발 간 등 록 번 호
11-1241140-100001-10

OPEN 공공누리 출처표시 상업용금지 변경금지 공공저작물 자유이용허락

2025년 연구보고서

의제주거소득 추정을 위한 자가주택의 월세평가 방법 연구

2026. 3.

https://mods.go.kr/dsri

국가데이터처
국가데이터연구원

연구보고서 2025-13

의제주거소득 추정을 위한 자가주택의 월세평가 방법 연구

송정현 · 변준석

Ministry of Data and Statistics
Data and Statistics Research Institute

발간사

“데이터의 가치는 분석과 활용을 통해 의사결정을 지원하고, 혁신과 효율성 향상 등 구체적인 성과를 창출하는 데서 비롯됩니다.”

급변하는 불확실성의 시대에 데이터는 더 이상 단순한 숫자의 기록이 아니라, 미래를 예측하고 사회 문제를 해결하는 핵심 나침반으로 자리매김하고 있습니다. 국가데이터연구원은 이러한 시대적 요구에 부응하여 국민의 삶을 실질적으로 개선하고 AI 기반의 공공 AX 대전환을 뒷받침하기 위한 데이터 기반 연구에 지속적으로 매진해 왔습니다.

2025년 연구보고서에는 우리 사회가 직면한 환경 변화에 능동적으로 대응하고자 첨단 기술을 국가통계에 접목하기 위해 치열하게 고민한 연구 성과를 담았습니다.

첫째, 인공지능(AI) 기반 국가통계 기술혁신을 선도하고자 노력하였습니다.

생성형 AI 기술을 현장조사에 적용하기 위한 기초연구를 통해 조사자료의 내용검토 및 자동분류, 질의응답에 활용 가능성을 모색하였으며, 이는 통계 생산의 신속성과 정확성을 획기적으로 제고하는 토대가 될 것입니다. 아울러 생성형 AI를 활용한 나우캐스트 지표 서비스 제공 방안 연구는 통계서비스의 새로운 가능성을 여는 의미 있는 첫걸음이라 할 수 있습니다.

둘째, 점차 열악해지고 있는 조사환경에 대응하기 위해 새로운 통계방법론 연구와 국가통계 품질제고를 위한 연구를 강화하였습니다.

확률표본과 자원자표본을 통합한 추정 방안 연구는 응답자 조사 부담을 완화하고 비확률표본의 병행 활용 가능성을 제시하였으며, 데이터 과학기술을 활용한 자료수집 개선 연구와 데이터 통합방법 연구는 다양한 데이터의 연계·통합 방법을 보다 체계화하였습니다.

셋째, 사회적 사각지대를 조명하고 지속가능한 미래를 지원하기 위한 데이터 기반 정책 연구에 집중하였습니다.

최근 심각한 사회 문제로 대두된 '고립·은둔 청년'의 실태 파악을 위한 조사 문항 개발 연구를 비롯하여, 돌봄 분야 국가통계 활용 방안과 국내 최초의 기후변화 통계·지표 분석 연구는 데이터가 사회안전망 강화에 기여할 수 있음을 보여줍니다. 또한 소득이동통계 심층 분석 연구와 생애과정 이행에 대한 중·고령기 비교 연구는 관련 정책의 실효성과 활용도를 한층 높일 것으로 기대됩니다.

아울러 가계동향조사의 소비지표 작성 연구와 퇴직연금 적립금 배분 방법 연구는 국민의 체감 경기를 보다 정확히 진단하고 합리적인 경제정책 수립을 지원하는 든든한 기반이 될 것입니다.

2025년 10월부터 새롭게 출발한 국가데이터처 국가데이터연구원은 앞으로도 최신 기술과 사람을 잇는 데이터 연구를 통해 국가통계의 지평을 지속적으로 확장해 나가겠습니다.

본 연구보고서가 통계 생산자와 이용자 모두에게 실질적인 도움이 되고, 각계각층의 의사결정자에게 깊이 있는 통찰을 제공하기를 기대합니다.

많은 관심과 성원을 부탁드립니다.

2026년 3월

국가데이터연구원장

목 차

제1장 서 론 ····· 1
제1절 연구 배경 ····· 1
제2절 연구 목적 ····· 3
제3절 연구 범위 및 보고서 구성 ····· 4

제2장 의제주거소득 개념 및 선행연구 ····· 5
제1절 개념 및 측정방법 ····· 5
제2절 선행연구 및 차별성 ····· 6
제3절 머신러닝 활용 연구 사례 ····· 8

제3장 전세보증금 예측모형 구축 및 성능 평가 ····· 9
제1절 활용자료 ····· 9
제2절 전세 신고자료 ····· 12
제3절 변수선택 및 전세데이터 전처리 ····· 13
제4절 머신러닝 모형 선정 및 평가지표 ····· 15
제5절 랜덤포레스트 모형 학습 및 평가지표 결과 ····· 17

제4장 가계금융복지조사 적용 및 의제주거소득 추정 ····· 28
제1절 가계금융복지조사 전세가구를 활용한 모형 검증 ····· 28
제2절 전월세 전환율 및 이자율 ····· 33
제3절 가계금융복지조사 자가·무상가구 전세보증금 예측 및 의제주거소득 추정 ··· 35
제4절 가계금융복지조사 자가·무상가구 예측보증금 간접 검증 ····· 36

제5장 결론 및 시사점 ····· 38
제1절 연구 요약 및 시사점 ····· 38
제2절 연구 한계 및 향후 과제 ····· 40

참고문헌 ····· 41
부 록 ····· 42
Abstract ····· 48

요 약

본 연구의 목적은 자가주택이 제공하는 주거서비스 가치를 가구 소득에 포함하기 위해, 전월세신고제 자료와 머신러닝 기법을 활용하여 의제주거소득을 추정하는 방법을 제시하는 데 있다.

2021~2023년 국토교통부 전월세신고제 자료 중 전세계약 정보를 기본으로, 한국부동산원 주택가격동향조사의 전세·매매 평균가격과 국토부 건축물대장 정보 등 단지 특성을 결합하였다. 분석 대상은 공동주택(아파트, 다세대·연립, 오피스텔)과 단독주택(단독·다가구)으로 구분하고, 주택 특성(면적, 층, 건축연도), 단지 특성(대지, 건축, 연면적, 가구수), 지역시장 변수(전세·매매 평균, 실거래 단위금액)를 설명변수로 설정하여 유형별 전세보증금 예측모형을 구축하였다.

예측모형은 트리 기반 앙상블 기법인 랜덤포레스트(Random Forest)를 사용하였으며, 이상치 처리를 위해 단위면적당 전세금에 대해 IQR 계수 검증, 트리 수(n_estimators)에 대해서도 교차검증을 실시하여 예측 성능이 일정 수준을 유지하고, 계산 시간이 과도하지 않은 구간을 선택하였다.

최종 구축 모형에 대해 첫째 가계금융복지조사 전세가구에 적용하여 조사전세보증금과의 차이를 점검하였으며, 둘째 예측모형을 활용하여 자가·무상가구의 전세보증금을 예측하였으며, 예측된 전세보증금에 전월세 전환율과 이자율을 적용하여 의제주거소득을 추정하였다.

본 연구는 행정자료와 머신러닝 기법을 활용해 자가주택의 주거서비스 가치를 추정하는 예측모형을 구축하고, 이를 가계금융복지조사에 시범 적용하여 성능과 활용 가능성을 점검했으며, 주택·단지·지역시장 특성을 보다 세밀하게 반영하는 방법을 제시하고, 공공임대주택의 사회적 현물이전 가치의 추정 가능성을 제시했다는 점에서 의의를 갖는다. 향후 설명력을 높일 수 있는 변수 추가와 모형 고도화를 통해 의제주거소득 추정 및 관련 통계 작성에 활용될 수 있기를 기대한다.

주요 용어 : 의제주거소득, 머신러닝, 전월세실거래

제 1 장

서 론

제1절 연구 배경

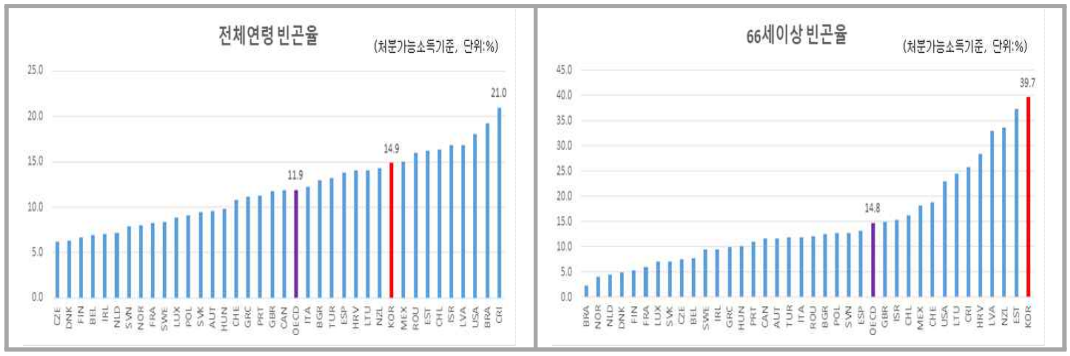
소득분배통계(지표)는 한 사회의 소득이 계층, 세대, 가구유형별로 어떻게 분포되어 있는지를 파악하고, 분배정책의 효과를 평가하기 위한 핵심 기초자료로 활용되어 왔다. 지니계수, 소득 5분위 배율, 상대적 빈곤율 등과 같은 대표적인 분배지표는 사회보험, 세금, 복지정책이 가계의 소득분포에 미치는 영향을 평가하는 데 사용되고 있으며, 고령층, 저소득층 등 특정 계층의 상대적 위치 변화를 살펴보는 데도 중요한 역할을 하고 있다.

우리나라에서는 이러한 공식 소득분배지표를 산출하기 위해 「가계금융복지조사」[1]를 활용하고 있다. 가계금융복지조사는 소득과 자산, 부채를 동시에 파악할 수 있는 조사로서, 분배지표뿐만 아니라 가계의 재무구조와 취약계층을 종합적으로 분석하는 데 활용되어 왔다. 그러나 기존 분배지표는 주로 임금, 사업소득, 이전소득 등 경상소득에 초점을 맞추고 있어, 자산 보유에 따른 서비스 흐름을 충분히 반영하지 못하였다고 볼 수 있다. 특히 주택과 같이 가계의 생활 수준에 큰 영향을 미치는 자산의 경우, 실제로는 주거서비스를 이용하고 있음에도 그 가치가 소득 개념에 포함되지 않는다는 한계가 존재한다.

이러한 한계는 고령층의 빈곤 문제를 해석할 때 더욱 뚜렷하게 드러난다. 고령층의 상당수는 주택 등 실물자산을 보유하고 있으나, 은퇴 이후 근로소득 감소와 공적연금 수준의 제약으로 인해 경상소득이 낮게 나타나는 경우가 많다. 이로 인해 소득 기준 빈곤율은 높게 나타나지만, 실제 생활 수준은 통계가 보여주는 것보다 더 나은 측면이 존재한다고 볼 수 있다. 반대로, 무주택이면서 전세(월세) 부담이 큰 계층은 주거비를 감안할 때 실질적인 후생 수준이 낮음에도, 단순 소득에 기반한 분배지표에서는 이러한 취약성이 충분히 드러나지 않는다는 문제가 있다.

---

1) 국가데이터처, 금융감독원, 한국은행이 공동 작성하는 국가승인통계(제93001호)



출저 : OECD data explorer('25. 12.)

<그림 1-1> OECD 회원국 상대적 빈곤율

한편 유럽연합(EU), 호주, 뉴질랜드, 미국 등 여러 OECD 국가에서는 이미 의제주거소득을 공식 통계 혹은 부가지표의 형태로 반영하고 있으며, 이를 통해 자가 보유에 따른 주거서비스 가치를 소득 개념에 포함시키고 있다.[2] 이러한 국제적 흐름과 국민계정(SNA)상의 주거서비스 계정 처리 방식 등을 고려할 때, 우리나라에서도 의제주거소득 도입을 검토할 필요성이 지속적으로 제기되어 왔다.

다만, 의제주거소득을 측정하는 작업은 개념적, 실무적 측면 모두에서 쉽지 않은 과제이다. 국민계정(SNA)에서 자가주거서비스를 포괄해야 한다는 원칙은 제시되어 있으나, 구체적인 산정 방식(임대료 상당액 접근법, 사용자비용 접근법 등)에 대해서는 국가 간에 통일된 국제 기준이 마련되어 있지 않아, 각국이 자료 여건과 정책 목적에 따라 상이한 방법을 선택·운영하고 있는 실정이다.

또한 의제주거소득을 구성하는 요소에 대한 조사(행정)자료의 가용성이 국가별, 시기별로 다르고, 전월세 전환율·이자율 등 핵심 가정에 따라 결과가 크게 달라질 수 있다는 점도 측정을 어렵게 만드는 요인이다.

위와 같은 제약에도 불구하고 의제주거소득을 계량적으로 측정하려는 노력은 충분한 의미가 있으며, 자가 보유 주택에서 발생하는 주거서비스 가치를 소득 개념에 반영하는 것은 고령층의 실질 생활 수준과 자산·소득 구조를 보다 정확하게 파악하는 데 필수적이라 할 수 있다.

2) Eurostat, Eusilc(소득/생활수준관련통계보고서), OECD설문(2015)

출저 : 「자산의 소득화를 반영한 노인빈곤율 산정 검토」(이병식 외, 2025), SRI 통계플러스.

<그림 1-2> 자산의 소득화 방법별 노인빈곤율

의제주거소득을 분배지표에 포함하여 분석할 경우, 단순 경상소득 기준으로는 포착되지 않던 계층 간 격차와 세대 간 불균형을 보다 다양하게 분석할 수 있으며, 이는 향후 소득분배 정책 등을 설계하는 과정에서 중요한 참고 지표로 활용될 수 있을 것이다.

제2절 연구 목적

본 연구의 목적은 국토교통부 임대차 실거래(전세) 자료와 머신러닝 기법을 활용하여 주택유형별(공동주택, 단독주택)로 전세보증금 예측모형을 구축하고, 이를 월세화하여 가계금융복지조사의 자가가구와 무상가구에 적용하는 것이다.[3] 이를 통해 자가가구의 주거서비스 가치를 시장 임대료 수준으로 평가할 수 있는 방법을 마련하고자 하였다.

구체적으로는 첫째, 전세 실거래 자료에 기반한 주택유형별 전세보증금 예측모형을 구축하고자 하였다. 이 과정에서 주택의 물리적 특성, 위치, 지역 주택시장 상황을 반영하는 설명변수를 구성하고, 머신러닝 모형을 활용하여 비선형 구조적인 전세보증금을 예측하고자 하였다.

둘째, 구축된 모형을 가계금융복지조사 전세가구에 적용하여 조사 전세금과 예측 전세금의 일치 정도를 검증함으로써, 예측모형의 신뢰도를 평가하고자 하였다.

3) 의제주거소득은 자가가구뿐만 아니라 전세(보증부월세)가구의 경우도 마찬가지로 전세금(보증금)은 주거비(월세)로 지출되지 않으므로 이에 대한 추가적인 의제주거소득 반영 필요성이 요구됨

셋째, 동일 모형을 자가가구에 적용하여 예측 전세금을 산출한 후, 전월세 전환율과 이자율을 적용해 월세 평가액을 추정함으로써 의제주거소득을 추정할 수 있는 체계를 마련하고자 하였다.

제3절 연구 범위 및 보고서 구성

본 연구의 범위는 2021년부터 2023년까지의 기간으로 한정하였으며, 국토교통부 임대차 실거래 자료 중 주거용 전세계약을 활용하여 전세보증금 예측모형을 구축하였다. 주택유형은 공동주택(아파트, 연립·다세대, 오피스텔)과 단독주택(단독·다가구)으로 구분하고, 각 유형별/연도별로 별도의 예측모형을 구축하였으며, 분석 범위는 전국 단위로 설정하였다.

예측모형에는 실거래자료에서 확인 가능한 주택 특성(면적, 층, 건축연도 등)과 건축물대장 기반 단지 특성, 주택가격동향조사의 전세/매매 평균가격 등 지역시장 변수를 포함하였고, 구축된 모형을 가계금융복지조사 소득연도 기준 2021년~2023년 전세, 자가, 무상 가구에 적용하여 의제주거소득을 추정하는 것까지를 본 연구의 범위로 정하였다.

보고서의 구성은 다음과 같다.

1장은 연구의 배경과 목적, 범위를 설명하여 전체 연구의 방향을 제시하였다.

2장에서는 의제주거소득의 개념과 측정 방법, 선행연구를 검토하여 본 연구가 어떤 점에서 차별화되는지를 정리하였다.

3장에서는 전세보증금 예측을 위한 자료 및 변수 구성, 전세 예측모형 작성, 평가지표를 바탕으로 최종 예측모형을 선정하는 과정을 제시하였다.

4장에서는 선정된 전세보증금 예측모형을 가계금융복지조사 전세가구에 적용하여 조사 전세금과의 비교를 수행하고, 최종적으로 예측모형에 자가·무상 가구를 적용하여 의제주거소득을 추정하는 절차를 단계적으로 제시하였다.

5장에서는 연구 결과를 요약하고 시사점을 제시하며, 작성 측면의 한계점과 향후과제 등을 제시하였다.

제 2 장

의제주거소득 개념 및 선행연구

제1절 개념 및 측정방법

의제주거소득(imputed rent)[4]은 자가주택에 거주하는 가구가 그 주택으로부터 얻는 주거서비스의 경제적 가치를 소득으로 간주하는 개념으로 정의할 수 있으며, 주택이 단순한 자산 취득을 넘어선 경제적 소득을 제공하고 있다는 점을 설명하고 있다.

소비자물가지수 작성 시에도 자가주거비지수(임대료 상당액)를 별도 항목으로 추정하여 주거비 부담에 반영[5]하고 있으며, 소득통계(가계금융복지조사)와 소비통계(소비자물가조사)에서의 차이만 있을 뿐 개념적으로는 의제주거소득과 자가주거비 모두 자가주택에서 발생하는 주거서비스 가치를 임대료 상당액 방식으로 평가한다는 점에서 동일한 개념으로 볼 수 있다.

의제주거소득 측정에는 크게 임대료 상당액 접근법, 사용자비용 접근법, 취득가격 접근법 등이 논의되어 왔다.

임대료 상당액 접근법은 시장에서 관측되는 유사 주택의 임대료를 기준으로 의제임대료를 산출하는 방법으로, 개념이 직관적이고 시장 정보를 직접 활용한다는 장점이 있다. 다만 충분한 임대료 자료가 확보되어야 하고, 자가주택의 특성과 정확히 일치하는 비교 대상이 존재하지 않는다는 한계가 있다.

사용자비용 접근법은 주택자산의 기대수익과 유지비용, 감가상각, 세제 등을 종합적으로 고려하여 순수한 이용비용을 계산하는 방식으로, 주거 유지의 실질적 비용을 측정한다는 장점이 있으나, 비용 산정 방식에 따라 편차 발생 및 실무적으로 계산이 복잡하다는 단점이 있다.

취득가격 접근법은 주택의 취득가격이나 공시가격을 기준으로 의제주거소득을 평가하는 방법으로, 직관적이고 자료 접근성이 높은 장점이 있으나, 현재 시장가치와 괴리가 발생할 수 있는 단점이 있다.

---

4) 의제주거소득(imputed rental income of owner-occupied housing)

5) 현재 소비자물가조사 주지표에는 미포함 되어있으나, 자가주거비 포함 지수를 보조지표로 공표 중임

<표 2-1> 의제주거소득 측정방법

방법	산출방식	장점	단점
임대료 상당액	유사한 조건의 주택 시세/임대료 기준	직관적, 시장정보 활용 용이	시장 임대료 자료 부족 시 사용 곤란
사용자비용	주택자산의 기대총수익에서 유지비용과 감가상각을 차감	주거 유지의 실질적 비용	비용 산정방식에 따라 편차 발생
취득가격	신규주택의 구매/취득가격	직관적	신규주택 중심, 현재가치 반영 미흡

출처 : ICLS2003, 호주통계청, 한국은행 OECD 자본측정 메뉴얼

국제 논의(ICLS 2003, 국제노동통계회의)에서는 실무적인 편의성과 국민계정(SNA)과의 정합성을 고려해 임대료 상당액 접근법을 주로 권고하고 있으며, 본 연구에서도 전월세 실거래 자료를 활용하는 임대료 상당액 접근법을 기본 방향으로 선택하였다.

특히, 전세 실거래 자료를 활용하여 자가주택의 의제주거소득을 추정하는 방식을 선택하였는데, 이는 전세제도가 우리나라에만 존재하는 특별한 임대차 방식이며, 전세가격이 자가주택의 가치와 해당 주택에서 소비되는 주거서비스 수준을 잘 반영한다고 판단하였다.[6)]

제2절 선행연구 및 차별성

가계금융복지조사에서 의제주거소득을 추정하는 연구는 과거 통계개발원(현 국가데이터연구원)과 복지통계과를 중심으로 진행되어왔다. 이들 연구는 자료 접근 가능성, 계산의 용이성, 국제 비교 가능성을 동시에 고려하여 다양한 시도를 수행하였다.

가. 가계금융복지조사에서 의제주거소득 반영을 위한 추정방법 연구 (통계개발원, 2017년)

가계금융복지조사 자가가구를 대상으로 지오코딩(주소→지도) 방법을 활용하여, 주변 전·월세 실거래 자료를 수집하고, 역거리 가중치를 사용하여 전세가격을 추정하는 방법을 제시하였다.

이 연구는 개별 자가주택 주변의 실제 임대시장 정보를 활용함으로써 임대료 상당액으로

6) 의제주거소득 추정 시 전세자료 활용에 대해서는 제3장 제2절에서 추가 논의

전세가격을 산출하려는 방법으로, 추정된 전세가격에 전월세 전환율을 적용해 의제주거소득을 추정하였다. 다만, 당시에는 일부 지역(부산, 충남)을 중심으로 시범 분석이 이루어졌고, 공간추정프로그램의 실무 적용 부담 및 거래건수 등 지역별 편향 발생 가능성이 한계점으로 나타났다.

나. 순의제주거소득 추정방법 연구(통계개발원, 2020년)

2020년 연구에서는 공시가격, 실거래 매매가격, 조사된 시장가격 등 다양한 기준을 활용하여 의제주거소득을 추정하고, 재산세와 수선비, 담보대출이자 등을 차감한 순의제주거소득을 산출하였다.

해당 연구는 소득 추정에 있어서는 공시가격 자료를 활용하는 방법의 장점을 강조하는 한편, 순의제주거소득을 위한 비용추정은 소득추정 결과를 이용해 재산세로 환산하는 것이 정합성이 높을 것으로 판단하였다.

다만, 공시가격 및 실거래 자료 매칭 시 시간 소요, 거래지역별 편향(실거래) 발생 가능성이 한계점으로 나타났다.

다. 복지통계과 자체 연구(2018년~2019년, ~현재)

복지통계과 자체 연구에서는 2016년, 2017년 조사에 모두 참여한 가구 중 자가 가구에 대해 월세평가액을 조사(2016년)한 후 2017년에 조사한 수선비와 재산세를 차감하여 순의제주거소득을 산출하였으며, 의제주거소득을 포함할 경우 노인가구와 자가 비중이 높은 계층의 소득 분포가 상향 이동하고, 지니계수, 상대적 빈곤율 등이 개선되는 경향을 보이고 있는 것으로 나타났다.[7)]

본 연구는 기존 연구가 제한된 지역 또는 특정 자료에 의존하였던 것과 달리, 국토교통부 임대차 실거래(전세) 자료를 활용하고, 국토부 건축물대장 정보를 추가하였으며, 머신러닝 기법을 결합하여 전세보증금 예측모형을 구축하고자 하였다.

또한 한국부동산원 주택가격동향조사에서 제공하는 전세평균, 매매평균 등 지역시장 변수를 주택 특성과 함께 고려함으로써, 각 주택이 위치한 지역의 시장 상황까지 예측모형에 반영되도록 설계하였다. 이로써 주택의 물리적 특성뿐만 아니라 지역 간 전세시장 구조와 가격 수준의 차이가 의제주거소득에 반영될 수 있도록 하였다.

마지막으로 본 연구는 트리 기반 앙상블 기법인 랜덤포레스트(Random Forest) 모델을 중심으로 머신러닝 방법을 활용함으로써, 전세보증금과 설명변수 간의 비선형 관계를 포착하고, 대규모 자료에서도 안정적인 예측력을 확보하고자 하였다.

---

7) 자체적으로도 과거 유사 연구 방법을 참고하여 국토부 매매가격 및 전월세실거래 자료를 활용하는 방안을 연구하고 있음

제3절 머신러닝 활용 연구 사례

가. XGBoost 기반 부동산 자동가치산정모형(AVM)의 실증분석(이선구, 2025년)

전통적인 감정평가(비교 사례 등) 한계를 보완하기 위해, 머신러닝 기반 AVM의 가능성을 검토하였으며, XGBoost 회귀모델을 활용하여 서울 은평구 다세대주택 실거래 자료를 활용하여 매매 거래금액을 예측하였다.

XGBoost 기반 AVM이 다세대주택과 같은 이질적 주택에 대해 실무에서 쓸 수 있는 정도의 예측력을 보인다고 하였다(R^2=0.747, MAPE=13.6%).

종속변수	매매거래금액(2023.11.~2024.10. 은평구 다세대 실거래 1,839건)
설명변수	물리적(면적, 층 등), 입지(인구수, 세대수 등) 시장및경제(매매실거래, 금리 등)

나. 기계학습과 설명가능한 인공지능 모형을 이용한 강남의 전세가격 결정요인 분석(김태영 외, 2023년)

전통적인 선형회귀 대신 XGBoost 같은 기계학습 모형을 활용하여 전세가격을 더 정확하게 예측하고, 어느 변수가 전세가격에 얼마나 영향을 주는지 설명하였다.

선형회귀, 의사결정나무, 랜덤포레스트, XGBoost와 예측 성능을 비교하였으며, XGBoost가 가장 높은 예측력을 보인 것으로 나타났다(R^2=0.79, MAPE=14%).

종속변수	단위면적당 전세가격(2018년~2022년 강남구 아파트 전세거래가 15,000여 개
설명변수	층, 세대수, 병원수, 학교수, 인근 공원 여부 등

다. 머신러닝을 활용한 김해시 주택가격변화 예측을 위한 모델링(이현기, 2024년)

경남 김해시의 주택매매가격지수 변화를 예측하기 위해 머신러닝 모델을 만든 사례 연구를 실시하였으며, 선형회귀, 랜덤포레스트, XGBoost 방법을 비교하였다. 분석결과 랜덤포레스트 방법이 가장 높은 예측력을 보인 것으로 나타났다(R^2=0.80, MSE=8.16).

종속변수	김해시 주택매매가격지수(2014년~2024년)
설명변수	인구, 세대수, 수출액, 기준금리, 코스피 지수 등 경제, 인구, 주택시장 지표

위와 같이 다양한 대상과 지역에 대한 연구결과 머신러닝 모델들은 물리적 특성부터 경제지표까지 폭넓은 변수를 학습하여 R^2=0.75~0.80 수준의 예측력을 보였다. 이는 복잡하고 이질적인 부동산 시장의 가격 결정 요인을 분석하는 데 있어 머신러닝 기반의 접근 방법이 실무적으로 활용 가능한 방법임을 시사한다.

제 3 장

전세보증금 예측모형 구축 및 성능 평가

제1절 활용자료

본 연구에서 사용한 자료는 다음과 같다.

가. 주택 임대차계약 신고제 자료(국토교통부 실거래가공개시스템)

국토교통부가 운영하는 주택 임대차계약 신고제(전월세 신고제) 자료를 주요 자료로 활용하였다. 주택 임대차 신고제는 임대차 시장의 정보 비대칭을 완화하고 임차인 권익을 보호하기 위해 도입된 제도로, 보증금 6천만 원 초과 또는 월 임차료 30만 원을 초과하는 주택 임대차 계약에 대해 계약 체결일로부터 30일 이내에 임대인 또는 임차인이 해당 계약 내용을 지자체 또는 부동산거래관리시스템(rtms.molit.go.kr)에 신고하도록 의무화한 제도이다.[8)]

<표 3-1> 주택 임대차계약 신고 대상

신고 대상	신고 예외
- 보증금 6,000만 원 초과 또는 월세 30만 원 초과인 주택 임대차 계약 (경기도 외 군 지역 제외) - 보증금 또는 월세가 변경된 갱신 계약	- 보증금 6,000만 원 이하 및 월세 30만 원 이하 계약 - 보증금과 월세 변동 없는 갱신 계약 - 공공임대주택, 기숙사, 고시원 등 일부 주택 유형 - 전대차(재임대)나 무상사용 계약

2021년 6월 1일 시행 이후 일정 기간 과태료 부과를 유예하는 계도기간이 시행되었으며, 계도기간이 2025년 5월 31일자로 종료됨에 따라 2025년 6월 1일 이후 체결되는 신고 대상 임대차 계약에 대해 미신고 시 법적 제재가 가능하게 되었다. 이러한 신고제 자료는 신고 대상 금액 기준을 충족하는 임대차 계약에 대한 실제 거래정보(보증금,

8) 부동산 거래신고등에 관한 법률(21.6.1.) 6조의 2(주택임대차 계약의 신고). 국토교통부 누리집.

월세액, 계약기간, 주택유형, 면적, 주소 등)를 체계적으로 축적하고 있어, 임대차 시장의 구조와 가격 수준을 분석하고 의제주거소득 수준을 추정하는 데 기초자료로 활용될 수 있다.

실거래가공개시스템에 등록된 2021년, 2022년, 2023년 전월세 실거래 신고자료 건수는 아래 표와 같다.

전국 국토부 아파트,연립,다세대, 오피스텔 전세 실거래가 조회 현황(2023.01.01.~2023.12.31.)

시군구	번지	도로조	계약면	단지명	전월세구	전용면적	계약년	계약일	보증금(만	층	건축년	도로명	계약기	계약구	주택유
경기도 부	104			트윈파크	전세	75.6518	202312	31	30,000	7	2004	부광로36번	202402~2	신규	아파트
경기도 부	104			트윈파크	전세	75.6518	202312	31	30,000	7	2004	부광로36번	202402~2	신규	아파트
경기도 고	2570			킨텍스휴먼	전세	84.3669	202312	31	28,000	14	2003	대화1로 6	-	-	아파트

<그림 3-1> 전월세 신고제 다운로드 자료

<표 3-2> 전월세 신고제 자료 건수

(단위: 건, %)

	주택 유형	월세	전세	합계	주택유형별(가로)			임차유형별(세로)	
					월세	전세	합계	월세	전세
2021년	단독	302,949	209,516	512,465	59.1	40.9	100.0	33.4	18.6
	다세대	103,141	178,507	281,648	36.6	63.4	100.0	11.4	15.8
	아파트	399,659	647,563	1,047,222	38.2	61.8	100.0	44.1	57.4
	오피스텔	100,370	92,689	193,059	52.0	48.0	100.0	11.1	8.2
	합계	906,119	1,128,275	2,034,394	44.5	55.5	100.0	100.0	100.0
2022년	단독	379,905	181,574	561,479	67.7	32.3	100.0	32.5	16.0
	다세대	121,357	164,604	285,961	42.4	57.6	100.0	10.4	14.5
	아파트	533,514	685,986	1,219,500	43.7	56.3	100.0	45.6	60.6
	오피스텔	134,837	99,941	234,778	57.4	42.6	100.0	11.5	8.8
	합계	1,169,613	1,132,105	2,301,718	50.8	49.2	100.0	100.0	100.0
2023년	단독	349,165	129,798	478,963,	72.9	27.1	100.0	29.4	12.2
	다세대	134,132	127,522	261,654	51.3	48.7	100.0	11.3	12.0
	아파트	556,140	720,779	1,276,919	43.6	56.4	100.0	46.9	67.6
	오피스텔	146,340	87,935	234,275	62.5	37.5	100.0	12.3	8.2
	합계	1,185,777	1,066,034	2,251,811	52.7	47.3	100.0	100.0	100.0

* 단독 : 다가구 포함, ** 다세대 : 연립 포함

2021년에는 전세 신고 건수가 더 많았으나, 2022년부터는 월세 신고 건수가 더 많아졌으며, 주택유형별로는 전세와 월세 모두 아파트의 비중이 높은 것으로 나타났다.

나. 한국부동산원 주택가격동향조사

시도 및 시군구별 지역 주택시장 수준을 반영하기 위해 한국부동산원이 작성하는 주택가격동향조사[9] 자료를 활용하였다. 아파트의 경우 시군구별 전세 및 매매 평균가격을 활용하였으며, 이외 주택(단독, 다세대(오피스텔 포함))의 경우는 시도별 전세, 매매 평균가격 자료만 활용하였다.

다. 국토부 건축물등록대장 행정자료

전월세 신고자료에는 건축연도 변수가 포함되어 있으나, 가계금융복지조사 조사항목에는 건축연도가 별도 조사되지 않으며, 추가로 건축물대장에 있는 대지면적, 건축면적, 연면적, 가구수 변수를 활용하기 위해 국토부 건축물대장 행정자료를 활용하였다.

대지면적, 건축면적, 연면적, 가구수 정보를 전월세 신고자료 및 가계금융복지조사 데이터에 PNU[10]코드를 활용하여 매칭하였으며, 각각 매칭률은 아래 표와 같다.

<표 3-3> 건축물대장 정보 매칭률

(단위: %)

공동주택	전월세 신고자료				가금복자료			
	대지면적	건축면적	연면적	가구수	대지면적	건축면적	연면적	가구수
다세대	72.0	81.6	83.7	76.6	54.0	73.0	73.8	73.3
아파트	73.0	81.7	83.8	76.3	69.0	77.7	79.5	75.8
오피스텔	73.0	81.3	83.3	76.6	67.5	70.0	70.3	22.8
총합계	72.7	81.5	83.6	76.5	65.8	76.5	78.1	74.0

9) 전국주택가격동향조사는 전국 263개 시·군·구를 대상으로 주택 매매·전세 가격지수를 산출하는 국가승인통계임

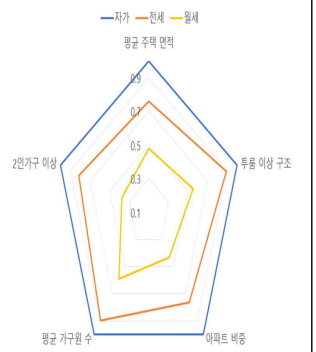
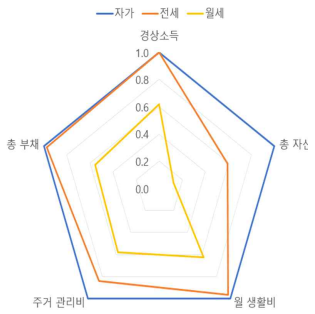
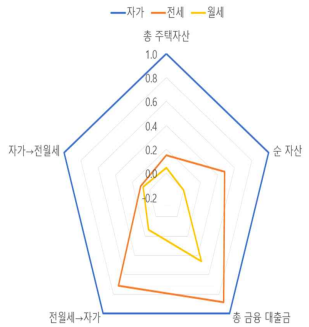
10) PNU코드(Parcel Number Code 필지별 토지 식별코드로서 이후 제5절에서 추가 설명)

제2절 전세 신고자료

실거래 자료 중에서도 본 연구는 전세 자료만을 활용하였다. <표 3-2>를 보면 연간 전월세 거래 건수(2023년)는 약 225만 건을 상회하며, 그중 전세는 약 106만 건(47.3%)으로 모형 학습에 충분한 표본을 제공한다.

전세가구는 자가가구와 주택 및 가구 특성, 경제적 특성 측면에서 월세가구보다는 상대적으로 유사한 집단을 이루고 있다는 점[11]에 주목하였기 때문이다. 자가가구가 현재 거주하는 주택을 동일한 조건으로 전세를 줄 경우 형성될 가격 수준이 자가가구의 주거서비스 가치를 잘 반영한다고 볼 수 있기 때문에 전세가격을 기준으로 자가가구의 의제주거소득을 산출하는 방법을 활용하였다.

<그림 3-2>는 점유형태별 주택, 가구, 경제적 특성을 비교해 파란색 자가가구특성을 1로 표준화하여 전세와 월세가구가 얼마나 유사한지를 나타낸 방사형 그래프이다. “주택/가구 특성”을 보면 파란색 자가가구와 주황색 전세가구는 유사한 반면, 노란색 월세가구는 매우 다른 형태를 보이고 있다. “경제적 특성”에서도 총자산에서 월세는 다른 형태이며, 전세는 자가와 유사한 형태를 보이고 있다(변준석 외, 2024).

주택/가구 특성	경제적 특성1	경제적 특성2
		

출저 : 행정자료활용 주거비지수 작성 및 실무적용방안(변준석 외, 2024)

<그림 3-2> 자가가구와 전세가구의 유사성

또한 전세가구는 월세가구에 비해 행정자료에 포착되는 비율이 상대적으로 높고, 신고 누락 가능성이 낮을 것으로 판단하였다. 아래 <표 3-4>를 보면 전세계약은 약 87%, 월세계약은 약 59% 정도의 신고율을 보이는 것으로 판단된다.[12] 전세계약은 일

11) 행정자료활용 주거비지수 작성 및 실무적용방안(변준석과 외, 2024), 주거실태조사 결과

12) 국토교통부에서도 임대차계약 신고율은 전체 임대차 건수를 알 수 없으므로 위와 같은 방법으로 추정함

정 규모 이상의 계약이 중심을 이루기 때문에 자신의 보증금을 안정적으로 지키기 위해서라도 더욱 적극적으로 신고를 할 것으로 판단되며, 이를 통해 보다 안정적으로 자료를 파악할 수 있다는 장점이 있다. 이러한 이유로 본 연구에서는 전월세 실거래 자료 중 전세 자료만을 활용하였다.

<표 3-4> 전세/월세 가구별 임대차 계약 신고율 추정

주택*		20년 인총(20% 표본)			확정일자 전기간 ('11.1 ~ '21.5)		확정일자 최근 2년('19.6 ~ '21.5)		임대차신고제 ('21.6 ~ '23.2)	
		가구수	비율	1/24(A)**	월평균 신고건수(B)	신고율 (B/A)	월평균 신고건수(C)	신고율 (C/A)	월평균 신고건수(D)	신고율 (D/A)
자가		11,989,186	57.3%							
무상		781,003	3.7%							
임차	월세	4,904,901	23.4%	204,371	59,663	29.2%	67,134	32.8%	121,041	59.2%
	전세	3,251,620	15.5%	135,484	87,905	64.9%	103,616	76.5%	117,504	86.7%
	소계	8,156,521	39.0%	339,855	147,568	43.4%	170,750	50.2%	238,545	70.2%
합계		20,926,710	100%							

* 신고대상 가구(8,156,521)= 20년 인총 총 가구(20,926,710) - 자가(11,989,186) - 무상(781,003) (주택 주택이외의 거처 모두 포함)

** 전월세 계약기간이 통상 2년(24개월)이므로 월평균 계약 가구로 추정

출저 : 행정자료활용 주거비지수 작성 및 실무적용방안(변준석 외, 2024)

제3절 변수 선택 및 전세데이터 전처리

전세가격에 영향을 미치는 요인은 ① 주택 특성(면적, 층, 건축연도, 주택유형 등), ② 단지 특성(대지면적, 건축면적, 연면적, 주차대수, 총 세대수 등), ③ 주변환경(교통접근성, 학군, 유해시설 존재 여부 등), ④ 지역시장 요인(지역 시세, 전세·매매 가격, 전세가율 등), ⑤ 거시정책 요인(금리 수준, 전세대출 규제 등)으로 나누어 볼 수 있다. 본 연구에서는 실제 통계 작성 및 행정자료 결합을 통한 업무 적용 가능성을 고려하여 ① 주택 특성, ② 단지 특성, ④ 지역시장 요인을 중심으로 설명변수를 구성하였다. 이를 통해 자료 가용성 범위 내에서 전세가격 형성에 중요한 요인을 최대한 반영하도록 모형을 설계하고자 하였다.

전세 실거래 자료를 기반으로 머신러닝 학습을 위해 설명변수(feature)와 종속변수(target)를 구분하였다. 분석 대상 주택은 공동주택(아파트, 다세대, 오피스텔)과 단독주택(다가구 포함)으로 구분하였으며, 단독주택의 경우 실거래 자료에 「층」 정보가 존재하지 않아 해당 변수는 공동주택에만 포함하였다.

<표 3-5> 설명변수 및 종속변수

구분		변수명
종속변수		전세보증금(만 원)
설명변수	지역시장	시도/시군구/읍면동
		전세평균, 매매평균
		실거래 단위금액 평균(전세금액/㎡)
	주택특성	면적(㎡)
		층(단독주택 제외)
		건축연도
		주택유형(단독, 아파트, 다세대, 오피스텔)
	단지특성	건축물대장(대지면적, 건축면적, 연면적, 가구수) (단독 제외)

* 다세대 : 연립포함, ** 단독 : 다가구포함, *** 실거래 단위금액 평균: 읍면동 및 주택유형별

전세 자료 취합 이후 전처리 단계에서는 결측치 및 이상치 제거를 통해 자료를 정제하였다. 건축연도와 층수 변수의 경우 일부 결측치는 삭제하였으며, 계약연도 이후에 건축된 것으로 신고된 거래 (예를 들어 2021년 신고자료임에도 건축연도가 2021년보다 나중인 것으로 기록된 사례) 등은 입력 오류로 간주하여 제거하였다. 아울러 향후 PNU 코드 생성을 위해 필수적인 지번 정보가 전혀 기재되지 않은 거래는 다른 행정자료와의 연계가 불가능하다고 판단하여 삭제하였다.[13] 연도별 결측치 제거 후 연도별 전세 실거래 자료 수는 아래 표와 같다.

<표 3-6> 결측치 및 이상치 제거 후 전세 자료 건수

(단위: 건, %)

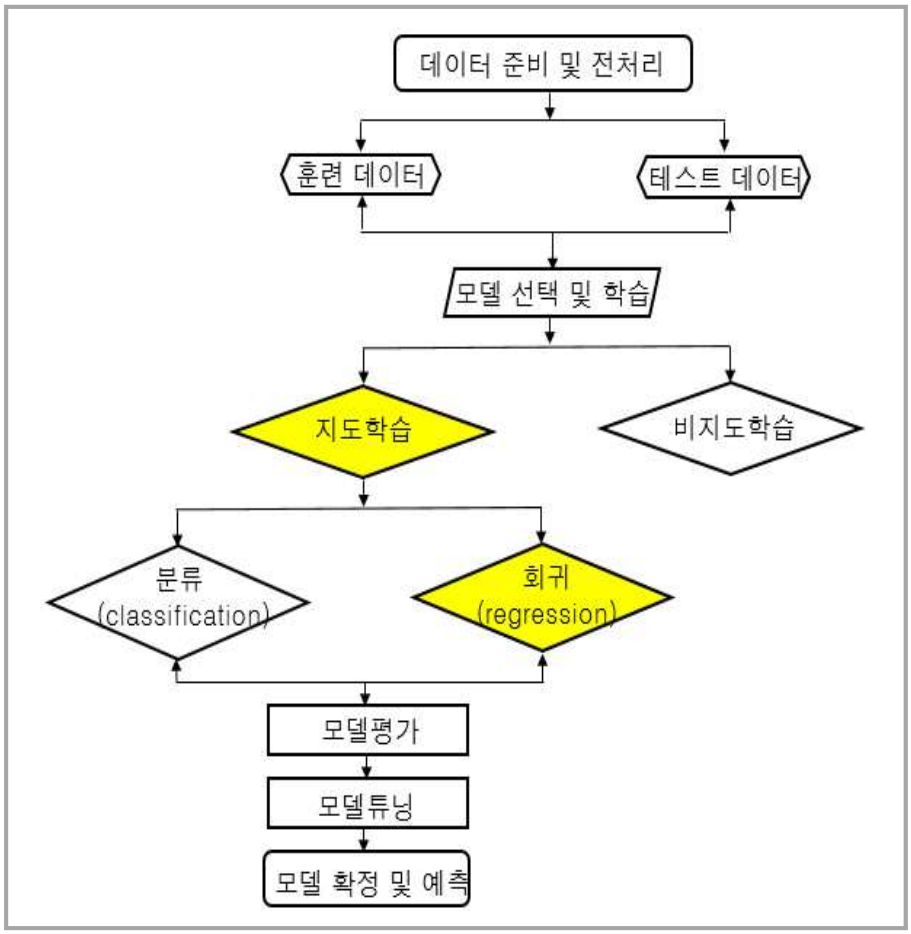
	주택 유형	2021년		2022년		2023년	
		건수	비율	건수	비율	건수	비율
전세	단독	191,470	17.4	166,980	15.0	116,607	11.1
	다세대	171744	15.6	164,424	14.8	127,507	12.1
	아파트	643,756	58.6	681,741	61.3	719,186	68.5
	오피스텔	91,490	8.3	98,935	8.9	87,029	8.3
	합계	1,098,460	100.0	1,112,080	100.0	1,050,329	100.0

* 활용 건수(3,260,869, 98.0%) = 전세 총 건수(3,326,414) - 결측 건수(65,545)

13) 신고금액 자체의 이상치 판단은 현 단계에서는 하지 않았으며, 예측모형 개선을 위한 단계에서 수행함

제4절 머신러닝 모형 선정 및 평가지표

머신러닝 분석의 일반적인 절차는 아래 그림과 같이 데이터 준비 및 전처리, 데이터 분할, 모델 선택 및 학습, 모델 평가, 모델 튜닝, 최종 모델 확정 및 예측의 단계로 정리할 수 있다.



<그림 3-3> 머신러닝 학습과정

전처리 이후 전체 자료를 훈련용(train)과 평가용(test)으로 분할하였으며, 본 연구에서는 훈련과 검증을 8:2 비율로 설정하였다. 훈련자료로 모형을 학습한 뒤 평가데이터를 이용하여 평가지표를 계산한다. 모형은 트리 기반 앙상블 기법인 랜덤포레스트(Random Forest)를 기본 모형으로 사용하였다.

본 연구에서 사용된 랜덤포레스트는 다수의 의사결정나무를 무작위로 생성한 후 그 예측값을 평균함으로써 개별 트리의 과적합을 완화하고 안정적인 예측 성능을 확보할

수 있는 모형이다. 또한 변수 중요도를 산출할 수 있어 전세보증금에 영향을 미치는 주요 요인을 파악하는 데에도 유용하다.[14)]

Probst 등(2019)은 랜덤포레스트의 하이퍼파라미터가 성능에 미치는 영향을 분석한 결과, 기본값만으로도 비교적 양호한 성능을 달성할 수 있으며, 제한적인 범위의 튜닝만으로 성능을 추가 개선할 수 있다는 점을 강조하였다.

Bentéjac 등(2021)은 XGBoost, LightGBM, CatBoost와 Random Forest를 비교한 결과 CatBoost가 평균적으로 가장 높은 정확도를 보였으나, 다른 앙상블 기법과의 차이는 크지 않으며, Random Forest는 기본 설정만으로도 안정적인 성능을 보인다는 점을 설명하였다.

이러한 이론적·실증적 근거를 볼 때, 본 연구와 같이 다양한 설명변수와 실제 통계작성 가능성 등을 감안하여 예측 성능, 안정성, 튜닝 부담, 해석 가능성 간의 균형을 고려하여 랜덤포레스트를 주요 예측모형으로 채택하였다.

본 연구에서는 실제 업무 적용 가능성을 고려하여 복잡한 하이퍼파라미터 탐색보다는 기본 옵션을 유지하되 트리 수(n_estimators) 변경(100→200)과 면적(㎡)당 전세금에 대한 이상치 제거 시 IQR 옵션값(±1.0, ±1.25)을 활용한 이상치 제거 수준 정도의 튜닝을 실시하여 성능 변화를 살펴봄으로써, 예측력과 안정성 간의 균형이 적절한 모형을 선택하였다. 계산된 평가지표에 따라 최종 모형을 선택하였으며, 선택된 모형을 사용하여 전세보증금을 예측하였다.

<표 3-6> 하이퍼파리미터 설정 값

옵션	값
n_estimators(트리수)	200
IQR 계수값	±1.0, ±1.25
max_depth(트리깊이)	None[15)]
min_samples_split(분할 최소샘플)	2
min_samples_leaf(리프노드 최소값)	1
max_features(요인)	None
random_state(재현성)	30

14) Random Forests(Breiman, 2001).
단독주택가격 추정을 위한 기계학습 모형의 응용(이창와 박기호, 2016).

15) 트리 깊이가 깊으면 과적합 가능성의 문제가 있으나 본 연구에서는 우선 전세보증금 예측모형의 활용 가능성과 의제주거소득 산정에의 적용 가능성을 검토하는 데 초점을 두었기 때문에, 트리 깊이에 별도의 제한을 두지 않은 상태에서 기본 설정을 유지하였으며, 향후 연구에서는 트리 깊이를 포함한 하이퍼파라미터를 보다 구체적으로 조정함으로써 과적합 위험을 조정할 필요가 있음

머신러닝에 의한 회귀예측모형 평가는 일반적으로 다음과 같이 4가지 지표를 중심으로 이루어지며, 지표를 종합적으로 활용하여 모형의 설명력과 예측오차 수준을 검토하였다.

<표 3-7> 회귀 예측모형 평가지표

평가지표명	설명	산식
결정계수(R^2)	종속변수의 총 변동 중 모형이 설명하는 비율을 나타내는 지표로, 값이 1에 가까울수록 설명력이 높다고 볼 수 있음	$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$
평균절대백분율오차 (MAPE)	실제값 대비 예측값의 상대적 오차를 백분율로 나타내어, 예측의 정확도를 직관적으로 보여주는 지표(%)	$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left\lvert\frac{y_i - \widehat{y_i}}{y_i}\right\rvert$
평균절대오차 (MAE)	예측오차의 절대값을 평균한 값으로, 전세보증금 단위에서 평균적으로 어느 정도의 오차가 발생하는지 파악	$MAE = \frac{\sum_{i=1}^{n}\lvert\widehat{y_i} - y_i\rvert}{n}$
평균제곱근오차 (RMSE)	오차 제곱의 평균에 제곱근을 취한 값으로, 큰 오차에 보다 큰 가중치를 부여한다는 특징을 가지며, MAE보다 큰 오차에 민감	$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\widehat{y_i} - y_i)^2}$

<y_i = 실제값, $\widehat{y_i}$= 예측값, $\overline{y}$= 실제값의 평균>

제5절 랜덤포레스트 모형 학습 및 평가지표 결과

전세 실거래 자료에 대해 기본 전처리 후 머신러닝 학습을 실시하고, 평가지표를 산출하였다. 1단계 방법은 실거래 자료를 기본 결측치만 제거하고 원자료 그대로 활용한 것이며, 2단계 방법은 이상치 제거를 위해 1.5×IQR 범위까지만 활용한 자료이다.[16] 3단계 방법은 추가로 전세평균, 매매평균, 실거래 평균 단위금액을 변수로 추가한 자료이다.

16) 전세보증금 자료 중 백만 원 미만인 자료도 있으며, 이는 착오, 공공임대 또는 사인 간 거래 등으로 판단되어 이러한 자료를 제거하기 위해 일반적으로 사용하는 사분위수 범위(IQR)의 ±1.5배를 기준으로 이상치를 제거하였다(금액기준: 시도/시군구/읍면동/주택유형별 ㎡ 금액).

1단계에서 2단계로 갈수록 R^2는 소폭 상승하고, MAPE, MAE, RMSE는 전반적으로 개선되는 것으로 나타났다.

3단계에서는 공동주택과 단독주택 모두에서 R^2가 가장 높고 오차지표(MAPE, MAE, RMSE)가 가장 낮아, 지역시장 변수 결합을 통해 모형 성능이 추가로 개선되는 효과가 확인되었다.

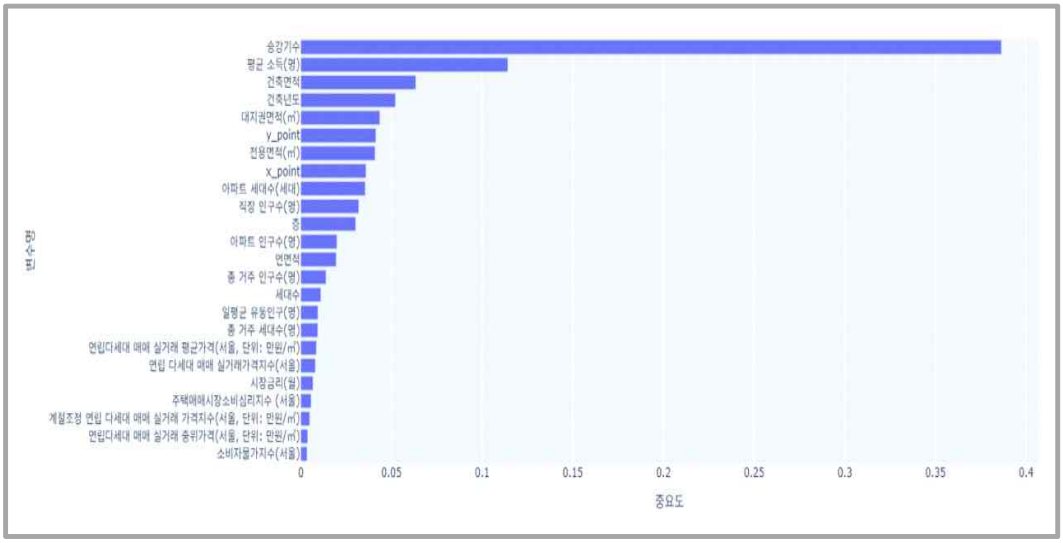
<표 3-8> 주택유형/연도별 전세보증금 예측모형(1-2-3단계) 성능비교

(단위: %, 만 원)

주택	년도	변수설정		R^2	MAPE	MAE	RMSE
공동주택	2022	1	실거래 원자료	0.913	14.455	3,582	6,643
		2	①+1.5*IQR	0.913	14.256	3,561	6,613
		3	②+(전세/매매 평균, 실거래 평균 단위금액)	0.921	12.968	3,386	5,982
	2023	1	실거래 원자료	0.939	11.558	2,929	5,534
		2	①+1.5*IQR	0.940	11.337	2,901	5,508
		3	②+(전세/매매 평균, 실거래 평균 단위금액)	0.947	10.168	2,726	4,968
단독주택	2022	1	실거래 원자료	0.751	26.931	2,506	4,121
		2	①+1.5*IQR	0.762	26.296	2,464	4,026
		3	②+(전세/매매 평균, 실거래 평균 단위금액)	0.778	25.154	2,318	3,727
	2023	1	실거래 원자료	0.746	26.473	2,579	4,244
		2	①+1.5*IQR	0.764	25.406	2,500	4,094
		3	②+(전세/매매 평균, 실거래 평균 단위금액)	0.776	24.070	2,350	3,741

성능개선을 위해 건축물대장 변수를 추가하고, IQR 계수값 조정, 트리수를 증가하여 머신러닝 학습을 실시하였다.

첫 번째로 건축물대장 변수를 추가하였다. 이선구(2025)는 「XGBOOST 기반 부동산 자동가치산정모형의 실증 분석」에서 연립/다세대의 거래금액을 예측하기 위해 여러 설명변수를 사용하였는데 변수 중 승강기수, 평균소득, 건축면적, 대지면적, 연면적 등이 주요 변수로 확인되었다.



출처 :「XGBOOST 기반 부동산 자동가치산정모형의 실증 분석」(이선구, 2025), 변수중요도 인용

<그림 3-4> 연립/다세대의 거래금액 예측변수 중요도

내부 업무시스템인 데이터융복합관리체계 시스템을 활용하여 국토교통부 건축물등록대장자료 중 대지면적, 건축면적, 전용면적, 세대수 변수를 공동주택에 한하여 추가하였다.[17)]

건축물대장자료와 전세 실거래 자료를 매칭시키는 key값으로는 PNU코드를 활용하였다. PNU코드(필지별 토지 식별코드)는 시도(2자리), 시군구(3자리), 읍면동(3자리), 리(2자리), 필지구분(1자리), 본번(4자리), 부번(4자리)을 나타내는 19자리 숫자 체계로 건물마다 부여되며, 행정동 체계와 달리 상대적으로 오랜기간 잘 변하지 않는 특성이 있다(변준석 외, 2023).

건축물등록대장과 전세 실거래 자료의 PNU코드는 위 코드 체계에 의해 쉽게 생성할 수 있으나, 가금복 자료의 경우 현장 조사를 위한 행정동 기반 주소체계로 되어 있기 때문에 법정동 지번 변경 및 PNU코드 생성을 위해 내부 시스템인 SGIS WORKS 시스템을 활용하였다.[18)]

일부 주택의 건축연도 정보는 건축물대장에 존재하지 않거나 확인이 어려워 한국토지주택공사(LH)가 운영하는 See:Real 부동산정보포털(seereal.lh.or.kr)을 참고하여 보완하였다.

---

17) 실거래 자료 중 공동주택에는 지번정보가 있지만, 단독주택에는 지번정보가 있지 않아 매칭 킷값(PNU코드)을 만들 수 없음

18) SGIS WORKS 시스템은 SGIS DB 구축에 필요한 데이터를 가공 및 적재, 관리하는 시스템으로, 행정주소 및 도로명 주소를 활용해서 PNU코드 생성이 가능함

단, 앞서 1절에서 살펴보았듯이 건축물대장 변수 자료와 실거래 변수 자료 사이에는 약 78% 정도의 매칭률을 보이므로, 건축물대장 변수 자료에 약 22%의 결측치가 존재한다.

이에 설명변수 중 결측이 확인된 건축물대장 변수에 한해서는 센티널(−1) 대치 + 결측 플래그(변수명_결측)방법을 적용하였다. 이는 랜덤포레스트 모델이 결측값을 직접 허용하지 않는 제약을 해소하기 위함이며, 학습 시 결측 표본은 값을(−1), 플래그를(1)로 입력하고, 비결측 표본은 실측값, 플래그(0)으로 입력하여 모형이 두 경우를 명확히 구분한다. 예측 단계에서도 동일 규칙을 적용하여, 개별 행에 결측이 있더라도 정상적으로 예측이 된다.

두 번째로 적정 IQR 계수값 선정 및 최적 트리수 산정을 위해 교차검증(cross validation) 기반 하이퍼파리미터를 실시하였다.

적정 IQR 계수값 선정 시 1차적으로 홀드아웃(hold-out) 방식으로 탐색을 한 후, 2차로 5-fold 교차검증 방식으로 훈련 폴드에서만 학습을 한 후 검증 폴드에서 검증하였다. 1차 탐색 시 IQR 계수 값이 0.5일 때 가장 좋은 평가지표를 보였으나, 이상치 제거 비율이 17.0%로 높게 나타나 2차 검증 시에는 제외하였다.

IQR 계수값은 이상치 제거비율 및 평가지표를 고려하여 1.0과 1.25 두 가지를 선정하였다.

적정 tree 값(n_estimators) 산정 역시 5-fold 교차검증으로 각 후보 트리 수에 대한 검증 성능을 비교해 최적값을 선택하였다. 트리 수를 50~500 범위에서 비교한 결과, R^2는 50에서 200까지 증가한 이후 200 이상에서는 평가지표의 개선 폭은 미미한 반면 계산 시간 부담만 증가한다고 판단하였다. 이후 분석에서는 예측 성능과 계산 효율성을 균형 있게 고려한 트리 수로 200을 최적값으로 선택하였다.

이상과 같은 방법으로 최적의 IQR 계수(±1.0, ±1.25) 및 트리 수(200)를 선정하였으며, 최종 모형 학습 시 기존 설명변수와 위 설정 값을 적용하여 예측모형을 구현하였다.

<표 3-9> IQR 계수 교차 검증

□ 2023년 공동주택

1차 홀드아웃(hold-out)								⇒	2차 교차검증(cross validation)		
IQR 계수	데이터수	제거율	정상치	R^2	MAE	RMSE	반복횟수		R^2	MAE	RMSE
0.5	933,722	17.0	775,209	0.9393	2,996	5,169	10		-	-	-
1.0		6.7	871,387	0.9238	3,416	5,890			0.9322	3439.1	5791.6
1.25		4.6	891,108	0.9238	3,490	5,975			0.9259	3586.3	6104.6
1.5		3.2	903,461	0.9171	3,573	6,193			0.9220	3660.2	6310.3
1.75		2.4	911,023	0.9114	3,672	6,502			0.9186	3707.3	6477.2
2.0		1.9	916,420	0.9142	3,642	6,430			0.9173	3747.2	6549.1

* 이상치 제거 : 시도/시군구/읍면동/주택유형별 ㎡ 금액 기준

□ 2023년 단독주택

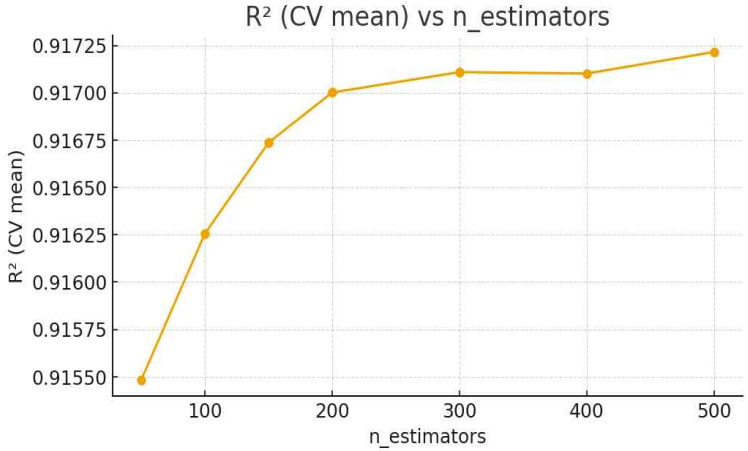
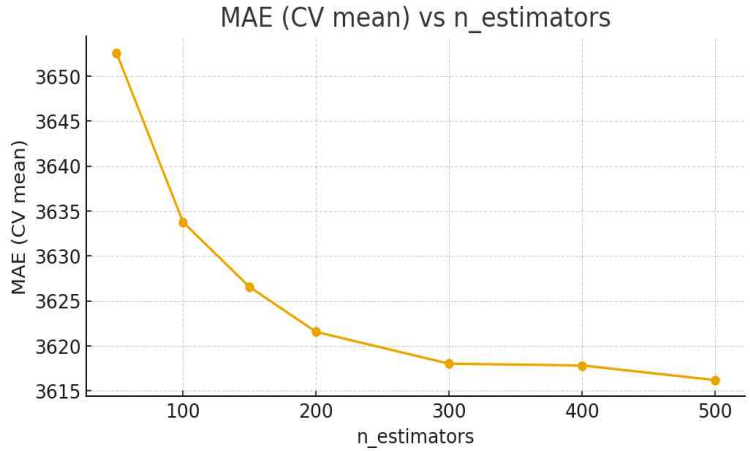
1차 홀드아웃(hold-out)								⇒	2차 교차검증(cross validation)		
IQR 계수	데이터수	제거율	정상치	R^2	MAE	RMSE	반복횟수		R^2	MAE	RMSE
0.5	116,607	18.6	94,889	0.8512	1,983	3,009	10		-	-	-
1.0		6.9	108,512	0.7849	2,350	3,608			0.7869	2408.3	3642.3
1.25		4.6	111,273	0.7708	2,417	3,741			0.7692	2500.7	3825.0
1.5		3.1	112,967	0.7615	2,495	3,903			0.7579	2564.4	3949.3
1.75		2.2	114,083	0.7505	2,520	3,992			0.7500	2603.5	4043.6
2.0		1.5	114,841	0.7451	2,553	4,091			0.7450	2642.2	4103.9

* 이상치 제거 : 시도/시군구/읍면동/주택유형별 ㎡ 금액 기준

<표 3-10> n_estimators(트리수) 교차검증

□ 2023년 공동주택

트리수	샘플수	폴드수	IQR	R^2	MAE	RMSE
50	80,000	5	1.25	0.9154	3,653	6,097
100				0.9162	3,634	6,069
150				0.9167	3,627	6,052
200				0.9170	3,622	6,043
300				0.9171	3,618	6,040
400				0.9171	3,618	6,040
500				0.9172	3,616	6,036

R^2 vs n_esitimators	MAE vs n_esitimators
	

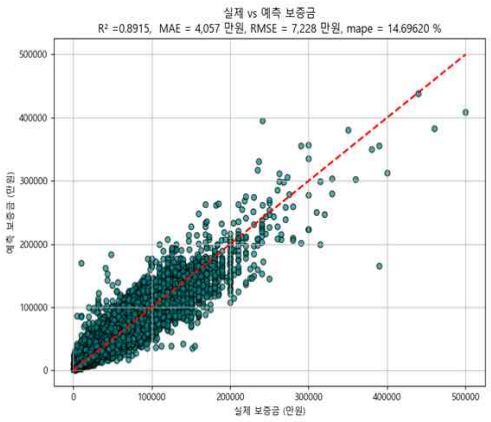
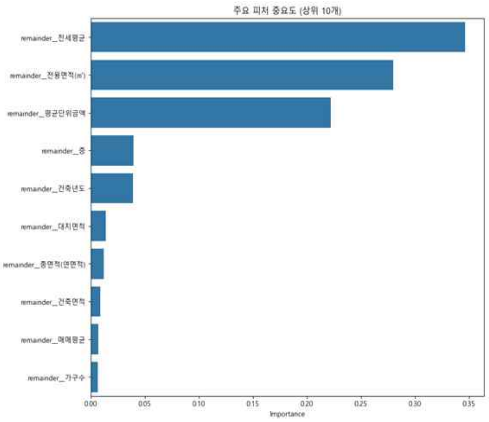
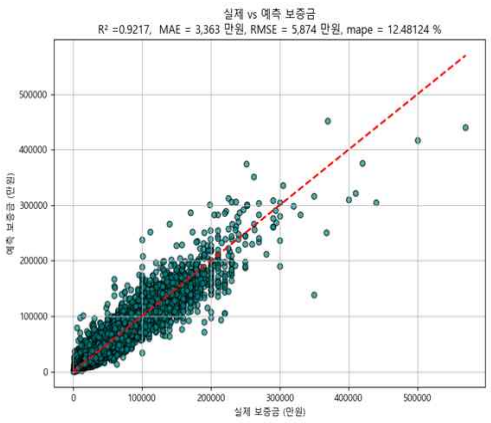
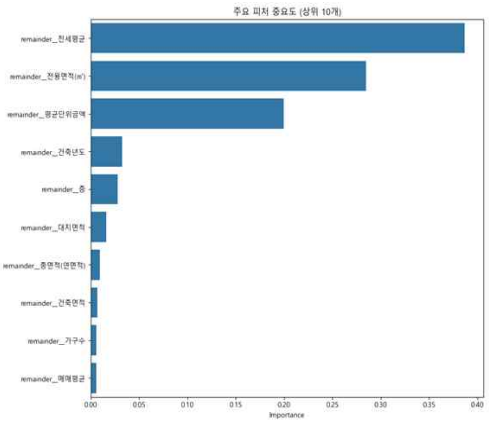
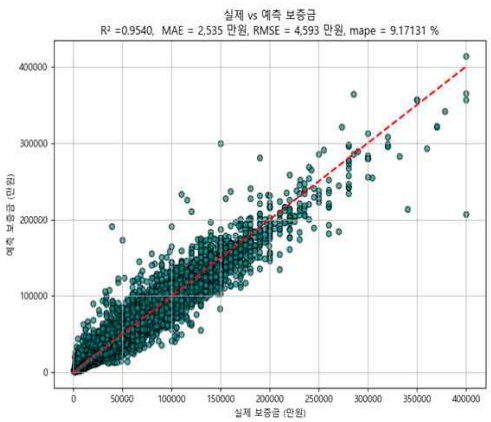
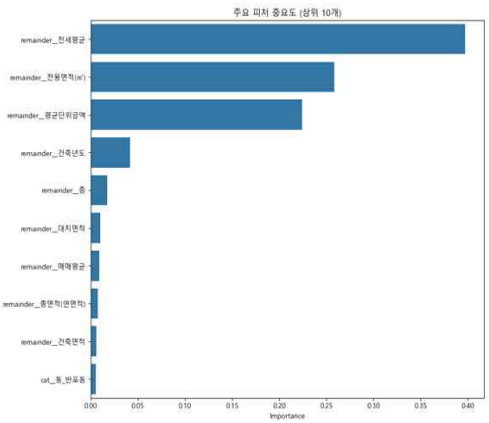
최종적으로 기존 설명변수에 건축물대장과 IQR 계수값, 적정 트리수를 반영한 최종 모형을 학습하였으며, 그 결과 모든 평가지표에서 예측 성능이 개선되는 효과를 확인할 수 있었다.

최종 모형에 대한 설명변수 중요도(Feature Importance)를 살펴보면 (공동주택) 기준으로는 전세평균-전용면적-단위금액-건축연도-층-대지면적-매매평균 순으로 나타났으며, (단독주택) 기준으로는 전용면적-단위금액-건축연도-전세평균-매매평균 순으로 나타났다.

<표 3-11> 주택유형/연도별 전세보증금 최종 예측모형(4, 5단계)

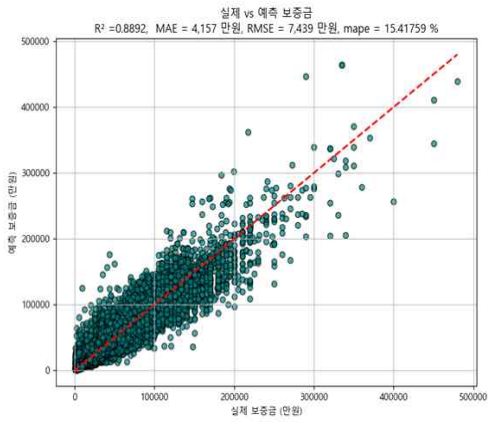
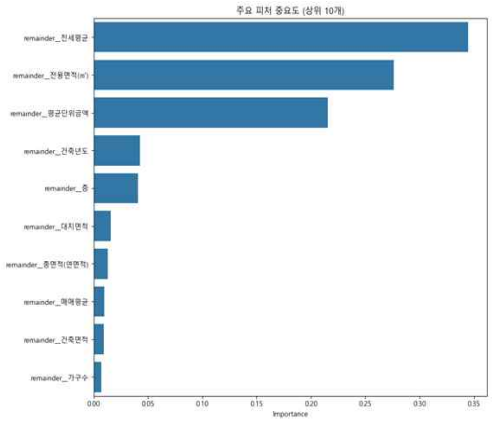
(단위: %, 만 원)

주택	년도	변수설정		R^2	MAPE	MAE	RMSE
공동주택	2021	4	③+건축물대장+(±1.25*IQR)	0.889	15.417	4,157	7,439
		5	③+건축물대장+(±1.0*IQR)	0.892	14.696	4,057	7,228
	2022	4	③+건축물대장+(±1.25*IQR)	0.919	12.781	3,434	6,027
		5	③+건축물대장+(±1.0*IQR)	0.922	12.481	3,363	5,874
	2023	4	③+건축물대장+(±1.25*IQR)	0.953	9.551	2,601	4,697
		5	③+건축물대장+(±1.0*IQR)	0.954	9.171	2,535	4,593
단독주택	2021	4	③+(±1.25*IQR)	0.779	26.476	2,210	3,682
		5	③+(±1.0*IQR)	0.798	25.119	2,153	3,539
	2022	4	③+(±1.25*IQR)	0.783	25.186	2,275	3,760
		5	③+(±1.0*IQR)	0.799	23.846	2,187	3,536
	2023	4	③+(±1.25*IQR)	0.796	23.789	2,283	3,561
		5	③+(±1.0*IQR)	0.803	22.669	2,210	3,482

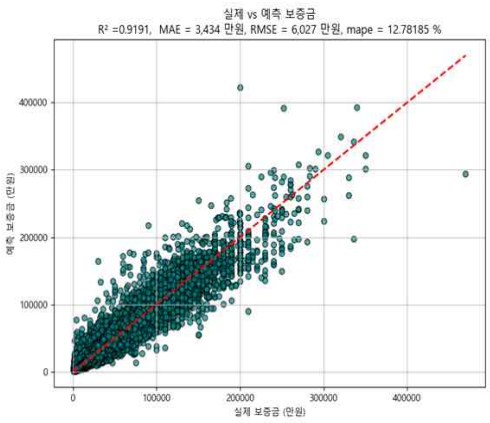
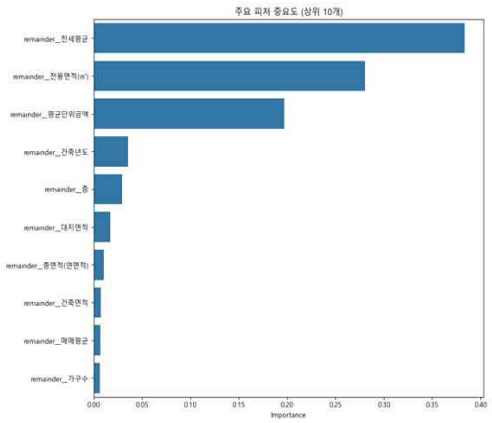
<table>
<tr><th colspan="2">2021년 공동주택 (IQR 1.0 기준)</th></tr>
<tr><th>실제보증금 vs 예측보증금</th><th>주요 피쳐 중요도</th></tr>
<tr><td> 
</td><td> 
</td></tr>
<tr><th colspan="2">2022년 공동주택 (IQR 1.0 기준)</th></tr>
<tr><th>실제보증금 vs 예측보증금</th><th>주요 피쳐 중요도</th></tr>
<tr><td> 
</td><td> 
</td></tr>
<tr><th colspan="2">2023년 공동주택 (IQR 1.0 기준)</th></tr>
<tr><th>실제보증금 vs 예측보증금</th><th>주요 피쳐 중요도</th></tr>
<tr><td> 
</td><td> 
</td></tr>
</table>

<그림 3-5> 연도별/주택유형별 예측보증금 및 주요 피쳐 중요도

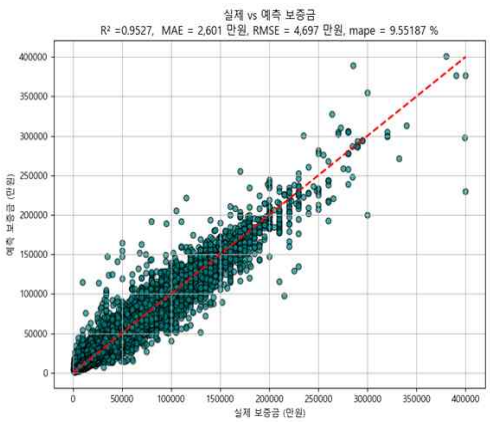
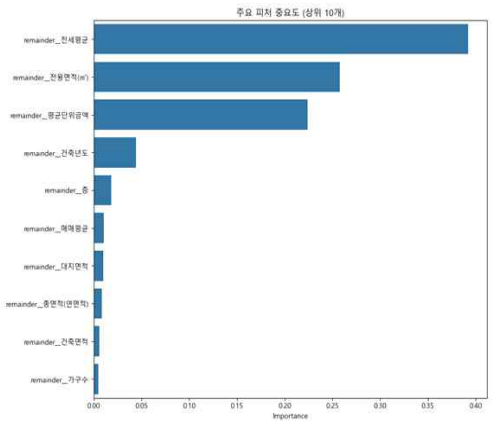
2021년 공동주택 (IQR 1.25 기준)

실제보증금 vs 예측보증금	주요 피쳐 중요도
	

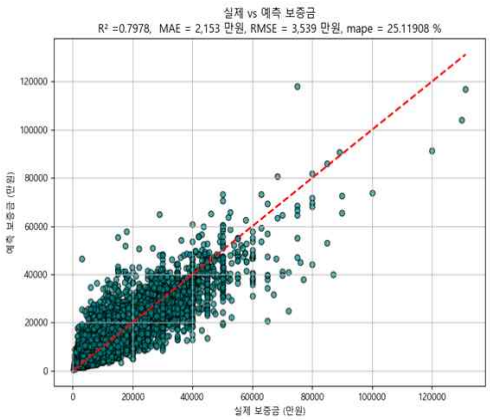
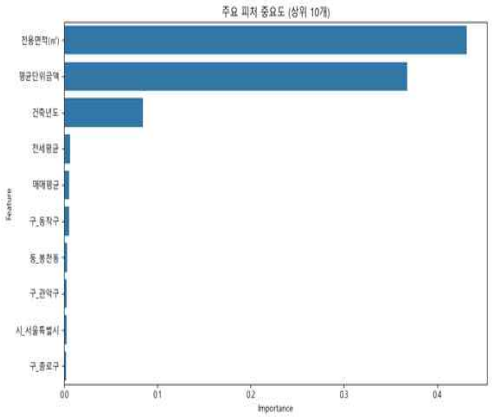
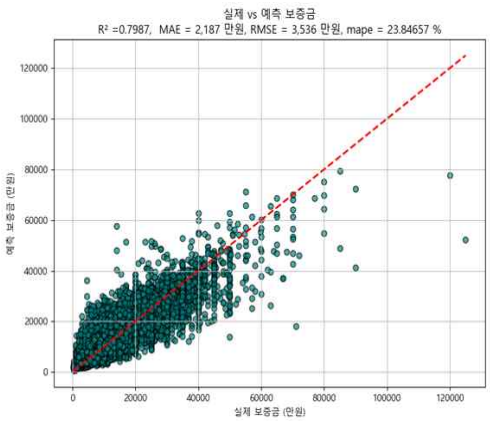
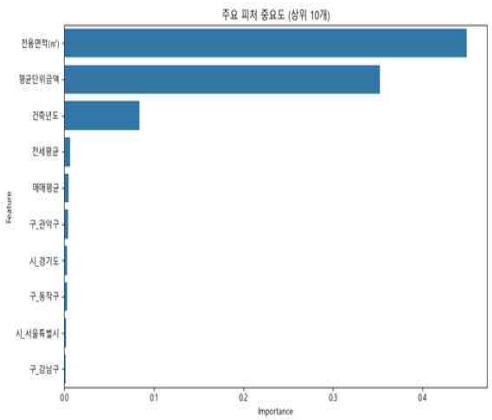
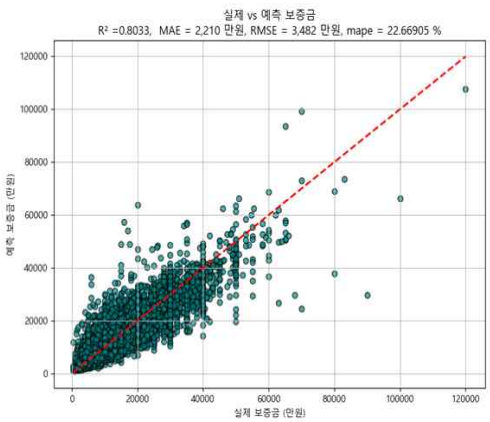
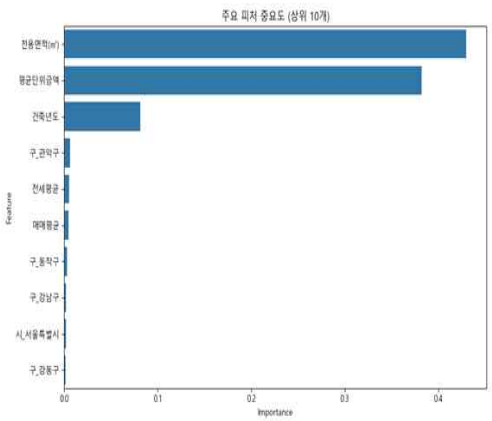
2022년 공동주택 (IQR 1.25 기준)

실제보증금 vs 예측보증금	주요 피쳐 중요도
	

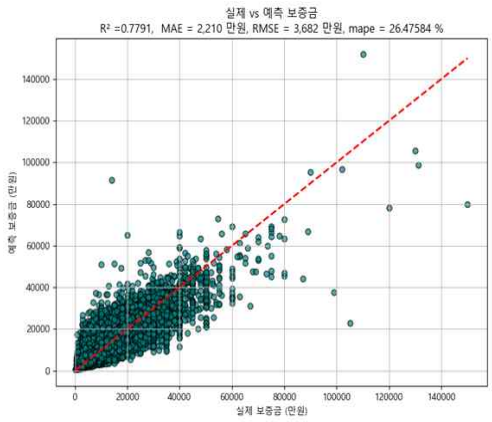
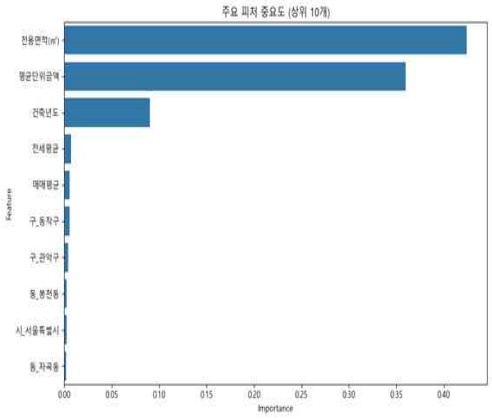
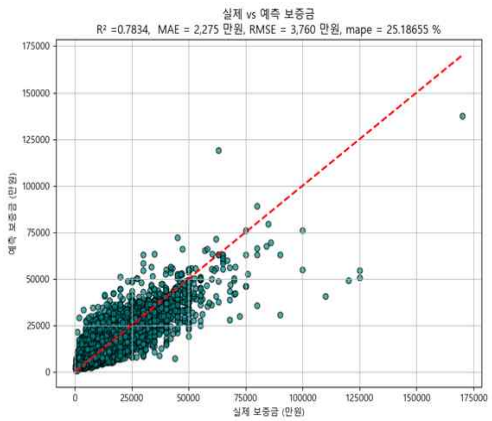
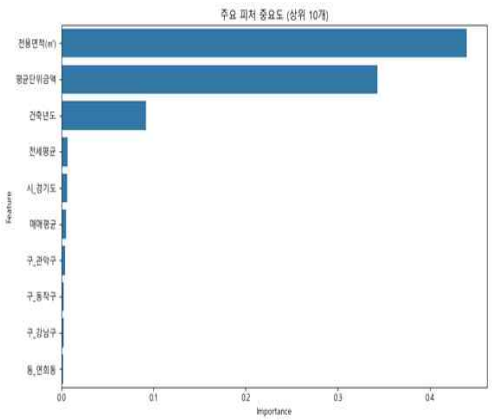
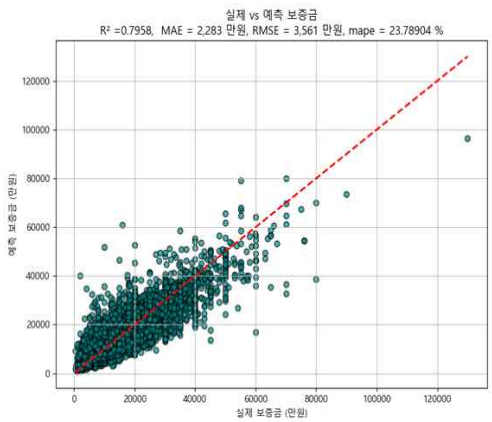
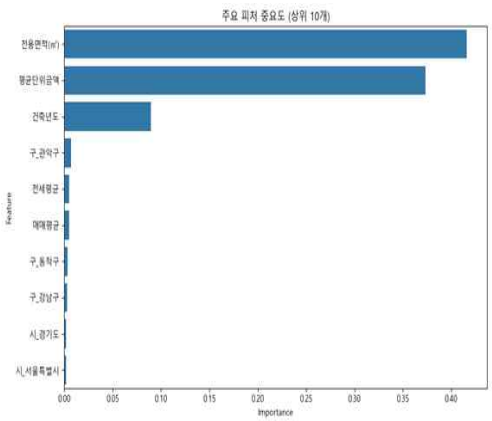
2023년 공동주택 (IQR 1.25 기준)

실제보증금 vs 예측보증금	주요 피쳐 중요도
	

<그림 3-5> 연도별/주택유형별 예측보증금 및 주요 피쳐 중요도 (계속)

2021년 단독주택 (IQR 1.0 기준)	
실제보증금 vs 예측보증금	주요 피쳐 중요도
	
2022년 단독주택 (IQR 1.0 기준)	
실제보증금 vs 예측보증금	주요 피쳐 중요도
	
2023년 단독주택 (IQR 1.0 기준)	
실제보증금 vs 예측보증금	주요 피쳐 중요도
	

<그림 3-5> 연도별/주택유형별 예측보증금 및 주요 피쳐 중요도 (계속)

2021년 단독주택 (IQR 1.25 기준)	
실제보증금 vs 예측보증금	주요 피쳐 중요도
	
2022년 단독주택 (IQR 1.25 기준)	
실제보증금 vs 예측보증금	주요 피쳐 중요도
	
2023년 단독주택 (IQR 1.25 기준)	
실제보증금 vs 예측보증금	주요 피쳐 중요도
	

<그림 3-5> 연도별/주택유형별 예측보증금 및 주요 피쳐 중요도 (계속)

제 4 장

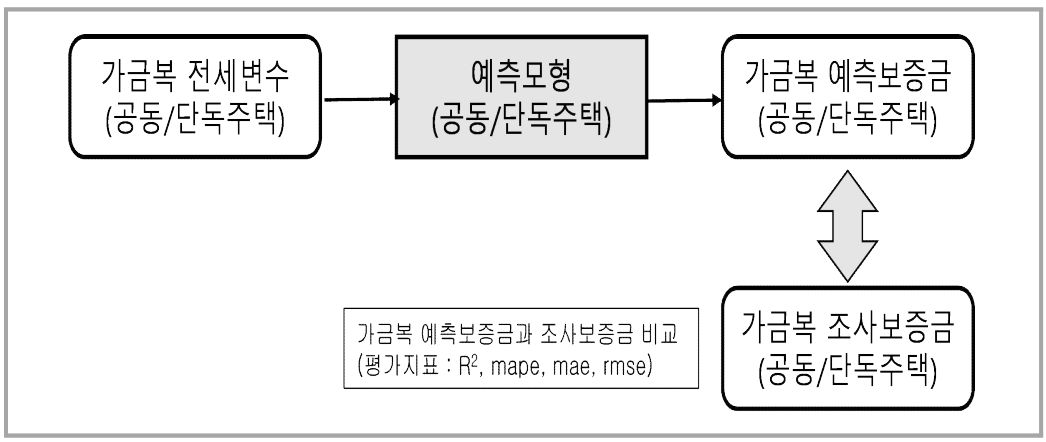
가계금융복지조사 적용 및 의제주거소득 추정

제1절 가계금융복지조사 전세가구를 활용한 모형 검증

연도별 예측모형을 공동주택과 단독주택으로 각각 구축한 후, 먼저 가계금융복지조사 전세가구 변수를 모형에 적용하여 성능을 확인하였다. 즉, 실거래 자료를 바탕으로 구축한 모형에 가금복 전세가구의 변수 특성을 입력하여 예측 전세보증금을 산출하고, 이를 기존에 조사된 전세보증금과 비교하였다.

그러기 위해서 앞서 언급했듯이 가금복 자료에 대지면적, 건축면적, 연면적, 가구수, 건축연도 정보를 국토부 건축물대장 자료를 활용하여 매칭하였다.

층 변수는 가금복 조사자료에 세대별 호수 정보가 존재하므로 호수에서 층 정보를 추출하여 생성하였다. 이와 같은 과정을 통해 가금복 자료의 주택 특성을 실거래 자료와 동일한 변수 구조로 정비하였다.



<그림 4-1> 가금복 전세보증금을 활용한 예측모형 검증 과정

<표 4-1>은 예측모형을 활용한 가금복 전세가구의 예측보증금과 조사보증금을 비교한 평가표이다. (좌측)가금복 원데이타는 모든 전세가구에 대해 예측한 결과값이며, (우측)이상치제거는 실제 가금복 가구에도 공공임대, 사인 간 거래 등으로 보이는 금액(천만 원 미만)이 있으므로 해당 금액을 제거[19]한 후 평가한 지표값이다.

(2023년 공동주택, IQR=1.0) 기준으로 보면 R^2=0.904, mape=30.2%, mae=3,656만 원으로 나타났으며, 예측모형 값은 R^2=0.954, mape=9.2%, mae=2,535만 원으로 나타났다. 이러한 차이는 학습데이터는 당해연도 금액이지만 가금복 전세금액은 1년~2년 전 금액인 점과 공공임대/사인 간 거래 같은 특이 금액을 정확히 예측하지 못하는 점, 학습모델의 트리 깊이(depth)를 지정하지 않아 과접합(overfitting)이 발생할 수 있는 점을 들 수 있다.[20]

(2023년 단독주택, IQR=1.0) 기준으로 보면 R^2=0.748, mape=23.8%, mae=2,416만 원으로 나타났으며, 예측모형 값은 R^2=0.803, mape=22.6%, mae=2,210만 원으로 나타나, 공동주택에 비해서는 예측모형과 차이가 상대적으로 크지 않은 것으로 나타났다.

예측모형에 가금복 전세변수를 적용하여 예측한 전세금액과 실제 조사된 전세금액을 비교한 결과 2023년 기준으로 공동주택은 약 90%, 단독주택은 약 75% 정도 설명력을 확인할 수 있었으며, 이는 기본 예측모형의 테스트 성능지표(<표 3-11>)보다 약간 낮은 값을 보였다.

19) 실거래 이상치 제거와 동일한 방법으로 (시도/시군구/읍면동/주택유형별 ㎡ 금액 기준)으로 사분위수 범위를 활용하여 제거

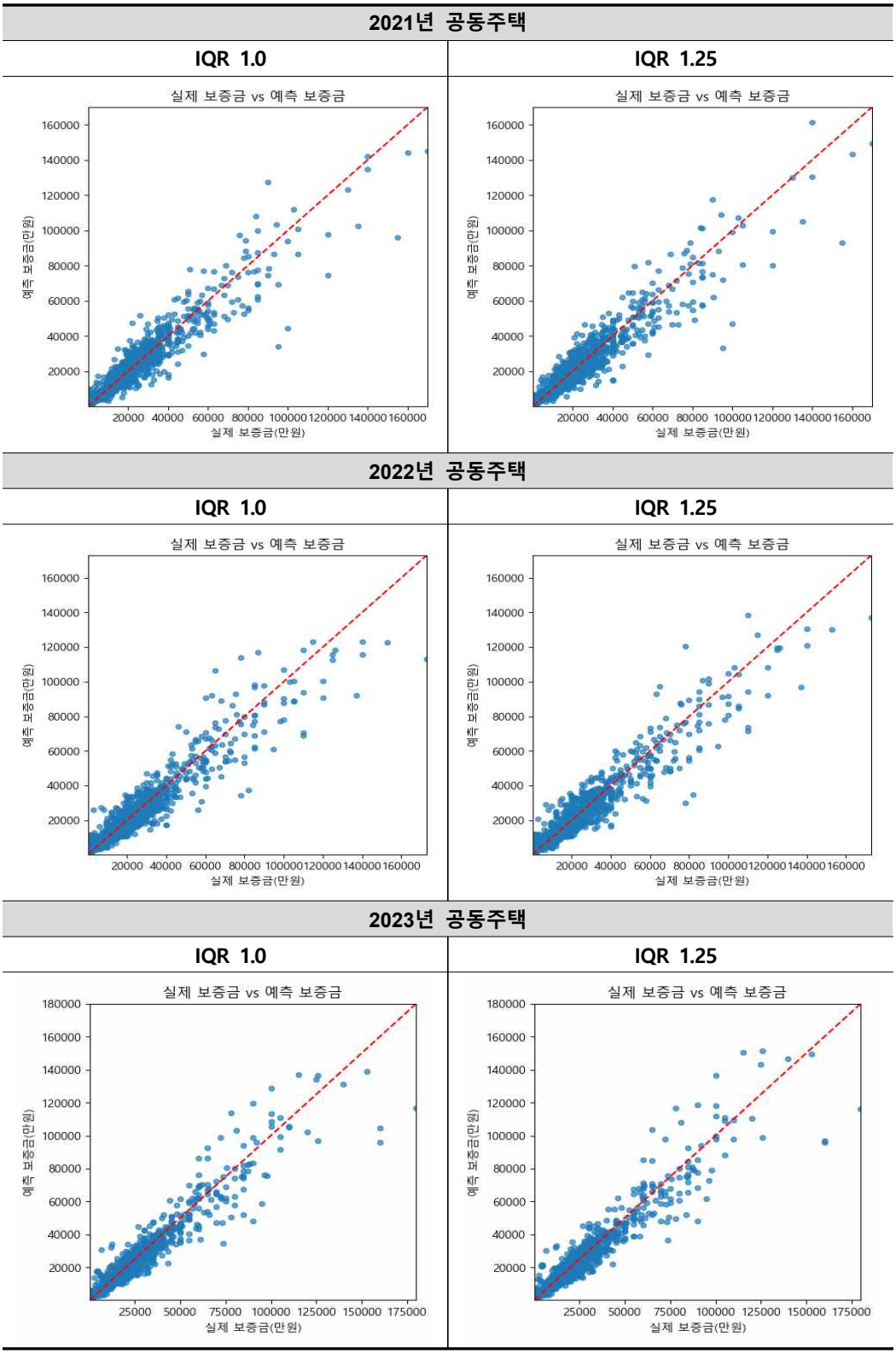
20) mape 값은 전세보증금이 천만 원 미만인 가구에 대해 실제 높게 예측하여 오차가 크게 발생하며, 해당 가구 삭제 후 평가지표 산정 시 mape 값은 크게 줄어듦

<표 4-1> 예측모형을 활용한 가금복 전세가구 (예측보증금 vs 조사보증금) 평가지표

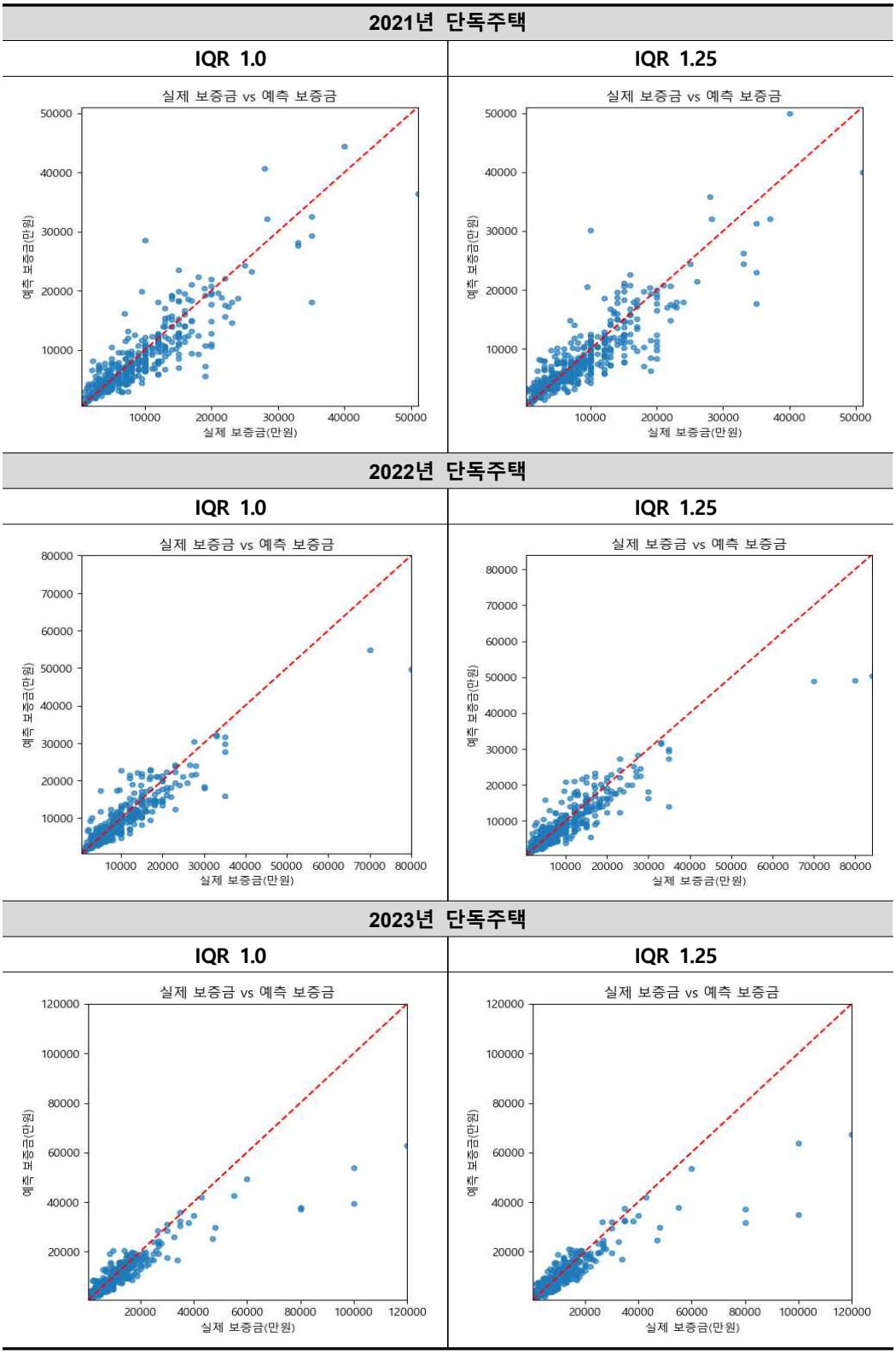
(%, 만원)

IQR=1.0 모델		가금복 원데이타					단위면적 이상치(1.0iqr) 제거(로그) (공공임대가구 제거위한)				
		가구수	R2(결정계수)	mape	mae	rmse	가구수	R2(결정계수)	mape	mae	rmse
2021년	공동	1,579	0.876	52.0	4,132	6,974	1,394	0.892	33.1	3,780	6,435
2022년	공동	1,576	0.869	94.8	4,369	7,526	1,403	0.893	34.0	4,020	6,837
2023년	공동	1,461	0.896	650.4	3,921	6,918	1,265	0.904	30.2	3,656	6,591
2021년	단독	562	0.763	710.3	1,957	3,180	501	0.787	25.5	1,800	2,935
2022년	단독	539	0.807	220.0	2,133	3,713	475	0.830	22.7	1,857	3,205
2023년	단독	547	0.750	290.0	2,587	5,886	480	0.748	23.8	2,416	5,991
2021년	전체	2,141	0.884	224.8	3,562	6,207	1,895	0.900	31.1	3,257	5,722
2022년	전체	2,115	0.877	126.7	3,799	6,762	1,878	0.900	31.2	3,473	6,126
2023년	전체	2,008	0.891	552.2	3,558	6,653	1,745	0.897	28.4	3,315	6,431
		6,264					5,518				

IQR=1.25 모델		가금복 원데이타					단위면적 이상치(1.25iqr) 제거(로그) (공공임대가구 제거위한)				
		가구수	R2(결정계수)	mape	mae	rmse	가구수	R2(결정계수)	mape	mae	rmse
2021년	공동	1,579	0.878	50.8	4,137	6,910	1,444	0.890	35.9	3,847	6,470
2022년	공동	1,576	0.876	92.6	4,349	7,323	1,442	0.899	36.1	4,043	6,620
2023년	공동	1,461	0.889	657.9	4,054	7,149	1,307	0.897	34.1	3,801	6,842
2021년	단독	562	0.747	715.2	2,021	3,290	524	0.769	30.2	1,925	3,098
2022년	단독	539	0.795	221.0	2,181	3,829	496	0.809	25.1	2,052	3,685
2023년	단독	547	0.751	347.2	2,574	5,867	496	0.749	25.1	2,456	5,904
2021년	전체	2,141	0.885	225.2	3,582	6,169	1,968	0.897	34.4	3,335	5,768
2022년	전체	2,115	0.883	125.3	3,797	6,611	1,938	0.904	33.3	3,534	6,007
2023년	전체	2,008	0.886	573.3	3,651	6,824	1,803	0.893	31.6	3,431	6,597
		6,264					5,709				



<그림 4-2> 가금복 전세가구 실제보증금 vs 예측보증금



<그림 4-2> 가금복 전세가구 실제보증금 vs 예측보증금 (계속)

제2절 전월세 전환율 및 이자율[21)]

예측모형을 활용한 전세보증금 예측 후, 의제주거소득을 추정하기 위해서는 전환율 또는 이자율을 적용하여 월세화해야 한다. 즉, 자가의 주거 서비스를 월세화하여 의제주거비 추정 시, 보증금 전환 방법에 따라 추정 결과의 변동성이 큰 만큼 이에 대한 적용 기준이 필요하다.

<표 4-2> 전월세 전환율 및 주요 이자율

이율(%)	구분		2020년	2021년	2022년	2023년
전월세 전환율	종합주택		5.81	5.68	5.78	6.05
이자율		수신금리(순수저축성, 잔액 기준)	1.14	1.30	3.61	3.87
		대출금리(주담대, 잔액 기준)	2.67	2.77	3.83	4.32
		평균	1.91	2.04	3.72	4.10
	대출금리(가계대출, 신규 기준)		2.75	3.10	4.60	4.96

* 출저 : 한국부동산원, 한국은행 경제통계시스템

의제주거소득, 자가주거비 추정 관련 통계는 다양하게 존재하며, <표 4-3>과 같이 정리할 수 있다.

전월세 전환율과 이자율은 자가주거서비스를 금액적으로 평가하는 서로 다른 접근 방식이므로, 어느 지표에 어떤 방법을 적용할 것인지에 대해 개념적 타당성, 시계열 연속성, 지표의 안정성이라는 세 가지 관점에서 검토할 필요가 있다.

첫째, 개념적 측면에서 보면, 전월세 전환율은 전세보증금이 시장에서 어느 정도의 월세 가치와 대응되는지를 나타내는 지표로서, 전세보증금을 월세 상당액으로 환산하는 데 직접적으로 사용된다. 이 방식은 “동일 주택을 임차할 경우 지불했을 임대료 상당액”이라는 국민계정(SNA)의 자가주거서비스 평가 원칙과 개념적으로 일치하므로, 소득(분배)지표에서 의제주거소득을 산정하는 데 적합하다.

21) 행정자료 활용 소비자물가의 주거비지수 작성 및 실무 적용방안 연구 인용(변준석 외, 2024)

<표 4-3> 주요 통계별 전월세 전환율 및 이자율 활용

구분	미시통계	거시통계
개념	동 주택을 타인에게 임차할 경우 임차인이 지불해야 할 임대료로 추정	
자가주거비 산출 방법	임대료 상당법 (인근의 유사한 주택의 전세 및 월세를 활용하여 임대료 계산	임대료 상당법 (지역별 임대가구의 유형별 임대료 평가액을 임대가구의 임대유형 비중으로 가중평균하여 가구당 귀속임대료를 계산 후 자가 가구수를 곱하여 계산)
전세의 월세전환 비율	전월세 전환율	가계대출 평균금리
주요 통계	가계수지동향과(자가주거비) 물가동향과(자가주거비)	한국은행 GDP(귀속임대료, 소득 및 지출) 소득통계과 GRDP(귀속임대료, 소득 및 지출)

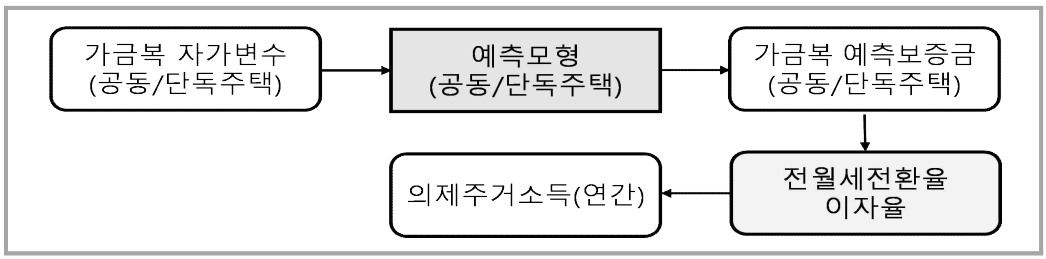
반면, 이자율을 적용하는 방식은 전세보증금을 주택에 묶어둠으로써 포기한 금융수익, 즉 자본의 사용자비용을 측정하는 것으로, 가계가 주거를 위해 부담하는 기회비용(이자비용)을 파악하는 데 더 적합하다. 따라서 개념상으로는 소득을 추정하는 경우(의제주거소득)에는 전월세 전환율을, 주거비 지출(부담)을 추정하는 경우(자가주거비)에는 이자율을 적용하는 것이 타당하다.

둘째, 시계열 연속성 관점에서 보면, 현재 전환율 기반 임대료 상당법(imputed rent)을 활용하는 미시통계(소비자물가, 가계동향지출)와 이자율 기반 사용자비용(user cost) 방법을 활용해 온 거시통계(GDP, GRDP 등)에 대해서는 방법 변경 시 시계열 단절과 해석상의 혼란을 초래할 우려가 크다. 반면, 가계금융복지조사의 의제주거소득은 아직 연구과정에 있으므로, 전월세 전환율을 활용한 새로운 산정 방식을 도입하더라도 시계열상의 부담이 상대적으로 적을 것으로 판단되며, 미시통계에서는 전환율 중심, 거시통계에서는 이자율 중심으로 사용되는 것이 현실적인 선택으로 보인다.

셋째, 지표의 안정성 측면에서는 전환율과 이자율 간 선택이 단순한 기술적 문제가 아니기 때문에 단일 방법으로의 일괄 전환에 대해 사회적 합의를 도출하기가 쉽지 않을 것이다. 이 경우 전환율과 이자율 중 하나만을 공식적으로 활용하기보다는 두 값의 평균값을 활용하는 방안도 고려해 볼 수 있다.

제3절 가계금융복지조사 자가·무상가구 전세보증금 예측 및 의제 주거소득 추정

예측모형을 통해 가금복 전세가구를 대상으로 검증을 실시한 이후 자가와 무상가구에 대한 전세보증금 예측을 실시하였다. 전세 예측 방식과 동일하게 자가/무상 가구의 주택 특성(위치, 면적, 층, 건축연도, 주택유형, 전세/매매 평균, 실거래 평균 단위금액, 건축물대장 변수 등)을 모형에 입력하여, 해당 주택을 전세로 임차할 경우 형성될 것으로 예상되는 전세보증금을 예측하였다. 예측된 전세보증금에 전월세 전환율과 이자율을 적용하여 연간 의제주거소득을 추정하였다.



<그림 4-3> 예측모형을 활용한 의제주거소득 추정

<표 4-4> 의제주거소득 추정 금액

<전체가구> (단위:만 원)

소득 년도	월세화(전월세 전환율 적용)		기회비용화(금리 적용)	
	머신러닝 (1.0)	머신러닝 (1.25)	머신러닝 (1.0)	머신러닝 (1.25)
2021	667	667	294	294
2022	762	758	531	529
2023	707	704	549	547

<자가/무상가구> (단위:만 원)

소득 년도	월세화(전월세 전환율 적용)		기회비용화(금리 적용)	
	머신러닝 (1.0)	머신러닝 (1.25)	머신러닝 (1.0)	머신러닝 (1.25)
2021	1,031	1,032	454	455
2022	1,189	1,184	828	825
2023	1,107	1,101	860	856

** 이자율은 한국은행 주택담보대출(잔액)과 순수저축성예금(잔액) 평균값 사용 <표 4-2> 참조

자가 및 무상가구에 대해 머신러닝 기반 전세보증금 예측모형으로 산출한 의제주거소득을 <전체가구> 기준으로 보면 전월세 전환율을 적용해 월세로 환산한 방식에서는 의제주거소득이 연간 약 700~760만원 내외 수준으로 나타난다.

금리를 적용한 기회비용화 방식의 의제주거소득은 300~550만원 수준으로 전월세 전환율 방식보다 낮은 것으로 나타났다.

전월세 전환율을 사용하는 방식의 추정치가 금리를 사용하는 방식보다 높은 이유는 일반적으로 주택 전월세 전환율이 주택담보대출 이자율이나 예금·저축이자율 등 금융 이자율보다 높은 수준에서 형성되기 때문이다.

제4절 가계금융복지조사 자가·무상가구 예측보증금 간접 검증

자가/무상 가구의 예측된 전세보증금은 실제 관측값이 존재하지 않아 개별 수준에서 예측 정확도를 직접 검증[22]하기는 어렵다. 이에 가금복조사 자료를 활용하여 집단 수준에서의 간접 검증을 수행하였다.

구체적으로 시도 단위를 기준으로, ① 가금복조사 전세가구의 실제 전세보증금 평균, ② 동일 집단에 대해 예측모형을 적용하여 산출한 전세가구의 예측 전세보증금 평균, ③ 같은 집단에 속한 자가가구의 예측 전세보증금 평균을 산출하여 비교하였다.

이와 같은 절차를 통해, 먼저 전세가구에 대해서는 실제 평균과 예측 평균이 어느 정도 일치하는지를 확인하였고, 이어서 같은 모형을 적용한 자가/무상가구의 예측 전세보증금 평균이 시도별 전세가구 수준과 비교하여 과도하게 높거나 낮지 않은지 검토하였다.

이러한 집단 단위 비교를 통해 자가/무상가구에 대한 전세보증금 예측값의 타당성을 간접적으로 확인하였으며, 개별 관측값이 존재하지 않는 상황에서 예측모형의 활용 가능성을 평가하는 보조적인 검증 도구로 활용하였다.

<표 4-5>를 보면 전세 예측의 경우, 전세 예측/전세 실제 비율이 전 시도에서 대체로 1 근처(약 0.94~1.18 범위)에 분포하고, 자가 예측의 경우, 자가 예측/전세 실제 비율이 대부분 1을 상회하는 것으로 나타났다.

자가가구의 예측 전세보증금이 전세가구 실제 전세보증금 평균을 상회하는 것은 자가가구가 동일 지역 내에서 전세가구에 비해 규모, 입지, 노후도 등 주택 특성이 상대적으로 우월한 주택에 거주하는 경향이 반영된 결과이며, 예측모델에 이러한 점이 반영되었다고 판단할 수 있다.

---

22) 자가가구의 전세보증금 예측 검증 방법은 ①전세가구 예측보증금과 자가가구 예측보증금 비교, ②매매가격 대비 전세금액 비율(전세가율) 등의 방법으로 간접 비교 가능할 것으로 판단되며 본 연구에서는 ①번 방법만 수행함

<표 4-5> 자가/무상 가구의 예측보증금 간접 검증

년도	시도	전세(1.0제거)				자가		
		n	실제평균 보증금(A)	예측평균 보증금(B)	예측/실제 (B/A)	n	예측평균 보증금(C)	자가_예측/ 전세_실제 (C/A)
2021	서울	512	24,916	23,418	0.94	1,217	43,769	1.76
	부산	128	13,807	13,785	1.00	835	18,159	1.32
	대구	96	12,097	12,040	1.00	745	20,938	1.73
	인천	103	13,882	14,224	1.02	538	20,926	1.51
	광주	66	11,850	12,261	1.03	517	18,928	1.60
	대전	113	14,688	14,948	1.02	619	20,701	1.41
	울산	53	13,096	13,317	1.02	454	19,084	1.46
	세종	32	19,528	21,906	1.12	249	20,107	1.03
	경기도	363	21,831	21,727	1.00	1,519	30,255	1.39
	강원	48	10,096	9,888	0.98	723	10,817	1.07
	충북	65	9,338	9,878	1.06	592	12,424	1.33
	충남	59	12,250	12,920	1.05	730	11,178	0.91
	전북	67	9,376	10,015	1.07	785	11,805	1.26
	전남	41	9,778	10,635	1.09	866	9,086	0.93
	경북	62	11,043	11,686	1.06	727	10,645	0.96
	경남	57	11,632	12,130	1.04	792	11,781	1.01
	제주	30	11,848	12,739	1.08	348	19,130	1.61
	전국	**1,895**	**17,598**	**17,389**	**0.99**	**12,256**	**19,722**	**1.12**
2022	서울	520	27,378	26,449	0.97	1,240	44,911	1.64
	부산	115	16,066	15,617	0.97	868	18,266	1.14
	대구	100	11,800	11,428	0.97	747	19,336	1.64
	인천	98	14,082	13,688	0.97	539	21,321	1.51
	광주	72	13,661	13,737	1.01	505	18,438	1.35
	대전	127	14,991	14,784	0.99	607	21,232	1.42
	울산	38	13,042	13,232	1.01	444	18,676	1.43
	세종	36	19,025	21,070	1.11	254	19,500	1.02
	경기도	359	23,268	22,716	0.98	1,554	29,856	1.28
	강원	48	11,580	11,601	1.00	716	13,229	1.14
	충북	67	10,752	10,521	0.98	575	13,801	1.28
	충남	66	14,191	14,520	1.02	761	13,005	0.92
	전북	63	10,798	11,181	1.04	789	12,215	1.13
	전남	35	11,136	13,170	1.18	849	9,819	0.88
	경북	52	9,519	9,670	1.02	760	11,564	1.21
	경남	57	12,659	13,164	1.04	789	12,469	0.99
	제주	25	14,014	14,575	1.04	357	18,834	1.34
	전국	**1,878**	**19,124**	**18,807**	**0.98**	**12,354**	**20,232**	**1.06**
2023	서울	524	29,223	27,611	0.94	1,329	42,543	1.46
	부산	104	15,139	14,453	0.95	879	17,253	1.14
	대구	102	11,260	10,814	0.96	783	16,146	1.43
	인천	89	14,763	13,682	0.93	566	17,922	1.21
	광주	65	11,678	11,515	0.99	483	18,170	1.56
	대전	116	15,288	15,045	0.98	585	20,588	1.35
	울산	36	14,542	13,729	0.94	458	16,212	1.11
	세종	37	18,756	20,012	1.07	229	18,759	1.00
	경기도	303	24,255	23,493	0.97	1,561	28,486	1.17
	강원	40	9,781	9,488	0.97	723	12,272	1.25
	충북	45	11,618	11,768	1.01	590	14,127	1.22
	충남	61	11,535	11,818	1.02	756	11,467	0.99
	전북	56	9,714	9,744	1.00	815	12,550	1.29
	전남	53	13,703	14,313	1.04	815	9,326	0.68
	경북	37	12,351	12,770	1.03	783	10,797	0.87
	경남	54	11,397	11,860	1.04	799	12,492	1.10
	제주	23	13,713	13,923	1.02	388	17,831	1.30
	전국	**1,745**	**19,899**	**19,201**	**0.96**	**12,542**	**19,180**	**0.96**

제 5 장

결론 및 시사점

제1절 연구 요약 및 시사점

본 연구는 국토교통부 임대차 실거래(전세) 자료, 한국부동산원 주택가격동향조사, 국토부 건축물대장 등 행정자료를 결합하고, 랜덤포레스트를 중심으로 한 머신러닝 기법을 활용하여 자가주택의 전세 예측모형을 구축하고, 가계금융복지조사에 적용함으로써 의제주거소득을 추정할 수 있는 방법을 제시하였다. 기존 연구가 제한된 지역이나 특정 자료에 의존하였던 것과 달리, 전국 단위 전세 실거래 자료와 주택, 단지, 지역시장 정보를 동시에 활용하였다는 점에서 보다 포괄적인 의제주거소득 추정 기반을 마련하였다고 볼 수 있다.

전세보증금 예측모형 측면에서는, 전세 실거래 자료에 대해 주택 특성(면적, 층, 건축연도, 주택유형), 단지 특성(대지면적, 건축면적, 연면적, 가구수), 지역시장 변수(전세/매매 평균, 실거래 평균 단위금액)를 결합하여 공동주택과 단독주택에 대한 별도의 예측모형을 구축하였다. 이상치 제거 수준은 IQR 계수와 홀드아웃 및 5-fold 교차검증을 통해 탐색하였고, 트리 수(n_estimators)는 50~500 범위에서 검증한 결과 예측 성능과 계산 효율성을 함께 고려하여 200을 최적값으로 선정하였다. 그 결과 공동주택의 경우 연도별 R^2가 약 0.89~0.95, 단독주택의 경우 약 0.78~0.80 수준으로 나타나, 행정자료에 기반한 전세보증금 예측모형이 통계작성에 활용할 수 있을 정도의 설명력을 확보하였음을 확인하였다.

구축된 예측모형을 가계금융복지조사 전세가구에 적용하여 조사 전세보증금과 예측 전세보증금을 비교한 결과, 공동주택은 약 90% 내외, 단독주택은 약 75% 내외의 설명력을 보였으며, 이는 동일 자료에서 산출한 테스트 성능보다는 다소 낮지만, 조사 시점 차이와 공공임대, 사인 간 거래 등 특이 거래가 포함된 점을 감안할 때 활용 가능한 수준의 예측력으로 평가되었다.

또한, 가금복 내 공공임대주택에 적용하여 동일 주택의 시장가치 기준 전세보증금을 추정하는 데에도 활용될 수 있다. 이를 실제 공공임대에서 부담하는 보증금·임대료

와 비교하면, 양 금액 간 차이를 통해 공공임대주택이 제공하는 주거서비스의 사회적 현물이전 가치를 계량적으로 산정할 수 있을 것으로 생각된다.

자가·무상 가구의 경우 실제 전세보증금이 관측되지 않아 개별 수준의 직접 검증은 불가능하므로, 시도별 단위에서 간접 검증을 수행하였다. 가계금융복지조사 전세가구에 대해 ① 실제 전세보증금 평균, ② 예측 전세보증금 평균을 비교하여 모형이 전세시장 수준을 어느 정도 재현하고 있는지를 확인하였고, 이어서 같은 집단에 속한 자가·무상 가구의 ③ 예측 전세보증금 평균을 비교하였다. 그 결과 전세가구의 (실제/예측) 비율은 대부분 1 근처(약 0.94~1.18 범위)에 분포하였으며, 자가가구의 예측 전세보증금은 전세가구 실제 전세보증금 평균을 전반적으로 상회하는 것으로 나타났다. 이는 자가가구가 동일 지역 내에서 전세가구에 비해 상대적으로 양호한 주택(규모, 입지, 노후도 등)에 거주하는 경향을 반영한 결과로 해석할 수 있으며 이는 예측모형의 적용 가능성을 간접적으로 뒷받침한다.

의제주거소득 산정 측면에서는, 예측 전세보증금을 전월세 전환율과 이자율을 활용하여 월세화하는 두 가지 방식을 검토하였다. 전월세 전환율을 적용한 경우에는 전세보증금이 시장에서 어느 정도의 월세 가치와 대응되는지를 반영함으로써, 국민계정(SNA)에서 자가주거서비스를 평가하는 임대료 상당액(rental equivalence)과 개념적으로 일치하는 의제주거소득을 산출할 수 있다.

반면 이자율을 적용한 경우에는 전세보증금을 주택에 묶어둠으로써 포기한 금융수익, 즉 자본의 사용자비용(user cost)을 측정함으로써, 가계가 주거를 위해 부담하는 이자비용 또는 기회비용을 파악할 수 있다. 본 연구는 이러한 차이를 고려하여 소득지표 산출에는 전월세 전환율을, 주거비 지출 측정에는 이자율을 적용하는 이원적 활용 기준을 제안하였다.

본 연구의 시사점은 행정자료와 머신러닝 기법을 활용하여 자가주택의 주거서비스 가치를 추정하는 예측모형을 구축하고, 이를 가계금융복지조사에 시범적으로 적용하여 모형의 성능과 활용 가능성을 점검했으며, 주택·단지·지역시장 특성을 보다 세밀하게 반영하는 방법을 제시하고, 공공임대주택의 사회적 현물이전 가치의 추정 가능성을 제시했다는 점을 들 수 있다.

제2절 연구 한계 및 향후 과제

첫째, 본 연구의 예측모형은 실거래 행정자료와 가계금융복지조사 자료로 관측 가능한 변수(소재지, 면적, 층, 건축연도, 주택유형, 건축물대장, 지역별 전세/매매 평균 등)에 기반하고 있어 주변 환경(교통, 학군, 생활편의시설 등)과 같은 공간정보는 반영하지 못하였다. 향후에는 지리정보시스템(GIS)을 활용한 입지 변수, 환경, 교통 자료 등의 변수를 추가하여 모형의 설명력을 높이는 고도화가 요구된다.

둘째, 랜덤포레스트 모형의 주요 하이퍼파라미터에 대해 IQR 계수와 트리 수(n_estimators)에 대해서는 교차검증 기반 탐색을 수행하였으나, 트리 깊이(max_depth) 등 기타 파라미터는 실제 업무 적용 가능성과 계산 부담을 고려하여 기본값을 유지하였다. 이로 인해 일부 과적합 가능성이 완전히 해소되지 않았으며, 다른 앙상블 모형(XGBoost, LightGBM, CatBoost 등)과의 체계적인 비교도 제한적으로만 수행되었다. 향후 연구에서는 다양한 머신러닝 모형과 보다 넓은 하이퍼파라미터 탐색을 통해, 예측 성능과 효율성을 더욱 높일 필요가 있다.

셋째, 본 연구는 전세 실거래 자료를 중심으로 예측모형을 구축하였으나, 향후 전세의 월세화가 진전될 경우 월세 및 보증부월세 자료를 전월세 전환율을 활용하여 전세보증금으로 환산한 후, 전세 실거래 자료와 통합하여 학습자료로 활용할 필요가 있다. 이를 통해 전세와 월세 혼합 구조에 따른 임차시장 변화가 전세보증금과 의제주거소득에 미치는 영향을 충실히 반영할 수 있을 것으로 판단된다.

넷째, 전세보증금 예측모델에 가금복 변수를 적용하여 예측하기 위해서는 가금복 조사항목에 층수와 건축연도를 추가하는 것을 제안한다. 기존 조사 항목에는 층수와 건축연도 정보가 없기 때문에 현장 조사된 호수와 행정자료를 활용해서 건축연도를 매칭하였으나, 실제 매칭이 안 된 사례도 있기 때문에 현장조사 시 해당 항목을 추가 조사하는 방법을 고려할 필요가 있다.

향후에는 이와 같은 변수 보완과 예측모형 고도화를 통해 의제주거소득 추정 및 관련 통계 작성에 활용될 수 있기를 기대한다.

참고문헌

국가데이터처. (2024). **가계금융복조조사 조사지침서**.

김선자, 임찬수, 김경미. (2020). **순의제주거소득 추정방법 연구**. 통계개발원(현 국가데이터연구원).

김정민. (2024). **국가통계 분야 인공지능 활용 체계 연구**. 통계개발원.

김태영, 류두진, 박은일. (2023). "기계학습과 설명가능한 인공지능 모형을 이용한 강남의 전세가격 결정요인 분석". **주택연구.** 31권 3호.

변준석, 송정현. (2024). **행정자료 활용 주거비지수 작성 및 실무적용 방안**. 통계개발원.

이선구. (2025). "XGBoost 기반 부동산 자동가치산정모형(AVM)의 실증분석". 동서울대학교.

이창로, 박기호. (2016). "단독주택가격 추정을 위한 기계학습 모형의 응용". **대한지리학회지**, 15권 2호.

이현기. (2024). **머신러닝을 활용한 김해시 주택가격변화 예측을 위한 모델링.** 경상남도빅데이타센터.

정규승, 김강영, 김서영, 김선자. (2017). **가계금융복지조사에서 의제주거소득 반영을 위한 추정방법 연구**. 통계개발원.

한국부동산분석학회. (2023). **자가주거비지수 산정방식 개선 및 주지표 전환 연구.**

홍정의. (2020). "기계학습 알고리즘을 이용한 주택가격감정 시스템의 구축 및 평가: XGBoost, LightGBM, CatBoost 알고리즘에 기반하여". **주택금융연구**, 4권.

Australian Bureau of Statistics (ABS). Owner-occupied Housing and Imputed Rent. Australian Bureau of Statistics, Canberra.

Bentéjac, C., Csörgő, A., & Martínez-Muñoz, G. (2021). A Comparative Analysis of Gradient Boosting Algorithms. *Artificial Intelligence Review*, 54(3), 1937–1967.

Breiman, L. (2001). Random Forests. *Machine Learning*, 45(1), 5–32.

Eurostat. Methodological Documents on Imputed Rent and the Treatment of Owner-Occupied Housing in EU-SILC and National Accounts. Eurostat, Luxembourg.

International Labour Organization (ILO). (2003). Resolution concerning household income and expenditure statistics. 17th International Conference of Labour Statisticians (ICLS).

Probst, P., Wright, M. N., & Boulesteix, A.-L. (2019). Hyperparameters and Tuning Strategies for Random Forest. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 9(3), e1301.

부 록

<부표 1> 예측모델 평가지표 (IQR=±1.0)

□ 2021년

(단위: %, 만 원)

시도	공동주택				단독주택			
	R^2	MAPE	MAE	RMSE	R^2	MAPE	MAE	RMSE
전국	0.892	14.696	4,057	7,228	0.798	25.119	2,153	3,539
서울	0.879	15.281	5,900	10,327	0.672	30.548	3,037	4,633
부산	0.903	12.498	2,405	4,111	0.496	43.977	1,994	3,053
대구	0.919	11.679	2,912	4,311	0.648	27.265	1,831	2,772
인천	0.856	14.887	3,078	5,154	0.799	22.644	1,692	2,820
광주	0.905	9.998	1,901	3,369	0.636	28.439	1,414	2,039
대전	0.832	17.671	3,849	5,558	0.781	20.442	1,531	2,298
울산	0.850	15.879	2,866	4,652	0.663	24.561	1,474	2,341
세종	0.620	20.388	4,613	6,443	0.895	14.008	986	2,062
경기도	0.857	16.668	4,507	6,901	0.853	23.146	2,218	3,513
강원도	0.931	10.816	1,333	2,266	0.782	18.990	1,105	1,819
충북	0.904	11.860	1,746	3,053	0.818	17.215	913	1,536
충남	0.907	12.792	1,968	3,215	0.813	14.254	918	1,497
전북	0.922	9.668	1,482	2,835	0.791	18.434	990	1,540
전남	0.952	6.455	751	1,587	0.728	31.525	1,565	2,393
경북	0.916	11.237	1,620	2,830	0.737	20.080	905	1,545
경남	0.874	12.452	2,224	3,743	0.748	19.700	1,140	1,782
제주도	0.905	10.555	1,871	3,194	0.718	34.602	4,066	7,125

□ 2021년

(단위: %, 만 원)

전국	R^2	MAPE	MAE	RMSE
단독	0.798	25.119	2,153	3,539
다세대	0.880	15.882	2,219	3,957
아파트	0.881	15.245	4,892	8,237
오피스텔	0.913	8.615	1,589	3,203

□ 2022년

(단위: %, 만 원)

시도	공동주택				단독주택			
	R^2	MAPE	MAE	RMSE	R^2	MAPE	MAE	RMSE
전국	0.922	12.481	3,363	5,874	0.799	23.847	2,187	3,536
서울	0.914	13.159	4,835	8,535	0.677	28.481	3,006	4,613
부산	0.909	11.371	2,339	3,847	0.465	36.218	1,916	2,944
대구	0.909	10.494	2,552	3,916	0.659	26.331	1,716	2,608
인천	0.875	13.338	2,741	4,208	0.749	26.676	1,905	3,236
광주	0.914	8.674	1,720	3,032	0.672	26.294	1,487	2,247
대전	0.902	12.150	2,563	3,682	0.787	19.977	1,519	2,139
울산	0.866	12.947	2,732	4,037	0.742	25.123	1,508	2,490
세종	0.788	11.581	2,566	3,575	0.825	15.298	945	1,318
경기도	0.901	13.451	3,556	5,286	0.846	21.819	2,311	3,557
강원도	0.919	10.856	1,471	2,437	0.736	19.812	1,339	2,265
충북	0.899	11.005	1,779	2,996	0.820	18.828	1,108	1,811
충남	0.920	11.334	1,803	2,761	0.810	16.439	1,055	1,762
전북	0.918	9.733	1,639	2,863	0.789	18.714	1,098	1,616
전남	0.943	6.354	816	1,809	0.757	24.042	1,398	2,323
경북	0.903	11.219	1,734	2,963	0.776	19.107	1,021	1,679
경남	0.882	11.280	2,106	3,501	0.760	19.845	1,361	2,292
제주도	0.864	11.311	2,395	4,329	0.739	31.862	3,890	5,640

□ 2022년

(단위: %, 만 원)

전국	R^2	MAPE	MAE	RMSE
단독	0.799	23.847	2,187	3,536
다세대	0.864	15.385	2,354	4,394
아파트	0.918	12.339	3,866	6,487
오피스텔	0.925	8.622	1,586	2,774

□ 2023년

(단위: %, 만 원)

시도	공동주택				단독주택			
	R^2	MAPE	MAE	RMSE	R^2	MAPE	MAE	RMSE
전국	0.954	9.171	2,535	4,593	0.803	22.669	2,210	3,482
서울	0.946	10.034	3,909	6,937	0.662	27.372	3,015	4,543
부산	0.924	9.035	1,765	2,937	0.544	41.717	1,873	2,684
대구	0.911	9.567	2,005	3,210	0.632	24.551	1,609	2,297
인천	0.918	9.765	1,950	2,966	0.794	20.573	1,678	2,581
광주	0.933	7.266	1,416	2,550	0.905	30.873	1,760	2,456
대전	0.913	9.948	2,082	3,050	0.721	18.579	1,556	2,268
울산	0.920	10.206	1,818	2,751	0.737	21.509	1,495	2,520
세종	0.825	9.633	2,092	2,822	0.794	19.719	1,011	1,490
경기도	0.945	9.258	2,494	3,814	0.863	18.801	2,138	3,232
강원도	0.944	7.371	1,069	1,825	0.644	23.908	1,788	3,238
충북	0.949	6.657	1,044	1,984	0.813	19.185	1,219	1,798
충남	0.935	8.608	1,368	2,166	0.785	15.984	1,150	1,977
전북	0.935	7.539	1,170	2,170	0.800	19.385	1,119	1,571
전남	0.959	5.232	585	1,348	0.730	40.403	1,825	2,845
경북	0.950	7.535	1,069	1,774	0.724	22.408	1,349	2,097
경남	0.939	7.583	1,391	2,269	0.783	20.990	1,408	2,202
제주도	0.892	10.136	2,000	3,544	0.827	21.238	3,442	5,028

□ 2023년

(단위: %, 만 원)

전국	R^2	MAPE	MAE	RMSE
단독	0.803	22.669	2,210	3,482
다세대	0.892	13.018	2,016	3,694
아파트	0.953	8.740	2,778	4,935
오피스텔	0.949	7.108	1,291	2,255

<부표 2> 예측모델 평가지표 (IQR=±1.25)

□ 2021년

(단위: %, 만 원)

시도	공동주택				단독주택			
	R^2	MAPE	MAE	RMSE	R^2	MAPE	MAE	RMSE
전국	0.889	15.418	4,157	7,439	0.779	26.476	2,210	3,682
서울	0.877	16.116	6,068	10,646	0.645	31.544	3,061	4,785
부산	0.897	13.007	2,445	4,294	0.458	41.554	2,108	3,385
대구	0.903	12.269	3,037	4,755	0.615	31.038	1,866	2,791
인천	0.859	15.721	3,102	5,134	0.766	30.327	1,833	2,770
광주	0.897	10.678	2,016	3,583	0.704	26.299	1,478	2,247
대전	0.831	18.025	3,891	5,607	0.775	20.980	1,556	2,382
울산	0.854	16.208	2,795	4,649	0.705	23.499	1,557	2,352
세종	0.579	22.481	4,975	6,835	0.929	17.562	1,224	2,535
경기도	0.853	17.263	4,602	7,079	0.835	24.396	2,289	3,696
강원도	0.920	12.002	1,453	2,472	0.708	21.962	1,295	2,182
충북	0.904	12.385	1,825	3,167	0.801	20.431	988	1,581
충남	0.901	13.733	1,999	3,398	0.763	17.336	1,043	1,834
전북	0.910	10.682	1,529	2,959	0.792	20.486	1,104	1,788
전남	0.944	7.311	816	1,723	0.747	26.944	1,291	1,951
경북	0.908	12.192	1,695	2,980	0.707	21.328	954	1,689
경남	0.872	12.745	2,235	3,816	0.691	21.137	1,234	2,039
제주도	0.896	13.381	2,054	3,441	0.647	31.415	4,371	6,569

□ 2021년

(단위: %, 만 원)

전국	R^2	MAPE	MAE	RMSE
단독	0.779	26.476	2,210	3,682
다세대	0.876	16.836	2,286	4,120
아파트	0.879	15.936	5,007	8,472
오피스텔	0.905	9.090	1,639	3,283

□ 2022년

(단위: %, 만 원)

시도	공동주택				단독주택			
	R^2	MAPE	MAE	RMSE	R^2	MAPE	MAE	RMSE
전국	0.919	12.782	3,434	6,027	0.783	25.187	2,275	3,760
서울	0.911	13.364	4,937	8,767	0.669	30.347	3,142	4,884
부산	0.903	11.981	2,428	4,096	0.367	43.425	2,052	3,483
대구	0.911	11.132	2,603	3,959	0.628	27.492	1,760	2,712
인천	0.874	13.893	2,801	4,292	0.777	25.559	1,999	3,240
광주	0.901	9.225	1,758	3,173	0.701	26.317	1,528	2,199
대전	0.895	12.732	2,629	3,809	0.761	20.151	1,560	2,314
울산	0.877	12.837	2,717	4,049	0.722	24.588	1,541	2,468
세종	0.787	11.464	2,531	3,456	0.842	15.505	1,310	3,124
경기도	0.897	13.748	3,642	5,419	0.829	23.298	2,403	3,765
강원도	0.921	10.152	1,481	2,460	0.795	20.303	1,248	1,994
충북	0.898	10.852	1,715	2,974	0.782	20.290	1,161	1,899
충남	0.918	11.467	1,782	2,708	0.798	16.505	1,110	1,857
전북	0.921	10.235	1,655	2,806	0.778	19.394	1,151	1,768
전남	0.949	6.241	810	1,723	0.843	25.893	1,416	2,479
경북	0.900	12.039	1,769	2,991	0.680	20.756	1,106	1,901
경남	0.881	11.676	2,206	3,669	0.764	21.045	1,270	2,080
제주도	0.837	12.892	2,734	5,174	0.631	26.688	4,420	8,463

□ 2022년

(단위: %, 만 원)

전국	R^2	MAPE	MAE	RMSE
단독	0.783	25.187	2,275	3,760
다세대	0.876	15.705	2,356	4,137
아파트	0.914	12.632	3,958	6,712
오피스텔	0.922	8.970	1,634	2,887

□ 2023년

(단위: %, 만 원)

시도	공동주택				단독주택			
	R^2	MAPE	MAE	RMSE	R^2	MAPE	MAE	RMSE
전국	0.953	9.552	2,601	4,697	0.796	23.789	2,283	3,561
서울	0.945	10.397	3,976	7,034	0.674	28.117	3,032	4,454
부산	0.925	9.404	1,795	2,959	0.536	36.952	1,796	2,576
대구	0.913	10.111	2,089	3,347	0.580	26.117	1,708	2,601
인천	0.911	10.290	2,051	3,165	0.813	22.510	1,666	2,537
광주	0.930	7.589	1,391	2,569	0.584	29.853	1,683	2,290
대전	0.908	10.383	2,189	3,211	0.679	19.737	1,656	2,585
울산	0.909	10.961	1,954	3,025	0.803	21.028	1,484	2,420
세종	0.812	10.498	2,261	3,074	0.759	22.926	1,364	2,117
경기도	0.942	9.576	2,562	3,952	0.847	20.994	2,257	3,471
강원도	0.943	8.471	1,118	1,900	0.722	23.432	1,684	2,811
충북	0.954	6.645	1,026	1,885	0.788	18.260	1,223	1,854
충남	0.931	8.825	1,439	2,307	0.794	17.271	1,273	2,030
전북	0.937	7.418	1,161	2,093	0.697	20.366	1,294	1,985
전남	0.965	4.844	569	1,256	0.681	32.104	1,828	3,067
경북	0.941	8.214	1,120	1,935	0.704	22.255	1,369	2,430
경남	0.932	8.408	1,479	2,532	0.749	22.022	1,554	2,692
제주도	0.896	10.705	2,131	3,490	0.801	27.245	3,770	5,117

□ 2023년

(단위: %, 만 원)

전국	R^2	MAPE	MAE	RMSE
단독	0.796	23.789	2,283	3,561
다세대	0.895	13.813	2,060	3,668
아파트	0.952	9.067	2,855	5,063
오피스텔	0.943	7.360	1,321	2,336

Abstract

A Study on the Measurement of Monthly Rent of Owner-occupied Dwellings for the Estimation of Imputed Rental Income of Housing

Song Jung-hyun, Byun Jun-seok

This study aims to develop a method for estimating imputed rental income from owner-occupied housing, utilizing data from the Jeonse and Monthly Rent Reporting System and applying machine learning techniques, in order to incorporate the value of housing services that owner-occupied dwellings provides into household income.

Based on jeonse contract information obtained from the Jeonse and Monthly Rent Reporting System of the Ministry of Land, Infrastructure and Transport (MOLIT) during the 2021-2023 period, this study combined the average jeonse and sale prices from the Korea Real Estate Board's Housing Price Trend Survey with housing complex characteristics from the MOLIT's Building Register. For analysis, housing units were classified into multi-family housing units (apartments, apartment units in a private house, row houses, officetels) and detached houses (single-family houses and multi-family houses). With housing characteristics (area, floor, year of construction), housing complex characteristics (land, building type, gross floor area, number of households) and variables in local markets as explanatory variables, this study developed a model to predict jeonse deposits by type.

Random Forest, a tree-based ensemble technique, was used for estimation. To deal with outliers, IQR coefficients were verified for jeonse deposits per unit area, and cross-validation was performed on the number of trees (n_estimators). These procedures ensured a consistent estimation performance and a moderate computation time.

For the final estimation model, the first applied it to jeonse households in the Survey of Household Finances and Living Conditions and examined the differences between the predicted and reported jeonse deposits. Second, we used the prediction model to estimate jeonse deposits for owner-occupied and rent-free households, and then applied the jeonse-monthly rent conversion rate and the interest rate to these predicted deposits to derive imputed rental income.

Using administrative data and machine learning techniques, this research developed an estimation model to measure the value of housing services of owner-occupied housing units. This model was applied to the Survey of Household Finances and Living Conditions to verify its performance and applicability. This study is meaningful in that the method of reflecting the characteristics of housing units, housing complexes and local markets in detail is proposed, and the possibility of measuring the value of social transfer in cash of public rental housing units is presented. It is expected that by adding variables that enhance explanatory power and further advancing this model in the future, this model will be utilized to estimate imputed rental income from owner-occupied dwellings and produce related statistics.

Key words: imputed rental income from owner-occupied housing units, machine learning, real transaction value of jeonse and monthly rental

연구진

○ 송정현(국가데이터처 국가데이터연구원 경제사회통계연구실 주무관)

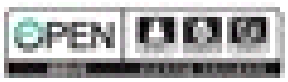
○ 변준석(국가데이터처 국가데이터연구원 경제사회통계연구실 사무관)

* 연구진의 소속 및 직급은 연구과제 완료 시 기준임을 알려드립니다.

연구보고서 2025-13

의제주거소득 추정을 위한 자가주택의 월세평가 방법 연구

인 쇄	2026년 3월
발 행	2026년 3월
발 행 인	김 진
발 행 처	국가데이터처 국가데이터연구원 35220 대전광역시 서구 한밭대로 713 TEL.(042)366-7100 Fax.(042)366-7123
홈페이지	https://mods.go.kr/dsri/
ISSN(Online)	2733-4120



국가데이터처
국가데이터연구원