

ModelOps 시작하기

진정한 비즈니스 가치를 창출하는 검증된 분석 모델 배포 방식



SAS는 지난 40년 동안 전 세계 고객들이 분석을 통해 더 나은 의사결정을 신속하게 내릴 수 있도록 지원해왔습니다. 기업들은 분석 기술을 개발하는 데 많은 시간과 비용을 투자합니다. 그러나 분석 모델을 배포하는 과정에서 어려움을 겪고 있습니다. 그래서 배포는 분석의 '마지막 고비'로 알려져 있습니다. 모델 배포에 걸리는 지연 시간이 길어지면 개발 노력이 무색하게도 기대했던 투자 효과가 물거품처럼 사라지기도 합니다.

이를 방지하기 모델 배포 방식을 재정립해야 합니다. 그 해결책이 바로 ModelOps입니다. ModelOps를 추구하는 기업은 고품질의 분석 모델을 개발, 검증, 배포 및 모니터링하는 데 걸리는 시간을 최소화할 수 있습니다.

ModelOps를 성공적으로 구현하려면 기존의 틀을 깨야 합니다. 물론 변화가 쉬운 것은 아닙니다. 그러나 지금과 같은 데이터 중심의 디지털 경제에서는 분석 모델을 더 빠르고 더 효과적으로 배포할수록 더 많은 가치 창출을 기대할 수 있습니다.

아래에서 ModelOps에 대한 내용을 확인해 보십시오. 분석의 마지막 고비를 극복할 수 있도록 SAS가 최선을 다해 돕겠습니다.

샤디 샤힌(Shadi Shahin)
제품 전략 사업부 부사장
SAS



목차

4

마지막 고비:
모델 배포



8

DevOps:
성공에 이상적인 모델



10

ModelOps:
분석 라이프사이클의
가속화



15

ModelOps를
시작하는 5단계



21

ModelOps는
장기전입니다.





마지막 고비:
모델 배포

오늘날 거의 모든 기업들이 디지털 혁신을 최우선 과제로 삼고 있습니다. 비즈니스 결정을 내리기 위해 데이터를 활용하고 실시간 분석 정보를 끊임없이 응용할 수 있어야 하기 때문입니다.

분석이 지닌 가능성과 잠재적인 비즈니스 가치가 무한하기에 기업들은 지난 10년 동안 적합한 인재, 데이터, 프로세스, 그리고 기술에 투자해왔습니다.

그리고 이 투자 비용은 날로 증가하고 있습니다. IDC에 따르면, 분석 투자 비용은 2019년에만 1,891억 달러에 달한다고 합니다. 그리고 2022년이면 2,743억 달러에 이를 것으로 예상하고 있습니다.¹

그러나 이러한 집중 투자가 무색하게도 현실은 처참합니다.

- 최적 모델 중 배포되는 모델은 50%가 채 되지 않습니다.
- 90%의 모델이 배포하기까지 3개월 이상이 소요됩니다.
- 44%의 모델이 프로덕션 환경에 투입되기까지 7개월 이상이 소요됩니다.



50% 미만

최적 모델 중 배포되는 모델은 50%가 채 되지 않습니다.



90%

90%의 모델이 배포하기까지 3개월 이상이 소요됩니다.

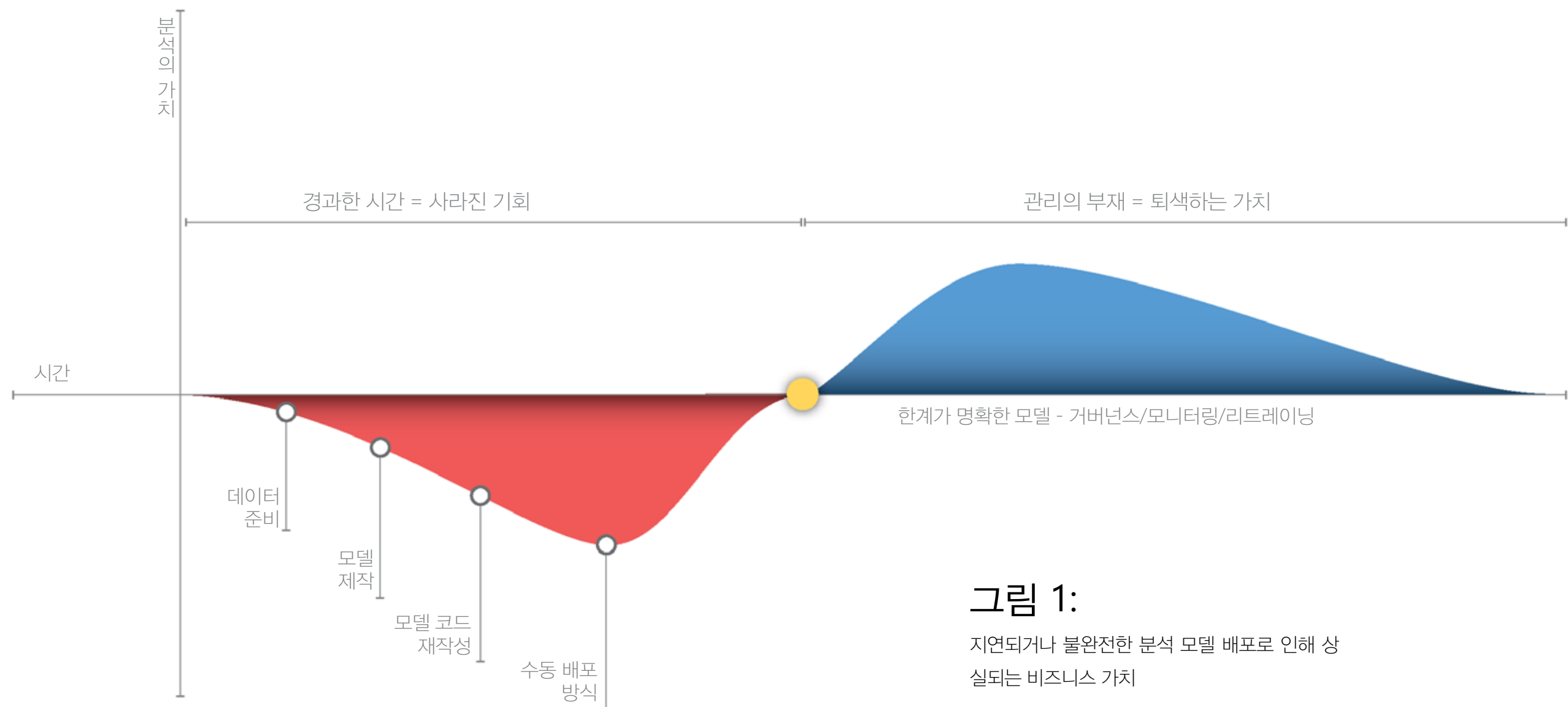


44%

44%의 모델이 프로덕션 환경에 투입되기까지 7개월 이상이 소요됩니다.

1. Worldwide Semiannual Big Data and Analytics Spending Guide, IDC. 2019년 4월

무수한 분석 모델이 아예 배포되지 않거나 제때 배포되지 않는 탓에 엄청난 자원과 투자 수익이 물거품처럼 사라집니다. 경영진으로서 용납할 수 없는 현실입니다. 모델을 프로덕션 환경에 투입하더라도 또 다른 난관이 기다리고 있습니다. 분석 라이프사이클의 번거로운 인계 과정과 더딘 프로세스에 지속적인 모델 관리의 부재라는 악재가 겹치면서 기회는 사라지고 분석 작업의 비즈니스 효과도 저하됩니다(아래의 그림 1 참조). 기업은 분석 모델을 배포하고, 적절히 관리하며, 새로운 데이터 또는 분석 기술로 시간이 지나도 적절한 성능을 유지하게 만드는 일을 최우선 과제로 삼아야 합니다.



모델 배포의 걸림돌

지금까지 투자를 어디에 집중(예: 데이터 사이언티스트 채용, 새로운 분석 솔루션 구축)했는지 돌아켜 보면 모델 배포가 절대 쉬운 일이라는 것을 알 수 있습니다. 기업이 일관적인 분석 플랫폼을 갖추거나 분석 기술력이 성숙기에 있지 않은 한은, 분석 기술을 운영하는 데 필요한 자원에 투자하더라도 새로운 요구를 충족하기가 어렵습니다.

최근 오픈 소스의 사용이 폭발적으로 증가하고 기업에서 사용 중인 솔루션의 환경이 계속 변화하고 확장되면서 배포 환경은 더욱 복잡해졌습니다.

예를 들어 데이터 사이언티스트가 더 우수한 모델을 제작하거나 최신 기술을 적용하는 방법을 어떤 식으로 찾는지 살펴볼 필요가 있습니다. 일반적으로 데이터 사이언티스트는 분석 모델을 프로덕션 환경에 효율적으로 배포하는 데 필요한 프로세스 중심 작업에 익숙하지도 않으며, 관련 교육을 받지도 않습니다.

마찬가지로 IT 부서 역시 분석 모델 제작 방식과 해석 방법에 대해 교육을 받지 않습니다. 그래서 이들은 데이터 사이언티스트가 분석 모델을 완성했을 때 배포 준비 과정에서 어려움을 겪을 수 있습니다. 결과적으로, 데이터 사이언티스트와 함께 모델을 다시 제작하거나 간소화하느라 시간을 낭비하는 경우가 많습니다.

이러한 유형의 작업에 대한 정형화된 프로세스가 없고 수행해야 할 작업에 대한 이해와 경험이 부족한 경우에는 배포 작업이 매우 어려워질 수 있습니다.



"업계에서 가장 큰 문제 중 하나는 분석 기술을 제대로 활용하지 못한다는 것입니다. 분석 기술 투자에서 진정한 비즈니스 가치를 창출하려면 ModelOps 방식을 채택해야 합니다."

SAS 제품 전략 사업부 부사장 샤디 샤힌

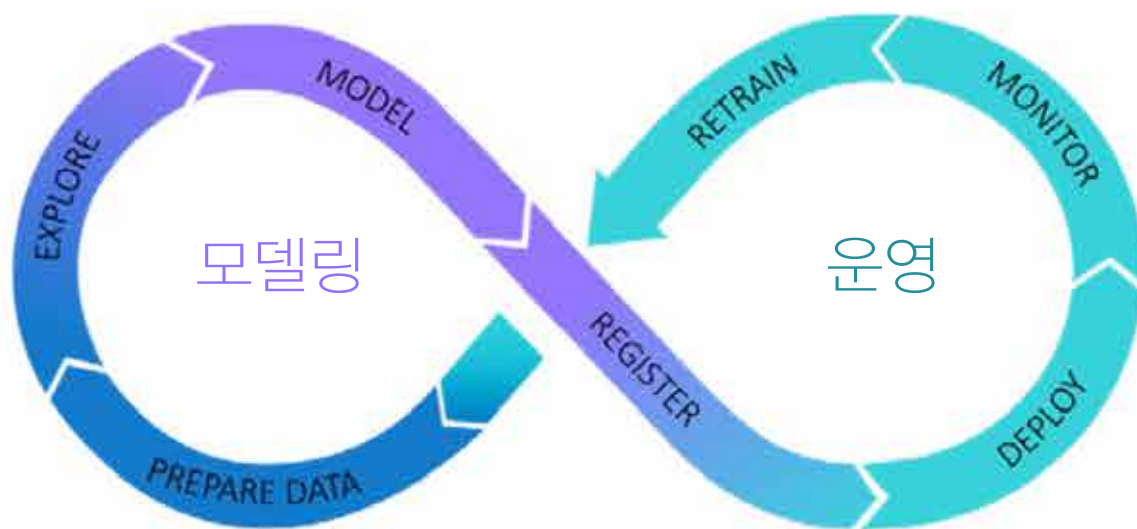


DevOps:
성공에 이상적인 모델

분석 팀들은 과거 10년 전에 애플리케이션 개발 팀이 경험했던 것과 유사한 문제를 안고 있습니다. 애플리케이션 개발 팀은 결국 DevOps로 문제를 해결했습니다. 이제 DevOps는 협업을 통해 보다 우수한 품질의 애플리케이션을 신속히 개발하여 비즈니스 요구와 고객의 가치를 충족하는 주요 방법으로 활용되고 있습니다.

애플리케이션 개발 팀과 운영 팀들은 업무와 목표가 서로 다르며 심지어 상충되는 부분이 있어 소통이 단절되는 경우가 많습니다. 이러한 문제로 인해 개발 주기가 길어지고, 고객 만족도가 떨어지며, 개발 작업은 수포로 끝나게 됩니다. DevOps를 도입하자 두 팀의 업무 진행 방향이 폐쇄적인 환경을 관리하는 일에서 기업의 목표를 달성하는 데 필요한 것, 즉 원하는 수준의 비즈니스 가치를 창출하는 애플리케이션을 관리하는 일로 바뀌었습니다.

DevOps를 제대로 도입하면 모든 관계자와 팀원들이 양질의 소프트웨어를 하루빨리 완성하기 위해 노력하게 됩니다. 또한 새로운 프로세스와 기술뿐만 아니라 서로의 차이를 이해하면서 원하는 수준의 비즈니스 가치를 제공하는 애플리케이션과 서비스를 협력적이고 민첩한 방식으로 구축할 수 있습니다. 소프트웨어는 기업과 사용자의 요구가 변화함에 따라 고객과의 소통을 통해 지속적으로 개발되고 보완됩니다. 덕분에 개발 프로젝트의 실패 확률이 줄어들며, 개발 팀이 문제 발생에 신속히 대응할 수 있습니다.



분석 모델의 운용에 따른 당면 과제를 해결하려면 DevOps와 유사한 개발 방식을 추구하여 데이터 사이언티스트, 일선 부서 및 IT 부서의 운영을 통합하고 다음과 같은 여건을 조성해야 합니다.

또, 모델을 손쉽게 테스트하여 운용하려면 분석 모델링 팀의 작업 결과가 일선 부서와 IT 부서의 요구에 부합해야 합니다.



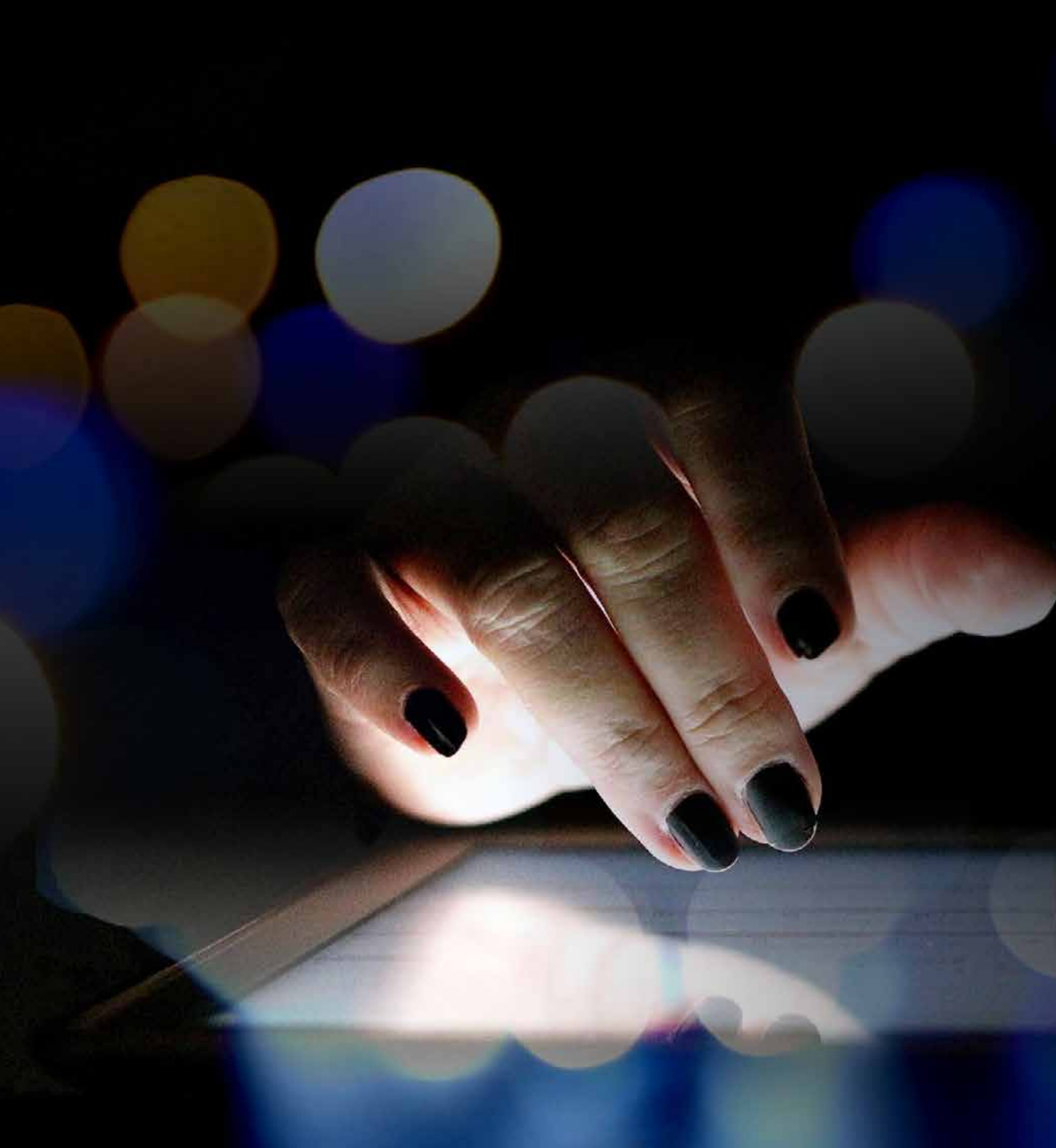
사용 사례(예: 데이터베이스, REST API, 장치를 통한 스트리밍)에 따라 다양한 통합 엔드포인트를 통해 분석 모델의 배포 속도와 확장성을 크게 개선해야 합니다.



모델이 적절한 성능을 계속 유지하고 부가 가치를 창출할 수 있도록 지속적으로 모니터링하고 업데이트해야 합니다.



유수의 기업들은 ModelOps 방식을 채택하고 있습니다.



ModelOps:

분석 라이프
사이클의 가속화

DevOps가 실제 비즈니스 가치를 창출하는 애플리케이션의 개발 시간을 단축하는 방식과 유사하게 ModelOps는 분석 모델의 연구부터 등록 및 배포까지 걸리는 시간을 최소화합니다. 덕분에 고품질의 분석 결과를 확보하고 원하는 수준의 비즈니스 가치를 실현할 수 있습니다.

ModelOps를 구현하면 배포 준비가 된 모델을 모든 단계에서 데이터 사이언티스트 팀과 IT 운영 팀이 교대로 관리할 수 있으며, 필요한 경우 모델 모니터링 과정에서 피드백을 토대로 모델을 적시에 리트레이닝할 수 있습니다. 이러한 방식으로 사람과 기술을 아우르는 검증된 통합 프로세스를 통해 더 많은 분석 모델을 일상 업무에 전사적으로 활용할 수 있습니다.

그리고 각 팀이 단절된 상황에서 애플리케이션을 개발하고 IT 부서에 배포 작업을 등한시하는 구조를 지양하는 DevOps와 마찬가지로, IT 운영 팀이 쉽게 배포할 수 있는 방식으로 모델을 관리하게 됩니다. 또한 모델 트레이닝에 사용되는 데이터가 프로덕션 환경에 사용되는 운영 데이터와 일치하게 됩니다. 그 밖에도 분석 팀과 IT 부서가 상대 팀과 최종 사용자의 입장에 대한 이해와 공감을 바탕으로 협업을 진행하게 됩니다. 이러한 문화는 초기 데이터와 아키텍처를 선정하는 데 자연스럽게 영향을 미치므로 향후 모델 배포 작업이 훨씬 수월해집니다. 즉, 모델이 아래와 같은 방향으로 개발 및 관리됩니다.

- 배포를 염두에 두고 개발됩니다.
- 모니터링을 염두에 두고 배포되므로 데이터 사이언티스트와 분석가가 모델을 모니터링하여 모델의 성능이 저하될 경우 리트레이닝할 수 있습니다.

오늘날처럼 가용 데이터의 유형과 분석 기술 및 프로그래밍 언어가 다양해지고 있는 상황에서는 이러한 방식이 더욱 중요한데, 이는 기업이 배포하는 모델의 유형과 사용하는 데이터의 출처가 다양할 수밖에 없기 때문입니다.

데이터도 빼놓을 수 없습니다.

모델을 최대한 효과적으로 운용하려면 견고한 데이터 관리 기반이 필수적입니다. 데이터에 관해 고려해야 할 주요 사항은 다음과 같습니다.



데이터 결합

기업이 모델을 성공적으로 개발하고 배포하려면 필요한 모든 관련 데이터를 결합하고 이에 액세스할 수 있어야 합니다. 예를 들어 모델이 학습을 통해 성능이 높아지려면 데이터가 필요합니다. 더 많은 데이터 출처를 확보하여 가용 데이터를 늘려야만 성공할 수 있습니다.



데이터 준비 및 데이터 품질

데이터 사이언티스트가 데이터를 준비하고 관리하는 데 대부분의 시간을 보내야 한다면 업무 만족도나 생산성이 떨어질 수밖에 없습니다. 기업은 직원들이 데이터에 빠르게 액세스하고 다운스트림 애플리케이션을 표준화, 대조, 연결할 수 있는 데이터 준비 도구를 갖춰야 합니다.



데이터 거버넌스

데이터 사이언티스트가 데이터 액세스 권한이 있는 사용자와 데이터의 위치, 그리고 시간의 추이에 따른 데이터의 변화 양상을 파악한다면 '견제와 균형'을 유지한 채 변화하는 데이터의 품질과 개연성을 모니터링하고 평가할 수 있습니다.

ModelOps의 차별화 요인

DevOps와 ModelOps에는 여러모로 유사한 점이 많지만, 분석 모델의 특성 때문에 ModelOps가 DevOps와 사뭇 다르다는 사실을 유념해야 합니다. 기존의 애플리케이션은 같은 일을 반복해서 수행하므로, 소프트웨어를 수정하거나 업데이트한 경우에만 코드 실행의 결과물이 달라집니다.

반면에 분석 모델은 기업이 확보하지 않은 데이터의 결과를 예측하는 알고리즘을 기반으로 합니다. 분석 모델은 결과와 기반 데이터 간의 관계가 달라지면 리트레이닝 또는 교체해야 합니다. 이로 인해 테스트 및 배포 프로세스가 더욱 복잡해지므로 지속적으로 모니터링, 리트레이닝, 재검증 및 재배포할 수 있는 환경을 구현해야 합니다.

ModelOps 방식의 최대 장점은 프로덕션 환경에 투입되는 모델의 수를 늘려서 투자 효과를 일관적으로 거둘 수 있다는 것입니다.

기업이 초기 투자 비용을 줄이면서도 거시적이고 전사적으로 더 우수한 분석 가치를 유지할 수 있다는 점 역시 중요합니다. 또한 ModelOps 방식을 이용하면 시간이 지나도 모델의 성능이 저하되지 않고 계속 유지됩니다.

데이터 사이언티스트와 소프트웨어 개발자는 실질적으로 ModelOps의 효과를 누릴 수 있습니다. 특히 ModelOps를 통해 두 그룹이 각자의 목적과 방식에 대한 공통된 이해를 바탕으로 작업하므로 제반 상황을 명확히 이해한 채 서로 소통할 수 있습니다.

ModelOps의 이점 실현

ModelOps는 다음과 같은 방식을 통해 기업이 분석 기술의 투자 가치를 실현하는 방법을 혁신하고 개선할 수 있습니다.

- 모델을 신속히 운용하여 비즈니스 가치를 더 빨리 창출합니다(13페이지의 그림 2 참조).
- 복잡성과 재코딩 작업을 최소화하여 생산성을 높입니다.
- 모델이 지속적으로 적절한 성능을 발휘할 수 있도록 배포 후에도 모델의 상태를 모니터링합니다.

ModelOps를 사용하여 분석 모델을 제작하는 데이터 사이언티스트는 다음과 같은 이점을 누릴 수 있습니다.

- 비즈니스 문제를 해결하는 데 원하는 솔루션을 사용할 수 있습니다.
- 더 많은 모델을 운용할 수 있습니다.
- IT 부서가 코드를 단순화하지 않고도 모델을 최대한 활용할 수 있습니다.
- 비효율성과 오류 발생 가능성이 높은 수동적이고 반복적이며 비생산적인 작업을 최소화할 수 있습니다.
- 최신 분석 기술을 전사적으로 보급하는 데 주력할 수 있습니다.



ModelOps를 사용하여 분석 모델을 배포하는 개발자는 다음과 같은 이점을 누릴 수 있습니다.

- 일회성 수작업을 수행하거나 모델 배포 방법을 다시 개발할 필요가 없습니다.
- 용량을 효과적으로 관리하고 계획하여 모델 활용 범위를 확장할 수 있습니다.
- 모델, 파라미터 및 관련 데이터를 체계화하여 통합 문제를 줄일 수 있습니다.
- 분석 작업의 저해 요인을 해소할 수 있습니다.
- 패키화된 알고리즘 코드와 데이터를 지속적으로 활용할 수 있습니다.



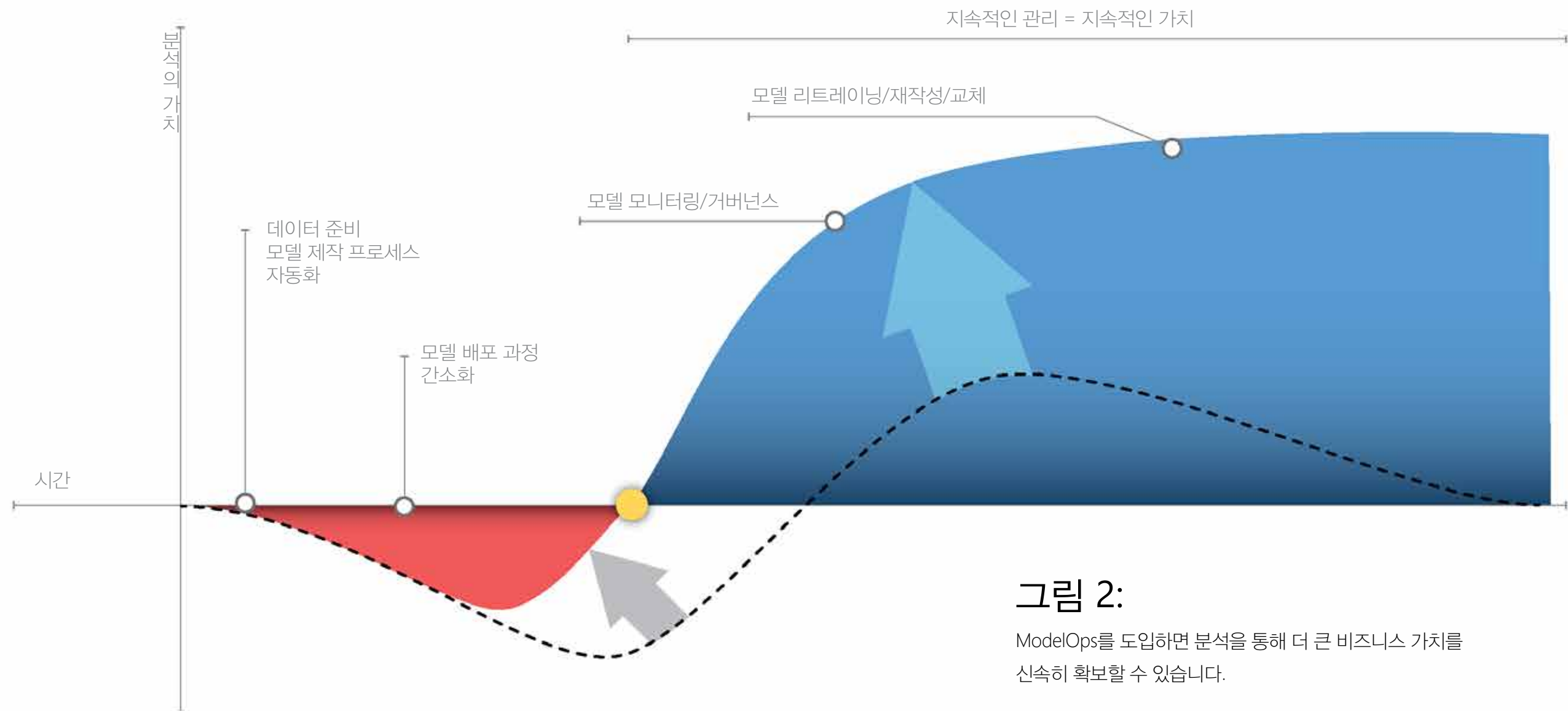
ModelOps의 필요성 진단

기업이 분석 모델을 사용하는 방식은 시간이 갈수록 변화하면서 성숙해지지만, 그 과정은 기본적으로 동일합니다. 대부분의 경우 초기에는 소수의 데이터 분석가만 동원됩니다. 그러다 경영진이 분석 결과에 관심을 갖기 시작하면 데이터를 정리하고 더 나은 기술을 도입하여 몇 가지 모델이 완성됩니다.

이와 같은 투자로 비즈니스 가치가 창출되면 분석 기술에 대한 수요가 증가하고 새로운 모델을 더 빨리 제작하라는 압력이 증가합니다. 하지만 분석 팀과 IT 팀은 데이터를 준비하고, 모델 파이프라인을 자동화하며, 우수 모델을 프로덕션 환경에 통합하는 데 상당한 문제를 겪게 됩니다.

이 시점에서 ModelOps가 필요합니다. ModelOps를 도입하면 단절성, 수동 프로세스, 가치 측정 방식에 대한 의견 충돌, 주요 관계자들 간의 화합 부족 등의 문제를 해결하고 개발 중인 분석 모델을 대규모로, 그리고 효율적으로 배포할 수 있습니다.

또한 ModelOps를 도입하면 분석 기술과 인공지능을 쉽게 확장할 수 있습니다. 귀사에 ModelOps가 필요한지 궁금하십니까? 다음 페이지에서 네 가지 상황 중 하나에 해당하는지 확인하십시오.



ModelOps가 필요한 4가지 상황



병목 현상

데이터 사이언티스트는 다양한 기술을 탑재한 상용 소프트웨어와 오픈 소스를 사용하여 많은 수의 모델을 제작합니다. 그러나 데이터 사이언티스트가 모델이 양호하다고 말한 시점과 실제로 배포되는 시점 사이에 워낙 많은 인계 절차와 확인 절차가 있다 보니 최종 심사까지 통과하는 모델은 극소수에 불과합니다. 기업이 분석 모델을 대규모로 제작할 능력을 갖췄다 하더라도 분석 모델을 대규모로 배포, 모니터링 및 관리할 능력은 그에 비해 떨어집니다.



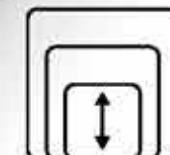
블랙박스 모델

운영 체제에는 다수의 분석 모델이 배포되어 있습니다. 그러나 모델을 마지막으로 리트레이닝한 시점, 모델의 성능, 모델 제작자에 대해서는 아무도 알지 못합니다. 모델은 기업이 작동 원리를 이해하지 못한 채 사용하는 일종의 블랙박스입니다. 배포한 모델과 관련된 IP가 남아 있지 않아서 모델을 업데이트하려면 처음부터 다시 작업해야 할 가능성이 높습니다.



빠듯한 일정

일선 부서가 새로운 문제를 들고 분석 팀을 찾아오면 분석 팀에서 데이터를 준비하고, 분석 정보를 도출하며, 모델을 배포하는 데 수개월이 걸립니다. 그런데 그 무렵이면 상황이나 요구사항이 크게 바뀌기 때문에 분석 팀은 처음부터 다시 시작해야 합니다. 마치, 성과를 전혀 거두지 못하고 영원히 끝나지 않는 개발 주기에 갇힌 듯한 모양새입니다. 모델의 분석 라이프사이클을 완료하는 데 걸리는 시간이 비즈니스의 변화 속도를 따라가지 못합니다.



버전 관리 문제

모델을 프로덕션 환경에 배포했지만 데이터를 평가하고 버전을 관리하는 프로세스는 모두 수작업으로 이뤄집니다. 모델의 버전이 너무 많아서 어떤 버전에 어떤 데이터를 사용하여 어떤 결과를 얻었는지 알아내기 어렵습니다. 데이터 사이언티스트는 같은 작업을 반복하기 일쑤이고, 분석 결과의 신뢰성을 검증해야 합니다.



ModelOps를
시작하는
5단계

이런 상황에 해당하는 기업이라면 단연코 ModelOps가 필요합니다. 다음 5단계에 따라 ModelOps 도입을 시작해보십시오.

1단계: 현재 상황 파악

ModelOps를 도입하기로 한 후 가장 먼저 해야 할 일은 기존의 공식/비공식적 분석 프로세스를 명문화하고 그 효용성을 평가하는 것입니다. 다음과 같은 질문을 통해 실태를 파악하십시오.

- 사용 중인 모델은 몇 개나 되며, 어디에 저장/보관되어 있는가?
- 각 모델을 누가, 언제, 어떻게 업데이트했는가?
- 모델을 누가 관리하는가?
- 프로덕션 환경에 적합한 모델이 사용되고 있는가? 적합하다는 것을 어떻게 알 수 있는가?
- 모델을 배포하는 데 어떤 작업이 필요한가? 해당 작업의 담당자는 누구인가? 문서화된 프로세스가 있는가?
- 모델을 배포하는 데 얼마나 걸리는가?
- 모델 학습에 얼마나 오래된 데이터를 사용하는가? 정확하고 신뢰할 수 있는 데이터가 사용되는가?
- 모델의 성능은 어떠한가?
- 시간이 흐르면서 다양한 모델을 동일한 용도로 사용하는 경우 이 모델들을 어떻게 비교할 수 있는가?
- IT 부서가 분석 팀과 협력하여 쉽게 배포 가능한 모델을 제작할 수 있는 환경을 구축하고 있는가?

이러한 질문들을 통해 병목 현상의 위치, 수정하거나 구현해야 하는 프로세스, 최우선적으로 관심을 기울여야 할 사항을 파악할 수 있습니다. 무엇보다 이러한 질문을 토대로 분석 모델을 제작 및 배포하는 사람들과 대화를 시작할 수 있습니다. 대화만으로도 ModelOps를 성공적으로 도입하는 데 결정적인 역할을 하는 문화적 변화의 계기를 마련할 수 있습니다.



2단계: 변화의 촉진

모델 배포 프로세스의 현재 상황을 파악했다면 이제 모델의 제작 및 배포 방식을 개선하는 데 필요한 문화적 변화와 자동화 및 프로세스 표준화를 촉진하는 데 주력해야 합니다. 대부분의 변화가 그러하듯 ModelOps를 도입하는 과정 역시 쉽지 않습니다. 그 과정은 반복적으로 진행되며, 문화와 프로세스의 변화를 결정하고 수용하는 데 시간이 걸리고 리더십도 필요합니다.

투자 효과가 확실한 영역부터 시작하고, 적절한 변화를 수용하며, 변화의 가치를 입증하고, 모든 것을 한 번에 바꾸려 하기보다는 지속적으로 되풀이하는 것이 가장 좋습니다. 일반적으로 투자 효과가 확실한 영역으로는 문화적 공감대 형성, 표준화, 그리고 자동화를 들 수 있습니다.

3단계: 문화적 공감대 형성

일관적이고 측정 가능한 방식으로 모델의 분석 라이프사이클을 완료하려면 원하는 수준의 비즈니스 가치를 제공하는 분석 모델을 개발하겠다는 공통된 목표 하에 완전히 새로운 차원의 부서 간 협업이 뒷받침되어야 합니다. 데이터 사이언티스트, 엔지니어, IT 개발자, 소스 시스템 전문가, IT 운영 인력, 그리고 경영진을 포함한 모든 구성원이 동일한 프로세스, 모델 기대치 및 규칙을 토대로 협업해야 합니다.

이러한 프로세스를 통해 ModelOps의 도입은 조직의 운영 방식 개선을 기대해볼 수 있습니다. 모델링 프로세스에서 단절성으로 인해 야기되는 마찰과 번거로운 인계 절차를 해소하고 운영 방식 개선을 통해 다기능 팀을 구성함으로써 분석 결과를 더 빨리 확보할 수 있습니다. 그 과정이 쉽지는 않겠지만 그에 따른 성과를 생각하면 노력할 가치가 충분합니다. 다른 성공적인 프로젝트와 마찬가지로 다음과 같은 요소가 성공에 결정적인 영향을 미칩니다.

- 경영진의 리더십
- 투자 효과가 확실한 영역 파악
- 가치 입증
- 신뢰 구축 및 이해
- 동기부여를 통해 가시적인 성과를 확보
- 직원 역량 강화
- 반복적 개선 문화 조성

4단계: 표준화와 자동화에 투자

분석을 적극적으로 활용하려면 모델 저장, 성능 모니터링, 경보, 모델 배포 및 리트레이닝 프로세스를 중심으로 표준화, 그리고 궁극적으로 자동화에 투자해야 합니다. 모든 수작업을 표준화하고 자동화하면 복잡성과 기술 공백 및 오류 지점이 줄어듭니다. 그림 3에서 볼 수 있듯이, 표준화와 자동화는 데이터 준비부터 모델 폐기에 이르는 전 과정에 영향을 미칩니다.

표준화

ModelOps를 통해 프로세스를 표준화하면 동일한 작업에 대한 접근 방식을 일원화하여 위험을 줄이고 처리 속도를 높일 수 있습니다.

표준화를 시작하려면 비효율적이고 KPI에 악영향을 미치는 프로세스, 방법, 기타 영역을 논의하여 파악해야 합니다. 그리고 가장 큰 문제가 있거나 KPI에 가장 큰 영향을 미치는 프로세스를 우선적으로 개선해야 합니다. 이러한 프로세스를 해결하고 나면 다음 프로세스를 수정해야 합니다.

표준화는 다음과 같은 작업에서 중요한 역할을 합니다.

- 환경 구성 및 설정 프로세스를 일관성 있게 확립하여 데이터 사이언티스트가 모델 개발, 테스트 및 릴리스 주기를 단축합니다.
- 사용되는 프로그래밍 언어와 관계없이 모든 분석 팀이 통제된 데이터를 중앙에서 프로비저닝합니다.
- 일관적이고 검증된 프로세스를 사용하여 모델을 전사적으로 배포하고 관리합니다.

- 사용되는 분석 언어나 기술과 관계없이 전사적으로 사용되는 분석 모델 전용 공통 저장소를 구축합니다.

특히 ModelOps를 구현하려면 중앙 저장소를 중심으로 표준화해야 합니다.

자동화

프로세스 표준화와 마찬가지로 분석 라이프사이클 전체를 자동화하면 분석 팀과 기술 팀의 수동 프로세스가 사라져 시간이 대폭 절약되므로 위험을 줄일 수 있습니다. 자동화에 주력할 영역은 다음과 같습니다.

- 데이터 프로비저닝을 간소화할 수 있는 데이터 액세스 및 데이터 준비 파이프라인 구축

- 분석 기술 및 머신 러닝의 분석 및 파이프라인 구축 프로세스 자동화
- 모델의 배포 및 모니터링
- 모델 정확도에 대한 모니터링 지표를 포함한 모델 성능 모니터링
- 시간의 추이에 따른 기반 입력 데이터 특성 벡터 분포의 변화 모니터링
- 모델의 시스템 안정성 모니터링
- 모델링 재작성 유도
- 모델 관리(예: 자동화된 워크플로 사용)
- 대시보드를 사용하여 모델 성능을 모니터링

이러한 유형의 자동화를 실현하려면 목적에 맞게 설계된 새로운 기술 솔루션이나 도구를 배포해야 합니다.

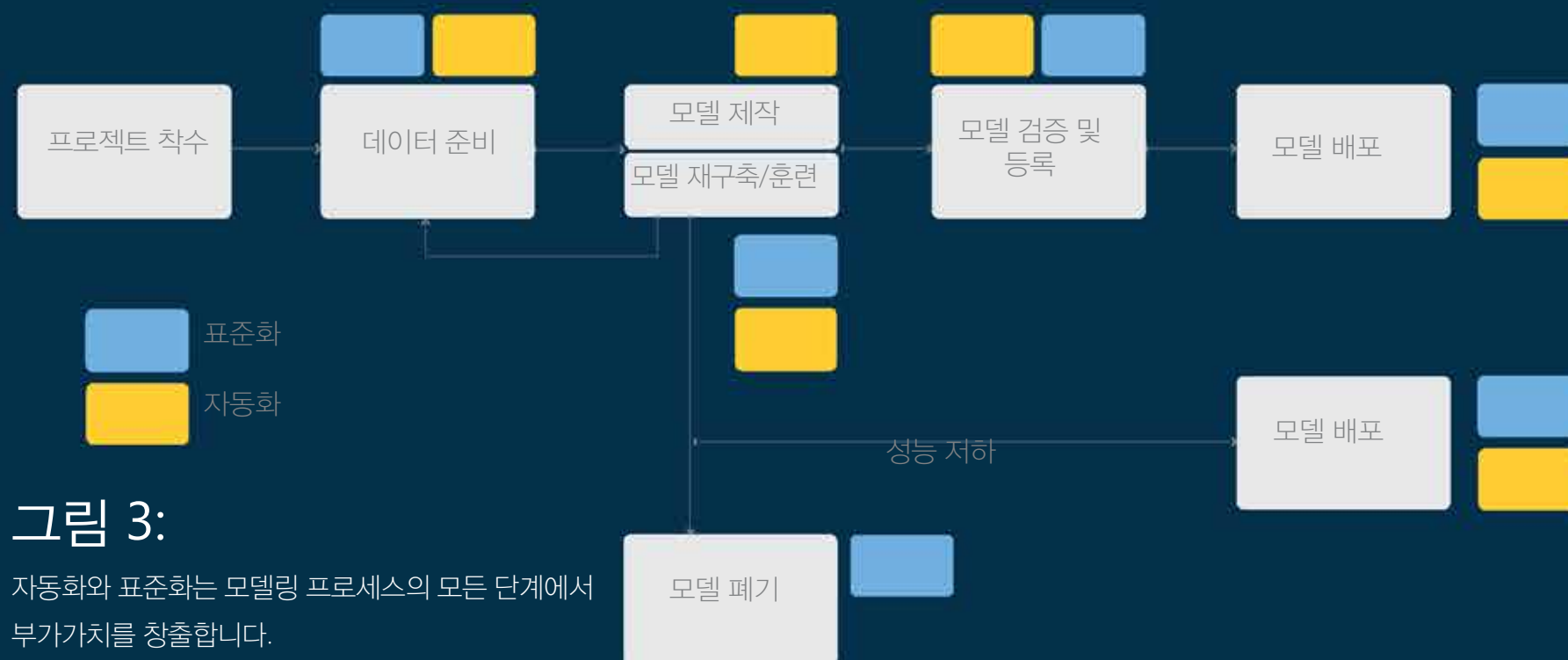


그림 3:

자동화와 표준화는 모델링 프로세스의 모든 단계에서 부가가치를 창출합니다.

5단계: 효과적인 모델 거버넌스 구현

다수의 기업들, 특히 다수의 코딩 언어를 사용하고, 수동 프로세스에 의존하며, 모델 배포 프로세스에 대한 가시성이 부족한 기업들이 거버넌스를 구현하는 데 어려움을 겪습니다. 그러나 문화적 공감대, 자동화, 표준화라는 ModelOps의 세 가지 기본 요건이 충족되면 매우 효과적인 모델 거버넌스를 구현할 수 있습니다. 예를 들어 다음과 같은 작업을 수행할 수 있습니다.

- 우수 모델과 후보 모델을 나란히 비교하여 평가합니다. 비즈니스 문제 해결 능력을 기준으로 분석 모델을 비교해보고, 해당 모델을 제작하는 데 사용된 언어와 관계없이 항상 최고의 모델을 선정합니다.

- 데이터 평가 빈도와 실행 중인 버전을 파악하고, 경보를 자동화하여 성능이 기준 이하로 하락한 경우 이를 즉시 인지하며, 적절한 시기에 쉽게 모델을 리트레이닝, 수정 또는 폐기함으로써 우수 모델의 성능을 항상 최적화합니다.
- 인종, 성별, 연령 또는 사회 경제적 수준 등 편향성을 유발할 수 있는 특성을 중심으로 입력 데이터 및 예측 결과를 검사하여 모델의 공정성을 평가합니다.
- 모델의 세부 정보를 비롯해 모델에 사용되는 데이터, 전사적으로 사용되는 데이터 등 각 모델의 기록과 계보를 추적합니다.

- 분석 라이프사이클과 관련된 KPI를 모니터링하여 프로세스가 비즈니스 속도에 맞게 진행되는지 확인하고 그렇지 않은 부분을 파악하여 개선합니다.

다시 말해서, 이런 데이터를 직접 수집하지 않고도 사용 중인 모든 모델과 각 모델의 성능, 그리고 각 모델에 취해진 조치에 관해 총체적인 가시성을 확보할 수 있습니다. 이러한 총체적인 가시성을 바탕으로 분석 모델의 상태와 기능, 목적, 효용성 및 가치를 파악할 수 있습니다. 결과적으로 분석 기술을 제대로 관리하여 분석 모델의 비즈니스 효과를 유지할 수 있습니다.





ModelOps는 장기전입니다. 따라서 제대로 된 준비가 필요합니다.

여느 좋은 솔루션들이 그러하듯 ModelOps 역시 하룻밤 새에 구현할 수 없습니다. ModelOps를 도입하는 과정은 시간이 걸리며, 조직의 규모, 기존의 시스템과 프로세스 및 문화에 따라 저마다 각기 다른 문제가 발생합니다. 그러나 ModelOps는 각 단계마다 데이터 사이언티스트와 IT 부서가 협력하여 분석 프로그램을 보완하는 새로운 방식을 제시하므로 훨씬 더 큰 비즈니스 가치를 창출할 수 있습니다.

이 백서에서 제시하는 단계를 거치다 보면 팀이 협업하고 모델을 배포하는 방식이 개선되기 시작할 것입니다. 수익에 실질적인 영향을 미치는 KPI와 비즈니스 의사 결정 능력이 점진적으로 개선되는 효과도 기대해볼 만합니다. 예를 들어 더 많은 모델을 더 빨리 배포하고, 더 많은 모델을 전사적으로 동시에 사용하면 최신 모델이 많아지므로 이를 토대로 더 나은 의사 결정을 신속히 내릴 수 있습니다.

뿐만 아니라 새로운 기술을 사용하여 모델의 변화와 성능 저하를 더 빨리 파악하고 모델을 더 빨리 업데이트하거나 수정함으로써 사용 중인 모델에서 더 많은 가치를 창출할 수 있습니다.

그 밖에도 작업을 자동화하고 가시성을 지원하며 분석 모델 데이터를 중앙에서 보관하고 관리하는 소프트웨어를 배포함으로써 문제를 일으키는 기존의 수동 프로세스를 꾸준히 혁신할 수 있습니다. 이러한 혁신은 지속적인 개선과 책임감으로 이어져, 모든 직원이 한 가지 목표에 집중할 수 있는 밑거름이 될 것입니다. 한 가지 목표란 더 우수한 품질의 모델을 더 빨리 배포하고 최상의 성능을 유지하여 가치 있는 비즈니스 결정을 내리는 것입니다.

ModelOps로 분석 기술을
운용하는 방법을
자세히 알아보십시오.



자세한 내용은 sas.com/kr에서 확인하실 수 있습니다.