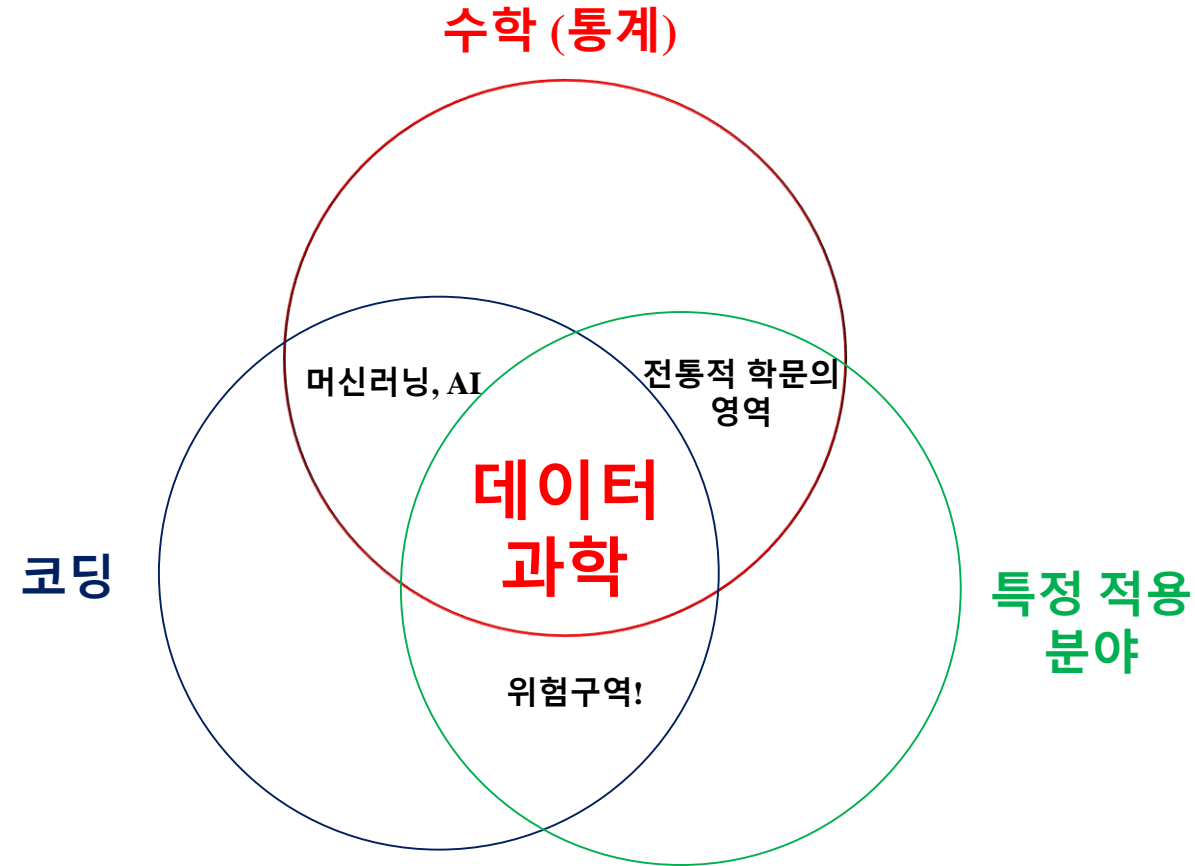


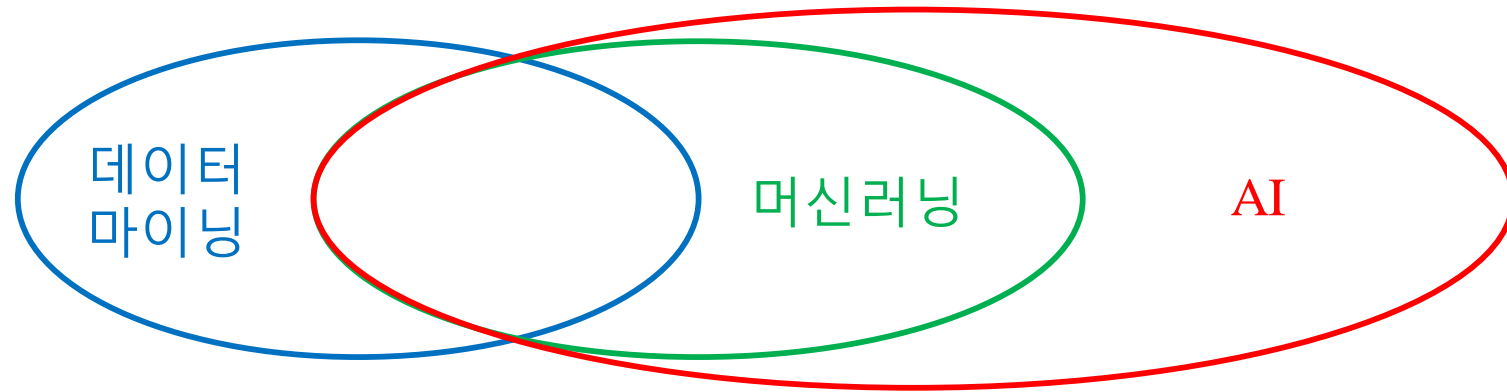
Python 머신러닝

데이터 과학?



→ 데이터 과학은 빅데이터, 머신러닝, AI (인공지능)을 다루는 학문이다.

Overview



기계학습?

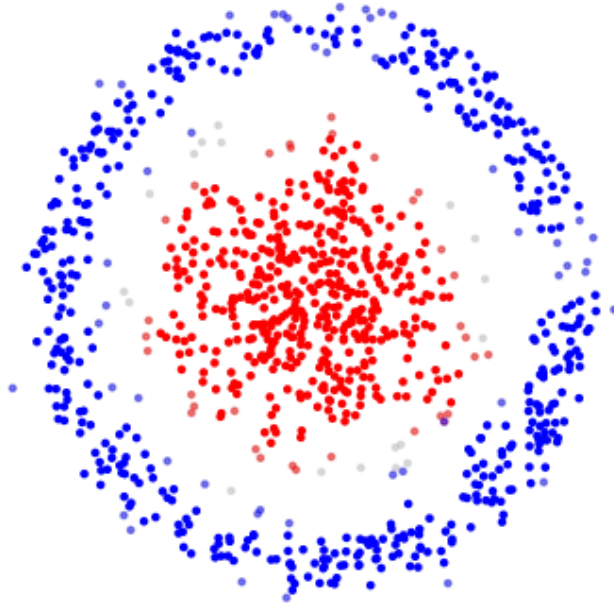
러닝머신 ???

머신러닝 ! (Machine Learning)

→ 통계 모형과 데이터를 사용한 학습과 예측을 의미한다.

기계학습 : 유형

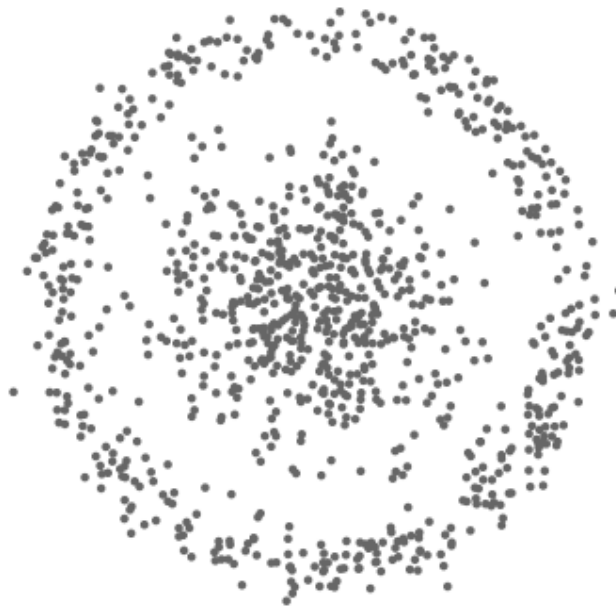
지도학습 (supervised learning):



- 학습목표와 내용이 분명하다. 학습패턴은 주어진다.

기계학습 : 유형

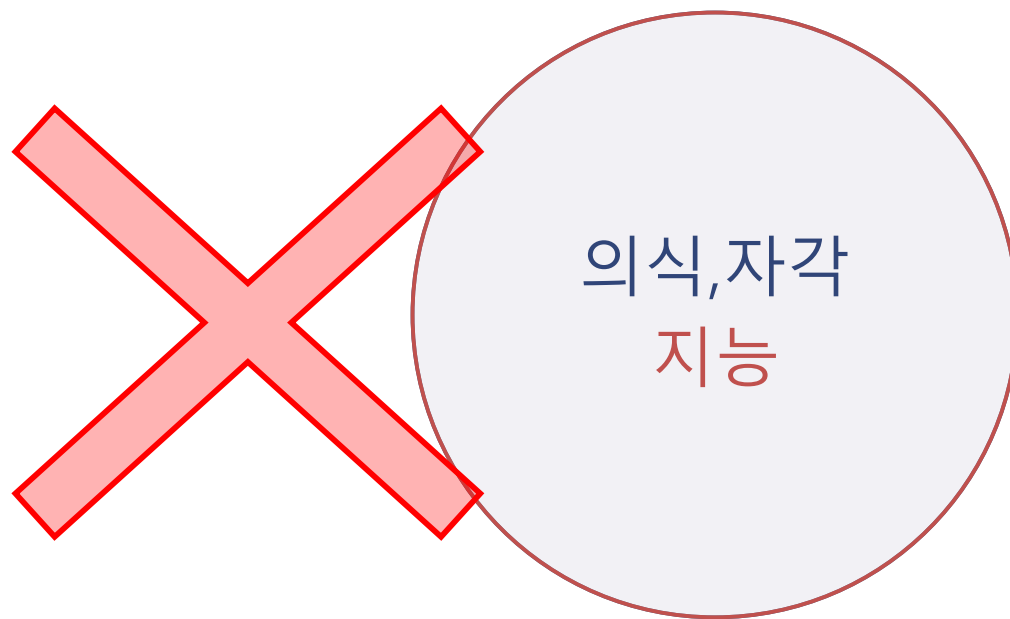
비지도학습 (unsupervised learning):



- 학습목표와 내용이 불분명하다. 스스로 패턴을 찾아 내어야 한다.

인공지능?

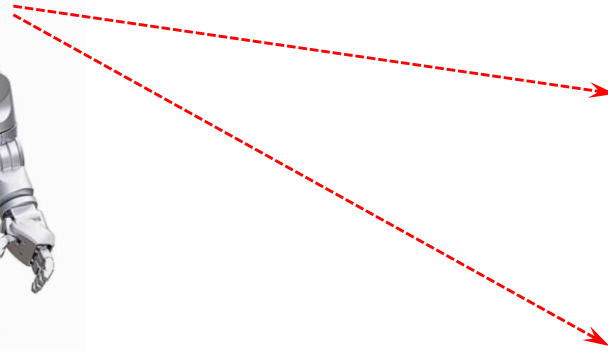
의식과 지능의 분리:



인공지능

인공지능

??
동물?? 고양이??



신문 기사 (2017-07-01)

▶ 데일리한국 > 경제 > 최신기사

국내 개발 인공지능(AI), 금융권서 가장 많이 쓴다

금융분야 26개·국방 23개·미디어 22개·유통 의료 17개 순

조진수 기자 rokmc4390@hankooki.com



사진=유토이미지

[데일리한국 조진수 기자] **국내** 개발 인공지능(AI) **기술**은 금융 분야에서 가장 많이 쓰이는 것으로 밝혀졌다.

30일 소프트웨어 정책연구소가 발간한 '국내 지능정보산업 실태에 대한 연구' 보고서에 의하면 35개 인공지능 **기술 개발기업**의 95개 제품(중복 포함 238개) 중 금융분야에 쓰인 경우가 26개로 가장 높았다.

- 국내 개발 인공지능 (AI), 금융권에서 사용 1위.
- 고객·민원상담 (20.8%), 마케팅 최적화 (18.8%), 이상거래 탐지 (15.8%).
- 머신러닝 (19.8%), 딥러닝·신경망 (18.0%), 자연어 처리 (13.5%), 상황인지 (10.8%).

미완성의 아름다움

스페인 바르셀로나에 위치한 Sagrada Familia 성당



- 건축가 가우디가 1882년 건축 시작.
- 136년째 건축 중. 2028년경 완공 예정.

미완성의 아름다움

- 데이터 분석은 시간과의 싸움입니다.
- 대다수의 경우에는 ‘정답’이 없습니다.
- 100% ‘정답’을 찾기 보다는 80% ‘**쓸만한**’ 답을 빠른 시간에 찾아내는 것이 목표입니다.

데이터 사이언스 도구

데이터 사이언스 도구 (tool):

- 데이터 **관리용** 도구: 입력, 저장, 보존.

예) MySQL, MongoDB, Hadoop, Excel, 등.

- 데이터 **분석용** 도구: 변형, 분석, 패턴학습.

→ 머신러닝을 가능케 하는 도구.

데이터 사이언스 도구

데이터 분석용 도구:

- R 프로그래밍 언어: caret, dplyr, tm, 등의 패키지.
- Python 프로그래밍 언어: pandas, numpy, scipy, sklearn, keras, theanos, 등의 패키지.
- 클라우드: 구글, 아마존, MS (Azure).

빅데이터 : 출현 배경

빅데이터의 출현 배경:

- 산업계: 양 ↔ 질 변환 법칙.
- 학계: 거대 데이터 활용 과학 확산. 예) 인간 게놈 프로젝트.
- 관련기술의 발전: +디지털화의 가속, 저장 기술의 발달과 가격 하락.
+인터넷, 모바일, 클라우드 등 관련기술의 보편화.

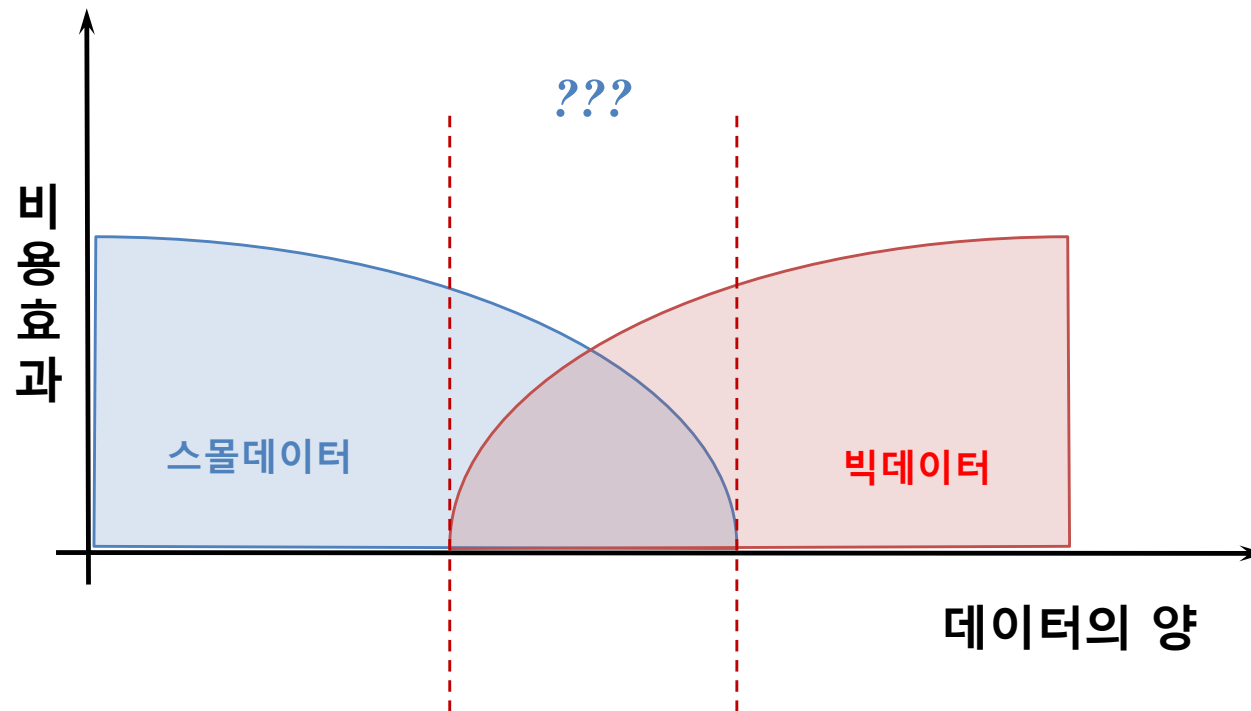
빅데이터 vs 스몰데이터

그러면, ‘스몰데이터’란?

- 사람이 직관적으로 이해할 수 있는 정도의 크기와 형태.
- 쉽게 접근 (accessible), 이해 (understandable), 실행 (actionable)할 수 있는 데이터.
- 일상 생활과 업무에서 자주 접하는 데이터의 유형.

빅데이터 vs 스몰데이터 : 비용효과

비용효과 비교:



빅데이터 시대의 본질적인 변화

사전처리에서 사후처리로 포커스 이동:

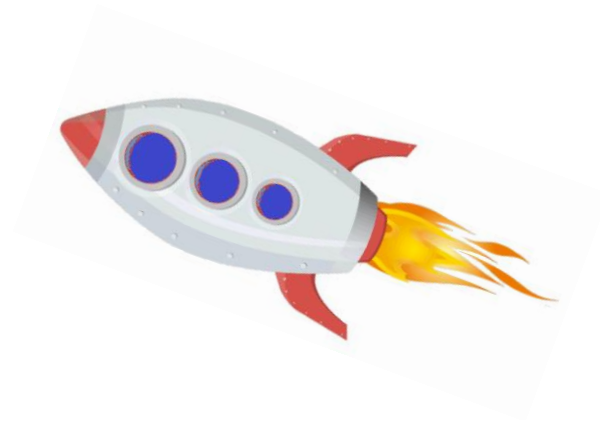
필요한 정보만 수집하는 시스템이 아니고 가능한 한 많은 데이터를 수집하여서 그 데이터를 다양한 방법으로 조합해 숨은 정보를 찾아낸다.



빅데이터 시대의 본질적인 변화

표본조사에서 전수조사로:

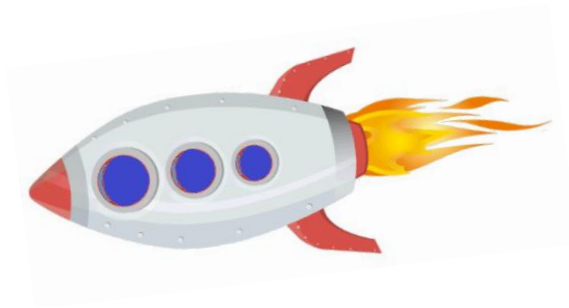
표본조사를 통해서 나타나지 않는 패턴이나 정보를 알아낼 수 있다.



빅데이터 시대의 본질적인 변화

질보다는 양:

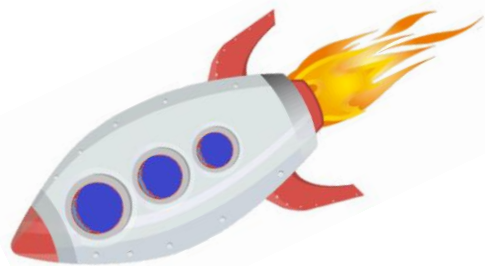
정보가 축적될 수록 오류 정보보다는 양질의 정보가 많아져서 전체적으로 좋은 결과 산출에 긍정적 영향을 미친다.



빅데이터 시대의 본질적인 변화

인과관계에서 상관관계로:

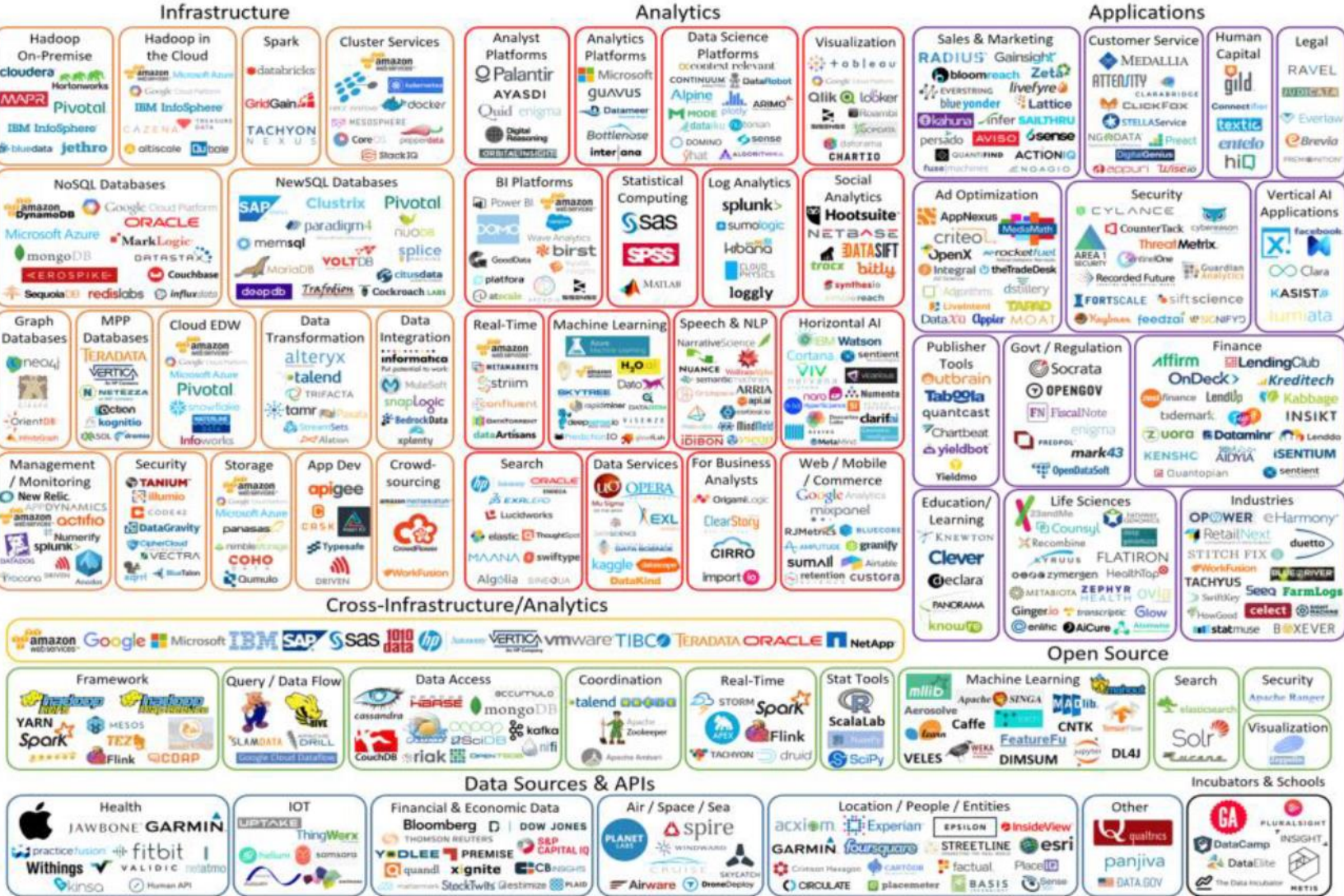
상관관계 분석이 주는 통찰력이 인과관계에 기초해서 할 수 있는 미래예측을 압도해가는 시대가 도래.



빅데이터 Landscape : 2012년도



빅데이터 Landscape : 2016년도



데이터 분석 기술: 정보 검색 (IR)

- 컴퓨터 시스템을 사용하여 데이터를 색인하고 주제와 관련된 자료를 빠르게 찾아 분석하는 기술.

데이터 분석 기술: 정보 수집 (Crawling)

- 기존의 웹 검색을 위한 수집기보다 매우 발전된 정보 수집 기술이 필요함.
- 예를 들어서, 트위터와 같은 SNS의 실시간 데이터 수집을 위한 스트림 처리 기술.

데이터 분석 기술

데이터 분석 기술: 기계 학습 (Machine Learning)

- 학습 데이터를 사용하여 통계 모델을 만들고 예측에 적용함.

데이터 분석 기술: 소셜 네트워크 분석

- SNS의 연결 구조 및 강도 등을 바탕으로 사용자의 인지도 및 영향력 측정.

데이터 분석 기술: 자연언어 처리 (NLP)

- 인간의 언어를 컴퓨터로 처리하기 위한 기술.
- 형태소 분석, 구문 분석, 개체명 인식 등의 기술을 포함.

데이터 분석 기술

데이터 분석 기술: 텍스트 마이닝 (Text mining)

- 비정형 텍스트 데이터에서 통계적, 연관적 특성을 추출하는 기술.

데이터 분석 기술: 클라우드 컴퓨팅 & NoSQL

- 빅데이터의 저장과 관리 운영을 위한 기술.
- NoSQL은 전통적인 관계형 데이터베이스의 틀을 벗어난 일관성 모델을 이용.
예) MongoDB, Hadoop, Hbase, Cassandra, 등.

데이터 분석 기술

데이터 분석 기술: 통계 기술 (Statistics)

- 빅데이터의 통계적 의미를 찾고, 그 패턴을 분석하기 위해서는 강력한 통계 기능 필요.

데이터 분석 기술: 시각화 (Visualization)

- 분석된 결과의 통찰력 있는 이해를 돕기 위한 기술.

데이터 분석의 유형

데이터 분석의 유형: 서술적 분석 (Descriptive Analytics)

- 주어진 상황에서 어떤 일이 벌어졌는지 설명하는데 사용.
- 예) 무슨 일이 일어났는가?, 누가 우리의 고객인가?, 고객은 유형별 어떻게 분류할 수 있는가?
- 시각화, 군집분석 (clustering analysis), 등 사용.

데이터 분석의 유형

데이터 분석의 유형: 진단형 분석 (Diagnostic Analysis)

- 어떤 이유로 특정 현상이 발생한지 원인을 밝히는 것이 주된 목적.
- 예) 왜 매달 이탈 고객이 늘어가는가?
- 군집분석 (clustering analysis), 의사결정트리, 등 사용.

데이터 분석의 유형: 예측 분석 (Predictive Analytics)

- 과거 데이터를 사용하여 미래에 대한 전망 제시.
- 예) 신용 카드 거래의 부정 여부, 고객이 높은 요금제로 전환할 확률 예측.
- 회귀분석, 몬테카를로 시뮬레이션, 의사결정트리, 랜덤포레스트, 인공신경망, 등 사용.

데이터 분석의 유형: 처방적 분석 (Prescriptive Analytics)

- 구체적인 실행 방안과 예상 효과 제시.
- 업무 방법과 예측 모형의 결합.
- 게임이론, 몬테카를로 시뮬레이션, 의사결정트리, 선형 및 비선형 프로그래밍, 등 사용.

데이터 분석을 잘 하려면...



먼저, **선입관**과 **편견**을 버린다.

데이터 분석을 잘 하려면...



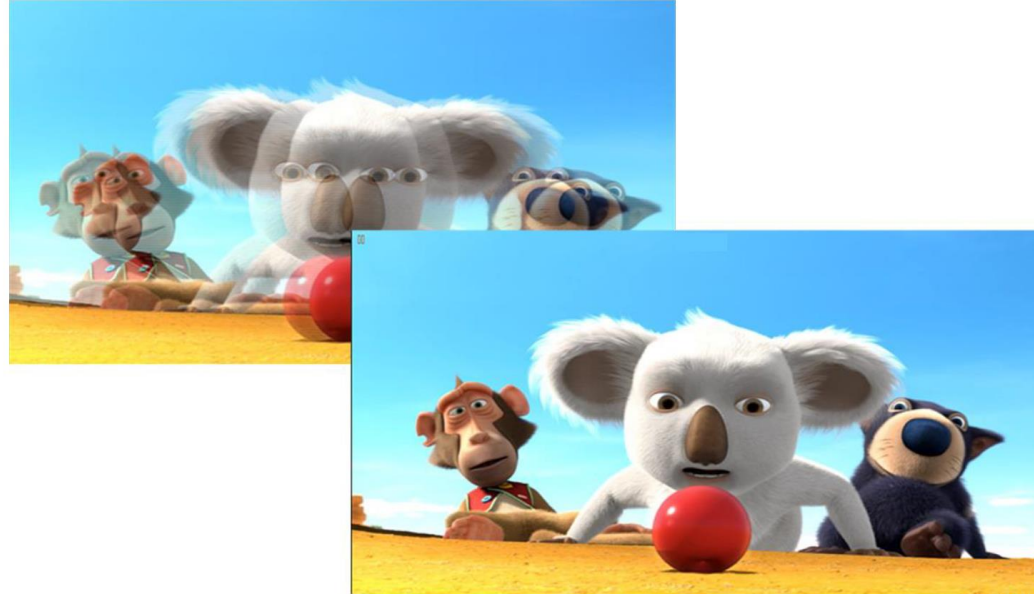
갈색 고양이???

데이터 분석을 잘 하려면...



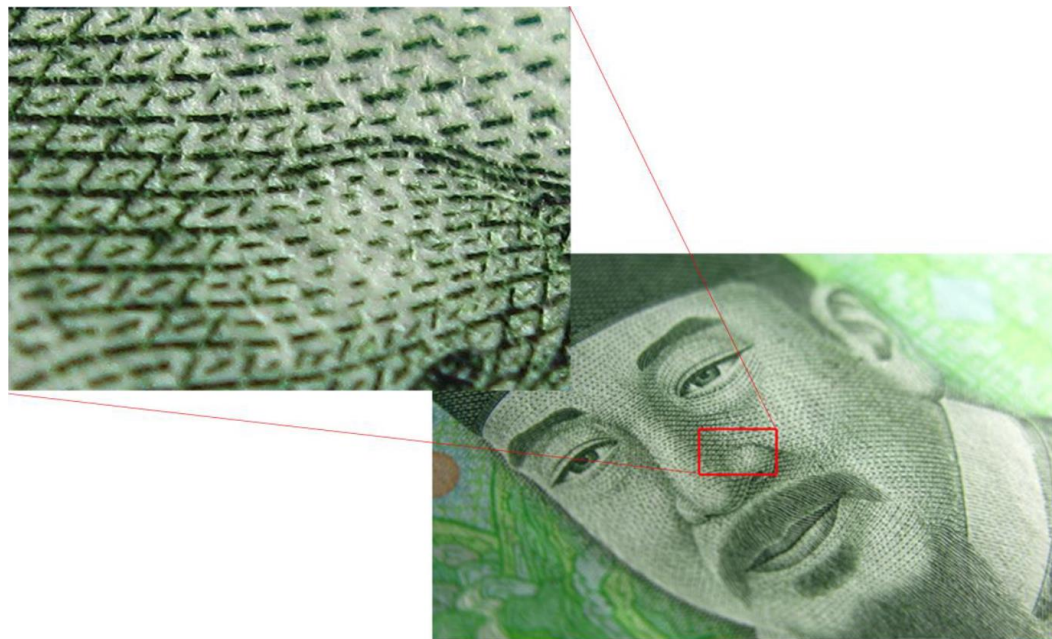
당장 눈앞에 보이는 것이 사실이 아닐 수도 있다.

데이터 분석을 잘 하려면...



사물을 명확하게 꿰뚫어 보려는 노력!

데이터 분석을 잘 하려면...



가끔은 디테일에 신경을 쓰도록 한다.

데이터 분석을 잘 하려면...



꾸준한 노력의 결과로 귀중한 통찰력을 얻는다.