

디지털미디어랩 머신러닝 여름캠프 4주차

(1) Classification 모델의 평가 – Confusion Matrix

목차

- Titanic호 생존자 분석
- Confusion Matrix 개념과 해석
 - 생성한 분류 모델의 평가는 어떻게 할까?

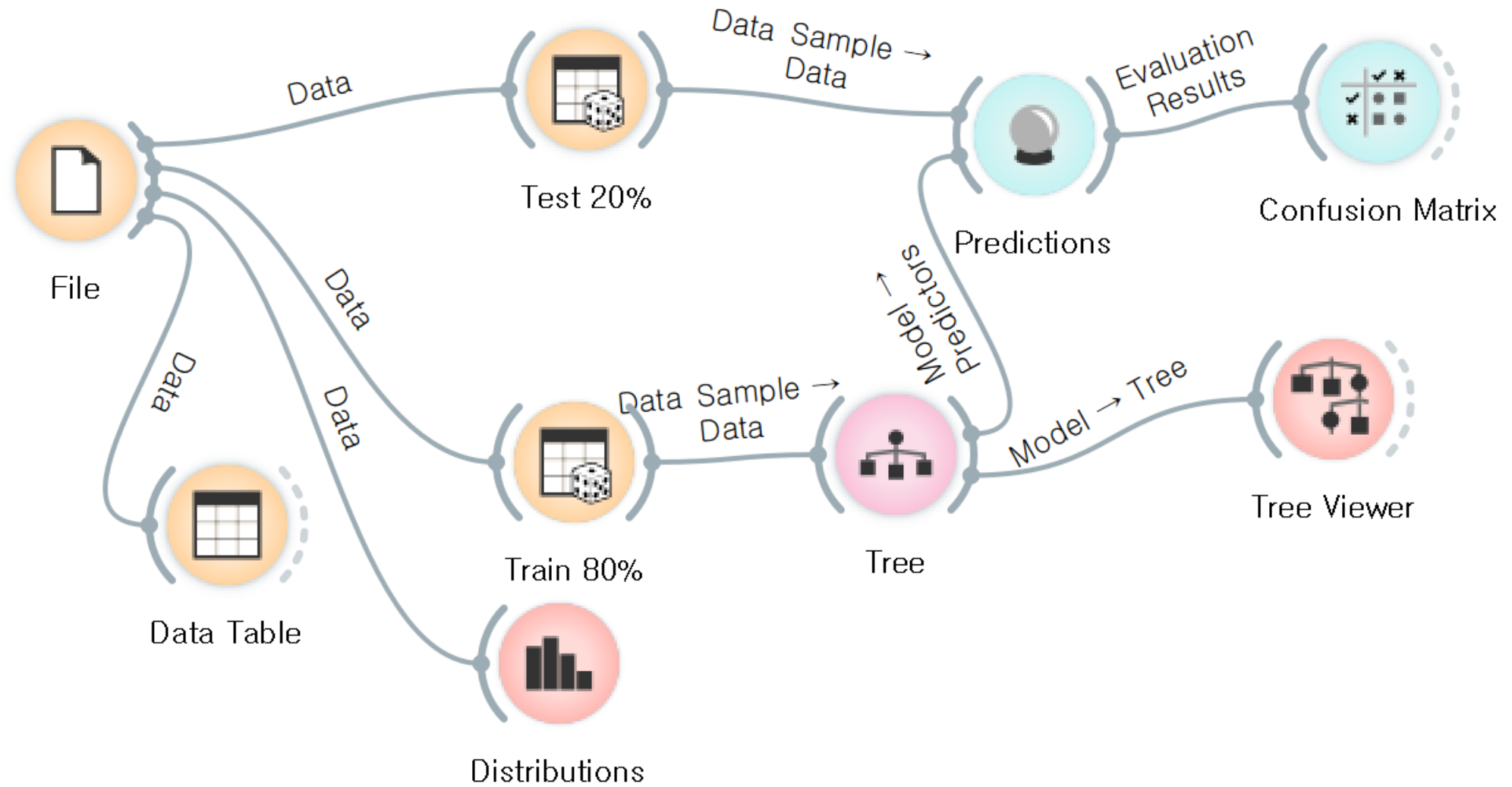
Titanic 호 생존자 분석

	survived	status	age	sex
1	no	third	child	female
2	no	second	adult	male
3	no	second	adult	female
4	no	crew	adult	male
5	no	first	adult	male
6	no	third	adult	female
7	yes	second	adult	female
8	no	crew	adult	male
9	no	third	adult	female
10	no	third	child	male
11	yes	third	adult	female
12	yes	first	adult	female
13	no	crew	adult	male
14	no	first	adult	male
15	yes	third	adult	female
16	no	crew	adult	male
17	no	third	adult	male
18	no	second	adult	male
19	no	third	adult	male

Titanic 호 생존자 분석

분류	탑승자	생존률	사망률	생존자	사망자
1등실, 어린이	6	83%	17%	5	1
2등실, 어린이	24	100%	0%	24	0
3등실, 어린이	79	34%	66%	27	52
1등실, 여성	144	97%	3%	140	4
2등실, 여성	93	86%	14%	80	13
3등실, 여성	165	46%	54%	76	89
1등실, 남성	175	33%	67%	57	118
2등실, 남성	168	8%	92%	14	154
3등실, 남성	462	16%	84%	75	387
승무원, 여성	23	87%	13%	20	3
승무원, 남성 ^[50]	885	22%	78%	192	693
어린이 총합	109	51%	49%	56	53
여성 총합	425	74%	26%	316	109
남성 총합	1690	20%	80%	338	1352
전체 총합	2224	32%	68%	710	1514

Titanic 호 생존자 분석



Confusion Matrix

		Predicted		Σ
		no	yes	
Actual	no	304	3	307
	yes	90	44	134
Σ		394	47	441

Confusion Matrix

		Predicted		Σ
		no	yes	
Actual	no	304	3	307
	yes	90	44	134
Σ		394	47	441

실제 no인 값을 no라고 예측 한 것

Confusion Matrix

		Predicted		Σ
		no	yes	
Actual	no	304	3	307
	yes	90	44	134
Σ		394	47	441

실제 yes인 값을 yes라고 예측 한 것

Confusion Matrix

		Predicted		Σ
		no	yes	
Actual	no	304	3	307
	yes	90	44	134
Σ		394	47	441

실제 yes인 값을 no라고 예측 한 것

Confusion Matrix

		Predicted		Σ
		no	yes	
Actual	no	304	3	307
	yes	90	44	134
Σ		394	47	441

실제 no인 값을 yes라고 예측 한 것

Confusion Matrix

		Predicted		Σ
		no	yes	
Actual	no	304	3	307
	yes	90	44	134
Σ		394	47	441

Accuracy (정확도)

- 전체 데이터 중 정확히 분류 된 것
- $(304 + 44)/441 * 100 = 78.91\%$

Confusion Matrix

		Predicted		Σ
		no	yes	
Actual	no	304	3	307
	yes	90	44	134
Σ		394	47	441

Precision (정밀도)

- No 로 분류 된 것 중 실제로 No인 것 (혹은 그 반대의 경우)
- $304/394 * 100 = 77.16\%$

Confusion Matrix

		Predicted		Σ
		no	yes	
Actual	no	304	3	307
	yes	90	44	134
Σ		394	47	441

Recall (재현율)

- 실제로 No 인 것 중에서 No로 분류된 정도 (혹은 그 반대)
- $304/307 * 100 = 99.02\%$

Confusion Matrix

		Predicted		Σ
		no	yes	
Actual	no	304	3	307
	yes	90	44	134
Σ		394	47	441

- Accuracy : 78.91%
- Precision (No) : 77.16%
- Recall (No) : 99.02%

- 목적에 따라 어느 척도가 더 중요한지 달라 진다.
- Ex) 틀리지 않는 게 더 중요한 경우 등

Multi-Labeled Classification에서의 Confusion Matrix

		Predicted				Σ
		unacc	acc	good	v-good	
Actual	unacc	230	18	0	1	249
	acc	4	66	0	1	71
	good	0	7	0	5	12
	v-good	0	2	0	12	14
Σ		234	93	0	19	346

- Accuracy : $(230+66+0+12)/346 * 100 = 89.02\%$

참고 자료

타이타닉 표 참조

<https://namu.wiki/w/%ED%83%80%EC%9D%B4%ED%83%80%EB%8B%89>