

데이터 분석의 유형

1. Descriptive analysis(기술통계)

주어진 데이터를 요약/집계하여 결과를 도출

과거의 데이터의 단순 계산/집계 Fact

분석 결과를 따로 해석하는 과정은 거치지 않음.

주로 사용되는 그래프는 Pie chart, Box plot, Bar plot, 요약 형식의 테이블

2. Exploratory analysis(탐색적 요인 분석)

여러 변수 간 trend, 패턴, 관계를 찾는 것

통계적 기법을 사용한 모델링이라기 보다는 그래프(주로 Plot)를 통한 사실 확인이 주된 작업

주로 데이터 분석 프로젝트 초기에 가설을 수립하기 위해 수행

데이터 탐색은 기본적으로 끝이 없는 작업이므로 분석의 목적이 가설 수립인지, 특정 변수들 간의 관계 파악인지, 트렌드 파악인지를 목적을 염두해두고 분석을 진행할 것

주로 사용되는 그래프는 Scatter plot, 시계열 차트 등

3. Inferential analysis (추론적 통계 분석)

샘플 - 모집단 간의 관계를 탐구하는 것이 이 분석 유형의 목적

샘플에서 얻어낸 정보가 모집단에도 적용될 수 있는지를 검토하는 것인데, 이는 자원이 한정되어 있어 샘플 크기를 무한정 늘릴 수 없기 때문에 발생하는 일반적인 문제를 해결하기 위한 분석 유형.

예) "95%의 확률로 모집단의 평균 점수는 80~85 점 사이라고 말할 수 있다",

예) "그룹 A의 성적이 그룹 B보다 유의미하게 높았다"는 식의 결론이 Inferential analysis의 결과

4. Causal analysis

독립 변수와 종속 변수 간의 인과관계가 있는지 여부를 확인하기 위한 분석.

선형 Regression이 가장 많이 사용되는 분석 방법

5. Predictive analysis

머신 러닝(Machine learning), 의사결정나무(Decision Tree) 등 다양한 통계적 기법을 사용하여 미래 혹은 발생하지 않은 어떤 사건에 대한 예측이 Predictive analysis의 주요 목표

"왜" 보다는 "정확한 예측"을 하는 것이 주요 관심

특정한 설명 변수가 어떤 매커니즘으로 목표 변수에 영향을 미치는가에 대한 설명력은 약하다.

수학 모델에 기반으로 함

6. Mechanistic analysis

독립 변수가 어떤 매커니즘으로 종속 변수에 영향을 미치는지를 분석하는 것

Causal analysis의 목적이 독립-종속 변수 간의 "인과 관계 여부"를 밝혀내는 것이었다면,

Mechanistic analysis의 목적은 이에 더해 어떠한 독립 변수가 어떠한 작용을 통해 독립 변수에 그러한 영향을 미치는지를 이해하는 것