



[자습자료] 복잡한 데이터 과학 프로젝트를 수행할 때 OOP 개념이 필요한 이유

소요시간 : 3 min

자습자료에서는 기존에 배웠던 부분, 재밌게 다뤄질 수 있는 새로운 부분, 개념 간의 복합적인 연관성을 초점으로 두고 있습니다. 정답을 찾는 것이 아니라 **개인학습** 시간에 집중을 하여 다양한 개념을 단기기억에서 장기기억으로 옮길 수 있도록 노력합니다.

OOP개념을 활용하면서 알 수 있는 장점을 살펴본다

코드분석을 진행합니다.

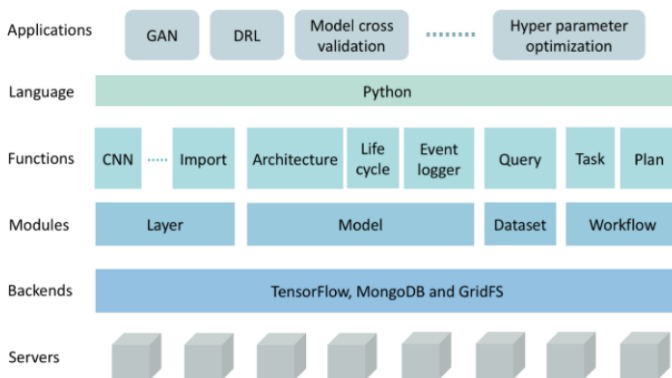
키워드 : Data Science Project, Code Reusability, Maintain Program

- 깃헙(ds-sa-oop)에서 제공되는 'InsuranceFraudDetection.zip' 파일을 활용하여 OOP 패러다임이 어떻게 활용될 수 있는지 코드를 분석해봅니다.
- 실제 서비스하는 프로덕트(product)에서는 OOP에서 중요한 키워드였던 '**재사용**'과 '**유지보수**'를 고려하며 수행해야 합니다.
- 대부분의 데이터 사이언스 프레임워크가 OOP 패러다임 내에서 수행되기때문에, OOP 패러다임에서 계속 개발되어야 재사용과 유지보수가 수월해집니다.
- 소규모 데이터 과학 프로젝트는 함수형 패러다임으로 작성되는 경우도 있습니다.
- 패러다임의 변화와 소스코드의 변화를 비교해보며 자신이 익힌 개념을 구체화합니다.
- 머신러닝과 딥러닝을 활용하는 경우 전체 라이프사이클에서 소스코드의 변화주기에 따라 OOP 패러다임의 활용정도가 다를 수 있습니다.

- 예를 들어, 변수, 함수, 클래스, 모듈, 라이브러리 파이썬파일(.py) Data파일

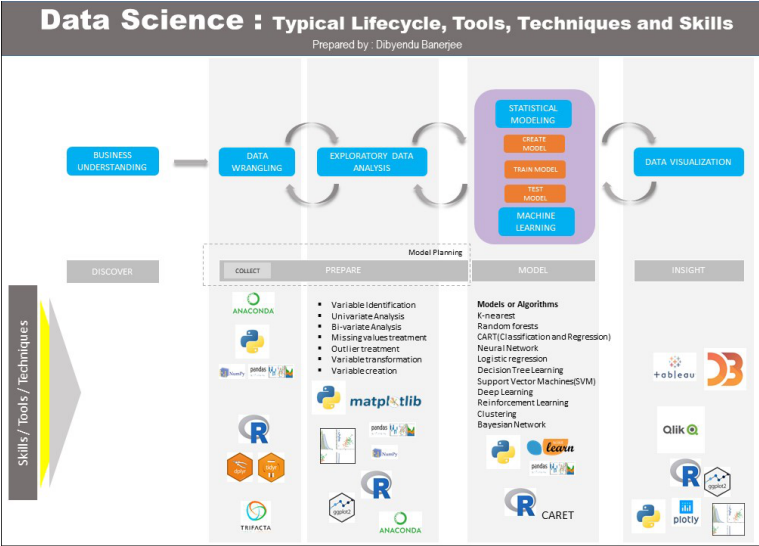


- 이런 경우 자주 사용할 수 있는 것, 재사용 가능한 것, 수정이 한 달에 3회 이상 되는 것 등 기준에 따라 OOP 개념을 접목시켜야 합니다.
- 텐서플로우에는 아래처럼 4가지 모듈에서 변화가 있을 수 있습니다.
 - 레이어 모듈 – CNN, RNN, Dropout, 배치 정규화 등과 같이 일반적으로 사용되는 신경망 레이어의 참조 구현을 제공합니다.
 - 모델 모듈 – train, evaluate, deploy 등 모델 수명주기 동안 발생하는 중간 상태를 관리하는데 도움이 됩니다. 상태는 유지, 캐시 및 리로드할 수 있습니다.
 - 데이터 세트 모듈 – 훈련 데이터 및 예측 결과를 관리합니다. TensorLayer는 다양한 기술을 사용하여 데이터 세트 성능을 최적화합니다.
 - 워크 플로우 모듈 – 동시 교육 작업에 대한 비동기 스케줄링 및 장애 복구를 지원합니다.



TensorLayer architecture

- 이처럼 분리된 모듈에 따라 별도의 구현이 필요하고 코드의 재사용이 지속적으로 발생하기 때문에 OOP에 대한 개념을 인지하셔야 합니다.
- 아래 그림은 데이터 과학 프로젝트를 수행하는 경우 발생할 수 있는 'Project LifeCycle' 입니다.
 - 단계별로 코드수정 및 유지보수를 고려하며 코드를 분석해봅니다.



☐ Mark as Completed

< Prev

Next >