

## AI/ML 기반 모델 개발, 과제와 해결 방안은?

ModelOps 도입으로 효율적인 협업 환경을 구축하고,  
모델 라이프사이클을 개선하십시오.



# 목차

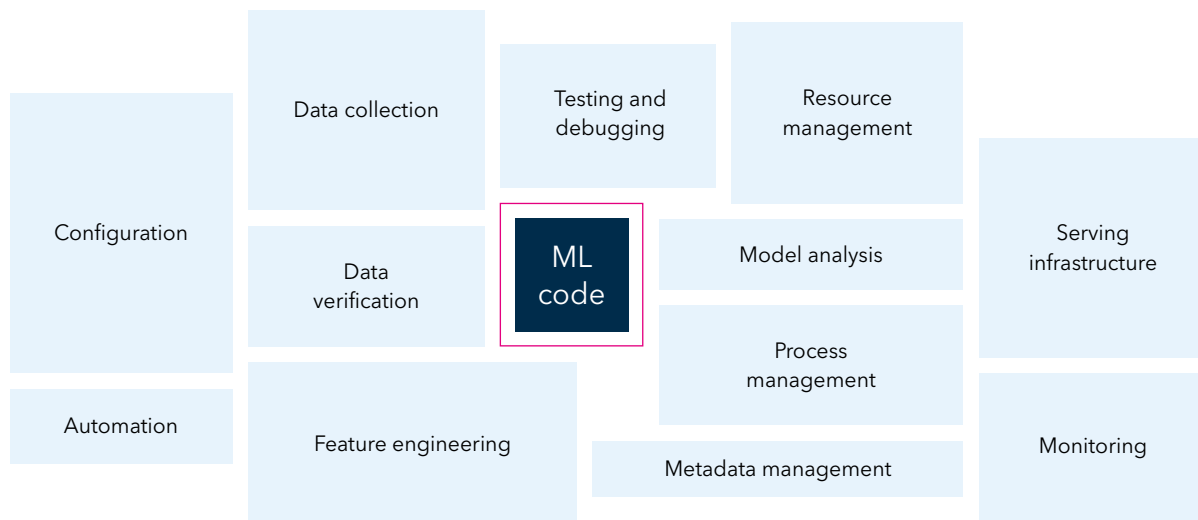
1. 최근 AI/ML 프로젝트 운영화(Operationalization)의 현황 및 문제점 .....	3
2. 운영화 이후 모델 운영관점에서 관측되는 문제점 .....	4
3. 당면한 AI/ML 문제의 해결 방안 .....	5
4. ModelOps 프레임워크와 X-Ops와의 연관 .....	7
5. ModelOps 구현 방법 .....	8
6. ModelOps의 도입 효과 .....	9



데이터 수집에서 모델 개발과 배포에 이르기까지 복잡한 과정을 보다 효율적으로 관리할 수 있는 방법은 없을까요? 체계적인 모델 관리로 생산성을 개선하고 모델의 품질을 보증하며, 데이터 과학자와 IT 팀원 모두가 모델 상태를 확인하면서 협업해 나갈 수 있는 시스템이 필요하지 않으신가요? 모델의 라이프사이클 관리를 향상시키고, 일상 업무에 더 많은 분석 모델을 더 빠르게 적용할 수 있게 해주는 ModelOps(모델 옴스)에 대해 알아보고 분석을 통해 더 나은 가치를 창출하는 방법을 알아보십시오.

## 1. 최근 AI/ML 프로젝트 운영화(Operationalization)의 현황 및 문제점

2010년 초반부터 AI/ML에 높은 관심을 보인 많은 회사들이 AI/ML을 활용한 모델 개발을 시도, 수행 또는 운영화하고 있습니다. 그러나 2019년 IDC 자료에 따르면, 모델의 50% 미만이 운영화 되고, 90% 이상의 모델이 배포에 3개월 이상 걸리며, 44%의 모델은 운영화 되어 실업무에 적용되는데 7개월 이상 소요된다고 합니다. 최근 국내에서도 “오픈 소스로 개발된 모델의 재학습, 평가 등의 모델관리가 힘들어요”, “모델 배포하기가 힘들고 오래 걸려요”, “서로 다른 오픈 소스 플랫폼이 많고 복잡해서 관리가 힘들어요” 등의 목소리가 흘러나오며, 이러한 문제를 해결하기 위해 플랫폼의 도입을 시도 하는 기업들이 늘어나고 있습니다. 하지만 단순 플랫폼의 도입만으로 이를 해결할 수는 없습니다. 위에 언급한 상황의 근본적인 문제는 AI/ML 도입 시기 초반에 많은 프로젝트에서 시도한 방법, 즉 예측 모델의 생성 및 도구 중심의 개발에 있기 때문입니다. 생성된 모델의 운영화를 위해서는 데이터의 수집, 점검, 테스트, 인프라, 시스템 표준 구성, 모델 관리를 위한 프로세스 및 규칙 정의 등의 많은 요소가 필요합니다. 초반에는 이러한 요소들을 무시 또는 간과하고 단순 운영화를 위한 필요 구성, 예를 들어, 스코어링 함수 생성 및 내부 호출만을 이용했습니다. 그러나 주변 구성 요소를 체계적으로 도입, 연계 및 관리하지 않으면 기술적 부채(그림1, 주1)가 발생되어 위에 언급한 문제가 나타나게 됩니다.



출처 : Elements for ML systems. Adapted from Hidden Technical Debt in Machine Learning Systems.

그림1) AI/ML의 기술적 부채

주1) 기술적 부채: 현 시점에서 시간 소요가 많은 더 나은 접근방식을 사용하는 대신, 쉬운(제한된) 솔루션을 채택함으로써 발생하는 추가적인 재작업 비용을 반영하는 소프트웨어 개발의 한 관점(출처 : 위키백과)

## 2. 운영화 이후 모델 운영관점에서 관측되는 문제점

AI/ML의 예측 모델은 과거 데이터에서 발생하는 패턴을 찾아 그 패턴 기반의 예측을 통해 의사결정과 실행을 합니다. 하지만 데이터의 학습 후에 발생된 데이터에서의 패턴이 항상 과거와 동일 할 수 없습니다. 예를 들어보겠습니다. 물건 구매의 세대 간 구매 패턴은 586, 밀레니엄 및 MZ 세대별로 다르죠. MZ세대의 물건 구매 패턴 데이터로 미래에 새롭게 명명될 ZZ세대(주2)의 구매 패턴을 정확하게 예측할 수가 있을까요? MZ세대의 구매 예측보다는 정확도가 떨어질 것이고, 정확도의 저하는 매출 감소와 비용 증대로 나타나게 될 것입니다. 이를 해결하기 위해서는 모델의 생애 주기, 즉 해결 또는 개선할 문제를 인식하는 것에서 시작하여, 데이터 수집과 전 처리, 학습, 모델의 개발과 배포, 운영 모니터링 및 성능 평가 등의 단계가 상호 끊임없이 연결되어 새롭게 생성된 ZZ세대의 행동 및 구매 데이터를 기반으로 예측 모델을 지속적으로 개선하여 운영화 해야 합니다.

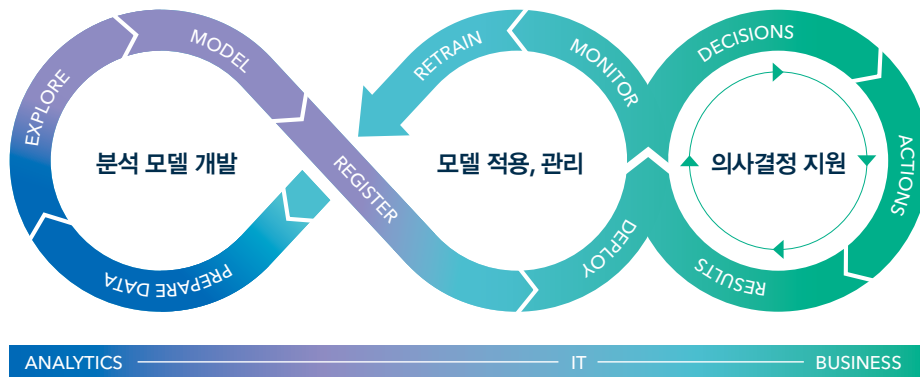


그림 2) AI/ML 모델의 생애 주기

또한 이러한 생애 주기는 체계적인 관리 절차(Governance)에 따라 모델의 성능이나 적용된 업무의 성과를 모니터링하여, 성능이나 성과 저하 시 재학습(주3) 또는 재모델(주4) 작업 후 생성된 새로운 모델을 운영화함으로써, 모델의 성능과 업무 성과를 지속적으로 유지해야 합니다.(그림 3)

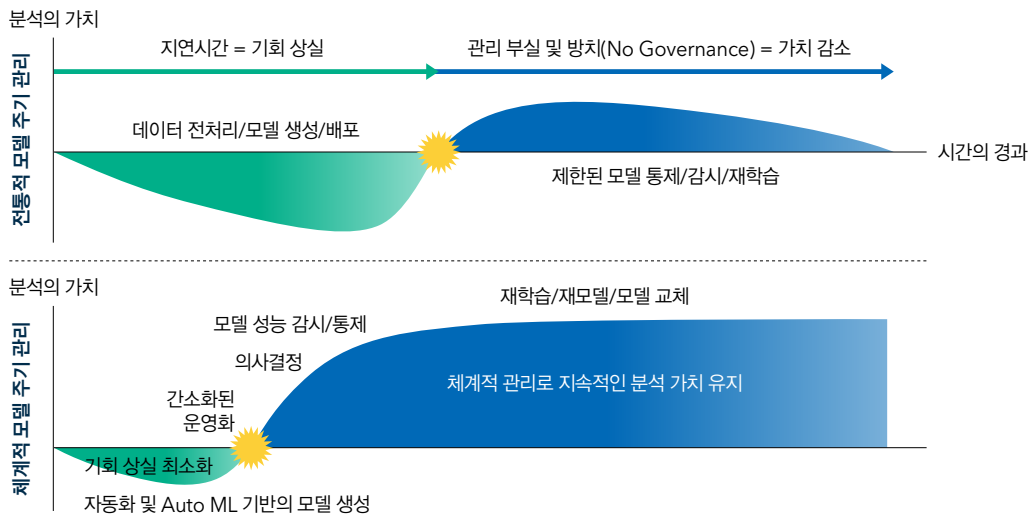


그림 3) 전통적 모델 주기 관리와 통제 기반의 관리

주2) ZZ세대 : 현재의 사회에서 언급되는 MZ세대와는 다른 미래의 세대로 MZ세대와는 다른 행동 및 의사결정 패턴을 갖는 세대의 의미로 임의로 ZZ세대로 표현했음

주3) 재학습 : 이전에 학습된 데이터 구조를 변경하여 새로운 데이터로 학습함으로써 새로운 모델을 생성하는 방식

주4) 재모델 : 연관된 속성의 추가 등을 통해 새로운 데이터를 생성하여 모델을 개발하는 방식

### 3. 당면한 AI/ML 문제의 해결 방안

AI/ML의 주요 영역인 신속한 모델 개발 및 운영화, 그리고 모델 관리와 관련된 문제점의 해결 방안에 대해 알아보겠습니다. 첫번째로 신속한 모델 개발 및 운영화에서 모델 개발은 시민 데이터 사이언티스트의 양성을 위한 환경 구축을 의미합니다. 즉, 최대한로우 코드 노코드로 개발하는 것을 지향합니다. ([데이터 문해력과 시민 데이터 사이언티스트의 필요역량 내용 보기](#)) 그리고 신속한 모델 배포와 운영화를 위해서는 최신의 S/W 개발 방법론인 DevOps의 적용이 필요합니다. 두번째로 모델 관리의 부재로 발생하는 문제점은 모델을 체계적으로 관리할 수 있는 방법론, 즉 Model Governance(이하 모델 거버넌스)를 적용해야 합니다. 모델 생성 과정 중 필요한 데이터의 조회 및 선택, 선택한 데이터에 대한 탐색, 품질 점검 및 정제 등의 전처리에 많은 시간이 소요됩니다. 이 시간을 최소화하기 위해서는 DataOps를 적용해야 합니다. 또한 모델 거버넌스 방법론을 구체화하여 실현하고, DataOps와 DevOps를 통합 및 연계하기 위해서는 ModelOps가 필요합니다. (그림 4)

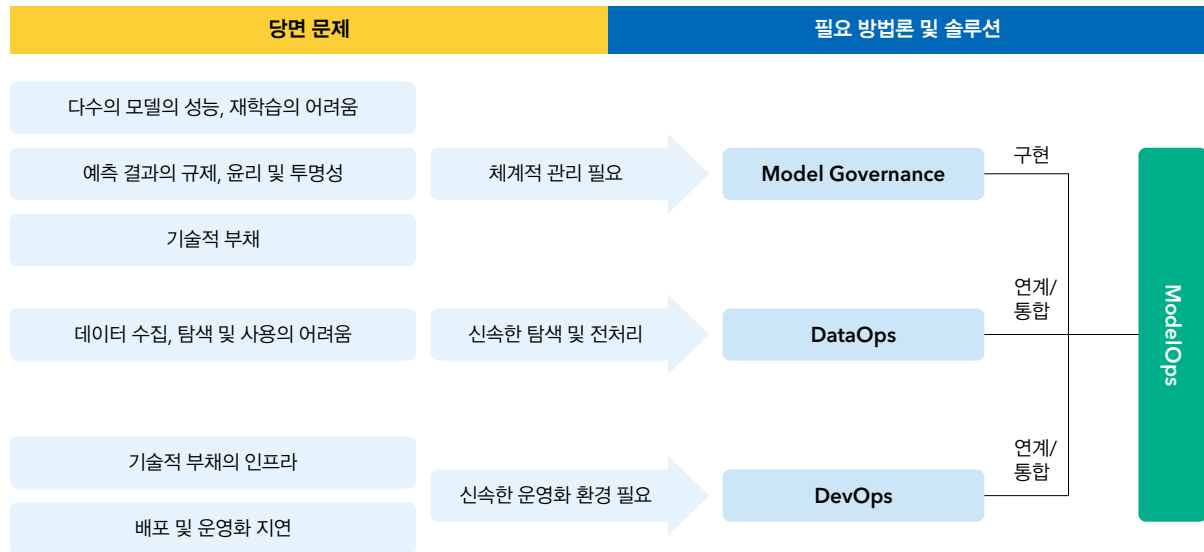


그림 4) 당면 문제와 해결 방안

DevOps는 소프트웨어 개발 방법론 중 하나로, 소프트웨어 개발과 운영을 가상 또는 조직 내 하나의 팀으로 결합하여 개발 주기를 단축하고 운영 중 발생된 문제나 개선 사항을 즉각적으로 개발에 반영하여 품질을 향상시키는 일련의 실행 방안입니다. 여기서 소프트웨어 개발에는 Agile 개발 방법론(주5)을 적용합니다.

주5) Agile 개발 방법론 : 효율적인 팀 협업을 중요시하는 방법론으로, 개발자와 고객이 서로 밀접한 협조를 통해 작은 기능들을 우선 순위에 따라 짧은 주기로 반복하여 개발 및 배포를 수행함으로써, 변화에 빠르고 유연하게 대응하며, 고객의 요구 사항에 신속하게 대응할 수 있다는 장점이 있음

DataOps는 데이터 분석의 품질을 개선하고 데이터의 분석 시간을 단축하기 위한 방법론으로, Agile 방법론을 통합하여 비즈니스 목표에 맞춰 개발을 촉진하고 통계적 프로세스 제어를 사용하여 데이터를 모니터링합니다.

모델 거버넌스는 조직 내에서 모델의 개발, 배포, 모니터링, 유지보수 등의 생명 주기를 통제하고 관리하는 프로세스, 조직 및 역할의 정의와 체계를 말합니다. 이는 모델의 안정성, 신뢰성, 안전성, 투명성을 보장하고, 규제 요구사항과 윤리적 측면을 준수하기 위해 필요합니다. 또한, 모델 거버넌스는 모델의 성과와 결과를 추적하고 측정하여 조직의 의사결정에 영향을 미치는 모델들을 효과적으로 관리하는 것을 목표로 합니다.

ModelOps는 모델 거버넌스에서 정의된 프로세스, 역할, 책임 및 규칙 등의 체계를 구현합니다. 또한 생성된 모델이 신속하게 등록, 배포 및 모니터링을 수행하고 재학습 및 재모델을 할 수 있도록 하며, 규제와 추적 가능성 등을 구현할 수 있게 해줍니다. (그림 5)



그림 5) ModelOps를 위한 Framework(구성 요소)

## 4. ModelOps 프레임워크와 X-Ops와의 연관

가트너에서는 ModelOps를 ‘머신러닝, 지식 그래프, 규칙, 최적화 및 언어를 포함한 광범위한 AI 및 의사 결정 모델의 거버넌스 및 수명주기 관리에 중점을 두고 있으며, AI 모델의 재조정, 재학습 또는 재구축을 지원하여 AI 기반 시스템 내에서 모델의 개발, 운영 및 유지 관리 사이에 중단 없는 프로세스를 제공하는 방법론’이라고 정의합니다. SAS는 이러한 정의에 맞추어 관련 컨설팅과 플랫폼 및 제품들을 제공합니다.

마이크로소프트 또는 구글 등에서는 ModelOps와 유사한 개념으로 MLOps (Machine Learning/Operation)를 정의하고, 이를 위한 프레임워크 및 제품을 제공하고 있습니다. 일부에서는 MLOps와 유사한 방법론을 AIOps라고도 합니다. AIOps와 MLOps를 동일한 것으로 보아도 무방합니다.

구글에서는 MLOps를 “ML 모델을 효과적으로 개발, 배포, 운영, 유지 보수하는 방법론”이라고 했으며, 가트너에서는 “DevOps 원칙에 따라 기계 학습(ML) 모델을 프로덕션에 적용하는 프로세스”로 정의했고, 가트너가 정의한 ModelOps의 일부분이라고 했습니다. 모델 거버넌스가 포함된 ModelOps가 MLOps 보다 포괄적인 방법론이라고 할 수 있겠습니다.

X-Ops의 시작은 DevOps입니다. DevOps 방법론에 데이터 수집, 처리, 저장 및 품질 등의 자동화 기능과 검색, 탐색 등을 포함한 것이 DataOps이고, 이 DataOps 기반에 모델의 생성, 배포, 운영 및 성능 관리 등의 기능과 각 기능들 간의 플랫폼 자동화를 포함한 것이 MLOps이며, MLOps에 모델 거버넌스를 융합한 것이 ModelOps라 볼 수 있습니다. 또한 DataOps 관련 작업을 Data Engineering이라고 하며, MLOps 또는 ModelOps 영역의 작업을 ML Engineering이라고 합니다. (그림 6) 각각의 작업을 수행하는 역할을 각각 Data Engineer와 ML Engineer라고 하는데, 요즘 데이터 사이언티스트보다 인기 있는 직업입니다.

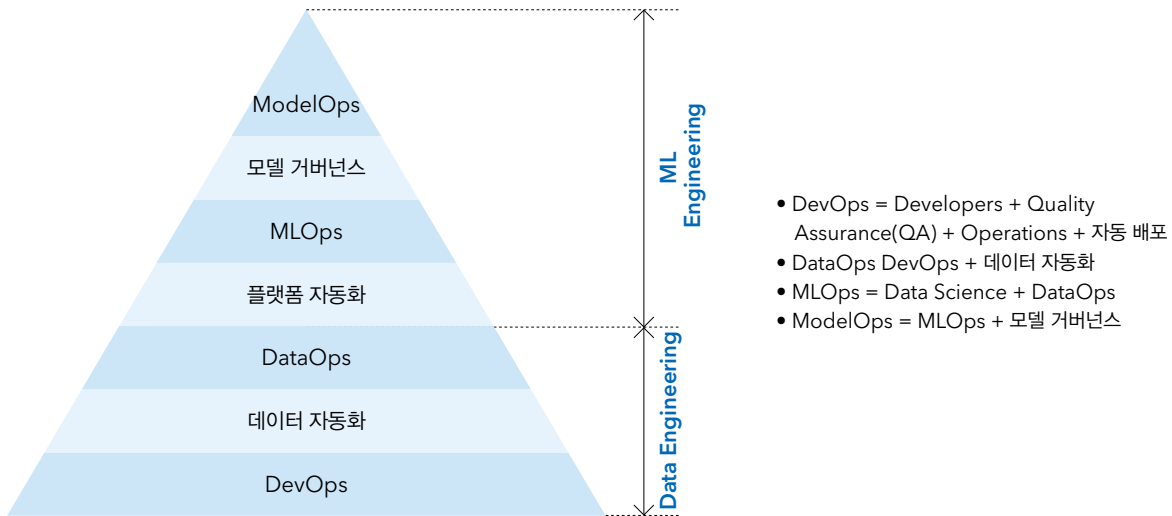


그림 6) DevOps 부터 ModelOps까지

ModelOps 프레임워크의 X-Ops는 데이터 영역에서는 DataOps, 배포 영역은 DevOps, 모델의 개발, 실행 및 성능 모니터링의 영역을 MLOps또는 AIOps, 프레임워크 전 영역은 ModelOps 가 적용된다고 요약할 수 있습니다.(그림 7).

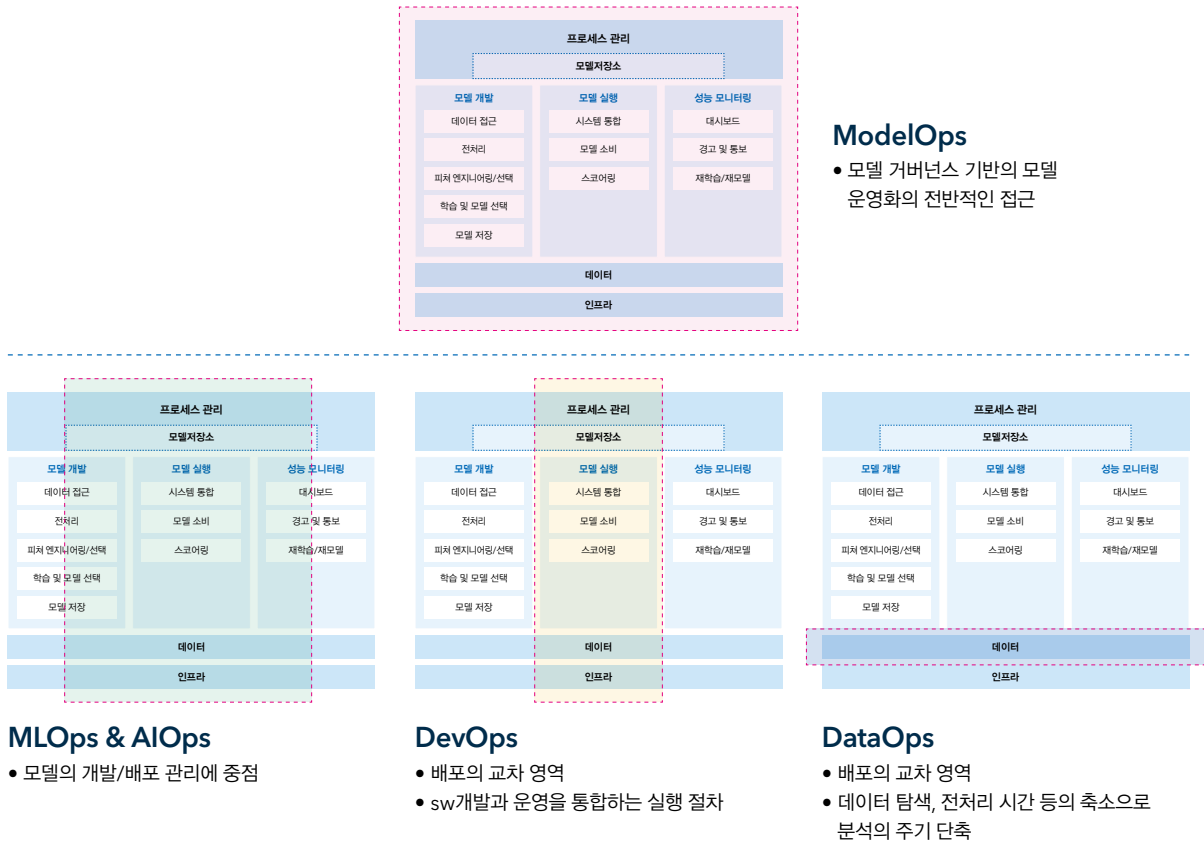


그림 7) ModelOps Framework에서의 X-Ops

## 5. ModelOps 구현 방법

ModelOps를 도입 및 실행하기 위해서는 다양한 역할과 활동이 필요합니다. 그 주요 내용은 아래와 같습니다.

- **업무 전문가(Business)** : 해당 분야의 전문 지식을 기반으로 개선 과제 및 목적 설정, 모델의 적용여부에 대한 손익을 분석하고 검토
- **데이터 사이언티스트(Data Scientist)** : 통계 및 수학적 전문 지식을 기반으로 모델 생성, 모니터링 등을 수행하고 전문 지식을 기반으로 시민 데이터 사이언티스트를 지원
- **시민 데이터 사이언티스트(Citizen Data Scientist)** : 업무 지식을 토대로 시각화 및 자동화 도구를 이용하여 데이터 사이언티스트가 주로 수행하는 모델의 생성, 모니터링 등을 수행
- **분석 책임자 (Head of Analytics)** : 업무 개선 결과와 모델의 생애 주기를 이해하고 프로젝트의 전 과정을 이끌며, 모델의 적용 여부 등의 주요 의사 결정
- **CxO** : 전사의 목표, 윤리 및 규제 등에 대한 제언, 지원 및 관련 의사 결정
- **IT/DevOps** : 모델 배포 및 성능 모니터링
- **Data Engineer/DataOps** : 데이터 연계, 통제 및 품질 관리

각각의 역할과 책임 기반의 모델 생애 주기 프로세스, 즉 모델 생성 및 적용 프로세스, 모델 모니터링 프로세스 등을 정의하여 ModelOps를 구현하게 되면, 이러한 다양한 부서의 담당자들이 투명하게 모델의 운영현황을 점검하며 효율적으로 협업해 나갈 수 있다는 장점이 있습니다. 각 부서 관점에서 필요한 사항들을 다면적으로 반영할 수 있다는 측면에서 매우 효과적이고 생산적인 접근법이라고 하겠습니다.

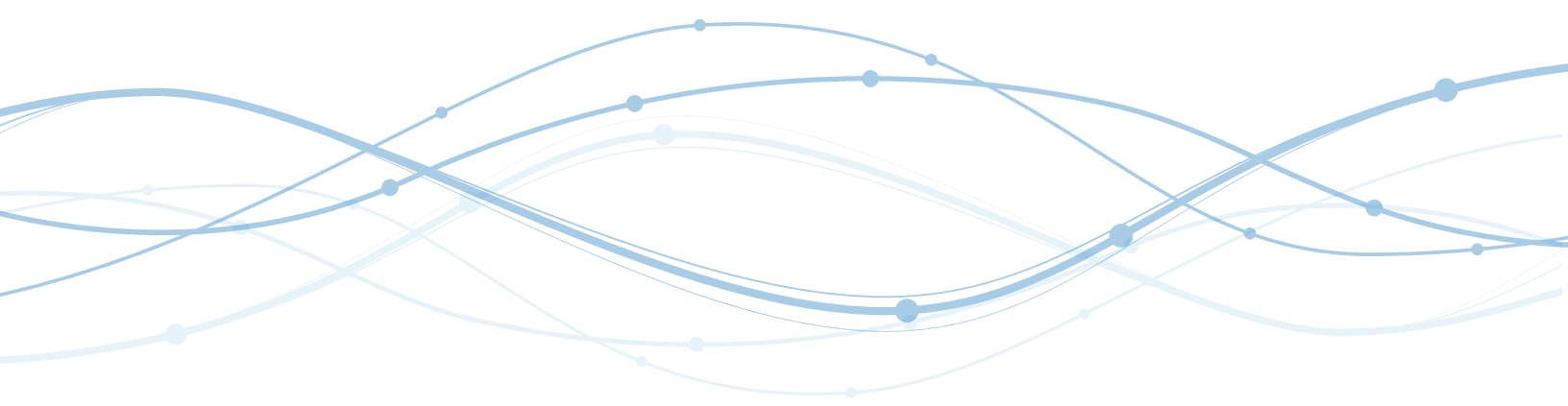
ModelOps의 실현을 위해서는 우선 이를 실행할 수 있는 플랫폼이 필요합니다. SAS Viya와 같은 데이터 분석 플랫폼 위에 각 영역을 지원하는 도구들을 구동해야 합니다. 다양한 오픈 소스 또는 밴더의 소프트웨어가 필요하며 이를 통합하기 위한 도구들도 있습니다. 반면 SAS Viya와 같은 플랫폼은 ModelOps를 구현하기 위한 기능들이 하나의 플랫폼으로 통합되어 있습니다. 이 외에, 써드파티 개발 플랫폼, 또는 외부 앱에서 오픈 소스로 개발된 모델을 통합 리파지토리에 등록 및 조회할 수 있는 Viya 서비스를 REST-API를 통해 제공하여 통합된 ModelOps를 실현할 수 있습니다. 아울러, Github와 같이 코드 리파지토리와 연계 기능도 제공합니다.

## 6. ModelOps의 도입 효과

ModelOps를 도입하면 사업 관점에서는 첫번째, 새로운 사업 또는 기존의 사업에 최신 기법을 적용한 신속한 모델 운영화로 기업의 경쟁 우위를 선점할 수 있습니다. 두번째, 지속적인 모델의 개선으로 업무의 성과가 개선되어, 매출 향상 또는 비용 감소로 나타납니다. 세번째, 정확한 데이터 기반의 분석 및 투명성 있는 모델 기반 예측으로 신속하고 신뢰성 있는 의사 결정이 가능하며 위험 또한 최소화할 수 있습니다.

IT 관점에서는 첫번째, 체계적인 프레임워크 기반에 안정적인 모델 개발과 운영이 가능합니다. 두번째, 지속적인 모니터링 및 관리를 통해 모델의 품질을 향상시킵니다. 세번째, 명확한 역할과 책임 기반의 분석 생애 주기 전과정을 최대한 자동화함으로써 신속한 개발 및 모델 배포를 지속할 수 있습니다.

지금까지 ModelOps의 필요성과 역할, 구현 방법 등에 대해 설명드렸습니다. 사업 관점과 IT 관리 관점 모두에서 기업의 경쟁력을 높여주는 ModelOps로 업무 생산성을 높이고 분석의 가치를 최대화하시기 바랍니다.



SAS Korea 정보는 [sas.com/kr](https://sas.com/kr)을 참조하시기 바랍니다.



SAS 및 기타 모든 SAS Institute Inc. 제품이나 서비스 이름은 미국과 기타 국가들에서 등록된 상표이거나 SAS Institute Inc.의 상표입니다. ® 기호는 미국에서 등록되었음을 나타냅니다.  
기타 브랜드 및 제품명은 각 해당 회사의 상표입니다. Copyright © 2023, SAS Institute Inc. All rights reserved. 113648.0923