

6. 세이지메이커 시작하기

2강. 세이지메이커 작동방식

학습목표

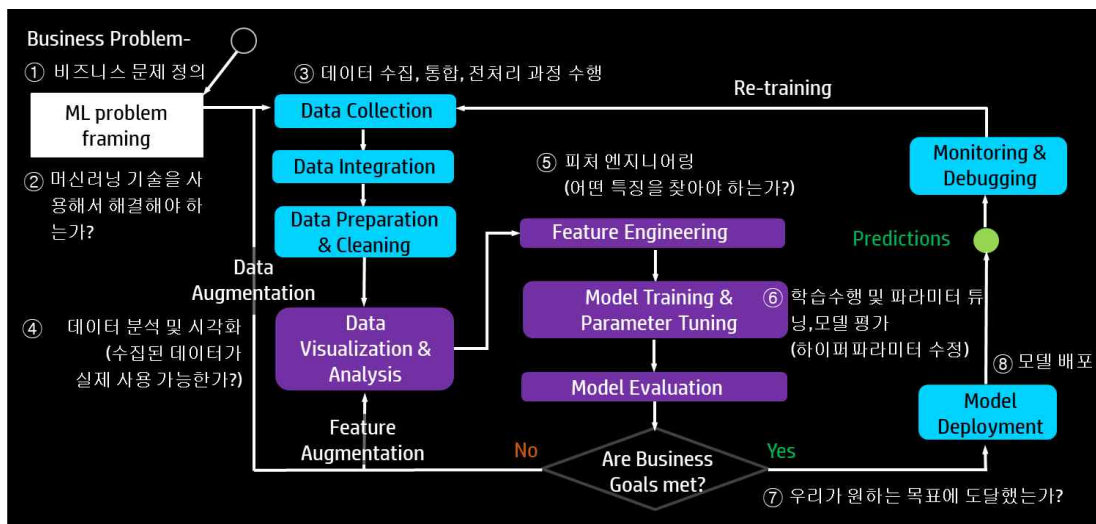
- 머신러닝 워크플로우와 세이지메이커를 비교하여 설명할 수 있다.
- 세이지메이커 워크플로우 단계를 열거할 수 있다.
- 세이지메이커 ML 알고리즘을 종류별로 분류할 수 있다.

학습내용

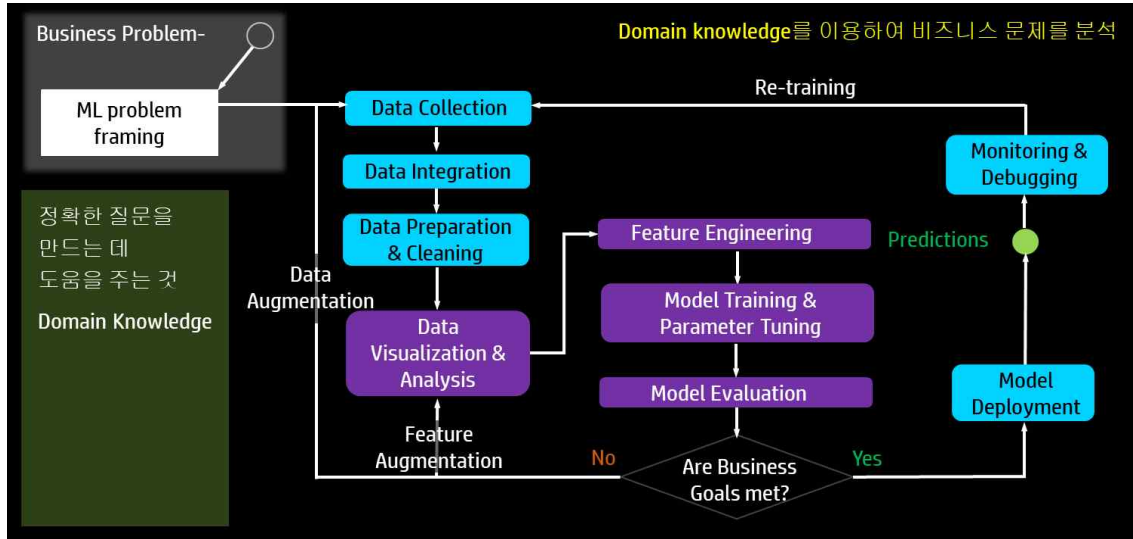
- 머신러닝 프로세스와 세이지메이커
- 세이지메이커 워크플로우
- 세이지메이커 ML 알고리즘

1. 머신러닝 프로세스와 세이지메이커

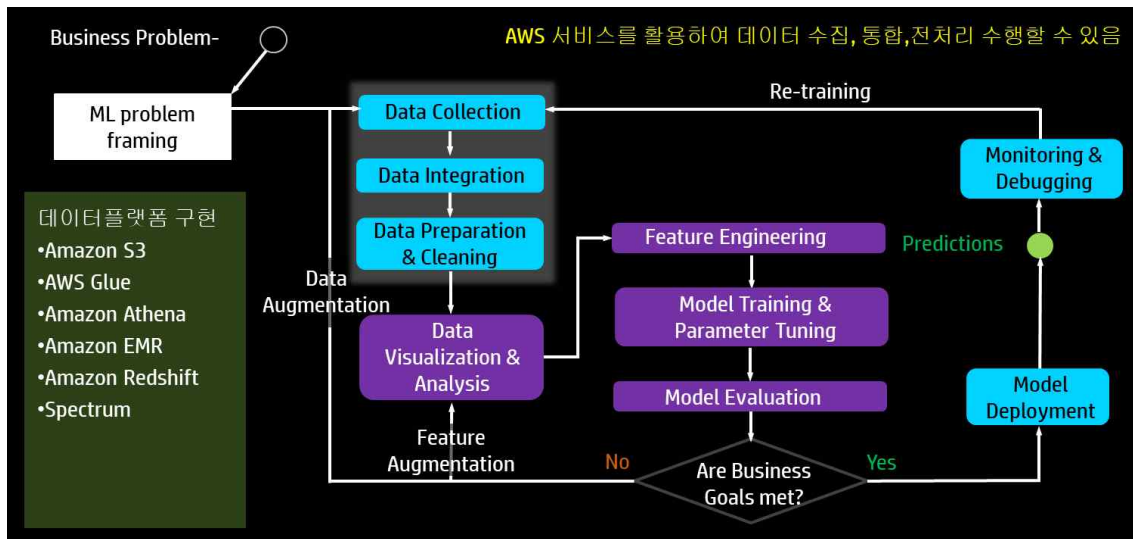
- 머신러닝 프로세스



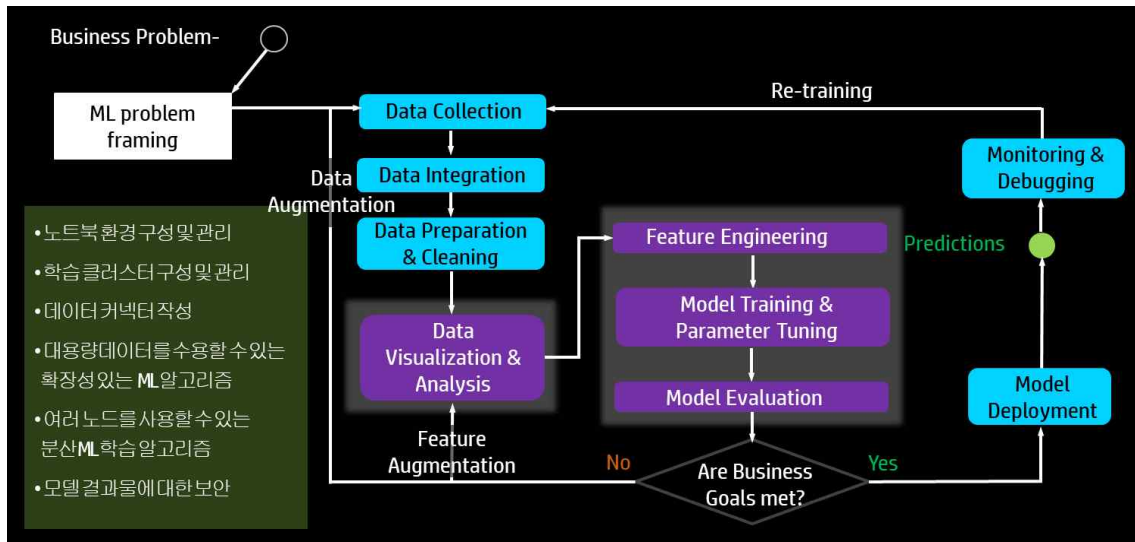
• 탐색 : 분석



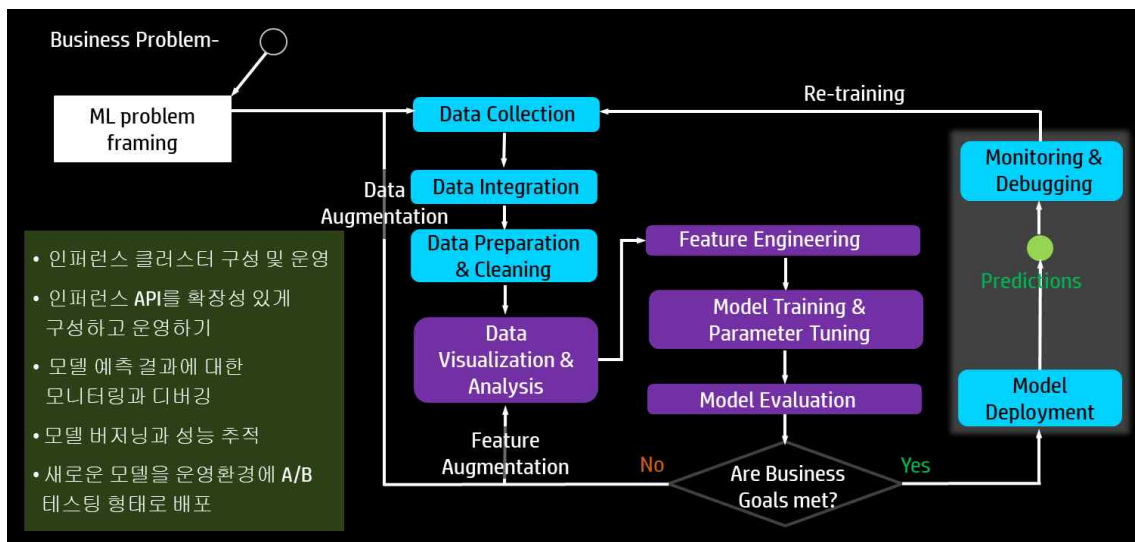
• 통합 : 데이터 아키텍처



- 왜 세이지메이커를 만들었나? : 학습환경 제공



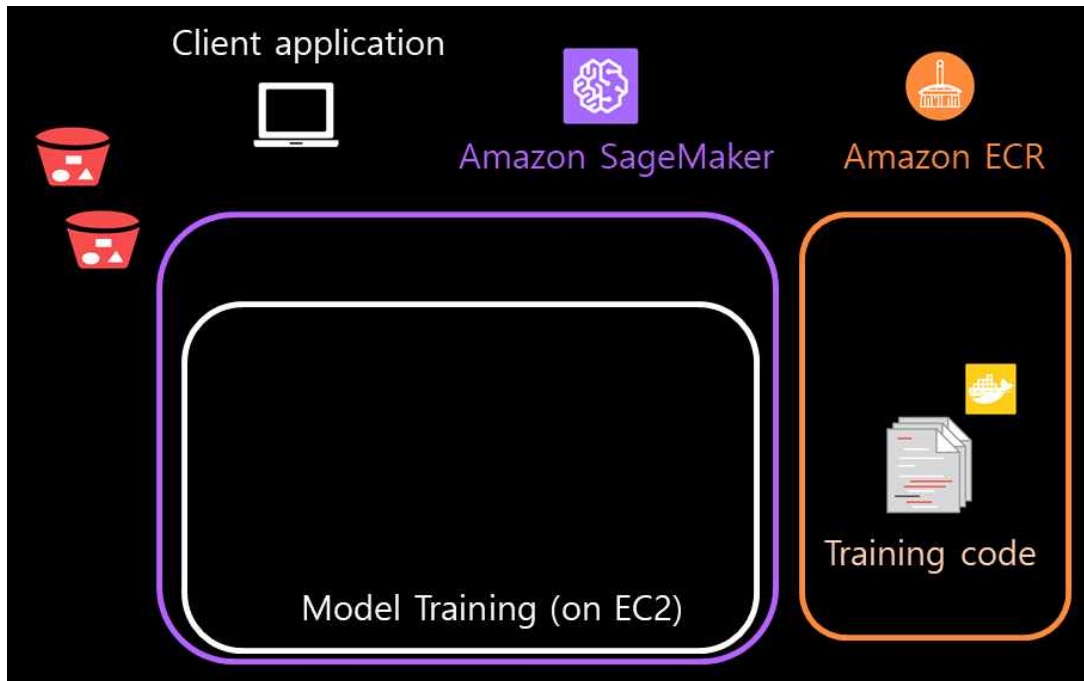
- 왜 세이지메이커를 만들었나? : 배포환경 제공



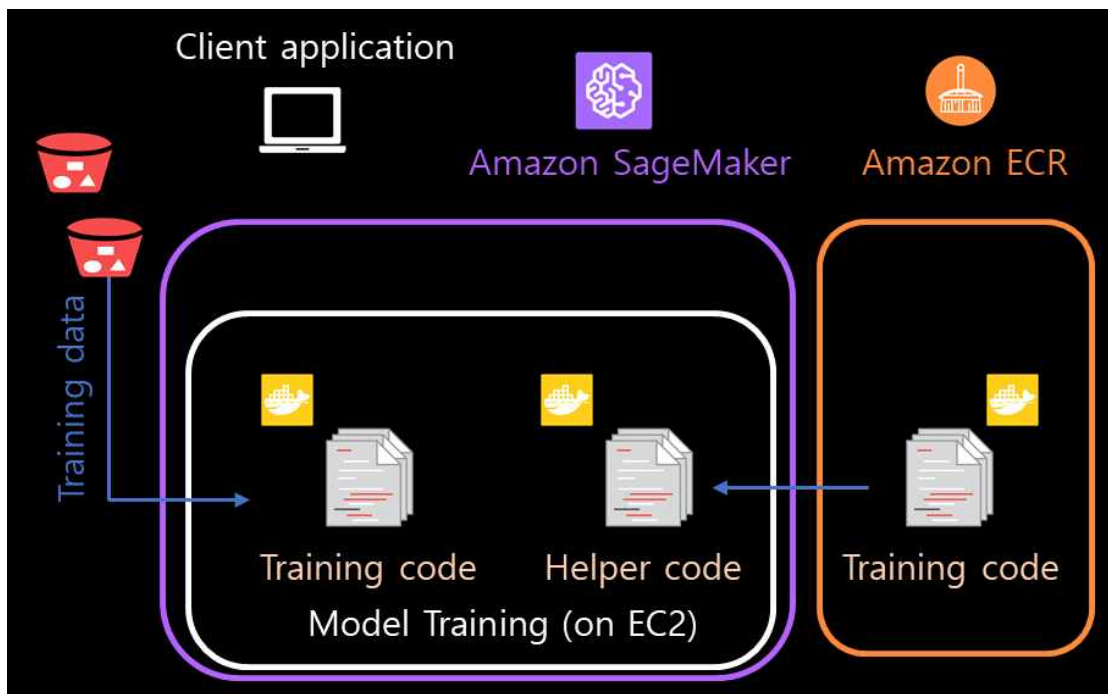
2. 세이지메이커 워크플로우

- 워크플로우 단계
 - ① 학습 코드 및 학습 데이터셋 준비
 - ② 학습환경 구성 및 학습 수행
 - ③ 학습완료된 모델 저장 및 예측 코드 준비
 - ④ 예측 환경 구성 및 모델 호스팅
 - ⑤ 예측 엔드포인트를 통한 API 서비스 제공
 - ⑥ 새로운 학습 데이터 수집 및 재학습, 배포

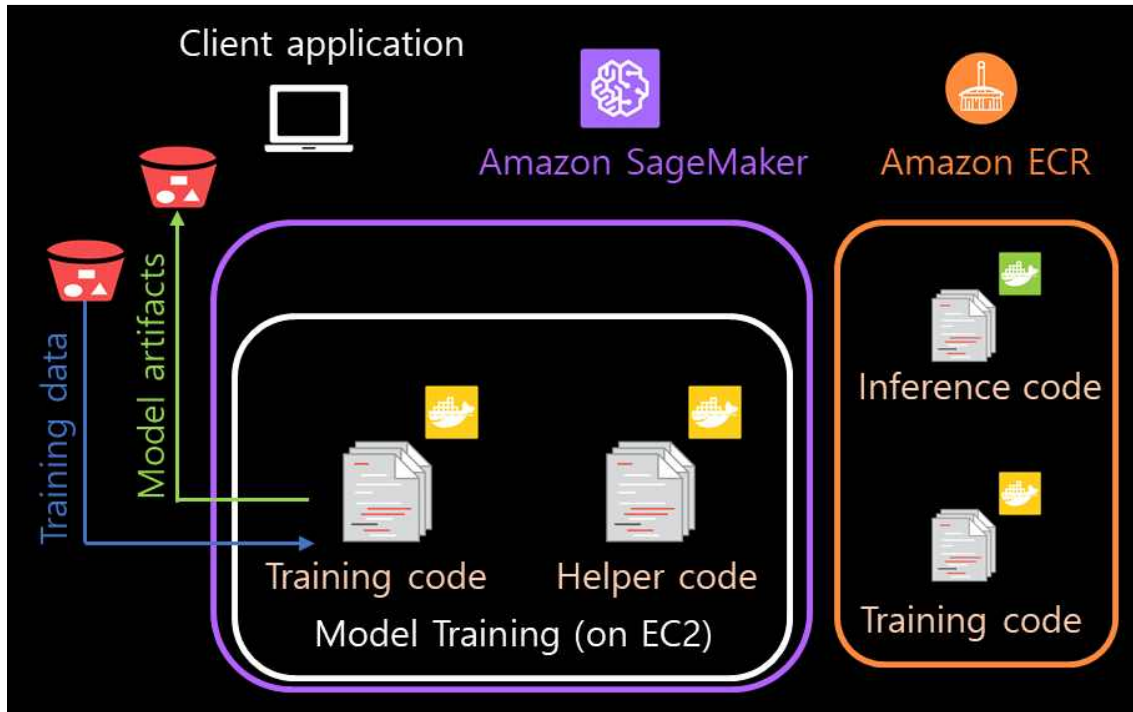
✓ 학습 코드 및 데이터셋 준비



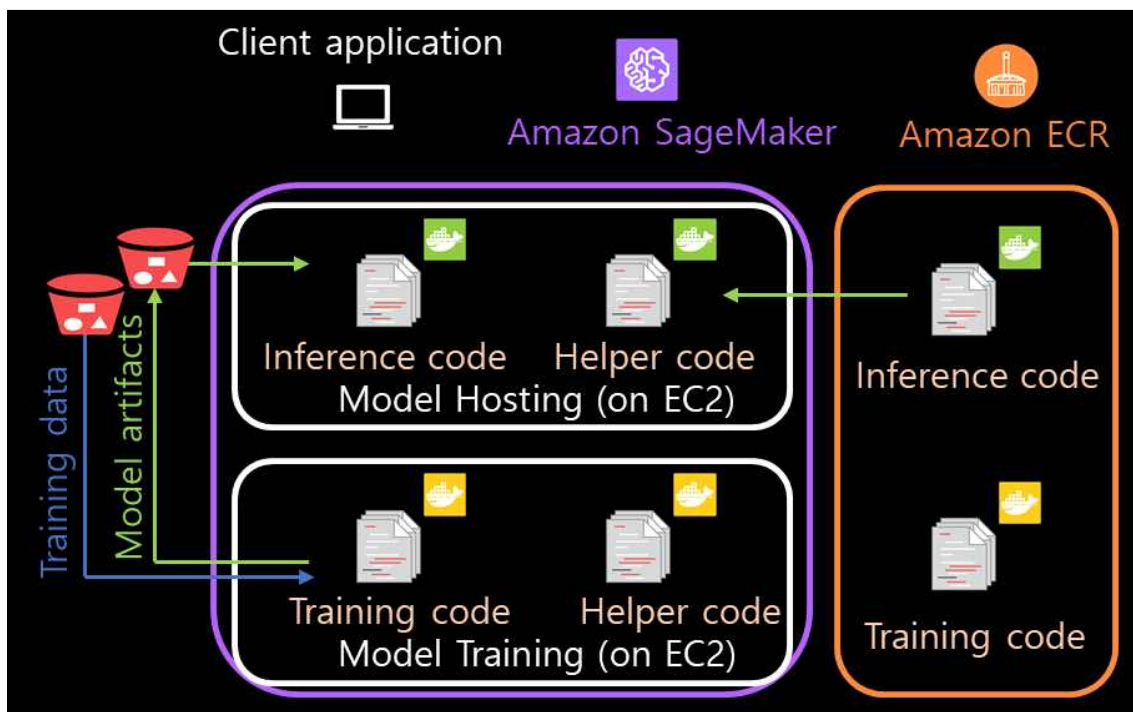
✓ 학습 환경 구성 및 학습 수행



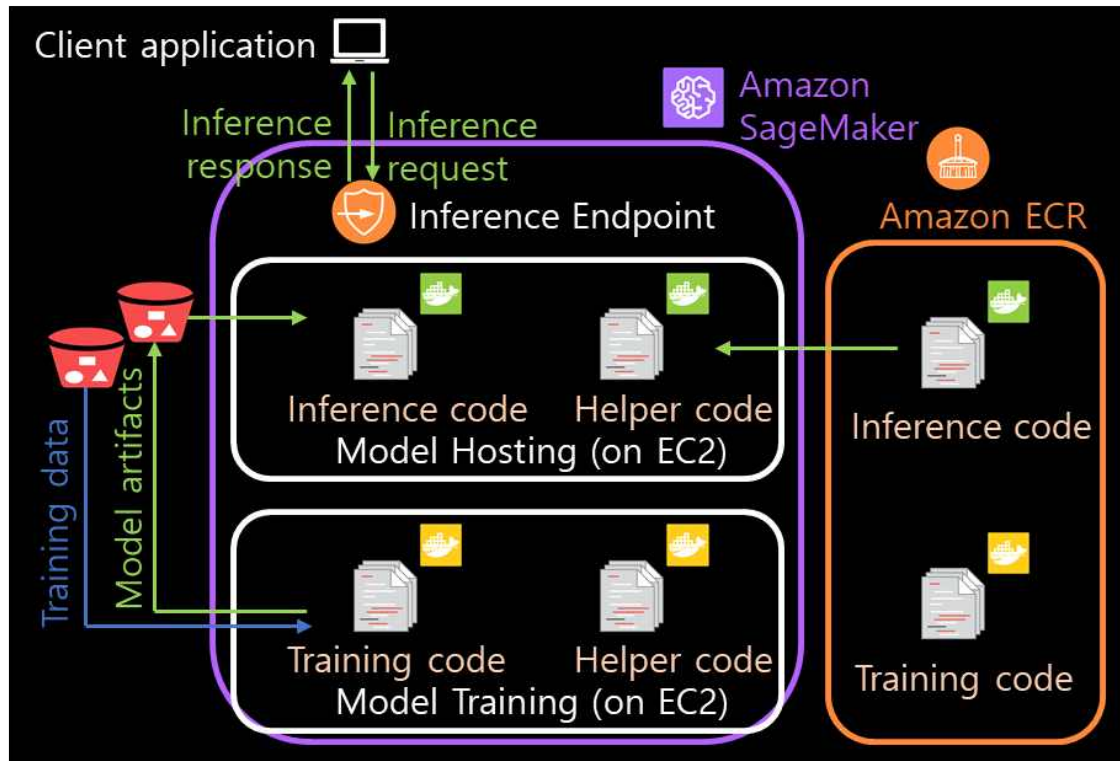
- ✓ 학습 완료된 모델 저장 및 예측 코드 준비



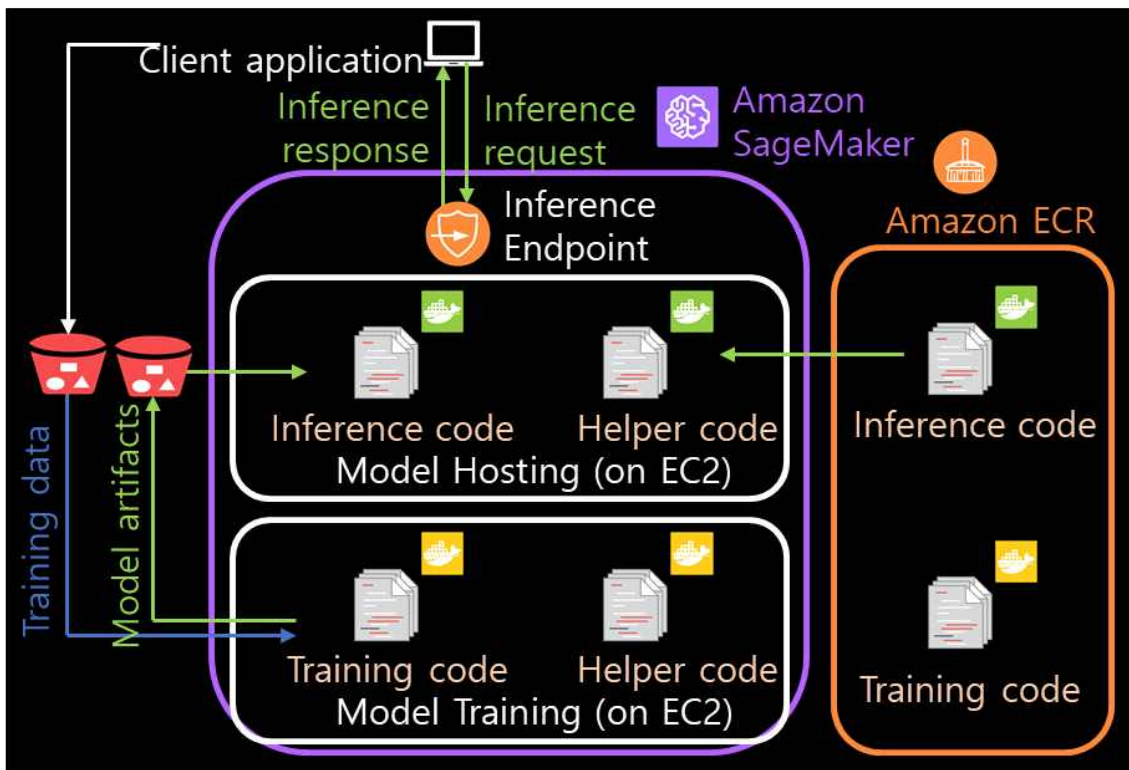
- ✓ 예측 환경 구성 및 모델 호스팅



- ✓ 예측 엔드포인트를 통한 API 서비스 제공



- ✓ 새로운 학습 데이터 수집 및 재학습, 배포





생각해보기

- 세이지메이커 사용 전과 사용 후

전

- ① AWS Deep learning AMI 선택
- ② EC2 인스턴스 생성
- ③ Jupyter 노트북 서비스 구동
- ④ SSH 터널링 설정
- ⑤ JUPYTER 노트북 접속

후

- ① Jupyter 인스턴스 생성 요청
- ② AWS 콘솔에서 Jupyter 노트북 접속

3. 세이지메이커 ML 알고리즘

- 세이지메이커 ML 알고리즘(내장)

알고리즘	설명
Blazing Text (Word2Vec)	<p>Word2vec 및 텍스트 분류 알고리즘을 최적화</p> <ul style="list-style-type: none"> • Word2vec 알고리즘은 감정 분석, 명명된 엔터티 인식, 기계 번역 등 여러 가지 다운스트림 자연 언어 처리(NLP) • 텍스트 분류는 웹 검색, 정보 검색, 순위 지정 및 문서 분류
DeepAR	반복 신경망(RNN)을 사용하여 스칼라(1차원) 시계열을 예상하는 지도 학습 알고리즘
Factorization Machines	분류 및 회귀 작업 모두에 대해 사용할 수 있는 범용 지도 학습 알고리즘
Gradient Boosted Trees(XGBoost)	<p>최적화된 분산형 경사 부스팅 라이브러리</p> <p>XGBoost는 Extreme Gradient Boosting의 약자</p>

알고리즘	설명
이미지분류 (ResNet)	이미지 분류 시스템 개발에서 인기 있는 신경망
IP Insights	악의적인 사용자를 탐지하거나 IP 주소의 사용 패턴을 학습하는 알고리즘
K-Means Clustering	레이블이 지정되지 않은 데이터 내 그룹을 찾는데 사용되는 머신러닝 알고리즘
K-NN(K-Nearest Neighbor)	분류 및 회귀 기반 문제를 해결하는 인덱스 기반 알고리즘

알고리즘	설명
LDA (Latent Dirichlet Allocation)	텍스트 파일 세트에 존재하는 기본 주제를 자동으로 발견하는데 적합한 모델
Linear Learner(분류)	객체의 특징을 사용해 객체가 속하는 적절한 그룹 식별
Linear Learner(회귀)	두 변수 사이의 선형 관계를 예측하는데 사용
NTM(Neural Topic Modeling)	텍스트 및 이미지 데이터셋에서 주제를 정하는 신경망 기반 접근 방식

알고리즘	설명
Object2Vec	가장 인접한 이웃을 계산하고 자연 클러스터를 시각화하는 신경 임베딩 알고리즘
Object Detection	이미지의 여러 개체를 탐지 및 분류하고 경계 상자를 배치
PCA(Principal Component Analysis)	흔히 데이터 사전 처리에 사용되는 알고리즘 많은 기능의 테이블 또는 매트릭스를 가져와 더 적은 수의 대표적 기능으로 줄임

알고리즘	설명
Random Cut Forest	이상 탐지를 하는 비지도형 기계학습 알고리즘
의미 세그멘테이션 (Semantic Segmentation)	이미지의 개별 픽셀에 레이블을 할당하여 관심 위치를 식별하도록 이미지에 파티션 지정
Sequence2Sequence	기계 번역, 텍스트 요약 등에서 자주 사용되는 텍스트용 범용 인코더-디코더

평가하기

1. 기존 EC2 인스턴스를 이용하여 ML 모델을 학습하는 방법보다 세이지메이커를 사용하였을 때 인스턴스 생성 과정이 복잡하다. (O/X)

- 정답 : X

해설 : 기존 EC2 인스턴스를 이용하여 ML 모델을 학습하는 방법보다 세이지메이커를 사용하였을 때 인스턴스 생성 과정이 복잡합니다.

2. 빈칸에 알맞은 단어를 고르시오.

보기 : 학습 환경, 내장 머신러닝 알고리즘, 배포 환경

① 세이지메이커는 머신러닝 과 을 제공한다.

② Amazon SageMaker는 여러 을 제공하며, 다양한 문제 유형에 대해 사용할 수 있다.

- 정답 : ① 학습 환경, 배포 환경, ② 내장 머신러닝 알고리즘

학습정리

1. 머신러닝 프로세스와 세이지메이커

- 탐색(분석) : Domain knowledge를 이용하여 비즈니스 문제를 분석
- 통합(데이터아키텍처) : 데이터 플랫폼을 구현 (Amazon S3, EMR, Athena 등)
- 학습환경 제공 : 세이지메이커에서 통합 제공
- 배포환경 제공 : 세이지메이커에서 통합 제공

2. 세이지메이커 워크플로우

- 학습 코드 및 학습 데이터세트 준비
- 학습 환경 구성 및 학습 수행
- 학습 완료된 모델 저장 및 예측 코드 준비
- 예측 환경 구성 및 모델 호스팅
- 예측 엔드포인트를 통한 API 서비스 제공
- 새로운 학습 데이터 수집 및 재학습, 배포

3. 세이지메이커 ML 알고리즘

- Blazing Text, DeepAR, Factorization Machines, Gradient Boosted Trees, 이미지분류, IP Insights, K-Means Clustering, K-NN, LDA, Linear Learner, NTM, Object2Vec, Object Detection, PCA, RCF, Semantic Segmentation, Sequence2Sequence