

산학협력 프로젝트 결과보고서

본 보고서를 SW중심대학 단기 산학협력 프로젝트 결과보고서로 제출합니다.

(연구과제명 : 빅데이터와 인공지능을 활용한 상권분석 서비스)

2020. 11. 30.

참여기업 : 와이폴라리스 김승연 (인)

와이폴라리스 이현순 (인)

참여학생 : 수학과 허정훈 (인)

경제학과 권기호 (인)

수학과 이가영 (인)

컴퓨터공학과 강다미 (인)

참여교수 : 소프트웨어교육센터 육동철 (인)

참여인원 : 학부생 4 명, 기업체 : 2 명, 교수 : 1 명

부산대학교 소프트웨어교육센터장 귀하

산학협력 프로젝트 결과보고서[10페이지 이내]

과제명	빅데이터와 인공지능을 활용한 상권분석 서비스			
수행기간	2020.08.24.~2020.11.30.		책임교수명	육동철
협력기관명	와이폴라리스(YPolaris)		기관담당자명	김승연
참여학생	이름	학번	팀명	AI STAR
	팀장: 허정훈	201511152		
	팀원: 권기호	201542104		
	팀원: 이가영	201611129		
	팀원: 강다미	201645801		
참여인원	참여교수 1명(육동철), 학부생 4명(허정훈, 권기호, 이가영, 강다미), 협력기관 2명(김승연, 이현수)			

추진 배경 및 목표

빅데이터와 인공지능을 이용하여 소매 및 도매점에 한하여 무인점포를 위한 상권분석 서비스 웹페이지를 구성한다. 카카오 API를 이용하여 지도를 띄우고 메인페이지를 구성하여 사용자가 동을 선택하면, 유동인구, 경쟁업체수, 총인구, 구매력, 매출액대비 영업이익 등 5개의 변수로 생성된 방사형그래프를 지도에 띄운다. 이 때 상권의 척도는 좋음, 양호, 고려, 나쁨으로 나타내어 사용자가 시각적으로 판단할 수 있게 해준다.

- 데이터 출처

1) 상가정보 - 공공데이터 포털

: 부산에 있는 행정동별 상가정보를 추출하여 편의점, 대형마트, 슈퍼마켓으로 구분하여 위도 경도와 각 카테고리별 점포 수와 경쟁업체 수를 얻었다.

2) 읍면동별 세대 및 인구 - 통계청(KOSIS)

: 2019년도 부산시의 총 인구 수를 최종데이터에 추가했다.

3) 구군별 총 카드이용금액 - 부산시 빅데이터포털

: 3개월간 총 카드이용금액을 이용하여 선형회귀에 의한 예측구매력을 최종데이터에 추가했다.

4) 부산시 읍면동별 매출액 - 통계청(KOSIS)

: 읍면동별 매출과 영업이익 데이터를 이용해 매출대비 영업이익 데이터를 얻었다.

5) 읍면동별 요일별 유동인구 - 부산시 빅데이터 포털

: 요일별 유동인구로 2019, 2020년도 총 유동인구 데이터를 얻어 사용했다.

수행 결과

- 최종데이터

데이터 출처에서 얻은 데이터를 가지고 읍면동별 총유동인구와 총인구, 편의점, 대형마트, 슈퍼마켓의 각 점포 수와 경쟁업체수, 영업이익대비 매출액, 구매력을 열로 가지는 데이터를 생성했다. 그리고 선형회귀를 이용하여 예측 유동인구와 예측 구매력을 얻었다.

- 선형 회귀

변수가 시간과 유동인구 2개뿐이었기에 LinearRegression을 통해 다음 달의 유동인구를 예측했다. 마찬가지로 구매력도 예측해볼 수 있었다. LinearRegression 이외의 다른 model도 사용해보려 했지만, 예를 들면 RandomForest, 변수가 2개뿐이고 기간도 크지 않아 제대로 된 예측이 이루어지지 않았다. 따라서 LinearRegression이 우리의 예측에 가장 좋은 model이라고 판단했다.

predict_people('중앙동', '총 유동인구')

153358명



RandomForestRegressor 사용

- 미래 시간에 대한 예측이 이루어지지 않아서 사용 불가

```
from sklearn.ensemble import RandomForestRegressor

predict_list=data[data['읍면동']!='장전3동']['읍면동']
predict_people_all2=[]
for i in predict_list:
    data['년월값']=0
    data['년월값']=pd.to_datetime(data['년월'])
    data['년월값']=data['년월값'].map(dt.datetime.toordinal)

    X=pd.DataFrame(data[data['읍면동']==i]['년월값'])
    y=pd.DataFrame(data[data['읍면동']==i]['총 유동인구'])

    X['년월']=X['년월값'].map(dt.date.fromordinal)
    X['년월']=pd.to_datetime(X['년월'])

    data_list=X['년월'].to_list()
    data_list=pd.DataFrame({'year_month': data_list})

    X=X.reset_index()
    X['년월']=df['year_month'].dt.strftime('%Y-%m')

    line_fitter=RandomForestRegressor(max_depth=2, random_state=1)
    line_fitter.fit(X['년월값'].values.reshape(-1,1),y)

    a=line_fitter.predict([[X['년월값'][11]+30]])
    predict_people_all2.append(int(np.round([a][-1])))

line_fitter.predict([[X['년월값'][11]+30]])

array([1101241.17316667])

line_fitter.predict([[X['년월값'][11]]])

array([1101241.17316667])
```

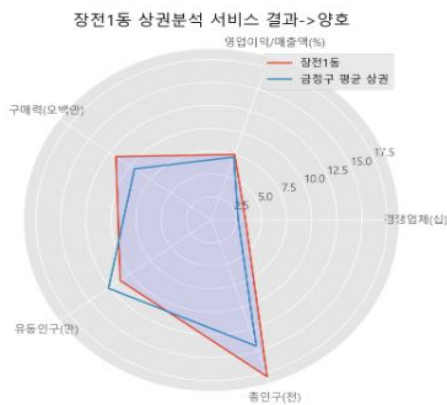
- 방사형 그래프

원하는 상권과 그 상권이 속한 구의 평균 상권의 차이를 구해서 각 항목의 합을 구했다.

단위가 차이 나기 때문에 개별 컬럼별로 Z-normalization을 해주고 가중치를 조절했다.

소상공인 상권분석 서비스와 비교하여 가장 비슷했던 [1.8, 1.0, 1.4, 1.0, 1.8]을 가중치로 쓰기로 결정했다. (경쟁업체, 영업이익/매출액, 구매력, 유동인구, 총인구) 하위 25% : 나쁨, 하위 25%-평균 : 고려, 평균-상위 25% : 양호, 상위 25% : 좋음이라고 평가했다.

장전1동



42.2입니다. [1.8, 1.0, 1.2, 1.8, 1.2]
 42.7입니다. [1.8, 1.0, 1.2, 1.8, 1.4]
 42.2입니다. [1.8, 1.0, 1.2, 1.8, 1.6]
 43.2입니다. [1.8, 1.0, 1.2, 1.8, 1.8]
 42.7입니다. [1.8, 1.0, 1.2, 1.8, 2.0]
 43.7입니다. [1.8, 1.0, 1.2, 2.0, 1.0]
 42.7입니다. [1.8, 1.0, 1.2, 2.0, 1.2]
 43.2입니다. [1.8, 1.0, 1.2, 2.0, 1.4]
 42.2입니다. [1.8, 1.0, 1.2, 2.0, 1.6]
 43.2입니다. [1.8, 1.0, 1.2, 2.0, 1.8]
 43.2입니다. [1.8, 1.0, 1.2, 2.0, 2.0]
 43.2입니다. [1.8, 1.0, 1.4, 1.0, 1.0]
 43.2입니다. [1.8, 1.0, 1.4, 1.0, 1.2]
 42.7입니다. [1.8, 1.0, 1.4, 1.0, 1.4]
 43.7입니다. [1.8, 1.0, 1.4, 1.0, 1.6]
 46.2입니다. [1.8, 1.0, 1.4, 1.0, 1.8]
 44.2입니다. [1.8, 1.0, 1.4, 1.0, 2.0]
 41.7입니다. [1.8, 1.0, 1.4, 1.2, 1.0]
 43.2입니다. [1.8, 1.0, 1.4, 1.2, 1.2]
 43.2입니다. [1.8, 1.0, 1.4, 1.2, 1.4]
 42.2입니다. [1.8, 1.0, 1.4, 1.2, 1.6]
 44.2입니다. [1.8, 1.0, 1.4, 1.2, 1.8]
 45.2입니다. [1.8, 1.0, 1.4, 1.2, 2.0]
 42.2입니다. [1.8, 1.0, 1.4, 1.4, 1.0]
 41.7입니다. [1.8, 1.0, 1.4, 1.4, 1.2]
 41.7입니다. [1.8, 1.0, 1.4, 1.4, 1.4]
 41.7입니다. [1.8, 1.0, 1.4, 1.4, 1.6]

[파일 구조]

yproject

- code
- static
- img : 상권분석 결과 출력을 위한 이미지
 - 고려.png
 - 나쁨.png
 - 양호.png
 - 좋음.png
- json : json 파일이 저장돼 있음
 - geo.json : polygon좌표가 포함된 json파일
- templates
- main.html : 웹 메인페이지 html
- __init__.py : main flask 코드

[테이블 사용]

```
mysql> show tables;
+-----+
| Tables_in_retail_info |
+-----+
| busan_dong_analysis    |
| busan_dong_analysis_graph |
| busan_dong_boundary    |
| busan_dong_population  |
| busan_store            |
| busan_store_select2    |
+-----+
6 rows in set (0.02 sec)
```

busan_store

: 구와 카테고리를 선택했을 때 마커 출력을 위한 상가 좌표, 주소, 이름을 출력하기 위해

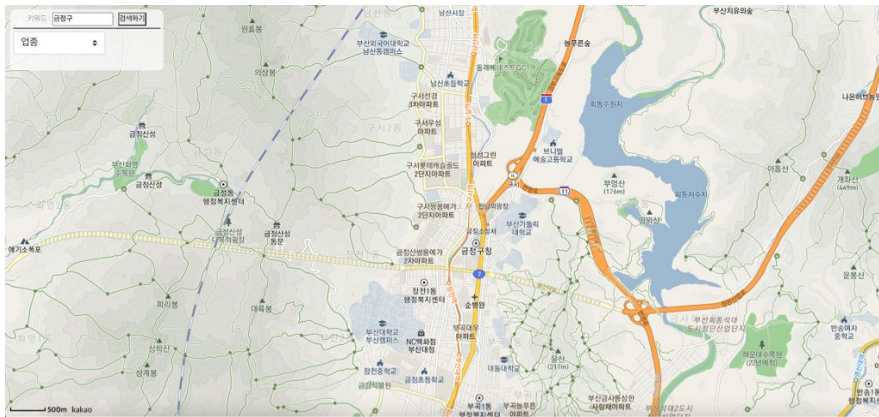
busan_dong_population

: 부산의 구에 속한 행정동 정보와 동별 인구를 출력하기 위해

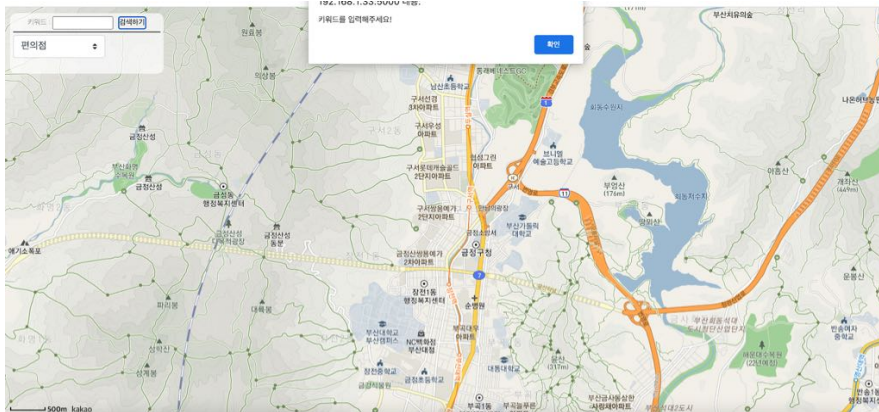
busan_dong_analysis_graph

: 상권 분석 결과를 방사형 그래프로 출력을 하기 위해 사용

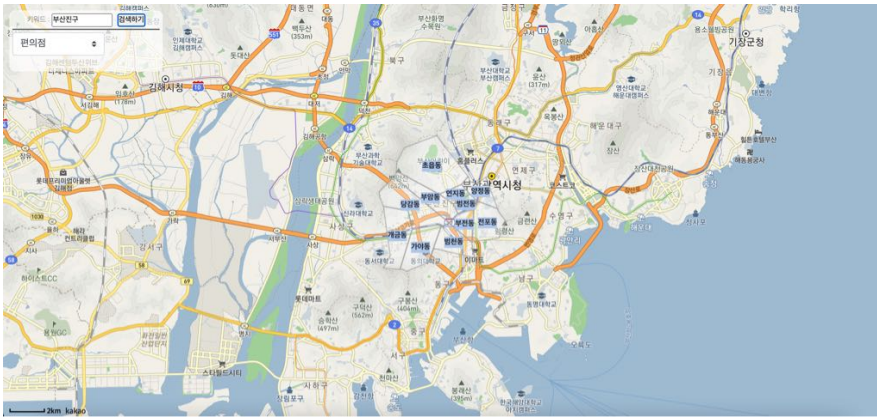
과제 성과 및 향후 계획



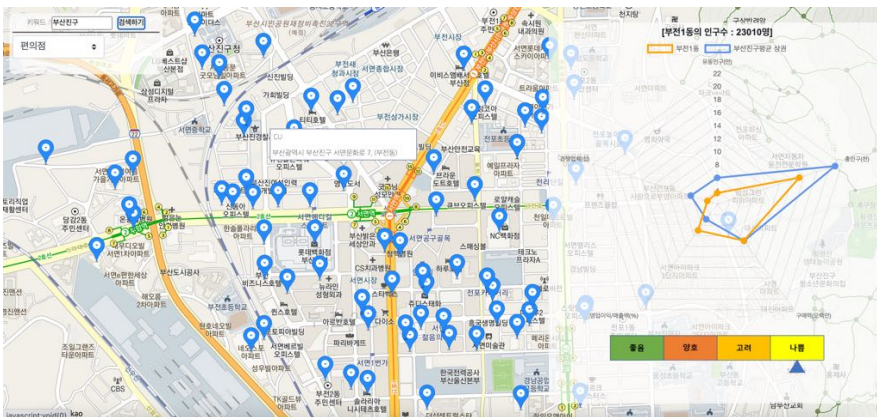
[그림1] 사이트에 처음 접속했을 때 볼 수 있는 메인 페이지이다.



[그림2] 사이트에 처음 접속했을 때 볼 수 있는 메인 페이지이다. 구를 입력하지 않고 카테고리만 설정하고 클릭했을 시 화면이다.



[그림3] 구와 카테고리를 선택하고 검색을 하면 해당 구의 경계와 해당 동을 표시한다.



[그림4] 동을 선택하여 지도를 확대하면 상가를 나타내는 마커가 출력되고 해당 동의 상권분석 결과를 그래프로 보여준다. 상권분석 평가는 위의 4가지 형태 중 하나로 나타낸다.



[그림5] 동을 선택하여 지도를 확대 하면 상가를 나타내는 마커가 출력되고 해당 동의 상권분석 결과를 그래프로 보여준다. 상권분석 평가는 위의 4가지 형태 중 하나로 나타낸다. 해당 그림은 금정구 구서1동에서 편의점을 선택했을 시 출력 화면이다.

< 결과 보고서 작성 주의사항 >

1. 산학협력 프로젝트 추진배경 및 목표

- 산학협력 프로젝트를 수행하게 된 배경 및 필요성, 중요성 등을 구체적으로 기술

2. 산학협력 프로젝트의 수행결과

- 프로젝트 수행 방법과 결과 내용을 구체적으로 기술
- 프로젝트 결과물 제출
 - 사용자 UI & 관리자 UI이 있을 경우 화면 캡처 제출
 - 실행되는 사이트 URL, 앱 apk 다운로드 URL제출
 - 데이터베이스 사용을 했을 경우 테이블 명세서 제출
 - 인공지능 학습에 해당되는 경우 학습 모델 관련 내용과 학습 데이터 일부 제출
- 프로젝트 수행 내용과 관련한 사진, 문헌(논문, 보고서 등), 데이터, 인증서, 제품 등을 포함

3. 과제성과 및 향후 계획

- 산학협력 프로젝트 성과 및 향후 계획 내용을 기술