

학습내용

- 1. 빅데이터 분석 프로세스
- 2. 빅데이터 분석 프로세스 적용 사례

학습목표

1. 데이터 분석 프로세스의 과정을 설명할 수 있다.

1주차. 빅데이터 분석 개요 1강. 강의 소개, 빅데이터 기초 개념 강의소개 및 데이터의 개요



빅데이터 분석 프로세스

데이터 분석 프로세스

문제 정의 데이터 수집 데이터 전처리 데이터 모델링 시각화 및 탐색

다양한 프로세스가 존재하지만 일반적으로 5단계로 진행됨



➡ 문제 정의 단계

- 분석하고자 하는 분야를 이해
- 해결해야 하는 문제를 객관적이고 구체적으로 명시





➡ 데이터 수집 단계

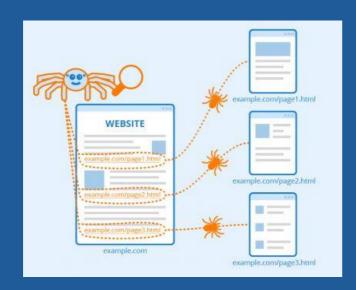
- 분석에 필요한 데이터 조건을 정의
- 데이터의 소재를 파악하고 확보



데이터수집



- 크롤링
 - 뉴스, 블로그, SNS, 카페 등 인터넷상에서 존재하는 웹문서의 정보를 수집하는 것



〈출처: seobility(연도미상). Data crawling 이미지.https://www.seobility.net/〉



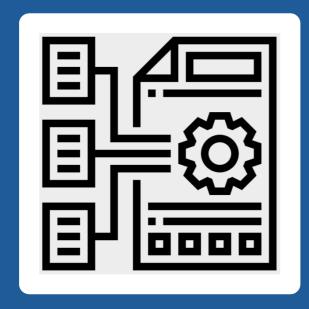
- Open API
 - 공급운영주체가 정보 및 데이터를 제공하기 위해 개발자와 사용자에게 오픈하는 것



〈출처: Flaticon(연도미상).api 이미지.https://www.flaticon.com/〉



- Log Aggregator
 - 각종 로그 데이터를 수집하는 오픈 소스 기술



〈출처: Flaticon(연도미상).data aggregator 이미지.https://www.flaticon.com/〉



- RDB Aggregator
 - 관계형 DB에서 정형 데이터를 수집해 Hadoop이나 NoSQL에 저장하는 오픈 소스 기술



〈출처: Flaticon(연도미상).Hierarchical structure 이미지. https://www.flaticon.com/〉



데이터 전처리 단계

- 수집한 데이터에 존재하는 결측치나 null값 등을 수정 및 보완
- 결측치 대체 및 이상치 존재 확인 및 제거를 하는 데이터 클리닝
- 분산된 데이터를 일관성 있는 형태로 변환하는 데이터 통합
- 적은 양으로도 전체 데이터 집합을 얻는 과정인 데이터 축소
- 그 외 다양한 데이터 전처리 방법이 존재



🥎 데이터 모델링 단계

- 데이터를 활용해 비즈니스에 맞는 모델을 만들어 내는 단계
- 통계 기반, 데이터마이닝 기반, 머신러닝, 딥러닝 등 다양한 모델방법이 존재



에이터 모델링 단계 (통계 기반)

- 데이터로부터, 응용 수학의 기법을 이용해 수치상의 성질, 규칙성 또는 불규칙성을 찾아냄
- 특정 모집단에서 구한 표본을 가지고 그 모집단의 특성 및 가능성 등을 추론해내는 추론 통계가 존재
- 측정이나 실험에서 수집한 자료의 정리, 포현, 예약, 해석 등을 통해 자료의 특성을 규명하는 기술 통계가 존재



데이터 모델링 단계 (데이터마이닝 기반)

데이터 마이닝 적용 기법	설명
Association (연관성)	데이터 간 연관성 발견
Sequence (연속성)	순서에 따른 이력을 분석해 이후의 가능성 예측
Classification (분류)	수집된 데이터의 패턴 및 속성으로 결합해 예측
Clustering (군집화)	특징 및 속성의 유사성으로 데이터 군집



데이터 모델링 단계 (머신러닝)

- 머신러닝은 경험을 통해 자동으로 개선하는 컴퓨터 알고리즘의 연구
- 데이터 마이닝과 도메인이 교차되는 경우가 있음
- 머신러닝은 예측에 초점, 데이터마이닝은 탐색적 분석
- 지도학습, 자율학습, 준지도학습, 강화 학습의 유형 존재



데이터 모델링 단계 (딥러닝)

- 딥 러닝은 인공신경망에서 발전한 형태, 은닉층이 2개 이상인 분석 알고리즘
- 딥 러닝으로 훈련된 시스템의 이미지 인식 능력은 이미 인간을 능가하게 됨
- 구글의 Alphago는 바둑의 기초를 배워 강화학습을 통해 스스로 학습



시각화 및 탐색

- 다양한 유형의 데이터 시각화
- 문제를 해결하기 위해 인사이트 도출

[1854년 브로드 스트리트의 콜레라] 콜레라의 피해를 가장 많이 입은 세대들이 모두 같은 우물을 식수로 사용하는 것을 발견하여, 질환과 우물간 관계를 밝혀내는 데 큰 도움



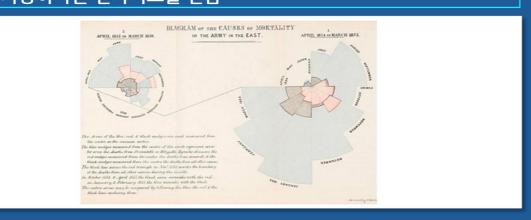
〈출처: 위키피디아 커먼스.Diagram of the causes of mortality in the army 이미지. https://commons.Wikimedia.org/〉



시각화 및 탐색

- 다양한 유형의 데이터 시각화
- 문제를 해결하기 위해 인사이트 도출

[크림전쟁 군인 사망률] 나이팅게일은 데이터 시각화를 통해, 크림 전쟁 중 높은 군인 사망률은 전장이 아닌 열악한 병원 상태란 걸 밝히고, 예방이 가능하다는 인사이트를 얻음



〈출처: 위키피디아 커먼스.Diagram of the causes of mortality in the army 이미지. https://commons.Wikimedia.org/〉

빅데이터 분석 프로세스 적용 사례

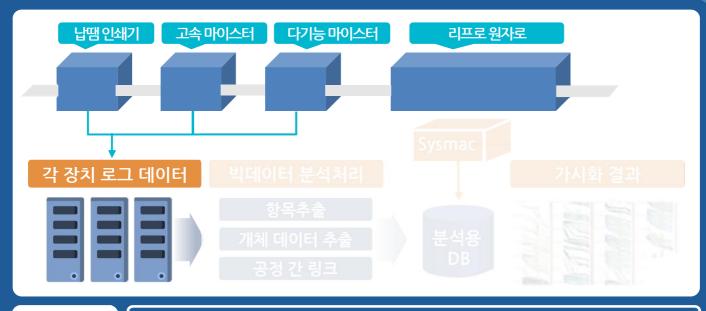


→ 오므론의 불량 원인 분석 개선 사례



문제 정의 플랫폼 도입 이전에는 공정상 불량 원인 분석에 6명의 숙련자가 필요



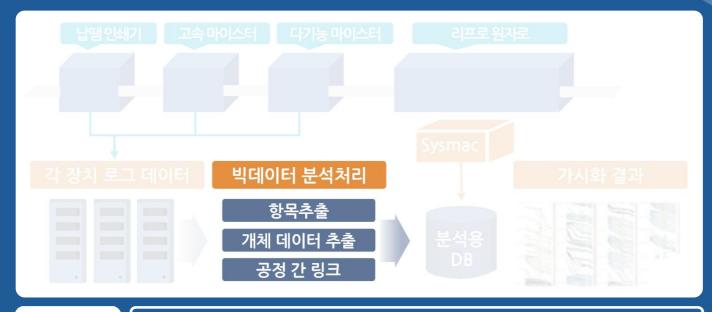


데이터 수집

인쇄기, 원자로에 해당하는 각 설비마다 발생하는 온도, 생산량 등의 로그데이터를 수집



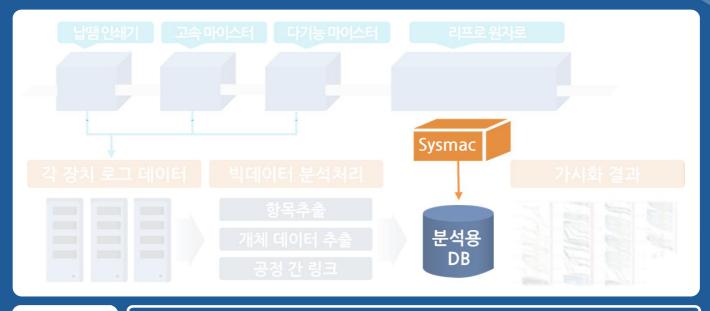
오므론의 불량 원인 분석 개선 사례



데이터 전처리 수집된 로그데이터를 분석하기 위해 시계열 데이터 변환, 데이터 프레임 구조화 등 처리 작업



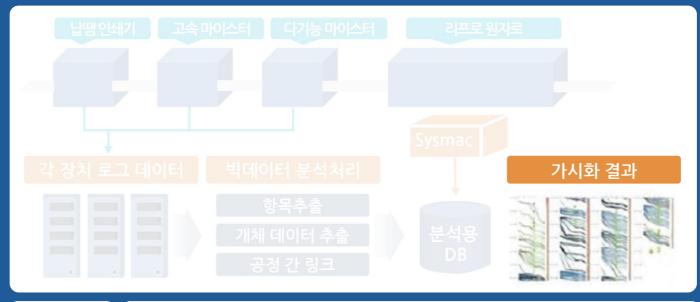
오므론의 불량 원인 분석 개선 사례



데이터 모델링 전처리된 로그 데이터를 활용하여 공정 상황 모니터링 및 이상 감지 시스템 구축



오므론의 불량 원인 분석 개선 사례



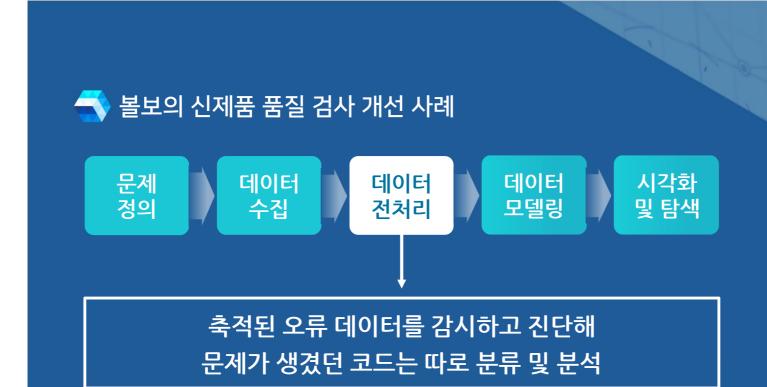
시각화 및 결론도출

데이터 분석 결과를 시각화하여 누구나 쉽게 실시간으로 각 공정의 정상작동 여부 파악 가능, 도입 후 필요 인력 1명으로 축소



신제품의 초기 불량 발견을 위해 필요한 생산대수 및 주행거리 측정 시간이 과다함







➡ 볼보의 신제품 품질 검사 개선 사례

데이터 데이터 전처리 문제 데이터 시각화 모델링 및 탐색 정의 수집

수집된 운행 정보, 운전자의 운전 습관 등을 분석해 불량감지, 소비자 보상 규모의 정확한 측정에 활용



➡ 볼보의 신제품 품질 검사 개선 사례

데이터 데이터 문제 데이터 시각화 전처리 모델링 및 탐색 정의 수집

신제품의 초기 불량 발견을 위해 필요한 생산대수가 기존 5만대 수준에서 2천대 수준으로 단축, 차량 주행거리 측정 시간이 기존 2시간에서 5분으로 단축

평가하기(1)

1번	문항	데이터 모델링 단계에서 "특징 및 속성의 유사성으로 데이터 군집"을 뜻하는 용어는 분류(Classification)이다.
	정답	X
	해설	특징 및 속성의 유사성으로 데이터 군집은 군집화(Clustering)이다.

평가하기(2)

1번	문항	볼보는 엔진, 변속기, 브레이크 등 자동차 부품에 대한 정보를 수기로 기록하여 데이터를 수집했다.
	정답	X
	해설	엔진, 변속기, 브레이크 등 자동차 부품에 센서를 탑재해 운행시 발생하는 차량 데이터 수집했다.

정리하기

1. 빅데이터 분석 프로세스

- 일반적 빅데이터 분석 프로세스는 문제 정의, 데이터 수집, 데이터 전처리, 데이터 모델링, 시각화 및 탐색 총 5단계

2. 빅데이터 분석 프로세스 적용 사례

- 오므론과 볼보는 빅데이터 기술을 활용해 그 프로세스에 따라 전체 공정관리 시스템을 개선해 단위 노동자당 생산성 향상을 도모

다음시간에는…

2주차. 빅데이터 분석 도구 소개

1강. 빅데이터 분석 도구 소개 및 활용 사례

에 대해 학습해 보겠습니다.

참고문헌

- 구글 이미지.seobility(연도미상). Data crawling 이미지.2021년 7월 12일 인용. https://www.seobility.net/en/wiki/Search_Engine_Crawlers. 1페이지.
- Flaticon(연도미상).api 이미지. 2021년 7월 12일 인용. https://www.flaticon.com/kr/free-icon/api_627558 . 1페이지.
- Flaticon(연도미상).data aggregator 이미지. 2021년 7월 12일 인용. https://www.flaticon.com/free-. 1페이지. icon/aggregate_2519263?term=data%20aggregator&page=1&position=2&page=1&position=2&related_id=2519263.1페이지.
- Flaticon(연도미상).Hierarchical structure 이미지. 2021년 7월 12일 인용. https://www.flaticon.com/free-icon/hierarchy-structure_2857303?term=structured%20data&page=1&position=1&page=1&position=1&related_id=2857303&origin=search . 1페이지.