

데이터로 같이, 가치 있게(With Value)!

데이턴십 해커톤 제 5회

부산광역시 정신건강 증진을 위한 빅데이터 기반
마음건강간편테스트 및 마음건강증진센터 최적 입지 선정

분석 결과보고서

참여조: 부산 4조

참여자: 김정훈(조장), 강수연, 김종화
류다미, 박기량, 박상현, 성호준

씨에스리 컨소시엄

CSLEE **kpc** 한국생산성본부

목 차

1. 분석 개요 -----	1
1.1. 분석 배경 및 개요	
1.2. 분석 목적 및 방향 -----	2
1.3. 분석 결과 활용 방안	
2. 분석 데이터 -----	3
2.1. 분석 데이터 목록	
2.2. 데이터 상세 설명 -----	4
2.3. 데이터 정제 방안 -----	9
3. 분석 프로세스 -----	10
3.1. 분석 프로세스	
3.2. 분석 내용 및 방법 -----	12
4. 분석결과 -----	36
4.1. 마음건강테스트	
4.2. 마음건강지수 -----	37
4.3. 마음건강증진센터 최적 입지 선정 -----	38
5. 활용 방안 -----	44
5.1. 문제점 개선 방안	
5.2. 업무 활용 방안	
6. 참고자료(Reference) -----	45
7. 부록 -----	49

1. 분석 개요

1.1. 분석 배경 및 개요

1.1.1. 개요

본 분석은 부산광역시의 정신건강 실태를 다각도로 분석하여 관련 시설이나 정책 및 제도의 활성화와 접근성을 높일 방안을 제시한다. 이를 통해 부산광역시의 정신건강을 증진시키고, 정책에 다방면으로 활용되고자 한다.

정신건강 요인(우울감, 불면증, 불안증)을 키워드로 갖는 유튜브 댓글을 웹 크롤링하여 텍스트 마이닝을 진행하였다. 이후, 감정 분석 및 빈도 분석을 통해 도출된 빈출 키워드들로 간편 마음 건강테스트를 제작하였다. 또한 텍스트 마이닝 결과 마음건강과 교우관계, 가족관계, 학업이 관계가 있는 것을 알 수 있었다.

지역사회보장조사 설문을 바탕으로 데이터 분석 모델을 이용해 마음건강지수를 도출하였다. 최적 입지 선정을 위해 Q-GIS상에서 다른 환경적 요인들과 결합하였다. 최적 입지 선정 분석 결과, 정신건강증진 시설이 필요한 자치구는 북구, 사하구, 해운대구로 나타났다.

감정분석 결과의 키워드는 마음건강증진 프로그램 및 정책을 기획할 때 활용될 수 있을 것이며, 마음건강간편테스트를 통해 인근 마음건강증진센터를 추천하여 관련 서비스에 대한 높은 심리적 장벽을 완화시킬 수 있을 것으로 기대한다. 최적 입지 선정 결과를 바탕으로 마음건강증진센터를 설립하거나 기존 시설을 확충할 때 활용할 수 있을 것이다.

1.1.2. 배경

- 부산시의 정신건강 문제로 인한 자살사망비율과 인구 10만명 당 자살률이 높아 개선이 필요하다.
 - 2022년 기준 인구 10만명 당 자살률 전국 행정구 중 1위 부산광역시 수영구, 3위 부산광역시 중구이다. **(참고자료 1)**
 - 지역 내 5개년(2013~2017) 자살사망자 중 정신건강문제로 인한 자살사망비율이 가장 높은 지역은 부산시(47.8%)이다. **(참고자료 1-1)**
 - 2019년 부산광역시 자살원인 분석 결과 1위 정신적, 정신과적 문제(307명, 33.0%), 2위 경제생활 문제(217명, 23.4%), 3위 육체적 질병문제(212명, 22.8%)로 정신적 문제로 인한 자살이 가장 높다. **(참고자료 1-2)**
- 부산시의 정신건강관련 정책 및 제도가 미흡하고 기존 시설과 정책 또한 이용률이 낮아 대책 마련이 시급하다.
 - 부산은 시민들과 부산의 미래 발전을 위해 핵심 정책들을 추진하고 있지만 정신건강 관련은 포함되어 있지 않은 실정. **(참고자료 2)**
 - 2018년 12월 말 기준 부산의 정신재활입소시설 정원은 인구 10만 명당 1.9명에 불과함. 이는 전국 17개 시·도 중 13위로 대전 15.1명, 충남 14.9명,

전북 12.7명 등에 비해 현저히 낮은 수준. (참고자료 2-1)

- 인천 정신건강복지센터 월별 평균 상담건수 약 8회에 비해 부산 정신건강복지센터 월별 평균 이용은 약 3회 (2022년 7월 기준) (참고자료 2-3)

1.2. 분석 목적 및 방향

1.2.1. 목적

정신건강 실태 다각도 분석을 바탕으로 관련 시설 접근성 재고와 정책 및 제도 활성화. 이를 통해 부산시민의 마음건강을 증진시키고 다양한 정책에 활용되고자 한다.

1.2.2. 방향

정신건강 관련 키워드를 제목으로 하는 유튜브 댓글 텍스트 마이닝 및 감정분석을 진행한다. 이를 토대로 마음건강테스트 문항을 도출해 심리적 접근성 개선.

지역사회보장조사 설문 데이터를 분류 모델을 통해 데이터 분석 실시하여 마음건강 지수 도출

QGIS 활용, 마음건강지수와 시설 접근성 및 환경등을 고려한 마음건강증진센터 최적 입지 선정으로 물리적 접근성 개선

1.3. 분석 결과 활용 방안

1.3.1. 텍스트 마이닝 결과를 바탕으로 마음건강간편테스트 제작

정신건강 요인 (우울감, 불안감, 불면증 등)을 키워드로 한 유튜브 콘텐츠 댓글 텍스트마이닝을 통한 마음건강간편 테스트 제작

1.3.2. 마음건강지수 개발

부산시 사회보장조사를 바탕으로 정신건강 및 각종 범주들의 분석을 통해 마음건강에 영향을 미치는 요인 선정 및 요인을 바탕으로 한 정신건강위험 수준 파악을 위한 마음건강지수 도출

1.3.3. 인구밀도, 교통시설현황, 정신건강지수를 통한 마음건강증진센터 입지분석
분석 데이터기반 마음건강증진센터 시설 건립 및 확충의 타당성 확보, 부산 시민의 시설 접근성 증대

2. 분석 데이터

2.1. 분석 데이터 목록

2.1.1. 활용 데이터

[표 2-1] 활용 데이터

활용 데이터	구분	중요도	생성 주기	지역 속성	데이터 소스
유튜브	비정형/외부	필수	실시간	전국	- 정신건강 요인(우울감, 스트레스, 불면증 등)을 키워드로 한 국내 유튜브 콘텐츠 댓글
감성 대화 말뭉치	정형	필수	-	-	AI-Hub 우울증 환자 대상 WOZ 대화 수집 텍스트 데이터
지역사회보장조사	정형	필수	4년	부산시	공공데이터포털 - 시군구, 읍면동, 정신건강, 신체건강, 문화여가, 주거 등 만족도 설문조사
주거인구	비정형	필수	-	전국	국가공간정보포털 - 광역시명, 광역시코드, 시군구명, 시군구코드, X좌표, Y좌표, 세대수, 인구수, 10대, 20 30대, 40대, 50대, 60대 이상 인구수
부산광역시_정신보건시설 현황	정형/외부	선택	1년	부산시	공공데이터포털 - 시설명, 도로명주소, 위도, 경도, 전화번호, 유형, 정원
행정구역_읍면동	비정형	필수	수시	부산시	국가공간정보포털 - 읍면동 단위 경계도면

부산시 버스 정류소 정보	비정형	필수	연간	부산시	공공데이터포털 - 부산광역시 버스정류소 정보(정류소 ID, 정류소명, GPS좌표, 정류소구분 등) 지도 표출용 공간정보
국가철도공단_부산 1,2,3,4호선 역위치	정형	필수	연간	부산시	공공데이터포털 - 도시광역시 철도역의 선명, 경도, 위도 데이터
수치표고모델 (DEM)_90M	비정형	필수	수시	전국	국토교통부 국가공간정보포털 오픈마켓, 국토지리정보원
부산광역시 공동주택현황	정형	선택	연간	부산시 (구별)	부산 공공데이터 포털 - 소재지, 층수, 동수, 세대수, 면적, 준공일자, 철거일

2.2. 데이터 상세 설명

2.2.1. 텍스트마이닝 사용 데이터

□ 유튜브

[표 2-2] 우울 키워드 URL

Index	제목	URL
1	우울증 #우울증증상 #우울증치료 - [PEOPLE in 세브란스] 몸이 보내는 우울증 신호 세 가지	https://www.youtube.com/watch?v=jME5_dk3mkQ
2	우울증을 경험할 때 나타나는 10가지 증상 [심리, 정신건강]	https://www.youtube.com/watch?v=RISEiEKBacY
3	우울증 직전 증상은? 심리상담사의 답변 3분 찌리얼 시선	https://www.youtube.com/watch?v=UmuwwDBZxGo
4	#우울증 #극복하기 #치료 - 마음의 감기, 우울증 극복하기! 해결의 법칙 -서울아산병원 김병수 정신과 중년 우울	https://www.youtube.com/watch?v=mHWV7ELxB5o

5	[PLAYLIST] 유독 우울한 밤에 듣기 좋은 감성팝송	https://www.youtube.com/watch?v=_ltzj3qRWV0
6	자살 생각들만큼 우울할 때 들으면 더 우울해지는 노래	https://www.youtube.com/watch?v=ic3i7MI69Ys

[표 2-3] 불면 키워드 URL

Index	제목	URL
1	10시간 잔잔한 수면음악 ♪ 스트레스 해소음악, 잠잘때 듣는 음악, 불면증치료음악, 수면유도음악 (My Dream)	https://www.youtube.com/watch?v=p2fxv3PAAtLU&t=22566s
2	# 밤에 듣기 좋은 감성 발라드 24곡 [가사첨부] PLAYLIST	https://www.youtube.com/watch?v=_qmcg4vqmVg
3	# □zzz 최상의 숙면을 경험하세요. 갸름 없는 깊은 잠을 돕는 수면단계별 뇌파 동조화 수면 사운드 - 90분 수면사이클 버전 & 기상알람 2회 (연구 논문 기반)	https://www.youtube.com/watch?v=NSNY8qDUKqw
4	# 10분 안에 마취시켜주는 수면유도 음악 잠 잘 때 듣는 음악	https://www.youtube.com/watch?v=Wlqe9vM4U34&t=121s
5	# ★불면증에 효과적인 뇌파소리 숙면델타파 2.0hz '회복수면' + 장작 불소리 2.0hz EEG + Calm Fire Sound	https://www.youtube.com/watch?v=Bb0d96fC7bc&t=1992s
6	# 잠 잘 때 듣는 음악 클래식 명곡 1시간 잠잘오는음악 수면음악 편안한 잔잔한 음악 잠오는 클래식 자장가 힐링음악 Classical Music for Sleeping	https://www.youtube.com/watch?v=bdjyo0qsejl
7	# 10분 안에 잠드는 꿀잠 수면명상 [숙면을 위한 수면유도 음악]	https://www.youtube.com/watch?v=iHCCwk3CV70&t=26s

8	# ☹빠르게 깊은 잠에 드는 뇌파 소리 1.8 Hz 델타파 - '8시간의 꿈 없는 잠' 1.8 Hz EEG	https://www.youtube.com/watch?v=9k0faMFph4w
9	# 아침이 달라지는 효과적인 수면법! 5시간 파워 숙면 뇌파소리!할 일이 많은 날 고효율 5시간 숙면!	https://www.youtube.com/watch?v=LysA1OKA5Tk

[표 2-4] 불안 키워드 URL

Index	제목	URL
1	불안한 마음을 위한 힐링음악☯스트레스해소음악,명상음악,이완음악,스파음악,수면음악 -"Warm Candle"	https://www.youtube.com/watch?v=cRbYxsLHyJU
2	불안한 하루를 보내고있는 너에게 (Playlist)	https://www.youtube.com/watch?v=EDsPY0d9s4E
3	[Playlist] 내가 불안할 때 듣는 노래	https://www.youtube.com/watch?v=j9Oa4GyVXOk
4	[PLAYLIST] 마음이 불안할 때, 아무 생각 없이 듣기 좋은 팝송 10곡	https://www.youtube.com/watch?v=2HhvS8A0RFQ
5	[Playlist] 괜찮아, 다 잘될거야. 불안한 당신을 위로해주는 노래♪	https://www.youtube.com/watch?v=uGa2cN1N5-Y
6	불안할 때, 마음의 안정을 찾아주는 노래 모음 PLAYLIST	https://www.youtube.com/watch?v=HfAhUxPSubs
7	(Playlist) 모든 관계가 불안하다고 느낄 때 듣는 노래	https://www.youtube.com/watch?v=0tyTmVswhjM

[표 2-5] 키워드별 댓글 개수

키워드명	불면	불안	우울	합계
개수	15,495	5,858	7,184	28,537

(특수문자, 영문 댓글, 불필요한 형태소 등 제외)

2.2.2. 마음건강지수 도출 사용 데이터

□ 지역사회보장조사

- 제5기 지역사회보장계획수립을 위한 지역주민욕구조사에 대한 설문조사 결과 데이터로 정신건강지수 도출을 위함
- 부산시 전체 8,000가구를 시군구별로 500가구씩 설문조사 진행
- 설문조사 구성 문항(5단계 리커트 척도, 숫자가 클수록 자주 겪음(부정적))

[표 2-6] 부산시 지역사회보장조사 데이터 상세

컬럼ID	컬럼명	상세 설명
ID	응답자	응답자 번호
시군구_CODE	시군구	부산광역시 14개 자치구
읍면동_CODE	읍면동	자치구 내의 동
B4_1_1 ~ B4_1_5	정신건강	1년 동안 정신건강, 정신질환, 중독, 자살관련 문제, 트라우마 등으로 인한 어려움 경험 정도
B5_1_1 ~ B5_1_2	신체건강	1년 동안 신체적 질환과 일상적인 신체 건강 관리의 어려움 경험 정도
B6_1_1 ~ B6_1_6	기초생활 유지	1년 동안 기초 생활을 유지하기 위한 경제적 어려움 경험 정도
B7_1_1 ~ B7_1_2	가족 및 사회적 관계	1년 동안 가족이나 이웃, 친인척 등과의 갈등 경험 정도
B8_1_1 ~ B8_1_2	보호안전	가족 구성원간 혹은 이웃, 친인척 등으로부터의 폭력 경험 정도
B9_1_1 ~ B9_1_4	교육	(초중고 재학생이 있는 가구만 응답) 학습에 관련한 어려움 경험 정도
B10_1_1 ~ B10_1_9	고용	(취창업 경험을 한 경우만 응답) 취창업과 고용 유지에 있어서 어려움 경험 정도
B11_1_1 ~ B11_1_11	주거	주거 시설의 열악함 경험 정도
B12_1_1 ~ B12_1_2	법률 및 권익보장	법률 문제와 권익침해 등으로 인한 어려움 경험 정도
B13_1_1 ~ B13_1_2	문화여가	문화여가와 체육 활동 관련 어려움 경험 정도

2.2.3. 마음건강증진센터 최적 입지 선정 사용 데이터

□ 주거인구

마음건강증진센터 최적 입지 선정을 위해 부산시 행정구역별 인구밀도 도출을 위함

□ 부산광역시_정신보건시설 현황

마음건강증진센터 최적 입지 선정을 위해 기존의 정신보건시설 현황을 파악하기 위함

□ 행정구역_읍면동

마음건강증진센터 최적 입지 선정을 위한 부산시 동/구별 레이어 생성을 위함

[표 2-7] 행정구역_읍면동 데이터 상세

컬럼ID	컬럼명
EMD_CD	읍면동코드
EMD_NM	읍면동명
SGG_OID	원천오브젝트ID
COL_ADM_SE	원천시군구코드
GID	도형ID

□ 부산시 버스정류소정보

마음건강증진센터 최적 입지 선정을 위한 부산시 대중교통 현황과 접근성 파악을 위함

[표 2-8] 부산시 버스정류소정보 데이터 상세

컬럼ID	컬럼명
bstoparsno	버스정류장 ARS number
bstopid	버스정류소ID
name	버스정류소명
stoptype	정류소구분
lng	경도
lat	위도

□ 국가철도공단_부산1,2,3,4호선 역위치

마음건강증진센터 최적 입지 선정을 위한 부산시 대중교통 현황과 접근성 파악을 위함

□ 수치표고모델(DEM)_90M

마음건강증진센터 최적 입지 선정을 위한 부산시 지형의 고도값을 수치로 저장함으로써 지형의 형상을 나타내는 지도로 입지 선정에 있어 접근성 파악을 위함

□ 부산광역시 공동주택현황

마음건강증진센터 최적 입지 선정을 위해 부산시 인구밀도 도출을 위함

2.3. 데이터 정제 방안

2.3.1 텍스트 마이닝 (마음건강간편테스트)

- 불필요한 문장 제외
- 영문, 특수문자, 자음 단순 반복과 같은 형태의 문장 제외
- 불용어 처리
- 불용어 사전에 포함된 단어 1,070개 제외
- 분석에 무의미한 단어 제외(형용사, 부사 등)

2.3.2. 마음건강지수

- 데이터 파악
- 설문지의 문항들이 컬럼으로, 응답한 가구들이 로우로 구성되어 있고 시/군/구별로 500가구씩 총 8,000가구가 포함되어 있음. 즉, 컬럼의 수는 496개, 로우의 수는 8,000개
- 데이터 척도 파악
- 5단계 리커트 척도로 작성된 설문지 결과를 등간척도로 간주하였음
- 컬럼명 변경
- 효율적인 분석을 위해 최소한의 의미를 담은 컬럼명으로 변경
- 결측치 제거
- 해당하는 가구만 응답해야 하는 문항의 경우, 응답수가 현저히 적어 제거
- 설문지 문항들의 유의성 검정
- 데이터를 구성하고 있는 설문지 문항들의 통계적 유의미 검정을 위해 크론바흐 알파 계수를 사용
- 파생변수 생성
- 리커트 척도의 합과 평균은 분석 목적에 따라 무엇을 사용할 지 결정
- 정신건강의 정도를 진단하기 위함이 목적이므로 척도의 합을 분석에 활용
- 문항의 합이 커질수록 해당 변수와 관련된 어려움을 자주 겪었다고 간주함

2.3.3. 마음건강증진센터 최적 입지분석

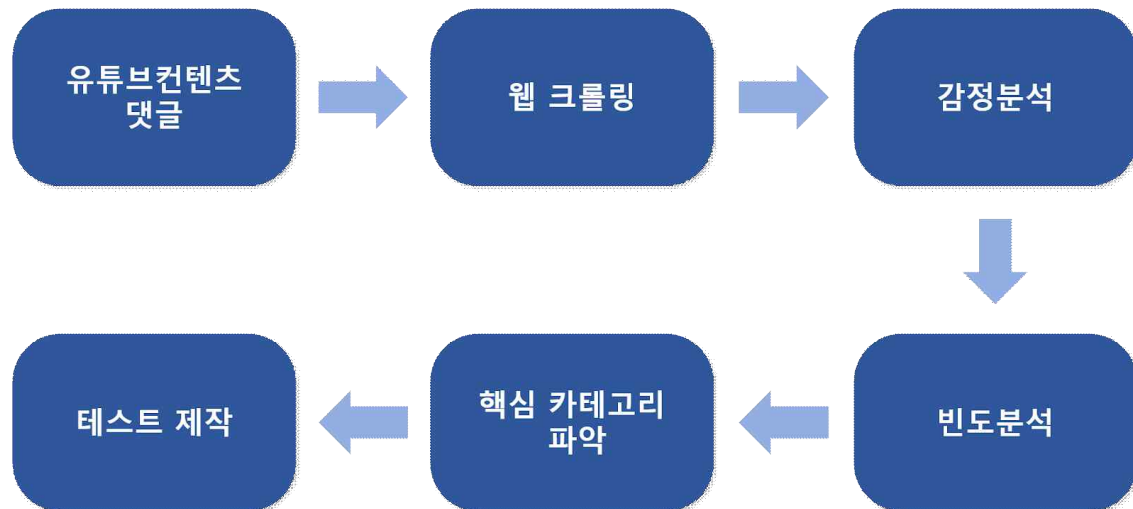
- 결측값 대체
- 행정동 기준의 마음건강지수의 결측값(금성동)을 주변 값의 산술평균으로 대체
- 데이터 전처리(정규화)
- Min-Max 정규화로 입지분석 변수 간 편차 축소
- csv 데이터 결합
- 세분화된 동별 인구 데이터를 103동 기준으로 동별 csv 데이터 결합
- 세분화된 동별 마음건강지수 데이터를 103동 기준으로 동별 csv 데이터 결합
- 폴리곤 결합
- 세분화된 동별 폴리곤 데이터를 103동 기준으로 동별 레이어 데이터 결합

3. 분석 프로세스

3.1. 분석 프로세스

3.1.1. 마음건강간편테스트 제작 과정(Python 활용)

- 정신건강 요인(우울감, 불안감, 불면증 등)을 키워드로 한 유튜브 콘텐츠 댓글 웹 크롤링
- 감정분석을 통한 부정적 댓글 추출
 - tensorflow keras 사용
 - 활성화 함수 : ReLU, Sigmoid
- 데이터 전처리 (불용어 사전 포함 단어, 분석에 무의미한 단어 제외)
- 빈도분석을 통한 핵심 카테고리 파악
- 테스트 제작

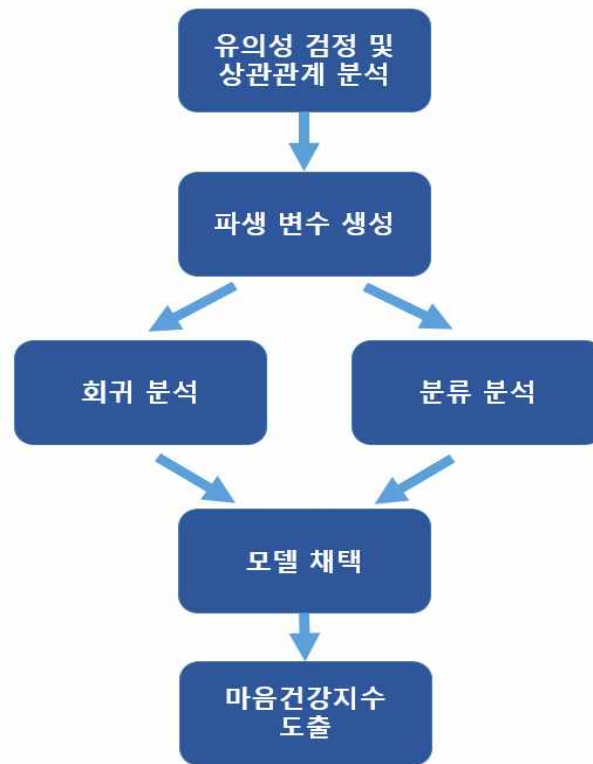


[그림 3-1] 마음건강간편테스트 제작 분석프로세스 도식화

3.1.2. 마음건강지수 도출 과정(Python 활용)

- 종속변수 및 독립변수 하위 문항 간의 상관관계 분석 및 시각화
 - 스피어만 상관계수 사용
 - 히트맵을 통한 상관관계 시각화
- 상관계수가 낮은 하위 문항들을 제거하고 합산하여 파생변수 생성
- 종속변수와 독립변수간의 상관관계 분석
 - 스피어만 상관계수 사용
 - 상관계수의 절대값을 기준으로 순위 분석
- 회귀분석 및 회귀계수 순위 분석
 - LinearRegression, RandomForestRegression, LGBMRegression 모델 사용
 - LightGBM Boosting 모델 사용
 - 모델 Feature Importance 시각화
- 분류분석 및 Confusion Matrix 결과 분석

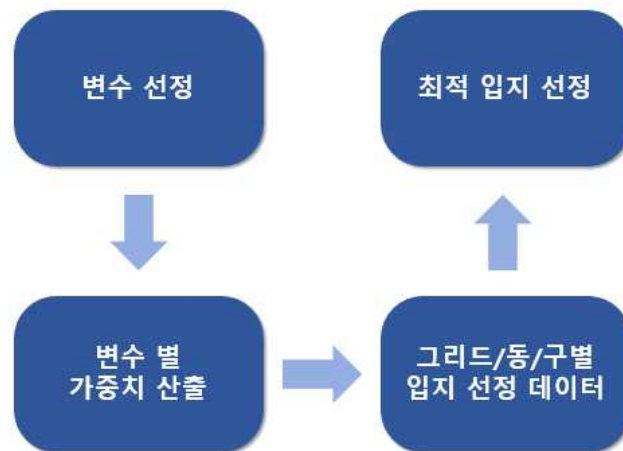
- RandomForestClassifier 모델 사용
- 종속변수의 가능한 값(2~10)으로 총 9개로 분류하여 예측
- 모델 Feature Importance 시각화
- 회귀분석과 분류분석의 모델 성능 비교 및 적용할 최종 모델 채택
- 모델 기반 부산시 시/군/구 및 읍/면/동별 마음건강지수 도출



[그림 3-2] 마음건강지수 도출 분석프로세스 도식화

3.1.3. 마음건강증진센터 최적 입지 선정 과정 (QGIS 활용)

- 마음건강증진센터의 최적 입지 선정 과정은 4단계로 최적 입지 선정을 위한 변수 선정, 최적 입지 선정에 영향을 주는 변수별 가중치 산출, 그리드/동/구별 입지 선정 데이터, 부산시 마음건강증진센터 최적 입지 선정 과정 순으로 진행
- 변수 선정 단계에서는 입지 선정에 영향을 주는 요소를 파악하고 변수를 선정
- 변수별 가중치 산출 단계에서는 계층적 의사결정법(AHP, Analytic Hierarchy process)을 사용하여 가중치를 산출
- 그리드/동/구별 입지 선정 데이터 단계에서는 QGIS툴을 활용하여 지도상에 시각화 및 지역, 구역별 데이터를 생성
- 최적 입지 선정 단계에서는 Python을 사용하여 앞의 과정에서 구한 지역, 구역별 데이터를 정규화 및 가중치 부여를 통해 순위를 구하고, 순위에 따라 지도상에 표시



[그림 3-3] 마음건강증진센터 최적 입지 선정 분석프로세스 도식화

3.2. 분석 내용 및 방법

3.2.1. 마음건강간편테스트

□ 감정분석을 통한 부정적 댓글 추출 (tensorflow keras 사용)

- 우울증 환자 1,500명의 인터뷰 데이터(원본 출처: Ai Hub)를 긍·부정적 문장으로 분류(전처리)하여 지도 학습을 위한 데이터셋 생성
- 해당 학습데이터를 ReLU, Sigmoid 활성화 함수를 활용하여 모델 제작

258/258 [=====] - 1s 5ms/step - loss: 0.2509 - accuracy: 0.8978

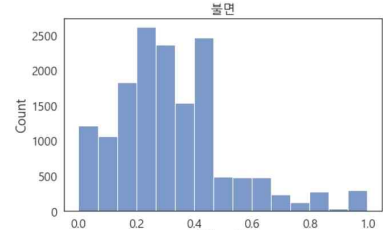
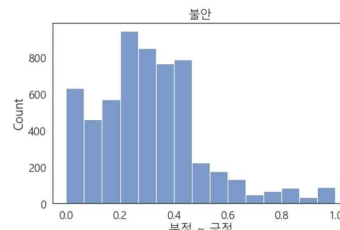
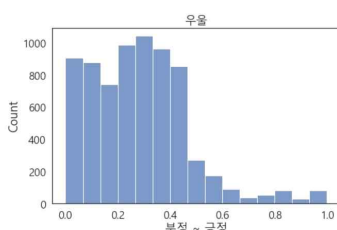
[0.25093764066696167, 0.8978182077407837]

[모델 정확도 89.78%]

□ 5개년 전국 자살사망자 분석 보고서 참고하여 마음건강 핵심 변수(우울, 불안, 불면) 3개 설정.

□ 핵심 변수 3개를 키워드로 하여 유튜브 영상 22개의 댓글 웹 크롤링 후 전처리 된 댓글 총 23,720개 수집

□ 감정분석을 통해 부정적 댓글 추출



[그림 3-4] 우울 감정지수
히스토그램

[그림 3-5] 불안 감정지수
히스토그램

[그림 3-6] 불면 감정지수
히스토그램

□ 부정 댓글 수 / 전체 댓글 수

- 우울 : 6,157 / 7,183 = 0.857
- 불안 : 4,853 / 5,858 = 0.828
- 불면 : 12,823 / 15,491 = 0.828

□ 부정 감정 지수 평균 (부정 댓글 감정지수 평균[0에가까울수록 부정])

- 우울 : 0.286
- 불안 : 0.307
- 불면 : 0.321

□ 빈도분석을 통해 관련성 가장 높은 키워드 추출

우울 단어	빈도수	비율
친구	854	0.109
엄마	682	0.087
마음	539	0.069
눈물	466	0.060
사랑	382	0.049
공부	364	0.047
아빠	359	0.046
감정	358	0.046
인생	356	0.046
세상	353	0.045
위로	353	0.045
부모	341	0.044
가족	335	0.043
상담	331	0.042
자살	325	0.042
우울	317	0.041
하루	296	0.038
매일	269	0.034
학교	268	0.034
기분	262	0.034
합계	7810	1.000



[그림 3-7] 우울 단어 빈도 및 워드클라우드

불안 단어	빈도수	비율
친구	759	0.124
위로	674	0.110
마음	476	0.078
하루	452	0.074
공부	433	0.071
사랑	429	0.070
인생	384	0.063
눈물	286	0.047
엄마	249	0.041
세상	238	0.039
내일	226	0.037
포기	198	0.032
시험	193	0.031
걱정	183	0.030
기분	183	0.030
수고	181	0.030
상처	154	0.025
학교	148	0.024
아빠	146	0.024
부모님	143	0.023
합계	6135	1.000



[그림 3-8] 불안 단어 빈도 및 워드클라우드

불문 단어	빈도수	비율
엄마	1333	0.136
아빠	1022	0.105
사랑	800	0.082
마음	664	0.068
할머니	521	0.053
아침	496	0.051
친구	434	0.044
매일	433	0.044
하루	421	0.043
공부	415	0.042
효과	400	0.041
수면	365	0.037
내물	352	0.036
나일	343	0.035
새벽	319	0.033
자고	305	0.031
걱정	297	0.030
할아버지	290	0.030
기억	283	0.029
세상	282	0.029
합계	9775	1.000



[그림 3-9] 불면 단어 빈도 및 워드클라우드

□ 키워드를 카테고리화하여 테스트 문항 제작

가족	엄마, 아빠, 부모님, 가족, 할머니, 할아버지
교우	친구
학업	시험, 공부, 선생님, 학교

[그림 3-10] 키워드별 카테고리 분류

☐ 질문 문항을 바탕으로 테스트 제작

3.2.2. 마음건강지수

□ 결측치 제거

- B9(교육) 영역에서 전체 8000명 중 결측치가 7,147명으로 확인되어 B9(교육) 변수는 제거하기로 결정
- B10(고용) 영역에서 전체 8000명 중 결측치가 4,639명으로 확인되어 B10(교육) 변수는 제거하기로 결정

□ 설문지 문항들의 유의성 검정

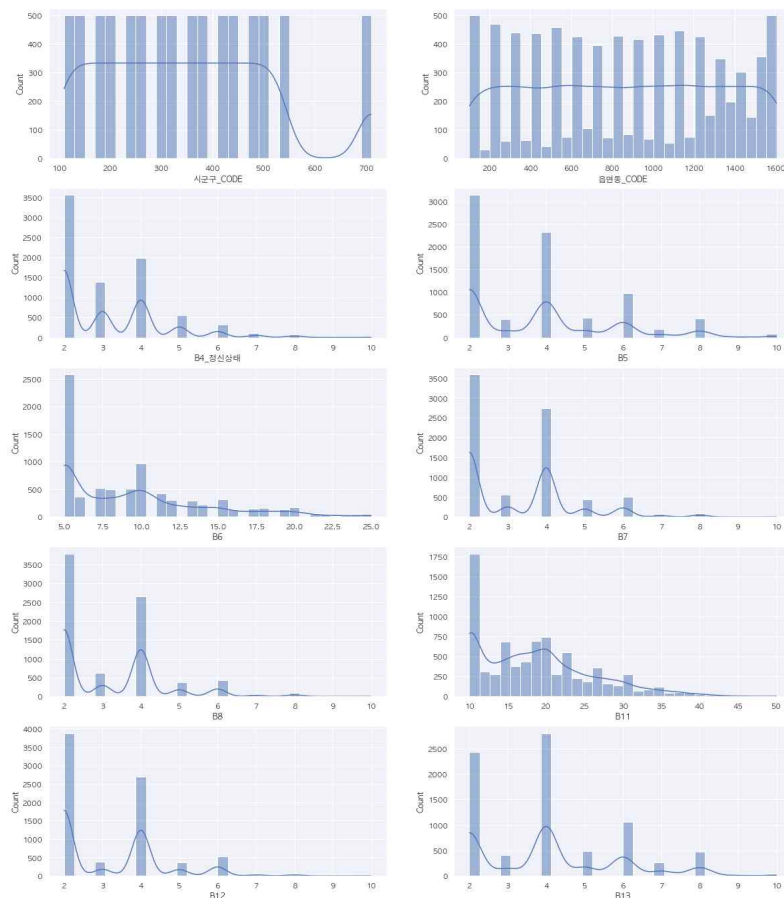
- 크론바흐 알파계수는 리커트척도로 작성된 설문조사 문항의 일관성을 평가하는 척도이다. 계수는 0~1 사이의 값을 가지며 1에 가까울수록 문항들 간의 신뢰도가 높다고 해석된다. 보통 0.8~0.9의 값이면 신뢰도가 매우 높은 것으로 보며, 0.7 이상이면 바람직한 것으로 본다.

[표 3-1] 크론바흐 알파 분석 결과

변수명	크론바흐 알파 계수
B4(정신건강)	0.9007
B5(신체건강)	0.9505
B6(기초생활유지)	0.9274
B7(가족 및 사회적 관계)	0.9110
B8(보호-안전)	0.8771
B11(주거)	0.9517
B12(법률 및 권익 보장)	0.9328
B13(문화-여가)	0.9352

- 분석 결과 모든 항목이 0.8을 넘으므로 유의미한 설문조사 데이터라고 할 수 있다.

□ 데이터 분포 확인



[그림 3-11] 데이터 분포 확인

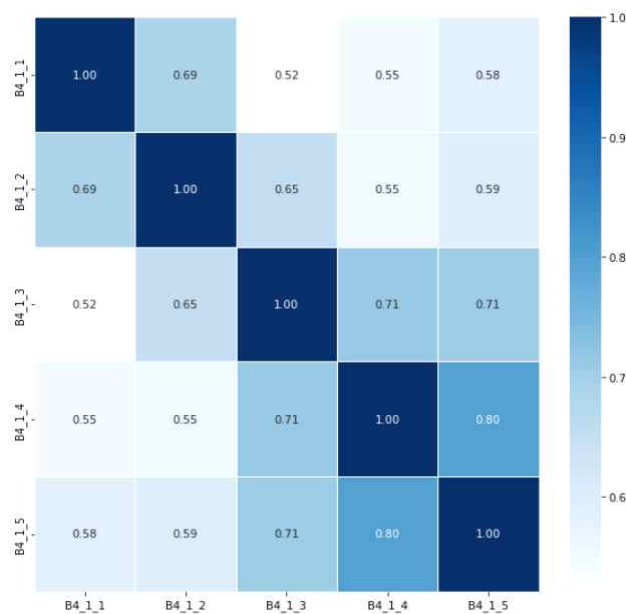
- 대부분의 문항에 1, 2로 응답했음을 알 수 있다.

□ 종속변수 하위 문항 간의 상관관계 분석

- B4(정신건강) 하위 문항 간의 상관관계 분석
- 스피어만 상관계수 사용(소수점 아래 다섯째 자리에서 반올림)
 - B4_1_1 : 정신건강증진의 어려움
 - B4_1_2 : 중증정신질환으로 인한 어려움
 - B4_1_3 : 중독으로 인한 어려움
 - B4_1_4 : 자살관련 문제로 인한 어려움
 - B4_1_5 : 트라우마로 인한 어려움

[표 3-2] B4(정신건강) 하위 문항 간의 상관계수

변수	B4_1_1	B4_1_2	B4_1_3	B4_1_4	B4_1_5
B4_1_1	1	0.6865	0.5211	0.5463	0.5813
B4_1_2	0.6865	1	0.6527	0.5516	0.5925
B4_1_3	0.5211	0.6527	1	0.7121	0.7075
B4_1_4	0.5463	0.5516	0.7121	1	0.7981
B4_1_5	0.5813	0.5925	0.7075	0.7981	1



[그림 3-12] B4(정신건강) 하위 문항 간의 상관관계 히트맵

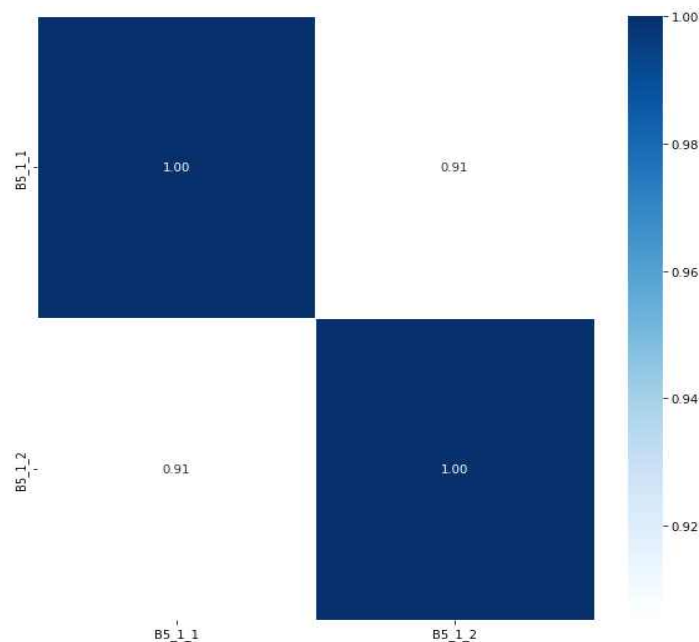
- B4_1_1과 B4_1_2은 정신건강/질환으로 인한 어려움에 대한 항목으로 서로 상관관계를 보이고, B4_1_3 ~ B4_1_5은 중독, 자살, 트라우마로 인한 어려움에 대한 항목으로 서로 상관관계를 보인다. 따라서 B4_1_3 ~ B4_1_5을 제외한 B4_1_1과 B4_1_2를 종속변수로 채택

□ 독립변수 하위 문항 간의 상관관계 분석

- B5(신체건강) 하위 문항 간의 상관관계 분석
- 스피어만 상관계수 사용(소수점 아래 다섯째 자리에서 반올림)
 - B5_1_1 : 신체적 질환으로 인한 어려움
 - B5_1_2 : 일상적인 신체 건강 관리의 어려움

[표 3-3] B5(신체건강) 하위 문항 간의 상관계수

변수	B5_1_1	B5_1_2
B5_1_1	1	0.9052
B5_1_2	0.9052	1



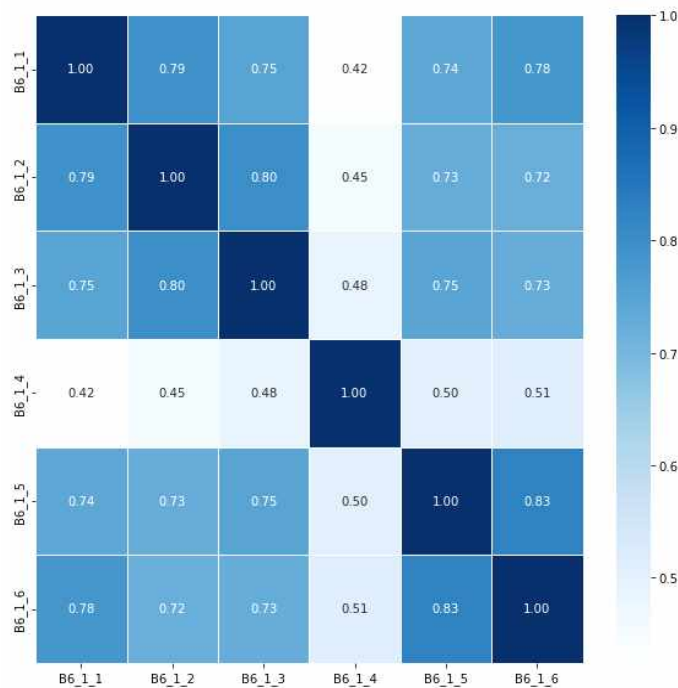
[그림 3-13] B5(신체건강) 하위 문항 간의 상관관계 히트맵

- B6(기초생활유지) 하위 문항 간의 상관관계 분석
- 스피어만 상관계수 사용(소수점 아래 다섯째 자리에서 반올림)
 - B6_1_1 : 생계비 부족
 - B6_1_2 : 주거비 부족(주택구입비 제외)

- B6_1_3 : 의료비 부족
- B6_1_4 : 교육비 부족
- B6_1_5 : 냉난방비 부족
- B6_1_6 : 가계 재정 관리의 어려움

[표 3-4] B6(기초생활유지) 하위 문항 간의 상관관계수

변수	B6_1_1	B6_1_2	B6_1_3	B6_1_4	B6_1_5	B6_1_6
B6_1_1	1	0.7928	0.7501	0.4244	0.7398	0.7809
B6_1_2	0.7928	1	0.7994	0.4550	0.7272	0.7228
B6_1_3	0.7501	0.7994	1	0.4842	0.7472	0.7260
B6_1_4	0.4244	0.4550	0.4842	1	0.5043	0.5083
B6_1_5	0.7398	0.7272	0.7472	0.5043	1	0.8260
B6_1_6	0.7809	0.7228	0.7260	0.5083	0.8260	1

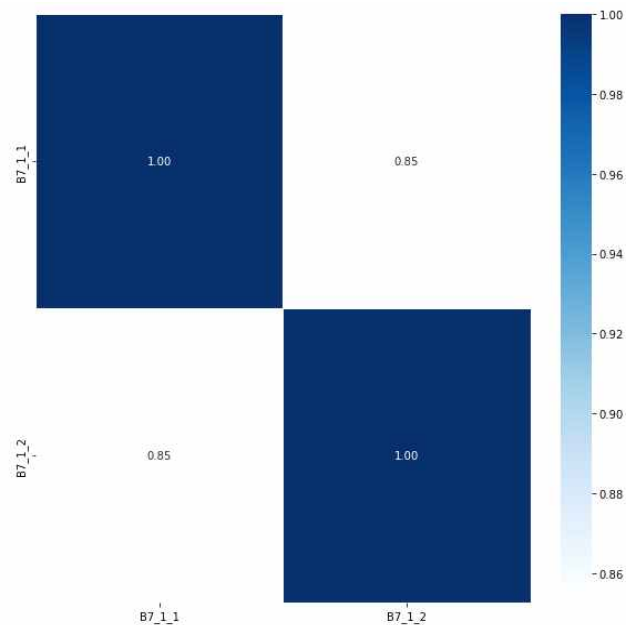


[그림 3-14] B6(기초생활유지) 하위 문항 간의 상관관계 히트맵

- B6_1_4 문항은 다른 문항과의 상관관계수가 0.7 이하이므로 제거
- B7(가족 및 사회적 관계) 하위 문항 간의 상관관계 분석
- 스피어만 상관관계수 사용(소수점 아래 다섯째 자리에서 반올림)
 - B7_1_1 : 가족관계에서의 어려움
 - B7_1_2 : 사회적 관계에서의 어려움

[표 3-5] B7(가족 및 사회적 관계) 하위 문항 간의 상관계수

변수	B7_1_1	B7_1_2
B7_1_1	1	0.8256
B7_1_2	0.8526	1

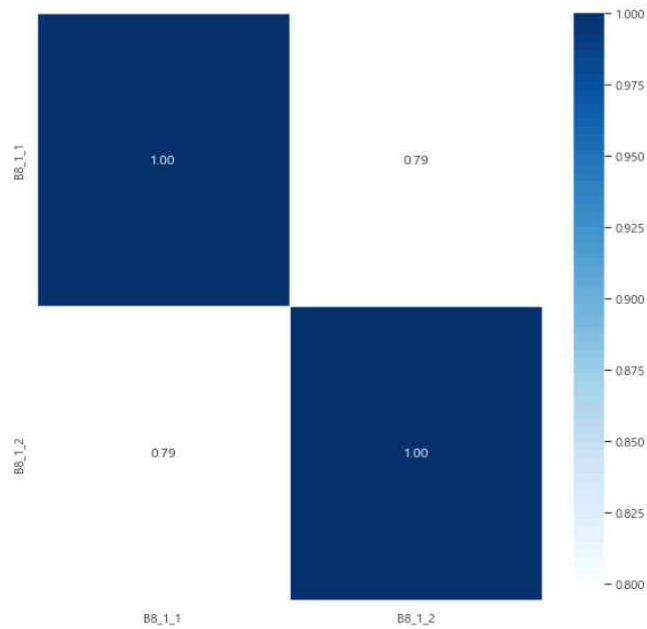


[그림 3-15] B7(가족 및 사회적 관계) 하위 문항 간의 상관관계 히트맵

- B8(보호안전) 하위 문항 간의 상관관계 분석
- 스피어만 상관계수 사용(소수점 아래 다섯째 자리에서 반올림)
 - B8_1_1 : 가족 내 안전유지의 어려움
 - B8_1_2 : 가족 외부로부터 안전유지의 어려움

[표 3-6] B8(보호안전) 하위 문항 간의 상관계수

변수	B8_1_1	B8_1_2
B8_1_1	1	0.7941
B8_1_2	0.7941	1



[그림 3-16] B8(보호안전) 하위 문항 간의 상관관계 히트맵

- B11(주거) 하위 문항 간의 상관관계 분석
- 스피어만 상관계수 사용(소수점 아래 다섯째 자리에서 반올림)
 - B11_1_1 : 화장실 열악
 - B11_1_2 : 주방시설 열악
 - B11_1_3 : 위생 상태 열악
 - B11_1_4 : 도배,장판 열악
 - B11_1_5 : 층간방음 열악
 - B11_1_6 : 냉난방 열악
 - B11_1_7 : 전기, 가스시설 열악
 - B11_1_8 : 상하수도 열악
 - B11_1_9 : 지붕노후(누수 등)
 - B11_1_10 : 벽, 담 등 노후
 - B11_1_11 : 주택 내 장애물로 인한 이동의 어려움

[표 3-7] B11(주거) 하위 문항 간의 상관관계수

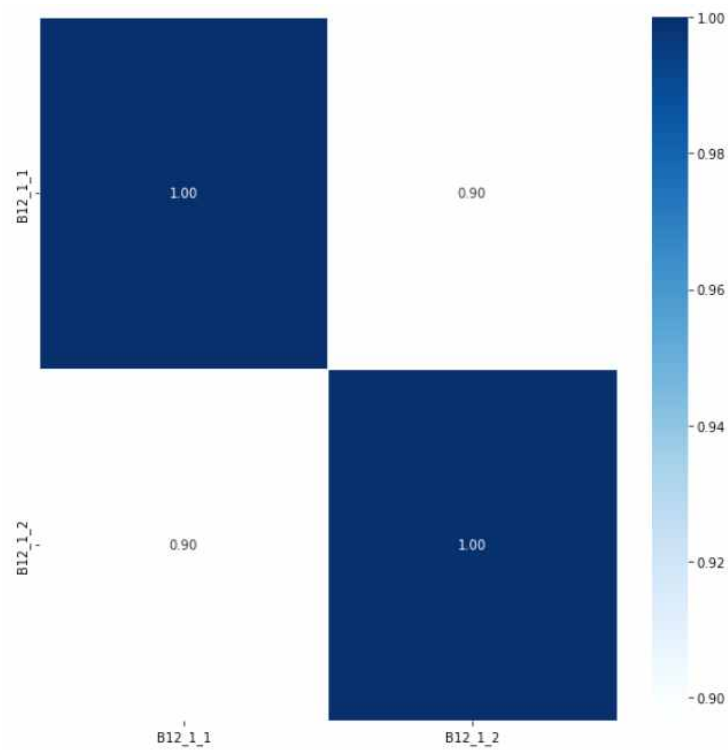
변수	B11_1_1	B11_1_2	B11_1_3	B11_1_4	B11_1_5	B11_1_6	B11_1_7	B11_1_8	B11_1_9	B11_1_10	B11_1_11
B11_1_1	1	0.7931	0.7084	0.6863	0.5227	0.7012	0.6520	0.6390	0.6577	0.6633	0.6634
B11_1_2	0.7931	1	0.7554	0.6523	0.4832	0.6932	0.6659	0.6425	0.6512	0.6581	0.6527
B11_1_3	0.7084	0.7554	1	0.7368	0.4573	0.6567	0.6551	0.6423	0.6487	0.6469	0.6448
B11_1_4	0.6863	0.6523	0.7368	1	0.5344	0.6154	0.5897	0.6206	0.6375	0.6372	0.6161
B11_1_5	0.5227	0.4832	0.4573	0.5344	1	0.5667	0.4579	0.4613	0.4792	0.4940	0.5170
B11_1_6	0.7012	0.6932	0.6567	0.6154	0.5667	1	0.7286	0.6246	0.6566	0.6789	0.6849
B11_1_7	0.6520	0.6659	0.6551	0.5897	0.4579	0.7286	1	0.7493	0.6726	0.7044	0.7287
B11_1_8	0.6390	0.6425	0.6423	0.6206	0.4613	0.6246	0.7493	1	0.7198	0.7118	0.7399
B11_1_9	0.6577	0.6512	0.6487	0.6375	0.4792	0.6566	0.6726	0.7198	1	0.8049	0.7923
B11_1_10	0.6633	0.6581	0.6469	0.6372	0.4940	0.6789	0.7044	0.7118	0.8049	1	0.8617
B11_1_11	0.6634	0.6527	0.6448	0.6161	0.5170	0.6849	0.7287	0.7399	0.7923	0.8617	1

- B11_1_5 문항은 다른 문항과의 상관관계수가 0.7 이하이므로 제거

- B12(법률 및 권익보장) 하위 문항 간의 상관관계 분석
- 스피어만 상관관계수 사용(소수점 아래 다섯째 자리에서 반올림)
 - B12_1_1 : 법률문제로 인한 어려움
 - B12_1_2 : 차별대우 및 권리침해로 인한 어려움

[표 3-8] B12(법률 및 권익보장) 하위 문항 간의 상관관계수

변수	B12_1_1	B12_1_2
B12_1_1	1	0.8966
B12_1_2	0.8966	1



[그림 3-17] B12(법률 및 권익보장) 하위 문항 간의 상관관계 히트맵

- B13(문화여가) 하위 문항 간의 상관관계 분석
- 스피어만 상관계수 사용(소수점 아래 다섯째 자리에서 반올림)
 - B13_1_1 : 문화여가 활동 관련 어려움
 - B13_1_2 : 체육 활동 관련 어려움

[표 3-9] B13(문화여가) 하위 문항 간의 상관계수

변수	B13_1_1	B13_1_2
B13_1_1	1	0.8750
B13_1_2	0.8750	1

□ 파생변수 생성

[표 3-10] 파생변수 생성 상세

컬럼ID	컬럼명	상세 설명
B4	정신건강	B4_1_1 ~ B4_1_2의 합계 (B4_1_3 ~ B4_1_5 제외)
B5	신체건강	B5_1_1 ~ B5_1_2의 합계
B6	기초생활유지	B6_1_1 ~ B6_1_6의 합계 (B6_1_4 제외)
B7	가족 및 사회적 관계	B7_1_1 ~ B7_1_2의 합계
B8	보호안전	B8_1_1 ~ B8_1_2의 합계
B11	주거	B11_1_1 ~ B11_1_11의 합계 (B11_1_5 제외)
B12	법률 및 권익보장	B12_1_1 ~ B12_1_2의 합계
B13	문화여가	B13_1_1 ~ B13_1_2의 합계

□ 종속변수(정신건강)와 독립변수 간의 상관관계 분석

- 연속형변수가 아니기 때문에 피어슨 상관계수가 아닌 스피어만 상관계수 사용

[표 3-11] 종속변수와 독립변수 간의 상관계수

독립변수	상관계수
B5(신체건강)	0.539608
B7(가족 및 사회적 관계)	0.502674
B8(보호안전)	0.465275
B12(법률 및 권익보장)	0.437430
B6(기초생활유지)	0.418855
B11(주거)	0.366676
B13(문화여가)	0.299688

□ 회귀분석

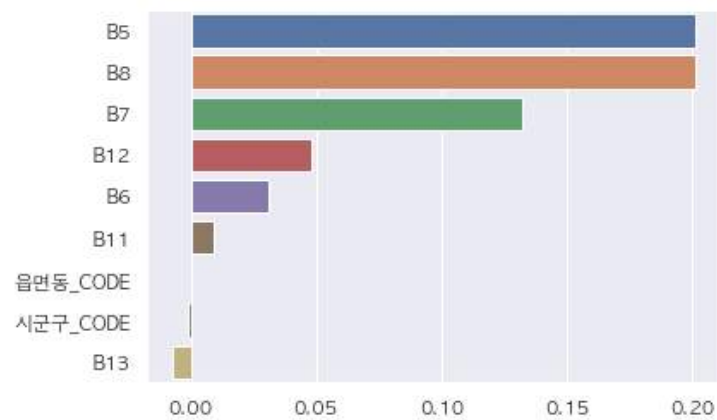
- 선형회귀모델의 가중치, 편향 확인

[표 3-12] 선형회귀모델의 가중치(weight) 확인

컬럼명	가중치(weight)
시군구_CODE	-0.00120101
읍면동_CODE	0.00045828

B5	0.2005338
B6	0.0306613
B7	0.13189806
B8	0.20109523
B11	0.00866712
B12	0.04818084
B13	-0.00711321

- 편향(bias) : 0.7644877102189689



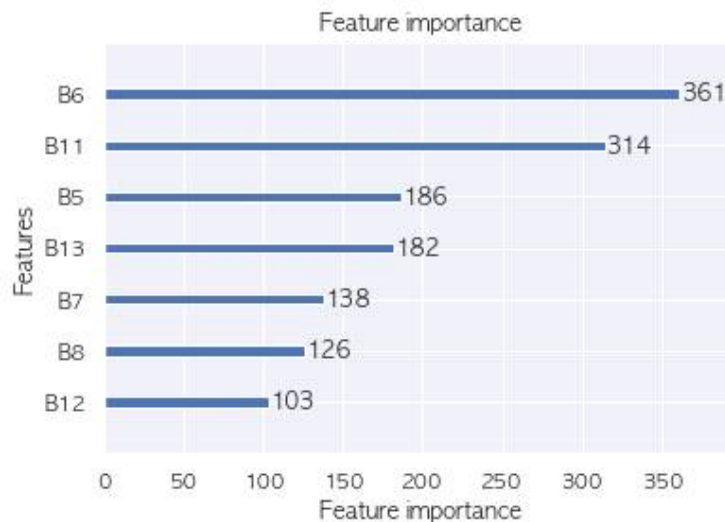
[그림 3-18] 선형회귀모델 가중치 막대그래프

[표 3-13] 4가지 회귀모델의 성능

	MSE	RMSE	R^2
LinearRegression	0.053	0.231	0.391
RandomForestRegressor	0.057	0.240	0.346
LGBMRegressor	0.052	0.228	0.407
Lightgbm Boosting	0.051	0.225	0.422

- 57번의 학습 끝에 성능 향상 종료

- Feature Importance 시각화



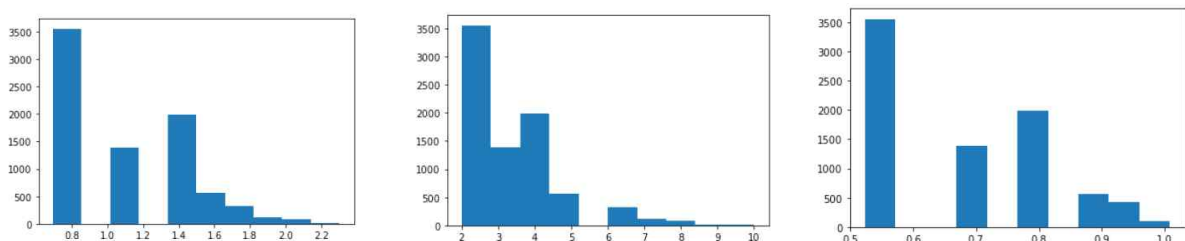
[그림 3-19] 선형회귀모델 Feature Importance 시각화

- 테스트 셋이 아닌 전체 셋을 대상으로 모델을 적용한 결과, R^2 스코어가 -0.725인 것으로 확인되었다. R^2 스코어가 음수가 나온 것은, 과적합이 발생한 모델로 평균으로 예측한 값보다 모델의 성능이 떨어지는 상황, 회귀 모델은 폐기하기로 결정하였다.

- 회귀 모델에서 시행착오를 겪은 이유는 우선 y 변수와 x 변수가 리커트 척도의 합으로 구성이 된 등간척도이고, 특히 x 변수가 2~10의 값을 갖기 때문에 선형회귀 모델은 적용하기가 어려운 부분이 있다. 만약 특정 값을 기준으로 0과 1로 나누는 로지스틱회귀 등 예측 모델은 적용이 가능할 것으로 보이나 데이터 분석 방향과 일치하지 않아 분류 모델을 사용하기로 결정하였다.

□ 분류분석

- B4(정신건강)에 대한 정규화(log변환, box-cox) 시도

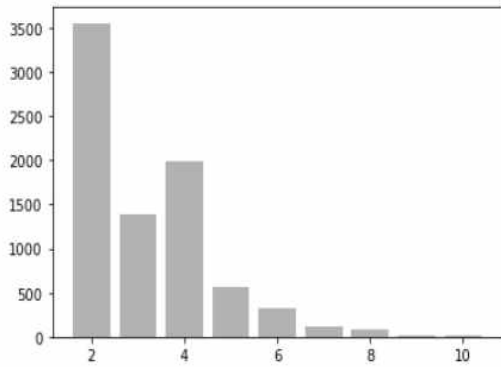


[그림 3-20] 기존 데이터의 종속변수, log 변환, box-cox 변환 후 히스토그램

- 여전히 좌측편향 분포양상이고 정신건강의 어려움을 겪는 사람들의 데이터가 적기 때문에 유의미한 결과를 해석할 수 없는 문제점이 있었다. 기존의 데이터를 정규화하

는 방법 (등간척도의 경우 정규화하기 까다로운 문제가 있었기 때문에) 대신 편향된 데이터의 양을 늘리는 방법을 사용해보기로 하였다.

□ 오버샘플링(SMOTE기법)



B4(정신건강)의 히스토그램

B4(정신건강)	인원 수
2	3555
3	1380
4	1984
5	558
6	317
7	113
8	77
9	6
10	10

[그림 3-21] 기존 데이터의 종속변수 히스토그램 및 도수분포표

- 4 이하 집단의 수는 6,919명, 5 이상 집단의 수는 1,081명으로 약 7:1이 비율로 데이터 불균형이 보인다. 이를 해결하기 위해 오버샘플링 기법인 SMOTE를 적용하여 5 이상 집단의 데이터를 4 이하 집단의 수만큼 늘려준다. 오버샘플링한 데이터를 8:2의 비율로 train 데이터와 test 데이터로 나눈 후 RandomForestClassifier 모델을 사용하여 예측한다.

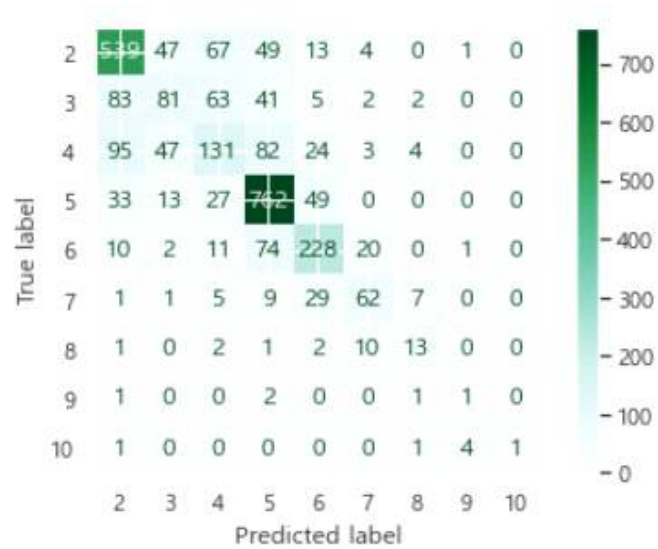
- accuracy : 0.6741329479768786

□ 정밀도(precision), 재현율(recall) 등 분류 모델의 성능지표 파악

- classification_report 결과

	precision	recall	f1-score	support
2	0.71	0.74	0.73	720
3	0.45	0.31	0.37	277
4	0.45	0.35	0.39	386
5	0.74	0.85	0.79	888
6	0.63	0.67	0.65	345
7	0.63	0.54	0.58	105
8	0.66	0.66	0.66	38
9	0.14	0.33	0.20	3
10	0.00	0.00	0.00	6
accuracy			0.66	2768
macro avg	0.49	0.49	0.48	2768
weighted avg	0.64	0.66	0.65	2768

[그림 3-22] test 데이터 RandomForestClassifier 모델 성능지표

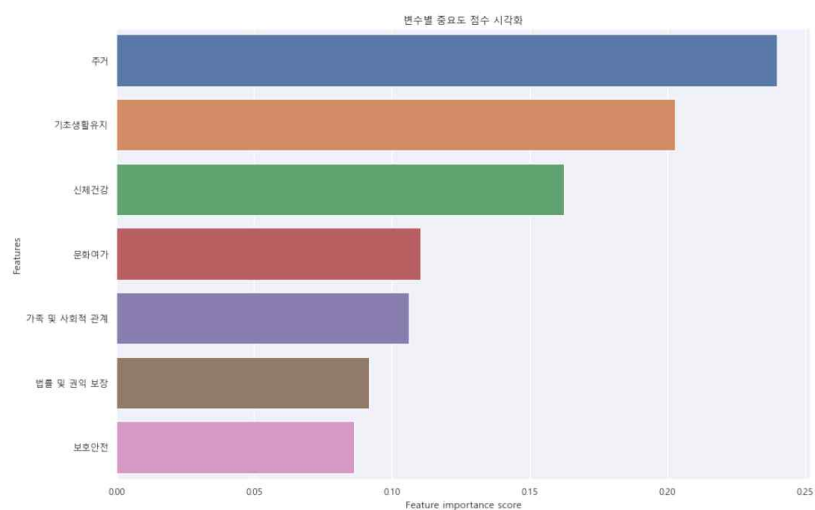


[그림 3-23] test 데이터 RandomForestClassifier 모델 Confusion Matrix

□ RandomForestClassifier 모델 feature 중요도 분석 결과

[표 3-14] RandomForestClassifier 모델 Feature Importance

B5(신체건강)	0.162409
B6(기초생활유지)	0.202807
B7(가족 및 사회적 관계)	0.106211
B8(보호-안전)	0.086363
B11(주거)	0.240009
B12(법률 및 권익 보장)	0.091865
B13(문화-여가)	0.110336



[그림 3-24] RandomForestClassifier 모델 막대그래프

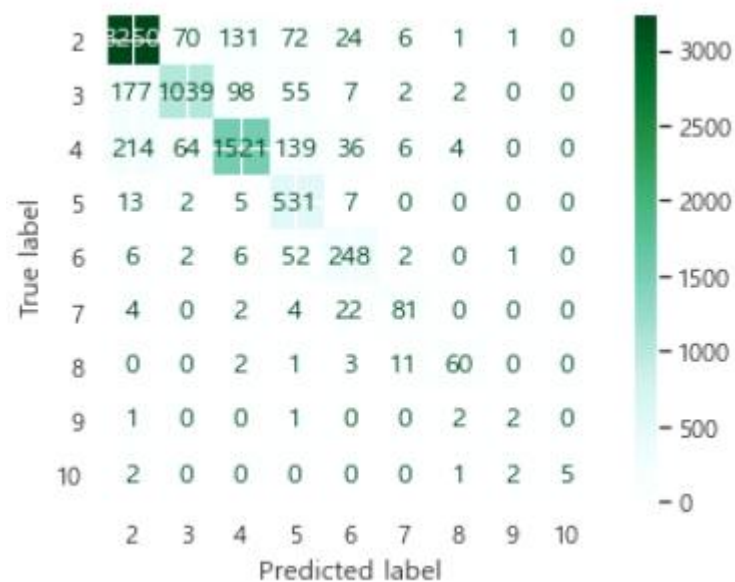
- 오버샘플링한 데이터를 기반으로 학습한 모델로 전체 8,000명 정신건강 예측
- accuracy : 0.842125

□ 정밀도(precision), 재현율(recall) 등 분류 모델의 성능지표 파악

- classification_report 결과

	precision	recall	f1-score	support
2	0.88	0.91	0.90	3555
3	0.88	0.76	0.81	1380
4	0.86	0.77	0.81	1984
5	0.62	0.95	0.75	558
6	0.73	0.79	0.76	317
7	0.73	0.72	0.72	113
8	0.86	0.73	0.79	77
9	0.29	0.33	0.31	6
10	1.00	0.50	0.67	10
accuracy			0.84	8000
macro avg	0.76	0.72	0.72	8000
weighted avg	0.85	0.84	0.84	8000

[그림 3-25] 기존 데이터 RandomForestClassifier 모델 성능지표



[그림 3-26] 기존 데이터 RandomForestClassifier 모델 Confusion Matrix

□ 모델 기반 부산시 시군구별 마음건강지수 예측 결과 비교

[표 3-15] 부산시 시군구별
마음건강지수 예측 결과[표 3-16] 부산시 시군구별
정신건강지수 실제 결과

시군구	마음건강지수	시군구	정신건강지수
북구	3.542	북구	3.576
서구	3.412	서구	3.362
영도구	3.402	해운대구	3.348
동래구	3.334	동래구	3.29
해운대구	3.31	영도구	3.264
사하구	3.306	사하구	3.264
중구	3.296	사상구	3.236
사상구	3.272	중구	3.206
동구	3.26	동구	3.184
수영구	3.152	강서구	3.172
기장군	3.118	수영구	3.158
강서구	3.1	남구	3.04
남구	3.042	연제구	3.034
연제구	3.02	기장군	2.988
금정구	2.944	금정구	2.91
부산진구	2.884	부산진구	2.846

3.2.3. 최적 입지

□ 최적 입지 선정을 위한 변수 선정

- 고려사항

1. 사람들이 많이 거주하고 있는가?
2. 사람들이 방문을 편리하게 할 수 있는가?
3. 기존의 유사한 시설이 위치하고 있는가?
4. 건물을 건설할 경우, 그 지역에 개발이 원활하게 이루어질 수 있는가?
5. 마음건강측정지수 분석 결과값이 높아 정신건강이 우려되는 지역인가?

- 최종적으로 입지 선정을 위한 인구, 버스정류소, 지하철정류장, 정신보건시설, 고도, 마음건강측정지수를 변수로 선정

□ 최적 입지 선정에 영향을 주는 변수별 가중치 산출

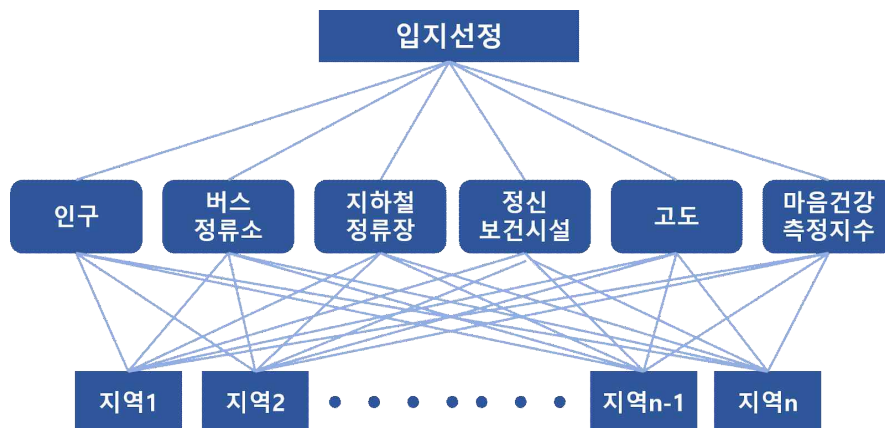
- 앞의 과정에서 선정한 총 6개의 변수가 입지 선정을 할 때 평가 요소로서 우선순위가 있다고 판단
- 우선순위에 따른 가중치를 다르게 부여하기로 결정
- 복잡한 평가 요소를 계층화하여 단계별 요소 간의 쌍대 비교를 하는 계층적 의사결정법(AHP, Analytic Hierarchy process)을 사용

중요도	정 의	설 명		A_1	A_2	A_3	A_4
1	비슷함 (Equal Importance)	두 요소가 비슷한 공헌도를 가짐	A_1	1	a_{12}	a_{13}	a_{14}
3	약간 중요함 (Moderate Importance)	한 요소가 다른 요소보다 약간 선호됨	A_2	a_{21}	1	a_{23}	a_{24}
5	중요함 (Strong Importance)	한 요소가 다른 요소보다 강하게 선호됨	A_3	a_{31}	a_{32}	1	a_{34}
7	매우 중요함 (Very strong Importance)	한 요소가 다른 요소보다 매우 강하게 선호됨	A_4	a_{41}	a_{42}	a_{43}	1
9	극히 중요함 (Extreme Importance)	한 요소가 다른 요소보다 극히 선호됨					

$$a_{ij} = 1/a_{ji}, \quad a_{ii} = 1, \text{ for all } i$$

[그림 3-27] 쌍대비교 방법

□ AHP 계층구조 작성



[그림 3-28] AHP 계층구조도

□ 쌍대 비교를 한 결과 비교 행렬을 작성

[표 3-17] 쌍대 비교 행렬

	인구	버스정류소	지하철정류장	정신보건시설	고도	마음건강 측정지수
인구	1	1	1	3	7	1/5
버스정류소	1	1	1/3	1/7	5	1/3
지하철정류장	1	3	1	1/7	5	1/3
정신보건시설	1/3	7	7	1	5	1/3
고도	1/7	1/5	1/5	1/5	1	1/7
마음건강 측정지수	5	3	3	3	7	1

- 쌍대 비교를 통해 구한 행렬로부터 자체적 행렬 곱 연산으로 상호 중요도 도출
 - n개의 요소에 대한 상대적 중요도를 w_i 로 놓는다. : w_i ($i = 1, \dots, n$)

$$A = \begin{bmatrix} 1 & a_{12} & a_{13} & \cdots & a_{1n} \\ a_{21} & 1 & a_{23} & \cdots & a_{2n} \\ a_{31} & a_{32} & 1 & \cdots & a_{3n} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ a_{n1} & a_{n2} & a_{n3} & \cdots & 1 \end{bmatrix} \quad \rightarrow \quad A = \begin{bmatrix} w_1/w_1 & w_1/w_2 & \cdots & w_1/w_n \\ w_2/w_1 & w_2/w_2 & \cdots & w_2/w_n \\ \cdots & \cdots & \cdots & \cdots \\ w_n/w_1 & w_n/w_2 & \cdots & w_n/w_n \end{bmatrix}$$

$$\sum_j^n a_{ij} \cdot w_j \cdot 1/w_i = n, \quad (i, j = 1, \dots, n) \rightarrow \sum_j^n a_{ij} \cdot w_j = n \cdot w_i, \quad (i, j = 1, \dots, n)$$

[그림 3-29] 쌍대 비교 행렬 곱 연산

- $A \cdot w = \lambda \cdot w$ 을 만족하는 λ 이 존재(선형대수학의 eigenvalue method) $(A - \lambda I) \cdot w = 0$ 을 만족하는 λ_{\max} (eigenvalue)와 w (eigenvector) 구할 수 있음

□ 상호 중요도에 따른 입지 선정에 영향을 주는 변수에 가중치 부여

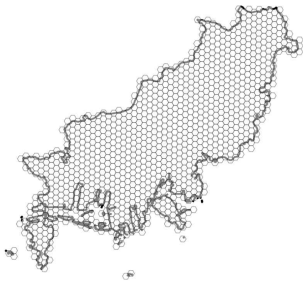


[표 3-18] 각 변수가 갖는 가중치

변수	가중치
정신건강지수	0.365
정신보건시설	0.246
인구	0.18
지하철정류장	0.107
버스정류소	0.077
고도	0.025

□ 그리드/동/구별 입지 선정 데이터

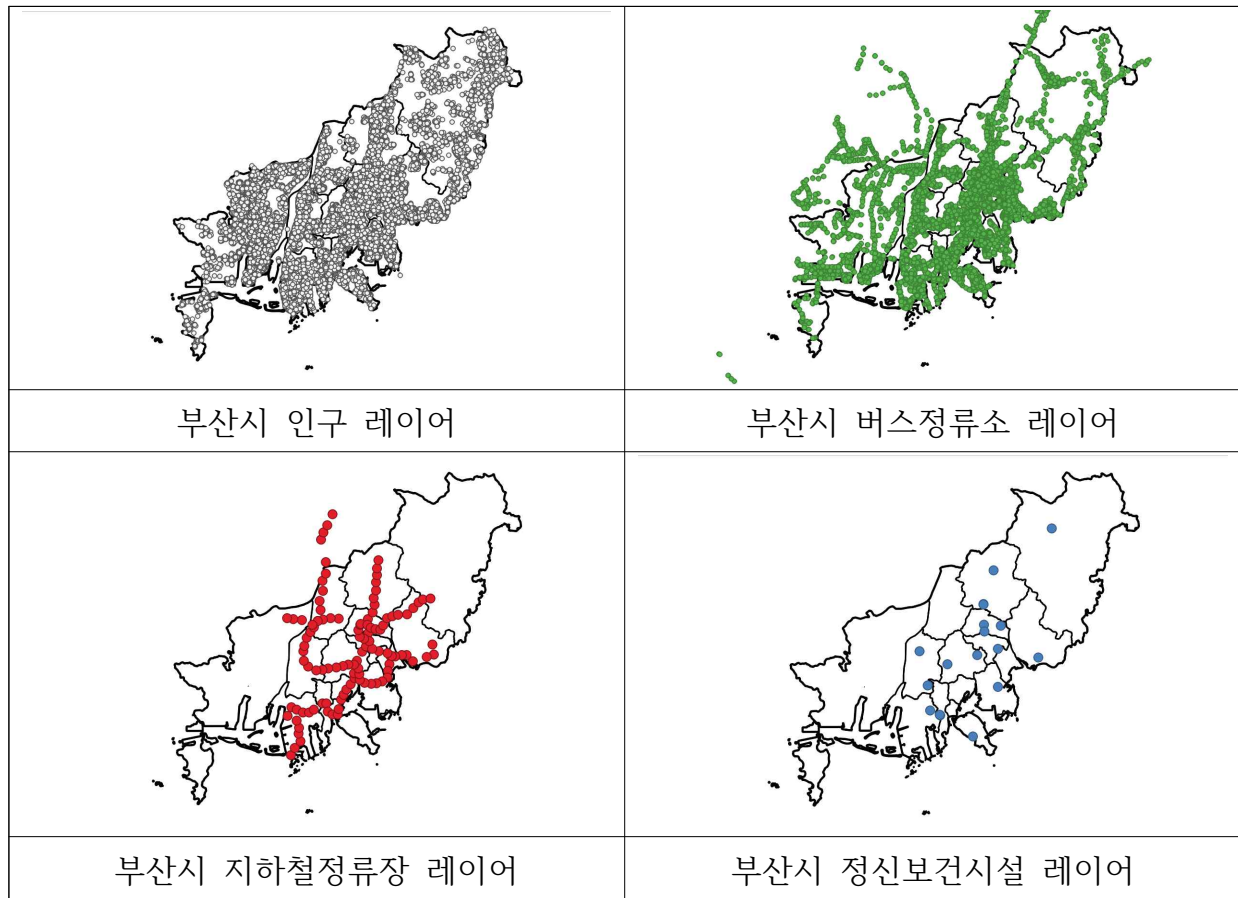
- 부산시 그리드 구역 데이터를 활용하여 부산시 1km*1km 그리드 구역 레이어 생성
- 부산시 동별 행정구역 데이터를 활용하여 부산시 103개 동을 기준으로 부산시 동 레이어 생성
- 부산시 구별 행정구역 데이터를 활용하여 부산시의 15개 구와 1개 군을 포함한 총 16개의 구/군을 기준으로 부산시 구 레이어 생성

[그림 3-30] 부산시 hexagon 그리드/동/구별 레이어

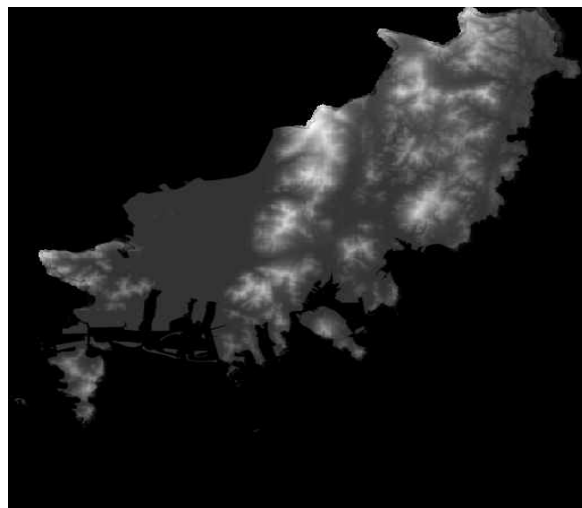
레이어 명	레이어 이미지
부산시 1km * 1km hexagon 그리드 구역 레이어	
부산시 동별 레이어	
부산시 구별 레이어	

- 그리드/동/구별 인구 데이터를 구하기 위해 그리드/동/구별 인구를 포함하는 19,439개 포인트를 구하고 포인트가 가지고 있는 3,418,122명의 총 인구 수를 통해 부산시 그리드/동/구별 인구 레이어 및 데이터셋 생성
- 부산시 버스정류소정보 데이터를 활용하여 그리드/동/구별 버스정류장 9,759개를 포함하는 부산시 그리드/동/구별 버스정류소 레이어 및 데이터셋 생성
- 국가철도공단_부산 1,2,3,4호선 역위치 데이터를 활용하여 그리드/동/구별 지하철정류장 개수를 포함하는 부산시 그리드/동/구별 지하철정류장 레이어 및 데이터셋 생성
- 부산광역시_정신보건시설 데이터를 활용하여 그리드/동/구별 정신보건시설 개수를 포함하는 부산시 그리드/동/구별 정신보건시설 레이어 및 데이터셋 생성

[그림 3-31] 부산시 수집 데이터 레이어

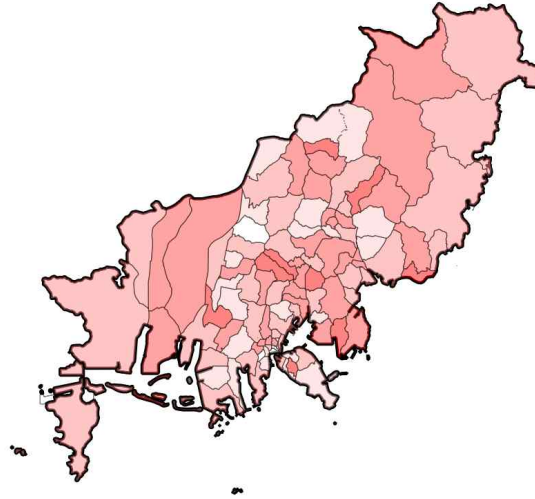


- 수치표고모델(DEM)_90M 데이터를 활용하여 부산시 그리드/동/구별 평균 고도값을 구하고 그리드/동/구별 고도 레이어 생성



[그림 3-32] 부산시 고도 래스터 레이어

- 구별 마음건강지수를 활용하여 그리드/동/구별 마음건강지수 레이어 생성



[그림 3-33] 부산시 마음건강지수 레이어

- 부산시 그리드/동/구, 인구, 버스정류장, 지하철정류장, 정신보건시설, 고도, 마음건강지수를 종합한 부산시 그리드/동/구별 분석 레이어 및 데이터셋 생성
- 그리드/동/구별 분석 데이터셋의 변수별 데이터 값의 범위가 다른 것을 MinMax 정규화를 통해 0~1 사이 값을 가지는 데이터로 변환

	id	버스정류장	지하철정류장	정신보건시설	인구	고도	지수
0	328	0.000000	0.0	0.0	0.000000	0.000568	0.000000
1	322	0.067568	0.0	0.0	0.000747	0.020246	0.728980
2	323	0.000000	0.0	0.0	0.000000	0.337168	0.728980
3	320	0.027027	0.0	0.0	0.000000	0.004022	0.728980
4	2234	0.000000	0.0	0.0	0.000000	0.373987	0.675378
...
1090	2406	0.013514	0.0	0.0	0.003278	0.307535	0.675378
1091	2407	0.000000	0.0	0.0	0.003436	0.210085	0.675378
1092	2404	0.027027	0.0	0.0	0.000000	0.451898	0.673515
1093	2405	0.054054	0.0	0.0	0.001230	0.258302	0.675378
1094	2399	0.000000	0.0	0.0	0.000000	0.438646	0.666530

1095 rows x 7 columns

[그림 3-34] 부산시 hexagon 그리드 정규화 데이터

	동	버스정류장	지하철정류장	정신보건시설	인구	고도	지수
0	반여동	0.237226	0.000000	0.0	0.595174	0.330582	0.617298
1	중동	0.167883	0.166667	0.0	0.233895	0.098668	0.222923
2	우동	0.239051	0.500000	0.5	0.425538	0.265661	0.438245
3	좌동	0.251825	0.333333	0.0	0.646525	0.328671	0.302570
4	송정동	0.060219	0.000000	0.0	0.072632	0.076174	0.523243
...
98	복산동	0.071168	0.000000	0.0	0.094007	0.089669	0.394595
99	반송동	0.135036	0.666667	0.0	0.396155	0.319775	0.340081
100	광복동	0.000000	0.000000	0.0	0.000877	0.034894	0.634234
101	부곡동	0.195255	0.166667	0.0	0.476724	0.210881	0.425579
102	청룡노포동	0.173358	0.333333	0.0	0.094083	0.616916	0.665065

103 rows × 7 columns

[그림 3-35] 부산시 동별 정규화 레이어

	구	버스정류장	지하철정류장	정신보건시설	인구	고도	지수
0	중구	0.000000	0.230769	0.000000	0.000000	0.000000	0.806377
1	서구	0.131271	0.230769	0.666667	0.231935	0.432677	0.787849
2	동구	0.109532	0.307692	0.000000	0.155730	0.268307	0.701968
3	영도구	0.223244	0.000000	0.333333	0.301061	0.288203	0.661992
4	부산진구	0.494983	0.846154	0.666667	0.993901	0.517276	0.000000
5	동래구	0.380435	0.846154	0.666667	0.675782	0.125395	0.629610
6	남구	0.339465	0.461538	0.333333	0.729373	0.121317	0.321341
7	북구	0.270903	1.000000	0.000000	0.756002	0.857094	1.000000
8	해운대구	0.464883	0.769231	0.333333	1.000000	0.513055	0.479928
9	사하구	0.472408	0.923077	0.333333	0.913496	0.313384	0.673681
10	금정구	0.358696	0.692308	1.000000	0.609533	1.000000	0.180849
11	강서구	1.000000	0.230769	0.000000	0.052746	0.021271	0.387118
12	연제구	0.178930	0.615385	0.333333	0.481836	0.192102	0.242287
13	수영구	0.093645	0.538462	0.333333	0.368382	0.070513	0.409298
14	사상구	0.301839	0.538462	0.666667	0.610015	0.390264	0.620053
15	기장군	0.639632	0.153846	0.333333	0.170248	0.718647	0.388358

[그림 3-36] 부산시 구별 정규화 데이터

- 각 변수별 변환된 데이터 값과 변수별 가중치의 곱을 구함
- 입지 선정에 (+)요소인 인구, 버스정류장, 지하철정류장, 마음건강지수와 (-)요소인 정신보건시설, 고도의 요소들의 계산을 통하여 최종값을 도출함
- 도출한 값에 따른 순위를 부여함으로써 마음건강증진센터 입지 선정 데이터를 구함
부산시 마음건강증진센터 최적 입지 선정
- 입지 선정 데이터와 부산시 그리드/동/구별 레이어를 결합을 통해 부산시 마음건강증진센터 레이어를 생성
- 마음건강증진센터 레이어를 활용하여 순위별 단계 구분을 통한 지도상 시각화
- 각 그리드, 동, 구별 시각화 자료를 통해 부산시 센터 최적 입지 선정 및 평가

4. 분석결과

4.1. 마음건강간편테스트

[표 4-1] 마음건강간편테스트 문항

분류	질문	응답
교우관계	친구나 주변인과 보내는 시간이 즐거운가요?	예/아니오
	내 마음을 편하게 터놓을 친구가 주변에 있나요?	예/아니오
	나는 친구들과 잘 어울린다고 생각하나요?	예/아니오
	친구들이 나를 어떻게 생각하는지 의식하게되나요?	예/아니오
가족관계	가족들과 함께 있을 때 마음이 편한가요?	예/아니오
	가족들과 원활한 의사소통을 하고 있나요	예/아니오
	현재, 주변에 의지할 가족이 있나요?	예/아니오
	가족들이 위로가 되는 존재인가요?	예/아니오
학업	보람찬 하루로 인생을 보내고 있는 것 같나요?	예/아니오
	자신이 원하는 결과를 이룰 수 있을 것이라는 믿음이 있나요?	예/아니오
	자신에게 주어진 일(공부, 일 등)을 기한 안에 해내기 어려운가요?	예/아니오
	직장/학교에 관해 상담을 받고 싶은 적이 있나요?	예/아니오
상태	기분이 가라앉거나 무기력할 때가 자주 있나요?	예/아니오
	자신이 죽는 것이 더 낫다고 생각할 때가 있나요?	예/아니오
	자신이 사랑받을만한 사람이라고 생각하나요?	예/아니오

4.2. 마음건강지수

[표 4-2] 부산시 읍면동별 마음건강지수 예측 결과(상위10개)

읍면동	마음건강지수
수정4동	6
구포3동	5
서대신3동	4.681818
남포동	4.545455
반여2동	4.5
초장동	4.428571
감천1동	4.4
구포2동	4.297297
괴정4동	4.272727
사직2동	4.255814

[표 4-3] 북구 읍면동별 마음건강지수 예측 결과

읍면동	마음건강지수
구포1동	4.028571
구포2동	4.297297
구포3동	5
금곡동	3.53125
화명1동	3.651163
화명2동	3.206897
화명3동	2.91667
덕천1동	3.342857
덕천2동	3.435897
덕천3동	4.185185
만덕1동	2.605263
만덕2동	3.508475
만덕3동	3.666667

4.3. 마음건강증진센터 최적 입지 선정

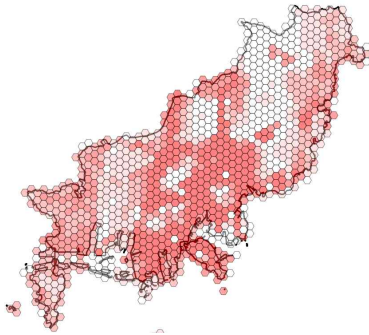
4.3.1 그리드/동/구별 부산시 마음건강증진센터 최적 입지 선정 결과

□ hexagon 그리드별 최적 입지 선정

	id	버스정류장	지하철정류장	정신보건시설	인구	고도	지수	합계	순위
314	1788	0.044743	0.107000	0.0	0.133787	0.000817	0.306845	0.591558	1
411	1445	0.029135	0.107000	0.0	0.105840	0.000725	0.301810	0.543061	2
338	1732	0.062432	0.000000	0.0	0.171312	0.001826	0.305740	0.537659	3
362	1559	0.039541	0.035667	0.0	0.180000	0.002297	0.268265	0.521176	4
261	1902	0.065554	0.071333	0.0	0.109509	0.000380	0.272640	0.518657	5
...
1048	264	0.000000	0.000000	0.0	0.000000	0.000018	0.000000	-0.000018	1084
67	208	0.000000	0.000000	0.0	0.000000	0.000095	0.000000	-0.000095	1085
584	3198	0.000000	0.000000	0.0	0.000000	0.000305	0.000000	-0.000305	1086
302	1664	0.000000	0.000000	0.0	0.000000	0.010249	0.000000	-0.010249	1087
383	1609	0.000000	0.000000	0.0	0.000000	0.013715	0.000000	-0.013715	1088

1095 rows × 9 columns

[그림 4-1] 부산시 hexagon 그리드 입지 선정 결과 데이터



[그림 4-2] 부산시 hexagon 그리드
최적 입지 시각화



[그림 4-3] 부산시 hexagon 그리드
최적 입지 상위 10개 지역 시각화

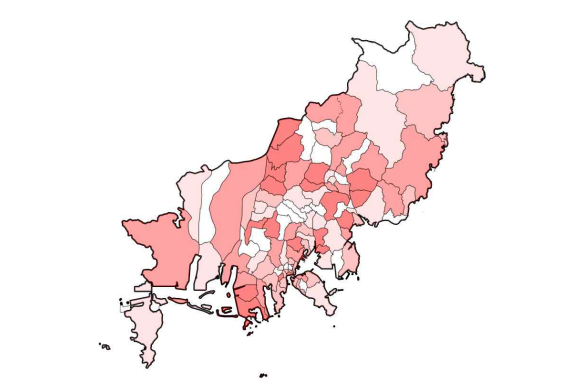
- 부산시 hexagon 그리드 최적 입지 상위 10개 지역을 시각화해 본 결과 북구, 동래구 쪽에 대다수의 분포가 집중되어 있었고, 금정구, 동구, 해운대구도 포함되어있는 것을 확인할 수 있었다.

□ 동별 최적 입지 선정

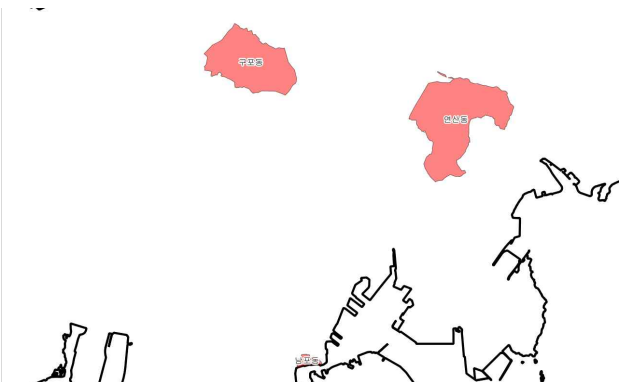
	동	버스정류장	지하철정류장	정신보건시설	인구	고도	지수	합계	순위
42	구포동	0.014754	0.071333	0.000	0.082512	0.009029	0.349277	0.508846	1
32	연산동	0.031193	0.089167	0.000	0.180000	0.003980	0.139245	0.435625	2
93	남포동	0.002810	0.035667	0.000	0.000000	0.000098	0.365000	0.403379	3
43	덕천동	0.011803	0.053500	0.000	0.065457	0.002949	0.229669	0.357481	4
54	온천동	0.026416	0.107000	0.123	0.104641	0.005792	0.243029	0.352294	5
...
27	망미동	0.011943	0.000000	0.123	0.052893	0.004261	0.117122	0.054698	99
17	정관읍	0.034987	0.000000	0.123	0.029577	0.011121	0.124246	0.054689	100
39	남산동	0.007869	0.017833	0.123	0.039133	0.011474	0.059004	-0.010634	101
36	장전동	0.011803	0.017833	0.246	0.049524	0.009618	0.120702	-0.055756	102
86	아미동	0.007307	0.017833	0.246	0.016623	0.005948	0.140344	-0.069841	103

103 rows x 9 columns

[그림 4-4] 부산시 동별 입지 선정 결과 데이터



[그림 4-5] 부산시 동별 최적 입지 시각화



[그림 4-6] 부산시 동별 최적 입지 상위 3개 지역 시각화

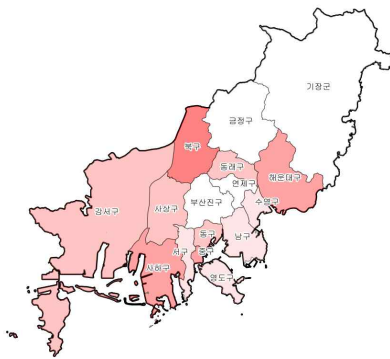
- 부산시 상위 3개 동 지역은 1순위 구포동, 2순위 연산동, 3순위 남포동 순으로 결과 도출

□ 구별 최적 입지 선정

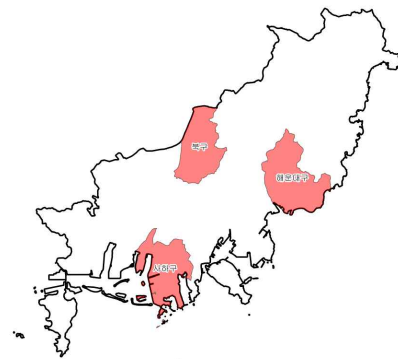
	구	버스정류장	지하철정류장	정신보건시설	인구	고도	지수	합계	순위
7	북구	0.020860	0.107000	0.000	0.136080	0.021427	0.365000	0.607513	1
9	사하구	0.036375	0.098769	0.082	0.164429	0.007835	0.245893	0.455633	2
8	해운대구	0.035796	0.082308	0.082	0.180000	0.012826	0.175174	0.378451	3
0	중구	0.000000	0.024692	0.000	0.000000	0.000000	0.294328	0.319020	4
2	동구	0.008434	0.032923	0.000	0.028031	0.006708	0.256218	0.318899	5
5	동래구	0.029293	0.090538	0.164	0.121641	0.003135	0.229808	0.304145	6
11	강서구	0.077000	0.024692	0.000	0.009494	0.000532	0.141298	0.251953	7
14	사상구	0.023242	0.057615	0.164	0.109803	0.009757	0.226319	0.243223	8
6	남구	0.026139	0.049385	0.082	0.131287	0.003033	0.117290	0.239067	9
3	영도구	0.017190	0.000000	0.082	0.054191	0.007205	0.241627	0.223803	10
13	수영구	0.007211	0.057615	0.082	0.066309	0.001763	0.149394	0.196766	11
1	서구	0.010108	0.024692	0.164	0.041748	0.010817	0.287565	0.189296	12
12	연제구	0.013778	0.065846	0.082	0.086731	0.004803	0.088435	0.167987	13
15	기장군	0.049252	0.016462	0.082	0.030645	0.017966	0.141751	0.138142	14
4	부산진구	0.038114	0.090538	0.164	0.178902	0.012932	0.000000	0.130622	15
10	금정구	0.027620	0.074077	0.246	0.109716	0.025000	0.066010	0.006422	16

[그림 4-7] 부산시 구별 입지 선정 결과 시각화

- 부산시 상위 3개 구 지역은 1순위 북구, 2순위 사하구, 3순위 해운대구 순으로 결과 도출

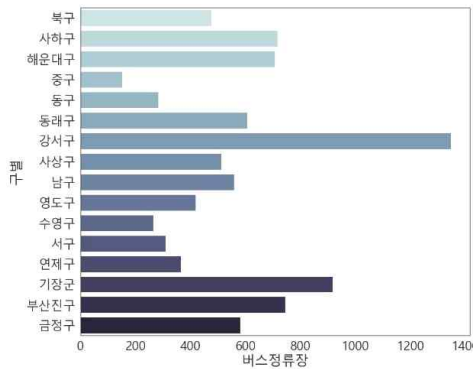


[그림 4-8] 부산시 구별 최적 입지 시각화

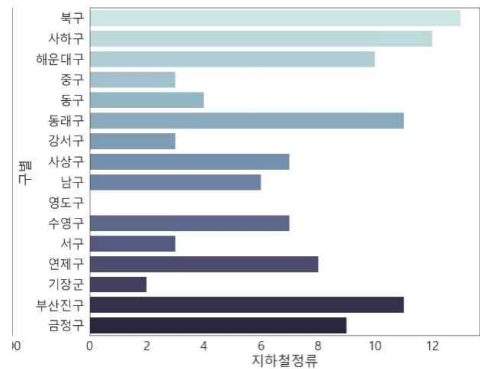


[그림 4-9] 부산시 구별 최적 입지 상위 3개 지역 시각화

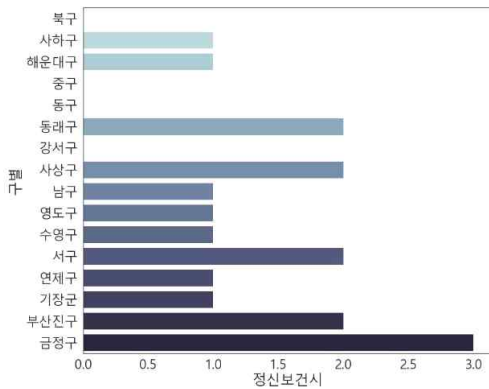
4.3.2. 구별 변수 데이터 그래프



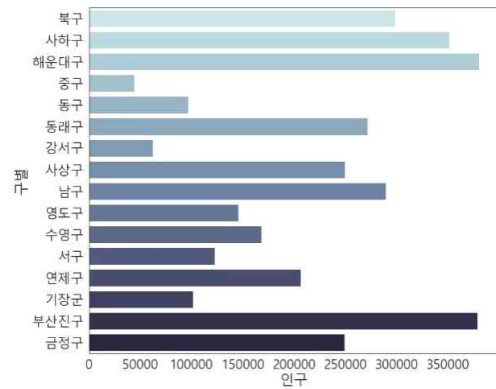
[그림 4-10] 부산시 버스정류장 막대그래프



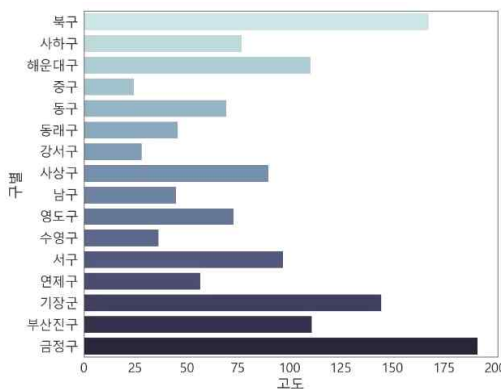
[그림 4-11] 부산시 지하철정류장 막대그래프



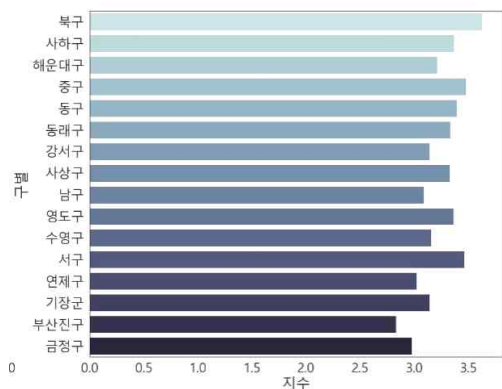
[그림 4-12] 부산시 구별 정신보건시설 막대그래프



[그림 4-13] 부산시 구별 인구 막대그래프



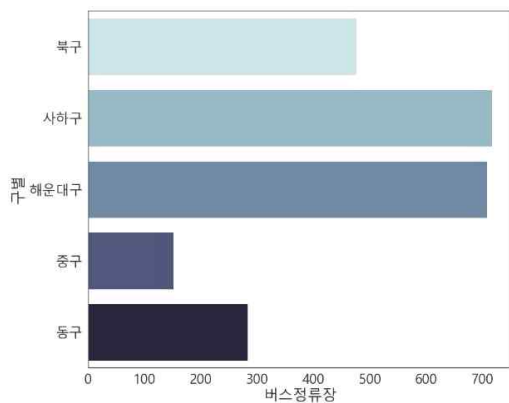
[그림 4-14] 부산시 구별 고도 막대그래프



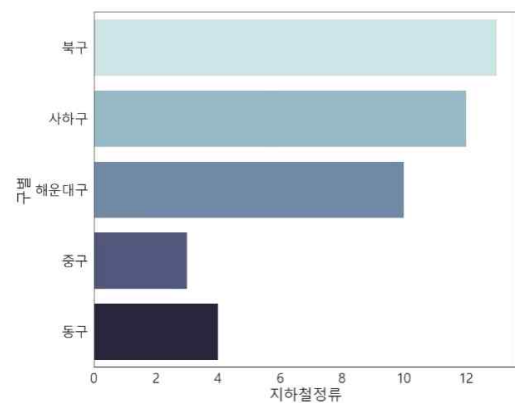
[그림 4-15] 부산시 구별 마음건강지수 막대그래프

4.3.3. 부산시 상위 5개 구 그래프 비교

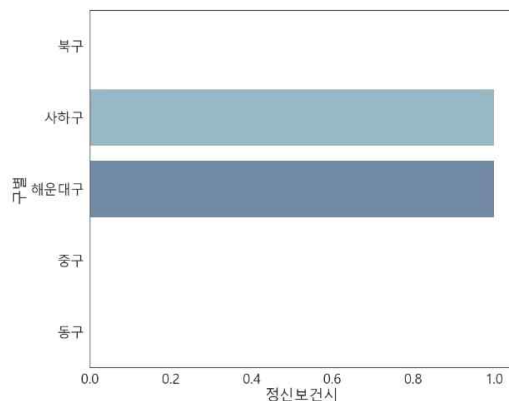
- 마음건강증진센터 입지 선정 결과 상위 5개 구는 북구, 사하구, 해운대구, 동래구, 동구 순으로 결과가 나왔다.
- 북구가 인구, 버스정류장 요소들이 평균적인 값을 가지고 있지만, 마음건강지수가 높아 정신건강 위험 지역에 속하다고 판단이 되지만 기존의 시설이 있지 않고 교통의 편리성을 고려해 보았을 때 1순위로 나온 것으로 볼 수 있다.
- 사하구는 기존의 시설이 있기는 하지만 인구, 버스정류장, 지하철정류장, 지수 요소들의 수치들이 평균보다 높게 형성되어있어 2순위로 나온 것으로 볼 수 있다.
- 해운대구는 다른 요소들이 높은 수치를 가지고 있지만 마음건강지수가 낮아 사하구 보다는 덜 위험지역으로 선정된 것으로 해석할 수 있다.



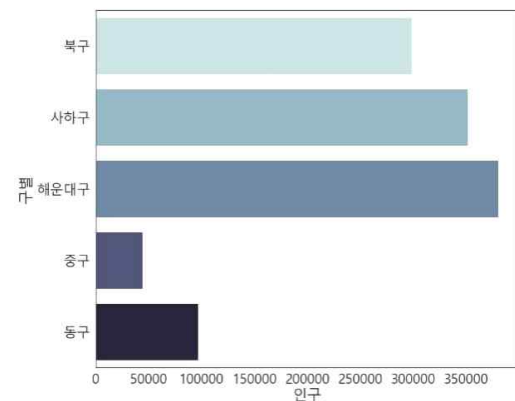
[그림 4-16] 부산시 상위 5개 구 버스정류장 막대그래프



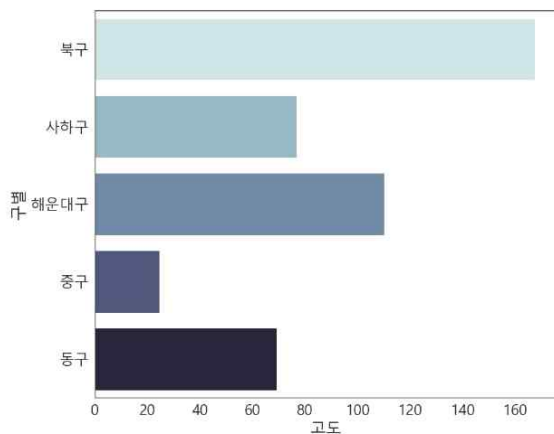
[그림 4-17] 부산시 상위 5개 구 지하철정류장 막대그래프



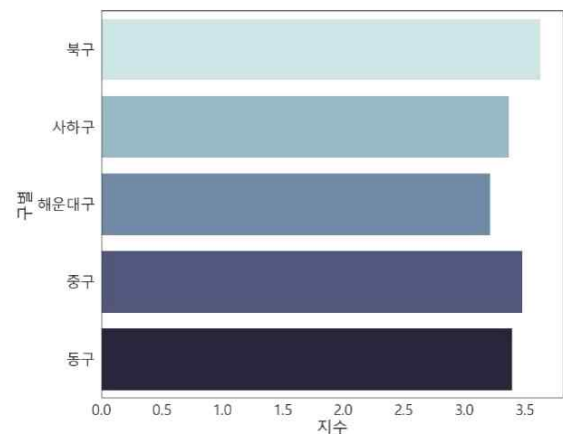
[그림 4-18] 부산시 상위 5개 구 정신보건시설 막대그래프



[그림 4-19] 부산시 상위 5개 구 인구 막대그래프



[그림 4-20] 부산시 상위 5개 구 고도 막대그래프

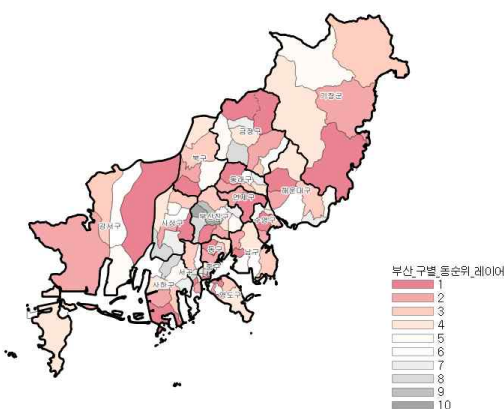


[그림 4-21] 부산시 상위 5개 구 마음건강지수 막대그래프

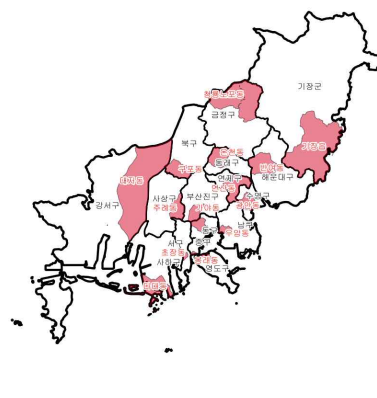
4.3.4. 구별 동순위 최적 입지 위치

동	구	버스정류장	지하철정류	정신보건시	인구	고도	지수	순위
0 대저동	강서구	300	3	0	19704.38	1.550388	3.085714	1
6 청룡노포동	금정구	95	2	0	13464.97	280.3415	3.740741	1
14 기장읍	기장군	251	0	0	48881.81	97.46499	3.173554	1
19 우암동	남구	66	0	0	24853.99	29.83908	3.851852	1
25 수정동	동구	101	1	0	29283.6	135.0135	4.04263	1
29 온천동	동래구	188	6	1	78435.65	105.8879	3.742586	1
36 가야동	부산진구	51	2	0	42356.17	135.1234	3.215862	1
46 구포동	북구	105	4	0	62042.71	164.5274	4.441956	1
51 주례동	사상구	142	2	0	56107.12	122.8277	3.43908	1
59 다대동	사하구	151	3	0	78441.38	37.00039	3.124275	1
67 초장동	서구	8	0	0	7833.17	112.9686	4.428571	1
75 광안동	수영구	72	4	0	62985.95	53.84919	3.257719	1
80 연산동	연제구	222	5	0	134260.1	73.06855	3.059431	1
82 봉래동	영도구	41	0	0	18397.49	46.39656	3.629109	1
88 남포동	중구	20	2	0	919.88	2.745097	4.545455	1
96 반여동	해운대구	130	0	0	80280.56	150.6782	3.625975	1

[그림 4-22] 부산시 구별 동 최적 입지 1순위 데이터



[그림 4-23] 부산시 구별 동 순위별 최적 입지 시각화



[그림 4-24] 부산시 구별 동 1순위 최적 입지 시각화

- 행정구 단위 내의 동별 순위를 도출해 시설이 특정 지역에 집중되는 현상을 방지할 수 있고 소외되는 지역을 최소화할 수 있다.

5. 활용 방안

5.1. 문제점 개선 방안

- (심리적 접근성 개선) 접근성 높은 인터넷을 통해 제공하는 텍스트마이닝 기반 심리 테스트로 심리적 접근성을 개선하여 정신건강관련 제도 및 정책의 이용률 증가
- (시설 확충) 정신건강증진센터 최적 입지 선정 지도를 활용하여 정신건강증진센터를 추가 선정 또는 기존 시설을 확충 시 활용
- (시설 이용률 증가) 정신건강증진센터 최적 입지 선정으로 시설 이용률을 증대시켜 자살원인 분석 결과 1위인 정신적, 정신과적 문제 해결에 기여

5.2. 업무 활용 방안

- 텍스트마이닝 및 감정분석 결과를 활용하여 정신건강 증진 프로그램 및 제도 만들 때 활용할 수 있을 것이다.
- 마음건강지수 도출 과정 중 상관관계 분석을 통해 높은 상관관계를 보이는 변수들을 활용하여 부산시민 정신건강 증진을 위한 프로그램이나 캠페인 제시
- 정신건강증진센터 최적 입지 주변에 약국이나 병원 등 의료시설 확충이 가능하므로 부산 시민의 복지와 편익 향상

6. 참고자료(Reference)

- 한국리서치 여론조사 사업1부 성현정 부장, 체계적이고 효과적인 의사 결정을 위한 AHP 분석 방법, 한국리서치 웹리서치 노트 No. 84
- 불용어 사전 출처 : (2022GBIG)분석참조모델_민원분석_실습(python) 강의 불용어 사전 사용

[참고자료 1. 자살예방백서]

[표 26] 2020년 연령표준화 자살률 상위 3개 시·군·구					[단위: 명, 인구 10만 명당 명, %]		
구분	순위	시·군·구		자살자 수	자살률	연령표준화 자살률	2019년 대비 연령표준화 자살률 증감률
구	1	부산광역시	수영구	70	40.1	33.5	72.7
	2	대전광역시	대덕구	56	31.9	32.6	30.4
	3	부산광역시	중구	12	29.7	32.2	-3.9

2020년 연령표준화 자살률(인구 10만 명당 명)이 가장 높은 구는 부산 수영구(33.5명), 대전 대덕구(32.6명), 부산 중구(32.2명)로 1위, 3위가 부산에 해당한다.

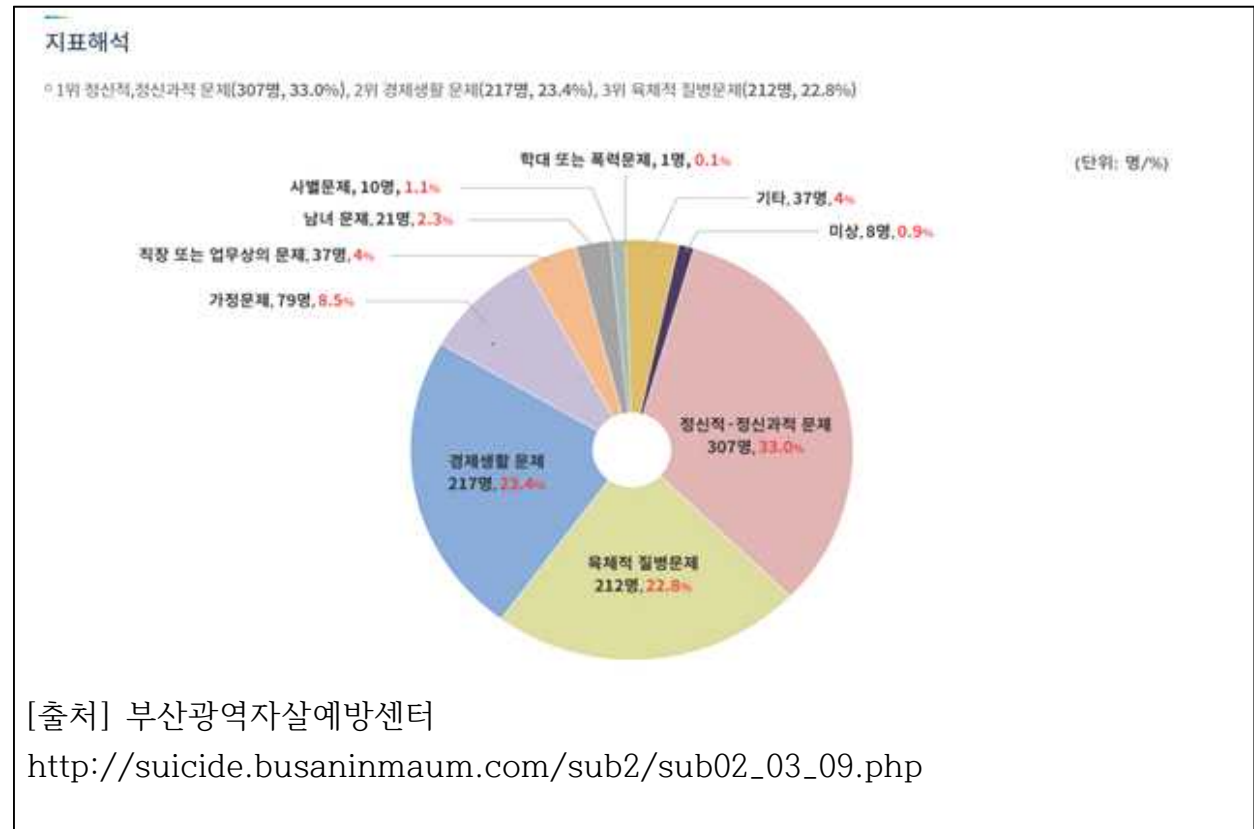
[출처] <https://www.korea.kr/archive/expDocView.do?docId=40007>

> 자살예방백서.pdf - p.57

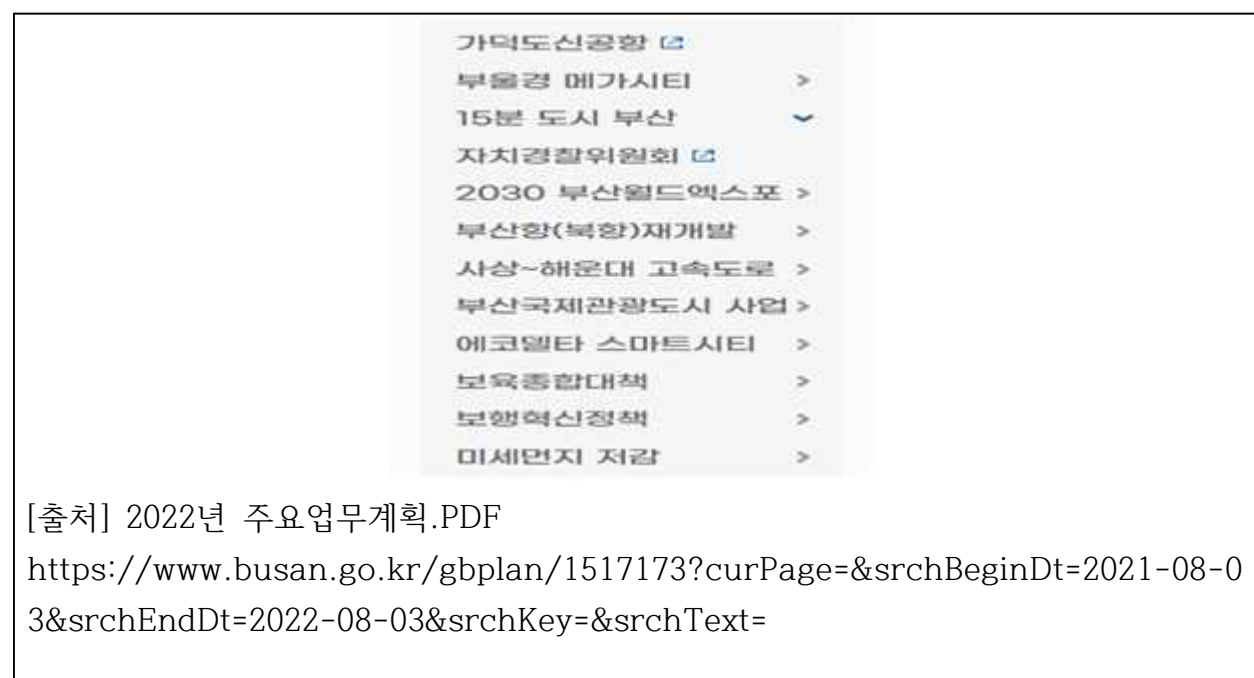
[참고자료 1-1. 5개년 전국 자살사망 분석 결과보고서]

<p>- 지역 내 5개년('13~'17) 자살사망자 중 정신건강문제로 인한 자살사망비율이 가장 높은 지역은 “부산광역시” (47.8%), 경제문제의 경우 “울산광역시” (23.8%), 신체건강문제의 경우 “충청남도” (22.8%)로 나타났다.</p> <p>* 정신건강: 부산(47.8%) > 광주(44.6%) > 세종(43.2%) / 평균 36.1% 경제문제: 울산(23.8%) > 인천(21.7%) > 경기(21.6%) / 평균 19.5% 신체건강: 충남(22.8%) > 전남(22.5%) > 전북(20.5%) / 평균 17.4%</p>	<p>[출처]</p> <p>http://www.mohw.go.kr/upload/viewer/skin/doc.html?fn=1630024032898_20210827092713.hwp&rs=/upload/viewer/result/202208/</p>
---	--

[참고자료 1-2. 부산광역시 자살 원인별 자살자 수/백분율]



[참고자료 2. 부산광역시 주요 업무계획 및 주요프로젝트]



[참고자료 2-1. 부산 정신재활시설 현황]

17일 중앙정신건강복지사업단이 발표한 '정신재활시설 현황'에 따르면 2018년 12월 말 기준 부산의 입소시설 정원은 인구 10만 명당 1.9명에 불과했다. 전국 17개 시·도 중 13위로 대전 15.1명, 충남 14.9명, 전북 12.7명 등에 비해 현저히 낮았다. 경남은 13명으로 16위를 기록했고, 울산은 입소시설이 전무했다. 입소 시설로는 공동생활가정, 지역사회전환시설, 중독자재활시설 등이 있다.

특히 부산은 입소시설과 이용시설 간 정원 격차 비율도 높은 편이었다. 제주 입소시설과 이용시설의 정원이 각각 1.5명과 18.3명으로 격차가 가장 컸고, 부산이 1.9명과 11명으로 그 뒤를 이었다. 이용시설에는 주간재활시설, 직업재활시설 등이 포함된다. 정신재활시설 송국클럽하우스의 유숙 소장은 "거주지를 마련하지 못하거나 가족 갈등 등으로 병원에서 퇴원하지 못하는 정신질환자가 많다"며 "3만 4000여 명의 부산시 정신질환자를 위해 특히 입소시설을 확대할 필요가 있다"고 밝혔다.

부산에서 정신질환자를 위한 입소시설이 부족하다는 점은 오래전부터 제기된 문제다. 김경일 사회복지연대사무국장은 "부산은 지자체의 관심이 적었던 데다 님비(NIMBY) 현상으로 입소시설이 가로막히기도 했다"며 "정신질환자의 탈시설 과정에서 공동생활가정과 같은 입소시설이 도움이 될 수 있다"고 밝혔다.

중앙정신건강복지사업단이 발표한 &정신재활시설 현황&에 따르면 부산의 정신재활 입소시설 정원은 인구 10만 명당 1.9명에 불과함. 전국 17개 시·도 중 13위로 대전 15.1명, 충남 14.9명, 전북 12.7명 등에 비해 현저히 낮은 수준

[출처] 부산일보, 부산 정신질환자 입소시설 수 바닥권 , 2019.09.17, 이우영 기자
<http://www.busan.com/view/busan/view.php?code=2019091717534969012>

[참고자료 2-2. 부산 정신건강복지센터 현황]



- 기장군정신건강복지센터
- 부산광역시정신건강복지센터
- 금정구정신건강복지센터
- 남구정신건강복지센터
- 동구정신건강복지센터
- 동래구정신건강복지센터
- 부산진구정신건강복지센터

- 북구정신건강복지센터
- 사상구정신건강복지센터
- 사하구정신건강복지센터
- 서구정신건강복지센터
- 수영구정신건강복지센터
- 연제구정신건강복지센터
- 영도구정신건강복지센터
- 해운대구정신건강복지센터
- 중구정신건강복지센터
- 강서구정신건강복지센터

[출처] 부산광역시 (보건,건강)<https://www.busan.go.kr/depart/ahmental01>

[참고자료 2-3. 정신건강복지센터 월별 평균 이용률]

부산 인구 부산광역시 전체 333만 6,737명 / 인천 인구 인천광역시 전체 295만 7,066명
부산광역시와 인구가 비슷한 광역시인 인천광역시와 비교하였을 때 부산 정신건강
복지센터 월별 평균 상담건수 약 3회에 비해 인천 정신건강복지센터 월별 평균 상담
건수 약 8회 (2022년 7월 기준)임을 알 수 있다.

- 기장군정신건강복지센터 www.gijangmhc.or.kr/
- 부산광역정신건강복지센터 www.busaninmaum.com
- 금정구정신건강복지센터 www.mhmc.kr
- 남구정신건강복지센터 www.namgumhc.or.kr
- 동구정신건강복지센터 www.dmhmc.co.kr
- 동래구정신건강복지센터 www.dnmhc.co.kr
- 부산진구정신건강복지센터 www.beautymind.or.kr
- 북구정신건강복지센터 www.bmhmc.co.kr/
- 사상구정신건강복지센터 www.ssmhc.co.kr
- 사하구정신건강복지센터 www.shmhc.co.kr
- 서구정신건강복지센터 <http://bsmhmc.com/>
- 수영구정신건강복지센터 www.growthmhc.or.kr
- 연제구정신건강복지센터 www.ymhc.or.kr
- 영도구정신건강복지센터 <http://yeongdomind.or.kr/main/index.php>
- 해운대구정신건강복지센터 www.hmhc.or.kr
- 중구정신건강복지센터 <http://jmhc.co.kr/>
- 강서구정신건강복지센터 www.brightmaum.or.kr

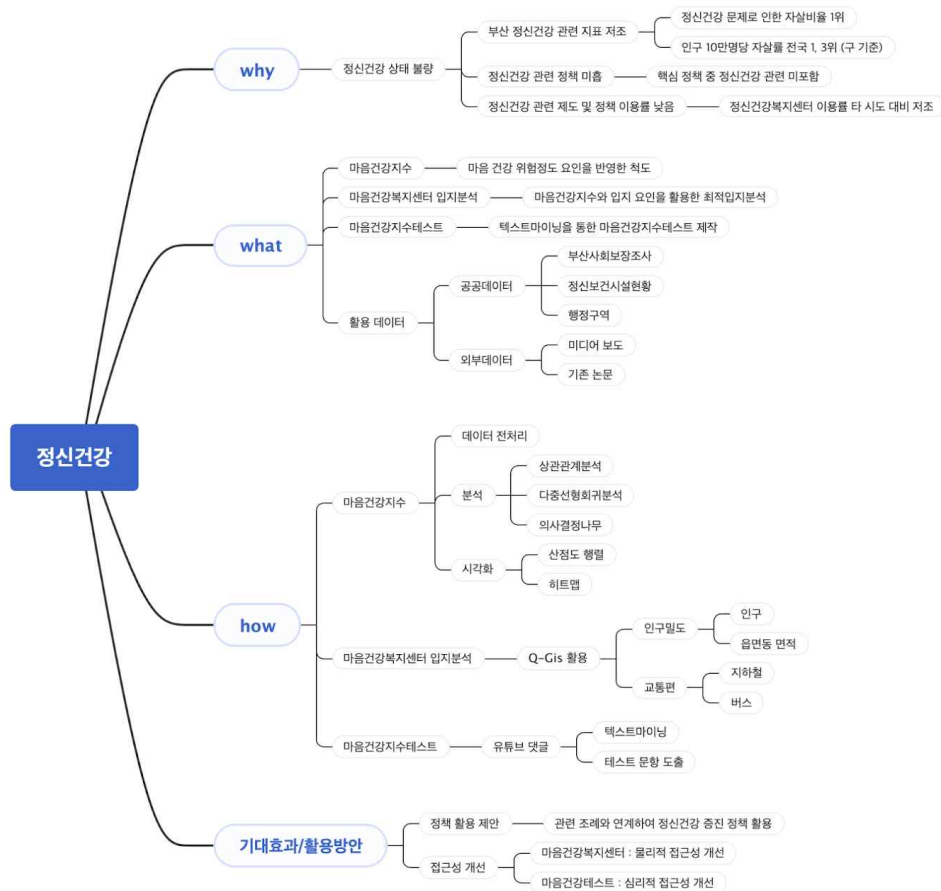
[출처] 각 구별 정신건강복지센터 사이트 상담 신청 현황

7. 부록

[부록]

1. 사례조사 정리

2. 주제설계를 위한 마인드맵



[그림 7-1] 정신건강 마인드맵

3, 분석 상세코드

댓글 웹크롤링

```
for url in range(len(you_list)):
    # 엑셀파일 쓰기
    # write_only=True 메모리 소모량 줄이기
    wb = Workbook(write_only=True)
    # 시트 생성
    ws = wb.create_sheet()
    # 구글 드라이브 위치
    driver_path="./chromedriver"
    driver = webdriver.Chrome(driver_path)
    time.sleep(3)
    # 다음에 사용....
    # keyword="잠 안올때 듣는 노래"
    # url = "https://www.youtube.com/"
    # driver.get(url)
    # driver.implicitly_wait(3)
    # time.sleep(1)

    driver.get(you_list[url])
    time.sleep(5)

    # execute_script 자바스크립트 코드사용
    # window.scrollTo(0, 800) 스크롤(0, N) 여기서 N은 heigjht
    driver.execute_script("window.scrollTo(0, 800)")
    time.sleep(2)

    # 스크롤 높이 가져오기 (초기 높이)
    l_ht = driver.execute_script("return document.documentElement.scrollHeight")

    while True:
        # 끝까지 스크롤 다운
        # window.scrollTo(0, N) 스크롤(0, N) 여기서 N은 heigjht
        driver.execute_script("window.scrollTo(0, document.documentElement.scroll
Height);")
        time.sleep(2)

        # 스크롤 다운 후 스크롤 높이 다시 측정
        n_ht = driver.execute_script("return document.documentElement.scrollHeigh
t")
        # 더 이상 스크롤 높이가 같다면 Break
```

```

        if n_ht == l_ht:
            break
        # 초기 높이 초기화
        l_ht = n_ht
    time.sleep(2)

    # 브라우저에 보이는 그대로의 HTML 가지고 오기
    html_source = driver.page_source
    soup = BeautifulSoup(html_source, 'html.parser')
    # 작성자 id 위치
    users_id = soup.select("div#header-author > h3 > #author-text > span")
    # 댓글 위치
    comment = soup.select("yt-formatted-string#content-text")

    id_list = []
    comment_list = []

    for i in range(len(comment)):
        user_id = users_id[i].text
        user_id = user_id.replace('\n', '')
        user_id = user_id.replace('\t', '')
        user_id = user_id.replace(' ', '')
        id_list.append(user_id) # 댓글 작성자
        user_comment = comment[i].text
        user_comment = user_comment.replace('\n', '')
        user_comment = user_comment.replace('\t', '')
        user_comment = user_comment.replace(' ', '')
        comment_list.append(user_comment) # 댓글 내용

    pd_data = {"아이디" : id_list , "댓글 내용" : comment_list}
    youtube_pd = pd.DataFrame(pd_data)
    youtube_pd.to_excel(f'./전처리/댓글{url + 1}.xlsx')
    driver.close()

#데이터 전처리
for n in range(len(file_list)):
    a = pd.read_excel(f'./전처리/댓글{n+1}.xlsx', index_col = '아이디')
    a.drop(['Unnamed: 0'], axis = 1, inplace=True)
    hangul = re.compile('[^ㄱ-ㅣ가-힣]+') # 한글과 띄어쓰기를 제외한 모든 글자
    result = hangul.sub('', a['댓글 내용'][10]) # 한글과 띄어쓰기를 제외한 모든 부분
    을 제거
    comment_list = []

```

```

for i in range(len(a['댓글 내용'])):
    user_comment = hangul.sub('', str(a['댓글 내용'][i]))

    # isspace() 문자열이 모두 공백이면 True 반환
    if not user_comment.isspace() and user_comment != '':
        comment_list.append(user_comment) # 댓글 내용
pd_data = {"댓글 내용" : comment_list}
youtube_pd = pd.DataFrame(pd_data)
youtube_pd.to_excel(f'./후처리/댓글{n + 1}.xlsx')

# 불용어 파일 불러오기
stopwords = pd.read_csv('stopwords.csv', encoding = 'cp949')
stop_words_list = [i for i in stopwords['pattern']]
len(stop_words_list)
from re import match # 전처리 위해서 정규표현식 관련 re 패키지 import
result = []
for c in ch:
    if len(c) >=2 and not c in stop_words_list:
        result.append(c)

result

#빈도분석
# Counter함수로 단어 빈도 카운트
from collections import Counter
# Counter(리스트의이름).most_common(숫자)
su = Counter(result).most_common(20)
tmp=0
for i in su:
    tmp+=i[1]
print(tmp)

#빈도분석 시각화
from wordcloud import WordCloud
import matplotlib.pyplot as plt
import numpy as np
from PIL import Image

```

```

image_file = './imgg.png'
img_file = Image.open(image_file)

alice_mask = np.array(img_file)

# WordCloud 객체에 전달할 폰트주소 생성
font_path = 'c:/Windows/Fonts/malgun.ttf'
wordcloud = WordCloud(
    font_path = font_path,
    width = 800,
    height = 800,
    background_color="white",
    contour_width=3,
    mask = alice_mask
)
wordcloud = wordcloud.generate_from_frequencies(tmp)
plt.figure(figsize=(10, 10))
plt.imshow(wordcloud, interpolation="bilinear")
plt.axis("off")
plt.show()

# 감정분석 머신러닝
# 에포크는 10으로 주어 학습을 진행
# 검증 정확도가 그전보다 낮아지면 학습을 멈춤

model = KR(**kargs)
model.compile(optimizer=tf.keras.optimizers.Adam(),
              loss = tf.keras.losses.BinaryCrossentropy(),
              metrics = [tf.keras.metrics.BinaryAccuracy(name='accuracy')])
#검증 정확도를 통한 EarlyStopping 기능 및 모델 저장 방식 지정
earlystop_callback = EarlyStopping(monitor='val_accuracy',
min_delta=0.0001, patience=2)
checkpoint_path = './감정분석데이터/DATA/weights.h5'
checkpoint_dir = os.path.dirname(checkpoint_path)
if os.path.exists(checkpoint_dir):
    print("{} -- Folder already exists \n".format(checkpoint_dir))
else:
    os.makedirs(checkpoint_dir, exist_ok=True)
    print("{} -- Folder create complete \n".format(checkpoint_dir))
cp_callback = ModelCheckpoint(
    checkpoint_path, monitor = 'val_accuracy', verbose=1, save_best_only =
True,
    save_weights_only=True
)
history = model.fit(train_input, train_label, batch_size=BATCH_SIZE, epochs

```

```
= NUM_EPOCHS,
    validation_split = VALID_SPLIT,
    callbacks=[earlystop_callback, cp_callback])
# 모델 저장하기
save_model(model, './감정분석데이터/Model/')
```


#크론바흐 알파 분석

```

from factor_analyzer import FactorAnalyzer
def CronbachAlpha(itemscores):
    itemscores = np.asarray(itemscores)
    itemvars = itemscores.var(axis=0, ddof=1)
    tscores = itemscores.sum(axis=1)
    nitems = itemscores.shape[1]
    return (nitems / (nitems-1)) * (1 - (itemvars.sum() / tscores.var(ddof=1)))
items=data.columns.tolist()
factors = ['B4', 'B5', 'B6', 'B7', 'B8', 'B11', 'B12', 'B13']
factors_items_dict = {}
for factor in factors:
    factors_items_dict[factor] = [x for x in items if factor in x]
#상관분석 히트맵 시각화
data.iloc[:,5].corr('spearman')
plt.figure(figsize=(10,10))
sns.heatmap(data = data.iloc[:,5].corr(method='spearman'), annot=True,
fmt = '.2f', linewidths=.5, cmap='Blues')

```

데이터 분포 확인

```

nrows = 5
ncols = 2
fig, axs = plt.subplots(nrows, ncols)
fig.set_size_inches(20,24)
for i in range(nrows):
    for j in range(ncols):
        num = i * ncols + j
        sns.histplot(x=df1.columns[num], data=df1, kde=True,
                    bins=30, ax=axs[i][j])

```

선형회귀모델, 가중치, 편향 확인

```

mlr = LinearRegression()
mlr.fit(x_train, y_train)
print('weight:', mlr.coef_)
print('bias:', mlr.intercept_)

```

회귀계수 순으로 정렬 및 시각화

```

coef_sort = coef.sort_values(ascending=False)
sns.barplot(x=coef_sort.values, y=coef_sort.index)

```

#회귀 분석

```

from sklearn.linear_model import LinearRegression
from sklearn.ensemble import RandomForestRegressor
from lightgbm import LGBMRegressor
from sklearn.metrics import mean_squared_error, r2_score
mlr = LinearRegression()
mlr.fit(x_train, y_train)
forest_model = RandomForestRegressor(n_estimators=100, random_state=42)
forest_model.fit(x_train, y_train)

```

```

lgb_model = LGBMRegressor(n_estimators=100, random_state=42)
lgb_model.fit(x_train, y_train)
model_list = [mlr, forest_model, lgb_model]
for model in model_list:
    model.fit(x_train, y_train)
    y_preds = model.predict(x_test)
    mse = mean_squared_error(y_test, y_preds)
    rmse = np.sqrt(mse)
    r2 = r2_score(y_test, y_preds)
    model_name = model.__class__.__name__
    print('{0} MSE : {1:.3f}, RMSE : {2:.3f}, r2 : {3:.3f}'.format(model_name, mse, rmse, r2))
#4보다 작으면 1, 크면 2의 값을 갖는 flag 변수 생성
df['flag']=[1 if i<=4 else 2 for i in df['정신건강']]
x=df.iloc[:,9:-1]
y=df['flag']
#SMOTE 이용해서 업샘플링
from imblearn.over_sampling import SMOTE
smote=SMOTE()
x_over,y_over=smote.fit_resample(x, y)
#Train, Test 8:2 비율로 쪼개기
from sklearn.model_selection import train_test_split
x=x_over.iloc[:,1:9]
y=x_over['정신건강']
train_x, test_x, train_y, test_y = train_test_split(x, y, test_size = 0.2, random_state = 42)
print(train_x.shape, test_x.shape, train_y.shape, test_y.shape) # 데이터 개수 확인
#RandomForestClassifier 모델(분류 분석)
from sklearn.ensemble import RandomForestClassifier
from sklearn.metrics import accuracy_score # 정확도 함수
clf = RandomForestClassifier(n_estimators=500, max_depth=30,random_state=0)
clf.fit(train_x,train_y)
pred_over_1 = clf.predict(test_x)
print(accuracy_score(test_y,pred_over_1))
from sklearn.metrics import mean_squared_error
RMSE = mean_squared_error(test_y, pred_over_1)**0.5
#분류모델 평가지표 확인
from sklearn.metrics import classification_report
print(classification_report(test_y,pred_over_1))
#Confusion Matrix 시각화
from sklearn.metrics import plot_confusion_matrix
plot_confusion_matrix(clf, test_x, test_y, cmap='BuGn')

#Feature Importance 시각화
forest_importances = pd.Series(clf.feature_importances_)
feature_scores = pd.Series(clf.feature_importances_,
index=train_x.columns).sort_values(ascending=False)
f, ax = plt.subplots(figsize=(15, 10))

```

```
ax = sns.barplot(x=feature_scores, y=feature_scores.index)
ax.set_title("변수별 중요도 점수 시각화", fontsize=18)
ax.set_xlabel("Feature importance score")
ax.set_ylabel("Features")
plt.show()
```

```

# 아규먼트로 이전에 만들었던
# matrix = 쌍대비교 값 행렬
# iterations = 반복 횟수
# eigenvector = 벡터
def compute_vector(self, matrix, iterations, comp_eigenvector=None):
    # 행렬거듭제곱연산
    sq_matrix = np.linalg.matrix_power(matrix, 2)
    row_sum = np.sum(sq_matrix, axis=1)
    total_sum = np.sum(sq_matrix)
    # 행렬 나눗셈
    principal_eigenvector = np.divide(row_sum, total_sum)
    # 0으로 채워진 Numpy array 생성
    if comp_eigenvector is None:
        comp_eigenvector = np.zeros(self.size)
    # 행렬 뺄셈
    remainder = np.subtract(principal_eigenvector,
                             comp_eigenvector).round(self.precision)
    # not(뽑아낼 요소가 있다면 True) = False
    if not np.any(remainder):
        return principal_eigenvector.round(self.precision)
    # 반복 횟수 1 차감
    iterations -= 1
    # 본인 함수 재호출
    if iterations > 0:
        return self.compute_vector(sq_matrix, iterations, principal_eigenvector)
    else:
        return principal_eigenvector.round(self.precision)

def compute(self):
    priority_vector = self.compute_vector(self.matrix, self.iterations)
    if self.cr:
        self.compute_consistency_ratio()
    # {컬럼 : 가중치, ...} 형태의 딕셔너리 생성
    weights = dict(zip(self.elements, priority_vector))
    # 가중치가 높은 순서로 정렬
    self.local_weights = dict(sorted(weights.items(),
                                     key=lambda item: item[1],
                                     reverse=True))
    self.node_weights = self.local_weights.copy()
    self.target_weights = self.node_weights

csv_data = pd.read_csv(f'{데이터}', encoding = 'CP949')

column = csv_data.columns.tolist()

#입지 선정 요소별 가중치 부여
index = column[1:]
print(index)

weight = [

```

```

        {
            index[0]: 0.061, #버스정류장
            index[1]: 0.09, #지하철정류장
            index[2]: 0.223, #정신보건시설
            index[3]: 0.217, #인구
            index[4]: 0.026, #고도
            index[5]: 0.383 #지수
        }
    ]

if (len(index) != len(weight[0])):
    print("오류")
else:
    print("진행")

#데이터 MinMax정규화 코드
from sklearn.preprocessing import MinMaxScaler
minmax_scaler = MinMaxScaler()

norm = csv_data.copy()

norm[index] = minmax_scaler.fit_transform(norm.loc[:, index])

norm

#정규화 데이터와 가중치를 사용하여 입지 선정 요소별 값 구하기
#인덱스 별 입지 선정 요소 계산
weight_data = norm.copy()

for i in range(len(index)):
    name = index[i]
    weight_data[name] = weight_data[name] * weight[0][index[i]]

weight_data['합계'] = weight_data['버스정류장'] + weight_data['지하철정류장'] -
weight_data['정신보건시설'] - weight_data['고도'] + weight_data['인구'] +
weight_data['지수']

weight_data['순위'] = weight_data['합계'].rank(method='dense',
ascending=False).astype(int)

```

```
weight_data
```

#입지 선정 데이터 그래프 시각화

```
import seaborn as sns
import matplotlib.pyplot as plt
sns.set(font="Malgun Gothic", font_scale=2.2,
rc={"axes.unicode_minus":False}, style='white')
fig, axs = plt.subplots(3, 2)
fig.set_size_inches(30,40)
n = 0
for i in range(3):
    for j in range(2):
        n += 1
        sns.barplot( y=gu_df_sum.columns[0], x = gu_df_sum.columns[n],
data=gu_df_sum,ax=axs[i][j], orient='h', palette="ch:start=.2,rot=-.3")

plt.show()
```