행정안전부 NIA 한국지능정보사회진흥원

비매품
03000
9 788984 836297
ISBN 978-89-8483-629-7

행정은 스마트하게, 국민은 편리하게!

공공부문 데이터 분석·활용 우수사례집

행정안전부 NIA 한국지능정보사회진흥원

공공부문 데이터 분석·활용 우수사례집



행정안전부 NIA 한국지능정보사회진흥원

공공부문
데이터 분석·활용
우수사례집

행정안전부 NIA 한국지능정보사회진흥원

CONTENTS



Part 1. 공공행정

빅데이터 분석으로 지능형 지방세정 실현 | 행정안전부 6
사용자 편의성 강화한 생활불편신고앱 개발 | 행정안전부 12
선제적인 AI 기반 제안요청서 진단모델로 품질향상 | 조달청 18
국유지 무단점유 꼼짝 마! 국·공유지 모니터링 책임지는 드론이 뜬다 | 통영시 24



Part 2. 교육복지

빅데이터 분석으로 공동주택 돌봄수요 미리 예측! | 복지부·교육부·국토부 32
공공데이터의 사회적 가치창출로 장애인 이동권에 공헌 | 일반 국민 38
사각지대 없이, 복지 앞에 모두 평등한 서비스 구축 | 안성시 44



Part 3. 재난안전

빅데이터 기반 해양안전 위험지수 도출로 안전한 바다를! | 해양경찰청 52
산불 예방과 피해절감 솔루션 소방용수시설 | 소방청 58
영사민원시스템에 기반해 선제적으로 재외국민 보호 | 외교부 64
교통 빅데이터 로드맵으로 맞춤형 교통안전지원 모델 개발 | 한국교통안전공단 70
생명 구하는 일등공신, 응급차량 출동을 책임지는 인공지능! | 대전광역시 76



Part 4. 보건의료

자동차 보험금 누수, 이상패턴 감지해 과학적으로 막는다 | 건강보험심사평가원 84
도서지역 건강권을 보장하는 응급시스템 스텝업 | 인천광역시 90
맞춤형 치매 진단 검사로 다각적인 치매 관리 | 김해시 96



Part 5. 산업고용

편리하게 사용하고 만족도 높이는 생활밀착형 서비스 | 한국전력공사 104
근로자 권리를 보호하는 믿음직한 임금체불금 지킴이 | 근로복지공단 110
취준생을 위한 맞춤형 취업 지침서 | 한국교육개발원 116

Part 1

공공행정

빅데이터 분석으로 지능형 지방세정 실현 | 행정안전부

사용자 편의성 강화한 생활불편신고앱 개발 | 행정안전부

선제적인 AI 기반 제안요청서 진단모델로 품질향상 | 조달청

국유지 무단점유 꼼짝 마! 국·공유지 모니터링 책임지는 드론이 뜬다 | 통영시

공공
행정

빅데이터 분석으로 지능형 지방세정 실현

정확한 분석과 예측으로 공정하고 신뢰받는 납세문화 조성



대부분은 성실히 지방세를 납부하고 있지만 약 400만 명의 국민은 여러 가지 이유로 지방세를 체납하고 있다. 체납징수를 위해 200여 종에 이르는 체납자의 과세정보를 확인하는 데는 약 3시간이 소요되고, 매년 40억 원 가량의 비용이 들어간다. 이런 비효율을 개선하기 위해 행정안전부는 빅데이터 분석으로 납부가능성을 미리 예측해 체납액을 징수하는 방안을 도입했다.

#징수가능성예측 #머신러닝 #스코어전환 #등급분류 #체납유형별징수활동

모델 개발 전

- 높은 지방세 체납률
- 체납자 상황 반영 없는 획일적 징수
- 사람이 직접 체납징수하는 방식으로 많은 시간과 비용 발생

모델 개발 후

- 체납유형과 등급 정도 등을 활용하여 맞춤형 징세활동 추진
- 연간 520시간과 17.4억 원의 비용 및 시간 절감 기대
- 지능형 지방세정 실현을 통한 공정한 납세문화 조성

빅데이터 분석으로 여는 미래형 납세시대

지금은 방대한 양의 데이터를 분석해 미래를 정확히 예측해내는 빅데이터 시대이다. 조세행정도 예외는 아니어서 과학적인 분석이 필요하다. 이에 따라 행정안전부는 지난 5년간 체납자 980만 명, 분석자료 9,500만 건을 수집하고 데이터를 분석했다. 이를 통해 징수가능성을 예측하고 정제해, 체납 유형별로 나타난 특성을 확인했다.

체납자 데이터 분석을 통해 단기체납자, 장기체납자, 생계형체납자 등 상황과 유형별로 총 5등급을 구분할 수 있었다. 체납자들은 저마다의 상황과 이유에 따라 체납 원인이 다르므로 빅데이터를 활용해 세정 정보를 꼼꼼하게 분석했다. 각 지자체별 과세자료를 분석하고, 신용정보 수집, 한국지역정보개발원과 연계해 납세 자료를 추출, 현장설명회를 통한 관련사항 공유와 의견 수렴에 이르기까지 다각적인 노력을 펼쳤다.

이를 토대로 어떤 요인에 따라 납부율이 변하는지 체납자별 분석을 통해 납부율을 예측함으로써 궁극적으로는 현장에서 활용할 수 있고, 국민과 소통하는 세무 서비스를 펼칠 수 있도록 차세대 시스템 구축을 추진했다.

신뢰성을 높인 데이터 추출 및 분석

행정안전부는 실제 데이터 분석 결과를 징수 현장에서 활용하여 실효성 있는 분석이 이루어지도록 경기도 31개 시군과 협업체계를 구축했고, 체납자별 빅데이터 분석을 수행한 뒤, 그 결과를 제공하기로 했다. 체납분석은 '필요데이터 항목 정의(데이터 수집, 시스템 구축) → 후보항목 설계(데이터 가공, 머신러닝) → 빅데이터 분석모델 마련(체납자 유형 분류, 6개월 이내 납부가능성 산출) → 납부율 예측 후 등급 생성(스코어 전환, 등급 부여) → 체납분석 보고서 제공(체납자별 요약자료 제공, 체납징수 현장 활용)' 과정으로 이루어졌다.

먼저 과거 5년간 체납 경험이 있는 980만 명을 대상으로 9,500만 건의 데이터를 수집했다. 각 지자체별로 부과, 납부, 체납, 결손 등 과세자료를 추출했고, 민간 신용정보사와 한국지역정보개발원의 협조하에 신용정보를 수집했다. 수집된 데이터는 암호화 처리를 통해 민감한 과세정보 취급에 각별히 유의했고, 수시로 설명회를 개최하여 관련 사항을 공유했다. 각 기관 및 지자체별로 보유한 데이터의 형태와 관리 형식이 상이해 수집 데이터를 분석하는 과정에는 상당한 어려움이 있었다. 데이터의 포맷이 통일되어 있지 않거나 잘못된 데이터가 포함되는 등의 문제를 제거하고 분석이 가능한 품질의 데이터를 만들기 위해 정보를 정제하고 조합하는 전처리 과정이 필요했다.



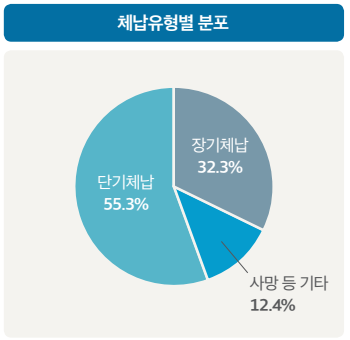
분석수행으로 도출한 유형별, 등급별 특성 확인

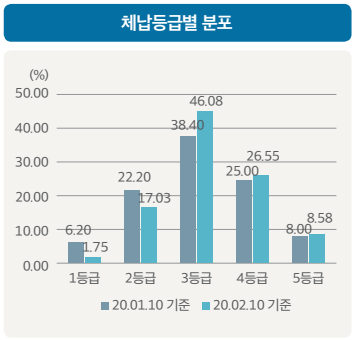
분석 결과를 통해 전체 체납자 중 단기 체납 55.3%, 장기 체납 32%, 사망 등 등급산출 제외 10%, 결손처분 2.5%로 구분이 가능했다. 이들을 각 유형별로 구분하여 특성을 파악한 결과, 단기 체납자는 납부 의지와 성실도가 가장 높은 특성을 나타냈다. 체납 즉시 문자 안내 등으로 집중관리할 경우 납부율이 상승한다는 결과를 얻을 수 있었다. 이들은 6개월 내 납부가능성이 53.2%에 이르는 것으로 예측할 수 있었다. 한편 장기 체납자의 납부 키포인트는 경제력인 것으로 확인됐다. 이 그룹에 속한 이들은 압류, 공매 등 적극적인 경제조치가 징수에 효과적인 것으로 분석되었으며, 6개월 내 납부가능성은 16.3%에 불과한 것으로 예측되었다. 마지막으로 생계형 체납자는 납세의지가 납부율에 큰 영향을 미치는 그룹으로 이들에게는 체납처분 유예, 복지 서비스 연계 등의 지원책을 마련할 필요가 있음이 확인되었다.

이러한 분석수행 결과를 확인함으로써 기존 수작업 위주의 행정처리와 획일적인 징수시스템에서 벗어나 조세행정에서도 빅데이터를 적극적으로 활용해야 한다는 사실에 공감할 수 있었다.

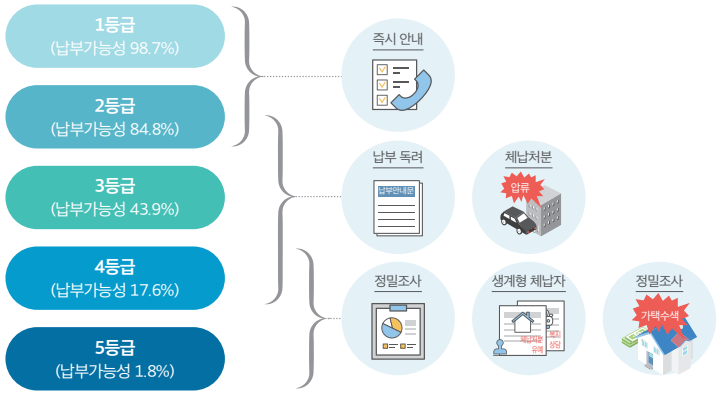
등급 특성에 따른 분류로 납부율 증대

행정안전부는 분석에 따라 체납 등급별 특성을 추출해냈다. 1~2등급(납부가능성 84.7%~98.7%)은 상담으로 납부를 안내하고, 3~4등급(납부가능성





체납 유형별 징수 활동



17.6%~43.9%)은 납부 독려 및 체납처분 조치를 취할 것을, 그리고 5등급(납부가능성 1.8%)은 체납처분을 병행한 실태조사를 실시하는 등의 징수활동을 추천했다. 또한 분석보고서를 통해 체납자 요약정보를 간편하게 확인하고, 현장 출장 시 사용할 수 있도록 제공했다.

이러한 소통과 개선으로 이제 매월 갱신되는 정보가 반영된 분석 결과를 통해 앞으로는 체납자 특성에 맞는 맞춤형 징수활동이 이루어질 수 있게 되었다. 체납 분석 시간과 관련 비용도 각각 연간 520시간과 17.4억 원씩 크게 절감될 것으로 기대된다.

현장에서 적극적인 활용으로 분석 가치 높여야

행정안전부는 이번에 구축한 분석데이터를 체납자의 투명하고 공정한 세금 납부 유도에 활용하도록 요약정보를 제공했다.

지자체는 분석결과를 기반으로 실제 현장에 적용하였다. 1등급 체납자에게는 독촉고지서를 발송하고, 2~3등급 체납자에게는 문자메시지를 발송했으며, 이 중 일정 금액 이상을 체납한 자에게는 압류예고서를 발송했다. 4~5등급 체납자는 가택수색을 실시하는 등 맞춤형 징수활동을 했다.

현장 활용 결과는 수시로 공유했고, 개선 의견을 청취하여 보고서를 보완하기도 했다. 또한 분석 결과를 제공하는 주기도 지방세 부과 일정에 따라 유연하게 적용하는 등 업무 활용도를 높이기 위한 방안을 모색하고 있다.

국민이 공감하는 세무행정 서비스

모두에게 똑같이 이루어지던 징수활동 및 체납처분도 이제 체납 유형에 따라 다르게 할 수 있게 되었다. 깜빡 잊고 세금을 내지 않은 국민이나 생계형 체납자에게는 친절한 안내로 납부를 돕고, 체납처분을 유예해주며, 복지상담과 연계해줄 수 있다. 하지만 고의로 세금을 피하는 경우 즉시 체납처분을 진행해 재산을 압류하고 출국을 금지하며 명단을 공개해 불필요한 민원을 줄임과 동시에 세금납부에 대한 의무를 다하도록 했다.

행정안전부는 여기에 머무르지 않고, 차세대 지방세 시스템으로 개편하기 위해 준비하고 있다. 차세대 시스템에서는 데이터가 단일화된 형태로 모이고, 동일한 형태로 관리되어 더욱 수월하게 체납자 분석이 가능할 것으로 기대된다.

지자체는 맞춤형 행정과 국민에게 친절한 세무 서비스를 제공할 수 있는 차세대 시스템을 갖출 수 있게 되었으며, 국민 누구나 책임감 있고 편리하게 세금을 납부하고, 공감하는 세무행정 구축을 통해 세금 잘 내는 나라로 나아갈 수 있게 되었다.

빅데이터 분석, 이렇게 달라졌어요!

- ☑ 등급과 유형 분석에 따른 납부가능성을 예측해 효율성을 강화한 차세대 조세행정 시스템으로 개편
- ☑ 체납 유형에 따라 맞춤형 징수방법을 적용하는 세무 서비스 제공
- ☑ 지자체와의 공유로 현장에서 활용도를 높이는 가치 있는 분석 데이터 활용

공공
행정

사용자 편의성 강화한 생활불편신고앱 개발

신속정확한 민원처리 서비스 제공



자동차 이용이 급증하면서 그에 따른 부작용이 사회적인 문제로 대두되고 있다. 대표적인 것이 바로 불법주정차와 관련된 것. 이를 해결하기 위해 개발된 생활불편신고앱은 신고 대비 과태료 부과율이 현저히 낮아 만족도와 효과성 면에서 부정적인 의견이 많았다. 이에 행정안전부는 민원처리 효과를 높이고자 생활불편신고앱 서비스를 개선하기로 했다.

#불법주정차감소 #단속차량 운행 최적화 #공간분석(GIS) #생활불편신고앱

모델 개발 전

- 민원유형 수동 선택
- 분류 체계 문제로 오 지정 가능성 존재

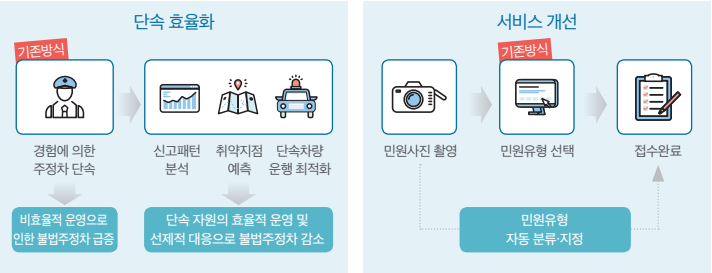
모델 개발 후

- 민원 유형 자동 분류 및 지정
- 사용자 편의성 강화
- 데이터 기반 단속 차량 운행 경로 최적화

효율적 운영과 서비스 개선으로 편리한 사용 여건 마련

불법주정차는 단순히 생활의 불편함을 초래하는 것에 그치지 않고 교통사고를 유발하거나 소방차 진입을 방해하는 등 시민의 안전과 재산권을 위협한다. 차량의 증가 속도에 비례해 불법주정차는 매년 증가하고 그에 따라 불편신고 건수도 꾸준히 늘어나고 있다. 2012년 생활불편신고앱을 통해 접수된 불법주정차 신고가 38,544건에 불과했던 것에 비해 2017년에는 957,618건으로 급격히 증가했다. 그러나 생활불편신고앱은 신고 시 민원 유형을 수동으로 선택해야 하고, 분류 체계가 비슷해 오 지정 문제도 존재했다.

이러한 불편함과 문제점을 해결하기 위해 행정안전부는 생활불편신고앱 신고 데이터를 분석했다. 이를 통해 지역별 불법주정차 빈발지 현황 및 예측, 이동형 단속 차량 운행 경로 최적화 분석을 진행해 운영의 효율성을 높였다. 더불어 상습 신고 지역에서 선제적으로 대응함으로써 불법주정차 감소를 유도해 단속 효율화를 추구하고자 했다. 또한 이미지나 텍스트로 신고 내용별 민원 유형을 자동으로 분류하는 모델을 개발하여 사용자의 편의성을 강화할 수 있었다.



데이터 융합과 다양한 알고리즘 활용한 최적의 모델 구현

생활불편신고앱을 통해 최근 6년간 접수된 민원은 약 460만 건에 이른다. 여기에 외부 데이터를 통해 접수된 민원도 있다. 서울과 경기도 지역의 CCTV 설치 지점 정보 약 12만 건, 업소 정보 약 100만 건 및 읍면동 단위 인구 정보, 버스 정류소 정보 등을 결합하여 분석에 활용했다.

행정안전부는 신고 데이터와 외부 데이터를 융합하여 공간분석(GIS)을 수행해 지역별·시간대별 불법주정차 현황을 시각화했다. 운행 거리와 소요시간, 지역별 신고 건수를 고려한 데이터 기반 단속 차량 운행 경로 최적화도 진행했다. 이미지와 텍스트 인식 분야에 널리 활용되는 모델을 구현했는데, 이 모델을 생활불편신고에 적용해 정확도를 높일 수 있었다.

데이터로 확인한 불법주정차 신고 증가와 미흡한 행정처분

행안부는 먼저 신고현황을 점검했다. 서울·경기 지역 불법주정차 신고 건수는 총 538,223건(15년 1월~18년 8월)이고, 경기도에서는 ○○시가 26,838건으로 가장 많았다. 서울에서는 ○○구가 평균보다 2.9배 높은 23,430건으로 나타났으며, 서울 ○○구의 경우 100명당 5.7건으로 인구 대비 신고 건수가 가장 높았다. 또한 불법주정차 신고는 매년 증가 추세로 2018년 상반기에는 2015년 상반기 대비 약 4.2배 증가했고, ○○시는 같은 기간 무려 17.2배가 증가한 것으로 확인됐다.

처리결과에 대한 데이터도 수집했다. 불법주정차 신고 민원의 45%는 기각되고, 35%는 과태료가 부과되었으며, 20%는 현장 단속 및 계도로 처리됐다. 신고된 민원의 65%가 행정처분이 이루어지지 않은 것이다. 특히 지자체마다 단속 정책과 처리율이 상이했는데, ○○구와 ○○시는 시민이 제출한 위반 증거를 기반으로, ○○구와 ○○시 △△△구는 단속 및 계도 위주로 처리하고 있었다.

서울시	행정처분	단속·계도	불충족	경기도	행정처분	단속·계도	불충족
평균	25%	28.1%	46.9%	평균	43.7%	13.9%	42.4%
○○구	81.6%	0.3%	18.2%	○○시	59.2%	1%	39.8%
○○구	3.9%	86.4%	9.7%	고양시 △△△구	6.8%	66.3%	26.9%

이러한 민원처리 결과는 고객 평점과도 높은 상관관계를 보였는데 행정처분이 이루어졌을 경우 평점이 높아지는 반면 불충족 시에는 낮아지는 현상이 나타났다.

통계분석과 시각화 작업

분석을 위해 지역별 불법주정차 현황에 근거해 신고 건수와 과태료 부과처리 등을 확인하여 이를 각 지자체별로 시각화했다. 이와 더불어 연도별·반기별 신고 추이와 월별·요일별·시간대별 신고 추이, 처리결과별 고객 평점 등록 비율과 평점 평균 및 등록 건수도 확인했다. 처리결과와 고객 평점 간 상관관계 분석 시에는 고객 평점의 분포가 통계적으로 유의한지 분산분석*을 통해 검정하여 집단 간 평균의 차이가 있음을 확인할 수 있었다. 또한 피어슨 상관관계 분석*을 통해 처리결과별 고객 평점의 상관계수도 도출했다.

* 분산분석(ANOVA): 두 개 이상으로 구성된 집단의 평균을 비교하는 목적으로 사용되는 통계분석 기법
* 피어슨 상관관계 분석: 두 변수 사이의 선형 관계 및 방향을 파악하기 위한 상관관계 분석 기법



효율적인 개선방향을 찾을 수 있는 정확한 결과 산출

신고데이터를 샘플링해 CNN* 기반 텍스트 분류 모델로 불법주정차 신고를 처리한 결과 답변 내용이 동일하고 빈도가 높은 데이터를 추출하여 약 6,000건의 학습 데이터 샘플링을 확보할 수 있었다. 더불어 600건의 평가 데이터도 확보했다. 레이블링과 결과 검증을 위해 분류 기준에 따라 레이블링 후 교차 검증을 실시함으로써 정확도를 높였다.

* CNN: 데이터(주로 이미지) 내 숨겨진 패턴이나 특성을 찾는데 유용한 알고리즘

상시 모니터링과 소요시간 단축에 활용

생활불편신고앱은 기존에는 행정안전부가 신고건수와 함께 공문을 작성해 17개의 광역시도에 발송하면 광역시도는 이를 다시 228개 시군구와 공유했다. 광역시도는 자료요청과 취합 과정을 거쳐 처리결과를 다시 행정안전부에 송부하는 방식이었다. 이는 약 2개월이 소요되어 단속과 계도, 불충족 비중을 확인하기 어려웠고, 불충족 요인 분석도 불가능했다.

하지만 이번 모델 개발을 통해 불충족 민원처리 결과를 자동으로 분류, 처리결과를 판별할 수 있게 되었다.

앱으로 간편하게 신고해 생활불편 해소

이번 모델 개발을 적용하면 간단한 방법으로 신고가 가능할 것으로 예상된다. 생활불편 스마트폰신고 앱을 설치한 후 앱에서 바로 촬영 후 업로드하는 방식으로, 신고내용에는 '신고위치'를 입력하는 것만으로 신고접수가 이루어진다.

세종시는 이러한 생활불편 스마트폰신고 앱을 활용한 단속을 전 지역을 대상으로 추진할 예정이다. 단속구역은 주정차가 엄격히 금지되는 6개 구역의 불법주정차로 제한했다. 해당 지역은 교차로, 횡단보도, 보도(인도) 전 구간에 걸쳐 실시되며 교차로의 가장자리나 도로의 모퉁이로부터 5미터 이내 구간을 포함한다. 버스정류장도 단속구역이다. 정류지임을 표시하는 기둥이나 표지판 또는 선이 설치된 곳으로부터 10미터 이내 구간이 해당된다. 소화전은 5미터 이내가 단속구역이며, 어린이보호구역의 경우 어린이보호구역 표시 구간에서 실시된다. 운영시간은 오전 8시부터 오후 8시까지이다.

이번 모델을 불법주정차 민원의 과태료 부과율과 단속 정책 등에 적용하면 처리율을 높이고, 고객만족도 향상을 꾀할 수 있을 것으로 전망된다.

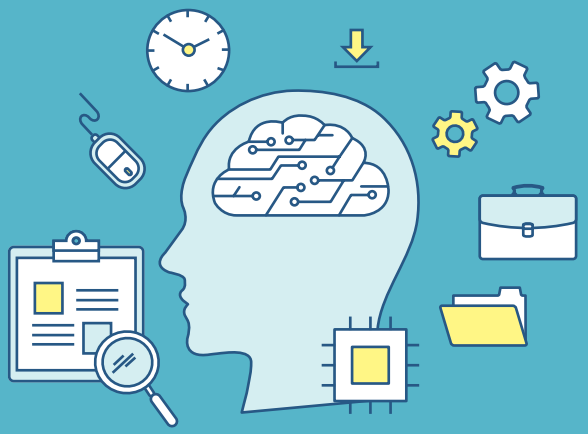
빅데이터 분석, 이렇게 달라졌어요!

- 불법주정차 민원 특성 및 처리 패턴 분석
- 불법주정차와 관련된 다양한 특성을 비교·분석하여 신속정확한 민원처리
- 빠른 민원처리를 통한 고객만족도 향상
- 생활불편신고앱 사용자 편의성 강화

공공
행정

선제적인 AI 기반 제안요청서 진단모델로 품질향상

AI로 제공하는 편리한 조달 서비스



최근 정보화(ICT) 사업과 같은 공공 발주 건이 늘어나고 있다. 그러나 카테고리별 법·제도 위반, 특정 업체에 유리한 조항이 포함되는 경우도 빈번해 조달청에서 제안요청서(RFP)를 검토하는 데만 건별 4~8시간이 소요되고 있다. 이에 조달청은 법규 위반 문장을 자동으로 판단하는 AI 기반 진단모델을 개발·적용하여 업무 효율성을 증가시켰다.

#제안요청서 #일하는방식 개선 #AI 기반 분석모델 #텍스트마이닝 #BERT

모델 개발 전

- 제안요청서 직접 검토로 건별 4~8시간이 소요되어 업무 효율성 저하
- 데이터 기반 제안요청서 사전검토 및 분석 서비스 필요성 증대

모델 개발 후

- 인공지능 기반 문장단위 비교분석으로 법규 위반 가능성 판단
- 효율적인 업무방식으로 검토 시간 단축
- 충분한 사업기간 확보로 사업의 품질력 향상과 입찰 경쟁력 강화

잘못된 조달요청을 방지하는 사건검토 분석 서비스 필요

정보화(ICT) 사업 등 공공발주 건이 매년 증가하고 있으나 발주기관의 경험이 없는 경우, 제안요청서(RFP) 검토에 상당한 시간이 소요되어 잘못된 법제도와 특정업체에 유리한 독소조항을 포함한 경우가 빈번하게 발생했다. 제안요청서에 수정사항이 많아지면 입찰공고가 늦어지고, 연내 완료 사업의 경우 사업기간이 절대적으로 부족해 공공SW 사업 품질이 저하되는 결과를 초래하게 된다. 이를 사전에 방지하기 위해 법규준수 진단에 반드시 검토해야 하는 용어사전을 구축할 필요가 있다. 더불어 제안요청서 검토단계에서 기존 데이터로부터 학습한 결과를 바탕으로 조달요청 발주를 지원하는 모델 개발이 절실했다. 조달청은 수요기관과 조달청 간의 검토와 반영시간을 단축하고 조달 품질 역시 자연스럽게 향상되는 효과를 창출하고자 했다. 이에 선제적인 AI기반 제안요청서 법규준수 진단모델을 개발함으로써 조달요청 내역 현황을 분석하고 조달요청 RFP 내 필수검토 용어사전을 구축하여 사전검토 모델을 구현하고자 했다.

01 현황분석 → 02 사전 구축 → 03 사전 검토 모델 개발

분석 활용 데이터

입찰정보 (xls)	• 공고명, 수요기관명, 입찰방법, 사전규격 공개일, 입찰개시일 등 • 공동계약허용 여부, 지역제한 여부, 분할납품 여부, 협상계약 여부 등 • 계약별 유형, 평가방식	조달요청일 기준 2016~2020 (10,526건)	최종 활용 조달요청건 (9,153건) hwp 파일 기준 18,306개
제안요청서 (hwp)	• 사전규격 공개용 제안요청서 • 최종 공개된 제안요청서	hwp 파일 기준 20,894개	

직접 검토는 그만, 이제 AI가 한다!

가장 큰 문제는 담당자의 업무 효율성이 떨어지고, 일정이 지연된다는 점이었다. 사업은 조달요청 내역 현황 분석, 조달요청서 내 필수 검토 용어사전 구축, 사전검토 모델 개발 순으로 진행됐다.

독소조항 등이 수정되기 전인 '사전규격공고' 파일과 위반내용이 수정된 '최종공고' 파일의 내용을 추출해 문장단위 비교 분석을 통해 위반가능성을 판단할 수 있도록 했다. 또한 법규위반 항목을 확인해 담당자가 정확한 수정을 하는데 참고하도록 과거 제안요청서 데이터를 탐색하고 유사한 문장을 제시하도록 했다. 이번에 구축한 모델은 담당자가 긴 시간동안 제안요청서를 검토해야 하는 업무방식에서 벗어나 인공지능 기반 사전검토 모델의 도움을 받아 법제도 위반 가능성 있는 문장들을 신속하고 정확하게 검토 완료함으로써 업무효율성을 높이는 것이 포인트였다. 업무방식의 전환을 이끌어내는 것은 물론 시간을 단축함으로써 발주기관 입장에서는 충분한 사업기간을 확보할 수 있게 되어 전반적인 사업의 품질향상을 이끌어낼 수 있고, 입찰경쟁력도 강화할 수 있는 선순환 구조를 정착시킬 수 있다.

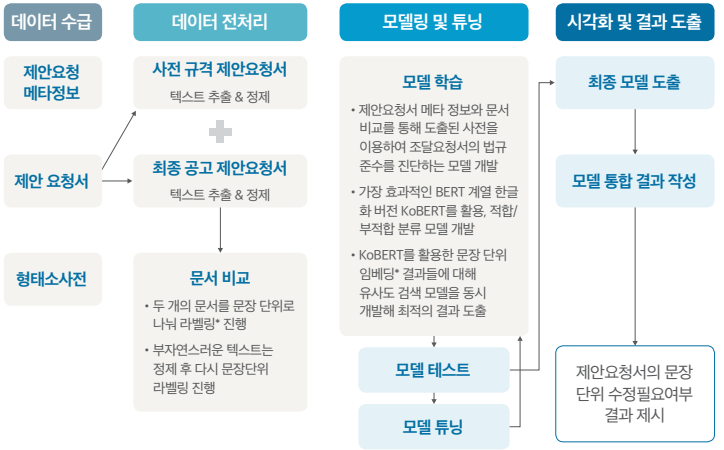
보다 효율적인 모델 개발을 위해 두 차례의 자문회의가 진행되었다. 1차 자문회의에는 문장 유사도를 도출할 수 있는 다양한 의견이 제시되었으며, 조달청은 이를 반영해 모델 개발을 진행했다. 더불어 품질을 향상시킬 수 있는 추가데이터도 확보했다. 2차 자문회의에서는 분류모델과 유사도 모델 결과를 활용해 공고 상에서 수정이 필요한 문장에 대해 관련 법규 삽입 방법을 제안해 향후 모델에 활용하도록 했다.

데이터 분석으로 정확해지는 조달업무

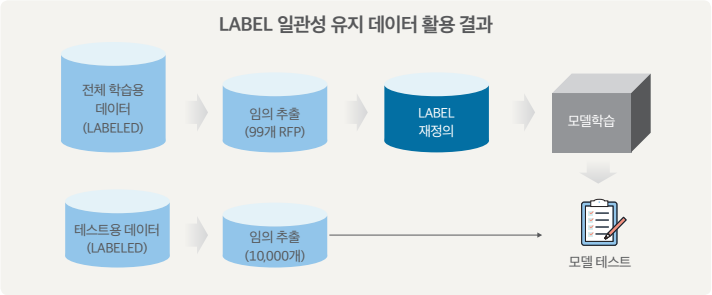
분석은 '데이터 수급→데이터 전처리→모델링 및 튜닝→시각화 및 결과 도출' 과정으로 이루어졌다.

발주공고 기관별(조달지청별) 조달요청 처리기간(공고일자-접수일자=일수차) 현황 분석으로 지역별 발주 특성을 도출하기도 했다. 이를 통해 제주와 충북이 평균 처리기간과 표준편차가 상대적으로 낮아 효율적인 처리가 이루어지는 것처럼 보이나, 전체 조달 건수가 낮은 편이라는 점도 확인할 수 있었다. 또한 서울지방조달청은 조달요청이 많음에도 불구하고 75% 이내에 처리해 기간이 제주, 충북과 유사하면서 평균 처리기간과 편차가 낮아 가장 효율적으로 처리되는 지청이라는 점도 밝힐 수 있었다.

텍스트(줄 단위) 분포 분석으로는 제안요청서(HWP)에서 추출된 텍스트의 문장 길이 분포를 확인해 검토가 불필요한 문장 길이가 어느 정도인지 파악했다.



* 임베딩(embedding): 사람이 쓰는 자연어를 컴퓨터가 이해할 수 있도록 숫자 형태(Vector)로 변환하는 것
* 라벨링(labeling): 각 문장별 적합/부적합을 태그



*수정이 필요한 경우 1, 수정이 불필요한 경우 0으로 분류

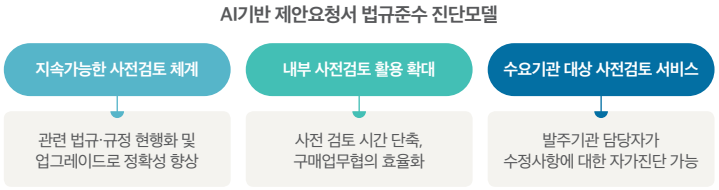
사용편의성을 높여 납품 품질에 기여할 방안 구축

더욱 우수한 결과를 얻기 위한 노력도 이어졌다. 담당자가 모델을 현업에서 사용해 볼 수 있도록 하여, 실제 활용 시 개선할 부분에 대한 피드백을 받아 이를 반영했다. 먼저 실행 프로그램 환경개선 및 결과 파일 포맷(Excel)을 변경할 것을 요청했으며, 유사도모델과 분류모델을 하나로 통합하는 것이 유익하다는 점을 밝혔다. 또한, 향후 활용 방안을 상세화할 것도 요청했다. 조달청은 HTML 형태로 제공되는 결과 파일을 엑셀에서 활용할 수 있도록 개선했으며, 분류모델 결과와 유사도모델 통합 리포트를 제공했다.

더 나은 조달서비스를 만드는 활용 방안

조달청은 분석 결과를 다양한 방면으로 활용할 계획이다. 먼저 조달요청 문서의 부적합 문장을 검출하는 데 활용한다. 이는 조달요청 문서의 부적합 가능성이 있는 문장을 선제적으로 제시함으로써 조달업무 효율화, 충분한 사업기간 확보, 사업품질 향상, 입찰 경쟁성 강화에 기여할 것이다. 조달 담당자 구매업무 협의 과정 중 사전검토에 소요되는 시간을 단축시키는 데도 일조할 것이다. 단기적으로, 관련 법규 변경 사항 등을 반영해 주기적인 모델 업그레이드로 정확성을 향상시킬 예정이다.

중장기적으로는 양질의 데이터를 추가로 학습해 모델 예측률을 향상시키고, 차세대 나라장터 시스템의 지능형 입찰정보 분석에 조달요청 문서 문장 검출



기능을 탑재하여 내부 서비스를 강화할 예정이다. 이외에도 발주기관에서 직접 수정사항을 반영하는 등 사용 편의성에 중점을 둔 기능을 추가하여 조달청과 발주기관 간의 업무를 효율적으로 처리할 수 있도록 할 계획이다. 이러한 조달서비스 기능 구현으로 수요기관의 사전검토 소요시간을 단축시킴에 따라 공공조달을 위한 제안요청서 품질 역시 향상될 것이다.

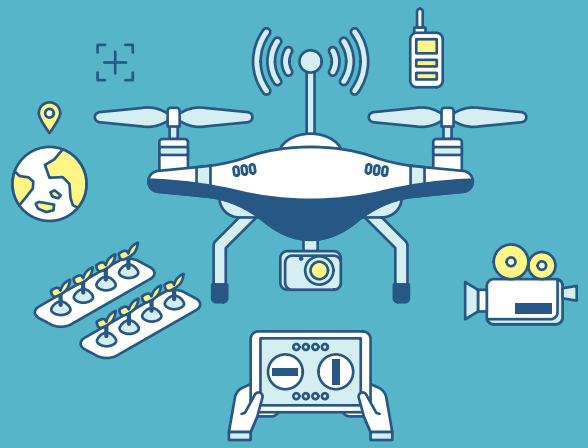
빅데이터 분석, 이렇게 달라졌어요!

- AI에 기반한 사전검토로 수요기관과 조달청 검토/반영 시간 단축 및 조달 품질 향상
- 요청된 RFP에 대해 조달청 업무협의 시간 단축
- 조달서비스 기능 구현으로 수요기관이 사전검토에 활용

교통
물류

국유지 무단점유 꼼짝 마! 국·공유지 모니터링 책임지는 드론이 뜬다

똑똑한 인공지능으로 무단 국·공유지 점유 분석



국가와 지방자치단체가 소유한 국·공유지는 개인이 별도의 허가 없이 무단으로 활용할 수 없다. 그러나 우리나라 국유재산 중 약 6.8%가 무단으로 점유되고 있으며, 이에 대한 변상금 또한 회수되지 못하는 실정이다. 이에 여러 지자체에서는 곳곳에 숨어 있는 국·공유지의 무단 점유 현장을 적발하기 위해 빅데이터 인공지능을 기반으로 한 국·공유지 모니터링 체계 구축 사업을 진행했다.

#국공유지 무단 점거지 관리 #국토영상 자동판독 기술 #Faster R-CNN 알고리즘

모델 개발 전

- 국공유지 면적에 비해 적은 관리 인력
- 국공유지의 무단 점유 및 불법 활용에 대한 단속이 어려움

모델 개발 후

- 국공유지 관리 감독 업무의 효율성 및 효과성 확보
- 국공유지 불법 무단점유지에 대한 관리 업무 효율화

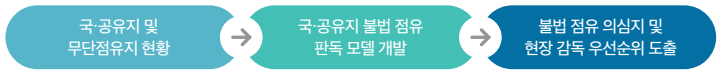
국·공유지 실태조사와 공공서비스 창출의 일등공신 빅데이터

빅데이터 인공지능을 기반으로 국·공유지 무단 점유 의심지를 판독하는 시스템은 국·공유지 항공정사영상과 한국국토정보공사의 고해상도 드론 영상을 인공지능이 학습하도록 해 사람이 관리·감독하기 용이하도록 만든 것이다.

인공지능이 불법 무단 점유하는 토지를 대규모로 조사하고, 고해상도의 영상을 자동판독·분석할 수 있다면 수작업에 의존하던 국·공유지 실태조사가 더 효율적으로 이뤄질 수 있다. 또 개인이 무단 점유한 토지를 효과적으로 파악해 다양한 행정 기초자료로 사용하면 무단 점유 문제 해결은 물론 누락된 세외 수입 확보라는 일거양득이 가능하다.

이러한 필요성에 따라 통영시는 빅데이터 인공지능 기반 국·공유지 모니터링 체계 구축 사업을 추진했다.

통영시는 먼저 'AI 기반 국토영상 자동판독 기술을 활용한 국·공유지 실태조사 체계로 드론과 인공지능을 활용한 공공서비스 창출'이라는 모토 아래 세 가지 목표를 세웠다. 첫째, 항공정사사진(원본 항공사진을 이용하여 지형의 기복 및 카메라로 인한 왜곡 등을 보정한 사진) 자동판독을 위한 학습 데이터를 구축해 AI학습 활용서비스 체계를 확보한다. 둘째, 통영시 전역의 무단 점



유 의심 지역에 드론영상DB를 구축해 국·공유지 실태조사 및 토지 특성 조사 업무의 편익을 높이는 한편 세외수입을 발굴하기 위한 체계를 확립한다. 마지막으로 토지정보 모니터링을 고도화해 토지 특성 자동 판독·분류 시스템을 만들고 이를 통해 인공지능 학습DB를 연계한 신(新) 공공서비스를 지원하는 것이다.

모델 개발의 신뢰성과 효용성, 향후 활용도 등 다각적인 요소를 고려해 전문가들의 자문을 거쳤다. 부동산 및 지적, 사회과학교육과 전문 교수들에게 여러 요인을 분석한 자문을 얻었다. 그 결과 국가표준정사영상 외에 통영시가 별도로 촬영한 영상 데이터를 추가 학습하여 활용할 필요가 있다는 것을 확인했다. 또한 분석참조모델에 근거해 현장조사 우선순위 항목을 지목상 ①도로 ②대지 ③임야순으로 정의했다. 이와 같이 우선순위를 지정하는 이유는 지목상 ①도로는 불법무단점유지가 많고, ②대지는 지목항목 중 가장 땅값이 비싼 항목으로 국가 및 지방자치의 세수확대 확보상 유의미하며, ③임야는 불법개간이 가장 많이 발생하기 때문이다.

학습용 데이터 수집해 토지별 특성 추출

국·공유지 모니터링 체계 구축을 위해 통영시는 2018년 항공사진을 수집해 행정 시스템 내 국·공유지 관리 기능을 추가하고, 이를 통해 AI 영상판독모델 개발 및 학습DB를 구축했다. AI 학습용 데이터로 개발한 영상판독 모델을 활용해 과거에 촬영된 항공 사진과 드론영상을 분석해 보았다. 그 결과 AI는 항공사진을 판독해 토지 특성별 현실지목 여러 종을 추출했다.

객체 식별과 공간화하는 과정 거쳐 분석 결과 수치화

분석 과정에 사용된 원본 이미지는 용량만 2.1GB, 해상도 23,890x28,850



의 대용량으로, 학습 모델로 직접 사용하기에는 어려움이 있었다. 이에 이미지를 학습에 적합한 최적의 사이즈(1024x1024)로 분할하여 활용했고, 손상 및 누락을 최소화 하기 위해 이미지 분할 시 50% 정도의 영역이 겹쳐지도록 했다.

시스템에 업로드된 영상은 AI가 판독을 거쳐 객체별로 위치 정보를 검출하게 된다. 판독 결과(두 이미지 간 유효한 차이가 있음)로 0과 1 사이의 실수로 값이 도출되는데, 임계치인 0.7을 넘는 값을 유효 판독 결과로 사용했다. 이 값을 활용해 전답 정보가 표시된 주제도에서 객체를 식별하고, 공간화하는 과정을 거쳤다.

데이터 베이스의 정확도를 검증하기 위해 진행한 검사에서 건물(지붕형·옥상형), 비닐하우스(백·흑), 도로, 답/전, 하천, 묘지, 주차장 등에 대한 정확도가 평균 80.50으로 나왔다. 높은 정확도에 근거해 통영시는 향후 나머지에 대한 정확도 검증을 진행해 완성도를 더욱 높여나갈 계획이다.

신뢰도 높은 평가 지표와 실제 적용 검증까지 확보

모델 분석 성능 평가를 통해 지표화하는 작업도 진행했다. 인식 모델 평가지표(M.I.O.U)*, 코헨의 카파 계수(Cohen's Kappa)* 등을 통해 모델의 분류 성능이나 평가자의 신뢰도 확보를 평가하는 지표를 추출할 수 있었다. 확보된 영상으로 국·공유지를 비교해 주요지목(도로, 대지, 농지(답/전), 임야 등)별 불법 무단점유지도 확인할 수 있었다.

또 오픈소스 기반 지리정보시스템 무료 공간 시각화 툴(QGIS)을 활용해 지역별 국·공유지 현황과 무단점유지역을 한눈에 확인할 수 있도록 했다. 현장감

독 우선순위별 지역을 표시하고, 불법 무단점유 의심지역의 변화 모습도 필지 단위별로 경계 및 정보 표시를 했다.

* 인식모델평가지표(M.I.O.U: Mean Intersection Over Union): 대표적인 이미지 분석모델 성능 평가 기법 중 하나로 정답 영역과 예측 영역의 겹치는 정도를 수치화한 지표
* 코헨의 카파(Cohen's Kappa): 평가자의 신뢰도를 확보하기 위한 대표적인 평가지표로 주로 범주형 자료 형태의 분석결과에서 사용함. 즉, 어떤 결과에 대한 통계적 신뢰도를 측정하는 통계지표

단축된 프로세스로 신뢰도 높고 신속한 행정 운영

이번 모델은 지목 불일치 토지에 대한 지목 체계 개선 등 부동산 행정정보 자료 정비 추진 시 활용할 수 있는 것은 물론 각 지방자치단체 담당자가 분석참조모델을 활용해 지목 불일치 토지 정비 대상을 쉽게 파악할 수 있다. 토지 소유자 외 이해관계자에게 부동산 정비 내용 확인 절차를 거쳐 정비 대장을 작성해 보고할 수 있다. 무단점유지와 시설 불법 토지형질변경 여부도 한결 수월하게 파악할 수 있게 됐다. 업무 담당자는 분석참조모델을 활용하여 무단점유지역 시설을 확인하고 원상복귀 시정명령과 고발조치를 할 수 있고, 이후 결과에 대한 재확인도 쉽게 진행할 수 있다. 이번 모델을 적용하면 국·공유지 내 불법 무단점유지의 관리감독 강화 및 업무처리 시 지자체별로 연간 1.9억 원, 전국적으로는 연간 422억 원의 업무 비용을 절감할 수 있을 것으로 전망된다.

지자체가 개발해 전국에서 활용가능한 모델로 보급

시스템을 활용해 시계열 분석을 한 결과 아무것도 없던 곳에 건물이 들어선 사례도 다수 확인할 수 있었다. 이처럼 시계열 영상을 통해 변화를 감지한다면 무단으로 점거, 사용되고 있는 국·공유지를 좀 더 효율적으로 관리할 수 있게 될 것이다. 새올, 표준지방세, 세외수입, 부동산종합공부 등의 시스템과 연계하여 국·공유지에 대한 실태 분석에도 활용할 수 있다.

통영시는 향후 분석모델을 적용해 추출된 의심지에 대한 드론 촬영 등 현장조사를 진행하고, 무단 점유지로 확인되면 측량을 통해 세외수입 확보 및 국공유지 관리를 체계화해 나갈 예정이다. 더 나아가 한국국토정보공사와 협업하여 국토조사 및 행정서비스 업무 고도화로 효율성을 증진하고, 다양한 공공데이터와 결합하여 국토센서스, 지목체계 개편 등 공공서비스의 품질 향상 기반도 마련할 계획이다. 이번 모델 구축을 계기로 앞으로 더 많은 토지 관련 담당자가 활용할 수 있도록 모델을 공유하고, 영상정보 처리 기술 등을 통해 다양한 분야에서 공공서비스 모델을 발굴할 수 있을 것으로 기대된다.

빅데이터 분석, 이렇게 달라졌어요!

- ☑ 무단 점거 국공유지 실태조사 및 실태분석 체계 구축
- ☑ 항공정사사진 자동판독을 위한 인공지능 학습DB 구축
- ☑ 드론영상 DB 구축으로 무단 점유 의심지역 조사 업무의 효율화 추구
- ☑ 토지 특성 자동판독/분류 모델을 통한 국토정보 모니터링 고도화
- ☑ 지자체에서 개발해 전국에서 활용가능한 모델 개발



Part 2

교육복지

빅데이터 분석으로 공동주택 돌봄수요 미리 예측! I 복지부·교육부·국토부

공공데이터의 사회적 가치창출로 장애인 이동권에 공헌 I 일반 국민

사각지대 없이, 복지 앞에 모두 평등한 서비스 구축 I 안성시

교육
복지

빅데이터 분석으로 공동주택 돌봄수요 미리 예측!

신규 공동주택 내 초등돌봄 수요예측 모델 개발



초등학생을 키우는 맞벌이 가정의 남편 A씨와 아내 B씨는 번갈아가며 아이를 보살피지만 때때로 생기는 돌봄 공백에 난처한 상황이 생기기 일쑤다. 하지만 새로 입주한 아파트 단지에 다함께돌봄센터가 있어 안심하고 아이들을 맡길 수 있게 되었다. 지자체가 아파트 신축 전 초등학생의 돌봄 수요를 예측하여, 수요에 맞는 다함께돌봄센터를 미리 준비한 덕분이다.

#초등돌봄 #예측분석 #군집분석 #주성분분석 #RandomForest·SVM 알고리즘

모델 개발 전

- 돌봄 수요 예측을 위해 필요한 다양한 정보를 개별 담당자가 모두 확인하기 어려움
- 돌봄 아동 수 예측이 어려워 돌봄시설 설치 여부의 사전 결정 곤란
- 학기 중이나 방학에 따라 달라지는 돌봄수요에 대응하기 위한 적정 돌봄시설 규모 산출 곤란

모델 개발 후

- 체계적인 알고리즘과 매뉴얼로 초보 분석자도 손쉽게 업무에 활용
- 신규 공동주택 준공 이전에 수요를 예측해 적절한 돌봄시설 규모 산정
- 수요에 맞는 온종일 돌봄시설 확충으로 돌봄 사각지대 해소

초등 돌봄 수요, 빠르게 예측하고 손쉽게 활용

국토부는 2021년 1월 신규 아파트 단지의 주민공동시설에 온종일 돌봄시설인 '다함께돌봄센터' 설치를 의무화했다. 그러나 돌봄이 필요한 초등학생 수를 사전에 정확히 예측해 적정 규모의 돌봄시설을 만드는 것은 쉽지 않은 일이었다. 이에 교육부와 복지부, 국토부는 아파트 및 주변 정보와 인구 현황을 고려해 신규주택 내 초등돌봄 수요를 예측할 수 있는 모델을 개발했다. 행정안전부가 청주시, 창원시를 대상으로 개발한 초등돌봄 수요예측 모델을 전국적으로 확산시켜 전국 신규 아파트에 대한 초등돌봄 수요를 예측하고, 이를 토대로 신규 공동주택 준공 이전에 돌봄이 필요한 초등학생 수에 맞는 적절한 돌봄센터 규모를 산정하는 것을 목표로 설정했다.

프로젝트는 크게 세 단계로 진행됐다. ①시군구별 지역 특징을 반영할 수 있는 변수를 발굴하여 전국을 군집화하는 군집분석, ②아파트 단지 단위로 초등돌봄의 수요를 예측하는 예측 모델 개발, ③초등돌봄 업무 담당자가 매뉴

얼을 보고 손쉽게 수요예측 모델을 활용할 수 있도록 모델 배포 및 분석환경을 지원하여 전국으로 확산하는 것이다.

프로젝트는 인터뷰와 자문을 통해 더욱 탄탄하게 보완됐다. 전문가들은 청주 및 창원 지역에 특화된 수요예측 모델을 전국 지자체로 확산할 수 있도록 전국을 특성별로 2~3개의 군집으로 나누고, 각 군집에 맞는 수요예측 모델의 최적화가 필요하다는 의견을 제시했다. 또 모든 지자체가 활용할 수 있도록 전처리 알고리즘을 자동화하는 것이 필요하다고 조언했다.

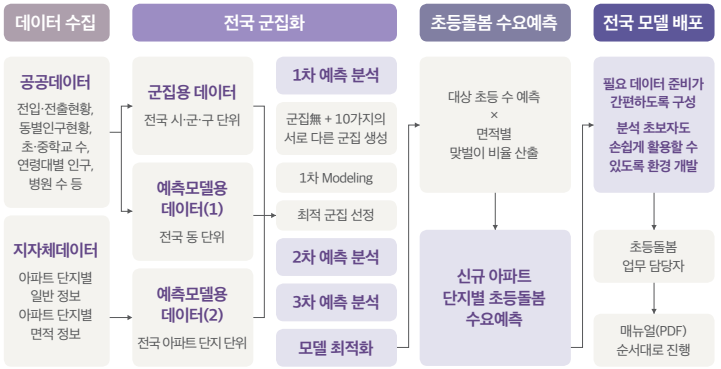
돌봄이 필요한 초등학생 수, 얼마나 될까? 수요예측 위한 데이터 수집

수요예측을 위해 활용된 데이터는 전국 군집화를 위한 데이터, 초등돌봄 수요예측을 위한 데이터로 구분된다. 먼저 전국을 분야별(교육환경, 인구 특성, 공동주택 특성, 생활편의 특성)로 군집화하기 위해 국가통계포털 데이터, 국내인구 이동통계, 공동주택 관리정보시스템, 보건의료 빅데이터 등 총 50개의 변수가 활용됐다. 초등돌봄 수요예측에는 290개의 변수가 활용됐다. 초등학생 수를 예측하기 위한 (아파트 구조)건폐율, 층수, 주차대수 등의 데이터, 맞벌이 비율 산출을 위한 통계청 가계동향조사(19,951건) 데이터 등이다.

분석 활용 데이터

추진단계(2)	목적	데이터 목록
01. 전국 군집화 총 50개 변수 활용	교육환경 기준 군집화	국가통계포털 데이터 초등학교 수, 중학교 수, 유치원 수 등
	인구 특성 기준 군집화	국내인구 이동통계 전출입현황, 동별인구 현황, 연령별 주민등록인구 통계
	공동주택 특성 기준 군집화	공동주택관리정보시스템 공동주택 호수, 주거전용면적, 단지 수
	생활 편의 특성 기준 군집화	보건의료 빅데이터개방 시스템 종합병원 수, 공공데이터포털 사설학원 수, 도서관 수, 은행 수, 보육시설 수 문화기반시설 수
02. 초등돌봄 수요예측 총 290개 변수 활용	초등학생 수 예측	(아파트 구조) 건폐율, 층수, 주차대수, 방 개수, 최고층, 최저층 등 (아파트 환경) 종합병원·대학·마트까지의 거리, 편의시설 유무 등 (인구 통계) 전입·전출입, 연령·세대별 전입·전출자 수, 동별 인구수 등
	맞벌이 비율 산출	통계청 가계동향조사(19,951건)

수요예측 모델 개발 프로세스



3차에 걸친 분석 결과, 실제와 근사한 결과 도출

예측분석은 군집 수를 다양하게 바꿔가며 분석해 예측 정확성이 가장 높은 군집을 선택해야 한다는 전문가들의 자문에 따라 3차에 걸쳐 이루어졌다. 1차 분석에서 군집을 나눴을 때와 나누지 않을 때를 비교하기 위해 전국을 하나의 모델로 예측한 '無군집'과 10가지의 서로 다른 군집 특성(기본, 인구, 교육, 편의, 공동주택, 수요(회귀), 지역구분)을 생성해 비교했다. 그 결과 종합병원, 도서관, 은행, 문화기반시설 등 생활편의 특성이 반영된 '편의 특성' 군집이 가장 예측도가 우수한 것으로 나타났다. 군집 평가는 평균 실제값과 예측값의 오차가 가장 적은 군집을 선정했다.

돌봄 수요를 예측하기 위한 분석에는 다수의 모델(의사결정나무)을 생성하고 평균값을 취하는 Random Forest 알고리즘과 서로 다른 집단 간 거리를 최대화하는 평면을 찾아내 분류하는 SVM(Support Vector Machine) 등의 알고리즘이 사용됐다.

편의 특성으로 세 개의 군집을 구분할 수 있었다. 거주 및 생활 편의성이 높은 도심 특성군집, 편의성이 낮은 비도심 특성군집, 중간 정도인 준도심 특성군집이다. 군집별로 비슷한 시기에 입주를 시작한 세 아파트 단지에 수요예측 모델을 적용한 결과, 실제 초등 수와 근사한 예측 초등 수가 도출됐다.

해당 단지의 아파트 데이터 수집



동수: 10개
층수: 40층
총 세대 수: 1,800세대
준공년월: 2019.00월
학군: OO초등학교
초등학교와의 거리: 380m
전용면적: 18~41평(평균 27평)

준공 전 수요예측('14) 준공 후 실제수요('19)



군집별 최종 수요 예측 모델 및 검증

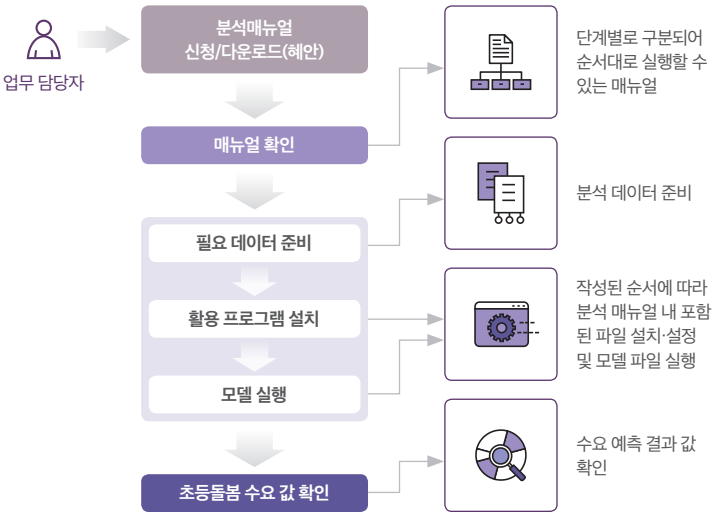
	주요변수	실제 초등 수	예측 초등 수
비도심 특성군집(A)	총 세대 수, 면적별 세대 수, 어린이집 유무, 동 수, 최고층	77명	81명
도심 특성군집(B)	총 세대 수, 면적별 세대 수, 평균 면적, 용적률, 초등학교와의 거리, 동별 6~11세 인구 차이(전년-재작년)	42명	50명
준도심 특성군집(C)	총 세대 수, 면적별 세대 수, 초등학교와의 거리, 동 수, 중학교까지의 거리	187명	190명

분석 초보자도 손쉽게 활용하는 수요예측 모델
현실에 맞는 행정 펼칠 수 있도록 지원

신규주택 내 초등돌봄 수요예측 모델은 매뉴얼, 분석코드, 샘플데이터 등 표준분석모델과 동일한 구성으로 업로드되어 있다. 분석 매뉴얼은 단계별로 구성돼 업무에 익숙치 않은 담당자여도 모델을 활용하면 손쉽게 결과값을 얻을 수 있도록 했으며, 분석 전 검토해야 할 사항을 도입부에 명시하였다.

수요예측 모델을 활용하고자 하는 담당자는 매뉴얼에 따라 필요한 데이터를 준비하고 프로그램을 설치한 후 모델을 실행하면 초등돌봄 수요 값을 확인할 수 있다. 이를 통해 향후 신규 주택단지 내 돌봄센터 규모 산정 시 국민의 돌봄 수요를 충족하고, 돌봄 사각지대를 해소하는 데 큰 도움이 될 것이다.

데이터 기반의 과학 행정 시대의 문을 연 돌봄 수요예측 모델은 돌봄 사각지대를 해소하고 돌봄 수요를 충족하는 데 도움이 될 것이다. 복지부는 지자체에 제공하는 「2021년 다함께돌봄 사업안내」 지침에 수요예측 모델을 활용하도록 명시하여 지자체에서 초등돌봄 수요를 예측, 다함께돌봄센터를 설치·운영하도록 권고했다. 국토부도 2021년 1월 500세대 이상의 신규공동주택



단지 내 다함께돌봄센터를 의무설치하는 「주택건설기준 등에 대한 규정」을 개정했다. 500세대 규모의 신규 공동주택 단지의 경우 25명 내외의 돌봄이 필요한 초등학생이 있다는 결과가 산출되었는데, 이를 위해서는 최소 한 개의 돌봄센터가 필요한 것으로 분석됐기 때문이다.

이처럼 수요예측 모델을 적절히 활용하면 신규 공동주택에 대한 돌봄 수요예측뿐 아니라 관련 규정을 현실에 맞도록 개정해 나가는 데에도 도움이 될 것으로 보인다.

빅데이터 분석, 이렇게 달라졌어요!

- ☑ 각 지자체 돌봄시설 신규 설치 시 가이드라인으로 제공
- ☑ 분석 매뉴얼을 활용해 누구나 손쉽게 수요 예측
- ☑ 지역별 특성에 맞춰 분석모델 변형 활용
- ☑ 학부모 수요에 부합하는 온종일돌봄시설 확충

공공데이터의 사회적 가치창출로 장애인 이동권에 공헌

장애인 이동편의 증진을 위한 새로운 배차 알고리즘



자유롭게 이동할 수 있는 이동권은 모든 국민이 누려야 하는 기본권이다. 하지만 혼자서 이동이 어려운 장애인들은 그동안 장애인 콜택시의 부족한 공급량과 번거로운 이용방법 등으로 이동권을 포기한 채 살아야 했다. 이러한 문제에 주목한 경희대학교 어셈블 팀은 공공데이터 분석으로 장애인 콜택시 배차 알고리즘을 개선하는 아이디어를 내놨다.

#장애인 콜택시 #긴 대기시간&부족한 대수 #CNN 알고리즘 #배차시간 감소 #장애인 이동권 보장

모델 개발 전

- 과도한 대기시간이 발생하는 배차 프로세스
- 수요량을 고려하지 않은 차고지별 배차

모델 개발 후

- 공차 운행 시간 및 최초 이동시간 단축을 통한 대기시간 감소
- 수요예측치를 반영한 시간대별/차고지별 차량 대수 배분

장애인 이동 발목 잡는 장애인 콜택시

우리나라 장애인 콜택시 이용자는 8만 5,000여 명으로, 활동이 불편한 지체 및 뇌병변 장애인과 휠체어를 이용하는 중증 장애인이 대다수다. 그러나 서울시에서 운행되고 있는 장애인 콜택시는 487대에 불과하다. 200명 당 한 대에 불과한 것으로, 이로 인해 장애인들은 그동안 이동에 많은 제약을 받았다. 비장애인이 흔히 사용하는 카카오 택시의 경우 콜을 예약한 후 평균 4분 미만의 대기시간을 보이지만 장애인 콜택시는 평균 60분으로, 무려 한 시간을 기다려야 했다. 또 장애인 콜택시 탑승에 걸리는 시간은 2016년 38분에서 2019년 55분으로 오히려 더 늘어났고, 30분 이내 탑승률도 2016년 51%에서 2019년 28%로 절반가량 줄어들었다.

전국장애인차별철폐연대 박승하 활동가는 “장애인은 지나친 대기시간으로 외출과 귀가에 큰 어려움을 겪고 있다”며 “매우 평범한 일상인 술자리, 데이트 등은 장애인들에게 전혀 평범하지 않은 일”이라고 장애인 이동의 어려움을 토로했다.

기다림과 불편함의 연속인 장애인 콜택시를 이용해야만 하는 이들의 불편함을 해소하고 문제점을 개선하기 위한 방법은 없을까? 경희대학교 어셈블 팀

은 장애인의 이동권과 생활 편의를 위해 장애인 콜택시 공공 데이터에 기반해 '배차 알고리즘 개선 및 앱 UI 구축' 분석을 진행했다. 이를 통해 불필요한 공차시간을 줄이고 주변 수요와 차고지 내 차량 대수도 확인했다. 장애인 콜택시를 보다 효율적으로 운영함으로써 장애인의 이동에 따른 피로감과 불편함을 줄이고자 한 것이다.

전용 앱 UI 개발로 대기시간 줄이는 배차 알고리즘 제안

현재 장애인 콜택시 운영체계는 자동 배차와 수동 배차 두 종류로 운영된다. 사용자가 전화나 문자, 인터넷, 스마트폰 등을 이용해 콜센터에 신청하면 GPS로 차량 위치 정보를 확인해 사용자 가까이 있는 택시가 배차되는 방식이다. 자동배차는 거리, 순서, 대기시간, 치료를 위한 이동 등에 따라 가중치를 부여해 배차된다. 자동배차가 5회 이상 실패할 경우 상담원이 직접 전화로 배차를 하는 수동배차가 이루어진다.

어셈블 팀은 장애인 콜택시 이용에 과도한 대기시간이 발생하는 원인을 공차시간을 불필요하게 야기하는 기존의 배차 알고리즘에서 찾았다. 장애인 콜택시는 주로 차고지 내에서 대기하는 경우가 많아 고객에게서 멀리 떨어져 있기 때문이다. 차량 대수가 부족한 것도 문제였다. 차량 대수가 부족한 문제는 택시회사와의 연계 등을 통해 근본적인 해결책이 필요한 일이지만, 이동시간 단축은 간단한 알고리즘 개선으로 충분히 보완할 수 있는 문제였다.

문제점 도출	• 불필요한 공차 시간 야기하는 배차 알고리즘 • 이동시간 증가
개선사항	• 새로운 배차 알고리즘 제시 • 차고지 내 차량 대수 조정
최종 목표	• 공차 운행 시간 감소 및 드라이버 전용 앱 UI 개발 • 최초 이동시간 감소를 통한 연쇄적 대기시간 감소

대기시간 감소를 위한 CNN기법 수요예측

어셈블 팀은 먼저 서울시의 장애인 콜택시 데이터와 장애인 유형별·등급별 등록현황, 장애인 시설 현황, 장애인 콜택시 차고지 등의 데이터를 활용해 대기시간 감소를 위한 수요예측 모델을 만들었다. 이미지를 분석하고 패턴을 찾는데 유용한 알고리즘인 CNN*기법으로 장애인 콜택시가 보이는 운행패턴을 학습하고 다른 변수로 인해 발생하는 수요의 이상치를 분석했다.

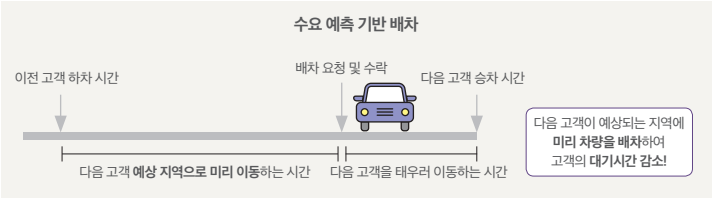
* CNN: 데이터(주로 이미지) 내 숨겨진 패턴이나 특성을 찾는데 유용한 알고리즘

92%대의 정확한 미래 수요예측까지 가능

기존 배차 방식은 이전 고객이 하차한 시점부터 다음 배차 요청을 받을 때까지 콜택시가 길 위에서 보내는 시간에 배차 요청을 수락하고 다음 고객이 있는 지점까지 이동하는 시간이 더해지는 구조다. 하지만 다음 고객이 있을 것으로 예상되는 지역에 미리 차량이 배차돼 있다면 고객의 대기시간을 절반 이상 획기적으로 줄일 수 있다.

이를 위해 어셈블 팀은 2017년~2018년 지역별로 정리된 장애인 콜택시의 시간대별 실수요치를 정리하고, 장애인 콜택시의 미래 수요를 시간대별로 예





측한 모델을 활용해 새로운 배차 알고리즘을 만들었다. 기존 배차 알고리즘인 고객-차량 거리(30%), 접수 순서(30%), 대기시간(30%), 치료 목적 이동(10%)을 수정해 수요예측에도 가중치를 부여한 것. 이로써 서울시 장애인 콜택시는 미래 수요를 시간대별로 92%나 정확하게 분석해낼 수 있었다.

사용자 편의 향상시킨 드라이버 앱 UI도 개발

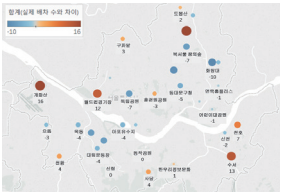
이렇게 만들어진 수요예측 모델에 기반하여 드라이버 앱 UI도 새롭게 개발했다. 보완된 배차 알고리즘에 지도 형식의 직관적인 UI를 더해 사용자의 편의성을 향상시키고 이동 거리를 시각화해 대기시간을 단축시킬 수 있도록 했다. 시간대별 차고지 차량 대수 분석 결과 콜택시 기사가 차고지로 출근함과 동시에 콜이 배정되고, 차고지에서 고객이 있는 곳으로 이동하기까지 약 13분의 시간이 감소되는 결과가 나왔다. 결과적으로 장애인 콜택시의 '첫 번째' 고객이 있는 곳으로 이동하는 시간을 감소시키면 그 다음부터는 연쇄적으로 고객의 대기시간을 감소시키는 효과가 나타났다.



기존 가중치 + 수요가 예측되는 지역에 가중치 부여

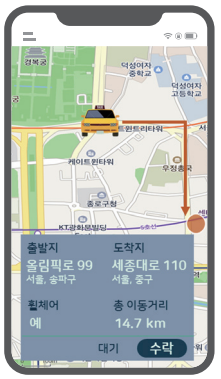
차고지별 배차량과 수요량 시각화로 편리성 극대화

이와 함께 어셈블 팀은 차고지에 대기 중인 차량 대수도 분석했다. 장애인 콜택시 기사는 조별로 시간을 정해 근무하고, 차고지에서 반경 5km 거리의 고객에게 우선 배차된다. 근무 시작 시간대의 차고지 주변 수


차고지별 배차량-수요량 시각화

요를 파악해 시간대별 차고지 차량 대수를 분석하자 차고지별 배차량과 수요량을 한눈에 볼 수 있었다.

차고지별 배차량과 수요량을 시각화해 제공하면 차량 수가 많은 곳에서 적은 곳으로, 배차 수요가 많은 곳에 더 많은 차량을 배차하는 방식으로 조정이 가능해진다.



공공데이터 활용해 장애인도 행복한 사회로 도약

새로운 배차 알고리즘을 통해 장애인 콜택시를 이용하는 장애인은 대기시간과 이동시간 감소라는 결과를 얻을 수 있다. 이는 곧 장애인 이동권의 실현이기도 하다. 새로운 배차 알고리즘은 또한 배차의 효율성을 제공하고 공공 차고지 간 차량 이동을 통해 추가비용 없이 서비스를 개선할 수 있으며, 드라이버에게는 개선된 앱 UI 제공으로 효과적인 의사결정을 지원할 수 있다.

공공데이터로 사회적 가치를 창출한 어셈블 팀의 배차 알고리즘 분석은 앞으로 민간과의 연계를 통해 공급 문제를 완화하고, 날씨·축제·행사 데이터 등을 추가적으로 학습시켜 보다 나은 서비스를 만들 수 있을 것으로 기대된다. 또 고객이 차량의 실시간 위치를 볼 수 있는 고객 전용 앱 UI를 구축해 고객 편의성을 증대시키는 등 다양한 방향으로 발전시켜 교통약자인 장애인 이동권 확보와 이동편의 증진에 기여할 수 있을 것이다.

빅데이터 분석, 이렇게 달라졌어요!

- ☑ 수요예측 모델 구축으로 이동시간 감소 및 장애인 이동권 실현
- ☑ 추가비용 없는 서비스 품질 향상
- ☑ 고객전용 UI 앱 구축으로 공공데이터의 사회적 가치 창출

교육
복지

사각지대 없이, 복지 앞에 모두 평등한 서비스 구축

빅데이터 기반 '안성맞춤형 복지정책' 수립



복지 사각지대에 놓여 위기를 겪는 사람이 없도록 정확한 데이터를 바탕으로 맞춤형 복지 서비스를 제공하는 것은 사회의 마땅한 역할이다. 경기도 안성시는 효율적인 사회복지 전달체계를 구축하고 시민의 복지 요구를 수용하기 위해 빅데이터 기반의 복지행정을 구현했다.

#지역맞춤형 복지정책 #복지욕구해소 #복지공감지도 #선제지원 #복지기관 이용편의성 향상

모델 개발 전

- 복지 사각지대 존재
- 불가피하게 장기간 경제적 빈곤 상황 발생

모델 개발 후

- 시민 복지체감도 향상
- 위기가정 선제지원 가능
- 지속가능한 복지시스템 구축

효과적이고 효율적인 사회복지 전달체계 필요

안성시의 복지예산은 2016년 1,581억 원에서 2018년 1,911억 원으로 지속적인 증가세를 보이고 있다. 그러나 이보다 중요한 것은 효과적이고 효율적으로 복지자원을 배분할 수 있도록 사회복지 전달체계를 개선하는 것이다. 이에 안성시는 복지 서비스 현황을 분석, 복지공감 지도 제작, 위기가정 선제지원, 복지기관 이용 편의성 향상을 모두 추구할 수 있도록 빅데이터 기반 복지행정을 구현하고자 했다. 이를 통해 지자체 맞춤형 복지정책을 수립하고, 복지 서비스 현황 시각화와 복지기관 접근성 분석 및 개선방안을 도출했으며, 집중관리 대상 가정(위기가정 → 기초·차상위) 예측 모델 개발을 통해 위기가정에 대한 선제적 지원 대책을 수립하고자 했다.

정확한 모델 분석을 위해 안성시는 공공 복지대상자 및 지원내역 61,000건(15.7월~18.7월/매년 1월, 7월 현황)과 민간 복지대상자 및 지원내역 63,700건(15.3월~18.8월/42개월분), 위기가정 사후조사보고서를 분석했다. 이와 더불어 안성시 행정구역별 인구 현황(18.7월 기준) 등도 분석했다.

정확한 분석 데이터를 도출하기 위해 안성시는 GBM(그래디언트 부스팅 머신)*알고리즘 기반 예측 모델을 개발하고, 등시선 알고리즘을 활용해 복지기관 접근성 취약지역을 도출했다.

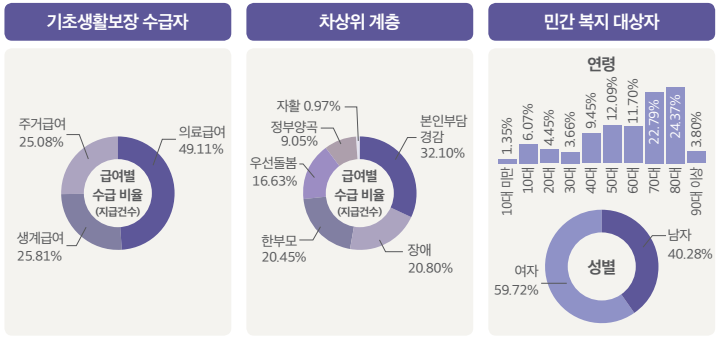
* GBM(그래디언트 부스팅 머신): 오차에 대한 모델링을 연속적으로 수행하여 예측도를 향상시키는 분석 기법

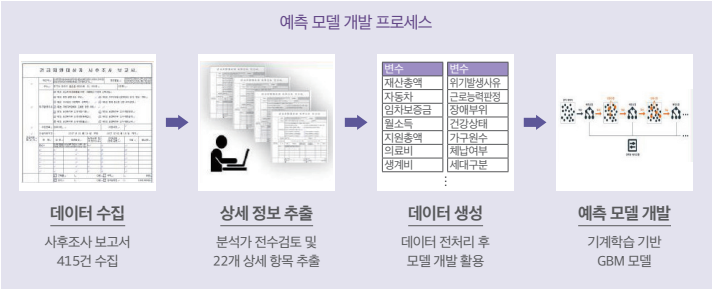
다각적인 접근으로 복지 현황 데이터 수집

안성시가 복지현황을 수집해 분석한 결과 공공복지 서비스의 경우 기초수급자와 차상위 계층 공통으로 의료비 지원 비율이 가장 높았으며, 기초수급자의 42.4%는 주거, 의료, 생계급여를 전부 수급 중인 것으로 확인되었다. 민간 복지대상자는 50대 이상, 기초수급가정이 가장 비율이 높았으며, 지원항목은 밑반찬 등 식품지원이 86.2%로 압도적으로 많았다. 또한 다수의 복지대상자는 복지기관 접근성이 취약했다.

지역별·상황별 예측 결과 토대로 복지 계획 수립

분석 결과 공공, 민간 복지대상자의 47%가 ○○읍 등 특정지역에 집중 거주하고 있었다. 복지기관 접근성이 좋지 않고 대상자가 넓게 분포되어 있는 4개 면 등에는 인력과 자원 재배치 등의 대책이 필요했다. 또한 복지기관 이용

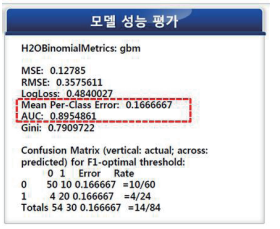




편의성 개선을 통해 복지체감도 향상이 절실하고, 위기가정에 대한 선제지원 필요성도 확인할 수 있었다. 최근 3년간(16년~18년) 지원을 받은 가정 중 약 30%가 수급가정으로 이동했으며, 이로 인한 장기간의 경제적 빈곤상황 방지 대책을 강구해 위기상황을 극복하고 자립할 수 있도록 지원해야 한다는 점을 확인했다.

중요 변수(요인) 목록

분류	변수명	변수설명
지원금액	지원금액(총액)	지원금액 합계
	지원금액(의료비)	의료비 지원금액
재산 및 소득	월소득	월소득 총액
	재산(임차보증금)	부동산 임차보증금
건강	건강상태	질병, 장애 등 4종 구분
	질병부위	질병 대분류 11종
	질병부위(상세)	질병 세분류 54종
세대 및 가족 구성	세대구분	한부모, 노인 등 8종 구분
근로능력	근로능력	근로능력 보유 여부
	근로능력판정 사유	연로, 질병 등 10종 분류
기타	위기발생 사유	12종의 위기발생 사유



지자체의 실정 반영한 현실적이고 효과적인 복지시스템

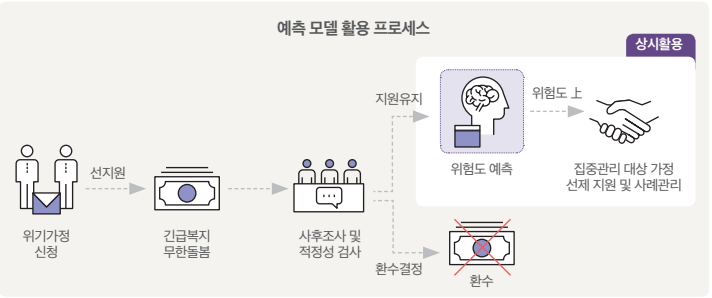
구축된 빅데이터 분석 모델은 세 가지 측면에서 활용성을 가진다. 우선 맞춤형 복지정책의 실현이다. 기초·차상위 대상자에 대한 민간 복지 서비스 지원을 확대하기 위해 정기 소식지 발송, 이동 푸드뱅크 운영 정례화 등 적극적으로 찾아가는 복지 서비스를 확대해 나갈 수 있다.

이와 함께 복지기관 접근성 면에서도 효율적인 복지 행정을 가능하게 한다. 원거리 거주 대상자를 위한 민간 복지기관 추가 설립, 복지셔틀버스 시범운영, 노선 최적화 등을 통해 안성시 복지대상자 중 약 25%의 복지기관 접근성 개선 효과가 나타날 것으로 기대된다.

마지막으로 위기가정 선제지원이다. 안성시 복지시스템에 계층 변동 예측 모델을 탑재하여 계층 변동 위험군을 지속 관리하고 지원 방안을 검토할 수 있다. 위기가정을 방문해 사후조사를 진행한 후 조사된 내용을 복지시스템에 입력하면 기초·차상위 변동 위험을 예측할 수 있기 때문이다. 이는 향후 민간 복지기관과 연계하여 식품지원, 의료 상담, 취업알선 등으로 서비스 분야를 넓혀 더 큰 시너지를 낼 수 있을 것으로 기대된다.

안성맞춤 복지정책으로 신뢰받는 복지행정 실현

안성시는 해당 모델을 활용해 안성맞춤형 복지정책 수립, 복지기관 접근성 향상, 집중관리 가정 예측 모델 상시 활용이라는 계획을 수립했다.



1단계(18년) 분석수행	2단계(19년) 정책적용 및 시범운영	3단계(20년 이후) 성과측정 및 환류
• 과제 발굴 및 분석의뢰 (18년 상반기) • 맞춤형 복지서비스 분석 (18.8~18.12)	• 분석결과 정책적용 (19.1~19.12) • 예측 모델 및 복지셔틀버스 시범운영(19년)	• 성과측정 및 결과의 정책 환류(20.1~) • 빅데이터 기반 정책 혁신을 통해 신뢰받는 복지행정 실현

시민맞춤형 복지정책 수립과 관련해 단기적으로는 이동 푸드뱅크 상시 운영과 방문 의료상담 등 찾아가는 복지 서비스를 강화하고, 복지셔틀버스 시범운행을 통한 보완사항을 도출해 중장기적으로는 복지기관 이용 편의성과 시민 복지체감도 향상을 목표로 삼았다. 이는 효과적인 사회복지전달체계 및 복지정책을 수립하는 데도 도움이 될 것으로 보인다. 또한 위기가정을 선제지원하기 위해 원스톱 복지시스템에 집중관리가정 예측 모델 탑재와 관련 기능을 개발하고, 운영환경의 특성을 반영해 모델을 최적화함으로써 향후 예측모델 상시 운영을 통한 위기가정 선제지원 체계 구축과 복지취약계층 발굴 모델로 발전시켜 나갈 계획이다.

더 나아가 맞춤형 복지 서비스 분석 결과의 정책적용과 시행은 물론 시범운영에 대한 성과 측정과 환류가 가능하도록 해 신뢰받는 복지행정을 실현할 수 있을 전망이다.

빅데이터 분석, 이렇게 달라졌어요!

- ☑ 시민 맞춤형 복지정책 수립 및 위기가정 선제지원
- ☑ 민간 복지기관 추가 설립 및 복지셔틀버스 운행을 통한 복지기관 이용 편의성 향상
- ☑ 원스톱 복지 시스템 탑재를 통한 상시 활용

Part 3

재난안전

빅데이터 기반 해양안전 위험지수 도출로 안전한 바다를! | 해양경찰청

산불 예방과 피해절감 솔루션 소방용수시설 | 소방청

영사민원시스템에 기반해 선제적으로 재외국민 보호 | 외교부

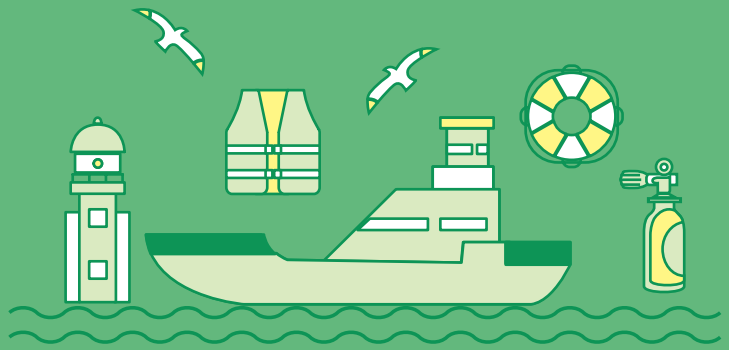
교통 빅데이터 로드맵으로 맞춤형 교통안전지원 모델 개발 | 한국교통안전공단

생명 구하는 일등공신, 응급차량 출동을 책임지는 인공지능! | 대전광역시

재난
안전

빅데이터 기반 해양안전 위험지수 도출로 안전한 바다를!

해양안전 세이프존 확대를 위한 빅데이터 분석



해양경찰청은 행정안전부와 함께 '해양안전 세이프존 확대를 위한 빅데이터 분석'을 실시했다. 목표는 공공 빅데이터를 활용한 데이터 기반 과학적 행정을 통해 연안 위험도를 예측하고, 해역별 맞춤형 안전관리 방안을 마련하는 것이다. 이를 통해 양 기관은 해양안전 가치를 실현하고자 했다.

#해양안전 #세이프존 #안전위험지수 #해양안전지도 #공간자기상관분석 #GIS분석

모델 개발 전

- 광활한 해양지역 특성상 체계화된 데이터 기반의 해양정보 구축 현황 미흡
- 유동인구 등 내·외부 정보를 융합한 연안사고 대응체계의 객관적 근거 필요

모델 개발 후

- 구조거점 파출소 설치와 운영 최적화로 사고 대응시간 개선과 위험도 기반 안전시설물 설치
- 빅데이터 플랫폼 탑재와 지속적 활용으로 대국민 해양안전 서비스 향상과 교육자료로 활용 가능
- 국민 여가활동 패턴을 반영한 해양안전 강화 활동 수행
- 지역별, 시기별 해양안전 활동 강화 및 해양안전지도 구축 사업 추진

해양안전을 위한 데이터 기반 근거 마련

2017년~2019년까지 3년간 연안사고 사망률은 약 17%로 매우 높다. 이는 교통사고 대비 10배에 달하는 수치이다.

해양경찰청은 지난 2020년 '국민이 주인인 정부 실현'을 위해 사회적 가치, 참여와 협력, 신뢰받는 정부라는 3대 전략을 담은 정부혁신 실행계획을 발표했다. 이 계획에 따라 연안사고 30% 감축, 해양안전 가치 실현, 해양안전을 위한 구조·안전 인프라 재정투자 확대 등의 과제들을 발굴·추진해 왔다. 해양경찰청은 이들 주요 과제를 추진하기 위해 데이터를 기반으로 근거를 먼저 확보하기로 했다.

다양한 조건과 상황별 위험지수 도출

해양안전 세이프존 분석은 과제 요구사항 정의, 데이터 수집 및 정제, 공간 데이터 변환 순으로 데이터 전처리가 이루어졌다. 이후 중첩분석을 통한 주요

관리지역 도출, 공간자기상관분석*을 통한 안전지수 도출, 분석 결과 검증 단계를 거쳤다.

분석을 위해서 해양안전 사건/사고, 구조세력 현황, 순찰활동 데이터, 연안이용객(유동인구), 기상데이터 등을 활용했다. 수집된 데이터를 지역별, 시기별로 분석해 해양 사고 현황과 사고 요인을 도출하고, 지역별 사고 건수 및 예측 결과를 활용해 해양사고 위험도를 공간지수화 했다. 또한 해양사고 발생 원인 및 사고다발 지역별 해양안전 위험지수를 도출해 지역별·시기별 맞춤형 해양안전 강화 활동을 수행하도록 했다.

데이터는 해양경찰청이 보유한 해상조난사고 상세 데이터와 구조세력(파출소 및 관할구역, 순찰일지) 데이터가 활용됐고, 개방형 공공 데이터인 기상 데이터, 민간 데이터인 유동인구 데이터가 활용됐다.

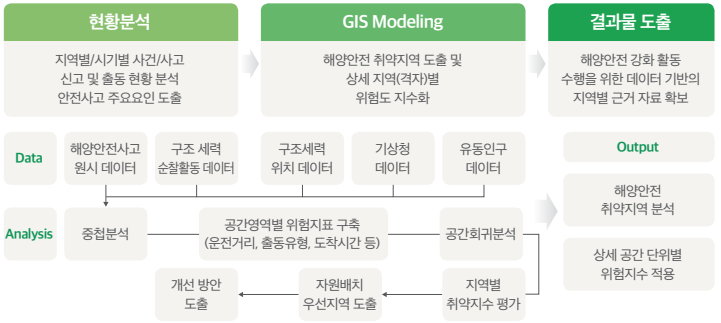
* 공간자기상관분석(spatial autocorrelation): 공간데이터 분석에서 특정 지역의 변수 값이 주변지역과 상관관계를 가질 것이라는 가정을 모델링에 반영하는 분석 기법

해양경찰청 데이터		개방형 공공 데이터	민간 데이터
해상사고 데이터	구조세력 데이터	기상 데이터	유동인구 데이터
• 해상조난사고 상세 데이터	• 파출소 명칭, 위치 및 관할구역 • 순찰일지 데이터	• 해양부이 관측지점데이터 • 해양부이 관측데이터	• 성·연령별 유동인구 • 시간대별 유동인구 • 요일별 유동인구

효율적 분석을 위한 자문회의와 방향 제시

보다 효율적인 분석을 위해 두 차례의 자문회의가 진행되었다. 1차 자문회의에서 전문가들은 분석 방법론은 공간자기상관분석(spatial autocorrelation)을 적용하고, 사고 발생 배경 및 원인 분석을 위해 순찰일지 데이터에 대한 텍스트 분석을 수행하자는 의견을 내놨다. 또한 향후 사고다발 지역 및 위험지수 상위 지역의 어민과 낚시객을 대상으로 안전 교육자료로 활용하는 방안을 검토하고, 해양경찰청과 지자체 공무원의 내부 정책자료로도 활용하자고 제안했다.

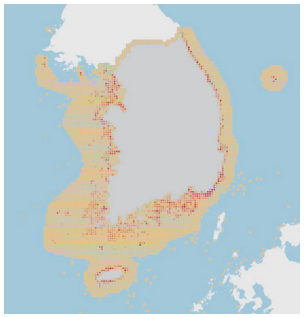
2차 자문회의에서는 전국 해양에 대한 데이터 기반 분석은 기존 현장조사 기반의 연구사례와 차별화되어야 한다는 점과 함께, 분석결과를 현장에 적용하여, 모델 개선 및 피드백 반영이 필요하다는 점을 강조하였다.



지역별·유형별 특성 분석으로 위험지수 시각화

구조세력 순찰활동 데이터, 유동인구 데이터 등 내·외부 데이터를 분석한 결과 해양안전 취약 지역과 상세 공간의 단위별 위험지수를 도출할 수 있었다. 먼저 해양사고 발생 시 위험지수를 도출해 10단계로 구분하여 6단계 이상인 지역을 위험지역으로 선정하였다. 6단계 이상 나온 지역은 해안가나 항포구 등에서 기타익수, 수상레저익수, 항포구차락익수 등이 발생했는데, 이들 사고 발생 상위지역은 전국적으로 발생한 해상 사고 건수의 17.37%를 차지했다.

상세 공간 단위별
해양사고 위험지수 도출



구분	위험지수	격자 수	비중
1	-6.5~0	46	0.1%
2	0	34,658	88.5%
3	0~13	3,568	9.1%
4	13~20	262	0.7%
5	20~35	311	0.8%
6	35~50	126	0.3%
7	50~75	99	0.3%
8	75~100	37	0.1%
9	100~200	34	0.1%
10	200~451.6	9	0.0%
총합계		39,150	100.0%

사고유형별로는 추락익수(물에 잠겨 구조된 상태), 고립, 기타익수, 추락, 표류 등이 가장 많았고, 사고는 항구와 포구, 해안가, 방파제 등에서 주로 발생했다. 사고 사망 시 구조세력이 파악한 인원은 평균 0.14명이며, 레저익수 등의 사고원인은 수중레저 활동인 것으로 비교분석이 가능했다.

지방청(중부·동해·서해·남해·제주)별 특징을 비교했을 때 갯벌, 무인도 사고가 많이 발행하는 지역과 추락이나 익수 등의 사고가 많이 발생하는 지역을 구분하여 파악할 수 있었다.

예측 변수까지 도출해 신뢰성 있는 근거 자료 확보

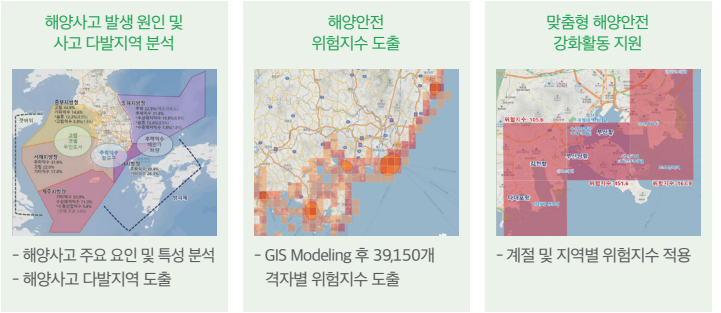
이번 모델을 개발하는 과정에서 격자별 사고건수를 모두 사용하기 위하여 다중회귀분석(Multiple Regresion Analysis)*이 활용됐다. 정확한 예측 결과를 도출하기 위해 약 200여 개의 변수 중 해양사고 변수에 유의미한 영향을 미치는 일평균 유동인구, 안전요원 유무 등 53개 변수를 찾아냈다. 이를 통해 해양사고 발생 시 상황조치의 신속성, 적정성 향상 등에 대한 설치기준을 최적화함으로써 해양안전 사각지대를 해소할 수 있다는 결론을 도출했다.

* 다중회귀분석(Multiple Regresion Analysis): 독립변수가 두 개 이상인 경우를 분석 대상으로 하는 회귀분석

해양안전 강화 활동에 폭넓게 적용 가능

해양안전 위험지수 분석은 해양안전 빅데이터 분석 기반의 선제적 공공서비스를 제공함으로써 해양에서 발생하는 다양한 사고로 인한 사회적 비용을 절감하는 효과를 창출할 수 있다. 해양사고 발생 원인과 사고다발지역을 분석하고, 위험지수를 도출함으로써 각 지역별, 시기별 맞춤형 해양안전 활동을 펼쳐나갈 수 있다.

사고발생 건수, 사망사고 발생 비율 등을 종합적으로 판단해 안전강화 활동을 펼쳐나간다면 해양안전을 강화해 나가는 데도 일조할 수 있을 것으로 기대된다.



해양안전 강화하는 빅데이터 분석

해양경찰청은 이번 분석 결과를 빅데이터 플랫폼에 탑재해 지속적으로 활용해 나간다는 계획을 세웠다. GIS 기반의 해양안전맵을 모니터링해 대국민 해양안전 서비스를 향상시키고, 안전의식을 함양하기 위한 교육자료로 활용하는 한편, 국민의 여가활동 패턴을 반영한 해양안전 위험지수를 도출하는 등 다각적인 해양안전 강화 활동을 수행할 수 있다.

예를 들어, 지역별, 시기별 맞춤형 해양안전 활동을 강화하고, 안전시설물 확충, 연안 위험구역 지정 및 관리, 안전 순찰 강화 및 인력 배치 등 해양안전 서비스의 효율성과 효과성을 극대화할 수 있을 것으로 기대된다. 더불어 해양경찰청은 어민, 낚시객, 관광객이 참여하는 해양안전지도(커뮤니티 맵핑) 구축 사업을 통해 국민과 함께 해양안전지대를 구축하겠다는 계획도 수립했다. 빅데이터를 통해 공공이 앞서 나가고 민간이 함께하는 해양 문화가 만들어지고 있다.

빅데이터 분석, 이렇게 달라졌어요!

- ☑ 빅데이터 분석을 통한 지역별·시기별 맞춤형 해양안전 강화 활동 수행
- ☑ 분석 결과를 기반으로 해양안전 서비스 향상과 해양안전 위험지역 관리 강화
- ☑ 빅데이터 기반의 해양안전 서비스 향상

재난
안전

산불 예방과 피해절감 솔루션 소방용수시설

골든타임 초기 대응에 유용한 소방용수시설 취약지수 개발



대형 산불은 산림과 건축물, 시설물 등에 막대한 피해를 초래한다. 산불 피해를 최소화하기 위해서는 무엇보다 초기 대응이 중요하지만 강원도는 지리적 특성상 소방용수시설을 확보하는 것이 어려웠다. 이에 소방청은 산림 인접 지역의 데이터를 분석하여 소방용수시설 취약지수를 도출, 용수시설물 격차 해소 및 국민 재산권 보호에 기여하게 됐다.

#소방용수시설 #화재진압 #공간데이터 #GIS시각화 #화재취약지수 #산불확산예측

모델 개발 전

- 초기 대응 어려움으로 산불 대형화
- 소방용수시설 저조로 피해 심각

모델 개발 후

- 초기 대응능력 향상으로 대형 화재 방지
- 소방용수시설의 확보 및 효율적 운영 지원
- 재산 및 환경 피해 최소화

산불 진압을 결정하는 초기 대응의 어려움 해소

한번 발생하면 걷잡을 수 없이 번져 큰 피해를 초래하는 산불은 대개 소방용수시설 부족으로 인해 초기 대응이 미흡하여 자연재해에 인재가 더해지는 상황이 빚어진다. 특히 산불로 인한 피해가 극심한 강원도 지역의 경우 상수도관이 설치되지 않은 지역과 산림인접지역에 소방용수시설 설치가 미흡한 실정이다. 소방청은 산불 발생 요인을 활용하여 취약지수를 도출해 소방용수시설이 우선적으로 설치되어야 할 지역을 선정했다. 데이터 분석과 화재 취약지수 모형을 개발해 더욱 원활하게 화재를 진압할 수 있도록 하기 위함이다.

소방청은 우선 현황을 분석한 후 분석 모형 개발에 돌입했다. 화재 발생 현황을 통한 산불 원인 및 발생 지역 파악, 산불 발생 요인과 취약지수를 도출, 소방용수시설 및 상수관로, 토지특성 정보 시각화, 산불 확산 모형 개발 등이 이뤄졌다. 개발된 분석 모형은 테스트를 통해 정확도를 높이고, 피드백과 수정·보완하는 작업을 거쳤다.



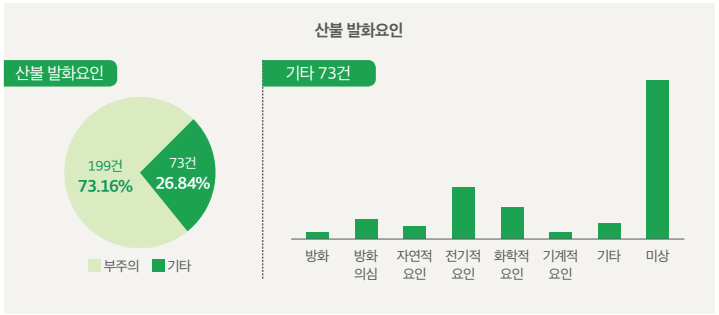
다양한 데이터와 자문의견 확보

원활한 분석을 위해 분석 범위를 동해안 인근 지역으로 한정하고 화재발생현황, 소방용수시설, 상수관로 등 내부 데이터와 도로명 주소 전자지도, 유동인구 등 외부 데이터까지 7개 기관, 10종의 분석 데이터를 확보해 분석을 진행했다. 분석 데이터의 경우 화재발생현황, 소방용수시설, 상수관로, 주거인구, 지역 분류, 건물 정보, 소방서 및 119 응급센터 위치, 유동 인구, 풍향 및 풍속 정보, 경사도와 고도 정보까지 분석해 다양한 화재 상황을 종합적으로 고려했다.

이 과정에서 소방청은 전문가 자문을 통해 시간대별로 지점을 분류하고, 각 변수별로 가중치를 산정하였으며, 건축 연령과 구조를 반영하고, 우선지역 선정 조건을 반영하여 분석을 진행했다.

산불 발생과 관련한 복합적인 요인 분석

분석 결과 총 7,302건의 화재발생 현황 데이터 중 '산불'은 총 272건이었으며, 봄(3~5월)에 화재가 가장 많이 발생했다. 평일보다는 주말에, 흐린 날보다는 맑은 날 화재가 더 많이 발생한다는 특징도 있었다. 시간대별로는 오전 8시를 기점으로 발생 빈도가 점점 높아져 오후 2시, 습도 30%일 때 가장 많은 화재가 발생했다. 또한 산불의 다수가 건조주의보 발효 시 발생했는데 발화 요인은 대다수가 부주의(73.16%)로 인한 것이었고, 전기적 요인, 화학적 요인, 방화 등이 뒤를 이었다. 시군구별로는 도심과 산림 모두 ○○시에서 가



장 많은 화재가 발생했고, 세부적으로는 도심 지역의 경우 ○○시 △△△읍(277건)에서, 산림 지역은 ○○시 △△면(17건)에서 많은 화재가 발생했다.

소방용수시설 설치 우선 지역을 선정하기 위한 조건도 설정했다. 우선 지역 선정 조건은 산림인접지역, 지속적인 건조주의보(10일 이상), 순간 풍속이 강함(최대 풍속 12m/s 이상), 반경 200m 이내 소방용수시설 없음, 가구 수 10호 이상, 건물의 노후화, 산불 빈도가 높은 계절(봄) 등이었다. 이 조건을 기준으로 분석 데이터셋을 구성하고 분석해 설치 우선 지역을 도출했다.

정확한 분석 위해 변수 설정과 가중치 부여

수집된 데이터 중 분석에 필요한 변수를 6가지(유동인구, 주거인구, 기온, 습도, 소방차 도착시간, 소방서와의 거리)로 선정하고, 자문 의견을 반영하여 지형과 시설물적 요인(경사, 고도, 건물 연령, 건물 구조, 도로와의 거리 등)을 변수로 추가했다.

최종적으로 선정된 변수는 기후적 요인(풍속, 습도, 기온), 지형적 요인(경사, 고도, 소방차 출동시간, 건물 연령, 건물 구조, 도로와의 거리), 인과적 요인(유동인구, 주거인구)이었다.

산불은 다양한 발생 요인이 존재하므로 체계적인 가중치 부여가 필요하다는 판단에 따라 소방청은 요인별 변수를 계층화하여 비교한 후 중요도를 산출하는 계층화분석법을 사용해 가중치를 부여했다.



지리정보시스템 통해 분석 결과 시각화

분석 결과, 동해안 인접 지역 중 산불 발생 위험이 높은 곳들은 평균 기온이 높고 습도가 낮아 화재가 발생할 확률이 높았다. 또 유동인구와 주거인구가 평균 이상으로 높은 곳이 화재 발생 빈도가 높았다. 노후화된 건물이 많아 화재 확산 피해가 증가할 것으로 예상되는 지역이기도 했다.

이러한 분석대상 예측 결과 ○○시, △△군, □□시 등 6개 지자체를 소방용수시설 설치 우선지역으로 도출할 수 있었다.

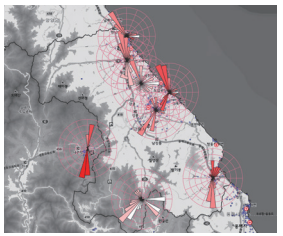
취약지수를 기반으로 우선 설치지역 선정에 활용

취약지수 분석 결과와 계절별 기상 정보를 활용하여 산림인접지역 소방용수시설 설치 우선지역을 시각화하기 위해 '지리정보시스템(QGIS)'을 이용했다. 기존에 설치된 소방용수시설의 최대 반경과 상수관로 및 건물의 위치를 지도에 표시했다. 인근에 산불 발생 지역이 많은 곳은 우선 설치지역으로, 유동인구와 건물이 많고 기존에 화재가 발생했었던 곳은 최우선 설치지역으로 분류했다.

강원도에 인접한 기상관측소에서 측정한 풍향, 풍속데이터를 바탕으로 계절별 바람의 풍속 및 풍향 정보도 시각화했다. 기상관측소에서 측정한 풍향, 풍속 데이터를 바탕으로 시각화한 바람의 평균 풍속과 풍향 정보는 향후 산불이 발생했을 경우 산불 확산을 막고 화재 진압과 주민 대피 등에 사용된다.



소방용수시설 우선 설치 지역 시각화


계절별 풍향 정보 지도 정보 시각화

산불 예방은 물론 다각적인 피해 최소화에 활용

소방청이 분석한 산림인접지역 소방용수시설 현황은 향후 정책을 결정하고 지원하는 데 효과적으로 활용될 수 있다. 단기적으로는 소방용수시설 설치 우선 지역 리스트를 만들어 상수도사업본부와 협업체계를 구축하고, 바람의 방향과 세기를 나타낸 풍향 정보 지도(Wind rose map)를 제공해 지자체에서 화재 피해 저감 대책을 마련하도록 할 수 있다. 장기적으로는 화재 확산 방지 정책 수립과 의사결정 지원, 거주민 대피, 화재 방어구역 지정 등 화재가 발생할 경우 확산 결과를 예측하여 화재가 번지는 것을 방지하고 국민의 생명과 재산을 보호할 수 있을 것이다.

빅데이터 분석, 이렇게 달라졌어요!

- ☑ 소방용수시설 취약지수 도출, 계절별 기상 정보 분석 및 시각화
- ☑ 상수도사업본부와 소방서 협업체계 구축
- ☑ 화재 관련 데이터를 활용하여 산림인접지역에 대한 소방용수시설 취약지수 산출
- ☑ 확산 예측으로 대피로 및 화재방어구역 지정

재난
안전

영사민원시스템에 기반해 선제적으로 재외국민 보호

재외국민 사건사고 위험도 예측 분석



매년 해외여행자 수가 증가함에 따라 해외에서 발생하는 국민 관련 사건사고 건수가 급격히 증가해 지난 2018년에는 약 2만 건에 달했다. 이에 따라 외교부는 재외국민 보호 관련 데이터를 분석, 처리하여 선제적인 재외국민 보호 및 사건사고 대응 방안을 마련하고자 재외국민 위험도 예측 분석을 진행했다.

#기계학습 #재외국민보호 #국가별위험도 예측 #사건사고 선제 대응

모델 개발 전

- 지속적으로 증가하는 재외국민 사건사고
- 신속한 대응과 보호대책 수립 필요

모델 개발 후

- 예측 정보 제공으로 재외국민 보호대책 마련
- 사건사고 현황 시각화
- 사건사고 예방을 위해 해외여행객을 대상으로 맞춤형 안전 문자 제공

다양한 사고 요인에 대한 위험도 분석으로 대응력 향상

해외에서 우리 국민이 사고를 당할 경우 신속하게 대응하기가 쉽지 않다. 나라별 상황이 다르고, 사건사고의 유형도 다양하기 때문이다. 하지만 우리 국민의 해외 체류가 많아지면서 여러 가지 사건사고가 발생하고 있다. 이에 따라 외교부는 연도별·월별, 국가별, 연령별, 사건사고·비자유형별, 신속해외송금 등의 빅데이터를 구축하여 다양한 유형별로 재외국민의 위험도 예측 분석을 진행해 선제적으로 대응함으로써 재외국민을 보호하고자 했다.

외교부는 2016년부터 2018년까지 발생한 5만 건의 재외국민 사건사고 정보, 같은 기간 맞춤형 안전정보 문자 일별 발송 건수, 2019년 기준 국가별 재외국민 통계 데이터를 활용했다. 또한 연도별 청년 실업률, 전체 실업률, 인플레이션 등 경제 데이터와 정치적 안정도를 측정할 수 있는 정치 데이터, 연도별 도로·항공 품질 수준, 연간 관광객 수에 기반한 관광인프라 데이터도 활용했다.

재외국민의 위험도를 예측하고 이를 기반으로 재외국민을 선제적으로 보호하기 위한 모델을 개발하고자 외교부는 여행객, 영주권자 등 단·장기 재외국

재외국민 위험도 예측 분석 절차



* G4K: 재외국민을 위한 통합전자행정시스템

민 통합 위험도를 가늠하고, 중국, 미국 등 전체 사건사고의 약 91.8%를 차지하는 상위 30개국을 대상으로 위험도 예측 분석을 실시했다. 위험도는 천 명당 사건사고 발생 건수를 기준으로 했으며 상위 20%를 고위험국으로 분류했다.

다양한 데이터를 수집·분석하여 재외국민 위험도 예측

다양한 상황을 모두 고려할 수 있는 모델을 구축하고자 외교부는 자체 데이터와 외부 데이터를 종합적으로 수집했다. G4K 데이터, 안전정보 문자 발송 내역 및 국가별 정치·경제 지표를 결합하고, GBM(그래디언트 부스팅 머신) 등 기계학습을 활용하여 익월의 국가별 재외국민 위험도를 예측해냈다. 이때 사건사고 상위 50개국에 한하여 예측을 진행했는데 이는 전체 사건사고 건수의 약 97%에 해당한다. 정치, 경제, 관광, 인프라 등 다양한 분야의 분석 데이터를 수집·분석하여 국가별·월별 재외국민 천 명당 사건사고 건수를 예측했으며, 위험도는 천 명당 사건사고 건수를 5단계로 구간화했다.

정확한 예측을 위해 다양한 분석 알고리즘 활용

분석에는 여러 가지 방법이 동원되었다. 선형회귀, 기계학습, 딥러닝 등 다수의 분석 알고리즘을 활용하여 훈련과 예측을 진행했다. 이러한 방법으로 예측 알고리즘 분석을 통해 예측 모델의 성능을 고도화해 정확도를 높이고 여러 가지 중요 변수를 포함한 분석 결과를 도출할 수 있었다.

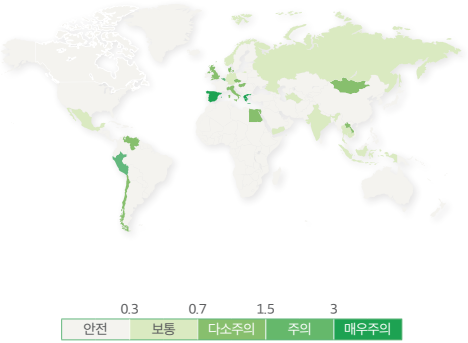
예측 목표 값: 국가별 월별 재외국민 천 명당 사건사고 건수	위험도: 천 명당 사건사고 건수를 5단계 위험도로 구간화
$Tc,m = \frac{\mid icidents \mid c,m}{\mid SMS \mid c,m + \mid Resident \mid c,m} \times 1,000$	0.3 / 0.7 / 1.5 / 3 안전 | 보통 | 다소주의 | 주의 | 매우주의

* Incidents: 사건사고 수, SMS: 로밍문자 발송 건수, Resident: 영주권 등 장기체류자('19년 기준)

일목요연하게 데이터화한 결과 도출

50개국의 정보를 분석해 다음 달을 예측하는 방법으로 위험도를 0.3까지는 안전, 0.7까지는 보통, 1.5까지는 다소 주의, 3까지는 주의, 3 이상이면 매우 주의로 구분했다. 위험도별 연령대 비율을 살펴보면 20~30대의 사건사고 비율이 높을수록 위험도가 높아지는 경향을 확인할 수 있었으며, 2019년의 예측 위험도와 실제 위험도는 거의 일치하는 것으로 나타났다. 2020년 1월의 예측 위험도를 살펴보면 그리스와 스페인은 매우 주의, 페루와 벨기에는 주의로 확인되었다. 그 외에 이탈리아, 체코, 덴마크, 칠레, 몽골, 헝가리, 라오스, 이집트, 영국은 다소 주의로 예측할 수 있었다.

예측 위험도



상위 국가	
국가	위험도
그리스	매우 주의
스페인	매우 주의
페루	주의
벨기에	주의
이탈리아	다소 주의
체코	다소 주의
덴마크	다소 주의
칠레	다소 주의
몽골	다소 주의
헝가리	다소 주의
라오스	다소 주의
이집트	다소 주의
영국	다소 주의

사건사고 위험도 예측을 통해 공관, 본부, 현지 정부, 민간 등 대응 주체별 대응분석과 사건사고 유형별 건수 순위를 집계할 수 있었고, 사건별 조치내용 빈출 키워드, 특정 사건에 대한 소셜 데이터 추이와 키워드를 추출할 수 있었다. 또한 소셜 데이터를 활용한 외교부, 영사 서비스와 관련한 국민 의견도 분석할 수 있었다. 이를 기반으로 국가·지역·시간·사고 유형·사고자 등 다양한 필터로 지도 시각화, 사고 유형과 발생 빈도에 기반한 히트맵, 공관·사고지점·주재국 긴급시설(병원, 소방, 경찰) 등 위치 시각화, 해외 수감자 지역별·사건별 현황 시각화를 구축할 수 있었다.

예측 정보를 활용해 선제적 대책 마련

해외에서 예기치 못한 사건사고가 발생할 경우 외교부에 도움을 청하는 것이 가장 바람직하다. 외교부가 구축한 재외국민 위험도 예측 분석은 이러한 상황에서 유용하게 사용하기 위해서 마련된 것이다.

외교부 시스템에 예측 정보를 미리 제공한다면 재외국민 보호 대책을 훨씬 효과적으로 마련할 수 있을 것이다. 또한 고위험 국가를 대상으로 사전에 정확하게 사건사고 유형을 분석하고 각 유형별로 맞춤형 안전 문자를 제공하는 등 적극적인 재외국민 보호에 활용할 수 있을 것이다.

해외공관의 대국민 사건사고 분석 서비스 제공에 활용

외교부는 이번에 분석한 재외국민 위험도 예측 시스템을 보다 정확하고 유용하게 활용하기 위해 추가적으로 재외국민 위험도 예측 분석 결과를 공유하고 의견을 수렴하는 과정도 진행했다. 무엇보다 분석 품질 향상을 위해 국가별 여행객, 재외국민 데이터를 추가적으로 제공하고 분류체계와 데이터 품질 개선 등이 필요하다는 요구가 있었다. 이에 외교부는 이번 예측 결과의 신뢰성을 높이고 보다 정확한 데이터로 적재적소에서의 활용이 가능하도록 데이터 입력 가이드를 제공하고 분류 체계 개선 등 지속적인 데이터 품질 향상을 추구해 나갈 계획이다.

이러한 데이터들은 여행경보 등급 지정 등과 관련해 해외공관에서 유용하게 사용될 것이다. 빅데이터 플랫폼과 외교부 데이터를 사용한 협업 모델을 구축함으로써 재외국민 보호 강화를 위한 예방/대응 매뉴얼, 대국민 홍보자료, 집행계획 수립에 두루 활용할 수 있을 것으로 전망된다.

빅데이터 분석, 이렇게 달라졌어요!

- ☑ 다양한 상황별 재외국민 위험도 예측
- ☑ 유형별 맞춤형 안전 문자 제공
- ☑ 대국민 사건사고 분석 서비스



교통
물류

교통 빅데이터 로드맵으로 맞춤형 교통안전지원 모델 개발

사업용 자동차를 위한 교통안전관리 컨설팅 시스템



한국교통안전공단은 운수회사의 교통안전 서비스 품질을 높이기 위해 교통 빅데이터를 기반으로 운수회사 맞춤형 컨설팅을 지원하고자 했다. 교통안전 관리를 효과적으로 진행하는 운수안전 컨설팅 지원 시스템을 통해 교통안전을 확보하고, 교통사고 예방을 위한 관리를 체계화함으로써 운수회사 및 종사자의 안전등급을 체계화할 수 있을 것이다.

#교통사고 예방 #도로별 안전도 평가 모델 #LSTM 알고리즘 #운수회사/운수종사자 안전 관리

모델 개발 전

- 공단이 보유한 정보 및 활용도 부족
- 운송회사 사고분석 서비스 및 교통안전관리를 위한 컨설팅 미흡

모델 개발 후

- 교통수단 안전점검에 대한 관리체계 수립
- 다양한 운수안전 정보 메뉴 구성으로 운수안전 통합정보 플랫폼 구축
- 교통 산업 부가가치 창출

풍부한 공단 정보 활용해 운수회사 안전점검 체계 확립

한국교통안전공단은 운수회사, 지자체와의 협업을 통해 교통사고 예방 및 안전관리를 위한 플랫폼이 필요하다는 점을 인식하고 컨설팅 지원 시스템 데이터를 활용하여 다양한 모델을 추가적으로 개발해, 향후 기존 시스템에 탑재할 예정이다. 이를 통해 한국교통안전공단은 운수안전생태계 운영으로 사전 위기관리를 할 수 있고, 운수회사는 자발적인 안전관리 수행을 촉진할 수 있으며, 지자체는 현황 모니터링을 통한 안전관리 업무에 효율화를 추구할 수 있다.

4차 산업혁명 시대에 대응한



교통 빅데이터 센터 로드맵을 만들고자 한국교통안전공단은 1단계로 전사적 마스터 데이터 체계를 수립하고 데이터 통합 시스템을 완성했다. 2단계로 데이터 통합 대상 확대 및 활용을 통해 빅데이터 처리 플랫폼을 확장했다. 마지막으로 3단계를 거치면서 데이터 개방을 위한 플랫폼 구축과 교통산업 부가가치 창출, 서비스 통폐합을 위한 효율성 제고라는 목표를 달성했다.

공단이 보유한 기존 보유 시스템을 활용했고, 운수안전 컨설팅을 위한 추가적인 컨설팅 활동을 펼친 것이다. 시스템이 구축되면 업무 생산성을 증대하는 것은 물론 새로운 관점의 교통안전 정보를 제공할 수 있게 된다. 또한 운수업체 맞춤형 컨설팅으로 성과관리가 가능하고, 서비스 품질을 높여 상향평준화를 이룰 것으로 기대된다.

운수회사의 안전 상황 관련 복합적인 데이터 수집

운수안전 컨설팅 지원 시스템을 구축하기 위해 다양한 데이터를 수집하는 노력도 아끼지 않았다. 운수회사와 관련해 상세한 정보를 수집했는데 운수종사자 정보, 취업예정자 정보, 위험운수회사 현황과 자동차검사 현황을 포함한 내용으로 구성했다. 컨설팅을 위해 실적과 관련된 데이터도 수집했다. 주별·월별 실적과 안정정보, 교통사고 등록현황 등도 조회했다. 이밖에 교통수단 안전점검과 교통안전관리규정 외에 관리자 기능 데이터를 수집했다.

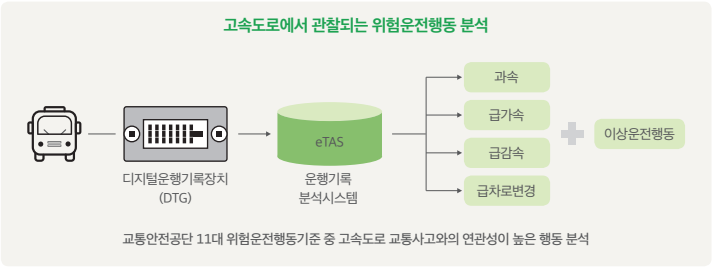
사업용 자동차 안전정보도 수집해 면허대수는 약 3만 5천 건, 등록대수는 5만 5천 건이라는 집계를 내고, 이를 토대로 특이 운전 행태(전체 DTG)들을 개별단위로 다양하게 도출(급브레이크, 급가속 등)했다. 이중 군집화를 통해 특이 운전 패턴을 정의(차선 변경 多, 급가속 및 제동 5회 이상 등)했으며, 사고와의 상관성을 분석해 사고가 날 확률이 높은 위험 운전 행태를 찾았다.

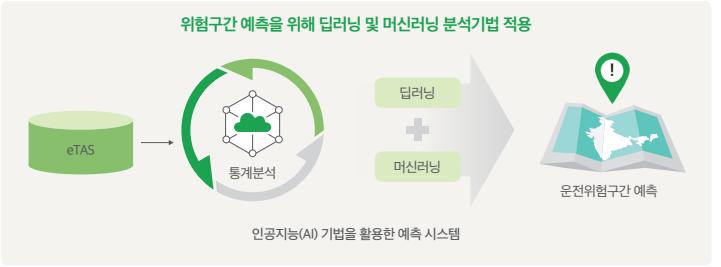


딥러닝과 머신러닝 분석으로 종합적인 도로안전 평가

한국교통안전공단은 다각도로 교통안전도 분석을 진행했다. 도로별 안전 등급을 산정하기 위해 종합적인 모델 구축에 앞서 세 가지 평가 모델을 만들었다. 운수회사 안전성능함수 모델과 운수종사자 안전등급 산정 모델, 위험운전행동과 사고의 상관관계 분석 모델을 적용해 운수회사와 종사자별 운행안전등급 평가 모델로 구축했다.

교통사고 발생 가능성이 높은 고속도로 구간을 예측하기 위하여 광역버스 DTG를 활용해 경부고속도로에서의 운전 위험구간도 분석했다. 여기에는 고속도로에서 관찰되는 위험운전행동 분석과 위험구간 예측을 위해 딥러닝 및 머신러닝 분석기법인 LSTM(Long Short-Term Memory model)을 적용했다.





운수회사의 안전등급 산정과 위험도 지표 추출

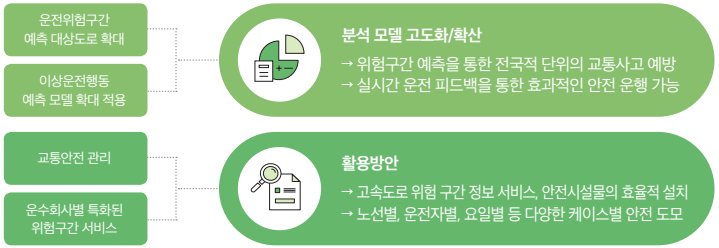
한국교통안전공단의 운수안전 컨설팅 지원 시스템을 활용한 사업용 자동차 교통안전 관리 모델을 통해 운수회사와 종사자, 도로의 위험도를 산출할 수 있었다. 도로별 안전 등급 평가 모델을 통해서는 한국교통안전공단의 자료를 활용 및 분석해 도로 안전등급을 제시하고 산정하는 방법으로 종합적인 도로 안전 평가를 진행할 수 있었다. 사고 빈도와 총 주행거리에 근거해 산출한 운수회사 안전성능함수(SPF)를 활용해서는 A등급에서 D등급, 그리고 산출불가라는 운수회사의 안전등급을 산정했다. 운전자의 여러 가지 속성을 반영해 안전등급도 산정했다. 모델 내에 있는 지리정보시스템을 활용해 도로안전정보와 위험지점, 사고지점, 위반차량에 대한 분석도 할 수 있었다.

교통수단 안전점검 통합 서비스 제공

한국교통안전공단은 이번 모델 개발로 얻은 결과에 운전안전 컨설팅 관련 정보를 담아 인포그래픽 형태로 제공한다. 한눈에 보기 좋게 정리한 자료를 통해 활용도를 높이기 위함이다. 인포그래픽에는 성과관리, 종합 현황, 운수회사 컨설팅 정보, 위험차량 분석, GIS 기반 도로안전정보를 담았다. 운수회사가 보다 안전한 운영을 하는 데 실질적인 도움이 되는 모든 분석 결과를 일목요연하게 정리할 수 있도록 한 것이다. 이와 더불어 해당 서비스가 제공하는 다양한 기능이 조화롭게 운영되도록 연결했다. 한국교통안전공단과 연구기관, 운수회사가 협업해 제공하는 운수안전 통합 정보 서비스는 교통수단 안전점검 전체 프로세스에 편리하게 적용할 수 있다.

운수회사 전용 포털 서비스 개발

무엇보다 중요한 것은 운수회사의 사용편의성을 도모해 폭넓은 영역에서 빈번하게 사용하는 것이다. 이를 위해 한국교통안전공단에서는 운수회사를 위한 전용 서비스도 개발했다. 운수회사가 자체 진단과 안전관리를 촉진할 수 있도록 하기 위해서다. 해당 서비스는 운수회사별 안전등급과 안전관련 정보



등의 운수회사 현황 정보, 운행 정보, DTG 제출 정보와 위험 운전 정보, 의무휴식 정보 등을 제공한다.

운수회사와 공단 간 교육자료 및 공지사항 소통채널도 지원한다. 원활하고 정확한 소통이 가능한 1:1 채널로 교육자료와 공지사항을 전달하고 운수회사는 피드백 기능을 적용해 공단에서 제공하는 교육자료에 대한 실효성을 높일 수 있다. 마지막으로 메일링과 SMS를 통한 정보 제공도 지원한다. 교통수단 안전점검 실시 여부와 운수회사의 종합현황을 담당자에게 정기·수시로 제공할 수 있다.

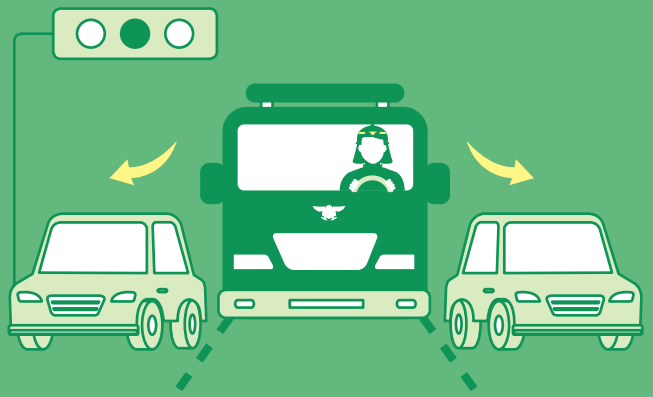
빅데이터 분석, 이렇게 달라졌어요!

- ☑ 운수안전 통합 정보 서비스 제공
- ☑ 교통수단 안전점검 활동 관리를 위한 최적화 서비스 제공
- ☑ 운수안전 컨설팅에 필요한 다양한 정보를 인포그래픽 형태로 제공
- ☑ 운수회사 전용 포털 서비스 개발

교통
물류

생명 구하는 일등공신,
응급차량 출동을 책임지는 인공지능!

긴급차량 출동지연 요인 분석으로 지키는 골든타임



소방차와 구급차는 사고현장에 도착하기까지의 골든타임이 중요하다. 하지만 대전시가 분석한 결과 평균 출동 커버리지(소방서와 안전센터가 5분 이내 출동할 수 있는 지역의 비율)는 소방차 20%, 구급차 31%로 매우 저조했다. 이에 대전시는 구급차의 평균속도를 기반으로 한 최단 경로 알고리즘을 활용하여 긴급상황 출동 효율화 및 화재 등 취약지역의 안전조치를 강화했다.

#로지스틱스회귀분석 #기계학습 #출동커버리지 #출동시간최적화 #골든타임

모델 개발 전

- 긴급차량 상습지연구간 발생
- 저조한 출동 커버리지
- 안전취약지역 발생

모델 개발 후

- 긴급차량 출동 시간 단축
- 취약지역 및 취약시간대 도출
- 상습정체구간에 대한 사전 대응

교통여건에 영향을 미치는 요인에 대한 치밀한 분석

국민의 안타까움을 사고 있는 대형 재난사고는 수많은 인명과 재산피해를 동반한다. 무엇보다 중요한 것은 초동대응인데 신속한 출동이 가장 중요한 역할을 한다. 하지만 증가하는 교통량과 도로, 골목길을 가득 메운 불법 주정차 등은 긴급차량의 출동을 가로막는 요인들이 되고 있다. 이러한 문제점을 해결하기 위해 중앙정부와 대전시가 손을 잡았다.

대전시는 긴급차량의 출동 시 지연되는 구간 및 출동 방해요인을 제거하기 위해 소방차와 구급차의 위치정보(GPS)를 분석하여 교통정책 개선에 활용하고자 했다. 소방차와 구급차가 출동시간을 단축하면 생명과 재산권을 보호하는 긍정적인 결과를 도출할 수 있다. 이에 대전시 소방본부에서 제공한 2016년 8월부터 2017년 1월까지에 해당되는 데이터를 분석했다. 구급차량 GPS 운행정보 3천만 건을 분석해 10초 간격으로 좌표를 생성하고, 재난별 출동차량 정보 210건에 대한 차량 무선호출 번호, 관할서, 센터, 차종 정보를 분석했다. 보다 정확하고 신뢰도 있는 결과를 도출하기 위해 지능형교통체계관리시스템(nodelink.its.go.kr)에 대한 분석도 진행했다. 도로 위치정보 11만 건을

비롯해 도로 특성 정보 9천 건을 도로명, 도로ID, 도로번호, 차선정보 등으로 분류해 분석했다. 이렇게 소방차와 구급차 출동에 영향을 미칠 수 있는 다양한 요인들을 빠짐없이 정확하게 분석하고자 다양한 알고리즘을 적용했다.

긴급차량 출동 최적화를 위한 동선 분석

기계학습으로 데이터를 분석한 결과, 긴급자동차가 5분 이내 출동하기 어려운 취약지역 7곳과 상습 지연구간 8백여 곳을 찾아낼 수 있었다. 또한 취약지역으로 신속하게 출동하기 위해 지연구간을 피해가는 최적경로도 도출할 수 있었다. 데이터 분석 결과 직선거리에 있는 센터보다는 최단경로에 위치한 센터에서 현장으로 출동하는 것이 효율적인 것으로 확인되었다. 예를 들어 용전사거리의 경우 직선센터는 가양센터이지만 최단경로는 삼성센터이고, 대전역은 직선거리센터에는 원동센터가 있지만 최단경로센터는 삼성센터인 것으로 나타났다. 이러한 자료를 토대로 신속한 출동이 가능한 센터를 지정할 수 있어 현장으로 빠르게 이동해 효율성을 높일 수 있게 되었다.



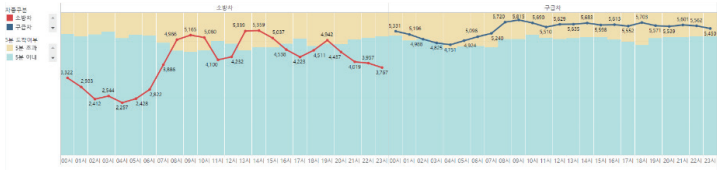
응급차량 최단경로 구축 예시

시간대별 커버리지 분석으로 취약지역 파악

다각적인 분석과 접근을 통해 추출한 결과를 살펴보면 소방차의 시간대별 커버리지는 매우 상이한 것으로 나타났다. 교통량이 적은 오전 0시부터 7시까지는 60%에 달하지만, 상대적으로 교통혼잡이 발생하는 오후 1시에서 2시 사이에는 10%에 불과한 것으로 분석되었다. 특히 10%의 저조한 커버리지를 보이는 지역은 ○○구 △△지역과 ○○구 △△동 등으로 오전 9시에서 10시 사이, 오후 1시부터 3시 사이의 취약시간대에 대해서는 적절한 대응방안을 모색할 필요가 있다는 것을 확인할 수 있었다.

함께 자료를 분석한 구급차의 커버리지 취약시간대는 오전 8시부터 10시로 나타났는데 이때는 불과 25%의 비율을 보였다. 가장 심각한 취약지역은 소방차와 마찬가지로 ○○구 △△지역으로 확인되었으며, 그 외에 중구의 남쪽 지역 및 동구 지역 등을 비롯한 7개 지역으로 분석되었다.

해당 시간대와 지역에 대해서는 정확한 데이터 기반의 출동 효율화를 꾀하고 화재 등 취약지역에 대한 안전조치가 강화되어야 할 필요가 있다는 결론을 내릴 수 있다. 이에 대한 정확한 방안을 강구하기 위해 대전시는 추가적인 분석을 진행했다. 구급차의 평균속도를 기반으로 최단경로 알고리즘을 적용하여 긴급상황 시 출동센터 배정 및 지연경로 회피 등의 시뮬레이션을 수행한 것이다. 그 결과 두 가지 조건을 적용했을 경우 커버리지 비율이 31%에서 60%로 거의 두 배 가까이 상승하는 것을 확인할 수 있었다.


시간대별 5분 이내 도착셀수 및 비율

무엇보다 빠르게 현장에 도착하는 것이 중요한 긴급차량. 이에 대한 정확하고 효과적인 방안을 강구하기 위해 확인한 대전시의 응급구급차 평균 속도는 24km, 소방차는 20km에 불과했다. 특히 소방차의 평일 주간 시간대 평균 속도는 18.8km에 그치는 것으로 확인되었다.

로지스틱 회귀 기계학습 적용해 긴급차량의 상습 지연구간 파악

데이터를 기반으로 응급차량의 상습 지연구간도 판단해 보았다. 지역별 평균 속도와 출동 건수 등 상습 지연 구간의 주요 변수를 파악해 로지스틱 회귀 기계학습을 통해 지연구간에 대한 등급 판단 모델을 적용했다. 평가 기준

은 A~B는 비상습 지연 구간, C는 주의 구간, D~E는 상습지연구간으로 분류했다. 기계학습 모델 판단 결과 ○○동 △△아파트 주변과 ○○시장 일대 등 무려 800여 곳의 상습지연구간이 있는 것으로 판단되었다. 또한, 대전지역 내 병원급 의료시설 86개소 중 5분 이내 소방차 출동이 가능한 곳은 12개소(14%)인 것으로 나타났다. 하지만 상습지연구간에 대한 해결법은 복잡하지 않다. 해당 지역에 대해서는 출동 GPS 데이터만 제공되면 기계학습을 통해 상습지연구간을 빠르고 정확하게 판단할 수 있다. 그를 통해 상습지연구간에 대한 지속 관리가 가능하다.

취약지역 안전조치 강화

대전시는 분석 결과를 활용해 긴급차량 출동 시스템을 개선하고 취약지역에 대한 안전조치를 강화했다. 기존 직선거리 기반 출동 배정 시스템에 최단거리 모델링 결과 적용 방안을 세우고 출동센터 인근 도로에는 좌회전 신호를 설치하는 등 지연요인을 적극적으로 제거할 수 있는 방안을 마련했다. 이와 함께 이동용 소방펌프를 구비하며, 병원급 의료시설에는 스프링클러를 설치하는 등 취약지역 내 안전조치에 심혈을 기울일 예정이다.

빅데이터 분석 결과에 근거해 체계적이고 구체적인 계획들을 실행하고 실질적으로 긴급차량의 빠른 출동을 가능하게 함으로써 전체적인 사회 안전망 구축을 꾀할 수 있을 것이다.

빅데이터로 국민의 생명과 재산을 지키는 모범사례

대전시가 수행한 이번 과제를 통해 얻은 가장 큰 수확은 빅데이터를 활용한 데이터 기반의 의사결정으로 긴급차량의 출동시간을 획기적으로 단축할 수 있는 가능성을 보았다는 점이다. 대전시는 빅데이터 분석결과를 바탕으로 소방차가 재난 현장에 신속하게 도착할 수 있도록 '소방차량 골든타임 확보계획'을 마련할 계획이다.

구체적인 내용을 살펴보면 상습지연구간의 주택과 상업지역에 거주하는 주민들이 화재가 발생했을 경우 우선 초기에 진압할 수 있도록 비상 소화 장치를 설치하는 것이다. 더불어 119 안전센터의 배정과 최적경로 빅데이터 분석결과를 반영해 국민의 생명과 재산을 지킬 수 있도록 신속한 출동여건을 확보할 계획이다. 이는 향후 중앙부처와 다양하고 지속적인 협력을 통해 국민생활과 밀접한 사회현안을 해결하는 모범사례로 활용될 것이다.

빅데이터 분석, 이렇게 달라졌어요!

- ☑ 데이터 기반 응급 출동센터 배정 및 경로 설정 등 효율화 가능
- ☑ 데이터 기반으로 취약지역 및 시간대를 빠르게 판단하여 지속 관리
- ☑ 로지스틱 회귀 기계학습 알고리즘을 활용하여 상습 정체 구간을 판단하여 시각화



Part 4

보건의료

자동차 보험금 누수, 이상패턴 감지해 과학적으로 막는다 I 건강보험심사평가원

도서지역 건강권을 보장하는 응급시스템 스텝업 I 인천광역시

맞춤형 치매 진단 검사로 다각적인 치매 관리 I 김해시

보건
의료

자동차 보험금 누수, 이상패턴 감지해 과학적으로 막는다

교통사고 환자 이상패턴 탐지 모델 개발



자동차 사용량이 증가함에 따라 교통사고 발생 건수도 늘어나고, 이에 비례해 자동차 보험금 청구도 많아지고 있다. 건강보험심사평가원은 보험금 누수를 방지하고, 과학적 심사와 관리체계를 마련하고자 탐지 및 예측 모델 개발에 나섰다. 자동차 보험 진료수가 청구내역 등 빅데이터를 활용하여 교통사고 다발생 환자와 의료기관에 대한 정보를 구축하고, 이를 적극적으로 활용해 자동차 보험금 누수를 막고 투명성을 높이기 위한 것이다.

#자동차 보험사기 #보험 청구 심사 #군집분석 #이상 패턴 분석 #고가도·장기도 지표

모델 개발 전

- 지능화 및 조직화되는 보험사기 단속에 한계
- 보험과 관련된 불필요한 진료가 다수

모델 개발 후

- 사고유발, 과다입원, 허위청구 등의 정보를 사전에 제공
- 교통사고 다발생 환자 및 의료기관 이력관리 가능
- 보험금 누수 방지를 위한 기반 마련 및 기관 간 연계 추진

모두의 부담 가중시키는 보험금 부당청구

자동차를 소유하고 운행하는 사람은 누구나 의무적으로 자동차 보험에 가입해야 한다. 매년 갱신해야 하는 자동차 보험은 사고 발생 시 유용하게 사용할 수 있는 안전장치이지만, 크고 작은 사고가 발생했을 경우 보험금이 올라간다는 부담이 있다. 게다가 자동차보험 편취 목적의 교통사고 유발 등 지능화·조직화 추세를 보이는 사례가 늘어나고, 환자의 도덕적 해이로 인한 과도한 보상 유발로 보험금 상승도 초래되고 있다. 이러한 현실에 따라 자동차보험 진료비 청구 내역을 기반으로 교통사고 환자 및 의료기관 관리 체제 구축에 대한 필요성이 높아지고 있다.

이에 건강보험심사평가원은 손해보험협회, 한국신용정보원, 금융감독원과의 인터뷰를 통해 분석 요인을 도출하고, 이들 기관과 협력해 자동차보험 진료수가 청구 내역 등 데이터를 활용한 교통사고 다발생 이상패턴 탐지 모델 개발에 나섰다. 건강보험심사평가원의 자동차보험 진료 정보 등을 활용하여 사고유발·과다입원·허위청구 등 교통사고 다발생 환자와 의료기관에 대한 관리체계를 마련하고, 데이터 분석을 통해 자동차보험 청구 심사업무를 과학화하는 동시에 교통사고 환자 및 의료기관의 이력관리를 통해 보험금 누수를

방지하기 위해서다. 해당 프로젝트를 진행하기 위해서 건강보험심사평가원은 메타데이터로 데이터 품질을 검증했다. 뿐만 아니라 방대한 자료를 분석해 타당성을 따져 분석모델을 개발했다.

심평원의 축적된 자료가 이상적인 분석의 저력 견인

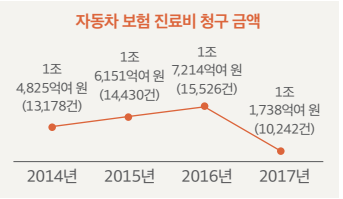
교통사고 다발생 이상패턴 탐지 모델은 방대한 데이터를 수집하고 분석하는 것이 우선적으로 필요했다. 가장 먼저 자동차보험심사시스템의 일반명세서, 진료내역, 상병내역, 처방내역, 의료기관, 사고환자에 대한 데이터들을 수집했다.

건강보험심사평가원에서는 2013년 7월 이후 건강보험심사평가원의 자동차보험 진료비 위탁 심사 수행에 따른 축적된 자료를 활용하기로 했다. 이를 위해 2014년 1월에서 2017년 6월까지의 무려 5,300만여 건의 데이터를 수집했다. 연도별로 자동차보험 진료비 청구내역을 살펴보면 2014년이래로 해마다 청구 규모가 증가하는 것을 확인할 수 있었다.

분석품질의 향상을 위해 해당분야의 전문가를 중심으로 자문단을 구성했다. 자문단은 자동차보험 관련 보험사기 개연성이 높은 행위를 적발하기 위해 심평원의 자동차보험 진료내역 투약정보와 세부 진료정보를 활용하는 것이 의미 있을 것이라는 의견을 제안하였다. 또 분석 시 병원의 행위특성에 따라 변수들의 중심점이 달라지므로 이를 분류하는 데 군집분석 기법을 적용하는 것이 적절하며, 이상지수를 생성할 때 누적 사고횟수 변수가 가장 중요하므로



이 변수에 높은 가중치를 두는 방법을 권고했다. 이상점수에 대한 임계치를 여러 번 시도해서 민감도(Sensitivity)를 높이는 것도 필요하다고 지적했다.



현실성 있는 데이터를 도출할 수 있는 다양한 분석 적용

정확한 분석을 위해 건강보험심사평가원은 아래와 같은 다양한 데이터를 분석했다.

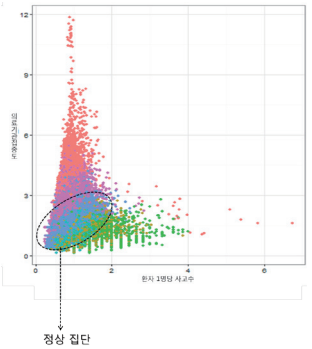
- 교통사고 관련 각 단계(사고접수, 조사, 종결 등)별 데이터
- 고객, 차량, 의료기관별 사고 및 진료 특성
- 인당치료비, 의사면허, 질병 정보, 비급여 항목
- 네트워크 의료기관, 비급여 항목, 경상환자 입원 비율
- 입원환자 부재율, 공휴일입원비율, 인당입원치료비, 일당입원비
- 특수검사비 비율, 합의일, 14일 이전 합의율, 한방비급여진료

이와 더불어 통계모형 및 룰 기반 분석, 고가도지표*·장기도지표* 순위와 추세 또는 평균기준치로 이상징후 감지, 의료기관 지역별/종별 분석과 사고 다발생 환자 단위 분석까지 진행했다.

현장의 목소리를 충분히 듣고 반영하는 노력도 아끼지 않았다. 유관기관 담당자들의 인터뷰를 거쳐 사고환자, 사고유형, 의료기관 측면에서 분석 방향을 분류했다. 이를 통해 ①사고환자의 성별, 연령, 교통사고 특성 분석 ②사고-환자 연관 특성 분석 ③의료기관 시설, 장비, 진료 행태 등의 특성 분석을 진행했다. 보험사기 관련 의료기관, 사고환자, 사고유형 관점의 이상패턴 분석과 통계모형 기반 외 최신 분석 기술 적용이 필요하다는 의견과, 최근 한방 진료가 증가해 이에 대한 패턴과 경미한 사고 현장에 대한 분석도 필요하다는 의견을 반영했다.

* 고가도지표: 진료비를 타 의료기관과 비교해 높은 지표
* 장기도지표: 입원일수가 타 의료기관과 비교해 긴 정도

이상패턴 놓치지 않고 감지·분석해 꼼꼼한 결과 도출



데이터를 분석한 결과 전국 의료기관은 총 13만 4,135개소이며 이 중에서 약 72.6%는 의원급(의원, 한의원)이고, 의료기관은 의원, 치과의원, 한의원, 병원, 한방병원, 종합병원, 치과병원, 종합전문병원 순으로 빈도수가 높은 것으로 나타났다. 2014년 1월부터 2017년 6월까지 교통사고로 의료기관에서 진료받은 사고환자는 약 530만 명이며, 이 중 누적사고 횟수가 4회 이상인 사고환자 약 5만 명을 분석 대상으로 한정한 결과 남성이 42,343명(85.2%), 여성이 7,358명(14.8%)인 것으로 확인되었다. 이는 다발성의 경우 남성 환자가 여성 환자보다 약 70.39%p 높은 비율을 차지하는 것으로 교통사고 전체에 있어 남성 환자가 여성 환자보다 약 13.9%p 높은 비율을 차지하는 것과 비교했을 때 현격한 차이를 보였다. 또한 2014년 1월부터 2017년 6월까지 의료기관에서 교통사고로 진료받은 환자 중 거의 대부분의 환자는 진료횟수가 1회 이상 3회 미만인 것에 비해 분석 대상 사고환자 중 약 60%는 4회이며, 10회 이상 환자도 약 3.5%에 달했다.

'k-means clustering'이라는 군집화 알고리즘을 사용하여, 사고환자들을 유사한 특징의 환자들끼리 군집화하였다

다양한 보험사기의 사전 예방 및 조기 적발을 위한 방안 마련

건강보험심사평가원이 폭넓은 자료를 분석하고 빅데이터를 활용해 도출한 결과는 교통사고정보, 심사참고자료, 건강보험 진료정보 데이터를 통합함으로써 다양한 데이터를 분석하는 데 활용할 수 있을 것이다. 또한 딥러닝 기술을 적용하여 문서, 이미지 등 심사참고자료를 분석하고, 탐색적 데이터분석과 군집분석으로 새로운 데이터의 특성 파악도 가능할 것으로 추측된다. 활용도와 정확도를 높이기 위해 파라미터(중심점, 최대 거리, 임계치)를 조정하여 교



통사고 다발생 이상패턴 탐지 모델을 갱신하고 검증하여 모델 고도화 진행도 필요할 것이다.

다양한 기관에서 두루 적용해 사용할 수 있는 모델

이번에 구축한 모델은 교통사고 다발생 이상패턴 탐지 결과를 심사시스템, 진료정보분석시스템에 적용할 수 있다. 건강보험심사평가원 내부적으로는 심사시스템에 적용해 1차 심사와 현지확인 심사에 활용 가능하고, 진료분석 시스템을 시각화해 교통사고 다발생 환자를 탐지하고 예측하는 종합모니터링을 진행할 수 있다. 또한 의료기관별 교통사고 다발생 환자와 사고환자별 탐지 및 현황을 구축할 수 있다. 외부적으로도 활용도가 높다. 외부 기관과의 정보 공유를 위한 법적, 제도적 근거 완비가 전제된다면 손해보험협회를 통해 각 손해보험회사에 공유할 수 있다. 더불어 금융감독원과 한국신용정보원에도 공유해 유용한 모델로 이용할 수 있을 것이다.

빅데이터 분석, 이렇게 달라졌어요!

- 교통사고 다발생 이상패턴 탐지 가능
- 자동차 보험금 청구에 대한 자동화 프로세스 구축 지원
- 자동차보험 심사자 관점의 분석 결과 시각화(진료정보분석시스템)
- 자동차보험 심사시스템에 분석 결과를 적용하여 자동차보험 청구 심사 업무 과학화

보건
의료

도서지역 건강권을 보장하는 응급시스템 스텝업

보건의료 서비스 분석으로 동등한 의료 혜택 제공



인천광역시는 강화도와 영종도, 백령도 등 크고 작은 섬들로 이루어져 있다. 대부분의 섬은 교통편이 부족하고 지리적으로 고립되어 있어 의료시설의 접근성이 비교적 낮다. 이에 인천시는 응급의료의 수요와 이송 수단별 의료센터 이동 시간 등을 분석해 응급환자의 초기대응 강화를 위한 의료자원 최적 활용 방안을 목표로 프로젝트를 진행했다.

#응급이송 #공공의료 서비스 #응급환자 초기 대응 #지역 맞춤형 예방활동 #응급헬기

모델 개발 전

- 섬이라는 지리적 고립으로 인한 의료시설 접근성 저하에 대한 고민
- 주민들의 정확한 건강 상태 분석 부족
- 응급환자에 대한 초기 대응 어려움

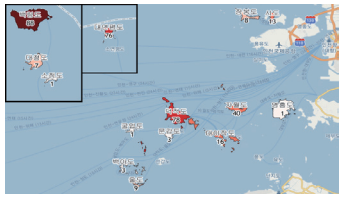
모델 개발 후

- 공공의료 서비스 개선 방안 제시
- 수단별 의료센터 이동시간 분석으로 응급환자 이송 최적 방안 마련
- 지역 맞춤형 공공의료 예방활동 방안 도출

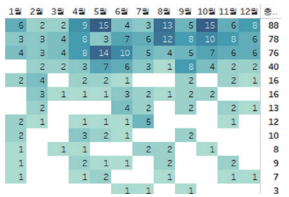
교통편의 부재와 응급의료의 제약

교통편이 열악해 대형 종합병원으로의 접근성이 떨어지는 도서지역에서는 응급 환자가 발생했을 때 어떻게 대처할까. □□시는 지난 2017~2018년 ○○군의 응급이송 주요 현황을 분석했다. 2년간 ○○군의 응급환자 헬기 이송은 총 385건으로 일평균 0.53회, 인구의 1.67%가 응급이송 서비스를 이용했다. ○○군 23개의 유인도 중 △△도는 88건으로 22.8%에 달하는 높은 수치를 보였다. 이어서 △△도가 78건, △△도가 76건, △△도가 40건으로 응급환자 헬기 이용률이 높은 4개 섬에 282건, 73.2%가 집중되었다. 이를 토대로 월별 파악도 이뤄졌다. 응급이송을 살펴보면 5월은 △△도와 △△도에서, 8월과 10월은 △△도와 △△도에서 응급환자가 집중적으로 발생했다. 응급이송이 중단된 36건은 기상 상황에 의한 18건과 사망 6건 등이 이유였으며, 응급이송이 가장 많은 △△도에서는 중단된 14건 중 11건이 기상악화로 인한 것이었다. 풍속, 운고, 시정, 안개 등 닥터헬기 기준에 따른 △△도 기상의 운항 중단 기능 비율은 24.3%에 달한다. 다음은 응급환자를 연령대에 따라 분류했다. 응급이송의 인구대비 비율은 연령대에 따라 증가하는 것으로

조사됐으며 50대와 90대는 35.4%로 비응급증상의 비중이 크지만, 60대에서 80대는 응급증상의 비율이 40%로 조사됐다. 중단 36건을 제외한 ○○군의 연령대별 인구대비 이송률을 조사한 결과 △△도와 △△도 등 핵심 도서지역의 관리가 필요하며, 응급이송 시기별 운영 효율화 방안 및 백령도의 기상, 연령대 분포 등을 고려한 응급의료 대책이 필요함을 알 수 있었다.



섬별 헬기이송 건수



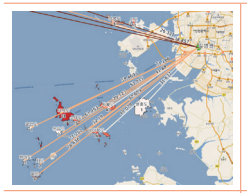
섬별, 월별 헬기이송 건수

<○○군 연령대에 따른 인구대비 이송률(중단 36건 제외)>

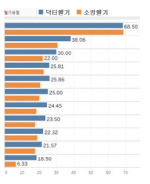
구분	20대	30대	40대	50대	60대	70대	80대	90대 이상	총계
이송건수(건)	18	18	44	65	77	53	45	14	349
인구대비 이송량(%)	0.63	0.81	1.72	1.65	2.15	2.41	4.2	6.09	1.66

시간·거리 단축으로 신속성에 연료비 절약은 덤

계류장에서 닥터헬기가 출발해 병원에서 의료진이 탑승한 후 섬까지 오가는 왕복 직선거리의 연간 합계는 약 11,500km이며, 평균 직선거리는 약 72km, 왕복 시 평균 78.1분이 소요되는 것으로 조사됐다. 2017년 응급의료통계에 따르면 닥터헬기의 출동거리 70km 초과 빈도는 □□시가 최대 45.6%로 □□시 도서지역 평균 21.4%에 비해 높은 것으로 나타났다. 또, 닥터헬기보다 △△도에 위치한 소방항공대에서 바로 출발하는 소방헬기가 거리와 시간을 단축하는 것으로 조사됐다. 특히, △△△ 지역에서 바로 출동·복귀가 가능할 경우 직선거리로 연간 3,468km를 단축할 수 있다. 이는 의료인력의 확보 및 계류지 변경 등에 기반해 닥터헬기 이동거리 단축으로 이어진다. 응급이송을 맡은 헬기는 소방헬기 183건, 닥터헬기 177건, 해경헬기 25건 순으로 집계되었다. 닥터헬기는 일출에서 일몰까지 운행시간에 제한이 있어 응급이송이 가장 많은 16시~17시 사이의 운행 비율이 낮고, 일몰이 빠른 10월에서 2월까지는 해당 시간대에 운영되지 않는다. 16시부터 06시 구간에서 닥터헬기를 이용하지 못한 경우는 113건으로 이 중 99건이 소방헬기로 이송되었기도 했다. 이는 응급헬기 이송실적이 집중되는 16시부터 17시까지의 닥터헬기 활용도 제고 및 소방헬기와 닥터헬기 간의 이용격차를 줄이기 위한 개선방안의 필요성을 뜻한다.



섬별 닥터헬기 이송 거리(km)



섬별 헬기 편도시간(분)



출발지 변경 시 이송 거리(km)

건강검진 및 진료내역 분석으로 도서지역 의료현황 파악

도서지역의 전반적인 의료상황을 분석하기 위해 인천시는 2016년부터 2018년에 해당하는 건강검진 결과와 2014년부터 2018년에 해당하는 진료내역 현황을 분석했다. 분석에는 고혈압 등 7개 목표질환(14개 검사항목)에 대한 데이터가 사용되었다. 7개의 질환에는 고혈압, 당뇨병, 이상지질혈증, 간장질환, 신장질환, 빈혈, 비만이 포함되었다. 해당 내용에 대한 전국('16년~'17년, 200만 건)과 ○○군의 검진정보('16년~'18년, 약 9천 건)의 평균 수치와 질환의심 판정비율 비교 등을 분석했다. ○○군의 연령대 비율을 표준화하기 위해 전국과 인천시의 데이터에 근거해 동일한 기준을 적용했으며, 지역별 평균 비교가 유의한지 확인을 위해 ANOVA(2개 이상의 다수의 집단 평균이 통계적으로 유의한 차이가 있는지 검정할 때 활용) 알고리즘을 활용했다. 또한 건강검진 실시기준(보건복지부 고시 행정규칙) 제공 판정기준 자료를 분석해 다양한 의료상황에 대한 분석을 진행했다.

응급헬기 운행 최적화 등 공공의료 인프라 개선

이번 프로젝트를 통해 응급헬기 이송과 건강검진 및 공공의료에 대한 정책제언이 도출되었다. 응급헬기 이송과 관련하여 첫 번째, 닥터헬기 계류 및 의료인력 대기 활용을 위한 도서지역 응급의료센터의 필요성이다. 닥터헬기 계류장의 연륙도서 이전 및 즉시 출동으로 신속성 제고가 필요하며 협력병원 확대를 통한 동승 의료인력풀 확대 및 이송 집중시기 대기인력을 확보해야 한다. 교통이 용이한 을왕리 인근에 의사와 헬기가 대기할 경우 연평균 약 3천 km(30.2%), 소방항공대를 활용할 경우 약 2천 km를 단축할 수 있으며, 대학병원 현지 대기 및 출발 시 약 8백 km 단축이 가능하다. 두 번째, 야간시간대 닥터헬기를 대체하는 소방헬기의 의료장비 보강 및 일몰 시간대 소방헬기의 의존도를 낮추기 위한 닥터헬기의 단계적 운행확대 검토가 필요하다. 세 번째, 앞서 도출한 결과를 토대로 주요 관리대상 섬을 선정하여 응급이송 집중기간에 병원 선진료, 주요 질환의 사전진단, 의사출장과 의료자원배치 등으로 사전관리를 강화해야 한다. 네 번째, 기상에 대비한 응급의료 인력 파견 및 의료자원을 재배치해야 한다.

건강검진 및 공공의료 관련 정책제언도 이어졌다. 첫 번째, 건강검진 결과의 질환의심수검자 비율에 비해 공공의료 이용률이 저조한 지역에 홍보, 방문 등 공공의료 서비스 활성화를 추진한다. 두 번째, 지역별 취약 만성질환을 고려해 관리정책을 강화한다. 지역의 취약질환 안내 및 권장신체활동 수행률을 위한 걷기대회, 주기적 고혈압 검진을 위한 출석 체크 이벤트, 건강 우수 섬 선정 등 맞춤형 교육과 행사를 추진한다. 또, 주기적 분석과 정책 활용을 위한 데이터 관리 강화 및 분석시스템을 확보한다.

도서지역 의료 서비스 향상으로 삶의 질 향상

인천시가 진행한 이번 도서지역 보건의료 서비스 분석 결과는 이러한 접근에서 유의미한 결과를 도출했다. 지리적으로 고립된 도서지역의 공공의료 보급에 대한 고민을 통해 구체적인 해결방안을 찾아낼 수 있었기 때문이다. 인천시는 이번 프로젝트를 진행하며 얻은 다양한 분석 결과를 활용해 공공의료 서비스 개선 방안을 제시하고 응급환자 이송 등에 있어 초기대응을 강화할 수 있는 의료자원에 대한 최적의 활용 방안을 제시하는 가이드라인을 마련할 수 있었다. 향후 인천시는 이러한 분석 결과를 바탕으로 도서지역 주민들의 건강권을 보장하고 삶의 질을 개선할 예정이다.

빅데이터 분석, 이렇게 달라졌어요!

- ☑ 공공의료 서비스 개선 방안 제시 및 응급환자 초기대응 강화를 위한 의료자원 최적 활용 방안 마련
- ☑ 공공의료 서비스 활성화 추진 및 지역 맞춤형 예방활동 강화
- ☑ 응급헬기 이송 및 건강검진과 공공의료 정책제언 마련



보건
의료

맞춤형 치매 진단 검사로 다각적인 치매 관리

AI 기반 치매 위험도 예측 모델 개발



본격적인 고령화사회로 접어들면서 노인 인구가 급격히 증가하고, 치매 인구도 늘고 있다. 당사자에게는 물론 가족과 사회적인 부담을 야기하는 치매를 잘 관리하기 위해서는 무엇보다 조기발견이 중요하다. 김해시는 노인을 대상으로 치매검사를 실시하고 위험도를 분석해 그 결과를 지도 내 공간정보로 시각화하는 작업을 수행했다.

#치매 조기발견 #가구별 Life-style 분석 #가구 돌봄 분석 #미래인구 추계 #치매 발생 위험도 예측 모델

모델 개발 전

- 치매환자 발생 원인이 모호
- 지자체 차원의 관리 체계 및 근거 부족

모델 개발 후

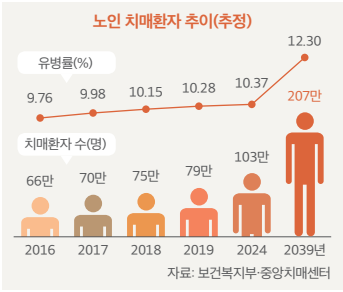
- 치매 환자 발생 위험 지역 도출
- 치매 환자 우선 검진 대상자 도출

절실한 치매 대응 및 관리

우리나라 국민이 가장 두려워하는 질병이 바로 치매이다. 고령화사회로 접어든 지금, 노인인구가 늘어나고 그에 따라 치매발병률도 높아지고 있다. 보건복지부 중앙치매센터의 발표에 따르면 노인 치매환자는 2016년에는 66만 명에서 2019년에는 79만 명으로 늘어났고, 2024년에는 103만 명, 2039년에는 207만 명에 이를 것으로 전망됐다.

하지만 사회적인 분위기는 치매에 대해서 드러내놓고 적극적으로 치료하는 것을 꺼리는 경우가 많다. 이러한 이유로 조기발견과 시기에 맞는 정확한 진단 및 치료에도 어려움이 따른다. 김해시는 이러한 상황을 고려해 치매를 조기발견하고 진행을 지연시킴으로써 치매환자에 대한 적극적인 관리와 치료가 이루어질 수 있도록 AI 기반 치매 위험도 예측 모델을 개발했다.

이를 위해 총 4개월 동안 지역 내 치매안심센터, 스마트도시담당관, 데이터 분석 전문가가 협업해 김해시 22



❶ 치매현상 분석	❷ 지역상관 분석	❸ 가구특성 분석	❹ 가구위험도 분석	❺ 시각화
· 치매환자 특성 분석을 통한 분석 인사이트 도출 · 분석 내용 - 분석 대상자료 선정 - 연도별 치매 분석 - 성연령별 치매 분석 - 인구당 치매 발생율 분석 · 거주환경과 치매는 상관성이 있는가를 파악하여 지역 관점의 분석 인사이트 도출	· 활용 데이터 - 인구수 - 가구 Life-Style - 지역 시설 - 인허가 사업체 - 사업체 총조사 · 치매 위험도 분석을 위한 치매 검진자별 가구와 생활권 특성 분석	· 분석 내용 - 행정동별 치매 유발 특성 분석 - 격자별 치매 유발 특성 분석 · 개인 검진자료 기반의 가구별 치매 위험도 예측모델 개발	· 분석 내용 - 치매 위험도 모델 개발 - 선정변수의 의미 해석(실무자와의 공감) - 성능평가 - 예측 정확도 검증 - 가구별/공동주택별/행정동별 치매 예측 · 위험도 예측 결과를 기반으로 치매 진단 홍보 장소 선정을 위한 지도 내 공간정보 시각화	· 분석 내용 - 치매 검진 현황 - 행정동별 - 격자별 - 사용자 편집(UX) 복합 주제도

만 가구별 라이프스타일 DB 추계를 반영해 분석하고 노인 연령에 대해 방문·집단 치매검사를 실시했다.

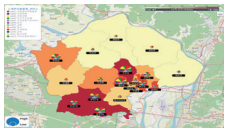
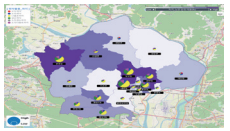
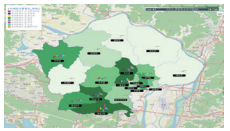
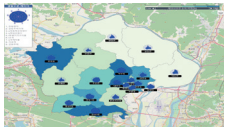
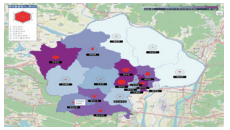
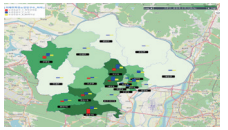
김해시는 치매환자 특성을 분석하여 인사이트를 도출하고 거주환경과 치매의 상관관계, 치매 검진자별 가구와 생활권 특성을 분석했다. 이를 토대로 개인 검진자료를 활용한 가구별 치매 위험도 예측 모델을 개발하고, 예측 결과를 기반으로 치매 진단 홍보를 진행할 지역 선정을 위해 지도 내 공간정보로 시각화하는 작업을 수행하였다.

정확한 데이터 확보를 위해 다각도의 자료 수집

우선 김해시는 유관기관과의 협업을 통해 정확한 자료를 수집하는 데 심혈을 기울였다. 치매진단과 관리 등은 단편적인 기준으로 평가하고 이루어지는 것이 아니기에 유관기관들이 힘을 모았다. 김해시 각 지역의 분석대상 자료를 선정해 연도별·성연령별 치매에 대한 자료를 취합하고, 인구당 치매 발생률도 확인했다. 또한 행정동별 치매 유발 특성과 가구별 라이프스타일, 지역 시설 등에 대한 데이터도 수집했다.

Market Segmentation 관점에서 494개 필드 생성

김해시는 정확한 모델 개발을 위해 인구정보 데이터와 통계청의 데이터를 융·복합하여 연령대별 미래인구 추계 및 가구별 거주수준과 소득수준 추계를 추출해 특성을 분석했다. 이를 위해 AI 학습용 데이터를 가구별 특성, 소득 수준 등으로 세분화하여 구축하였다. 우선 주민등록인구 정보와 공동주택 정보, 설문조사 결과 등의 원천 데이터를 분석해 통계 데이터를 구축하고 분류했다. 이 과정에서 연령별 인구, 가구별 거주수준과 소득수준, 가구 유형별 소득과 소비 통계 등을 바탕으로 돌봄대상을 분류하고 통계를 집계했다. 총 494개의 필드를 생성해 지리, 인구특성, 행위, 심리 등에 대한 가구 라이프스타일별 구축 관점을 수립해내고 각 가구별 연령구성코드, 혈연구성코드, 가구주 연령구성코드를 부여하여 돌봄유형을 분류할 수 있었다.

가구별 Life-Style 분석	가구 돌봄 분석	미래 인구 추계
		
가구 소득과 소비지출 분석	독거노인과 취약돌봄 가구수 분석	미래 5년차 생애주기별 인구수 변화 분석
		
가구 생활수준 종합 평가 분석	돌봄 유형별 취약도 평가 분석	취학 전/노인 인구 변화 분석

노인 인구 수와 치매 위험도가 비례하지 않는다는 결과 도출

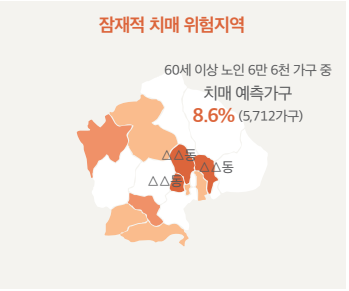
데이터 분석을 통한 예측 결과 소득, 독신, 결혼, 취약 돌봄관점에서 총 144종의 돌봄유형을 분류할 수 있었으며, 노인의 주 생활공간(행정동, 격자 1km)별 인구, 가구 라이프스타일, 가구 유형, 지역 시설, 인허가, 사업체 등 다양한 조사 변수를 선정할 수 있었다. 상관계수를 통해 치매환자 발생과 지역 특성(748개)과의 상관성도 분석할 수 있었다. 치매 검진자별 가구와 생활권 특성

에 따라 치매 발병 확률 분석을 위한 변수 554개도 선정했다. 또한 기계학습 및 AI 알고리즘을 활용해 치매 발생 위험도 예측 모델 개발을 통한 우선 검진 대상자를 도출할 수 있었다. 예측 결과 치매 검진 결과에 근거한 위험지역은 △△면, △△△△동, △△면으로 확인되었으며, 잠재적 치매 위험지역은 △△동, △△동, △△동인 것으로 나타났다. 읍면동별 노인 인구수 대비 위험 가구 비율도 확인할 수 있었는데 이 중 세 개 지역은 노인 인구수는 적지만 치매 위험도가 높은 지역이라는 결과가 나와 노인 인구 수와 위험도가 비례하지 않는다는 것을 알 수 있었다.

치매에 대한 지자체의 정책 수립 근거 확립

김해시가 개발한 이번 모델은 치매와 관련한 다양하고 방대한 자료를 분석해 추출한 결과이다. 지자체는 치매 관련 정보를 정확히 분석하고 그것을 근거로 정책을 수립하는 데 중요한 역할을 할 수 있을 것으로 보인다. 무엇보다 독거노인과 취약돌봄 대상을 찾아낼 수 있으며, 돌봄유형별 취약도 평가를 분석하고, 미래 5년차 생애주기별 인구수 변화와 취학 전 아동과 노인 인구 변화, 가구의 생활수준 종합 평가를 분석해 전반적인 복지 서비스를 높일 수 있다. 뿐만 아니라 치매 발생 위험도 예측 모델 개발을 통해 우선 검진 대상자를 도출하는 데에도 유용하게 활용하여 사전에 대비책을 구상할 수 있을 것으로 예상된다.





분석 모델 활용이 가능한 다양한 기관과 공유

김해시가 개발한 치매 위험도 예측 모델은 치매환자 관리가 필요한 여러 기관에도 꼭 유용한 자료이다. 위험도를 반영한 조기진단으로 치매 관리율을 제고할 수 있고, 치매안심마을을 비롯해 협력병원 등 관련 정책이 필요한 기관에 정책 근거로 제공될 수 있다. 치매를 정책적으로 잘 관리하기 위해서는 다양한 활동을 동시에 진행해야 하는 어려움이 있다. 이번 모델을 통해 맞춤형 예방, 공공후견인, 치매가족지원사업 등 도움이 절실하고 제대로 알려야 할 필요성이 있는 모든 사업에 적용해 치매관리에 대한 체계적인 지원이 가능할 것으로 전망된다. 마지막으로 김해시 내 관련 기관과 분석 결과를 공유해 연계협력을 강화함으로써 보다 안정적이고 효율적으로 치매에 대응하는 사회적 안전장치를 마련할 수 있을 것이다. 포스트 코로나 시대를 맞이해 치매 진단검사 맞춤형 도구를 마련한 김해시의 AI 기반 치매 위험도 예측 모델은 사회적 문제로 부상하고 있는 치매에 대해 긍정적이고 안전한 대책을 마련하는 데 큰 도움이 될 것이다.

빅데이터 분석, 이렇게 달라졌어요!

- 조기진단 및 치매 관리율 제고
- 지도 내 공간시각화를 통해 치매 발생 위험도 한눈에 파악 가능
- 치매 관련 다양한 지원사업 활용 및 연계 기관과 협력 강화

Part 5

산업고용

편리하게 사용하고 만족도 높이는 생활밀착형 서비스 I 한국전력공사

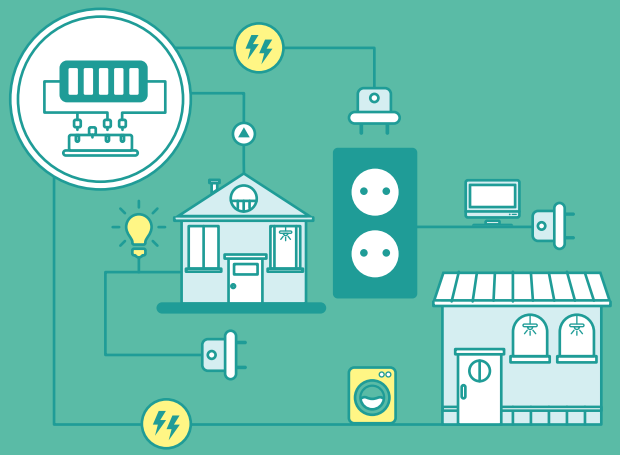
근로자 권리를 보호하는 믿음직한 임금체불금 지킴이 I 근로복지공단

취준생을 위한 맞춤형 취업 지침서 I 한국교육개발원

산업
고용

편리하게 사용하고 만족도 높이는 생활밀착형 서비스

전력사용 실시간 정보로 상업시설 영업 여부 파악



상업시설의 영업 여부를 확인하려면 전화를 하거나 직접 방문해야 하는 등 다소 불편한 절차를 밟아야 한다. 또 인터넷에 제공된 정보는 정확도가 떨어져 헛걸음을 하는 경우도 종종 발생한다. 이에 한국전력공사는 전력사용량을 통해 상업시설의 영업 여부를 파악하는 서비스를 개발했다.

#지능형 전력 계량 인프라(양방향 통신) #CNN 알고리즘 #영업 여부 실시간 예측 서비스

모델 개발 전

- (고객) 영업 여부 확인에 번거로움(전화/방문)
- (상업시설) 영업 여부 및 영업 시간에 대한 정보 제공에 어려움

모델 개발 후

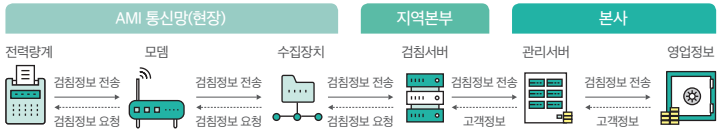
- (고객) 파워온 앱을 통한 상업시설 영업 정보를 실시간으로 확인
- (상업시설) 영업 여부 및 영업 시간에 대한 실시간 정보 제공

전력사용량 확인으로 영업 여부 판단하는 획기적인 모델

바쁜 일상을 추구하는 현대인들은 다양한 목적을 위해 상업시설을 방문한다. 우리나라는 세계 어느 나라보다 인터넷을 통해 많은 정보를 얻을 수 있지만 폭넓은 상업시설에 대한 정확한 정보를 제공받는 데에는 한계가 있었다. 특히 상업시설의 영업 여부를 정확히 확인하기 위해서는 해당 시설에 전화를 해 문의하거나 번거로움을 무릅쓰고 직접 방문하기도 해야 한다.

한국전력공사는 이러한 문제를 간단하고 효과적으로 해결할 수 있는 방법을 제안했다. 상업시설의 영업 여부는 전력사용량과 밀접한 상관관계를 가지고 있다는 점에 착안한 것이다. 빅데이터를 활용해 전력사용량을 확인하는 것만으로도 영업 여부에 대한 정확한 정보를 제공할 수 있는 획기적인 예측서비스를 제공하게 됐다.

한국전력공사는 활용도 높은 서비스를 개발하기 위해 머신러닝 학습 및 예측 서비스 모델을 구현했다. 총 5단계에 걸쳐 제공되는 예측 서비스는 우선 전력사용 데이터를 전처리(보정 및 정렬)하고, 일정량의 데이터를 확보한 후 인공신경망 학습을 적용했다. 일정수준의 정확도가 확보되면 예측 서비스를 가동



해 신규 데이터를 전처리하는 과정을 거쳐 기존에 학습한 데이터와 병합하여, 인공신경망 모델을 업데이트하는 방식이다.

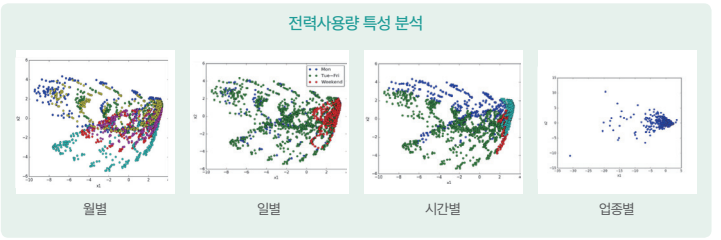
양방향 통신이 가능한 지능형 전력계량 인프라로 수집한 데이터

한국전력공사가 구축한 상업시설 영업 여부 실시간 예측 서비스를 제공하기 위해 구축한 모델은 AMI*를 통하여 전력사용 데이터를 분석하고 수집한 정보를 지역본부에 보내 검침한 후 정제된 정보를 다시 본사에 보내 관리서버를 통하여 영업 여부를 확인해 이용객에게 제공하는 방식이다. 이를 위해 상업시설의 영업 데이터를 활용하였다.

* AMI(Advanced Metering Infrastructure): 양방향 통신이 가능한 지능형 전력계량 인프라

전력사용량에 근거해 영업현황에 대한 정확한 분석

우선적으로 한국전력공사는 데이터를 분석하고 전처리하는 작업을 진행했다. 기존의 스마트미터는 전력사용량 형태와 수집주기의 차이가 발생해 일관성을 유지하는 것이 필요했다. 계절·요일·시간·업종 등의 상관관계에 따라 특성을 분석해 전력사용량의 변화와 절대사용량의 차이를 표준화하는 작업도 진행했다. 또한 스마트미터의 H/W 장애 및 전송 에러로 인한 특이 데이터와 비주기성 데이터에 대한 필터링도 했다. 상업시설 영업현황에 대한 전처리 역시 놓치지 않았다. 업주의 부정확한 응답으로 인해 발생할 수 있는 노이즈를 제거하기 위해 필터링 작업도 시행했다. 이와 더불어 609개소의 3개월 전력사용 표본을 활용했다. 인공신경망(RNN, FC, CNN)의 성능을 분석해 연산시간과 예측 정확도를 분석해 가장 적합한 방식인 CNN(Convolutional Neural



Network)을 선정했다. CNN은 이미지를 분석하기 위해 패턴을 찾는데 유용한 알고리즘이다.

예측 정확도 향상과 안정성 강화 동시에 확립

한국전력공사는 설문조사를 통한 실측데이터를 추가 학습시켜 예측 정확도를 98%까지 끌어올렸다. 또한 하이퍼 파라미터* 튜닝과 앙상블 기법*을 적용하였다.

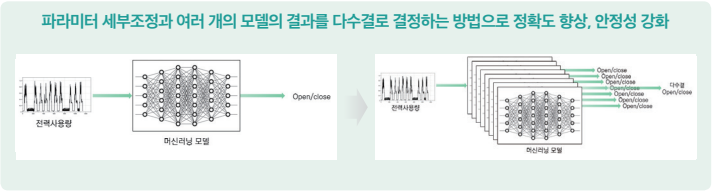
* 하이퍼 파라미터: 정확도 향상을 위해 사용자가 조절 가능한 파라미터

* 앙상블 기법: 동일 목적의 여러 개 분석 모델 결과를 활용하여 향상된 예측결과를 도출하는 모델



이용자의 편리성과 만족도 높이는 정보 제공

전국 2,250만 호에 대한 AMI 보급을 목표로 진행한 한국전력공사의 상업시설 영업 여부 실시간 예측 서비스 사업은 하루 약 2TB 수준의 대규모 빅데이



터가 자동으로 수집되는 시스템이다. 이번 모델 구축으로 실시간 영업여부, 해당 요일 평균 영업 시간 등이 추가되어 95%에 이르는 정확한 정보를 제공해 만족도를 높이게 될 것이다.

이를 통해 이용객들은 전화로 확인하거나 직접 방문해 영업 여부를 확인하는 시간적·경제적 불편함으로부터 자유로워질 수 있다. 편리하고 정확한 정보제공이 가능한 예측 서비스가 마련됨에 따라 한국전력공사는 서비스 제공을 위한 웹 애플리케이션도 제작할 예정이다.

'파워온' 앱으로 정확한 정보 제공과 전국적 범위 확대 동시 추구

이번 모델은 전력사용량을 활용해 상업시설의 영업여부를 정확히 예측하는 서비스다. 활용도를 높이기 위해 한국전력공사는 '파워온'이라는 이름의 애플리케이션을 제작해 시범 서비스를 시행하고 있다. 파워온은 강남구와 서초구 상업시설 1,500개소를 대상으로 필요한 상업시설을 찾고 검색결과를 통해 상세정보를 얻을 수 있는 방식을 취하고 있다. 현재 시범 운영 중인 파워온을 통해서는 근린생활시설의 실시간 영업 여부 정보는 물론 향후 3일 이내 상업시설의 영업시간 예측 정보까지 동시에 제공하고 있다.

이번 분석 결과로 구축한 서비스의 활용 범위를 넓히기 위해 한국전력공사는 향후 1단계로 기능 고도화 및 커버리지 확대, 2단계로 수도권 서비스 확대 및 연계프로그램 개발, 3단계로 전국단위 서비스로 확대를 진행할 예정이다. 이를 통해 전 국민이 편리하고 유익한 상업시설 영업 여부 실시간 예측 서비스를 제공받을 수 있을 것이다.





빅데이터 분석, 이렇게 달라졌어요!

- 상업시설 영업에 대한 정확한 정보 실시간 제공
- 파워온 서비스 고도화 및 전국으로 서비스 확대
- 모델 업데이트를 통한 정확한 정보 제공 가능

산업
고용

근로자 권리를 보호하는 믿음직한 임금체불금 지킴이

임금체불 근로자 보호를 위한 임금채권보장기금 운영 효율화



근로자는 당연히 일한 대가를 받아야 할 권리를 가지지만 실제로 여러 사업장에서 임금체불 사태가 빚어지고 있다. 근로자가 이를 직접 해결하기란 쉽지 않은 것이 사실이다. 근로복지공단은 그간의 데이터를 기반으로 임금채권보장기금을 더욱 효율적으로 운영할 수 있는 변제금 회수율 예측 모델을 구축했다.

#임금체불 #체당금 지급 #근로자 보호 #변제금 회수 #변제금 회수 가능성 모델

모델 개발 전

- 매년 증가하는 대금 지급과 변제금
- 처리인력 부족 현상 심화
- 변제금 회수를 위한 동일한 행정 처리

모델 개발 후

- 회수 가능성 모델에 기반한 등급별 대응 가능
- 변제금 회수 업무 효율화 및 임금채권보장기금의 효율적 운영

우리나라 임금체불금의 현주소에 맞는 회수 프로세스 절실

현재 우리나라의 전체 임금체불금은 2020년도 기준 1조 6천억 원이고, 임금체불 근로자 수는 30만 명에 육박함에 따라 나날이 대지급금 지급액과 변제금 회수 대상이 증가하고 있다. 최근 코로나19로 도산하는 기업의 수가 늘어나며, 관리 사업장 수가 급증해 관리 인력도 부족해졌다. 체당금 변제금 회수 업무 프로세스의 효율화·집중화·고도화를 통해 채권 회수율 제고가 절실했다.

이에 근로복지공단은 빅데이터 분석을 통해 임금체불 근로자 보호를 위한 체당금 운영 효율화 방안을 마련했다. 사업장 및 사업주의 특성을 반영하여 변제금 회수가능성을 예측해 체당금 징수활동을 개선하고 업무체계의 효율화를 위한 모델로, 사업장 및 사업주의 특성을 분석하여 머신러닝 기법으로 회수가능성 모델을 개발하고 변제금 회수가능 순위 및 회수요인 영향도를 도출해 내는 것을 목표로 했다.

이를 토대로 총 두 차례에 걸쳐 자문 및 개선방안을 토의했다. 1차 자문에서는 사업장별 변제 능력 변화를 확인할 수 있는 신용평점(3개월 단위, 6개월



단위 변경) 데이터를 추가 활용하자는 의견이 제시되었다. 2차 자문에서는 근로자 1인 기준 사업장별 체불금 등과 같은 비율 변수를 추가 활용할 필요가 있으며, 분석대상을 크게 두 그룹으로 구분할 필요가 있다는 의견이 제시되었다.

다양한 데이터 확보가 회수가능성 예측의 관건

사업장과 사업주의 특성을 반영한 변제금 회수가능성 예측을 통해 회수 업무 체계의 효율화를 추구하는 모델을 개발하고자 근로복지공단은 관련한 다양한 데이터를 수집했다. 사업장 정보 및 체당금 지급 정보, 변제금 회수 정보, 재산 정보, 사업주 신용 정보, 경제동향 정보 등 임금을 체불하는 요인을 도출하기 위한 다각적인 데이터를 동원했다. 이를 통해 사업장 특성을 고려해 변제금 회수율을 제고하고, 회수율을 높이기 위해 우선 순위 선정 방안 마련이 필요하다고 진단했다.

체당금 및 변제금의 안정적인 운영방안

근로복지공단은 먼저 사업장 정보 및 체당금 지급 정보, 변제금 회수 정보, 재산 정보, 사업주 신용정보 등의 주요 데이터를 추출하여 현황분석 및 변제요



인을 분석했다. 임금체불 변제금 회수에 영향을 미치는 사업장 재산 정보를 융합하여 사업장의 재산 지수를 개발했다. 사업장 재산 지수, 사업장별 특성 등을 활용하여, 회수 또는 미회수 등 각 사업장의 회수율을 예측할 수 있는 변제금 회수 가능성 예측 모형을 개발했다. 이때 약 6,000여 개의 체당금지급 사업장 특성을 고려해 변수 중요도에 따라 회수 가능 예측 모델의 개발과 영향도를 산출했다.

사업장 재산 지수 개발을 위해 먼저 사업장별 재산 데이터를 구성했다. 재산 상세정보 데이터에서 재산 종류별 수량 및 면적 합으로 데이터를 구성하고, 면적합 계산 시 거래 빈도가 높은 지목의 경우 가중치를 부과하여 계산했다. 마지막으로 계산된 각 재산변수를 정리하여 재산별 등급을 부여했다.

이러한 데이터 구성을 통해 도산대지급금*과 간이대지급금*이 지급된 사업장 중 회수율이 높은 사업장에 대한 여러 가지 특징을 도출할 수 있었다. 이러한 특징에는 사업주의 법인격(자연인, 법인), 사업장 재산등급, 신규창업 여부, 변제금 회수율, 신용등급, 카드 이용금액, 납·독촉 도달 여부, 대출 상환 여부, 상시 근로자 수, 연매출 추정 금액, 체불퇴직금 합계, 부동산 보유 여부 등이 있는 것으로 확인되었다.

* 도산대지급금: 기업의 도산이 인정되어 지급되는 대지급금
* 간이대지급금: 기업이 도산하지 않았더라도 체불임금 중 1천만 원까지 소액재판 또는 고용노동지청의 진정서(21년 10월 이후)를 통해 지급되는 대지급금

빅데이터로 도출한 안정적인 변제금 확보

임금채권보장사업 체당금 변제금에 대한 회수 영향요인을 분석한 결과 개인 신용등급, 카드 이용금액, 대출 상환금 등 개인 신용등급과 보유 부동산 자산 가치 등이 회수율에 영향을 미치는 것으로 나타났다.

고용보험 업종과 관련하여서는 고용보험 업종 77종에 대한 업종별 체당금 지급액 및 회수율 현황을 확인했다. 금속 제조업, 보건업, 출판업 등 업종별로 지급액 순위를 파악하였다.

2020년 1월부터 8월까지 최초로 체당금이 지급된 사업장 약 1만 개소의 임금채권 체당금 변제금 회수가능성 확률 예측모델을 실제 적용한 결과, 회수가능성 확률이 40% 이상인 사업장을 집중 회수 시 실익이 가장 큰 것으로 확인되었다. 이때 전체 사업장 중 회수가능성 확률이 40% 이상인 사업장은 79%이며, 해당 사업장의 자진 납부 금액은 90%를 차지한다.

업무 프로세스의 효율화, 집중화, 고도화

이번 공공 빅데이터 분석사업을 통해 근로복지공단은 임금체불 근로자 보호를 위한 임금채권보장기금 운영 및 관리방안 개선으로 업무체계 효율화를 이루어냈다. 회수확률에 따른 맞춤형 관리로 연간 5천만 원의 예산과 담당자 1인당 매월 40시간을 절감할 수 있게 되었다. 채권 관리대상에 대한 효율적인 분류체계 구축으로 행정력 및 예산을 집중하여 업무 효율성 증대도 추구할 수 있다. 체불환경의 변화를 반영한 데이터 활용과 분석으로 채권회수액을 증액하고 효율적인 임금채권보장기금을 운영해 기금 운영 효율화와 근로자 보호정책을 강화할 수 있게 된 것도 이번 모델 구축을 통해 얻은 성과다. 이는 빅데이터에 기반해 회수체계 고도화를 지속적으로 추진할 수 있는 버팀목이 될 것이다.

비용은 줄이고 효율은 극대화한 예측 모델의 구축

이번 프로젝트는 크게 세 가지 측면에서 의의를 가진다. 첫째, 회수가능성 예측 모델의 운용을 통해 회수 인력과 자원에 집중할 수 있게 되었다. 변제금 회수가능성 고확률군에는 납부독촉장 발송과 강제집행을 수행하고, 저확률군에는 채권 관리정지에 대한 요건을 검토할 예정이다. 둘째, 빅데이터 모델 결과물의 공단 내부 전산망 도입이 가능해졌다. 주기적으로 본부가 일괄 운용하여 1차적인 모델 운용 계획 추진 완료 후 공단 내부 전산망을 개선할 전망이다. 회수가능성 예측 확률의 이력 및 현황 조회 화면을 추가 개발하여 채권관리 담당자가 주기적으로 관리하는 것이 가능해졌다. 셋째, 모델 운용의 전문성이 확보되었다. 데이터 분석 전문업체에 위탁을 통해 지속적인 장애관리 및 모델 유지보수를 해나갈 예정이며, 임금채권보장법 개정에 따른 사용자 요구사항 수정을 통해 지속적으로 모델 활용 방안도 확보할 것이다.

분석 결과는 단기적으로 회수가능성 고확률군 선별 사업장에 대한 독촉장을 집중 발송하여 채권 회수율을 제고할 예정이다. 또한, 회수가능성 저확률군 선별 사업장에 대해서는 채권 관리정지 요건을 검토하여 집중 행정을 통한 행정력을 절감하고 업무 체계의 효율성을 확보해 나갈 예정이다. 이를 바탕으로 중장기적 모델 결과물의 이력을 관리하고 지속적인 모델 유지보수 및 모델 인터페이스를 개선해나갈 전망이다.

빅데이터 분석, 이렇게 달라졌어요!

- 채권 관리대상에 대한 효율적인 분류체계 구축
- 임금체불 근로자 보호를 위한 임금채권보장기금 운영 효율화
- 관리방안 개선을 통한 업무체계 효율화 및 빅데이터 기반 회수체계 고도화 지속 추진



산업
고용

취준생을 위한 맞춤형 취업 지침서

교육 빅데이터 분석으로 취업 및 진로 가이드 제공



한국교육개발원은 다양한 외부데이터를 확보하고 있으나 이를 다차원적으로 분석한 자료가 부족하고, 졸업생 특성과 관련된 취업통계를 조회하는 것에 어려움이 있어 취업 상담 과정에서 보유 데이터를 효과적으로 활용할 수 없었다. 이에 한국교육개발원은 실효성 있는 취업·진로 가이드를 제공하기 위해 각종 데이터와 개별 특성을 연계·분석하는 빅데이터 분석 사업을 진행했다.

#취업 및 채용 패턴 #Career Path 모델 #Decision Tree #취업 추천 #취업 지원 정책

모델 개발 전

- 다차원적 취업 관련 정보 부족
- 취업통계 정보 부족
- 채용패턴 정보 제공 부재
- 취업유지 관련 분석 정보 부족

모델 개발 후

- 다차원 취업 패턴분석을 통한 정확한 정보 제공
- 채용패턴 분석을 통한 기업유형별 정보 제공
- 정보를 활용한 맞춤식 취업준비 가능
- 지속가능한 취업유지 모델 구축

데이터 분석으로 좁은 취업의 문을 뚫는 해법 제시

장기적인 경기침체로 인해 취업의 문턱을 넘는 것이 점점 어렵고 치열해지고 있다. 이러한 현실을 반영해 한국교육개발원은 학생들의 특성을 입력하면 진로 및 취업패턴 정보가 생성되는 Career Path 모델을 개발하는 것을 목표로 세웠다. 이와 더불어 지속적으로 정보를 유용하게 활용할 수 있도록 취업통계 현황을 분석하고, 취업 대상자의 전공 또는 졸업 대학의 특성에 따라 유사한 그룹을 세분화할 수 있는 다차원 취업통계 분석모델을 개발했다.

이를 위해 한국교육개발원은 졸업생, 학교, 취업기업 등의 특성을 분석한 후 졸업생의 취업 계층모델과 취업기업의 채용패턴 분석모델(대학과 학과에 따라 어느 정도 연봉 수준의 기업군에 취업하는지에 대한 분석)을 개발했다. 분석 대상으로는 한국교육개발원 내부 데이터와 국민건강보험관리공단, 보건복지부, 고용노동부 등이 선정되었다.

Career Path 모델은 선정된 여섯 개 기관의 데이터를 수집해 품질을 분석하고 정제·변환한 후, 분석항목 생성 및 예측모형 도출을 통한 데이터베이스 구

성으로 개발이 진행됐다. 이는 향후 3~5년간 단계적 고도화를 지속해, 각 대학교에 API 형태로 취준생의 정보를 확인하고 내부 학사 시스템 및 외부 워크넷 등과 연계된 시스템을 구축할 예정이다. 다차원 취업통계 분석모델 개발에 대해서는 그룹 세분화를 통해 그룹별 계층적 구조의 취업패턴을 분석하고 학생들에게 알맞은 취업 노선을 추천할 수 있을 것으로 전망된다.

실효성 있는 정보 제공을 위한 분석 시나리오

Career Path 모델로 졸업생을 계층적으로 분류한 후 체계적으로 취업/진로 현황 정보를 정의해 계층별 취업현황을 비교하고, 취업 지역별·취업 기업규모별·급여수준별 비율 등 취업상담을 위한 다양한 비교정보를 생성했다. Career Path 모델로 취업통계를 시각화하는 작업도 진행했다. 데이터 시각화 툴로 취업 패턴 정보에서 중요한 통계정보를 선정하여 시각화 페이지를 만들고, 해당 속성을 선택하면 그에 따라 시각화 화면이 달라진다.

자문회의에서 제기된 전문가의 의견을 반영해 분석모델을 개선하는 과정도 거쳤다. 자문회의 의견을 반영하여, 최근 5개년 취업률에 대한 외부영향 요인 상관관계를 분석하고, 졸업 후 취업한 기업 유형이 다른 산업으로 이직한 졸업생 분석모델을 개발했다. 이를 통해 한국교육개발원은 학생들에게 도움이 되는 더욱 효율적인 모델을 개발할 수 있었다.

다양한 경우의 수를 흡수할 수 있는 폭넓은 데이터 취합

한국교육개발원은 고등교육기관 졸업생들의 니즈를 수용할 수 있는 다양한 정보를 주고자 건강보험공단 자료를 연계해 산업분류별 취업률 등의 다양한 취업패턴을 탐색했다. 기업 규모별 취업률은 취업기업의 상시 종업원 수를 기준으로 5개 유형으로 분류했으며, 지역별, 전공계열별, 학교 종류별, 산업분류별 취업비율 정보를 제공했다. 개별 학교에 제공되는 Career Path는 산업별 취업정보와 급여 수준을 비교하여 각 전공별로 취업자 수와 평균연봉, 연봉 구간별 구성비 등의 상세하고 다양한 정보를 포함했다.

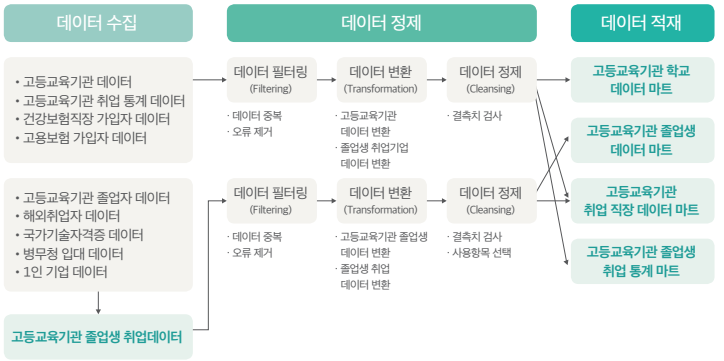
또 전공계열별, 학교 종류별, 학교 소재지별 등 취업 증감률과 고등교육기관

취업생들의 졸업 소요 기간(입학일부터 졸업일까지의 경과 기간)에 따라 취업패턴이 어떻게 다르게 나타나는지를 분석해 공개했다. 고등교육기관 취업생 중에서 취업기업에 계속 근무하는 졸업생과 퇴사한 졸업생의 특성도 함께 분석해 기업 규모별, 학교 종류별, 급여 수준별에 따라 취업 유지율 차이가 어떻게 나타나는지 상세하게 제공했다.

고등교육기관 취업률과 외부환경 요인과의 상관관계를 분석한 결과도 있었다. 분석 결과 전체 취업률과 실업률은 역의 상관관계를 보였으나, 지역별 취업률과 실업률은 상관관계가 낮았다. 전체 취업률과 1인당 GRDP 총계 역시 역의 상관관계를 보였으나, 지역별 취업률과 1인당 GRDP의 상관관계는 낮은 것으로 나타났다.

취업률을 제고하기 위한 과학적인 분석 적용

정확한 분석으로 효과적인 결과가 도출될 수 있도록 수집된 데이터에 대한 전처리 작업도 진행했다. 데이터셋 확인, 데이터 결측치 및 오류데이터 처리, 이상값 처리 등의 3단계를 거쳤다. 우선 데이터셋은 변수확인을 통해 변수의 유형(범주형/연속형)과 변수의 데이터 타입(Date/Character/Numeric)을 분석했다. Raw Data 분포분석을 통해서는 단변량 분석(Histogram, Box Plot)과 이변량 분석(상관관계 분석, 교차분석)을 진행해 정확한 분석을 도출하기 위한



데이터 결측치와 오류 데이터를 처리했다. 또한 변수를 삭제하며, 결측값을 대체하고, 중요항목은 추정값을 생성하는 꼼꼼한 분석과정을 거쳤다.

취준생과 취업 상담 담당자 모두 유익한 모델 프로세스

다양한 데이터를 분석해 도출한 정보는 다각도로 활용할 수 있다. 고등교육기관 상담 담당자는 취업 상담 과정에서 개인별 특성을 고려한 취업통계 정보를 활용하여 효과적인 취업 및 진로 상담이 가능하다. 졸업생 즉, 취업준비생(고등교육기관 졸업생, 대학생, 대학원생)도 이 모델을 활용해 효과적이고 철저한 취업준비를 할 수 있다. 목표기업의 채용패턴과 개인별 특성을 모두 고려한 Career Path 정보를 활용하여, 현실적인 취업 현황을 파악해 주체적인 Career Path 계획을 세울 수 있다. 교육정책 수립에 있어서도 활용도가 높다. 다차원적인 취업 통계정보와 최근 5년 증감율 분석정보에 기반해 청년 취업 정책의 효과를 모니터링할 수 있고 이를 근거로 정책을 수립할 수 있다.

세부적으로는 졸업생 개인 특성을 반영한 산업별 취업 가능성 예측모델 개발을 통해 최적의 Career Path 시뮬레이션이 가능하다. 이를 통해 고등교육기관 졸업생의 산업별 또는 지역별 취업 가능성 예측 스코어를 활용해 최적의 취업 추천이 가능하며, 목표로 하는 기업의 취업 가능성을 높이는 데 자격증, 토익, 해외연수 등의 영향력이 있는지 분석해 자원낭비를 방지할 수 있다.

취준생의 취업 가능성과 취업 유지율을 참고하여 취업기업 또는 취업산업을 세분화해 맞춤형 진로상담에 활용할 수 있어 중장기적인 취업 서비스를 제공할 수 있다.

취업 관련 여러 상황에 대응할 수 있는 정책 수립에 기여

고등교육기관 졸업생들의 취업은 개인적인 삶의 성취를 높이는 것은 물론 적재적소에 필요한 인재를 등용함으로써 기업과 사회, 그리고 경제 발전을 이끌 수 있는 중요한 일이다. 한국교육개발원이 이번에 구축한 교육 빅데이터 분석을 통한 취업 및 진로 가이드 모델은 다양한 효용성을 가지고 적용대상도 폭넓다. 고등교육기관 취업패턴을 분석해 학교 종류별, 전공계열별 등 다르게 나타나는 취업정보를 제공할 수 있다. 기업유형별 채용패턴 분석 결과는 기업 규모별, 분야별, 채용 패턴 등 취업정보에 유용하게 활용할 수 있다. 기업의 채용패턴을 반영하면 취업증감률에도 효과적으로 대응할 수 있다. 또한 남녀 졸업소요기간에 따른 취업 정보, 전공과 기업 규모에 따른 취업 유지율 정보, 외부환경이 취업에 미치는 상관관계 정보도 제공해 상황과 조건에 따른 적절한 대응과 해결책을 수립할 수 있을 것으로 전망된다.

후속사업 및 진행 상황

한국교육개발원은 행정안전부의 빅데이터 분석사업을 통해 개발된 모델을 고도화해 청년 개인의 특성, 진학경로 및 희망진로에 따라 맞춤형 진로·취업패턴 추천(Curation)을 위한 청년 진로정보 시스템 구축계획(4단계)을 수립하였다. 2020년에는 1단계 과정으로 분석항목 도출을 위한 기초연구를 실시하였으며 연구결과를 바탕으로 고등교육기관(대학)에서 활용할 수 있는 취업진로지도를 위한 정보체계를 구축하였다. 이로 인해 기존에 단순 장표로 제공하였던 통계분석 결과를 조건변수와 결과변수를 다중 선택하여 다차원 분석 자료로 보여줄 수 있는 데이터 시각화 서비스가 가능해졌다. 2021년에는 2단계 과정으로 맞춤형 진로추천 서비스를 위한 분석모델 개발 사업을 추진하고 있다. 개발되는 분석모델에 외부 데이터를 연계하여 개인별 특성 및 역량정보에 따라 진로를 추천하는 서비스를 고등교육기관에 제공할 예정이다. 추후 3단계 및 4단계에서는 수요자 맞춤형 서비스를 위해 분석 모델과 시스템을 고도화할 예정이며, 최종적으로 대국민 서비스 제공을 목표로 사업을 추진하고 있다.

빅데이터 분석, 이렇게 달라졌어요!

- 근로형태·임금수준·고용유지기간 등의 데이터와 취준생 및 기업별 특성을 연계 분석
- 다양한 니즈에 부합하는 Career Path 정보 제공
- 실효성 있는 취업·진로 가이드 제공
- 목표기업의 채용패턴 정보를 활용한 맞춤식 취업준비 가능



공공부문 데이터 분석·활용 우수사례집

발행일 2021년 12월
발행처 행정안전부, 한국지능정보사회진흥원
ISBN 978-89-8483-629-7 03000
디자인·제작 늘품플러스 070-7090-1177

• 본 자료는 대한민국 저작권법의 보호를 받습니다. 작성된 모든 내용의 권리는 작성자에게 있으며, 저작자의 승인이 없는 모든 사용이 금지됩니다.

• 이 자료의 일부 혹은 전체 내용을 무단으로 복제, 배포, 2차적 저작물을 작성할 경우 5년 이하의 징역 또는 5천만 원 이하의 벌금과 민사상 손해배상을 청구합니다.