

[1] 제안발표 요약

• 산업 배경

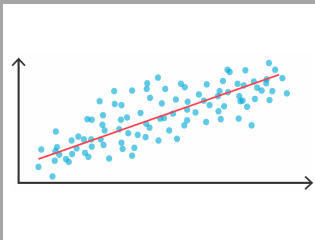


지방대학의 중도탈락 심화

중도탈락은 사회적 비용 낭비

대학의 교육여건 하락

• 분석 방법



대학 알리미의 14가지 변수
데이터 수집

다중회귀모형을 통한 예측

위계적 회귀분석으로

- 1) 대학 조치 가능요인과
- 2) 대학 조치 불가능요인 분석

• 기대효과



비용분석을 통해 현실적인
해결방안 제시

도출된 요인을 사용하여
중도탈락율 개선

대학의 수익/지원 증대로 인한 교육의
질 향상

[1] 제안발표 요약_분석 변수 변경

<상관계수 히트맵 일부>

num_sdnt	1	-0.41	0.91
Grad_Job	-0.41	1	0.38
Professor	0.91	0.38	1

특정 변수간 높은 상관계수, **다중공선성** 확인

변수통합 및 추가

기존 변수 14개



최종 변수 11개

No.	변수명	단위
①	졸업생 취업률(Get_Job)	%
②	등록금(Tuition)	원
③	1인당 장학금(scholar)	원
④	기숙사 수용률(Dorm_Capa)	%
⑤	교수 1인당 논문 수(1prof_paper)	개
⑥	A학점 비율(A_Ratio)	%

No.	변수명	단위
⑦	연간 학교 교육 지원금(Edu_Fund)	원
⑧	교수 1인당 학생 수(1Prof_Student)	명
⑨	직원 1인당 학생 수(1Work_Stud)	명
⑩	신입생 총원률(Fresh_Ratio)	%
⑪	신입생 경쟁률(Fresh_Comp)	%
Y	중도 탈락율(Fail)	%

[1] 제안발표 요약_변경 사항 소개

#1. 분석 변수 변경(과정)

1. 변수 제거

⑨	학생 창업 지원액(sdnt_fund)
⑩	학생 창업 지원 인력(sdnt_man)
⑬	교지 확보 현황(Area)

결측치
많은
변수 3개

2. 기존 변수 병합

⑦	연간 학교 지원금(Edu_Fund)
⑧	교수 1인당 학생 수(1Prof_Student)
⑨	직원 1인당 학생 수(1Work_Stud)

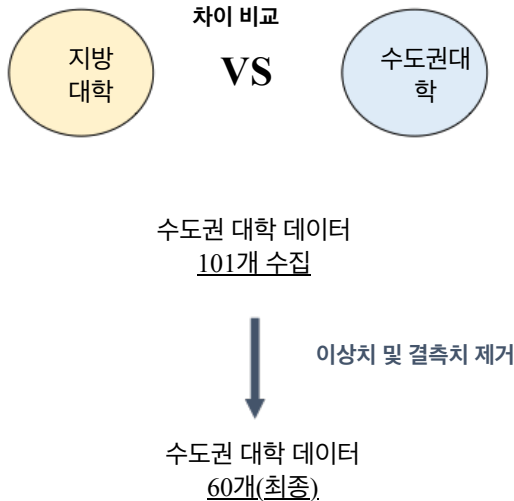
상관성 띄는 변
수 통합

3. 변수 추가

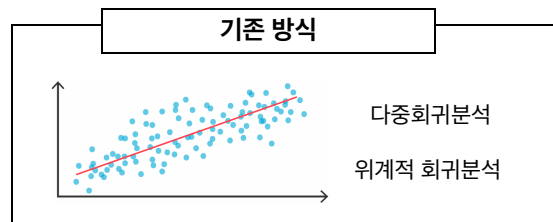
⑩	신입생 충원률(Fresh_Ratio)
⑪	신입생 경쟁률(Fresh_Comp)

새 변수
2개 추가

#2. 수도권 대학 데이터 수집

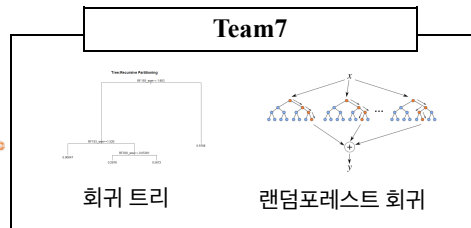


[1] 제안발표 요약_제안 발표와 차이점



차이점

분석방법



수도권 대학과 지방 대학을 함께 분석
대학 소재지 비수도권=0, 수도권=1

분석대상

수도권 대학 VS 지방 대학

학생 수 외 13개 (총 14개)

변수선정

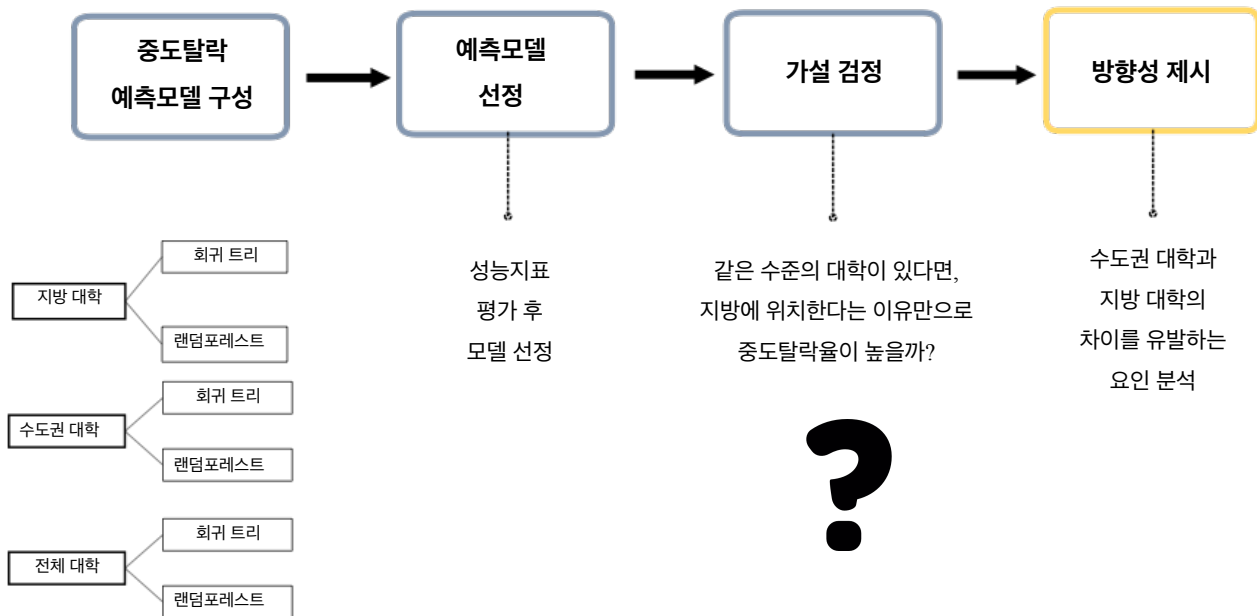
직원 1인당 학생 수 외 10개 (총 11개)

회귀계수를 이용한 비용분석식 정립

해결방안

변수 중요도, PDP를 이용한 요인분석

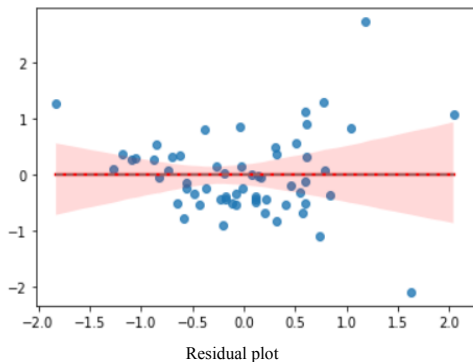
[2] 분석 과정_Summary



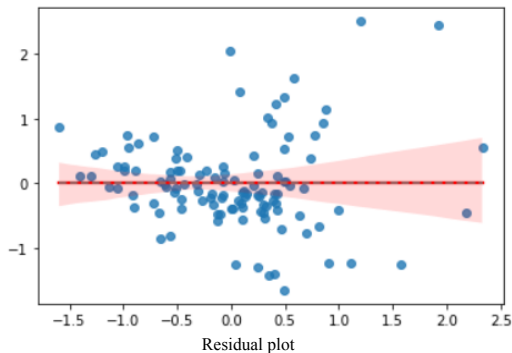
[2] 분석 결과_모델 선정 이유

#. 데이터의 선형성 확인

[수도권 대학 비선형성]



[지방 대학 비선형성]



선형 회귀 분석에
적합하지 않은
데이터

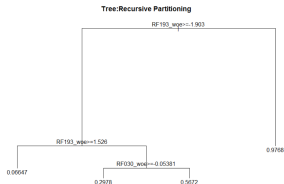
비선형 회귀 분석 모델 선정

의사결정나무 회귀
or
랜덤포레스트 회귀

[2] 분석 결과_모델 소개

Regression Tree

구간을 나누어, 각 leaf 노드에 대한
평균으로 회귀를 수행하는 모델



모델 구성 단계

1. Train data를 재귀이진분할하여 전체 트리를 만든다.

- **Greedy 방법** : 특정 변수에서 임계값을 찾는 최선해 도출 알고리즘
- **정지 기준** : Leaf node들의 RSS가 0이거나 임계값 이하.

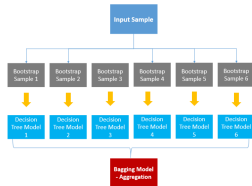
2. 트리 복잡도를 떨어뜨리기 위한 가지치기

3. 교차검증(CV)을 통한 최종 트리 확정

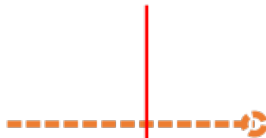
Random Forest Regressor

• 앙상블 학습

여러 개의 분류기 생성, 이
후 결합



• Bagging



Bootstrap : 복원 추출 샘플링

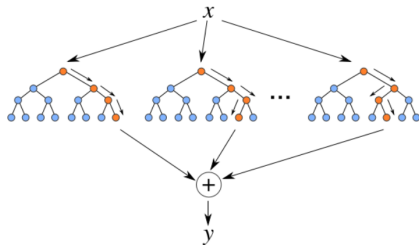
Aggregating : 집계

[2] 분석 결과_모델 소개

Random Forest Regressor

여러 개의 결정 트리의 결과를 바탕으로 예측하는 모델

- 분류 – Voting(다수결)
- 회귀 – 각 트리의 평균값



모델 구성 단계

1. 앞서 설명한 방식으로 회귀트리 생성
2. Bootstrap을 이용한, 변수 샘플링으로
지정한 수만큼 회귀트리 생성 반복
3. 다수의 트리의 예측값들의 평균을 최종값으로 선정

모델 성능 지표

MSE (평균 제곱 오차)	MAE (평균 절대 오차)
이상치에 영향을 많이 받음	이상치에 둔감, 강건함



이상치를 포함하였기에 선정

[2] 분석 결과_모델 선정

1. 하이퍼 파라미터 도출 모델

```
# 그리드검색 최적화
model = RandomForestRegressor()
param = {'criterion': ['mae'], 'n_estimators': [10, 20, 30, 40, 50, 60, 70, 80, 90],
        'max_depth': [4, 5, 6, 7, 8, 9, 10],
        'min_samples_split': [2, 3, 4, 5, 6, 7, 8]}
grid = GridSearchCV(model, param, grid=param, cv=10, n_jobs=-1)
grid.fit(X_train, y_train)
```

2. 도출된 하이퍼 파라미터

```
Best Param : {'criterion': 'mae', 'max_depth': 10, 'min_samples_split': 5, 'n_estimators': 40}
```

3. 최적값 이용하여 랜덤 포레스트

```
# 랜덤포레스트 모델 생성
trf = RandomForestRegressor(criterion='mae', n_estimators=40, max_depth=10, min_samples_split=5)
trf.fit(X_train, y_train)
```

조정한 파라미터

n_estimators	결정트리의 개수
max_depth	트리의 최대 깊이
min_samples_split	노드를 분할하기 위한 최소한의 샘플 데이터 수

모델 간 Train/Test MAE 비교

지방 대학

	의사결정나무	랜덤 포레스트
MAE_test	0.92	0.66
MAE_train	0.67	0.59

MAE

수도권 대학

	의사결정나무	랜덤 포레스트
MAE_test	0.97	0.84
MAE_train	0.72	0.71

MAE

전체 대학

	의사결정나무	랜덤 포레스트
MAE_test	0.88	0.57
MAE_train	0.53	0.52

MAE

[2] 분석 결과_가설 정의

같은 수준의 대학이 있다면,
단지 지방에 위치한다는 이유만으로
중도탈락율이 높을까?

?

궁금증에 대해
가설 정의

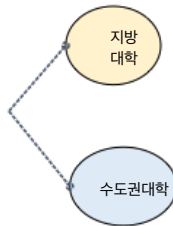


<가설>

동일 수준의 대학임에도 불구하고
지방에 위치한다면 중도탈락율이 높을 것이다.

<검정>

변수 값이
모두 동일한
하나의
대학 데이터



중도탈락율



중도탈락율



<검정>

지역 별 지역 거점 국립 대학 포함 2-3개씩 랜덤 추출하여 각각의 모델에 대입

지역	대학명	1work_std	Get_Job	Tuition	Scholar	1prof_std	1Prof_Paper	Edu_Fund	A_Ratio	Dorm_Capa	Fresh_Ratio	Fresh_Comp
강원	가톨릭관동대	47.4978166	56.6	7165674	3378575	828	0.4296	12292404	46.8	34.4	79.8	3.8
강원	강릉원주대학	18.0664962	51.7	4253234	2982597	644	0.5435	15727439.2	51.9	26.4	98	5.8
전북	군산대학교	28.6258065	59.4	3934276	2780279	636	0.5485	16372609.4	59.1	21.6	83.3	3.4

수도권 대학 결과

```
array([3.9, 3.9, 3.9, 8.4, 5.8, 3.9, 5.8, 8.4, 4.2, 2.8, 2.8, 3.9,
       3.9, 3.9, 5.8, 5.8, 8.4, 8.4, 2.8, 5.8, 4.2, 5.8, 3.9])
```



지방 대학 결과

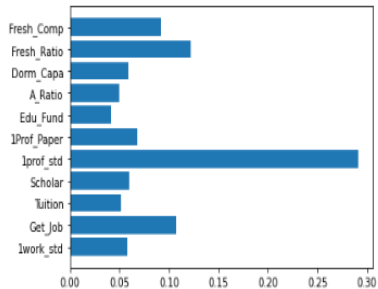
```
6.5765, 3.9375, 7.0995, 3.554, 8.071, 5.7295, 7.309,
6.6115, 10.3235, 2.662, 3.7335, 4.716, 5.7295, 5.6255,
5.0125, 6.072, 6.324, 2.7265, 13.203, 4.5935, 4.5205,
2.899, 8.228, 5.1275])
```



24개 중, 19개의 대학이
지방 대학 모델에서
중도탈락율↑

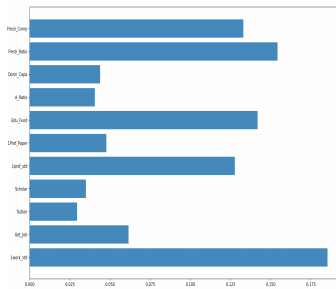
[3] 문제 해결 방안_변수 중요도

수도권 대학



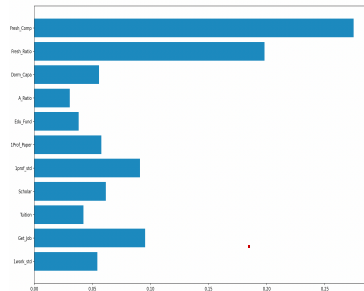
1	교수 1인당 학생 수
2	신입생 총원률
3	졸업생 취업률

지방 대학



1	직원 1인당 학생 수
2	신입생 경쟁률
3	신입생 총원률
4	교수 1인당 학생 수
5	연간 학교 교육 지원금

전체 대학



1	신입생 경쟁률
2	신입생 총원률
3	졸업생 취업률

※ 10% 넘게 관여하는 변수들 선정

[3] 문제 해결 방안_해결방안 도출

변수	중도탈락율과의 관계	해석	졸업과의 직접적인 연관성
교수 1인당 학생 수	교수의 수 ↑ 탈락율 ↓	다양한 교육의 폭은 학생의 졸업의지를 높임	O
신입생 총원률	총원율↑ 탈락율 ↓	총원율이 높은대학은 학생의 졸업의지를 높임	O
신입생 경쟁률	경쟁률↑ 탈락율 ↓	수요가 많은 대학입학은 학생의 졸업의지를 높임	O
졸업생 취업률	취업률↑ 탈락율 ↓	취업이 잘되는 대학은 학생은 졸업의지를 높임	O
직원 1인당 학생 수	직원↑ 탈락율 ↓	많은 직원수가 늘어나면 학생의 졸업의지를 높임	X, 재정
연간 학교 교육 지원금	지원금↑ 탈락율 ↓	높은 연간 학교 교육 지원금은 학생의 졸업 의지를 높임	X, 재정

수도권 대학

1	교수 1인당 학생 수
2	신입생 총원률
3	졸업생 취업률

→ 학생 졸업과 직접적인 문제

지방 대학

1	교수 1인당 학생 수
2	신입생 경쟁률
3	신입생 총원률
4	직원 1인당 학생 수
5	연간 학교 교육 지원금

→ 1. 학생 졸업과의 직접적인 문제

→ 2. 재정 문제

전체 대학

1	신입생 경쟁률
2	신입생 총원률
3	졸업생 취업률

→ 학생 졸업과의 직접적인 문제

앞선 검증 : 지방대학의
중도탈락율이 더 높다



지방대학은 수도권대학과 달리 **재정
문제**가 중도탈락율에 중요한 영향을
미침

[3] 문제 해결 방안

“위기의 지역대학, 새정부의 정책은 어떻게 해야 하는가?”

교수노조 | 승인 2022.05.02 10:30 | 댓글 0

지역대학 정책은 큰 기초가 “지역대학 경쟁력 제고를 위한 지자체 자기책임성 강화”에서 알 수 있듯이 지역대학은 알아서 생존해야 하는 상황

다. 대학 평가에 관한 정책으로 “정부 주도의 획일적 평가 중단 및 사업별 대학지원을 포괄적 대학지원으로 전환”이라는 표현이 있는데 이는 기존의 사업별 보조금 지원 방식에서 교부금 지원 방식으로 전환하는 것으로 볼 수도 있지만, 구체적 방안이 나와 있지 않다. 개방과 공유에 기반한 고등교육 생태계 조성 정책 역시 기존 교육

①

②

보조금 □ 지원금액의 사용목적 및 수준을 제한
교부금 □ 지원금액의 사용목적 제한하지 않음

③

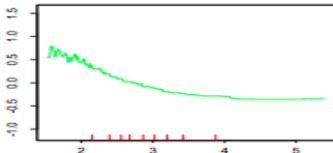
어떤 지역 대학에 얼마를 지원할 것인가?
□ 각 대학별 교부금의 사용 목적이 불분명

④

랜덤포레스트 **PDP** (Partial Dependence Plot) 분석을 통해
지방 대학의 교부금 사용 목적을 분명히 하고자 함

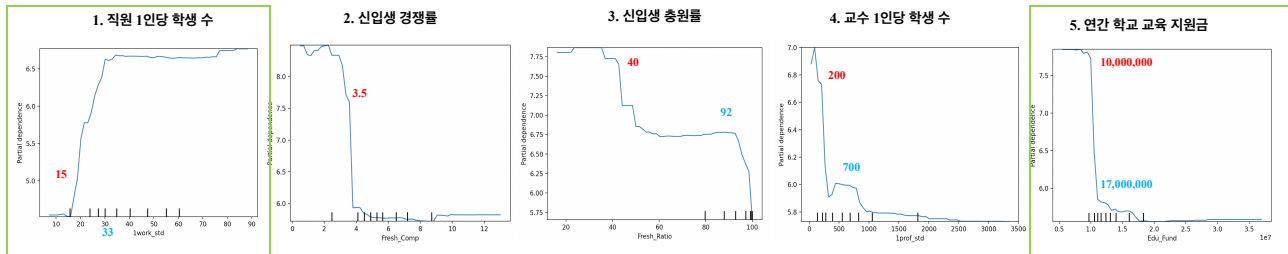
PDP : Partial Dependence Plot

□ 각 변수가 종속 변수에 미치는 영향 그래프



[3] 문제 해결 방안

지방 대학 PDP



변수명	분석 결과
직원 1인당 학생 수	15명 이상에서 급상승, 33명 이상에서 수렴
신입생 경쟁률	3.5 : 1 이상에서 급감, 9 : 1 이상에서 수렴
신입생 총원률	40% 이상에서 급감, 92% 이상에서 급감
교수 1인당 학생 수	200명 이상에서 급감, 700명 이상에서 급감
연간 학교 교육 지원금	1천만원 이상에서 급감, 1.7천만원 이상에서 수렴

지역 대학 교부금 우선 사용 목적과 기준을 정의

교부금 사용항목	사용 목적
직원 1인당 학생 수	직원 1인당 학생수가 33명 이하를 유지하도록 직원 수를 조절한다. (학생수는 조절 불가)
연간 학교 교육 지원금	1천만원 이상의 교육 관련 지원금 투자를 목표

□ 지역 대학들이 교부금 사용 목적을 분명히 하고 교부금 투자를 받는 효과 기대