

ارزیابی عملکرد الگوریتم‌های یادگیری ماشین و مدل هدانیک در برآورد ارزش ملک با تاکید بر شاخص‌های ساختاری و مکانی

مجید صائبی، محمد کریمی*

^۱ دانشجوی دکتری رشته سیستم‌های اطلاعات مکانی دانشگاه صنعتی خواجه نصیرالدین طوسی، تهران

m.saebi@email.kntu.ac.ir

^۲ دانشیار گروه سیستم‌های اطلاعات مکانی دانشگاه صنعتی خواجه نصیرالدین طوسی، تهران

mkarimi@kntu.ac.ir

(دریافت: دی ۱۴۰۴، تصویب: خرداد ۱۴۰۵)

چکیده

برآورد دقیق ارزش املاک مسکونی از عناصر کلیدی برای تصمیم‌گیری ذی‌نفعان، کارشناسان بازار مسکن و مدیران شهری به شمار می‌آید و به دلیل ماهیت چندوجهی و پیچیده عوامل مؤثر بر ارزش ملک، ترکیب شیوه‌های یادگیری ماشین با تحلیل‌های مکانی می‌تواند به بهبود دقت پیش‌بینی منجر شود. در این پژوهش، به‌طور هم‌زمان از مجموعه‌ای ترکیبی شامل ویژگی‌های ساختاری و شاخص‌های مکانی استخراج‌شده در محیط GIS استفاده شده و با به‌کارگیری چهار الگوریتم یادگیری ماشین شامل جنگل تصادفی، رگرسیون گرادیان تقویت‌شده، XGBoost و CatBoost، فرایند استخراج و بهینه‌سازی بردار ویژگی برای ارزش‌گذاری املاک انجام گرفته است. چارچوب پیشنهادی در دو رویکرد ارزش‌گذاری طراحی شده است که رویکرد اول رویکرد برآورد ارزش کل ملک است که قیمت نهایی واحد مسکونی را مدل می‌کند و رویکرد دوم رویکرد تعیین قیمت هر مترمربع است که با نرمال‌سازی نسبت به مساحت، امکان مقایسه‌پذیری بیشتر بین املاک را فراهم می‌سازد؛ در هر دو رویکرد، عملکرد مدل‌های یادگیری ماشین با مدل کلاسیک هدانیک مبتنی بر رگرسیون OLS مقایسه شده است. نتایج این مطالعه که در منطقه ۵ شهر تهران طی یک دوره دوماهه انجام شده است نشان می‌دهد که مدل‌های یادگیری ماشین در هر دو رویکرد، عملکردی برتر نسبت به مدل هدانیک ارائه کرده‌اند و در این میان، الگوریتم CatBoost با ثبت مقادیر ۰/۳۰ برای ریشه میانگین مربعات خطا، ۰/۸۲ برای ضریب تعیین و ۱۰/۰۴ درصد برای میانگین درصد خطای مطلق، در مجموع پایدارترین و دقیق‌ترین عملکرد را نشان داده است؛ در حالی که مدل هدانیک در بهترین حالت به RMSE برابر با ۰/۳۶ و MAPE معادل ۱۲/۱۳ درصد دست یافته است. این نتایج بیانگر تأثیر قابل‌توجه ترکیب تحلیل‌های مکانی با روش‌های یادگیری ماشین و به‌ویژه الگوریتم CatBoost در ارزش‌گذاری املاک مسکونی است.

واژگان کلیدی: برآورد قیمت املاک مسکونی، سیستم‌های اطلاعات مکانی، یادگیری ماشین، مدل هدانیک، شاخص‌های مکانی

۱- مقدمه

مطالعه در زمینه برآورد ارزش املاک مسکونی به دلیل تأثیری که بر تصمیم‌گیری‌های سرمایه‌گذاران در خرید، فروش و سرمایه‌گذاری دارد، از اهمیت بسیار زیادی برخوردار است. این مطالعات به فرایندهای تصمیم‌گیری مالی از جمله تعیین ارزش، ارائه وام، تعیین مالیات و انجام تجزیه و تحلیل سرمایه‌گذاری مرتبط هستند. علاوه بر این، نقش مهمی در حل اختلافات قانونی دارند و به مشکلات مربوط به بیمه نیز کمک می‌کنند؛ بنابراین، ارزش‌گذاری املاک به عنوان ابزاری مهم برای کلیه افراد جامعه و همچنین نهادهایی که در تصمیم‌گیری‌های مالی و سرمایه‌گذاری دخیل هستند شناخته می‌شود [۱]، [۲].

ارزش‌گذاری املاک مسکونی فرایندی چندوجهی و پیچیده است که تحت تأثیر عوامل مختلفی قرار دارد. در این میان، ویژگی‌های فیزیکی و مکانی ملک نقش بسیار مهمی ایفا می‌کنند [۳]. موقعیت مکانی ملک، تعداد اتاق خواب، سال ساخت، تعداد طبقات، متراژ و نوع خانه از جمله عوامل اصلی در تعیین قیمت هستند [۳]، [۴]، [۵]. دسترسی به امکانات و خدمات شهری، نزدیکی به مراکز آموزشی و رفاهی و وضعیت ترافیکی می‌تواند تأثیر بسزایی بر ارزش ملک داشته باشند. علاوه بر ویژگی‌های فیزیکی و مکانی، عواملی مانند عرضه و تقاضا در بازار، نرخ بهره و سلامت کلی اقتصاد نیز مؤثرند؛ با این حال، تمرکز بر تحلیل‌های مکانی می‌تواند به برآورد دقیق‌تر و واقع‌بینانه‌تر قیمت املاک کمک کند و اطمینان از صحت پیش‌بینی‌ها را افزایش دهد. ارزیابی مستمر شرایط مکانی و بهره‌مندی از اطلاعات جدید در تحلیل‌های مکانی، به قیمت‌گذاری دقیق‌تر املاک مسکونی کمک می‌کند [۳].

از روش‌های مهم و سنتی برآورد ارزش املاک می‌توان به روش‌های مقایسه فروش، هزینه و درآمد اشاره کرد. در روش مقایسه فروش، با بررسی املاک مشابهی که در سال‌های اخیر مورد معامله قرار گرفته‌اند، ارزش ملک برآورد می‌شود. در روش هزینه، ارزش ملک بر اساس اختلاف هزینه خرید زمین و ساخت ملک از هزینه‌های ناشی از استهلاک به دست می‌آید. روش درآمد بیشتر برای املاک با کاربری صنعتی و تجاری استفاده شده و ارزش را بر پایه میزان درآمد یا مخارج املاک مشابه مانند هزینه نگهداری، هزینه بیمه و موارد دیگر مورد بررسی قرار می‌دهد [۶].

معروف‌ترین روش مبتنی بر مقایسه فروش برای برآورد ارزش املاک مسکونی، روش هدایتگر است. در سال‌های اخیر، در کنار این روش سنتی، از روش‌های دیگری مانند روش‌های یادگیری ماشین و روش‌های آماری نیز برای برآورد ارزش املاک مسکونی استفاده شده است.

فوسینگ و همکاران (۲۰۲۱) از مدل تقویتی برای پیش‌بینی قیمت مسکن در سنگاپور استفاده کرده‌اند. در این تحقیق علاوه بر عواملی مانند فاصله تا نزدیک‌ترین سیستم حمل و نقل عمومی سریع و مرکز تجاری شهر، ویژگی‌های دارایی مثل اندازه واحد، سطح طبقه و سال و ماه فروش بررسی شده و نتایج نشان داده است که مدل تقویتی نسبت به درخت تصمیم و رگرسیون چندگانه عملکرد بهتری دارد [۷]. سرگویان و همکاران (۲۰۲۲) در ارمنستان به عواملی مانند موقعیت مکانی، مشخصات ملک شامل تعداد اتاق‌ها، نوع ساختمان و وضعیت آن، و ویژگی‌های محله توجه کردند و نشان دادند که عملکرد مدل XGBoost از مدل جنگل تصادفی (RF) بهتر است [۸]. گوئرا و همکاران (۲۰۲۳) در پاناما درختان تقویت‌کننده گرادیان را بهترین مدل یافتند و از معیارهایی مانند مساحت، تعداد اتاق خواب، تعداد حمام، تعداد پارکینگ و موقعیت مکانی استفاده کردند [۹]. ایروکاس و همکاران (۲۰۱۷) از رگرسیون خطی و از ویژگی‌های مکانی و مشخصات فیزیکی ساختمان در قبرس استفاده کردند [۱۰]. کانگ و همکاران (۲۰۲۱) در بوستون، مدل‌های مختلف را بررسی و ماشین تقویت گرادیان را بهتر از رگرسیون خطی دانستند و بر ویژگی‌های ساختاری ملک، محیط ساخته شده، الگوهای تحرک انسانی و ویژگی‌های اجتماعی-اقتصادی محله‌ها تأکید کردند [۳]. کالمر و همکاران (۲۰۱۹) در لهستان از مدل‌های رگرسیون مکانی استفاده کردند [۱۱]. یانگ و همکاران (۲۰۲۲) از داده‌های حجیم در کالیفرنیا بهره بردند [۱۲]. چوی و همکاران (۲۰۲۳) در کره جنوبی از عواملی مانند ویژگی‌های فیزیکی، محیطی و دسترسی استفاده کرده و در مطالعه خود از الگوریتم‌های مختلف یادگیری ماشین بهره گرفتند. فان (۲۰۱۸) در ملبورن، روش SVM را مؤثرترین مدل دانست و بر عواملی مانند متغیرهای فیزیکی و موقعیتی ملک تمرکز کرد [۱۳]، [۱۴]. لورنز و همکاران (۲۰۲۲) نیز از یادگیری ماشین قابل تفسیر (IML) برای تحلیل ارزش ملک استفاده کرده و روش‌های خطی، مکانی و غیرخطی را مقایسه نمودند [۱۵].

جمع‌بندی مطالعات پیشین نشان می‌دهد که پژوهش‌های انجام‌شده، هرچند از نظر نوع داده، محدوده جغرافیایی و الگوریتم‌های مورد استفاده متفاوت هستند، در چند محور مشترک قابل مقایسه‌اند. بخش قابل‌توجهی از این مطالعات بر بهبود دقت پیش‌بینی از طریق الگوریتم‌های یادگیری ماشین تمرکز داشته‌اند و نشان داده‌اند که مدل‌های غیرخطی و گروهی، در بسیاری از موارد، عملکرد بهتری نسبت به مدل‌های کلاسیک دارند. همچنین، در شماری از پژوهش‌ها، نقش متغیرهای مکانی و محیطی مانند فاصله از مراکز خدماتی، حمل‌ونقل عمومی، ویژگی‌های محله و موقعیت مکانی ملک مورد توجه قرار گرفته است. با این حال، در اغلب این مطالعات، ویژگی‌های مکانی خردمقیاس و قابل استخراج از داده‌های شهری، مانند عرض معبر دسترسی و جهت استقرار ساختمان، کمتر وارد فرایند مدل‌سازی شده‌اند. افزون بر این، بیشتر پژوهش‌ها یا بر ارزش کل ملک تمرکز داشته‌اند یا قیمت واحد سطح را به‌صورت مستقل بررسی کرده‌اند و مقایسه نظام‌مند این دو رویکرد کمتر مورد توجه قرار گرفته است.

بر این اساس، مرور ادبیات نشان می‌دهد که مطالعات قبلی در زمینه ارزش‌گذاری املاک مسکونی به سه جنبه مهم توجه کمتری داشته‌اند و دقیقاً در همین سه محور، خلأ پژوهشی و نوآوری اصلی این تحقیق شکل می‌گیرد. نخست، در بسیاری از پژوهش‌ها به‌جای دسته‌بندی و انتخاب هدفمند ویژگی‌ها، صرفاً همه متغیرهای در دسترس وارد مدل شده‌اند؛ در حالی که تجمیع ویژگی‌های همبسته، مانند مساحت، تعداد اتاق، طبقه ملک، تعداد پارکینگ، وضعیت انباری، وضعیت آسانسور، وضعیت بالکن و وضعیت حیاط، در قالب معیارهای خلاصه‌تری مانند معیار مساحت می‌تواند به کاهش بعد، جلوگیری از هم‌خطی و در نهایت عملکرد سریع‌تر و بهینه‌تر مدل‌ها منجر شود [۳]، [۱۰]، [۱۱]، [۱۳]، [۱۵]. دوم، بسیاری از مدل‌های موجود تنها برخی ویژگی‌های ساختاری و مکانی را در نظر گرفته و سایر عوامل مؤثر از جمله جهت خانه و عرض معبر دسترسی را کمتر مورد بررسی قرار داده‌اند؛ موضوعی که می‌تواند ناشی از کمبود داده مناسب یا بسنده کردن به داده‌های موجود باشد و به نادیده گرفتن نقش واقعی برخی ویژگی‌های مکانی در قیمت منجر شود [۳]، [۱۱]، [۱۵]. سوم، تمرکز قسمت عمده‌ای از تحقیقات بر برآورد قیمت بر اساس «ارزش کل ملک» بوده و به برآورد «ارزش ملک بر اساس قیمت هر

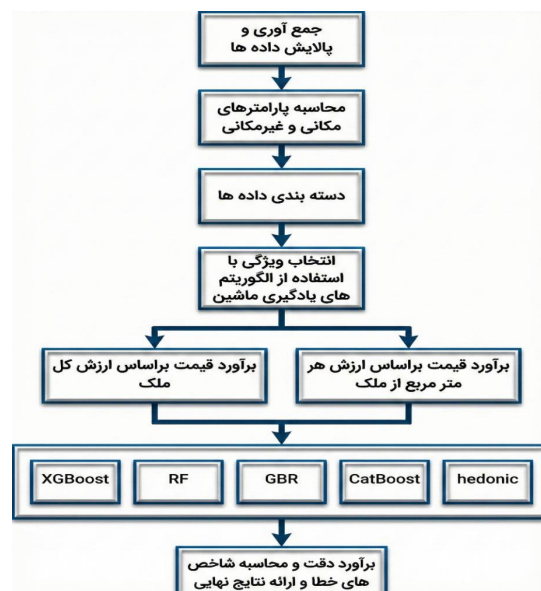
مترمربع» و مقایسه نظام‌مند این دو رویکرد توجه کمتری شده است؛ در حالی که هر دو شاخص در تصمیم‌گیری خریداران، فروشندگان و نهادهای مرتبط اهمیت دارند [۱۴]، [۱۲]، [۹]، [۷].

مطالعه حاضر با هدف برطرف کردن این ضعف‌ها طراحی شده است. در این تحقیق، ابتدا با استفاده از تکنیک‌های پیشرفته یادگیری ماشین و تحلیل‌های آماری، فرایند دسته‌بندی و انتخاب ویژگی‌ها انجام شده و ویژگی‌های همبسته در قالب معیارهای مناسب‌تری سازمان‌دهی شده‌اند. سپس، با استفاده از داده‌های موجود و به‌کارگیری تحلیل‌های مکانی، بدون تحمیل هزینه زیاد برای جمع‌آوری داده، مجموعه گسترده‌تری از معیارهای مؤثر از جمله ویژگی‌هایی مانند جهت ملک و عرض معبر دسترسی در مدل‌سازی برآورد املاک به‌کار گرفته شده است. در نهایت، کارایی چهار روش متداول یادگیری ماشین شامل جنگل تصادفی، رگرسیون تقویت‌شده با گرادیان، XGBoost و CatBoost در مقایسه با روش سنتی هدائیک، هم در زمینه برآورد بر اساس کل ارزش ملک و هم در زمینه برآورد بر اساس ارزش هر مترمربع ارزیابی شده تا تصویر جامع‌تری از توان مدل‌ها در دو رویکرد مختلف ارزش‌گذاری ارائه شود. به این ترتیب، نوآوری این پژوهش در ترکیب هم‌زمان سه محور یادشده و پرکردن خلأ موجود میان ادبیات کلاسیک ارزش‌گذاری و رویکردهای نوین مبتنی بر یادگیری ماشین و تحلیل‌های مکانی است.

۲- روش تحقیق

در این مطالعه، ابتدا با استفاده از بررسی نتایج تحقیقات پیشین، مشورت با کارشناسان حوزه املاک و بررسی داده‌های موجود در منطقه مورد مطالعه، ویژگی‌های مکانی و غیرمکانی مؤثر بر ارزش املاک شناسایی شدند. سپس داده‌های موردنیاز جهت برآورد ارزش املاک مسکونی جمع‌آوری و پالایش شدند و مقادیر ویژگی‌های مکانی و غیرمکانی مؤثر بر ارزش املاک محاسبه شدند. در گام بعدی، با بررسی خودهمبستگی بین ویژگی‌های مکانی و غیرمکانی مؤثر بر ارزش املاک، داده‌ها دسته‌بندی شدند. سپس برای کاهش ابعاد بردار ویژگی‌ها، از الگوریتم‌های یادگیری ماشین استفاده شد تا مهم‌ترین ویژگی‌ها در این زمینه انتخاب شوند. در نهایت، با تنظیم فرآیندهای هر

الگوریتم، فرایند برآورد ارزش املاک مسکونی انجام شده و دقت هر الگوریتم با استفاده از شاخص‌های تعیین دقت محاسبه شد. ساختار این مطالعه در شکل (۱) نمایش داده شده است.



شکل ۱. ساختار کلی مطالعه

۲-۱- جمع آوری و پالایش داده‌ها

در این پژوهش از مجموعه داده مربوط به املاک مسکونی منطقه ۵ شهرداری تهران استفاده شده است که پیش‌تر توسط جعفری و همکاران [۱۶] گردآوری و در مطالعه برآورد قیمت املاک مسکونی به کار گرفته شده بود. این مجموعه شامل ۲۶۹۶ نمونه آپارتمان مسکونی در بازه زمانی دی و بهمن ۱۳۹۹ است و منبع اولیه آن، بنا بر گزارش مطالعه مذکور، وبگاه بنگاه معاملات ملکی دودوتا بوده است. داده‌ها شامل قیمت پیشنهادی/تخمینی بازار و مشخصات ساختاری واحدهای مسکونی نظیر مساحت، تعداد اتاق، سال ساخت، تعداد پارکینگ، انباری، بالکن، حیاط و سایر امکانات هستند. از آنجا که دسترسی عمومی به قیمت‌های قطعی و ثبت شده معاملات املاک با محدودیت همراه است، قیمت‌های مورد استفاده در این پژوهش الزاماً معادل قیمت نهایی معاملات رسمی نیستند. با این وجود، به دلیل ماهیت بازارمحور داده‌ها و پوشش مکانی مناسب آن‌ها در منطقه مطالعه، این مجموعه داده برای ارزیابی نسبی عملکرد مدل‌های برآورد ارزش املاک مناسب در نظر گرفته شد. همچنین، در پژوهش حاضر،

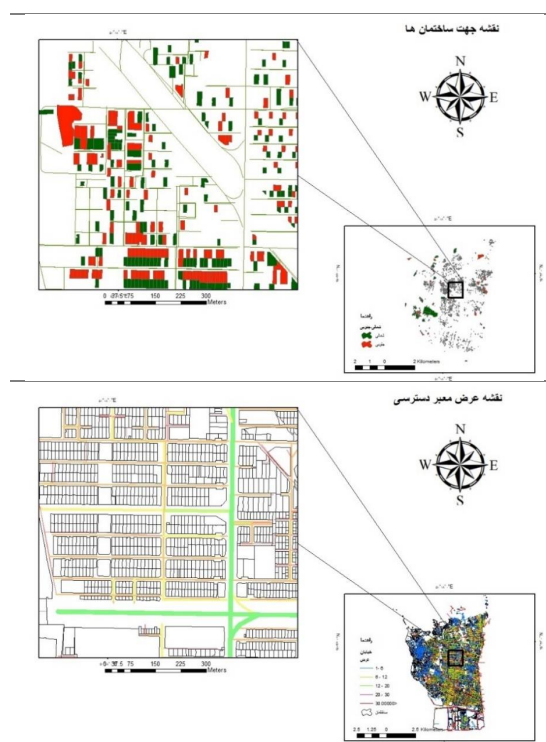
برای غنی‌سازی داده‌ها از منظر مکانی، لایه‌های پارسل، شبکه معابر، کاربری اراضی، بلوک‌های جمعیتی، مراکز خدماتی و ایستگاه‌های حمل‌ونقل عمومی در محیط GIS آماده‌سازی و به رکوردهای ملکی متصل شدند.

۲-۱-۱- محاسبه پارامترهای مکانی و غیرمکانی

پس از گردآوری و پالایش داده‌ها، کلیه معیارهای مؤثر بر ارزش املاک مسکونی محاسبه شد. برای استخراج لایه‌های اطلاعاتی موردنیاز، از مجموعه ابزارهای Analysis Tools، Spatial Analyst Tools، Network Analyst Tools در محیط نرم‌افزار ArcGIS استفاده گردید. در این چارچوب، ابتدا نقشه‌های مربوط به معیارهای ساختاری ملک شامل مساحت، تعداد اتاق، تعداد پارکینگ، تعداد انباری، وضعیت آسانسور، وضعیت بالکن، وضعیت حیاط، سال ساخت و وجود امکانات اضافی تهیه شد. همچنین لایه‌های تراکم جمعیت، فاصله از مراکز حمل‌ونقل عمومی (ایستگاه‌های مترو، اتوبوس و تاکسی)، فاصله از کاربری‌ها و مراکز مهم شهری (درمانی، صنعتی، نظامی، تجاری، آموزشی، ورزشی، مذهبی، فرهنگی و فضای سبز)، فاصله از جاده‌های اصلی و میزان آسیب‌پذیری لرزه‌ای تولید و به‌عنوان ورودی مدل‌ها آماده گردید.

در گام بعد، دو معیار مکانی مهم و کمتر به کاررفته در مطالعات پیشین، یعنی عرض معبر دسترسی و جهت استقرار ساختمان، استخراج شد. از آنجا که در داده‌های پایه شهری، لایه یا فیلد مستقلی برای این دو معیار وجود نداشت، فرایند استخراج آن‌ها با استفاده از ترکیب ابزارهای تحلیل مکانی در محیط ArcGIS و کدنویسی در زبان Python انجام گرفت. برای محاسبه عرض معبر، ابتدا لایه شبکه معابر و پارسل‌های شهری از نظر هندسی و توپولوژیکی آماده‌سازی شد. سپس برای هر قطعه معبر، پارسل‌های واقع در دو سوی آن شناسایی و فاصله کوتاه‌ترین عمود میان مرز پارسل‌های روبه‌رو محاسبه شد. میانگین این فاصله‌ها به‌عنوان عرض تقریبی معبر در نظر گرفته و به‌صورت یک فیلد توصیفی در لایه معابر ذخیره گردید. در ادامه، هر ملک بر اساس نزدیک‌ترین معبر دسترسی به آن متصل شد و مقدار عرض معبر متناظر به رکورد ملکی اختصاص یافت. برای تعیین جهت استقرار ساختمان نیز ابتدا نزدیک‌ترین معبر دسترسی هر پارسل مشخص شد و سپس با مقایسه موقعیت نسبی مرکز هندسی پارسل نسبت به راستای معبر مجاور، وضعیت قرارگیری ملک در یکی از دو گروه

شمالی یا جنوبی تعیین گردید. این متغیر به صورت طبقه‌ای کدگذاری و به لایه املاک افزوده شد. نتایج این فرایند در شکل (۲) ارائه شده و نشان می‌دهد که بهره‌گیری از تحلیل‌های مکانی می‌تواند امکان استخراج شاخص‌های خردمقیاس و مؤثر بر ارزش املاک را فراهم کند؛ شاخص‌هایی که در بسیاری از مطالعات پیشین به دلیل نبود داده آماده، کمتر مورد توجه قرار گرفته‌اند.



شکل ۲. نقشه عرض معبر دسترسی و جهت ساختمان‌ها

۲-۱-۲- دسته‌بندی داده‌ها

در این مطالعه، فرایند دسته‌بندی ویژگی‌ها با اتکا به ترکیبی از نظرات کارشناسان خبره در حوزه املاک و شهرسازی و تحلیل آماری (همبستگی میان ویژگی‌ها) انجام شده است. به طور کلی، دو الگوی دسته‌بندی برای داده‌ها در نظر گرفته شده است. در الگوی نخست، ۱۴ معیار تعریف شده‌اند که هر یک از این معیارها خود بعضاً از چند زیرمعیار تشکیل می‌شوند. به عنوان نمونه، مطابق جدول (۱)، معیار مساحت به عنوان یک معیار ساختاری شامل زیربنای واحد مسکونی، تعداد اتاق، تعداد پارکینگ، تعداد انباری، وضعیت آسانسور، وضعیت بالکن و وضعیت حیاط است.

در الگوی دوم، کلیه معیارهای مطالعه در پنج گروه اصلی «مساحت»، «تراکم»، «آسیب‌پذیری لرزه‌ای»، «فاصله

از مراکز خدماتی» و «فاصله از شبکه حمل‌ونقل» سازمان‌دهی شده‌اند. اطلاعات مربوط به دسته‌بندی ویژگی‌های ساختاری و مکانی املاک مسکونی مورد استفاده در این تحقیق در جداول (۱) و (۲) ارائه شده است.

جدول ۱. جدول معیارهای ارزیابی در مطالعه (دسته بندی نوع اول)

ردیف	متغیر	شرح
۱	مساحت	متراژ زیربنای ملک
		تعداد اتاق
		تعداد پارکینگ
		تعداد انباری
		وجود آسانسور
		وجود بالکن
		حیاط
۲	سال ساخت	سال ساخت ملک
۳	امکانات	وجود امکانات اضافی (وجود سالن‌های ورزشی، استخر و...)
۴	آسیب‌پذیری	میزان آسیب‌پذیری لرزه‌ای محله ملک
۵	فاصله - خیابان	فاصله اقلیدسی تا نزدیک‌ترین خیابان
۶	فاصله - درمانی	فاصله اقلیدسی تا نزدیک‌ترین مرکز درمانی
۷	فاصله - صنعتی	فاصله اقلیدسی تا نزدیک‌ترین مرکز صنعتی
۸	فاصله - نظامی	فاصله اقلیدسی تا نزدیک‌ترین مرکز نظامی
۹	فاصله - حمل‌ونقل	فاصله اقلیدسی تا نزدیک‌ترین ایستگاه تاکسی
		فاصله اقلیدسی تا نزدیک‌ترین ایستگاه مترو
		فاصله اقلیدسی تا نزدیک‌ترین ایستگاه اتوبوس
		فاصله اقلیدسی تا نزدیک‌ترین مرکز آموزشی
۱۰	فاصله - خدمات	فاصله اقلیدسی تا نزدیک‌ترین مرکز ورزشی
		فاصله اقلیدسی تا نزدیک‌ترین مکان مذهبی
		فاصله اقلیدسی تا نزدیک‌ترین مرکز فرهنگی
		فاصله اقلیدسی تا نزدیک‌ترین مرکز خرید
		فاصله اقلیدسی تا نزدیک‌ترین فضای سبز
		میزان تراکم جمعیت
۱۱	تراکم جمعیت	میزان تراکم جمعیت
۱۲	استهلاک	میزان استهلاک زمانی ملک
۱۳	جهت	جهت شمالی یا جنوبی بودن ملک در معبر دسترسی
۱۴	عرض معبر	متراژ عرض معبر دسترسی ملک

جدول ۲. جدول معیارهای ارزیابی در مطالعه (دسته بندی نوع دوم)

ردیف	متغیر	شرح
۱	مساحت	متراژ زیربنای ملک
		تعداد اتاق با ضریب
		تعداد پارکینگ
		تعداد انباری
		وجود آسانسور
		وجود بالکن
		حیاط
		امکانات اضافه
		جهت ملک
۲	تراکم جمعیت	سال ساخت
		میزان تراکم جمعیت
		عرض معبر دسترسی با ضریب تاثیر ۱ درصد در متر
۳	آسیب پذیری لرزه ای	میزان آسیب پذیری لرزه ای محله ملک
۴	فاصله - خدمات	فاصله اقلیدسی تا نزدیکترین مرکز آموزشی
		فاصله اقلیدسی تا نزدیکترین مرکز ورزشی
		فاصله اقلیدسی تا نزدیکترین مکان مذهبی
		فاصله اقلیدسی تا نزدیکترین مرکز فرهنگی
		فاصله اقلیدسی تا نزدیکترین مرکز خرید
		فاصله اقلیدسی تا نزدیکترین فضای سبز
		فاصله اقلیدسی تا نزدیکترین مرکز درمانی
		فاصله اقلیدسی تا نزدیکترین مرکز صنعتی
		فاصله اقلیدسی تا نزدیکترین مرکز نظامی
		فاصله اقلیدسی تا نزدیکترین ایستگاه تاکسی
		فاصله اقلیدسی تا نزدیکترین ایستگاه مترو
		فاصله اقلیدسی تا نزدیکترین ایستگاه اتوبوس
۵	فاصله - حمل و نقل	فاصله اقلیدسی تا نزدیکترین خیابان

افزایش ابعاد بردار ویژگی و تعداد پارامترهای مدل تخمین‌گر شده و در نهایت ممکن است به کاهش تعمیم‌پذیری مدل منجر شود [۱۷]، [۱۸].

برای مواجهه با این چالش، از دو گروه اصلی روش‌ها، یعنی نگاشت ویژگی و انتخاب ویژگی برای کاهش بعد استفاده می‌شود. در این تحقیق، از الگوریتم‌های یادگیری ماشین معرفی شده، در قالب روش‌های نوع Wrapper برای انتخاب ویژگی بهره گرفته شده است. این الگوریتم‌ها با ارزیابی تکراری ترکیب‌های مختلف ویژگی‌ها و انتخاب مهم‌ترین آن‌ها، ضمن کاهش پیچیدگی مدل، به بهبود دقت تخمین نیز کمک می‌کنند. در واقع، روش‌های یادگیری ماشین مذکور با استفاده از الگوریتم‌های جستجو، مناسب‌ترین زیرمجموعه ویژگی‌ها را برای برآورد ارزش املاک شناسایی می‌کنند [۱۸].

۲-۲- روش‌های یادگیری ماشین

در این پژوهش، برای برآورد ارزش املاک مسکونی از چهار الگوریتم متداول و کارآمد یادگیری ماشین شامل CatBoost، جنگل تصادفی (Random Forest)، رگرسیون تقویت‌شده با گرادیان (GBR) و XGBoost استفاده شده است. وجه اشتراک این الگوریتم‌ها، تکیه بر درخت‌های تصمیم و رویکردهای گروهی (Ensemble) است که آن‌ها را برای داده‌های جدولی با ترکیبی از ویژگی‌های ساختاری و مکانی به گزینه‌ای مناسب برای مدل‌سازی روابط پیچیده و غیرخطی بین متغیرها و قیمت ملک تبدیل می‌کند [۱۸]. [۱۹]، [۲۰]، [۲۱]، [۲۲]، [۲۳]، [۲۴]، [۲۵]، [۲۶]. در ادامه، خلاصه‌ای از سازوکار هر الگوریتم و منطق استفاده از آن‌ها در این مطالعه ارائه می‌شود.

۲-۲-۱- الگوریتم CatBoost

CatBoost الگوریتمی مبتنی بر تقویت گرادیان روی درخت‌های تصمیم است که توسط شرکت Yandex توسعه داده شده و به‌طور ویژه برای کار با داده‌های جدولی ناهمگن شامل ترکیبی از متغیرهای کمی و کیفی طراحی شده است. این الگوریتم با به‌کارگیری طرح‌های خاص برای تبدیل متغیرهای طبقه‌ای و استفاده از «تقویت ترتیبی»، مشکل برازش بیش از حد ناشی از پردازش نادرست داده‌های طبقه‌ای را تا حد زیادی کاهش می‌دهد.

۲-۱-۳- انتخاب ویژگی

برآورد قیمت املاک مسکونی مسئله‌ای پیچیده است که تحت تأثیر مجموعه‌ای از عوامل کمی و کیفی مرتبط با خود ملک و همچنین ویژگی‌های منطقه‌ای که ملک در آن واقع شده قرار دارد. به دلیل دشواری در شناسایی دقیق تمامی پارامترهای مؤثر، معمولاً تعداد زیادی ویژگی به‌عنوان نماینده این عوامل استخراج می‌شود. این امر اگرچه می‌تواند اطلاعات بیشتری را وارد مدل کند، اما در عین حال باعث

در CatBoost، مدل نهایی از ترکیب مجموعه‌ای از درخت‌های تصمیم تشکیل می‌شود که هر درخت به‌طور متوالی برای کاهش خطای باقی‌مانده مراحل قبلی آموزش داده می‌شود. پایداری عددی، نیاز کمتر به پیش‌پردازش و کارایی مناسب در داده‌های با ابعاد متوسط تا بزرگ باعث شده‌است که این الگوریتم در سال‌های اخیر در مطالعات گوناگون مورد توجه قرار گیرد [۲۰]، [۱۹].

۲-۲-۲- جنگل تصادفی

جنگل تصادفی یک روش گروهی مبتنی بر بوت‌استرپ و درخت‌های تصمیم است که در آن مجموعه‌ای از درخت‌ها روی نمونه‌های تصادفی (با جایگزینی) از داده‌های آموزشی و با انتخاب تصادفی زیرمجموعه‌ای از ویژگی‌ها در هر گره ساخته می‌شوند. پیش‌بینی نهایی برای هر نمونه معمولاً از طریق میانگین‌گیری (در مسائل رگرسیونی) از خروجی درخت‌ها به دست می‌آید. این رویکرد با کاهش واریانس مدل و افزایش تنوع درخت‌ها، منجر به دقت تعمیم مناسب و مقاومت در برابر بیش‌برازش می‌شود. توانایی مدیریت داده‌های با ابعاد بالا، برخورد نسبتاً خوب با داده‌های ناقص و امکان استخراج اهمیت نسبی ویژگی‌ها از جمله مزایای مهم جنگل تصادفی است که آن را به یکی از الگوریتم‌های پرکاربرد در مدل‌سازی قیمت املاک تبدیل کرده است [۸]، [۲۱]، [۲۰].

۲-۲-۳- الگوریتم رگرسیون تقویت شده با گرادیان

رگرسیون تقویت شده با گرادیان (Gradient Boosting Regression) الگوریتمی گروهی است که در آن درخت‌های تصمیم کم‌عمق به‌عنوان یادگیرنده‌های ضعیف به‌صورت ترتیبی آموزش داده می‌شوند. در هر تکرار، درخت جدید برای تقریب خطاهای باقی‌مانده یا گرادیان تابع زیان مدل قبلی ساخته شده و به مدل افزوده می‌شود. به این ترتیب، مدل نهایی ترکیبی خطی از درخت‌های متوالی است که هر یک بخشی از ساختار غیرخطی رابطه بین ویژگی‌ها و متغیر هدف را به‌خوبی بازنمایی می‌کند. انتخاب مناسب تعداد درخت‌ها، نرخ یادگیری و عمق درخت‌ها نقش کلیدی در کنترل بیش‌برازش و دستیابی به دقت مناسب دارد. به‌دلیل انعطاف‌پذیری بالا و توان مدل‌سازی روابط پیچیده، الگوریتم

GBR در بسیاری از مطالعات ارزش‌گذاری املاک با موفقیت به‌کار رفته است [۲۲]، [۲۳]، [۲۵]، [۲۴].

۲-۲-۴- الگوریتم XGBoost

XGBoost نسخه‌ای بهینه‌سازی‌شده و مقیاس‌پذیر از روش تقویت گرادیان بر پایه درخت تصمیم است که با تمرکز بر کارایی محاسباتی، منظم‌سازی (Regularization) و مدیریت مؤثر داده‌های پراکنده توسعه یافته است. این الگوریتم با استفاده از تابع هدف شامل جزء خطا و عبارات‌های منظم‌سازی، کنترل خوبی بر پیچیدگی مدل اعمال کرده و از بیش‌برازش جلوگیری می‌کند. قابلیت‌هایی مانند پردازش موازی، پشتیبانی از داده‌های حجیم و امکان تنظیم دقیق فرآیندها باعث شده است که XGBoost در رقابت‌های علمی و کاربردهای عملی متعددی از جمله مدل‌سازی قیمت املاک به‌عنوان یکی از الگوریتم‌های قدرتمند شناخته شود [۲۶]، [۲۵].

۲-۲-۵- روش هدانیک

روش ارزش‌گذاری هدانیک یکی از رویکردهای کلاسیک در تحلیل قیمت املاک است که در آن، قیمت بازار یک ملک به‌عنوان تابعی از مجموعه‌ای از ویژگی‌های فیزیکی، ساختاری، خدماتی و شرایط محیطی آن در نظر گرفته می‌شود. به‌صورت کلی، این رابطه را می‌توان به شکل زیر نمایش داد [۲۹]، [۲۸].

$$P = P(S, A, E, D) \quad (۱)$$

در این رابطه، P قیمت فروش ملک، S بردار ویژگی‌های ساختاری ملک (مانند مساحت، تعداد اتاق و کیفیت ساخت)، A متغیرهای مرتبط با نزدیکی و دسترسی به مراکز خدمات شهری، E ویژگی‌های محلی و محیطی و D سایر ویژگی‌های مؤثر (از جمله ویژگی‌های خاص موقعیت یا بنا) را نشان می‌دهد.

با جمع‌کردن همه این ویژگی‌ها در یک بردار کلی X ، تابع قیمت هدانیک را می‌توان به‌صورت زیر بازنویسی کرد:

$$P = f(x), X = (X_1, X_2, X_3, \dots, X_n) \quad (۲)$$

که در آن X_k نشان‌دهنده‌ی k امین ویژگی ملک است. در چارچوب نظریه هدانیک، «قیمت ضمنی» هر ویژگی از

طریق مشتق جزئی تابع قیمت نسبت به آن ویژگی تعریف می‌شود؛ یعنی:

$$\frac{\partial P}{\partial X_K}(X) = \frac{\partial f}{\partial X_K}(X). k = 1, \dots, k \quad (3)$$

این مقدار، تغییر نهایی در قیمت تعادلی ملک را نسبت به تغییر در ویژگی موردنظر، در حالی که سایر ویژگی‌ها ثابت فرض شده‌اند، نشان می‌دهد. به‌صورت شهودی، با افزایش ویژگی‌های مطلوب یک ملک، قیمت بازار آن نیز در شرایط برابر سایر عوامل افزایش خواهد یافت.

در عمل، تابع قیمت هدائیک اغلب با استفاده از یک فرم تابعی مشخص (معمولاً خطی یا شبه‌خطی) و روش حداقل مربعات معمولی (OLS) برآورد می‌شود. در این مطالعه، با الهام از تحقیقات پیشین، عوامل مؤثر بر قیمت به دو دسته اصلی ویژگی‌های ساختاری ساختمان و عوامل محیطی و دسترسی تفکیک شده‌اند و مدل قیمت هدائیک به‌صورت خطی زیر فرموله شده است [۲۷]، [۲۹]:

$$PRICE = a + \sum (b_i \times X_i) + \sum (c_j \times Y_j) \quad (4)$$

در این رابطه، PRICE متغیر وابسته (قیمت ملک)، a ثابت رگرسیون، X_i ویژگی‌های ساختاری ساختمان (مانند مساحت، تعداد اتاق و سایر مشخصات فیزیکی) و Y_j متغیرهای مربوط به شرایط محیطی و دسترسی (مانند فاصله تا مراکز خدماتی و شبکه حمل‌ونقل) را نشان می‌دهند. ضرایب b_i بیانگر تأثیر نسبی هر یک از ویژگی‌های ساختاری بر ارزش ملک و ضرایب c_j نشان‌دهنده اثر عوامل محیطی و دسترسی بر تغییرات قیمت هستند. در این پژوهش، مدل هدائیک با این فرم خطی برآورد شده و به‌عنوان مدل مرجع برای مقایسه عملکرد با الگوریتم‌های یادگیری ماشین مورد استفاده قرار گرفته است [۳۰]، [۲۷].

در این پژوهش، مجموعه ویژگی‌های ورودی مدل هدائیک شامل ۹ ویژگی ساختاری و ۱۸ ویژگی مکانی و محیطی است. ویژگی‌های ساختاری شامل مساحت زیربنا، تعداد اتاق، تعداد پارکینگ، تعداد انباری، وجود آسانسور، وجود بالکن، وجود حیاط، سن بنا و وجود امکانات اضافی هستند. ویژگی‌های مکانی و محیطی نیز شامل متغیرهایی مانند تراکم جمعیت، فاصله تا ایستگاه‌های حمل‌ونقل

عمومی (مترو، اتوبوس و تاکسی)، فاصله تا کاربری‌های مهم شهری (درمانی، تجاری، آموزشی، صنعتی، نظامی، فرهنگی، ورزشی و فضای سبز)، فاصله تا جاده‌های اصلی، آسیب‌پذیری لرزه‌ای، جهت ملک و عرض معبر دسترسی می‌شوند. تمام این متغیرها مطابق چارچوب نظری مدل هدائیک در دو دسته X_i (ویژگی‌های ساختاری) و Y_j (ویژگی‌های محیطی و دسترسی) در معادله (۴) سازمان‌دهی شده‌اند. این ساختار امکان تحلیل جامع تأثیر هم‌زمان ویژگی‌های فیزیکی و مکانی بر قیمت املاک را فراهم می‌کند و پایه‌ای منسجم برای مقایسه عملکرد مدل هدائیک با الگوریتم‌های یادگیری ماشین ایجاد می‌نماید.

۳-۲- معیارهای ارزیابی

در مطالعات مربوط به برآورد ارزش املاک، از شاخص‌های مختلفی مانند ضریب تعیین (R^2)^۱، جذر میانگین مربعات خطا (RMSE)^۲، میانگین درصد مطلق خطا (MAPE)^۳، دقت و صحت برای ارزیابی عملکرد مدل‌ها استفاده شده است [۳]، [۷]، [۸]، [۱۱]. در این پژوهش، با توجه به پیوسته‌بودن متغیر وابسته (قیمت ملک) و ماهیت مقایسه‌ای مطالعه، سه شاخص RMSE، MAPE و ضریب تعیین R^2 به‌عنوان معیارهای اصلی ارزیابی مدل‌ها به‌کار گرفته شده‌اند.

معیار RMSE میزان ریشه‌ی میانگین مربعات خطا را نشان می‌دهد و در واقع، اندازه‌ی خطای پیش‌بینی را در همان واحد متغیر وابسته (قیمت) بیان می‌کند. مقادیر کمتر RMSE بیانگر عملکرد بهتر مدل و تمرکز بیشتر پیش‌بینی‌ها در اطراف مقادیر واقعی است. این شاخص به‌صورت زیر محاسبه می‌شود:

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (y_i^{pred} - y_i)^2}{n}} \quad (5)$$

که در آن، n تعداد مشاهدات، y مقدار واقعی قیمت ملک و y_i^{pred} مقدار پیش‌بینی‌شده برای مشاهده‌ی i ام است.

^۱ R-square

^۲ Root Mean Squared Error

^۳ Mean Absolute Percentage Error

با توجه به اختلاف مقیاس قیمت املاک، متغیر هدف پیش از آموزش مدل‌ها مقیاس‌بندی شد و مقادیر RMSE گزارش‌شده در جداول نتایج بر اساس مقیاس نرمال‌شده متغیر هدف ارائه شده‌اند. از این‌رو، این مقادیر برای مقایسه نسبی عملکرد مدل‌ها در سناریوهای مختلف استفاده شده‌اند و تفسیر اقتصادی خطا در متن مقاله عمدتاً بر مبنای MAPE، به‌عنوان شاخص بدون واحد و درصدی، انجام شده است.

MAPE یک معیار رایج برای ارزیابی دقت مدل در مقیاس نسبی است و میانگین درصد خطای مطلق را نشان می‌دهد. این شاخص امکان مقایسه‌ی خطا را در سناریوهای مختلف فراهم می‌کند. تعریف ریاضی MAPE به‌صورت زیر است:

$$MAPE = \left(\frac{1}{n}\right) * \sum \left(\frac{|(y_i - y_{hat-i})|}{y_i}\right) \times 100 \quad (6)$$

در این رابطه، n تعداد کل مشاهدات، y_i مقدار واقعی و y_{hat-i} مقدار پیش‌بینی‌شده برای مشاهده‌ی i ام است و قدر مطلق کسر داخل جمع، درصد خطای مطلق برای هر مشاهده را نشان می‌دهد.

ضریب تعیین R^2 شاخصی آماری است که نسبت واریانس توضیح‌داده‌شده‌ی متغیر وابسته توسط مدل را نشان می‌دهد و در واقع بیان می‌کند که چه سهمی از تغییرات قیمت ملک توسط متغیرهای ورودی مدل تبیین شده است. این شاخص به‌صورت زیر محاسبه می‌شود:

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - y_i^{pred})^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (7)$$

در این معادله، $\bar{y}$ میانگین مقادیر واقعی قیمت است. مقدار R^2 در بازه‌ی $[0,1]$ قرار دارد؛ به‌طوری‌که مقادیر نزدیک به ۱ نشان‌دهنده‌ی برازش بهتر و توان بیش‌تر مدل در توضیح واریانس متغیر وابسته و مقادیر نزدیک به ۰ نشان‌دهنده‌ی توان توضیحی کم مدل است.

۲-۴- تقسیم داده‌ها و تنظیم فرآپارامترها

برای افزایش پایداری و قابلیت اتکای ارزیابی مدل‌ها، فرایند آموزش و آزمون تنها به یک بار تقسیم‌بندی داده‌ها محدود نشد. در این مطالعه از تقسیم‌های تصادفی

تکرارشونده استفاده شد؛ به این صورت که داده‌ها در چند مرحله و به‌صورت تصادفی به مجموعه‌های آموزش و آزمون تفکیک شدند و هر مدل در هر مرحله مجدداً آموزش یافت. در هر تکرار، شاخص‌های RMSE، MAPE و R^2 برای مجموعه آزمون محاسبه و در نهایت، میانگین نتایج این تکرارها به‌عنوان عملکرد نهایی هر مدل گزارش شد. این رویکرد وابستگی نتایج به یک تقسیم خاص از داده‌ها را کاهش داده و مقایسه مدل‌ها را پایدارتر و قابل‌اعتمادتر کرده است.

برای تنظیم فرآپارامترهای اصلی مدل‌های یادگیری ماشین، از روش جستجوی تصادفی^۱ در ترکیب با اعتبارسنجی متقاطع^۲ استفاده شد. در این فرایند، برای هر یک از الگوریتم‌های CatBoost، جنگل تصادفی، XGBoost و رگرسیون تقویت‌شده گرادیان، مجموعه‌ای از فرآپارامترهای اثرگذار مانند تعداد درخت‌ها، عمق درخت، نرخ یادگیری، حداقل تعداد نمونه در برگ و پارامترهای کنترل‌کننده بیش‌برازش در یک فضای جستجوی مشخص ارزیابی شد. ترکیب مناسب فرآپارامترها بر اساس کمینه‌سازی خطای پیش‌بینی و بیشینه‌سازی ضریب تعیین در داده‌های اعتبارسنجی انتخاب شد و سپس مدل‌های بهینه روی داده‌های آموزشی برازش داده شدند. در نهایت، عملکرد مدل‌ها بر اساس داده‌های آزمون و با استفاده از شاخص‌های RMSE، MAPE و R^2 گزارش شد. این رویکرد باعث شد مقایسه مدل‌ها صرفاً متکی بر تنظیمات پیش‌فرض نرم‌افزار نباشد و ارزیابی عملکرد آن‌ها در شرایط کنترل‌شده‌تر و قابل‌اعتمادتر انجام گیرد.

۳- برآورد ارزش املاک مسکونی

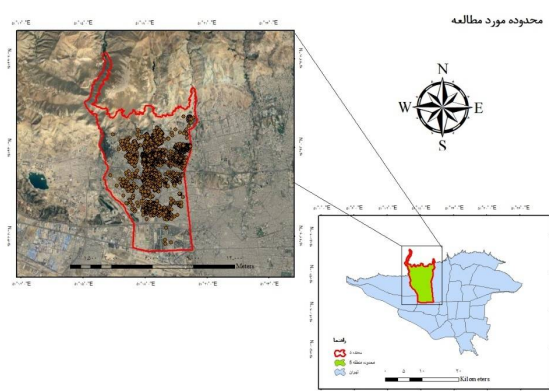
این بخش از پژوهش به تشریح ویژگی‌های منطقه مورد مطالعه و تحلیل داده‌های گردآوری‌شده برای برآورد ارزش املاک مسکونی اختصاص دارد. در این بخش، ابتدا خصوصیات ساختاری و مکانی منطقه تشریح شده و سپس ساختار داده‌ها، نوع متغیرها و ویژگی‌های استخراج‌شده برای مدل‌سازی قیمت با استفاده از الگوریتم‌های یادگیری ماشین ارائه می‌شود. این مرور، مبنای اصلی برای آماده‌سازی داده‌ها و اجرای مراحل مدل‌سازی در ادامه تحقیق است.

۱ Random Search

۲ Cross-Validation

۳-۱- منطقه مورد مطالعه^۱

کلان‌شهر تهران با جمعیتی در حدود ۹ میلیون نفر و مساحتی نزدیک به ۶۱۵ کیلومترمربع، به‌عنوان بزرگ‌ترین کانون جمعیتی و اقتصادی کشور شناخته می‌شود. این شهر بر اساس تقسیمات رسمی شامل ۲۲ منطقه، ۱۲۳ ناحیه و ۳۵۴ محله است (سالنامه آماری شهرداری تهران، ۱۳۹۹) [۳۱]. در میان این مناطق، منطقه ۵ به‌دلیل وسعت زیاد، جمعیت بالا، غلبه کاربری مسکونی و نقش پررنگ در پویایی بازار مسکن شهر، به‌عنوان منطقه مورد مطالعه در این پژوهش انتخاب شده است. این منطقه از مهم‌ترین کانون‌های توسعه آپارتمان‌نشینی و معاملات مسکن در غرب تهران به‌شمار می‌رود و از این رو، بررسی الگوی مکانی و قیمتی املاک در آن می‌تواند تصویر مناسبی از رفتار بازار مسکن شهری ارائه دهد. شکل (۳) موقعیت مکانی منطقه ۵ و توزیع مکانی داده‌های مورد استفاده در این تحقیق را نشان می‌دهد.

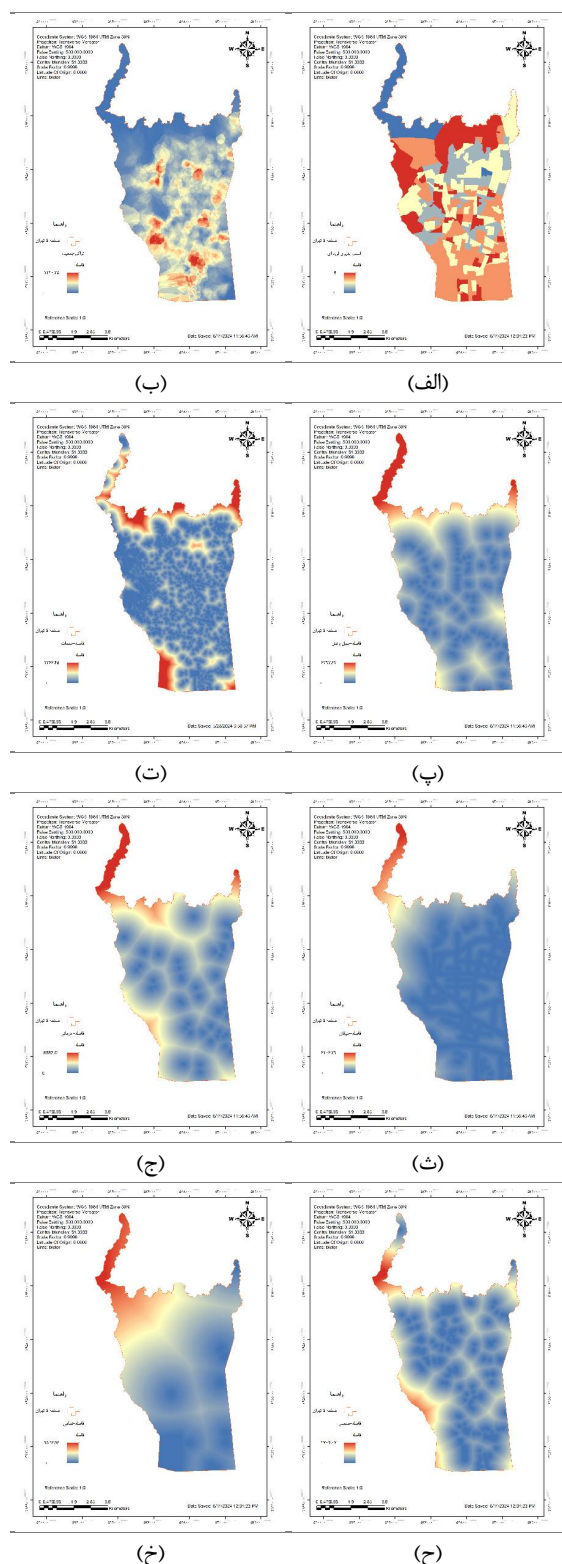


شکل ۳. منطقه مورد مطالعه و توزیع مکانی داده‌ها

۳-۲- معیارهای مؤثر در برآورد ارزش املاک

در این تحقیق، پس از شناسایی و انتخاب مناسب‌ترین پارامترهای مکانی مؤثر بر ارزش املاک مسکونی بر اساس ادبیات تحقیق و نظرات کارشناسان حوزه املاک و شهرسازی فرایند تهیه و تولید لایه‌های اطلاعاتی مورد نیاز انجام شد. در این چارچوب، لایه‌های مکانی شامل نقشه‌های توزیع تراکم جمعیت، فاصله تا کاربری‌ها و مراکز مهم شهری، فاصله تا جاده‌های اصلی و شاخص آسیب‌پذیری

لرزه‌ای به‌عنوان پارامترهای کلیدی تأثیرگذار بر قیمت املاک مسکونی تولید گردید.



شکل ۴. الف آسیب‌پذیری لرزه‌ای، ب) تراکم جمعیت، پ) فاصله از مراکز حمل و نقل، ت) فاصله از خدمات، ث) فاصله از خیابان، ج) فاصله از مراکز درمانی، ح) فاصله از مراکز صنعتی، خ) فاصله از مراکز نظامی

علاوه بر این، دو پارامتر جهت استقرار ساختمان و عرض معبر دسترسی که در اغلب مطالعات پیشین نادیده گرفته شده‌اند، با استفاده از تحلیل‌های مکانی ArcGIS و کدنویسی در محیط پایتون استخراج شدند و به‌عنوان معیارهای جدید در مدل‌سازی به‌کار گرفته شدند؛ امری که غنای مکانی مدل را به میزان قابل توجهی افزایش می‌دهد.

در ادامه، با تلفیق لایه‌های مکانی با داده‌های معاملاتی، مقادیر هر یک از پارامترها برای تمامی نمونه‌های املاک محاسبه شد. تمامی محاسبات فاصله با استفاده از ابزارهای تحلیل مکانی نرم‌افزار ArcGIS و بر مبنای فاصله اقلیدسی انجام شده است. لایه‌های اطلاعاتی به‌دست‌آمده برای املاک مسکونی منطقه مورد مطالعه در شکل (۴) ارائه شده است.

در مبحث ارزش‌گذاری املاک، تحلیل داده‌ها صرفاً به توصیف‌های سطحی محدود نمی‌شود و بررسی روابط آماری بین متغیرها از اهمیت ویژه‌ای برخوردار است. یکی از جنبه‌های کلیدی در این زمینه، تحلیل همبستگی میان معیارهای مختلف و متغیر هدف (ارزش ملک) است. در این پژوهش، برای سنجش شدت و جهت رابطه‌ی بین متغیرها از ضریب همبستگی پیرسون استفاده شده است تا درکی دقیق‌تر از نحوه‌ی اثرگذاری هر معیار بر ارزش املاک به‌دست آید و الگوهای پنهان در رفتار بازار مسکن آشکار شود [۵]، [۱۳]. علاوه بر این، نتایج این تحلیل همبستگی به‌عنوان مبنایی برای دسته‌بندی ویژگی‌ها و کاهش ابعاد بردار ویژگی نیز مورد استفاده قرار گرفته است؛ به‌گونه‌ای که ویژگی‌های دارای همبستگی بالا در گروه‌های مفهومی مشترک سازمان‌دهی شده‌اند.

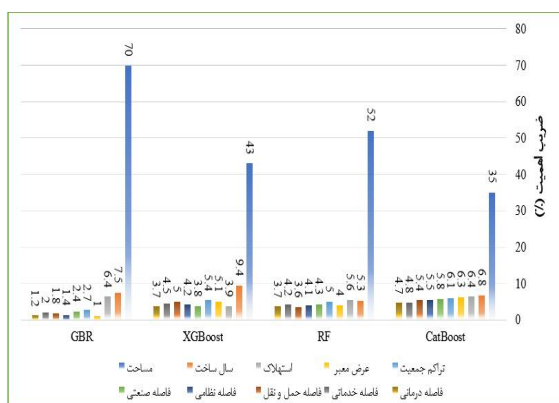
۴- نتایج

۴-۱- نتایج انتخاب ویژگی

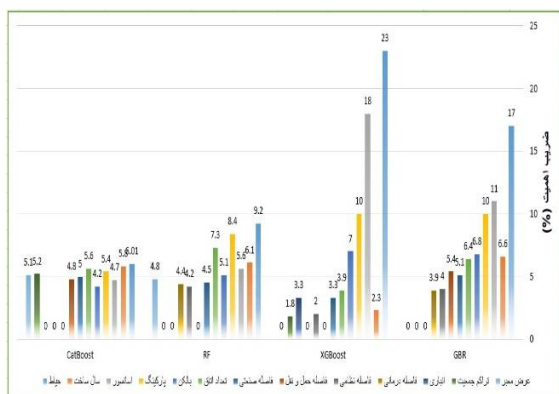
در این بخش، نتایج انتخاب ویژگی بر اساس دو رویکرد برآورد بر اساس ارزش کل ملک و برآورد بر اساس ارزش هر مترمربع از ملک و با استفاده از چهار الگوریتم یادگیری ماشین CatBoost، RF، XGBoost و GBR ارائه شده است. انتخاب ویژگی یکی از مراحل کلیدی در فرایند برآورد ارزش دارایی است؛ زیرا با شناسایی مرتبط‌ترین متغیرها نسبت به متغیر هدف، هم دقت مدل را بهبود می‌بخشد و هم از افزایش غیرضروری ابعاد مدل جلوگیری می‌کند. در این پژوهش، از قابلیت‌های الگوریتم‌های یادشده برای استخراج

پیش‌بینی‌کننده‌های اصلی استفاده شده است؛ به‌گونه‌ای که ۱۰ ویژگی با بالاترین اهمیت نسبی در هر رویکرد شناسایی و در نمودارهای (۱) و (۲) نمایش داده شده‌اند.

بر اساس نتایج نمودار (۱) و تفسیر آن با کمک کارشناسان خبره، معیار مساحت مهم‌ترین شاخص در برآورد ارزش کل ملک است و پس از آن به‌ترتیب معیارهای سال ساخت، استهلاک، تراکم جمعیت و عرض معبر دسترسی از بالاترین اهمیت برخوردارند. در مقابل، نتایج نمودار (۲) نشان می‌دهد که در رویکرد برآورد ارزش هر مترمربع از ملک، مهم‌ترین معیار، حیاط‌دار بودن ملک است و پس از آن به‌ترتیب سال ساخت، تعداد پارکینگ، تعداد اتاق و وجود آسانسور بیشترین نقش را در تبیین تغییرات قیمت هر مترمربع ایفا می‌کنند. این نتایج تأیید می‌کند که در تحلیل ارزش کل ملک، متغیرهای مرتبط با اندازه و مقیاس ملک نقش پررنگ‌تری دارند، در حالی که در تحلیل ارزش هر مترمربع، ویژگی‌های کیفی مرتبط با امکانات و فضای باز واحد مسکونی اهمیت بیشتری می‌یابند.



نمودار ۱. نمودار نتایج انتخاب ویژگی بر اساس ارزش کل مساحت ملک



نمودار ۲. نمودار نتایج انتخاب ویژگی بر اساس ارزش هر مترمربع از ملک

۴-۲- نتایج برآورد ارزش املاک

این مطالعه با هدف ارزیابی نظام‌مند عملکرد چهار الگوریتم پیشرفته یادگیری ماشین CatBoost، جنگل تصادفی (RF)، XGBoost و رگرسیون تقویت‌شده با گرادیان (GBR) در مقایسه با مدل رگرسیونی هدایک مرسوم (OLS) در برآورد ارزش املاک مسکونی انجام شده است. برای دستیابی به تحلیلی جامع، مدل‌سازی در دو چارچوب «ارزش کل ملک» و «قیمت هر مترمربع» و در قالب سه سناریوی متفاوت صورت گرفت: (۱) سناریوی پایه شامل تمامی پارامترهای ساختاری و مکانی، (۲) سناریوی حذف مکانی با کنار گذاشتن همه متغیرهای مکانی و (۳) سناریوی حذف هدفمند با حذف دو پارامتر بااهمیت یعنی «جهت استقرار ساختمان» و «عرض معبر دسترسی». داده‌های مورد استفاده شامل ۲۶۹۶ نمونه معاملاتی بود که با روش تقسیم‌بندی تصادفی به ۲۱۸۴ نمونه آموزشی و ۵۱۲ نمونه آزمون تفکیک شد.

در میان همه مدل‌های ارزیابی‌شده، الگوریتم CatBoost به‌عنوان مدلی برتر و پایدار در اغلب سناریوها ظاهر شد. این الگوریتم در چالش‌برانگیزترین شرایط، یعنی رویکرد ارزش کل تحت سناریوی پایه، به بالاترین سطح دقت دست یافت و مقادیر $MAPE = ۱۰.۰۴\%$ ، $R^2 = ۰.۸۲$ ، $RMSE = ۰.۳۰$ را ثبت کرد. عملکرد برتر CatBoost را می‌توان به قابلیت‌های ذاتی آن در پردازش کارآمد داده‌های طبقه‌بندی بدون نیاز به پیش‌پردازش گسترده و همچنین مقاومت در برابر بیش‌برازش نسبت داد. الگوریتم‌های RF و XGBoost نیز با عملکردی بسیار رقابتی و نزدیک به CatBoost، خود را به‌عنوان گزینه‌های قدرتمند و قابل‌اعتماد برای این‌گونه مسائل تثبیت کردند.

شاید یکی از یافته‌های کلیدی این پژوهش، که بر اهمیت آن در ادبیات موضوع تأکید شده است، نمایش آشکار تأثیر حیاتی و غیرقابل‌انکار پارامترهای مکانی بر دقت تمامی مدل‌ها باشد. تجزیه و تحلیل نتایج سه سناریو، روایتی گویا از این وابستگی ارائه می‌کند. حذف کلیه پارامترهای مکانی (سناریوی ۲) منجر به افت شدید و سیستماتیک دقت در همه مدل‌ها شد. برای مثال، در مدل برتر (CatBoost رویکرد ارزش کل)، این حذف باعث افزایش ۳۰٪ در RMSE و کاهش ۱۳ واحدی در ضریب تعیین (R^2) گردید. این افت محسوس به‌تنهایی گواهی

کافی بر ضرورت گنجاندن داده‌های مکانی در هر مدل ارزش‌گذاری املاک است.

جدول ۳. جدول مقادیر معیار ارزیابی برای هر مدل در رویکرد اول

مدل	سناریوی مورد ارزیابی	RMSE	R^2	MAPE (%)
الگوریتم CatBoost	وجود پارامترهای مکانی و ساختاری	۰.۳۰	۰.۸۲	۱۰.۰۴
	عدم وجود پارامترهای مکانی	۰.۳۹	۰.۶۹	۱۲.۱۸
	عدم وجود پارامترهای جهت و عرض معبر ملک	۰.۳۴	۰.۷۵	۱۱.۰۶
	وجود پارامترهای مکانی و ساختاری	۰.۳۲	۰.۷۹	۱۰.۴۳
	عدم وجود پارامترهای مکانی	۰.۳۹	۰.۶۸	۱۲.۴۰
	عدم وجود پارامترهای جهت و عرض معبر ملک	۰.۳۵	۰.۷۵	۱۱.۱۶
الگوریتم RF	وجود پارامترهای مکانی و ساختاری	۰.۳۲	۰.۷۹	۱۰.۴۸
	عدم وجود پارامترهای مکانی	۰.۴۱	۰.۶۶	۱۳.۲۴
	عدم وجود پارامترهای جهت و عرض معبر ملک	۰.۳۵	۰.۷۶	۱۱.۱۳
	وجود پارامترهای مکانی و ساختاری	۰.۳۳	۰.۷۷	۱۰.۹۴
	عدم وجود پارامترهای مکانی	۰.۳۹	۰.۶۷	۱۲.۵۳
	عدم وجود پارامترهای جهت و عرض معبر ملک	۰.۳۵	۰.۷۴	۱۱.۳۷
الگوریتم XGBoost	وجود پارامترهای مکانی و ساختاری	۰.۳۳	۰.۷۷	۱۰.۹۴
	عدم وجود پارامترهای مکانی	۰.۳۹	۰.۶۷	۱۲.۵۳
	عدم وجود پارامترهای جهت و عرض معبر ملک	۰.۳۵	۰.۷۴	۱۱.۳۷
	وجود پارامترهای مکانی و ساختاری	۰.۳۶	۰.۷۲	۱۲.۱۳
	عدم وجود پارامترهای مکانی	۰.۴۲	۰.۶۳	۱۳.۵۵
	عدم وجود پارامترهای جهت و عرض معبر ملک	۰.۴۰	۰.۶۷	۱۳.۰۰
الگوریتم GBR	وجود پارامترهای مکانی و ساختاری	۰.۳۶	۰.۷۲	۱۲.۱۳
	عدم وجود پارامترهای مکانی	۰.۴۲	۰.۶۳	۱۳.۵۵
	عدم وجود پارامترهای جهت و عرض معبر ملک	۰.۴۰	۰.۶۷	۱۳.۰۰
	وجود پارامترهای مکانی و ساختاری	۰.۳۶	۰.۷۲	۱۲.۱۳
	عدم وجود پارامترهای مکانی	۰.۴۲	۰.۶۳	۱۳.۵۵
	عدم وجود پارامترهای جهت و عرض معبر ملک	۰.۴۰	۰.۶۷	۱۳.۰۰

نکته مهم و قابل‌تأمل، نتایج سناریوی سوم است که در آن تنها دو پارامتر «جهت» و «عرض معبر» حذف شدند. مشاهده شد که حتی این حذف هدفمند نیز به افت محسوس و قابل‌اندازه‌گیری در دقت مدل‌ها منجر گردید. این یافته، به‌طور خاص بر اهمیت در نظرگیری این‌گونه ویژگی‌های مکانی بااهمیت ولی غالباً مغفول در مدل‌سازی‌ها

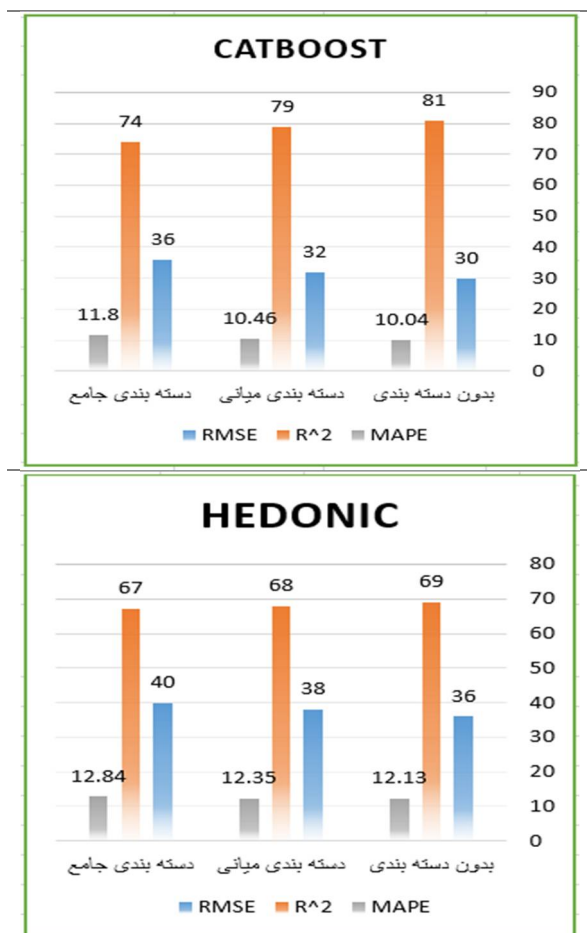
تأکید می‌کند و نشان می‌دهد که دقت یک مدل، علاوه بر قدرت الگوریتم، به شدت به غنای مجموعه داده‌های ورودی، به‌ویژه داده‌های مکانی، وابسته است.

جدول ۴. جدول مقادیر معیار ارزیابی برای هر مدل در رویکرد دوم (قیمت هر مترمربع از ملک)

مدل	سناریوی مورد ارزیابی	RMSE	R ²	MAPE (%)
الگوریتم CatBoost	وجود پارامترهای مکانی و ساختاری	۰.۳۳	۰.۷۷	۱۰.۶۰
	عدم وجود پارامترهای مکانی	۰.۴۰	۰.۶۷	۱۲.۹۰
	عدم وجود پارامترهای جهت و عرض معبر ملک	۰.۳۴	۰.۷۵	۱۱.۰۲
	وجود پارامترهای مکانی و ساختاری	۰.۳۲	۰.۷۷	۱۰.۴۲
	عدم وجود پارامترهای مکانی	۰.۴۱	۰.۶۷	۱۲.۹۵
	عدم وجود پارامترهای جهت و عرض معبر ملک	۰.۳۵	۰.۷۵	۱۱.۲۶
الگوریتم XGBoost	وجود پارامترهای مکانی و ساختاری	۰.۳۴	۰.۷۶	۱۰.۹۱
	عدم وجود پارامترهای مکانی	۰.۴۱	۰.۶۶	۱۳.۵۲
	عدم وجود پارامترهای جهت و عرض معبر ملک	۰.۳۷	۰.۷۳	۱۲.۲۱
	وجود پارامترهای مکانی و ساختاری	۰.۳۴	۰.۷۵	۱۱.۲۶
	عدم وجود پارامترهای مکانی	۰.۴۲	۰.۶۶	۱۲.۸۰
	عدم وجود پارامترهای جهت و عرض معبر ملک	۰.۳۶	۰.۶۹	۱۱.۹۴
الگوریتم GBR	وجود پارامترهای مکانی و ساختاری	۰.۳۶	۰.۷۲	۱۱.۵۳
	عدم وجود پارامترهای مکانی	۰.۴۱	۰.۶۷	۱۲.۸۵
	عدم وجود پارامترهای جهت و عرض معبر ملک	۰.۳۸	۰.۷۱	۱۲.۱۳
	وجود پارامترهای مکانی و ساختاری	۰.۳۶	۰.۷۲	۱۱.۵۳
	عدم وجود پارامترهای مکانی	۰.۴۱	۰.۶۷	۱۲.۸۵
	عدم وجود پارامترهای جهت و عرض معبر ملک	۰.۳۸	۰.۷۱	۱۲.۱۳

در مقایسه کلی بین دو رویکرد «ارزش کل ملک» و «قیمت هر مترمربع»، مدل‌ها عموماً در رویکرد اول (ارزش کل) عملکردی اندکی بهتر از خود نشان دادند. این امر می‌تواند ناشی از اثر تثبیت‌کننده و کاهش نویز در داده‌ها

هنگام استفاده از ارزش کل باشد، در حالی که قیمت هر مترمربع ممکن است نوسانات موضعی و پرنوسان‌تری داشته باشد. با این حال، شایان‌ذکر است که این تفاوت عملکرد در همه مدل‌ها چشمگیر نبود و الگوی کلی برتری یادگیری ماشین و اهمیت پارامترهای مکانی در هر دو رویکرد کاملاً یکسان و مشهود بود.



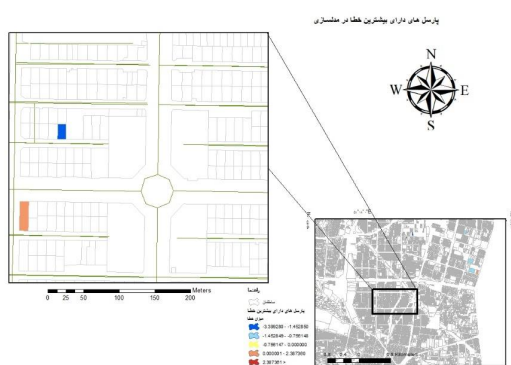
شکل ۷. نمودار شاخص‌های ارزیابی در رویکرد اول برای مدل‌های Catboost و هدانیک

با توجه به شاخص‌های ارزیابی به‌دست‌آمده و نتایج نمایش‌داده‌شده در جداول (۳) و (۴)، در رویکرد اول (برآورد ارزش کل ملک) مشاهده می‌شود که مدل‌های یادگیری ماشین در حالت دسته‌بندی میانی به‌طور بهینه‌تری عمل کرده‌اند و مقادیر RMSE، MAPE و R² آن‌ها بسیار نزدیک به حالت بدون دسته‌بندی ویژگی‌ها است. در مقابل، در مدل هدانیک، دسته‌بندی جامع عملکردی نزدیک به حالت بدون دسته‌بندی داشته و نسبت به سایر حالت‌های دسته‌بندی نتایج مناسب‌تری ارائه کرده است. در رویکرد دوم (برآورد ارزش هر مترمربع

۴-۲- تحلیل مکانی خطاها

برای بررسی دقیق‌تر عملکرد مدل‌ها، ۸۰ رکورد از معاملات ملکی که در میان نمونه‌های آزمون بیشترین مقدار خطا را داشتند انتخاب شد. این رکوردها روی نقشه منطقه جانمایی و از نظر موقعیت مکانی و شرایط اطراف ملک بررسی شدند تا مشخص شود چه عواملی می‌تواند باعث خطای زیاد در برآورد قیمت شده باشد.

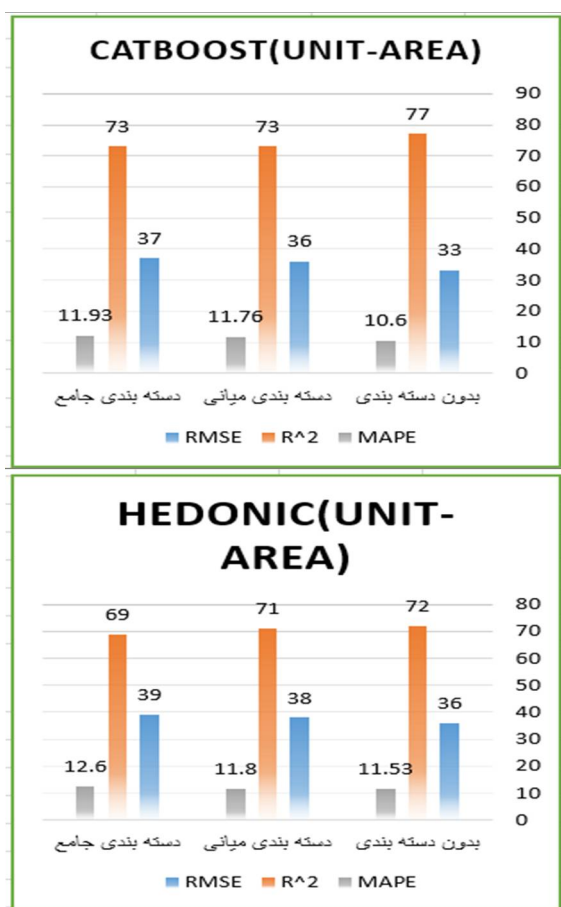
نتیجه این تحلیل مکانی نشان داد که بخشی از خطاهای بزرگ به ویژگی‌هایی مربوط است که در این پژوهش به‌صورت صریح وارد مدل نشده‌اند. به‌عنوان نمونه، بن‌بست بودن کوچه دسترسی در بسیاری از موارد با کاهش قیمت ملک همراه بوده، در حالی که این متغیر در مجموعه ویژگی‌ها وجود نداشته است. در مقابل، دوبر یا سه‌بر بودن پلاک (قرار گرفتن ملک در نبش یا داشتن چند جبهه دسترسی) در بعضی از پارسل‌ها باعث افزایش قیمت شده و یکی از دلایل اختلاف بین قیمت واقعی و قیمت برآوردشده است. در چند مورد دیگر نیز مجاورت ملک با بزرگراه‌ها و معابر پرتردد، به‌دلیل سر و صدا و شلوغی، تأثیر منفی بر ارزش ملک داشته، اما این عامل نیز به‌طور مستقیم در مدل لحاظ نشده است.



شکل ۹. پارسل‌های دارای بیشترین خطا (رنگ متمایل به آبی خطای منفی و قرمز خطای مثبت دارند)

در شکل (۹) نمونه‌ای از پارسل‌های دارای خطای مثبت و منفی نشان داده شده است؛ پارسلی که در کوچه بن‌بست قرار دارد (با رنگ آبی) قیمتی کمتر از مقدار پیش‌بینی شده دارد، در حالی که پارسل دوبر (با رنگ قرمز) به‌دلیل موقعیت بهتر، قیمتی بالاتر از مقدار برآوردشده توسط مدل داشته است. علاوه بر این موارد، در تعدادی از نمونه‌ها علت خطا از روی موقعیت مکانی به‌طور دقیق قابل تشخیص نبود

از ملک)، برای سه مدل یادگیری ماشین CatBoost، RF و XGBoost، حالت بدون دسته‌بندی ویژگی‌ها به‌طور محسوسی دقت بالاتری نسبت به دسته‌بندی‌های میانی و جامع نشان می‌دهد، در حالی که در مدل GBR، دسته‌بندی جامع توانسته است عملکردی قابل قبول و رقابتی ارائه کند. در مدل هدائیک نیز نتایج نشان می‌دهد که «دسته‌بندی میانی» از «دسته‌بندی جامع» کاراتر است و مقادیر شاخص‌های آن بسیار نزدیک به حالت بدون دسته‌بندی ویژگی‌ها قرار می‌گیرد.



شکل ۸. نمودار شاخص‌های ارزیابی در رویکرد دوم برای مدل‌های CatBoost و هدائیک

در مجموع می‌توان گفت مدل‌های یادگیری ماشین نسبت به نحوه دسته‌بندی ویژگی‌ها حساس‌تر از مدل هدائیک هستند و دسته‌بندی‌های پیچیده‌تر لزوماً به بهبود دقت منجر نمی‌شود؛ به‌گونه‌ای که در بسیاری از موارد، استفاده از مجموعه‌ی کامل ویژگی‌ها یا حداکثر دسته‌بندی میانی بهترین توازن بین دقت پیش‌بینی و سادگی مدل را فراهم کرده است.

و احتمالاً به ویژگی‌های ساختاری و ظاهری بنا مانند نوع نما، کیفیت ساخت، مصالح به‌کاررفته و ویژگی‌های اجتماعی محله مربوط می‌شود که داده‌ای از آن‌ها در این مطالعه در دسترس نبوده است.

به‌طور کلی، این نتایج نشان می‌دهد که بخشی از خطای مدل‌ها ناشی از نبود برخی متغیرهای مکانی و ساختاری مهم است و در پژوهش‌های بعدی می‌توان با افزودن متغیرهایی مانند بن‌بست بودن معبر، تعداد برپلاک و شاخص‌های دقیق‌تر کیفیت بنا و محیط، دقت برآورد ارزش املاک را بهبود داد.

۵- بحث و نتیجه‌گیری

برآورد ارزش املاک مسکونی به دلیل تأثیرپذیری هم‌زمان از ویژگی‌های ساختاری، موقعیت مکانی و شرایط محیطی، مسئله‌ای پیچیده و چندبعدی است. در این پژوهش، با تلفیق داده‌های ساختاری و شاخص‌های مکانی استخراج‌شده در محیط GIS، عملکرد چهار الگوریتم یادگیری ماشین شامل CatBoost، جنگل تصادفی، XGBoost و رگرسیون تقویت‌شده گرادیان در مقایسه با مدل هدانیک مبتنی بر OLS برای برآورد ارزش املاک مسکونی منطقه ۵ تهران ارزیابی شد. مدل‌سازی در دو رویکرد ارزش کل ملک و قیمت هر مترمربع انجام گرفت تا توان مدل‌ها در دو شیوه رایج ارزش‌گذاری املاک به‌صورت هم‌زمان بررسی شود.

نتایج نشان داد که مدل‌های یادگیری ماشین در مقایسه با مدل هدانیک، توان بالاتری در بازنمایی روابط غیرخطی میان ویژگی‌های ملک و قیمت دارند. در میان مدل‌های بررسی‌شده، CatBoost در اغلب سناریوها پایدارترین و دقیق‌ترین عملکرد را ارائه کرد و در سناریوی استفاده هم‌زمان از ویژگی‌های ساختاری و مکانی، بهترین نتیجه را در رویکرد ارزش کل ملک به دست آورد. عملکرد نزدیک مدل‌های جنگل تصادفی و XGBoost نیز نشان داد که روش‌های مبتنی بر درخت‌های تصمیم و یادگیری گروهی، به دلیل توانایی در مدل‌سازی برهم‌کنش‌های پیچیده میان متغیرها، گزینه‌هایی مناسب برای ارزش‌گذاری داده‌محور املاک محسوب می‌شوند. در مقابل، عملکرد ضعیف‌تر مدل هدانیک بیانگر محدودیت مدل‌های خطی در مواجهه با ساختار پیچیده و ناهمگن بازار مسکن است.

یکی از یافته‌های اصلی این پژوهش، نقش تعیین‌کننده شاخص‌های مکانی در افزایش دقت برآورد ارزش املاک است. مقایسه سناریوها نشان داد که حذف متغیرهای مکانی موجب افت محسوس عملکرد مدل‌ها می‌شود؛ بنابراین، اتکا به ویژگی‌های ساختاری ملک به‌تنهایی برای برآورد دقیق ارزش املاک کافی نیست. افزون بر شاخص‌های متداولی مانند فاصله از مراکز خدماتی، شبکه حمل‌ونقل، تراکم جمعیت و آسیب‌پذیری لرزه‌ای، نتایج نشان داد که متغیرهای خردمقیاس‌تر مانند عرض معبر دسترسی و جهت استقرار ملک نیز می‌توانند در بهبود دقت مدل‌ها مؤثر باشند. این یافته از آن جهت اهمیت دارد که چنین متغیرهایی در بسیاری از مطالعات پیشین به دلیل نبود داده آماده یا دشواری استخراج مکانی، کمتر مورد توجه قرار گرفته‌اند.

مقایسه دو رویکرد ارزش‌گذاری نشان داد که برآورد ارزش کل ملک در بیشتر سناریوها عملکردی اندکی بهتر از برآورد قیمت هر مترمربع ارائه می‌دهد. این تفاوت را می‌توان به ماهیت متغیر هدف نسبت داد؛ زیرا قیمت هر مترمربع بیش از ارزش کل ملک تحت تأثیر نوسانات محلی، ویژگی‌های خاص واحد مسکونی و تفاوت‌های کیفی مشاهده‌نشده قرار می‌گیرد. با این حال، نتایج هر دو رویکرد به‌طور کلی مؤید برتری مدل‌های یادگیری ماشین نسبت به مدل هدانیک و اهمیت استفاده هم‌زمان از داده‌های ساختاری و مکانی است. همچنین، بررسی سناریوهای مربوط به دسته‌بندی ویژگی‌ها نشان داد که تجمع گسترده ویژگی‌ها لزوماً به بهبود دقت منجر نمی‌شود و در بسیاری از موارد، حفظ جزئیات ویژگی‌ها یا استفاده از دسته‌بندی محدود، توازن مناسب‌تری میان دقت، سادگی و قابلیت تفسیر مدل فراهم می‌کند.

با وجود دستاوردهای پژوهش، نتایج آن باید با توجه به محدودیت‌های داده‌ای و مکانی تفسیر شود. داده‌های مورد استفاده به بازه زمانی دی و بهمن ۱۳۹۹ و محدوده منطقه ۵ تهران محدود بوده‌اند و قیمت‌های به‌کاررفته الزاماً معادل قیمت نهایی معاملات رسمی نیستند؛ از این‌رو، امکان تحلیل روندهای زمانی، اثرات فصلی و تعمیم مستقیم نتایج به سایر مناطق تهران یا شهرهای دیگر محدود است. همچنین، برخی متغیرهای اثرگذار مانند طبقه واحد، تعداد واحدهای ساختمان، کیفیت ساخت، وضعیت نورگیری، بن‌بست بودن معبر، تعداد برپلاک و ویژگی‌های اجتماعی-اقتصادی محله

در مدل لحاظ نشده‌اند و می‌توانند بخشی از خطاهای باقی‌مانده را توضیح دهند.

بر این اساس، در پژوهش‌های آینده پیشنهاد می‌شود از داده‌های زمانی گسترده‌تر و منابع متنوع‌تر، از جمله داده‌های رسمی معاملات، اطلاعات مشاوران املاک و داده‌های به‌روز سامانه‌های آنلاین مانند دیوار، برای ارزیابی پایداری زمانی مدل‌ها استفاده شود. همچنین، افزودن متغیرهای مکانی و

ساختاری تکمیلی، توسعه مدل‌های مکانی-زمانی، و تکرار مطالعه در سایر مناطق تهران و شهرهای مختلف می‌تواند به سنجش تعمیم‌پذیری چارچوب پیشنهادی کمک کند. در مجموع، یافته‌های این پژوهش نشان می‌دهد که تلفیق تحلیل‌های مکانی با الگوریتم‌های یادگیری ماشین می‌تواند مبنایی کارآمد برای توسعه سامانه‌های هوشمند و داده‌محور ارزش‌گذاری املاک مسکونی فراهم سازد.

مراجع

- [۱] Y. Yu, J. Lu, D. Shen, and B. Chen, "Research on real estate pricing methods based on data mining and machine learning," *Neural Comput & Applic*, vol. 33, no. 9, pp. 3925–3937, May 2021, doi: 10.1007/s00521-020-05469-3.
- [۲] W. K. O. Ho, B.-S. Tang, and S. W. Wong, "Predicting property prices with machine learning algorithms," *Journal of Property Research*, vol. 38, no. 1, pp. 48–70, Jan. 2021, doi: 10.1080/09599916.2020.1832558.
- [۳] Y. Kang *et al.*, "Understanding house price appreciation using multi-source big geo-data and machine learning," *Land Use Policy*, vol. 111, p. 104919, 2021.
- [4] R. Cellmer and R. Trojanek, "Towards increasing residential market transparency: Mapping local housing prices and dynamics," *ISPRS International Journal of Geo-Information*, vol. 9, no. 1, p. 2, 2019.
- [۵] S. Smirnov and V. Tlostanov, "Moscow Real Estate: Pricing Analysis Through the Prism of Statistics and Machine Learning," *Available at SSRN 3374436*, 2019, Accessed: Oct. 29, 2023. [Online]. Available: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3374436
- [۶] P. Ai, "Residential land value modelling," *Faculty of Geo-Information Science and Earth Observation of the University of Twente*. 2005. [An electronic resource]. Access mode: https://www.itc.nl/library/papers_2005/msc/upla/ai_ping.pdf (دما مراجعه: 29.12. 2017), free. Heading from the screen, 2005, Accessed: Jun. 17, 2024. [Online]. Available: <https://www.muthar-alomar.com/wp-content/uploads/2013/01/Residential-Land-Value.pdf>
- [۷] T. F. Sing, J. J. Yang, and S. M. Yu, "Boosted Tree Ensembles for Artificial Intelligence Based Automated Valuation Models (AI-AVM)," *J Real Estate Finan Econ*, vol. 65, no. 4, pp. 649–674, Nov. 2022, doi: 10.1007/s11146-021-09861-1.
- [۸] H. T. Sergoyan and G. V. Bezirganyan, "Automated Real Estate Valuation With Machine Learning: A Case Study on Apartment Sales in Yerevan," *Journal of Architectural and Engineering Research*, vol. 2, pp. 83–91, 2022.
- [۹] A. Guerra, A. Garces, and L. Chacon, "Apartment Pricing Prediction in the Panamanian Real Estate Market: A Machine Learning Perspective," *Available at SSRN 4493347*, 2023, Accessed: Jun. 02, 2024. [Online]. Available: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4493347
- [۱۰] C. Yiorkas and T. Dimopoulos, "Implementing GIS in real estate price prediction and mass valuation: the case study of Nicosia District," in *Fifth International Conference on Remote Sensing and Geoinformation of the Environment (RSCy2017)*, SPIE, 2017, pp. 112–128. Accessed: Oct. 29, 2023. [Online]. Available: <https://www.spiedigitallibrary.org/conference-proceedings-of-spie/10444/104440F/Implementing-GIS-in-real-estate-price-prediction-and-mass-valuation/10.1117/12.2280255.short>
- [۱۱] R. Cellmer, A. Cichulska, and M. Belej, "Spatial analysis of housing prices and market activity with the geographically weighted regression," *ISPRS International Journal of Geo-Information*, vol. 9, no. 6, p. 380, 2020.
- [۱۲] Y. Xiao, "BIG DATA FOR COMPREHENSIVE ANALYSIS OF REAL ESTATE MARKET," 2022, Accessed: Jun. 17, 2024. [Online]. Available: <https://scholarworks.lib.csusb.edu/etd/1596/>
- [۱۳] L. H. Choy and W. K. Ho, "The Use of Machine Learning in Real Estate Research," *Land*, vol. 12, no. 4, p. 740, 2023.
- [۱۴] T. D. Phan, "Housing price prediction using machine learning algorithms: The case of Melbourne city, Australia," in *2018 International conference on machine learning and data engineering (iCMLDE)*, IEEE, 2018, pp. 35–42. Accessed: Oct. 29, 2023. [Online]. Available: <https://ieeexplore.ieee.org/abstract/document/8614000/>
- [۱۵] F. Lorenz, J. Willwersch, M. Cajias, and F. Fuerst, "Interpretable machine learning for real estate market analysis," *Real Estate Economics*, vol. 51, no. 5, pp. 1178–1208, Sep. 2023, doi: 10.1111/1540-6229.12397.
- [۱۶] A. Jafari and M. R. Delavar, "Estimating the price of residential properties based on the optimal support vector machine," in *International Conference of GIScience: Basis and Trans/Interdisciplinary Applications, Mashhad, Iran*, 2021. Accessed: Nov. 09, 2024. [Online]. Available: <https://www.researchgate.net/profile/Ali-Jafari->

- 47/publication/361433368_Estimating_the_price_of_residential_properties_based_on_the_optimal_support_vector_machine/links/62b14230dc817901fc6ef931/Estimating-the-price-of-residential-properties-based-on-the-optimal-support-vector-machine.pdf
- [۱۷] K. Koutroumbas and S. Theodoridis, *Pattern Recognition*. Academic Press, 2008.
- [۱۸] M. V. A. William, S. Ramesh, R. Cep, S. K. Mahalingam, and M. Elangovan, "DPGWO Based Feature Selection Machine Learning Model for Prediction of Crack Dimensions in Steam Generator Tubes," *Applied Sciences*, vol. 13, no. 14, p. 8206, 2023.
- [۱۹] J. T. Hancock and T. M. Khoshgoftaar, "CatBoost for big data: an interdisciplinary review," *J Big Data*, vol. 7, no. 1, p. 94, Dec. 2020, doi: 10.1186/s40537-020-00369-8.
- [۲۰] X. Ding, W. Wang, Y. Zhang, and X. Zhong, "Machine Learning-based Models for House Price Prediction in Provincial Administrative Regions of China," in *2022 7th International Conference on Financial Innovation and Economic Development (ICFIED 2022)*, Atlantis Press, 2022, pp. 221–229. Accessed: Jun. 04, 2024. [Online]. Available: <https://www.atlantis-press.com/proceedings/icfied-22/125971699>
- [۲۱] S. Levantesi and G. Piscopo, "The importance of economic variables on London real estate market: A random forest approach," *Risks*, vol. 8, no. 4, p. 112, 2020.
- [۲۲] J. Cai, K. Xu, Y. Zhu, F. Hu, and L. Li, "Prediction and analysis of net ecosystem carbon exchange based on gradient boosting regression and random forest," *Applied Energy*, vol. 262, p. 114566, 2020.
- [۲۳] J. H. Friedman, "Greedy function approximation: a gradient boosting machine," *Annals of Statistics*, pp. 1189–1232, 2001.
- [۲۴] K. Baur, M. Rosenfelder, and B. Lutz, "Automated real estate valuation with machine learning models using property descriptions," *Expert Systems with Applications*, vol. 213, p. 119147, 2023.
- [۲۵] Q. Gao, V. Shi, C. Pettit, and H. Han, "Property valuation using machine learning algorithms on statistical areas in Greater Sydney, Australia," *Land Use Policy*, vol. 123, p. 106409, 2022.
- [۲۶] J. Avanijaa, "Prediction of house price using xgboost regression algorithm," *Turkish Journal of Computer and Mathematics Education (TURCOMAT)*, vol. 12, no. 2, pp. 2151–2155, 2021.
- [۲۷] T. Chen and C. Guestrin, "XGBoost: A Scalable Tree Boosting System," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, San Francisco California USA: ACM, Aug. 2016, pp. 785–794. doi: 10.1145/2939672.2939785.
- [۲۸] S. Rezaeian, H. Asgari, and B. Darvishi, "The study of determinants of rent housing in Ilam City based on hedonic spatial econometrics," *Journal of urban economics and management*, vol. 7, no. 26, pp. 17–30, 2019.
- [۲۹] H. Amirnejad, "The impact of Rasht Eynak Lagoon on housing price of the region by using hedonic pricing method," in *The impact of Rasht Eynak Lagoon on housing price of the region by using hedonic pricing method: Amirnejad, Hamid*, 2016. Accessed: Jun. 18, 2024. [Online]. Available: <https://www.zbw.eu/econis-archiv/bitstream/11159/602/1/iueam-v4n16p37-en.pdf>
- [۳۰] M. A. Milani and M. Haddadi, "Metro and Its Effect on Residential Property Prices (Tehran)," *Journal of Economic Research (Tahghighat-E-Eghtesadi)*, vol. 47, no. 4, pp. 79–96, 2013.
- [۳۱] سازمان فناوری اطلاعات و ارتباطات شهرداری تهران، امارنامه شهرداری تهران (سالنامه شهر تهران سال ۱۳۹۹). ۱۳۹۹.