

# 서울시 행정동별 상권변화지표 예측 기반 예비 소상공인 시장진입전략 기초자료 제공 및 정책적 지원방안 제시

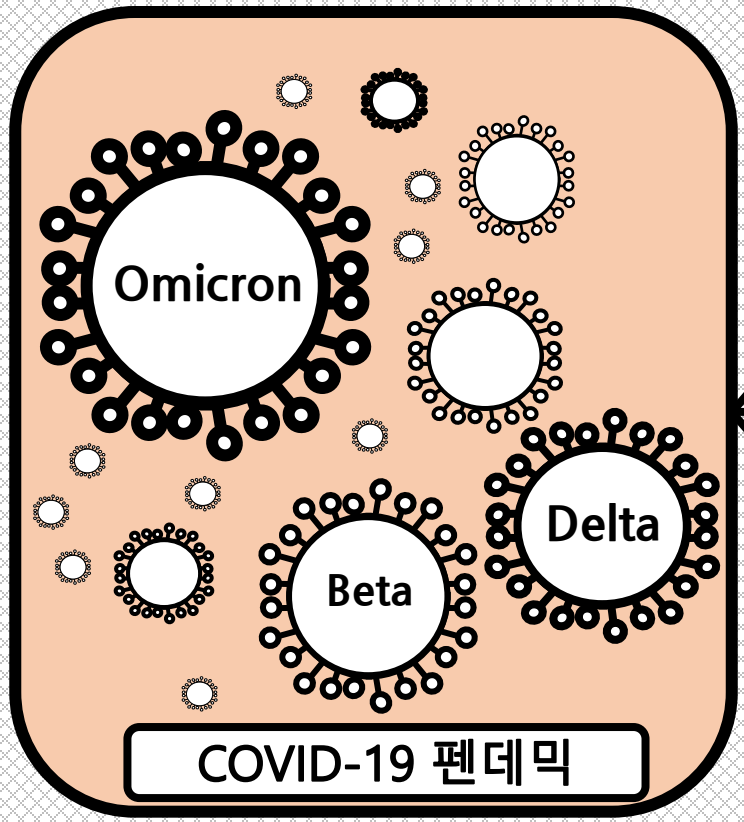
2022년 통계 데이터 분석 활용대회

<b>1. 연구 주제 및 분석 필요성</b> <ul style="list-style-type: none"><li>- 문제 상황</li><li>- 분석의 필요성</li><li>- 연구 주제 및 전략</li></ul>	<b>2. 연구 설계 (데이터 처리 방식, 프로세스)</b> <ul style="list-style-type: none"><li>- 활용 데이터</li><li>- 1 단계 전략 (상권지표예측모델 구축)</li><li>- 2 단계 전략 (예측 실시)</li><li>- 3 단계 전략 (급변동 예측 지역 분석)</li></ul>	<b>3. 분석 결과</b> <ul style="list-style-type: none"><li>- 1 단계 전략 분석 결과</li><li>- 2 단계 전략 분석 결과</li><li>- 3 단계 전략 분석 결과</li></ul>	<b>4. 결론 및 제안</b> <ul style="list-style-type: none"><li>- 상권변화 대응 방식 및 지원 방 안 결론</li><li>- 기대 효과</li></ul>
---	---	---	--

## **1. 연구 주제 및 분석의 필요성**

1. 연구 주제  
및 분석 필요성

<문제 상황>



1차부터 4차  
대유행까지,  
팬데믹의 장기  
적 영향력  
존재

비대면으로의  
전환에 상대적  
으로 취약한  
수익창출 구조  
+  
사회적 거리두  
기

경제적 충격  
경기 침체 → 각  
국별 양적완화 →  
인플레이션 → 긴  
축정책+원자재  
값 등 상승 →...

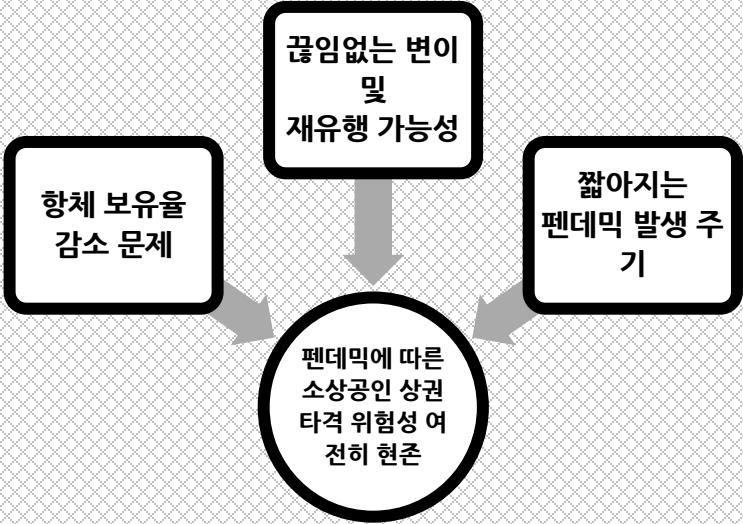
**소상공인 경제적 피해**

소상공인이란?  
:소상공인이란 대체로 취업자 가운데 비임금근로자 중에  
서는 자영업자, 그리고 임금 근로자 중에서는 소규모 기  
업(서비스업, 도소매업 등에서는 상시근로자 4인 이하)에  
서 근로하는 임금 근로자를 의미.

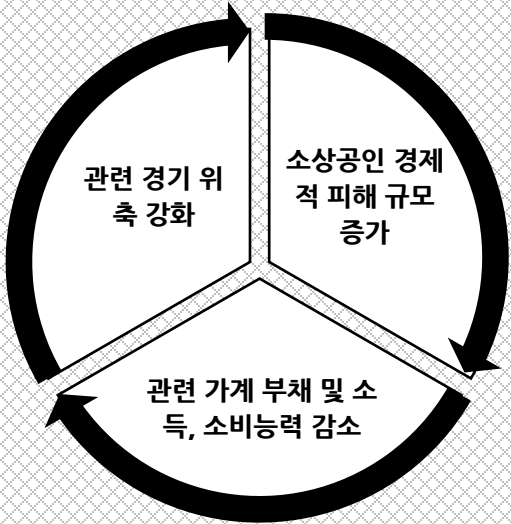
지난 2년은 COVID-19 팬데믹의 장기화와 이에 따른 경제적, 사회적 충격이 중심이 된 시간이었다. 이때 근 2년의 코로나-19 팬데믹은 단순히 현대 사  
회에서 질병이 갖는 파괴력과 이에 대한 탄력적 대응의 필요성 뿐만 아니라, 방역과 관련된 거시적 사회변동 발생 시 비임금근로자, 특히 자영업 소상공  
인들의 활동이 여타 산업에 비하여 비대면으로의 전환이 상대적으로 어렵다는 점에서 강한 경제적 타격을 받게 된다는 것을 보여주었다.

1. 연구 주제  
및 분석 필요성

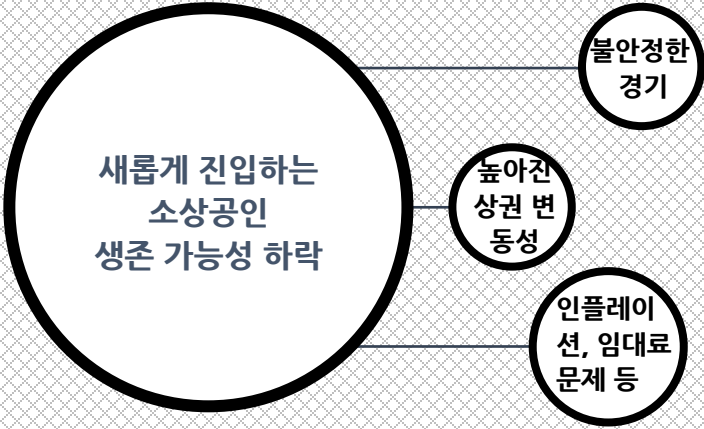
<분석의 필요성>



코로나-19와 같은 펜데믹으로 인한 소상공인 상권 피해에 대한 대응책 필요



순환 고리를 끊어낼 필요 존재

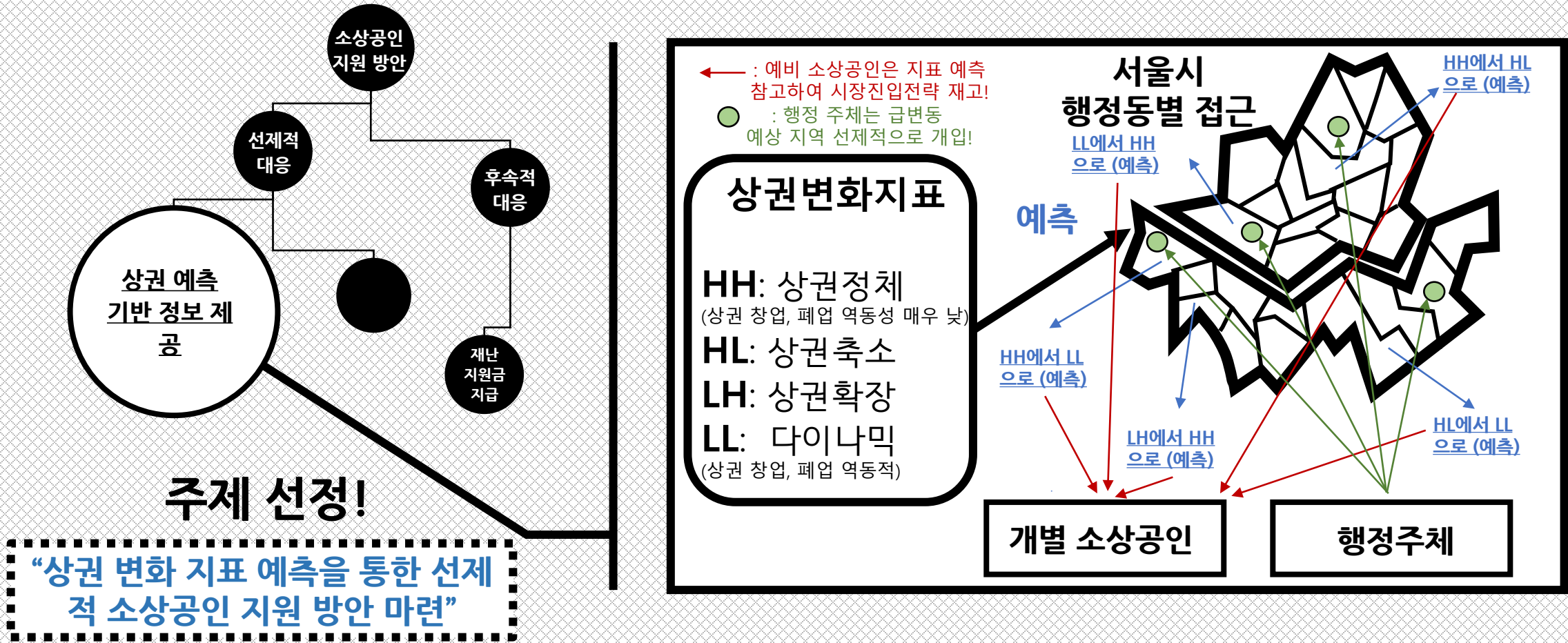


신생 소상공인 생존 가능성 증진 위한 개입 필요

이때 코로나-19로 인한 경제적 타격, 인플레이션, 높아진 상권 변동성, 재유행 가능성 등의 문제는 팬데믹에 따른 상권 피해 지속의 문제, 순환 고리 발생의 문제, 신생 소상공인의 생존 가능성 문제 등을 발생시키고 있다. 이에 갈수록 높아져만 가는 상권 피해 규모에 따라 소상공인을 지원하기 위한 방안에 대한 연구의 필요성이 발생하고 있으며, 또한 거시적인 차원에서도 경기에 영향을 미치는 소상공인의 생존은 국민 복지 및 후생 증진과도 밀접하게 연결되어 있기에 적절한 대응이 필요하다.

# 1. 연구 주제 및 분석 필요성

## <연구 주제 및 전략>



본 연구에서는 소상공인 지원 방안 가운데 선제적 대응에 초점을 맞추었으며, 구체적으로 개인 차원에서는 특정 지역에 진입하려는 예비 소상공인들이 향후 관련 상권 특성을 고려할 때 현재 자신의 진입 전략이 합리적인지를 선제적으로 판단할 수 있는 기초자료를 생성하는 한편, 행정 차원에서는 특정 상권에서 발생할 문제를 확인하고 선제적으로 영향요인을 조정함으로써 관련 소상공인의 경제적 타격을 줄일 수 있도록 하는 데 기여하는 선제적 대응 방안이 있을지를 고민하였다. 그 결과, 행정동별 상권변화지표 예측을 기반으로 관련 상권 특성을 선제적으로 제시하여 예비 소상공인들의 판단을 돕는 동시에, 개입이 필요할 것으로 예상되는 지역(상권지표 급변동 지역, (∵ 상권 급락, 급등 시 각각 단기적 차원, 장기적 차원에서 안정성 저해 및 경제적 피해 발생→개선 필요))을 선별하고 해당 문제의 영향요인을 추출함으로써 정책적 방향성을 제시하는 것을 연구 주제로 선정하게 되었다(범위는 서울특별시로 제한).

## **2. 연구 설계 (데이터 처리 방식, 프로세스)**

## 2. 연구 설계

### 1) [선정된 데이터]

데이터 종류	선정 이유	자료 기반
1) 서울시 우리마을가게 상권분석서비스_행정동별 상권변화지표	2014년부터 2021년 분기별 서울시 행정동별 상권변화지표	서울시 열린데이터 광장
2) 서울시 행정동별 세대 수	행정동별 인구 특성 반영	서울시 열린데이터 광장
3) 서울시 행정동별 인구 수	행정동별 인구 특성 반영	서울시 열린데이터 광장
4) 서울시 행정동별 여성 수	행정동별 성별 변수 효과 반영	서울시 열린데이터 광장
5) 서울시 행정동별 한국인 수	행정동별 인구 특성 반영	서울시 열린데이터 광장
6) 서울시 행정동별 노인 인구	행정동별 연령 변수 효과 반영	서울시 열린데이터 광장
7) 서울시 자치구별 이동자 수_순이동	자치구별 인구이동 특성 반영	통계청 국내인구이동통계
8) 서울시 자치구별 이동자 수_시도 내 이동	자치구별 인구이동 특성 반영	통계청 국내인구이동통계
9) 서울시 자치구별 재정자주도	자치구별 재정자주도 반영	통계청 e-지방지표 재정자주도(시도/시/군/구)
10) 서울시 자치구별 1인당_자동차등록대수	자치구별 소득수준 반영	통계청 e-지방지표 1인당 자동차 등록대수(시도/시/군/구)
11) 서울시 자치구별 지가변동률_평균	자치구별 공시지가 수준 반영	서울시 열린데이터 광장
12) 서울시 자치구별 지가변동률_주거용	자치구별 공시지가 수준 반영	서울시 열린데이터 광장
13) 서울시 자치구별 지가변동률_상업용	자치구별 공시지가 수준 반영	서울시 열린데이터 광장
14) 서울시 도로 현황_도로연장	자치구별 도로 환경 수준 반영	서울시 열린데이터 광장
15) 서울시 도로 현황_도로면적	자치구별 도로 환경 수준 반영	서울시 열린데이터 광장
16) 서울시 도로 현황_도로비율	자치구별 도로 환경 수준 반영	서울시 열린데이터 광장
17) 서울시 자치구별 실업급여(미용, 영업 등)	자치구별 실업 흐름 반영	Nowcast (빅데이터활용지표) 서울시 자치구별 실업 급여)
18) 창업기업실태(개인)	2020년 창업기업 실태조사	MDIS (마이크로데이터 통합서비스)

## <활용 데이터>

### 2) [구축한 데이터셋]

변수 종류	변수명	설명	기간
반응변수	target	특정 연도 t번째 분기 행정동별 상권 변화지표	2015년 1/4분기부터 2021년 4/4분기
	house	전년도 행정동별 t번째 분기 세대 수	2014년 1/4분기부터 2020년 4/4분기
	housep	전년도 행정동별 t번째 분기 인구 수	2014년 1/4분기부터 2020년 4/4분기
	ratio	전년도 행정동별 t번째 분기 세대 당 인구 비율	2014년 1/4분기부터 2020년 4/4분기
	elder	전년도 행정동별 t번째 분기 65세 이상 고령자 인구	2014년 1/4분기부터 2020년 4/4분기
	elderatio	전년도 행정동별 t번째 분기 고령자 인구 비율	2014년 1/4분기부터 2020년 4/4분기
	fratio	전년도 행정동별 t번째 분기 여성 인구 비율	2014년 1/4분기부터 2020년 4/4분기
	kratio	전년도 행정동별 t번째 분기 한국인 비율	2014년 1/4분기부터 2020년 4/4분기
	districtnum	자치구 식별번호	2014년 1/4분기부터 2020년 4/4분기
	quarter	분기 t	2014년 1/4분기부터 2020년 4/4분기
설명변수	netshift	전년도 자치구별 t번째 분기 순이동량	2014년 1/4분기부터 2020년 4/4분기
	indistrict	전년도 자치구별 t번째 분기 시도 내 이동량	2014년 1/4분기부터 2020년 4/4분기
	averageprice	전년도 자치구별 t번째 분기 대표 월별(1,4,7,10) 공시지가(평균)	2014년 1/4분기부터 2020년 4/4분기
	residuse	전년도 자치구별 t번째 분기 대표 월별 주거용도 공시지가	2014년 1/4분기부터 2020년 4/4분기
	commerceuse	전년도 자치구별 t번째 분기 대표 월별 상업용도 공시지가	2014년 1/4분기부터 2020년 4/4분기
	rdextend	전년도 자치구별 도로 연장	2014년 1/4분기부터 2020년 4/4분기
	rdarea	전년도 자치구별 도로 면적	2014년 1/4분기부터 2020년 4/4분기
	rdratio	전년도 자치구별 도로 비율	2014년 1/4분기부터 2020년 4/4분기
	publicadm	행정동 식별번호	2014년 1/4분기부터 2020년 4/4분기
	finindep	전년도 자치구별 재정자주도	2014년 1/4분기부터 2020년 4/4분기
	cratio	전년도 자치구별 인당 자동차 등록 대수	2014년 1/4분기부터 2020년 4/4분기

- 본 연구에서는 상권변화지표 예측을 위한 설명변수를 구축하기 위해, 상권 활성화 영향 요인과 관련한 여러 선행연구를 검토. 그 결과, 인구 측면에서는 소비자의 성별, 연령, 배우자의 인구 특성(규모, 밀도 등), 그리고 유동 인구 규모가, 소득·경제의 측면에서는 소비자의 소득수준과 상권의 공시지가가, 접근성의 측면에서는 도로 환경, 교통 거점의 특성 등이 대표적 유의 요인으로 확인됨(\*참고문헌 참조). 이에 해당 유의 요인들에 대응될 수 있는 자료들을 수집함으로써 설명변수를 구성하였으며, 이와 동시에 선행연구와의 차별성을 위해 본 연구에서는 새로운 접근으로서 자치구별 재정자주도 및 인구이동을 설명변수로서 또한 포함시킴.

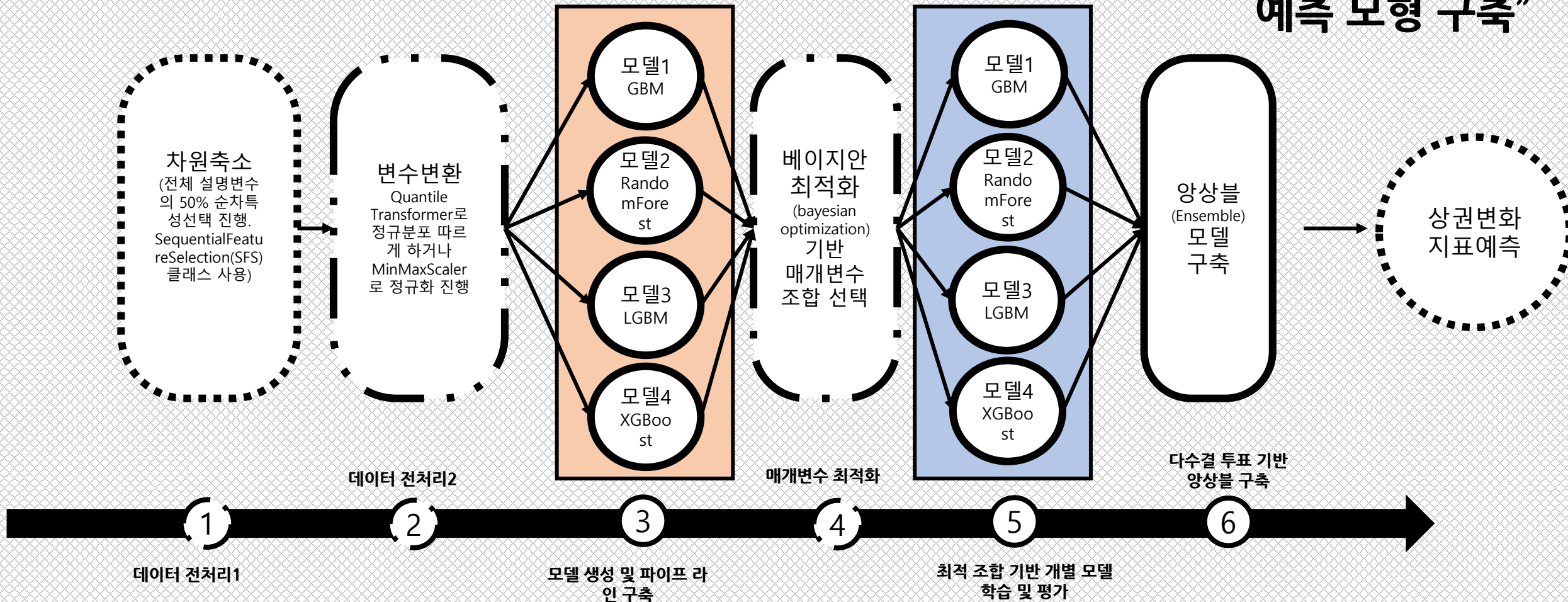
-수집된 데이터 지표는 위의 '1) [선정된 데이터]', 이를 전제로 구축한 설명변수 및 반응변수 데이터셋은 위의 '2)[구축한 데이터셋]'과 같음. 2)의 경우, 본 연구의 주목적인 선제적 상권변화지표 예측을 위해 반응변수와 설명변수 사이 1년 시차를 발생시켰다.



## 2. 연구 설계

### <1단계 전략 프로세스>

“머신러닝 기반  
예측 모형 구축”



먼저 1단계 전략으로서는 상관변화지표(HH,LL,LH,HL)을 예측할 수 있는 머신러닝 기반 모델을 구축하는 것을 목표로 하였다. 해당 프로세스는 위와 같이 요약할 수 있다. 먼저 데이터 전처리 과정에서는 특성선택 기반 차원축소와 변수변환(정규화 등)을 활용하였으며, 모델 생성 과정에서는 의사결정트리기법 기반 모형인 GradientBoostingMachine, RandomForest, LightGBM, XGBoost 모델을 개별 모델로서 사용하였다. 이때, 개별 모델의 최적 매개변수(파라미터) 조합 탐색을 위해서는 효율성을 높이기 위해 베이지안 최적화(평가함수: 5-겹 교차검증 성능)를 활용하여 적절 매개변수를 추출한 다음, 모델 학습을 진행하여 개별 모델의 성능을 증진하였다. 이후 각 모델을 기반으로 한 앙상블 모델을 구축하고 해당 모델의 성능을 개별 모델의 성능과 비교하여 더 나은 접근법을 기반으로 예측 모형을 구축하였다.

## 2. 연구 설계

### \* 베이지안 최적화 예시 화면)

본 연구 RF 최적화 단계 일부

In[23]: B01.maximize(init\_points=20, n\_iter=20)

iter	target	max_depth	n_esti...
1	0.942	26.68	172.0
2	0.8593	10.0	130.2
3	0.9435	15.87	109.2
4	0.9455	17.45	134.6
5	0.9456	2.87	153.9
6	0.9465	26.77	107.5
7	0.9475	18.18	187.1
8	0.8879	11.1	167.0
9	0.9464	16.69	155.9
10	0.9453	15.62	119.8
11	0.9477	42.0	190.8
12	0.9467	22.54	169.2
13	0.9467	45.06	189.5
14	0.9317	13.4	103.9
15	0.9452	16.79	117.8
16	0.9396	13.93	142.1
17	0.9463	48.32	153.3
18	0.9467	37.68	131.6
19	0.9467	37.46	183.5
20	0.8916	10.73	175.0
21	0.949	25.68	137.6

Parameter2: 트리 개수

Parameter1: 트리 최대 깊이

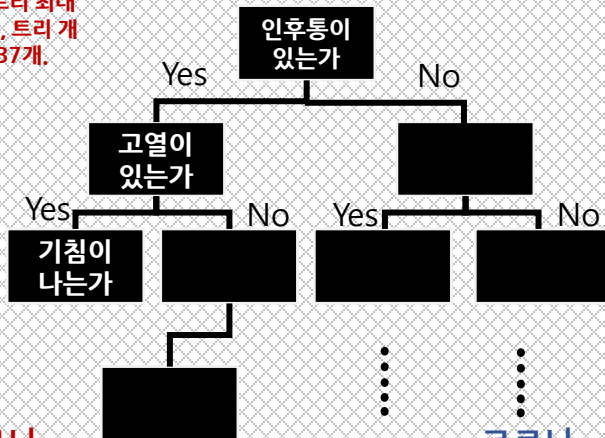
조합 평가 점수 (5-겹 교차검증 정확도)

공간탐색 시행횟수

현재까지 최적 조합: 21번째에 발생, 트리 최대 깊이: 25, 트리 개수: 137개.

### \* 의사 결정트리 예시)

코로나 감염 여부 분류



코로나  
감염자

코로나  
비감염자

## [1단계 전략 활용 개념]

### 베이지안 최적화란?

(Bayesian optimization)

: SMBO 기반 공간 탐색 방식 가운데 하나. 탐색 이력과 기대향상(Expected Improvement)을 전제로 하여 성능이 높은 조합 공간을 집중적으로 탐색하는 방식으로, 최적 매개변수 조합을 찾아내는 접근법. GridSearch(설정 후보공간 모두 탐색) 또는 RandomSearch(후보공간 내 랜덤 조합 탐색)에서 발생하는 비효율성을 극복하는 동시에 탐색이력(\*본 연구에서는 모델 검증을 위해 5겹 교차검증 정확도를 점수로 사용)을 활용함으로써 실제 데이터 양상을 반영한다는 점에서 최적 파라미터 조합을 추출하는데 효율적인 알고리즘으로서 인정되어 있음. 예) TPE 등

### 의사 결정트리 기법이란?

(Decision Tree Classifier)

: 정보이득(Information Gain, 엔트로피, 지니 계수 등이 활용됨)의 최대화를 전제로 하여 특성의 수준 또는 종류에 따라 계층적으로 노드를 분할하는 식으로 논리 체계를 구축하는 방법. 이때 형성되는 논리 구조들이 나무에서 가지가 뻗어 나가는 모습과 유사하여 트리(tree)라 명명된다. 본 연구에서 사용되는 GBM, LGBM, RF, 그리고 XGBoost 모두 여러 의사결정트리를 연결하여 판단을 도출하는 방식의 머신러닝 모델(부스팅 등 기법 적용됨).

### 순차 특성 선택이란?

(Sequential feature selection)

: PCA, LDA, SVD 등과 같이 행렬 분해를 기반으로 특성을 추출해내는 차원축소 방식과는 달리, 연구 초반에 설정된 설명변수 집합 내에서 모델 성능 기여도를 고려하여 정해진 개수만큼의 특성을 순차적으로 선택하는 방식으로 차원축소를 진행하는 전처리 방법. (본 연구에서는 평가 지표로서 RF(트리 100개)를 사용하였으며, 분석과 모형의 복잡도를 줄이기 위해 전체 특성 가운데 50%에 해당 하는 특성을 선택하는 경우를 가정하고 연구를 진행).

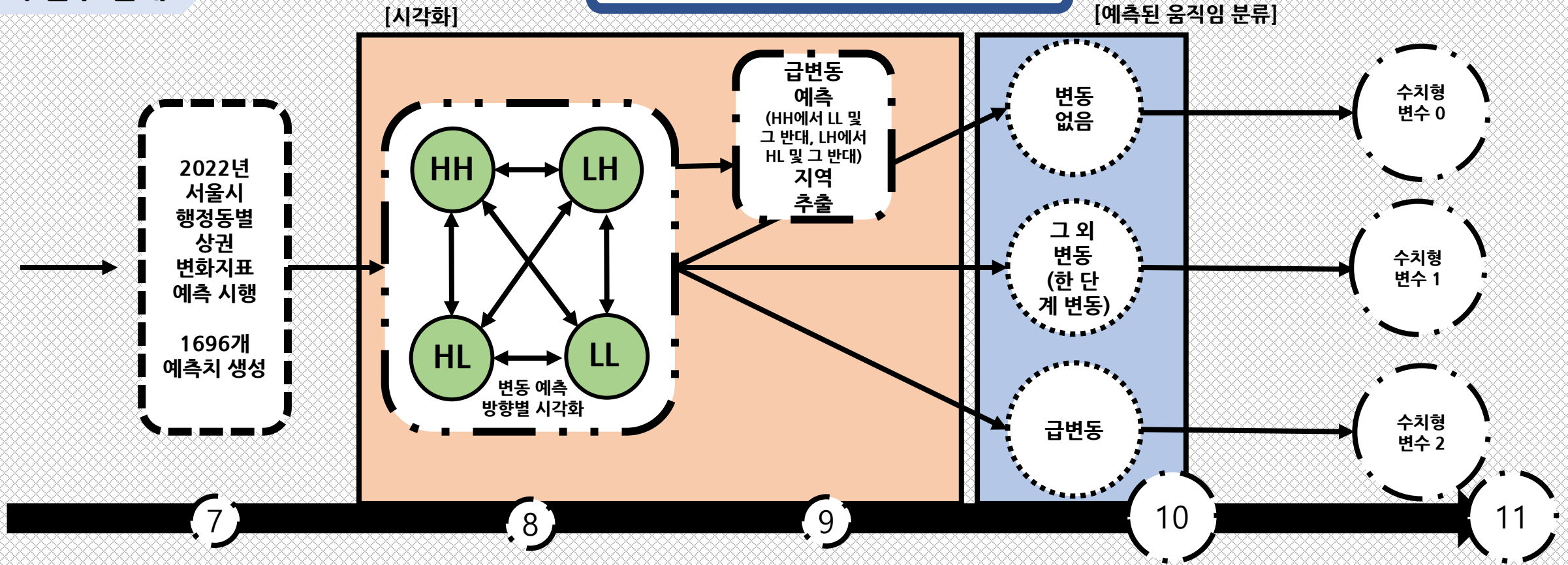
### 다수결 투표 기반 앙상블 기법이란?

(Voting Classifier Model)

: 앙상블 기법 가운데 하나로, 여러 모델을 엮어 각 모델에서 발생하는 각 범주형 자료 예측치를 도출한 후, 각 예측치에 신뢰도를 곱한 다음 범주별로 해당 값들의 총합을 도출하였을 때, 더 높은 값이 도출되는 범주를 최종적인 예측치로 선택하는 방법. 여러 모델을 동시에 고려함으로써 더욱 높은 일반화 성능을 갖출 수 있다는 장점이 있음. 본 연구에서는 ensemble 클래스의 Voting Classifier ("soft")를 사용함으로써 확률 기반 다수결 투표 기법을 시행함.

## 2. 연구 설계

### <2단계 전략 프로세스>



- 2단계 전략 프로세스에서는 먼저 1단계 전략 프로세스를 통해 도출된 모형을 전제로, 2021년 데이터를 활용하여 2022년 상권변화지표예측을 시행. 이에 따라 생성된 총 1696개의 예측치는 직전 연도 동일 행정동 상권변화지표와 비교되었으며, 본 연구자는 그 가운데 2022년 3, 4분기에 전년 동분기 대비 상권 특성 변화가 발생할 것으로 예측되는 행정동들을 선별하여 추출하였다. 이때 상권지표변화 양상 별로 행정동을 분류 및 시각화 함으로써 예비 소상공인의 진입전략 구축을 위한 참고용 기초자료 생성. 또한, 해당 기초자료 기반으로 급변동 예측 지역 추출.

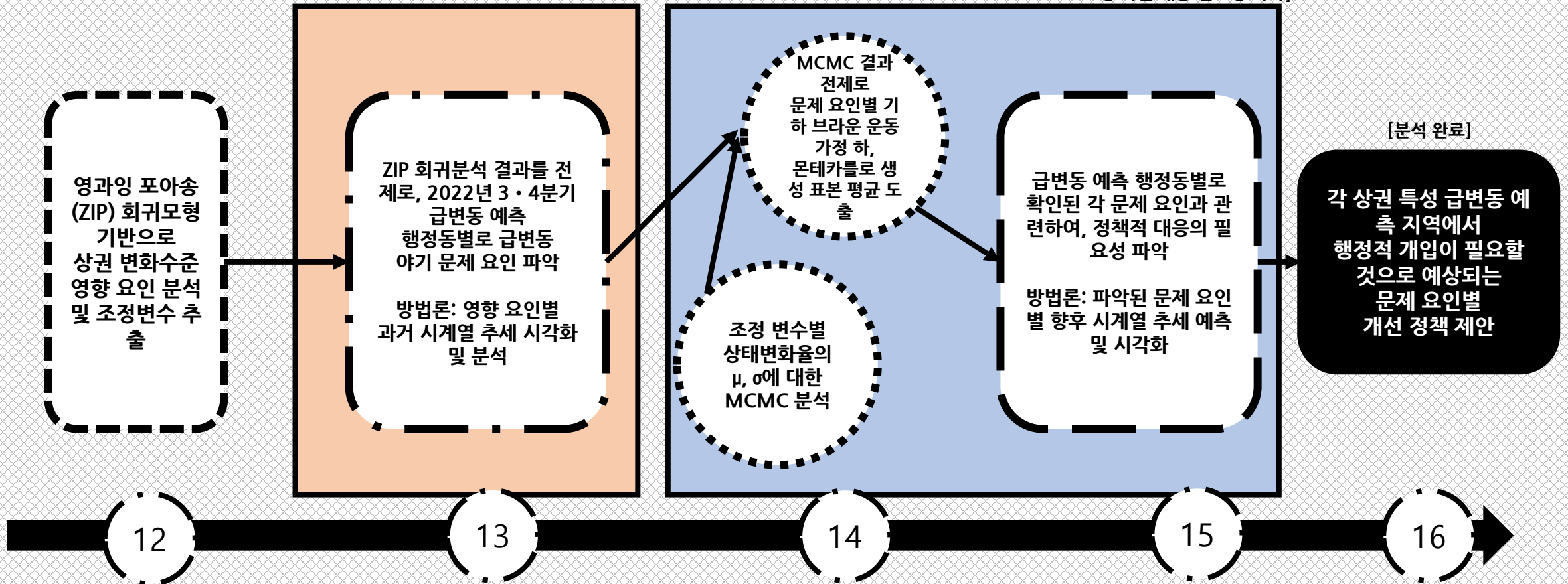
- 한편, 추후 진행할 급변동 지역 영향 요인 분석을 위해 변화지표의 질적 변동 여부뿐만 아니라 양적 변화 수준에도 집중하여 2016년부터 2021년까지의 상권변화지표 데이터에 대하여 변동 수준에 따른 변수를 생성. 이때, 상권의 변화는 특정 범주에서 다른 범주로 이동하는 방식으로 이뤄지는 현상이 아니라, 연속적 변화가 누적된 결과물이라는 점에서(가령 상권 정체(HH)에서 상권 다이내믹(LL)으로의 상권 지표변화는 단순히 범주 간 이동이 아니라, HH에서 LH(상권 확장), 그다음 LH에서 LL 등 단계적 변화를 거쳐 발생하는 변동이라는 점에서) 순서형 범주가 아닌 수치형 변수(특정 기간 내에 단계적 변화가 발생한 횟수)로 정리. ('0번: 지표 변동 없음, 2번: HH에서 LL, LL에서 HH, LH에서 HL, HL에서 LH(본 연구에서는 급변동으로 정의), 1번: 그 외 지표 변동'으로 정의).

## 2. 연구 설계

### <3단계 전략 프로세스>

[급변동 야기 문제 요인 추출]

[행정동별 문제 요인 추세 예측 기반 행정 차원 대응 필요성 파악]



-3단계 프로세스에서는 추출된 급변동 예측 행정 구역들 각각에서의 급변동 원인을 분석함으로써 정책적 개입 방안을 제안. 이에 먼저 기존 데이터셋에서 target 변수를 앞서 생성한 수치형 변수로 대체한 후, 생성된 새로운 데이터셋(\*이때 기간은 2016년에서 2021년으로 축소)을 전제로 변동 수준에 유의한 영향을 미치는 요인에 대한 회귀분석을 시행함. 이때, 본 연구에서는 회귀분석의 반응변수가 되는 데이터가 이산형 변수라는 점, 상권 특성의 경우 변동이 발생하는 경우보다는 현상 유지가 발생하는 빈도가 압도적으로 높음에 따라 수치형 변수 0이 기타 변수에 비하여 과하게 발생하는 현상이 존재한다는 점, 두 문제점을 해결할 수 있는 회귀모형으로서 ZIP (Zero -Inflated Poisson) 모형 (영과잉 포아송 회귀모형)을 활용하여 분석을 진행(R의 pscl 패키지 이용).

-다음으로, 변수별 유의성 및 전체 회귀모형의 유의성(우도비 검정 활용)을 유의 수준 0.05에서 확인한 후 정리함으로써 변화 수준 영향 요인 및 각 요인의 영향 방향성을 확인하고 영향 요인 가운데 정책적 개입이 이뤄지는 것이 가능할 것으로 파악되는 요인을 조정 변수로서 추출.

-마지막으로 조정변수 가운데 각 급변동 지역에서 어떠한 요인의 움직임이 문제가 되었던 것인지를 확인하고 해당 문제의 지속 여부를 파악하기 위해 영향 요인별 과거 시계열 시각화(추세 분석)와 향후 시계열 예측을 활용함. 이때, 향후 시계열 예측치는 MCMC(MAP, MH-알고리즘 활용)를 통해 추정된 영향 요인별 상태변화율의  $\mu, \sigma$ , 그리고 기하 브라운 운동 가정을 기반으로 한 몬테카를로 생성 표본 300개(300개 시계열)의 평균치를 통해 도출하였음. 이후 해당 과정을 통해 파악한 정책적 개입의 필요성과 실업급여지급 변화 추세 예측을 통해 파악한 노동시장 흐름, 그리고 MDIS의 2020 창업기업실태 조사를 바탕으로 각 급변동 예측 행정동별로 최종적인 정책적 제안 진행.

### 영과잉 포아송 회귀모형이란?

(Zero-Inflated Poisson Regression Model)

\* ZIP 모형이란 0 이상의 정수로 구성된 영과잉 데이터에 대하여 균형 있는 확률 모델을 상정하기 위해 고안된 회귀 모델로, 혼합형 분포에 대한 확률함수를 가정한 다음 0인 부분과 가산 부분을 분리한 후 반응변수에 영향을 미칠 것으로 예상되는 잠재변수들의 집합을 각 부분의 모수와 연결함으로써 서로 다른 조건부 확률  $P(Y|X)$ 을 도출하는 구조를 형성한다(김주영, 이성건, 2008). 이후 ZIP에서는 해당 구조를 기반으로 최대우도추정을 시행하여 각 잠재변수에 대한 계수를 추정하며, 영과잉 부분과 그렇지 않은 부분에 대응되는 잠재변수들의 유의성 여부를 제시하는 방식으로 영향 요인을 추출한다. 이때 영과잉 확률  $p$ 를 위한 연결함수로서는 로짓 모형 또는 프로빗(Probit) 모형, complementary log-log 모형 등이 활용될 수 있는데, 본 연구에서는 오차가 정규분포를 따른다는 가정하에 프로빗 모형을 활용하여 분석을 진행하였다.

\* 이때, 본 연구에서  $X, Z$ 를 구성하는 변수들을 선택 시, 다중공선성 발생 가능성을 최소화하기 위해 의미가 다소 중복되는 특성들의 경우 일부만을 선택하는 방식으로 잠재변수들을 가정한 것이며, 그 결과  $X$ 를 구성하는 잠재변수로서는 행정동별 인구수, 자치구별 상업용 지구 공시지가 변동률, 행정동별 65세 이상 고령자 비율, 자치구별 시도 내 이동 수, 자치구별 도로 비율, 그리고 행정동별 여성 인구 비율이 선택되었으며,  $Z$ 를 구성하는 잠재변수로는 행정동별 세대 수, 세대 당 인구수, 65세 이상 고령자 수, 자치구별 시도 내 이동 수, 상업용 지구 지가변동률, 행정동별 여성 인구 비율, 자치구별 1인당 자동차 등록대수, 도로면적, 그리고 재정자주도가 선택되었다.

### [ZIP 회귀모형 수식]

*ZIP(Zero-inflated Poisson) Model*

(\*Y: 변동 정도,  $x = \langle x_1, x_2, \dots, x_k \rangle$ : k개 설명 변수 벡터,  $p$ : 영과잉부분 발생 확률)

$$ZIP \text{ Mixed pmf } f(y; p, \lambda) = p \times h_1(y) + (1-p) \times h_2(y; \lambda)$$

$$h_1: \begin{cases} 1 & (y=0) \\ 0 & (else) \end{cases}, h_2: \text{Poisson pmf}$$

$$\therefore ZIP \text{ probability } P(Y=y|x) = \begin{cases} p + (1-p)e^{-\lambda} & (y=0) \\ (1-p) \frac{e^{-\lambda} \lambda^y}{y!} & (otherwise) \end{cases}$$

1) 가산부분, 영과잉부분 공통

반응함수: 지수 함수

$$E(Y|x) = \mu = \exp(X\beta) \text{ (반응함수)}$$

$$\therefore \ln \mu = \ln \mu = X\beta \text{ (} \because E(Y) = \mu \text{)} (X: \text{설명 변수}, \beta: \text{계수})$$

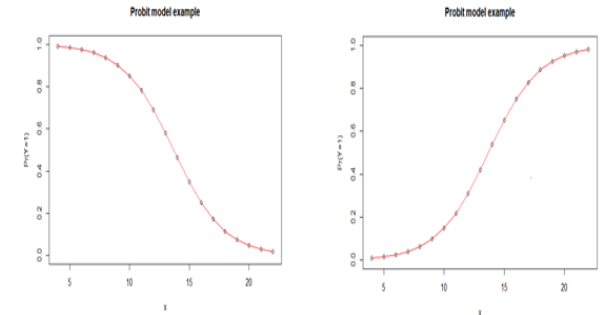
2) 영과잉부분 ( $y=0$ )

발생 확률  $p$ 와 설명 변수 사이 연결함수: 프로빗 모형

$$\therefore \Phi^{-1}(p) = Z'\gamma \text{ (} Z: \text{설명 변수}, \gamma: \text{계수})$$

(1)

### [프로빗 모형 예시]



좌: 프로빗 모형의 계수  $\beta_1$ 이 음수,

우: 프로빗 모형의 계수  $\beta_1$ 이 양수

\* 프로빗 평균반응 함수는 범주형 자료 가운데 이항 반응변수가 존재하는 상황에서 활용되는 대표적 시그모이드 반응함수로, 표준정규분포의 cdf인  $\Phi$ 를 활용해 0과 1 사이에 반응 함수값이 존재하도록 하여 특정 범주(1,0)의 발생 가능성을 표현한다. ( $\Phi(p) = \Pr(Y=1) = Z'\gamma$ )



## 2. 연구 설계

### 기하 브라운 운동이란?

(Geometric Brownian Motion)

: 기하 브라운 운동이란 브라운 운동을 기반으로 한 확률미분방정식과 이토의 보조정리(Ito's Lemma)를 활용할 때 특정 시점의 대상이 취할 상태에 대한 확률과정을 말하는 것(Jun Mo, S., 2011). 여기서 브라운 운동이란 꽃가루 등의 입자가 유체에서 확산되는 양상에 관한 Robert Brown의 연구(1827)를 시작점으로 한 개념으로, 정규분포와 확률과정을 전제로 어떠한 대상과 관련하여 특정 상태 또는 값이 발생할 가능성을 제시한다. 브라운 운동 아래서는 특정 시점부터 다음 시점까지의 경과 기간을  $s$ 라 정의 시, 어떠한 대상의 위치 변동 수준  $B(s+t)-B(t)$ 의 평균이 0, 표준편차가  $\sigma\sqrt{t}$  ( $=\sqrt{t}$ )인 정규분포를 따른다. ( $B(0)=0$ )

### MCMC 과정이란?

(Markov Chain Monte Carlo process)

:MCMC 과정이란 베이지 정리, 마코프 과정, 그리고 몬테카를로 적분을 전제로 사후분포에서 직접 추출한 표본과 비교불가한 표본을 반환함으로써 특정 모수의 분포에 대한 정보를 취득할 수 있는 통계적 접근법을 의미한다. 구체적으로, MCMC에서는 아래와 같은 detailed balance equation을 충족시키며 후보 모수값을 정해진 횟수만큼 생성하는 한편 수용-기각 알고리즘에 따라 해당 모수값을 수용 또는 기각함으로써 표본을 생성하며, 대표적 수용-기각 알고리즘으로는 깁스 샘플러, HMC sampler, MH(Metropolis Hastings)-알고리즘 등이 활용된다. 본 연구에서는 MH 알고리즘을 활용하여 연구를 진행했다. (pymc3의 sample 함수 사용)

$\pi_i P_{ij} = \pi_j P_{ji}$  \* P: 마코프 연쇄,  $\pi: \pi(\pi^{(s)} = [P(X_s = 1), \dots, P(X_s = r)])$

[Detailed balance equation]

## [3단계 전략 활용 개념]

### 몬테카를로 생성이란?

(Monte Carlo)

: 몬테카를로 생성은 균등분포, 정규분포 생성자 등을 기반으로 특정 분포나 표본에서 관측값을 인위적으로 만들어내는 통계적 방법으로서, 본 연구에서는 정규분포 관측값 생성자로서 R의 rnorm() 함수를 활용하였다. (해당 함수의 활용은 기하 브라운 운동의 수식에서  $B_t$  값 생성 역할 수행)

### [기하 브라운 운동 수식]

기하브라운 운동과정

$t$ 시점 상태(가격 등)를  $X_t$ 로 정의 시,  $X_t = X_0 e^{(\mu - 0.5\sigma^2)t + \sigma B_t}$

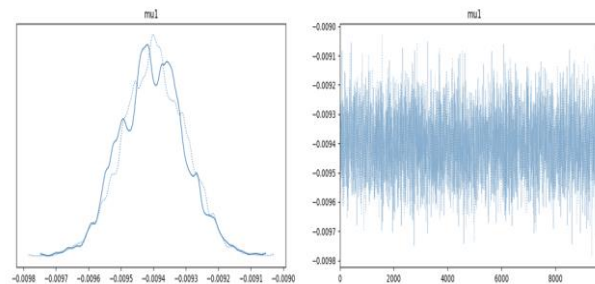
즉,  $X_{t+1} = X_t e^{(\mu - 0.5\sigma^2) + \sigma B_t}$  ( $\mu$ : 상태변화율 평균,  $\sigma$ : 상태변화율 표준편차)

전제 1)  $dX_t = \mu(t, X_t)dt + \sigma(t, X_t)dB_t$  (확률 미분방정식)의 성립

전제 2)  $d \ln X_t = \frac{dX_t}{X_t} - 0.5 \times \frac{(dX_t)^2}{(X_t)^2}$  (Ito's Lemma 성립)

\*여기서  $\mu, \sigma$ 는 mcmc 추정을 통해 도출

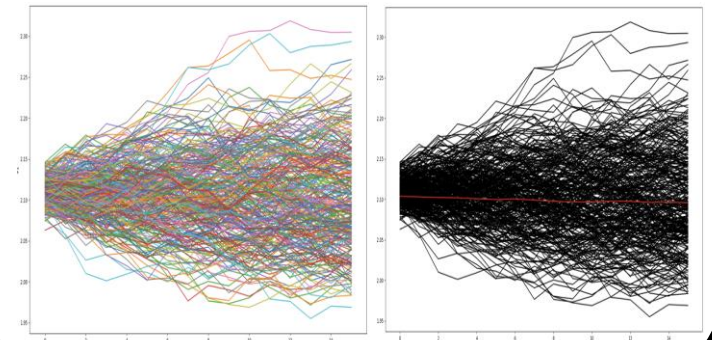
### [MCMC 예시] 휘경1동 동대문구 재정자주 도 상태변화율의 $\mu$ 사후 분포 및 trace]



### [몬테카를로 생성 예시]

동대문구 휘경1동 세대 당 인구 예측

좌: 300개 시계열, 우: 평균치(빨간선)]



### 3. 분석 결과

### 3. 분석 결과

## <1단계 전략 분석 결과>

## 개별 모델 구축 및 결과

```
[Parallel(n_jobs=2)]: Using backend ThreadingBackend with 2 concurrent workers.
[Parallel(n_jobs=2)]: Done 46 tasks | elapsed: 0.3s
[Parallel(n_jobs=2)]: Done 46 tasks | elapsed: 0.3s
[Parallel(n_jobs=2)]: Done 100 out of 100 | elapsed: 0.8s finished
[Parallel(n_jobs=2)]: Using backend ThreadingBackend with 2 concurrent workers.
[Parallel(n_jobs=2)]: Done 46 tasks | elapsed: 0.0s
[Parallel(n_jobs=2)]: Done 100 out of 100 | elapsed: 0.7s finished
[Parallel(n_jobs=2)]: Using backend ThreadingBackend with 2 concurrent workers.
[Parallel(n_jobs=2)]: Done 100 out of 100 | elapsed: 0.0s finished
[Parallel(n_jobs=2)]: Done 46 tasks | elapsed: 0.0s
[Parallel(n_jobs=2)]: Done 100 out of 100 | elapsed: 0.0s finished
[Parallel(n_jobs=2)]: Using backend ThreadingBackend with 2 concurrent workers.
[Parallel(n_jobs=2)]: Done 46 tasks | elapsed: 0.2s
[Parallel(n_jobs=2)]: Done 100 out of 100 | elapsed: 0.0s finished
[Parallel(n_jobs=2)]: Using backend ThreadingBackend with 2 concurrent workers.
[Parallel(n_jobs=2)]: Done 46 tasks | elapsed: 0.0s
[Parallel(n_jobs=2)]: Done 100 out of 100 | elapsed: 0.0s finished
Out[19]:
SequentialFeatureSelector(estimator=RandomForestClassifier(criterion='entropy',
n_jobs=2,
random_state=44,
verbose=1),
n_features_to_select=10, n_jobs=2)
In[20]: mask = SFS.support_...
...: mask
...: dataframe.columns[0:20][mask]
...:
Out[20]:
Index(['house', 'housep', 'ratio', 'elder', 'fratio', 'districtnum',
'rnextend', 'rdarea', 'rdratio', 'publicadm'],
dtype='object')
```

-실행화면

**LGBM**  
베이지안 최적화 결과

**RF**  
베이지안 최적화 결과

**XGBoost**  
베이지안 최적화 결과

**GBM**  
베이지안 최적화 결과

**특성선택 결과(10개):** 행정동별 세대 수, 인구 수, 세대 당 인구 수, 65세 이상 고령 인구, 여성 인구 비율, 자치구 식별 번호, 자치구별 도로 연장, 도로 면적, 도로 비율, 행정동 식별 번호 **변수 선택됨**

```
19 | 0.9421 | 647.5 |
20 | 0.9444 | 872.6 |
21 | 0.9451 | 3.282e+0 |
22 | 0.9452 | 3.767e+0 |
23 | 0.9453 | 3.999e+0 |
24 | 0.9451 | 2.497e+0 |
25 | 0.9447 | 1.983e+0 |
26 | 0.945 | 3.104e+0 |
27 | 0.945 | 1.108e+0 |
28 | 0.9453 | 5.60e+0 |
29 | 0.9425 | 324.9 |
30 | 0.9452 | 3.888e+0 |
31 | 0.9451 | 3.402e+0 |
32 | 0.945 | 2.395e+0 |
33 | 0.945 | 2.601e+0 |
34 | 0.9447 | 1.875e+0 |
35 | 0.945 | 2.092e+0 |
36 | 0.9449 | 3.01e+03 |
37 | 0.9446 | 1.54e+03 |
38 | 0.9451 | 3.195e+0 |
39 | 0.9452 | 1.189e+0 |
40 | 0.9449 | 1.017e+0 |
=====
Out[11]: {'n_estimators': 3999, 4638543141646}
In[12]:
```

```
19:58:44] WARNING: C:\Users\Administrator\workspace\sguon\win64
40 | 0.9462 | 0.8646 | 34.91 | 195.4 |
=====
Out[14]:
{'learning_rate': 0.49629263711353,
'max_depth': 12, 134501804483214,
'n_estimators': 336, 11764862380194}
```

-실행화면

<표1> 모델별 성능 평가

모델 종류	Training Accuracy	Mean Cross Validation Score (10-fold)	Recall Score(macro)	Test Data F1-score(macro)	Test Data Accuracy
RandomForest	100%	94.82%	93.36%	93.65%	94.8%
GBM	100%	95.05%	93.25%	93.50%	94.72%
LGBM	100%	94.81%	93.62%	93.82%	95.00%
XGBoost	100%	94.81%	93.04%	93.4%	94.69%

차원 축소를 위한 10개(50%)의 특성 선택 결과, house, housep, ratio, elder, fratio, districtnum, rnextend, rdarea, rdratio, 그리고 publicadm이 선택되었다. 이에 해당 특성을 기반으로 설명변수 집합을 재설정 한 다음, 파이프라인을 활용하여 RF, GBM, XGBoost에서는 Quantiletransformer, LGBM에서는 MinMaxScaler을 기반으로 특성을 전처리한 후 모델별 파라미터에 대한 베이지안 최적화를 진행하였다. 그 결과 RF 파라미터의 최적 조합으로서는 (트리 개수: 155, 트리 최대 깊이: 26), GBM에서는 (학습률: 0.7, 트리 개수:91, 트리 최대 깊이:10), LGBM에서는 (트리 개수: 3999), 그리고 XGBoost에서는 (학습률: 0.497, 트리 개수: 136, 트리 최대 깊이:12)가 선택되었다. 이후 해당 조합을 기반으로 각 모델을 학습하고 평가를 진행한 결과, 학습 데이터에 대한 정확도 및 교차검증 결과, 그리고 평가 데이터에 대한 모델의 정확도는 <표 1>과 같이 도출되었다.

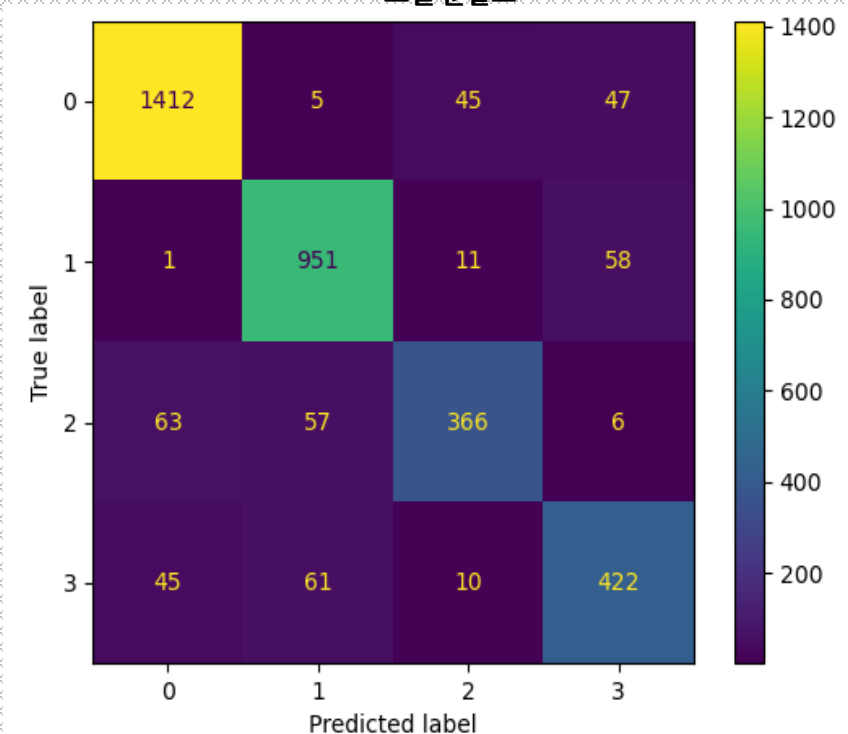


### 3. 분석 결과

## <1단계 전략 분석 결과>

### 양상블 모델 구축 및 결과

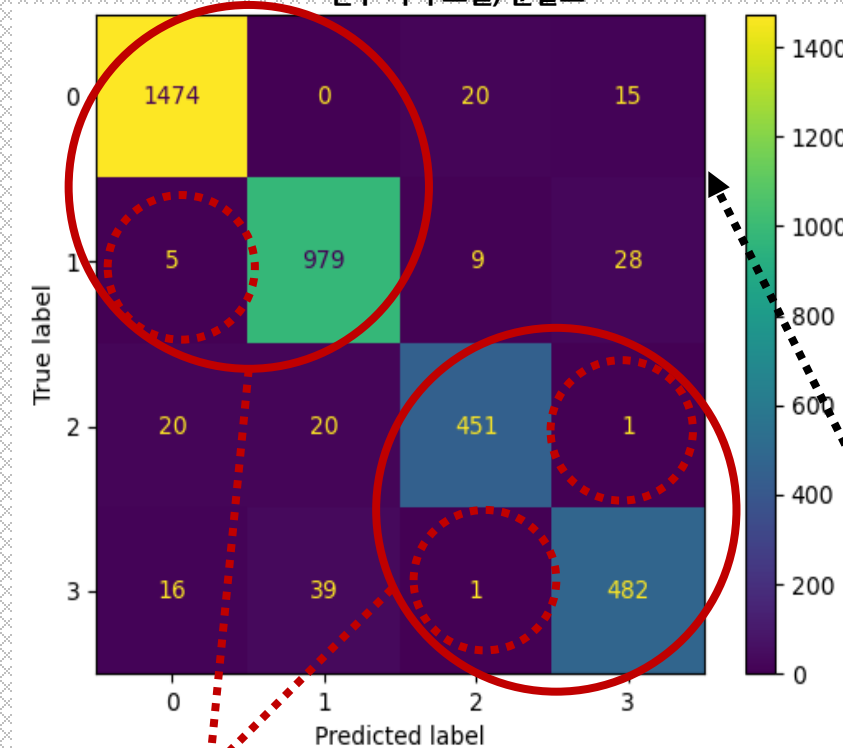
(테스트 데이터 예측) 행정구역만을 설명변수로 하는 양상블 모델 분할표



-실행화면

{\* 'LL':0, 'HH':1, 'HL':2, 'LH':3}

\*(테스트 데이터 예측) 순차 특성 선택(10개) 기반 양상블 모델(본 연구 기저 모델) 분할표



-실행화면

양 극단 예측(LL과 HH, HL과 LH)만 개별적으로 고려 시, 오류 비율 각각 약 0.002, 0.0021.(=정확도 각각 99.8%, 99.79%)

∴ 구축한 양상블 모델을 연구 기저 모델로 채택, 예측 진행.

정확도: **95.11%**  
정밀도: **94.25%**  
재현율: **93.71%**  
F1-score: **93.97%**

\*산출 기준: macro

다음으로, 활용된 네 개의 모델을 기반으로 양상블(다수결 투표 기반)을 구성한 후 학습 및 평가를 진행한 결과, training accuracy는 100%, mean cross validation score는 95.1%, test score(accuracy)는 약 95.11%(<그림 4>), 정밀도(macro)는 94.25%, 재현율은 93.71%, f1-score는 93.97%로 도출되었으며, 예측 성능이 상당히 높은 수준임이 확인되었다. 또한, 양상블 모델은 개별 모델의 성능에 비하여 test score, 재현율, f1-score에서 상당 수준 비교 우위가 있음이 확인되었다. 한편, 행정동의 특성 자체에 높게 의존하는 상권의 특성상 행정구역 외 변수를 고려해야 할 실효성이 확인될 필요가 존재함에 따라 행정구역만을 설명변수로 하는 양상블 모델을 사용한 결과, 정확도는 88.51%, 재현율은 84.89%, f1 score로는 85.50%가 도출되었으며, 이에 따라 본 연구에서 활용되는 SFS 특성 선택 기반 양상블 모델은 실효성과 정확도 두 측면 모두에서 매우 유의함을 확인하였다. 따라서 양상블 모델은 상권지표예측 기저 모델로 채택되었다.

### 3. 분석 결과

## <2단계 전략 분석 결과>

예측 실행 기반 소상공인 지원 기초자료 생성

2022년 행정동별  
예측치 생성(자료 화면)

<예측치 기반 상권변화지표 변동 분할표(소상공인 진입전략 기초자료) 생성>

	HH_2022	HL_2022	LH_2022	LL_2022		HH_2022	HL_2022	LH_2022	LL_2022
HH_2021		개포2동(3분기), 대방동(3), 여의동(3)		서초3동(3)	HH_2021		개포2동(4분기), 마천2동(4)	하계1동(4), 금호4가동(4), 신당5동(4)	서초3동(4), 휘경1동(4)
HL_2021	개포4동(3), 휘경1동(3), 용문동(3)			등촌2동(3), 자양2동(3), 상계9동(3), 신정1동(3), 망우본동(3), 창4동(3)	HL_2021	개포4동(4), 신길5동(4), 응암3동(4)			등촌2동(4), 자양2동(4), 상계9동(4), 신정1동(4), 망우본동(4)
LH_2021	수궁동(3), 정릉1동(3)	신길4동(3)		상일동(3), 신정3동, 양평1동(3), 응암1동(3), 일원본동(3)	LH_2021	수궁동(4), 전농1동(4), 홍은1동(4)	신길4동(4)		상일동(4), 신정3동, 양평1동(4), 응암1동(4), 명일2동(4), 번2동(4)
LL_2021		신길7동(3), 가양3동(3), 대림3동(3)	능동(3)		LL_2021		신길7동(4)	능동(4)	

### 분할표(기초자료) 해석 방법

예) 신길7동(3,4분기): LL(2021)에서 HL(2022).  
-의미: 신길7동의 상권은 전년도에는 진출입이 매우 빈번, 즉 상권 역동성이 높았으나, 다가오는 3,4분기에는 이와 달리 상권이 축소되는 양상이 발생할 것으로 예측됨.

급변동 예측 지역: 영등포구 신길4동(급락), 동대문구 휘경1동(급등), 서초구 서초3동(급등)

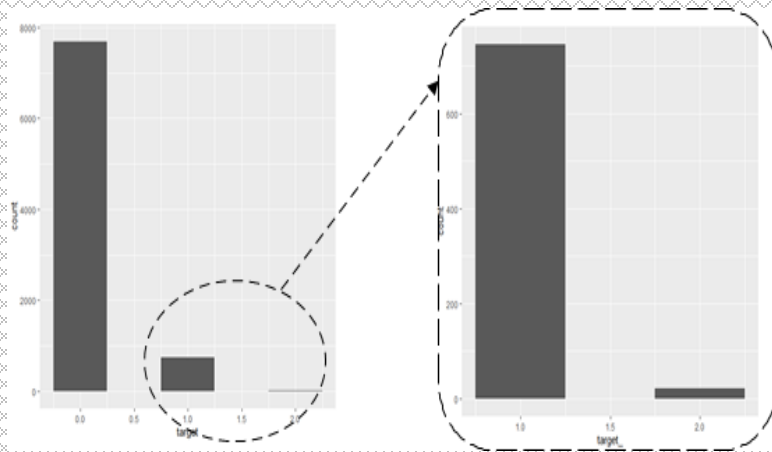
양상별 모델을 기반으로 지표 3, 4분기 예측을 수행한 결과, 전년 동분기 대비 변화가 예측되는 행정동의 지표변동 방향 정보를 시각화 시 위의 분할표가 도출됨(3분기: 25개 행정동, 4분기: 27개 행정동). 예측 결과, 자치구 기준 접근 시 3분기에는 영등포구(5개), 강남구(3개), 그 외 자치구 순으로 지표변동 발생 예측 행정동의 개수가 많았음. 4분기의 경우에는 영등포구(4개)를 제외한 나머지 자치구의 경우 대부분 2개 또는 1개 행정동에서 지표변동이 예측됨. 한편 행정동 기준으로 접근 시, 3, 4분기 공통으로 변화가 예측되는 행정동으로는 강남구 개포2동, 강남구 개포4동, 강동구 상일동, 강서구 등촌2동, 광진구 능동, 광진구 자양2동, 구로구 수궁동, 노원구 상계9동, 동대문구 휘경1동, 서초구 서초3동, 양천구 신정1동, 양천구 신정3동, 영등포구 신길4동, 영등포구 신길7동, 영등포구 양평1동, 은평구 응암1동, 중랑구 망우본동이 존재. 이때 지표상 극단적 변동이 보이는 행정동으로는 서초3동(3,4)(HH에서 LL), 신길4동(3,4)(LH에서 HL), 휘경1동(4)(HH에서 LL)이 존재함이 파악됨에 따라 세 행정구역을 급변동 예측 행정동으로 분류하고 서초3동, 휘경1동은 급등지구로, 신길4동은 급락지구로 정의하여 연구를 재개. 다음으로, 변동 수준 영향 요인에 대한 회귀분석 반응변수를 구축하기 위해 2016년부터 2021년 사이 전년 동분기 대비 상권지표 변동 정도에 따른 수치형 변수(0, 1, 2)를 생성.

### 3. 분석 결과

## <3단계 전략 분석 결과>

### ZIP 회귀분석 결과 (변화수준 영향 요인 도출)

<2단계에서 진행한 반응변수 생성 결과>



(0: 7686개(약 91%), 1: 746개(약 8.8%), 2: 21개(약 0.2%). 반응변수에서 **영과잉 발생 확인**).

먼저 영과잉 부분의 p에 영향을 미치는 유의 요인(Z)으로서는 **house, ratio, elder, indistrict, fratio, rdarea, finindep**가 확인되었으며, 포아송 모형의  $\lambda$ 에 영향을 미치는 유의 요인(X)으로는 **housep, elderatio, indistrict, fratio**가 확인되었다. 한편, 모형 자체의 유의성 측면에서는 우도비 검정(축소 모형으로서의 Null 모형과 완전 모형으로서의 ZIP 회귀모형에 대한 우도비검정)을 시행한 결과,  $p\text{-value} = 2.685875e-28$  ( $< 0.05$ ,  $df=17$ )이 도출되었으며, 귀무가설( $H_0: \forall \beta_i = 0$  ( $i = 1, 2, \dots$ ))이 기각됨에 따라 해당 ZIP 모형은 유의수준 0.05에서 유의하다고 말할 수 있음이 확인됨.

Count model coefficients ( $\lambda$  관련 X,  $\beta$ 분석)

$\hat{\lambda}$	Estimate	S.E.	z-value	P-value	
(Intercept)	2.744316	1.31150	2.093	0.03634	*
housep	-0.014426	0.004433	-3.254	0.00114	**
commerceuse	-0.186226	0.195136	-0.954	0.33991	
elderatio	-6.066137	1.401696	-4.328	1.51e-05	***
indistrict	-0.037882	0.017099	-2.215	0.02673	*
rdratio	-0.020063	0.012349	-1.625	0.10422	
fratio	-6.94343	2.405519	-2.533	0.01129	*

Zero-inflation model coefficients (p 관련 Z,  $\gamma$ 분석)

$\hat{p}$	Estimate	S.E.	z-value	P-value	
(Intercept)	25.1897	8.2101	3.068	0.002154	**
house	1.2825	0.4249	3.018	0.002545	**
ratio	-36.7497	11.3788	-3.230	0.001239	**
elder	-9.6241	3.1143	-3.090	0.002000	**
indistrict	1.0844	0.4224	2.567	0.010247	*
commerceuse	-0.0669	2.4905	-0.244	0.807472	
fratio	-99.9589	28.0297	-3.566	0.000362	***
cratio	-8.5820	10.1077	-0.849	0.395849	
rdarea	7.4201	2.5868	2.868	0.004125	**
finindep	139.0653	41.3217	3.365	0.000764	***

\*:  $p < .05$ , \*\*:  $p < .01$ , \*\*\*:  $p < .001$

Number of iterations in BFGS optimization: 113, ZIP Log-likelihood: -2587 (df=17).

영과잉 부분 유의 영향 요인: 행정동별 세대 수, 세대 당 인구 수, 고령 인구 수, 자치구별 시도 내 인구이동 수, 행정동별 여성 인구 비율, 자치구별 도로 면적, 자치구별 재정자주도  
포아송 모형 부분 유의 영향 요인: 행정동별 인구 수, 고령 인구 비율, 시도 내 이동, 여성 인구 비율

### 3. 분석 결과

## <3단계 전략 분석 결과>

ZIP 회귀분석 결과  
(변화수준 영향 요인 도출)

Count model coefficients ( $\lambda$  관련 X,  $\beta$  분석)

$\hat{\lambda}$	Estimate	S.E.	z-value	P-value	
(Intercept)	2.744316	1.31150	2.093	0.03634	*
housep	-0.014426	0.004433	-3.254	0.00114	**
commerceuse	-0.186226	0.195136	-0.954	0.33991	
elderatio	-6.066137	1.401696	-4.328	1.51e-05	***
indistrict	-0.037882	0.017099	-2.215	0.02673	*
rdratio	-0.020063	0.012349	-1.625	0.10422	
fratio	-6.94343	2.405519	-2.533	0.01129	*

Zero-inflation model coefficients ( $p$  관련 Z,  $\gamma$  분석)

$\hat{p}$	Estimate	S.E.	z-value	P-value	
(Intercept)	25.1897	8.2101	3.068	0.002154	**
house	1.2825	0.4249	3.018	0.002545	**
ratio	-36.7497	11.3788	-3.230	0.001239	**
elder	-9.6241	3.1143	-3.090	0.002000	**
indistrict	1.0844	0.4224	2.567	0.010247	*
commerceuse	-0.0669	2.4905	-0.244	0.807472	
fratio	-99.9589	28.0297	-3.566	0.000362	***
cratio	-8.5820	10.1077	-0.849	0.395849	
rdarea	7.4201	2.5868	2.868	0.004125	**
finindep	139.0653	41.3217	3.365	0.000764	***

\*:  $p < .05$ , \*\*:  $p < .01$ , \*\*\*:  $p < .001$

Number of iterations in BFGS optimization: 113, ZIP Log-likelihood: -2587 (df=17).

“본 연구에서는 확인된 영향 요인 가운데 정책적 지원 또는 개입의 대상이 될 수 있을 것으로 파악되는 요소인 **행정동별 고령자 비율, 세대당 인구 수, 시도 내 이동 수, 재정자주도, 그리고 도로면적**을 **정책적 조정 변수**로 선별”

종합적인 차원에서 정리 시, 잠재변수 가운데 상권 변동 수준에 영향을 미치는 영향 요인으로서는 크게 **“행정동별 세대당 인구수, 행정동별 고령 인구 비율, 자치구별 시도 내 이동 수, 자치구별 재정자주도, 자치구별 도로면적, 그리고 행정동별 여성 인구 비율”**을 상정할 수 있음 (전체 유의 요인 가운데 중복성이 높은 요인들은 대표 요인 하나를 선택하여 고려)

유의 변수 영향 방향성

[영과잉 부분 개별 변수의 측면]

세대 수 ↑, 세대 당 인구 비율 ↓, 65세 이상 고령자 인구 ↓, 시도 내 인구이동 ↑, 여성 인구비율이 ↓, 도로면적이 ↑, 재정자주도가 ↑  
→ 변화 자체가 발생하지 않을 확률  $p$  ↑

[포아송 모형 부분(변화 존재 상태) 개별 변수의 측면]

행정동 인구 ↓, 고령자 비율 ↓, 시도 내 인구이동 ↓, 여성 인구 비율 ↓ → 단계적 변화 발생 횟수 ↑

“변화 자체의 발생 측면에서는 **세대 수 급락, 고령자 인구의 급증, 인구이동의 축소, 재정자주도의 감소** 등이 변화의 존재 가능성을 증가시키는 문제로 상정될 수 있음이 확인되었으며, 변화의 존재를 가정할 때 양적인 수준에서는 **행정동 총인구의 감소, 고령 인구 비율의 감소, 인구이동의 축소** 등이 변동의 정도를 더욱 심화하는 문제로 상정될 수 있음이 확인되었다.”

### 3. 분석 결과

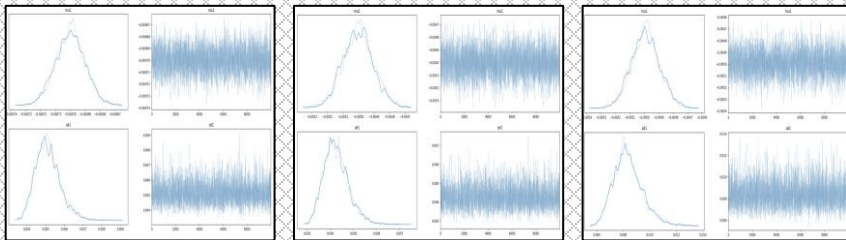
## <3단계 전략 분석 결과>

### 조정 변수별 MCMC 결과

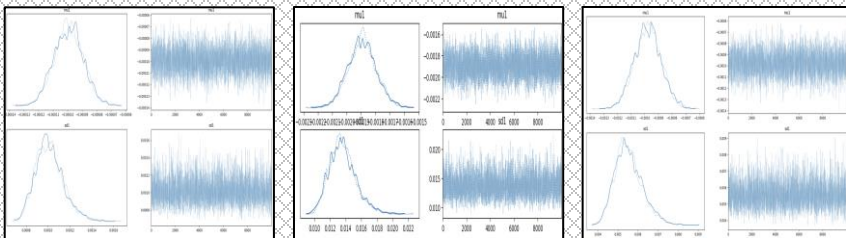
#### <표본 평균 기반 모수별( $\mu, \sigma$ ) 추정치 결과>

	동대문구 휘경1동	서초구 서초 3동	영등포구 신길4동
시도 내 이동( $\hat{\mu}$ )	-0.0079	-0.0083	-0.0055
시도 내 이동( $\hat{\sigma}$ )	0.1385	0.2661	0.2018
재정자주도( $\hat{\mu}$ )	-0.0094	-0.001	-0.01
재정자주도( $\hat{\sigma}$ )	0.292	0.0558	0.0355
고령 인구 비율( $\hat{\mu}$ )	0.008	0.0087	0.0074
고령 인구 비율( $\hat{\sigma}$ )	0.0055	0.0273	0.0071
세대 당 인구 수( $\hat{\mu}$ )	-0.005	-0.005	-0.0007
세대 당 인구 수( $\hat{\sigma}$ )	0.0043	0.0084	0.0051
도로 면적( $\hat{\mu}$ )	-0.0001	-0.0019	-0.001
도로 면적( $\hat{\sigma}$ )	0.001	0.0136	0.0055

세대 당 인구수 상태변화율 ( $\mu, \sigma$ )

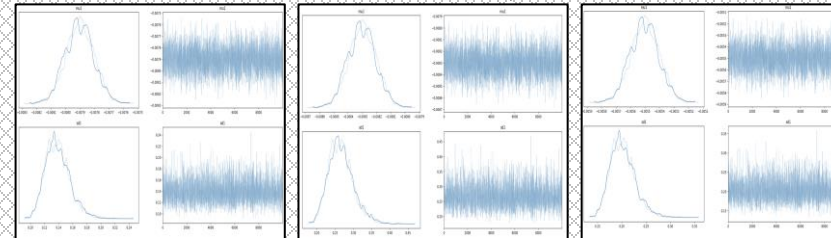


도로면적 상태변화율 ( $\mu, \sigma$ )

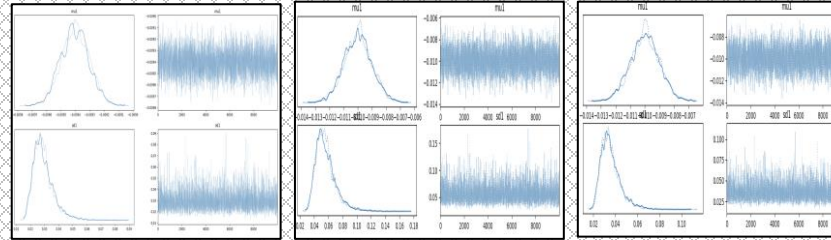


MH-알고리즘 병렬 샘플링  
시행: 10000번, 튜닝:5000번  
결과(좌: 사후분포, 우: trace)

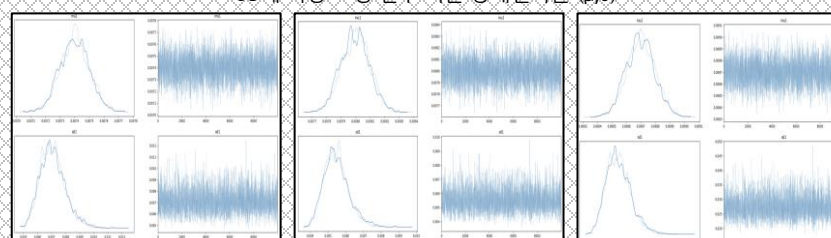
시도 내 이동 상태변화율 ( $\mu, \sigma$ )  
(좌측부터 동대문구 휘경1동, 서초구 서초3동,  
영등포구 신길4동)



재정자주도 상태변화율 ( $\mu, \sigma$ )



65세 이상 고령 인구 비율 상태변화율 ( $\mu, \sigma$ )

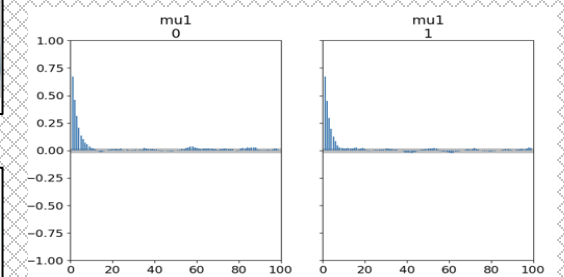


모든 mcmc 과정에서  
수렴 발생 확인!

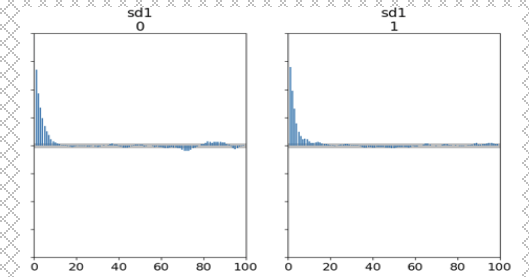
(\*자기상관도 그래프에서 0으로의  
수렴을 기준으로 판단)

아래는 자기상관도 그래프 예시(동대문구 휘  
경1동 재정자주도 상태변화율

재정자주도 상태변화율  $\mu$ 의 자기상관도



재정자주도 상태변화율  $\sigma$ 의 자기상관도





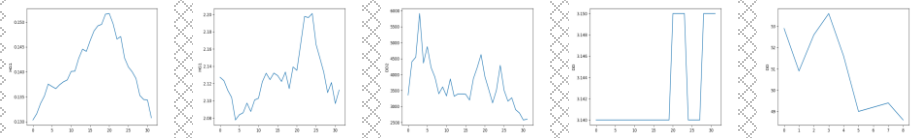
### 3. 분석 결과

## <3단계 전략 분석 결과>

행정동별 급변동 야기 문제  
요인 파악(조정 변수 내)

#### 1) 동대문구 휘경1동

과거, 즉 현재까지의 조정변수별 추세(시계열) 시각화



좌측부터 순차적으로 고령 인구 비율, 세대 당 인구 수, 시도 내 이동, 도로 면적, 재정자주도  
**최근 급락** **최근 급락** **점진적 감소** **진동** **점진적 감소**

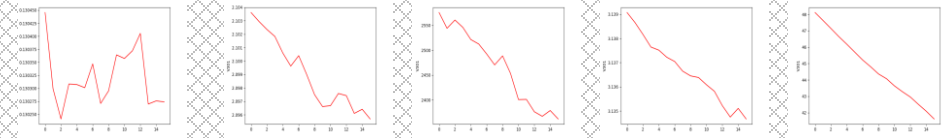
앞선 ZIP 회귀 분  
석 결과 반영 시  
...

급변동 문제 요인

(문제가 된 조정변수 움직임):

행정동별 고령 인구비율의 감소, 자치구별 시도 내  
이동의 감소, 재정자주도의 감소

기하 브라운 운동 및 몬테카를로 생성 기반 조정변수별 추세(시계열) 예측 평균치 시각화



좌측부터 순차적으로 고령 인구 비율, 세대 당 인구 수, 시도 내 이동, 도로 면적, 재정자주도  
**미약한 반등** **감소 추세 유지** **감소 추세 유지**

∴ 휘경1동 고령 인구비율, 동대문구 시도 내 이동,  
동대문구 재정자주도 모두에 대한  
정책적 대응의 필요성 존재

휘경1동의 경우, 과거 시계열 시각화 시 elderatio, ratio, indistrict, rdarea, finindep에서 각각 최근 급락 추세, 최근 급락 후 진동 추세, 점진적 감소 추세, 진동 추세, 초기 반등 후 감소 추세가 발견되었으며, 이는 휘경1동에서는 행정동별 고령 인구 비율, 자치구별 시도 내 이동, 그리고 재정자주도의 움직임이 지표 급변동에 영향을 미쳤음을 내포한다. 이때 기하 브라운 운동을 기반으로 해당 움직임들의 향후 추세를 시각화한 결과, indistrict와 finindep에서는 최근 추세가 유지되는 양상이, elderatio에서는 추세가 매우 미약하게 반등하는 양상이 예측되었으며, 따라서 동대문구 휘경1동에서는 급변동을 야기한 모든 요인(조정변수)의 움직임에 대하여 정책적 차원에서 대응할 필요성이 존재함을 확인하였다.

### 3. 분석 결과

## <3단계 전략 분석 결과>

행정동별 급변동 야기 문제  
요인 파악(조정 변수 내)

### 2) 영등포구 신길4동

과거, 즉 현재까지의 조정변수별 추세(시계열) 시각화



앞선 ZIP 회귀 분  
석 결과 반영 시  
...

**급변동 문제 요인**  
(문제가 된 조정변수 움직임):  
행정동별 고령 인구비율의 감소, 자치구별 시도 내  
이동의 감소, 재정자주도의 감소

기하 브라운 운동 및 몬테카를로 생성 기반 조정변수별 추세(시계열) 예측 평균치 시각화



∴ 영등포구 시도 내 이동과  
영등포구 재정자주도에 대한  
정책적 대응의 필요성 존재

신길4동의 경우, 과거 시계열 시각화 시 elderatio, ratio, indistrict, rdarea, finindep에서는 각각 최근 급락 추세, 반등 이후 다시 감소 추세, 최근 급락 추세, 최근 급증 추세, 감소 이후 진동 추세가 발견되었으며, 이는 신길4동에서는 조정변수 가운데 자치구별 고령자 비율, 자치구별 시도 내 이동, 재정자주도의 움직임이 급변동에 영향을 미쳤음을 내포한다. 이때 기하 브라운 운동을 기반으로 해당 움직임들의 향후 추세를 시각화한 결과, elderatio에서는 이전 수준이 회복되는 양상, indistrict과 finindep에서는 최근 추세가 유지되는 양상이 예측되었으며, 따라서 영등포구 신길4동에서는 조정변수 가운데 시도 내 이동과 재정자주도의 움직임에 대해 정책적 차원에서 대응할 필요성이 존재함을 확인하였다.

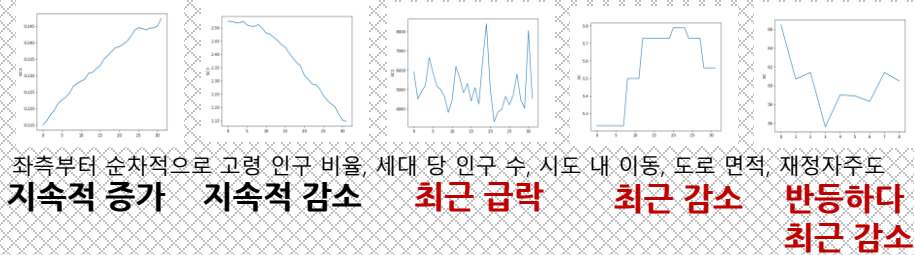
### 3. 분석 결과

## <3단계 전략 분석 결과>

행정동별 급변동 야기 문제  
요인 파악(조정 변수 내)

### 3) 서초구 서초3동

과거, 즉 현재까지의 조정변수별 추세(시계열) 시각화



앞선 ZIP 회귀 분  
석 결과 반영 시

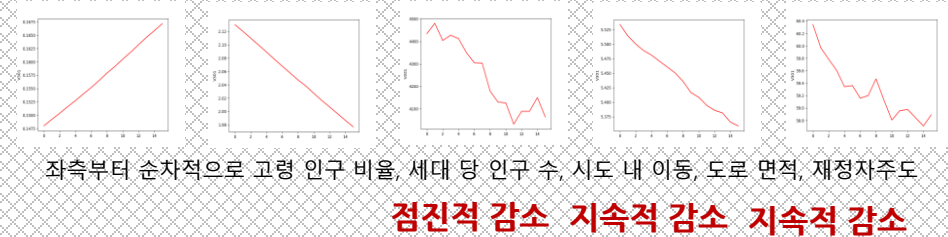
...

급변동 문제 요인

(문제가 된 조정변수 움직임):

자치구별 시도 내 이동의 감소, 자치구별 도로 면적  
감소, 재정자주도의 감소

기하 브라운 운동 및 몬테카를로 생성 기반 조정변수별 추세(시계열) 예측 평균치 시각화



∴ 서초구 시도 내 이동, 서초구 도로 면적, 그리고  
서초구 재정자주도 모두에 대한  
정책적 대응의 필요성 존재

서초3동의 경우, 과거 시계열 시각화 시 elderatio, ratio, indistrict, rdarea, finindep에서 각각 지속적 증가 추세, 지속적 감소 추세, 최근 급락 추세, 최근 감소 추세, 최근 감소 추세가 발견되었으며, 이는 서초3동에서는 조정변수 가운데 자치구별 시도 내 이동, 재정자주도, 그리고 도로면적의 움직임이 급변동에 영향을 미쳤음을 내포한다. 이때 기하 브라운 운동을 기반으로 해당 움직임들의 향후 추세를 시각화한 결과, indistrict, finindep, rdarea 모두에서 최근 추세가 유지되는 현상이 예측되었으며, 따라서 서초구 서초3동에서는 급변동을 야기한 모든 움직임에 대하여 정책적 차원에서 대응할 필요성이 존재함을 확인하였다.



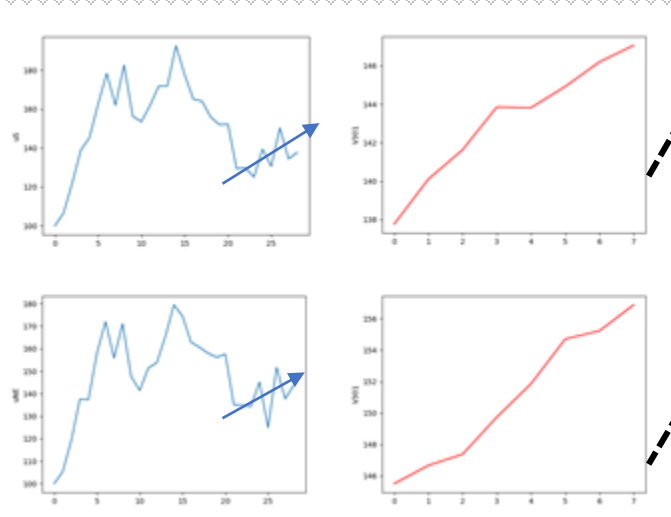
### 3. 분석 결과

## <3단계 전략 분석 결과>

정책적 차원 대응 제안  
-정책 제안의 전제

### 노동시장의 흐름 예측

상/하: 서울시 남부/북동부 실업급여지급 시계열 시각화  
과거(파란색)·예측(빨간색) 시계열



급변동 예측 지역 가운데 서초구, 영등포구가 속한 서울시 남부(Nowcast 분류 기준상 남부)는 6개월간 약 4%의 실업급여 지급 변화에 대응되는 수준의 정책적 개입 필요

급변동 예측 지역 가운데 동대문구가 속한 서울시 북동부(Nowcast 분류 기준상 북동부)는 6개월간 약 7%의 실업급여 지급 변화에 대응되는 수준의 정책적 개입 필요

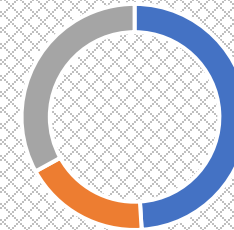
### \*이후 제안할 정책적 대응의 전제

- 급변동 예측 지역의 경우, 선제적 정책의 구성 및 예산의 책정을 위해서는 해당 변동의 여파로 인한 실질적인 노동시장 타격 양상 및 정도를 함께 제시할 필요가 존재하여 실업 흐름에 대한 예측이 추가적으로 필요하다고 판단되었으며, 이에 앞선 기하 브라운 운동 기반 예측과 같은 방식으로 실업급여 지급건수 예측을 진행. 그 결과 영등포구 신길4동과 서초구 서초3동에 대응되는 서울시 남부 실업급여지급 양상, 그리고 동대문구 휘경1동에 대응되는 서울시 북동부 실업급여지급 양상 모두에서 실업 규모 증가(악화) 추세가 확인 및 예측되었으며, 구체적으로 남부의 경우에는 향후 6개월간 약 4%의 실업급여 지급 변화에 대응되는 수준의 정책적 개입이, 북동부에서는 약 7%의 실업급여 지급 변화에 대응되는 수준의 정책적 개입이 필요할 것으로 예상됨.

- MDIS 창업기업실태조사 분석 결과 개인들이 경험하는 가장 큰 창업 시 어려움은 '창업자금의 확보'로 예측되는 어려움(49%)이었으며, 기업 운영 과정에서 경험하는 가장 큰 어려움으로는 '영업마케팅의 어려움(32%)'. 다음, 개인 창업자들이 상권입지를 선택하는 요인 가운데 가장 높은 비중을 차지한 요인은 '본사, 영업장 등의 취득 비용 저렴'(32%). 이는 창업자들(소상공인 등)을 위한 효율적 경영 지원은 근본적으로 기초자금 문제, 그리고 소비자와의 소통·교류 활성화 문제를 대상으로 하여야 함을 내포. 즉 한정된 예산을 고려할 때, 생존 기간상 어려움이 대두되는 급등지구의 재정적 지원은 마케팅 방안 및 소비자와의 교류 증진에 집중하는 것이, 진입상 어려움이 대두되는 급락지구의 재정적 지원은 기초자금 제공에 집중하는 것이 효과적일 것이 예상됨.

### MDIS 창업실태조사(개인 부문)

창업 시 가장 큰 어려움



■ 창업자금의 확보 ■ 실패 및 재기 두려움 ■ 기타

기업 운영 시 가장 큰 어려움



상권입지 선택 요인



■ 영업마케팅 어려움 ■ 판매납품단가 인하 ■ 재무자금관리 ■ 기타 ■ 본사, 영업장 등 취득 비용 저렴 ■ 기타 이유로 ■ 기타

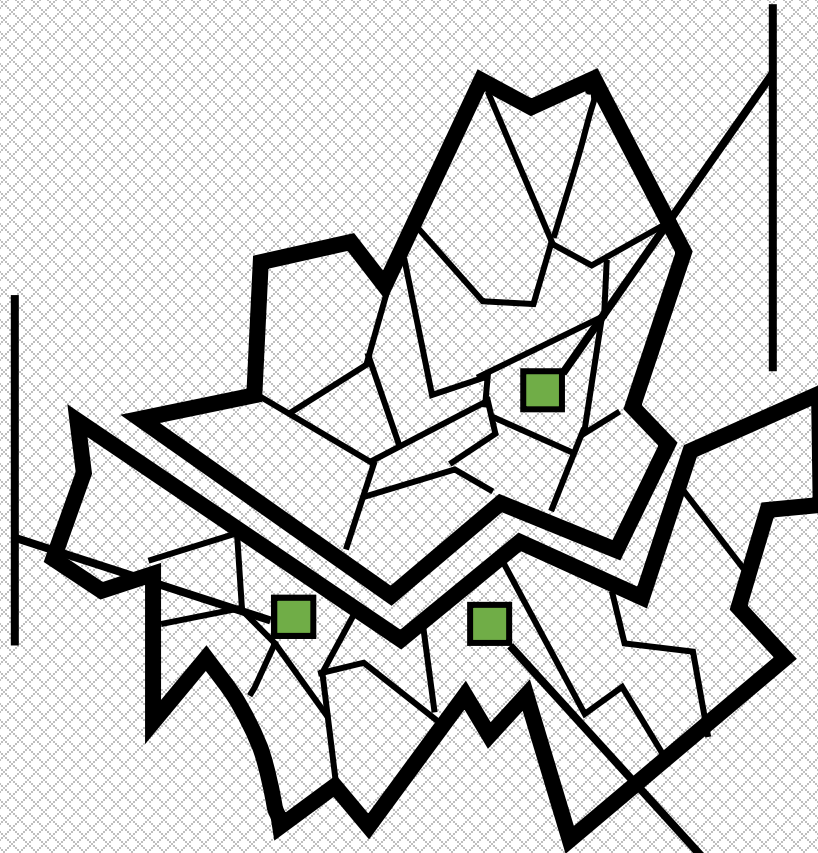
### 3. 분석 결과

## <3단계 전략 분석 결과>

### 정책적 차원 대응 제안 -급변동 행정동별 제안

#### 급변동 지역1: 영등포구 신길4동 정책 제안

영등포구 신길4동의 경우 향후 전년 동분기 대비 상권 급락이 예측되며, 자치구 행정 차원에서는 재정자주도를 높이려는 시도를 전개함으로써 창업 지원 종류 가운데 기초 자금 지원 규모를 증진하는 정책을 도입하는 것이 필요할 것이라 예상되며 서울시 자체의 행정 차원에서는 간접적으로 시도 내 이동을 증진하려는 시도가 필요할 것으로 판단된다.



■ : 상권지표 급변동 예상 서울시 행정동 위치

동대문구 휘경1동의 경우에는 상권 향후 전년 동분기 대비 상권 급증이 예측되며, 시도 내 이동과 재정자주도와 관련하여서는 앞서 서초3동과 관련하여 제시된 접근과 동일한 (시·구) 행정 차원의 노력이 요구될 것으로 파악되며, 추가적으로 자치구 행정 측면에서 노인 친화적 환경을 증진함으로써 고령 인구 비율을 증진하려는 노력이 필요할 것으로 판단된다.

#### 급변동 지역2: 동대문구 휘경1동 정책 제안

서초구 서초3동의 경우 향후 전년 동분기 대비 상권 급증이 예측되며, 자치구 행정 차원에서는 재정자주도를 높이려는 시도를 전개함으로써 창업 지원 종류 가운데 마케팅 전략 개선 등 교류 활성화 지원의 규모를 증진하는 정책을 도입하려는 노력이, 그리고 낙후 도로의 개선 및 유흥지 구입 후 도로화 또는 무주지 도로화 등을 통해 실질 도로 면적을 증진하려는 노력이 필요할 것으로 파악되며, 서울시 자체의 행정 차원에서는 간접적으로 시도 내 이동을 증진하려는 시도가 필요할 것으로 파악된다.

#### 급변동 지역3: 서초구 서초3동 정책 제안

### 기대효과

1. 첫째, 구축한 양상블 예측 모델을 기반으로 도출한 기초자료(p.18. 분할표)를 통해 다가오는 3, 4분기 서울시 행정동별로 진입을 고려하고 있는 예비 창업자 및 소상공인들이 효율적 시장 진입 전략 및 경영 전략을 수립할 수 있게 될 것으로 기대된다.

예> 4분기 구로구 수궁동으로의 진입을 고려하고 있는 소상공인의 경우는 <표5>에 따라 상권지표가 LH(상권 확장)에서 HH (정체, 기존 생존상권 강화)으로 변화할 것으로 예측되므로 기존 상권이 전혀 시도하지 않은 기대 분야를 공략하거나 다른 행정동을 고려 해봄으로써 효율적으로 대응 가능.

2. 둘째, 추출한 급변동 예측 지역에 대하여 앞서 제시한 여러 선제적인 정책적 개입을 시도함으로써 시장 과열 또는 침체의 정도를 낮출 수 있을 것으로 기대된다.

예>급변동 예측 지역 가운데 동대문구 휘경 1동은 노인 친화적 정책과 함께 재정 자주도를 높이려는 시도를 지속하며 예비 창업자 및 소상공인들을 대상으로 한 마케팅 전략 교육 지원 및 소비자 교류 증진 지원을 강화할 경우, 시장 과열로 인한 실업을 악화를 방지하거나 경제적 피해 규모를 축소할 수 있을 것.

3. 셋째, 향후(2022년 이후) 상권 변화 또는 타 지역과 관련하여서도 본 데이터 분석을 통해 확립된 지표변동에 대한 모델과 ZIP 회귀모형을 통해 도출된 영향 요인을 활용함으로써, 선제적인 상권변화 대응이 가능할 것으로 기대된다.

### 방향제시

\*예비 창업자가 특정 시장(행정동) 진입과 관련하여 추구해야 할 방향성과 관련하여서는 p.18.의 분할표를 참고하여 해당 행정동의 상권 지표변동 예측 종류에 따라 HH에서 LH, LH에서 LL 등의 이동이 예상될 시에는 상권확장 예상에 따라 보다 적극적인 진입 시도를 전개하되 변동성을 고려하여 단기적 경영 전략을 수립하는 반면, LL에서 HL, LH에서 HH 등으로의 이동이 예상될 시에는 상권축소 예상에 따라 자신의 사업 분야의 독창성 등을 기준으로 진입에 신중하게 접근하는 동시에 다른 행정동 지역을 탐색하는 방식의 경영 전략 등을 수립하는 것이 적절할 것이다.

\*행정 및 정책의 방향성 차원에서는 ZIP 모형을 통해 도출된 유의 요인들에 대한 정책적 개입의 필요성을 지속적으로 점검하는 한편, 급변동 지역으로 선택된 세 가지 행정동과 관련하여서는 p.26.에서 제시된 접근을 전개함으로써 시장 충격 완화를 시도하는 것이 적절할 것이라 판단된다.

- 강성호, 양성수, 김성준, 주형준, 강연준(2020), 하이브리드 모델링 기법과 베이지안 최적화를 통한 엔진 마운트 응답개선, **한국소음진동공학회논문집**, 30(6):546~552.
- 강수철, 남승용, 김만배, 직장인의 차량보유 결정요인 및 통근수단 선택행태 분석, **한국정책과학학회보**, 13(1):287~302.
- 강현모, 이상경(2019), 시계열 군집분석과 로지스틱 회귀분석을 이용한 골목상권 성장요인 연구, **한국측량학회지**, 37(6):535~543.
- 김주영;이성건(2008), 국내 은행의 신용평가모형에 관한 사례 연구-영과잉-포아송분포의 활용을 중심으로, **Journal of the Korean Data Analysis Society**, 10(6):3255~3265.
- 김현수(2022), 투자자 예상가격 기반 자산가격 시뮬레이션, **비즈니스융복합연구** 7(1): 93~101.
- 김현철, 안영수, 상점 밀도와 업종 다양성을 이용한 서울시 골목상권의 동태적 변화 모니터링 연구, **서울도시연구**, 20(4):149~170.
- 나우캐스트(빅데이터활용지표) 실업급여(<https://data.kostat.go.kr/nowcast/main.do#>)
- 박성익;조장식(2019), 이변량 프로빗 모형을 활용한 인력유출 결정요인 분석, **한국데이터정보과학회지**, 30(1):23~31.
- 서울시 열린데이터 광장(<https://data.seoul.go.kr/>)
- 석혜은;방희정;김수영(2017), 영과잉 순서형 프로빗 모형의 베이지안 추론, **한국심리학회지**, 36(2):215~239.
- 송지현, 이재영(2020), GIS 기반 집적원리에 기초한 도시지역 상권의 구역별 특징 분석-서울시 홍대 지역 상권을 중심으로-, **주거환경 통권**, 18(4):117~129.
- 수학백과, 『기하브라운 운동』,  
(<https://terms.naver.com/entry.naver?docId=5669366&cid=60207&categoryId=60207>)
- 오영기, 이병길, 이상경(2022), 서울시 상권 위계와 중심지체계 비교분석, **한국지적정보학회**, 24(1): 65~76.
- 이다솜;이은지;조성일;최태련(2020), 베이지안 순서형 프로빗 준모수 회귀 모형: 국민건강영양조사 2016 자료를 통한 흡연 양태와 커피섭취 간 관계 분석, **응용통계연구**, 33(1):25~46.
- 이대륙, 김종진(2019), 대형할인마트와 전통시장의 상생발전방안에 관한 연구, **주거환경**, 17(2):17~35.4
- 이서구(2016), 베이지안 추정을 이용한 국내 자동차 시장의 다항 프로빗 모형 분석, **상품학연구**, 34(5): 65~ 74.
- 전희주(2017), 영과잉 음이향회귀 모형을 이용한 보험설계사들의 이직횟수 적합, **한국데이터정보과학회지**, 28(5): 1087~1097.
- 정은애;김영기(2022), 상권활성화 성공 요인 분석 및 정책 방향-18개 상권 르네상스 사업 대상지를 중심으로-, **주거환경**, 20(1):193~210
- 제승욱, 강정규, 김영근(2018), 상가 업종 다양성에 따른 상권 변화 실증 분석, **주거환경**, 16(2):357~369.
- 홍성두 외 4인(2009), 지적장애 연구를 위한 영과잉 모형의 활용방안 탐색, **지적장애연구**, 11(4):179~197.
- Chib S., Greenberg E.(1995), Understanding the Metropolis-Hastings Algorithm, **The American Statistician**, 49(4), 327~335.
- Jun Mo, S.(2011), Deciding a sampling Length for estimating the parameters in Geometric Grownina Motion, **한국데이터정보과학회지**, 22(3):549~553.
- KOSIS 국가통계포털(<https://kosis.kr/index/index.do>)
- MDIS 마이크로데이터 통합서비스(<https://mdis.kostat.go.kr/index.do>)
- Pinsky, M.A., Karlin, S.(2010), *An Introduction to STOCHASTIC MODELING*, Academic Press.
- Raschka, S., Mirjalili, V.(2021), *Python Machine Learning Third Edition*, 길벗.

