

텍스트 마이닝과 데이터 마이닝

Part 07. 데이터 마이닝

정 정 민

Chapter 18. 데이터 마이닝이란?

1. 데이터 마이닝 소개
2. 데이터 마이닝 사례

데이터 마이닝 소개

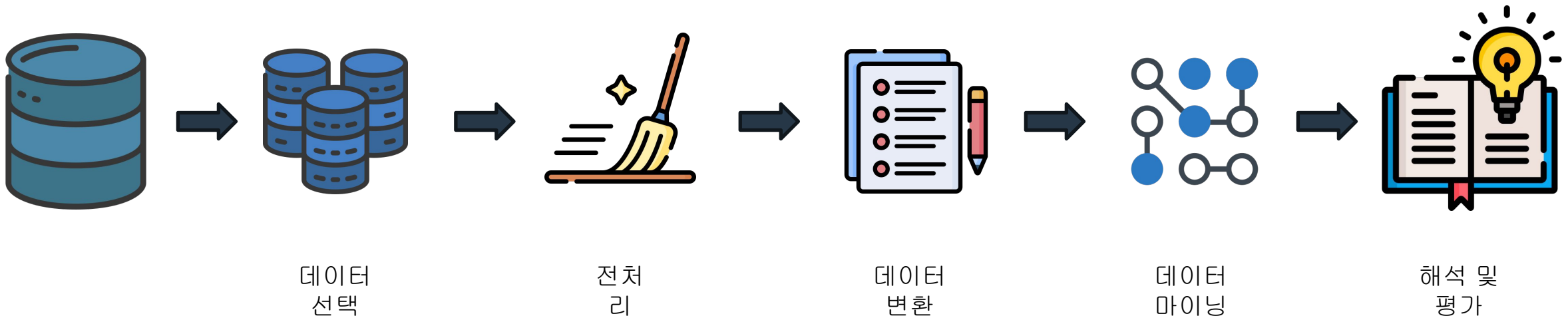
[RECAP] 마이닝(Mining, 채굴)

- 광물에서 귀중한 금속이나 광석을 **채굴**하는 작업
- 대량의 광물에서 소량의 귀중한 자원을 발견하고 추출하기 위해서는
- 광물을 면밀히 살펴봐야 함
- 비슷하게, **대량의 데이터(광물)에서**
- **유용한 정보와 패턴(자원)을 찾기 위해 채굴이 필요**
- 이렇게 **채굴된 정보와 패턴으로 통찰력을 얻고, 의사 결정을 진행**
- 정보를 얻고자 하는 원천 데이터로
- 텍스트 데이터가 주어진다면 > **텍스트 마이닝**
- 이미지 데이터가 주어진다면 > **이미지 마이닝**
- 특정되지 않은 일반적인 데이터라면 > **데이터 마이닝!**



데이터 마이닝이란?

- 특정 데이터에 한정하지 않고
- 대용량의 데이터 내에 존재하는 관계, 패턴, 규칙을 탐색
- 이로부터 유용한 지식을 추출하는 과정
- 일반적으로 아래의 과정으로 진행 됨



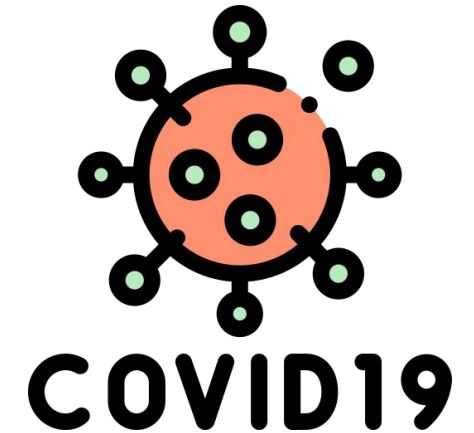
데이터 마이닝의 중요성

- 데이터 마이닝은 아래의 과정에서 / 과정을 위해 사용
- **의사결정 강화**
 - 비지니스와 조직에서 의사 결정을 강화 (당위성을 확인)
 - 데이터로부터의 통찰력으로 전략 계획 수립 진행
- **효율성 증대**
 - 조직 운영 효율성 상승 (비용 절감, 리스크 관리)
- **고객 이해**
 - 고객의 요구 행동 이해
 - 맞춤형 서비스 제품 개발
- **시장 동향 예측**

데이터 마이닝 사례

코로나 19 데이터 마이닝

- 세계적 수준에서 경제, 사회, 보건 전반에 걸쳐 큰 영향을 미친 코로나 19 상황에서 데이터 마이닝이 활용
- 전반적인 흐름을 추적하기 위해
- **감염자 수, 회복률, 백신 접종 현황, 인구 통계학 정보** 등이 사용
- 위 정보를 이용해
- 전파 속도, 패턴, 감염 차이, 백신의 효과성 등을 확인
- 그리고, 이러한 정보를 바탕으로
 - 공중 보건 정책의 영향력 파악 및 이에 대한 의사 결정
 - 백신의 효과 확인 및 위급 지역 우선 배포
 - 변이 바이러스 예측 및 추적
 - 등 **의사 결정에 활용**



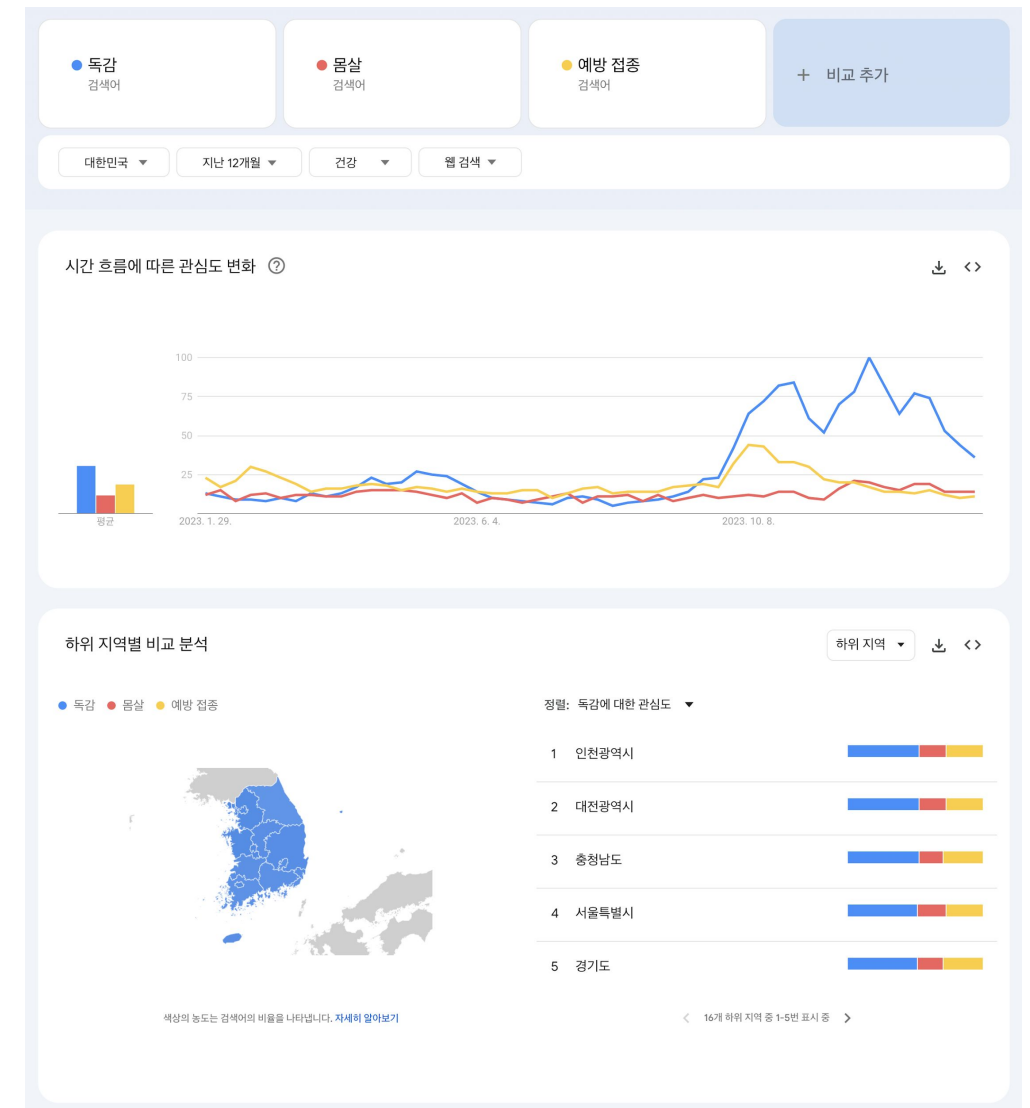
실시간 GPS 데이터를 통한 최적 교통 정보 도출

- GPS란, 위치 추적 기술로 얻은 지리적 위치 정보
- 실시간 차량의 위치 데이터, 움직이는 물체의 속도, 이동 방향 등의 데이터를 사용
- 이러한 정보를 바탕으로
 - 전반적인 차량의 교통 패턴, 교통량, 시간대별 교통 상황 등
 - 사용자 행동 패턴
- 이를 사용한다면
 - 어떤 경로로 주행을 할지
 - 교통 체계 최적화 및 체증 완화 전략 개발
 - 대규모 행사와 같은 신속한 교통 대응 계획 수립
 - 등의 의사 결정에 사용



구글 검색어 기반 독감 예보

- 감기, 독감과 같은 질병의 유행성을 파악
- 전통적인 감시 방법의 한계 및 실시간 데이터 활용 필요성
- 구글에 검색하는 **유행성 질병 검색어 데이터** 활용
- 특정 조건의 검색어 정보 활용
 - **시간대별, 지역별**
- 독감의 활동성 비교적 정확하게 예측
 - 독감 발생률 증감
 - 지역별 위험도 평가
- 의사 결정 과정에도 활용
 - 공중 보건 예방 조치
 - 의료 기관 및 약국 자원 배분
 - 방역 대책 수립



DNA 데이터를 통한 범죄자 추적

- 현대 범죄 수사에서 사용하는 첨단 분석 방법
- 여러 사람의 DNA 데이터베이스와
- 수집한 비교 대상의 DNA 의 유사도를 분석
- 빠른 범죄자 추적 및 일치 여부 확인
- 정확한 범인 확인 과정에서 사용되며
- 증거 자료로서의 역할



E.O.D