1. Value Chain 전반의 빅데이터 활용 동향

1. 정보통신 기술의 발전과 빅데이터

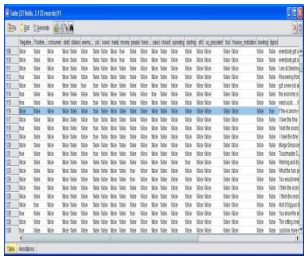
H/W, S/W와 이와 연관된 서비스 기술의 발전이 빅데이터의 출현을 가져오게 되었음.

	PC	인터넷	스마트폰	?
패러다임 변화	디지털화, 전산화	온라인화, 정보화	소셜화, 모바일화	지능화, 개인화, 사물정보화
IT 이슈	PC, PC통신 데이터베이스	초고속 인터넷, www, 웹서버	모바일 인터넷, 스마트폰	빅데이터, 인공지능, 사물인터넷(IoT)
핵심분야 (서비스)	PC, OS	인터넷 포털, 검색엔진	SNS, 앱서비스	클라우드, AI기반 제품, 개인맞춤형 서비스
IT 비전	1인 1PC	인터넷 카페, 가상화	손안의 PC, 정보공유, 개인화	사물인터넷, 新가치창출
리더 기업	Microsoft, IBM	Google, NAVER	facebook, twitter	?

참고 : 정지선, 빅 데이터의 새로운 가능성과 대응전략, 2012. 3

2. 빅데이터란?

빅데이터란 규모가 방대하고, 생성 주기도 짧으며, 형태도 수치데이터 이외에 문자와 영상 데이터를 포함하여 기존의 방법으로는 수집/저장/검색/분석이 어려운 데이터를 말함.

규모(Volume)	다양성(Variety)	속도(Velocity)
MB, GB단위 Data ↓ 테라급 이상의 Data 	정형 데이터 → 비정형 데이터  숫자Data (R-DB[1]관리) 텍스트, 동영상, 음악, 소셜미디어, 위치정보, 센서 Data	발생과 활용의 Gap (몇시간~몇주) ↓ 실시간 분석/대응 real time
표본분석 → 전수분석 정보의 왜곡이 줄어듦	SNS상의 데이터를 확보하고 새로운 분석방법을 적용하여 새로운 Insight발견	실시간 이슈 감지/대응 (Nowcasting)

가치(Value)

주1) Relational Database의 약어로 관계형 데이터 베이스를 의미함

3. 제조업에서 빅데이터의 활용분야

제조업 Big Data 출현

글로벌 생산지 전략

Value Chain을 모니터링하는 다양한 IT시스템은 대량의 복잡한 데이터를 생성함

센서데이터의 증가

공정복잡도 증가	센서기술 발전

데이터가 증가할수록
기존 RDB[1)]관리방식에서는
처리효율에 문제가 발생함

활용분야

R&D	▪ 제품탑재 센서 log data 분석 - 발견 어려운 결함 감지 및 고객의 자사제품 사용패턴 추출/분석 → 제품개발 반영 (차기 모델의 성능개선 및 차별화 전략 수립 반영) ▪ **개발시험 log data 분석** - 설계/성능테스트 관련 정보의 연계 분석 → 既개발 이슈사항 도출 (개발초기 예상되는 이슈사항 先대응)
생산	▪ **생산설비의 log data 분석** - **생산라인의 설비데이터 실시간 분석/대응** (분석시간 단축 및 분석대상 확대 : **불량 원인 분석 → 생산중단 최소화)** ▪ **생산설비 이상감지/대응** - **생산현장의 장비의 고장을 사전에 감지하여 대응함 → 장비실동률 개선**
품질	▪ **고객 Claim 데이터 분석** - **품질불량 분류 및 관련부서 Warning &** 신속대응
SCM	▪ **Demand Forecast & Supply Planning** - **다양한 수요관련 데이터 분석 → 수요변동에 대한 정확한 예측**
마케팅	▪ **제품에 대한 소비자와 전문가의 사용경험/느낌 분석** - **제품에 대한 포지셔닝 전략 수립 → 마케팅 캠페인 실시**

1) Relational Database의 약어로 관계형 데이터 베이스를 의미함

4. 빅데이터 기술

빅데이터를 다루는데 필요한 기술은 하기와 같음.

기능	기술
원본데이터 저장	대용량 분산 파일 시스템(Hadoop File System)
구조적 데이터 저장	대용량 분산 데이터 저장소(NoSQL : Hbase, MongoDB 등)
배치분산 병렬처리	MapReduce(Hadoop)
실시간 데이터 처리	Spark, Storm
데이터 분석	Python, R, Deep Learning Tool(TensorFlow, Keras, Pytorch)
분산관리	ZooKeeper
전통적 솔루션	RDBMS, BI/DW, SAS, Minitab
데이터 분석 알고리즘	기초통계분석, 다변량분석, 기계학습, AI(딥러닝, 강화학습), 패턴인식, 데이터 마이닝

5. 적용사례

시장상황 예측	트위터 분석을 통하여 확보한 투자심리를 활용하여 '트위터 펀드' 개발·운용
숨은니즈 발견	기업의 서비스에 대한 고객의 사용패턴을 분석하여 맞춤형 신규서비스 개발
실시간 니즈분석	전 세계 매장의 판매와 재고 데이터를 실시간으로 분석하여 최대매출 달성
행동변화 예측	모바일 데이터를 활용해 시간단위로 고객의 이동경로 분석 및 물품 배치 최적화
실시간 모니터링 및 이상감지 대응 (제품,생산라인)	제품에 대한 원격 진단 및 응급 상황 대응 서비스 제공 생산라인 설비이상감지 및 대응 필드불량 분석 및 대응 Recipe, 설비, 품질이상 감지 및 조기대응(EWS = RMS + FDC + R2R)
마케팅 (고객감성)	기업이 자사 이미지에 부정적 영향을 실시간으로 파악해 원인 분석/대응
공공 (재난대응)	기상 센서에서 실시간으로 취합되는 데이터 분석을 통한 재해 예측 대응

6. Big Data의 추진방향

소규모라도 실제 가치를 만들어 낼 수 있는 분야에서 시작하여 장기적 관점에서 Big Data 활용환경을 만들어가야함.

Big Data 활용기반 마련	
Data 확보	양질의 대량 데이터를 실시간으로 확보할 수 있는 Channel 마련
분석 인력 양성	개별 Domain과 연계하여 데이터 분석을 수행하고 Value를 창출할 수 있는 인력 양성
분석 Infra 확보	각 Domain에 맞는 사용성이 편리한 분석 Solution 개발 및 확보

Big Data 분석 적용 및 Value창출	
개발	상품기획 : 고객 요구사항 분석, 경쟁사 분석, 제품가격 포지셔닝
	제품개발 : 성능, 신뢰성 예측 및 개선Point도출
	UX분석 : 사용자 Log분석 → 사용패턴 분석 → 차기제품 반영
생산	품질개선 : 생산경로/조건분석 → 품질이상 원인 규명
	설비진단 : 설비상태 데이터 분석 및 이상예측 및 대응
품질	고객Claim분석 : 불량분류 → 불량 원인규명 → 신속대응
	대외평판 관리 : Consumer Union순위, 소셜미디어 평가 분석
기타	마케팅 효과 분석, 자재 가격동향 예측, 구매비용 사용분석

7. Data Science란?

컴퓨터 도구를 효율적으로 이용하고, 적절한 통계학 방법과 기계학습 방법을 사용하여 실제 문제에 대한 해답을 찾아내는 활동을 말함.

데이터 과학이란?

❶ 실제적인 문제를 통계적으로 표 현하고,

❷ 컴퓨터 도구를 사용하여 시각화와 데이터 가공과 모형화를 한 후에,

❸ 이러한 결과물을 이용하여

실제적인 언어로 의미 있는 결과를 만들어내는 능력의 조합이다.

여러 조직의 사람들과 협업을 해야 하는 작업이며
치밀한 분석 역량과 Insight를 도출할 수 있는
사고를 할 수 있어야 함

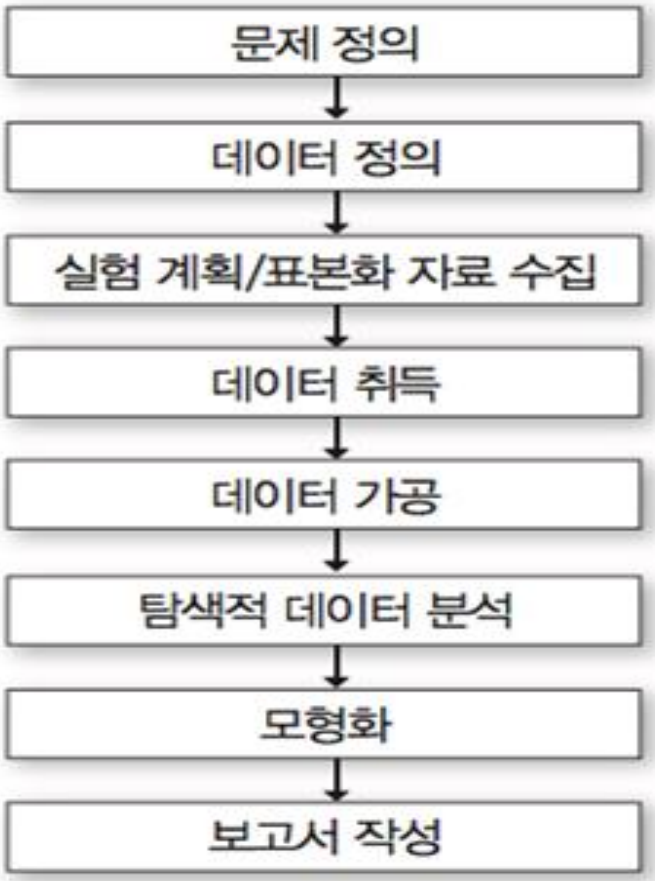
대상과제

- 주택 가격을 예측하는 방법은?
- 초등생 자녀의 수학 능력과 상관 관계가 높은 변수는 무엇일까?
- 훌륭한 직원을 뽑는 인터뷰 방법은 무엇일까?
- 괴혈병의 치료법은 무엇일까?
- 산욕열의 원인은 무엇일까?
- 웹사이트를 개선하는 방법은?
- 신약이 혈압을 낮추는 데 효과가 있을까?
- TV 광고가 제품 판매에 얼마만큼의 영향을 줄까?
- 온라인 광고에서 클릭 여부를 예측하는 방법은?
- 집 안에 있는 수영장과 권총 중 어느 것이 어린이에게 더 위험할까?
- 대학 입학에 남녀의 성차별이 있을까?
- 비싼 와인이 더 맛있을까?
- 아버지의 키가 180cm라면 아들의 키는 얼마일까?
- 대학 진학을 할 때 전공이 중요할까, 학교가 중요할까?
- 흡연은 몸에 해로울까?
- 투자신탁과 ETF 인덱스 펀드 중 어느 곳에 투자하는 것이 좋을까?

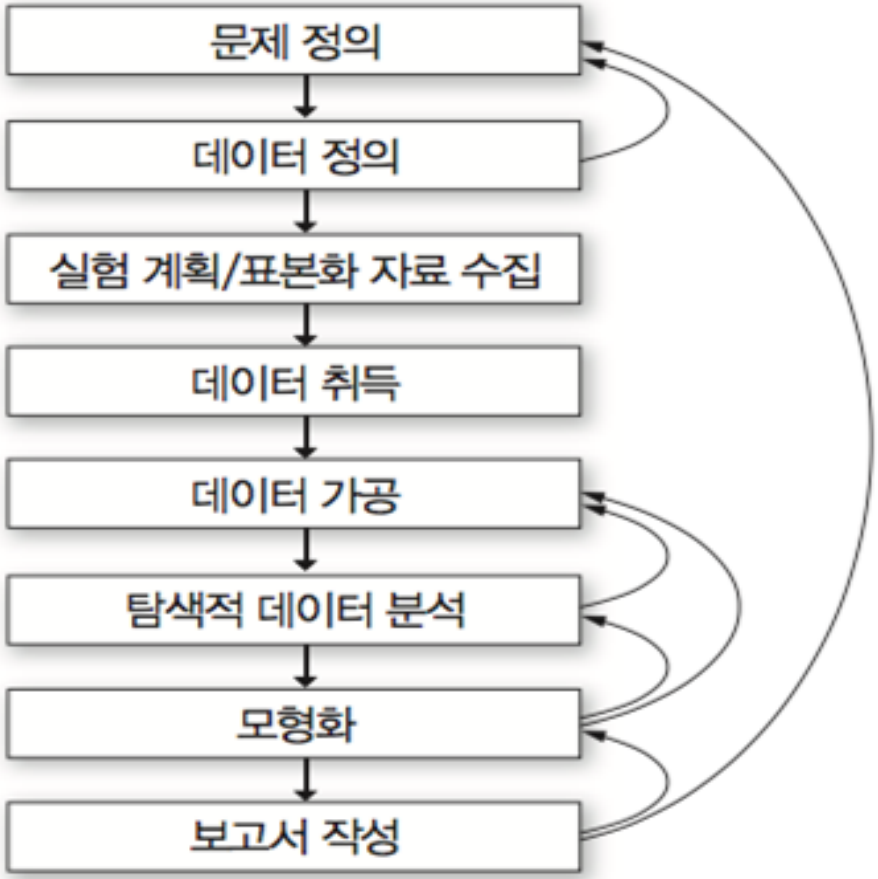
8. Data Science 업무 과정

Data Science업무 수행 과정은 왼쪽과 같지만 실제 현실에서는 오른쪽과 같은 과정을 거쳐 수행함.

이상적인 관점

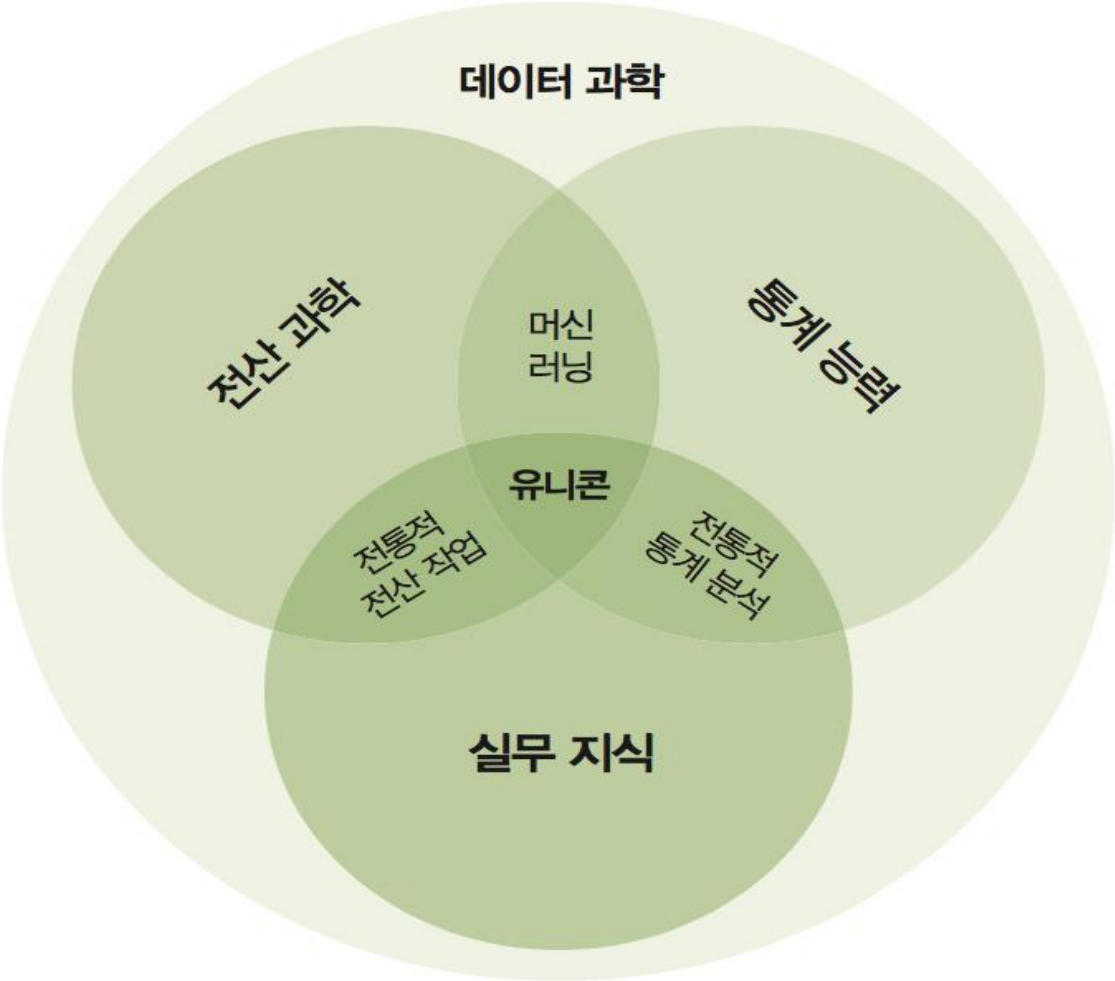


현실적 관점



9. Data Scientist 소양

Data Scienist는 하기와 같이 여러 방면에 걸친 역량을 갖고 있어야 하며, 이를 위해서는 다양한 분야의 지식과 기술역량을 보유하고 있어야 함.



주가선행 지수로 트윗상의 감성지표 분석

S&P500보다 22시간, 구글 주가보다 14시간 선행패턴을 보여줌 → 미래상황 예측 대응

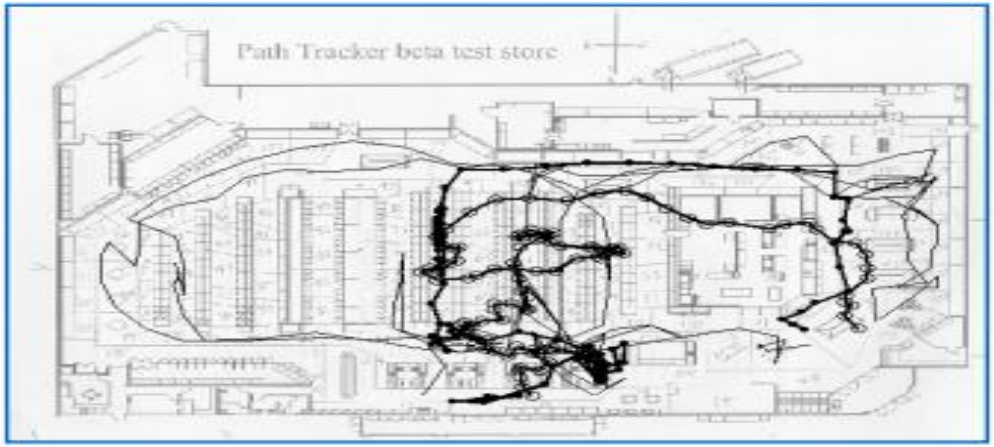
RFID를 활용한 Location Tracking Data를 활용한 고객이동패턴 분석

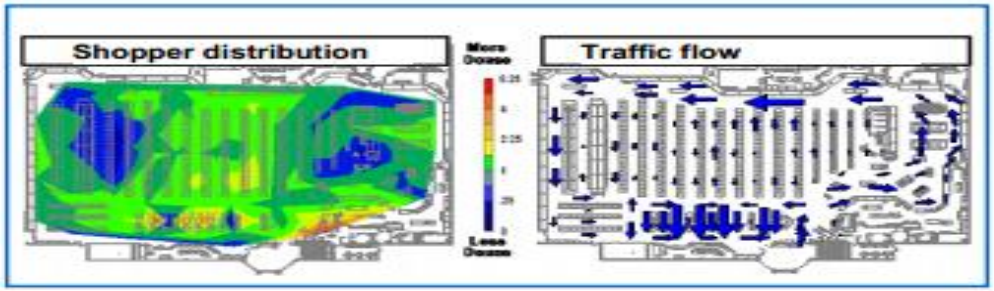
Input

RFID Tag

+ Mobile phones

Output

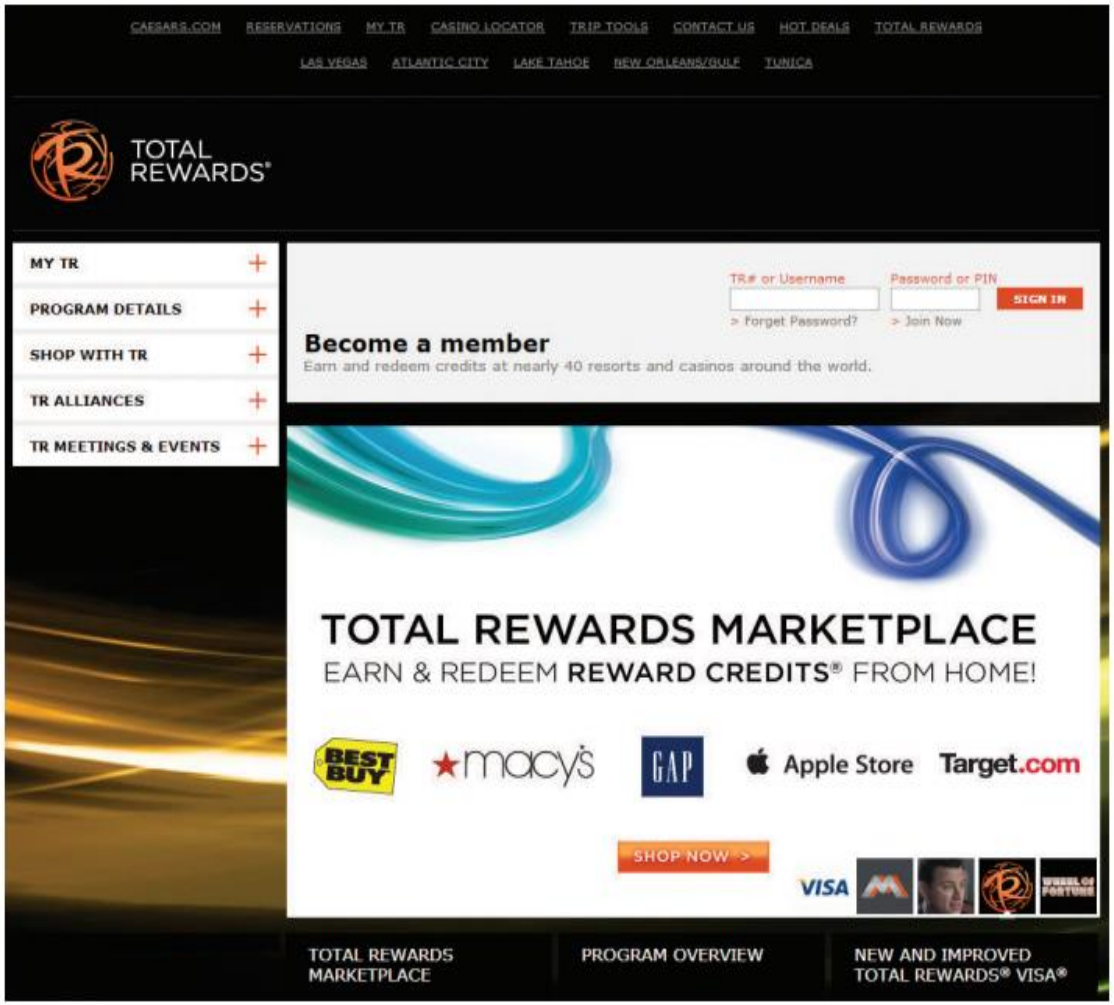




상품 배치 최적화를 통한 매출 증대에 활용

유첨. 적용사례 – 하라스 카지노

개별 고객의 행동 패턴 분석을 통한 마케팅 고도화



Total Rewards Card 사용

↓

실시간 고객의
카지노 사용 Log수집

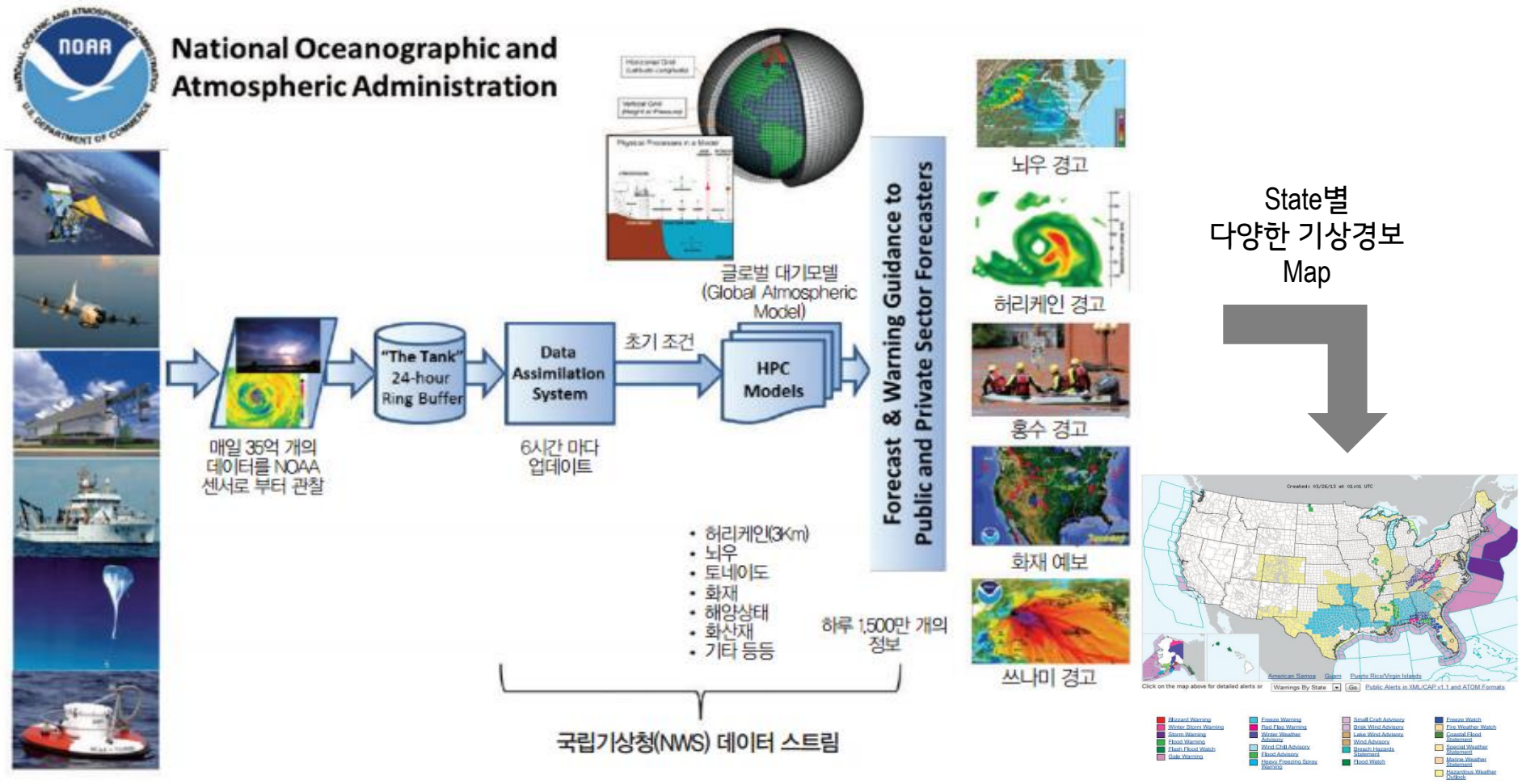
↓

고객의 행동패턴 분석

↓

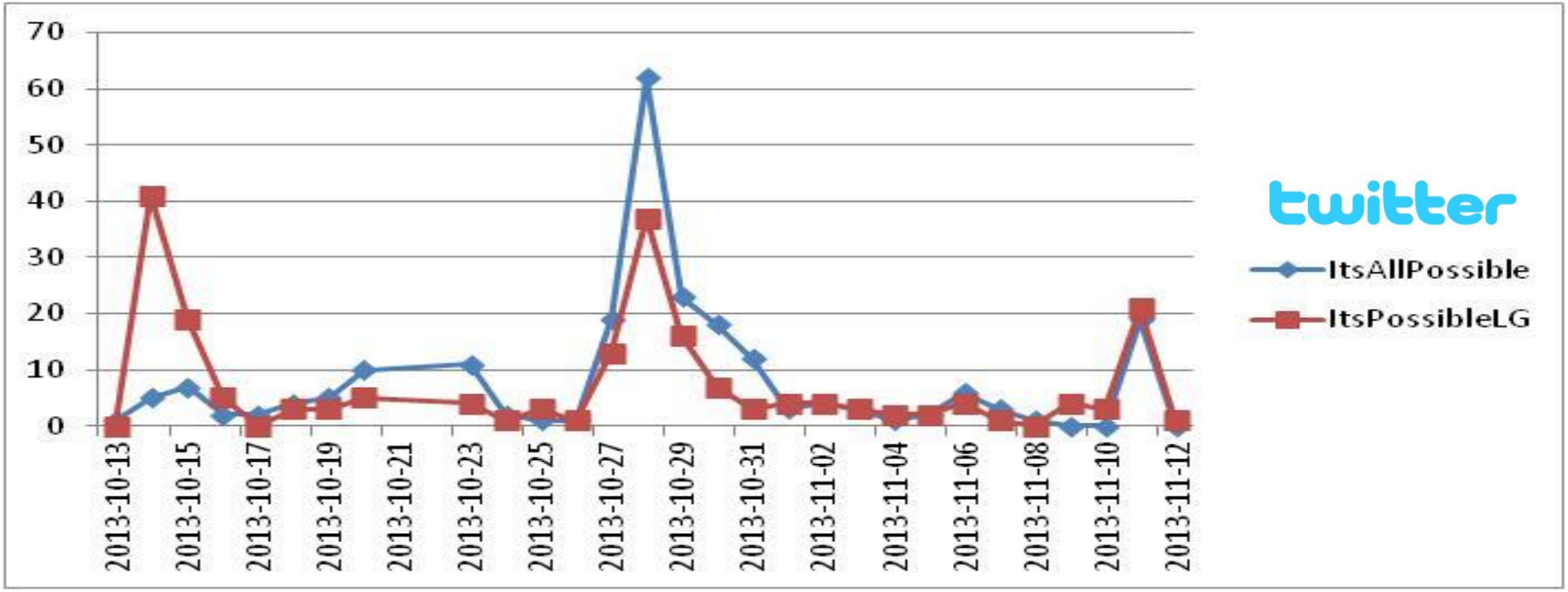
신규서비스 개발

35억 개의 센싱 데이터를 이용한 날씨 정보 예측 및 재해 대응



유첨. 적용사례 – LG 북미 Brand Campaign

SNS상에서 고객반응 분석(Sentiment Analysis)



IAP키워드가 증가한 부분에 대한 상세 원인 분석이 필요하며
향후 마케팅 전략을 수립하는 경우 해당 시점에서 Marketing홍보를 극대화함으로써
최소의 투입자원으로 최대의 효과를 기대할 수 있음.

유첨. 적용사례 - 볼보자동차

커넥티드 드라이브를 통한 원격 차량 서비스

텔레매틱스 시스템 장착

⇩

실시간 교통/차량정보 수집

⇩

빅데이터 분석

⇩

신규 서비스 창출 및 응급대응

50만대 이상 생산 후에 발견할 수 있는 결함을 1000대 출고만으로 파악가능하게 됨

유첨. 적용사례 - 설비예지보전

공장내 노후화 장비의 OEE개선을 위해 전력계측 모듈을 설치하여 실시간 설비상태 모니터링과 이상징후를 사전에 감지하고 대응함.

현상

개요

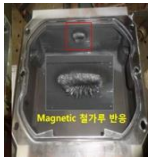
- 대상 : 솔라셀 Back Contact라인
  2개사 제품 대상(PFEIFFER, LOT베큠)
- 이슈 : PECVD 진공펌프 고장 정지

문제점

- 대당 6회 고장 발생 / 년
  : 대당 년 4천만원 수리비용 발생
- 수리기간 2~3일 소요

주요 불량 현상

- NDE in/Out 베어링 손상 (Ball Flaking)
- 구동축 흔들림

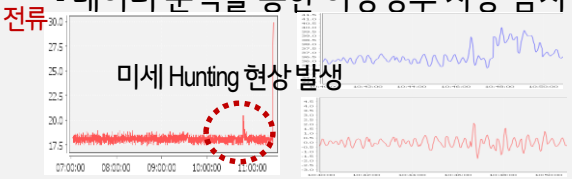


적용기술

클라우드 | 센서 | IoT | AR/VR | 로봇 | CAD/CAE
머신러닝 | 딥러닝 | 비전 | 진동소음 | 마이닝 | 시뮬레이션

- 전력 계측 모듈 설치
  - 설비 주요부품에 전력센서 설치

- 설비 이상징후 감지 알고리즘 개발
  - 설비 상태 실시간 모니터링 시스템 개발
  - 데이터 분석을 통한 이상징후 자동 감지



대상데이터 분석 알고리즘 개발
1) 미세이상변화 감지
- 고주파 분석 : 미세변화 감지
- 저주파 분석 : 패턴변화 감지
2) 이상패턴 감지
- Time Series Classification : 단위 전류패턴별 유사도 분석

적용 결과

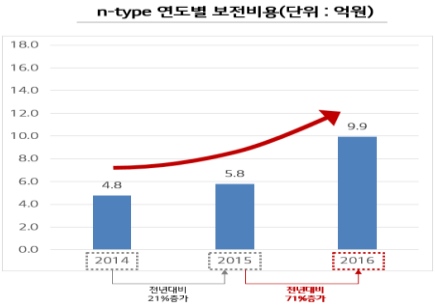
❑ 진공 펌프 이상징후 감지 자동화

- 고장발생 2~3시간전 이상징후 감지
  (데이터 마이닝 알고리즘 적용)

❑ 무인 설비보전 운영 가능

- 순시방문 통한 설비보전업무 제거
  (전력Check로 인한 감전위험 방지)
- 자동 Alarm가능
  (문자 or e-mail을 통한 Alarm 전송)

개선 효과



- 설비노후화로 인한 보전비용 저감가능
  ('16년 기준 20%저감시 2억원 저감예상)

유첨. Field 품질불량 Claim 분석 및 대응 시스템

Field Failure Analytics

FFR[1] Data 취합 → Data 전처리 → FFR 분류 → FFR 빈도계산 → FFR통계값 저장 → 시간기준 FFR추이 변동 분석

- Data 전처리: ▪ 형태소분석 ▪ 표준용어 적용
- FFR 분류: ▪ 텍스트분석 ▪ 자동분류
- FFR 빈도계산: ▪ 국가별 ▪ Defect Type별

시스템 제공기능

❑ **FFT 불량률 예측(장/단기간)**

- 시장특성에 따른 불량률
- 기간특성에 따른 불량률
- 제품특성에 따른 불량률

❑ **실시간 FFR분석**

- 시장별 불량발생 현황 정리
- 급대응 불량항목 감지

❑ **생산품질 연계 FFR원인 규명**

- 기간특성에 따른 불량원인 추적
- 라인특성에 따른 불량원인 추적
- 제품특성에 따른 불량원인 추적

❑ **설계품질 연계 FFR 원인 규명**

- 해당 모델기준 ECO정보 분석
- 급대응 불량항목 원인 분석/대응 (구매, 설계, 생산 정보와 연계)

일반패턴 → FFR패턴 분석 (시계열, 신경망 등) → FFR 예측모델링 개발

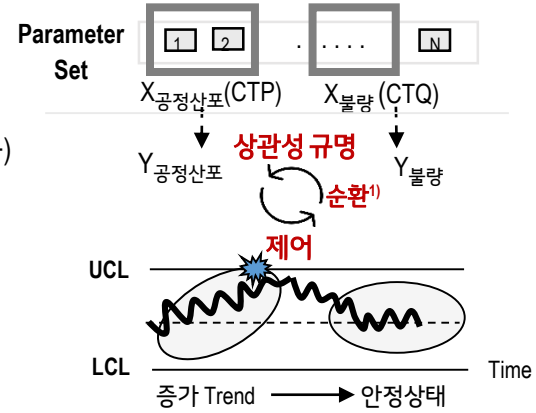
이상패턴 → 모델명 기준 생산정보와 FFR정보간 연계 분석 (Text, Network분석) → 생산품질정보와 연계분석기능 개발 (라인/모델별 불량원인추적)

1) FFR : Field Failure Rate의 약어로 필드불량 정보임

유첨. EWS 개요

제품 품질(Y)에 영향을 미치는 공정 원인(X)과 최적 운영 기준을 규명하고 현장에서 이슈사항 발생시 이를 통제하는 관리 체계를 의미함.

	Parameter 영역	**제품 Data 영역**
	원인 : X	결과 : Y
Data	공정/장비 Parameter Fab. 환경 Parameter 설비 Event 재료 자재	공정 산포 (공정/장비/환경 등) Defect 검사 (품질측정값)
Process	변경/변화 공장1,2…	계측 → 검사 → Field Claim $Y_{공정산포}$ $Y_{불량}$ $Y_{불량}$
Activity	• **CTP, CTQ 정의 및 모니터링** - 주요 공정산포 및 불량 인자(Parameter) 선정 (공정/장비, 이물/환경, 재료 등) - 실시간 변동사항 모니터링 • **Data Tracking 체계 구축** - 공정 변동/변화 이력 추적성 확보 (공정/장비, 환경, 설비, 재료, 공정간 진행 결과 등) • **설비 Event 모니터링 및 가동률 향상** - 이물, 공정/장비 Parameter 상태 모니터링 - 설비 이상징후 모니터링/예지 대응 • **Parameter 표준화** - 통합 품질 Code 생성 및 관리	• **공정 산포 및 품질 Monitoring** - 실시간 계측 및 검사 데이터 모니터링 • **공정 및 품질 이상 Alarm** - 이상원인 규명 - 이상 발생시 Alarm 방법 설정 (시스템 or 모바일 문자메시지 등) • **공정 및 품질 이상 발생시 대응** - 계측 및 검사결과 분석 - 설비 Parameter 보정



1) 4M 1E 변경 / 변화에 따른 지속적인 순환 Loop가 운영되어야 함

유첨. EWS 구성 요소

EWS의 핵심 구성요소는 RMS, FDC, R2R, PD Set으로 하기와 같은 관계를 갖고 있음.

구성요소

❑ **RMS(Receipe Mgt. System)**

- 설비의 Recipe를 편집(생성, 수정, 삭제)하고, 설비간 Recipe를 비교 및 검증하여 사고를 예방함

❑ **FDC(Fault Detection & Classification)**

- 설비에서 Sensing 되어 올라오는 data를 수집하여 허용된 범위를 벗어나는 경우 Alarm을 발생시켜 조기에 문제 해결이 가능하게 함

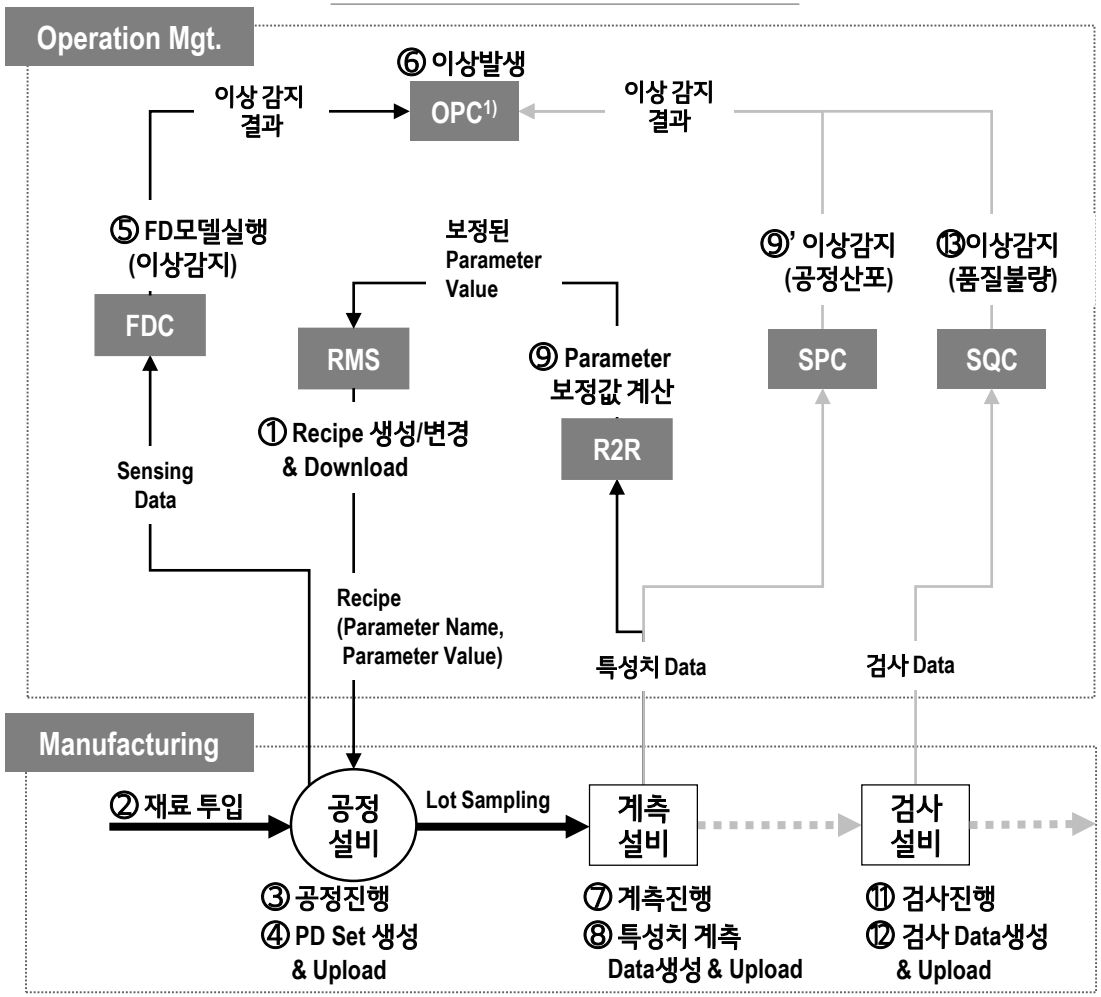
❑ **R2R(Run To Run)**

- 가공된 제품의 특성치와 목표치와의 차이를 감지하여 공정 설비의 Setting값을 자동 보정하고 지속적으로 설비에 자동 반영 및 제어함

❑ **PD Set(Process Data Set)**

- 공정 Data를 분석에 활용하기 위해 만든 DB로 기준(제품) ID로 모든 공정 정보를 조회할 수 있음

관계도



1) OPC : Operation & Process Control

유첨. EWS 구성 요소

EWS의 핵심 구성요소는 FDC, R2R, PD Set으로 하기와 같은 관계를 갖고 있음.

구성요소

❑ FDC(Fault Detection & Classification)

- 설비에서 Sensing 되어 올라오는 data를 수집하여 허용된 범위를 벗어나는 경우 Alarm을 발생시켜 조기에 문제 해결이 가능하게 함

❑ R2R(Run To Run)

- 가공된 제품의 특성치와 목표치와의 차이를 감지하여 공정 설비의 Setting값을 자동 보정하고 지속적으로 설비에 자동 반영 및 제어함

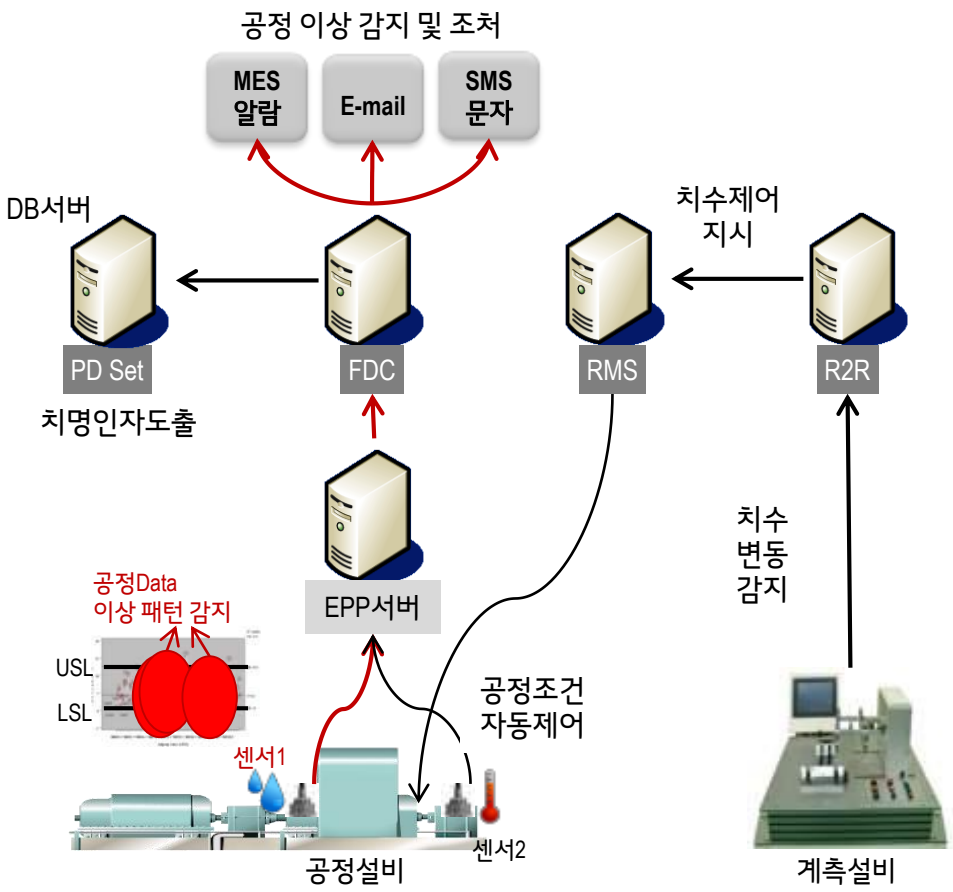
❑ RMS(Receipe Mgt. System)

- 설비의 Recipe를 편집(생성, 수정, 삭제)하고, 설비간 Recipe를 비교 및 검증하여 사고를 예방함

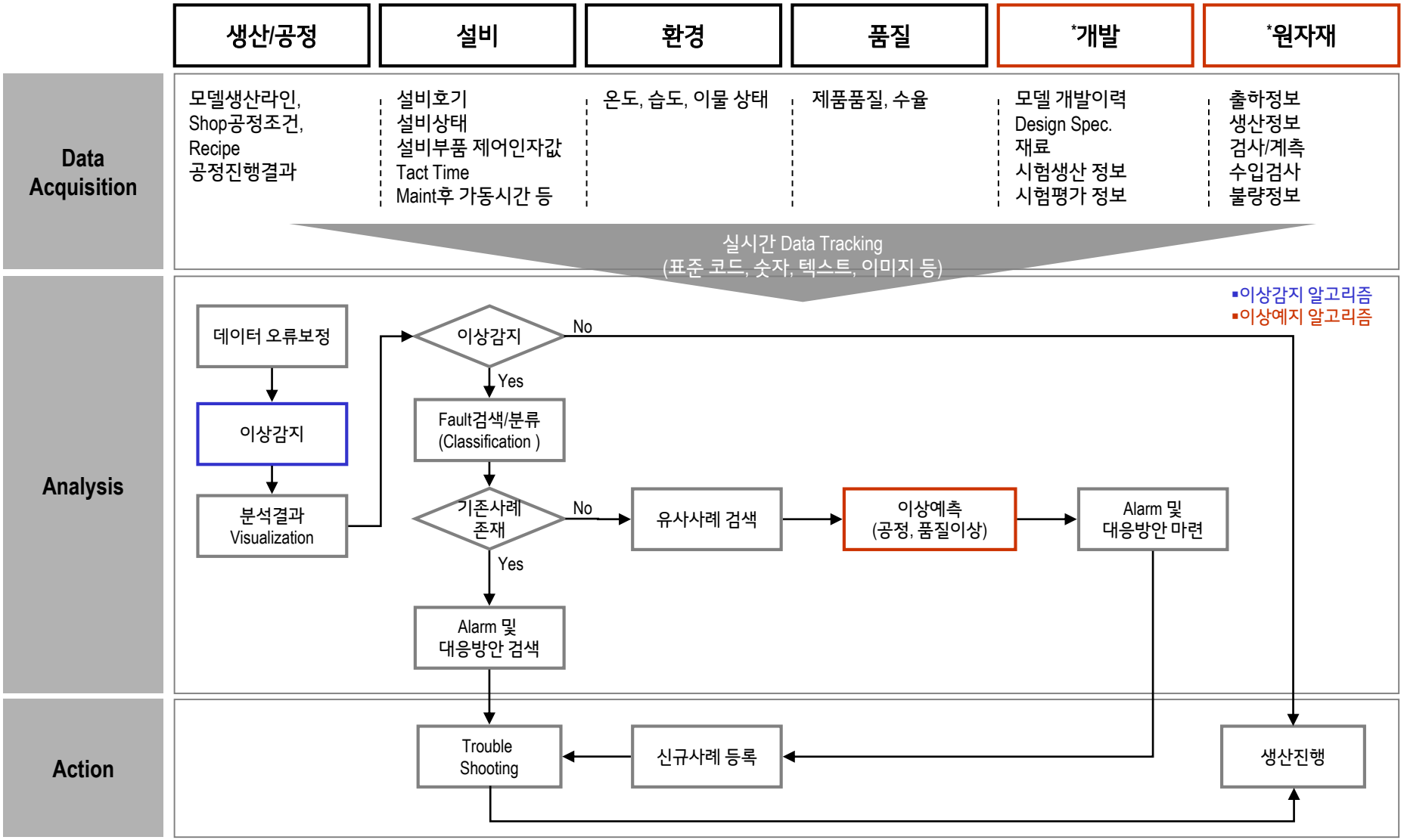
❑ PD Set(Process Data Set)

- 제품ID와 Matching 되는 공정 Data를 조회하여 불량 원인 분석에 활용할 수 있도록 만들어진 데이터

관계도



유첨. EWS 프로세스



* 생산/공정외 요인에 의함, 품질불량 원인규명의 왜곡 가능 요소

유첨. 적용사례 - ZARA

Fast Fashion : 소량생산 적기판매 → 재고 최소화

MIT의 빅데이터 분석 기술을 사용하여 적용함으로써 Fast Fashion 비즈니스를 대표하는 기업으로 성장함

- **경쟁사 : 6개월 전 기획한 의류 3,000여종** 판매
- **ZARA : 인기있는 의류 11,000여종 생산, 적기판매**
  (빅데이터 분석으로 생산비용 저감)

전세계 매장에서
실시간 판매 & 재고 데이터 수집

⇩

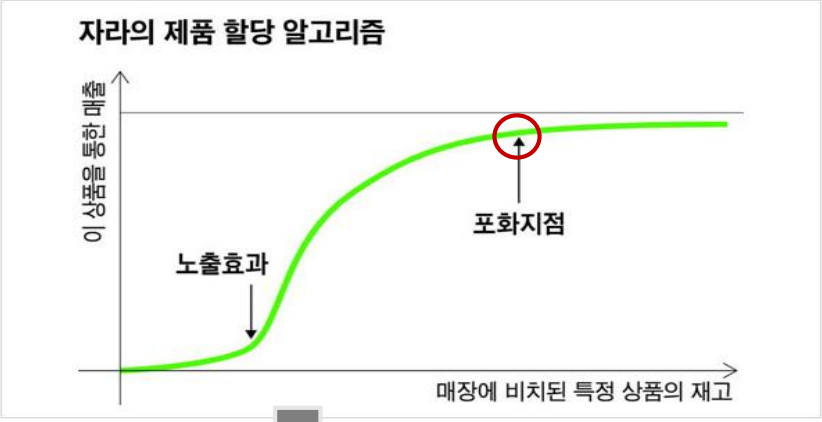
빅데이터 기반 고객 니즈 파악, 인기제품 분석

⇩

최적화 생산전략 적용
(잘팔리는 제품으로 바로 생산판매)

⇩

최적화 제품 할당 알고리즘에 근거한
재고최소화 & **소량생산 & 적기판매**



- **적정 재고량 : 노출효과(10개),** 포화지점(50개)
- **현 생산** 제품량 : 100개
- 대안1 : 전세계 지점 50개 매장에 2개씩 배정
- **대안2 : 번화가에 있는 2개의 매장에 50개 배정**

※ Fast fashion : 현재 유행하는 패션 트렌드를 포착해 패스트 푸드처럼 신속하고 저렴하게 고객에게 제공하는 패션 비즈니스 모델

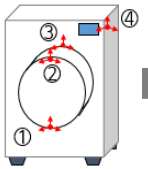
유첨. 제품개발업무 자동화

개발시험 자동화를 통한 시험빅데이터를 확보하고, 딥러닝을 이용하여 제어설계를 무인 최적화하여 업무효율성을 제고함

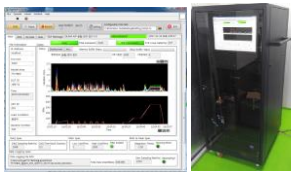
❑ 편심방지 제어 개발시험 자동화를 통한 시험 빅데이터 확보

편심(UB)발생 검토시험 자동진행

3축 4point 진동측정



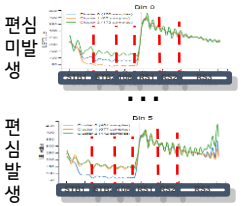
시험 빅데이터 수집 및 실시간 모니터링



❑ 편심발생 예측모델링 개발 → 편심방지 최적 제어사양 설계 자동화

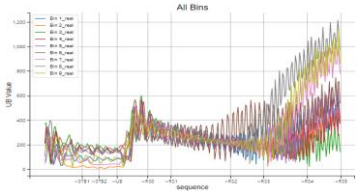
현 제어사양에서 발생가능한 현상예측 시뮬레이션

시험Log값 준비 (사례샘플링)



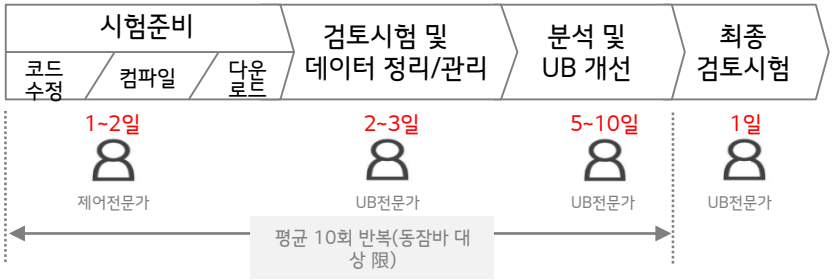
UB발생여부 예측

UB방지를 위한 최적제어사양 도출

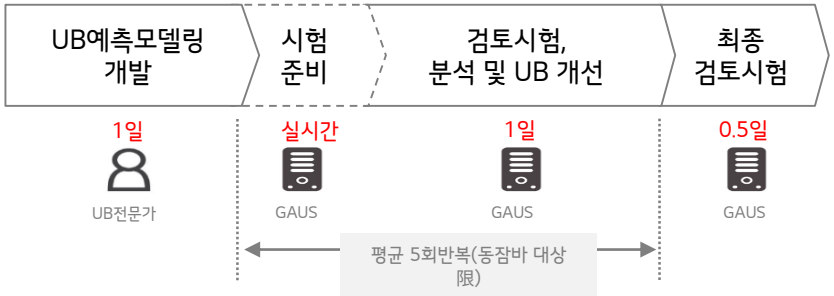


대상 제어구간 : 10개
구간별 모터단락 설정값 범위 : 0~∞
적용가능 제어사양 설정대안 개수 : ∞

❑ 기존 : 전문 Eng'r가 편심방지 제어사양 개발



❑ 개선 : AI를 이용한 편심방지 제어사양 개발 무인화



성과

- 편심방지 제어사양 개발 자동화(무인 시험 & 설계 可)
- 개발기간 단축(기존 3개월 → 1.5개월 이하 可)

적용기술

센서 · IoT · 클라우드 · 빅데이터 · 인공지능 · VR/AR · CPS · 로봇 · 3DP · 정보화

적용현황

개발 → 적용 → 확산 → 협력회사

LG전자 HA본부 세탁기

향후계획

- 편심방지 제어사양 개발 자동화 기능 확장(북미향 → 유럽, 내수향)
- 해석 시뮬레이션 연계를 통한 검토시험 가상화

유첨. 설비 이상 모니터링

생산 설비의 Log정보와 센서계측 정보를 수집/분석하여 이상징후를 미리 감지하여 대응함으로써 운영효율을 극대화할 수 있음.

데이터 수집 및 모니터링

▪**설비 보전 이력**
- 부품교체이력 및 수명
- 고장종류 및 원인
- 고장빈도(MTBF)
- 수리시간(MTTR)
- 진단시간(MTTD)
- 고장전 운영시간(MTTF)

▪**센서 데이터**
- 전력 계측값
(전류, 전압, 전력량, 역률, 고조파, 전류불평형도 등)
- 진동 계측값
- 초음파, 압력, 열분포, 비전 계측값

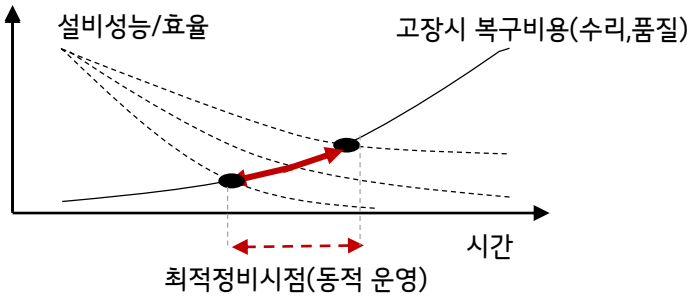
▪**설비 운전 데이터**
- 설비 운전 방법
(구동 부하, 소요시간)
- 설비 운전 패턴
(구동 부하량 이력)

▪**품질 데이터**
- 제품품질 데이터
(제품 불량유형, 수율)
- 제품공정진행 경로
(설비작업이력, 제품품질)

센서 계측 데이터 조회, 처리 및 가시화

이상진단/예지 모델 개발 및 설비보전이력 분석

▪**설비 보전 주기 분석**
- 수선비, 품질Loss 정량화
- 최적정비시점 or 주기 도출



▪**설비 이상진단/예지 분석**
- 실시간 설비상태 및 운영 부하분석
- 고장발생전 예지보전

Data처리	- 센서신호처리/해석 (주파수 파형 Pattern)
이상진단 및 예지	- 시계열 Data(AR) - 선형 Data(LDA) - 비선형 Data(DT, ANN)
상황별 설비보전	- 이상유형별 보전방안 수립 (부품교체, S/W개선 등) - 설비와 품질이상을 연계한 보전방안 수립

데이터분석 기반 설비보전 방안정립

▪**1단계 : 기간기준 보전(TBM[1])**
- 최적보전기간분석
- 정적보전시점 Rule정립
(부품평균수명 기준)
- 주기별 보전실시

▪**2단계 : 상태기준 보전(CBM[2])**
- 설비 이상진단방법 개발
- 이상판단 Rule정립
- 고장발생예상시 보전실시

▪**3단계 : 운영기준 보전(OBM[3])**
- 설비 운영부하패턴 분석
- 동적 보전시점 Rule정립
(부하패턴 기준)
- 제품품질/수율을 고려한 보전방안 수립
- 고장발생예상시 보전실시

1) TBM : Time Based Maintenance, 2) CBM : Condition Based Maintenance, 3) OBM : Operation Based Maintenance

유첨. 빅데이터 플랫폼

빅데이터 플랫폼은 빅데이터에서 가치를 추출하기 위한 일련의 과정을 지원하기 위한 프로세스를 규격화한 기술과 서비스의 집합으로 하기와 같음.

	내용	방법	솔루션
수집	디지털 서비스의 발전에 의해 폭증하는 정형/비정형 데이터가 SNS/센서/공장/제품 등에서 발생하며 이를 수집함	비정형/정형 데이터 수집 ETL(Extract -Transform-Loading) Web Crawling Log 수집기, IoT 센싱 등	비정형/정형 데이터 수집 ETL(Extract -Transform-Loading) Web Crawling Engine Log data 수집기
저장	빅데이터의 3V특성을 반영하여 데이터를 저장하는 기술로 단순저장이 아니라 분석을 감안하여 Query업무 사전 반영	원본 데이터 저장 NoSQL 저장 R-DB 저장 메모리 저장	Hadoop System, Google File System Hbase, Cassandra, MongoDB
처리	저장된 빅데이터에 대해 수요발생시 적시에 가공하고 분석을 지원함	배치처리 실시간 처리(CEP)	Hadoop Mapreduce Hadoop Pig Apache Hive, Spark
분석	저장되어 있는 빅데이터에서 잠재된 가치와 통찰력을 추출하기 위한 작업	통계분석, 기계학습 데이터/텍스트 마이닝 소셜분석 인공지능(딥러닝, 강화학습)	R Python(Scikit-Learn, StatsModels, Scipy) Mahout Deep learning(TensorFlow, Keras, Pytorch) KNIME, SAS
표현	분석정보의 의미를 가장 빠르고 정확하게 전달하기 위해 효율적인 그림이나 이미지로 제공하는 방법임	시각화 (통계그래프, Weblog/SNS/지도 시각화)	Splunk, Tableau, Qlik Plotly, R(ggplot) Python(matplotlib)

유첨. 하둡 생태계

빅데이터 플랫폼은 빅데이터에서 가치를 추출하기 위한 일련의 과정을 지원하기 위한 프로세스를 규격화한 기술과 서비스의 집합으로 하기와 같음.

하둡관련 오픈소스 프로젝트

분산 코디네이션 **(Zookeeper)**	**관리 및** 모니터링 **(Ambari)**	분산 데이터베이스 (Cassandra, Hbase)	기계학습**(Mahout, R, Python)**	데이터직렬화(Avro)
			데이터플로우(Pig)	SQL(Hive)
			메타데이터 관리(HCatalog)	
			분산데이터 처리**(MapReduce)**	
		분산 파일 시스템(HDFS)		
		정형데이터 수집(Sqoop)	비정형 데이터 수집**(Chukwa)**	

- Zookeeper : 분산환경에서 노드 간의 정보 공유, 락, 이벤트 등 보조기능을 제공하는 프레임워크
- Ambari : 하둡 클러스터를 프로비저닝, 관리, 모니터링하는 웹기반 도구
- Avro : 이 기종간에 RPC(Report Procedure Call)와 데이터 직렬화(Serialization)를 지원하는 프레임 워크
  (JSON을 이용, 작고 빠른 바이너리 포맷으로 데이터 직렬화)
- Cassandra : 하나의 에러 때문에 전체 시스템이 정지되지 않도록 설계된 확장 가능한 멀티-마스터를 갖는 데이터 베이스
- Chukwa : 분산환경에서의 로그수집 및 저장, 관리, 분석을 위한 플랫폼
- Hbase : 대용량 데이터베이스 테이블을 위한 구조적 데이터 저장 공간을 지원하는 확장 가능한 분산 데이터 베이스
- Hive : HDFS에 저장된 데이터를 관리할 수 있도록 쿼리를 제공하는 데이터 웨어하우스 솔루션
- Mahout : 확장가능한 머신러닝과 데이터 마이닝을 위한 라이브러리
- Pig : 병렬계산을 위한 고수준 데이터 플로우 언어와 실행 프레임워크
- Sqoop : SQL + Hadoop의 합성어
- YARM : 리소스 매니저(리소스 관리 플랫폼으로 사용자의 애플리케이션 스케줄링 사용 관리)