

I. 금융 빅데이터와 데이터 분석

(Big data in finance & financial time series data analysis)

④ 학습 데이터 유형

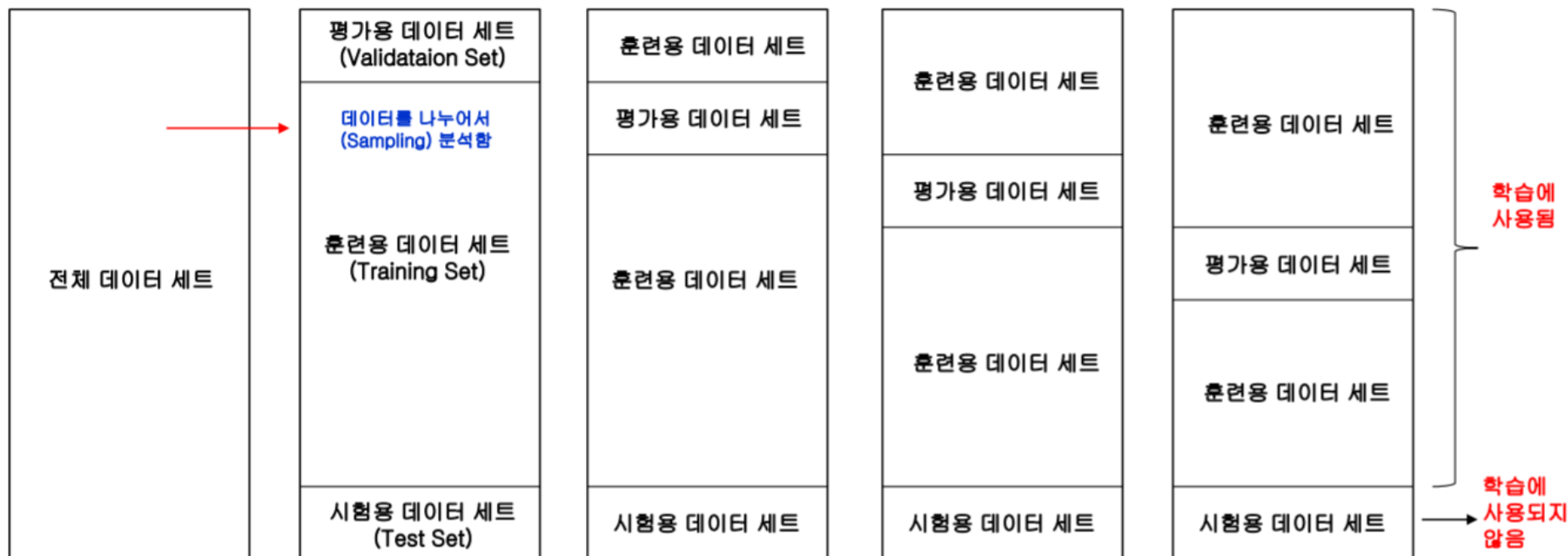
입력 데이터의 구분

- 입력 데이터는 **학습 데이터 (Training Data Set)**, **평가 데이터 (Validation Data Set)**, **시험 데이터 (Test Data Set)**로 구분됨.
- **학습 데이터**는 실제 학습 (분석)을 위한 데이터이고, **평가 데이터**는 과적합 (Overfitting)을 방지하고, 일반화 특성을 향상시키기 위해 사용됨 (학습을 어느 수준까지 할 것인가를 결정함). 그리고 **시험 데이터**는 결과의 신뢰성을 확인하기 위해 사용됨. 학습 데이터와 평가 데이터는 학습 과정에서 사용된 데이터이므로 결과의 신뢰성 확인은 학습에 사용되지 않은 시험 데이터를 이용함.
- 데이터 마이닝 알고리즘을 이용하여 입력 데이터를 학습하고 신뢰성 검증 과정을 거치면 **결과**가 나옴. 학습 결과는 데이터 후처리 과정을 거쳐 결과 해석, 시각화 과정 등을 통해 의사 결정에 사용 가능한 형태의 **지식**으로 변환됨.
- 최종 산출물인 지식은 의사 결정 시스템의 하나의 **입력**으로 사용될 수 있음

학습 데이터 유형

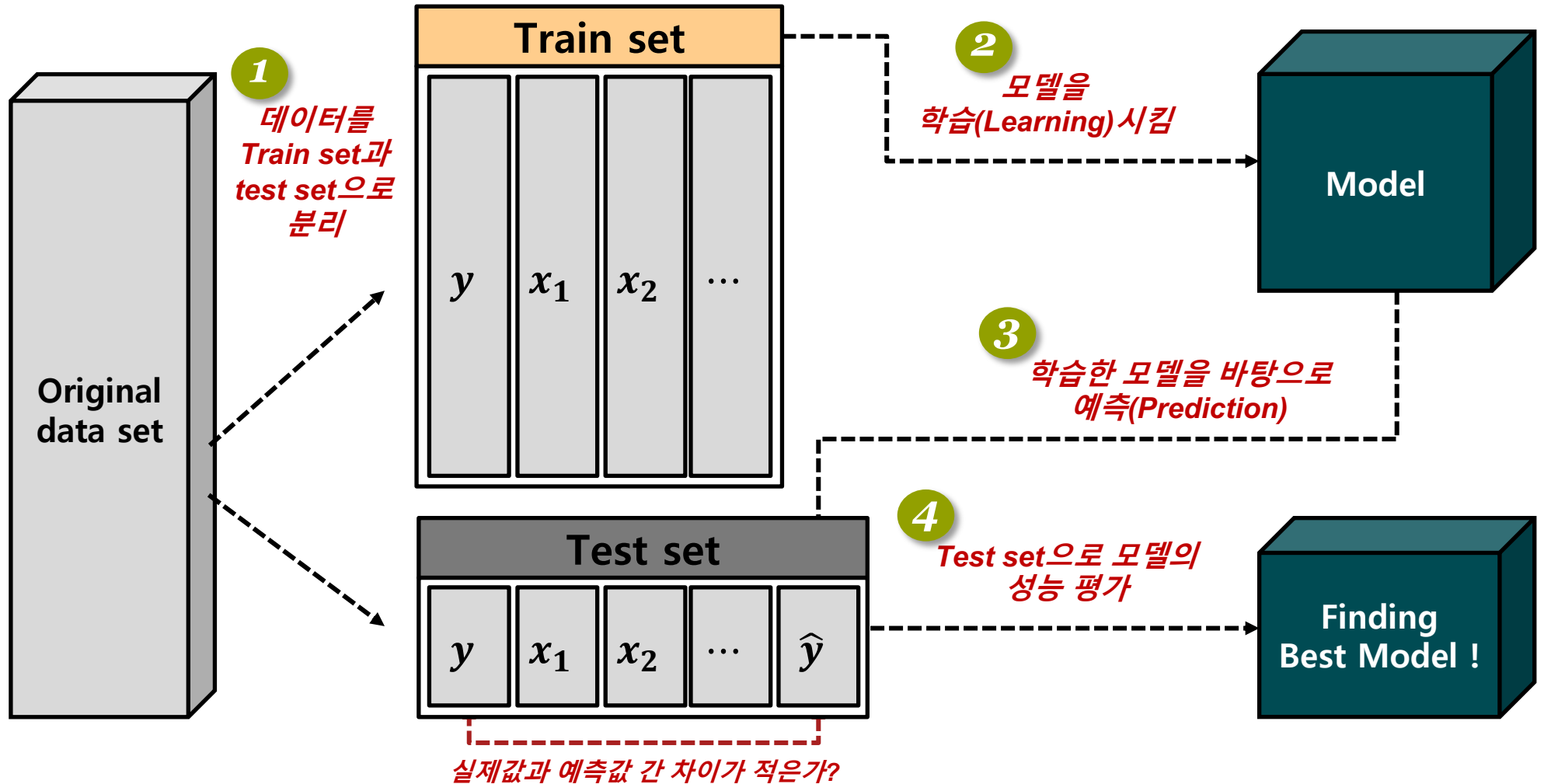
- Training Data (훈련용 데이터) : 실제 학습할 데이터
- Validation Data (평가용 데이터) : 어느 정도의 수준으로 학습하는 것이 적당한지를 결정
- Test Data (시험용 데이터) : 학습의 최종 성과를 평가하기 위해 사용

-> 분석할 데이터 세트를 위 세 가지로 나누어서 분석함.



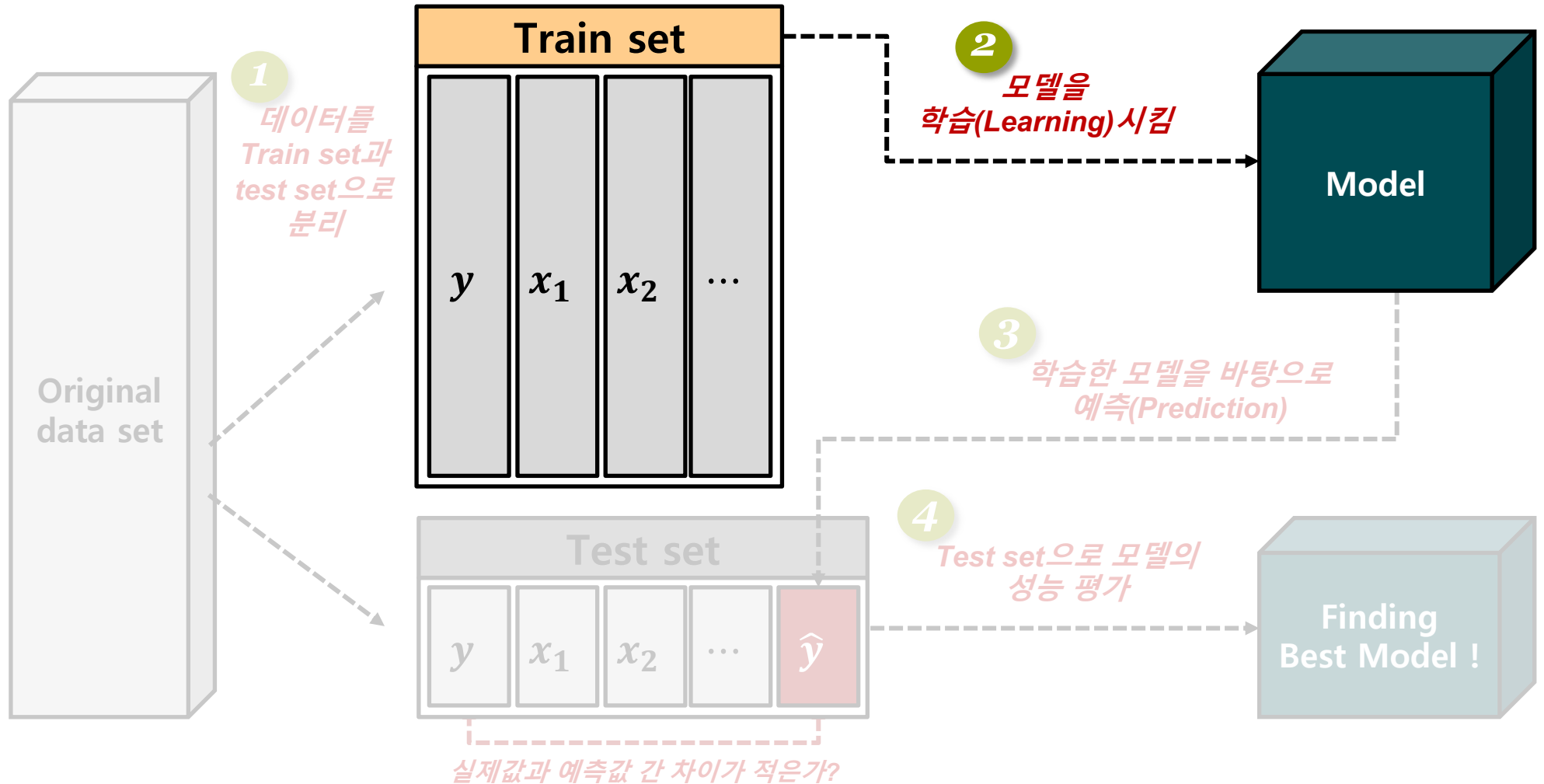
훈련(Training) and 검증(Testing)

예측모형을 도출하기 위해서는 내가 갖고 있는 원래의 자료(Original data set)를 훈련데이터(train set)와 검증데이터(test set)으로 나누고, 이를 지속적으로 학습시켜 최적의 모형을 찾아야 함



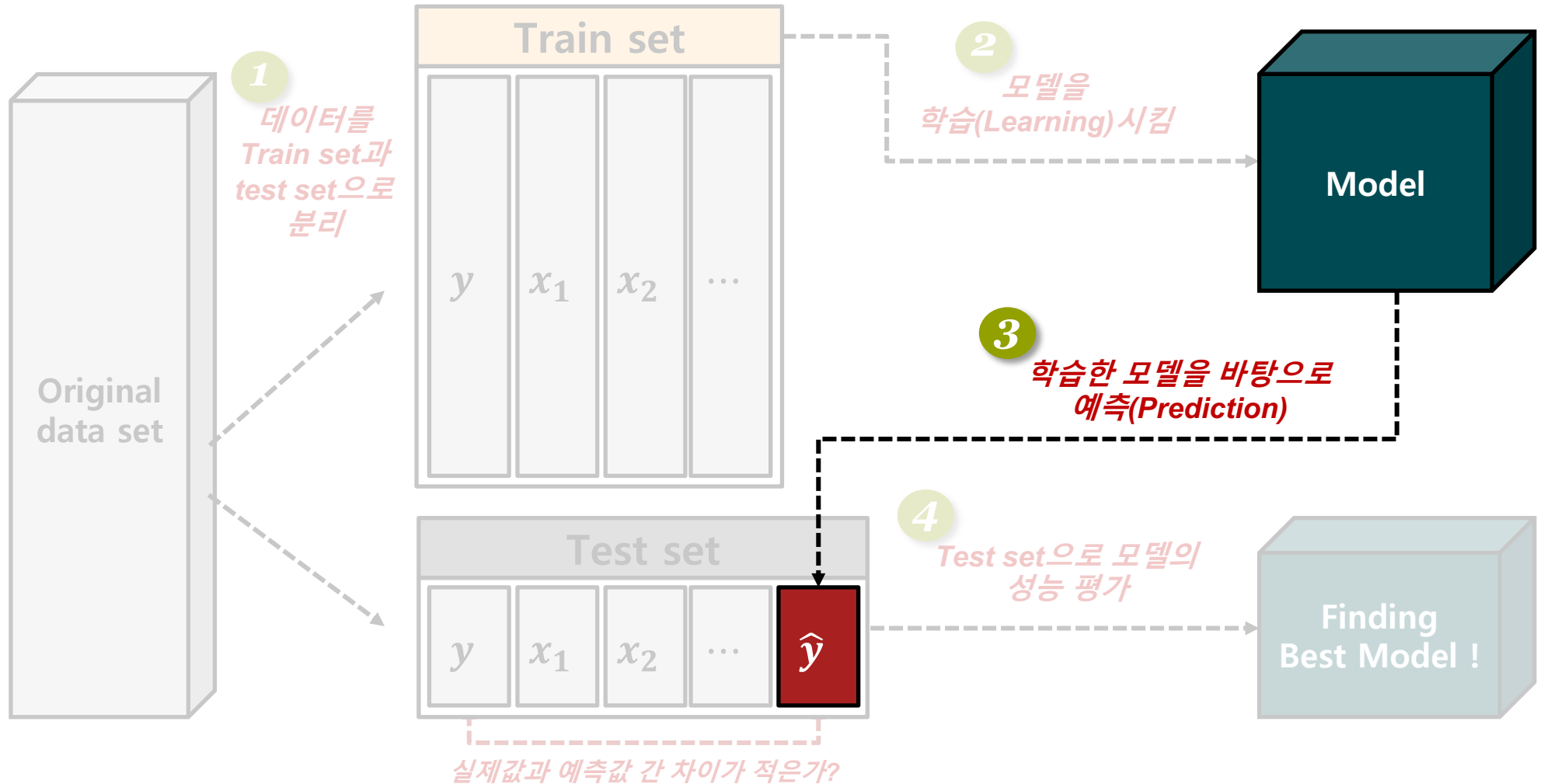
훈련(Training) and 검증(Testing)

예측모형을 도출하기 위해서는 내가 갖고 있는 원래의 자료(Original data set)를 훈련데이터(train set)와 검증데이터(test set)으로 나누고, 이를 지속적으로 학습시켜 최적의 모형을 찾아야 함



훈련(Training) and 검증(Testing)

예측모형을 도출하기 위해서는 내가 갖고 있는 원래의 자료(Original data set)를 훈련데이터(train set)와 검증데이터(test set)으로 나누고, 이를 지속적으로 학습시켜 최적의 모형을 찾아야 함



훈련(Training) and 검증(Testing)

예측모형을 도출하기 위해서는 내가 갖고 있는 원래의 자료(Original data set)를 훈련데이터(train set)와 검증데이터(test set)으로 나누고, 이를 지속적으로 학습시켜 최적의 모형을 찾아야 함

