

**빅데이터 플랫폼 기반  
분석서비스 지원 사업  
요약 결과보고서**

**2022.12.**

**NIA 한국지능정보사회진흥원**

**금융·소방안전·지역경제**  
**(7개 과제)**

# 빅데이터를 활용한 사용자 인테리어 스타일에 적합한 인공지능 홈퍼니싱 자동추천 서비스(갤러리티)

## □ 과제 목표

- 주거 공간의 인테리어 스타일에 어울리는 최적의 홈퍼니싱 제품 자동 추천서비스 구현
  - 홈퍼니싱 자동추천 서비스 구현을 위한 딥러닝 기반의 공간(거실), 소품(의자, 소파 등) 스타일 분석 AI 모델 개발

## □ 분석주제

- 인테리어에 대한 소비자 개념이 개보수에서 개인별 취향을 반영한 집 꾸미기로 변화함에 따라 스타일 공간분석 기반 추천 서비스 필요
  - AI 모델 개발 및 운영을 위한 컨설팅
  - 홈퍼니싱 자동 추천 서비스 구현을 위한 스타일 분석 AI 모델 개발
  - 클라우드 기반으로 최적화된 AI 모델 개발

### <활용 데이터 현황>

순번	수집채널	데이터명	기간
1	금융빅데이터플랫폼 (비씨카드)	온라인 결제 기반 분석 데이터	2022. 03. ~ 2022. 08.
2	내부 데이터	인테리어티 거실 공간 데이터	2022. 01. ~ 2022. 10.
3	외부데이터	소품(의자/소파) 데이터	-

## □ 분석 결과

- AI 모델 개발 및 운영을 위한 컨설팅
  - 라벨링 작업을 위한 컨설팅과 모델 활용을 위한 홈퍼니싱 자동 추천 서비스 기획 컨설팅 및 MLOps 플랫폼 적용 컨설팅 진행
  - 추천 알고리즘, 초기 구축 컬럼 구성 등 추천 서비스 데이터셋 컨설팅
  - 컨설팅을 통한 인테리어티 추천시스템 연계방안 구상 및 추천 서비스 시나리오 지원

- 홈페이지 자동 추천 서비스 구현을 위한 스타일 분석 AI 모델 개발
  - 학습 모델중 처리량(throughput), 지연속도(latency) 등을 고려하여 매개변수가 18.9M인 ViT\_18.9M 모델을 최종모델로 선정
  - 공간 이미지 5,173개에 스타일 분류 AI모델을 적용하여 6개 범주에 대해 52.25%의 분류 성능 달성
  - 소품 이미지 26,133개에 스타일 분류 AI모델을 적용하여 6개 범주에 대해 54.48%의 분류 성능 달성

구분	공간 스타일 모델	소품 스타일 모델
모델 이미지		
SOTA	-	49.20%
정확도	52.28%	54.48%

<그림 2> 스타일 분석 AI 모델 개발 결과 요약

- 클라우드 기반으로 최적화된 AI모델 개발
  - 개인용 PC에서 운영할 수 있도록 3단계 경량화 및 최적화 과정 진행

## □ 주요 성과

- 사용자 취향을 반영한 “인테리어 AI큐레이팅” 서비스 로드맵 완성
  - 스타일 분석 AI모델을 적용하여 ‘인테리어 AI큐레이팅’ 서비스 제공 예정
  - 컨설팅을 통해 추천 서비스 고도화 및 플랫폼 추천 부문 개선을 위한 구체적인 방안 및 로드맵 완성
- MLOps 플랫폼 적용 방안 컨설팅을 통한 운영비용 절감
  - 홈페이지 인테리어 소품 자동추천 서비스 구현을 통한 거래 활성화 및 고객 유입 증대 및 스타일 사전 분석을 통한 업무 효율화와 운영비용 절감 기대

# 자사 온라인몰, 인터넷, 오픈마켓, 경쟁사별 소비패턴 분석을 통한 마케팅 전략 수립(내츄럴엔도텍)

## □ 과제 목표

- 수요기관 판매 품목의 시장진출을 위한 마케팅 전략 수립
  - 판매 품목별 주 소비층 분석 및 매출추이 분석
  - 수요기관 판매 품목의 온라인 광고 SNS 분석
  - 소비자 관심 유발을 위한 온라인 광고 효율화 전략 컨설팅

## □ 분석주제

- 판매품목별 주 소비층 분석 및 매출 추이 분석
- 온라인 광고 SNS 분석
- 온라인 광고 효율화

### <활용 데이터 현황>

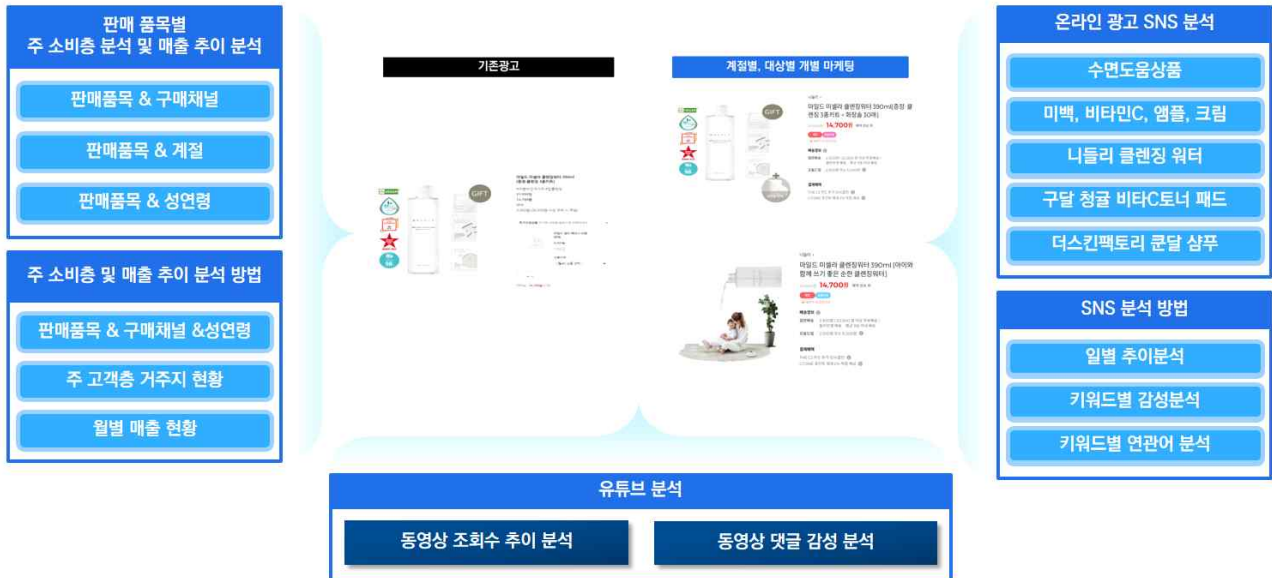
순번	수집채널	데이터명	기간
1	금융빅데이터플랫폼 (비씨카드)	온라인채널별 구매품목별 인구통계학별 소비데이터	2021. 06. ~ 2022. 03.
		온라인채널별 구매품목별 구매시점별 데이터	2021. 06. ~ 2022. 03.
2	외부 데이터	수면 도움 상품 관련	2021. 10. ~ 2022. 09.
3		미백, 비타민C, 애플, 크림 관련	2021. 10. ~ 2022. 09.
4		니들리 클렌징 워터 관련	2021. 10. ~ 2022. 09.
5		B브랜드 토너패드 관련	2021. 10. ~ 2022. 09.
6		더스킨팩토리 A브랜드 샴푸 관련	2021. 10. ~ 2022. 09.

## □ 분석 결과

- 수요기관 판매 품목별 주 소비층 분석 및 매출 추이 분석
  - 구매 예상지수가 높은 고객층의 유입을 위한 마케팅 실행
- 온라인 광고 SNS 분석
  - 수집된 각 키워드들을 16개 키워드로 세분화하여 인사이트 도출

## ○ 온라인 광고 효율화

- 매출이 급증한 상위 3개의 월을 SNS 연관어, 감성분석을 진행하여 공통점 도출



## □ 주요 성과

### ○ 효과적인 유튜브 광고로 인한 매출 증가

- 구매데이터와 유튜브 데이터를 병합하여 효율적인 유튜브 광고를 선정에 따라 내츨엔도텍의 자사 제품 영업 전략이 다양해짐
- '23년 매출액은 전년도에 비해 10% 증가할 것으로 예상되며, 추가적으로 내츨엔도텍의 판매 제품을 영상으로 제작한 경험이 있는 유튜버 리스트를 확보하여 소비자들에게 영향력 있는 유튜브 광고 진입이 원활할 것으로 예상됨

## 초단기 일자리 매칭, 커리어 개인 매니저 플랫폼 [(주)더뉴아크]

### □ 과제 목표

- 단기 일자리 정보 제공으로 구직자와 구인 사업장 일자리 매칭 및 온라인 경력 증명서 발급을 통한 금융상품 매칭 플랫폼 개발

### □ 분석주제

- 타겟 구체화를 위한 일자리 구인·구직 시장 현황 분석
- 소셜데이터 기반 마케팅 활용 컨설팅
- 금융상품 개발을 위한 기존 대출 시장 흐름 및 금융상품 특성 분석

#### <활용 데이터 현황>

순번	수집채널	데이터명	기간
1	금융빅데이터플랫폼 (비씨카드)	BC카드 가맹점 업종 분류	2021.10.
2		연령대별 대출정보 익명화 데이터	2019.03. ~ 2022.04.
3		환율_주식시장과 대출정보	2019.11. ~ 2022.04.
4		금융 사 카드이용 평균정보	2021.11. ~ 2022.04.
5		담보/신용 대출금리정보	2021.11. ~ 2022.04.
6		대출상품정보	2021.11. ~ 2022.04.
7		지역, 연령, 성별 카드대출과 현금서비스 이용금액	2020.01, 2021.01, 2022.01
8	외부 데이터	소셜데이터	2019.07. ~ 2022.06.
9		공공데이터(통계청, 워크넷)	2017.08 ~ 2022.08.

### □ 분석 결과

- 목표시장 진입 관련 분석 수행
  - 공공데이터 분석을 통해 수도권에 거주하는 비정규직 '보건업 및 사회복지 서비스업에서 종사하는 고졸 출신의 60대 이상의 여성'의 단기 일자리 수요가 높아질 것을 예상
  - BC카드 가맹점 업종 분석을 통해 코로나 이후에 가맹점 개업 수가

가파르게 증가하였으며 상위 개업 업종은 식당 및 카페/편의점/ 화물 운송 순으로 파악됨



< 그림 4 > 목표시장 진입 관련 분석 수행 결과

## ○ 서비스 개발 관련 컨설팅

- 20대의 대출 현황을 파악한 결과, 대출 건수는 많지만, 평균 대출 금액이 낮은 것으로 보아 소액 대출을 받는 것으로 파악되며 신용 등급이 5~7등급에 위치한 고객을 주 잠재고객으로 예상

## □ 주요 성과

### ○ 타겟 도출에 따른 마이크로 마케팅을 통해 이용자 유입 확대

- 구체화한 타겟이 밀집한 채널을 대상으로 마이크로 마케팅 진행 및 소셜 데이터 분석을 통해 도출한 초단기 일자리 연관키워드를 마케팅에 활용 예정

### ○ 초단기 일자리 매칭을 제공하는 어플 출시

- 내년 상반기, 어플 출시를 앞두고 있으며 해당 플랫폼을 통해 일용직 근로자, 플랫폼 노동자, 프리랜서 등 앱 내 활동기록으로 경력과 능력을 증명이 가능한 기능을 포함할 예정
- 해당 앱의 기능을 통해 근무 경력과 활동 수준을 보장하며, 단기 노동도 하나의 직업군으로 자리 잡아 사회적인 인식 변화를 기대

# AI 객체 인식 자동 모니터링 기술과 초 단위 시청률 융합, 간접광고 효과 분석 서비스 개발 (세븐스타커뮤니케이션)

## □ 과제 목표

- AI 객체 인식 자동 모니터링 기술
  - AI 모니터링 기술을 통해 정확한 노출 위치 및 크기를 측정함으로써 수작업 모니터링으로 인한 위치값 오차 발생 문제점 해결
- 초 단위 시청률 융합, 간접광고(PPL) 효과분석 서비스 개발
  - 짧은 시간 노출되는 간접광고(PPL)의 특성을 고려하여 기존 분 단위 시청률 간접광고(PPL) 효과분석의 문제점 해결

## □ 분석주제

- AI 로고 인식 자동 모니터링 모델 개발
  - 수요기관에서 선정한 브랜드 로고를 대상으로 객체 인식 기술을 통해 화면 내 로고를 인식하여 정확한 로고 위치 정보를 제공하는 모델 개발
- 초 단위 시청률 융합을 통한 간접광고(PPL) 효과분석 보고서 개발
  - AI 로고인식 모델과 금융 빅데이터 플랫폼 초 단위 시청률 데이터를 융합하여 실효성 있는 간접광고(PPL) 효과분석을 위한 보고서 개발

### <활용 데이터 현황>

순번	수집채널	데이터명	기간
1	금융빅데이터플랫폼	간접광고(PPL) 효과분석을 위한 시청률 데이터	2021.05. ~ 2021.12.
2	내부 데이터	시청률 데이터	2020.01. ~ 2022.04.
3		광고량 데이터	2020.01. ~ 2022.04.

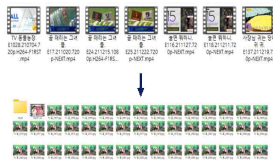
## □ 분석 결과

- AI 로고 인식 자동 모니터링 모델 개발
  - 5개의 브랜드 로고\* 학습데이터로 모델 학습 및 성능 검증을 통한

## 최종 모델 선정

\*아이시스 8.0, 하림펫푸드 더리얼, 아디다스, 새마을금고, 코카콜라

- 개인용 PC에서 운영할 수 있도록 AI 모델 경량화 및 최적화



[[1.079, 311.5997314455, 201.8512725830078, 1262.2225189208964, 593.3440399169922, 0.9814618229566028, 'Harim\_petfood\_thereal']]

그림 5 test sample



Confidence score = 98%

그림 6 test sample 결과

- 초 단위 시청률 융합을 통한 간접광고(PPL) 효과분석 보고서 개발
- AI 로고 인식 자동 모니터링 모델을 통한 로고 위치 정보와 초 단위 시청률 데이터를 기반으로 컨설팅을 통해 간접광고(PPL) 효과 분석 보고서 개발

## PPL 효과분석

3,300 %GPRs (누적시청률) 목표

### ▶ 시청률 대비 광고기대효과 분석

구분	전체 발달률수	광고시청률	GPR(%) (발달률 대비)	간접광고비	광고발달률 (발달률 대비)	GPP
드라마 60분대 (광고 30초 포함)	150회(개)	22%	3,300	800,000,000원	4,000,000	181,818
TV-DM 15" (광고 30초 포함)	150회	6.75%	1,012	1,532,260,000원	10,215,000	1,514,081

[주] 1) 광고시청률은 본 프로그램 시청률의 1수준

2) GPP(Cost per rating Point)=간접광고비/GPP

3) 드라마상에서 노출시간은 CF 15"보다 길게 나타났을 경우 고려하지 않음

4) 예고 SPOT의 C.I.노출

## 기존 간접광고(PPL) 효과분석 보고

- 모니터링 요원들의 육안 기반
- 프로그램 평균 시청률 기반
- 도표 기반의 분석 결과 제시

## 간접광고(PPL) 효과분석 보고서 sample

- AI 모델을 통해 객관적 위치 정보 기반
- 브랜드 로고 노출시 시청률 기반
- 다양한 시각화 기반의 분석 결과 제시

## □ 주요 성과

- 간접광고(PPL) 관련 업계 시장규모 확대 기대
- 정확한 간접광고(PPL) 효과분석을 통한 광고주 및 광고대행사 계약 단가 산정의 객관적 기반 제공

# 강원도 안심안전관광을 위한 블랙박스반사경 설치 최적지 분석 (시프트미러)

## □ 과제 목표

- 강원지역 등산로, 자연휴양림 등 산림지역 범죄 예방 환경조성
- 산림지역과 방범 시설물 미흡 지역, 구조·구급·화재 출동이 빈번한 지역을 분석하여 스마트 반사경 설치 최적지 도출

## □ 분석주제

- 강원도 내 방범 시설물 설치 미흡 지역 분석
  - CCTV, 보안등, 안전비상벨 데이터를 활용한 기술통계분석
- 강원도 내 구조·구급·화재 출동 지역 분석
  - 구조·구급·화재 출동 데이터를 활용한 기술통계분석
- 감시취약지수, 격자가중치모델 통한 스마트 반사경 최적지 분석

### <활용 데이터 현황>

순번	수집채널	데이터명	기간
1	소방안전 빅데이터 플랫폼 (강원도청)	격자별 교통사고·구급출동 데이터	2020.01. ~ 2020.12.
2		격자별 건축물·구급출동 데이터	2020.01. ~ 2020.12.
3		격자별 인구·구급출동 데이터	2020.01. ~ 2020.12.
4		일자별 기상·구급출동 데이터	2020.01. ~ 2020.12.
5		일자별 질병·구급출동 데이터	2020.01. ~ 2020.12.
6		구조 종별 신고·재난 데이터	2020.01. ~ 2020.12.
7		구급 종별 신고·재난 데이터	2020.01. ~ 2020.12.
8		화재 종별 신고·재난 데이터	2020.01. ~ 2020.12.
9		차량 화재 종별 신고·재난 데이터	2020.01. ~ 2020.12.
10	산림 빅데이터 플랫폼	전국 국립 자연휴양림별 이용정보	2020.01. ~ 2020.12.
11		전국 국립 자연휴양림별 생태계 정보	2020.01. ~ 2020.12.
12	외부 데이터	공공데이터	2020.12. ~ 2022.04.

## □ 분석 결과

- 강원도 내 방범 시설물 설치 미흡 지역 분석
  - CCTV설치 위치 도심 편중, 등산로/산야지역 내 보안등 부족, 안전비상벨 설치 수 지역별 편차 큼
- 강원도 내 구조·구급·화재 출동 지역 분석
  - 탈수/실족추락 긴급 구조, 산/바다 등 특정 주소가 없는 지역 구급 신고, 화재신고 등의 발생 시 빠른 출동을 위한 시설물 필요
- 감시취약지수, 격자가중치모델 통한 스마트 반사경 최적지 분석
  - 분석단위 면적 : 500셀(500\*500), 분석 대상지 : 동해시와 원주시
  - 각 격자에 존재하는 CCTV 대수를 이용하여 감시취약지수 산출
  - 변수와 표준화된 가중치를 통해 안심안전지수 도출 및 격자 가중치 모델로 스마트 반사경 최적지 각각 상위 10개 셀 도출



## □ 주요 성과

- 스마트 반사경 5대 설치 예정
  - 강원도 내 스마트 반사경 최적지에 설치할 스마트 반사경 5대 계약 완료로 2023년 설치 및 효과검증 예정

### 강원도청 소방안전 빅데이터 활용 고도화 계획

#### □ 배경

- 데이터 수요기관 및 활용방안 발굴위해 소방안전 빅데이터와 외부 데이터 융합 필요
- 데이터 융합 모델, 활용 효과 검증으로 데이터 활용 확대 추진

#### □ 추진계획

- 분석대상 : 산불예방 및 소방안전 안전시설물 최적지 도출(강원)
- 내용 :
  - 방범 시설물(CCTV, 가로등, 방화벽 등) 중 구조, 구급, 방범 지역 분석
  - 소방안전(화재, 유출사고, 시설장애)에 대한 안전시설물 분석 최적지 도출
  - 강원소방 빅데이터 센터 컨소시엄 '빅데이터 분석 플랫폼'에 연계 추진



※데이터 : 원주시, 동해시 감시 카메라 지점 시각화 및 안전시설 최적지 분석

○ 활용 방안(예시)

- 스마트 반사경 : 조명, 센서, CCTV, 디스플레이를 결합한 스마트 반사경
- 방범 효과 : 방범 시설물 중 소방안전 안전시설물 최적지 도출, CCTV, 가로등, 조명 등

(좌) 방범 효과 분석 (우) 효과 검증

(가) 방범 효과 분석 (나) 효과 검증

○ 기대 효과(향후계획)

- 소방안전 : 방범효과를 증진한 스마트 반사경 설치로 산악 안전시설물 촉진
- 소방안전 빅데이터 분석과 예비화재기법 상생 모델 구현
- 강원도내 선의 운영 효과 분석 및 지자체 확산 협의 추진

### 커스터마이징 제품 예시



# 기업 데이터 기반 포지셔닝 진단시스템 구축 (와우커뮤니케이션)

## □ 과제 목표

- 기업 데이터 기반 포지셔닝 진단시스템 구축
  - 정부지원사업 신청대상 요건을 충족하는 기업정보 데이터 분석
  - 정부지원사업에 선정(낙찰)된 기업의 주요 특성 분석
  - 기업 포지셔닝 진단 AI 모델 알고리즘 제안

## □ 분석 주제

- 기업정보 데이터 분석
  - 데이터 분석이 가능한 형태로 병합·가공 후 기업정보 분석
  - 기업규모, 기업형태, 휴폐업여부, 각종 인증여부 등 다차원 분석
- 낙찰기업 및 대조군 비교 및 낙찰기업 주요 특성 도출
  - 지원사업 선정 낙찰기업과 대조군 비교 분석
  - 낙찰기업의 주요 특성 도출
- 기업 포지셔닝 진단 AI 모델 알고리즘 제안
  - 정량적·정성적 포지셔닝 평가지표 설계 및 산출
  - 랜덤포레스트, XGBoost, Ridge 등 ML모델 활용결과 성능비교 분석

### <활용 데이터 현황>

□ 지역경제빅데이터플랫폼 구매데이터 및 외부 데이터 활용 (시간적범위: 2021년)

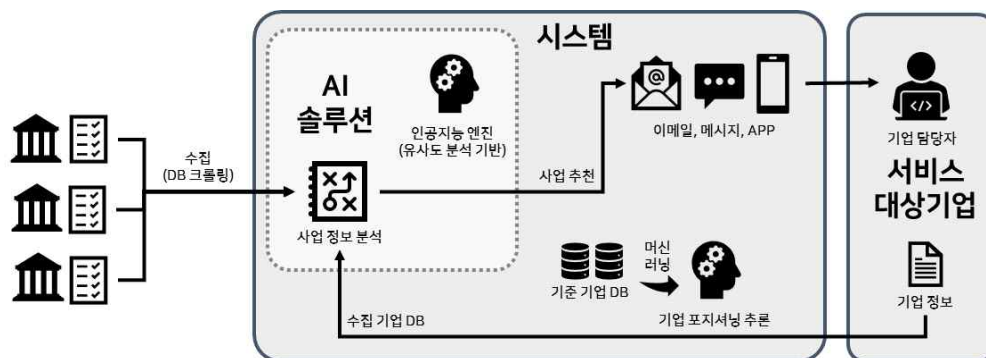
순번	수집채널	데이터명	기간
1	지역경제빅데이터플랫폼 (한국평가데이터)	기업현황 DB	2016.01. ~ 2021.12.
2		주요재무현황 DB	2019.01. ~ 2021.12.
3		신용등급 DB	2021.01. ~ 2021.12.
4	외부 데이터	신기술(NET) 인증DB	2021.01 ~ 2021.12.
5		신제품(NEP) 인증DB	2021.01 ~ 2021.12.
6		나라장터 입찰DB	2021.01 ~ 2021.12.
7		나라장터 낙찰DB	2021.01 ~ 2021.12.

## □ 분석 결과

- 기업정보 데이터 분석
  - 재무항목 및 재무비율 상대변화, 기업규모, 기업형태, 휴폐업여부, 각종 인증보유여부 등 여러 기준에서 다차원 분석
- 낙찰기업 및 대조군 비교 및 낙찰기업 주요 특성 도출
  - EDA 분석결과 매출액, 영업이익, 자산총계 등의 재무정보가 좋은 기업일수록 기업 낙찰과 상관관계가 높으며, 입찰기업의 신용등급이 실제 낙찰여부에 영향을 미치는 상관관계는 낮음
- 기업 포지셔닝 진단 AI 모델 알고리즘 제안
  - 포지셔닝 분류설계 및 정량적(수익성, 성장성), 정성적(구체성, 실현가능성, 확장성) 평가지표 산출
  - 특허/실용신안권 등의 기술정보, 각종 인증정보 데이터를 포함하여 학습데이터 및 검증데이터 생성하여 10개 이상의 변수생성
  - 총 4개의 머신러닝 기법 활용 RMSE값 성능테스트 및 결과 분석

## □ 주요 성과

- 기업 포지셔닝 진단서비스 초기 MVP모델 출시를 위한 AI모델 제안
  - 랜덤포레스트 XGBoost, Ridge회귀, Lasso회귀 총 4개 머신러닝 모델을 활용한 성능비교 및 RMSE값 산출, 우수모델 제안



- 지속적인 기업데이터 제공을 위한 활용 컨설팅
  - 추후 기업 포지셔닝 진단서비스 개발을 고려하여 지속적인 기업 정보가 연계될 수 있도록 API 방식 송수신단 개발 및 DB연계

# 서울생활권 아파트 엘리베이터TV 입주민 특성 고도화 상품 [포커스미디어코리아]

## □ 과제 목표

- 엘리베이터TV 설치 단지 입주민 특성 기반 소비 클러스터 도출

## □ 분석주제

- 분석이 가능한 형태로 데이터 병합 및 가공
  - 금융빅데이터플랫폼에서 구매한 온라인상세매출정보와 수요기관 내부데이터를 분석이 가능한 형태로 병합 및 가공
- 엘리베이터TV 설치단지별 입주민 소비 특성 분석
  - 아파트 구조와 입주민 인구 통계, 입주민 생애주기 등을 고려해 입주민의 소비 특성에 따른 군집 도출
- 설치단지별 미디어 송출 적합 업종 스코어 도출
  - 1개의 아파트 단지와 N개의 소비 특성 군집 간의 유사도 스코어 도출 필요

### <활용 데이터 현황>

순번	수집채널	데이터명	기간
1	금융빅데이터플랫폼 (비씨카드)	온라인상세매출데이터	2021. 06. ~ 2022. 03.
2	외부 데이터	해당 없음	-
3	내부데이터	아파트 구조DB	2021. 06. ~ 2022. 03.
4		인구 통계DB	2021. 06. ~ 2022. 03.
5		생애주기DB	2021. 06. ~ 2022. 03.
6		라이프스타일DB	2021. 06. ~ 2022. 03.
7		단지조사DB	2021. 06. ~ 2022. 03.

## □ 분석 결과

- 금융 빅데이터 플랫폼의 온라인상세매출정보를 내부 데이터와 병합함으로써 입주인 인구 특성, 입주인 생애주기 특성, 아파트 단지 내 인프라 등의 정보를 군집 분석에 활용할 수 있게 됨
- K-means를 활용해 3,501개 단지를 9개의 군집으로 분류하여 각 군집의 입주인 인구 특성, 입주인 생애주기 특성, 아파트 단지 내 인프라 특성을 파악해 군집 간 차별성 파악
- 1개 아파트 단지와 9개 군집의 중심값 간의 거리를 산정하여 소비 유형이 유사한 군집 추천



<그림 14> 군집별 입주인 소비 특성 도출

## □ 주요 성과

- 서울생활권 클러스터 상품 개발
  - 군집 비교 후 카테고리의 소비 차별성을 서울생활권 클러스터에 반영하여 상품을 고도화해 파트너사에게는 주목도 향상을, 입주민에게는 맞춤형 콘텐츠를 제공
- 신제품 출시를 통한 매출 증가
  - 엘리베이터TV 콘텐츠 송출 영업 전략이 다양해짐에 따라 상품 단가 인상 및 협업 가능 파트너사 범위 확대에 의한 매출액 증가

**유통소비·산림·중소기업·농식품**  
**[6개 과제]**

# 인공지능 운동 재활 디지털 치료제 서비스 개발 [넥스트도어]

## □ 과제 목표

- 이미지 데이터를 활용한 신체 24개 관절에 대한 Annotation
- 인공지능 운동(모션) 학습(테스트) 데이터셋 구축 컨설팅

## □ 분석주제

- 이미지 Annotation
  - 구매 데이터(영상)에서 이미지 추출 및 분류, 수요기업의 스튜디오 촬영 이미지 활용
  - 신체(전신) 24개 포인트를 규정하여 Annotation 수행
- 컨설팅
  - 3D 메쉬 데이터 가공 관련 컨설팅
  - 학습용(테스트) 데이터셋 구축 지원

### <활용 데이터 현황>

구분	데이터명	데이터 유형	데이터 수(개)
구매데이터	국대는 국대다	MBN '국대는 국대다' 방송VOD 영상	9
	매경골프	MBN '골프 포 위민' 방송(레슨) VOD 영상	51
	병아리 하이킥	MBN '병아리 하이킥' 방송VOD 영상	12
	백 투 더 그라운드	MBN '백 투 더 그라운드' 방송VOD 영상	16
	엄지의 제왕	MBN '엄지의 제왕' 방송VOD 영상	22
제공데이터	재활데이터(이미지)	수요기업(넥스트도어) 스튜디오 촬영 이미지	49,993

## □ 분석 결과

- 구매데이터(100개 VOD 영상) 이미지 추출
  - 구매한 5개 방송VOD에서 Frame 기준 10만장의 이미지 추출 후 Annotation에 활용(약 1만장)
  - 부족한 Annotation 작업 수량을 늘리기 위해 수요기업 스튜디오에서 촬영한 이미지(약 5만장)를 추가로 활용

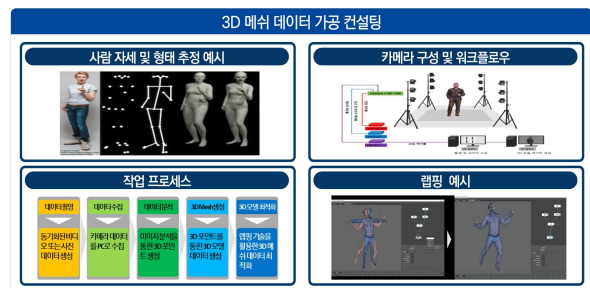
## ○ 이미지 Annotation 수행

- 신체 전신 24개 포인트를 규정(수요기업 요청사항 반영)
- 결과물에 대해 정의 테이블과 함께 수요기업에 제공



## ○ 컨설팅

- 기존에 수요기업에서 보유한 3D 메쉬데이터 가공 관련 기술 고도화
- Annotation 결과물을 활용한 학습용 테스트 데이터 셋 구축 지원



## □ 주요 성과

- 구축된 데이터셋 및 컨설팅 내용을 바탕으로 수요기업의 관련 산업기술 고도화
- 재활서비스 제공 고도화 기초 마련
  - 비대면 서비스로 시간적, 공간적 제약 없이 다양한 재활서비스 제공
  - 사용자 요구에 맞춘 다양하고 신속한 재활서비스 제공 및 대응
  - 디지털 재활서비스 제공 분야의 산업 활성화

# 구리시 상권 활성화 성과 분석 서비스 지원사업 [구리시상권활성화재단]

## □ 과제 목표

- 상권 활성화 테마거리별 유동인구, 매출 데이터 분석을 통한 사업 지원 성과분석
- 성과분석을 바탕으로 사업지원 정책 수립에 활용할 수 있도록 지원

## □ 분석주제

- 상권활성화 관리구역 내 현황분석(업종별, 폐업, 교통 등)
- 상권활성화 관리구역 내 트렌드 분석
  - 유동인구 데이터 현황분석
  - 카드 소비데이터 현황분석

### <활용 데이터 현황>

구분	데이터명	데이터 정보	범위	형식
구매 데이터	지역별, 업종별 카드 소비데이터	상권활성화 관리구역(6개) 내 발생 카드소비 데이터	2021.08 ~ 2022.07	xls
	유동인구 데이터	상권활성화 관리구역 내 일별 유동인구 데이터 (50m×50m, 10세 단위)	2021.08 ~ 2022.07	xls
수집 데이터	상가활성화 업종 상가정보	분석을 위한 업종 구분 정보	2021.10	xls
	주차장 데이터	상권활성화 관리구역 인근 공영주차장 이용정보	2021.08 ~ 2022.07	xls
	상가정보 데이터	상권활성화 관리구역 내 상가등록 정보	2022.06	csv
	건물데이터	상권활성화 관리구역 내 건물정보 데이터	2022.06	shp
	버스정류장	경기도 관내 버스정류장 정보	2022.10	xls
	인허가 데이터	상권활성화 관리구역 내 상가 인허가 정보 데이터	2022.08	csv
	현장조사 데이터	상호명(현장조사를 통해 상권활성화 관리구역 내 모든 상호명 조사)	2022.10	xls

## □ 분석 결과

### ○ 유동인구 분석

- 여자(43.1%) 보다 남자(56.9%) 비율이 높으며, 60대 이상이 43.2%를 차지
- 토요일 유동인구 16.1%, 저녁(16 ~ 21시) 시간대 38.3%를 차지
- 30분 이상 비정기방문인구는 남자(49.4%)보다 여자(50.7%) 비율이 높음
- 관외 인구(49.8%) 보다 관내 인구(50.3%)가 많음
- 관내 및 관외 60대 이상 인구가 많았으며, 관내는 저녁(41.5%) 시간대, 관외는 오후(37.9%)와 저녁(37.5%) 시간대가 많음
- 60대 이상 인구가 높게 나타난 이유는 인구 구조적 특징 때문으로 판단됨

### ○ 카드소비 분석

- 사용건수는 남자(51.1%)가 많지만, 사용금액은 여자(53.5%) 높으며, 주로 20~34세 연령층(사용건수 39.2%, 사용금액 33.4%)
- 저녁 시간대(16~21시)에 사용건수 47.3%, 사용금액 45.4%를 차지
- 남자는 꽃길문화거리, 여자는 전통시장저잣거리 사용 비율이 높으며, 사용금액은 남자 꽃길문화거리, 여자 낭만청춘거리비율 높음
- 낭만청춘거리, 생활저잣거리 그리고 선술집포차거리는 사용건수와 사용금액 업종이 서로 상이
- 관내는 사용건수와 사용금액에서 음/식료품/마트/편의점의 비중이 높으며, 관외의 경우 사용금액은 의류 및 패션잡화에서 사용금액 비율이 높음

## □ 주요 성과

- 분석 및 시각화 정보를 통한 사업지원 효과 검증
- 상권활성화 구역 내 상가업종에 대한 DB구축
- 실질적인 상권활성화 지원 정책 마련을 위한 기초자료 활용

# 회계사/소상공인 대상 재무관리 업무 혁신 서비스 [내일을향해]

## □ 과제 목표

- 회계사 업무 자동화 및 효율화
- 중소기업 및 소상공인 창업 및 재무 리스크 관리
- 맞춤형 지표 제공을 통한 의사결정 효율화

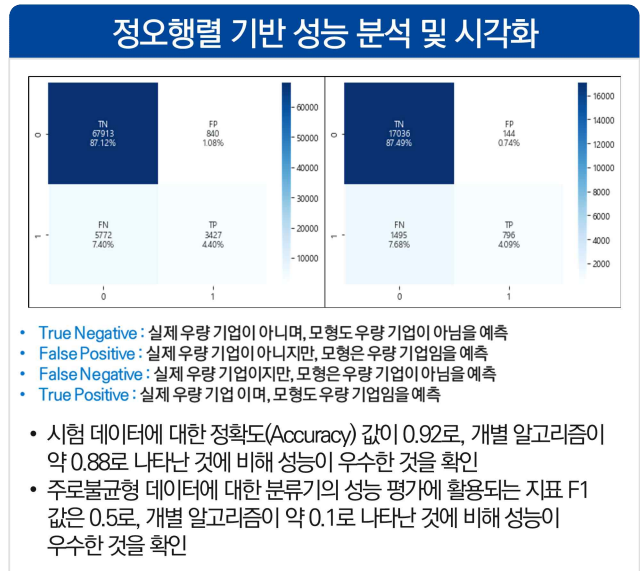
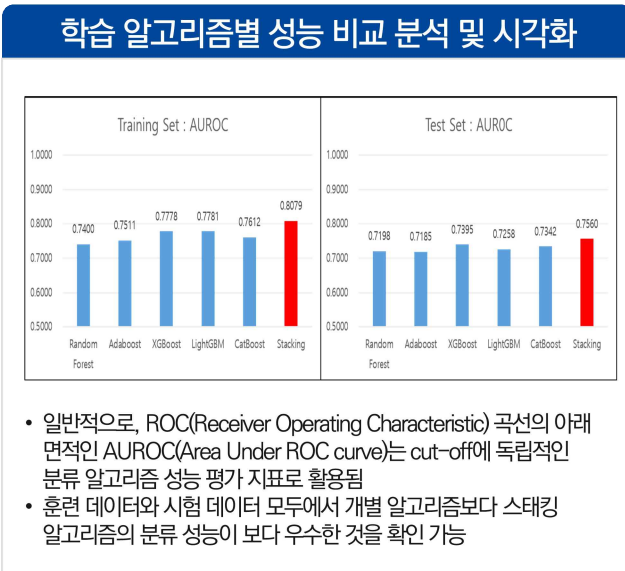
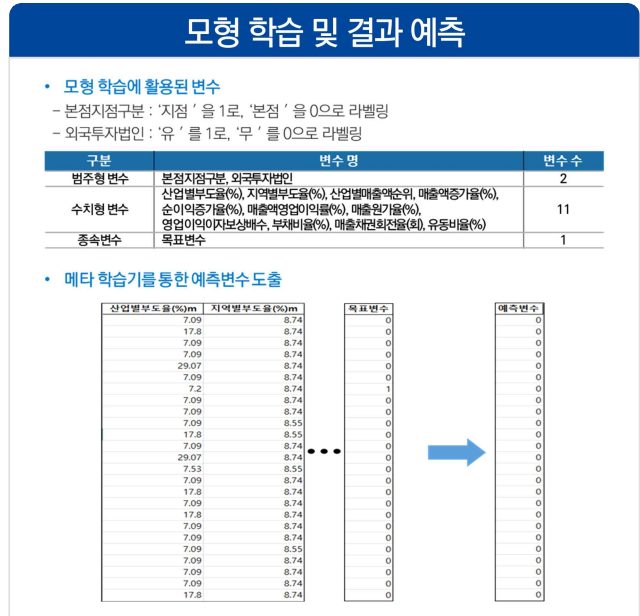
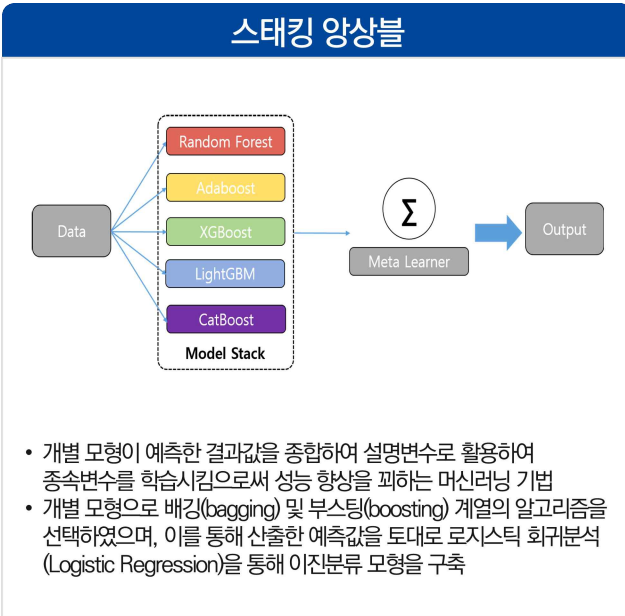
## □ 분석주제

- 타겟 기업(사업자번호) 기준 Data Mart 생성
  - 기업개요 기준 테이블 조인
  - 주요파생변수 생성을 위한 전처리
  - 업무 효율화를 위한 파생변수 생성
- 우량기업 선별을 위한 분류모형 구축
  - 학습 및 평가 데이터 구축
  - 통계 기반(회귀모형) 모형 구축
  - 머신러닝 및 딥러닝 분류모형 구축
  - 최종 분류모형 선정
  - 평가 결과 산출 및 최종 산출물 생성

### <활용 데이터 현황>

구분	데이터명	데이터 정보	비고
구매 데이터	기업개요	사업자번호, 기업규모, 산업코드 등 국내 기업정보	일부 칼럼(기업규모, 산업코드 등) 기본적인 클렌징 가공
	휴/폐업 정보	휴폐업 상태와 상태 변경에 대한 일자 정보	분석 기준일에 따라 현재 또는 과거 휴/폐업 여부 칼럼 추가
	채무불이행 정보	기업 연체 발생일, 연체금액 등 신용불량정보	채무불이행 여부 및 일자 정보
	재무정보	사업자번호 기준의 주요 재무정보	데이터 건수에 따라 협의 후 2개년 또는 3개년도 결산 재무 정보 추가
	국내 기업 신용등급	재무분석, 모형 산출, 조사로 산출되는 신용등급	기업번호(PK), 재무 결산 연도에 따른 신용등급 활용
수집 데이터	폐업자 정보	국내 폐업기업의 사유별통계 정보	

## □ 분석 결과



## □ 주요 성과

- 회계 업무 효율성 증대
  - 업무 비효율 개선
  - 회계 투명성 및 기업 평가 적정성 확보
- 소상공인 업무 및 창업 생태계 효율성 증대
  - 소상공인 의사결정 비용 절감
  - 중소기업/소상공인 창업 및 재무관리에 효율적

# 온실가스 상쇄사업 참여확대를 위한 데이터기반 분석서비스 [에이아이블루]

## □ 과제 목표

- 딥러닝을 활용 항공영상 이미지의 수종분류 모델 구축 지원
- 수종분류 AI 모델의 정확도 향상을 위한 컨설팅

## □ 분석주제

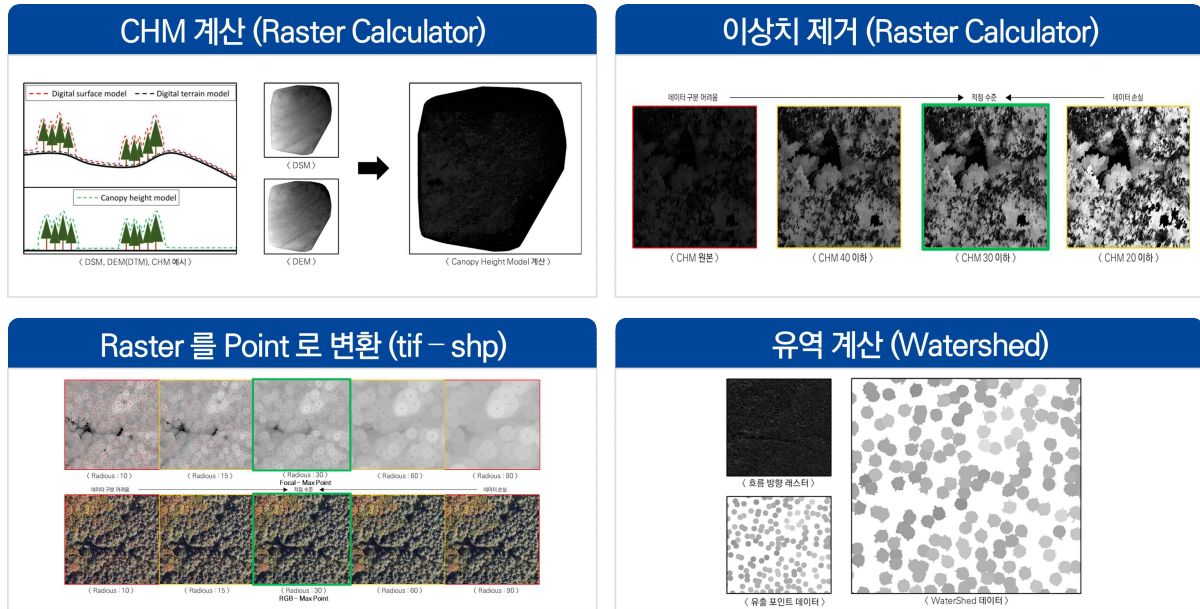
- 구매데이터 탐색적 자료 분석 (EDA) 진행
- 수종분류 모델구축 컨설팅
  - CNN 영상 판독 기술 관련 자료 수집, 공유
  - 드론 영상 전처리·가공, 학습 데이터셋 구축 지원, AI모델 활용방안

### <활용 데이터 현황>

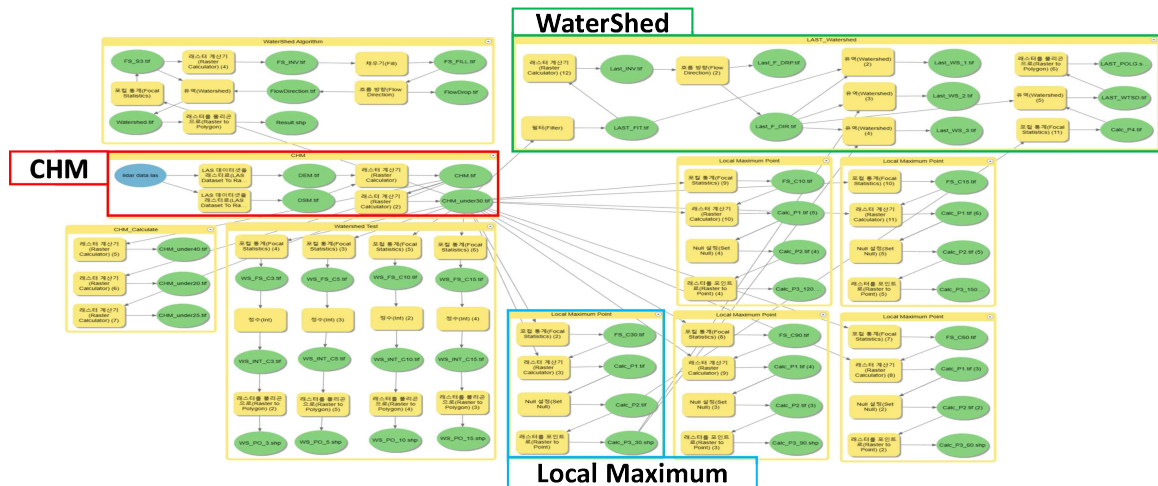
구분	보유기관(출처)	데이터명	데이터 정보	범위
구매 데이터	KDX (우림엔알)	산림탄소흡수정보	지상, 지하, 고사목, 토양 등 어노테이션데이터	설악산, 덕유 산, 무등산
	KDX (시선아이티)	탄소저장량지도_ 시도_전국	필지별로 산림 수종별탄소흡수량의등급정보 (1~5등급 표현, 낮은 등급일수록 좋은 등급)	전국
	KDX (데이터 사이언스랩)	격자별건축물.구 급출동 데이터	위치정보, 시군구코드, 건축물 용도, 사고유형	강원도
		격자별인구.구급 출동 데이터	위치정보, 발생시간, 성별 · 나이 분류, 사고 유형	강원도
		일자별기상구급 출동데이터	지점정보, 측정시간, 기상정보, 사고발생건수	강원도
		일자별질병 구급 출동 데이터	발생일자, 행정동 코드, 사고유형 등	강원도
		구조 종별 신고 재난 데이터	신고시간, 출동시간, 도착시간, 사고원인 등	강원도
		구급 종별 신고 재난 데이터	시간, 환자유형, 소방서, 위치정보, 병원명등	강원도
		화재 종별 신고 재난 데이터	접수시간, 상황종료시간, 위치정보 등	강원도
		차량화재 종별 신고재난 데이터	접수시간, 신고자명 및 전화번호, 사고위치 등	강원도
수집 데이터	자체수급	드론촬영 영상데 이터	정사영상, L1 Lidar data, 20ha 면적	서울대학교 태화산학술림

## □ 분석 결과

### ○ 드론촬영 데이터 전처리



### ○ 복잡한 전처리 과정 자동화(워크플로우 생성)



## □ 주요 성과

- AI 모델을 활용한 수종 분류 모델 고도화를 위한 기반 마련
- 추후 다양한 연구 및 모델학습 고도화를 통한 수종 분류 모델을 활용한 온실가스 상쇄율 분석 및 수익성 정보제공을 통한 사업참여 확대

# 중소기업 빅데이터 플랫폼을 통한 수출통합지원 플랫폼 구축 (피알마케팅)

## □ 과제 목표

- 기업정보 데이터를 활용한 기업 현황 및 산업 분석을 통한 산업 동향 정보제공
- 회원기업을 대상으로 지원사업 정보와의 매핑을 통해 정보 및 지원사업 기회 제공

## □ 분석주제

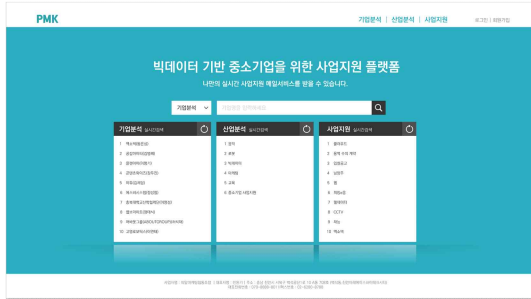
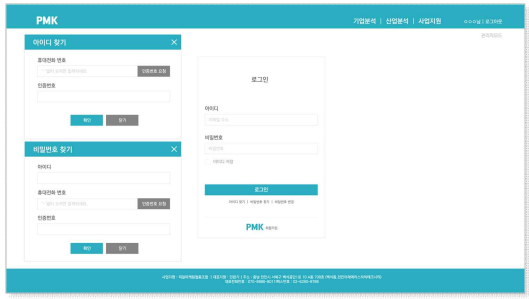

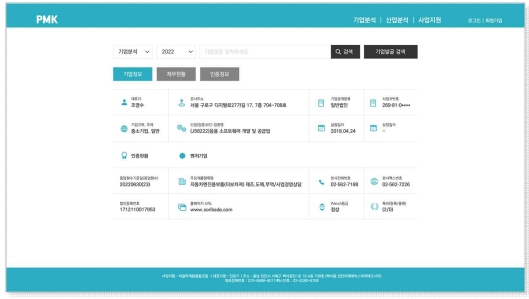
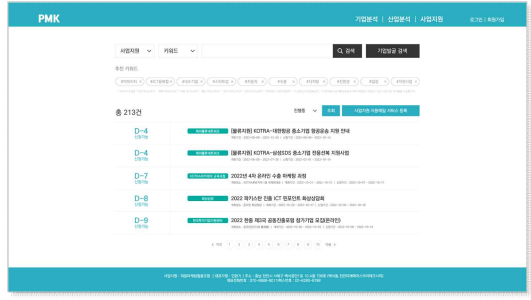
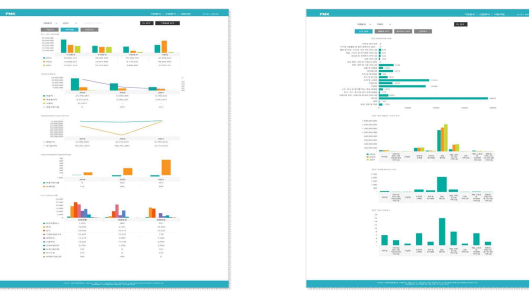
- 기업발굴, 산업분석 서비스 제공
  - 수요기업 구축 플랫폼 내에 서비스 제공 페이지 구축 지원
- 사업지원 서비스 제공
  - 문서분석(웹크롤링, NLP 등)을 통해 사업지원 공고 필수 참여 요건과 회원 기업의 보유 요건 매칭을 통한 정보제공 서비스 구축 지원

### <활용 데이터 현황>

구분	보유기관 (출처)	데이터명	데이터 정보	비고
구매 데이터	KDX (나이스디앤비)	기업정보 _마스터개요	10차 산업분류코드 기준 수요기업 요청 기업 데이터 제공(민감정보 제외)	xlsx
		기업정보 _개별재무GAAP	기업재무 DB(자산총계, 부채총계, 자본금, 자본 총계, 매출액, 수출액, 영업이익 등)	xlsx
		기업정보 _연결재무GAAP	연결재무상태표, 연결(포괄)손익계산서, 연결자 본변동표, 연결현금 흐름표등	xlsx
		기업정보 _인증정보	경영혁신(메인비즈), 기술혁신(이노비즈), 벤처 인증, 노색인증, 사회적기업 등 인증정보	xlsx

## □ 분석 결과

- 일반적인 분석서비스 지원에서 데이터를 활용한 서비스 플랫폼 구축 지원으로 과제 수행
- 기존에 수요기업에서 진행중인 플랫폼 구축 사업과 연계하려 하였으나 개발환경, 기간 등의 문제로 별도의 페이지를 구축 지원 하는 것으로 과제 요건 정리
- DB설계부터 화면 구축 등 사업화를 위한 플랫폼 구축 각 단계별 지원 수행

메인페이지	로그인 페이지
	
관리자 페이지	기업발굴 검색 페이지
	
사업지원 페이지	기업분석, 산업분석
	

## □ 주요 성과

- 기업정보 및 수집정보 매핑 작업을 통한 조건에 맞는 회원기업에 수집정보를 제공함으로써 지원사업참여 기회 확대

# 인공지능 기반 레시피 및 외식/배달메뉴의 칼로리, 영양성분정보 제공 서비스 구축 (더에이치랩)

## □ 과제 목표

- 레시피 재료명, 단위, 중량(수치) 표준화를 통한 자동분류 체계 구축
- 표준화된 정보와 영양성분정보를 결합하여 칼로리, 체중조절, 질병식 등 영양관련 전문정보 서비스 제공 지원

## □ 분석주제

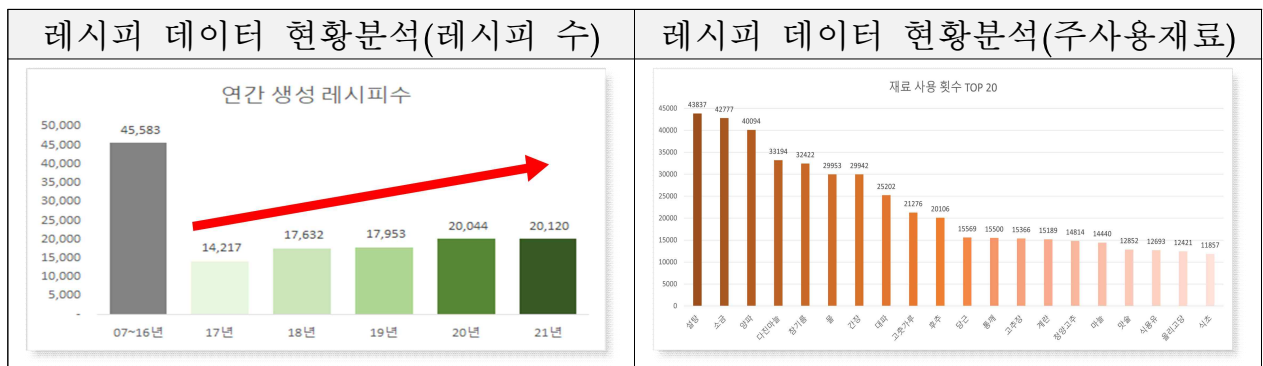
- 수집 데이터(레시피) 현황 분석(EDA)
- 레시피 기반 재료명 표준화 및 함량단위 환산 자동화 구축 지원

### <활용 데이터 현황>

구분	보유기관 (출처)	데이터명	데이터 정보	범위	형식
구매 데이터	KDX (만개의레시피)	레시피 데이터	요리명, 레시피, 재료내용, 인분수등	2007 ~2021(15년)	csv
제공 데이터	수요기업 (더에이치랩)	농축산물분류체계	재료명, 식품명, 식품_대분류 등	2022년기준	xlsx
		표준재료명 리스트	표준재료명, 식품명, 영양성분, 분류 등	2022년기준	xlsx
		재료 Master	Raw_Ingredients, Ingredients_std.	2022년기준	xlsx

## □ 분석 결과

- 레시피 데이터 현황 분석(EDA)



## ○ 자동화 프로세스 구축

### - 재료명/중량/단위 분류 및 표준화 자동화



## ○ 레시피 표준화 방안 컨설팅

- 농작물 도감 기반 작물 리스트 샘플 활용 제안
- 레시피 표준화에 필요한 데이터(재료) 리스트 제공
- 표준재료명 정의 가이드 제공

## ○ 영양성분 합산 시 테이블 활용 가이드 제시

- 재료정보품질 flag, 가식부flag 컬럼을추가하여 영양성분 합산 시 직관적인 필터링 가능
- 레시피상 필수 재료 여부 판별 방안 제시
- 재료 분류 컬럼 추가, 최소한의 기준으로 반영하도록 구현
- 인분 컬럼의 숫자 부분을 분리
- 1인분당 레시피의영양성분 연산이 가능하도록 가이드

## □ 주요 성과

- 표준화된 정보(재료명, 단위, 중량)에 따른 신규 레시피 분류 자동화 체계 구축
- 수요기업 사업화 지원
  - 구축된 자동분류 모델 활용 방안 컨설팅을 통해 수요기업에서 구축하고자 하는 앱의 기능 고도화

**라이프로그·교통·통신**  
**[6개 과제]**

# 난청환자 특징을 고려한 난청 자가진단 및 청각재활시스템 개발 [지엔히어링]

## □ 과제 목표

- 난청환자의 유형별, 연령별 및 연도별 유병률을 파악하고, 난청 환자의 의료행태 및 보청기 사용 행태 (보청기 사용률, 청각장애율)를 산정하여 구체적인 타겟층을 설정
- 보청기 착용 후 실질적인 보청기 효과를 예측할 수 있는 예측 모델을 개발하여, 보청기 신규 상담 및 재활의 보조도구로 활용

## □ 분석주제

- 건강보험공단 데이터와 청각 빅데이터 확보 및 통합분석
- 난청 분류에 대한 기준
- 난청 분류 및 보청기 착용에 따른 의료 패턴 분석
  - 난청 유형별 정보 (실인원, 분류별 현황, 유병률)
  - 장애진단 (분류별 장애진단 현황, 건강검진을 통한 장애진단)
  - 보청기 착용(보청기 사용률, 보청기 구매 기간)
- 보청기 착용 후 순음청력검사 예측 모델 개발

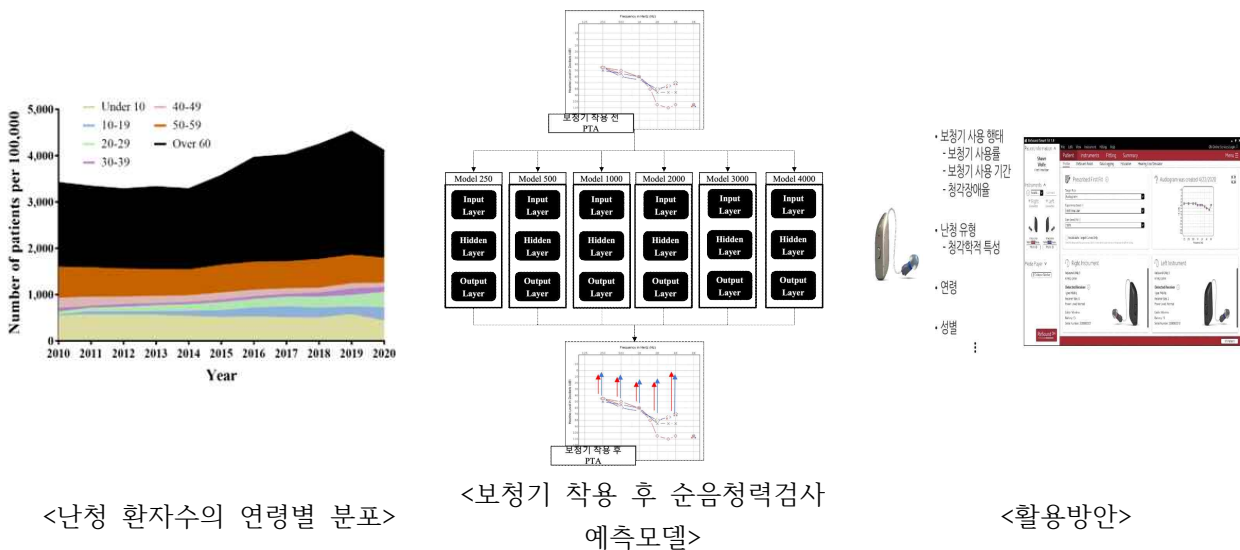
### <활용 데이터 현황>

- 라이프로그 플랫폼 데이터 (기도청각검사, 골도청각검사, 환자프로파일, 보청기 착용 후 순음청력검사, 청성뇌간반응역치, 청성뇌간반응 잠복기)
- 건강보험공단 난청 데이터 (전도성난청, 감각신경성난청, 돌발성난청, 이독성난청, 노인성난청, 소음성난청, 혼합성난청, 기타난청)

## □ 분석 결과

- 난청 유병률과 연평균 증가율
  - 2010~2020년간의 난청 유형별, 연령별 및 연도별 난청 유병율에 대한 산출
  - 2015년을 기점으로, 난청의 유병률과 보청기 처방인구, 청각 장애인구의 증가 추세간의 회귀분석 결과, 난청의 유병률에 유의미한 영향 확인

- 보청기 사용 패턴 및 의료소비행태 분석
  - 청년층의 높은 유병률과 대조적인 보청기 의료 행태 확인
  - 보청기 사용 패턴을 근거로, 보청기 사용의 취약층 및 취약예비층 산정하고, 이를 활용한 마케팅 전략 수립 완료
- 보청기 착용 후 순음청력검사 예측 모델 개발
  - 대상자 정보 및 순음청력검사 결과를 기반으로, 보청기를 착용하였을 시, 얼마나 청력의 향상이 있는지 예측하는 모델을 구현 (환자의 보청기 구매 보조를 하기 위한 모델)



## □ 주요 성과

- 전체적인 현 보청기 시장의 시장 조사, 난청 유병률과 각 의료 행태 등 보청기 시장에서 데이터를 활용한 분석결과를 도출하여 마케팅 기반 자료로 제시
- 보청기 사용의 취약층 / 미래고객을 산정하여 해당 그룹의 의료 행태에 기반한 보청기 판매/마케팅 전략 수립, 컨설팅 완료
- 보청기를 착용하지 않아도 예상되는 보청기 효과를 근거 기반으로 산출할 수 있으며, 보청기의 신규상담 및 적합, 재활의 보조 도구로써 활용성 제시 완료('23년 보청기 신규상담에 활용 예정)

# 골전도 보청기의 성능인증 및 표준구축을 통한 근거 기반의 임상가이드라인 구축 (히어라이프)

## □ 과제 목표

- 수요기업 히어라이프는 골전도 보청기 기업으로, 골전도 보청기의 국내 시장 안착 및 활성화를 목적으로 청각데이터의 확보 및 분석을 희망하였음
- 골전도 보청기의 중심대상자군의 선정을 토대로, 건강보험공단 난청데이터 및 청각빅데이터를 활용한 전체적인 시장조사, 골전도 보청기 환자의 의료비용 및 행태를 분석하고, 실제 임상 환자를 대상으로 골전도 보청기의 임상적 이점을 분석하고자 함

## □ 분석주제

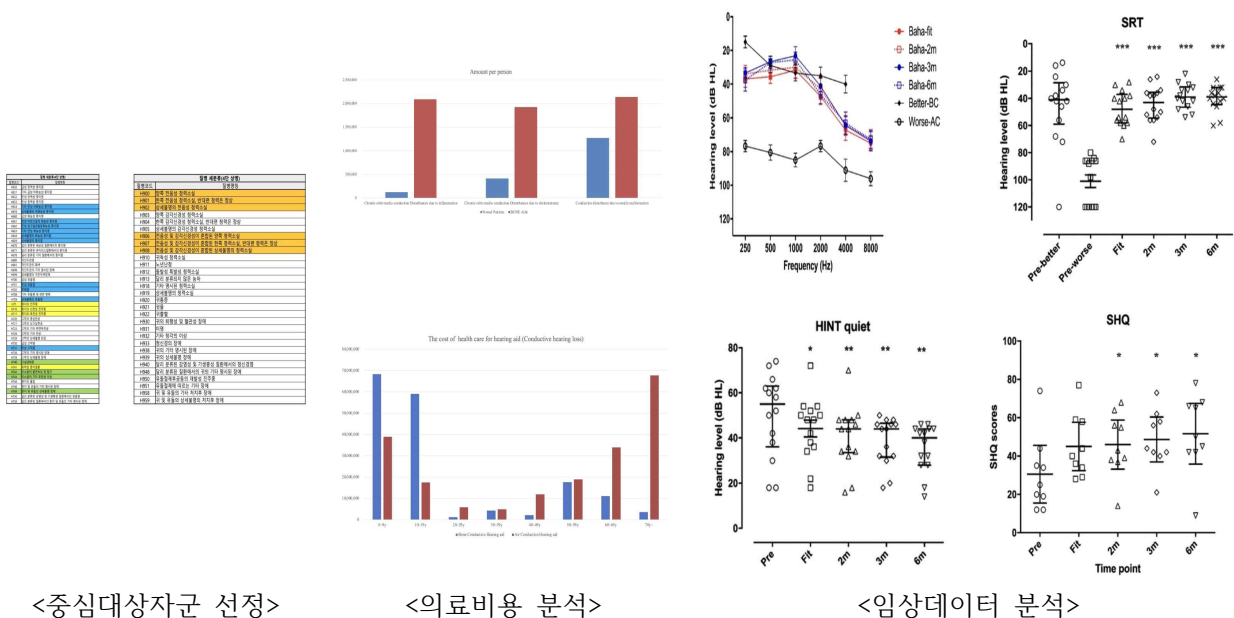
- 건강보험공단 및 청각빅데이터의 통합 분석
- 난청 분류에 대한 기준
  - 전음성, 일측성, 선청성 외이 기형의 3가지 환자군 설정
  - 조작적 정의 및 분석 진행
- 보청기 착용 여부, 임플란트 유무 등에 대한 빈도 분석
  - 각 그룹 환자의 의료비용, 환자 수 조사
- 임상 환자의 청각검사 데이터 분석
  - 실제 골전도 보청기 착용 환자의 임상 청각 검사 분석
  - 수술 전/후, 6개월 데이터를 기반으로 임상 청각 결과 분석

### <활용 데이터 현황>

- 라이프로그 플랫폼 데이터 (기도청각검사, 골도청각검사, 환자프로파일, 보청기 착용 후 순음청력검사, 청성뇌간반응역치, 청성뇌간반응 잠복기)
- 건강보험공단 난청 데이터 (전도성난청, 감각신경성난청, 돌발성난청, 이독성난청, 노인성 난청, 소음성난청, 혼합성난청, 기타난청)

## □ 분석 결과

- 중심 대상자군 선정
  - 건강보험공단 난청 데이터를 활용한 중심대상자군의 시장 조사
  - 중심환자군을 대상으로, 보험수가 산정 및 보조금 확대의 근거 수립을 통해, 골전도 보청기 활성화 전략 수립 완료
- 골전도 보청기의 의료비용 분석
  - 분류된 환자군별 치료효과 및 청각학적 이점 산출 필요
- 임상데이터 분석
  - 종합적으로, 골전도 보청기를 착용하였을때 모든 영역에서 유의미하게 청력이 개선됨을 확인하며, 이를 기반으로 골전도 보청기의 임상적 근거 수립 및 활용



## □ 주요 성과

- 근거 기반의 중심 대상자군의 산정과, 해당 대상자의 유의미한 개선점을 도출함. 이를 기반으로, 골전도 보청기의 수가 산정 및 보조금 확대의 기반 자료로 활용
- 또한, 실제 골전도 보청기 착용자의 결과를 기반으로, 골전도 보청기의 임상 가이드라인 구축 완료

# 빅데이터분석 신규반려동물 간식개발, 제조 및 제품 추천서비스(오션)

## □ 과제 목표

- SNS 데이터 기반 미래신호 분석 및 신제품 개발 컨설팅
- 카드 데이터 기반 반려동물 보유자 소비행태 분석을 통한 소비  
행태별 마케팅 방안 컨설팅

## □ 분석주제

- SNS 데이터 분석
  - SNS 데이터 분석으로부터 미래신호 추출, 신제품 키워드 추천
- 카드 데이터 분석
  - 카드 데이터 분석으로부터 소비행태를 파악하고 마케팅 방안 도출

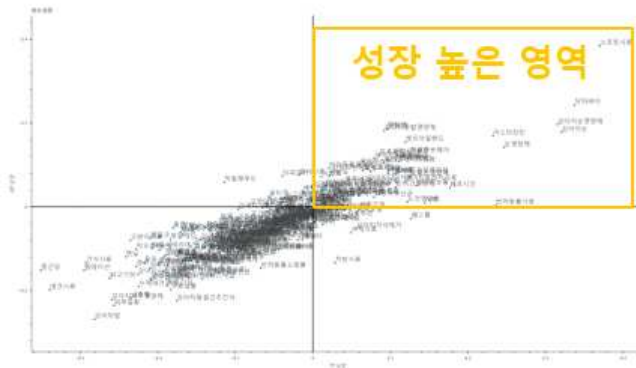
### <활용 데이터 현황>

- SNS데이터: 카페, 블로그 데이터 약 160만개 데이터 활용(12개월)
- 카드데이터: 삼성카드 데이터(32개월)

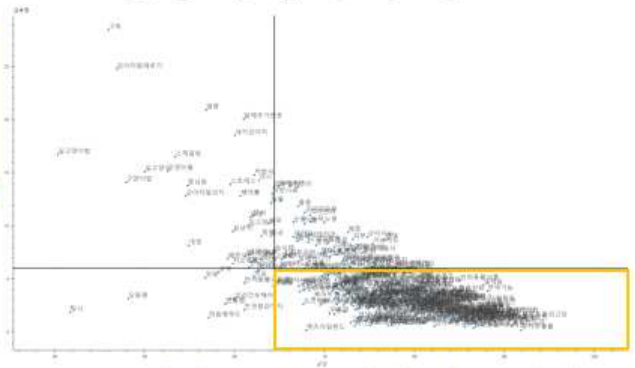
## □ 분석 결과

- SNS 데이터 분석
  - 블로그/카페/커뮤니티의 SNS 데이터에서 반려동물 관련 키워드를  
추출하고 그 단어빈도와 문서빈도, 각 빈도의 증가율, 키워드간  
연관분석, 키워드 감성분석을 진행
  - 각각의 분석결과를 종합하여 빈도의 증가율과 키워드의 긍부정을  
기준으로 신제품 추천 키워드를 도출
- 카드 데이터 분석
  - 반려동물에 관련된 카드 소비 데이터의 구매횟수 및 구매금액을  
지역별, 생애주기별(추정), 성별, 연령별, 업종별 등으로 시각화하고  
핵심 그룹을 도출

## TF와 DF의 성장



## 긍정과 부정의 정도



## 긍정이 높고 부정이 낮은 영역

단어	TF성장	DF성장	TF성장	DF성장	미래신호	태그	네트워킹태그	긍정	중립	부정
강아지영양제	0.044	0.053	0.053	0.053	380.3 S	매뉴	[급여, 건강, 강아지]	78.28	19.19	2.53
프로바이오틱스	0.053	0.053	0.053	0.053	470.8 S	매뉴, 재료	[급여, 제품, 건강, 강아지]	83.66	13.28	3.06
강력	0.004	0.017	0.017	0.017	236.9 S	매뉴	[강아지]	73.43	22.93	3.64
눈건강	0.006	0.031	0.031	0.031	211.2 S	매뉴	[급여, 동물, 건강, 강아지, 간식]	87.42	10.75	1.83
강아지관장영양제	0.108	0.079	0.079	0.079	114.0 S	매뉴	[강아지]	78.61	18.57	2.82
이더드	0.011	0.072	0.072	0.072	159.1 S	매뉴	[급여, 제품, 건강]	69.92	24.95	5.13
글루텐	0.008	0.015	0.015	0.015	166.1 S	매뉴		85.44	13.09	1.47
강아지다이어트	0.071	0.062	0.062	0.062	63.4 W	매뉴		73.68	23	3.31
강아지눈물자극	0.079	0.116	0.116	0.116	52.6 W	매뉴		75.3	19.09	5.61
강아지영양제	0.024	0.085	0.085	0.085	58.3 W	매뉴		82.53	14.15	3.32
혈사포기	0.065	0.087	0.087	0.087	62.8 W	매뉴		71.89	23.57	4.54
반려견영양제	0.107	0.047	0.047	0.047	63.9 W	매뉴		81.62	16.73	1.64
눈영양제	0.247	0.142	0.142	0.142	44.2 W	매뉴		79.37	18.19	2.44
강아지츄르	0.029	0.029	0.029	0.029	37.2 W	매뉴		78.64	19.7	1.66
반려견츄르	0.050	0.032	0.032	0.032	49.8 W	매뉴		81.46	15.45	3.09
영양제추진	0.060	0.100	0.100	0.100	45.4 W	매뉴		76.62	20.77	2.61
과일	0.033	0.023	0.023	0.023	528.3 S	재료		76.65	17.96	5.39
요거트	0.011	0.035	0.035	0.035	224.9 S	재료		79.13	17.22	3.65
블루베리	0.041	0.058	0.058	0.058	256.1 S	재료		81.02	16.68	2.29
고양이유산균	0.079	0.023	0.023	0.023	107.7 S	재료		76.1	20.47	3.43
영양분	0.022	0.023	0.023	0.023	197.4 S	재료	[급여, 건강]	84.74	12.24	3.02
피부건강	0.008	0.015	0.015	0.015	140.1 S	재료	[건강, 강아지]	84.6	13.46	1.95
강아지눈물	0.134	0.091	0.091	0.091	64.0 W	재료		74.23	20.29	5.38
강아지보양식	0.001	0.004	0.004	0.004	64.2 W	재료		80.44	17.76	1.8
프리바이오틱스	0.011	0.045	0.045	0.045	93.9 W	재료		84.04	13.99	1.97
강아지치석	0.027	0.042	0.042	0.042	61.3 W	재료		73.83	21.18	5
반려동물사료	0.226	0.004	0.004	0.004	66.6 W	재료		76.22	16.16	5.62
프로폴리스	0.096	0.124	0.124	0.124	88.8 W	재료		83.8	14.21	1.98
강아지눈영양제	0.316	0.197	0.197	0.197	40.6 W	재료		77.37	20.57	2.06
강아지	0.028	0.031	0.031	0.031	104.8 W	재료		84.13	13.08	2.79
강아지종합영양제	0.092	0.182	0.182	0.182	48.3 W	재료		83.31	15.37	1.31
소프트사료	0.370	0.384	0.384	0.384	35.4 W	재료		84.11	14.95	0.93
육안그레이트블루	0.092	0.106	0.106	0.106	92.9 W	재료		87.82	11.23	0.95
치아건강	0.031	0.025	0.025	0.025	77.6 W	재료		80.85	14.34	4.81
강아지알라지사료	0.119	0.061	0.061	0.061	32.0 W	재료		80.78	16.67	2.55
강황	0.095	0.054	0.054	0.054	45.0 W	재료		84.27	12.72	3.01
피로나민	0.007	0.080	0.080	0.080	55.3 W	재료		85.99	11.77	2.24
피부관리	0.061	0.043	0.043	0.043	46.3 W	재료		80.77	15.12	4.11
노견영양제	0.122	0.011	0.011	0.011	32.2 W	재료		79.96	16.79	3.24
스피루리나	0.003	0.010	0.010	0.010	44.0 W	재료		86.36	11.98	1.65
노랑강영양제	0.125	0.072	0.072	0.072	35.7 W	재료		77.99	18.54	3.47
생채밥	0.104	0.078	0.078	0.078	36.5 W	재료		87.69	7.86	4.44
아스타잔틴	0.232	0.169	0.169	0.169	34.3 W	재료		88.31	10.02	1.67
알코세올	0.075	0.084	0.084	0.084	34.0 W	재료		85.87	13.21	0.92
유카추출물	0.107	0.111	0.111	0.111	34.5 W	재료		91.88	7.94	0.18
오리독백	0.059	0.028	0.028	0.028	59.6 W	기호성		87.63	29.78	2.59

성장 0.05이상

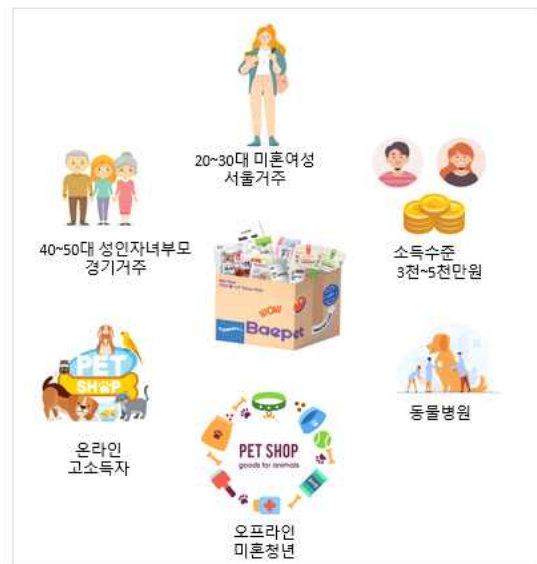
성장 0.1이상

농영양제, 강아지눈영양제, 소프트사료, 아스타잔틴, 유카추출물

자주 출현

강아지 영양 건강

강아지에 영양을 공급해 건강하게 만들자!



[타겟마케팅 이미지]

## □ 주요 성과

### ○ 수요기업과의 워크숍 진행

- 키워드 선택 기준에 대한 토의를 진행하였으며 마케팅팀, R&D팀, 제조영업팀 등 다양한 분야의 시각과 의견이 제시됨
- 추천키워드 중 제형에 관련된 '강아지츄르'에 집중하여 신제품을 기획

### ○ 신제품 기획

- 추천키워드를 사용하여 상품기획안이 작성됨, 영양과 건강에 집중하였으며 타겟으로는 카드 데이터에서 도출된 핵심 그룹을 사용
- '23년도 3월 출시 예상

# 데이터를 활용한 국내 반찬류 소비자의 특성 파악 및 마케팅 전략 수립(네이피)

## □ 과제 목표

- 구매 데이터 분석을 통한 B2C, B2B 마케팅 전략수립
- 소비자 특성 파악을 통한 신상품 기획 방향 도출

## □ 분석주제

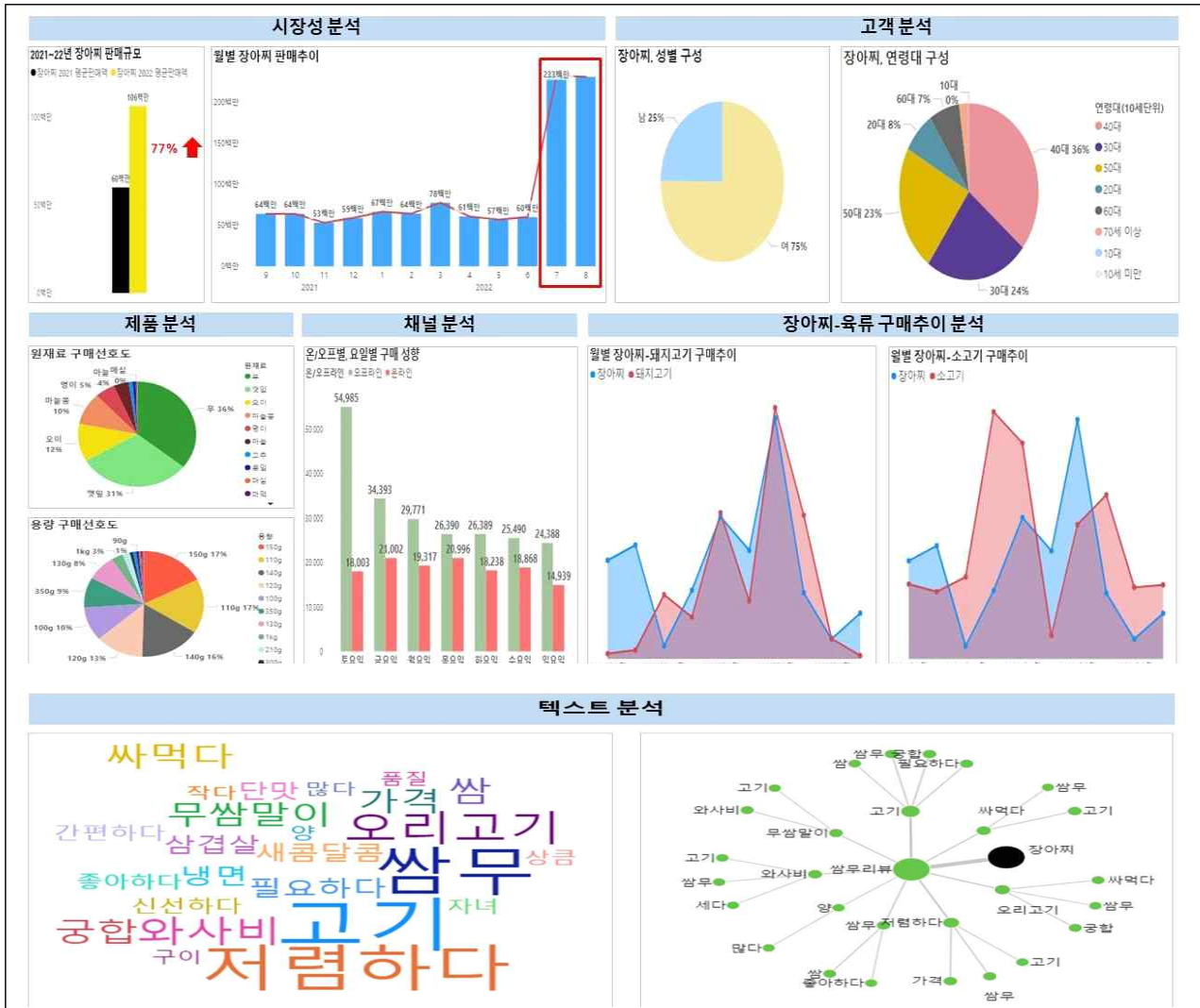
- 절임류의 B2C, B2B 마케팅 전략 수립을 위한 데이터 분석
  - 고객 분석, 제품 분석, 채널 분석, 장아찌-육류 구매추이 분석
- 장아찌 제품군의 인기 상품 파악 및 고객반응 분석
  - 장아찌 인기 재료별 고객 리뷰데이터 텍스트 분석

### <활용 데이터 현황>

- 빅데이터 플랫폼 데이터 : 롯데마트/롯데온의 품목별 구매 데이터
- 외부 수집 데이터 : 장아찌 인기 재료 8종에 대한 오픈마켓 고객 리뷰데이터

## □ 분석 결과

- 장아찌 제품 마케팅 전략도출을 위한 마케팅 분석
  - 주 고객층은 3040 여성. 이에 따른 타겟 마케팅이 필요함
  - 인기 재료는 무와 깻잎이며 100g대의 소포장 전략이 중요함
  - 오프라인에서는 주말효과가 강함. 요일별 프로모션 필요
  - 온라인에서는 주말효과 없고 3040 여성 이용율이 더욱 높음
- 육류와 장아찌간의 구매 추이 연관성 확인
  - 소고기 보다는 돼지고기에서 장아찌와 구매추이의 연관성이 더욱 높게 나타남. 육류 프랜차이즈업체 판로개척에서 돼지고기 프랜차이즈를 우선적으로 공략하여야 함
- 장아찌 제품 고객리뷰 분석을 통해 신제품 개발 방향 도출
  - 장아찌는 종류에 따라 고기와 곁들이는 것과 밥과 곁들이는 것이 다름. 따라서 종류별로 조리방법, 포장방법, 곁들임 대상을 달리 하고 포장 및 마케팅 메시지를 차별화할 것 컨설팅



## □ 주요 성과

- 마케팅 전략 수립으로 인한 매출 증대 예상
  - 분석결과 기반 장아찌 제품 마케팅 전략 제시
  - 목표고객을 대상으로 한 자사물 개편 방향 제시(홍보모델 및 제품 메시지 등)
- 육류 프랜차이즈 판로 확보를 통한 매출 증대 예상
  - 장아찌-육류 연관성 데이터를 기반으로 육류 프랜차이즈에 장아찌 제품 납품업체 지위 획득
- 신상품 개발을 통한 매출 증대 예상
  - 버섯장아찌 단일 재료에서 무, 깻잎 등 제품 라인업 확대

# 이동형 매트리스 살균 건조 서비스(홀템)

## □ 과제 목표

- 사업확장을 위한 합리적 서비스 운영 모델 정립
  - 지역별 잠재수요량 기반 서비스 권역 분할 군집화 모델 개발
  - 서비스 권역별 최적 차고지 선정

## □ 분석주제

- 지역별 서비스 잠재수요량 예측 모형 개발 및 서비스 권역 군집화
  - 서비스 수요예측 모형 개발 및 행정동별 잠재수요 산출
  - 지리적 인접성과 잠재수요를 반영하여 권역 군집화 진행
- 서비스 권역별 최적 차고지 선정 분석
  - 잠재수요량과 교통 인접성 등을 고려한 권역별 차고지 선정 분석

### <활용 데이터 현황>

#### □ 빅데이터 플랫폼 데이터

유통소비(KDX) 플랫폼 LG유플러스 생활인구 가구연계 데이터(서울/인천경기)

#### □ 홀템 서비스주문데이터: 일자, 주소, 주문사유

#### □ 외부 수집 데이터

행정동별 평균소득, 전국주차장정보표준데이터, 서울시 역사마스터 정보 등

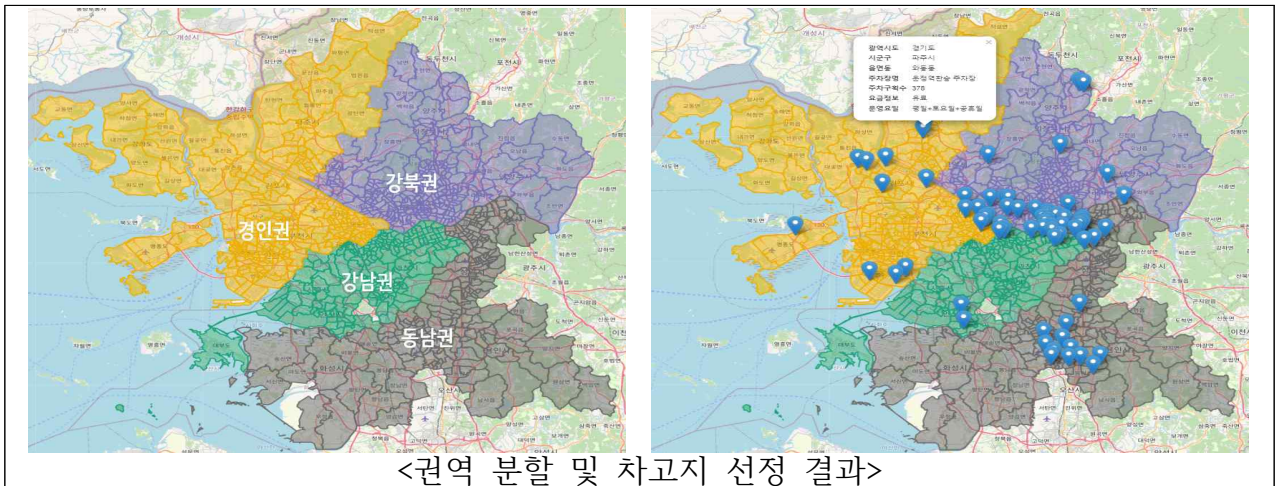
## □ 분석 결과

Parameter Estimates					
Parameter	DF	Estimate	Standard Error	t Value	Approx Pr >  t
Intercept	1	-16.088875	1.575773	-9.58	<.0001
fam2_L	1	0.353574	0.123551	2.86	0.0042
midage2_L	1	-0.511762	0.149689	-3.42	0.0006
income_L	1	0.712289	0.156575	4.55	<.0001
Inf_Intercept	0	-3.750450			
Inf_fam2_L	0	-0.891411			
Inf_midage2_L	0	-1.093569			
Inf_income_L	0	-0.896561			
_Alpha	1	0.408557	0.098042	4.17	<.0001

시군구	행정동	서비스 잠재수요량
화성시	동탄7동	5,171
파주시	운정3동	3,936
화성시	동탄5동	3,762
화성시	동탄4동	3,603
화성시	동탄1동	3,570
파주시	운정2동	3,569
부천시	신중동	3,568
남양주시	다산1동	3,517
서초구	반포1동	3,389
강남구	역삼1동	3,273

<잠재수요량 회귀식 및 잠재수요량 예측 결과>

- 영과잉 음이항 회귀분포를 이용한 서비스 수요량 예측 모델 생성
  - 주문건수 데이터가 영과잉 자료이므로 해당 회귀모형 사용
  - 회귀 분석에 사용할 독립변수로 2인 가구, 중년 2인 가구, 소득 수준을 채택하여 분석 진행, 서비스 수요량 일반식 획득
  - 서비스 목표 수요량을 100만 가구로 설정했을 때의 행정동별 잠재수요량을 통해 지역별 서비스 이용 추세를 파악할 수 있음



- 서비스 잠재수요량 기반 구역 분할 및 차고지 선정
  - 교통 인접성을 고려하고, 구역간 잠재수요량 차이가 크지 않게끔 서비스 지역 내에 서비스 권역을 4개로 분할함
  - 잠재수요량 및 예상 침투율\* 을 활용하여 구역별 서비스 중점 행정동을 선정, 해당 행정동에서 지하철역과 버스 정류장 등의 교통역 인접도가 제일 높은 주차장을 차고지로 선정

\*예상 침투율: 지역의 전체 가구수에서 잠재수요가 있는 가구가 차지하는 비율

## □ 주요 성과

- 지역별 기회시장 포착 및 시장침투로 매출 증대
  - 지역별 잠재수요량 및 예상 침투율을 활용하여 행정동 단위에서 잠재 시장규모 파악 및 최우선 침투시장 파악
- 차량 최적 차고지로 서비스 리드타임 단축으로 운영 효율성 증대
  - 최적 차고지 배치로 차량 이동시간을 단축함으로써 일일 서비스 제공 건수 및 차량당 서비스 매출 증대

# 사물인터넷(IoT) 기반 웨어러블 안전장비 및 통합관리솔루션 개발(코렐테크)

## □ 과제 목표

- 철도작업자 상태 파악을 위한 '작업자 행동패턴 파악 방안' 마련
- 대피 알람 정확도 향상을 위한 '열차위치 정확도 향상 방안' 마련

## □ 분석주제

- IMU 센서데이터 기반 행동 패턴 분류 알고리즘 가이드 컨설팅
  - 공개데이터(UCI)를 기반으로 행동패턴 알고리즘 개발과정을 프로토타입으로 개발
  - 분류 알고리즘 개발 방법 및 과정에 대한 컨설팅
- 열차 위치 추정방안 컨설팅
  - 열차 위치 추정의 접근방법을 도출
  - 가상의 열차 위치를 생성하고 CTC 및 GPS 열차 위치 정보를 활용하여 통과 시간을 산출하는 처리 방법 도출

### <활용 데이터 현황>

#### □ 내부 데이터

- 1) 열차 위치 데이터(코렐 CTC 데이터): 주요컬럼(열차번호, 궤도위경도 등)

#### □ 외부 데이터

- 1) UCI IMU 센서데이터\*: 레코드수(100만), 주요컬럼(가속도 3개, 각속도 3개 등)
- 2) 자체 수집 IMU 센서데이터: 레코드수(4만), 주요컬럼(가속도 3개, 각속도 3개 등)

\*캘리포니아어바인대학교(UCI)의 Machine Learning Repository의 공개데이터

## □ 분석 결과

- IMU 센서데이터 기반 행동 패턴 분류 알고리즘 가이드 컨설팅
  - 작업자 행동 패턴을 분류하기 위해 XGBoost, Random Forest, SVM, Logistic Regression, Naive Bayes를 사용하였으며 [표 1]과 같이 결과를 얻었으며, 그 중 정확도가 0.9 이상인 XGBoost를 최종 선택함

- 아래 [표 2]에서 정적인 행동(앞기, 늡기, 서기)의 정확도가 점차 향상된 것으로 보아 향후 열차 접근 대피 프로세스의 효율적 적용이 가능한 것으로 보임

알고리즘	Step①	Step②	Step③
XGBoost	0.86	0.90	0.97
Random Forest	0.81	0.86	0.95
SVM	0.79	0.83	0.90
Logistic Regression	0.78	0.80	0.89
Naive Bayes	0.51	0.58	0.65

[표 1] 알고리즘별 예측정확도

XGBoost	Step①	Step②	Step③
전체	0.86	0.90	0.97
앞기	0.74	0.76	0.93
계단 내려가기	0.96	0.98	0.97
서기	0.79	0.86	0.95
계단 오르기	0.94	0.95	0.98
걸기	0.97	0.98	0.99
늡기	0.82	0.88	1

[표 2] XGBoost의 행동별 예측정확도

#### ○ 열차 위치 추정방안 건설팅

- 대전, 대구 지역의 열차번호 183번으로 분석 실행한 결과는 아래 그림과 같음
- 고정좌표의 개수가 늘어나고, 좌표 간의 최대거리가 줄어들어 따라 GPS 기기가 없더라도 작업자의 안전도가 높아질 것이라고 기대됨



[그림 1] AS-IS



[그림 58] TO-BE

### □ 주요 성과

- 철도작업자 안전상태 모니터링 통합 솔루션의 핵심모듈 개발
  - 사내 스타트업의 서비스 솔루션의 핵심 모듈 개발로 투자 유치
  - 코레일 등을 대상으로 안전관리 솔루션 판매를 통한 매출 확보