

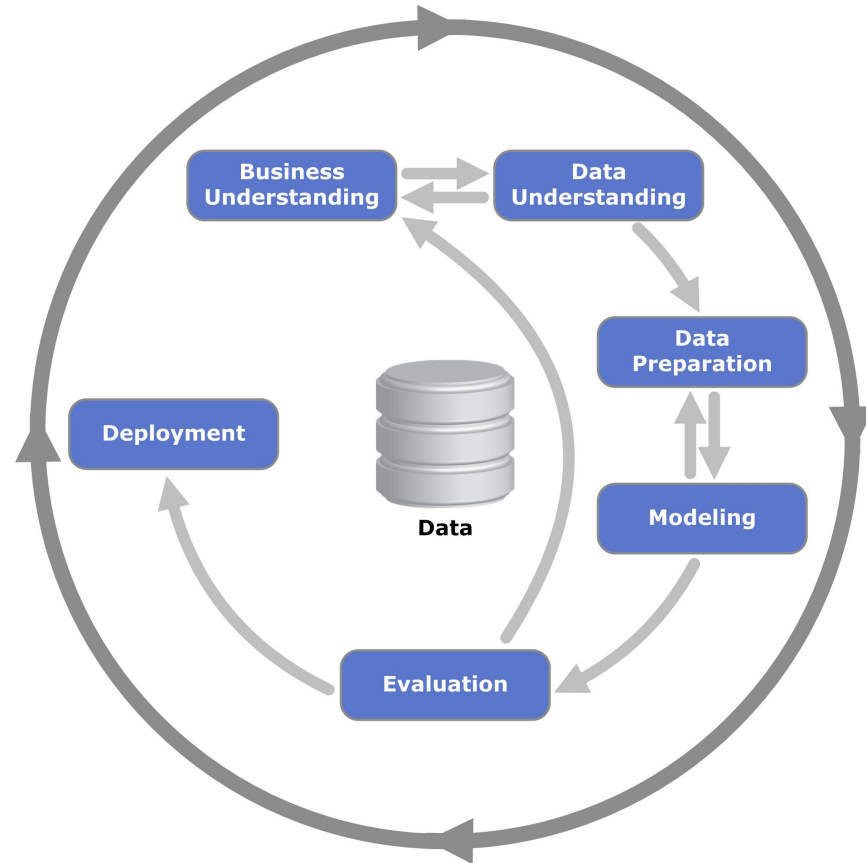


Extra Material for P1W2D3



P1 W2 D3 AM

CRISP-DM (CRoss Industry Standard Process for Data Mining)



Generalization vs Overfitting - Example

Sumber : <https://bulletin-cyberbully-detection.herokuapp.com/>


BULLETIN - Cyberbullying Detector

Determine if your tweet is considered as cyberbullying or not:

Your tweet:

Submit!

Stop Cyberbullying, right now!



The illustration depicts a person with long brown hair, wearing a yellow shirt and blue pants, sitting on the floor with their head buried in their arms, appearing distressed. They are surrounded by several speech bubbles containing cyberbullying messages: 'I hate you!', 'LOSER', '%!#@!', 'Ugly', 'You suck!', and a skull and crossbones icon. A green speech bubble with a red angry face icon is also present. The word 'Cyberbullying' is written in green, stylized text on the left. The background is a dark teal color with a white wavy line on the left.

Overfitting on the Test-Set

Analogi : Ujian Praktek SIM C di Indonesia

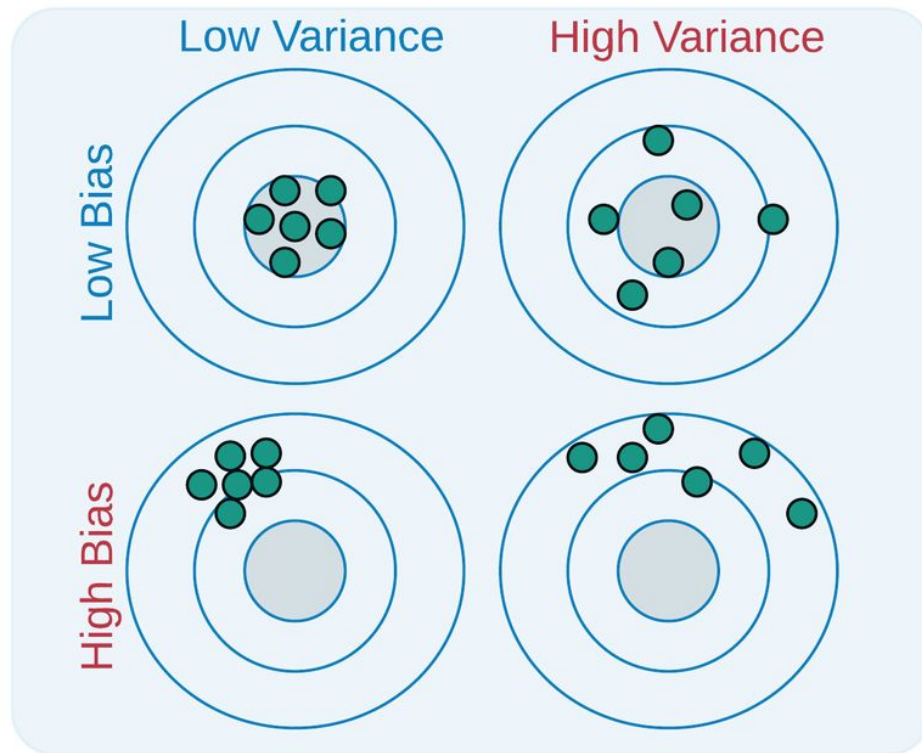


Overfitting on the Test-Set

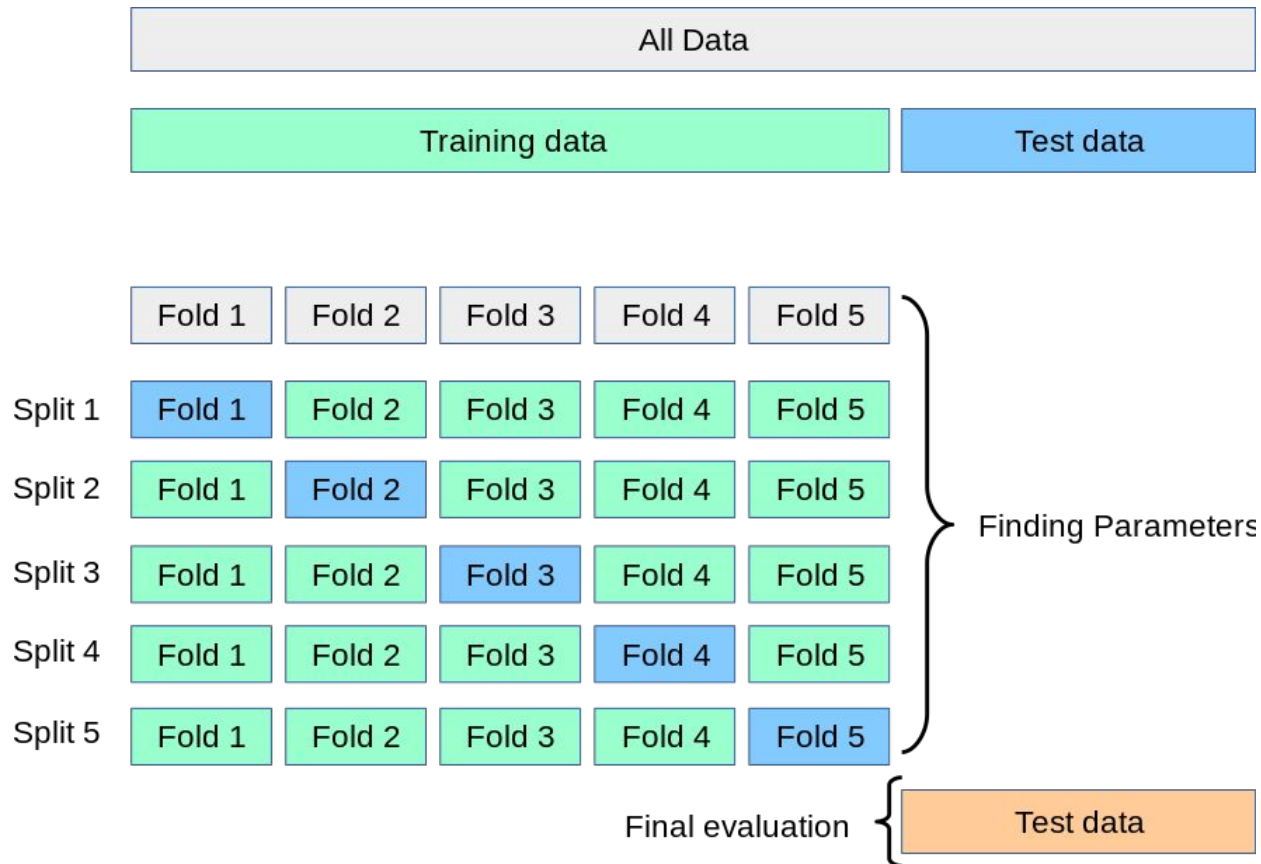
Analogi : Ujian Praktek di Malaysia



Bias vs Variance



Cross Validation



Leave One Out (LOOCV)



: Training Set



: Test Set

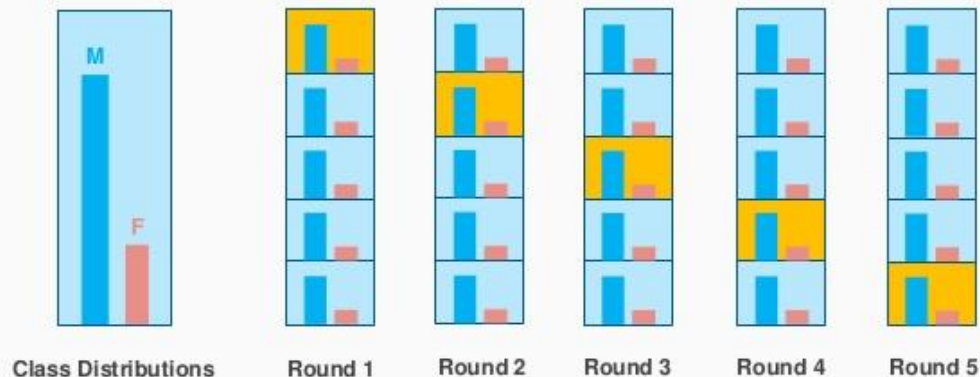


•
•
•



Stratified K-Fold

Stratified K-fold Cross Validation (K = 5)

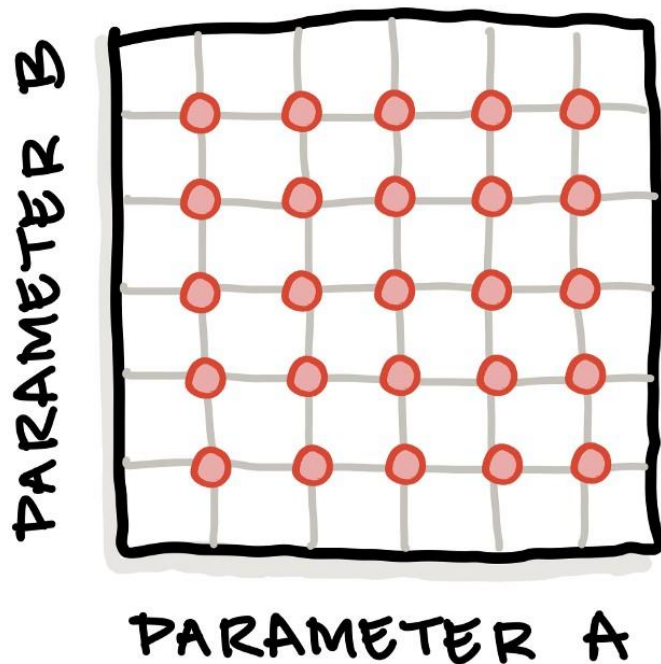


Keep the distribution of classes in each fold

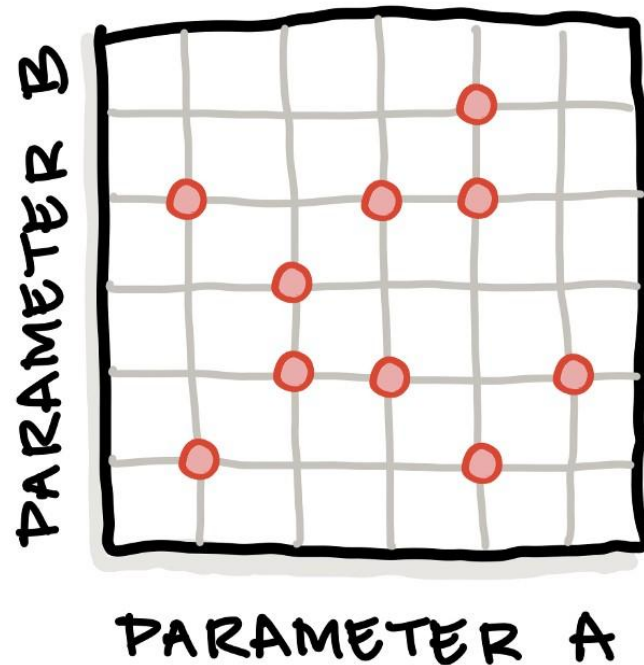
Training Data
Validation Data

Grid Search vs Random Search

GRID SEARCH



RANDOM SEARCH



The background of the slide features a complex, abstract geometric pattern. It consists of a network of thin, light gray lines that connect various small, dark gray dots. These dots are scattered across the slide, with a higher concentration in the upper half. The lines form a series of interconnected triangles and polygons of varying sizes, creating a mesh-like structure. The overall effect is a modern, technical, and somewhat organic-looking design.

P1 W2 D3 PM

Pipeline (Illustration)



GridSearch without/with Pipeline (Illustration)

GridSearch without Pipeline



GridSearch with Pipeline

