



ARTIFICIAL INTELLIGENCE 501

Lesson 3

AI in the Enterprise

Learning Objectives

You will be able to:

- 데이터 과학 워크 플로우의 단계를 확인.
- AI 분야에서 주요 역할과 기술 세트를 확인.
- AI 팀 구성 방법을 설명.
- 일반적인 데이터 과학 오해를 식별.
- 배포 후 모델 유지 관리의 구성 요소를 확인.



DATA SCIENCE WORKFLOW

Data Science Workflow

Problem Statement

어떤 문제를 해결하려고합니까?

Data Collection

문제 해결을 위해 어떤 데이터가 필요합니까?

Data Exploration & Preprocessing

모델에서 사용할 수 있도록 데이터를 어떻게 정제 해야 합니까?

Modeling

문제를 해결하기 위해 어떤 모델을 구축해야합니까?

Validation

문제를 해결 했습니까?

Decision Making & Deployment

이해 관계자와 의사 소통하거나 생산에 투입 할 수 있습니까?

Problem Statement

어떤 문제를 해결하려고합니까?

- 데이터 과학자들은 먼저 해결할 문제를 식별해야합니다.
- 영향력있는 기회를 파악하려면 비즈니스에 대한 지식이 필요합니다.
- 기술 지식은 올바른 질문을하고 가능한 것을 알기 위해 필요합니다.

Data Collection

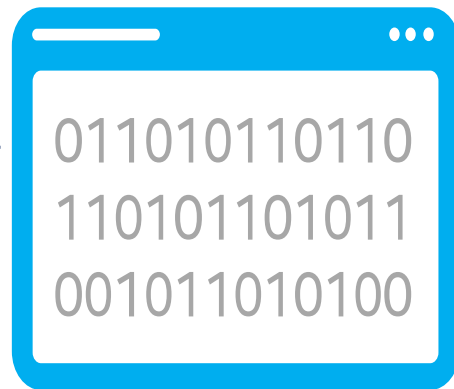
문제 해결을 위해 어떤 데이터가 필요합니까?

- 문제를 해결하는 데 필요한 데이터를 확인하고 수집해야 합니다.
- 여러 소스의 데이터를 수집하고 통합하려면 데이터 및 엔지니어링 기술이 필요합니다.

Data Exploration and Preprocessing

모델에서 사용할 수 있도록 데이터를 어떻게 정제 해야 합니까?

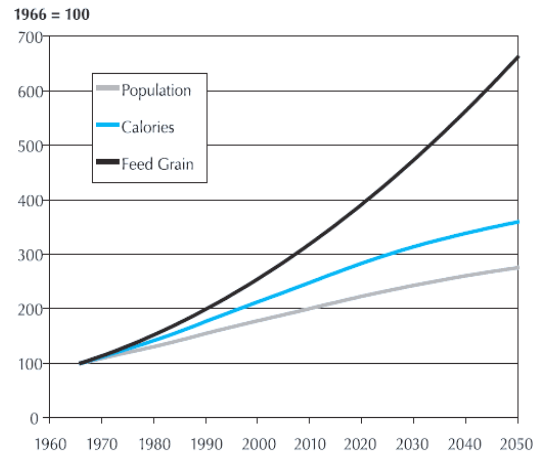
- 모델링을 위해 사용 가능한 형식이 되도록 데이터를 정리하고 처리해야 합니다.
- 탐구는 데이터 내의 중요한 요소를 식별하고 데이터 품질 문제를 식별하는 데 필요합니다.
- 데이터, 엔지니어링 및 통계 기술은 데이터를 적절히 처리하고 추론을하는 데 필요합니다.



Modeling

문제를 해결하기 위해 어떤 모델을 구축해야합니까?

- 복잡성, 필수 데이터, 속도 및 성능과 같은 여러 요인이 모델링에 포함됩니다.
- 이 단계에는 엔지니어링, 모델링 및 통계 기술이 필요합니다.



Validation

문제가 충분히 해결 되었습니까?

- 원래 문제가 해결되었는지 확인하려면 검증이 필요합니다.
- 모델 성능을 정확하게 측정해야 합니다.
- 통계 및 모델링 기술뿐만 아니라 도메인 지식은 결과가 비즈니스 문제와 일치하는지 확인하는 데 필요합니다.



Decision Making and Deployment

이해 관계자와 의사 소통하거나 생산에 투입하십시오.

- 비즈니스 결정을 내릴 필요가 있거나 제품을 생산에 투입해야 비즈니스에서 프로젝트의 가치를 볼 수 있습니다.
- 이를 위해 도메인 지식, 의사 소통 및 스토리 텔링 기술이 필요합니다.
- 코드를 백엔드 소프트웨어 시스템에 통합하려면 엔지니어링 기술이 필요합니다.

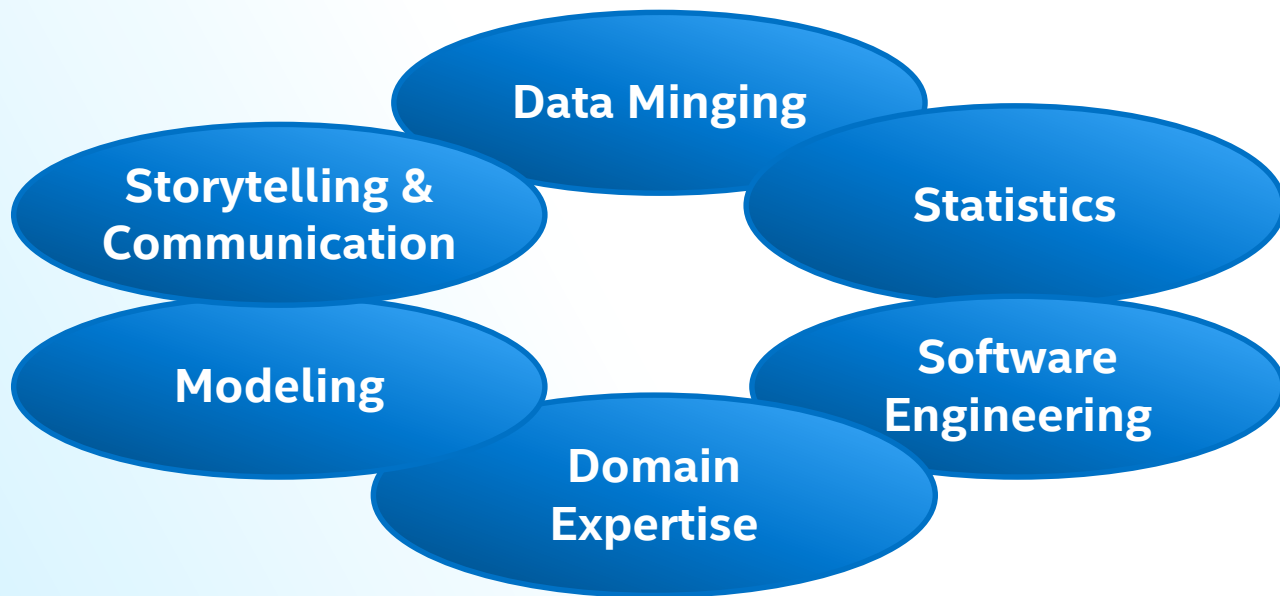




DATA SCIENCE SKILL SET

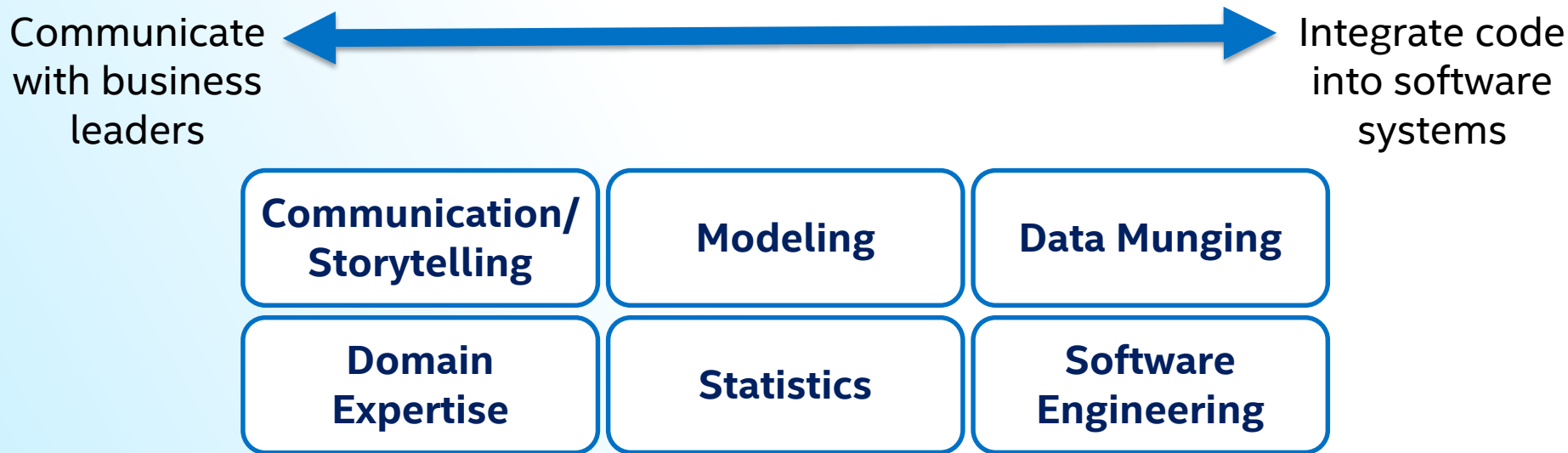
Data Science Skill Sets

데이터 과학 팀은 성공하기 위해 다양한 기술이 필요합니다.



Data Science Team Skills

데이터 과학 팀은 성공하기 위해 다양한 기술이 필요합니다.



Communication, Storytelling, Domain Expertise

비즈니스 요구 사항을 이해하고 해결 방법을 알려줍니다.

- 도메인 전문 지식을 통해 프로세스 및 비즈니스 문제를 이해하여 비즈니스를 지원하십시오.
- 의사 결정자들에게 그들의 생각을 뒷받침하도록
- 설득하십시오.
- 복잡한 개념을 명확하게 전달하고 이야기를 전하십시오.



Modeling and Statistics

데이터를 사용하여 모델을 통해 예측을하고 통계를 사용하여 예측의 유효성을 평가합니다.

- 회귀에서 DL에 이르는 다양한 모델링 기법을 사용하십시오.
- 통계를 사용하여 한 모델의 실적을 평가합니다.
- 실험을 설계하고 A / B 테스트를 수행하십시오.



Data Munging/Software Engineering

지저분한 데이터를 깨끗하고 사용 가능한 데이터로 변환하고 모델을 배포하기 위한 소프트웨어 시스템을 구축하십시오.

- 원시 데이터는 복잡하고 구조가 복잡 할 수 있습니다.
- 데이터는 사용하기 전에 조작되고 데이터 베이스에 저장되어야합니다.
- 모델을 배치해야 합니다.

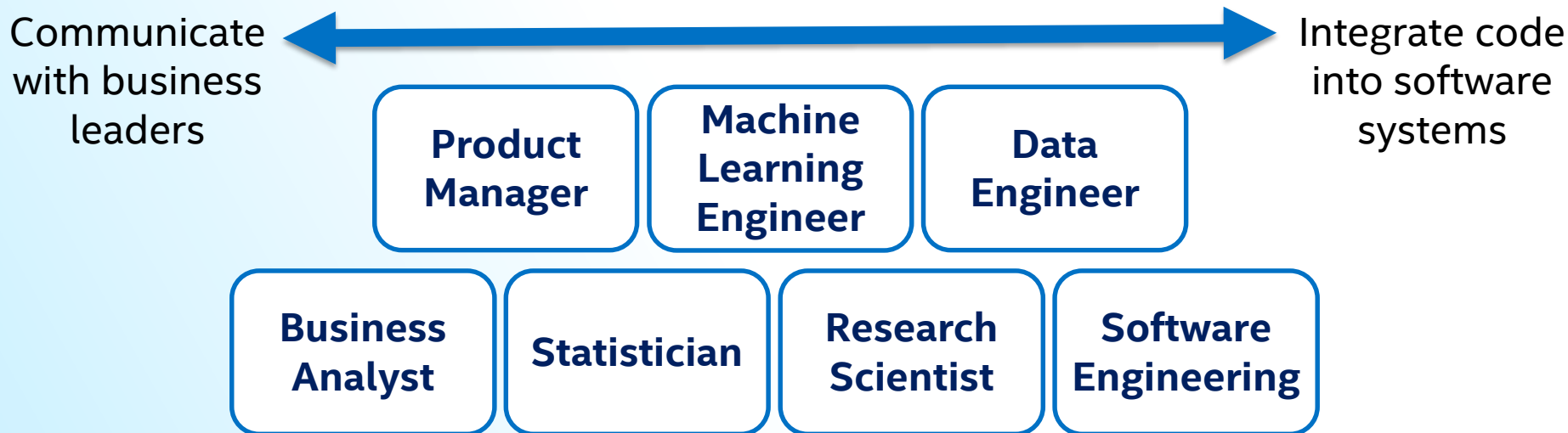




DATA SCIENCE ROLES

Roles on Data Science Teams

Roles have evolved that fit on different places on this spectrum.



Business Analysts

비즈니스 분석가는 의사 결정권자와 상호 작용합니다.

- 보고서를 작성하고 통찰력을 제공하십시오.
- 주요 제품 KPI를 표시하는 대시 보드를 만듭니다.
- 분석을 수행하여 신제품 / 기능의 비즈니스 영향을 판별하십시오.
- Excel *, PowerBI * 및 Tableau *는 사용 된 도구의 예입니다.



Product Managers

제품 관리자는 비즈니스 요구 사항을 얻습니다.

- 비즈니스 아이디어를 제품 아이디어로 변환하십시오.
- 비즈니스 문제를 해결할 가능성을 결정하십시오.
- 새로운 제품이나 모델이 주요 비즈니스 지표에 미치는 영향을 고려하십시오.
- 프로젝트와 작업의 우선 순위를 정하십시오.
- 사용 된 도구의 예 : Microsoft Project *, Trello * 및 JIRA *.



Statisticians

통계학자는 모델의 타당성을 결정합니다.

- 통계 개념을 적용하여 필요한 데이터의 양을 결정합니다.
- 특이점과 추세를 탐구합니다.
- 결과가 통계적으로 중요한지 여부를 결정합니다.
- 사용 된 도구의 예 : R *, SAS *, Python *.



Machine Learning Engineers

기계 학습 엔지니어는 많은 양의 고차원 데이터가 관련된 문제를 해결합니다.

- 기계 학습 기술을 적용하십시오.
- 모델을 정확하고 빠르게 만드는 엔지니어링에 중점을 둡니다.
- 사용 된 도구의 예 : Python, R 및 MATLAB.



Research Scientists

연구 과학자들은 출혈이 심한 분야에서 문제를 해결합니다.

- 빅데이터 및 머신러닝에서 가장 힘든 문제를 해결하십시오.
- 특정 하위 분야의 전문가.
- 알고리즘이 어떻게 작동하는지 이해하십시오.
- 필요한 경우 데이터 과학 팀과 상호 작용하는 별도의 연구 팀의 일원이 될 수 있습니다.
- 사용 된 도구의 예 : Python *, MATLAB *, R *, C ++ 및 Java *



Data Engineers

데이터 엔지니어는 데이터 처리, 스토리지 및 인프라를 구축합니다.

- 데이터베이스 전문가.
- 속도, 안정성 및 크기 사이의 균형을 알아야 합니다.
- 데이터 정리를 자동화 하십시오.
- ETL (추출, 변환, 로드) 파이프 라인을 구축하여 일반 종단에서 데이터를 사용할 수 있도록 합니다.
- 사용 된 도구의 예 : Java *, SQL 및 noSQL.



Software Engineers

소프트웨어 엔지니어는 코드 최적화 및 배포를 담당합니다.

- 프로덕션으로 코드를 가져옵니다.
- 코드 분할 및 버그를 탐지하는 테스트를 작성하십시오.
- 모델 코드를 관리 할 수 있는지 확인하십시오.
- 사용 된 도구의 예 : Python *, Ruby *, C ++ 및 Java *.



Skills and Roles

	Business Analyst	Product Manager	Statistician	Machine Learning Engineer	Research Scientist	Data Engineer	Software Engineer
Communication/Story	X	X					
Domain Expertise	X	X					
Modeling		X	X	X	X		
Statistics			X	X	X	X	
Data Munging				X	X	X	X
Software Engineering							X



COMMON MISCONCEPTIONS

Misconception #1: Data Science “Unicorn”

모든 분야의 전문가인 데이터 과학자를 유니콘 (unicorns)이라고합니다.

- 성공적인 팀에는 다양한 기술 및 배경을 가진 사람들이 있습니다.
- 어떤 사람들은 의사 소통에 능하며, 다른 사람들은 통계학에 능하다.
- 성공적인 팀은 비즈니스, 과학 및 엔지니어링이라는 세 가지 주요 영역의 전문가를 보유하고 있습니다.

Misconception #2: Research and Algorithms Focus

데이터 과학 팀은 연구 및 알고리즘에만 집중할 수 없습니다.

- 효과적인 팀에는 다음과 같은 메커니즘이 있습니다.
 - 문제 식별
 - 결과 통보
 - 엔지니어링과 협력하여 모델을 생산에 적용하는 방법을 이해합니다.



Misconception #3: Complex and Advanced Systems

가장 복잡한 솔루션이 항상 최선은 아닙니다.

- 팀은 단순 해지기 시작한 후 더 복잡한 모델링 기술로 넘어갈 때 성공하는 경향이 있습니다.
- 복잡한 모델은 더 정확할 수도 있지만 해석하기 어렵고 예측할 수 없는 방식으로 실패 할 가능성이 높으며 유지 관리가 더 어려울 수 있습니다.
- 간단한 시작은 팀이 구축하는 것이 비즈니스 요구 사항과 일치하도록 합니다.

Misconception #4: Industry Differences

기본 모델링 및 데이터 분석 기술은 산업간에 크게 이전 가능합니다.

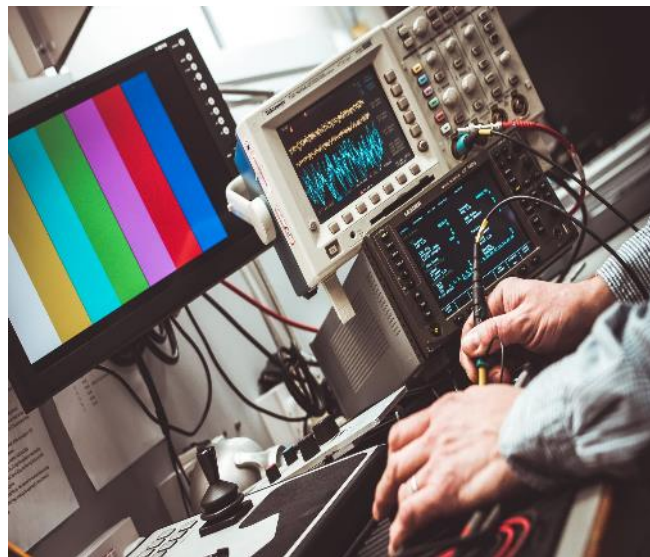
- 도메인 전문 지식은 관련성이 높은 데이터와 해결해야 할 가장 중요한 문제를 이해하는 데 필요합니다.
- 데이터를 정리하고 저장하며 유용한 통찰력을 추출하고 모델링하는 기술은 매우 유사합니다.



Misconception #5: Projects Begin Well-Defined

데이터 과학 프로젝트는 종종 실험적이며
실험적입니다.

- 데이터를 탐구하는 시간을 투자 할 때까지 문제를 해결하는 것이 얼마나 어려운지는 분명하지 않을 수 있습니다.
- 제품 관리자는 기대를 관리하기 위해 팀 및 비즈니스 이해 관계자와 적극적으로 협력해야 합니다.



Misconception #6: Best Prediction Models are Best

예측 능력보다 모델을 선택할 때 더 많은 어려움이 있습니다.

- 일부 모델은 생산에 포함하기에는 너무 느리거나 복잡 할 수 있습니다.
- 일부 모델은 해석이 불가능하고 의사 결정권자와의 판매가 어려울 수 있습니다.





AFTER DEPLOYMENT

After Deploying a Model

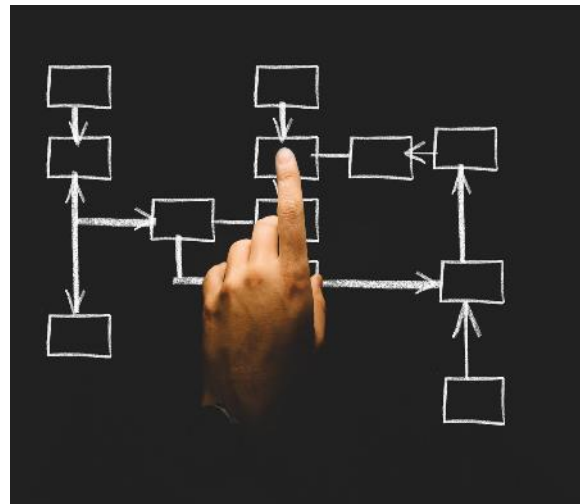
모델이 배치되면 관련 팀은 유용한 모델을 모니터링하고 관리해야 합니다.

- 비즈니스 인텔리전스 팀은 모델 결과를 표시하는 보고서 / 대시 보드를 작성해야 합니다.
- 비즈니스 팀 (예 : 운영, 머천다이징) 또는 고객은 모델 출력을 사용해야 합니다.
- 데이터 과학 팀은 적절한 주기(예 : 월간)로 모델을 업데이트해야 합니다.

Business Intelligence Team

비즈니스 인텔리전스 팀은 모델 예측이 시간이 지남에 따라 어떻게 변하는 지 모니터링합니다.

- 모델 사용 및 출력을 모니터링합니다.
- 주요 성과 지표 및 대시 보드 개발.
- 시각화는 모델을 사용할 기능 팀 (예 : 작업)의 요구 사항과 일치해야 합니다.



Functional Business Team

의사 결정을 향상시키기 위해 대시 보드를
사용하십시오.

- 예를 들어, 마케팅 팀은 변동 모델을 사용하여
고객 그룹에 연락 할시기를 결정할 수 있습니다.
- 예 : 운영 팀은 모델을 사용하여 운송 / 물류
시간을 예측할 수 있습니다.



Data Science Team

지속적인 타당성과 유용성을 보장하기 위해 모델을 모니터하십시오.

- 예를 들어, 마케팅 변경으로 인해 고객 유형이 달라질 수 있으므로 고객 행동에 대한 구형 모델의 정확성이 떨어집니다.
- 비즈니스 조건의 변화 속도에 따라 데이터 과학 팀은 적절한 빈도로 모델을 업데이트해야 합니다.
- 새로운 데이터로 모델 재교육.