

빅데이터 활용사례

— 해외

- 01 · 노스페이스, 기계학습을 통한 지식 기반 상품 추천 서비스
- 02 ·페이팔, 딥러닝을 통한 온라인 사기 방지 시스템 구축 사례
- 03 · 스타캐스트, 빅데이터 기반의 MLB 중계 시스템
- 04 · Travelbasys, 소셜 및 데이터 기반의 안전 여행 알림 서비스
- 05 · DB Systel, 센서데이터 분석을 통한 기관차운행관리 개선 사례
- 06 · 오므론, 빅데이터 분석을 이용한 제품 생산성 향상

01

BIGDATA
해 외 사 례

AI를 이용한 'Expert Personal Shopper'로 보다
스마트한 상품 추천을 주도

노스페이스

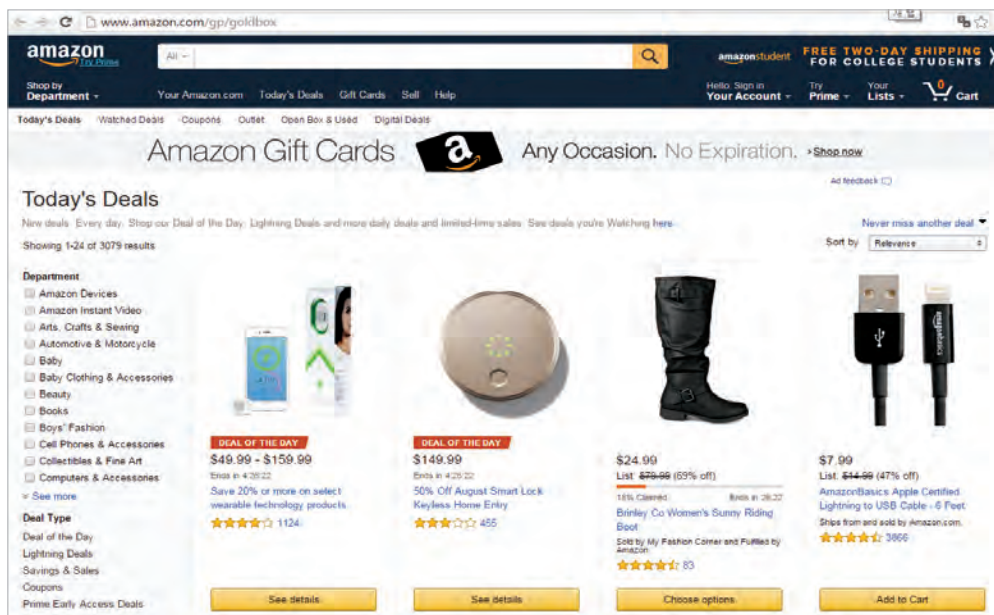
❖ 사업의 추진 배경 및 목적

1. 슬슬 한계를 드러내고 있는 온라인 쇼핑몰 추천 서비스

• 기본적인 온라인 추천 서비스 제공은 이제 더 이상 새롭지 않아

- 온라인 쇼핑몰에서는 데이터를 활용해 고객들이 많이 찾는 제품들만 선정해서 보여주거나 평소와 비교하여 저렴하게 책정된 가격의 상품을 한 곳에 모아 보여주는 핫딜(Hot-deal) 추천 서비스가 대다수
- 또는 소비자가 이전에 선택했던 상품을 모아서 한 번에 제공하는 개인 큐레이션 서비스를 제공하기도 함

[그림] Amazon Hotdeal(Today's deal)



[출처 : Amazon.com]

전문 지식 전달이 상품 구매로 이어지는 특성을 가진 아웃도어 상품 - 추천의 새로운 장을 열다

• 상품의 기능성에 대한 이해가 필요한 아웃도어용품

- 아웃도어 용품의 경우 오프라인 매장을 방문하는 소비자가 많은 이유는 일반 소비자에게 수많은 기능이 집약되어 있어 상품에 대한 지식이 대부분 없기 때문
- 의류 매장에 경우 단순히 소비자가 원하는 디자인을 고르면 되지만 아웃도어용품의 경우 산악 관련 상품의 기능에 대한 전문적 지식이 필요하기 때문에 사이트를 방문한 개인이 관심 제품 정보만을 노출하는 것은 실제 구매까지 연결하는데 한계를 가짐
- 이처럼 단순한 개인화된 사이트 구성뿐만 아니라 방문하고자하는 지역과 용도, 그에 맞는 기능을 갖춘 상품을 전문 지식과 함께 제공하는 추천 서비스가 필요

❖ 추진 내용

1. 소비자 욕구에 특화된 목적 지향 상품 추천 서비스 구현

• 기능성 이해도 증진을 위해 상품과 관련된 빅데이터 수집 · 분석

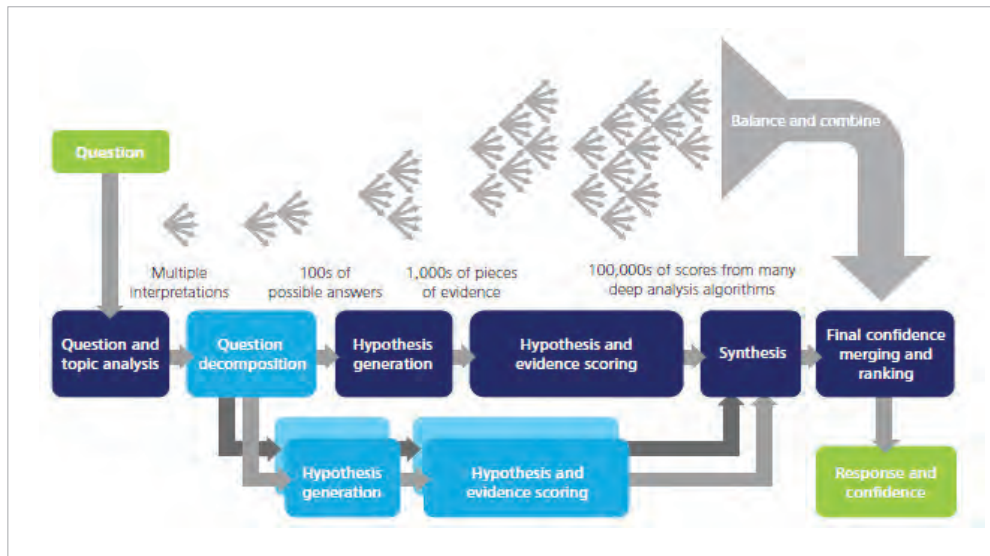
- 해당 상품에서 제품을 선택할 때 소비자가 따라가는 상품에 대한 쇼핑몰 고객의 판매 내역과 해당 사용자의 리뷰 등 상품 관련 데이터에 대한 정보 수집 및 분석
- 쇼핑몰 내에 수집되는 데이터 외에도 기존 오프라인 매장 판매자의 상품 판매와 관련된 소비 패턴을 데이터화하고, 상품과 관련된 개인블로그 및 산악 잡지, 아웃도어 간행물의 데이터 수집을 통한 상품 분석

• 인공지능(AI) 대화상자를 통한 특화된 상품 추천

- 구매자가 원하는 특정 상품을 바탕으로 인공지능 기반의 대화 상자에서 구매하고자 하는 상품의 세부 내용을 입력하여 상품을 추천

- 인공지능 대화 상자를 통해 방문하고자 하는 지역과 필요한 상품의 요건 등 구매하고자 하는 상황을 파악하고, 해당 지역의 날씨 및 지형 등의 환경적 요인을 분석하고, 전문가의 사용 평가 및 산악용품 사용 후기 등을 종합적으로 분석하여 조건에 충족하는 상품을 화면에 제공
- 오프라인 매장에서만 얻을 수 있었던 정보와 지식을 온라인 쇼핑몰에서도 소비자에게 제공하게 되면서 온라인 쇼핑의 이용률을 높일 수 있고, 사이트를 방문하는 고객의 쇼핑물 이동, 클릭, 탐색 데이터와 인공지능 커뮤니케이션 데이터를 통해 보다 적합도가 높은 상품 및 지식 제공이 가능하도록 시스템을 구축

[그림] 인공지능 질의를 통한 데이터 분석 프로세스



[출처 : IBM]

❖ 효과 및 향후 적용 확대 방안

1. 추천 상품 클릭률 평균 60% 달성

• 재방문 의사를 밝힌 소비자가 대다수

- 2014년 카카오에서 발표한 자료에 따르면 평균 업종별 광고 클릭률은 7.4%, 최적화 광고 클릭률은 8.5%를 기록하였는데, Fluid의 'Expert Personal Shopper'의 경우 추천 상품에 대한 클릭률은 60%로 소비자의 취향을 정확하게 저격
- 온라인 쇼핑몰에서 고객들이 대화박스에 참여한 시간은 평균 2분을 기록

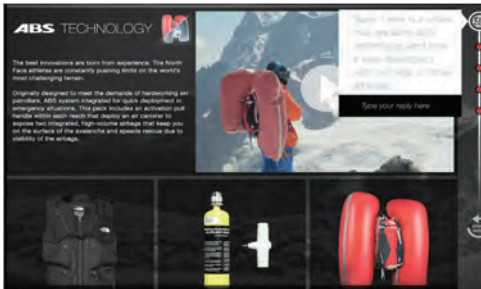
〈아웃도어 제품 구매의 질문 타이핑〉



〈필요한 제품의 추천〉



〈아웃도어 제품 구매의 질문 타이핑〉



〈필요한 제품의 추천〉



– Fluid의 “Expert Personal Shopper”를 사용했던 소비자들은 대부분 높은 만족도를 나타내며 해당 사이트를 다시 사용할 것이라고 평가

[그림] Expert Personal Shopper 화면



[출처 : Northface]

02

BIGDATA
해 외 사 례

딥러닝 갑옷 입고 온라인 사기 ‘수호신’ 자처 페이팔

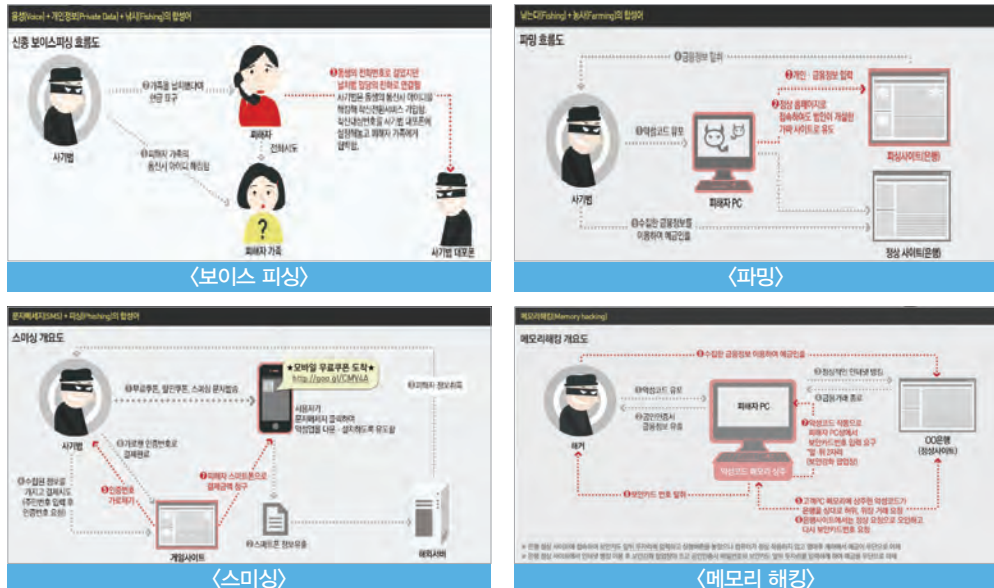
❖ 사업의 추진 배경 및 목적

1. 진화하는 온라인 사기, 소비자는 혼란 속으로

• 온라인 사기 갈수록 증가하고, 지능적으로 진화

- 소비자의 개인정보를 불법으로 취득하려는 피싱사기가 유명 온라인 쇼핑몰이나 은행 그리고 SNS사이트로 교묘하게 위장하며 소비자들을 혼란에 빠뜨리고 있음
- 신종 온라인 피싱은 불특정 다수를 향한 기존 형태에서 벗어나, 피해 대상이 되는 개인의 이메일 주소 등 개인정보를 미리 확보한 후 해당 사이트와 동일한 로고와 이메일 형식으로 개인정보의 변경을 요청하거나 결제를 요구하는 행태를 보여 피싱사이트와 정식 웹사이트의 구분이 갈 수로 어려워지고 있는 실정

[그림] 다양해진 사기 수법



[자료 : 조선일보, 2015]

빅데이터 활용한 딥러닝 기술, 온라인 결제 안전 ‘지킴이’

• 다양해진 사기수법으로 각별한 주의 필요

- 지능화된 피싱메일의 링크를 의심하지 않고 눌렀을 경우 일차적인 정보누출의 피해뿐 아니라 스파이웨어나 해킹프로그램에 감염될 위험성이 높아 소비자들의 각별한 주의가 필요
- 스마트폰을 통한 피싱사기도 점차 증가하고 있는데, 대부분의 소비자들이 스마트폰의 앱을 다운로드 받을 때 특별한 경계를 하지 않는 점과 스마트폰에서는 웹페이지에 입력된 정보를 지우지 않고 저장하는 습관이 있는 점에 착안하여 범죄자들이 스마트폰 앱에 해킹프로그램을 설치하는 사례가 증가하고 있음

2. 온라인 사기 피해 - 사전 예방이 ‘중요’

• 현실과 동떨어진 예방 방법은 ‘문제’

- IT업계 전문가들은 이러한 온라인 피싱사기를 사전에 방지하기 위해서는 대부분의 상품구매나 금융거래와 관련된 이메일을 받을 경우 그 내용이 거래내역확인 이상을 요구한다면 무시할 것을 주문
- 받은 이메일에 포함된 링크 URL을 클릭하지 말고 해당 사이트에 직접 접근하여 로그인을 하거나, 링크 URL을 클릭하지 말고 커서를 올려만 놓게 되면 링크의 주소를 확인할 수 있어 피싱메일을 소비자가 직접 걸러낼 수 있다고 조언

• 소비자의 눈높이에 맞지 않는 조언, 소비자는 “글쎄...”

- 피해를 입은 소비자는 다양한데, 전문가의 조언은 전문용어만 나열하여 IT와 친숙하지 않은 소비자들은 이해하기도 어려움
- ‘사후약방문(死後藥方文)’ 격의 조언으로 피해를 입은 후에야 대책을 마련해 소비자들이 이미 입은 피해는 되돌릴 수 없음

●● 추진 내용

1.페이팔의 결제 사기 대응책 - '딥러닝' 도입

- **페이팔, 훨씬 복잡해진 사기수법을 분석 및 예방하기 위해 '딥러닝(Deep Learning)' 기술 도입**
 - 국내에서 논의가 한창인 이상금융거래탐지시스템(FDS, Fraud Detection System)의 원조격으로 페이팔은 사람의 뇌와 유사한 방식으로 분석업무를 수행하는 '딥러닝(Deep Learning)' 기술까지 적극 도입
 - 페이팔 글로벌 리스크 사이언스 담당 선임디렉터인 후이 왕(Hui Wang)은 “사기 수법이 이전보다 훨씬 복잡해지면서 사기분석 및 예방을 위해 딥러닝 기술을 도입하고 있다.”고 밝힘

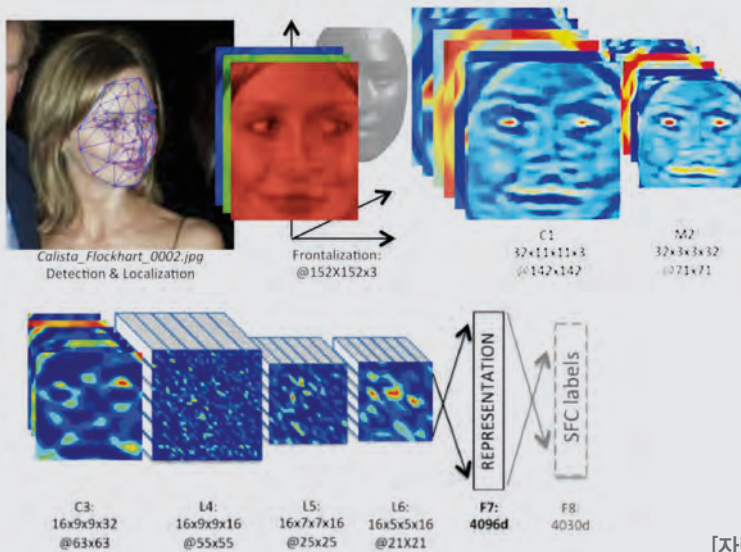
- **빅데이터를 활용한 딥러닝 기술은 빅데이터 융합 인식기술**

- 딥러닝 기술은 머신러닝(Machine Learning) 혹은 인공지능(Artificial Intelligence)에 대한 또 다른 접근법 중 하나이며, 구글, 페이스북, 마이크로소프트, 바이두 등이 수년간 연구해 온 분야
- 사람의 뇌가 정보를 처리하는 방식과 유사한 인공 신경세포 네트워크 알고리즘을 사용하며, 이 알고리즘은 뇌에서 영감을 얻어 유사한 분석시스템을 마련하는 것으로 뇌 자체를 모델링하는 것과는 차이를 보임
- 딥러닝을 구성하는 인공 신경세포 네트워크는 이를 구현한 시스템이 얻은 데이터의 패턴이나 특징을 이해하는 데 뛰어나며, 컴퓨터 비전, 음성인식, 텍스트 분석, 비디오 게임 등의 발달이 모두 여기에 영향을 미치고 있음. 따라서 복잡한 패턴과 특징을 가진 사이버범죄 혹은 온라인 결제사기를 확인하는 데에도 도움

※ 참고 : 딥러닝 기술

- 딥러닝은 사물이나 데이터를 군집화하거나 분류하는 데 사용하는 기술이다. 예를 들어 컴퓨터는 개와 고양이를 구분하지 못하나 사람은 구분할 수 있다. 이를 위해 '기계학습'이라는 방법이 고안되었는데, 이는 많은 데이터를 컴퓨터에 입력하고 비슷한 것끼리 분류하도록 하는 기술이다. 이때, 데이터를 어떻게 분류할 것인가를 놓고 많은 기계학습 알고리즘이 등장했는데 '의사결정나무', '베이지안망', '서포트벡터머신', '인경신경망'이 그 예이다.

[그림] 딥러닝 기술을 활용한 '딥 페이스'



[자료 : GIGAOM, 2015]

- 딥러닝은 인경신경망의 한계를 극복하기 위해 제안된 기계학습 방법으로 인간의 두뇌가 수많은 데이터 속에서 패턴을 발견한 뒤 사물을 구분하는 정보처리 방식을 모방해 컴퓨터가 사물을 분별하도록 기계를 학습시킨다.
- 딥러닝 기술을 적용하면 사람이 모든 판단 기준을 정해주지 않아도 컴퓨터가 스스로 인지·추론·판단할 수 있게 되며, 음성·이미지 인식과 사진 분석 등에 광범위하게 활용된다. 이에 따라 구글, 페이스북, 마이크로소프트, 바이두 등의 기업들이 딥러닝 기술을 적용하여 데이터를 분석하는 등 서비스에 활용하고 있다.

2. 페이팔의 딥러닝 기술, 효과는 '톡톡' 전략 가치는 '쑥쑥'

• 페이팔이 도입한 딥러닝 기술, 사이버범죄와 온라인사기를 확인하는데 뛰어난 효과

- 사기방지 전문가와 결합해 '탐정이 하는 것과 같은 방법론(Detective-like Methodology)'을 적용할 수 있게 함
- 전 세계에서 이뤄지는 온라인 결제에서 발견된 수 만개의 잠재적인 특징을 분석해 특정 사기유형과 비교하거나 사기방식을 탐지하고, 다양한 유사수법을 파악할 수 있게 함

- 다양한 사기수법들은 누군가 X라는 일을 했을 때, Y라는 결과가 나오는 것보다 훨씬 복잡하며 이에 따라 사람이 할 수 있는 것보다 훨씬 높은 수준의 분석이 필요
- 딥러닝을 통해 사기가 가능한 모델이 탐지되면 사기방지 전문가는 현실에서 일어날 수 있는 일인지, 다음에 무슨 일이 발생할 수 있는지를 파악

[그림] 페이팔의 온라인 사기 필터링 화면

Edit My Filter Settings

Fraud Management Filters flag potentially fraudulent transactions.

You can indicate whether you would like to **Review**, **Deny**, or **Flag** transactions.

To restore PayPal's default settings, click "Restore Defaults" below.

1. Select your Fraud Management Filters.
2. Choose how to handle filtered transactions and set values (if necessary).
3. Click Save to activate filters.

Activate	Filter Name	Filter Action	Value
Basic Fraud Management Filters			
<input type="checkbox"/>	Maximum Transaction Amount	- Select Action -	0 USD
<input type="checkbox"/>	Unconfirmed Address	- Select Action -	0 USD
<input type="checkbox"/>	Country Monitor	- Select Action -	Edit
Advanced Filters - Card & Address Validation			
<input type="checkbox"/>	AVS No Match	- Select Action -	
<input type="checkbox"/>	AVS Partial Match	- Select Action -	
<input type="checkbox"/>	AVS Unavailable/Unsupported	- Select Action -	
<input type="checkbox"/>	Card Security Code (CSC) Mismatch	- Select Action -	
<input type="checkbox"/>	Billing/Shipping Address Mismatch	- Select Action -	

[자료 : GIGAOM, 2015]

• 페이팔의 딥러닝 모델, '왕'의 자리에 점점 다가서

- 페이팔은 '챔피언-챌린저(Champions-and-Challengers)'방식의 접근법을 사용
- 전통적인 챔피언-챌린저 방식에서, 새로운 전략(챌린저)이 기존에 사용 중인 전략(챔피언)보다 뛰어나다면, 새로운 챔피언 전략으로 선택
- 이런 과정은 결과에 대한 챌린저 모델을 적용해봄으로써 새로운 전략이 나타낼 결과를 미리 예상할 수 있으며, 프로그램의 90%가 챔피언 모델을 통해 처리되는 동안 나머지 10%는 챌린저 모델을 통해 처리됨. 최종적으로 결과를 비교해서 더 우수한 전략을 선택함
- 페이팔은 이런 접근법을 통해 어떤 사기탐지 모델에 무게를 둘지를 결정하고 있으며, 딥러닝에 의한 방식은 새로운 전략(챌린저)에서 출발하여 점점 새로운 챔피언 모델로 자리를 잡음

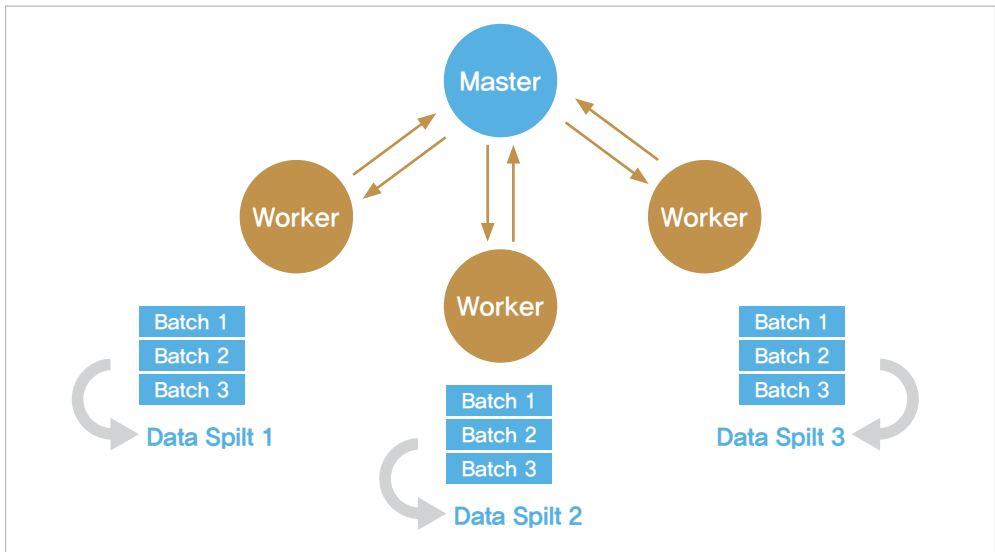
●● 효과 및 향후 적용 확대 방안

1. '딥러닝' 도입으로 온라인 사기 차단 가능성 높여

• 치밀함과 호환성을 담보한 알고리즘

- 다중시스템 클러스터에서의 실행을 위한 알고리즘의 분배는 필요조건이기 때문에 힌튼(Hinton)의 딥러닝 알고리즘을 사용하였으며, 여러 시스템에서의 활용성을 높이기 위해 그라지아(Grazia)가 제시한 가이드라인을 따름

[그림] 페이팔의 딥러닝 알고리즘



[자료 : 페이팔 Engineering, 2015]

- ① 마스터 노드를 통해 RBM(Restricted Boltzmann Machines)의 가중치를 초기화
- ② 마스터 노드를 통해 가중치를 부여하고, 작업자 노드를 쪼갬
- ③ 작업자 노드는 1개의 데이터셋 사건을 위해 1개의 RBM계층을 다룸, 즉 쪼개져 있는 전체적인 작업자 노드를 통해 완성된 부분(업데이트 된 가중치)을 다시 보냄
- ④ 마스터노드는 하나의 주어진 사건을 위해 모든 작업자 노드로부터 얻은 가중치를 평균함
- ⑤ ③단계 ~ ④단계의 과정을 미리 정의된 사건의 집합에 대해 반복
- ⑥ ⑤단계까지 거치면 RBM 1개의 계층이 완성되며, 동일한 과정을 RBM 전체 계층에 걸쳐서 반복
- ⑦ 모든 계층이 완료된 후, 잘 조율된 결과를 다시 전파

- 사이버범죄 및 온라인 사기 패턴 분석의 정확도 제고

- 60,000개의 트레이닝 사진과 10,000개의 실험 사진을 통한 모의실험을 RBM 계층 당 히든 유닛의 수 (1개의 작업 노드에 들어가는 사진의 수)를 증가시킨 결과 히든유닛이 많을수록 Error Rate(%)가 점차 감소
- 온라인 사기 사이트의 데이터를 딥러닝의 알고리즘 통해 분석하여, 사기 사이트 여부를 보다 정확히 판단할 것으로 기대

[그림] 딥러닝 테스트 결과

Hidden Units per layer	Total Units	Error Rate(%)
100 - 100 - 100	300	2.78
500 - 250 - 100	850	2.33
500 - 500 - 250	1250	2.13
500 - 500 - 500	1500	1.98
500 - 500 - 1000	2000	1.77
500 - 500 - 2000	3000	1.66

[자료 : 페이지팔 Engineering, 2015]

03

BIGDATA
해 외 사 례

Statcast로 구현된 빅데이터 야구 실현 메이저리그

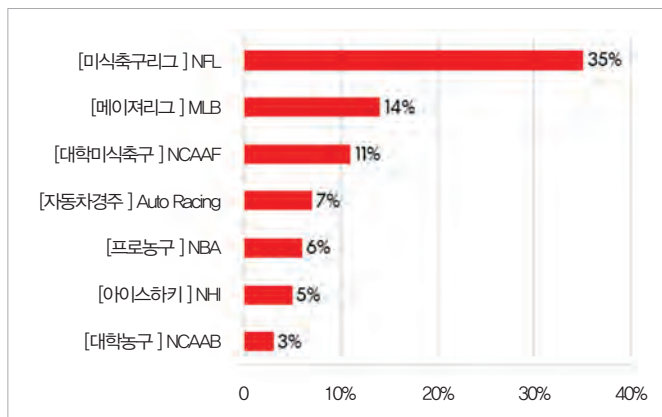
●● 사업의 추진 배경 및 목적

1. 팬들의 외면, 시청자의 고령화, 새로운 대안마련 필요

• 야구에 대한 관심 하락, 젊은 층의 외면

- 146년 역사(1869년 설립)를 지닌 MLB(Major League Baseball)는 1950년대 까지 가장 인기 있는 스포츠였으나 TV의 등장 이후 NFL(National Football League)이 등장하면서 전세가 역전
- NFL의 역동성은 시청률, 스폰서십, 구단용품 판매 등 다양한 분야에서 메이저리그를 앞서기 시작했으며 야구팬의 고령화 및 시청자 수 감소로 향후 대책 마련이 필요
- 2014년 TV로 MLB를 시청하는 팬의 50%는 55세 이상으로 10년 전 41%보다 9%p 증가하였고, ESPN의 시청률 조사에서도 MLB팬의 평균 연령은 53세로 NFL(47세), NBA(37세) 보다 높은 것으로 나타남
- 1978년~1982년 3,800만 명이 월드시리즈를 시청하였으나 2014년 시청자 수는 1,380만 명으로 64%가 급감

[그림] 2014년 미국 내 가장 인기 있는 스포츠 순위



[출처 : FTW(USA Today Sports site)]

경기영상과 기록 – 빅데이터 분석을 통해 야구를 보는 새로운 재미 제공

●● 추진 내용

1. ICT 기술과의 접목을 통해 보다 역동적인 야구 중계를 실현

• 투구, 타구, 선수들의 움직임을 모두 포착하는 Statcast 시스템 구축

- MLB는 2015년 시즌부터 투구는 물론 타구와 선수들의 움직임을 모두 포착하는 Statcast 시스템을 30개 구장 모두에 설치하고, 공의 궤적을 추적할 수 있는 레이더 장비업체 트랙맨(Trackman)과 영상 장비업체 카이론히고(ChyronHego)와 협력
- 덴마크의 레이더 회사 트랙맨(Trackman)에서 들여온 도플러 효과¹⁾를 이용한 레이더 카메라를 이용하여 공의 궤적을 분석하고, 카이론히고(ChyronHego)의 카메라는 모든 선수들을 1초당 30개의 사진으로 찍어 움직임을 추적하고 분석

[그림] Statcast를 통한 투수간 비교 분석



[출처 : MLB]

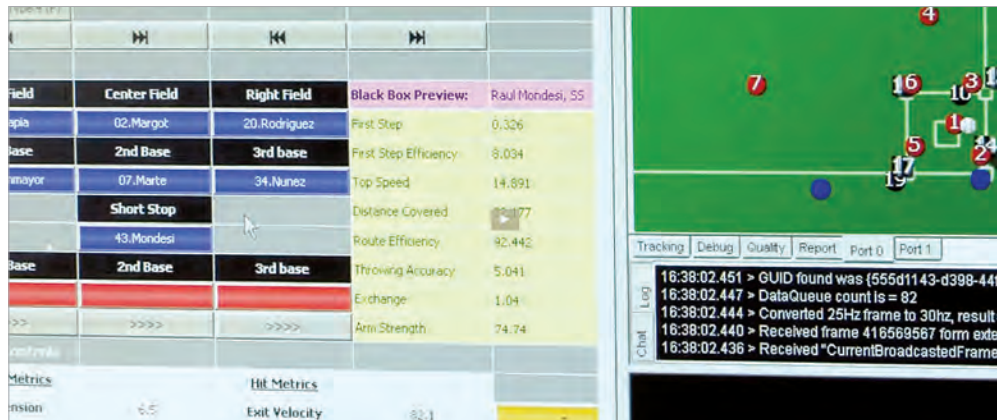
1) 도플러 효과 : 레이더 카메라가 전파를 발송하면 날아오는 공으로 인해 반사전파의 진동수가 증가하는데, 진동수 변화로 속도와 궤적을 측정

- 메이저리그는 Statcast 시스템을 통해 모든 경기 이닝마다 투수의 피칭, 타자의 배팅, 타구에 대한 수비수들의 움직임 등을 추적하고 기록
- 이 엄청난 양의 데이터는 메이저리그를 분석하고 운영하는 방송사는 물론 통계 및 분석 전문 회사를 통해 가공되어 공급됨에 따라 경기를 즐기는 방식이 크게 변화할 것으로 기대

• 데이터 야구에 대한 새로운 기대

- 축적된 기록은 세밀한 통계 분석을 가능하게 해 야구의 흥미를 배가했고 그 흐름은 과학적 통계로 야구를 분석하여 의미 있는 인사이트를 찾아내는 것에 초점을 맞춤
- 투구 분석뿐만 아니라 타구와 선수의 움직임을 모두 처리한 데이터양은 경기당 3TB~7TB에 이르며 미국 메이저리그 사무국의 자회사인 MLBAM(MLB Advanced Media)을 통해 모든 데이터를 무료로 공개하고 있으며 누구나 접근이 가능
- ※ 빅데이터가 공공재가 되면 야구에 대한 지식과 방법론이 리그나 하나의 구단이 주도할 때보다 훨씬 발전될 것으로 기대

[그림] 실시간 데이터 분석을 통한 경기 중계 및 기록 관리



2. 빅데이터를 통해 빠르게 변화하는 MLB

• 수치화된 실시간 데이터, 새로운 야구 중계를 선보여

- Statcast 시스템은 투구의 속도와 궤적, 공의 회전 방향부터 투수의 보폭과 자세를 보고 타자가 예측하는 속도와 어떻게 다른지 까지 분석

- “투수의 공 끝이 밋밋하다”, “배트의 스피드가 부족하다” 등 해설자의 개인적인 표현이 중심이던 이전 중계와 달리 구속(km/h), 타구속도(km/h), 비거리(m) 등 실질적인 수치 중심의 중계가 가능

• 보다 역동적인 방송 중계를 실현

- Statcast 시스템은 모든 선수들을 1초당 30개의 사진으로 찍어 움직임을 추적· 분석, 타격 후에는 기 후에 따라 어떻게 공이 날아가며, 수비수들의 반응과 공이 떨어지는 지점까지 얼마나 효율적으로 움직였는지도 추적하여 보다 역동적인 중계가 가능
- 또한, 경기장 안에서 선수와 공의 움직임을 다양한 각도에서 시각화된 이미지를 추가해 보여주어 보다 흥미롭게 TV 중계
- 팬들의 경우 스마트폰을 통한 심판의 아웃/세이프 판정 등이 정확하게 이루어졌는지 곧바로 살펴볼 수 있는 실시간 인스턴트 리플레이 등 이전보다 고품질, 고차원의 야구 중계 경험이 가능

[그림] Statcast를 통한 TV 중계 시연



[출처: Phys.org]

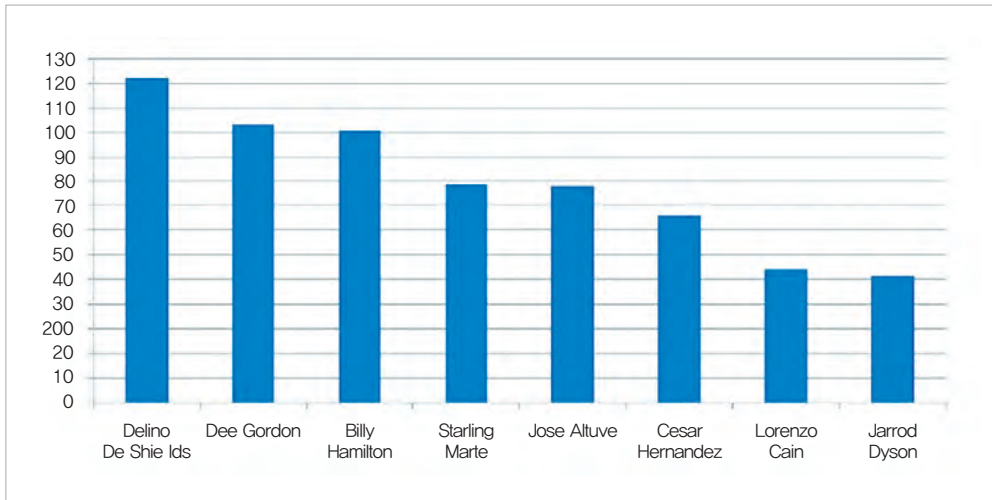
●● 효과 및 향후 적용 확대 방안

1. 빅데이터, MLB 구단의 새로운 동력으로 부각

• Statcast, 선수에 대한 객관적 평가 자료로 활용

- 첨단 ICT 기술과 빅데이터가 결합한 Statcast 시스템은 선수 개개인의 객관적 평가 자료로 활용 가능
- 예를 들어 2015년 MLB 선수들 중 가장 빠른 선수로는 도루왕 디고든으로 알려져 있으나 Statcast를 통해 실질적인 통계를 분석하니 텍사스의 델리노 드실즈가 가장 빠른 선수로 나타남
- 델리노 드실즈는 Statcast로 스피드를 측정한 결과 121경기에서 110번 출장하여 21마일(33.8km/h) 이상 속도의 주루플레이를 총 132번 기록
- Statcast를 통한 데이터는 선수의 기량을 정확히 측정·파악하거나 향상하는데 도움을 주고, 상대팀의 투수와 타자를 분석하여 경기전략 마련이 가능하며, 선수영입에 있어 빅데이터를 이용하여 시장에서 저평가된 선수를 영입하여 예산의 효율적 사용이 가능

[그림] MLB 주요 선수들의 21마일 이상 주루 플레이 현황



[출처 : BaseBallGen.com]

• 빅데이터 기술은 경기, 선수 운영 측면뿐만 아니라 대형 스포츠의 팬 서비스와 마케팅 목적으로 활용이 가능

- 빅데이터 기술은 경기분석과 선수의 경기력 향상, 팀의 승률 예측, 경기 전략 수립, 스포츠 방송, 광고 등 다양한 영역에서 적용되면서 스포츠 비즈니스의 양상을 변화

- 팬들은 개방된 빅데이터를 통해 직접 대량의 정형/비정형 기초데이터로부터 새로운 의미를 추출하고, 가상적으로 조작하여 결과를 분석하는 등 새로운 차원의 경험이 가능

2. MLB, 빅데이터를 통해 새로운 비즈니스 창출을 기대

- 빅데이터 효용은 데이터 분석을 통해 고객이 만족하는 새로운 상품이나 서비스 개발에 있음
- Statcast를 통해 만들어진 빅데이터는 다양한 채널을 통한 야구 중계는 물론 게임, 마케팅, 스포츠 교육 등 다양한 분야에서 활용이 가능
- Statcast는 아직 초기 단계로 향후 기술이 진보하면서 야구의 새로운 진화는 물론 다양한 산업에서 활용될 것으로 예상

04

BIGDATA
해 외 사 례

안전한 여행을 위한 빅데이터 분석

Travelbasys

●● 사업의 추진 배경 및 목적

1. Travelbasys 소개

- 1972년 설립되어 '효율적인 관광 프로세스에 대한 귀하의 IT 파트너'이라는 슬로건 아래 혁신적인 IT 시스템 개발 솔루션을 여행 관련 산업에 공급
- 1천 2백 개 이상의 여행 회사와 IT 통합 및 자동화 서비스를 성공적으로 지원해 왔으며, 80억 유로 이상의 매출액과 1천 2백만 관광지 데이터와 2천 2백만 명 이상의 여행자 프로파일을 보유하고 있으며 매 월 새로운 정보가 갱신
- 여행 산업에 대한 중요한 IT 시스템 솔루션을 간단하고 저렴하게 공급하여 고객이 사용할 수 있도록 세상의 모든 여행 업체와 제휴하여 상호 성공하는 비즈니스 라이프 관계를 추구

2. 걱정 없이 여행을 즐길 수 있는 방법이 있을까

• 여행 중 사고를 당하는 여행자가 증가함에 따라 실시간 여행객 보호를 위한 자동 경보 시스템 개발 필요성 대두

- 소셜 네트워크에서의 여행 관련 데이터는 불확실한 정보가 많아 신뢰성 확보가 어려움
- 보도 자료나 기관이 보유한 자료를 실질적으로 유용한 데이터로 가공하여 신뢰성과 정확성을 갖춘 여행지 관련 위험 정보 확보

• 대한민국 해외여행 정보 안내 서비스

- Travelbasys의 빅데이터 기반 여행 정보 제공의 큰 특징은 실시간으로 관련 정보를 최신화하여 공급하기 때문에 여행자들이 자신의 환경에 맞도록 상황 반영 가능

여행객의 안전을 위한 실시간 여행 자동 경보 시스템

- 국내에서는 외교부 주관으로 해외여행 안전과 관련된 정보를 제공하고 있지만 업데이트가 일정 간격으로 진행되기 때문에 최신 정보 반영은 상대적으로 속도가 느림

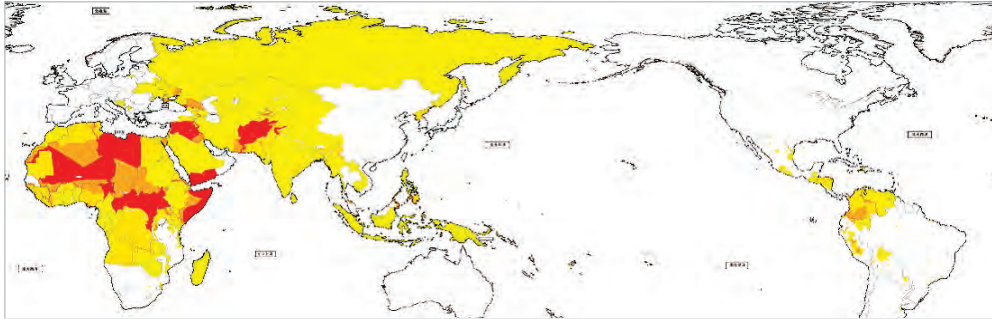
[그림] 해외여행 안전 단계 안내 서비스(외교부)



[출처 : 대한민국 외교부 해외안전여행(www.0404.go.kr)]

- 안전한 여행을 위한 정보 안내를 위하여 오랜 시간 축적된 데이터를 환경에 따라 체계적으로 분석하여 여행객에게 위협이 되는 요소를 사전에 인지
 - 다양한 경로에서 수집되는 자료를 통해 여행 위험 데이터를 확보하고, 기존의 데이터와 비교하면서 신뢰도 향상에 밑거름이 될 수 있는 데이터로 가공

[그림] 2015년 전 세계 여행 위험 지역 표시



노랑 - 여행 주의(한국으로 치면 여행 유의) 특별한주의 요청
 귤색 - 여행 연기(한국으로 치면 여행 자제) 불필요한 여행 자제
 주황 - 여행 중단(한국으로 치면 옛 여행 제한 & 철수 권고) 여행을 멈추어야하는 지역
 빨강 - 대피 권고(한국으로 치면 여행 금지) 머물러있다면 즉시 나가야 할 지역

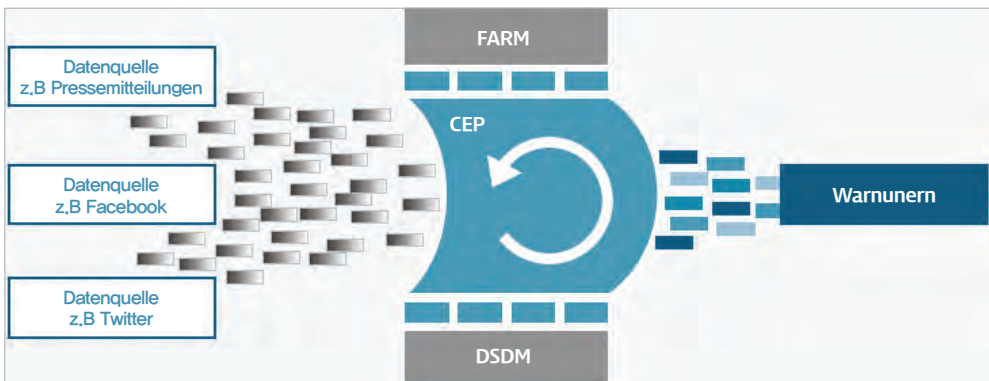
[출처 : 나무위키]

●● 추진 내용

1. 지역 정보 및 관련 정보를 통해 위험도 예측 알고리즘 개발

- 언론 보도나 소셜 네트워크에서 확보된 데이터에 폭동, 자연 재해, 전염병 등의 위험 요소를 입력하여 실시간 데이터 처리를 통해 여행 경로에 대한 위험도 확인
 - 데이터 분석을 거쳐 위험 요소를 도출하여 현재 여행객의 위치 등 상황에 맞는 정보를 제공하고 실시간 업데이트를 통해 신속하게 적용하여 실시간 안전 상황을 확보

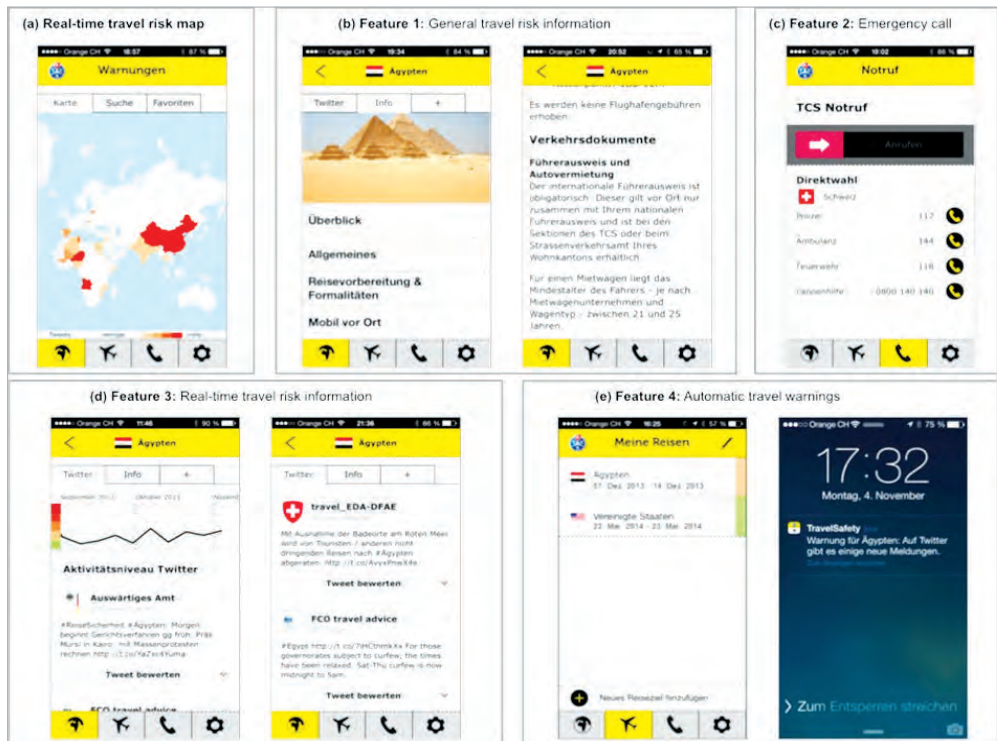
[그림] 개별 소스의 대량 데이터 처리 모델



[출처 : Travelbasys, 2015]

- 여행 일정과 경로에 따른 위험도를 분석하고 관리하는 Back-office Solution을 결합하여 전체 여행 위험도에 대한 데이터 분석 시스템 구축
- 수집된 자료는 DB화하여 빅데이터 기반의 맞춤형 정보로 재가공 되고, 이를 토대로 여행객에게 필요한 정보를 제공
- 자동화된 데이터 분석을 통한 실시간 상황인식으로 여행 위험성을 경고(날짜, 지역 등)
 - 페이스북, 트위터 등의 사회 관계망 서비스와 보도 자료에서 수집된 최신 데이터를 자체 알고리즘에 따라 분류 - 검증하고 이들 대량의 데이터를 분석하여 여행의 대체 노선(경로) 제시
 - 여행에 위험을 미칠 수 있는 요소들을 특정 주제나 키워드를 통해 실시간으로 메시지 탐지
- 특정 날짜에 대한 위험 정보와 환경 안내
 - 여행자의 안전과 관련된 데이터가 필요성과 시기에 따라 자동적으로 필터링됨
 - 여행 위험에 영향을 미칠 수 있는 관련된 행동 패턴이나 정보를 구체화하여 제공

[그림] 모바일로 제공되는 여행지 위험 정보



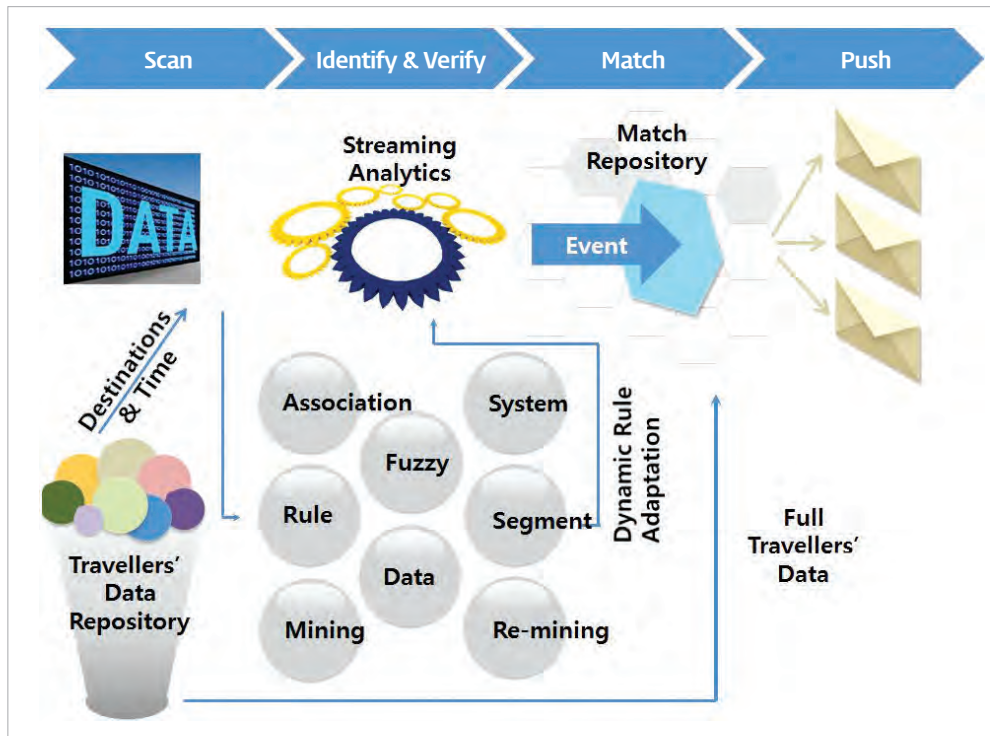
[출처 : 베를린 하이델베르크]

2. 작은 것도 놓치지 않는 세심함

- 새로운 서비스를 기존 보다 저렴한 비용으로 제공하여 많은 여행객이 혜택을 받을 수 있는 체계 구성

- 소셜 기반 데이터 및 자체 구축 DB 뿐만 아니라 지리 정보 및 지역 정보 관련 기업과의 데이터 교환을 통해 위험 정보의 정확도를 향상시키고 예측력을 보다 강화하는 방향으로 분석을 진행
- 수집된 여행 관련 데이터를 저장하여 각각의 성격에 맞게 분류하고, 적절하게 가공하여 새로운 정보를 원하는 이에게 도움이 될 수 있도록 즉시 발송하는 체계 구축

[그림] 수집된 데이터가 가공을 거쳐 이용자에게 전달되는 과정



[출처 : ITESA]

●● 효과 및 향후 적용 확대 방안

1. 데이터의 축적 및 알고리즘 개선을 통한 서비스 확장

- 누적 데이터 분석을 통해 보다 정밀한 위험 예측 알고리즘을 개발하여 관련된 서비스의 확장 및 개선 계획
- 개별 여행에 한해 단편적 정보만을 다룰 수 있는 수준에서 벗어나 향후에는 다양한 검색 및 조건으로 보다 신속한 여행 위험 경고가 가능한 단계로 발전시킬 계획

05

BIGDATA
해 외 사 례

센서를 이용한 기관차 관리 DB Systel

❖ 사업의 추진 배경 및 목적

1. 철로 관련 사고는 언제나 대형사고 - 사전 예방 필요성 대두

- 시설 유지 및 기관차 문제에 따른 철도 운송 중단 예방을 위해 실시간 상황 탐지 필요성 대두
 - 독일 내 5천 7백 개의 기차역과 3만 4천여 km에 이르는 철도 네트워크가 원활히 작동하기 위해서는 철도 및 기관차의 유지와 보수를 위한 예측과 분석이 반드시 필요하며, 이 문제를 매우 긴급하고 중요한 문제로 인식

[그림] 철로, 기관차 보수 및 정비



[출처 : DB Systel, 2015]

- 정비 관련 숙련자에 의지하는 예방 및 정비를 데이터 분석에 기반한 시스템으로 변경하기 위해 다양한 데이터 수집 방법과 분석 방법을 연구

기관차의 원활한 운행과 순환 시스템 확립을 위한 빅데이터 활용

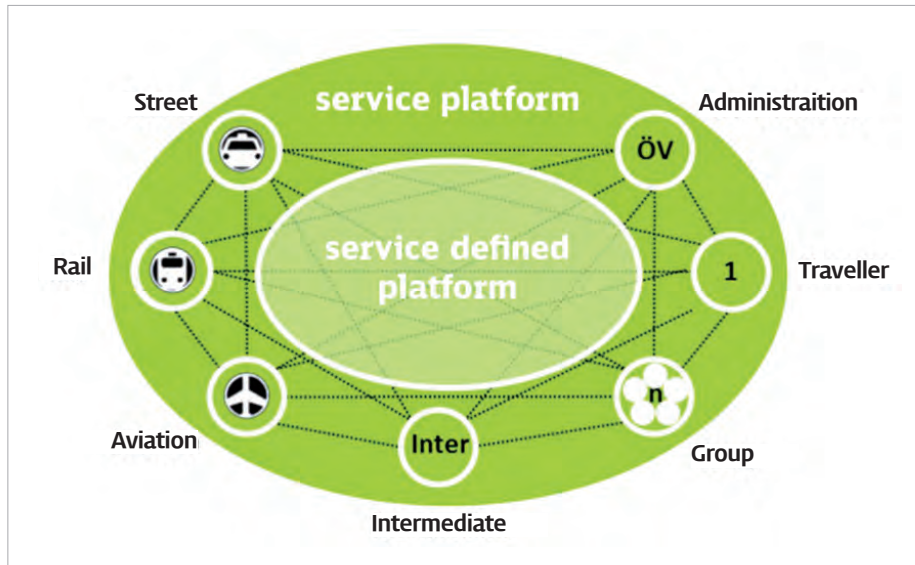
[그림] DB Systel의 빅데이터 센서를 부착한 기관차의 역내 진입



[출처 : DB Systel, 2015]

- 철로에서의 열차 사고는 상당한 비용, 운송 지연 그리고 만족/신뢰도 하락의 원인
 - 철도가 독일 사회에서 차지하는 비중을 고려할 때, 한 기관차의 연착은 전체 철도 스케줄에 악영향
 - 철로나 기관차 손상을 복구하기 위한 부품 교체 값은 매우 고가이며 단시간 내에 수리를 기대하기 어려운 시스템

[그림] 타 교통 및 이용객과 연계된 시스템 구축



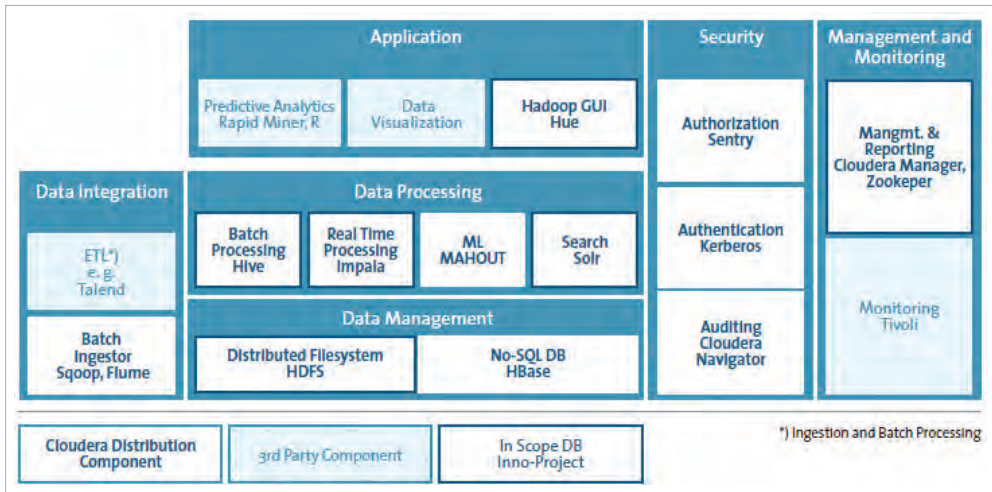
[출처 : DB Systel, 2015]

❖ 추진 내용

2. 기관사의 '경험'보다 데이터 분석 기반의 관리시스템 추진

- 운영 상황 정보를 실시간으로 담아 인지 가능한 빅데이터로 즉시 가공
 - 상황 진단 과정에서 사람의 예감이 아닌 근거에 기반한 효과적 조치 가능
 - 부착된 센서가 기관차나 시설의 상태 변화를 감지하거나 환경을 신속하게 분석하여 빅데이터화
 - 문제 발생 시, 센서가 패턴을 감지하여 관리자가 디스플레이 형태로 인식 가능하도록 전송
- 기관차에 부착된 센서를 통해 기관차 유지, 보수에 필요한 온라인 진단 데이터 시스템 및 관련 인프라 구축
 - 유지와 보수가 필요한 부분에 대해 일정 기간 그 내용을 리스트로 만들어 예방 · 정비에 활용하며 작업장에서 유지 · 보수를 진행하는 순간에도 데이터 수집이 가능

[그림] 기관차 관리 빅데이터 플랫폼 체계



[출처 : DB Systel, 2015]

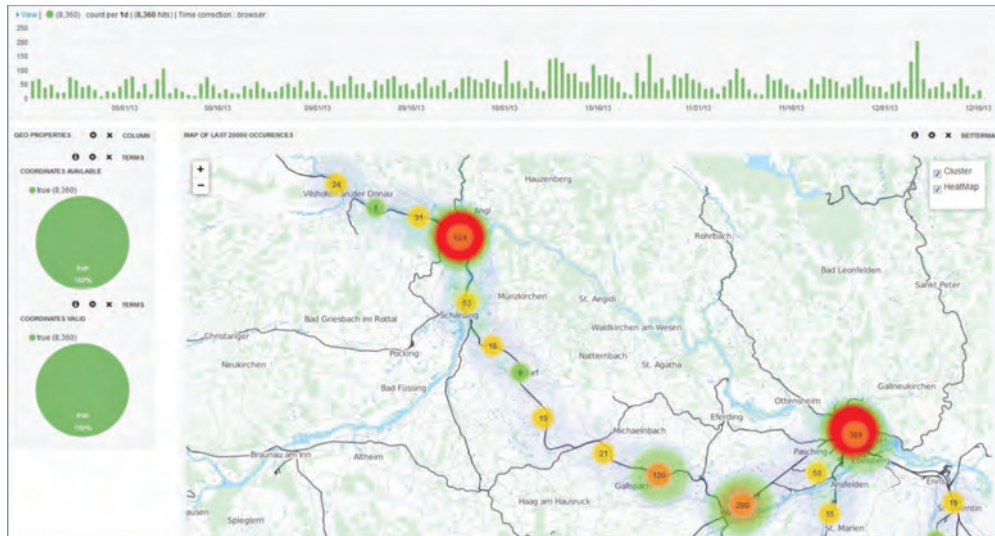
- 기관차의 운행·상황 관련 데이터를 지속적 생성하고 관리자가 인지
 - 기관차 센서에서 데이터가 일정 간격으로 전송되나 관리자는 상황에 따라 수동 컨트롤 가능
 - 문제를 빠르게 파악하여 유지와 보수에 정기적으로 적용하고, 정확한 대처가 진행 될 수 있는 관리 분야의 유용성 확보
 - 실시간 모니터링으로 상황을 신속, 정확하게 파악한 후 원격 진단이 가능한 체계 수립

❖ 효과 및 향후 적용 확대 방안

1. 우리 곁에 안전한 철도

- 센서에서 생성된 데이터에 기반한 실시간 통합형 운행 스케줄 관리
 - 네트워크, 기상조건, 에너지 공급, 작업장 데이터, 스케줄 데이터 등을 통합한 실시간 관리로 운행 스케줄 관리 가능
 - 축적되는 데이터의 특징을 분석하여 발생 가능한 사고나 재발을 차단할 수 있도록 잠재적인 사고에 지속적으로 대비할 수 있는 체계 마련

[그림] 기관차의 문제 상황을 실시간으로 안내하는 디스플레이



[출처 : DB Systel, 2015]

- 여객·화물 수송의 안정적 운영을 위해 과거에 진단된 문제의 특징이 보관된 빅데이터를 토대로 예측 모델 개발 계획 추진
 - 사고 발생 시, 그 특징과 유형을 즉시 데이터로 저장하여 과거 사례와의 공통점이나 차이점을 면밀히 분석 후 적합한 조치로 이어지는 매뉴얼 마련
- 날씨나 노선 상태, 환경 등의 요소를 인위적으로 설정하고, 주기적 시뮬레이션 테스트를 진행하여 문제 발생을 최소화
 - 예측 자체가 매우 어려운 천재지변 상황이 발생했을 경우에 현실적으로 가능한 대비 체계와 방침 마련
 - 피해가 예상되는 돌발적인 상황에서의 피해를 최소화하고, 신속한 정상 운영을 위한 컨트롤 타워 구성

빅데이터 활용으로 생산성 향상
오므론

❖ 사업의 추진 배경 및 목적

1. 오므론 소개

- 산업용 제어기기 및 헬스 케어 제품을 주력으로 개발 생산하는 1948년에 설립된 일본 교토 소재 전자 회사로 생산 라인에 사용되는 기기를 통합 제어하는 Sysmac²⁾ 자동화 플랫폼을 통해 제어기기 분야를 선도하는 지위에 있는 기업
- 제어기기 · FA(공장 자동화) 시스템 사업이 일본 내 점유율 40%를 차지하고, 유럽, 북미, 아시아 등 세계 80여 개국에서 사업 전개 중이며 최근 다양화되는 제조 현장의 경영 과제를 해결하기 위한 '솔루션 사업'에 주력

2. 단위 시간 내 더 높은 생산성 확보 방안 고민

• 숙련자에 의한 제조 공정 관리상의 비효율성 개선 필요

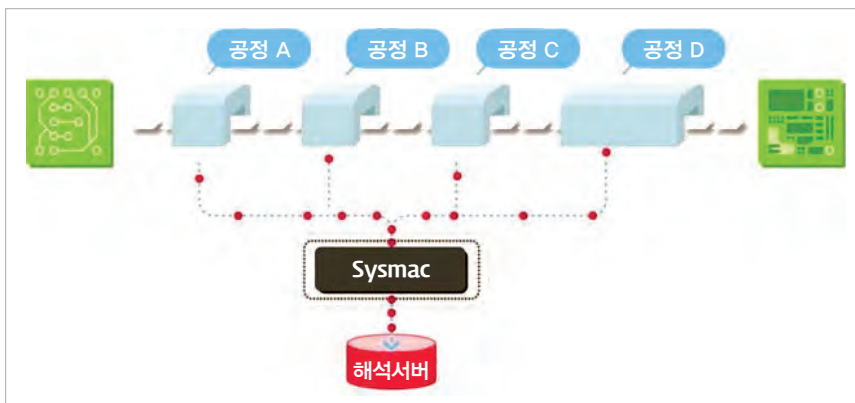
- 생산라인의 상태를 쉽게 파악하여 현장 직원이나 관리자 누구든지 시스템의 오류를 인지할 수 있는 환경을 조성하고, 전체적인 진행 상황을 파악할 수 있는 시스템의 고안 필요성 대두
- 가동률 및 생산 리드타임(Lead Time)³⁾의 결과에 대해 현장에서 시각적으로 즉시 확인하거나 필요한 경우 자료를 Excel이나 Web에서 확인할 수 있는 형태로 변환하여 분석할 수 있는 체계를 구성

2) 기존의 PLC(Programmable Logic Controller) 기능과 모션 제어 기능을 하나로 묶은 오므론 주식회사의 국내외에서 사용되는 등록 상표 명칭

3) 생산을 계획하여 공정에 착수하는 시점부터 제품이 완성되어 완제품 창고에 입고되는 시점까지의 기간

생산라인을 통과하는 모든 제품의 상황과 모습을 ‘타임 라인 데이터 시각화’로 가시화하여 관리

[그림] 생산라인의 상황을 최신화 하는 데이터 실시간 전송 구조도

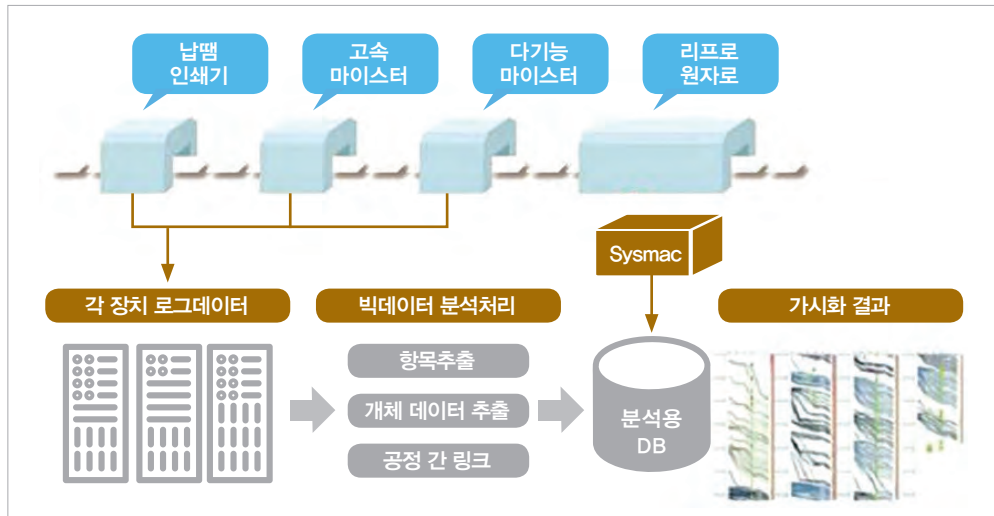


●● 추진 내용

1. 비효율성이 나타나기 시작하는 생산라인

- 생산라인의 장치에서 시계열 데이터를 단계별로 변환하여 생산 상황을 개체별로 분석할 수 있는 여건 마련
 - 로그 수집 구조를 구축하여 유효하게 실시간으로 생산 상황을 파악하고, 데이터를 분석
 - 각 장치가 방출하는 로그 데이터를 통합하여 생산 공정을 일련의 흐름으로 시각화
- 프로토타입 시스템 도입으로 생산라인 움직임과 개선점 감지
 - 실증 실험에서 얻은 노하우를 바탕으로 실시간 이상(異常)을 감지
 - 생산성에 문제가 발생하는 시간이나 공정을 정확하고 상세하게 파악

[그림] 생산라인 실증 실험 진행도



[출처 : 오므론, 2015]

- PCB 제조 라인은 부품의 위치와 납땜이 필요한 부분을 그림으로 나타낸 '납땜 인쇄기', 전자 부품을 회로 기판에 접합하는 '고속 마이스터', 부품이 접합된 회로 기판을 필요한 부분에 위치시키는 '다기능 마이스터', 그리고 위치된 회로 기판에 솔더(납땜)를 녹여 접합하는 '리프로 원자로'의 네 단계로 구성
- 제조 시설마다 제각각 존재하는 단편적 기록들을 특징 있는 데이터로 재생산 가능

• 품질 및 생산성 향상을 위한 솔루션-하드웨어 기술 접목

- 전기회로가 편성되어 있는 기판 표면 라인의 품질 및 생산성 향상을 위해 마이크로소프트의 솔루션과 후지쯔의 하드웨어 기술 접목
- 생산라인의 빅데이터 분석을 통한 시각화 및 개선점 추출
- 생산라인의 각 장치에서 발생하는 데이터와 조립라인의 흐름을 연계하여 시각화 구조 구축
- 실증 실험에서 얻은 노하우를 바탕으로 실시간 단위의 이상(異常)을 감지하여 라인 제어 방법을 연구하고 적용

• 분석 플랫폼 제공으로 생산 정보를 다양한 측면에서 분석 할 수 있는 다양한 데이터베이스 구축

- 데이터 시각화로 현장 직원들의 의식에 가장 큰 변화가 나타남
- 생산라인을 살펴보는 시야가 넓어짐에 따라서 개인 담당 부분뿐만 아니라 생산라인 전체의 최적화를 위한 생각을 제안하기도 하여 1시간 당 생산 능력 향상에 도움

[그림] 실시간 생산라인 현황 파악이 가능한 시스템



[출처 : 오므론, 2015]

- 생산라인의 현재 상태가 나타나는 시각화 데이터를 보면 누구나 자신의 역할과 오류발생에 주의할 수 있으며, 생산라인의 현재 상황을 매우 수월하게 파악 가능하도록 구축

2. 비효율화 요소 및 지점 파악 수월

- 모든 제품의 생산 상황이 파악 가능한 시각화 화면 제공으로 효율과 비효율 부분의 포인트 파악 가능

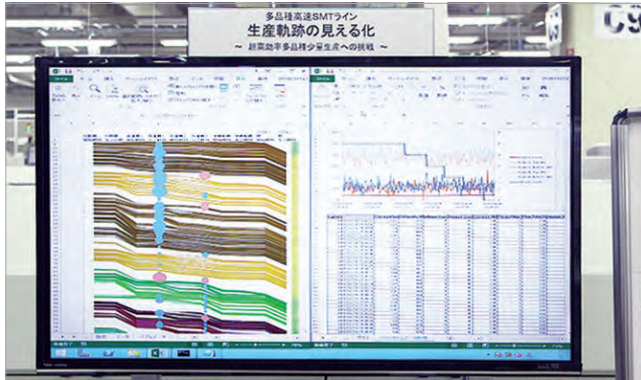
- 시각화 테이블에 나타나는 라인의 간격이 좁아 간격이 빠곡한 경우 생산라인이 순조롭게 작동하는 상황인 반면, 라인의 간격이 넓다면 어떤 문제가 발생했음을 나타냄. 이러한 상황은 숙련자의 직감과 경험으로 쉽게 파악이 불가능했던 것으로, 개선점 파악이 보다 수월해짐

[그림] 생산성 효율성을 나타내는 시각화 결과 이미지



[출처 : 오므론, 2015]

[그림] 타임 라인 데이터 시각화에서의 생산 상황 화면



[출처 : 오므론, 2015]

- 제품 생산에 어느 정도의 시간이 필요한지 표현하며 공정마다 발생한 오류 수 등을 버블 차트로 축적하여 생산라인에서의 원활한 작동이 가능한지 여부를 실시간으로 확인

● 효과 및 향후 적용 확대 방안

1. 효율성 6배 상승효과

- 기존 6명 이상의 숙련자가 필요하던 문제 원인 분석이 빅데이터 시스템 도입으로 1명의 인력으로 충분히 대응 (효율 6배 이상 증가)

[그림] 효율성 전후 비교



[출처 : 오므론, 2015]

– 생산성 저하 발생 공정과 그 원인이 보다 쉽게 파악되어 시간당 생산성이 30% 향상 되었고, 향후 개선 가능성은 더욱 높은 수준이 될 것으로 기대

• 보다 혁신적인 성능 향상을 목표로 Upgrade된 라인 관리 방법 연구 개발 계획

– 검사 공정 이미지 데이터 연계, 로그 품질 등 다양한 데이터를 매시업 하여 현장 기술 개선 또는 분석 등 다양한 분야로 활용 영역을 확대해 나갈 계획

[그림] 시간 당 생산성 증가율

