

[주거 이동과 정착 결정 요인 분석 보고서]

1. 배경

☐ 주제 선정

☐ 지방 소멸 대응과 농촌 지역 공동화

- 1975~2015년 기간 동안 전국에서 차지하는 군 지역의 인구 비중이 25.1%에서 8.3%로 급감하는 등 대부분의 군 지역은 이미 소멸 위험단계에 진입했으며, 최근 3년간 소멸 고 위험지역이 25곳이나 증가

- 정부는 2021년 11월 관계부처 합동으로 '지역소멸 선제 대응 방안'을 발표하며 비수도권 정착 및 이동 유도 필요성의 부각(이상명, 2023)

- 우리나라는 현재 대도시로 인구가 편중되며 농촌의 공동화 현상이 급격히 진행. 이에 따른 농촌 인구의 감소와 노령화가 지방자치단체마다 극복을 위한 대책 필요성의 증가(박재현, 2019)

농촌 인구의 감소와 노령화 문제를 극복하기 위한 지방자치단체의 대응 필요성이 더욱 커지고 있으며, 수도권으로의 인구 쏠림 현상이 지속됨에 따라 비수도권 지역의 공동화와 불균형 문제는 심화하고 있다. 지역의 지속가능성을 확보하기 위해서는 주거 이동과 정착 결정 요인에 대한 체계적인 분석이 필수적이다.

이에 본 보고서는 2023년도 주거실태조사 마이크로데이터를 활용하여, 주거 이동과 정착 결정에 영향을 미치는 요인을 데이터로 규명해 보고자 한다.

☐ 분석 필요성(문제점) 및 전략

☐ 비수도권 정착 유인 정책의 한계

- 비수도권 지역에서 정착한 가구의 평균 주택 면적은 약 82m² 로, 수도권(약 71m²)보다 넓음에도 불구하고 비수도권 정착 유인 정책은 주로 재정 지원에 집중되어 있어, 실제 주민이 체감하는 주거 만족, 이전 이력, 가구 특성, 주택 물리적 요인 등은 충분히 반영되고 있지 않다는 데 한계가 존재

○ 데이터 기반 의사결정의 필요성

- 효과적인 정책 설계와 대상군 선별을 위해서는 실제 거주 이동 패턴과 정착 여부를 예측할 수 있는 데이터 기반 모형 구축이 필수적이지만 다소 제한적으로 이뤄짐.

따라서, 실제 거주 이동 패턴과 정착 여부를 예측할 수 있는 데이터 기반의 모형이 필요하며, 이를 통해 정책 대상군 식별과 맞춤형 지원 정책 설계가 가능해지게 하고자 한다. 본 분석은 다양한 머신러닝 기법을 통해 주거 이동 예측 모델을 비교하고, 가장 성능이 우수한 모델(LightGBM)을 선정하여 정착 결정 요인을 SHAP 분석으로 해석해 보고자 한다.

2. 데이터 분석

□ 데이터 선정

본 연구는 통계청에서 제공한 2023년도 주거실태조사 마이크로데이터를 사용해 분석을 진행했다.

주거실태조사는 전국 가구의 주거환경, 주택 유형, 점유 형태, 주거 만족도, 이전 이력, 향후 계획 등 다양한 주거이동 관련 항목을 포함하는 대표적인 국가 승인 통계로, 수도권 집중 문제 해소와 비수도권 정착 관련 정책 분석에 적합하다.

□ 데이터 개요 및 핵심 내용 요약

- 데이터명: 2023년도 주거실태조사 마이크로데이터 (통계데이터센터 제공)
- 표본 수: 전국 약 6.1만 가구 대상
- 주요 변수: 가구주 속성(연령, 성별), 주택 유형, 점유 형태, 무주택 기간, 주택가격, 이전 거주지, 주거지원 프로그램 인지 여부 등
- 종속 변수 정의: 현재 거주지가 수도권(서울·인천·경기)이면 0, 비수도권이면 1로 설정하여 '비수도권 정착 여부'를 종속 변수로 정의
- 데이터 특징: 전국 가구 대상 표본 설계로 지역별 대표성을 확보하였으며, 다양한 주거환경 요인을 종합적으로 포함

○ 데이터 변수 설명

주택 유형 및 소유 정보	주택 가격 및 임차료	주택 만족도 및 환경 평가	이사 관련 이력 및 계획	소득 및 자산	가구 구성 및 특성	주거복지 서비스 인지 및 만족도	주거환경 개조 경험 및 필요성	아동·청소년 관련 주거 조건
• 현재 주택 유형, 점유형태, 자가/전세 여부, 상해·최초 주택 구입 여부 등	• 현재 주택 가격, 임차료(보증금, 월세), 구입 및 임차자금 조달 방식	• 구조, 방수, 채광, 소음, 한기 등 주택 상태 • 상업시설, 의료, 공공기관, 교통 등 접근성 • 종합 만족도 포함	• 과거 이사 횟수, 이사 사유, 이전 주택 정보 • 향후 이사 계획 및 지역, 주택유형, 이유	• 월평균 소득 • 항목별, 총 경장/비경장 소득 • 총 자산/부채, 주택/금융/기타 자산, 대출	• 구성원별 관계, 나이, 성별, 실제 거주 여부	• 공공임대, 전세자금, 월세보증금 등 정책 인식 및 이용 만족도	• 손잡이, 문턱 제거, 비상벨, 수전 등 고령자 및 장애인 고려 요소	• 돌봄 주체, 형사 이용, 위생 문제 등

- 코드북에 따라 수치형·범주형 변수로 분류하고, 불필요한 결측치는 수치형은 0, 범주형은 '없음'으로 대체.
- 범주형 변수는 pandas get_dummies()로 One-Hot Encoding 인코딩 후, 특수문자를 제거하여 모델 입력에 적합하게 전처리 진행.
- 시도 칼럼의 11(서울), 28(인천), 41(경기)을 수도권으로 정의하여 비수도권_정착 변수 생성.

□ 데이터 분석 (분석 프로세스, 분석 방법, 접근방법 등)

○ 데이터 전처리

- 수치형/범주형 결측치 대체
- One-Hot Encoding 인코딩 및 변수명 정제
- SMOTE로 소수 클래스 보정

○ 분석 프로세스



○ 분석 방법

- 다중 분류 문제에 Random Forest, XGBoost, LightGBM, CatBoost, SVM, KNN 비교 적용.
- SMOTE로 수도권 거주자 대비 비수도권 정착자 샘플 불균형을 해소.
- 성능 평가는 Accuracy, Precision, Recall, F1-score, ROC AUC로 비교.
- LightGBM을 최종 모델로 선정하고, Optuna 기반 튜닝 및 SHAP 해석 수행.

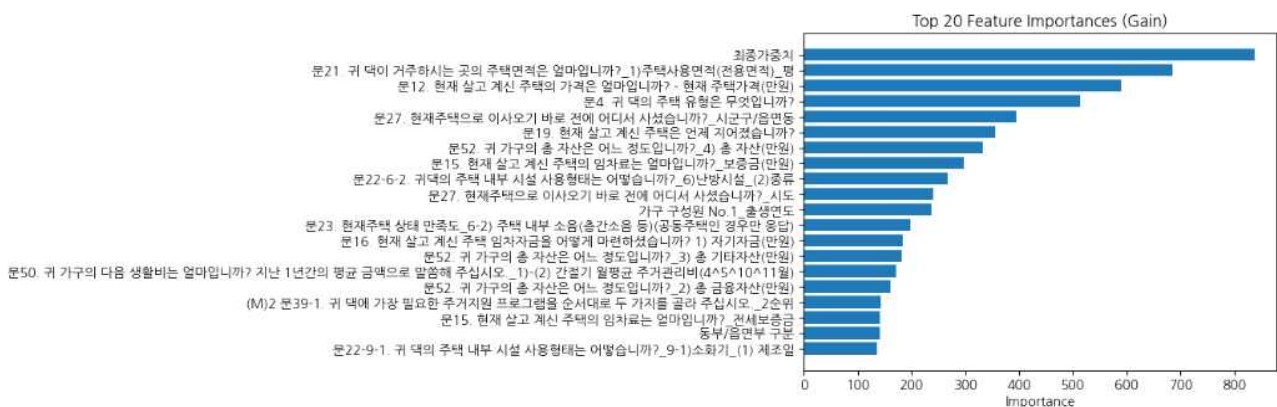
□ 분석 결과 및 해석

Train shape: (85,790, 4460)

Model	Accuracy	Precision	Recall	F1-score	ROC AUC
Random-Forest	0.946	0.968	0.971	0.969	0.973
XGBoost	0.973	0.983	0.986	0.984	0.994
LightGBM	0.981	0.988	0.990	0.989	0.996
CatBoost	0.976	0.985	0.987	0.986	0.994
SVM	0.829	0.912	0.891	0.902	0.673
KNN	0.780	0.948	0.792	0.863	0.800

- LightGBM이 Accuracy(0.981), Precision(0.988), Recall(0.990), F1-score(0.989), ROC AUC(0.996)로 가장 높은 성능을 보였다. 이에 따라 이후의 변수 중요도 분석과 정책 시사점 도출은 LightGBM 결과를 중심으로 진행하였다.

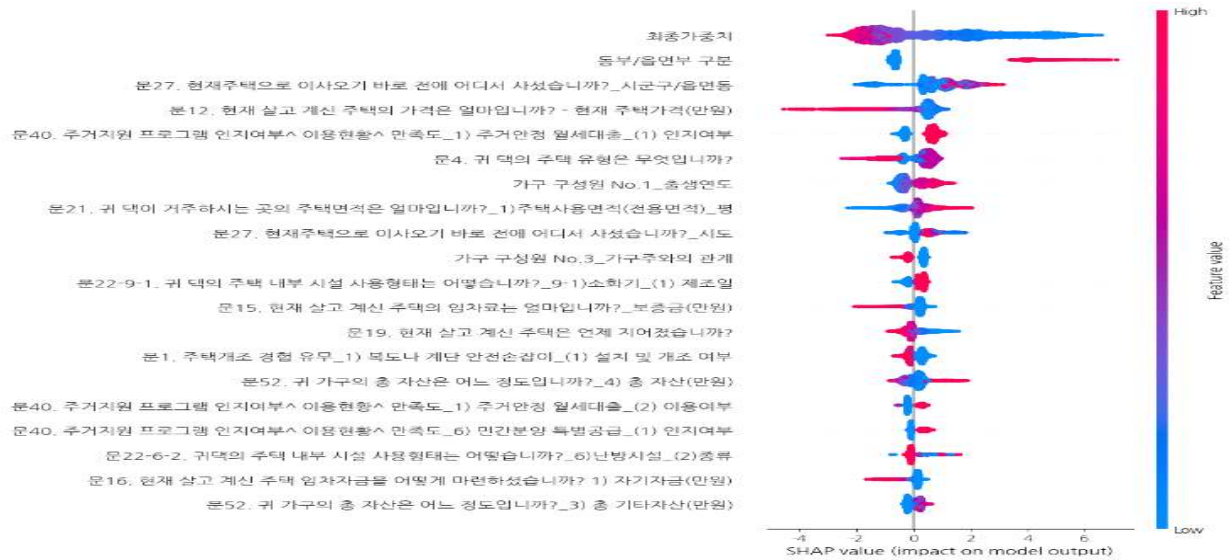
○ [그래프 1] Feature Importance Top 20 바 그래프



LightGBM의 변수 중요도 상위 20개 변수와 중요도 값을 시각화한 막대그래프

- 주요 변수로는 최종가중치, 동부/읍면부 구분, 이사 전 주거지, 현재 주택가격, 주거지원 프로그램 인지 여부 등이 상위에 위치했다. 하지만 분석에 포함된 최종가중치는 조사 표본의 대표성을 바로잡기 위해 설계된 변수로, 개별 가구의 주거이동 결정 요인을 직접 설명하지는 않는다. 따라서 SHAP 분석 결과에서 최종가중치 변수의 중요도는 모집단 보정 영향일 뿐, 정책 요인으로는 활용되지 않는다.

○ [그래프 2] SHAP Summary Plot



변수별 SHAP값 분포를 나타내어 각 변수가 모델 출력에 어떤 방향으로 영향을 미쳤는지 보여줌

- 해당 그래프는 LightGBM 모델에서 주요 변수들이 비수도권 정착 여부 예측에 미치는 기여도를 SHAP 값을 기준으로 시각화한 결과이다.상단의 최종가중치는 표본의 대표성을 바로잡기 위한 가중치로, 개별 가구의 특성을 직접 설명하지 않으므로 해석에서 제외하였다.

- 실제로는 동부/읍면부 구분, 이사 전 주거지(시군구/읍면동), 현재 주택가격, 주거지원 프로그램 인지 여부 등이 예측에 크게 기여하고 있으며, 이는 해당 변수 값의 높고 낮음에 따라 비수도권 정착 확률이 달라질 수 있음을 시사한다. 특히, 다수 변수에서 변수 값이 높을수록 SHAP 값이 양(+)방향으로 나타나 비수도권 정착 확률을 높이는 경향을 보였다.

3. 분석 활용 전략

□ 기대효과

- 본 연구는 주거 이동과 정착 결정 요인을 데이터로 실증하여, 지자체가 인구 분산 정책을 설계할 때 신뢰성 있는 근거로 활용할 수 있다.

특히 시군구 단위 맞춤형 이사 지원금, 청년·신혼부부 장기임대 연계, 주거지원 프로그램 설계 등 정책 실현성이 높다.

□ 방향 제시

- 분석 결과 확인된 핵심 변수(주택가격, 주거지원 프로그램, 이사 전 주거지 등)에 따라 맞춤형 지원 정책 설계
- 비수도권 정착 가구가 필요로 하는 시설·서비스에 대한 추가 지원 방안 제시
- 향후 동일 데이터와 추가 패널데이터 연계 분석을 통해 시계열 모니터링 기반 정책 효과성 평가
- 분석 결과에 따라, 특정 지역(읍면부/동부)이나 주택유형별로, 가구 군을 분류하여 맞춤형 이사 지원금 또는 장기임대주택 연계 방안 추진

참고문헌

- 이상명(2023). 지방소멸대응과 국가균형발전을 위한 법정책적 방향, 한양법학, 34(3), 3-34.
- 박재현(2019). 농촌 지역 공동화 현상에 따른 활성화 방안 연구, 한국디자인리서치, 4(1), 32-40.
- 최예슬(2021). 지방 중소도시의 유출인구 직종 특성 분석, 국토연구원 국토연구 통권 제108권, 3-18.
- 김진영(2005). 한국의 수도권 집중의 원인과 대책, 한국정책학회보, 17(2), 385-404.