



기계학습을 활용한 소득계층별 통근시간 영향요인*

Influencing Factors of Commuting Time by Income Class : A Machine Learning Approach

김성주** · 송재민***
Kim, Sungjoo · Song, Jaemin

Abstract

This study aims to investigate determinants of commuting time among diverse income groups in the Seoul Metropolitan Area, contributing to a reduction in carbon emissions and an enhancement in the quality of life for workers. The XGBoost machine learning technique, coupled with data from the Household Travel Survey, was used to evaluate the effect of factors such as housing price, attributes of residential and workplace locales, and individual parameters on commuting time. The goal is to highlight the key influences on commuting time for each income stratum. The findings underscore a compensatory relationship between housing prices and commuting time across all income classes. In addition, workplace proximity and access to public transportation emerged as cardinal determinants of commuting duration for low-income earners. For individuals in the middle- and high-income brackets, however, the quality of the residential setting assumes greater importance in shaping commuting time. The results prompted the recommendation of cost-effective housing near employment nodes as an effective strategy for reducing commuting durations across all income categories. It is important to note that providing such affordable housing near public transportation appears more pivotal for low-income earners. For middle- and high-income earners, the improvement of the residential milieu proximity to workplaces is critical in curtailing commuting duration.

주제어 통근시간, 주택가격, 거주지 특성, 기계학습, 비선형 관계
Keywords Commuting Time, Housing Price, Residential Characteristics, Machine Learning, Non-linear Relationship

1. 서론

통근시간을 줄이는 것은 다양한 환경적, 사회적 이점이 있다. 통근시간 감소는 통근에 소요되는 교통에너지 소비와 탄소배출을 줄이며, 통근시간 감축으로 인한 개인 시간의 증가는 삶의 질 향상에 기여한다(장재민 외, 2019). 도시계획 분야에서 직주근접의 실현을 통한 통근거리 및 통근시간 축소는 오랜 기간 공간계획의 주요한 목표로 남아있다.

통근시간을 줄이기 위해서는 다양한 요인들이 고려되어야 한

다. 다양한 요인들 중 주거비용은 통근시간 결정에 큰 영향을 끼칠 수 있다. 어디서 살 것인가와 어디서 일할 것인가는 통합된 결정일 수 있기 때문이다(Molho, 1986). 오랜 기간 도시경제 이론에서는 주거비용과 통근비용이 상쇄관계를 지니고 있음을 주장해 왔다. 현대 도시경제이론의 토대가 되는 Alonso의 주거입지 이론에 따르면 가계는 교통비용과 주거비용의 상쇄관계(trade-off) 속에서 효용을 극대화하는 주거입지를 선택한다(전명진·안현주, 2016). 특히, 대도시를 중심으로 한 부동산 가격의 상승은 근로자를 일자리로부터 떨어진 주택 가격이 낮은 지역으로 이동

* 이 연구는 서울대학교 환경계획연구소에서 지원되는 연구비에 의하여 수행되었음.

** Graduate Student, Graduate School of Environmental Studies, Seoul National University (First Author: sungjoo@snu.ac.kr)

*** Associate Professor, Graduate School of Environmental Studies, Seoul National University (Corresponding Author: jaemins@snu.ac.kr)

하게 만들어 통근시간 및 거리가 더 길어지는 양상이 나타날 수 있다(Blumenberg and King, 2021). 또한, 주거입지를 설명하는 시장중심 접근방법 이론인 접근성-공간 맞교환 이론(Access-Space Trade-off)에 따르면, 가구의 주거입지 선택은 고용 접근성에 대한 고려와 삶의 질에 대한 고려의 상쇄관계에 의해 결정된다(Phe and Wakely, 2000; Kim et al., 2005; Moos and Skarburskis, 2010; 김승남·안건혁, 2011). 따라서 주거비용 이외에도 지역의 교육(Kim et al., 2005), 안전(Weisbrod et al., 1980), 도시 어메니티(Rapoport, 1980; Tyrvaenen and Vaananen, 1998; Rouwendal and Meijer, 2001; Parkes et al., 2002) 등 거주지의 환경적 질을 나타내는 특성 또한 거주지 선택의 주요 요인으로 고려될 수 있다(김승남·안건혁, 2011).

이와 같이, 주거비용과 거주지의 환경적 질을 나타내는 거주지 특성은 고용접근성과 긴밀한 관계를 가지고 있다. 이에 있어, 소득계층별로 주거비용 및 거주지 특성이 통근시간 결정에 미치는 영향력은 다를 수 있다. 왜냐하면, 저소득 계층은 한정된 재원으로 인해 주거비 및 교통비 상승에 매우 민감하게 반응하지만, 고소득층은 거주지 선택 시 자가용을 주로 사용하기 때문에 대중교통 접근성에 대한 고려 우선순위가 낮은 반면 거주지의 다양한 어메니티에 대한 선호가 높기 때문이다(Adair et al., 2000; 하재현, 2020).

우리나라를 대상으로 진행한 이민주·박인권(2016) 연구에 따르면 고소득층보다 저소득층의 통근시간이 길게 나타나며, 주택가격이 높은 지역일수록 해당 지역으로 통근하는 저소득층과 고소득층의 통근시간 격차가 큰 것으로 나타났다. 하지만, 일부 선행연구에서는 상반된 결과가 도출되기도 했다. 이병승(1998)의 연구에서는 고소득자가 원거리 통근을 하고 있는 것으로 나타났으며, Park and Quercia(2015)는 미국 고소득 근로자의 경우 통근시간의 감소보다 임대료의 상승이 더욱 가파른 경향이 있음을 보여주었다. 이는 고소득층의 직장이 고용중심지, 즉 지대가 더욱 높은 지역에 위치할 가능성이 있으므로, 주택가격이 통근시간에 미치는 영향이 저소득층보다 더욱 영향이 클 수도 있음을 의미하고 있다. 이처럼, 주거비와 통근의 관계는 여전히 분명하지 않으며(Islam and Saphores, 2022), 특히 소득계층별 주거비용을 포함한 거주지 특성이 통근시간에 미치는 영향 차이를 분석한 연구는 매우 부족하다.

이와 같은 배경에서 본 연구의 목적은 소득계층별 거주지 특성과 통근시간과의 관계 및 그 차이를 분석하는 것이다. 본 연구에서는 기계학습 분석을 통해 전통적 회귀모형에 비해서 설명력이 높은 모델을 구축하고 요인 간의 비선형 관계를 파악하고자 하였다. 이를 위해, 본 연구는 2016년 가구통행실태조사 자료를 활용하여 수도권을 대상으로 기계학습 기법 중 하나인 XGBoost를 사용하여 소득계층별 통근시간 영향요인에 대하여 분석하였다. 본 연구는 수도권에서 주거비 및 거주지 특성이 통근시간에 미치

는 영향의 상대적 크기를 실증적으로 파악하고, 이를 기반으로 직주근접을 위한 소득계층별 주거정책의 차별화에 중요한 시사점을 제시하고자 한다.

II. 이론 및 선행연구 고찰

1. 통근비용과 주거비용의 관계

고전 도시경제이론에서는 단핵도시 가정하에, 도시의 중추기능과 핵심 업무지역이 위치한 도심을 기준으로 거리가 멀어질수록 주거비용은 감소하는 반면, 통근비용은 증가하게 되어 주거비용과 통근비용이 상쇄관계(Alonso, 1964)를 지닌다고 설명한다. 이에 따라 가구는 효용을 극대화하는 최적의 주거입지를 선택하게 된다(최준성·노승한, 2021). 입찰지대이론을 바탕으로 주거비용과 통근비용의 관계를 살펴본 연구는 다음과 같다. Islam and Saphores(2022)는 로스앤젤레스 카운티를 대상으로 일반구조방정식 모형(Generalized Structural Equation Model)을 사용하여 거주지와 직장지 주변의 토지이용 특성을 통제하고, 주택가격이 통근거리에 미치는 영향을 살펴보았다. 연구 결과, 주거비용이 높은 동네에 거주하는 가구는 그렇지 않은 가구보다 통근거리 및 통근시간이 짧게 나타났다. Park and Quercia(2015)는 미국 8개 대도시권을 대상으로 주거비용과 통근시간의 관계를 파악하였다. 이 연구에서는 현대 대도시권의 고용중심지 다핵화와 분산에도 불구하고, 가계는 여전히 근본적인 주거비용과 통근비용의 상쇄관계가 존재한다고 밝혔다. 특히, 고소득 근로자는 통근시간의 감소보다 임대료의 상승이 더욱 가파르게 나타나, 오히려 더 긴 통근시간을 가지고 있는 것으로 나타났다. 한편, Hass and Osland(2013)은 통근, 이주, 주택과 노동 시장의 복잡한 상호관계에 대한 문헌연구를 통해서 소득과 소득 전망은 주택비 부담과 밀접하게 관련이 있으며, 따라서 주택 가격 또는 소득 전망의 공간적 변동은 통근 거리를 결정할 수 있다고 설명하고 있다.

우리나라를 대상으로 한 주요 연구 결과는 다음과 같다. 전명진·안현주(2016)는 수도권 통근자를 대상으로 주거비와 통근기간 상쇄관계의 존재 여부를 실증적으로 검증했다. 통근시간에 대한 소득별 기회비용을 추정하고, 통근시간 기회비용을 주거비용 결정 요인으로 반영하였다. 분석 결과, 통근비용이 한 단위 증가할 때 주거비용이 3.5% 감소하는 것을 확인하여 주거비용과 통근비용 간의 상쇄관계가 존재함을 입증하였다. 최준성·노승한(2021)은 수도권 신도시를 대상으로 통근시간의 가치를 산출해 주거비 결정 요인모형의 주요 변수인 통근기회비용을 계산하고, 이를 통해 통근비용과 임차료 간의 상쇄관계를 규명하였다. 이 연구에 따르면 신도시의 경우 단핵도시 모형의 가정과 달리 직장은 도시 외부에 분포하고 거주지는 도시 내부에 위치한 경우가 많은데, 이러한 공간 특성을 가진 신도시에서는 통행비용과 주거비

용 간 상쇄관계가 존재하지 않는 것으로 나타났다. 이는 신도시 내에 밀집된 도심이 존재하지 않는다는 것을 의미하며 더 나아가 신도시의 자족수준이 낮다는 것으로 해석할 수 있다고 설명하고 있다. 박미선 외(2018)는 수도권 주거비 부담은 월소득 대비 19.6% 수준이었으나 교통비를 포함하는 경우 23.3%까지 상승하는 것을 확인하였다. 소득수준별로는 소득수준이 낮을수록 주거비 부담수준이 높았으며, 가구소득이 300만원 미만인 경우 주거교통부담 수준은 30%에 육박한다고 밝혔다.

2. 주거비용 외 통근시간 및 통근행태 결정요인

통근시간 및 통근행태는 주거비용 이외에도 지역 및 개인적 특성에 의해 영향을 받는다. 장재민·김태형(2016)은 서울시 25개 자치구를 대상으로 10년간의 통근시간 변화에 미치는 영향변수를 도출하였다. 아파트 시세차익과 종사자수는 통근시간 변화량에 영향을 미치는 변수로 도출되었으며, 이는 해당지역에 종사자수가 늘어 통근시간이 감소하여 통근시간의 시간가치가 아파트 시세에 반영된 것으로 해석하고 있다. 이와 같은 결과는 통근시간과 아파트 가격의 연관성을 시사하고 있다. 이반송(1998)은 소득, 주거지 소유 형태 등을 포함한 통근자의 특성, 거주지 이동, 통근목적지와 주거지의 특성이 통근거리 결정에 어떠한 영향을 미치는지를 연구하였다. 연구 결과, 고소득자가 원거리 통근을 하는 것으로 나타났으며, 도심 및 부도심에 속하는 서울 중심지에 거주하는 통근자의 통근거리가 비중심지에 거주하는 통근자에 비해 단거리임을 확인하였다. 또한, 서울시 내에서 이주한 서울 거주 통근자는 이주 후에 오히려 통근거리가 멀어지는 것으로 나타났다. 경기도 등 신도시로의 이주는 자가소유의 증대와 주택의 질 향상에는 도움이 되었으나, 통근거리를 현저히 상승시켰다. 전명진·이지현(2020)은 수도권을 대상으로 통근자의 직장 접근성과 주택 접근성이 통근시간에 미치는 영향을 분석하였다. 분석 결과, 서울에서 일하는 통근자의 경우 실제 거주하는 주택 유형과 동일한 주택에 대한 접근성이 높음에도 직장에서 먼 곳에 주거지를 결정하는 것으로 나타났다. 반면, 경기 및 인천의 교외지역에서 일하는 통근자의 경우 직장 주변의 주택 접근성이 높으면 통근시간이 짧아지는 것으로 나타났다. 이와 같은 결과는 거주지와 직장지의 특성이 통근시간에 영향을 미칠 수 있으며, 특히 직장지의 높은 주택 가격은 실질적인 주택 접근성을 저해하는 것을 의미한다. 송재민(2021)은 정량 분석 기반의 선행연구를 통합적으로 분석하기 위한 메타분석을 통해, 도시형태 특성 중 인구밀도가 승용차 사용, 대중교통 및 비동력 통행에 가장 큰 영향을 미친다고 밝히고 있다. Zheng et al.(2022)은 도시 건조 환경은 유입통근거리와 유출통근거리에 서로 다른 영향을 미치고 있음을 보였다. 거주 인구, 오락 시설, 복합 개발은 유출통근거리를 증가시키는 요인이지만, 노동 인구, 교통 시설은 유입통근거리를 증

가시키는 것으로 나타났다.

한편, 최근 다수의 선행연구에서 통근행태와 도시 건조 환경 간의 비선형적 관계가 있음이 연구되고 있다. Liu et al.(2021)은 출발지와 도착지의 모든 건조 환경 변수는 통근과 쇼핑 목적의 실제 통행에서 비선형적 관계를 지니고 있음을 확인했다. Ding et al.(2018a)과 Ding et al.(2018b)의 연구에서도 연구 대상지는 워싱턴과 오슬로로 상이하나 두 연구 모두에서 건조환경이 통근 수단 및 통근 거리에 미치는 영향이 비선형적임을 보이고 있다.

3. 소득계층별 통근시간 및 통근행태 결정요인

이민주·박인권(2016)은 서울시를 대상으로 소득에 따른 직주 불일치 현상을 확인하기 위해 개인 특성과 통근목적지의 지역 특성이 통근시간에 미치는 영향을 위계선형모형을 활용하여 분석하였다. 분석 결과, 고소득층보다 저소득층의 통근시간이 길게 나타났으며, 주택가격이 높은 지역일수록 해당 지역으로 통근하는 저소득층과 고소득층 간 통근시간 격차가 커지는 것을 확인했다. 이는 소득에 따라 통근행태가 실제로 다르게 나타나고 있으며, 주택 가격이 통근시간에 미치는 영향이 소득계층에 따라 차이를 시사하고 있다. 하재현(2020)은 미국 대도시권 지역을 대상으로 한 실증연구에서 고소득 통근자의 경우 자동차 의존도가 높아 긴 통행거리를 통근하고 직장이나 주거지 선택에 있어 입지 이외의 다양한 요인이 고려되고 있음을 보이고 있다. 하지만, 중소득 이하의 통근자는 대중교통 수단에 대한 의존도가 높아, 도심 집중화 정도가 높을수록, 대중교통 결정지 주변으로 주거지를 결정하는 경향이 파악되었다. 이 결과는 도시공간구조 특성이 통근 거리에 미치는 영향이 소득수준에 따라 차별적이라는 점을 시사하고 있다. 김희철·안건혁(2011)은 고밀개발, 용도혼합, 직주근접 등과 같은 압축도시 계획요소들이 통행거리에 미치는 영향이 소득계층별로 어떻게 차별적으로 나타나는지 분석하였다. 분석 결과, 저소득층에서는 밀도, 용도혼합, 고용접근성 변수가 통근거리 감소에 영향을 미쳤지만, 고소득층은 고용접근성 변수만 통근거리 감소에 영향을 미쳐 소득계층별로 압축도시 계획 요소의 통근거리 감소효과와 차이가 존재함을 입증했다. 특히, 고용접근성의 통근거리 감소효과는 저소득층보다 고소득층에서 크게 도출되었다. 주거지의 고용접근성이 증가한다는 것은 주거지가 고용 중심지에 가까워지면서 통근거리가 감소하게 되는 것을 의미하는데, 이러한 관계를 주거비용, 통행비용, 개인의 시간가치 사이의 교환관계(trade-off)로 해석했다. 고용접근성의 통근거리 감소 효과가 고소득층에서 더욱 크게 나타났다는 것은 시간가치가 높은 개인일수록 고용 중심지 근처에 거주하여 주로 통행시간이나 통행거리로 나타나는 통행비용을 줄이려는 행동에서 기인한다고 이야기했다. 더 나아가 이는 고소득층이 고용중심지와 가까운 주거입지 선정에 있어서 보다 유리함을 시사한다.

4. 연구의 차별성

선행연구를 종합해보면 통근비용과 주거비용의 상쇄가 명확하게 나타나지 않아 통근비용으로 대변될 수 있는 통근시간과 주거비용과의 관계를 살펴볼 필요가 있으며, 또한 소득계층별 차등적 영향에 대한 심도 깊은 연구가 필요함을 확인할 수 있다. 이와 같은 배경에서 본 연구는 다음과 같은 점에서 선행연구와 차별된다.

첫째, 통근시간에 영향을 미치는 요인과 통근비용과 주거비용의 관계를 분석한 연구는 다수 있었으나, 소득계층별로 나누어 주거비용을 포함한 주거지 특성과 통근시간과의 관계를 실증적으로 분석한 연구는 매우 부족하다. 선행연구에 따르면 소득은 통근시간 결정에 중요한 영향을 미치며, 통근 목적지의 지역 특성, 특히 통근 목적지 주거비용과 통근시간이 관계있는 것으로 나타났다. 이는 통근시간 결정에 있어 소득계층별로 주거비용의 영향이 있음을 시사하고 있으나, 선행연구에서는 소득계층별로 영향을 미치는 요인이 어떻게 달라지는지, 또한 목적지와 주거지 간의 주거비 차이가 어떠한 영향을 미치는지에 대한 분석이 미비하다. 따라서, 소득계층을 구분하여 통근시간에 미치는 영향요인 차이를 식별하는 것은 직주근접에 대한 정책적 시사점을 도출하는 데에 매우 중요한 연구 주제이다. 둘째, 대부분의 선행연구에서는 주택가격과 통근시간 및 통근거리의 관계를 선형 관계에 기반하여 분석해 왔다. 본 연구에서는 기계학습 기법 중 하나인 XGBoost를 활용하여 주택가격과 통근시간 간 비선형적 관계를 파악하고, 나아가 주택가격 및 거주지 특성이 통근시간에 영향을 미치는 임계치를 확인하고자 한다.

III. 분석방법

1. 연구의 범위 및 분석 자료

본 연구의 시간적 범위는 2016년, 공간적 범위는 수도권 중 도시교통정비 촉진법에서 지정한 서울특별시의 교통권역을 대상으로 한다. 도시교통정비 촉진법 제2장 제4조에 따르면, 교통권역이란 도시교통정비지역 중 같은 교통생활권에 있는 둘 이상의 인접한 도시교통정비지역 간 연계된 교통관련 계획을 수립할 수 있도록 국토교통부장관이 지정·고시한 것으로, 서울특별시를 중심으로 교통 및 통행에 있어 밀접한 관계를 지닌 지역이기에 이에 기반하여 본 연구의 공간적 범위를 설정하였다. 연구 대상지는 <표 1>과 같다.

본 연구에서는 2016년 가구통행실태조사 응답자료 중 연구의 공간적 범위 내 거주지 및 직장지가 모두 위치한 응답자를 대상으로 진행하였으며 거주지와 직장지가 같은 행정동에 위치한 응답자는 제외하였다. 결측치를 제외하고 총 69,124건의 자료가 연구에 활용되었다. 소득계층은 선행연구(오지에 외, 2019; 이해승·

Table 1. Spatial scope of the study

Urban transportation maintenance area	Transportation area (excluding Ongjin-gun, Incheon)	
	State/City	City/Gun/Gu
Incheon		Jung-gu, Dong-gu, Nam-gu, Yeonsu-gu, Namdong-gu, Bupyeong-gu, Gyeong-gu, Seo-gu, Ganghwa-gun
		Seongnam-gu, Seongnam-si, Jungwon-gu, Seongnam-si, Bundang-gu, Seongnam-si, Uijeongbu-si, Anyang-si Manan-gu, Anyang-si Dongang-gu, Bucheon-si Wonmi-gu, Bucheon-si Sosa-gu, Bucheon-si Ojeong-gu, Gwangmyeong-si, Goyang-si Deokyang-gu, Goyang-si Ilsan-dong-gu, Goyang-si Ilseo-gu, Gwacheon-si, Guri-si, Namyangju-si, Hanam-si, Gimpo-si, Yangju-si

이희연, 2009)를 바탕으로 가구 전체의 월 소득 200만원 미만의 응답자는 저소득층, 200만원 이상 500만원 미만의 경우 중소득층, 500만원 이상의 응답자는 고소득층으로 분류하여 분석하였다.

2. 변수 구성

본 연구에서는 선행연구를 반영하여 주요 변수를 선정하였다. 분석에 사용된 변수는 <표 2>와 같으며, 각 변수의 구축방법은 다음과 같다. 종속변수인 통근시간은 2016년 가구통행실태조사에 기재된 응답자의 출발시각과 도착시각을 바탕으로 산정하였다. 독립변수인 주택가격은 2016년 국토교통부 실거래가 자료를 활용하여 구축하였다. 2016년 연구 범위 내 아파트 241,664건, 오피스텔 23,482건, 연립, 다세대 및 다가구주택 114,824건, 단독주택 13,663건의 거래를 활용하여, 연구 범위 내 각 행정동의 주택유형별 평균 평당거래가격을 산출하였다. 주택가격변화는 2011년 대비 2016년에 아파트 매매가격이 얼마나 변화하였는지를 측정하였으며, 전희정·정수영(2021)의 평균 주택가격 변화 측정 방법을 사용하였다. 전희정·정수영(2021)은 물가 차이 및 시기별 인플레이션에 따른 주택가격 상승을 통제하고자 광역권의 m²당 평균 주택가격에 대한 행정동의 m²당 평균 주택가격의 상대적 비율을 계산하였다. 즉, 수도권 대비 행정동별 평균 주택가격을 기준으로, 두 시기의 상대적 비율의 변화를 평균 주택가격 변화로 측정하는 것이다. 예를 들어, 서울시 한 행정동의 2011년 수도권 대비 상대적 평균 주택가격이 100%, 2016년 수도권 대비 상대적 평균 주택가격이 120%일 경우 해당 근린은 상승한 것으로 설명할 수 있다. 최종적으로 주택가격변화는 2011년, 2016년의 상대적 비율의 변화에 로그를 취하여 식 (1)과 같이 분석하였다.

Table 2. Variables definition

Category		Data		Source	
Dependent variable	Commute times	Respondent's commute time (min)		2016 Household Travel Survey	
	Housing price	Housing prices at home	Average price per 3.3 m ² by housing type in respondent's administrative district (KRW 10,000/3.3 m ²)		
		Housing price difference between residential and workplace	(Housing price at work)-(Housing price at home) (KRW 10,000/3.3 m ²)	Ministry of Land, Infrastructure, and Transport	
Neighborhood characteristics at home	Population	Total number of people		KOSIS	
		Population density (number/km ²)			
	Economic characteristics	Number of businesses			
		Number of employees per business			
		Housing price change (2011 vs. 2016 apartment price change)			Ministry of Land, Infrastructure, and Transport
	Physical characteristics	Percentage of apartments (%)			KOSIS
		Land mix			Ministry of Land, Infrastructure, and Transport
		Distance from the center of the administrative district of the residence to the nearest subway station (km)			Calculation
		Number of private institutes			KEDI
	Neighborhood characteristics at work	Population	Total population		KOSIS
Population density (number/km ²)					
Economic characteristics		Number of businesses			
		Employees per business			
Physical characteristics		Percentage of apartments (%)		Calculation	
		Distance from the center of the administrative district of the workplace to the nearest subway station (km)			
Distance from the center of the administrative district of the workplace to the nearest Seoul 3-city center (km)					
Individual characteristics	Gender	Gender of respondents	1: Male 0: Female	2016 Household Travel Survey	
	Age	Respondent's age			
	Household size	Number of people in the respondent's household			
	Relationship	Head of household	1: Yes 0: No		
	Occupation	Administrative/clerical worker	1: Yes 0: No		
		Service/sales worker	1: Yes 0: No		
		Semi-skilled worker	1: Yes 0: No		
	Housing types	Apartment	1: Yes 0: No		
	Residence area	Seoul	1: Yes 0: No		
	Commuting mode	Public transportation	1: Yes 0: No		
Walking/Biking		1: Yes 0: No			

$$\log(y_{ij,t_2}/y_{ij,t_1}) \quad (1)$$

여기서, y_{ij} 는 수도권 j 의 3.3m²당 평균 아파트 매매가격에 대

한 읍면동 i 의 3.3m²당 평균 아파트 매매가격의 상대적 가격이며 t_1 은 2011년, t_2 는 2016년을 의미한다. 식을 통해 산출된 주택가격 변화가 음수일 경우 상대적 가격이 하락, 양수일 경우 상대적 가

격이 증가했음을 나타낸다.

토지혼합도는 엔트로피로 측정하였으며, 국토교통부에서 제공하는 토지특성 자료를 활용하였다. 노태욱·강창덕(2009)의 엔트로피 산출식을 참고하여 4개 토지이용(주거, 상업, 주상용, 공업용)으로 엔트로피 값을 식 (2)와 같이 구하였다.

$$-\sum_{i=1}^4 (p_i) \ln(p_i) / \ln(s) \quad (2)$$

여기서, p_i 는 주거, 상업, 주상용, 공업용 건물의 연면적을 모두 더하여 각 용도의 건물 연면적을 나눈 값이며, s 는 토지이용 용도의 개수로 본 연구에서는 4이다(노태욱·강창덕, 2009). 엔트로피 값이 0에 가까울수록 단일 토지이용이 높은 반면, 1에 근접할수록 다양한 토지이용이 균등한 면적으로 분포하고 있음을 의미한다(김태형·고준호, 2016). 거주지 및 직장지 행정동 중심에서 가장 가까운 지하철역까지의 거리, 직장지 행정동 중심에서 가장 가까운 서울 3도심까지의 거리는 GIS를 사용하여 2016년 도로망 기반으로 거리를 측정하였다. 서울 3도심은 2030 서울도시기본계획이 정의하는 여의도, 강남, 한양도성 3도심이다. 도심까지의 거리는 각 행정동 중심에서 서울 3도심의 주요 교통거점인 여의도역, 강남역, 광화문역까지의 거리로 측정하였다. 범주형 변수

는 모든 범주형 변수를 선택하지 않고 주요 범주형 변수만 채택하여 더미변수로 변환하였다.

독립변수 간 상관성이 지나치게 높은 변수는 최대한 제거하여 분석의 정확도를 높였다. 기계학습 기법은 다중공선성 문제를 다룰 수 있다는 데 큰 강점이 있지만, 각 독립변수와 종속변수와의 관계를 살펴볼 때에는 독립변수 간 상관성이 어느 정도 영향을 미칠 수 있기 때문이다. 분석에 사용된 변수 간 상관성은 <그림 1>과 같다.

3. 분석 방법

본 연구에서는 Python의 XGBoost 패키지를 이용하여 기계학습 기법의 하나인 Extreme Gradient Boost(XGBoost)를 활용하였다. 머신러닝을 활용할 경우, 전통적인 통계학으로 분석하기 힘든 특정 의도 없이 수집되거나 다양한 형태를 가진 자료를 분석할 수 있으며, 비선형을 보이는 변수 사이의 복잡한 상호작용을 파악할 수 있게 된다(유진은, 2017; 박준상·이수기, 2022). 비선형적 관계를 파악하는 것은 특정 변수가 영향력을 미치는 임계치를 파악하여 정책 도입을 보다 효과적으로 제한할 수 있다는 데 의의가 있다. 다만, 기계학습은 모델마다 설정되는 매개변수

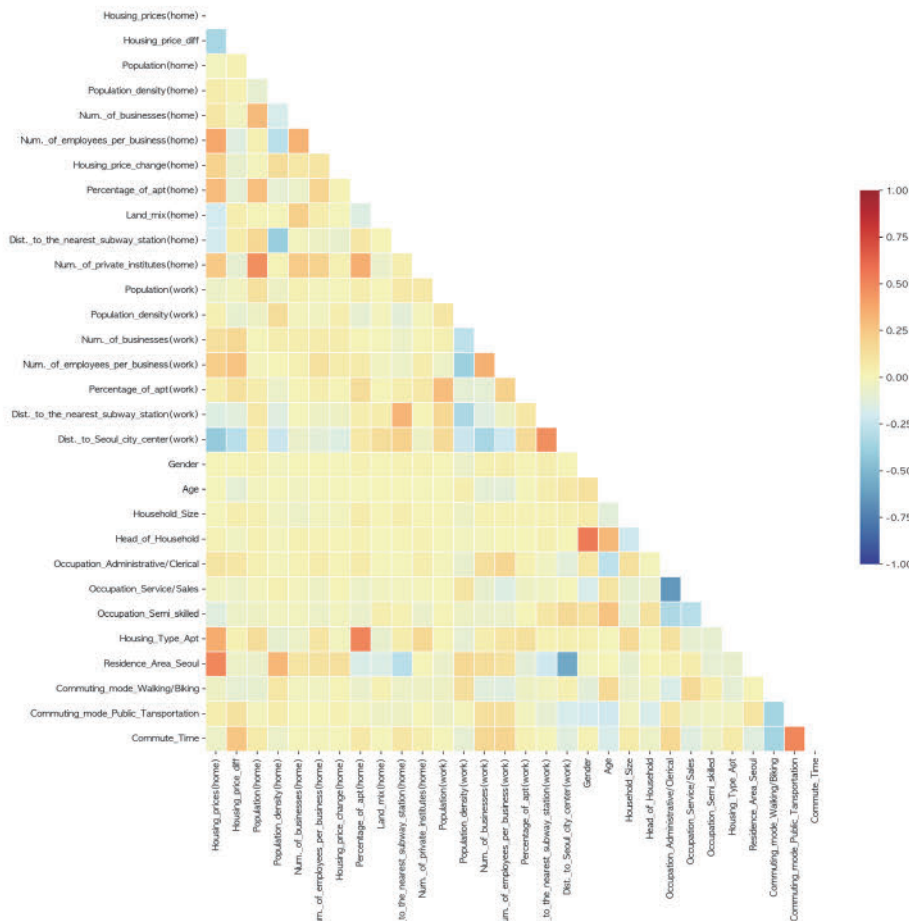


Figure 1. Correlation between variables

(hyper parameter)의 영향과 블랙박스(black-box)적인 특성으로 분석 과정에 대한 이해가 어려우며 이로 인해 재현성(Reproducibility)의 한계를 지닌다(권오익·김영일, 2023). 본 연구에서는 분석 과정을 이해하기 위해 분석에 사용한 매개변수를 공개하고, 변수의 기여도를 나타내는 Shapley Value와 Partial Dependence Plot(PDP)를 활용하여 분석 결과를 해석하였다.

본 연구에서는 선형 회귀모형, 기계학습 모형 중 Random Forest, Histogram-based Gradient Boost, Gradient Boost, CATBoost, XGBoost 모형의 평가 결과를 비교하였으며, 그중 가장 우수한 성능을 보인 XGBoost를 분석 모형으로 선정하였다. XGBoost는 Gradient Boosting Decision Tree(GBDT)를 효율적으로 구현하기 위한 시스템으로, 속도를 크게 향상시킬 수 있도록 분산 컴퓨팅을 사용하여 GBDT에 비해 일반적으로 10배 이상 처리 속도가 빠르다(Chen and Guestrin, 2016; Liu et al., 2021). 본 연구에서는 과적합(overfitting)을 방지하고, 모델의 정확도를 높이기 위해 Optuna 프레임워크를 사용하여 매개변수를 최적화하였다.

XGBoost를 활용한 분석 결과를 토대로, 주택가격 및 거주지 특성이 각 소득계층별 통근시간에 미치는 영향력을 파악하기 위해 변수별 상대적 기여도(Feature Importance)를 살펴보았다. GBDT 모델에서 의사결정트리를 구축하기 위해 독립변수는 반복적으로 무작위로 선택되며, 이때 특정 변수가 선택된 횟수에 따라 상대적 중요도를 측정한다(Breiman et al., 1984; Fried-

man, 2001). 따라서 상대적 기여도가 더 높다는 것은 변수가 성공적인 예측에 기여함을 의미한다(Zhang and Haghani, 2015). 본 연구에서는 Shapley Value를 통해 변수의 기여도를 파악하였으며, 소득계층별 분석모형의 Global Shapley Value를 분석하고, PDP를 통해 각 변수가 통근시간에 미치는 영향과 비선형적 관계를 분석하였다. PDP는 기계 학습 모델의 예측 결과에 대해 하나 혹은 두 개의 변수가 갖는 한계 효과를 보여준다(Friedman, 2001).

IV. 분석결과

1. 기술 통계

소득계층별 주요 변수의 기술 통계량은 <표 3>과 같다. 통근시간의 경우, 저소득층에서 고소득층으로 갈수록 평균 통근시간은 증가하는 것으로 나타났다. 거주지 주택가격 또한 고소득층으로 갈수록 높아지나, 직장지와 거주지의 주택가격 차의 평균은 중소득층에서 가장 크게, 저소득층에서 가장 작게 나타났다.

한편, 거주지 지역변수의 경우, 사업체수, 아파트 비율, 사설학원 수는 고소득층으로 갈수록 증가하고 있었다. 반면, 행정동 중심지로부터 가까운 지하철역까지의 거리는 고소득층으로 갈수록 줄어드는 것으로 나타났다. 응답자의 직장지 지역변수를 살펴보면 고소득층으로 갈수록 직장지의 사업체수 및 사업체당 종사자

Table 3. Descriptive statistics by income class

Category	Low income		Middle income		High income			
	mean, %	std.	mean, %	std.	mean, %	std.		
Observation	6,936		41,848		20,340			
Dependent variable	Commute time (min)		38.40	24.54	40.79	25.07	42.24	25.17
Housing price	Housing prices at home		1189.30	669.41	1419.39	732.73	1714.25	967.83
	Housing price difference between residential and workplace		160.38	597.73	241.72	761.26	177.13	923.47
Neighborhood characteristics at home	Population		24809.66	11032.41	26531.05	11363.94	27120.92	10982.79
	Population density		20522.68	12290.49	20895.54	12573.56	20409.05	12498.65
	Number of businesses		1659.24	1236.98	1758.66	1415.15	1791.45	1383.55
	Number of employees per business		4.61	2.21	4.89	2.44	5.38	2.89
	Housing price change (2011 vs. 2016 apartment price change)		-0.019	0.03	-0.017	0.04	-0.017	0.03
	Percentage of apartments (%)		55.51	29.82	61.18	29.23	66.02	29.02
	Land mix		0.55	0.15	0.54	0.15	0.54	0.15
Dist. from the residence to the nearest subway station (km)		1.47	1.62	1.44	1.52	1.41	1.39	
Number of private institutes		32.03	29.58	38.06	36.31	44.58	41.39	

다음 페이지에 계속(Continue on next page)

Category		Low income		Middle income		High income		
		mean, %	std.	mean, %	std.	mean, %	std.	
Observation		6,936		41,848		20,340		
Neighborhood characteristics at work	Total population	24889.49	11128.87	24982.71	11186.35	24881.82	11181.86	
	Population density	18495.44	12057.28	17643.48	11713.87	17078.10	11555.29	
	Number of businesses	2728.82	2729.10	3187.77	3216.27	3486.04	3487.41	
	Employees per business	5.79	3.27	6.50	3.80	7.14	4.20	
	Percentage of apartments (%)	56.69	28.91	57.70	28.70	59.51	28.76	
	Dist. from the workplace to the nearest subway station (km)	1.33	1.53	1.28	1.63	1.20	1.42	
	Dist. from the workplace to the nearest Seoul city center (km)	14.94	9.47	12.72	9.49	11.56	9.38	
Individual characteristics	Gender	Male	53.42%		66.05%		60.15%	
		Female	46.58%		33.95%		39.85%	
	Age	50.83	15.02	44.26	11.59	44.22	11.71	
	Household size	2.06	0.99	3.01	1.02	3.58	0.89	
	Relationship	Head of household	72.23%		61.81%		47.61%	
		Spouse	12.46%		19.12%		24.26%	
		Child	12.85%		16.77%		25.15%	
		Other	1.90%		1.48%		1.62%	
	Occupation	Professionals	3.46%		5.96%		10.16%	
		Service and sales	45.63%		36.96%		31.69%	
		Administrative and clerical	17.89%		42.56%		49.50%	
		Semi-skilled	29.84%		12.72%		6.91%	
		Low-skilled	0.62%		0.40%		0.32%	
		Other	2.55%		1.39%		1.41%	
	Housing types	Apartments	45.00%		61.97%		75.92%	
		Multi-units	34.41%		26.44%		16.07%	
		Officetel	2.02%		1.27%		0.74%	
		Single-family	18.57%		10.32%		7.27%	
	Residence areas	Seoul	41.48%		50.25%		51.60%	
		Gyeonggi	28.62%		28.53%		31.49%	
Incheon		29.90%		21.22%		16.90%		
Commuting mode	Walking and biking	22.91%		12.15%		9.67%		
	Car and motorcycle	23.62%		40.73%		43.16%		
	Public transportation	51.77%		45.88%		46.23%		
	Taxi	0.69%		0.49%		0.51%		
	Other	1.01%		0.75%		0.44%		

수는 증가했다. 반면, 직장지 행정동 중심지로부터 가까운 지하철역 및 서울 3도심까지의 거리는 고소득층으로 갈수록 감소했다.

개인 특성 변수의 경우, 저소득층으로 갈수록 서비스 및 판매 종사자, 기능원 및 단순노무종사자, 농림어업 및 숙련종사자의 비율은 증가하고 있었다. 이와 반대로 전문가 및 관련종사자, 관

리자 및 사무종사자의 비율은 고소득층으로 갈수록 증가하는 것으로 나타났다. 주택유형의 경우, 고소득층으로 갈수록 아파트 거주비율은 현저히 증가하는 반면, 연립, 다세대, 다가구주택 거주비율은 감소하고 있었다.

통근수단의 경우, 저소득층은 대중교통 이용비율이 가장 높았

고, 다음으로 승용차 및 오토바이, 도보 및 자전거 순으로 이용비율이 낮아지고 있었으나, 고소득층으로 갈수록 승용차 및 오토바이 이용비율이 현저히 증가한 반면, 도보 및 자전거 이용률은 눈에 띄게 감소했다. 연구 범위 내 수도권에 거주하는 저소득층, 중소득층, 고소득층 모두 대중교통 이용 비율이 가장 높았다.

2. 기계학습 모형

소득계층별 분석에서는 저소득층 6,936건, 중소득층 41,848건, 고소득층 20,340건의 응답자료를 활용하였으며, 각 계층의 데이터를 학습 데이터와 평가 데이터로 분류하였다. 학습 표본은 소득계층별 표본 수의 70%, 평가 표본은 30%로 할당해 분할하였다. 기계학습 모형 중 Random Forest, Histogram-based Gradient Boost, Gradient Boost, CATBoost, XGBoost 모형의 평가 결과를 비교한 결과는 <표 4>와 같다. 분석모형의 R-Squared, 표준절대오차(Mean Absolute Error), 표준제곱근오차(RMSE)를 통해 비교 평가하였으며, 그중 모든 소득계층에서 가장 우수한 성능을 보인 XGBoost를 분석 모형으로 선정하였다. 동일한 변수를 사용하여 전통적으로 선형 관계를 가정하는 회귀분석 결과와 비교해보면, 변수 간 비선형 관계를 파악할 수 있는 기계학습 기법 중 하나인 XGBoost의 설명력이 매우 높은 것을 확인할 수 있다.

본 연구에서는 과적합을 방지하고 모델의 정확도를 높이기 위

해 Optuna 프레임워크를 활용하여 XGBoost의 매개변수(hyper-parameter)를 조정하였다. Optuna 프레임워크는 매개변수를 최적화하는 방법으로, 반복되는 최적화 과정에서 반복적으로 학습하여 최적의 매개변수를 도출한다(Srinivas and Katarya, 2022). Optuna 프레임워크를 통해 조정된 소득계층별 모형의 주요 매개변수는 <표 5>와 같다. max_depth는 의사 결정 트리의 최대 깊이, learning_rate는 학습률, colsample_bytree와 subsample은 트리가 과도하게 복잡하게 생성되지 않도록 과적합을 방지하기 위한 매개변수, alpha와 lambda는 과적합을 방지하기 위한 정규화 매개변수, min_child_weight는 의사 결정 트리에서 가지를 추가할지를 결정하는 데 필요한 가중치의 총합을 의미한다.

3. 소득계층별 변수의 기여도 비교

변수별 통근시간에 미치는 상대적 기여도(Feature Importance)를 보기 위해 Shapley Value를 도출하였다. Local Shapley Value에 따라 소득계층별 주요 변수의 기여도 순위를 소득계층별로 비교한 결과는 <표 6>, Local Shapley Value Summary Plot은 <그림 2>와 같다. Local Shapley Value는 모든 데이터를 점으로 표현하여 해당 변수의 영향을 보여주는 그래프이다. 점의 색이 진할수록 해당 변수의 값이 큰 데이터를 의미하며 0을 기준으로 좌측 영역은 종속변수와 음의 관계, 우측 영역은 종속변수

Table 4. Model evaluation result

Model	Low income			Middle income			High income		
	R ²	RMSE	MAE	R ²	RMSE	MAE	R ²	RMSE	MAE
Linear Regression	0.401	18.333	13.286	0.394	19.379	14.079	0.381	20.271	14.684
Random Forest	0.495	17.473	12.156	0.535	16.983	11.900	0.512	17.445	12.480
Histogram-based Gradient Boost	0.502	17.348	12.227	0.521	17.243	12.274	0.507	18.092	12.664
Gradient Boost	0.478	17.767	12.566	0.491	17.765	12.743	0.488	18.429	13.049
CATBoost	0.512	17.183	11.996	0.542	16.862	11.886	0.511	17.455	12.356
XGBoost	0.524	16.466	11.838	0.551	16.956	11.677	0.524	16.941	12.248

Table 5. Hyper-parameter used in XGBoost model

Hyper-parameter	Low income	Middle income	High income
max_depth	12	11	8
learning_rate	0.0482	0.0268	0.0069
colsample_bytree	0.2578	0.2152	0.2095
subsample	0.4278	0.6820	0.6625
alpha	0.0546	0.3102	0.7028
lambda	0.0165	0.0799	1.3323
min_child_weight	24.8235	16.8640	175.9004

Table 6. Top 10 feature importance by income class

Rank	Low income	Middle income	High income
1	Commuting mode - Public Transportation	Commuting mode - Public Transportation	Commuting mode - Public Transportation
2	Housing price differences between residential and workplace	Housing price differences between residential and workplace	Residence area - Seoul
3	Commuting mode - Walking/Biking	Residence area - Seoul	Housing price differences between residential and workplace
4	Residence area - Seoul	Commuting mode - Walking/Biking	Distance from the workplace to Seoul city center
5	Distance from the workplace to Seoul city center	Distance from the workplace to Seoul city center	Housing price at home
6	Housing price at home	Housing price at home	Commuting mode - Walking/Biking
7	Distance from home to the nearest subway station	Number of businesses in the neighborhood at work	Gender
8	Number of employees per business in the neighborhood at work	Distance from home to the nearest subway station	Number of employees per business in the neighborhood at work
9	Population density in the neighborhood at work	Number of employees per business in the neighborhood at work	Number of private institutes in the neighborhood at home
10	Number of businesses in the neighborhood at work	Percentage of apartments in the neighborhood at home	Number of employees per business in the neighborhood at home

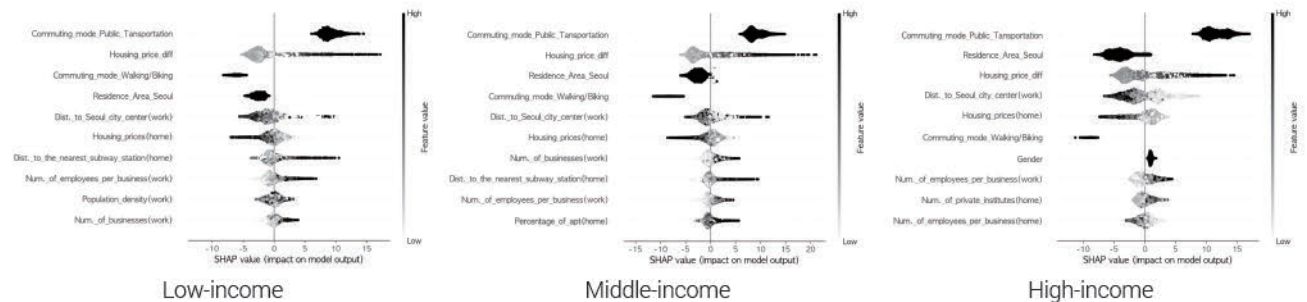


Figure 2. Local shapley value summary plot of top 10 features on the commuting time

와 양의 관계를 지니고 있음을 나타낸다.

〈그림 2〉에서 통근시간을 설명하는 상위 5위까지의 주요 변수는 통근수단과 거주지 및 직장지의 지역 특성과 관련이 있는 것으로 나타났다. 우선, 모든 소득계층에서 대중교통 이용 여부가 통근시간을 가장 잘 설명하는 변수로 나타났으며, 대중교통을 이용하는 경우 통근시간이 길게 나타나는 경향을 보인다. 반면, 도보 및 자전거를 이용할 경우 통근시간이 짧아지는 경향이 있다. 이는 일반적으로 근거리 이동의 경우 도보 및 자전거를 선호하는 경향으로 인한 결과로 해석될 수 있다. 한편, 모든 소득계층에서 거주지와 직장지 간 주택가격의 차이, 서울 거주 여부와 직장지로부터 서울 3도심까지의 거리가 통근시간에 매우 중요한 요인으로 나타났다. 한편, 서울 거주 변수가 통근시간과 음의 관계가 있어 경기도 및 인천 거주자의 통근 시간이 서울 거주자에 비해서 길게 나타났다. 서울 거주 여부는 고소득층으로 갈수록 통근시간 결정에 미치는 설명력이 크고, 도보/자전거 사용의 설명력은 고소득층으로 갈수록 낮게 나타났다. 거주지의 주택가격은 저, 중소득

층에서는 6번째, 고소득층에서는 5번째 주요 변수로 도출되어 상대적으로 고소득층에서 높게 나타났다. 한편, 거주지의 대중교통 접근성 변수는 소득계층별로 기여도가 다르게 도출되었다. 거주지에서 가장 가까운 지하철역까지의 거리는 소득이 낮아질수록 중요한 변수로 나타났다. 반면, 거주지의 아파트 비율은 중소득층에서 유일하게 10번째, 거주지의 사설학원 수는 고소득층에서 유일하게 9번째로 변수의 기여도 상위 10개 변수에 포함되었다.

종합하면, 전반적으로 대중교통 이용 여부, 직장지와 거주지의 주택가격 차, 서울 거주 여부, 직장지에서 가장 가까운 서울 3도심까지의 거리, 거주지의 주택가격, 직장지의 사업체당 종사자수가 통근시간을 설명하는 주요한 변수로 도출되었다. 저소득층의 경우 대중교통 접근성을 나타내는 가장 가까운 지하철역까지의 거리가 다른 소득층에 비해 주요한 설명 요인으로 도출되었다. 성별과 거주지의 사설학원 수는 다른 소득계층에 비해서 고소득층의 통근시간의 주요한 설명변수로 나타났다. 한편, 대중교통, 도보 또는 자전거 이용, 서울 거주 여부는 통근 시간에 미치는

영향 방향이 비교적 일관되게 나타났다. 즉, 대중교통 이용자는 대체로 통근시간이 길게 나타난 반면 도보 또는 자전거 이용, 서울 거주자는 인천이나 경기 거주자에 비하여 통근시간이 짧게 나타났다. 대부분의 주요 변수는 <그림 2>에서 보이는 바와 같이 영향 관계가 일관된 방향성을 보이지 않아 주요 변수에 대한 통근시간 의존도를 다음에서 상세하게 볼 필요가 있다.

4. 주요 변수의 통근시간 의존도 분석

본 연구에서는 변수의 기여도가 높게 나온 요인을 중심으로 소득계층별 분석 모형의 PDP를 살펴보았다. 모든 소득계층에서 가장 중요한 변수로 도출된 직장지와 거주지의 주택가격 차이, 직장지에서 가장 가까운 서울 3도심까지의 거리에 대해서 각 영향요인이 소득계층별 통근시간에 미치는 비선형적 관계를 분석하였다. PDP의 x축은 해당 변수, y축은 종속변수인 통근시간을 의미한다.

우선, 통근시간을 설명하는 연속형 변수 중 변수의 기여도가 가장 높게 나타난 직장지와 거주지의 주택가격 차의 PDP는 <그림 3>에서 보이는 바와 같이 V자 형태로 나타나고 있다. 직장지와 거주지의 주택가격 차는 직장지의 평당 주택가격에서 거주지의 평당 주택가격을 뺀 값으로, 음의 값은 거주지가 직장지의 주택가격보다 높고 양의 값은 거주지가 직장지의 주택가격보다 저렴하다는 것을 의미한다. 소득계층에 따라 V자 형태는 조금 상이하다. 첫째, x축의 음의 영역에 해당하는 범위(range)가 저소득층이 가장 짧고 고소득층으로 갈수록 길어진다. 이는 고소득층으로 갈수록 직장지보다 거주지의 주택가격이 약 평당 1,500만원까지 높은 경우까지도 존재하는 반면, 저소득층의 경우에는 직장지의 주택가격이 주거지보다 높은 경우 그 차이가 평당 약 500만원 이내인 것으로 나타났다. 또 전반적으로 중, 고소득층으로 갈수록 x축의 범위가 넓어져 주거지와 직장지의 주택가격 차이의 분포 범위가 넓음을 알 수 있다. 둘째, 그림에서 보는 바와 같이 직장지의 주택가격이 더 높은 경우(x축 양의 값 부분), 직장지와 거주지의 주택가격 차가 변함에 따라 통근시간의 변화는 대략적인 기울기

를 고려할 때 저소득층 > 중소득층 > 고소득층 순으로 나타났다. 또한, 저소득층의 경우 상대적으로 분명하게 식별되는 변곡점 없이 주거지에 비해 직장지의 주택가격 차이가 커질수록 통근시간 증가 정도가 크다. 고소득층의 경우 다른 소득계층에 비해 넓은 V자 형태로 나타났으나, 주거지의 주택가격이 직장지보다 높은 경우에는 직장지의 주택가격이 높은 경우보다는 비교적 완만하게 통근시간이 변화하는 것으로 나타났다.

직장지와 거주지의 주택가격 차이가 통근시간에 미치는 영향은 실제 직장지와 거주지의 주택가격 분포와 연결시켜 이해할 필요가 있다. 소득계층별 직장지와 거주지의 주택가격 차 분포를 나타낸 <그림 4>를 보면, 저소득층의 경우 직장지와 거주지의 주택가격 차이가 크지 않은 경우가 대다수로 나타났다. 중, 고소득층 또한 <그림 4>에서 보는 바와 같이 직장지와 거주지의 주택가격 차가 0에 가장 가까운 구간의 비율이 높지만, 상대적으로 저소득층에 비해서 낮은 편이다. 또한, 고소득층은 다른 소득계층에 비해 직장지와 거주지의 주택가격 차 분포가 상대적으로 넓게 분포하고 있다. 이는 저소득층의 경우 주거지 및 직장지가 다른 계층에 비해서 상대적으로 제한되어 있어 직장지와 거주지의 주택가격 차이 분포 범위가 넓지 않으며, 주거지와 직장지의 주택가격이 비교적 유사한 곳 또는 인접한 지역 일자리에 종사하는 경향

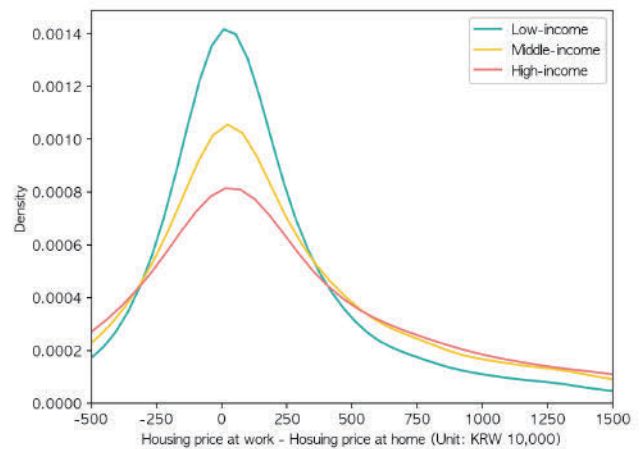


Figure 4. Distribution of housing price differences by income class

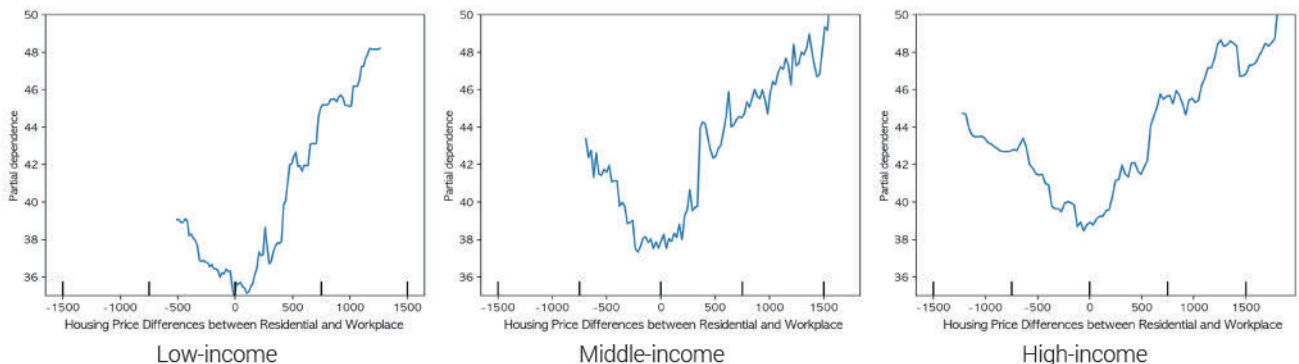


Figure 3. PDP of housing price differences between residential and workplace by income class

이 높은 것으로 나타났다. 한편, 고소득층으로 갈수록 거주지 주택가격이 직장지 주택가격에 비해서 높은 곳에 거주하는 비율도 높아지지만, 동시에 직장지 주택가격이 거주지 주택가격에 비해서 높아지는 경우도 증가하고 있다. 이는 저소득층의 직장지와 주거지의 불일치(mismatch) 가설과는 달리 고소득층일수록 직장지 지대가 높은 고용중심지에 가까이 위치하기 때문에, 통근시간의 감소보다 임대료의 상승이 더욱 가파르기(Park and Quercia, 2015) 때문이라고 해석가능하다.

반면, 직장지가 거주지보다 저렴한 경우는 주거비용 외에도 지역의 교육(Kim et al., 2005), 안전(Weisbrod et al., 1980), 도시 어메니티(Rapoport, 1980; Tyrvaenen and Vaananen, 1998; Rouwendal and Meijer, 2001; Parkes et al., 2002) 등 거주지의 환경적 질을 우선적으로 선택한 것으로 이해될 수 있다. 주거 선택에 있어서 고용 접근성 이외에 다양한 요인들이 고려된 것으로, 이러한 경향은 고소득층으로 갈수록 강하게 나타나고 있다. 이는 가구의 주거입지 선택이 고용 접근성에 대한 고려와 삶의 질에 대한 고려의 상쇄관계로 설명된다는(Phe and Wakely, 2000; Kim et al., 2005; Moos and Skaburskis, 2010; 김승남·안건혁, 2011) 접근성-공간 맞교환 이론(Access-Space Trade-off)과도 같은 맥락이며, 특히 본 연구에서는 이러한 경향이 고소득층으로 갈수록 심화된다는 것이 확인되었다.

본 연구에서 통근시간에 영향을 미치는 중요한 변수 중 하나인 직장지로부터 가장 가까운 서울 3도심까지의 거리 변수는 <그림 5>와 같이 직장지가 도심에서 멀어질수록 통근시간은 감소하는 것으로 나타났다. 특히, 세 소득계층 모두 직장지가 도심에 가까울 때 통근시간이 가장 긴 것으로 나타났다. 고소득층의 경우 다른 소득계층과는 달리 직장지에서 서울 3도심까지의 거리 약 5km 반경에서 통근시간이 비교적 일정하게 높게 나타나는 것으로 나타났다. 이와 같은 결과는 다른 소득계층에 비해서 고소득층의 경우 비교적 도심 중심부 인접지역에 근무하는 비율이 높으며, 도심부에 직장지가 있는 경우 통근시간이 매우 높음을 나타낸다. 한편, 모든 계층에서 직장지가 가장 가까운 도심에서 약 16~19km 떨어져 있는 경우 통근시간이 다소 증가하는 경향이

나타났다. 이는 해당 구간의 통행 중 약 65%가 경기도 지역 내 통근에 해당하는데 상대적으로 경기도의 경우 내부통행의 통근시간이 높게 나타나면서 다소 증가하는 형태가 도출되었다.

V. 결론

본 연구에서는 2016년 가구통행실태조사 자료를 활용해 수도권을 대상으로 소득계층별 통근시간 영향요인을 비교·분석하였다. 주택가격 등 주거비용을 나타내는 변수, 거주지 및 직장지 특성 변수, 응답자 개인 특성 변수를 분석에 활용하였다. 본 연구에서는 분석의 설명력을 높이고자 기계학습 기법 중 하나인 XGBoost를 활용했으며, 분석을 통해 소득계층별로 통근시간을 설명하는 변수를 식별하였다. 주요 연구 결과 및 정책적 시사점은 다음과 같다.

첫째, 서울시 교통권역에 해당하는 수도권 내 거주자의 통근시간은 고소득층으로 갈수록 길게 나타나, 통근에서 취약계층, 즉 저소득층의 주거-직장 공간 불일치(spatial mismatch) 정도가 다른 계층에 비해 심화되는 경향은 확인되지 않았다. 또한, 저소득층의 경우 다른 계층에 비하여 거주지와 직장지 간 평균 주택가격 차이가 가장 낮게 나타났는데, 이는 기존 선행연구에서 저소득층의 경우 주거환경보다 직장지로의 접근성에 대한 선호도가 중요(Zhao, 2015)한 결과로 나타난 것과 일치하는 결과이다. 따라서 저소득층의 경우 주거입지 선택의 주요한 결정요인이 직장지로의 접근성이며 이로 인해 평균적인 통근시간이 다른 계층에 비해서 상대적으로 짧게 나타나는 것으로 이해될 수 있다.

둘째, 각 소득계층의 요인별 통근시간 설명력을 비교한 결과, 응답자의 대중교통 이용 여부 및 서울 거주 여부, 주택가격, 직장지의 업무중심지적 특성이 모든 소득계층의 통근시간을 설명하는 주요한 요인으로 나타났다. 서울 거주 여부와 거주지의 주택가격의 경우 고소득층에서 그 중요도가 가장 높았다. 이처럼 주택가격이 고소득층의 통근시간에 더욱 민감한 이유는 고소득층으로 갈수록 직장지가 서울 3도심에 가까워지면서 통근시간의 감소보다 임대료의 상승이 더욱 가파르기(Park and Quercia,

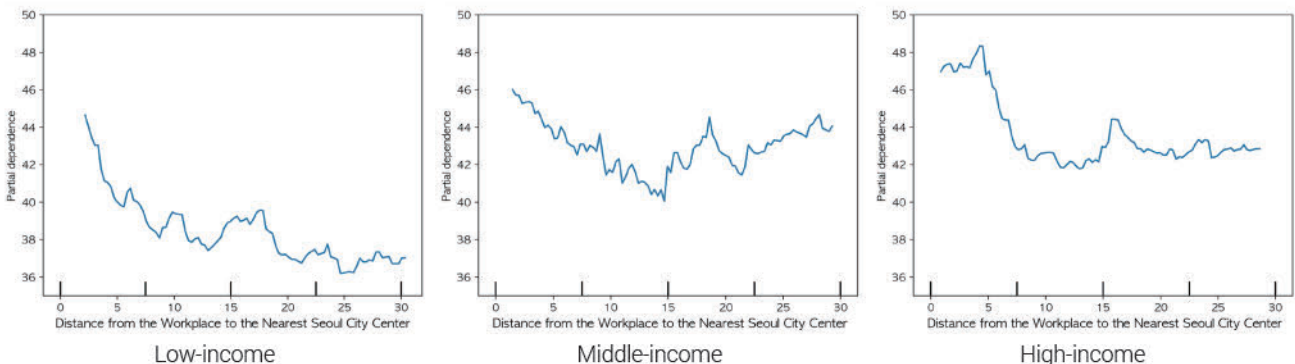


Figure 5. PDP of the distance from the workplace to the nearest Seoul city center by income class

2015) 때문으로 보인다. 한편, 소득계층별로 중요도 차이가 있는 주요한 요인은 대중교통 접근성의 설명력이다. 저소득층의 경우 거주지의 대중교통 접근성이 중요한 요인으로 나타난 반면, 고소득층은 저소득층에 비해 양호한 거주지 환경을 나타내는 거주지의 사설학원 수가 통근시간 결정에 중요하게 작용하고 있다.

셋째, 소득계층별 통근시간과 직장지와 거주지의 주택가격 차간의 관계는 V자 형태로 나타났다. 이는 직장지와 거주지의 평당 주택가격이 유사한 곳의 경우 통근시간이 매우 짧아지나 차이가 커질수록 통근시간이 증가함을 의미하는데, 직장지와 거주지의 주택가격 차가 변함에 따라 통근시간의 변화는 고소득층보다 저소득층에서 더욱 민감하게 증가하는 것으로 나타났다. 한편, 고소득층의 경우 거주지의 주택가격이 직장지보다 높은 구간이 상대적으로 넓게 나타났는데 이는 고용 접근성과 함께 주거환경 및 삶의 질(Phe and Wakely, 2000; Kim et al., 2005; Moos and Skaburskis, 2010; 김승남·안건혁, 2011)을 고려한 것으로 이해될 수 있다. 또한, 모든 소득계층에서 직장지가 업무중심지의 성격을 지닐수록 통근시간이 증가하는 것으로 나타났다.

연구 결과의 주요한 정책적 시사점은 다음과 같다. 우선, 저소득층의 경우 직장지에 비해서 거주지의 주거 가격이 높은 구간이 매우 짧은 편으로 이는 저소득층의 경우 예산의 제약으로 거주지 선택의 폭이 다른 계층에 비해서 좁다는 것을 의미한다. 또한 주택가격의 차가 커졌을 때 통근 시간의 편차가 다른 계층에 비해서 가장 크게 나타나고 있다. 이와 같이 저소득층의 경우 주거환경에 대한 선호보다 직장에 대한 근접성이 중요한 주거 선택 요인이라는 점을 고려할 때 저소득층을 위한 주거 복지 정책 및 주거 입지 정책 고려 시 주거환경에 대한 요인뿐만 아니라 일자리로의 접근성이 중요하게 고려되어야 함을 시사하고 있다. 특히 주택가격이 높은 주거 지역일 경우 적극적으로 다양한 소셜믹스를 실현할 수 있는 도시계획 인센티브, 임대주택 구성 방식 등을 도입할 필요가 있다. 둘째, 모든 소득계층에서 직장지가 도심에 가까울수록 통행시간은 급격하게 증가하는 것으로 나타나 서울 3도심 및 주요 고용 중심지에 주택 공급이 중요함을 시사한다. 직장지가 도심에 가까울수록 통행시간이 증가하는 것은 도심의 높은 주택가격으로 인한 주거입지 제약과 일자리 대비 주거가 부족한 직주비로 인한 공간적 제약으로 인한 영향으로 이해될 수 있다. 실제로 서울 도심은 전통적으로 상업과 업무의 중심지 역할을 수행하나 직주비율은 해외 대도시와 비교했을 때 불균형이 매우 심각한 형편이다(서울시, 2023). 직주비를 거주취업자수 대비 일자리수로 보면, 서울 종로·중구의 경우 4.93, 강남 2.41 등으로 파리 0.76, 런던 6구 1.38, 뉴욕맨해튼 1.41, 도쿄 11구 2.23 등으로 주거 비율이 매우 낮다(서울시, 2023). 따라서, 수도권 도심 및 업무 중심지에 다양한 주거 비용과 유형의 주거를 공급하는 직주연계형 주거 공급 정책이 적극적으로 추진될 필요가 있다.

한편, 본 연구는 다음과 같은 점에서 한계를 지닌다. 첫째, 본

연구에서 사용한 가구통행실태조사의 소득 자료의 경우 범주형으로 구성되어 있기 때문에 소득계층을 보다 세밀하게 구분하는데 한계가 있다. 또한, 분석 자료의 소득 구분은 가구 평균 월 소득을 기준으로 하여, 자산 등 정기적으로 발생하지 않는 소득을 계층 구분에 포함하지 못했다. 이와 같은 이유로 평균 월 소득은 적으나 자산을 많이 보유하고 있는 일부 계층이 과소평가되었을 수 있다. 둘째, 가구통행실태조사 자료에 기입된 응답자의 정보는 행정동 단위로 거주지 및 직장지가 표기되어 있다. 따라서 주택유형별, 행정동별 주택가격을 구분하였음에도 불구하고 개별 응답자의 주택가격 및 가장 가까운 지하철역까지의 거리와 같은 지역변수가 보다 일반화되어 적용되었을 수 있다. 셋째, 본 연구의 기계학습 모형은 변수 간의 인과관계를 도출하기에는 한계가 있다. 이와 같은 한계에도 불구하고 본 연구에서는 기존 연구의 한계를 보완하기 위하여 직장지와 거주지의 특성을 세밀하게 분석하고, 이에 소득계층별 통근시간 영향요인 차이를 기계학습 방법을 활용하여 비교 분석했다는 데에 의의가 있다.

인용문헌 References

- 권오익·김영일, 2023. “건물에너지 예측을 위한 기계학습 모델 검토”, 『대한건축학회논문집』, 39(5): 133-140.
Kwon, O.I. and Kim, Y.I., 2023. “Review of Machine Learning for Building Energy Prediction”, *Journal of the Architectural Institute of Korea*, 39(5): 133-140.
- 김승남·안건혁, 2011. “주택근무와 주거입지의 관계에 대한 실증 연구: 수도권 거주 임금근로자를 중심으로”, 『국토계획』, 46(7): 37-55.
Kim, S.N. and Ahn, K.H., 2011. “The Relationship between Home-based Telecommuting and Residential Location: Focused on the Salaried Workers in the Seoul Metropolitan Area”, *Journal of Korea Planning Association*, 46(7): 37-55.
- 김태형·고준호, 2016. “대도시 토지이용 압축도 지표의 개발 및 적용: 서울시를 대상으로”, 『서울도시연구』, 17(1): 1-21.
Gim, T.H. and Ko, J.H., 2016. “Developing a Land Use Compactness Index for a Large City: A Case of Seoul, Korea”, *Seoul Studies*, 17(1): 1-21.
- 김희철·안건혁, 2011. “압축도시 계획요소가 소득계층별 통근 거리에 미치는 영향”, 『한국도시계획학회지 도시설계』, 12(1): 55-70.
Kim, H.C. and Ahn, K.H., 2011. “The Effects of Compact City Planning Strategies on Commuting Distance of Different Income Levels: Focused on Seoul, Korea”, *Journal of the Urban Design Institute of Korea Urban Design*, 12(1): 55-70.
- 노태욱·강창덕, 2009. “도시환경이 주거용 토지가격에 미치는 영향에 관한 연구: 서울시 강북지역을 중심으로”, 『부동산학연구』, 15(1): 81-101.
Rho, T.U. and Kang C.D., 2009. “The Impacts of Urban Envi-

- ronment on the Price of Residential Land”, *Journal of the Korea Real Estate Analysts Association*, 15(1): 81-101.
6. 도시교통정비 촉진법, “법률 제17975호, 2022년 3월 24일 시행”, Urban Traffic Improvement Promotion Act, “Act No. 17975, Mar. 24, 2022”.
 7. 박미선·김호정·강미나·김동형·권기현, 2018. 「교통비용을 고려한 주거부담 지표의 개발 및 활용방안 연구», 국토연구원. Park, M.S., Kim, H.J., Kang, M.N., Kim, D.H., and Kwon, K.H., 2018. *Calculation and Application of Location Housing Affordability Index*, Korea Research Institute for Human Settlements.
 8. 박준상·이수기, 2022. “해석가능한 기계학습을 활용한 보행목적별 보행만족도 영향요인 분석”, 「국토계획」, 57(1): 26-41. Park, J. and Lee, S., 2022. “Analysis of Influencing Factors of Walking Satisfaction by Purpose Using Interpretable Machine Learning”, *Journal of Korea Planning Association*, 57(1): 26-41.
 9. 서울시, 2023. 「서울도시 기본계획」, 서울시. Seoul Metropolitan Government, 2023. *Seoul Comprehensive Plan*, Seoul Metropolitan Government.
 10. 송재민, 2021. “도시형태가 통행행태에 미치는 영향 메타분석”, 「국토계획」, 56(7): 103-114. Song, J.M., 2021. “The Impact of Urban Form on Travel in Korea: A Meta Analysis”, *Journal of Korea Planning Association*, 56(7): 103-114.
 11. 오지에·이영호·정예원·홍성연, 2019. “가구통행실태조사 자료를 활용한 소득계층별 공간적 분리 탐색”, 「교통연구」, 26(1): 17-29. Oh, J.Y., Lee, Y.H., Jeong, Y.W., and Hong, S.Y., 2019. “An Analysis of the National Household Travel Survey Data for Exploring Spatial Segregation of Income Groups”, *Journal of Transport Research*, 26(1): 17-29.
 12. 유진은, 2017. “기계학습을 통한 TIMSS 2011 중학생의 수학 성취도 관련 변수 탐색”, 「교원교육」, 33(1): 43-56. Yoo, J.E., 2017. “TIMSS 2011 Predictors Relating to Korean 8th Graders’ Mathematics Achievement, Explored Via Machine Learning”, *Korean Journal of Teacher Education*, 33(1): 43-56.
 13. 이민주·박인권, 2016. “지역 특성에 따른 소득별 직주불일치에 관한 연구”, 「지역연구」, 32(1): 67-82. Lee, M.J. and Park, I.K., 2016. “A Study on the Spatial Mismatch by Income and Regional Characteristics”, *Journal of the Korean Regional Science Association*, 32(1): 67-82.
 14. 이변송, 1998. “서울 거주자의 통근거리 결정요인 분석”, 「국토계획」, 33(3): 241-263. Lee, B.S., 1998. “Determinants of Commuting Distance for Seoul Residents”, *Journal of Korea Planning Association*, 33(3): 241-263.
 15. 이해승·이희연, 2009. “서울시 대중교통체계 개편 이후 통근 교통수단 선택의 차별적 변화”, 「대한지리학회지」, 44(3): 323-338. Lee, H.S. and Lee, H.Y., 2009. “Differential Changes in Commuter’s Mode Choice after the Intergrated Public Transit System in Seoul Metropolitan City”, *Journal of the Korean Geographical Society*, 44(3): 323-338.
 16. 장재민·김태형, 2016. “통근시간 변화에 영향을 미치는 개인 및 지역변수 분석: 서울시 25개 자치구를 중심으로 (2006~2014)”, 「서울도시연구」, 17(3): 105-116. Jang, J.M. and Gim, T.H., 2016. “An Analysis of Individual and Regional Characteristics in Relation to Commuting Time Changes: Focusing on the 25 Districts of Seoul (2006~2014)”, *Seoul Studies*, 17(3): 105-116.
 17. 장재민·이병호·고준호, 2019. “근로자의 통근시간 만족도 결정요인 연구: 경기도 거주자를 중심으로”, 「대한교통학회지」, 37(4): 290-301. Jang, J.M., Lee, B.H., and Ko, J.H., 2019. “Determinants of Commute Time Satisfaction: Focusing on the Residents of Gyeonggi Province”, *Journal of Korean Society of Transportation*, 37(4): 290-301.
 18. 전명진·안현주, 2016. “수도권 통근자의 주거비용과 통근비용의 상쇄관계에 대한 실증분석”, 「한국지역개발학회지」, 28(2): 25-40. Jun, M.J. and Ahn, H.J., 2016. “An Empirical Test of Trade-off Between Housing Rents and Commuting Costs for The Commuters in Seoul”, *Journal of The Korean Regional Development Association*, 28(2): 25-40.
 19. 전명진·이지현, 2020. “수도권 통근자의 직장 및 주택접근성이 통근시간에 미치는 영향에 대한 분석”, 「한국지역개발학회지」, 32(2): 119-135. Jun, M.J. and Lee, J.H., 2020. “The Effects of Commuter’s Job and Housing Accessibility on Commuting Time in the Seoul Metropolitan Area”, *Journal of The Korean Regional Development Association*, 32(2): 119-135.
 20. 전희정·정수영, 2021. “근린의 경제적 변화에 대한 영향요인과 공간적 상호의존성: 읍면동 수준의 주택가격 변화를 중심으로”, 「국토계획」, 56(5): 19-29. Jun, H.J. and Jung, S.Y., 2021. “Determinants of Neighborhood Economic Change and Spatial Dependence: An Analysis of Housing Price Change at the Neighborhood Level”, *Journal of Korea Planning Association*, 56(5): 19-29.
 21. 최준성·노승환, 2021. “수도권 신도시의 통근비용을 고려한 주거비용 분석에 관한 연구”, 「주택연구」, 29(3): 93-116. Choi, J.S. and Ro, S.H., 2021. “A Study on the Housing Cost Analysis Considering Commuting Cost of New Towns in Seoul Metropolitan Area”, *Housing Studies Review*, 29(3): 93-116.
 22. 하재현, 2020. “도시공간구조가 소득수준별 통근거리에 미치는 영향에 관한 연구: 미국 대도시권 지역의 LEHD 자료를 중심으로”, 박사학위논문 한양대학교 대학원. Ha, J.H., 2020. “A Study on the Effects of Urban Spatial Structure on Commuting Distances by Income Level: Evidence from the LEHD of US Metropolitan Areas”, Doctoral Dissertation, Hanyang University.
 23. Adair, A., McGreal, S., Smyth, A., Cooper, J., and Ryley, T., 2000. “House Prices and Accessibility: The Testing of Relationships within the Belfast Urban Area”, *Housing Studies*, 15(5): 699-716.
 24. Alonso, W., 1964. *Location and Land Use: Toward a General Theory of Land Rent*, Harvard University Press.
 25. Blumenberg, E. and King, H., 2021. “Jobs-housing Balance

- Re-re-visited”, *Journal of the American Planning Association*, 87(4): 484-496.
26. Breiman, L., Friedman, J., Olshen, R., and Stone, C.J., 1984. *Classification and Regression Trees*, CRC Press.
 27. Chen, T. and Guestrin, C., 2016. “Xgboost: A Scalable Tree Boosting System”, *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785-794.
 28. Ding, C., Cao, X., and Næss, P., 2018a. “Applying Gradient Boosting Decision Trees to Examine Non-linear Effects of the Built Environment on Driving Distance in Oslo”, *Transportation Research Part A: Policy and Practice*, 110: 107-117.
 29. Ding, C., Cao, X., and Wang, Y., 2018b. “Synergistic Effects of the Built Environment and Commuting Programs on Commute Mode Choice”, *Transportation Research Part A: Policy and Practice*, 118: 104-118.
 30. Friedman, J.H., 2001. “Greedy Function Approximation: A Gradient Boosting Machine”, *Annals of Statistics*, 1189-1232.
 31. Haas, A. and Osland, L., 2014. “Commuting, Migration, Housing and Labour Markets: Complex Interactions”, *Urban Studies*, 51(3): 463-476.
 32. Islam, M.R. and Saphores, J.D.M., 2022. “An LA Story: The Impact of Housing Costs on Commuting”, *Journal of Transport Geography*, 98: 103266.
 33. Kim, J.H., Pagliara, F., and Preston, J., 2005. “The Intention to Move and Residential Location Choice Behaviour”, *Urban Studies*, 42(9): 1621-1636.
 34. Liu, J., Wang, B., and Xiao, L., 2021. “Non-linear Associations between Built Environment and Active Travel for Working and Shopping: An Extreme Gradient Boosting Approach”, *Journal of Transport Geography*, 92: 103034.
 35. Molho, I., 1986. “Theories of Migration: A Review”, *Scottish Journal of Political Economy*, 33(4): 396-419.
 36. Moos, M. and Skaburskis, A., 2010. “Workplace Restructuring and Urban Form: The Changing National Settlement Patterns of the Canadian Workforce”, *Journal of Urban Affairs*, 32(1): 25-53.
 37. Park, K.A. and Quercia, R., 2015. “Housing Costs and Commuting Distance”, *Journal of the Center for Real Estate Studies*, 3(2): 6-25.
 38. Parkes, A., Kearns, A., and Atkinson, R., 2002. “What Makes People Dissatisfied with their Neighborhoods?”, *Urban Studies*, 39(13): 2413-2438.
 39. Phe, H.H. and Wakely, P., 2000. “Status, Quality and the Other Trade-off: Towards a New Theory of Urban Residential Location”, *Urban Studies*, 37(1): 7-35.
 40. Rapoport, A., 1980. “Environmental Preference, Habitat Selection and Urban Housing”, *Journal of Social Issues*, 36(3): 118-134.
 41. Rouwendal, J. and Meijer, E., 2001. “Preferences for Housing, Jobs, and Commuting: A Mixed Logit Analysis”, *Journal of Regional Science*, 41(3): 475-505.
 42. Srinivas, P. and Katarya, R., 2022. “hyOPTXg: OPTUNA Hyper-parameter Optimization Framework for Predicting Cardiovascular Disease using XGBoost”, *Biomedical Signal Processing and Control*, 73: 103456.
 43. Tyrvaainen, L. and Vaananen, H., 1998. “The Economic Value of Urban Forest Amenities: An Application of the Contingent Valuation Method”, *Landscape and Urban Planning*, 43(1-3): 105-118.
 44. Weisbrod, G.E., Lerman, S.R., and Ben-Akiva, M., 1980. “Tradeoffs in Residential Location Decisions: Transportation Versus Other Factors”, *Transport Policy and Decision Making*, 1(1): 13-26.
 45. Zhang, Y. and Haghani, A., 2015. “A Gradient Boosting Method to Improve Travel Time Prediction”, *Transportation Research Part C: Emerging Technologies*, 58: 308-324.
 46. Zhao, P., 2015. “The Determinants of the Commuting Burden of Low-income Workers: Evidence from Beijing”, *Environment and Planning A: Economy and Space*, 47(8): 1736-1755.
 47. Zheng, Z., Zhou, S., and Deng, X., 2022. “The Spatially Heterogeneous and Double-edged Effect of the Built Environment on Commuting Distance: Home-based and Work-based Perspectives”, *Plos One*, 17(3): e0262727.

Date Received	2023-02-23
Reviewed(1 st)	2023-05-30
Date Revised	2023-08-07
Reviewed(2 nd)	2023-08-13
Date Accepted	2023-08-13
Final Received	2023-09-19