

# 가계대출 연체율 예측모델을 통한 시도별 맞춤 정책 제안

# 목차



## 서론

- 연구 배경
- 분석 목적
- 분석 과정



## EDA

- 독립변수 선정
- 시차적용
- 데이터 구성



## 모델링

- 1차 모델링
- 2차 모델링



## 결론

- 최종 분석 결과
- 지역별 맞춤 정책 제안
- 한계점 및 의의



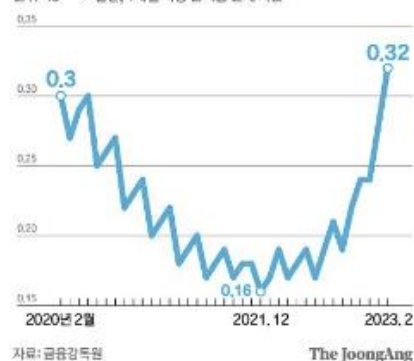
## 최근 경제동향\_가계부채연체율

금융불안지수 '위기' 지속... GDP 대비 가계·기업 빚 사상 최대  
가계대출 늘며 금융 불안... 취약차주 40%,  
빚이 연소득보다 많아

'빚의 역습' 시작됐다...가계대출 연체율, 코로나 이전 수준 돌파

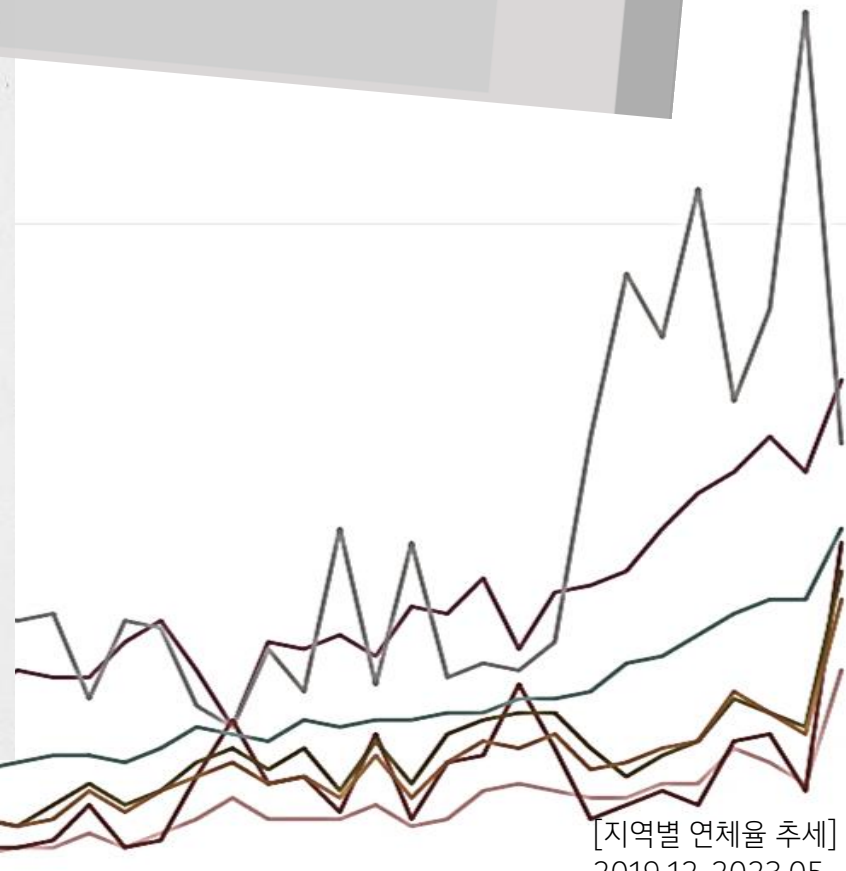
### 가계대출 연체율

단위: % ※ 월말, 1개월 이상 원리금 연체 기준



대출 연체율이 무섭게 치솟고 있다. 아직 절대 수치로는 낮은 수준이다. 하지만 고금리 기간이 길어지면서, 상승 폭을 키우고 있다는 점은 문제다. 특히 신종 코로나바이러스 감염증(코로나19) 때 빌린 돈의 연체율 상승이 하반기로 갈수록 가팔라질 수 있어, 금융사 부담이 커질 전망이다.

25일 금융감독원은 2월 말 국내은행의 원화 대출 연체율이 1월 말과 비교해 0.05%포인트 상승한 0.36%를 기록했다고 밝혔다. 지난해 2월 말 연체율(0.25%)과 비교하면 0.11%포인트 급등했다. 2020년 8월(0.38%) 이후 가장 높은 연체율이다. 다만 2월 신규 대출의 연체율은 0.09%로 전달과 비슷한 수준이었다. 금감원은 전체 은행 계정 원화 대출금 및 신탁 대출금에서 한 달 이상 원리금을 갚지 못한 비율로 연체율을 집계했다.



# 연구 배경

## 전국단위 정책



### 50년 만기 주택담보대출

기존 30. 40년 만기에서 50년 만기로 늘린 정책



### 금융위원회, 저금리 대한 프로그램

: 대출 한도와 대상을 늘리는 정책



### 연체감소 '상생금융' 프로그램

: 연체 이자를 납부하면 원금 줄여주는 프로그램

## 하지만, 현재 효과 미미

### 가계부채 낙달 연속 늘자...금융당국 책임 회피 공방 본격화

금감원장은 예정에 없던 가계대출 현황을 설명하는 기자회견담회를 열었다.

'상생금융' 행보가 가계대출 증가로 이어졌다는 비판을 의식하였기 때문이다.

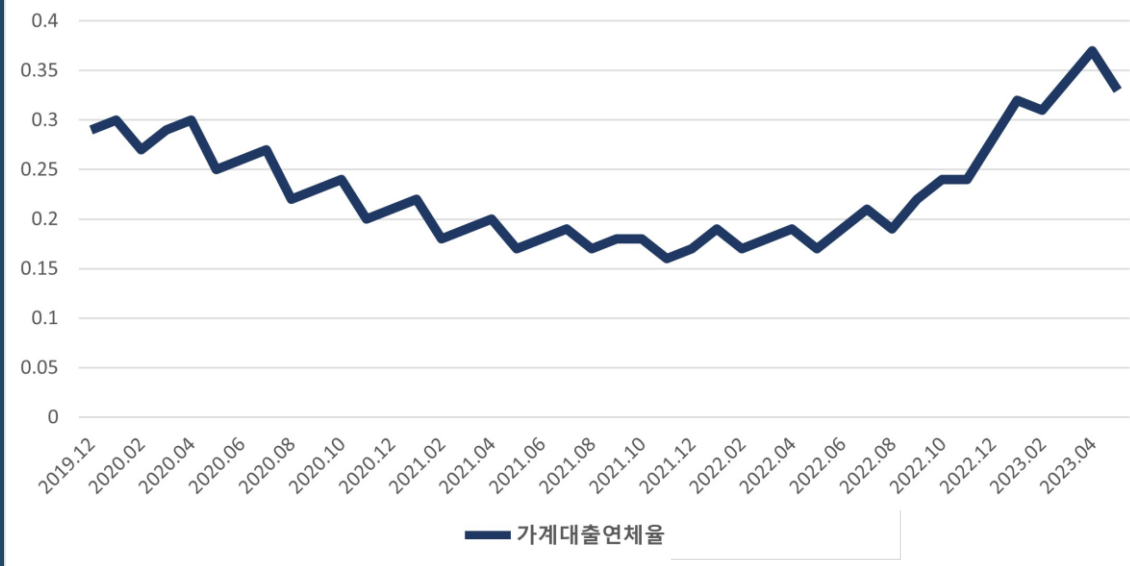
앞서 한은 금융통화위원회에서는 금감원의 '상생금융' 행보가 통화정책 효과를 **감아먹는다**는 비판이 나온 바 있다.

### 50년만기 주담대 사라지나...가계부채 증대 원인으로 지목

NH농협은행 50년 주담대 상품 판매 2달 만에 중단

가계대출 증가 원인으로 꼽혀

가계대출연체율추이



[금융감독원, 가계대출연체율추이]

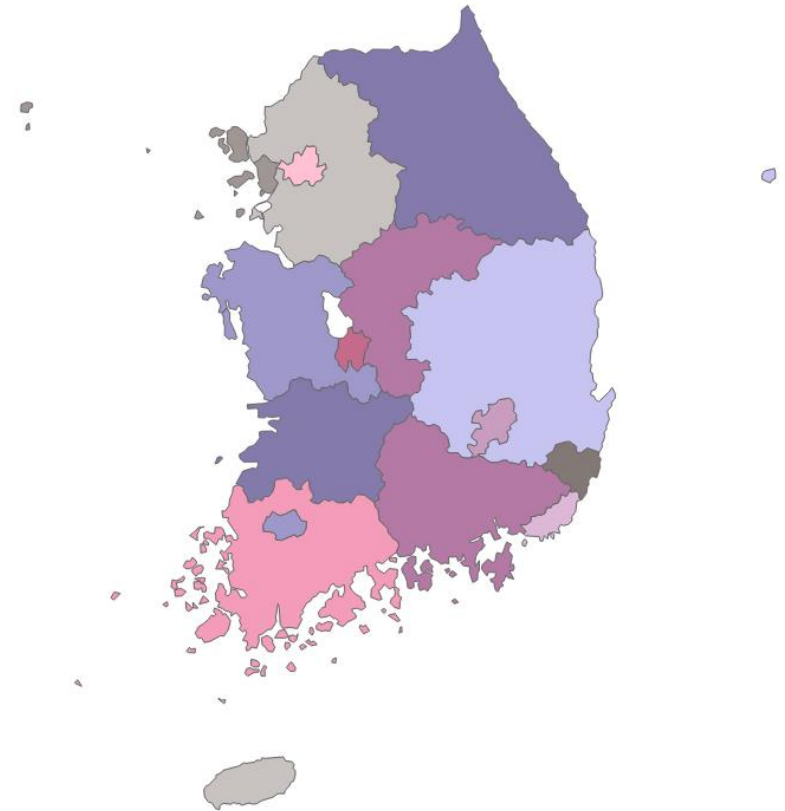
## "가계대출연체율 예측모델을 통한 시도별 맞춤 정책 제안"

### 분석 구체화

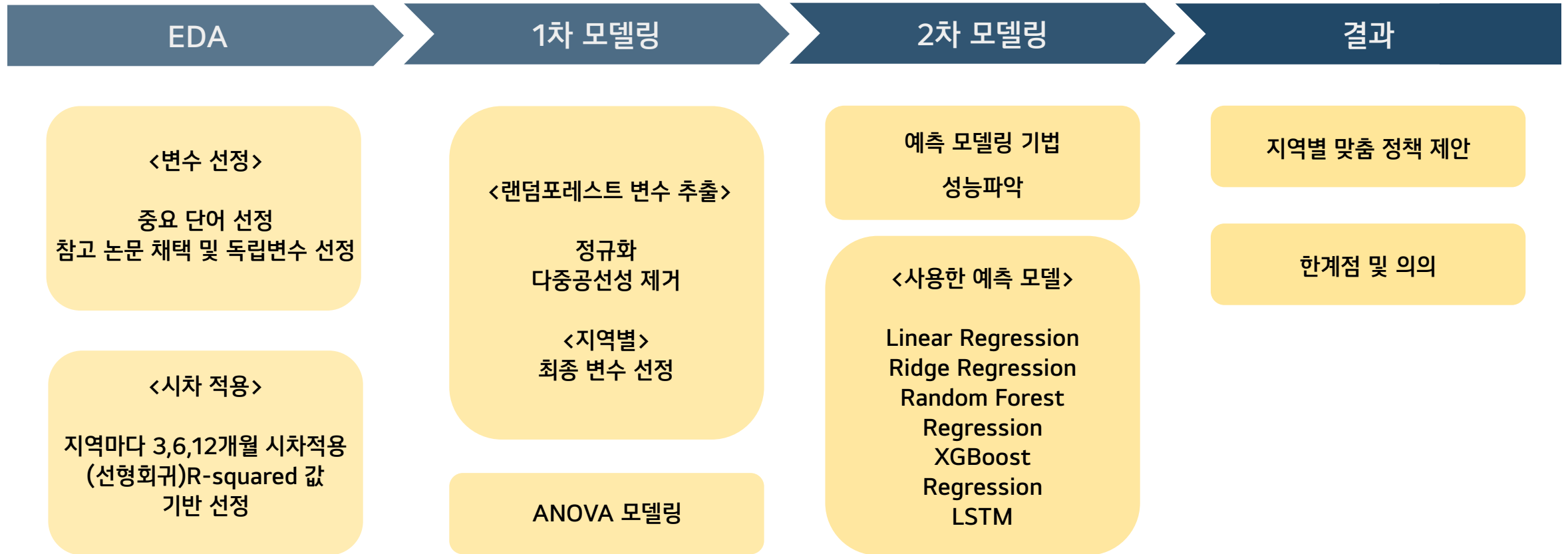
1. 지역별 분석을 통해 가계대출연체율의 증감 원인을 파악한다.
2. 독립변수가 가계대출연체율에 영향을 미치는 시점을 알 수 있다.

### 분석 목표

1. 가계대출연체율 사전적 인지에 도움
2. 연체율 관련 정책 도출까지 시간적 여유 제공
3. 지역별 맞춤 정책 가능



# 분석 과정

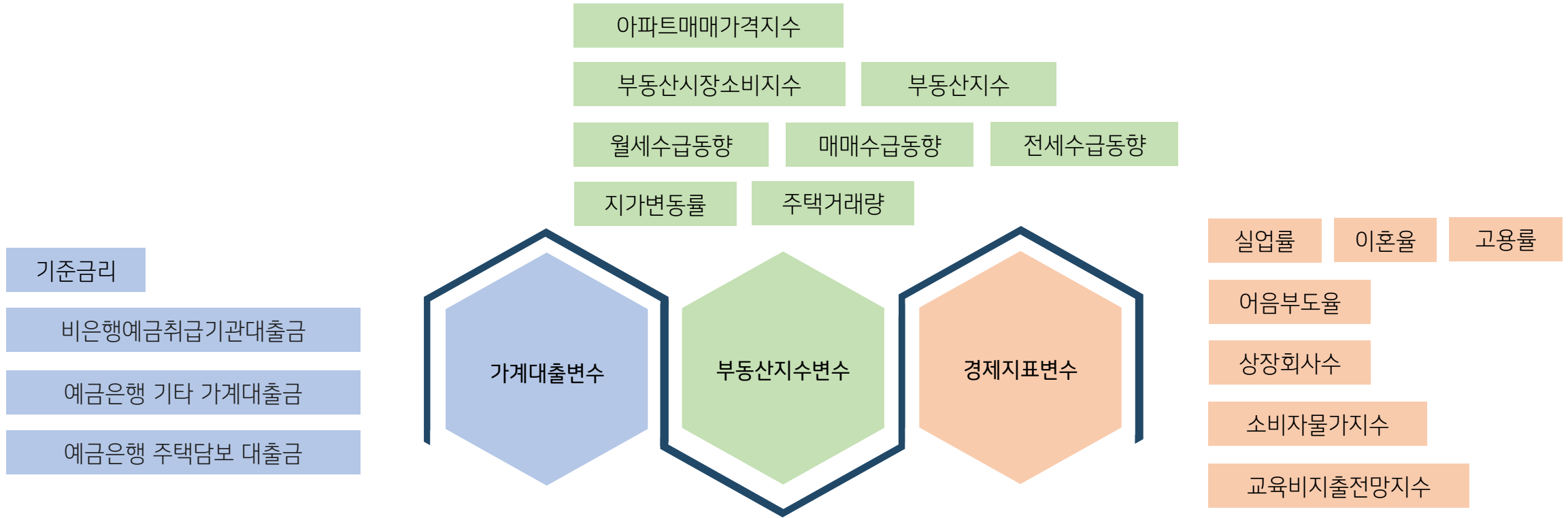








# 독립변수 선정

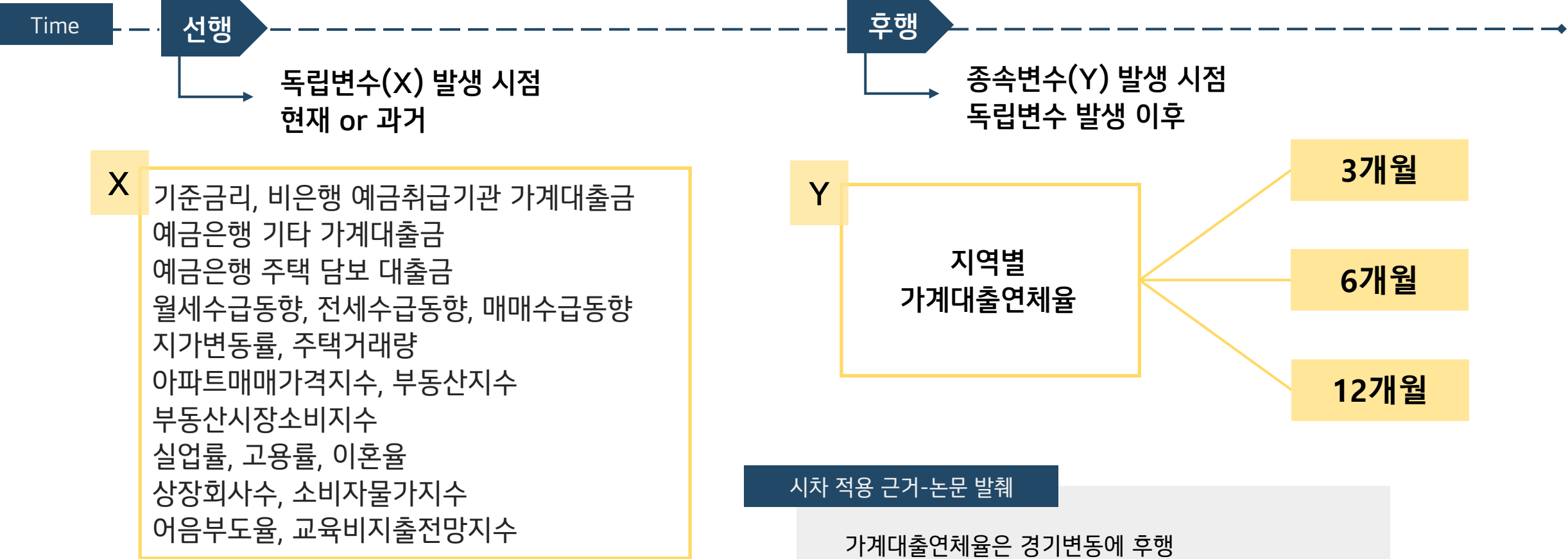


## 변수 선정 근거-논문 발췌

- 1) 가계대출 변수  
주택담보대출 연체율과 관련하여 가장 밀접하다고 판단되는 요소는 금리를 비롯한 가계부채와 관련된 요소들이다.
- 2) 부동산지수 변수  
부동산과 관련한 여러 지수들은 현재 부동산시장의 상황을 측정할 수 있는 요소들이다.
- 3) 경제지표 변수  
거시경제지표는 국내 경기의 수준을 보여주고 있어 경기변동이 주택담보대출 연체율에 영향을 주는 것은 당연하다고 판단된다.

# 시차 적용

독립변수가 가계대출연체율에 영향을 미치는 시점을 알 수 있다.

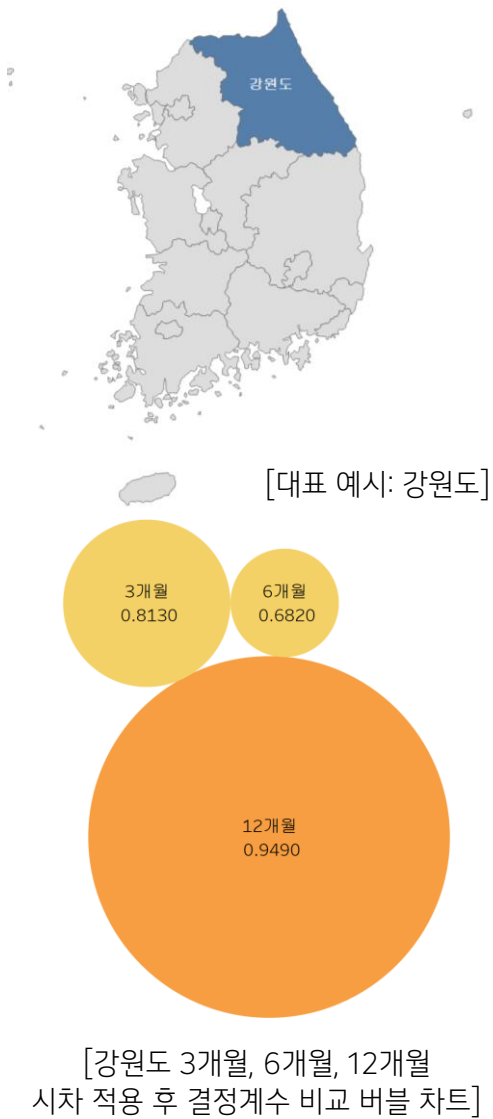


## 시차 적용 근거-논문 발췌

가계대출연체율은 경기변동에 후행  
즉, 경기악화->실업증가->가계파산->연체증가  
등의 경로를 거쳐 나타날 가능성이 크다.

[전홍배, 이정진, 최운열, 2008.06.03., '스트레스테스트를 활용한은행 가계대출부문의 안정성 연구',  
한국은행 금융안정분석국 리스크분석팀]

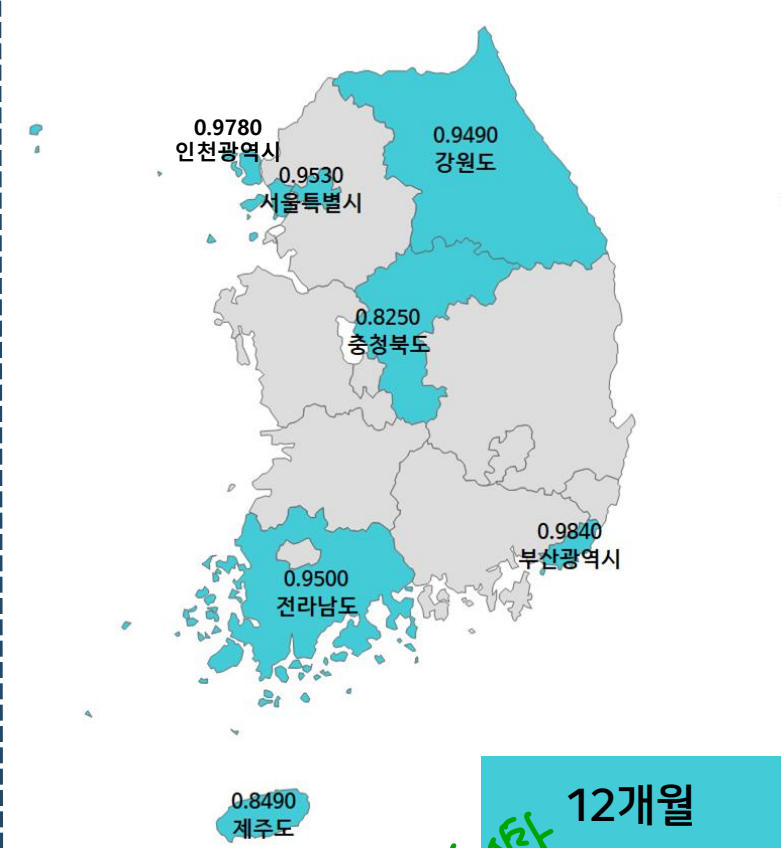
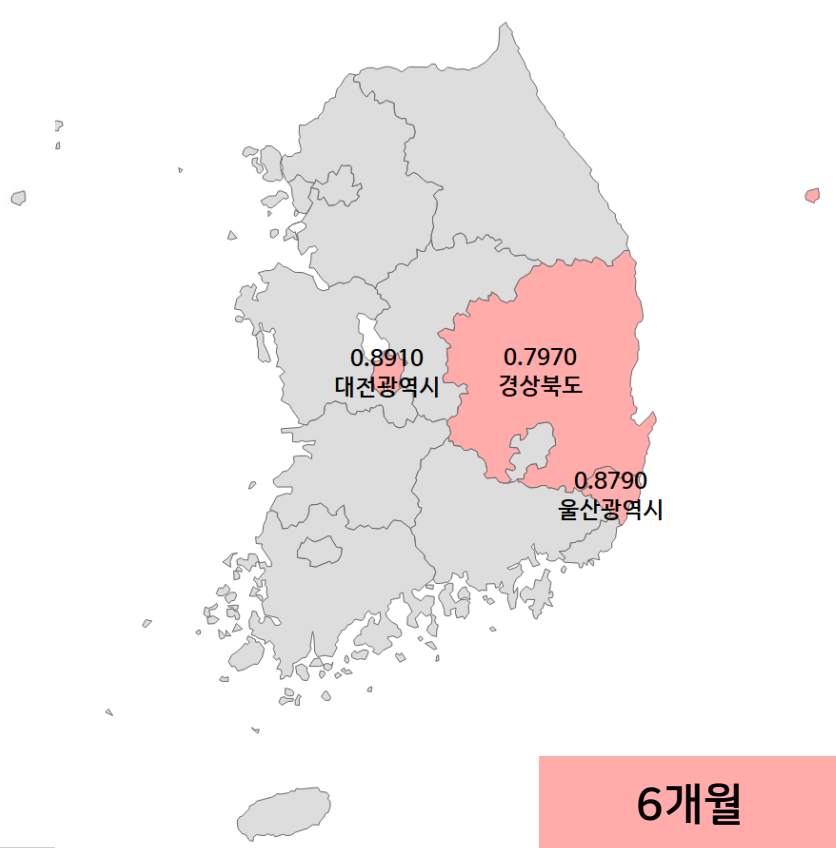
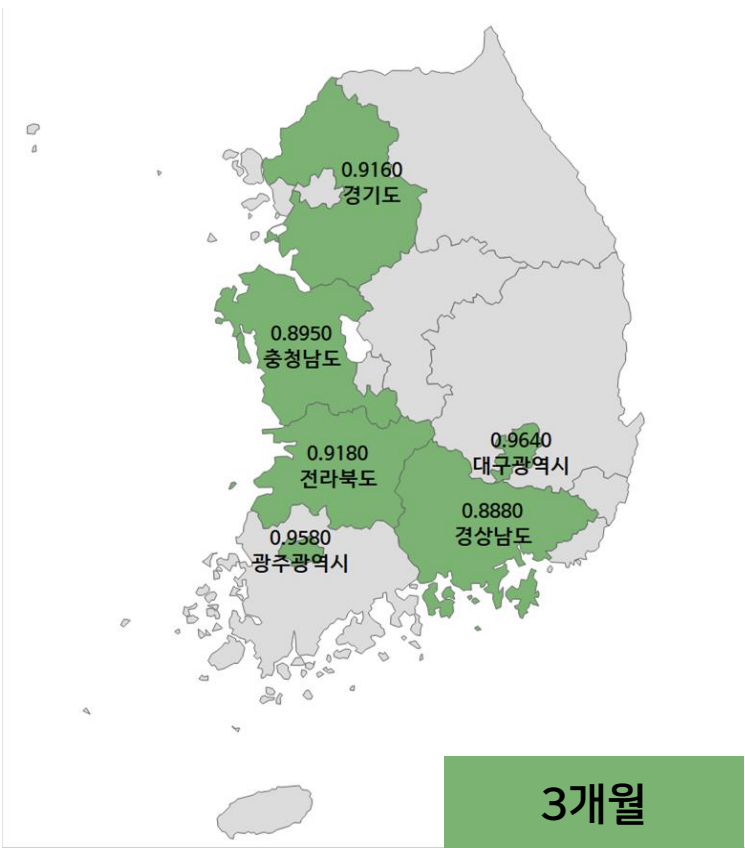
# 시차 적용



종속변수인 가계대출연체율을 3/6/12개월 미루며 시차 적용.  
3개월, 6개월, 12개월 미룬 데이터를 가지고 변수 각각을 상관관계 & 회귀분석 하여 결정계수를 살펴봄.

시차	요인별 상관계수
3개월 결정계수: 0.813	<div><div>지가변동률0.54743</div><div>월세수급동향0.59022</div><div>0.520.530.540.550.560.570.580.590.6</div></div>
6개월 결정계수: 0.682	<div><div>기준금리0.5125</div><div>00.10.20.30.40.50.6</div></div>
12개월 결정계수: 0.940	<div><div>통화량0.7659</div><div>전세수급동향0.8087</div><div>예금은행주택담보대출금0.9071</div><div>아파트매매가격지수0.8058</div><div>소비자물가지수0.8922</div><div>비은행예금취급기관가계대출금0.8139</div><div>부동산지수0.7986</div><div>기준금리0.7301</div><div>00.20.40.60.81</div></div>

# 시차 적용 결과



- 1) 종속변수인 가계대출연체율을 지역별로 3,6,12개월 미룸
- 2) 회귀분석 후 결정계수가 가장 높은 개월로 해당 지역의 시차를 결정
- 3) 결정계수의 수치가 시차 별로 비슷한 경우: 변수 별로 상관관계를 고려 ✓

x-y 간의 관계가 없는 시차  
→ 보편의 미/ X

# 데이터 구성

[독립변수(총 19개) : 예시는 2019.12~2020.02 강원도 수치]

가계대출 변수

기준 금리	비은행 예금취급 기관 가계대출금	예금은행 기타 가계대출금	예금은행 주택담보 대출금
1.25	16498.4	4308.3	5755.5
1.25	16500.9	4289.6	5807
1.25	16550.4	4307.6	5817.8

경제지표 변수

실업률	전월 대비 이혼률	고용률	상장회사수	어음부도율	소비자물가수	교육비지출 전망지수
3.3	48.8	60.4	0	0.03	99.81	100
5	-10.4	59.2	0	0.04	100.39	104
2.3	-2.8	60.7	0	0.01	100.37	101

부동산 변수

지가변동률	주택거래량	월세수급동향	매매수급동향	전세수급동향	아파트매 가격지수	부동산지수	부동산시장 소비지수
0.182	2719	66.07	73.47	81.38	86.6	95.1	125.6
0.174	2318	70.65	78.69	81.63	86.7	95	117.1
0.171	2765	73.02	97	77.81	87.3	95	88.3

[종속변수  
: 시차를 미룬 가계대출 연체율]



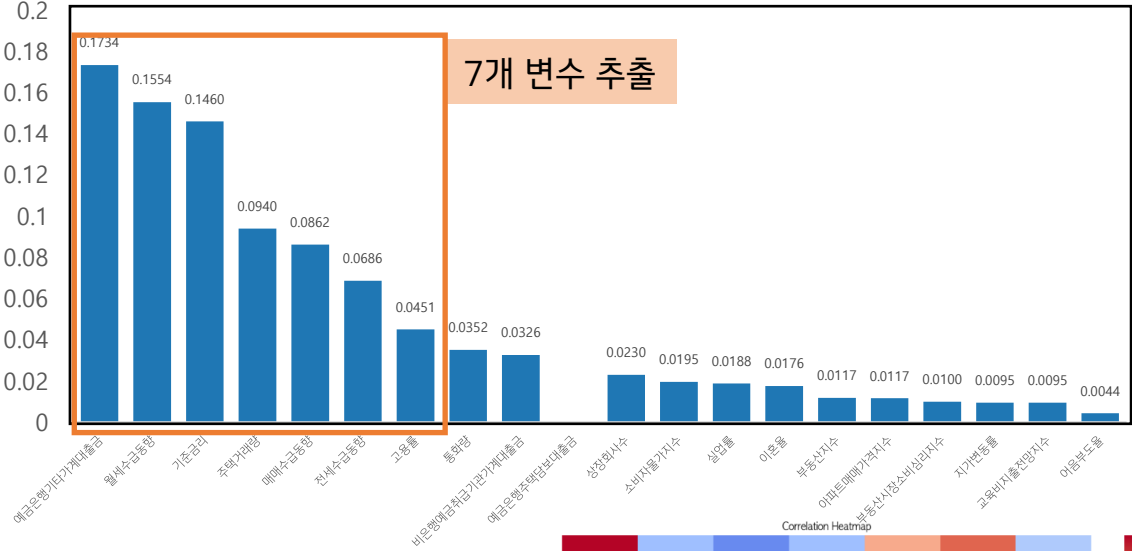
# 모델링 #1

✓ 다양한 데이터셋  
4. 다양한 모델링 방법론의 비교

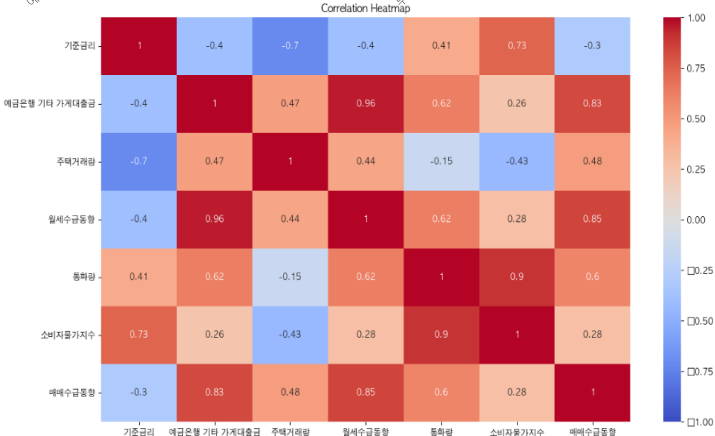
# Random Forest 변수 중요도

[Ex. 경상북도, 시차: 6개월]

## 1 Random forest 회귀 모델링을 통해 독립변수 중요도 추출



독립변수의 상관계수 히트맵



## 2 채택된 독립변수 정규화(Z-score)

기준금리	예금은행 기타 가계대출금	주택 거래량	월세수급동향	통화량	소비자 물가지수	매매수급동향
1.75	5672.7	2498	59.37	2813933	99.992	58.97
1.5	5718.8	3147	61.21	2807362	99.628	60.03



기준금리	예금은행 기타 가계대출금	주택 거래량	월세수급동향	통화량	소비자 물가지수	매매수급동향
0.848235	-1.780889	-1.187369	-1.931262	-1.543938	-0.82708	-1.961954
0.498962	-1.676707	-0.626066	-1.816237	-1.564459	-0.92603	-1.90288

**Z- score 정규화:**  $z = (x - \mu) / \sigma \sim N(0, 1)$   $\sigma =$  표준편차,  $\mu =$  평균.

z-score는 어떤 데이터 포인트가 평균으로부터 몇 표준편차만큼 떨어져 있는지를 표준화된 값으로 나타내어, 데이터의 상대적 위치나 비교 분석에 활용됩니다.



# Random Forest 변수 중요도

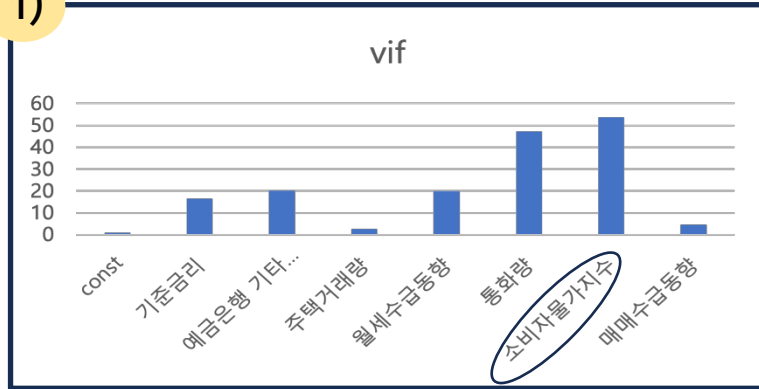
[Ex. 경상북도, 시차: 6개월]

$VIF : 1/(1-R^2)$ .  $R^2$  = 결정계수

: 다중공선성을 평가하는 통계적 지표로서, 회귀 분석에서 독립 변수 간의 상관 관계로 인해 발생할 수 있는 다중공선성을 측정하는데 사용됩니다.

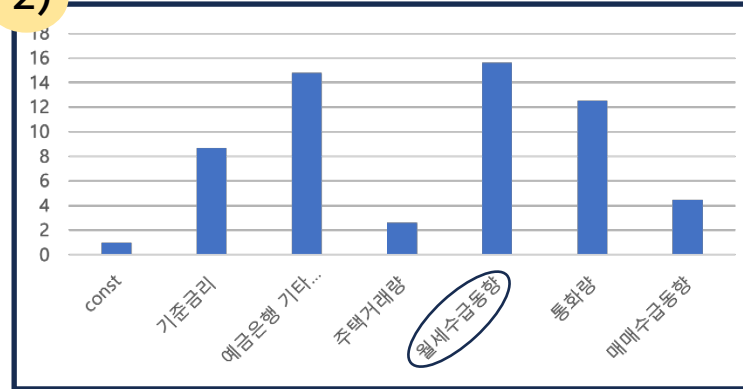
## 3 다중공선성 제거( $VIF \geq 10$ )

1)



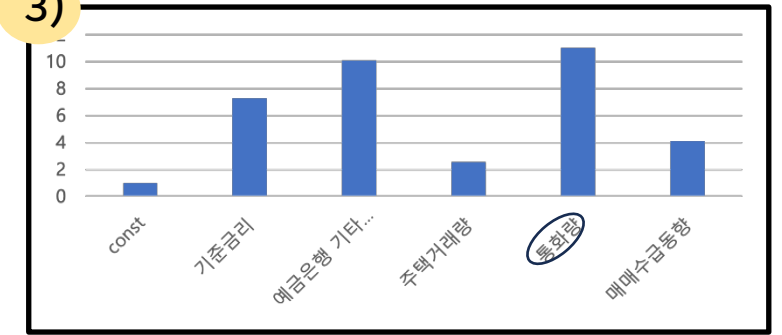
소비자 물가지수 변수 제거

2)



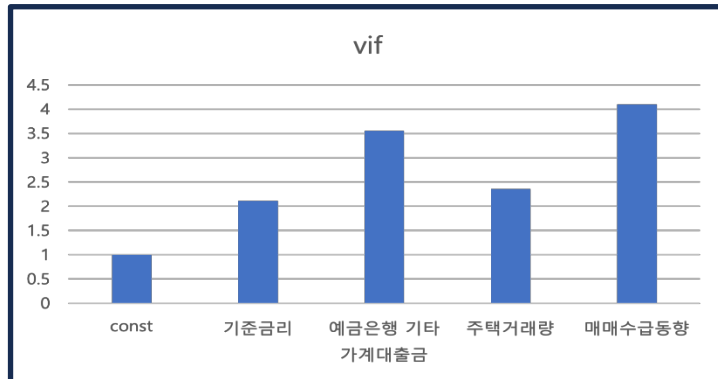
월세수급동향 변수 제거

3)



통화량 변수 제거

## 4 최종 독립변수 선정



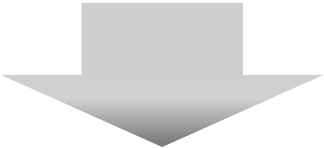
기준금리	예금은행 기타 가계대출금	주택거래량	매매수급동향
0.848235	-1.780889	-1.187369	-1.961954
0.498962	-1.676707	-0.626066	-1.90288
0.498962	-1.56891	-1.142396	-1.616423
0.498962	-1.556255	-1.132017	-1.572395

# Random Forest 변수 중요도

5

시차별/지역별 최종 데이터 프레임 구성

날짜	실업률	고용률	어음부도율	지가변동률	상장회사수	비은행예금 취급기관 가계대출금	주택거래량	예금은행 기타 가계 대출금	월세수급동 향	매매수급동 향	전세수급동 향	예금은행 주택담보 대출금	아파트매매 가격지수	소비자물가 지수	이혼율	교육비지출 전망지수	부동산지수	부동산시장 소비심리지수	통화량	기준 금리	연체율
----	-----	-----	-------	-------	-------	------------------------	-------	----------------------	------------	------------	------------	---------------------	---------------	-------------	-----	---------------	-------	-----------------	-----	----------	-----

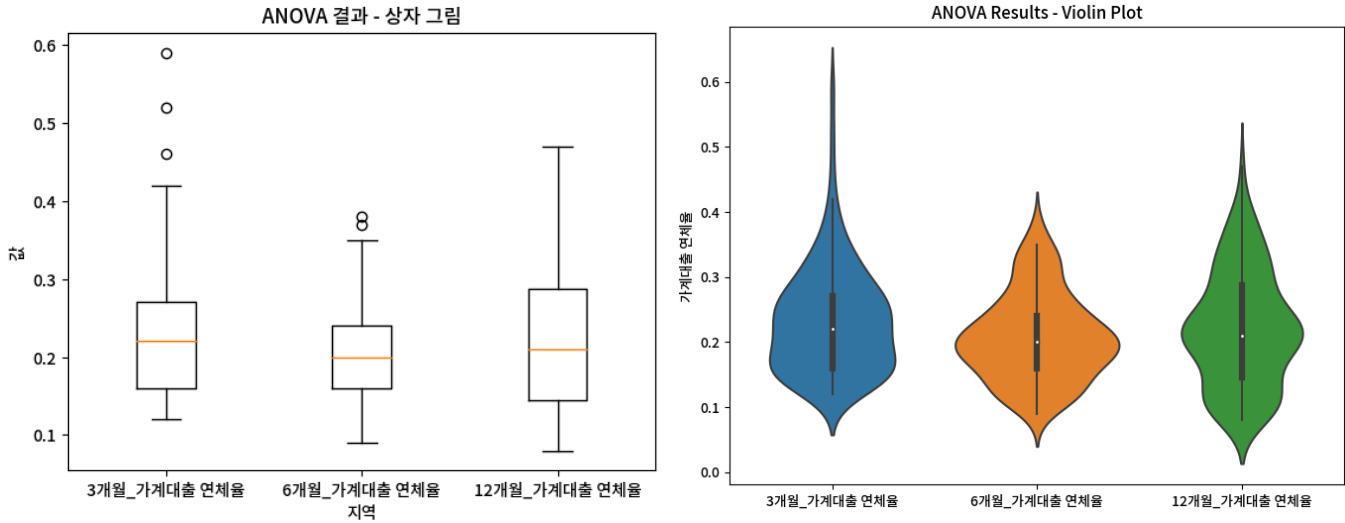


예시지역: 전라북도 변수 개수: 5개	지가변동률	예금은행 기타 가계대출금액	주택거래량	비은행예금취급기관_대출금액	예금은행 주택담보 대출금액	3개월
	0.622192	-1.810829	-1.187369	-1.165221	-1.938821	
	0.909182	-1.423192	-0.626066	-1.137273	-1.890192	
예시지역: 경상북도 변수 개수: 4개	기준금리	예금은행 기타 가계대출금액	주택거래량	매매수급동향		6개월
	0.848235	-1.780889	-1.187369	-1.961954		
	0.498962	-1.676707	-0.626066	-1.90288		
예시지역: 서울특별시 변수 개수: 4개	기준금리	주택거래량	월세수급동향	소비자물가지수		12개월
	0.962485	-0.875874	-1.032042	-1.167259		
	1.472462	-1.075059	-1.396835	-1.143337		

# ANOVA 모델링

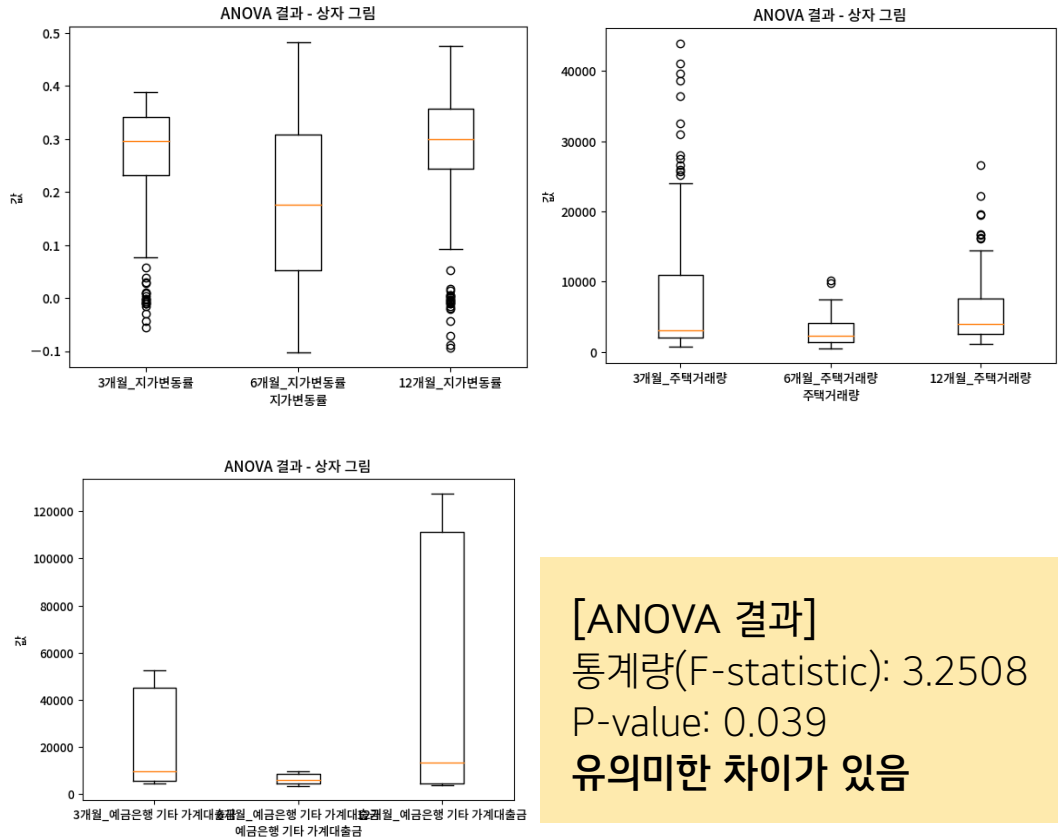
시차별 대표지역 선정  
> 3개월: 경기, 광주, 대구 / 6개월: 대전, 경북, 울산 / 12개월: 서울, 전남, 부산

## 1차 시차/ 지역별 연체율(Y) 변수의 차이가 유의미함을 확인



[ANOVA 결과]  
통계량(F-statistic): 3.2508  
P-value: 0.0398  
유의미한 차이가 있음

## 2차 독립변수간의 유의미한 차이가 있음을 확인



[ANOVA 결과]  
통계량(F-statistic): 3.2508  
P-value: 0.039  
유의미한 차이가 있음



## 2차 모델링 기법

Linear  
회귀모델

Ridge  
회귀모델

Random  
Forest

XGBoost

LSTM

R-squared

모델의 설명력을 평가하는 지표로 1에 가까울수록 설명력이 좋다.

MSE

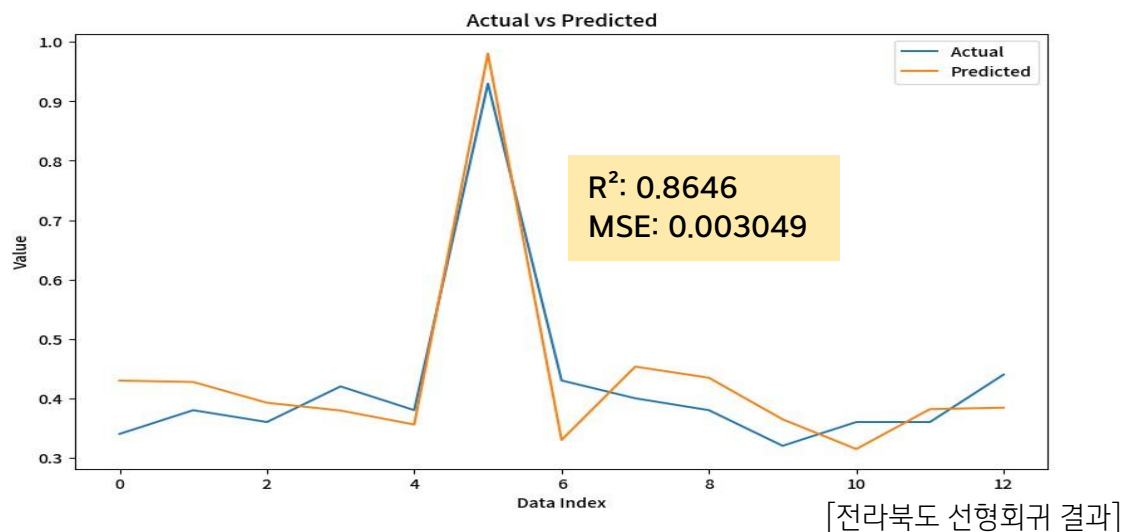
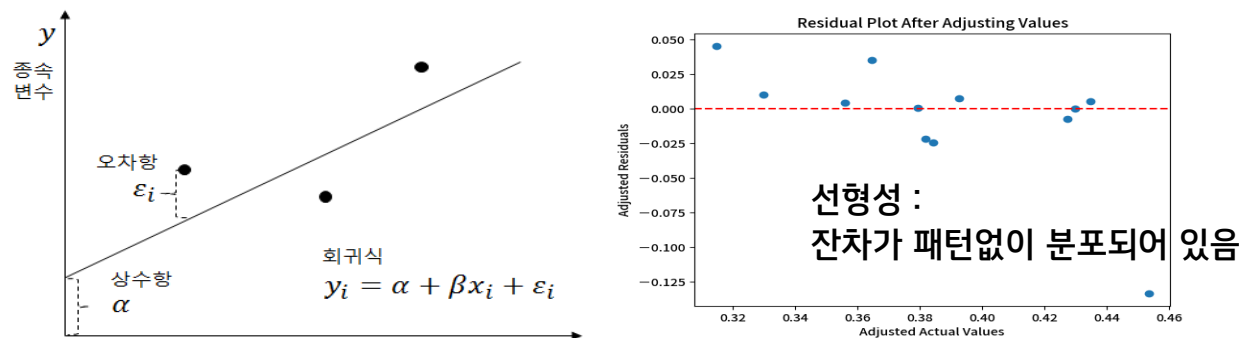
예측값과 실제값의 차이를 나타내는 지표로 작을수록 예측력이 좋다.

# Linear&Ridge Regression

[Ex. 전라북도, 시차: 3개월]

## Linear Regression(선형회귀)

- 학습과 예측이 빠름
- 데이터 개수에 비해 특성이 많을 때 학습 용이

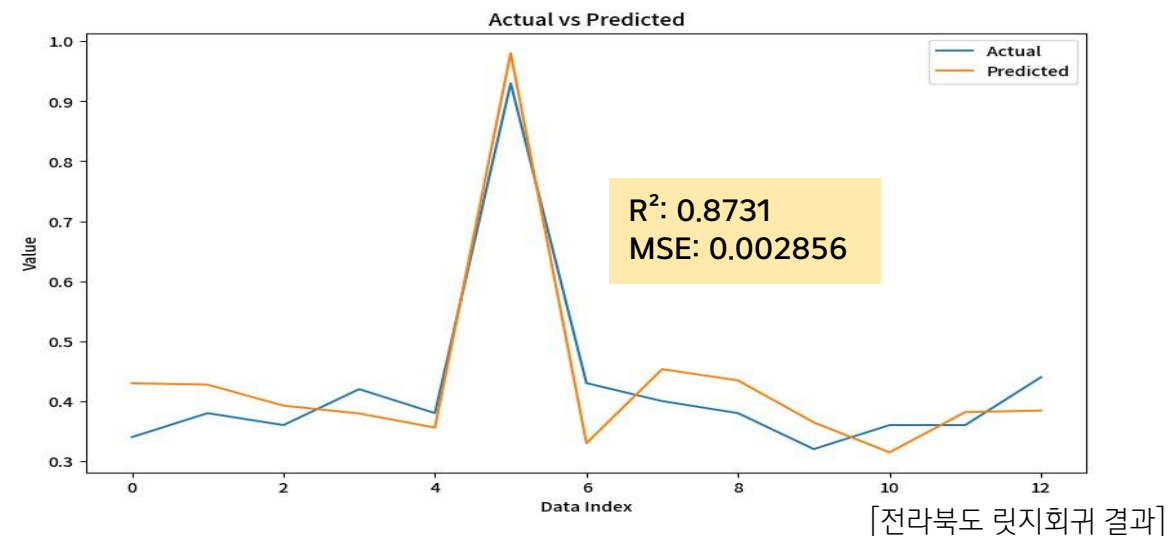


## Ridge Regression(릿지회귀)

- 선형회귀 모델의 과적합을 방지함
- 모델을 일반화 하기 용이

$$J(\theta) = \text{MSE}(\theta) + \alpha \frac{1}{2} \sum_{i=1}^n \theta_i^2$$

[릿지회귀 수식]

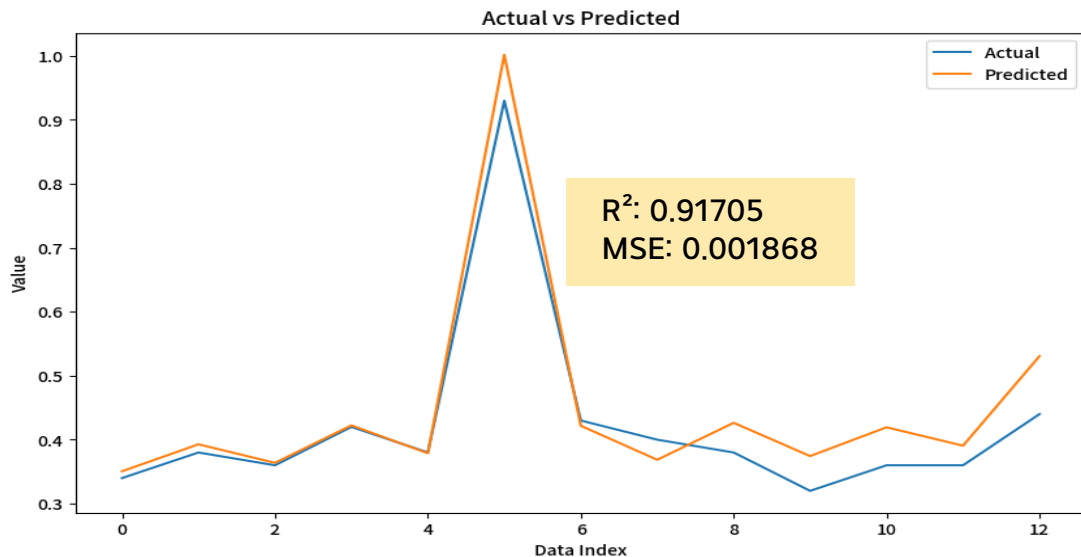


# Random Forest Regression

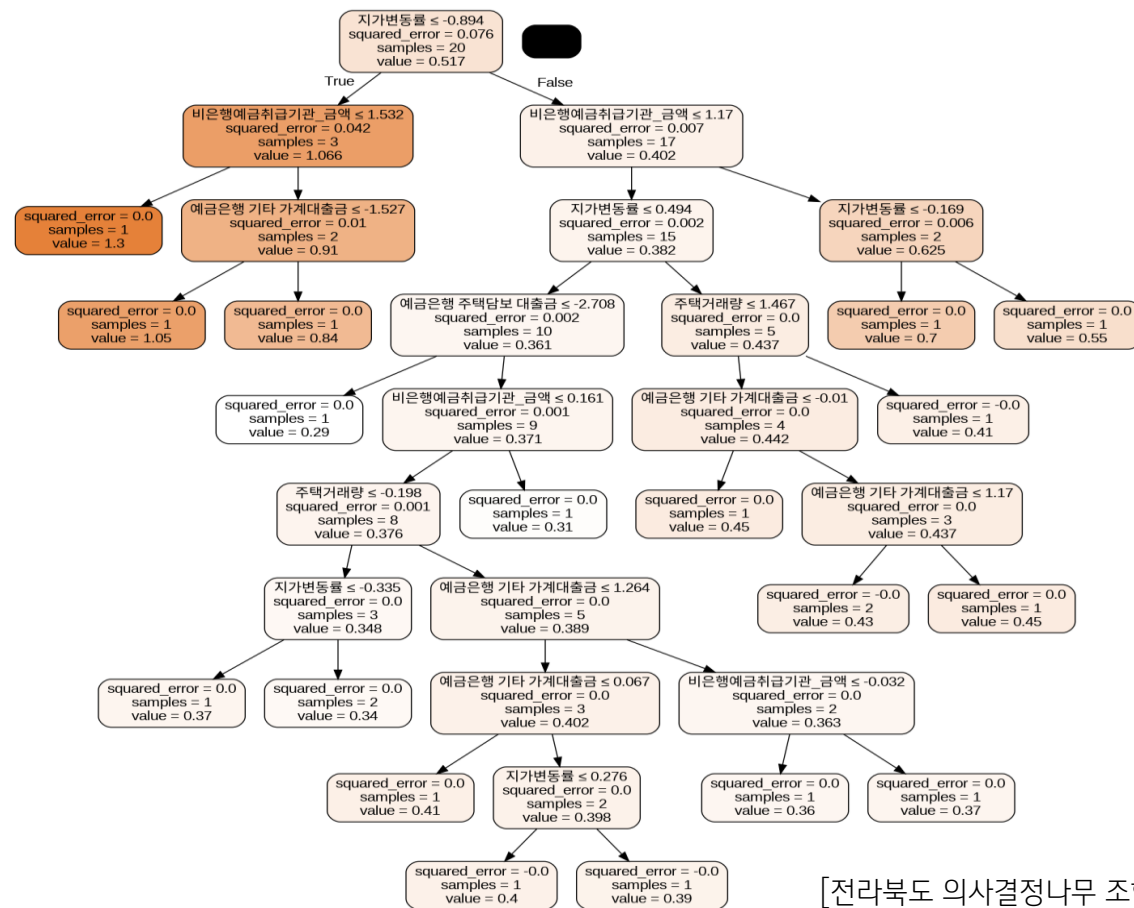
[Ex. 전라북도, 시차: 3개월]

## 랜덤포레스트 회귀 모델

- 여러 개의 의사결정나무를 조합해 더 정확한 모델을 만드는 앙상블 기법
- 랜덤으로 선택한 데이터로 학습
- 다양한 학습 결과를 종합해 더 안정적인 예측 수행
- 과적합 방지 효과
- GridSearchCV 통해 하이퍼파라미터 튜닝 후 예측율 향상



[전라북도 랜덤포레스트 회귀 결과]



[전라북도 의사결정나무 조합 과정]

## GridSearchCV

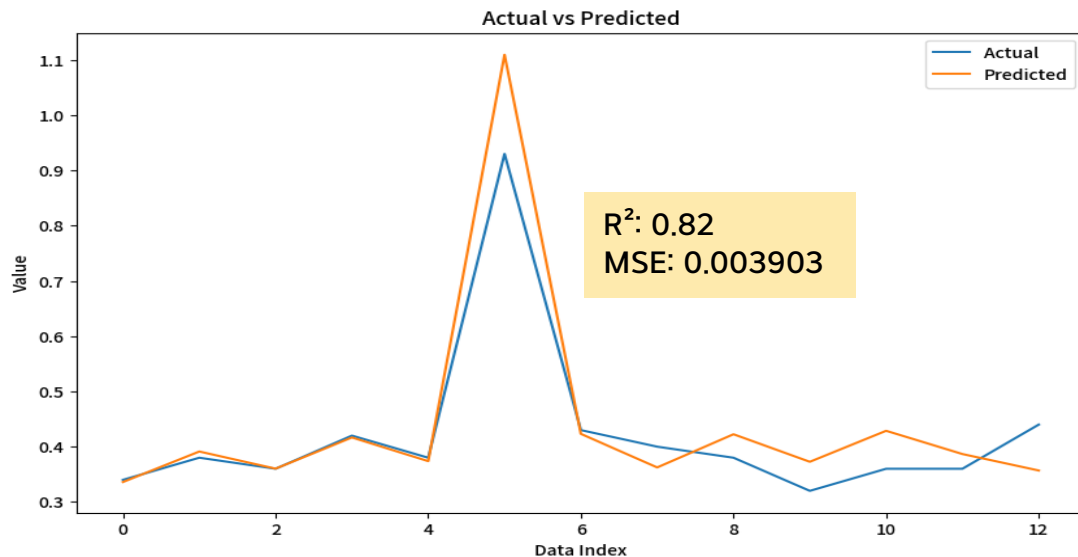
: 모델의 최적 파라미터를 검색할 수 있는 클래스 중 하나

# XGBoost Regression

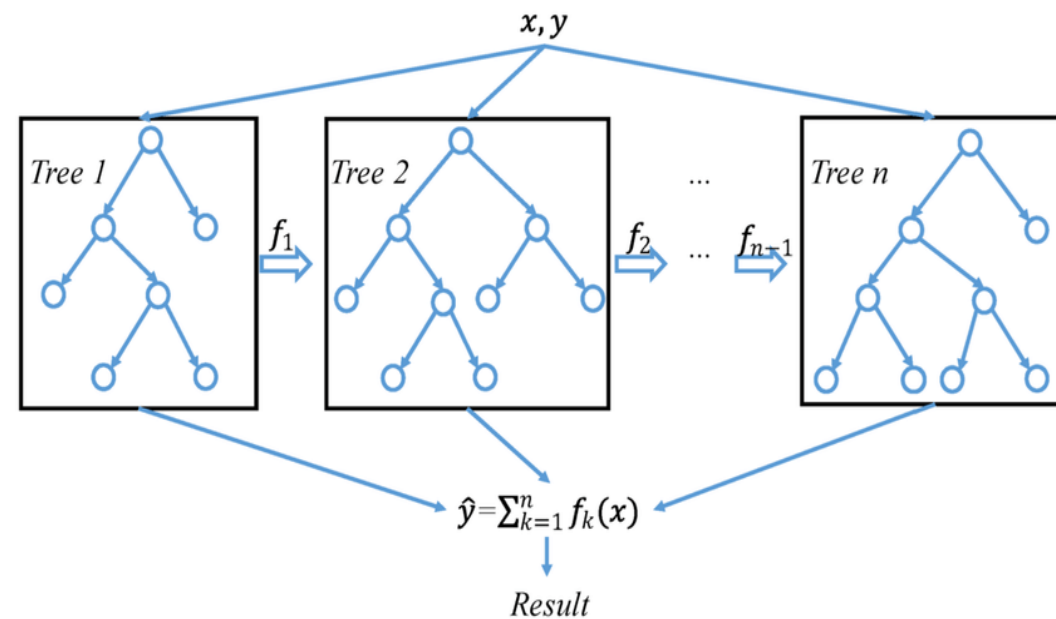
[Ex. 전라북도, 시차: 3개월]

## XGBoost(Regressor)

- 빠른 예측 속도와 높은 정확도
- 다른 모델에 비해 최적화된 성능을 보여주는 모델입니다.
- Gradient Boosting 알고리즘을 개선
- 금융 분야에서 활용도가 높습니다.



[전라북도 XGBoost 결과]



## 부스팅(Boosting)

: 성능이 약 50%인 학습기들을 결합하여 90% 이상의 성능을 보여주는 강한 학습기로 결합하는 기법

- 기존 모델의 오차를 최소화하도록 새로운 모델을 계속해서 추가&결합

## 원리

- 의사결정나무를 다양하게 만들어 집계
- 데이터의 노이즈로부터 보호
- 전체적으로 예측모델의 정확도 향상

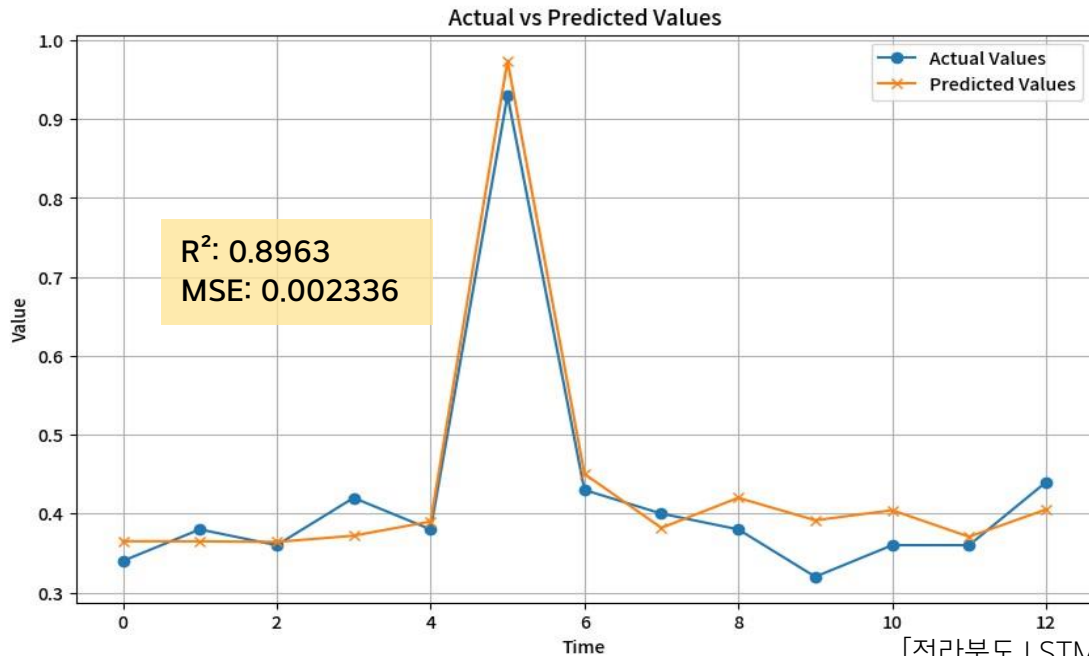


# LSTM

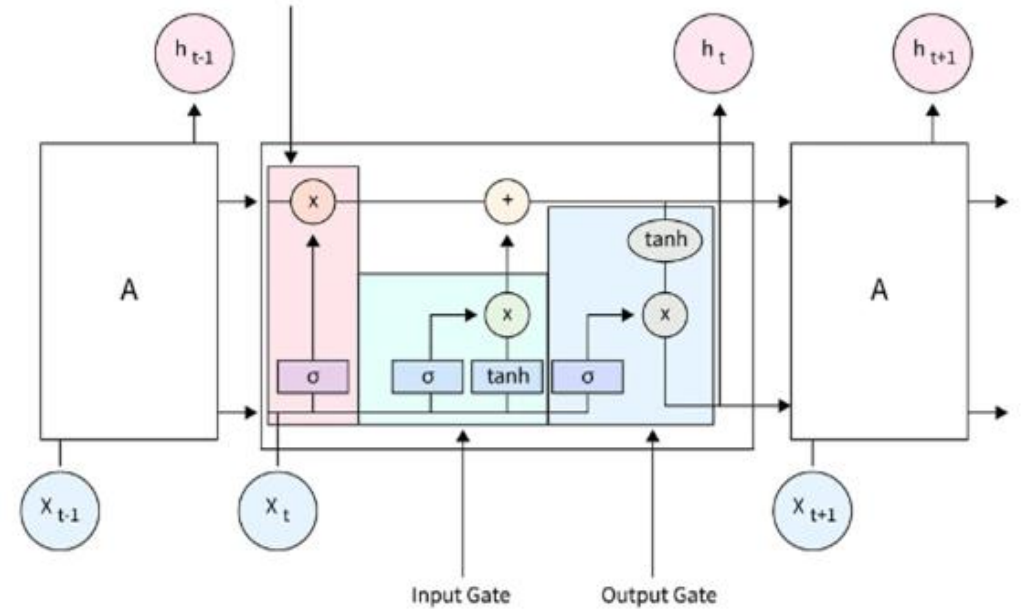
[Ex. 전라북도, 시차: 3개월]

## LSTM(Long Short-Term Memory)

- 순환신경망(RNN)의 종류
- 시계열 데이터나 순차적 데이터 처리에 특화된 딥러닝 아키텍처
- 과거 정보를 기억하며 현재 입력에 따른 패턴 학습으로 장기적인 의존 관계 처리 가능



[전라북도 LSTM 결과]



## 그리드 서치(Grid Search)

: 사용자 지정 매개변수(하이퍼파라미터)를 설정하면 성능이 좋아지는 최적의 조합을 찾아줌  
원리

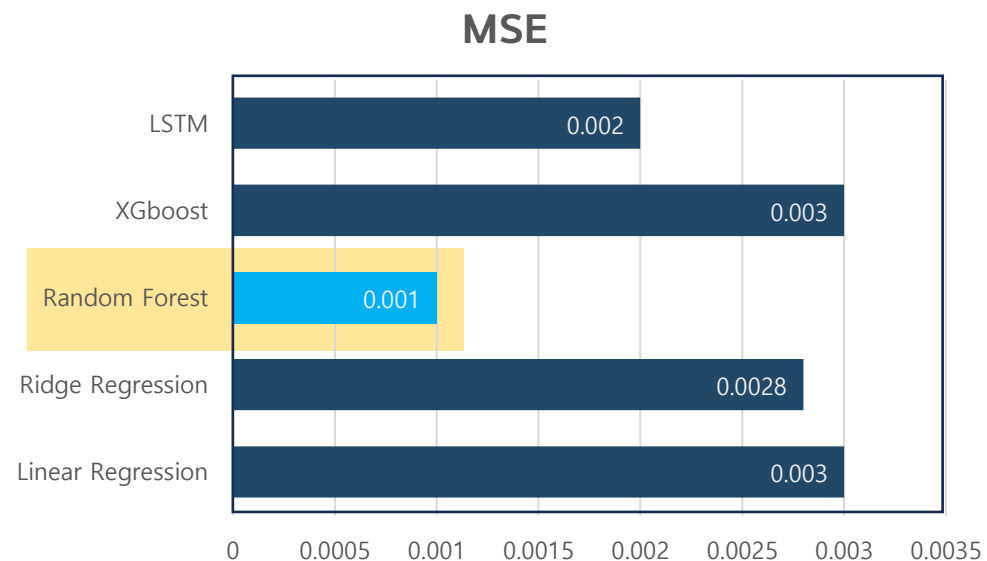
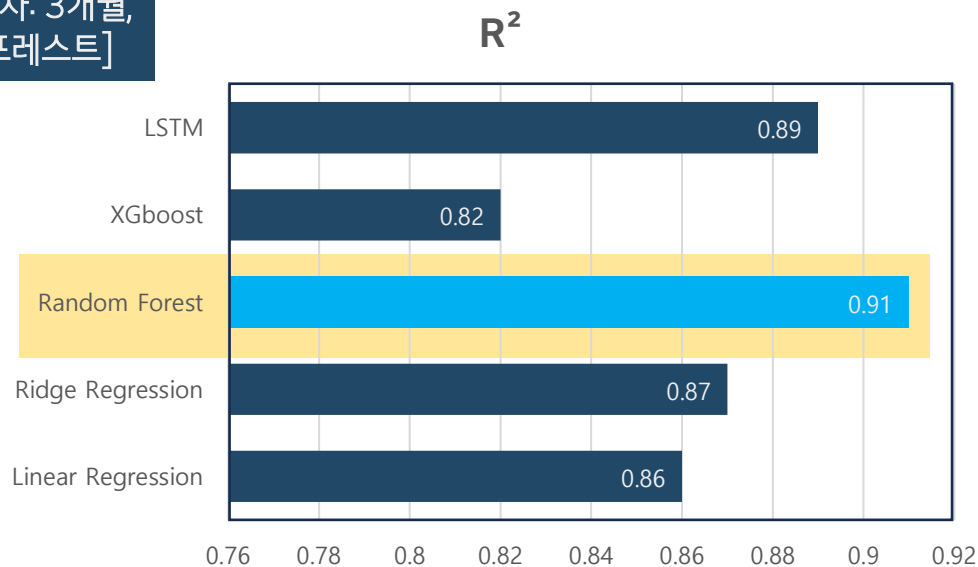
- 순환신경망의 일종
- 이전 출력 값이 다음 입력 값으로 사용, 장기적 기억 가능

## 파라미터 서치(Parameter Search)

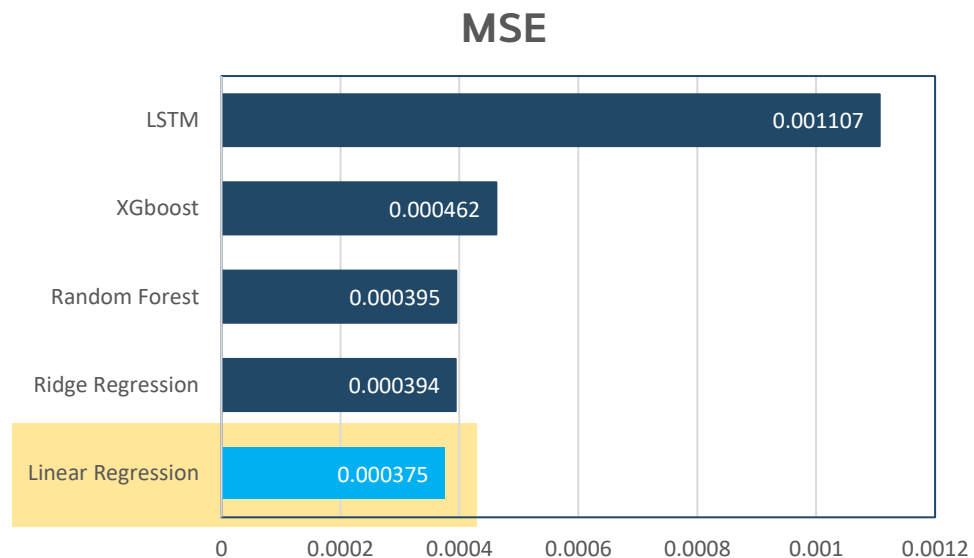
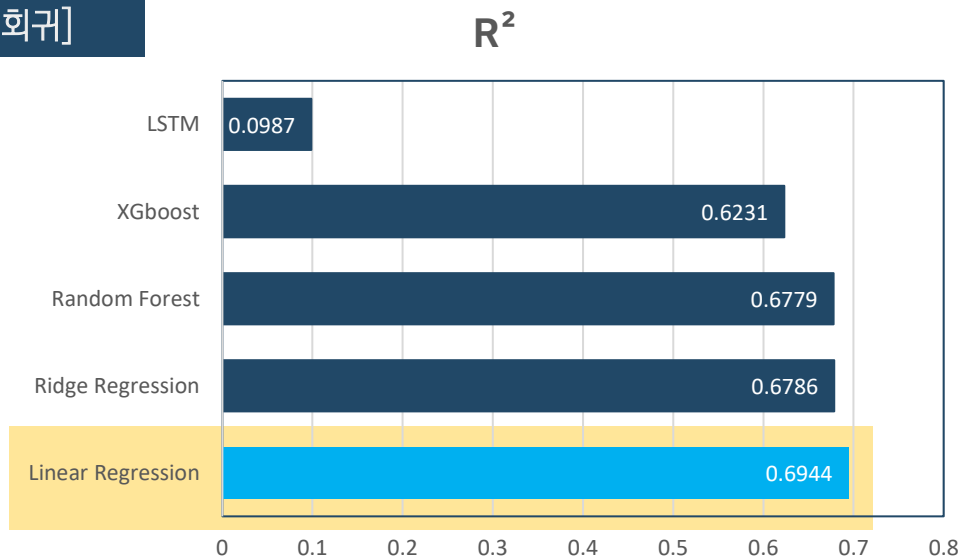
: 사용자가 지정한 매개변수(하이퍼파라미터) 중 가장 성능이 좋은, 다시 말해 mse가 가장 낮은 최적의 조합을 찾아준다

# 2차 모델링 성능

[Ex. 전라북도, 시차: 3개월,  
최적모델: 랜덤포레스트]

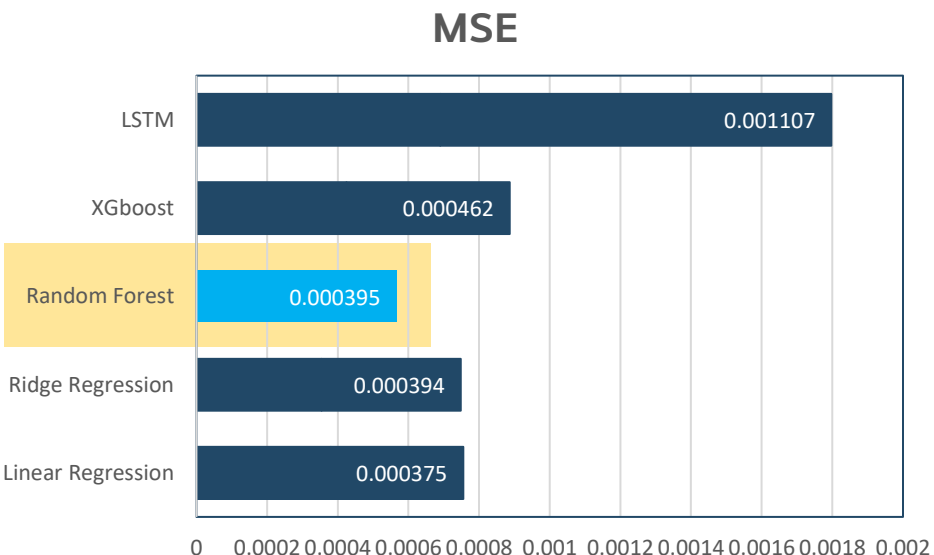
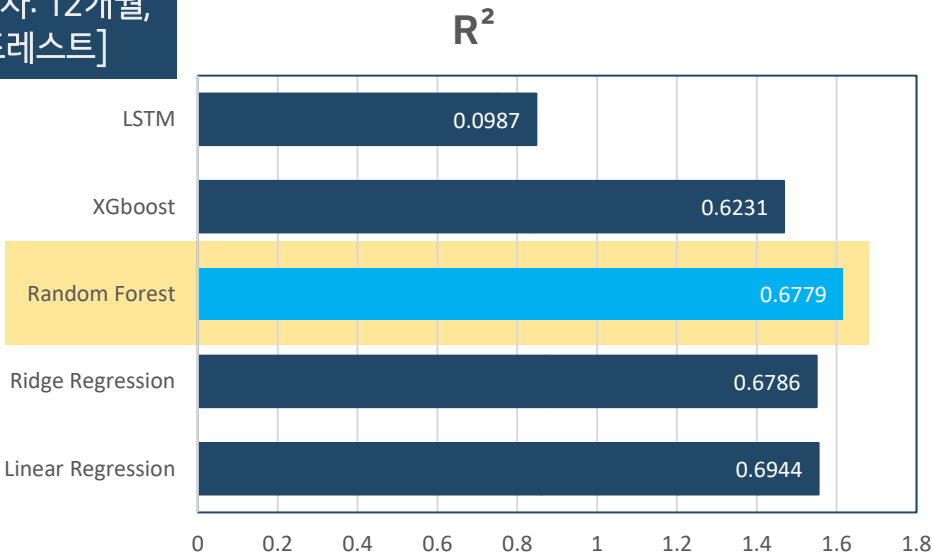


[Ex. 경상북도, 시차: 6개월,  
최적모델: 선형회귀]



# 2차 모델링 성능

[Ex. 서울특별시, 시차: 12개월,  
최적모델: 랜덤포레스트]



[지역별 최적 모델]

[최적모델: LSTM]

지역	R <sup>2</sup>	MSE
부산광역시	0.87	0.000481
충청남도	0.80	0.000127

[최적모델: 랜덤포레스트]

지역	R <sup>2</sup>	MSE
경기도	0.91	0.000153
광주광역시	0.84	0.000617
대구광역시	0.84	0.000462
울산광역시	0.84	0.000388

[최적모델: XGboost]

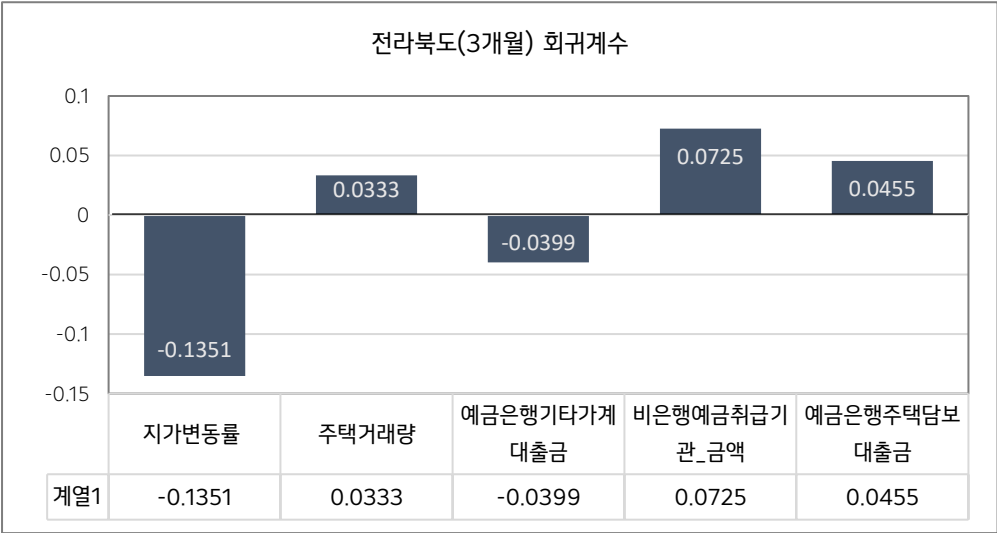
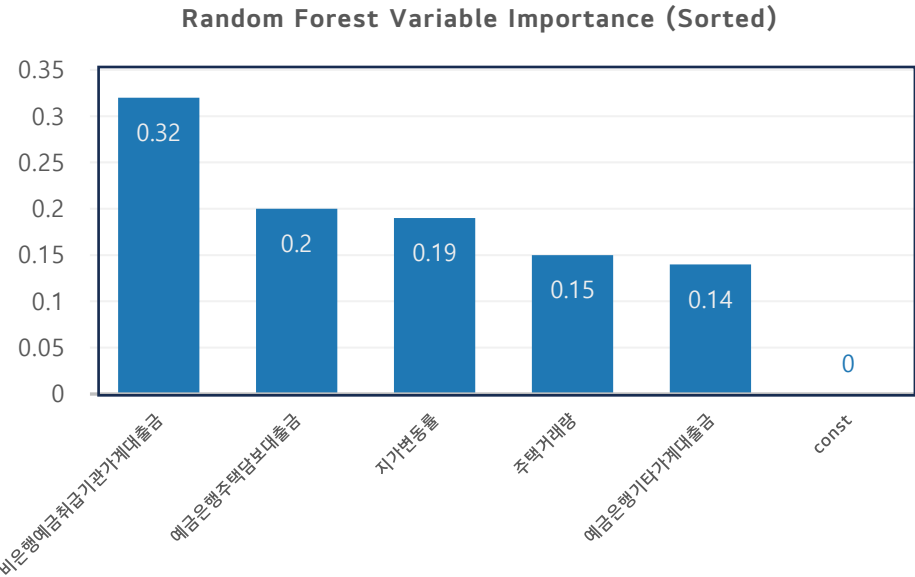
지역	R <sup>2</sup>	MSE	지역	R <sup>2</sup>	MSE
강원도	0.64	0.000119	인천광역시	0.77	0.000254
경상남도	0.56	0.000792	전라북도	0.92	0.001868
서울특별시	0.94	0.000171	제주특별자치도	0.48	0.002733
			충청북도	0.14	0.000297

[사회과학 분야에서는 결정계수(R<sup>2</sup>)가 0.13 이상이면 유효한 결과라고 해석-Cohen J]



# 분석 결과 및 맞춤 정책

[Ex. 전라북도, 시차: 3개월]



## [전라북도 결과]

### 부동산지표 변수

자가변동률(-), 주택거래량(+)

### 가계대출지표 변수

비은행 예금취급기관 가계대출금(+)  
예금은행 주택담보 대출금(+)  
예금은행 기타 가계 대출금(-)

### 변수 원인 분석

- 1. 타 지역에 비해 중저신용자가 다수 거주
- 2. 중저신용자 대상 대출 프로그램 진행(JB햇살론 البنك)
- 3. 높은 수치의 주택 관련 지표로 주택 가격 인상 예상

### 변수 결과 분석

- 1. 중저신용자의 가계대출 증가
- 2. 주택거래량 증가

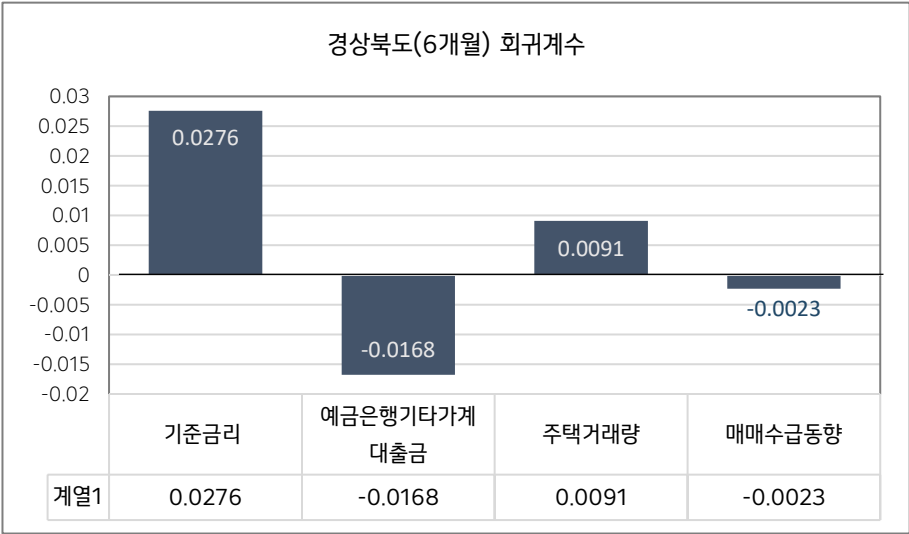
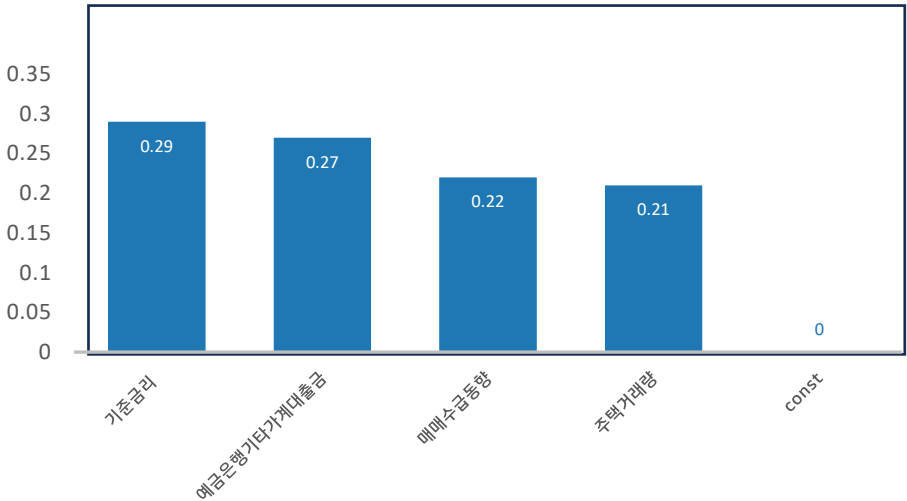
### 실행 정책

- 1. 중저신용자에 대한 대출한도조정
- 2. 신용 개선 프로그램
- 3. 지자체 금융교육강화
  - 금융상품 이용방법, 대출원리, 상환 방법 이해도 증가는 중저신용자 대출 금액줄일 수 있음
- 4. 대출 상담 및 평가 강화
  - 대출 상담 프로그램을 확대하고, 대출 신청 시 신중한 평가를 통해 대출 부담을 최소화할 수 있도록 도움

# 분석 결과 및 맞춤 정책

[Ex. 경상북도, 시차: 6개월]

Random Forest Variable Importance (Sorted)



## [경상북도 결과]

### 부동산지표 변수

매매수급동향(-), 주택거래량(+)

### 가계대출지표 변수

기준금리(+)  
예금은행기타가계대출금(-)

### 변수 원인 분석

- 1. 가계소득과 안정성이 높음
- 2. 고위험대출자가 금리에 민감
- 3. 미분양 주택 수가 많아 주택 공급이 수요를 따라가지 못함

### 변수 결과 분석

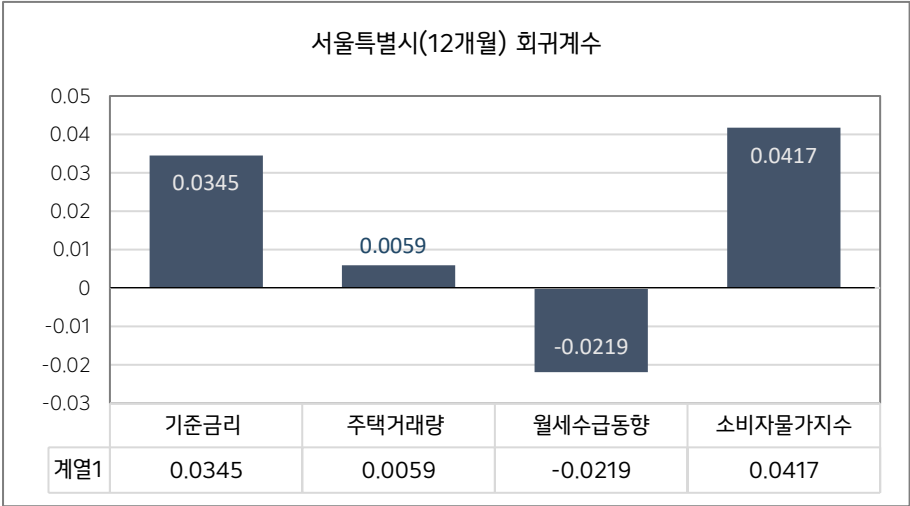
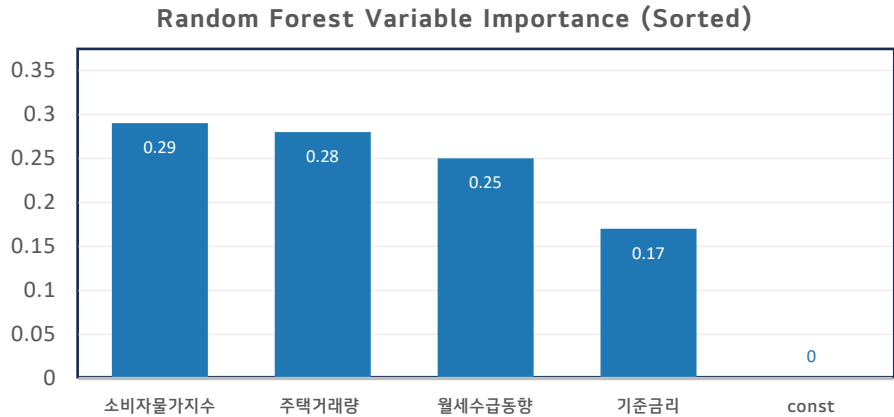
- 1. 중저신용자의 가계대출 증가
- 2. 미분양 주택으로 인한 불안정성

### 실행 정책

- 1. 예금은행 기타 가계대출금의 시차 분석해 대출규제 및 지원 정책  
- 대출 수요의 변화 예측 후 부도 위험이 높은 가계에 대한 대출 규제 강화
- 2. 고위험대출자에 대한 리스크 관리 강화 방안 마련  
- 중저신용자 대출 고객들에게 대출 후 초반 3개월 이상 프로그램을 제공

# 분석 결과 및 맞춤 정책

[Ex. 서울특별시, 시차: 12개월]



[서울특별시 결과]

### 부동산지표 변수

월세수급동향(-), 주택거래량(+)

### 가계대출지표 변수

기준금리(+)

### 경제지표 변수

소비자물가지수(+)

### 원인

1. 소비자물가지수 상승
2. 소득수준과 무관하게 대출해주는 '특례보금자리론' 실행으로 2030세대 주택 거래 증가
3. 금리 동결로 인한 주택담보대출 증가

### 결과

1. 가계부담 증가주택거래량 증가
2. 상환 능력이 적은 2030에 의한 연체율 상승 가능성

### 실행 정책

1. 2030을 타겟으로 한 부동산 정책 마련
  - 대출 조건을 완화시키는 대출 장려 대신 저축 장려 시스템 구축
2. 무리한 투자를 하는 사람들에게
  - 연체율 상승 위험에 대한 경각심을 심어주는 금융교육 추진
  - 일자리를 구하는 노력이 없는 청년에게 불이익
  - 자신의 경제력을 이해하고 그에 맞는 소비를 할 수 있게끔 도움

# 한계점 및 개선방안

## 한계점

### 소표본

- 지역별로 42개 관측치들로 구성된 소표본들이기에
- 학습, 검증 데이터 분할과정에서 이 둘의 분포가 달라질 수 있음
- 최종 모델의 성능에서 과적합, 과소적합 혹은 일반화의 오류 문제 발생 가능

### 잠재 변수

- 대출 연체율에는 사람의 심리 및 국가의 정책들이 영향을 미침
- 자연 재해나, 팬데믹 등 변수들도 고려해야 함

### 데이터 수집의 한계

- 연체율을 확실히 분석하기 위해서는 채무자의 미시 데이터가 필요
- 개인정보의 문제 발생

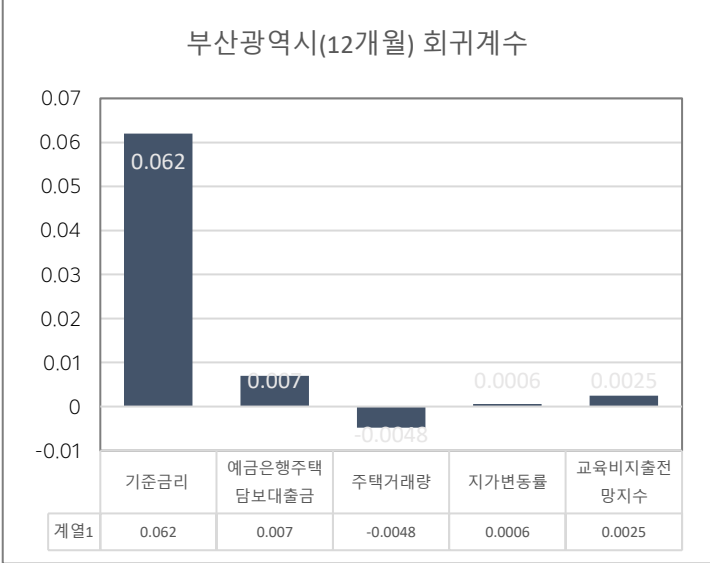
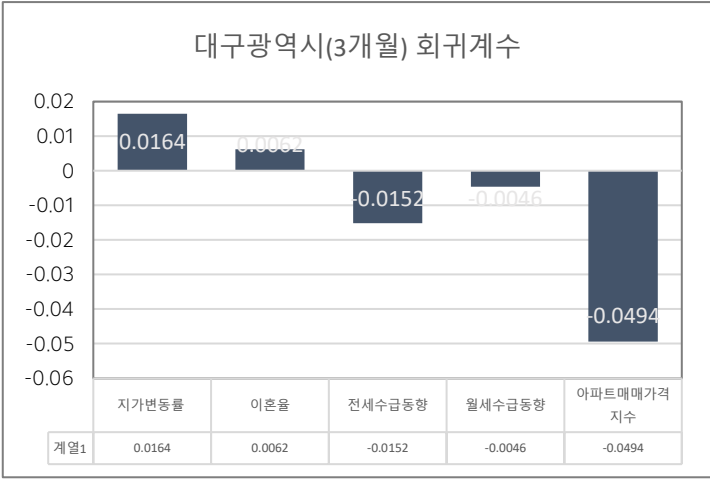
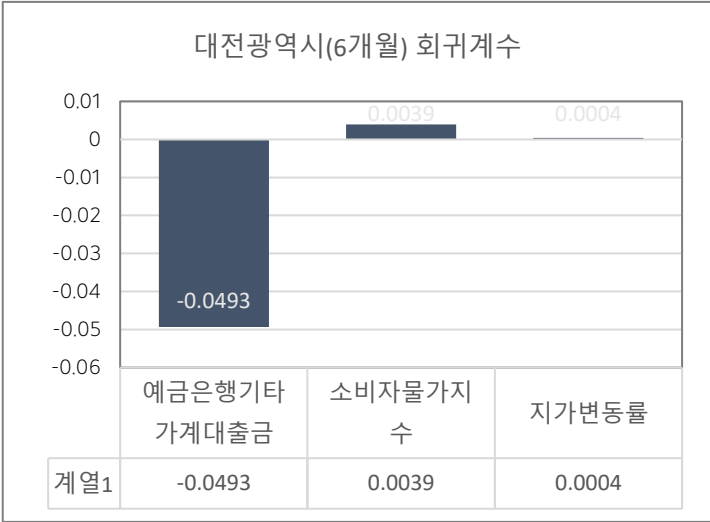
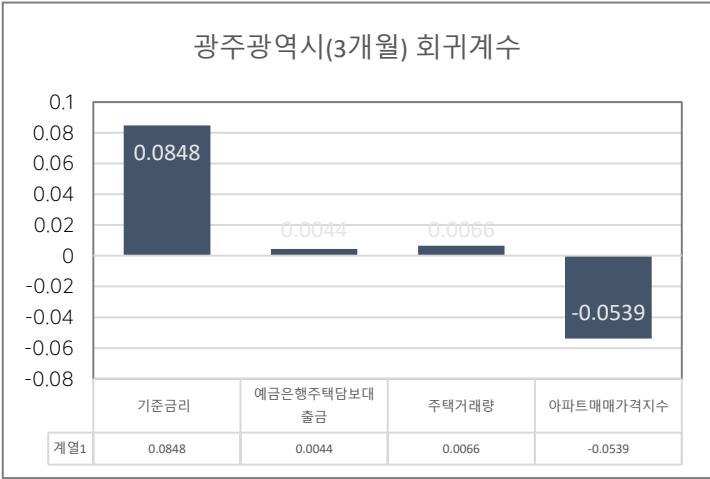
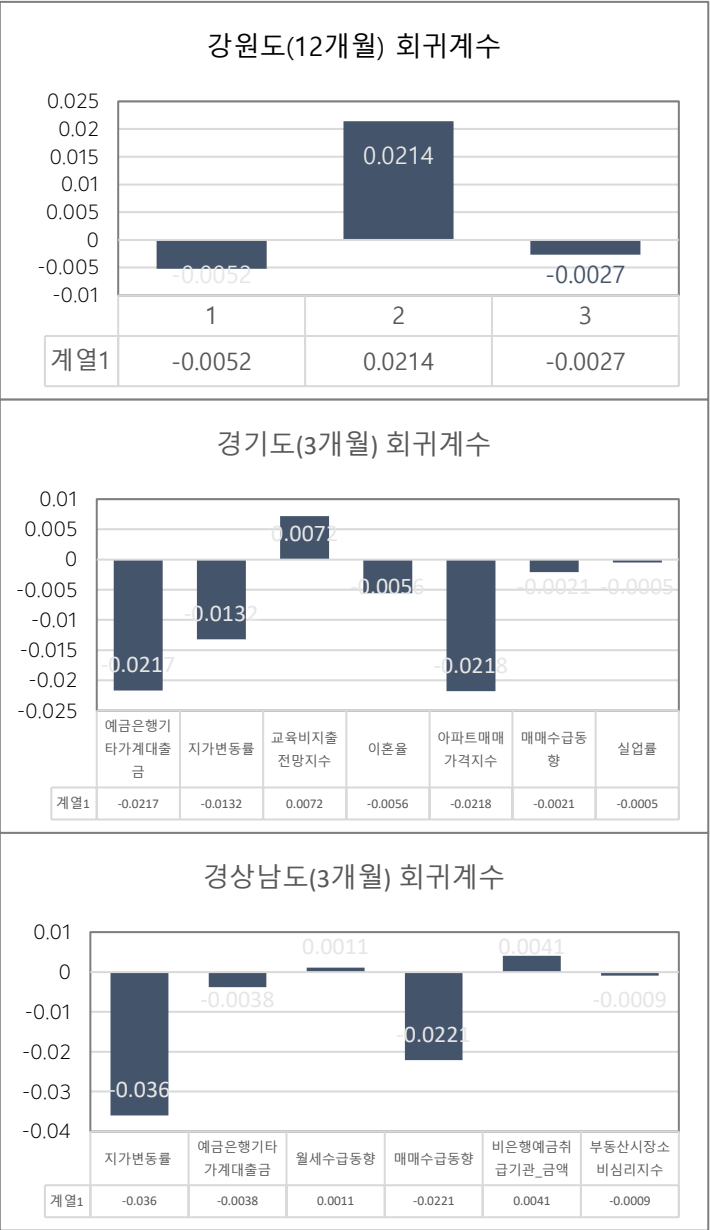
## 개선방안 및 의의

1. 앞으로 더욱 많은 금융 자료들이 체계적으로 정리되어 공개된다면, 보다 다양한 연체율 관련 연구의 기반이 마련될 것이다.
2. 본 분석에서 사용한 변수 외에 타 변수를 이용한다면, 가계대출연체율과의 새로운 관계성도 모색할 수 있다.
3. 미시 데이터와 결합하여 분석한다면 더 좋은 성능을 보여줄 것
4. 더욱 세세한 시차 분석을 통해 지역뿐만 아니라 전국에 가계대출연체율이 영향을 주는 시점을 파악해, 전국 단위 정책 도출에도 기여할 수 있다.

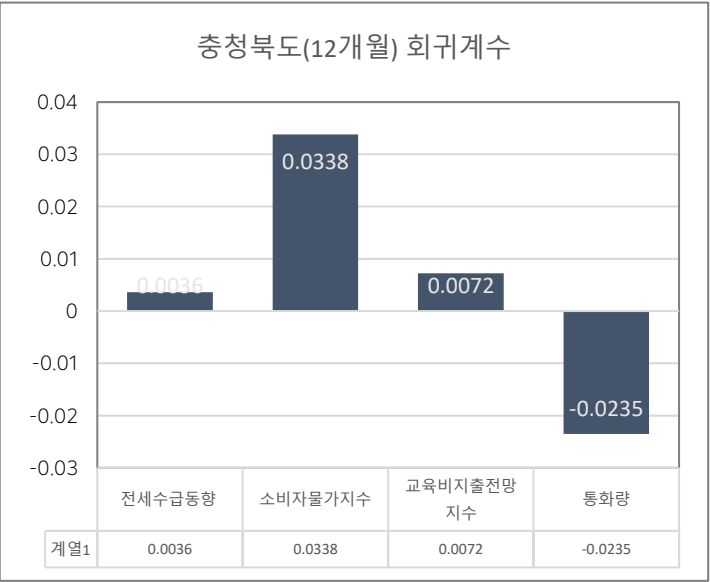
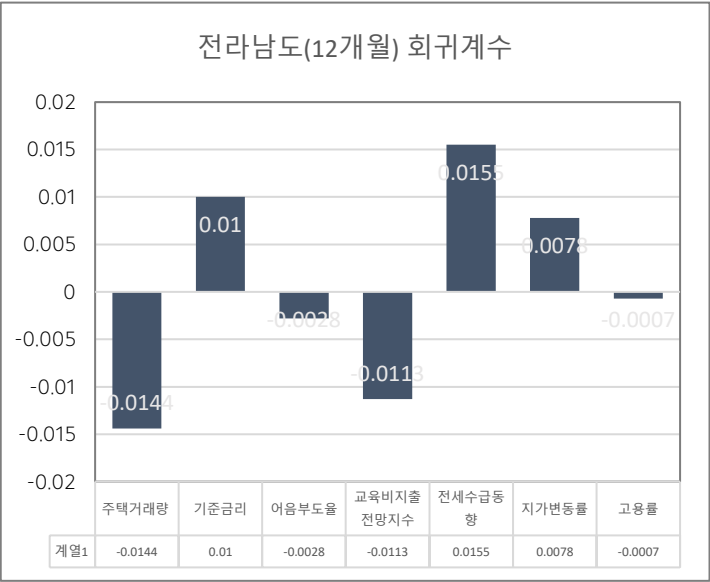
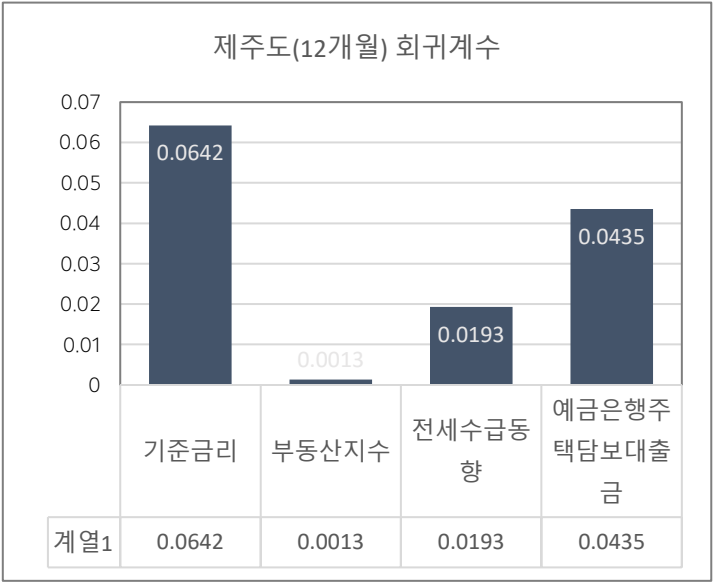
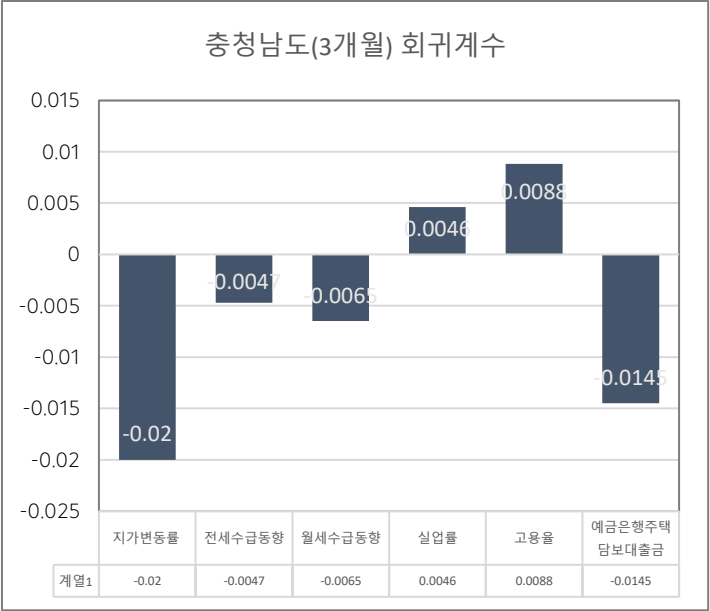
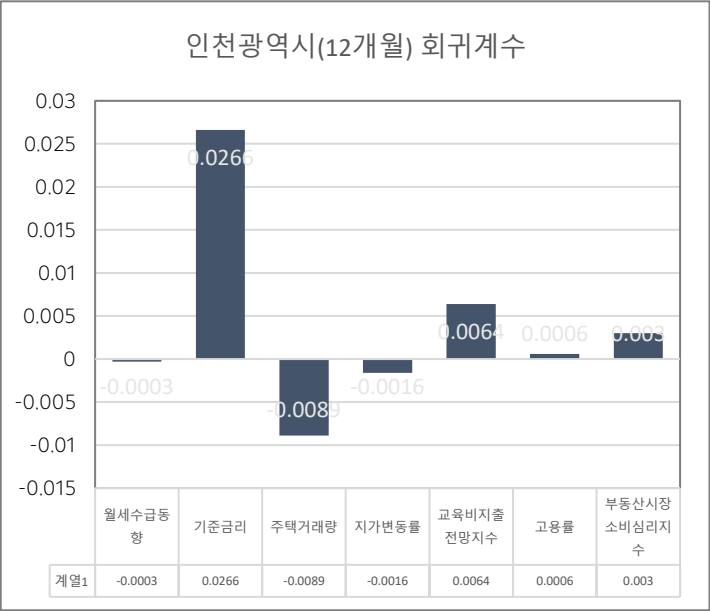
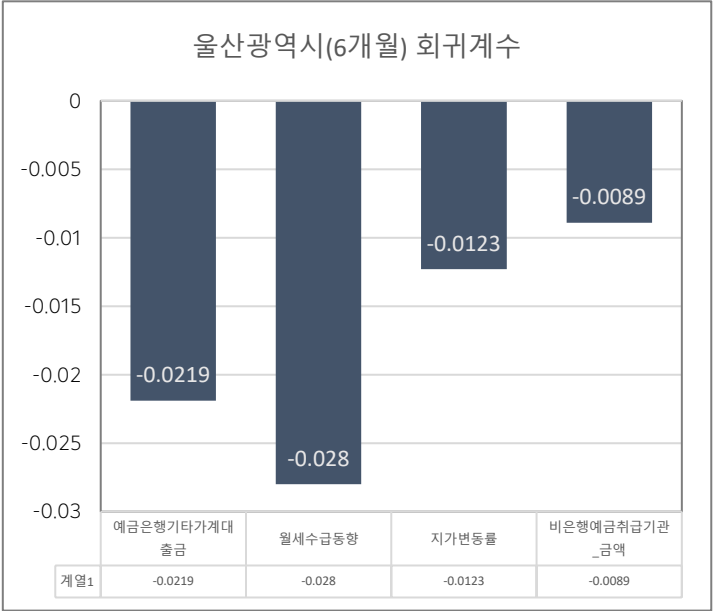




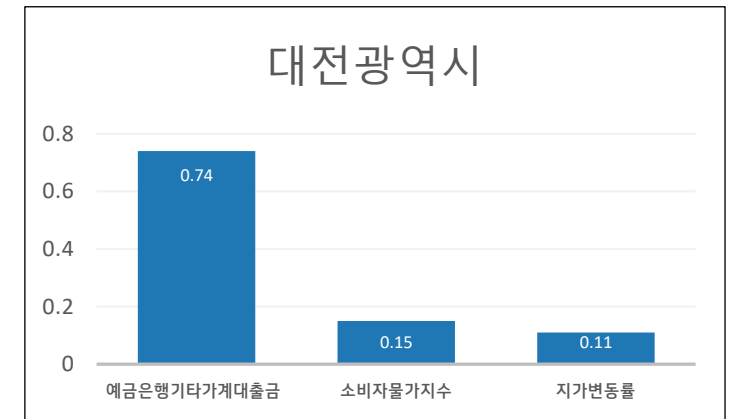
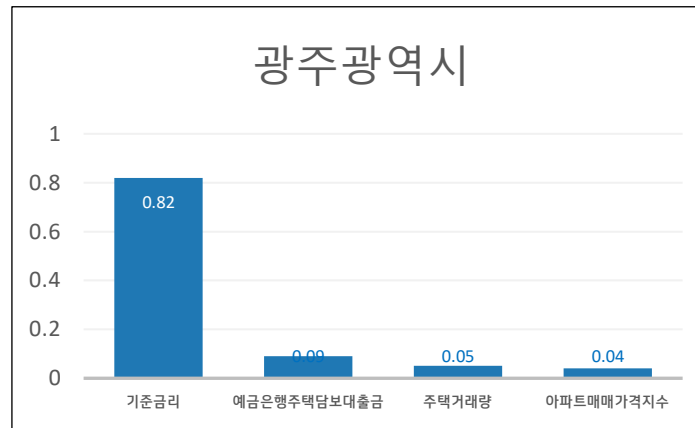
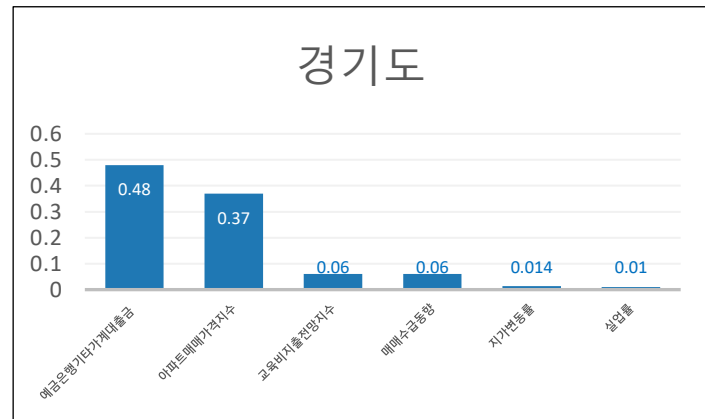
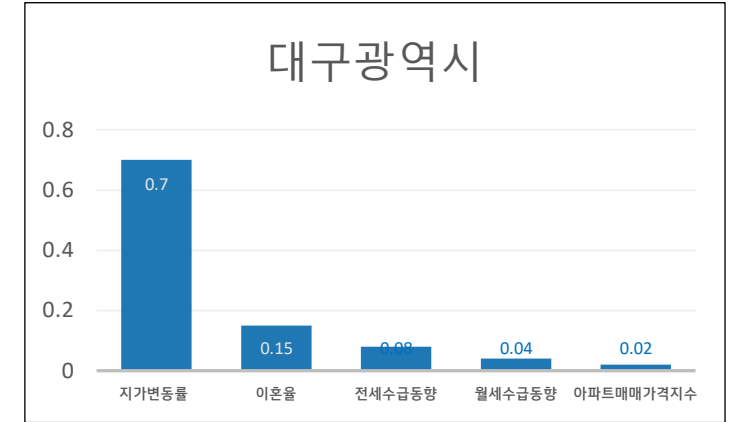
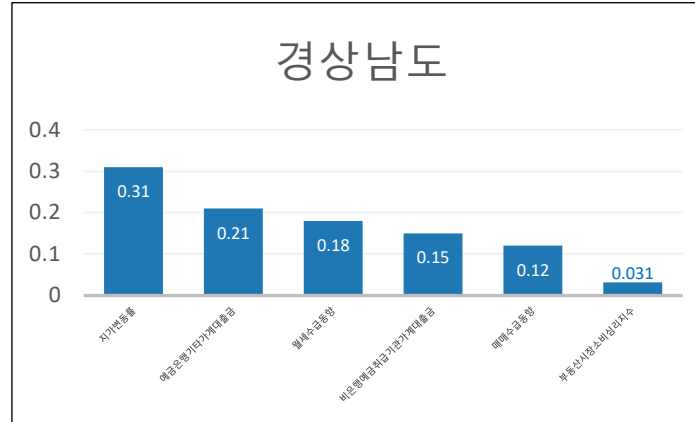
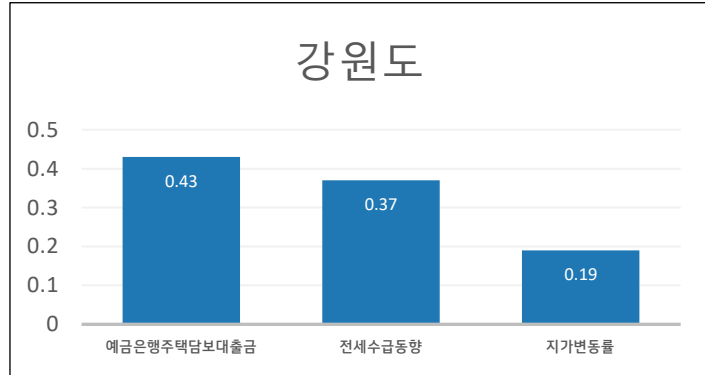
# 2차 모델링 회귀계수(지역별)



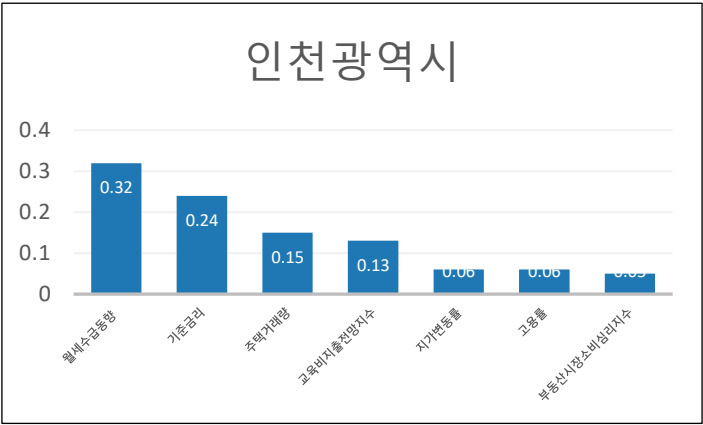
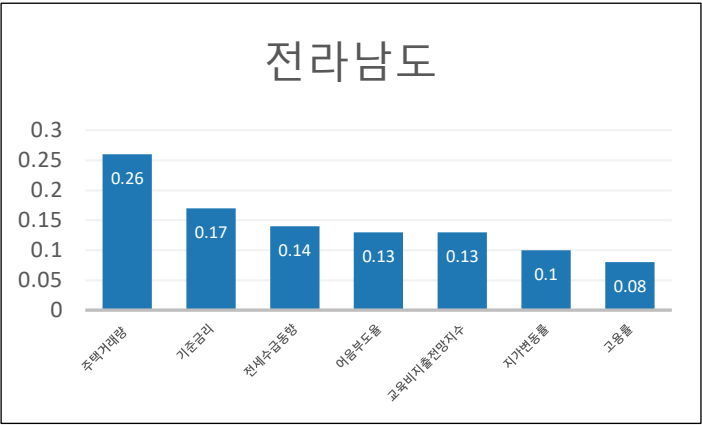
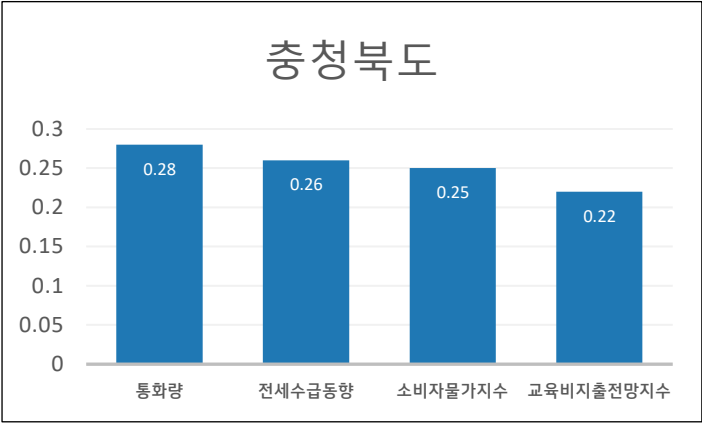
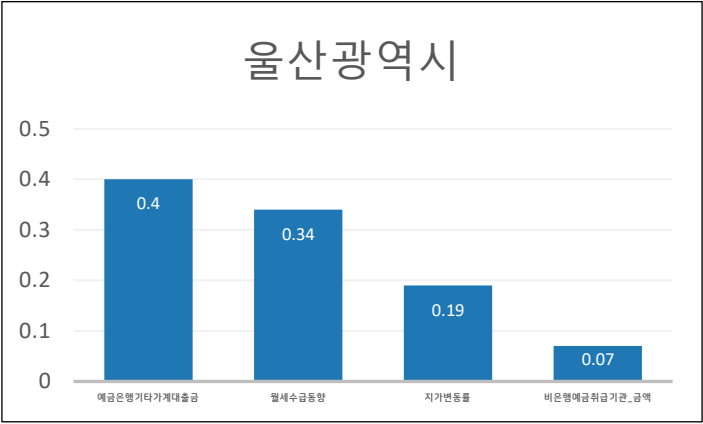
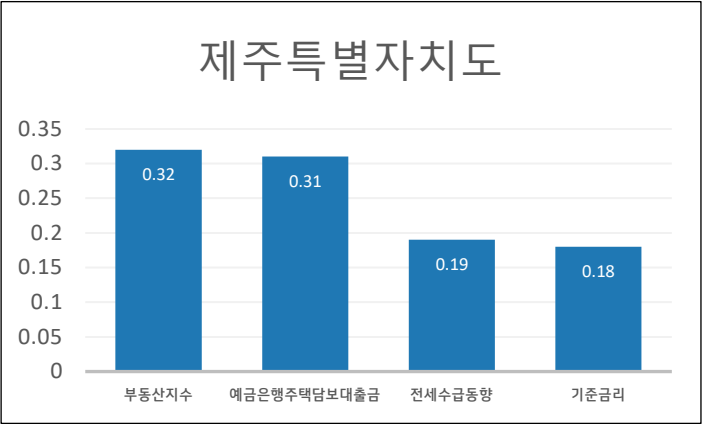
# 2차 모델링 회귀계수(지역별)



# 2차 모델링 변수중요도(지역별)



# 2차 모델링 변수중요도(지역별)



## 데이터

데이터명	출처
연체율 데이터	한국은행 경제통계시스템
어음부도율 데이터	한국은행 경제통계시스템
지가변동률 데이터	한국은행 경제통계시스템
비은행예금취급기관 데이터	한국은행 경제통계시스템
통화량	한국은행 경제통계시스템
기준금리	한국은행 경제통계시스템
부동산지수	한국은행 경제통계시스템
예금은행 주택담보대출금 데이터	한국은행 경제통계시스템
예금은행 기타가계대출금 데이터	한국은행 경제통계시스템
실업률, 고용률 데이터	KOSIS
주택수급동향	KOSIS
아파트매매가격지수 데이터	KOSIS
심리지수 데이터	KOSIS
주택거래량	한국부동산원

## 참고 문헌

[1] 김영준 ( Young Jun Kim ),남주하 ( Joo Ha Nam ),and 김상봉 ( Sang Bong Kim ). "국제경제 : 한국의 개인파산 증가원인 분석 : 2000년 이후를 중심으로." 국제지역연구 17.1 (2013): 149-170.

[2] 김정렬. "가계부채 규모와 연체율 결정요인에 관한 연구." 韓日經濟論集 66.- (2015): 25-46.

[3] 김영일. "거시경제 환경 변화에 따른 가계부채 리스크 점검과 시사점." 한국경제포럼 16.1 (2023): 1-37.

[4] 김희철(Hee-Cheul Kim),and 신현철(Hyun-Cheul Shin). "주택 담보 가계 대출액 결정요인 추정에 관한 패널 데이터 모형 연구." 韓國컴퓨터情報學會論文誌 16.7 (2011): 183-190.

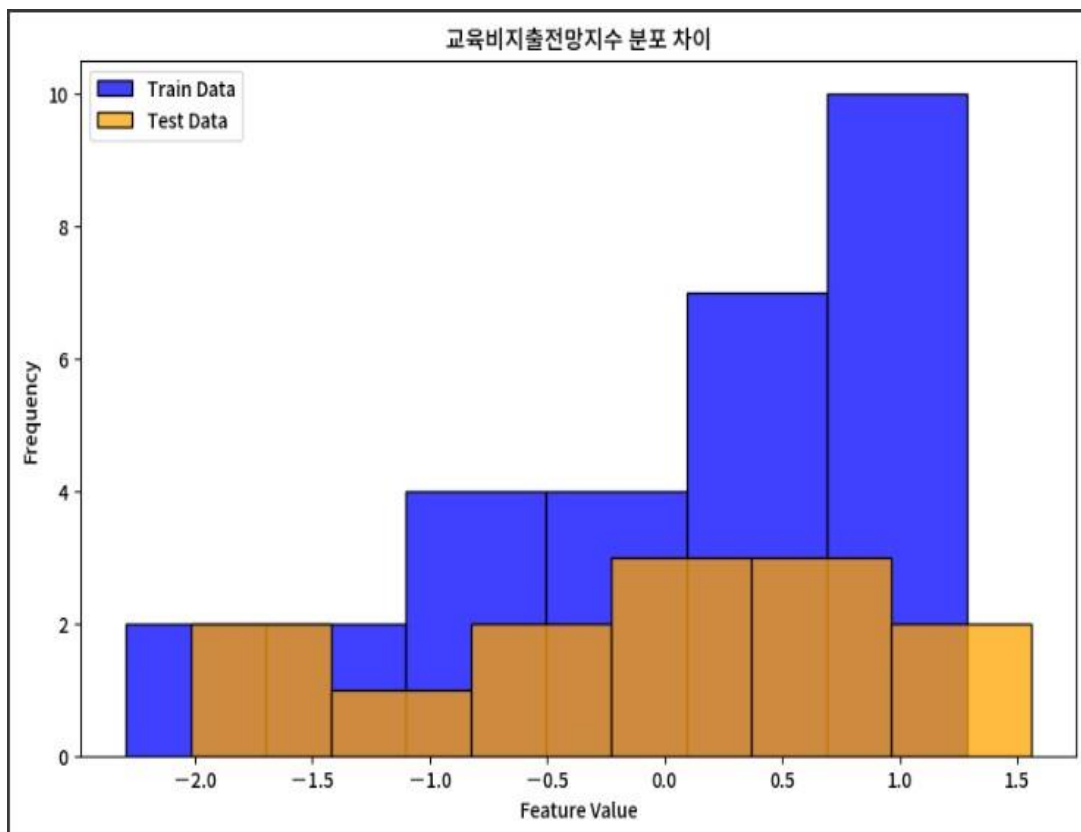
[5] 오규만, 이명훈. "가계부채 연체율에 영향을 주는 요인 연구- 의사결정나무를 활용한 변수 중요도를 이용하여 -." 주거환경(한국주거환경학회논문집) 17.3 (2019): 149-162.

[6] 이보미. "가계대출의 연체율에 대한 이해." 주간 금융 브리프 26.25 (2017): 3-7.

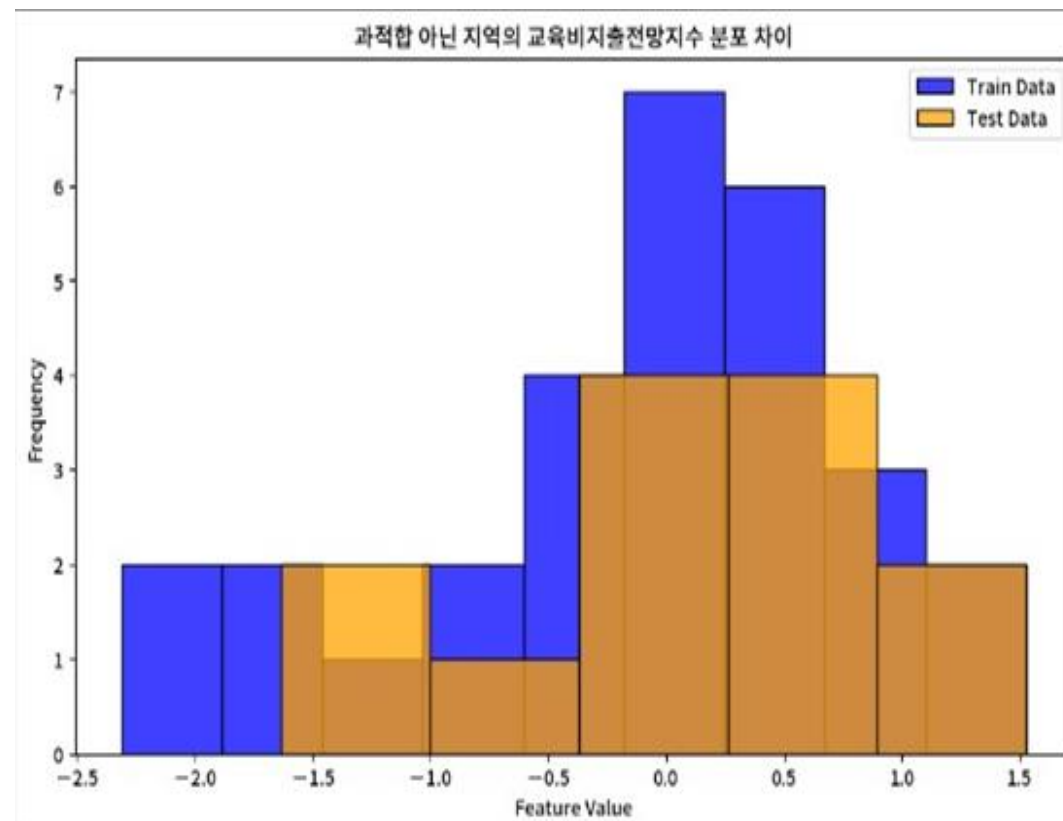
[8] Cohen, J.(1988), Statistical Power Analysis for the Behavioral Sciences(2nd Ed.), Lawrence Erlbaum Associates, Inc.

# 소표본이라는 단점에서 부각되는 성능 저하 요소

전라남도가 과적합인 이유



▶ 전라남도 교육비지출전망지수

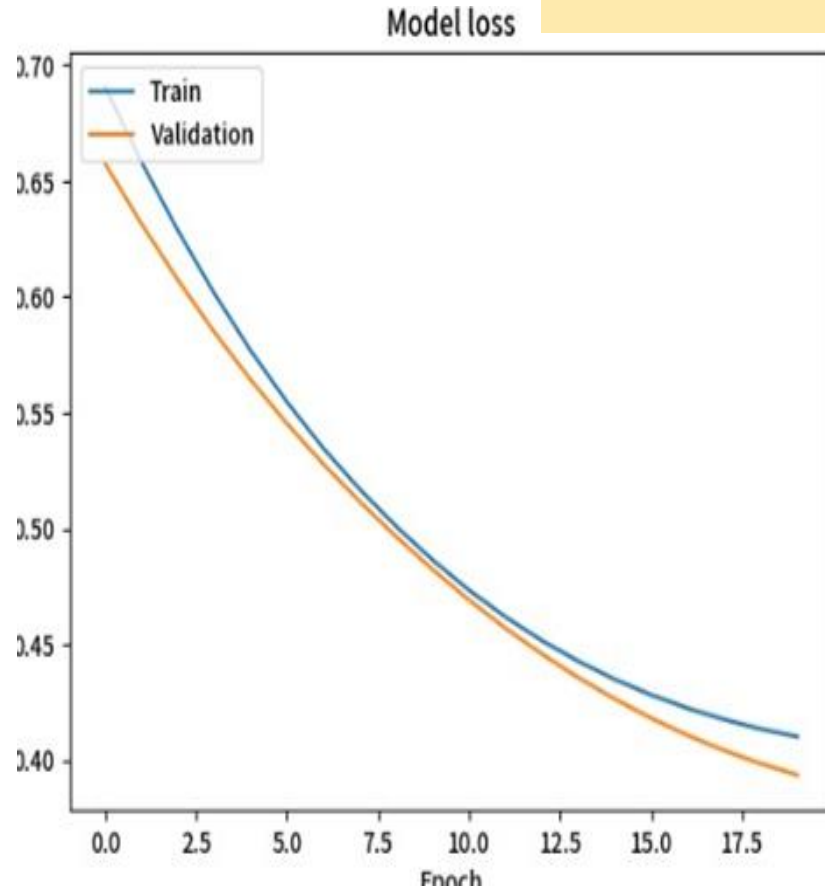


▶ 인천광역시 교육비지출전망지수

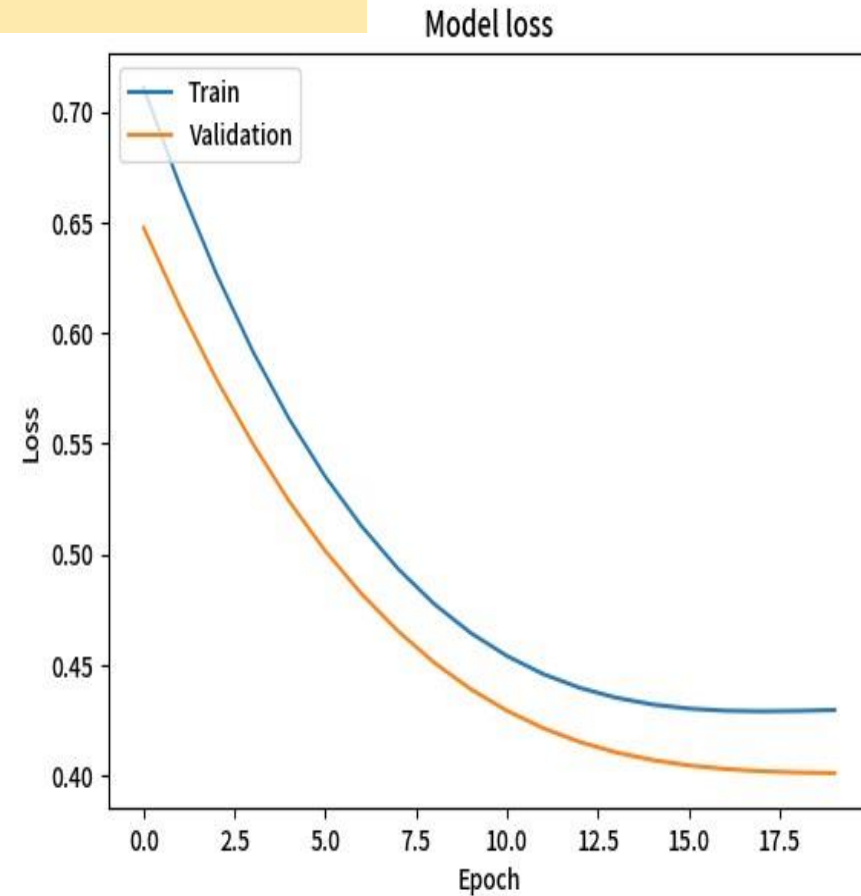
# 소표본이라는 단점에서 부각되는 성능 저하 요소

충청북도가 성능이 떨어지는 이유: 과소적합

검증데이터의 MSE > 학습데이터의 MSE



▶ 충청북도의 손실함수

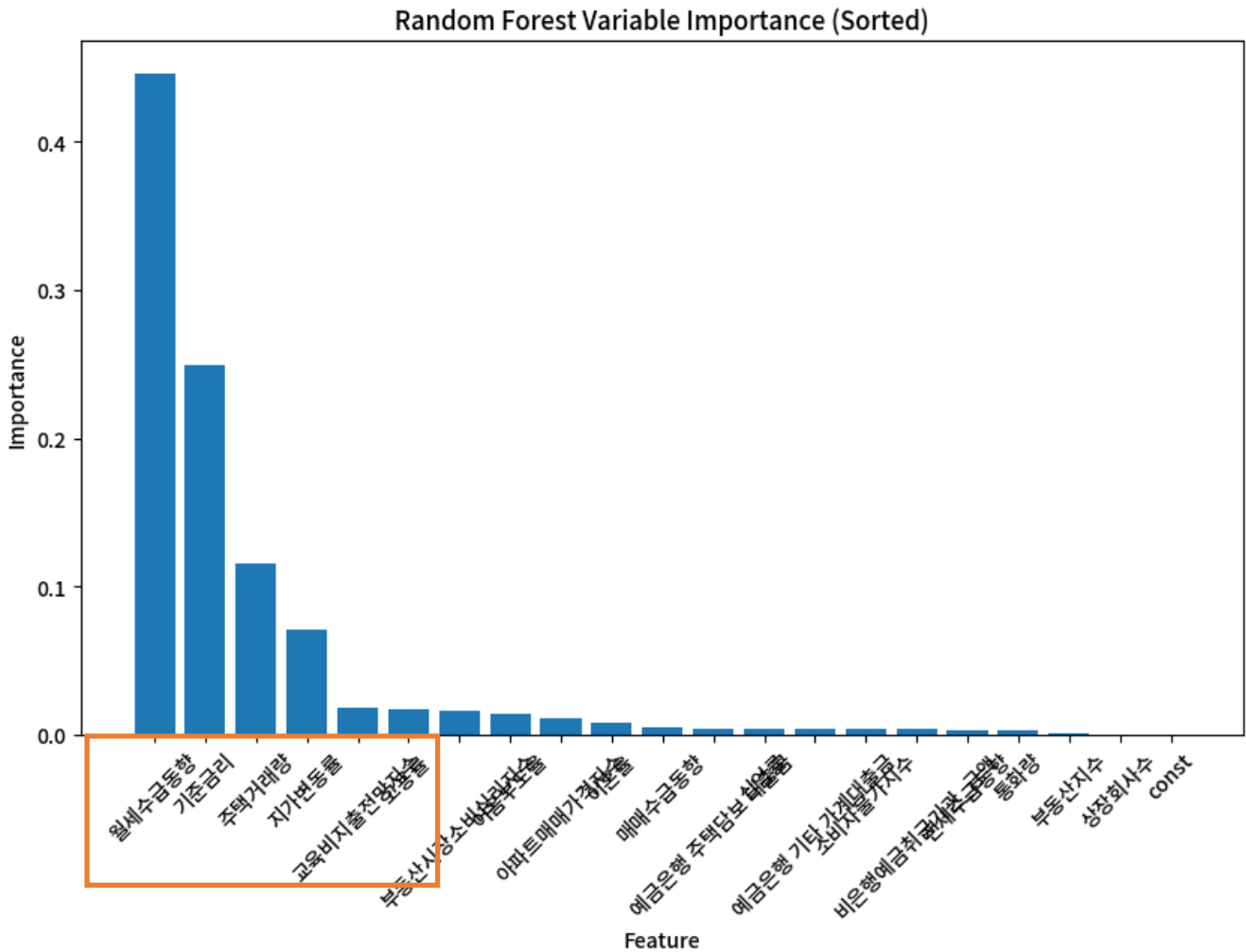


▶ 인천광역시의 손실함수



# 소표본이라는 단점에서 부각되는 성능 저하 요소

변수 중요도는 높지만 통계적으로 유의하지 않은 문제 발생



OLS Regression Results

Dep. Variable: 12개월 후 연체율

R-squared: 0.886

Model: OLS

Adj. R-squared: 0.863

Method: Least Squares

F-statistic: 37.89

Date: Fri, 25 Aug 2023

Prob (F-statistic): 3.02e-14

Time: 04:23:45

Log-Likelihood: 122.81

No. Observations: 42

AIC: -229.6

Df Residuals: 34

BIC: -215.7

Df Model: 7

Covariance Type: nonrobust

	coef	std err	t	P> t	[0.025	0.975]
const	0.1448	0.002	64.947	0.000	0.140	0.149
월세수급동향	-0.0003	0.006	-0.059	0.953	-0.012	0.011
기준금리	0.0266	0.006	4.189	0.000	0.014	0.040
주택거래량	-0.0089	0.004	-2.537	0.016	-0.016	-0.002
지가변동률	-0.0016	0.003	-0.540	0.593	-0.007	0.004
교육비지출전망지수	0.0064	0.003	1.922	0.063	-0.000	0.013
고용률	0.0006	0.003	0.196	0.846	-0.005	0.007
부동산시장소비심리지수	0.0030	0.003	1.107	0.276	-0.002	0.008

Omnibus: 4.677

Durbin-Watson: 1.503

Prob(Omnibus): 0.096

Jarque-Bera (JB): 3.381

Skew: -0.595

Prob(JB): 0.184

Kurtosis: 3.720

Cond. No. 6.69