

# DS 06주차 수업 팀과제

자유주제

팀 AOA

# 00 서론

1. 모델을 생성할 때, 학습데이터 셋의 비율과 테스트 데이터 셋의 비율에 각각 오분류 비용을 곱해서 모델 생성 시 오분류 비용을 반영하는 모델을 만들었다.
2. 교차검증(k-cross validation)을 이용하여 모델의 정확도를 높여 보았다.

# 01

## 오분류 비용 측정

case	Cost	Ratio	result
정분류→오분류	1	70%	70
오분류→정분류	5	30%	150

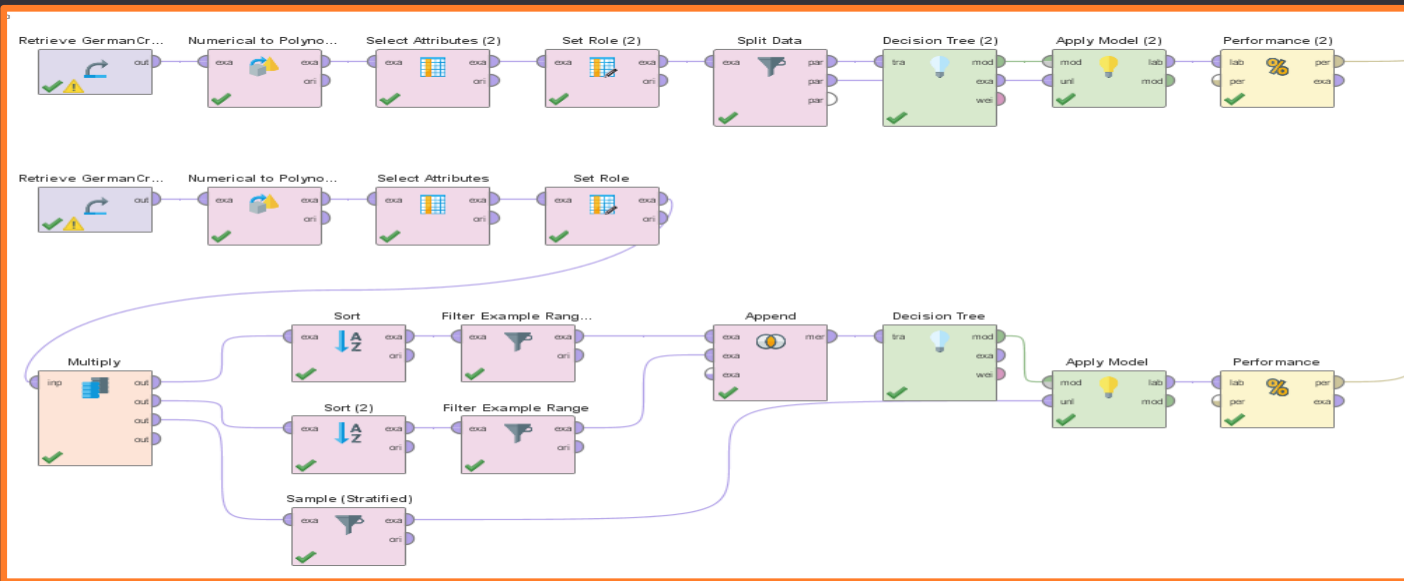
German credit dataset의 오분류 비용을 측정하기 위해, 표를 작성해 보았다.

정분류를 오분류로 착각하는 것을 cost 1로, 오분류를 정분류로 착각하는 비용은 cost 5로 측정하였다.

학습 모델의 정분류 오분류의 비율을 cost와 곱한 값을 result 로 정하고 cost model을 작성해 보았다.

## 02

# 오분류 비용 Process



Result 1은 원래 의사결정나무-error model

Result2는 오분류 비용을 고려한 프로세스-cost model

## 03

## cost model vs error model

accuracy: 70.00%			
	true 1	true 0	class precision
pred. 1	175	55	76.09%
pred. 0	35	35	50.00%
class recall	83.33%	38.89%	

〈Error model: 오분류를 정분류로 착각하는 경우= 55/300〉

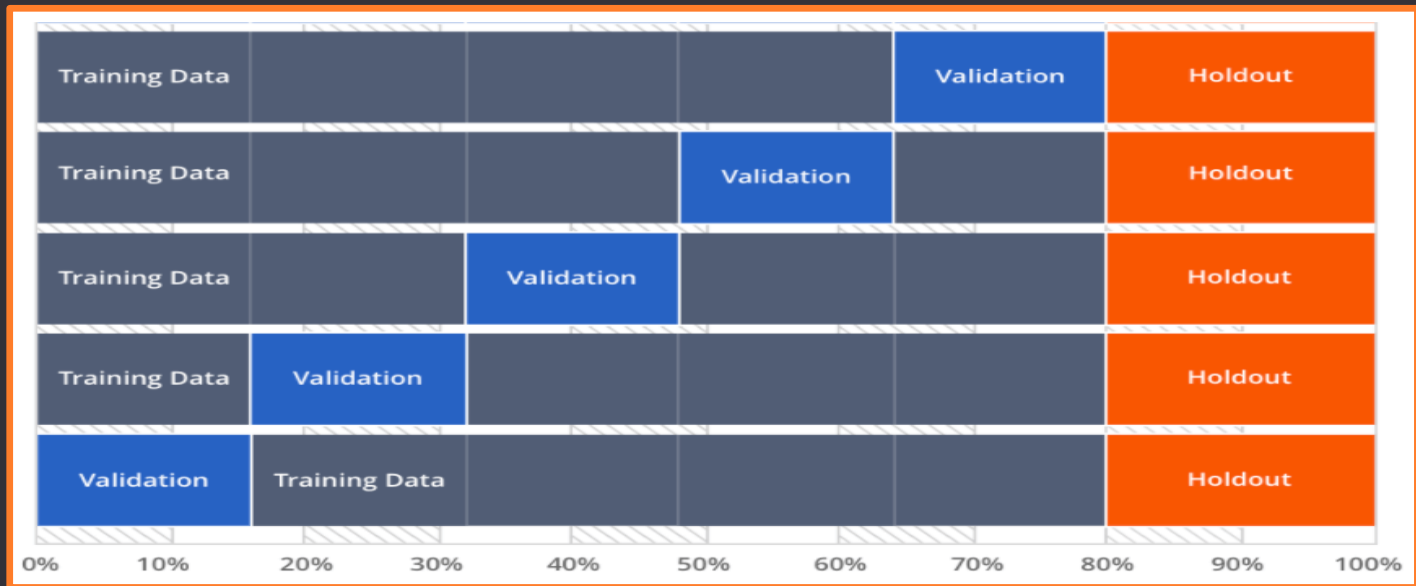
accuracy: 42.86%			
	true 1	true 0	class precision
pred. 1	10	1	90.91%
pred. 0	39	20	33.90%
class recall	20.41%	95.24%	

〈Cost model: 오분류를 정분류로 착각하는 경우= 1/70〉

Cost model은 기존 모델에 비해 정확도는 떨어지지만, 비용을 준만큼 오분류를 정분류로 착각하는 경우는 현저히 낮아졌다.

## 04

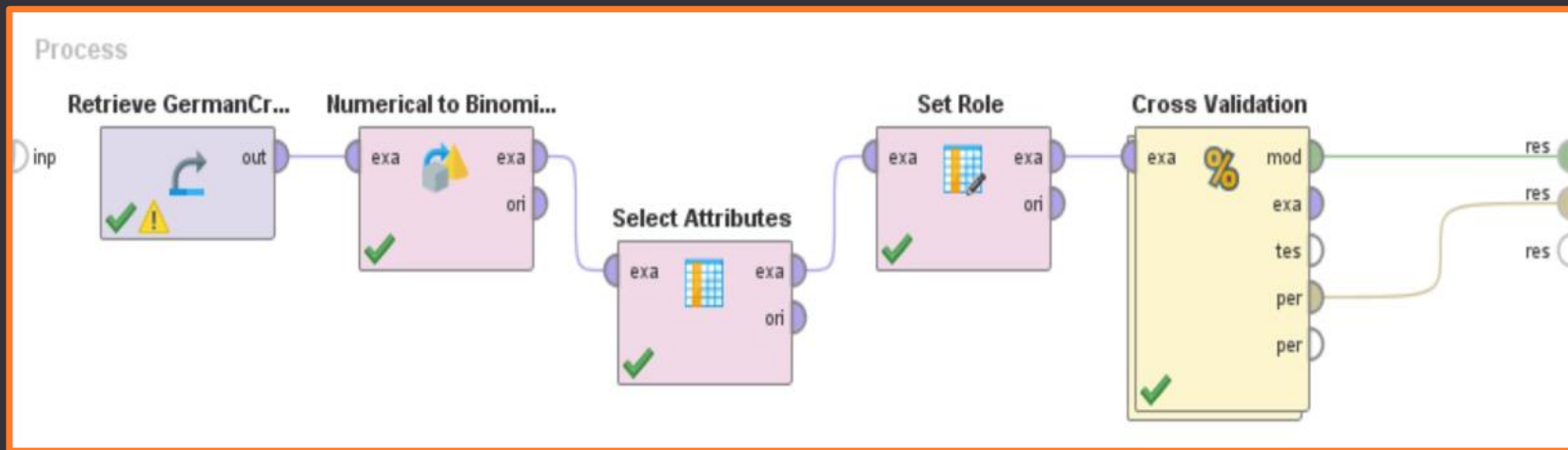
# K-Cross Validation 목적



분류 모델, 모델 생성 알고리즘, 데이터 셋을 일반화 하여 성능을 측정하고,

사용할 수 있는 데이터 양이 적을 경우에 한정된 데이터셋을 여러번 이용할 수 있게 한다.

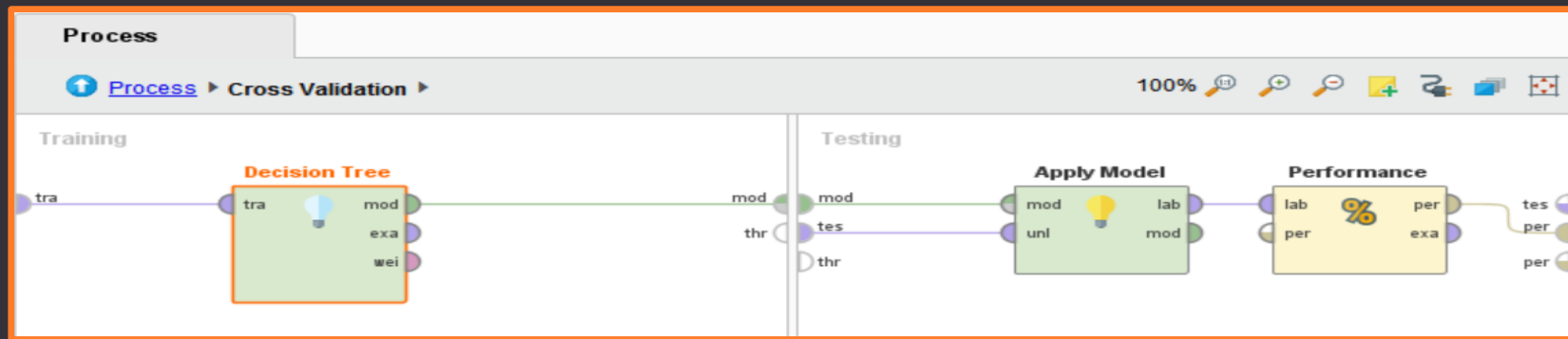
## K-Cross Validation Process



Cross Validation operator를 활용하여 모델의 정확도를 높이는 프로세스를 구현해보았다.

## 06

# K-Cross Validation Process



Cross Validation operator를 클릭하면 위와같은 sub-process가 나온다.

이 부분에서 결정 나무 모델을 만들고, 모델에 적용하며 평균 정확도의 결과를 얻어낼 수 있다.

Parameter 부분에서 몇 개의 부분을 만들 것인지 결정할 수 있는데 오른쪽 위 사진과 같이 number of folds에서 10으로 설정하였다.



## 07 Result

accuracy: 71.10% +/- 2.56% (micro average: 71.10%)

	true false	true true	class precision
pred. false	108	97	52.68%
pred. true	192	603	75.85%
class recall	36.00%	86.14%	

Cross Validaton 수행 결과 정확도가 71.10%가 나왔음을 확인할 수 있다. 또한 그 범위는 +/-2.56을 지닌다.

교차 검증한 결과의 정확도는 교차검증을 하지않은 것보다 신뢰도가 높다.