

Part.01

Machine Learning의 개념과 종류

# | 모형의 적합성 평가 및 실험 설계 (데이터분석과정)

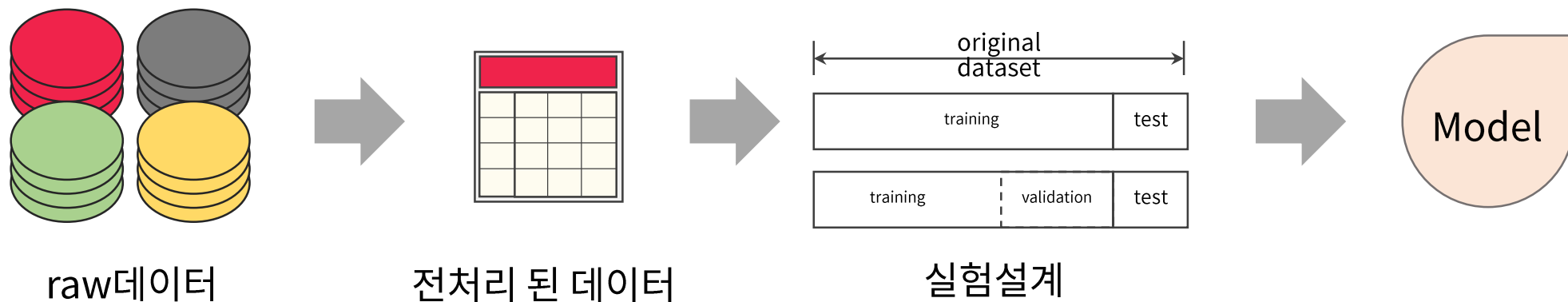
FASTCAMPUS  
ONLINE

머신러닝과 데이터분석 A-Z

강사. 이경택

# I 모형의 적합성 평가 및 실험 설계

## ■ 데이터 분석과정



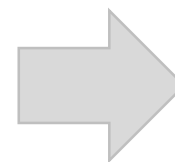
# I 모형의 적합성 평가 및 실험 설계

## ■ 전처리

- Raw데이터를 모델링 할 수 있도록 데이터를 병합 및 파생 변수 생성
- Ex) 게임 유저 이탈 예측 문제

Userid	게임접속일	획득경험치	던전입장수	던전클리어수
1	3	100	10	7
1	5	150	5	3
1	6	10	3	2
1	10	200	2	1
2	4	-200	10	2
2	11	100	5	5

Userid	지불금액
1	10
2	1



Userid	변수1	변수2	이탈여부
1	2	100	이탈
2	5	20	잔존
3	6	10	이탈
4	1	550	이탈
...			

전처리 된 데이터  
(Model에 들어 갈 수 있는 데이터)

# I 모형의 적합성 평가 및 실험 설계

## ■ 전처리

Userid	게임접속일	획득경험치	던전입장수	던전클리어수
1	3	100	10	7
1	5	150	5	3
1	6	10	3	2
1	10	200	2	1
2	4	80	10	2
2	11	100	5	5

Userid	지불금액
1	10
2	1



Userid	게임누적접속일	누적획득경험치	평균획득경험치	누적던전입장수	누적던전클리어수	...
1	4	460	115	20	13	
2	2	180	90	15	7	
...						

# I 모형의 적합성 평가 및 실험 설계

## ■ 전처리

“어떤 변수가 게임의 이탈에 영향을 끼칠까?”

노력한것 대비 성과가 부족 할 때, 투자한 비용 대비 성과가 떨어 질 때

Userid	게임누적접속일	누적획득경험치	평균획득경험치	누적던전입장수	누덕전던클리어수	...
1	4	460	115	20	13	
2	2	180	90	15	7	
		...				

Userid	지불금액
1	10
2	1

비용대비

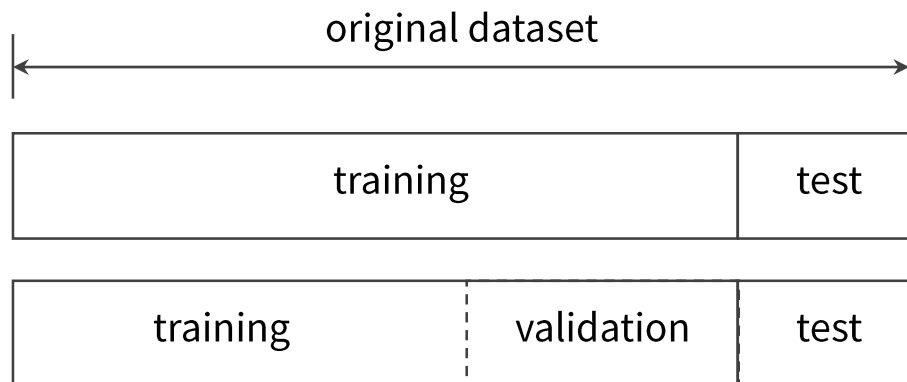
던전클리어비율

Userid	게임누적접속일	비용대비 누적획득경험치	비용대비 평균획득경험치	비용대비 누적던전 클리어비율	...
1	4	46	11.5	0.185	
2	2	180	90	0.466	
		...			

# I 모형의 적합성 평가 및 실험 설계

## ■ 실험설계

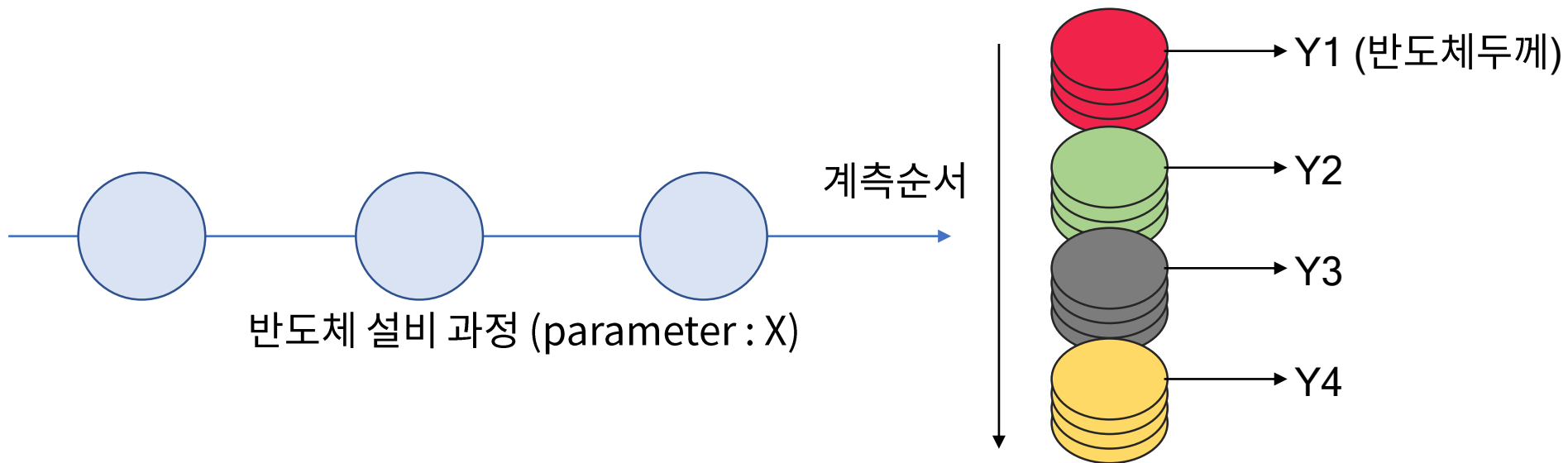
- 실험설계에서 test데이터는 실제로 우리가 모델을 적용을 한다는 가정하여야 함
- Train, validation 데이터에 test정보는 없어야 함



# I 모형의 적합성 평가 및 실험 설계

## ■ 실험설계

Ex) 반도체 설비 parameter를 통해 반도체 두께를 예측하고 싶은 경우



# I 모형의 적합성 평가 및 실험 설계

- 실험설계

Ex) Class imbalance 한 경우

