

(참고) 피처 중요도 산출 방법론, SHAP 알고리즘의 비교

구분	피처 중요도(Feature Importance)	샐플리(SHAPely) 값
원리	<ul style="list-style-type: none">① Permutation (치환) 기법 이용② 특정 피처의 값을 임의의 값으로 치환했을 때 원래보다 예측 에러가 얼마나 커지는지 비교하는 방식으로 계산<ul style="list-style-type: none">- 모델 에러가 증가하면 필요 없는 피처로 보고 무시	<ul style="list-style-type: none">① 어떠한 결론에 도달하게 된 각 피처의 기여도를 계산② 전체 고객 간의 상대적 피처 중요도를 표시 (0~1)<ul style="list-style-type: none">- 어떤 고객의 소득구간 예측에 특정 피처가 결정적이었다면 샐플리 영향력 값은 크게 계산됨
장점	<ul style="list-style-type: none">① 계산 속도가 빠름② 피처의 변량에 기반하여 측정되기 때문에 극단치 영향 少③ 널리 알려진 개념이며 각 분석 패키지 내장함수로 존재	<ul style="list-style-type: none">① 변수들이 서로 영향을 미칠 가능성 고려② 음(-)의 영향력을 계산하여 긍·부정 기여 파악③ 개별 데이터의 분류 결정근거 설명 가능<ul style="list-style-type: none">- 트리 기반일 경우 분기(pruning)별 설명 가능
단점	<ul style="list-style-type: none">① 피처들이 상호의존적일 때 결과가 왜곡될 수 있음② 음(-)의 영향력은 계산 X, 실제보다 과대평가 가능③ 개별 피처 및 데이터에 대한 설명이 제한적<ul style="list-style-type: none">- 특히 피처의 범주별, 크기별 영향력에 대한 설명이 불가	<ul style="list-style-type: none">① 기본 계산량이 많아 측정 속도가 느림② 아웃라이어 데이터에 취약③ 측정치가 직관적이지 않을 수 있으며 분석 편이도 낮음
활용도	<ul style="list-style-type: none">① 피처의 추가와 삭제가 빠른 모델② 모델 기반 설명이 필요할 때③ 피처 선별과 축소를 통한 학습모델 개선④ 피처간 상관관계가 낮은 분야 (예: 이미지)	<ul style="list-style-type: none">① 이미 구축되어 있는 모델② 개별 피처, 개별 데이터에 대한 해석이 필요할 때③ 새로운 데이터 분류를 위한 예측모델 개선④ 피처간 상관관계가 높은 분야 (예: 인구통계, 금융)