2014년도 빅데이터 활용 스마트서비스 사례집







〈목차〉

1.	2014년도 빅데이터 활용 스마트서비스 시범사업 개요	1
2.	유통 빅데이터를 통한 중소상인 지원	3
3.	빅데이터 분석 기반 외국인 관광산업 지원	12
4.	맞춤형 유의질병 및 병원정보 제공	22
5.	자동차 부품기업 공동활용 빅데이터 플랫폼	26
6.	조류 인플루엔자(AI) 확산 조기대응	35
7.	국도 비탈면 붕괴사고 예측	40

2014년도 빅데이터 활용 스마트서비스 시범사업 개요

1. 추진 목적

- o 비즈니스 중심의 빅데이터 선도 서비스 모델 발굴 및 개발 지원을 통해 빅데이터 서비스 수요 창출 및 관련 산업 활성화
- o 빅데이터 서비스 확산을 통해 새로운 고부가가치를 창출하고, 창조 경제 기반의 국가 신 성장동력 발굴

[빅데이터 시범사업 추진방향]

O 빅데이터 활용 촉진과 파급효과가 큰 산업 분야를 중심으로 서비스 발굴

O 빅데이터 서비스 수요 창출 및 관련 산업 활성화



2. 2014년도 주요 추진 내용

o 사업명 : 빅데이터 활용 스마트서비스 시범사업

o 사업예산 : 총 12억원(과제당 2~4억원 지원)

o 선정과제 수 : 6개

o 참여방식 : 컨소시엄 구성에 의한 자유공모

o 지원방식 : 매칭펀드(현물 포함)

o 과제 수행기간 : 2014.5.21 ~ 12.31

3. 2014년 추진과제 과제

컨소시엄	서비스명	내 용	활용 데이터		
대한상공회의소 (한일네트웍스, 클루닉스, 리테일테크, 디노플러스)	빅데이터 기반 유통시장 상품판매 분석 서비스	대형유통사의 판매정보를 분석하여 지역 슈퍼마켓 등 중소상인을 위한 시즌별 인기상품 분석정보, 날씨/지역 맞춤형 상품추천, 연관 구매상품 추천 등 데이터 기반 마케팅 정보 제공	o 대한상공회의소 : 유통사 매출데이터 o 공공데이터 : 기상정보, 지역/인구정보, 지역관리정보 등		
(주) 오픈메이트 (BC카드㈜, NICE평가정보㈜, 한국관광공사, KT㈜)	빅데이터 분석 기반 외국인 관광산업 지원	외국인 관광·소비 패턴과 관광객 트렌드를 복합 분석하여 추가 관광지 개발, 관광지 추천 등 지도 기반의 맞춤형 관광정보 제공을 통해 관광객 만족도를 향상시키고, 분석 정보를 관광정책에 활용	 ○ 오픈메이트 : 유동인구정보 ○ 비씨카드 : 내외국인 소비패턴 ○ 나이스평기정보 : 상기업소 DB ○ 한국관광공사 : 관광정보 ○ KT : 내외국인 유동인구 패턴 		
메디벤처스㈜ (㈜라인웍스, 건강보험심사평가원)	의료정보 빅데이터 기반 맞춤형 유의질병 및 병원정보 제공	진료정보 빅데이터 분석을 통해 발생 질환별 예상 유의 질병정보 및 맞춤형 병원 정보 제공	o 심평원 : 건강보험 청구 데이터, 전국 병원정보		
메타빌드㈜ (㈜솔바테크놀로지, 자동차부품연구원)	자동차 부품기업 공동활용 빅데이터 플랫폼 구축	자동차 부품 제조사가 공동활용할 수 있는 빅데이터 플랫폼을 제공하여 생산 공정 데이터의 정밀 분석을 통한 제품 품질향상에 활용	o 자동차부품생산기업 : 품질자료(생산, 품질, 공정 등 15개종 130여 데이터)		
㈜ KT (농림축산검역본부)	조류 인플루엔자(AI) 확산 대응	국 가 동 물 방 역 통 합 시 스 템 (KAHIS) 데이터와 KT의 통화로 그 데이터를 연계 분석하여 조 류인플루엔자(AI) 확산 경로 예 측	o 축사 차량 방문 기록, 전 국 농장정보, AI발병 농 장정보, 축산차량등록 정 보: 농림축산검역본부 o 통화로그 데이터: KT		
대한지적공사 (한국건설기술연 구원, 모바일팩토리, SKT)	국도 비탈면 붕괴사고 예측 서비스	국도 비탈면 데이터(절토사면, 지형정보 등)와 기상정보의 복 합 분석을 통해 급경사지 위험 도 산정 모델 개발 및 예측 서 비스 제공	o 지형, 지질, 붕괴이력, 위 치정보: 국토교통부 o 강우량, 온도, 습도 등 기상데이터: 기상청		

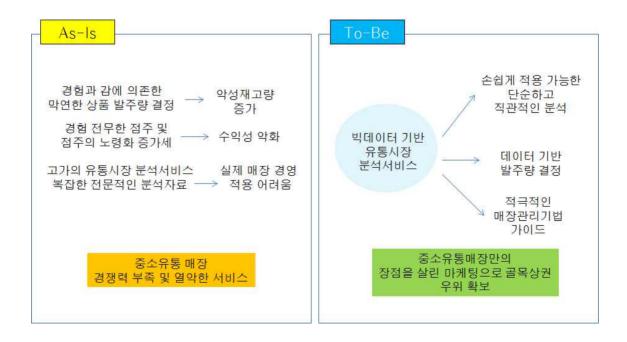
유통 빅데이터를 통한 중소상인 지원

대형유통사의 판매정보를 분석하여 지역 유통매장 중소상인을 위한 맞춤형 상품추천과 시즌·기온별 데이터 기반 마케팅 정보 제공

추진 목적 및 배경

■ 사업 추진의 배경

- 빅데이터를 적극 활용하는 상위 3개의 유통기업의 생산성과 수익률이 경쟁 기업 대비 6%이상 높음 (대한상공회의소 유통마스터플랜 분야별 정책과제, 2012)
- 대형유통 업계에서는 빅데이터 분석을 활용하여 마케팅 및 매장운영 정책에 반 영하여 상당한 효과를 누리게 되었으나, 빅데이터의 혜택을 받지 못하는 중소 유통기업은 경쟁에서 소외되는 현상 발생
- 전체 기업의 60%이상을 차지하며 외부 시장 정보를 전혀 활용하고 있지 않은 중소 업체의 경쟁력 강화를 위한 위한 데이터 분석 서비스가 필요



■ 사업 추진의 필요성

- 본 사업 주관기관인 대한상공회의소는 유통사로부터 매월 매장별 매출 데이터를 수집할 수 있는 PDS(Pos Data Service)시스템을 운영하고 있으며 2011년 3월부 터 유통시장 분석정보 시스템을 구축 운영 중
- 중소 유통 상공인에게 실질적인 도움을 주고자 기존 보유 데이터와 외부 개방 데이터의 메쉬업(Mash-Up) 분석을 통한 본격적인 빅데이터 서비스 추진

추진 내용

□ 참여기관: 대한상공회의소, 한일네트웍스, 리테일테크, 클루닉스, 디노플러스

□ 주요 활용데이터

구분	데이터	데이터 규모	보유기관		
매출	POS(Point of Sale)	2EAMD/ス	대한상공회의소		
데이터	* 전국 약 700여개 점포,2015 2월 기준	350MB/주 점포,2015 2월 기준			
	점포정보	210MB/주			
매출메타	점포속성	350MB/주	대한상공회의소		
데이터	상품분류	상품분류 70MB/주			
	상품속성	560MB/주			
	기상데이터	3MB/건	기상청		
બે ચો	주민등록인구	10MB/건	안전행정부		
연계	공시지가	5MB/건	한국감정원		
데이터	용지별 면적	5MB/건	국토해양부		
	Social Network Service	1GB/년	대형포털/커뮤니티		

□ 분석 내용 및 기법

- 빅데이터 분석 서비스 모델 선정
- PDS 자문위원 및 중소 유통사 대상 인터뷰, 벤치마킹, 선진사례 분석을 통해 아래 와 같이 서비스와 관련된 주요 진행방향을 도출
- 다양한 외부 데이터와 유통 데이터를 메쉬업하여 실질적인 도움이 가능한 분석

결과를 산출

- SNS 데이터를 활용한 빅데이터와 자연어 처리 기술을 적극 활용하여 연관 상품 과 관련된 분석 결과를 도출
- 분석 결과는 IT 접근성과 가독성이 떨어지는 중소 상공인을 고려하여 접근이 쉽고 단순한 유저 인터페이스(UI)를 통해 제공
- 도출된 시사점과 현재 보유한 데이터, 수집 가능한 외부 데이터를 고려하여 다음과 같이 5개의 빅데이터 주요 분석 서비스 시나리오를 선정함
- 주요 시즌 / 이벤트별 상품군 추천지수 산출
- 기온대별 상품군 추천지수 산출
- 지역특성별 상품군 추천지수 산출
- 신상품 라이프 사이클(Life-Cycle) 매출 추이 분석
- SNS 데이터 기반 연관 상품 분석
- 분석 서비스 시나리오별 상세 내용
- 주요 시즌 / 이벤트별 상품군 추천지수 산출
- 주요 이슈별 시즌 및 이벤트를 정의하고, 매출 데이터와 시즌 정의 데이터를 연 관분석하여 시즌 이벤트별 상품군 추천 지수를 산출
- 기온대별 상품군 추천지수 산출
- 회귀분석 기법을 통해 기온대와 매출량과의 상관관계를 분석하여 5도 단위 기온 대를 정의 한 후, 정의된 기온대와 매출 데이터를 연관분석하여 기온대별 상품 군 추천 지수를 산출
- 지역 특성별 상품군 추천지수 산출
- 지역 부동산 데이터, 인구 데이터를 종합하여 법정 동 단위의 지역 특성을 선정한 후, 상품군별 전체 평균 매출과 지역 특성에 따른 매출을 비교하여 지역 특성에 따른 상품군 단위의 추천 지수를 산출
- 신상품 라이프 사이클 분석
- 신상품의 출시 후 매출 추이를 동일 상품군 내 다른 상품의 매출 추이 패턴과 비교하여 추후 매출 추이를 예측
- 모든 상품에 대해 출시 후 2년간의 매출 추이 데이터를 산출하고, 이를 상품군 단위로 분류한 후 매출 패턴을 일반화한 4개의 클러스터 매출 추이를 산출
- 최종적으로 신상품에 대한 매출 추이를 해당 상품군의 4개 클러스터 매출 추이 와 비교·분석한 데이터를 리포트 형식으로 제공

- SNS 데이터 기반 연관 상품 분석
- SNS 데이터를 수집하여 검색 대상 키워드에 대한 연관 키워드를 필터링하고 해당 키워드 중 음식료품과 관련된 데이터를 추출하여 검색 키워드에 대한 연관 상품을 추출 (ex. 맥주의 연관 상품: 불닭, 소세지, 새우짱)
- SNS 크롤러(Crawler)를 통해 수집된 데이터는 자연어 분석 처리 기법에 따라 연관 키워드를 추출하며, 별도로 작성된 유사어 사전 참조에 의해 최종적인 연관 상품 단어를 추출

■ 데이터 처리 및 분석 기법

방대한 양의 매출 빅데이터와 정형/비정형 데이터가 혼재된 데이터를 고속으로 수집 및 분석하기 위하여, 데이터 전처리 기술과 다양한 빅데이터 분석 기법들 을 동원하여 데이터를 처리 및 분석함

■ ETL 처리

- 스크립트 또는 프로그래밍 언어를 통해 개발한 분석도구를 사용하거나, ETL 전용 분석도구를 사용할 수 있으며, 본 사업에서는 클루닉스사의 MDP 솔루션을 사용

■ 클러스터 분석

- 신상품의 매출 패턴 분석 시 동종 상품군의 다른 상품 출시 이후 매출 패턴과의 비교를 위해 비교 대상 상품 전체를 클러스터 분석으로 몇 가지 클러스터(군집) 으로 분류

■ 자연어 처리

- SNS 크롤러를 통해 수집된 텍스트에서 추출 키워드인 음식료품 단어들을 선별하 기 위해 사용됨

< 빅데이터 처리·분석기술 개요 > -

- ETL(Extraction, Transformation, Load) 처리는 데이터의 수집, 정제, 단순 변환, 최종 storage로의 데이터 적재 과정을 지칭하고, 데이터를 효율적으로 이전하거나 고속 분석을 위해 불필요한 데이 터의 필터링 또는 필요한 데이터의 통합 작업 등을 수행하는 것을 의미하며, 정형 타입의 빅데이터 분석을 위한 전처리 과정으로 수 행하는 경우가 많음
- 글러스터(Cluster, 군집) 분석이란 데이터들의 특성을 고려해 데이터 집단을 정의하고 집단을 대표 할 수 있는 대표점을 찾는 데이터 마이닝 기법중 하나임. 클러스터란 비슷한 특성을 지닌 데이터들의 집단으로 클러스터 분석을 통해 같은 클러스터 내에 특성을 정의 할 수 있고 클러스터 간의 차이를 명확히 볼 수 있음. 이러한 클러스터 분석을 통해 새로운 데이터의 특성을 예측하고 인간이 예측하기 힘든 데이터 그룹간의 차이도 쉽게 파악 할 수 있음.

■ 데이터 분석 과정

각 분석 시나리오 수행을 위한 데이터의 수집, 저장, 분석 등의 단계에서 필요한 데이터 저장소 및 분석 처리 소프트웨어들을 선정하고, 이를 효율적으로 조작 및 관리할 수 있는 Hadoop 기반의 통합 데이터 저장 분석 프레임웍(G-PAS)을 구축함

① 수집

- OpenAPI, FTP, Sqoop을 활용해 내·외부 데이터를 수집시스템으로 수집하는 단계
- 데이터별 수집 주기에 따라 수집을 실시하며, 수집 프로세스는 빅데이터 시스템 의 워크플로우 도구로 설정 및 관리

② 저장

- Hadoop 기반의 분산 파일시스템을 활용하여 데이터를 저장
- Hadoop은 별도의 백업이 필요 없으며, 일부 구성 노드의 장애시에도 서비스 지속이 가능하며, Hadoop 기반의 분산 병렬 처리 S/W 활용이 용이함

③ 정제/변환/매핑

- 분석 과정에 불필요한 데이터를 소거 하고, 고속 처리를 위해 필요한 데이터를 변환 또는 병합하는 과정을 수행, ETL 솔루션을 활용함

④ 절차적 분석

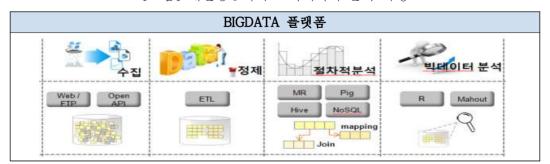
- 빅데이터에 대한 요약, 평균 산출 등의 기본 통계 수치 산출을 수행하며, ETL 솔루션과 Pig tool을 활용

⑤ 클러스터 분석 / 자연어 처리

- 신상품 라이프 사이클 분석과 연관 상품 분석에 필요한 분석을 수행하며, 분석 S/W R을 사용함

⑥ 결과 업로딩

- 웹서비스나 openAPI를 통해 외부에 제공할 최종적인 분석 결과를 MySQL로 저장하는 과정을 의미하며, Sqoop tool을 사용함



[그림] 대한상공회의소 빅데이터 분석 과정

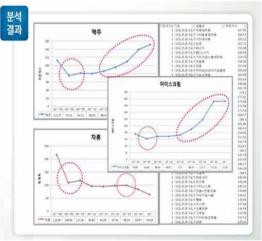
주요 분석 결과 및 활용방안

□ 주요 분석 결과

- 시즌 분석(계절, 명절, 발렌타인데이 등 주요 시즌이나 이벤트 분석)
- 35개 주요 시즌에 대해 114개 상품군 분석결과 쌈장, 김치류 상품군은 여름에 판매량이 가장 높음. 설탕의 경우 초여름(6월말 경)에 판매량이 가장 높음
 - ※ 여름철 시원한 음식선호, 여행 증가 등과 설탕절임(잼류 등) 음식 생산 관련된 현상으로 추정
- 기온별 분석(전체 기온을 5도 단위로 분할하여 매출 분석)
- 35개 아이스크림은 혹한기에 오히려 매출이 증가세로 변함
 - ※ 혹하기 실내외 온도차에 따른 실내건조 현상으로 인해 소비가 증가하는 것으로 판단
- 차 상품군은 -5도 이하에서 최고 매출을 보이며, 25도 이상에서는 기온상승에 따라 급격한 하락세를 보임

[그림] 주요 분석결과



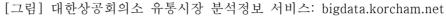


■ 지역별 분석

- 농업지역은 탈취제가 최고 추천 상품, 부유층 지역은 젓갈류, 성인용기저귀, 와 인이 잘 팔리는 상품군임
- 연관상품 분석(주요 키워드에 대해 연관 판매되는 상품 제시)
- 등산에는 '막걸리, 김밥, 도시락' 등, 이사에는 '맥주, 치킨, 커피' 등이 연 관 상품으로 분석됨

□ 서비스 계획 및 활용방안

- 유통시장 분석정보 제공
- 대한상공회의소가 직접 운영하며 3월부터 정식서비스 제공
- 상품군별 인기상품, 가격동향, 상품구색리스트와 같은 단순 집계 분석과 이벤트/ 시즌별 추천상품, 지역특성별 추천상품, 기온대별 추천상품, 상품 라이프 사이 클, 연관어분석 등 빅데이터 분석 결과를 조회 또는 다운로드 할 수 있는 기능 제공





- POS 사업자와 협력을 통해 중소매장 POS 단말에 상용 서비스(2015년 1/4분기)
- 데이터 파일 전송 및 API 연계를 통한 중소매장용 POS 시스템으로의 정보제공



[그림] 대한상공회의소 Open API 서비스

효과 및 향후 적용 확대 방안

■ 사업 의미

- 본 사업은 중소 유통매장 점주들이 원하는 분석 수준과 산출물을 제공
- 최우선 서비스 대상자인 중소 유통매장의 이익 실현, 전체 유통산업 활성화를 통한 경제 활성화에 기여할 수 있는 공공서비스 창출
- 향후 데이터 제공업체/기관의 확대를 통해 전체 유통산업의 세부적인 흐름까지 분석 활용하여 유통·소매업 경쟁력을 강화

■ 활용 및 발전 방안

- 고도화를 통한 서비스 및 리포팅 확장 데이터 추가 확보 및 분석 시나리오 추가 개발을 통해 중소 유통 상공인을 위한 맞춤형 서비스를 지속적으로 확대할 예정
- 데이터 제공을 통한 연계 사업 지원 유통분야 빅데이터 분석 결과 및 빅데이터 분석용 요소데이터들을 필요로 하는 외부 기업이나 기관에 제공하여 2차 사업 활성화 및 빅데이터 서비스 사업에 기여

빅데이터 분석 기반 외국인 관광산업 지원

내외국인 관광·소비 패턴, 중국인 관광 트렌드를 분석하여 개인 맞춤형 관광정보 제공. 추가 관광지 개발. 관광지 추천 정보 제공

추진 목적 및 배경

■ 사업 추진의 배경

- 중국인 관광객 10년전 대비 500% 증가, 외래 관광객 중 가장 높은 비중과 소비 규모, 관광 제도 개선으로 인한 자유여행객의 증가 예상
- 외국인 관광객 1000만 시대, 중국인 관광객은 전체 관광객의 33%로 1위
- 1인당 경비 지출이 중국인관광객이 가장 높음 (약256만원, 타 관광객의 40% 이 상)
- 중국여유법(旅遊法) 개정으로 단체 여행객의 감소, 개별 여행객 수가 증가하고 있음





■ 사업 추진의 필요

- 중국인 관광객의 획일화된 관광지 및 관광패턴, 관광일정으로 인해 재방문률의 정체와 재방문 의향이 낮은 상황

- 국내 입국 관광객의 모바일 인프라를 활용하여 보다 다양한 국내 관광정보의 제 공을 통한 관광 만족도를 높일 필요성
- 중국인 관광객의 여행 패턴 및 소비 패턴 분석을 통해 중국인 관광객 대상 정책 활용 및 사업 환경 개선에 활용

추진 내용

□ 참여기관: ㈜오픈메이트, 비씨카드(주), 한국관광공사, 나이스평가정보, KT

□ 주요 활용 데이터

구분	데이터	데이터 양	제공 기관	
	내국인 거래 패턴	연 약 24억건		
소비	외국인 거래 패턴	연 약 1,300만건		
/거래 패턴	고객 유형 정보 (성, 연령, 주소의 비식별 정보)	약 2,900만건	비씨카드	
관광	전국 블록 및 유형 정보	366,999 건		
권역 및 공간 정보	주요 상권 영역	1,200 건	오픈메이트	
중국어 관광 컨텐츠	중국번체/간체 관광정보 (공통, 이미지, 소개정보, 위치기반 관광정보, 지역기반 관 광정보, 숙박, 행사정보 등)	N/A	한국관광공사	
유동인구 패턴	내국인 통화 데이터	연 약 1.5억건	KT	
비 이 신 티 케인	중국인 로밍 데이터	연 약 180만건	17.1	
상가/업소 정보	상가/업소 DB	약 300만건	나이스평가정보	

□ 분석 내용 및 기법

■ 데이터 처리 및 분석 기법

중국인 관광객의 소비패턴, 이동패턴을 파악하기위해 유동인구 지수를 개발하고 중국 관광활성화 지역을 추출하기 위한 분석기법 적용

• 유동인구 지수 개발

- 교통개발연구원에서 수행한 「전국교통DB구축사업 교통유발원단위조사 및 기초 분석」를 바탕으로 해당 지점 주변의 인구유발시설(백화점, 영화관, 아파트, 지 하철역 등)을 찾고 각각의 인구유발시설들이 통행을 유발시키는 영향력을 중첩, 거리에 따른 가중치를 적용하여 도보 가능한 도로에 대해 10m 간격으로 값을 산 출, 지수화

■ 소비패턴 분석

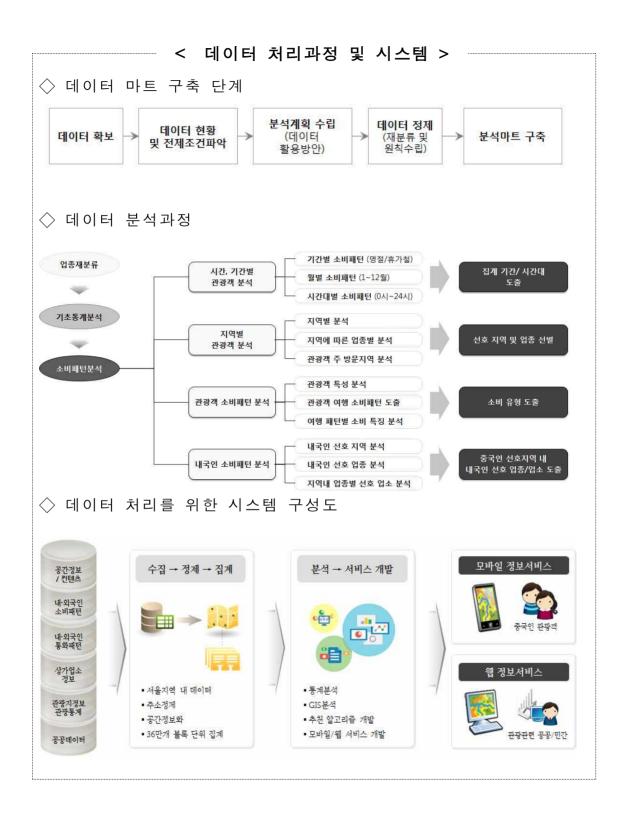
- 중국인 여행유형별, 여행시즌별, 시간/요일대별 소비패턴 분석, 선호지역 비교 분석, 내국인과의 여행 패턴을 비교하기 위한 내국인 소비패턴 분석도 병행

■ 이동패턴 분석

- 카드 거래 데이터 기반으로 소비형태에 의한 이동패턴 분석(여행이동거리, 여행 시 먹거리 이동거리 등)

■ 관광활성화지역 추출

- 서울시 주요여행지역 및 소비밀집도, 통화밀집도 등을 통한 주요 관광권역 분석 /추출(6개 주요 권역: 강남/서초, 동대문, 명동/남대문, 이태원, 종로/인사, 홍 대/신촌)



- 데이터 분석 과정
- KT 로밍데이터를 활용한 중국인 유동인구 산출

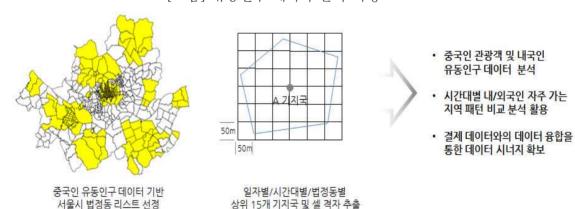
① 유동인구 추출

- 일단위 국제로밍호 테이블에서 서울시 지역에 대한 법정동 단위 유동인구 추출
- BC카드 결제내역 및 동별 시설물 정보 등의 비교 분석을 통한 유의미한 176개 법정동 선정

② 기지국 추출

- 일자별/시간대별/법정동별 유동인구가 높은 상위 15개 기지국 추출
- 통화 발생 기지국 커버리지를 포함하는 50m x 50m 격자셀 단위의 유동인구 정보 분석

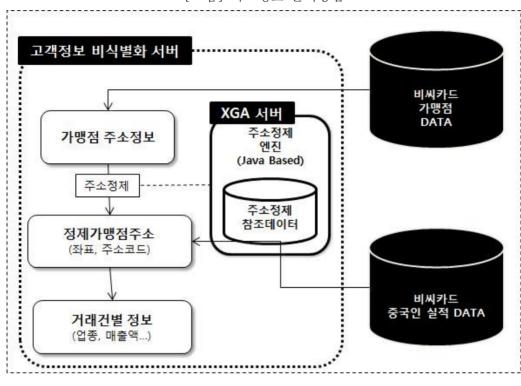
[그림] 유동인구 데이터 분석 과정



③ 제공 단위 가공 및 분석

- 유의미한 시간구간별 분류 : 00-07시, 07-11시, 11-14시, 14-17시, 17-20시, 20-00시(각 시간 구간별 50m*50m 셀 단위 집계)
- BC카드 데이터를 활용한 상권별 중국인 매출액 분석
- ① 고객유형정보 분석: 블록 단위(거주지 주소 기준) 성, 연령 고객 유형 ※ 본격적인 정보분석을 위한 고객정보의 비식별화를 비롯한 거래건별 정보 추출
- ② 내국인 거래패턴 분석: 블록 단위 카드 거래의 성별/연령대별, 시간대별/요일별 패턴

③ 중국인 거래패턴: 블록 단위 은련카드 거래실적 데이터



[그림] 카드정보 분석방법

주요 분석 결과 및 활용방안

□ 주요 분석 결과

- 중국인 관광패턴 분석
- 외래 관광객은 2008년 이후 매년 10%씩 증가하고 있으며, 특히 중국인 관광객은 500% 이상 증가
- ※ 일본 관광객수를 넘어 우리나라 관광산업에서 차지하는 중요도가 크게 상승





- 중국인 관광객 구매횟수는 전년 대비 181% 상승하였고, 구매금액은 전년 대비 120% 상승

□ 중국인 1인당(카드당) 거래 현황

구분	평균 체류기간(일)	평균 방문매장수	평균 구매횟수	평균 구매금액(원)
2013년	2.8	5.9	9.7	2,565,689
2013년 전반기	2.8	5.9	9.5	2,552,406
2014년 전반기	2.9	6.3	10.5	2,235,100
증감	0.1일	0.4개	1회	-317,306

- 명동/동대문 인근 쇼핑타운을 중심으로 밀집도가 높은 것으로 확인되나, 최근에는 강남/서초, 종로권역의 거래증가율이 높게 나타남
- ※ 명동/동대문의 경우 백화점/면세점이 월등히 높았으며, 강남/서초는 의복의류
- 중국관광객 평균 구매횟수는 9.7회, 1인당 소비금액은 256만원으로 전체 관광객 평균 183만원 보다 중국관광객이 40% 추가 소비(2013년 기준)
- 관광객은 주로 명동/남대문(2,458,948건)에서 자주 소비하였으며, 종로/인사 (288%), 강남/서초(239%)의 거래 건수가 전년대비 크게 증가
- ※ 명동/남대문(2,458,948건) → 홍대/신촌(325,006건) → 동대문(322,604건) → 종로/인사(242,882 건) → 강남/서초(223,691건)
- 소비금액 역시 명동/남대문(6조 44억 원)이 가장 높게 나타났으며, 종로/인사동 (200%)이 전년대비 가장 큰 폭으로 증가
- 명동/남대문(6조 44억 원) → 홍대/신촌(979억 원) → 강남/서초(882억 원) → 종 로/인사(450억 원)
- 관광객은 화장품, 의류 등을 주로 구매하고 화장품은 명동/남대문, 종로/인사 동, 홍대/신촌에서 의류는 강남/서초, 동대문에서 주로 구매
- ※ 업종별 거래건수 : 화장품 판매점 → 여성의류점 → 인삼제품판매점 → 토산품/기념품점 → 할인점
- 강남/서초, 종로/인사동은 한식, 동대문은 갈비·삼겹살, 홍대/신촌에서는 닭갈 비를 주로 먹으며, 명동/남대문에서는 커피를 많이 마시는 것으로 나타남

□ 서비스 계획 및 활용방안

- 중국인 대상 모바일 앱서비스에 데이터 분석 결과를 활용한 지도기반의 여행지 안내, 여행코스 추천, 인기 가맹점 소개 등 정보 제공
 - * BC카드의 중국인 대상 앱서비스 '완쭈안한궈'에 기능 탑재, 한국관광공사의 중국인 앱서비스 '한국자유여행'에 기능 탑재



■ '유통시장 분관광서비스 관련 기업, 기관, 지자체 등의 의사결정자들을 위한 트렌드 분석 서비스 제공



효과 및 향후 적용 확대 방안

- 지역 소상공인 또는 소규모 지역 단위 현황분석 요구 사용자 지원
- 해당 사업장 주변 또는 원하는 위치의 중국인 관광객의 소비 규모, 트렌드 파악을 용이하게 하여 중국인 관광객 소비에 대한 이해를 도움
- 관광업 관련 종사자 지원
- 중국인 관광객을 대상으로 하는 관광업 관련 종사자에게 중국인 관광객의 유동 흐름, 여행패턴별, 휴가시즌별 관광객 특징, 관광지별 유동흐름 및 주변 소비특 징 등의 분석을 통해 서울지역 관광패턴과 선호 관광지 등의 정보를 제공함으로 써 중국인 관광객의 관광 및 소비 트렌드를 반영한 관광 상품 개발에 활용

■ 관광관련 정책수립 지원

- 관광관련 정책수립을 위한 각 지역별 통계정보를 제공함으로써 해당 공공기관이 관할하고 있는 지역 내 소비규모, 시간대별 유동객, 선호 업종 등을 분석하여 정책 수립 시 활용함으로써 중국인 관광객의 해당 지역 내 관광 편의 및 지역경제 활성화에 활용

< 참고 >

강남스타일 즐기는... 멋 좀 아는 유커들

서동일 dong@donga.com ·김재형 기자 2015-01-21

中관광객 서울 소비패턴 빅데이터 분석해보니



동아일보가 단독 입수한 미래창조과학부·한국정보화진흥원(NIA)의 '빅데이터 기반의 외국인 관광 산업 지원 시범사업' 결과 보고서에 따른 해석이다. NIA는 지난해 1월부터 올 6월까지 18개월간 중국의 은련(銀聯)카드 거래 명세를 기반으로 유커들이 서울 어느 지역에서 어떤 물건을 사고, 어떤 음식을 먹는지 분석했다. 은련카드는 중국인 90% 이상이 사용하는 것으로 알려진 신용카드다. (중략)

유커들의 관광 행태 중 권역별 차이가 뚜렷한 것은 음식이었다. 강남에서 가장 많이 먹는음식은 양식(43%)이었다. 강남역 인근 파스타·바비큐 전문점 'Big PLATO' 직원 이민영 씨(26·여)는 "중국인들은 주로 모바일 앱을 통해 우리 가게 위치를 확인한 다음 찾아온다"며 "최근에는 유커들을 대상으로 한 음식점 추천 앱을 만드는 관계자들도 '식당 정보 업데이트를 하고 싶다'며 찾아오고 있다"고 말했다. 강남권에서 눈에 띄는 것은 '게장 전문 음식점'(4.3%)이 5위를 차지했다는 것. 강남의 일부 간장게장 전문점은 중국 홍콩 대만 등 동아시아권 관광객들에게 관광 명소로 알려져 있다.

신촌·홍대앞을 찾은 유커들은 닭요리(17.5%)를 가장 많이 먹었고 한식(9.8%) 전문점을 많이 찾았다. 반면 유커들의 '메카' 명동과 동대문에선 카페(39.8%)에서 쓴 돈이 가장 많았다. (중략)

동아일보 기사원문: http://news.donga.com/3/02/20141120/68005542/1

맞춤형 유의질병 및 병원정보 제공

진료정보 빅데이터 분석을 통해 발생 질환별 예상 유의 질병 정보 및 맞춤 형 병원 정보 제공

추진 목적 및 배경

- 사업 추진의 배경
- 정부기관 및 지방자치단체가 보유한 데이터를 누구나 손쉽게 활용하고 이를 통해 새로운 가치를 창조하고자 하는 '정부 3.0'의 방침에 따라 접근이 어려웠던 대규모 보건의료 공공데이터가 개방

As - Is To - be

- 법·제도적 이슈,개인 질병이력 확인 및 체계적 관리 불가능
- 사전예방이 아닌 사후관리 중심으로 인한 의료비 증대
- 평균연령 증가로 인한 노후 의료비 부담 지속적 증대
- 검증되지 않은 병원 및 질환 관련 검색정보증가
- ✔ 국민의료비 약 100조원 (2013년 기준)
- ✔ 국민의료비 가계 직접부담비율 OECD 평균 2배
- 한국 35.8% ↔ OECD 평균 19%
- ✔ 의료비 증가율 OECD 회원국중 최고



- 검진 결과, 처방 기록 및 진료 기록을 확인하고, 체계적으로 관리 가능
- 건강위험요인 데이터를 통해 유해성 질병에 대한 주의 안내를 개인별 맞춤으로 제공



- 의약품 처방·조제 시 병용금기 등 의약품 안정 성과 관련된 정보를 실시간으로 제공하여 부적 절한 약물사용을 사전에 접검
- \odot
- 개인의 검진결과 및 진료기록을 기반으로 필 요한 맞춤 건강정보를 선별하여 제공
- ✔ 건강보험심사평가원, 약 50TB DB 활용
- ✔ 예방중심, 환자중심, 의료서비스 개선을 통해 국민의료비 8조 2000억 절감









■ 사업 추진의 필요성

- 진료/질병/처방 등의 데이터가 지속적으로 생성되고 있는 보건의료 분야의 방대 한 데이터를 활용하여 국민건강 증진을 위한 서비스를 제공하고자 함
- 또한 Open API를 통해 빅데이터 분석 결과를 외부에서 참고 가능하도록 하여, 정부 3.0 및 창조경제 부합하는 새로운 고부가가치 서비스를 창출하고자 함

추진 내용

□ **참여기관**: 메디벤처스㈜, 건강보험심사평가원, ㈜라인웍스

□ 주요 활용데이터

- 건강보험심사평가원의 건강보험청구 데이터베이스 등 60여종의 데이터를 기반으로, 다양한 공공 개방 데이터를 수집 활용함
- ※ 개인정보를 비식별화하여 분석

구분	데이터	데이터 양	제공 기관
건강보험청구 데이터	건강보험청구 데이터베이스 등 60여종 (요양급여비용청구명세서 / 요양기관정 보 / 의약품처방정보)	50 TB	건강보험심사 평가원

□ 분석 내용 및 기법

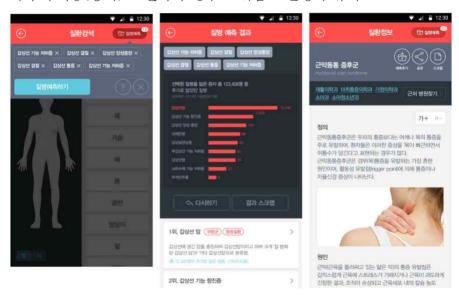
- 데이터 처리 및 분석 기법
- 유사그룹 기반 데이터 선별 (Collaborative Filtering)
- 유사그룹 기반 데이터 선별은 가상의 사용자 그룹을 생성하고, 그 안에서 가장 빈도수가 높은 데이터를 검색하는 기법
- 건강보험청구 데이터에서 주상병-부상병의 연결성을 정렬하여 데이터 분석 기계 학습(Machine Learning)을 위한 학습데이터를 생성

- 질병기준 인구통계학적 데이터 선별 (Demographic Curating)
- 인구통계란 사용자의 부가 속성, 즉 나이, 성별, 거주지 등의 일반적인 정보를 의미하며, 이러한 인구통계학적 데이터와 질병은 연관도가 매우 높아 유행성 질병군의 예방 및 치료를 위한 데이터 추출에 활용
- 서비스를 통해 입력 받은 인구통계학적 정보를 데이터 선별의 기준 조건으로 카이스퀘어 분석(Chi-Square Test)을 통해 속성과 질병과의 연관성을 분석

주요 분석 결과 및 활용방안

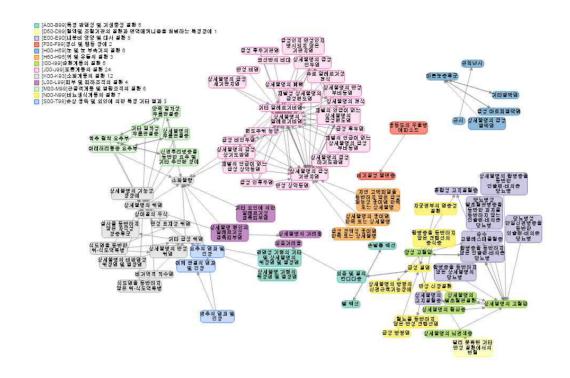
□ 주요 분석 결과

- 국민질환 통계자료를 기반으로 개인 발생 질환별 유의질병 정보 제공
- 예) '치아 우식증(충치)' 환자의 경우 '위염' 발생에 유의



- 의료정보 빅데이터 분석모델 개발
- 본 시범사업에서는 건강보험심사평가원의 대용량 보험청구데이터를 분석하여 유 의질병 빅데이터 모델을 개발
- 분석방법으로 데이터마이닝분야에 저명한 기법인 유사그룹 기반 데이터 선별기 법 (Collaborative Filtering)을 사용하였고, 주상병-부상병의 연결성을 정렬하 여 기계학습(Machine Learning)에서 사용할 데이터 셋을 생성
- 아래의 그림은 추출한 데이터로 생성한 연관질병 그래프이며 해당 그래프의 연결성을 의사결정트리의 기준으로 분류하여 사용자에게 유의질병을 선별 제공

[그림] 연관질병 데이터 그래프



효과 및 향후 적용 확대 방안

- 보건의료 빅 데이터를 기반으로 국민건강 증진을 위한 공익 서비스 제공
- 의료정보의 공개·개방·활용 확대를 통한 맞춤형 서비스 제공 및 정보이용 활성화를 통해 어렵고 복잡하게 만들어진 데이터를 사용자가 보다 쉽게 접할 수 있도록 함으로써 국민건강 증진에 기여함
- 정부 3.0 및 창조경제에 부합하는 새로운 부가가치 및 일자리 창출
- 데이터 공개 및 Open API 방식으로 빅데이터 분석결과를 외부에서 참고 가능하 도록 개방함으로써 신규 비즈니스 창업 및 일자리 창출에 기여함

자동차 부품기업 공동 활용 빅데이터 플랫폼

데이터 분석 기반의 제품 품질향상을 위해 자동차 부품 제조사가 공동 활용할 수 있는 빅데이터 플랫폼 구축 운영

추진 목적 및 배경

■ 사업 추진의 배경

- 한국의 자동차 산업은 2013년도 전체 452만대 생산하여 세계 점유율 5.2%로 9년 연속 세계5위로 산업의 10% 이상 점유하여 가장 큰 제조 분야를 차지
- 빠르게 성장하고 있는 중국 자동차 생산량은 2013년 2212만대로 전체 1위를 하 였으며 최초 2000만대 돌파
- 빅데이터 서비스 가치에 대한 인식부족으로 자동차산업의 경쟁력을 좌우하는 자동차부품 업체의 공장/품질 정보에 대한 분석 서비스 부재





■ 사업 추진의 필요성

- 2012년 한국자동차산업은 유럽 발 경제위기 및 고유가 속에서도 한국차의 품질 및 신뢰도 향상, FTA 발효 확대 등으로 수출 3백17만대, 수출금액(부품포함) 718억 달러로 사상 최대 수출실적을 기록하였지만, 국내 시장의 경기 불확실성 지속 및 가계부채 증가에 따른 소비심리 위축으로 국내 생산은 3년 만에, 내수 판매는 4년 만에 각각 감소세를 나타냄

※ 자동차 산업 현황

*생산: 내수부진 및 임단협 기간 중 공급차질로 3년 만에 감소세로 전환

*내수: 경제성장률 하향과 가계부채 부담, 신차효과 약화 등으로 감소세를 보임

*수출: 국산차의 품질 및 브랜드 이미지 상승, 주요국과의 FTA 체결로 인해 경쟁력 상승 *수입: 최근 수입차에 대한 소비자 인식 제고 등으로 국내 수입차 시장규모가 급격히 증가

- 자동차산업의 국민경제적 비중(2만 여개 부품으로 생산되는 전후방 연관효과가 가장 큰 산업)은 제조업 생산의 11.4%, 고용의 10.7% 및 부가가치의 10.6% 차지 (2011 통계청 광업제조업통계 조사보고서 참고)
- 2012년 기준 자동차산업의 무역수지는 616.5억불로 반도체, 선박류 등 주력 기 간산업과 비교해 볼 때 국가 경제에 이바지 하는 바가 큼
- 연계 산업의 파급효과와 고용효과가 큰 자동차산업의 경쟁력 강화를 위해 강소 자동차 부품기업 육성이 필요하고, 이를위해 자동차부품들이 공동으로 활용할 수 있는 데이터 분석을 통한 품질 및 생산성 향상 공통 서비스 제공 필요
 - ※ 데이터 분석은 기존 시스템의 정보를 활용하여 추가 투자가 많이 필요치 않고, 공정개선 만으로 품질과 생산성 향상이 가능

추진 내용

- □ 참여기관: 메타빌드㈜, 자동차부품연구원, 솔바테크놀로지
 - * Data 제공업체(자동차 부품기업 K 사)
 - K사는 3차 기업으로 주조 가공 전문 업체이며 다이캐스팅 주조 가공 전문기 업이다. 칠천만불 수출의 탑을 수상하였으며 동종업계최초 ISO 16949 인증, 중소기업인 대통령상 수상을 한 바 있다.
 - 매출 중 내수 및 해외가 54%, 공조, 엔진, 변속기, 조향 부품에서 46%를 차지하고 있으며, 내부 주조 관리 MES(Manufacturing Execution System)가 가동 중에 있다. 전문 가공 공장 시스템을 운영하고 있으며 추가 도입을 계획중에 있다. 본 사업을 위해 품질관련 데이터 제공 및 업무 공조 협약 맺어데이터를 제공하였다.

□ 주요 활용데이터

구분	데이터	데이터 규모	보유기관
	설비 설정값 실측정보	564,946건(540MB) *표본 샘플링 6개월 치	
설비 및 품질데이터		2만건(14MB) *표본 샘플링 6개월 치	K 사
	작업장 온습도 정보	24만건(70MB) *표본 샘플링 6개월 치	

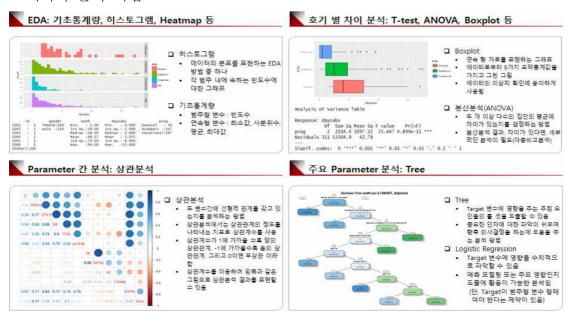
- 다이캐스팅 주조공정 장비 설정데이터, 실측데이터 6개월치(MES에서 추출)

0.463 m/s 1.537 m/s 254 mm

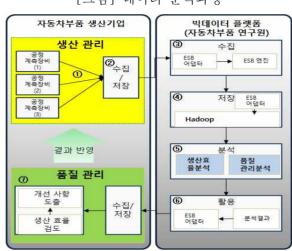
[그림] 분석대상 알루미늄 다이캐스팅 설비

□ 분석 내용 및 기법

■ 데이터 분석 기법



■ 데이터 분석과정



[그림] 데이터 분석과정

① 데이터 수집

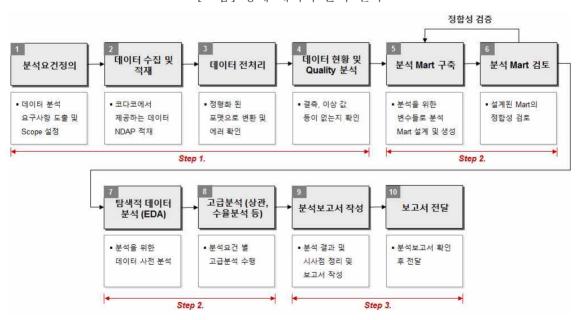
- 분석된 결과 데이터를 해당 부품기업의 전용 웹 사이트에 정보 제공
- 필요한 내부 시스템으로 연계 전송 반영

② 데이터 저장

- 자동차부품 공통 분석 플랫폼의 하둡서버에 추출된 데이터 저장

③ 데이터 분석

[그림] 상세 데이터 분석 절차



■ 탐색적 데이터분석

- 분석 대상으로 선정된 데이터의 탐색적 분석을 통해 다양한 분석 포인트를 도출 하고 변수별 기초통계량 파악, 변수 별 패턴 및 분포 확인

■ 각 호기 별 차이 분석

- 변수 별 차이 분석: 작업 호기, 프로세스, 작업시간, 작업자 등 조건에 따른 불량률 및 수율차이 분석

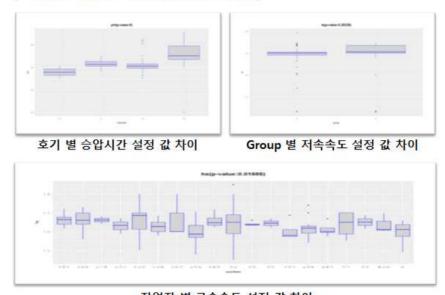
₹/0‡ 至7 작업자 코아 11.6096 11.00% 22 34.60% 22.70% 12.20% 4 김걸 12,40% В 11,1096 23 12.50% 22.40% 7.90% 24 12.20% 국형진 8.40% 25 7.70% 김면수 4.20% MUS NUS BXL - D 불량율 cavity 주야 그룹 작업자 • 호기 불발용 0.223 0.130 0.158 0.126 0.129 0.25 0.140 0.130 0.120 night 0.130

[그림] 작업조건별 불량률 분석

- 설비 설정 값 간 상관 분석
- 6개의 설비 설정 값의 현황 분석, 시간 별 패턴 분석
- 주조에 사용되는 설비 설정 값 간의 연관성을 상관분석으로 분석

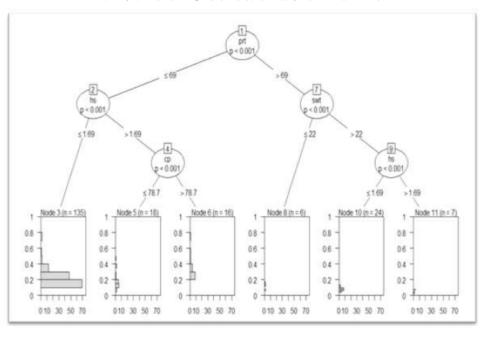
[그림] 속성별 불량률 분석

[속성 별 Parameter 설정 값에 대한 ANOVA]

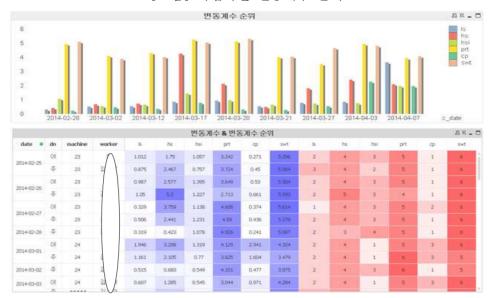


- 작업자 별 고속속도 설정 값 차이
- 불량률에 영향을 주는 요인분석
- 통계 모델링을 통하여 불량률에 영향을 주는 요인을 분석

[그림] 의사결정나무기법을 활용한 요인 분석



- 설비설정값 중 수율에 영향을 주는 주된 요인 분석
- 설비설정값과 실측값 사이의 차이분석, 설비설정값 과 실측값의 변동성 분석을 통해 작업자 별로 각 설정 값을 적절히 조절하고 있는지 파악



[그림] 작업자별 변동계수 분석

④ 분석결과 활용

- 분석결과의 현장적용
- 인터뷰 조장의 주조조건 설정 방안을 수평교육 후 결과 확인

[표] 전/후 조건설정치 비교

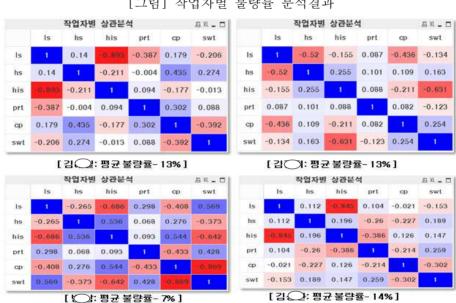
구 분	भी भी जी भी					전체평균				
一 一 七	생산장비	생산수량	불량합계	불량률	저속속도	고속속도	고속구간	승압시간	주조압력	탕구두께
1~6월	23,24,25	334.2	34.5	10.3%	0.59	1.64	244.9	63.4	77.9	24.5
9~10월	24	405.8	26.7	6.6%	0.61	1.58	232.1	58.3	69.2	23.5
차이	*	+71.6	-7.8	-3.7%	+0.02	-0.06	-12.8	-5.1	-8.7	-1

- 전체적인 주조조건의 변경 차이가 나타남
- 주조압력을 기준으로 고속속도부터 변경 적용 후 나머지 조건을 미세하게 변경
- 외형적인 결과를 고려했을 때, 약 3.7%의 불량률이 감소하였음
- 현장 적용 후의 성과: 불량률 감소에 따른 생산비용 절감 기대

주요 분석 결과 및 활용방안

□ 주요 분석 결과

- 불량률에 영향을 주는 장비 설정 변수 도출
- 분석 대상으로 결정된 '자동차 변속기 하우장' 생산장비인 알루미늄 다이캐스팅 장비의 생산 수율에 영향을 미치는 변수 도출
- ※ 사례) 장비 승압시간 설정값이 69 이하이고, 고속 속도가 1.69를 초과하며, 주조압력이 78.7을 초과하는 경우, 0.2 이상의 불량률이 발생
- 동일 공정에 대해 작업자별 운영방법 상이
- 불량률 저하를 위한 장비 운영 매뉴얼 없이 숙련자의 직감에 의존하는 형태로 운영
- 작업자별 변수 설정값이 다르고, 불량률 차이도 심함



[그림] 작업자별 불량률 분석결과

□ 서비스 계획 및 활용방안

- 장비 운영 매뉴얼 제작
- 분석 대상을 확대하고, 정밀 분석을 거친 후 장비 운영 매뉴얼 및 가이드 제작 에 활용 예정
- ※ 분석결과를 반영하여 장비운영 조건을 실험한 결과 3.7%의 불량률 감소
- 분석 지원 대상 기업 확대 및 자동차 부품기업 공동활용 빅데이터 플랫폼 운영 계획 수립·추진

ndap3 ndap4 ndap5 ndap6 ndap1 ndap2 Name Node Name Node Hadoop Cluster Data Node Data Node Data Node Data Node NDAP Hbase Master TaskTracker TaskTracker | TaskTracker | TaskTracker Rstudio Zookeepe Server Region Region Node 1 Server 1 Server 2 Server 4 Hive Store

[그림] 자동차 부품기업 공동활용 빅데이터 플랫폼

※ KT NexR - NDAP 시스템(하둡 기반)

효과 및 향후 적용 확대 방안

■ 기대효과

- 부품 품질 향상은 생산 및 공급 수율을 증가시켜 직접 매출 상승에 기여할 것으로 기대(약 10% 예측)
- 부품데이터 분석 기반의 생산품질 향상(불량률 저하를 통한 수익향상)
- 제조업 빅데이터 활용 우수사례 제시

■ 사업 활용 방안

- 자동차 부품 기업 공동 활용 빅데이터 분석 플랫폼은 자동차 부품 연구원에 등록된 중소기업 대상의 공공 인프라로 사용을 원하는 중소 자동차 부품 기업 사용 신청 시 일정 조정 및 승인 및 데이터 분석 업체 리스트 제공 등 모든 행정적 절차를 자동차 부품 연구원이 전담 운영 인력을 배치하여 관리
- 빅데이터 분석 컨설팅 사업을 원하는 기업을 대상으로 현 분석 상황 및 진행 절차, 과정, 리소스 등의 다양한 정보를 구체적으로 제공 창업을 유도 지원
- 센서를 통한 데이터 수집을 연구 기획하여 효율적인 공통 데이터 수집 방안과 체계를 마련하여 제시 하는 과제 개발 기획
- 동종업계 시범 사업자와 협조하여 분석 방법론의 기술적인 교육 기회 부여하고 기술 개발을 통한 추가 확장 서비스 개발 시 플랫폼 시범 활용 서비스 기회 제공

조류 인플루엔자(AI) 확산 조기대응

국가동물방역통합시스템(KAHIS) 데이터와 KT의 통화로그 데이터를 연계 분석하여 조류인플루엔자(AI)의 확산 경로 예측

추진 목적 및 배경

- 사업 추진의 배경
- 현재 탐문조사와 방역활동 등 AI 발병 후 대응을 중심으로, 광범위 확산에 대한 효과적 방지에 한계
- KT 기지국 통계 및 농림축산검역본부 KAHIS 축산차량 이동 데이터를 활용하여 AI 확산 지역 예측 모델을 개발





- 사업 추진의 필요성
- AI확산의 명확한 규명을 통한 2차 감염피해 예방
- 축산농가의 막대한 피해액 최소화
- 효율적 방역을 통한 국가 예산의 절감
- 빅데이터 기반 농축산 분야 ICT 융복합 촉진을 통한 공공 이익 극대화

추진 내용

□ 참여기관: ㈜KT, 농림축산검역본부

□ 활용데이터

구분	데이터	데이터 양	제공 기관
통화로그(CDR)	기지국 정보	67TB/월	KT
	기지국별 통화량 통계 정보		
KAHIS-축산차 량, 농장정보	축산차량 농장방문 기록 정보	200MB/월	
	전국 농장 정보	90MB/월	농림축산검역본 부
	AI 발병 농장 정보		
	축산 차량 등록 정보		

□ 분석 내용 및 기법

- 데이터 처리 및 분석 기법
- DATA 수집·전처리
- KT 데이터 분석 전문가가 내부 분석 플랫폼 활용하여 데이터 전처리, 통계 테이블 생성 등 분석을 위한 형태로 데이터 가공 작업 진행
- 가금류 축산업 관련자 인터뷰 수행
- 관련자 심층 인터뷰를 통해, AI 확산 매커니즘 규명 및 예측모델 개발에 활용
- 인터뷰 대상: 동물 감염병 전문가, 가금류 축산업 종사자, 가금류 농장주, 가축 운반, 사료운반 등 기능별 관련 종사자, 방역 담당 공무원 등
- AI 확산 매커니즘 규명
- AI 발생 농가별 발생원인 및 확산 매개체 분석
- 기지국 통계 데이터와 가금류 농장 관련자들의 차량이동데이터를 분석하여 농가 별 발생원인 분석
- AI 확산의 핵심 요인으로 추정되는 확산 매개체 선정
- AI 확산 관련 핵심 요인 추출
- 발병 농가 사이의 연관성 분석

- GIS 기반 분석 및 시각화
- KT의 GIS 전문가 인력 투입하여 확산 매커니즘 분석 결과를 지도상에 시각화
- 공간분석, 방문 매개체 분석, 거리측정, 농가 위험도 분석 및 시각화 작업 수행

■ 데이터 분석 과정

① AI 확산 예측모델 개발

- KT 기지국 통계 데이터 및 농림축산검역본부 KAHIS 축산차량 데이터를 통한 차량·사람 이동과의 연관성 분석
- 발병 농가 특성 및 발병 농가 방문 차량 특성별 연관성 분석 확산 매개체의 핵심 요인 선별 및 가중치 부여
- 이를 적용한 AI 확산 예측 모델 개발

② AI 확산 예측모델 정확도 검증

- 과거 AI 발병 이력 데이터를 활용한 가상 예측 시뮬레이션 수행
- 예측 모델에 의한 결과를 실제 발병이력과 비교하여 성능 검증

③ AI 확산 위험지역 선별 제공

- 농림축산검역본부 위기대응센터에 일별 확산 예상지역 지도 및 방역 대상 농가 리스트 전달
- Web 기반 GIS 시각화 및 레포트 제공

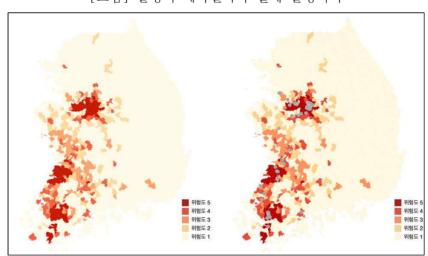
[그림] AI 확산경로 예측 및 대응 프로세스



주요 분석 결과 및 활용방안

□ 주요 분석 결과

- 가축전염병의 주요 전염확산 요인은 차량으로 차량 이동경로를 따라 질병이 확산 되는 것을 확인
- 사료운반, 가축운반, 분뇨차량, 시료채취·방역 차량 등 차량 유형별 확산 확률이 다름
- 이들 차량의 이동 데이터를 토대로 지역별 전염 위험도 시뮬레이션
- AI 발병농가의 약 83%가 고위험 추정지역(위험도5 : 전국영토의 약 4%) 내에서 발생한 것을 확인



[그림] 발생지 예측결과와 실제 발생지역

□ 서비스 계획 및 활용방안

- 매월·필요시 분석리포트를 농림축산검역본부에 제공
- 향후 가축전염병 확산예측 시스템으로 확대 구축하여 운영하여 지자체 공무원이 방역을 위한 세부 행동계획을 제시



[그림] 가축 전염병 확산예측 시스템

효과 및 향후 적용 확대 방안

- 가축전염병에 대한 초동대처가 가능해져 확산 조기방지 및 피해최소화 ※ 구제역, 돼지콜레라 등 타 전염병으로 확대 가능, 일본중국 등에 서비스 모델 수출 가능
- 타 사람/가축 전염병 확산에 대한 확대 적용
- 현재 발병/확산되고 있는 구제역과 같은 피해 규모가 큰 가축 전염성 질병에 확대 적용
- 가축 전염병은 AI 적용사례와 유사하므로 축산 차량 이동분석을 통해 확산 지역 예측가능
- AI 확산 예측모델 자동화를 통한 실시간 대응체계 구현
- 서버 내 AI 확산 지역 예측 모델 적용 모듈 개발
- 확산 위험지역을 확인할 수 있도록 지도상의 표현기능 제공(Web 기반)
- 확산 위험 지역 리스트 확인 가능 리포트 daily 제공

국도 비탈면 붕괴사고 예측

급경사지 지형 데이터와 기상정보를 연계 분석하여 국도 비탈면의 위험도 산정 모델 개발 및 예측 서비스 제공

추진 목적 및 배경

■ 사업 추진의 배경

- 우리나라는 매년 장마철이나 태풍 내습과 관련된 집중호우에 의해 산사태나 비탈면 붕괴가 빈번하며, 최근 15년간 매년 평균 53번의 비탈면 붕괴로 인해 20여명의 사상자가 발생함



■ 사업 추진의 필요성

- 강우는 급경사지 붕괴의 중요요소임에 따라, 실제 강우가 많이 발생하는 우기철 에 급경사지의 불안정화로 인한 붕괴가 빈번히 발생하고 있음
- 강우침투에 의한 급경사지 안정해석 모델 연구가 많이 진행되어 있으므로 이러한 모델을 기반으로 기상자료와 연계한 급경사지 특성, 사고이력에 대한 빅데이터 분석을 통해 사고예측 필요

추진 내용

□ 참여기관: 대한지적공사, 한국건설기술연구원, 모바일팩토리, SKT

□ 활용데이터

구분	데이터	데이터규모	보유기관
도로비탈면데이터	도로비탈면 기초조사자료	6,587 지점	. 건설기술연구원
	도로비탈면 정밀조사자료	936 지점	
강우량등 기상정보	일강수량, 3일 누적 강수량, 7일 누적 강수량	1년간 데이터	. 기상청
	일평균기온	1년간 데이터	

□ 분석 내용 및 기법

- 데이터 처리 및 분석 기법
- 도로비탈면유지관리시스템의 기초 및 정밀 조사 데이터 분석
- 기상조건 및 공간정보 인자 도출
- 빅데이터 분석 기법을 활용한 비탈면 붕괴 위험 지수 모델 개발
- 위험 경보 서비스를 위한 위험 등급 설정 및 지오펜싱(geofencing)1) 기술 개발
- 실시간 기상 정보 연계 비탈면 위험 예측 경보 모바일 서비스 개발

¹⁾ Geographic과 Fencing의 합성어로 특정 구역에 대한 사용자 출입현황을 알려주는 API

대국민 비탈면 정보 제공 모바일 서비스 비탈면 사고 예측 시스템 6번 국도 상행선 용행에서 준내방향 14K 낙석 및 용괴 위원이 안스다다 비탈면 위험 예측 모델 개발 기상청 $y = w_0 * (rainfall) + w_1 * (slope) + w_2 * (aeration) + w_3 * (rock) + ... + e$ 실시간 기상정보 1×81 A0001 20° F Massive 45 A0002 Tabular 78 지오펜싱을 활용한 위치기반 실시긴 A0003 Irregular 32 한국건설기술연구원 비탈면 붕괴위험 모니터링/시뮬레이션 시스템 **CSMS** 비탈면 정보 10-30 Α 31~50 51~70 71...00 91~100

[그림] 비탈면 사고 예측 시스템 구성도

■ 데이터 분석 과정

빅데이터를 활용한 국도 비탈면 붕괴 위험도 예측 모델 수립 및 서비스 방안 도출은 DMAPM 5단계를 거쳐 수행됨. DMAPM 단계는 분석의 수행 범위 및 세부 목표를 정의하는 Define 단계, 분석 대상 데이터의 수집 및 특성 파악과 전처리를 수행하는 Measure 단계, 기존 예측 모델 및 핵심 인자 검토 그리고 새로운 핵심 인자 도출을 위한 그래픽분석 및 통계분석을 수행하는 Analyze 단계, 분석된 결과를 토대로 위험도 예측 모델 수립 및 위험도 등급별 인자 값 결정과 타당성 검증을 실시하는 Predict 단계, 붕괴 위험도 예측 모델을 활용한 기상정보 변화에 따른 위험 경보 발생 체제 수립 및 서비스 방안 도출을 위한 Monitor 단계로 구성

[그림] 데이터 분석 과정



- ① Define 단계: 분석의 수행 범위 및 세부 목표 정의
- 국도 비탈면을 여러 특성(지질학적 특성, 환경적 특성, 지리적 특성 등)에 따라 구분하고 이를 바탕으로 실제적인 분석의 수행 범위를 결정함. 또한 기존의 비탈면 붕괴 위험성을 평가할 수 있는 비탈면 재해 예측법 및 기준을 고찰하여 이러한 방법론의 한계점을 파악하고 개선방안을 도출
- 최종적으로 분석하여 개발할 비탈면 위험도 모델의 유형(수리적 모형, 인과형 모형 등)을 결정하며 개발되는 모델의 수준(모델의 정확도, 정밀도 등)을 결정 하여, 이를 토대로 제공되는 서비스의 유형(위험도 수준에 따른 경보 서비스, 위험도 확률 제공 서비스, 실시간 위험도 제공 서비스 등)과 제공되는 형태 (PUSH알림, 붕괴 위험 지도 등)를 결정
- ② Measure 단계: 도로비탈면유지관리시스템 데이터와 기상 데이터의 메쉬업
- 도로비탈면유지관리시스템의 비탈면 정밀조사 데이터와 기상청에서 제공하는 국가 기후자료센터로부터 획득한 기상 데이터를 위도와 경도를 기준으로 메쉬업 실시
- 데이터 메쉬업을 통해 국도 비탈면 붕괴 위험도 분석에 비탈면의 고유특성을 나타내는 인자(비탈면 관리시스템의 정밀조사 데이터)와 실시간으로 변화되는 인자 (기상청의 기상 데이터)를 동시 고려하는 실시간 붕괴 위험도 예측 모델 수립
- 분석의 용이성을 위한 데이터 정제(Cleaning) 및 전처리(Preprocessing)
- ③ Analyze 단계: 그래프 분석 및 문헌조사를 통한 붕괴 주요 인자 도출
- 데이터를 붕괴 레코드와 미붕괴 레코드로 분류하여 각각 그래픽 분석을 실시. 분석되는 인자의 데이터 종류가 연속형인 경우 Box Plot을 활용하고 범주형의 경우 Dot Plot을 사용하여 붕괴와 미붕괴 간에 확연한 차이 보이는 인자들을 선 별한 후 통계분석으로 확정
- 비탈면 붕괴와 관련된 50편의 논문에 대한 문헌조사를 통해 비탈면 붕괴에 영향을 미치는 인자들을 확인하고 이를 토대로 비탈면 붕괴에 유의한 후보 인자 리스트를 작성함. 또한 비탈면 붕괴를 예측하는 기존의 예측 모델을 검토
- ④ Predict 단계: 데이터 분석 기법
- 붕괴 이력 기반 빅데이터를 활용한 비탈면 분석 모델 도출
- 연관분석을 통한 붕괴 패턴 파악, 군집분석 및 의사결정나무 분석을 통한 패턴 도출과 전체 레코드의 군집화
- 의사결정나무 분석을 이용한 군집별 붕괴와 미붕괴 분류분석
- 기상정보에 따른 비탈면 위험도 등급 제시
- 최종 위험도 산정 모델에 대해 기존 data set 및 외부 data set을 활용하여 모델 검증을 실시하고 정확도를 개선할 수 있는 방안 제시

- ⑤ Monitor 단계: 서비스
- 비탈면 위험도 경보 서비스를 위한 분석 결과 DB화
- 위험 경보 발생 순서도 작성 및 시스템 흐름도 제시

주요 분석 결과 및 활용방안

□ 주요 분석 결과

- 붕괴여부에 영향을 미치는 인자 도출
- 강우량, 비탈면의 수분 함유량, 지하수 맥 여부, 풍화도 등이 붕괴에 영향을 미치는 주요 인자로 판명
- 인자들 간 영향도 분석(상관분석)을 수행 하여 붕괴/미붕괴의 특성 패턴화

[그림] 붕괴영향인자 분석결과



□ 서비스 계획 및 활용방안

- 대한지적공사의 대국민 모바일 서비스 앱 "LX토지알림e"에 비탈면 위험알람 서비스 및 주변 비탈면 정보 조회서비스를 적용하여 배포
 - ※ 구글 Play 스토어 및 앱 스토어(iOS)를 통해 배포(2015년 2월)
 - ※ 인터넷기사, 홈페이지, SNS, 블로그 등 다양한 매체를 통해 홍보 추진
- 향후 T맵 등 내비게이션 서비스와 협력하여 우회경로 안내 서비스 추진
- 한국건설기술연구원과 대한지적공사가 협업체계 구축하여 국토교통부 도로운영과 에 정책 제안
- 강원영서 시범사업 지역 외 지자체로 확대 추진
- 지자체 2016년 예산 반영 추진

■ 활용 서비스

[그림] 국토비탈면 붕괴사고 예측 서비스



- 대국민 모바일 서비스
- : 주변 비탈면 위험정보 조회 및 지오펜싱 기술을 활용한 위험지역 진입 시 위험알람 서비스 제공



- 국도 비탈면 위험예측시스템
- 기상정보와 연계하여 실시간 비탈면 붕괴 위험 도 조회, 시뮬레이션, 정보 검색으로 관리체계 고도화 가능
- 예측 모델 검증 후 비탈면 보강공사 우선순위 결정 등 의사결정에 활용
- 3차원 지형분석기술 활용 등 붕괴예측 서비스 고도화

효과 및 향후 적용 확대 방안

- 성과 및 시사점
- 공공데이터와 민간데이터가 융합되어 가치있는 정보를 국민에게 제공하여 안전 사회 구현
- 과학적인 비탈면 보강공사 우선순위 선정에 활용하여 예산 절감
- 지자체가 관리하는 지방도에 대한 실태조사 추진과 예측 모델 적용으로 비탈면 관리사업 확산
- 국가재난재해서비스와 연계한 서비스 확대(추후 관계 기관 협의를 통해 반영)

■ 향후 발전방향

- 국토비탈면 붕괴관련 모바일 활용 시범서비스의 제공 후 분석 데이터와 서비스 종류를 확대하고 서비스 확산 및 홍보

[그림] 국도비탈면 관리 서비스의 발전방향



- 3차원 지형분석기술을 활용하여 고도화된 붕괴예측 서비스 추진

[그림] 국도비탈면 붕괴예측서비스(안)

