

# AI를 만드는 AI, 'AutoML'…데이터 과학자 부족 해소하나

긱 김재호 기자 | 입력 2020.11.11 08:02 | 수정 2021.03.04 10:54 | 댓글 0 | 좋아요 0

AutoML, 데이터 과학자 '대체' vs '보완'  
데이터 과학자 인력난 해소와 ML 모델 업무 경감 기대  
AI 지식 갖춘 산업별 전문가 중요할 것

머신러닝(ML) 개발 도구인 'AutoML'이 인공지능(AI) 분야 인력 부족을 보완할 것이라는 전망이 나왔다. 빠른 데이터 분석과 ML 제작 능력을 바탕으로 업무 효율성을 높이면서 데이터 과학자의 역할을 대체할 수 있다는 설명이다.

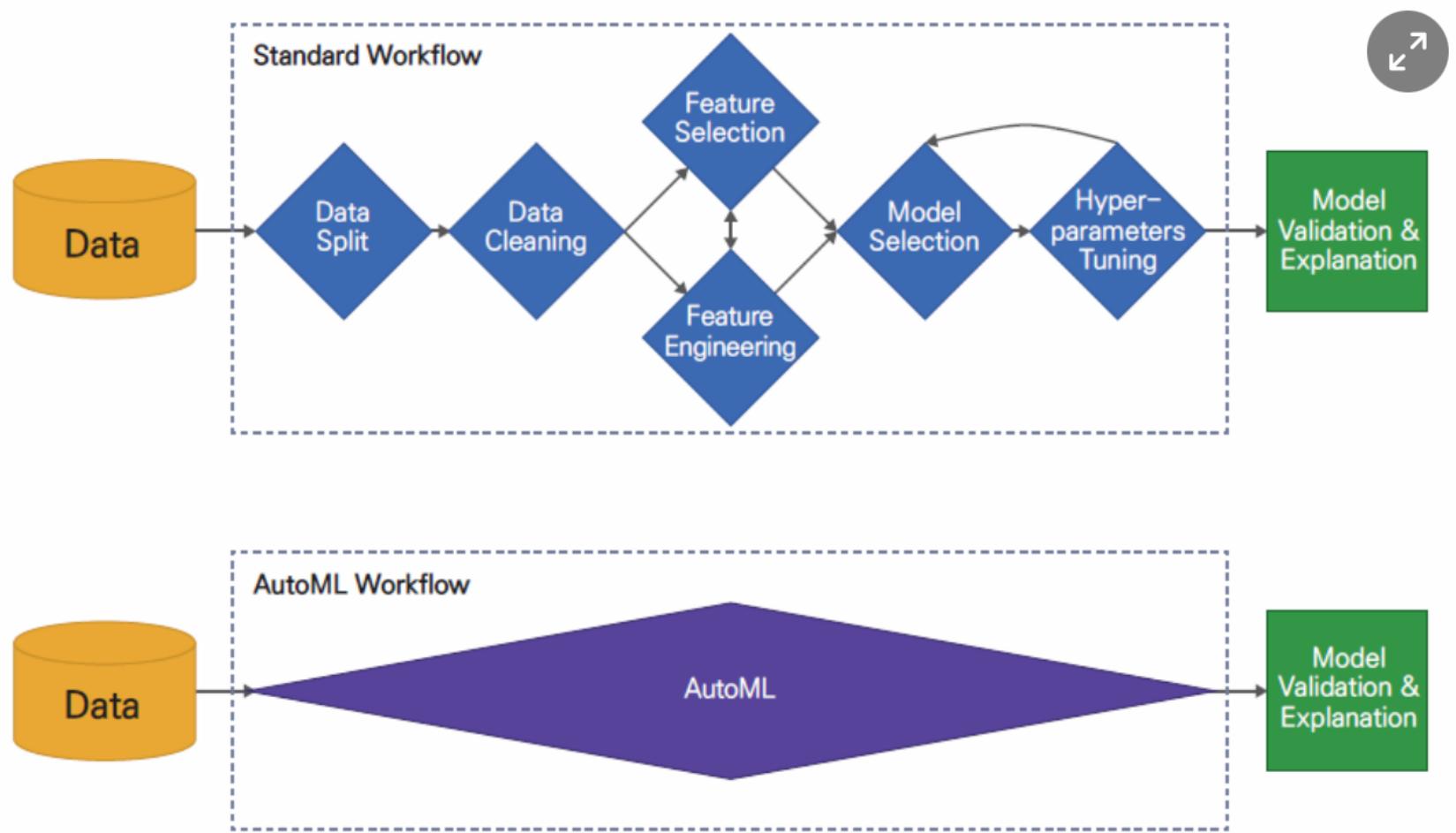
지난 달 15일 정보통신정책연구원(KISDI · 원장 권호열)은 '인공지능 기술 발전이 인재 양성 정책에 주는 시사점 : AutoML의 사례' 보고서를 통해 AutoML 기술이 AI 인재 수요에 영향을 미칠 것이라고 전망했다.

기존 ML 모델 개발 과정에서 데이터 과학자는 데이터 분석 · 전처리 과정부터 실제 모델 구축까지 다양한 역할을 수행했다. 최근 ML을 비롯한 AI 기술 수요가 높아지며 데이터 과학자의 수요도 높았다.

하지만 보고서는 AutoML을 적용할 경우 앞선 과정을 자동화할 수 있어 데이터 분석의 효율을 높일 것이며 향후 데이터 과학 분야 인재 부족을 해소할 수 있을 것으로 예상했다.

## ◆ 최적의 AI를 AI가 만든다…'AutoML'은 무엇인가

AutoML(Automatic Machine Learning)은 데이터 과학자 도움없이 기계 스스로 ML 개발 과정에 필요한 분석 · 수정 행위를 반복하며 고품질의 ML 모델을 자동 생성하는 AI 기술이다. 쉽게 말해 새로운 AI를 AI가 만드는 것이다.



ML의 표준 워크플로우와 AutoML 도입시 워크플로우 비교(사진=KISDI, '인공지능 기술 발전이 인재 양성 정책에 주는 시사점 : AutoML의 사례')

AutoML을 활용할 경우 반복적 작업을 자동 실행하기 때문에 효율성을 향상시키고 데이터 과학자가 문제에 더 집중할 수 있도록 돕는다. 또 자동화한 ML 파이프라인으로 수동 작업의 잠재적 오류를 최소화할 수 있으며, ML 기술의 대중화를 가능하게 해 비전문가도 AI의 편익을 쉽게 누릴 수 있다.

이전까지 최적의 성능을 갖춘 ML 모델을 개발하는 데 '수동 선택법(Manual Search)'을 적용했다. 실무자가 ML의 초매개변수를 일일이 변경하고 ML 모델의 성능 측정값을 평가해 비교적 높은 성능의 모델을 선택하는 방법이다.

초매개변수는 학습 프로세스 자체를 제어하는 변수다.

하지만 이 방식은 반복적인 시행착오를 수반한다. 데이터 획득 · 분석 · 전처리 · 설계, 적합한 ML 모델 선택, 모델의 초매개변수 최적화 등 많은 과정을 거쳐 최종 모델의 성능을 예측하며 이 과정에서 사람이 직접 다양한 기법을 적용해 모델 성능을 평가해야 한다.

또 최적의 ML 모델 예측 · 선택을 사람이 담당하는 만큼, 실무자 직관에 의존하기 때문에 고도의 전문지식을 갖춘 전문가가 필요했다.

주민식 삼성 SDS 데이터분석그룹 프로는 "반복적인 시행착오는 높은 컴퓨터 연산 능력을 요구하기 때문에 컴퓨터 성능을 충분히 확보한 1990년대 오픈소스에서 AutoML이 드러나기 시작했다"고 짚었다.

이후 AutoML의 시초 격인 '하이퍼 파라메터 최적화(HPO)'가 등장했다. HPO는 수동 선택법보다 자동화한 ML 모델 제작법이다. 기존보다 지능화한 기법을 이용해 최소한의 반복적 시행착오를 거쳐 필요조건을 만족하는 초매개변수를 스스로 탐색한다.

주민식 프로는 "2017년 구글이 '보강 학습을 통한 신경 아키텍처 탐색' 논문을 통해 '신경망 아키텍처 탐색(NAS)' 기술을 제시하면서 AutoML 연구가 활발해졌다"고 봤다.

NAS는 목표 데이터 · 태스크에 적합한 인공 신경망 구조를 자동 탐색하는 기술이다.

주민식 엔지니어는 "구글과 아마존 등 해외 IT 기업은 보다 실용화한 자체 AutoML 솔루션을 제공하고 있다"며 "삼성 SDS에서도 데이터 이해 · 전처리 영역에 강점을 보이는 'Brightics AI'를 서비스하고 있다"고 짚었다.

이어 그는 "국내 IT 전문 기업도 알고리즘 자동 설계 기술을 지원할 수 있도록 검증된 해외 소스를 이용하거나 자체 확보한 기술로 AutoML를 출시 · 준비하고 있다"고 덧붙였다.

## ◆ 데이터 과학자 수요 감소는 불가피…산업 분야별 전문가 역할 중요할 것

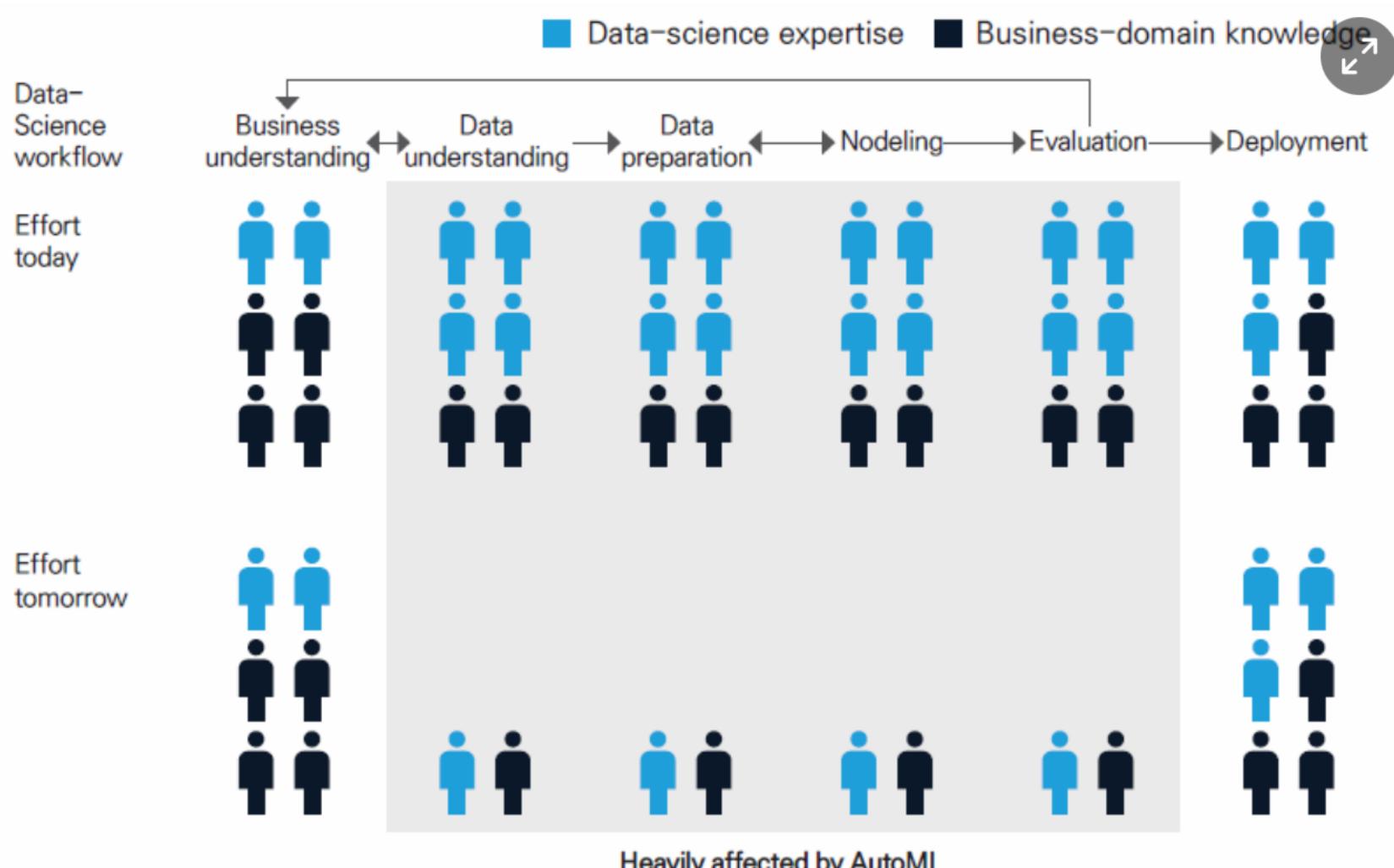
업계는 AutoML 등장에 따라 데이터 과학자 수요 감소는 불가피할 것으로 전망하고 있다. 데이터 과학자 구인난을 겪은 기업 입장에서 AutoML을 투입해 이를 해소할 것이란 예측이다.

KISDI 보고서에 따르면, 데이터 과학자가 거대 IT기업에 몰리며 일반 기업의 데이터 과학자 인력 수요가 높은 상황이라고 설명했다. 이에 개발 생산성을 높일 수 있도록 ML 모델 개발 과정에 AutoML 적용 시도가 활발하다고 덧붙였다.

AutoML 기술에 의한 데이터 과학자 수요 감소를 바라보는 시각은 ①'전면 대체'와 ②'보완적 수준' 두 가지로 나뉜다.

① 전면 대체의 경우 향후 AutoML이 데이터 과학자의 업무 영역인 'AI 모델링과 튜닝' 부문까지 수행 할 것인 전망이다. 지난 'TensorFlow Dev Summit 2018'에서 제프 딘(Jeff Dean) 구글 AI 연구팀 대표는 "현재의 100배 수준의 컴퓨팅 파워를 AutoML에 적용할 경우 AI 전문가를 대체할 것"이라고 예측한 바 있다.

KISDI 보고서는 AutoML 기술을 이용해 기존보다 ML 모델 구현이 쉬워지면서 AI 전문가를 향한 의존도가 약해지고 AI 지식을 갖춘 산업 영역별 전문가 수요가 증가할 것으로 봤다.



산업에서 AI 기술을 효율적으로 운용하기 위해서 각 분야별 전문 지식을 함께 갖춰야 한다. 이전까지 산업별 전문가와 AI 전문가가 팀을 이뤄 산업 분야에 맞는 AI 모델을 개발했다. 하지만 AutoML 기술 고도화에 따라 AI 지식을 갖춘 비전문가도 ML 모델 개발을 기대할 수 있게 됐다.

하지만 고도로 훈련된 전문 데이터 과학자의 필요성은 존재한다고 평가했다. AutoML은 현재 일반적 모델 개발을 간소화하는 데 적합하지만, ML 모델의 공정성과 신뢰를 구축하기 위해 훈련된 데이터 과학자의 전문 지식이 필요하다는 지적이다.

주민식 프로는 AutoML을 보조적 수단으로 인식하는 시각에 무게를 실었다. AutoML이 다양한 조건에 따른 실험과 시행착오를 대신 수행해 데이터 과학자의 데이터 분석과 특징 설계 등 창의적 활동에 집중할 수 있도록 할 것이란 분석이다.

ML 모델을 개발하는 데 ▲도메인 · 데이터 이해 ▲데이터 전처리 ▲모델 개발 ▲모델 평가 · 배포의 과정을 거친다. 주로 도메인 전문 지식은 데이터 이해 · 전처리 단계에서 활용하고, AI 전문 지식의 경우 모델 개발 이후 과정부터 중요하다.