AWS AL/ML

Conference 리포트

Commax 인턴

오 수찬

목차

서론 1

연구분야 1.1

동기 1.2

목적 1.3

본론: 아마존의 AI/ML 서비스 구조 및 활용 2

구조 분석 2.1

활용 방식 2.2

결론 3

참고 자료 4

1. 서론

1.1 연구 분야

* AWS에서 제공하는. AI(Artificial Intelligence)/ML(Machine Learning) 서비스의 구조 및 활용법.

1.2 동기

* AI/ML 기술의 시장 관심 증가 및 수요가 확대됨. 그러한 AI/ML 기술을 활용한 개인화 서비스의 선두주자인 아마존의 AI/ML 관련 서비스를 알아보고자 함.

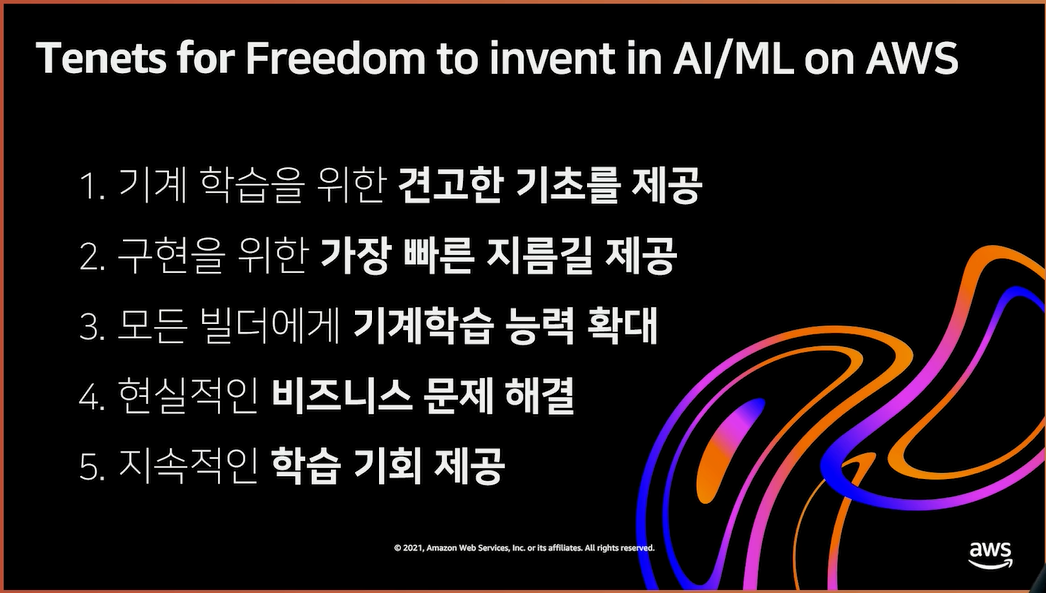
1.3 목적

* AI/ML 기술을 활용한 서비스의 종류와 구조를 파악하는 것.
* AI/ML 기술을 효율적으로 사용할 수 있는 분야와 능동적인 사용방안을 파악하는 것.

2. 관련연구

아마존은 클라우드 서비스, 온라인 마켓, O2O등 다양한 분야의 선두주자이다. 뿐만 아니라 AI/ML 관련 분야에서도 그 두각을 나타내고 있다. 그 대표 사례가 AWS에서 제공하는 AI/ML 서비스이다.

AWS는 아래와 같은 5가지 사상적 기조 하에 AI/ML서비스를 제공한다.



위의 5가지 기조를 풀어 설명하면 아래와 같다.

1. ML을 위한 견고한 기초를 제공.

쉽게 말해서 인스턴스와 같은 ML을 위한 견고한 인프라를 제공하고자 한다.

이를 위해 현재 아마존이 제공하는 대표적인 인스턴스는 pd4다. 이는 고성능의 기계학습 인스턴스로써 GPU를 최대 4000개까지 확장이 가능하며 기존 제공되던 인스턴스인 p3대비 최대 3배 빠른 속도를 자랑한다.

훌륭한 인스턴스를 제공하는 데서 그치지 않는다. 최근에는 ML에도 프레임 워크를 사용하는 것이 추세인데. 이러한 추세에 따라 AWS에서 제공하는 인스턴스는 프레임워크를 선택하는데 있어서 자유를 준다. 따라서 사용자는 TensorFlow, Pytorch등 다양한 프레임 워크를 입맛에 맞게 적용할 수 있다.

1. 구현을 위한 가장 빠른 길 제공.

아마존은 이를 위해 End to End로 이루어진 완전처리형 학습 서비스를 제공하고자 한다. 풀어 말하면 데이터 레이블링, 모델 학습, 평가 및 디버깅, 재 학습, 최적화가 한번의 클릭으로 해결되는 서비스를 제공하고자 한다.

이를 위해 AWS는 자동으로 스케일링 되는 서비스를 제공하며 이 서비스를 이용할 통합 IDE환경 및 콘솔을 제공한다. 그 결과 사용자는 IDE 혹은 웹을 통해 단순한 클릭으로 복잡한 모델이나 인프라 또는 ML에 관련된 복잡한 지식 없이 ML서비스를 배포 및 모니터링 할 수 있다. 더하여 IDE 및 콘솔 환경은 DevOps 모델을 추구하여 배포까지의 일련의 과정에서 ML전문가와 클라우드 개발자 사이의 쉬운 협업을 지원하고자 한다.

1. 모든 빌더에게 ML 능력 확대

AWS에서 제공하는 AI/ML 서비스는 일반 개발자, 데이터분석가 뿐만 아니라 현업 전반에 ML 기술을 확대 적용가능 하게 하고자 한다.

기존 ML관련 기술을 현업에 적용하려면 ML사이언티스트와 현업 전문가 사이의 협업 필요했다. 이러한 방식을 개선하고자 SQL만 던지면 학습된 데이터 추출해서 결과를 보완 Aurora ML이라는 서비스를 제공한다.

1. 현실적인 비즈니스 문제 해결

다양한 요구에 맞춤 학습모델 제공. 개인화에도 많은 경험으로 서비스 제공하고자 한다.

관련 AWS서비스로는 비즈니스 데이터 이상탐지, 인프라 운영 데이터 이상탐지 등 다양한 형태의 이상탐지 서비스가 있다. 대표적인 예로 센서를 통해 기계 이상을 탐지(한국 제조기업 참여)하는 서비스인 모니트론이 있다. 센서 + 데이터 전송장치로 구성된 이 서비스는. 센서를 설비에 붙이면 앱으로 설비의 장애상황을 모니터링 할 수 있다.

클라우드를 위한 DevOps 관련 기능 들도 다수 존재한다. DevOps Guru와 같은 서비스들이 대표적이다.

Lookout for vision 또한 비전기술 이용해서 문제상황을 탐지하는 서비스로 Panorama Appliance를 Lookout for vision과 병합적으로 활용하면 여러 이상들을 탐지해 내거나 제품의 불량 등을 확인할 수 있다.

5.) 지속적인 학습 기회 제공

ML과 관련된 지속적인 학습 자료와 정보를 제공함으로써 현업 전반에 ML에 대한 관심과 이해를 넓히고자 한다.

이를 위하여 DeepRacer와 같은 강화학습용 툴을 제공한다. 이는 자율 레이싱 학습도구로써 AI/ML기술을 이용한 자율 주행과 같은 분야에 지식을 넓힐 수 있다.

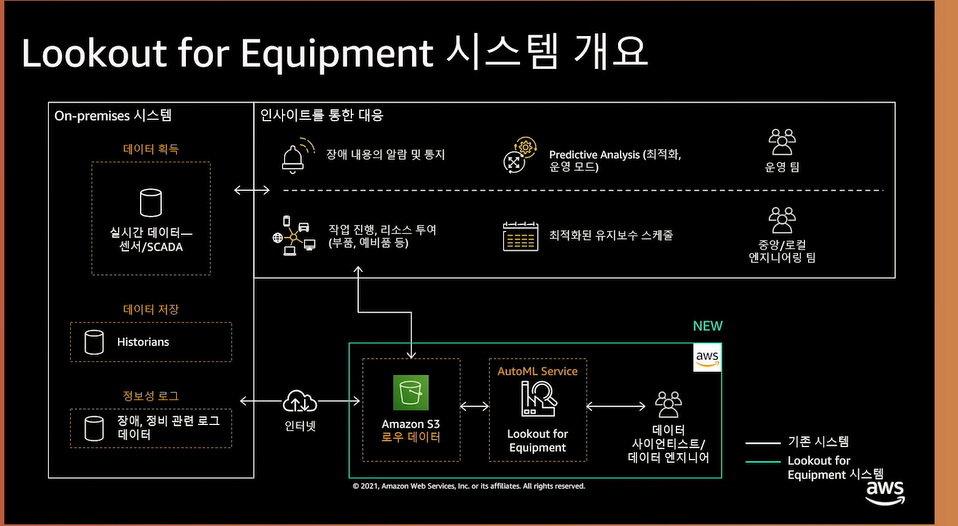
뿐만 아니라 손쉽게 ML을 배울 수 있는 툴과 매뉴얼 및 교육 등이 AWS 웹사이트에 존재한다. 위와 같은 온라인 교육을 수료하면 취득할 수 있는 자격증 제도 또한 존재한다.

3. 본론: 아마존의 AI/ML 서비스 구조 및 활용

**2.1 구조 분석**

**2.1.1 Amazon Lookout for Equipment 및 Amazon Lookout for Vision**

앞서 살펴볼 Amazon Lookout for Equipment란 설비의 장애상태나 유지보수 스케쥴을 ML기술을 활용하여 탐지해주고 모니터링 할 수 있게 지원해주는 서비스이다.

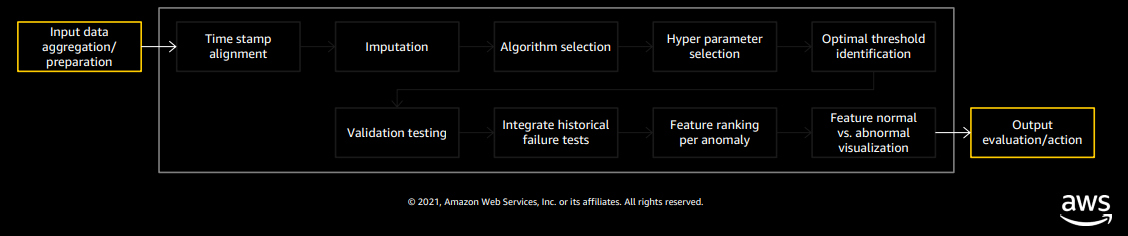
****

Amazon Lookout for Equipment의 구조를 살펴보자. 위의 사진은 Lookout for Equipment 동작 방식을 도식으로 표현한 것이다.

풀어 설명하면 다음과 같다.

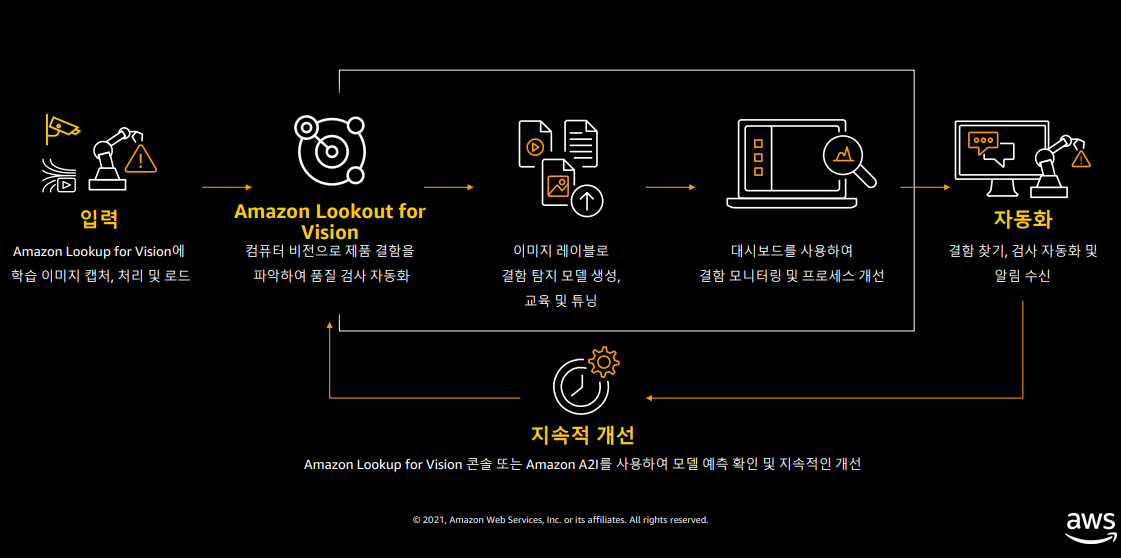
* 실시간 센서데이터를 인터넷을 통해 AWS 서비스로 전송하고 이 이미지들을 AWS S3버킷에 저장한다.
* Lookout for Equipment는 이 이미지들을 이용하여 학습한다.
* 학습을 통해 만들어진 모델을 현업에 적용하여 유지보수, 운영 및 장애상황을 탐지한다.

모델 학습부분을 더 세분화하여 살펴보자.



기존 ML에서 활용하던 학습과정과 크게 다르지 않다. 하지만 사용자는 정제된 데이터만 입력하면 결과가 나오는 것처럼 느낄 수 있도록 학습 일련의 과정을 자동화한 것을 볼 수 있다.

기존 ML에서 전문가가 필요했던 알고리즘 선택 부분이나 결과에 큰 영향을 주는 매개변수 선정 그리고 테스트와 재학습의 과정을 자동화함으로써 사용자가 접근하기 쉬워졌다. 뿐만 아니라 ML에 필요한 시간과 비용 또한 절약될 수 있다.

Amazon Lookout for Vision의 동작과정이다.

설치된 카메라를 통해 전달된 이미지를 S3버킷으로 전송. 이 이미지를 통해 Lookout for Vision이 학습한다. 학습을 통해 만들어낸 모델을 활용하여 공정과정에 불량이나 장애를 파악하고 탐지 실패사례를 통한 재 학습으로 성능이 자체적으로 개선되는 일련의 과정을 반복한다

학습 과정에서는 API에 적당한 커스텀 모델을 사용자가 직접 제공할 수도 있다 또한 알고리즘도 직접 선택하거나 제공할 수 있다. 개인화된 모델을 만들어서 사용하고 싶다면 아래와 같은 방법을 제공한다.

1. AWS SageMaker를 이용하여 직접 모델을 생성하는 방식.

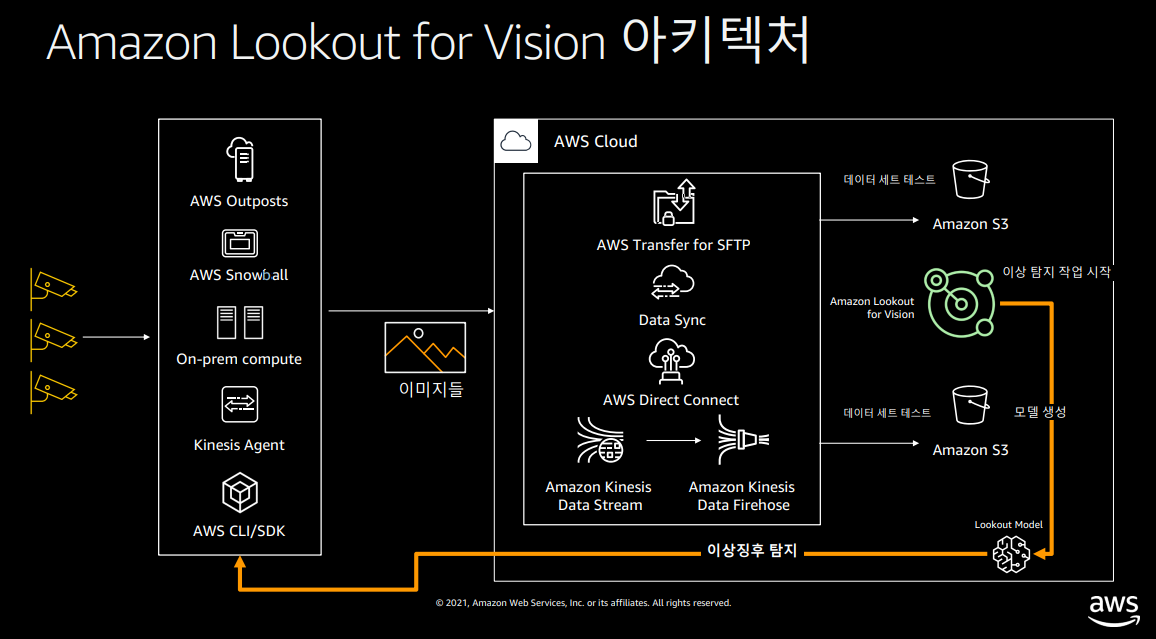
이 방식을 위해서는 사내 ML담당조직 필요하다 SageMaker는 위의 원 클릭 방식의 ML 서비스와 달리 ML과정 전반을 관리하고 감시할 필요하 있기 때문이다. 하지만 본인 서비스에 가장 최적화된 모델을 생성하고 싶다면 이 방법이 권장된다.

1. Marketplace에서 비전모델을 구매.

Marketplace에는 다른 사용자들이 구성해둔 여러 모델들이 존재한다. 이러한 모델 중에 본인이 제공하고자 하는 서비스에 가장 적합한 모델을 구매하여 사용할 수 있다.

1. Custom Labels 이용.

이미지에 레이블링을 하고 AutoML을 이용에 학습. 여러 모델을 사용해서 동시에 학습. Pre Train 데이터 이용해서 모델 선택 후 그 모델로 서비스를 제공하는 방법이다. 쉽게 말하면 본인이 거의 모든 과정을 커스텀 하는 방식이다.

위는 이미지를 전송하고 받아서 저장하고 활용되는 부분이 조금 더 자세하게 표현된 도식이다.

Lookout for Vision이 대표적으로 제공되는 기능 5개는 다음과 같다.

1) 제조라인 통합: API식으로 기존 제조라인에 통합가능.

2) 대시보드 뷰: 대시보드를 통해 생산라인, 프로세스별로 식별가능

3) 레이블링 간략화: ML과정중 데이터 레이블링 작업이 단순 해진다.

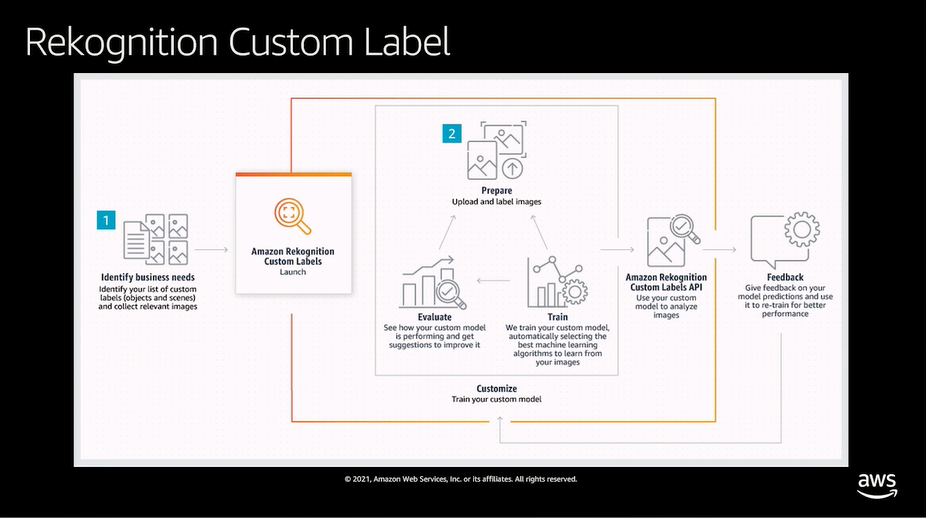
4) 빠른 평가: 테스트 데이터셋을 자동 생성하여 빠르게 평가 가능하다.

5) 테스트 감지 및 피드백: 사용자에게 피드백을 제공하여 결함 탐지 모델 개선에 긍정적인 영향을 줄 수 있다.

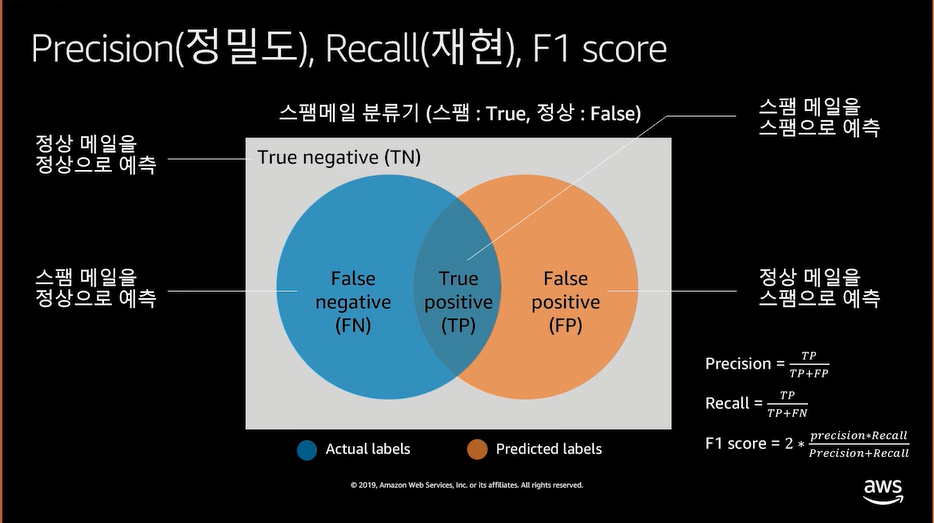
**2.1.2 Amazon Rekognition**

ML 전문지식 없이도 딥러닝 기술을 사용하여 이미지 및 비디오 분석을 추가할 수 있도록 하기위해 제공되는 서비스다. API 호출 방식으로 ML 모델을 활용할 수 있다는 것이 장점이다.

기본적으로 제공되는 모델은 이미지 분석부분은 사물, 장면, 액티비티, 얼굴탐지 분석, 유명인사 인식, 폭력적이거나 성적인 컨텐츠 탐지가 있고 비디오 분석부분에는 인물 추적, 영상 통계 등의 모델을 제공한다.

위가 Rekgnition의 동작 과정이다. 사용자의 관점에서는 원하는 API를 선택하고 레이블링 된 이미지들을 전송하면 Rekognition가 분석, 최적의 알고리즘을 결정한다. 이후 그 알고리즘으로 학습하여 결과를 도출하고 모델을 API 형태로 반환한다.

Rekgnition의 학습 과정을 조금 더 자세히 살펴보자. 사용자는 이미지에 레이블링을 하고 그 이미지들을 전송한다. Rekognition은 이런 이미지들을 AutoML을 통해서 학습한다. 이 때 학습의 과정은 두가지 동작으로 나뉜다. 첫째 선택한 API에 권장되는 여러 모델을 모두 뽑아내어 일부 데이터만 이용하여 뽑아낸 모델들의 성능을 평가한다.

모델 평가 부분에 있어서는 모델마다 False negative(FN)이 낮은 것이 중요할 수 있고 FP가 낮은 것이 중요할 수 있다. 이러한 평가 기준이 위의 사진 우측 하단에 기재된 Precision, Recall, F1 score다. 위와 같은 평가 기준을 사용하여 모델을 평가하고 그중 가장 적합한 모델을 선택한다.

이후 선택된 모델을 이용하여 학습 -> 평가 -> (재 학습) -> 배포의 과정을 거친다.

이미지나 데이터를 레이블링 하는 과정에서 SageMaker Ground Truth 이용하면 ML훈련용 데이터 레이블을 쉽게 할 수 있다. 묶음으로 이미지를 레이블링 하는 방법이나 인력을 소싱 할 수 있도록 지원해주고 또한 자동화된 레이블링 기술까지 지원해 준다.

샘플 이미지가 부족한 경우 Data Augmentation방식을 이용해서 보완할 수 있다. Data Augmentation방식이란 한 샘플 이미지의 부분을 잘라내거나 색을 바꾸고 회전시키고 왜곡시키는 등의 방식으로 여러 다른 이미지처럼 보이게 하는 방식이다. Data Augmentation을 위한 여러 오픈 소스 또한 존재한다.

PPE detection이라는 서비스도 제공한다. 이 서비스는 배포나 학습의 과정 없이 API형태로 제공되기 때문에 API를 어플리케이션에 적용하는 것 만으로 이용이 가능하다.

이 기능은 Head cover, Face cover 등을 탐지하거나 얼굴을 인식하는 수준의 기능을 제공한다. 사용자는 이를 보호장구 확인 등의 기능에 이용할 수 있다.

**2.1.3 Amazon Personalize**

Amazon Personalize란 ML에 대한 전문지식이 없이도 데이터만 있으면 손쉽게 사용자 경험(UX)에 기반한 개인화를 통해 검색 성능, 결제, 사용자 만족도, 매출 등의 향상에 영향을 줄 수 있는 추천 ML 모델을 제공해주는 서비스이다.

동작 과정은 단순화하면 데이터파악 -> 데이터필드 선택 -> 영향이 큰 매개변수 선택 -> 추론 -> 평가의 과정을 반복해서 모델을 생성하는 과정으로 정리된다.

조금 더 자세히 살펴보면 사용자가 인터랙션, 유저, 아이템 3개의 데이터를 AWS S3 버킷을 통해 Personalize에 전송하는 것부터 시작한다. 그러면 전송된 데이터를 바탕으로 결과에 큰 영향을 주는 매개변수, 알고리즘을 이용하여 학습하고 모델을 만들어 주고 이 모델과 연동할 수 있는 커넥터를 API형태로 제공해준다. 사용자는 이 API를 본인의 애플리케이션에 적용하여 서비스를 이용할 수 있다. 따라서 사용자는 복잡한 ML과정에 대해 알 필요가 없다. 이렇게 솔루션이 완성되었으면 배포를 하면 된다. 배포한 솔루션은 캠페인이라는 이름으로 관리되는데 캠페인은 매번 자동 업데이트되지만 API를 통해 수동으로 업데이트 할 수 있다.

부가적으로 살펴보자면 데이터는 두가지 데이터로 나뉘어 처리되는데 그중Interaction 데이터가 학습에 필수적이다. Interaction Data란 평점, 언제, 어디서, 누가 이러한 고객의 활동에서 나오는 데이터를 말한다. 이외에 유저의 메타데이터 아이템에 대한 메타데이터는 추가적인 데이터로 본다. 매개변수 설정 시에 Interaction데이터를 더 중요한 데이터로 보고 부수적으로 메타데이터를 고려하는 방식으로 구성되어 있다.

데이터를 전송하는 방법은 단순하다. CSV형태의 자료를 입력하면 자동으로 JSON형태로 저장되어 활용되기 때문에 CSV파일을 업로드 하는 것으로 간단하게 처리할 수 있다.

그리고 본인 서비스에 알고리즘을 선택. 유저 개인화, 유사제품, 개인화 랭킹 중 하나의 서비스를 선택하여 학습을 시작할 수 있다. 대부분의 경우 AWS Personalization에서 가장 최근 개발된 User personalization알고리즘이 사용된다. 위의 3가지 서비스를 모두 커버할 수 있을 뿐 아니라 성능 또한 뛰어나다.

서비스 종류별로 사용하는 API가 다르다 예를 들어 유사제품 서비스의 경우 유사제품 서비스에 맞는 API를 선택해 주어야 한다. 이 점을 유의하자. 이 API를 이용해서 실제 어플리케이션에서 사용이 가능하다.

Amazon Personalize 콘솔창에서 일련의 과정을 모니터링 할 수도 있다.

**2.1.4 시계열 데이터 예측 지원 서비스**

시계열 데이터의 예측에는 AWS에 3가지 서비스가 사용될 수 있다. 3개의 서비스는 각각의 장, 단점을 가지고 있으니 환경에 따라 적합한 서비스를 골라 사용하면 된다.

1. Amazon Forecast

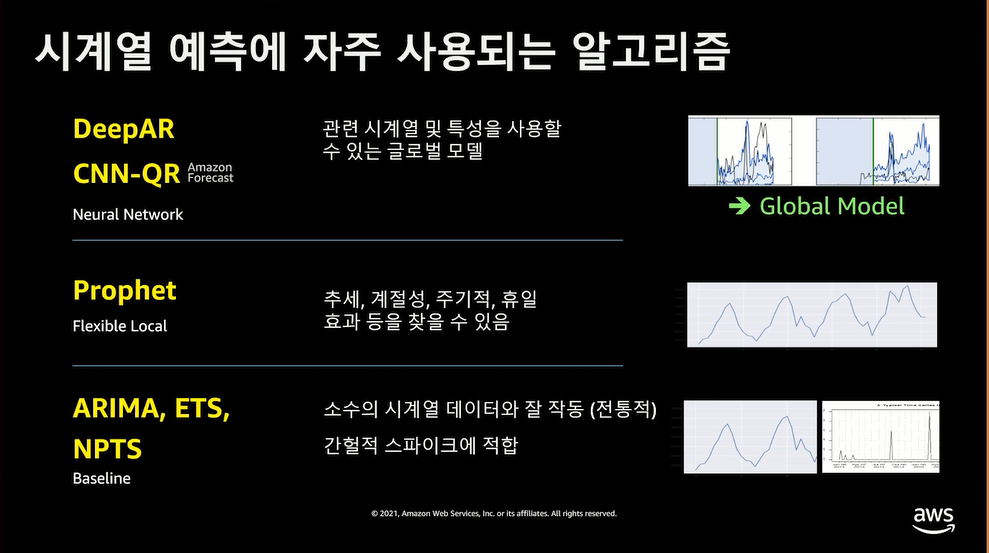
웹 환경 위에서 클릭 몇 번으로 학습된 모델을 제공한다. ML에 대한 기반이 없는 조직일 수록 효과가 크다. AutoML을 이용해서 최적의 모델을 자동으로 만들어낸다. AutoML은 wQL이라는 평가 기준을 이용해서 모델을 평가하고 최선의 모델을 이용해서 학습한 이후 결과를 도출해낸다. 단점으로는 학습 시간이 길고 예측 가능한 최대 개수가 제한적이며 시각화가 불편하다는 점이 있다. 하지만 ML에 대한 전문적인 지식이 필요하지 않고 EDA과정이 불필요하다는 장점이 있다.

1. SageMaker

조직내에 ML에 전문적인 부서가 존재할 경우 효과적이다. 학습, 평가, 재 학습 배포의 파이프라인을 직접 조작하여 가장 효율적인 모델을 도출해 낼 수 있다. 즉 고객환경에 좀더 최적화된 모델을 선택할 수 있다는 장점을 가진다.

1. Gluon TS

알고리즘부터 개발이 가능한 MXNet(ML 프레임워크)위에서 동작하는 라이브러리다. 오픈소스 기반이기 때문에 다른 유저가 개발해 둔 알고리즘을 가져다 쓸 수도 있다. SageMaker에서 사용할 수 있는 시계열모델을 직접 생성하는데 쓰이는 가장 나이브한 오픈소스 Toolkit이라고 할 수 있다. 장점은 빠르게 최신의 딥러닝 모델 구축, 테스트가 가능하다는 점이 있다.

일반적으로 시계열 예측에 자주 사용되는 알고리즘 3개는 다음과 같다.

**2.2 활용 방식**

**2.1.1 Amazon Lookout for Equipment 및 Amazon Lookout for Vision**

Lookout for Equipment의 경우.

자동화된 설비 유지보수 방식에 주로 활용되고 있다. 유지보수 방식은 다음과 같이 발전되어 왔다. 고장 시 수리(다운타임 존재) -> 정기적 유지보수(비용 커짐) -> 통계기반 예측(정확도 및 시간적 비용) -> 자동화.

ML을 사용한 설비 장애 예측 방식은 공정라인 정지 리스크 감소, 품질보증, 기기사용기간 증가 등의 많은 장점을 가지고 있다. 아래와 같은 조건이 갖춰졌을 때 더 큰 효용성을 낳을 수 있다.

* 적합한 적용분야일 때.
* 다양한 자산과 데이터 인프라의 양과 질이 충분할 때.
* 시간, 규모, 능력이 충분할 때.

Lookout for Vision

비전 기술로 품질검사를 자동화하여 결함 재작업, 폐기, 소비자 반품 등의 비용을 최소화하여 운영 비용을 최소화하는 식으로 활용되고 있다. 기존방식 = 육안검사 인라인검사 -> 최종 결과물 검사 방식을 사용한다. 위와 같은 방식은 유연, 민첩하지만 처리량에 한계가 존재했고. 개인간 기준의 차이로 인한 일관성이 떨어진다는 단점이 존재했다.

컴퓨터비전 = 인라인검사 -> 오프라인 검사 방식을 사용한다. 이는 빠르고 반복이 가능하며 비용이 절감된다는 장점을 가지지만 높은 초기비용이 필요하고 유연성이 떨어지며 적용범위가 협소하다는 단점을 가진다.

**2.2.2 Amazon Personalize**

세션기반 상품 추천 알고리즘을 제공해 줄 뿐 아니라. 유사 아이템, 제품추천, 알림 이메일 또는 SNS서비스 등의 개인화 서비스에 이용 가능하다. 세션기반의 추천은 유저의 취향 변화에도 민감하게 대응 가능하다.

실제 활용 사례를 들면 VOD서비스에서 제품과 유저, 유저 활동 메타데이터를 활용하여 Reranking즉 가져온 아이템 리스트의 순서를 재배열하고 재배열된 개인화된 리스트를 고객에게 제공할 수 있다.

**2.2.3 산업자산을 위한 ML**

산업용자산을 위해 만들어진 기계학습 모델들을 총칭한다. AWS ML모델들의 경우 대부분 자산종류와 크게 관련이 없으며 기계학습 플로우가 자동화 되어있다. 덕분에 전문가의 도움 없이 쉽게 구성할 수 있다. 일반적으로 정상적인 동작을 학습해두고 범주를 벗어가면 비정상 동작으로 파악 사용자에게 경고를 해 주는 방식으로 이용된다.

실례로 하수처리 시설에 사용되는 모델을 살펴보자. 펌프에 RPM 유량 측정 센서를 장착하고 정상범위를 벗어난 유량을 보이는 데이터들을 감지하여 비효율적인 동작이나 비정상 동작임을 사용자에게 경고해주는 방식으로 동작한다. 또한 이를 모니터링하기 쉽도록 그래프나 가시적인 결과물로 보여준다.

4. 결론.

아마존이 쌓아온 AI/ML에 관한 노하우를 살펴볼 수 있었다. AI/ML 분야의 다양한 기능을 산업에 활용될 수 있게 하는데도 적극적일 뿐 아니라 AI/ML기술을 위한 인프라를 제공하는 부분에서도 선두에 있었다.

이번 Conference를 보고 가장 눈에 띄는 점은 두가지가 있었다.

첫째 ML과 관련된 지식이 부족한 사용자 또는 조직에게도 서비스를 제공하기 위해 가장 쉬운 원 클릭 방식으로 ML서비스를 제공할 뿐만 아니라 정 반대되는 매개변수의 가중치 및 ML프레임워크까지 조작할 수 있는 세분화된 ML서비스와 인프라를 제공한 다는 점이 눈에 띄었다. 두 부분은 상당히 상반된 부분인데 다른 두가지 서비스로 제공되기도 하고 두 부분을 적절히 녹여내어 하나의 서비스로 제공하는 부분은 혁신적으로 다가왔다. 사용자의 입장에서는 하나의 웹 애플리케이션처럼 보이는 구조로 ML서비스를 제공하지만 그 안에서 세밀하게 가중치나 알고리즘을 연계된 AWS서비스를 이용하여 조절할 수 있었다. 뿐만 아니라 AWS 서비스를 유기적으로 활용하면 데이터 레이블링이나 일련의 과정의 편의성을 향상시킬 수 있게 해 두었다. 이러한 방식은 ML에 대한 지식이 부족한 사용자는 낮은 복잡도로 서비스를 이용할 수 있게 해 줄 뿐 아니라 고수준의 사용자가 원하는 세밀한 조정까지 충족시킬 수 있는 것이다.

둘째 정말 넓은 산업분야에 적용시도를 하고 있다는 점이었다. 웹 쇼핑몰부터 시작하여 VOD 서비스, 공장 등 다양한 서비스를 다양한 산업분야에 센서를 이용한 방식부터 Vision방식까지 다양한 적용되고 있었다. 또 그 용도는 단순한 분류에서 끝나지 않고 예측, 분석까지 넓은 분야에서 이용되고 있었다. 이러한 비즈니스 모델은 눈에 띄는 성과를 내고 있었으며 또 제공된 서비스들은 자체적으로 재 학습을 거치거나 사용자의 조정을 통해서 점점 발전된 서비스로 진화하고 있었다. 또 기존의 몇 달이 걸리는 서비스 제공시간이 시간단위로 줄어들어 있는 모습을 보여줬다. 이러한 모습을 보며 음악이나 스포츠 등의 더 다양한 범위에 또 지금처럼 큰 조직이 아닌 개인에게 서비스로서 제공되는 일이 곧 다가올 것이라는 예측을 해볼 수 있었다.

5. 참고자료.

AWS INNOVATE Confernce AI/ML트랙 중

* AWS 기반 기계학습 서비스의 5가지 주요목표.
* Amazon Lookout for Equipment 및 Amazon Lookout for Vision.
* Amazon Rekognition으로 컴퓨터 비전 서비스 만들기.
* Retail 기업의 Amazon Forecast 사용기.
* Amazon Personalize를 활용한 더 빠르고 정확한 추천서비스 구축.

위의 5가지 트랙.