

인적사항 및 주요경력



지원분야 비즈니스 빅데이터 분석 트랙

인적사항 이 름 김성범

생년 월일 1996년 5월 8일 (만 25세, 남)

연 락 처 010-9591-7640

이 메 일 tjdqja0508@naver.com

블 로 그 <https://blog.naver.com/tjdqja0508>

| | |
|---------------|--|
| 학 력 | 2021. 08 연세대학교 정보대학원 디지털서비스 연구실 2021. 08 순천향대학교 경영학과 졸업 (학점 3.59) * 복수전공 : 빅데이터공학(학점 4.29) * 동아리 : 빅데이터 분석 / 마케팅기획 / 창업 * 성적우수 장학금 2회, 특별장학금 3회 수상 2015. 02 신송고등학교 졸업 |
| 보유자격 | 2020 데이터 분석 준전문가 (ADsP), 컴퓨터활용능력 1급 2019 전산회계 1급, 회계관리 2급 2018 한국사검정능력1급, 워드프로세스 1급, 정보처리기능사 2016 MOS word |
| 수상경력 | 2021. 02 학내 프로젝트 중심 창업동아리 경진대회 장려상 2021. 02 공공 빅데이터 청년인턴십 우수인턴 NIA 원장상 2020. 11 서울특별시 빅데이터 캠퍼스 공모전 최우수상 (주제 : 서울시 응급의료 전용헬기를 위한 인계점 최적입지 분석) 2020. 12 학내 창업경진대회 장려상(주제 : 스마트폰 액세서리) 2020. 07, 12 학내 창업 결과 발표회 금상(주제 : 스마트폰 액세서리) 2019. 12 우체국 예금 서포터즈 개인 최우수 |
| 교육훈련 이력 | 2020. 12 행정안전부, 2020 공공 빅데이터 청년인턴십 확대운영 - 데이터 전문교육(200시간), 인천광역시청 인턴십(3개월) |
| 핵심 역량 | (활용SW) R, Python, Q-GIS (분석 및 관심 역량) 빅데이터 분석과제 기획, 변수·요인 탐색, 데이터 수집/정제, 통계, Business Intelligence, Machine·Deep Learning, Data·Text Mining, Digital Marketing, Recommender System |
| 프로젝트 수행 이력 | - [2021. 01] 해양수산 빅데이터 활용 혁신 아이디어 제안 - [2020. 12] 지방세 체납징수 효율화를 위한 빅데이터 분석 - [2020. 11] 서울시 응급의료 전용헬기를 위한 인계점 최적입지 분석 - [2020. 07] 인스타그램 해시태그를 통한 사회 연결망 분석 - [2020. 07] 의정부 지역화폐를 활용한 소비패턴 분석 시각화 - [2019. 07] 해외 소비자 타겟 선정 및 시장 조사 Study Abroad |

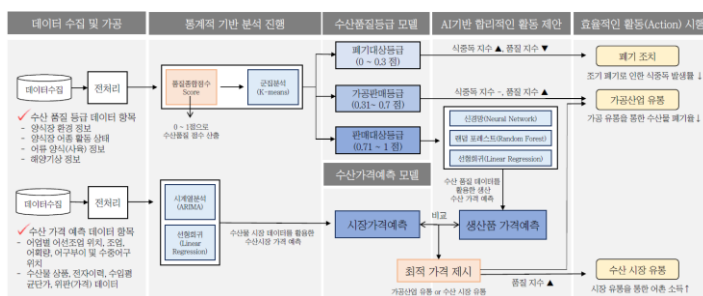
수산물품질 등급화를 통한 효율적인 양식장 가이드

| | |
|-----------|---|
| 프로젝트명 | 수산물품질 등급화를 통한 효율적인 양식장 가이드 |
| 배경 및 목적 | (배경) 다양한 외부 요인으로 양식장 관리의 어려움으로 인한 문제 발생 (목적) 문제를 사전에 예측하여 선제적 대응을 하는 AI기반 솔루션 제안 |
| 수행기간 | 2021. 01 ~ 2021. 01(약 1개월) |
| 팀 구성 및 역할 | 4인 1팀 / 프로젝트 매니저 및 분석 제안서 작성 |
| 사용언어 및 도구 | Power Point(분석 제안서) |
| 주요 활용 데이터 | 양식장 환경 정보, 해양기상정보, 어획정보, 수산물 데이터 등 |
| 주요 분석 기법 | 군집분석(K-means), 선형회귀(Linear Regression), 시계열분석(ARIMA) |
| 산출물 | 해양수산 빅데이터 활용 혁신 아이디어 제안발표자료 |

■ 분석 방법

- 양식장 환경 정보와 해양기상 정보 등을 융합하여 비지도 학습인 **K-means**를 통해 3가지의 품질등급으로 '수산물 품질 등급'을 정의함
- 수산물 시장가격과 어획량 데이터를 융합하여 지도 학습인 **선형 회귀**를 통해 '수산물 가격 예측'을 제공하여 합리적인 시장 활동을 위함

■ 분석 결과



- 품질종합점수의 분포에 따라 **군집분석(K-means)**를 이용하여 시장 유통, 가공산업 유통, 폐기 등급의 **3단계**로 제시
- 수산 산업 종사자에게 **시장가격 예측**을 통한 가격을 제안하여 합리적인 시장 활동 제안

- 지속가능한 AI 기반 수산 산업 가이드를 위하여 분석 모델의 **오류 최소화 필요**
- 데이터 축적**과 주요한 **파생변수 도출**을 통해 등급 분류 및 가격 예측 모델을 제안

■ 활용 및 기대효과

- ✓ 효율적인 양식장 관리 가이드를 통한 수산물 식중독 문제 해결 및 집단 폐기 예방
- ✓ 품질경쟁력 강화를 통한 수산 시장 소비 활성화
- ✓ 양식장 생산력 증가와 어촌소득 7,000만원 조기 달성 기대

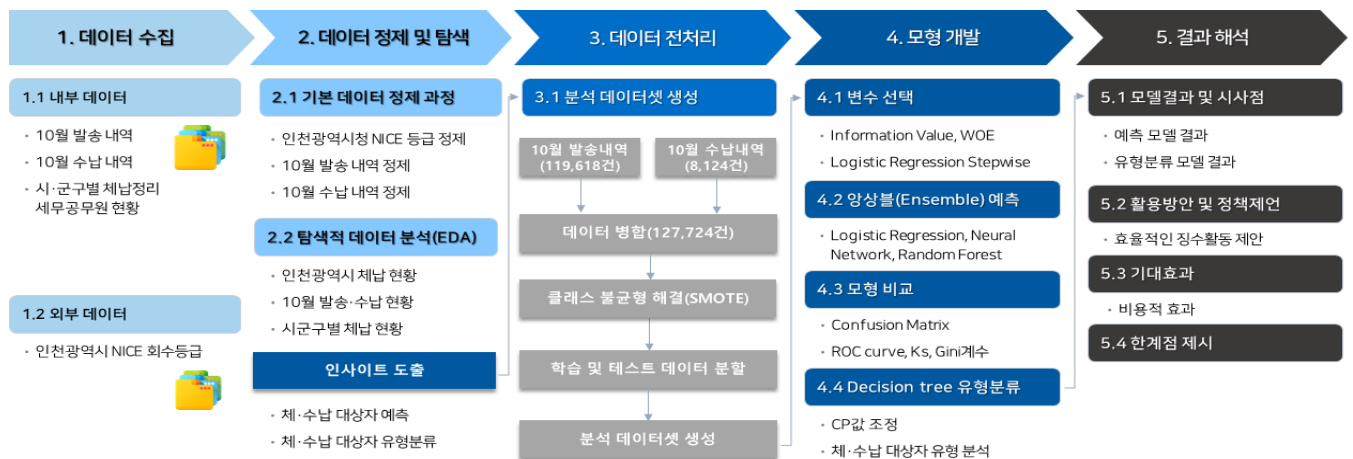
지방세 체납징수 효율화를 위한 빅데이터 분석

| | |
|-----------|---|
| 프로젝트명 | 지방세 체납징수 효율화를 위한 빅데이터 분석(1/2) |
| 배경 및 목적 | (배경) 지방세 체납액 증가에 따른 데이터 기반 정책 마련의 필요성 (목적) 체·수납 대상자 예측과 유형분류를 통한 효율적 징수활동 제안 |
| 수행기간 | 2020.10.5 ~ 2020.12.31(약 3개월) |
| 팀 구성 및 역할 | 개인 / 데이터 정제, 모델 구축, 모형 검정, 결과 해석 |
| 사용언어 및 도구 | R, Python, Excel |
| 주요 활용 데이터 | 체납·결손의 내부데이터와 NICE 신용정보의 외부데이터 |
| 주요 분석 기법 | Logistic Regression, Neural Network, Random Forest, Decision tree |
| 산출물 | [별첨] 인천광역시 지방세 체납 효율화 빅데이터 분석 보고서 |

■ 분석 방법

- 내·외부 데이터를 융합하고 회수등급을 중심으로 데이터 정제
- 탐색적 데이터 분석(EDA)을 통한 체납징수 현황을 파악 후 문제 정의
- 앙상블(Ensemble)과 의사결정 나무(Decision tree) 모형을 통해 효율적인 징수 활동 전개 제안하는 등 구체적인 해결 방안 제시

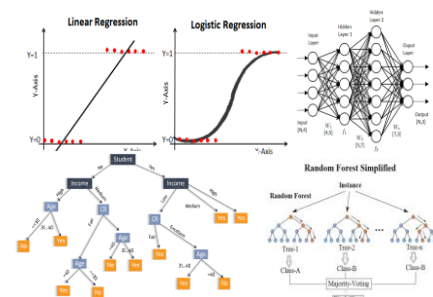
■ 분석 프로세스



- 내·외부 데이터를 활용하여 데이터 탐색 후 인사이트를 도출하고, 알맞은 분석 데이터 셋을 생성하여 모형 개발을 통해 결과를 해석하였음.

■ 주요 활용 분석 방법

- 앙상블(Ensemble) 모형을 통한 체·수납 대상자 예측
- 의사결정 나무(Decision tree) 모형을 통한 체·수납 유형분류

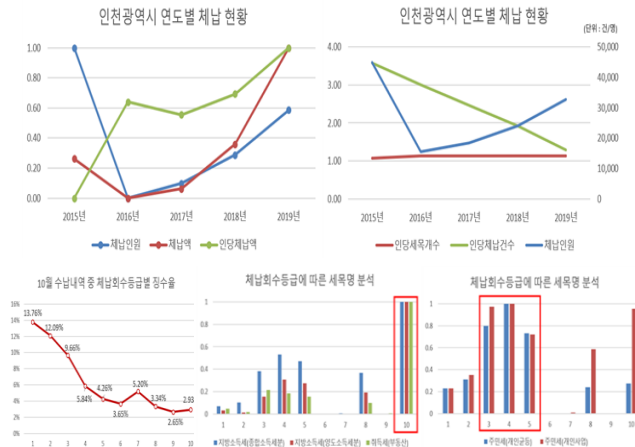


지방세 체납징수 효율화를 위한 빅데이터 분석

| | |
|-------|-------------------------------|
| 프로젝트명 | 지방세 체납징수 효율화를 위한 빅데이터 분석(2/2) |
| 수행기간 | 2020.10.5 ~ 2020.12.31(약 3개월) |

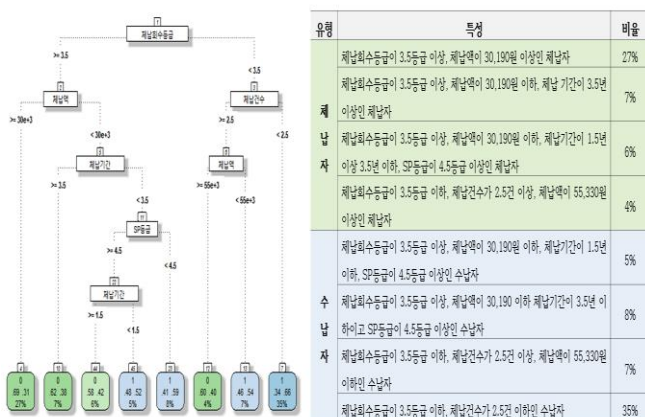
■ 분석 결과

<탐색적 데이터 분석>



- 인천시 체납징수 현황 파악을 통한 **인사이트(Insight)** 도출
- 회수등급에 따른 체·수납 대상자에 대한 특성이 유의미한 것으로 파악

<체·수납 대상자 유형분류>



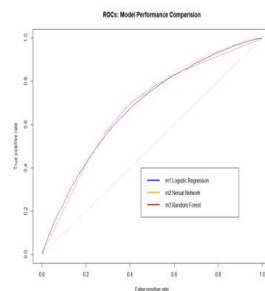
- 체·수납 대상자를 **총 8개의 유형** 분류를 통해 **납세자의 패턴**을 확인
- 이를 통한 납세자의 특성을 반영하여 **효율적인 징수활동** 전개

■ 활용 및 기대효과

- ✓ **연령, 지역, 체납정보, 신용정보**를 통한 고지방법으로 **차별화된 징수활동** 전개
- ✓ 체납 대상자 유형 특성에 **효율적인 징수 활동**을 통한 **민원 감소, 행정 효율 증대**
- ✓ 체납·신용 데이터를 융합한 **회수등급 갱신**으로 데이터 기반 **세무행정 실현**

<체·수납 대상자 예측>

| 모형구분 | 우·불량 예측(A) | 우·불량 실제(B) | 정확도(A/B) |
|---------------------|------------|------------|----------|
| Logistic Regression | 45,499 | 71,592 | 63.55% |
| Neural Network | 46,097 | 71,592 | 64.39% |
| Random Forest | 46,420 | 71,592 | 63.85% |

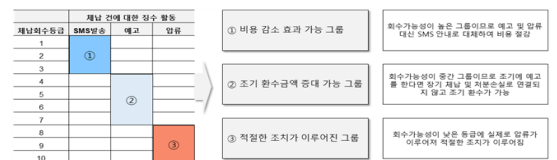


| Model | AUC | KS | Gini |
|---------------------|--------|--------|--------|
| Logistic Regression | 0.6419 | 0.6549 | 0.6600 |
| Neural Network | 0.6355 | 0.6439 | 0.6485 |
| Random Forest | 0.6219 | 0.6164 | 0.6201 |

| Confusion Matrix | Logistic Regression | Neural Network | Random Forest |
|------------------|---------------------|----------------|---------------|
| Precision | 0.6419 | 0.6549 | 0.6600 |
| Accuracy | 0.6355 | 0.6439 | 0.6485 |
| Recall | 0.6219 | 0.6164 | 0.6201 |
| Specificity | 0.6493 | 0.6717 | 0.6771 |
| FP_rate | 0.3126 | 0.3099 | 0.3117 |
| F1_Score | 0.6317 | 0.6351 | 0.6394 |
| Kappa | 0.2712 | 0.288 | 0.2971 |

- 앙상블(Ensemble) 모형** 구축을 통한 체·수납 대상자 **예측**
- 정확도는 **약 64%**로 회수등급과 다양한 요인에 대한 연구가 필요

<기대효과와 시뮬레이션>



| 단위 : 백만원 | | | |
|----------|-----------|---------|-------|
| 회수회수등급 | 입금X | 입금Y | 입금Z |
| 1 | 105,306.2 | 1,768 | 165 |
| 2 | 809.2 | 1,768 | 165 |
| 3 | 809.2 | 1,768 | 165 |
| 4 | 142,799.2 | 2,748.2 | 2,968 |
| 5 | 142,799.2 | 2,748.2 | 2,968 |
| 6 | 142,799.2 | 2,748.2 | 2,968 |
| 7 | 142,799.2 | 2,748.2 | 2,968 |
| 8 | 142,799.2 | 2,748.2 | 2,968 |
| 9 | 142,799.2 | 2,748.2 | 2,968 |
| 10 | 142,799.2 | 2,748.2 | 2,968 |

- 실제 징수업무 적용 시 **절감비용, 환수금액** 시뮬레이션 결과 도출
- 징수 활동 시 **약 400만원**의 비용절감 효과 기대

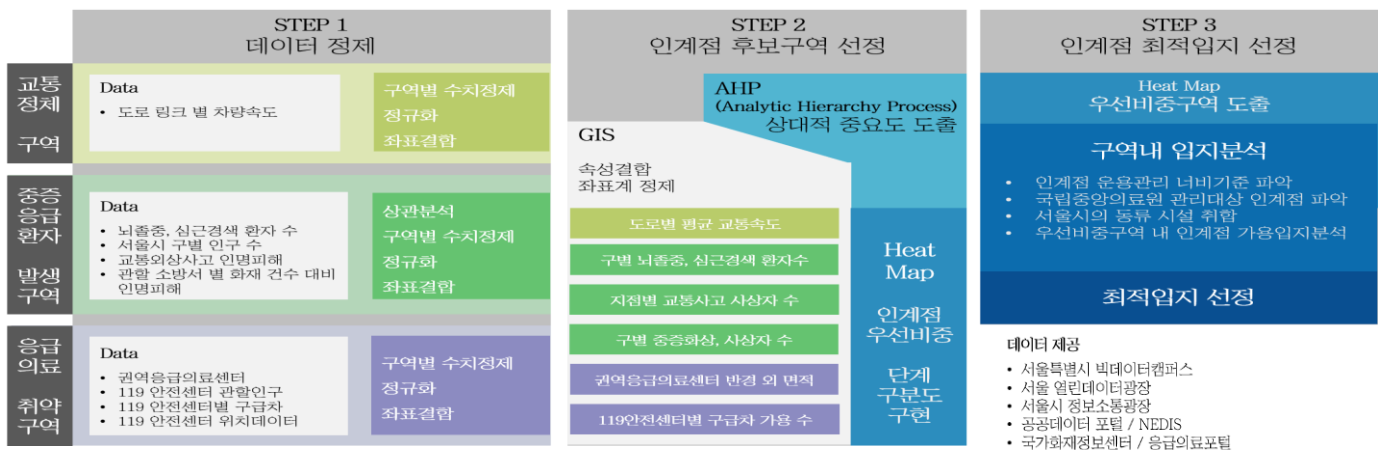
서울시 응급의료 전용헬기를 위한 인계점 최적입지 분석

| | |
|-----------|--|
| 프로젝트명 | 서울시 응급의료 전용헬기를 위한 인계점 최적입지 분석(1/2) |
| 배경 및 목적 | (배경) 서울시 응급의료 전용헬기 도입에 따른 인계점 선정 필요 (목적) 과학적 의사결정을 통한 효율적인 인계점 입지 분석 제안 |
| 수행기간 | 2020.08.24 ~ 2020.11.20(약 3개월) |
| 팀 구성 및 역할 | 9인 1팀 / 팀장 및 프로젝트 매니저, 프로세스 기획 관리, 요인 선정 |
| 사용언어 및 도구 | R / Python / Excel |
| 주요 활용 데이터 | 교통정체, 중증응급환자, 응급의료 취약구역 관련 공공데이터 |
| 주요 분석 기법 | 계층 분석 과정(Analytic Hierarchy Process : AHP) |
| 산출물 | [별첨] 「2020 서울시 빅캠공모전」_「Dr.헬기팀」_분석결과서 |

■ 분석 방법

- 인계점 선정요인 데이터를 활용하여, 계층적 분석 과정(AHP)후 산출한 가중치를 이용하여 요인에 대한 상대적 중요도 도출
- 인계점 선정 요인에 대한 가중치 종합 → 인계점 최적입지 선정

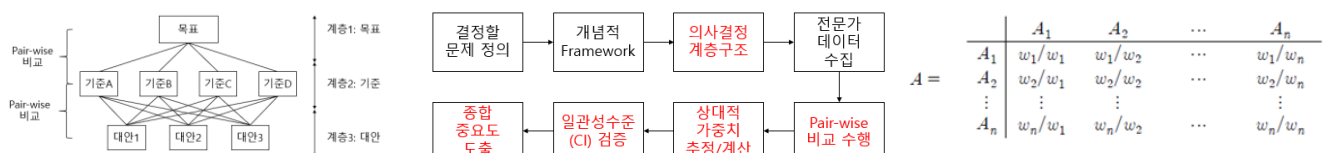
■ 분석 프로세스



- 요인들의 데이터를 수집 및 정제 후 계층 분석 과정(AHP)을 통해 가중치 산출하고, 우선비중구역을 도출하여 최적입지를 선정하고자 함.

■ 주요 활용 분석 방법

- 계층 분석 과정(Analytic Hierarchy Process, AHP)을 이용한 요인 가중치 산출



서울시 응급의료 전용헬기를 위한 인계점 최적입지 분석

| | |
|-------|------------------------------------|
| 프로젝트명 | 서울시 응급의료 전용헬기를 위한 인계점 최적입지 분석(2/2) |
| 수행기간 | 2020.08.24 ~ 2020.11.20(약 3개월) |

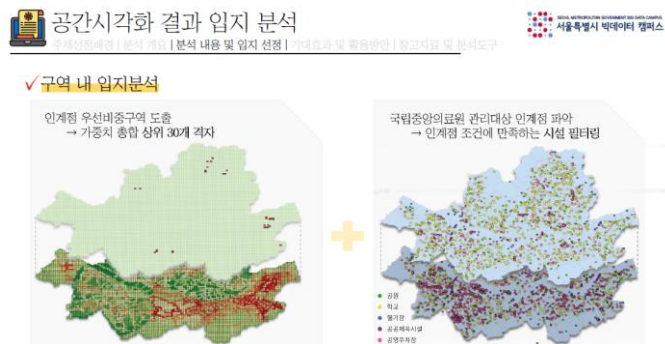
분석 결과

<인계점 선정 요인>



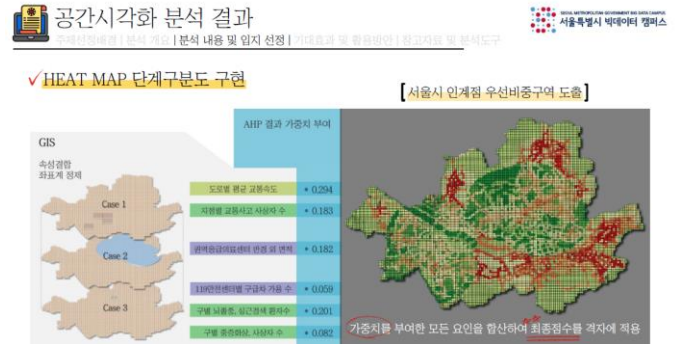
- 소방청 범부처 매뉴얼과 관련 연구논문 기반 정성적 자료로 접근
- '교통정체구역', '중증응급환자 발생구역', '응급의료 취약구역' 3가지 요인 선정

<인계점 입지 분석>



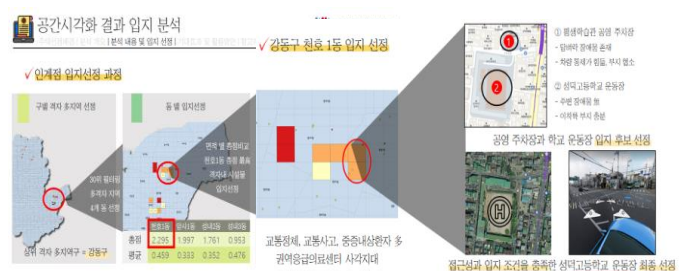
- 인계점 우선비중구역 중 상위 30개의 격자 나열 후 **4개의 구로 구분**
- 인계점 조건에 해당하는 시설 중 실제 **활용가능한 시설을 필터링**하여 위치표시

<공간 분석 시각화>



- AHP를 통한 가중치를 '우선비중구역'로
하나 종합점수인 **히트맵**으로 산출
- 서울 외곽 지역 및 도심지역에 **인계점
우선설치** 구역을 확인

<인계점 최적입지 선정>



- 예시로 '강동구'에서 면적 별 총점 비교를 통해 '천호 1동'을 **최적입지 분석**
- '천호 1동'은 인계점 선정 **요인 3가지의** 특성을 보이며, 인계점 **활용 기준에 충족**

■ 활용 및 기대효과

- ✓ 서울시 내 과학적인 인계점 선정으로 응급의료 전용헬기 운영활동 강화
- ✓ 서울시 골든타임 이수율 33%에서 42%으로 상승할 것으로 기대
- ✓ 지역별 특성을 고려하고, 공공 데이터를 활용한 맞춤형 표준분석모델로 활용 가능

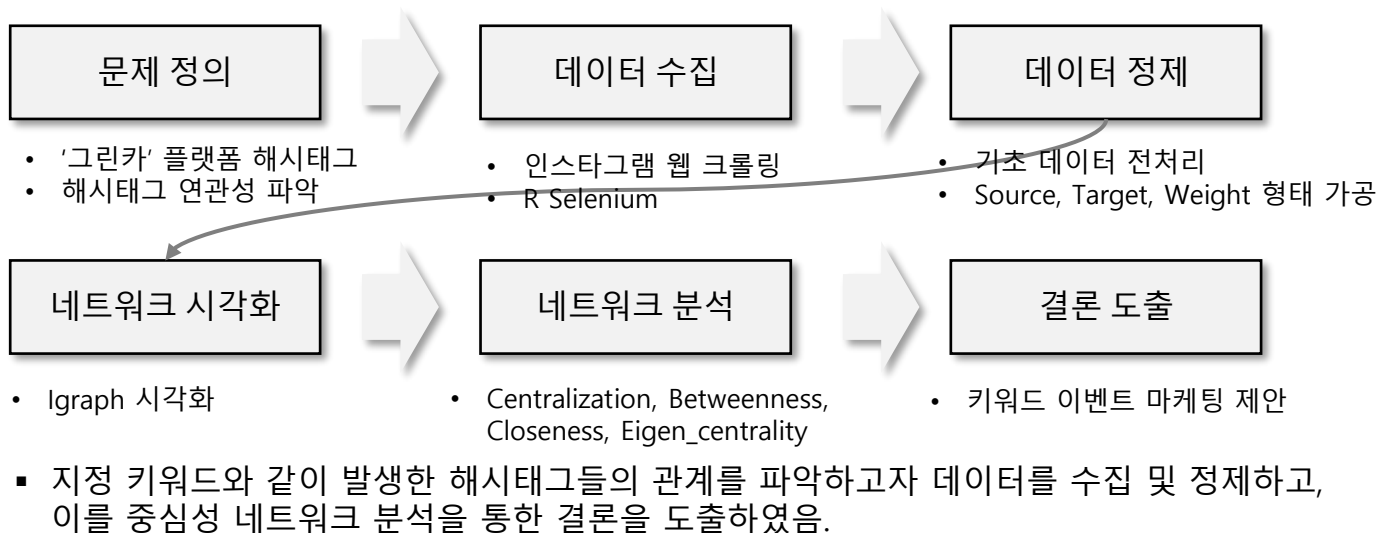
인스타그램 해시태그를 통한 사회 연결망 분석

| | |
|-----------|---|
| 프로젝트명 | 인스타그램 해시태그를 통한 사회 연결망 분석(교내 과제)(1/2) |
| 배경 및 목적 | (배경) SNS 사용자의 해시태그 분석을 통한 과학적 마케팅 필요 (목적) 연관 발생하는 키워드를 파악하여 新 마케팅 제안 |
| 수행기간 | 2020.04 ~ 2020. 07 (약 3개월) |
| 팀 구성 및 역할 | 2인 1팀 / 팀장, 데이터 수집 및 정제, 네트워크 분석 및 시각화 |
| 사용언어 및 도구 | R |
| 주요 활용 데이터 | 인스타그램 해시태그 크롤링 |
| 주요 분석 기법 | SNA(Social Network Analysis) |
| 산출물 | 인스타그램 해시태그를 통한 사회 연결망 분석(교내과제) |

■ 분석 방법

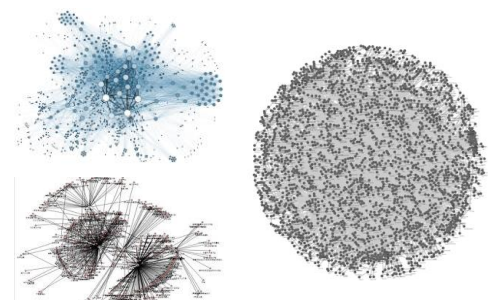
- 키워드를 중심으로 인스타그램 해시태그를 크롤링하여, 데이터 수집 및 정제
- R의 Igraph 패키지를 활용하여 네트워크를 시각화하고, 연결, 매개, 근접, 아이겐벡터 중심성을 이용한 네트워크 분석 시행
- 중심성에 위치한 키워드를 중심으로 과학적 의사결정 제안

■ 분석 프로세스



■ 주요 활용 분석 방법

- 사회 연결망 (Social Network Analysis, SNA)

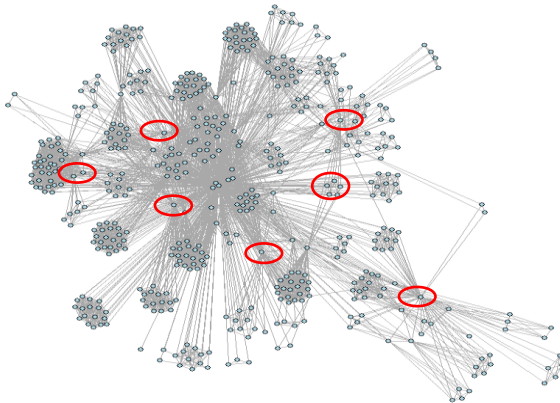


인스타그램 해시태그를 통한 사회 연결망 분석

| | |
|-------|--------------------------------------|
| 프로젝트명 | 인스타그램 해시태그를 통한 사회 연결망 분석(교내 과제)(2/2) |
| 수행기간 | 2020.04 ~ 2020. 07 (약 3개월) |

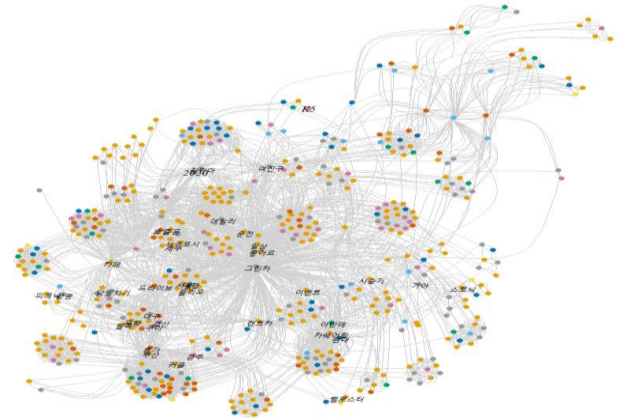
■ 분석 결과

<해시태그 네트워크 시각화>



- '그린카' 키워드를 중심으로 서브그룹 형성된 것을 확인
- 서브그룹간 연결하는 특정 노드를 파악

<네트워크 키워드 출력>



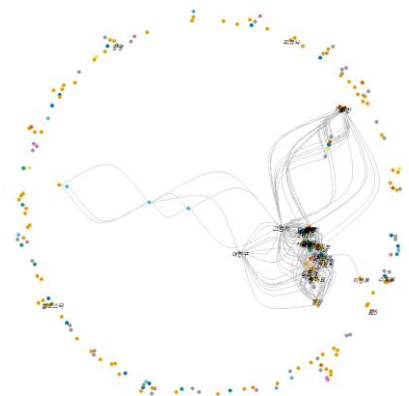
- 형성된 네트워크 연결망 라벨을 띄워 해시태그 명을 파악
- 이를 통한 특정 노드들의 키워드를 파악

<네트워크 중심성 분석>

| 중심성(평균) | 그린카 | 드라이브 | 여행 | 아반테 | 커플 |
|-------------------------|----------|---------|---------|---------|--------|
| Edges | 4363 | | | | |
| centralization | 0.9105 | | | | |
| betweenness(269.7) | 108982.2 | 1546.87 | 366.96 | 341.73 | 123.82 |
| closeness(0.0009685) | 0.00187 | 0.00105 | 0.00102 | 0.00102 | 0.0010 |
| eigen_centrality(0.071) | 1 | 0.2578 | 0.1954 | 0.1525 | 0.1948 |

- 관심있는 특정 노드의 키워드를 선정하여 중심성 확인 필요
- 중심성의 평균을 임계 값으로 선정하고, 키워드들이 기준 이상인지 확인

<네트워크 최종 중심성 분석>



- 실제 관심있는 키워드들이 임계 값 조정 후 노드 간 연결이 되어있는지 출력
- 해당 키워드는 해시태그 간 중심성을 확인하고 新 마케팅 제안 시 참고

■ 활용 및 기대효과

- ✓ 기간별 분석을 통해 인사이트를 창출하여 新 마케팅 의사결정 제안
- ✓ 시즌 별 SNS 해시태그를 통한 마케팅으로 플랫폼 이용률 증가 기대
- ✓ SNS 마케팅 퍼포먼스 검증 및 해시태그 연관성 분석을 통해 향후 마케팅 활동 제안

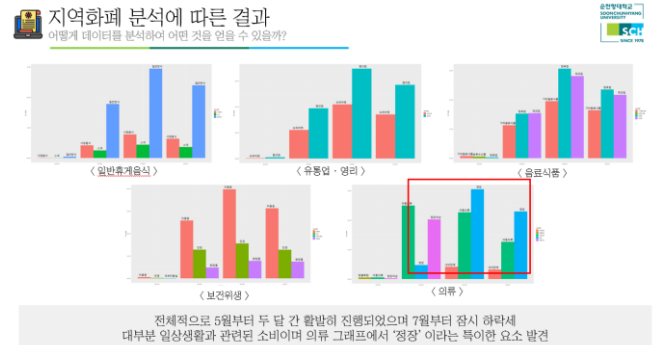
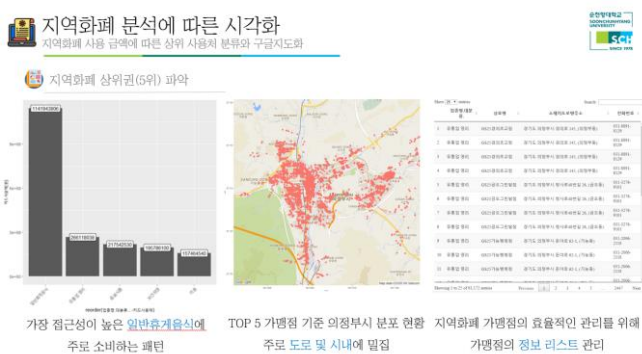
의정부 지역화폐를 활용한 소비패턴 분석 시각화

| | |
|-------|-----------------------------------|
| 프로젝트명 | 의정부 지역화폐를 활용한 소비패턴 분석 시각화 (교내 과제) |
| 수행기간 | 2020.04 ~ 2020.07 (약 3개월) |

■ 분석 방법

- 지역화폐 공공데이터를 R의 **ggplot2**을 활용한 데이터 시각화
- 월별, 금액별, 업종별, 매출 금액별에 따른 의정부 지역화폐 소비 패턴 파악

■ 분석 결과



- 업종별 매출액을 확인하고, 이를 지도 시각화하여 지역 소비 특성을 파악
- 시간에 따른 업종별 사용금액을 통한 지역화폐 소비 추이 파악

| | |
|-------|---|
| 프로젝트명 | 해외 소비자 타겟 선정 및 시장 조사 Study Abroad (교내 과제) |
| 수행기간 | 2019.04 ~ 2019.07 (약 3개월) |

■ 분석 방법

- '동물가위'는 해외 시장에서 어떤 소비자 타겟 층인지 시장조사 활동
- '구매의사'와 '가격결정'에 대한 인구통계학 기반 가설을 수립

■ 분석 결과



- 연령, 성별, 자녀의 유무에 따라 제품의 구매의사의 차이를 확인
- 디자인, 용도에 따른 제품의 가격선정의 차이를 확인

교육훈련 이력

지속적인 경쟁력 확보를 위하여 지원분야에 대한 연구 및 트렌드 파악 등 적극적으로 활동하였습니다!

| | | |
|-----------|----|---|
| 교육과정명 | | 2020년 공공 빅데이터 청년인턴십 확대운영 |
| 주관 / 시행기관 | | 행정안전부/ 한국지능정보사회진흥원 |
| 교육훈련 | 기간 | 2020. 08. 24 ~ 2020. 09. 29/ 총 200시간(약 2개월) |
| | 기관 | 씨에스리 컨소시엄(씨에스리, KPC) |
| 인턴훈련 | 기간 | 2020. 10. 05 ~ 2020. 12. 31 |
| | 기관 | 인천광역시청 데이터혁신담당관 빅데이터팀(납세협력담당관 배치) |

교육내용

공공빅데이터 기획 분석 과정(144시간) : 공공 빅데이터 직무 기본 및 분석 기술 배양

- 빅데이터 직무 기본교육: 4차산업과 공공 빅데이터 개요, 디자인씽킹 창의적 문제 해결법, 데이터 이해 및 데이터셋 모델링, 공공빅데이터 관련 법/제도 및 표준분석모델이해
- 빅데이터 과제 기획: 공공 빅데이터 분석 과제 기획, 빅데이터 분석·시각화 방법론(R, Q-GIS)
- 빅데이터 분석: 외부데이터 수집/정제(Python), 통계분석(R), 공간분석(Q-GIS)
- 행안부 표준분석모델 실습 : 민원분석, 관광/축제 유동인구분석, 전기차 충전소 입지선정 텍스트분석, 최적의사결정, 예측 등 분석모델 주요유형 집중실습

실무형 프로젝트 (56시간) : 문제 해결형 프로젝트(해커톤)로 실무 역량 배양

- 빅데이터 분석과제 기획부터 분석, 시각화까지 분석 Life Cycle 전반에 대한 프로젝트
- 도메인에 대한 이해를 바탕으로 문헌조사 및 사례조사, 현황분석, 핵심 원인 도출, 가설 설정, 해결안 제시까지 진행
- 빅데이터 분석과제 기획, 분석 모델 개발, 데이터 정의, 수집/정제, 데이터 분석/시각화, 결과보고서 작성/발표

취업지원특강 (16시간, 선택교과) : 비즈니스 매너 배양, 취업 경쟁력 확보

- 취 · 창업을 위한 비즈니스 매너 및 이미지 메이킹
- 나를 알리는 이력서 작성과 면접전략