

프로젝트 요청서

프로젝트 명	빅데이터와 인공지능을 활용한 상관분석 서비스
주관기관	와이폴라리스(ypolaris)

1. 프로젝트 목표

취득한 데이터를 재가공하여 행정구역, 업종을 기본으로 하는 상권분석 웹 서비스를 구성한다.

세부 목표는 아래와 같다.

- a. 데이터를 취득하고 가공(전처리)하여 빅데이터화
- b. 빅데이터화 된 데이터를 머신러닝을 이용하여 회기분석 하여 상권분석 서비스를 위한 데이터베이스화
- c. b에서 구성된 데이터베이스를 기준으로 상권분석 웹 서비스 구축

세부 목표 a, b를 수행하여 지역이 추가되거나 해당 지역의 상권정보가 업데이트 되었을 때 데이터베이스를 추가/업데이트 하여 웹 서비스에 반영할 수 있는 프로세스를 구축되어야 한다.

또한 회기분석으로 분석된 상권 예측 정보에 의해 선택한 지역의 업종이 좋지 않은 상권에 속하는 경우 부족한 요소를 알려주고, 사업자 과밀 등에 의한 경우라면 비슷한 다른 지역 추천하는 등 기존 상권분석 서비스에서 제공되지 않는 서비스를 추가 될 수 있도록 한다.

2. 프로젝트 요구 스킬/툴

- | |
|--|
| <ol style="list-style-type: none">1) 수집/취득된 데이터에 대한 전처리 기술2) R (R-Studio) |
|--|

- 3) Python
- 4) 데이터베이스 사용 기술 (RDBMS – MariaDB, MySQL 등)
- 5) 웹 서비스 개발 기술(Linux, Python, Django/Flask 등)

3. 프로젝트 요구 정의서

- 1) 프로젝트 결과물은 웹 서비스로 구성
 - a. 서비스용 OS는 Linux를 사용하며 Python을 사용하여 Django/Flask등으로 구성
 - b. 빅데이터 저장소와 서비스용 데이터베이스는 RDBMS인 MariaDB 또는 MySQL을 사용한다
- 2) 서비스 구성
 - a. 시작페이지
 - i. 시도, 시군군으로 이루어진 지역과 업종을 선택할 수 있는 검색바와 서비스 사용 안내 페이지 구성
 - b. 검색 페이지
 - i. 페이지 레이아웃은 검색에 사용된 정보를 보여주는 검색바(좌측 패널)와 지도로 구성된 메인 콘텐츠로 구성한다
 - ii. 업종과 상권정보 분류 옵션은 수집/취득된 데이터를 전 처리하여 구성한다.
 - iii. 검색에 사용되는 주소는 도로명주소 안내시스템의 도로명주소 DB의 체계를, 지도 표기는 Kakao 지도 API을 이용한다.
 - iv. 검색 결과는 아래와 같이 데이터를 구성한다.
 - 1. 해당 지역의 인구밀집도와 편의점(슈퍼마켓, 마트 등 음식료품 판매 소매업종)수를 마커 클러스터리를 사용하여 표기한다.
 - 2. 행정구역명, 업종별 사업자 과밀정도, 기준 분기, 창업 위험도

등의 객관적 지표를 표기한다

3. 인공지능을 활용한 상권 예측치 정보

- a. 상권이 좋지 않다면 상권을 살리기 위한 요소 및 조건 등을 배치
- b. 사업자 과밀에 의해 창업추천도가 낮다면 조건이 비슷한 다른 지역을 추천 다른 지역을 추천
- c. 주변 역과의 거리, 사무실/학교/아파트 등의 주변 환경 데이터로 유동인구 추정치 산출하여 마커 클러스터링 구성하고 검색바에서 보임/숨김 관리

4. 사용자가 클릭한 위치의 건물상가 1층의 임대료(월)를 기준으로 매출(사용자 입력)/비용(임대료 자동 기입) 계산기 구성

c. 관리자 페이지

- i. 관리자 페이지는 관리자 계정으로만 로그인 가능해야 한다
- ii. 검색 페이지에 필터로 사용하는 업종과 상권정보 분류 옵션은 관리자 기능에서 추가/수정이 가능해야 한다.

3) 데이터 관리

- a. 모든 수집된 데이터는 데이터베이스화 되어야 하며, 필요시 가공/전처리를 거쳐 인공지능 학습/분석 기법에 사용되어야 한다
- b. 데이터베이스화 된 데이터는 RDBMS인 MariaDB 또는 MySQL에 저장한다.

4) 데이터의 수집 및 재가공

- a. 수집/취득해야 할 데이터는 아래와 같이 정의한다
 - i. 지역: 서울시 관악구, 강북구 등 (구 단위)
 - ii. 업종: 소매점 예) 편의점, 슈퍼마켓, 마트 등
- b. 데이터 소스는 아래 사이트들의 데이터를 이용한다
 - i. 도로명주소 DB - 도로명주소 안내시스템(개발자센터)
<https://www.juso.go.kr/addrlink/addrlinkJusoDBUse.do?menu=main&cPath=99MD>
 - ii. 시도/시구군별 인구현황 - 공공데이터 포털 (eg, 부산의 인구)
<https://www.data.go.kr/tcs/dss/selectDataSetList.do?keyword=%EC%>

9D%B8%EA%B5%AC+%22%EB%B6%80%EC%82%B0%22

- iii. 인구통계 - 국가통계포털
http://kosis.kr/statisticsList/statisticsListIndex.do?menuId=M_01_01&vwcd=MT_ZTITLE&parmTabId=M_01_01&statId=1962001&themald=A#A_4.2
- iv. 상가(상권)정보 - 공공데이터 포털
<https://www.data.go.kr/data/15012005/openapi.do>
- v. 상업용 부동산 임대동향 조사 > 임대정보 - 한국감정원
https://www.r-one.co.kr/rone/resis/statistics/statisticsViewer.do?menuId=RCS_51510
- vi. 상권별 소규모 상가 임대료 - 국가통계포털
http://kosis.kr/statHtml/statHtml.do?orgId=408&tblId=DT_40801_N4203_02&conn_path=I2
- vii. 전국주요상권현황표준데이터 - 공공데이터 포털
<https://www.data.go.kr/data/15029180/standard.do>
- viii. 소상공인시장진흥공단_상가(상권)정보_부산 - 공공데이터 포털
<https://www.data.go.kr/data/15059997/fileData.do>
- ix. 중소벤처기업부_전국주요상권현황 - 공공데이터 포털
<https://www.data.go.kr/data/3034800/fileData.do>
- x. 가맹사업정보제공시스템 > 가맹점 창업비용 - 공정거래위원회 가맹사업거래
<https://franchise.ftc.go.kr/user/extra/main/70/firHope/listBrand/jsp/LayOutPage.do>
- xi. 지역/업종별 사업자수, 총매출, 건당결제, 요일매출, 시간매출, 연령매출, 주거인구, 가구수 - 네모 웹서비스 스크랩
<https://www.nemoapp.kr/Search?ArticleType=1>
- xii. 상권평가등급, 점포이력 등 - 소상공인마당 스크랩
<http://sg.smba.go.kr/>
- xiii. 서울시 공공데이터: 서울시 열린데이터 광장 “골목상권” 검색
<https://data.seoul.go.kr/>
- xiv. 벤치마크 서비스
1. 통계지리정보서비스 > 대화형 통계지도

<https://sgis.kostat.go.kr/view/map/interactiveMap/companyView>

2. 네이버 부동산

<https://new.land.naver.com/complexes?ms=35.180143,129.075413,13&a=APT:ABYG:JGC&e=RETAIL>

3. 네모

<https://www.nemoapp.kr/Search?ArticleType=1>

c. 데이터셋의 구조는 아래와 같이 구성한다.

i. 상가(상권)정보

상가업소번호, 상호명, 지점명, 상권업종대분류코드, 상권업종대분류명, 상권업종중분류코드, 상권업종중분류명, 상권업종소분류코드, 상권업종소분류명, 표준산업분류코드, 표준산업분류명, 시도코드, 시도명, 시군구코드, 시군구명, 행정동코드, 행정동명, 법정동코드, 법정동명, 지번코드, 대지구분코드, 대지구분명, 지번본번지, 지번부번지, 지번주소, 도로명코드, 도로명, 건물본번지, 건물부번지, 건물관리번호, 건물명, 도로명주소, 구우편번호, 신우편번호, 동정보, 층정보, 호정보, 경도, 위도 등

ii. 지역/업종별 정보

1. 사업자 수

지역구분, 기간(년도와 분기), 업종, 사업자 밀집도(매우낮음|낮음|보통|높음|매우높음)

2. 총매출

지역구분, 기간(년도와 분기), 업종, 상권(발달상권|골목상권|관광특구|전통시장)

3. 건당결제

지역구분, 기간(년도와 분기), 업종, 상권(발달상권|골목상권|관광특구|전통시장)

4. 요일결제

지역구분, 기간(년도와 분기), 업종, 요일, 상권(발달상권|골목상권|관광특구|전통시장)

5. 시간매출

지역구분, 기간(년도와 분기), 업종, 시간대, 상권(발달상권|골목상권|관광특구|전통시장)

6. 연령매출

지역구분, 기간(년도와 분기), 업종, 연령대, 상권(발달상권|골목 상권|관광특구|전통시장)

7. 주거인구

지역구분, 기간(년도와 분기), 성별, 연령대

8. 가구수

지역구분, 기간(년도와 분기), 가구원수

5) 인공지능 학습법

- a. 데이터베이스화 된 데이터는 회기분석으로 학습하고 분석한다.

6) 보안 및 개인정보 준수

- a. 사용자 및 관리자 권한 설정을 통한 접근 제한 기능이 구현되어야 한다(관리자 페이지)
- b. 보안 코딩(Secure Coding)가이드라인을 준수하여 소스코드 작성
- c. 서비스는 한국 인터넷진흥원의 웹 서버 고안 관리 가이드를 준수하도록 한다.

7) 프로젝트 비밀 보장

- a. 본 프로젝트를 수행하는 과정을 통하여 취득한 제반정보를 비밀로 하여야 하며, "와이폴라리스"의 승인 없이는 이를 제3자에게 공개 또는 제공할 수 없다. 이를 위반할 경우, 와이폴라리스는 프로젝트 수행 인원 모두에게 손해배상을 청구할 수 있다.

8) 결과물의 저작권

- a. 본 프로젝트를 수행하는 과정에서 산출된 결과물 및 수정/개선을 통해 제작된 2차적 결과물에 대한 저작권 일체는 "와이폴라리스"가 소유한다.

9) 기타

- a. 프로젝트 요구사항 내 항목 변경이나 수정이 필요한 경우 양사간의 협의를 통해 진행한다.

4. 프로젝트 검수 사항

1) 프로젝트 검수 결과물
a. 서비스 데모
b. 프로젝트 산출물
c. 서비스 코드 및 데이터베이스 덤프
d. 수집/취득된 데이터, 가공된 데이터 등
2) 프로젝트 검수 방법
a. 제출된 결과물 검수
i. 서비스 데모, 코드 및 데이터베이스 덤프
ii. 프로젝트 산출물
iii. 수집/취득된 데이터, 가공된 데이터 등
3) 기타
a. 검수에 불합격된 결과물은 즉시 재 제출하여야 한다.

5. 프로젝트 팀 구성

구분	인원	역할
팀장	1	- 프로젝트 및 산출물 문서 작성 - 프로젝트 품질 관리 - 상권 데이터 수집 및 분석 - 상권 분석 서비스 개발
		팀원1:

팀원	3	<ul style="list-style-type: none"> - 상권 데이터 수집 및 분석 <p>팀원2/3:</p> <ul style="list-style-type: none"> - 상권 분석 서비스 개발
----	---	---

6. 프로젝트 커리큘럼

일정	커리큘럼	비고
1주차	1. 상권분석 방법과 데이터셋 I - 상권분석 방법 - 상권분석 조건과 관련 요소	자료조사 I
2주차	1. 상권분석 방법과 데이터셋 II - 데이터셋 수집 및 전처리 - 데이터 모델링 기법 - 회기분석 - 데이터 처리 언어 - R, Python	자료조사 II
3주차	1. 데이터 분석 기획 및 프로젝트 플랜 I - 상권분석 데이터 수집 및 분석 방법 기획	프로젝트 기획 I
4주차	1. 데이터 분석 기획 및 프로젝트 플랜 II - 수집, 분석된 데이터를 이용한 상권분석 서비스 플랜 작성	프로젝트 기획 II
5주차	1. 프로젝트 환경 구성	데이터 수집, 분석, 개발

	<ul style="list-style-type: none"> - 데이터 분석 및 개발 환경 구성 2. 데이터 수집 및 분석 I 3. 상권분석 서비스 개발 I 	
6주차	<ul style="list-style-type: none"> 1. 데이터 수집 및 분석 II 2. 상권분석 서비스 개발 II 3. 프로젝트 중간 점검 	
7주차	<ul style="list-style-type: none"> 1. 데이터 수집 및 분석 III 2. 상권분석 서비스 개발 III 	
8주차	<ul style="list-style-type: none"> 1. 데이터 수집 및 분석 IV 2. 상권분석 서비스 개발 IV 	
9주차	<ul style="list-style-type: none"> 1. 프로젝트 시연 및 개선 I 	
10주차	<ul style="list-style-type: none"> 1. 프로젝트 시연 및 개선 II 	
11주차	<ul style="list-style-type: none"> 1. 프로젝트 산출물 작성 	