostic explanations (LIME)
- Partial Dependence Plot (PDP)
- Individual Conditional Expectation (ICE)
- Feature Interaction
- Scoped Rules (Anchors)



# SHAP Values [\(Referensi\)](#)

Secara teknis, metode Shapley Value menggunakan konsep [game theory](#) yang melihat bagaimana gain & cost dari masing-masing aktor dalam sebuah koalisi (dalam hal ini pengaruh dari feature dalam prediksi targetnya).



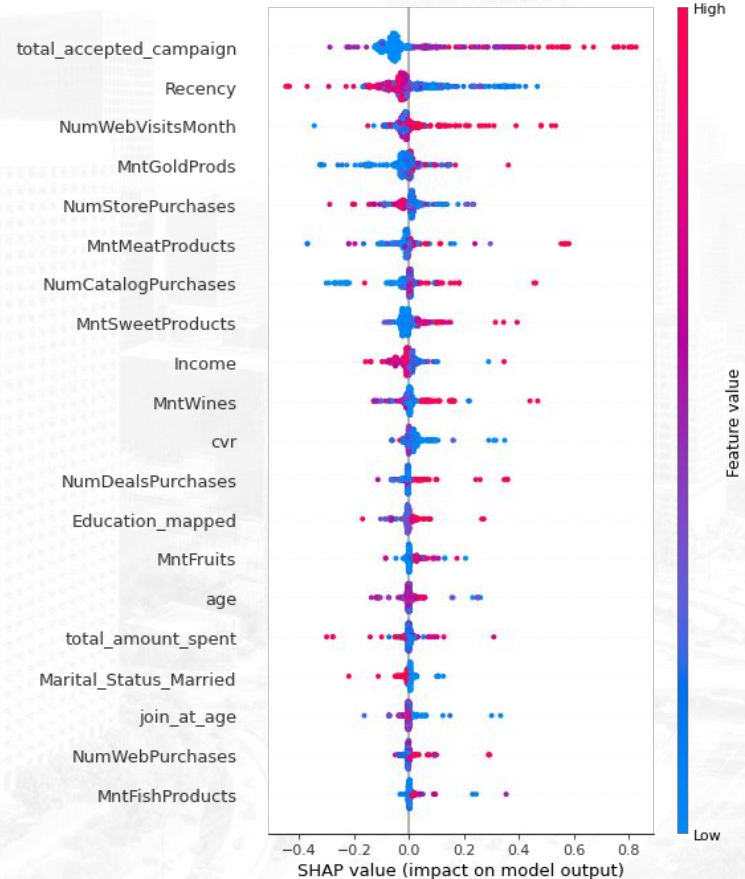


# SHAP Values

```
import shap
explainer = shap.TreeExplainer(trained_model)
shap_values = explainer.shap_values(X_test)
shap.summary_plot(shap_values[1], X_test)
```

## Interpretasi:

- Masing-masing data point adalah interpretasi dari masing-masing data test
- Feature total\_accepted\_campaign berkorelasi secara positif (merah). Semakin besar total\_accepted\_campaign, semakin besar shap valuenya (probabilitas untuk respon campaign)
- Semakin kecil nilai recency (biru), semakin besar probabilitas untuk respon campaign



# SHAP Values - Force Plot

```
shap.force_plot(explainer.expected_value[1], shap_values[1][0:1], X_test.iloc[0,:])
```



Interpretasi:

- Force plot melakukan plot terhadap pengaruh feature untuk melakukan prediksi di masing-masing data point
- Pada hasil diatas, terlihat bahwa feature yang berpengaruh terbesar adalah MntMeatProducts, total\_accepted\_campaign, Recency, dan Education\_mapped yang mempengaruhi customer tidak respon campaign

# Topik Supervised Learning: Ensemble Methods

- |   |                               |   |                        |
|---|-------------------------------|---|------------------------|
|  | Introduction: Ensemble Method |  | Stacking               |
|  | Bagging                       |  | Hands On - Stacking    |
|  | Hands On - Bagging            |  | Explainable AI         |
|  | Boosting                      |  | Hands On - Shap Values |
|  | Hands On - Boosting           |   |                        |

# Topik Supervised Learning: Ensemble Methods



Introduction: Ensemble Method



Bagging



Hands On - Bagging



Boosting



Hands On - Boosting



Stacking



Hands On - Stacking



Explainable AI



Hands On - Shap Values

# Hands On

# Topik Supervised Learning: Ensemble Methods

- ✓ Introduction: Ensemble Method
- ✓ Bagging
- ✓ Hands On - Bagging
- ✓ Boosting
- ✓ Hands On - Boosting
- ✓ Stacking
- ✓ Hands On - Stacking
- ✓ Explainable AI
- ✓ Hands On - Shap Values