

6. 세이지메이커 시작하기

1강. 세이지메이커란?

학습목표

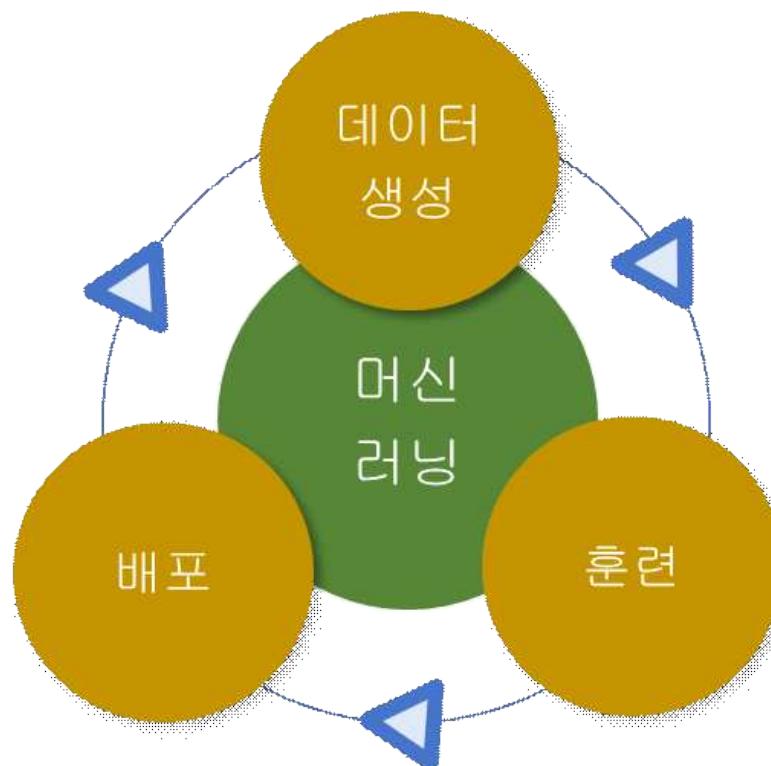
- 머신러닝 워크플로우에 대해 설명할 수 있다.
- 세이지 메이커의 개념, 장점 등에 대해서 설명할 수 있다.
- 세이지메이커의 기능에 대해서 열거할 수 있다.

학습내용

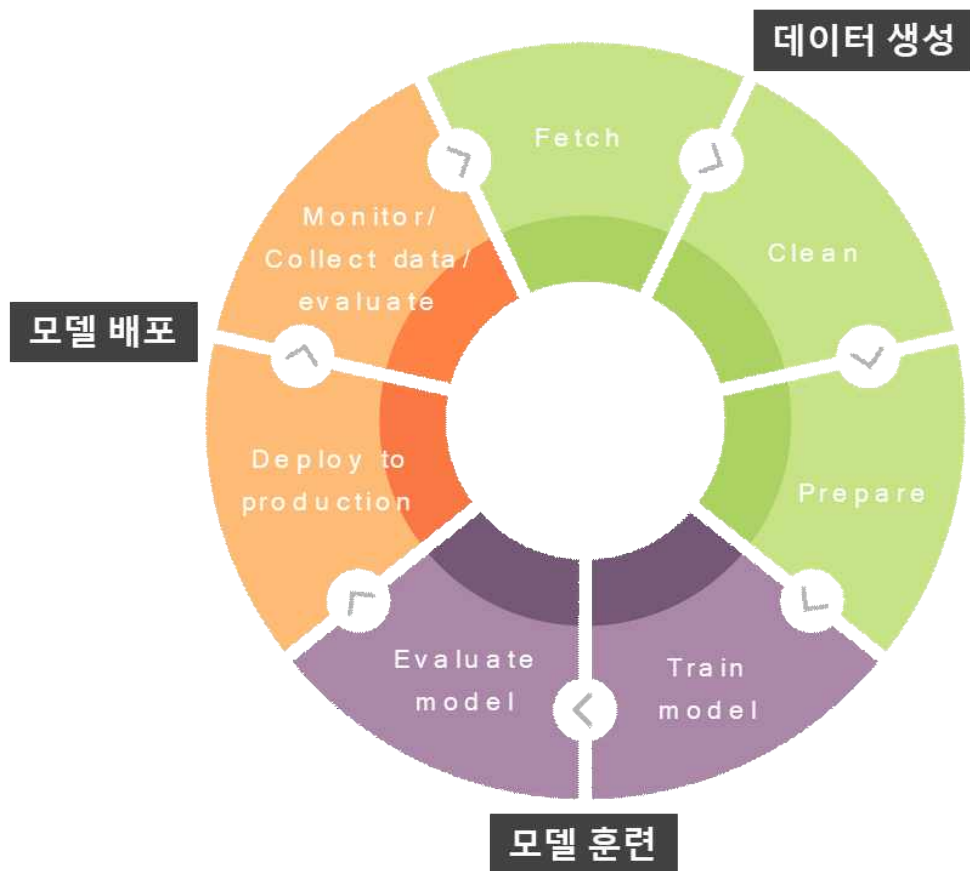
- 머신러닝 워크플로우
- 세이지메이커란 무엇인가?
- 세이지메이커의 기능

■ 세상을 잇(IT)다!

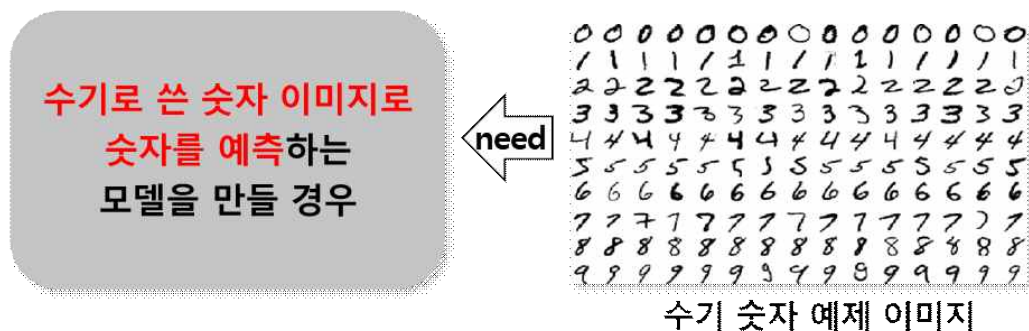
- 머신러닝 생성 워크플로우



1. 머신러닝 워크플로우



- 데이터 생성
 - ✓ 모델을 훈련시키려면 예제 데이터가 필요
 - ✓ 필요한 데이터 유형은 모델을 사용하여 해결하고자 하는 비즈니스 문제에 따라 다름

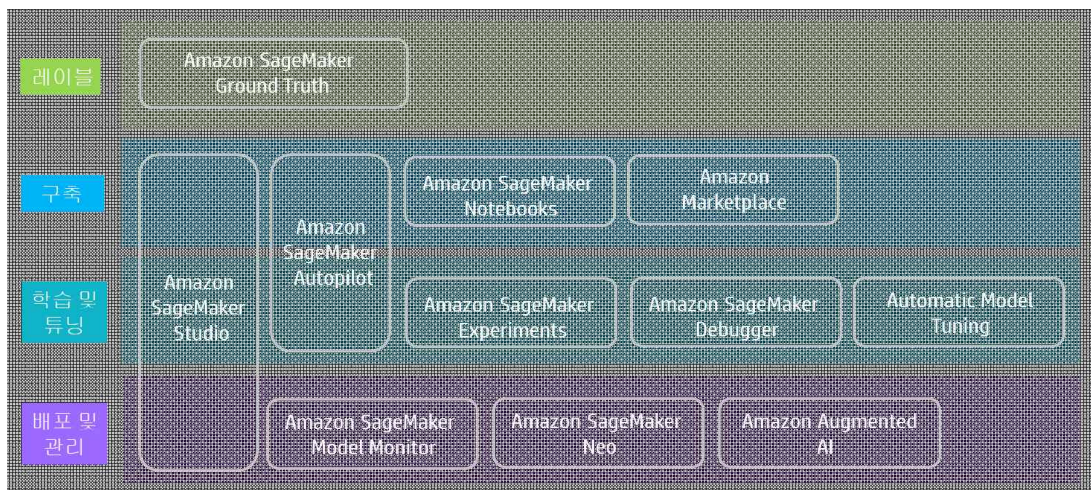


- ✓ 데이터 과학자는 모델 교육에 사용하기 전에 예제 데이터를 찾고 사전 처리하거나 이에 대해 '논쟁'하는데 많은 시간을 소비
- ✓ 데이터 사전 처리 단계

- ① 데이터 가져오기(Fetch)
 - ② 데이터 정리(Clean)
 - ③ 데이터 준비 또는 변환(Prepare)
- 모델 훈련
 - ✓ 모델 훈련에는 모델을 훈련시키고 평가하는 작업이 모두 포함
 - ✓ 모델 훈련(Train Model) : 모델을 훈련시키기 위해서는 알고리즘이 필요하며, 선택하는 알고리즘은 요인의 수에 따라 다름
 - ✓ 모델평가(Evaluate Model) : 모델을 훈련한 후 모델을 평가하여 추론의 정확도가 허용 가능한지 여부를 결정
 - 모델 배포
 - ✓ 모델을 애플리케이션과 통합하고 배포하기 전에 모델을 리엔지니어링(하이퍼파라미터, 히든레이어 등)
 - ✓ 배포, 모니터링, 평가, 데이터 수집

2. 세이지메이커란 무엇인가?

- 세이지메이커의 정의
 - ✓ Amazon SageMaker는 완전 관리형 머신러닝 서비스
 - ✓ 머신러닝 모델을 빠르게 구축, 훈련, 최적화, 배포할 수 있도록 서비스 내에서 프로세스화

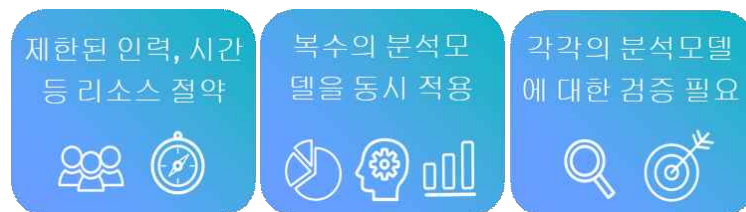


[출처] AWS(<https://aws.amazon.com/ko/sagemaker>)

- 세이지메이커의 연혁
 - ✓ 2017년
 - AWS re:Invent 컨퍼런스에서 SageMaker 시작
 - ✓ 2018년도 전반기
 - 2018SageMaker에서 Managed TensorFlow 및 MxNet 심층 신경망 학습 및 추론 지원
 - 반복신경망 학습, word2vec 학습, 다중클래스 선형 학습, 분산 심층 신경망 학습 지원
 - ✓ 2018년도 하반기
 - 데이터에 라벨링 작업을 쉽게 하도록 SageMaker Ground Truth를 지원
 - 개발자와 데이터 과학자가 강화 학습 모델을 빠르고 쉽게 개발할 수 있도록 SageMaker RL(강화학습) 지원
 - 스마트폰 및 스마트 카메라와 같은 에지 장치에 심층 신경망 모델을 배포할 수 있도록 SageMaker Neo 지원
 - SageMaker Marketplace를 통해 SageMaker에서 학습 및 배포할 수 있는 머신러닝 모델을 사고 팔 수 있음
 - ✓ 2019년
 - SageMaker Neo가 오픈 소스 소프트웨어로 출시

생각해보기

- Mycelebs - 세이지메이커 도입 배경
 - ✓ 마이셀렙스는 자체 인공지능 솔루션인 '빅데이터 스튜디오'를 기반으로 비즈니스에 즉시 적용 가능한 서비스를 제작 운영하는 AI 어플리케이션 그룹



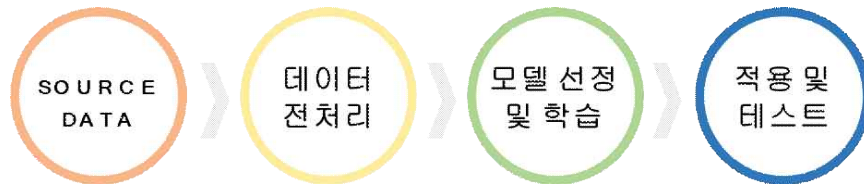
세 가지 목표를 동시에 달성해야 하는 상황

 BIGDATA STUDIO

 Amazon SageMaker

[출처] AWS

- Mycelebs - 세이지메이커 활용 전(자체 진행)
 - ✓ 마이셀랩스는 자체 인공지능 솔루션인 ‘빅데이터 스튜디오’를 기반으로 비즈니스에 즉시 적용 가능한 서비스를 제작 운영하는 AI 어플리케이션 그룹



- Mycelebs - 세이지메이커 활용 후



- ✓ Lambda와 SageMaker를 활용하여 분석 및 결과 확인 과정을 자동화
- ✓ SageMaker에서 제공되지 않는 모델은 Docker를 활용하여 처리 가능했음
- ✓ 분석결과 및 검증 등은 자체 솔루션을 활용하여 수행함

- Mycelebs - 도입효과

- ✓ 별도의 시스템 구축 및 관리 없이 바로 사용 가능
- ✓ SageMaker에서 기본제공중인 모델 성능이 뛰어남
- ✓ 기존에 도입/활용 중인 AWS 제품들과 연동이 용이함
- ✓ 분석 프로세스를 포함한 내부 통합 솔루션 개발 기간 단축
- ✓ 여러 프로젝트에서 각각 필요한 만큼의 자원을 활용 가능



- ✓ 전체적인 분석 프로세스의 효율성 개선에 기여
- ✓ 내부 업무 프로세스 기준으로 소요시간 5배 가량 단축

[출처] AWS

- 세이지메이커의 장점

- ✓ 구축

- 사전에 빌드된 노트북 인스턴스
 - 고도로 최적화된 머신러닝 알고리즘 제공

※ 노트북 인스턴스 : Jupyter 노트북 앱을 실행하는종합 관리형 ML 컴퓨팅 인스턴스

- ✓ 학습

- 한번 클릭으로 ML, DL, 커스텀 알고리즘 학습
 - 하이퍼파라미터 최적화를 통한 손쉬운 학습
 - 어떤 프레임워크이던지 모두 수행 가능

- ✓ 세이지메이커 지원 프레임 워크



- ✓ 배포

- ✓ 엔지니어링 노력(서버 생성, 웹 서비스 구축, 엔드포인트 생성 등)이 필요 없이 배포 가능
 - ✓ 확장성이 있는 완전 관리형 모델 호스팅

3. 세이지메이커의 기능



- Amazon SageMaker Notebook Instance
 - Jupyter 노트북 앱을 실행하는 종합 관리형 ML 컴퓨팅 인스턴스
 - Amazon SageMaker 콘솔 및 Amazon SageMaker API를 사용

```
def convert_data():
    data_partitions = [('train', train_set), ('validation', valid_set), ('test', test_set)]
    for data_partition_name, data_partition in data_partitions:
        print('{}: {}'.format(data_partition_name, data_partition[0].shape, data_partition[1].shape))
        labels = [t.tolist() for t in data_partition[1]]
        features = [t.tolist() for t in data_partition[0]]

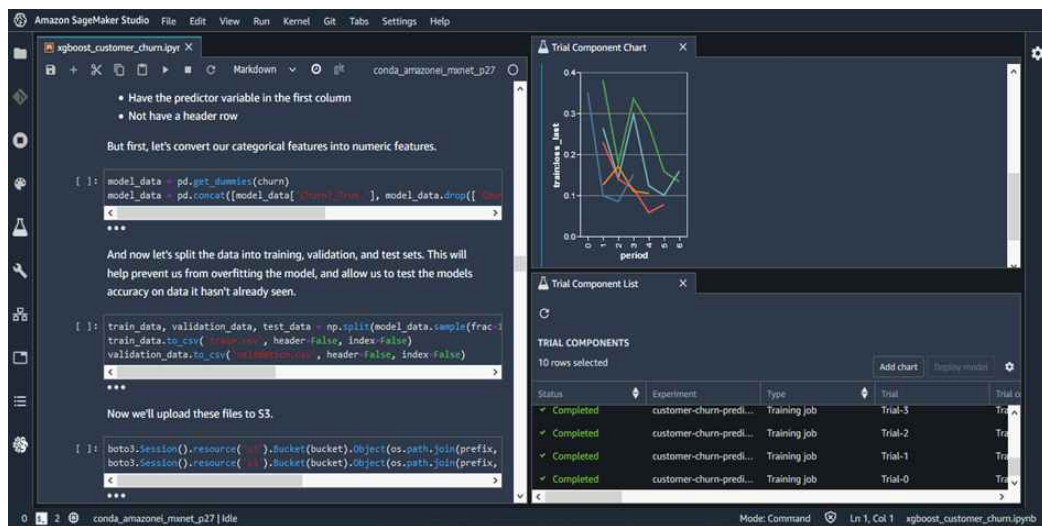
        if data_partition_name != 'test':
            examples = np.insert(features, 0, labels, axis=1)
        else:
            examples = features
        #print(examples[50000, :])

        np.savetxt('data.csv', examples, delimiter=',')

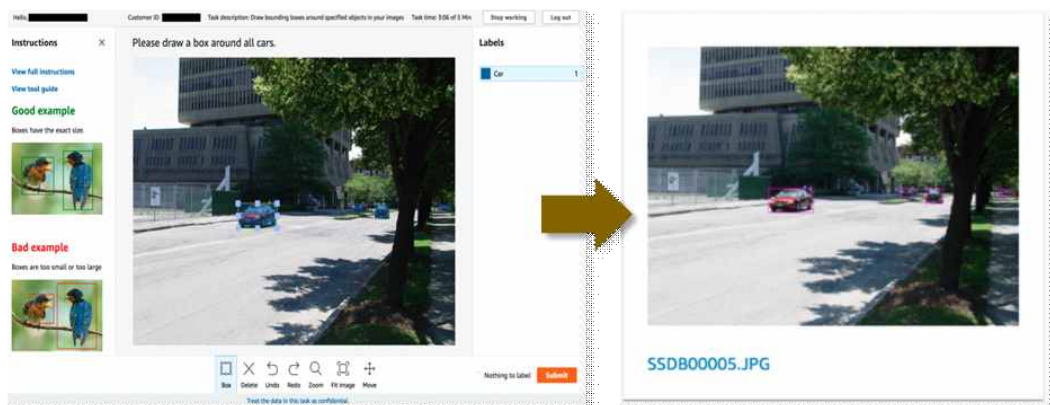
        key = "{}/{}/examples".format(prefix, data_partition_name)
        url = 's3://{}{}'.format(bucket, key)
        boto3.Session().resource('s3').Bucket(bucket).Object(key).upload_file('data.csv')
        print('Done writing to {}'.format(url))

convert_data()
```

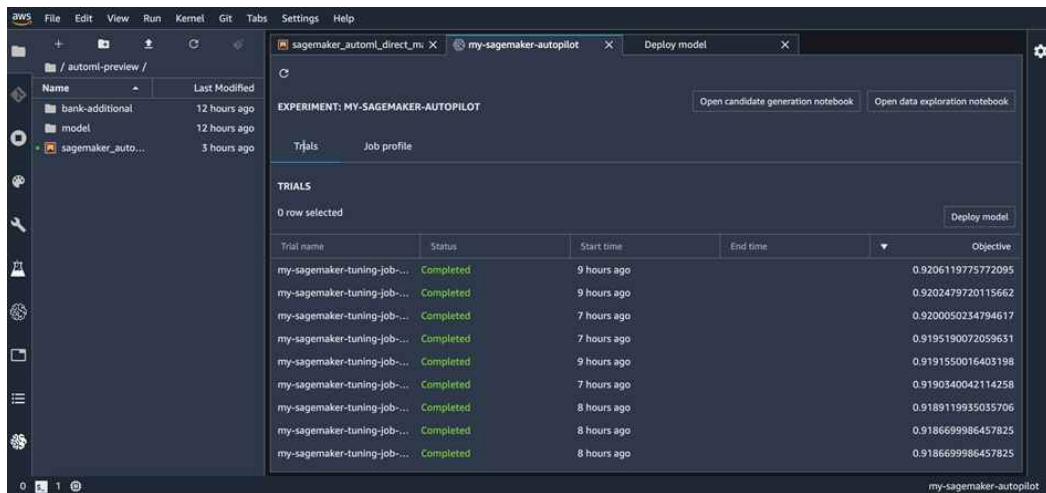
- Amazon SageMaker Studio
 - 머신러닝 모델을 쉽게 빌드, 훈련, 디버깅, 배포 및 모니터링 할 수 있게 해주는 머신러닝을 위한 웹 기반의 통합 개발 환경(IDE)



- Amazon SageMaker Ground Truth
 - 학습 데이터셋을 만들기 위한 레이블링 작업에 대한 워크플로우와 인터페이스 제공



- Amazon SageMaker Autopilot
 - 머신러닝 프로세스를 자동화하여 머신러닝 환경을 단순화
 - 인프라, 알고리즘, 데이터 전처리, 모델 튜닝이 자동 처리

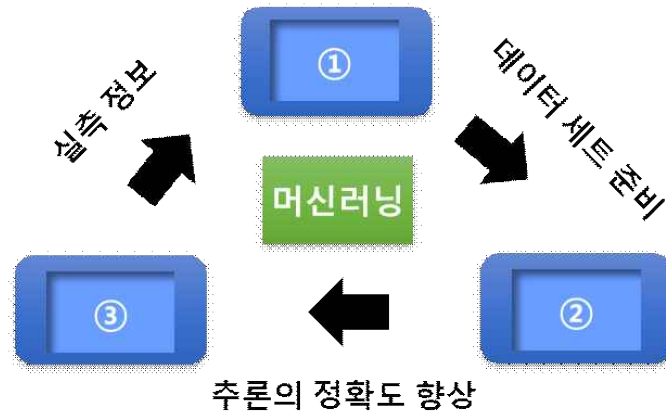


- 그 외 Amazon SageMaker 기능

SageMaker Neo	<ul style="list-style-type: none"> • 한 번 훈련시킨 머신러닝 모델을 클라우드와 엣지 어디서나 실행할 수 있음
Amazon SageMaker Model Monitor	<ul style="list-style-type: none"> • 프로덕션 환경에서 Amazon SageMaker 머신러닝 모델의 품질을 지속적으로 모니터링
Amazon SageMaker Experiments	<ul style="list-style-type: none"> • 머신러닝 실험을 구성, 추적, 비교 및 평가할 수 있음
Amazon SageMaker Debugger	<ul style="list-style-type: none"> • 훈련 프로세서 전반에 걸쳐 훈련 파라미터 및 데이터를 검사 • 파라미터 값이 너무 크거나 작아지는 등 일반적으로 발생하는 오류를 자동으로 감지하여 사용자에게 알림
AWS Marketplace	<ul style="list-style-type: none"> • 많은 알고리즘과 사전에 학습된 모델을 제공
Amazon Augmented AI	<ul style="list-style-type: none"> • ML 예측의 인적 검토에 필요한 워크플로를 쉽게 구축할 수 있는 서비스
Automatic Model Tuning	<ul style="list-style-type: none"> • 한번의 클릭으로 하이퍼파라미터 최적화

평가하기

1. 머신러닝 워크플로우의 단계를 보기를 보고 나열해 보시오.



보기 : 모델 훈련, 모델 배포, 데이터 생성(수집)

- 정답 : ① 데이터 생성(수집), ② 모델 훈련, ③ 모델 배포

2. 세이지메이커에 대한 설명으로 옳지 않은 것은?

- ① 완전 관리형 머신러닝 서비스이다.
- ② 최적화된 머신러닝 알고리즘을 제공한다.
- ③ 하이퍼파라미터를 최적화할 수 있다.
- ④ 텐서플로우 프레임워크만 지원한다.

- 정답 : ④ 번

해설 : 세이지메이커는 다양한 프레임워크(텐서플로우, mxnet, PyTorch, Keras 등)을 지원합니다

학습정리

1. 머신러닝 워크플로우

- 데이터 생성(수집) → 모델 훈련 → 모델 배포 (순환)

2. 세이지메이커란 무엇인가?

- 완전 관리형 머신러닝 서비스 : 머신러닝 모델을 빠르게 구축, 훈련, 최적화, 배포할 수 있도록 서비스 내에서 프로세스화
- 2017년에 세이지 메이커 시작, 그 이후, 지속적으로 발전
- 세이지 메이커 장점
 - 구축 : 최적화된 머신러닝 알고리즘 제공
 - 학습 : 원클릭으로 ML/ DL/커스텀 알고리즘 학습, 하이퍼파라미터 최적화를 통한 손쉬운 학습, 다양한 프레임워크 지원
 - 배포 : 엔지니어링 노력없이 배포 가능, 확장성이 있는 완전 관리형 모델 호스팅

3. 세이지메이커의 기능

- Amazon SageMaker Notebook Instance : 종합 관리형 ML 컴퓨팅 인스턴스
- Amazon SageMaker Studio : 머신러닝을 위한 웹 기반의 IDE
- Amazon SageMaker Ground Truth : 레이블링 작업에 대한 워크플로우와 인터페이스 제공
- Amazon SageMaker Autopilot : 인프라, 알고리즘, 데이터전처리, 모델 튜닝 자동 처리
- 기타 : SageMaker Neo, Amazon SageMaker Model Monitor, Amazon SageMaker Experiments, Amazon SageMaker Debugger, AWS Marketplace, Amazon Augmented AI, Automatic Model Tuning