

분석기획서 작성 방법

분석기획서는 비즈니스 문제를 식별하고 이를 해결하기 위한 방안 수립을 넘어, 데이터를 활용한 분석방법으로 해결 방안에 도움이 되는 결과를 도출하는 것을 기획하는 문서이다. 즉, 분석기획서 안에는 분석을 위한 데이터의 검토, 변수들의 역할, 적용할 분석기법, 활용 방안 및 기대 효과의 내용이 포함되어야 한다.

분석기획서가 작성되면 실제 데이터를 진단하는 단계를 거쳐, 문제가 없는 경우 분석을 실행하는 단계로 넘어간다. 데이터의 품질이 확보되지 않는 경우 PoC 또는 Pilot 프로젝트를 통해 사전 검증을 하는 경우도 있다.

1 비즈니스 배경 및 문제 식별

- 현재 어떤 업무에 어떤 문제가 발생하고 있고, 이 문제가 미치는 영향에 대해 상세히 기술하여 문제를 제기한다. 필요한 경우 현업 또는 전문가와의 인터뷰를 적극 활용한다.
- 문제가 되고 있는 사항에 대한 단순 현상뿐만 아니라 보다 객관적인 근거와 참고할 수 있는 정량화된 수치로 표현하면 좋다. 특정 기간의 예시를 들어 설명하면 보다 문제가 명확하게 이해된다.
ex) ~연구결과에 따르면~, 지난 동기간 대비 XX% 증가/감소,
- 범위가 넓은 문제에서 좁은 문제로 좁혀가며 문제점을 정의하는 것이 좋다.
ex) 상반기 불량율 증가 vs. A제품 생산 라인에서 불량건수 증가

2 해결방안 수립 및 주제 확정

2.1 해결방안 수립

- 식별된 문제를 해결하기 위해 필요한 정보를 도출하고 이에 따른 기대효과를 파악하여 해결방법을 수립하는 단계이다.

- 해결방안은 실현 가능해야 하며 구체적일수록 좋은데, 무엇보다도 활용가치(효용성)가 분명히 있어야 하므로 활용방안 및 기대효과가 명확히 파악되어야 한다.

2.2 후보 분석주제 확정

- 해결방안이 있는 경우 최종 분석주제를 확정하며, 후보 분석주제를 여러 개 정의하기도 한다.
→ 최종 후보 주제가 확정되기 때문에 아래와 같이 정리가 가능해야 한다.

No	분석주제명	비즈니스 문제	활용방안	기대효과
1				
2				

3 데이터 정의 및 식별

- 앞서 정의된 후보 분석주제별로 필요로 하는 데이터를 정의하고 파악하는 단계이다.
- 필요한 데이터를 영역별로 정의하고 구체적으로 어떤 컬럼(변수)이 있는지 조사한다.
- 조사하는 항목으로는 컬럼명, 생성 규칙, 관리 상태 등이 있다.
 - 1) **컬럼명** : 컴퓨터가 식별할 수 있는 수준의 항목을 식별
ex) 온도, 압력, 흡집 종류 등
 - 2) **생성 규칙** : 어떤 이벤트에 의해 생성되는지, 생성 주기는 어떻게 되는지 등
ex) 온도, 압력 등의 데이터 발생 주기(초단위, 밀리초단위 등)
 - 3) **관리 상태** : 얼마 기간의 데이터를 저장하고 관리하는지, 데이터 접근이 가능한지 등
ex) 최근 6개월의 데이터 저장 및 다운로드 가능
- 추가적으로 데이터의 품질을 평가할 수 있으면 좋다.
ex) 결측값 유/무, 결측값 비율, 분포 탐색(변동의 정도 탐색) 등

4 분석정의

4.1 변수의 역할 정의

- 분석방법론의 최상위 구분기준은 목표변수의 유/무이다. 문제해결 방법과 식별된 데이터를 기반으로 어떤 분석에 해당하는 것인지를 판단한다.
 - 1) **목표변수 有** : 예측하고자 하는 변수가 과거데이터에 있는 경우(지도학습/통제학습/Supervised Learning)
설명변수와 목표변수가 필요하다.
ex) 결함여부(유형), 회원탈퇴 여부 등
분석기법 : 회귀분석 계열, Random Forest, XGBoost, GBM, DNN 등
 - 2) **목표변수 無** : 만들고자 하는 변수가 과거데이터에 없는 경우(비지도학습/비통제학습/Unsupervised Learning)
목표변수는 필요없다.
ex) 유사성을 기준으로 개체를 그룹화하는 경우
분석기법 : 군집분석(K-Means), 이상치 탐지(DBScan, LoF) 등
- 지도학습인가 비지도학습인가에 따라 필요로하는 변수의 역할이 결정된다.
 - ex1) 공정 각 단계의 조건에 따라 결함여부의 패턴 분석(지도학습)
 - 설명변수 : 공정 각 단계에서의 조건 변수들
 - 목표변수 : 결함 유형, 여부
 - ex2) 공정 각 단계의 조건 조합의 이상패턴 분석(비지도학습)
 - 설명변수 : 공정 각 단계에서의 조건 변수들

4.2 분석기간 정의

4.2.1 목표변수 기간정의(지도학습의 경우)

- 목표변수가 있는 경우 얼마기간에 발생한 해당 목표변수에 주목할 것인가를 정

의한다.

ex) 최근 6개월간의 불량 상품을 분석

4.2.2 활용기간 정의(지도학습의 경우)

- 분석결과를 적용하는데 있어, 최종 Action을 취해야 하는 시점과 분석결과를 제 공받아야 하는 시점의 term을 결정하는 단계이다.

ex) 고객 이탈방지 캠페인 실행 단계에 타겟팅하는 시점

불량율을 감소시키기 위한 최적 래시피를 적용시킬 수 있는 시점

4.2.3 설명변수기간 정의

- 지도학습의 경우, 목표변수의 상태를 얼마기간의 설명변수를 가지고 분석할 것 인지를 정의한다. 이 때, 마케팅 활용기간을 제외한 과거 시점으로 정의한다.
- 비지도학습의 경우, 얼마 기간의 설명변수의 상태를 기준으로 분석할 것인지를 정의한다.

4.3 평가기준 정의

- 앞서 정의한 분석방법에 따라 분석결과와 신뢰성과 적용가능한 수준의 품질 등을 평가하기위한 적절한 기준을 정의하는 단계이다.
- 평가 방법과 더불어 추가적으로 KPI가 있는 경우 기술하기도 한다.
- 지도학습의 경우 과거 정답이 있기 때문에 차이(맞고 틀림)를 정량적으로 평가한다.
ex) 평균절대오차율(MAPE), 평균제곱오차(MSE), 정확율(precision), 포괄율(recall), F1-Score 등
- 비지도학습의 경우 과거 정답이 없기 때문에 약간의 정성적 평가가 포함되는 것이 보통이다.
ex) 기초통계량(그룹별 평균, 중위수 등의 분포)의 탐색

5 분석결과 활용 방안

- 어떤 업무에 어떤 시점에 어떤 방법으로 적용할 것인지를 결정하고 [구체화](#)하여 기술한다.
- 해결방안 단계에서 작성한 활용방안보다는 구체적으로 시스템 구성도, 적용 주기, 취해야 하는 Action 등의 [As-Is vs. To-Be](#) 정보가 들어가야 한다.
- 추후 진행되어야 할 진단 추진일정이 들어가면 좋다.