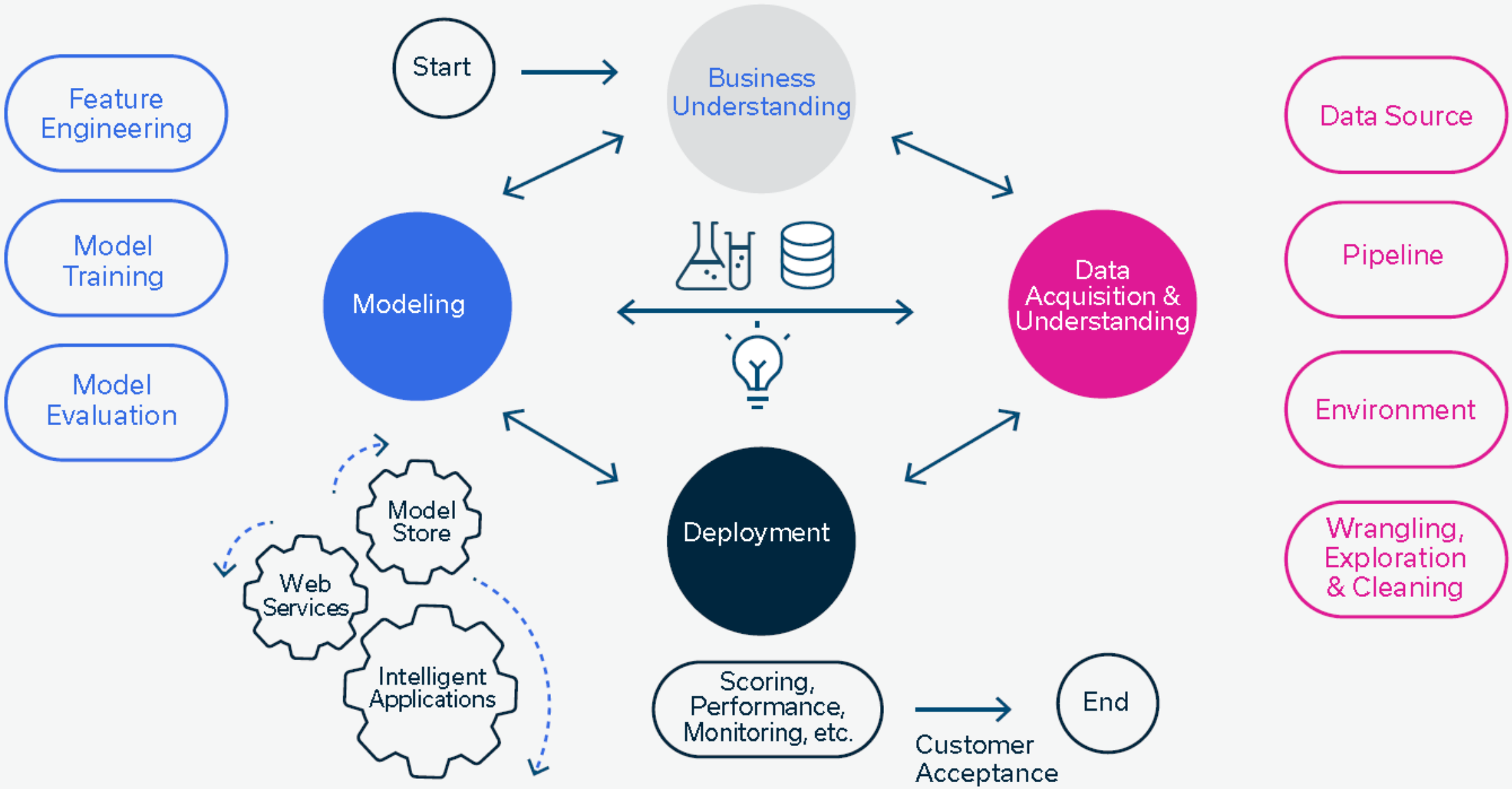
MLOps 구축가이드

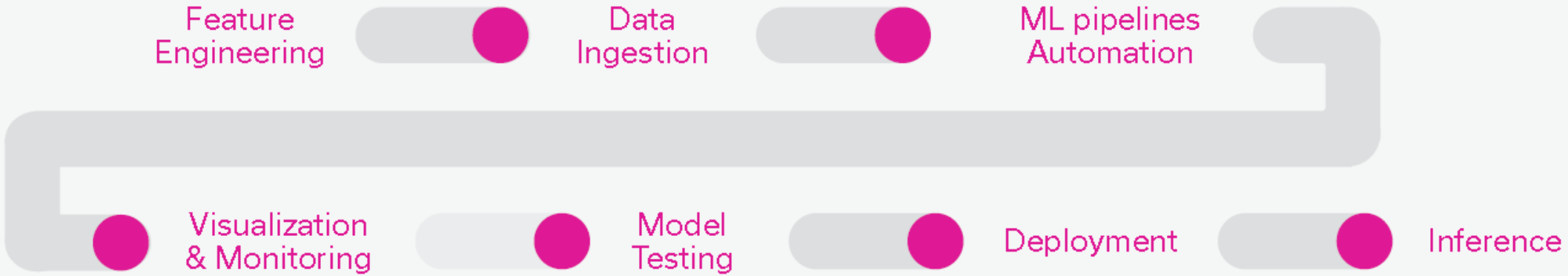
내용 출처 : Complete Guide to MLOps
(https://www.run.ai/white-papers)

박경규

MLOps


Feature Engineering
Model Training
Model Evaluation
Start
Business Understanding
Modeling
Data Acquisition & Understanding
Deployment
Model Store
Web Services
Intelligent Applications
Scoring, Performance, Monitoring, etc.
Customer Acceptance
End
Data Source
Pipeline
Environment
Wrangling, Exploration & Cleaning

ML 인프라


Feature Engineering
Data Ingestion
ML pipelines Automation
Visualization & Monitoring
Model Testing
Deployment
Inference

ML인프라 구성 요소

■ 모델 선택

머신러닝 모델 선택은 잘 맞는 모델을 선택하는 과정입니다.

어떤 데이터를 수집할지, 어떤 도구를 사용할지, 어떤 어떤 구성 요소가 필요한지,

구성 요소가 어떻게 상호 연결되는 방식을 결정합니다.

■ 데이터 수집

데이터 수집 기능은 모든 기계 학습 인프라의 핵심입니다.

이러한 기능은 모델 교육, 적용 및 개선을 위한 데이터를 수집하는 데 필요합니다.

데이터 수집 도구를 사용하면 상당한 사전 처리 없이 광범위한 소스의 데이터를 집계하고 저장할 수 있습니다.

이를 통해 팀은 실시간 데이터를 활용하고 효과적으로 데이터 세트 생성에 협력합니다.

■ ML 파이프라인 자동화

스크립트 및 이벤트 트리거에 따라 기계 학습 워크플로를 자동화할 수 있는 다양한 도구가 있습니다.

파이프라인은 데이터 처리, 모델 교육, 모니터링 수행에 사용됩니다. 작업을 수행하고 결과를 배포합니다.

이러한 도구를 사용하면 효율성을 높이고 프로세스 표준화를 보장하는 동시에 더 높은 수준의 작업에 집중할 수 있습니다.

인프라를 개발할 때 도구를 개별적으로 통합하고 조정하여 처음부터 도구 체인을 생성할 수 있습니다.

ML인프라 구성 요소

■ 시각화 및 모니터링

ML 시각화 및 모니터링은 워크플로가 얼마나 원활하게 진행되고 있는지, 모델 훈련이 얼마나 정확한지에 대한 관점을 얻고, 모델 결과에서 통찰력을 도출하는 데 사용됩니다.

팀이 시스템 측정항목을 신속하게 해석할 수 있도록 시각화를 기계 학습 워크플로의 어느 시점에나 통합할 수 있습니다.

모니터링은 전체적으로 통합되어야 합니다.

시각화 및 모니터링을 기계 학습 인프라에 통합할 때 도구가 데이터를 일관되게 수집하는지 확인해야 합니다.

■ 모델 테스트

ML모델을 테스트하려면 훈련 단계와 배포 단계 사이에 도구를 통합해야 합니다. 이 도구는 결과가 예상대로인지 확인하기 위해 수동으로 레이블이 지정된 데이터 세트에 대해 모델을 실행하는 데 사용됩니다.

철저한 테스트에는 다음이 필요합니다.

- 정성적, 정량적 데이터를 수집하고 분석합니다.
- 동일한 환경에서 여러 훈련이 실행됩니다.
- 오류가 발생한 위치를 식별하는 기능을 설정하려면 인프라에 모니터링, 데이터 분석 및 시각화 도구를 추가해야 합니다.

  또한, 자동화된 환경 생성 및 관리를 설정해야 합니다. 통합 테스트를 수행해야 합니다.

ML인프라 구성 요소

■ 배포

배포는 아키텍처에서 고려해야 할 마지막 단계입니다.이 단계에서는 모델을 패키지화하여 개발 팀이 서비스나 애플리케이션에 통합할 수 있도록 합니다.

MLaaS(서비스형 기계 학습)를 제공하는 경우 모델을 프로덕션 환경에 배포하는 것을 의미할 수도 있습니다.

이 배포를 통해 사용자로부터 데이터를 가져오고 결과를 사용자에게 반환할 수 있습니다.

일반적으로 MLaaS에는 모델 컨테이너화가 포함됩니다.

모델이 컨테이너에 호스팅되면 최종 환경에 관계없이 확장 가능한 분산 서비스로 모델을 제공할 수 있습니다.

■ 추론

배포 단계에서는 딥 러닝 프레임워크를 평가하고 새로운 데이터의 지속적인 추론을 위해 요구 사항에 가장 적합한 프레임워크를 선택하는 것이 중요합니다.

하드웨어 리소스를 고갈시키지 않고 프로덕션 성능 요구 사항을 충족하는 프레임워크를 선택하고 최적화해야 합니다.

예를 들어, 자율 주행 자동차에서 실행되는 컴퓨터 비전 모델은 자동차에 탑재된 하드웨어를 고려하면서 밀리초 속도로 추론을 수행해야 합니다. 최근 몇 년 동안 범용 모델 파일 형식이 개발되면서 생산 요구 사항에 따라 프레임워크 간에 모델을 이동하는 프로세스가 더욱 쉬워졌습니다.

주요 고려 사항

■ 위치

기계 학습 워크플로우가 수행되는 위치에 주의를 기울이십시오.

온프레미스 운영과 클라우드 운영에 대한 요구 사항은 크게 다를 수 있습니다.

또한, 선택한 위치는 모델의 목적을 지원해야 합니다.

Training 단계에서는 주로 비용 고려 사항과 운영 편의성에 중점을 두어야 합니다.

훈련 데이터를 저장할 위치를 결정할 때 데이터와 관련된 보안 및 규정도 중요한 고려 사항입니다.

온프레미스나 클라우드에서 Train을 수행하는 것이 더 저렴하거나 쉬울까요?

대답은 모델 수, 수집되는 데이터의 크기와 특성, 인프라 자동화 능력에 따라 달라질 수 있습니다.

추론 단계에서는 성능 및 대기 시간 요구 사항과 대상 위치에서 사용 가능한 하드웨어 간의 균형을 맞추는 데

중점을 두어야 합니다.

빠른 응답 또는 매우 낮은 대기 시간이 필요한 모델은 로컬 또는 에지 인프라에 우선 순위를 두고

저전력 로컬 하드웨어에서 실행되도록 최적화되어야 합니다.

약간의 지연 시간을 견딜 수 있는 모델은 클라우드 인프라를 활용할 수 있으며, "더 무거운" 추론 워크플로를 실행하는 데

필요한 경우 확장할 수 있습니다.

주요 고려 사항

■ 컴퓨팅 요구 사항

기계 학습에 사용되는 하드웨어는 성능과 비용에 큰 영향을 미칩니다.

일반적으로 딥러닝 모델을 실행하는 데 GPU가 필요합니다. GPU 또는 CPU의 효율성은 운영 및 클라우드 비용, 프로세스가 완료될 때까지 기다리는 데 소요되는 시간, 더 나아가 출시 시간에 영향을 미칩니다.

기계 학습 인프라를 구축할 때 리소스의 성능을 저하시키는 것과 리소스를 과도하게 사용하는 것 사이의 균형을 찾아야 합니다. 성능이 부족하면 초기 비용을 절약할 수 있지만 추가 시간이 필요합니다. 그리고 효율성을 감소시킵니다.

압도적인 성능을 발휘하면 하드웨어에 의한 제한을 받지 않지만 사용하지 않은 리소스에 대한 비용이 발생합니다.

■ 네트워크 인프라

효율적인 기계 학습 작업을 보장하려면 올바른 네트워크 인프라가 중요합니다.

병목 현상 없이 외부 소스와 데이터를 수집하고 전달해야 합니다.

네트워킹 기능이 처리 및 스토리지 기능과 얼마나 잘 일치하는지 주의 깊게 측정해야 합니다.

■ 스토리지 인프라

자동화된 ML 파이프라인은 모델의 데이터 요구 사항에 따라 적절한 스토리지 볼륨에 액세스할 수 있어야 합니다.

온프레미스 또는 클라우드 중 이 스토리지를 어디에 배치할지 미리 고려해야 합니다.

주요 고려 사항

■ 데이터 센터 확장

기계 학습을 기존 비즈니스 운영에 통합하는 경우 현재 인프라를 확장하기 위해 노력해야 합니다.

처음부터 시작하는 것이 더 쉬워 보일 수도 있지만 이는 비용 효율적이지 않고 생산성에 부정적인 영향을 미칠 수 있는 경우가 많습니다. 더 나은 옵션은 보유하고 있는 기존 인프라 리소스와 도구를 평가하는 것입니다.

기계 학습 요구 사항에 적합한 모든 자산을 통합해야 합니다.

■ 보안

모델을 훈련하고 적용하려면 막대한 양의 데이터가 필요하며, 이는 종종 가치 있고 민감한 데이터입니다.

예를 들어 금융 데이터나 의료 이미지 등이 있습니다.

기계 학습 인프라를 생성할 때 데이터를 적절하게 보호하기 위해 모니터링, 암호화 및 액세스 제어 기능을 구축하는 데 주의를 기울여야 합니다.

또한 데이터에 어떤 규정 준수 표준이 적용되는지 확인해야 합니다.

결과에 따라 데이터 저장의 물리적 위치를 제한하거나 데이터를 처리하여 민감한 정보를 사용하기 전에 제거해야 할 수도 있습니다.

머신러닝 자동화

■ 데이터 과학 파이프라인 속도 향상

기계 학습 자동화를 통해 데이터 과학자는 기계 학습 프로세스 생성을 자동화할 수 있습니다.

기계 학습 자동화가 없으면 ML 프로세스는 데이터 준비부터 교육, 실제 배포까지 몇 달이 걸릴 수 있습니다.

기계 학습 파이프라인의 속도를 높이는 데 도움이 되는 기계 학습 자동화 도구가 만들어졌습니다.

경우에 따라 이는 모델 선택과 같은 특정 작업만 자동화하는 것을 의미합니다.

다른 경우에는 전체 기계 학습 운영 프로세스를 자동화하는 것을 의미합니다.

■ AutoML

기계 학습에서 모델 선택은 기계 학습 구현에 적합한 후보 모델을 선택하는 프로세스입니다.

모델 성능, 복잡성, 유지 관리 가능성은 물론 사용 가능한 리소스에 따라 결정됩니다.

모델 선택 프로세스는 모델 개발 파이프라인의 구조를 결정합니다.

모델 선택 자동화는 하이퍼파라미터 최적화와 거의 동일한 방식으로 수행됩니다.

이는 둘 다 본질적으로 동일한 최종 목표를 추구하기 때문입니다.

차이점은 모델 선택에 AIC(Akaike Information Criterion) 또는 BIC(Bayesian Information Criterion)와 같은 방법을 통한 보다 광범위한 필터링이 포함될 수도 있다는 것입니다.

머신러닝 자동화

■ 하이퍼파라미터 최적화

하이퍼파라미터는 모델이 훈련되기 전에 정의되는 값입니다.

이러한 값은 모델 훈련을 관리하고 모델의 최종 정확도에 영향을 미칩니다.

하이퍼파라미터에는 학습률, 활성화 함수, 유닛 및 레이어 수, 에포크 수가 포함됩니다.

모델을 개선하려면 하이퍼파라미터를 최적화해야 합니다.

이는 일반적으로 Random search, Grid search,r Bayesian optimization와 같은 알고리즘을 적용하여 수행됩니다.

■ 모델 선택

기계 학습에서 모델 선택은 기계 학습 구현에 적합한 후보 모델을 선택하는 프로세스입니다.

모델 성능, 복잡성, 유지 관리 가능성은 물론 사용 가능한 리소스에 따라 결정됩니다.

모델 선택 프로세스는 모델 개발 파이프라인의 구조를 결정합니다.

모델 선택 자동화는 하이퍼파라미터 최적화와 거의 동일한 방식으로 수행됩니다.

이는 둘 다 본질적으로 동일한 최종 목표를 추구하기 때문입니다.

차이점은 모델 선택에 AIC(Akaike Information Criterion) 또는 BIC(Bayesian Information Criterion)와 같은 방법을 통한 보다 광범위한 필터링이 포함될 수도 있다는 것입니다.

머신러닝 자동화

■ Feature 선택

Feature 선택은 기계 학습 모델에 사용되는 예측 변수 수를 구체화하는 프로세스입니다.

모델에 포함된 feature 의 수는 학습, 이해 및 실행의 난이도에 직접적인 영향을 미칩니다.

Feature 선택을 자동화할 때 테스트는 래퍼, 필터 또는 임베디드와 같은 다양한 알고리즘 방법 중 하나 이상을 사용하도록 스크립트로 작성됩니다.

Feature 선택 테스트를 수행한 후 오류율이나 프록시 측정값이 가장 낮은 기능이 선택됩니다.

■ 데이터 전처리

데이터 전처리에는 사용 전 데이터 정리, 인코딩 및 확인이 포함됩니다.

자동화된 작업은 하이퍼파라미터 및 모델 최적화 단계를 수행하기 전에 기본적인 데이터 전처리를 수행할 수 있습니다.

- 기본적인 데이터 전처리: 열 유형 감지, 숫자 데이터로 변환, 누락된 값 처리
- 고급 전처리 : Feature 선택, Feature 생성, 인코딩, 데이터 압축, 텍스트 콘텐츠 처리, 데이터 정리

머신러닝 자동화

■ 전이 학습 및 사전 훈련된 모델

기계 학습에서 전이 학습에는 유사한 데이터 세트에 대해 이미 훈련된 모델을 가져와 기계 학습 이니셔티브에 사용하는 것이 포함됩니다.

일반적으로 이 모델은 기본으로 사용되며 정확한 요구 사항에 맞게 추가 학습됩니다.

기계 학습 자동화 측면에서 이 초기 모델은 최종 모델에 대한 데이터 세트를 수집하거나 준비하는 동안 최종 모델과 동일한 방식으로 학습될 수 있습니다.

이는 특히 매우 정확한 모델이 필요하지 않은 경우 상당한 시간을 절약할 수 있습니다.

■ 네트워크 아키텍처 검색

준비 및 모델 선택 프로세스를 넘어 기계 학습 알고리즘의 동적 개발로 확장할 수도 있습니다.

새로운 개발로 인해 네트워크 아키텍처 검색이 일부 자동화되었습니다.

특히 신경구조탐색(NAS) 방법은 경사하강법, 강화학습, 진화알고리즘을 기반으로 한 문제에 탐구 및 적용되고 있다.

이 방법은 이미 다음을 포함한 여러 도구에 통합되었습니다.

오픈 소스 라이브러리인 AutoKeras와 그 결과는 자율주행차를 포함한 여러 프로젝트에 통합되었습니다.

유연한 ML 워크플로 구축

ML 워크플로는 기계 학습 프로젝트 중에 구현되는 단계를 정의합니다.

일반적인 단계에는 데이터 수집, 데이터 사전 처리, 데이터 세트 구축, 모델 교육 및 개선, 평가, 프로덕션 배포가 포함됩니다.

모델 및 기능 선택 단계와 같은 기계 학습 작업 워크플로의 일부 측면을 자동화할 수 있지만 전부는 아닙니다.

이러한 단계는 일반적으로 표준으로 받아들여지지만 변경의 여지도 있습니다.

기계 학습 워크플로를 만들 때 먼저 프로젝트를 정의한 다음 작동하는 접근 방식을 찾아야 합니다.

모델을 엄격한 작업 흐름에 맞추려고 하지 마십시오.

오히려 소규모로 시작하여 프로덕션급 솔루션으로 확장할 수 있는 유연한 워크플로를 구축하세요.

ML 워크플로 모범사례

■ 프로젝트 정의

프로젝트 목표를 신중하게 정의하십시오.
프로젝트를 정의할 때 다음 측면을 고려하십시오.

▪ 현재 프로세스는 무엇입니까?

일반적으로 모델은 기존 프로세스를 대체하도록 설계됩니다.
기존 프로세스가 어떻게 작동하는지, 목표가 무엇인지, 누가 수행하는지, 무엇이 성공으로 간주되는지를 이해하는 것이 모두 중요합니다. 이러한 측면을 이해하면 모델이 충족해야 하는 역할, 구현에 존재할 수 있는 제한 사항, 모델이 충족하거나 초과해야 하는 기준을 알 수 있습니다.

▪ 무엇을 예측하고 싶나요?

예측하려는 대상을 주의 깊게 정의하는 것은 수집해야 할 데이터와 모델 교육 방법을 이해하는 데 중요합니다.
이 단계를 최대한 자세히 설명하고 결과를 정량화 하십시오. 목표를 측정할 수 없으면 목표 달성에 어려움을 겪게 됩니다.
Research - 접근 방식을 구현하기 전에 다른 팀이 유사한 프로젝트를 어떻게 구현했는지 조사하는 데 시간을 투자해야 합니다. 그들이 사용한 방법을 빌리거나 실수로부터 배울 수도 있고,시간과 돈을 절약하십시오.
A/B 테스트 - 현재 모델을 기존 프로세스와 비교할 수 있습니다.이를 통해 모델이 효과적이며 팀과 사용자에게 가치를 더할 수 있는지 여부를 확인하거나 거부할 수 있습니다.

▪ 데이터 소스는 무엇입니까?

현재 프로세스가 어떤 데이터에 의존하는지, 어떻게 수집되는지, 어느 정도의 양인지 평가하세요.
이러한 소스에서 예측을 형성하는 데 필요한 특정 데이터 유형과 포인트를 결정해야 합니다.

ML 워크플로 모범사례

■ 효과적인 접근 방식 찾기

기계 학습 워크플로 구현의 목표는 현재 프로세스의 효율성 및/또는 정확성을 향상시키는 것입니다.

이 목표를 달성하는 접근 방식을 찾으려면 다음을 수행해야 합니다.

- Research - 접근 방식을 구현하기 전에 다른 팀이 유사한 프로젝트를 어떻게 구현했는지 조사하는 데 시간을 투자해야 합니다. 그들이 사용한 방법을 빌리거나 실수로부터 배울 수 있어 시간과 비용을 절약할 수 있습니다.

- Experiment - 시작하기 위한 기존 접근 방식을 찾았든, 자신만의 접근 방식을 만들었든, 이를 실험해야 합니다.

ML 워크플로 모범사례

■ Full-Scale 솔루션 구축

PoC 솔루션에서 배포 가능한 솔루션으로 전환하려면 다음이 필요합니다.

- A/B 테스트
  현재 모델을 기존 프로세스와 비교할 수 있습니다.
  이를 통해 모델이 효과적이며 팀과 사용자에게 가치를 더할 수 있는지 여부를 확인하거나 거부할 수 있습니다.

- 기계 학습 API
  모델 구현을 위한 API를 생성하면 데이터 소스 및 서비스와 통신할 수 있습니다.
  모델을 기계 학습 서비스로 제공하려는 경우 이러한 접근성은 특히 중요합니다.

- 사용자 친화적인 문서
  코드, 메소드, 모델 사용 방법에 대한 문서가 포함됩니다.
  시장성이 있는 제품을 만들려면 모델을 활용하는 방법, 결과에 액세스하는 방법, 기대할 수 있는 결과의 종류를 사용자에게 명확하게 설명해야 합니다.

Thank you