

Python 분석 기초

모듈 - 1

강사: 장순용 박사

강사 소개



장순용 박사

현) 휘안리서치 연구소장: 빅데이터/인공지능 컨설팅.

약력) 일리노이대(UIUC) 물리학 PhD. 워싱턴大, 오하이오 주립大 연구원.

저서) SCI급 국제 저널에 다수의 논문 출간. 알고리즘 트레이딩 시리즈 (총 3권). Data Science 101.

연락처) sychang1@gmail.com

강사 소개

- 취업준비생 대상 교육:

- 아시아 경제 청년취업 아카데미: 2019/7월 ~ 8월, 2019/11월~12월.

- 4차 산업혁명 선도인력 양성 4기 메인 강사: 2018/11월 ~ 2019/5월 (멀티캠퍼스).

- 경기 빅데이터: 2018/7월 ~ 2018/8월 (가천대, 명지대).

- 기업 출강 및 컨설팅:

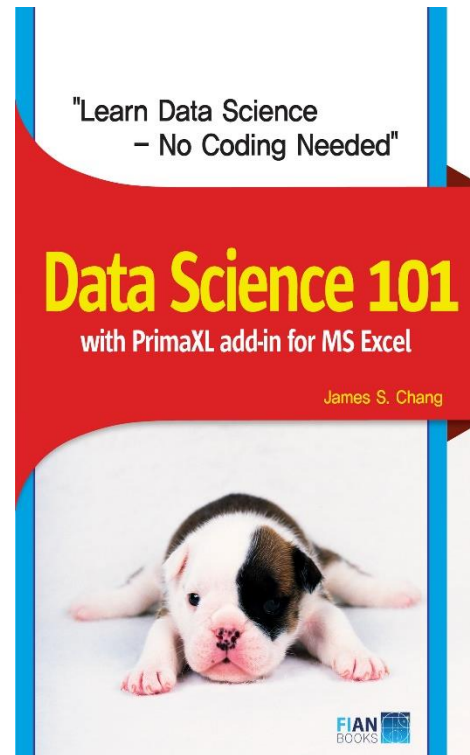
- 삼성전자, KT, 한국 IBM, 삼성 SDS, POSCO, LG U+, LG CNS, KB 국민은행, 대구은행, Investec (영국) 등.

- 교육기관:

- 아시아경제, 멀티캠퍼스, 패스트 캠퍼스, 비트아카데미, 한국 생산성 본부 (KPC), 한국기술대, 등.

저서 소개

“Data Science 101” (아마존 이북)



저서 소개

“알고리즘 트레이딩” (총 3권)

제 1권: 머신러닝, 시계열 예측, 트레이딩 전략.

제 2권: 리스크 관리, 페어트레이딩.

제 3권: 선물과 옵션. 트레이딩 전략.

저자 : 장순용.



금융 공학부터 데이터 수집/분석, 실제 적용까지

금융 공학 / 퀀트 올인원 패키지 Online.



 **fast campus**

교육과정의 키 포인트

- ✓ **빅데이터 분석**의 이해와 활용.
- ✓ **인공지능** (AI) 알고리즘의 이해와 활용.
- ✓ **파이썬** 주로 사용.
- ✓ **금융 프로젝트** 도출과 수행 능력 함양.

기본에 충실한 교육!



1. 빅데이터 금융 사이언티스트 개요.

1.1. 빅데이터 개요.

1.2. 데이터의 이해.

1.3. 빅데이터의 가치와 인사이트.

1.4. 데이터 분석의 이해.

1.5. 빅데이터와 데이터 과학의 미래.

1.6. 데이터 사이언스와 금융.

1.7. 활용 사례 (금융 & 비금융).

시대적 배경

지금은 소위 빅데이터의 시대라고 하는데....

데이터의 홍수로군!



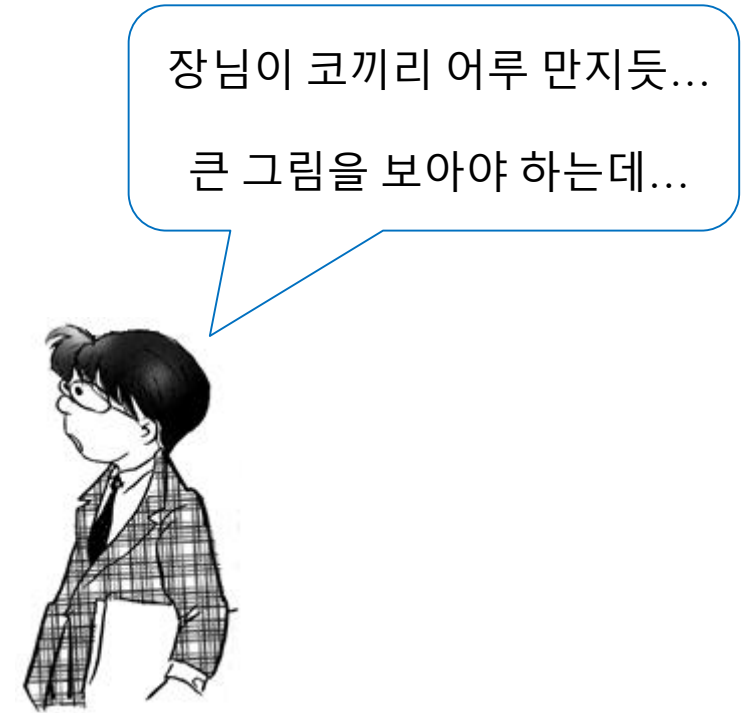
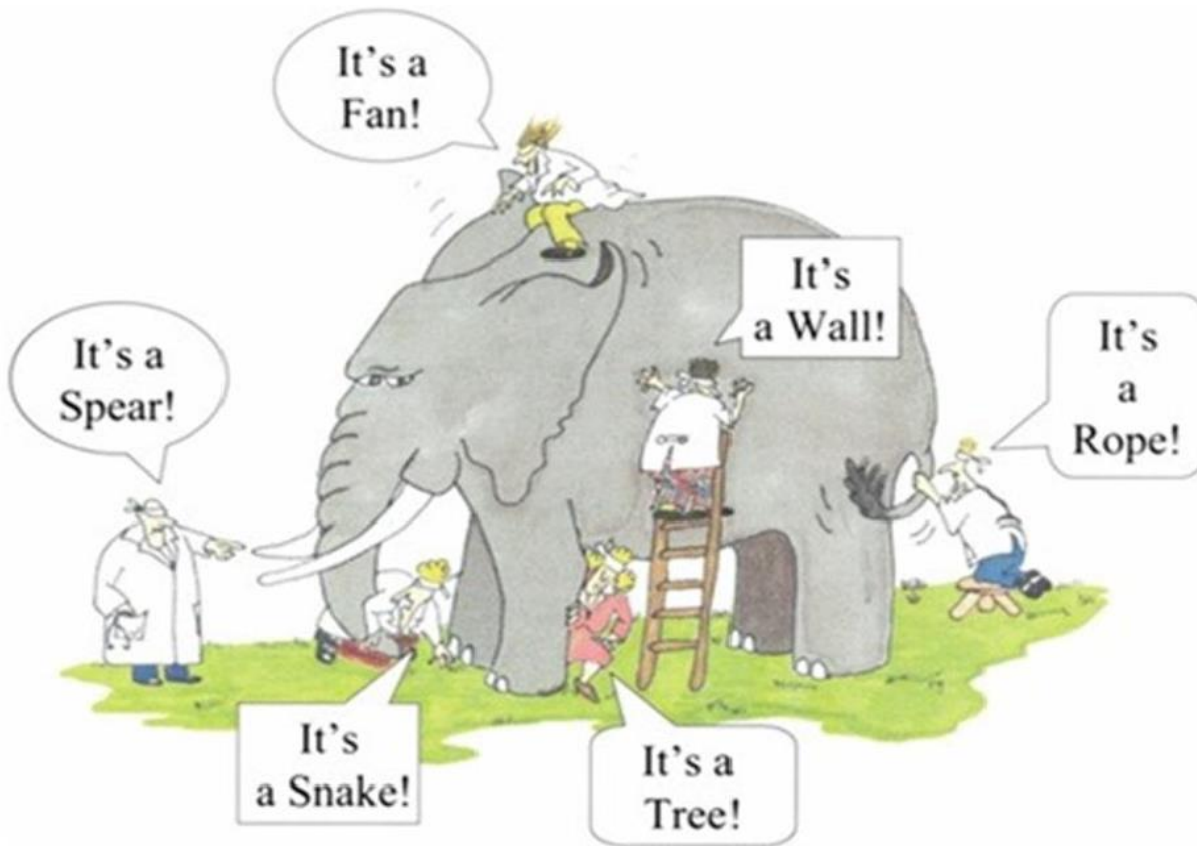
시대적 배경

빅데이터는 이미 생활의 일부가 되어 있습니다:



실상

그런데 빅데이터를 제대로 이해하고 있는 사람은 많지 않습니다.



주목해야할 트렌드

이세돌 대 알파고의 충격적인 결과:

- 이세돌 9단 대 알파고: 1승 4패.



이세돌 대 알파고의 충격적인 결과:

- 이세돌 9단 대 알파고: 1승 4패.
- 인간의 고유 영역에 대한 AI의 거침없는 도전.
- 예상했던 것 보다 빠른 AI의 발전속도.
- 기업체(Google, Facebook, Amazon 등) 주도로 발전, 경제적 가치 창출!

주목해야할 트렌드

알리바바의 마윈 회장:

- 제 4차산업 혁명의 핵심은 데이터 이다.



주목해야할 트렌드

알리바바의 마윈 회장:

- 제 4차산업 혁명의 핵심은 데이터 이다.
- 앞으로 30년간 Data Technology (DT) 혁명에 기반한 새로운 시장이 열릴 것이다.

주목해야할 트렌드

‘MIT 디지털 경제 이니셔티브’ 연구소장 Andrew McAfee:

- 스마트화와 자동화에 인한 일자리 감소.

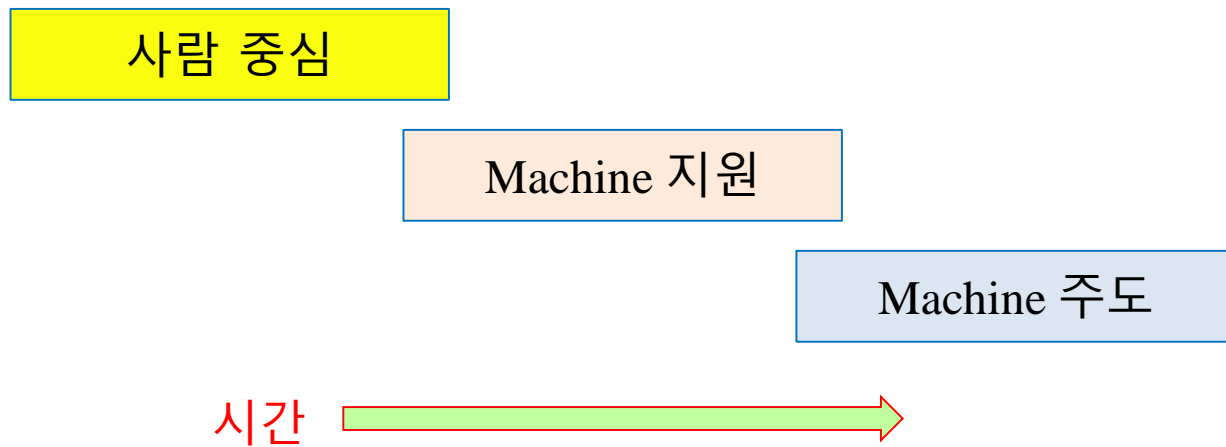


주목해야할 트렌드

‘MIT 디지털 경제 이니셔티브’ 연구소장 Andrew McAfee:

- 스마트화와 자동화에 인한 일자리 감소.
- 급진적인 AI의 발전 때문에 ‘화이트 칼라’ 일자리도 줄어들고 있습니다.
- 오히려 블루 칼라 일자리의 감소세는 둔화되고 있습니다.
- 소득의 **평균**은 꾸준히 증가하지만 **중위수**는 정체되어 있습니다 → **양극화!**

현재 진행중인 변화



- Machine 주도형의 장점: 빅데이터를 사용한 합리적이고 최적화된 판단 가능.
- 데이터가 사용되는 모든 산업 활동에 적용 가능 : 엄청난 파급효과.
- 인간의 인지력이 개입하는 곳에 새로운 기회가 있습니다 ← **엘리트**.

현재 진행중인 변화 : 4차 산업 혁명

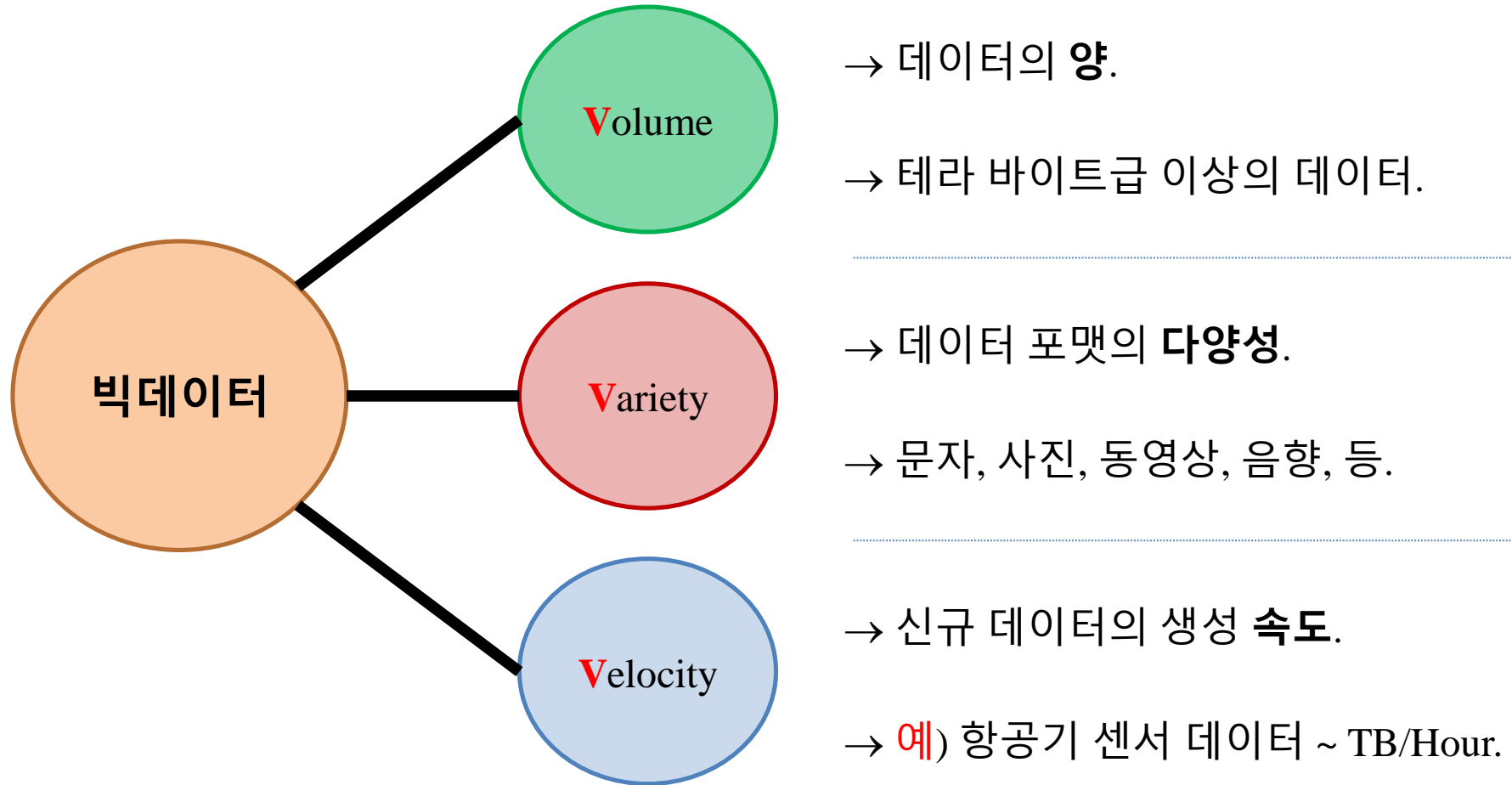


현재 진행중인 변화 : 4차 산업 혁명

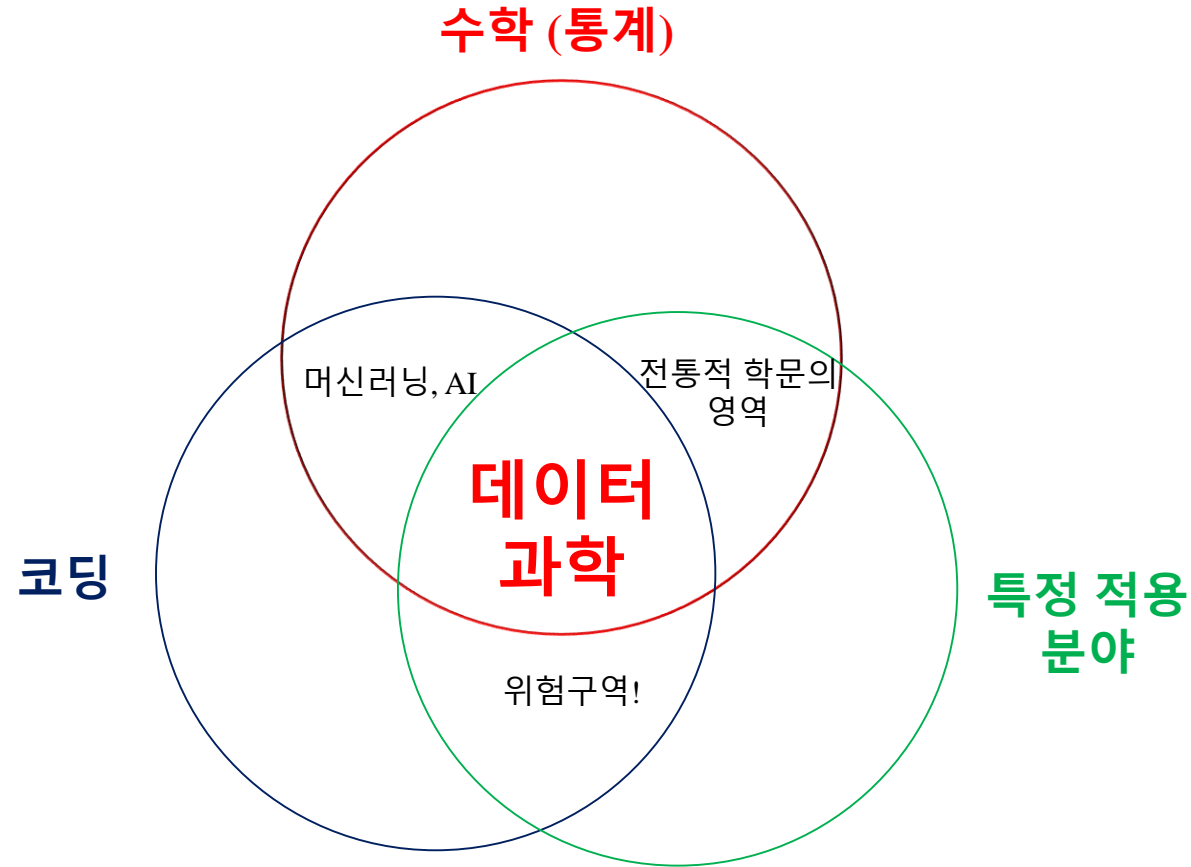


빅데이터?

빅데이터의 정의 (3V): Gartner 그룹의 Doug Laney.



데이터 과학?



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

데이터 과학 (data science):

- 데이터로부터 의미 있는 정보를 추출해내는 학문으로 이것을 효과적으로 구현하고 전달하는 과정까지를 포괄한 개념.
- 정형, 비정형을 막론하고 인터넷, 센서, 모바일 디바이스 등에서 생성된 다양한 유형의 데이터를 대상으로 함.
- 융합학문: 공학, 수학, 통계학, 컴퓨터공학, 도메인 지식.
- 3가지 분야: 분석, IT, 비즈니스.

데이터 과학자 (data scientist):

- 대용량 데이터를 검색, 발굴하여 구조화 하고 불완전한 데이터를 연결할 수 있어야 함.
- 호기심에 기초하여 문제의 이면을 파고들고, 이슈를 찾아내고 검증 가능한 가설을 세울 수 있어야 함.
- 분석에 필요한 하드스킬과 커뮤니케이션 등의 소프트스킬 겸비.

기계학습?

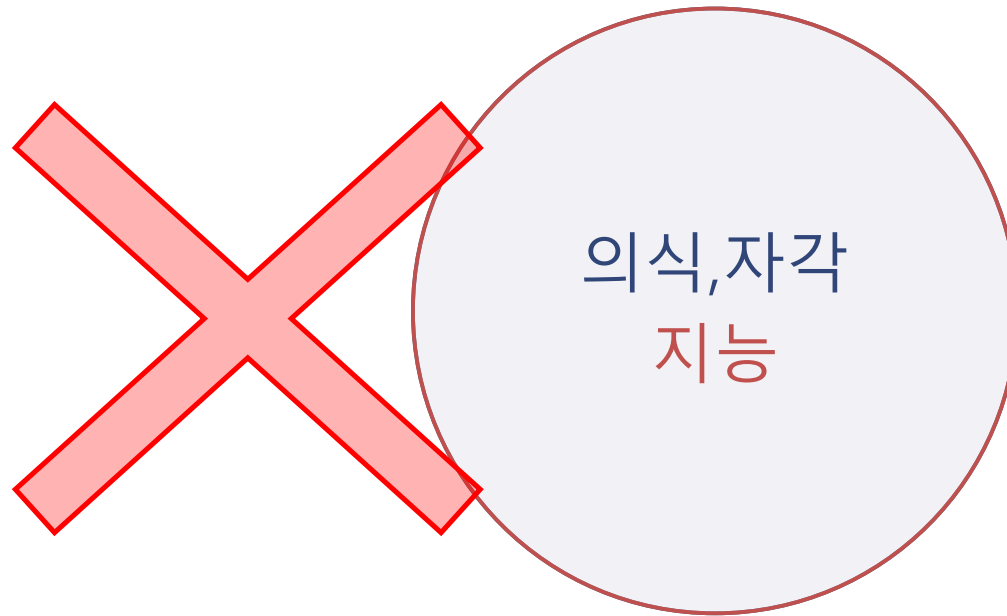
~~러닝머신 ???~~

머신러닝 ! (Machine Learning)

→ 통계 모형과 데이터를 사용한 학습과 예측을 의미합니다.

인공지능?

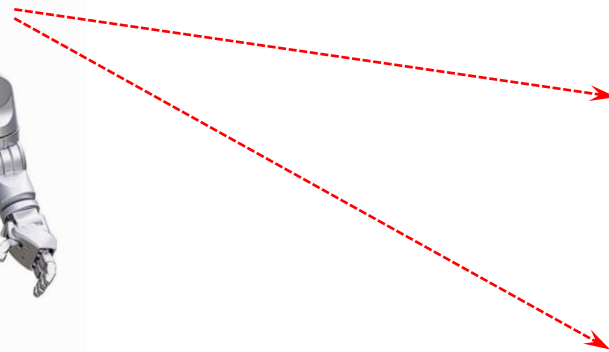
의식과 지능의 분리.



인공**의**지능

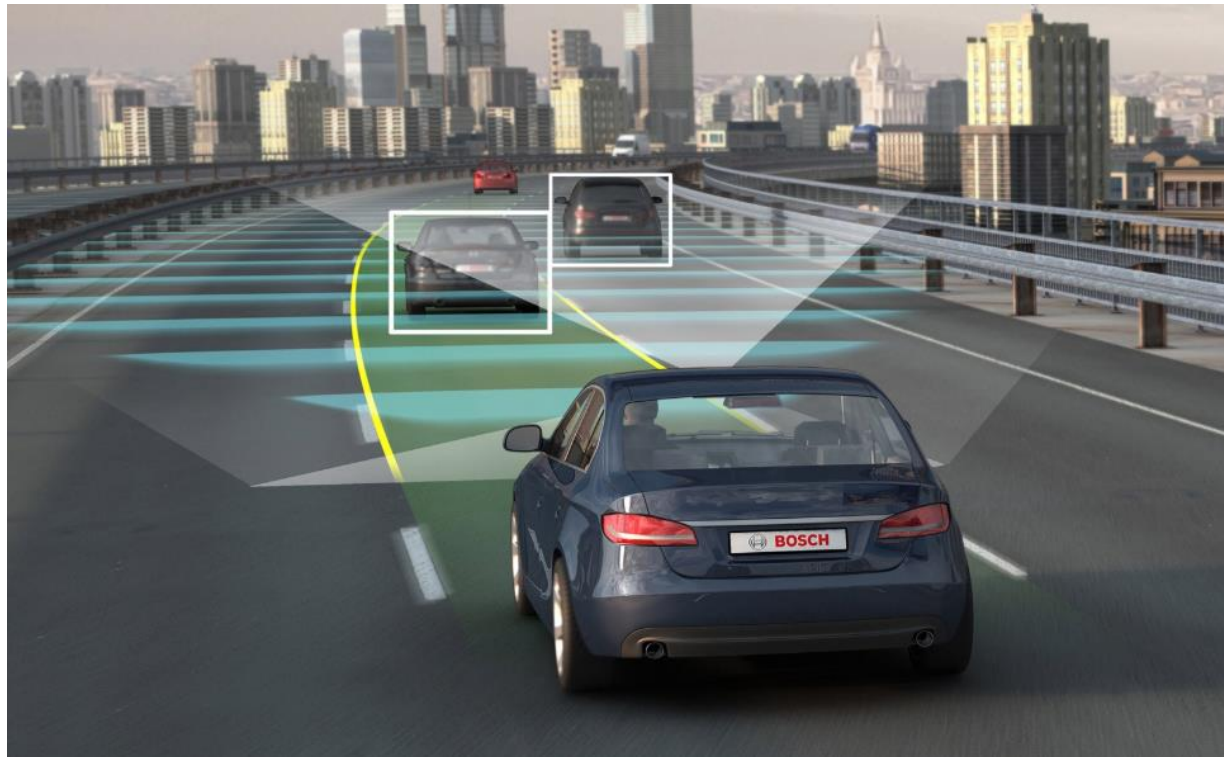
인공지능

??
동물?? 고양이??



인공지능 : 응용 분야

자율주행 차량:



통역/번역:



인공지능 : 응용 분야

금융, 투자, 핀테크:



Kaggle 대회

잘 알려진 데이터 관련 사이트 [Kaggle.com](https://www.kaggle.com):

- 2010년에 설립된 예측 및 분석 플랫폼.
- 기업/단체에서 데이터와 과제를 등록하면, 참가자들이 해결하는 모델을 개발하고 경쟁한다.

→ 명예.

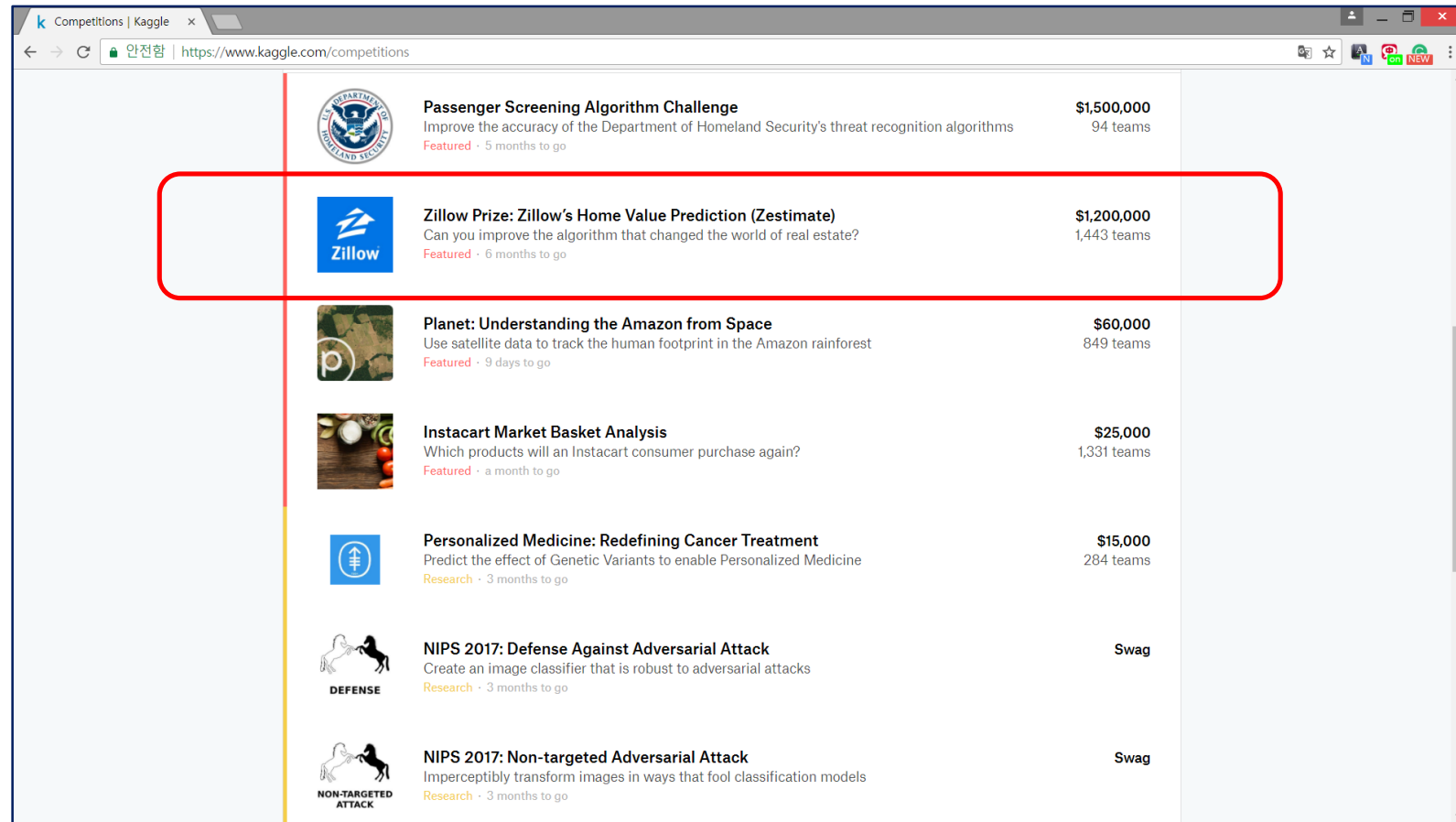


→ **상금!**



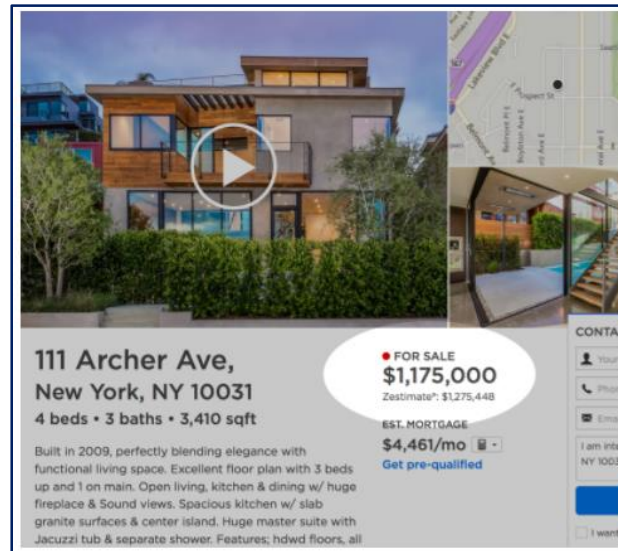
Kaggle 대회

잘 알려진 데이터 관련 사이트 Kaggle.com:



Kaggle 대회

미국 최대의 부동산 가격 감정사이트 Zillow:



→ 1억 1천만 건 이상의 부동산 리스팅.

→ 수백여 가지의 변수를 사용하여 가격 감정. 14% → 5% 이하의 오차.

1. 빅데이터 금융 사이언티스트 개요.

1.1. 빅데이터 개요.

1.2. 데이터의 이해.

1.3. 빅데이터의 가치와 인사이트.

1.4. 데이터 분석의 이해.

1.5. 빅데이터와 데이터 과학의 미래.

1.6. 데이터 사이언스와 금융.

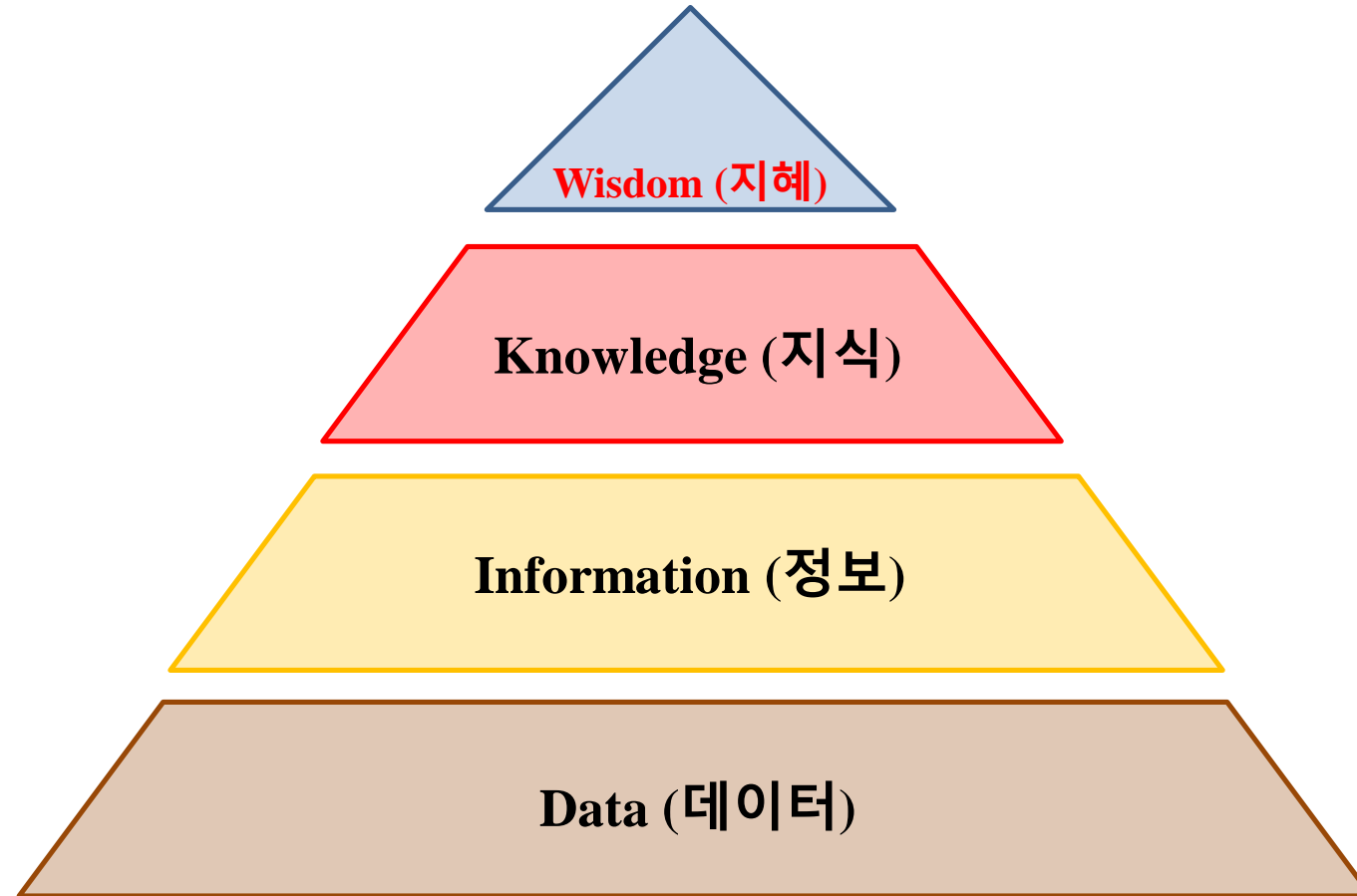
1.7. 활용 사례 (금융 & 비금융).

데이터의 정의:

- Data는 1646 영국 문헌에 처음 등장. 라틴어의 ‘주어진 것’이라는 의미.
- 컴퓨터 시대가 시작되고 추상적인 개념에서 기술적이고 사실적 의미로 변화.
- 데이터의 유형:
 - 질적 데이터, 정성적 데이터 (qualitative data): 혈액형, 성별, 국적, 지방, 등.
 - 양적 데이터, 정량적 데이터 (quantitative data): 몸무게, 신장, 가격, 등.
- 지식 경영의 핵심 이슈인 암묵지식과 명확지식 사이를 이어주는 데이터:
 - Tacit Knowledge (암묵지식): 경험으로 내면화 되어있어서 공유하기 어려움.
 - Explicit Knowledge (명확지식): 문서화 되어 있어서 공유하기 쉬움.

데이터의 이해 : 데이터와 정보

데이터와 정보와의 관계 : DIKW 피라미드



데이터 베이스의 특징:

- 1950년대 미군이 처음 구축 (컴퓨터 도서관).
- 복수 사용자의 요구에 대응해서 데이터를 받아들이고 저장, 공급하기 위해서 일정한 구조에 따라서 편성된 데이터의 집합.
- 데이터 베이스의 구조적 특징:
 - 통합성 (integrated): 동일 데이터가 중복되지 않음.
 - 저장됨 (stored): 컴퓨터 기술을 바탕으로 저장 매체에 저장됨.
 - 공용성 (shared): 여러 사용자가 다양한 목적으로 사용할 수 있음.
 - 변화성 (changeable): 새로운 데이터의 삽입, 기존 데이터의 삭제 및 변경 가능.

데이터 베이스의 특징:

- 데이터 베이스의 다른 특징:
 - 정보의 축척 및 전달 측면: 가독성, 원격 조작성, 등.
 - 정보 이용 측면: 접속 속도, 검색, 등.
 - 정보 관리 측면: 데이터 베이스의 구조.
 - 정보기술 발전 측면: 데이터 베이스 관련 소프트웨어와 하드웨어 및 네트워크.
 - 경제, 산업 측면: 데이터 인프라로서의 특성. 경제, 산업의 효율성 제고. 편의 증진.

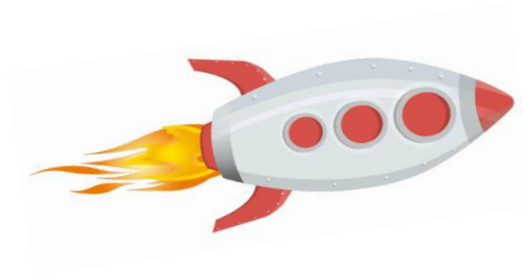
데이터 베이스의 활용: CRM (Customer Relationship Management)

“고객관계관리” 라고도 한다. 고객과 관련된 내외부 자료를 분석, 통합하여 고객 중심 자원을 극대화 하고 이를 토대로 고객 특성에 맞게 마케팅 활동을 계획, 지원, 평가하는 과정.



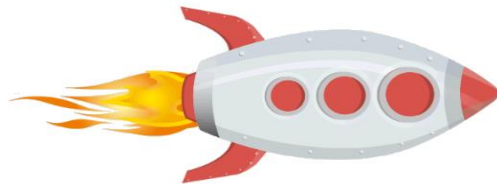
데이터 베이스의 활용: SCM (Supply Chain Management)

기업에서 원재료의 생산, 유통 등 모든 공급망 단계를 최적화해서 수요자가 원하는 제품을 원하는 시간과 장소에 제공하는 “공급망 관리”를 뜻한다.



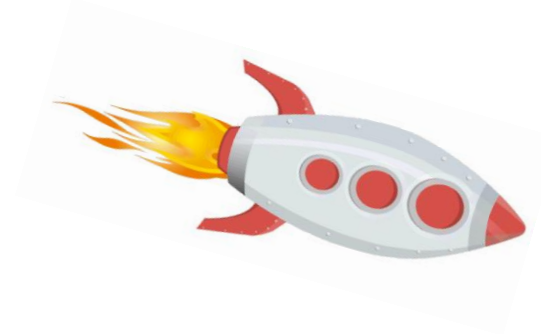
데이터 베이스의 활용: ERP (Enterprise Resource Planning)

인사, 재무, 생산 등 기업의 전 분야에 걸쳐 독립적으로 운영되던 관리 시스템의 경영자원을 하나의 통합 시스템으로 재구축함으로써 생산성을 극대화 하려는 경영혁신기법을 의미한다.



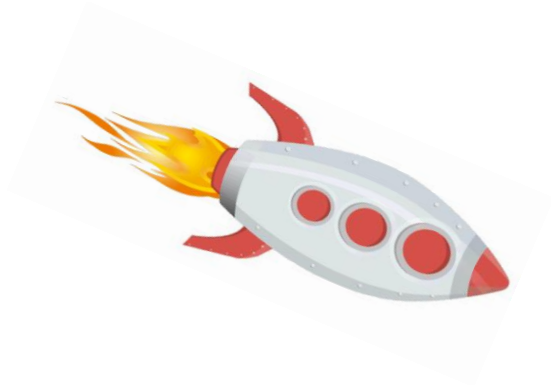
데이터 베이스의 활용: BI (Business Intelligence)

기업이 보유하고 있는 수많은 데이터를 정리하고 분석해 기업의 의사결정에 활용하는 일련의 프로세스를 말함. 기업의 사용자가 더 좋은 의사결정을 하도록 데이터를 수집, 저장, 분석, 접근을 지원하는 응용 시스템과 기술이다.



데이터 베이스의 활용: KMS (Knowledge Management System)

기업의 환경이 물품을 주로 생산하던 산업사회에서, 지적 재산의 중요성이 커지는 지식사회로 급격히 이동함에 따라, 기업 경영을 지식이라는 관점에서 새롭게 조명하는 접근방식. 의사결정의 주체인 인적 자원이 떠나면 그가 갖고 있던 지식 자원도 함께 떠나가고 기업의 지적 자원이 소실되니, 이것을 방지하는 차원에서 KMS를 활용함.



1. 빅데이터 금융 사이언티스트 개요.

1.1. 빅데이터 개요.

1.2. 데이터의 이해.

1.3. 빅데이터의 가치와 인사이트.

1.4. 데이터 분석의 이해.

1.5. 빅데이터와 데이터 과학의 미래.

1.6. 데이터 사이언스와 금융.

1.7. 활용 사례 (금융 & 비금융).

빅데이터의 출현 배경:

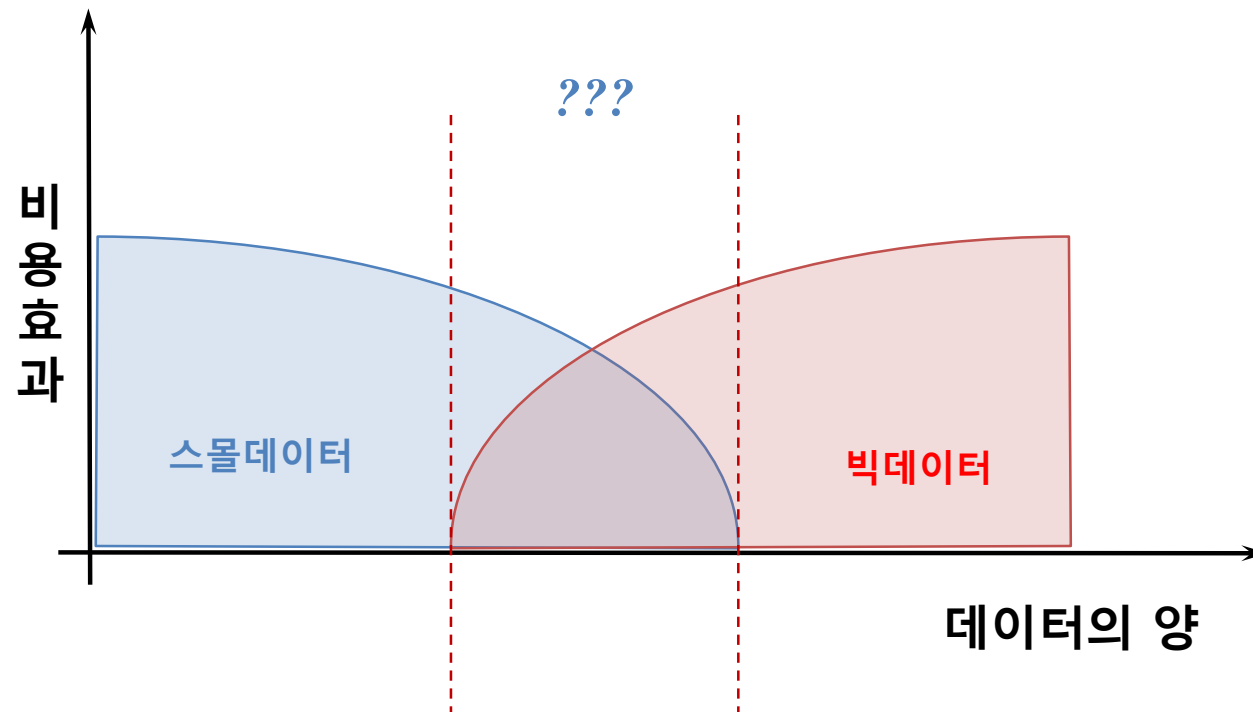
- 산업계: 양 ↔ 질 변환 법칙.
- 학계: 거대 데이터 활용 과학 확산. 예). 인간 게놈 프로젝트.
- 관련기술의 발전: +디지털화의 가속, 저장 기술의 발달과 가격 하락.
+인터넷, 모바일, 클라우드 등 관련기술의 보편화.

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:

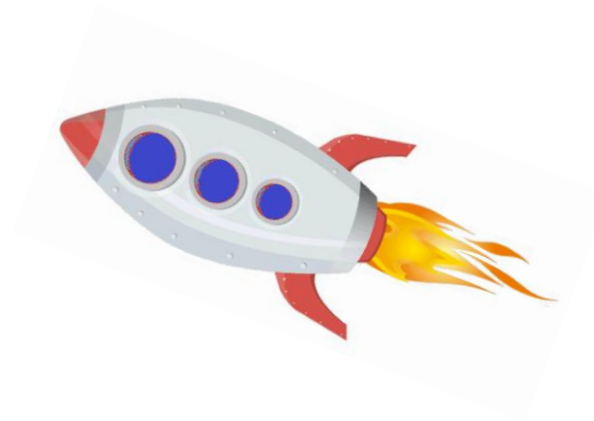
필요한 정보만 수집하는 시스템이 아니고 가능한 한 많은 데이터를 수집하여서 그 데이터를 다양한 방법으로 조합해 숨은 정보를 찾아낸다.



빅데이터 시대의 본질적인 변화

표본조사에서 전수조사로:

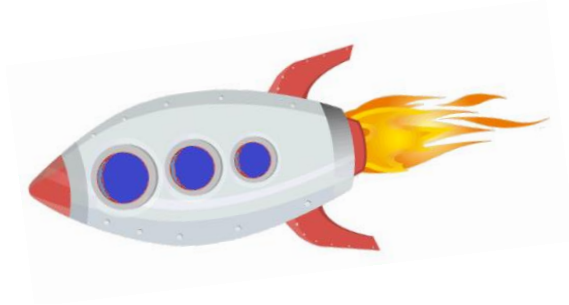
표본조사를 통해서 나타나지 않는 패턴이나 정보를 알아낼 수 있다.



빅데이터 시대의 본질적인 변화

질보다는 양:

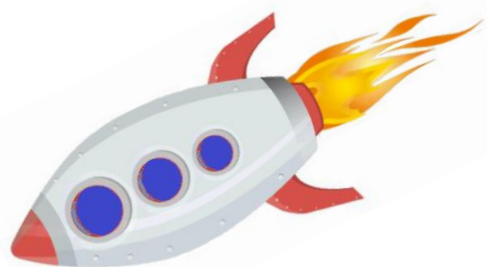
정보가 축적될 수록 오류 정보보다는 양질의 정보가 많아져서 전체적으로 좋은 결과 산출에 긍정적 영향을 미친다.



빅데이터 시대의 본질적인 변화

인과관계에서 상관관계로:

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



빅데이터 열풍과 회의론:

- “빨리 끓어오른 냄비가 빨리 식는다”라는 일종의 거품현상 우려.
- 성급한 투자에 비해서 미미한 성과.
- 빅데이터 사례 과대 포장 논란 (예: CRM). 굳이 빅데이터가 필요없다는 인식 확산.

빅데이터에서 'Big'이 핵심이 아니다!

- 단순히 분석을 많이 하는 것이 경쟁 우위를 가져다 주지는 못한다.
- 이미 여러 분야에서 일차원적인 분석 애플리케이션은 많이 있다.
예). 금융: 신용점수 산정, 이상거래 탐지, 가격 책정, 트레이딩, 클레임 분석, 고객수익성 분석, 등.
- 싸이월드의 실패 사례:
 - 웹로그 분석과 같은 일차원적 분석에만 집중.
 - 소셜네트워크 분석을 위한 관련 지표나 프레임워크 미적용.
 - 데이터 분석을 의사결정과 전략적 통찰에 접목시키지 못함.
- 전략적 통찰이 없는 분석의 함정: 사업 환경의 변화, 새로운 기회 발굴 등에 악영향.

빅데이터 시대 : 인재의 역량

필요한 인재의 역량:

하드 스킬

(코딩 + 통계 + 공학)



소프트 스킬

(창의력 + 스토리텔링 + 커뮤니케이션)

필요한 인재의 역량:

- 데이터 과학자: 이공계와 인문계의 교차점.
- 전략적 통찰을 완성 시키는 것은 인문학적 요소인 소프트 스킬이다!
- 여러 관점 (성향적, 행동적, 상황적)에서 인간을 바라볼 수 있을 때 새로운 가치 창조.

빅데이터 시대 : 인재의 역량



기능 (스킬)



기술



예술

+ 감성

+ 창의력

1. 빅데이터 금융 사이언티스트 개요.

1.1. 빅데이터 개요.

1.2. 데이터의 이해.

1.3. 빅데이터의 가치와 인사이트.

1.4. 데이터 분석의 이해.

1.5. 빅데이터와 데이터 과학의 미래.

1.6. 데이터 사이언스와 금융.

1.7. 활용 사례 (금융 & 비금융).

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: 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: 텍스트 마이닝 (Text mining)

- 비정형 텍스트 데이터에서 통계적, 연관적 특성을 추출하는 기술.



데이터 분석 기술: 클라우드 컴퓨팅, MySQL & NoSQL

- 빅데이터의 저장과 관리 운영을 위한 기술.
- MySQL은 관계형 데이터 베이스. NoSQL은 관계형 데이터베이스의 틀을 벗어난 일관성 모델을 이용.
예). MongoDB, Hadoop, 등.



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: 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: 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.



1. 빅데이터 금융 사이언티스트 개요.

1.1. 빅데이터 개요.

1.2. 데이터의 이해.

1.3. 빅데이터의 가치와 인사이트.

1.4. 데이터 분석의 이해.

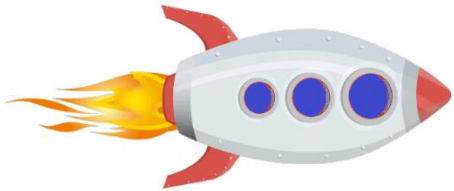
1.5. 빅데이터와 데이터 과학의 미래.

1.6. 데이터 사이언스와 금융.

1.7. 활용 사례 (금융 & 비금융).

외부 환경적 측면의 인문학 열풍:

- 단순 세계화에서 복잡한 세계화로의 변화: 컨버전스 → 디버전스.
- 비즈니스 중심이 제품생산에서 서비스로 이동: 견고한 제품 → 뛰어난 서비스 응대.
- 경제와 산업의 논리가 생산에서 시장창조로 변황: 기술경쟁 → 무형자산의 경쟁.

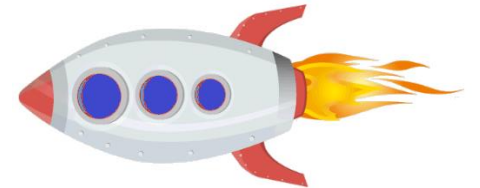


가치 패러다임의 변화:

제1 단계) 디지털화: 과거 문서의 디지털화를 통한 가치의 형상화와 표준화.

제 2단계) 연결: 현재 진행중인 인터넷, 모바일 등의 기술을 활용한 정보의 연결과 제공.

제 3단계) 에이전시: 미래에는 개인, 기기, 사물을 서로 연결하여 방대한 정보를 제공하고 관리하는 시대로 발전할 것이 예상됨.



1. 빅데이터 금융 사이언티스트 개요.

1.1. 빅데이터 개요.

1.2. 데이터의 이해.

1.3. 빅데이터의 가치와 인사이트.

1.4. 데이터 분석의 이해.

1.5. 빅데이터와 데이터 과학의 미래.

1.6. 데이터 사이언스와 금융.

1.7. 활용 사례 (금융 & 비금융).

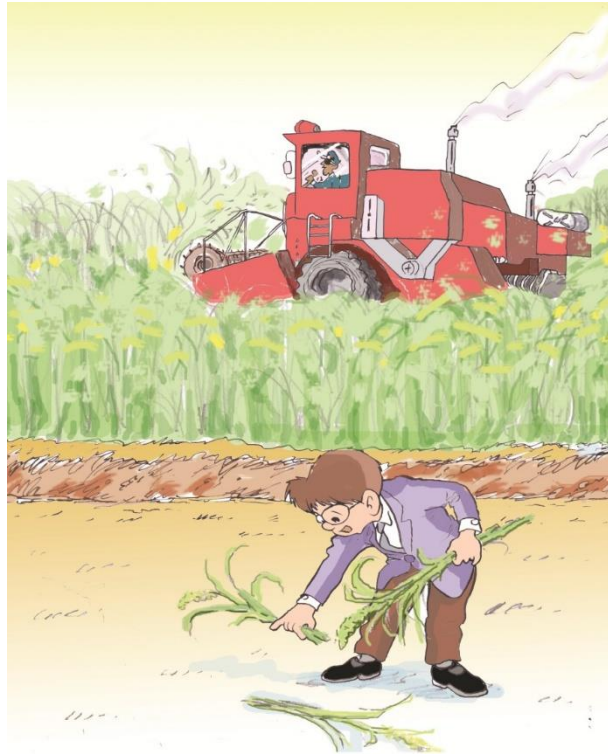
인간의 선입견 보다는 데이터에 기반한 AI의 판단에 의존:



AI와 금융은 대박?



AI와 금융은 대박?



- 소홀히 지나칠 수 있는 가치의 발견!
- 육중한 몸집의 시장이 버리고 간 이삭을 줍는다.

AI와 금융은 대박?



- 리스크 관리는 선택이 아닌 필수.

AI와 금융은 대박?



- 분석을 통한 전략의 도출.

확률과 통계의 중요함

난제를 풀어봅시다.



난제 = *brainteaser*

난제를 풀어봅시다.

질문: 다음과 같은 룰의 동전 던지기 게임에 참가하겠습니까?

동전 던지기 게임의 룰:

- 앞면이 나오면 수익 100\$ 이고, 반대로 뒷면이 나오면 수익 0\$이다.
- 매번 게임에 참가하는 비용은 40\$이다.

확률과 통계의 중요함

난제를 풀어봅시다.

그런데, 동전은 앞면이 나올 확률과 뒷면이 나올 확률이 같습니다.
1회 기대수익은 다음과 같습니다.



$$\text{기대수익} = \frac{1}{2} \times 100\$ + \frac{1}{2} \times 0\$ = 50\$$$



기대수익 50\$가 비용 40\$보다 크니까 물론 게임에 참가?

확률과 통계의 중요함

난제를 풀어봅시다.

잠깐만, 1회 수익의 리스크(변동성, 표준편차)를 계산해봅시다.



$$\text{리스크} = \sqrt{\frac{1}{2} \times (100 - 50)^2 + \frac{1}{2} \times (0 - 50)^2} = 50\$$$

확률과 통계의 중요함

난제를 풀어봅시다.

샤프지수는 리스크 대비 초과수익을 나타냅니다.



$$\begin{aligned}\text{샤프지수} &= \frac{\text{초과수익}}{\text{리스크}} \\ &= \frac{(\text{기대수익} - \text{참가비용})}{\text{리스크}} \\ &= \frac{(50 - 40)}{50} = 0.2\end{aligned}$$



1 미만의 샤프지수는 불만족스럽습니다.

확률과 통계의 중요함

난제를 풀어봅시다.

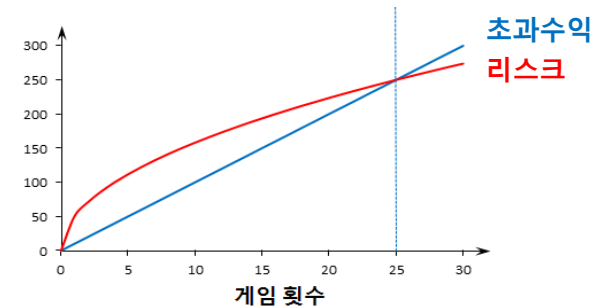
그런데, 1 회 이상 N 회 동전 던지기를 할 수 있다면?

- 초과수익은 N 에 비례해서 증가합니다.
- 리스크(표준편차, 변동성)은 \sqrt{N} 에 비례해서 증가합니다.
- 이항 분포 (binomial distribution)의 특성입니다.

↓

$$\text{샤프지수} = \frac{(50 - 40) \times N}{50 \times \sqrt{N}} = 0.2 \times \sqrt{N}$$

↓



25 회 이상 플레이하면 샤프지수는 1을 초과하게 됩니다. → 만족 !

1. 빅데이터 금융 사이언티스트 개요.

1.1. 빅데이터 개요.

1.2. 데이터의 이해.

1.3. 빅데이터의 가치와 인사이트.

1.4. 데이터 분석의 이해.

1.5. 빅데이터와 데이터 과학의 미래.

1.6. 데이터 사이언스와 금융.

1.7. 활용 사례 (금융 & 비금융).

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%).

사례 #1 : 영국계 대형은행

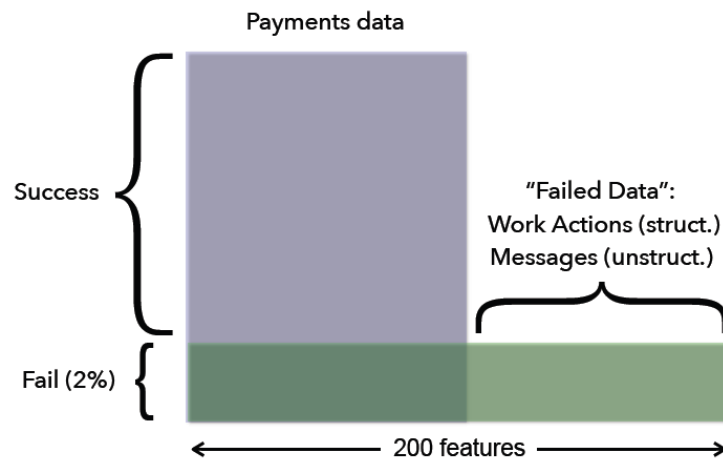
목표:



- 해외 송금중 2% FAIL분석 (이상거래 분석).
- FAIL의 이유를 분석, 예측하여 대응력 증강.
- FAIL의 경우 처리에 소요되는 시간과 비용의 절감.

사례 #1 : 영국계 대형은행

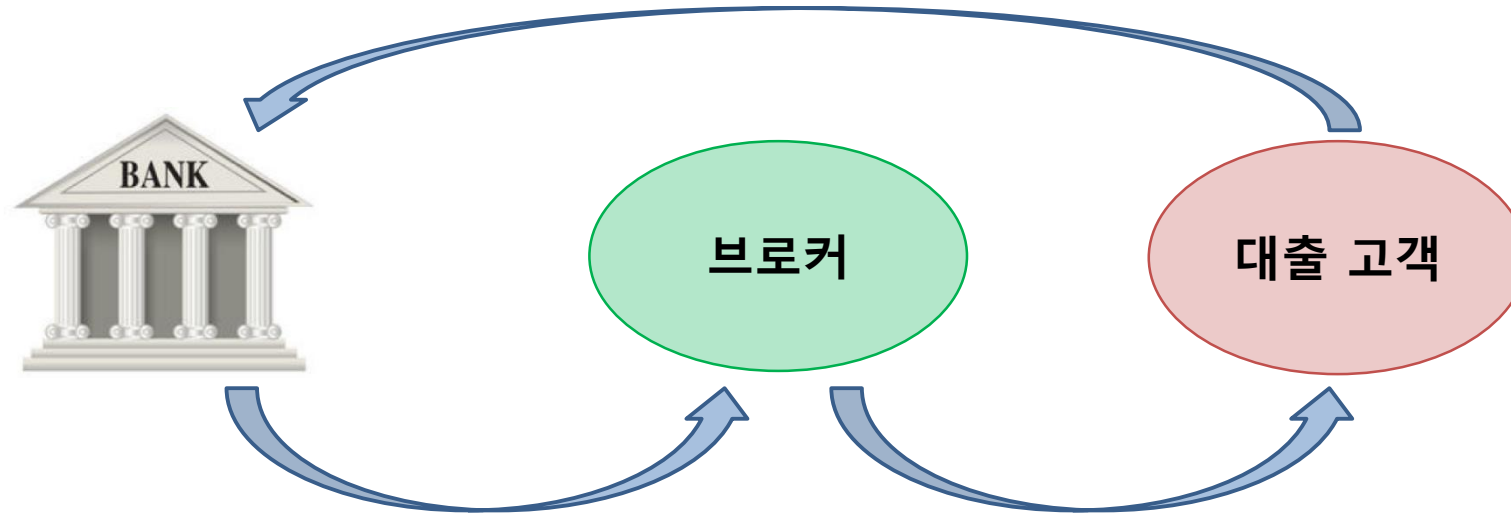
내용:



- 데이터 가공을 통한 유효 특성 추출 (feature engineering).
- 자율학습 (unsupervised learning)을 통한 고객군 분류.
- 지도학습 (supervised learning)을 통한 FAIL 예측.

사례 #2 : 영국계 중견은행 A사

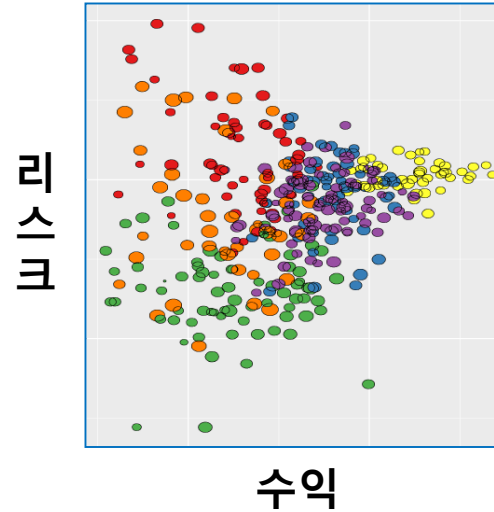
목표:



- Asset Finance Group (AFG)의 **브로커** 프로파일링 시스템 구축.
- 고객과 브로커 특성에 의한 군집화 분석. 효율적인 커미션 군집 인식.
- 리스크 관리 및 수익률 최적화.

사례 #2 : 영국계 중견은행 A사

내용:



- 데이터 테이블의 연결과 유효 특성 추출 (feature engineering).
- 자율학습 (unsupervised learning)을 적용한 브로커 분류.
- 이메일과 전화통화 내용의 텍스트 마이닝을 통한 무드 분석.

사례 #2 : 영국계 중견은행 A사

AGILE 개발 **사이클** 도입:

결과 평가 & FEEDBACK



탐색과 유효 특성 추출

분석 적용

사례 #3 : 영국계 중견은행 B사

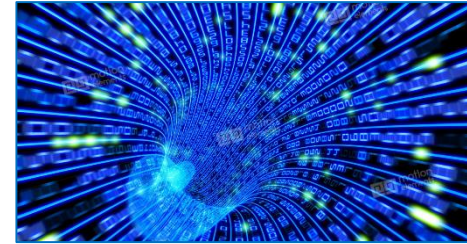
목표:



- 외환 (FOREX) 고객의 주문 패턴 분석.
- 고객의 주문 예측에 대한 대응력 최적화.

사례 #3 : 영국계 중견은행 B사

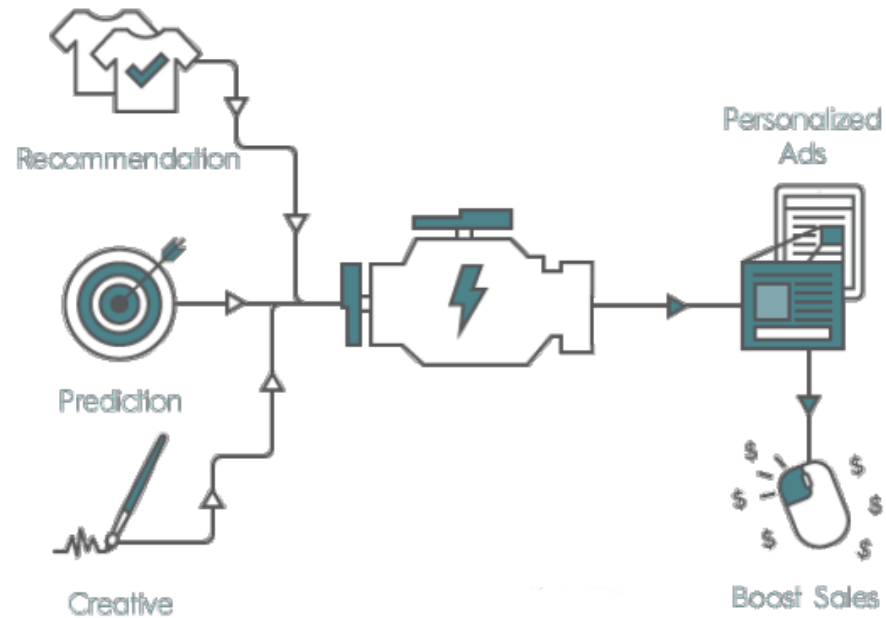
내용:



- 각종 지표 (원유가격, 금값 등)와 고객의 과거 주문 데이터 사용.
- 주문시기와 주문량 예측.
- 통계적 유의미한 예측성 확인.

사례 #4 : 인터넷 마케팅 C사

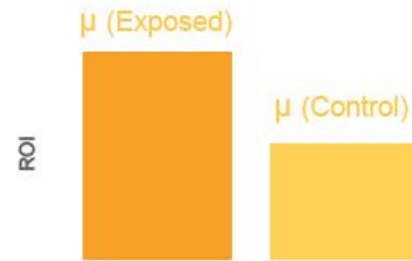
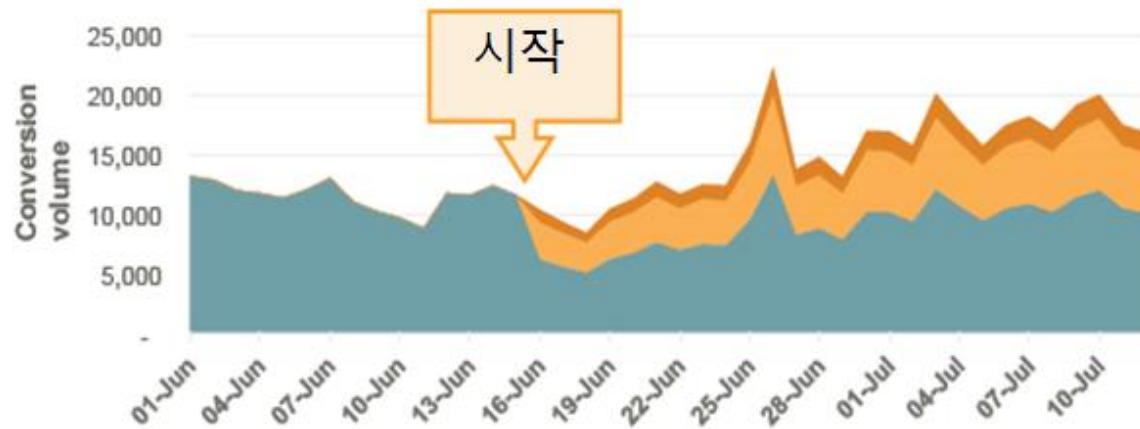
목표:



- 마케팅 캠페인의 ROI분석.
- 사이트 디스플레이 효과 분석.

사례 #4 : 인터넷 마케팅 C사

내용:



Happy client

- 데이터를 사용한 가설검정.
- 통계적 유의미한 마케팅 효과 추적.

사례 #5 : 업스케일 유통업 M사

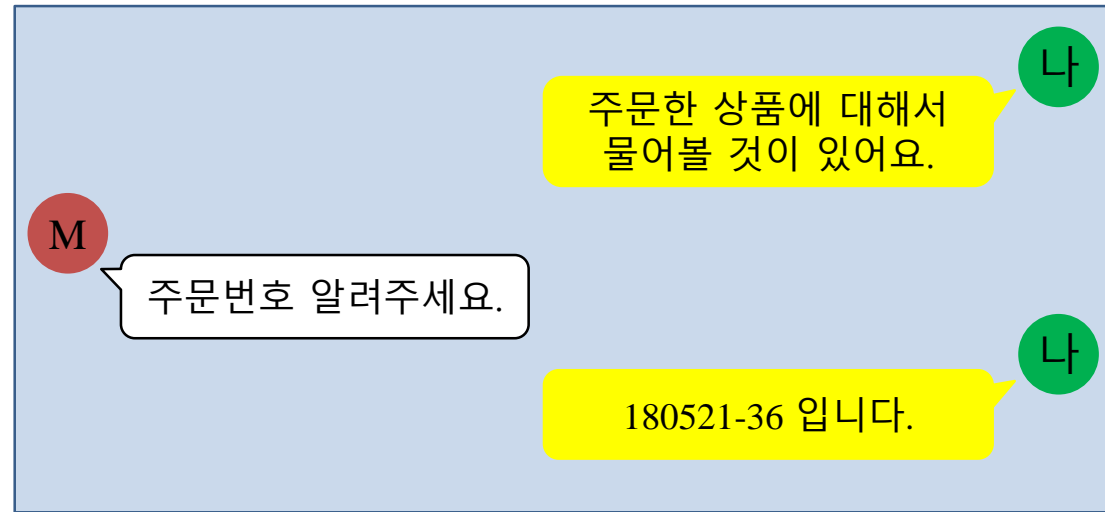
목표:



- 홈페이지에 챗봇 (chatbot) 구축.
- 고객 문의 중 ~30% 응대 처리 목표.
- CS의 반자동화. 최적화된 채널링 (channeling) 시스템 구축.

사례 #5 : 업스케일 유통업 M사

내용:



- 자연언어 인식 (NLP) 기술 적용.
- 문맥과 고객의 무드 인식.
- 메모리 효과.

사례 #6 : 스마트 시티

목표:



시간과 위치에 따라서 심한 **불균형** 발생

- 시내 곳곳에 설치되어 있는 자전거 stand의 사용/대기 최적화.
- 시간대별로 자전거 플로우 예측 & 대응 (특히 피크타임).

사례 #6 : 스마트 시티

내용:

- 4년간 축적된 데이터 (2012~2016) + 실시간 데이터 사용.
- 자전거 한대 하루 평균 2.4회, 50분간 대여됨 (under-utilization).
- 클라우드 병렬 컴퓨팅과 인공지능망 적용.
- 10분 단위로 stand별 inflow와 outflow 예측.

사례 #7 : 차량 관제 서비스 S사

목표/내용:



- 렌터카 fleet의 운영 데이터 분석.
- 운영 데이터에 기반한 보유 fleet size의 최적화.

명사의 명언 한마디

“Everything should be made as simple as possible but not simpler !”

→ *Albert Einstein* ←

“완벽을 추구함에 있어서는 간결함을 목표로해야 하지만 단순화는 피해야 한다”

→ 알베르트 아인슈타인 ←

모듈 #1: 끝

문의:

sychang1@gmail.com