Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

aws training and certification

Machine Learning Engineering on AWS (KO)

Student Guide

버전 1.0.3

200-MLMLEA-10-KO-SG

DO NOT COPY
ssleeeee1@gmail.com

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

본 내용은 Amazon Web Services, Inc.의 사전 서면 허가 없이 전체 또는 일부를 복제하거나 재배포할 수 없습니다. 상업적인 복제, 임대 또는 판매는 금지됩니다.

본 과정에 대한 수정 사항이나 피드백, 문의사항이 있으면 https://support.aws.amazon.com/#/contacts/aws-training 을 통해 연락해 주십시오.

모든 상표는 해당 소유자의 자산입니다.

DO NOT COPY

ssleeeee1@gmail.com

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

목차

모듈 0: 과정 소개	4
모듈 1: AWS의 기계 학습(ML) 소개	18
모듈 2: ML 관련 당면 과제 분석	51
모듈 3: ML용 데이터 처리	88
모듈 4: 데이터 변환 및 특성 엔지니어링	132
모듈 5: 모델링 방식 선택	195
모듈 6: ML 모델 훈련	243
모듈 7: ML 모델 평가 및 튜닝	279
모듈 8: 모델 배포 전략	322
모듈 9: AWS ML 리소스 보호	374
모듈 10: 기계 학습 운영(MLOps) 및 자동화된 배포	403
모듈 11: 모델 성능 및 데이터 품질 모니터링	449
모듈 12: 과정 요약	507

DO NOT COPY
ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

DO NOT COPY
ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

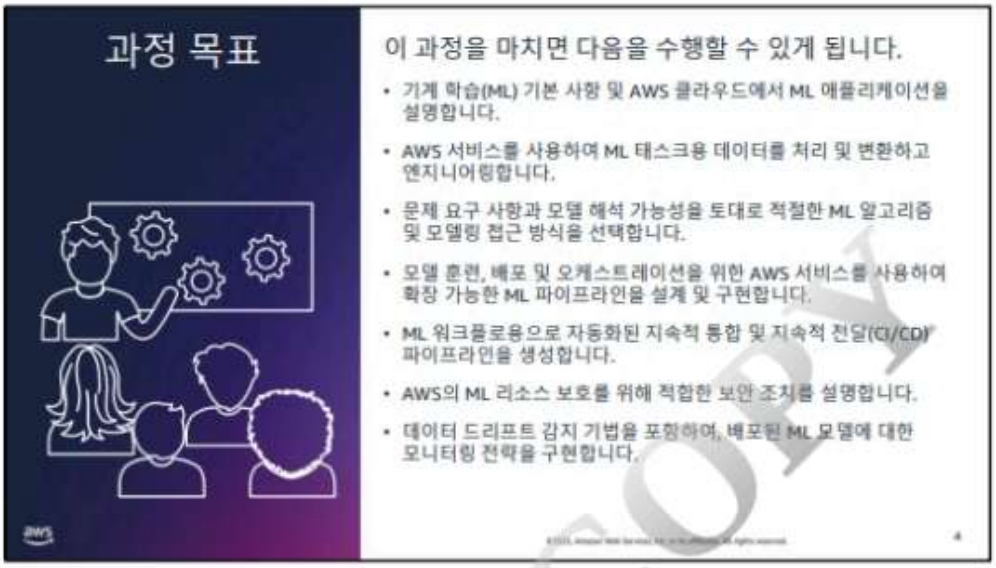
DO NOT COPY
ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

DO NOT COPY
ssleeeee1@gmail.com

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



DO NOT COPY
ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

일정

1일 차
- 모듈 1: AWS의 기계 학습(ML)
- 모듈 2: ML 도전 과제 분석 + 활동
- 모듈 3: ML을 위한 데이터 처리
- 모듈 4: 데이터 변환 및 피처 엔지니어링 + 활동 + 실습

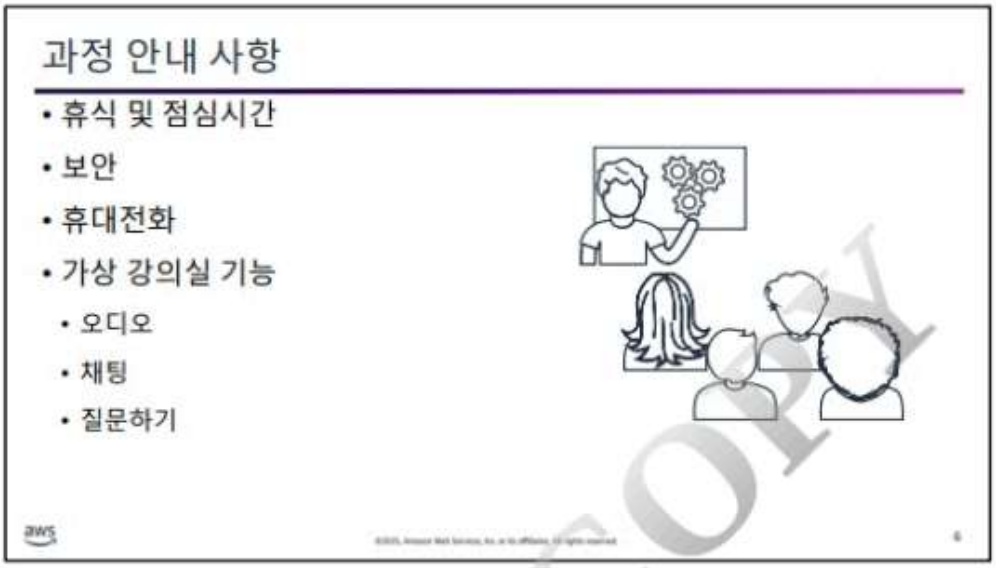
2일 차
- 모듈 5: 모델링 방식 선택 + 활동
- 모듈 6: ML 모델 훈련 + 실습
- 모듈 7: ML 모델 평가 및 튜닝 + 실습
- 모듈 8: 모델 배포 전략 + 실습

3일 차
- 모듈 9: AWS ML 리소스 보호
- 모듈 10: ML 운영(MLOps) 및 자동화된 배포 + 활동 + 실습
- 모듈 11: 모델 성능 및 데이터 품질 모니터링+ 실습

DO NOT COPY
ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



강사는 휴식 시간 및 점심시간 일정을 설명합니다. 가상으로 교육에 참여하는 경우 담당 강사와 함께 가상 강의실 소프트웨어 기능을 검토하십시오. 담당 강사에게 이러한 기능을 사용하여 강의에 참여하는 방법에 대한 설명을 들을 수 있습니다.

DO NOT COPY
ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

과정 사전 조건

- 기본적인 ML 관련 개념 및 알고리즘 유형에 친숙함
- 일반 데이터 과학 라이브러리 및 Python 프로그래밍 언어에 대한 실무 지식
- 클라우드 컴퓨팅 개념에 대한 기본적인 이해
- AWS 클라우드에 친숙함

유용하지만, 필수 조건 아님:

- SQL 및 관계형 데이터베이스에 친숙함
- Git 등의 버전 제어 시스템 사용 경험

이 과정은 AWS에서 기계 학습 모델을 구축, 배포 및 운영하는 방법을 알아보려는 전문가용입니다. 이는 현직 그리고 AWS 스택을 사용한 경험이 거의 없는 훈련 중인 ML 엔지니어도 해당될 수 있습니다. 이 교육의 혜택을 받을 수 있는 다른 역할로는 DevOps 엔지니어, 개발자, 시스템 운영(SysOps) 엔지니어 등이 있습니다.

사전 조건

이 과정의 참석자는 아래를 갖추는 것이 좋습니다:

- 지도 및 비지도 학습, 회귀, 분류, 클러스터링 알고리즘 등의 기본적인 ML 관련 개념에 친숙함
- NumPy, Pandas, Scikit-Learn 등 일반 데이터 과학 라이브러리 및 Python 프로그래밍 언어 관련 실무 지식
- 기본적인 클라우드 컴퓨팅 이해 및 AWS 클라우드에 친숙함

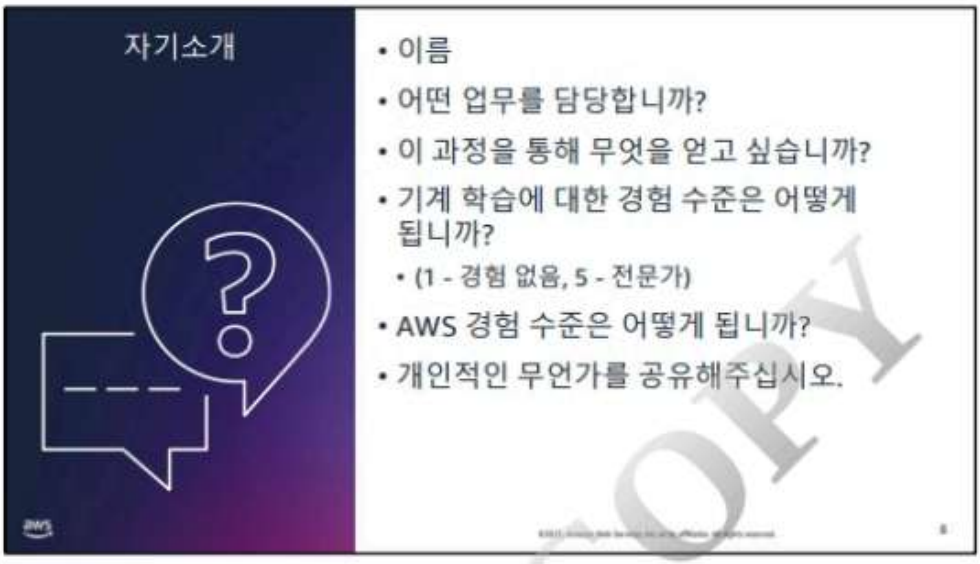
권장하지만, 필수 조건 아님

- SQL 및 관계형 데이터베이스에 익숙할 것을 추천하지만 필수 아님
- Git 등의 버전 제어 시스템 사용 경험이 있음 유리하지만 필수 아님

NO COPY ssleeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



자기소개를 하면 강사가 수강생의 숙련도 수준과 수강생이 이 과정을 통해 얻고자 하는 것을 파악하는 데 도움이 됩니다.

DO NOT COPY
ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

본 과정에 적극적으로 참여하여 기계 학습(ML) 실무 경험을 쌓으십시오. 토론, 채팅, 화이트보드, 활동을 통해 지식을 점검하고 생각을 공유하십시오. 데모를 통해 기계 학습(ML)의 다양한 측면에 대한 추가적인 인사이트를 얻을 수 있습니다. 실습을 통해 학습한 내용을 강화하고 적용하여 실제 기계 학습(ML)을 준비합니다.

질문도 언제든지 환영합니다.

DO NOT COPY
ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

DO NOT COPY
ssleeeee1@gmail.com

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

받은 편지함에서 강사가 보낸 환영 이메일을 확인하십시오. 이 이메일에서 강의용 고유한 수강생 등록 URL을 찾을 수 있습니다. 이 URL 링크를 사용하여 계정을 생성하거나 기존 AWS Builder Lab 계정에 로그인하십시오. AWS Builder Lab에서 실습 환경, Lab Guide 및 Student Guide에 액세스할 수 있습니다.

DO NOT COPY
ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

이제 AWS Builder Lab에 로그인되었을 것입니다. 여기서 실습 가이드 및 수강생 가이드에 액세스할 수 있습니다. 가이드는 eVantage Bookshelf(VitalSource)에 있습니다. Lab Guide 및 Student Guide 버튼은 AWS Builder Lab 대시보드에 있습니다. 강의가 시작될 때까지는 실습과 버튼이 비활성화됩니다.

수업이 시작된 후 두 버튼 중 하나를 선택하여 가이드를 이용할 수 있습니다. 기존 eVantage Bookshelf(VitalSource) 계정으로 로그인하거나 새 계정을 생성하라는 메시지가 표시됩니다. eVantage Bookshelf(VitalSource)에 로그인하면 강의의 수강생 가이드 및 실습 가이드에 액세스할 수 있습니다. 가이드는 온라인으로 액세스하거나 다운로드할 수 있습니다. 이러한 가이드를 참조하여 과정을 진행하고 교육 후에는 가이드를 참조용으로 활용하십시오.

DO NOT COPY
ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

실습 요구 사항

- 다음 운영 체제를 실행하는 컴퓨터:
  - Windows
  - macOS
  - Linux: Ubuntu, SUSE 또는 Red Hat
- 권장 웹 브라우저:
  - Google Chrome
  - Mozilla Firefox
  - Microsoft Edge
- HTTPS를 사용하여 인터넷을 탐색할 수 있는 안정적인 인터넷 연결
- AWS Builder Lab에 등록:
  - 광고 차단기 및 스크립트 차단기 끄기

14

DO NOT COPY
ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

이 모듈에서 알아본 데이터 준비 및 특성 추출 관련 개념을 잠시 생각해 보시기 바랍니다. 이 과정에서 다룬 자료 또는 이 과정에서 배운 기술을 실제 업무에 적용하는 방법과 관련하여 궁금한 점이 있으면 지금 질문을 할 수 있습니다.

학습은 꾸준히 진행해야 하는 과정이며 특히 급격하게 발전하고 있는 기계 학습 분야에서는 새로운 정보를 지속적으로 습득해야 합니다. AWS에서는 지속적인 자기 계발 지원을 위해 다음과 같은 여러 리소스를 제공합니다.

- **AWS 설명서**: 서비스 및 모범 사례와 관련된 자세한 정보 확인 가능
- **AWS 교육 및 자격증**: 기술 역량 강화 및 전문 자격증 취득 지원
- **AWS Support**: AWS 서비스 관련 기술 지원 제공

더 정확하게 파악해야 하는 부분이 있거나 세션 후에 추가로 문의할 사항이 있는 경우 반드시 질문을 하시기 바랍니다. 수강생의 피드백은 향후 이 과정을 진행하는 학습자들을 위해 교육 자료와 세션을 개선하는 데 유용하게 활용됩니다.

DO NOT COPY ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

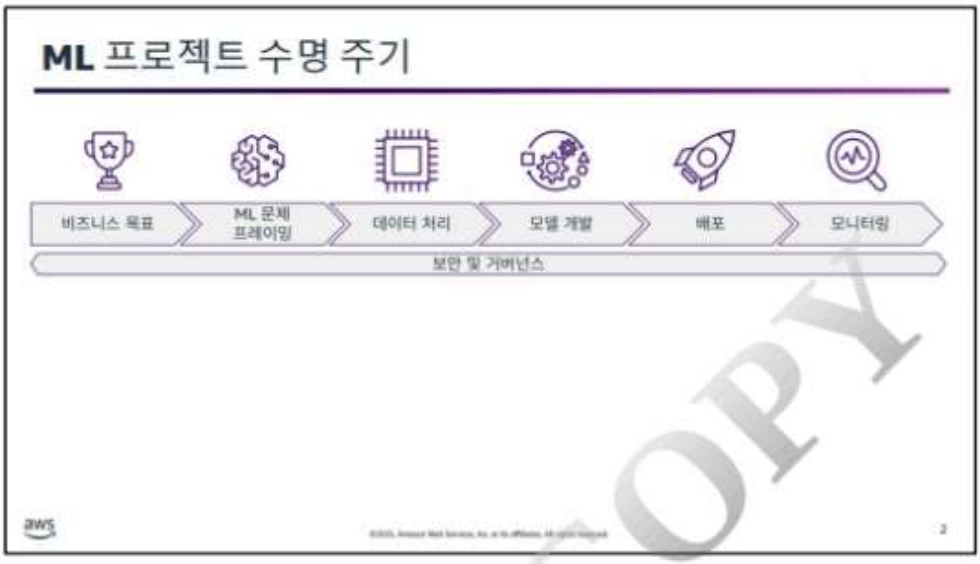
이 모듈에서는 AWS의 기계 학습(ML) 엔지니어링의 전반적인 소개를 제공합니다. 구체적으로는 알고리즘, 모델, 그리고 딥 러닝과 생성형 AI를 비롯한 고급 주제를 포함한 ML 핵심 개념을 살펴봅니다. ML 솔루션에 적합한 비즈니스 문제를 식별하고, AWS 서비스를 활용하여 ML 개발을 간소화하는 방법을 이해하기 위한 지식을 습득합니다.

이 모듈을 완료하면, ML 관련 기본 사항을 이해하여 AWS의 특화된 구현 방식을 살펴볼 준비가 될 것입니다. 이런 기본 사항은 ML 프로젝트에서 Amazon SageMaker AI 기능을 손쉽게 사용하는데 도움이 될 것입니다. 이 모듈에서는 AWS에서 ML과 관련된 실제 도전 과제를 해결할 수 있도록, ML의 이론적인 개념과 실제 적용을 모두 설명합니다.

DO NOT COPY
ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



기계 학습(ML) 수명 주기는 ML 솔루션을 개발하는 엔드 투 엔드 프로세스를 이해 할 수 있는 프레임워크입니다. 다이어그램에는 비즈니스 목표와 ML 문제 프레이밍에 중점을 둔 주요 단계가 있습니다. 이러한 초기 단계는 조직의 목표에 맞게 ML 프로젝트를 맞추고, 올바른 문제를 해결하는데 있어 중요합니다.

SageMaker AI는 데이터 준비부터 모델 배포와 모니터링에 이르기까지, 전체 ML 수명 주기를 지원합니다. 이런 수명 주기를 이해하면 각 단계에서 SageMaker AI 도구를 효율적으로 사용할 수 있습니다. 예를 들어 SageMaker Studio에서는 모든 단계를 수행할 수 있는 통합 환경을 제공합니다. 그리고 Amazon SageMaker Data Wrangler, Amazon SageMaker Feature Store 등의 기능을 사용하면 데이터 준비 및 피처 엔지니어링을 원활하게 진행할 수 있습니다.

이 모듈을 진행하면서, ML 수명 주기의 각 단계에서 활용 가능한 AWS 서비스를 살펴봅니다. 이 정보는 강력하고, 확장 가능한 ML 솔루션을 구축하는데 도움이 될 것입니다.

DO NOT COPY
ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

이 모듈의 아젠다는 ML 기본 사항과 AWS 환경의 전반적인 이해를 제공합니다.

A. ML 소개: 이 섹션에서는 ML 및 ML의 주요 구성 요소를 정의합니다. 이는 AWS ML 서비스를 이해할 토대를 마련합니다.
B. ML 알고리즘 및 모델: 이 섹션에서는 다양한 유형의 알고리즘과 모델을 살펴보고 SageMaker AI 등의 AWS 서비스를 사용하여 어떻게 이들을 구현할 수 있는지 설명합니다.
C. 책임감 있는 ML: 이 섹션에서는 윤리적인 AI 사례, 그리고 어떻게 AWS가 책임감 있는 ML 개발을 지원하는지 설명합니다.

이 구조는 기본적인 개념부터 고급 주제로의 논리적인 진행을 제공하며, AWS 및 SageMaker AI 기능의 맥락에서 이루어집니다.

DO NOT COPY ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

기계 학습은 컴퓨터 시스템이 경험을 통해 특정 태스크의 성능을 개선하도록 돕는 AI의 하위 집합입니다. AWS에서의 ML은 데이터에서 인사이트를 도출하고 예측하고 의사 결정 프로세스를 자동화하는 데 사용됩니다.

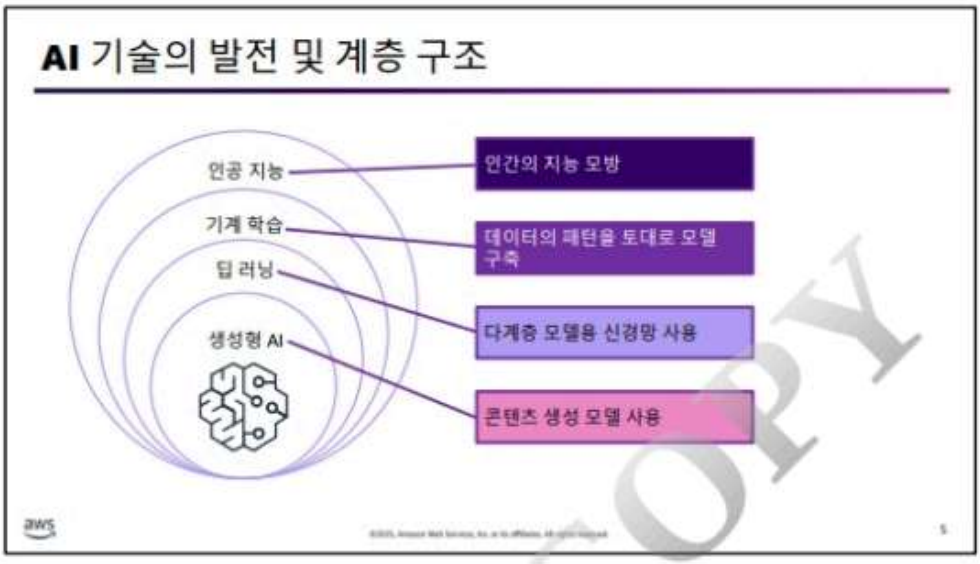
ML은 추천 시스템에서 사기 행위 탐지, 자연어 처리, 컴퓨터 비전까지 폭넓은 애플리케이션에 활용됩니다. AWS는 Amazon SageMaker AI를 중심으로 하는 포괄적인 ML 서비스 제품군을 제공하여, 기업이 대규모로 ML 모델을 구축, 훈련, 배포할 수 있도록 도와줍니다.

이 섹션에서는 ML 수명 주기, 알고리즘 유형, AWS 서비스가 전체 ML 워크플로를 지원하는 방식 등의 ML 기본 사항에 대해 설명합니다. 이러한 지식은 프로젝트에서 AWS의 ML 기능을 효율적으로 사용할 수 있도록 합니다.

DO NOT COPY ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



슬라이드의 다이어그램에는 AI 기술 간의 중첩 관계를 보여주는데, 각 기술이 이전 기술을 기반으로 어떻게 구축되었는지를 확인할 수 있습니다. 가장 광범위한 수준인, AI는 인간의 의사 결정 과정을 시뮬레이션 할 수 있는 모든 지능형 시스템을 포함합니다. 이는 후속 혁신의 토대를 형성합니다.

기계 학습은 통계적 예측 및 패턴 인식을 전문적으로 수행하는 AI의 하위 집합입니다. **딥 러닝**은 인간의 뇌 기능을 모방한, 복잡한 신경망을 사용하여 방대한 양의 데이터를 처리합니다. 신경망 아키텍처는 이미지 인식, 자연어 처리 등의 복잡한 태스크를 다룰 수 있는 딥 러닝을 가능하게 합니다.

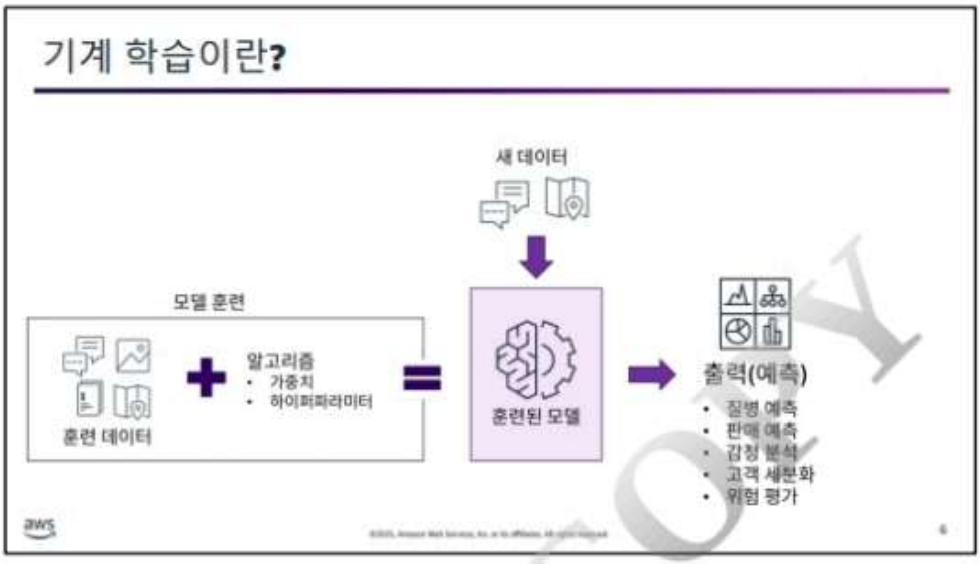
이 계층 구조의 핵심에는 가장 최신이자, 전문적인 형태의 **생성형 AI**가 있습니다. 생성형 AI는 이전의 모든 계층을 기반으로 텍스트, 이미지, 코드 및 기타 미디어에서 새 콘텐츠를 생성할 수 있습니다. 이 기술은 어떻게 각 AI 계층이 계속 발전하면서 더욱 정교하고 창의적인 애플리케이션에 적용 가능한지 보여주었는지를 보여줍니다.

비즈니스가 모든 AI 및 ML 기술을 사용하는데 돕도록, AWS 서비스가 각 영역을 어떻게 지원할 수 있는지 탐색할 시에, 이들의 차이점들을 이해하는 게 중요합니다.

DO NOT COPY ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



ML은 컴퓨터 프로그래밍에 접근하는 방식의 근본적인 변화에 해당하는 기술이라 할 수 있습니다. 모든 시나리오에서 명시적인 규칙과 지침을 작성하는 대신, 데이터로 학습을 하여 장기적으로 성능을 개선할 수 있는 시스템을 생성합니다. 이런 패러다임 변화는 인간의 경험과 패턴 인식을 통한 학습 방식을 반영하지만 컴퓨팅 파워를 가지고 매우 더 큰 규모와 더 빠른 속도로 동작합니다.

슬라이드의 다이어그램에 나와 있듯이, 프로세스를 진행하려면 훈련 데이터부터 준비해야 합니다. 이는 학습을 위한 기초 정보로 사용되는 과거의 예제입니다. 이 데이터는 면밀하게 설계된 알고리즘을 통해 처리되는데, 해당 알고리즘에 포함된 2가지 주요 요소는 가중치와 하이퍼파라미터입니다. 가중치는 데이터 내 여러 특성의 중요도를 결정하는 요소이며, 하이퍼파라미터는 모델의 학습 방식을 제어하는 요소입니다. 결과는 과거 데이터에서 학습된 걸 바탕으로 패턴을 인식하고 예측을 할 수 있는 훈련된 모델입니다.

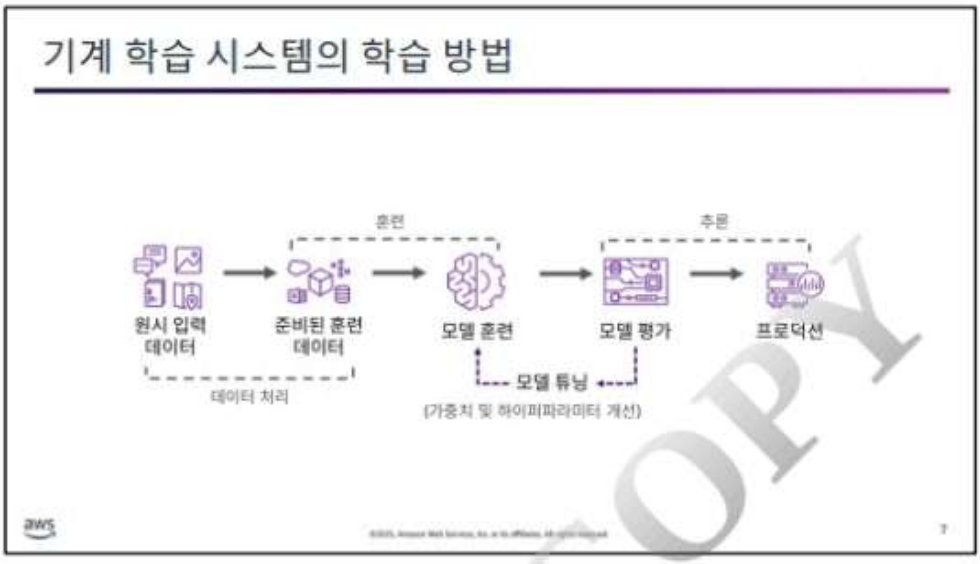
주요 용어는 다음과 같습니다.

- 훈련 데이터: 모델을 훈련하는 데 사용되는 준비된 과거 예제와 그 결과
- 알고리즘: 학습 과정의 기초를 만드는 수학적인 접근법
- 가중치: 데이터 내 여러 특성의 중요도를 결정하는 파라미터
- 하이퍼파라미터: 모델 학습 방식을 제어하는 구성 설정
- 훈련된 모델: 학습한 패턴을 토대로 예측을 수행할 수 있는 결과 시스템
- 새 데이터: 모델이 예측을 추론하는 데 사용하는 이전에는 보지 않았던 정보
- 결과: 질병 판별, 판매 예측 등 모델의 예측 또는 인사이트

이렇게 훈련된 모델을 배포하면 기존 프로그래밍 방식으로는 해결하기가 어려웠거나 불가능했던 광범위한 실제 문제를 해결할 수 있습니다. 이러한 모델은 새 데이터에 맞게 학습 방식을 지속적으로 조정할 수 있으므로 장기적으로 성능이 개선됩니다. 방대한 데이터를 처리하고 복잡한 패턴을 식별하는 능력과 결합한 기계 학습은 질병 예측, 판매 예측, 고객 행동 분석 등의 태스크에서 특히 유용합니다.

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



기계 학습 시스템은 새로운 데이터를 토대로 성능을 반복적으로 개선하는 방식으로 학습을 합니다.

학습 과정에서는 대개 다음 작업이 진행됩니다.

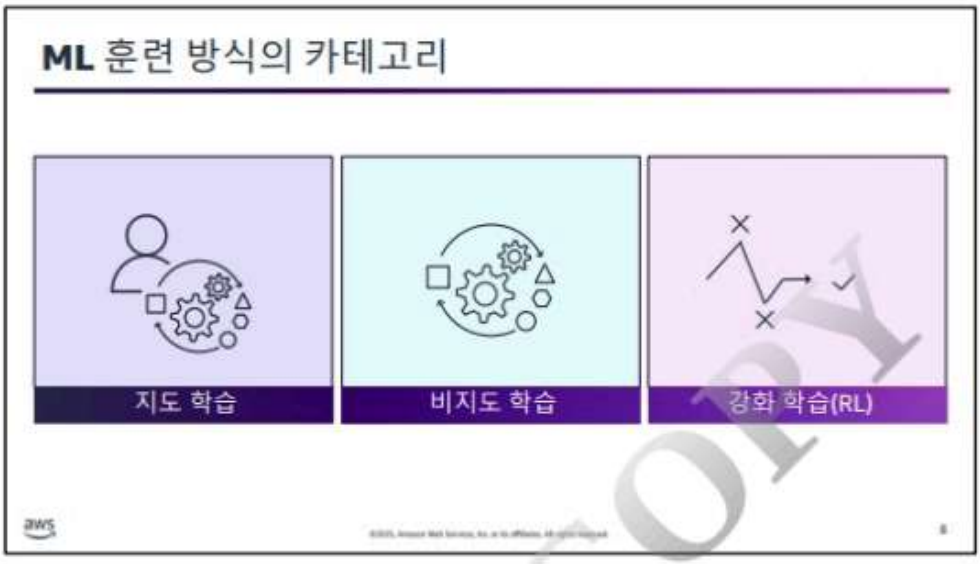
1. 데이터 입력: 데이터를 대량으로 공급하는 프로세스
2. 모델 훈련: 리소스를 많이 사용하는 모델 구축 프로세스
3. 검증: 처음 보는 데이터로 테스트 프로세스
4. 반복: 성능을 기반으로 모델을 개선하는 프로세스
5. 추론: 훈련된 모델을 사용하여 예측을 하는 간소화된 프로세스

AWS의 SageMaker AI에서는 데이터 준비에서 모델 훈련과 배포에 이르기까지 이 프로세스의 모든 단계를 진행할 수 있는 엔드 투 엔드 솔루션이 제공됩니다. AWS에서 ML 솔루션을 구현하려면 이 프로세스를 정확하게 파악해야 합니다.

DO NOT COPY ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



기계 학습 훈련 메커니즘의 3가지 주요 유형은 지도 학습, 비지도 학습 및 강화 학습(RL)입니다. 각 카테고리의 메커니즘 특성은 각기 다른 문제 유형을 해결하는 데 적합합니다.

지도 학습에서는 레이블이 지정된 데이터를 사용하여, 알려진 출력에 입력을 매핑합니다. 일반적인 애플리케이션은 스팸 탐지, 이미지 분류, 판매 예측이 있습니다. 이 방식은 정확도는 높지만, 레이블이 지정된 데이터세트가 많아야 합니다. SageMaker AI에서는 다양한 지도 학습 태스크용 알고리즘이 기본 제공됩니다.

비지도 학습에서는 레이블이 지정되지 않은 데이터를 사용하여, 은닉 패턴이나 구조를 찾아냅니다. 일반적인 애플리케이션은 탐색적 데이터 분석, 고객 세분화, 이상 탐지가 있습니다. 이 방식을 활용하면 인간은 놓칠 수 있는 인사이트를 파악할 수는 있지만, 제공된 결과를 해석하기는 쉽지 않습니다. Amazon Comprehend 등의 AWS 서비스는 텍스트 분석에 비지도 학습을 사용합니다.

강화 학습(RL)에서는 주위 환경과의 상호 작용하고, 결정을 내리고 피드백을 수신하는 방식으로 학습을 합니다. 게임 플레이, 로보틱스, 자율 주행 차량 등 순차적 의사 결정 문제를 해결할 때는 강화 학습이 효율적입니다. 강화 학습 방식은 효율성은 높지만 여러 번 반복해야 하는 경우가 많으며 계산을 많이 수행해야 할 수도 있습니다. 강화 학습(RL) 응용 분야의 한 예가 AWS DeepRacer입니다.

사용 가능한 데이터, 문제의 특성 및 원하는 결과에 따라 각기 다른 알고리즘을 선택하게 됩니다. AWS는 이러한 모든 방식을 지원하는 도구와 서비스를 제공하므로 구체적인 사용 사례에 가장 적합한 방법을 선택할 수 있습니다.

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

이어지는 모듈에서 이러한 방법에 대해 자세히 다룰 예정입니다.

DO NOT COPY
ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

기계 학습은 유용한 기능을 제공하지만 이 기술을 사용하는 과정에서는 여러 가지 도전 과제를 해결해야 합니다.

데이터 관련 도전 과제:
- 제한된 데이터: 영역 특화의 데이터세트 필요
- 데이터 품질: 불완전하거나, 일관성이 없거나, 불필요한 정보가 많은 데이터 해결
- 데이터의 양: 대규모 데이터세트의 효율적인 관리와 처리

기술 구현 관련 도전 과제:
- 피처 엔지니어링: 관련 특성을 선택하여 유의미한 데이터 표현 생성
- 모델 선택: 특정 문제를 해결하는 데 적합한 알고리즘 선택
- 하이퍼파라미터 튜닝: 성능을 최대한 높일 수 있도록 모델 파라미터 최적화

윤리적이고 실용적인 고려 사항:
- 편향성 및 공정성: 특정 대상을 차별하거나 과거의 편향을 계속 제시하지 않는 모델 생성
- 해석 가능성: 모델의 결정 방식 이해, 특히 규제 대상 업계의 경우 중요함
- 프라이버시 및 보안: ML 프로세스에서 사용되는 민감한 데이터 보호

인프라 도전 과제:
- 확장성: 계속 늘어나는 데이터의 양 및 계산 요구 사항 처리
- 모델 배포: 프로덕션 환경에 ML 모델 통합
- 모니터링 및 유지 관리: 지속적인 모델 성능 및 관련성 보장

데이터 준비, 모델 개발 및 배포용 도구를 제공하는 SageMaker AI 등의 AWS 서비스를

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

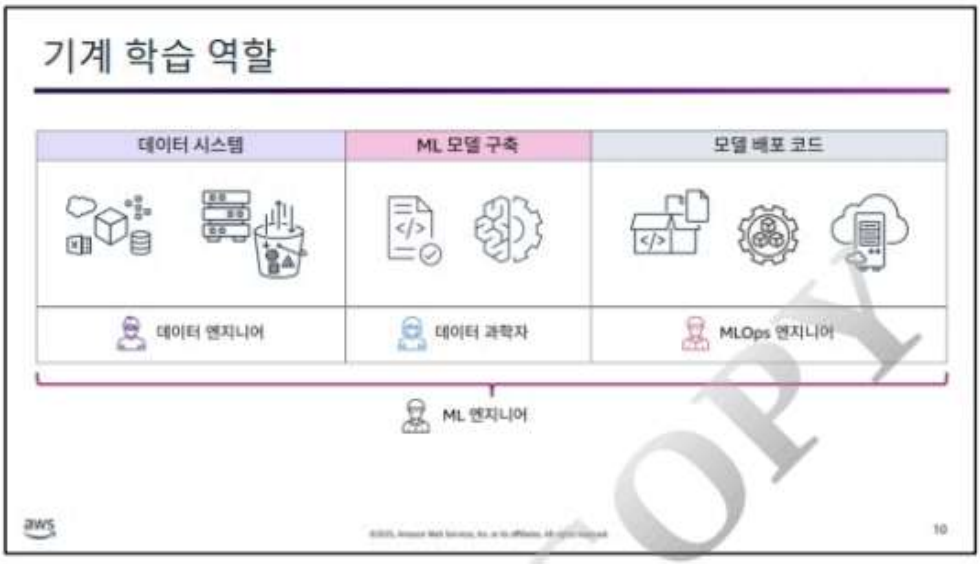
Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

활용하면 이러한 당면 과제를 해결할 수 있습니다.

DO NOT COPY
ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



기계 학습 프로젝트에는 대개 ML 파이프라인의 각기 다른 작업을 처리하는 서로 다르지만 상호 연관되어 있는 세가지 역할이 포함됩니다. 데이터 엔지니어는 데이터 시스템 구축과 유지 관리, 데이터 품질 보장, 데이터 인프라 관리 업무를 주로 담당합니다. 데이터 과학자는 피처 엔지니어링, 모델 선택, 최적화를 포함하는 ML 모델 구축에 집중합니다. MLOps 엔지니어는 주로 프로덕션 환경의 모델 배포, 모니터링 및 유지 관리에 초점을 맞추어 배포 코드를 처리합니다.

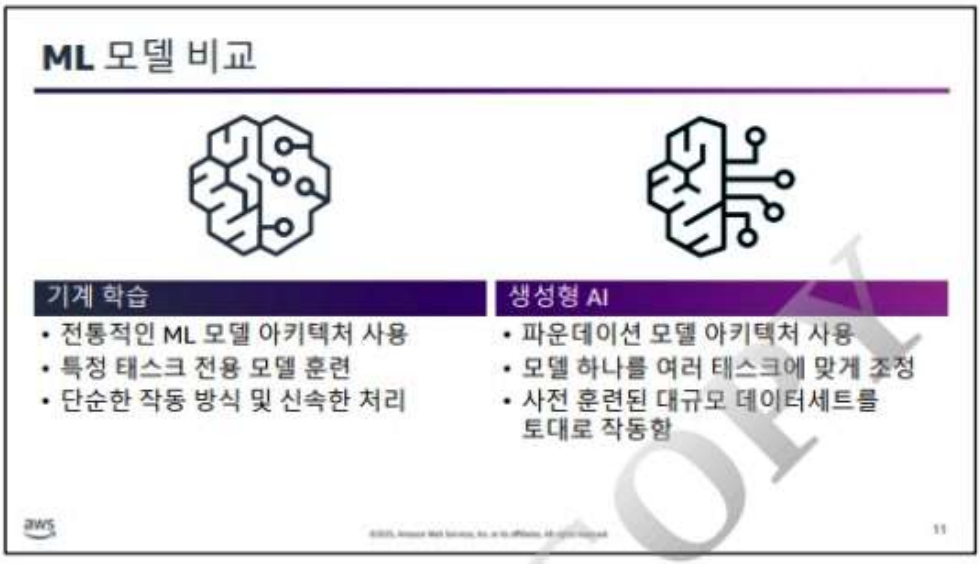
이 세가지 전문 분야를 연결하는 역할인 ML 엔지니어는, 전체 ML 파이프라인에 대해 파악하고 있어야 합니다. 이 역할은 데이터 시스템, 모델 개발 및 배포 인프라의 원활한 통합을 보장하는데 중요합니다.

ML 엔지니어는 각 팀의 의견을 조율하는 경우가 많으며, 데이터 수집 단계에서 프로덕션 환경에 모델을 제공하는 단계까지 전체 기계 학습 워크플로에 영향을 주는 아키텍처 관련 사항을 결정합니다.

DO NOT COPY
ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



전통적인 기계 학습 모델과 생성형 AI는 장점이 서로 다른 2가지 독자적인 AI 방식이라 할 수 있습니다. 전통적인 ML 모델은 특정 태스크용을 위해서 특별히 제작되었으며, 레이블이 지정된 훈련 데이터 필요하고 해당 태스크 중심의 출력을 생성합니다. 이는 분류하고 패턴 인식 같은 잘 정의된 작업에 뛰어나서 빠른 성능과 효율적인 리소스 사용을 제공합니다. 반대로, 생성형 AI, 특히 대형 언어 모델은 레이블이 지정되지 않은 방대한 양의 데이터에 대해 사전 훈련되어, 여러 작업에 적용할 수 있는 폭넓은 지식을 축적합니다.

가장 큰 차이점은 그들의 문제 해결 방식입니다. 전통 ML 모델은 특수 도구와 같이 효율성은 높지만 사용 가능 범위는 제한적입니다. 반면 생성형 AI 모델은 언어와 데이터에서 패턴을 인식하여 다양한 태스크를 처리할 수 있는 다용도 도구입니다. 이러한 모델은 유연성을 제공하지만, 특정 태스크에서 특화된 ML 모델의 정밀도를 맞출 수 있도록 더 많은 리소스와 미세 조정이 필요할 수 있습니다.

이러한 모델과 관련된 주요 용어는 다음과 같습니다.

- 전통 ML: 레이블이 지정된 데이터로 훈련해야 하는 태스크별 모델
- 레이블이 지정된 데이터: 훈련용으로 명시적인 태그나 카테고리가 지정되어 있는 정보
- LLM: 다양한 태스크를 처리할 수 있는 모델
- 사전 훈련: 방대한 일반 데이터를 사용하는 초기 학습 단계
- 태스크별 조정: 학습한 지식을 새로운 문제에 적용하는 기능
- 리소스 요구 사항: 필요한 컴퓨팅 성능 및 메모리
- 모델 범위: 모델이 효율적으로 처리할 수 있는 태스크의 범위

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

Amazon SageMaker Studio는 데이터 준비에서 모델 배포 및 모니터링에 이르기까지 전체 기계 학습 워크플로를 간편하게 진행할 수 있는 완전 통합 개발 환경(IDE)입니다. SageMaker Studio에서는 데이터 과학자와 개발자가 효율적으로 협업을 진행하고 ML 모델 구축, 훈련, 배포를 진행할 수 있도록 포괄적인 도구 및 기능 세트가 제공됩니다.

SageMaker Studio를 사용하면 ML 모델을 구축, 훈련, 배포에 소요되는 시간과 작업량을 대폭 줄여 인프라 관리보다는 비즈니스 문제 해결에 주력할 수 있습니다.

DO NOT COPY
ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

AWS ML 및 생성형 AI 스택에서는 다양한 요구 및 전문 지식에 맞춰 다양한 추상화 수준의 솔루션이 제공됩니다. 스택은 3가지 주요 계층으로 구성되어 있습니다.

생산성 향상 애플리케이션:
이 최상위 계층은 광범위한 기술 전문 지식 없이도 인공 지능과 기계 학습 (AI/ML) 기능을 사용할 수 있도록 도와주는 특수 애플리케이션과 서비스를 제공합니다. 비즈니스 인텔리전스를 위한 Amazon Q Business, 코드 생성을 위한 Amazon Q Developer를 예로 들 수 있습니다.

생성형 AI 앱 구축용 모델 및 도구:
중간 계층은 생성형 AI 애플리케이션 구축을 위한 파운데이션 모델(FM)과 도구에 대한 액세스를 제공합니다. Amazon Bedrock과 같은 서비스는 사용자 지정 및 통합 기능과 더불어 AWS와 서드 파티의 FM을 제공합니다. Amazon SageMaker AI를 사용하여 사용자 지정 ML 모델을 개발, 훈련 및 배포할 수 있습니다.

ML 인프라 도구 및 서비스:
최하위 계층은 AWS Trainium과 AWS Inferentia 같은 AWS 인프라 서비스로 구성되어 있는데, 이들은 ML 모델을 대규모로 훈련하고 실행하는 데 최적화되어 있습니다. 이 계층은 전문 사용자와 연구원에게 제어 가능성 및 성능으로 고도로 전문화된 ML 솔루션을 구축하고 배포할 수 있게 해줍니다.

이 모듈식 설계는 요구 사항, 전문 지식, 원하는 제어 수준에 따라 적절한 추상화 수준을 선택할 수 있도록 도와줍니다. 이러한 유연성은 즉시 사용 가능한 솔루션, 맞춤형 애플리케이션 또는 고급 연구 개발(R&D)을 위한 하위 수준 인프라가 필요하든 상관없이 AI/ML 기능을 효과적으로 사용하는 데 도움이 됩니다.

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

Amazon은 컴퓨터 비전, 음성, 텍스트 등의 다양한 유형의 데이터를 처리할 수 있는 포괄적인 AI 서비스 제품군을 제공합니다. 각 카테고리는 다른 AWS 서비스와의 통합 기능을 유지하면서 특정 사용 사례를 다루기 때문에, AI 기반 애플리케이션 구축을 위한 강력한 도구가 됩니다.

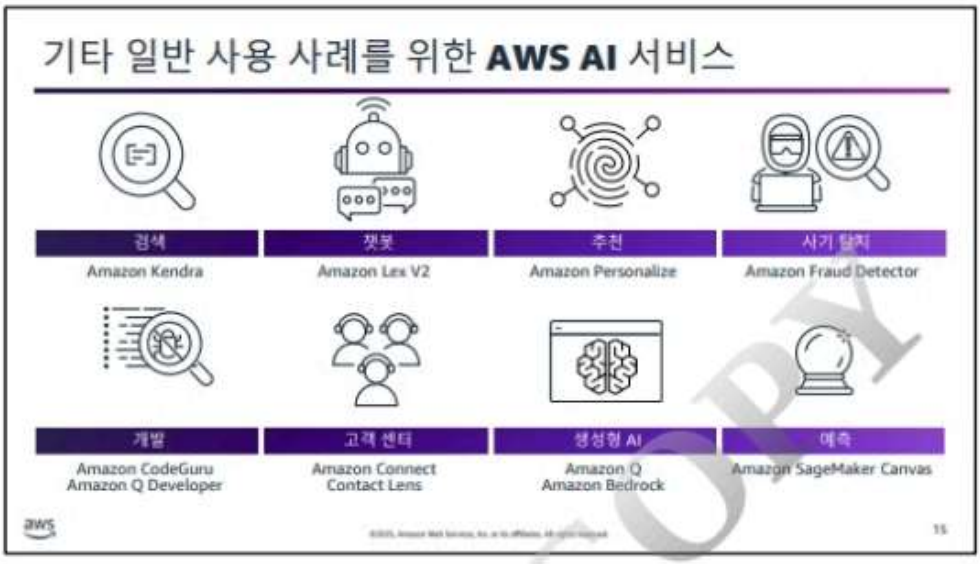
컴퓨터 비전 카테고리인, Amazon Rekognition에서는 정교한 이미지 및 비디오 분석 기능을 제공하여 애플리케이션이 객체를 감지하고, 장면을 분석하고, 시각적 콘텐츠에서 얼굴과 텍스트를 인식할 수 있도록 합니다.

음성 처리와 관련해서 Amazon 두 가지 주요 서비스를 제공합니다. 텍스트를 음성으로 변환하는 Amazon Polly와 음성을 텍스트로 변환하는 Amazon Transcribe를 제공해서 애플리케이션이 음성 상호 작용과 오디오 콘텐츠 처리를 효과적으로 처리할 수 있도록 합니다.

텍스트 처리 카테고리에는 세 가지 강력한 서비스가 포함됩니다. 자연어 처리 및 이해를 위한 Amazon Comprehend, 언어 번역을 위한 Amazon Translate, 문서에서 텍스트와 데이터를 추출하는 Amazon Textract입니다. 이러한 텍스트 기반 서비스는 다양한 형식과 언어로 작성된 콘텐츠를 이해, 분석 및 처리하는 애플리케이션을 구축하는 데 도움이 되므로 문서 처리 및 언어 기반 애플리케이션에 필수적인 도구가 됩니다.

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



AWS는 다양한 비즈니스 요구를 해결하도록 설계된 몇 가지 다른 AI 서비스를 제공합니다,.

- 검색: Amazon Kendra에서는 지능형 엔터프라이즈 검색 기능이 제공됩니다.
- 챗봇: Amazon Lex V2에서는 대화형 인터페이스를 구축할 수 있습니다.
- 추천: Amazon Personalize에서는 AI 기반 추천 엔진이 제공됩니다.
- 사기 탐지: Amazon Fraud Detector에서는 사기 가능성이 있는 활동을 식별하는데 도움이 됩니다.
- 개발: Amazon CodeGuru에서는 코드 리뷰와 애플리케이션 성능 추천을 도와줍니다.
- 고객 센터: Amazon Connect Contact Lens에서는 고객 센터 상호 작용을 위한 분석 기능이 제공됩니다.
- 생성형 AI: Amazon Q 및 Amazon Bedrock에서는 생성형 AI 애플리케이션 구축용 도구가 제공됩니다.
- 예측: Amazon SageMaker Canvas에서는 시계열 예측을 위한 코드 없는 도구가 제공됩니다.

사전 구축된 이러한 솔루션을 사용하면 사용자 지정 ML 모델을 처음부터 구축할 필요 없이, 고객 경험을 향상시키고, 운영 효율성을 개선하고, 새로운 AI 기반 애플리케이션을 만들 수 있습니다. 이런 서비스를 사용하면 특정 비즈니스 사용 사례에 맞게 조정된 AI 기능을 신속하게 구현하여 혁신과 디지털 전환 노력을 가속화할 수 있습니다.

DO NOT COPY ssleeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

Amazon SageMaker AI는 모델 구축 및 교육부터 대규모 모델 배포에 이르기까지 전체 ML 수명 주기를 간소화하도록 설계된 포괄적인 기계 학습 서비스입니다. 이 서비스의 주 목표는 데이터 과학자와 ML 엔지니어에게 강력한 도구를 제공하면서 기계 학습 프로세스를 단순화하는 것입니다.

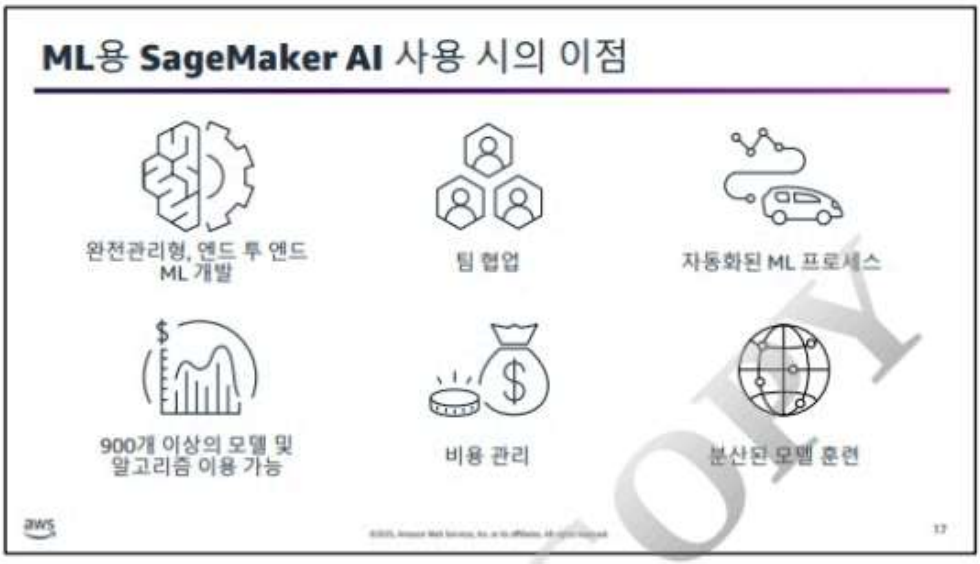
SageMaker AI는 다섯 가지 주요 구성 요소로 이루어져 있는데, 각 구성 요소는 ML 워크플로우에서 특정 목적을 수행합니다. Amazon SageMaker Studio는 ML 개발을 위한 통합 개발 환경(IDE)을 제공하는 반면, SageMaker Autopilot은 다양한 수준의 전문 지식을 가진 사용자를 위해 자동화된 ML 기능을 제공합니다. SageMaker Canvas는 코드 없는 ML 모델 개발 환경을 제공하여 비즈니스 분석가가 기계 학습에 접근할 수 있게 해줍니다. SageMaker Data Wrangler 구성 요소는 효율적인 데이터 준비와 기능 엔지니어링을 용이하게 하고, 모델 훈련 구성 요소는 ML 모델 학습을 위한 강력한 인프라와 도구를 제공합니다.

이러한 구성 요소가 원활하게 작동하여 데이터 준비, 모델 훈련, 배포 같은 ML 개발의 일반적인 문제를 해결합니다. SageMaker AI는 포괄적인 도구 모음을 제공함으로써, 조직에 ML 이니셔티브를 가속화하는 동시에 개발 프로세스에 대한 유연성과 제어를 유지할 수 있는 도구를 제공합니다. 이런 통합 접근 방식은 프로덕션 환경에서 ML 모델을 빌드하고 배포하는 데 일반적으로 연관된 복잡성을 줄이는 데 도움이 됩니다.

Amazon SageMaker AI 기능 및 서비스의 전체 목록은 Amazon SageMaker AI 설명서 (https://docs.aws.amazon.com/sagemaker/latest/dg/whatis.html)를 참조하십시오.

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



ML 워크플로에 SageMaker AI를 사용하는 경우의 주요 이점은 다음과 같습니다.

- 통합형 엔드 투 엔드 ML: SageMaker Studio에서는 데이터 준비에서 배포까지의 모든 ML 태스크를 수행할 수 있는 통합 환경이 제공되므로 ML 프로세스를 간편하게 진행할 수 있습니다.
- 팀 협업: 실험, 모델, Notebook의 더욱 개선된 공유 기능이 제공되므로 협업과 일관성을 높일 수 있습니다.
- 자동화된 ML 프로세스: SageMaker Autopilot을 통해 피처 엔지니어링, 알고리즘 선택, 하이퍼파라미터 튜닝 등의 엔드 투 엔드 워크플로를 자동화할 수 있습니다.
- 사전 구축된 모델 이용: SageMaker AI에서는 AWS Marketplace에서 제공되는 사전 훈련된 모델 및 알고리즘을 이용할 수 있으므로 개발 시간이 단축됩니다.
- 비용 관리: 종량제 요금이 책정되며, 리소스 모니터링과 오토 스케일링 기능도 제공되므로 ML 개발을 유동적으로 진행하면서 비용을 제어할 수 있습니다.
- 분산 모델 훈련: 자동 클러스터 관리 및 스케일링 기능을 통해 분산 훈련 설정 및 구성을 간소화합니다.

이러한 이점을 활용하면, ML 개발을 신속하게 진행하,고 운영 오버헤드를 줄일 수 있으며, ML 프로젝트에서 비즈니스 가치를 창출하는 과정에 주력할 수 있습니다. SageMaker AI의 포괄적인 도구 세트를 통해 ML 수명 주기에서 일반적인 도전 과제를 해결할 수 있으므로 더욱 효율적으로 모델을 구축 및 배포할 수 있습니다.

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

SageMaker AI는 전체 ML 수명 주기를 촉진합니다.

1. 데이터 준비: SageMaker Data Wrangler가 데이터 준비 태스크를 자동화하므로 데이터 클리닝 및 피처 엔지니어링에 소요되는 시간이 단축됩니다.
2. 특성 저장: SageMaker Feature Store에서 모델 훈련 및 추론을 위한 특성을 구성하고 저장하여 프로젝트 간 일관성을 유지하도록 도와줍니다.
3. Notebook으로 구축: 통합 Jupyter Notebook은 컴퓨팅 리소스에 쉽게 액세스하면서 실험하고 협업할 수 있는 공간을 제공합니다.
4. 모델 훈련: 기본 제공 알고리즘, 프레임워크 및 분산 훈련 기능을 활용하여 대규모 모델 훈련을 효율적으로 진행할 수 있습니다.
5. 파라미터 튜닝: 자동 모델 튜닝은 하이퍼파라미터를 최적화하므로 수동 시행착오 없이 모델 성능을 개선합니다.
6. 프로덕션 환경에 배포: 원클릭 모델 호스팅은 확장 가능 엔드포인트로 모델을 배포할 수 있으므로 개발 단계에서 프로덕션 단계로의 전환을 단순화합니다.
7. 관리 및 모니터링: SageMaker Model Monitor가 프로덕션 환경에서 모델 품질을 추적하여 드리프트를 탐지하고 필요시 재훈련을 시작합니다. 설명 가능성 및 거버넌스 도구를 활용하면 투명성과 규정 준수를 보장하는데 도움이 됩니다.

반복 루프는 모델 훈련과 하이퍼파라미터 튜닝 단계를 연결하는데, 이는 최적의 성능에 도달할 때까지 모델을 훈련하고, 튜닝하고, 재훈련하는 지속적인 미세 조정 사이클을 나타냅니다.

SageMaker AI의 통합된 접근 방식은 데이터 수집에서 배포 및 모니터링에 이르는 전체 ML 워크플로우를 간소화합니다. 이 엔드-투-엔드 솔루션은 여러 도구와 환경을 관리하는 복잡성을 줄여 주므로 ML로 비즈니스 문제를 해결하는 데 집중할 수 있습니다.

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

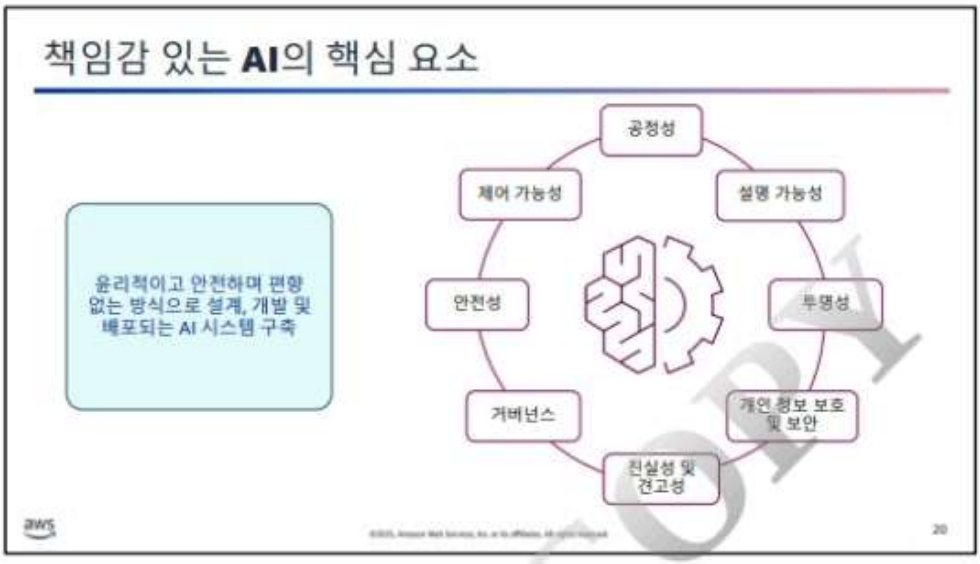
책임감 있는 기계 학습은 AI 시스템이 윤리적이고 안전하며 편견이 없도록 하는 데 도움이 됩니다. AI의 영향력이 커질수록 잠재적 위험을 해결하는 것이 중요합니다.

AWS는 편향 탐지를 위한 SageMaker Clarify 같은 도구, 공정성과 설명 가능성 기능을 갖춘 서비스, 윤리적 AI 개발에 관한 문서를 제공하여 책임 있는 관행을 장려합니다. 책임감 있는 ML을 우선시하면 신뢰가 쌓이고 규정 준수를 보장하며 긍정적인 AI 발전에 기여합니다.

DO NOT COPY
ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



책임감 있는 AI의 8가지 핵심 요소는 아래와 같습니다.

- 공정성: 여러 집단에 미치는 영향을 고려해야 합니다.
- 설명 가능성: 어떻게 출력이 생성되었는지 이해할 수 있어야 합니다.
- 투명성: AI 시스템에 대한 명확한 정보를 제공해야 합니다.
- 개인 정보 보호 및 보안: 데이터와 모델을 적절하게 확보, 사용 및 보호해야 합니다.
- 진실성 및 안정성: 적대적 입력에도 불구하고 정확한 출력을 보장해야 합니다.
- 거버넌스: 책임 있는 관행을 정의하고 시행해야 합니다.
- 안전: 유해한 출력 및 오용을 방지해야 합니다.
- 제어 가능성: 인간이 AI를 감독하고 개입 해야 합니다.

AWS는 편향 탐지 및 설명 가능성을 위한 SageMaker Clarify, 보안 및 액세스 제어를 위한 AWS Identity and Access Management (IAM), 이러한 요소를 지원하는 책임 있는 AI 모범 사례 문서 같은 도구를 제공합니다.

DO NOT COPY ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

책임감 있는 AI 사례는 상당한 비즈니스 이점을 제공합니다.

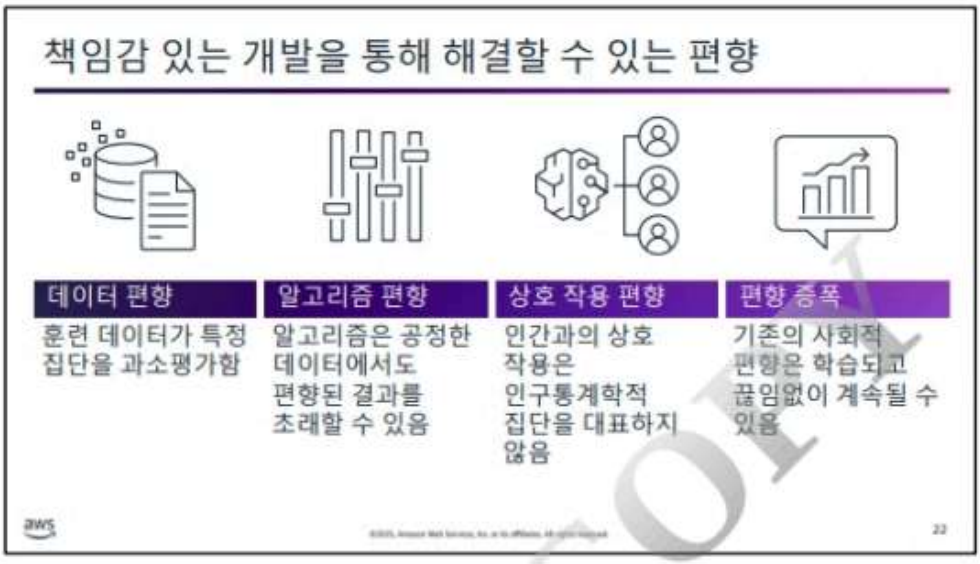
- 신뢰 및 평판 향상: 윤리적 측면을 최우선으로 고려함을 입증하면 고객과 파트너 관계와 브랜드 충성도를 강화할 수 있습니다.
- 규정 준수 개선: 책임 있는 관행을 채택하면 기업이 AI 관련 규정을 발빠르게 준수하고 법적 문제를 방지할 수 있습니다.
- 향상된 위험 완화: 편향, 개인 정보 침해 등 잠재적 AI 위험을 식별하고 해결하는 기업은 재정적 손해와 평판 실추를 방지하는데 도움이 됩니다.
- 경쟁 우위 향상: 책임감 있는 AI를 우선시하면 더욱 많은 고객과 인재를 확보하여 제품과 서비스를 차별화할 수 있습니다.
- 향상된 의사 결정: 공정하고 투명하며 책임감 있는 AI를 보장하면, 정보를 토대로 더욱 윤리적인 결정을 내릴 수 있습니다.
- 제품 및 프로세스 개선: 다양한 관점과 윤리적 영향을 고려하면 더 포괄적으로 효과적인 솔루션이 나올 수 있습니다.

편향 탐지용 SageMaker Clarify 및 거버넌스용 AWS CloudTrail 같은 AWS 도구는 책임감 있는 AI 관행을 지원합니다.

DO NOT COPY ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



책임감 있는 AI는 AI 시스템의 다양한 편향을 해결하는 것을 목표로 합니다. AI 시스템은 책임 있는 개발 과정에서 신중하게 해결해야 하는 다양한 유형의 편향이 보일 수 있습니다. 이러한 편향은 네 가지 주요 영역, 즉 데이터 편향, 알고리즘 편향, 상호 작용 편향, 편향 증폭으로 나타나는데, 각각 특정한 완화 전략이 필요한 고유한 도전 과제를 제시합니다.

데이터 편향은 훈련 데이터 세트가 모든 인구나 지역을 적절하게 나타내지 못할 때 발생합니다. 예를 들어, 주로 뉴욕이나 샌프란시스코 같은 대도시 지역의 데이터를 기반으로 훈련한 급여 예측 모델은 시골 지역이나 소도시에 적용하면 성과가 좋지 않아 급여 기대치가 왜곡될 수 있어요. 마찬가지로, 주로 학술 텍스트를 기반으로 훈련된 언어 처리 모델은 다양한 교육 배경의 일상적인 의사소통 패턴에서 어려움을 겪을 수 있어요.

적절하게 균형이 잡힌 데이터 세트로 작업하더라도, 알고리즘 편향과 상호 작용 편향은 이런 문제를 더 복잡하게 만들 수 있습니다. 예를 들어, 추천 시스템은 과거의 상호 작용 패턴을 기반으로 특정 인구통계학적 집단에 불균형적으로 유리할 수 있습니다. AI 시스템이 구인 추천이나 대출 승인 같은 영역에서 기존의 사회적 격차를 학습하고 강화할 때 편향 증폭이 발생될 수 있습니다. 책임감 있는 AI 개발을 하려면 모든 인구통계학적, 사회경제적 집단에 걸쳐 공정하고 공평한 대우를 확립하는 데 도움이 되도록 이러한 시스템을 지속적으로 모니터링하고 조정해야 합니다.

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

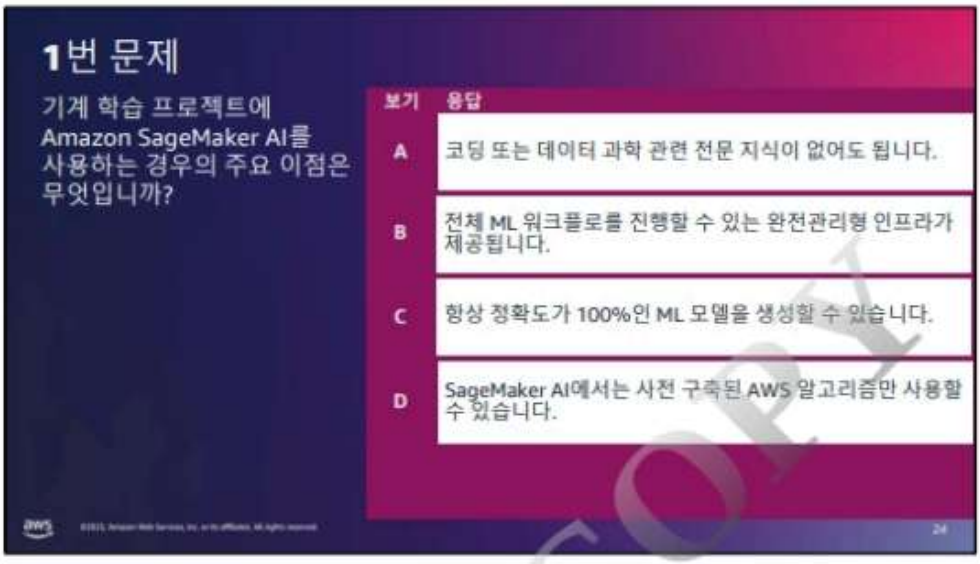
Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

이 단원에서 배운 주요 개념을 더욱 명확하게 파악할 수 있도록 간략한 모듈 종료 지식 확인을 진행하여 단원 내용을 빠르게 복습해 보겠습니다.

DO NOT COPY
ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

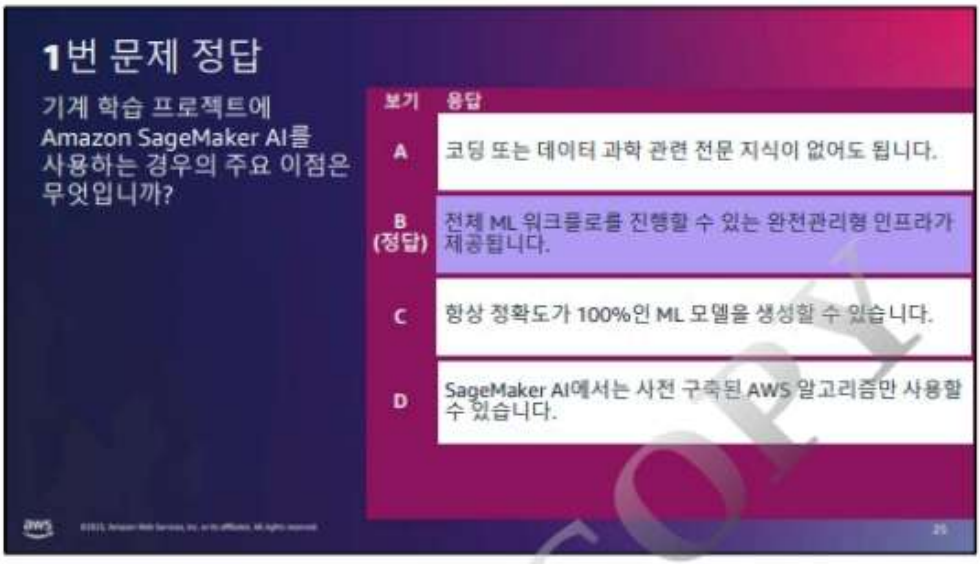


SageMaker AI의 이점과 기능을 이해하고 있는지를 확인하기 위한 문제입니다.

DO NOT COPY
ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



정답은 B: '전체 ML 워크플로를 진행할 수 있는 완전관리형 인프라가 제공됩니다.'입니다.

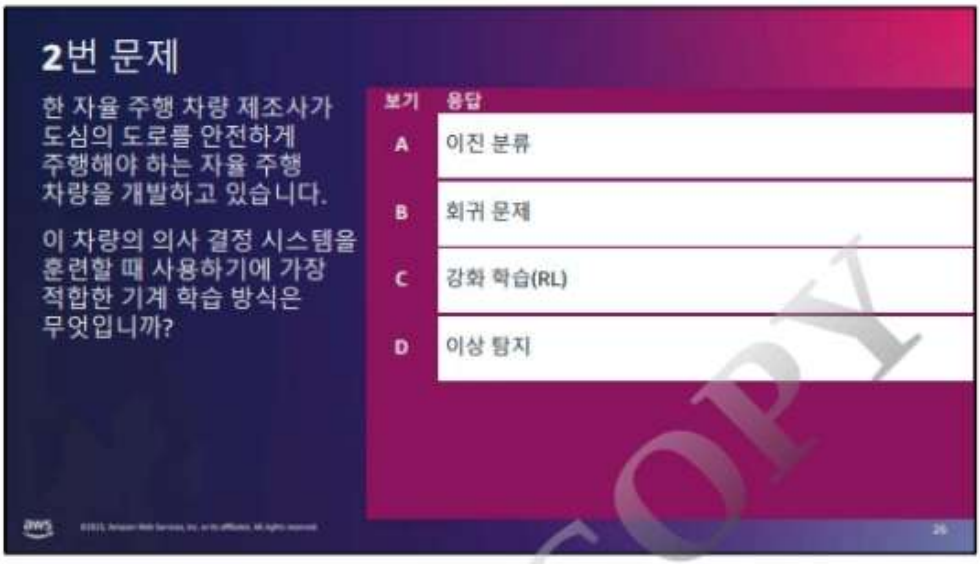
설명: SageMaker AI는 데이터 준비에서 모델 배포 및 모니터링에 이르기까지 전체 ML 수명 주기의 인프라를 관리합니다.

오답:
A. SageMaker AI를 사용하더라도 코딩 및 데이터 과학 관련 지식이 어느 정도 있어야 합니다.
C. 모든 모델의 완벽한 정확도를 보장할 수 있는 ML 플랫폼은 없습니다.
D. SageMaker AI에서는 AWS 알고리즘과 사용자 정의 알고리즘이 모두 지원됩니다.

DO NOT COPY ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

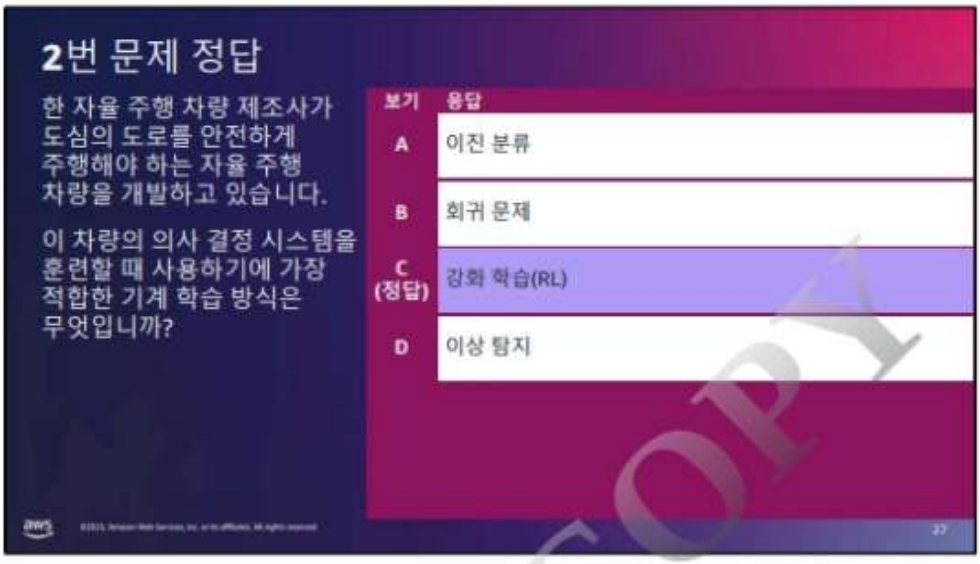
Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



선택한 모델을 훈련시키기 위한 다양한 기계 학습 접근법에 대한 기억을 평가하기 위한 문제입니다.

DO NOT COPY
ssleeeee1@gmail.com

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



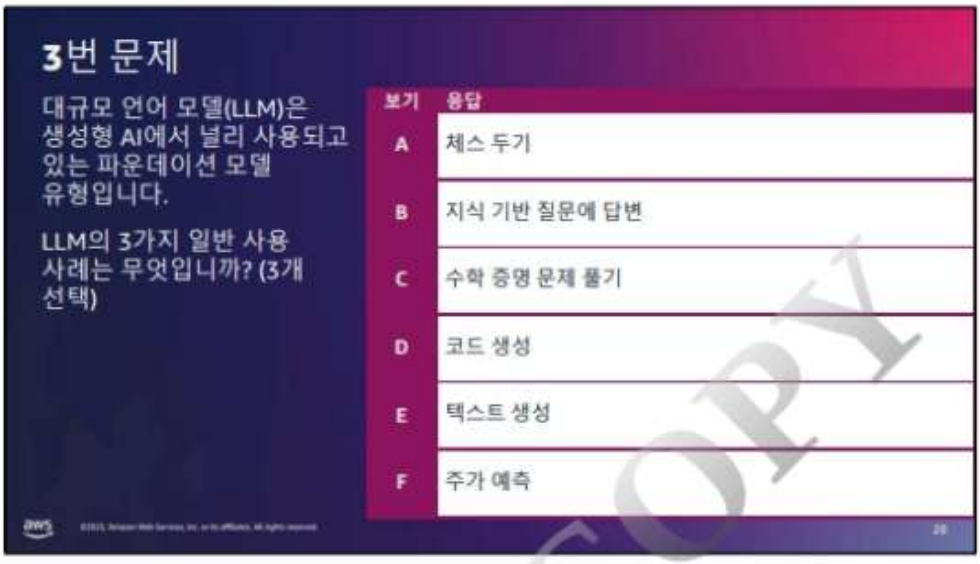
정답은 C. 강화 학습(RL)입니다.

설명: 자율 주행 차량 훈련에 적합한 방식은 강화 학습(RL)입니다. 강화 학습을 진행하면 시스템이 시뮬레이션 된 환경에서 시행착오를 거치면서 최적의 행동을 학습할 수 있게 하여 안전 주행 시에는 보상을 받고 위험한 주행 시에는 페널티를 받을 수 있습니다.

오답:
A. 이진 분류는 너무 단순하여 자율 주행에 필요한 복잡한 의사 결정을 진행할 수 없습니다.
B. 회귀는 연속 값 예측 시에는 유용하지만 순차적 의사 결정 태스크에서는 유용하지 않습니다.
D. 이상 탐지는 해당 시스템의 구성 요소 중 하나일 수는 있지만 기본적인 주행 훈련 방식으로는 적합하지 않습니다.

DO NOT COPY ssleeeee1@gmail.com

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



대규모 언어 모델(LLM)의 기능 및 적절한 사용 사례를 이해도를 평가하기 위한 질문입니다. 방대한 텍스트 데이터로 훈련한 유용한 언어 모델인 LLM은 폭넓은 언어 관련 태스크를 수행할 수 있습니다.

DO NOT COPY
ssleeeee1@gmail.com

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

정답은 B, D, E입니다.

B. 지식 기반 질문에 답변: LLM은 대량의 정보를 처리하고 관련 답변을 검색하는데 탁월해서 질문 답변 시스템에 적합합니다.
D. 코드 생성: LLM은 방대한 코드 리포지토리로 훈련한 모델이므로 다양한 프로그래밍 언어로 작성된 코드 조각을 생성하고, 설명하고, 디버그를 할 수 있습니다.
E. 텍스트 생성: 콘텐츠 생성, 요약, 창작물 작성 등의 애플리케이션을 가능하게 하는 LLM의 핵심 기능입니다.

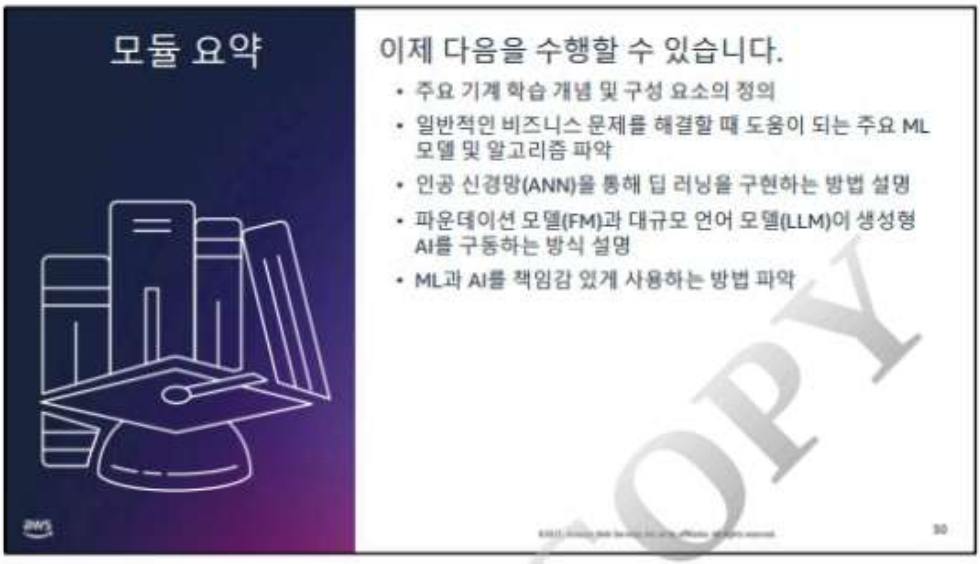
오답 및 오답에서 LLM 사용이 제한되는 이유는 다음과 같습니다.
A. 체스 두기: LLM에는 수준 높은 체스 게임을 위해 필요한 전략적 추론 및 계획 기능이 부족합니다. 이 태스크에는 특수 게임용 AI가 더 적합합니다.
C. 수학 증명 풀기: LLM도 기본적인 수학 문제를 풀 수는 있지만 복잡한 수학적 추론과 형식 증명 구성은 어려움을 겪습니다.
F. 주가 예측: 이는 현재 LLM 기능을 넘어서는 실시간 데이터 분석 및 복잡한 재무 모델링 기능이 필요합니다.

실제 애플리케이션에서 이처럼 강력한 모델을 효율적으로 사용하려면 LLM에 적합한 태스크와 그렇지 않은 태스크를 파악해야 합니다.

DO NOT COPY ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



이 모듈을 통해서 AWS 기반 기계 학습 엔지니어링에 대한 필수 지식과 기술을 습득했습니다. 이제 다음을 수행할 수 있습니다.

- 주요 기계 학습 개념 및 구성 요소의 정의
- 일반적인 비즈니스 문제를 해결할 때 도움이 되는 주요 ML 모델 및 알고리즘 파악
- 인공 신경망(ANN)을 통해 딥 러닝을 구현하는 방법 설명
- 파운데이션 모델(FM)과 대규모 언어 모델(LLM)이 생성형 AI를 구동하는 방식 설명
- ML과 AI를 책임감 있게 사용하는 방법 파악

이러한 목표에는 광범위한 ML 원칙, 고급 주제 및 실제 고려 사항이 포함됩니다. 이 지식은 기계 학습을 더 깊이 살펴보고 실제 시나리오의 기계 학습 애플리케이션에 필요한 견고한 기반을 형성합니다.

DO NOT COPY
ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

이 모듈에서 알아본 데이터 준비 및 피처 엔지니어링 관련 개념을 잠시 생각해 보시기 바랍니다. 이 과정에서 다룬 자료 또는 이 과정에서 배운 기술을 실제 업무에 적용하는 방법과 관련하여 궁금한 점이 있으면 지금 질문을 할 수 있습니다.

학습은 꾸준히 진행해야 하는 과정이며 특히 급격하게 발전하고 있는 기계 학습 분야에서는 새로운 정보를 지속적으로 습득해야 합니다. AWS에서는 지속적인 자기 계발 지원을 위해 다음과 같은 여러 리소스를 제공합니다.

- **AWS 설명서**: 서비스 및 모범 사례와 관련된 자세한 정보 확인 가능
- **AWS 교육 및 자격증**: 기술 역량 강화 및 전문 자격증 취득 지원
- **AWS Support**: AWS 서비스 관련 기술 지원 제공

더 정확하게 파악해야 하는 부분이 있거나 세션 후에 추가로 문의할 사항이 있는 경우 반드시 질문을 하시기 바랍니다. 수강생의 피드백은 향후 이 과정을 진행하는 학습자들을 위해 교육 자료와 세션을 개선하는 데 유용하게 활용됩니다.

DO NOT COPY ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

이 모듈에서는 **기계 학습(ML) 훈련 알고리즘**, 그리고 Amazon SageMaker AI에서의 응용을 소개합니다. 다양한 유형의 ML 알고리즘, 특정 문제에 적합한 알고리즘을 선택하는 방법, 그리고 SageMaker AI에서 기본 제공 알고리즘 및 사용자 정의 훈련 옵션을 사용하는 방법을 알아봅니다.

이번 주제에서 지도 학습, 비지도 학습, 강화 학습(RL)을 다룹니다. 선형 회귀, 의사 결정 트리, 신경망 같은 인기 알고리즘에 대한 통찰력을 얻을 수 있습니다. 모듈을 마치면, 비즈니스 문제를 적절한 ML 기술을 맞추는 방법과 효율적인 모델 훈련을 위해 SageMaker AI 기능을 사용하는 방법을 이해할 수 있을 것입니다.

DO NOT COPY
ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

기계 학습 수명 주기는 여러 주요 단계로 구성된 반복적인 프로세스입니다.

이것은 네비게이션 다이어그램입니다. 과정을 계속 진행하면서 해당 수명 주기를 가로지르게 되고 완료한 주제는 보라색으로 바뀝니다. 이 모듈에서는 기계 학습 워크플로의 비즈니스 문제를 평가하는 방법을 설명합니다.

DO NOT COPY
ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

이 모듈은 4가지 주요 주제를 다룹니다.

A. ML 비즈니스 도전 과제 평가: 비즈니스 문제가 ML 솔루션에 적합한지를 평가하는 방법과 해당 문제를 ML 태스크로 구성하는 방법을 알아봅니다.
B. ML 훈련 접근 방식: SageMaker AI를 사용하여 ML 모델 구축, 훈련 및 배포하고, Amazon SageMaker Studio로 ML 솔루션을 개발하기 위한 통합 개발 환경을 살펴봅니다.
C. ML 훈련 알고리즘: 기본 제공 알고리즘과 사용자 지정 훈련 옵션을 사용하는 걸 포함하여, SageMaker AI를 사용한 ML 모델 훈련의 구체적인 방법을 살펴봅니다.

이러한 주제에는 과제 평가에서 실제 구현까지 논리적으로 진행되며, AWS에서 ML을 개발하는 포괄적인 개요를 제공합니다.

DO NOT COPY
ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

기계 학습에 대한 비즈니스 과제를 평가할 때는 비즈니스의 주요 과제와 성공 기준을 이해해야 합니다.

DO NOT COPY
ssleeeee1@gmail.com

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

ML 성공 기준 프레임워크에서는 ML 프로젝트의 잠재적 성공을 평가하는데 도움이 됩니다.

1. **비즈니스 문제 정의** - 해결하려는 도전 과제를 명확하게 기술합니다. 예를 들어 "위험 상태의 고객을 파악하여 고객 이탈 감소"와 같은 과제입니다.
2. **측정 가능한 목표 설정** - 구체적이고, 정량화 가능한 목표를 수립합니다. 예를 들어 "6개월 내에 이탈률 15% 감소"와 같은 목표를 설정합니다.
3. **성공 지표 파악** - 진행 상황을 측정하는 방법을 결정합니다. 예를 들어 '월별 이탈률, 고객 평생 가치, 고객 유지 캠페인 효과를 추적하세요."입니다.
4. **ML 실행 가능성 평가** - ML이 적절한 접근 방식인지를 평가합니다. 아래를 생각해야 합니다.
   - 데이터 가용성 및 품질
   - 문제의 복잡성
   - 구현 비용 대비 잠재적 영향

이 프레임워크는 ML 프로젝트가 비즈니스 목표에 부합하고 측정 가능한 성공으로 가는 명확한 경로를 갖도록 도와줍니다. ML 이니셔티브의 우선순위를 지정하고 이해관계자들에게 현실적인 기대치를 설정하는 데 도움이 됩니다.

DO NOT COPY
ssleeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

AnyCompany Ecommerce 사기 행위 방지 사용 사례

성장하고 있는 온라인 마켓플레이스인, AnyCompany Ecommerce에서 평판과 재정적 안정성을 모두 위협하는 사기 행위가 크게 증가하고 있습니다. 보안팀이 두 가지 주요 문제를 파악했습니다. 해킹 당한 사용자 계정 수가 증가하고 악의적인 공격자가 만든 가짜 제품 목록이 쏟아지는 것입니다. 이러한 사기 행위로 인해 수많은 고객 불만이 제기되었습니다. 합법적인 사용자들이 자신의 계정으로 무단 구매를 확인하였습니다. 그뿐만 아니라, 인지하지 못한 구매자들은 허위 등록된 제품을 구매한 후 위조품을 받거나 아예 아무것도 받지 못했습니다. 지불 거절이 전년 대비 150% 증가하고 고객 유지율이 18% 감소하는 등 재정적 피해가 상당합니다.

해당 상황을 분석한 후, AnyCompany는 이러한 사기 문제를 해결하기 위해 기계 학습 솔루션을 구현하기로 결정했습니다. 데이터 과학팀은 ML이 이 과제에 특히 적합하다고 판단했습니다. 사기 행위의 패턴이 복잡하고 끊임없이 진화하기 때문입니다.

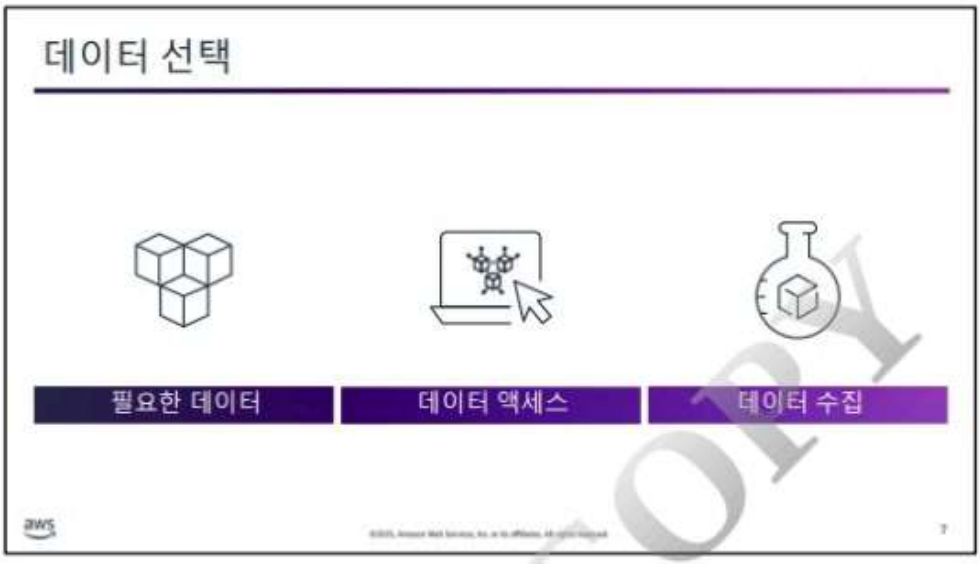
구현 계획에는 다음 작업이 포함됩니다.
- 실시간으로 계정 도용 시도를 탐지하는 모델 구축
- 의심스러운 제품 목록에 자동으로 플래그를 지정하는 알고리즘 생성
- 거래 및 판매자 계정을 위한 위험 평가 시스템 개발

AnyCompany는 손상된 계정 수를 90% 줄이고 허위 등록 제품 중 95%를 제거한다는 구체적인 목표를 수립했으며, 고객 신뢰도를 25% 향상시킬 것으로 기대합니다. 고객 만족도 설문 조사 및 유지율을 토대로 이를 측정할 예정입니다. 회사는 재정적 손실을 70% 줄이는 걸 예상하고 있으며, 지불 거절 및 사기 행위 관련 환불을 토대로 측정합니다.

회사는 매주 진행 상황을 추적하고 위험도가 가장 높은 제품 카테고리부터 시작하여 3개월에 걸쳐 단계적으로 구현할 계획입니다.

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



ML 프로젝트에 사용할 데이터를 선택할 때는 다음과 같은 주요 사항을 고려해야 합니다.

- 필요한 데이터 – 비즈니스 문제를 해결하려면 구체적으로 어떤 데이터가 필요합니까? AnyCompany Ecommerce의 경우, 고객 인구 통계, 구매 내역, 제품 속성이 포함될 수 있습니다.
- 데이터 액세스 – 필수 데이터는 어디에 저장됩니까? 쉽게 접근할 수 있습니까? 고객 관계 관리(CRM), 전사적 자원 관리 (ERP), 웹 분석 등 여러 시스템에 분산된 데이터를 고려할 수 있습니다.
- 데이터 수집 - 데이터를 어떻게 수집하고 중앙 집중화할 것입니까?

아래와 같은 옵션을 사용할 수 있습니다.

  - 데이터 레이크: 원시 데이터가 네이티브 포맷으로 저장됩니다.
  - 데이터 웨어하우스: 처리를 완료한 구조화된 데이터가 저장됩니다.
  - 추출, 변환, 로드(ETL) 프로세스: 여러 출처로부터 데이터를 추출, 변환 및 로드합니다.

ML 프로젝트의 성공을 위해서 데이터 품질은 중요합니다. 데이터가 다음 기준을 충족하는지 확인하십시오.

- 정확하며 오류가 없는 상태임
- 누락된 값이 거의 없이 완성됨
- 당면한 문제와 관련성이 있음
- 실제 시나리오를 대표할 수 있을 만큼 다양함

Amazon Simple Storage Service(Amazon S3), AWS Glue, Amazon Redshift 등의 AWS 서비스를 사용하면 ML 프로젝트용 데이터를 수집, 저장 및 준비할 수 있습니다.

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

고객	거래 날짜	공급업체	청구 금액	사기 여부
ABC	10/5	스토어 1	10.99	아니요
DEF	10/5	스토어 2	999.99	예
GHI	10/5	스토어 2	15.00	아니요
JKL	10/6	스토어 2	699 dls	?
MNO	10/6	?	999.99	예

기계 학습에서, **특성**과 **레이블**은 훈련 데이터의 중요한 구성 요소입니다. 특성은 예측을 하는 데 사용되는 입력 변수입니다. AnyCompany Ecommerce에서, 특성은 아마 아래와 같을 것입니다:

- 고객 연령, 위치, 구매 내역
- 제품 가격, 카테고리, 등급
- 거래 시간, 거래 요일

레이블은 다음 예와 같이 예측하려는 대상 변수입니다.

- 제품 추천의 경우, 고객이 추천 품목을 구매하는지 여부 (이진 레이블)
- 수요 예측의 경우, 판매된 제품 수 (연속 레이블)
- 사기 행위 탐지의 경우, 거래가 사기인지 아닌지 여부 (이진 레이블)

레이블이 지정된 데이터가 중요한 이유는 다음과 같습니다:

- 레이블이 지정된 데이터로 지도 학습 알고리즘이 가능합니다.
- 레이블을 사용하여, 패턴 및 관계를 학습할 수 있습니다.
- 모델 성능을 평가하려면 레이블이 필요합니다.

아래와 같이 데이터 품질 고려 사항을 준수하십시오:

- 특성이 관련성이 있고 유용한 정보를 제공하는지 확인합니다.
- 레이블의 클래스 불균형이 있는지 확인합니다. (예: 드문 사기성의 거래)
- 대규모 데이터세트에 레이블을 지정하는데 소요되는 비용과 시간을 고려합니다.

Amazon SageMaker Ground Truth는 ML 프로젝트의 대규모 데이터세트에 레이블을 효율적으로 지정하는데 도움이 됩니다.

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

AnyCompany Ecommerce의 사기 방지 도전 과제에 ML 성공 기준 프레임워크를 적용하십시오.

1. **비즈니스 도전 과제 파악.** 계정 도용과 허위 등록 제품으로 비즈니스에서 사기 증가를 겪고 있습니다.
2. **명확한 비즈니스 목표 정의**. 계정 손상률을 90% 줄이고 허위 등록 제품 중 95%를 제거합니다.
3. **성공 지표 수립**.
   - 고객 신뢰도를 25% 높입니다. (설문 조사, CSAT)
   - 재정적 손실을 70% 줄입니다.
4. **기계 학습이 적절한 해결 방법인지 결정**. 여러 데이터 포인트와 복잡한 패턴 분석 결과를 토대로 결정을 내립니다.
   - 데이터 가용성: 다양한 고객 및 제품 데이터 사용 가능
   - 문제의 복잡성: ML에 적합함 (협업 필터링, 콘텐츠 기반 필터링)
   - 잠재적 영향: 높음(판매 및 고객 경험에 직접적인 영향)
   - 구현 가능성: Amazon Personalize 또는 SageMaker AI의 사용자 정의 모델로 가능

이 프레임워크는 ML 프로젝트를 구조화하고, 명확한 목표를 설정하고, 비즈니스 목표에 맞게 조정하는 데 도움이 됩니다. 이는 적절한 알고리즘 선택을 안내하고 프로젝트 성공을 측정하는데 도움이 됩니다.

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

기계 학습 훈련은 데이터를 토대로 예측을 하거나 결정을 내리도록 알고리즘을 훈련하는 프로세스입니다. ML의 3가지 주요 카테고리가 있습니다: **지도 학습, 비지도 학습** 및 **강화 학습(RL)**입니다. 각 카테고리는 각각 다른 유형의 문제와 데이터에 적합합니다.

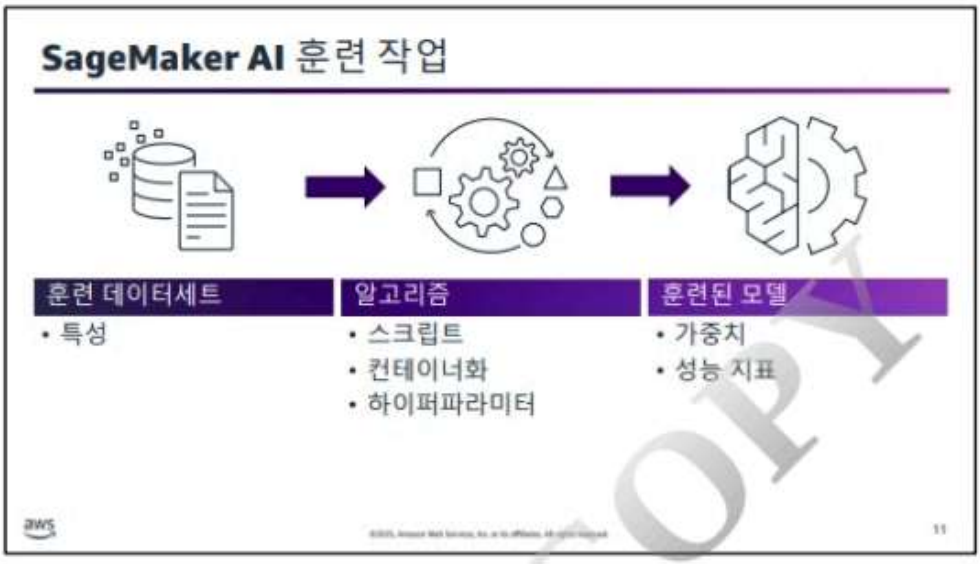
ML **모델**은 데이터에 대한 알고리즘 훈련의 결과입니다. 이는 데이터에서 학습한 패턴과 관계를 나타내며, 새로운, 처음 보는 데이터를 예측하는 데 사용할 수 있어요. 알고리즘의 선택은 모델이 학습하는 방식과 발견할 수 있는 패턴 유형에 영향을 줍니다.

알고리즘과 모델의 관계가 중요합니다. 알고리즘이 다르면 같은 데이터로 학습해도 다른 모델을 만들 수 있습니다. 왜냐면 패턴을 찾고 예측하는 접근 방식이 각각 다르기 때문입니다. 이 관계를 이해하면 주어진 문제와 데이터세트에 가장 적합한 알고리즘을 선택하는 데 도움이 됩니다.

DO NOT COPY ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



SageMaker AI 훈련 작업은 모델 훈련 프로세스의 핵심 구성 요소입니다. 훈련 작업에서는 특성과 레이블이 포함된 **훈련 데이터세트**를 입력으로 사용합니다. SageMaker AI의 기본 제공 알고리즘 중에서 선택할 수도 있고, 모델 아키텍처와 훈련 프로세스를 정의하는 자체 사용자 정의 스크립트를 제공할 수도 있습니다.

작업은 Amazon Elastic Compute Cloud (Amazon EC2) 인스턴스의 **컨테이너**에서 실행되고, SageMaker AI가 구성에 따라 컴퓨팅 리소스를 자동으로 조정합니다. 훈련 중에 **하이퍼파라미터**를 지정해서 모델 성능을 최적화합니다. SageMaker AI는 강력한 실험 추적 기능도 제공해서 다양한 실행과 구성을 비교할 수 있습니다. 파라미터는 모델 훈련 중에 데이터에서 학습이 되는 반면, 하이퍼파라미터는 훈련 전에 설정되고 훈련 과정을 제어합니다.

훈련 작업의 결과는 학습된 가중치 및 성능 지표를 가진 **훈련된 모델**입니다. 그런 다음, 추론을 위해 배포하거나, 더 다듬거나, 앙상블 기법에서 사용할 수도 있습니다. SageMaker AI의 관리형 환경은 인프라 복잡성을 제거하여 교육 프로세스를 간소화하므로, 모델 개발 및 최적화에 집중할 수 있습니다.

DO NOT COPY
ssleeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

ML 훈련 메커니즘은 크게 3가지로 분류됩니다. **지도 학습, 비지도 학습** 및 **강화 학습(RL)**입니다. 각각 목적이 다르고 다양한 유형의 문제에 적합합니다.

지도 학습에서는 레이블이 지정된 데이터를 사용하여 입력을 알려진 출력에 매핑하는 모델을 훈련합니다. 이 메커니즘은 스팸 탐지, 이미지 분류, 판매 예측 등의 예측 작업에 적합합니다. 정확도는 매우 높지만, 레이블이 지정된 데이터가 상당히 필요한데, 이를 얻는데 비용이 많이 들 수 있습니다.

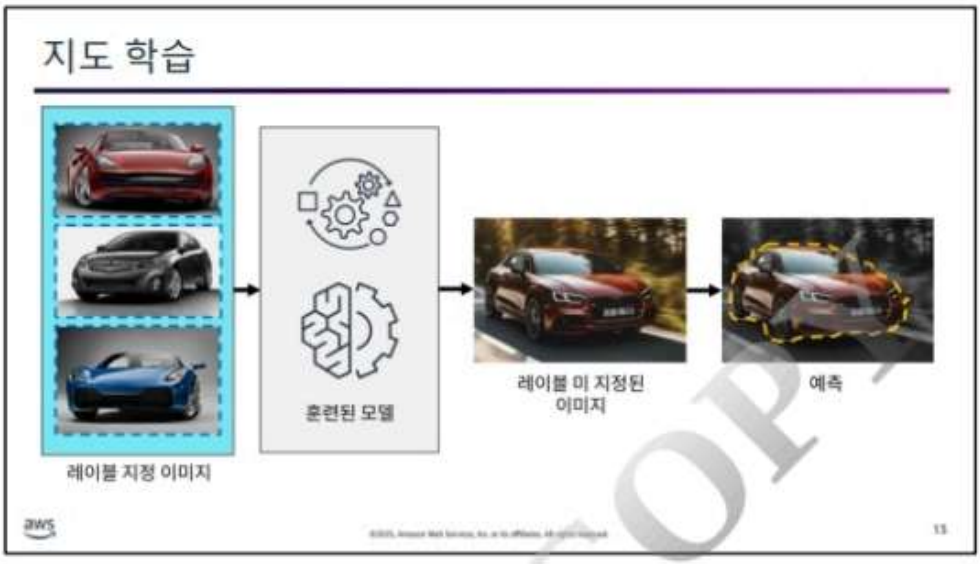
비지도 학습에서는 레이블이 지정되지 않은 데이터에서 패턴을 찾아냅니다. 이는 탐색적 분석, 고객 세분화, 이상 탐지 등에서 유용합니다. 비지도 학습 시에는 레이블이 지정된 데이터는 필요하지 않지만 결과를 해석하기가 까다로울 수 있습니다.

강화 학습(RL)에서는 환경과의 시행착오 상호 작용을 통한 에이전트 학습을 포함합니다. 게임 플레이, 로보틱스, 자율 주행 차량 등의 순차적 의사결정 작업에 효과적입니다. 하지만 계산을 많이 수행해야 할 수 있으며 성능을 최적화하려면 학습을 여러 번 반복해야 할 수도 있습니다.

SageMaker Python SDK의 로컬 모드를 통해 SageMaker의 관리형 서비스를 사용하기 전에 테스트용으로 추정기, 처리기 및 파이프라인을 로컬 환경에 생성하고 배포할 수 있습니다. 이는 특정 종속성과 Docker Compose V2의 설치가 필수이고, YAML 구성 파일을 사용하여 LocalSession의 기본값을 정의합니다.

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



지도 학습 알고리즘은 레이블이 지정된 훈련 데이터에서 학습을 합니다. 여기서 각 입력은 해당 출력과 쌍을 이룹니다. 목표는 새로운, 처음 보는 입력을 올바른 출력에 매핑할 수 있는 함수를 학습하는 것입니다.

지도 학습 문제의 2가지 주요 유형은 **분류**(개별 카테고리 예측)와 **회귀**(연속 값 예측)입니다. 예시로는 이메일 스팸 탐지(분류)와 주택 가격 예측(회귀)등이 있습니다.

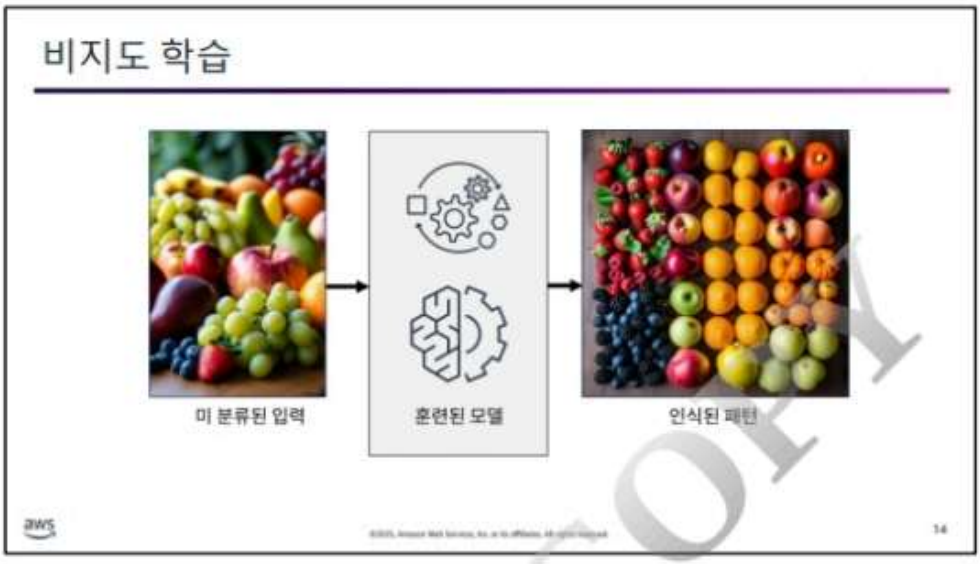
지도 학습은 다양한 영역에서 응용될 수 있습니다.
- 의료: 질병 결과 예측
- 금융: 신용 평가 및 사기 탐지
- 마케팅: 고객 이탈 예측 및 제품 추천

지도 학습을 진행하면 높은 정확도를 달성할 수 있지만 레이블이 지정된 데이터가 상당히 필요합니다. 그런데 이러한 데이터를 얻는데 비용과 시간이 많이 걸릴 수 있습니다. 지도 학습으로 훈련한 모델의 성능은 훈련 데이터의 품질과 대표성에 따라 크게 좌우됩니다.

DO NOT COPY ssleeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



비지도 학습 알고리즘은 레이블이 지정되지 않은 데이터를 사용하여 사전 정의된 대상 변수 없이 숨겨진 패턴이나 구조를 찾아냅니다. 비지도 학습을 진행하는 2가지 주요 기법은 다음과 같습니다.

- **클러스터링:** 비슷한 데이터 포인트를 함께 그룹화 (예: 고객 세분화 및 이상 탐지)
- **차원 감소:** 주요 특성은 유지하면서 고차원 데이터 단순화 (예: 데이터 시각화 및 사전 처리)

비지도 학습의 응용 분야는 다음과 같습니다.

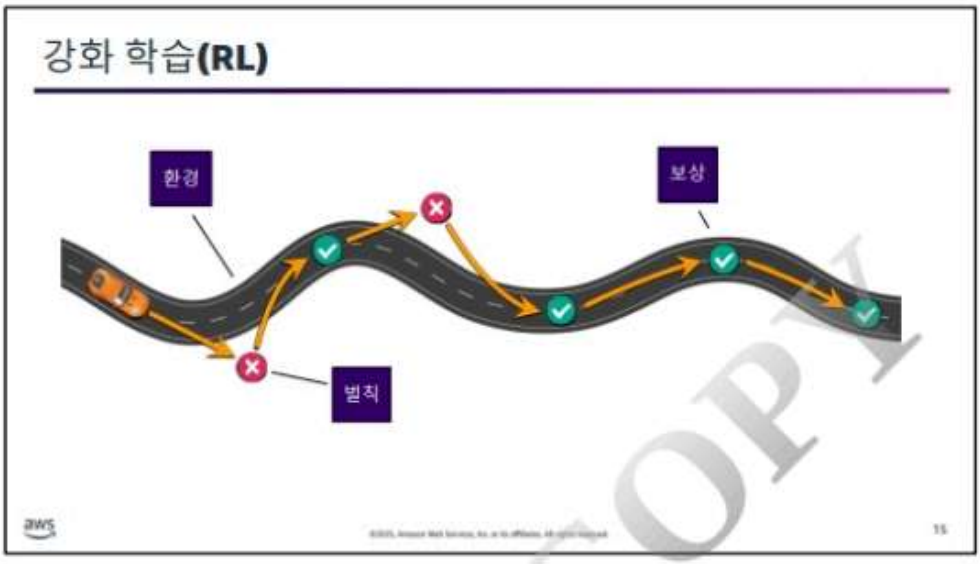
- 마케팅: 고객 세그먼트 기반으로 타겟 캠페인
- 사이버 보안: 네트워크 트래픽에서 비정상적 패턴 탐지
- 생물학: 유전자 발현 패턴 식별

비지도 학습의 주요 이점은 풍부하고 저렴한 레이블 없는 데이터로 작업할 수 있다는 것입니다. 하지만, 정해진 정답의 레이블이 없으면 결과를 해석하기가 어려울 수도 있습니다. 그럼에도 불구하고, 비지도 학습은 탐구적 분석과 인간이 놓칠 수도 있는 인사이트를 파악하는 데 유용한 도구입니다.

DO NOT COPY ssleeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



강화 학습(RL)에서는 에이전트가 환경과 상호 작용을 하면서 의사 결정을 내리는 방법을 학습합니다. 에이전트는 수행한 작업에 대해 보상이나 벌칙을 받으며, 시간 경과에 따른 누적 보상 극대화를 목표합니다.

RL의 주요 구성 요소에는 다음이 포함됩니다.
- **에이전트:** 의사 결정 주체
- **환경:** 에이전트가 작동하는 세계
- **상태:** 환경의 현재 상황
- **작업:** 에이전트가 내리는 결정
- **보상:** 행동의 적합성에 대한 환경 피드백
- **정책:** 행동 선택을 위한 에이전트의 전략

RL은 다음과 같은 다양한 애플리케이션에서 성공을 거두었습니다.
- 게임 플레이: 바둑, 체스와 같은 복잡한 게임에서 인간을 능가하는 실력 발휘
- 로보틱스: 시행착오를 반복하면서 로봇에게 작업 수행을 지도
- 금융: 거래 전략 및 포트폴리오 관리 방식 최적화

RL의 중요한 장점은 명시적인 프로그래밍 없이 복잡한 행동을 학습하고 변화하는 환경에 적응할 수 있다는 것입니다. 하지만, RL은 계산을 많이 수행해야 할 수 있으며 최적의 정책을 찾기 위해 많은 반복이 필요할 수 있습니다. 자율 시스템, 맞춤형 추천, 프로세스 최적화 분야에서의 유망한 애플리케이션으로 여전히 활발히 연구되고 있습니다.

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

Amazon SageMaker AI는 사용 편의성과 사용자 지정 요구 사항 간의 균형을 유지하면서 기계 학습 알고리즘 구현을 위한 다양한 옵션을 제공합니다.

- 기본 제공 알고리즘 - 이미지 분류, 자연어 처리(NLP)와 같은 일반적인 작업을 위한 빠른 모델 배포를 제공합니다. 최소한의 코드만 필요하기 때문에 프로토타입을 빠르게 생성하려는 경우와 ML을 처음 접하는 사용자에게 적합합니다. 널리 사용되는 기본 제공 알고리즘으로는 표 형식 데이터를 위한 XGBoost, 시계열 예측을 위한 DeepAR, NLP 작업을 위한 BlazingText 등이 있습니다. 이들은 광범위하게 사용되고 있는 ML 기술을 최적화해서 구현해줍니다.
- 사용자 정의 스크립트 - TensorFlow, PyTorch 등의 인기 프레임워크와 직접 작성한 코드를 사용할 수 있습니다. 이 옵션은 SageMaker AI의 인프라 및 관리 기능을 사용하면서 유연성을 제공합니다.
- 사용자 정의 컨테이너 - 특수 알고리즘 및 종속성으로 자체 보유 Docker 이미지를 패키징하여 원하는 대로 제어할 수 있습니다. 고급 사용 사례 또는 기존 ML 워크플로를 이식할 때 적합합니다.

문제의 복잡성, 사용자 지정 요구 사항, 팀의 전문 지식성과 같은 요인에 따라 선택하면 됩니다. SageMaker AI의 유연성은 ML 프로젝트 진행 과정에서 더욱 수준 높은 기능을 점진적으로 채택하는데 도움이 됩니다.

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

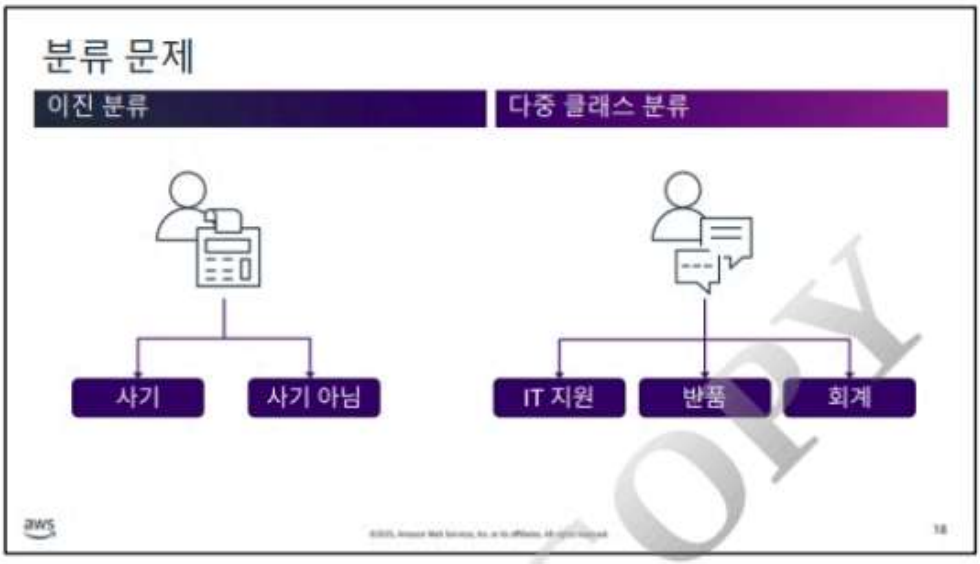
Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

프로젝트 성공을 위해서는 적절한 ML 알고리즘을 선택해야 합니다. 이 섹션에서는 다양한 훈련 알고리즘에 대해 살펴봅니다.

DO NOT COPY
ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



분류는 모델이 입력 데이터에서 범주형 레이블을 예측하는 지도 학습 작업입니다. 분류의 2가지 주요 유형은 다음과 같습니다.

- **이진 분류** – 2개 클래스 중 하나를 예측합니다. (예: 스팸 또는 스팸 아님, 사기 또는 사기 아님) 이 분류가 적합한 알고리즘으로는 로지스틱 회귀, 서포트 벡터 머신(SVM), 의사결정 트리 등이 있습니다.
- **다중 클래스 분류** – 3개 이상의 클래스 중 하나를 예측합니다. (예: 이미지 인식과 감정 분석) 이러한 문제에 적합한 알고리즘은 랜덤 포레스트, 신경망, 나이브 베이즈 등입니다.

알고리즘의 선택은 데이터세트의 크기, 특성 수, 해석 가능성 요구 사항 같은 요인에 따라 달라집니다. 분류의 평가 지표로는 정확도, 정밀도, 재현율, F1 점수 등이 있습니다.

분류는 다음을 비롯한 여러 업계에서 실제로 응용되고 있습니다.

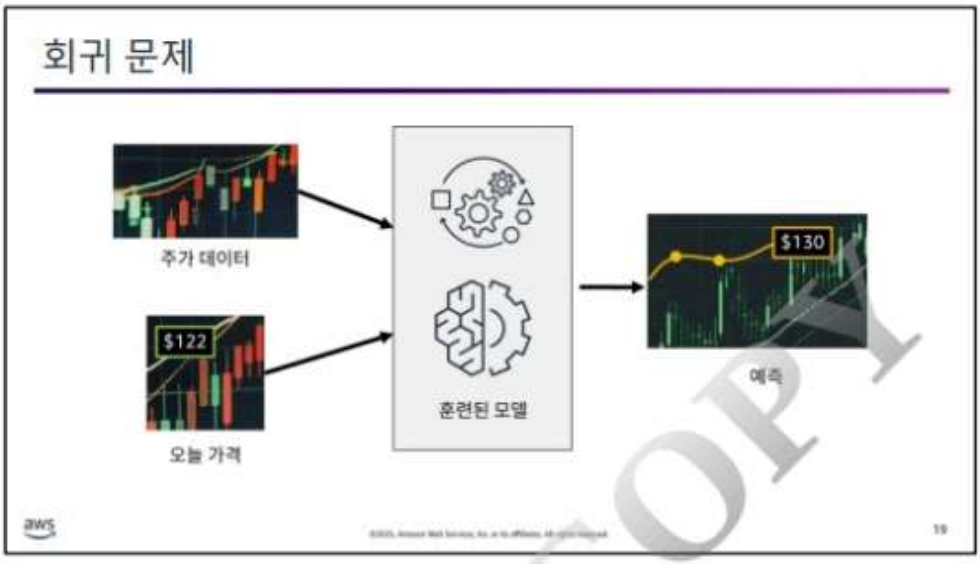
- 의료: 질병 진단
- 금융: 신용 평가, 사기 행위 탐지
- 마케팅: 고객 세분화, 고객 이탈 예측

이러한 모델은 과거의 예제에서 학습한 패턴을 토대로 새 데이터를 분류하므로, 데이터를 기반으로 결정을 할 수 있습니다.

DO NOT COPY ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



회귀는 입력 특성을 토대로 연속적인 수치 값을 예측하는 지도 학습 작업입니다. 변수 간의 관계를 모델링하고 특정 요인의 영향을 예측 및 이해하는 데 사용됩니다.

흔히 사용되는 회귀의 유형은 다음과 같습니다.
- 선형 회귀: 선형 관계를 모델링합니다.
- 다항식 회귀: 비선형 패턴을 캡처합니다.
- 다중 회귀: 여러 입력 변수를 사용합니다.
- 릿지와 라쏘 회귀: 과대적합 방지를 위해 정규화를 추가로 수행합니다.

회귀는 다음과 같은 애플리케이션에 적용되고 있습니다.
- 금융: 주가 예측, 위험 평가
- 부동산: 부동산 평가
- 마케팅: 고객 평생 가치 추정
- 환경 과학: 기후 모델링

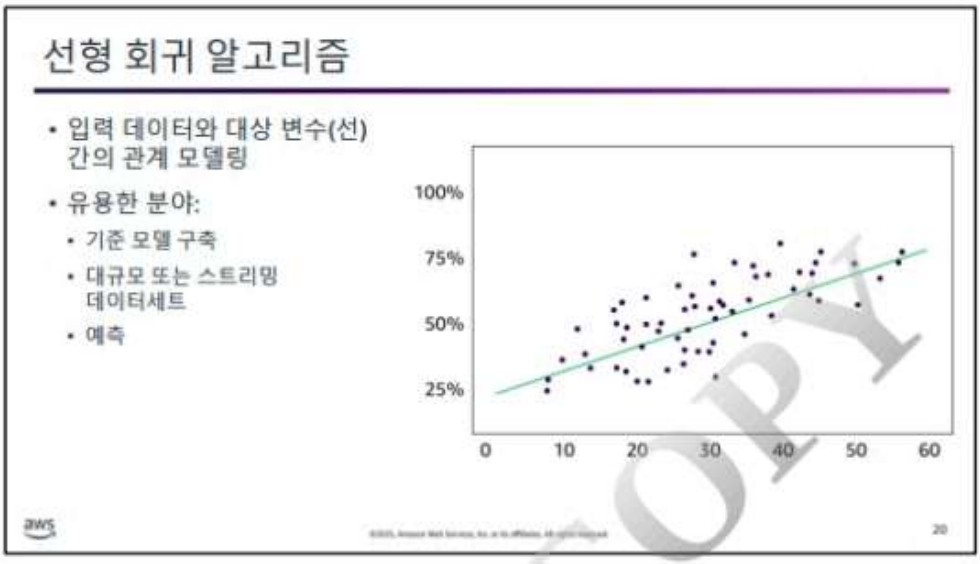
회귀를 적용할 때 주요 고려 사항은 다음과 같습니다.
- 선형성 가정
- 이상치 유무
- 특성 간의 다중 공선성

회귀 모델의 평가 지표로는 평균 제곱 오차(MSE), 평균 제곱근 오차(RMSE), R-제곱이 포함됩니다. 이는 모델의 성능과 적합성을 평가하는데 도움이 됩니다.

DO NOT COPY ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



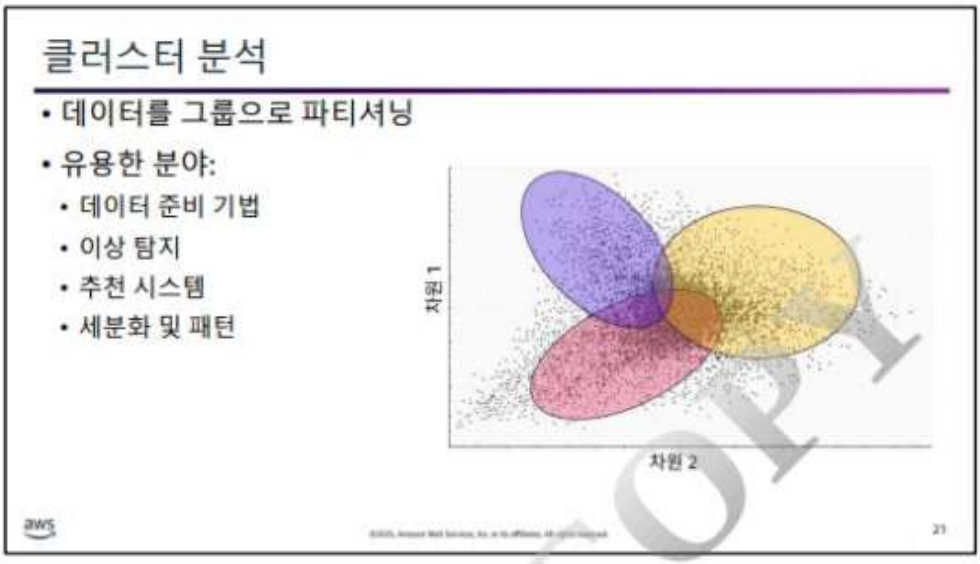
선형 회귀는 관찰된 데이터에 선형 방정식을 맞추어 입력 변수와 연속 대상 변수 간의 관계를 모델링하는 기초적인 기계 학습 알고리즘입니다. 가장 단순한 형태로, 예측 값과 실제 값 간의 제곱근의 차이 합이 가장 적은 직선(또는 더 높은 차원의 초평면) 찾기를 시도합니다. 이 접근 방식은 입력과 출력 간의 관계를 선형 함수로 적절하게 근사화할 수 있는 문제에 특히 적합합니다.

선형 회귀를 위한 훈련 과정에서는 비용 함수, 보통 평균 제곱 오차를 최소화하는 각 특성의 최적 계수(가중치)을 찾는 단계가 포함됩니다. 이 최적화는 일반 최소 제곱 같은 분석 방법이나 경사 하강법 등의 반복 접근법을 통해 계수를 최적화할 수 있습니다. 후자는 데이터세트가 크거나 스트리밍 형식인 기계 학습 상황에서 특히 중요합니다. 새 데이터가 유입되면 모델이 점진적으로 적응할 수 있기 때문입니다.

단순한 알고리즘이기는 하지만, 선형 회귀는 현재 기계 학습 워크플로에서 기준 모델, 특성 중요도 분석기, 더 복잡한 아키텍처 내의 구성 요소로 계속 사용되고 있습니다. 해석이 용이하기 때문에, 각 입력 특성이 예측에 어떻게 기여하는지 이해할 수 있어서 설명이 예측 정확도만큼 중요한 영역에서 유용합니다. 선형성, 독립성, 상수 분산 요건이 합리적으로 충족된다면 선형 회귀를 통해 최소한 계산 오버헤드로 정확한 예측을 생성할 수 있습니다.

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



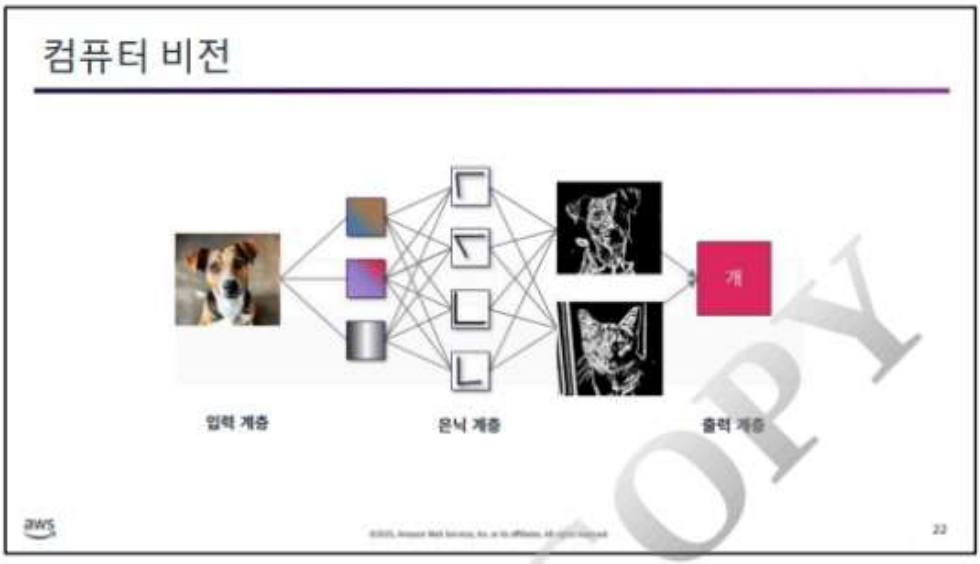
클러스터 분석은 데이터를 그룹(클러스터)으로 분할하여 같은 클러스터 내의 객체가 다른 클러스터에 있는 객체보다 더 유사하도록 하는 것을 목표로 하는 기계 학습의 기본 기술입니다. 지도 학습과는 달리 클러스터 분석은 사전 정의된 클래스 레이블을 사용하지 않고 작동하는 '비지도 학습' 접근 방식입니다. 이 알고리즘은 데이터의 고유한 특성과 관계를 기반으로 자연스러운 그룹화 방식을 찾아냅니다. 유사도는 보통 유클리트, 맨해튼, 코사인 거리 등의 거리 지표를 통해 측정됩니다.

클러스터 분석 애플리케이션은 고객 세분화, 이미지 인식, 이상 탐지, 문서 분류 등의 여러 영역에 걸쳐 있습니다. 예를 들어, 클러스터링을 사용하여 구매 행동이 비슷한 고객 세그먼트를 식별해서 타켓 마케팅 전략을 사용할 수 있습니다. 생의학 연구 시에는 클러스터링으로 여러 차례 실험에서 발현 패턴이 비슷한 유전자 그룹을 식별하는데 도움이 됩니다. 이 기법의 다양성은 대규모 데이터세트를 유의미한 구조로 구성하여 복잡성을 줄이는 능력에서 비롯됩니다.

각 클러스터링 알고리즘은 클러스터를 형성하기 위해서 다양한 전략을 사용합니다. **K-평균**에서는 가장 가까운 중심에 점을 반복 할당하고 그 중심을 업데이트해서 클러스터 내의 분산을 최소화합니다. **계층적 클러스터링**에서는 상향식(응집) 또는 하향식(분할) 접근 방식으로 중첩 클러스터를 구축합니다.
DBSCAN와 같은 밀도 기반 방법에서는 클러스터를 희소 영역으로 구분된 밀집 영역으로 파악하여 노이즈를 효과적으로 처리하고 임의의 모양으로 형성된 클러스터를 찾습니다. 알고리즘의 선택은 구체적인 데이터 특성, 계산 제약 조건 및 결과 클러스터 용도에 따라서 달라집니다.

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



컴퓨터 비전용 딥 러닝에서는 '계층'을 통해 계층적 특성을 추출합니다.

- 입력 계층: 원본 이미지의 원시 픽셀 데이터 처리
- 은닉 계층: 갈수록 복잡해지는 특성 (가장 자리, 모서리, 객체 부분) 탐지
- 출력 계층: 최종 이미지 분류(예: '강아지') 생성

이와 같은 계층적 특성 학습은 **합성곱 신경망(CNN)**의 특징입니다, 이는 분류, 객체 탐지, 세분화 등의 이미지 관련 태스크를 매우 효율적으로 처리합니다.

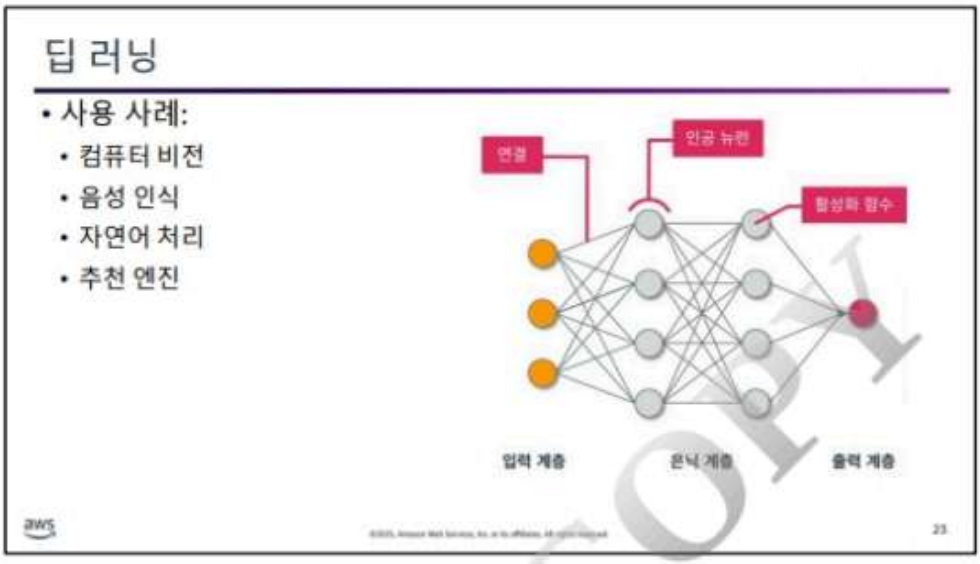
Amazon Rekognition 등의 AWS 서비스는 이와 비슷한 아키텍처를 사용해 강력한 컴퓨터 비전을 제공합니다. 이러한 기능은 다양한 업계에서 콘텐츠 조정 및 객체 탐지가 포함됩니다.

컴퓨터 비전을 사용하는 경우, 기계가 세상의 시각적 정보를 해석 및 분석할 수 있으므로 자율 주행 차량, 의료 영상 촬영, 증강 현실 등의 분야에서 가능성을 열고 있습니다.

DO NOT COPY
ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



딥 러닝(DL)은 컴퓨팅 하드웨어의 진전으로 인해 가능해진 AI의 상당한 발전이라 할 수 있습니다. DL 알고리즘은 인간 두뇌의 작동 방식을 토대로 개발된 **인공 신경망(ANN)**을 사용합니다.

- 인공 뉴런 계층 (수학 함수)
- 각 계층이 정보를 요약하여 다음 계층으로 전달함
- 알고리즘이 훈련 중에 특성을 직접 도출함

딥 러닝의 주요 사용 사례는 다음이 포함됩니다:

- 컴퓨터 비전 (이미지 인식, 객체 탐지)
- 음성 인식 및 합성
- 자연어 처리 (번역 및 감정 분석)
- 추천 시스템

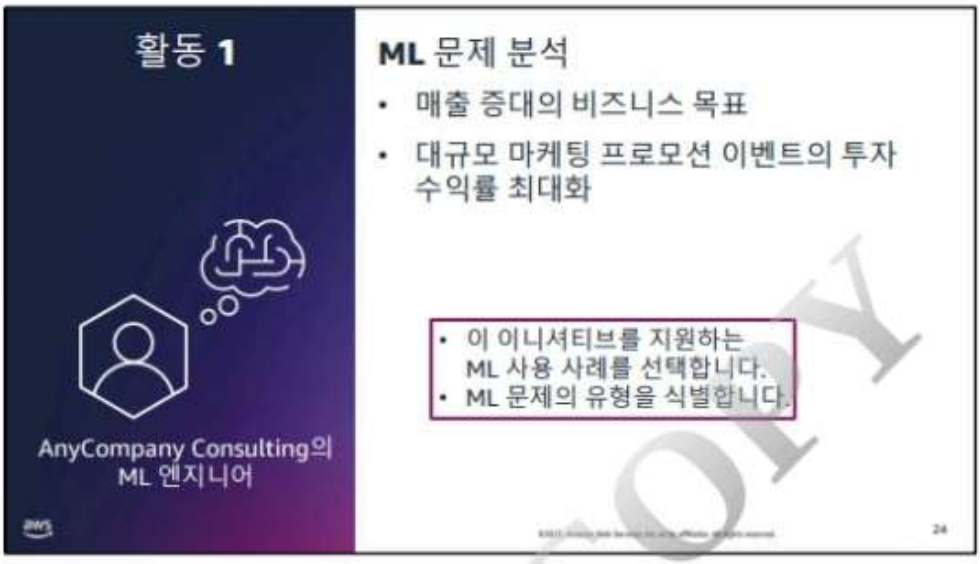
AWS는 포괄적인 DL을 제공합니다:

- 모델 빌드, 훈련, 배포를 위한 Amazon Sagemaker AI
- 사전 훈련된 AI 서비스(Amazon Rekognition, Amazon Transcribe, Amazon Comprehend)
- 고성능 훈련을 위한 그래픽 처리 장치(GPU)로 가속되는 EC2 인스턴스

데이터에서 복잡한 특성을 자동으로 학습하는 딥 러닝의 기능으로 많은 AI 애플리케이션에 획기적인 발전이 일어났고, 이전에는 어려웠던 영역에서도 혁신을 이루는데 도움이 되었습니다.

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



사용 사례 시나리오:

여러분은 획기적인 프로모션 캠페인을 전문으로 하는 유명 마케팅 업체인 AnyCompany Consulting 소속 ML 엔지니어입니다. AnyCompany는 다음 회계 연도에 고객 매출을 20% 늘리는 야심찬 목표를 세웠습니다. 이 목표를 달성하기 위해, AnyCompany는 가전 업계의 주요 고객을 대상으로 대규모 다중 채널 마케팅 캠페인을 계획하고 있습니다.

AnyCompany의 CEO는 중요한 과제를 가지고 당신을 찾아왔습니다. 바로 다가오는 프로모션 이벤트의 투자 수익률(ROI)을 최대한 높일 수 있는 인공 지능 및 기계 학습(AI/ML) 솔루션을 파악하여 구현해야 합니다. 캠페인에서는 소셜 미디어, 이메일 마케팅, 타겟 온라인 광고, 매장 내 프로모션 등의 여러 마케팅 채널이 포함됩니다.

캠페인의 규모와 복잡성을 감안할 때, 경영진은 AI/ML 기술을 활용하면 마케팅 전략 효과를 대폭 높임으로써 궁극적으로는 고객 매출을 높일 수 있을 것으로 보고 있습니다. 사용할 수 있는 데이터세트에는 인구 통계 정보와 소득 데이터가 포함되어 있습니다. 이는 타겟 마케팅 활동에서 유용할 수 있습니다.

여러분의 과제는 마케팅 캠페인 요구 사항, 사용 가능 데이터 및 잠재적인 AI/ML 애플리케이션을 분석하여 이 이니셔티브를 지원할 수 있는 솔루션을 제안해야 합니다. 고객 타겟팅, 콘텐츠 개인화, 성과 예측 등 캠페인의 다양한 측면을 개선하기 위해 ML을 적용할 수 있는 구체적인 사용 사례를 파악해야 합니다.

제안서 작성 시에는 각 사용 사례에 해당하는 ML 문제의 유형(예: 분류, 회귀, 클러스터링), 그리고 그것이 매출 증대와 ROI 극대화라는 전반적인 비즈니스 목표에 어떻게 부합하는지도 고려해야 합니다.

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

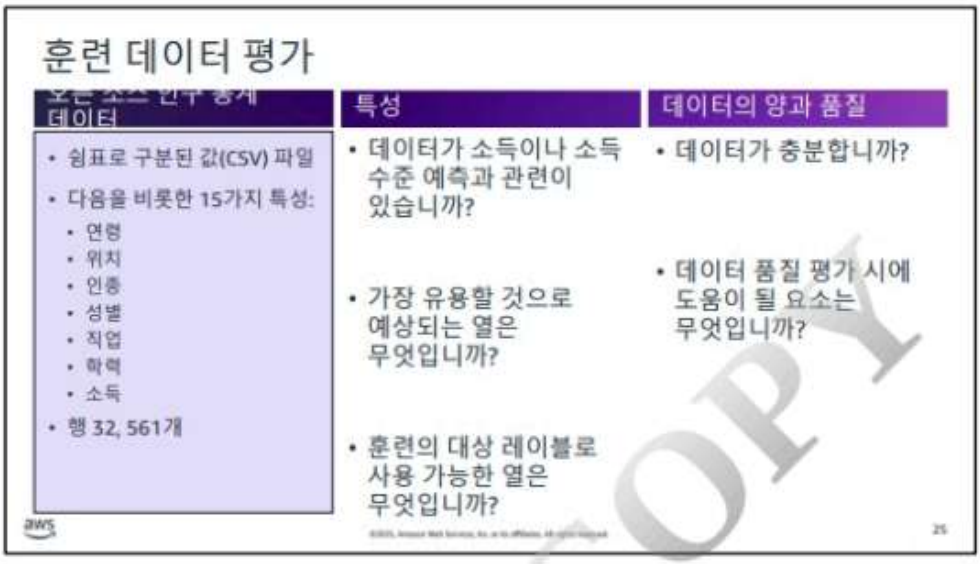
AI/ML 통합 목표 달성 여부는 고객 참여율, 전환율, 궁극적으로는 마케팅 캠페인에 기인한 판매 증가와 같은 핵심 성과 지표에 미치는 영향으로 측정됩니다.

여러분은 AnyCompany 경영진에게 이 대규모 마케팅 프로모션을 지원하고 효과를 높이는 데 AI/ML을 사용할 수 있는 방법을 명확하게 설명하면서 발견된 사항과 권장 사항을 제시하는 것입니다.

DO NOT COPY
ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



제공된 데이터세트를 평가합니다.

특성: 데이터세트에는 연령, 위치, 인종, 성별, 직업, 학력, 소득 등의 15개 열이 포함되어 있습니다. 구매 방식 예측과 관련성이 가장 높은 것으로 보이는 주요 특성은 소득, 직업, 위치, 학력 등입니다.

데이터의 양: 특성 15개에 32,561개 행으로 구성되어 있습니다. 이는 ML 모델을 훈련하기에 충분한 데이터세트를 제공합니다.

품질 관련 고려 사항:
- 대표성: 데이터에 목표 고객층을 반영하는지 평가합니다.
- 분산: 소득 수준과 기타 주요 특성의 균형을 확인합니다.
- 완전성: 누락된 값을 확인하여 처리합니다.
- 관련성: 특성이 구매 패턴과 연관지을 수 있는지 판단합니다.

다음 단계:
- 탐색적 데이터 분석을 수행하여 심층 인사이트를 파악합니다.
- 피처 엔지니어링 또는 추가 데이터 소스 필요 여부를 평가합니다.
- 인구 통계 데이터 사용 시의 윤리적 측면을 고려합니다.

기계 학습에 적용되는 일반적인 지침 중 하나인 10X 규칙에서는 특성보다 최소 10배 많은 샘플을 확보할 것을 제시하고 있습니다. 하지만 엄격한 요구 사항인 것은 아니며 특정 문제, 모델 복잡도 및 데이터 특성에 따라 달라질 수 있습니다. 즉, 이 규칙을 기준점으로 사용할 수는 있지만, SageMaker나 특정 서비스에서 국한된 것이 아니며, 실제로는 데이터 품질, 피처 엔지니어링, 적절한 교차 검증 등의 요인이 모델 성능에 더 중요한 영향을 주는

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

경우가 많습니다.

DO NOT COPY
ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

ML 성공 기준 프레임워크에서는 업무상의 특정 당면 과제를 해결하는 데 ML이 적합한지 여부를 평가할 수 있습니다.

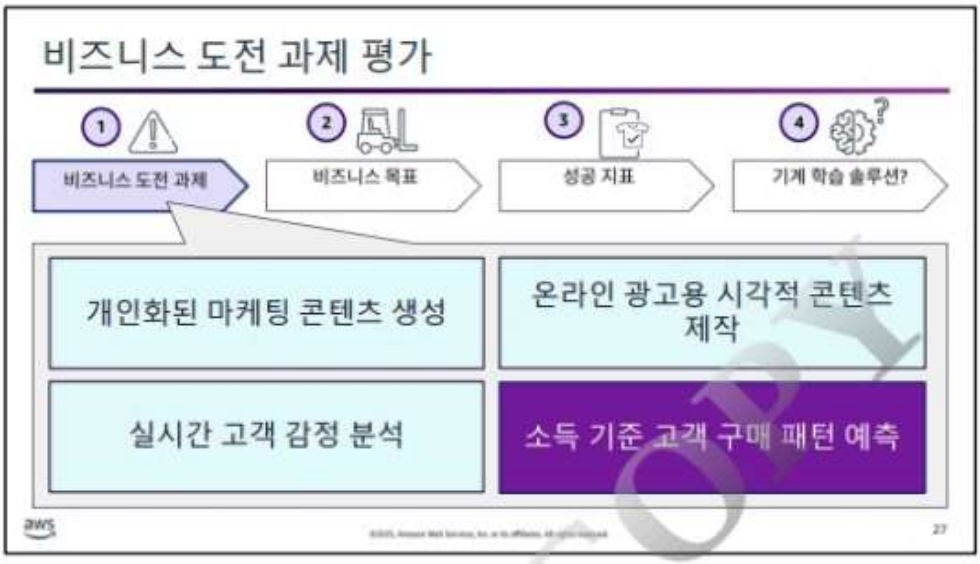
1. **비즈니스 도전 과제 파악.** AnyCompany에서 사기 행위가 증가하여, 신뢰, 재정 그리고 평판을 위협하고 있습니다.
2. **명확한 목표 정의.** 강력한 사기 탐지 기능을 구현하여 재정적 수익을 보호하고 마켓플레이스 보안을 유지해야 합니다.
3. **성공 지표 수립.**
   - 6개월 내에 사기 거래를 80% 줄입니다.
   - 손상된 계정 수를 분기 내에 90% 줄입니다.
   - 30일 내에 허위 등록 제품 중 95%를 찾아냅니다.
4. **ML 사용 적합 여부 결정.** ML이 적합한 이유는 다음과 같습니다.
   - 대량의 다양한 데이터 사용 가능
   - 여러 데이터 소스에서 패턴을 탐지해야 함
   - 갈수록 발전하는 사기 수법에 대응할 수 있는 적응형 알고리즘 필요

ML 기반 사기 탐지를 구현하면 데이터를 사용하여 보안, 신뢰, 재정적 성과를 향상시켜 AnyCompany의 목표와 일치합니다.

DO NOT COPY ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



AnyCompany의 마케팅 캠페인에 ML 성공 기준 프레임워크를 적용합니다.

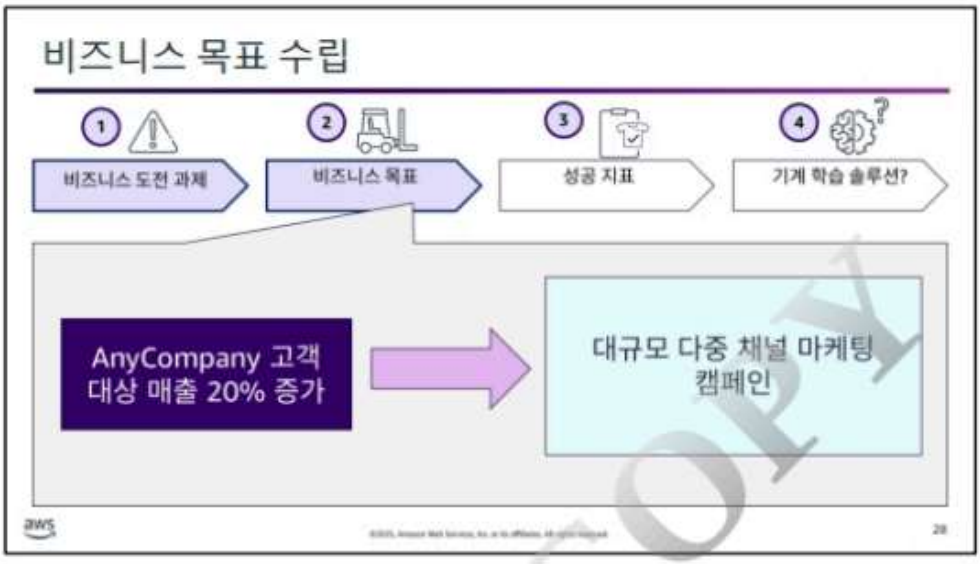
1. **업무상의 당면 과제**:
   a) 개인화된 마케팅 콘텐츠 생성 (생성형 AI)
   b) 온라인 광고용 시각적 콘텐츠 제작 (딥 러닝)
   c) 실시간 고객 감정 분석 (딥 러닝)
   d) 소득 기준 고객 구매 패턴 예측 (기계 학습)
2. **명확한 목표**: 다중 채널 마케팅 캠페인을 진행하여 다음 회계 연도의 AnyCompany 고객의 매출 20% 증가
3. **성공 지표**: (다음 슬라이드에서 정의함)
4. **ML 사용 적합 여부**:
   - 기존 ML에 가장 적합한 도전 과제는 (d)입니다.
   - 도전 과제 (a)~(c)는 딥 러닝이나 생성형 AI를 활용하면 더욱 효율적으로 해결할 수 있습니다.

ML을 사용하여 소득 기준으로 구매 패턴을 예측하는데 집중합니다. ML은 매출 성장 목표를 직접 지원하고, 사용 가능한 인구통계학적 데이트를 사용하기 때문입니다.

DO NOT COPY ssleeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



비즈니스 목표와 ML 접근 방식을 더 자세히 설명합니다.

주요 사업 목표: 대규모 다 채널 마케팅 캠페인을 진행하여 다음 회계 연도의 AnyCompany의 고객 매출 20% 증가

ML 엔지니어의 태스크: 다가오는 소비자 가전 업계에서 진행 예정인 프로모션 이벤트의 ROI를 최대한 높일 수 있는 AI/ML 솔루션 파악 및 구현

AI/ML 통합 목표:
- 고객 타겟팅 정확도 향상
- 여러 채널에서 개인화된 콘텐츠 제공
- 캠페인 성과 예측
- 실시간으로 마케팅 전략 최적화

채널: 소셜 미디어, 이메일, 온라인 광고, 매장 내 프로모션

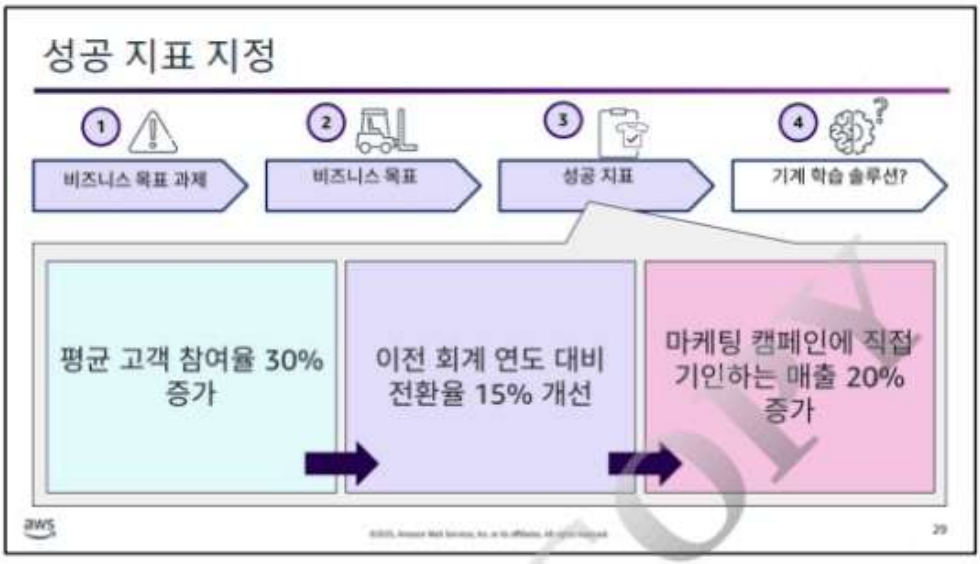
예상 결과:
- 참여율 개선
- 전환율 개선
- 전반적인 매출 증가

ML을 사용해 소득 및 기타 인구 통계적 요인을 토대로 구매 패턴을 예측하면 AnyCompany는 캠페인 ROI를 극대화하면서 판매를 늘리겠다는 가장 중요한 목표에 맞춰 마케팅 활동을 더 효과적으로 조정할 수 있습니다.

DO NOT COPY ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



구체적이고 측정 가능한 성공 기준을 수립합니다.

고객 참여율:
- 지표: 모든 채널의 평균 참여율
- 임곗값: 전년 대비 30% 증가
- 측정 방법: 주간 추적, 기준과 비교

전환율:
- 지표: 참여한 고객 중 구매를 하는 고객 비율
- 임곗값: 이전 회계 연도 대비 15% 증가
- 측정 방법: AI 기반 전략의 A/B 테스트, 예측 분석

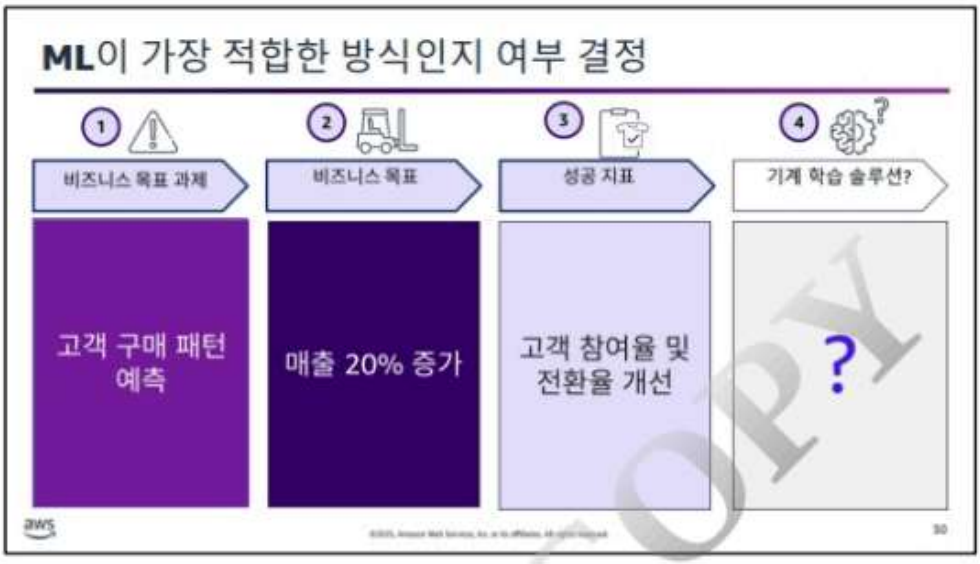
매출 증가 기여도:
- 지표: 캠페인에 직접 연결된 매출 증가
- 임곗값: 판매 수익 최소 20% 증가
- 측정 방법: 속성 모델링, 회귀 분석

이 프레임워크를 체계적으로 적용하면 ML 이니셔티브가 비즈니스 목표에 부합하며 성공 가능성을 높일 수 있습니다.

DO NOT COPY ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



실제 사례에 ML 성공 기준 프레임워크를 적용합니다.

1. **업무상의 도전 과제**: 소득 기준 고객 구매 패턴 예측. 이는 다양한 소득 구간의 소비자 행동 관련 인사이트를 파악할 수 있습니다.
2. **목표**: 타겟 마케팅 캠페인을 진행하여 매출 20% 증가. 이처럼 정량화할 수 있는 목표를 명확하게 수립하면 ML 진행 방향을 결정할 수 있습니다.
3. **성공 지표**: 고객 참여 및 전환율 개선. 목표와 직접적으로 연관된 이러한 지표를 캠페인 전반에서 측정할 수 있습니다.

이 프레임워크는 ML 솔루션이 명확한 목표와 측정 가능한 결과로 특정 비즈니스 요구를 해결하도록 접근 방식을 구조화 하는데 도움이 됩니다. ML 솔루션이 적합한지 여부를 결정하는 마지막 단계에서는 사용 가능한 데이터, 문제의 복잡성 및 비즈니스 목표에 미칠 수 있는 잠재적 영향을 신중하게 고려해야 합니다.

DO NOT COPY ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

시나리오에 ML 성공 기준 프레임워크를 적용합니다.

1. **비즈니스 도전 과제**: 다중 채널 마케팅 캠페인을 위한 콘텐츠를 최적화하고 고객 타겟팅을 개인화합니다.
2. **목표**: 마케팅 효과 개선을 통해 다음 회계 연도에 매출을 20% 늘립니다.
3. **성공 지표**:
   - 고객 참여율을 30% 높입니다.
   - 전환율을 15% 높입니다.
   - 캠페인으로 인한 매출 수익을 20% 늘립니다.
4. **ML 솔루션의 적합성 여부**: ML은 이 도전 과제에 적합합니다. 복잡한 고객 데이터를 처리하여 행동 방식을 예측하고 콘텐츠를 최적화하고, 대규모로 마케팅 전략을 개인화할 수 있기 때문입니다.

대규모 데이터세트를 분석하여 고객 행동 방식의 패턴을 식별할 수 있으므로 기계 학습이 특히 효과적입니다. 따라서 더 타겟팅되고, 효과적인 마케팅 전략을 수립할 수 있습니다. ML 모델의 지속적인 학습 기능은 캠페인을 실시간으로 최적화하여, 매출 증가라는 목표를 직접적으로 지원할 수 있습니다.

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

솔루션 - ML 문제 유형 이해

태스크	태스크 설명	ML 영역	훈련 데이터의 레이블 설명
1	소득 값을 예측합니다.	지도	소득 숫자 값
2	소득이 지정된 범위 내에 포함인지, 초과인지, 미만인지를 예측합니다.	지도	소득 범위에 [초과, 포함, 미만]
3	소득이 임곗값을 초과하는지 아니면 미만인지를 예측합니다.	지도	소득 임곗값에 [초과, 미만]
4	소득 수준 포함, 고객과 가장 유사한 개별 그룹을 식별합니다.	비지도	레이블 없음 소득 특성 - 숫자 값

32

ML 문제 유형을 정확하게 파악하는 것은 적절한 알고리즘과 접근법을 선택하는데 필수입니다.

소득 관련 예측에서는 다음과 같은 여러 문제 유형을 작성할 수 있습니다.

1. 회귀 - 정확한 소득 값을 예측합니다. 숫자 레이블이 있는 지도 학습 태스크입니다.
2. 멀티클래스 분류 - 소득 카테고리(예: 저소득, 중위, 상위)를 예측합니다. 범주형 레이블이 있는 지도 학습 태스크입니다.
3. 이진 분류 - 소득이 임곗값을 초과하는지 아니면 미만인지를 예측합니다. 이진 레이블을 이용한 지도 학습 태스크입니다.
4. 클러스터링 - 소득을 포함한 유사성을 토대로 고객을 그룹화합니다. 사전 정의된 레이블이 없는 비지도 학습 태스크입니다.

이진 분류는 컴퓨팅 리소스 및 훈련 시간 측면에서 일반적으로 더 비용 효율적입니다. 그러나 데이터세트가 작으면 큰 차이가 없을 수도 있습니다. 선택은 또한 사용 가능한 데이터에 따라서 달라집니다. 만약 데이터가 자연스럽게 여러 카테고리로 나뉜다면, 멀티클래스 분류가 더 적합할 수도 있습니다.

문제 유형의 선택은 특정 비즈니스 요구와 사용 가능한 데이터에 따라 달라집니다. 마케팅 캠페인의 경우, 소득 카테고리를 예측하거나 소득이 임곗값을 초과하는지 여부를 예측하는 이진 분류가 타겟 마케팅 전략에 가장 유용할 수 있습니다.

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

아래는 프로젝트를 위한 ML 성공 기준 프레임워크입니다.

1. **비즈니스 도전 과제**: 구매 패턴을 기반으로 고객 소득을 예측합니다.
2. **목표**: 타겟 마케팅을 진행하여 매출을 20% 늘립니다.
3. **성공 지표**: 고객 참여 및 전환율을 개선합니다. 클릭 방문 지표가 필요합니다.
4. **ML 접근 방식**: 예, 지도 학습을 사용하여 구매 내역을 토대로 고객 소득을 예측합니다.

이 프레임워크에서 제안하는 방식은 **지도 학습**, 특히 소득 예측 정의 방식(정확한 값 또는 범주)에 따라 회귀 또는 분류 태스크를 진행할 수 있습니다. 지도 학습을 선택하는 것이 적절한 이유는 다음과 같습니다.

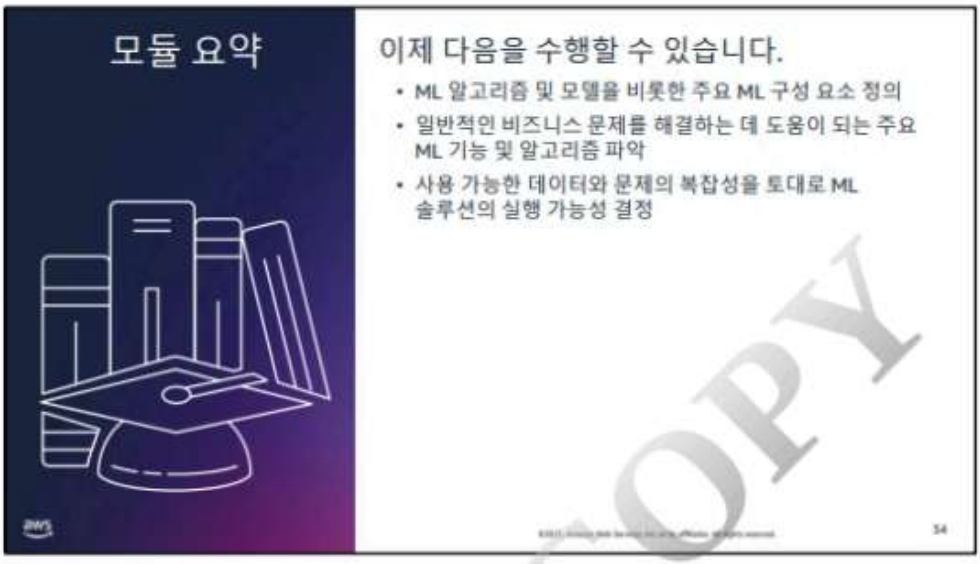
- 레이블이 지정된 데이터(과거 고객의 알려진 소득 값)가 있습니다.
- 학습해야 하는 명확한 관계(구매 패턴과 소득)가 있습니다.
- 결과가 타겟 마케팅 목표 달성을 직접적으로 지원합니다.

고객 소득을 예측하면, 마케팅 전략을 조정하고, 제품 추천을 최적화하고, 가격을 개인화하여, 이 모든 것이 판매 및 참여 증가에 기여할 수 있습니다.

DO NOT COPY ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



이 모듈을 완료했으므로 이제 다음을 수행할 수 있습니다.

- 알고리즘과 모델을 비롯한 주요 ML 구성 요소를 정의하고, 비즈니스 문제 해결에서의 역할 이해
- 분류, 회귀, 클러스터링과 같은 일반적인 비즈니스 문제를 해결하는 ML 기능과 알고리즘 파악
- 인공 신경망(ANN)이 딥 러닝 애플리케이션을 지원하는 방식 설명
- 생성형 AI 기술의 토대가 되는 파운데이션 모델(FM) 및 대규모 언어 모델(LLM) 설명
- 사용 가능한 데이터, 문제의 복잡성 및 비즈니스 요구 사항에 따라 ML 솔루션 실행 가능성 평가
- Amazon SageMaker 및 SageMaker Studio 관련 주요 개념 및 엔드 투 엔드 ML 워크플로에서 이러한 서비스가 제공하는 이점 파악

이러한 기술은 비즈니스 가치를 창출할 수 있는 효과적이며 책임감 있는 솔루션을 개발하려는 ML, 데이터 과학 및 AI 분야의 전문가에게 필수입니다. 이 과정을 진행하면서 실제 시나리오에 이러한 개념을 적용해 보고, 이 분야의 최신 기술 관련 정보를 지속적으로 파악하십시오.

DO NOT COPY
ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

DO NOT COPY
ssleeeee1@gmail.com

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

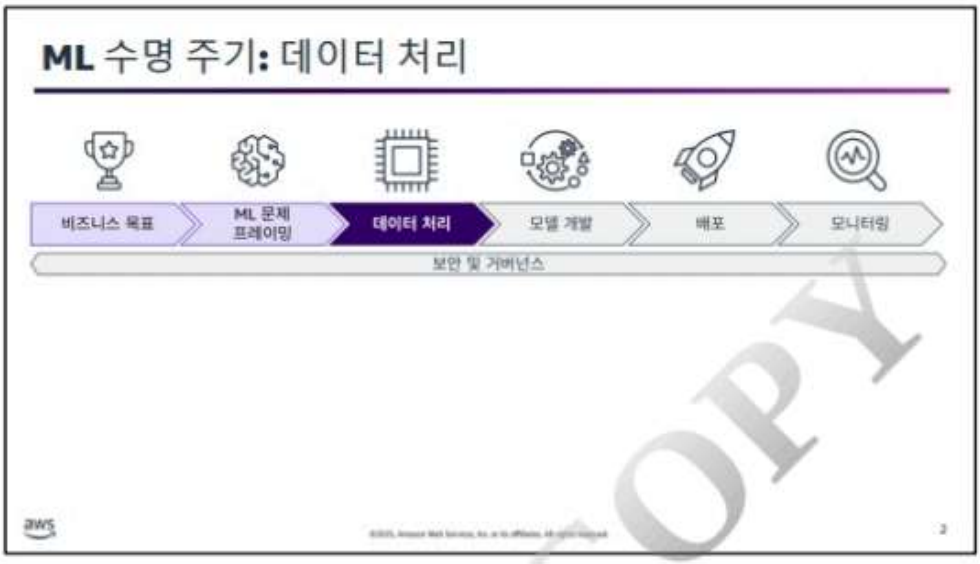
이 모듈에서는 기계 학습(ML) 수명 주기의 초기 단계 중 하나인 데이터 처리에 대해 중점적으로 살펴봅니다. 효과적인 데이터 처리는 성공적인 ML 모델 개발 및 배포의 토대를 마련합니다. 이 모듈에서는 유용한 ML 모델을 훈련하기 위해서 데이터를 준비하는 주요 개념과 기술을 살펴봅니다.

데이터 처리에서는 원시 데이터를 수집 및 정리, 변환하고 분석과 모델 훈련에 적합한 형식으로 구성합니다. 데이터 품질과 관련성이 모델 성능에 직접적인 영향을 주기 때문에 이 단계가 중요합니다. 이 모듈에서는 AWS에서 더 정확하고 신뢰도가 높은 ML 모델을 구축하는데 도움이 되는 데이터 준비 모범 사례를 배우게 됩니다.

DO NOT COPY
ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



데이터 처리는 성공적인 모델 개발을 위한 토대를 마련하는 기계 학습 수명 주기의 중요한 단계입니다. 이 단계에서는 분석 및 모델 훈련을 위한 원시 데이터를 준비하는 작업을 중점적으로 수행합니다. 효과적인 데이터 처리는 ML 모델의 품질과 성능에 상당한 영향을 미칩니다

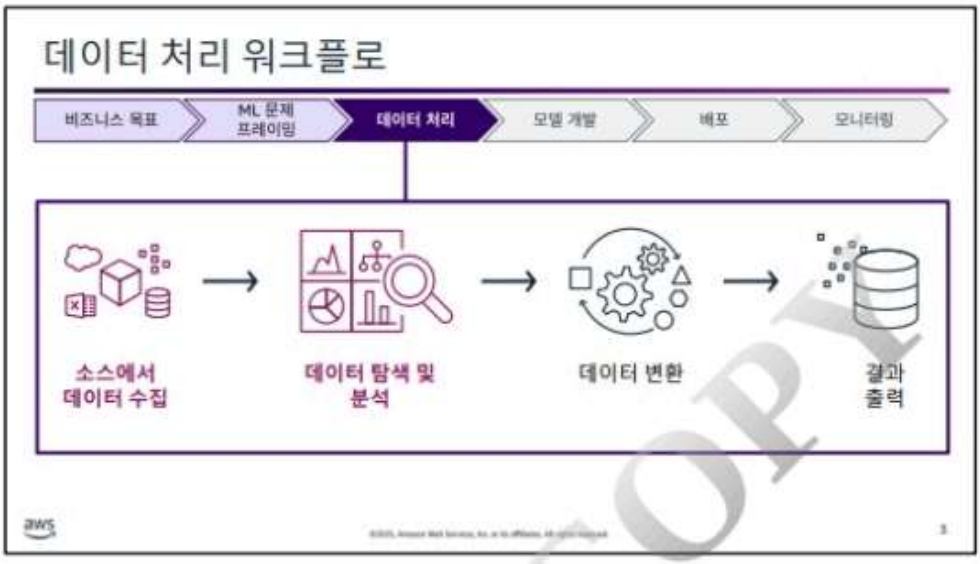
조직의 규모와 구조에 따라 데이터 과학자, 데이터 엔지니어, ML 엔지니어 등의 여러 역할이 데이터 처리 업무를 분담할 수 있습니다.

철저한 데이터 처리에 시간과 노력을 투자하면 더 정확하고 신뢰도가 높은 기계 학습 모델을 위한 기반을 마련할 수 있습니다. 이 단계는 모델 개발, 배포, 모니터링 등 ML 수명 주기의 후속 단계 효율성에 직접적인 영향을 줍니다.

DO NOT COPY
ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



ML 워크플로의 데이터 처리 단계에서는 다음이 포함됩니다:

1. **소스에서 데이터 수집**: 원시 데이터를 수집 및 저장합니다.
2. **데이터 탐색 및 분석**: 시각화 도구를 사용하여 인사이트를 파악하고 변환된 데이터를 검증합니다.
3. **데이터 변환**: 데이터 클리닝 및 피처 엔지니어링 단계를 지정합니다. 전체 데이터세트에 변환을 적용하기 전에 데이터 샘플에서 변환을 테스트합니다.
4. **결과 출력** : 처리된 데이터를 선호하는 스토리지 솔루션으로 내보냅니다.

이 모듈에서는 데이터 수집 및 탐색적 데이터 분석에 대해 살펴봅니다. 다음 모듈에서는 데이터 클리닝, 피처 엔지니어링, 특성 선택을 포함한 데이터 변환에 대해서 더 자세히 살펴봅니다.

DO NOT COPY
ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

이 모듈은 기계 학습을 위한 데이터 처리에 관한 포괄적인 이해를 위해서 3개 주요 주제로 구성되어 있습니다:

주제 A: 데이터 준비 및 유형

다양한 데이터 유형을 인식하고 효과적인 데이터와 그렇지 않은 데이터를 구별하는 것을 포함하여, 데이터 준비와 관련된 기초적인 개념을 살펴봅니다. ML 모델이 관련성이 높은 고품질 정보를 동작할 수 있도록 하려면 이러한 지식이 꼭 필요합니다.

주제 B: 탐색적인 데이터 분석

데이터를 시각화하고 분석하여 유용한 인사이트를 파악하는 기법에 대해 알아봅니다. 이 섹션에서는 모델 훈련을 시작하기 전에 데이터세트의 패턴, 이상치 및 잠재적인 문제를 파악하는 방법에 대해 다룹니다.

주제 C: AWS 스토리지 옵션 및 스토리지 선택

ML 워크플로에서 일반적으로 사용되는 Amazon Simple Storage Service(Amazon S3) 등의 핵심 AWS 스토리지 서비스를 살펴봅니다. 그리고 비용, 성능, 데이터 구조 등의 요인을 고려하여 구체적인 ML 요구 사항에 따라 가장 적합한 스토리지 솔루션을 선택하는 방법을 알아봅니다.

이러한 주제의 내용을 다루면, 데이터 처리 기술과 AWS 기반 기계 학습 모델 모범 사례의 견고한 기반을 다질 수 있습니다.

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

데이터 준비는 기계 학습 프로세스의 첫 단계입니다. 이 단계에서는 ML 모델 훈련용으로 대표적이면서도 유용한 고품질 데이터를 수집하는 걸 중점적으로 진행합니다. 이처럼 세심하게 큐레이팅된 데이터는 모델의 기반이 되어, 모델이 정확한 예측을 하거나, 정보를 분류하거나, 결정을 내리는 데 도움이 됩니다.

데이터의 품질과 특성은 알고리즘 선택 및 모델 성능에 큰 영향을 줍니다. 아래는 예시입니다:

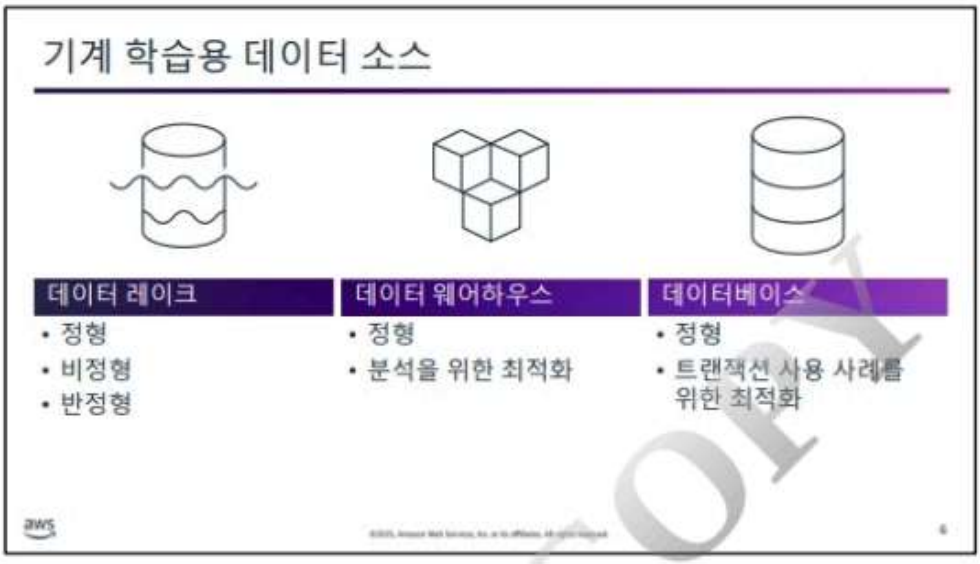
- **수치 데이터:** 선형 회귀 모델에 적합합니다.
- **이미지 데이터:** 특정 신경망 유형에서 더 잘 처리될 수 있습니다.
- **시계열 데이터:** 예측에 특수 알고리즘이 필요합니다.

데이터 전처리, 피처 엔지니어링 및 기타 변환 기법이 모델의 최적 성능에 중요한 역할을 합니다. 철저한 데이터 준비에 시간을 투자하면, 정확도와 신뢰도가 더욱 높은 기계 학습 결과를 생성하기 위한 기반을 마련할 수 있습니다.

ML 프로젝트의 성공은 훈련 데이터의 품질과 관련성에 따라 크게 좌우된다고 해도 과언이 아닙니다. 데이터 수집 및 사전 준비를 신중하게 고려하면 알고리즘 훈련 및 배포 시에 큰 이점을 얻을 수 있습니다.

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



ML 모델을 구축할 때는 관련성이 높은 고품질 데이터를 충분히 수집해야 합니다. 아래는 기계 학습에서 흔히 사용되는 데이터 소스입니다.

데이터 레이크: 여러 형식의 방대한 원시 데이터를 저장하는 중앙 리포지토리로 사용됩니다.
- 정형 데이터 (예: 쉼표로 구분된 값(CSV) 파일)
- 비정형 데이터 (예: 이미지 및 텍스트 문서)
- 반정형된 데이터 (예: JSON 및 XML)

데이터 레이크는 유연성을 제공하지만, ML에 사용하려면 데이터 전처리 작업을 많이 수행해야 합니다.

데이터 웨어하우스: 분석용으로 최적화된 정형 데이터가 저장됩니다. 데이터 웨어하우스의 특징은 다음과 같습니다.
- 잘 정의된 스키마 단위로 구성됩니다.
- 복잡한 쿼리 및 집계 시에 효율적입니다.
- 성능 개선을 위해 열 기반 스토리지를 자주 사용합니다.

데이터베이스: 트랜잭션 작업용으로 최적화된 정형 스토리지를 제공합니다. 데이터베이스의 특성은 다음과 같습니다.
- 관계가 정의된 테이블로 구성됩니다.
- 잦은 업데이트 및 검색 시에 유용합니다.
- 각기 다른 데이터 모델에 적합한 여러 유형(예: 관계형, NoSQL)으로 구성됩니다.

ML용 데이터를 수집할 때 고려해야 하는 사항은 다음과 같습니다.
- 모든 관련 데이터 소스를 파악합니다.
- 중앙 스토리지 위치(예: Amazon S3 데이터 레이크)를 선택합니다.
- 적절한 수집 방법을 선택합니다.

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

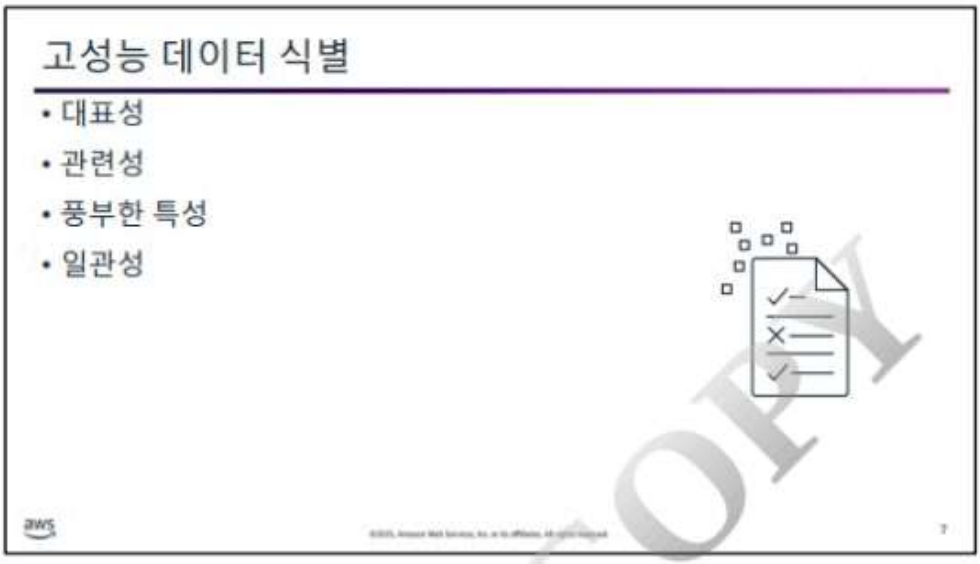
Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

- 데이터를 정리 및 준비하고 보호합니다.

DO NOT COPY
ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



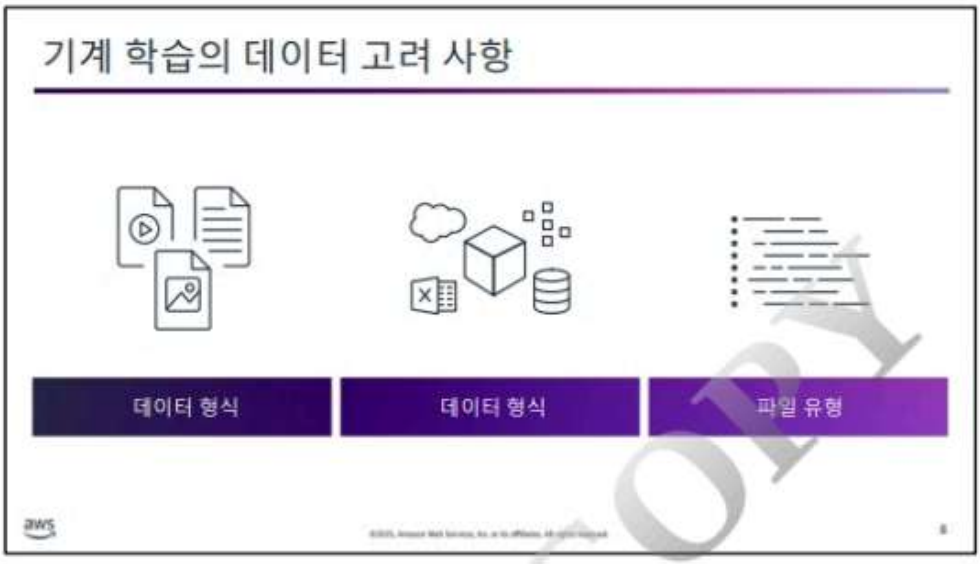
데이터 품질을 미리 평가하면 문제에 가장 적합한 데이터세트를 만드는데 도움이 됩니다. 고성능 데이터는 대표적이고, 관련성이 높으며, 특성이 풍부하고, 일관적입니다.

- 대표성: 모델의 의미 있는 예측 능력을 향상하려면 훈련 데이터가 실제 시나리오를 반영해야 합니다. 예를 들어 고객 중 20%가 보통 1년 후에 멤버십을 취소한다면, 데이터가 해당 이탈률을 정확하게 나타내야 합니다.
- 관련성: 예측 목표와 관련 있는 패턴을 보이는 속성이 포함되어야 합니다. 고객 이탈 예측의 경우 관련 데이터에는 멤버십 기간, 사용 빈도, 고객 지원 상호 작용이 포함될 수 있습니다. 관련성이 없는 정보는 모델 성능에 부정적인 영향을 끼칠 수 있습니다.
- 풍부한 특성: 특성의 포괄적인 세트는 모델이 기본 패턴을 학습하는데 도움이 됩니다. 다양하고, 관련된 속성이 포함되어 있으면 의미 있는 상관 관계 파악 가능성이 높아집니다. 예를 들어 인구 통계 정보를 통해 고객 행동의 중요한 추세를 확인할 수 있습니다.
- 일관성: 데이터세트 전반에서 일관된 형식, 단위 및 메타데이터를 사용해야 합니다. 일관성이 낮으면 알고리즘을 혼동하고 정확도를 떨어뜨릴 수 있습니다. 여러 소스의 데이터가 표준 형식을 준수하는지 확인해야 합니다.

입력 데이터 품질은 ML 모델의 출력 품질에 직접적인 영향을 주는 걸 기억하십시오. 고품질의 대표 데이터를 수집하고 준비하는데 시간을 충분히 투자하면 더 정확하고 신뢰할 수 있는 결과를 얻을 수 있습니다.

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



기계 학습용 데이터를 준비할 때는 데이터의 다양한 측면을 고려하는 것이 필수입니다.

데이터 유형은 알고리즘 선택 및 모델 성능에 큰 영향을 미칩니다. 일반적인 유형은 다음과 같습니다.

- 텍스트 데이터(예: 문서 및 소셜 미디어 게시물)
- 테이블 형식 데이터(예: 스프레드시트 및 데이터베이스)
- 시계열 데이터(예: 주가 및 센서 판독값)
- 이미지 데이터(예: 사진 및 의료 영상)

다양한 ML 알고리즘은 특정 데이터 유형에 최적으로 작동하도록 설계가 되어 있습니다. 예를 들어, 합성곱 신경망은 이미지 처리 태스크를 효율적으로 처리하는 반면, 순환 신경망은 텍스트나 시계열 등의 순차적 데이터에 적합합니다.

데이터 형식은 ML 알고리즘이 데이터를 처리하고 데이터를 통해 학습을 하는 방식에 영향을 줍니다. 일반적인 데이터 형식은 다음과 같습니다.

- 행 기반
- 열 기반
- 객체 표기법

형식의 선택은 스토리지 효율성, 처리 속도, 분석 용이성에 영향을 미칠 수 있습니다. 예를 들어 Parquet 등의 열 기반 형식은 대규모 데이터세트의 분석 쿼리에 더 효율적인 경우가 많습니다.

파일 유형에 따라 데이터 저장 및 액세스 방식이 결정됩니다. 파일 유형의 예는 다음과 같습니다.

- CSV

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

- Parquet (열 기반 스토리지 형식)
- Avro (스키마로 구성된 행 기반 형식)
- JSON 및 JavaScript Object Notation Lines (JSONL)

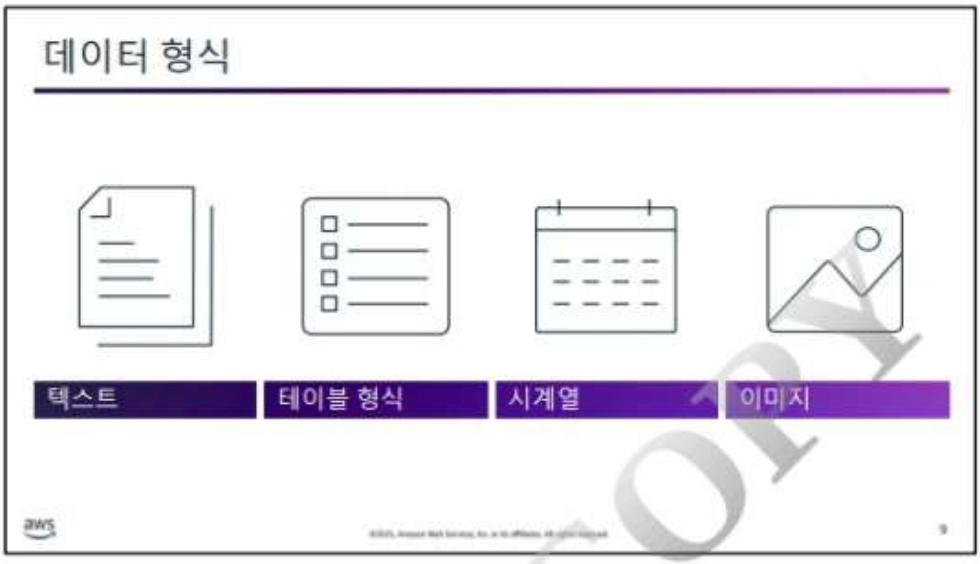
적절한 파일 유형을 선택하는 것은 데이터 크기, 쿼리 패턴, ML 도구 및 인프라와의 호환성과 같은 요인에 따라 달라집니다.

이와 같은 데이터 측면을 면밀하게 고려하면 ML 워크플로를 최적화하여 성능과 효율성을 높일 수 있습니다.

DO NOT COPY
ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



기계 학습에서 흔히 사용되는 4가지 데이터 유형은 다음과 같습니다.

텍스트 데이터는 문서, 기사, 소셜 미디어 게시물 및 기타 서면 콘텐츠가 포함됩니다. ML에서 텍스트는 보통 처리를 위해서 숫자 표현(예: 단어 임베딩)으로 변환됩니다. 모델은 이 숫자로 표현된 텍스트를 사용하여 데이터를 분석합니다. 감정 분석 또는 텍스트 분류와 같은 태스크를 위해 텍스트 데이터에 자연어 처리(NLP) 기법을 자주 적용됩니다.

테이블 형식 데이터는 스프레드시트 또는 관계형 데이터베이스처럼 행과 열로 구성됩니다. 각 행은 인스턴스를 나타내고 열은 특성이나 속성을 나타냅니다. 테이블 형식 데이터는 고객 이탈 예측부터 사기 탐지 등의 많은 ML 애플리케이션에서 일반적으로 사용됩니다. 테이블 형식 데이터는 선형 회귀 모델에 적합합니다.

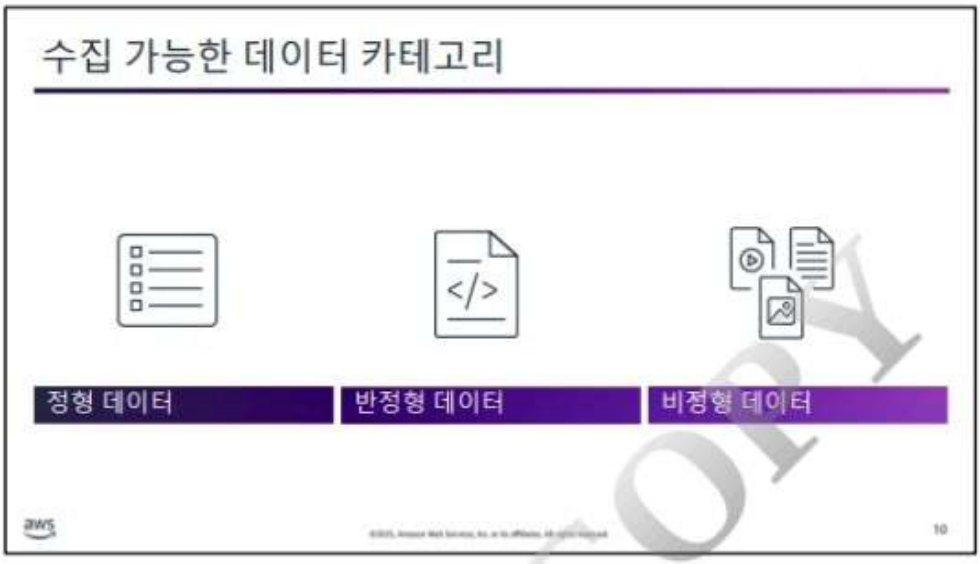
시계열 데이터는 일정 간격으로 시간 경과에 따라 수집된 데이터 포인트로 구성됩니다. 주가, 날씨 측정, 센서 측정값 등을 예로 들 수 있습니다. 시계열 데이터 사용 시에는 시간적 패턴과 추세를 포착하기 위해서 특수 전처리 및 모델링 기술을 사용해야 하는 경우가 많습니다.

이미지 데이터는 사진, 의료 스캔, 위성 이미지 및 기타 시각 정보 등이 포함됩니다. ML에서 이미지는 보통 다차원 픽셀 값 배열로 표현됩니다. 이러한 데이터 유형은 기계 학습에서 객체 인식, 자율 주행, 이미지 분류 등에 자주 사용됩니다.

작업 중인 데이터 유형을 파악하면 적절한 ML 알고리즘과 피처 엔지니어링 기법 및 평가 지표를 선택하여 효율적인 모델을 구축할 수 있습니다.

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



수집 가능한 데이터의 3가지 주요 카테고리는 다음과 같습니다.

정형 데이터는 스키마가 잘 정의되어 있으므로 알고리즘이 쉽게 해석할 수 있는 데이터입니다. 정형의 특성은 다음과 같습니다.

- 행과 열이 포함된 테이블로 구성되어 있습니다.
- 각 필드에 일관된 데이터 유형이 포함되어 있습니다.
- 데이터 요소 간의 명확한 관계가 제공됩니다.

정형 데이터의 예로는 관계형 데이터베이스, CSV 파일 등이 있습니다. 정형 데이터는 보통 처리하고 분석하기는 쉽지만 복잡한 실제 정보를 표현할 수 있는 유연성이 부족할 수 있습니다.

반정형 데이터는 정형 데이터와 비정형 데이터의 요소를 결합합니다. 조직적인 속성이 어느 정도 있지만 엄격한 스키마를 따르지는 않습니다. 아래는 반정형 데이터의 예제입니다:

- JSON 및 XML 파일
- 이메일 메시지(정형의 헤더 및 비정형의 본문 포함)
- 일부 일정한 필드를 가진 로그 파일

반정형 데이터는 유연성과 구조 사이의 균형을 제공하므로 많은 ML 애플리케이션에 적합합니다.

비정형 데이터는 사전 정의된 스키마 또는 구조가 없는 데이터입니다. 비정형 데이터의

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

예는 다음과 같습니다.
- 자유 형식 텍스트 문서
- 이미지 및 비디오
- 오디오 녹음

비정형 데이터는 풍부하지만, ML 모델에 유의미한 특성을 추출하려면 전처리를 많이 수행해야 합니다. 하지만 제대로 분석하면 가치 있는 인사이트로 이어질 수 있는 풍부한 정보가 종종 포함되어 있습니다.

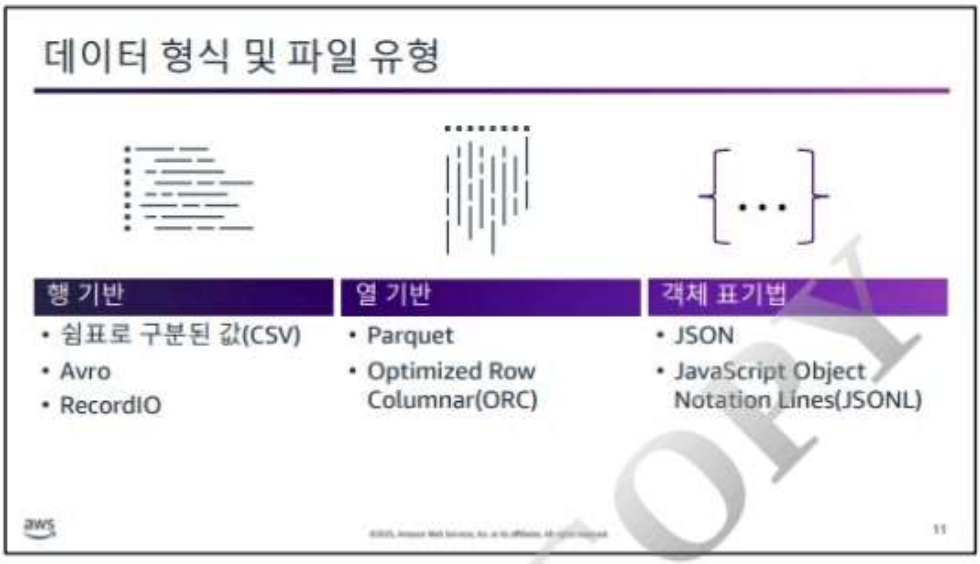
데이터 형식을 선택할 시 다음 요소를 고려해야 합니다.
- 처리 및 분석 용이성
- 스토리지 효율성
- ML 도구 및 인프라와의 호환성
- 데이터 내의 복잡한 관계 표현하는 능력

적절한 데이터 형식을 선택하면 ML 워크플로를 최적화하여 성능과 효율성을 높일 수 있습니다.

DO NOT COPY
ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



행 기반, 열 기반, 객체 표기법과 같은 일반적인 데이터 형식을 사용하여 기계 학습에서 데이터를 구조화할 수 있습니다. 형식은 분석의 효율성에 영향을 미치며 필요한 데이터 구조 및 작업과 일치해야 합니다.

CSV, Avro, RecordIO 등의 행 **기반 형식**은 행에서 자주 액세스하는 밀집된 테이블 형식 데이터에 적합합니다.

Parquet, Optimized Row Columnar(ORC) 등의 열 **기반 형식**은 열별로 집계를 수행하는 희소한 데이터에 더 적합합니다. 분석 워크로드에서 압축 및 성능의 이점을 제공합니다.

JSON, JSONL 등의 **객체 표기 형식**은 테이블 형식이 아닌 계층적 데이터 구조에 적합합니다. 이는 유연성을 제공하지만 전처리를 많이 해야 할 수도 있습니다.

DO NOT COPY
ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

행 기반 데이터 형식은 각 행이 단일 레코드나 엔티티를 나타내고 열이 해당 엔티티의 각 특성을 나타내도록 데이터를 구성합니다. 이 형식은 관계형 데이터베이스 및 스프레드시트에서 일반적입니다. 이는 특성 간의 관계를 보여줍니다. 이 형식이 디스크에 데이터를 쓰는 가장 빠른 방식이지만, 관련 없는 데이터는 건너 뛰어야 하므로 반드시 가장 빠른 읽기 옵션은 아닙니다.

예를 들어, 데이터 분석을 통해 구독 상태 및 고객 연령과 같은 열 특성 간의 관계를 확인할 수 있습니다.

DO NOT COPY
ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

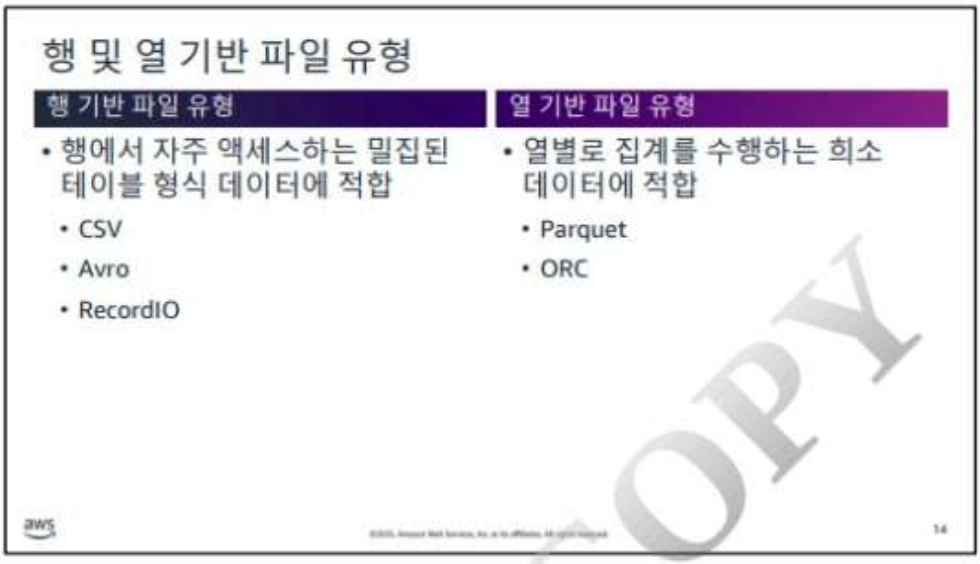
열 기반 데이터 형식은 열을 기본 구조로 하는 정보를 저장합니다. 이 형식에서는 쿼리가 전체 레코드가 아닌 열 내의 패턴에서 인사이트를 추출하므로 대규모 데이터세트의 추세를 효율적으로 분석할 수 있습니다. 또한 관심이 없는 열은 건너뛸 수 있으므로 일반적으로 읽기 성능도 개선됩니다.

예를 들어 구독 상태와 마지막 고객 서비스 문의 이후 경과한 시간이라는 2가지 특성을 사용하여 고객 이탈을 분석할 수 있습니다. 전체 레코드를 로드하지 않고도 많은 레코드에서 이탈을 분석할 수 있습니다. 행 기반 데이터가 여전히 일반적이지만 열 기반 데이터는 압축 기능을 제공하고 ML 워크로드의 관련 데이터에 대한 액세스 속도가 높습니다.

DO NOT COPY
ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



행 기반 파일 유형은 행에서 자주 액세스하는 밀집된 테이블 형식 데이터에 적합합니다.
- CSV 파일은 열 값이 쉼표로 구분되어 있는 테이블 형식 데이터에 해당하는 경량 텍스트 파일입니다. 이 형식은 단순하지만 다른 형식에 비해 덜 효율적입니다.
- Avro 및 RecordO는 스키마를 사용하여 레코드를 순차적으로 저장하므로 전체 데이터세트를 반복하는 ML 워크로드에서 도움이 됩니다.

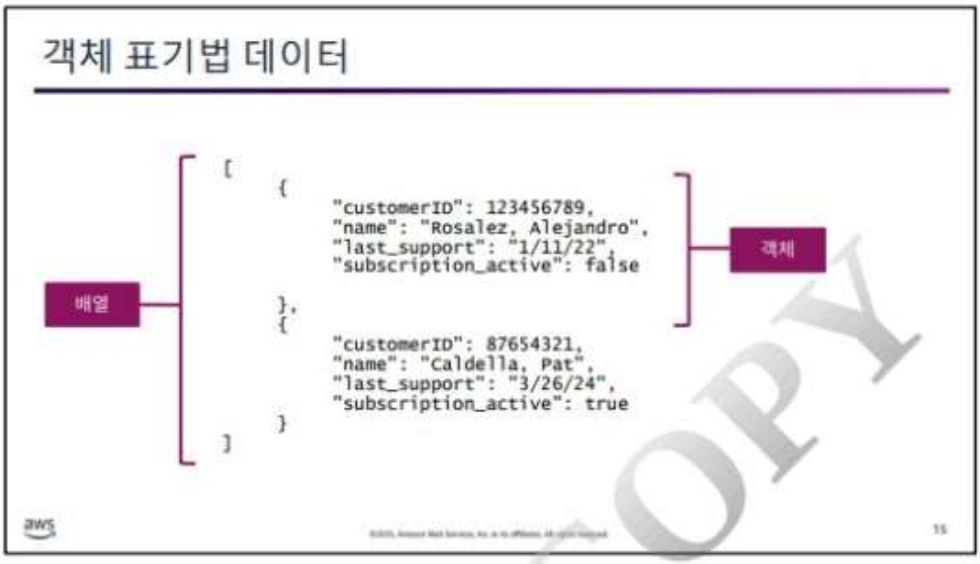
열 기반 파일 유형은 열 단위로 집계하는 스마스 데이터에 더 적합합니다.
- Parquet은 일반적으로 분석과 데이터 웨어하우징에 사용되는 열 기반 형식으로, 압축 및 성능 이점을 제공합니다.
- ORC는 Apache Hive, Spark 같은 빅 데이터 워크로드에 사용되는 열 기반 형식인데, 압축 및 성능 개선 이점이 있습니다.

이 선택은 ML 워크로드의 데이터 구조, 액세스 패턴 및 성능 관련 요구에 따라 달라집니다.

DO NOT COPY ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



객체 표기법은 그래프 또는 텍스트 데이터와 같이 테이블 형식이 아닌 계층 데이터에 적합합니다. 객체 표기법 데이터는 행과 열이 아닌 특성과 키-값 페어가 있는 계층적 객체로 구조화됩니다.

코드 블록에는 객체 표기법 형식의 데이터의 예가 나와 있습니다. 배열에 객체 두 개가 포함되어 있는데, 각 객체에는 고객 ID, 이름, 마지막 지원 연락 날짜, 구독 상태 등의 속성에 해당하는 키-값 페어를 가진 고객을 나타냅니다.

객체 표기법은 계층적이고 중첩된 특성 때문에 대다수 ML 워크로드에 적합한 유연한 데이터 구조를 제공합니다.

DO NOT COPY ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

객체 표기법 파일 유형

- JSON
  - JSON은 데이터 구조가 유동적입니다.
  - 데이터는 객체와 배열로 표시됩니다.
  - JSON에는 중첩된 계층적 객체가 포함될 수 있습니다.
- JSONL
  - 효율성이 더 높습니다.
  - 객체를 한 줄에 하나씩 입력합니다.

```
{"customerID": "12345678","last_support": "1/12/22","subscription_active": "false"}
{"customerID": "87654321","last_support": "3/26/24","subscription_active": "true"}
```

16

객체 표기법 파일 유형은 객체와 유사한 구조를 사용하여 데이터를 인코딩하는 방법입니다.

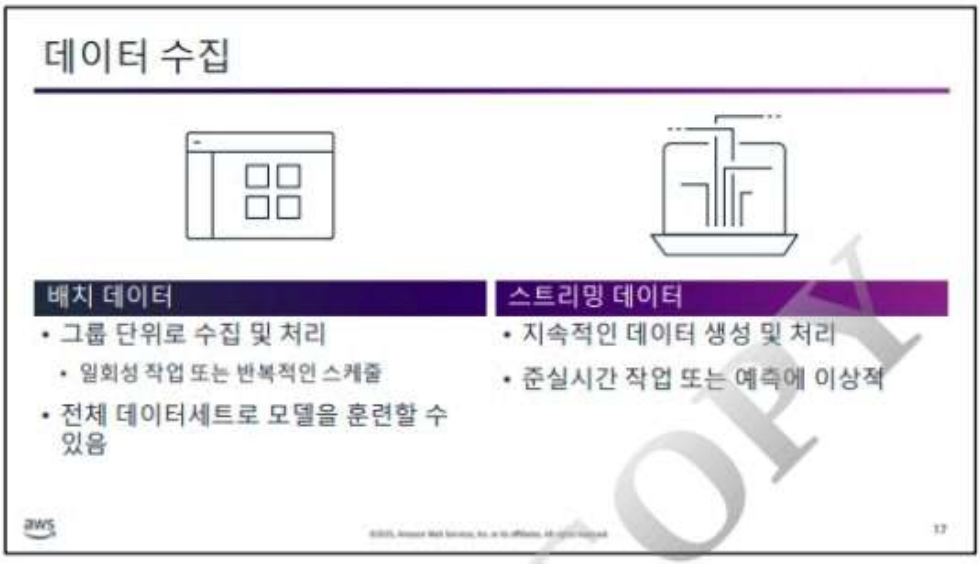
JSON은 사람과 기계 모두가 읽을 수 있는 문서 기반 데이터 형식입니다. ML 모델은 데이터 구조가 유연하기 때문에 JSON에서 학습할 수 있습니다. 데이터가 간결하고 계층적이며 파싱하기 쉬워서 많은 ML 워크로드에 적합합니다. JSON은 객체와 배열로 표현됩니다. 객체는 키-값 페어로 정의되고 중괄호 {}로 묶인 데이터입니다. 데이터는 문자열, 숫자, 부울, 배열, 객체 또는 null일 수 있습니다. 배열은 대괄호 []로 묶인 값 모음이며 쉼표로 구분된 값을 포함할 수 있습니다.

JSONL은 줄바꿈으로 구분된 JSON이라고도 합니다. 중첩되는 대신 새 줄로 구분되는 JSON 객체를 인코딩하는 형식입니다. 각 JSON 객체는 한 줄에 하나씩 작성됩니다. JSONL은 더 큰 JSON 배열을 로드하지 않고도 개별 객체를 처리할 수 있기 때문에 효율성이 향상됩니다. 이렇게 객체 파싱의 효율성이 향상되어 ML 워크로드에 대한 대규모 데이터 세트를 더 잘 처리할 수 있습니다. 또한 JSONL 구조는 Parquet 같은 열 형식에 매핑할 수 있어 해당 파일 유형의 추가 이점을 제공합니다.

JSON과 JSONL 사용 시에는 다양한 정보를 유동적으로 캡처할 수 있습니다. 그러나 JSONL은 ML 워크로드에서 대규모 데이터세트를 처리할 때 추가적인 이점을 제공합니다.

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



기계 학습용 데이터 수집은 보통 배치 처리나 스트리밍의 2가지 방식 중 하나를 사용합니다.

배치 데이터는 일회성 작업에서 또는 반복 일정에 따라 그룹별로 수집 및 처리합니다. 이 접근법은 전체 데이터세트를 미리 사용할 수 있는 기존 ML 태스크에 일반적으로 사용됩니다. 이 접근법은 실시간 처리가 필요하지 않은 과거 데이터세트에서 ML 모델을 훈련할 때 적합합니다. 예를 들어 레이블이 지정된 이미지의 대규모 데이터 세트를 대상으로 이미지 분류 모델을 훈련할 때 배치 처리를 사용할 수 있습니다. 배치 데이터는 보통 시간이 중요하지 않으며 장기 분석에 사용됩니다.

반면, **스트리밍 데이터**에는 거의 실시간으로 데이터를 지속적으로 생성하고 처리하는 작업이 포함됩니다. 이 접근 방식은 토대로 즉시 작업을 수행하거나 데이터를 기반으로 예측을 해야 하는 애플리케이션에 이상적입니다. 일반적인 사용 사례로는 사기 탐지가 있습니다. 거래를 즉시 분석하여 잠재적 사기 활동을 식별해야 합니다. 스트리밍 데이터는 배치로 처리할 수 있지만, 보통 데이터 버퍼링, 실시간 모니터링 그리고 지속적인 흐름을 처리하기 위한 추가 내결함성 및 확장성과 같은 사항을 추가로 고려해야 합니다.

배치 처리와 스트리밍 중 선택은 구체적인 사용 사례와 요구 사항에 따라 달라집니다. 실시간으로 분석하지 않아도 되는 대규모 데이터세트를 사용할 때는 배치 처리가 구현하기가 더 간단하며 비용 효율성도 더 높습니다. 스트리밍에서는 최신 인사이트를 제공하지만 지속적인 데이터 흐름을 처리하는 더욱 복잡한 인프라가 필요할 수도 있습니다. 둘 중에서 사용할 방식을 결정할 때는 데이터의 볼륨, 대기 시간 요구 사항, 처리 작업의 복잡도와 같은 요인을 고려해야 합니다.

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

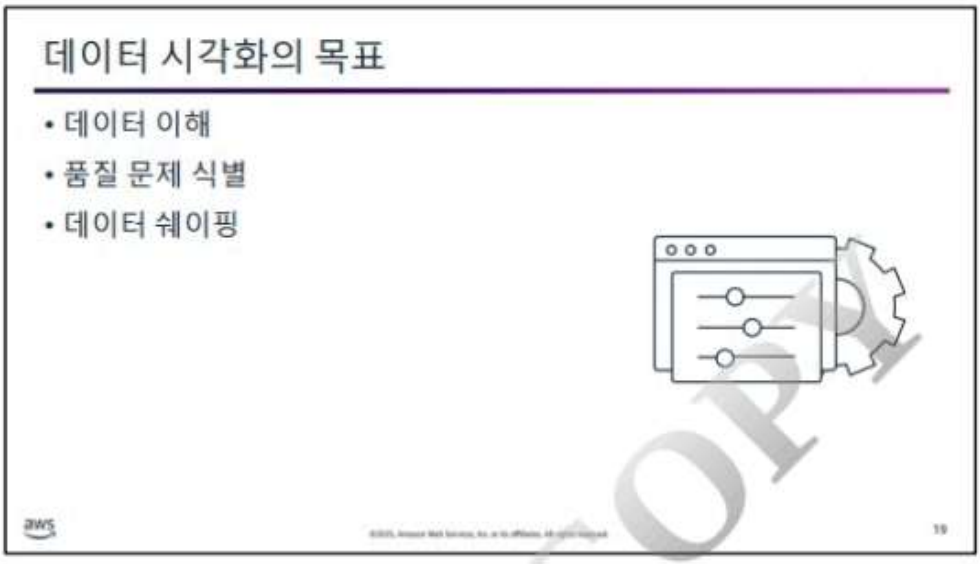
Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

기계 학습 워크플로의 탐색적 데이터 분석 단계는 원시 데이터에서 실행 가능한 인사이트 사이의 가교 역할을 합니다. 데이터 시각화는 ML 모델을 훈련하기 전에 데이터의 품질 및 관련성에 대한 인사이트를 수집하는 데 도움이 됩니다. 데이터를 이해하면 편차, 왜도, 이상치와 같은 문제를 식별할 수 있습니다. 이러한 문제를 파악하면 모델 정확도가 저하될 가능성이 줄어듭니다.

DO NOT COPY
ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



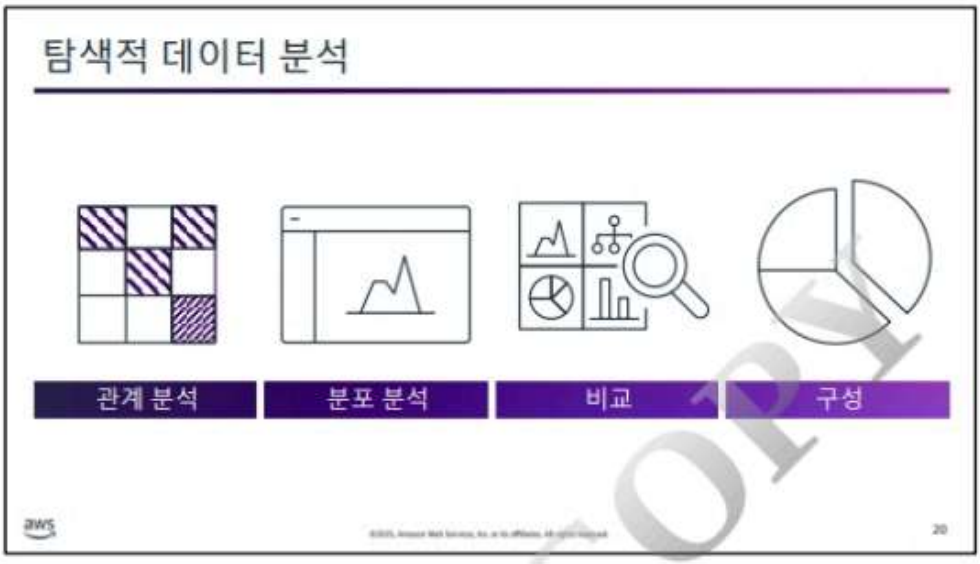
기계 학습의 데이터 시각화는 3가지 주요 목표에 도움이 됩니다. 각 목표는 더 정확하고 신뢰할 수 있는 모델 개발에 기여합니다.

- 데이터 이해: 데이터를 시각화하면 데이터세트의 전반적인 구조와 특성을 파악할 수 있습니다. 데이터 이해란 데이터의 패턴, 추세, 변수 간의 관계 등을 파악하는 것입니다. 가령 상관관계 행렬 열 지도를 생성하면 대상 변수와 가장 관련성이 높은 특성을 빠르게 확인할 수 있으므로 적절한 특성을 선택할 수 있습니다.
- 품질 문제 식별: 시각 검사를 통해서 테이블 형식에서는 놓칠 수도 있는 데이터 품질 문제를 확인할 수 있는 경우 종종 있습니다. 품질 문제 식별이란 데이터의 이상치, 누락된 값, 불일치를 감지하는 것이 포함됩니다. 예를 들어 박스 플롯은 모델 훈련 중에 해결해야 할 수 있는 이상치를 명확하게 파악할 수 있습니다.
- 데이터 쉐이핑: 시각화에서 파악한 인사이트에 따라 데이터 전처리 및 변환 방법에 대해 정보에 입각하여 결정할 수 있습니다. 데이터 쉐이핑에는 피처 엔지니어링, 스케일링, 범주형 변수 인코딩이 포함될 수 있습니다. 모델링할 데이터를 최적화할 때 유용합니다. 예를 들어 히스토그램에서 수치적 특성의 분포가 심하게 치우쳐 보인다면, 로그 변환을 적용하여 해당 특성을 정규화 할 수 있습니다.

이러한 목표에 집중하면, 기계 학습 알고리즘에 데이터를 입력하기 전에 데이터를 잘 이해하고 적절하게 준비할 수 있습니다. 이러한 기초 작업은 종종 더 강력하게 정확한 모델과 해결하고자 하는 문제에 대한 더 깊은 이해로 이어집니다.

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



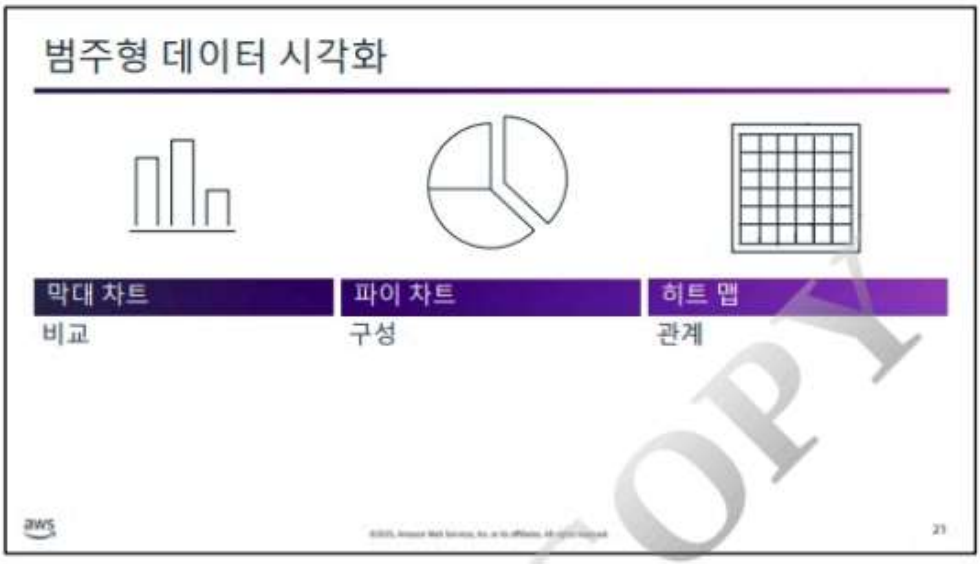
탐색적 데이터 분석(EDA)에서는 데이터세트 관련 인사이트를 파악하기 위해서 다양한 시각화 기법을 사용합니다. 다음은 4가지 주요 분석 방법입니다:

- 관계 분석: 이 방법은 변수 간의 상관 관계와 종속성을 파악하는데 초점을 맞춥니다. 이러한 작업에 흔히 사용되는 도구로는 산점도와 상관 행렬이 있습니다. 예를 들어 주택 크기와 가격 간의 명확한 양의 상관 관계가 산점도에 표시된다면 주택 가격 예측 모델의 특성 선택에 도움이 될 수 있습니다.
- 분포 분석: 이 기법은 개별 변수의 데이터 분산과 형태를 조사합니다. 분포를 파악할 때는 히스토그램과 박스 플롯이 유용합니다. 이러한 시각화를 사용하면 데이터의 왜도(비대칭도), 이상치 및 중심 경향을 파악하는데 도움이 되며, 정규화, 이상치 제거 등의 전처리 결정에 영향을 줄 수 있습니다.
- 비교: 이 방법은 데이터의 하위 그룹 간 차이를 수치로 환산하여 시각화하는 작업이 포함됩니다. 막대형 차트와 그룹 박스 플롯은 비교에 효과적입니다. 예를 들어, 그룹형 막대 차트를 사용하여 여러 제품 카테고리의 고객 이탈률을 비교할 수 있습니다.
- 구성: 이 접근법은 각 파트가 전체에 어떻게 기여하는지 이해하는 데 초점을 맞춥니다. 파이 차트와 누적 막대 차트가 구성 분석에 흔히 사용합니다. 분류 문제나 시장 점유율 분석에서 클래스 분포를 시각화하려는 경우 이러한 시각화가 유용할 수 있습니다.

각 방법은 데이터에 대해 다른 관점을 제공하며, 조합해서 사용하면 데이터세트를 포괄적으로 파악할 수 있습니다. 데이터 시각화에 대한 이런 다각도 접근 방식은 데이터 전처리에서 모델 선택과 평가에 이르기까지 기계 학습 워크플로 전체에서 정보에 입각하여 결정을 내릴 수 있습니다.

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



범주형 데이터는 그룹이나 범주로 구분할 수 있는 정성적 정보에 해당됩니다. 이 데이터에는 성별, 제품 유형, 고객 상태 같은 변수가 포함됩니다. 범주형 데이터 시각화는 정성적 변수 내의 분포와 관계를 이해하는 데 중요합니다.

범주형 데이터에 대한 통계 분석에는 빈도, 비율 및 분할표 계산이 포함됩니다. 범주형 변수에는 고유한 순서가 없는 가능한 값의 수가 제한되어 있습니다.

다음은 일반적인 3가지 기법입니다.

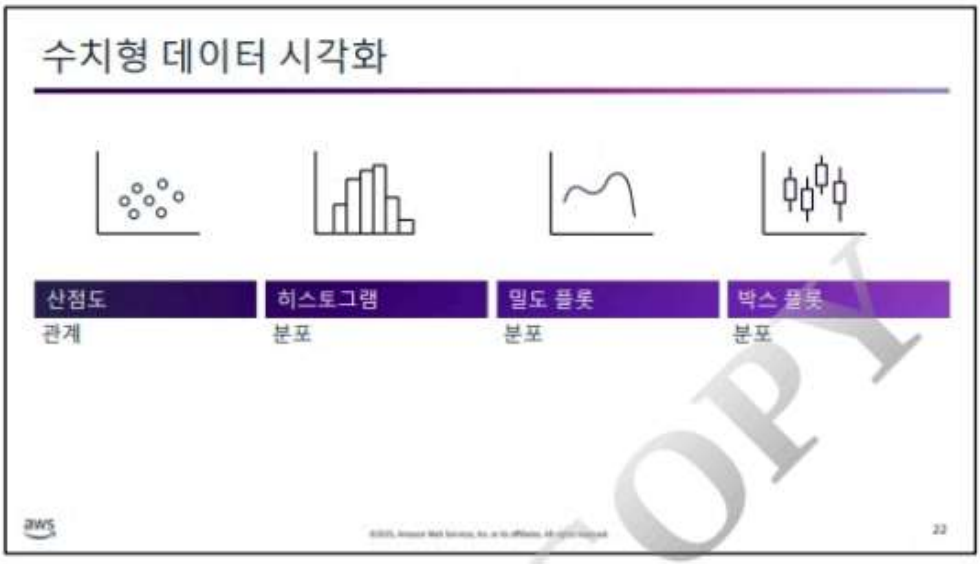
- **막대 차트**는 여러 카테고리에서 빈도나 비율을 비교할 때 매우 좋습니다. 카테고리가 많은 경우 또는 시간에 따른 변화를 보여주고 싶을 때 특히 유용합니다. 예를 들어, 고객 이탈 예측 모델에서 막대 차트를 사용해 다양한 제품군이나 고객 세그먼트의 이탈률을 시각화할 수 있습니다.
- **파이 차트**는 카테고리 수가 적을 때 (이상적으로는 5~7개 이하) 전체의 구성을 보여주는 데 가장 적합합니다. 시장 점유율 또는 분류 문제의 클래스 분포 등을 표시하려는 경우 더 효과적일 수 있습니다. 하지만 파이 차트 사용 시 비율을 정확하게 비교하기가 어려운 경우도 있습니다.
- **히트 맵**은 2개 이상의 범주형 변수 간의 관계를 시각화할 수 있는 강력한 도구입니다. 이는 색상의 농도로 빈도나 기타 지표를 표현합니다. 기계 학습에서 히트 맵은 상관 행렬을 시각화하는 데 자주 사용되며, 서로 다른 특성들이 얼마나 밀접하게 연관되어 있는지 보여줍니다.

범주형 데이터로 작업할 때는, 카테고리 수와 파악하려는 구체적인 인사이트를 고려해야 합니다. 막대 차트는 여러 용도로 활용 가능하며 많은 카테고리를 처리할 수 있습니다. 반면, 파이 차트는 카테고리 수가 적으며 전체와 부분 간의 관계를 명확하게 보여줍니다. 히트 맵에서는 여러 변수 간의 복잡한 관계를 쉽게 표시할 수 있습니다.

이러한 시각화의 목표는 특성 선택에서 모델 해석까지의 전체 기계 학습 프로세스에 정보를 제공할 수 있는 인사이트를 발견하기 위해서입니다.

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



반면, **수치형 데이터**는 숫자로 표현할 수 있는 정량적 측정값을 나타냅니다. 이러한 데이터로는 연령, 소득, 기온 등의 변수가 포함됩니다. 수치형 데이터는 고급 수학 분석에 유용합니다.

수치형 데이터를 시각화 하는 것은 정량적 변수 내의 분포, 관계 및 패턴을 이해하는 데 중요합니다. 수치형 데이터의 시각화에는 평균, 중앙값, 표준 편차와 같은 통계적 방법이 사용됩니다.

4가지 주요 기법은 다음과 같습니다:

- **산점도**는 두 수치형 변수 간의 관계를 시각화하는 데 사용됩니다. 각 점은 관찰 값을 나타내며, 그 위치는 두 변수의 값에 의해 결정됩니다. 산점도는 선형 관계, 클러스터, 이상치 등의 패턴을 나타낼 수 있습니다. 기계 학습에서 산점도는 특성 관계를 살펴보고 대상 변수에 잠재적 예측 변수를 식별하는 데 자주 사용됩니다.
- **히스토그램**는 값의 범위를 빈으로 나누고, 각 빈도의 관측치 빈도를 표시하여 단일 수치형 변수의 분포를 보여줍니다. 데이터 분포의 형태 (예: 정규, 왜곡, 이중모달) 를 이해하고 이상치일 수 있는 항목을 식별할 때는 히스토그램이 매우 유용합니다. 이 정보는 데이터 변환 또는 특성 스케일링에 대한 결정의 지침이 될 수 있습니다.
- **밀도 플롯**: 히스토그램과 비슷하지만 불연속형 빈이 아닌 평탄화된 분포 곡선을 보여줍니다. 같은 플롯에 있는 여러 변수나 그룹의 분포를 비교하는데 특히 유용합니다. 기계 학습에서 밀도 플롯을 활용하면 분류 문제의 클래스 간 특성 분포가 어떻게 다른지 이해하는 데 도움이 됩니다.
- **박스 플롯** (상자수염도라고도 함)은 중위수, 사분위수 및 이상치일 가능성이 있는 수치형 변수의 분포 요약을 제공합니다. 여러 그룹 또는 범주의 분포를 비교하려는 경우 박스 플롯이 효율적입니다. 기계 학습에서 박스 플롯은 클래스가 명확하게

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

구분되어 있는 특성을 식별하거나 특수 처리가 필요할 수 있는 이상치를 탐지하는 데 도움이 됩니다.

이 시각화 기법들은 수치형 데이터와 관련된 다양한 관점을 제공합니다. 산점도에서는 변수 간의 관계에 초점을 맞추고, 히스토그램과 밀도 플롯에서는 개별 변수 분포를 보여주고, 박스 플롯에서는 분포의 간결하게 요약해서 보여줍니다. 이러한 도구들을 조합해서 사용하면, 데이터 특성을 포괄적으로 파악할 수 있습니다. 이는 기계 학습 프로젝트에서 효과적인 특성 엔지니어링, 모델 선택 및 해석에 중요합니다.

DO NOT COPY
ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

올바른 스토리지 솔루션을 선택하는 것은 AWS에서 기계 학습 워크플로를 설계할 때 중요한 결정입니다. 선택한 스토리지 옵션이 ML 프로젝트의 성능, 비용 효율성 및 확장성에 큰 영향을 미칠 수 있습니다.

기계 학습용 스토리지 옵션을 고려할 때는 몇가지 요인이 작용됩니다:

- 성능: ML 태스크별로 I/O 요구 사항이 다릅니다. 예를 들어 대규모 모델을 훈련하는 경우 처리량이 높은 스토리지가 필요할 수도 있고, 실시간 추론 시에는 지연 시간이 짧은 액세스를 더 우선 순위로 둘 수 있습니다.
- 확장성: 데이터세트가 커지고 ML 모델이 더 복잡해지면, 스토리지 솔루션도 성능 저하 없이 그에 따라 규모가 확장해야 합니다.
- 비용: 특히 대규모 데이터세트라면 스토리지 비용이 급격히 늘어날 수 있습니다. 성능 요구와 비용 효율성의 균형을 맞추는 것이 중요합니다.
- 데이터 액세스 패턴: 데이터에 액세스하는 방법에 (예를 들어, 순차적 액세스, 랜덤 액세스 비교) 따라 가장 적합한 스토리지 옵션에 영향을 미칠 수 있습니다.
- 통합: 스토리지 솔루션은 Amazon SageMaker AI 등 ML에 사용 중인 다른 AWS 서비스와 원활하게 통합되어야 합니다.

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

AWS에서는 ML 워크플로의 다양한 요구를 충족하는 여러 가지 스토리지 서비스를 제공됩니다. 다음 슬라이드에서 Amazon Simple Storage Service(S3), Amazon Elastic Block Store(Amazon EBS), Amazon Elastic File System(Amazon EFS), Amazon FSx 등의 주요 옵션을 소개합니다. 각 서비스는 강점이 있고 ML 파이프라인 내의 다양한 사용 사례에 활용할 수 있습니다.

- Amazon S3는 다양한 애플리케이션에 사용되는 유연하고 확장 가능한 객체 스토리지 서비스입니다. 이러한 애플리케이션에는 데이터 레이크, 웹 사이트, 클라우드 중심 애플리케이션, 백업, 아카이브, 분석, ML이 포함됩니다.
- Amazon EBS는 Amazon Elastic Compute Cloud(Amazon EC2) 인스턴스에 영구 블록 수준 스토리지 볼륨을 제공하므로 필요에 따라 스토리지 성능 및 비용을 확장할 수 있습니다.
- Amazon EFS는 확장성이 뛰어난 서버리스 클라우드 중심 파일 스토리지 서비스입니다. 스토리지 용량을 프로비저닝하거나 관리할 필요 없이 자동으로 페타바이트까지 확장되는 파일 시스템을 설정할 수 있습니다.
- Amazon FSx는 Lustre, NetApp ONTAP, OpenZFS, Windows File Server와 같은 인기 파일 시스템에 대한 액세스를 제공하는 완전관리형 서비스입니다.

이러한 스토리지 옵션 및 특성을 이해하면 특정 사용 사례에 맞게 최적의 성능과 비용 효율성을 위해 AWS에서 ML 인프라를 설계할 방법에 대해 정보에 입각한 결정을 내리는 데 도움이 됩니다.

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

Amazon S3

이점	사용 사례	비용 최적화
• API를 통해 액세스 • 버킷과 타 서비스 간에 데이터 스트리밍 또는 복사 • 고가용성 및 다중화를 유지하면서 스토리지 규모 조정 • 대용량 데이터에 비용 효율적인 스토리지 솔루션 • 적절한 저장 사례 적용	• 데이터 수집 및 저장 • 모델 훈련 및 평가 • 타 AWS 서비스와 통합 가능	• 액세스 패턴을 토대로 Amazon S3 스토리지 계층 사용 • 알 수 없는 액세스 패턴의 경우 S3 Intelligent-Tiering 사용 • 액세스 빈도가 낮은 데이터는 아카이브하거나 미사용 데이터 삭제 • 수명 주기 구성을 사용하여 이전 버전 객체 만료 처리

25

Amazon S3는 태스크 처리에 사용되는 다른 AWS 서비스에서 데이터를 수집, 추출, 변환하기 위한 중앙 데이터 레이크 역할을 합니다. 이러한 태스크는 대부분의 ML 워크로드에서 없어서는 안 될 부분입니다. 어디서나 데이터를 저장하고 검색할 수 있는 기능이 있기 때문에 Amazon S3는 확장 가능하고 내구성이 우수하며 안전한 데이터 저장 및 관리가 필요한 워크플로의 핵심 구성 요소입니다.

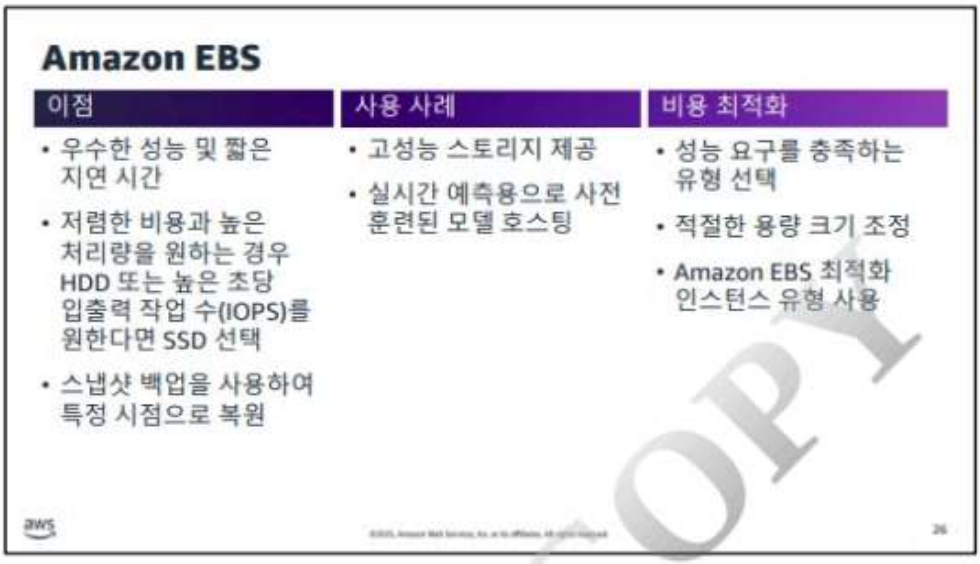
Amazon S3에서는 ML 워크플로와 관련된 다양한 이점이 제공됩니다. 이러한 이점으로는 API 액세스, 스트리밍 기능, 고가용성, 대규모 데이터 세트에 사용 가능한 비용 효율적인 스토리지를 포함합니다. Amazon S3는 대개 데이터 수집과 저장, 모델 훈련과 평가, 그리고 SageMaker AI 및 AWS Glue와 같은 기타 AWS 서비스와의 통합에 사용됩니다.

비용을 최적화하려는 경우에는 액세스 패턴에 따라 적합한 Amazon S3 스토리지 계층을, 미확인 패턴용으로는 S3 Intelligent-Tiering을 사용하고 미사용 데이터는 아카이브하거나 삭제합니다. Amazon S3는 확장성과 내구성이 매우 우수하지만 특정 워크로드에서는 지연 시간이 길어질 수도 있음을 고려해야 합니다.

DO NOT COPY
ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



Amazon EBS에서는 EC2 인스턴스에 연결할 수 있는 지연 시간이 짧은 고성능 블록 스토리지를 제공합니다. 다양한 성능 및 비용 요구 사항을 충족할 수 있도록 하드 디스크 드라이브(HDD) 및 SSD 볼륨 유형이 모두 제공됩니다. Amazon EBS는 대규모 데이터세트에 빠르게 액세스해야 하는 데이터베이스, 웹 애플리케이션, 분석 및 ML 워크로드에 매우 적합합니다.

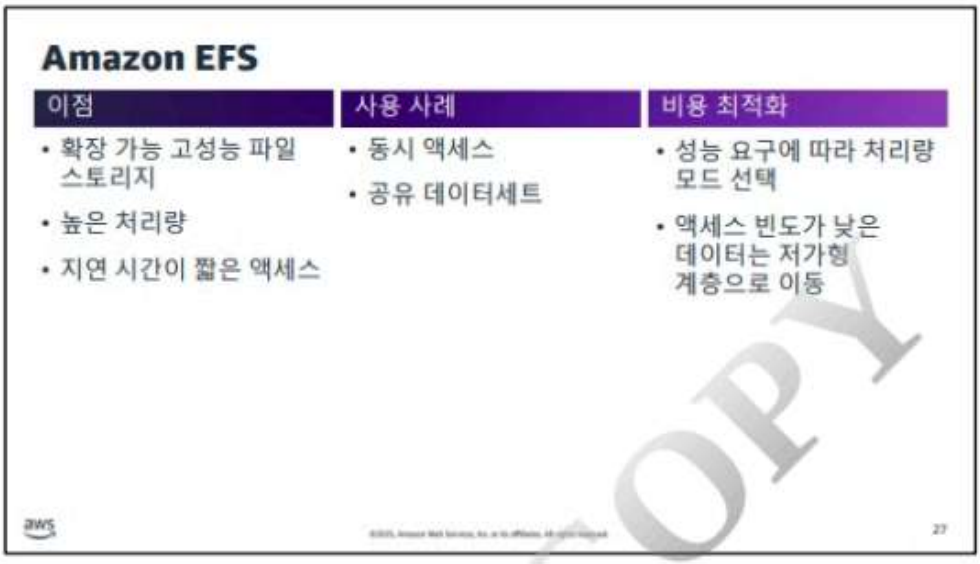
EBS의 주요 이점으로는 우수한 성능, 짧은 지연 시간, 백업 및 재해 복구용으로 특정 시점의 스냅샷을 생성하는 기능 등이 있습니다. ML 워크로드에서 고성능 스토리지를 제공하며 실시간 예측용으로 사전 훈련된 모델을 호스팅려는 경우 Amazon EBS가 특히 유용합니다.

비용을 최적화하려면, 워크로드 요구 사항에 따라 적절한 EBS 볼륨 유형을 선택하고, 용량을 적절하게 조정하고, 최적의 성능을 위해 Amazon EBS 최적화 인스턴스를 사용합니다. 스토리지 아키텍처를 계획할 때는 Amazon EBS와 인스턴스 스토어 간의 절충점을 고려하십시오.

DO NOT COPY
ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



Amazon EFS는 여러 EC2 인스턴스에서 동시에 액세스할 수 있는 고성능 확장 가능 파일 스토리지 시스템을 제공합니다. 파일을 추가하거나 제거하면 스토리지가 자동으로 확장 및 축소되므로 사용량이 변화하더라도 높은 성능이 유지됩니다. Amazon EFS는 NFSv4 프로토콜을 사용하므로 온프레미스 NFS 서버를 사용하는 기존 애플리케이션을 마이그레이션할 수 있습니다.

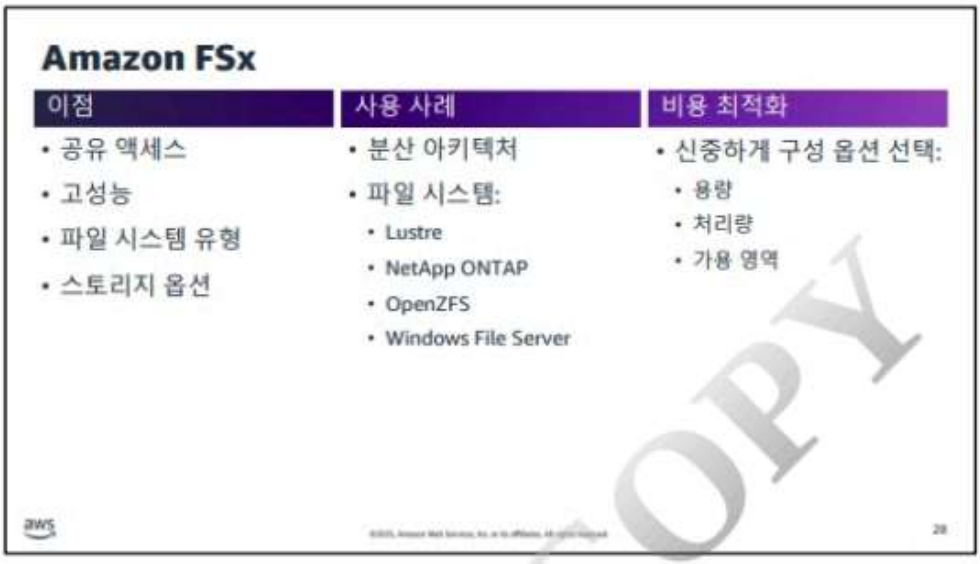
EFS의 주요 이점으로는 높은 처리량, 짧은 액세스 지연 시간, 페타바이트급 공유 파일 시스템 처리 능력 등이 있습니다. ML 워크로드의 경우, 여러 컴퓨팅 인스턴스의 공유 데이터세트에 동시 액세스해야 하는 시나리오의 경우 특히 유용합니다.

비용을 최적화하려는 경우에는 성능 요구에 따라 적절한 처리량 모드를 선택하고 Amazon EFS 공유의 크기를 적절하게 조정합니다. Amazon EBS와 같은 다른 스토리지 옵션과 비교하여 Amazon EFS의 더 높은 가격과 간소화된 규모 간의 절충점을 생각해 보십시오.

DO NOT COPY
ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



Amazon FSx는 Lustre, NetApp ONTAP, OpenZFS, Windows File Server와 같은 널리 사용되는 파일 시스템에 대한 액세스를 제공하는 완전관리형 서비스입니다. 초당 수백만 개의 입출력 작업(IOPS)과 1밀리초 미만의 지연 시간으로 고성능을 제공하므로 ML, 분석, 고성능 컴퓨팅 애플리케이션 등에 적합합니다.

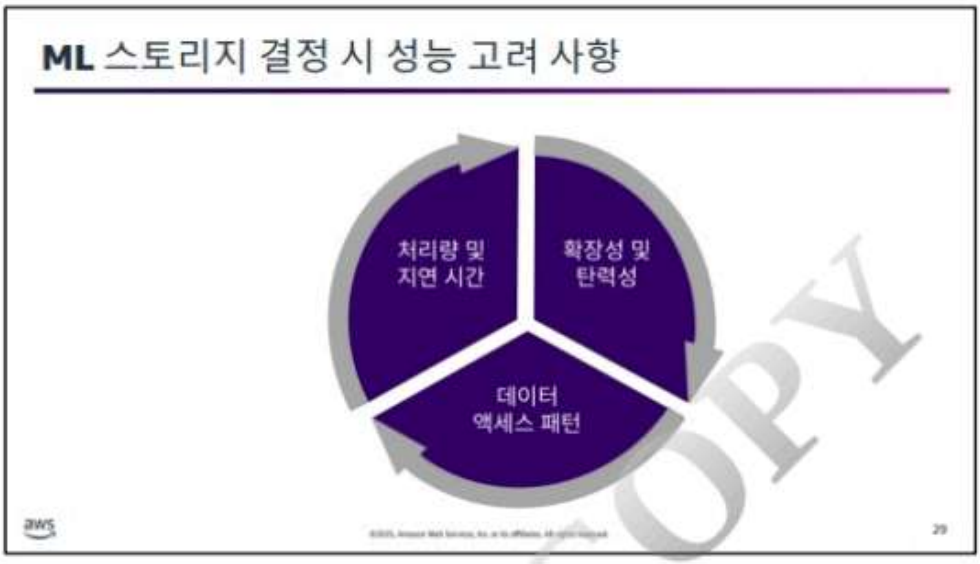
주요 이점으로는 여러 컴퓨팅 인스턴스에서 공유 액세스, 고성능, 다양한 파일 시스템 유형 지원 등이 있습니다. ML 워크로드의 경우, 분산 아키텍처와 모델 훈련에 사용하는 대량의 처리량이 높은 데이터 세트를 호스팅하는데 특히 유용합니다.

비용을 최적화하려는 경우 필요에 따라 스토리지 용량, 처리량 및 배포 옵션(단일 가용 영역 또는 다중 가용 영역)을 신중하게 선택해야 합니다. ML 워크플로를 특정 파일 시스템 유형에 긴밀하게 결합할 때는 복잡성과 벤더 종속의 잠재적인 절충점을 고려해 보십시오.

DO NOT COPY
ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



성능은 보통 ML 워크로드의 스토리지 결정을 좌우하는 요소입니다. 성능 관련 주요 고려 사항으로는 I/O 처리량, 지연 시간, 확장성, 탄력성, 데이터 액세스 패턴 등이 있습니다. 구체적인 요구 사항은 데이터의 크기와 복잡도, 그리고 데이터에 액세스하는 ML 애플리케이션이나 모델에 따라 달라집니다.

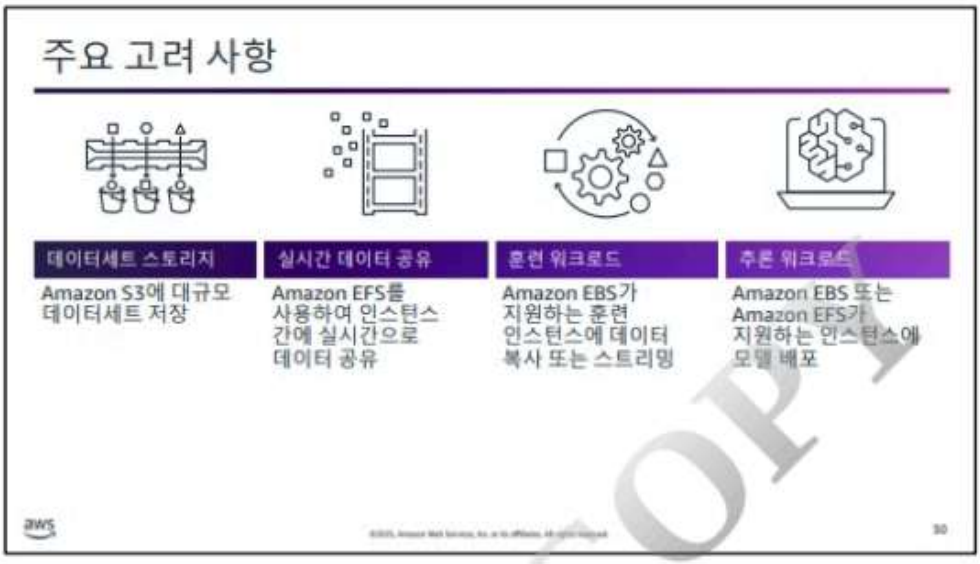
ML 워크로드용 스토리지 서비스를 선택할 때는 특정 애플리케이션 및 데이터세트의 컨텍스트에서 이러한 성능 특성을 파악해야 하는 것이 중요합니다. 각 AWS 스토리지 서비스가 처리량, 지연 시간, 확장성 측면에서 서로 다른 강점을 제공하므로, 이러한 기능을 워크로드 수요에 맞추는 것이 최적의 성능을 위해서 중요합니다.

스토리지 옵션을 평가할 때는 데이터 액세스 빈도, 동시 액세스 필요 여부, 데이터세트 크기 등의 요인을 고려해야 합니다. 성능 요구 사항과 비용 고려 사항의 균형을 맞추면 ML 워크플로에 가장 적합한 스토리지 솔루션을 선택할 수 있습니다.

DO NOT COPY ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



ML 워크로드에는 최적의 스토리지 시스템을 선택하는 것이 중요합니다. 스토리지를 선택할 때 고려해야 하는 주요 사항에는 데이터세트 스토리지 요구 사항, 실시간 데이터 공유, 훈련 또는 추론 워크로드를 위한 데이터 저장 여부 등이 있습니다.

데이터세트 스토리지

Amazon S3를 사용하여 대규모 데이터세트를 저장할 수 있습니다.

실시간 데이터 공유

Amazon EFS 파일 시스템은 공유 데이터세트에 효율적인 동시 데이터 액세스와 높은 처리량을 제공합니다. Amazon EFS는 여러 EC2 인스턴스에서 동일한 데이터세트를 공유함으로써 데이터 공유가 필요한 애플리케이션의 요구 사항을 충족하는 데이터세트 액세스 기능을 제공합니다. 이러한 기능을 제공하는 Amazon EFS는 미디어 처리, 웹 서비스, 콘텐츠 관리 등의 워크로드에 적합합니다.

훈련 워크로드

훈련 워크로드에는 고성능이 필요하고 데이터에 대한 빈번한 랜덤 I/O 액세스가 필요합니다. Amazon EBS 볼륨은 훈련 워크로드에 필요한 랜덤 IOPS를 제공하는 데 적합합니다. 또한, Amazon EC2 인스턴스 스토어 볼륨에서는 데이터가 네트워크에 연결된 볼륨이 아닌 인스턴스 자체에 저장되기 때문에 데이터 액세스 지연 시간이 매우 짧습니다. EC2 인스턴스 스토어 볼륨은 일시적이라는 점을 기억해야 합니다.

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

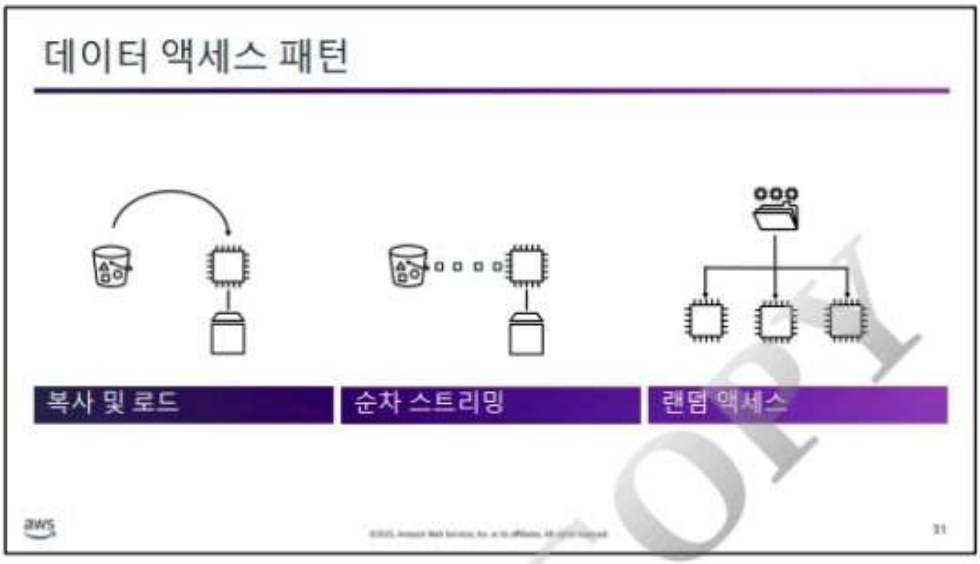
추론 워크로드

추론 워크로드는 예측을 제공하기 위해 응답 시간이 짧아야 하지만 대개 I/O 성능은 높지 않아도 됩니다. Amazon EBS gp3 볼륨 또는 Amazon EFS 스토리지 옵션이 이러한 요구 사항을 충족하는 데 적합합니다. 짧은 지연 시간 요구 사항이 증가함에 따라 Amazon EBS io2 볼륨 유형으로 업그레이드하면 짧은 지연 시간 성능을 개선할 수 있습니다.

DO NOT COPY
ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



ML 워크로드에 적합한 스토리지 서비스를 선택하려면 데이터 액세스 패턴을 이해해야 합니다. 3가지 일반적인 패턴은 다음과 같습니다.

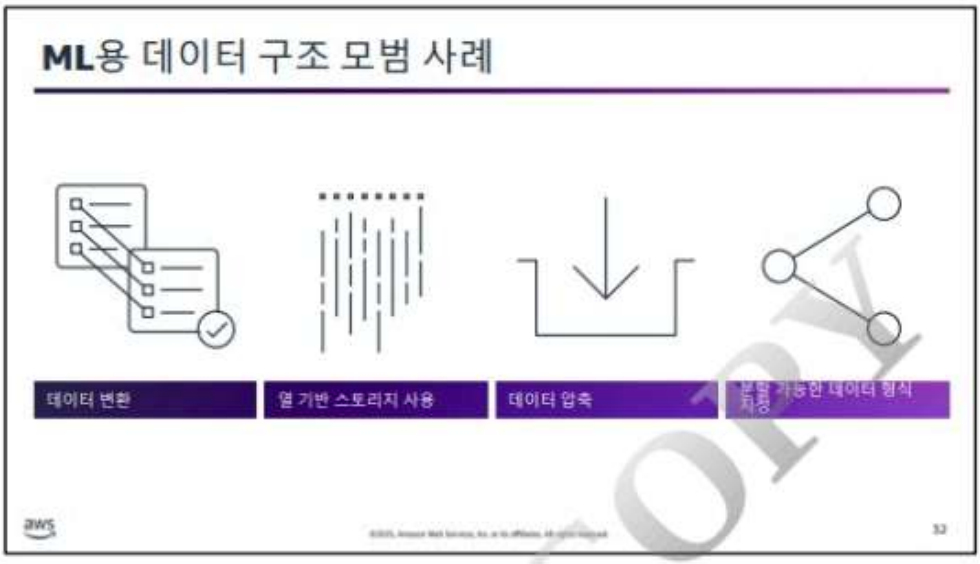
- 복사 및 로드: 데이터가 Amazon S3에서 Amazon EBS가 지원하는 훈련 인스턴스로 한꺼번에 복사됩니다. 이는 메모리에 저장할 수 있으며 자주 변경되지 않는 데이터세트에 적합합니다.
- 순차 스트리밍: 데이터가 배치 또는 개별 레코드로, 일반적으로는 Amazon S3에서 Amazon EBS 지원 인스턴스로 스트리밍 됩니다. 이 패턴은 메모리에 저장할 수 없는 대규모 데이터세트의 경우 잘 맞습니다.
- 랜덤 액세스: 데이터는 무작위로 액세스되는데, 대개 Amazon FSx나 Amaon EFS 등의 공유 파일 시스템을 사용합니다. 이 패턴은 공유 데이터세트에 동시에 액세스해야 하는 워크로드의 경우 적합한 옵션입니다.

스토리지를 적절한 액세스 패턴에 맞추면 ML 워크플로의 성능과 효율성을 대폭 개선할 수 있습니다. 특정 사용 사례에 가장 적합한 방식을 결정할 때는 데이터세트 크기, 업데이트 빈도, 동시 액세스 요구 사항 등의 요인을 고려해야 합니다.

DO NOT COPY ssleeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



데이터 구조를 최적화하면 ML 워크로드의 성능을 개선하고 비용을 줄일 수 있습니다. 고려할 수 있는 모범 사례는 다음과 같습니다.

- ML 태스크와 관련이 있는 정보만 포함되도록 데이터를 변환합니다. 이 사례는 처리 시간을 단축되고 효율성은 높아집니다.
- Parquet, ORC 등의 열 기반 저장 형식을 사용합니다. 특히 분석 쿼리에서 이러한 형식을 사용하면 필요한 스토리지 공간을 크게 줄이고 데이터를 더욱 빠르게 처리할 수 있습니다.
- 각 데이터 유형에 적합한 방법을 사용하여 데이터를 압축합니다. 그러면 스토리지 요구 사항이 감소하며 읽기 성능이 개선될 수 있습니다.
- 데이터를 분할 가능한 형식으로 만듭니다. 그러면 여러 인스턴스에서 분산 훈련을 진행할 수 있습니다. 이 사례는 데이터 병렬 처리 속도가 빨라지며 전반적인 성능도 개선됩니다.

이러한 사례를 구현하면 스토리지 요구 사항 감소, 컴퓨팅 시간 감소, 데이터 전송 비용 감소, 액세스 요청 감소 등 여러 영역에서 비용을 줄일 수 있습니다. 데이터 형식과 사용 방식을 최적화하면, 전반적인 효율성을 높이고 ML 워크플로의 데이터가 차지하는 공간을 줄일 수 있습니다.

NOT COPY
ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

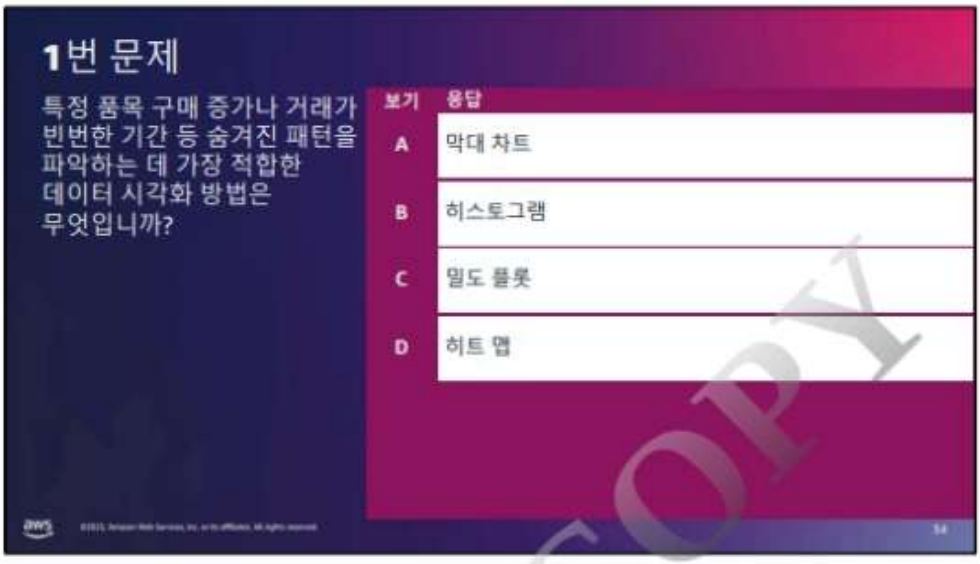
이 슬라이드에서는 지식 확인 섹션을 소개합니다.
이제 간단한 지식 확인을 진행하여 이 모듈에서 알아본 ML용 데이터 처리 관련 주요 개념을 복습해 보겠습니다.

이 지식 확인을 통해 다음 모듈로 넘어가기 전에 수업 내용을 더욱 정확하게 이해하고 헷갈리는 부분을 확실하게 파악할 수 있습니다.

DO NOT COPY
ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



거래 데이터에서 숨겨진 패턴을 식별하는 데 가장 적합한 탐색적 데이터 분석 시각화 방법을 찾는 문제입니다. 정답을 선택하려면 각 시각화 유형의 장점과 한계를 고려해야 합니다.

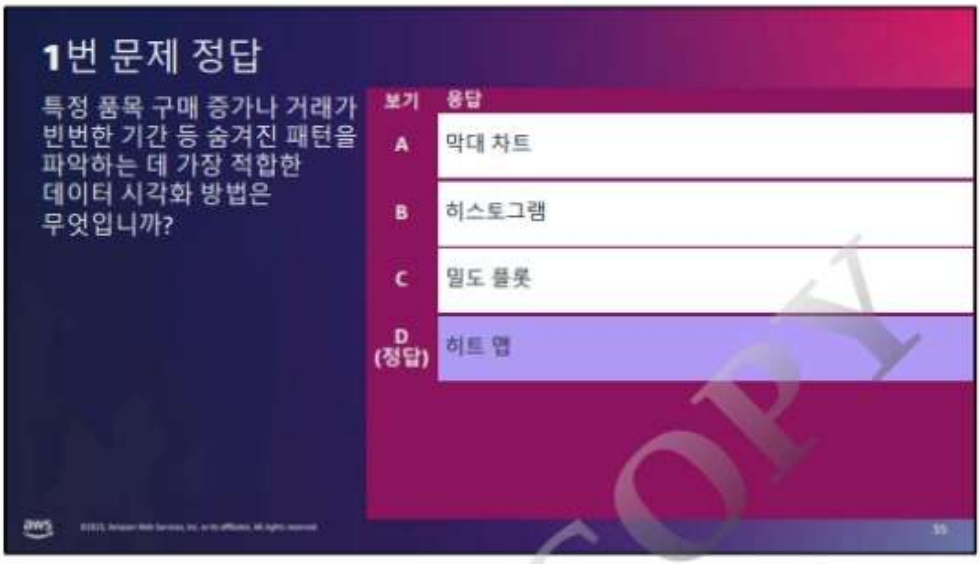
- 막대 차트는 범주형 데이터 비교 시에 적합합니다.
- 히스토그램에는 수치 데이터 분포가 표시됩니다.
- 밀도 플롯에는 변수 하나의 분포가 표시됩니다.
- 히트 맵에서는 2개 차원에서 패턴을 표시할 수 있습니다.

시간의 경과에 따른 거래 빈도나 품목 구매 패턴을 가장 정확하게 표시할 수 있는 방법이 무엇인지를 생각해 보십시오. '숨겨진 패턴을 식별'할 수 있는 방법이 정답입니다. 이러한 패턴을 식별하려면 여러 차원의 데이터를 동시에 시각화해야 할 수도 있습니다.

DO NOT COPY ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



정답은 D, 히트 맵입니다. 히트 맵이 거래 데이터에서 숨겨진 패턴을 식별하는 데 적합한 이유는 다음과 같습니다.

- 색상을 사용하여 2개 데이터 차원을 시각화함으로써 시간 경과에 따른 거래 패턴을 표시할 수 있습니다.
- 다른 시각화 유형에서는 명확하게 표시되지 않을 수도 있는 클러스터나 추세가 표시될 수도 있습니다.
- 대규모 데이터세트에 적합해서 전반적인 패턴을 한 눈에 확인할 수 있습니다.

그 외의 답안이 적합하지 않은 이유는 다음과 같습니다.

A. 막대 차트는 시간 경과에 따른 패턴은 막대 차트을 보여주지 않고, 범주형 데이터 비교 시에 더 적합합니다.
B. 히스토그램에는 변수 하나의 분포가 표시되지만 시간 차원은 표시되지 않습니다.
C. 밀도 플롯은 데이터 분포를 이해하는데 유용하지만 시간 경과에 따른 패턴은 명확하게 표시되지 않습니다.

DO NOT COPY ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

2번 문제

ML 모델 훈련을 위한 데이터 스토리지 요구 사항 및 해당 요구 사항을 충족하는 데 가장 적합한 AWS 스토리지 솔루션을 연결하십시오.

	기계 학습 요구 사항	AWS 스토리지 솔루션
A	동시 액세스 및 공유 데이터세트	Amazon S3
B	스냅샷 백업을 사용하여 특정 시점으로 복원	Amazon FSx
C	Windows 파일 서버 공유 액세스	Amazon EBS
D	적절한 스토리지 클래스 사용	Amazon EFS

36

ML 모델 훈련을 위한 데이터 스토리지 요구 사항 및 해당 요구 사항을 충족하는 데 가장 적합한 AWS 스토리지 솔루션을 연결하는 문제입니다.

DO NOT COPY
ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

2번 문제 정답

ML 모델 훈련을 위한 데이터 스토리지 요구 사항 및 해당 요구 사항을 충족하는 데 가장 적합한 AWS 스토리지 솔루션을 연결하십시오.

기계 학습 요구 사항	AWS 스토리지 솔루션
A 동시 액세스 및 공유 데이터세트	Amazon EFS
B 스냅샷 백업을 사용하여 특정 시점으로 복원	Amazon EBS
C Windows 파일 서버 공유 액세스	Amazon FSx
D 적절한 스토리지 클래스 사용	Amazon S3

기계 학습 모델 훈련용 스토리지를 선택할 때는 워크로드의 구체적인 요구 사항에 맞는 기능을 제공하는 AWS 스토리지 솔루션을 선택해야 합니다.

Amazon S3는 사전 훈련된 모델이나 아카이브 데이터 등 자주 액세스하지 않아도 되는 대규모 데이터세트를 저장하기에 적합한 옵션입니다. 확장성과 내구성 때문에 이런 용도에 적합합니다.

고성능 공유 액세스가 필요한 워크로드의 경우 Amazon EFS는 여러 EC2 인스턴스에서 지연 시간이 짧은 동시 액세스를 제공합니다. 이러한 이점을 제공하는 Amazon EFS는 ML 파이프라인의 실시간 데이터 공유에 좋은 선택입니다.

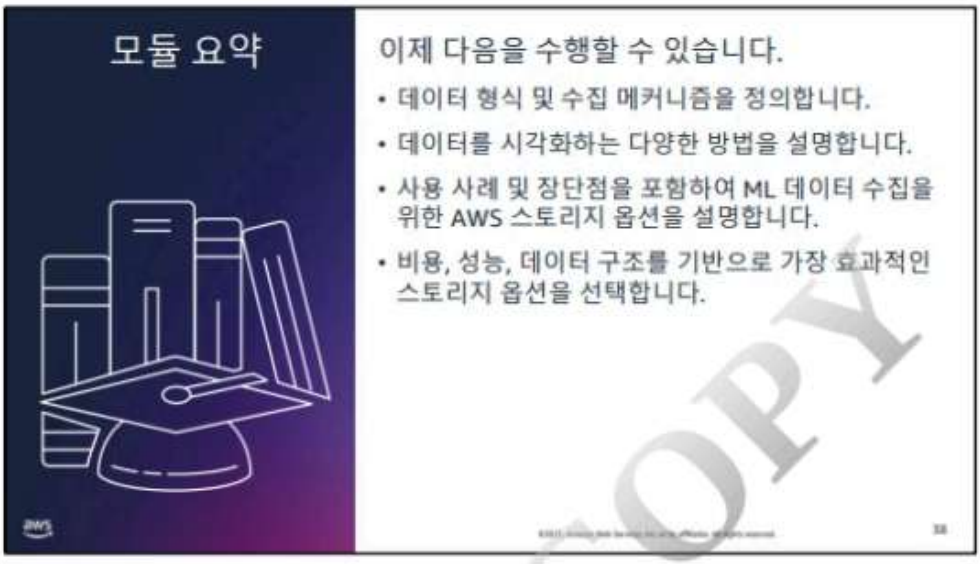
Amazon EBS는 EC2 인스턴스에 직접 연결되어 있어 IOPS는 높고 지연 시간은 짧은 스토리지를 제공하므로 이는 모델 훈련의 집중적인 I/O 수요에 적합합니다.

Windows 기반 워크로드 또는 특정 파일 시스템 기능이 필요한 워크로드의 경우, Amazon FSx는 성능과 호환성에 최적화된 완전관리형 파일 스토리지를 제공합니다.

용량, 성능, 액세스 패턴 및 파일 시스템 요구 사항 측면에서 ML 워크로드의 스토리지 요구 사항을 면밀하게 평가하면 모델 훈련 프로세스를 가장 효율적으로 지원하는 AWS 스토리지 솔루션을 선택할 수 있습니다.

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



DO NOT COPY
ssleeeee1@gmail.com

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

DO NOT COPY
ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

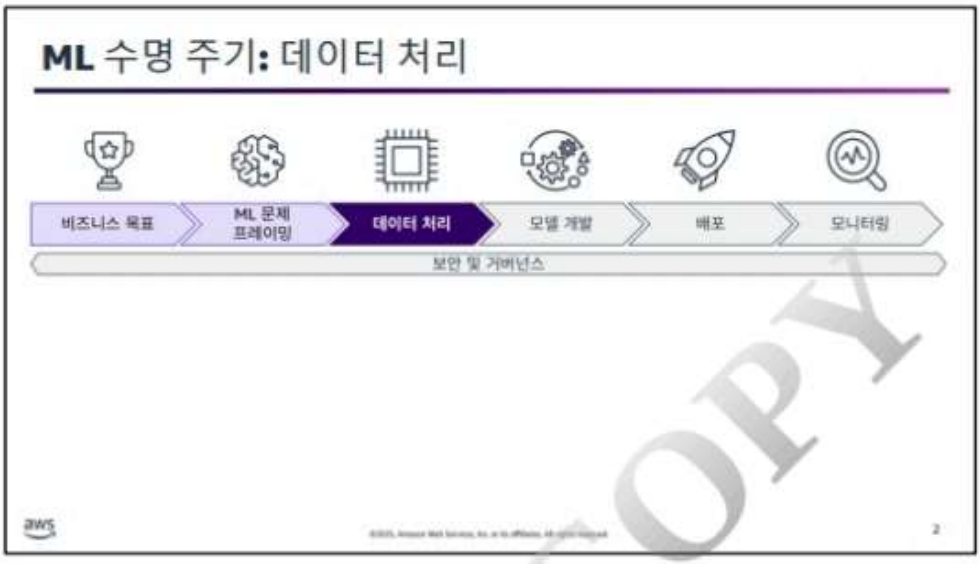
기계 학습 수명 주기의 이 시점에서는 데이터 수집, 흡수, 저장에 익숙하실 것입니다. 데이터 유형과 형식과 같은 데이터 기초를 다루었고, 핵심 AWS 스토리지 서비스도 파악했을 것입니다. 따라서 특정 기계 학습 요구 사항을 기반으로 가장 효과적인 AWS 스토리지 솔루션과 데이터 형식을 선택할 수 있습니다.

이 모듈에서는, ML 모델 훈련 준비 시에 데이터를 변환하는 방법을 알아봅니다. 먼저 데이터 클리닝, 인코딩 및 피처 엔지니어링 관련 개념과 기법을 살펴봅니다. 그리고 Amazon SageMaker Data Wrangler, Amazon SageMaker Processing, Amazon SageMaker Feature Store, Amazon EMR 등의 AWS 서비스를 사용하여 효율적인 방식을 통해 대규모로 데이터를 변환하는 방법을 알아봅니다.

DO NOT COPY
ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



회사의 ML 요구에 적합한 방식으로 데이터를 수집, 흡수, 추출 및 저장하였으며, 이젠 그 데이터를 깔끔하고 인코딩 되고 엔지니어링 된 데이터세트로 변환할 준비가 되었습니다.

데이터 클리닝, 범주형 인코딩, 피처 엔지니어링은 보다 정확하게 모델에 바로 사용할 수 있는 데이터세트를 생성하는 데 사용할 수 있는 중요한 기법입니다. 적절한 데이터 처리는 모델의 성능과 정확도에 큰 영향을 미칠 수 있습니다.

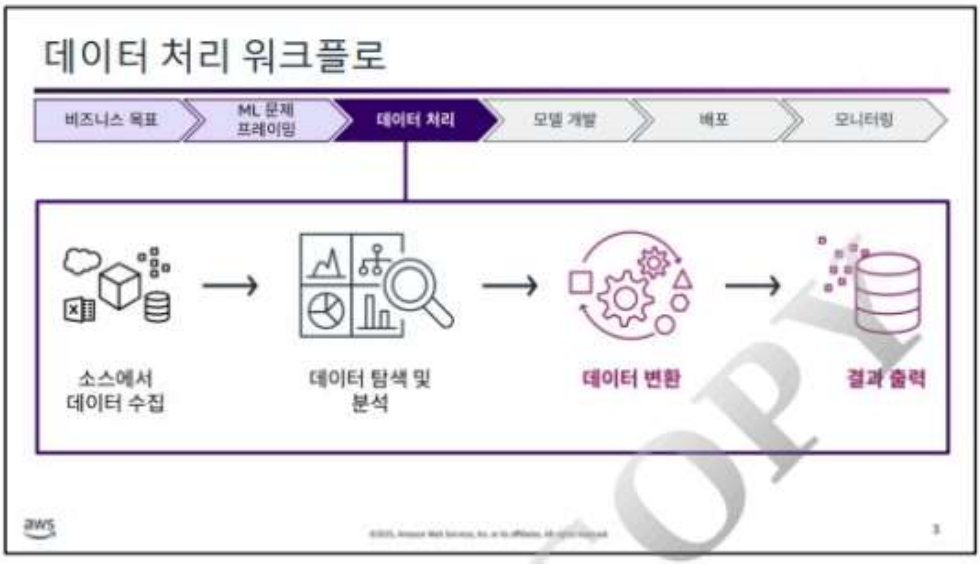
이 단계에서는 다음을 주로 수행합니다.
- 데이터를 정리하여 잘못된 데이터, 중복 데이터 또는 누락된 데이터 처리
- 범주형 변수를 수치 표현으로 인코딩
- 새 특성을 엔지니어링하여 데이터의 중요한 관계 캡처

목표는 ML 알고리즘이 효과적으로 학습을 하고 정확한 예측을 할 수 있도록 고품질 데이터세트를 생성하는 것입니다.

DO NOT COPY
ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



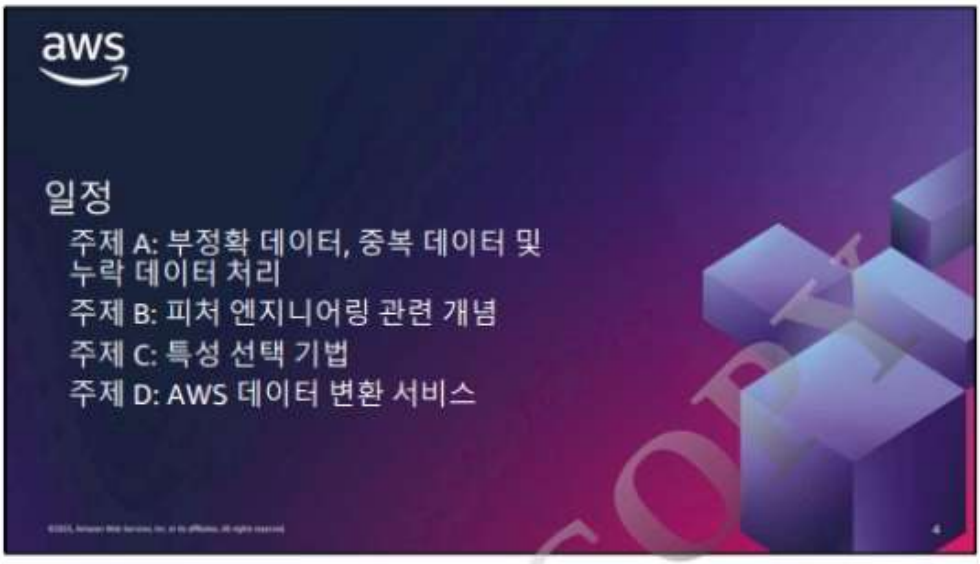
이전 모듈에서는 데이터 처리 워크플로의 처음 두 단계인 소스에서 데이터 수집과 데이터 탐색 및 분석에 대해 알아보았습니다. 이전 모듈에서는 데이터세트 형식 지정과 관련된 변환 시나리오만을 다루었습니다. 이 모듈에서는 데이터 클리닝, 피처 엔지니어링, 특성 선택을 비롯하여 데이터 변환에 대해 더 자세히 살펴봅니다.

1. **데이터 변환**: 데이터 클리닝 및 피처 엔지니어링 단계를 지정하고 전체 데이터세트에 변환을 적용하기 전에 데이터 샘플에서 변환을 테스트합니다.
2. **결과 출력** : 원하는 스토리지 솔루션으로 처리된 데이터를 내보냅니다.

DO NOT COPY
ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



이 모듈에서 다룰 주요 주제는 다음과 같습니다.

A. 부정확 데이터, 중복 데이터 및 누락된 데이터 처리: ML 모델의 품질과 신뢰성을 보장하는 데이터를 정리하고 전처리하는 기술을 살펴봅니다.
B. 피처 엔지니어링 개념: 새 특성을 생성하거나 기존 특성을 수정하여 모델 성능을 개선하는 방법을 알아봅니다.
C. 특성 선택 기법: ML 모델과 가장 관련성이 높은 특성을 선택하는 기법에 대해 논의합니다.
D. AWS 데이터 변환 서비스: 데이터 변환 프로세스를 간소화하고 자동화하는 데 도움이 되는 AWS 도구와 서비스를 살펴봅니다.

이러한 개념과 기법을 파악하면 AWS에서 기계 학습 프로젝트용 고품질 데이터세트를 준비하여 정확도와 효율성이 더욱 높은 모델을 만드는 데 도움이 됩니다.

DO NOT COPY ssleeeee1@gmail.com

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

데이터를 수집, 흡수, 저장한 후에 데이터세트를 준비할 때 첫 번째 단계 중 하나는 데이터 정리입니다. 이는 시간이 많이 걸릴 수도 있지만 효과적인 ML 모델을 훈련하기 위해서는 중요합니다.

데이터 품질은 모든 ML 프로젝트 성공의 기본입니다. 품질이 떨어지는 데이터는 부정확한 모델, 편향된 예측, 리소스 낭비로 이어질 수 있습니다. 이 섹션에서는 데이터 품질과 관련하여 흔히 발생하는 3가지 문제를 중점적으로 살펴봅니다.

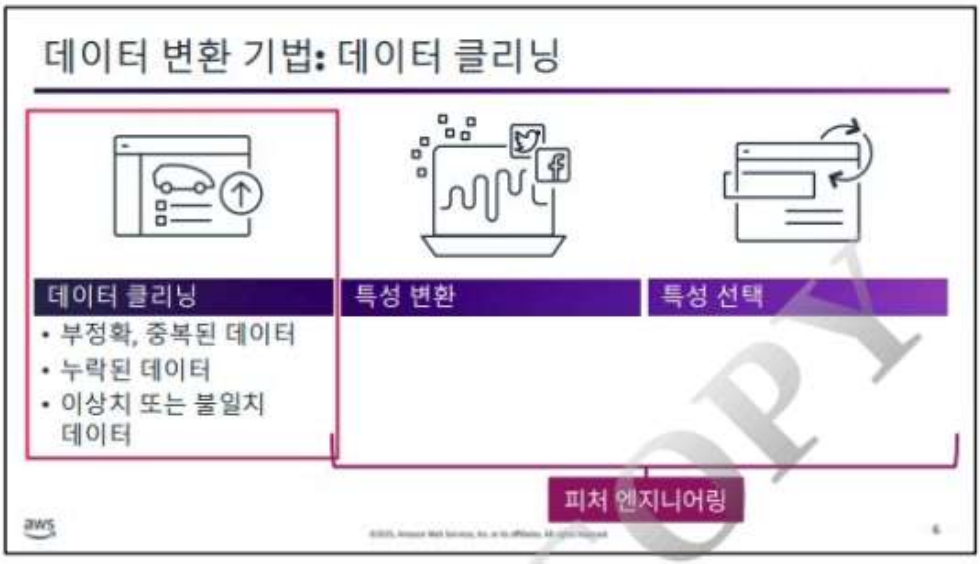
- 부정확한 데이터: 정확하지 않거나 필요한 형식과 일치하지 않는 값
- 중복 데이터: 모델 훈련 방식의 왜곡을 초래할 수 있는 중복 레코드
- 누락된 데이터: 적절하게 처리해야 하는 불완전한 정보

이러한 문제를 해결하면 ML 모델을 위한 더 안정적인 토대를 만들 수 있습니다. 이 섹션에서는 이러한 데이터 품질 문제를 효율적으로 확인하여 해결하는 데 사용할 수 있는 기술 및 모범 사례를 살펴봅니다.

DO NOT COPY ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



데이터 클리닝에서는 데이터 오류, 불일치, 부정확성을 식별하고 수정(또는 제거)하는 방식으로 ML에 맞게 데이터세트를 준비합니다. 데이터를 정리한 후에는 ML 알고리즘이 패턴을 찾고 ML 모델을 훈련하는 데 사용할 수 있는 특성으로 데이터세트를 변환할 수 있습니다. 피처 엔지니어링은 특성을 수정하거나 새 특성을 생성하여 데이터의 중요한 관계를 캡처하는 프로세스입니다.

이 주제에서는 흔히 발생하는 데이터 품질 문제 및 각 문제를 해결할 수 있는 데이터 클리닝 기법에 대해 다룹니다. 데이터 품질 문제에는 아래가 포함됩니다.

- 부정확 데이터 또는 중복 데이터: 잘못된 항목과 중복 레코드 또는 특성을 식별하여 제거하거나 수정합니다.
- 누락된 데이터: 대치 또는 기타 기법을 통해 데이터세트의 격차를 해소합니다.
- 이상치 또는 불일치 데이터: 모델 왜곡을 초래할 수 있는 극단적인 값을 감지하여 처리합니다.

DO NOT COPY
ssleeeee1@gmail.com

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

ML 모델 훈련용으로 유용한 데이터는 일관성이 있어야 합니다. 부정확하거나 일관성이 없는 훈련 데이터는 ML 모델의 정확한 예측을 하기 어렵게 합니다. 다음은 잘못된 데이터의 몇 가지 예입니다.

- 언어 차이: 용어나 번역 사용에 일관성이 없음
- 오타: 철자 오류가 있거나 값이 잘못 입력됨
- 일관성 없는 형식: 날짜 형식이나 측정 단위가 다양함

이러한 문제를 식별한 후 잘못된 데이터를 처리하는 기법을 사용할 수 있습니다. 예를 들어, 데이터 샘플을 수동으로 검토를 통해 오류 패턴을 식별할 수 있습니다. 다음은 잘못된 데이터를 처리하는 몇 가지 기법입니다.

- 올바른 값 채우기: 잘못된 값을 (알려진 경우) 올바른 값으로 대체합니다.
- 형식 수정: 데이터세트 전반에서 형식을 표준화 합니다. (예: 날짜 형식)
- 데이터세트에서 제외: 심각하게 부정확하거나 관련성이 거의 없는 데이터 포인트는 제거합니다.

잘못된 데이터를 다룰 때는 오류의 원인을 이해하고 오류를 수정할 체계적인 접근 방식을 개발해야 합니다. 자동화된 도구와 데이터 검증 규칙은 특히 데이터세트가 대규모인 경우 이 프로세스를 간소화하는데 도움이 됩니다.

DO NOT COPY
ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

중복 데이터는 ML 모델 품질에 영향을 줄 수 있는 또 다른 일반적인 문제입니다. 중복 레코드가 있으면 데이터세트가 왜곡되어 편향된 결과가 나오고 과대적합 현상이 나타날 수 있습니다. 다음은 중복 데이터의 몇 가지 예입니다.

- 여러 소스의 복사본: 다양한 시스템이나 데이터베이스에서 데이터를 수집
- 잘못된 데이터 입력: 수동 입력 오류로 인해 레코드가 중복

중복 레코드를 식별하여 제거하는 프로세스가 **중복 제거**입니다. 중복 제거는 데이터세트를 스캔하여 중복된 정보가 있는지 확인하고, 정보의 복사본 하나를 보존하고 다른 인스턴스를 저장된 복사본으로 연결하는 포인터로 대체하는 방식으로 작동합니다. 이 프로세스는 스토리지 용량을 대폭 늘리고 데이터 품질을 개선할 수 있습니다.

데이터 중복 제거 시에는 다음 작업을 수행해야 합니다.

- 고유 식별자로 사용할 필드를 결정합니다.
- 중복 레코드에서 상충하는 정보를 처리할 방법을 결정합니다.
- 대규모 데이터세트의 경우 자동 도구 사용을 고려합니다.

중복을 제거하면 ML 모델 성능이 개선되긴 하지만, 데이터가 중복된 이유를 파악하고 근본 원인을 해결하여 나중에 데이터 중복 현상을 방지하지 않도록 하는 것도 중요합니다.

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

불완전하거나 누락된 데이터 문제는 기계 학습 프로젝트에서 흔히 발생하는 문제입니다. 누락된 값을 적절하게 처리하면 데이터 무결성을 유지하고 모델을 훈련하여 더 정확한 예측을 하는데 도움이 됩니다. 누락된 데이터 문제를 체계적으로 해결하는 방식은 다음과 같습니다.

- 데이터세트에서 누락된 값을 식별합니다.
- 값이 누락된 이유를 확인합니다.
  - 무작위 누락(MAR): 관찰 데이터에 따라 다름
  - 완전 무작위 누락(MCAR): 데이터와는 무관
  - 비무작위 누락(MNAR): 관찰되지 않은 데이터에 따라 다름

급여에 대한 설문조사 질문을 공란으로 남겨둔 설문조사 데이터를 분석해 봅니다.

- 고학력 조사 대상은 해당 질문을 건너뛸 가능성이 더 높음: MAR
- 예상치 못한 소프트웨어 결함으로 인해 값이 입력되지 않는 상황이 있음: MCAR
- 급여 수준이 높은 조사 대상이 해당 질문의 답변을 거부함: MNAR

누락된 데이터 처리를 위한 기법을 적용합니다.

- 값이 누락된 행이나 열 삭제: 영향을 받는 행이 거의 없거나 열의 많은 값이 영향을 받는 경우 적합합니다.
- 값 교체: 누락된 값을 추정치로 대체합니다. 기법에는 평균값, 중앙값, 회귀에 기반한 대치 또는 다중 대치가 포함됩니다.

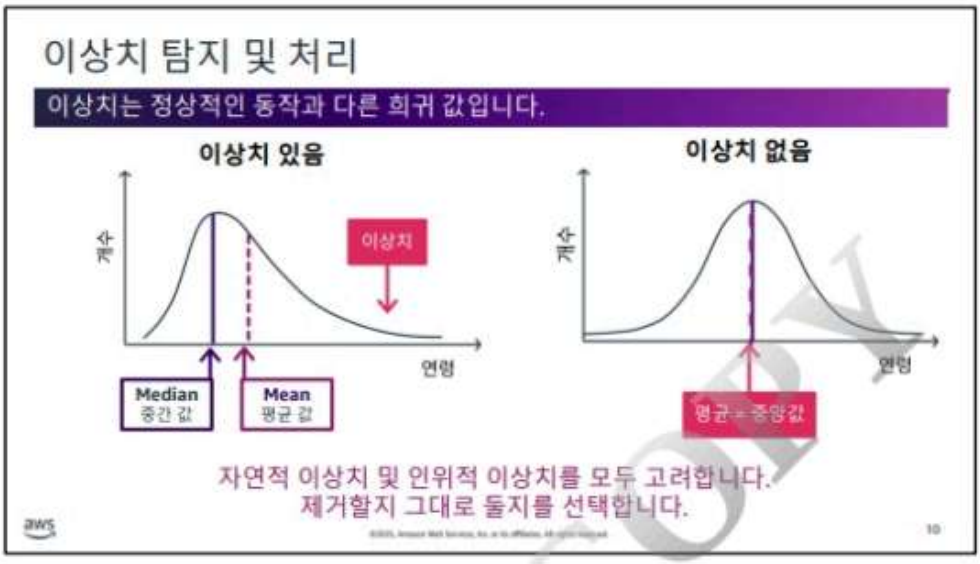
누락된 데이터 처리 방법을 결정할 때는 다음을 수행합니다.

- 누락된 데이터의 양과 패턴을 고려합니다.
- 영향을 받는 변수의 중요도를 평가합니다.
- 분석이나 모델에 미치는 잠재적 영향을 평가합니다.

선택하는 기법에 따라 결과가 크게 달라질 수 있습니다. 여러 접근법을 시도하고 그것이 모델 성능에 미치는 영향을 비교하는 것이 유용할 때가 많습니다.

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



이상치는 데이터세트 내의 다른 관찰 데이터와 크게 다른 데이터 포인트입니다. 어떤 시나리오에서는 이상치가 중요한 통찰력을 제공하지만 다른 시나리오에서는 이상치가 분석을 왜곡되고 모델 성능에 영향을 미칠 수 있습니다. 데이터의 이상치를 식별하고 처리할 방법을 신중하게 고려해야 합니다.

데이터세트 분포와 관련된 2가지 통계를 비교하면 이상치를 식별할 수 있습니다.

- **평균**은 데이터세트 내 모든 값의 평균입니다. 이상치는 평균에 상당한 영향을 미칠 수 있습니다.
- **중앙값**은 정렬된 데이터세트의 중간 값입니다. 중앙값은 극단적인 이상치에 덜 민감합니다.

이상치 처리 시에는 다음을 수행합니다.

- 평균과 중앙값을 모두 계산합니다.
- 두 값을 비교합니다. 두 값이 매우 다르면 이상치가 있는 것일 수 있습니다.
- 상자 플롯이나 산점도 등의 시각화 기법을 사용하여 이상치가 있을 가능성을 파악합니다.

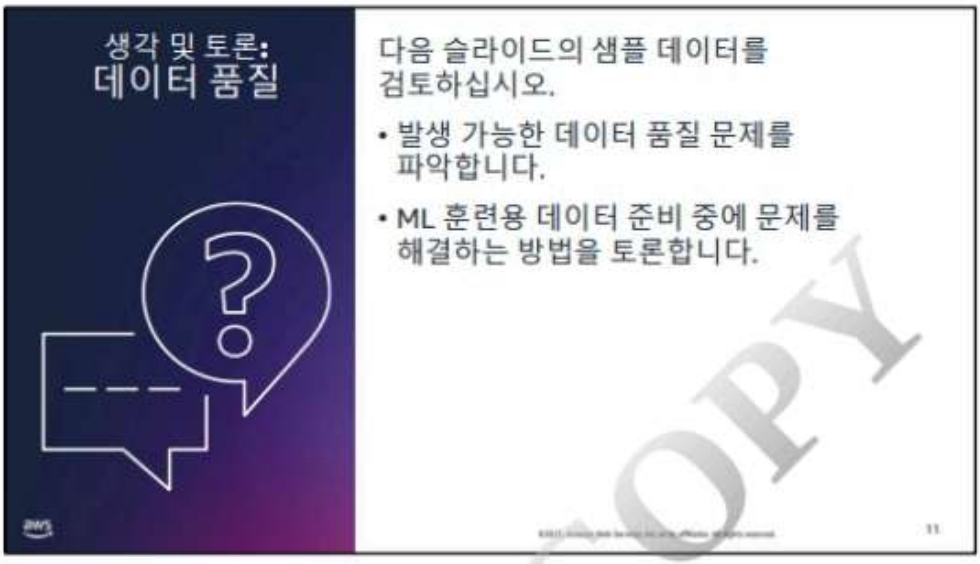
자연적 이상치 및 인위적 이상치를 모두 고려합니다.

- **자연적 이상치**는 유효한 데이터 포인트를 나타내는 극단적인 값입니다. 키 데이터세트에서 키가 매우 큰 사람이 여기에 해당됩니다.
- **인위적 이상치**는 데이터 수집 또는 처리 오류를 나타냅니다(예: 발생할 수 없는 값).

구체적인 사용 사례 및 이상치의 특성에 따라 이상치를 그대로 둘지 제거할지를 결정하게 됩니다. 데이터세트에서 이상치를 처리할 방법을 결정하기 전에 항상 이상치가 발생한 원인을 조사해야 합니다.

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



DO NOT COPY
ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

데이터 품질 문제 살펴보기

참가자 ID	연령	소득	학력	주	활성 구독 여부	구독 유형
123456	39	45,000/년	4년제 학위	NY	예	연간
123457	23	1,500/월	고졸(프랑스어)	MN	아니요	없음
123458	78		석사, 박사		예	매월
123459	20	3,000,000/년	고졸	CA	아니요	없음
123450	154	53,000/년		석사, 박사		
123456	39	45,000/년	4년제 학위	NY	예	연간
...	...	...	...	...	...	...

©2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.
12

어느 연구 팀이 참가자 정보가 포함된 데이터세트를 사용하고 있습니다. 이 데이터를 조사할 때는 정리해야 할 수 있는 여러 가지 문제 유형을 확인해야 합니다.

- 부정확한 데이터: 명백히 잘못되었거나 문맥상 말이 안되는 값을 찾아봅니다.
- 중복된 데이터: 행이 다른 행과 정확히 같거나 거의 중복되지 않는지 확인합니다.
- 일관성 없는 형식: 같은 유형의 데이터라도 행 마다 형식이 다르게 지정되어 있는 필드를 확인합니다.
- 잘못 배치된 데이터: 잘못된 열에 입력되어 있는 값을 찾습니다.
- 누락된 데이터: 비어 있는 셀이나 자리 표시자 값을 확인합니다.

활동:
각 열과 행을 검토하여 이와 같은 데이터 품질 문제일 가능성이 있는 항목을 찾습니다. 분석 또는 기계 학습용으로 이 데이터세트를 준비할 때 각 문제를 해결하는 방법을 고려합니다.

DO NOT COPY
ssleeeee1@gmail.com

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

샘플 데이터세트에는 여러 가지 데이터 품질 문제가 있으며, 이러한 문제를 해결하려면 데이터를 정리해야 합니다.

- **중복 데이터**: 마지막 행은 참가자 ID가 123456인 첫 번째 행과 똑같은 복제본입니다.
- **잘못 배치된 데이터**: 주 열에 '석사, 박사' 값이 잘못 입력되어 있습니다.
- **형식 일관성이 없는 데이터**:
  - 나머지 항목은 모두 연 소득인데 두 번째 행의 값만 월 소득입니다.
  - 나머지 학력 값은 모두 영어인데 'Baccalauréat(학사의 프랑스어)' 학력 값만 프랑스어입니다.
- **누락된 데이터**:
  - 3번 행에서 소득 및 거주지가 누락되었습니다.
  - 참가자 ID가 123450인 5번 행의 3개 열에서 데이터가 누락되었습니다.
- **이상치**: 연 소득 3,000,000은 기타 항목에 비해 너무 큽니다.

데이터 정리에서는 분석 또는 기계 학습 태스크에 사용할 수 있도록 데이터세트를 일관성과 정확도가 더 높은 상태로 변환합니다.

이러한 데이터 품질 문제를 해결하기 위해 사용할 수 있는 기법을 생각해 보십시오.

DO NOT COPY ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

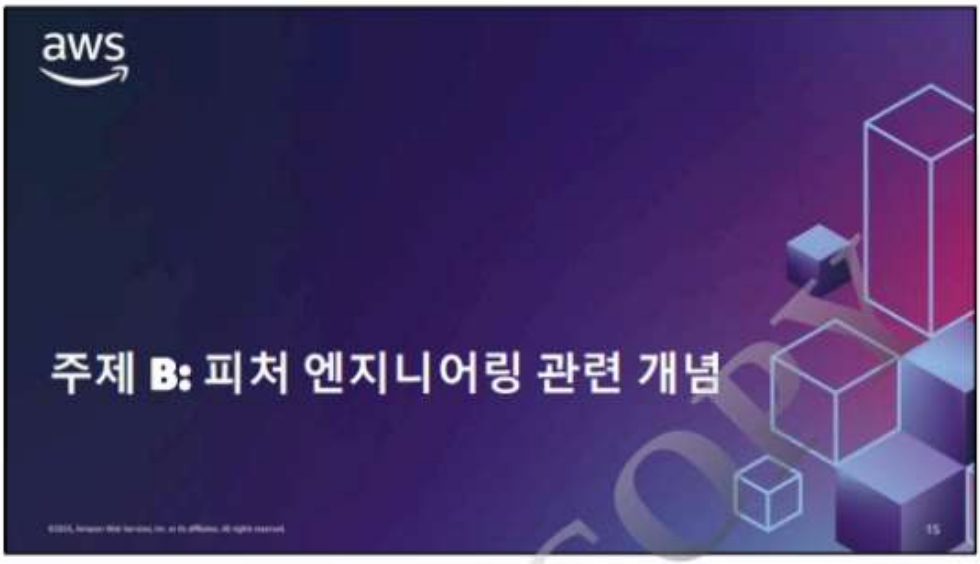
Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

각 데이터 품질 문제는 이 모듈에서 설명하는 기법으로 '정리'할 수 있습니다.

- **중복 데이터 제거**: 마지막 행은 첫 번째 행과 똑같은 복제본이므로 데이터세트의 편향 방지를 위해 전체 행을 제거합니다.
- **잘못 배치된 데이터 재배치**: '석사, 박사' 값을 학력 열로 재배치합니다. 그러면 이 항목의 거주지 필드에서 데이터 누락이 발생합니다.
- **일관성 없는 형식 표준화**:
  - 급여 값 '1,500/월'을 연봉으로 변환합니다.
  - 일관성 유지를 위해 학력 값 'Baccalaureat'를 영어 '고졸'로 번역합니다.
- **누락된 데이터 제거 또는 대치**:
  - 3번 행에서 누락되어 있는 급여 데이터와 거주지는 기존 데이터에서 대치할 수 있습니다.
  - 참가자 ID가 123450인 행에서는 3개 열의 데이터가 누락되어 있는데, 해당 행은 제거할 수 있습니다. 이 기법 사용 시 데이터세트에서 행이 많이 제거된다면 값 대치나 대체를 대신 사용할 수 있습니다.
- **이상치 검사 및 유지**:
  - 소스 데이터 검토 결과 이상치인 '3,000,000/년'은 실제 데이터임이 확인되었습니다. 전체 데이터세트의 유사한 범위에 여러 예제 항목이 있었기 때문입니다. 그러므로 해당 이상치는 유의미한 데이터로 유지하면 됩니다. ML 문제 및 데이터세트의 특성에 따라 이상치 유지 여부를 선택합니다.

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



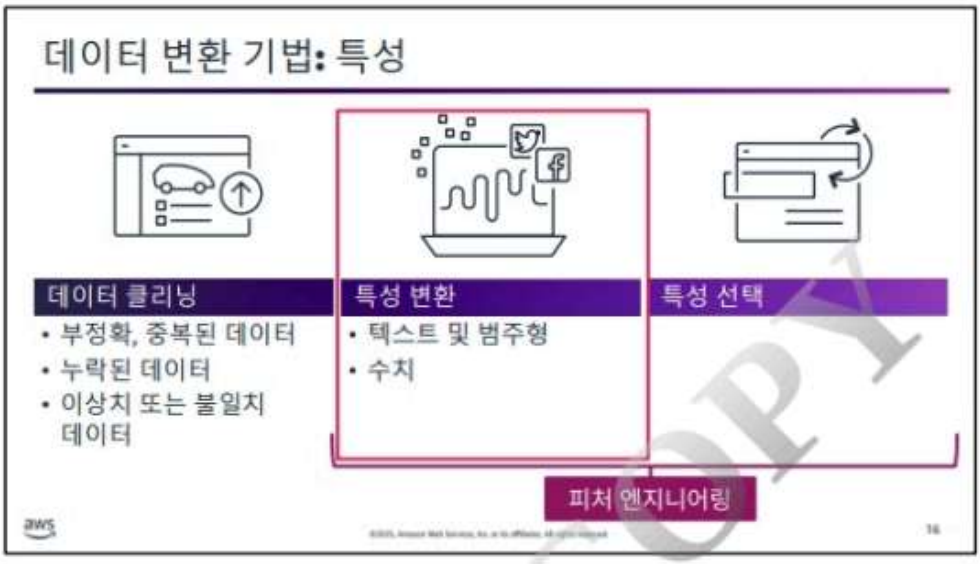
특성은 모델이 데이터의 기본적인 관계를 더 잘 포착하는데 도움이 되는 방식으로 변환된 데이터입니다. **피처 엔지니어링**은 새 특성을 생성하거나 기존 특성을 수정하여 모델 성능과 예측력을 높이는 것을 목표로 합니다. 대부분의 ML 알고리즘에서는 기본적으로 수치 입력을 선호하므로 텍스트 기반 변수를 숫자 기반 변수로 조작하는 경우가 많습니다.

효과적인 피처 엔지니어링을 위해서는 해결하려는 데이터와 비즈니스 문제를 모두 이해해야 합니다. 피처 엔지니어링의 올바른 방법이 하나도 없는 경우가 종종 있습니다. 실험과 직관이 필요합니다.

DO NOT COPY
ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



피처 엔지니어링은 ML 모델 훈련에 사용할 정리된 데이터세트를 준비합니다. 특성은 카테고리 등의 텍스트 기반 항목일 수도 있고 수치일 수도 있습니다. ML 모델에서 사용하려면 범주형 데이터를 숫자 형식으로 인코딩해야 하는 경우가 많습니다. 수치 특성을 엔지니어링하여 모델 성능을 개선할 수도 있습니다. 특성을 선택하거나 결합하면 데이터세트의 패턴을 효율적으로 인식하는 ML 알고리즘의 능력을 더욱 개선할 수 있습니다.

피처 엔지니어링 기법을 적용하면 원시 데이터에서 더 많은 값을 추출할 수 있으며 잠재적으로 ML 모델의 정확도와 효과를 개선할 수 있습니다.

DO NOT COPY
ssleeeee1@gmail.com

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

텍스트 및 범주형 특성 추출

도전 과제:

- 카테고리 또는 레이블의 의미 표현
- 대량의 텍스트에서 의미 추출
- 의미론적 및 문맥적 관계 캡처

기법: 범주형 인코딩

17

텍스트 피처 엔지니어링은 ML 모델이 이해할 수 있으며 적절하게 균형이 유지된 정보를 추론하는 데 효율적으로 사용할 수 있는 형식으로 원시 텍스트 데이터를 변환합니다. 대부분의 원시 텍스트 데이터 유형은 텍스트 피처 엔지니어링의 이점을 누릴 수 있습니다.

텍스트 피처 엔지니어링 관련 도전 과제는 다음과 같습니다.

- **범주형 텍스트**: 카테고리나 레이블을 나타내는 텍스트 데이터 처리
- **대량 텍스트**: 긴 텍스트 구절에서 의미 처리 및 추출
- **복잡한 관계**: 텍스트 내의 의미론적, 문맥적 정보 캡처

텍스트 피처 엔지니어링 기법에는 다음과 같은 항목이 있습니다.

- **범주형 인코딩**: 범주형 텍스트 변수를 숫자 표현으로 변환
- **토큰화**: 텍스트를 더 작은 단위(단어, 보조 단어, 문자)로 나누기
- **백 오브 워즈**: 텍스트를 개별 단어 및 그 빈도 모음으로 표현
- **N-gram**: 문맥을 보존하기 위해 N-인접 단어 시퀀스 캡처

이러한 기법을 적용하면 ML 모델이 더욱 효율적으로 처리할 수 있는 형식으로 텍스트 데이터를 변환할 수 있습니다. 그러면 텍스트 기반 태스크에서 모델 성능이 잠재적으로 개선될 수도 있습니다. 이 모듈에서는 텍스트 특성 추출용 범주형 인코딩 기법을 주로 살펴봅니다.

DO NOT COPY ssleeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

범주형 인코딩은 범주형 텍스트 값을 ML 알고리즘이 해석할 수 있는 형식으로 변환하는 프로세스입니다.

각 유형의 범주형 데이터에는 해당하는 인코딩 기법이 있습니다.

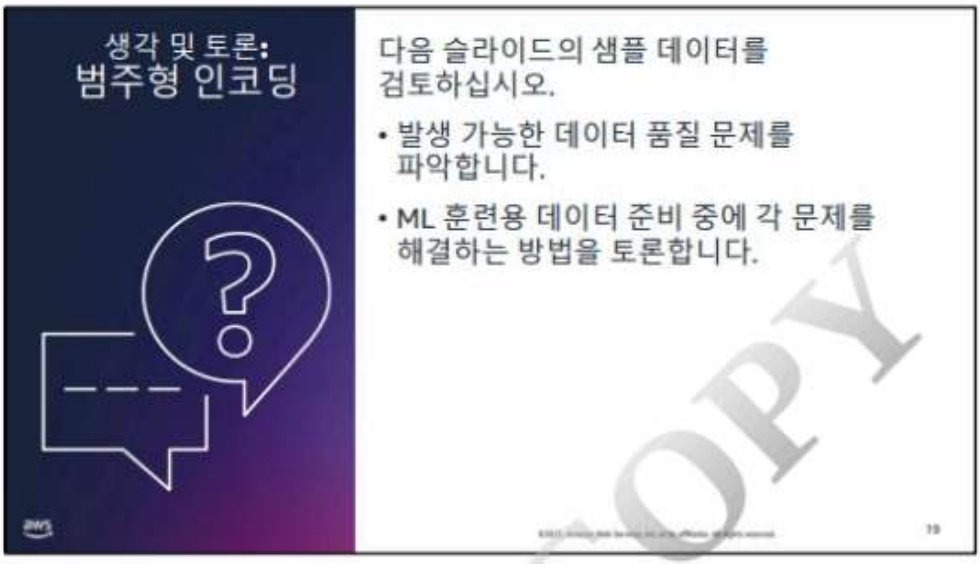
- **이진 데이터** – 참과 거짓, 예와 아니요 등의 가능한 값 2개를 사용하는 기법입니다. 이진 데이터는 대개 0과 1로 인코딩됩니다.
- **순서형** – 소, 중, 대 등 카테고리에 유의미한 순서가 있는 경우 사용하는 기법입니다. 이러한 데이터 유형은 서수 인코딩 또는 정수 인코딩을 사용하여 숫자로 인코딩할 수 있습니다. 단, 레이블 인코딩이 더 효율적인 경우도 있습니다(특히 카디널리티가 높은 특성). ML 알고리즘이 숫자 레이블을 관계로 해석할 수도 있습니다. 해당 상황에서는 숫자가 아닌 레이블이나 원-핫 인코딩을 사용해 보십시오.
- **명목형** – 색상이나 도시 등 기본적인 순서가 없는 여러 카테고리 값에 사용하는 기법입니다. 레이블 인코딩 기법을 사용하면 명목형 데이터를 단일 특성 내의 숫자로 인코딩할 수 있는데, 일부 알고리즘은 이 접근법에서 순서를 유추합니다. **원-핫 인코딩** 기법을 사용하면 각 카테고리 값에 해당하는 새 특성으로 인코딩할 수 있습니다. 이렇게 하면 정렬된 값으로 잘못 해석되는 상황을 방지할 수는 있지만 특성 수를 늘립니다, 특히 고유 카테고리 수가 많은 (카디널리티가 높은) 특성의 경우입니다.

ML 모델은 예측이나 분류에서 적절하게 인코딩 된 데이터를 더욱 효과적으로 사용할 수 있습니다. 인코딩 기법은 구체적인 알고리즘과 다음과 같은 범주형 변수의 특성에 따라 달라집니다.

- 인코딩이 불필요한 알고리즘도 있습니다. 예를 들어, 의사 결정 트리와 랜덤 포레스트는 범주형 변수를 직접 처리할 수 있습니다.
- 카디널리티가 높은 원-핫 인코딩 사용 시에는 모델 훈련 효율성이 떨어질 수 있습니다.
- 인코딩 값이 모델 해석 가능성에 영향을 줄 수도 있습니다.

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



DO NOT COPY
ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

슬라이드의 표에는 고객 데이터세트의 샘플 데이터(인구 통계 데이터 및 구독 정보 포함)가 나와 있습니다.

데이터세트에서 강조 표시되어 있는 각 열의 인코딩 유형을 파악합니다.
- 연령
- 학력
- 주
- 활성 구독 여부
- 구독 유형

인코딩해야 하는 각 열에서 다음 항목을 고려합니다.
- 범주형 데이터 유형(이진, 명목, 서수)
- 가장 적절한 인코딩 기법(예: 원-핫 인코딩, 레이블, 서수)
- 선택한 ML 알고리즘의 요구 사항

ML 모델이 효과적으로 처리할 수 있는 형식으로 데이터를 변환하고 각 특성의 내재적인 정보는 유지해야 합니다.

DO NOT COPY ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

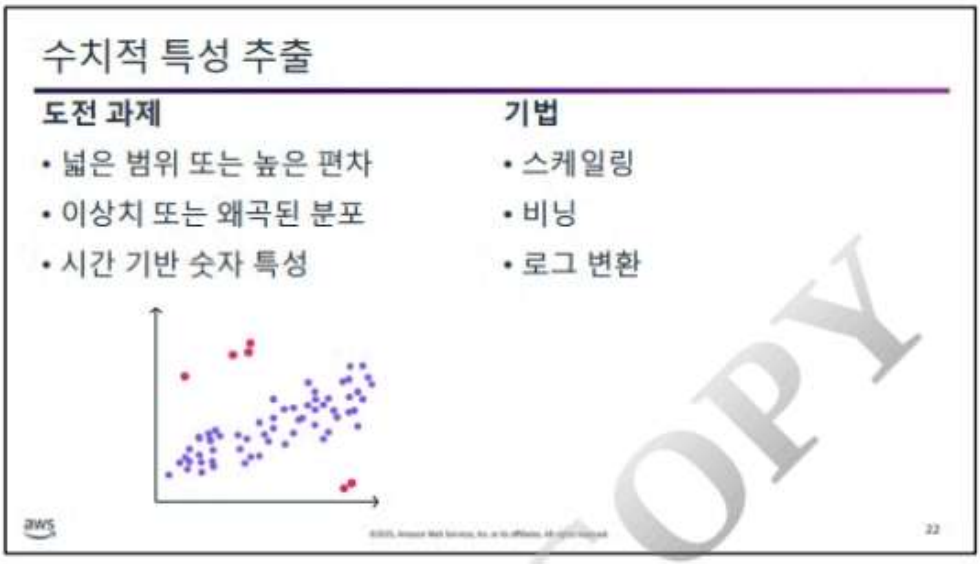
각 열의 인코딩 결정을 검토하고 그 의미를 생각해 봅니다:

- **연령** - 인코딩할 필요가 없습니다. 수치형 데이터는 이미 대다수 ML 알고리즘에서 적합합니다.
- **학력** - 순서형 인코딩에서는 카테고리의 고유 순서를 반영하는 증가하는 정수 값이 할당됩니다. 대안인, 원-핫 인코딩으로는 각 학력 수준에 해당하는 별도의 이진 열을 만들어 모든 정보가 유지되기는 하지만 차원은 증가할 수 있습니다. 인코딩 기법의 선택은 선택한 알고리즘 및 데이터의 특성과 의미에 따라 달라집니다.
- **거주지** - 명목형 인코딩 기법이 사용되었습니다. 이 데이터에는 학력과 비슷한 고려 사항이 있습니다. 거주지의 순서가 지정되는 상황을 방지하려는 경우 원-핫 인코딩이 더 적합할 수도 있습니다. 레이블 인코딩에서는 각 수준에 고유한 정수가 할당되는데, 알고리즘은 질서 있는 관계를 갖는 걸로 해석할 수도 있습니다.
- **활성 구독 여부** - 이진 인코딩 기법이 사용되었습니다. 예 또는 아니요 값을 0이나 1로만 인코딩해도 모든 정보가 유지되며, 대다수 알고리즘은 해당 값을 이해할 수 있습니다.
- **구독 유형** – 명목형 인코딩 기법이 사용되었습니다. 사용 사례와 데이터 특성에 따라 달라집니다. 카디널리티가 높은 특성에는 레이블 인코딩이 더 효율적일 수도 있습니다. 단, 각 유형에 고유한 정수가 할당되므로 ML 알고리즘이 이를 서수 관계로 잘못 해석할 수도 있습니다. 원-핫 인코딩에서는 각 구독 유형에 해당하는 개별 이진 열이 생성됩니다. 그러므로 레이블 인코딩 값들 간의 관계가 잘못 해석되는 상황은 방지할 수 있지만 차원은 증가하게 됩니다.

ML 알고리즘은 이제 학력 및 활성 구독 여부 특성 내의 값 간 관계가 반영된 수치 데이터를 사용하여 전체 데이터세트의 패턴을 식별할 수 있습니다. 인코딩 기법의 선택은 모델의 성능과 해석 가능성에 상당한 영향을 미칠 수 있다는 걸 기억하십시오. 그러므로 인코딩을 결정할 때는 항상 데이터의 특성 및 선택한 ML 알고리즘의 요구 사항을 고려해야 합니다.

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



수치 특성 엔지니어링에서는 ML 모델에 쓸 수치형 데이터를 준비합니다. 모델 성능을 개선하고 다양한 데이터 문제를 처리하기 위해서 수치 값을 변환하는 작업이 포함됩니다.

일반적인 도전 과제는 다음과 같습니다.

- **넓은 범위 또는 높은 편차**: 척도가 크게 다른 특성들이 모델 훈련에서 크게 작용될 수 있습니다.
- **이상치 또는 왜곡된 분포 발생 가능성**: 극단적인 값이 있으면 모델 동작에 불균형적으로 영향을 미칠 수 있습니다.
- **시간 기반 숫자 특성**: 날짜 및 시간은 특수하게 취급해야 하는 경우가 많습니다.

수치 특성 추출용 기법에는 다음이 포함됩니다.

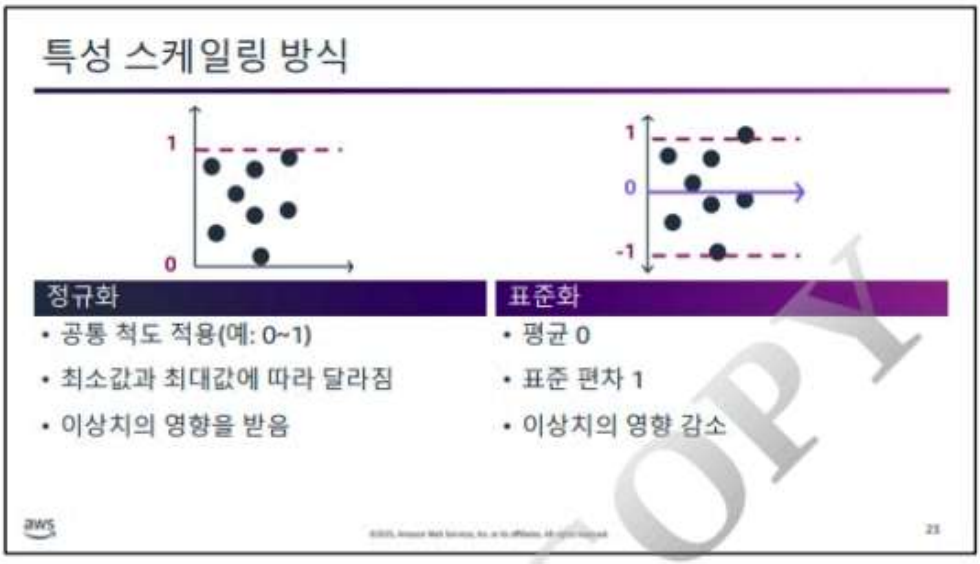
- **스케일링**: 특성 값을 일반 척도로 조정합니다.(예: 0~1 또는 -1~1)
- **비닝**: 연속 값을 개별 카테고리로 그룹화하여 비선형 관계를 캡처하고 이상치의 영향을 줄입니다.
- **로그 변환**: 로그 함수를 적용하여 광범위한 값을 압축하고 오른쪽으로 왜곡된 분포에서 이상치의 영향을 줄입니다.

이러한 기법을 사용하면 숫자 특성을 ML 알고리즘이 효과적으로 처리할 수 있는 형식으로 변환하여 잠재적으로 모델 성능을 개선되고 문제가 있는 데이터 특성의 영향을 줄여줍니다.

DO NOT COPY ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



대다수 기계 학습 알고리즘, 특히 입력 특성의 규모에 민감한 알고리즘에는, 특성 스케일링을 수행해야 합니다. 일반적인 스케일링 접근 방식 2가지는 **정규화** 및 **표준화**입니다.

정규화:
- 값의 규모를 고정 범위(대개 0~1)로 조정합니다.
- (x - min(x)) / (max(x) - min(x)) 공식을 사용합니다.
- 희소 데이터의 0 값이 보존됩니다.
- 값 경계를 지정해야 할 때 유용합니다.
- **이상치 민감도 높음**: 극단적인 값이 있으면 다른 값의 범위가 압축될 수 있습니다.

표준화:
- 데이터를 변환하여 평균을 0으로, 표준 편차를 1로 조정합니다.
- (x - mean(x)) / std(x) 공식을 사용합니다.
- 데이터는 0을 중심으로 배치됩니다(데이터 위치가 중요한 알고리즘도 있음).
- 정규화에 비해 이상치의 영향 정도가 더 적습니다.
- 값 경계를 지정할 필요가 없을 때 유용합니다.

기법 선택 방법:
- 정규화를 사용하여 이상치가 없는 데이터를 0~1 사이의 경계가 있는 값으로 스케일링 합니다.
- 표준화를 사용하여 이상치의 영향을 줄이고 데이터의 중심을 0으로 잡습니다.
- 특정 ML 알고리즘의 요구 사항을 고려합니다. 트리 기반 알고리즘 같은 일부 알고리즘은 스케일링이 전혀 필요하지 않습니다.

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

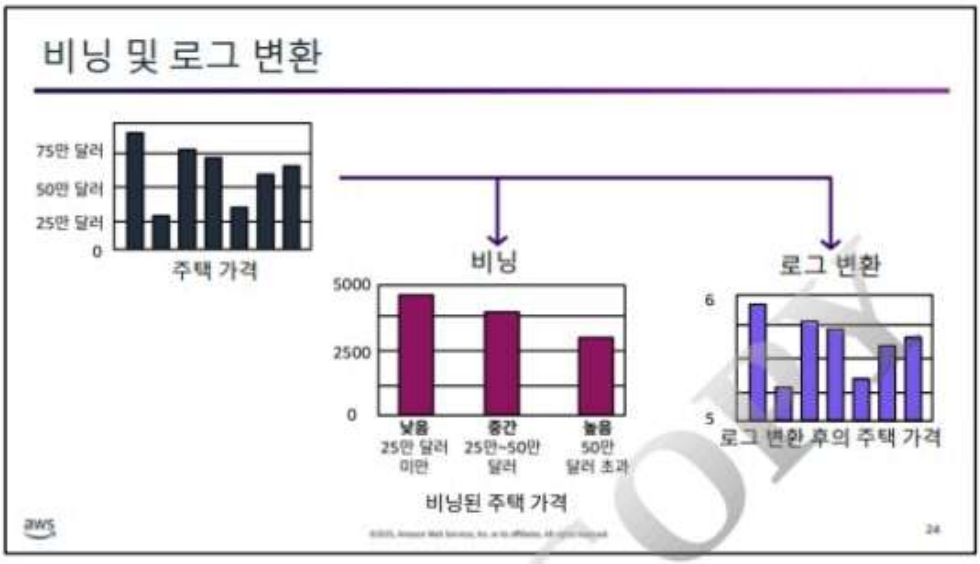
Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

훈련 및 테스트 데이터세트를 일관되게 스케일링하면 특성 간의 관계를 유지하고 데이터 유출을 방지하는 데 도움이 됩니다.

DO NOT COPY
ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



2가지 추가적인 수치 피처 엔지니어링 기법은 이상치의 영향을 줄일 수 있습니다.

비닝

비닝 기능은 이상치의 영향을 줄여 모델의 더 강력하게 만들 수 있습니다. 비닝은 또한 특성과 목표 값 간의 비선형 관계를 더 쉽게 캡처하는 데 도움이 됩니다. 정확한 숫자 차이는 중요하지 않지만 일반적인 범위가 요인인 경우 수치형 데이터에 효과적인 전략입니다.

예를 들어, 주택 데이터세트에서 주택 가격이 150,000~151,000달러인 데이터 포인트는 구분할 필요가 없지만 가격이 150,000달러인 주택과 500,000달러인 주택은 서로 다른 특성이 있습니다. 비닝을 사용하면 이러한 고유 값 범위 관련 정보는 그대로 유지하면서 특성을 단순화합니다. 하지만 포함되는 데이터 포인트의 수는 빈별로 크게 다를 수도 있습니다.

사분위수 비닝 기법을 사용하면 빈 간에 값을 균등 분배할 수 있습니다. 예를 들어, 주택 데이터를 3개 카테고리로 비닝하면 최하위 33.3%에 해당하는 데이터, 다음 33.3%에 해당하는 데이터, 최상위 33.3%에 해당하는 데이터가 각기 다른 빈에 분배됩니다.

로그 변환

로그 함수는 지수 함수의 역함수입니다. 인구 증가나 감소 등 시간의 경과에 따라 증가율이 낮아지는 현상을 모델링할 때 유용합니다. 특성 로그 변환에서는 로그 함수를 사용해 왜곡된 수치 데이터를 정규화하고, 이상치의 영향을 줄이고, 값을 정규 분포에 가깝게

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

유지합니다. 가장 일반적인 로그 함수의 밑수는 10 또는 e이며, 여기서 e는 2.71828과 거의 같습니다. 분포가 심하게 치우치거나, 이상치가 있는 수치 특성 사용 시 로그 변환을 수행하면 모델 성능을 개선할 수 있습니다.

예를 들어, 주택 데이터세트에 지리적 위치가 여러 개 포함되어 있어 이상치가 매우 높다고 가정해 보겠습니다. 한 지역에는 시골 부지가 10,000달러에 등록되어 있고 다른 지역에는 도시의 고급 주택이 10,000,000달러에 등록되어 있습니다. 주택 가격의 로그를 사용하면 큰 이상값이 줄어듭니다. 예를 들어 10,000달러의 로그는 약 4이고 10,000,000달러의 로그는 약 7이 될 것입니다. 이 기법을 사용하면 이상치가 나머지 데이터세트의 정상 값에 훨씬 더 가까워집니다.

DO NOT COPY
ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

예: 수치적 피처 엔지니어링

가격(달러)	정규화	표준화	비닝	로그 변환
최소값: 100,000	0.1	-1.15	낮음	5
중앙값: 437,000	0.437	-0.289	중간	5.64
최대값: 1,000,000	1	1.15	높음	6

25

예를 들어, 주택 데이터를 고려한 데이터세트에는 침실 수, 주택 가격 등의 특성이 포함될 수 있습니다. 침실 수는 숫자 값으로 1~5개일 수 있지만 가격은 100,000달러에서 1,000,000달러 사이일 수 있습니다.

- 값이 100,000에서 1,000,000 사이인 주택 가격 특성을 **정규화**하면 값이 0~1 범위로 변환됩니다. 이 기법은 가장 작은 값을 0에 할당하고 가장 큰 값을 1에 할당하고 계산을 사용하여 나머지 값을 스케일 다운합니다.
- **표준화**에서는 평균 주택 가격을 정수 0으로 설정하고 계산을 사용하여 표준 편차가 1이 되도록 나머지 값에 점수를 매깁니다. 대다수 값은 -1~1 범위 내에 있습니다. 계산을 통해 값이 매끄러워지므로 표준화를 사용하면 정규화를 사용할 때만큼 이상값이 부정적인 요소가 되지 않습니다.
- **비닝**에서는 숫자 값을 낮음, 보통, 높음 등의 텍스트 카테고리로 변환합니다.
- **로그 변환**에서는 평균과 이상치 간의 '거리'를 단축합니다. 이 시나리오에서는 모든 데이터 포인트가 5~6 사이의 범위로 변환되므로 이상치의 영향이 감소합니다.

이러한 수치 피처 엔지니어링 기법 중 사용할 항목을 선택할 때는 사용 사례, 이상치 또는 왜곡 포함 여부와 같은 데이터 분포의 특성 및 ML 알고리즘의 기능을 고려해야 합니다.

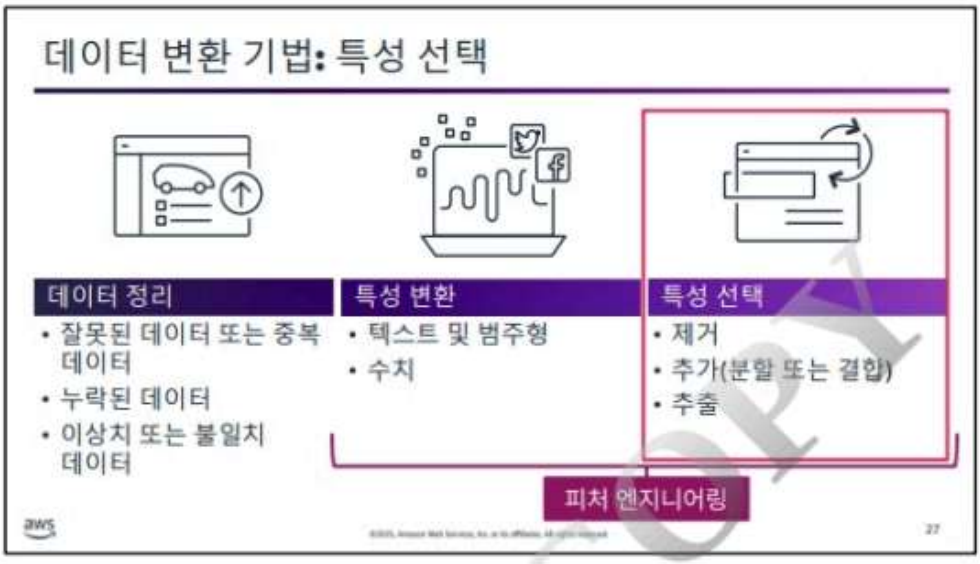
© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

특성 선택은 ML 모델에서 사용할 가장 관련성이 높은 특성을 선택하는 프로세스입니다. 즉, 모델 성능을 개선하고 훈련 시간을 줄일 수 있는 데이터 준비 과정의 중요한 단계입니다.

DO NOT COPY
ssleeeee1@gmail.com

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



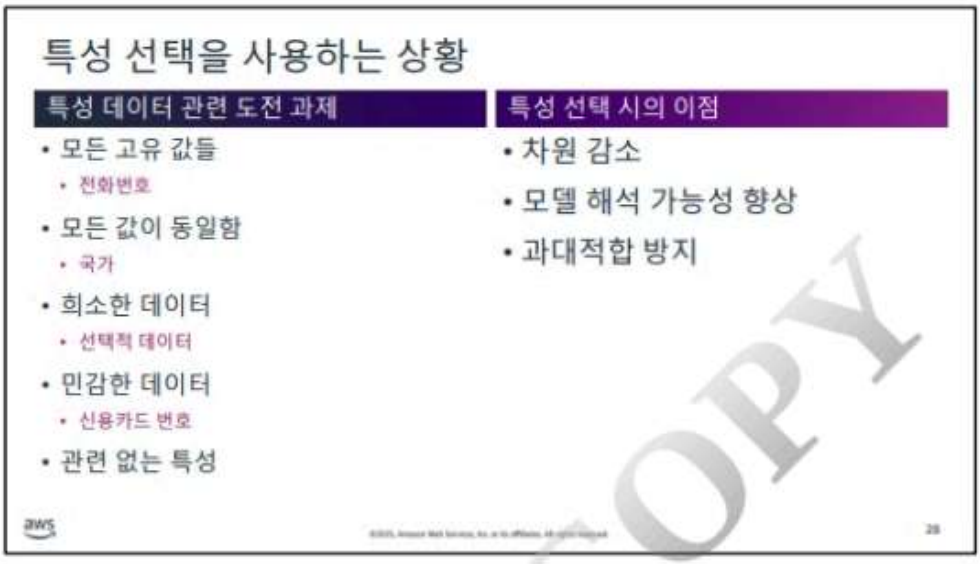
지금까지, 데이터를 정리한 후 ML 알고리즘이 더욱 정확한 예측을 하는 데 사용할 수 있는 형식으로 특성으로 변환하였습니다. 피처 엔지니어링의 다음 단계는 특성 선택입니다. 특성 선택은 훈련된 모델에서 오류를 줄이기 위해 가장 관련성이 높은 특성 하위 집합을 선택하는 프로세스입니다.

이 프로세스에서는 다음과 같은 기법이 포함될 수 있습니다:
- 모델 정확도에 큰 영향을 주지 않는 특성 제거
- 기존 특성을 분할하거나 결합하여 새 특성 추가
- 일부 하위 집합에서 모델 정확도에 영향을 주는 원래 데이터 특성 추출

특성 선택 시에는 다음과 같은 기법이 포함됩니다.
- 모델 단순화와 관련이 없거나 중복된 특성 제거
- 기존 특성을 다르게 표현하는 새 특성 추가
  - 특성을 여러 파생 특성으로 분할하여 더 세부적인 정보 캡처
  - 특성을 결합하여 관련 정보 집계
- **주성분 분석(PCA)** 등의 기법을 사용해 여러 특성에서 데이터 추출하여 데이터의 변동을 대부분 유지하면서 차원을 감소

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



어떤 특성은 ML 모델 훈련에 필요한 리소스의 수를 불필요하게 늘리거나 ML 모델의 유용성에 영향을 미칩니다. 각 값이 고유한 전화번호 특성이나 한 국가 내 사용자의 데이터세트에 포함된 국가 특성을 예로 들 수 있습니다. 모델 훈련은 이러한 특성에서 유용한 패턴을 찾을 것 같지 않습니다. 해당 특성은 제거할 수 있습니다. 마찬가지로 설문 조사의 선택적 데이터 필드와 같이 누락된 값이 많은 특성도 모델 훈련에서 별로 유용하지 않습니다. 개인 식별자 또는 신용 카드 번호와 같은 민감한 원시 데이터를 포함하는 경우에는 모델 출력에서 민감한 데이터가 유출될 위험이 있습니다. 모델 학습에 더 의미 있게 이러한 예제와 같은 기능이나 ML 문제와 관련 없는 다른 특성을 제거하거나 수정할 수 있습니다.

특성 선택은 ML 모델에 다음과 같은 다양한 주요 이점을 제공합니다:

- 특성을 제거하거나 결합하여 차원을 줄여서 훈련 시간과 계산 비용을 줄일 수 있습니다. 이것은 고차원 데이터세트를 다룰 때 특히 중요합니다.
- 관련성이 가장 높은 특성만 선택하여 모델 해석 가능성을 높일 수 있습니다. 많은 비즈니스 애플리케이션에서는 모델 결정 사항의 타당성을 입증해야 합니다. 특성이 적고, 더 의미 있는 모델일 수록 이해하고 설명하기가 더 쉬울 때가 많습니다.
- 특성 선택은 데이터세트의 노이즈를 줄여서 모델이 보이지 않는 데이터도 더욱 정확하게 일반화할 수 있습니다. 가장 유익한 특성에 초점을 맞추면 모델이 훈련 데이터 전용이지만 잘 일반화되지 않는 패턴을 덜 학습하게 됩니다. 이러한 상황을 **과대적합**이라고 합니다.

특성을 효과적으로 선택하면 모델의 초점을 데이터의 가장 중요한 측면에 집중시켜 정확도가 높아지고 과대적합 현상을 줄일 수 있습니다.

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

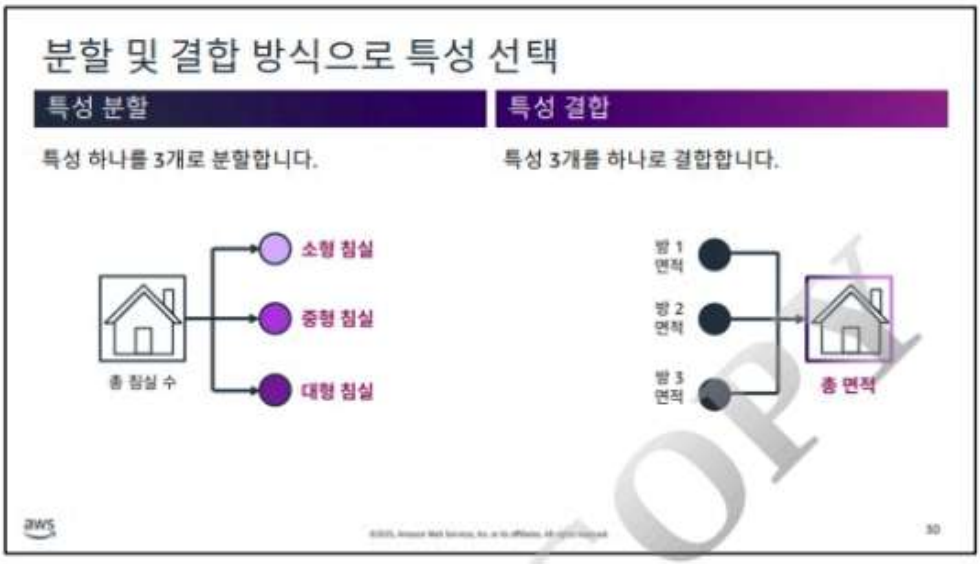
Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

예를 들어, 주택 가격 예측 모델에서는 각 주택과 관련된 여러 특성들로 시작할 수 있습니다. 이 예시에서는 특성 선택을 통해서 가격과 침실 수가 가장 예측 가능성이 높은 특성으로 확인되었습니다. 훈련 데이터세트에서 예측성이 떨어지는 특성을 제거하면 더욱 단순하면서도 안정적인 모델을 완성할 수 있습니다. 가장 예측 가능한 특성은 사용 사례 및 데이터세트의 특성에 따라 달라질 수 있습니다.

DO NOT COPY
ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



특성 분할 및 결합은 모델 성능을 대폭 높일 수 있는 특성 추출의 유용한 기법입니다.

특성 분할에서는 특성 하나를 여러 파생 특성으로 구분합니다. 특성 하나에 정보가 여러 개 포함되어 있을 때 유용할 수 있습니다. 예를 들어 '날짜' 특성을 '년', '월', '일' 특성으로 분할하면 모델이 특정 시간 구성 요소와 관련된 패턴을 학습할 수 있습니다.

특성 결합에서는 여러 특성을 결합된 새 특성으로 집계합니다. 이는 여러 특성이 상호 관련되어 있으며 그 결합한 정보가 더 의미 있을 때 유용합니다. 예를 들어 주택 가격 예측 모델에서 개별 방 크기를 '총 면적' 특성으로 결합하는 경우를 예로 들 수 있습니다.

이러한 두 기법을 활용하면 모델이 데이터에서 더 미묘한 패턴을 포착하는 데 도움이 될 수 있습니다. 하지만, 과도하게 분할하거나 결합하면 과대적합이 발생하거나 중요한 정보가 손실될 수도 있으므로 신중하게 사용해야 합니다. 데이터와 해결하려는 문제에 따라 적합한 기법을 선택하면 됩니다.

DO NOT COPY
ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

주성분 분석(PCA)

- 특성 수가 감소합니다.
- 특성을 주성분으로 변환합니다.
- 원래 정보가 대다수 유지됩니다.
- 분산을 기준으로 성분 순위가 지정됩니다.

효율적인 ML 워크플로우와 추론을 위해서 데이터 과학자는 데이터세트의 가장 큰 변동을 나타내는 가장 수가 적은 특성을 찾는 경우가 많습니다. 주성분 분석(PCA)은 기계 학습의 **차원 축소를** 위한 유용한 기법입니다. PCA에서는 원래 특성을 **주성분**이라는 상관관계가 없는 특성인 신규 집합으로 바꾸는 방식으로 작동합니다.

PCA의 핵심 아이디어는 **데이터가 가장 많이 변하는 방향(주성분)을 식별**하는 것입니다. 이러한 주성분은 데이터에서 설명하는 분산의 양에 따라 정렬됩니다. 첫 번째 주성분이 가장 많은 분산을 설명하고, 두번째 주성분이 다음으로 많은 분산을 설명하는 등의 방식입니다.

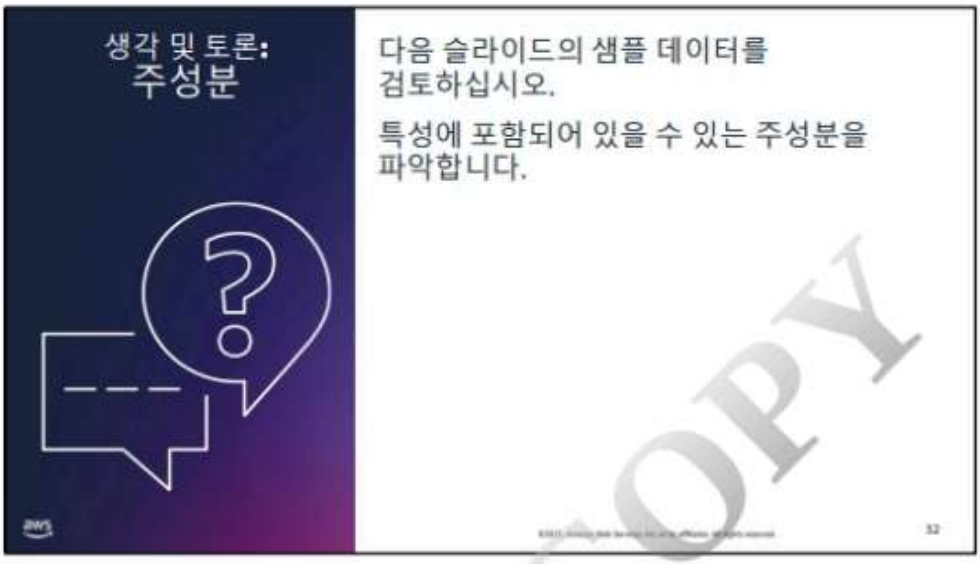
PCA에서는 여러 가지 이점을 제공합니다:

1. 원본 정보는 대부분 유지하면서, **특성 수를 줄이므로** 모델 훈련 속도가 빨라질 수 있습니다.
2. 고차원 데이터를 저차원 공간에 표시하여 **시각화하는 데 도움**이 될 수 있습니다.
3. 일부 알고리즘에서 문제를 야기할 수 있는 특성 간의 **다중 공선성을 줄일** 수 있습니다.

PCA 사용 시에는 먼저 특성을 표준화해야 합니다. 입력 특성의 규모에 따라 PCA 결과가 달라질 수도 있기 때문입니다. 그리고 PCA 진행 시의 이점과 결과 특성 해석 가능성 간의 적절한 절충점을 고려해야 합니다.

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



DO NOT COPY
ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

이 활동에서는, 주택 가격을 예측하도록 ML 모델을 훈련시킵니다. 샘플 교육 데이터세트를 검토한 후 가장 중요한 정보는 유지하면서 PCA를 적용하여 이 데이터세트를 단순화할 수 있는 방법을 고려해 보십시오.

원래 특성은 다음과 같습니다.
- 면적
- 침실 수
- 욕실 수
- 가격
- 세율
- 지역

이 데이터세트에 PCA를 적용할 때, 데이터의 가장 큰 변동을 포착하는 주성분으로 이러한 특성 중 일부를 결합하는 방식을 파악해야 합니다. 모델은 크기에 따라 주택 가격을 예측하므로, 어떤 특성이 크기와 관련이 있고 어떤 특성이 가격과 관련이 있는지 생각해야 합니다.

DO NOT COPY
ssleeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

부동산 번호 및 국가 열은 삭제할 수 있습니다.
- 모든 부동산 번호 값은 고유 값이며 가격이나 크기에 대한 정보를 제공하지 않습니다.
- 모든 국가 값은 동일하며 모델에 어떤 정보도 제공하지 않습니다.

나머지 열 내에서 도출할 수 있는 주성분은 크기와 비용입니다.

주성분 1: 크기
다음과 같은 주택의 물리적 속성이 결합된 주성분입니다.
- 면적
- 침실 수
- 욕실 수

주성분 2: 비용
주택의 재정적 특성을 캡처하는 주성분입니다.
- 가격
- 세율
- 지역

DO NOT COPY
ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

PCA는 원래 특성 5개를 주성분 2개로 줄임으로써 데이터의 가장 중요한 정보는 그대로 캡처하면서 데이터 세트를 단순화합니다. 데이터의 변동을 가장 잘 설명하는 특성 조합을 찾는 것입니다. 이렇게 축소된 차원은 다음과 같은 여러 이점이 제공될 수 있습니다.

- 모델 성능 향상: 특성이 적을수록 모델 훈련 속도가 빨라지며 과대적합 가능성은 낮아집니다.
- 더욱 쉬운 시각화: 이제 2개 차원으로 데이터를 표시할 수 있으므로 패턴을 더 쉽게 시각화할 수 있습니다.
- 다중 공선성 감소: 관련 특성을 결합하면 입력 변수 간에 설정될 수 있는 상관관계의 수가 감소합니다.

이 기법을 사용하면 후속 분석 및 모델링을 더 효율적으로 만들 수 있습니다. 정확한 주성분은 특성 이름을 기반으로 하지 않고, 실제 데이터 값의 변동을 최대화해서 수학적으로 결정됩니다. ML 실무자는 각 주성분이 나타내는 데이터를 파악할 수 있어야 합니다.

DO NOT COPY
ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

데이터 수집은 데이터베이스, 데이터 레이크, 스트리밍 소스와 같은 다양한 소스에서 데이터를 수집하는 프로세스입니다. 데이터 통합의 목표는 기계 학습 솔루션에 사용할 데이터를 통합하는 것입니다.

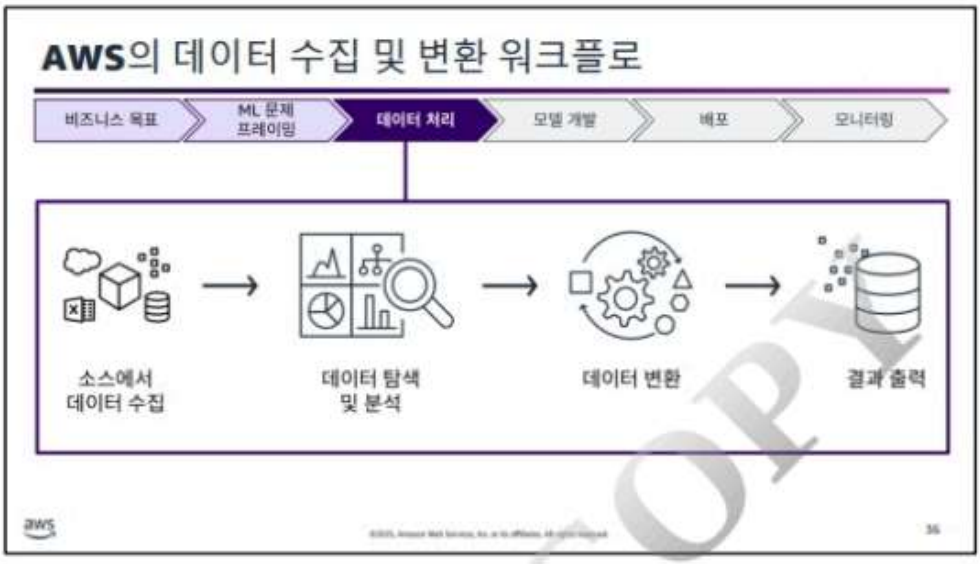
이 섹션에서는 Amazon SageMaker Data Wrangler와 같은 AWS 서비스를 사용하여 데이터를 수집하는 방법을 배웁니다. 구체적으로는 Amazon SageMaker Feature Store를 사용하여 데이터를 수집하고 특성을 생성 및 관리하는 방법을 배우게 됩니다.

그리고 이러한 AWS 서비스의 데이터 변환 사용 사례에 대해 알아봅니다. 또한 이러한 서비스를 사용하여 데이터 변환 프로세스의 일부 또는 전체를 자동화하는 방법에 대한 기본 단계도 알아봅니다.

DO NOT COPY ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



AWS에서 기계 학습용 데이터를 수집 및 변환하는 프로세스에서는 대개 다음과 같은 단계를 수행합니다.

1. **소스에서 데이터 수집**. Amazon Simple Storage Service(Amazon S3), 데이터베이스 또는 스트리밍 서비스와 같은 소스에서 데이터를 가져올 수 있습니다. AWS Glue 등의 AWS 서비스를 사용하면 데이터 카탈로그 작성과 검색을 수행할 수 있습니다.
2. **데이터 탐색 및 분석**. SageMaker Data Wrangler 등의 도구에서는 데이터를 더 정확하게 파악할 수 있는 데이터 품질 보고서와 시각화가 제공됩니다. Amazon Athena를 사용하여 데이터세트를 쿼리하거나 Amazon QuickSight를 사용하여 데이터를 시각화할 수도 있습니다.
3. **변환 정의**. 이 단계에서는 데이터를 정리 또는 정규화 하거나 피처 엔지니어링 방법을 지정합니다. SageMaker Data Wrangler에서는 코드를 작성하지 않고도 이러한 변환을 정의할 수 있는 시각적 인터페이스가 제공됩니다. SageMaker Processing은 사용자 정의 변환 코드가 포함된 스크립트 또는 Notebook을 실행합니다. Amazon EMR 같은 분석 서비스는 빅 데이터 변환과 분석과 ML 워크로드 간의 데이터 파이프라인 통합에 효과적입니다.
4. **변환 미리 보기 및 적용**. 전체 데이터세트에 변환을 적용하기 전에 샘플에서 결과를 미리 볼 수 있습니다. 그러면 프로세스 초반에 문제를 파악할 수 있습니다. SageMaker Data Wrangler 등의 도구에서는 이 프로세스가 기본적으로 지원됩니다. 다른 도구에서 스크립트 또는 Notebook을 실행할 때 데이터세트 샘플을 대상으로 변환을 실행할 수도 있습니다.
5. **결과 출력**. 변환된 데이터는 추가 처리 또는 모델 훈련에 사용할 수 있도록 Amazon S3 등의 대상으로 내보낼 수도 있고 특성 관리를 위해 SageMaker Feature Store로 내보낼 수도 있습니다.

이 프로세스는 반복 수행해야 하며 탐색 및 분석 결과에 따라 변환을 미세 조정해야 할

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

수도 있습니다. AWS 서비스를 활용하면 데이터 준비 워크플로를 유동적으로 수정하여 다시 실행할 수 있습니다.

DO NOT COPY
ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

AWS 서비스에는 기계 학습 및 분석 워크플로를 지원하기 위해 데이터를 변환하는 데 사용할 수 있는 목적별 도구가 포함되어 있습니다.

Amazon SageMaker AI에서 제공되는 통합 데이터 변환 기능을 사용하면 복잡한 데이터 수집 파이프라인을 간편하게 진행할 수 있습니다.

- **SageMaker Data Wrangler**에서는 시각적인 노코드 방식으로 여러 소스의 데이터를 정리 및 변환할 수 있습니다.
- **SageMaker Processing**에서는 스크립트나 Notebook을 실행하여 ML 훈련 준비를 위해 데이터세트를 처리, 변환 또는 분석할 수 있습니다.
- **SageMaker Feature Store**에서는 ML 모델 개발용으로 큐레이팅된 특성을 생성, 관리 및 공유할 수 있습니다.

AWS는 다음과 같이 포괄적인 분석 워크로드용 기능을 제공합니다.

- **Amazon EMR**에서는 Apache Spark, Apache Hive 등의 프레임워크를 사용하여 확장 가능 빅 데이터 워크로드를 실행할 수 있습니다.
- **AWS Glue**는 서버리스 데이터 통합 서비스입니다. **AWS Glue Data Brew 및** AWS Glue Studio는 대화형 데이터 준비를 지원합니다.
- **Amazon SageMaker Unified Studio**는 차세대 SageMaker AI의 기능입니다. 분석 및 ML 워크플로 전반의 기능을 합쳐진 데이터 액세스와 통합합니다. 여기에는 Amazon Athena, Amazon EMR, AWS Glue, Amazon Managed Workflows for Apache Airflow(Amazon MWAA)와 같은 서비스의 기능을 사용하여 데이터 워크플로를 분석, 준비 및

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

오케스트레이션하는 **SageMaker Data Processing**이 포함되어 있습니다.

적절한 서비스를 선택하는 것은 구체적인 요구 사항, 데이터의 양 및 선호하는 워크플로우에 따라 달라집니다. SageMaker Data Wrangler는 대화형 탐색에 적합하고, Amazon EMR은 빅 데이터 처리에, SageMaker Processing은 데이터 준비를 ML 파이프라인에 통합하려는 경우에 적합합니다. 이러한 서비스들을 조합하여 사용하면, 데이터를 수집하고 랭글링할 때 많이 하게 되는 수동 코드 작성 작업을 상당량 줄일 수 있습니다. 이러한 서비스는 서로 연동되어 원활한 데이터 준비 파이프라인을 제공합니다. 예를 들어 초기 대규모 데이터 처리에는 Amazon EMR을, 피처 엔지니어링에는 SageMaker Data Wrangler를 사용하고 마지막으로 SageMaker Feature Store에 특성을 저장할 수 있습니다.

ML 데이터 준비 시에 AWS 서비스를 사용하는 경우의 이점은 다음과 같습니다.

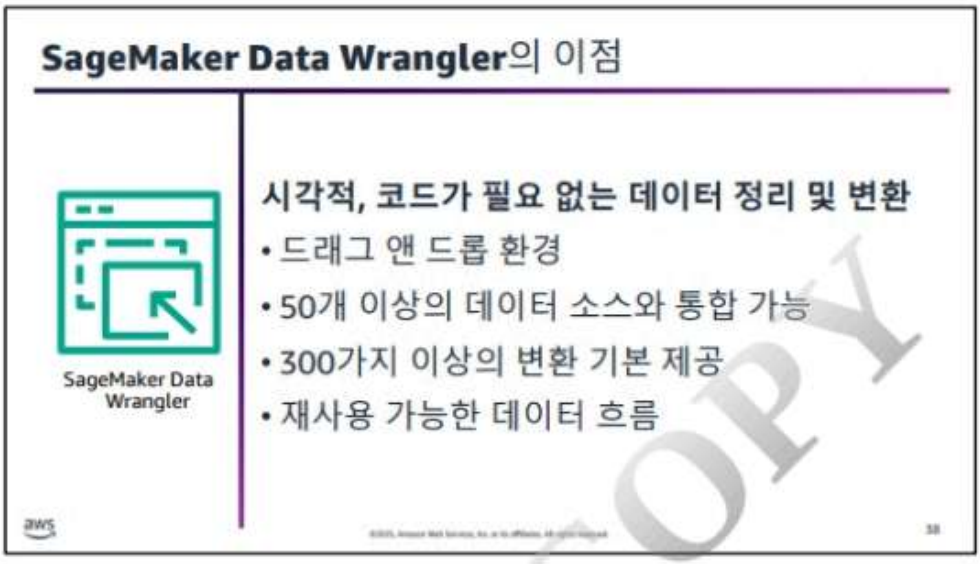
- **확장성**: 대용량 데이터세트 처리
- **통합**: 원활한 워크플로를 위한 서비스 간
- **관리형 인프라**: 운영 오버헤드 감소
- **기본 제공 모범 사례**: 데이터 보안 및 거버넌스 관련

기계 학습 엔지니어가 이러한 AWS 서비스를 사용하면 인프라를 관리하는 작업 대신 데이터에서 인사이트를 도출하는 작업에 집중할 수 있습니다.

DO NOT COPY
ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



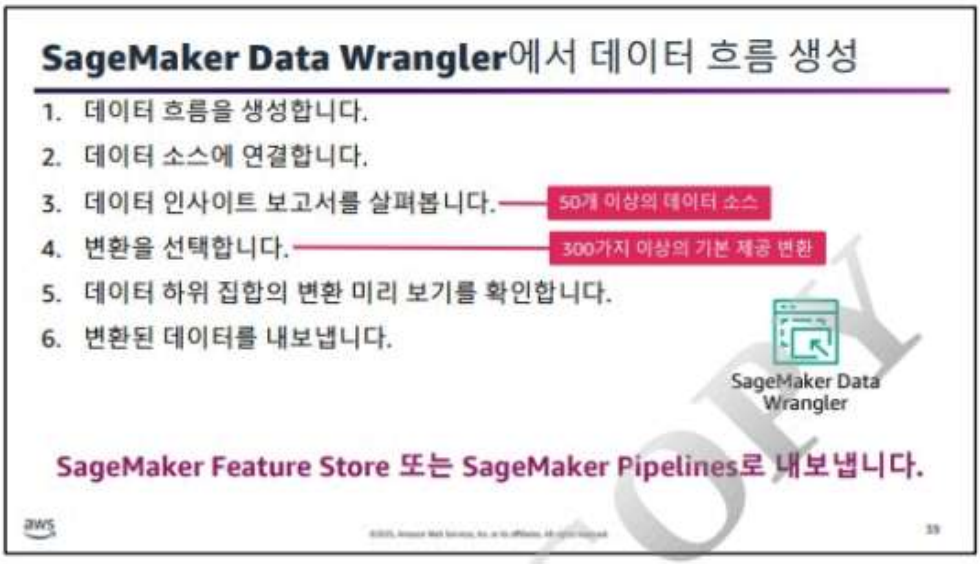
SageMaker Data Wrangler는 기계 학습용 데이터 준비 프로세스를 간소화합니다. 데이터 과학자와 분석가가 데이터를 프로파일링, 정리 및 변환하는 데 사용할 수 있는 **코드가 필요 없는 시각적 인터페이스**를 제공합니다. 예를 들어 SageMaker Data Wrangler는 자동 데이터 품질 검사 기능을 제공하며 데이터세트 관련 통계를 생성합니다. 그러므로 데이터를 신속하게 파악하고 발생 가능한 문제를 확인할 수 있습니다.

- **드래그 앤 드롭 환경**이 제공되므로 코드를 작성하지 않아도 데이터 변환 워크플로를 구축할 수 있습니다. 이렇게 하면 데이터 준비 프로세스를 크게 빨라지고, 프로그래밍에 익숙하지 않은 팀원들이 더 쉽게 액세스할 수 있습니다.
- SageMaker Data Wrangler는 Amazon S3, Amazon Redshift, Amazon Athena를 비롯한 **50가지 이상의 AWS 데이터 소스**에 직접 연결할 수 있습니다. 그러므로 데이터 수집 프로세스를 간편하게 진행할 수 있습니다.
- SageMaker Data Wrangler에서는 누락된 값 처리, 범주형 변수 인코딩, 특성 스케일링 등 일반적인 데이터 준비 태스크에 사용할 수 있는 **300개 이상의 알고리즘이 기본 제공**됩니다. 따라서 흔히 수행하는 데이터 변환에 해당하는 사용자 정의 코드를 작성해야 하는 경우가 감소하며 모범 사례도 더욱 철저하게 준수할 수 있습니다.
- SageMaker Data Wrangler에서는 **재사용 가능 데이터 흐름**을 생성할 수 있으므로 여러 데이터세트나 프로젝트에 일관성 있는 변환을 적용할 수 있습니다. 그러므로 비슷한 데이터세트 사용 시 일관성은 높아지고 시간은 절약됩니다.

SageMaker Data Wrangler는 데이터 정리, 피처 엔지니어링, 데이터세트 크기 축소, 자동화를 비롯한 일반적인 데이터 처리 태스크를 지원합니다. SageMaker Data Wrangler는 ML 모델용 데이터를 빠르게 탐색하고 준비해야 하는 경우, 특히 여러 소스의 구조화된 데이터를 사용할 때 특히 유용합니다. 반복적인 데이터 준비를 위한 유용한 도구로, 다양한 변환을 빠르게 시도하고 그 효과를 볼 수 있습니다.

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



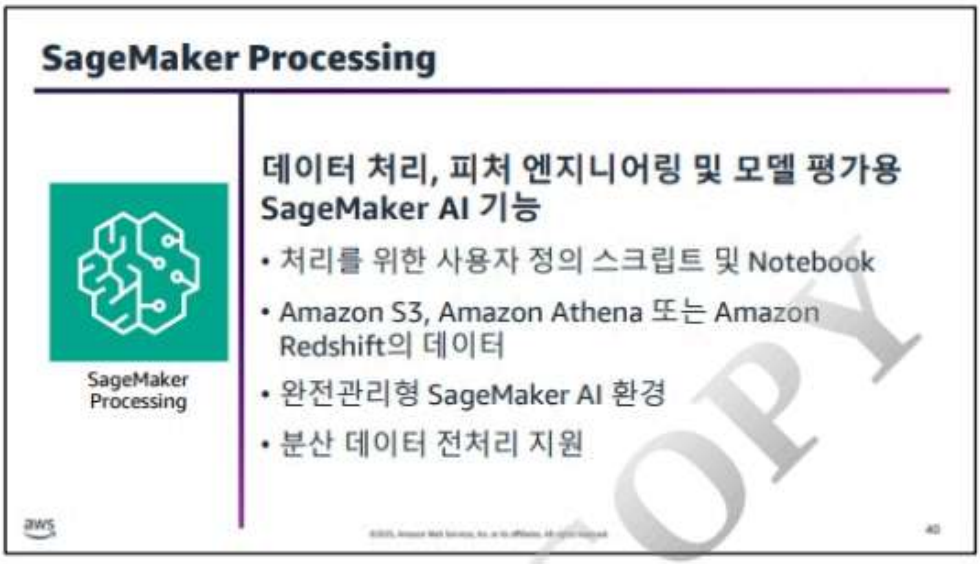
SageMaker Data Wrangler에서 데이터 흐름을 생성할 때는 다음 단계를 수행합니다.

1. **데이터 흐름 생성**: 먼저 SageMaker Studio 인터페이스에서 새 데이터 흐름을 시작합니다.
2. **데이터 소스에 연결**: Amazon S3, Amazon Redshift, Amazon Athena 등 50개 이상의 지원되는 소스에서 데이터를 가져옵니다. SageMaker Data Wrangler는 CSV(쉼표로 구분된 값), JSON, Parquet을 비롯한 여러 파일 형식을 지원합니다.
3. **데이터 인사이트 보고서 살펴보기**: SageMaker Data Wrangler는 데이터의 특성을 이해하는 데 도움이 되는 통계와 시각화가 포함된 보고서를 자동으로 생성합니다.
4. **변환 선택**: 300개 이상의 기본 제공 변환 중 특성 정리, 정규화 및 추출에 사용할 변환을 선택합니다. 필요한 경우 사용자 정의 Python 또는 PySpark 코드를 추가할 수도 있습니다.
5. **변환 미리 보기**: 전체 데이터세트를 처리하기 전에 데이터 하위 집합에 변환을 적용하여 결과를 확인합니다.
6. **변환된 데이터 내보내기**: 변환 결과가 적절하다고 생각되면 처리된 데이터를 Amazon S3, SageMaker Feature Store로 내보낼 수도 있고 SageMaker AI 모델 훈련 작업으로 직접 내보낼 수도 있습니다.

이 프로세스는 대개 반복적이므로, 각 단계에서 파악한 인사이트에 따라 변환을 개선할 수 있습니다. Data Wrangler를 기타 SageMaker AI 기능과 통합하면 데이터 준비 단계에서 모델 훈련 및 배포 단계로 원활하게 전환할 수 있습니다.

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



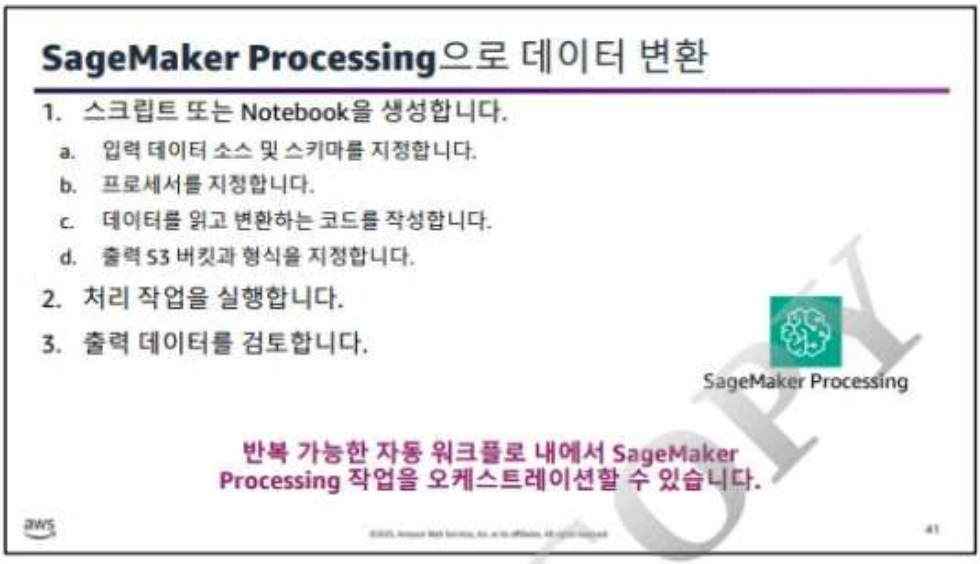
SageMaker Processing에서는 데이터 처리, 피처 엔지니어링 및 모델 평가를 위한 SageMaker AI 기능을 제공합니다. SageMaker Processing은 SageMaker AI 확장 가능 인프라 및 기본 제공 알고리즘을 사용하여 이러한 태스크를 실행합니다. 이 기능은 데이터 준비 및 분석에 여러 가지 주요 이점이 제공됩니다.

- Amazon SageMaker Processing에서는 데이터 처리를 위한 사용자 정의 스크립트와 Notebook을 실행할 수 있습니다. 이러한 유연성 덕분에 관리형 SageMaker AI 환경의 이점과 함께 친숙한 도구와 라이브러리를 사용할 수 있습니다.
- Amazon S3, Amazon Athena, and Amazon Redshift 등의 여러 소스에서 데이터를 처리할 수 있습니다. 이렇게 통합하면 여러 AWS 스토리지 및 분석 서비스에서 데이터 작업을 단순화 합니다.
- 처리 작업이 인프라 프로비저닝과 스케일링이 자동으로 처리되는 기계 학습용 완전관리형 SageMaker AI 환경에서 실행됩니다. 이렇게 하면, 인프라 관리가 아닌 데이터 처리 로직에 집중할 수 있습니다.
- SageMaker Processing에서는 분산 데이터 전처리가 지원되므로 워크로드를 여러 인스턴스에 분배하는 방식으로 대규모 데이터세트를 효율적으로 처리할 수 있습니다.

Amazon SageMaker Unified Studio에도 이름이 비슷한 SageMaker Data Processing의 기능이 포함되어 있습니다. 이 두 도구에서는 모두 유용한 데이터 처리 기능이 제공됩니다.

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



SageMaker Processing으로 데이터를 변환할 때는 몇 가지 주요 단계가 포함됩니다.

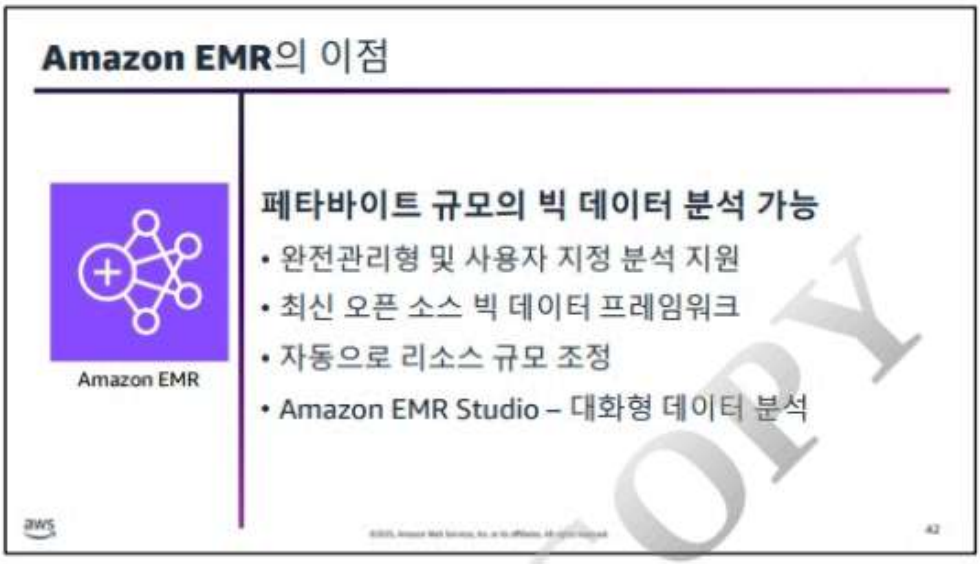
1. 데이터 처리 로직을 정의하는 **스크립트 또는 Notebook을 생성합니다**. Python, R 또는 기타 지원되는 언어를 사용할 수 있습니다. Notebook 또는 스크립트에 포함해야 하는 정보는 다음과 같습니다.
   a. 입력 데이터 소스 및 스키마를 지정합니다. Amazon S3, Athena, Amazon Redshift의 데이터를 지정할 수 있습니다.
   b. SKLearn, Spark 등의 프로세서 또는 필수 라이브러리가 포함된 사용자 정의 컨테이너를 선택합니다.
   c. 데이터를 읽고 변환하는 코드를 작성합니다. 이 코드에서 데이터 클리닝, 피처 엔지니어링 및 기타 사전 처리 단계를 구현합니다.
   d. Amazon S3의 출력 위치 및 원하는 출력 형식을 지정합니다.
2. SageMaker Python SDK 또는 AWS Management Console을 사용하여 **처리 작업을 실행합니다.**
3. 출력 데이터 및 처리 로그를 검토하여 **결과를 검증합니다**.

SageMaker Processing 작업은 SageMaker Pipelines에 통합하거나, 자동화되고 반복 가능한 워크플로우를 위한 AWS Step Functions 등의 서비스를 사용해 오케스트레이션할 수 있습니다. 이렇게 하면 데이터 준비, 모델 훈련 및 배포 단계가 포함된 엔드 투 엔드 ML 파이프라인을 생성할 수 있습니다.

SageMaker Processing을 사용하면 SageMaker AI의 관리형 인프라와 다른 AWS 서비스와 통합을 활용하면서 확장 가능 데이터 변환을 수행할 수 있습니다.

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



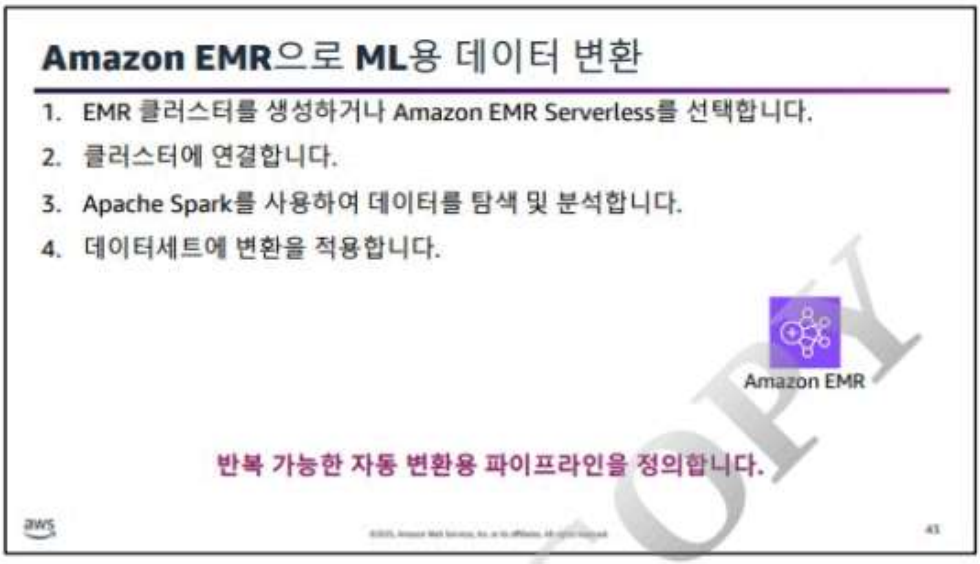
Amazon EMR은 Apache Spark, Hive, Presto 등의 분산 컴퓨팅 프레임워크를 사용해 **페타바이트급 빅 데이터 분석**을 수행하고 방대한 양의 데이터를 효율적으로 처리할 수 있도록 도와줍니다. 이 서비스는 빅 데이터 처리와 분석에 다음과 같은 여러 가지 주요 이점이 제공됩니다.

- Amazon EMR은 **사용자 지정 가능 완전관리형** 서비스이므로 복잡한 클러스터 설정, 관리, 스케일링 작업은 모두 AWS에서 처리합니다. 클러스터에 설치하는 소프트웨어와 구성은 사용자가 전적으로 제어할 수 있습니다.
- **최신 오픈 소스 프레임워크 릴리스**에서 사용자 정의 스크립트를 실행하여 Spark, Hive 등의 도구에서 최신 기능과 성능 향상을 사용할 수 있습니다.
- Amazon EMR 클러스터는 워크로드에 따라 **리소스를 자동으로 확장 및 축소**할 수 있으므로 리소스 사용량이 최적화되며 비용 효율성이 높아집니다.
- Amazon EMR Studio에서는 **대화형 데이터 분석**용 통합 개발 환경이 제공되므로 데이터를 살펴보고 분석 애플리케이션을 개발할 수 있습니다.
- Amazon EMR은 기계 학습 워크플로용 서비스인 Amazon SageMaker AI와 원활하게 통합됩니다. 데이터 준비와 피처 엔지니어링에 Amazon EMR을 사용하고, 모델 훈련 및 배포를 위해 처리된 데이터를 SageMaker AI에 전달할 수 있습니다.

Amazon EMR을 사용하는 경우에는 인프라 관리가 아닌 데이터에서 인사이트를 추출하는 작업을 중점적으로 수행할 수 있습니다. Amazon EMR는 AWS의 빅 데이터 분석 서비스급 확장성과 유연성을 갖춘 서비스입니다.

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



Amazon EMR을 사용하는 ML용 데이터 변환에서는 몇 가지 주요 단계가 포함됩니다.

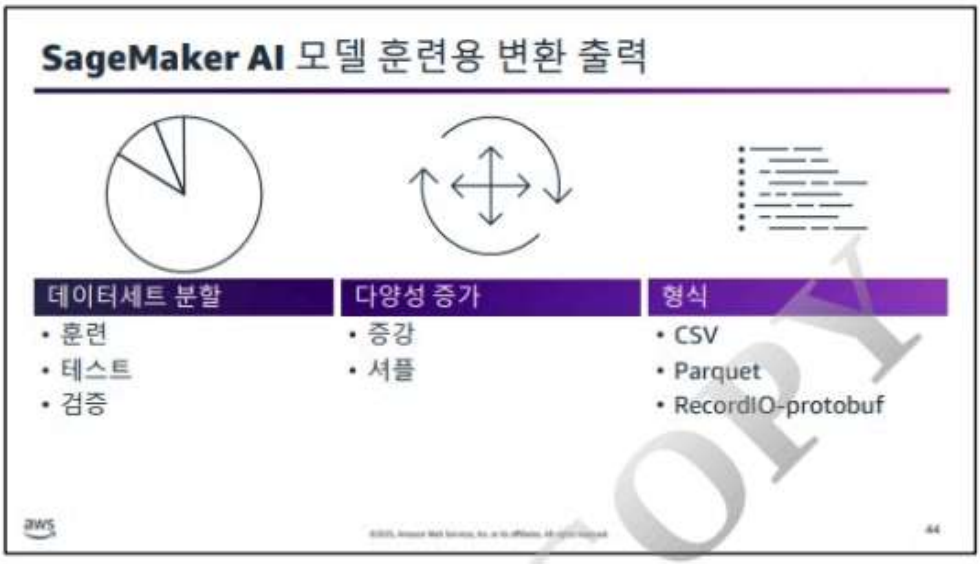
1. **EMR 클러스터를 생성하거나 Amazon EMR Serverless를 선택합니다.** 영구 클러스터가 필요한지 아니면 더욱 유동적인 스케일링과 비용 관리를 위해 서버리스 옵션을 선택할지를 결정합니다.
2. **클러스터에 연결합니다.** Secure Shell(SSH)을 사용하거나, EMR Studio를 사용하거나, 다른 AWS 서비스에서 연결할 수도 있습니다. 이후 진행할 실습에서는 SparkMagic PySpark 커널을 사용하여 SageMaker Studio Notebook에서 연결합니다.
3. **Apache Spark를 사용하여 데이터를 탐색 및 분석합니다.** Spark의 유용한 분산 컴퓨팅 기능을 사용하여 대규모 데이터세트에서 초기 데이터 탐색 및 분석을 수행합니다.
4. **데이터세트에 변환을 적용합니다.** Spark API를 사용하여 데이터를 정리하고 피처 엔지니어링하고, 기계 학습용 데이터를 준비합니다. 이 과정에서 필터링, 집계, 데이터세트 결합, 파생 특성 생성 등의 작업을 수행할 수 있습니다.
5. **변환된 데이터를 출력합니다.** 후속 ML 워크플로에 사용할 수 있도록 Amazon S3 또는 기타 적절한 스토리지 서비스에 처리된 데이터를 다시 저장합니다.

Amazon EMR은 다른 AWS 서비스와 통합할 수 있으므로 엔드 투 엔드 데이터 처리 및 기계 학습 파이프라인을 생성할 수 있습니다. 예를 들어 Amazon EMR을 사용하여 대규모 데이터 준비를 수행한 후 모델 훈련 및 배포용으로 SageMaker AI에 결과를 전달할 수 있습니다.

Amazon EMR 사용 시에는 Spark와 같은 분산 컴퓨팅 프레임워크의 우수한 확장성과 성능을 활용하여 대규모 데이터세트를 효율적으로 처리 및 변환할 수 있습니다.

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



SageMaker AI에서 ML 모델을 훈련하는 데 사용할 데이터를 변환할 때는. 모델 성능 평가, 편향 감소, 훈련 효율성 개선 기회를 고려해야 합니다. 예를 들어, 데이터세트의 편향을 줄이는 것은 데이터 준비의 주요 목표 중 하나입니다. 균형 잡힌 데이터를 사용하면 모델은 내재적인 노이즈가 아닌 데이터세트의 의미 있는 패턴에 집중할 수 있습니다.

변환 출력을 정의할 때 사용할 수 있는 3가지 주요 기법은 다음과 같습니다.

- **데이터세트 분할** – 균형 잡히고 정확한 데이터세트를 얻기 위한 일반적인 전략은 훈련, 테스트, 검증을 위해 데이터를 여러 개의 별도 하위 집합으로 나누는 것입니다. 테스트 및 검증 하위 집합을 사용해 처음 보는 데이터에 대한 모델 성능을 평가할 수 있습니다. 이 하위 집합이 이때 훈련 데이터세트와 프로덕션 데이터세트를 대표하는지 확인하면 모델을 프로덕션 환경에 배포하기 전에 모델 성능의 편향을 식별할 수 있습니다.
- **다양성 개선** – 과소평가된 하위 그룹을 가진 불균형한 데이터세트를 생각해보십시오. 증강이나 셔플링 등의 기법을 통해 데이터세트의 다양성을 높이면 훈련 모델을 더 균형 있게 표현할 수 있습니다.
  - **데이터 증강**은 모델의 데이터 분포에 대한 이해를 넓히는 새롭고 현실적인 훈련 사례를 만드는 방식으로 작동합니다. 데이터세트 증강은 기존 데이터 샘플에 다양한 변환과 수정을 적용하여 데이터세트의 크기와 다양성을 인위적으로 확장하는 것을 포함합니다. 이러한 변환은 학습 태스크와 관련된 필수적인 특성을 보존하는 동시에 데이터에 변형을 도입합니다.
  - **데이터 셔플링**은 ML 모델을 훈련하기 전에 데이터세트의 데이터 포인트 순서를 재정렬합니다. 데이터를 셔플링하면 훈련 예제가 고정된 순서가 아닌 무작위 순서로 모델에 제공됩니다. 이는 고정된 순서로 제공되는 경우에는 모델 성능에 편향이 발생하거나 정확도가 낮은 결과가 해석될 수 있기 때문입니다.

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

- **형식** - SageMaker AI는 특정 형식의 데이터와 잘 작동하도록 설계되었습니다. SageMaker AI 기본 제공 알고리즘에서 지원하는 가장 일반적인 2가지 형식은 CSV 형식과 RecordIO-protobuf 형식입니다. 그러므로 이 2가지 형식과 관련된 기본적인 개념과 구조를 파악해 두면 유용합니다.
  - CSV 파일은 테이블 형식 데이터를 저장하는데, 각 행은 데이터 샘플을 나타내고 열은 해당 샘플의 특성을 나타냅니다. SageMaker AI 기본 제공 알고리즘의 경우 훈련 데이터세트의 레이블은 왼쪽의 첫 번째 열이어야 하고 특성은 오른쪽에 있어야 합니다. XGBoost, 선형 학습기, DeepAR과 같은 SageMaker AI의 대다수 기본 제공 알고리즘은 모델 훈련을 위한 입력으로 CSV 형식의 데이터를 허용합니다.
  - Apache Parquet 형식은 성능 중심의 오픈 소스 열 기반 데이터 형식입니다.
  - RecordIO-protobuf는 SageMaker AI에서 효율적인 데이터 저장 및 전송을 위해 사용하는 이진 데이터 형식입니다. 이 형식은 특히 대용량 데이터세트를 처리하거나 자동 데이터 변환을 수행해야 할 때 유용합니다. RecordIO-protobuf 형식은 일련의 레코드로 구성되어 있는데, 각 레코드에는 이진 인코딩된 protobuf 메시지가 포함됩니다. 이 형식은 각 레코드와 관련 메타데이터를 나타내는 이미지 데이터에 일반적으로 사용됩니다.

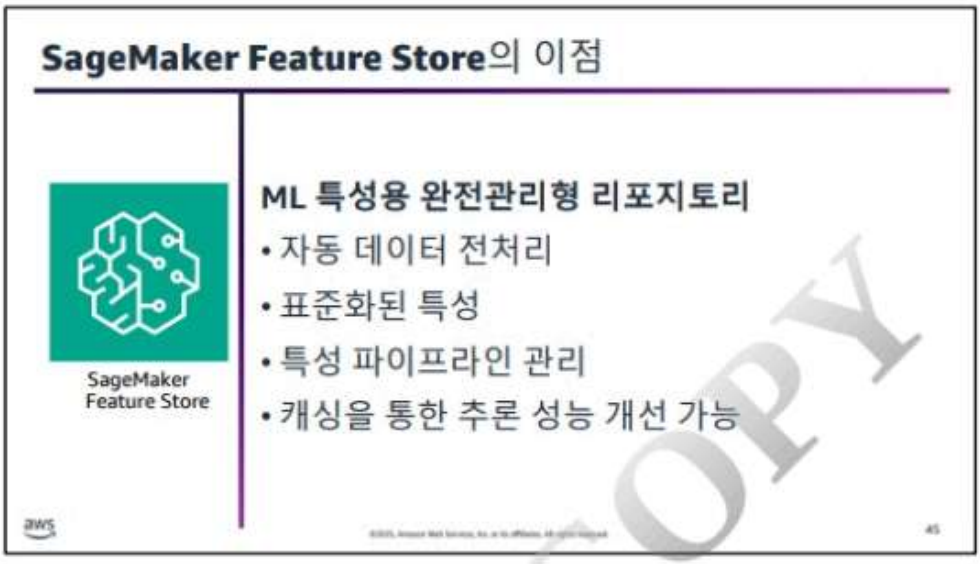
기계 학습 수명 주기의 모델 훈련 및 미세 조정 단계에서 특정 형식 유형을 준수하는 데 더 집중합니다.

SageMaker AI에서 지원되는 형식과 관련된 추가 정보는 Amazon SageMaker AI Developer Guide의 Common Data Formats for Training(훈련에 사용되는 일반적인 데이터 형식)(https://docs.aws.amazon.com/sagemaker/latest/dg/cdf-training.html)을 참조하십시오.

DO NOT COPY ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



SageMaker Feature Store는 ML 모델에 사용되는 특성을 저장, 공유, 관리하기 위한 관리형 리포지토리입니다. 여러 팀에서 특성은 재사용할 수 있으며 SageMaker Feature Store는 일괄 훈련과 실시간 추론을 위해 해당 특성을 동기화된 상태로 유지합니다. SageMaker Feature Store는 안전한 저장소에 특성을 중앙 집중화하여 ML 수명 주기 전반에 걸쳐 대규모로 특성 품질 및 모델 정확도를 개선합니다. 특성을 저장하면 중복되는 피처 엔지니어링 노력이 없어져 시간을 절약할 수 있습니다.

SageMaker Feature Store는 ML 모델용 특성을 저장, 공유 및 관리할 수 있는 완전관리형 서비스입니다. 이 서비스에서는 ML 워크플로를 간편하게 진행할 수 있는 중앙 집중식 피처 엔지니어링 관리 기능이 제공됩니다. 데이터 수집부터 모델 훈련 및 예측에 이르는 검색, 재사용 및 자동화가 지원하는 메타데이터가 포함된 특성 그룹의 구성 가능한 온라인 및 오프라인 스토리지가 제공됩니다.

DO NOT COPY
ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

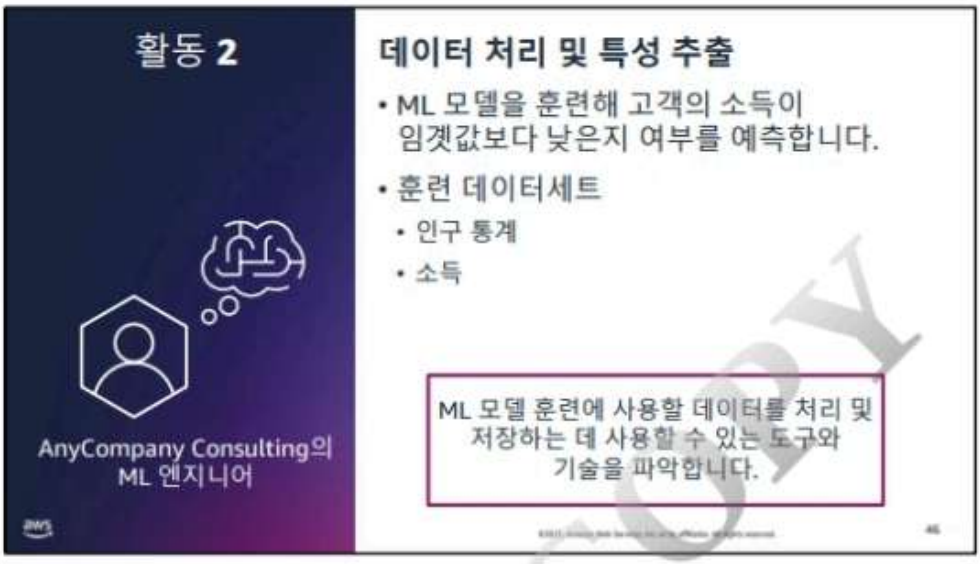
SageMaker Feature Store는 특성을 엔지니어링, 모니터링, 제어 및 재사용할 수 있는 중앙 집중화되고 자동화된 방법을 제공하여 팀 전체의 모델 훈련 및 성능을 개선합니다. 다음은 몇 가지 일반적인 사용 사례입니다.

- **자동 데이터 전처리** – SageMaker Feature Store는 원시 데이터가 수집될 때 자동으로 사전 처리할 수 있습니다. 여기에는 모델 훈련에 사용할 수 있도록 데이터를 스케일링, 인코딩, 변환 및 정리하는 작업이 포함됩니다.
- **중앙 집중식 특성 리포지토리** – SageMaker Feature Store는 사전 처리 및 추출된 큐레이팅된 특성을 위한 중앙 집중식 리포지토리 역할을 합니다. 여러 팀과 모델에서 특성을 재사용할 수 있습니다. 이렇게 하면 반복적인 사전 처리를 피할 수 있어 성능이 향상됩니다.
- **표준화된 특성** – SageMaker Feature Store는 조직 전체에서 일관된 특성화된 데이터에 대한 표준화된 정보 소스를 설정합니다. 그러므로 여러 팀에서 자체 특성 추출 코드를 사용할 때 발생하는 문제를 방지할 수 있습니다.
- **특성 파이프라인 관리** – SageMaker Feature Store는 데이터 수집부터 사전 처리까지 특성 파이프라인의 엔드 투 엔드 수명 주기를 관리합니다. 이는 특성 추출의 모니터링 및 거버넌스를 제공합니다.
- **성능을 위한 캐싱** – 자주 사용하는 특성은 추론 중에 더 빠르게 검색할 수 있도록 SageMaker Feature Store에 캐시됩니다.

DO NOT COPY
ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



이 활동에서는 다음 시나리오에 적용할 기법과 도구를 선택합니다.

여러분이 근무하고 있는 마케팅 업체에서 판매량을 늘리고자 합니다. 이 업체는 대규모 마케팅 캠페인을 계획 중이며 해당 캠페인의 투자 수익률을 최대한 높이고자 합니다. 마케팅 캠페인의 일환으로, 연 소득이 5만 달러 미만인 고객은 프로모션 혜택을 받을 수 있습니다. 사용 가능한 인구 통계 데이터를 토대로 특정 고객의 연 소득이 이 임곗값보다 낮은지 여부를 예측하도록 ML 모델을 훈련해야 합니다. 데이터세트에는 대규모 성인 모집단의 인구 통계 및 급여 데이터가 포함되어 있습니다.

DO NOT COPY
ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

1. 데이터를 살펴보면서 데이터 처리 단계를 파악합니다.

CSV 훈련 데이터세트의 샘플
- 인구 통계
- 소득

ID	소득	위치	연령
1	50K	뉴욕 시	52
2	42,000	플로리다주 마이애미	30
3	$37000	91809	19
4	25000000	매사추세츠주 보스턴	47
5	62000	뉴욕주 뉴욕	23
6	38K	60609	2
7	57,250.25	텍사스 오스틴	
8	시간당 15달러		29
9	25000	60609-9999	35

47

슬라이드에는 CSV 형식 훈련 데이터세트의 샘플이 나와 있습니다. 각 열과 행을 검토하여 데이터 품질 문제를 파악하십시오.

각 문제 유형을 해결할 수 있는 기법은 무엇입니까?

DO NOT COPY
ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

4가지 데이터 품질 오류 유형이 강조 표시되어 있습니다.

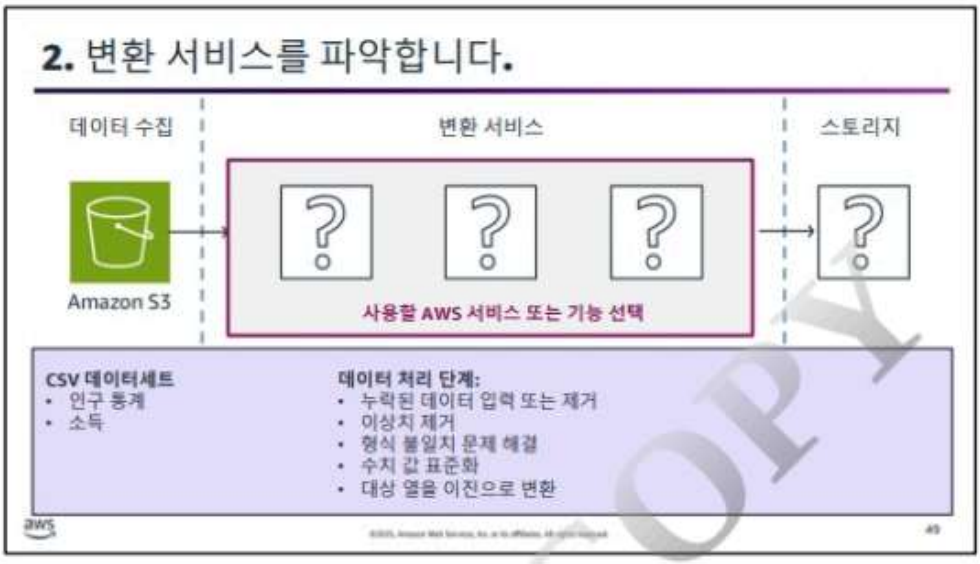
- 위치 형식 불일치 - 문자열을 우편 번호로 변환하고 5자리 형식으로 표준화합니다.
- 연령 및 위치의 데이터 누락 – 입력 데이터로 누락된 값을 채우거나 누락된 데이터가 있는 항목을 제거합니다.
- 소득 및 연령의 이상치 - 잘못된 값을 수정하거나 이상치가 포함된 항목을 제거합니다.
- 소득 정보를 이진 데이터로 변환 - 모든 데이터를 기호가 없는 연간 소득 수치 값으로 변환합니다. 그런 다음 해당 수치 값을 대상 소득 임곗값인 5만 달러보다 큰지 아니면 작은지를 나타내는 이진 데이터 0이나 1로 변환합니다.

시각화 또는 책임감 있는 AI 측면에서 이 데이터세트와 관련하여 고려할 수 있는 다른 사항은 무엇입니까?

DO NOT COPY ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



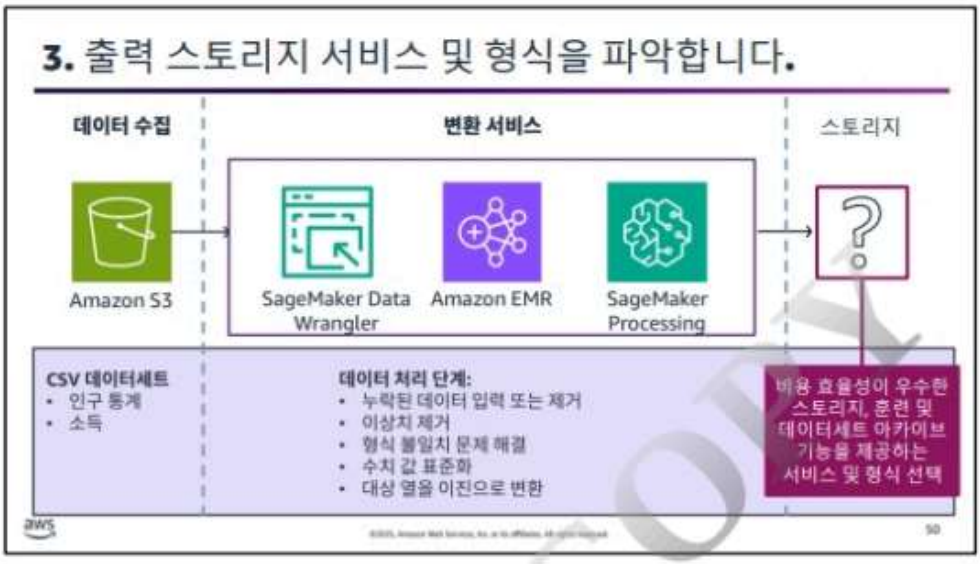
이전 슬라이드에서 살펴본 데이터세트는 CSV 형식으로 Amazon S3에 저장됩니다. 누락된 데이터와 이상치 처리 등 데이터세트에 대해 수행해야 하는 변환을 확인했습니다.

데이터 처리 단계를 완료하기 위해 사용할 수 있는 3가지 AWS 서비스를 제시하십시오.

DO NOT COPY
ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



이 모듈에서 살펴보았던 ML 훈련용 데이터세트를 변환하는 데 사용할 수 있는 3가지 서비스는 다음과 같습니다.

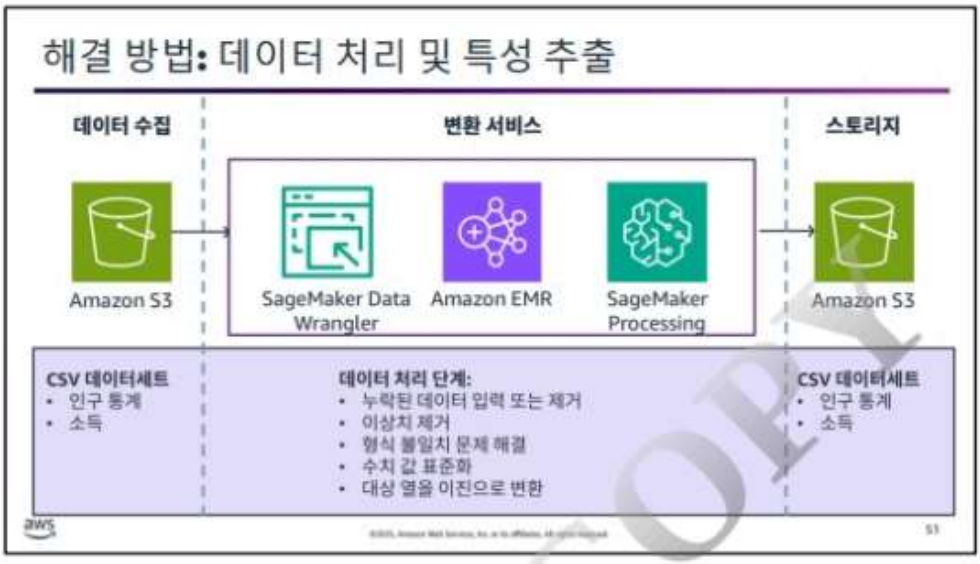
- SageMaker Data Wrangler
- Amazon EMR
- SageMaker Processing

이 시나리오에서는 SageMaker AI에서 ML 모델을 훈련할 때 사용할 변환된 데이터세트를 위한 비용 효율성이 우수한 스토리지가 필요합니다. 또한 프로덕션 환경에서 사용하는 모든 ML 모델용으로 훈련 데이터세트의 아카이브를 유지 관리해야 한다는 규정 준수 요구 사항도 충족해야 합니다.

비용 효율성이 우수한 솔루션을 제공하는 AWS 서비스는 무엇입니까? SageMaker AI에서 ML 훈련을 효율적으로 진행하는 데 사용할 수 있는 데이터 형식은 무엇입니까?

DO NOT COPY
ssleeeee1@gmail.com

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



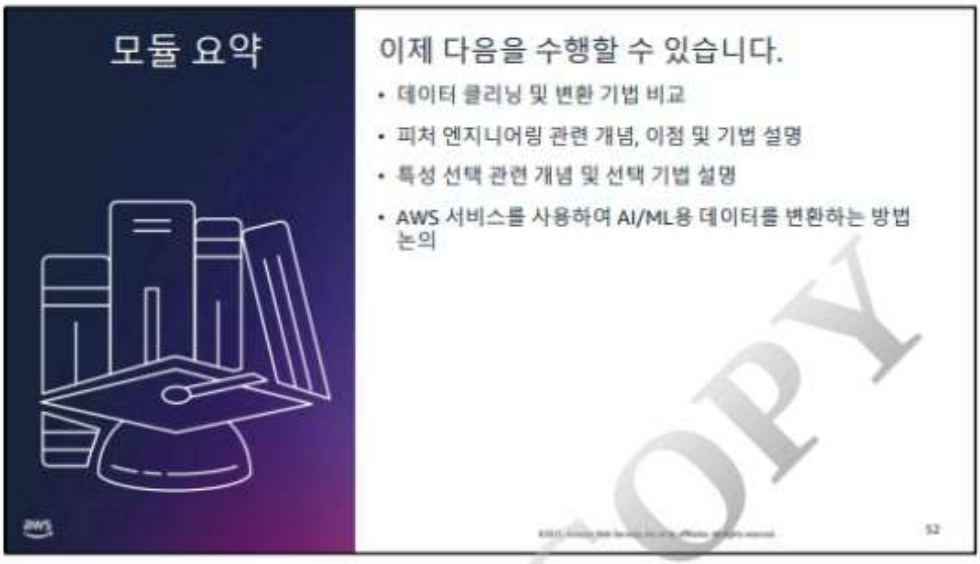
이 슬라이드에는 앞서 수행한 활동의 샘플 해결 방법이 나와 있습니다. 이 샘플 해결 방법에는 **데이터 처리 및 피처 엔지니어링 워크플로**에 사용할 수 있는 아키텍처 하나가 대략적으로 표시되어 있습니다. 해당 워크플로에서는 먼저 Amazon S3에서 **데이터 수집**을 수행합니다. Amazon S3에는 인구 통계 및 소득 정보가 포함된 원시 CSV 데이터세트가 저장되어 있습니다.

3가지 주요 AWS 변환 서비스가 포함되어 있습니다: Amazon SageMaker Data Wrangler, Amazon EMR, SageMaker Processing입니다. 이러한 도구 중 하나 이상을 사용하여 일반적인 **데이터 처리** 단계를 수행합니다. 이러한 단계로는 누락된 데이터 처리, 이상치 제거, 형식과 단위 표준화, 대상 열을 이진으로 변환 등이 있습니다.

변환의 마지막 단계에서는 처리된 데이터세트를 내보냅니다. 이 예에서는 모델 훈련에 바로 사용할 수 있는 CSV 형식으로 Amazon S3에 데이터가 저장됩니다. 이 워크플로에서는 **여러 AWS 서비스를 함께 사용**하여 기계 학습 프로젝트를 위한 효율적이고 확장 가능한 데이터 준비 파이프라인을 생성하는 방법을 제시합니다.

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



이 모듈을 통해 기계 학습의 데이터 준비 작업에 꼭 필요한 지식과 기술을 습득했습니다.

- 데이터 클리닝 및 변환: ML 훈련 데이터의 품질을 개선하기 위해 부정확한 데이터, 중복 데이터, 이상치 및 누락된 값을 처리하는 것이 중요하다는 것을 알아보았습니다.
- 인코딩 기법: 범주형 데이터를 ML 알고리즘에 적합한 수치 형식으로 변환하는 방법을 파악했습니다.
- 피처 엔지니어링: 새 특성을 생성하거나 기존 특성을 수정하여 모델 성능을 높이는 기법을 살펴보았습니다.
- 특성 선택: ML 모델과 가장 관련성이 높은 특성을 선택하는 기법에 대해 알아보았습니다.
- 데이터 준비를 위한 AWS 서비스: 효율적인 데이터 변환과 피처 엔지니어링을 위한 SageMaker Data Wrangler, Amazon EMR, SageMaker Processing 등의 도구를 실제로 경험해 보았습니다.

이러한 기술을 습득한 ML 엔지니어와 팀은 안정적이며 정확한 ML 모델을 개발할 수 있습니다. 이러한 개념을 업무에 적용할 때는 ML 프로젝트의 목표를 달성하려면 효율적인 데이터 준비 작업이 중요한 경우가 많다는 점을 꼭 기억해 두시기 바랍니다.

DO NOT COPY
ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

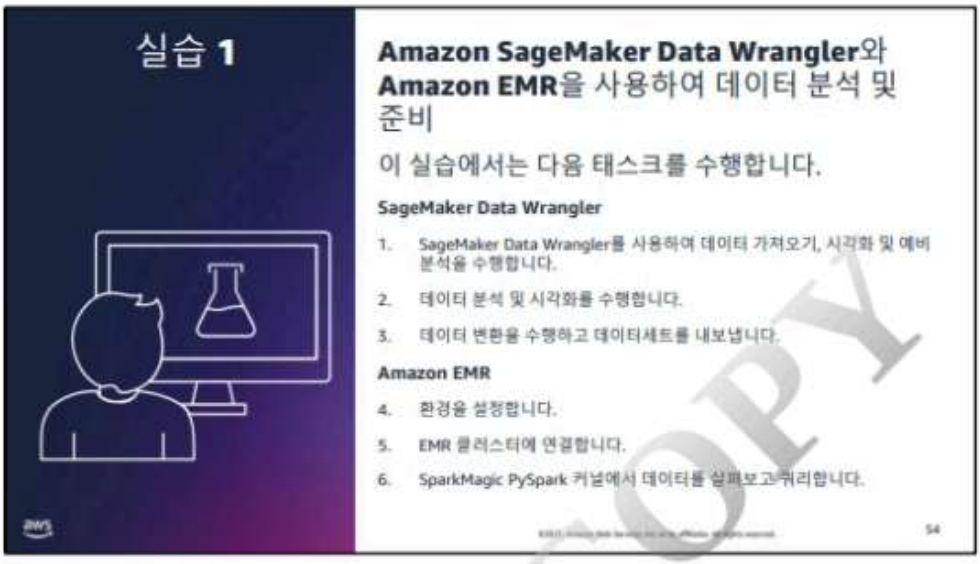
Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

이 모듈에서 알아본 데이터 준비 및 특성 추출 관련 개념을 잠시 생각해 보시기 바랍니다. 이 과정에서 다룬 자료 또는 이 과정에서 배운 기술을 실제 업무에 적용하는 방법과 관련하여 궁금한 점이 있으면 지금 질문을 할 수 있습니다.

DO NOT COPY
ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



이 실습에서는 데이터 분석 및 준비에 사용할 수 있는 2가지 유용한 AWS 서비스인 SageMaker Data Wrangler 및 Amazon EMR을 실제로 사용해 보게 됩니다.

SageMaker Data Wrangler에서는 다음 작업을 수행하는 방법을 알아봅니다.
- 데이터세트 가져오기 및 시각화
- 예비 분석 및 데이터 변환 수행
- 나중에 사용할 수 있도록 처리된 데이터세트 내보내기

Amazon EMR에서는 다음 작업을 수행하게 됩니다.
- 빅 데이터 처리 환경 설정
- EMR 클러스터에 연결
- SparkMagic 및 PySpark 커널에서 데이터 살펴보기 및 쿼리

이 실습에서는 데이터 준비 기법에 대한 실무 경험을 제공하고 대규모 데이터 처리 태스크 처리용 AWS 도구 사용법도 숙지할 수 있습니다. 학습 경험을 최대한 효율적으로 진행하려면 해당 프로세스의 진행 방식을 중점적으로 파악하면서 궁금한 점을 질문하시기 바랍니다.

NOT COPY
ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

이 실습에서는 데이터 준비 태스크에 SageMaker Processing 및 SageMaker Python SDK를 사용하는 방식을 중점적으로 살펴봅니다. 다음 작업에 대한 실무 경험을 쌓을 수 있습니다:

1. **환경 설정**: SageMaker Processing을 사용할 수 있도록 개발 환경을 구성합니다.
2. **데이터 처리 수행**: SageMaker Processing을 사용하여 데이터 변환 및 특성 추출 태스크를 수행합니다.

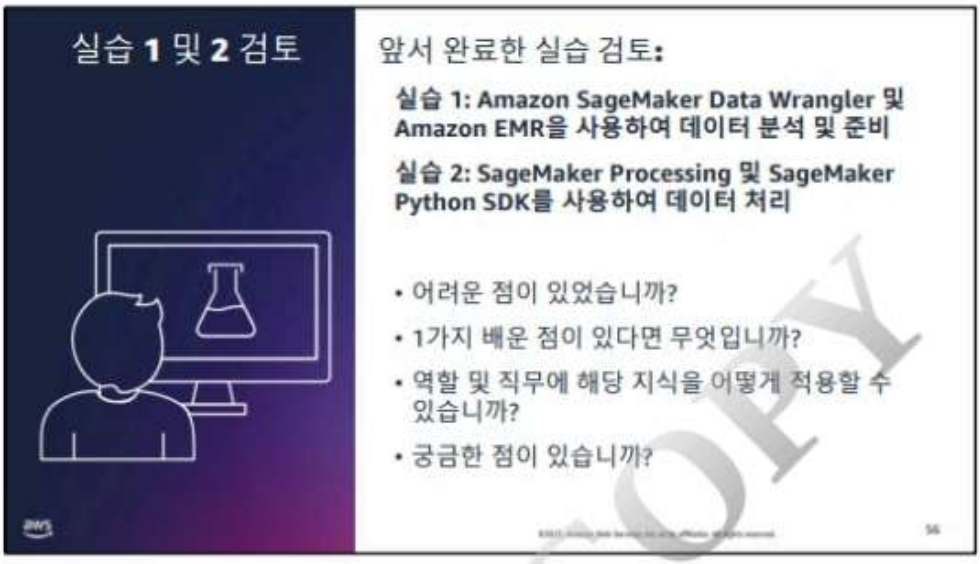
SageMaker Python SDK는 SageMaker AI 작업을 위한 고급 인터페이스가 제공하여 기계 학습 워크플로를 Python 애플리케이션에 통합하는 데 도움이 됩니다. 이 실습은 대규모 데이터 준비 파이프라인을 더 유연하고 재현 가능한 SageMaker AI의 처리 기능을 프로그래밍 방식으로 사용하는 방법을 이해하는 데 도움이 됩니다.

현재 진행 중인 ML 프로젝트에서 이러한 도구를 응용할 수 있는 방법 및 해당 도구 관련 개념을 중점적으로 파악하십시오.

DO NOT COPY
ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



DO NOT COPY
ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

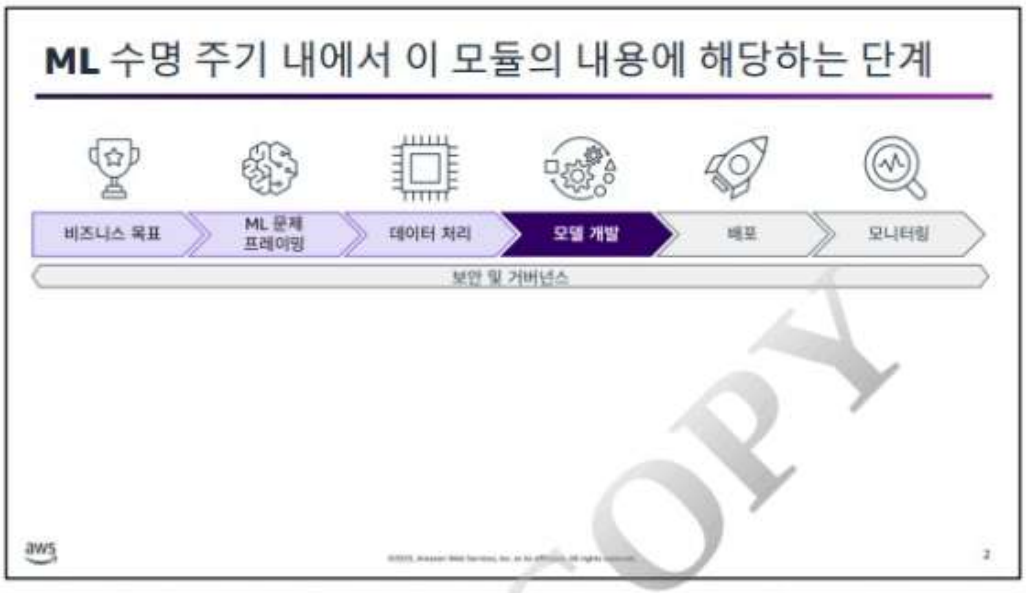
Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

이 모듈에서는 기계 학습(ML) 엔지니어링의 중요한 단계인 올바른 모델링 방식 선택에 대해 중점적으로 살펴봅니다. 구체적으로는 AWS ML 스택의 여러 계층을 살펴보고 AWS AI 서비스를 사용해 일반적인 비즈니스 도전 과제를 해결하는 방법을 알아봅니다. 또한 Amazon SageMaker AI를 사용하여 적절한 모델을 선택하는 전략을 설명하고, 모델 선택 과정에서 해석 가능성을 고려해야 하는 이유도 살펴봅니다. 이 모듈을 마치면 프로젝트용 ML 모델을 선택할 때 정보를 토대로 적절한 결정을 내리는 방법을 포괄적으로 파악할 수 있습니다.

DO NOT COPY
ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



보안은 전체 **기계 학습 수명 주기**에서 중요한 고려 사항입니다. 이 모듈에서는 AWS에서 진행하는 ML 개발 및 배포의 여러 단계에서 보안 조치를 구현하는 방법을 알아봅니다.

- 데이터 수집 및 준비: ML 모델용 데이터를 수집하고 사전 처리할 때는 데이터를 안전하게 저장하고 액세스를 제어할 수 있어야 합니다.
- 모델 개발 및 훈련: 훈련 환경을 보호하고 개발 중에는 모델 무결성을 보장하는 것이 필수입니다.
- 모델 배포 및 추론: 엔드포인트를 보호하고 배포된 모델에 대한 액세스를 관리하면 모델 무단 사용 또는 변조를 방지할 수 있습니다.
- 모니터링 및 유지 관리: ML 시스템의 무결성과 성능을 장기적으로 유지하려면 지속적인 보안 모니터링과 업데이트를 수행해야 합니다.

ML 수명 주기 전반에 보안 사례를 통합하면 AWS에서 기계 학습 프로젝트를 진행하는 데 사용할 안정적이며 신뢰할 수 있는 환경을 구축할 수 있습니다.

DO NOT COPY
ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

이 모듈에서는 AWS에서 모델링 방식을 선택할 때 적용되는 주제를 포괄적으로 살펴봅니다.

- 주제 A: Amazon SageMaker AI 기본 제공 알고리즘
- 주제 B: Amazon SageMaker Autopilot
- 주제 C: 기본 제공 훈련 알고리즘 선택
- 주제 D: 모델 선택 시의 고려 사항
- 주제 E: ML 비용 관련 고려 사항

이러한 주제 관련 사항을 숙지하면 AWS의 모델 선택 프로세스를 전반적으로 파악할 수 있으므로 ML 프로젝트에서 정보를 토대로 적절한 결정을 내릴 수 있습니다.

DO NOT COPY
ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

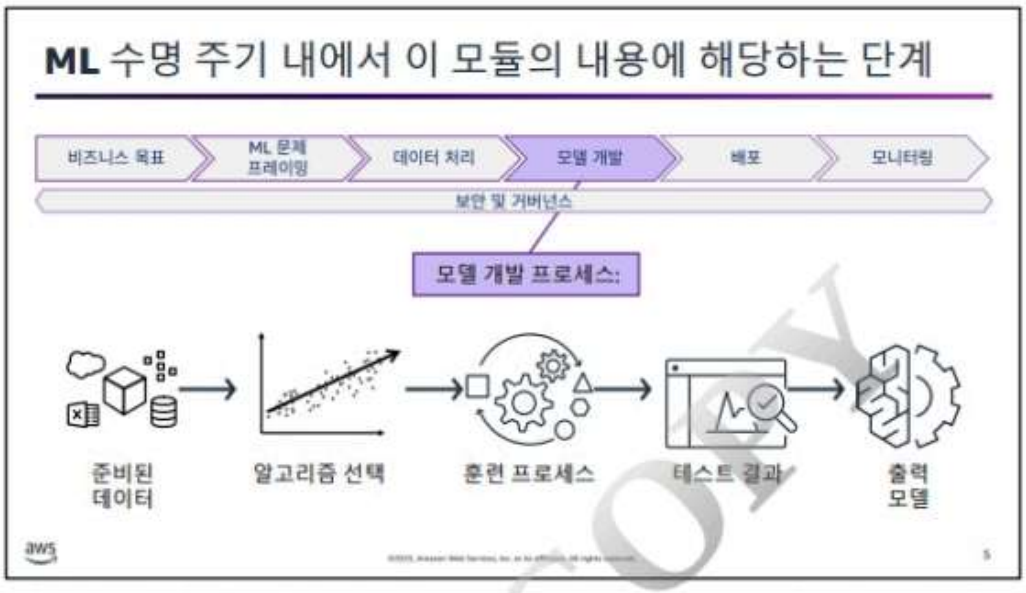
Amazon SageMaker AI에서는 폭넓은 ML 태스크에 활용할 수 있는 포괄적인 기본 제공 알고리즘 집합이 제공됩니다. 이러한 알고리즘은 훈련과 배포를 효율적으로 진행할 수 있도록 사전 구성 및 최적화되어 있으므로 모델을 자세히 사용자 지정할 필요 없이 빠르게 구축하여 배포할 수 있습니다.

기본 제공 알고리즘에 더하여, SageMaker AI에서는 기존 보유 사용자 정의 알고리즘 또는 AWS Marketplace의 사전 훈련된 모델도 사용할 수 있습니다. 이러한 유연성으로, 구체적인 업무상의 요구에 가장 적합한 솔루션을 선택할 수 있습니다.

DO NOT COPY
ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

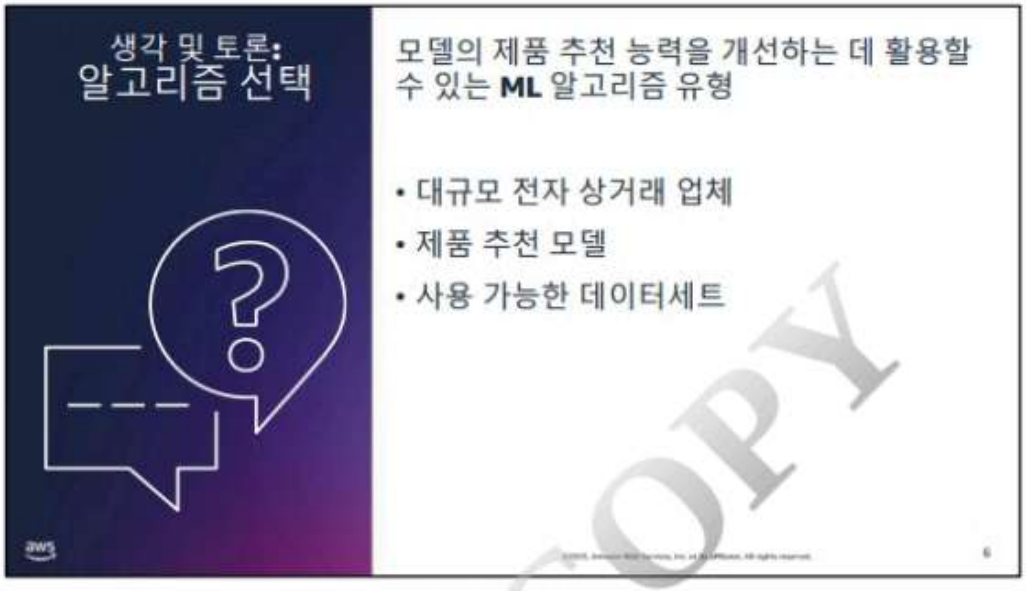


ML 수명 주기란 비즈니스 목표 정의에서 배포된 솔루션 모니터링까지의 모든 단계가 포함되는 기계 학습 프로젝트의 전체 과정에 해당됩니다. 이 모듈에서는 모델 개발 단계를 중점적으로 살펴보면서 Amazon SageMaker AI의 기본 제공 알고리즘, 그리고 다양한 업무상의 문제에 적합한 알고리즘을 선택하는 방법을 알아봅니다.

모델 개발 프로세스 자체는 특정 워크플로에 따라 진행됩니다. 해당 워크플로에서는 먼저 데이터를 준비하고 적절한 알고리즘을 선택한 후 훈련 프로세스를 구현하고 테스트 결과를 평가한 다음 출력 모델을 생성합니다. SageMaker AI의 기능을 가장 적절하게 활용하여 ML 관련 당면 과제를 해결하려면 이 프로세스의 각 단계를 면밀하게 고려해야 합니다.

DO NOT COPY
ssleeeee1@gmail.com

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



활동 시나리오: 전자 상거래 업체의 추천 모델 관련 토론
대형 온라인 소매 업체에서 제품 추천 시스템을 개선하려고 합니다. 이 업체가 이용할 수 있는 정보는 다음과 같습니다.
- 고객 구매 내역(함께 구매한 품목)
- 검색 패턴(각 페이지를 살펴본 시간 및 구경한 품목)
- 클릭 방문 데이터(클릭한 추천 항목)

다양한 모델링 방식 고려:
특정 알고리즘을 구현하기 전에 다양한 모델링 패러다임을 평가하여 고유한 데이터 특성 및 업무상의 요구 사항에 가장 적합한 모델링 방식을 결정해야 합니다.

지도 학습 방식:
- 과거 구매 데이터를 사용하여 향후 구매 예측
- 성공적인 추천에 대한 레이블이 지정된 데이터에서 학습
- 알려진 고객-구매 매칭 정보로 모델 훈련

비지도 학습 방식:
- 유사 제품으로 구성된 자연스러운 그룹 파악
- 고객 검색 동작의 패턴 파악
- 제품 간의 숨겨진 관계 파악

고려할 수 있는 모델 유형:
- 선형 모델(단순 예측 패턴)
- 트리 기반 모델(복잡한 의사결정 규칙)
- 신경망(심층 패턴 인식)
- 행렬 기반 모델(고객과 제품 간 관계 파악)

DO NOT COPY
ssleeeee1@gmail.com

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

고려 사항:

1. 데이터의 특징
   - 정형 데이터 또는 비정형 데이터
   - 수치, 범주형, 혼합
   - 데이터세트의 특성 수

2. 데이터세트의 크기
   - 데이터의 양(소량, 보통, 대량)
   - 시간 경과에 따라 데이터세트가 매우 커질 가능성

3. 데이터 관계의 복잡도
   - 관계의 선형 또는 비선형
   - 특성 간의 상호 작용 정도

4. 필요한 모델 해석 가능성 수준
   - 이해관계자에게 모델 결정 사항을 설명해야 하는지 여부
   - 블랙 박스 예측이 가능한지 아니면 모델 투명성을 유지해야 하는지 여부

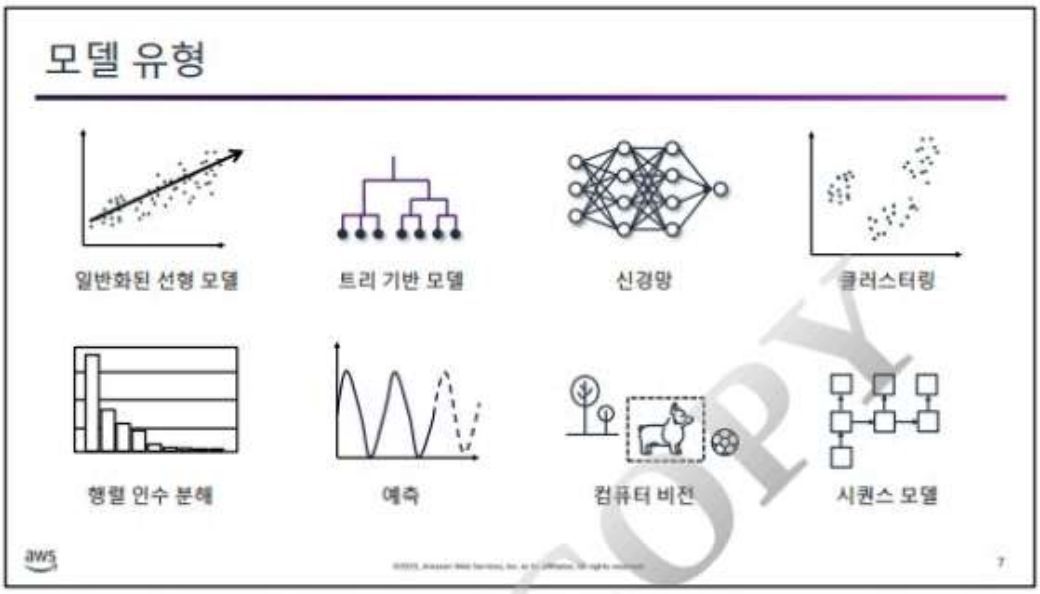
5. 컴퓨터 예산
   - 고성능 컴퓨팅 리소스 이용 가능 여부
   - 훈련 시간이 알고리즘 선택을 결정하는 요인인지 여부

6. 원하는 예측 속도
   - 실시간 예측 필요 여부
   - 배치 처리 허용 여부

DO NOT COPY
ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



일반화된 선형 모델: 선형/로지스틱 회귀와 SVM 등 해석이 가능한 고속 모델입니다. 활용 가능한 방식은 더 복잡한 모델에 비해 적습니다.

트리 기반 모델: 부스팅 트리, 랜덤 포리스트, 의사결정 트리 등이 포함되는 카테고리입니다. 선형 모델에 비해 정확도는 높지만 훈련 시간이 약간 더 오래 걸립니다.

신경망: MLP, CNN, RNN, 트랜스포머 등의 신경망은 복잡한 관계를 모델링할 수는 있지만 대규모 데이터세트를 사용해야 하고 훈련 속도가 느리며 해석하기도 더 어렵습니다.

클러스터링: K-평균 및 DBSCAN과 같은 알고리즘은 그룹 내의 유사성은 높아지고 그룹 간의 유사성은 낮아지도록 데이터를 그룹화합니다. 이러한 알고리즘은 탐색적 분석 및 비지도 학습에 주로 사용됩니다.

행렬 인수 분해: 추천 시스템, 차원 감소, 특성 학습 등에 사용되는 PCA와 같은 기법이 포함됩니다.

예측: ARIMA, DeepAR 등의 시계열 모델은 과거 데이터를 토대로 향후 값을 예측하는 데 사용됩니다.

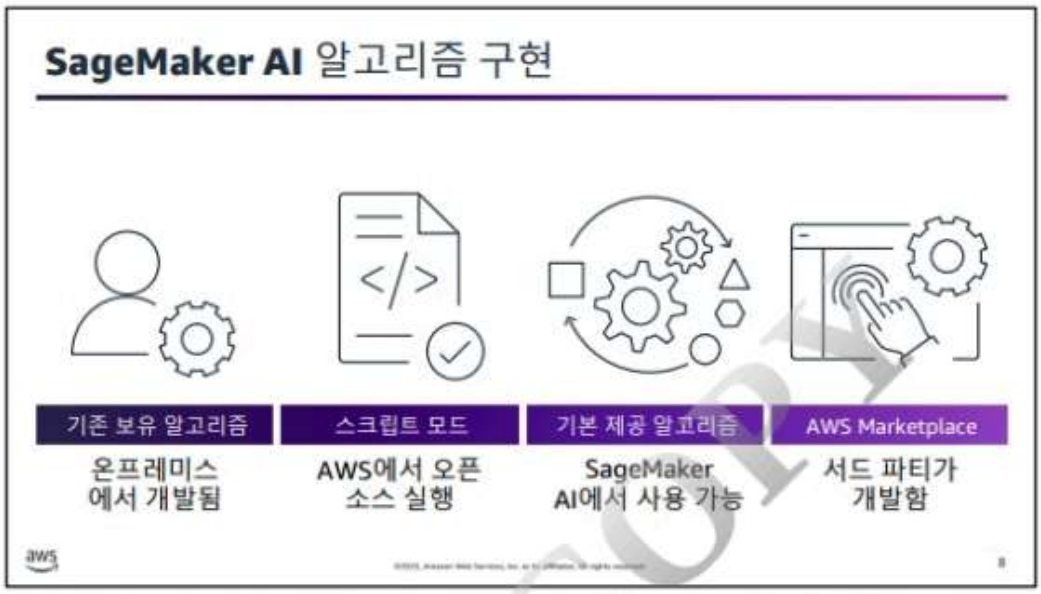
컴퓨터 비전 모델: CNN, ResNet 및 비전 트랜스포머는 이미지와 비디오 처리, 분류, 탐지, 세분화 태스크를 처리하는 특수 모델입니다.

시퀀스 모델: RNN, LSTM 및 트랜스포머는 특히 NLP 및 시계열 분석에서 순차 데이터를 처리하는 데 적합합니다.

SageMaker AI에서는 정형 데이터 분석, 컴퓨터 비전, NLP, 시계열 예측 등의 여러 사용 사례에 활용할 수 있는 다양한 모델 유형이 제공됩니다. 예를 들어 이미지 및 비디오 처리용 CNN, ResNet 및 비전 트랜스포머가 제공되므로 분류, 객체 탐지, 세분화 등의 태스크를 수행할 수 있습니다. ResNet은 잔여 학습 프레임워크를 통해 심층 신경망을 훈련할 수 있게 되어 이미지 인식 성능이 더욱 개선되었습니다.

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



SageMaker AI로 ML 모델을 구축할 때는, 훈련 알고리즘 구현 방법을 선택할 수 있는 다양한 옵션이 있습니다. 이러한 옵션을 통해 특정 요구 사항에 가장 적합한 접근 방식을 유동적으로 선택할 수 있습니다.

- **기존 보유 알고리즘 사용:** 사내에서 개발한 전용 알고리즘이나 모델이 있다면 SageMaker AI로 가져온 후 훈련, 배포 및 관리 기능을 사용할 수 있습니다.
- **스크립트 모드:** 자체 사용자 정의 코드를 사용하거나, TensorFlow, PyTorch, Apache MXNet 등의 널리 사용되는 오픈 소스 프레임워크를 사용하려는 경우에는 SageMaker AI의 관리형 환경 내에서 스크립트를 작성한 후 실행할 수 있습니다.
- **기본 제공 알고리즘:** SageMaker AI에서는 지도 학습, 비지도 학습, 컴퓨터 비전, 자연어 처리 같은 다양한 ML 태스크에 사용할 수 있는 사전 구축 및 최적화된 알고리즘을 제공합니다.
- **AWS Marketplace:** SageMaker AI는 AWS Marketplace에 통합되므로 업계의 유명 기업과 연구 기관을 비롯한 여러 공급자가 개발한 서드 파티 알고리즘과 모델 패키지를 구독 및 사용할 수 있습니다.

이처럼 다양한 옵션이 제공되므로 구체적인 사용 사례, 기존 코드베이스 및 전문 지식에 가장 적합한 방식을 선택할 수 있습니다.

DO NOT COPY
ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

알고리즘 선택을 위한 고려 사항

- 문제 유형
- 데이터 특성
- 성능 요구 사항
- 훈련 시간 및 컴퓨팅 리소스
- 해석 가능성
- 모델 복잡성
- 확장성
- 도메인 지식

적절한 ML 알고리즘 선택은 정확하고 효율적인 모델을 구축하는 과정에서 꼭 필요한 단계입니다. SageMaker AI에서 알고리즘을 선택할 때 고려해야 하는 몇 가지 핵심 요인은 다음과 같습니다.

- **문제 유형** – 첫 번째 단계는 분류, 회귀, 클러스터링, 추천 시스템 등 해결하려는 문제의 유형을 식별하는 것입니다. 어떤 알고리즘은 특정 문제 유형에 더 적합합니다. 문제 특성을 파악하면 알고리즘을 적절하게 선택할 수 있습니다.
- **데이터 특성** – 데이터를 분석하여 크기, 차원, 노이즈 수준, 잠재적 편향을 파악해야 합니다. 이 정보를 파악하면 데이터 특성을 효율적으로 처리할 수 있는 적절한 모델 아키텍처와 알고리즘을 선택하는데 도움이 됩니다.
- **성능 요구 사항** – 정확도, 정밀도, 재현율 또는 F1-점수 등 모델에 필요한 성능 지표를 고려해야 합니다. 일부 알고리즘은 특정 지표를 다른 지표보다 우선시할 수 있으므로 구체적인 성능 목표에 따라 적절한 알고리즘을 선택해야 합니다.
- **훈련 시간 및 컴퓨팅 리소스** – 모델마다 훈련 단계 중의 계산 요구 사항은 다릅니다. 알고리즘을 선택할 때는 CPU, 그래픽 처리 장치(GPU), 메모리 등의 사용 가능한 리소스와 프로젝트의 시간 제약을 고려해야 합니다.
- **해석 가능성**: – 사용 사례에서 해석 가능성이 중요한 경우 해석과 설명이 더 까다로울 수 있는 복잡한 딥 러닝 모델보다 선형 또는 트리 기반 알고리즘처럼 더 단순한 모델을 선택하는 것이 좋습니다.

이러한 요인을 신중하게 고려하면 ML 태스크에 가장 적합한 알고리즘을 선택하여 모델의 성능과 효율성을 최적화함으로써 비즈니스 목표에 부합하는 모델을 구축할 수 있습니다.

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

적절한 ML 모델 선택은 프로젝트의 성공에 영향을 좌우하는 중요한 결정 사항입니다. 몇 가지 주요 고려 사항이 선택의 지침이 됩니다.

중요한 역할을 하는 사항은 **데이터 특성**입니다. 데이터세트의 크기, 차원 및 품질에 따라 실현 가능하고 성능이 좋을 것 같은 모델이 결정되기 때문입니다. 예를 들어, 딥 러닝 모델의 경우 데이터가 많이 필요한 경우가 많은 반면, 소규모 데이터 세트는 단순한 모델이 더 적합할 수 있습니다.

성능 요구 사항 역시 반드시 고려해야 하는 사항입니다. 특성 사용 사례에서 가장 중요한 지표 (정확도, 정밀도, 재현율 등) 그리고 다양한 모델 유형에 어떻게 맞춰지는지 생각해야 합니다. 단, **해석 가능성** 및 **계산 리소스** 등의 다른 고려 사항과 성능의 균형을 맞추는 것이 중요합니다.

해석 가능성은 의료, 금융 등 모델의 의사결정 프로세스를 이해하는 것이 필수적인 영역에서 특히 중요한 고려 사항입니다. 의사결정 트리와 같은 모델은 해석 가능성이 높은 반면 신경망 등의 모델은 해석이 까다로울 수 있습니다.

학습 및 추론에 사용할 수 있는 **컴퓨팅 리소스**도 고려해야 합니다. 높은 수준의 처리 능력이나 많은 메모리가 필요한 모델도 있으며, 이러한 모델을 선택하는 경우 배포 옵션과 비용이 달라질 수도 있습니다.

요약하자면, 최적 모델을 선택하려는 경우 이러한 요인의 적절한 절충점을 찾아야 하는 경우가 많습니다. 그리고 최적 모델은 프로젝트의 구체적인 요구 사항 및 제약 조건에 따라 달라집니다.

DO NOT COPY ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

SageMaker Autopilot

SageMaker Autopilot

자동 기계 학습(AutoML) 단순화

- 신속한 ML 모델 개발 진행
- 최소한의 ML 관련 전문 지식
- 투명한 ML 파이프라인 제공
- 자동화를 통해 비용 최적화 가능
- 테이블 형식 데이터로 회귀 및 분류 문제 해결

11

Amazon SageMaker Autopilot의 강력한 자동화 기능이 SageMaker AI 시스템 전반에 긴밀하게 통합되어 있으므로 여러 인터페이스에서 고급 ML 기술에 액세스할 수 있습니다.

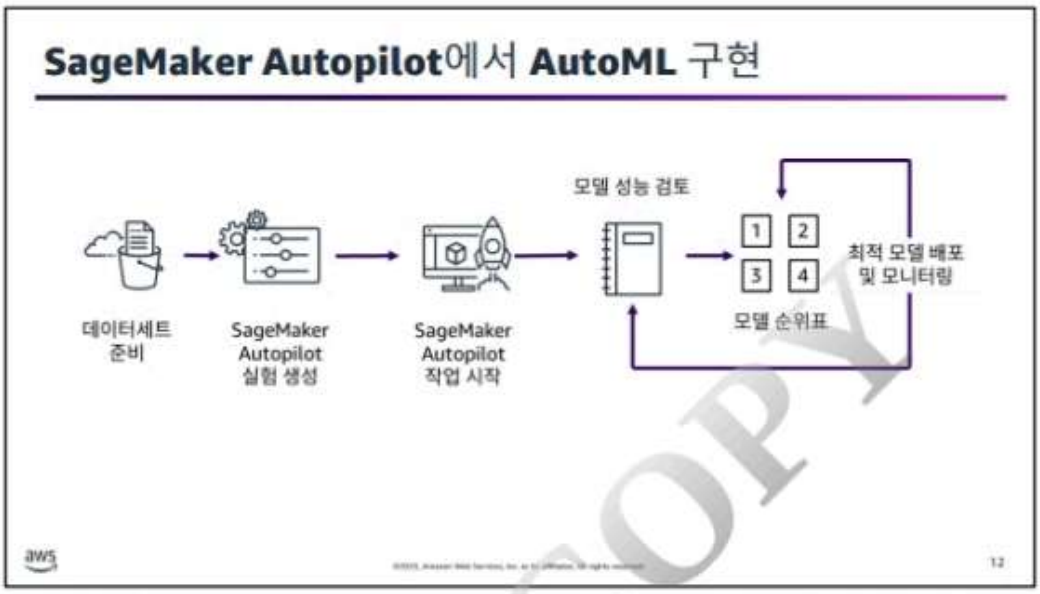
Amazon SageMaker Canvas에서, SageMaker Autopilot은 노코드 방식 모델 구축 환경을 구동합니다. 그래서 비즈니스 사용자는 코드를 전혀 작성하지 않고도 시각적 인터페이스를 통해 예측을 생성할 수 있습니다. Amazon SageMaker Studio 에서 데이터 과학자는 Notebook을 통해 SageMaker Autopilot 작업을 바로 시작할 수 있으며 SageMaker Python SDK를 통해서도 동일한 기능을 프로그래밍 방식으로 사용할 수 있습니다. 이러한 **통합 자동 기계 학습(AutoML)** 엔진은 어떤 진입점을 선택하던 관계없이 피처 엔지니어링, 알고리즘 선택 및 하이퍼파라미터 최적화를 처리하여 ML 모델 개발 속도를 대폭 높일 수 있습니다.

SageMaker Autopilot을 사용하면 기계 학습 분야의 전문 지식이 없는 팀도 프로덕션급 모델을 구축할 수 있습니다. 알고리즘 선택 또는 특성 변환 방법을 잘 모르는 비즈니스 분석가, 개발자 및 도메인 전문가도 직관적인 인터페이스를 통해 AutoML 기능을 사용할 수 있습니다. 이러한 민주화 덕분에, 조직은 이전에는 기술 리소스를 제대로 활용하지 못했던 부서에도 기계 학습을 적용할 수 있습니다.

블랙 박스 솔루션과는 달리, SageMaker Autopilot은 전체 ML 파이프라인을 문서화되는 Notebook을 생성하여 완전한 투명성을 제공하므로 사용자가 ML 프로세스를 쉽게 이해하고 사용자 지정할 수 있습니다. 이 시스템은 리소스 할당을 자동으로 최적화하고 실험을 병렬 처리하므로 수동 방식에 비해 비용이 대폭 감소합니다. 배포할 준비가 되면, 추론 코드 및 스케일링 구성을 갖춘 된 모델은 클릭 한 번으로 프로덕션 엔드포인트로 푸시할 수 있으므로 기계 학습 모델을 운영하려면 수행해야 했던 기존의 복잡한 작업을 수행할 필요가 없습니다.

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



SageMaker Studio에서 Amazon SageMaker Autopilot을 사용하여 AutoML을 구현할 때 수행해야 하는 대략적인 단계는 다음과 같습니다.

1. **데이터 준비**: 쉼표로 구분된 값(CSV) 또는 Parquet 형식의 테이블 형식 데이터를 Amazon Simple Storage Service(Amazon S3)에 업로드하거나 로컬 파일에서 SageMaker Studio로 해당 데이터를 직접 가져옵니다.
2. **SageMaker Autopilot 실험 생성**: SageMaker Studio에서 데이터세트 위치와 예측할 대상 열을 지정하고 원하는 경우 문제 유형(분류 또는 회귀)을 선택합니다. 그리고 최대 런타임 및 리소스 제한 등의 추가 설정을 구성합니다.
3. **시작 및 모니터링**: 실험을 시작한 후 SageMaker Studio 인터페이스를 통해 진행 상황을 추적합니다. SageMaker Autopilot이 자동으로 데이터를 분석하여 특성을 생성하고 알고리즘을 선택한 후 하이퍼파라미터를 최적화합니다.
4. **후보 탐색**: 실험 대시보드에서 여러 모델 후보의 성능 지표를 검토합니다. 자동 생성된 데이터 탐색 및 모델 설명 Notebook에서 인사이트를 파악합니다.
5. **최적 모델 배포**: 클릭 몇 번으로 선택한 모델을 SageMaker AI 엔드포인트에 배포할 수도 있고 배치 처리를 위해 예측을 내보낼 수도 있습니다.

SageMaker Studio 내에서 SageMaker Autopilot을 사용하면 프로세스의 각 단계를 문서화하는 생성된 Notebook의 투명성과 함께 통합 데이터 과학 환경 내에서 자동 ML의 이점을 얻을 수 있습니다. SageMaker Canvas 내에서도 이와 비슷한 프로세스가 사용됩니다.

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

Amazon SageMaker Autopilot에 액세스

- **SageMaker Canvas**에서 시각적으로 모델 생성
- **SageMaker Studio**에서 실험 구축
- **Amazon SageMaker Pipelines**와 워크플로 통합
- **SageMaker AI Python SDK** 및 **SageMaker AI API** 사용
- **Amazon SageMaker Model Monitor**를 사용하여 모니터링

여러 SageMaker AI 기능에서 Amazon SageMaker Autopilot을 실행할 수 있습니다.

SageMaker Canvas는 SageMaker Autopilot의 자동 기계 학습 기능이 통합된 노코드 방식 시각적 인터페이스가 제공되므로 코드를 작성해 본 적이 없는 비즈니스 분석가와 도메인 전문가도 모델을 구축할 수 있습니다. 사용자는 데이터를 업로드하고 예측 대상을 지정할 수 있으며, 그러면 SageMaker Canvas는 SageMaker Autopilot을 사용하여 최적화된 모델을 생성합니다.

SageMaker Studio에서는 데이터 과학자와 개발자가 활용할 수 있는 포괄적인 통합 개발 환경(IDE)이 제공됩니다. 이 IDE에는 SageMaker Autopilot 실험 생성, 관리 및 분석용 시각적 인터페이스가 포함되어 있습니다. 모델 순위표와 성능 지표가 표시되는 SageMaker Studio 대시보드에서는 자동 생성된 Notebook 및 모델 설명에 액세스할 수 있습니다.

Amazon SageMaker Pipelines에서는 SageMaker Autpilot 작업을 자동화된 ML 워크플로의 단계로 통합할 수 있습니다. 이렇게 하면 데이터 준비에서 모델 배포까지의 전체 ML 수명 주기를 오케스트레이션할 수 있으며, 모델 구축 단계는 SageMaker Autopilot이 처리합니다.

프로그래밍 방식 액세스 시에는 SageMaker AI Python SDK에서 제공되는 AutoML 클래스를 사용하면 SageMaker Autopilot 실험을 세부적으로 제어할 수 있습니다. 반면 SageMaker AI API를 사용하면 지속적 통합 및 전달(CI/CD) 시스템과 사용자 정의 애플리케이션에 AutoML 작업을 통합할 수 있습니다.

모델 배포 후에는, Amazon SageMaker Model Monitor를 사용하여 데이터 및 예측 드리프트를 추적할 수 있습니다. 그러면 SageMaker Autopilot에서 생성된 모델의 품질을 프로덕션 환경에서도 유지할 수 있습니다. 드리프트가 감지되면 새 SageMaker Autopilot 실험을 시작하여 모델을 재훈련할 수 있습니다.

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

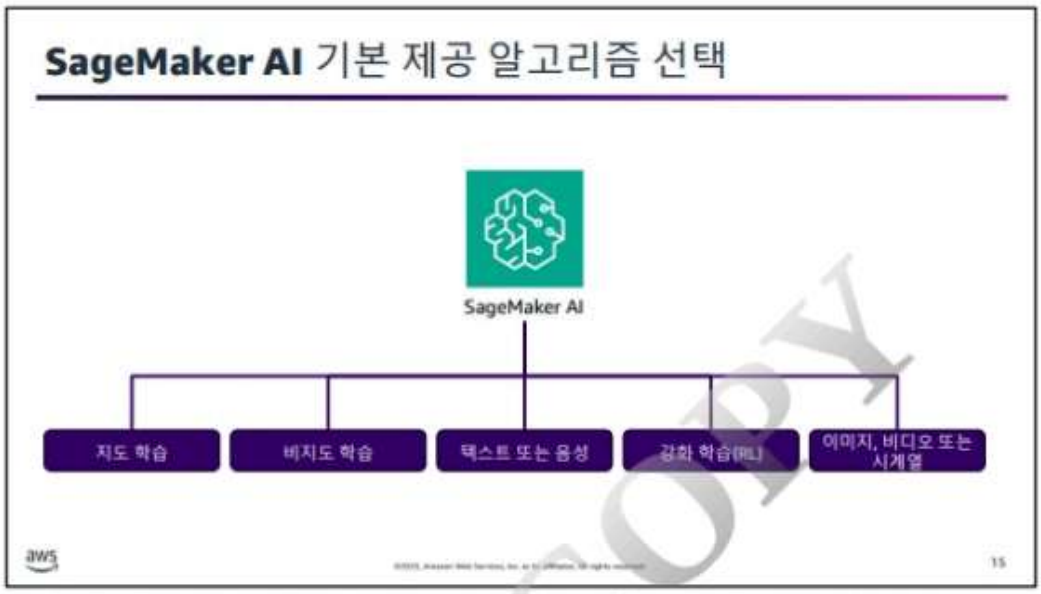
데이터 탐색과 분석이 완료되면 ML 모델을 구축할 수 있습니다. 적절한 훈련 알고리즘 선택은 이 프로세스의 중요한 단계입니다. 훈련 알고리즘은 모델의 성능, 정확도, 그리고 구체적인 사용 사례에서 모델이 적합한지 여부에 직접적인 영향을 주기 때문입니다.

SageMaker AI는 데이터 과학자와 ML 실무자가 모델 구축을 빠르게 시작할 수 있도록 기본 제공 알고리즘, 사전 훈련된 모델 및 사전 구축된 솔루션 템플릿이 포함되어 있는 포괄적인 모음이 제공됩니다. 이러한 기능을 사용하면 알고리즘을 광범위하게 사용자 지정하거나 개발하지 않고도 ML 모델을 효율적으로 훈련해 배포할 수 있습니다.

이러한 기본 알고리즘을 사용하는 경우 ML 모델 개발과 배포에 걸리는 시간을 대폭 단축하고 작업량도 줄일 수 있습니다.

DO NOT COPY
ssleeeee1@gmail.com

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



SageMaker AI 기본 제공 알고리즘은 효율적인 훈련과 배포를 위해 사전 구성 및 최적화되어 있으므로 모델을 자세히 사용자 지정할 필요 없이 빠르게 구축하여 배포할 수 있습니다.

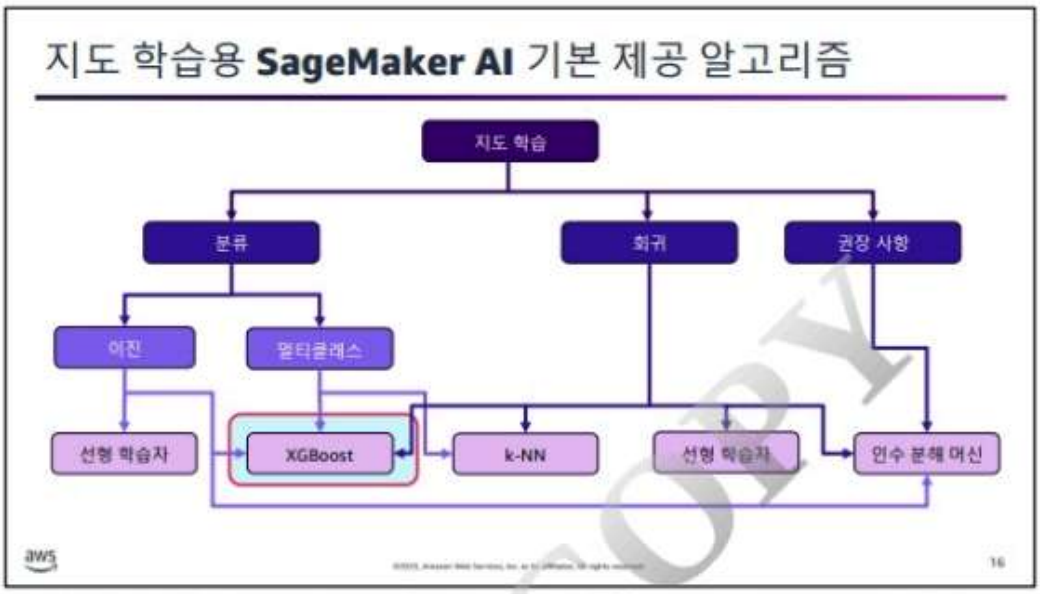
SageMaker AI 기본 제공 알고리즘은 여러 카테고리로 구성되어 있습니다. 구체적으로는 분류 및 회귀 태스크용 **지도 학습**, 클러스터링 및 차원 감소용 **비지도 학습, 텍스트 및 음성 분석, 컴퓨터 비전, 강화 학습** 등의 카테고리가 있습니다. 각 카테고리에서는 특정 사용 사례에 적합한 여러 알고리즘 옵션이 제공됩니다.

SageMaker AI에서는 기본 제공 알고리즘을 사용할 수 있을 뿐 아니라 기존 보유 사용자 정의 알고리즘 또는 AWS Marketplace의 사전 훈련된 모델도 사용할 수 있습니다. 이처럼 알고리즘을 유동적으로 선택할 수 있으므로 구체적인 업무상의 요구에 가장 적합한 솔루션을 선택할 수 있습니다.

SageMaker AI 기본 제공 알고리즘에 대한 자세한 내용은 **Amazon SageMaker AI 개발자 가이드**의 'Amazon SageMaker의 기본 제공 알고리즘 및 사전 훈련된 모델'(https://docs.aws.amazon.com/sagemaker/latest/dg/algos.html)을 참조하십시오.

DO NOT COPY
ssleeeee1@gmail.com

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



SageMaker AI는 개발자가 분류 및 회귀 문제와 같은 지도 학습 태스크에 사용할 수 있는 몇 가지 기본 제공 범용 알고리즘을 제공합니다. 이러한 알고리즘은 훈련과 배포를 효율적으로 진행할 수 있도록 사전 구성 및 최적화되어 있으므로 광범위하게 사용자 지정할 필요 없이 빠르게 구축하여 배포할 수 있습니다.

사전 훈련된 모델 및 솔루션 템플릿

Amazon SageMaker JumpStart에서는 이미지 분류, 테이블 형식 분류와 회귀, 텍스트 분석, 객체 탐지, 자연어 처리 등 인기 있는 문제 유형에 대해 여러 가지 사전 훈련된 모델과 사전 구축된 솔루션 템플릿이 제공됩니다. 이와 같은 사전 훈련된 모델과 템플릿은 프로젝트의 출발점이 되어 개발 프로세스를 빠르게 진행할 수 있습니다.

지도 학습 알고리즘

선형 모델, 트리 기반 모델, 신경망, 앙상블 메서드 등의 광범위한 모델 유형에 다음과 같은 기본 제공 알고리즘을 사용할 수 있습니다.

특정 사용 사례 및 데이터 특성에 가장 적합한 알고리즘을 선택할 수 있습니다:

- **AutoGluon-Tabular**: 분류 및 회귀용 앙상블 모델 (앙상블 학습)
- **CatBoost**: 순차적 부스팅으로 그래디언트 부스팅 트리 구현 (그래디언트 부스팅 트리)
- **인수 분해 머신**: 고차원 희소 데이터세트에서 상호 작용 캡처 (인수 분해 머신)
- **k-최근접 이웃(k-NN)**: 레이블이 지정된 가장 가까운 점을 분류와 회귀에 사용 (인스턴스 기반 학습)
- **라이트 그래디언트 부스팅 머신(LightGBM)**: 그라디언트 기반 단면 샘플링(GOSS) 및 독점적인 기능 번들링 (EFB)이 적용된 효율적인 그래디언트 부스팅 트리 적용 (그래디언트 부스팅 트리)
- **선형 학습기**: 회귀 또는 분류용 선형 함수 사용(선형 모델)
- **TabTransformer**: 트랜스포머에 심층 테이블 형식 데이터 모델링 활용(트랜스포머)
- **eXtreme Gradient Boosting(XGBoost)**: 그래디언트 부스트 트리의 앙상블 구현 (그래디언트 부스팅 트리)

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

1단계. ML 영역과 문제 유형 파악

태스크	태스크 설명	문제 유형	SageMaker AI 기본 제공 알고리즘
A	소득이 임계값 미만인지를 예측합니다. CSV 데이터세트: • 인구 통계 • 소득	ML 영역과 문제 유형 파악	

17

가이드 활동 - 1단계

1. ML 문제의 유형을 식별합니다.

소득이 임곗값을 초과하는지를 예측합니다.

CSV 데이터세트에는 값이 0이나 1인 레이블이 지정된 열이 포함되어 있습니다.

DO NOT COPY
ssleeeee1@gmail.com

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

2단계. 문제 유형과 알고리즘 파악

태스크	태스크 설명	문제 유형	SageMaker AI 기본 제공 알고리즘
1	소득이 임계값 미만인지를 예측합니다. CSV 데이터세트: • 인구 통계 • 소득	지도 학습/이진 분류	해당 ML 문제 유형에 사용할 수 있는 SageMaker AI 기본 제공 알고리즘 목록 작성

Amazon SageMaker의 기본 제공 알고리즘 및 사전 훈련된 모델

18

이 안내 활동에서는 ML을 흔히 사용하는 태스크, 즉 개인의 소득이 특정 임계값 미만인지를 예측하는 태스크에 초점을 맞춥니다. 인구 통계 정보 및 소득 데이터가 포함된 CSV 데이터세트가 제공됩니다. 이 시나리오는 타겟 마케팅 또는 금융 서비스와 같은 여러 실제 애플리케이션에서의 전형적인 사용 사례입니다.

이 문제를 해결하려면 다음 단계를 진행해야 합니다.

1. ML 영역 및 문제 유형 식별: 제시된 정보를 분석하여 이것이 지도 학습인지 비지도 학습 문제인지, 분류인지 회귀 태스크인지 확인합니다.
2. 관련 SageMaker AI 기본 제공 알고리즘 나열: 제공된 QR 코드를 사용하여 SageMaker AI 설명서에 액세스한 후 파악한 문제 유형에 적합한 알고리즘을 파악합니다.

알고리즘 선택 시에는, 구체적인 사용 사례와 관련이 있는 데이터 특성, 모델 해석 가능성 요구 사항, 성능 지표 등 요소를 고려해야 합니다. 이 프로세스는 SageMaker AI에서 효율적인 ML 솔루션을 구축하는 데 중요합니다.

DO NOT COPY
ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

해결 방법 - 가이드 활동

태스크	태스크 설명	문제 유형	SageMaker AI 기본 제공 알고리즘
1	소득이 임계값 미만인지를 예측합니다. CSV 데이터세트: • 인구 통계 • 소득	지도 학습/이진 분류	샘플 알고리즘: • CatBoost • 선형 학습자 • XGBoost

19

가이드 활동의 정답에 따르면 소득이 임계값보다 많은지를 예측하는 태스크는 지도 학습 문제이며, 구체적으로는 이진 분류 태스크입니다. 데이터(소득 값)에 레이블이 지정되어 있으며 2개 결과 중 하나(임계값 초과 또는 미만)를 예측할 것이기 때문입니다.

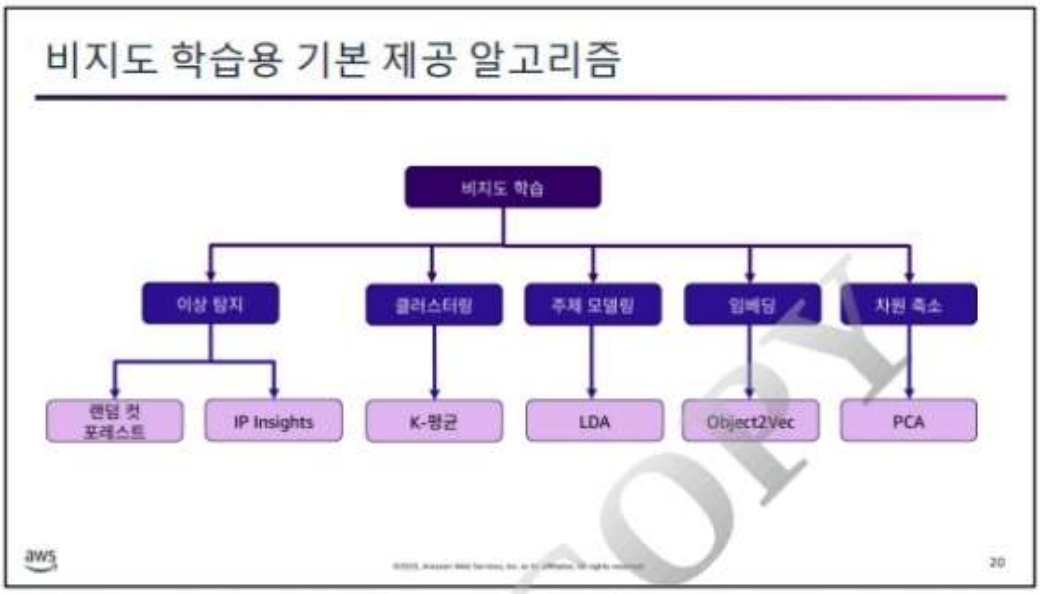
이 태스크에 적합한 샘플 알고리즘은 다음과 같습니다.
- CatBoost: 범주형 특성을 처리할 때 효율적입니다.
- 선형 학습기: 선형으로 분리 가능한 데이터 및 해석 가능한 모델에 적합합니다.
- XGBoost: 데이터 관계가 복잡할 때 효율적입니다.

이러한 알고리즘이 다양한 데이터 유형을 처리할 수 있으며 유사한 태스크에서 성능이 검증되었기 때문에 SageMaker AI의 이진 분류에 적합합니다. 이러한 알고리즘 중 하나를 선택할 때는 데이터세트 크기, 특성 유형, 모델 설명 가능성 요구 사항, 사용 가능한 컴퓨팅 리소스 등의 요인을 고려해야 합니다.

DO NOT COPY
ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



SageMaker AI는 개발자는 레이블이 지정되지 않은 데이터에서도 인사이트를 추출할 수 있도록 비지도 학습 태스크용 다양한 알고리즘이 기본 제공됩니다. 이러한 알고리즘은 여러 카테고리로 분류됩니다.

클러스터링에서는, K-평균 알고리즘이 유사한 데이터 포인트를 그룹화합니다. 주제 모델링은 문서 컬렉션에서 추상적인 주제를 발견하는 **잠재 디리클레 할당(LDA)**과 신경망을 사용하여 기존 방법에 비해 더 높은 정확도로 잠재적인 주제를 식별하는 **신경망 주제 모델(NTM)**을 사용할 수 있습니다. Object2Vec 알고리즘은 임베딩을 생성하여 고차원 공간의 객체 간 관계를 캡처합니다.

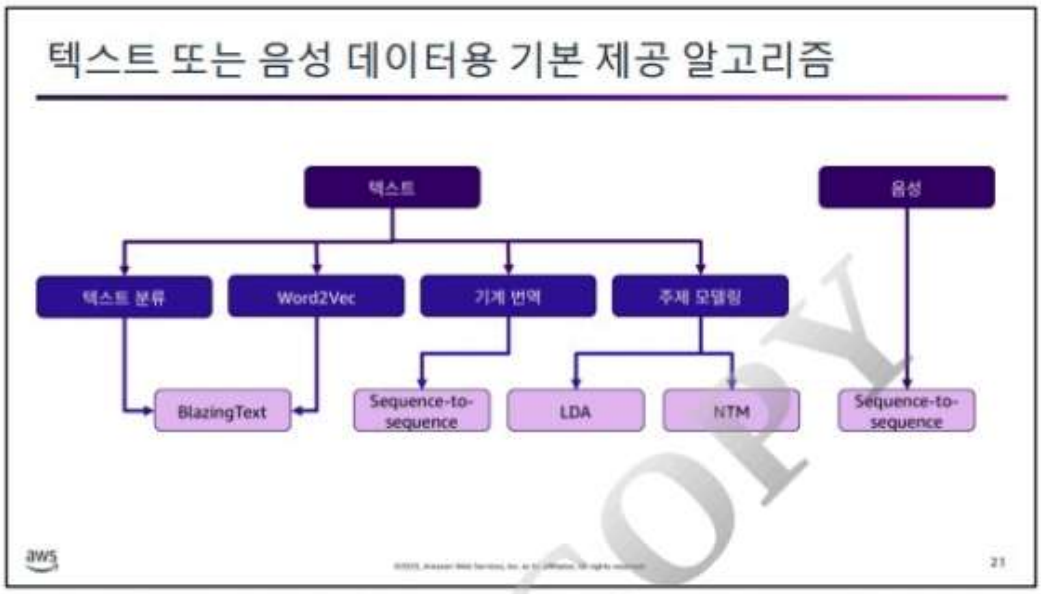
이상 탐지를 처리할 때는 데이터 스트림의 이상치를 식별하는 **랜덤 컷 포레스트(RCF)** 알고리즘과 IP 주소의 사용 패턴을 학습하는 **IP 인사이트**를 사용할 수 있습니다. 차원 축소 시, SageMaker AI는 중요한 정보는 그대로 유지하면서 특성 수를 줄이는 **주성분 분석(PCA)** 기능을 사용할 수 있습니다.

비지도 학습 알고리즘 선택 시에는 해결하려는 구체적인 문제와 데이터 구조 및 결과 해석 가능성 요구 사항을 고려해야 합니다. 비지도 기법은 탐색적 데이터 분석을 수행하고 복잡한 데이터세트에서 은닉 패턴을 찾아내려는 경우 특히 유용합니다.

DO NOT COPY
ssleeeee@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



SageMaker AI에서는 텍스트 및 음성 데이터 처리용 특수 알고리즘을 제공합니다, 개발자가 이를 사용하여 광범위한 자연어 처리(NLP) 태스크를 처리할 수 있습니다. SageMaker AI는 여러 가지 유용한 텍스트 분석용 옵션을 제공합니다.

BlazingText는 스킵 그램 및 Continuous Bag of Words(CBOW)와 같은 Word2Vec 및 텍스트 분류 알고리즘을 고도로 최적화한 구현입니다. BlazingText는 대규모 NLP 태스크에 적합합니다. Sequence-to-Sequence(seq2seq) 알고리즘은 다재다능해서 텍스트와 음성 데이터 모두에 대한 기계 번역과 텍스트 요약 같은 애플리케이션을 지원합니다.

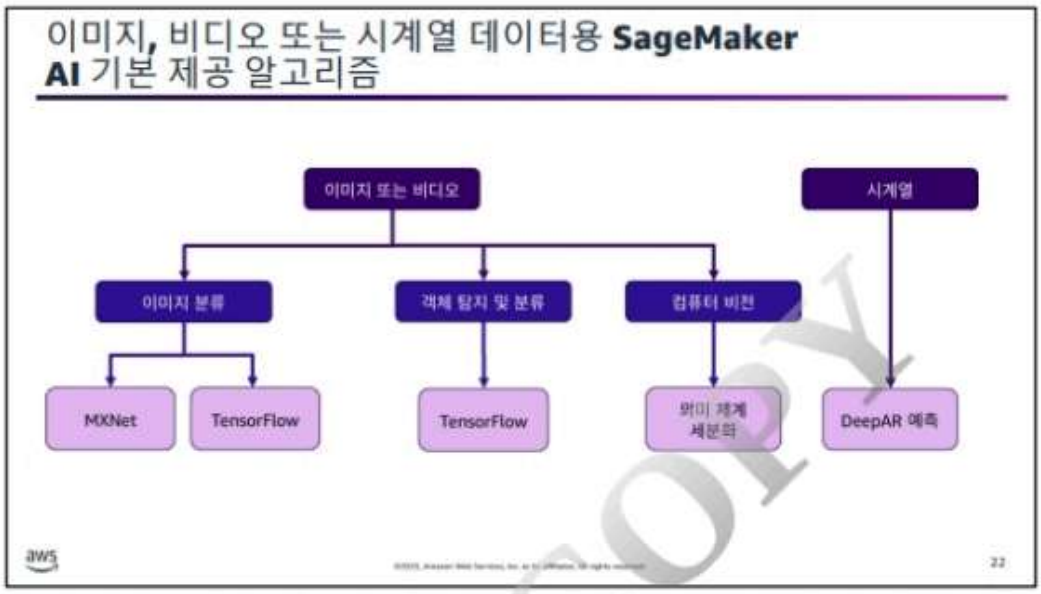
주제 모델링을 위해서 SageMaker AI에서 제공되는 2가지 옵션은 LDA 및 NTM입니다. LDA는 확률적 방식인 반면 NTM은 주제를 더욱 유연하게 검색하기 위해서 신경망을 사용합니다.

음성 데이터 사용 시에는 Sequence-to-Sequence 알고리즘을 음성 인식, 음성 합성 등의 태스크에 적용할 수도 있습니다. 텍스트 또는 음성 처리용 알고리즘을 선택할 때는 구체적인 NLP 태스크, 데이터세트의 크기와 특성, 모델 복잡도와 해석 가능성 간의 절충점과 같은 요인을 고려해야 합니다.

DO NOT COPY
ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



SageMaker AI에서는 이미지, 비디오 및 시계열 데이터 처리용 특수 기본 제공 알고리즘이 제공됩니다. 이러한 알고리즘은 다음과 같은 개별 카테고리로 구성되어 있습니다.

이미지 또는 비디오 처리:
- 딥 러닝 기반 분류용 MXNet 구현
- TensorFlow 구현, 대안적인 접근 방식 제공

객체 탐지 및 분류:
- 객체 식별 및 위치 파악용 MXNet 기반 탐지
- TensorFlow 기반 탐지, 또 다른 구현 옵션 제공

컴퓨터 비전:
- 픽셀 수준 이미지 파악용 의미 체계 세분화

시계열 분석:
- 시계열 데이터의 향후 값 예측을 위한 DeepAR 예측

이러한 알고리즘을 사용하면 다음과 같은 태스크용으로 다양한 ML 모델을 훈련할 수 있습니다.
- 이미지를 사전 정의된 카테고리로 분류
- 이미지 또는 비디오 내 객체 탐지 및 식별
- 상세한 장면 파악용으로 이미지 세분화
- 시계열 데이터로 정확한 예측 생성

구체적인 사용 사례와 선호하는 딥 러닝 프레임워크(MXNet 또는 TensorFlow)에 따라 적절한 알고리즘을 선택하면 됩니다.

DO NOT COPY ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

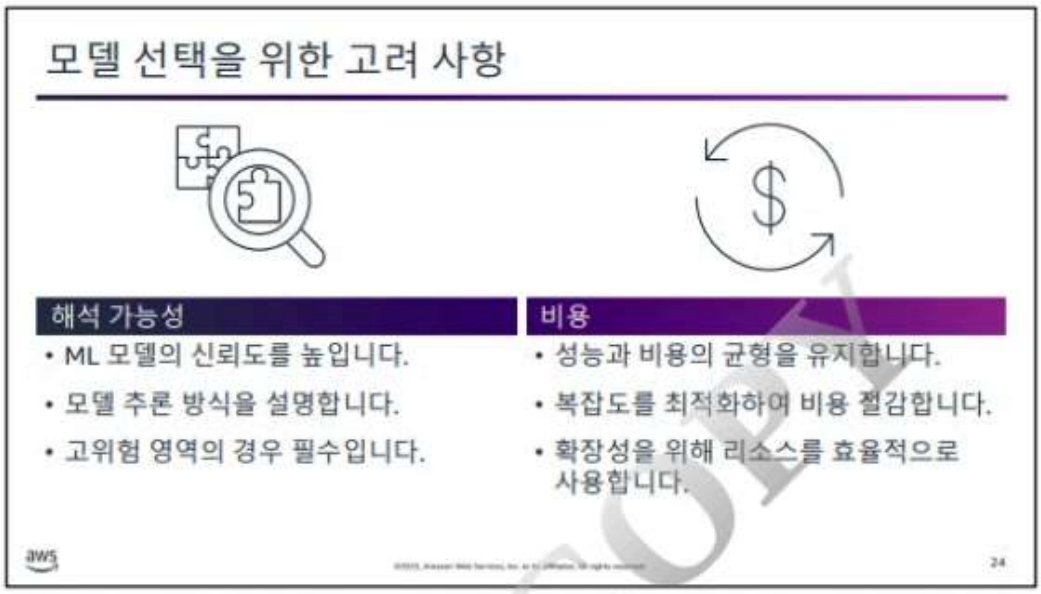
Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

진행 중인 기계 학습 태스크의 특성(특히 해당 태스크에서 사용하는 데이터의 특성, 거버넌스 및 규정 준수 요구 사항, 비용 제약 조건)에 따라 선택하는 모델과 리소스 요구 사항은 크게 달라질 수 있습니다.

DO NOT COPY
ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



해석 가능성은 기계 학습 모델의 예측 과정에 사용된 추론 과정을 이해 및 설명할 수 있는 능력을 지칭합니다. 특히 의료, 금융, 사법 등의 고위험 영역에서는 알고리즘을 선택할 때 해석 가능성을 반드시 고려해야 합니다.

투명성과 설명 가능성, 도메인 지식과 모델 복잡도, 규제 및 법적 요구 사항, 이해관계자의 신뢰 및 수용 여부 등 여러 요인이 해석 가능성의 필요성에 영향을 미칩니다. 의사결정 트리 및 선형 회귀와 같은 모델은 본질적으로 해석이 더 쉬운 반면, 복잡한 신경망은 투명성이 낮은 학습 모델인 블랙 박스로 간주되는 경우가 많습니다.

보통 모델의 해석 가능성과 예측 성능을 적절하게 절충점이 있습니다. 딥 러닝 모델은 정확도는 매우 높을 수 있지만 특정 응용 분야에서 필요한 투명도 요구 사항은 충족하지 못할 수도 있습니다. 그러므로 구체적인 사용 사례 및 요구 사항에 따라 이러한 요인의 균형을 신중하게 맞춰야 합니다.

모델 선택 시에는 애플리케이션에 필요한 해석 가능성 수준을 고려해야 합니다. 해석 가능성보다 정확도를 더 우선적으로 고려해야 하는 경우도 있는 반면, 모델의 결정을 설명하는 능력이 더 중요한 경우도 있기 때문입니다. 효율성과 신뢰도가 높은 기계 학습 솔루션을 구축하려면 이러한 균형을 이해하는 것이 필요합니다.

DO NOT COPY
ssleeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

해석 가능성은 기계 학습 모델의 예측 과정에 사용된 추론 과정을 이해 및 설명할 수 있는 능력을 지칭합니다. 특히 의료, 금융, 사법 등의 고위험 영역에서는 알고리즘을 선택할 때 해석 가능성을 반드시 고려해야 합니다.

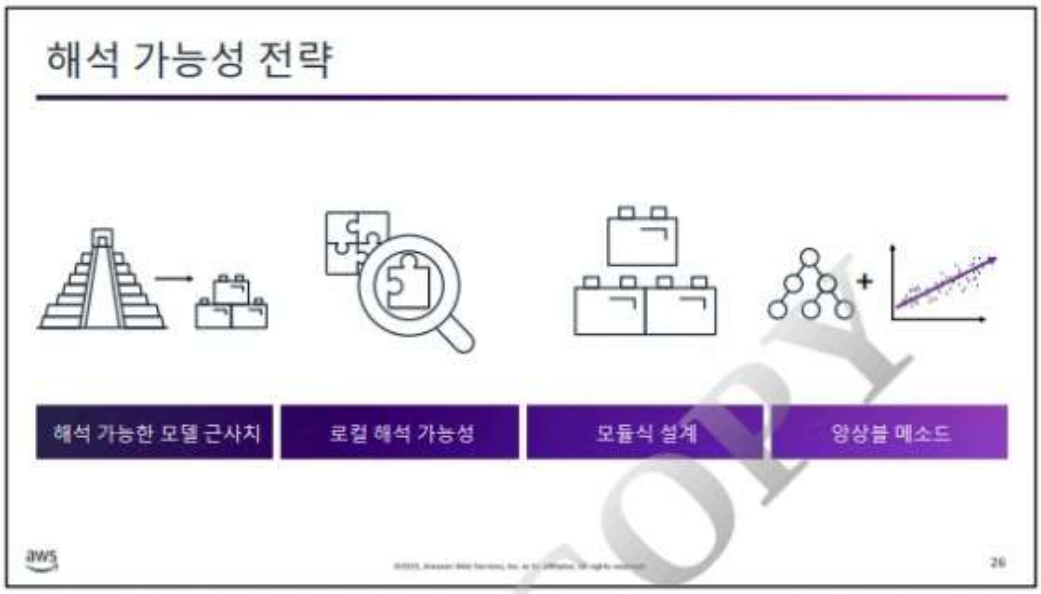
투명성과 설명 가능성, 도메인 지식과 모델 복잡도, 규제 및 법적 요구 사항, 이해관계자의 신뢰 및 수용 여부 등 여러 요인이 해석 가능성 필요 여부에 영향을 미칩니다. 의사결정 트리 및 선형 회귀와 같은 모델은 본질적으로 해석이 더 쉬운 반면, 복잡한 신경망은 투명성이 떨어지는 학습 모델인 블랙 박스로 간주되는 경우가 많습니다.

보통 모델의 해석 가능성과 예측 성능 사이에는 절충점이 있습니다. 딥 러닝 모델은 정확도는 매우 높을 수 있지만 특정 응용 분야에서 필요한 투명도 요구 사항은 충족하지 못할 수도 있습니다. 그러므로 구체적인 사용 사례 및 요구 사항에 따라 이러한 요인의 균형을 신중하게 맞춰야 합니다.

모델 선택 시에는 응용 분야에 필요한 해석 가능성 수준을 고려해야 합니다. 해석 가능성보다 정확도를 더 우선적으로 고려해야 하는 경우도 있는 반면 모델의 결정을 설명하는 능력이 더 중요한 경우도 있기 때문입니다. 신뢰도가 높고 효과적인 기계 학습 솔루션을 구축하려면 이러한 균형을 이해하는 것이 필요합니다.

DO NOT COPY
ssleeeee1@gmail.com

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



모델 해석 가능성과 성능의 균형을 맞추기 위해 개발자는 다양한 전략을 채택할 수 있습니다.

해석 가능한 모델 근사치에는 복잡한 블랙 박스 모델의 예측을 모방하도록 더 간단하고 해석 가능성한 모델을 훈련하는 것이 포함됩니다. 이것은 다음과 같은 기법을 활용할 수 있습니다.

- 지식 증류 (대형 모델에서 소형 모델로 지식 이전)
- 모델 근사화 (핵심 예측 동작이 그대로 유지된 단순화된 버전 생성)

Local Interpretable Model-agnostic Explanation(LIME), SHapley Additive exPlanations(SHAP) 등의 **로컬 해석 가능성** 방법을 사용하면 모델의 전체적인 동작은 불투명하더라도 모델이 데이터의 특정 인스턴스나 하위 집합에서 예측을 하는 방식과 관련된 인사이트는 파악할 수 있습니다. AWS는 Amazon SageMaker Clarify를 사용하여 SHAP를 통한 설명 가능성 기능을 제공합니다.

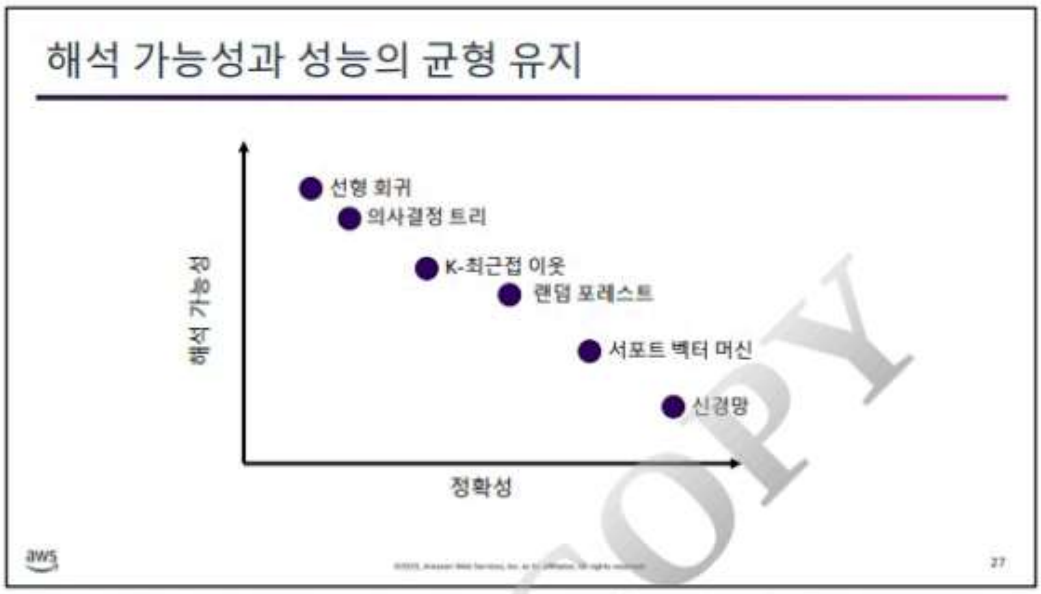
모듈식 설계 방식에서는 복잡한 모델을 더 작고 해석 가능한 구성 요소로 분해해서, 서로 다른 태스크에서 활용할 수 있도록 여러 모델의 장점을 결합합니다.

앙상블 메소드는 의사결정 트리, 규칙 기반 모델 등의 여러 해석 가능 모델을 결합합니다. 그러면 개별 구성 요소의 해석 가능성은 유지하면서 복잡한 모델급 성능을 제공하는 '앙상블'을 만들 수 있습니다.

DO NOT COPY
ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



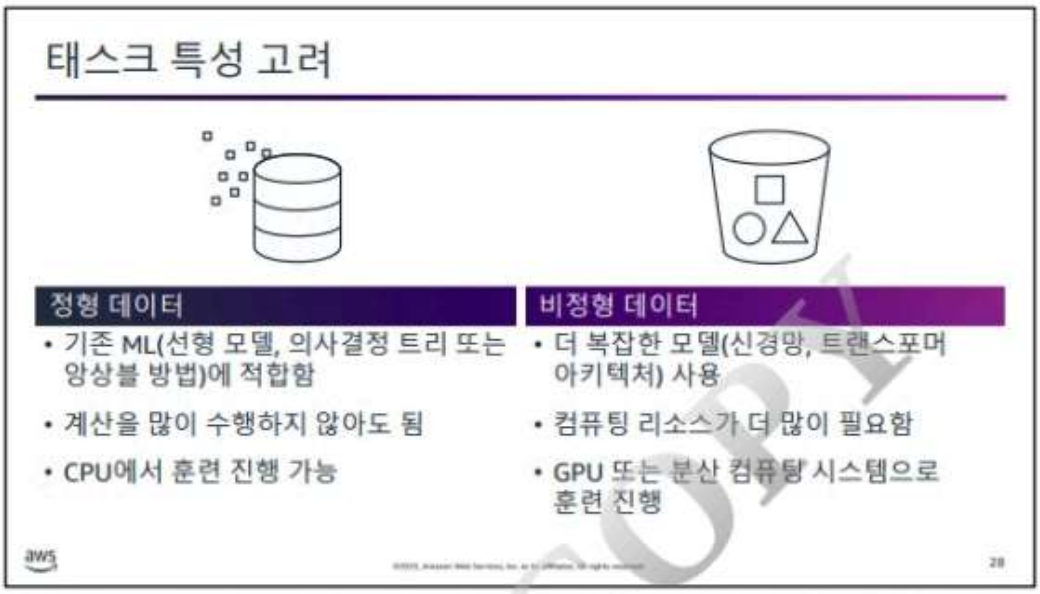
모델의 해석 가능성과 예측 정확도는 반비례하는 경우가 많습니다. 가령 회귀 등의 선형 모델은 해석 가능성은 높지만 문제가 복잡할 때는 정확도가 낮아질 수 있습니다. 반면 심층 신경망의 경우 성능은 매우 높을 수 있지만 일반적으로는 투명성이 낮은 블랙 박스로 간주됩니다.

구체적인 사용 사례와 요구 사항에 따라 해석 가능성과 성능 간의 적절한 균형을 유지해야 합니다. 모델의 결정이 매우 중요한 역할을 하는 의료, 금융 등의 영역에서는 미미한 정확도 향상보다 해석 가능성을 최우선으로 고려해야 할 수 있습니다. 하지만 정확도가 가장 중요하고, 이해관계자의 신뢰가 덜 중요한 애플리케이션의 경우에는 복잡한 블랙 박스 모델이 더 적합할 수 있습니다.

요약하자면, 해석 가능성과 성능의 절충 가능성을 평가하고 해결하려는 문제와 규제에 따른 제약 조건 및 이해관계자의 요구에 따라 두 요소 간의 균형을 적절하게 유지해야 합니다.

DO NOT COPY
ssleeeee1@gmail.com

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



진행 중인 기계 학습 태스크의 특성이 모델이나 알고리즘의 선택과 관련 계산 요구 사항에 영향을 미칠 수 있습니다.

테이블 형식 데이터, 시계열 데이터 등의 **정형 데이터를 사용하는 태스크**에서는 선형 모델, 의사결정 트리, 앙상블 메소드 등의 기존 기계 학습 알고리즘이 적합한 경우가 많습니다. 이러한 알고리즘은 계산 집약도가 낮고 대개 CPU에서 효율적으로 훈련할 수 있습니다.

반대로 이미지, 오디오 또는 텍스트와 같은 **비정형 데이터와 관련된 태스크**에는 심층 신경망 또는 트랜스포머 아키텍처와 같은 더 복잡한 모델이 필요할 수 있습니다. 이러한 모델은 계산이 많이 필요할 수 있으며 GPU 가속화 또는 분산 컴퓨팅 리소스를 활용하므로 전반적인 비용이 늘어날 수 있습니다.

적절한 모델을 선택하고 리소스 사용률 및 비용을 최적화하려면 데이터 특성 및 계산 요구 사항을 파악해야 합니다.

DO NOT COPY
ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

기계 학습 모델이나 알고리즘을 선택할 때는 모델 성능 및 해석 가능성과 더불어 비용도 반드시 고려해야 합니다. 효율적인 확장 가능 솔루션의 수요가 늘어남에 따라, 이제는 예측 성능과 비용 효율성이 모두 적절한 수준인 솔루션을 제공해야 합니다.

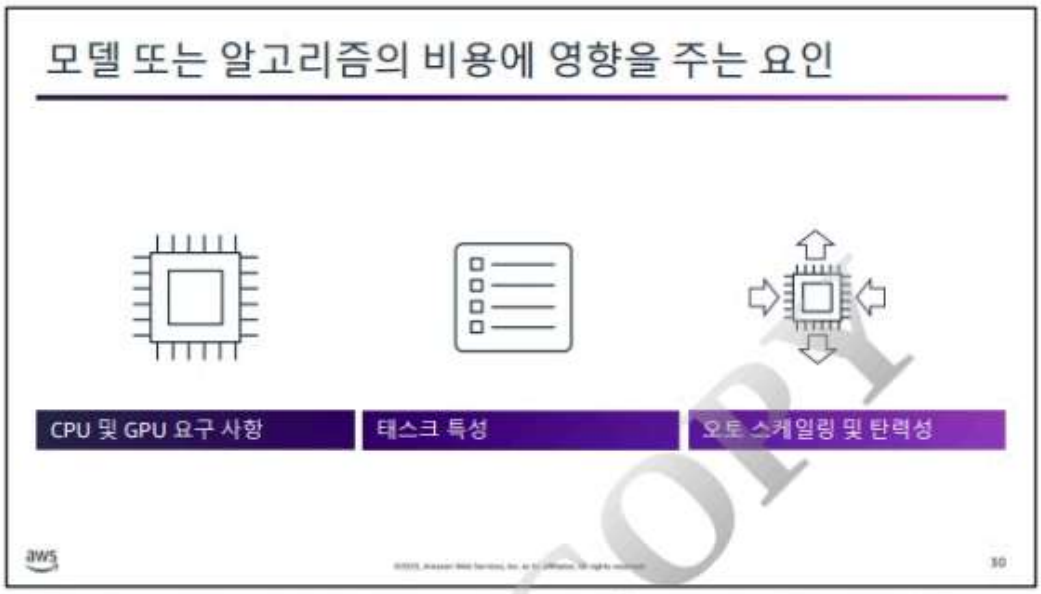
모델의 복잡도와 컴퓨팅 리소스 사용 방식을 최적화하면 운영 비용을 대폭 줄이는 동시에 적절한 성능 표준을 유지할 수 있습니다. CPU 및 GPU 요구 사항, 작업 특성, 오토 스케일링 및 탄력성 지원 등의 요인이 기계 학습 모델 배포와 실행 시의 전반적인 비용에 영향을 줄 수 있습니다.

이러한 요인을 면밀하게 평가하면, 조직은 구체적인 사용 사례에 가장 비용 효율적인 모델이나 알고리즘을 선택하고, 효율적인 리소스 활용을 보장하고, 비용을 최대한 절약하는데 있어 정보를 토대로 결정을 내릴 수 있습니다.

DO NOT COPY
ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



여러 가지 요인이 기계 학습 모델이나 알고리즘 배포 및 실행 비용에 영향을 줄 수 있습니다. 정보를 토대로 적절한 모델이나 알고리즘을 선택하고 비용을 최적화하려면 다음 사항을 고려해야 합니다.

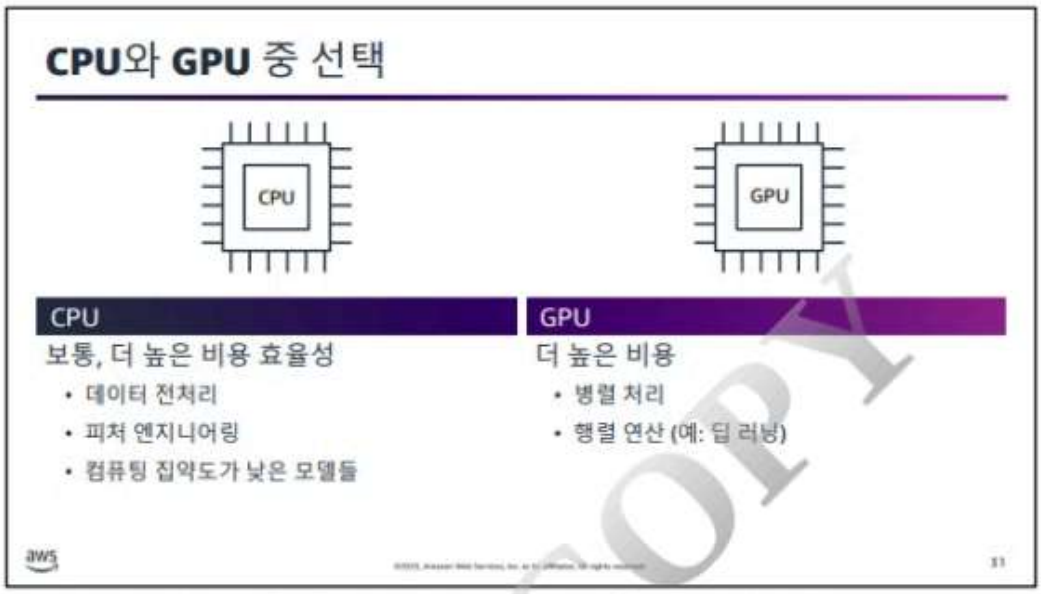
- **CPU 및 GPU 요구 사항**: GPU 가속이 지원되면 성능이 대폭 개선되어 더 쉽게 처리할 수 있는 태스크도 있지만 GPU 솔루션은 CPU 기반 솔루션에 비해 대개 비용이 더 많이 듭니다.
- **태스크 특성**: 문제의 복잡도, 데이터세트 크기, 실시간 추론 필요 여부 등 ML 태스크의 구체적인 요구 사항에 따라 필요한 계산 리소스가 달라질 수 있으며 필요한 리소스의 양에 따라 관련 비용도 늘어나거나 줄어들 수 있습니다.
- **오토 스케일링 및 탄력성**: 워크로드 수요에 따라 컴퓨팅 리소스를 동적으로 조정할 수 있는 경우 비용을 최소화할 수 있습니다. 리소스를 효율적으로 사용할 수 있으며 수요가 적은 기간 동안 불필요한 비용 발생을 방지할 수 있기 때문입니다.

이러한 요인을 면밀하게 평가하면 성능 요구 사항과 제약된 예산의 균형을 맞추면서 가장 비용 효율적인 모델이나 알고리즘을 선택할 수 있습니다.

DO NOT COPY
ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



기계 학습 모델이나 알고리즘을 배포할 때 결정해야 하는 중요한 사항 중 하나는 CPU 또는 GPU 중 컴퓨팅 워크로드에 사용할 프로세서를 선택하는 것입니다.

CPU는 일반적으로 더 비용 효율적이며 집중적인 계산 능력이 필요하지 않은 태스크에 적합합니다. 데이터 전처리, 피처 엔지니어링 및 의사결정 트리나 선형 모델 같은 특정 유형의 기계 학습 모델에는 CPU를 사용할 수 있습니다.

GPU는 병렬 처리용으로 설계되었으며 딥 러닝 및 기타 계산 집약적인 알고리즘과 같은 행렬 연산과 관련된 태스크에서 탁월한 성능을 발휘합니다. GPU는 대개 비용은 더 많이 들지만 성능은 훨씬 더 높을 수 있으므로 특정 사용 사례의 경우 장기적으로 비용 효율성이 더 높습니다. 프로젝트의 구체적인 요구 사항에 따라 GPU 사용 여부를 결정해야 합니다. 태스크에 대규모 데이터세트, 복잡한 모델 또는 실시간 추론이 포함되는 경우 GPU 리소스를 구매하면 초기 비용은 높지만 성능이 향상되므로 비용 효율성이 더 높아질 수 있습니다.

DO NOT COPY
ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

사용 사례 평가

- 분산 컴퓨팅 필요 여부를 결정합니다.
- 모델 또는 알고리즘을 최적화합니다.
- 모델 압축 또는 양자화와 같은 기법을 사용합니다.
- 전이 학습 또는 페더레이션 학습과 같은 대안적 접근 방식을 검토합니다.

32

ML 사용 사례를 평가할 때는 먼저 분산 컴퓨팅 필요 여부를 확인합니다. 이 단계에서는 데이터의 양과 모델 복잡도 및 훈련 시간 요구 사항을 평가하여 단일 인스턴스 훈련만 진행해도 충분한지 아니면 다중 노드 처리를 진행해야 하는지를 결정합니다.

다음에는, 모델과 알고리즘 최적화를 중점적으로 진행합니다. 구체적으로는 하이퍼파라미터를 미세 조정하고, 정확도와 계산 효율성의 균형을 맞추고, 솔루션이 제약된 리소스 내에서 성능 요구 사항을 충족하는지를 확인합니다.

모델 압축이나 양자화 등의 고급 최적화 기법을 구현할 수도 있습니다. 이러한 방식을 활용하면 모델 크기를 대폭 줄이고 효율성을 개선하는 동시에 적정 수준의 정확도를 유지할 수 있으므로 더욱 현실적이며 비용 효율적인 방식으로 모델을 배포할 수 있습니다.

마지막으로 전이 학습 또는 페더레이션 학습과 같은 대체 방식을 살펴봅니다. 특정 시나리오에서는 이러한 기법이 매우 유용할 수 있습니다. 특히 사용 가능한 데이터가 제한되어 있거나, 개인 정보 보호 관련 제약 조건이 있거나, 기존의 사전 훈련된 모델을 사용하여 개발을 신속하게 진행하려는 경우에는 이러한 기법이 효율적입니다.

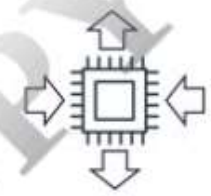
DO NOT COPY ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

분산 컴퓨팅 리소스 비용 절감

- 개발 및 실험용 절감형 플랜을 선택합니다.
- 온디맨드 또는 스팟 인스턴스를 사용합니다.
- 워크로드 수요를 충족하는 동시에 비용을 최적화할 수 있는 방식으로 용량 규모를 조정합니다.



대규모 기계 학습 태스크를 처리하거나 훈련 및 추론 프로세스를 신속하게 진행하려면 클러스터와 같은 분산 컴퓨팅 리소스를 사용해야 할 수 있습니다. 이러한 리소스를 사용하는 경우 성능은 크게 높아질 수 있지만 비용도 추가로 발생합니다.

비용 효율성을 최적화하려는 경우 **온디맨드 인스턴스** 사용을 고려할 수 있습니다. 이러한 인스턴스는 꼭 필요한 경우에만 실행되며 더 이상 필요하지 않을 때는 스케일 다운할 수 있습니다. 이러한 **오토 스케일링(탄력성)** 방식을 사용하면 워크로드 수요에 따라 컴퓨팅 리소스를 동적으로 조정하여 비용을 최소화하고 리소스를 효율적으로 활용할 수 있습니다.

오토 스케일링과 온디맨드 인스턴스를 활용하면 분산 컴퓨팅 리소스를 효율적으로 관리하여 불필요한 비용을 줄이는 동시에, 필요시에는 이러한 리소스가 제공하는 우수한 성능 이점도 활용할 수 있습니다.

DO NOT COPY
ssleeeee1@gmail.com

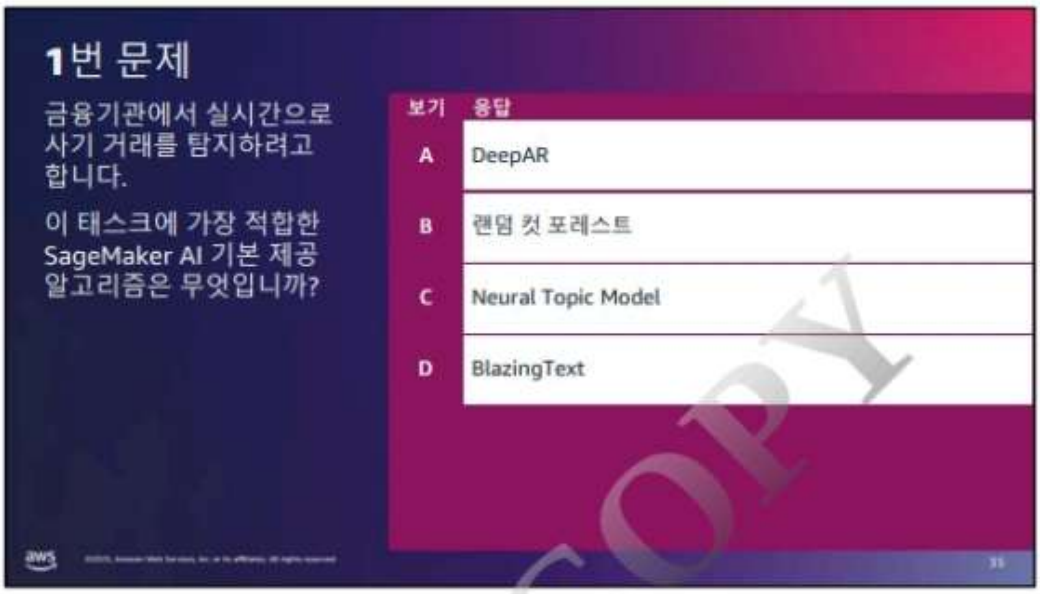
Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

이 섹션에서는 이 모듈에서 다루었던 주제를 정확하게 이해하고 있는지를 평가하는 지식 확인 문제를 풀어 봅니다. 이 모듈에서 학습한 개념을 적용해 보고 관련 자료를 정확하게 이해했는지를 확인할 수 있습니다.

DO NOT COPY
ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



DO NOT COPY
ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

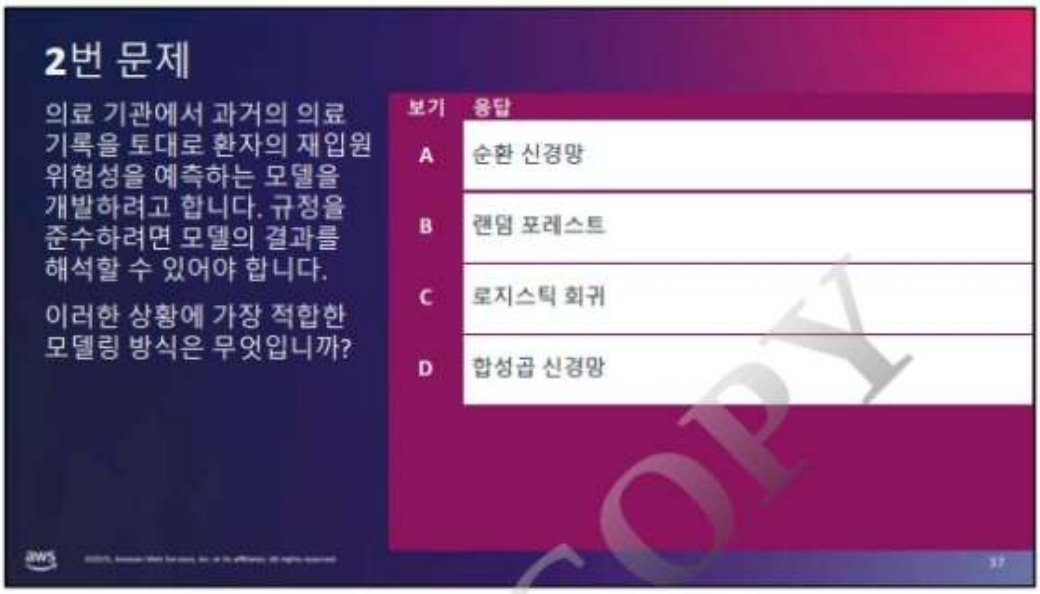
정답은 B, 랜덤 컷 포레스트입니다.

설명: 랜덤 컷 포리스트는 스트리밍 데이터에서 실시간으로 이상을 탐지하는 알고리즘입니다. 이 알고리즘은 거래 데이터의 비정상적인 패턴이나 이상치를 효율적으로 식별할 수 있으므로 사기 행위 탐지에 적합합니다.

오답:
A. DeepAR은 실시간 이상 탐지용이 아닌 시계열 예측용 알고리즘입니다.
C. Neural Topic Model은 사기 행위 탐지용이 아닌 텍스트 데이터의 주제 모델링용 알고리즘입니다.
D. BlazingText는 텍스트 분류 및 단어 임베딩용 알고리즘이며 거래 데이터에는 적합하지 않습니다.

DO NOT COPY
ssleeeee1@gmail.com

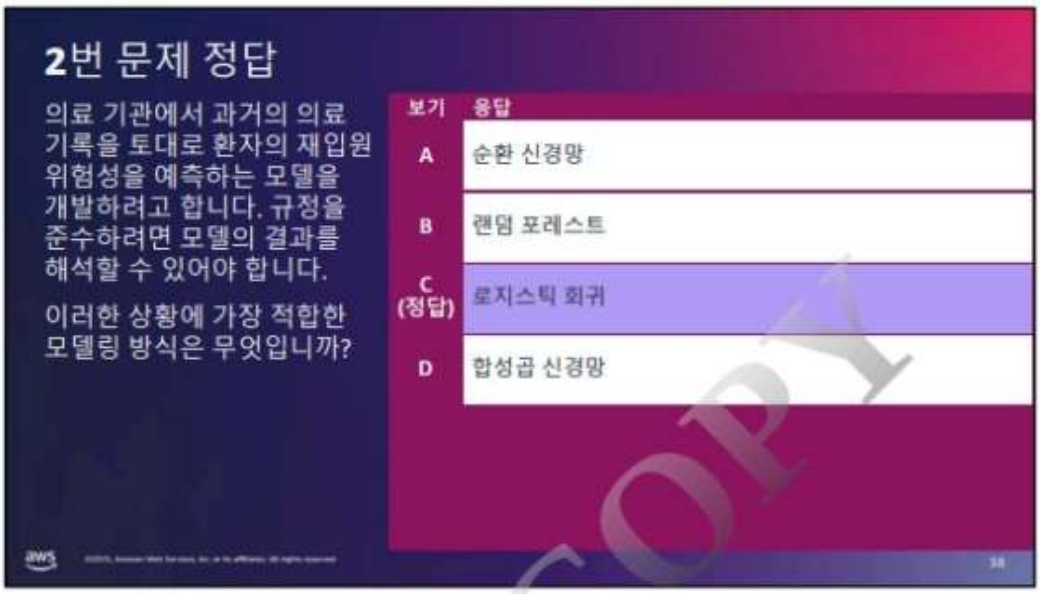
Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



모델 해석 가능성을 파악해야 하는 이유 및 특정 영역에서 모델 해석 가능성이 중요한 이유를 이해하고 있는지를 확인하기 위한 문제입니다.

DO NOT COPY
ssleeeee1@gmail.com

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



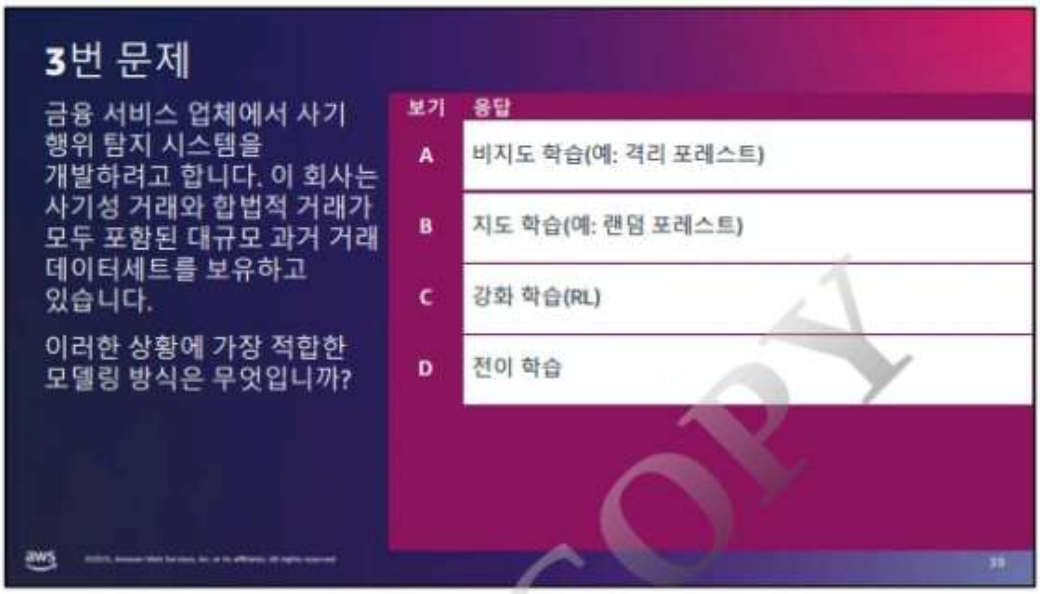
정답은 C, 로지스틱 회귀입니다.

설명: 해석 가능성이 우수한 선형 모델은 로지스틱 회귀는 재입원 위험 예측 등의 이진 분류 태스크에도 효율적입니다. 이 모델을 사용하면 특성의 중요도와 모델의 결정 방식을 이해할 수 있습니다.

오답:
A. 심층 신경망은 복잡하며 해석 가능성이 낮으므로 엄격한 규제가 적용되는 이 시나리오에는 적합하지 않습니다.
B. 랜덤 포레스트는 신경망보다는 해석 가능성이 높지만 로지스틱 회귀에 비해서는 투명성이 떨어집니다.
D. 컨볼루션 신경망은 이미지 처리 태스크용이며 해석 가능성이 낮습니다.

DO NOT COPY ssleeeee1@gmail.com

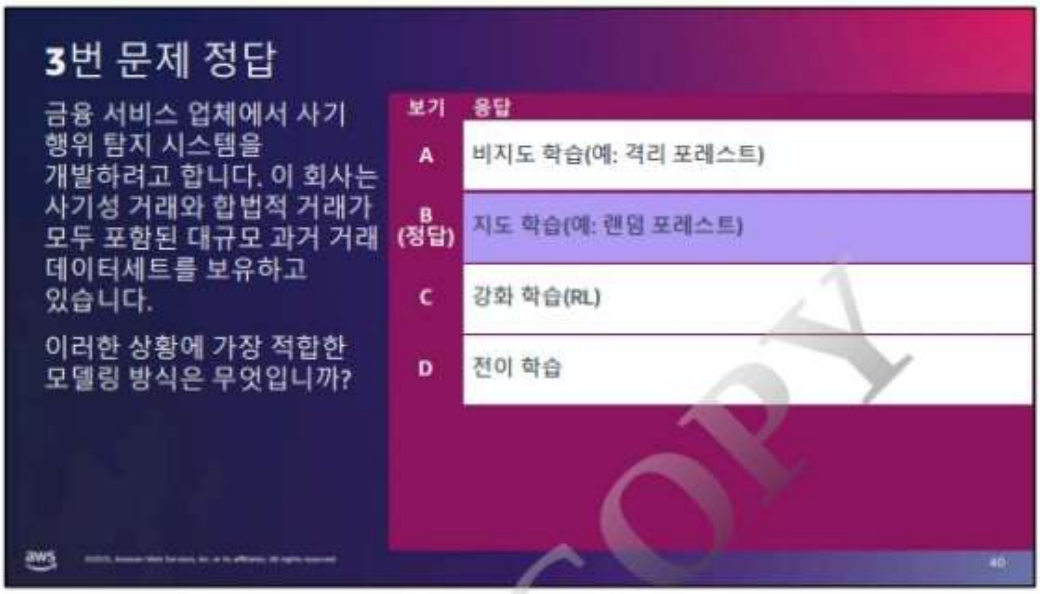
Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



다양한 기계 학습 패러다임 및 각 패러다임을 사용하기에 적합한 사용 사례를 이해하고 있는지를 평가하기 위한 문제입니다.

DO NOT COPY
ssleeeee1@gmail.com

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



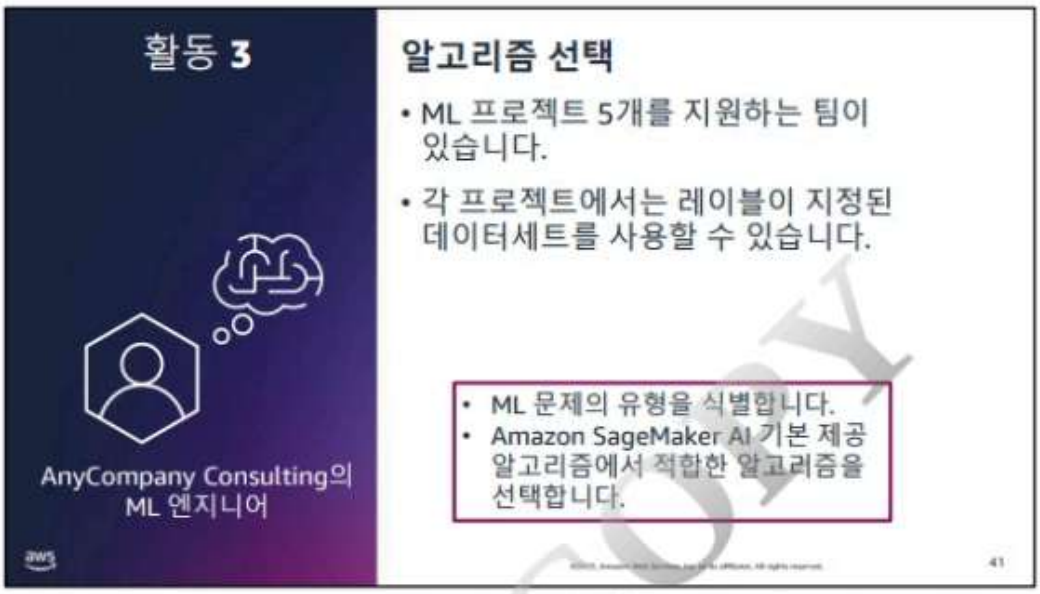
정답은 B, 지도 학습(예: 랜덤 포레스트)입니다.

설명: 이 시나리오에서는 랜덤 포레스트 알고리즘 등의 지도 학습을 사용하는 것이 가장 적합합니다. 지도 모델은 레이블이 지정된 과거의 사기성 거래 및 합법적인 거래 데이터를 사용하여 새 거래를 효율적으로 분류하는 방법을 학습할 수 있습니다.

오답:
A. 레이블이 지정된 데이터를 사용할 수 있는 시나리오에서는 비지도 학습이 적합하지 않습니다. 비지도 학습에서는 알려진 사기 행위 레이블을 사용하지 않기 때문입니다.
C. 강화 학습은 대개 시간 경과에 따른 의사결정 프로세스에 사용되며 사기 행위 탐지 등의 분류 태스크에는 사용되지 않습니다.
D. 전이 학습은 대상 영역에서 사용 가능한 레이블이 지정된 데이터가 제한되어 있는 경우 유용한데 이 시나리오는 그러한 상황에 해당되지 않습니다.

DO NOT COPY ssleeeee1@gmail.com

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



이 활동에서 여러분은 AnyCompany Consulting의 ML 엔지니어 역할을 맡아 작업을 진행합니다. 이 활동의 주요 목표는 각 프로젝트의 ML 문제 유형을 파악한 후 Amazon SageMaker AI의 기본 제공 옵션 중에서 적절한 알고리즘을 선택하는 것입니다.

비즈니스 문제: 한 소매업체가 타겟 마케팅 캠페인을 진행하여 매출을 늘리고자 합니다. 이 업체는 인구 통계, 과거 구매 내역, 고객이 이전 캠페인에 반응했는지 여부에 대한 정보 등의 과거 고객 데이터를 보유하고 있습니다.

잠재적인 ML 접근 방식:
- 고객 평생 가치 예측(회귀 분석)
  - 장점: 일정 금액에 따라 고객 우선 순위 지정
  - 단점: 모델 복잡도가 높으며 과거 데이터가 더 많이 필요함
- 캠페인에 반응할 가능성 예측(이진 분류)
  - 장점: 캠페인 반응 목표 달성 여부를 바로 파악할 수 있으며 더 단순한 모델 사용 가능
  - 단점: 창출 가능한 구체적인 수익 정보가 제공되지 않음
- 고객을 그룹으로 세분화(비지도 클러스터링)
  - 장점: 고객층에서 예기치 않은 패턴이 확인될 수 있음
  - 단점: 캠페인에 대한 반응을 직접 예측할 수 없음
- 구체적인 제품 추천 예측(멀티클래스 분류)
  - 장점: 매우 정확한 타켓 마케팅 진행 가능
  - 단점: 방대한 제품 및 고객 데이터 필요

권장 접근 방식: 고객이 캠페인에 반응할 가능성을 예측할 때는 이진 분류 방식을 사용하는 것이 좋습니다.

근거:
- 이 방식을 사용하면 캠페인 효율성 개선이라는 목표를 직접적으로 해결합니다.
- 훈련에 사용할 과거의 캠페인에 대한 반응 관련 데이터에 레이블이 지정되어 있습니다.
- 이 방식에서는 회귀에 비해 필요한 데이터가 더 적은 단순한 모델을 사용합니다.

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

- 이 방식에서는 고객의 반응 여부를 정확하게 예측하여 캠페인 대상 고객의 우선 순위를 지정할 수 있습니다.
- 모델 출력을 토대로 비즈니스 조치를 취할 수 있습니다.

다음 단계:

1. 과거의 고객 및 캠페인 반응 데이터를 준비합니다.
2. 테스트할 이진 분류 알고리즘 (예: 로지스틱 회귀 및 랜덤 포레스트)을 선택합니다.
3. 모델 성능 및 비즈니스 영향을 평가합니다.
4. 선택한 모델을 배포하여 진행 예정인 캠페인 대상 고객의 점수를 매깁니다.

이 활동에서는 해당 비즈니스 문제에 사용할 수 있는 ML 방식을 파악하고 각 방식의 장단점을 평가한 후 적절한 프레이밍 및 알고리즘 유형을 선택합니다. 여기서 파악한 ML 방식이 처음 수립했던 사업 목표에 부합하는지도 다시 확인해야 합니다.

DO NOT COPY
ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

지도 학습 문제의 유형 파악

태스크	태스크 설명	문제 유형	SageMaker AI 기본 제공 알고리즘
A	소득이 임계값 미만인지를 예측합니다(이전 슬라이드).	이진 분류	• CatBoost • 선형 학습자 • XGBoost
B	고객의 향후 구매 날짜를 예측합니다.		각 태스크에 적합한 지도 학습에 해당하는 ML 문제 유형 선택: • 이진 분류 • 멀티클래스 분류 • 회귀 • 시계열 예측
C	소액 할인, 대폭 할인, 할인 미제공에 고객이 가장 잘 반응할까요?		
D	10% 할인 제공 시 고객의 구매 금액을 예측합니다.		

42

이 슬라이드에는 고객 행동 예측과 관련된 3가지 지도 학습 태스크가 나와 있습니다. 각 태스크의 문제 유형을 정확하게 식별해야 합니다.

A. 소득이 특정 임계값 미만인지 여부 예측: 이 태스크가 비즈니스 사례이며 이진 분류 유형이라는 것은 이미 확인되었습니다.

B. 향후 구매 날짜 예측: 이 태스크에서는 특정 시점을 예측합니다.

C. 할인 유형에 대한 고객 반응 예측: 이 태스크에서는 예상되는 반응을 토대로 고객을 여러 그룹으로 분류해야 합니다.

D. 할인이 적용된 구매 금액 예측: 이 태스크에서는 연속하는 수치 값을 예측합니다.

적절한 알고리즘을 선택하고 효율적인 솔루션을 설계하려면 ML 문제 분류 방식을 이해해야 합니다. 태스크 B, C, D의 문제 유형을 선택합니다. 문제 유형을 결정할 때는 출력의 특성(범주형, 연속, 시간 기반) 및 제공될 수 있는 결과의 수를 고려해야 합니다.

DO NOT COPY ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

모델 및 알고리즘 설명서 참조

태스크	태스크 설명	문제 유형	SageMaker AI 기본 제공 알고리즘
A	소득이 임계값 미만인지를 예측합니다.	이진 분류	• CatBoost • 선형 학습자 • XGBoost
B	고객의 향후 구매 날짜를 예측합니다.	시계열 예측	Amazon SageMaker의 기본 제공 알고리즘 및 사전 훈련된 모델
C	소액 할인, 대폭 할인, 할인 미제공에 고객이 가장 잘 반응할까요?	멀티클래스 분류	
D	10% 할인 제공 시 고객의 구매 금액을 예측합니다.	회귀	

aws 41

각 태스크의 문제 유형은 다음과 같습니다.

B. 향후 구매 날짜 예측: 특정 시점을 예측하는 이 태스크는 시계열 문제입니다.
C. 할인 유형에 대한 고객 반응 예측: 예상되는 반응을 토대로 고객을 여러 그룹으로 분류해야 하는 이 태스크는 멀티클래스 분류 문제입니다.
D. 할인이 적용된 구매 금액 예측: 연속하는 수치 값을 예측하는 이 태스크는 회귀 문제입니다.

문제 유형을 식별한 후의 다음 단계는 각 태스크에 적합한 SageMaker AI 기본 제공 알고리즘을 선택하는 것입니다.

제공된 QR 코드를 사용하여 기본 제공 알고리즘 관련 정보가 포함된 SageMaker AI 설명서에 액세스할 수 있습니다. 알고리즘 선택 시 고려해야 하는 요인은 다음과 같습니다.

- 데이터 특성(예: 크기, 특성 유형, 희소성)
- 모델 해석 가능성 요구 사항
- 사용 가능한 컴퓨팅 리소스
- 각 알고리즘의 구체적인 장점(예: 범주형 변수 처리와 비선형 관계 캡처 기능)

SageMaker AI에서 효과적인 ML 모델을 구축하려면 각 문제 유형에 적합한 알고리즘을 맞추는 것이 중요합니다.

SageMaker AI 기본 제공 알고리즘에 대한 자세한 내용은 'Amazon SageMaker의 기본 제공 알고리즘 및 사전 훈련된 모델'(https://docs.aws.amazon.com/sagemaker/latest/dg/algos.html)을 참조하십시오.

DO NOT COPY ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

샘플 해결 방법 - 활동 3

태스크	태스크 설명	문제 유형	SageMaker AI 기본 제공 알고리즘
A	소득이 임계값 미만인지를 예측합니다.	이진 분류	• CatBoost • 선형 학습자 • XGBoost
B	고객의 향후 구매 날짜를 예측합니다.	시계열 예측	• DeepAR 예측
C	소액 할인, 대폭 할인, 할인 미제공에 고객이 가장 잘 반응할까요?	멀티클래스 분류	• K-최근접 이웃(k-NN) • 선형 학습자 • XGBoost
D	10% 할인 제공 시 고객의 구매 금액을 예측합니다.	회귀	• K-최근접 이웃(k-NN) • 선형 학습자 • XGBoost

44

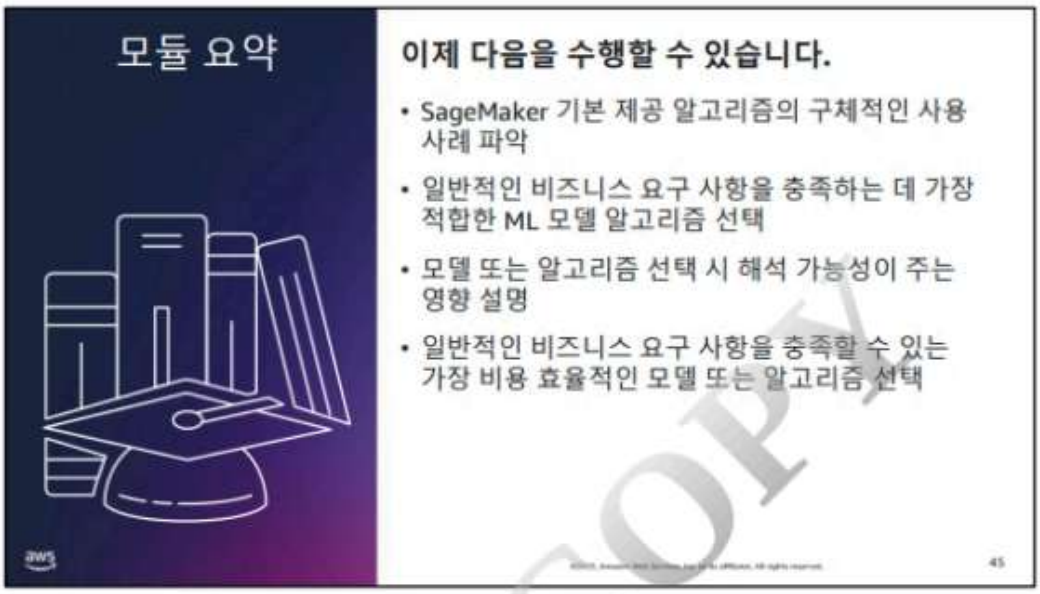
이 슬라이드는 액티비티 3의 샘플 솔루션을 제시하는데, 3개 지도 학습 태스크에 적합한 SageMaker AI 기본 제공 알고리즘을 보여줍니다. 향후 구매 날짜를 예측하는 시계열 태스크의 경우에는 DeepAR 예측 알고리즘을 사용하는 것이 좋습니다. 시계열 데이터 전용으로 제작된 이 알고리즘은 여러 관련 시계열을 동시 처리할 수 있습니다.

각 할인 수준에 대한 고객 반응을 확인하는 멀티클래스 분류 문제에서는 K-최근접 이웃(k-NN), 선형 학습기, XGBoost 등의 여러 알고리즘을 사용할 수 있습니다. 이러한 알고리즘은 여러 클래스를 사용하는 분류 작업을 효율적으로 처리할 수 있습니다.

구매 금액을 예측하는 회귀 태스크에서도 K-최근접 이웃, 선형 학습기, XGBoost 등의 알고리즘이 적합합니다. 즉, XGBoost 등의 일부 알고리즘은 다재다능하고 여러 문제 유형에 적용할 수 있다는 것을 보여줍니다.

DO NOT COPY ssleeeee1@gmail.com

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



이 모듈을 완료했으므로 이제 다음을 수행할 수 있습니다.

- 일반적인 비즈니스 요구 사항을 해결하기 위한 AWS AI 서비스 선택. 업무상의 구체적인 요구 사항에 맞는 적절한 AWS AI 서비스를 파악하여 선택하면 다양한 업계 시나리오에 가장 적합한 솔루션을 사용할 수 있습니다.
- SageMaker 기본 제공 알고리즘의 구체적인 사용 사례 파악. SageMaker AI 기본 제공 알고리즘을 확인하여 관련 실무 사용 사례에 적용하면 해당 알고리즘의 기능을 사용하여 문제를 효율적으로 해결할 수 있습니다.
- 일반적인 업무상의 요구를 충족하는 데 가장 적합한 ML 모델 알고리즘 선택. 정확도, 확장성, 성능 등의 요인을 고려하여 AWS가 제공하는 알고리즘 중에서 업무 과정에서 일반적으로 발생하는 당면 과제를 해결하는 데 가장 적합한 ML 알고리즘을 선택할 수 있습니다.
- 모델 또는 알고리즘 선택 시 해석 가능성이 주는 영향 설명. 모델 또는 알고리즘 선택 프로세스에서 AI 기반 결정 과정의 투명성과 신뢰도를 높이려면 해석 가능성을 반드시 고려해야 하는 이유를 설명할 수 있습니다.
- 일반적인 업무상의 요구를 충족할 수 있는 가장 비용 효율적인 모델 또는 알고리즘 선택. 성능과 예산 제한을 모두 적절하게 고려하여 AWS에서 제공되는 비용 효율성이 가장 높은 모델이나 알고리즘을 파악하여 구현할 수 있습니다.

이 모듈에서 습득한 지식과 기술을 적용하면 AWS에서 기계 학습 솔루션을 구축할 때 정보를 토대로 적절한 결정을 내릴 수 있으며, 그러면 모델의 성능과 확장성 및 비용 효율성을 모두 최적화할 수 있습니다.

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

이상으로 AWS의 기계 학습용 모델링 방식 선택 관련 모듈을 마치겠습니다. 다른 질문이 있거나 피드백을 제공하고 싶거나 추가 지원을 받으려는 경우 언제든지 AWS 교육 팀(https://support.aws.amazon.com/#/contacts/aws-training)에 문의하시기 바랍니다.

이 세션의 전 과정에서 적극적으로 참여해 주신 데 대해 감사의 인사를 드립니다.

DO NOT COPY
ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

이 모듈에서는 모델 훈련 프로세스의 핵심 요소에 대해 다룹니다. 구체적으로는 훈련 프로세스를 효율적이며 원활하게 진행하기 위해 모델 훈련 소요 시간을 단축하는 데 사용되는 일반적인 방식을 알아봅니다.

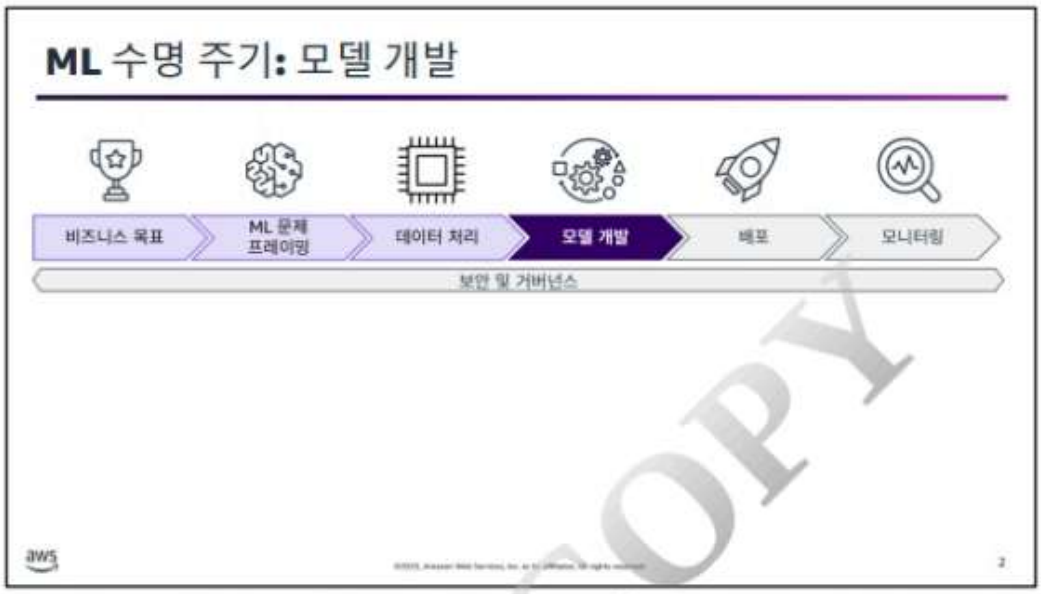
먼저 편향 완화 및 모델 성능 방법을 검토하고, 모델 과대적합과 과소적합을 방지하는 방법을 알아봅니다. 그런 다음 여러 방법을 함께 활용하여 모델 성능을 개선하는 방법과 하이퍼파라미터 튜닝을 사용하여 최적화된 모델 결과를 생성하는 방법을 살펴봅니다.

그리고 다양한 모델 크기 및 모델 버전 관리 방법을 알아보고, 모델 구체화 프로세스에서 Amazon SageMaker AI 기능을 활용할 수 있는 방법을 살펴봅니다.

DO NOT COPY
ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



이 모듈에서는 ML 수명 주기의 **모델 개발** 단계를 중점적으로 살펴봅니다. 이 단계에서는 준비된 데이터와 이전 단계에서 파악한 실무 정보를 사용하여 ML 모델을 정의합니다. 이 모듈에서는 모델 학습 프로세스를 검토하는 동시에 모델 훈련 컴퓨팅 환경을 선택하는 방법, 그리고 모델 훈련 소요 시간을 단축하는 방식에 대해 설명합니다.

이 모듈에서는 모델 훈련 프로세스의 구성 요소를 소개합니다. 구체적으로는 모델 훈련, 평가 및 등록을 지원하는 AWS 서비스 및 기능에 대해 알아봅니다. 이러한 내용을 숙지하면 훈련 및 추론용 컴퓨팅 환경을 선택하는 방법과 관련된 인사이트를 파악할 수 있습니다.

그리고 마지막으로 훈련 프로세스를 효율적으로 진행하기 위해 모델 훈련 소요 시간을 단축하는 데 사용되는 방식을 살펴봅니다.

모델 개발 단계에서는 하이퍼파라미터를 사용하여 훈련 작업 구성을 튜닝해 모델 훈련 방식을 조정하는 작업도 수행합니다.

DO NOT COPY
ssleeeee1@gmail.com

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

이 모듈에서 다룰 3가지 주요 주제는 다음과 같습니다.

주제 A: 모델 훈련 관련 개념

이 섹션에서는 기계 학습(ML) 모델 훈련과 관련된 기본적인 개념을 살펴봅니다. 이러한 개념을 이해하면 ML 작업을 위한 기초 지식을 습득할 수 있습니다.

주제 B: SageMaker AI에서 모델 훈련

이 섹션에서는 실제 구현 과정을 중심으로 Amazon SageMaker AI를 사용해 모델을 효율적으로 훈련하는 방법을 설명합니다.

DO NOT COPY
ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

ML 엔지니어로서, 모델 훈련과 관련된 핵심 개념 및 기법을 이해하는 것은 정확하고 유용한 ML 솔루션을 개발하는데 중요합니다.

이 섹션에서 소개하는 개념은 다음과 같습니다.

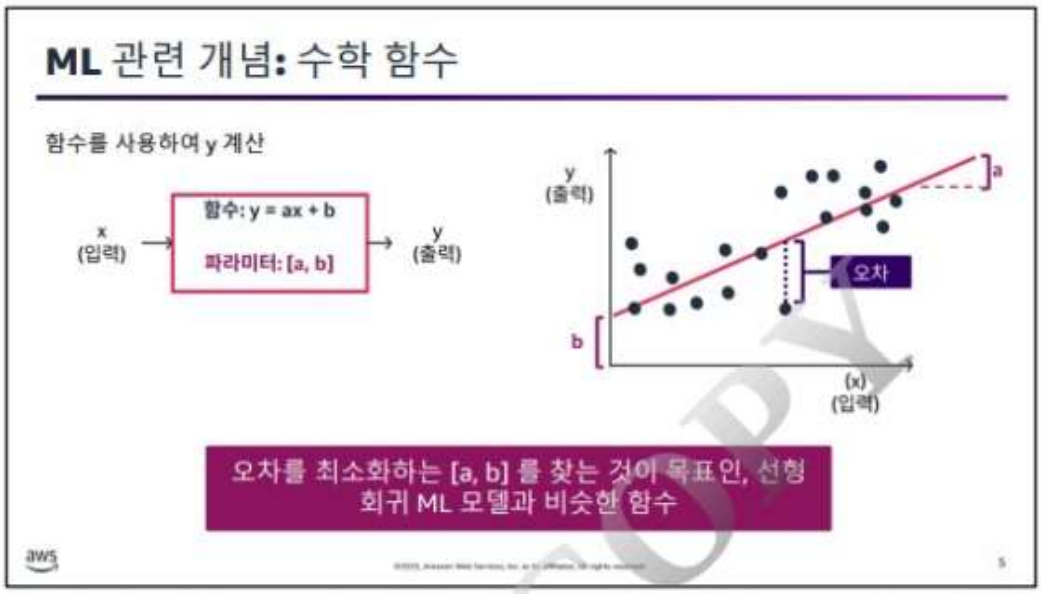
- 모델 훈련 프로세스: 데이터 준비부터 모델 평가에 이르기까지 ML 모델을 훈련할 때 진행하는 단계에 대해 알아봅니다.
- ML 모델의 핵심 요소: ML 모델 최적화 시에 적용되는 알고리즘, 가중치, 손실 함수 등의 개념을 살펴봅니다.

이러한 기본 개념을 파악하면 이 모듈의 뒷부분에서 살펴볼 고급 주제를 이해하는 데 필요한 기초 지식을 확실하게 습득할 수 있습니다.

DO NOT COPY
ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



ML 모델은 매우 복잡한 다차원일 수 있지만, 2차원 선으로 표시된 수학 함수를 살펴보면 ML 모델의 개념, 모델 훈련이 무엇을 포함하는지에 대한 통찰력을 얻을 수 있습니다.

슬라이드에 나와 있는 y = ax + b 방정식에는 다음 요소가 포함되어 있습니다.

- x는 입력입니다.
- y는 출력입니다.
- 파라미터 [a, b]는 선의 위치와 방향을 정의합니다.

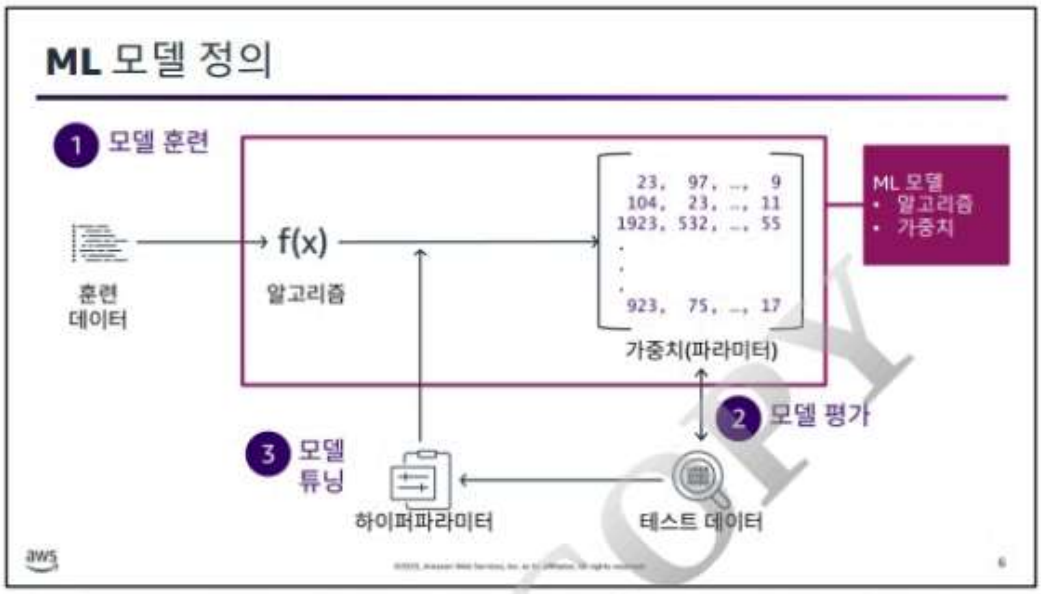
이 두 파라미터를 조정하여 여기 점으로 표시된 (x, y) 데이터에 가장 적합한 선을 정의할 수 있습니다. **오차**는 선에서 데이터 포인트까지의 거리를 기준으로 정의됩니다. 오차가 최소화되는 파라미터 세트에 따라 최적 적합성이 결정됩니다. 파라미터를 확인하고 나면 추가 입력 x 값의 y 값을 계산할 수 있습니다.

이 시퀀스는 오류를 최소화하고 정확한 예측을 하기 위해서 파라미터인 **가중치**를 조정하여 ML 모델을 훈련하는 방식과 개념적으로 비슷합니다. 여기에 설명되어 있는 방식은 **선형 회귀** 알고리즘(**스크립트**)과 비슷합니다. 여기서 목표는 예측 출력과 실제 출력 간의 오차를 최소화하여 훈련 데이터를 가장 정확하게 대표하는 파라미터(가중치) 세트를 찾는 것입니다.

DO NOT COPY ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



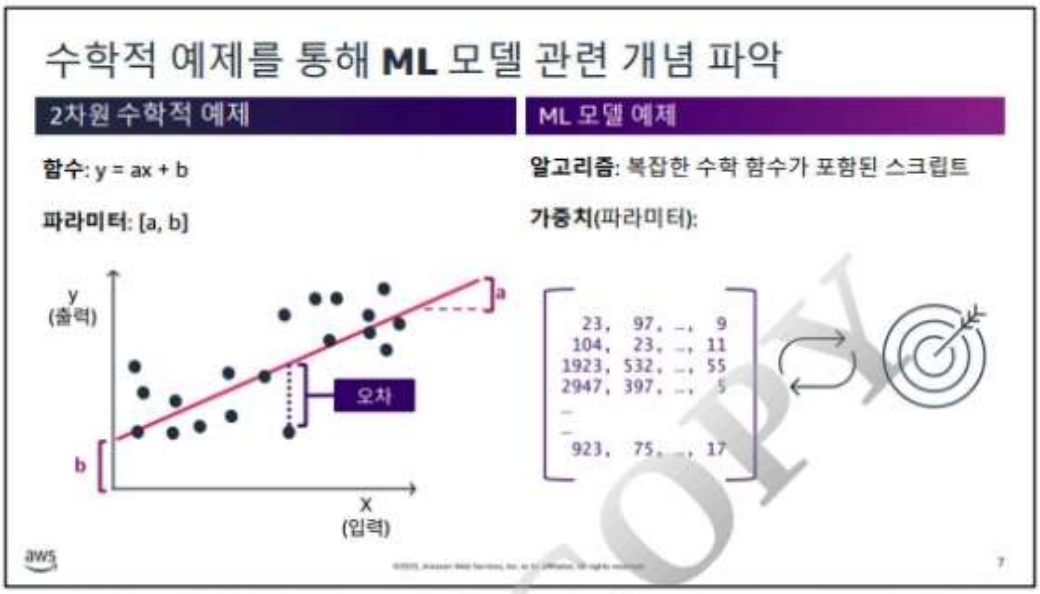
1. **ML 모델**은 실제 프로세스의 수학적 표현이라 할 수 있습니다. 입력 데이터를 가져와 예측이나 결정을 생성하기 때문입니다.
   - **알고리즘**은 모델이 데이터에서 패턴을 학습하고 예측하는 데 사용하는 핵심 로직 또는 규칙 세트입니다. 알고리즘의 예로는 선형 회귀, 의사결정 트리, 신경망 등이 있습니다.
   - **파라미터**라고도 하는 **가중치**는 훈련 중에 모델이 학습하는 값입니다. 가중치에 따라 예측을 위해 입력 특성이 결합되는 방식이 결정됩니다.
   - **훈련**은 모델 가중치를 조정하여 훈련 데이터의 예측 값과 실제 값 간 차이를 최소화하는 프로세스입니다.
   - ML 모델에는 알고리즘과 모델 가중치가 모두 포함됩니다.
2. **모델 평가**는 처음 보는 데이터에 대한 모델 성능을 평가하여 모델의 일반화 능력을 측정하는 프로세스입니다.
   - ML 모델을 사용하여 테스트 데이터세트로 예측을 합니다. 훈련에서는 테스트 데이터가 사용되지 않았으므로 이 예측의 성능을 통해 모델이 프로덕션 데이터로 예측을 하는 성능을 예상할 수 있습니다.
3. **모델 튜닝**은 훈련 프로세스를 최적화하여 모델 성능을 개선하는 프로세스입니다. 하이퍼파라미터는 훈련 프로세스를 제어하는 설정입니다. 모델 튜닝에서는 광범위한 하이퍼파라미터 값을 적용해 보면서 여러 모델 훈련 단계를 진행하는 경우가 많습니다.

이러한 구성 요소가 반복 프로세스에서 함께 작동해 ML 모델을 생성 및 개선하는 것입니다.

DO NOT COPY ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



ML 알고리즘은 단순한 수학 함수보다는 복잡할 수도 있지만, 이 비유는 가중치, 모델 평가, 책임감 있는 AI 등의 핵심 개념을 이해하는 데 도움이 될 수 있습니다.

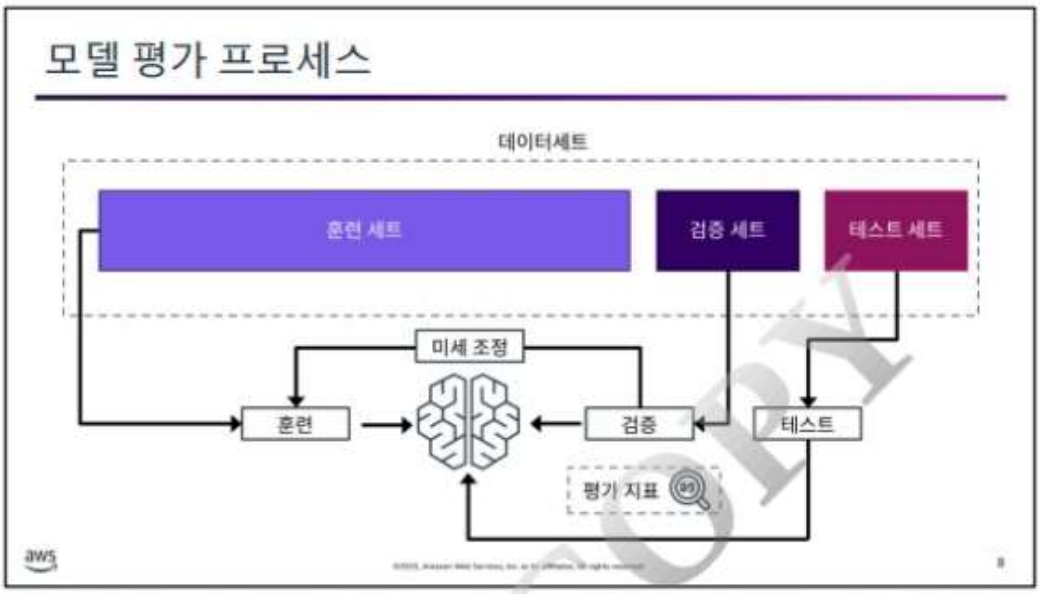
훈련 프로세스에서 알고리즘은 훈련 데이터에 가장 적합한 가중치를 확인해야 합니다. 이러한 각 시도를 평가할 때는 가중치 조정 결과를 알려주는 손실 함수를 사용하여 오차를 계산합니다. 이렇게 훈련을 여러 번 반복하면 오차가 최소화되는 방향으로 성능이 점점 높아지며, 오차가 가장 작은 지점에서 훈련 데이터를 더욱 정확하게 대표하는 예측이 생성됩니다. 이 프로세스는 오차가 임곗값보다 낮아지거나 중지 기준이 충족될 때까지 계속 진행됩니다. 이러한 훈련의 목표는 손실을 최소한으로 줄이는 것입니다.

ML 모델을 훈련할 때 고려해야 하는 사항은 다음이 포함됩니다.

- 차원: 고차원 데이터를 사용하는 경우가 많음
- 복잡도: 비선형 함수 및 다중 계층 사용
- 규모: 파라미터 수십억 개를 최적화

DO NOT COPY ssleeeee1@gmail.com

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



모델을 훈련한 후에는 모델 성능이 요구 사항을 충족하는지를 확인해야 합니다. **모델 평가**에서는 모델이 훈련 과정에서 훈련 데이터의 패턴과 관계를 캡처했는지 여부를 평가하는 처음 보는 데이터를 기반으로 합니다.

기계 학습 수명 주기 중 이 단계는 성능 기준을 설정하고 모델 평가 데이터세트를 준비하는 것으로 시작합니다. 성능 기준 설정은 안정성과 신뢰성이 높은 모델을 구축하는 과정에서 기본적으로 수행해야 하는 단계입니다. 이러한 성능 기준을 기준점으로 사용하여 향후 모델 개선 사항을 측정할 수 있습니다.

모델 평가 데이터세트는 전체 데이터세트 중 하위 집합으로, 훈련에서는 사용되지 않았으며 기계 학습 모델 성능 평가하는 용도로만 사용합니다. 데이터세트를 나누는 일반적인 방식은 훈련용 데이터로 80%, 검증용 데이터로 10%, 테스트용으로 10%를 배정하는 것입니다. 이 비율은 모델 훈련 및 검증 중에 사용할 수 있는 계산 능력과 모델의 복잡도에 따라서도 달라집니다. 데이터 준비와 피처 엔지니어링은 훈련 및 평가 데이터세트 모두에서 수행됩니다.

- **검증 데이터세트**는 모델 훈련 프로세스 중에 하이퍼파라미터를 미세 조정하고, 최상의 모델 아키텍처를 선택하고, 모델 성능을 모니터링하는 데 사용됩니다. 이 데이터세트는 과대적합을 방지하고 모델이 새로운 데이터에도 일반화되도록 하기 위한 훈련 데이터와는 별개입니다.
- **테스트 데이터세트**는 최종 모델의 성능을 평가하기 위해 보류해 둔 데이터의 일부입니다. 이 데이터세트를 훈련 또는 하이퍼파라미터 튜닝에 사용해서는 안 됩니다. 그렇게 하면 편향이 발생하여 성능 추정치를 인위적으로 상향 조정될 수도 있기 때문입니다. 테스트 데이터세트는 모델이 배포될 실제 시나리오를 나타내야 합니다.

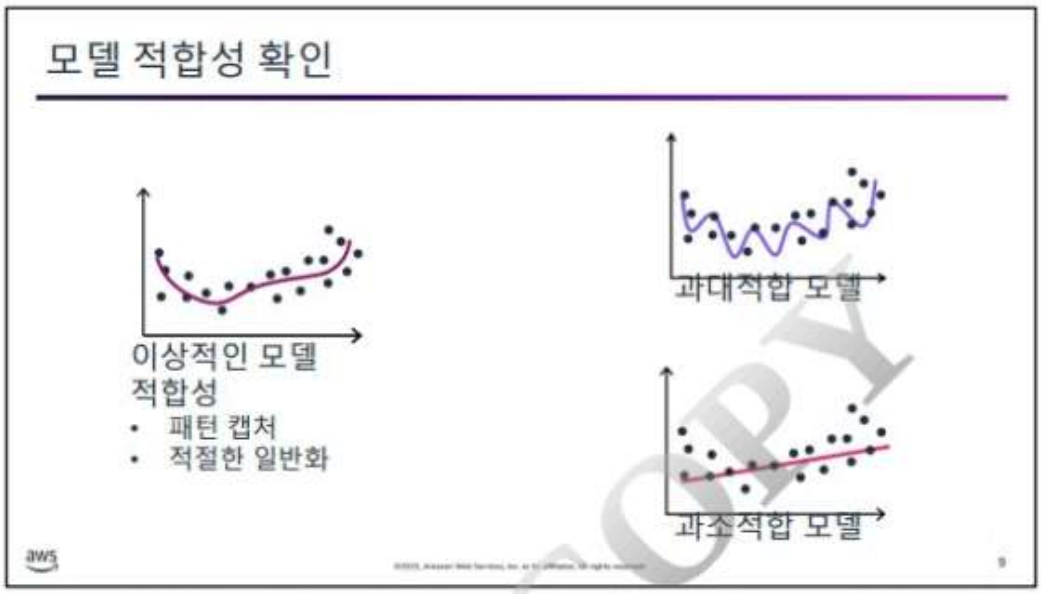
모델을 평가하는 동안, 훈련 작업과 관련된 지표를 검사합니다. 모델의 오차 정도를 나타내는 지표를 특히 자세하게 확인해야 합니다. 이렇게 하면 훈련 작업 설정을 최적화하여 모델 품질을 개선할 수 있기 때문입니다. 평가 프로세스에 사용되는 지표를 통해 모델 편향 및 모델 분산을 파악할 수 있습니다. 구체적인 지표는 선택한 알고리즘 유형에 따라 달라집니다. 알고리즘 유형의 예는 다음과 같습니다.

- 분류: 정확도, 정밀도, 재현율, F1-점수
- 회귀: 평균 제곱 오차(MSE), R-제곱

이러한 모델 평가 프로세스를 진행하면 처음 보는 데이터도 효율적으로 일반화할 수 있는 모델을 완성할 수 있습니다.

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



보이지 않는 새로운 데이터로 정확하게 예측을 할 수 있는 기계 학습 모델의 능력을 **모델 적합성**입니다. 이상적으로, 모델 적합성에는 모델 훈련이 훈련 데이터의 패턴을 캡처했음이 반영되어야 합니다. 결과 모델은 이러한 패턴을 사용하여 새로운 데이터와 관련된 유의미한 예측을 생성합니다. 모델 적합성은 데이터와 복잡도와 너무 단순한 표현 여부가 모두 반영된 지표입니다.

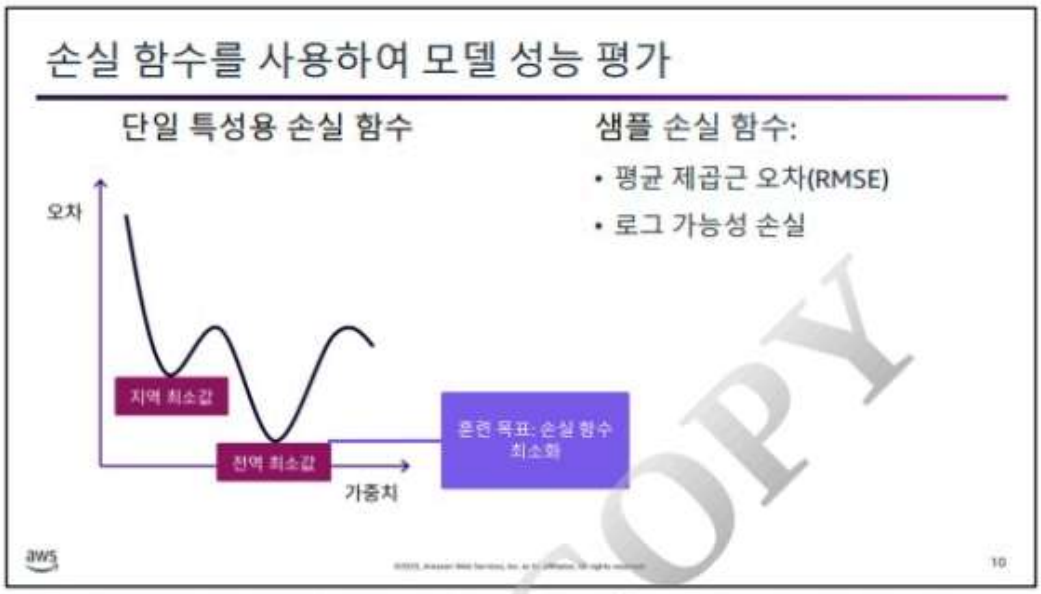
과대적합 모델은 훈련 데이터에 너무 가깝게 튜닝되어 해당 데이터세트 특유의 노이즈나 예측과는 관련이 없는 세부 정보도 예측에 반영하는 모델입니다. 이는, 모델에서 고유한 패턴과 불규칙성을 포함하여 훈련 데이터를 너무 잘 학습했으므로 보이지 않는 새로운 데이터에 잘 일반화되지 않을 수 있습니다. 훈련 데이터의 다양성과 샘플 수를 늘리면 과대적합 가능성을 줄일 수 있습니다. **정규화** 기법을 활용하면 모델 복잡도를 낮출 수 있습니다.

과소적합 모델은 훈련 데이터의 기본적인 관계나 패턴을 정확하게 캡처하지 못한 모델입니다. 너무 단순하여 데이터의 실제 복잡성을 정확하게 나타내지 못하므로 훈련 데이터와 새로운 데이터 모두 성능이 낮아집니다. 특성을 추가하는 등의 방식으로 모델 복잡도를 높이면 과소적합 모델 성능을 개선할 수 있습니다.

모델 적합성을 높이기 위한 모범 사례는 다음과 같습니다.
- 충분한 양의 고품질 훈련 데이터를 사용하여 모델을 훈련시킵니다.
- 훈련 및 검증 오류를 모니터링합니다.
- 정규화 등의 기법을 사용하여 예측 결과를 개선합니다.

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



모델 훈련에서는, **목표 함수**라고도 하는 **손실 함수**는 훈련 데이터의 실제 값 대비 모델 예측 성능을 측정합니다. 모델 훈련의 목표는 손실 함수를 최소화하는 것입니다. 훈련을 반복하면서 손실이 계속 최소화되는 과정에서 모델은 훈련 데이터에서 패턴을 학습하며, 그러면 이러한 패턴을 '일반화'하여 새로운 미확인 데이터를 토대로 정확한 예측을 할 수 있습니다.

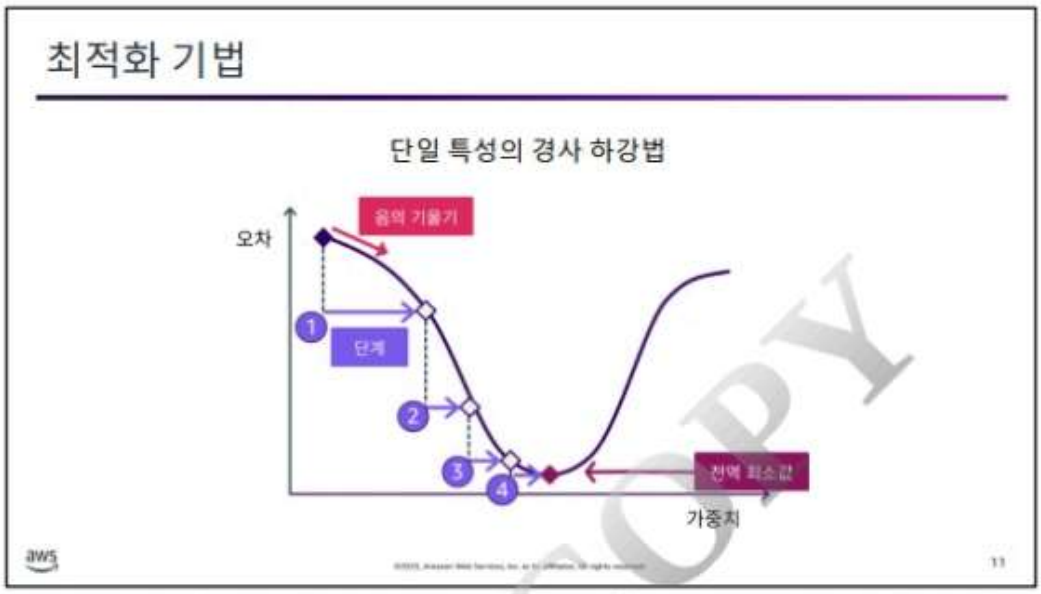
ML 태스크에서 흔히 사용되는 2가지 손실 함수는 다음과 같습니다.

- **평균 제곱근 오차(RMSE)**: 예측 오류의 표준 편차를 설명합니다. 실제 출력과 계산된 출력 간의 차를 제곱하고 전체 데이터세트에서 해당 값의 평균을 구하는 방식으로 RMSE를 계산합니다. RMSE 손실 함수는 주택 가격 예측 등의 선형 회귀 태스크에서 모델 정확도를 측정하려는 경우 유용합니다.
- **로그 가능성 손실(교차 엔트로피 손실)**: 예측의 확률을 기준으로 계산됩니다. 입력이 둘 이상의 클래스 중 하나에 속하는지 여부를 예측하는 것이 목표인 분류 태스크에 사용됩니다. 가령 이메일이 스팸인지 여부를 예측하는 로지스틱 회귀 모델의 정확도를 평가할 때 로그 가능도 손실을 사용할 수 있습니다.

대상 특성이 하나일 때 손실 함수의 모양은 구불구불한 곡선입니다. 이 곡선의 최저점인 **전역 최소값**은 오차를 최소화하는 최적 모델 가중치 세트에 해당됩니다. 모델 훈련은 일련의 **에포크**로 구성되어 있습니다. 각 에포크에서는 전체 데이터세트에서 알고리즘이 반복 실행되어 손실 함수가 최소화되는 가중치 값을 찾습니다. 각 반복 중에 수집되는 피드백을 토대로 가중치가 업데이트됩니다. 모델은 훈련 중에 전역 최소값에 가까워지는 방향으로 가중치를 반복 조정하며, 그러면 각 단계에서 모델 성능이 개선됩니다. 이 과정에서 해결해야 하는 문제 중 하나는 손실 함수 곡선에서 지역 최소값을 피하도록 모델 훈련을 구성하는 것입니다.

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



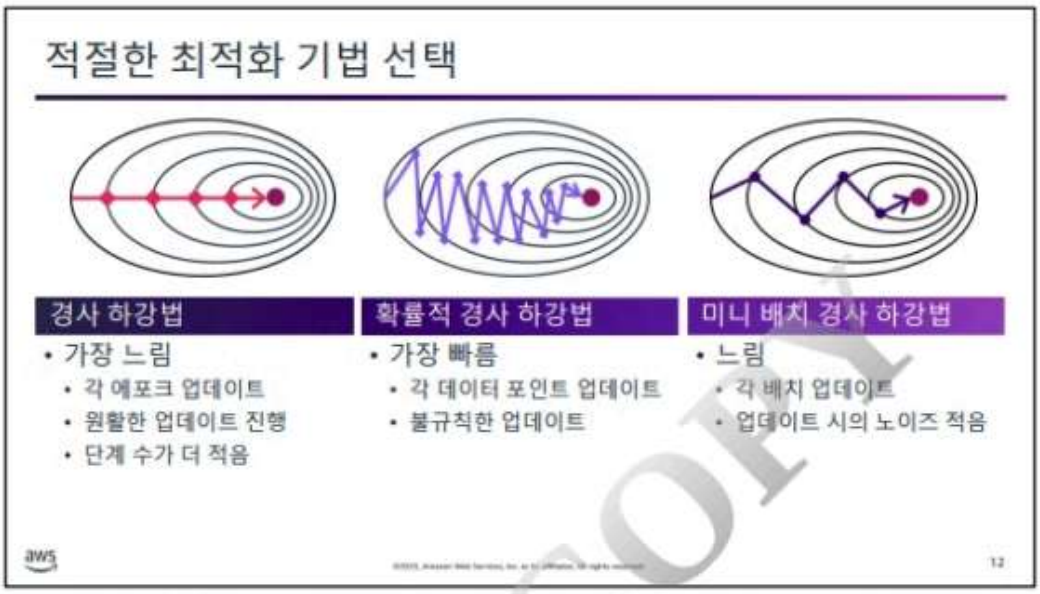
수 많은 특성에서 손실 함수를 최적화하는 접근 방식을 생각해 보십시오. 가능한 모든 가중치 조합을 테스트하는 것은 비효율적입니다. **최적화 기법**을 활용하면 손실 함수 값을 더욱 효율적으로 최소화할 수 있습니다.

경사 하강법은 손실 함수의 최저치를 구하는 데 널리 사용되는 최적화 기법입니다.

기울기는 현재 가중치에서의 오차 곡선의 기울기입니다. **음의 기울기**는 손실이 최저치에 가까워지는 것입니다. 경사 하강법에서는 이 기울기를 계산한 후 기울기가 점차 아래쪽을 향하도록 가중치 값을 업데이트합니다. 훈련의 오차가 전역 최소값에 조금씩 가까워지도록 이 프로세스가 **반복적으로** 진행됩니다. 즉, 경사 하강법은 오차를 반복적으로 하향 조정함으로써 손실 함수가 최소화되는 최적 가중치 값을 식별하는 기법입니다.

DO NOT COPY
ssleeeee1@gmail.com

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



업데이트 속도와 컴퓨팅 요구 사항, 그리고 불규칙한 업데이트 허용 오차 요구 사항은 최적화 기법 선택에 영향을 줄 수 있습니다. 3가지 일반적인 최적화 기법은 다음과 같습니다.

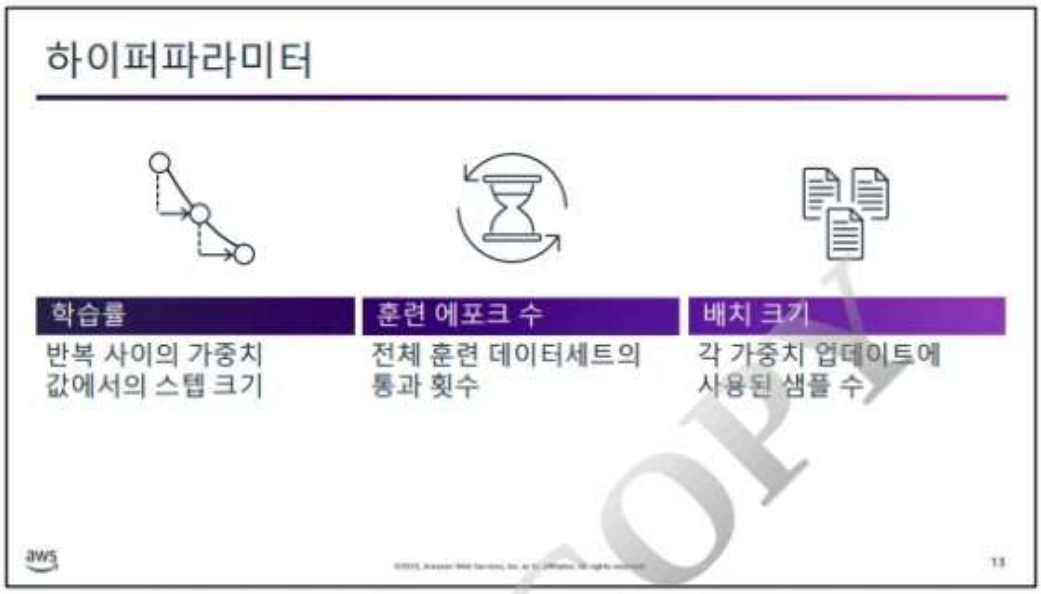
- **경사 하강법**: 각 **에포크** 후에, 훈련 데이터를 전체 통과한 후에, 가중치를 업데이트합니다. 이 기법에서는 최저치에 도달할 때까지 여러 하위 단계를 사용하며 모든 단계는 최저치에 가까워지는 방향으로 진행됩니다. 하지만 타 최적화 기법에 비해 속도는 느립니다.
- **확률적 경사 하강법(SGD)**: 데이터세트의 각 레코드에서 가중치를 업데이트합니다. 슬라이드에 설명되어 있는 3개 기법 중 속도는 가장 빠르지만 각 단계가 다른 방향으로 진동합니다. 이 진동으로 인해 최저치로 접근할 때 잡음이 심하고 곡선이 불규칙해집니다. 가령 데이터세트의 데이터 포인트 수가 1,000개라면 SGD에서는 가중치가 1,000번 업데이트됩니다.
- **미니 배치 경사 하강법**: 경사 하강법과 SGD가 모두 적용되는 하이브리드 방식입니다. 이 방식에서는 각 **배치**, 혹은 훈련 데이터 레코드의 하위 집합, 후에 성능을 평가합니다. 미니 배치 경사 하강법은 경사 하강법보다는 업데이트 횟수가 많으며 SGD보다는 불규칙성이 낮습니다. SGD의 계산에 필요한 것보다 메모리를 더 적게 사용하도록 **배치 크기**를 정의할 수 있습니다.

구체적인 요구 사항, 데이터세트 크기, 사용 가능한 계산 리소스에 따라 이러한 기법 중 적절한 항목을 선택하면 됩니다. 각 기법의 특성을 파악하면 기계 학습 프로젝트에 가장 적합한 방법을 선택할 수 있습니다.

DO NOT COPY ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



훈련 작업을 생성할 때 해당 작업에서 모델을 훈련할 방법을 구성해야 합니다. 이러한 모델 훈련 방법을 정의하는 설정이 **하이퍼파라미터**입니다. 훈련 중에 모델이 학습을 하는 모델 파라미터와는 달리 하이퍼파라미터는 훈련이 시작되기 전에 설정하는 값입니다.

예를 들어, 아래 하이퍼파라미터는 앞에서 설명한 최적화 기법과 관련이 있습니다. 이러한 기법은 흔히 사용되는 대다수 알고리즘에 적용되며 Amazon SageMaker AI에서도 지원됩니다.

- **학습률**은 경사 하강법 사용 시 가중치를 업데이트할 때 진행하는 스텝 크기를 정의합니다. 학습률에 따라 각 반복에서 최적의 솔루션을 향해 나아가는 단계의 크기가 결정됩니다. 학습률을 너무 낮게 설정하면 수렴이 느려질 수 있고, 학습률을 너무 높게 설정하면 알고리즘이 목표를 초과하여 수렴에 실패할 수 있습니다.
- **훈련 에포크 수**는 훈련이 전체 훈련 데이터세트에서 완전히 통과한 횟수에 해당됩니다. 훈련 에포크 수를 늘리면 모델이 학습을 추가로 진행할 수 있지만 과대적합 위험성도 높아집니다.
- **배치 크기**는 각 가중치 업데이트 전에 훈련 데이터에서 처리되는 샘플 수입니다. 각 에포크 내에서는 데이터가 배치 단위로 처리되며 가중치는 각 배치 후에 업데이트됩니다. 배치 크기가 클수록 성능이 향상될 수 있지만 더 많은 메모리가 필요합니다.

이상적인 하이퍼파라미터 값 세트는 모델 훈련이 모델이 불안정해지거나 제대로 수렴되지 않는 현상이 발생하지 않고 최적의 솔루션 방향으로 빠르게 진행되도록 도와줍니다. 초기 모델 훈련 후에는 하이퍼파라미터를 **튜닝**하여 모델이 새로운, 처음 보는 데이터에서 패턴을 일반화하는 능력을 개선할 수 있습니다. 그리고 비슷한 평가 기법을 사용하여 하이퍼파라미터 튜닝을 위한 정보를 제공할 수 있습니다. 튜닝 관련 주제는 다음 모듈에서 살펴볼 예정입니다.

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

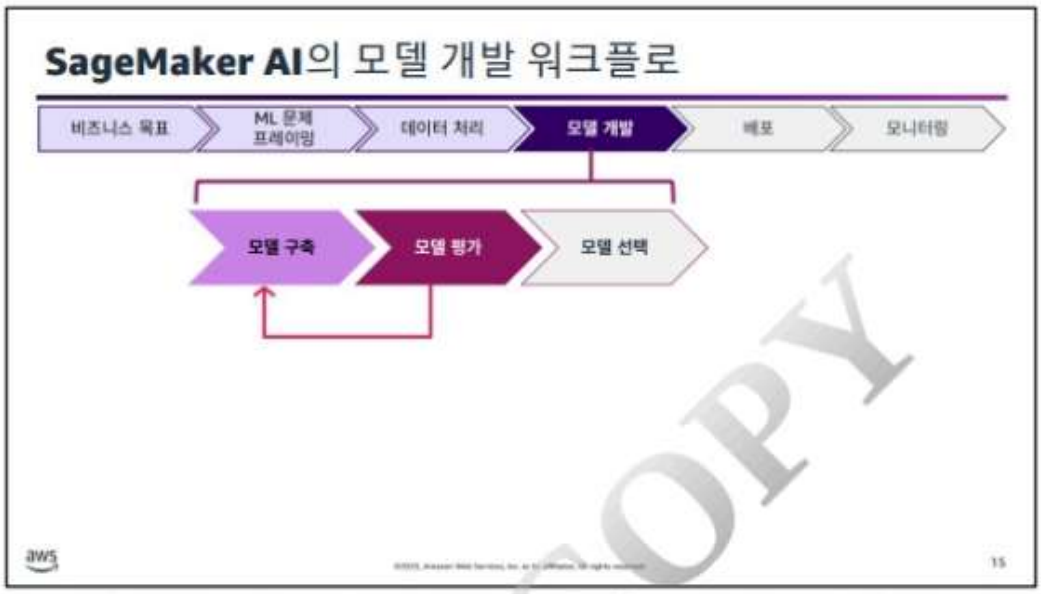
Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

이 섹션에서는 Amazon SageMaker AI를 사용하여 기계 학습 모델을 효율적으로 훈련하는 방법을 살펴봅니다. SageMaker AI에서는 완전관리형 서비스, 사용자 지정 가능 환경 등 모델 훈련에 사용할 수 있는 여러 가지 옵션이 제공됩니다. 이 섹션에서는 SageMaker AI를 사용하여 ML 워크플로를 원활하게 진행하고, 훈련 프로세스 규모를 조정하고, 모델 성능을 개선하는 방법을 알아봅니다. 예를 들어 선택하는 컴퓨팅 환경에 따라 훈련 프로세스의 성능, 비용 및 효율성이 크게 달라질 수 있습니다. 이 섹션에서는 적합한 컴퓨팅 환경을 선택하는 데 사용할 수 있는 옵션과 고려 사항을 살펴보겠습니다.

DO NOT COPY
ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



SageMaker AI에는 다음을 비롯한 모델 개발의 모든 단계를 지원하는 기능이 포함되어 있습니다.

1. **모델 구축**: 다양한 기본 제공 알고리즘 중에서 선택할 수도 있고 인기 있는 ML 프레임워크를 통해 사용자 지정 스크립트를 정의할 수도 있습니다. 분산 훈련 및 하이퍼파라미터 튜닝용 옵션이 포함된 완전관리형 확장 가능 훈련 인프라를 사용할 수 있습니다.
2. **모델 평가**: 모델 평가, 편향 탐지 및 설명 가능성용으로 기본 제공되는 도구를 사용할 수 있습니다.
3. **모델 선택**: 모델 메타데이터를 비교하여 배포하기에 가장 적합한 모델 버전을 파악하고 승인 상태를 관리할 수 있습니다.

모델 구축의 경우, Amazon SageMaker AI의 기본 제공 알고리즘과 컨테이너 또는 사용자 지정 알고리즘과 컨테이너 중에서 선택할 수 있습니다. 하이퍼파라미터 및 인프라 설정을 사용자 지정할 수 있습니다. 필요시에는 흔히 사용되는 ML 프레임워크와 진입점 스크립트에 바로 액세스하여 모델 구축을 유연성을 높일 수 있습니다. SageMaker AI에서는 하이퍼파라미터 튜닝, 분산 훈련 라이브러리 및 관찰성 도구가 기본 제공됩니다. ML 모델을 반복적으로 구축 및 평가하는 과정에서는 SageMaker AI에는 모델 개발과 거버넌스의 실험과 오케스트레이션을 지원하는 추가적인 기능이 포함됩니다.

DO NOT COPY ssleeeee1@gmail.com

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

Amazon SageMaker AI에서 ML 모델 훈련

Amazon SageMaker 훈련

대규모 완전관리형 ML 모델 훈련:

- Amazon SageMaker Canvas
- Amazon SageMaker Studio
- Amazon SageMaker Pipelines

16

Amazon SageMaker Training은 SageMaker AI에서 제공되는 완전관리형 ML 서비스입니다. 이를 통해 광범위한 ML 모델을 대규모로 효율적인 방식으로 훈련할 수 있습니다. SageMaker AI 작업에서는 ML 워크로드를 컨테이너화하며 AWS 컴퓨팅 리소스와 인프라를 프로비저닝 및 관리합니다.

이 과정에서는 ML 모델 훈련용으로 제공되는 3가지 SageMaker AI 기능을 살펴봅니다.

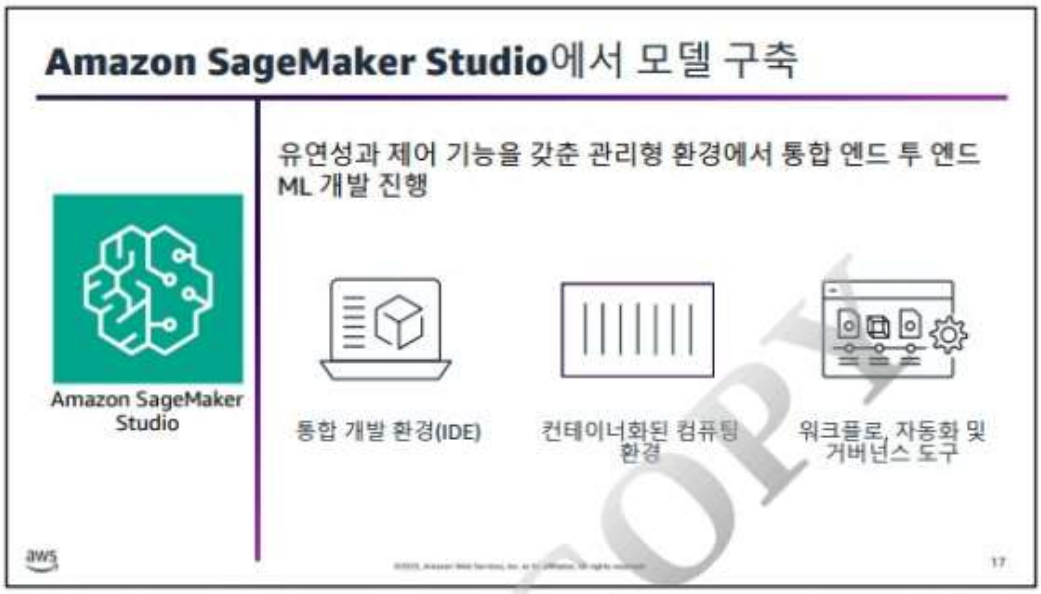
- Amazon SageMaker Canvas: 노코드 또는 로우코드 환경에서 ML 모델을 훈련시킵니다.
- Amazon SageMaker Studio: 유연성과 제어 기능을 갖춘 완전관리형 환경에서 코드를 사용해 대규모로 ML 모델을 훈련시킵니다.
- Amazon SageMaker Pipelines: SageMaker Studio는 ML 개발 자동화를 위해 특별히 구축된 워크플로 오케스트레이션 서비스인 SageMaker Pipelines와 통합됩니다. SageMaker Pipelines에서는 서버리스 오케스트레이션 인프라가 제공되므로 시각적 편집기, SDK, API, JSON 중에서 원하는 방식을 선택하여 파이프라인을 정의하고 관리할 수 있습니다.

종합하면, SageMaker AI의 이러한 기능은 AWS 기반 ML 모델 훈련, 튜닝 및 평가를 지원합니다.

DO NOT COPY
ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

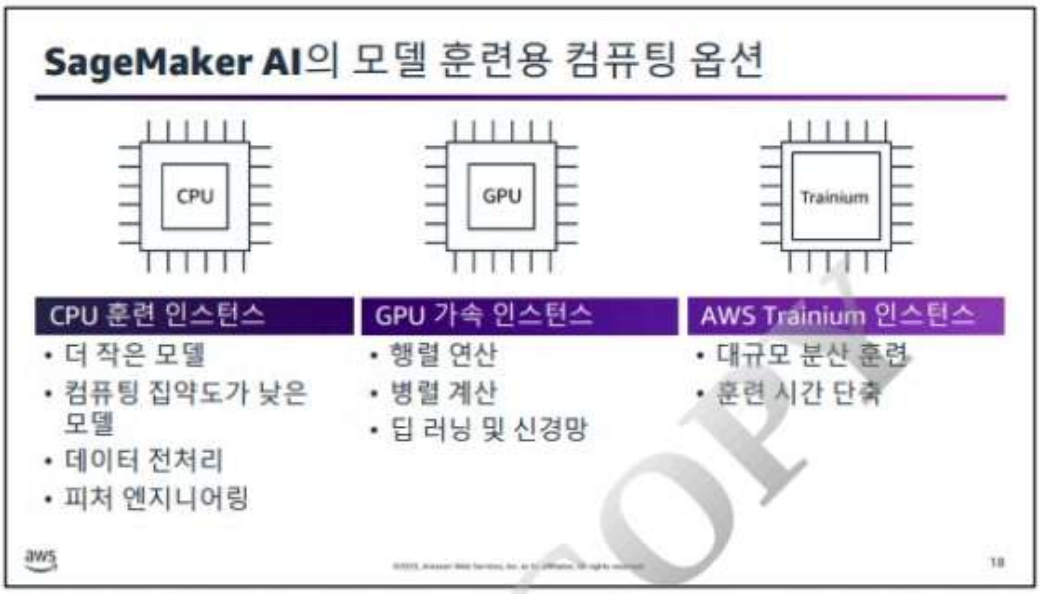


SageMaker Studio는 통합 엔드 투 엔드 ML 개발 환경을 제공하는 SageMaker AI의 기능입니다.

- SageMaker Studio에서는 JupyterLab, CodeEditor, RStudio 등의 **통합 개발 환경(IDE)** 중 원하는 환경을 선택할 수 있습니다.
- **SageMaker AI 관리형 컨테이너 이미지**에는 여러 ML 프레임워크를 사용하여 구축되며, 훈련용 기본 제공 알고리즘, 설정 및 필요한 라이브러리와 종속성이 포함되어 있습니다. 기본 제공 알고리즘을 사용할 때는 실험을 실행하는 데 코딩이 필요하지 않습니다. SageMaker AI 노트북 인스턴스 또는 SageMaker Studio에서 이 이미지를 사용하여 훈련 컨테이너를 생성할 수 있습니다. 훈련 데이터, 하이퍼파라미터, 컴퓨팅 리소스만 입력으로 제공하면 됩니다. 그러므로 결과 추적 및 코드 변경 관련 오버헤드를 줄이면서 실험을 실행하고 모델 아티팩트를 생성할 수 있습니다.
- SageMaker Studio를 사용하여 기본 제공 관찰 기능 및 추적 기능을 갖춘 목적별 워크플로, 자동화 및 거버넌스 도구를 구성 및 모니터링할 수 있습니다. 코드와 데이터를 모두 AWS 보안 서비스 및 기능도 통합을 통해서 보호할 수 있습니다.

DO NOT COPY ssleeeee1@gmail.com

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



Amazon SageMaker AI에서는 컨테이너화된 ML 훈련 워크로드용으로 최적화된 여러 가지 Amazon Elastic Compute Cloud(Amazon EC2) 컴퓨팅 환경이 제공됩니다. 즉, 최신 CPU, 그래픽 처리 장치(GPU) 및 사용자 지정 액셀러레이터 중 적합한 옵션과 업계 최고 수준의 네트워킹 및 스토리지 옵션을 선택할 수 있으므로 모델의 예산 관련 요구와 성능 요구 사항을 충족할 수 있습니다.

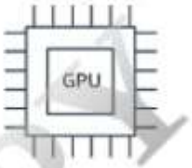
- **CPU 기반 훈련 인스턴스**는 컴퓨팅 집약도가 낮은 소형~중형 모델에 적합합니다. CPU 인스턴스는 다양한 병렬화나 특수 하드웨어 가속이 필요하지 않은 훈련 모델에 적합합니다. 데이터 전처리, 피처 엔지니어링 및 데이터세트가 비교적 작은 모델 훈련 등의 작업에서는 CPU 인스턴스를 선택하는 것이 좋습니다. AWS는 고성능 컴퓨팅 파워와 메모리 구성을 제공하는 최신 세대의 EC2 인스턴스를 포함하여 다양한 CPU 인스턴스를 제공합니다.
- **GPU 가속 인스턴스**는 행렬 연산 및 병렬 계산을 효율적으로 처리하므로 딥 러닝과 신경망에 적합한 옵션입니다. 이 인스턴스에서는 복잡한 모델과 대규모 데이터세트의 훈련 속도가 대폭 높아집니다. TensorFlow, PyTorch, MXNet, XGBoost 등의 ML 프레임워크를 사용하여 신경망으로 딥 러닝 모델을 훈련할 때는 GPU 인스턴스를 선택하면 효율적입니다. 성능과 비용의 균형을 고려한 다양한 GPU 인스턴스 중에서 선택합니다.
- **AWS Trainium 인스턴스**는 대규모 분산 훈련 전용으로 제작된 맞춤 제작형 ML 액셀러레이터를 사용합니다. Trainium 인스턴스는 고속 로컬 상호 연결을 사용하여 병렬 처리 계산 방식으로 분산 훈련을 스케일 아웃합니다. 대규모 분산 훈련용으로 최적화된 이러한 인스턴스에서는 기존 GPU 인스턴스에 비해 모델 수렴 속도가 빨라지며 훈련 시간을 단축합니다. 특히 대규모 언어 모델(LLM) 및 자연어 처리(NLP) 태스크에서 성능과 비용 효율성이 매우 우수합니다.

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

훈련용 GPU 선택 시 고려 사항

- 딥 러닝 모델 (CNN, RNN, Transformers)
- 대규모 데이터세트
- 반복 방식 모델 개발
- GPU 가속화에 최적화된 모델 또는 프레임워크
- 성능, 속도, 비용 간의 균형



19

ML 워크로드에 적합한 컴퓨팅 환경을 선택하면 최적의 성능, 비용 효율성 및 확장성을 달성하는 데 도움이 됩니다. CPU로도 처리할 수 있는 ML 태스크도 있지만, 딥 러닝 모델 훈련 등의 태스크에서는 대개 GPU를 사용하는 것이 좋습니다. ML 훈련에 GPU 인스턴스를 할지 여부를 결정할 때 고려해야 하는 요인은 다음과 같습니다.

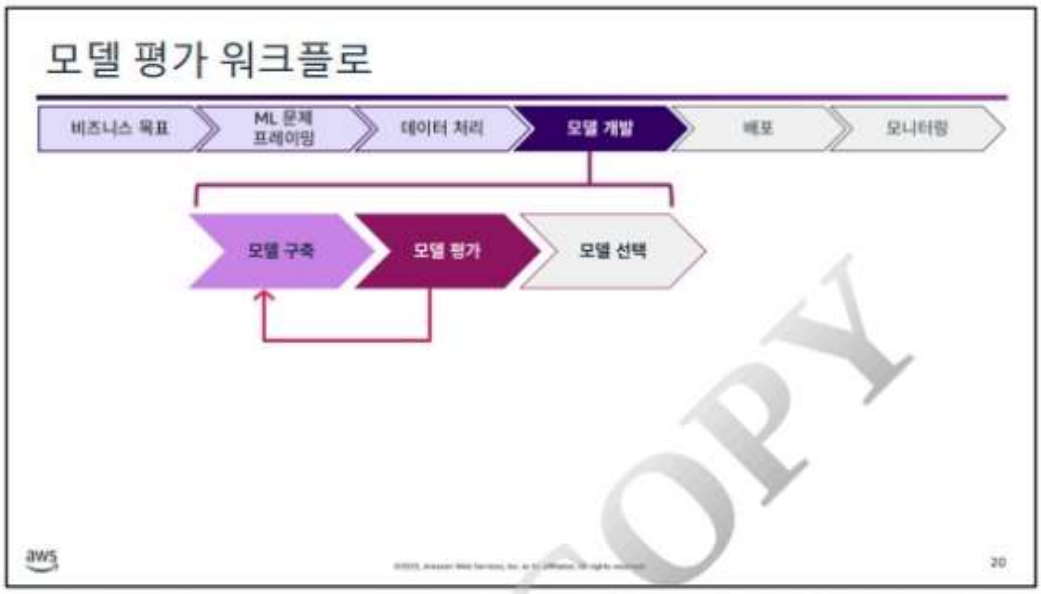
- 모델 복잡도: GPU는 합성곱 신경망(CNN), 순환 신경망(RNN) 등의 딥 러닝 모델을 훈련할 때 특히 유용합니다.
- 데이터세트 크기: 대용량 데이터세트(특히 고해상도 이미지 또는 대량의 텍스트 모음) 사용 시에는 GPU 가속으로 인해 훈련 성능이 대폭 개선됩니다.
- 훈련 시간 요구 사항: 개발 주기에서 훈련 속도를 높여야 하는 경우 GPU를 선택하면 반복 모델 개발을 훨씬 더 빠르게 진행할 수 있습니다.
- 프레임워크 최적화: 대다수 ML 프레임워크는 GPU 가속을 지원하도록 최적화되어 있으므로 이러한 인스턴스에서 효율성이 더 높아집니다.
- 가성비: GPU를 사용하는 경우 성능이 높아지는 동시에 비용도 늘어납니다. 그러므로 구체적인 사용 사례에서 필요한 성능과 사용 가능한 비용을 적절히 절충하여 GPU 사용 여부를 평가해야 합니다.

기계 학습 엔지니어는 이러한 요인을 면밀히 평가하여 GPU 인스턴스 사용으로 인한 성능 개선이 특정 프로젝트에서 추가로 발생하는 비용을 감수할 수 있는 수준인지를 결정해야 합니다.

DO NOT COPY ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



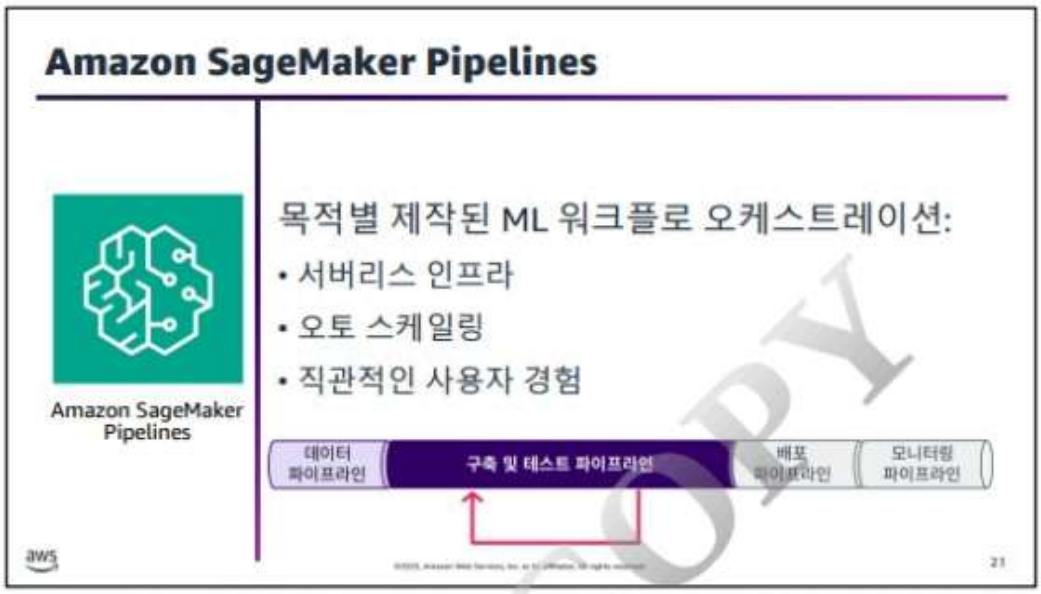
SageMaker AI에서는 모델 평가가 지원되므로 모델이 새 데이터에서 정확한 예측을 생성하는지 여부를 파악할 수 있습니다. 모델을 평가할 때는 모델 훈련 중에 사용되지 않은 테스트 데이터세트에서 실제 값과 예측 값을 비교할 수 있습니다. ML 문제의 유형에 따라 손실 함수와 지표를 선택합니다. 예를 들어, 어떤 사용 사례는 더 많은 오류를 허용할 수 있고, 다른 사용 사례는 거짓 긍정이나 거짓 부정에 민감합니다. 비즈니스 사용 사례를 토대로 요구 사항을 충족하는 지표 값을 결정해야 합니다. 진행하려는 사용 사례에서 예측과 평가 지표의 인적 검토가 필요한지 여부도 고려해야 합니다.

평가 기준에 맞는 모델 개발은 반복 프로세스로 진행해야 하는 경우가 많습니다. 모델 구축 및 테스트용 파이프라인을 생성하면 모델 훈련 및 평가를 반복적으로 수행할 수 있습니다.

DO NOT COPY
ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

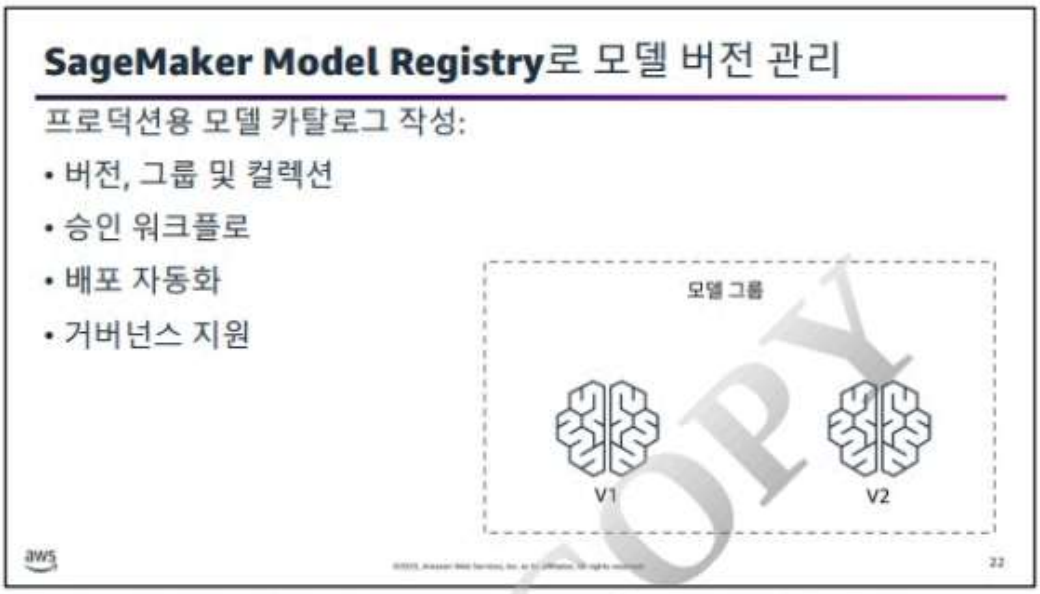


모델 개발은 일정한 방향에 따라 진행되는 순차적인 프로세스가 아닙니다. 즉, 여러 단계를 반복하면서 각 단계의 결과 및 확인된 인사이트에 따라 개발 방식을 개선해야 하는 경우가 많습니다.

SageMaker AI에는 ML 워크플로를 오케스트레이션하는 데 사용할 수 있는 Amazon SageMaker Pipelines 등의 기능이 포함되어 있습니다. 반복적인 하이퍼파라미터 튜닝과 같은 태스크 중에, SageMaker Pipelines는 오케스트레이션 컴퓨팅 리소스 프로비저닝, 규모 조정, 종료를 자동으로 수행합니다. SageMaker Studio 콘솔의 시각적 편집기에서 파이프라인을 생성 및 관리할 수도 있고 SDK, API 또는 JSON을 사용하여 프로그래밍 방식으로 파이프라인 생성과 관리를 수행할 수도 있습니다.

DO NOT COPY
ssleeeee1@gmail.com

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



모델 개발의 마지막 구성 요소에 해당하는 단계는 비즈니스 요구 사항을 충족하는 모델을 선택하여 배포를 준비하는 것입니다.

모델 훈련, 튜닝 및 평가 결과에 따라 요구 사항에 가장 잘 맞는 모델을 선택합니다. 그리고 거버넌스와 모니터링을 위해 기준 지표와 데이터 계보 등의 모델 메타데이터를 추적해야 합니다. SageMaker AI에는 모델 버전을 추적할 수 있는 Amazon SageMaker Model Registry와 거버넌스를 지원하는 추가 기능이 포함되어 있습니다.

예를 들어, 기계 학습 엔지니어는 반복성과 감사를 위해 모델 버전을 관리해야 합니다. SageMaker Model Registry는 기계 학습 모델을 관리하기 위한 중앙 집중식 리포지토리입니다. SageMaker Model Registry에서는 수명 주기 전반에 걸쳐 모델의 **버전 관리, 추적** 및 **거버넌스**를 수행할 수 있는 기능이 제공됩니다.

SageMaker Model Registry의 주요 기능 중 하나는 모델 버전을 등록하고 배포하는 기능입니다. 모델을 훈련하거나 미세 조정할 때마다 새 모델 버전을 생성할 수 있습니다. 이 방식에서는 각 버전을 고유하게 식별 및 추적합니다.

SageMaker Model Registry에서 포함된 몇 가지 기능은 아래와 같습니다:

- 모델 등록: 모델 파일 및 관련 메타데이터를 포함한 모델 아티팩트를 제공합니다. SageMaker Model Registry가 모델에 고유한 버전 번호를 할당합니다.
- 모델 승인 워크플로: ML 파이프라인 내에서 등록된 모델의 승인 상태를 관리할 수 있습니다. 프로덕션 환경에 배포하도록 승인된 버전을 제어함으로써, 철저한 테스트와 검증이 완료된 모델만 릴리스할 수 있습니다. 시간의 경과에 따라 모델 변경 사항을 감사하면서 필요시에는 이전 버전으로 롤백할 수 있습니다.
- 모델 그룹: 다양한 모델 버전을 적절한 그룹에 포함할 수 있습니다. 이 기능을 사용하여 특정 문제를 해결하도록 훈련한 모든 모델을 그룹화하고 추적할 수 있습니다.
- 컬렉션: 컬렉션은 여러 모델 그룹을 포함할 수 있는 상위 수준의 조직 단위입니다. 이러한 컬렉션 계층 구조에서 모델 카탈로그를 관리하고 살펴볼 수 있습니다.
- 모델 배포: 지속적 통합 및 지속적 전달(CI/CD) 파이프라인을 통해 모델 배포를 자동화할 수 있으므로

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

프로덕션 환경에서 항상 승인된 최신 모델 버전을 실행할 수 있습니다.

DO NOT COPY
ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

SageMaker AI를 사용하는 자동 모델 튜닝

하이퍼파라미터 튜닝 자동 수행:

- 훈련 작업 실행
- 성능이 가장 좋은 모델 확인
- 최적 하이퍼파라미터 값 제공

23

Amazon SageMaker AI에는 이 프로세스를 매우 간편하게 진행할 수 있는 **기능이 포함되어** 있습니다. **SageMaker AI 자동 모델 튜닝(AMT)**은 ML을 사용하여 하이퍼파라미터 튜닝을 수행해 모델의 최적 버전을 찾습니다. 기본적으로 베이지안 최적화를 사용하여 데이터세트와 알고리즘과 지정한 하이퍼파라미터 범위를 사용하는 여러 훈련 작업을 실행하여 최적의 하이퍼파라미터 구성을 효율적으로 찾습니다. 랜덤 검색, 그리드 검색, 하이퍼밴드 기법도 지원됩니다. 그런 다음, 선택한 지표에 따라 성능이 가장 좋은 모델을 식별하고 해당 모델을 생성하는 데 사용되는 하이퍼파라미터 값을 제공합니다. SageMaker AI AMT를 사용하면 시간을 절약하고 리소스를 효율적으로 사용할 수 있으며 모델 성능을 개선할 수 있습니다.

SageMaker AI AMT 작업을 설정하려면 각 하이퍼파라미터의 적절한 범위를 선택합니다. 모델의 목표에 맞는 목표 지표를 선택하고, 시간 및 리소스 제약에 따라 최대 훈련 작업 수를 지정합니다. 리소스를 절약하려면 조기 중지 등의 기능을 사용합니다.

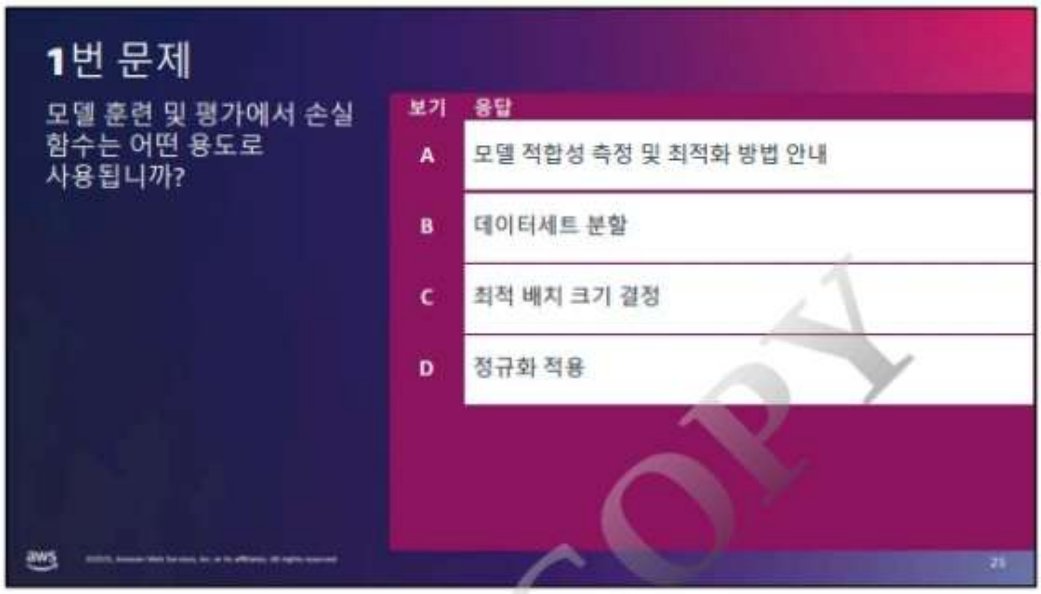
DO NOT COPY ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

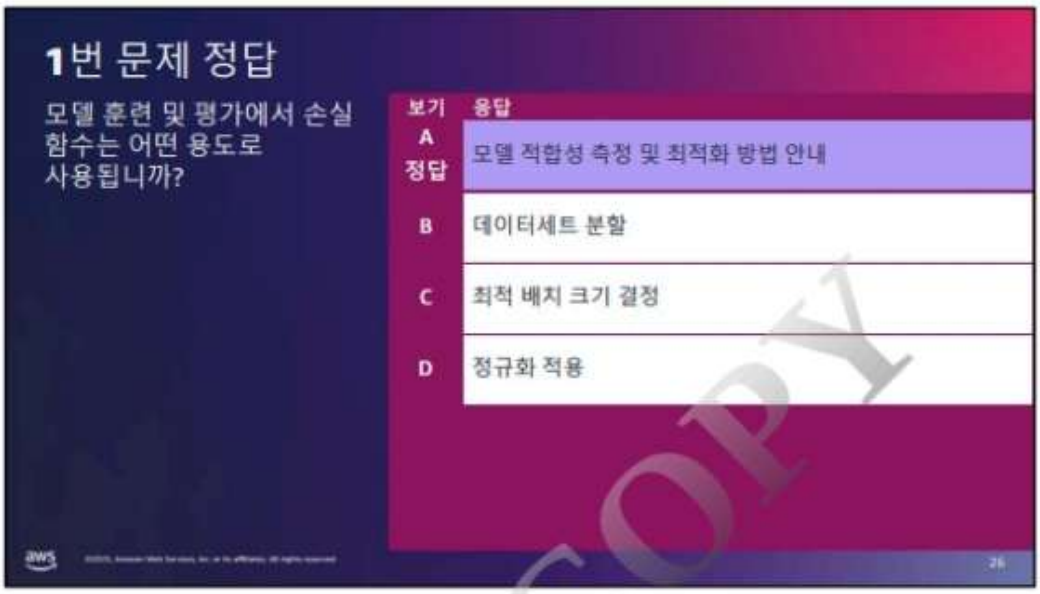
DO NOT COPY
ssleeeee1@gmail.com

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



DO NOT COPY
ssleeeee1@gmail.com

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



정답은 다음과 같습니다.

- A: 손실 함수는 모델의 예측과 실제 값 간의 차이를 수치로 계산합니다 그리고 모델 적합성을 측정하여 이 오차가 최소화되도록 모델 가중치를 조정할 것을 최적화 알고리즘에 알립니다.

오답은 다음과 같습니다.

- B: 데이터세트를 훈련 세트, 검증 세트, 테스트 세트로 분할합니다. 테스트 세트와 검증 세트에서는 모델 훈련에 사용되지 않은 데이터에서 모델 성능을 측정합니다. 이렇게 측정한 성능이 미확인 프로덕션 데이터에 대한 모델 성능입니다. 데이터 분할은 손실 함수와는 관계가 없는 별도의 단계입니다.
- C: 배치 크기, 학습률 등의 하이퍼파라미터도 튜닝할 수는 있지만 별도의 작업을 통해 설정하는 항목이며, 모델의 손실 함수가 이러한 하이퍼파라미터의 값을 결정하는 것은 아닙니다.
- D: 정규화는 과대적합 방지에 사용되는 기법이며 손실 함수와는 용도가 다릅니다.

DO NOT COPY ssleeeee1@gmail.com

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

2번 문제

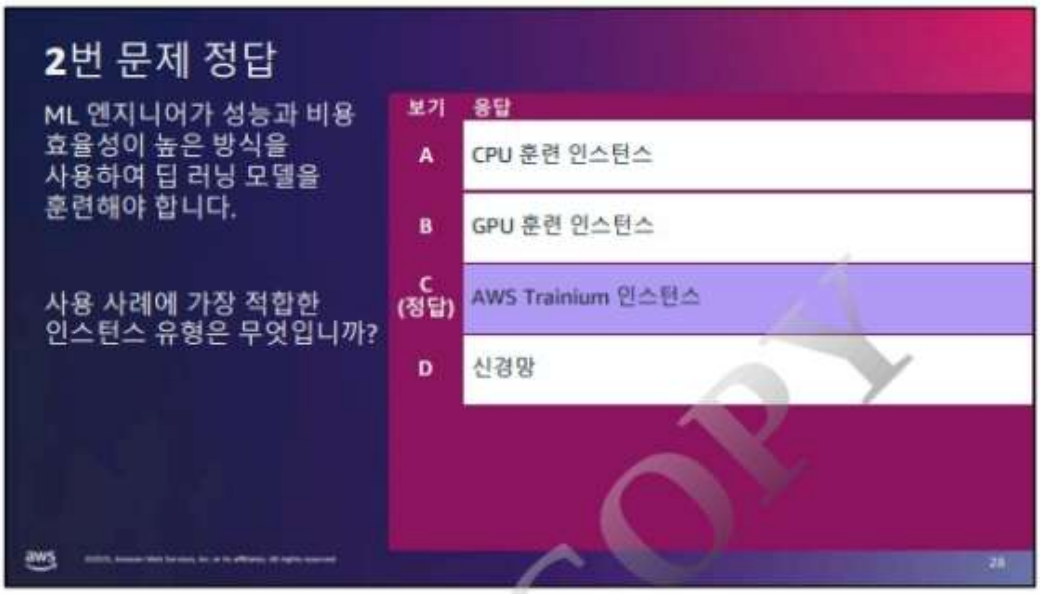
ML 엔지니어가 성능과 비용 효율성이 높은 방식을 사용하여 딥 러닝 모델을 훈련해야 합니다.

사용 사례에 가장 적합한 인스턴스 유형은 무엇입니까?

보기	응답
A	CPU 훈련 인스턴스
B	GPU 훈련 인스턴스
C	AWS Trainium 인스턴스
D	신경망

DO NOT COPY
ssleeeee1@gmail.com

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



정답은 다음과 같습니다.

- C: AWS Trainium은 클라우드에서 딥 러닝 모델을 훈련할 수 있도록 특별히 설계된 사용자 지정 빌드된 기계 학습 액셀러레이터입니다. Trainium 인스턴스는 훈련 워크로드, 특히 대규모 언어 모델과 자연어 처리 태스크에 탁월한 성능과 경제성을 제공합니다.

오답은 다음과 같습니다.

- A: CPU 인스턴스는 다양한 병렬화나 특수 하드웨어 가속이 필요하지 않은 훈련 모델에 적합합니다. 이러한 인스턴스는 데이터 사전 처리, 특성 추출, 상대적으로 작은 데이터세트나 컴퓨팅 집약적이지 않은 모델을 사용하는 훈련 모델과 같은 태스크에 적합합니다.
- B: GPU 인스턴스는 병렬 처리의 이점을 활용할 수 있는 딥 러닝 모델과 기타 기계 학습 알고리즘 훈련과 같은 컴퓨팅 집약적인 워크로드를 가속화하도록 설계되었습니다.
- D: 대량의 데이터를 처리하도록 다양한 컴퓨팅 파워가 필요한 딥 러닝 모델입니다.

DO NOT COPY
ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

3번 문제

개발자가 통합 개발 환경에서 모델 구축 및 훈련 시에 사용할 수 있는 Amazon SageMaker AI 기능은 무엇입니까?

보기	응답
A	SageMaker 자동 모델 튜닝(AMT)
B	SageMaker Model Registry
C	SageMaker Pipelines
D	SageMaker Studio

DO NOT COPY
ssleeeee1@gmail.com

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

정답은 다음과 같습니다.

- D: SageMaker Studio는 ML 모델 구축 및 훈련을 비롯한 ML 워크플로용으로 웹 기반 환경을 제공합니다. ML 실무자는 JupyterLab, 코드 편집기, RStudio 등 사용법을 잘 알고 있는 통합 개발 환경 중에서 선택할 수 있습니다.

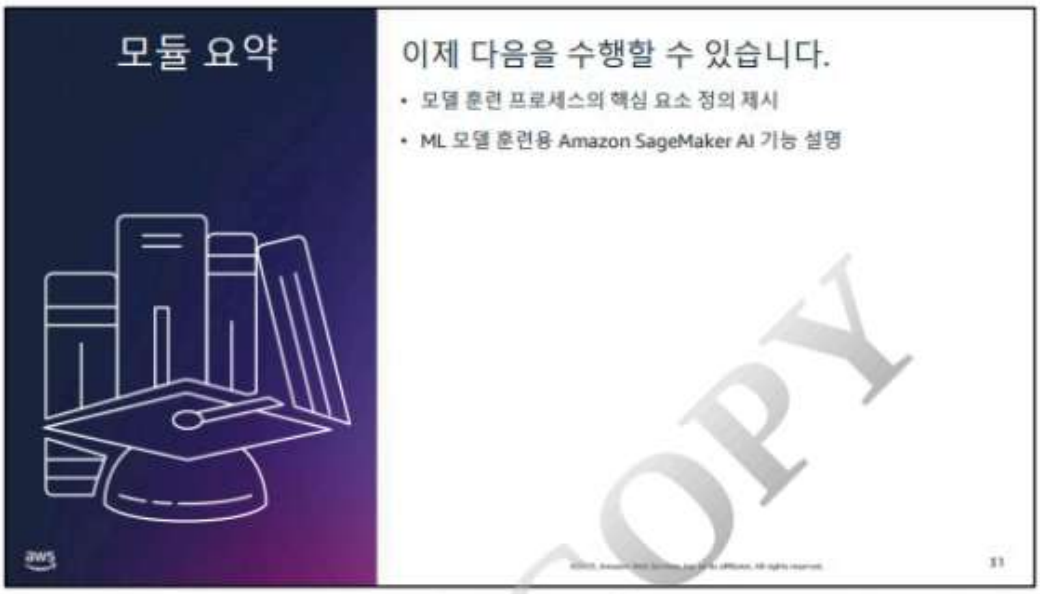
오답은 다음과 같습니다.

- A: SageMaker AMT는 모델 성능을 최대한 높일 수 있는 하이퍼파라미터 세트를 자동 식별합니다.
- B: SageMaker Model Registry는 모델 버전, 배포 상태, 계보 및 메타데이터를 관리할 수 있는 중앙 집중식 모델 카탈로그를 제공합니다.
- C: SageMaker Pipelines는 개발자가 모델 개발 워크플로를 자동화하는 데 사용할 수 있는 오케스트레이션 기능입니다.

DO NOT COPY
ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



이 모듈에서는 모델 훈련 프로세스의 핵심 요소 및 모델 훈련에 SageMaker AI를 사용하는 방식을 살펴보았습니다.

DO NOT COPY
ssleeeee1@gmail.com

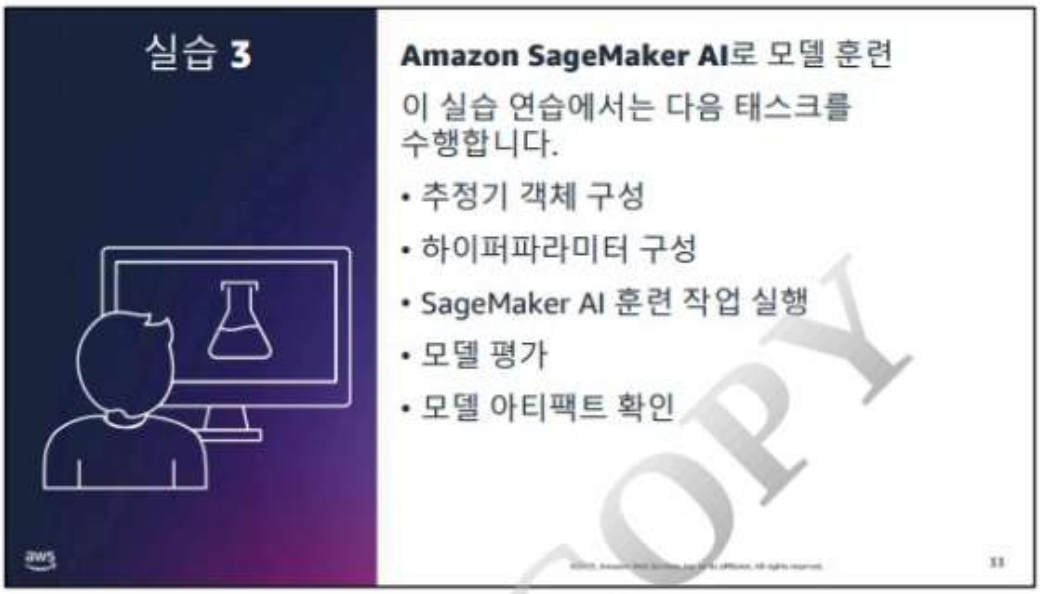
© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

DO NOT COPY
ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



실습 개요

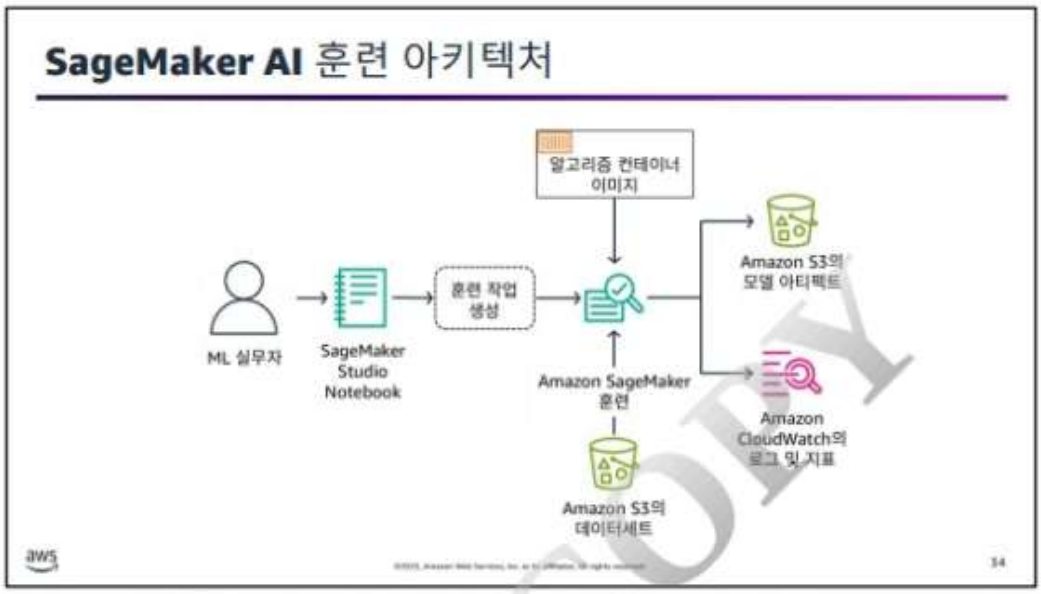
이 모듈에서는 Amazon SageMaker AI로 모델을 훈련할 때 수행하는 여러 태스크를 검토했습니다.

이 실습을 마치면 모듈에서 살펴보았던 SageMaker AI의 기능을 더욱 정확하게 이해할 수 있습니다. 또한 하이퍼파라미터를 구성하고 ML 모델이 성공적으로 훈련되었는지 검증하는 방법을 더 잘 이해할 수 있습니다.

DO NOT COPY
ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



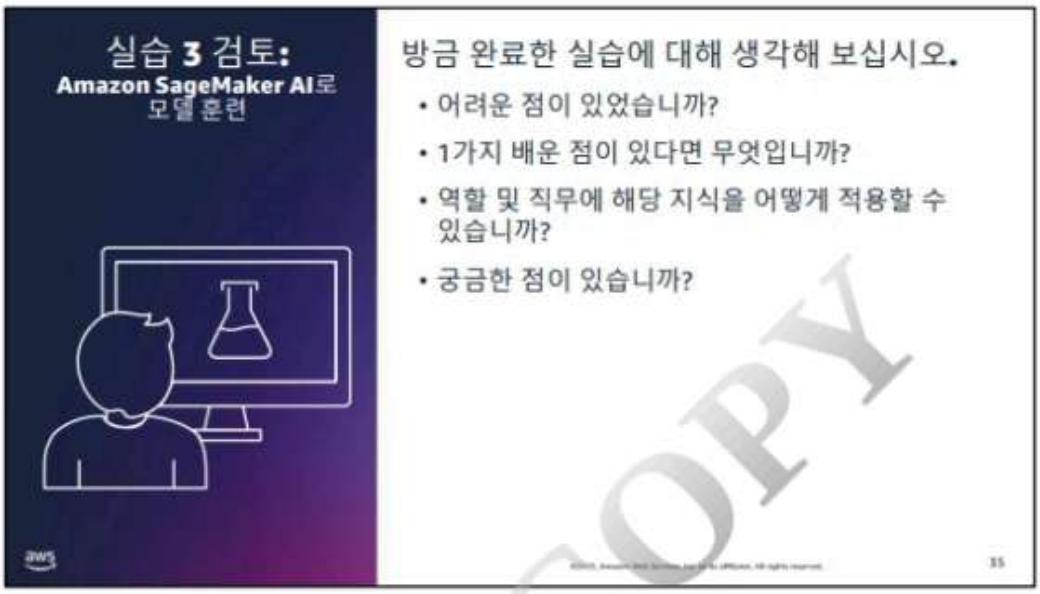
이제 논의된 도구 중 나머지 수업 동안 사용하게 될 일부를 살펴보겠습니다.

이 실습에서는 훈련 작업을 정의합니다. 모델 아티팩트는 Amazon Simple Storage Service(Amazon S3)로 전송되며 로그, 지표 및 훈련 작업은 Amazon CloudWatch에 캡처됩니다. Amazon SageMaker Studio는 는 전체 ML 수명 주기를 지원하는 통합 사용자 인터페이스를 제공합니다. AWS 리전 내의 Amazon SageMaker AI 도메인 내에서 사용자 프로파일이 정의된 후 AWS Identity and Access Management(IAM) 리소스와 연결됩니다. 실습에서는 고유한 URL을 통해 Amazon SageMaker Studio에 직접 연결하게 됩니다.

DO NOT COPY
ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



DO NOT COPY
ssleeeee1@gmail.com

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

이 모듈에서는 모델 훈련 프로세스의 핵심 요소에 대해 다룹니다. 구체적으로는 간소화되고 효율적인 훈련 프로세스를 위해 모델 훈련 기간을 단축하는 데 흔히 사용되는 방식에 대해 알아봅니다.

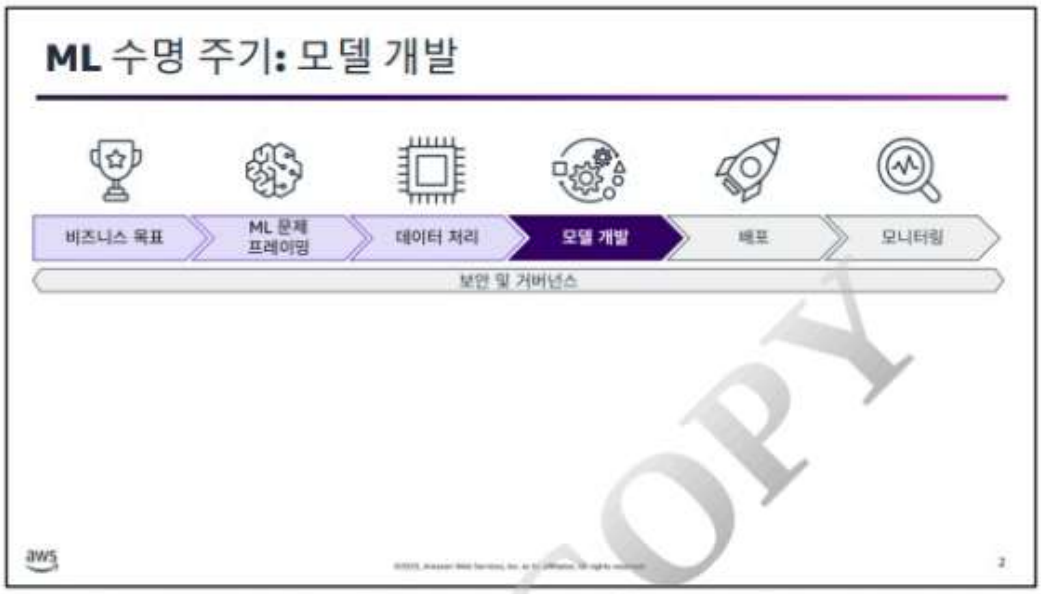
먼저 편향 완화 및 모델 성능 방법을 검토하고, 모델 과대적합과 과소적합을 방지하는 방법을 알아봅니다. 그런 다음, 여러 방법을 함께 활용하여 모델 성능을 개선하는 방법과 하이퍼파라미터 튜닝을 사용하여 최적화된 모델 결과를 생성하는 방법을 살펴봅니다.

또한 이 모듈에서는 다양한 모델 크기 및 모델 버전 관리를 살펴보고, Amazon SageMaker AI 서비스에서 모델 구체화 프로세스를 지원할 수 있는 방법을 살펴봅니다.

DO NOT COPY
ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

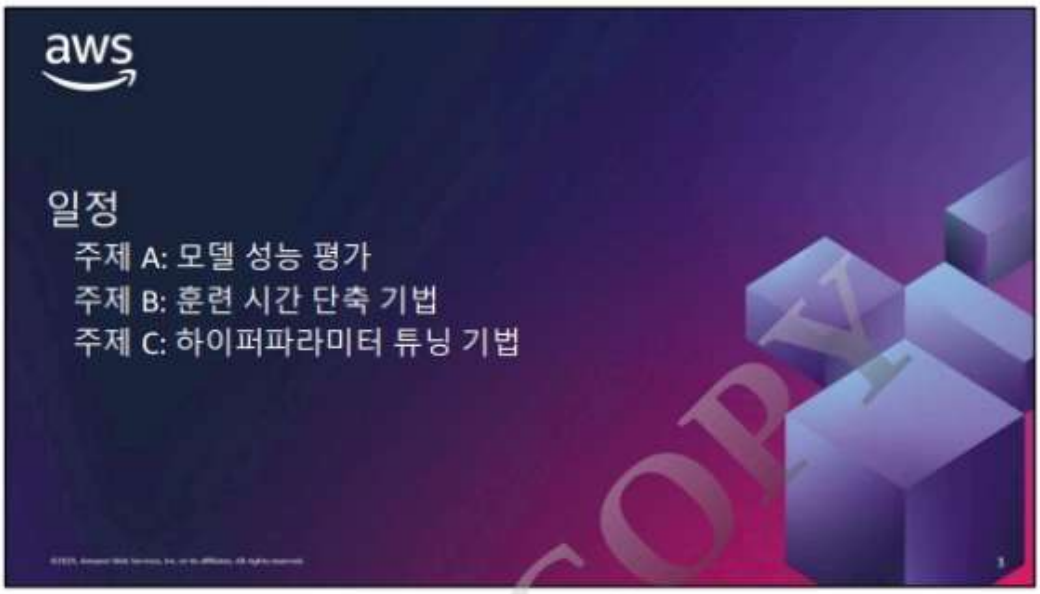


이 모듈에서는 기계 학습(ML) 수명 주기의 **모델 개발 단계**에 계속 중점적으로 살펴봅니다. 모델 훈련은 가장 성능이 가장 우수한 모델을 찾아내면서 데이터세트를 반복합니다. 이 훈련 과정에서 모델은 예측에 사용할 가중치 즉, 파라미터를 **학습**합니다. 훈련 작업이 완료되면 모델 성능이 비즈니스 목표를 달성하기에 적합한 수준인지를 평가해야 합니다. 예를 들어, 거짓 긍정 비율이 매우 낮아야 하는 사용 사례도 있는 반면 거짓 부정 비율이 매우 낮아야 하는 사용 사례도 있습니다.

하이퍼파라미터를 사용하여 훈련 프로세스를 구성할 수도 있습니다. 이상적인 하이퍼파라미터 세트는 훈련이 수렴을 향해 더 빠르게 진행되도록 도와주고 모델의 일반화 능력도 개선됩니다. 이 모듈에서는 효율적인 하이퍼파라미터 튜닝을 지원하는 기법 및 AWS 서비스와 기능을 살펴봅니다.

DO NOT COPY
ssleeeee1@gmail.com

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



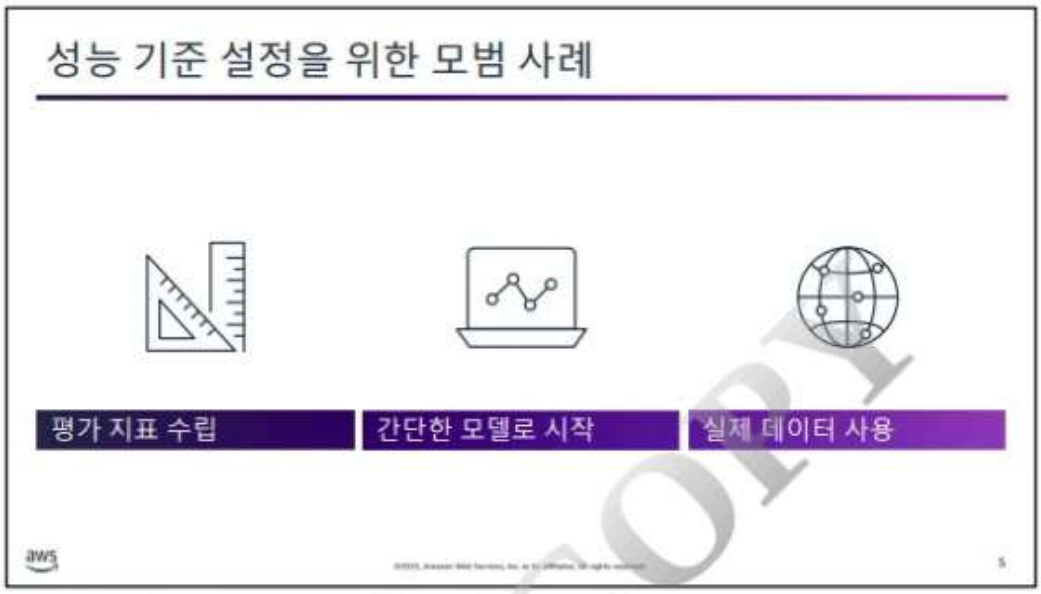
DO NOT COPY
ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

DO NOT COPY
ssleeeee1@gmail.com

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

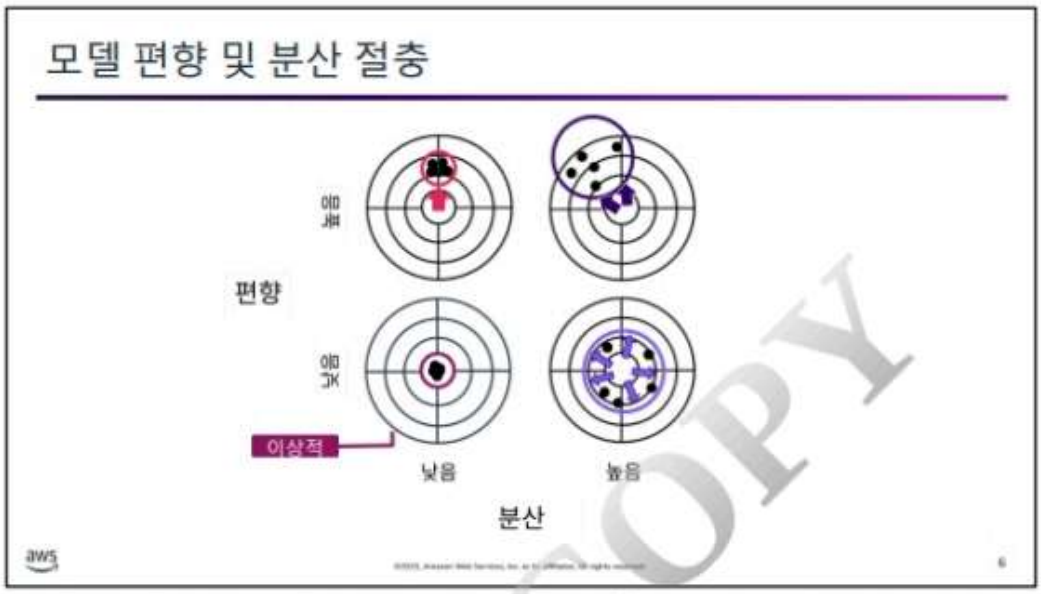


성능 기준 설정을 위한 3가지 모범 사례는 다음과 같습니다.

- 해결하려는 문제와 비즈니스 목표에 부합하는 적절한 평가 지표를 수립합니다.
- 선형 회귀와 같은 간단한 모델로 시작하여 검증 및 테스트 데이터세트에서 모델의 성능을 평가합니다. 이 초기 기준은 기준점 역할을 하며 문제의 복잡성에 대한 인사이트를 제공합니다.
- 실제 데이터 분포를 대표하고 유출 또는 편향이 없는 평가 데이터세트를 선택합니다.

DO NOT COPY
ssleeeee1@gmail.com

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



평가 지표를 선택할 때는 성능과 정확도 간의 절충점을 고려해야 합니다. 딥 러닝 모델이 단순한 모델보다 성능은 우수할 수 있지만 단순한 ML 모델은 설명 가능성이 더 높습니다.

모델 편향과 분산은 또 다른 절충점입니다.

- 모델 편향: 예측 동작의 불균형성. 편향이 높은 모델은 특정 부분군에서만 성능이 우수합니다. 훈련 데이터나 알고리즘에서도 편향이 발생할 수 있습니다.
- 모델 분산: 정확한 예측을 일관되게 제공하는 모델의 능력

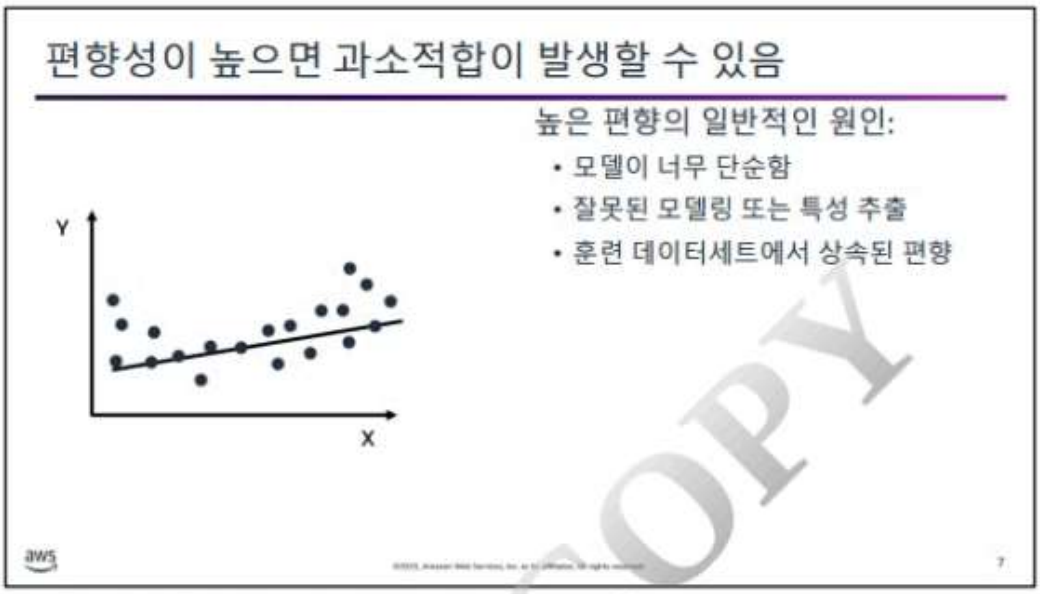
편향과 분산 간의 상호 작용 방식을 파악하려면 각기 다른 편향 및 분산 조합에 해당하는 슬라이드의 4가지 대상을 살펴보십시오.

- 높은 편향, 높은 분산: 결과가 중심부에서 벗어나 넓게 분산됩니다.
- 낮은 편향, 높은 분산: 결과가 중심부 근처에 있지만 넓게 분산됩니다.
- 높은 편향, 낮은 분산: 결과가 중심부에서 떨어진 위치에 밀집됩니다.
- 낮은 편향, 낮은 편차: 결과가 중심부와 가까운 위치에 밀집됩니다.

편향과 분산을 모두 최대한 낮추기보다는 편향과 분산의 적절한 균형점을 찾는 것이 가장 좋습니다.

DO NOT COPY ssleeeee1@gmail.com

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



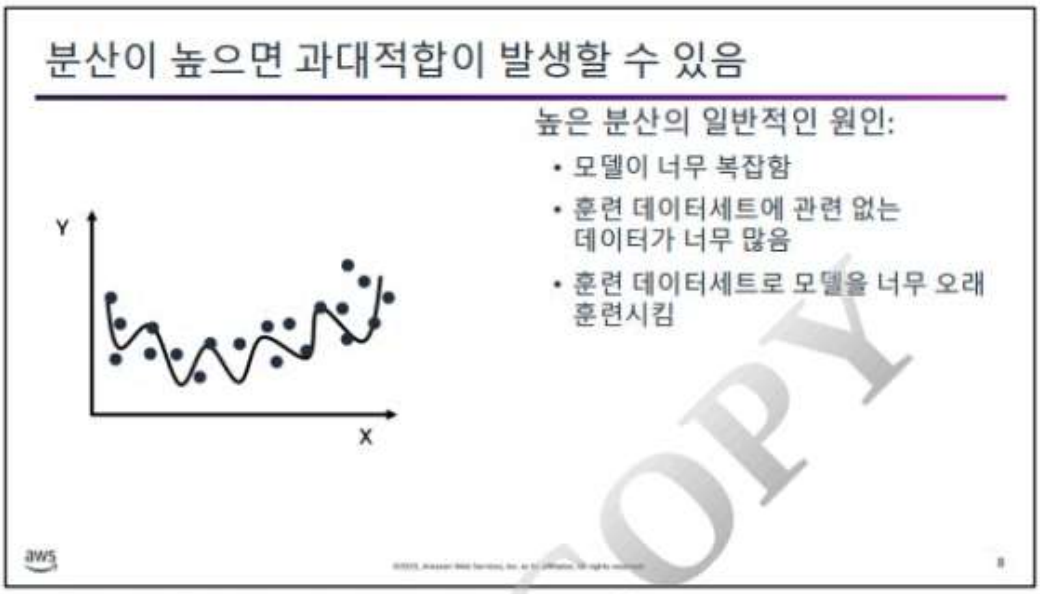
높은 수준의 모델 편향이 발생하는 일반적인 원인은 다음과 같습니다.

- 모델이 너무 단순함
- 잘못된 모델링 또는 특성 추출
- 훈련 데이터세트에서 상속된 편향

모델 편향을 측정하여 모델 과소적합을 탐지할 수 있습니다. 모델이 훈련되면 검증 집합을 사용하여 모델 성능을 측정할 수 있습니다. 모델의 성능이 훈련 집합과 검증 집합 모두에서 저조한 경우 높은 편향과 과소적합을 나타냅니다.

DO NOT COPY
ssleeeee1@gmail.com

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



높은 수준의 모델 분산이 발생하는 일반적인 원인은 다음과 같습니다.

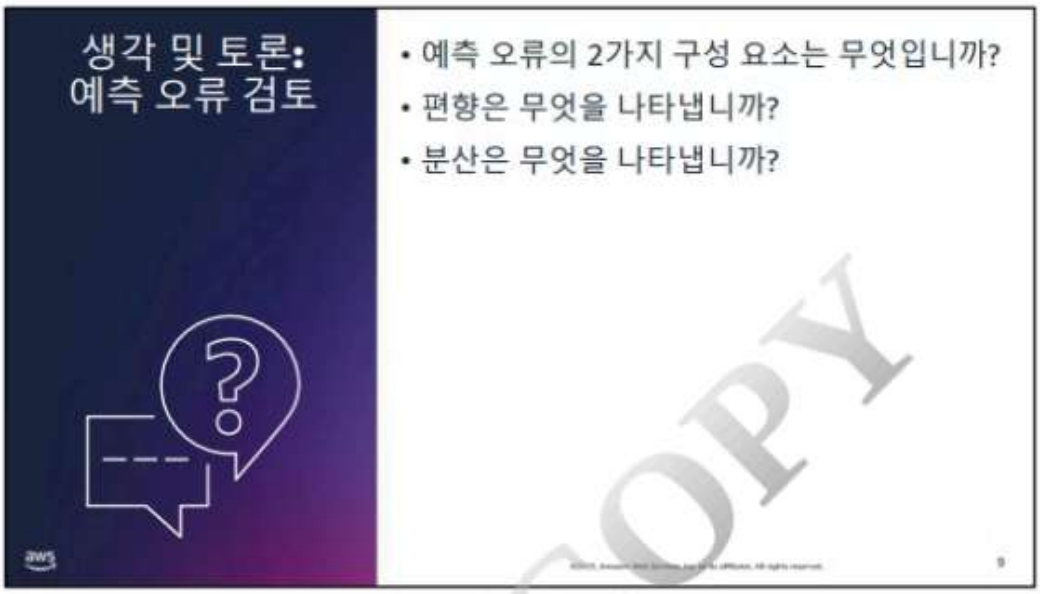
- 모델이 너무 복잡함
- 훈련 데이터세트에 관련 없는 데이터가 너무 많음
- 훈련 데이터세트에서 모델이 너무 오래 훈련됨

모델 분산을 측정하여 과대적합을 탐지할 수 있습니다. 모델이 훈련되면 검증 집합을 사용하여 모델 성능을 측정합니다. 모델의 성능이 훈련 집합에서 잘 수행되었지만 검증 집합에서는 잘 수행되지 않았으면 높은 분산과 과대적합을 나타냅니다.

DO NOT COPY
ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



예측 오류와 편향 분산의 절충을 빠르게 검토하면 분류 및 회귀 문제에 대한 평가 지표를 확인할 수 있습니다.

샘플 응답:

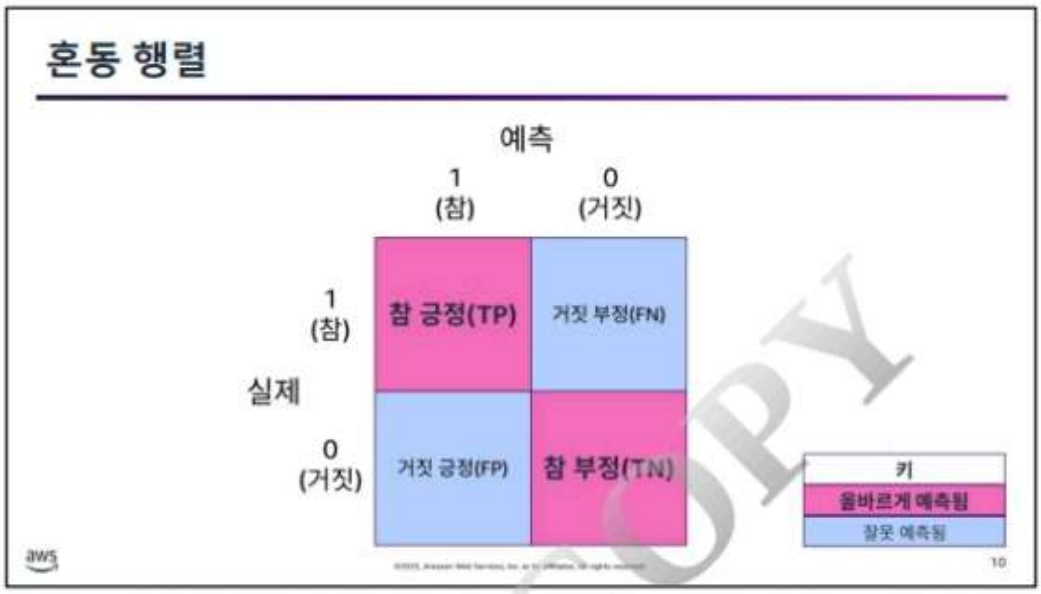
예측 오류의 2가지 구성 요소는 무엇입니까?
답변: 예측 오류는 편향 및 분산으로 구성됩니다. 이 두 구성 요소는 모델 성능의 서로 다른 측면을 나타내며 데이터 과학자가 모델 정확도의 절충점을 이해하는 데 도움이 됩니다.

편향은 무엇을 나타냅니까?
답변: 편향은 모델의 평균 예측과 정확한 값 간의 차이를 나타냅니다. 편향이 높다는 것은 모델이 지나치게 단순화되어 데이터의 중요한 패턴을 캡처하지 못한다는 것을 나타냅니다.

분산은 무엇을 나타냅니까?
답변: 분산은 주어진 데이터 포인트에 대한 모델 예측의 변동성을 나타냅니다. 분산이 높다는 것은 모델이 지나치게 복잡하고 훈련 데이터의 작은 변동에 민감하다는 것을 나타냅니다.

DO NOT COPY
ssleeeee1@gmail.com

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



테이블 형식 데이터에 대한 지도 학습에서는 분류 문제 또는 회귀 문제를 처리합니다. 분류 문제에 대한 평가 지표를 보는 것으로 시작합니다.

이진 분류 문제를 고려해 보십시오. 목표는 참 또는 거짓입니다. 레이블이 지정된 데이터세트의 경우 모델의 예측 오류는 혼동 행렬에서 요약될 수 있습니다.

이 2 x 2 행렬은 다음 4개의 가능한 출력에 대한 예측 결과를 캡처합니다.

- 참 긍정(TP)은 모델이 예측한 값과 실제 값이 모두 참인 경우입니다. 이 모델은 주민의 소득이 5만 달러 미만이므로 적격 주민으로 예측하며, 해당 예측은 올바릅니다.
- 참 부정(TN)은 모델이 예측한 값과 실제 값이 모두 거짓인 경우입니다. 이 모델은 주민의 소득이 5만 달러이므로 부적격 주민으로 예측하며, 해당 예측은 올바릅니다.
- 거짓 긍정(FP)은 모델이 예측한 값은 참인데 실제 값은 거짓인 경우입니다. 이 모델은 해당 주민이 적격이라고 예측하는데 실제 소득은 5만 달러가 넘으므로 해당 주민은 부적격입니다.
- 거짓 부정(FN)은 모델이 예측한 값은 거짓인데 실제 값은 참인 경우입니다. 이 모델은 해당 주민이 부적격이라고 예측하는데 실제 소득은 5만 달러 이하이므로 해당 주민은 적격입니다.

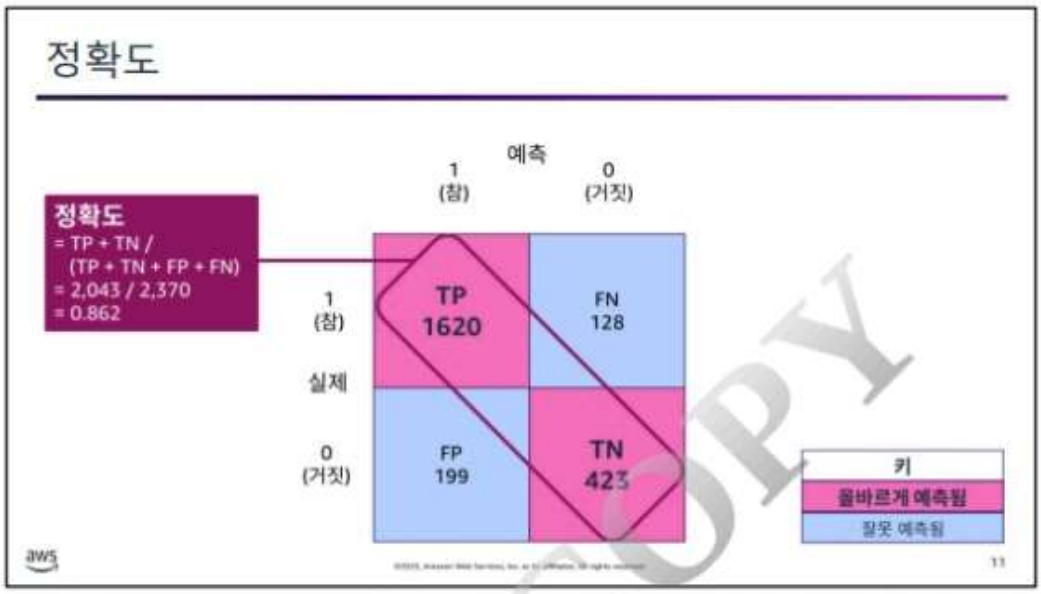
일반적으로 분류 문제에 대한 혼동 행렬은 n x n 행렬이고 이 경우 대상에는 n개의 고유 클래스가 있습니다. 자세한 내용은 '멀티클래스 모델 인사이트'(https://docs.aws.amazon.com/machine-learning/latest/dg/multiclass-model-insights.html)를 참조하십시오.

이어지는 슬라이드에서는 분류 문제와 관련하여 혼동 행렬을 기준으로 다음과 같은 성능 지표를 소개합니다.

- 정확도
- 정밀도
- 재현율
- F1 점수
- 곡선 아래 면적: 수신기 작동 특성(ROC) 곡선

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



정확도는 총 예측 수(TP + TN + FP + FN) 대비 일치하는 예측 수(TP + TN)의 비율입니다. 정확도는 데이터세트에 심각한 대상 불균형이 없는 한, 좋은 지표입니다. 이진 분류의 경우 대부분의 인스턴스는 true 또는 false라는 것을 의미합니다.

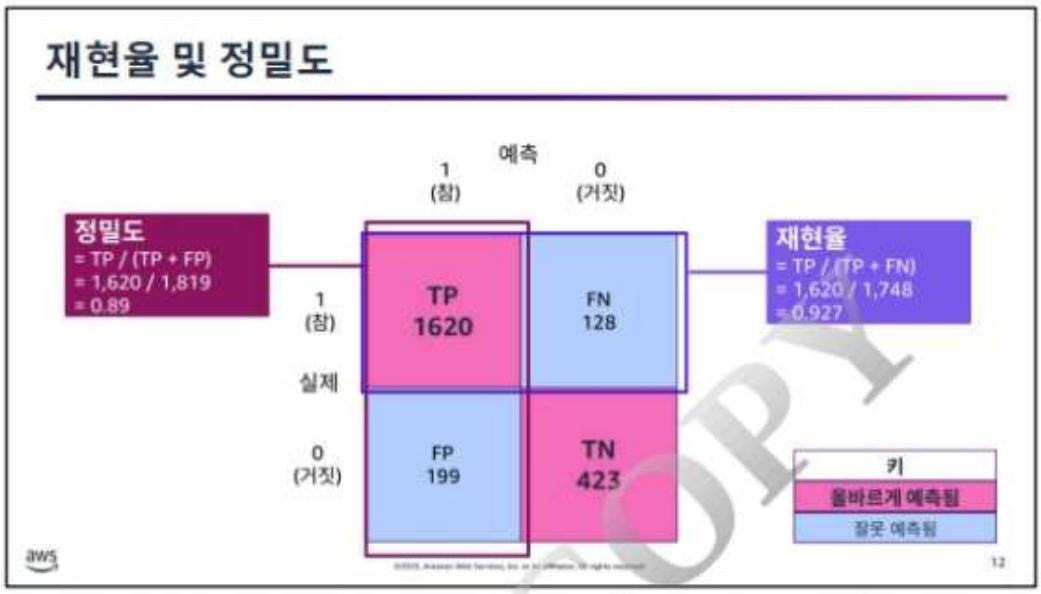
0.2%만 사기로 레이블이 지정된 대규모 재무 거래 데이터세트를 고려해 보십시오. 항상 부정이 아니라고 예측하는 것보다 더 나은 결과를 얻으려면 매우 높은 정확도가 필요합니다. 따라서 예측은 99.8% 정확합니다.

이 사례 연구에서 연간 소득이 1백만 달러를 초과하는 경우처럼 목표가 소득이 아주 높은 사람을 예측하는 것이라면 어떻습니까? 이 기준을 충족하는 데이터세트의 비율은 얼마나 됩니까?

표시된 혼동 행렬의 경우, 이러한 예측은 모델을 훈련 데이터세트로 훈련하고, 검증 데이터세트를 사용하여 평가하여 얻을 수 있습니다. 이 실습 중에는 비슷한 값을 포함하는 보고서를 검토해 볼 것입니다.

DO NOT COPY
ssleeeee1@gmail.com

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



목표 불균형이 있는 데이터세트나 거짓 긍정이나 거짓 부정 관련 특정 요구 사항을 충족해야 하는 사용 사례의 경우에는 2가지 추가로 지표가 도움이 됩니다.

정밀도는 실제로 맞는 긍정적인 예측의 비율입니다. 정밀도를 계산할 때는 참 긍정 수를 참 긍정 + 거짓 긍정의 합으로 나눕니다. 특정 비즈니스 상황에서 거짓 긍정으로 인한 비용이 높은 경우 정밀도가 좋은 지표가 될 수 있습니다.

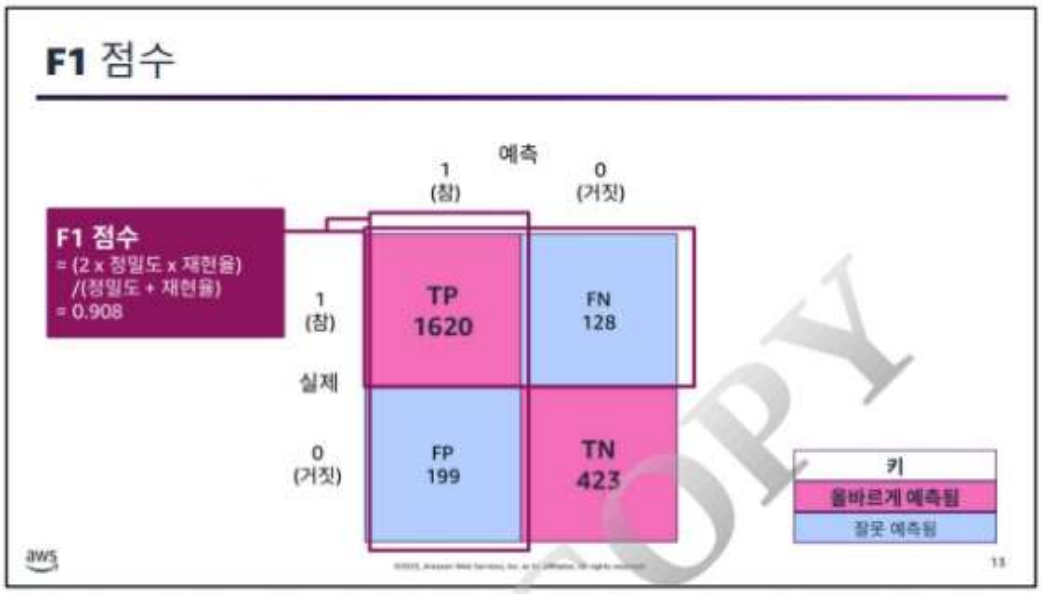
이메일을 '스팸' 또는 '스팸 아님'으로 식별하는 분류 모델을 생각해 보십시오. 이 사례를 보면 실제로는 이메일이 타당한데도 모델이 이메일을 스팸으로 레이블을 지정함으로써 사용자가 해당 이메일을 보지 못하게 되는 일은 없어야 합니다. 이 사례 연구에서는 거짓 긍정 항목이 많으면 부적격 주민을 대상으로 프로모션 예산이 낭비되는 결과가 초래됩니다. 이렇게 하는 데는 마케팅 비용이 발생하며, 주민들이 좀처럼 놓치지 않는 서비스를 대상으로 마케팅을 진행하게 되므로 해당 인구의 비율을 화나게 할 수도 있습니다.

재현율은 긍정으로 식별되는 정답 집합의 비율입니다. 재현율을 계산할 때는 참 긍정 수를 참 긍정 + 거짓 부정의 합으로 나눕니다. 특정 비즈니스 상황에서 거짓 부정으로 인한 비용이 높은 경우 재현율이 좋은 지표가 될 수 있습니다.

환자가 불치병을 가지고 있는지 여부를 예측해야 하는 모델을 생각해 보십시오. 이 경우 정밀도를 평가 지표로 사용하면 모델의 거짓 부정이 고려되지 않습니다. 이런 상황에서는, 이러한 병을 앓고 있는 환자에게 실제로 해당 질환에 걸렸다고 예측하는 것이 모델의 성공에 절대적으로 중요합니다. 이 사례 연구에서는 적격 주민에게 이용 가능한 지원 프로그램을 적절하게 제공하지 않으면 조직에 벌금이 부과될 수 있습니다. 이러한 상황에서는 재현율이 더 적합한 지표입니다.

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



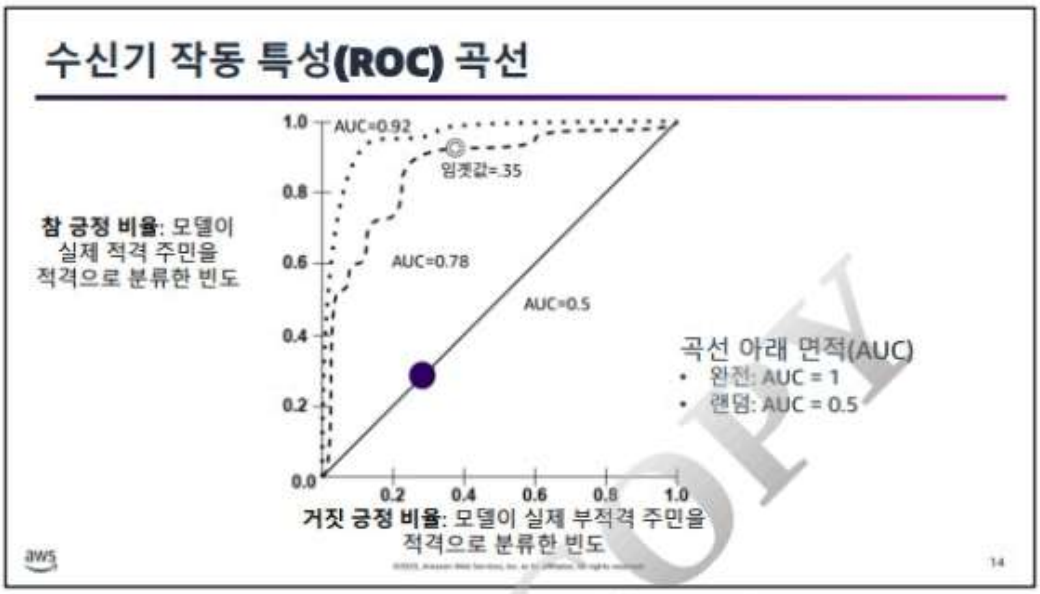
F1 점수는 정밀도와 재현율 사이의 균형 잡힌 조화를 나타내며, 두 측정값을 결합하여 0에서 1 사이의 값 하나를 생성하는 단일 지표라 할 수 있습니다. 이 예시의 F1 점수인 0.908은 만점인 1.0과 근접한 수치이므로 모델 성능이 매우 좋은 것입니다. F1 점수 계산 공식은 (2 x 정밀도 x 재현율)/(정밀도 + 재현율)인데, 정밀도와 재현율 사이의 극심한 불균형에 불이익을 주는 가중 평균을 구합니다.

슬라이드의 혼동 행렬에 나와 있는 모델의 경우 참 긍정(TP) 수가 1,620개이고 참 부정(TN) 수가 423개이므로 성능이 매우 좋은 것입니다. 모델이 긍정 사례와 부정 사례를 모두 정확하게 예측하는 것입니다. 반면 거짓 긍정과 거짓 부정 수는 비교적 적으므로(각각 199개, 128개) F1 점수도 높습니다. 즉, 두 오류 유형이 모두 최소한으로만 발생하도록 모델의 균형이 적절하게 유지되어 있는 것입니다.

F1 점수는 정확도만 사용하는 경우 오해의 소지가 있는 불균형 데이터세트를 사용할 때는 특히 유용합니다. 이 지표를 혼동 행렬과 함께 살펴보면 모델이 모든 카테고리에서 잘 작동한다는 것을 확인할 수 있습니다. F1 점수가 높다는 것은 정밀도나 재현율 중 하나가 높아지면 다른 하나는 낮아지는 문제가 발생하지 않는다는 의미이기 때문입니다. 그러므로 F1 점수는 정밀도나 재현율 중 하나만 확인하는 방식에 비해 신뢰도가 더 높은 지표라 할 수 있습니다.

DO NOT COPY
ssleeeee1@gmail.com

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



참고로, 분류 문제를 해석하는 데 사용되는 모델은 실제로 인스턴스가 특정 대상 분류에 속할 확률을 반환한다는 점입니다. 이진 분류 시나리오를 계속 살펴보면, 이러한 모델은 결과가 참일 확률을 반환합니다. 이 확률은 0과 1 사이의 값입니다. 확률 임곗값은 예측이 true로 해석될지 또는 false로 해석될지를 결정합니다. 예를 들어, 임계값 0.7의 경우 0.7보다 큰 예측은 true로 해석되지만 0.7보다 작은 값을 반환하는 예측은 false로 해석됩니다.

그렇다면 최선의 임계값은 어떻게 결정됩니까?

이 지점에서 수신기 작동 특성(ROC) 곡선이 작동하기 시작합니다. 이 곡선은 0에서 1 사이의 모든 임곗값에 대해 X축의 거짓 긍정 비율(FPR)과 비교하여 Y축의 참 긍정 비율(TPR)을 플로팅합니다.

TPR이 이전에는 재현율로 정의되었습니다. 이것을 민감도라고도 합니다. FPR 또는 폴아웃 = FP / (FP + TN)(실제 부정 값이 잘못 예측된 비율)

서로 다른 모델에 대한 ROC 곡선과 비교할 때 이 곡선은 경사가 클수록 모델 성능이 더 좋습니다. 이 품질을 캡처하는 지표는 ROC 곡선의 곡선 아래 면적 또는 AUC-ROC입니다.

이 그래픽은 3개의 다른 ML 모델에 대한 3개의 ROC 곡선을 보여 줍니다. 곡선 아래 면적이 가장 큰 모델이 가장 성능이 좋은 것으로 간주됩니다. 또한 각 곡선의 '변곡점'도 알아야 합니다. 이 변곡점은 모델에서 사용할 최선의 임곗값을 알려줍니다.

AUC-ROC 및 비주얼 표현에 대한 자세한 내용은 'ROC & AUC'(https://mlu-explain.github.io/roc-auc)를 참조하십시오.

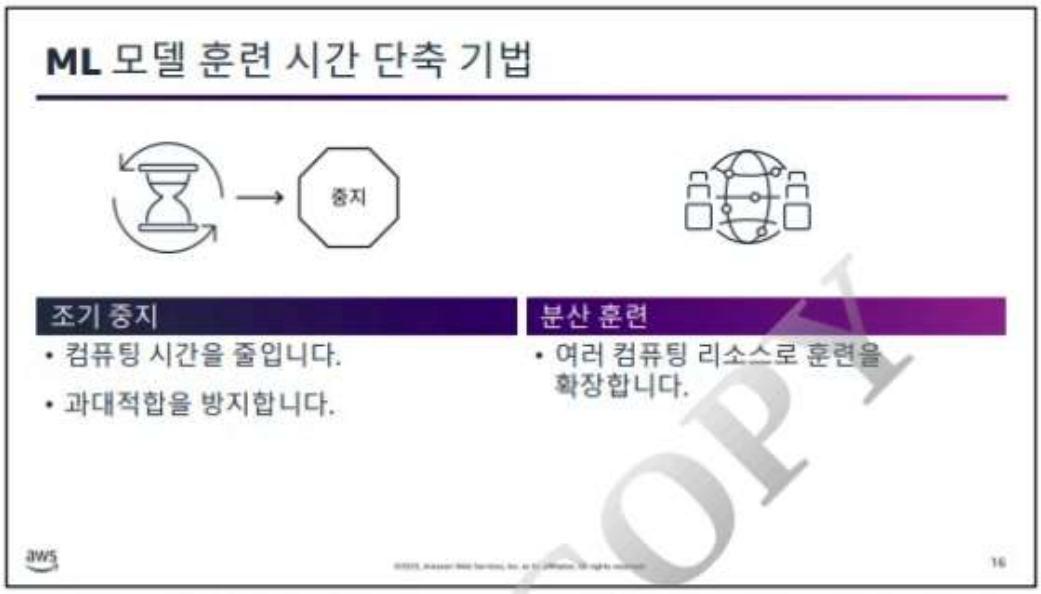
© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

ML 모델은 점점 더 복잡해지고 데이터 집약적이 됩니다. 이로 인해 훈련 시간이 길어져 모델 개발 주기가 크게 느려질 수 있습니다. 훈련에 필요한 시간을 줄이면 반복 및 실험 주기가 빨라져 모델을 더욱 효율적으로 개발할 수 있습니다. 그리고 컴퓨팅 리소스 사용 시 발생하는 비용도 줄일 수 있습니다. 따라서 모델 성능을 저하시키지 않으면서 훈련 시간을 안정적으로 줄일 수 있는 기법이 현장에서 많이 요구되고 있습니다. Amazon SageMaker AI는 이러한 문제를 해결할 수 있는 유용한 기능(예: 조기 중지 및 분산 훈련)을 제공합니다.

DO NOT COPY
ssleeeee1@gmail.com

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



ML 모델 훈련은 시간이 많이 걸릴 수 있으며 특히 여러 하이퍼파라미터 범위에서 훈련을 실행하거나 대규모로 모델을 훈련할 때는 시간이 매우 오래 걸릴 수도 있습니다. 모델 훈련 시간을 줄이면 훈련을 더 빠르게 반복하여 더 높은 품질의 모델을 완성하거나, 프로덕션 환경에 ML 모델을 더욱 신속하게 배포하거나, 모델 훈련 시의 컴퓨팅 비용을 줄일 수 있습니다. 이 섹션에서는 ML 모델 훈련 시간 단축을 위해 흔히 사용되는 2가지 기법을 살펴봅니다.

- **조기 중지**를 사용하면 과대적합을 방지하고 컴퓨팅 시간을 줄일 수 있습니다. 조기 중지는 검증 세트에서 모델 성능 개선이 멈출 때 ML 모델의 훈련 프로세스를 중단하는 정규화 기법입니다. 훈련 중에 홀드 아웃 검증 세트에서 모델의 성능을 모니터링하면 조기 중단으로 모델이 최적점에 도달한 시점을 감지할 수 있습니다. 이 단계에서는 훈련 프로세스가 중단되어 모델이 지나치게 복잡해지는 것을 방지하고 일반화 기능을 보존합니다. 중지 기준은 신중하게 선택해야 합니다.
- **분산 훈련**은 GPU 및 인스턴스와 같은 여러 컴퓨팅 리소스로 딥 러닝 모델 훈련 규모를 조정하는 데 사용되는 기법입니다. 모델이 커지고 데이터세트가 방대해지면 단일 머신에서 훈련하는 것이 불가능해집니다. 분산 훈련을 사용하면 여러 GPU, 인스턴스 또는 클러스터를 사용하여 훈련 시간을 가속화할 수 있습니다. 그러므로 단일 GPU 메모리에 저장하기에는 너무 큰 모델을 쉽게 훈련할 수 있습니다.

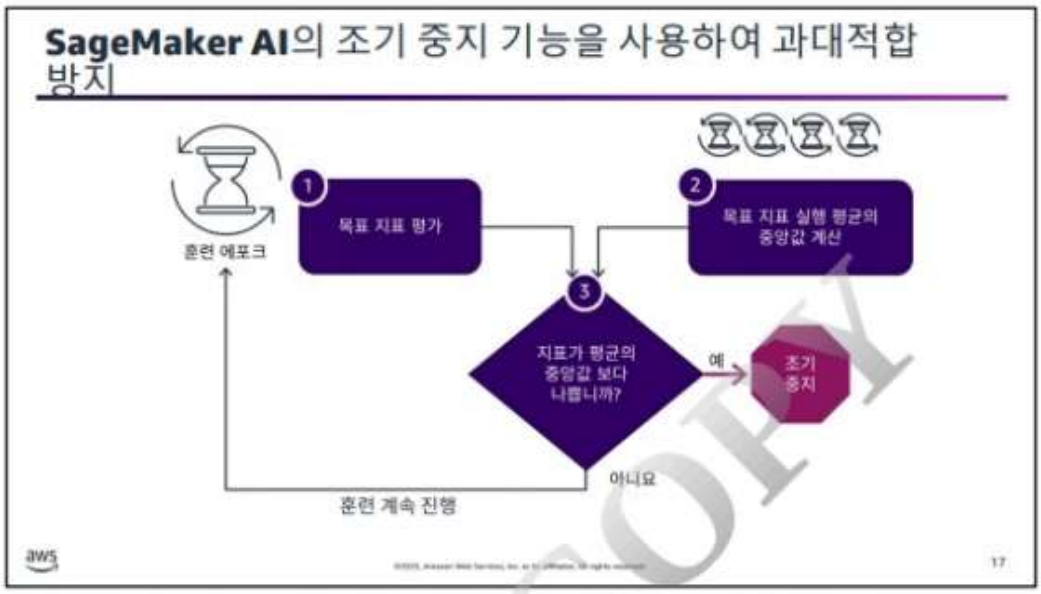
Amazon SageMaker AI에서는 이 두 기법이 모두 기본적으로 지원됩니다.

- SageMaker Automatic Model Tuning에 조기 중지 기능이 포함되어 있습니다.
- SageMaker 분산 훈련 라이브러리에서 데이터 및 모델 병렬 처리를 효율적으로 구현할 수 있습니다.

이러한 기법을 사용하면 훈련 시간을 줄일 수 있습니다. 특히 복잡한 대규모 모델이나 대용량 데이터세트를 사용할 때는 이러한 기법이 유용합니다.

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



Amazon SageMaker AI는 하이퍼파라미터 튜닝 중에 지원되는 기본 제공 알고리즘(CatBoost, 선형 Learner, XGBoost, Sequence to Sequence) 및 사용자 정의 알고리즘용 조기 중지 기능이 통합 제공됩니다. 사용자 정의 알고리즘 사용 시에는 각 에포크/반복 이후 코드가 목표 지표 값을 내보내야 합니다.

조기 중지 사용 시 SageMaker AI에서 수행되는 작업은 다음과 같습니다.

1. 각 에포크/반복에 대해 지정된 목표 지표(정확도, 손실, F1 점수 등)를 평가합니다.
2. 동일 에포크 시점까지의 이전 훈련 작업 지표 실행 평균과 중앙값을 계산합니다.
3. 지표 값이 중앙값에 비해 바쁘면 (최소값의 경우 지표 값이 더 높고 최대값의 경우 지표 값이 더 낮음) 현재 작업을 조기 중지합니다.

조기 중지를 구성할 수 있는 방법은 다음과 같습니다.

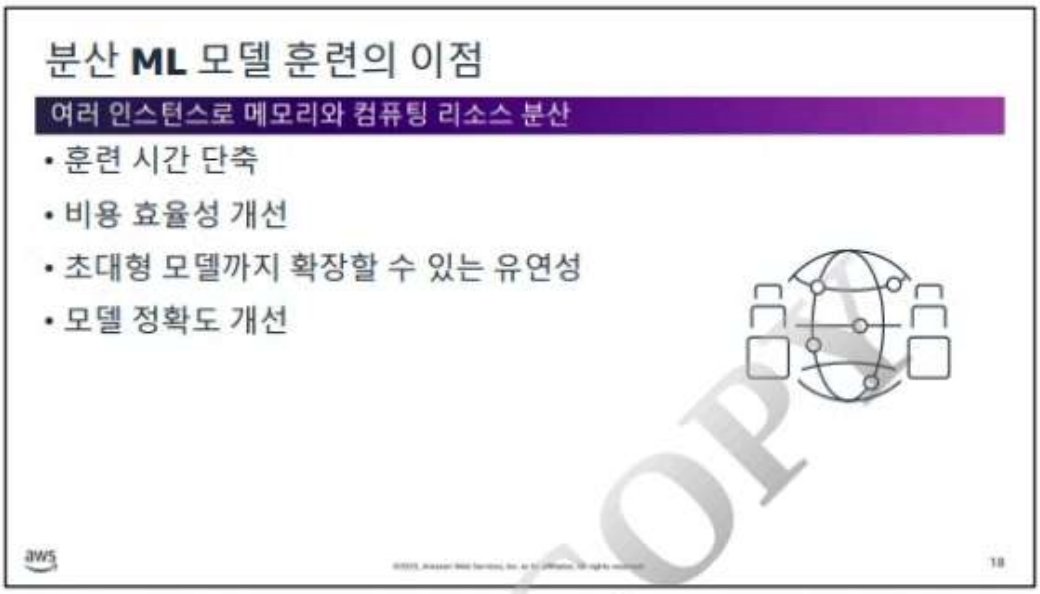
- SageMaker AI 콘솔(하이퍼파라미터 작업 생성 시)
- AWS SDK for Python(Boto3): TrainingJobEarlyStoppingType을 AUTO로 설정
- SageMaker AI Python SDK: early_stopping_type을 AUTO로 설정

조기 중지를 적절하게 활용하면 훈련에 필요한 계산 리소스와 시간이 감소하므로 더욱 효율적인 모델을 개발하고 출시 시간을 단축할 수 있습니다.

DO NOT COPY ssleeeee@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



분산 훈련에서는 대규모 기계 학습 모델을 빠르게 훈련하기 위해 여러 GPU나 인스턴스를 사용합니다. 분산 훈련의 주요 이점은 훈련 시간이 단축될 수 있도록 컴퓨팅 리소스를 확장할 수 있다는 점입니다. 병렬 처리는 데이터와 모델 파라미터를 여러 디바이스에서 분할하여 단일 GPU에서 훈련에 비해 훈련 시간을 크게 줄일 수 있습니다. 이처럼 컴퓨팅 리소스를 확장할 수 있으므로 메모리 제한으로 인해 단일 GPU에서 다루기 어려운 초대형 모델을 훈련할 수 있습니다. 또한 분산 훈련을 통해 GPU 리소스를 최대한 활용하여 사용 가능한 컴퓨팅 파워를 효율적으로 사용할 수 있습니다.

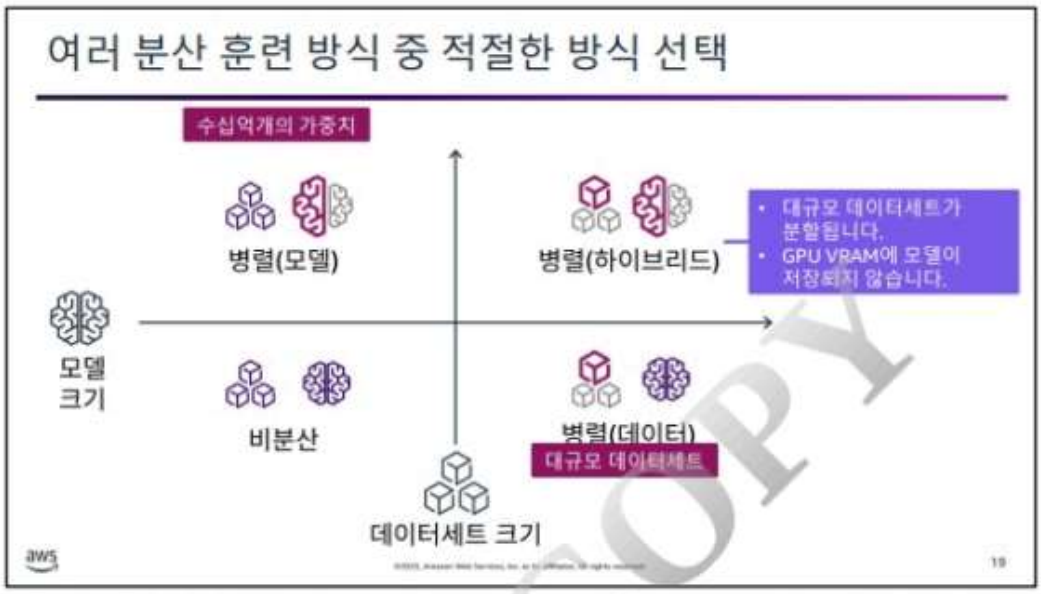
그 외 주요 이점은 다음과 같습니다.

- GPU 활용률 극대화를 통한 경제성
- 필요에 따라 컴퓨팅 리소스 규모를 유동적으로 조정하여 최첨단 대규모 언어 모델과 컴퓨터 비전 모델 훈련 진행
- 배치 크기가 커져도 분산 훈련을 통해 모델 정확도 향상

DO NOT COPY
ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



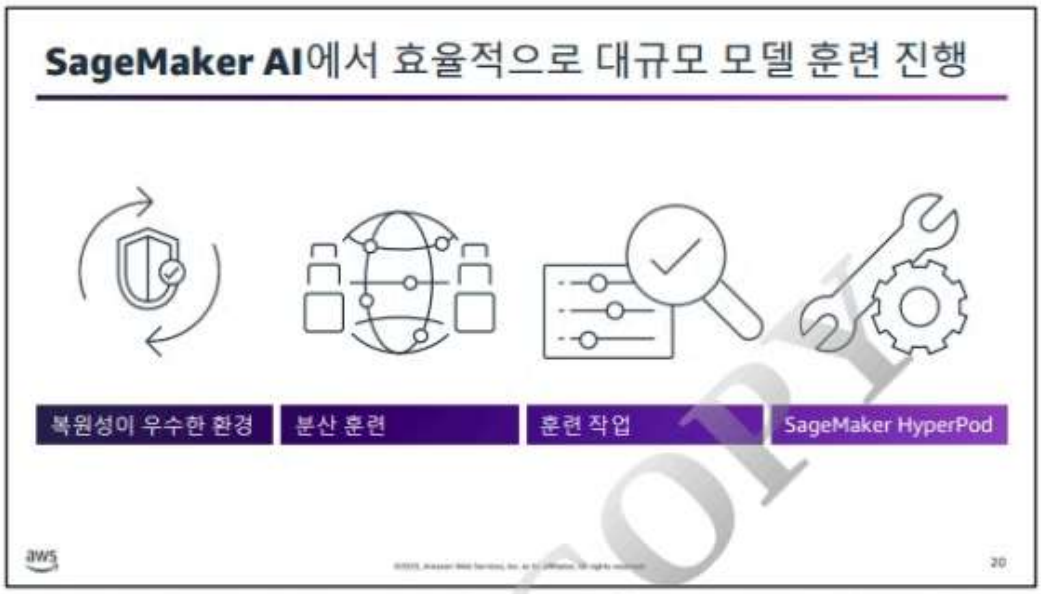
분산 훈련 방식을 사용할 때는 여러 컴퓨팅 리소스를 사용하도록 모델이나 데이터세트 중 하나 또는 둘 다를 분할할 수 있습니다. 적절한 방식은 모델과 데이터세트의 크기에 따라 달라집니다. 분산 훈련 방식에는 다음과 같은 항목이 포함됩니다.

- **비분산**: 소규모 모델 및 데이터세트에 적합한 방식입니다. 단일 CPU나 GPU에 맞는 훈련 작업의 경우 분산 컴퓨팅의 오버헤드가 발생하지 않습니다.
- **병렬(데이터)**: 데이터 병렬 처리에서는 훈련 데이터세트가 여러 GPU 또는 인스턴스로 분산되는 더 작은 배치로 분할됩니다. 각 디바이스는 모델의 복제본을 가져오고 할당된 데이터 배치에서 순방향 및 역방향 패스를 계산합니다. 각 배치 후 모델 가중치를 업데이트하기 전에 결과 그래디언트가 여러 디바이스에서 집계됩니다. 여러 디바이스에서 데이터를 분할하여 각 배치에 대한 계산을 병렬화할 수 있습니다. 이 방식에서는 단일 GPU의 Video RAM(VRAM)에 들어가는 것보다 더 큰 전체 배치 크기가 지원되므로 훈련 안정성과 모델 정확도가 향상될 수 있습니다. AWS에서 간편하게 데이터 병렬 처리 훈련을 진행할 수 있도록 Amazon SageMaker AI는 SageMaker AI 분산 데이터 병렬 처리(SMDDP) 라이브러리를 제공합니다.
- **병렬(모델)**: 모델 병렬 처리에서는 모델 자체를 여러 디바이스로 분할합니다. 훈련 모델이 너무 커서 단일 샘플도 한 디바이스의 메모리(예: GPU 메모리)에 적합하지 않을 수 있는 경우 이러한 분할이 필요합니다. 모델 병렬 처리에서는 모델 파라미터 또는 가중치의 다른 부분이 서로 다른 디바이스에 할당됩니다. 입력 데이터는 여러 디바이스에서 모델의 서로 다른 부분에서 처리되며 작업은 디바이스 전체에서 분할됩니다. 텐서 병렬 처리와 같은 기법은 해당 레이어의 텐서(가중치 또는 활성화)를 분할하여 디바이스 간에 큰 행렬 곱셈을 개별적으로 분할합니다. 모델을 복제하는 데이터 병렬 처리에 비해 모델 병렬 처리는 더 복잡합니다. 모델의 분산 구성 요소를 조정해야 하기 때문입니다. SageMaker AI 모델 병렬 처리 라이브러리와 같은 프레임워크는 모델 병렬 처리 구현을 간소화하는 기능을 제공합니다.
- **병렬(하이브리드)**: 대규모 데이터 세트가 여러 컴퓨팅 리소스로 분할됩니다. 각 데이터 하위 집합으로 각 모델 부분을 훈련하며, 훈련 결과는 모델 하나에 컴파일됩니다. 모델을 GPU RAM에 저장할 수 없는 대용량 데이터세트를 사용할 때는 이 방식이 적합합니다.

대규모 데이터세트로 훈련하는 경우 데이터 병렬 방법부터 시작합니다. 훈련 중에 메모리가 부족하면 모델 병렬 방식으로 전환하거나 하이브리드 및 데이터 병렬 처리를 시도해 볼 수 있습니다.

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



Amazon SageMaker AI에서는 대규모 모델을 효율적으로 훈련할 수 있는 여러 가지 기능이 제공됩니다.

- SageMaker AI는 훈련 시간을 최대 20%까지 단축할 수 있는 자체 복구 클러스터를 갖춘 탄력적인 환경을 제공합니다.
- SageMaker AI는 성능을 최대 20%까지 높일 수 있는 분산 훈련 라이브러리가 포함됩니다.

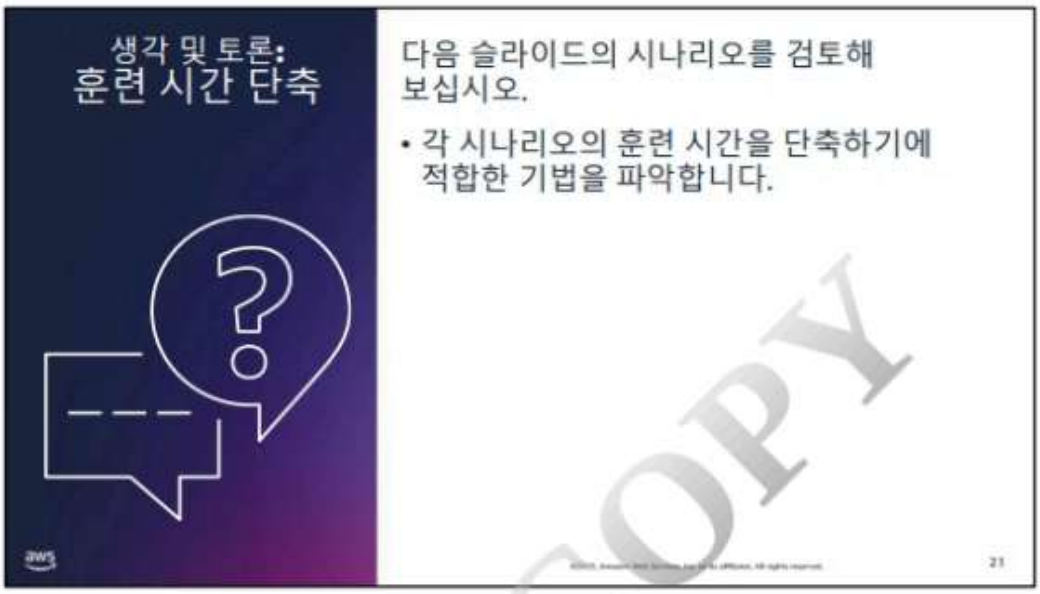
SageMaker AI에서는 대규모 모델의 훈련 작업을 실행할 수 있는 다양한 환경이 제공됩니다.

- SageMaker AI는 대다수 훈련 작업용 리소스의 프로비저닝과 규모 조정이 자동으로 처리되는 관리형 훈련 환경을 제공합니다.
- 더 많은 제어가 필요한 대규모 작업을 Amazon SageMaker HyperPod에서 실행하여 리소스 사용을 최적화할 수 있습니다. 컴퓨팅 환경과 워크로드 스케줄링을 계속 제어할 수 있습니다.

DO NOT COPY
ssleeeee1@gmail.com

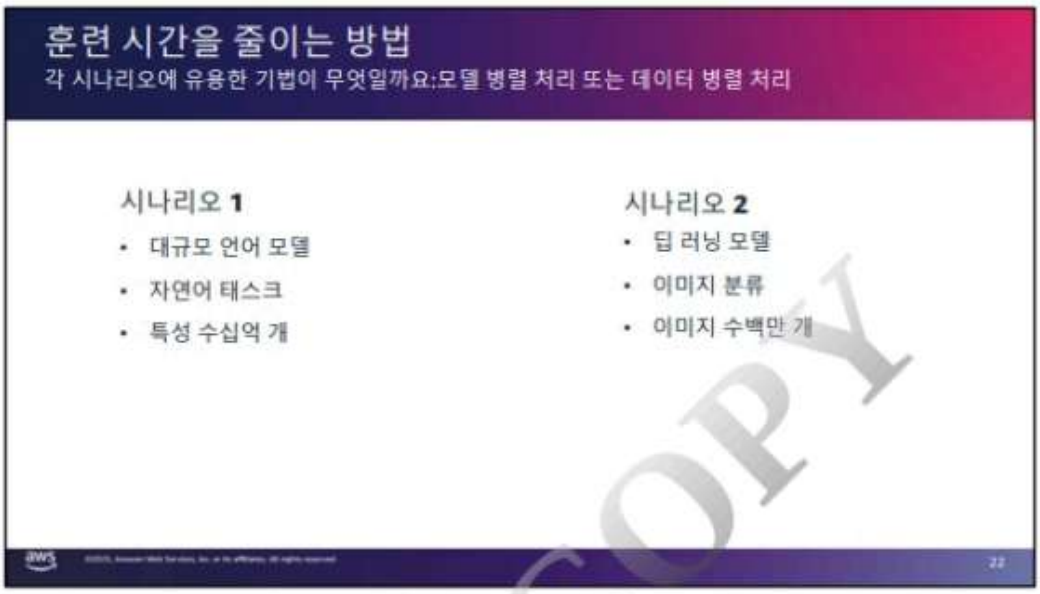
© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



DO NOT COPY
ssleeeee1@gmail.com

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

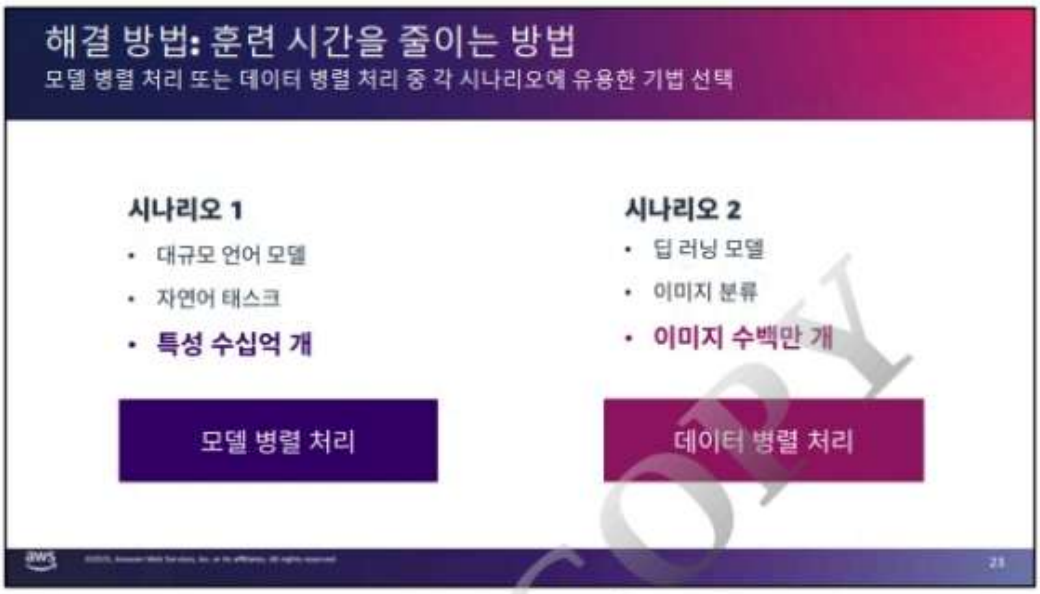


각 시나리오에서 모델 병렬 처리 또는 데이터 병렬 처리를 통해 훈련 시간을 줄일 수 있는 방법을 생각해 보십시오.

- 시나리오 1: 기계 학습(ML) 회사가 자연어 처리(NLP) 태스크를 위한 대규모 언어 모델(LLM)을 개발하고 있습니다. 모델에는 수십억 개의 가중치(파라미터)가 있으며 메모리 제약으로 인해 단일 GPU에서 모델을 훈련할 수 없습니다. 이 회사는 Amazon SageMaker AI를 사용하여 모델을 효율적으로 훈련하려고 합니다.
- 시나리오 2: 한 회사에서 이미지 수백만 개로 구성된 대규모 데이터세트를 대상으로 이미지 분류를 위한 딥 러닝 모델을 훈련하고 있습니다. 훈련 프로세스가 너무 오래 걸리고 회사는 모델 성능을 저하시키지 않으면서 훈련 시간을 줄이려고 합니다. SageMaker AI에서 GPU 인스턴스 여러 개에 액세스할 수 있습니다.

DO NOT COPY
ssleeeee1@gmail.com

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



시나리오 1에서는 수백만 개 파라미터의 가중치가 수렴되는 게 포함됩니다. 이렇게 큰 모델은 GPU 하나의 메모리에 맞지 않을 수도 있으며 처리 시간도 매우 많이 걸립니다. 모델 병렬 처리 방식을 사용하면 가중치를 여러 컴퓨팅 인스턴스로 분산할 수 있습니다. 이 기법을 사용하는 경우 여러 가중치 하위 집합으로 모델을 병렬 훈련할 수 있으므로 훈련 시간을 줄일 수 있습니다.

시나리오 2에서는 이미지 수백만 개를 훈련 데이터세트로 사용됩니다. 데이터 병렬화는 데이터세트를 여러 컴퓨팅 인스턴스로 분산하는데, 여기서 모델 훈련은 각 하위 집합에서 병렬로 일어납니다. 이 기법을 사용하는 경우 대용량 데이터세트의 모델 가중치에 수렴하는 시간을 줄일 수 있습니다.

DO NOT COPY
ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

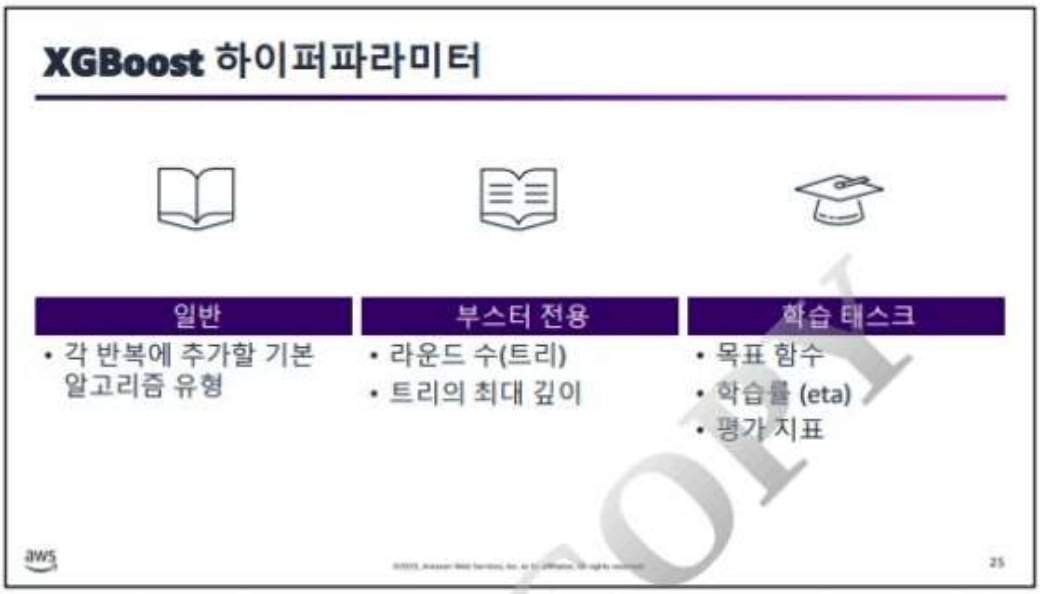
Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

모델을 튜닝하면 특정 알고리즘과 다양한 하이퍼파라미터 값에 대해 가장 성능이 좋은 모델을 선택할 수 있습니다. 하이퍼파라미터는 훈련 프로세스를 조정하는 데 사용되는 알고리즘별 노브로 간주할 수 있습니다. 동일한 훈련 데이터를 사용하면서 알고리즘의 하이퍼파라미터에 대해 다른 값을 적용하면 다음 모델 변형이 구현됩니다.

ML 모델을 최적화하려면 최적의 하이퍼파라미터 세트를 찾아야 합니다. 하이퍼파라미터를 튜닝하면 편향과 분산의 균형을 적절하게 조정하고 과대적합을 방지할 수 있으며 시간과 비용을 제어할 수 있습니다. 이 섹션에서는 몇 가지 하이퍼파라미터 튜닝 전략을 제공하며, 각 전략에는 고유한 강점과 약점이 있습니다.

DO NOT COPY
ssleeeee1@gmail.com

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



XGBoost 하이퍼파라미터는 3개 주요 카테고리로 구성할 수 있습니다: 일반 하이퍼파라미터, 부스터 전용하이퍼파라미터 및 학습 태스크 하이퍼파라미터 입니다. 각 카테고리의 하이퍼파라미터는 특정 사용 사례에서 모델 성능을 최적화하는 과정에서 꼭 필요합니다.

일반 하이퍼파라미터에 따라 각 반복에 추가할 기본 알고리즘 유형이 결정되어 모델이 훈련 태스크에 접근하는 기초를 형성합니다. 부스터 전용 하이퍼파라미터에는 모델의 복잡도 및 데이터에서 패턴을 캡처하는 모델의 능력에 직접적인 영향을 미치는 라운드(트리) 수 및 최대 심도와 같은 중요한 설정이 포함됩니다.

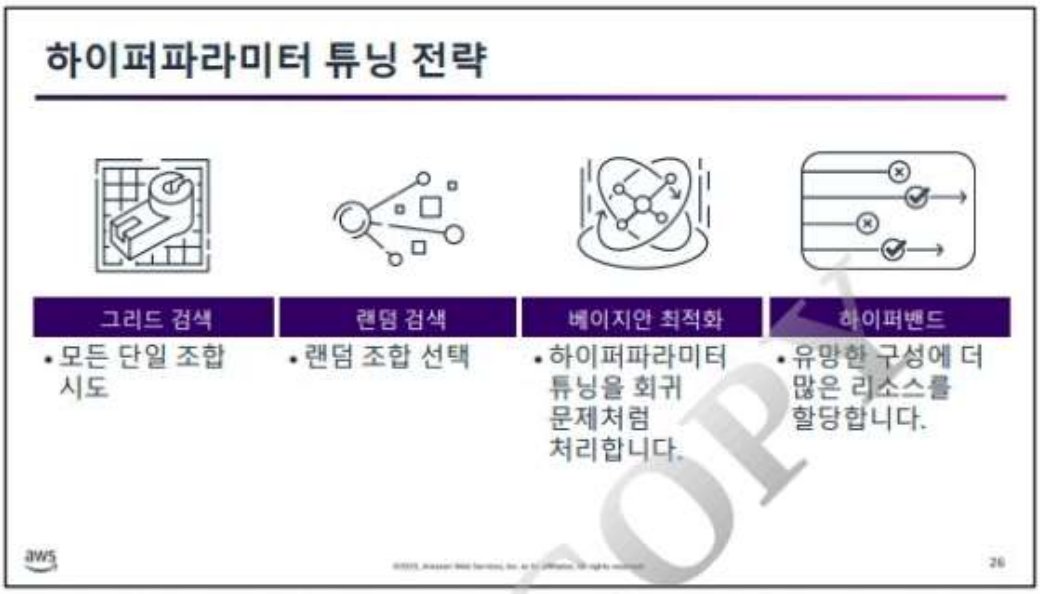
학습 태스크 하이퍼파라미터는 최적화 목표 및 평가 기준에 초점을 맞춥니다. 여기에는 모델 성능을 측정하는 데 사용할 평가 지표 및 적절한 목표 함수 선택이 포함됩니다. 이러한 하이퍼파라미터를 함께 활용하면 XGBoost가 트리 앙상블을 구축 및 최적화하는 방식을 세밀하게 제어할 수 있습니다.

XGBoost와 관련하여 반드시 고려해야 하는 구성 관련 사항은 다음과 같습니다.

1. **일반 하이퍼파라미터:** 모델 아키텍처 전반을 제어합니다. 각 반복에 사용되는 모델의 유형(예: gbtree)을 설정하는 주요 하이퍼파라미터가 부스터입니다. 최종 모델에서는 각 반복 단계의 출력이 통합 제공됩니다.
2. **부스터별 하이퍼파라미터:** 선택한 부스터의 작동 방식을 정의합니다. 부스터에는 max_depth(트리 심도), min_child_weight(리프에 필요한 최소 합), alpha(모델 보존성을 제어하는 L1 정규화) 등의 중요 설정이 포함됩니다.
3. **학습 태스크 하이퍼파라미터:** 모델이 학습을 하고 성능을 평가하는 방식을 결정합니다. 구체적으로는 다음과 같은 항목을 결정합니다.
   - 목표: 최소화할 손실 함수(예: binary:logistic, binary:hinge)를 정의합니다.
   - 학습률:
   - EVAL_Metric: 모델 성능을 평가할 방법을 지정합니다(예: AUC-ROC).

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



하이퍼파라미터 튜닝은 매우 노동 집약적인 작업일 수 있습니다. 기존에는 이러한 작업을 수동으로 수행했습니다. 해당 하이퍼파라미터 및 사용 사례와 관련된 도메인 경험이 있는 사람이 직관과 경험에 따라 하이퍼파라미터를 수동으로 선택합니다. 그런 다음, 모델을 훈련하고 검증 데이터에 점수를 매깁니다. 이 프로세스는 만족스러운 결과가 나올 때까지 반복됩니다.

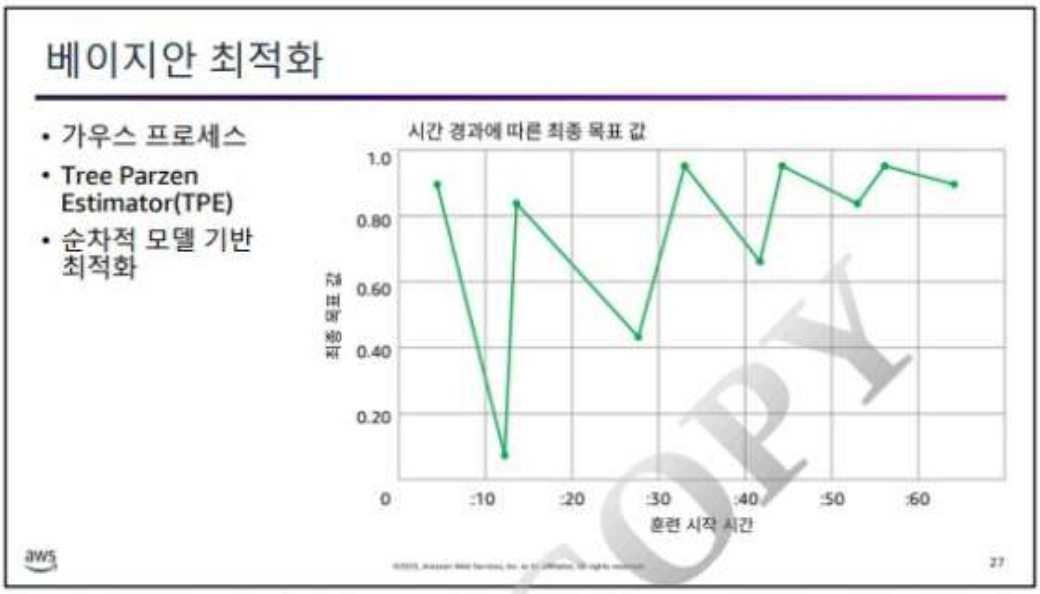
하이퍼파라미터 튜닝 전략은 이 수동 프로세스의 효율성을 높일 수 있습니다.

- **그리드 검색**은 체계적이며 포괄적인 하이퍼파라미터 튜닝 방식입니다. 여기에는 가능한 모든 하이퍼파라미터 값을 정의하고, 이러한 값의 모든 조합에 대해 모델을 훈련하고 평가하는 작업이 포함됩니다. 이 방식은 다른 기법에 비해 컴퓨팅 비용이 많이 듭니다.
- **랜덤 검색**은 그리드 검색보다 더 효율적인 대안입니다. 랜덤 검색은 가능한 모든 하이퍼파라미터 조합을 평가하는 대신 하이퍼파라미터 값 세트를 무작위로 선택합니다. 이로 인해 최적 하이퍼파라미터 세트가 선택되지 않을 수도 있습니다.
- **베이지안 최적화**는 이전 하이퍼파라미터 선택의 성능을 사용하여 후속 값 중 가장 좋은 결과를 생성할 수 있는 값을 예측합니다. 이 방식은 수렴 속도는 빠르지만 복잡하며 규모 조정이 어렵습니다.
- **하이퍼밴드**는 리소스를 성능이 좋은 구성에 동적으로 할당하고, 성능이 낮은 구성을 조기에 중지하는 하이퍼파라미터 튜닝 기법입니다. 하이퍼밴드는 신경망과 같은 반복적 알고리즘에 유용합니다. 이러한 알고리즘에서는 각 에포크 후에 중간 결과를 사용하여 다음 튜닝 프로세스를 가이드할 수 있습니다.

SageMaker AI에서는 이러한 하이퍼파라미터 튜닝 기법이 모두 지원됩니다.

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



베이지안 최적화는 고가의 블랙 박스 함수를 효율적으로 최적화할 수 있는 기법입니다. 이 기법에서는 확률 모델(대개 가우스 프로세스)을 사용하여 미지의 목적 함수를 모델링하고 최적 파라미터를 찾도록 안내합니다. 각 반복에서는 '탐색'과 '활용'을 적절하게 진행하는 획득 함수를 기준으로 평가할 새 포인트를 선택합니다.

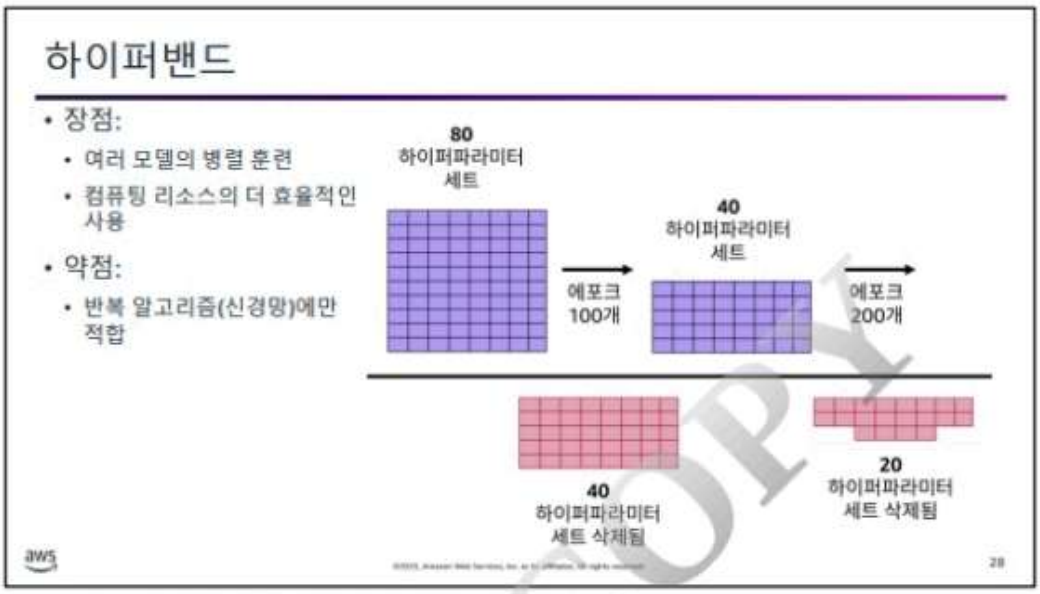
기계 학습에서는 하이퍼파라미터 튜닝에 베이지안 최적화를 사용하는 경우가 많습니다. 특히 평가에서 계산 비용이 많이 발생할 때는 베이지안 최적화를 사용하면 그리드 검색이나 랜덤 검색에 비해 더욱 효율적으로 적절한 하이퍼파라미터 구성을 찾을 수 있습니다. Tree Parzen Estimator(TPE) 및 순차적 모델 기반 최적화 방식은 특정 최적화 시나리오에서는 가우스 프로세스 대신 사용할 수 있는 대안적 접근법입니다.

Amazon SageMaker AI의 자동 모델 튜닝 기능에는 베이지안 최적화가 사용됩니다. 사용자가 하이퍼파라미터 범위와 목표 지표를 지정합니다. 그러면 SageMaker AI가 여러 훈련 작업을 실행하고 베이지안 최적화를 사용하여 각 반복에서 사용해 볼 하이퍼파라미터 값을 선택합니다. 따라서 최적 모델 구성을 찾는 프로세스가 자동으로 진행되므로 수동 튜닝에 비해 시간과 계산 리소스가 절약됩니다.

DO NOT COPY
ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



하이퍼밴드는 탐색 단계와 활용 단계의 2단계로 작동합니다.

- 탐색 단계: 탐색 단계에서는 고정된 수의 에포크를 사용하여 다양한 하이퍼파라미터 구성 내에서 모델을 훈련시킵니다. 다음 조건 중 하나가 충족되면 탐색이 중지됩니다.
  - 성능 개선이 관찰됨. 이 시점에서 탐색을 중지하면 과대적합을 방지할 수 있습니다.
  - 지정된 에포크 수에 도달함.
- 활용 단계: 활용 단계에서는 검증 지표를 사용하여 성능이 가장 낮은 구성을 식별합니다. 초기 구성의 절반은 삭제되고 나머지 구성에는 다음 라운드를 위한 추가 에포크가 할당됩니다. 하이퍼밴드 알고리즘은 연속적인 반감기 실험을 계속 시작합니다. 이 프로세스는 최종 하이퍼파라미터 구성 세트를 선택할 때까지 계속됩니다.

강점

하이퍼밴드는 여러 모델을 동시에 훈련할 수 있으므로 분산 컴퓨팅 환경에 적합합니다. 이렇게 하면 수렴 시간을 개선할 수 있습니다. 또한 하이퍼밴드는 리소스를 더 높은 성능의 하이퍼파라미터 구성에 집중합니다. 이렇게 하면 컴퓨팅 리소스를 그리드 검색이나 랜덤 검색보다 더 효율적으로 할당할 수 있습니다.

약점

하이퍼밴드는 신경망과 같은 반복적 알고리즘에만 사용해야 합니다. 이러한 알고리즘에서는 각 에포크 후에 중간 결과를 사용하여 다음 튜닝 프로세스를 안내할 수 있습니다.

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

SageMaker AI Automatic Model Tuning(AMT) 사용

- 튜닝 전략
- 하이퍼파라미터 범위
- 지표
- 완료 기준

29

SageMaker AI에는 자동 하이퍼파라미터 튜닝 최적화에 사용할 수 있는 Automatic Model Tuning(AMT) 기능이 포함되어 있습니다. 이 기능을 사용하면 모델 정확도를 실시간으로 높일 수 있습니다.

베이지안, 랜덤, 하이퍼밴드, 그리드 중 어떤 하이퍼파라미터 튜닝 전략을 사용할지 지정해야 합니다.

하이퍼파라미터의 정적 값을 지정할 수도 있고 동적 하이퍼파라미터 범위를 정의할 수도 있습니다. 범주형, 정수 또는 연속 값의 지정된 범위 내에서 하이퍼파라미터를 30개까지 지정할 수 있습니다. 특정 조건 충족 시 완료 기준을 설정하여 AMT 튜닝 작업을 중단할 수 있습니다. 이러한 기준의 예로는 최대 훈련 작업 수, 최대 런타임(초), 대상 목표 지표의 구체적인 값 등이 있습니다.

SageMaker AI AMT는 최적 하이퍼파라미터 값을 찾기 위해 데이터세트에서 많은 훈련 작업을 실행합니다. AMT는 ML을 사용하여 사용자가 지정한 범위와 지표에서 최적 하이퍼파라미터를 확인합니다. 자동 튜닝을 사용해 하이퍼파라미터 범위, 리소스 또는 목표 지표를 검색하지 않고 최적 값을 찾을 수도 있습니다.

DO NOT COPY
ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

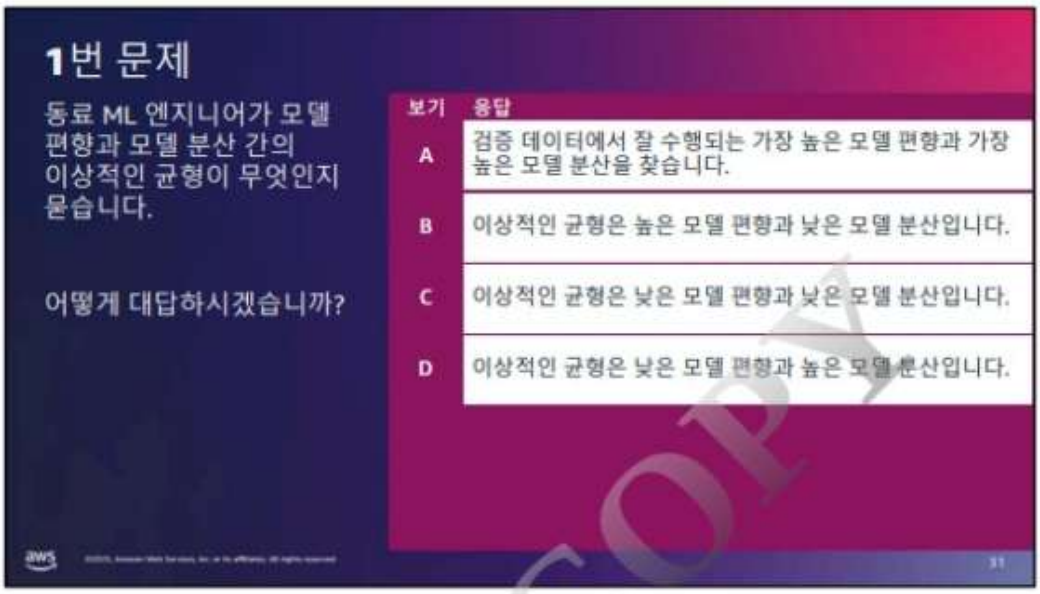
이제 간단한 지식 확인을 진행하여 이 모듈에서 알아본 모델 배포 전략 관련 주요 개념을 복습해 보겠습니다. 다음과 같은 주제와 관련된 문제를 풀어 볼 예정입니다.

- 모델 배포의 정의
- 배포 인프라 관련 고려 사항
- 배포 대상 선택
- 블루/그린 배포, 롤링 배포 등의 배포 전략
- 블루/그린 배포용 트래픽 이동 모드

이 지식 확인을 통해 다음 모듈로 넘어가기 전에 수업 내용을 더욱 정확하게 이해하고 헷갈리는 부분을 확실하게 파악할 수 있습니다.

DO NOT COPY
ssleeeee1@gmail.com

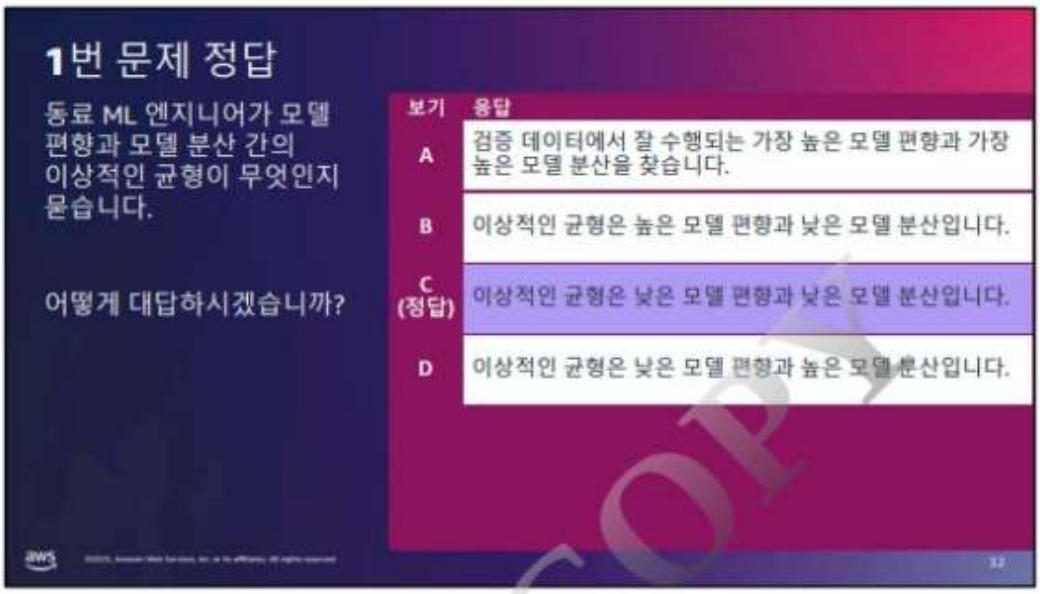
Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



DO NOT COPY
ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



정답: C

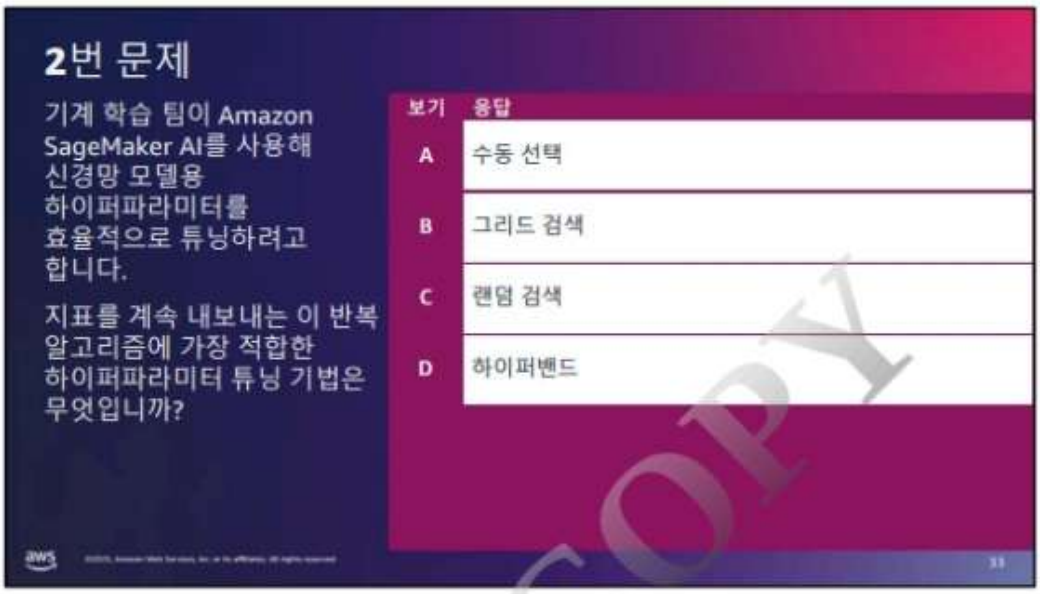
설명: 정답은 C입니다. 모델 편향이 낮고 모델 분산이 낮으면 일반화 성능이 우수하고 극단적인 예측을 방지할 수 있습니다. 이 균형은 모델에서 데이터의 실제 기본 패턴을 학습했음을 의미합니다.

오답:

- A: 이 조합은 모델에서 데이터의 기본 관계를 학습하지 않았지만 데이터의 노이즈는 학습했을 수 있음을 의미합니다.
- B: 높은 모델 편향과 낮은 모델 분산은 모델에서 지속적으로 잘못 가정하고 데이터의 주요 관계를 학습하지 않았음을 나타냅니다.
- D: 모델 편향이 낮고 모델 분산이 높으면 모델이 훈련 데이터에 매우 잘 맞지만 잘 일반화되지는 않습니다.

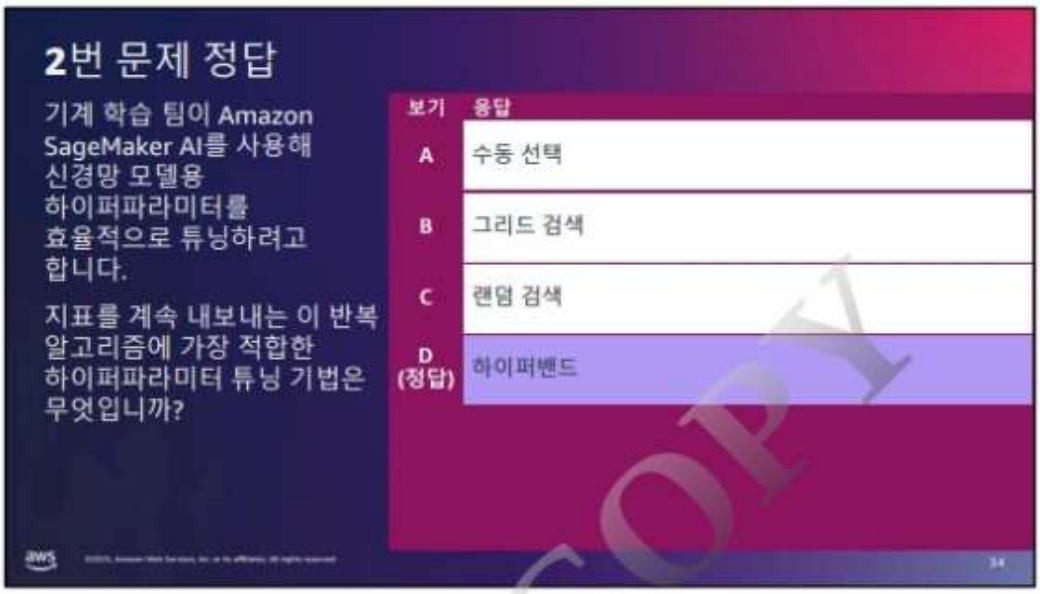
DO NOT COPY ssleeeee1@gmail.com

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



DO NOT COPY
ssleeeee1@gmail.com

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



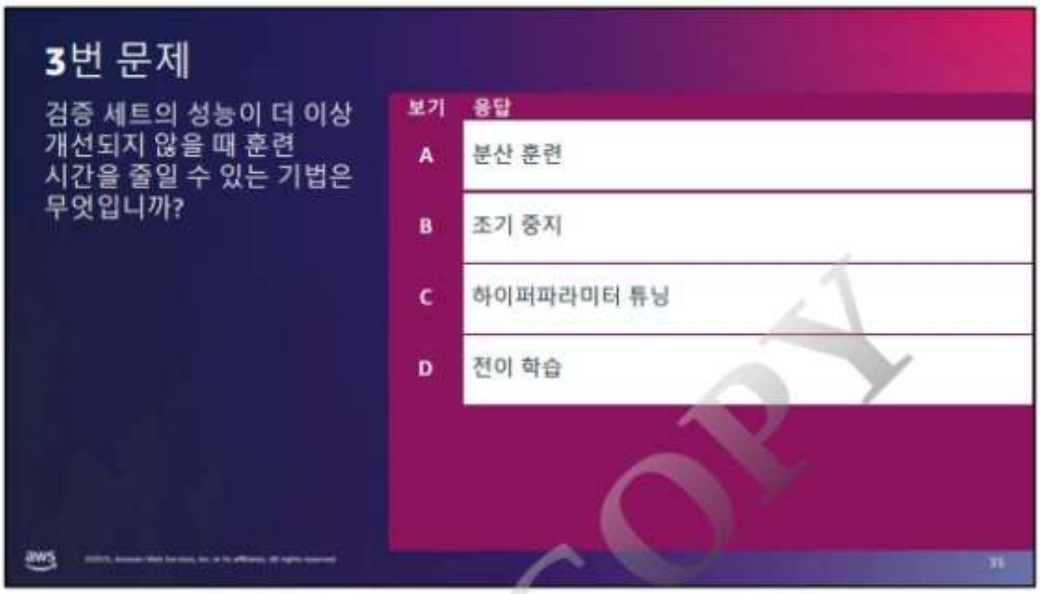
정답: D

설명: 정답은 D, 하이퍼밴드입니다. 하이퍼밴드는 신경망과 같은 반복적 알고리즘에 특히 유용합니다. 각 에포크 후에 중간 결과를 사용하여 다음 튜닝 프로세스를 진행할 수 있기 때문입니다.

오답:
- A: 포괄적으로 진행되며 계산 비용이 많이 드는 그리드 검색은 신경망에서는 효율성이 낮습니다.
- B: 랜덤 검색은 그리드 검색보다는 효율적이지만 하이퍼밴드처럼 중간 결과를 활용하지는 않습니다.
- C: 수동 선택은 시간이 많이 걸리며 최적 하이퍼파라미터를 효율적으로 찾지 못할 수도 있습니다.

DO NOT COPY ssleeeee1@gmail.com

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

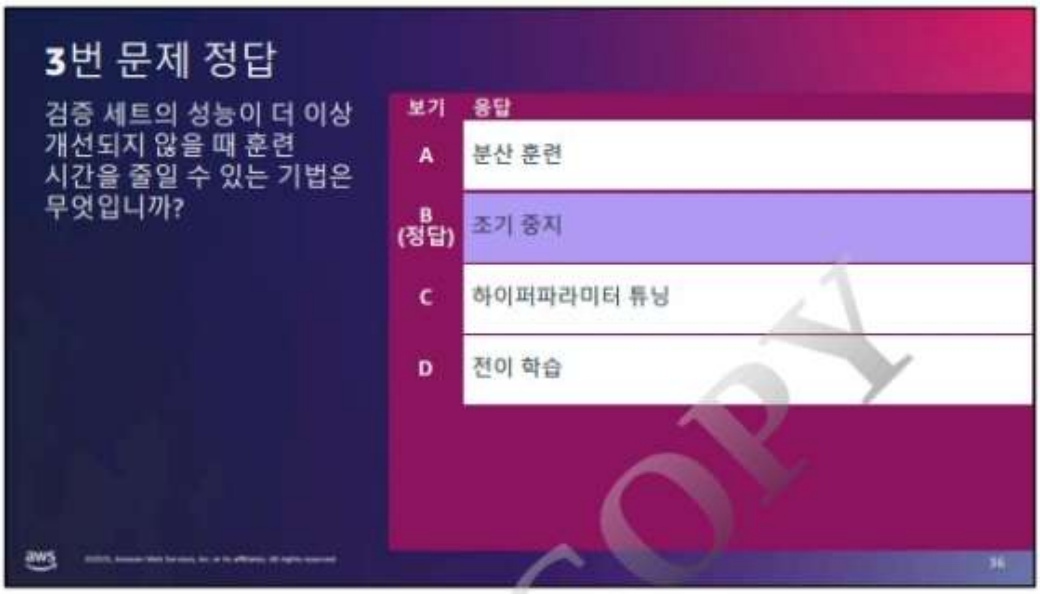


훈련 시간 단축 기법인 조기 중지에 대해 이해하고 있는지를 평가하기 위한 문제입니다.

DO NOT COPY
ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



정답: B

설명: 정답은 B, 조기 중지입니다. 조기 중지는 훈련 중에 검증 세트에서 예측을 하는 모델의 성능을 모니터링하면서 성능이 더 이상 개선되지 않으면 프로세스를 중지하는 기법입니다. 이 기법을 활용하면 훈련 시간을 줄이고 과대적합을 방지할 수 있습니다.

오답:
- A: 분산 훈련을 진행하는 경우 훈련이 여러 시스템에서 병렬 처리되므로 시간은 단축되지만 훈련이 조기 중지되지는 않습니다.
- C: 하이퍼파라미터 튜닝에서는 모델 파라미터가 최적화될 뿐 튜닝 자체에서 훈련 시간이 단축되지는 않습니다.
- D: 전이 학습에서는 사전 훈련된 모델을 활용하지만 훈련 시간이 그다지 단축되지는 않습니다.

DO NOT COPY ssleeeee1@gmail.com

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

SageMaker AI의 자동 모델 튜닝 기능 관련 지식을 평가하기 위한 문제입니다.

DO NOT COPY
ssleeeee1@gmail.com

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

정답: C

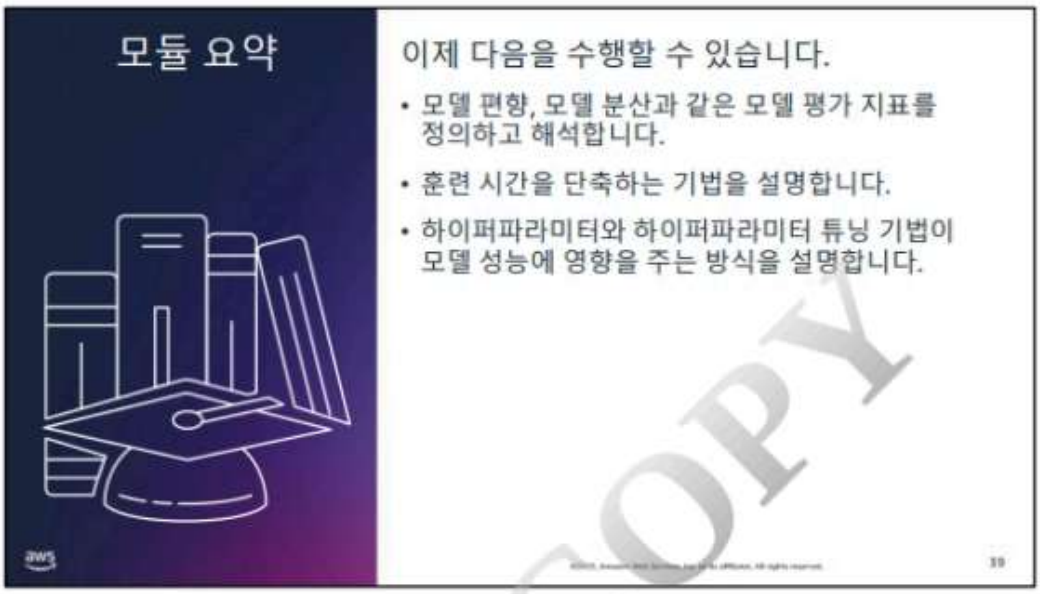
설명: 정답은 D, SageMaker Automatic Model Tuning입니다. SageMaker Automatic Model Tuning은 베이지안 최적화, 하이퍼밴드 등의 기법을 사용하여 모델에 가장 적합한 하이퍼파라미터 구성을 자동으로 검색하는 기능입니다.

오답:
- A: SageMaker Pipelines는 하이퍼파라미터 튜닝이 아닌 ML 워크플로 구축용 도구입니다.
- B: SageMaker Model Monitor는 하이퍼파라미터 튜닝이 아닌 배포된 모델 모니터링용 도구입니다.
- D: SageMaker Feature Store는 모델 튜닝이 아닌 특성 저장 및 관리용 도구입니다.

DO NOT COPY ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

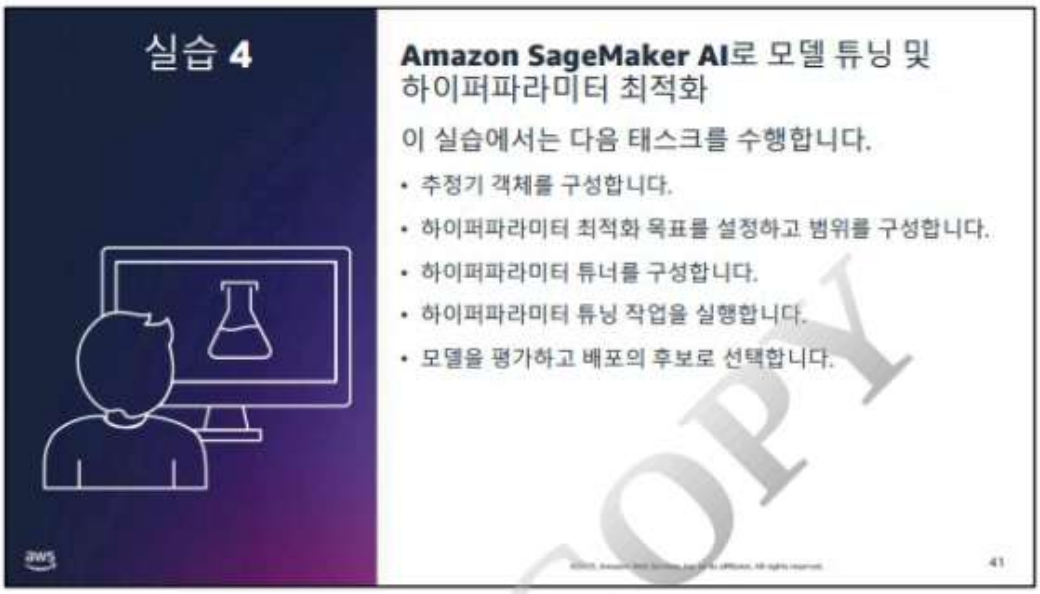


DO NOT COPY
ssleeeee1@gmail.com

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

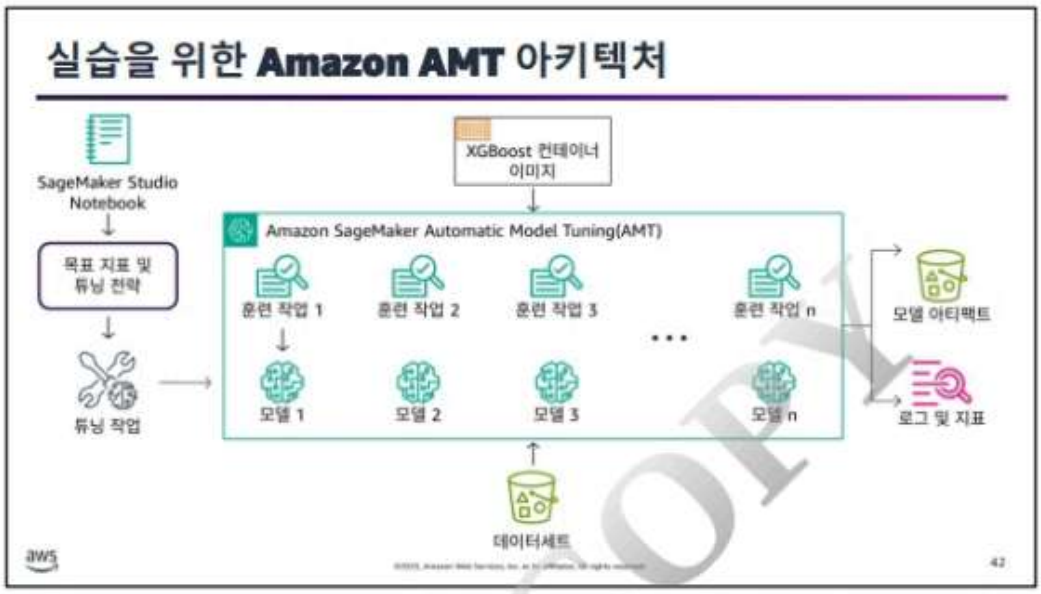
DO NOT COPY
ssleeeee1@gmail.com

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



DO NOT COPY
ssleeeee1@gmail.com

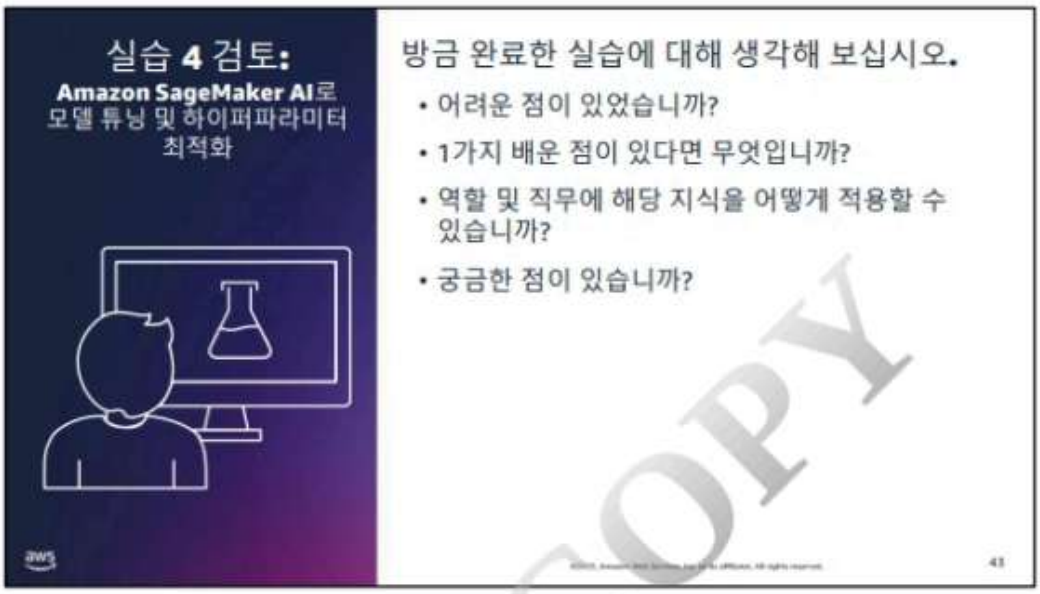
Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



실습 연습에서는 SageMaker Studio Notebook에서 튜닝 작업을 정의합니다. 구체적으로는 목표 지표와 튜닝 전략, 그리고 Amazon Simple Storage Service(Amazon S3)의 훈련 데이터세트 위치를 정의하게 됩니다. 실행 시 튜닝 작업은 XGBoost 컨테이너 이미지를 사용하여 많은 수의 훈련 작업을 생성하고 실행합니다. 각 훈련 작업에서는 모델 아티팩트가 Amazon S3에 출력되며, 튜닝 작업 및 모든 훈련 작업의 로그 및 지표는 Amazon CloudWatch에 캡처됩니다.

DO NOT COPY
ssleeeee1@gmail.com

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



DO NOT COPY
ssleeeee1@gmail.com

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

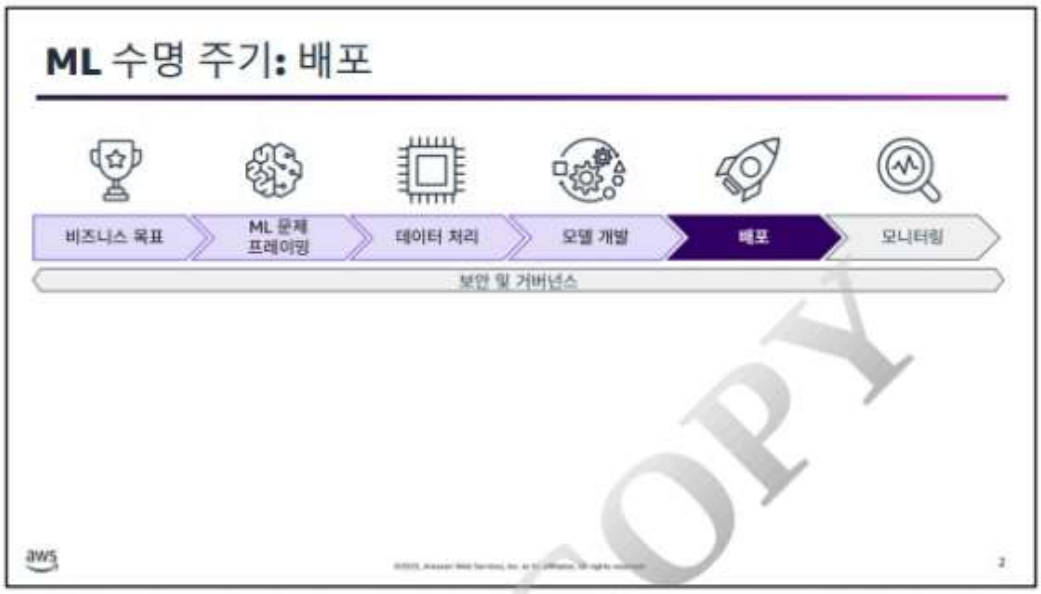
모델 배포는 기계 학습(ML) 수명 주기에서 학습된 모델을 프로덕션 환경에 통합하여 추론을 통해 새 데이터에 대한 예측을 생성하는 단계입니다. 이 모듈에서는 주요 고려 사항, 배포 목표, 모범 사례를 포함하여 모델 배포 전략을 포괄적으로 살펴봅니다.

이 모듈에서는 성능 요구 사항, 필요한 확장성 수준 및 운영 오버헤드 등의 요인을 고려하여 적절한 배포 인프라를 선택하는 방법도 알아봅니다. 그리고 Amazon SageMaker AI, Amazon Elastic Kubernetes Service(Amazon EKS), Amazon Elastic Container Service(Amazon ECS), AWS Lambda 등 효율적인 모델 배포를 원활하게 진행하는 데 사용할 수 있는 여러 AWS 서비스와 도구도 살펴봅니다.

이 모듈에서는 블루/그린 배포, 카나리 릴리스, 롤링 업데이트와 같은 배포 전략 관련 인사이트도 파악할 수 있습니다. 이러한 전략은 새 모델 버전의 통제되고 위험 완화 출시를 위한 메커니즘을 제공하고 프로덕션 시스템의 중단을 최소화하는 데 도움이 됩니다. 이 모듈에서는 배포 워크플로 최적화 시에 적용할 수 있는 오토 스케일링, 모델 변형, 배치 변환 등의 필수 개념에 대해서도 알아봅니다.

DO NOT COPY
ssleeeee1@gmail.com

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



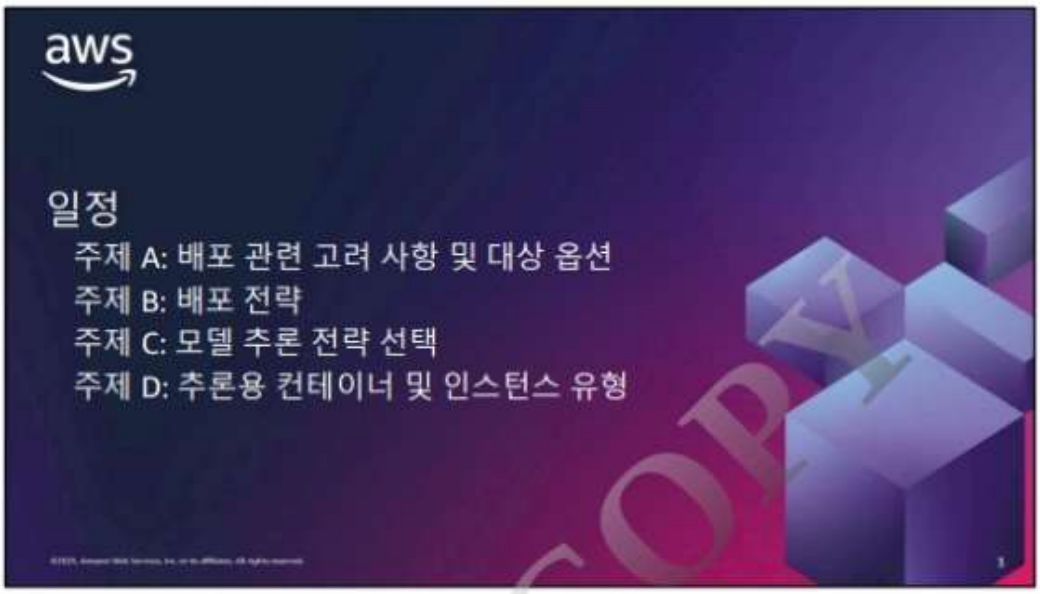
ML 수명 주기에서는 여러 단계가 진행되며 각 단계는 ML 솔루션 개발 및 운영 과정에서 중요한 역할을 합니다. 배포 단계 이전에 완료한 모델 개발 단계에서는 ML 모델 훈련, 튜닝, 테스트를 수행했습니다.

모델 배포는 ML 수명 주기에서 개발 단계의 다음 단계로, 새 데이터에서 예측을 생성할 수 있도록 프로덕션 환경에 훈련된 모델을 통합합니다. 이 프로세스에서는 모델의 예측 기능이 필요한 사용자, 애플리케이션 또는 시스템에 모델을 제공하게 됩니다.

ML 모델을 프로덕션 환경에 배포하는 것은 기계 학습 엔지니어의 주요 업무입니다. 구체적인 사용 사례와 비교하여 컴퓨팅 환경의 성능과 비용 간 균형을 이해하는 것이 중요합니다. 이 모듈은 효율적이고 효과적인 모델 배포를 위해 배포 인프라에 대해 정보를 토대로 결정하는 데 도움이 됩니다.

DO NOT COPY
ssleeeee1@gmail.com

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



이 모듈에서 다루는 4가지 주요 주제는 다음과 같습니다.

주제 A: 배포 관련 고려 사항 및 대상 옵션
주제 B: 배포 전략
주제 C: 모델 추론 전략 선택
주제 D: 추론용 컨테이너 및 인스턴스 유형

DO NOT COPY
ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

이 주제에서는 기계 학습 모델 배포 시의 주요 고려 사항 및 배포에 사용할 수 있는 대상 옵션을 살펴봅니다. 모델을 효과적으로 배포하려면 성능 요구 사항, 필요한 확장성, 운영 오버헤드와 같은 여러 요인을 면밀하게 계획 및 고려해야 합니다.

이 주제에서는 모델 배포와 연관된 다음과 같은 당면 과제에 대해 알아봅니다.

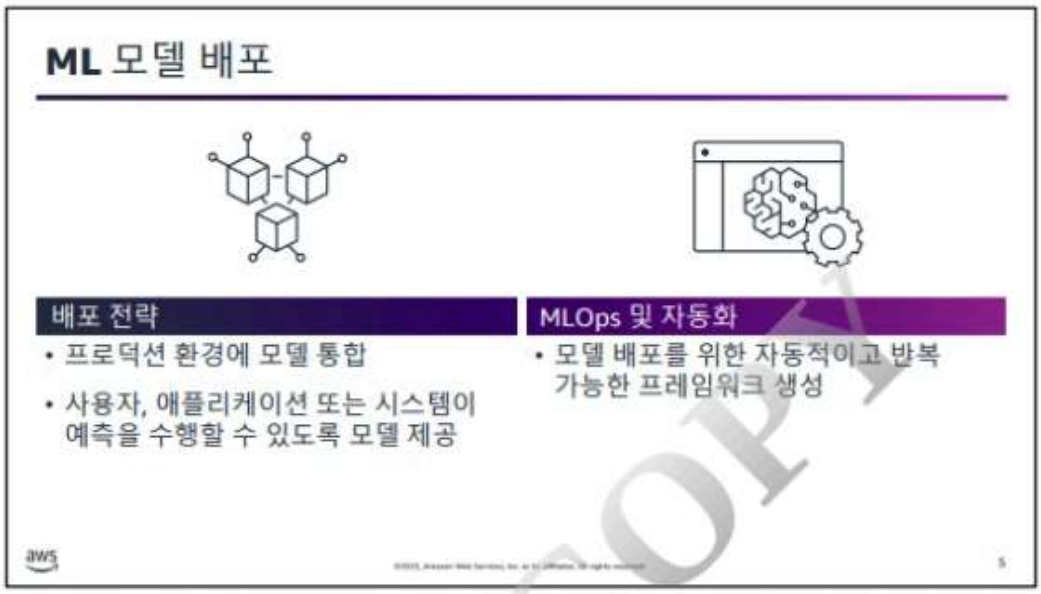
- 데이터 품질 및 사전 처리
- 모델 모니터링 및 유지 보수
- 확장성 및 성능
- 인프라 및 DevOps
- 지속적 전달 및 업데이트

이 주제에서는 SageMaker AI 엔드포인트, Amazon EKS, Amazon ECS, Lambda 등 AWS에서 사용 가능한 여러 배포 대상 관련 인사이트도 파악할 수 있습니다. 이러한 각 대상의 이점과 고려 사항은 각기 다릅니다. 이 주제에서는 구체적인 사용 사례에 따라 가장 적절한 옵션을 선택하는 방법을 알아봅니다.

DO NOT COPY ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



모델 배포에서는 훈련된 기계 학습 모델 및 관련 리소스를 프로덕션 환경에 통합하여 새 데이터에서 예측을 생성할 수 있습니다. 다시 말해서, 모델 배포는 최종 사용자, 애플리케이션 또는 시스템에서 ML 모델을 사용할 수 있도록 하는 프로세스입니다.

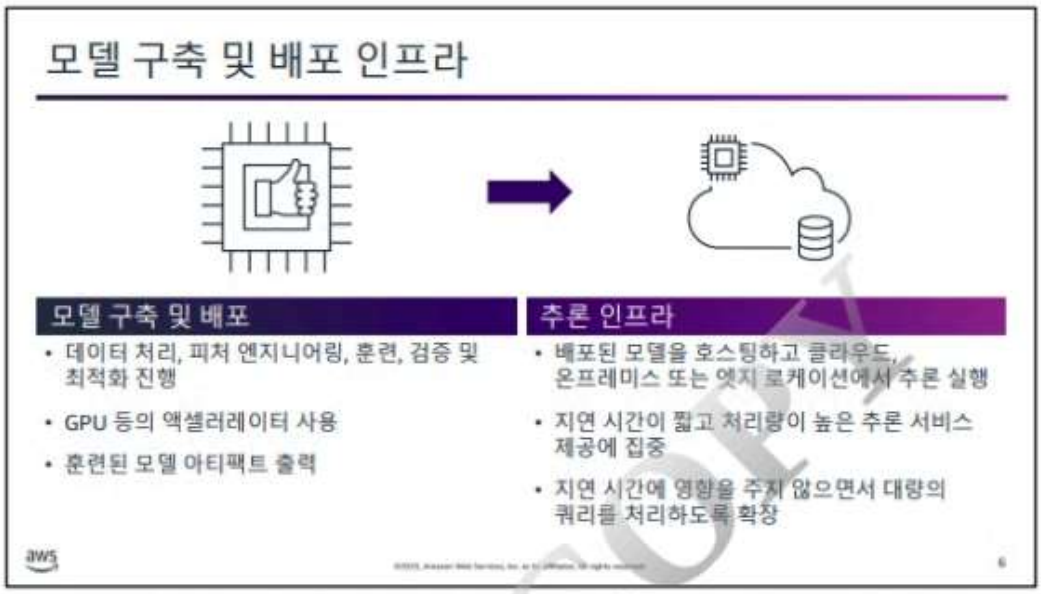
모델 배포의 주요 목표는 ML 모델 배포용으로 자동화되고 반복 가능한 프레임워크를 생성하는 것입니다. 이 프레임워크에는 입력 데이터세트에서 추론 요청 처리에 이르기까지 전체 파이프라인이 포함되어 있어야 합니다. 반복 가능한 배포 과정을 자동화하면 프로덕션 환경에서 모델 성능을 항상 안정적으로 유지할 수 있습니다.

효율적인 모델 배포 인프라의 목표는 ML 솔루션용 엔드 투 엔드 파이프라인을 구축하는 것입니다. 이 파이프라인에서는 모델 배포와 스케일링 및 배포된 모델 관리 프로세스가 자동으로 진행되어야 합니다. 그래야 모델의 예측 기능이 필요한 사용자 애플리케이션 또는 시스템과 모델을 원활하게 통합할 수 있기 때문입니다.

DO NOT COPY
ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



배포는 기계 학습 프로세스의 다양한 파트에 적용할 수 있는 광범위한 용어입니다. 마찬가지로, **인프라**는 ML 수명 주기의 다양한 구성 요소를 설명하는 데 사용되는 용어이므로 혼동을 일으킬 수 있습니다. 모델 배포 인프라를 선택하고 오케스트레이션할 때에는 여러 계층을 고려해야 합니다.

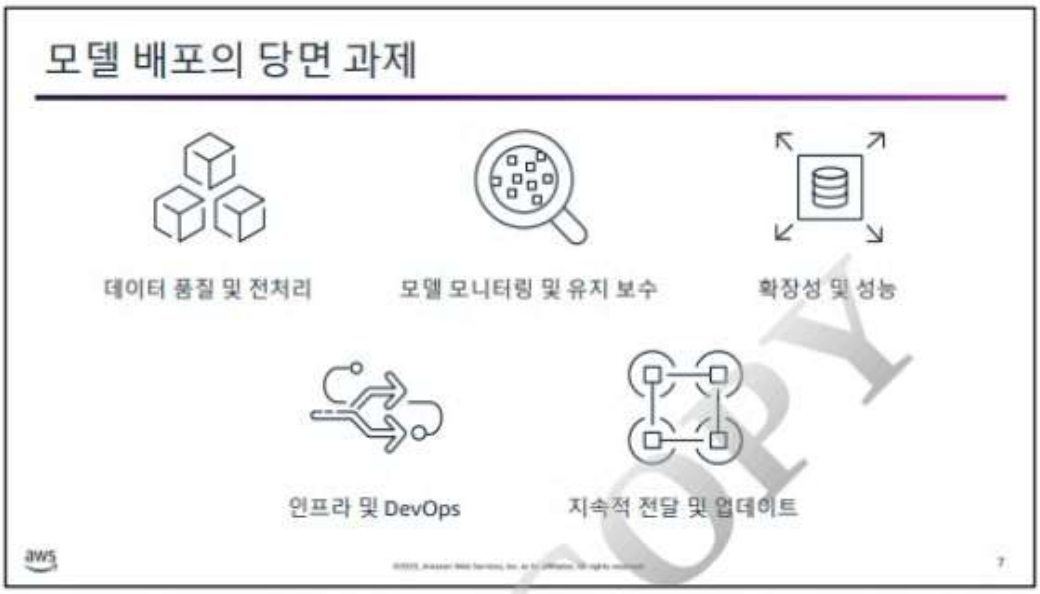
모델 구축 및 배포 인프라:
- 기계 학습 모델 훈련 및 최적화 처리
- 데이터 처리, 피처 엔지니어링, 모델 훈련, 검증 및 최적화 수행
- 복잡한 모델을 빠르게 훈련할 수 있는 GPU 등의 컴퓨팅 리소스 필요
- 추론에 사용할 최종 훈련된 모델 아티팩트 출력

추론 인프라:
- 훈련 후 배포된 모델을 호스팅하고 실행 중인 추론 처리
- 새로운 처음 보는 데이터를 수신하여 모델에서 실행한 후 예측 반환
- 지연 시간이 짧고 처리량이 높은 추론 서비스 제공에 집중
- 지연 시간에 영향을 주지 않으면서 대량의 추론 요청을 처리하도록 규모 조정 가능
- 클라우드, 온프레미스 또는 엣지 로케이션에 호스팅 가능

모델 구축과 추론 인프라를 분리하면 각 구성 요소를 특정 요구 사항에 따라 최적화함으로써 모델 배포 워크플로를 더욱 효율적이고 효과적으로 만들 수 있습니다.

DO NOT COPY
ssleeeee1@gmail.com

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



기계 학습 모델을 효율적으로 배포하려면 모델이 실행될 환경 및 모델이 특정 시스템과 통합되는 방식을 신중하게 계획하고 고려해야 합니다. 모델 배포 시 문제가 발생할 수 있지만 잠재적 장애물을 인식하면 효과적인 배포 워크플로를 보다 잘 준비하는 데 도움이 될 수 있습니다.

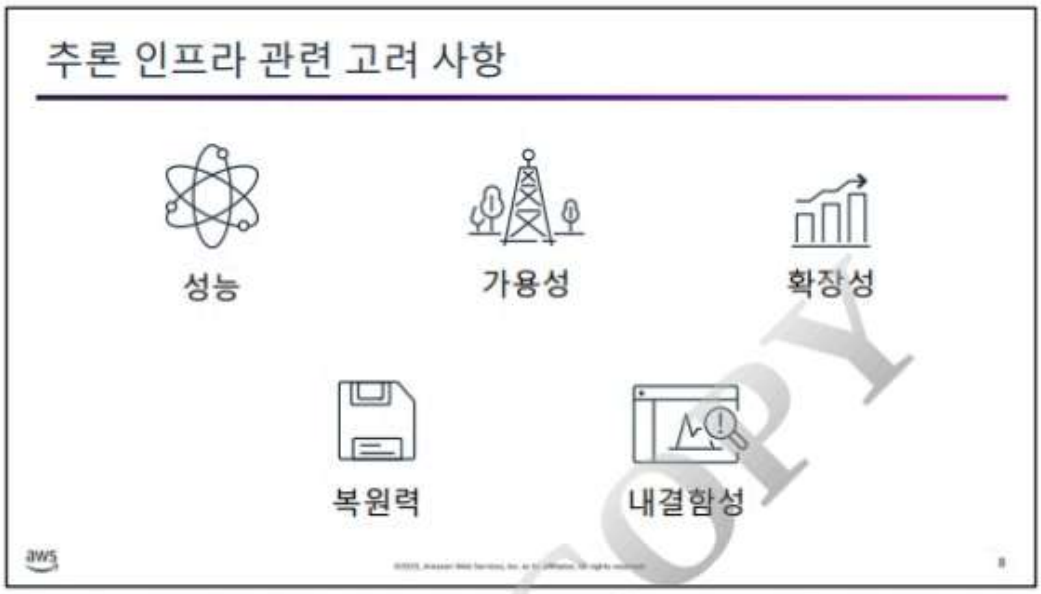
배포 전략을 고려할 때는 배포 시에 발생할 수 있는 다음 당면 과제를 기억하고 있어야 합니다.

- 데이터 품질 및 전처리: 모델 훈련 및 배포에 사용할 깔끔하고 올바른 형식의 고품질 데이터를 확보하기란 쉽지 않을 수도 있습니다. 그러나 이 과정은 ML 프로세스에서 꼭 완료해야 하는 중요한 단계입니다.
- 모델 모니터링 및 유지 보수: 배포된 모델은 시간이 지나면서 데이터 및 개념 변화, 데이터 분산 변화 또는 환경 변화로 인해 성능이 저하될 수 있으므로 지속적인 모니터링과 유지 보수가 필요합니다.
- 확장성 및 성능: 모델을 대규모로 배포하기란 쉽지 않을 수도 있습니다. 짧은 지연 시간과 높은 처리량을 유지하면서 대량의 데이터를 효율적으로 처리해야 하기 때문입니다.
- 인프라 및 DevOps: ML 모델을 기존 시스템 및 워크플로에 통합하고, 종속성을 관리하고, 배포 프로세스를 자동화할 때 기술적 문제가 발생할 수 있습니다.
- 지속적 전달 및 업데이트: 진행 중인 작업을 중단하지 않고 모델 업데이트와 새 버전을 배포하려면 신중하게 계획을 세우고 프로세스를 진행해야 합니다.

배포 중에 발생할 수 있는 이러한 도전 과제를 파악하면 모델 배포와 연관된 위험에 더욱 효율적으로 대비하고 해당 위험이 발생하는 경우 더 쉽게 완화할 수 있습니다. 그러면 프로덕션 환경으로 효율적이며 원활하게 전환할 수 있습니다.

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



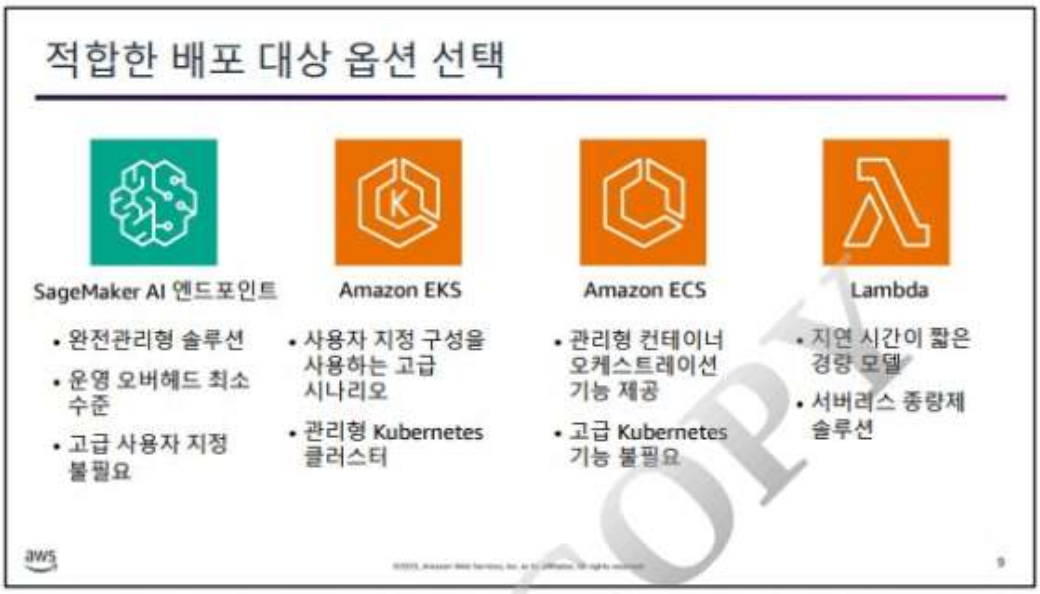
ML 모델을 구축하고 훈련한 후에는 모델을 배포하여 예측 또는 추론을 시작합니다. 추론 인프라를 설계할 때는 여러 가지 요인을 고려해야 합니다. 이러한 요인으로는 성능, 가용성, 확장성, 복원력, 내결함성 등이 있습니다.

- 성능: 시간이 지나도 모델 예측 성능이 저하되지 않아야 합니다. 높은 모델 성능을 유지하려면 효율적인 계산과 추론이 가능하도록 모델 아키텍처를 최적화하고 적절한 Amazon Elastic Compute Cloud(Amazon EC2) 인스턴스 유형과 크기를 사용해야 합니다.
- 가용성: 서비스가 중단되지 않도록 중복성 및 장애 조치 메커니즘을 갖춘 시스템을 설계해야 합니다. 가용성을 높이려면 워크로드를 여러 인스턴스에 분산하도록 로드 밸런싱 및 수평적 스케일링을 구현해야 합니다.
- 확장성: 증가하는 워크로드를 수용할 수 있도록 모듈식 분산 아키텍처로 시스템을 설계해야 합니다. 우수한 확장성을 유지하려면 오토 스케일링 기능을 사용하는 동시에, 구성 요소 분리 및 확장성 개선을 지원하는 비동기 및 이벤트 기반 처리 방식을 구현해야 합니다.
- 복원력: 데이터 손실을 방지하도록 데이터 중복성 메커니즘과 백업 메커니즘을 통합해야 합니다. 복원력을 높이려면 장애 발생 시 전체 운영을 중단하는 것이 아니라 부분적인 장애와 기능 저하를 정상적으로 처리하도록 시스템을 설계해야 합니다. 또한 수동 개입 없이 장애에서 정상 상태를 자동으로 복구할 수 있도록 해야 합니다.
- 내결함성: 문제를 신속하게 식별하고 진단하도록 강력한 오류 처리 메커니즘과 로깅 메커니즘을 구현해야 합니다. 내결함성을 높이려면 수동 작업을 수행하지 않아도 장애에서 정상 상태로 복구되도록 오토 스케일링, 재시작 또는 장애 조치와 같은 자가 복구 기능을 구현해야 합니다.

설계 단계에서 이러한 고려 사항을 해결하면 ML 솔루션의 성능, 가용성, 확장성, 복원력 및 내결함성 요구 사항을 충족하는 안정적이며 신뢰할 수 있는 추론 인프라를 구축할 수 있습니다.

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



ML 모델 배포 대상을 선택할 때는 AWS에서 여러 가지 옵션을 사용할 수 있습니다. 배포 대상 선택은 확장성, 운영 오버헤드, 비용, 모델 복잡성 등 다양한 요인에 따라 달라집니다. 사용 사례별로 다음과 같은 특정 AWS 서비스를 사용해야 하는 상황을 고려할 수 있습니다.

- SageMaker AI 엔드포인트: 운영 오버헤드를 최소화할 수 있으며 고급 인프라 사용자 지정을 진행하지 않아도 되는 완전관리형 솔루션을 사용하려는 경우 SageMaker AI 엔드포인트를 선택합니다. 기본 제공 모니터링 및 로깅 기능이 포함되어 있는 ML 모델을 배포하고 규모를 조정하려는 경우 SageMaker AI 엔드포인트가 편리합니다.
- Amazon EKS: 고급 배포 시나리오와 사용자 지정 구성이 필요하며, Kubernetes 클러스터를 관리할 리소스가 있다면 Amazon EKS를 선택합니다. Amazon EKS는 ML 모델을 컨테이너화된 애플리케이션으로 배포하고 관리하는 데 사용할 수 있는 유용한 컨테이너 서비스를 제공합니다.
- Amazon ECS: 다른 AWS 서비스와 잘 통합되는 관리형 컨테이너 오케스트레이션 서비스를 원하며 고급 Kubernetes 기능은 필요하지 않은 경우 Amaazon ECS를 선택합니다. Amazon ECS를 사용하면 ML 모델용 Docker 컨테이너를 쉽게 배포 및 관리할 수 있습니다.
- Lambda: 지연 시간이 짧은 경량 모델이 있으며 서버리스 종량제 솔루션을 사용하려는 경우 AWS Lambda를 선택합니다. 이벤트 기반 아키텍처에 적합한 Lambda를 사용하면 요청이 없을 때 컨테이너 규모를 0으로 자동 조정하여 비용을 절약할 수 있습니다.

각 배포 대상의 이점과 고려 사항을 파악하면 효율적이고 효과적인 모델 배포 워크플로를 위해 특정 요구 사항에 맞는 정보에 입각한 결정을 내릴 수 있습니다.

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

SageMaker AI 엔드포인트

이점	고려 사항
• 완전관리형 서비스 • 편리한 배포 및 규모 조정 • 기본 제공 모니터링 및 로깅 기능 포함 • 다양한 ML 프레임워크 지원	• 다른 옵션보다 사용자 지정 가능한 범위가 좁음

10

ML 모델 배포 시에 SageMaker AI 엔드포인트를 사용하는 경우 여러 가지 주요 이점이 제공됩니다.

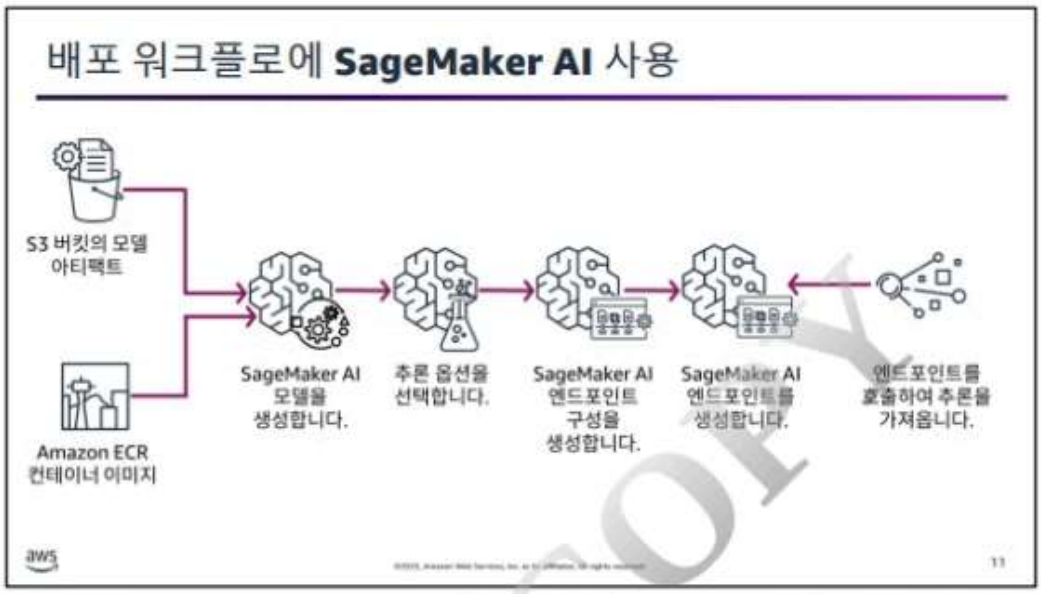
이점:
- 운영 오버헤드를 줄일 수 있는 완전관리형 서비스
- 편리한 모델 배포 및 규모 조정
- 기본 제공 모니터링 및 로깅 기능 포함
- 다양한 ML 프레임워크 기본 지원

고려 사항:
- 일부 기타 옵션보다 사용자 지정 가능한 범위가 좁음

운영 오버헤드가 최소화되는 완전관리형 솔루션을 원하는 경우 SageMaker AI 엔드포인트가 적합합니다. 인프라 관리가 아닌 모델 개발에 주력하려는 팀의 경우 SageMaker AI 엔드포인트가 특히 유용합니다. 하지만 매우 전문적인 배포 방식이 필요하거나 인프라를 세부적으로 제어해야 하는 경우에는 다른 옵션이 더 적절할 수도 있습니다.

DO NOT COPY
ssleeeee1@gmail.com

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



실시간 추론 호스팅에 Amazon SageMaker AI를 사용하는 경우 SageMaker AI 엔드포인트를 생성해야 합니다. 엔드포인트는 API 통신 채널의 대상 끝입니다. 클라이언트 애플리케이션 API는 이 채널에 추론 요청을 보내고 예측과 기타 응답이 이 채널에서 반환됩니다. 인프라 작업의 대부분이 추상화되어 있으므로 인프라 관리가 아닌 애플리케이션에 집중할 수 있습니다.

모델 배포에 SageMaker AI를 사용할 때 수행하는 단계는 다음과 같습니다.

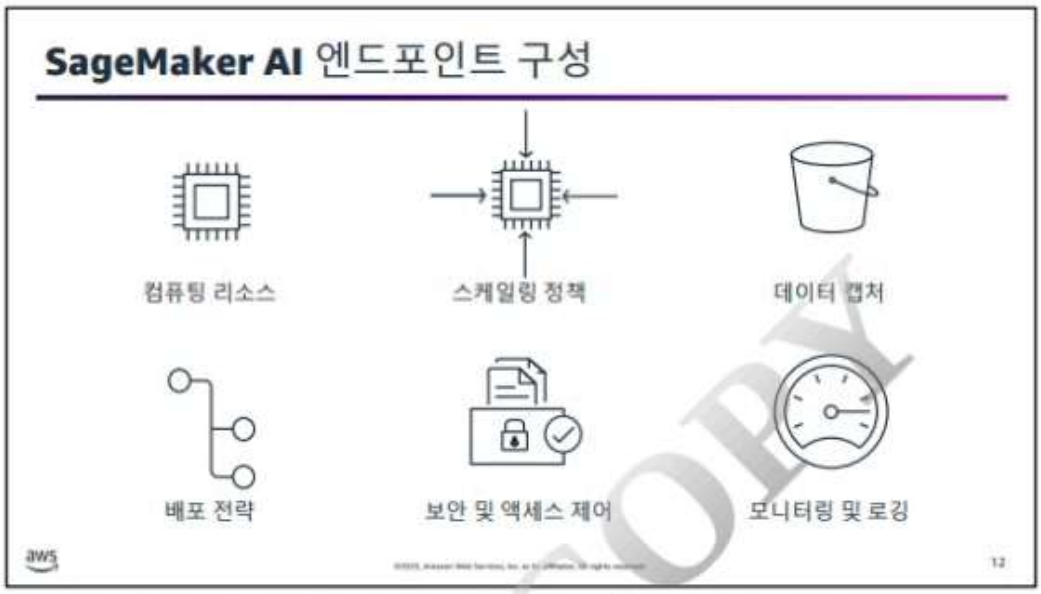
1. SageMaker AI 모델 생성: SageMaker AI에서 모델을 생성합니다. 모델 생성에 사용하려는 S3 버킷에 저장된 모델 아티팩트와 Amazon Elastic Container Registry(Amazon ECR) 컨테이너 이미지를 가리키면 됩니다.
2. 추론 옵션 선택: 요구 사항을 토대로 적합한 추론 옵션을 선택합니다. Inference Recommender를 사용하여 최적 추론 옵션이 포함된 권장 사항을 확인할 수 있습니다.
3. SageMaker AI 엔드포인트 구성 생성 엔드포인트 다음에 필요한 인스턴스 유형과 인스턴스 수를 선택하여 엔드포인트를 구성합니다. 이 구성에 따라 모델 호스팅용 컴퓨팅 리소스가 정의됩니다.
4. SageMaker AI 엔드포인트 생성: 구성된 설정에 따라 SageMaker AI 추론 엔드포인트를 생성합니다.
5. 엔드포인트 호출: 생성된 엔드포인트를 호출하여 추론을 응답으로 수신할 수 있습니다.

Amazon SageMaker AI를 모델 배포에 활용하는 경우 완전관리형 기능과 기본 제공 모니터링 및 로깅 기능을 활용할 수 있으며 타 AWS 서비스와 모델을 원활하게 통합할 수 있습니다. SageMaker AI에서는 대다수 인프라 관리 작업이 추상화되므로 애플리케이션 로직과 ML 워크플로 실행에 주력할 수 있습니다.

DO NOT COPY
ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



SageMaker AI 엔드포인트를 구성할 때는 몇 가지 주요 요소를 고려해야 합니다.

컴퓨팅 리소스는 모델에 사용할 수 있는 처리 능력을 결정하는 요소입니다. 모델에서 예상되는 추론 부하를 처리하기에 적합한 인스턴스 유형 및 인스턴스 수를 선택해야 합니다. SageMaker AI 사용 시에는 고가용성 유지를 위해 동일 유형의 인스턴스가 2개 이상 필요합니다.

스케일링 정책을 적용하면 트래픽 패턴이나 사용자 지정 지표에 따라 엔드포인트 용량을 자동 조정할 수 있습니다. 그러면 엔드포인트가 계속 변화하는 부하를 효율적으로 처리할 수 있습니다.

데이터 캡처 기능이 있으면 디버그, 감사 또는 규정 준수용으로 입력 데이터와 모델 예측을 저장할 수 있습니다. 이 기능을 사용할 때는 캡처된 데이터의 저장 위치와 샘플링 속도를 지정할 수 있습니다.

배포 전략은 새 모델 버전 롤아웃 방식을 제어하는 요소입니다. 블루/그린 배포, 롤링 배포 등의 기법을 사용하면 위험을 최소화할 수 있습니다.

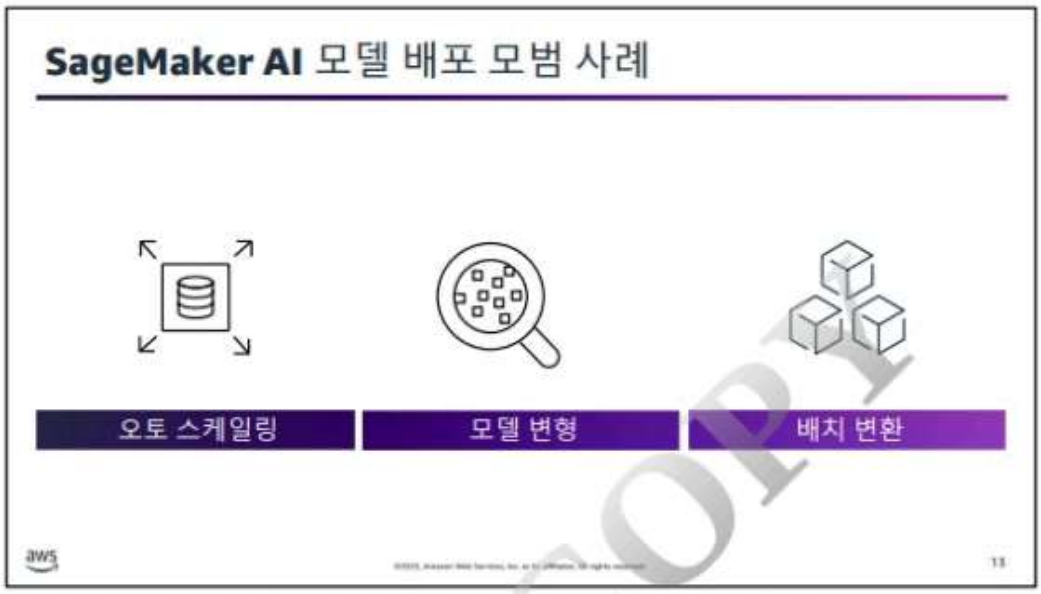
보안 및 액세스 제어는 모델과 데이터를 보호하려면 반드시 고려해야 하는 요소입니다. 구체적으로는 권한, 암호화 옵션 및 네트워크 액세스 제어 등을 구성해야 합니다.

모니터링 및 로깅 기능을 사용하면 엔드포인트 성능 및 상태를 추적할 수 있습니다. 세부 로깅을 활성화하고 보존 기간을 지정하면 로깅된 정보를 문제 해결 및 최적화 시에 참조할 수 있습니다.

DO NOT COPY
ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



SageMaker AI에서는 모델 배포 최적화를 위한 여러 모범 사례가 제공됩니다.

오토 스케일링은 워크로드에 따라 모델에 대해 프로비저닝되는 인스턴스의 수를 동적으로 조정합니다. 이처럼 크기가 자동으로 조정되면 엔드포인트가 급증하는 트래픽도 정상적으로 처리할 수 있으며, 작업량이 적은 기간 동안에는 비용을 최소화할 수 있습니다. 스케일링 정책을 구성하여 엔드포인트 규모를 조정할 시기와 방법을 정의할 수 있습니다.

모델 변형을 사용하면 같은 엔드포인트에서 한 모델의 여러 버전을 테스트할 수 있습니다. A/B 테스트를 진행하는 경우 또는 새 모델 버전을 점진적으로 롤아웃하려는 경우에는 이러한 여러 버전 테스트 방식이 특히 유용합니다. 트래픽 중 일정 비율을 각 변형에 라우팅하여 성능을 비교한 후 원래 모델을 완전 교체할 수 있습니다.

배치 변환은 대규모 데이터세트를 처리할 때 실시간 엔드포인트 대신 사용할 수 있는 옵션입니다. 개별 실시간 요청의 응답으로 예측을 생성하는 것이 아니라 전체 데이터세트와 관련된 예측을 생성해야 하는 시나리오에서는 배치 변환이 적합합니다. 여러 컴퓨팅 인스턴스에 워크로드가 자동 분산되는 배치 변환은 대규모 추론 태스크에 적합합니다.

이러한 모범 사례를 적용하면 SageMaker AI에서 더욱 안정적이며 비용 효율적인 확장 가능 모델 배포를 생성할 수 있습니다.

DO NOT COPY
ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

이 섹션에서는 SageMaker AI에서 지원되는 구체적인 배포 전략을 자세히 살펴봅니다. 이러한 전략을 채택하면 프로덕션 환경에서 기계 학습 모델을 효율적이면서도 안전하게 업데이트할 수 있습니다.

이 섹션에서 설명하는 전략은 다음과 같습니다.
- 블루/그린 배포 및 트래픽 이동 모드
- 롤링 배포

각 전략과 관련하여 알아볼 사항은 다음과 같습니다.
- 작동 방식
- 적합한 상황
- 이점 및 발생 가능한 단점

이러한 배포 전략을 파악하면 위험 허용 범위, 업데이트 긴급도, 리소스 제약과 같은 요인을 고려하여 구체적인 사용 사례에 가장 적합한 방식을 선택할 수 있습니다.

DO NOT COPY ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

SageMaker AI의 배포 가드레일

- 완전관리형 배포
- 자동 롤백
- 진행 상황 추적

15

SageMaker AI의 배포 가드레일에서는 프로덕션 환경에서 ML 모델을 업데이트하는 데 사용할 수 있는 완전관리형 배포 옵션 집합이 제공됩니다.

가드레일의 주요 기능은 다음과 같습니다.

- 완전관리형 배포: SageMaker AI가 엔드포인트 업데이트 메커니즘과 통합되어 배포 오케스트레이션을 처리합니다. 이와 같은 관리형 배포에서는 복잡한 배포 프로세스 관리 시의 운영 오버헤드가 감소합니다.
- 자동 롤백: 기본 제공 보호 기능이 배포 중에 발생하는 문제를 자동 탐지한 후 이전의 안정적 버전으로 롤백하는 과정을 시작할 수 있습니다. 이렇게 자동 롤백이 진행되면 문제가 있는 업데이트로 인해 장기간 모델 작동이 중단되는 상황을 방지할 수 있습니다.
- 진행 상황 추적: DescribeEndpoint API 또는 Amazon CloudWatch 이벤트를 통해 배포 진행 상황을 모니터링할 수 있습니다. 이와 같이 진행 상황을 모니터링하면 모델 업데이트 상태를 지속적으로 파악할 수 있습니다.

배포 가드레일은 여러 가지 이점을 제공합니다:

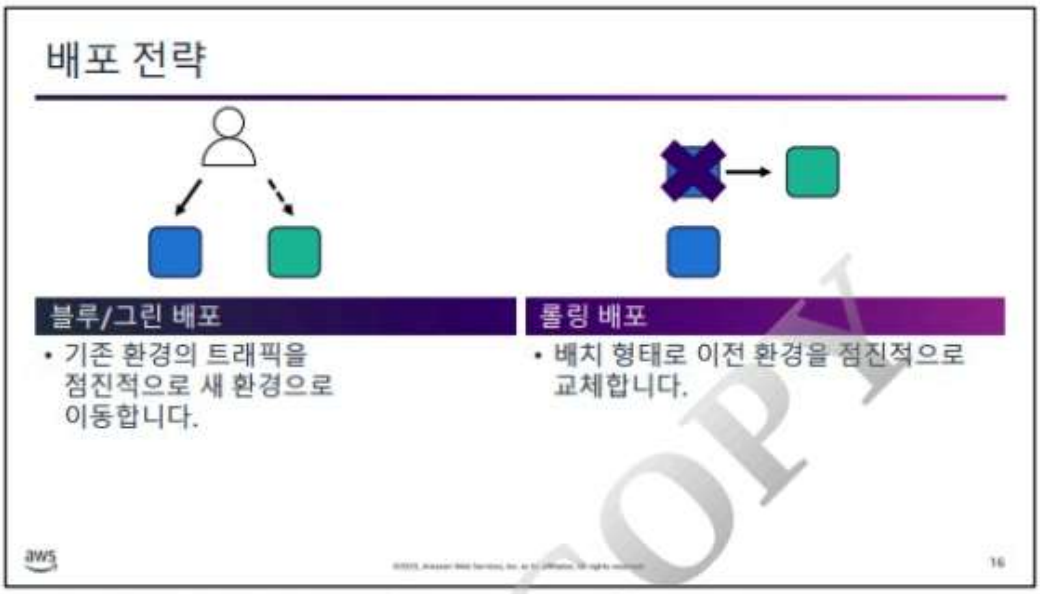
- 더욱 안전하게 프로덕션 환경 업데이트 가능
- 계획되지 않은 가동 중단 시간 또는 성능 문제 발생 위험 감소
- 배포 프로세스가 간소화되므로 ML 애플리케이션 개발에 주력할 수 있음
- 업데이트 중 트래픽 이동을 더욱 철저하게 제어할 수 있음

배포 가드레일은 비동기 추론 및 실시간 추론 엔드포인트 유형에만 적용됩니다.

이러한 가드레일을 활용하면 최소한의 작업만으로 모델 배포용 모범 사례를 구현할 수 있습니다.

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



SageMaker AI에서는 지원되는 2가지 주요 배포 전략 유형은 다음과 같습니다.

블루/그린 배포:
- 동일한 2개 프로덕션 환경, 즉 블루(기존) 환경과 그린(환경) 환경이 유지됨
- 블루 환경에서 그린 환경으로 트래픽 점진적 이동
- 문제 발생 시 블루 환경으로 트래픽을 리디렉션하여 빠르게 롤백할 수 있음
- 위험을 최소화하면서 새 버전을 배포 및 테스트할 수 있는 안전한 방식 제공

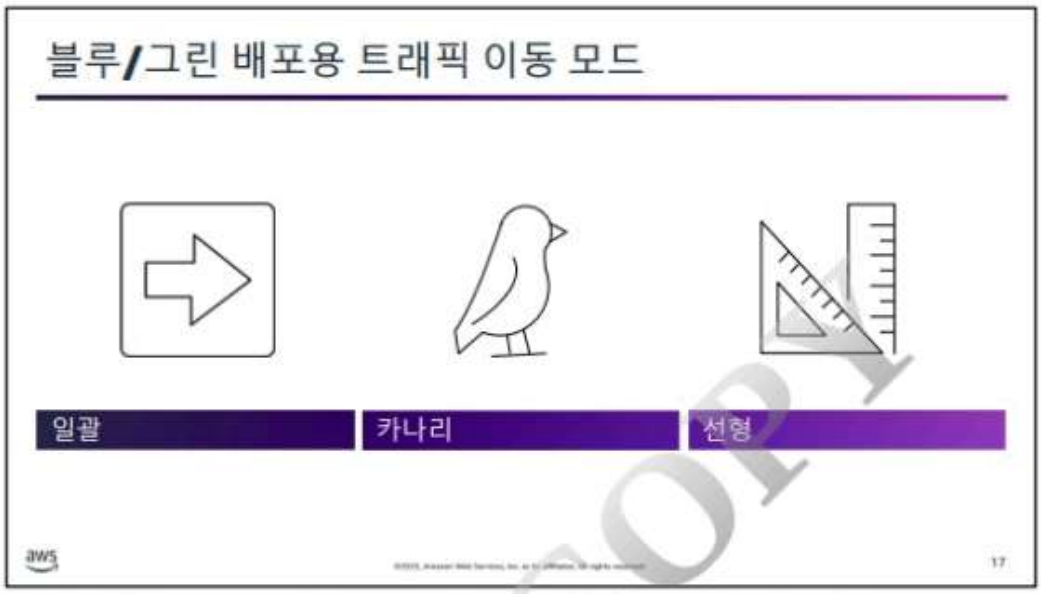
롤링 배포:
- 이전 버전 인스턴스를 새 버전으로 점진적 배치 단위로 교체
- 증분 방식으로 엔드포인트 업데이트에 따라, 블루/그린에 비해 필요한 전체 용량이 감소
- 인스턴스 수가 많아서 전체 환경을 복제하려면 비용이 많이 드는 대규모 모델이나 엔드포인트에 적합한 전략

두 전략의 주요 차이점은 다음과 같습니다.
- 블루/그린 배포의 경우 리소스는 더 많이 필요하지만 롤백이 더 빠르게 진행됩니다.
- 롤링 배포의 경우 사용되는 리소스는 더 적지만 업데이트 프로세스가 더욱 점진적으로 진행됩니다.

두 전략 모두 배포 중에 가동 중지 시간과 위험을 최소한으로 줄이는 것이 목표이지만 모델 크기, 리소스 제약, 위험 허용 범위 등의 요인에 따라 선택하는 전략이 달라집니다.

DO NOT COPY ssleeee1@gmail.com

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



SageMaker AI에서는 블루/그린 배포용으로 3가지 트래픽 이동 모드가 제공됩니다.

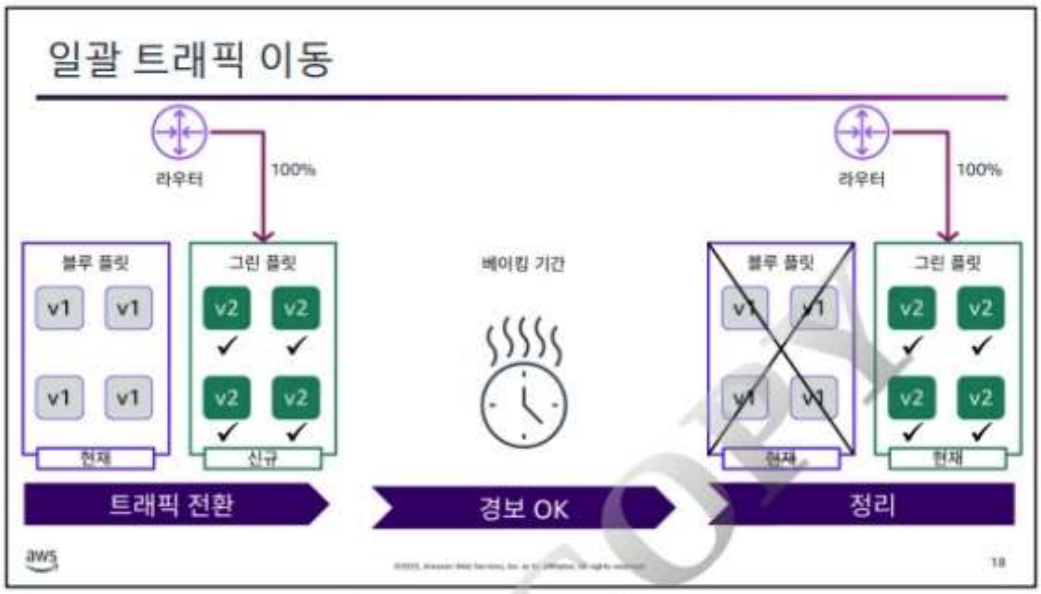
- **일괄**:
  - 한 번에 블루 플릿에서 그린 플릿으로 전체 트래픽 이동
  - 속도가 가장 빠른 동시에 위험성도 가장 높은 업데이트 방법
  - 업데이트의 중요도가 높지 않은 경우 또는 배포를 신속하게 진행해야 하는 경우에 적합
- **카나리**:
  - 두 단계로 트래픽 이동:
    1. 소량의 트래픽(**카나리**)을 그린 플릿으로 이동
    2. 이동이 정상적으로 완료되면 나머지 트래픽 이동
  - 배포에 대한 영향을 최소화하면서 문제 조기 탐지 가능
  - 중요한 변경 사항을 신중하게 롤아웃하려는 경우 적합
- **선형**:
  - 여러 단계(동일 크기)를 진행하여 점진적으로 트래픽 이동
  - 배포 프로세스를 가장 세부적으로 제어할 수 있음
  - 위험을 방지하려는 시나리오 또는 각 단계를 철저하게 모니터링해야 하는 시나리오에 가장 적합

각 모드에서는 트래픽 이동 후의 베이킹 기간을 사용하여 배포를 계속 진행하거나 마무리하기 전에 문제를 모니터링합니다. 그러므로 문제가 탐지되는 경우 이전 버전으로 안전하게 롤백할 수 있습니다.

위험 허용 범위, 업데이트 중요도 및 모니터링 기능에 따라 선택하는 트래픽 이동 모드가 달라집니다.

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



일괄 트래픽 이동은 SageMaker AI에서 속도가 가장 빠른 동시에 위험성도 가장 높은 블루/그린 배포 전략입니다. 일괄 트래픽 이동의 작동 방식은 다음과 같습니다.

1. 초기 상태: 모든 트래픽이 블루(현재) 플릿으로 전송됩니다.
2. 트래픽 이동: 전체 트래픽이 그린(신규) 플릿으로 즉시 리디렉션됩니다.
3. 베이킹 기간: 트래픽이 이동되고 나면 사전 정의된 베이킹 기간이 시작됩니다. 베이킹 기간은 전체 전환을 완료하기 전에 그린 플릿의 성능을 모니터링하기 위해 설정된 시간으로, 경보가 울리면 롤백할 수 있습니다. 이 기간은 영구적인 전환 이전에 신규 배포에 대한 확신을 심어줍니다.
4. 마무리: 베이킹 기간 동안 경보가 확인되지 않으면 배포는 성공한 것으로 간주되며 블루 플릿은 제거됩니다.
5. 롤백: 경보가 확인되면 트래픽이 블루 플릿으로 즉시 다시 이동됩니다.

일괄 트래픽 이동 시 제공되는 주요 이점은 다음과 같습니다.

- 업데이트 소요 시간 최소화
- 2개 플릿이 모두 실행되는 데 걸리는 시간이 단축되므로 비용이 절약됨

일괄 트래픽 이동 시에 발생 가능한 단점은 다음과 같습니다.

- 모든 사용자에게 새 버전이 즉시 표시되므로 위험성이 더 높음
- 점진적 롤아웃 또는 손쉬운 부분 롤백 불가

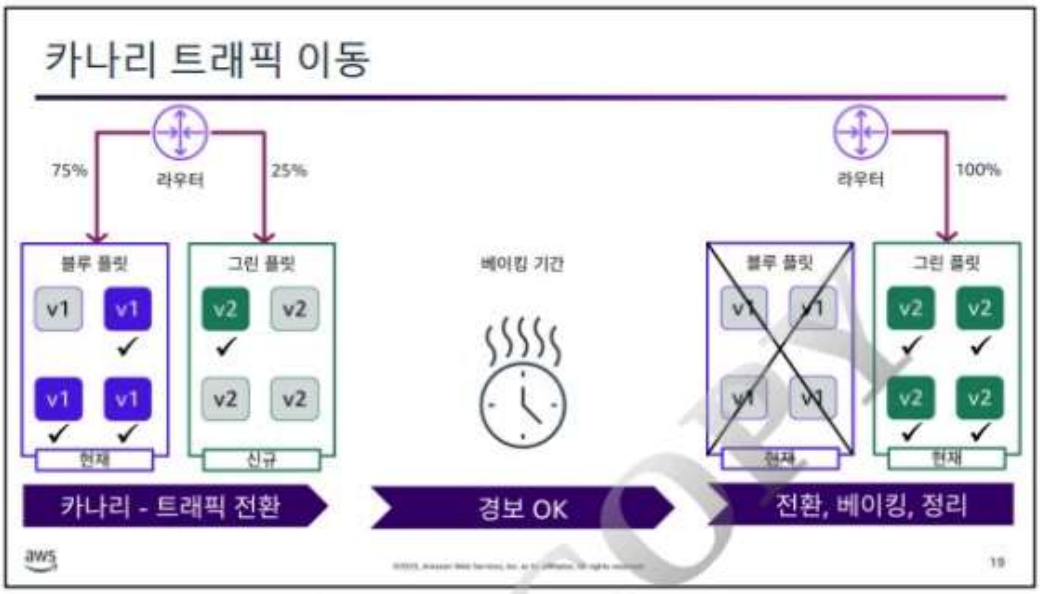
이 전략이 가장 적합한 상황은 다음과 같습니다.

- 중요도가 높지 않은 업데이트
- 빠른 배포를 진행해야 하는 시나리오
- 안정적인 모니터링 및 빠른 롤백 기능이 있는 환경

이 전략을 사용할 때는 항상 안정적인 롤백 계획을 세우고 트래픽 이동을 모니터링해야 합니다.

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



카나리 트래픽 이동은 일괄 이동에 비해 트래픽을 좀 더 신중하게 이동하는 Amazon SageMaker AI의 블루/그린 배포 전략입니다. 이 전략에서 수행하는 2가지 주요 단계는 다음과 같습니다.

1. **카나리 릴리스**:
   - 소량의 트래픽(**카나리**)을 그린 플릿으로 이동합니다.
   - 이 카나리 트래픽은 대개 총 트래픽 양의 5~10% 정도입니다.
   - 베이킹 기간 동안 카나리 플릿을 모니터링합니다.
2. **전체 배포**:
   - 카나리 단계에서 문제가 탐지되지 않으면 나머지 트래픽을 그린 플릿으로 이동합니다.
   - 그러면 베이킹 기간이 다시 시작됩니다. 이 베이킹 기간이 끝날 때까지 경보가 확인되지 않으면 그린 플릿이 모든 트래픽을 처리하며 SageMaker AI가 블루 플릿에 대한 모든 트래픽 전송을 중지합니다.

카나리 트래픽 이동 시 제공되는 주요 이점은 다음과 같습니다.

- 처음에는 트래픽의 일부만 새 버전으로 이동하는 방식을 통해 위험 감소
- 배포에 대한 영향을 최소화하면서 문제 조기 탐지 가능
- 안전성과 배포 속도의 적절한 균형 유지

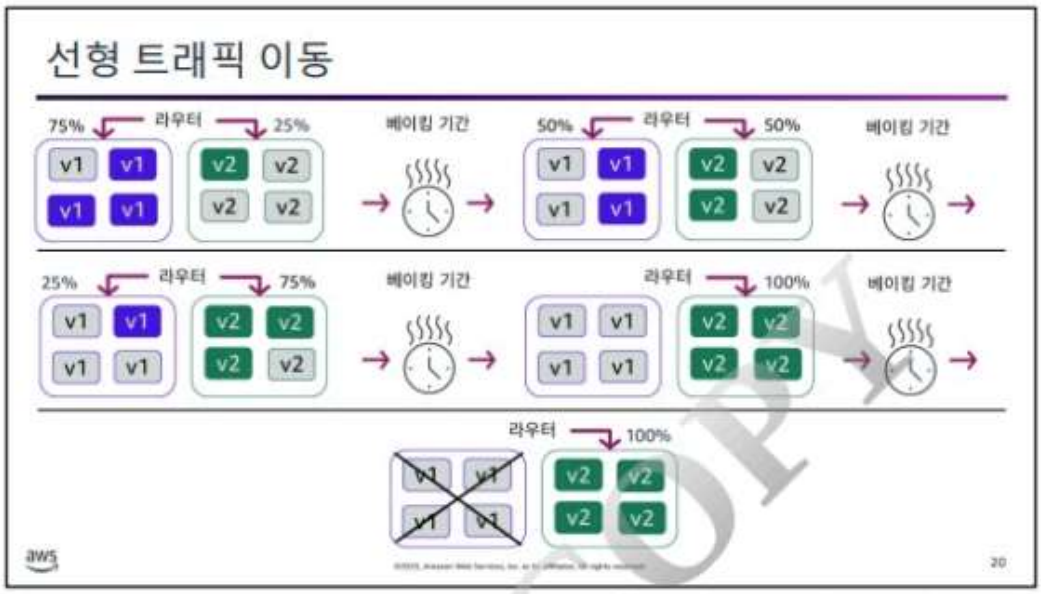
카나리 트래픽 이동 관련 고려 사항은 다음과 같습니다.

- 트래픽의 특성을 대표할 수 있을 정도로는 크되 배포에는 제한적인 영향만 주는 작은 카나리 크기를 결정해야 합니다.

카나리 배포를 구현할 때는 배포가 성공했음을 결정하거나 롤백을 시작할 기준으로 사용하려는 지표와 임곗값을 신중하게 고려해야 합니다.

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



선형 트래픽 이동은 SageMaker AI에서 트래픽이 가장 점진적으로 이동되는 블루/그린 배포 전략입니다. 선형 트래픽 이동에서는 크기가 동일한 여러 단계를 통해 블루 플릿에서 그린 플릿으로 트래픽을 이동하게 됩니다.

1. **초기 상태**:
   - 모든 트래픽이 블루 플릿에 있습니다.
2. **점진적 이동**:
   - 트래픽이 사전 정의된 증분 단위로 이동됩니다. 증분 크기는 그린 플릿 용량의 10~50% 사이여야 합니다.
   - 각 단계 후에는 문제 모니터링을 위한 베이킹 단계가 시작됩니다.
   - 문제가 탐지되지 않으면 다음 단계가 시작됩니다.
3. **완료**:
   - 전체 트래픽이 그린 플릿으로 이동되면 최종 베이킹 기간이 진행된 후 블루 플릿시 제거됩니다.

다음은 선형 트래픽 이동의 주요 특징입니다.
- 단계 수와 각 단계에서 이동되는 트래픽의 비율을 구성할 수 있습니다.
- 일반적으로 사용되는 단계 크기는 전체 트래픽의 10~50%입니다.
- CloudWatch 경보가 각 단계를 모니터링하여 문제 발생 가능성을 파악합니다.

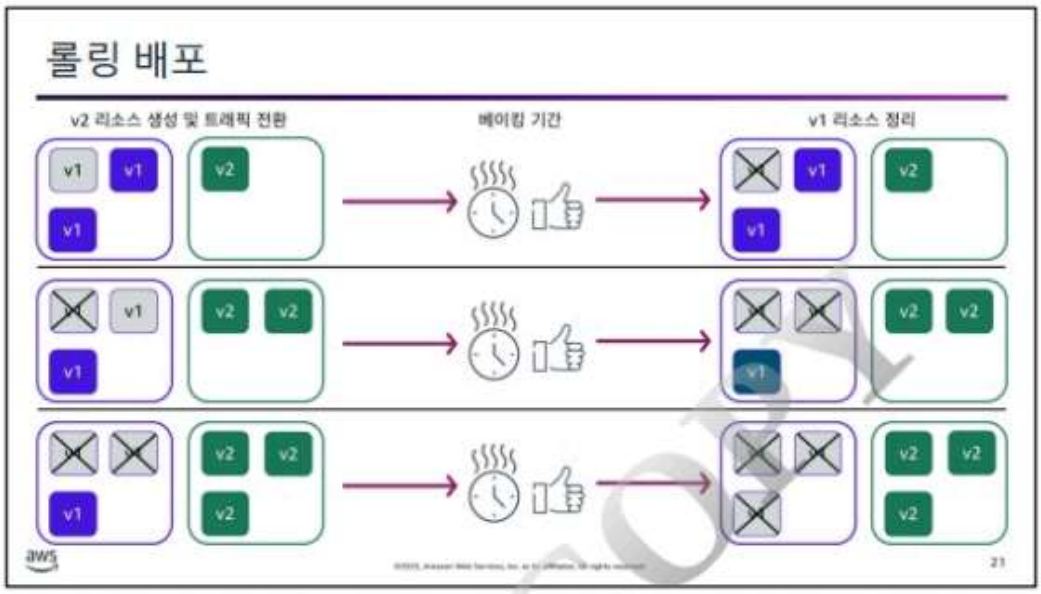
선형 트래픽 이동의 이점은 다음과 같습니다.
- 배포 프로세스를 가장 세부적으로 제어할 수 있음
- 롤아웃의 각 단계에서 트래픽을 면밀하게 모니터링할 수 있음
- 새 버전이 표시되는 사용자가 제한되므로 위험이 최소화됨

선형 트래픽 이동 관련 고려 사항은 다음과 같습니다.
- 다른 전략에 비해 전체 배포에 시간이 더 오래 걸림
- 프로세스 전반을 더욱 적극적으로 모니터링해야 함
- 두 플릿이 모두 장시간 실행되므로 비용이 더 많이 발생할 수 있음

선형 이동 사용 시에는 위험 허용 범위와 리소스 제약을 고려하여 적절한 단계 수를 선택해야 합니다.

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



롤링 배포에서는 인스턴스를 '배치' 단위로 교체하여 모델 버전을 점진적으로 업데이트할 수 있습니다. 롤링 배포의 주요 단계는 다음과 같습니다.

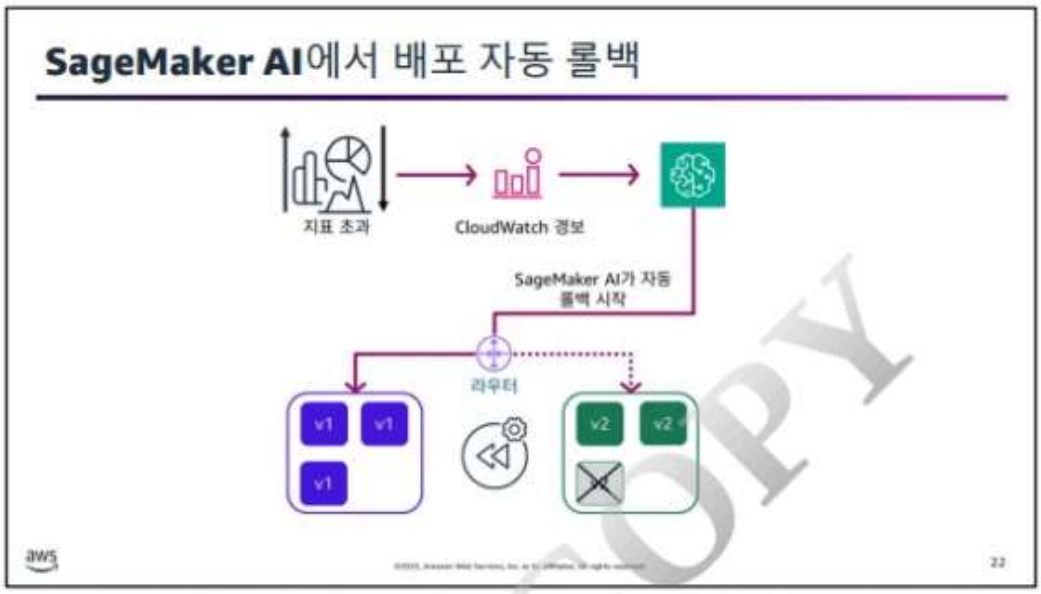
1. 새 v2 리소스를 생성한 후 트래픽 중 일부를 해당 리소스로 이동합니다.
2. 베이킹 기간에서 문제가 없음이 확인되면 새 리소스로 교체된 v1 리소스를 삭제합니다.
3. 점진적으로 단계를 수행하여 추가 v1 리소스를 v2 리소스로 교체합니다.
4. 배포가 성공하면 이전 v1 리소스를 정리합니다.

이 방법을 사용하는 경우 가동 중지 시간을 최소화하면서 롤아웃을 제어할 수 있습니다. 롤링 배포는 선형 블루/그린 배포와 비슷하지만 용량 요구 사항이 더 낮습니다. 각 단계가 성공하고 나면 이전 인스턴스가 제거되기 때문입니다.

DO NOT COPY
ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



SageMaker AI에서는 자동 롤백 기능이 제공됩니다. 자동 롤백은 새 모델 버전을 안전하게 배포할 수 있는 중요한 기능입니다.

- CloudWatch 경보가 새로 배포한 플릿의 주요 지표를 지속적으로 모니터링합니다.
- 베이킹 기간 동안 지표가 정의된 임계값을 초과하는 경우 문제가 있을 수도 있는 것입니다.
- 이 경우 SageMaker AI가 롤백을 자동으로 시작하여 원래 안정적 버전으로 트래픽을 다시 리디렉션합니다.

이 기능을 사용하면 애플리케이션을 사용 가능한 상태로 유지하고 배포 문제의 영향을 최소화할 수 있습니다.

자동 롤백 기능은 SageMaker AI에서 관리형 배포 옵션을 사용하는 경우의 주요 이점이라 할 수 있습니다. 프로덕션 환경에 업데이트를 롤아웃할 때 위험이 감소하기 때문입니다.

DO NOT COPY
ssleeeee1@gmail.com

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

Amazon SageMaker AI 배포를 위한 트래픽 이동 전략을 선택할 때 고려해야 하는 사항은 다음과 같습니다.

- **일괄:**
  - 한 단계를 통해 전체 트래픽을 새 플릿으로 이동
  - 가장 속도가 빠르고 비용이 저렴한 방법
  - 문제 발생 시 모든 트래픽이 영향을 받으므로 위험성도 가장 높음
- **카나리:**
  - 소규모 카나리 트래픽을 이동한 후 전체 트래픽을 이동하는 2단계 프로세스
  - 속도도 비교적 빠르며 위험도 다소 완화됨
  - 전체 롤아웃 전에 하위 집합에서 테스트 진행 가능
- **선형:**
  - 여러 동일 단계를 진행하여 점진적으로 트래픽 이동
  - 위험성이 가장 낮으며 롤아웃을 가장 철저하게 제어할 수 있음
  - 배포 소요 시간이 가장 길며 두 플릿이 모두 실행되므로 비용도 가장 많이 발생함

위험 허용 범위, 업데이트 긴급성 및 비용 고려 사항에 따라 적절한 전략을 선택해야 합니다. 위험도가 낮은 긴급 업데이트의 경우 일괄 전략이 가장 적합합니다. 카나리 전략 사용 시에는 대다수 시나리오에서 속도와 위험도의 균형을 유지할 수 있습니다. 선형 전략은 중요한 시스템에 사용할 수 있는 가장 신중한 방식입니다.

DO NOT COPY
ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

엔드포인트 또는 엔드포인트 구성을 생성할 때 추론 옵션을 선택해야 합니다. 모델을 효율적이고 효과적으로 배포하려면 올바른 모델 추론 전략을 선택해야 합니다. Amazon SageMaker AI는 여러 가지 모델 추론 옵션을 제공하며, 옵션마다 고유의 장점과 단점이 있습니다.

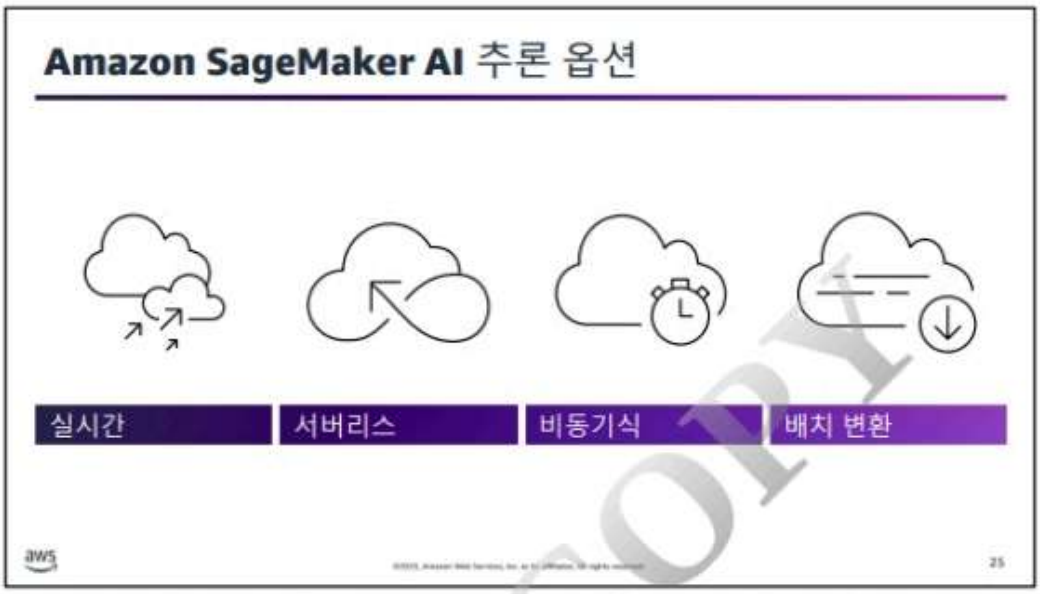
추론 전략을 선택할 때는 다음 요소를 고려하십시오.

- **지연 시간 요구 사항**: 애플리케이션이 예측을 수신해야 하는 속도
- **처리량 요구**: 처리해야 할 것으로 예상되는 요청의 수
- **페이로드 크기**: 입력 및 출력 데이터의 크기
- **처리 시간**: 예측을 생성하는 데 걸리는 시간
- **워크로드 패턴**: 트래픽이 일정한지 간헐적인지

이러한 옵션을 파악하면 애플리케이션 요구 사항에 부합하는 추론 전략을 선택하고 ML 배포를 최적화할 수 있습니다.

DO NOT COPY ssleeeee1@gmail.com

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



Amazon SageMaker AI에서는 여러 가지 워크로드 요구 사항을 충족할 수 있는 4가지 주요 추론 옵션이 제공됩니다.

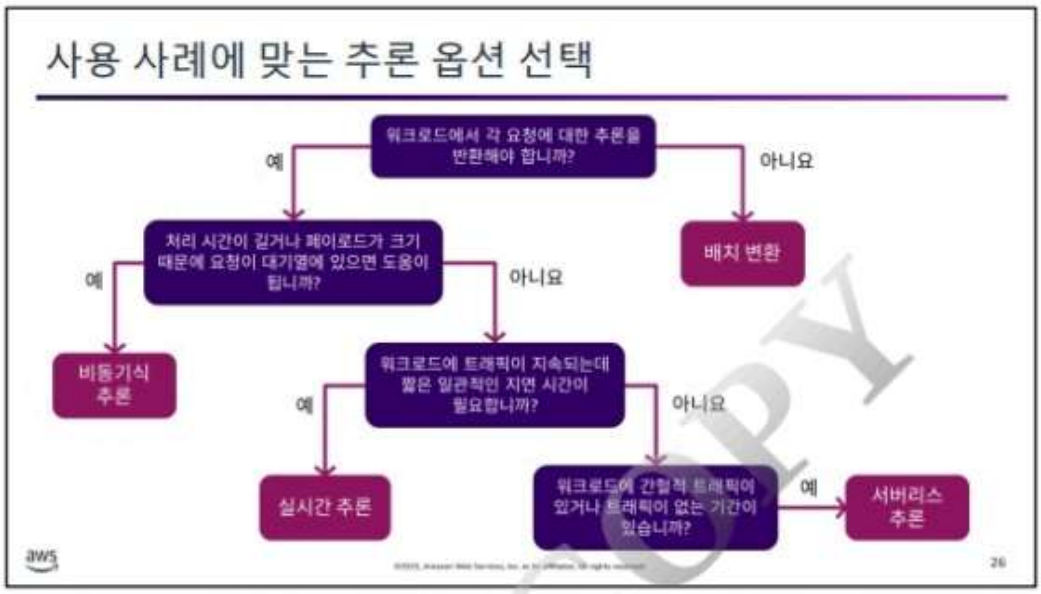
- **실시간 추론** - 지연 시간이 짧은 대화형 응답을 제공해야 하는 워크로드에 적합합니다. 이 옵션을 사용할 때는 지연 시간 1밀리초 미만 요구 사항을 충족하면서 소량의 데이터로 예측을 할 수 있는 영구 엔드포인트가 제공됩니다.
- **서버리스 추론** - 트래픽 패턴이 간헐적이거나 예측 불가능한 워크로드에 적합합니다. 이 기능은 AWS Lambda와 통합되어 SageMaker AI 추론 워크로드의 고가용성, 기본 제공 내결함성, 오토 스케일링을 제공합니다. 인스턴스 유형을 선택하거나 스케일링 정책을 관리할 필요가 없습니다.
- **비동기 추론** - 페이로드 크기가 크거나(최대 1GB) 처리 시간이 긴(최대 1시간) 요청에 가장 적합합니다. 비동기 추론에서는 수신 요청을 대기열에 추가하고 비동기식으로 처리합니다.
- **배치 변환** - 영구 엔드포인트가 필요하지 않을 때 대규모 데이터세트를 오프라인으로 처리하거나 데이터를 사전 처리하거나 추론을 실행하려는 경우 유용합니다. 입력 레코드를 추론과 연결해야 하는 시나리오에서는 배치 변환이 적합합니다.

각 옵션의 장점과 사용 사례는 서로 다릅니다. 구체적인 지연 시간, 처리량, 페이로드 크기 및 처리 시간 요구 사항에 따라 선택하는 옵션이 달라집니다.

DO NOT COPY
ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



사용 사례에 맞는 추론 옵션을 선택할 때 고려해야 하는 지침은 다음과 같습니다.

배치 변환이 적합한 상황:
- 대규모 데이터세트의 오프라인 처리
- 훈련 전 데이터 사전 처리
- 영구 엔드포인트가 필요 없는 시나리오

비동기 추론이 적합한 상황:
- 대규모 페이로드(최대 1GB) 처리
- 긴 처리 시간(최대 1시간)
- 지연 시간이 더 길어져도 되는 워크로드

실시간 추론이 적합한 상황:
- 짧은 지연 시간 및 높은 처리량 요구 사항이 적용되는 경우
- 지속적인 트래픽 보관을 위한 영구 API 엔드포인트가 필요한 경우
- 페이로드 크기 최대 6MB, 처리 시간 최대 60초

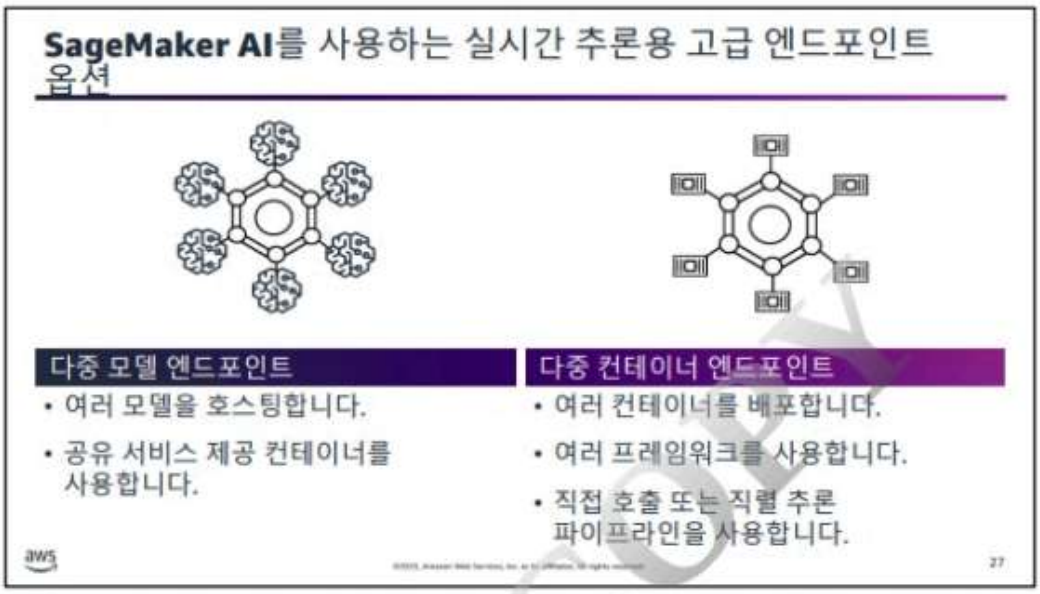
서버리스 추론이 적합한 상황:
- 트래픽 패턴이 간헐적이거나 패턴 예측이 불가능한 경우
- 인프라를 관리하지 않으려는 시나리오
- 트래픽에 따라 자동으로 크기를 조정하려는 경우
- 페이로드 크기 최대 4MB, 처리 시간 최대 60초

지연 시간 요구 사항, 페이로드 크기, 처리 시간, 트래픽 패턴 등의 요인을 고려하여 구체적인 사용 사례에 따라 가장 적합한 옵션을 선택해야 합니다.

DO NOT COPY ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



ML 사용 사례가 점점 복잡해짐에 따라 다양한 유형의 요청을 처리하거나 다양한 워크로드를 처리하기 위해 여러 모델 또는 여러 컨테이너를 배포해야 할 때가 있습니다. AWS는 다중 모델 및 다중 컨테이너 배포를 위한 여러 가지 옵션을 제공하므로 인프라 및 리소스 사용 방식을 최적화할 수 있습니다.

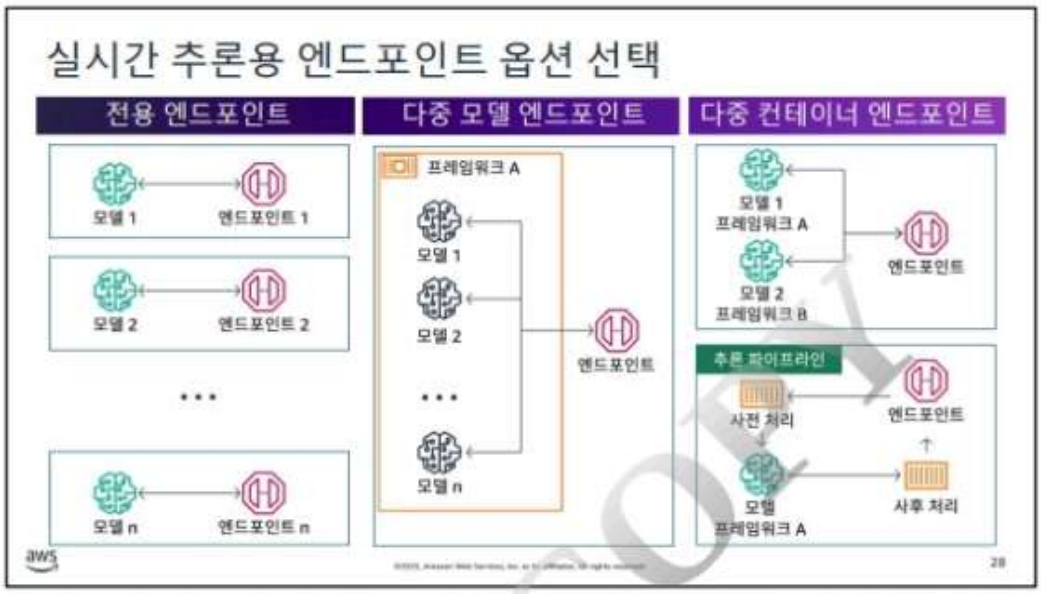
다중 모델 엔드포인트
- 단일 엔드포인트 내에서 여러 모델을 호스팅합니다.
- 공유 서비스 제공 컨테이너를 사용합니다.

다중 컨테이너 엔드포인트
- 단일 엔드포인트 내에 여러 컨테이너를 배포합니다.
- 다양한 프레임워크 또는 모델 유형이 지원됩니다.
- 특정 컨테이너를 직접 호출할 수 있습니다.
- 직렬 추론 파이프라인(컨테이너 2~15개)을 사용할 수 있습니다.

다중 모델 엔드포인트는 리소스 요구 사항이 비슷하며 컨테이너를 공유할 수 있는 모델이 여러 개일 때 적합합니다. 다중 컨테이너 엔드포인트는 사용하는 프레임워크나 처리 요구 사항이 서로 다른 모델을 배포해야 할 때 유용합니다.

DO NOT COPY ssleeeee1@gmail.com

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



SageMaker AI에서는 여러 모델을 배포할 때 사용 가능한 3가지 주요 엔드포인트 옵션이 제공됩니다.

전용 엔드포인트
- 각 모델에 자체 엔드포인트가 있습니다.
- 성능은 가장 우수하지만 비용이 더 많이 듭니다.
- 처리량 요구 사항이 엄격하거나 필요한 리소스의 양이 각기 다른 모델이 여러 개일 때 적합합니다.

SageMaker AI 다중 모델 엔드포인트
- 단일 엔드포인트에서 여러 모델을 호스팅합니다.
- 공유 서비스 제공 컨테이너와 리소스 플릿을 사용합니다.
- 필요시 모델을 메모리에 동적으로 로드합니다.
- 호스팅 비용이 낮아지며 엔드포인트 사용률이 높아집니다.
- 리소스 요구 사항이 비슷한 여러 모델을 배포할 때 적합합니다.

SageMaker AI 다중 컨테이너 엔드포인트
- 단일 엔드포인트에 여러 컨테이너를 배포합니다.
- 각 컨테이너가 서로 다른 모델이나 프레임워크를 사용할 수 있습니다.
- 추론 파이프라인 또는 개별 컨테이너 호출이 지원됩니다.
- 유동적으로 추론을 하면서 비용을 최적화할 수 있습니다.
- 워크플로가 복잡하거나 여러 프레임워크를 사용할 때 적합합니다.

추론 파이프라인을 사용하여 전처리, 예측, 사후 처리 데이터 과학 과제를 결합할 수 있습니다. 추론 파이프라인은 완전관리형입니다.

성능 요구, 비용 제약, 모델이나 프레임워크의 종류 등의 요인을 고려하여 구체적인 요구 사항에 따라 적합한 옵션을 선택해야 합니다.

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

사용 사례에 적합한 엔드포인트 옵션 선택

다중 모델 엔드포인트	다중 컨테이너 엔드포인트
• A/B 테스트	• 데이터 전처리
• 앙상블 모델	• 후처리 및 해석
• 동적 모델 로드	• 다중 모달 추론
• 멀티 테넌트 애플리케이션	• 복잡한 워크플로

29

구체적인 사용 사례에 따라 적절한 엔드포인트 옵션도 달라집니다. 다중 모델 및 다중 컨테이너 엔드포인트가 유용한 몇 가지 시나리오가 아래에 나와 있습니다.

다중 모델 엔드포인트 사용 사례
- A/B 테스트: 비교용으로 여러 모델 버전 호스팅
- 앙상블 모델: 서로 연동되는 여러 특수 모델 배포
- 동적 모델 로드: 자주 사용하지 않는 여러 모델을 효율적으로 관리
- 멀티 테넌트 애플리케이션: 단일 엔드포인트에서 여러 고객에게 서비스 제공

다중 컨테이너 엔드포인트 사용 사례
- 데이터 전처리: 기본 모델에서 사전 처리 태스크 실행
- 후처리 및 해석: 모델 예측 관련 추가 태스크 수행
- 다중 모달 추론: 특수 컨테이너를 사용하여 다양한 데이터 유형 처리
- 복잡한 워크플로: ML 파이프라인을 재사용 가능한 모듈식 구성 요소로 분할

이러한 옵션 중 적절한 옵션을 결정할 때는 모델 수, 리소스 요구 사항, 프레임워크 종류, ML 워크플로의 복잡성과 같은 요소를 고려해야 합니다. 다중 모델 엔드포인트는 대개 비슷한 모델 여러 개를 호스팅할 때 비용 효율성이 높은 반면 복잡한 다단계 추론 프로세스에서는 다중 컨테이너 엔드포인트를 사용하면 더욱 유동적으로 프로세스를 진행할 수 있습니다.

DO NOT COPY
ssleee@gmail.com

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

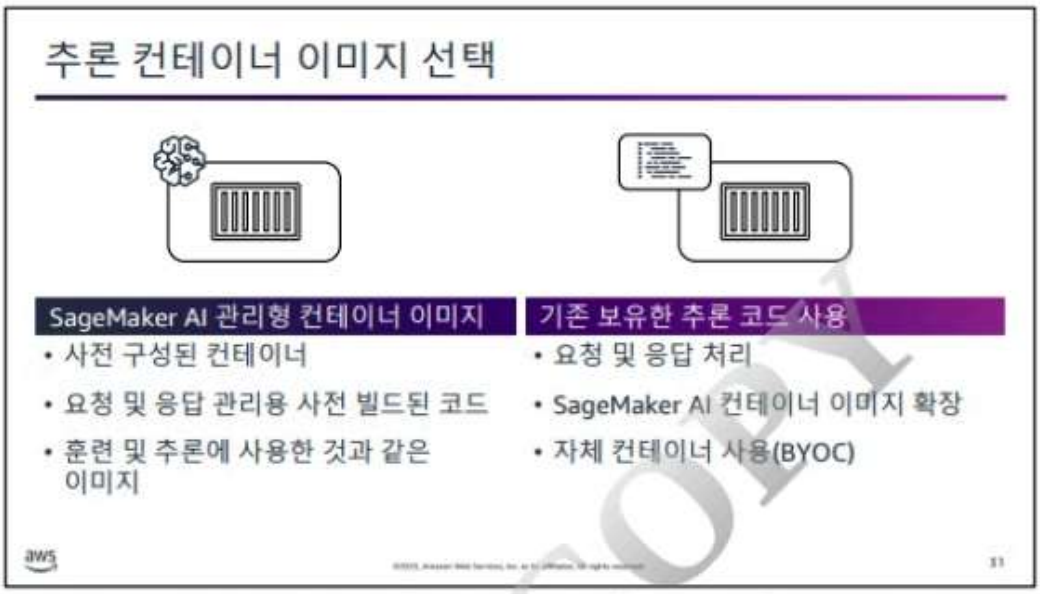
이 섹션에서는 추론 배포의 2가지 주요 요소를 살펴봅니다.

- **컨테이너 옵션**:
  - SageMaker AI에서는 널리 사용되는 ML 프레임워크용으로 사전 구축된 컨테이너 이미지가 제공됩니다.
  - 특수한 요구 사항을 충족하는 사용자 지정 컨테이너 이미지를 사용할 수도 있습니다.
- **인스턴스 유형**:
  - SageMaker AI에서는 다양한 ML 워크로드용으로 최적화된 여러 가지 인스턴스 유형이 제공됩니다.
  - 선택해야 하는 적합한 인스턴스 유형은 모델에서 수행해야 하는 계산 및 성능 요구 사항에 따라 달라집니다.

이러한 옵션을 이해하면 ML 모델을 배포할 때 정보를 토대로 결정을 내리는 데 도움이 됩니다. 이 섹션에서는 SageMaker AI의 관리형 컨테이너 이미지를 사용하는 방식과 기존 보유 컨테이너 이미지를 사용하는 방식 중 적절한 옵션을 선택하는 방법을 알아봅니다. 그리고 추론 요구 사항에 가장 적합한 인스턴스 유형을 선택하는 방법도 알아봅니다.

DO NOT COPY ssleeeee1@gmail.com

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



SageMaker AI는 비즈니스 요구 사항을 충족하는 여러 컨테이너 옵션을 제공합니다. 비즈니스 요구 사항을 가장 효과적인 컨테이너 이미지를 선택할 때 사용할 수 있는 2가지 접근 방식은 다음과 같습니다.

- SageMaker AI 관리형 컨테이너 이미지에 포함된 추론 로직 사용
- 기존 보유한 추론 코드 사용

SageMaker AI 관리형 컨테이너 이미지 사용

- 배포 - SageMaker AI 관리형 컨테이너 이미지를 사용하여 모델을 배포합니다. 사전 구성 컨테이너는 모델 호스팅 및 추론 처리를 포함하여 필요한 설정을 처리합니다.
- 사전 작성 코드 - 대다수 SageMaker AI 관리형 컨테이너 이미지에는 추론 요청 및 응답을 관리하기 위한 사전 작성 코드가 포함되어 있습니다.
- 모델 훈련 - 훈련과 추론에 같은 컨테이너 이미지를 사용할 수도 있습니다. SageMaker AI 관리형 컨테이너 이미지를 사용하여 모델을 훈련하는 경우 모델에 이미 추론 로직이 포함되어 있을 수 있습니다.

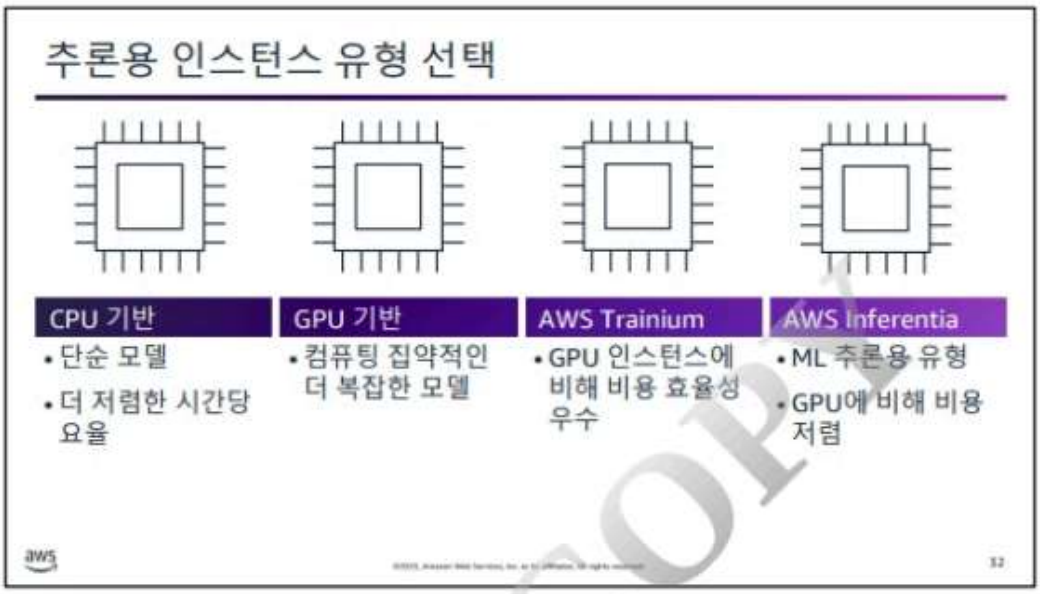
기존 보유한 추론 코드 사용

- 인바운드 및 아웃바운드 - 고객의 자체 추론 코드를 사용하면 인바운드 요청과 아웃바운드 응답을 유연하게 처리할 수 있습니다.
- SageMaker AI 컨테이너 확장 - 고객의 자체 추론 코드를 사용하여 SageMaker 관리형 컨테이너 이미지를 확장할 수 있습니다. 컨테이너 이미지는 프레임워크, 종속성, 라이브러리 등의 필수 구성 요소로 사전 구성된 상태로 제공됩니다.
- 자체 컨테이너 사용(BYOC) - 고객의 자체 컨테이너를 가져와 자체 추론 코드를 사용할 수도 있습니다. 모델, 종속성 및 기타 필요한 구성 요소가 포함된 사용자 지정 컨테이너 이미지를 생성하십시오. 이 옵션은 최고의 유연성을 제공합니다.

이러한 2가지 접근 방식 중 하나를 결정하려면 추론 로직의 복잡성, 사용자 지정의 필요성, 팀의 컨테이너 관리 기능 등의 요소를 고려하십시오. SageMaker AI 관리형 컨테이너는 즉시 사용 가능한 솔루션이라 할 수 있습니다. 반면, 사용자 지정 컨테이너는 관리형 컨테이너에 비해 유동적으로 사용 가능하지만 직접 처리해야 하는 컨테이너 관리 업무가 많습니다.

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



CPU

CPU 인스턴스는 기존 ML 모델에 효과적입니다. 이러한 인스턴스는 모델 프로토타입 제작, 소규모 데이터세트 미세 조정 또는 시간과 비용의 절충이 필요한 사용 사례에 적용됩니다. CPU 인스턴스는 다음과 같은 이유로 효과적입니다.

- 소규모 데이터세트에 사용: CPU 인스턴스는 소규모 데이터세트, 덜 복잡한 모델 또는 계산 집약적이지 않은 훈련 프로세스에 적합합니다.
- 비용 효율성: CPU 인스턴스는 일반적으로 그래픽 처리 장치(GPU) 가속 인스턴스에 비해 시간당 요금이 저렴하기 때문에 특정 사용 사례에서 비용 효율적일 수 있습니다.

사용 사례는 다음과 같습니다.

- 텍스트 분류 또는 감정 분석
- 재무 데이터
- 소규모 및 중규모 모델

GPU

GPU 인스턴스는 TensorFlow, PyTorch, Apache MXNet, 익스트림 그래디언트 부스팅(XGBoost)과 같은 ML 프레임워크를 훈련하는 데 효과적인 선택입니다. 그리고 신경망을 사용하여 딥 러닝(DL) 모델을 훈련하는 데에도 효과적입니다. 성능과 비용의 균형을 고려한 다양한 GPU 인스턴스 중에서 선택합니다. GPU 인스턴스는 다음과 같은 이유로 효과적입니다.

- 훈련에 자주 사용: GPU 인스턴스는 병렬 처리를 통해 계산을 가속화할 수 있기 때문에 대부분의 ML 훈련 태스크에 자주 사용됩니다.
- 훈련 시간 단축: GPU 인스턴스를 사용하면 훈련 시간이 단축되며 더 복잡한 모델에 사용되는 더 큰 데이터세트를 처리할 수 있습니다.

사용 사례는 다음과 같습니다.

- 시퀀스 모델링
- 객체 탐지 및 비디오 분석
- 대규모 모델 또는 생성적 적대 신경망(GAN)

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

AWS Trainium

추론에 AWS Trainium 인스턴스를 사용할 수도 있습니다. Trainium 인스턴스는 다음과 같은 이유로 효과적입니다.

- 비용 효율성: Trainium 인스턴스는 특히 대규모 ML 훈련 또는 추론 태스크를 처리하는 사용 사례에서 GPU 인스턴스보다 비용 효율적일 수 있습니다.
- 가용성이 제한될 수 있음: Trainium은 널리 제공되는 CPU 및 GPU 인스턴스에 비해 가용성이 제한되거나 리전이 제한적일 수 있습니다.

사용 사례는 다음과 같습니다.

- 컴퓨터 비전 및 음성 인식
- 권장 모델
- 시뮬레이션

AWS Inferentia

Amazon Elastic Compute Cloud(Amazon EC2) Inf2 인스턴스는 딥 러닝 추론용 목적별 인스턴스입니다. 이 인스턴스는 Amazon EC2에서 가장 저렴한 비용으로 대규모 언어 모델과 비전 트랜스포머를 비롯한 생성형 AI 모델에 고성능을 제공합니다.

사용 사례는 다음과 같습니다.

- 자연어 이해(고급 텍스트 분석, 문서 분석, 대화형 에이전트)
- 번역
- 이미지 및 비디오 생성
- 음성 인식

DO NOT COPY
ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

SageMaker AI 비용 최적화 옵션

- SageMaker AI Savings Plans
  - 일정 용량을 사용할 수 있는 유동적인 약정 요금제
- SageMaker AI 훈련 플랜
  - 훈련 작업
  - Amazon SageMaker HyperPod 클러스터
- 프로비저닝된 동시성(서버리스 추론용)
  - 트래픽 버스트 예측 가능 웜 엔드포인트

Amazon SageMaker AI Savings Plans에서는 1년 또는 3년 기간 동안 일정량의 사용량 (시간당 달러 단위)을 약정하는 대가로 Amazon SageMaker AI에 유연한 가격 책정 모델을 제공합니다. 이러한 플랜은 적격 SageMaker ML 인스턴스 사용량에 자동 적용됩니다.

SageMaker AI 훈련 플랜을 선택하면 SageMaker AI 훈련 작업이나 Amazon SageMaker HyperPod 클러스터 등 리소스가 필요한 구체적인 상황에 맞게 컴퓨팅 용량을 예약할 수 있습니다. 이 서비스에서는 신속한 컴퓨팅 리소스 프로비저닝, 인프라 설정, 워크로드 실행, 인프라 장애 시 복구가 자동으로 처리됩니다.

프로비저닝된 동시성을 이용한 서버리스 추론은 트래픽이 예측이 가능한 버스팅시 비용 효율적인 옵션입니다. 프로비저닝된 동시성 사용 시에는 성능을 예측할 수 있는 서버리스 엔드포인트에 모델을 배포한 후 엔드포인트를 활성 상태로 유지하여 확장성을 높일 수 있습니다.

DO NOT COPY
ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

ML 컴퓨팅 리소스 선택 시의 고려 사항

요구 사항	테스트 환경	프로덕션 환경
비용 관리	비용 효율적인 인스턴스 유형 스팟 인스턴스 수명이 짧은 리소스	안정성과 가용성을 최우선으로 고려
성능 및 확장성	유동적인 리소스 할당 엄격하지 않은 확장성 요구 사항	우수한 성능 및 신뢰성 확장성
배포 및 테스트	변경 및 배포 빈도 높음 다양한 테스트 지원	통제식 배포 롤백 메커니즘

ML 컴퓨팅 리소스 선택 시에 고려해야 하는 사항은 다음과 같습니다.

테스트 환경:

비용 관리
- 비용 효율적인 인스턴스 사용: 비용이 저렴한 Amazon EC2 스팟 인스턴스는 중요하지 않은 태스크에 사용합니다. 크기가 더 작고 저렴한 인스턴스 유형(T3 또는 T3a)을 사용할 수 있습니다.
- 수명이 짧은 리소스: 불필요한 비용이 발생하지 않도록 필요할 때만 리소스를 프로비저닝하고 테스트가 완료되면 종료합니다.

성능 및 확장성
- 유연한 리소스 할당: 프로덕션 환경과 거의 비슷하지만 스케일 다운할 수 있는 리소스를 프로비저닝합니다. 다양한 인스턴스 유형과 구성을 실험하여 최적의 성능을 선택합니다.
- 낮은 확장성 요구 사항: 일반적으로 부하 처리량이 적으므로 오토 스케일링의 중요성이 낮습니다.

배포 및 테스트
- 잦은 변경 및 배포: 새로운 기능과 수정 사항을 테스트하기 위해 변경 사항을 신속하게 반복하고 배포합니다. 지속적 통합 및 전달(CI/CD) 파이프라인을 사용하여 테스트 및 배포 프로세스를 자동화합니다.

DO NOT COPY ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

- 테스트 다양성: 단위 테스트, 통합 테스트, 성능 테스트를 비롯한 다양한 테스트를 수행합니다.

프로덕션 환경:

비용 관리
- 안정성과 가용성에 높은 우선 순위: 예측 가능하고 안정적인 성능을 제공할 수 있도록 Amazon EC2 온디맨드 인스턴스 또는 예약 인스턴스를 사용합니다. 오래 실행되고 안정적인 워크로드에는 비용을 절감할 수 있는 AWS 절감형 플랜을 고려해 보십시오.

성능 및 확장성
- 높은 성능과 신뢰성: 필요한 성능과 신뢰성을 제공하는 인스턴스 유형 및 구성을 선택합니다. 리소스가 최대 로드를 처리하고 짧은 지연 시간을 유지할 수 있어야 합니다.
- 확장성: Amazon EC2 Auto Scaling 그룹을 구현하여 수요에 따라 용량을 자동으로 조정합니다. Elastic Load Balancing(ELB)을 사용하여 트래픽을 여러 인스턴스에 효율적으로 분산합니다.

배포 및 테스트
- 통제된 배포: 블루/그린 배포 또는 카나리 릴리스를 사용하여 업데이트 중 위험을 최소화합니다. 프로덕션 환경에 배포하기 전에 스테이징 환경에서 업데이트를 철저하게 테스트합니다.
- 롤백 메커니즘: 문제가 발생하면 이전의 안정된 상태로 빠르게 되돌릴 수 있도록 강력한 롤백 메커니즘을 구현합니다.

DO NOT COPY
ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

ML 컴퓨팅 리소스 선택 시의 고려 사항

- 모델 복잡성
- 데이터세트 크기
- 실시간 추론
- 확장성 및 탄력성
- 비용 최적화

35

ML 컴퓨팅 리소스 옵션, 유형 및 사용 사례에 대해 알아보았습니다. 기계 학습을 위한 컴퓨팅 리소스를 프로비저닝할 때에는 상황에 따라 어떤 컴퓨팅 리소스를 어떻게 프로비저닝할 것인지 신중하게 고려해야 합니다.

- **모델 복잡도** – 심층 신경망이나 앙상블 방법과 같은 더 복잡한 모델은 훈련 및 추론에서 더 높은 계산 성능을 요구하는 경향이 있습니다. 따라서 까다로운 워크로드를 처리하기에 충분한 CPU, GPU 및 메모리 리소스를 할당하십시오.
- **데이터세트 크기** – 대규모 데이터세트는 메모리 및 스토리지 리소스에 상당한 부담을 줄 수 있습니다. 데이터 처리 태스크를 여러 노드에 분산하거나 병렬 처리 기술을 사용하여 성능을 높이십시오.
- **실시간 추론** – 전용 GPU 인스턴스와 같이 지연 시간이 짧아야 하는 추론을 위한 리소스를 특별히 프로비저닝합니다.
- **확장성 및 탄력성** – 오토 스케일링 기능 또는 로드 밸런싱 기술을 사용하여 워크로드 수요에 따라 컴퓨팅 리소스를 동적으로 조정합니다.
- **비용 최적화** – 필요한 성능과 비용 간의 적절한 균형점을 찾습니다. 스팟 인스턴스 또는 오토 스케일링 옵션을 사용하여 성능 저하 없이 비용을 최적화할 수 있습니다.

DO NOT COPY
ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

Amazon SageMaker Inference Recommender

- 추천 작업 실행
- 추론 유형 및 구성 평가
- 부하 테스트 자동화
- 선택 가능한 요소:
  - 최적 인스턴스 유형
  - 구성

Amazon SageMaker Inference Recommender에서는 ML 모델에 가장 적합한 인스턴스 유형과 구성을 선택할 수 있습니다.

Inference Recommender는 여러 SageMaker AI ML 인스턴스에서 부하 테스트와 모델 튜닝을 자동화합니다. 그리고 인스턴스 수, 컨테이너 파라미터, 모델 최적화, 최대 동시성, 메모리 크기 등의 요인를 고려하여 다양한 추론 유형 및 구성을 평가합니다. Recommender에서는 ML 모델 및 워크로드용으로 최적 인스턴스 유형과 구성을 선택할 수 있습니다.

Inference Recommender 사용 시에는 프로덕션 환경에 ML 모델을 배포하는 데 걸리는 시간과 배포 관련 작업을 대폭 줄일 수 있습니다.

DO NOT COPY
ssleeeee1@gmail.com

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

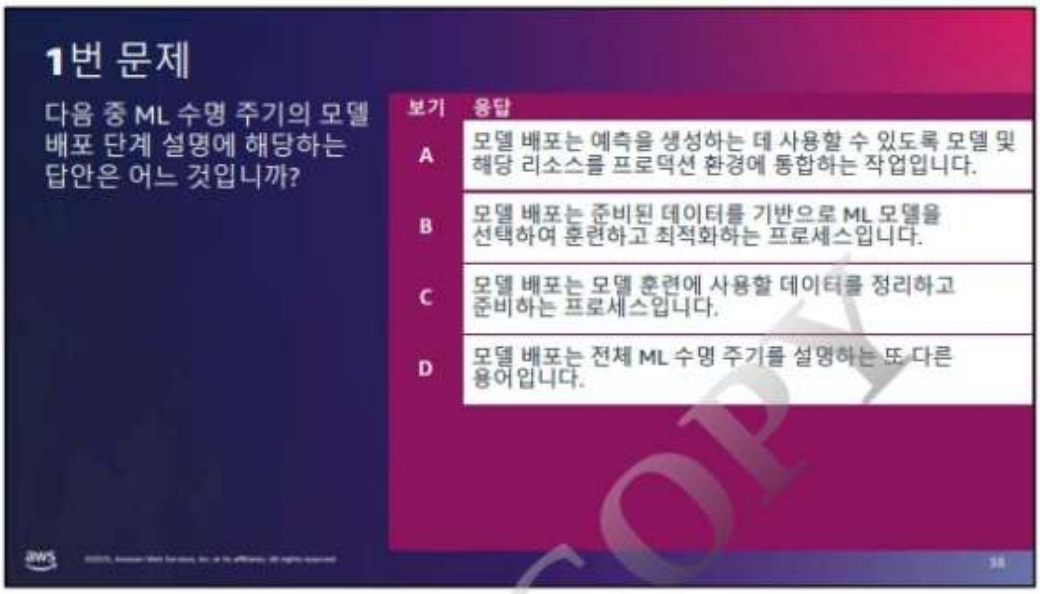
This slide introduces a knowledge check section
이제 간단한 지식 확인을 진행하여 이 모듈에서 알아본 모델 배포 전략 관련 주요 개념을 복습해 보겠습니다.

이 지식 확인을 통해 다음 모듈로 넘어가기 전에 수업 내용을 더욱 정확하게 이해하고 헷갈리는 부분을 확실하게 파악할 수 있습니다.

DO NOT COPY
ssleeeee1@gmail.com

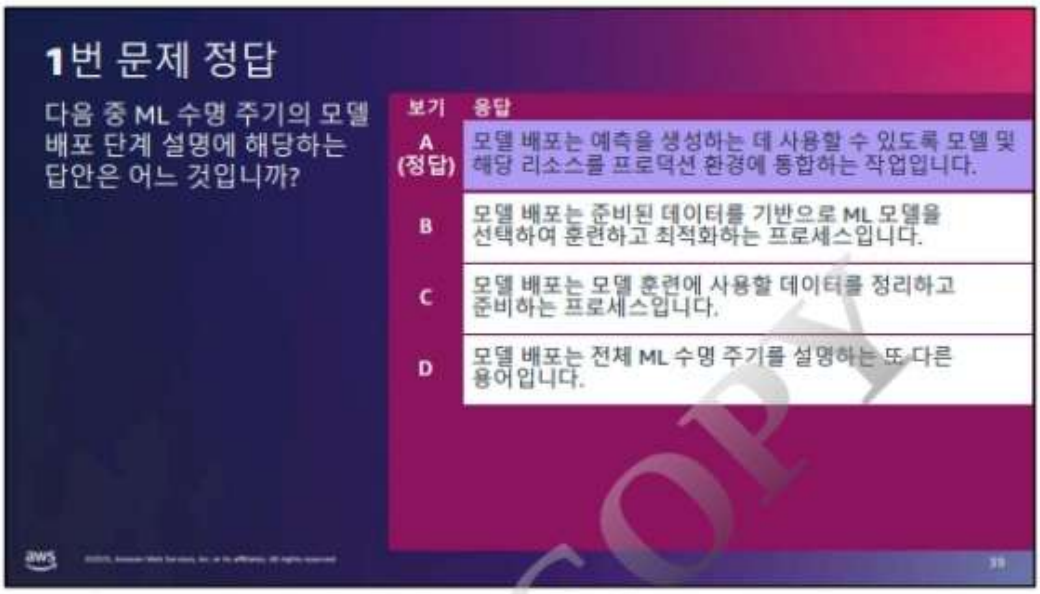
© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



DO NOT COPY
ssleeeee1@gmail.com

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



정답은 A입니다. 모델 배포는 예측을 생성하는 데 사용할 수 있도록 모델 및 해당 리소스를 프로덕션 환경에 통합하는 작업입니다.

설명:

A. (정답) 훈련된 모델을 사용할 수 있도록 제공하는 배포 단계에 대한 정확한 설명입니다.
B. (오답) 이 답안은 배포가 아닌 모델 훈련 단계에 대한 설명입니다.
C. (오답) 이 답안은 훈련 전에 진행되는 데이터 준비 단계에 대한 설명입니다.
D. (오답) 배포는 전체 프로세스가 아니라 ML 수명 주기의 구체적인 한 단계입니다.

DO NOT COPY
ssleeeee1@gmail.com

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

2번 문제

기계 학습(ML) 엔지니어가 고객의 검색 및 구매 기록을 기반으로 고객에게 제품을 추천하는 추천 시스템용 ML 모델을 설계하고 있습니다. 애플리케이션은 지연 시간이 짧고 처리량이 많아야 하는 사용자 요청을 실시간으로 처리해야 합니다.

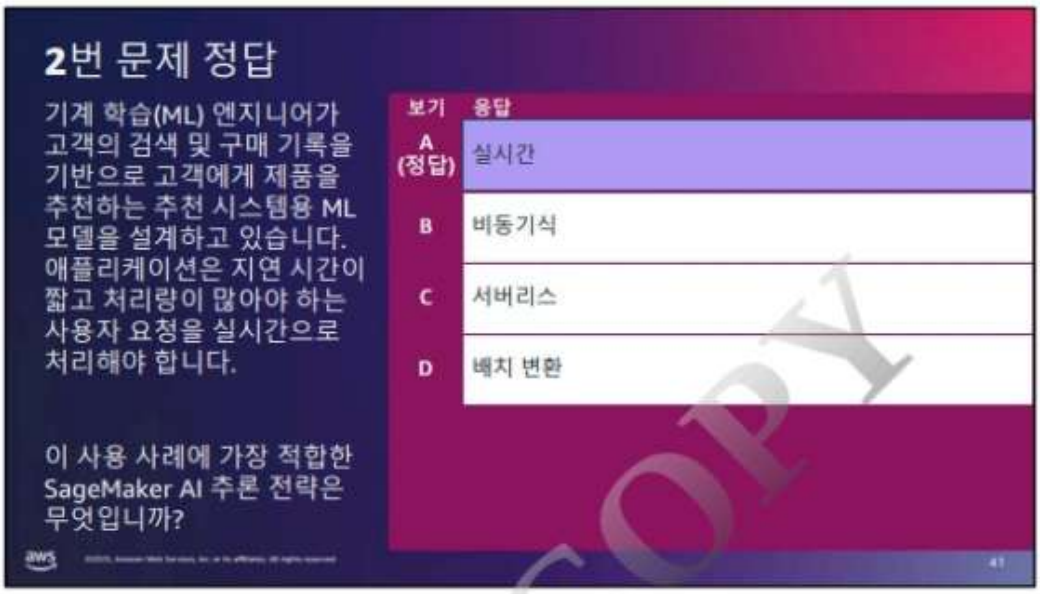
이 사용 사례에 가장 적합한 SageMaker AI 추론 전략은 무엇입니까?

보기	응답
A	실시간
B	비동기식
C	서버리스
D	배치 변환

40

DO NOT COPY
ssleeeee1@gmail.com

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



정답은 다음과 같습니다.

- A, 실시간 추론. 짧은 지연 시간 및 높은 처리량을 유지하면서 사용자 요청을 실시간으로 처리해야 하는 추천 시스템에 가장 적합한 답안은 실시간 추론입니다. Amazon SageMaker AI의 실시간 추론은 수신되는 요청에 즉시 응답을 해야 하는 시나리오에 사용할 수 있습니다. 실시간 추론에서는 지연 시간이 짧은 요청을 대량으로 처리할 수 있는 영구 엔드포인트가 제공됩니다. 따라서 추천 시스템 등의 대화형 응용 분야에 적합합니다.

그 외의 답안이 이 시나리오에 적합하지 않은 이유는 다음과 같습니다.

- B. 비동기 추론은 실시간 상호 작용이 아닌 페이로드가 크거나 처리 시간이 긴 시나리오에 적합합니다.
- C. 서버리스 추론은 유동적으로 진행할 수 있기는 하지만 실시간 추천에 필요한 짧은 대기 시간이 일정하게 제공되지는 않을 수도 있습니다.
- D. 배치 추론은 실시간 사용자별 추천이 아닌 대규모 데이터세트의 오프라인 처리용 전략입니다.

DO NOT COPY
ssleeeee1@gmail.com

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

3번 문제

데이터 사전 처리, 모델 훈련, 모델 제공과 같은 다양한 태스크를 처리하기 위해 여러 컨테이너가 필요한 기계 학습(ML) 모델을 개발 중인 팀이 있습니다.

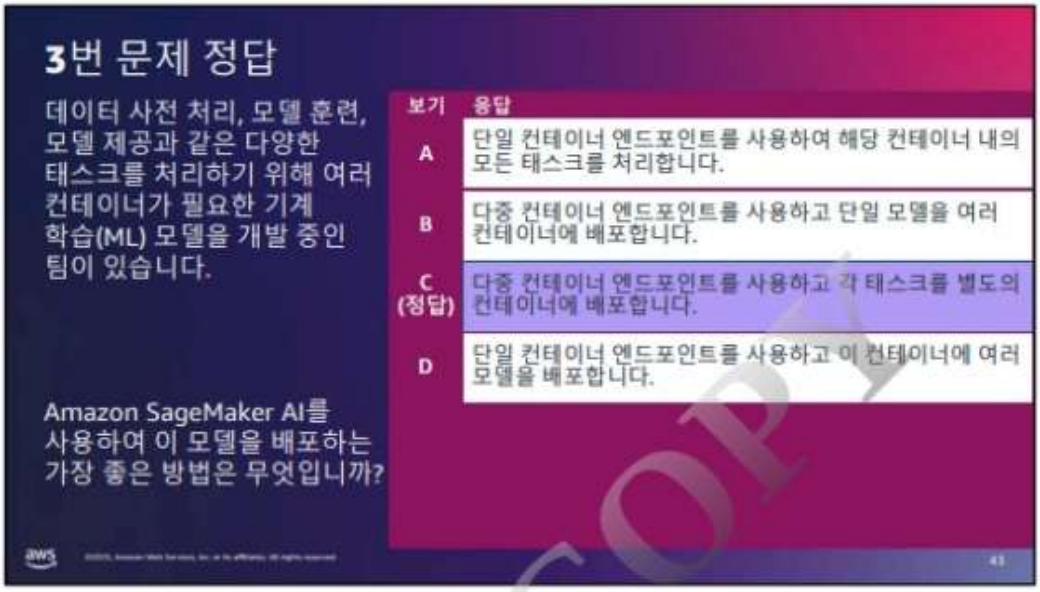
Amazon SageMaker AI를 사용하여 이 모델을 배포하는 가장 좋은 방법은 무엇입니까?

보기	응답
A	단일 컨테이너 엔드포인트를 사용하여 해당 컨테이너 내의 모든 태스크를 처리합니다.
B	다중 컨테이너 엔드포인트를 사용하고 단일 모델을 여러 컨테이너에 배포합니다.
C	다중 컨테이너 엔드포인트를 사용하고 각 태스크를 별도의 컨테이너에 배포합니다.
D	단일 컨테이너 엔드포인트를 사용하고 이 컨테이너에 여러 모델을 배포합니다.

42

DO NOT COPY
ssleeeee1@gmail.com

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



정답은 다음과 같습니다.

- C: Amazon SageMaker AI에서 다중 컨테이너 엔드포인트를 가장 적절하게 사용하는 방식에 해당되는 답안입니다. 각 태스크(데이터 사전 처리, 모델 훈련, 모델 제공)를 별도의 컨테이너에 배포하면 유동적인 모듈식 아키텍처를 구축할 수 있습니다.

오답은 배포 방식과 관련하여 흔히 잘못 알려져 있는 내용에 해당됩니다.

- 복잡한 워크플로를 진행할 때 단일 컨테이너 방식(보기 A와 D)을 사용하면 다중 컨테이너 배포처럼 워크플로를 유동적으로 진행하고 확장할 수가 없습니다.
- 여러 컨테이너에 모델 하나를 배포하는 방식(보기 B)은 효율적이지 않으며 다중 컨테이너 아키텍처의 이점도 제공되지 않습니다.

DO NOT COPY
ssleeeee1@gmail.com

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

DO NOT COPY
ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

정답은 A, Amazon SageMaker AI 엔드포인트입니다.

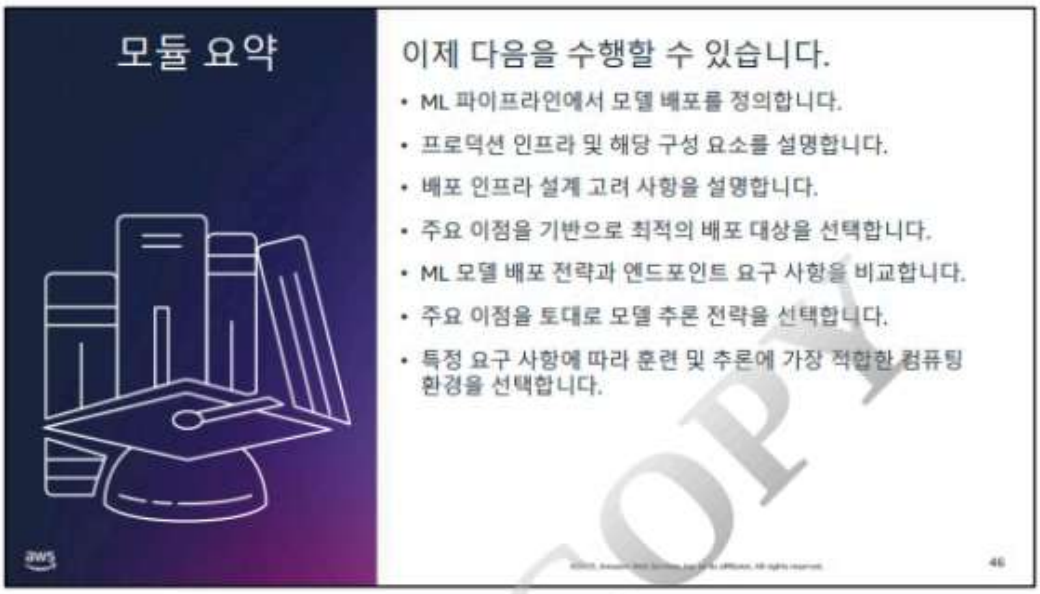
설명:

A. (정답) 지연 시간이 짧으며 운영 오버헤드를 최소화할 수 있는 고성능 완전관리형 솔루션인 Amazon SageMaker AI 엔드포인트가 이 시나리오에 가장 적합한 대상입니다.
B. (오답) Amazon EKS도 확장 및 사용자 지정이 가능하기는 하지만 SageMaker AI 엔드포인트에 비해 운영 오버헤드가 많이 발생합니다.
C. (오답) Amazon ECS도 Amazon EKS와 마찬가지로 이 사용 사례에서 SageMaker AI 엔드포인트에 비해 관리 작업을 더 많이 처리해야 합니다.
D. (오답) Lambda는 서버리스 서비스이므로 오버헤드는 적지만 이 시나리오에서 필요한 높은 성능을 일관되게 제공하지 못할 수도 있습니다(특히 대규모 모델의 경우).

이 시나리오에서는 성능이 우수하고 쉽게 관리할 수 있으며 빈번한 업데이트도 지원되는 SageMaker AI 엔드포인트가 가장 적합합니다.

DO NOT COPY ssleeeee1@gmail.com

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



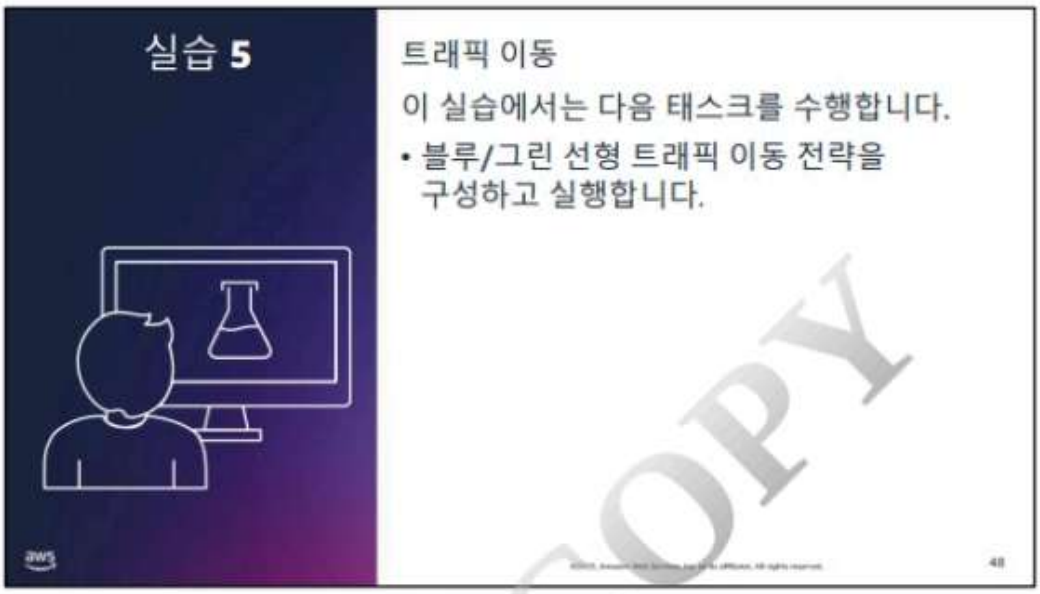
DO NOT COPY
ssleeeee1@gmail.com

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

DO NOT COPY
ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



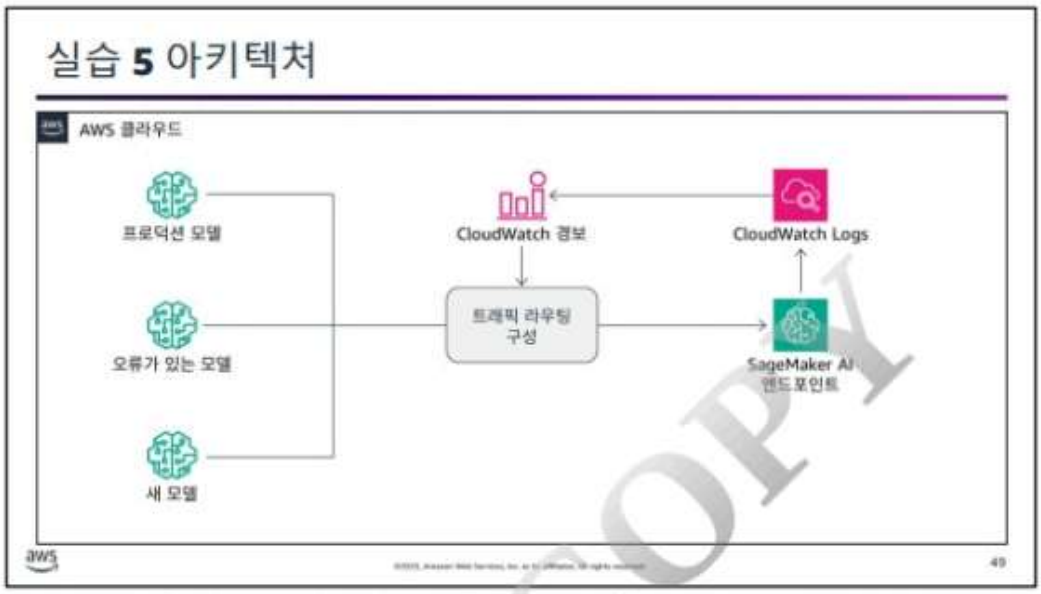
본 실습을 마치면 다음을 수행할 수 있습니다.

- 블루/그린 선형 트래픽 이동 전략을 구성하고 실행합니다.

DO NOT COPY
ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



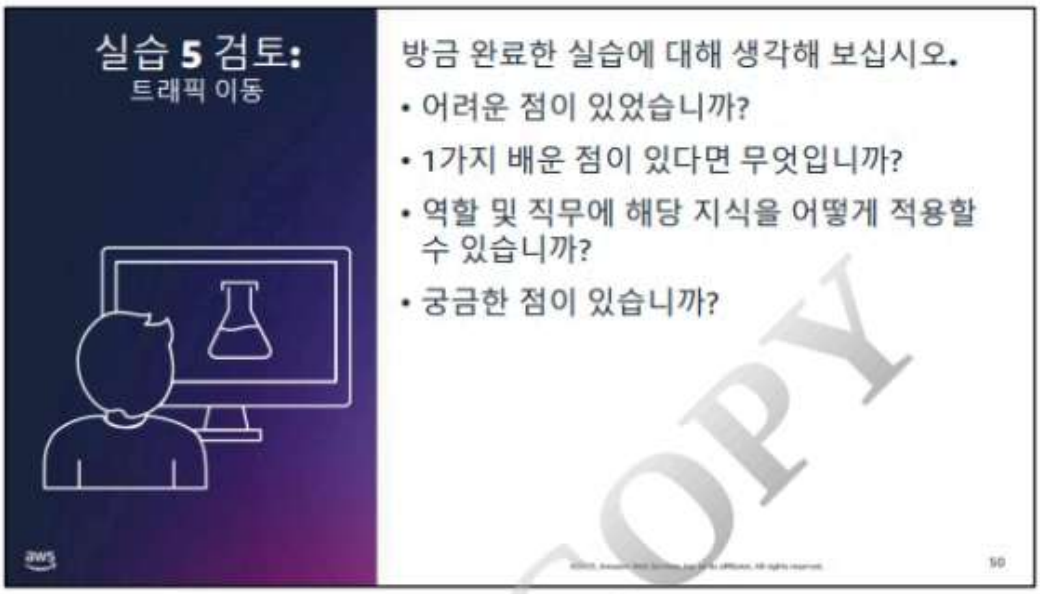
AnyCompany Consulting에서는 전복 회귀 모델을 개발하고 있습니다. 이전 실습에서 테스트한 새 모델은 현재 프로덕션 모델보다 성능이 더 우수한 모델입니다. 이제 회사는 블루/그린 배포를 사용하여 안정적인 방식으로 새 모델을 프로덕션 엔드포인트에 배포하려고 합니다. 이 태스크에서는 선형 트래픽 이동 전략을 사용하여 블루/그린 배포를 구성해 프로덕션 모델에서 새 후보 모델로 트래픽을 이동합니다. 또한 자동 롤백 배포 가드레일 기능이 활성화되도록 CloudWatch 경보를 설정합니다.

이 실습에서는 프로덕션 엔드포인트와 CloudWatch 경보를 생성하고, 선형 트래픽 이동 전략을 구성하고, 트래픽 이동 전략을 구현하여 특정 모델의 트래픽을 다른 모델로 이동합니다. 이 전략에서는 프로덕션 모델, 오류가 있는 모델 및 새 모델이 트래픽 라우팅 구성을 사용하여 SageMaker AI 엔드포인트와 연결됩니다. SageMaker AI 엔드포인트는 CloudWatch 로그에 로그 데이터를 전송합니다. 트래픽 라우팅 구성에는 트래픽의 동작에 따라 트래픽 라우팅 구성에 영향을 미칠 수 있는 CloudWatch 경보가 있습니다.

DO NOT COPY
ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



DO NOT COPY
ssleeeee1@gmail.com

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

이 모듈에서는 기계 학습 엔지니어링 시에 꼭 필요한 요소인 Amazon Web Services(AWS) 기계 학습(ML) 리소스를 보호하는 방법을 중점적으로 살펴봅니다. 구체적으로는 AWS에서 ML 워크플로, 데이터 및 모델을 보호하기 위해 안정적인 보안 조치를 구현하는 방법을 알아봅니다.

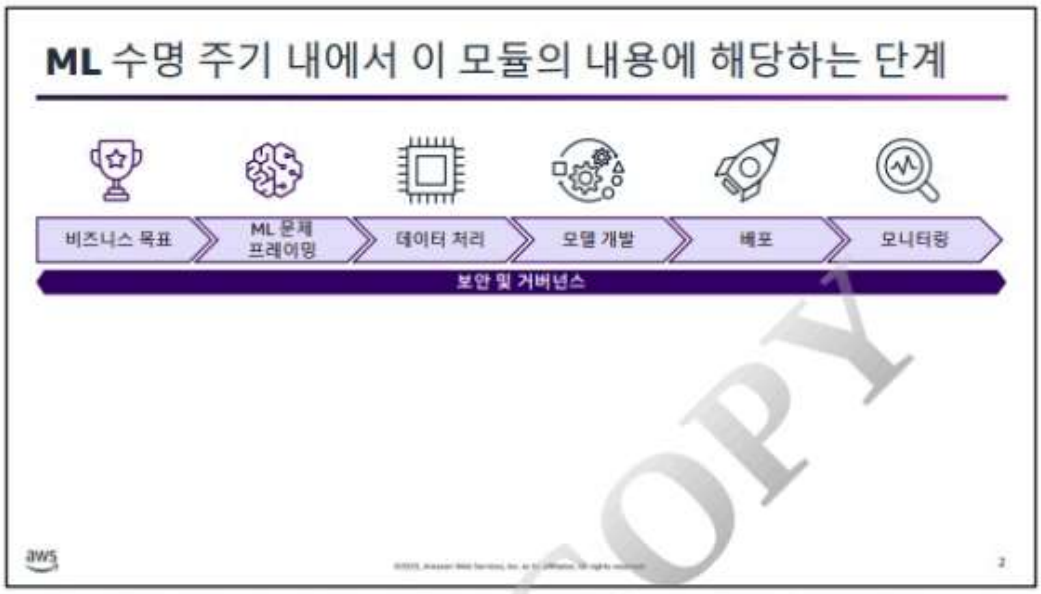
특히 기계 학습에서 민감한 데이터나 업무상 중요한 모델을 사용할 때는 보안이 매우 중요합니다. 이 모듈에서는 AWS의 ML 리소스 보호를 위한 모범 사례를 적용하는 방법과 관련된 지식을 습득할 수 있습니다. 그러면 발생 가능한 위협으로부터 리소스를 보호하고 관련 규정을 철저하게 준수할 수 있습니다.

이 모듈에서는 ML 인프라의 보안 강화를 위해 사용할 수 있는 여러 AWS 서비스 및 기능에 대해 살펴봅니다. 이 모듈을 마치면 AWS에서 진행하는 ML 프로젝트용 보안 컨트롤을 안정적으로 구현할 수 있게 됩니다.

DO NOT COPY
ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



보안은 전체 **기계 학습 수명 주기**에서 중요한 고려 사항입니다. 이 모듈에서는 AWS에서 진행하는 ML 개발 및 배포의 여러 단계에서 보안 조치를 구현하는 방법을 알아봅니다.

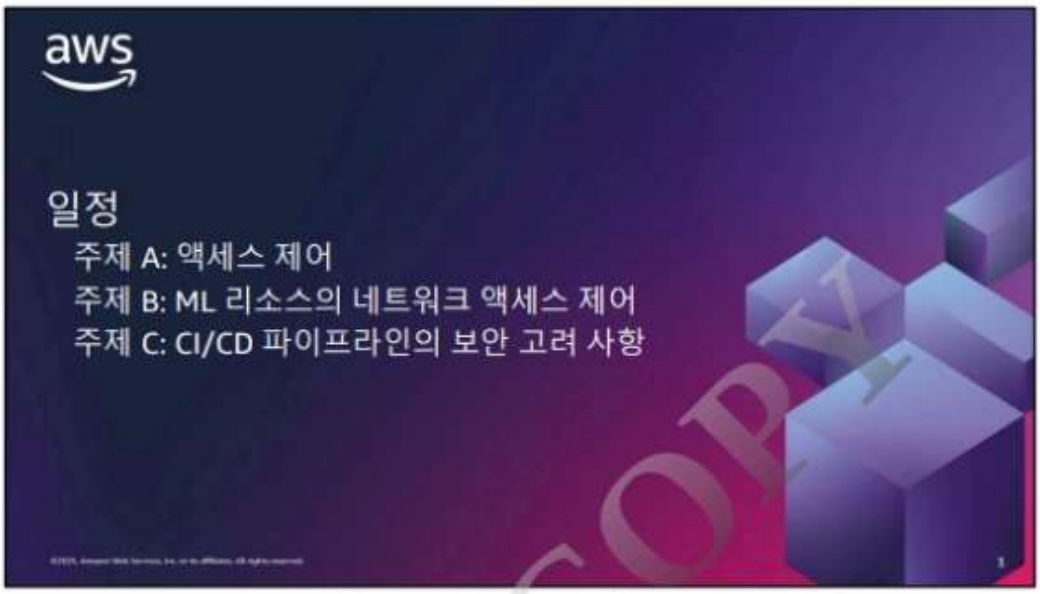
- 데이터 수집 및 준비: ML 모델용 데이터를 수집하고 사전 처리할 때는 데이터를 안전하게 저장하고 액세스를 제어할 수 있어야 합니다.
- 모델 개발 및 훈련: 개발 중에는 훈련 환경을 보호하고 모델 무결성을 유지해야 합니다.
- 모델 배포 및 추론: 엔드포인트를 보호하고 배포된 모델에 대한 액세스를 관리하면 모델 무단 사용 또는 변조를 방지할 수 있습니다.
- 모니터링 및 유지 관리: ML 시스템의 무결성과 성능을 장기적으로 유지하려면 지속적인 보안 모니터링과 업데이트를 수행해야 합니다.

ML 수명 주기 전반에 보안 사례를 통합하면 AWS에서 기계 학습 프로젝트를 진행하는 데 사용할 안정적이며 신뢰할 수 있는 환경을 구축할 수 있습니다.

DO NOT COPY
ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



이 모듈에서는 AWS ML 리소스를 보호하려면 꼭 알아 두어야 하는 다음의 3가지 주요 주제에 대해 다룹니다.

주제 A: 액세스 제어

AWS Identity and Access Management(IAM)를 사용하여 ML 리소스에 대한 액세스를 제어하는 방법을 알아봅니다. 여기서 액세스 제어에는 ML 워크플로와 관련된 IAM 정책, 역할, 사용자를 생성하고 관리하는 작업이 포함됩니다.

주제 B: ML 리소스의 네트워크 액세스 제어

Virtual Private Cloud(VPC), 보안 그룹, 네트워크 액세스 제어 목록(네트워크 ACL)을 사용하여 ML 인프라용 네트워크 수준 보안을 구현하는 방법을 살펴봅니다.

주제 C: CI/CD 파이프라인의 보안 고려 사항

ML 프로젝트용 지속적 통합 및 지속적 전달(CI/CD) 파이프라인을 보호하는 모범 사례를 파악합니다. 이러한 모범 사례를 준수하면 모델 개발 및 배포 프로세스의 보안과 무결성을 유지할 수 있습니다.

이 주제에서는 AWS에서 ML 리소스를 보호할 수 있는 포괄적인 방식을 소개하는 동시에 자격 증명 관리, 네트워크 보안 및 안전한 개발 관행에 대해 설명합니다.

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

네트워크 액세스 제어는 AWS의 ML 리소스를 보호하려면 꼭 필요한 기능입니다. 이러한 기능을 사용하면 ML 워크로드를 격리하고 무단 액세스나 위협 발생 가능성으로부터 보호할 수 있습니다.

이 섹션에서는 AWS에서 제공되는 다음과 같은 여러 네트워크 보안 메커니즘에 대해 알아봅니다.

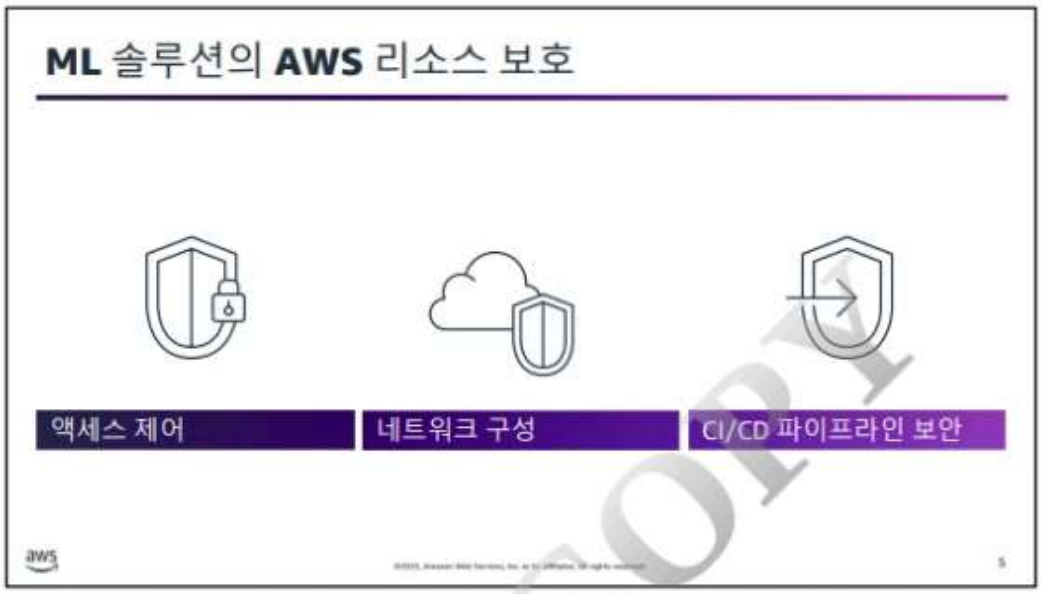
- VPC: ML 리소스용 격리 네트워크 환경을 생성합니다.
- VPC 엔드포인트: AWS 네트워크를 벗어나지 않고 AWS 서비스에 안전하게 액세스할 수 있습니다.
- 보안 그룹: 인바운드 및 아웃바운드 트래픽을 제어하는 가상 방화벽 역할을 합니다.
- 네트워크 ACL: 서브넷 수준에서 네트워크 보안 계층을 추가로 제공합니다.

이러한 제어 기능을 정확하게 파악하여 구현하면 ML 프로젝트용으로 안전한 네트워크 아키텍처를 생성할 수 있습니다. 그러면 민감한 데이터와 모델을 무단 액세스로부터 보호할 수 있습니다.

DO NOT COPY
ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



기계 학습 솔루션에서 AWS 리소스를 보호하려면 안정적인 액세스 제어 전략을 추진해야 합니다. IAM을 활용하면 리소스 액세스 가능성 및 허용되는 작업을 정밀하게 관리할 수 있습니다. 그리고 역할 기반 액세스 제어를 통해 최소 권한의 원칙을 구현하면 ML 워크로드 전반에서 효율적인 보안 경계를 생성하여 공격 대상 영역 발생 가능성을 최소화할 수 있습니다.

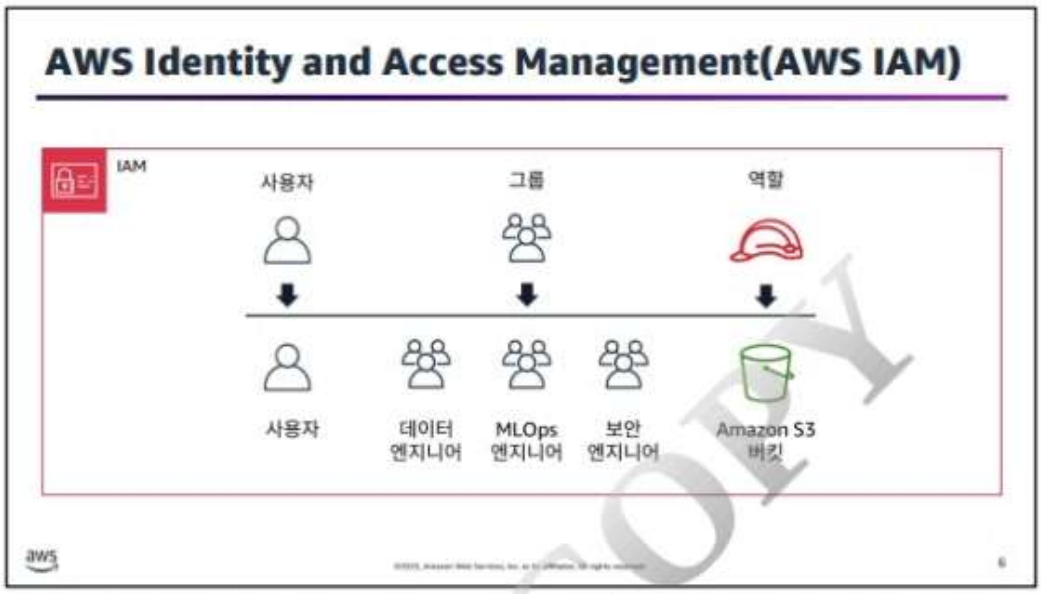
AWS ML 보호를 위한 두 번째 주요 기능은 네트워크 보안입니다. 조직은 VPC, 보안 그룹 및 네트워크 ACL을 사용하여 ML 워크로드용 격리 환경을 생성해 트래픽 패턴을 제어할 수 있습니다. 이렇게 네트워크 수준 보안을 설정하면 자격 증명 기반 보호 방식을 보완하는 데 꼭 필요한 경계를 구축할 수 있습니다.

ML 보호를 위한 마지막 구성 요소는 보호되는 CI/CD 파이프라인을 통해 ML 개발 수명 주기를 보호하는 것입니다. 빌드, 테스트 및 배포 프로세스 전반에 걸쳐 엄격한 보안 관행을 구현하면 전체 ML 워크플로의 무결성을 유지할 수 있습니다. 이러한 3가지 보안 차원(액세스 제어, 네트워크 구성 및 CI/CD 파이프라인 보안)을 구현하려면 ML 프로젝트 진행 과정을 지속적으로 모니터링하고 조정하여 새롭게 등장하는 위협으로부터 ML 리소스를 효율적으로 보호해야 합니다.

DO NOT COPY
ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



IAM은 ML 리소스를 보호하는 중요한 서비스입니다. 이 서비스를 통해 AWS 서비스와 리소스에 대한 액세스를 안전하게 관리할 수 있습니다. IAM의 주요 구성 요소는 다음과 같습니다.

- **사용자**: AWS 액세스 권한이 필요한 사용자나 서비스의 개별 자격 증명입니다. 사용자의 예로는 데이터 과학자, ML 엔지니어 등이 있습니다.
- **그룹**: IAM 사용자 컬렉션입니다. 그룹을 사용해 여러 사용자의 권한을 관리할 수 있습니다.
- **역할**: 신뢰할 수 있는 엔터티에 특정 권한을 부여하는 데 사용되는 자격 증명입니다. 훈련 데이터가 포함된 Amazon Simple Storage Service(Amazon S3) 버킷에 Amazon Elastic Compute Cloud(Amazon EC2) 인스턴스가 액세스하도록 허용하는 역할을 예로 들 수 있습니다.

ML 워크플로에서는 태스크별로 각기 다른 역할을 생성할 수 있습니다.

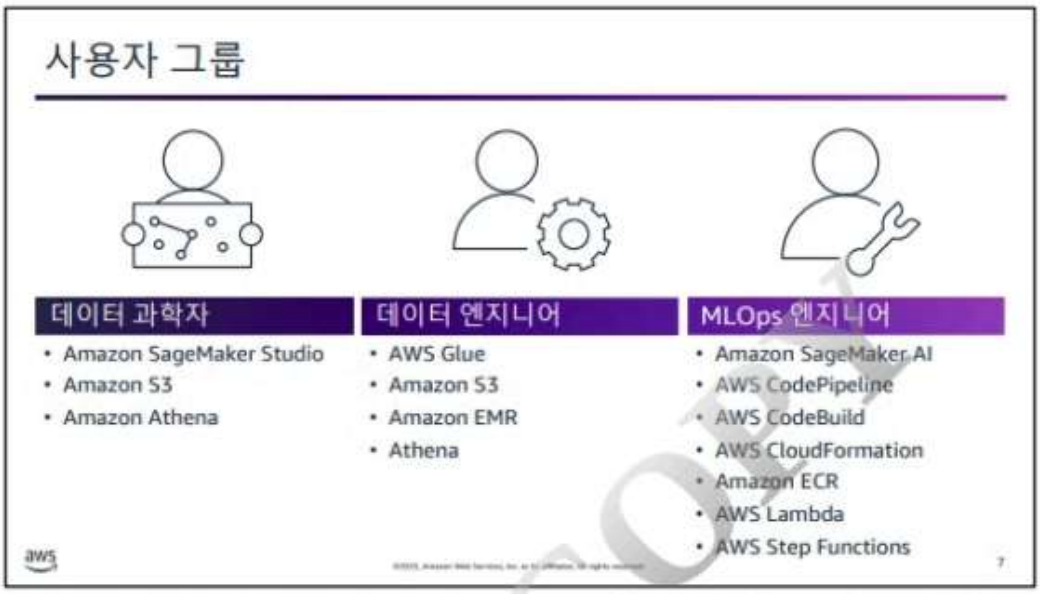
- Amazon S3의 훈련 데이터 액세스용 역할
- Amazon SageMaker AI 훈련 작업 시작용 역할
- SageMaker AI 엔드포인트로의 모델 배포용 역할

IAM 정책을 생성할 때는 항상 **최소 권한의 원칙**에 따라 구체적인 ML 태스크에 필요한 권한만 부여해야 합니다. 이 방식을 사용하면 ML 프로젝트에서 발생 가능한 보안 위험을 최소화할 수 있습니다.

DO NOT COPY ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



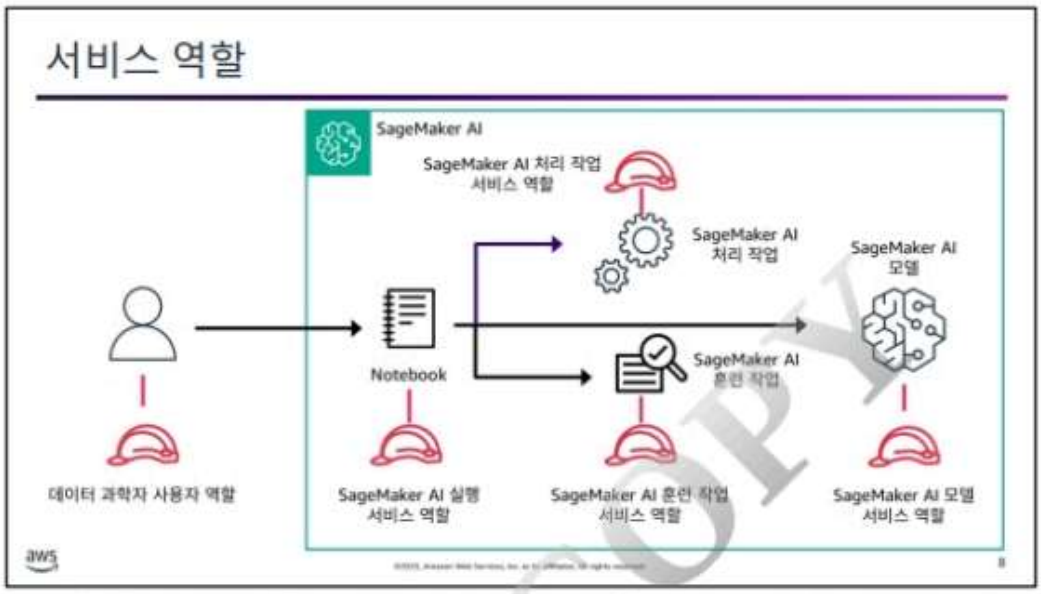
ML 프로젝트를 진행할 때는 팀에 소속된 전문 역할의 AWS 서비스 액세스를 체계적이며 면밀하게 관리해야 합니다. 데이터 과학자와 ML 엔지니어는 주로 모델 개발용 Amazon SageMaker Studio를 사용해야 하며 데이터 저장용 Amazon S3 및 쿼리 기능이 제공되는 Amazon Athena도 사용합니다. 이처럼 주로 사용하는 서비스에 대한 액세스 권한만 제공되면 불필요한 서비스가 표시되지 않으므로 모델 생성 작업만 집중적으로 수행할 수 있습니다.

반면 데이터 엔지니어에게는 데이터 파이프라인 관리를 중심으로 하는 다른 액세스 프로파일이 필요합니다. 데이터 엔지니어의 도구 키트에는 대개 추출, 변환, 로드(ETL) 작업용 AWS Glue, 데이터 레이크 관리용 Amazon S3, 분산 처리용 Amazon EMR, 분석 쿼리용 Athena 등이 포함됩니다. 이러한 서비스 사용을 위한 권한을 제공하면 데이터 엔지니어가 데이터를 효율적으로 준비 및 변환할 수 있으며, 모델 배포 기능 사용 권한은 부여할 필요가 없습니다.

ML 워크플로의 운영 측면을 관리하는 MLOps 엔지니어는 이러한 역할 중 가장 많은 서비스에 액세스해야 합니다. MLOps 엔지니어는 업무 과정에서 배포와 모니터링용 SageMaker AI, CI/CD 구현용 AWS CodePipeline 및 AWS CodeBuild, 인프라 관리용 AWS CloudFormation, 컨테이너 레지스트리 액세스용 Amazon Elastic Container Registry(Amazon ECR), 오케스트레이션용 AWS Lambda 및 AWS Step Functions를 사용해야 합니다. IAM을 통해 이러한 역할 기반 액세스 패턴을 구현하면 최소 권한의 원칙을 준수하여 자격 증명 손상 가능성으로 인한 보안 위험을 대폭 줄일 수 있습니다.

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



AWS의 **서비스 역할**은 사용자 대신 작업을 수행할 권한을 AWS 서비스에 부여하는 IAM 역할입니다. ML 워크플로의 보안을 유지하려면 서비스 역할을 파악하여 적절하게 구성해야 합니다. 주요 SageMaker AI 서비스 역할이 아래에 나와 있습니다.

- SageMaker AI 실행 서비스 역할: SageMaker AI가 데이터 검색 및 저장용 Amazon S3와 같은 타 AWS 서비스에 사용자 대신 액세스하는 데 사용하는 역할입니다.
- SageMaker AI 처리 작업 서비스 역할: SageMaker AI 처리 작업이 필요한 리소스에 액세스하는 권한을 부여하는 역할입니다.
- SageMaker AI 훈련 작업 서비스 역할: SageMaker AI가 Amazon S3의 훈련 데이터 등 모델 훈련에 필요한 리소스에 액세스하는 데 사용하는 역할입니다.
- SageMaker AI 모델 서비스 역할: SageMaker AI에 모델을 배포하고 엔드포인트를 생성하는 권한을 제공하는 역할입니다.

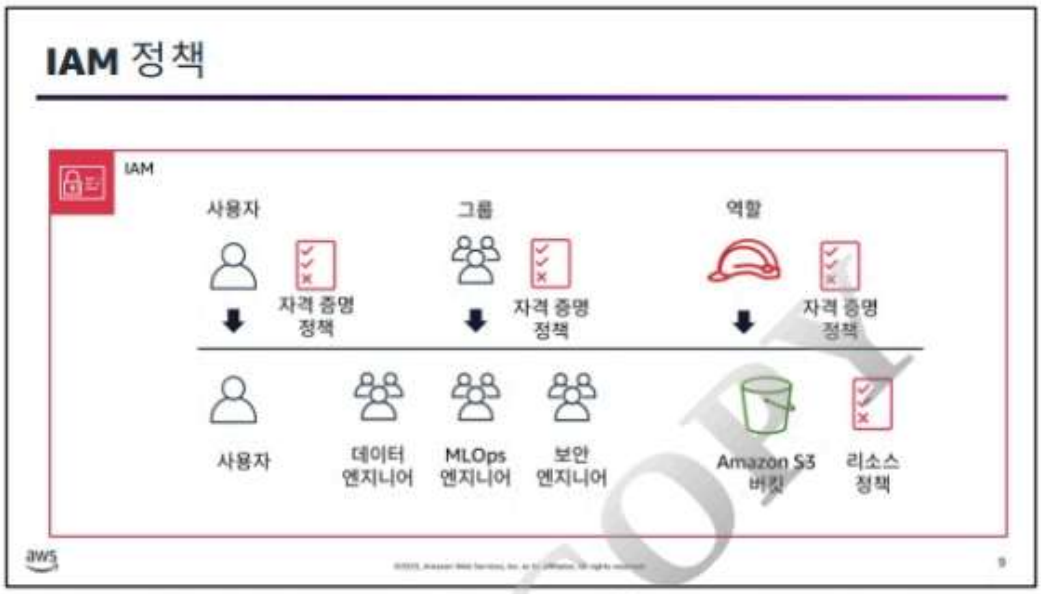
구체적인 태스크를 수행하는 데 필요한 최소 권한을 사용하여 이러한 각 역할을 구성해야 합니다. 이 방식을 사용하면 역할 손상 시 발생 가능한 영향을 제한함으로써 보안을 강화할 수 있습니다.

ML 워크플로를 진행하는 과정에서 안정적인 보안 태세를 유지하려면 이러한 서비스 역할을 정기적으로 검토 및 업데이트해야 합니다.

DO NOT COPY ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



IAM 정책은 AWS의 권한을 정의하는 JSON 문서입니다. 액세스를 세부적으로 제어하는 IAM 정책은 ML 리소스를 보호하려면 반드시 사용해야 하는 중요한 요소입니다. IAM 정책에는 2가지 주요 유형이 있습니다.

- 자격 증명 기반 정책: IAM 사용자, 그룹 또는 역할에 연결되는 자격 증명 기반 정책은 자격 증명이 각 리소스에 대해 수행할 수 있는 작업을 지정합니다. 데이터 과학자에게 SageMaker AI 훈련 작업 시작 권한은 부여하되 모델 배포 권한은 부여하지 않는 정책을 예로 들 수 있습니다.
- 리소스 기반 정책: S3 버킷 등의 리소스에 직접 연결되는 리소스 기반 정책은 리소스에 액세스할 수 있는 사용자 및 해당 사용자가 수행할 수 있는 작업을 정의합니다. 특정 SageMaker AI 역할이 훈련 데이터 버킷에서 데이터를 읽도록 허용하는 정책을 예로 들 수 있습니다.

ML 워크플로용 IAM 정책을 생성할 때는 다음 지침을 준수해야 합니다.

- 최소 권한의 원칙에 따라 필요한 권한만 부여합니다.
- 조건 키를 사용하여 보안을 강화합니다. 가령 암호화를 적용하거나 특정 VPC 액세스를 제한할 수 있습니다.
- 정책을 정기적으로 검토 및 감사하여 정책이 계속 적절하며 안전한 상태인지를 확인합니다.

리소스 기반 정책은 SageMaker AI에서 직접 지원되는 것은 아니지만 훈련 데이터 또는 모델 아티팩트가 포함된 S3 버킷 등의 관련 리소스에서는 사용할 수 있습니다.

AWS에서 ML 프로젝트의 보안과 무결성을 유지하려면 IAM 정책을 적절하게 구성해야 합니다.

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

ML용 IAM 정책

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": [
                "s3:GetObject",
                "s3:PutObject"
            ],
            "Resource": [
                "arn:aws:s3:::ml-data
                 -bucket/*"
            ]
        },
        {
            "Effect": "Allow",
            "Action": [
              "sagemaker:CreateTrainingJob",
              "sagemaker:DescribeTrainingJob",
              "sagemaker:StopTrainingJob"
            ],
            "Resource": "*"
        },
```

```
            "Effect": "Allow",
            "Action": ["cloudwatch:PutMetricData"],
            "Resource": "*",
            "Condition": {
                "StringEquals": {
                    "cloudwatch:namespace":
                    "SageMaker"
                }
            }
        },
        {
            "Effect": "Allow",
            "Action": [
                "logs:CreateLogGroup",
                "logs:CreateLogStream",
                "logs:PutLogEvents"
            ],
            "Resource": [
                "arn:aws:logs:*:*:
                 log-group:/aws/sagemaker/*" ]
        }
    ]
}
```

10

|강사 참고 사항:
슬라이드의 예제에는 기계 학습 워크플로용으로 조정된 자격 증명 기반 IAM 정책이 나와 있습니다. 이 정책은 Amazon S3 액세스, SageMaker AI 훈련 작업 관리, CloudWatch 지표 및 Amazon CloudWatch Logs와 관련된 구체적인 권한을 부여합니다.

그리고 최소 권한의 원칙에 따라 이 특정 ML 사용 사례에 필요한 권한만 부여합니다. 가령 이 정책은 특정 S3 버킷에서만 `s3:GetObject` 및 `s3:PutObject` 작업을 허용하며 훈련 작업을 생성, 설명, 중지하는 SageMaker AI 작업은 제한합니다.

참고로 슬라이드의 코드는 간소화된 예제이며 실제 시나리오에서는 구체적인 요구 사항에 따라 Amazon Resource Name(ARN)을 조정하거나 권한을 더 추가하거나 조건을 조정해야 할 수도 있습니다. 항상 IAM 정책을 철저하게 검토 및 테스트하여 정책이 보안을 유지하면서 적절한 액세스 권한을 제공하는지를 확인하십시오.

DO NOT COPY
ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

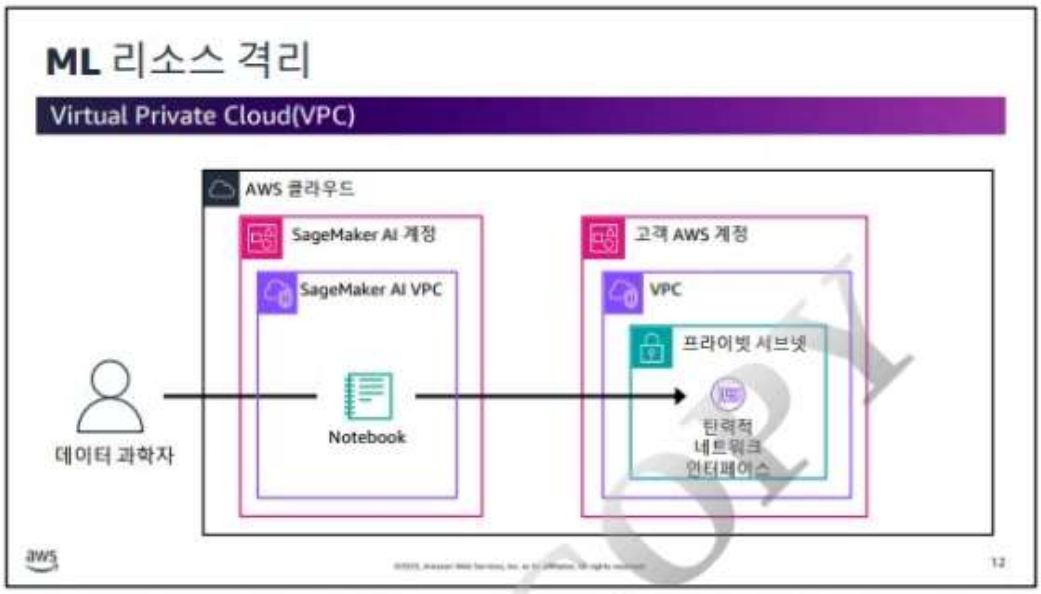
네트워크 액세스 제어는 AWS의 기계 학습 리소스를 보호 및 격리하려면 꼭 필요한 기능입니다. 이 섹션에서는 무단 액세스 및 위협 발생 가능성으로부터 ML 시스템을 보호하는 데 사용할 수 있는 여러 메커니즘을 살펴봅니다.

먼저 VPC에 대해 알아봅니다. VPC에서는 ML 리소스를 시작할 수 있는 AWS 클라우드의 논리적으로 격리된 섹션이 제공됩니다. 그리고 Amazon VPC 엔드포인트에 대해서도 살펴봅니다. VPC 엔드포인트를 사용하면 퍼블릭 인터넷에 트래픽을 노출하지 않고 지원되는 AWS 서비스로의 프라이빗 연결을 설정할 수 있습니다.

그 후에는 NAT 게이트웨이, 보안 그룹, 네트워크 ACL 및 AWS Network Firewall을 사용하여 ML 인프라에 여러 보안 계층을 적용하는 방법도 살펴봅니다. 이러한 도구는 서로 연동되어 ML 리소스에서 송수신되는 네트워크 트래픽을 심층 방어하며 세부적으로 제어합니다.

DO NOT COPY ssleeeee1@gmail.com

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



VPC는 AWS 내에서 ML 리소스를 격리할 때 기본적으로 사용하는 도구입니다. 구체적으로 설명하자면 VPC는 가상 네트워크를 직접 정의하고 AWS 리소스를 시작할 수 있는 AWS 클라우드의 논리적으로 격리된 섹션입니다.

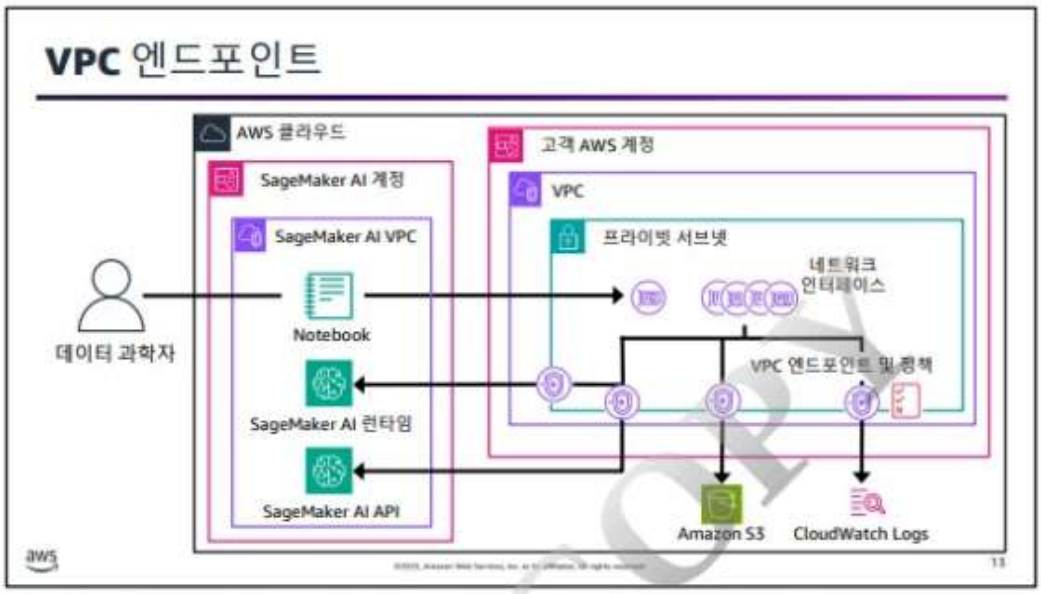
사기 탐지 모델 등의 민감한 ML 워크로드를 진행할 때는 인터넷에 직접 액세스할 수 없는 **프라이빗 서브넷** 사용 VPC를 생성할 수 있습니다. 이처럼 워크로드를 격리하면 AWS 클라우드 내에서만 네트워크 통신이 진행되므로 민감한 데이터와 모델의 공격 대상 영역이 축소됩니다.

탄력적 네트워크 인터페이스를 통해 VPC와 SageMaker AI를 통합할 수 있습니다. 이 경우 VPC에서 시작하는 SageMaker Studio 또는 Notebook 인스턴스는 이러한 네트워크 인터페이스를 통해 SageMaker AI 서비스와 통신합니다. 이러한 설정을 통해 ML 개발 환경에 대한 네트워크 액세스를 제어할 수 있으며, 격리된 네트워크 내에서만 데이터 처리와 모델 훈련을 진행할 수 있습니다.

VPC와 프라이빗 서브넷을 사용하면 ML 워크플로용으로 안전한 기반 환경을 구축할 수 있습니다. 이 환경에서는 민감한 데이터가 무단 액세스로부터 보호되는 상태로 유지됩니다.

DO NOT COPY ssleeeee1@gmail.com

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



VPC 엔드포인트를 사용하면 인터넷 게이트웨이, NAT 디바이스, VPN 연결 또는 AWS Direct Connect를 사용하지 않고도 지원되는 AWS 서비스에 VPC를 비공개로 연결할 수 있습니다. 필요한 AWS 서비스 액세스는 허용하면서 ML 리소스의 보안과 격리 상태를 유지하려면 이 기능을 사용해야 합니다.

ML 워크플로에서는 VPC 엔드포인트를 사용하여 프라이빗 서브넷 내에서 SageMaker AI API, SageMaker AI 런타임, CloudWatch Logs, Amazon S3 등의 서비스에 안전하게 액세스할 수 있습니다. 이 설정을 적용하면 VPC와 이러한 서비스 간에 전송되는 트래픽을 보호할 수 있습니다. 해당 트래픽이 퍼블릭 인터넷을 통해 전송되지 않으며 AWS 네트워크 내에 유지되기 때문입니다.

엔드포인트 정책을 사용하면 엔드포인트에 액세스할 수 있는 사용자 및 해당 사용자가 수행할 수 있는 작업을 제어함으로써 보안을 더욱 강화할 수 있습니다. 예를 들어 SageMaker AI 런타임 엔드포인트를 통해 배포된 ML 모델을 특정 IAM 역할만 호출할 수 있도록 허용하는 정책을 생성할 수 있습니다.

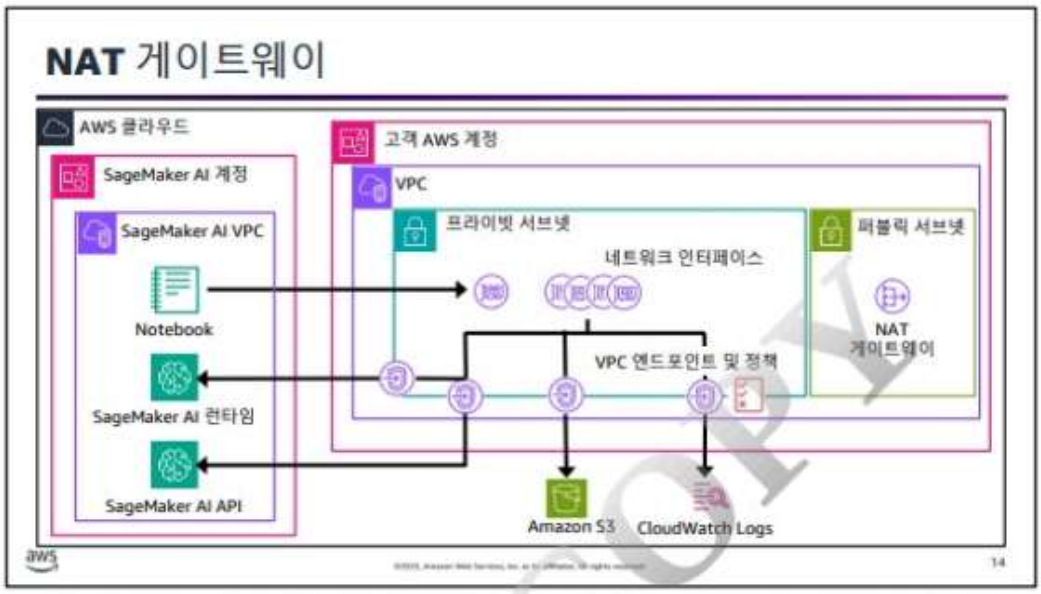
VPC 엔드포인트 및 관련 정책 사용 시의 이점은 다음과 같습니다.
- ML 데이터 및 모델의 보안 강화
- 데이터 전송 비용 절감(트래픽이 AWS 네트워크 내에 유지됨)
- 네트워크 아키텍처 간소화(인터넷 게이트웨이 또는 NAT 디바이스 불필요)

엔드포인트 정책을 정기적으로 검토 및 업데이트하여 정책이 보안 요구 사항 및 최소 권한의 원칙에 부합하는지 확인하십시오.

DO NOT COPY ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



NAT 게이트웨이를 사용하면 인터넷의 인바운드 연결은 차단하면서 프라이빗 서브넷의 인스턴스가 인터넷이나 기타 AWS 서비스에 액세스하도록 할 수 있습니다. 그러므로 격리 상태는 유지하되 아웃바운드 인터넷 액세스도 어느 정도는 필요한 ML 워크플로에서 특히 유용합니다.

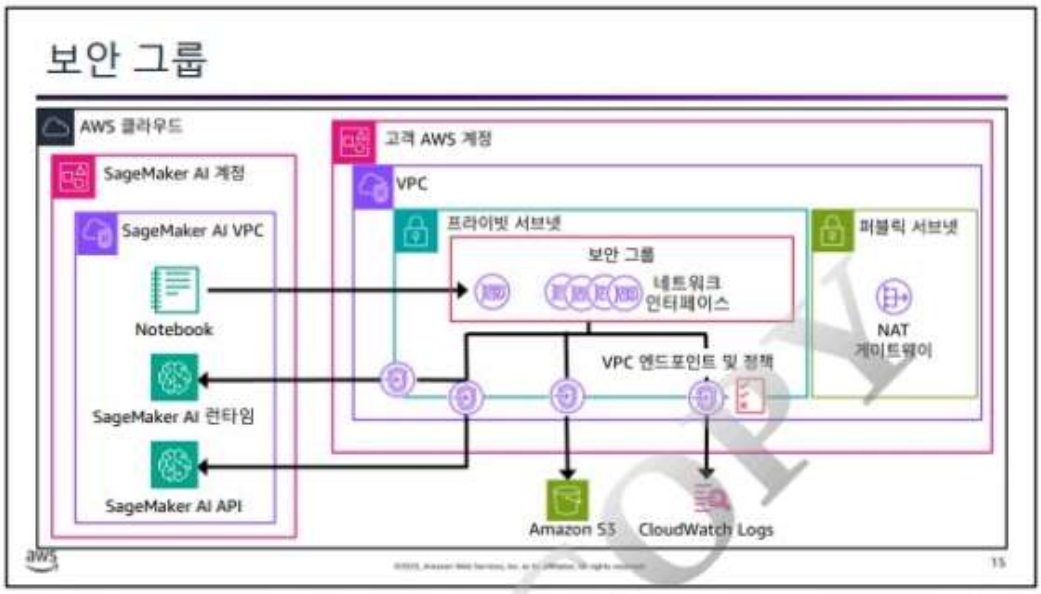
ML 워크플로에서 NAT 게이트웨이를 흔히 사용하는 사용 사례는 다음과 같습니다.
- 퍼블릭 리포지토리에서 업데이트 또는 종속성 다운로드
- 모델 훈련용으로 외부 데이터세트 가져오기
- 외부 서비스에 알림 또는 로그 전송

콘솔을 사용하여 Studio 도메인을 생성하려면 다음 단계를 수행합니다.
1. Amazon SageMaker AI 콘솔을 열고 SageMaker Studio를 선택합니다.
2. 메시지가 표시되면 Standard 설정을 선택합니다.
3. Authentication method에서 AWS Identity and Access Management (IAM)를 선택합니다.
4. Execution role에서 Notebook IAM 역할(기본값: studiovpc-notebook-role)을 선택합니다.
5. Network 섹션에서 다음을 수행합니다.
- VPC에서 원하는 VPC를 선택합니다(기본값: studiovpc-vpc).
- Subnet에서 서브넷을 선택합니다(기본값: studiovpc-private-subnet).

중요: 보안 정책을 준수하려면 인터넷 액세스가 필요한 리소스를 면밀하게 고려하고 트래픽을 모니터링해야 합니다.

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



EC2 인스턴스와 기타 AWS 리소스용 가상 방화벽 역할을 하는 **보안 그룹**은 인스턴스 수준에서 인바운드 및 아웃바운드 트래픽을 제어합니다. ML 워크플로에서는 리소스를 보호하고 승인된 트래픽만 ML 인스턴스 및 엔드포인트로 전송되도록 하려면 보안 그룹을 사용해야 합니다.

보안 그룹 관련 주요 정보가 아래에 나와 있습니다.
- 인스턴스 수준에서 작동하여 트래픽을 '스테이트풀' 방식으로 필터링합니다.
- 허용 규칙을 지정할 수 있지만, 거부 규칙은 지정할 수 없습니다.
- 모든 규칙은 트래픽 허용 여부를 결정하기 전에 평가됩니다.

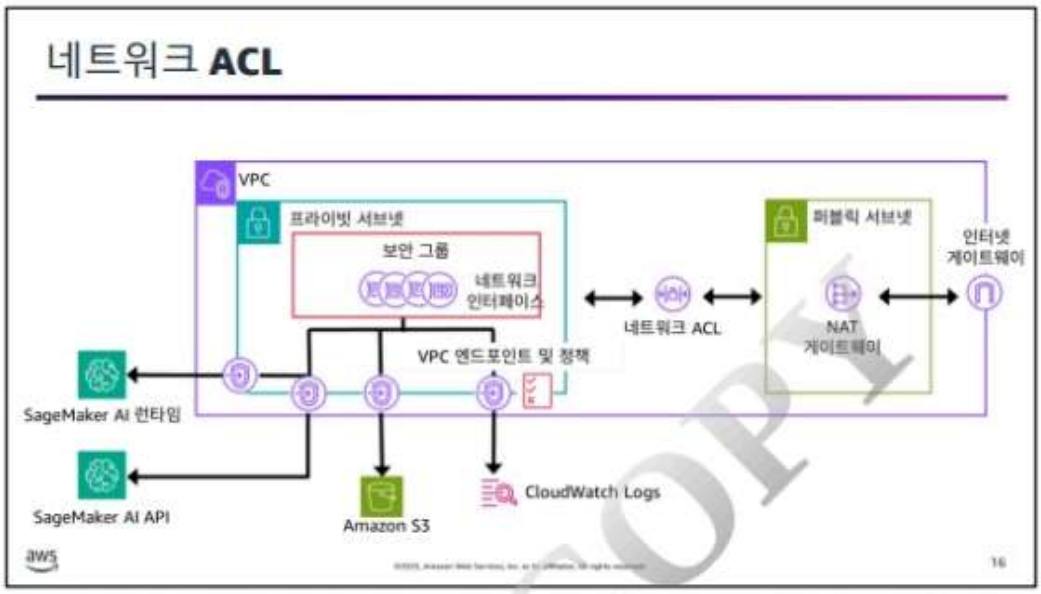
ML 리소스용 보안 그룹은 다음과 같이 구성할 수 있습니다.
- 신뢰할 수 있는 IP 범위의 특정 포트(예: Jupyter Notebook의 경우 포트 8888)에서 전송되는 인바운드 트래픽만 허용
- ML 워크플로에 필요한 특정 AWS 서비스 또는 엔드포인트로 전송되는 아웃바운드 트래픽 허용
- ML 파이프라인의 각 구성 요소(예: 데이터 사전 처리, 모델 훈련, 추론) 간 통신 허용

가령 SageMaker AI Notebook 인스턴스를 보호하려는 경우 다음과 같은 보안 그룹을 생성할 수 있습니다.
- 기업 네트워크 IP 범위에서 전송되는 인바운드 HTTP 트래픽(포트 443)만 보호
- Amazon S3 및 SageMaker AI API 엔드포인트로 전송되는 아웃바운드 트래픽 허용
- 기타 모든 인바운드 및 아웃바운드 트래픽 차단

보안 그룹을 신중하게 구성하면 ML 리소스의 안정적인 '보안 경계'를 구축함으로써 무단 액세스 또는 데이터 유출 위험을 최소화할 수 있습니다.

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



네트워크 ACL을 사용하면 VPC에 네트워크 보안 계층을 추가할 수 있습니다. 인스턴스 수준에서 작동하는 보안 그룹과 달리 서브넷 수준에서 작동하는 네트워크 ACL은 서브넷에서 송수신되는 트래픽을 제어하는 방화역 역할을 합니다.

네트워크 ACL의 주요 특성은 다음과 같습니다.
- 스테이트리스 방식이므로 인바운드 규칙과 아웃바운드 규칙을 모두 지정해야 합니다.
- 규칙은 가장 낮은 번호부터 순서대로 평가됩니다.
- 허용 및 거부 규칙을 모두 지정할 수 있습니다.

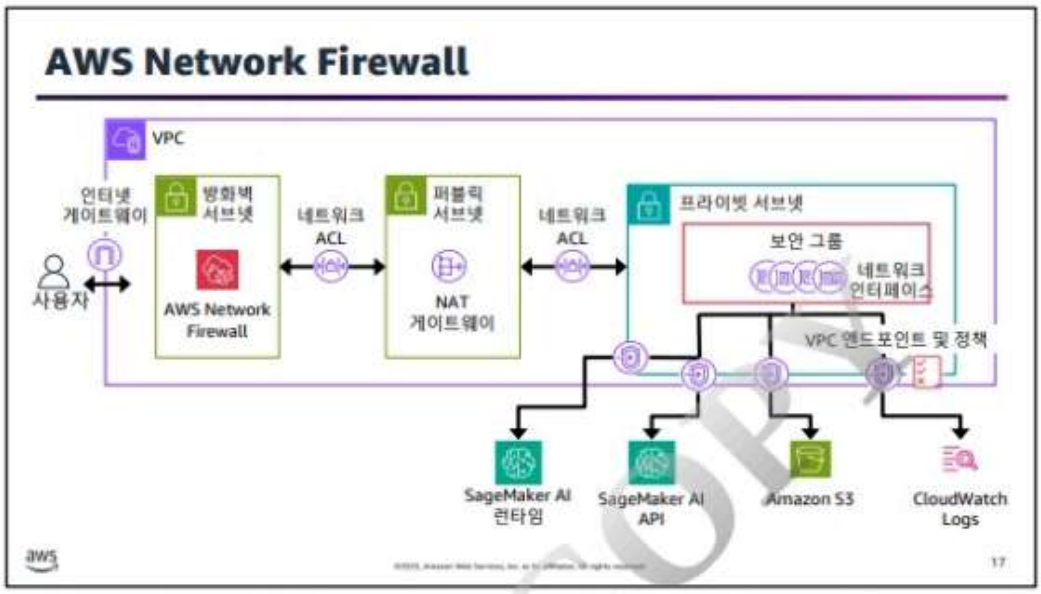
ML 워크플로에서 네트워크 ACL을 활용할 수 있는 방식은 다음과 같습니다.
- 서브넷 수준에서 특정 IP 범위 또는 포트 차단
- ML 리소스용 심층 방어 계층 추가 구현
- ML 인프라의 서브넷 간 트래픽 제어

예를 들어 SageMaker AI 훈련 작업이 포함된 서브넷의 보안을 강화하려는 경우 다음과 같이 네트워크 ACL을 구성할 수 있습니다.
- 특정 IP 범위(예: 기업 네트워크)에서 전송되는 인바운드 트래픽만 허용
- 알려진 악성 IP 주소에서 전송되는 모든 인바운드 트래픽 거부
- 필수 AWS 서비스(예: Amazon S3 및 Amazon ECR)로 전송되는 아웃바운드 트래픽 허용
- 기타 모든 아웃바운드 인터넷 액세스 차단

네트워크 ACL과 보안 그룹을 함께 활용하면 ML 리소스용 다계층 보안 방식을 적용할 수 있습니다. 네트워크 ACL로는 서브넷 수준에서 광범위하게 트래픽을 제어하고, 보안 그룹으로는 인스턴스 수준에서 리소스를 더 자세하게 보호할 수 있기 때문입니다. 즉, 이 두 기능을 동시에 사용하면 워크플로 전반에서 민감한 ML 데이터와 모델을 안전한 상태로 유지할 수 있습니다.

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



AWS Network Firewall은 VPC용 네트워크 보호 기능을 제공하는 관리형 서비스입니다. 스테이트풀 침입 방지 및 탐지 기능이 제공되는 이 서비스를 사용하면 광범위한 위협으로부터 ML 인프라를 보호할 수 있습니다.

Network Firewall의 주요 기능은 다음과 같습니다.
- 스테이트풀 방식 네트워크 트래픽 검사
- 침입 방지 및 탐지
- 웹 필터링 및 도메인 이름 기반 제어
- 사용자 지정 규칙 생성 및 관리

ML 워크플로에서 Network Firewall을 활용할 수 있는 방식은 다음과 같습니다.
- SageMaker AI 엔드포인트에서 송수신되는 트래픽 흐름 검사 및 제어
- 알려진 위협과 새로운 위협으로부터 ML 인프라 보호
- 세분화된 네트워크 통신 제어 구현

예를 들어 ML 추론 엔드포인트를 보호하려는 경우 다음과 같이 Network Firewall을 구성할 수 있습니다.
- SageMaker AI 엔드포인트로 전송되는 HTTPS 트래픽(포트 443)만 허용
- 알려진 악성 도메인으로의 아웃바운드 연결 차단
- ML 인프라의 공격 가능성을 식별하는 침입 탐지 규칙 구현
- 의심스러운 트래픽 패턴 기록 및 해당 알림 표시

Network Firewall을 배포하면 ML 리소스에 고급 보호 계층을 추가할 수 있습니다. 심층 패킷 검사 및 위협 인텔리전스 기반 필터링을 제공하는 이 서비스를 활용하면 보안 그룹, 네트워크 ACL 등의 기존 보안 제어 기능을 보완할 수 있기 때문입니다.

ML 환경에서 새롭게 등장하는 위협에 대응하고 계속해서 변화하는 요구 사항을 충족하려면 Network Firewall 규칙을 정기적으로 검토 및 업데이트해야 합니다. 이처럼 위험을 사전에 예방하면 장기적으로 ML 워크플로의 보안과 무결성을 유지할 수 있습니다.

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

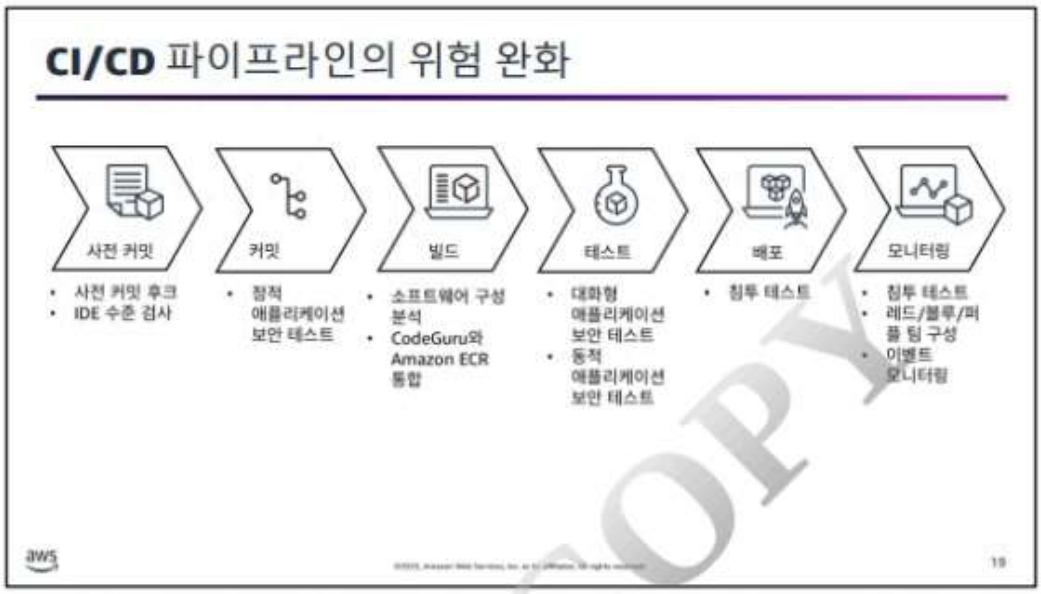
Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

기계 학습 워크플로의 개발 및 배포 프로세서 전반에 걸쳐 ML 모델과 데이터의 무결성, 기밀성 및 가용성을 높이려면 CI/CD 파이프라인을 보호해야 합니다.

DO NOT COPY
ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



CI/CD 파이프라인에서는 코드 작성, 테스트 실행, 새 버전 배포를 비롯한 소프트웨어 제공 프로세스가 자동화됩니다. 그런데 적절한 보안 조치를 취하지 않으면 이 파이프라인이 취약해져 데이터 침해, 모델 변조, 인프라 손상 등의 결과가 발생할 수 있습니다.

위험을 완화하려면 다음과 같이 모든 파이프라인 단계에 보안 관행을 통합해야 합니다.

1. 사전 커밋: 사전 커밋 후크 및 통합 개발 환경(IDE) 도구를 사용하여 코드 작성 표준을 적용하고 문제를 조기에 탐지합니다.
2. 커밋: 정적 애플리케이션 보안 테스트(SAST)를 구현하여 소스 코드의 취약성을 분석합니다.
3. 빌드: 소프트웨어 구성 분석(SCA)을 사용하여 오픈 소스 패키지 및 종속성의 취약성을 파악합니다.
4. 테스트: 동적 애플리케이션 보안 테스트(DAST) 및 대화형 애플리케이션 보안 테스트(IAST)를 적용하여 실행 중인 애플리케이션을 평가합니다.
5. 배포: 침투 테스트를 수행하여 최종 제품의 취약성을 파악합니다.
6. 모니터링: 레드/블루/퍼플 팀을 구성하여 공격을 시뮬레이션하고 지속적으로 방어 방식을 개선합니다.

이러한 보안 사례를 통합하면 개발 수명 주기 전반에 걸쳐 위험을 크게 줄이고 ML 모델 및 인프라를 보호할 수 있습니다.

DO NOT COPY ssleeeee1@gmail.com

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

CI/CD 파이프라인을 보호해야 하는 동시에 **인프라 보안**도 유지해야 합니다. 인프라 보안 유지를 위해 준수해야 하는 주요 지침은 다음과 같습니다.

- 코드형 인프라(IaC)를 AWS CloudFormation과 함께 사용하여 안전하고 일관된 환경을 조성합니다. 버전 관리 및 인프라 자동 배포 기능을 활용하면 수작업으로 인한 오류를 줄이고 재현 가능성을 높일 수 있습니다.
- AWS Config 및 AWS CloudTrail을 사용하여 인프라 변경 사항을 모니터링하고 감사합니다. 이러한 서비스에서는 리소스 구성과 API 활동을 확인하여 보안 문제를 빠르게 탐지하고 적절하게 대응할 수 있습니다.
- AWS Secrets Manager를 사용하여 보안 정보를 안전하게 저장 및 교체합니다. 이 서비스를 사용하면 API 키 및 데이터베이스 암호와 같은 민감한 정보를 보호할 수 있으며 자격 증명을 자동으로 교체하여 손상 위험을 줄일 수 있습니다.
- IAM을 사용하여 액세스 및 권한을 관리합니다. 그리고 최소 권한의 원칙을 구현하여 사용자와 서비스에 필요한 권한만 부여합니다.

이러한 관행을 통합하면 개발 및 배포 프로세스 전반에서 ML 모델과 인프라를 보호하는 안정적인 보안 태세를 갖출 수 있습니다.

DO NOT COPY ssleeeee1@gmail.com

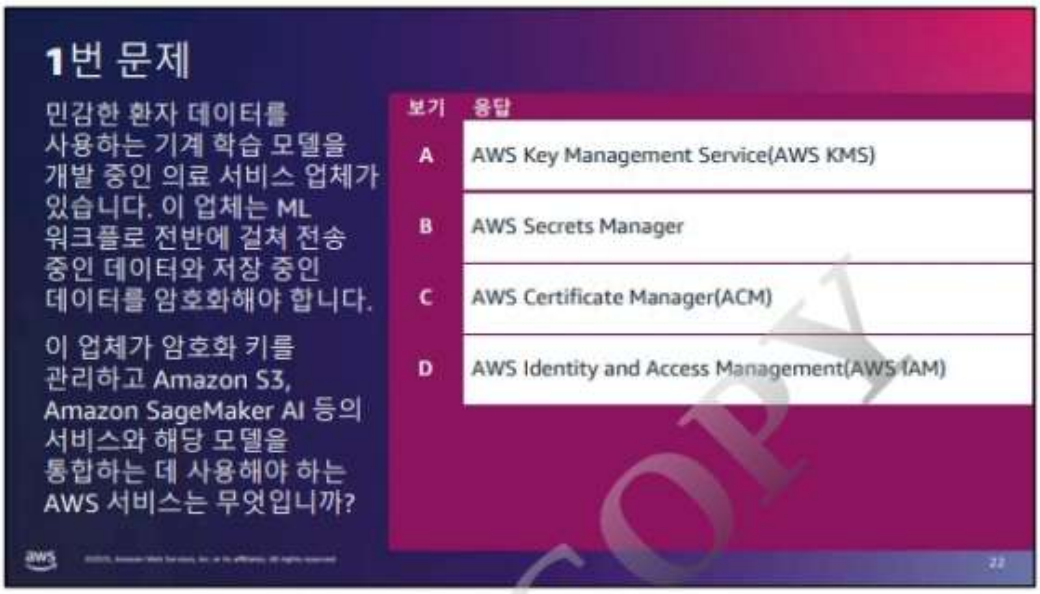
Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

이제 이 모듈에서 다루었던 개념을 정확하게 이해하고 있는지를 평가하는 문제를 풀어 봅니다. 모듈에서 배운 내용에 따라 각 문제의 정답을 선택하십시오. 이러한 문제를 풀어 보면 주요 개념을 더욱 정확하게 이해하고 더 복습을 해야 하는 부분을 파악할 수 있습니다.

DO NOT COPY
ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



ML 워크플로의 데이터(특히 민감한 의료 데이터)를 보호하는 데 사용되는 AWS 서비스를 이해하고 있는지를 평가하기 위한 문제입니다.

DO NOT COPY
ssleeeee1@gmail.com

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

정답은 A, AWS Key Management Service(AWS KMS)입니다.

설명:
이 시나리오에서 암호화 키를 관리하는 데 적합한 서비스는 AWS KMS입니다. AWS KMS는 Amazon S3 및 SageMaker AI와 원활하게 통합되어
저장된 데이터와 전송 중인 데이터를 암호화합니다. AWS KMS를 사용하면 민감한 데이터 보호에 사용되는 암호화 키를 중앙 집중식으로 제어할 수 있습니다.

오답:
B. Secrets Manager는 암호화 키 관리용 서비스가 아니라 데이터베이스 자격 증명 등의 보안 정보 저장 및 교체용 서비스입니다.
C. ACM은 암호화 키 관리용 서비스가 아니라 SSL/TLS 인증서 프로비저닝 및 관리용 서비스입니다.
D. IAM은 AWS 리소스에 대한 액세스 관리용 서비스이며 암호화 키 관리는 처리하지 않습니다.

DO NOT COPY ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

2번 문제

데이터 과학자가 SageMaker AI Notebook, 훈련 데이터가 포함된 S3 버킷, 장기 자격 증명을 사용하여 모니터링할 CloudWatch 로그에 액세스할 수 있는 권한을 팀원들에게 부여해야 합니다.

이러한 권한을 관리하는 데 사용할 수 있는 가장 안전한 방식은 무엇입니까?

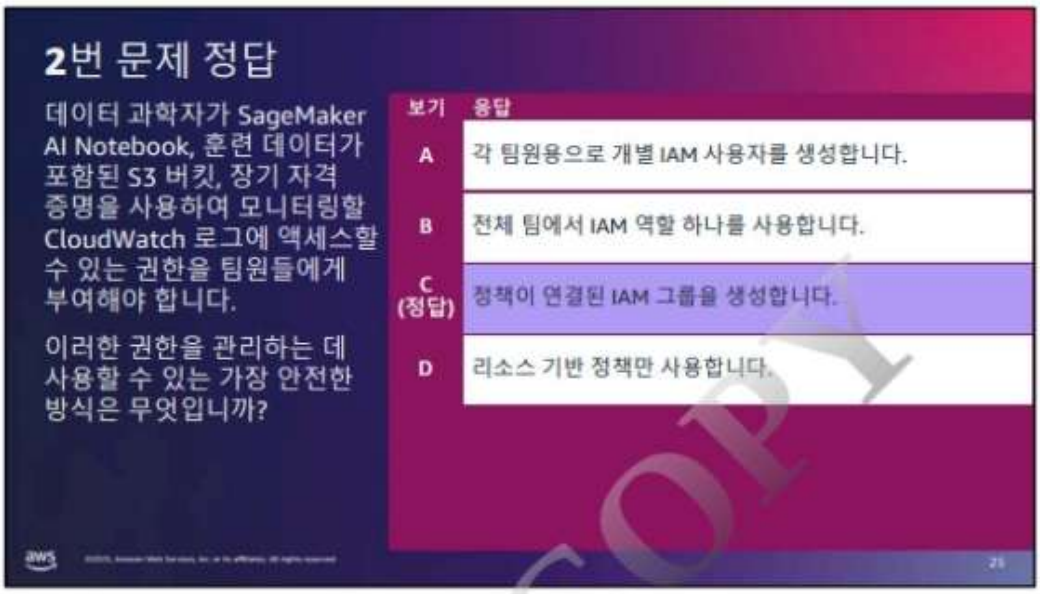
보기	응답
A	각 팀원용으로 개별 IAM 사용자를 생성합니다.
B	전체 팀에서 IAM 역할 하나를 사용합니다.
C	정책이 연결된 IAM 그룹을 생성합니다.
D	리소스 기반 정책만 사용합니다.

24

ML 워크플로용 IAM 역할 및 정책을 이해하고 있는지를 평가하기 위한 문제입니다.

DO NOT COPY
ssleeeee1@gmail.com

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



정답은 C, 정책이 연결된 IAM 그룹을 생성합니다.입니다.

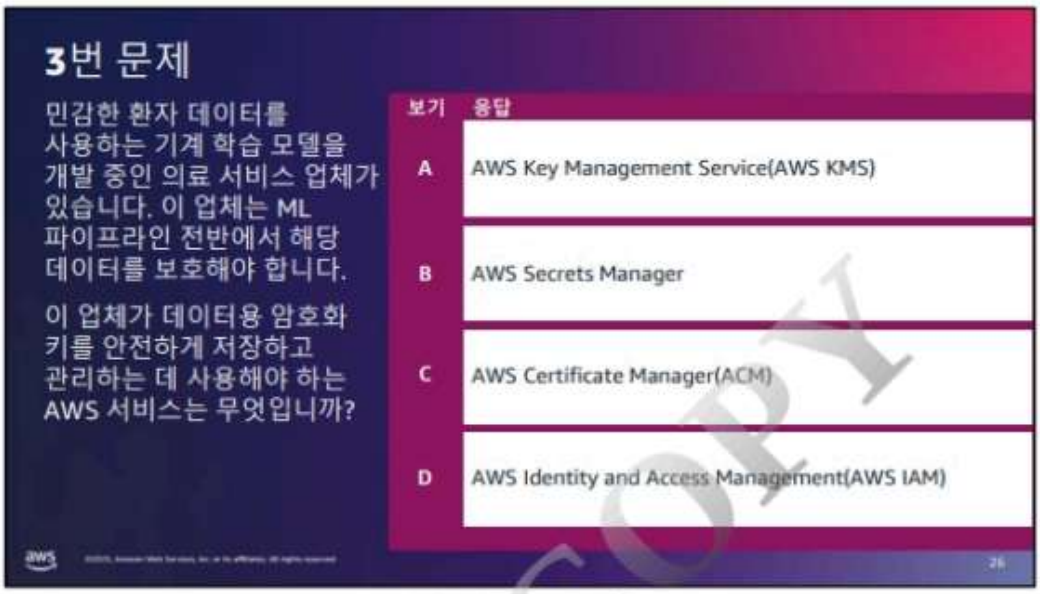
설명: 이 시나리오에서 가장 안전하며 쉽게 관리할 수 있는 방식은 정책이 연결된 IAM 그룹을 생성하는 것입니다. IAM 그룹을 사용하면 여러 사용자에게 권한을 할당하고 중앙에서 해당 권한을 관리할 수 있습니다. 최소 권한의 원칙을 준수하는 이 방식을 적용하면 팀원이 프로젝트에 새로 합류하거나 프로젝트에서 빠질 때 권한을 더욱 쉽게 업데이트할 수 있습니다.

오답:
A. 개별 IAM 사용자를 생성해도 되기는 하지만 팀 규모가 커지면 권한을 업데이트해야 할 때 관리가 어려워질 수 있습니다.
B. 전체 팀용으로 IAM 역할 하나를 사용하는 방식은 최소 권한의 방식 위반에 해당합니다. 모든 팀원에게 동일 수준의 액세스 권한이 제공되기 때문입니다.
D. 리소스 기반 정책만 사용하는 경우 여러 서비스와 팀원의 권한을 세부적으로 관리할 수 없습니다.

DO NOT COPY ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

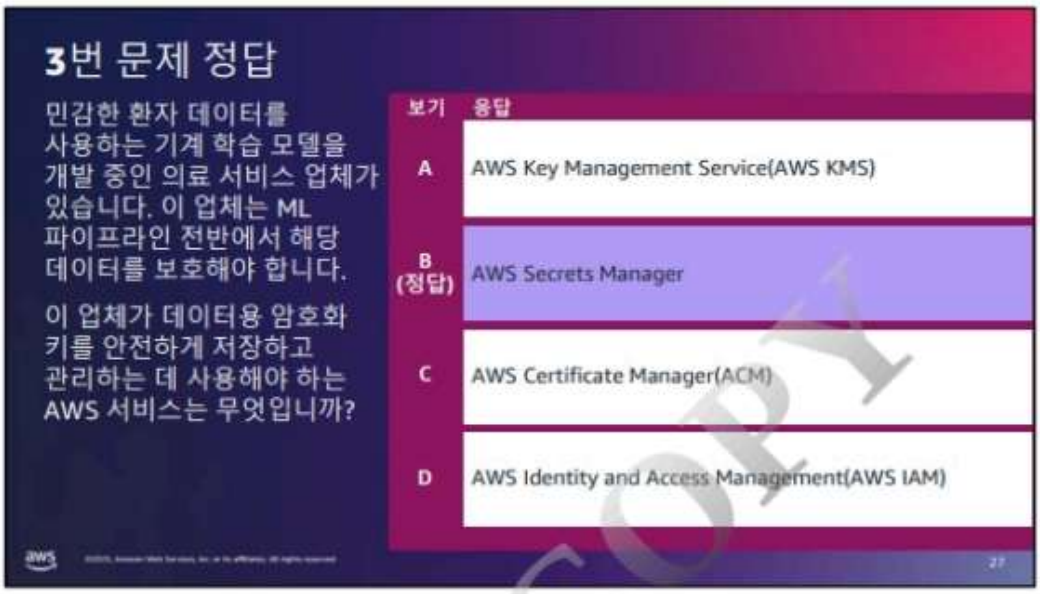
Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



ML 워크플로의 데이터 보호 관련 지식을 평가하기 위한 문제입니다.

DO NOT COPY
ssleeeee1@gmail.com

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



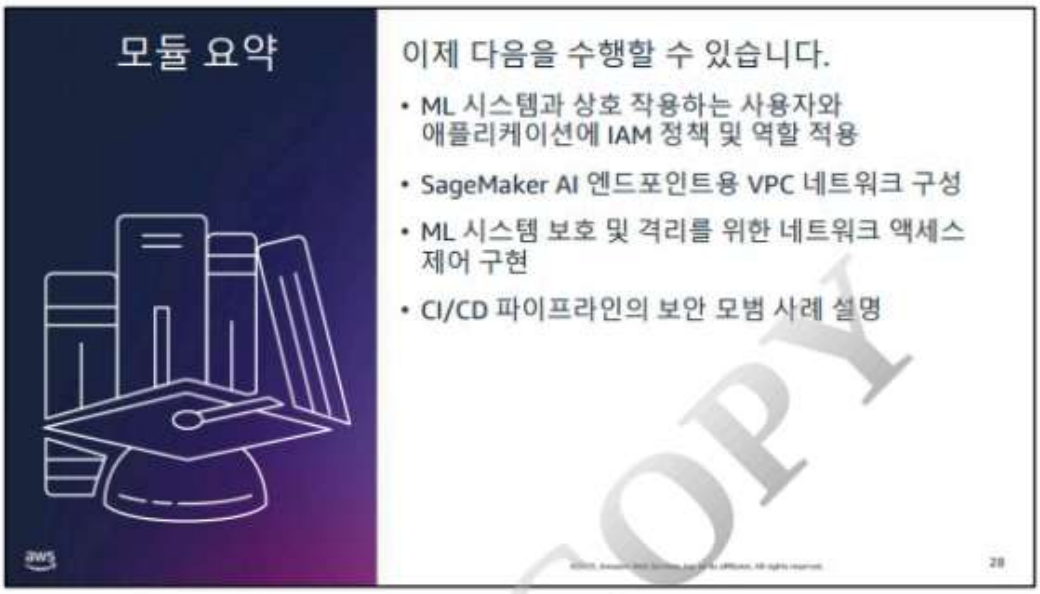
정답은 B, AWS Secrets Manager입니다.

설명: AWS Secrets Manager는 데이터베이스 자격 증명 및 API 키와 같은 보안 정보와 기타 민감한 데이터를 안전하게 저장, 관리 및 검색할 수 있는 서비스입니다. 이 서비스를 사용하면 이러한 보안 정보와 민감한 데이터를 수명 주기 전반에 걸쳐 쉽게 교체하고 해당 데이터에 대한 액세스를 관리할 수 있습니다. 또한 이 서비스는 Amazon S3, SageMaker AI 등 ML 워크플로에서 사용되는 타 AWS 서비스와도 원활하게 통합됩니다.

오답:
B. AWS Key Management Service(AWS KMS)는 암호화 키 저장 및 교체용 서비스입니다.
C. ACM은 데이터 보호를 위한 보안 정보 관리용 서비스가 아니라 SSL/TLS 인증서 프로비저닝 및 관리용 서비스입니다.
D. IAM은 자격 증명 및 보안 정보 저장과 관리용 서비스가 아니라 AWS 리소스에 대한 액세스 관리용 서비스입니다.

DO NOT COPY ssleeeee1@gmail.com

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



이제 다음을 수행할 수 있습니다.

- ML 시스템과 상호 작용하는 사용자 및 애플리케이션에 IAM 정책 및 역할 적용. 이렇게 하면 액세스를 적절하게 제어하고 최소 권한을 원칙을 준수할 수 있습니다.
- SageMaker AI 엔드포인트용 VPC 네트워크 구성. 이렇게 하면 ML 리소스용 격리 환경을 구축하고 네트워크 액세스를 제어할 수 있습니다.
- 보안 그룹, 네트워크 ACL, VPC 엔드포인트와 같은 도구를 사용하여 ML 시스템 보호 및 격리를 위한 네트워크 액세스 제어 구현
- CI/CD 파이프라인의 보안 모범 사례 설명(통신 보호, 코드 분석, 개발 수명 주기 전반에 걸친 지속적 모니터링 포함)

이러한 기술을 습득하면 AWS에서 안전한 ML 시스템을 구축하는 데 필요한 준비를 갖출 수 있습니다. 계속해서 현재 진행 중인 프로젝트에서 안정적인 보안 태세를 갖출 수 있도록 이러한 개념을 연습하고 관련 개념을 적용해 보시기 바랍니다.

DO NOT COPY ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

AWS ML 리소스 보호 관련 사항을 살펴본 이 세션에 참석해 주셔서 감사합니다. 궁금한 점이 있으면 지금 질문하실 수 있습니다. 나중에 문의 사항이 있거나 피드백을 제공하려는 경우에는 강사가 제공하는 링크를 통해 AWS 교육 지원 팀에 문의하십시오. ML의 보안 유지는 지속적으로 진행해야 하는 프로세스이며, 데이터와 모델을 보호하려면 모범 사례 관련 정보를 지속적으로 파악해야 한다는 점을 꼭 기억해 주시기 바랍니다.

DO NOT COPY
ssleeeee1@gmail.com

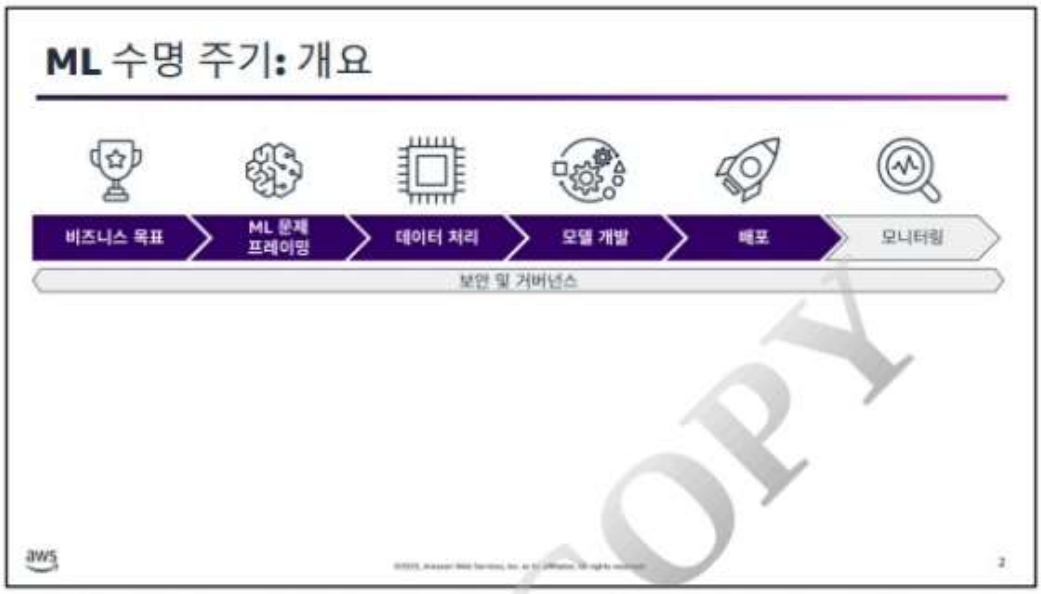
Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

이 모듈에서는 **기계 학습 운영(MLOps)** 및 기계 학습(ML) 모델용 **자동 배포 전략을** 중점적으로 살펴봅니다. 또한 엔드 투 엔드 ML 워크플로를 간편하게 진행하고 자동화할 수 있는 **지속적 통합 및 지속적 전달(CI/CD)**과 같은 개발 운영(DevOps) 방법론의 원칙 및 해당 방법론 적용 시의 모범 사례를 살펴봅니다.

DO NOT COPY
ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

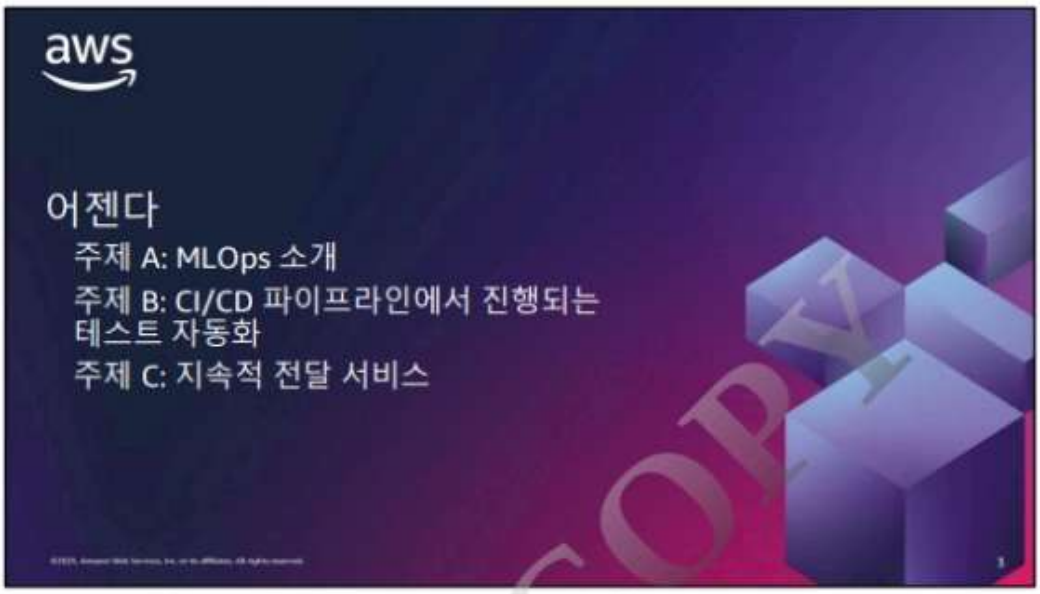


슬라이드의 다이어그램에는 전체 ML 수명 주기가 대략적으로 표시되어 있습니다. 이전 모듈에서는 기본적인 배포 관련 개념을 살펴보았습니다. 이 모듈에서는 ML CI/CD 워크로드의 배포에 대해 중점적으로 살펴봅니다.

DO NOT COPY
ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



이 모듈에서는 CI/CD 파이프라인 구축에 사용할 수 있는 자동 오케스트레이션 도구를 살펴봅니다. 이 모듈에서 소개하는 주제는 다음과 같습니다.

A. MLOps 소개: 기계 학습 워크로드에 DevOps 관행과 CI/CD 원칙을 적용할 때 사용되는 방법론
B. CI/CD 파이프라인에서 진행되는 테스트 자동화: 문제를 조기에 찾아내고 모델 성능을 유지할 수 있도록 CI/CD 파이프라인 내에서 자동 테스트 및 검증을 구현하기 위한 전략
C. 지속적 전달 서비스: 기계 학습 모델의 자동 배포와 전달을 지원하는 AWS 서비스(예: AWS CodePipeline, AWS CodeBuild, AWS CodeDeploy) 관련 개요 정보

DO NOT COPY
ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

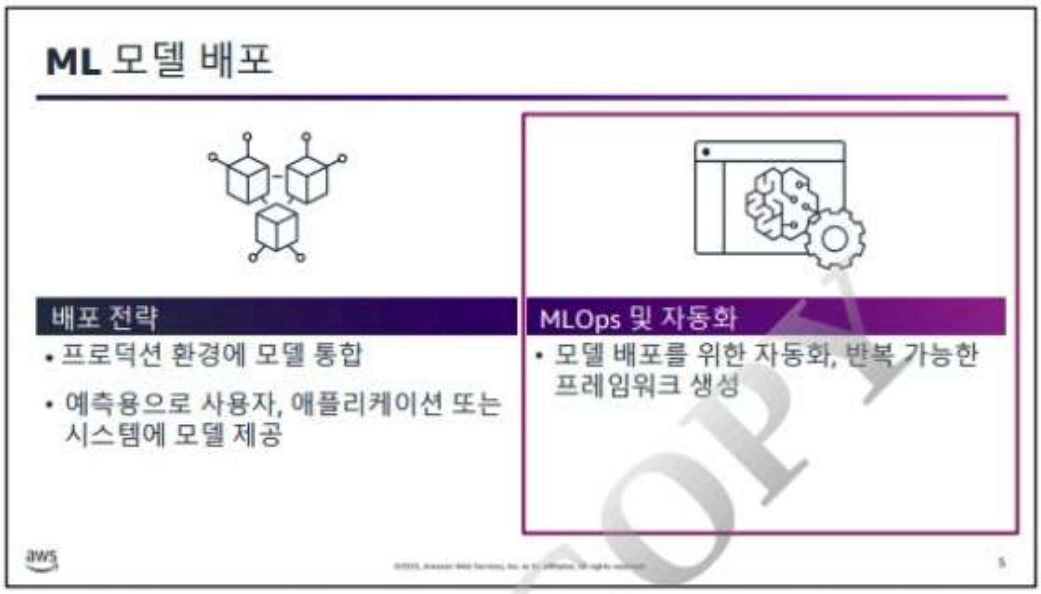
Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

MLOps란 자동화, 모니터링, 등의 DevOps 원칙을 기계 학습 시스템에 적용하는 방식을 지칭합니다. 그러나 MLOps에는 개발부터 배포까지 ML 모델 수명 주기 전반의 고유한 당면 과제를 관리하기 위한 맞춤형 프로세스와 절차가 필요합니다.

DO NOT COPY
ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

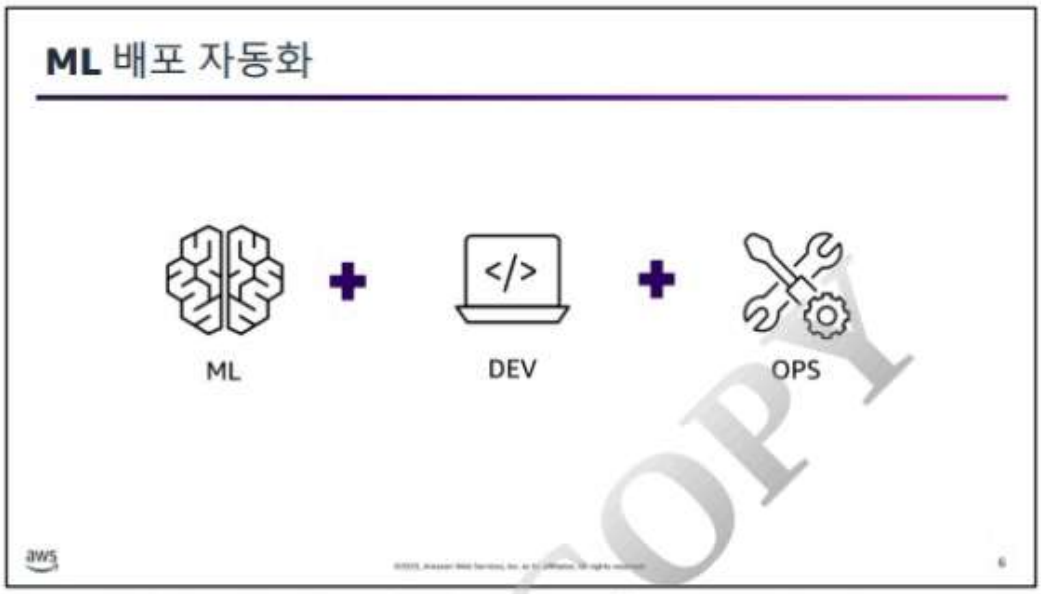


ML 모델 배포의 2가지 주요 요소는 배포 전략과 MLOps 자동화입니다. 배포는 훈련된 모델을 프로덕션 환경에 적절하게 통합하고, 예측을 수행해야 하는 최종 사용자, 애플리케이션 및 시스템이 해당 모델에 액세스할 수 있는지를 중점적으로 확인하는 과정입니다.

MLOps 구성 요소는 반복 가능한 모델 배포용 자동 프레임워크를 생성하는 데 주로 사용됩니다. 이처럼 체계적인 방식을 활용하면 배포 일관성을 높이고 수작업으로 인한 오류를 줄일 수 있으며, 조직 전체에서 기계 학습 작업의 규모를 효율적으로 조정할 수 있습니다. 자동화 프레임워크에서는 모델 검증에서 모니터링과 업데이트에 이르기까지 ML 수명 주기의 모든 작업이 처리됩니다.

DO NOT COPY
ssleeeee1@gmail.com

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



ML 워크플로를 자동화하면 몇 주가 걸리는 모델 배포 시간을 몇 시간으로 단축할 수 있으며, 프로덕션 모델에 영향을 주기 전에 데이터 품질 문제를 찾아낼 수 있습니다.

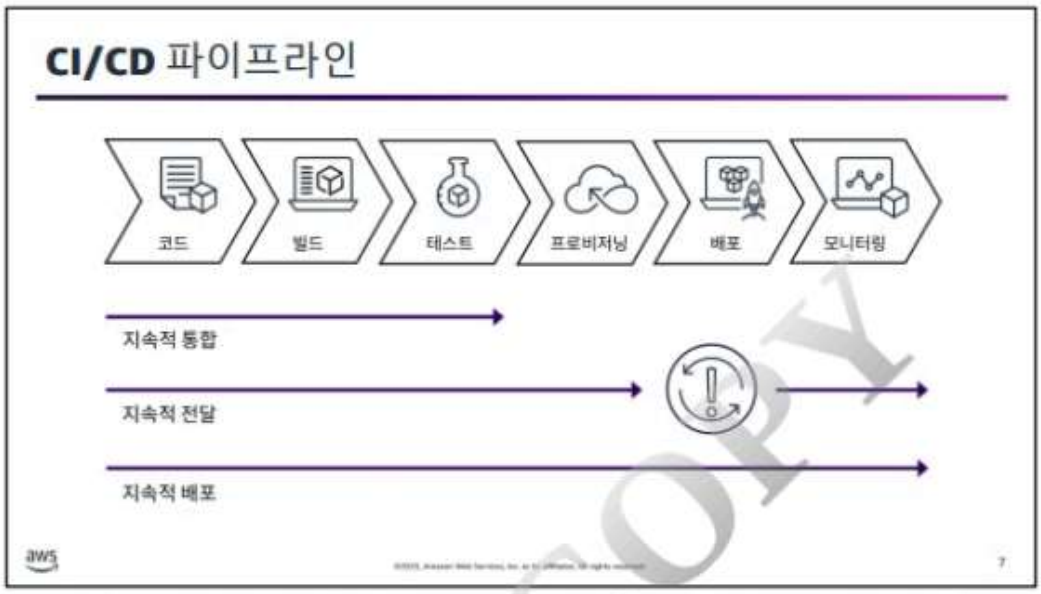
기계 학습 모델 개발은 다음과 같은 몇 가지 주요 측면에서 기존 소프트웨어 개발과 다릅니다.

- 복잡한 수학 방정식과 여러 번의 반복 실험이 포함됩니다.
- 새 데이터가 유입되면 훈련된 모델은 구형이 될 수 있으므로 지속적인 모니터링과 재훈련 및 재배포를 진행해야 합니다.

이처럼 고유한 당면 과제를 해결하려면 ML 워크플로에 맞게 조정된 모범 사례를 채택해야 합니다. MLOps와 DevOps는 몇 가지 유사점을 공유하지만 ML 모델이 복잡하기 때문에 배포 및 지속적인 운영 시 추가 관리가 필요합니다.

DO NOT COPY
ssleeeee1@gmail.com

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



CI/CD 파이프라인에서는 구축, 테스트 및 배포 프로세스가 자동화하되므로 기계 학습 모델 릴리스 속도와 신뢰도가 높아집니다.

DevOps의 필수 단계인 소프트웨어 릴리스 프로세스에는 다음과 같은 전체 소프트웨어 개발 수명 주기가 포함됩니다.

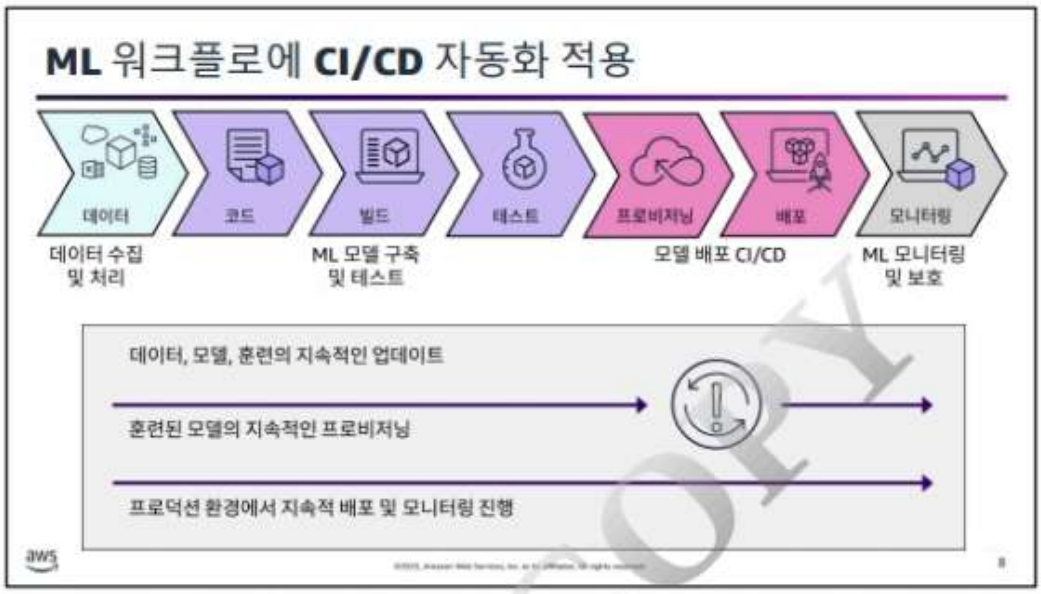
1. **코드**: 개발자가 애플리케이션 코드를 작성하여 소스 코드 리포지토리에 저장합니다.
2. **빌드**: 코드를 컴파일하고 배포에 적합한 패키지로 빌드합니다.
3. **테스트**: 빌드된 애플리케이션의 기능 및 성능을 테스트하고 버그가 있으면 수정합니다.
4. **프로비저닝**: 서버 및 데이터베이스와 같은 인프라가 배포를 위해 프로비저닝됩니다.
5. **배포**: 애플리케이션을 스테이징 및 프로덕션 환경에 배포합니다.
6. **모니터링**: 프로덕션 환경에서 성능과 오류를 모니터링합니다.

CI/CD 파이프라인에서 이러한 단계가 모두 포함된 전체 프로세스를 자동화하면 소프트웨어 변경 사항을 신뢰할 수 있는 방식으로 빠르게 전달할 수 있습니다.

DO NOT COPY ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



ML 워크플로에 CI/CD 자동화 원칙을 효율적으로 적용하면 데이터 처리부터 모니터링까지의 전 과정을 간편하게 진행할 수 있는 파이프라인이 구축됩니다. 이러한 통합 방식에서는 ML 수명 주기가 데이터 처리, 구축과 테스트, 배포(CI/CD), 모니터링과 보호 등의 여러 개별 단계로 구분되며 각 단계에서는 특정 자동 프로세스와 검사가 진행됩니다.

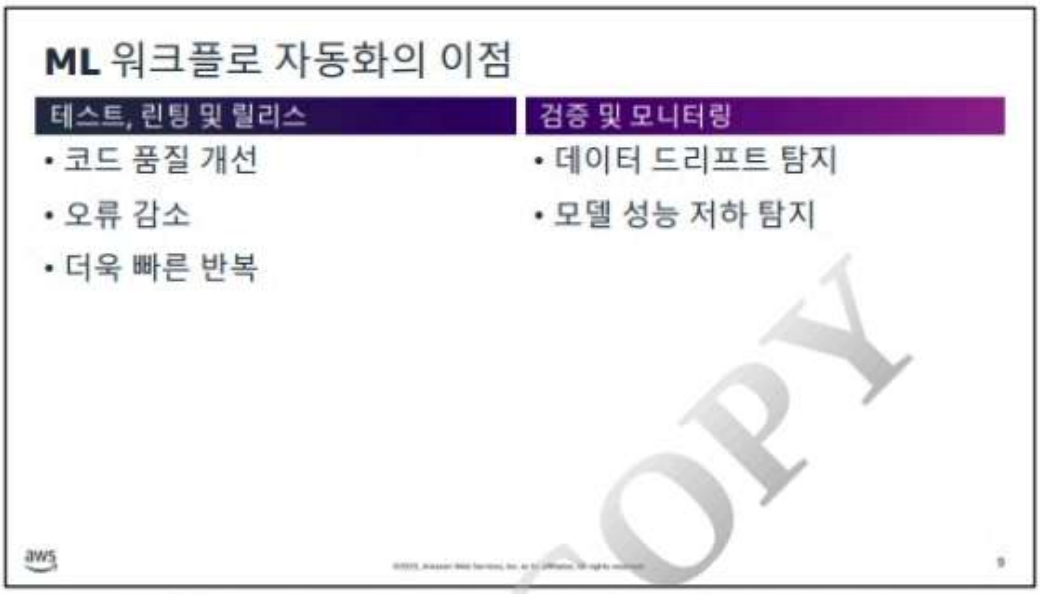
ML 워크플로를 자동화하면 다음의 2가지 주요 카테고리에서 많은 이점이 제공됩니다. 우선 개발 단계를 자동화하면 일관성 있는 테스트와 린팅을 진행하여 코드 품질을 개선할 수 있습니다. 개발 작업을 자동화하면 인적 오류가 감소하며 모델 개발 반복 주기 속도도 빨라집니다. 그러면 높은 품질 표준을 유지하면서 전체 개발 프로세스를 신속하게 진행할 수 있습니다.

운영 측면에서는 검증 및 모니터링 시스템을 자동화하면 데이터 드리프트, 모델 성능 저하 등의 중요한 문제를 탐지할 수 있습니다. 이러한 자동 검사에서는 프로덕션 환경에서 모델의 정확도와 신뢰성이 유지되는지를 검증할 수 있으므로 사업 결과에 영향을 주기 전에 문제를 미리 해결할 수 있습니다. 이처럼 체계적인 ML 운영 방식을 채택하면 프로덕션 ML 시스템의 효율성과 신뢰성을 높일 수 있습니다.

DO NOT COPY
ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



ML 워크로드에 CI/CD 원칙을 통합하면 기계 학습 시스템 구축, 배포 및 유지 관리 과정에서 다음과 같은 획기적인 이점이 제공됩니다.

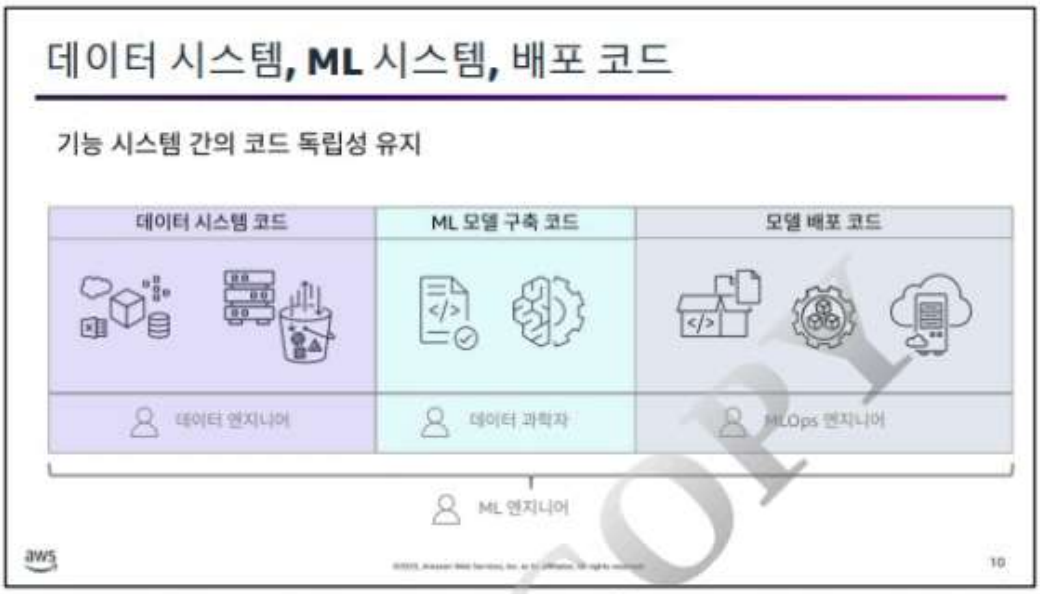
- **테스트, 린팅, 릴리스:** CI/CD 파이프라인을 통해 이러한 프로세스를 자동화하면 코드 품질을 개선하고 오류를 줄이는 동시에 반복 주기를 더욱 빠르게 진행할 수 있습니다.
- **검증 및 모니터링:** CI/CD 파이프라인에서는 데이터 드리프트 및 성능 저하 모니터링과 데이터세트 및 모델 검증이 자동화됩니다. 따라서 장기적으로 ML 시스템의 안정성과 높은 성능이 유지됩니다.
- **문제 조기 탐지:** 테스트 및 모니터링을 자동화하면 모델을 프로덕션 환경에 배포하기 전에 개발 프로세스 초반에서 문제를 파악할 수 있으므로 더욱 품질이 우수한 모델과 코드를 배포할 수 있습니다.

데이터 드리프트는 시간의 경과에 따라 입력 데이터의 통계적 속성과 패턴이 점진적으로 변화하는 현상입니다. 원래 훈련 데이터세트와 전혀 다른 데이터를 배포된 기계 학습 모델에 제공하는 경우 이러한 현상이 뚜렷하게 나타납니다.

DO NOT COPY
ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

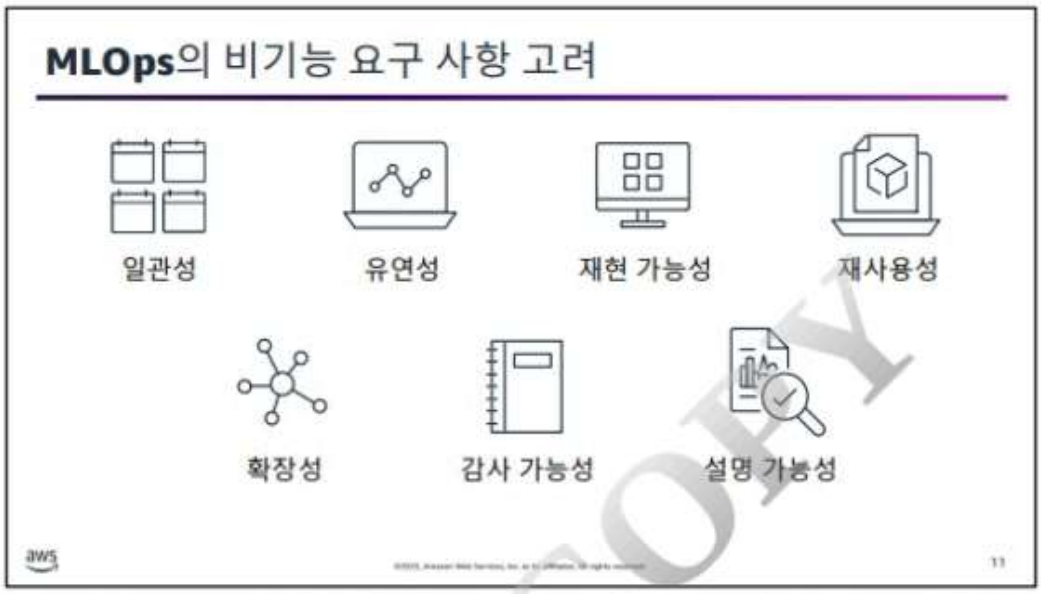


데이터 및 ML 코드의 독립성은 3가지 주요 구성 요소인 데이터 관리, 모델 빌드, 배포 간의 명확한 책임 분리를 통해서 설명됩니다. 이러한 체계에서는 전문 담당자가 각 구성 요소를 취급합니다. 데이터 엔지니어는 데이터 인프라와 파이프라인을 관리하며, 데이터 과학자는 모델 개발과 훈련을 중점적으로 진행합니다. 그리고 MLOps 엔지니어는 배포 및 프로덕션 인프라 관련 작업을 처리합니다.

이처럼 각 구성 요소를 분리하면 워크로드 효율성을 높이고 ML 시스템을 더욱 원활하게 유지 관리할 수 있습니다. 데이터 엔지니어는 모델 변경 여부에 관계없이 데이터 파이프라인을 최적화할 수 있습니다. 데이터 과학자는 세부적인 배포 작업에 신경 쓸 필요 없이 다양한 모델링 방식을 적용해 볼 수 있습니다. 그리고 MLOps 엔지니어는 모델 개발 과정에 영향을 주지 않고 배포 인프라를 개선할 수 있습니다. 즉, 이러한 모듈식 방식을 사용하는 경우 개별 구성 요소 담당자가 명확하게 정의되므로 전체 팀이 동시에 작업을 진행할 수 있습니다.

DO NOT COPY
ssleeeee1@gmail.com

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



기계 학습 시스템에는 MLOps와 관련하여 해결해야 하는 다음과 같은 몇 가지 **비기능 요구 사항**이 있습니다.

- **일관성**: 모델 성능을 유지하기 위해 개발, 스테이징, 프로덕션 등 다양한 환경 간의 편차 최소화
- **유연성**: 다양한 ML 프레임워크와 기술을 수용하여 변화하는 요구 사항에 적응
- **재현 가능성**: 과거 실험 및 훈련 프로세스를 재현하여 모델 신뢰성과 추적성 강화
- **재사용성**: 다양한 ML 프로젝트에서 재사용할 수 있는 구성 요소를 설계하여 효율성 개선 및 개발 시간 단축
- **확장성**: 증가하는 수요와 데이터 볼륨을 처리하기 위해 컴퓨팅, 스토리지, 메모리와 같은 리소스의 탄력성 강화
- **감사성**: 투명성과 규정 준수를 위해 모든 ML 아티팩트에 대한 포괄적인 로그, 버전 관리 및 종속성 추적 제공
- **설명 가능성**: 의사 결정의 투명성과 모델 해석 가능성을 촉진하는 기법 통합

MLOps 프레임워크 내에서 안정성과 신뢰성이 높은 확장 가능 ML 시스템을 구축하려면 이러한 비기능 요구 사항을 충족해야 합니다.

DO NOT COPY
ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

DevOps와 MLOps 기능 비교

기능	DevOps	MLOps
코드 버전 관리	✔	✔
컴퓨팅 환경	✔	✔
CI/CD	✔	✔
프로덕션에서의 모니터링	✔	✔
데이터 출처		✔
데이터세트		✔
모델		✔
워크플로를 통한 모델 구축		✔
모델 배포 워크플로		✔

12

기계 학습 프로젝트는 기존 소프트웨어 개발과 비교할 때 독특한 복잡성을 가지고 있습니다. DevOps와 MLOps에서 동일하게 적용되는 원칙도 있지만 MLOps에서는 다음과 같은 사항을 추가적으로 고려해야 합니다.

- **데이터 버전 관리**: 훈련 데이터 및 데이터세트의 변경 사항 추적 및 관리
- **모델 드리프트**: 시간 경과에 따른 배포된 모델의 성능 저하 모니터링 및 해결
- **실험 추적**: 모델 훈련 실험, 하이퍼파라미터 및 결과 레코드 유지 관리

MLOps와 DevOps는 몇 가지 유사점을 공유하지만 ML 모델이 복잡하기 때문에 배포 및 지속적인 운영 시 추가 관리가 필요합니다. 기계 학습 실무자로서 모델을 프로덕션 환경으로 원활하게 전환하고 성공적으로 관리하려면 이러한 고유한 문제를 고려해야 합니다.

DO NOT COPY
ssleeeee1@gmail.com

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

이 섹션에서는 CI/CD 파이프라인에서 진행되는 테스트의 자동화에 대해 중점적으로 살펴봅니다. 개발 프로세스 초반에 문제를 찾아내고 모델 품질을 유지하려면 테스트를 자동화해야 합니다. 이 주제에서는 CI/CD 파이프라인을 사용하여 ML 워크플로의 여러 단계에서 자동화된 테스트를 구현하는 전략을 살펴봅니다.

DO NOT COPY
ssleeeee1@gmail.com

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

ML 시스템 신뢰성

- 수동 테스트는 오류가 발생하기 쉬우며 규모 조정이 어려움
- 개발의 여러 단계에서 자동 테스트 구현
- 워크플로 오케스트레이션에 AWS 서비스 사용

14

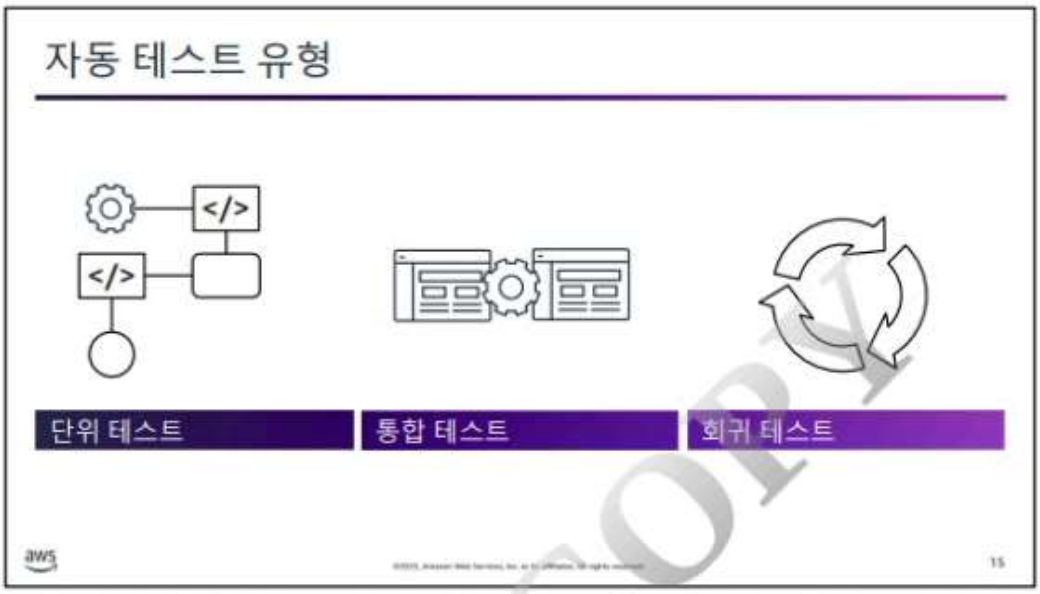
수동 테스트는 오류가 발생하기 쉬우며 ML 프로젝트가 복잡해지면 규모를 조정하기도 어렵습니다. 시스템의 신뢰성을 유지하려면 **개발 과정의 여러 단계에서 자동화된 테스트를 구현**해야 합니다.

워크플로 오케스트레이션에 AWS 서비스를 사용하면 핵심 CI/CD 개념을 구현함으로써 개발 프로세스 전반에서 자동화된 테스트를 적극 활용할 수 있습니다. 이 방식을 채택하는 경우 오류를 조기에 찾아내 피드백을 신속하게 제공할 수 있으며 전반적인 품질도 개선할 수 있습니다.

DO NOT COPY
ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



ML 개발 프로세스 초반에 문제를 찾아내고 모델 품질을 유지하려면 테스트를 자동화해야 합니다. 자동 테스트의 3가지 주요 유형은 다음과 같습니다.

- **단위 테스트**: 데이터 전처리 단계, 모델 예측 로직 등의 개별 구성 요소나 기능을 검증합니다. 이러한 테스트에서 각 단위가 격리 상태로 정상 작동하는지를 확인할 수 있습니다.
- **통합 테스트**: 전체 데이터 수집 및 모델 훈련 파이프라인과 같이 서로 결합되어 있는 여러 구성 요소의 전반적인 기능을 확인합니다. 이러한 테스트에서는 시스템의 여러 부분이 올바르게 연동되는지를 확인합니다.
- **회귀 테스트**: 코드 변경 후 다시 실행하여 기존 기능이 부적절하게 손상되었거나 성능이 저하되지 않았음을 확인합니다. 이러한 테스트를 실행하면 모델 성능이나 데이터 품질 저하를 방지할 수 있습니다.

이처럼 여러 수준에서 단위, 통합, 회귀 등의 자동 테스트를 구현하면 ML 시스템을 포괄적으로 검증할 수 있습니다. 이 방식을 사용하는 경우 오류를 조기에 찾아내고 수동 작업을 줄일 수 있으며 ML 수명 주기 전반에서 높은 품질 표준을 유지할 수 있습니다.

DO NOT COPY
ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

CI/CD 파이프라인에 자동 테스트를 통합하면 기계 학습 프로젝트에서 다음과 같은 여러 가지 주요 이점이 제공됩니다.

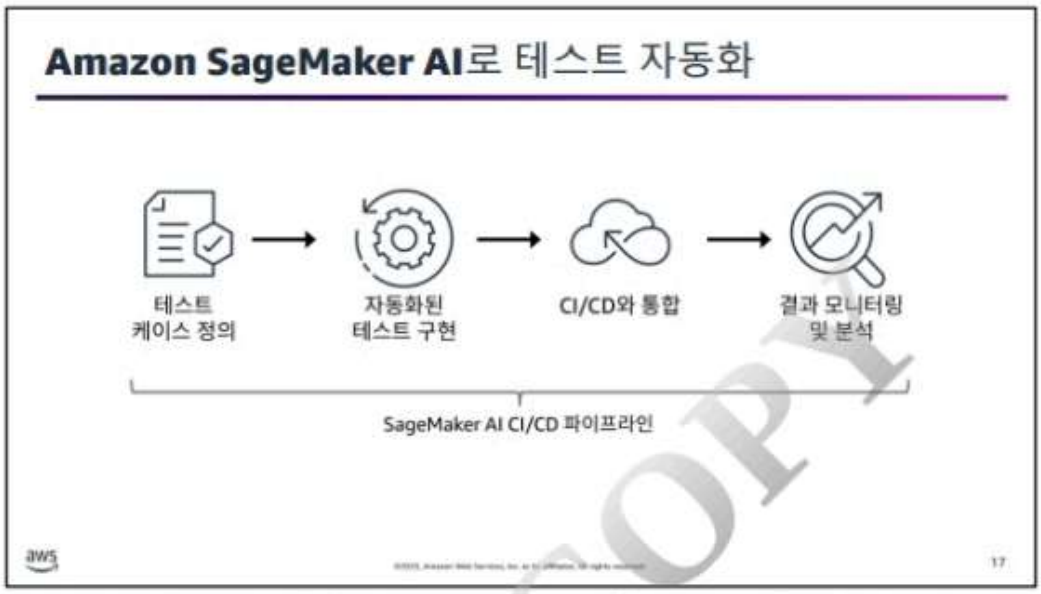
- **속도**: 테스트가 신속하게 실행되므로 빠른 피드백 루프가 제공됩니다. 이로 인해 데이터 과학자와 엔지니어가 문제를 더 빨리 파악하여 해결할 수 있으므로 개발 주기가 더욱 신속하게 진행됩니다.
- **신뢰성**: 자동 테스트는 일관된 방식으로 실행되므로 인적 오류가 발생하지 않으며 여러 환경에서 반복 활용 가능한 믿을 수 있는 결과가 제공됩니다.
- **적용 범위**: 테스트에서 다양한 시나리오, 입력, 하이퍼파라미터 및 성능 지표를 평가할 수 있으므로 ML 시스템과 엣지 사례를 철저하게 검증할 수 있습니다.
- **오류 조기 탐지**: 프로덕션에 배포하기 전에 개발 프로세스 초반에 버그와 회귀를 찾아내면 비용이 많이 드는 문제를 방지하고 모델의 높은 품질을 유지할 수 있습니다.

요약하자면, CI/CD 파이프라인을 통해 테스트를 자동화하는 경우 ML 워크플로를 간편하게 진행하고 코드 품질을 높일 수 있습니다. 그뿐만 아니라 기술 부채를 줄이고 모델 업데이트와 릴리스의 빈도 및 신뢰성을 높일 수 있습니다.

DO NOT COPY
ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



Amazon SageMaker AI에서는 ML 워크플로의 테스트 자동화에 사용할 수 있는 포괄적인 솔루션이 제공됩니다. Amazon SageMaker Pipelines를 사용하여 자동화된 테스트 솔루션을 설정하는 주요 단계는 다음과 같습니다.

- **테스트 케이스 정의**: ML 프로젝트에서 자동화하려는 구체적인 테스트 케이스를 파악합니다. 데이터 품질 검사, 모델 성능 지표, 엣지 케이스 시나리오 등을 예로 들 수 있습니다.
- **자동 테스트 구현**: pytest 등의 테스트 프레임워크를 사용하여 자동 테스트를 작성하거나 AWS에서 제공하는 사전 구축된 테스트를 사용합니다. 이러한 테스트는 파이프라인에 통합되며 코드 커밋 또는 모델이 업데이트되고 나면 자동 실행됩니다.
- **CI/CD와 통합**:AWS CodePipeline, Jenkins 등의 서비스를 사용하여 지속적 통합 및 지속적 전달(CI/CD) 파이프라인에 자동 테스트를 통합합니다.
- **결과 모니터링 및 분석**: 테스트 결과를 검토한 후 자동 테스트에서 탐지된 문제를 빠르게 식별하여 대응할 수 있도록 경보나 알람을 설정합니다.

SageMaker Pipelines 및 CI/CD 워크플로에 자동 테스트를 통합하면 **데이터 드리프트**, 모델 회귀 및 기타 문제를 조기에 찾아낼 수 있으므로 개발 수명 주기 전반에서 ML 시스템의 높은 신뢰성과 품질을 유지할 수 있습니다.

DO NOT COPY
ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

이 섹션에서는 ML 워크플로의 지속적 전달 및 자동화를 위해 사용 가능한 여러 가지 AWS 서비스를 살펴봅니다. 지속적 전달(CD)은 MLOps에서 중요한 역할을 합니다. 프로덕션 환경에 검증된 모델과 코드 변경 사항을 효율적이며 안정적으로 배포할 수 있도록 지원하기 때문입니다.

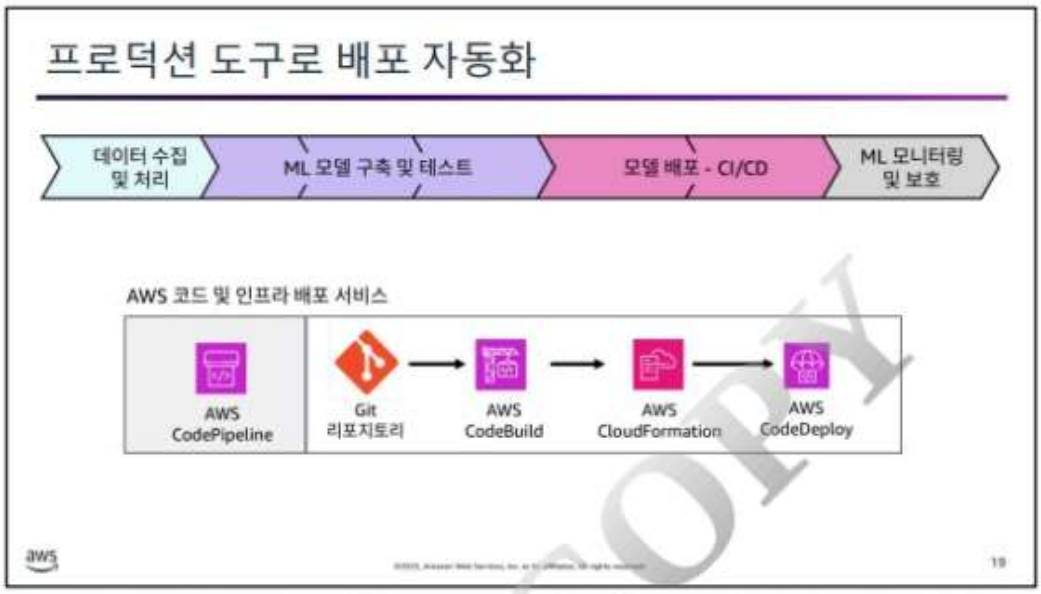
이 섹션에서는 AWS CodePipeline, AWS CodeBuild, AWS CloudFormation, AWS CodeDeploy 등의 서비스를 살펴봅니다. 이러한 서비스를 함께 활용하는 경우 ML 수명 주기의 구축, 테스트 및 배포 단계를 자동화할 수 있는 포괄적인 솔루션이 제공됩니다. 이러한 AWS 서비스를 활용하여 CD 사례를 구현하는 경우 여러 가지 이점이 제공됩니다. 해당 이점으로는 더욱 신속한 릴리스 주기 진행, 수동 작업 감소, 여러 환경의 일관성 개선, 필요시 신속한 변경 사항 롤백 등이 있습니다. 우수한 성능 수준을 유지하면서 계속 변화하는 데이터 패턴에 맞게 모델의 작동 방식을 조정하기 위해 모델을 자주 업데이트하고 재훈련해야 하는 기계 학습에서는 자동화가 특히 중요합니다.

이러한 AWS 서비스를 사용하면 ML 배포 프로세스를 간편하게 진행하고 다운타임을 최소한으로 줄일 수 있으며 일관성과 신뢰도가 높은 모델을 프로덕션 환경에 배포할 수 있습니다. 그러면 조직이 고객에게 획기적인 AI/ML 솔루션을 더욱 효율적으로 제공할 수 있습니다.

DO NOT COPY ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



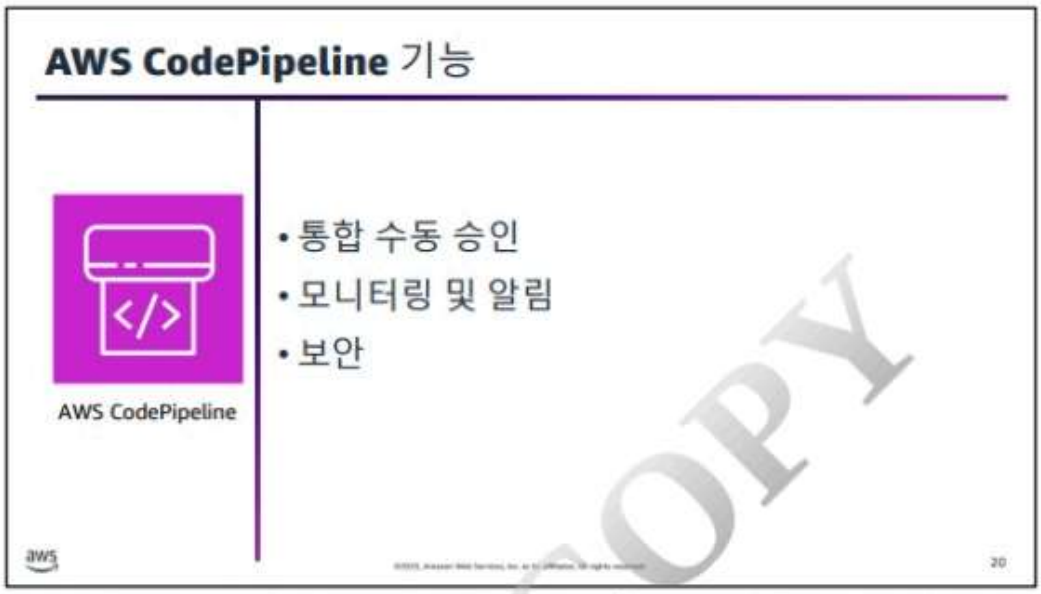
ML 모델 배포를 자동화하려면 프로덕션급 배포 도구에 통합할 수 있는 체계적인 파이프라인이 필요합니다. 이 프로세스에서 진행되는 포괄적인 워크플로에서는 초기 데이터 처리, 모델 구축과 테스트, 배포와 모니터링 등의 단계를 수행합니다. 이러한 워크플로를 따르면 프로덕션 환경에 ML 모델을 체계적으로 배포할 수 있습니다.

AWS에서는 이러한 자동화 파이프라인을 지원하는 유용한 서비스 제품군이 제공됩니다. 우선 오케스트레이션용 AWS CodePipeline 서비스를 사용하여 Git 리포지토리에서 코드를 끌어온 후 AWS CodeBuild를 사용하여 모델을 구축합니다. 그런 다음 AWS CloudFormation을 통해 인프라를 프로비저닝한 후 AWS CodeDeploy를 사용하여 최종 배포합니다. 이 통합 도구 체인을 활용하면 신뢰할 수 있는 일관된 방식으로 자동 배포를 진행하는 동시에 버전 제어와 인프라 관리도 적절하게 수행할 수 있습니다.

DO NOT COPY
ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



AWS CodePipeline은 애플리케이션 및 소프트웨어 업데이트 릴리스 프로세스를 자동화하는 완전관리형 지속적 전달 서비스입니다. 이 서비스에서는 배포 수명 주기 전반의 작업을 더욱 정확하게 제어 및 확인하고 보안을 강화할 수 있는 여러 가지 주요 기능이 제공됩니다.

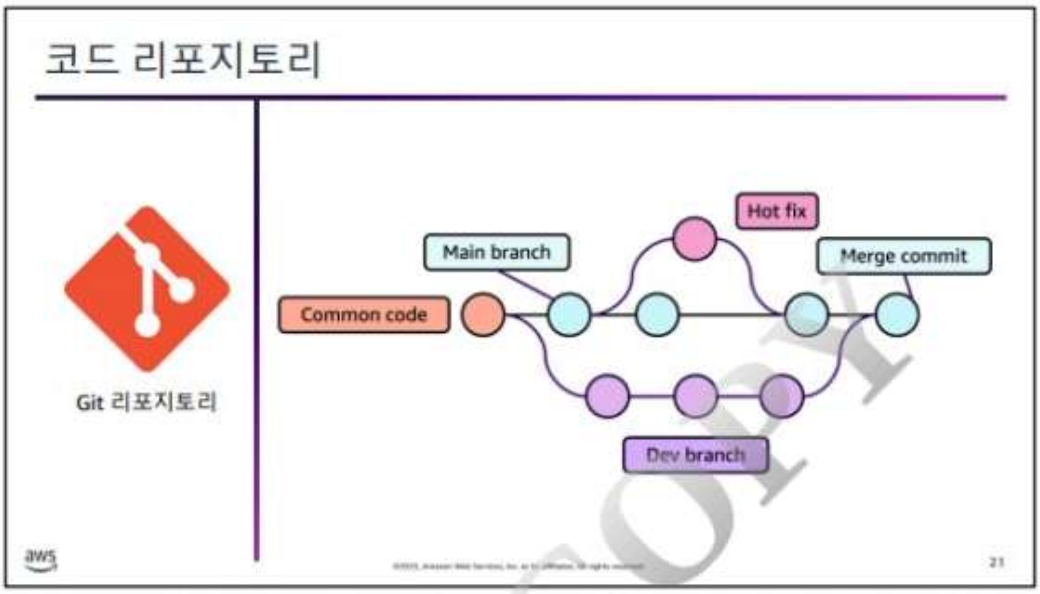
- **수동 승인** – CodePipeline에서는 파이프라인의 특정 단계에 수동 승인 게이트를 구성할 수 있습니다. 이렇게 하면 지정된 사용자나 그룹이 명시적으로 승인을 해야 배포를 진행할 수 있는 통제된 릴리스가 가능해집니다.
- **모니터링 및 알림** – 이 서비스는 세부 모니터링 및 로깅 기능을 제공하므로 파이프라인 상태를 추적하고 중요한 이벤트나 문제와 관련된 알림을 수신할 수 있습니다. 이처럼 파이프라인 상태를 확인할 수 있으므로 문제를 미리 해결하여 파이프라인 성능을 최적화할 수 있습니다.
- **보안** – CodePipeline에서는 AWS Identity and Access Management(IAM) 정책을 통한 세분화된 액세스 제어가 지원됩니다. 그러므로 파이프라인에서 특정 작업을 수행할 수 있는 사용자나 역할을 지정하여 담당 업무를 적절하게 구분하는 동시에 보안 관련 모범 사례를 준수할 수 있습니다.

이러한 기능을 활용하면 애플리케이션과 ML 모델용으로 안정적이며 안전한 감사 가능 CD 프로세스를 구현할 수 있습니다. CodePipeline 사용 시에는 릴리스 주기를 더욱 빠르게 진행하고 협업 방식을 개선하는 동시에 배포 워크플로를 더욱 효율적으로 통제할 수 있습니다.

DO NOT COPY
ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



최신 방식으로 코드를 관리할 때 기본적으로 사용되는 Git 리포지토리에서는 버전 관리 및 협업 방식 개발을 위한 안정적인 시스템이 제공됩니다. 리포지토리 구조에는 보통 프로덕션에 바로 사용할 수 있는 안정적인 코드를 보관하고 프로젝트의 공식 버전 기록을 나타내는 메인 브랜치(마스터라고도 함)가 포함됩니다. 2024년부터 AWS의 소스 제어 전략을 신규 고객 등록이 중지된 자체 CodeCommit 서비스 보다는 서드 파티 Git 공급자를 지원하는 데 초점을 맞추는 쪽으로 바꾸었습니다.

기본 코드베이스의 무결성을 보호하기 위해 개발 작업은 보통 별도의 브랜치에서 진행됩니다. 새 기능과 버그 수정을 테스트하는 기본 위치는 개발 브랜치입니다. 개발자는 해당 브랜치에서 변경 사항을 기본 브랜치에 병합하기 전에 안전하게 적용해 보면서 검증할 수 있습니다. 테스트는 개발 브랜치에서 수행되는 경우가 많지만 대규모 프로젝트의 경우에는 품질 보증 프로세스를 테스트와 분리하고 문제를 더욱 명확하게 파악하기 위해 전용 테스트 브랜치를 활용할 수도 있습니다.

Git 리포지토리에서 흔히 수행되는 작업이 슬라이드의 워크플로에 나와 있습니다. 이 워크플로에서는 긴급하게 해결해야 하는 문제용 핫픽스를 생성하는 작업, 기본 브랜치에 커밋을 다시 병합하는 프로세스 등을 진행하게 됩니다. 이러한 브랜칭 전략을 채택하면 안정적인 상태의 공통 코드를 유지하면서 여러 개발 작업을 병렬로 진행할 수 있습니다. 개발 수명 주기에서 각 브랜치가 특정 용도로 사용되기 때문입니다. 병합 커밋은 검증된 변경 사항의 테스트와 검토가 정상적으로 완료되면 기본 코드베이스에 해당 변경 사항을 다시 통합하는 지점에 해당됩니다.

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

AWS CodeBuild 기능

- 환경 및 사양
- 로깅 및 아티팩트
- 확장성 및 가용성
- AWS 서비스와 통합
- 모니터링 및 보안

22

AWS CodeBuild는 소스 코드를 컴파일하고 테스트를 실행하며, 배포 준비가 된 소프트웨어 패키지를 생성하는 완전관리형 지속적 통합 서비스입니다. 다음과 같은 여러 가지 주요 기능을 제공하는 이 서비스는 구축 및 테스트 프로세스를 자동화할 수 있는 유용한 도구입니다.

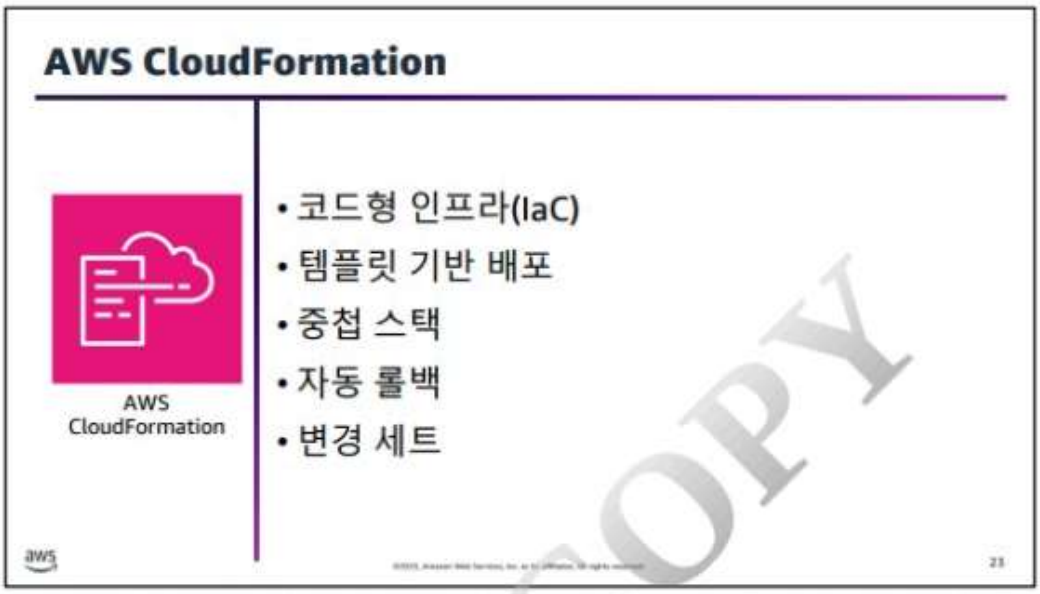
- 코드 변경 사항이 푸시되면 단위 테스트, 통합 테스트, 기능 테스트, 엔드 투 엔드 테스트 등의 자동 테스트 모음을 실행합니다.
- 관리형 구축 환경과 Docker 이미지를 사용하는 사용자 정의 환경이 모두 지원되므로 YAML buildspec 파일을 사용하여 빌드를 구성할 수 있습니다.
- 구축 프로세스의 세부 로깅 기능이 제공되며, Amazon S3 버킷 또는 기타 대상에 빌드 아티팩트를 저장할 수 있습니다.
- 컴퓨팅 리소스의 규모가 구축 요구를 충족하도록 자동 조정되므로 높은 가용성을 유지하고 대기 시간을 최소한으로 단축할 수 있습니다.
- AWS CodePipeline, Git 코드 리포지토리, AWS CodeDeploy 등의 타 AWS 서비스와 원활하게 통합되므로 엔드 투 엔드 CI/CD 워크플로를 구축할 수 있습니다.
- AWS CodeBuild 콘솔, Amazon CloudWatch 및 기타 도구를 통해 구축 비표와 로그를 모니터링하는 기능을 제공하며 IAM 정책을 사용한 세분화된 액세스 제어도 지원합니다.

ML 워크플로에서 CodeBuild를 사용하는 경우 구축 및 테스트 프로세스를 자동화하여 코드 품질의 일관성과 신뢰성을 높일 수 있습니다. 또한 타 AWS 서비스와 CodeBuild를 원활하게 통합하여 전체 프로세스를 자동화할 수 있습니다.

DO NOT COPY
ssleeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



AWS CloudFormation은 ML 인프라 배포 및 관리를 자동화하는 과정에서 꼭 필요한 서비스입니다. ML 파이프라인에서 CloudFormation이 제공하는 이점은 다음과 같습니다.

- **코드형 인프라(IaC)** – IaC에서는 SageMaker AI 엔드포인트, 훈련 인스턴스, 데이터 처리 리소스를 비롯한 전체 ML 환경의 버전을 관리할 수 있습니다. 그러므로 개발, 테스트 및 프로덕션 환경 전반에서 ML 인프라의 재현 가능성이 높아집니다.
- **템플릿 기반 배포** – 템플릿 기반 배포 기능을 사용하면 ML 환경을 표준화할 수 있습니다. 분산 훈련 클러스터, 실시간 추론 엔드포인트 또는 배치 변환 작업 등의 여러 시나리오용으로 템플릿을 생성할 수 있습니다.
- **중첩 스택** – 중첩 스택을 활용하면 복잡한 ML 아키텍처를 손쉽게 구성할 수 있습니다. 가령 데이터 사전 처리 리소스, 훈련 인프라, 모델 배포 엔드포인트 관리용 스택을 하나씩 포함할 수 있습니다. 모든 스택은 원활하게 연동됩니다.
- **자동 롤백** – 배포 실패 시 자동 롤백 기능이 배포를 마지막으로 알려진 정상 상태로 되돌리므로 ML 서비스를 보호할 수 있습니다. 높은 모델 가용성을 유지하고 추론 엔드포인트 중단을 방지하려면 자동 롤백 기능이 꼭 필요합니다.
- **변경 세트** – 변경 세트에서는 인프라 수정 사항을 적용 전에 미리 볼 수 있습니다. 프로덕션 ML 엔드포인트를 업데이트하거나 훈련 리소스의 규모를 조정할 때는 이러한 변경 세트가 꼭 필요합니다. 변경 세트에서 수정 사항을 미리 보면 ML 인프라 개선 시의 위험이 감소합니다.

이러한 기능을 활용하면 AWS 환경 전반에서 신뢰할 수 있는 일관된 확장 가능 방식으로 ML 워크로드를 배포할 수 있습니다. 따라서 수명 주기 전반의 ML 작업을 효율적으로 자동 관리할 수 있습니다.

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

AWS CodeDeploy 기능

- 자동 배포
- 유연한 배포 유형
- 롤백 기능
- 서비스 통합
- 보안 및 규정 준수

24

AWS CodeDeploy는 AWS 컴퓨팅 서비스와 온프레미스 서버 전반에서 소프트웨어 릴리스를 간편하게 진행할 수 있는 배포 자동화 서비스입니다. 이 서비스의 주요 기능은 다음과 같습니다.

- **자동 배포** - 수동 프로세스를 진행할 필요가 없으므로 더욱 신뢰도가 높은 방식으로 애플리케이션을 신속하게 업데이트할 수 있습니다.
- **유동적인 배포 유형** - 블루/그린, 카나리, 현재 위치 배포 등의 유형 중에서 적절한 방식을 선택할 수 있으므로, 요구 사항에 따라 통제된 방식으로 롤아웃을 진행할 수 있습니다.
- **롤백 기능** - 배포 과정에서 문제가 발생하면 릴리스를 마지막 정상 작동 버전으로 신속하게 되돌릴 수 있습니다.
- **서비스 통합** - Amazon Elastic Compute Cloud(Amazon EC2), Amazon Elastic Container Service(Amazon ECS), AWS Lambda 등의 서비스와 통합할 수 있으므로 인프라 전반에서 전체 배포 과정을 자동화할 수 있습니다.
- **보안 및 규정 준수 기능** - 규제 대상 환경에서 세분화된 액세스 제어와 감사 추적을 진행할 수 있습니다.

CodeDeploy의 자동화 기능을 활용하면 배포 관련 위험과 인적 오류를 줄여 소프트웨어 릴리스 주기를 더욱 신속하게 진행할 수 있습니다.

DO NOT COPY ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

Amazon SageMaker Projects는 다음과 같은 통합 구성 요소 집합을 활용하여 주요 CI/CD 사례를 구현합니다.

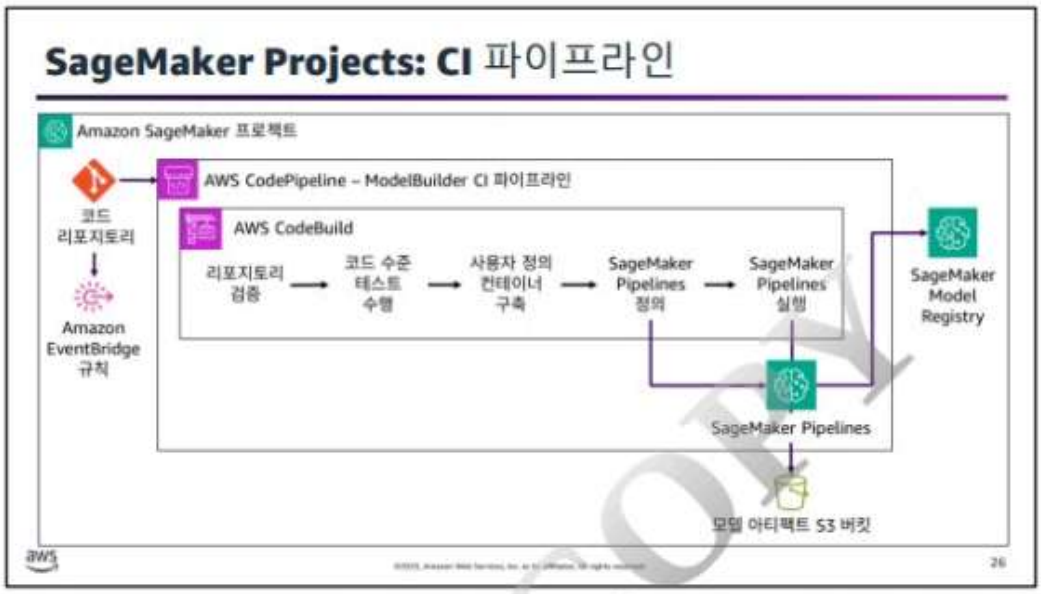
- **소스 코드 제어** - Git 리포지토리를 사용하여 ML 코드 버전을 관리합니다.
- **기본 제공 이벤트** - 코드 커밋에 따라 자동 워크플로를 시작합니다.
- **자동 모델 구축 파이프라인** - 훈련, 튜닝, 검증을 처리합니다.
- **중앙 모델 리포지토리** - 버전이 지정된 모델 아티팩트가 저장됩니다.
- **배포 파이프라인** - 여러 환경으로 모델을 자동 롤아웃합니다.

이 통합 프레임워크에서는 전체 ML 수명 주기를 추적 및 통합할 수 있습니다. 표준화된 템플릿 기반 프로세스 내에서 개발에서 배포까지의 모든 단계를 자동화하고 버전을 지정할 수 있으며 완벽하게 재현할 수 있습니다.

SageMaker Projects가 CI/CD 모범 사례를 기본적으로 적용하므로 일관성 있는 MLOps 워크플로를 유동적으로 진행할 수 있습니다.

DO NOT COPY
ssleeeee1@gmail.com

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



SageMaker AI의 ModelBuilder 워크플로는 ML 모델에 사용할 수 있는 포괄적인 CI/CD 파이프라인에 해당됩니다. 이 파이프라인에는 여러 주요 단계 및 구성 요소가 포함됩니다. 이 워크플로가 진행되는 구체적인 방식은 다음과 같습니다.

프로세스가 시작되면 먼저 데이터 과학자나 ML 엔지니어가 파이프라인 워크플로용 코드를 생성하여 커밋합니다. 그러면 Amazon EventBridge가 시작되며, EventBridge는 AWS CodeCommit 및 CodeBuild를 통해 파이프라인을 시작합니다. 전체 워크플로를 조율하는 AWS CodePipeline이 파이프라인 오케스트레이션을 관리합니다.

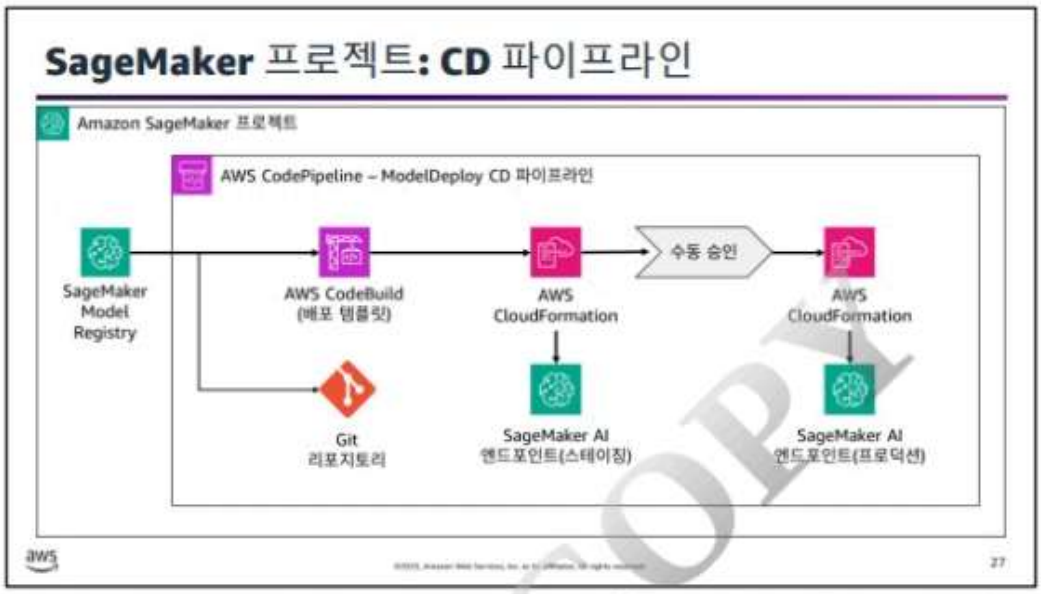
핵심 SageMaker AI 파이프라인에는 순서대로 진행되는 작업 3개가 포함되어 있습니다. 해당 작업은 데이터 사전 처리용 처리 작업, 모델 훈련용 처리 작업, 모델 평가용 처리 작업입니다. 이러한 단계가 완료되면 훈련된 모델이 Amazon S3에 아티팩트로 저장됩니다. 마지막 단계에서는 SageMaker Model Registry에 모델을 등록합니다. 모델을 등록하려면 권한이 있는 팀원이 배포 전에 모델을 승인해야 합니다. 이처럼 체계적인 방식에서는 ML 모델 배포 프로세스의 품질 관리, 재현 가능성 및 적절한 거버넌스 요구 사항이 적용됩니다.

이와 같은 자동 워크플로에는 DevOps 사례와 MLOps가 통합되어 있습니다. 즉, 소프트웨어 개발 관련 모범 사례와 ML 시스템의 구체적인 요구 사항이 모두 적용되어 있는 것입니다. 이 파이프라인에서는 모델 구축 및 배포 프로세스의 일관성, 추적 가능성 및 신뢰성을 강화하는 동시에 승인 시스템을 통해 프로세스를 적절하게 제어할 수 있습니다.

다음 슬라이드에서는 ModelBuilder CD 파이프라인에서 진행되는 이와 동일한 프로세스를 살펴봅니다.

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



SageMaker AI의 모델 배포 단계는 체계적인 CD 파이프라인에서 진행됩니다. 해당 파이프라인에서는 ML 모델이 통제된 방식으로 안전하게 프로덕션 환경에 배포됩니다. 이 프로세스에서는 먼저 SageMaker Model Registry가 승인된 모델을 저장하고 버전을 지정합니다. 즉, Model Registry가 프로덕션 환경에서 사용 가능한 모델의 중앙 리포지토리로 사용됩니다.

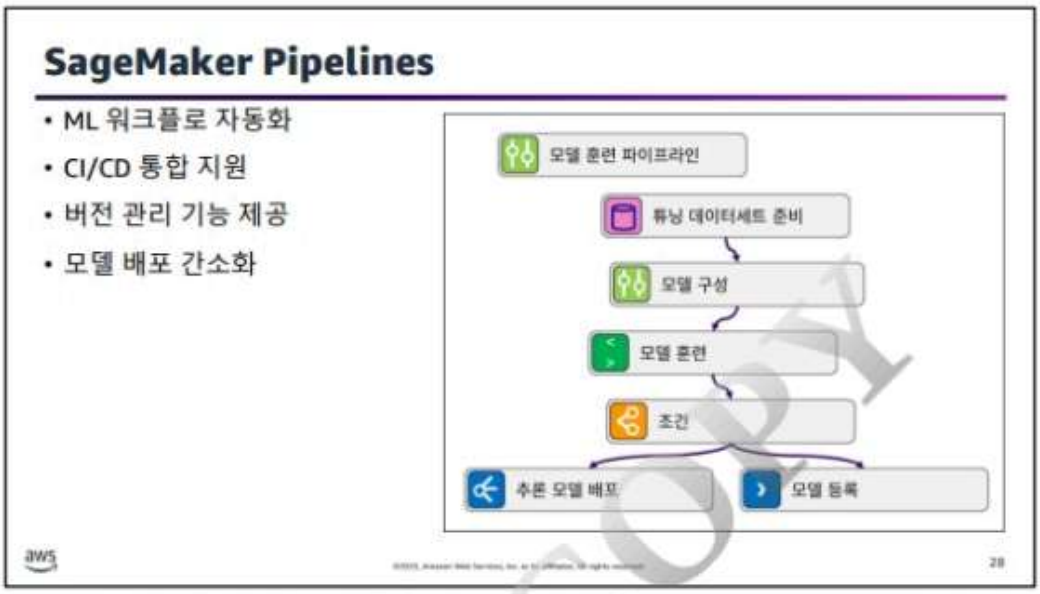
ModelDeploy CD 파이프라인은 여러 AWS 서비스를 통해 배포 프로세스를 오케스트레이션합니다. 먼저 Git 리포지토리가 배포 코드를 관리하며, 그 후에는 AWS CodeBuild가 CloudFormation 배포 템플릿 생성을 처리합니다. 그러고 나면 스테이징 및 프로덕션 엔드포인트로의 배포가 진행됩니다. 이 과정에서는 먼저 AWS CloudFormation을 통해 모델을 스테이징 환경으로 배포합니다. 그러면 테스트 및 검증용 SageMaker AI 엔드포인트가 생성됩니다.

이 파이프라인에는 스테이징 배포와 프로덕션 배포 단계 사이의 수동 승인 게이트가 필수적인 안전 기능으로 포함되어 있습니다. 스테이징 환경에서 검증 결과 정상으로 확인되어 명시적으로 승인된 모델만 프로덕션 환경으로 배포할 수 있습니다. 모델이 프로덕션 환경으로 배포되면 프로덕션 SageMaker AI 엔드포인트가 생성됩니다. 이처럼 수동 감독 단계가 포함되어 있는 단계별 방식을 채택하면 자동 배포 프로세스를 간편하게 진행하면서 문제가 있는 모델이 프로덕션 환경으로 배포되는 상황을 방지할 수 있습니다.

DO NOT COPY
ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



ML 워크플로 자동화에 사용할 수 있는 포괄적인 솔루션인 Amazon SageMaker Pipelines에서는 ML 개발 및 배포 프로세스 간소화를 위한 4가지 주요 이점이 제공됩니다. SageMaker Pipelines는 전체 ML 워크플로를 자동화하므로 수동 개입과 오류 발생 가능성을 줄이는 동시에, CI/CD를 원활하게 통합하여 MLOps에 소프트웨어 개발 모범 사례를 적용할 수 있습니다.

슬라이드의 예제 파이프라인에 나와 있는 전형적인 ML 워크플로에서는 먼저 모델 훈련과 데이터세트 준비를 진행합니다. 그러고 나면 Nova Pro를 사용하여 모델을 미세 조정하고 추가 훈련을 진행합니다. 이러한 과정을 통해 복잡한 ML 프로세스를 자동으로 오케스트레이션할 수 있습니다. 파이프라인 끝부분의 조건부 브랜칭은 시스템의 유동적인 작동 방식을 나타냅니다. 즉, 모델 성능이나 기타 기준에 따라 시스템이 자동으로 결정을 내릴 수 있습니다.

워크플로 다이어그램에 나와 있듯이 SageMaker Pipelines에는 버전 관리와 간편한 모델 배포가 필수 기능으로 포함되어 있습니다. 추론 모델을 배포할지 아니면 사전 정의된 조건에 따라 모델을 등록할지를 파이프라인에서 자동으로 선택할 수 있으므로 검증된 모델만 프로덕션 환경으로 배포됩니다. 이와 같은 체계적인 ML 작업 방식을 채택하면 재현 가능성과 추적 가능성을 높이고 조직 내에서 ML 프로젝트 규모를 효율적으로 조정할 수 있습니다.

DO NOT COPY
ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

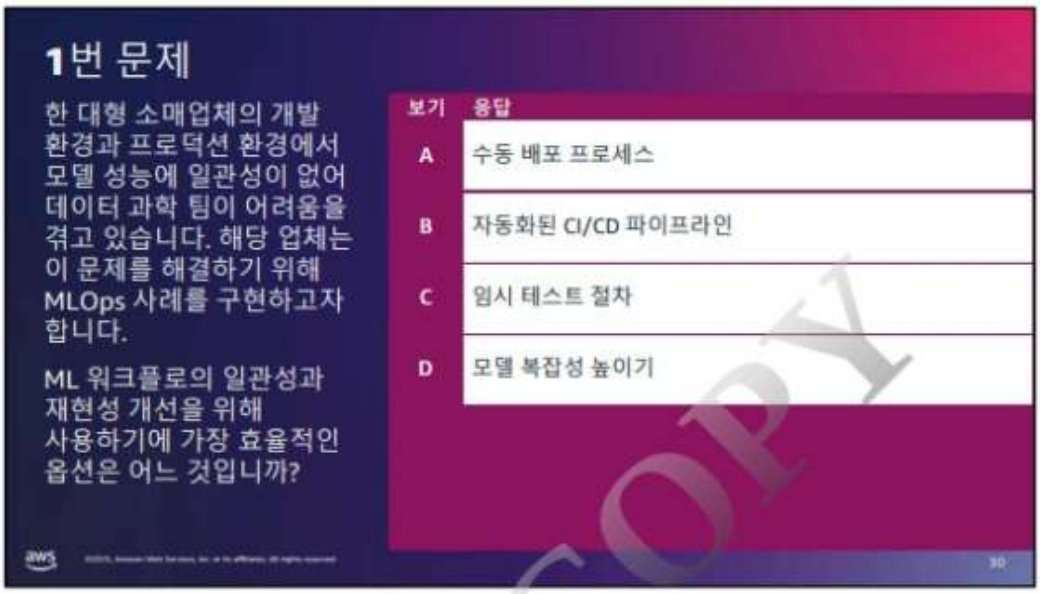
Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

이 섹션에서는 이 모듈에서 다루었던 개념을 정확하게 이해하고 있는지를 평가하는 문제를 풀어 봅니다. 모듈에서 배운 내용에 따라 각 문제의 정답을 선택하십시오. 이러한 문제를 풀어 보면 주요 개념을 더욱 정확하게 이해하고 더 복습해야 하는 부분을 파악할 수 있습니다.

DO NOT COPY
ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

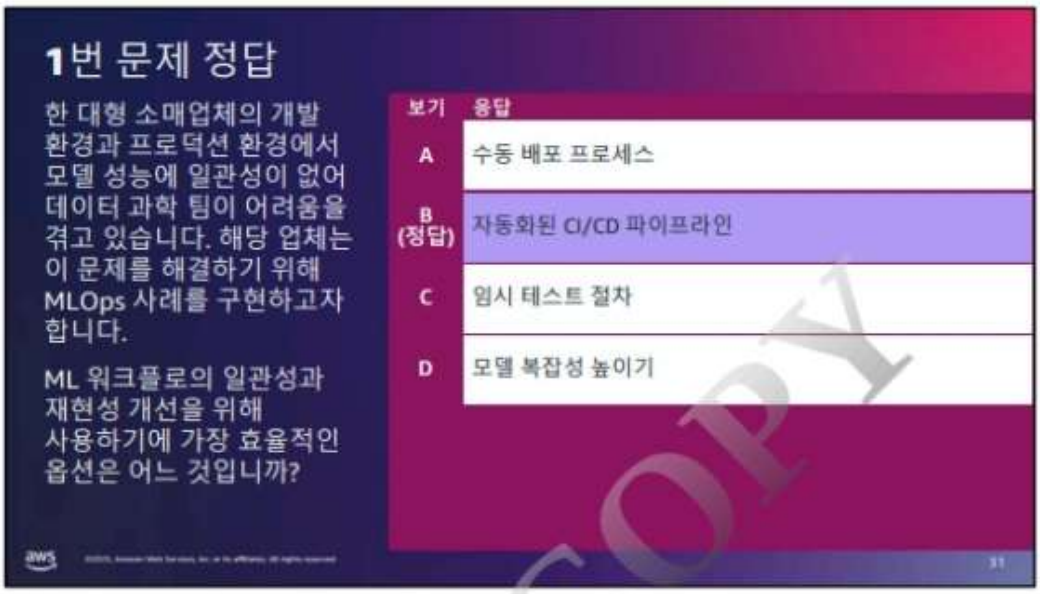
Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



ML 워크플로 자동화의 이점을 이해하고 있는지를 평가하기 위한 문제입니다.

DO NOT COPY
ssleeeee1@gmail.com

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



정답은 B, 자동화된 CI/CD 파이프라인입니다.

ML 워크플로의 일관성과 재현 가능성을 높일 수 있는 가장 효율적인 방식은 자동화된 CI/CD 파이프라인입니다. 이러한 파이프라인에서는 항상 동일한 구축, 테스트 및 배포 프로세스가 진행되므로 개발 환경과 프로덕션 환경의 불일치 사항이 감소합니다.

오답은 다음과 같습니다.
A. 수동 배포 프로세스는 인적 오류가 발생하기 쉬우며 일관성도 떨어지므로 재현 가능성이 개선되지 않습니다.
D. 임시 테스트 절차는 표준화되어 있지 않으므로 여러 환경에서 일관된 결과가 제공되지 않습니다.
E. 더 복잡한 모델을 생성한다고 해서 핵심적인 문제인 환경 간 불일치가 해결되지는 않습니다.

DO NOT COPY ssleeeee1@gmail.com

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

ML 워크플로 자동화용 AWS 서비스에 대한 지식을 평가하기 위한 문제입니다.

DO NOT COPY
ssleeeee1@gmail.com

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

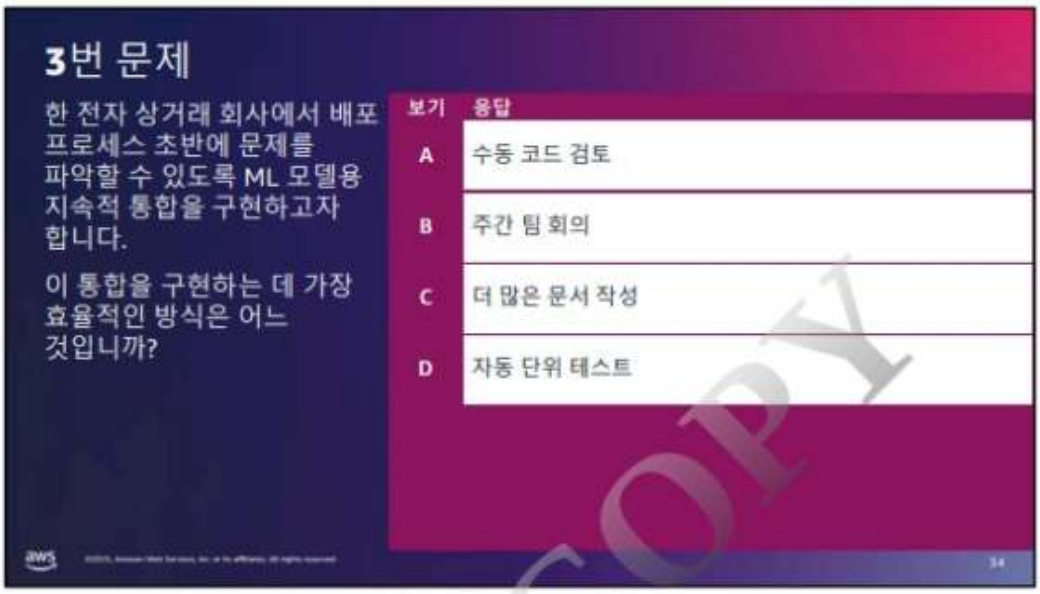
정답은 B, Amazon SageMaker Pipelines입니다.

Amazon SageMaker Pipelines는 데이터 준비, 모델 훈련, 평가를 비롯한 ML 워크플로를 자동화할 수 있는 서비스입니다. SageMaker Pipelines에서는 새 데이터가 제공될 때 자동으로 시작할 수 있는 재사용 가능 ML 파이프라인을 생성할 수 있습니다.

오답은 다음과 같습니다.
A. AWS CodePipeline은 ML 워크플로 전문 서비스가 아닌 일반 소프트웨어 전달 및 배포에 주로 사용되는 서비스입니다.
D. AWS CloudFormation은 ML 워크플로 자동화용 서비스가 아닌 코드형 인프라 서비스입니다.
E. Amazon EMR은 ML 파이프라인 자동화용 서비스가 아닌 빅 데이터 처리용 서비스입니다.

DO NOT COPY
ssleeeee1@gmail.com

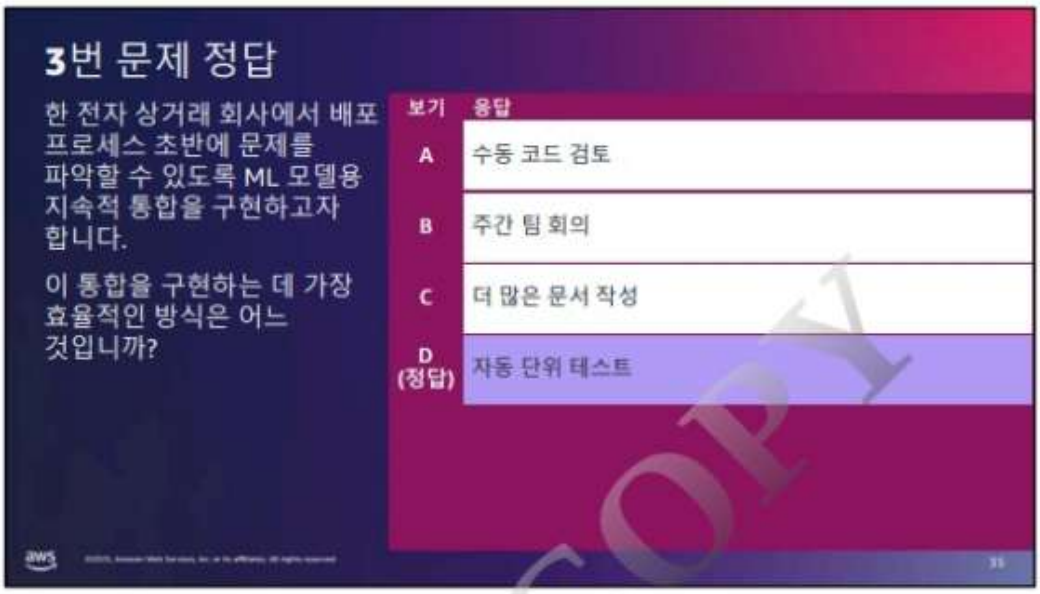
Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



ML 프로젝트의 CI/CD 사례 관련 지식을 평가하기 위한 문제입니다.

DO NOT COPY
ssleeeee1@gmail.com

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



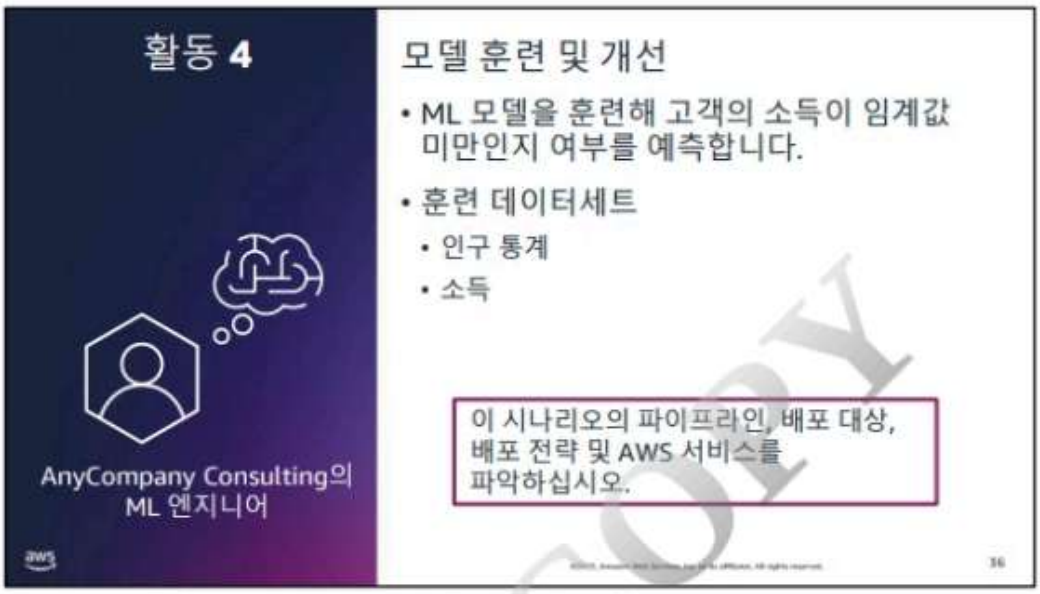
정답은 D, **자동 단위 테스트입니다.**

ML 프로젝트의 지속적 통합을 구현할 때 가장 효율적인 방식은 자동 단위 테스트입니다. 이러한 테스트를 실행하면 변경을 수행하는 즉시 코드, 데이터 처리 단계 또는 모델 구성 요소의 문제를 빠르게 찾아낼 수 있으므로 문제를 조기에 탐지하여 해결할 수 있습니다.

오답은 다음과 같습니다.
A. 수동 코드 검토도 유용하기는 하지만 시간이 많이 걸리며 특히 복잡한 ML 시스템에서는 모든 문제를 찾아내지 못할 수도 있습니다.
D. 주간 팀 회의는 협업을 진행하는 경우 꼭 필요한 요소이기는 하지만 지속적 통합에 필요한 피드백이 즉시 제공되지는 않습니다.
E. 문서를 더 많이 작성하는 방식도 효율적이기는 하지만 문서 수가 늘어난다고 해서 개발 프로세스 초기에 문제를 찾아낼 수 있는 것은 아닙니다.

DO NOT COPY ssleeeee1@gmail.com

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

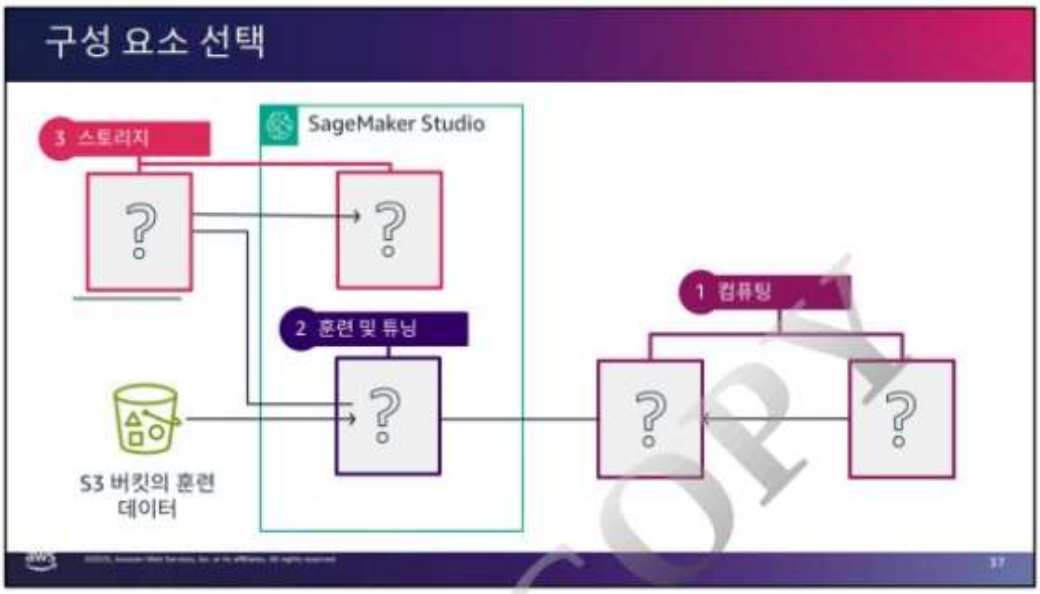


이 활동에서는 다음 시나리오에 적용할 기법과 도구를 선택합니다.

여러분이 근무하고 있는 마케팅 업체에서 판매량을 늘리고자 합니다. 이 업체는 대규모 마케팅 캠페인을 계획 중이며 해당 캠페인의 투자 수익률을 최대한 높이고자 합니다. 연간 소득이 5만 달러 미만인 고객이 프로모션 대상입니다. 제공된 인구 통계 데이터를 토대로 특정 고객의 연 소득이 이 임곗값보다 낮은지 여부를 예측하도록 ML 모델을 훈련해야 합니다. 데이터세트에는 대규모 성인 모집단의 인구 통계 및 급여 데이터가 포함되어 있습니다.

DO NOT COPY
ssleeeee1@gmail.com

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

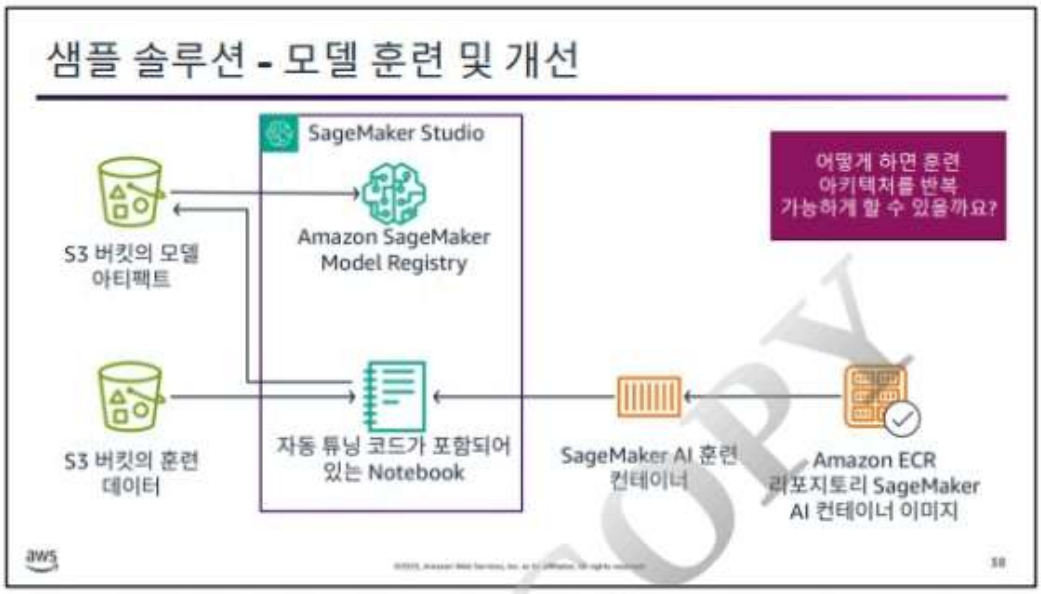


훈련 및 튜닝 단계가 시작되면 SageMaker Studio에서 모델 개발 전용 환경이 제공됩니다. 컴퓨팅 리소스와 데이터 스토리지가 '연결'되는 이 중간 단계에서 훈련 과정의 중요 태스크인 모델 훈련 및 파라미터 최적화 프로세스가 진행됩니다.

마지막으로 스토리지 구성 요소에는 2개 요소가 연결되어 있습니다. 이러한 요소는 S3 버킷에 저장되어 있는 훈련 데이터를 사용합니다. 이 스토리지 인프라를 채택하면 기계 학습 워크플로 전반에서 데이터를 적절하게 관리하고 액세스할 수 있습니다. 이와 같이 구성 요소를 선택하면 SageMaker Studio 환경 설정이 완료됩니다.

DO NOT COPY ssleeeee1@gmail.com

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



이 슬라이드에는 이전 활동 슬라이드의 샘플 해결 방법이 나와 있습니다. 이 샘플 해결 방법에는 모델 훈련 및 워크플로 개선에 사용할 수 있는 아키텍처 하나가 대략적으로 표시되어 있습니다. 해당 워크플로에서는 먼저 Amazon S3에서 데이터 수집을 수행합니다. Amazon S3에는 인구 통계 및 소득 정보가 포함된 원시 CSV 데이터세트가 저장되어 있습니다.

이 샘플 해결 방법에 포함되어 있는 3가지 주요 AWS 변환 서비스는 Amazon SageMaker Data Wrangler, Amazon EMR, SageMaker Processing입니다. 이러한 도구 중 하나 이상을 사용하여 일반적인 데이터 처리 단계를 수행합니다. 이러한 단계로는 누락된 데이터 처리, 이상치 제거, 형식과 단위 표준화, 대상 열을 이진으로 변환 등이 있습니다.

변환의 마지막 단계에서는 처리된 데이터세트를 내보냅니다. 이 예에서는 모델 훈련에 바로 사용할 수 있는 CSV 형식으로 Amazon S3에 데이터가 저장됩니다. 이 워크플로에서는 여러 AWS 서비스를 함께 사용하여 기계 학습 프로젝트용으로 효율적인 확장 가능 데이터 준비 파이프라인을 생성하는 방법을 제시합니다.

DO NOT COPY
ssleeeee1@gmail.com

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

방식 1의 단일 파이프라인에서는 모든 아티팩트가 연속 프로세스 하나를 통해 전달됩니다. 그러므로 각 아티팩트가 생성되거나 변환되는 정확한 위치를 찾아내기가 더 어려워집니다. 파이프라인 시작 부분에서 원시 데이터가 입력되고 끝부분에서 평가 지표가 제공된다는 것은 알 수 있지만 중간 아티팩트(데이터 클리닝, 특성, 모델 아티팩트)가 통합 '데이터 준비 및 모델 개발' 단계에서 처리되는 시점은 정확하게 알 수 없습니다.

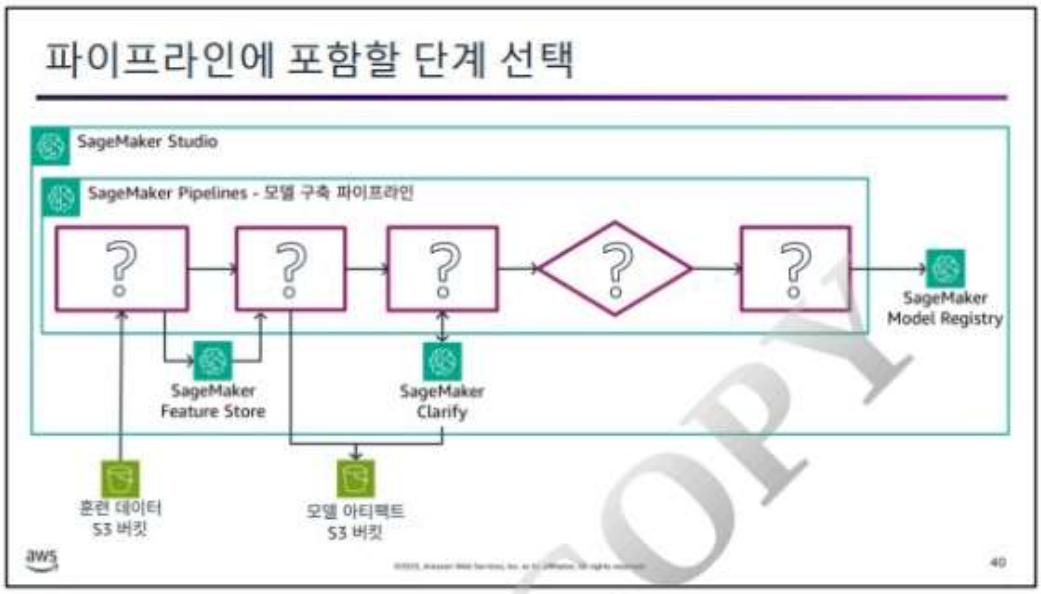
방식 2에서는 아티팩트를 좀 더 명확하게 추적할 수 있습니다. 원시 데이터와 클린 데이터 및 특성은 첫 단계인 '데이터 준비 및 특성 추출'에서 처리되며 모델 아티팩트와 평가 지표는 두 번째 단계인 '모델 구축 및 평가'에서 관리됩니다.

아티팩트를 가장 정확하게 매핑할 수 있는 파이프라인은 방식 3입니다. 원시 데이터와 클린 데이터는 '데이터 준비' 단계에서, 특성은 '특성 추출' 단계에서, 모델 아티팩트는 '모델 구축' 단계에서, 평가 지표는 '모델 평가' 단계에서 처리되므로 각 아티팩트의 수명 주기를 가장 쉽게 모니터링 및 관리할 수 있습니다.

DO NOT COPY
ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

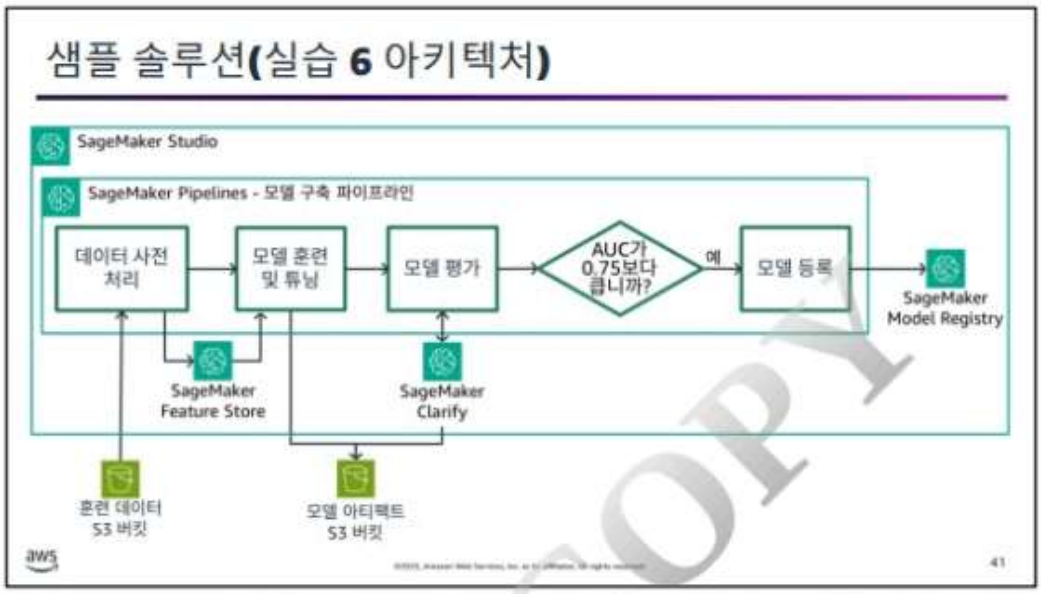


1. 데이터 사전 처리
   - 훈련 데이터로 시작
   - 모델용 데이터 정리 및 준비
   - 데이터 품질 및 형식 보장
2. 모델 훈련 및 튜닝
   - 초기 모델 훈련
   - 하이퍼파라미터 조정
   - 모델 성능 최적화
3. 모델 평가
   - 모델 성능 테스트
   - 임계값과 지표 비교(AUC가 0.75보다 큰지 확인)
   - 결과 검증
4. 모델 등록
   - 지표가 임계값을 충족하면 SageMaker Model Registry에 모델 등록
   - 모델 버전 및 파라미터 정보를 문서로 작성
   - 모델 아티팩트 저장

참고: 각 단계는 이전 단계의 결과를 토대로 진행되므로 원시 데이터 제공 단계에서 프로덕션 환경에 배포 가능한 모델 등록까지의 전체 파이프라인이 생성됩니다.

DO NOT COPY
ssleeeee1@gmail.com

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

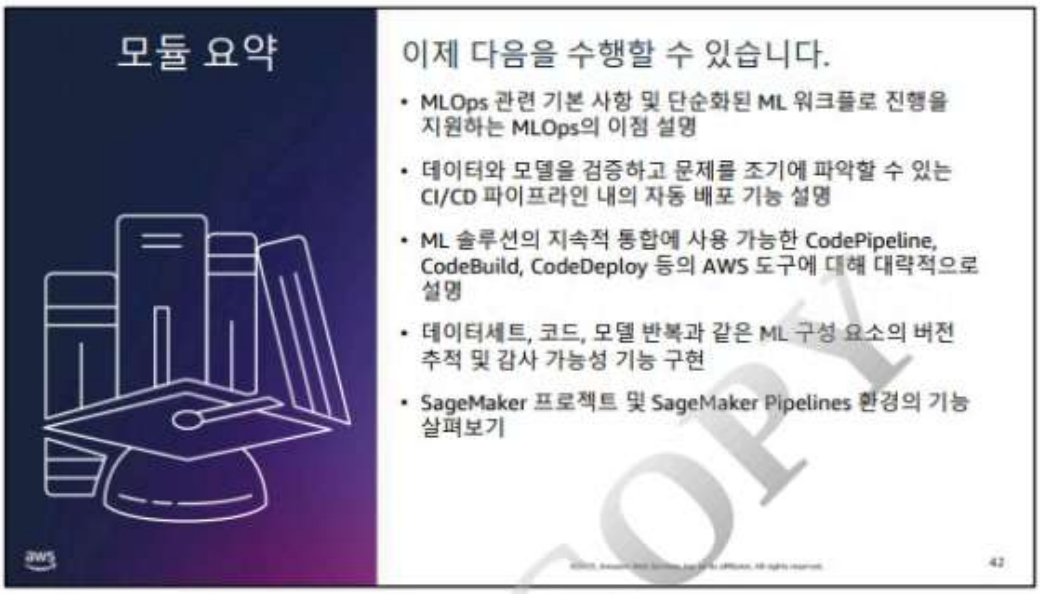


SageMaker Model Build Pipeline은 훈련 데이터 및 모델 아티팩트용 S3 버킷 2개를 활용하여 SageMaker Studio 내에서 전체 ML 워크플로를 구축합니다. 특성 관리 서비스인 Feature Store 및 모델 투명성 보장 서비스인 Clarify와 통합되는 이 파이프라인에서는 사전 처리에서 훈련과 평가로 이어지는 체계적인 과정이 진행됩니다. 그리고 모델 등록 전에 중요 결정 지점(AUC 점수가 0.75보다 높아야 함)을 거칩니다.

실습 6의 템플릿이기도 한 이 아키텍처는 통합 품질 검사 기능, 중앙 집중식 특성 관리 기능 및 기본 제공 모니터링 도구가 포함된 자동 워크플로를 제시합니다. 이 파이프라인에서 제공하는 프레임워크에서는 성능 임곗값을 충족하는 모델만 Model Registry에 등록됩니다. 그러므로 개발 프로세스 전반에서 모델의 투명성과 관리 용이성이 유지됩니다.

DO NOT COPY
ssleeeee1@gmail.com

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



이 모듈을 완료했으므로 이제 다음을 수행할 수 있습니다.

- MLOps 관련 기본 사항 및 단순화된 ML 워크플로 진행을 지원하는 MLOps의 이점 설명
- 데이터와 모델을 검증하고 문제를 조기에 파악할 수 있는 CI/CD 파이프라인 내의 자동 배포 기능 설명
- ML 솔루션의 지속적 통합에 사용 가능한 CodePipeline, CodeBuild, CodeDeploy 등의 AWS 도구에 대해 대략적으로 설명
- 데이터세트, 코드, 모델 반복과 같은 ML 구성 요소의 버전 추적 및 감사 가능성 기능 구현
- SageMaker 프로젝트 및 SageMaker Pipelines 환경의 기능 살펴보기

DO NOT COPY
ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

이 모듈에서 알아본 데이터 준비 및 특성 추출 관련 개념을 잠시 생각해 보시기 바랍니다. 이 과정에서 다룬 자료 또는 이 과정에서 배운 기술을 실제 업무에 적용하는 방법과 관련하여 궁금한 점이 있으면 지금 질문을 할 수 있습니다.

DO NOT COPY
ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

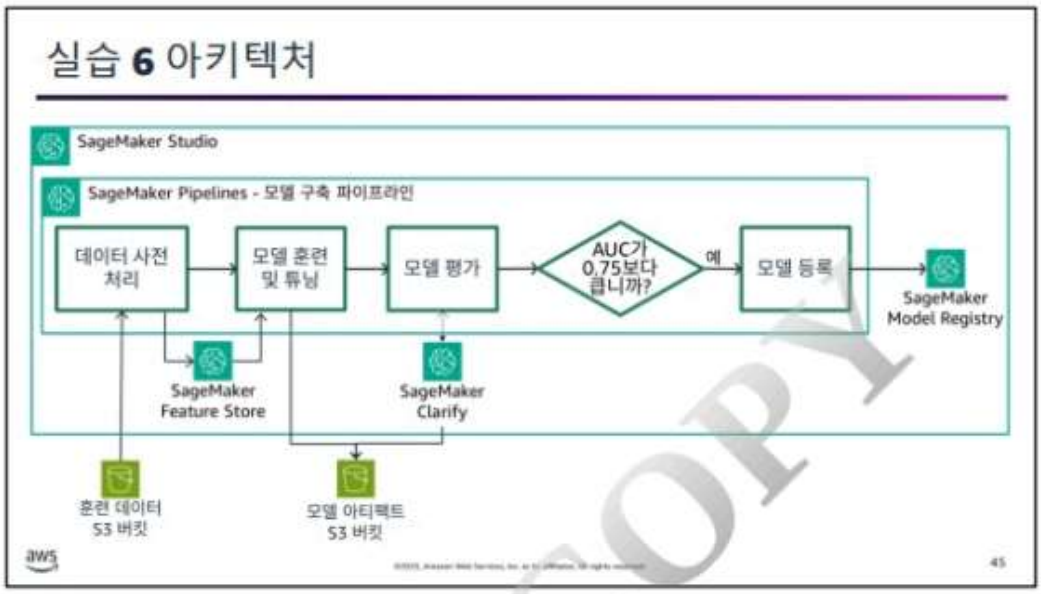
Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

이전 실습에서 프로덕션 환경에 모델을 배포하고 새 데이터로 모델을 테스트했으므로 이제 모델 크기를 조정합니다. 고객에게 가장 적합한 업무용 워크플로를 생성하려고 합니다. 모델 크기를 조정하기 위해 SageMaker Studio와 함께 SageMaker Pipelines 및 SageMaker Model Registry를 사용할 예정입니다.

이 실습에서는 데이터 로드, 데이터 변환, 훈련 및 튜닝 등 자동 기계 학습 워크플로의 여러 단계를 관리합니다. 그리고 Model Registry를 사용하여 훈련된 모델을 저장합니다.

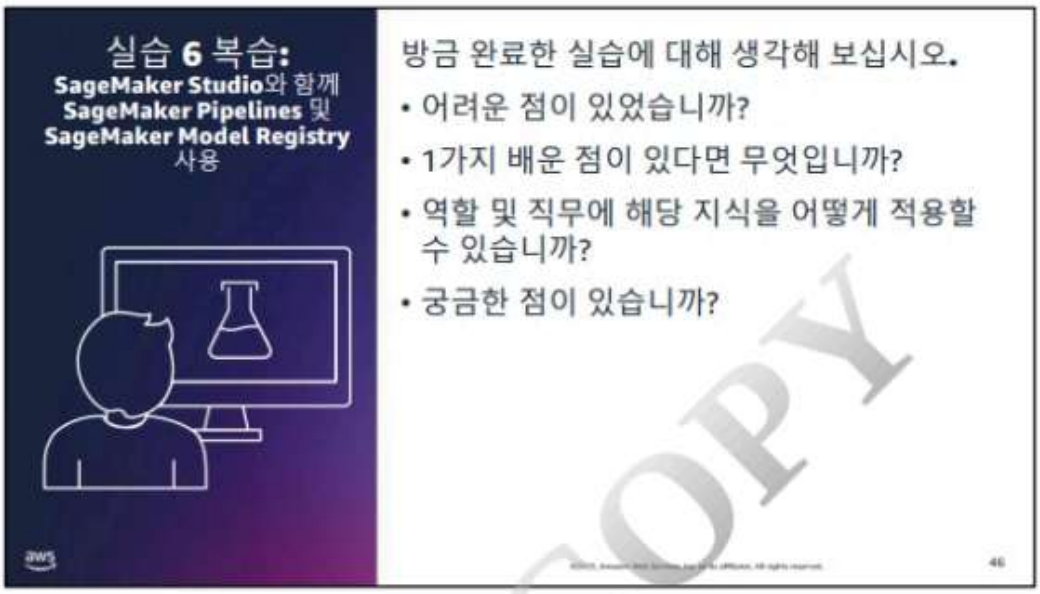
DO NOT COPY
ssleeeee1@gmail.com

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



DO NOT COPY
ssleeeee1@gmail.com

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



DO NOT COPY
ssleeeee1@gmail.com

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

기계 학습(ML) 수명 주기는 데이터 준비, 모델 개발, 배포, 모니터링 등의 단계를 수행하는 지속적인 프로세스입니다. 모니터링 단계에서는 시간이 지나도 ML 솔루션의 높은 효율성이 유지되는지를 확인할 수 있습니다.

이 모듈에서는 모델의 성능 및 데이터 품질을 모니터링하는 데 사용할 수 있는 도구와 기법에 대해 알아봅니다. 먼저 4가지 드리프트 유형인 데이터 품질, 모델 품질, 통계적 편향 및 특성 기여도 드리프트를 탐지하는 방법을 살펴봅니다. 이러한 모니터링 기법을 활용하면 모델의 정확도와 성능에 영향을 줄 수 있는 문제를 파악하여 해결할 수 있습니다.

이 모듈에서는 지속적인 모니터링, 알림 설정, 자동화된 문제 해결 전략 구현 관련 모범 사례에 대해서도 살펴봅니다. 이러한 모범 사례를 준수하면 프로덕션 환경에서 ML 솔루션의 무결성과 효율성을 유지할 수 있습니다.

DO NOT COPY ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

기계 학습 수명 주기의 이 단계를 시작할 때는 배포된 모델이 프로덕션 환경에서 실행되고 있는 상태입니다. 이 단계에서는 해당 솔루션을 모니터링해야 합니다.

DO NOT COPY
ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

이 모듈에서는 다음의 4가지 주요 주제에 대해 설명합니다.

A. ML 모델의 드리프트 탐지: 모델 성능에 영향을 줄 수 있는 다양한 드리프트 유형을 살펴보고 드리프트를 식별하는 방법을 알아봅니다.
B. SageMaker Model Monitor: Amazon SageMaker Model Monitor에서 ML 모델을 모니터링하는 방법을 알아봅니다.
C. 데이터 품질 및 모델 품질 모니터링: 시간이 지나도 입력 데이터와 모델 출력의 우수한 품질이 유지되도록 하는 데 사용할 수 있는 구체적인 기법을 살펴봅니다.
D. 자동 문제 해결: 마지막으로 발생하는 문제를 자동으로 해결할 수 있는 전략 및 문제를 해결하는 방법에 대해 알아봅니다.

이 모듈을 마치면 프로덕션 환경에서 ML 솔루션을 효율적으로 모니터링 및 유지 관리하는 방법을 포괄적으로 파악할 수 있습니다.

DO NOT COPY ssleeeee1@gmail.com

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

기계 학습에서 **드리프트**란 입력 변수와 출력 변수 간의 관계가 점진적으로 변화하는 현상입니다. 드리프트가 발생하면 시간 경과에 따라 모델 성능이 저하될 수 있습니다. 드리프트를 탐지해야 하는 이유는 모델 예측 정확도나 신뢰도가 낮아지는 상황을 파악할 수 있기 때문입니다.

이 섹션에서는 다음과 같은 여러 유형의 드리프트를 살펴봅니다.

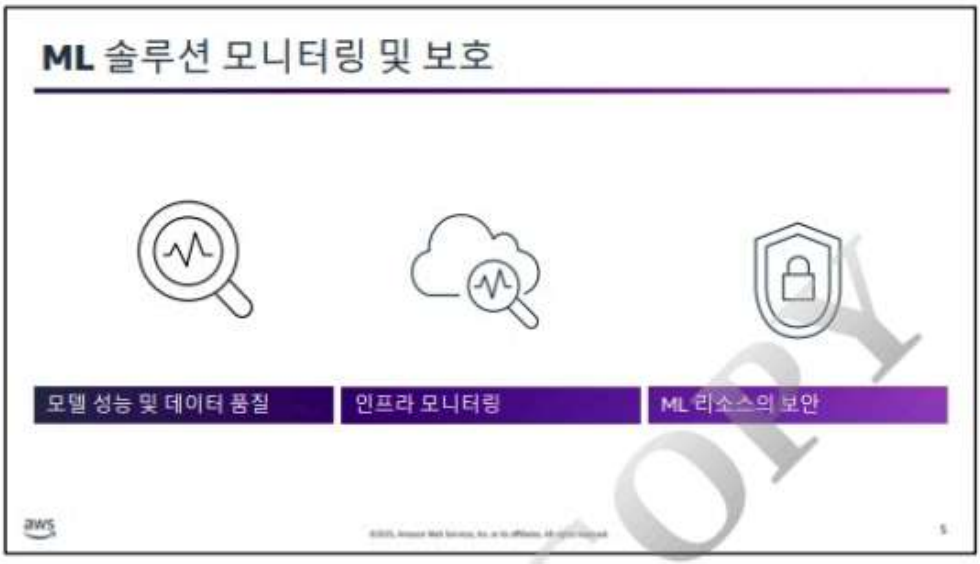
- 데이터 품질 드리프트: 입력 데이터의 통계적 속성 변화
- 모델 품질 드리프트: 모델의 예측 성능 저하
- 편향 드리프트: 여러 하위 그룹에서 걸친 모델의 공정성 변화
- 특성 기여도 드리프트: 예측 생성 시 각 특성의 중요도 변화

이러한 여러 드리프트 유형을 파악하면 더욱 효율적인 모니터링 전략을 구현할 수 있으며 장기적으로는 ML 모델의 우수한 성능을 유지할 수 있습니다.

DO NOT COPY ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



ML 솔루션이 안정적으로 작동하도록 하려면 효율적인 모니터링 및 보안 전략이 필요합니다. 중점적으로 모니터링해야 하는 3가지 주요 영역은 다음과 같습니다.

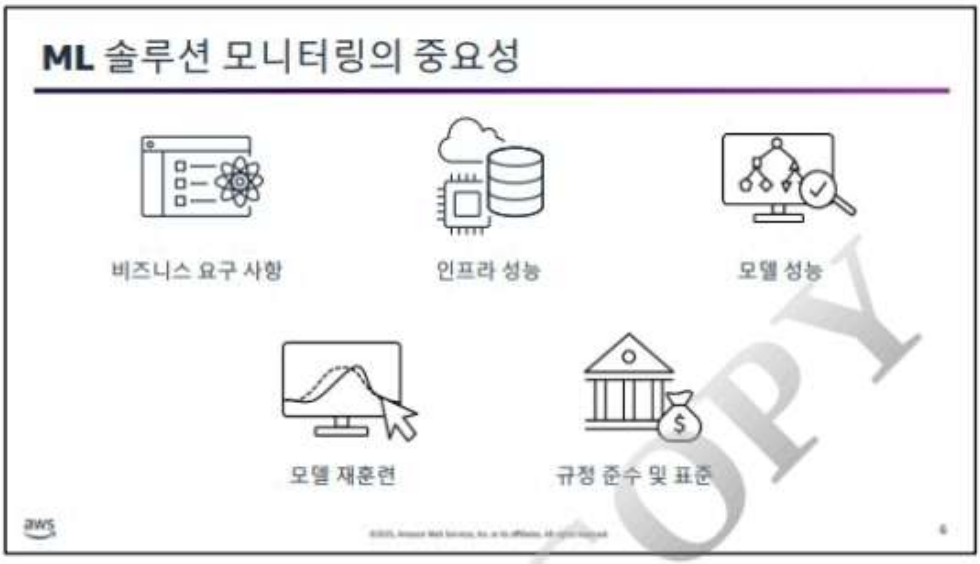
- 모델 성능 및 데이터 품질: 모델의 정확도와 신뢰성을 유지하기 위해 입력 데이터 품질과 모델 예측을 지속적으로 모니터링해야 합니다. 구체적으로는 정밀도, 재현율, F1 점수 등의 지표를 추적하는 동시에 데이터 드리프트도 모니터링해야 합니다.
- 인프라 모니터링: ML 시스템이 사용하는 컴퓨팅 리소스를 모니터링해야 합니다. 구체적으로는 CPU 및 그래픽 처리 장치(GPU) 사용률, 메모리 사용량, 네트워크 처리량 등을 모니터링해야 합니다.
- ML 리소스 보안: 강력한 액세스 제어, 암호화 및 규정 준수 조치를 구현하여 ML 수명 주기 전반에서 데이터, 모델 및 인프라를 보호해야 합니다. 그래야 ML 솔루션의 무결성과 기밀성을 유지할 수 있습니다.

이러한 영역을 중점적으로 모니터링하면 ML 시스템의 효율성과 신뢰성을 유지할 수 있는 포괄적인 모니터링 및 보안 전략을 수립할 수 있습니다.

DO NOT COPY
ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



ML 모델과 솔루션을 도입하는 영역이 늘어나고 있는 오늘날에는 이러한 모델을 적절하게 모니터링 및 감독해야 합니다. 실제 응용 분야에서 ML 솔루션을 배포할 때는 모니터링도 반드시 수행해야 합니다. 그래야 모델의 성능과 신뢰성을 최적 상태로 유지하고 윤리적 표준도 준수할 수 있기 때문입니다.

비즈니스 요구 사항 충족
모니터링은 ML 솔루션을 조직의 비즈니스 목표에 맞추는 데 중요한 역할을 합니다. 모니터링은 조직이 원하는 결과를 달성하고 있는가?라는 질문에 답하는 데 도움이 됩니다. 기업은 종합적인 시스템 성능 지표를 분석하여 투자 수익률(ROI)을 평가하고, 리소스 할당 및 전략적 조정과 관련하여 데이터에 기반한 결정을 내릴 수 있습니다.

인프라 성능 모니터링
ML 솔루션에는 상당한 컴퓨팅 리소스뿐 아니라 시스템의 효율적 운영과 수요에 따른 적절한 규모 조정을 위한 인프라 성능 모니터링이 필요한 경우가 많습니다. 이 프로세스에는 CPU 및 GPU 사용률, 메모리 소비, 네트워크 처리량 같은 주요 인프라 지표의 모니터링이 포함됩니다.

모델 성능 평가
모델 모니터링은 실제 시나리오에서 기계 학습 모델의 성능을 평가하는 프로세스입니다. 모델 모니터링은 관련 지표를 사용하여 모델이 처음 보는 데이터에 얼마나 잘 일반화되는지 측정함으로써 모델의 정확성, 효율성, 전반적인 효과를 평가합니다. 사용되는 구체적 지표는 회귀나 분류 같은 문제 유형에 따라 달라집니다. 예를 들어 분류 태스크의 주요 지표에는 정밀도, 재현율, F1 점수, 추론 시간이 포함될 수 있습니다. 또한 모니터링은 엣지 케이스와 잠재적 장애 모드에서도 모델의 성능을 평가해야 합니다.

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

모니터링을 통해 잠재적 문제를 조기에 탐지하면 적시에 개입하여 문제를 해결하고 최적의 성능을 유지할 수 있습니다.

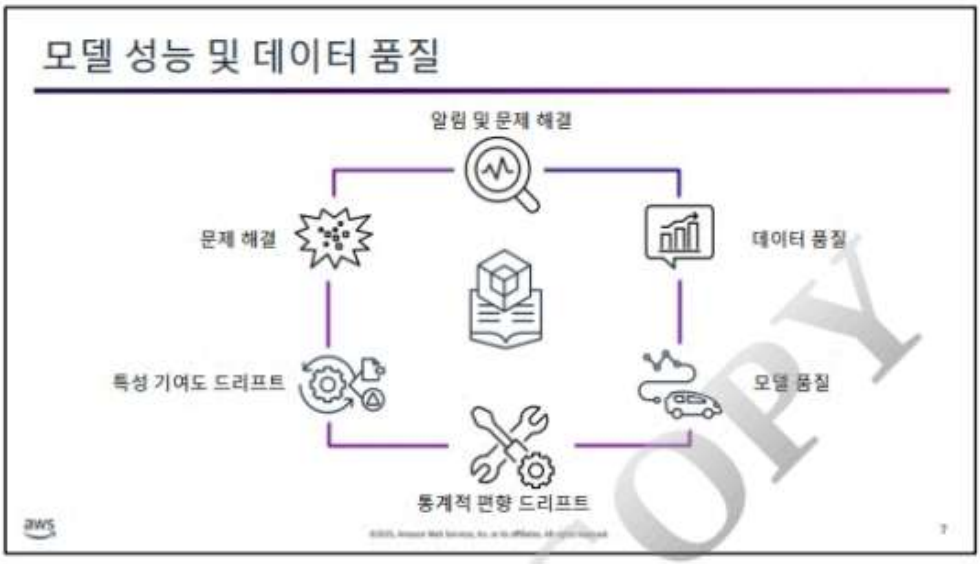
모델 재훈련 촉진
데이터가 진화하고 외부 요인이 변화함에 따라 재훈련되지 않는 모델은 낡고 신뢰성이 떨어질 수 있습니다. 조직은 데이터 분포 변화와 예측 성능 같은 주요 지표를 추적하여 모델의 효과를 유지하기 위한 시기적절한 재훈련 일정을 예약할 수 있습니다.

규정 및 산업 표준 준수 보장
ML 솔루션은 관련 규칙과 규정을 준수해야 합니다. 모니터링 프로세스는 사용자 개인 정보, 보안, 공정성을 보호합니다. 이는 AI와 ML이 엄격한 규칙과 감독에 따라야 하는 의료 및 금융 같은 산업에서 중요합니다. 조직에는 사용자 데이터를 보호하는 모니터링 프레임워크가 있어야 합니다. 이 프레임워크는 민감한 정보의 무단 액세스나 악용을 방지합니다. 이 프레임워크는 데이터 보호와 윤리적 AI 및 ML 원칙을 따릅니다. 조직은 책임을 지고 신뢰를 구축함으로써 AI 및 ML 기술을 책임감 있게 사용할 수 있습니다.

DO NOT COPY
ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

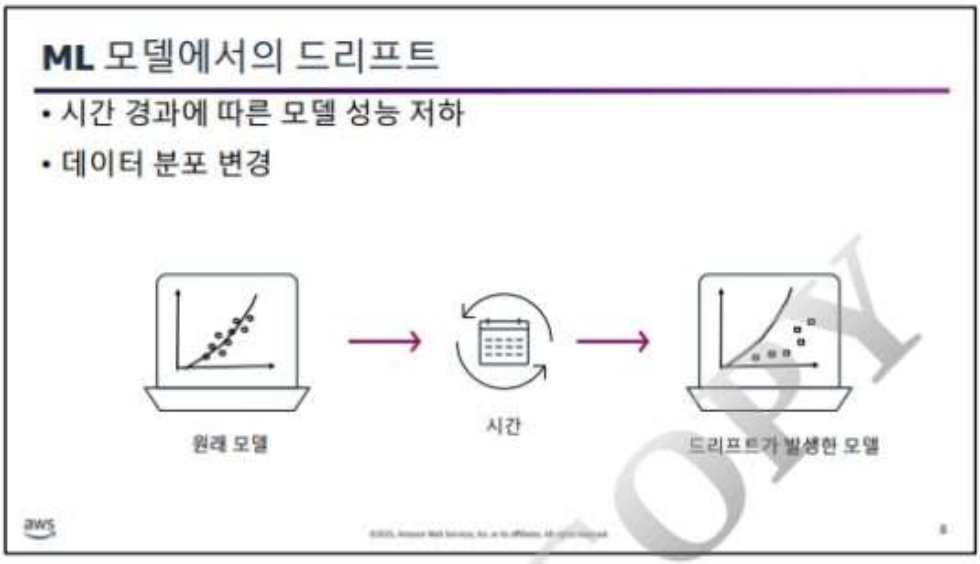


ML 모델이 정확한 결과를 생성하는지 여부는 입력 데이터 품질의 품질에 따라 달라집니다. 데이터는 시간에 따라 변하기 때문에 모델의 정확도와 성능 문제를 지속적으로 탐지, 수정, 완화하려면 모니터링이 필요합니다. 이 모니터링 단계에서 정제된 최신 데이터를 사용하여 모델을 재훈련해야 할 수도 있습니다.

포괄적인 모니터링 프로세스를 구현하면 데이터 분포 변화, 모델 성능 저하 또는 편향 유입 같은 잠재적 문제를 조기에 탐지할 수 있습니다. 이 프로세스를 진행하면 ML 솔루션의 신뢰성과 효율성 유지를 위한 사전 조치를 취할 수 있습니다.

DO NOT COPY
ssleeeee1@gmail.com

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



ML 모델에서 **드리프트**란 기본 데이터 분포나 특성 및 대상 변수 간의 관계 변화로 인해 시간 경과에 따라 모델 성능이 저하되는 현상입니다. 이러한 현상은 ML 솔루션의 정확도와 신뢰성에 큰 영향을 줄 수 있습니다.

드리프트가 발생되는 일반적인 원인이 아래에 나와 있습니다.
- 사용자 행동 방식이나 선호도 변화
- 시장 상황 변화 또는 외부적인 요인
- 계절별 데이터 패턴 변화
- 데이터 수집 프로세스 수정

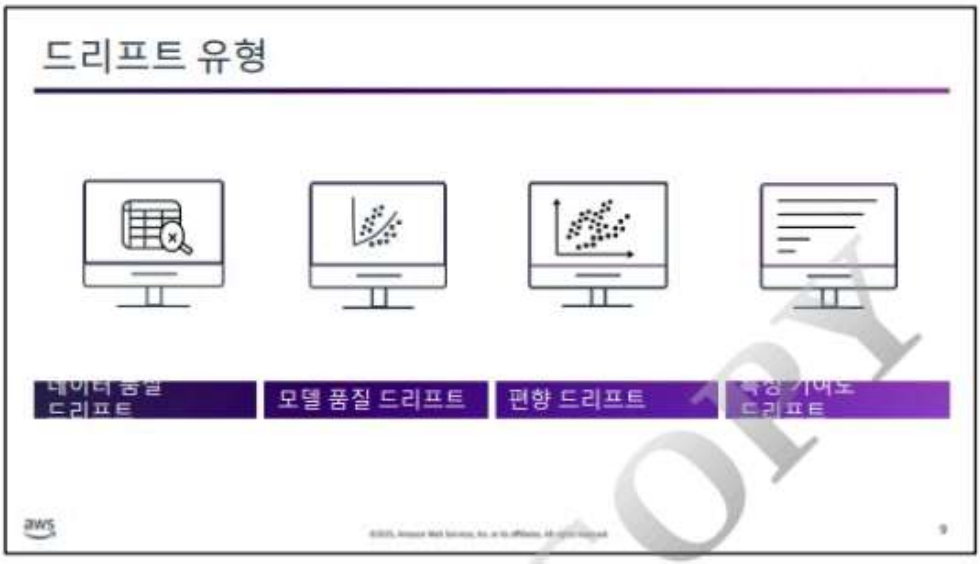
데이터 분포가 바뀌면 모델이 다른 데이터세트에서 훈련되었기 때문에 모델 성능이 저하되기 시작할 수 있습니다. 이처럼 성능이 저하되면 정확도, 정밀도, 재현율 또는 기타 관련 지표가 낮아질 수 있습니다.

정기적인 모니터링 및 재훈련 전략을 수립하면 드리프트의 영향을 완화할 수 있으며, 그러면 시간이 지나도 모델이 정확한 예측을 제공할 수 있습니다.

DO NOT COPY ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



ML 모델에 영향을 줄 수 있는 4가지 주요 드리프트 유형은 다음과 같습니다.

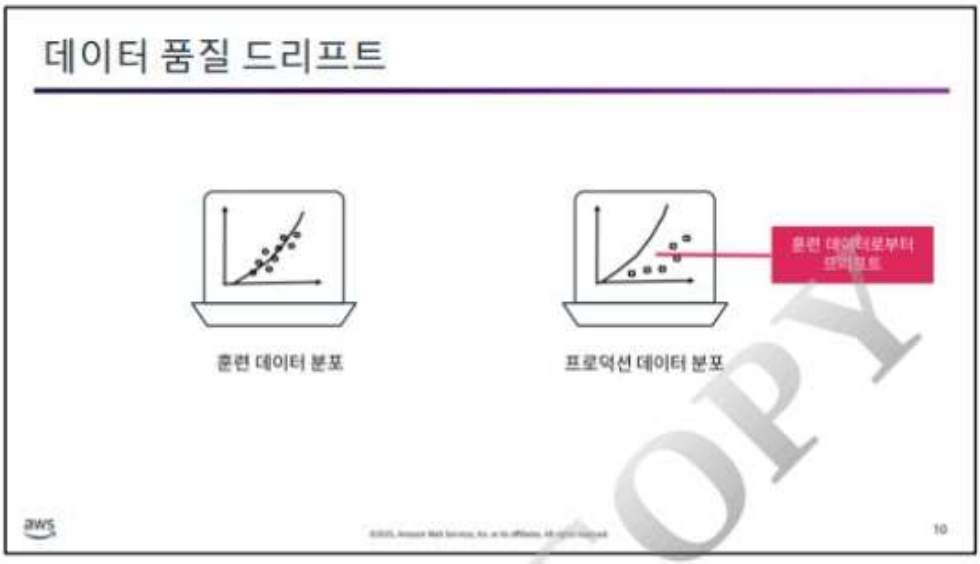
- **데이터 품질 드리프트**: 시간 경과에 따라 입력 데이터의 통계적 속성이 바뀌는 경우에 발생합니다. 특성 분포 변화, 새로운 범주 추가, 데이터 수집 프로세스 변경 등을 예로 들 수 있습니다.
- **모델 품질 드리프트**: 모델의 예측 성능이 저하되는 현상입니다. 예측 성능은 정해진 정답 레이블에 예측을 비교하는 방식으로 측정하는 경우가 많습니다.
- **편향 드리프트**: 시간이 경과하면서 모델이 불공평하거나 차별적인 예측을 점점 많이 반환하는 현상입니다. 이러한 드리프트는 특히 모집단 내의 특정 하위 그룹에 영향을 주게 됩니다.
- **특성 기여도 드리프트**: 개념 드리프트라고도 합니다. 시간이 경과하면서 예측 시에 반영되는 특성의 상대적 중요도가 바뀌면 이러한 드리프트 유형이 발생합니다.

이러한 드리프트 유형을 탐지하려면 데이터 및 모델 성능을 지속적으로 모니터링해야 합니다. 통계 테스트, 시각화 도구, 자동화된 모니터링 시스템 등의 기법을 활용하면 드리프트를 조기에 식별하여 시정 조치를 취함으로써 ML 솔루션의 신뢰성을 유지할 수 있습니다.

DO NOT COPY ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



데이터 품질 드리프트는 시간이 경과하면서 입력 데이터의 통계적 속성이 변경되어 모델 훈련에 사용된 데이터의 특성과 달라지는 현상입니다. 이러한 드리프트 유형은 모델 성능에 큰 영향을 줄 수 있습니다. 데이터 품질 드리프트가 발생한 모델은 새로운 데이터 분포에 사용하기에 더 이상 적합하지 않을 수도 있기 때문입니다.

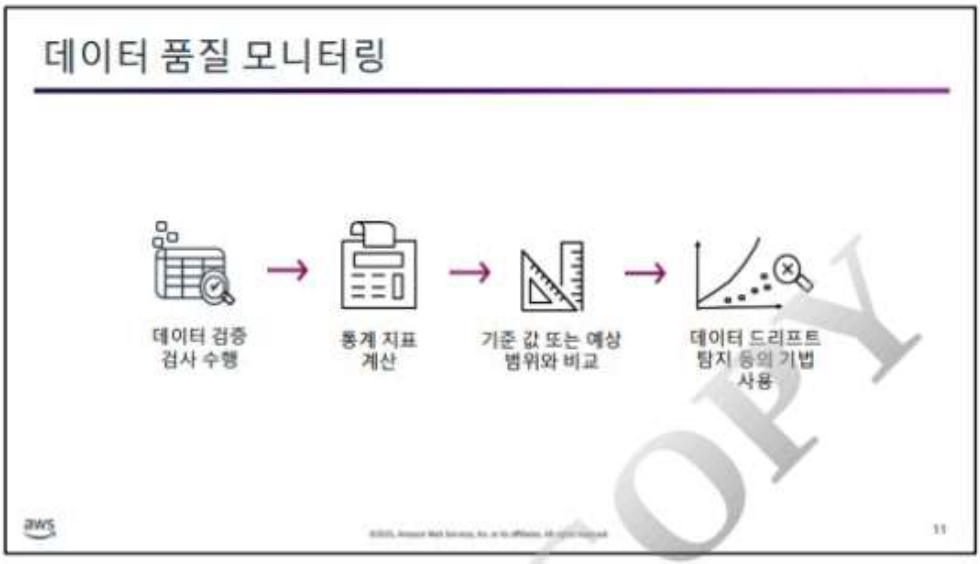
데이터 품질 드리프트를 발생시키는 일반적인 원인이 아래에 나와 있습니다.
- 데이터 수집 프로세스 또는 출처 변화
- 사용자 행동 방식이나 선호도 변화
- 계절별 데이터 패턴 변화
- 새 범주 또는 특성 추가

데이터 품질 드리프트를 탐지하기 위해 수행할 수 있는 작업은 다음과 같습니다.
- 수신 데이터의 통계적 속성(예: 평균, 분산, 분포)을 모니터링합니다.
- 중요한 변경 사항 관련 알림을 제공하는 자동 모니터링 시스템을 구현합니다.

데이터 품질 드리프트를 해결할 때는 최신 데이터로 모델을 재훈련하거나 새로운 데이터 특성을 반영하여 데이터 사전 처리 파이프라인을 업데이트하는 경우가 많습니다. ML 모델을 정기적으로 모니터링하고 사전 조치를 취하면 데이터 분포가 변화하더라도 ML 모델의 정확도와 신뢰성을 유지할 수 있습니다.

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



기계 학습 모델의 우수한 신뢰성과 성능을 장기적으로 유지하려면 데이터 품질을 모니터링해야 합니다. 데이터 품질 모니터링의 주요 단계는 다음과 같습니다.

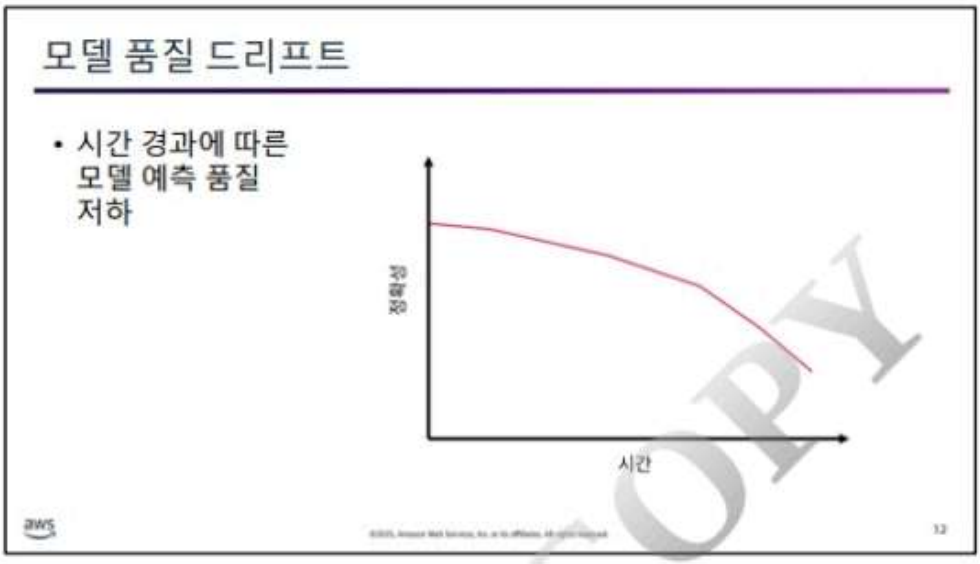
1. 입력 데이터를 모델에 공급하기 전에 입력 데이터에서 데이터 검증 검사를 구현하여 데이터가 예상 형식, 범위, 품질 요구 사항을 충족하는지 확인합니다. 여기에는 누락된 값, 이상값 또는 일관성 없는 데이터 유형 검사가 포함될 수 있습니다.
2. 입력 데이터에서 평균, 표준 편차, 누락된 값 수, 이상치 수 등의 통계 지표를 계산합니다.
3. 이러한 지표를 모델 개발 단계에서 설정된 기준 값 또는 예상 범위와 비교합니다. 기준과의 상당한 편차는 데이터 품질 문제나 분포 변화를 나타낼 수 있습니다.
4. 데이터 드리프트 탐지 같은 기법을 사용하여 시간 경과에 따른 입력 데이터 분포의 변화를 식별합니다. 여기에는 평균, 분산 또는 왜도 등 데이터의 통계적 속성을 모니터링하는 것이 포함될 수 있습니다. 콜모고로프-스미르노프 검정 또는 최대 평균 불일치 같은 고급 기법을 사용하여 분포 변화를 탐지할 수도 있습니다.

데이터 품질을 지속적으로 모니터링하면 문제를 빠르게 파악하여 모델 성능에 큰 영향을 주기 전에 문제를 해결할 수 있습니다.

DO NOT COPY
ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



모델 품질 드리프트는 시간이 갈수록 모델 예측이 레이블이 지정된 데이터(**정해진 정답**) 레이블과 달라지는 현상입니다. 이러한 드리프트는 기본 데이터 분포 변경, 사용자 행동 방식 변화, 시장 상황 변화 등의 여러 가지 요인으로 인해 발생할 수 있습니다.

모델 품질 드리프트가 발생하면 예측의 정확도와 신뢰성이 낮아집니다. 모델 출력이 비즈니스 프로세스 또는 고객과의 상호 작용 관련 정보를 제공하는 데 사용된다면 이러한 드리프트로 인해 잘못된 결정을 내리게 될 수도 있습니다. 추천 시스템에서 관련성이 없는 제품 추천을 시작하거나 사기 행위 탐지 모델이 새로운 유형의 사기 행위를 놓치는 경우를 예로 들 수 있습니다.

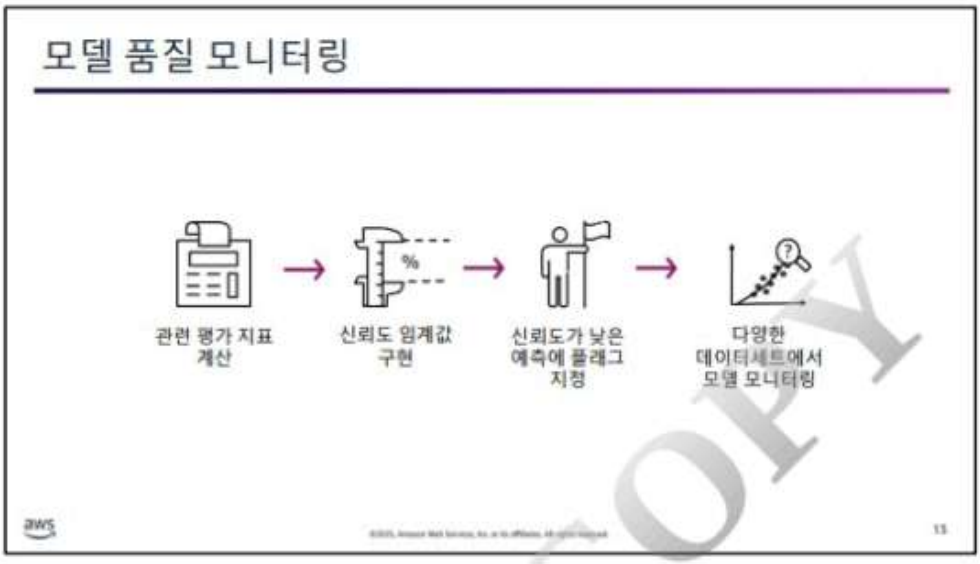
드리프트를 조기에 탐지하고 시정 조치를 취하려면 모델 품질을 지속적으로 모니터링해야 합니다. 구체적으로는 모델의 예측을 레이블이 지정된 데이터 세트(정해진 정답)와 정기적으로 비교해야 합니다. 시간 경과에 따른 성능 지표를 추적하면 모델 품질 저하가 시작되는 시기를 파악하여 모델 재훈련이나 업데이트가 필요한 상황을 결정할 수 있습니다.

안정적인 모니터링 시스템을 구현하면 모델의 효율성을 유지할 수 있으며, 그러면 모델이 프로덕션 환경에서 정확한 예측을 계속 제공하게 됩니다.

DO NOT COPY ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



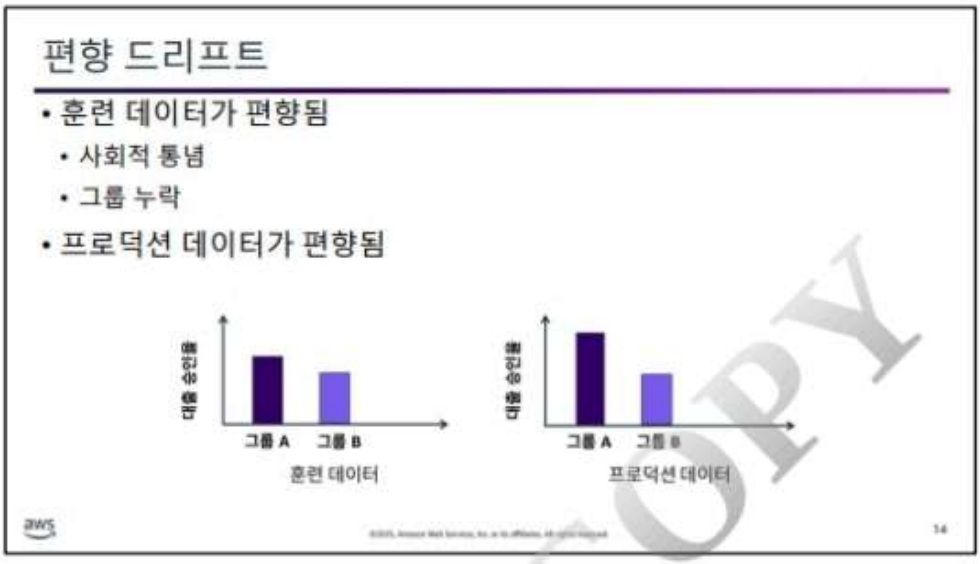
모델 품질을 모니터링하려는 경우 아래 단계를 진행합니다.

1. 홀드아웃 테스트 세트 또는 프로덕션 데이터의 대표 샘플에서 정확도, 정밀도, 재현율, F1 점수, 수신기 작동 특성 곡선(AUC-ROC) 아래 면적 등 관련 평가 지표를 계산합니다. 이렇게 하면 처음 보는 데이터 사용 시의 모델 성능을 평가한 후 훈련 또는 이전 모니터링 주기의 성능과 비교할 수 있습니다.
2. 신뢰도 임계값 또는 불확실성 추정 기법을 구현하여 신뢰성이 낮을 가능성이 있는 예측을 식별합니다. 가령 모델 신뢰도 점수에 임계값을 설정하거나 Monte Carlo Dropout 등의 방법을 사용하여 예측 불확실성을 예상할 수 있습니다.
3. 추가 검토를 위해 또는 애플리케이션에서 특수하게 처리할 수 있도록 신뢰도가 낮거나 불확실성이 높은 예측에 플래그를 지정합니다. 이렇게 하면 중요한 결정에서 신뢰도가 낮은 모델 출력이 사용되는 상황을 방지할 수 있습니다.
4. 여러 하위 모집단 또는 데이터 슬라이스에서 모델 성능을 모니터링합니다. 이렇게 하면 특정 그룹이나 데이터 클러스터에만 영향을 줄 수 있는 성능 저하 또는 편향 문제를 탐지할 수 있습니다.

이러한 모니터링 방법을 구현하면 모델 품질 관련 문제를 사전에 파악하여 해결할 수 있으며, 그러면 ML 모델이 프로덕션 환경에서 계속 효율적으로 작동하게 됩니다.

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



라이브 데이터의 분포가 바뀌면 **편향 드리프트**가 발생할 수 있습니다. 이러한 드리프트를 야기할 수 있는 몇 가지 이유는 다음과 같습니다.

- 훈련 데이터의 편향: 원래 훈련 데이터가 너무 적거나 라이브 데이터를 대표하지 않을 수도 있고 기존의 사회적 통념이 훈련 데이터에 반영되어 있을 수도 있습니다.
- 데이터 포인트 누락: 중요한 데이터 요소나 그룹이 중요한 정보임을 인지하지 못하여 훈련 데이터에서 제외되었을 수도 있습니다.
- 실제 데이터 변경: 모델이 상호 작용하는 라이브 데이터에서 훈련 데이터에는 없던 새로운 편향이 발생했을 수 있습니다.

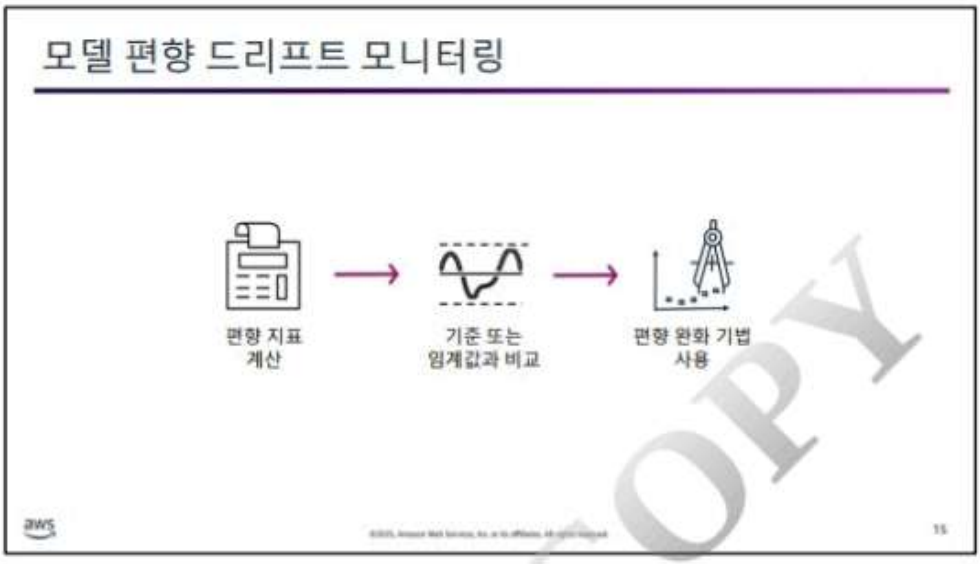
편향 드리프트가 발생하면 특정 그룹을 대상으로 불공평하거나 정확하지 않은 예측이 생성될 수 있으며, 그러면 기존의 불평등 상황이 더욱 심화되거나 잘못된 결정을 하게 될 수도 있습니다. 훈련 데이터에 정확하게 반영되지 않았던 특정 인구 통계 그룹과 관련된 편향이 발생하는 대출 승인 모델을 예로 들 수 있습니다.

프로덕션 환경에서 편향 드리프트를 모니터링하면 모델이 모든 그룹을 대상으로 공정하며 정확한 예측을 계속 제공할 수 있습니다. 이 모니터링에서는 여러 하위 모집단에서 모델 예측을 정기적으로 분석한 후 설정되어 있는 공정성 지표와 비교하는 작업을 수행합니다.

편향은 분산과는 다른 개념입니다. 편향이 발생하면 모델 예측이 과도하게 일반화되는 경우가 많으며, 분산이 발생하면 모델이 예측을 과소 일반화하게 됩니다. 편향 드리프트를 해결할 때 편향의 영향을 줄이기 위해 모델 분산 수준을 높여야 하는 경우도 있습니다.

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



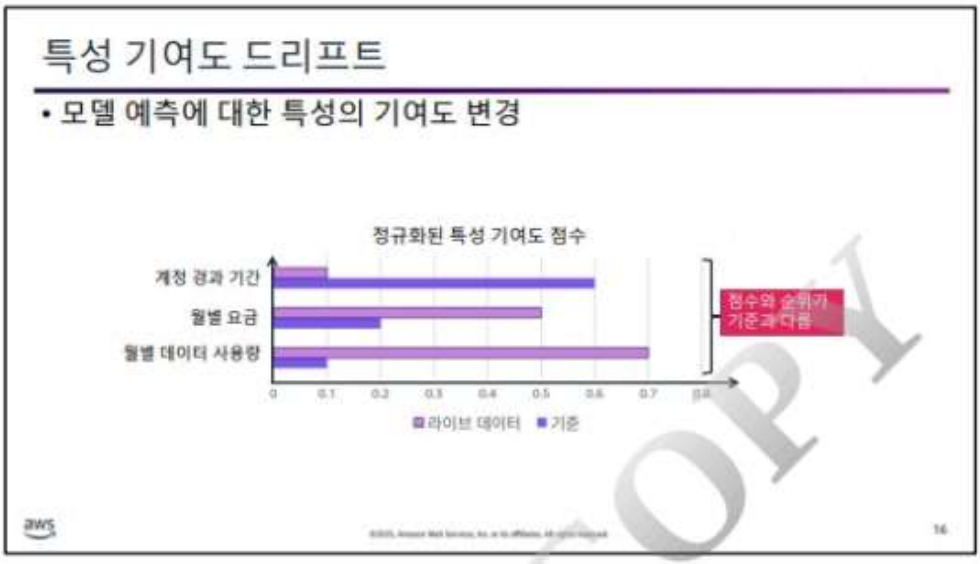
모델 편향 드리프트를 모니터링하려는 경우 아래 단계를 진행합니다.

1. 여러 민감한 그룹 또는 하위 모집단을 대상으로 편향 지표를 계산합니다. 일반적으로는 영향력 불균형, 기회 균등 차이, 평균 승산 차이 등의 지표를 계산하게 됩니다. 이러한 지표는 인종, 성별 또는 연령 같은 보호 속성에 대한 모델 예측의 편향이나 불공정성의 정도를 정량화합니다.
2. 모델 개발 중에 설정된 기준 값 또는 허용 임계값과 편향 지표를 비교합니다. 기준과의 상당한 편차는 모델 예측에 편향이 유입되거나 증폭되었음을 나타낼 수 있습니다.
3. 허용되지 않는 수준의 편향이 탐지되면 편향 완화 기법을 구현합니다. 가령 다음과 같은 방법을 사용할 수 있습니다.
   - 적대적 편향 제거: 민감한 속성을 고려하지 않고 예측을 수행하도록 모델을 훈련시킵니다.
   - 균등하게 보정된 승산 적용: 모든 그룹에서 거짓 긍정 비율과 거짓 부정 비율이 같아지도록 모델 예측을 조정합니다.

바이어스 드리프트를 지속적으로 모니터링하고 필요시 완화 전략을 구현하면 프로덕션 환경에서 기계 학습 모델이 공정하며 공평한 예측을 계속 제공하게 됩니다. 대출, 채용, 법무 등 모델의 예측이 사회적으로 큰 영향을 주는 영역에서는 이러한 모니터링이 특히 중요합니다.

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



특성 기여도는 각 특성이 모델의 예측 또는 출력에 미치는 영향 또는 기여도를 이해하고 정량화하는 것을 말합니다. 특성 기여도 드리프트는 **모델 예측에 대한 개별 특성의 기여도나 중요도가 시간이 경과하면서 크게 바뀌면** 발생합니다. 이러한 드리프트 유형은 기본 데이터 분포의 변화 또는 특성과 대상 변수 간의 관계 변화를 나타낼 수 있습니다.

특성 기여도를 통해 모델의 의사결정 프로세스에서 가장 관련성이 높은 특성과 해당 특성의 상대적 중요도를 파악할 수 있습니다. 프로덕션 환경에서는 라이브 데이터 분포가 변화할 수 있으며, 이로 인해 특성 기여도 값에서 해당 드리프트가 발생할 수 있습니다.

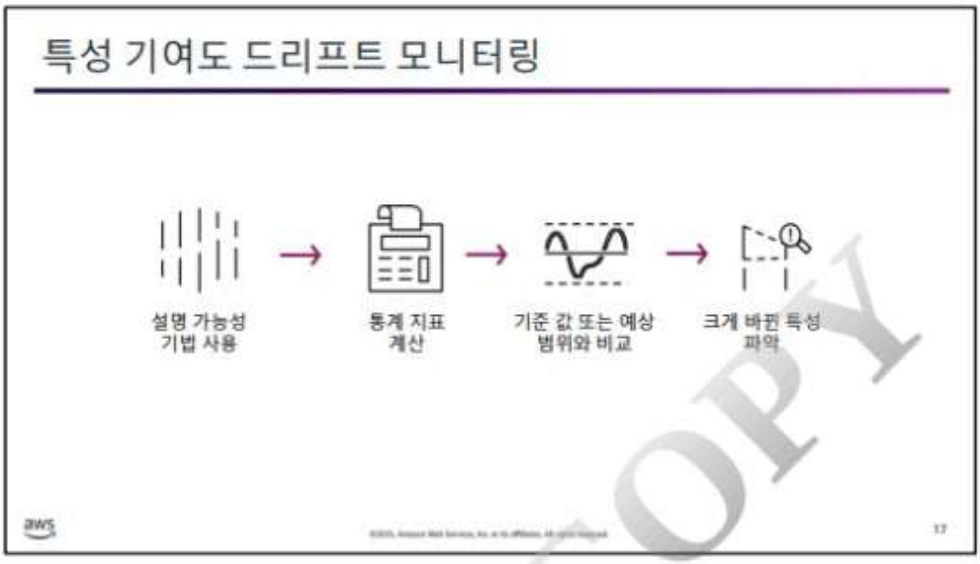
예를 들어 고객 이탈 예측 모델에서는 다음과 같이 특성 기여도 드리프트가 발생할 수 있습니다.
- 처음에는 계정 사용 기간 특성의 중요도가 가장 높을 수 있습니다.
- 시간이 경과하면서 고객 행동 방식이나 시장 상황이 바뀌어 월별 데이터 사용량의 중요도가 더 높아질 수 있습니다.

특성 기여도 드리프트를 모니터링해야 하는 이유는 다음과 같습니다.
- 모델을 업데이트해야 할 수도 있는 문제 영역의 변화를 파악할 수 있습니다.
- 발생 가능한 데이터 품질 문제나 데이터 수집 프로세스의 변화를 파악할 수 있습니다.
- 특성 중요도를 실무 과정에서 확인된 수준에 일치하는 상태로 유지함으로써 높은 모델 해석 가능성을 유지할 수 있습니다.

데이터 과학자는 특성 속성 변화를 추적하여 데이터의 변화하는 데이터 패턴에 대한 인사이트를 파악할 수 있으며, 관련 정보를 토대로 모델 유지 관리 및 재훈련 전략을 결정할 수 있습니다.

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



특성 기여도 드리프트를 모니터링할 때는 다음과 같은 몇 가지 주요 단계를 수행하여 모델의 예측 과정에서 각 특성의 중요도 변화를 추적하게 됩니다.

1. SHAP(SHapley Additive exPlanations) 같은 설명 가능성 기법을 사용하여 개별 예측의 특성 기여도 또는 중요도 점수를 계산합니다. 이러한 기법은 다양한 특성이 모델의 출력에 어떻게 기여하는지에 대한 인사이트를 제공합니다.
2. 특성 기여도에 대한 통계 지표를 계산합니다(예: 평균 또는 표준 편차). 그러면 각 특성 기여도의 전반적인 중요도와 가변성을 정량화할 수 있습니다.
3. 계산한 지표를 모델 개발 중에 설정된 기준 값 또는 예상 범위와 비교합니다. 기준과의 상당한 편차는 시간이 지남에 따라 특정 특성의 중요도 또는 기여도의 변화를 나타낼 수 있습니다.
4. 시간이 경과하면서 기여도가 크게 바뀐 특성을 파악합니다. 이러한 특성이 있다면 기본 데이터 분포가 바뀌었거나 모델의 작동 방식이 변경된 것일 수 있습니다.

특성 기여도 드리프트를 정기적으로 모니터링하는 경우의 이점은 다음과 같습니다.

- 모델 의사결정 프로세스의 미묘한 변화를 탐지할 수 있습니다.
- 모델이 업무상 적절한 방식으로 작동하도록 조율합니다.
- 재훈련 또는 모델 업데이트를 수행해야 할 수 있는 시기를 파악합니다.

이 프로세스를 진행하면 프로덕션 환경에서 기계 학습 모델의 해석 가능성과 신뢰성을 유지할 수 있으며 사업 목표에 부합하는 방식으로 모델을 사용할 수 있습니다.

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

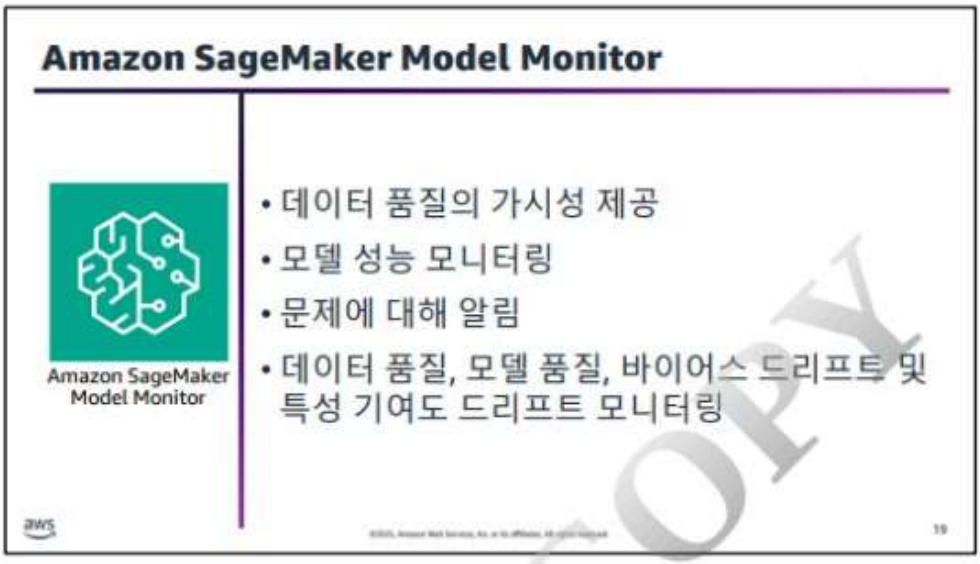
Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

이 주제에서는 SageMaker Model Monitor의 작동 방식, 그리고 Model Monitor의 기능을 사용하여 기계 학습 모델의 신뢰성과 효율성을 지속적으로 유지할 수 있는 방법을 알아봅니다.

DO NOT COPY
ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



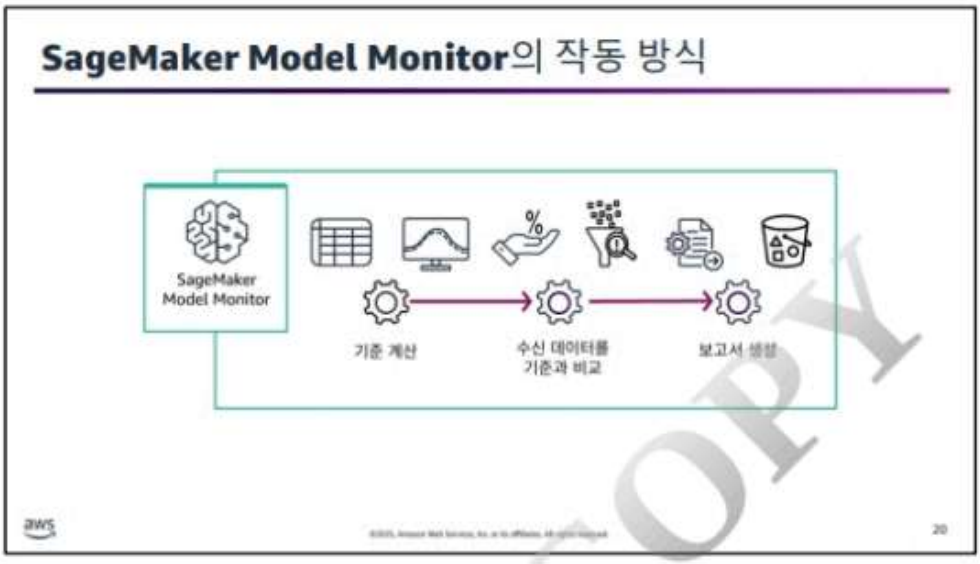
Amazon SageMaker Model Monitor는 프로덕션 환경에서 기계 학습 모델을 모니터링할 수 있는 포괄적인 솔루션입니다. SageMaker Model Monitor에서는 다음과 같은 주요 기능이 제공됩니다.

- 데이터 품질 모니터링: SageMaker Model Monitor는 수신 데이터의 통계적 속성을 지속적으로 분석하여 훈련 중에 설정된 기준과 비교합니다. 그러면 모델 성능에 영향을 줄 수 있는 데이터 드리프트 또는 품질 문제를 탐지할 수 있습니다.
- 모델 성능 모니터링: SageMaker Model Monitor는 시간 경과에 따른 모델의 주요 성능 지표를 추적합니다. 따라서 모델의 정확도 또는 기타 관련 지표가 떨어지기 시작하는 시기를 파악할 수 있습니다.
- 알림 시스템: 데이터 품질이나 모델 성능의 큰 변화가 탐지되면 알림을 전송하도록 SageMaker Model Monitor를 구성할 수 있습니다. 이러한 알림을 확인하면 해당 상황을 즉시 조사하여 문제를 해결할 수 있습니다.
- 포괄적인 드리프트 탐지: SageMaker Model Monitor는 다음과 같은 여러 드리프트 유형을 모니터링합니다.
  - 데이터 품질 드리프트
  - 모델 품질 드리프트
  - 바이어스 드리프트
  - 특성 기여도 드리프트

모델 작동 방식과 관련된 이러한 여러 요소를 확인할 수 있는 SageMaker Model Monitor를 활용하면 필요한 조치를 즉시 취하여 배포한 모델의 품질을 일정하게 유지 및 개선할 수 있습니다.

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



SageMaker Model Monitor의 작동 과정에서는 다음과 같은 일련의 주요 단계가 진행됩니다.

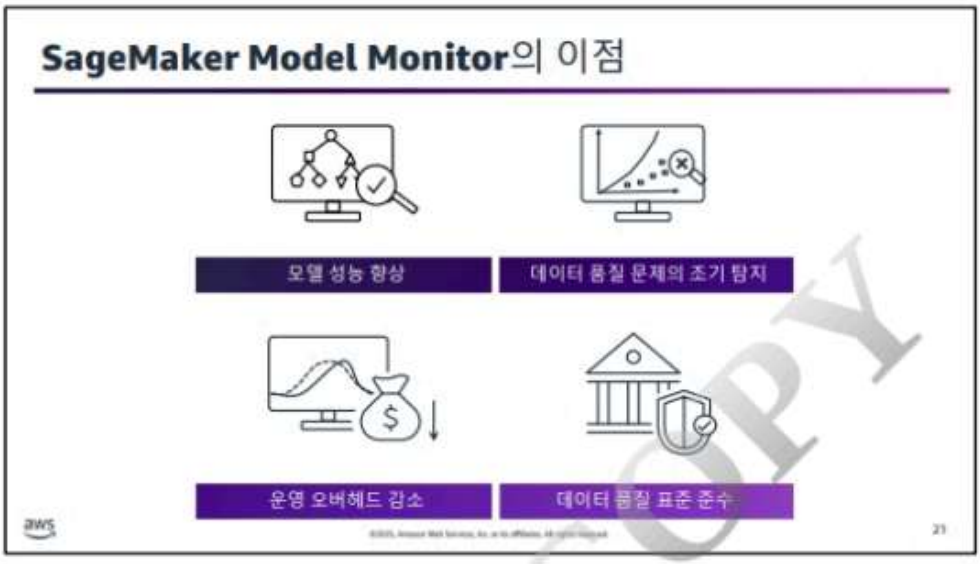
1. 기준 생성: 훈련 데이터세트를 사용하여 데이터 관련 통계와 제약 조건을 캡처하는 방식으로 기준을 설정합니다. 이 기준은 모니터링용 참조 수치로 사용됩니다.
2. 데이터 캡처: Model Monitor 서비스가 SageMaker AI 엔드포인트에서 데이터 캡처 특성을 구성하고 추론 요청과 예측을 수집합니다.
3. 비교 프로세스: 수신 데이터를 기준과 비교하여 편차나 품질 문제를 탐지합니다.
4. 보고서 생성: 문제가 탐지되면 SageMaker Model Monitor가 문제를 자세히 설명하는 세부 보고서를 생성합니다.

이러한 보고서는 액세스 및 분석용으로 Amazon Simple Storage Service(Amazon S3) 버킷에 저장됩니다. SageMaker Model Monitor에서는 이 프로세스가 자동으로 진행되므로 ML 모델의 품질을 지속적으로 보증할 수 있습니다.

DO NOT COPY ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



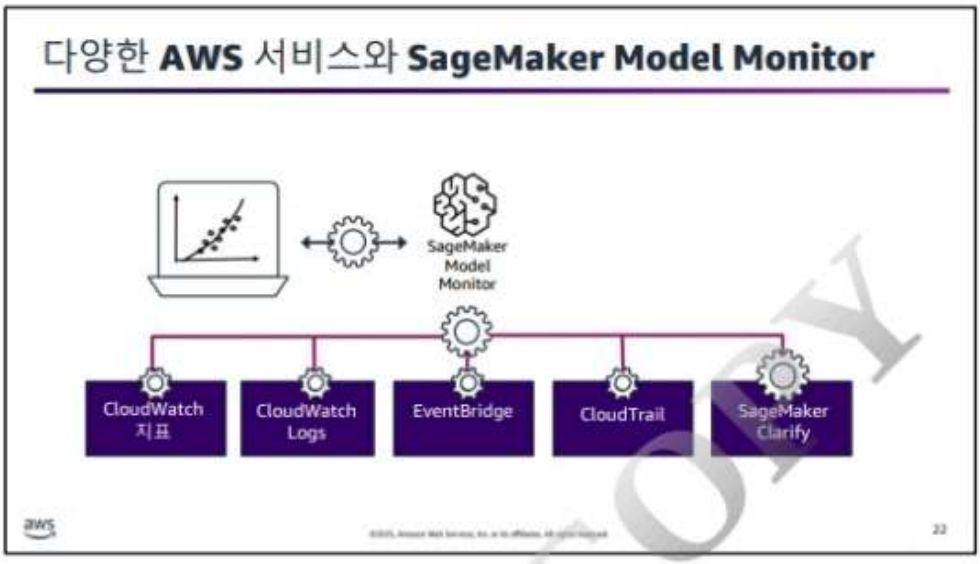
SageMaker Model Monitor는 Amazon SageMaker AI에서 호스팅되는 기계 학습 모델을 지속적으로 평가할 수 있습니다. SageMaker Model Monitor는 모델 성능 편차를 탐지하고 사용자에게 알리도록 설계되었습니다. SageMaker Model Monitor는 ML 작업에서 다음과 같은 여러 가지 중요한 이점을 제공합니다.

- 모델 성능 및 정확도 개선: SageMaker Model Monitor가 데이터 품질을 지속적으로 모니터링하므로 모델이 작동 과정에서 신뢰할 수 있는 데이터를 사용하도록 할 수 있으며, 그러면 모델 성능도 개선됩니다.
- 데이터 품질 문제 조기 탐지: SageMaker Model Monitor가 수신 데이터의 문제를 빠르게 식별하므로 이러한 문제가 모델 성능에 중대한 영향을 주기 전에 즉시 해결할 수 있습니다.
- 운영 오버헤드 감소: 모니터링 프로세스가 자동화되므로 데이터 과학자와 엔지니어가 더 중요한 태스크를 중점적으로 수행할 수 있는 시간이 늘어납니다. 따라서 전반적인 생산성도 높아집니다.
- 데이터 품질 표준 준수: SageMaker Model Monitor가 데이터 품질 및 모델 성능을 지속적으로 검사하므로 업계 및 규정 표준 준수 상태를 유지할 수 있습니다.

이러한 혜택은 집계적으로 더욱 안정적이고 효율적이며 신뢰할 수 있는 ML 운영에 기여하여, 조직은 자신 있게 ML 이니셔티브를 확장할 수 있습니다.

DO NOT COPY
ssleeee1@gmail.com

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



SageMaker Model Monitor는 여러 AWS 서비스와 원활하게 통합할 수 있으며, 통합 시에는 다음과 같은 포괄적인 모니터링 및 관리 기능이 제공됩니다.

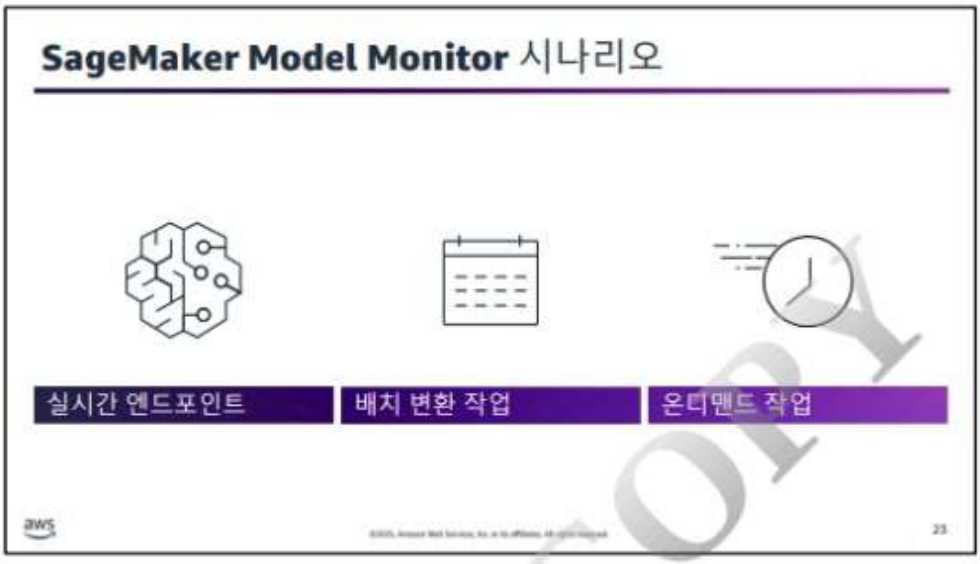
- Amazon CloudWatch 지표: SageMaker Model Monitor는 CloudWatch로 지표를 내보내므로 실시간으로 지표를 모니터링하고 알림을 확인할 수 있습니다.
- Amazon CloudWatch Logs: 모니터링 작업의 세부 로그가 문제 해결 및 분석용으로 CloudWatch Logs에 저장됩니다.
- Amazon EventBridge: 이 서비스와 Model Monitor를 통합하면 모니터링 이벤트를 토대로 자동화된 작업을 수행할 수 있습니다.
- AWS CloudTrail: SageMaker Model Monitor에서 수행하는 모든 API 호출이 기록되므로 규정 준수 및 보안용 감사 추적 정보가 제공됩니다.
- Amazon SageMaker Clarify: 이 서비스와 Model Monitor를 통합하면 모델에서 발생 가능한 편향 및 특성 기여도 변화를 확인할 수 있습니다.

이러한 통합은 ML 모델 모니터링 및 관리를 위한 안정적인 시스템을 구축됩니다. 이러한 시스템에서는 자동화된 워크플로, 세부 감사, 포괄적인 성능 추적을 진행할 수 있습니다.

DO NOT COPY ssleeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



SageMaker Model Monitor에서는 다음과 같은 여러 ML 워크플로를 진행할 수 있는 다양한 모니터링 시나리오가 지원됩니다.

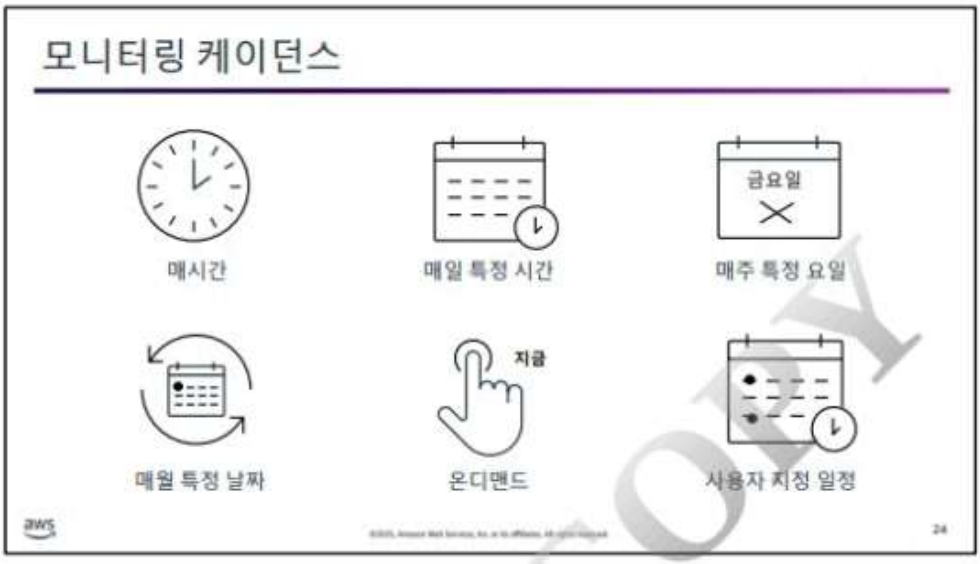
- 실시간 엔드포인트 모니터링: 실시간 엔드포인트에 배포된 모델의 지속적 모니터링을 설정할 수 있습니다. SageMaker Model Monitor는 모델 입력 및 출력에서 데이터를 캡처하고 기준과 비교한 편차를 모니터링합니다.
- 배치 변환 작업 모니터링: 이 워크플로를 정기적으로 실행하거나 일정에 따라 실행하는 배치 변환 작업에 사용되는 모델을 모니터링할 수 있습니다. SageMaker Model Monitor는 이러한 배치 작업에서 입력 데이터와 출력 예측을 캡처하고 드리프트를 모니터링합니다. 데이터 가용성과 원하는 빈도에 따라 일정을 조정할 수 있습니다.
- 온디맨드 모니터링 작업: SageMaker Model Monitor에서는 온디맨드 모니터링 작업도 지원됩니다. 이러한 작업을 수동으로 시작하여 언제든지 모델 성능을 평가할 수 있습니다.

이처럼 SageMaker Model Monitor에서는 여러 워크플로를 유동적으로 진행할 수 있으므로 프로덕션 환경에서 모델이 사용되는 방식에 관계없이 다양한 ML 배포 패턴에 맞게 SageMaker Model Monitor를 활용할 수 있습니다.

DO NOT COPY
ssleeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



구체적인 사용 사례와 데이터 빈도에 맞게 SageMaker Model Monitor에서 모니터링 케이던스를 사용자 지정할 수도 있습니다.

- 실시간 엔드포인트의 경우 데이터 품질 문제를 빠르게 탐지할 수 있도록 지속적인 모니터링을 수행하는 것이 좋습니다.
- 배치 변환 작업은 작업 빈도와 데이터 사용 가능 여부를 토대로 모니터링할 수 있습니다.
- cron 표현식으로 사용자 지정 일정을 생성하면 타이밍 옵션을 유동적으로 조정할 수 있습니다.

다음과 같은 케이던스로 실행되도록 모니터링 일정을 설정할 수 있습니다.

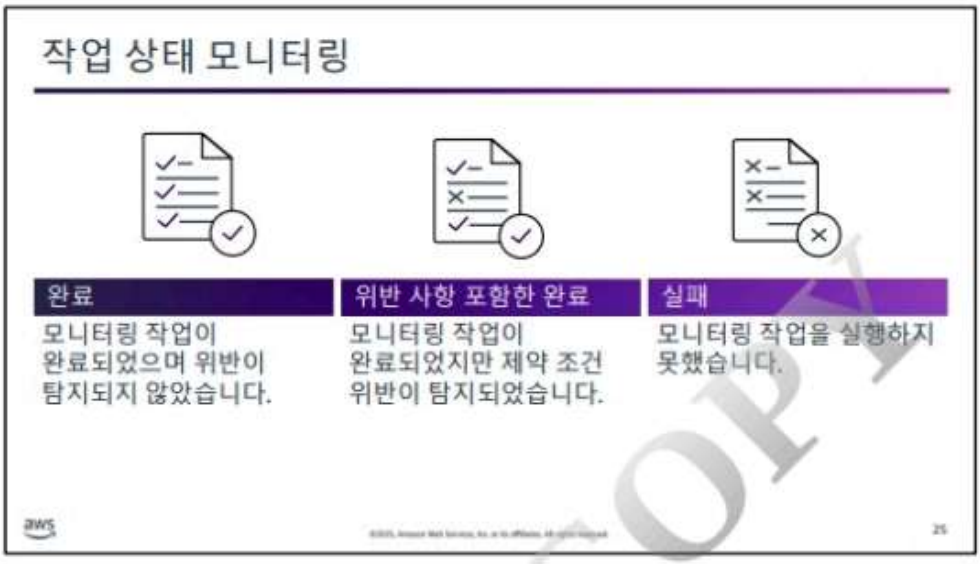
- 매시간
- 매일 특정 시간
- 매주 특정 요일
- 매월 특정 날짜
- 필요시 온디맨드

SageMaker AI의 CronExpressionGenerator 유틸리티를 사용하면 이러한 사용자 지정 일정을 생성할 수 있습니다. 적절한 모니터링 주기를 선택하면 문제를 제때 탐지하고 리소스 사용량을 최적화할 수 있습니다.

DO NOT COPY
ssleeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



모니터링 작업이 시작된 후 해당 작업을 나열하거나 작업 상태를 볼 수 있습니다. 모니터링 작업은 3가지 상태 중 하나로 완료됩니다.

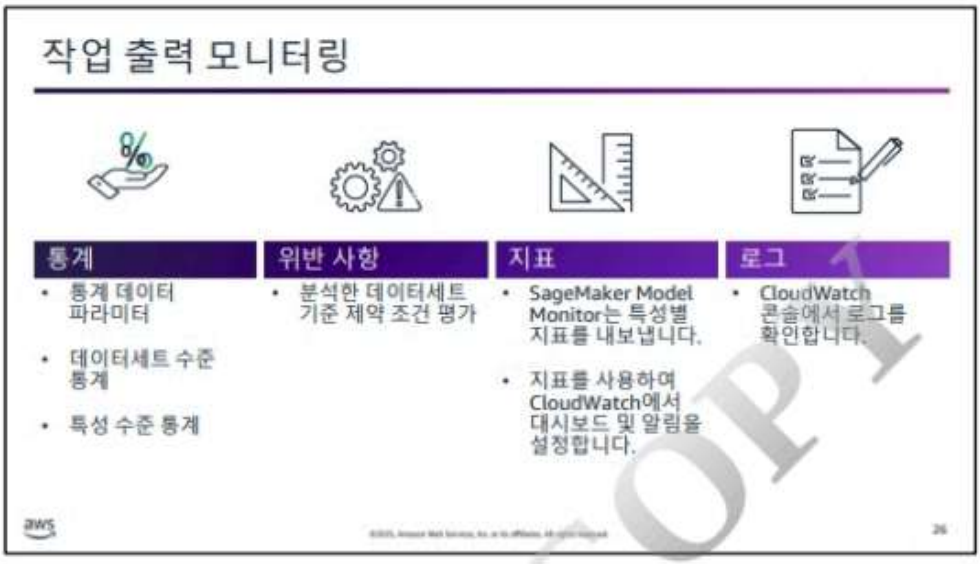
- 완료: 작업이 정상적으로 완료되었으며 위반이 탐지되지 않았습니다. 모델과 데이터 품질이 예상 파라미터 범위 내에 있음을 나타내는 상태입니다.
- 위반 사항 포함한 완료: 작업이 완료되었지만 제약 조건 위반이 탐지되었습니다. 이 상태가 표시되면 추가 조사를 진행하여 확인된 문제를 파악하고 해결해야 합니다.
- 실패: 모니터링 작업에서 오류가 발생하여 작업을 완료할 수 없습니다. 이 상태는 잘못된 권한, 인프라 문제, 잘못된 구성 등의 여러 가지 원인으로 인해 나타날 수 있습니다.

작업 상태를 정기적으로 확인하면 데이터 품질 또는 모델 성능 문제를 제때 탐지하여 해결함으로써 ML 솔루션의 신뢰성을 유지할 수 있습니다.

DO NOT COPY ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



SageMaker Model Monitor는 모델과 수신 데이터를 분석합니다. SageMaker Model Monitor는 모델 및 데이터 모니터링에 사용되는 기준 통계와 제약 조건을 정의합니다. 모니터링 작업의 출력에는 다음 4가지 유형이 포함됩니다.

- 통계: SageMaker Model Monitor 출력에는 다음과 같은 다양한 유형의 통계가 포함될 수 있습니다.
  - 기준 데이터와 캡처되는 데이터에 대해 계산할 통계 파라미터
  - item_count(행 또는 인스턴스 수) 같은 데이터세트 수준 통계
  - 데이터세트의 각 특성 또는 열에 대한 특성 수준 통계
  - null 개수, 최댓값, 최솟값, 평균 등의 기타 통계
- 위반: 위반 파일에는 분석된 데이터세트를 기준으로 제약 조건(constraints.json 파일에 지정됨)을 평가한 결과가 나열됩니다. constraints_violations.json으로 보고서가 출력됩니다.
- 지표: SageMaker Model Monitor는 대시보드와 알림을 설정하는 데 사용할 수 있는 특성별 지표를 CloudWatch로 내보냅니다. CloudWatch의 요약 지표는 Amazon SageMaker Studio에서도 볼 수 있으며, 모든 통계, 모니터링 결과, 수집된 데이터를 노트북에서 보고 추가 분석할 수 있습니다.
- 로그: Model Monitor는 시작 시간, 종료 시간, 상태 등 모니터링 작업에 대한 정보가 포함된 로그를 생성합니다. 이 로그는 CloudWatch 콘솔에서 볼 수 있습니다.

이러한 출력에서 모델 상태를 전체적으로 확인할 수 있으므로 데이터 과학자와 엔지니어가 정보를 토대로 모델 유지 관리 및 재훈련과 관련된 결정을 내릴 수 있습니다. 이러한 출력을 정기적으로 분석하면 프로덕션 환경에서 ML 모델의 높은 품질을 유지할 수 있습니다.

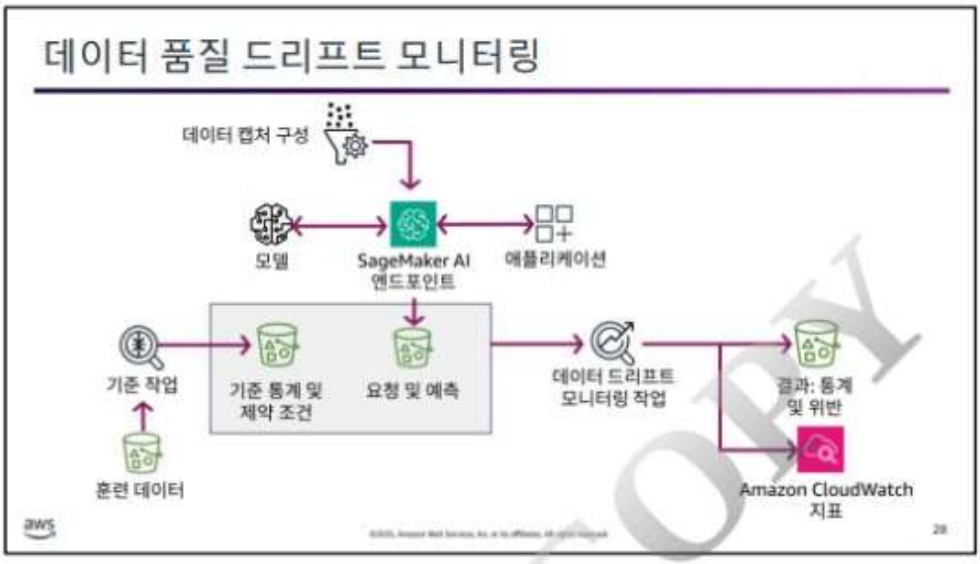
© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

이 주제에서는 프로덕션 환경에서 기계 학습 모델과 관련하여 모니터링해야 하는 2가지 요소인 **데이터 품질**과 **모델 품질**에 대해 중점적으로 살펴봅니다. 데이터 품질을 모니터링하면 예측에 사용되는 입력 데이터를 훈련 데이터와 일치하는 상태로 유지할 수 있습니다. 따라서 모델 성능에 영향을 줄 수 있는 데이터 드리프트를 탐지할 수 있습니다. 또한 모델 품질을 모니터링하면 모델 예측이 Ground Truth 레이블과 일치하는 정도를 평가하여 시간 경과에 따른 성능 저하도 탐지할 수 있습니다.

DO NOT COPY
ssleeeee1@gmail.com

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



데이터 품질 모니터링에서는 대략적으로 다음과 같은 단계를 수행합니다.

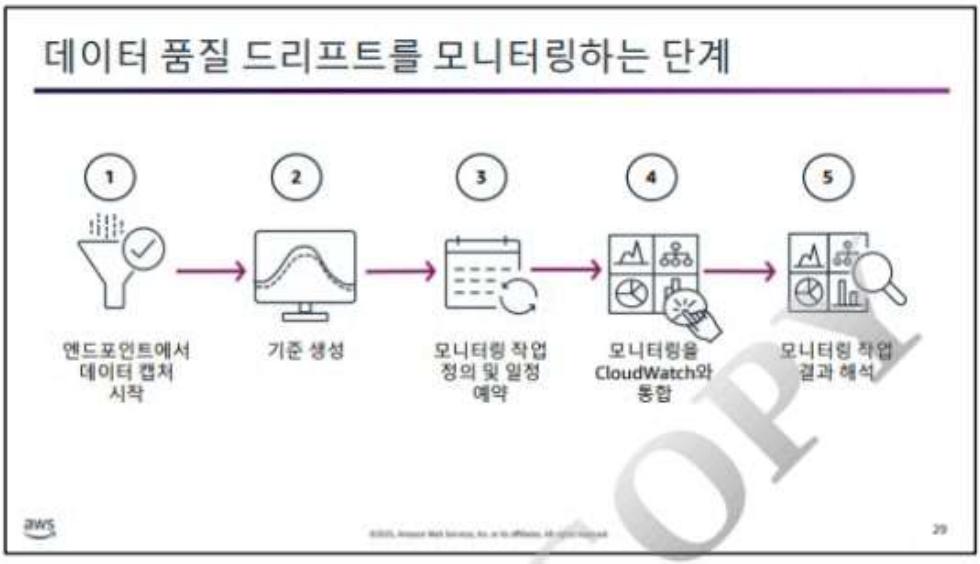
- 데이터 캡처 구성: SageMaker AI 엔드포인트에서 데이터 캡처 구성을 설정합니다.
- 기준 생성: 기준 제약 조건 세트를 제안하고 설명 통계를 생성하도록 SageMaker AI에 요청할 수 있습니다.
- 기준 통계 및 제약 조건: 기준이 실행되고 나면 반환되는 출력에는 훈련 데이터의 각 특성 관련 통계가 포함됩니다.
- 데이터 드리프트 모니터링 작업: 각 모니터링 작업은 배포된 엔드포인트에서 새로운 추론 데이터를 캡처합니다. 그리고 새 데이터를 기준과 비교합니다. 모든 위반 또는 데이터 품질 문제가 보고됩니다.
- Amazon CloudWatch 지표: CloudWatch 지표용으로 기본 제공되는 SageMaker Model Monitor 컨테이너를 사용할 수 있습니다.

데이터 품질을 지속적으로 모니터링하면 데이터 분포의 변화 또는 예상치 못한 값 추가 등의 문제를 **사전에 파악**하고 제때 해결하여 모델 성능을 유지할 수 있습니다.

DO NOT COPY
ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



SageMaker Model Monitor에서 데이터 품질 드리프트를 모니터링할 때는 다음의 5개 주요 단계를 수행합니다.

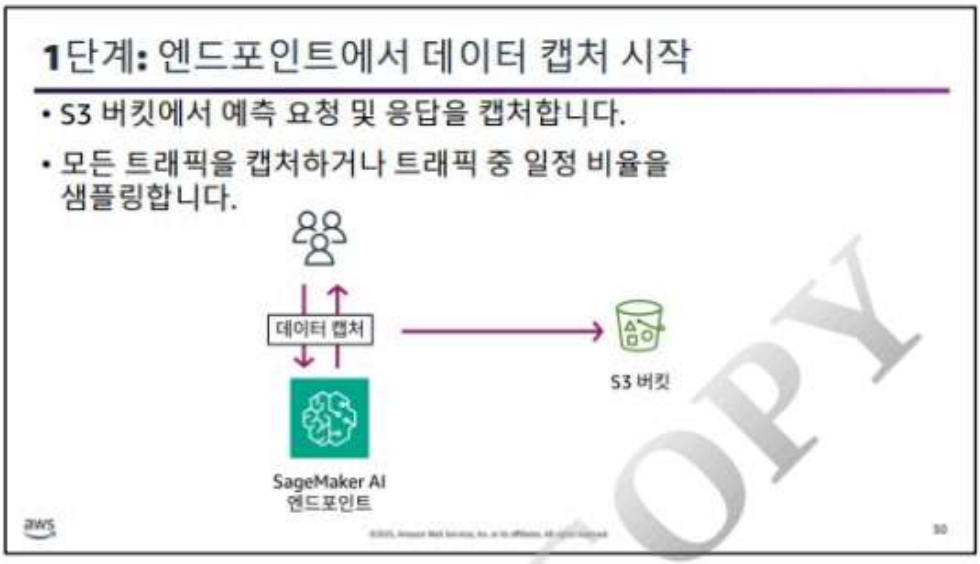
1. 엔드포인트에서 데이터 캡처 시작: 추론 요청과 응답을 캡처한 다음, S3 버킷에 저장하도록 SageMaker AI 엔드포인트를 구성합니다.
2. 기준 생성: 훈련 데이터를 사용하여 비교 기준점으로 사용되는 기준 통계와 제약 조건을 생성합니다.
3. 모니터링 작업 정의 및 예약: 캡처한 데이터를 분석하고 기준 데이터와 비교할 정기적인 모니터링 작업을 설정합니다.
4. CloudWatch와 통합: 데이터 품질이 지정된 임계값을 벗어나면 알리도록 CloudWatch 지표와 경보를 구성합니다.
5. 결과 해석: 모니터링 작업의 출력을 분석하여 발생 가능한 문제를 파악하고 적절한 조치를 취합니다.

이러한 단계를 수행하면 시간이 지나도 ML 모델의 품질과 신뢰성을 유지할 수 있는 안정적인 모니터링 시스템을 구축할 수 있습니다.

DO NOT COPY ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

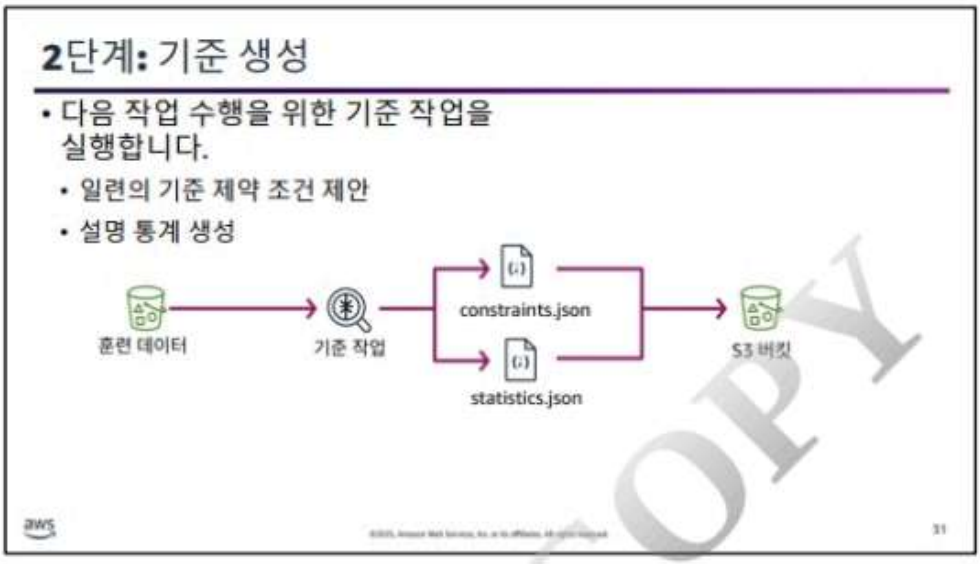


데이터를 모니터링하려면 모델을 호스팅하는 엔드포인트에서 추론 요청을 캡처해야 합니다. 엔드포인트를 구성할 때 데이터 캡처 기능을 활성화합니다. 데이터 캡처가 활성화되면 수신 예측 요청과 응답이 캡처되어 S3 버킷에 유지됩니다. 요청 중 일정 비율을 샘플링하려는 경우 캡처하려는 트래픽의 비율인 샘플링 백분율을 지정할 수 있습니다.

DO NOT COPY
ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



모니터링용 기준을 생성하려는 경우 기준 제약 조건 세트를 제안하고 설명 통계를 생성하도록 SageMaker AI에 요청할 수 있습니다. 이렇게 하려면 Amazon SageMaker Python SDK에서 `ModelQualityMonitor` 객체의 `suggest_baseline` 메서드를 사용하여 기준 처리 작업을 시작합니다.

기준 작업은 훈련 데이터세트를 분석하고 키 파일 2개(constraints.json 및 statistics.json)를 생성합니다. 기준이 실행되고 나면 반환되는 출력에는 훈련 데이터의 각 특성 관련 통계가 포함됩니다. 특성에 따라 기준 작업에서 관련 통계가 생성됩니다.

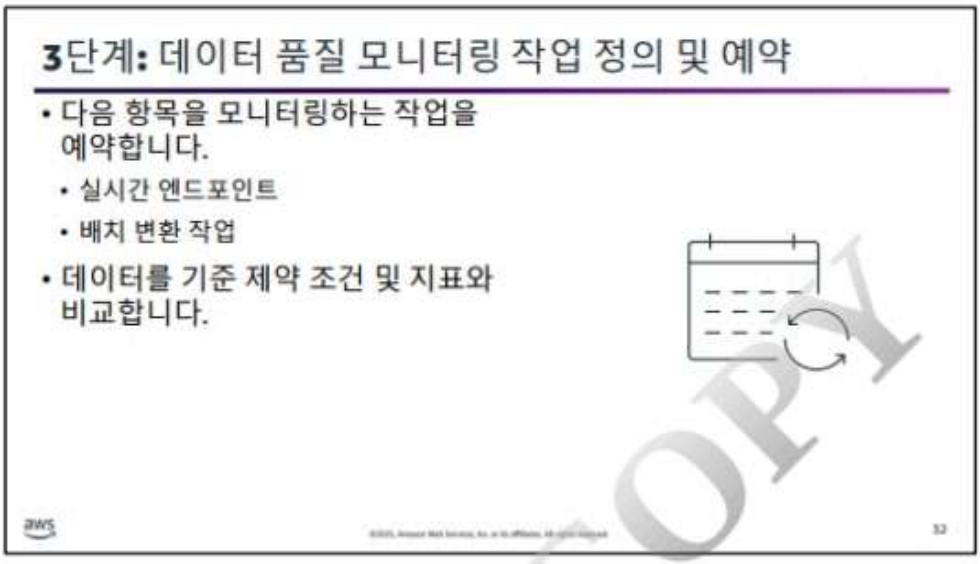
표시되는 통계는 다음과 같습니다.
- 수치 데이터: 최소값 또는 최대값
- 문자열 또는 범주형 데이터: 누락된 값 또는 고유 값

이러한 제안된 제약 조건을 모니터링에 사용하기 전에 검토하고 수정할 수 있습니다. 예를 들어 도메인 지식이나 사용 사례에 따라 제약 조건을 강화하거나 완화할 수 있습니다. 이러한 조정은 모니터링 중에 식별되는 위반에 영향을 미칠 수 있습니다. 이 기준을 후속 모니터링 작업의 비교 기준으로 사용할 수 있습니다.

DO NOT COPY ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



모니터링 일정을 생성하려면 일정 이름과 모니터링 파라미터를 정의합니다. 실시간 드래픽과 비교할 기준 제약 조건과 통계를 사용하여 엔드포인트의 모니터링 일정을 생성합니다. 모니터링 일정을 생성하면 SageMaker AI가 지정된 빈도로 처리 작업을 시작하여 특정 기간 동안 수집된 데이터를 분석할 수 있습니다.

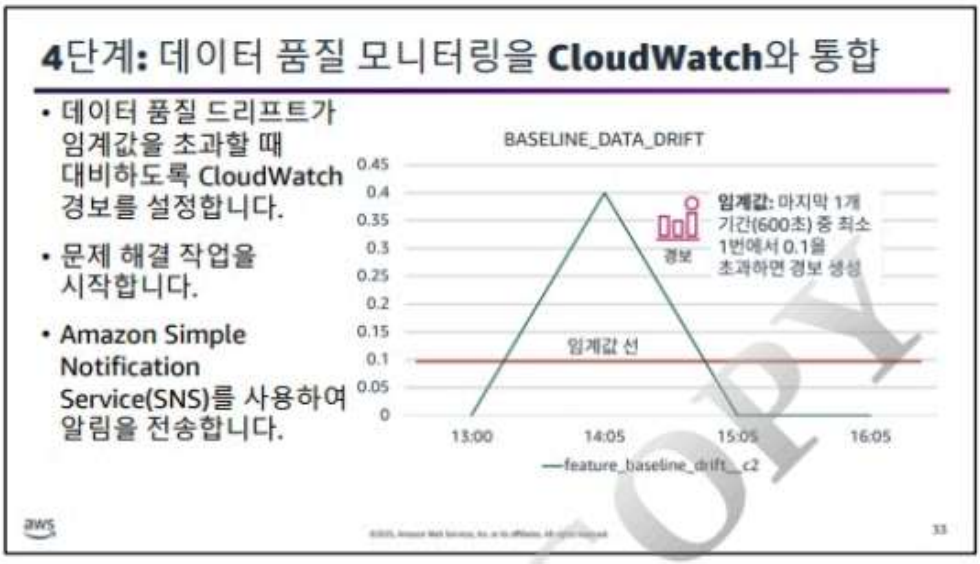
각 모니터링 작업은 배포된 엔드포인트에서 새로운 추론 데이터를 캡처합니다. 그리고 새 데이터를 기준과 비교합니다. 모든 위반 또는 데이터 품질 문제가 보고됩니다.

정기적인 모니터링 일정을 설정하면 SageMaker AI가 **지정된 간격으로 처리 작업을 자동 시작**하여 각 모니터링 기간 동안 수집된 데이터를 분석할 수 있습니다. 이처럼 데이터 상태를 지속적으로 확인하면 프로덕션 환경에서 ML 모델의 품질과 신뢰성을 유지할 수 있습니다.

DO NOT COPY ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



데이터 품질 모니터링을 Amazon CloudWatch와 통합하여 데이터 품질 관련 지표를 확인하려는 경우 기본 제공 SageMaker Model Monitor 컨테이너 이미지를 사용할 수 있습니다. 데이터 품질이 지정된 임계값을 벗어나면 경고하도록 CloudWatch 경보를 구성할 수 있습니다. CloudWatch 경보를 사용하여 모델 재훈련 또는 훈련 데이터 업데이트 같은 문제 해결 작업을 시작할 수도 있습니다.

CloudWatch는 **Amazon Simple Notification Service(Amazon SNS)**를 사용하여 알림을 보냅니다. 알림을 수신하려면 **Amazon SNS 주제**를 생성하고 구독합니다. CloudWatch 경보를 생성할 때는 경보 상태 변경 시 알림이 전송되도록 이 Amazon SNS 주제를 추가할 수 있습니다. 알림은 푸시 알림, 문자 메시지 또는 이메일로 보낼 수 있습니다.

DO NOT COPY ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

5단계: 결과 해석 및 분석

constraint_violations.json으로 생성된 출력

- data_type_check
- completeness_check
- baseline_drift_check
- missing_column_check
- extra_column_check
- categorical_values_check

34

데이터 품질 모니터링 프로세스의 마지막 단계는 결과를 해석하고 분석하는 것입니다. 각 모니터링 작업에서는 위반 파일(대개 이름이 `constraint_violations.json`인 파일)이 생성됩니다. 해당 파일에는 현재 데이터세트를 기준으로 제약 조건을 평가한 결과가 나열됩니다.

SageMaker Model Monitor의 사전 구축 컨테이너에서는 다음과 같은 여러 유형의 위반 검사가 제공됩니다.

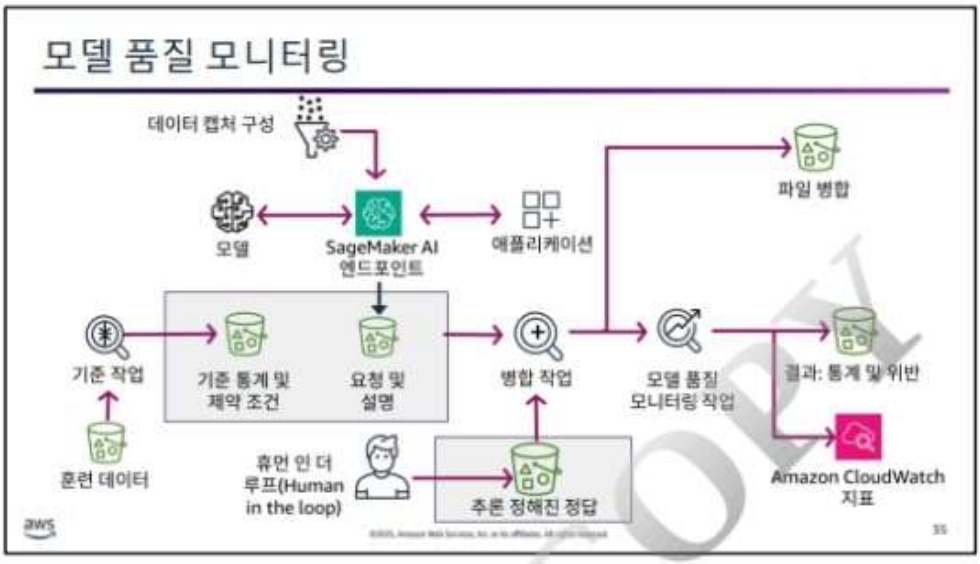
- data_type_check: 데이터 유형이 기준과 일치하는지 검증
- completeness_check: 누락된 값 유무 확인
- baseline_drift_check: 기준과 크게 다른 결과 탐지
- missing_column_check: 현재 데이터에서 누락된 열 식별
- extra_column_check: 기준에 없는 새 열에 플래그 지정
- categorical_values_check: 범주형 값이 기준의 값과 일치하는지 확인

이러한 결과를 일관된 방식으로 분석하면 입력 데이터의 품질을 유지함으로써 모델이 프로덕션 환경에서도 계속 효율적으로 작동하도록 할 수 있습니다.

DO NOT COPY ssleeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



모델 품질 모니터링도 데이터 품질 모니터링과 비슷하지만 시간 경과에 따른 모델 성능을 중점적으로 평가합니다. 데이터 캡처, 기준 설정, 모니터링 일정 설정 등을 수행하는 전반적인 모니터링 방식은 동일하지만, 이 두 모니터링 작업에는 다음과 같은 중요한 몇 가지 차이점이 있습니다.

- 휴먼 인 더 루프(Human in the loop) - 두 모니터링 작업의 중요한 차이점 중 하나는 실제 레이블(Ground Truth)을 확인하는 휴먼 인 더 루프(Human in the loop)의 유무입니다. Ground Truth란 기계 학습 모델이 예측 또는 분류하려는 데이터의 실제 레이블 또는 값을 말합니다. Ground Truth가 있으면 모델 예측과 실제 사실의 일치 정도를 평가할 수 있습니다.
- 병합 작업 - 이 예제에서 모니터링 작업은 예측을 Ground Truth 데이터와 병합하여 비교를 수행합니다.
- 병합 파일 - 병합된 데이터는 모델 품질 드리프트 모니터링 실행의 중간 결과입니다. 병합된 데이터는 각 줄이 캡처된 데이터와 Ground Truth 데이터를 결합한 유효한 JSON 객체인 JSON 줄 파일에 저장됩니다.
- Amazon S3로 결과 전송 - 보고서, 통계, 위반은 자동으로 Amazon S3에 전송됩니다. 기준 성능과 비교한 모든 위반은 위반 파일에 캡처됩니다.
- CloudWatch – 모델 품질이 지정된 임계값을 벗어나면 알리도록 CloudWatch 경보를 구성할 수 있습니다. CloudWatch 경보를 사용하여 모델 재훈련 또는 훈련 데이터세트 업데이트 같은 문제 해결 작업을 호출할 수도 있습니다.

모델 품질을 모니터링하려면 SageMaker Model Monitor에 다음과 같은 입력이 필요합니다.

- 기준 데이터
- 배포된 모델이 하는 추론 입력 및 예측
- 모델에 대한 입력과 연결된 Amazon SageMaker Ground Truth

모델 품질 모니터링은 모델이 하는 예측을 SageMaker Ground Truth 레이블과 비교하여 모델의 품질을 측정합니다. 그리고 새로 레이블이 지정된 데이터를 사용하여 훈련 중에 최적화한 성능 지표를 기준으로 모델을 평가합니다. 또한 새 값을 모델 훈련 중에 확인된

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

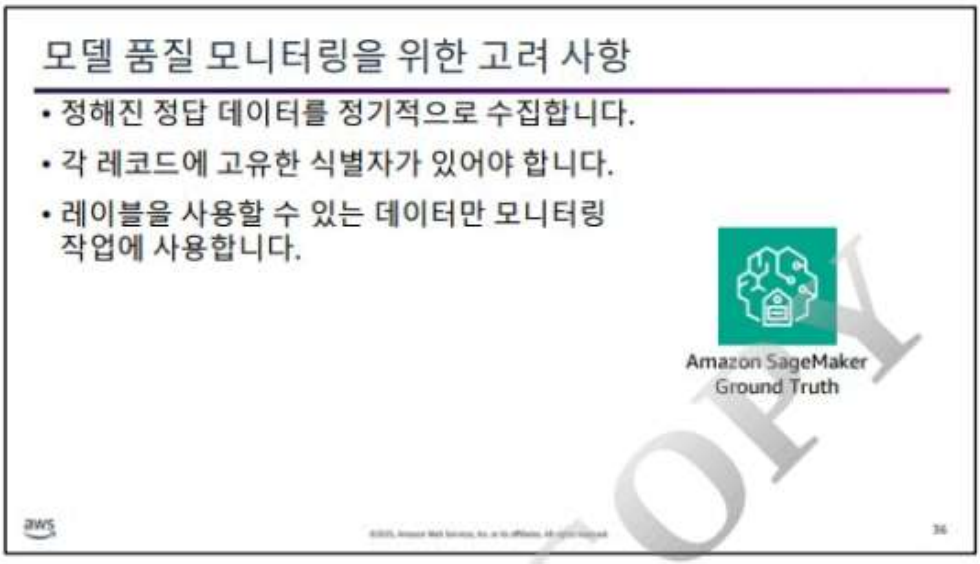
Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

값과 비교하여 정확도가 낮아질 가능성이 있는지를 파악합니다.

DO NOT COPY
ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



SageMaker Ground Truth는 ML 모델의 성공과 모델 품질 모니터링에서 중요한 역할을 합니다.

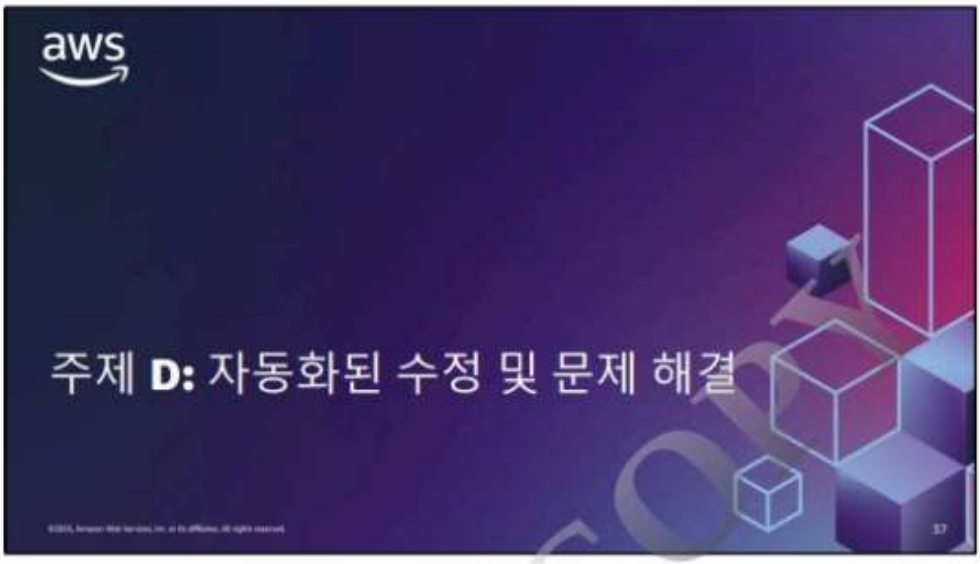
다음 3가지 고려 사항을 염두에 두어야 합니다.

- SageMaker Ground Truth 데이터 수집 - 실제 사용 사례에서는, SageMaker Ground Truth 데이터를 정기적으로 수집하여 지정된 Amazon S3 위치에 업로드해야 합니다. 이렇게 하면 모니터링 작업이 정확한 평가를 위해 레이블이 지정된 최신 데이터에 액세스할 수 있습니다.
- 고유 식별자 - SageMaker Ground Truth 레이블을 캡처된 예측 데이터와 일치시키려면 데이터세트의 각 레코드에 고유 식별자가 있어야 합니다. 이 식별자는 예측을 해당 Ground Truth 레이블과 올바르게 연결하는 데 사용됩니다.
- SageMaker Ground Truth 통합 - 데이터 레이블링에 SageMaker Ground Truth를 사용하는 경우 모니터링 작업이 레이블을 사용할 수 있는 데이터만 사용하도록 해야 합니다. 'EndpointInput' 구성의 'start_time_offset' 및 'end_time_offset' 파라미터를 조정하면 이를 수행할 수 있습니다. 이러한 파라미터는 모니터링 작업이 데이터를 선택하는 기간을 정의하여 선택한 데이터에 SageMaker Ground Truth 레이블을 사용할 수 있도록 합니다.

DO NOT COPY ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

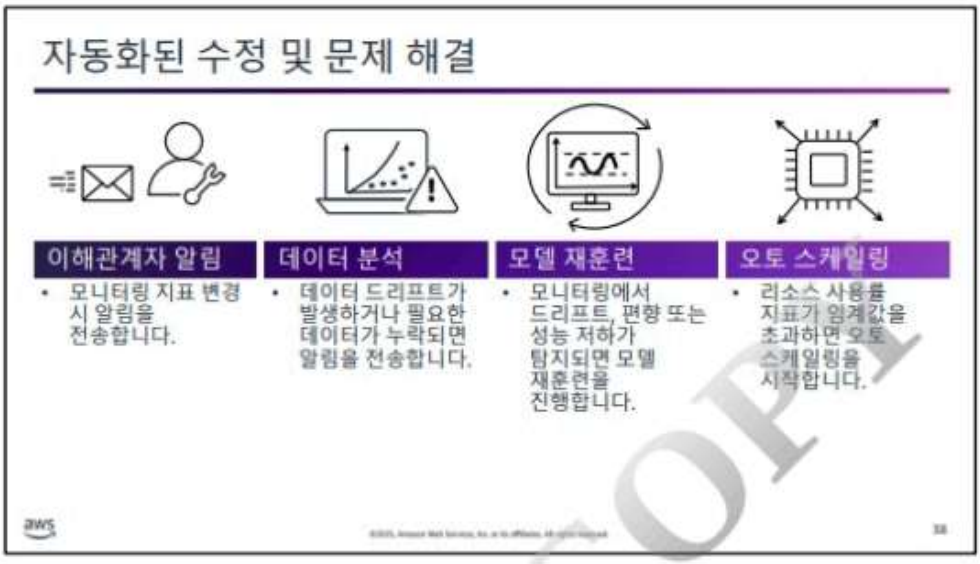


자동 문제 해결은 프로덕션 환경에서 안정적인 기계 학습 시스템을 유지하려면 꼭 필요한 구성 요소입니다. 이러한 프로세스에서는 포괄적인 모니터링을 통해 문제를 사전에 식별한 다음 해당 문제를 해결하는 자동 솔루션을 구현합니다.

이러한 자동 프로세스를 구현하면 최소한의 수동 작업만 수행하여 ML 모델의 성능과 신뢰성을 유지할 수 있습니다. 그러면 모델 성능을 일정한 상태로 유지할 수 있을 뿐 아니라 데이터 과학자와 ML 엔지니어가 더 복잡한 태스크와 획기적인 기능 개발에 더 많은 시간을 할애할 수 있습니다.

DO NOT COPY ssleeeee1@gmail.com

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



프로덕션 모니터링 구성이 완료되면 지표를 사용하여 자동화된 솔루션을 제공함으로써 신뢰성을 높일 수 있습니다. 모니터링 시스템이 모델이나 데이터에서 이상 또는 성능 저하를 탐지하면 자동화된 해결 전략을 사용하여 이러한 문제를 사전에 해결할 수 있습니다.

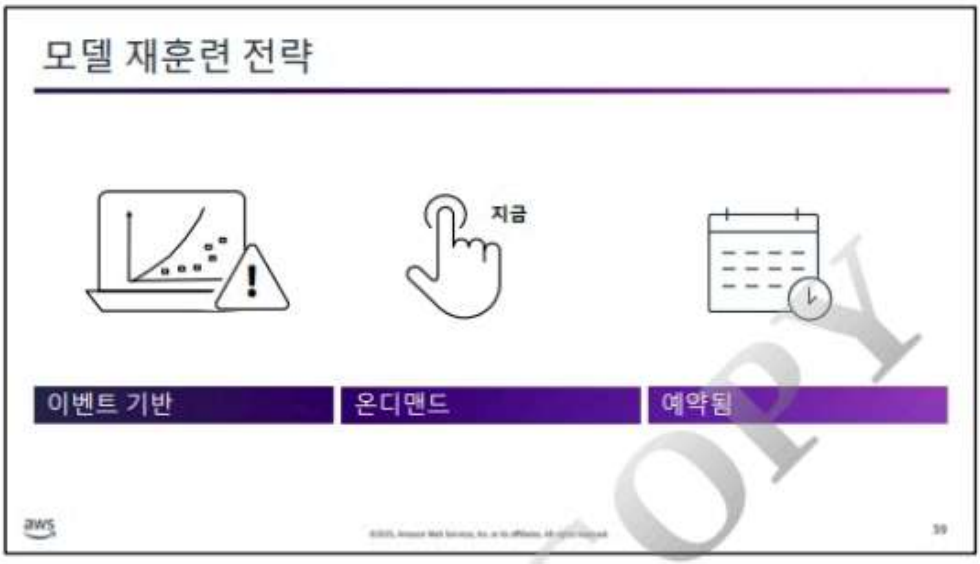
문제 시에 사용할 수 있는 몇 가지 일반적인 방식은 다음과 같습니다.

- 이해관계자에게 알림 전송 - 비즈니스 핵심 성과 지표(KPI) 또는 기본 문제에 영향을 미치는 변화가 모니터링 지표에 나타나면 관련 이해관계자에게 알림을 보내도록 메시징 시스템을 구성합니다. 이 방식을 적용하는 경우 문제 발생 시 적절한 관련자에게 즉시 알릴 수 있습니다.
- 데이터 분석 알림 - 모니터링에서 데이터 드리프트가 탐지되거나 예상 데이터가 누락된 경우 데이터 과학자에게 전송할 자동 알림을 설정합니다. 그러면 데이터 관련 문제를 빠르게 조사하여 해결할 수 있습니다.
- 모델 재훈련 - 모니터링에서 드리프트, 편향 또는 성능 저하가 탐지되면 모델을 자동으로 재훈련하도록 모델 훈련 파이프라인을 구성합니다. 그러면 시간 경과에 따라 데이터 분포가 변화하더라도 모델 정확도를 유지할 수 있습니다.
- 오토 스케일링 - 인프라 모니터링이 수집한 리소스 사용률 지표를 사용하여 오토 스케일링 작업을 시작합니다. 이 방식을 적용하면 ML 시스템이 다양한 로드를 효율적으로 처리할 수 있습니다.

이러한 자동 수정 전략을 구현하면 문제가 모델 성능이나 업무 결과에 큰 영향을 주기 전에 미리 해결할 수 있습니다. 이 방식을 적용하면 수동 작업은 최소한으로 줄이면서 ML 시스템의 신뢰성과 효율성을 유지할 수 있습니다.

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



모니터링 전략은 모델 성능이 표준 이하가 되는 시기를 식별하고 재훈련을 통해 시정 조치를 시작하는 데 중요합니다. 모델 및 데이터 모니터링 지표를 사용하여 모델 재훈련 방식에 필요한 정보를 제공합니다.

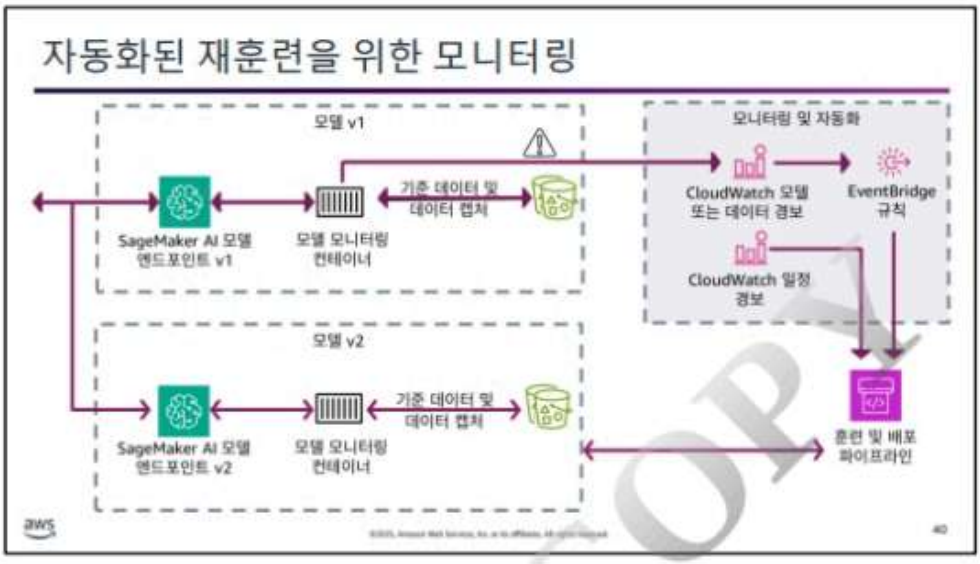
일반적인 재훈련 방식은 다음과 같습니다.

- 이벤트 기반 재훈련 - 모니터링 지표를 사용하여 특정 조건이 충족되면 재훈련을 시작하는 방식입니다. 예를 들어 특정 지표에 대한 경보를 생성하여 CloudWatch가 드리프트를 탐지하도록 구성할 수 있습니다. 이러한 경보는 모델을 재훈련하는 파이프라인을 시작할 수 있습니다. 이렇게 하면 업데이트는 제때 제공되지만, 재훈련에서는 리소스가 많이 사용될 수 있으므로 드리프트가 탐지될 때마다 모델을 재훈련하는 방식은 효율적이지 않을 수도 있습니다.
- 온디맨드 재훈련 - 업무상의 특정 이벤트나 변화하는 조건에 대응하여 재훈련을 진행하는 전략입니다. 예를 들어 새로운 경쟁업체가 등장하는 경우 변화하는 시장 상황에 맞게 모델을 재훈련해야 할 수 있습니다. 이 방식을 사용하는 경우 재훈련을 유동적으로 진행할 수 있지만 업무상의 상황을 적극적으로 모니터링해야 합니다.
- 예약된 재훈련 - 사전에 결정된 간격으로 모델을 재훈련시킵니다. 계절적 변화를 예측할 수 있는 환경에서 작동하는 모델의 경우 이 방식이 특히 유용합니다. 과거 모니터링 데이터를 분석하면 추론 데이터 및 모델 성능의 패턴을 파악하여 특정 이벤트가 발생하거나 특정 시기가 되기 전에 재훈련을 예약할 수 있습니다.

동적인 환경에서 작동하는 모델(예: 경쟁 가격에 의존하는 모델 또는 이자율이 변화하는 재무 모델)은 좀 더 유연한 방식으로 훈련해야 합니다. 이러한 상황에서는 모델에서 비즈니스 지표를 기준으로 성능 변화를 모니터링해야 합니다. 프로덕션 모델의 성능이 갑자기 저하되면 재훈련을 자동으로 시작해야 합니다. 평균 예측이 x이고 이 값이 특정 시간 간격 동안 10% 하락하는 경우 갑작스러운 개념 드리프트 문제가 발생했음을 의미할 수 있습니다.

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



실시간 추론 구성을 적용하면 온디맨드 및 예약 방식을 모두 사용하여 새 모델을 자동으로 재훈련하고 배포할 수 있습니다.

SageMaker AI 엔드포인트는 SageMaker Model Monitor를 사용하여 배포됩니다. SageMaker Model Monitor는 모니터링을 수행하기 위해 사전 구축된 컨테이너를 제공합니다. 모델 모니터링은 데이터 캡처와 기준 데이터 저장에 S3 버킷을 사용합니다.

다음의 2가지 방법 중 하나를 통해 모델 재훈련 작업을 자동화할 수 있습니다.

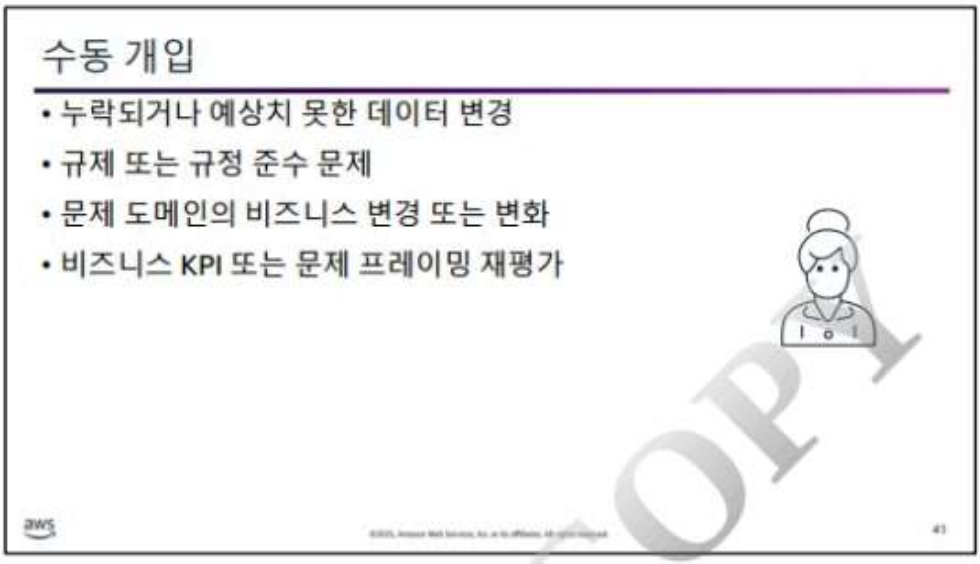
- 이벤트 기반 - 모델 또는 데이터 모니터링 지표가 특정 임계값을 위반하면 CloudWatch가 경보 상태를 시작합니다. 이 경보 상태로 인해 EventBridge 규칙이 훈련 및 배포 파이프라인을 시작할 수 있습니다.
- 예약 - CloudWatch가 특정 날짜와 시간에 경보 상태를 시작할 수 있습니다. 그러면 이 경보 상태가 훈련 및 배포 파이프라인을 시작할 수 있습니다.

훈련 및 배포 파이프라인은 모델을 재훈련하고, 엔드포인트 리소스를 배포하고, 새 모델의 기준을 설정합니다. 자동화된 테스트가 완료되면 트래픽이 새 엔드포인트로 리디렉션됩니다.

이러한 유형의 시스템을 구현하면 데이터 분포가 변화하더라도 장기적으로 성능을 유지할 수 있는 **자체 적응형 ML 솔루션**을 생성할 수 있습니다. 이 방식을 적용하면 수동 작업을 최소화할 수 있으며 계속해서 변화하는 환경에서도 모델의 효율성과 관련성을 유지할 수 있습니다.

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



자동화된 수정은 강력하지만 모니터링에서 식별된 일부 문제는 사람의 분석과 개입이 필요할 수 있습니다. Amazon SNS 같은 서비스를 통한 자동 알림은 이러한 문제를 조사하고 해결하도록 관련 직원에게 알릴 수 있습니다.

수동 개입이 필요할 수 있는 몇 가지 시나리오는 다음과 같습니다.

데이터 누락 또는 예상치 못한 데이터 변경
모니터링 도구에서 예상 데이터가 누락된 것을 발견할 수 있습니다. 예를 들어 추론 데이터를 생성하는 고객 양식이 업데이트되어 더 이상 특정 데이터 포인트를 수집하지 않는 것을 발견할 수 있습니다. 데이터의 특성도 변화할 수 있습니다. 예를 들어 교통 상황을 캡처하는 데 사용되는 카메라가 새 버전으로 업그레이드될 수 있습니다. 이 변경으로 인해 동영상 데이터가 크게 변화할 수 있습니다.

규제 또는 규정 준수 문제
데이터 또는 모델의 변경이 규제 및 규정 준수에 영향을 미치는 경우를 고려하는 것도 중요합니다. 예를 들어 애플리케이션이 익명화된 데이터를 사용하도록 설계되었을 수 있습니다. 그렇다면 애플리케이션이 보호되거나 민감한 정보를 수신하고 있는지 여부를 수동으로 검토해야 할 수 있습니다. 데이터 과학자는 의료나 금융과 같이 규제가 엄격한 업계의 맥락에서 모델이 내리는 추론을 조사해야 할 수도 있습니다.

문제 도메인의 비즈니스 변경 또는 변화
모델 성능을 평가할 때 비즈니스의 일부 측면이 변경되었음을 발견할 수 있습니다. 모기지 채무 불이행을 예측하는 애플리케이션은 이자율이나 대출 승인 요구 사항의 변경으로 인해 영향을 받을 수 있습니다. 또한 추론 응답이 오용되고 있음을 발견할 수도 있습니다.

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

비즈니스 KPI 또는 문제 프레이밍 재평가

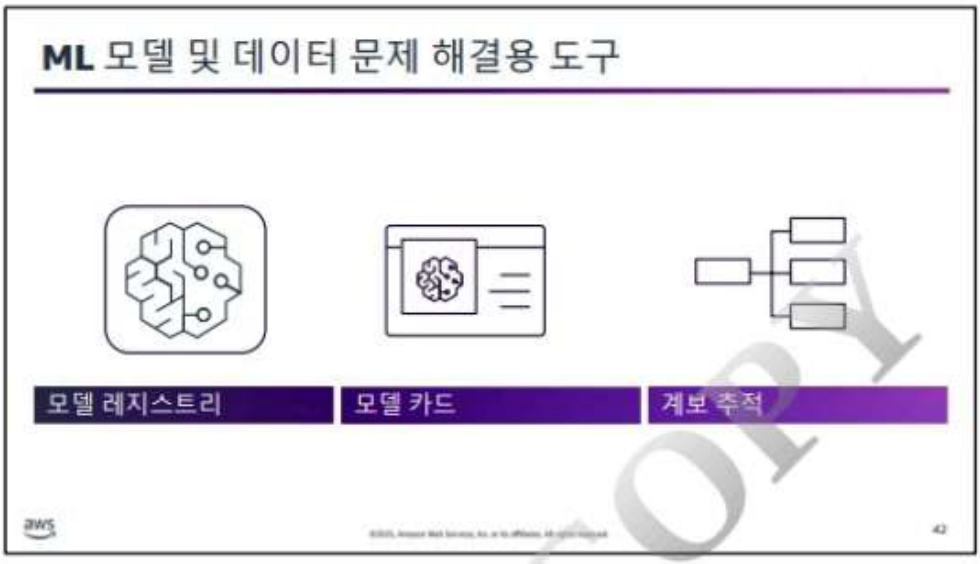
비즈니스 문제를 다시 살펴봐야 할 수도 있습니다.

- 애플리케이션이 사업 목표를 달성하고 있는지 판단하기 위해 올바른 KPI를 측정하고 있습니까?
- 올바른 항목을 측정하고 있습니까?
- 비즈니스에 변화가 발생했지만 설계한 ML 솔루션이 계속 적합한 상태입니까?"

DO NOT COPY
ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



ML 모델과 데이터의 문제를 조사할 때 유용하게 활용 가능한 몇 가지 주요 도구가 있습니다.

Amazon SageMaker Model Registry를 사용하면 프로덕션 환경에서 사용할 모델의 카탈로그 작성, 버전 관리, 메타데이터 연결, 승인 상태 제어 등의 작업을 수행할 수 있습니다. 중앙 리포지토리 역할을 하는 모델 레지스트리에서는 모델 반복 및 배포 기록을 추적할 수 있습니다.

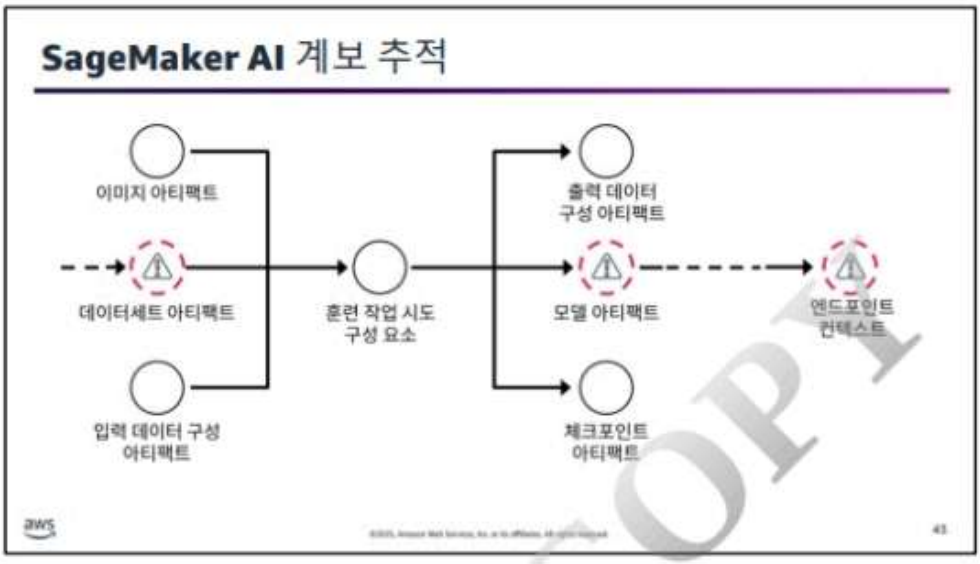
Amazon SageMaker 모델 카드를 사용하면 표준화된 방식을 통해 훈련 세부 정보, 평가 결과, 적절한 사용 방식 등의 모델 정보를 문서로 작성할 수 있습니다. 비즈니스 KPI에 대한 모델의 영향을 평가하고 모델 개발 과정을 명확하게 파악하려는 경우 이러한 정보가 특히 유용합니다.

Amazon SageMaker 계보 추적을 사용하면 데이터 준비에서 모델 배포까지의 ML 아티팩트 전체 수명 주기를 추적할 수 있습니다. 계보 추적 사용 시에는 문제가 있는 아티팩트의 출처를 추적하고 해당 아티팩트의 영향을 받았을 수 있는 후속 이벤트나 아티팩트를 확인하는 방식으로 문제의 원인을 조사할 수 있습니다.

모델 레지스트리는 기술 메타데이터 추적용이고 모델 카드는 비즈니스 및 규정 준수 메타데이터 추적용입니다. SageMaker AI에서는 모델 레지스트리와 모델 카드를 편리하게 통합하여 모델 카드 대시보드에서 모든 기술, 비즈니스 및 규정 준수 메타데이터를 직접 확인할 수 있습니다. 이러한 도구를 활용하면 ML 워크플로에서 문제를 더욱 쉽게 해결할 수 있습니다.

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



문제를 해결할 때 SageMaker AI 계보 추적 을 사용할 수 있습니다. 모니터링 시스템에서 아티팩트 중 하나의 문제에 대한 알림을 받으면 해당 아티팩트를 생성한 리소스를 찾을 수 있습니다. 해당 아티팩트로 생성된 리소스도 찾을 수 있습니다.

훈련 작업을 실행하고, SageMaker AI 모델 레지스트리에 모델을 등록하고, 모델을 엔드포인트에 배포하는 워크플로를 생각해 보십시오. 이 워크플로에는 훈련에 사용되는 데이터세트, 훈련 작업 및 시험 구성 요소, 모델 레지스트리의 모델 패키지, SageMaker AI 엔드포인트가 포함됩니다. 이제 데이터세트에 오류가 있어서 해당 데이터를 사용하여 해당 데이터를 사용하여 훈련된 모델이 있는 모든 엔드포인트를 롤백해야 한다고 가정해 보겠습니다.

이 경우 SageMaker AI 계보 추적을 사용하여 다음을 수행할 수 있습니다.
- 문제가 있는 모델을 훈련하는 데 사용된 데이터세트를 찾습니다.
- 해당 모델을 사용하여 배포한 엔드포인트 또는 애플리케이션을 식별합니다.

이는 문제의 근본 원인과 잠재적 다운스트림 영향을 조사하는 데 도움이 되므로 집중적 수정 작업이 용이해집니다.

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

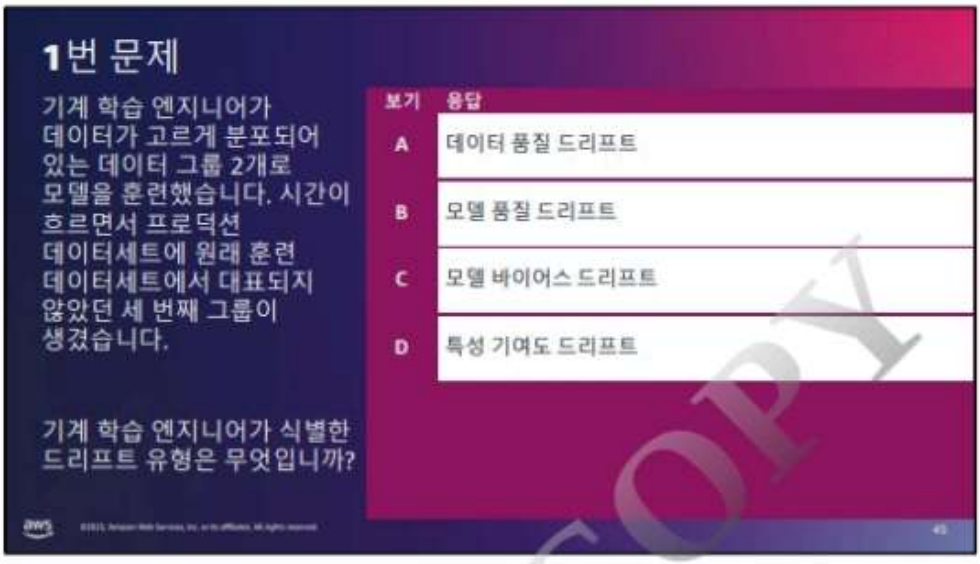
Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

이 슬라이드에서는 지식 확인 섹션을 소개합니다.
이제 간단한 지식 확인을 진행하여 이 모듈에서 알아본 모델 성능 및 데이터 품질 모니터링 관련 주요 개념을 복습해 보겠습니다.

이 지식 확인을 통해 다음 모듈로 넘어가기 전에 수업 내용을 더욱 정확하게 이해하고 헷갈리는 부분을 확실하게 파악할 수 있습니다.

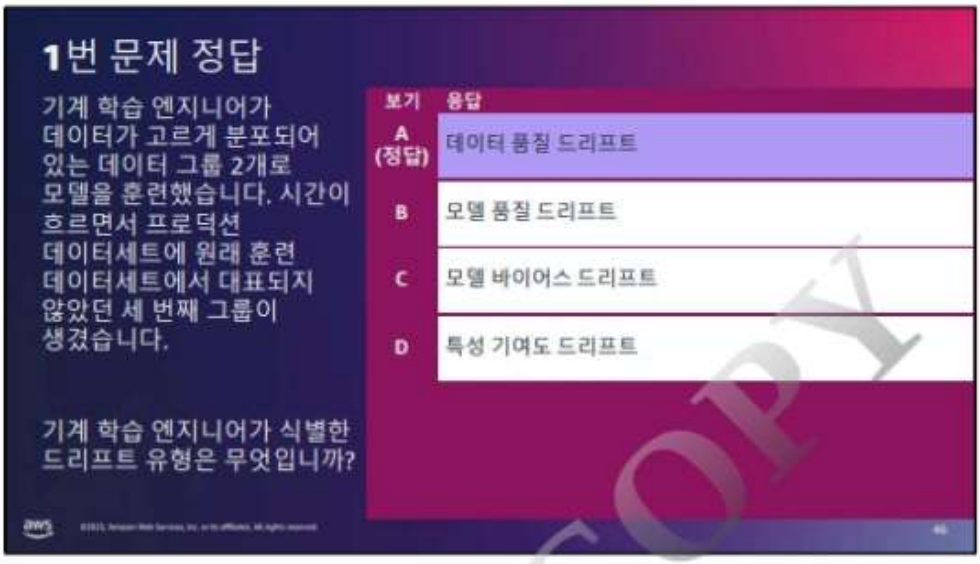
DO NOT COPY
ssleeeee1@gmail.com

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



DO NOT COPY
ssleeeee1@gmail.com

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



정답은 다음과 같습니다.
A. 데이터 품질 드리프트에는 기존 특성에서 표시되는 새 범주와 통계 속성 또는 데이터 유형의 변화가 모두 포함됩니다.

오답은 다음과 같습니다.
B. 모델 품질 드리프트는 새 그룹 추가로 인해 편향이 발생하는 현상이 아니라 정확도나 정밀도 등 모델의 전반적인 성능 변화와 관련된 현상입니다.
C. 모델 바이어스 드리프트는 그룹의 분산 증가와 같은 프로덕션 데이터 분포의 변화에 해당하는 현상입니다. 이 드리프트가 발생하면 모델을 훈련하는 데 사용된 데이터가 그룹을 과소 대표하는 상태가 됩니다.
D. 특성 기여도 드리프트(개념 드리프트)란 입력 특성과 대상 변수 간의 관계 변화를 지칭하는 개념이며 데이터에 새 그룹이 추가되어 발생하는 현상은 아닙니다.

DO NOT COPY ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

2번 문제

사용자 행동 데이터를 토대로 고객 유지 확률을 예측하는 기계 학습 모델이 있습니다. 그런데 일정 시간이 경과하자 특정 인구 통계 그룹의 예측 정확도가 낮아졌습니다.

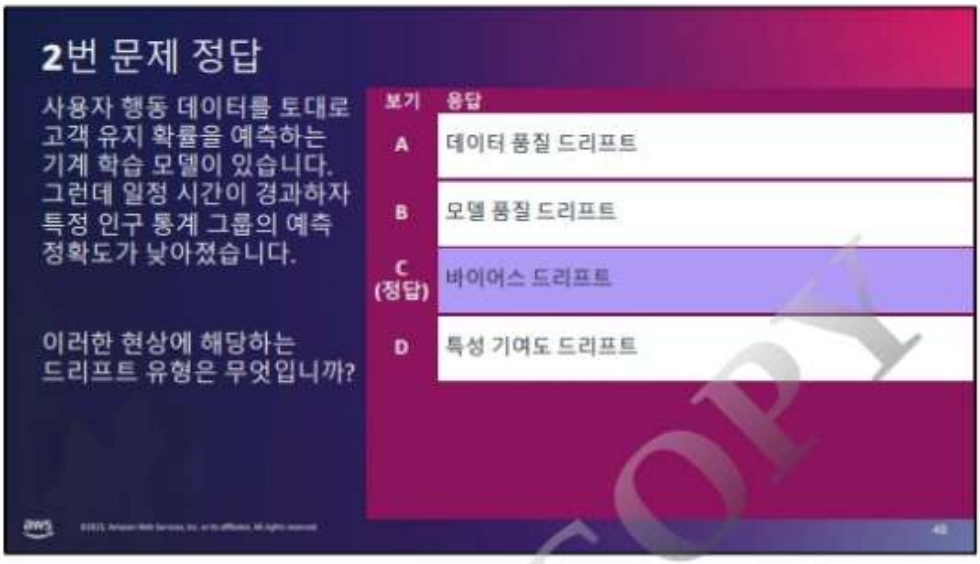
이러한 현상에 해당하는 드리프트 유형은 무엇입니까?

보기	응답
A	데이터 품질 드리프트
B	모델 품질 드리프트
C	바이어스 드리프트
D	특성 기여도 드리프트

47

DO NOT COPY
ssleeeee1@gmail.com

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

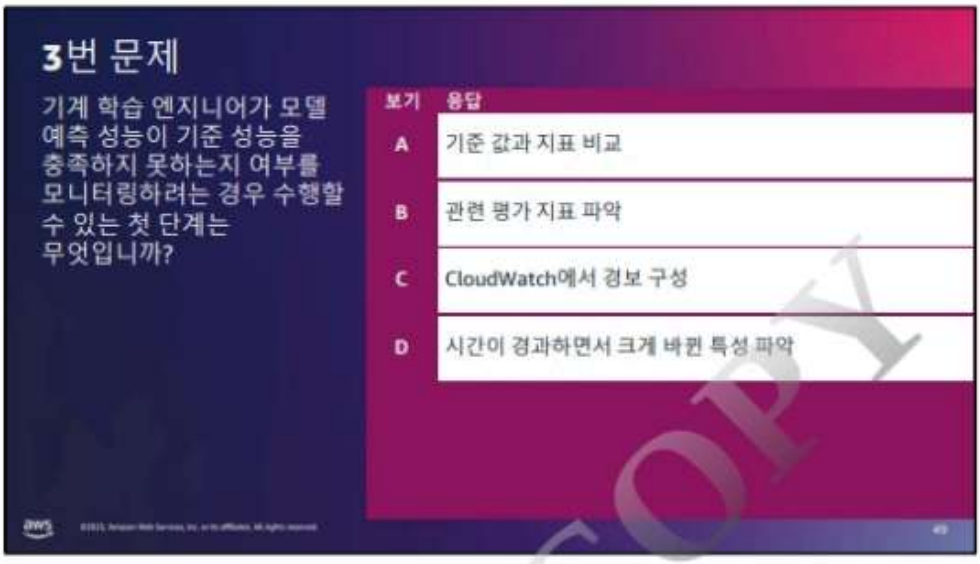


정답은 다음과 같습니다.
C. 바이어스 드리프트는 시간 경과에 따라 특정 인구 통계 그룹에 대한 모델 예측의 정확도나 공정성이 떨어지는 현상입니다. 이 시나리오는 특정 인구 통계 그룹과 관련된 모델 예측 정확도가 낮아진 상황에 해당하므로, 해당 그룹과 관련하여 모델 예측의 편향이 발생했거나 심화되었다고 할 수 있습니다.

오답은 다음과 같습니다.
A. 데이터 품질 드리프트는 시간 경과에 따라 특성 분포가 변화하거나 새 범주가 추가되어 입력 데이터의 통계적 속성이 변경되는 경우에 발생합니다. 그러나 이 시나리오의 문제는 입력 데이터의 전반적인 변화가 아니라 특정 인구 통계 그룹에 대한 모델 성능과 관련된 것입니다.
B. 모델 품질 드리프트는 모델의 전반적인 예측 성능이 저하되는 현상입니다. 예측 성능은 Ground Truth 레이블에 예측을 비교하는 방식으로 측정하는 경우가 많습니다. 이 시나리오에서도 예측 정확도가 낮아지기는 하지만 해당 문제는 전체 데이터세트에서 발생하는 것이 아니라 특정 인구 통계 그룹에서만 발생합니다.
D. 특성 기여도 드리프트(개념 드리프트)는 시간이 경과하면서 예측 시에 반영되는 특성의 상대적 중요도가 바뀌는 경우에 발생합니다. 이 시나리오는 특정 인구 통계 그룹에 대한 예측 정확도 저하와 관련된 것이며 특성 중요도 변화에 대한 정보는 제공하지 않습니다.

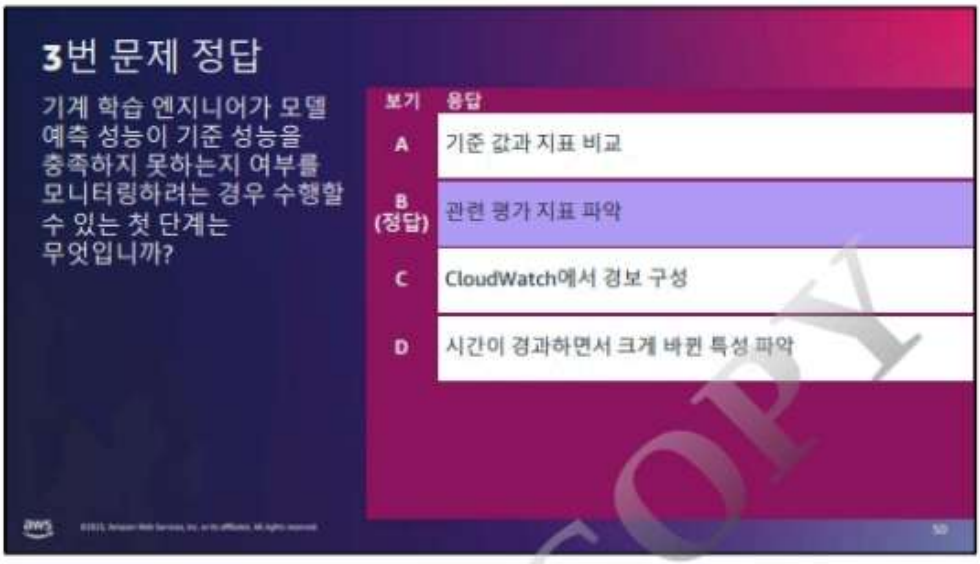
© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



DO NOT COPY
ssleeeee1@gmail.com

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

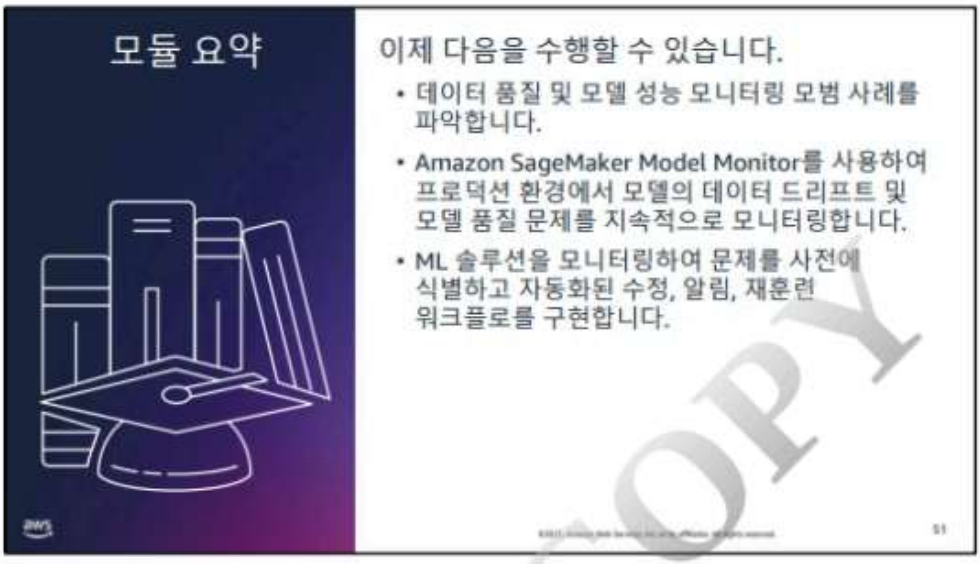


정답은 다음과 같습니다.
B: 비즈니스 요구 사항을 토대로 평가 지표(예: 정확도, 정밀도, 재현율) 및 임계값을 식별합니다. 그리고 이러한 지표를 사용하여 모델 성능의 기준을 생성한 후 시간이 경과함에 따라 선택한 임곗값보다 큰 변화가 나타나는지를 모니터링합니다.

오답은 다음과 같습니다.
A. 프로덕션 환경의 지표를 기준과 비교하기 전에 비즈니스 요구 사항을 기준으로 비즈니스 지표와 임곗값을 확인한 후 기준을 설정해야 합니다.
C. CloudWatch 경보를 구성하기 전에 지표와 임곗값을 확인해야 합니다.
D. 이 단계는 특성 기여도 드리프트를 모니터링할 때 더 적합하며 모델 품질 문제를 모니터링하는 단계는 아닙니다.

DO NOT COPY ssleeeee1@gmail.com

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



이 모듈에서는 기계 학습 시스템에서 데이터 품질 및 모델 성능을 모니터링할 때 준수해야 하는 모범 사례를 살펴보았습니다. 그리고 Amazon SageMaker M1odel Monitor를 사용하여 프로덕션 모델의 데이터 드리프트 및 모델 품질 문제를 지속적으로 모니터링하는 방법을 알아보았습니다. 또한 ML 솔루션을 모니터링하여 문제를 사전에 식별하고 자동화된 문제 해결, 알림, 재훈련 워크플로를 구현하는 방법도 살펴보았습니다. 프로덕션 환경에서 ML 모델의 성능과 신뢰성을 유지하려면 이러한 기술을 반드시 습득해야 합니다.

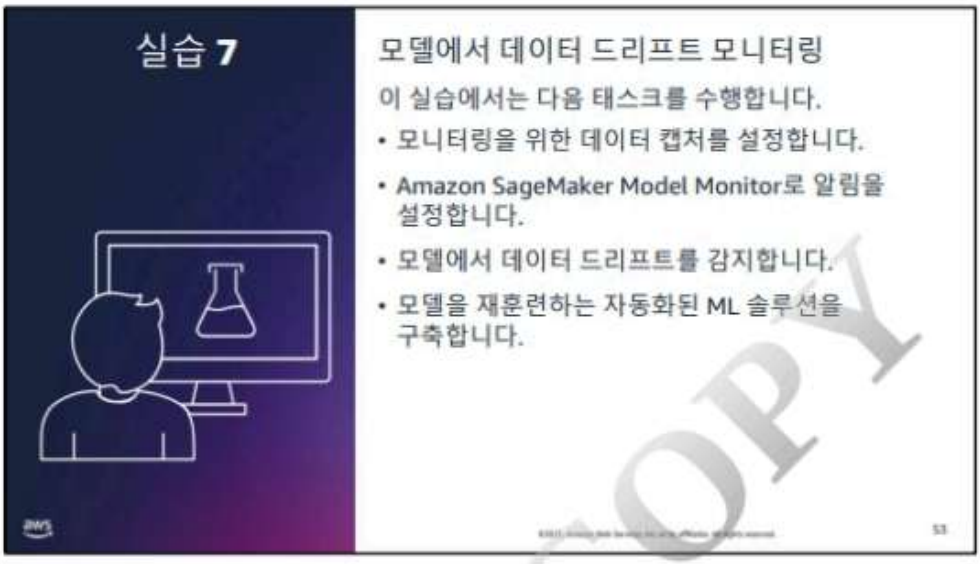
DO NOT COPY
ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

DO NOT COPY
ssleeeee1@gmail.com

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

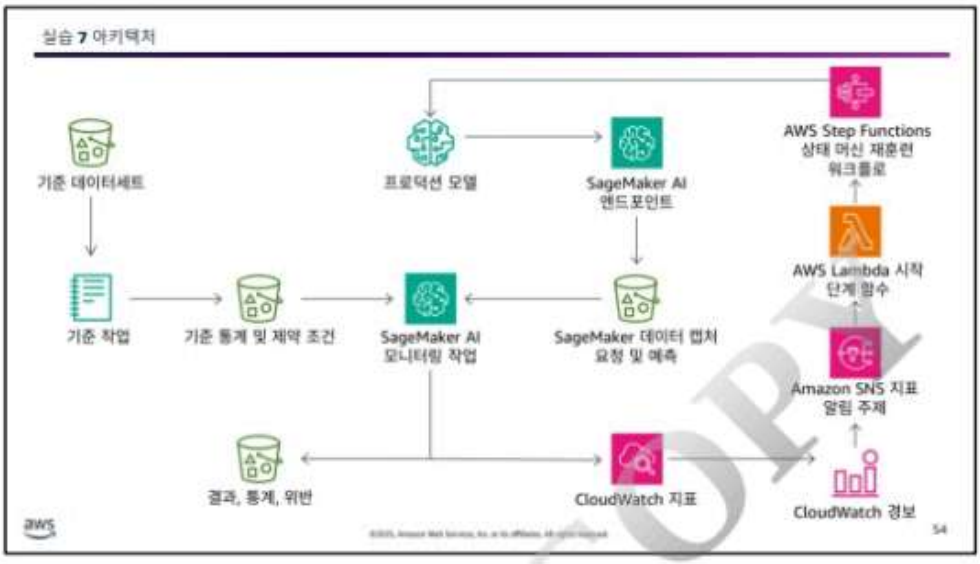


이 실습에서는 모듈에서 알아본 개념을 실제 시나리오에 적용합니다. 구체적으로는 모니터링을 위한 데이터 캡처를 설정하고 Amazon SageMaker Model Monitor를 사용하여 알림을 구성합니다. 또한 모델의 데이터 드리프트를 탐지하는 방법과 필요시 모델을 재훈련하는 자동화된 ML 솔루션을 구축하는 방법을 알아봅니다. 이 실습에서는 프로덕션 환경의 ML 모델에 사용할 포괄적인 모니터링 솔루션을 구현해 볼 수 있습니다. 이러한 연습은 실제 문제를 해결할 때 큰 도움이 됩니다.

DO NOT COPY
ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

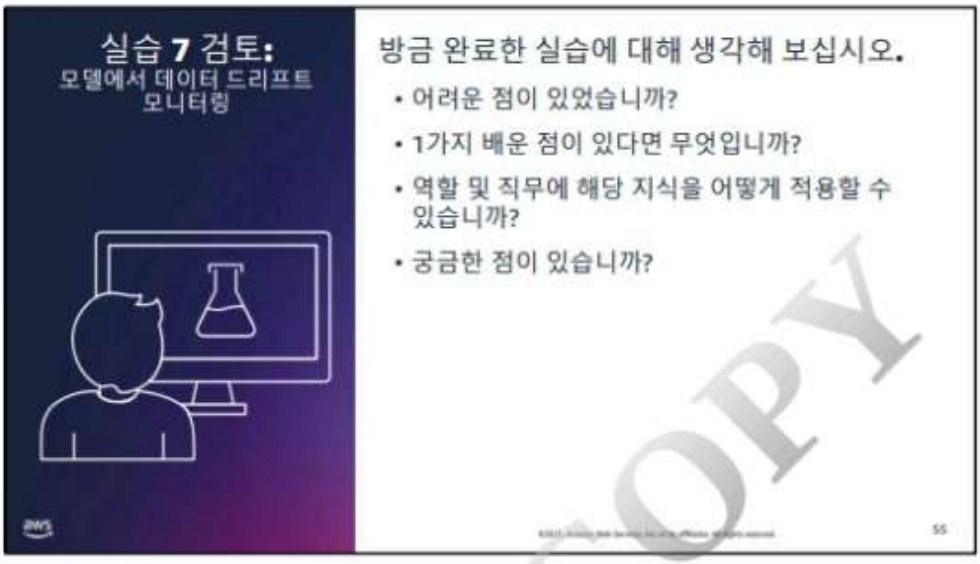


AnyCompany Consulting에서는 현재 전복 회귀 모델을 개발하고 있습니다. 회사는 실시간 추론에 사용되는 데이터를 모니터링하고 새 데이터에서 발생하는 데이터 드리프트를 감지하여 MLOps 안정성을 지원하려고 합니다. 데이터 드리프트가 감지되면 자동으로 모델을 재훈련해야 합니다. 실습 아키텍처에는 ML 모델을 모니터링하고 자동으로 재훈련하는 전체 워크플로가 나와 있습니다.

이미지 설명:
SageMaker AI 엔드포인트에 연결된 프로덕션 모델이 나와 있는 다이어그램. 엔드포인트는 SageMaker AI 데이터 캡처를 사용하여 데이터를 캡처하도록 구성됩니다. SageMaker AI 모니터링 작업에서는 요청과 예측을 사용하여 기준 처리 작업을 통해 데이터세트에서 처리되는 기준과 새 추론 레코드를 비교합니다. SageMaker AI 모니터링 작업은 결과, 통계, 위반을 생성하며 CloudWatch 로그 그룹으로 구성됩니다. CloudWatch 경보는 새 추론 레코드가 기준에서 설정된 범위를 벗어나는 경우 Amazon Simple Notification Service(SNS) 지표 알림 주제에 알림을 보내도록 구성됩니다. Lambda 함수는 SNS 주제를 구독하고 새 데이터로 모델을 재훈련하고 새 모델을 프로덕션 엔드포인트에 배포하는 Step Functions 상태 머신 워크플로를 호출합니다.

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



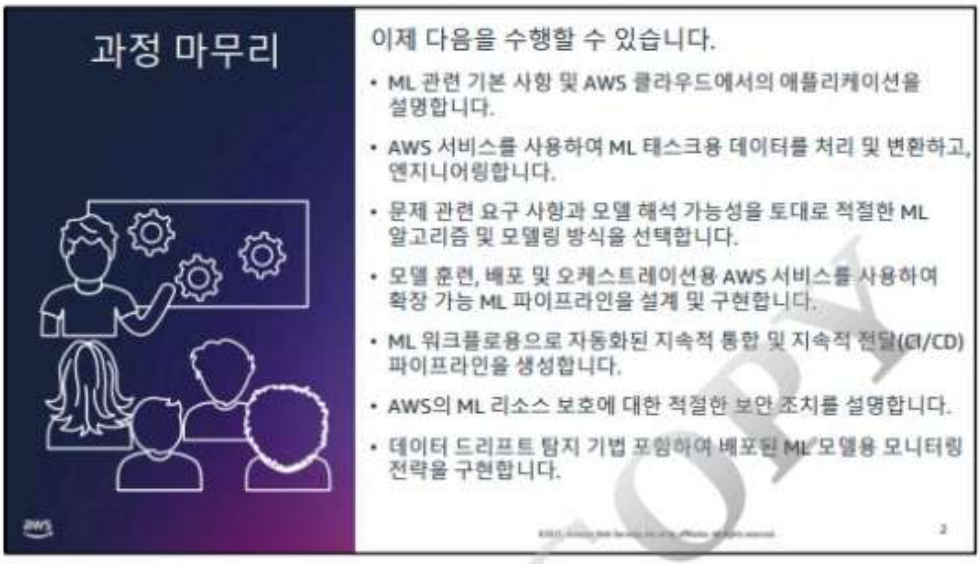
DO NOT COPY
ssleeeee1@gmail.com

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

이 모듈에서는 과정에서 배운 내용을 다시 생각해 보고 궁금한 점을 추가로 질문할 수 있습니다. 이 주제, AWS 서비스 및 AWS 자격증에 대해 자세히 알아볼 수 있는 리소스도 제공됩니다.

DO NOT COPY
ssleeeee1@gmail.com

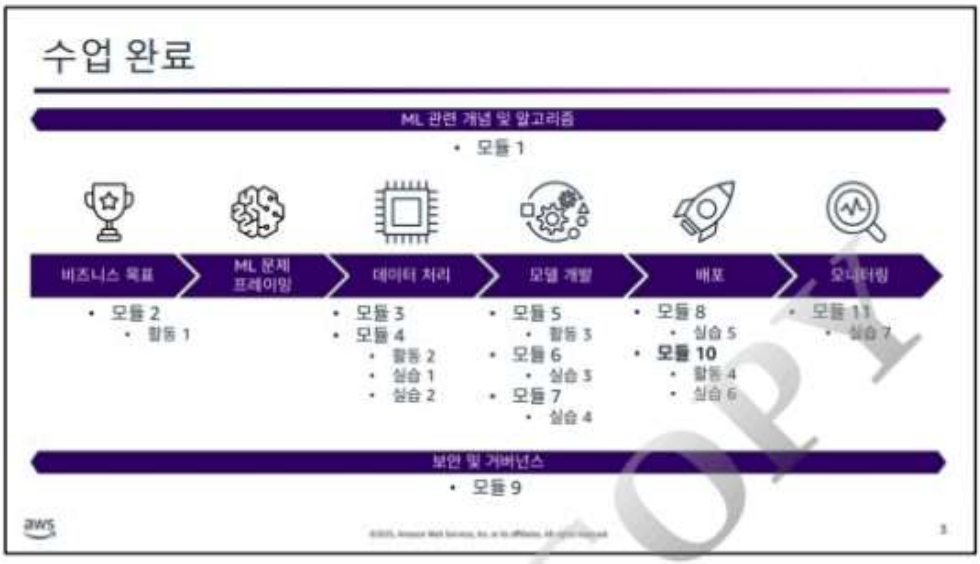
Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



이 과정에서는 모델 훈련, 배포 및 오케스트레이션용 AWS 서비스를 사용하는 방법을 연습해 보았습니다. ML 워크플로용으로 자동화된 CI/CD 파이프라인을 생성하면 반복적인 모델 개발 프로세스의 효율성을 높일 수 있습니다. 그리고 ML 리소스 보호를 위해 AWS에서 제공되는 보안 조치도 파악했습니다. 마지막으로 배포된 ML 모델용 모니터링 전략(데이터 드리프트 탐지 기법 포함)을 구현하는 방법을 알아보았습니다.

DO NOT COPY
ssleeeee1@gmail.com

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



이 과정에서는 ML 관련 개념과 알고리즘부터 대략적으로 살펴보았습니다. 그런 다음 ML 수명 주기, 그리고 AWS 서비스가 해당 수명 주기의 각 단계를 지원하는 방식을 살펴보았습니다.

이 과정에서 다룬 주제 및 개념은 다음과 같습니다.

- 아이디어 구상에서 프로덕션 환경으로의 배포 및 모니터링까지 전체 ML 수명 주기에 대해 알아보았습니다.
- 업무상의 문제를 토대로 ML 문제를 프레이밍하는 방법을 살펴보았습니다.
- AWS 서비스를 사용하여 ML 훈련에 사용할 데이터를 처리하는 방법(서식 지정, 정리, 특성 추출 포함)을 알아보았습니다.
- 모델을 구축, 훈련, 평가하고 이 과정을 반복하여 최종적으로 요구 사항을 충족하는 모델을 선택하는 모델 개발 과정을 살펴보았습니다. Amazon SageMaker AI를 사용하여 모델 훈련, 평가, 튜닝을 수행하는 방법을 알아보았습니다.
- 배포 대상 및 기법, 그리고 A/B 테스트를 사용하여 모델 버전을 배포하는 방법을 알아보았습니다.
- 전체 수명 주기에서 AWS 서비스가 보안 및 거버넌스 요구 충족을 지원하는 방식을 확인했습니다.
- ML 모델에 모니터링이 적용되는 방식, 그리고 데이터 드리프트를 식별하도록 Amazon SageMaker Model Monitor를 구성하는 방법을 알아보았습니다.

프레젠테이션, 토론, 활동 및 실습을 진행하여 전체 ML 수명 주기에서 AWS의 ML 엔지니어링 기능을 활용하는 방법에 대한 인사이트를 파악했습니다.

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

DO NOT COPY
ssleeeee1@gmail.com

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

온라인 학습 센터인 AWS Skill Builder로 학습을 계속합니다.

원하는 속도로 목표를 달성할 준비가 되었습니까? AWS Skill Builder의 무료 디지털 교육은 필요한 기술을 원하는 방식으로 학습할 수 있도록 500개가 넘는 온디맨드 과정과 학습 플랜을 제공합니다.

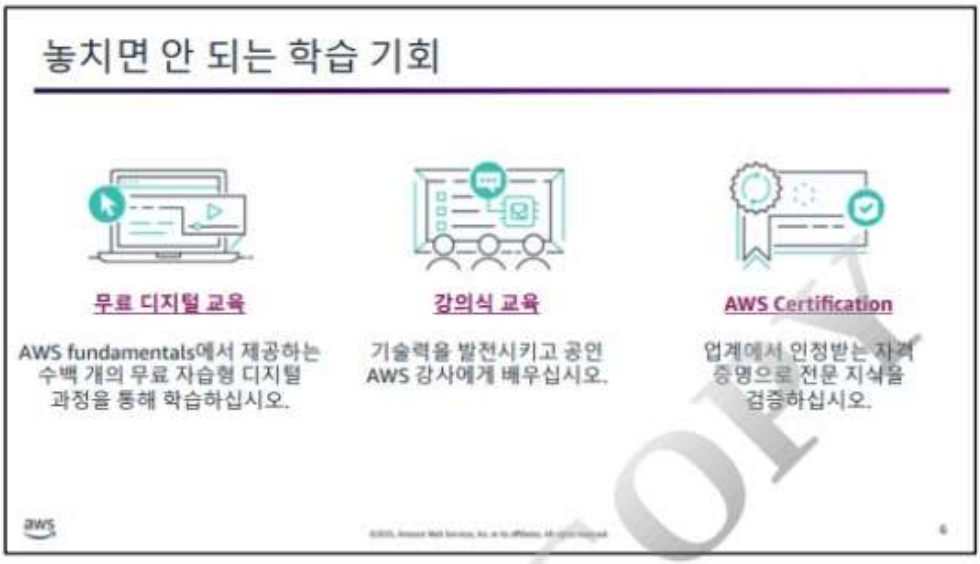
참여를 유도하는 대화형 환경을 통해 문제 해결 클라우드 기술을 학습하고 싶습니까? Skill Builder를 구독하면 자습형 실습(SPL), 연습 시험, 역할 기반 게임 및 실제 과제를 통해 빠르게 배울 수 있습니다.

자세히 알아보고 시작하려면 AWS Skill Builder(https://aws.amazon.com/training/digital)를 참조하십시오.

DO NOT COPY ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



AWS Training and Certification은 AWS의 지식을 확장 및 심화하고 AWS 서비스를 널리 퍼트리기 위해 노력하는 조직입니다. 이 프로그램은 고객, AWS 파트너 및 AWS 직원을 위해 설계되었습니다. 지난 수개월 동안 고객과 파트너를 위한 여러 신규 과정, 교육 실습, 자격증을 선보였습니다.

AWS 클라우드 스킬을 확장하십시오. 자세한 내용은 다음 페이지를 참조하십시오.

- **디지털 교육** - https://explore.skillbuilder.aws/
- **강의식 교육** - https://aws.amazon.com/training
- **AWS Certification** - https://aws.amazon.com/certification
- **AWS 워크숍** - https://workshops.aws/
- **Tech Talks** - https://aws.amazon.com/events/online-tech-talks/on-demand/
- **AWS 램프업 가이드** - https://aws.amazon.com/training/ramp-up-guides/

DO NOT COPY ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

AWS Certification은 업계에서 인정하는 보안 자격증으로 클라우드 전문 지식을 검증함으로써 신뢰와 확신을 쌓는 데 도움이 됩니다. 이 자격증은 조직에서 Amazon Web Services(AWS)를 사용하여 클라우드 이니셔티브를 이끌 수 있는 숙련된 전문가를 파악하는 데 도움이 됩니다.

슬라이드에는 현재 제공되는 AWS Certification이 나와 있습니다. AWS Certification을 취득하려면 감독관 참관 시험에서 합격 점수를 받아야 합니다. 역할 기반 자격증의 경우 AWS Cloud Services의 각 자격증 레벨에 권장되는 경험 수준은 다음과 같습니다.

- **Professional:** AWS 클라우드를 통한 솔루션 설계, 운영, 문제 해결 경험 2년
- **Associate:** AWS 클라우드를 통한 문제 해결 및 솔루션 구현 경험 1년
- **Foundational:** 기초 AWS 클라우드 경험 및 업계 지식 습득 경험 6개월

Specialty 자격증은 특정 기술 분야에 중점을 두고 있습니다. Specialty 시험 응시에 권장되는 경험은 시험 가이드에 지정된 영역에서의 기술 경험입니다.

AWS는 자격증 시험에서 다루는 모든 서비스 또는 기능의 목록을 게시하지 않습니다. 그러나 각 시험의 시험 가이드에는 시험에서 다루는 현재 주제 영역 및 목표가 나열되어 있습니다. 자세히 알아보고 시험 가이드와 기타 준비 자료를 복습하려면 AWS Certification 시험 준비(https://aws.amazon.com/certification/certification-prep/)를 참조하십시오.

이 정보는 2024년 10월 기준으로 작성되었습니다. 그러나 시험은 자주 업데이트되며 제공되는 시험과 각 시험의 테스트 항목에 관한 세부 정보는 변경될 수 있습니다. 최신 AWS Certification 시험 정보는 AWS Certification(https://aws.amazon.com/certification/)을 참조하십시오.

3년마다 자격증을 갱신(또는 재인증)해야 합니다. 자세한 내용은 AWS 자격증 갱신 (https://aws.amazon.com/certification/recertification/)을 참조하십시오.

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

이 과정에는 AWS Certification 시험과 관련이 있을 수 있는 콘텐츠가 포함됩니다. 시험 준비를 계속하려면 핵심적인 다음 4개 단계를 따르십시오.

각 시험에 관한 자세한 내용을 보려면 QR 코드를 스캔하여 AWS Certification 시험 살펴보기(https://aws.amazon.com/certification/exams/)를 참조하십시오.

DO NOT COPY
ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

1단계는 시험 및 시험 유형의 문제에 대해 알아보는 것입니다.

AWS Certification 시험 페이지에서 각 시험의 시험 안내서를 검토할 수 있습니다. 자세한 내용은 **모든 AWS Certification 시험 살펴보기**(https://aws.amazon.com/certification/exams/)를 참조하십시오.

샘플 시험 문제를 보려면 AWS Skill Builder에 가입합니다. Skill Builder 내에서 Official Practice Question Set에 등록할 수 있습니다. 자세한 내용은 **AWS Skill Builder: 수요가 많은 클라우드 기술을 구축하기 위한 학습 센터**(https://explore.skillbuilder.aws/learn)를 참조하십시오.

실제 AWS Certification 시험에 나올 수 있는 문제와 동일한 프로세스에 따라 연습 세트의 문제가 생성됩니다. 여기에는 시험을 준비하는 데 도움이 되는 자세한 피드백과 권장 리소스가 포함되어 있습니다.

DO NOT COPY ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

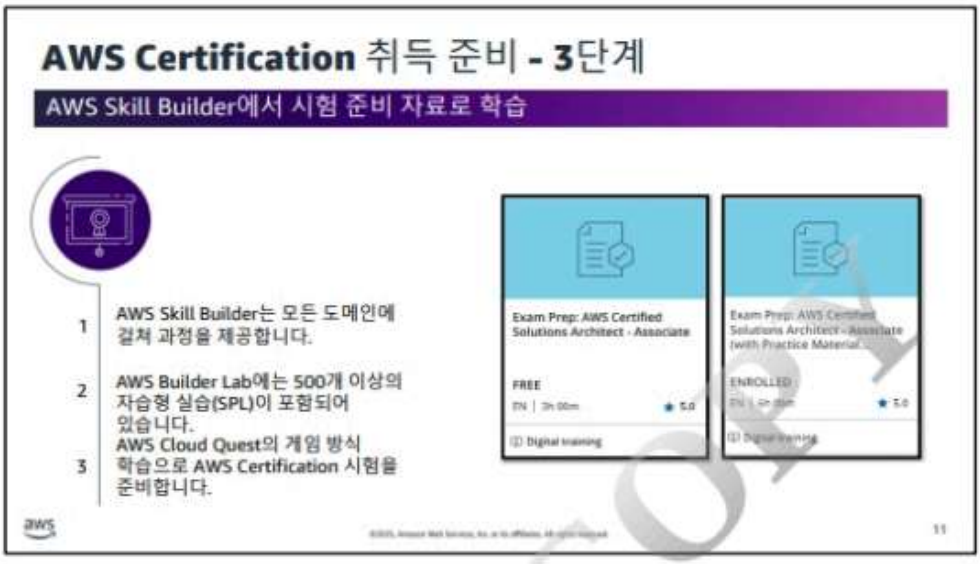
2단계에서는 시험 주제를 복습합니다. AWS Skill Builder에서 시험 가이드를 검토하고 자습형 실습 과정에 등록하는 것 외에도 AWS Builder Lab을 살펴보면서 AWS에 대한 실습을 진행해 볼 수 있습니다.

자세한 내용은 AWS Builder Lab(https://aws.amazon.com/training/digital/aws-builder-labs/)을 참조하십시오.

DO NOT COPY
ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



다음으로 AWS Skill Builder에서 시험 준비 과정을 수강할 수 있습니다. Skill Builder는 여러분의 지식에서 부족한 부분을 해결하는 데 사용할 수 있는 많은 리소스도 제공합니다.

Skill Builder는 모든 영역의 교육 과정을 제공합니다. AWS Builder Lab에는 500개 이상의 자습형 실습(SPL)이 포함되어 있습니다. 그리고 게임을 플레이하여 AWS 서비스를 직접 경험하고 싶은 경우에는 AWS Cloud Quest를 시도해 볼 수 있습니다.

참고: 이러한 리소스 중 일부는 디지털 구독이 필요합니다.

DO NOT COPY
ssleeeee1@gmail.com

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

마지막으로, 공식 연습 시험에 응시하여 실전 시험 준비를 확인하십시오.

각 연습 시험에는 실제 시험과 동일한 수의 문제가 포함되어 있습니다. 연습 시험은 자격증 시험과 동일한 문제 스타일과 난이도에 같은 수준의 엄격함을 적용하여 연습을 진행합니다. 여기에는 시험 방식의 채점과 통과 또는 실패가 포함됩니다. 또한 각 문제의 선택 답안에 대한 피드백과 함께 핵심 주제에 대한 이해 심화를 도울 수 있는 권장 리소스가 함께 제공됩니다. 또한 제한 시간이 있고 시험이 끝난 후에만 정답이 공개되는 모의 시험에 응시하여 시험을 미리 경험해 볼 수 있습니다. 또는 제한 시간이 없는 시험이나 각 문제의 답을 제출할 때마다 정답이 공개되는 시험 등의 다른 옵션을 선택할 수도 있습니다.

DO NOT COPY
ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

AWS는 유연하고 편리한 시험 응시 옵션을 제공합니다.

자세히 알아보고 자신에게 가장 적합한 시험 옵션을 선택하려면 '시험 예약'(https://aws.amazon.com/certification/certification-prep/testing/)을 참조하십시오.

DO NOT COPY
ssleeeee1@gmail.com

Printed by: ssleeeee1@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

DO NOT COPY
ssleeeee1@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.