Microsoft

MLOps 101

Episode 4: ML 생애주기 (3) 모델 해석

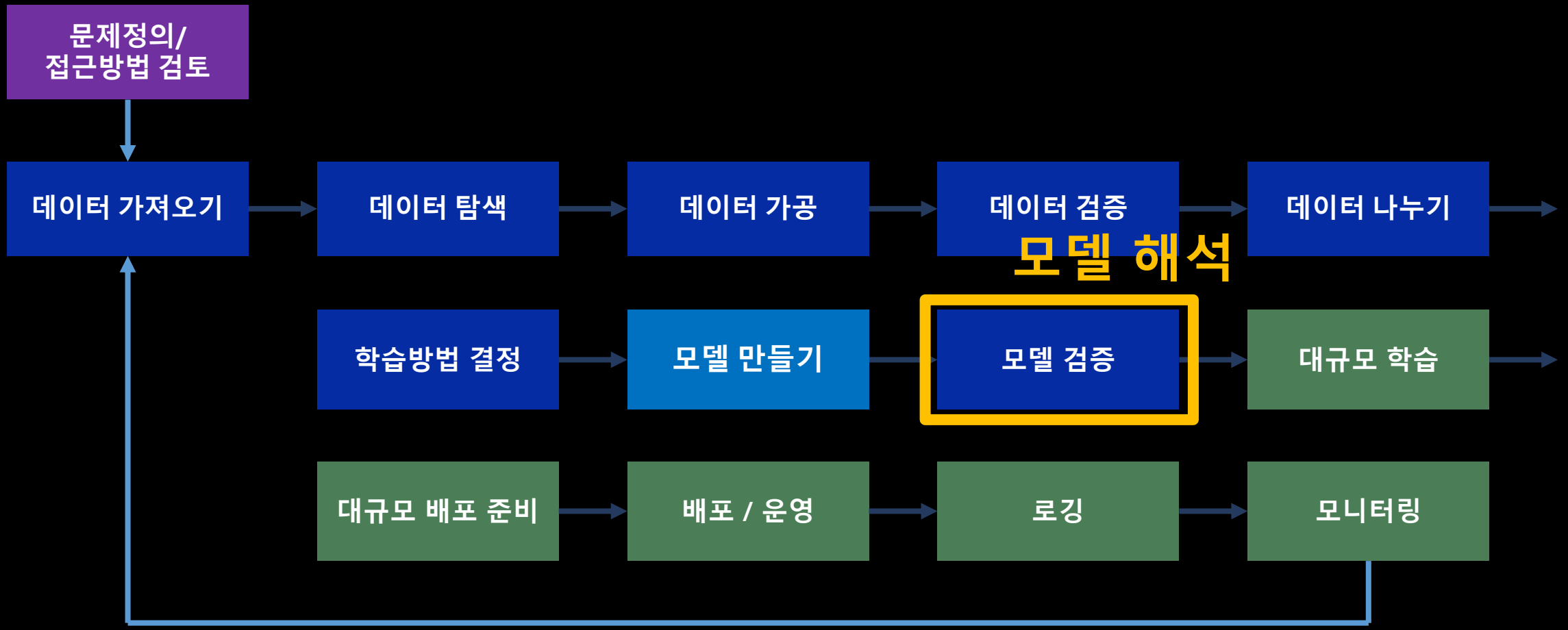
한석진
마이크로소프트

Episode 4
ML 생애주기 (3)
모델 해석

ML 생애주기 (3) 모델 해석

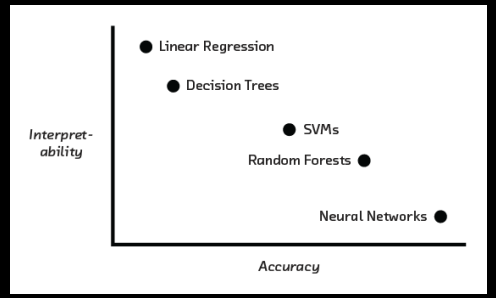
- 모델 해석이 왜 중요한가
- 모형을 해석하려는 시도
- azureml.interpret 들여다보기
- Explainer 관련 시각화 예시
  - 애저머신러닝에서 모델 해석 *DEMO*

ML 생애주기

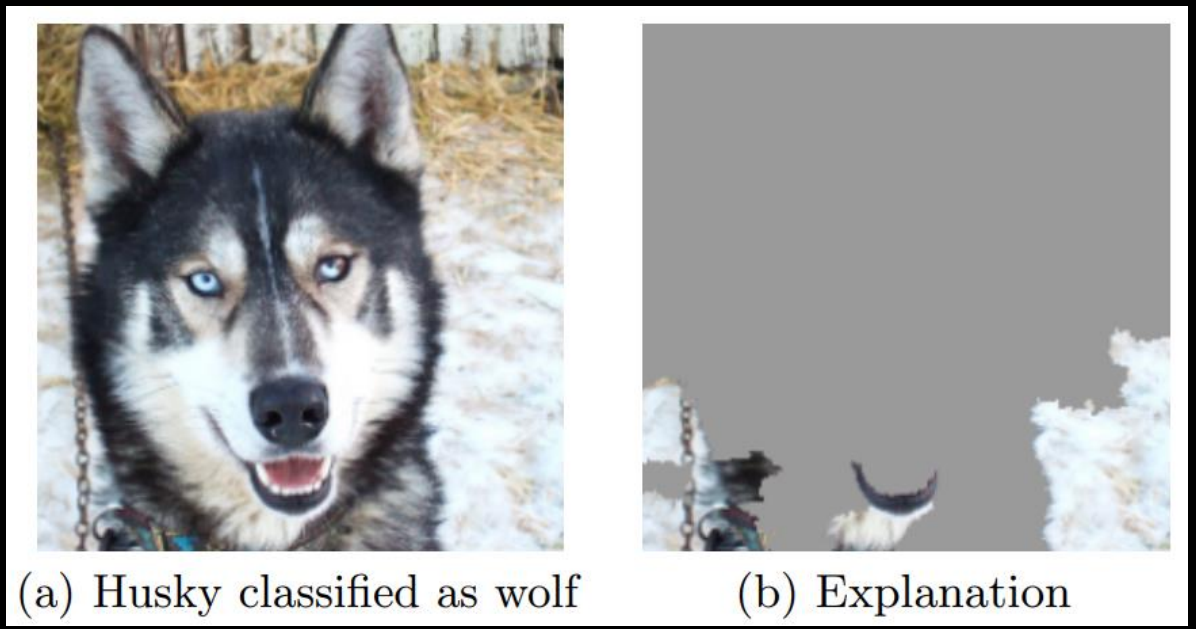

문제정의/
접근방법 검토
데이터 가져오기
데이터 탐색
데이터 가공
데이터 검증
데이터 나누기
모델 해석
학습방법 결정
모델 만들기
모델 검증
대규모 학습
대규모 배포 준비
배포 / 운영
로깅
모니터링

모델 해석이 왜 중요한가?

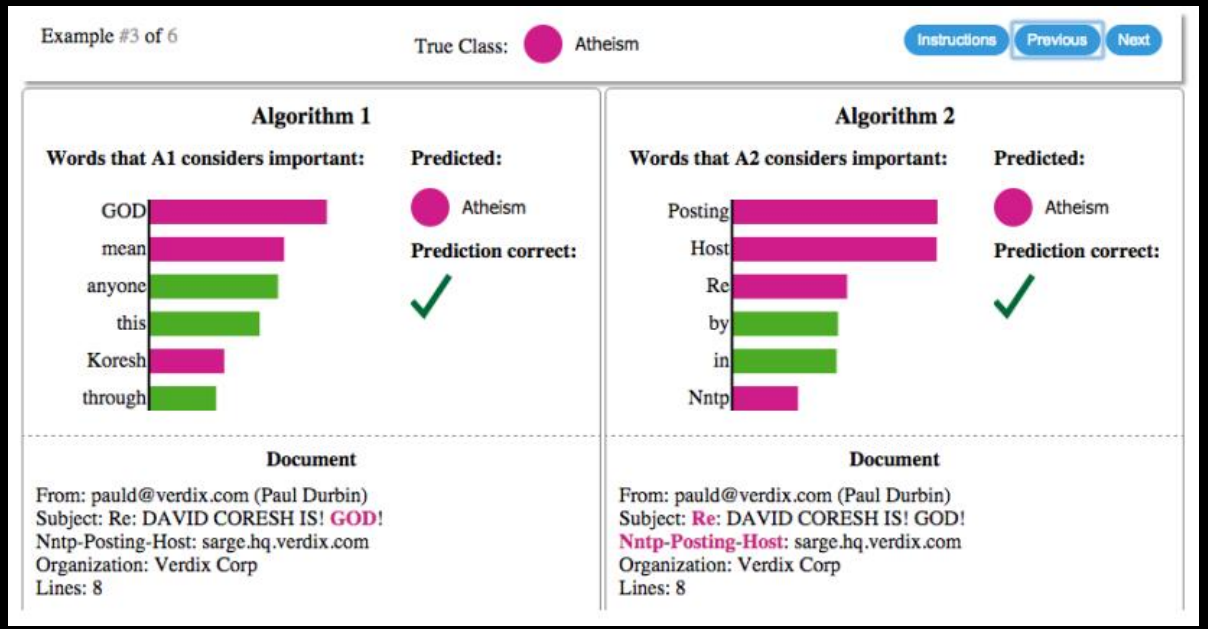
- "단순하게 설명할 수 없다면, 충분히 제대로 이해한 것이 아니다" — Albert Einstein
- 정확도만으로는 더 이상 충분하지 않다
- 내 모델이 예측을 잘 하는 것 같지만 사실은 완전 착각하고 있지는 않는가



해석이 더 어렵다고 포기?



(a) Husky classified as wolf (b) Explanation



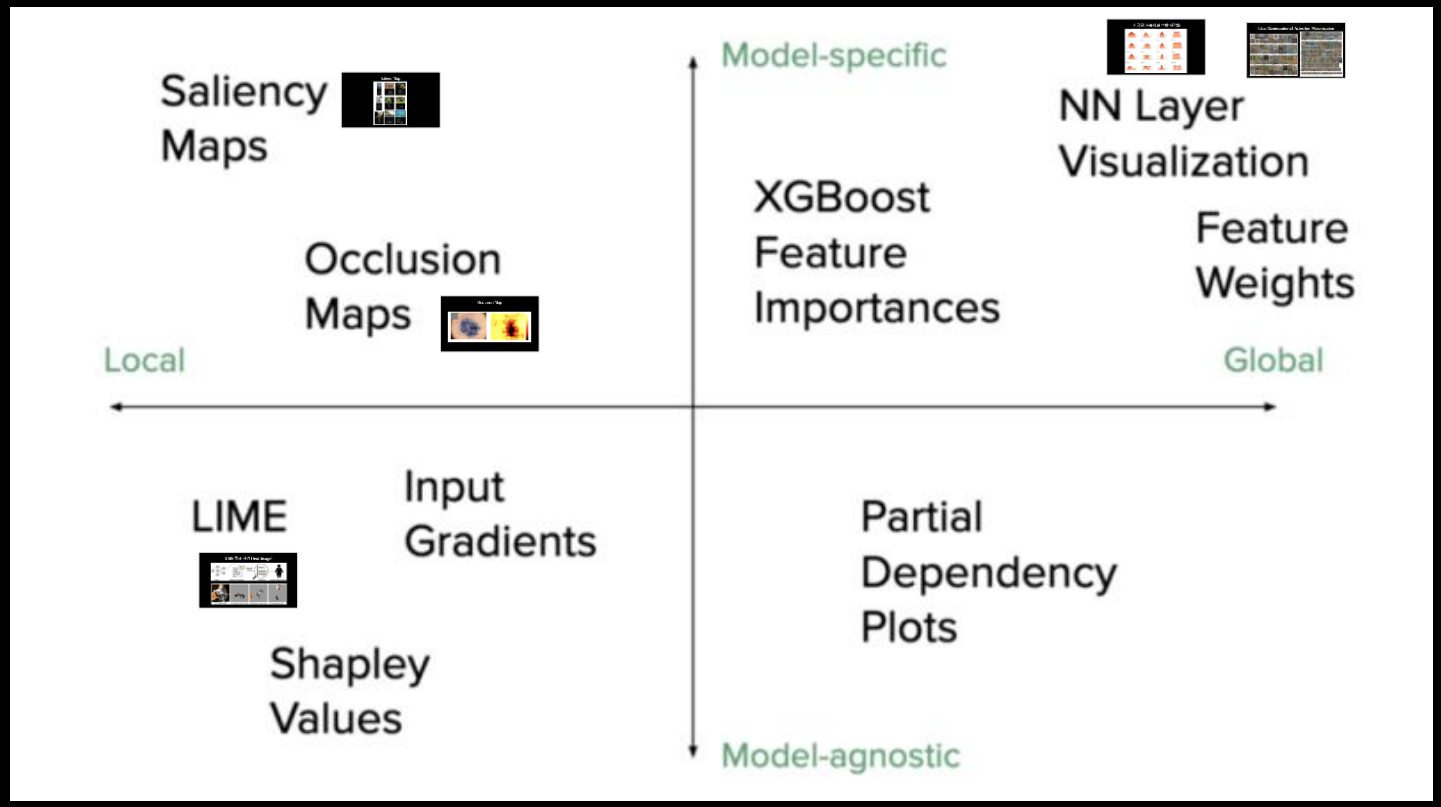
출처: https://arxiv.org/abs/1602.04938, "Why Should I Trust You?": Explaining the Predictions of Any Classifier

모델 해석이 왜 중요한가?

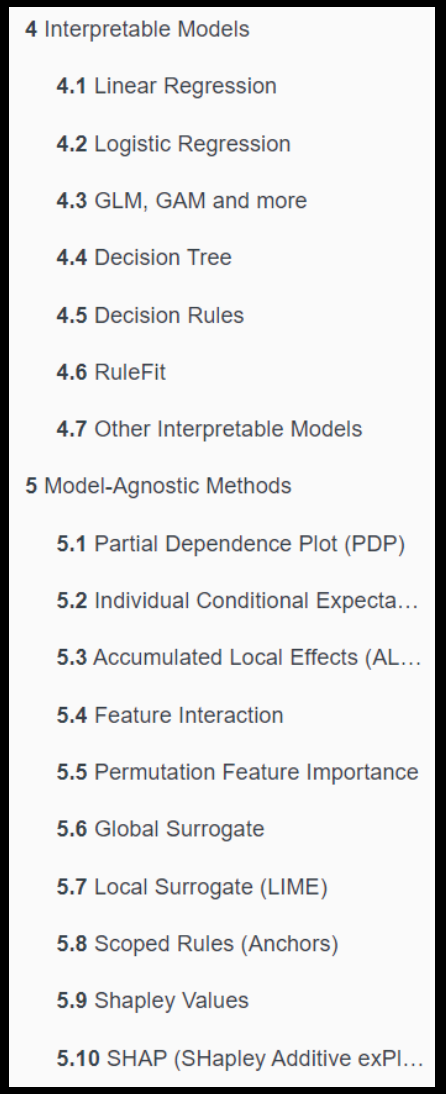
모델 디버깅	왜 내 모델이 잘못 예측했나?
모델 공정성 판단	내 모델이 알게 모르게 차별을 하고 있지는 않나?
사람과 AI의 협력	나는 어떻게 모델의 예측을 이해하고 믿을 수 있나?
규제 및 컴플라이언스	내 모델이 법적인 요건을 만족하는가?
高리스크 영역	의료, 금융, 법률 등

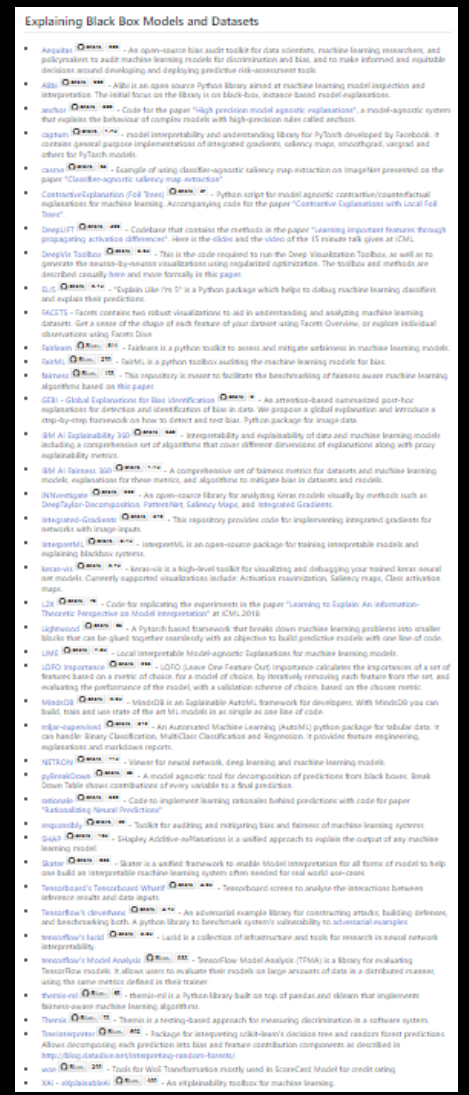
출처: https://github.com/interpretml/interpret

모델을 해석하려는 시도



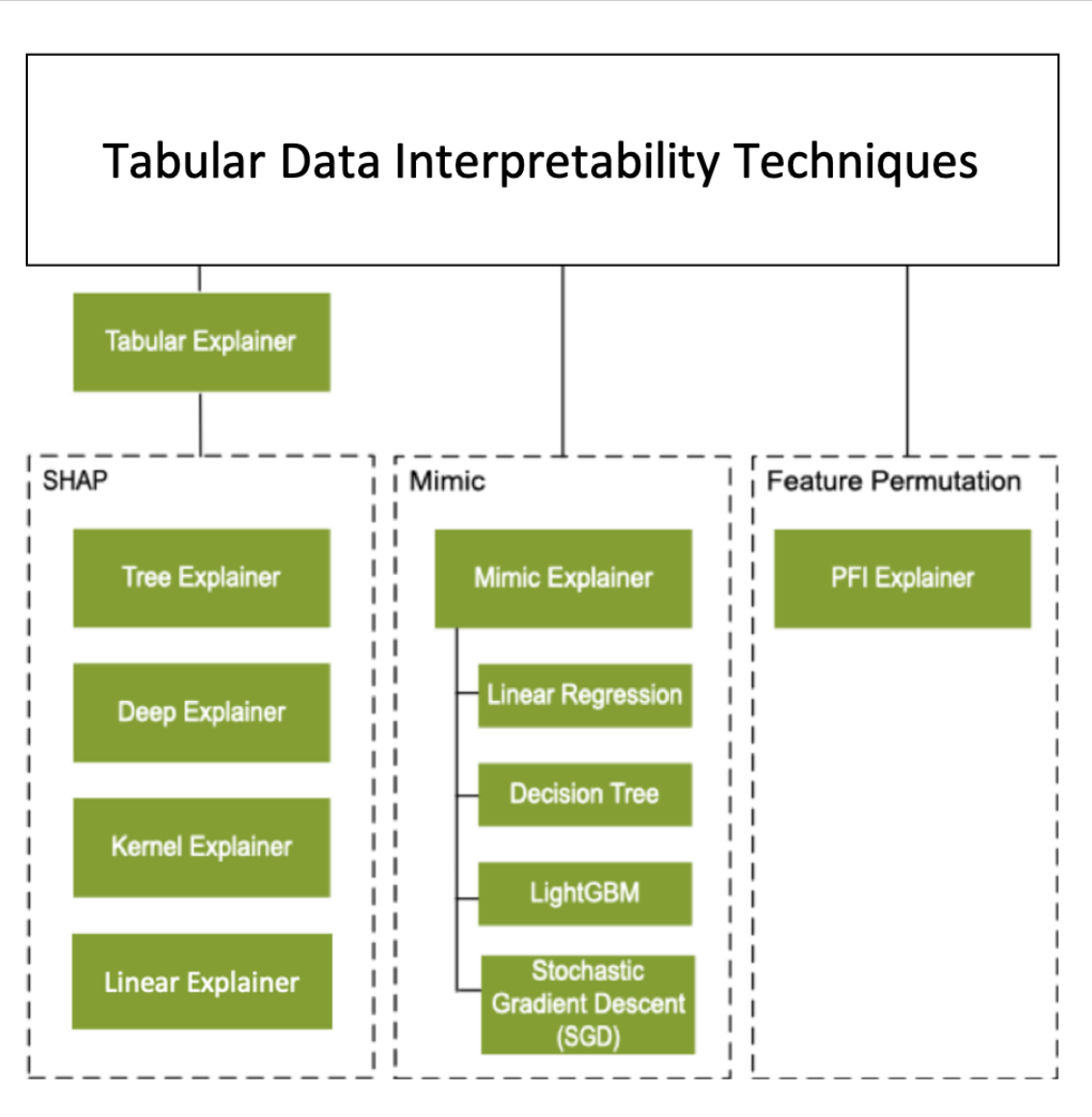
X축	**Local**	일부를 뜯어보기 (신경망의 개별 필터 등)
	Global	전체를 보기 (신경망의 가중치 분포나 전파를 시각화)
Y축	**Model-agnostic**	다수의 모델 유형에 적용 가능
	Model-specific	특정 유형의 모델에만 적합





출처: https://github.com/Harvard-IACS/2020-ComputeFest/, https://christophm.github.io/interpretable-ml-book/, https://github.com/EthicalML/awesome-production-machine-learning#explaining-black-box-models-and-datasets

azureml.interpret 들여다보기

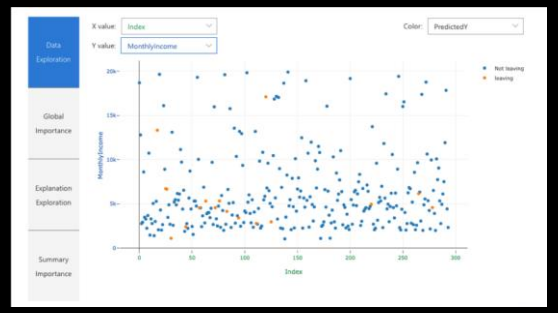


Explainer 종류	적용 모델	접근 방법
SHAP Tree Explainer	Tree, Forest 계열	Polynomial time fast SHAP value estimation
SHAP Deep Explainer	TensorFlow (+ Keras), PyTorch	딥러닝에서 SHAP value 추정을 DeepLIFT 기법과 연계하여 빠르게 수행
SHAP Linear Explainer	선형 모델	선형 모델을 위한 SHAP value 계산 (inter-feature correlation을 고려)
SHAP Kernel Explainer	임의의 모델	Local 선형 회귀모델에 특정 가중치를 적용하여 SHAP value 추정
Mimic Explainer	임의의 모델	LightGBM, Linear Regression, SGD, Decision Tree 등 설명하기 쉬운 전역 대체 모델(global surrogate model)을 추정하여 원 모델 대신 해석
Permutation Feature Importance Explainer	임의의 모델	Breiman's Random Forests paper (section 10)에 기반하여, 특정 feature를 변경시 최종 모델 성능에 미치는 영향을 측정하여, 임의 모델에서 성능에 중요한 feature를 추출
Tabular Explainer	임의의 모델	SHAP Explainer들에 대한 wrapper로서, 모델의 형태에 따라서 Tree, Deep, Linear, Kernel Explainer 중 선택하여 수행함 (예를 들어 Tree 모델이면 무조건 Tree Explainer를 수행하고, Tree가 아닌 딥러닝 모델이면 Deep Explainer를 실행)

DEMO: 애저머신러닝에서 모형 해석

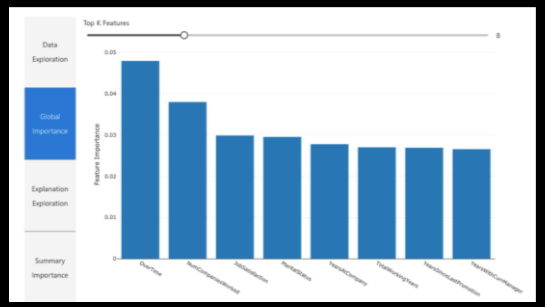
Explainer 관련 시각화 예시

데이터 탐색



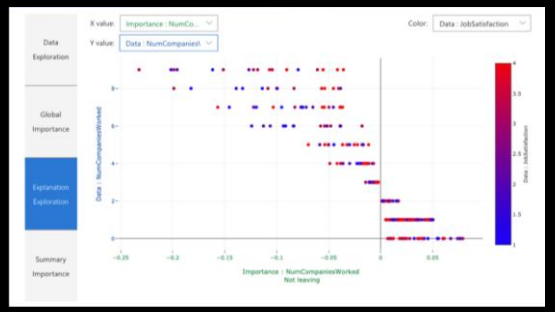
데이터셋을 예측값과 함께 표시

Global Feature Importance



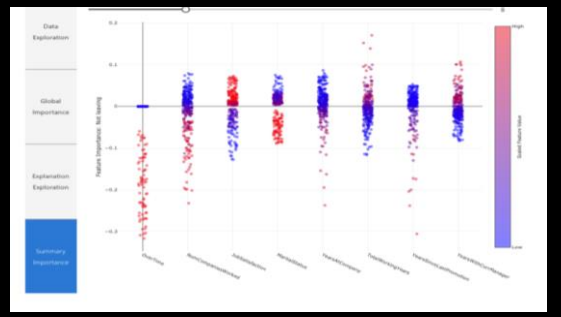
중요한 데이터 속성 N개를 글로벌 관점에서 표시

Explanation 탐색



개별 feature가 모델 예측에 미치는 영향 분석 (feature interaction 확인가능)

Summary Importance

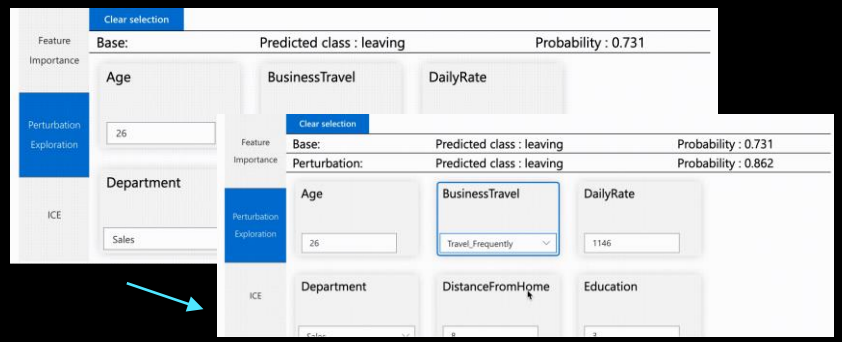


local feature importance를 각 data point에 대해 표시하여 예측값에 미치는 영향의 분포 확인

Feature Importance

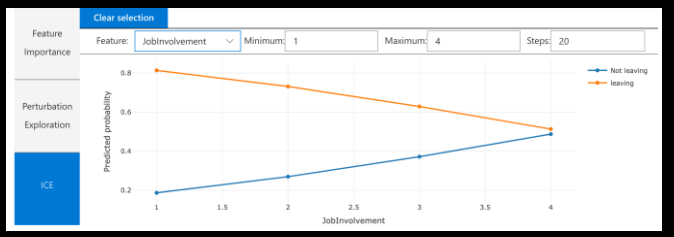
Data 탐색화면에서 특정 데이터 포인트를 선택하면, 해당 데이터 포인트에 대해 해당 모델의 지역적(local) feature importance 시각화

Perturbation 탐색



feature 값을 특정 값으로 변경했을 때 예측값에 영향을 주는 결과를 시뮬레이션할 수 있는 기능.

ICE (Individual Conditional Expectation)



feature 값을 특정 값이 아닌 최소에서 최대까지의 값으로 단계적으로 변화시켰을 때 예측값에 미치는 영향 확인

Episode 4
ML 생애주기 (3)
모델 해석

ML 생애주기 (3) 모델 해석

- 모델 해석이 왜 중요한가
- 모형을 해석하려는 시도
- azureml.interpret 들여다보기
- Explainer 관련 시각화 예시
  - 애저머신러닝에서 모델 해석 *DEMO*

{다음 시간에는}

Episode 5
ML 생애주기 (4)
배포/서빙

ML 생애주기 (4) 배포/서빙

- 패키징, 배포 (서빙)
- 모델의 모니터링: Data Drift

데모

- 애저머신러닝에서 배포
- 애저머신러닝에서 Data Drift 모니터링