

개발자와 데이터 과학자가 협업하는 5가지 주요 방식

데이터 과학자 이해하기

데이터 과학자의 작업 방식과 효과적인 인공 지능(AI) 기반 애플리케이션을 빌드하기 위해 데이터 과학자와 협력하는 방식을 알아야 합니다. 즉, AI에 관한 기본 지식을 갖추고 데이터 과학자인 동료들과 협업하는 방법을 잘 알아야 합니다. 다음은 데이터 과학자와 협력하고 AI 기반 지능형 애플리케이션을 빌드할 때 알아야 할 다섯 가지 주요 사항입니다. 이 체크리스트를 원활한 협력 관계를 형성하고 탁월한 애플리케이션 개발 협업을 진행하기 위한 가이드로 활용하시기 바랍니다.

1 데이터 과학자의 작업 방식 이해

일반적으로 데이터 과학자는 애플리케이션을 개발하거나 모델을 소프트웨어 요소로 통합하는 것보다 모델을 구축하고 개선하는 데 더 많은 관심이 있습니다. 지속적 통합/지속적 제공(CI/CD) 파이프라인을 구축하거나 애플리케이션 코드를 작성하는 데 관여하고 Python, R, Jupyter Notebook과 같은 친숙하지 않은 툴을 사용하는 것은 거의 원치 않습니다.

개발자와 동일한 목표로 작업하고 있다 하더라도 데이터 과학자가 개방형 협업을 먼저 제안하지는 않을 것입니다. 하지만 우수한 애플리케이션을 빌드하기 위해선 개발자와 데이터 사이언스 팀 간의 견고하고 일관된 팀워크가 필수적입니다. 적극적인 협업을 통해 다음과 같은 이점을 누릴 수 있습니다.

- ▶ AI를 효과적으로 활용하는 데이터 기반 지능형 애플리케이션 배포
- ▶ 데이터 과학자가 자신의 모델링 작업을 회사 및 고객에 더 많은 가치를 제공하는 배포 가능 솔루션에 사용할 수 있는 기회

데이터 과학자인 동료와의 협업 경험을 먼저 제안하고 촉진하는 역할은 개발자의 몫이 될 가능성이 높습니다. 이 체크리스트의 지침을 참고하여 유익한 방식으로 데이터 사이언스 팀과 관계를 형성하는 방법에 대해 알아보세요.

2 공통 기반 찾기

마찰 없는 협업과 모든 사람에게 효과적인 경험을 창출하려면 개발 사례에 대해 설명하고 이러한 사례가 데이터 과학자들의 노력을 어떤 방식으로 보완하는지 확인하는 것이 중요합니다. 이를 위해 권장되는 사항은 다음과 같습니다.

- ▶ **빈번한 점점 장려.** 빈번하고 정기적인 점점을 설정하면 함께 작업 중인 프로젝트가 제대로 진행되도록 하는 데 가장 큰 도움이 됩니다.
- ▶ **경계 존중.** 데이터 과학자는 애플리케이션을 프로덕션 단계로 발전시키는 방법에 대해 알기를 원치 않거나 알아야 할 필요가 없을 수 있습니다. MLOps가 인기 있는 개념이기는 하지만 일부 과학자는 개발자에게 자신의 Jupyter Notebook을 이메일로 보내는 방법을 선호합니다. 이들의 관심사와 선호하는 작업 방식을 존중하면 이에 화답할 것입니다.
- ▶ **서로의 프로세스 공유.** 데이터 과학자의 작업 방식을 이해하는 것 외에도 Git, Tekton 또는 쿠버네티스와 같이 프로덕션 단계에서 사용하는 프로세스와 툴을 공유합니다. 오픈소스 정신으로 데이터 과학자가 프로세스를 엿볼 수 있게 하세요.
- ▶ **협업을 위한 공통 플랫폼 사용.** Red Hat® OpenShift® Data Science와 같은 공통 클라우드 네이티브 AI 개발 플랫폼은 개발자와 데이터 사이언스 팀 간의 협업을 지원하고 촉진합니다. 이 플랫폼을 통해 AI 툴 사용이 대중화되고 팀은 지능형 애플리케이션 개발을 구현하고 가속화할 수 있습니다.

3 모델 교육 툴로 작업하는 방법 학습

데이터 과학자가 자주 사용하는 몇 가지 모델 교육 툴에 대해 최소한의 기본 지식을 습득합니다. 모델 교육 툴에 대한 실용적 지식이 있으면 모델이 구축되는 방식을 이해하는 데 도움이 됩니다. 가장 널리 사용되는 몇 가지 모델 교육 툴과 라이브러리는 다음과 같습니다.

- ▶ Jupyter 및 PyCharm 개발 환경
- ▶ PyTorch
- ▶ scikit-learn
- ▶ TensorFlow

이를 비롯한 여러 툴에 친숙해지면 배포된 모델 애플리케이션을 성공적으로 생성하는 데 도움이 됩니다. 또한 모델 생성에 수반되는 작업을 더 깊이 이해할 수 있고 모델이 지능형 애플리케이션에 원활히 통합되지 않는 경우 발생하는 문제를 해결할 수 있습니다.

4 가장 선호하는 툴과 프로세스 사용 지식

데이터 과학자 및 AI와 협력할 때는 다수의 새로운 프로세스와 몇 가지 새로운 툴에 대해 학습해야 합니다. 하지만 가장 선호하는 여러 가지 툴을 계속 사용하여 로직을 수행할 수 있습니다. 애플리케이션 코드 및 모델링은 선호하는 언어나 프레임워크로 수행할 수 있습니다.

예를 들어, Quarkus 개발자라면 Quarkus에서 애플리케이션 로직을 수행하여 REST(Representational State Transfer) 엔드포인트로 애플리케이션 프로그래밍 인터페이스(API)를 호출하도록 할 수 있습니다. 한편 데이터 과학자는 Python 또는 R과 같은 툴을 통해 실제 데이터 처리 및 예측을 처리합니다. AI와 데이터 사이언스는 복잡하지만 친숙한 툴과 프로세스를 사용해 작업을 간소하게 유지하세요.

5 모델이 애플리케이션의 일부임을 기억

모델링이 중요하듯이 모델 이전의 MLOps도 중요합니다. 다음은 모델이 프로덕션 단계에 배치될 때 계속해서 성능을 유지하도록 보장하기 위해 수행해야 하는 4가지 사항입니다.

- ▶ 개발 중인 애플리케이션에 효과적으로 작동하는 모델 제공 인프라 구축
- ▶ 모델 교육 및 제공 작업을 처리할 신규 CI/CD 파이프라인을 생성하거나 기존 CI/CD 파이프라인 확장
- ▶ 모델 제공 애플리케이션 확장
- ▶ Apache Kafka와 같은 스트리밍 데이터 서비스와 기타 데이터 수집 구성 요소 통합

마지막으로, 애플리케이션 배포는 시작에 불과합니다. 모델은 계속 변화하므로 모니터링해야 합니다. 모델 변동을 방지하기 위해 데이터 사이언스 팀과 긴밀히 협력하여 개발자 또는 개발자의 운영 담당자가 어떤 메트릭을 모니터링해야 하는지 정의하세요. 문제나 변경 사항이 발생하면 해당 데이터 과학자와 협업하여 모델을 개선하세요.

자세히 알아보기

Red Hat OpenShift Data Science를 방문하고 [e-book](#)을 읽어보세요.

사용해 보기

Red Hat OpenShift Data Science를 [사용해](#) 보고 데모를 미리 확인하세요.



Red Hat 소개

Red Hat은 권위 있는 어워드를 수상한 지원, 교육, 컨설팅 서비스로 고객이 여러 환경에 걸친 표준화를 진행하고, 클라우드 네이티브 애플리케이션을 개발하고, 복잡한 환경을 통합, 자동화, 보안, 관리할 수 있도록 지원합니다.

f www.facebook.com/redhatkorea
 구매문의 080 708 0880
buy-kr@redhat.com