

이삭줍이

수확행

맞춤형커농전락추천모델



CONTENTS



소개



배경



데이터
수집 및 활용



모델
구축 결과



시연



결론

The background is a light beige color with a subtle texture. On the right side, there are several golden wheat stalks with long awns, arranged diagonally. On the left side, there is a large white geometric shape that resembles a stylized 'L' or a corner cutout, with a curved top edge.

소개

수확행이란?

나만의 텃밭을 가꾸는
소소하지만 **확실한 행복**부터



풍성한 곡식을
수확하는 **행복**까지



사람[人]에게 확실한 행복을 플러스[+]해주는
“ **맞춤형 귀농 전략 추천 모델** ”

CHONBUK UNIV.

지역 문화·산업 융복합 빅데이터 전문가 과정

#DB #R #지역문화콘텐츠 #파이썬

비IT계열 전공자 우대

접수 마감

빅데이터 교육과정

빅데이터 기초 역량 오효정 교수님
양동민 교수님
김순태 교수님
우선미 선생님

- Python, R, DB
- 통계학개론
- 엑셀 기반 데이터 분석

빅데이터 처리&분석 역량 이말레 교수님

- 기계학습
- 인공지능
- 고급 빅데이터 분석

빅데이터 프로젝트 박태연 박사님
한희정 박사님

- 지역문화콘텐츠
- 지역 문화·산업 융복합 빅데이터 프로젝트

좋아요 5650개

CHONBUK UNIV._ #DB #R #지역문화콘텐츠 #파이썬 **#비IT계열 전공자**

댓글 45개 모두 보기

유호선 [기록관리학과]

- 팀장(전체 총괄)
- 기술 부분
 - 데이터 학습
 - API 연동

강주연 [기록관리학과]

- 발표
- 모델 설계
 - 타당성 및 근거 확보
 - 데이터 수집

김명규 [도시공학과]

- 모델 설계
 - 데이터 수집
 - 데이터 정제

김태영 [기록관리학과]

- 모델 설계
 - 타당성 및 근거 확보
 - 데이터 수집

심서령 [IT응용시스템공학과]

- 기술 부분
 - 데이터 학습
 - API 연동

좋아요 350개

이삭줍이_ #한알 한알 이삭 줍듯이 필요한 정보를 차곡차곡 모아 맞춤으로 제공

#성공적 #귀농 #멘토 오효정 교수님 #멘토 채홍석 선생님



배경

• 최근 귀농인구 증가 추세

- 최근 도시생활을 접고 농사를 짓기 위해 **귀농하는 인구가 급증**하고 있음
- 통계청에 의하면 2017년 기준 전국 귀농인 수는 12,763명으로, **2013년 대비 약 24%가 증가**하였음
- **BUT**, 치밀한 준비없이 막연한 기대감만 갖고 귀농했다가 **다시 도시로 돌아가는 사례(역귀농)도 다수**임

소득원 못 찾고 원주민과 갈등 준 비 안 된 귀농인 '다시 도시로'

역귀농·귀촌을 선택한 주원인은 소득원이나 **귀농·귀촌 환경에 대한 사전 준비나 교육이 부족**한 것이다.

..... 중략

역귀농·귀촌 사례를 연구했던 한 연구원 역시 '**준비 기간이 비교적 짧고, 지역에 대한 충분한 조사 없이 떠나는 경우**'가 가장 흔한 도시로의 재이주 원인이라고 밝혔다. **농사에 필요한 노동력, 투자 비용, 판로 및 수익에 대한 충분한 시뮬레이션이 없었다는 뜻이다.**

중앙선데이(2016.09.25)

어설픔게 덤볐다간 낭패...귀농인 10명 중 1명 짐 싣다

실패 없는 귀농...철저한 준비가 중요

귀농을 고민한다면 먼저 **농업 관련 기관·단체나 선배 귀농인을 방문해 필요한 정보를 수집**해야 한다.

..... 중략

가족 합의를 통해 귀농이 결정되면 **영농기술을 충분히 배우고 익힌 뒤 선택한 작목의 재배여건이나 생활환경 등을 고려해 정착지를 선택**해야 한다. 귀농·귀촌 관련 박람회를 찾아 정보를 모으는 것도 중요하다. 정부나 지자체의 지원 시책도 눈여겨봐야 한다.

연합뉴스(2017.02.04)

귀농·귀촌, 실패하는 유형이 있다

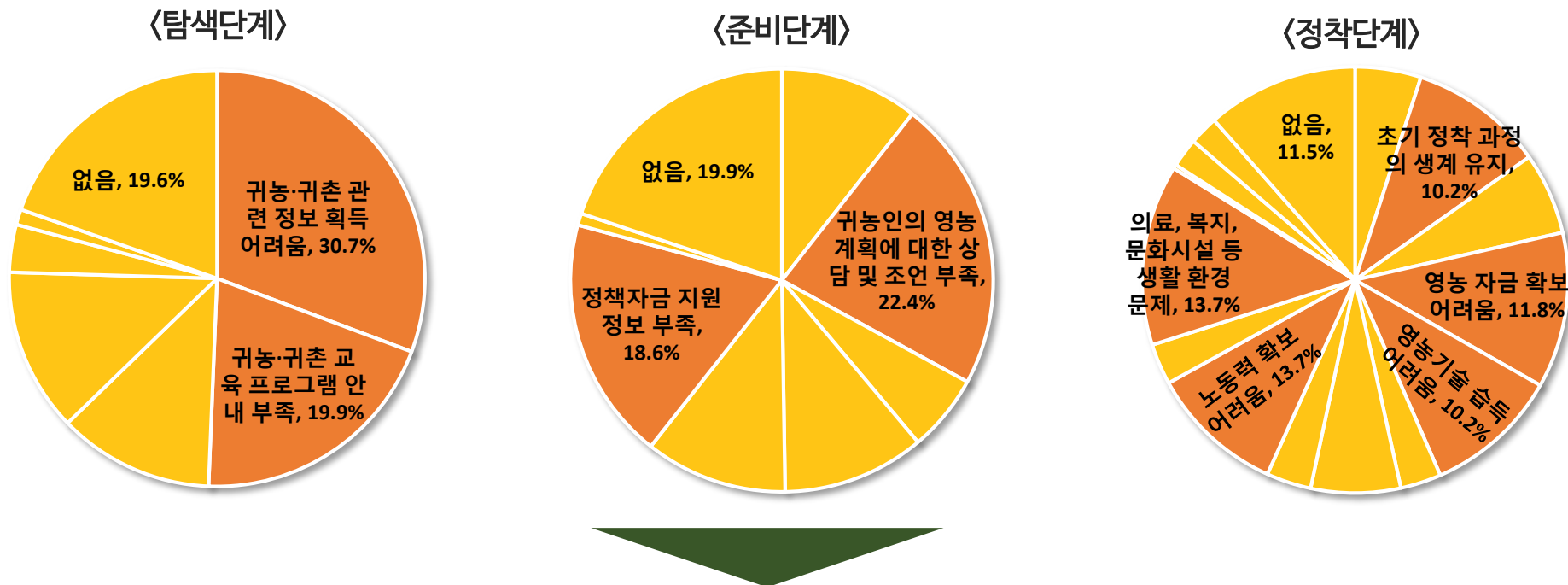
귀농·귀촌은 **준비과정을 제대로 거치지 않으면 실패할 수밖에 없다.**

..... 중략

두번째 유형은 지역 선정에 실패한 경우다. 젊은 층이 자녀가 다닐 학교를 고려하지 않고 농촌으로 이주했다가 자녀가 입학연령이 돼 어려움을 겪어 재이주하는 사례가 더러 있다. 가령 농사가 제법 자리를 잡았는데 자녀가 다닐 학교가 없어 난처한 상황에 부딪치는 것이다. 나이가 들면서 병원과의 접근성 문제로 역귀농을 고민하는 경우 역시 마찬가지다. 이는 **처음부터 귀농지역을 선정할 때 여러가지 상황을 고려하는 신중함이 필요**하다는 뜻이다.

농민신문(2018.06.18)

- 귀농 실패 요인_귀농 시 어려움 (2017 통계청 창녕군 귀농귀촌 실태조사)



귀농 시 겪는 어려움으로 대부분이 정보 부족을 지목

귀농 관련 정보를 종합적으로 이용자에게 제공한다면
귀농 실패 확률을 줄일 수 있지 않을까?

• 맞춤형 귀농 전략 추천 모델



Step 1) 이용자 정보 수집

이용자의 정보(특징)를
(EX:나이, 성별, 환경요인 등)
입력 받는다



Step 2) 적합한 정착지역 추천

이용자의 정보(특징)에 적합한
정착지를 3곳 추천한다



Step 3) 귀농정보 제공

이용자가 입력한 정보(특징)에 가장 적합한
집과 농지, 농지에 맞는 **작물**과 최종
귀농비용을 제공한다

귀농 전략 추천 모델

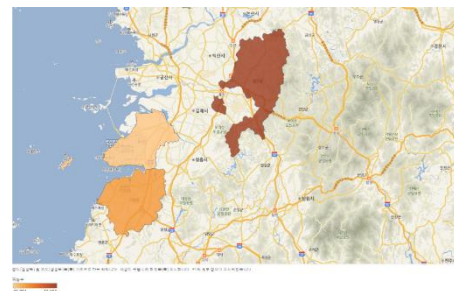
귀농사유: 그 밖의 사유

| | | |
|------|----|----|
| 세대원1 | 선택 | 선택 |
| 세대원2 | 선택 | 선택 |
| 세대원3 | 유아 | 선택 |
| 세대원4 | 선택 | 선택 |

시설 선택

| | | | |
|-------|--------|--------|-----|
| 공공시설 | 관공서 | 경찰서 | 소방서 |
| 교육시설 | 유치원 | 학교 | 학원 |
| 편의시설 | 음식점 | 카페 | 편의점 |
| | 은행 | 시장 | 마트 |
| 복지시설 | 노인복지회관 | 지역아동센터 | |
| 의료시설 | 병원 | 약국 | |
| 문화시설 | 영화관 | 공연장 | |
| 교통 시설 | 기차 | 여객선 | 터미널 |

Click Me



지역추천

1순위 : 완주군 ✓

2순위 : 고창군 ○

3순위 : 부안군 ○

추천 결과 (완주군)

- 추천 지역
- 농지/주택구매 정보
- 추천 작물/축산물 정보
- 작물/축산물 판매 정보
- 주변 편의시설

[귀농비용 산출]

| | |
|--------|-------|
| 필요 자본금 | 기대 수익 |
|--------|-------|

• 기존 연구와의 차별성

❖ 서울대학교 (2016). 빅데이터를 이용한 귀농수요 분석과 귀농인의 성공적 모델 연구

- 귀농에 대한 인식과 정책 수요 파악
- 지자체별 귀농인구 유입규모 예측
- 귀농인 특성, 귀농 패턴, 귀농 단계별 특성파악
- 귀농인들이 주로 선택하는 품목의 성공 및 실패요인 분석 후 성공모형 제시
 - 빅데이터 분석은 일부(인구통계)에서만 사용
 - 설문조사를 실시한 값을 사용하여 실제 행위와 연구결과 간의 차이가 있을 수 있음

❖ 전라북도 (2017). 귀농, 귀촌 서비스 모델 빅데이터 분석 연구

- 귀농 귀촌 데이터 구축 및 분석
- 귀농, 귀촌을 구분하여 의사결정에 영향을 미치는 요인 분석
- 귀농, 귀촌 서비스 모델(알고리즘) 개발
 - 귀농, 귀촌 추천 서비스를 제공하고자 하였으나 단일변수만 고려한 결과를 제공함

VS

수확행

맞춤형 귀농 전략 추천 모델

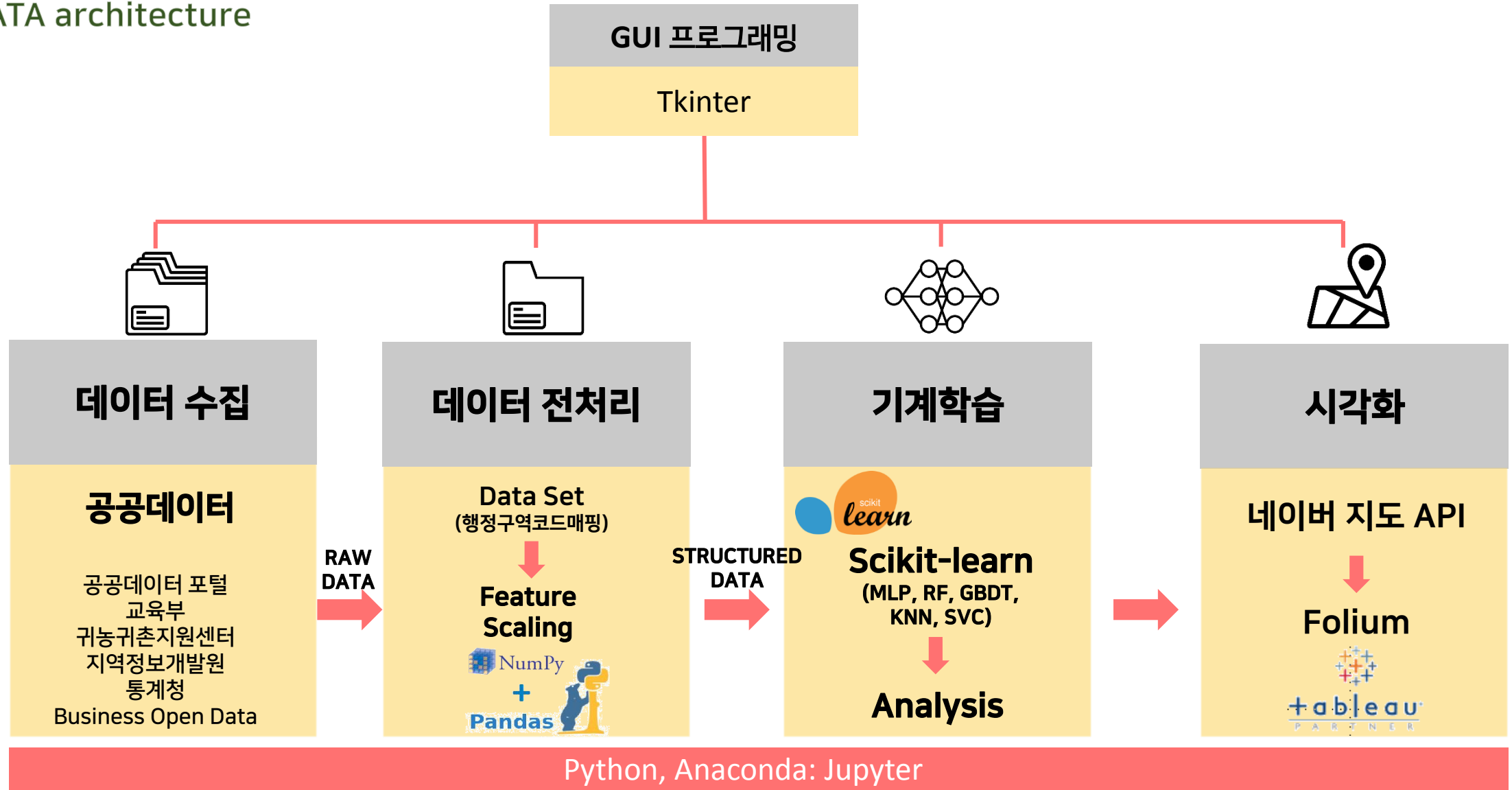
- ✓ 귀농지 추천 과정 전반에 걸쳐 필요로 한 빅데이터를 활용했으며 실제 과거 이력에 기반한 연령대별 맞춤 귀농지 추천이 가능함
- ✓ 농사 관련 데이터 이외 실제 생활권역별 빅데이터 활용을 통해 귀농시 직면하는 현실적인 문제 반영
- ✓ 귀농, 귀촌 시 고려되어지는 다양한 요인을 다차원 변수로 치환하여 종합적으로 반영한 결과를 추천함

3

데이터 수집 및 활용

3. 데이터 수집 및 활용

- DATA architecture



수집 데이터 목록



| 범주 | | 데이터명(주요 속성) | 출처 |
|----------|---------|---|---------------------|
| 정착지(지역) | 인구정보 | 국내인구이동통계(전입행정구역, 전입자 연령), 농업총조사(총가구수, 농가수) | 마이크로데이터-통계청 |
| | 지원정책 | 귀촌귀농정책정보 (지원 예산 정책, 교육 지원 정책) | 귀농귀촌지원센터 |
| | 교육시설 | 초·중·고 학교 정보, 학원(교과, 평생교육) | 교육부 |
| | 의료시설 | 병원, 약국 | Business Open Data |
| | 편의시설 | 마트 및 편의시설, 카페, 일반음식점 정보 | |
| | 문화/복지시설 | 문예회관, 문화의집, 노인복지회관, 지역아동센터 | |
| | 공공시설 | 경찰서, 소방서, 관공서 | 공공데이터포털 |
| | 교통정보 | 버스정류장 현황 | 전국버스운송사업조합연합회 |
| 작물 및 축산물 | 수익성 | 지역별작물정보(연간 운영비) 특용작물생산현황, 농특산물재배현황 | 지역정보개발원 공공데이터포털 |
| | | 축산업등록DB | 지역정보개발원 |
| | | 농가경제조사(농가소득, 가계지출) | 통계청 |
| | 재배/생육환경 | 작물별토양적성도 농지전용현황도, 농지정보도 농림수산물식품 생산·가공 R&D 정보 축산농장정보, 농작업일정 정보 자동기상관측 조회 API, 농업기상정보 | 귀농귀촌지원센터 공공데이터포털 |
| 농지 및 주택 | 농지 | 공시지가 API | 공공데이터 포털 |
| | 주택 | 빈집 DB | 전라북도 귀농귀촌지원센터 |
| | | 귀농인의 집 DB | |
| | | 건축물 정보 | 공공데이터 포털 |

3. 데이터 수집 및 활용

• 데이터 활용



Step 1) 이용자정보 수집



Step 2) 적합한 정착지역 추천



Step 3) 귀농정보 제공



작물 및 축산

지역별작물정보(연간 운영비)
특용작물생산현황, 농특산물재배현황
농가경제조사(농가소득, 가계지출)
작물별토양적성도
농지전용현황도, 농지정보도
농림수산물식품 생산·가공 R&D 정보
축산농장정보, 농작업일정 정보
자동기상관측 조회 API, 농업기상정보



농지정보

공시지가 API



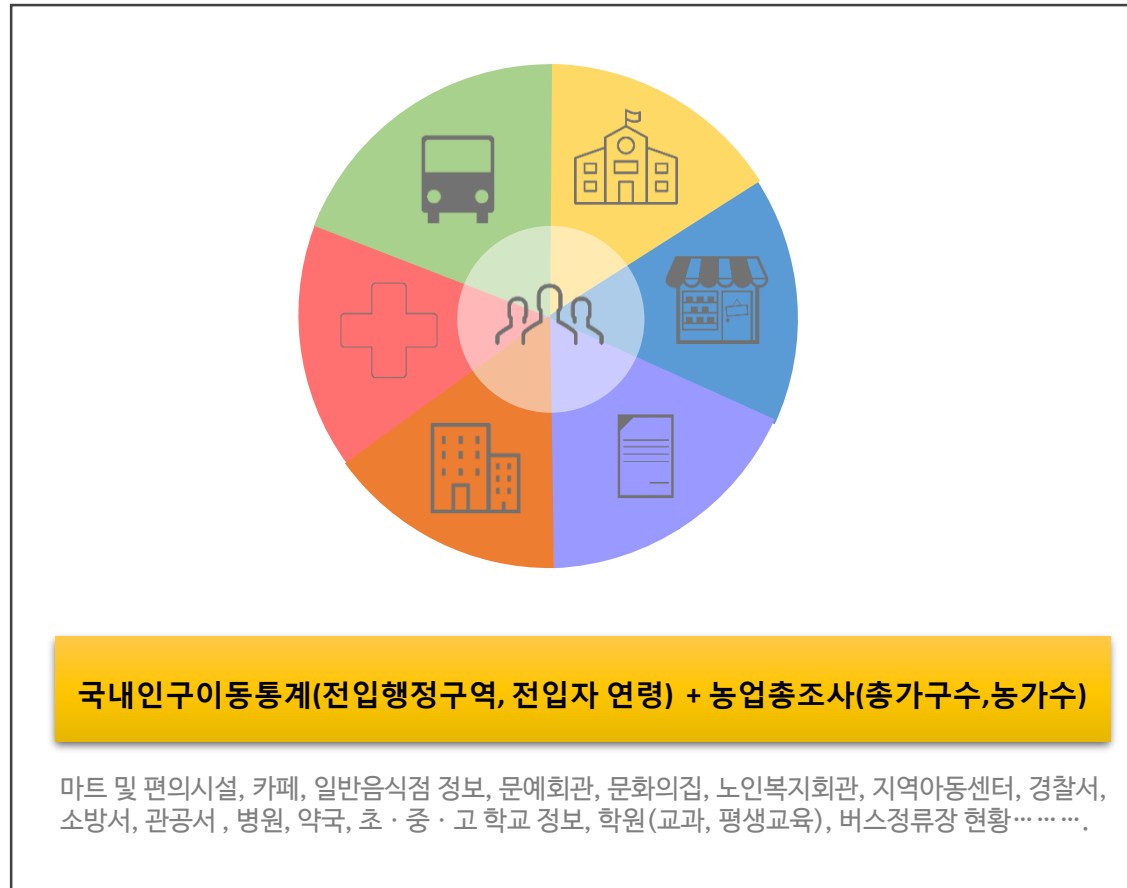
주택정보

빈집 DB
귀농인의 집 DB
건축물 정보

3. 데이터 수집 및 활용



- 기계학습(Machine Learning)
 - 전체데이터(196, 320세대)



Training Set
147,240세대, 75%

Test Set
49,080세대, 25%

4

모델 구축 결과



```
# 데이터 읽어 들이기
row_data = pd.read_csv("edu_row.csv", header=1)
```

```
# 데이터 내부의 분류 변수 전개하기
label = []
data = []
```

```
for row_index, row in mr.iterrows():
    label.append(str(row.ix[0]))
    row_data=[]
    if row.ix[2]==1:
        row_data.extend(a1)
    if row.ix[2]==2:
        row_data.extend(a2)
    if row.ix[2]==3:
        row_data.extend(a3)
    if row.ix[2]==4:
        row_data.extend(a4)
    if row.ix[2]==5:
        row_data.extend(a5)
    if row.ix[2]==6:
        row_data.extend(a6)
    if row.ix[2]==9:
        row_data.extend(a7)
    for v in row.ix[3:21]:
        row_data.append(round(v,1))
    data.append(row_data)
```

변수 코딩(전처리)

- * 이산형 변수
one-hot-encoding
- * 연속형 변수
→ 단위 통일, 반올림

데이터 분류

```
#학습 전용 데이터와 테스트 전용 데이터로 나누기
data_train, data_test, label_train, label_test = train_test_split(data, label,
                                                                    train_test_split=0.75)
```

- *Train : 75%
- *Test : 25%

#데이터 학습시키기 및 테스트

```
clf = RandomForestClassifier(n_estimators=1000,max_depth=None,min_samples_split=10)
clf.fit(data_train, label_train)
predict = clf.predict(data_test)
ac_score = metrics.accuracy_score(label_test, predict)
print("RF정답률 =", ac_score)

gbdt = GradientBoostingClassifier(n_estimators=100,max_depth=4, min_weight_fraction_leaf=0.1)
gbdt.fit(data_train, label_train)
predict2 = gbdt.predict(data_test)
ac_score2 = metrics.accuracy_score(label_test, predict2)
print("GBDT정답률 =", ac_score2)

mlp=MLPClassifier(hidden_layer_sizes=[100,100])
mlp.fit(data_train, label_train)
predict3 = mlp.predict(data_test)
ac_score3 = metrics.accuracy_score(label_test, predict3)
print("mlp정답률=",ac_score3)

svc=SVC()
svc.fit(data_train, label_train)
predict4 = svc.predict(data_test)
ac_score4 = metrics.accuracy_score(label_test, predict4)
print("svc정답률=",ac_score4)

neigh=KNeighborsClassifier()
neigh.fit(data_train, label_train)
predict5=neigh.predict(data_test)
ac_score5 = metrics.accuracy_score(label_test, predict5)
print("knn정답률=",ac_score5)
```

학습 및 테스트

- * Scikit-learn
 - Support Vector Machine
 - K-Nearest Neighbors
 - Gradient Boosting Decision Trees Algorithms
 - Random Forest
 - Multi Layer Perceptron

4. 모델 구축 결과

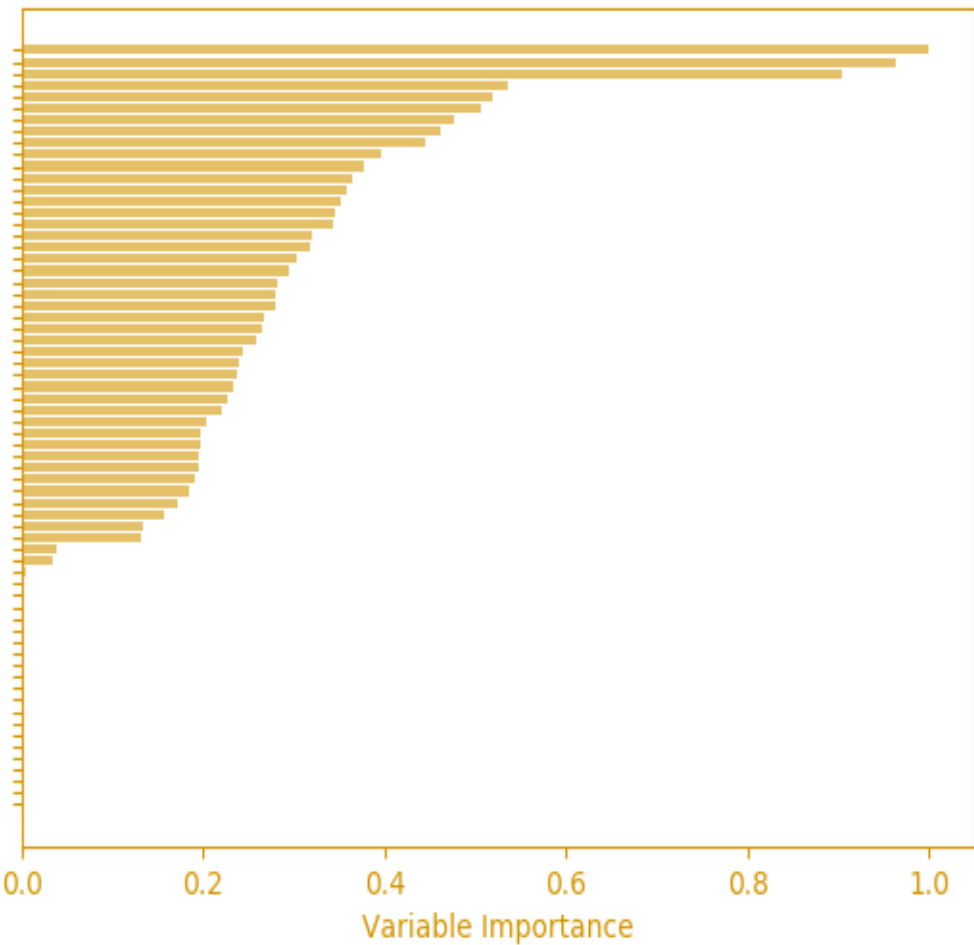
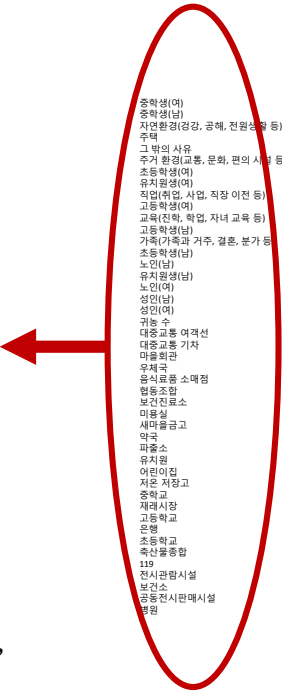


• 요인별 중요도 확인

- 인구정보
- 교육시설
- 교통시설
- 의료/편의/문화시설
- 기타

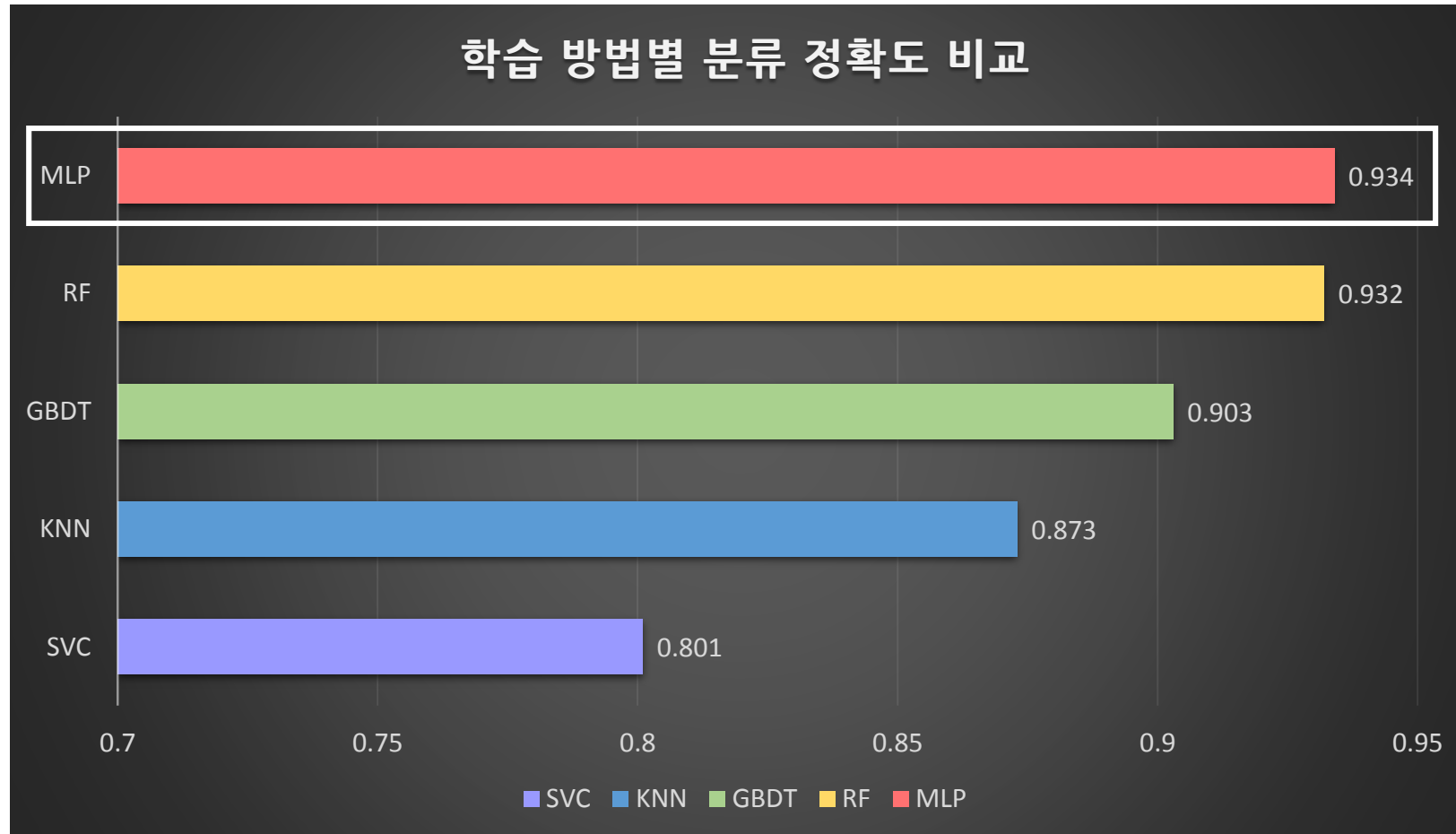
귀농 수
대중교통 여객선
대중교통 기차
마을회관
우체국
음식료품 소매점
협동조합
보건진료소
미용실
새마을금고
약국
파출소
유치원
어린이집
저온 저장고
중학교
재래시장
고등학교
은행
초등학교
축산물종합
119
전시관람시설
보건소
공동전시판매시설
병원

중학생(여)
중학생(남)
전입사유-자연환경(경강, 공해, 전원생활 등)
전입사유-주택
전입사유-그 밖의 사유
전입사유=주거 환경(교통, 문화, 편의 시설 등)
초등학생(여)
유치원생(여)
전입사유-직업(취업, 사업, 직장 이전 등)
고등학생(여)
전입사유-교육(진학, 학업, 자녀 교육 등)
고등학생(남)
전입사유-가족(가족과 거주, 결혼, 분가 등)
초등학생(남)
노인(남)
유치원생(남)
노인(여)
성인(남)
성인(여)



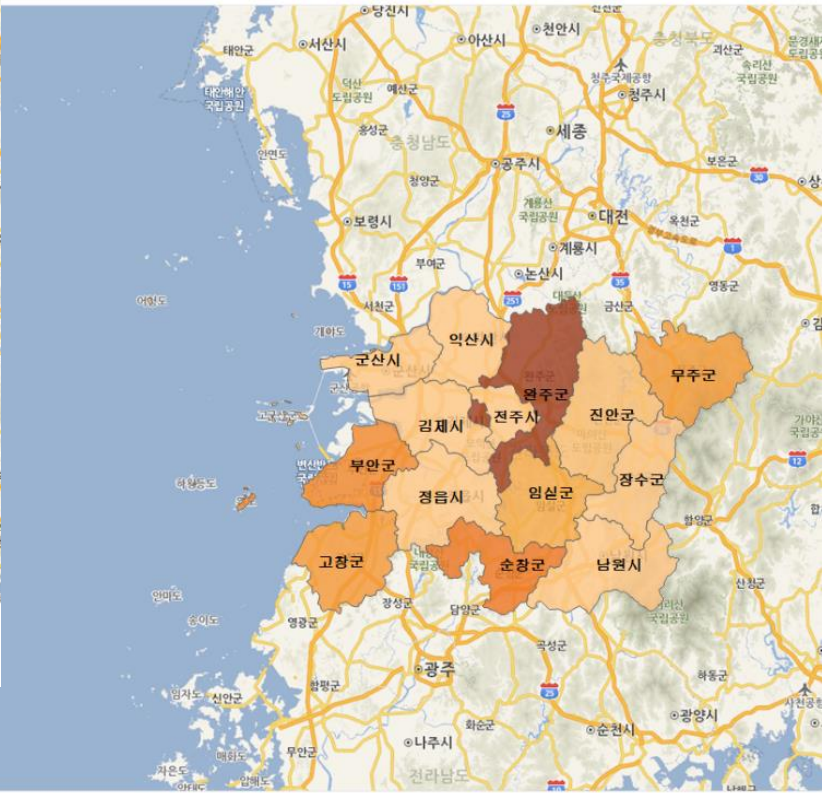
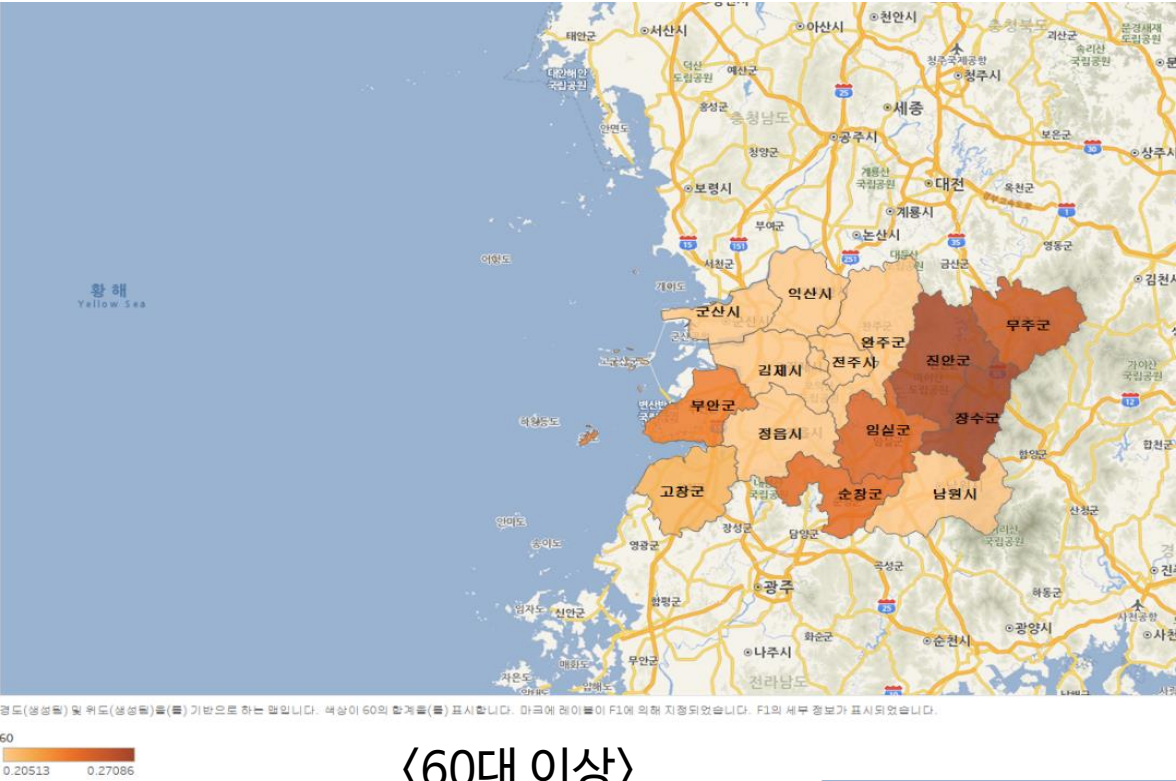


- 학습(Machine & Deep Learning) 방법별 분류 정확도 비교





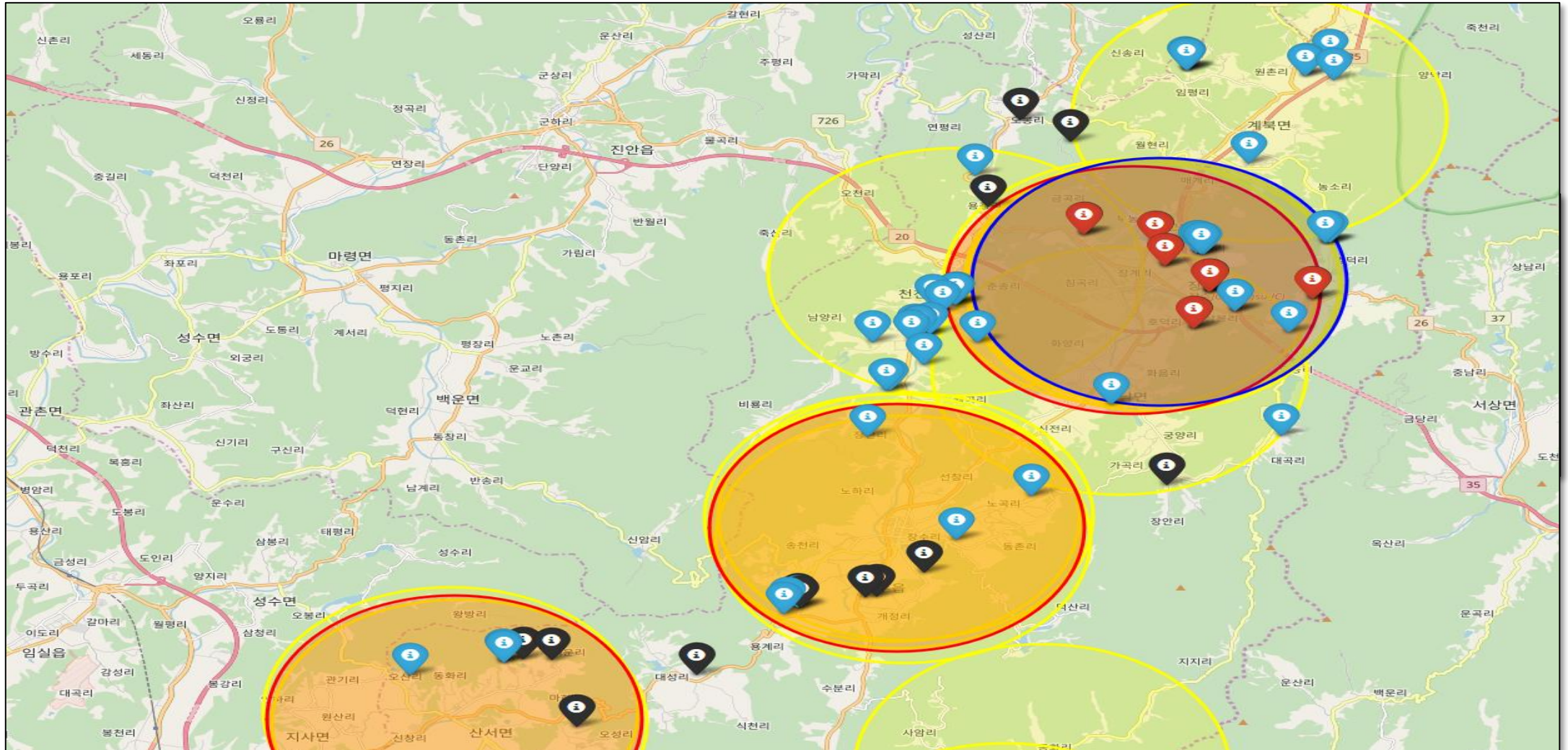
• 학습 결과(연령별 귀농 지역)



4. 모델 구축 결과



- 학습 결과 시각화 단계



5

시연



6

결론



- 기대효과

- 이용자 요구에 적합한 귀농 모델을 제안함으로써 귀농 실패 확률 감소
- 고려할 조건이 많아 귀농이 어려운 20~30대에게 문화생활 등을 고려한 귀농 적합지를 추천함으로써 농촌 내 젊은 귀농 인구의 유입을 유도할 수 있음



- ✓ 청년실업 해소에 기여
- ✓ 지역 내 경제활동 인구 증가
- ✓ 농가인구의 고령화 추세 감소
- ✓ 로컬푸드 생산 증가로 안전한 농식품에 대한 지역민의 만족도 향상

• 전라북도 적용 시 기대효과

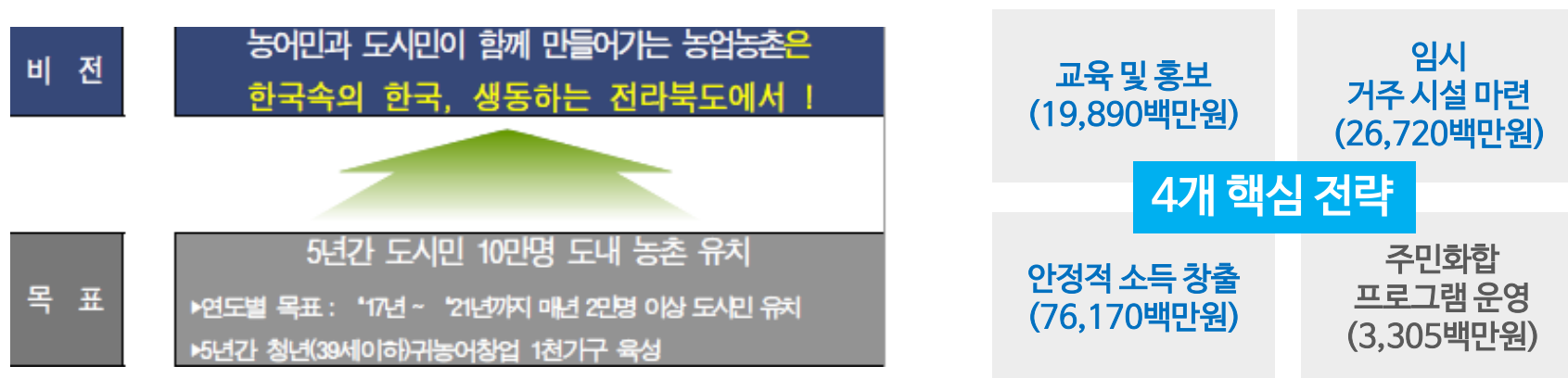
✓ 전라북도에 위치한 농촌진흥청 및 산하 연구기관과의 협조를 통해 공공기관과의 상생 구조 구축

귀농귀촌 적지를 추천함에 있어 기존 데이터와 딥러닝을 통해 적지선정의 편의를 제공하고자 하는 노력은 필요하고 시의적절하다고 생각합니다. 그러한 맥락에서 실제 이 프로젝트가 현실화되었을 때, 1)현재 농촌진흥청에서 관리하고 있는 정보들(휴토람 등)을 연계하여 활용할 수 있고, 2)농업관련 신기술들을 소개할 수 있다면, 그리고 이런 것들을 귀농인들이 받아들일 수 있다면 서로가 윈윈할 수 있는 구조가 될 수 있을 것이라고 생각합니다.

농촌진흥청(지식정보화담당관실 최윤지 연구관)

2018.08.27 서면인터뷰 내용 일부 발췌

✓ 전라북도에서 계획하고 있는 귀농귀촌 종합계획(2017~2021) 지원 가능



- 확장 가능성

타 시도군으로 확장

전국 인구데이터 활용

귀촌, 귀어 모델로 확장

환경요인 조정
어류(양식류) 데이터 활용

수확행

맞춤형 귀농 전략 추천 모델

[웹페이지]

공공데이터 포털. <https://www.data.go.kr/>

귀농귀촌종합센터. <http://www.returnfarm.com/>

농민신문 (2018.06.18). 귀농·귀촌, 실패하는 유형이 있다. <https://www.nongmin.com/>

연합뉴스 (2017.02.04). 어설프게 덤뻐다간 낭패...귀농인 10명 중 1명 짐 싣다. <http://www.yonhapnews.co.kr/>

중앙선데이 (2016.09.25). 소득원 못 찾고 원주민과 갈등 준비 안 된 귀농인 ‘다시 도시로’. <https://news.join.com/>

중앙일보 (2016.11.21). 농업의 중요성과 과제. <https://news.join.com/>

통계청. <http://kostat.go.kr/>

한국지역정보개발원. <https://www.klid.or.kr/>

[간행물]

김정섭, 이정해 (2017). 최근 귀농·귀촌 실태와 시사점. 한국농촌경제연구원 농정포커스, 151, 1-20.

농림축산식품부 (2017). 귀농·귀촌 종합계획 수립 방향 연구. 세종: 농림축산식품부.

농림축산식품부 (2018). 2018년 농림축산식품부 업무계획. 세종: 농림축산식품부.

박대식 외 (2017). 귀농·귀촌이 농촌 지역사회에 미치는 사회경제적 영향 분석. 나주: 한국농촌경제연구원.

전라북도 (2017). 귀농, 귀촌 서비스 모델 빅데이터 분석 연구

통계청 (2017). 2017년 귀농어·귀촌인통계. 대전: 통계청.

서울대학교 (2016). 빅데이터를 이용한 귀농수요 분석과 귀농인의 성공적 모델 연구

송호, 박성준 (2016). 전북 농업의 고부가가치화 방안. 전주: 한국은행 전북본부.

감사합니다.

