머신러닝 파이프라인

모델분석

WIT What-If-Tool

송호연





목차

- 모델 분석 WIT
 - o 13-1. What-If-Tool 소개
 - 13-2. What-If-Tool 모델 분석 실습



모델 분석 WIT

- O1. What-If-Tool의 특징과 기능을 이해한다.
 - What-If-Tool이 어떤 특징과 기능을 갖고 있는 지 이해한다.
 - 02. What-If-Tool 모델 분석 실습을 통해 사용법을 익힌다.

What-If-Tool을 활용하여 직접 모델을 분석하는 실습을 진행한다.

What-If-Tool

탭

작업공간

모듈, 플레이 그라운드





모델 분석이 필요한 이유

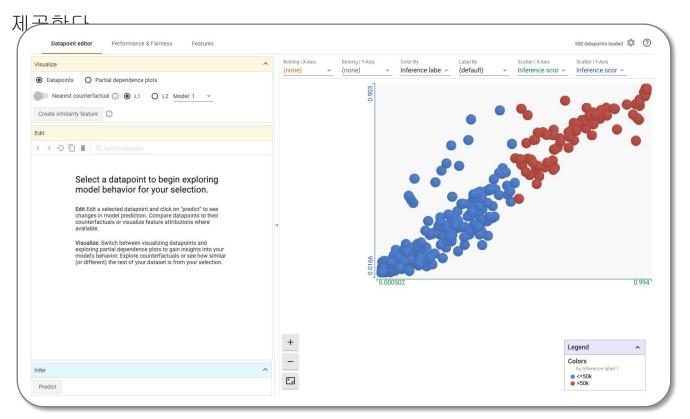
전통적인 소프트웨어 개발과 머신러닝 소프트웨어 개발 방식의 차이

	전통적인 소프트웨어 개발	머신러닝 소프트웨어 개발
개발 접근법	개발자는 명시적으로 컴퓨터에게 해야 할 일을 알려줘야 합니다.	개발자는 데이터를 통해서 특정 작업을 수행하도록 알고리즘을 학습시킵니다.
데이터 요구사항	데이터가 거의 필요하지 않습니다. 인간이 직접 데이터로부터 논리를 만들어냅니다.	프로젝트의 성공이 데이터의 품질과 양에 크게 의존하고 있습니다.
적응력	환경이 변했을 때 소프트웨어 디자인을 다시 해야 합니다.	새로운 환경이 오더라도 새로운 데이터를 보여주면 적응합니다.



What-If-Tool 소개

What-If-Tool 도구는 훈련 된 ML 모델의 동작을 분석하는 시각화 기반 도구다. What-If-Tool 도구를 시작할 때 예상 할 수 있는 사항에 대한 빠른 안내를





탭

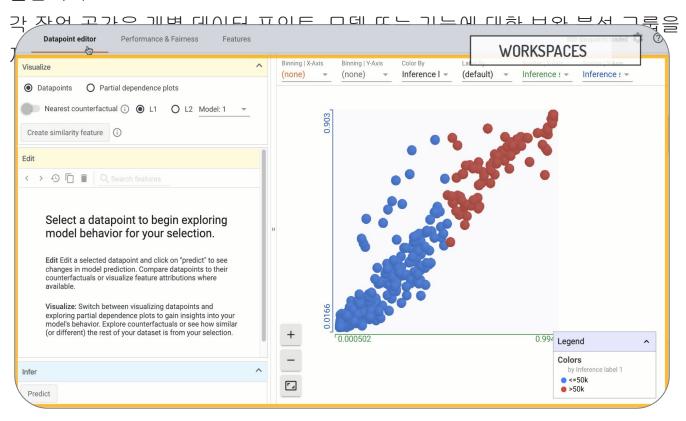
What-If Tool의 인터페이스는 What-If Tool에 로드 된 모델의 종류에 따라세 개의 탭으로 구성됩니다. 탭은 인터페이스의 오른쪽 상단에 있는 What-If 도구의 여러 영역으로 이동하는 데 사용할 수 있는 대화 형요소입니다.

- Datapointer Editor : 기본적으로 What-If 도구는 항상이 탭에서 열립니다.
- Performance : 멀티 클래스와 회귀 모델에 대한 성능 및 공정성 이진 분류 모델
- Feature



▶ 작업 공간

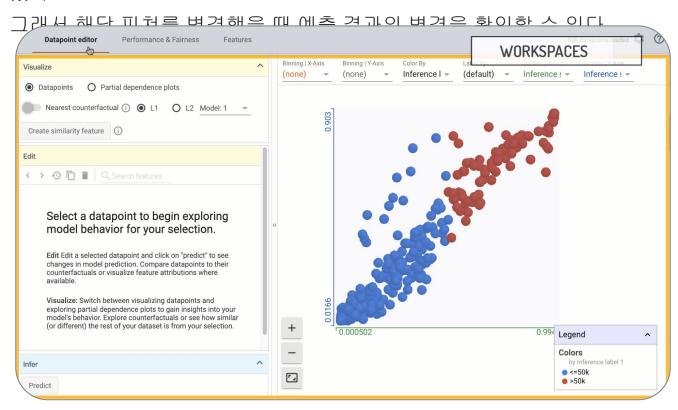
탭을 클릭하면 다른 모델 이해 작업을 수행 할 수 있는 환경 인 해당 작업 공간이 열립니다.





작업 공간 - Datapoint Editor

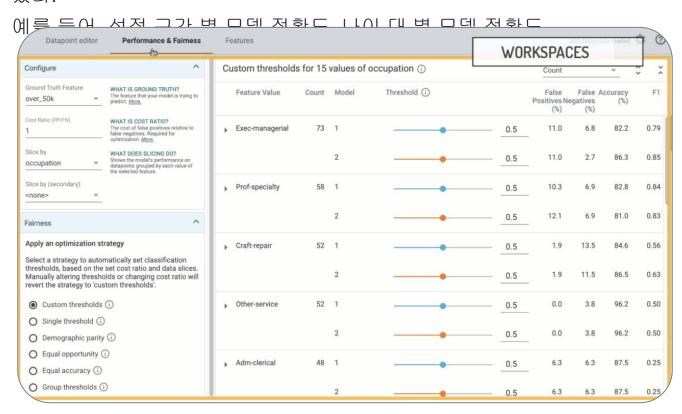
Datapoint Editor : 각각의 데이터 포인트를 우측 시각화 화면에 적용될 수 있다.





작업 공간 - Performance

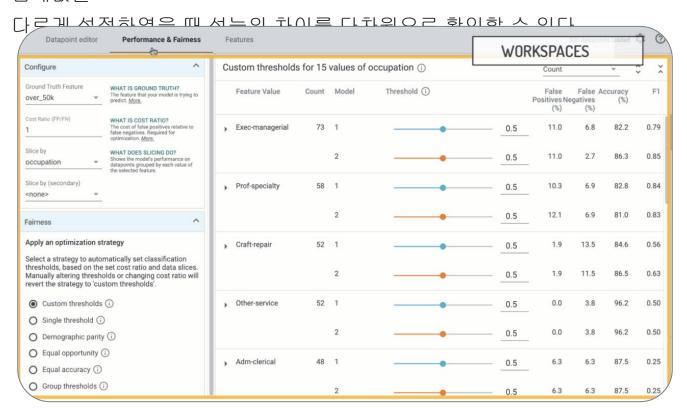
Performance : 각 피쳐를 슬라이스 한 모델의 성능을 확인할 수 있다.





작업 공간 - Performance and Fairness

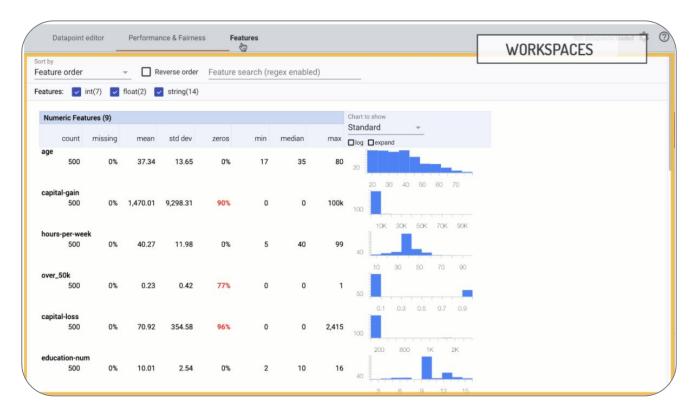
Performance and Fairness : 이진 분류 모델의 경우 각각의 피쳐에 대한 임계값을





작업 공간 - Feature

Feature : 높은 수준에서 피쳐의 분포를 확인할 수 있다. 데이터셋의 특성을 확인할 때 유용함.



데이터 세트 및 모델

노트북의 WIT

간단한 시각적 분석

가장 가까운 Counterfactuals 탐색

비용 비율 및 결정 임계 값 최적화





데이터 세트 및 모델

UCI 인구 조사 데이터 세트가 각 레코드는 사람을 표현하는 데이터 집합이다. 각 기록에는 1994 년 미국 인구 조사 데이터베이스에서 한 사람에 대한 14 개의 인구 조사 정보가 포함되어 있다.

여기에는 연령, 결혼 상태 및 교육 수준과 같은 정보가 포함된다.

예측 작업은 사람이 고소득 (연간 \$ 50ka 이상 벌어 들인 것으로 정의 됨)인지 확인하는 것이다.



🔘 노트북의 WIT

노트북에서 다음 명령어를 통해 WitWidget을 생성한다.

0

from witwidget.notebook.visualization from WitConfigBuilder from witwidget.notebook.visualization from WitWidget

노트북에서 WIT는 Cloud AI Platform Prediction, set_ai_platform_model메서드를 통해 또는 set_custom_predict_fn 메서드를 통해 Python에서 쿼리 할 수 있는 모든 모델을 통해

제공되는 모델에서도 사용할 수 있다.

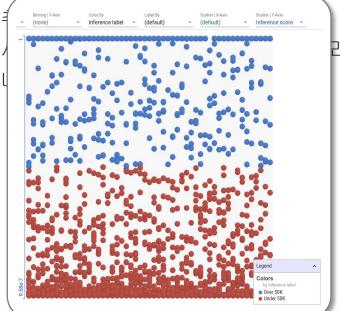


데이터 훑어보기

What-If-Tool을 모델 및 데이터 세트에 지정하면 첫 번째로 보이는 것은 Facets Dive 에서

개별 포인트로 시각화 된 데이터 세트다. 각 데이터 포인트는 모델이 예측 한 카테고리

(예 : 추론 레이블)에 따라 색상이 지정된다. 파란색 포인트는 모델이 고소득층이라고



민델이 저소득층이라고 추론 한 사람들을

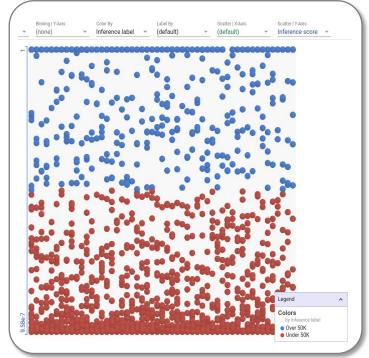


초기 데이터에서 확인할 수 있는 인사이트

파란색 점보다 빨간색 점이 더 많으며, 이는 모델이 고소득보다 저소득으로 더 많은 사람들을

예측 한다는 것을 의미한다. 시각화의 맨 아래와 맨 위에 많은 포인트가 모여 있다.

즉, 이 모델은 종종 사람이 저소득 또는 고소득임을 매우 확신한다.



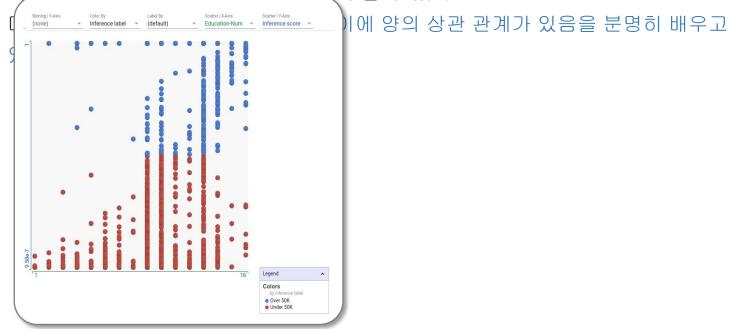


간단한 시각적 분석

이 차트에 데이터를 더 추가해 보겠다. x 축 분산을 교육 수준과 같은 데이터 세트의 기능으로

설정할 수 있다. 우리는 교육 수준이 증가함에 따라 (그림에서 오른쪽으로 이동함에 따라)

파라색 전이 스가 증가하다는 것을 즉시 알 수 있다.

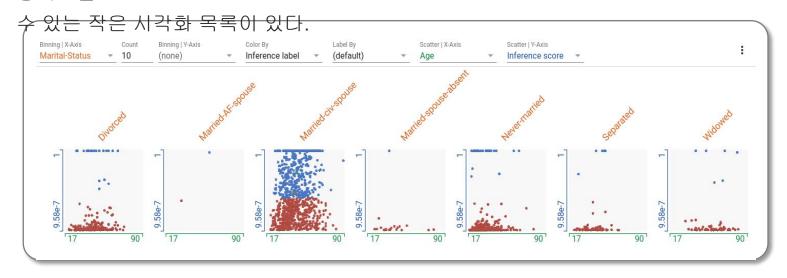




🔘 간단한 시각적 분석

Facets Dive는 매우 유연하며 데이터 포인트를 버킷 화, 분산 및 색상 지정하는 기능을 통해

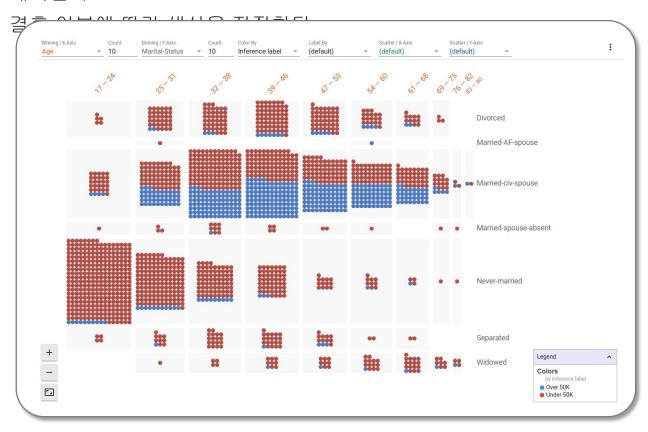
여러 흥미로운 시각화를 만들 수 있다. 가능성은 무궁무진하지만 여기에 이 모델에서 흥미로울





히스토그램 | 표시된 혼인 상태로 근무한 시간

x 축의 비닝을 주당 시간으로 설정한다. "(기본값)"을 사용하여 산점도 설정을 제거한다.





가장 가까운 Counterfactuals 탐색

사람에 대한 변경으로 인해 분류가 어떻게 변경

되는지 확인하는 또 다른 방법은 선택한 데이터

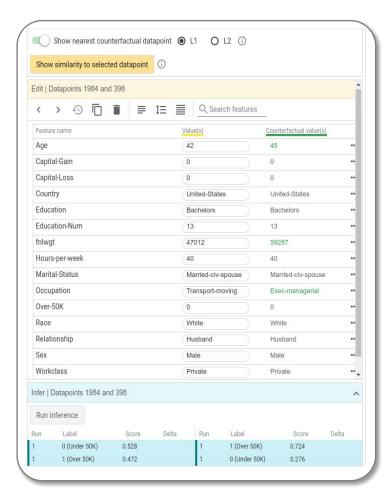
포인트에 가장 가까운 Counterfactual을 찾는 것이다.

가장 가까운 Counterfactual은 다른 추론결과

또는 다른 분류를 가진 가장 유사한 데이터 포인트

이다. WIT는 데이터 포인트 간의 유사성을 계산

하는 두 가지 방법 인 L1, L2 거리 중 하나를 사용하여 가장 가까운 반 사실을 찾을 수 있다.





가장 가까운 Counterfactuals 탐색

또한 "Show similarity to selected datapoint" 버튼을 통해 데이터 세트의 모든 포인트가 선택한 데이터 포인트와 얼마나 유사한 지 확인할 수 있다.

여기서 도구는 선택한 데이터 포인트에서 로드 된 다른 모든 데이터 포인트까지의 거리를 계산하고

해당 거리를 데이터 포인트 시각화에 사용한다. 선택한 데이터 포인트까지의 L1 거리를

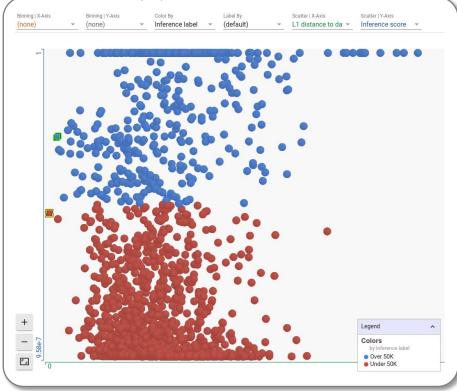
Show similarity to selected da	tapoint	
Metric name		
L1 distance to datapoint 1984		
Dictance type		
Distance type		
Apply to datapoints visuali		
X-Axis Scatter -		
		Apply



가장 가까운 Counterfactuals 탐색

노란색으로 강조 표시된 선택한 데이터 포인트와 가장 유사하므로 가장 왼쪽에 있는 데이터 포인트가

Counterfactual이다.





가장 가까운 Counterfactuals 탐색

부분 종속성 플롯을 사용하면 데이터 포인트의 변경이 모델의 예측에 어떻게 영향을 미치는지

탐구하는 원칙적인 접근 방식을 사용할 수 있다. 각 부분 종속성 그림은 데이터 포인트에서

단일 특성이 조정될 때 모델의 양성 분류 점수가 어떻게 변하는 지 보여준다.

오른쪽 컨트롤에서 "부분 의존성 플롯"버튼을 클릭하면 이 사람의 나이가 최소 17세에서 최대

72 세로 변경되면 선택한 데이터 포인트에 대한 플롯이 즉시 표시된다.



부분 종속성 플롯 탐색

모델은 나이와 소득 사이에 양의 상관 관계를 배운 것으로 보이며, 이는 사람들이나이가

들어감에 따라 더 많은 돈을 버는 경향이 있기 때문이다. 물론,이 가정은 사람들이 나중에

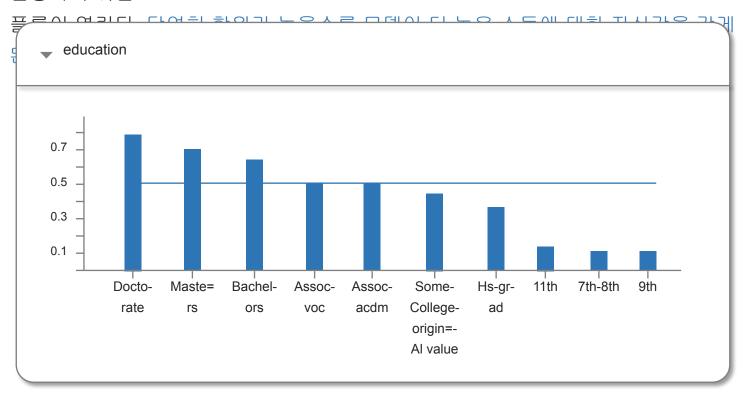
Partial Dependence Plots ① Global partial dependence plots

Set range of values to test
19 - 59



부분 종속성 플롯 탐색

부분 의존도 영역에서 교육 헤더를 클릭하면 범주 형 (숫자가 아닌) 교육 기능을 변경하기 위한





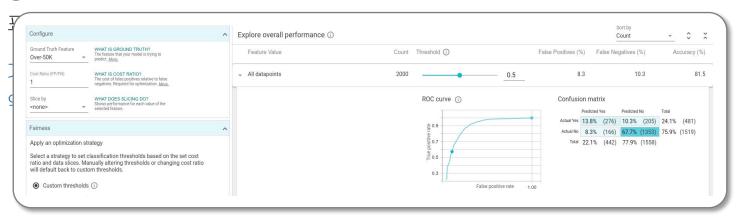
모델 성능 분석

이제 모델에 대한 양의 분류 임계 값 슬라이더, 혼동 행렬 및 ROC 곡선을 볼 수 있습니다.

혼동 행렬은 현재 설정된 분류 임계 값에 대해 데이터 세트에 대해 모델이 예측 한참 양성,

참 음성, 거짓 양성 및 거짓 음성 수를 보여준다. ROC 곡선은 긍정 분류 임계 값의 가능한

모든 설정에 대한 참 양성률과 거짓 양성률을 보여 주며 현재 임계 값은 곡선에서 강조





○ 모델 성능 분석

이 탭에는 "비용 비율"과 "최적화 임계 값"버튼에 대한 설정이 있다. WIT는 로드 된데이터

세트와 가양 성 대가 음성의 상대적 비용을 지정하는 비용 비율이 주어지면 분류 임계 값을

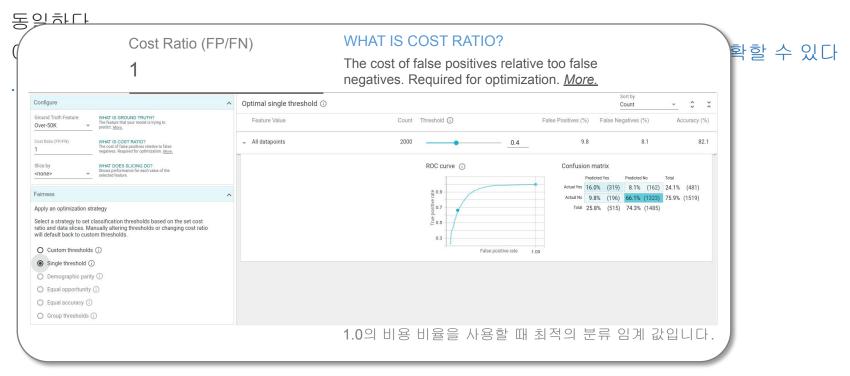
<u> 치적이 지적으로 자동 석정학 수 있다. 이 비용은 사용자가 스스로 격정해야하다</u> Sort by Explore overall performance ① Count Count Threshold (i) Feature Value False Positives (%) False Negatives (%) Accuracy (%) 14.4 5.3 80.3 All datapoints 2000 0.25 ROC curve (i) Confusion matrix Predicted Yes Predicted No Actual Yes 18.7% (374) 5.3% (107) 24.1% (481) e.0 g Actual No 14.4% (288) 61.6% (1231) 75.9% (1519) 0.7 Total 33.1% (662) 66.9% (1338) E 0.5 0.3 False positive rate



비용 비율 및 결정 임계 값 최적화

기본 비용 비율이 1 인 경우 "최적화 임계 값"을 클릭하면 포지티브 분류 임계 값이 0.4로 변경된다.

비용 비율 1로 최적화하는 것은 오 탐지 및 오 탐지의 총 수를 최소화하므로 정확성 최적화와

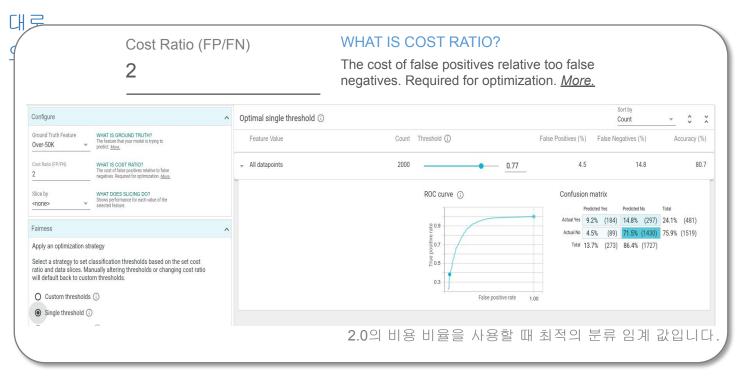




비용 비율 및 결정 임계 값 최적화

비용 비율을 2로 변경하고 임계 값 최적화 버튼을 클릭하면 최적 임계 값이 0.77까지 올라간다.

이 임계 값의 전체 정확도는 이전 임계 값보다 낮지 만이 비용 비율 설정에서 원하는





○ 피쳐 개요

피쳐 탭에서 데이터 세트의 각 피쳐에 대한 값 분포를 볼 수 있다. **5,000** 개의 테스트

데이터 포인트 중 3,300 개 이상이 남성이고 4,200 개가 넘는 데이터 포인트가인과

관계임을 알 수 있다. 이 데이터 세트에서 여성과 소수 민족은 과소 표현 된 것





실습

https://link.chris-chris.ai/ai-lecture-13

https://colab.research.google.com/github/PAIR-code/what-if-tool/blob/master/WIT_Smile_Detector.ipynb#scrollTo=A1s1_SiOyS0I



짚어보기

- 모델 분석 WIT
- O1. What-If-Tool의 특징과 기능을 이해한다.
 What-If-Tool이 어떤 특징과 기능을 갖고 있는 지 이해한다.
 - 02. What-If-Tool 모델 분석 실습을 통해 사용법을 익힌다.
 What-If-Tool을 활용하여 직접 모델을 분석하는 실습을 진행한다.

머신러닝 파이프라인

모델분석

WIT What-If-Tool

송호연



머신러닝 파이프라인

감사합니다.

THANKS FOR WATCHING

