



LEARNING PROGRESS REVIEW

Week 13

Entropy Team

A large, light gray semi-circle graphic located in the bottom right corner of the slide.

DAFTAR ISI

1.

Communication and Presentation Skill

Presentation tips and data storytelling

2.


Evaluation Metrics and Hyperparameter Tuning

Confusion matrix, cross validation, grid search

3.

Advanced Machine Learning

Feature selection and explainable AI



01

COMMUNICATION AND PRESENTATION SKILL

Presentation tips and
data storytelling

Public Speaking

- Orang lain umumnya memahami apa yang kita katakan
- Agar **orang lain mau mendengarkan** kita, itu yang menjadi tantangan
- Oleh sebab itu, konten dalam *public speaking* **kebanyakan** merupakan **unsur psikologis** dibanding unsur teknis
- Tips dalam *public speaking*
 - **Percaya diri**, anggap diri sendiri sebagai orang yang **paling memahami materi** yang akan kita sampaikan
 - **Jangan sampaikan** persis seperti apa **yang tertulis** pada *slide*



Presentation Tips

- Buat presentasi yang **sederhana**
- **Hindari** penggunaan **animasi** yang tidak diperlukan
- Perhatikan penggunaan **jenis font**
- Perhatikan **warna**, sesuaikan dengan tema perusahaan atau *client*
- Sesuaikan konten dengan **pendengar**
- Buat **alur presentasi** yang jelas

Valuable Insight



- Banyak perusahaan yang kesulitan **mengubah data** yang banyak **menjadi *insight*** yang berguna
- Tidak peduli *tools* apa yang kita gunakan, yang penting adalah **bagaimana cara kita menggunakannya**
- Akan tetapi, *tools* yang bagus juga memiliki keunggulan tersendiri

Valuable Insight

Langkah agar kita mendapatkan *insight* yang berguna

- Kenali calon **audience**
- Tentukan **tujuan** yang jelas
- Tentukan **data** yang akan digunakan
- Tentukan cara menyampaikan **cerita**
- Gunakan *feedback* untuk **perbaiki** di kemudian hari

Data Storytelling

Langkah untuk menyampaikan cerita dengan data

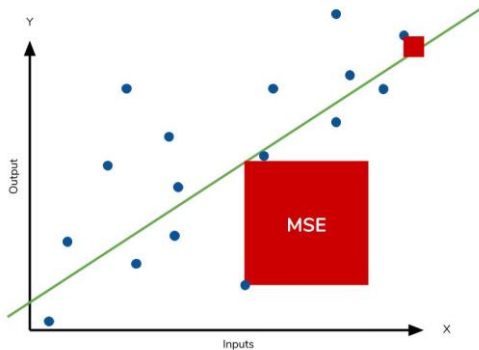
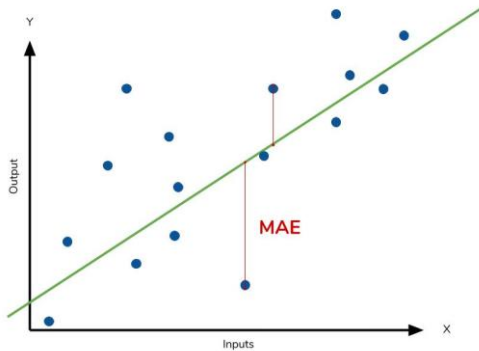
- Pahami **konteks**
- Gunakan jenis **visualisasi** yang sesuai
- Hilangkan bagian dari visualisasi yang **tidak penting**
- *Highlight* bagian yang ingin dijadikan **fokus utama**
- Berpikirlah seperti **designer**
- **Sampaikanlah cerita** yang menarik

02

EVALUATION METRICS AND HYPERPARAMETER TUNING

Confusion matrix, cross
validation, grid search

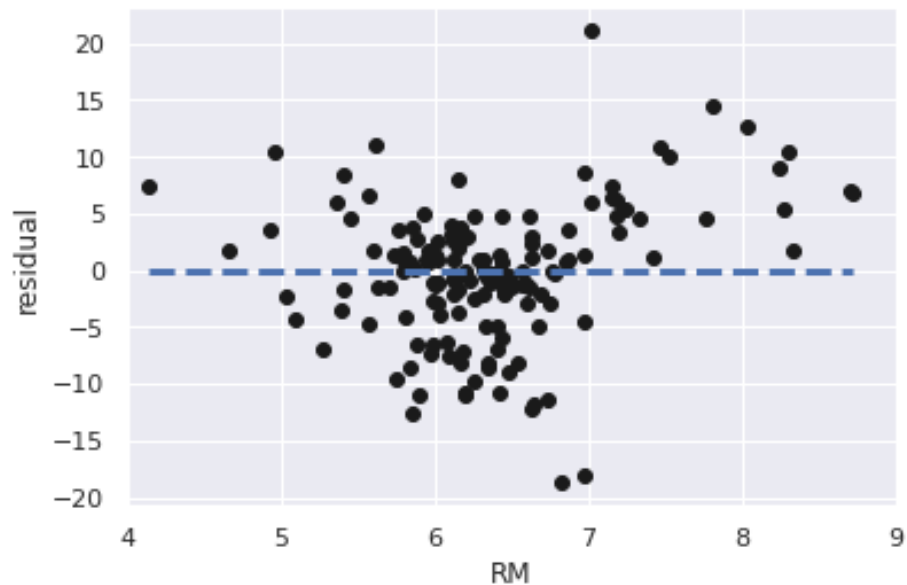
Regression Case



Metrics yang umum digunakan untuk menilai performa model regresi:

- *Mean Absolute Error (MAE)*
- *Mean Squared Error (MSE)*
- *Root Mean Squared Error (RMSE)*
- *R-squared (R^2)*

Regression Case



Model regresi dikatakan bagus jika:

- Nilai MAE, MSE, atau RMSE **relatif rendah**
- Nilai R-squared **mendekati 1**

Classification Case

Metrics yang umum digunakan untuk menilai performa model klasifikasi:

- *Confusion matrix*
- *Accuracy*
- *Precision*
- *Recall*
- *F1 score*
- *ROC – AUC curve*

Confusion Matrix

		Prediction	
		Positive	Negative
Actual	Positive	True Positive	False Negative
	Negative	False Positive	True Negative

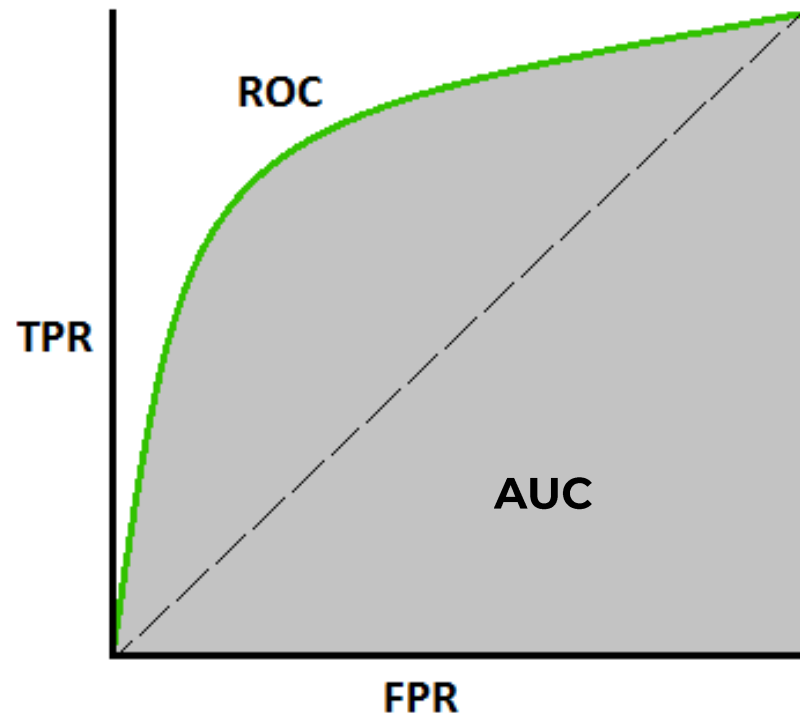
- *Confusion matrix* adalah tabel yang berisikan kelompok hasil prediksi
- Dalam *confusion matrix*, terdapat 2 **jenis error**
 - *Type 1 (false positive)*
 - *Type 2 (false negative)*

Confusion Matrix

Dari *confusion matrix*, kita dapat menghitung beberapa *metrics* seperti:

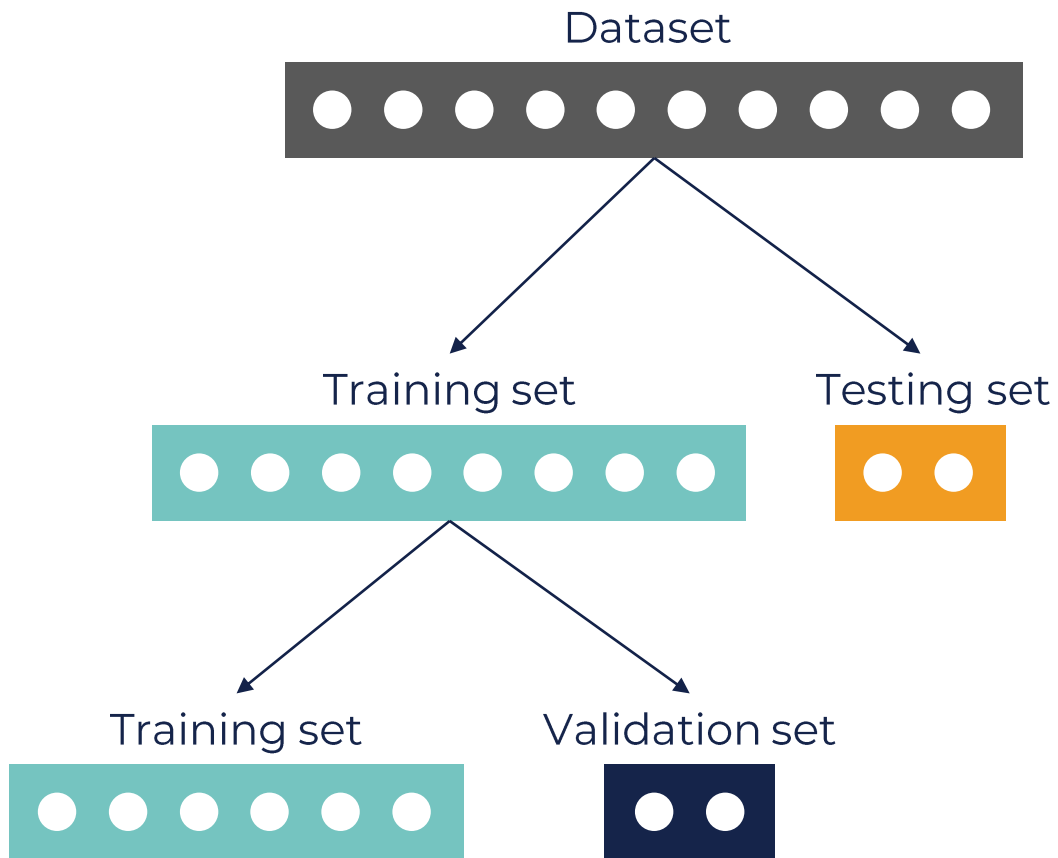
- **Accuracy**, yaitu kemampuan model untuk memprediksi nilai positif (*true positive*) dan negatif (*true negative*) secara tepat
- **Precision**, yaitu kemampuan model untuk memprediksi nilai positif dan meminimalkan *false positive*
- **Recall**, yaitu kemampuan model untuk memprediksi nilai positif dan meminimalkan *false negative*
- **F1 score**, yaitu rata-rata antara *precision* dan *recall*

ROC – AUC Curve



- Sumbu dari grafik ROC dan AUC terbentuk dari nilai *true positive rate* (TPR) dan *false positive rate* (FPR)
- Model klasifikasi dikatakan bagus jika:
 - ROC mendekati **pojok kiri atas**
 - AUC **mendekati 1**

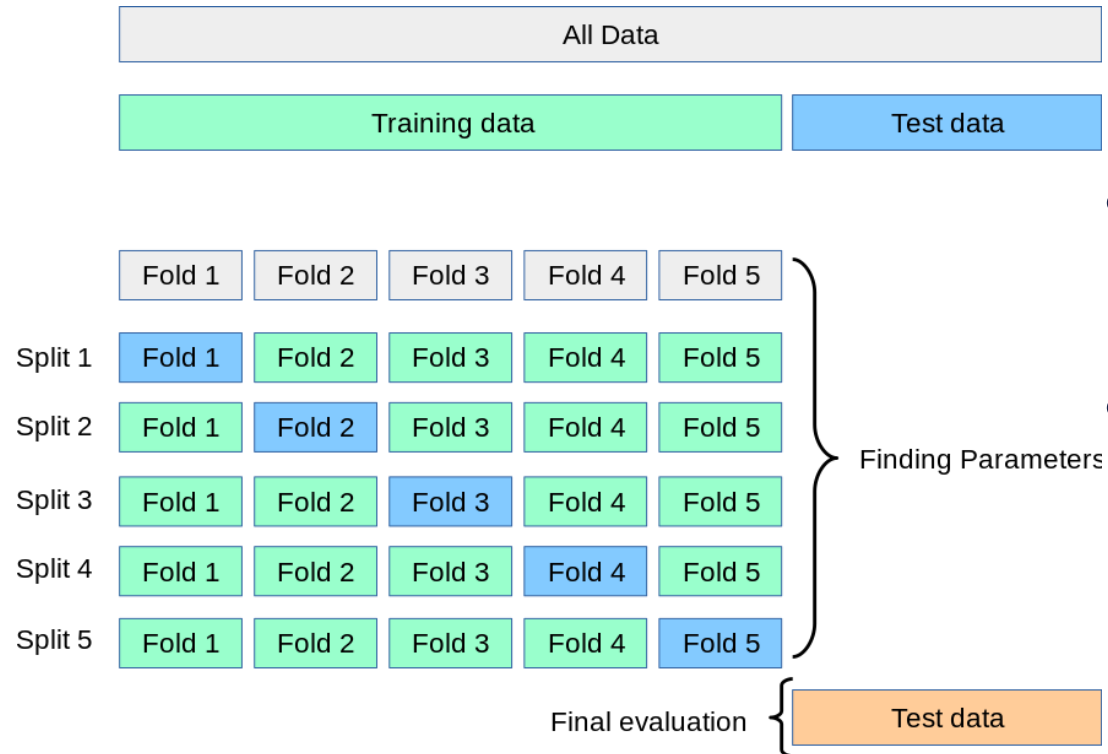
Cross Validation



Langkah melakukan *cross validation*:

- Pecah *dataset* menjadi **training set** dan **testing set**
- Pada **training set**, ubah sebagian data menjadi **validation set**
- **Validation set** digunakan untuk menguji apakah model sudah **sesuai harapan** atau belum
- **Testing set** digunakan untuk **pengujian akhir**

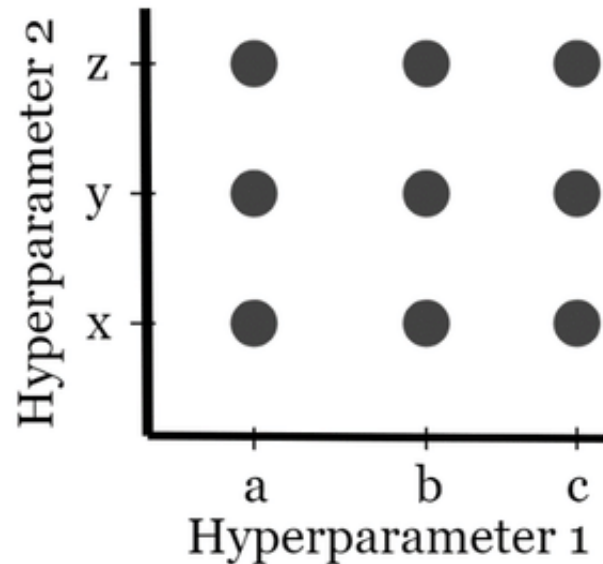
K-Fold Cross Validation



- Langkahnya mirip dengan *cross validation*
- Bedanya, pada *training set* dilakukan sebanyak K-kali pembuatan ***validation set yang berbeda***

Grid Search

Pseudocode
Hyperparameter_One = [a, b, c]
Hyperparameter_Two = [x, y, z]



- *Grid search* merupakan salah satu metode untuk **mencari hyperparameter yang terbaik**
- Secara konsep, *grid search* akan membuat beragam model dengan **semua kombinasi hyperparameter** yang telah ditentukan
- Setelah proses *grid search* selesai, model dengan nilai terbaik dapat dilihat *hyperparameter*-nya



03

ADVANCED MACHINE LEARNING

Feature selection and
explainable AI

Feature Selection

All Features



Feature Selection



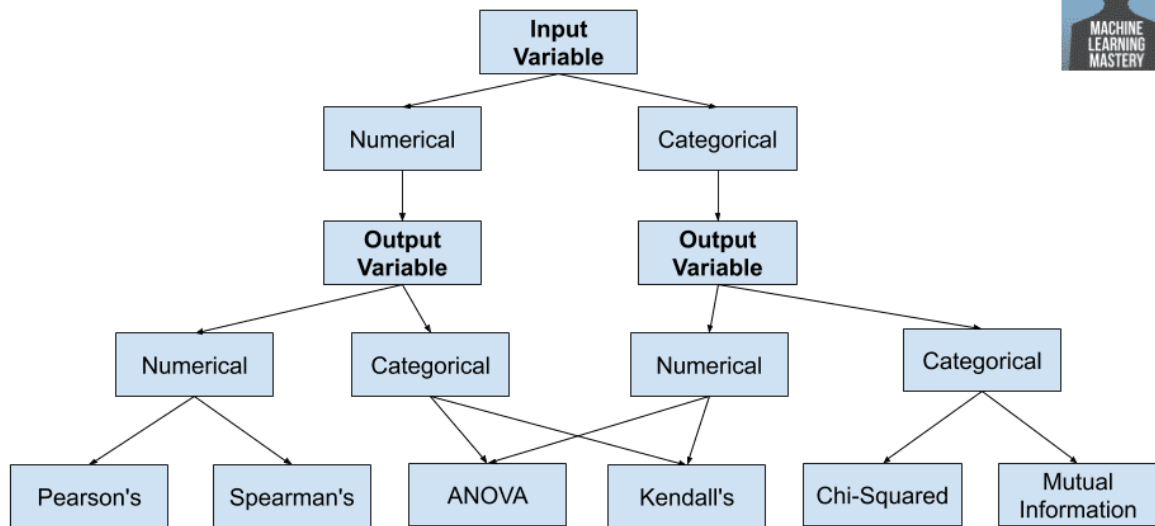
Final Features



- **Manfaat** melakukan *feature selection*:
 - Mengurangi beban **komputasi**
 - Mengurangi data **redundant**
 - Mengurangi **overfitting**
 - Mengurangi **kompleksitas** model
 - Umumnya meningkatkan **performa model**
- **Contoh** metode *feature selection*:
 - *Filter method*
 - *Wrapper method*
 - *Embedded method*

Filter Method

How to Choose a Feature Selection Method



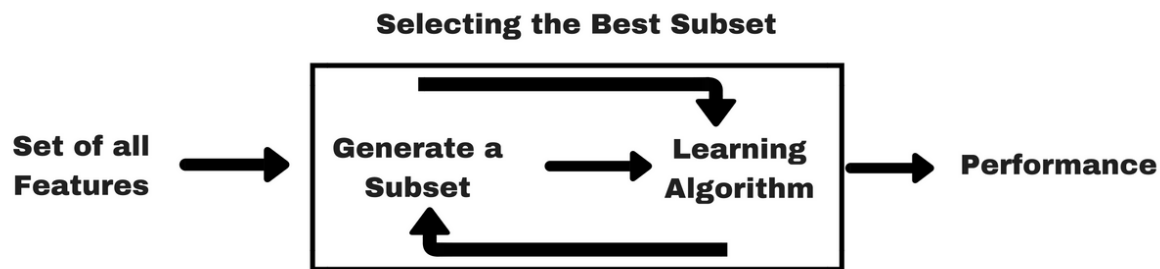
Copyright © MachineLearningMastery.com



- Pada *filter method*, *feature selection* dilakukan dengan menggunakan **metode uji statistik**
- Contoh uji statistik pada scikit-learn:
 - ANOVA
 - Chi-square
 - *Mutual information*

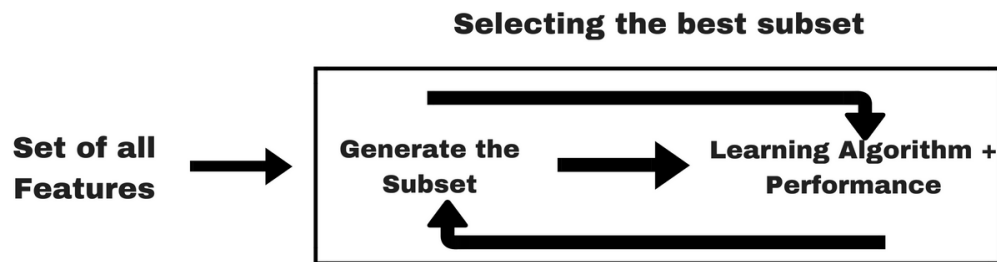
Sumber: <https://machinelearningmastery.com/feature-selection-with-real-and-categorical-data/>

Wrapper Method



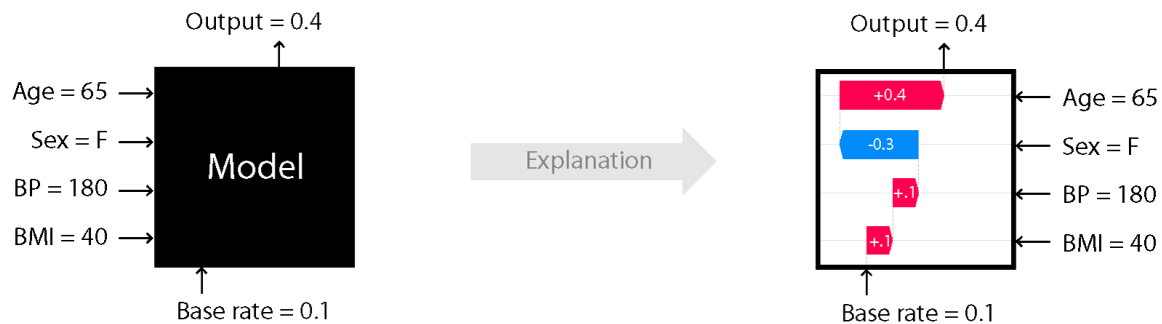
- Pada *wrapper method*, *feature selection* dilakukan dengan **membuat beberapa model** berdasarkan **kombinasi *feature***
- Contoh *wrapper method*:
 - *Forward selection*
 - *Backward elimination*
 - *Recursive feature elimination* (RFE)

Embedded Method



- Pada *embedded method*, *feature selection* dapat dilakukan oleh **algoritma model itu sendiri**
- Model yang dapat digunakan yaitu model yang memiliki **atribut *feature importance*** atau **koefisien**
- Contoh model yang dapat digunakan:
 - *Lasso regression*
 - *Ridge regression*
 - *Decision tree*

Explainable AI



- SHAP (SHapley Additive exPlanations) adalah salah satu metode untuk **menginterpretasikan hasil prediksi** dari model
- SHAP dapat digunakan untuk mengetahui **pengaruh** dari setiap **feature terhadap hasil prediksi**

THANKS

Entropy Team

CREDITS: This presentation template was originally created by **Slidesgo**, including icons by **Flaticon**, and infographics & images by **Freepik**