

✓ EDA(Exploratory Data Analysis, 탐색적 데이터 분석 - 메타, 메타러닝이란 뭘까?)

✓ EDA(Exploratory Data Analysis, 탐색적 데이터 분석)

확증적 데이터 분석(CDA)



탐색적 데이터 분석(EDA)



- 1) 정의

- 수집한 데이터가 들어왔을 때, 이를 다양한 각도에서 관찰하고 이해하는 과정(데이터를 분석하기 전에 그래프나 통계적인 방법으로 자료를 직관적으로 바라보는 과정)

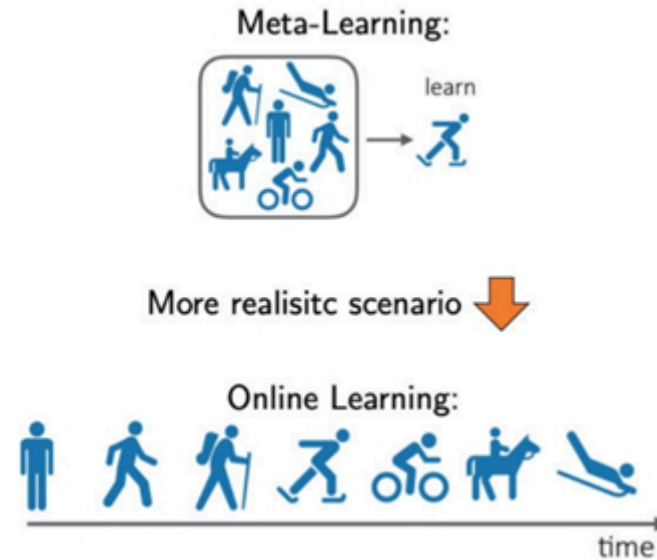
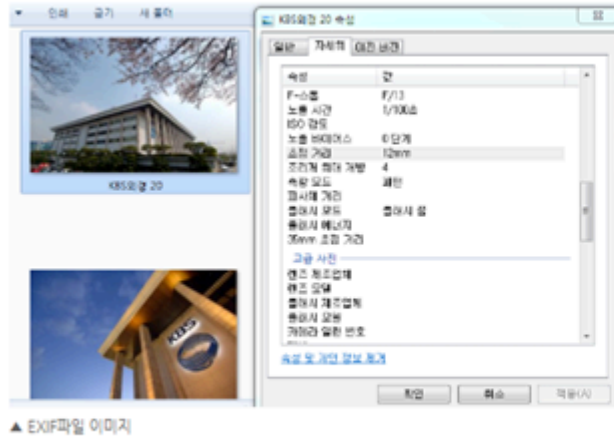
- 2) 필요한 이유

- 데이터의 분포 및 값을 검토함으로써 데이터가 표현하는 현상을 더 잘 이해하고, 데이터에 대한 잠재적인 문제를 발견할 수 있음
- 이를 통해, 본격적인 분석에 들어가기에 앞서 데이터의 수집을 결정할 수 있음
- 다양한 각도에서 살펴보는 과정을 통해 문제 정의 단계에서 미처 발생하지 못했을 다양한 패턴을 발견하고, 이를 바탕으로 기존의 가설을 수정하거나 새로운 가설을 세울 수 있음

• 3) 과정

- 기본적인 출발점은 문제 정의 단계에서 세웠던 연구 질문과 가설을 바탕으로 분석 계획을 세우는 것
- 분석 계획에는 어떤 속성 및 속성 간의 관계를 집중적으로 관찰해야 할지, 이를 위한 최적의 방법은 무엇인지가 포함되어야 함
- 분석의 목적과 변수가 무엇이 있는지 확인. 개별 변수의 이름이나 설명을 가지는지 확인
 - 데이터를 전체적으로 살펴보기 - 데이터에 문제가 없는지 확인. head나 tail부분을 확인, 추가적으로 다양한 탐색(이상치, 결측치 등을 확인하는 과정)
 - 데이터의 개별 속성값을 관찰 - 각 속성 값이 예측한 범위와 분포를 갖는지 확인. 만약 그렇지 않다면, 이유가 무엇인지를 확인.
 - 속성 간의 관계에 초점을 맞추어, 개별 속성 관찰에서 찾아내지 못했던 패턴을 발견(상관관계, 시각화 등)

✓ 메타, 메타러닝이란 뭘까?



- 메타라는 용어

- 한 단계 더 위에 있는 것을 가리킬 때 사용
- 대표적인 예시로, 데이터에 대한 정보를 나타낼 때, 데이터에 대한 데이터인 meta-data가 있음

- Meta-Learning?

- 첫 번째 학습

- 모델이 학습이라는 것은 주어진 데이터로부터 패턴이나 특징을 찾고 새로운 데이터가 주어졌을 때, 그 특징을 찾는 것(이는 모델의 파라미터 θ 를 찾는 것과 동일)
- Supervised Learning에서는 데이터에 대한 Label이 주어지기 때문에, 새로운 데이터에 대해서 Label을 찾는 것을 목적으로 파라미터를 학습(러닝)

- Unsupervised Learning에서는 데이터에 대한 레이블이 없으므로, 데이터 자체에 대한 특징을 알아서 찾아야 함
- 두 가지 Learning 방법은 데이터에 대한 특징을 찾는 것을 목표로 모델 파라미터를 학습
 - 두 번째 학습
 - 딥러닝을 하다보면, Learning rate를 조절하면서 Loss가 떨어지는 것을 비교해본 적이 있으실 텐데, 이 경우도 Meta Learning에 해당함
 - 모델을 훈련하는 learning보다 한 단계 위인, Hyperparameter에 대해서 적합한 값을 찾는 Learning을 진행하기 때문임
 - 또 다른 예시는 Model을 Ensemble하는 것 입니다. 이미 learning이 된 모델들을 가지고, 모델들의 예측을 기반으로 결정하는 새로운 모델을 Stacking으로 학습시킨다면, 1차적인 학습을 넘어서 학습을 진행하기 때문에 Meta-Learning임

✓ <<<참조자료 사이트>>>

1. [메타, 메타러닝이란 뭘까?](#)
2. ["하나를 알려주면 열을 안다" \[특별기획 AI 2030\] ② 메타학습](#)
3. [EDA \(Exploratory Data Analysis\) 탐색적 데이터 분석](#)
4. [범주형 데이터 다루기 : Encoding 기법 \(1\)](#)
5. [데이터 전처리 : 레이블 인코딩과 원핫 인코딩](#)
6. [범주형 데이터 다루기 : Encoding 기법 \(2\)](#)
7. [Confusion Matrix\(혼돈 행렬\)과 분류 성능 평가 지표](#)
8. [파이썬 머신러닝 완벽가이드 - 평가, 정확도](#)
9. [파이썬에서의 머신러닝 모델 평가 방법](#)

10. [평가\(정확도, 오차 행렬, 정밀도, 재현율\)](#)

11. [의미를 이해하는 통계학과 데이터 분석](#)

12. [ROC Curve](#)

13. [파이썬 ROC 커브, AUC 면적 구하기 예제](#)
