



قیمت گذاری املاک مسکونی به کمک الگوریتم ترکیبی یادگیری عمیق-فازی

محمد سعدی مسگری^۲

زهرا مرادی^۱

تاریخ پذیرش مقاله: ۱۴۰۱/۱۲/۱۷

تاریخ دریافت مقاله: ۱۴۰۱/۰۳/۱۵

چکیده

اهمیت مسئله مسکن، لزوم برآورد قابل اعتماد از قیمت املاک را برای امر سیاست گذاری در این زمینه ایجاب می نماید. لذا ضروری است قیمت گذاری املاک با در نظر گرفتن صدها عامل تأثیرگذار بر آن، به صورتی کارآمد انجام شود. با توجه به ماهیت پیچیده بازار املاک در تحقیقات انجام شده از الگوریتم های متداول یادگیری عمیق مانند DNN، RNN، CNN و ... استفاده شده است، اما این الگوریتم ها در خصوص داده های جدولی چندان مناسب نمی باشند. از طرفی مدل های یادگیری عمیق موجود در قیمت گذاری ملک نیز کاملاً قطعی هستند و عدم قطعیت داده ها را لحاظ نمی کنند. در این مقاله سعی شده است که در به کارگیری روش های یادگیری عمیق به ساختار جدولی داده های املاک توجه شود. برای این منظور معماری عمیق جدید TabNet به کار گرفته شده است. این الگوریتم برخلاف سایر الگوریتم های متداول یادگیری عمیق داده های جدولی خام را بدون هیچ گونه پیش پردازشی دریافت می کند. در پژوهش حاضر، همچنین با استفاده از تکنیک های ترکیب موجود، منطق فازی با الگوریتم های یادگیری عمیق ترکیب شده است تا ضمن یادگیری سریع و دقیق تر مسائل پیچیده، بر کاستی های قطعی بودن مدل های یادگیری عمیق و در نظر نگرفتن عدم قطعیت ذاتی داده ها در این مدل ها غلبه شود. همچنین با به کارگیری سیستم اطلاعات مکانی (GIS) ارزیابی شفاف تری ارائه شد تا بصری سازی کامل الگوی مکانی ویژگی های ملک و همچنین ارتباط این ویژگی ها و قیمت گذاری تضمین و متغیرهای مکانی نیز در مدل ارزش گذاری لحاظ شوند. به منظور ارزیابی روش های پیشنهادی از داده های املاک منطقه ی پنج تهران استفاده شده است. ترتیب و اولویت بندی تأثیرگذاری ویژگی ها در قیمت گذاری املاک مسکونی تهران توسط الگوریتم TabNet نشان دهنده ی تأثیر قابل توجه عوامل مکانی است. به طوری که در این رتبه بندی پس از مساحت، دو ویژگی مکانی طول و عرض جغرافیایی به ترتیب رتبه ی دوم و سوم را به خود اختصاص می دهند. از این رو برای مجموعه داده ی تهران الگوریتم های TabNet، DNN، CNN، RNN، LSTM، خود رمزگذار و همچنین الگوریتم یادگیری ماشین XGBoost به کار گرفته شده و معیارهای ارزیابی RMSE، MAE و R^2 مقایسه شدند که بر اساس معیار R^2 ، با به کارگیری TabNet پنج درصد بهبود دقت حاصل شد. در نهایت RMSE الگوریتم ترکیبی FuzzyTabNet برای داده ی تهران نسبت به الگوریتم پایه ی TabNet ۴/۶۵ درصد کاهش یافت. همچنین شبکه ی خود رمزگذار فازی نیز نسبت به شبکه ی خود رمزگذار معمولی ۶/۵۲ درصد بهبود یافت.

واژه های کلیدی: قیمت گذاری املاک مسکونی، یادگیری عمیق، شبکه های عصبی عمیق، فازی، TabNet

۱- دانشجوی کارشناسی ارشد مهندسی عمران نقشه برداری، گرایش سیستم های اطلاعات جغرافیایی، دانشگاه خواجه نصیرالدین طوسی (نویسنده مسئول)
Zahra-moradi@email.kntu.ac.ir

۲- دانشیار دانشکده نقشه برداری، دانشگاه صنعتی خواجه نصیرالدین طوسی mesgari@kntu.ac.ir

۱- مقدمه

یادگیری عمیق همچنین از توانایی نگاشت غیرخطی زیادی برخوردار است که انعطاف پذیری بیشتری را در درک ماهیت پیچیده بازار املاک فراهم می کند. به علاوه این الگوریتم ها باعث کاهش نیاز به مهندسی ویژگی^۳ نیز خواهد شد که این مسئله در حال حاضر در یادگیری داده های جدولی^۴ جنبه ای کلیدی دارد. داده های جدولی یک نوع از انواع داده های موجود در جهان واقعی هستند که اطلاعات در قالب ترتیب ساختاریافته ای از سطرها و ستون ها جمع آوری شده است. از طرفی اگرچه الگوریتم های متعارف یادگیری عمیق موفقیت قابل توجهی را با تصاویر، متن و صدا نشان داده اند اما داده های جدولی که رایج ترین نوع داده در جهان واقعی هستند، چندان مناسب نبوده و روش های دیگر عملکرد بهتری داشته اند. معماری جدید یادگیری جدولی تفسیرپذیر توجه (TabNet)^۵ که خاص داده های جدولی است، برخلاف سایر الگوریتم های متداول یادگیری عمیق، داده های جدولی خام را بدون هیچ گونه پیش پردازشی دریافت می کند. به علاوه هم زمان با فرایند یادگیری انتخاب ویژگی را به صورت کاملاً تفسیرپذیر انجام می دهد و فرایند یادگیری بهتر انجام می شود، به این دلیل که کل ظرفیت یادگیری صرف ویژگی های با اهمیت تر می شود. با توجه به ارزیابی TabNet در سایر موضوعات انتظار می رود که این الگوریتم بر روی داده های رده بندی شده ای املاک نیز نتیجه ی قابل قبولی داشته باشد. از طرفی به علت عدم قطعیت ذاتی در فرایند قیمت گذاری املاک و مستغلات، به کارگیری منطق فازی در کنار یادگیری عمیق سودمند بوده و شفافیت ارزیابی را در قالب قوانین فازی معنادار، بیشتر می کند.

ترکیب این دو روش علاوه بر اینکه گزینه ی مناسبی برای تجزیه و تحلیل داده ی زیاد با سطح بالای پیچیدگی است، سیستم استخراج فازی را نیز بهبود می بخشد. چراکه استخراج قوانین فازی مبتنی بر دانش در روش های فازی مسئله ای زمان بر است و یادگیری عمیق می تواند از جفت داده های ورودی/

مسکن به عنوان سرمایه ای که سهم زیادی از بودجه خانوارها، هزینه ها و سرمایه گذاری ثابت ناخالص ملی را به خود اختصاص داده است، نقش زیادی در اشتغال و ارزش افزوده کشورها دارد (عبده کلامچی و دیگران، ۱۳۹۳). اهمیت روزافزون مسکن به لحاظ تأثیرات عمیق و قابل توجهی که بر ابعاد مختلف اجتماعی، سیاسی و اقتصادی کشورها می گذارد، بر کسی پوشیده نیست (Afonso et al., 2019)؛ بنابراین برآورد دقیق و قابل اعتماد قیمت به طور قطع امر سیاست گذاری در این زمینه را آسان می نماید. در شرایط مختلف ممکن است صدها عامل به صورت زیرمجموعه ای از عوامل ساختاری، مکانی و اجتماعی - اقتصادی بر قیمت املاک تأثیر بگذارد. بنابراین بایستی با در نظر گرفتن این عوامل، قیمت گذاری املاک به طور کارآمد انجام شود (Yiorkas & Dimopoulos, 2017). از جمله روش های معمول برای قیمت گذاری املاک، روش های مبتنی بر رگرسیون چندگانه (MRA)^۱ است که در بسیاری از پژوهش ها به کار گرفته می شود. پارامترها و عوامل دخیل در قیمت املاک بسیار متعدد و زیاد هستند. به علاوه، نوع و ذات این پارامترها بسیار مختلف و متنوع، و مقادیر آنها از نوع عدد صحیح، اعشاری، اسمی، باینری و غیره هستند. ضمناً ارتباط قیمت با این عوامل مؤثر غیرخطی می باشد، لذا استفاده از رویکردهای متداول رگرسیون دشوار و ناکارآمد خواهد بود. همچنین، به دلیل تأثیرپذیری قیمت ملک از ویژگی های مختلف پارامتری و غیر پارامتری با روابط آماری ناشناخته، استفاده از پیشرفته ترین الگوریتم ها برای تجزیه و تحلیل و مدل سازی ضروری است. از این رو به کارگیری روش های داده کاوی و یادگیری ماشین اهمیت بیشتری پیدا می کند. یادگیری عمیق^۲ به عنوان شاخه جدیدی از یادگیری ماشین، به دلیل توانایی در یادگیری خودکار و عدم محدودیت در خصوص الگوهای ذاتی داده ها و قابلیت کار با مدل های پیچیده، مورد توجه جوامع هوش مصنوعی قرار گرفته است.

1- Multiple Regression Analysis

2- Deep Learning

3- Feature Engineering

4- Tabular Data

5- Attentive Interpretable Tabular Learning

مهم‌ترین هدف این تحقیق طراحی و مدل‌سازی ساختار فازی برای مسئله و در نهایت پیاده‌سازی و ارزیابی مدل یادگیری عمیق فازی به‌عنوان یک روش ترکیبی بهبودیافته در قیمت‌گذاری املاک مسکونی می‌باشد. به‌علاوه با در نظر گرفتن موقعیت نسبتاً دقیق املاک و فاصله‌ی مکانی از مراکز مهم تأثیرگذار بر قیمت، بر روی جنبه‌ی مکانی مسئله نیز تمرکز شده است. در ادامه این مقاله، بعد از بیان مرور کارهای گذشته، روش‌های پیشنهادی توضیح داده شده و سپس نحوه پیاده‌سازی و ارزیابی روش‌ها بیان شده است. در انتها هم نتایج و پیشنهادهایی برای کارهای آتی در بخش نتیجه‌گیری ارائه شده است.

۲- مروری بر روش‌های قیمت‌گذاری املاک مسکونی

مدل‌سازی و پیش‌بینی قیمت ملک سابقه‌ای طولانی دارد (McCord et al., 2012). با توجه به اهمیت روزافزون مسکن و لزوم قیمت‌گذاری دقیق آن، در طول دهه‌های گذشته تحقیقات زیادی در این زمینه انجام شده است. از طرفی تعدد پارامترهای تأثیرگذار و پیچیدگی ارتباط قیمت با این پارامترها، انتخاب بهترین روش قیمت‌گذاری را با چالش روبه‌رو کرده است. مرور تحقیقات پیشین در این زمینه و روش‌های موردنظر این مقاله، در ادامه ارائه شده است.

۲-۱- الگوریتم‌های سنتی متداول

ازجمله روش‌های معمول برای قیمت‌گذاری املاک، روش‌های مبتنی بر رگرسیون چندگانه بود که در پژوهش‌های بسیاری استفاده شده است. در پژوهش بنجامین^۲ و همکاران (۲۰۰۴) تمرکز مطالعه بر مبانی قیمت‌گذاری املاک و تحلیل رگرسیون چندگانه است. به این صورت که تحلیل رگرسیون چندگانه توسط ارزیابان و دانشجویان مبتدی انجام شد و به فهم چگونگی استفاده از رویکرد مقایسه بازار و تحلیل رگرسیون چندگانه برای قیمت‌گذاری املاک انجامید. در پژوهش سوری و منیری^۳ (۲۰۱۱) ترکیب نسبی قیمت‌های

خروجی، این قوانین فازی را تولید نماید؛ بنابراین به‌نظر می‌رسد روش مناسب برای دستیابی به مزایای هر دو روش و حل مشکلات مربوط به هرکدام، ترکیب آن‌ها در یک سیستم یکپارچه است که ضمن یادگیری سریع و دقیق‌تر مسائل پیچیده، بر کاستی‌های قطعی بودن مدل‌های یادگیری عمیق و در نظر نگرفتن عدم قطعیت ذاتی داده‌ها در این مدل‌ها غلبه می‌کند. این روش ترکیبی در حوزه‌های دیگر مانند پیش‌بینی جریان ترافیک و طبقه‌بندی، به‌کار گرفته شده و عملکرد خوبی حاصل شده است (Chen et al., 2015; Chen et al., 2018; Talpur et al., 2020). به‌علاوه رایج‌ترین استنباط از دنیای املاک و مستغلات این است که ارزش املاک اساساً بر اساس موقعیت مکانی آن‌ها مشخص می‌شود. به گفته رابرت والنر مکان، مهم‌ترین عامل تعیین‌کننده ارزش املاک مسکونی در بازار مسکن است. یک ملک معمولاً به‌عنوان یک دارایی غیرمنقول شناخته می‌شود که در فضای اجتماعی و فیزیکی وجود دارد، جایی که مکان، ویژگی برتری است که بر ارزش آن‌ها تأثیر می‌گذارد (Yiorkas & Dimopoulos, 2017). با این وجود اکثر مدل‌های قیمت‌گذاری سنتی وجود همبستگی مکانی در قیمت‌ها را که منعکس‌کننده‌ی نقش ویژگی‌های مکانی است، در نظر نمی‌گیرند (Rambaldi & Rao, 2011). یورکاس و دیموپولوس^۱ (۲۰۱۷) استدلال می‌کنند که املاک مجاور اغلب ویژگی‌های ساختاری مشابهی دارند و ویژگی‌های مکانی را به اشتراک می‌گذارند، در نتیجه وابستگی مکانی ایجاد می‌شود. ویژگی‌های مکانی مانند فاصله تا نزدیک‌ترین امکانات، می‌تواند محیط ناملموس اطراف را توصیف کند (Kang et al., 2021) و بالعکس غفلت از این ویژگی‌ها باعث کاهش دقت قیمت‌گذاری خواهد شد (McCord et al., 2012). هدف این مقاله، مدل‌سازی مسئله‌ی قیمت‌گذاری املاک مسکونی در قالب تعدادی از الگوریتم‌های متداول یادگیری عمیق است. از دیگر اهداف تحقیق حاضر، توجه کامل بر ساختار داده‌های مسئله و به‌کارگیری الگوریتم جدید TabNet در کنار دیگر الگوریتم‌های یادگیری عمیق می‌باشد.

2- Benjamin

3- Souri and Moniri

1- Yiorkas and Dimopoulos

این دو روش از دقت خوبی برخوردار هستند. به علاوه آن‌ها دریافتند که شبکه عصبی المان می‌تواند دقیق‌تر و سریع‌تر از روش‌های دیگر پیش‌بینی کند. پژوهش تابالز^۷ و همکاران (۲۰۱۳) نیز نشان داد که در صورت وجود اطلاعات آماری کافی و مجموعه گسترده‌ای از داده‌ها، استفاده از شبکه ANN مناسب است.

۲-۳- شبکه‌های عصبی فازی

لی و ژانگ^۸ (۲۰۰۷) قیمت‌های املاک و مستغلات را در راستای تأیید اهمیت استدلال و یادگیری فازی با استفاده از شبکه عصبی فازی پیش‌بینی کردند. در واقع آن‌ها برای پیش‌بینی دقیق‌تر، منطق فازی را با شبکه عصبی ترکیب کردند تا یک شبکه عصبی فازی که توانایی استدلال و یادگیری فازی دارد، ایجاد نمایند. تحقیقات آن‌ها نشان داد از آنجایی که عوامل نامطمئن و غیرمعمول زیادی در برآورد قیمت مؤثر هستند، شبکه عصبی به‌تنهایی قادر به مدیریت این اطلاعات فازی نیست، لذا شبکه عصبی فازی می‌تواند بهتر از رویکردهای شبکه عصبی سنتی عمل کند. مطالعه‌ی گان^۹ و همکاران (۲۰۰۸) نیز استفاده از شبکه عصبی فازی را در زمینه‌ی ارزیابی انبوه املاک و مستغلات بررسی کرده و نتایج قابل قبولی حاصل شد. با این حال این مطالعه از حجم نمونه‌ی بسیار کمی استفاده کرده است. آن‌ها ضمن استفاده از این روش برای یک مجموعه داده‌ی بزرگ‌تر، تعریفی منحصر به فرد از همسایگی ارائه دادند که به بهبود نتایج قیمت‌گذاری منجر شد (Guan et al., 2008). با این وجود مدل‌های شبکه عصبی برای همگرایی، بسته به پیچیدگی مسئله به هزاران یا میلیون‌ها تکرار نیاز داشتند و نمی‌توانستند مسائل تشخیص الگوهای پیچیده را به‌طور مؤثر و در یک مدت‌زمان محاسباتی معقول مدل‌سازی کنند (Rafiei & Adeli, 2016). با این حال وجود مدل‌های شبکه عصبی برای همگرایی بسته به پیچیدگی مسئله به هزاران یا میلیون‌ها

فروش مکرر املاک به‌عنوان شاخص قیمت در نظر گرفته شد. در نهایت تکنیک‌های استاندارد تحلیل رگرسیون برای تخمین این شاخص مورد استفاده قرار گرفت. تحلیل رگرسیون ضمن اینکه از اطلاعات مربوط به شاخص قیمت دوره‌های قبلی موجود در قیمت‌های فروش در دوره‌های بعدی استفاده کرده است، خطاهای استاندارد شاخص تخمینی را به‌آسانی محاسبه و حذف کرده و به این طریق کارآمدتر از روش‌های دیگر به نظر می‌رسید.

۲-۲- شبکه‌های عصبی مصنوعی

به تدریج روش‌های شبکه عصبی مصنوعی (ANN)^۱ به‌عنوان جایگزینی مناسب‌تر پیشنهاد شدند چراکه عملکرد MRA با افزایش حجم داده و متغیرهای وابسته کاهش می‌یافت (Nghiep & Al, 2001). به کارگیری شبکه‌های عصبی در قیمت‌گذاری ملک قابل توجه بوده و موفقیت‌های بسیاری حاصل شده است (Guan et al., 2014). مطالعه‌ی بورس^۲ (۱۹۹۱) یکی از اولین مواردی است که از روش شبکه عصبی برای پیش‌بینی قیمت ملک استفاده کرد. نگویت و کریپس^۳ (۲۰۰۱) عملکرد پیش‌بینی شبکه‌های عصبی مصنوعی و تحلیل رگرسیون چندگانه را برای فروش مسکن تک خانواده مقایسه کردند. این مقایسه بین دو مجموعه داده با اندازه‌ی نمونه‌ی متفاوت انجام شد که ANN در مجموعه داده با حجم نمونه‌ی بزرگ‌تر، بهتر از MRA عمل کرد. به این ترتیب آن‌ها دریافتند که وقتی داده‌های کافی برای آموزش در دسترس باشد، شبکه‌های عصبی می‌توانند عملکرد بهتری نسبت به رگرسیون‌های خطی متعدد داشته باشند. خیاولونگ و مینگ^۴ (۲۰۱۰) برای پیش‌بینی قیمت املاک و مستغلات از شبکه‌های عصبی پس انتشار خطا^۵ و شبکه عصبی المان (ENN)^۶ استفاده کرده و تأیید می‌کنند که

1- Artificial Neural Network

2- Borst

3- Nghiep and Cripps

4- Xiaolong and Ming

5- Back Propagation

6- Elman Neural Network

7- Tabales

8- Li and Zhang

9- Guan

اجاره آپارتمان در ژاپن با روش کریجینگ^۷ که یک روش رگرسیون فضایی است مقایسه کرد تا تفاوت دقت این دو روش را با یک مجموعه داده‌ی بزرگ نشان دهد. این مطالعه نشان داد که با افزایش حجم نمونه، دقت پیش‌بینی برون نمونه‌ای^۸ DNN به Kriging نزدیک شده و در حجم نمونه‌ی^۹ ۱۰، برابر می‌شود. علاوه بر این در مواردی که انحراف از میانه‌ی قیمت ملک زیاد است، دقت پیش‌بینی DNN بهتر خواهد بود. چراکه این روش غیرخطی بودن فرم تابع را نیز در نظر می‌گیرد.

در مطالعه‌ی چن^۹ و همکاران (۲۰۱۷) برای پیش‌بینی قیمت ملک از شبکه عصبی حافظه طولانی کوتاه‌مدت (LSTM)^{۱۰} ساده که شاخه‌ای از شبکه عصبی بازگشتی (RNN)^{۱۱} محسوب می‌شود و توالی زمان را نیز در قیمت‌گذاری املاک لحاظ می‌کند، استفاده شد. نتیجه نشان داد که مدل LSTM با توجه به پیش‌بینی‌های سری زمانی خود در زمینه‌ی قیمت ملک عملکرد خوبی دارد. همچنین برای مطالعه‌ی بیشتر و بهبود دقت مدل، از دو نوع دیگر این روش استفاده شد. بانه و یو^{۱۲} (۲۰۱۷) ثابت کردند که در پیش‌بینی قیمت املاک مسکونی اگرچه مدت‌زمان آموزش DNN طولانی‌تر است اما در مقایسه با LSTM دقت بهتری دارد. مطالعه‌ی کیم^{۱۳} و همکاران (۲۰۲۰) با استفاده از داده‌های ۱۶۷ ماه و به‌کارگیری LSTM ساده، به پیش‌بینی دقیق و قابل‌اعتمادی دست‌یافت. در پژوهشی دیگر در زمینه‌ی بازار املاک یک رویکرد جدید مبتنی بر شبکه عصبی پیش‌بینی (CNN)^{۱۴} به‌کار گرفته شد که در مقایسه با رویکردهای سنتی برآوردی منطقی و مناسب ارائه داد (Piao et al., 2019).

تکرار نیاز داشتند و نمی‌توانستند مسائل تشخیص الگوهای پیچیده را به‌طور مؤثر و در یک مدت‌زمان محاسباتی معقول مدل‌سازی کنند (Rafiei & Adeli, 2016).

۲-۴- الگوریتم‌های یادگیری عمیق

امروزه الگوریتم‌های یادگیری عمیق تخمین پیشرفته و قوی‌تری را ارائه می‌دهند. پژوهش رافعی و عادل^۱ (۲۰۱۶) برای برآورد قیمت املاک در مرحله طراحی یا شروع ساخت‌وساز، یک مدل جدید و جامع ارائه داد. در این پژوهش با ادغام مبتکرانه‌ی ماشین بولتزمن محدود^۲ باور عمیق^۳ و الگوریتم ژنتیک یک ساختار داده‌ی مؤثر ارائه شد که ضمن در نظر گرفتن تعداد زیادی شاخص اقتصادی، تغییرات فصلی و وابسته به زمان متغیرها را نیز در بر گرفت. از طرفی با در نظر گرفتن تعداد زیادی از متغیرها، آموزش مدل به مجموعه داده‌های آموزشی بزرگ و منابع محاسباتی قابل‌توجهی نیاز داشت و قسمت دوم این پژوهش با ارائه‌ی راهبردی هوشمندانه، بر کاهش ابعاد داده تمرکز داشت که در نتیجه‌ی آن برآوردی منطقی حاصل شد. در مطالعه‌ی دیگر برای ارزیابی هوشمندانه‌ی ارزش املاک، یک شبکه عصبی عمیق (DNN)^۴ ساختاریافته ارائه شد که در مقایسه با متدهای مبتنی بر رگرسیون و حتی DNN با اتصال کامل، کارآمدتر و دقیق‌تر بود (Xu & Gade, 2017). به‌طور دقیق‌تر این پژوهش برای طراحی یک شبکه عصبی عمیق ساختاریافته یک روش سیستماتیک ارائه می‌نماید و اتصال ساختاریافته‌ی گره‌ها، منجر به کاهش قابل‌توجه گره‌ها و اتصالات آن‌ها خواهد شد که به تبع نیاز به داده نیز کمتر خواهد شد. این پژوهش برای نشان دادن اثربخشی رویکرد پیشنهادی، به مطالعه‌ی موردی قیمت‌گذاری املاک واقعی در شهرهای کوچک پرداخت. مطالعه‌ی سیا و شیروی^۵ (۲۰۱۹) دقت شبکه عصبی عمیق را برای پیش‌بینی قیمت

7- Kriging

8- Out of sample

9- Chen

10- Long Short-Term Memory

11- Recurrent Neural Network

12- Bae and Yu

13- Kim

14- Convolutional Neural Networks

15- Ayan and Eken

1- Rafiei & Adeli

2- Restricted Boltzmann Machines (RBM)

3- Deep Belief

4- Deep Neural Network

5- Node

6- Seya and Shiroy

تفسیرپذیری محلی برای نشان دادن اهمیت ویژگی‌ها و اینکه چطور با یکدیگر ترکیب می‌شوند، استفاده می‌شود و تفسیرپذیری سرتاسری سهم هر ویژگی در مدل آموزش‌دیده را مشخص می‌کند. در این مدل یک معماری منفرد یادگیری عمیق نیز برای مهندسی ویژگی به کار گرفته می‌شود (Arak & Pfister, 2021). این الگوریتم جدید با ساختار داده‌های این مقاله متناسب است و با معماری متعارف خود به کار گرفته می‌شود. بنابراین نیازی به تغییر در ساختار الگوریتم برای داده‌های مسئله نیست. همچنین از آنجایی که معماری جدید TabNet خاص داده‌های جدولی است، لزوم انجام پیش‌پردازش‌های لازم بر روی داده‌های مسئله را به حداقل می‌رساند. الگوریتم DNN نیز با معماری متعارف خود به کار گرفته می‌شود و نیازی به تغییر در ساختار الگوریتم برای داده‌های مسئله نیست. الگوریتم‌های LSTM و RNN برای داده‌هایی به صورت سری زمانی طراحی شده‌اند اما برای ارزیابی و مقایسه‌ی جامع‌تر الگوریتم‌های یادگیری عمیق، استفاده می‌شوند. بنابراین به کارگیری این شبکه برای داده‌های مسئله، نیاز به تغییرات جزئی در ساختار داده‌ها دارد. هم چنین اگرچه ساختار CNN در استخراج ویژگی به خوبی عمل می‌کند، اما به ندرت در داده‌های جدولی استفاده می‌شود، زیرا ترتیب ویژگی‌ها در این نوع داده مهم نیست. این در حالی است که ارتباط بین پیکسل‌های همسایه در تصاویر حائز اهمیت است. باین وجود اخیراً CNN یک بعدی بهترین عملکرد را با داده‌های جدولی به دست آورده است (Shwartz-Ziv & Armon, 2022). بنابراین با وجود اینکه الگوریتم CNN برای داده‌هایی با ساختار تصویر، سیگنال و ... طراحی شده است، در مسئله‌ی حاضر CNN یک بعدی با کمی تغییر در ابعاد و شکل داده‌های جدولی مسئله، به کار گرفته می‌شود. به این ترتیب با توجه به نقطه‌ی قوت این الگوریتم در استخراج ویژگی‌ها، میزان اثرگذاری پارامترهای مؤثر بر قیمت نیز لحاظ می‌شود. رمزگذارهای خودکار نیز غالباً در کاربردهای بدون نظارت استفاده می‌شوند. در مسئله‌ی پیش رو برای کاربرد یادگیری تحت نظارت و

استانبول را از سال ۲۰۰۷ تا ۲۰۱۹ با استفاده از یک مدل خود رمزگذار^۱ بررسی کرد. نتایج تجربی آن پژوهش نشان داد که برای تشخیص مؤثر حساب‌های مسکن می‌توان از مدل خود رمزگذار LSTM که ترکیبی از دو مفهوم قوی LSTM و خود رمزگذار است، استفاده کرد.

۳- روش پیشنهادی

روش پیشنهاد شده در این مقاله بر پایه الگوریتم‌های یادگیری عمیق توسعه داده شده است. ساختار پیشنهادی این روش به صورت ترکیب بهبودیافته‌ی الگوریتم‌های متداول یادگیری عمیق و سیستم استنتاج فازی می‌باشد. قبل از مدل‌سازی مسئله با الگوریتم‌های پیشنهادی، بایستی آماده‌سازی و پیش‌پردازش‌های مورد نیاز بر روی داده‌های ورودی انجام پذیرد. این پیش‌پردازش‌ها شامل شناسایی و حذف داده‌های پرت، شناسایی و محاسبه‌ی مقادیر نامعلوم و بررسی همبستگی میان ویژگی‌ها است. در صورت زیاد بودن پارامترهای مسئله، حذف ویژگی‌های اضافی و غیر مهم با روش‌های انتخاب ویژگی انجام می‌شود. سپس شبکه‌های عصبی عمیق متداول، LSTM، RNN، DNN، CNN و AutoEncoder در کنار معماری جدید TabNet، به عنوان روش‌های حل مسئله‌ی قیمت‌گذاری و همچنین الگوریتم‌های پایه‌ی روش ترکیبی به کار گرفته می‌شوند. TabNet یک مدل انتها به انتها^۲ و خاص داده‌های جدولی است که در مقایسه با سایر مدل‌های پیشرفته عملکرد خوبی نشان داده است. این مدل، داده‌های خام را بدون هیچ پیش‌پردازشی می‌گیرد و به کمک الگوریتم‌های بهینه‌سازی مبتنی بر گرادیان کاهشی^۳ آموزش می‌دهد. TabNet از توجه متوالی^۴ برای انتخاب اینکه هر ویژگی در چه مرحله تصمیم‌گیری به کار گرفته شود، استفاده می‌کند که به تفسیرپذیری مدل منجر خواهد شد. سطح تفسیرپذیری TabNet به دو سطح محلی و سرتاسری تقسیم می‌شود.

- 1- Autoencoder
- 2- End to End
- 3- Gradient Descent
- 4- Sequential Attention

برای هر متغیر ورودی (m) می‌باشد. خروجی هر گره به صورت رابطه (۱) است (Guan et al., 2014):

$$O_{ij} = \mu_{ij}(x_i), \text{ for } i = 1, \dots, n, \quad j = 1, \dots, m \quad \text{رابطه (۱)}$$

μ_{ij} تابع عضویت z_m برای ورودی x_i می‌باشد. مقدار بهینه‌ی پارامترهای توابع عضویت به‌عنوان بخشی از پارامترهای شبکه طی فرایند آموزش و توسط الگوریتم پس انتشار^۷ به‌دست می‌آید. به‌این ترتیب در مسئله‌ی پیش رو، به‌ازای هر یک از پارامترهای مؤثر بر قیمت، توابع تعیین درجه‌ی عضویت اعمال شده و داده‌های ورودی به ورودی‌های فازی شده تبدیل می‌شوند. در قسمت طراحی قوانین فازی همه‌ی قوانین ممکن، توسط الگوریتم یادگیری عمیق به‌صورت خودکار ایجاد می‌شود. هر گره در این لایه یک گره ثابت است که تنها به یک قانون فازی TSK ارتباط دارد. درجه‌ی فعال هر قاعده‌ی k تحت‌عنوان خروجی این لایه از ضرب تمام ورودی‌های آن طبق رابطه (۲) به‌دست می‌آید (Guan et al., 2014):

$$O_k = r_k = \mu_{1k}(x_1) \times \mu_{2k}(x_2) \dots \mu_{mk}(x_m) \quad \text{رابطه (۲)}$$

for $k = 1, \dots, R$

تعداد قواعد R به‌صورت m^n است که کلیه ترکیبات ممکن از حالت‌های مختلف ورودی را شامل می‌شود. قوانین ایجادشده توسط الگوریتم لایه‌ی عمیق در قسمت نرمال‌سازی، نرمال می‌شود. هر گره در این لایه یک گره دایره‌ای است و نرمال‌سازی هر قاعده‌ی ورودی را به‌صورت رابطه (۳) محاسبه می‌کند (Jang, 1993):

$$O_k = \bar{r}_k = \frac{r_k}{r_1 + r_2 + \dots + r_R}, \quad k = 1, \dots, R \quad \text{رابطه (۳)}$$

بنابراین در مقاله حاضر، با تقسیم هر قاعده بر مجموع قواعد ۱ تا R ، قواعد قسمت فازی نرمال می‌شوند. در نهایت فازی‌زدایی به‌صورت یک لایه‌ی تک گره، خروجی‌های تمام قوانین لایه‌ی قبلی را به‌صورت رابطه (۴) ادغام می‌کند:

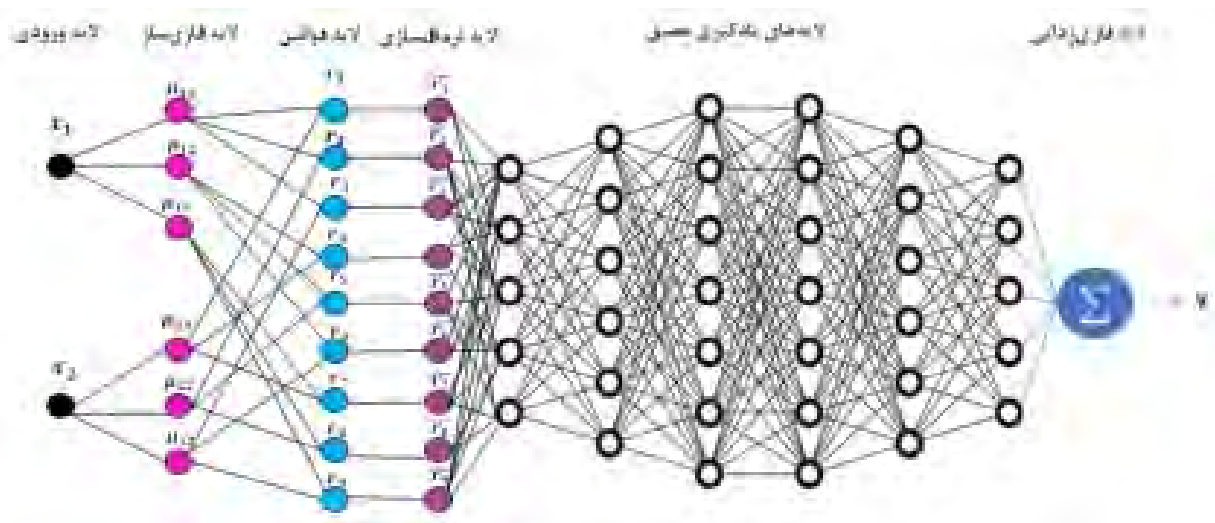
$$O = \text{SUM}(\bar{r}_k), \quad k = 1, \dots, R \quad \text{رابطه (۴)}$$

برآورد قیمت، یک‌لایه‌ی برآوردگر به آخرین لایه پنهان^۱ در خود رمزگذار متصل می‌شود. به‌این ترتیب با توجه به قابلیت حذف نویز این الگوریتم، هدف قیمت‌گذاری همراه با حذف نویز داده‌ها انجام می‌شود. در ادامه، ساختار پیشنهادی برای روش ترکیبی بهبودیافته یادگیری عمیق و سیستم استنتاج فازی شرح داده خواهد شد. در این روش ترکیبی، سیستم فازی تاکاگی-سوگنو-کانگ (TSK)^۲ در نظر گرفته می‌شود (Yazdanbakhsh & Dick, 2019). از طرفی برای ترکیب الگوریتم یادگیری عمیق و سیستم استنتاج فازی در قالب یک سیستم یکپارچه با ساختار هماهنگ از ترکیب هیبریدی^۳ این دو روش استفاده می‌شود که پارامترهای فازی توسط الگوریتم یادگیری عمیق به دست می‌آید (Yazdanbakhsh & Dick, 2019). از آنجایی‌که در چنین سیستمی الگوریتم‌های یادگیری عمیق معمولی را مستقیماً نمی‌توان اعمال کرد، بنابراین در ساختار این الگوریتم ترکیبی به طراحی لایه‌های جدید فازی می‌پردازیم. در ادامه به شرح مدل‌سازی ساختار فازی پرداخته می‌شود. سپس نحوه ترکیب الگوریتم‌های پایه‌ی یادگیری عمیق با لایه‌های ساختار فازی، در قالب یک سیستم یکپارچه بیان می‌شود.

۳-۱- مدل‌سازی ساختار فازی

مدل‌سازی ساختار فازی چهار لایه‌ی فازی ساز^۴، قوانین فازی، نرمال‌سازی و فازی‌زدایی^۵ را شامل می‌شود. مدل‌سازی فازی ساز، به‌عنوان اولین لایه‌ی قسمت فازی است. هرکدام از گره‌های این لایه، یک گره تطبیقی با یک تابع عضویت^۶ مناسب است که با ورودی گره مطابقت دارد و درجه‌ی عضویت آن ورودی را مشخص می‌کند. بنابراین، این لایه شامل N گره است که به‌صورت $m \times n$ یا حاصل‌ضرب اندازه‌ی ورودی (n) و تعداد توابع عضویت

- 1- Hidden Layer
- 2- Takagi-Sugeno-Kang
- 3- Hybrid
- 4- Fuzzification
- 5- Defuzzification
- 6- Membership Function



نگاره ۱: ساختار ترکیب یادگیری عمیق متوالی و سیستم فازی

به جای ورودی‌های معمولی، ورودی‌های فازی شده منجر به قیمت‌گذاری خواهد شد. هم‌چنین با طراحی قوانین فازی شیوهی جدیدی از محاسبات در ساختار یادگیری عمیق به کار گرفته می‌شود.



نگاره ۲: ساختار ترکیب یادگیری عمیق و سیستم فازی به روش Stacking

۴- پیاده‌سازی و ارزیابی

به منظور پیاده‌سازی و ارزیابی روش‌های پیشنهادی در این مقاله، ۳۰ درصد داده‌ها به عنوان داده‌ی ارزیابی^۳ و ۱۰ درصد مابقی داده‌ها به عنوان داده‌ی اعتبارسنجی^۴ در نظر گرفته شده است. باقیمانده‌ی داده‌ها نیز به عنوان داده‌ی آموزشی^۵ هستند. هم‌چنین از روش اعتبارسنجی متقابل K-

3- Test Data
 4- Validation Data
 5- Train data

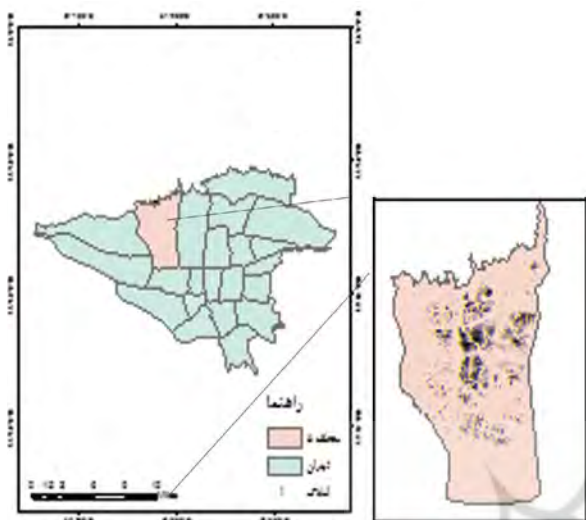
۲-۳- مدل‌سازی الگوریتم ترکیبی یادگیری عمیق-فازی برای اجرای سیستم فازی در ساختار الگوریتم‌های یادگیری عمیق متوالی مانند LSTM، RNN، CNN، DNN و خود رمزگذار، تنها می‌بایست لایه‌های فازی‌ساز، قوانین و نرمال‌سازی طراحی شده در ابتدای این الگوریتم ترکیبی قرار گیرد و در نهایت بعد از لایه‌های شبکه عصبی عمیق، لایه فازی‌زدایی به کار گرفته شود (نگاره ۱).

TabNet نیز با همین روش قابل ترکیب است؛ اما با توجه به اینکه قرارگیری لایه‌های فازی در ساختار الگوریتم‌های عمیق انتها به انتها کمی دشوار است، بایستی لایه‌های فازی در ساختار زیر کلاس‌بندی^۱ لایه‌ها و کلاس‌های مدل TabNet، قبل از دیگر لایه‌ها قرار گیرد. به این ترتیب هر لایه، وزن زیر لایه‌های خود را ردیابی می‌کند. هم‌چنین برای این الگوریتم انتها به انتها، از روش ترکیب تعمیم انباشته^۲ یا انباشتن را نیز می‌توان به کار برد (Ribeiro & dos Santos Coelho, 2020). در این روش لایه‌های فازی و لایه‌های شبکه‌ی عمیق در قالب دو مدل مجزا و به صورت دوسطحی، در یک مدل واحد ترکیب می‌شوند (نگاره ۲).

به این ترتیب در مسئله‌ی پیش رو سیستم استنتاج فازی در ساختار شبکه‌های عمیق به کار گرفته می‌شود. بنابراین

1- Subclassing
 2- Stacked Generalization

بر تمایل و تقاضای خریداران به خرید املاک و به تبع آن بر قیمت املاک، مورد توجه است.



نگاره ۳: منطقی مطالعه به همراه نمایش پراکندگی داده‌های املاک

۲-۴- پیش پردازش

در این پژوهش با استفاده از ArcGIS مکان و در واقع طول و عرض جغرافیایی با اطلاعات ساختاری ملک مرتبط شده است تا مدل قیمت‌گذاری به همراه جزئیات مکانی ایجاد شود. سپس برای در نظر گرفتن ویژگی‌های مکانی املاک دسترسی به پارک، بیمارستان، مراکز خرید و ایستگاه‌های مترو به صورت نقشه‌های فاصله به دست آمد. برای ساده‌تر شدن روند حل مسئله، فاصله‌ی اقلیدسی^۸ در نظر گرفته شده است. همپوشانی داده موقعیت مکانی املاک با لایه دسترسی به پارک، فاصله هر ملک از پارک به عنوان یک ویژگی به جداول اطلاعات املاک اضافه شد. دقیقاً مشابه با این کار برای لایه‌های دسترسی به مرکز خرید، دسترسی به مترو و دسترسی به بیمارستان انجام شده است. این دسترسی‌ها به عنوان نزدیکی املاک به امکانات و خدمات مهم و با نظر کارشناسی مشاورین املاک شهر مورد مطالعه انتخاب شده است. در نتیجه جدول اطلاعات توصیفی املاک شامل این پارامترها نیز می‌شود.

8- Euclidean Distance

لایه^۱ با $K=10$ برای ارزیابی دقیق‌تر مدل‌ها استفاده شده است (Fan et al., 2015). برای ارزیابی هر روش از معیارهای آماری متداول به کار گرفته شده در بسیاری از پژوهش‌های قیمت‌گذاری املاک، استفاده شده است. این معیارها شامل معیارهای ارزیابی خطای جذر میانگین مربعات^۲ (RMSE)، میانگین خطای مطلق^۳ (MAE) و ضریب تعیین^۴ (R^2) می‌شوند. برای آماده‌سازی داده‌ها و ایجاد متغیرهای مکانی از محیط نرم‌افزار ArcGIS استفاده شده است. پیاده‌سازی الگوریتم‌ها در محیط برنامه‌نویسی Python به کار گرفته شده است. ساختار یکپارچه‌ی لایه‌های در شبکه‌های عصبی عمیق کتابخانه‌ی آماده‌ی ندارد. لذا در این مقاله مدل‌سازی قسمت فازی به صورت لایه‌های سفارشی انجام شده است. هم چنین با توجه به چالش‌های مربوط به ترکیب این الگوریتم با فازی که در قسمت ۳-۲ شرح داده شد، از ساختار کلاسه‌بندی شده‌ی TabNet استفاده شده است.

۱-۴- داده مورد استفاده

داده‌های ساختاری ۲۷۹۶ ملک مسکونی واقع در منطقه ۵ تهران به همراه موقعیتشان در فرودین ۱۴۰۰ از سایت "دو دوتا"^۵ جمع‌آوری شده است (نگاره ۳). اطلاعات املاک شامل قیمت، طول و عرض جغرافیایی، مساحت، تعداد پارکینگ، سال ساخت، تعداد اتاق و وجود یا عدم وجود حیاط، انباری، بالکن، آسانسور و امکانات ویژه می‌شود. برای سادگی به جای قیمت مترمربع، قیمت کل املاک در نظر گرفته شده است. داده‌های مکانی شامل نقشه‌ی پارک‌ها، مراکز خرید، بیمارستان، ایستگاه‌های مترو به کمک سرویس گوگل ارث^۶ و گوگل‌مپ^۷ تهیه شده است. در این نقشه‌ها، امکانات اطراف و خارج از مرز منطقه‌ی ۵ نیز لحاظ شده است. داده‌ی ترافیک نیز از سازمان حمل‌ونقل و ترافیک تهیه شده است. ترافیک به عنوان یک عامل تأثیرگذار

1- K-Fold Cross Validation

2- Root Mean Square Error

3- Mean Absolute Error

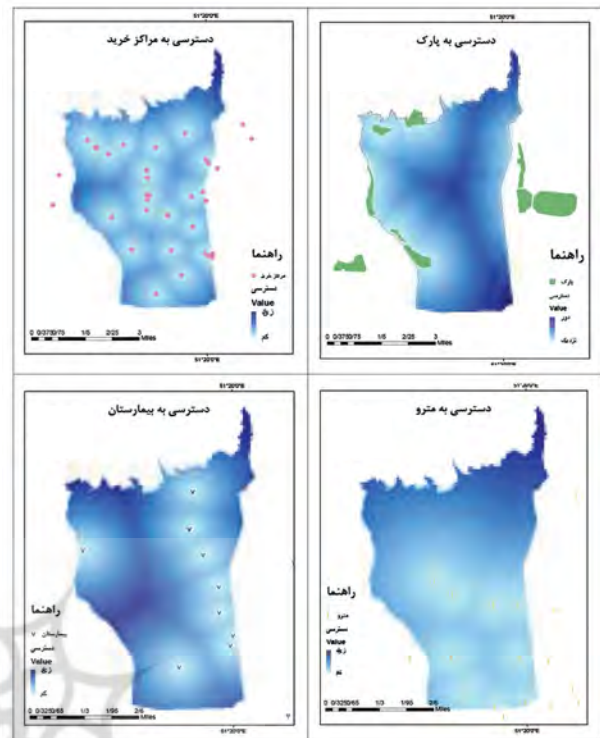
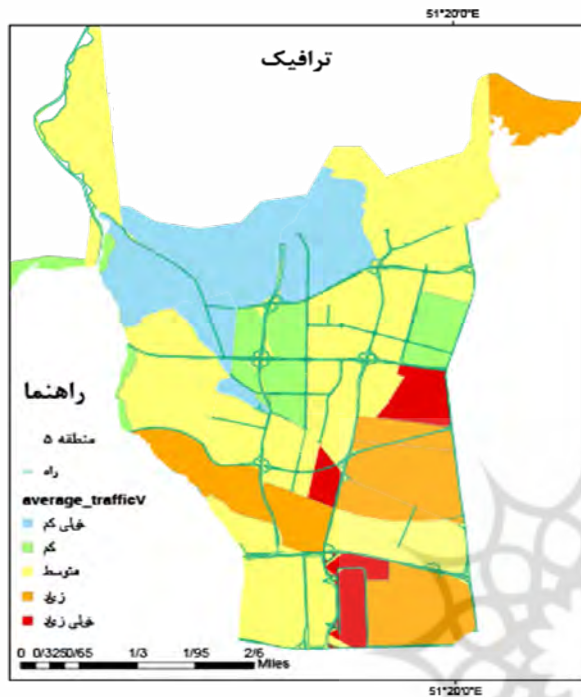
4- Coefficient of determination (R^2)

5- <https://dodota.com>

6- Google Earth Engine

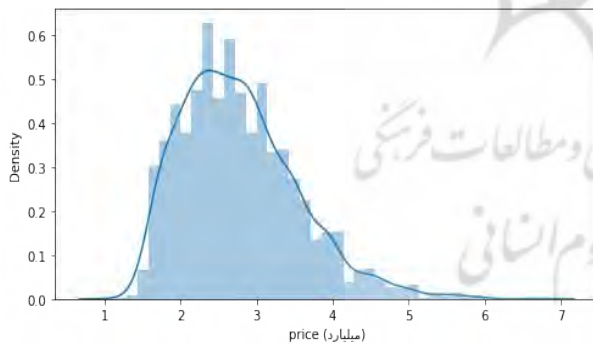
7- Google Map

متغیر پیوسته‌ی قیمت املاک است را نشان می‌دهد. طبق این نمودار قیمت املاک در حدود ۱/۱۵ تا ۶/۷ میلیارد می‌باشد.



نگاره ۴: نقشه‌های دسترسی برای املاک

نگاره ۵: نقشه‌ی ترافیک حوزه‌های ترافیک منطقه‌ی ۵ تهران



نگاره ۶: نمودار توزیع متغیر پیوسته‌ی قیمت املاک

نگاره ۷ ماتریس همبستگی^۲ بین ویژگی‌های ساختاری و مکانی املاک که در بخش ۴-۲ و ۴-۱ شرح داده شد و متغیر هدف (قیمت) را نشان می‌دهد. این ماتریس از نتایج ضریب همبستگی پیرسون^۳ تشکیل شده است، به طوری که مقادیر در بازه منفی یک تا مثبت یک قرار دارند.

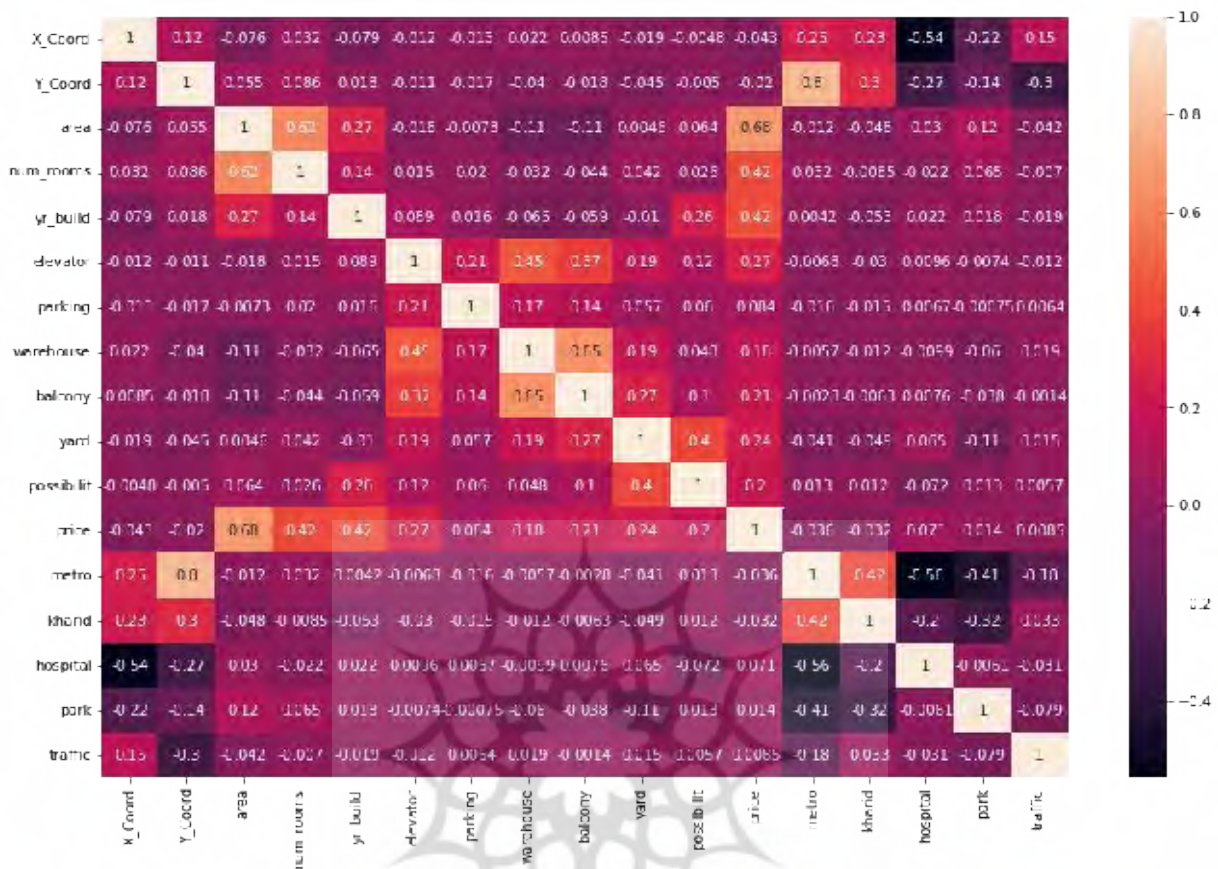
2- Correlation Matrix

3- Pearson Correlation Coefficient

اطلاعات ترافیک نیز شامل حوزه‌های ترافیک و راه‌ها است. ابتدا برای هر پلی‌گون مربوط به حوزه‌های ترافیک، راه‌های عبوری از آن شناسایی شده است. سپس ترافیک برای هر حوزه به صورت میانگین حجم ترافیک راه‌های عبوری به دست آمده است. حجم ترافیک هر راه در واقع میزان سفرهای تولید شده است. در نهایت با همپوشانی داده موقعیت مکانی املاک با این لایه ترافیک، ترافیک هر حوزه به املاک واقع در آن منتسب شده و به عنوان یکی دیگر از ویژگی‌های مکانی املاک در نظر گرفته می‌شود (نگاره ۵).

در این پژوهش برای سادگی در مدل‌سازی، ترافیک که یک معیار متغیر است به صورت ویژگی ثابت املاک، فرض شده است. پس از آماده‌سازی ورودی، پیش‌پردازش‌های لازم در محیط Python انجام شد. اطلاعات ۱۵ ملک به دلیل پرت بودن با روش فاصله‌ی کوک^۱ شناسایی و حذف شدند. وجود مقادیر نامعلوم در ویژگی‌های املاک بررسی شد. نگاره ۶ نمودار توزیع قیمت که در واقع توزیع کلی

1- Cook's Distance



نگاره ۷: ماتریس همبستگی ویژگی‌های ملک

شبکه در مسائل مشابه و اضافه و کم کردن لایه‌ها، به دست آمد. به منظور تنظیم^۱ ابر پارامترهای شبکه‌ها متناسب با داده‌های مسئله، از الگوریتم بهینه‌سازی بیزین^۲ استفاده شده است (Victoria & Maragatham, 2021). همان‌طور که گفته شد برای DNN نیازی به تغییر در ساختار ورودی‌ها نبوده و ابعاد لایه‌ی ورودی نشان‌دهنده‌ی ۱۶ پارامتر مؤثر بر قیمت ملک می‌باشد. معماری مناسب شبکه نیز شامل لایه ورودی، سه لایه پنهان تماماً متصل^۳ و لایه خروجی می‌شود. برای به‌کارگیری شبکه RNN و LSTM برای داده‌های مسئله، نیاز به تغییرات جزئی در ساختار داده‌ها وجود دارد، به طوری که بایستی شکل داده‌های ورودی تغییر کرده

مثبت نشان‌دهنده‌ی ارتباط مستقیم و مقادیر منفی رابطه‌ی معکوس دو ویژگی را نشان می‌دهد. هرچه مقادیر از صفر فاصله می‌گیرند، همبستگی بیشتر می‌شود. بنابراین در این ماتریس میزان همبستگی متوسط تا قوی بین ویژگی‌های ملک با متغیر هدف قابل مشاهده است. از طرفی همبستگی ویژگی‌های ملک با یکدیگر کم و نسبتاً مناسب است. این بدان معنی است که متغیرها یا ویژگی‌های انتخاب‌شده برای برآورد قیمت املاک مناسب و گویای تغییرات قیمت املاک می‌باشند.

۴-۳- پیاده‌سازی و مقایسه‌ی الگوریتم‌های یادگیری عمیق

ابتدا برای پیاده‌سازی الگوریتم‌ها، داده‌های تهران به سه قسمت آموزشی، اعتبارسنجی و ارزیابی تفکیک شده است. معماری شبکه‌های عمیق با آزمون و خطای معماری‌های

1- Tuning

2- Hyper Parameter

3- Bayesian Optimization

4- Fully Connected

لایه‌های پیچشی در لایه‌های سطح^۶ بار دیگر به صورت خطی تغییر شکل یافته و به بردار تبدیل شدند. همچنین در پژوهش پیش رو برای کاربرد یادگیری تحت نظارت و برآورد قیمت، یک لایه‌ی برآوردگر به آخرین لایه پنهان در خود رمزگذار متصل شده است. پس از حذف نویز ورودی‌ها در لایه‌های خود رمزگذار، برآورد قیمت در این لایه انجام می‌شود. هر دو بخش رمزگذار^۷ و رمزگشای^۸ این شبکه از لایه‌های تماماً متصل تشکیل شده‌اند. در این شبکه نیز از لایه نرمال‌سازی دسته استفاده شده است. الگوریتم جدید TabNet بدون نیاز به تغییر در ساختار داده‌ها یا شبکه، با ساختار متعارف خود به کار گرفته شده است. از طرفی ابر پارامترهای این الگوریتم یادگیری عمیق با دیگر الگوریتم‌ها متفاوت است که در جدول ۱ به تعدادی از آن‌ها پرداخته شده است. باقی پارامترها به صورت پیش فرض انتخاب شده‌اند. به منظور مقایسه‌ی قدرت پیش‌بینی الگوریتم‌های یادگیری عمیق نتایج سه معیار ارزیابی بر روی مجموعه داده‌ی ارزیابی در جدول ۲ ارائه شده است. نتایج الگوریتم افزایش گرادیان اکستریم^۹ (XGBoost) نیز به عنوان یکی از متداول‌ترین الگوریتم‌های یادگیری ماشین برای داده‌های جدولی، لحاظ شده است. این الگوریتم در سال ۲۰۱۶ توسط چن و گواسترین معرفی شد (Chen & Guestrin, 2016). در کاربردهای قیمت‌گذاری املاک، XGBoost در مقایسه با الگوریتم‌های سنتی‌تر مانند رگرسیون خطی نتایج بهتری داشته است (Peng et al., 2019). بنابراین به عنوان نماینده‌ی قابل قبول سایر الگوریتم‌ها با الگوریتم‌های یادگیری عمیق مقایسه شده است. مقادیر کمتر RMSE و MAE ضمن در نظر گرفتن بازه‌ی مقادیر قیمت املاک، بهتر هستند. اگرچه برای معیار R² مقادیر بیشتر از ۵۰ درصد قابل قبول است، ولیکن افزایش این مقدار و نزدیکی آن به ۱۰۰ درصد، موجب افزایش تناسب مدل‌ها خواهد شد.

و با ساختار ورودی در این شبکه‌ها منطبق شود. بنابراین به داده‌های جدولی املاک، بعد سوم زمان اضافه شد که تنها اطلاعات املاک در یک‌زمان ثابت را شامل می‌شود. به عبارتی بعد سوم تنها شامل یک لایه بوده و اطلاعات املاک در زمان ثابت فروردین ۱۴۰۰ را نشان می‌دهد. بنابراین برای به‌کارگیری RNN و LSTM در مسئله، ساختار ورودی تغییر کرده و ابعاد لایه‌ی ورودی نشان‌دهنده‌ی ۱۶ پارامتر مؤثر بر قیمت ملک و یک بعد زمان می‌باشد. الگوریتم RNN بعد از لایه‌ی ورودی، از دو لایه RNN و یک لایه‌ی تماماً متصل تشکیل شده است و همانطور که گفته شد این معماری بهینه با آزمون و خطای معماری‌های شبکه در مسائل مشابه و اضافه و کم کردن لایه‌ها، به دست آمده است. به منظور حفظ مقادیر خروجی لایه‌های RNN در محدوده نرمال، بعد از این لایه‌ها از لایه‌ی نرمال‌سازی دسته^۱ نیز استفاده شده است.

معماری LSTM به عنوان خانواده‌ی شبکه‌ی RNN، مشابه با همان شبکه در نظر گرفته شده است (Zaremba et al., 2014). با این تفاوت که لایه‌های LSTM جایگزین لایه‌های RNN شده‌اند. در مسئله‌ی حاضر CNN یک بعدی با کمی تغییر در ابعاد و شکل داده‌های جدولی مسئله، به کار گرفته می‌شود. بنابراین شکل داده‌ها با اضافه شدن بعد سوم تغییر یافته و داده‌های ورودی با ارتفاع ۱ و عرض ۱۶ که شامل پارامترهای تأثیرگذار بر قیمت هستند، وارد شبکه می‌شوند. معماری مناسب این الگوریتم شامل دو لایه‌ی پیچشی^۲ و یک لایه‌ی ادغام^۳ با عملگر بیشینه است. همچنین از یک لایه‌ی انصراف^۴ استفاده شده است که با حذف تصادفی تعدادی از گره‌ها در هر تکرار آموزش، آموزش مستقل گره‌ها را بهبود بخشیده و از برازش بیش از حد^۵ جلوگیری می‌نماید. در انتها پیش از لایه‌های تماماً متصل، ویژگی‌های استخراج‌شده‌ی

6- Flatten Layer

7- Encoder

8- Decoder

9- Extreme Gradient Boosting (XGBoost)

1- Batch Normalization

2- Convolutional Layer

3- Pooling Layer

4- Dropout Layer

5- Overfitting

جدول ۱: ابر پارامترهای TabNet

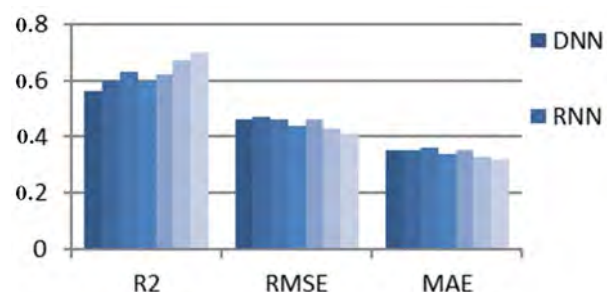
مقدار بهینه	تعریف	ابر پارامتر
۲۴	عمق لایه‌ی پیش‌بینی تصمیم/ مقادیر بزرگ‌تر ظرفیت بیشتری را به مدل می‌دهد.	n_d
۲۴	عمق به‌کارگیری مکانیسم توجه برای هر ماسک/ معمولاً n_d=n_a مقدار خوبی است.	n_a
۶	تعداد مراحل معماری	n_steps
۲	این ضریب برای استفاده‌ی مجدد ویژگی در ماسک‌ها می‌باشد. مقدار نزدیک به ۱ باعث می‌شود که انتخاب ماسک کمترین همبستگی را بین لایه‌ها داشته باشد.	Gamma
۲	تعداد لایه‌های واحدهای خطی دروازه‌ای مستقل در هر مرحله	n_independent
مقدار بهینه	تعریف	ابر پارامتر
۲	تعداد واحدهای خطی دروازه‌ای مشترک در هر مرحله	n_shared
۰/۴	تکانه برای نرمال‌سازی دسته‌ای	Momentum
۱e-۴	ضریب اتلاف پراکندگی اضافی است. هرچه این ضریب بزرگ‌تر باشد، مدل از نظر انتخاب ویژگی محدودتر خواهد بود. بسته به پیچیدگی مسئله، کاهش این مقدار کمک‌کننده می‌باشد.	lambda_sparse
"entmax"	تابع ماسک کردن برای انتخاب ویژگی‌ها	mask_type
۵۰	تعداد دوره‌های متوالی بدون بهبود برای انجام توقف اولیه	Patience
۵۱۲	تعداد نمونه در هر دسته	batch_size

با توجه به نتایج جدول کلیه معیارها نشان‌دهنده برتری مدل TabNet نسبت به دیگر مدل‌های یادگیری عمیق است. براساس نتایج به‌دست‌آمده TabNet در بین شبکه‌های عمیق خطای کمتر و در نتیجه کارایی بیشتری در پیش‌بینی قیمت املاک داشته است. مقدار R^2 و نزدیکی مقادیر پیش‌بینی‌شده‌ی TabNet و مقادیر واقعی قیمت در نگاره ۹، تأییدی بر این ادعاست. البته با توجه به اینکه اختلاف بین R^2 این الگوریتم‌ها کم است، به‌طور دقیق قابل‌شناسایی و تحلیل نمی‌باشد، اما به‌طور کل ماحصل این بصری‌سازی برای کل داده، در مقدار این معیار نشان داده شده است.

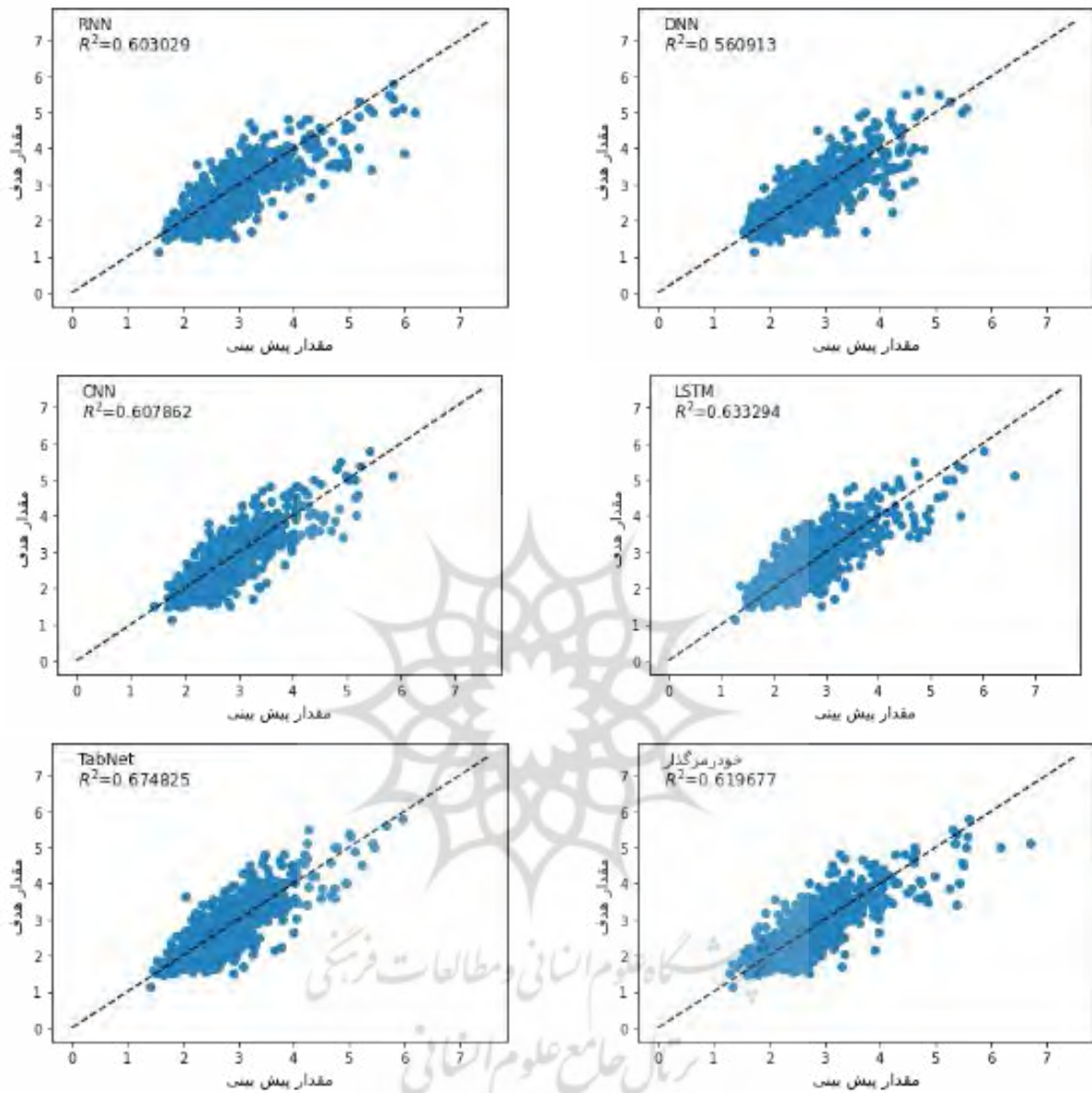
بر اساس نتایج نشان داده‌شده در جدول ۲ در این ارزیابی LSTM بر اساس معیار R^2 به‌عنوان دومین الگوریتم یادگیری عمیق مناسب شناخته‌شده است. با وجود به‌کارگیری داده‌های جدولی مقطعی همچنان پیچیدگی بلوک‌های LSTM موجب بهبود دقت بلوک‌های RNN معمولی شده است. همچنین دو بخش رمزگذار و رمزگشای شبکه‌ی خود رمزگذار در مقایسه با یک شبکه‌ی عمیق معمولی دقت را افزایش داده است. از طرفی تعداد کم داده‌های املاک تهران موجب شده

جدول ۲: مقایسه‌ی الگوریتم‌های یادگیری عمیق (مجموعه داده تهران)

R^2	MAE	RMSE	
۰/۵۶	۰/۳۵	۰/۴۶	DNN
۰/۶۰	۰/۳۵	۰/۴۷	RNN
۰/۶۳	۰/۳۶	۰/۴۶	LSTM
۰/۶۰	۰/۳۴	۰/۴۴	CNN
۰/۶۲	۰/۳۵	۰/۴۶	خود رمزگذار
۰/۶۸	۰/۳۳	۰/۴۳	TabNet
۰/۷۰	۰/۳۲	۰/۴۱	Xgboost



نگاره ۸: نمودار مقایسه‌ی روش‌های یادگیری عمیق و XGBoost بر اساس معیارهای ارزیابی



نگاره ۹: نمودار مقدار واقعی و پیش‌بینی‌شده‌ی الگوریتم‌ها

موردی کمتر مورد توجه است. همچنین داده‌های پرت املاک غالباً بیانگر مقدارهای غیرمعمول بوده و کمتر حاصل مشاهدات مخدوش است. از طرفی معیار MAE اثر این داده‌های پرت را در برآورد مدل از بین می‌برد. لذا به‌نظر می‌رسد که همچنان استفاده از معیار RMSE به نسبت MAE مفیدتر باشد. کمتر بون معیار MAE به نسبت RMSE نیز به همین دلیل قابل توجه است. در ادامه برتری الگوریتم XGBoost نسبت به TabNet و همچنین سایر شبکه‌های

است که برخلاف انتظار، از لحاظ معیار R^2 شبکه‌ی CNN در مواجهه با داده‌های جدولی ملک برتری قابل توجهی را کسب نکنند. با این حال این روش از لحاظ دو معیار دیگر در جایگاه دومین الگوریتم یادگیری عمیق مناسب قرار گرفته است. همان‌طور که قبلاً اشاره شده است، در ارزیابی نهایی به دلیل اهمیت کاهش متوسط خطای پیش‌بینی در مسئله، معیار خطای RMSE ملاک قرار گرفته است و معیار R^2 به دلیل تأثیرپذیری از انحراف معیار و توجه به نتیجه‌ی

لایه‌ی قوانین شده است که تعداد گره‌های ثابت آن به تعداد قواعد یا به عبارتی تعداد ویژگی‌ها به توان تعداد توابع عضویت، می‌باشد. در این لایه بخش مقدم قوانین به صورت همه‌ی ترکیبات ممکن از ورودی‌های فازی شده‌ی لایه‌ی قبل، ایجاد می‌شود. بخش نتیجه‌ی قوانین نیز مقادیر غیرفازی قیمت املاک است. لایه بعد و یا لایه نرمال‌سازی نیز شامل گره‌های دایره‌ای است که خروجی‌های لایه‌ی قبل را نرمال کرده است. در انتها لایه‌ی فازی‌زدایی با تک گره، قوانین نرمال‌شده‌ای را که چه به صورت مستقیم یا غیرمستقیم به دست آمده است، به صورت مجموع وزنی ادغام می‌کند. منظور از محاسبه‌ی مستقیم قوانین لایه‌ی نرمال‌سازی این است که خروجی لایه‌ی نرمال‌سازی بدون وجود لایه‌ی دیگر، مستقیماً وارد لایه‌ی فازی‌زدایی شود. در حالت غیرمستقیم نیز بعد از لایه‌ی نرمال‌سازی قوانین، تعدادی از لایه‌های شبکه مانند لایه تماماً متصل، پیچشی و ... قرار داشته است و سپس لایه‌ی فازی‌زدایی به کار گرفته می‌شود. در نهایت وزن هر قاعده به عنوان پارامترهای این لایه در نظر گرفته شده است. باید توجه داشت که برای عمیق‌تر شدن قسمت فازی از لایه‌های تماماً متصل پس از لایه فازی‌زدایی استفاده شده است. ابتدا به روش ترکیب انباشتن، لایه‌های فازی و TabNet بهینه‌ی قسمت قبل، در قالب دو مدل مجزا و به صورت دوسطحی، در یک مدل واحد ترکیب شده‌اند. سپس برای ارزیابی دقیق‌تر، الگوریتم ترکیبی به روش زیر کلاس‌بندی نیز ایجاد شد. برای این ترکیب لایه‌های فازی‌ساز، قوانین و نرمال‌سازی در ساختار زیر کلاس‌بندی لایه‌ها و کلاس‌های مدل TabNet، قبل از دیگر لایه‌ها قرار گرفت. لایه‌ی فازی‌زدایی نیز پس از همه‌ی لایه و قبل از لایه‌ی خروجی TabNet به کار گرفته شد. پارامترهای n_d ، n_a ، n_steps و γ به ترتیب ۲، ۳۶، ۳۶ و ۱/۲ انتخاب شده است. در ادامه به ترکیب شبکه‌ی خود رمزگذار با لایه‌های فازی پرداخته شد. برای اجرای این سیستم فازی در بخش رمزگذار این شبکه، لایه‌های فازی به روش متوالی به کار گرفته شد. در ابتدا برای کاهش

عمیق به کاررفته، بر این احتمال تأکید می‌کند که کم بودن تعداد داده‌ها مانع از دستیابی به نتیجه‌ی مطلوب و ایده‌آل روش‌های یادگیری عمیق شده است. در نهایت با در نظر گرفتن نتایج کلیه روش‌ها به نظر می‌رسد که قیمت املاک در تهران به بسیاری از پارامترهای دیگر همچون نمای ساختمان، تعداد واحدهای آپارتمانی موجود در هر طبقه و در کل ساختمان، سیستم‌های گرمایشی و سرمایشی و موضوع پرستیژ، زیبایی و کلاس اجتماعی کوجه‌ها و خیابان‌ها ربط دارد. اضافه بر این، نیازهای شدید شخص فروشنده به فروش ملک و یا نیاز خاص خریدار به خرید عجولانه، گاهی باعث تغییرات شدید در قیمت هر ملک خاص می‌شود. این عوامل باعث کاهش دقت در نتایج همه‌ی الگوریتم‌ها می‌شود.

۴-۴- پیاده‌سازی الگوریتم ترکیبی یادگیری عمیق فازی و مقایسه با الگوریتم‌های پایه

برای پیاده‌سازی الگوریتم ترکیبی ابتدا با توجه به نتایج قسمت قبل، TabNet به عنوان بهترین الگوریتم یادگیری عمیق، انتخاب شده است. سپس لایه‌های فازی به صورت لایه‌های سفارشی^۱ ایجاد شد. در ساختار قسمت فازی بعد از لایه‌ی ورودی لایه‌ی فازی‌ساز قرار گرفته است. تعداد گره‌های تطبیقی این لایه به تعداد ویژگی‌ها و همین‌طور تعداد توابع عضویت بستگی دارد. با توجه به افزایش قواعد و به تبع آن مشکل افزایش محاسبات، دو تابع عضویت به ازای هر ویژگی در نظر گرفته شده است. به این ترتیب برای داده‌های تهران با توجه به ۱۶ ویژگی مسئله، تعداد گره‌ها ۳۲ بوده است. بر اساس آزمون و خطا، تابع عضویت گوسی با دو پارامتر μ ، σ به عنوان تابع عضویت مناسب انتخاب شد که خروجی هر گره با توجه به این توابع عضویت، درجه‌ی عضویت هریک از ویژگی‌ها را به مجموعه‌ی فازی مربوطه نشان می‌دهد. در نهایت خروجی این لایه شامل ورودی‌های فازی شده می‌شود. خروجی‌های این لایه وارد

است. از آنجایی که ساختار فازی به‌تنهایی نتیجه‌ی بسیار دقیقی نداشته است عدم بهبود معیار سوم ارزیابی، قابل توجه به‌نظر می‌رسد. با توجه به نتایج، استفاده از زیر کلاس‌بندی برای ایجاد این ساختار ترکیبی، به افزایش این معیار منجر می‌شود. اگرچه بازهم معیار موردنظر در مقایسه با الگوریتم پایه بهبودی نداشته است، اما با این حال در رقابت با سایر الگوریتم‌های یادگیری عمیق مقدار قابل قبولی دارد. البته این ترکیب از لحاظ دو معیار RMSE و MAE بهبودی نداشته است. بازهم بر اساس دو معیار RMSE و MAE، ترکیب الگوریتم خود رمزگذار و فازی موجب بهبود دقت و برتری این الگوریتم نسبت به همه‌ی الگوریتم‌های یادگیری عمیق شده است. البته باید توجه داشت که به‌طور کل ترکیب TabNet و فازی از این ترکیب بهتر عمل کرده است. همچنان این ترکیب از لحاظ معیار سوم بهبودی نداشته است. با این حال این معیار مقدار قابل قبولی داشته است. بنابراین با وجود اینکه ترکیب سیستم استنتاج فازی و الگوریتم‌های یادگیری عمیق همواره و بر اساس تمام معیارهای ارزیابی بهبود قابل توجهی نداشته است، با این حال منجر به بهبود قابل قبول دو معیار RMSE و MAE شده است. همان‌طور که اشاره شد در ارزیابی نهایی مسئله به دلیل اهمیت کاهش متوسط خطای پیش‌بینی، معیار خطای RMSE ملاک قرار گرفته است و معیار R^2 به دلیل تأثیرپذیری از انحراف معیار و توجه به نتیجه‌ی موردی کمتر مورد توجه است. در نتیجه با توجه به این ارزیابی در مسائلی که با عدم قطعیت و لزوم استفاده از منطق فازی همراه است، بهتر است از الگوریتم ترکیبی یادگیری عمیق- فازی استفاده شود.

۴-۵- تحلیل اهمیت پارامترهای مکانی

ترتیب و اولویت‌بندی تأثیرگذاری ویژگی‌ها در قیمت‌گذاری املاک مسکونی تهران توسط الگوریتم TabNet که دقت بهتری داشته، مشخص شده است (نگاره ۱۰). این ویژگی‌ها در بخش ۴-۲ و ۴-۱ شرح داده شد.

قوانین و حجم محاسبات فازی از روش کاهش بعد تحلیل مؤلفه‌های اصلی (PCA) ^۱ استفاده شد (Khalid et al., 2014) و با آزمون و خطا ویژگی‌ها ادغام و تعداد هشت ویژگی استخراج شد. به‌عنوان روش جایگزین، قرار دادن یک‌لایه تماماً متصل با تعداد گره کمتر، پیش از لایه‌ی فازی‌ساز به افزایش دقت منجر شد.

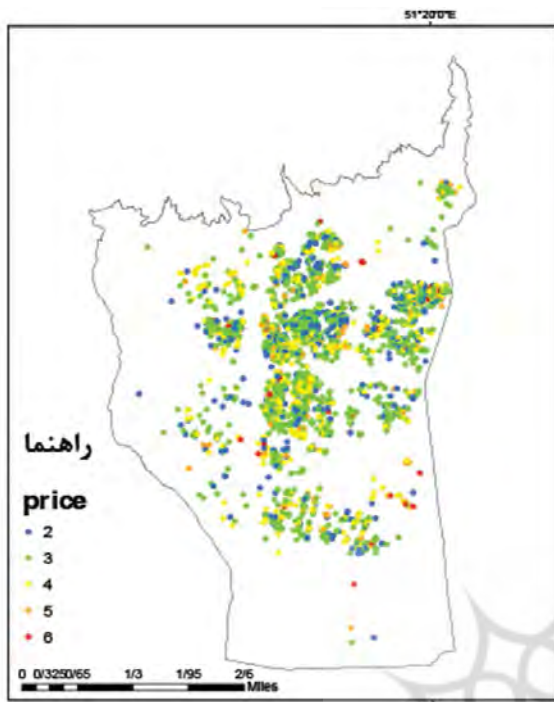
در نهایت به‌جز قسمت‌های فازی، ساختار بهینه‌ی الگوریتم ترکیبی خود رمزگذار فازی مشابه الگوریتم پایه‌ی خود رمزگذار است. برای ارزیابی روش ترکیبی یادگیری عمیق-فازی برای داده‌های تهران، در جدول ۳ به مقایسه‌ی نتایج الگوریتم‌های ترکیبی با شبکه‌های پایه پرداخته شده است.

جدول ۳: مقایسه‌ی الگوریتم ترکیبی یادگیری عمیق-فازی با

الگوریتم‌های پایه

R ²	MAE	RMSE	
۰/۶۸	۰/۳۳	۰/۴۳	TabNet
۰/۵۵	۰/۳۲	۰/۴۱	TabNet (ترکیب به روش انباشتن)
۰/۶۲	۰/۳۵	۰/۴۷	TabNet (ترکیب به روش زیر کلاس‌بندی)
۰/۶۲	۰/۳۵	۰/۴۶	خود رمزگذار
۰/۵۹	۰/۳۳	۰/۴۳	خود رمزگذار فازی

با توجه به جدول ۳ و بر اساس دو معیار RMSE و MAE، ترکیب الگوریتم TabNet به روش انباشتن موجب بهبود دقت شده است. همچنین مقایسه‌ی این دو معیار بهبودیافته و نتایج جدول ۳ نشان‌دهنده‌ی برتری این روش ترکیبی نسبت به همه‌ی روش‌های یادگیری عمیق و برابری با XGBoost می‌باشد. با این حال این روش در معیار سوم نتیجه‌ی قابل قبولی نداشته است. همان‌طور که پیش‌ازین گفته شد در این ترکیب سیستم استنتاج فازی و TabNet به‌صورت دو مدل مجزا در نظر گرفته شده‌اند؛ بنابراین دقت هر دو مدل پایه بر روی فرا مدل نهایی تأثیرگذار



نگاره ۱۱: نقشه منطقه‌ای املاک قیمت‌گذاری شده

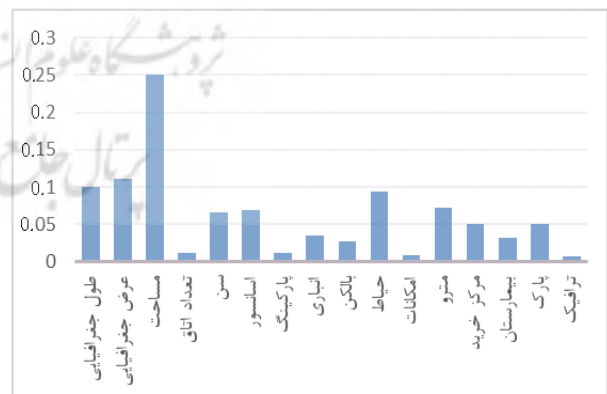
به روش TabNet

۵- نتیجه‌گیری و پیشنهادات

با توجه به اهمیت روزافزون مسکن و لزوم قیمت‌گذاری دقیق آن، در این مقاله به ساختار داده‌ها توجه شده است و ضمن به‌کارگیری الگوریتم‌های متداول یادگیری عمیق از معماری جدید TabNet که خاص داده‌های جدولی است، استفاده شده است که از جمله تفاوت‌های این مقاله با سایر مطالعات صورت گرفته، محسوب می‌شود. در این مقاله سعی شده است با در نظر گرفتن موقعیت نسبتاً دقیق املاک و فاصله‌ی اقلیدسی از مراکز مهم تأثیرگذار بر قیمت مانند بیمارستان، مراکز خرید، مترو و پارک، بر روی جنبه‌ی مکانی مسئله نیز تمرکز شود. در نهایت برای مدل‌سازی عدم قطعیت ذاتی پارامترهای تأثیرگذار بر قیمت املاک به‌عنوان یکی از اهداف اصلی، منطق فازی با الگوریتم‌های یادگیری عمیق TabNet و خود رمزگذار ترکیب شده است. برای یک ترکیب یکپارچه، در ساختار این الگوریتم ترکیبی به طراحی لایه‌های جدید فازی پرداخته شده است.

الگوریتم TabNet با توجه به سطح تفسیرپذیری سرتاسری خود، سهم هر ویژگی در مدل آموزش‌دیده را مشخص کرده است. نتایج نشان‌دهنده‌ی آن است که اکثر متغیرها معنادار هستند. با وجود اینکه بعضی از عوامل ساختاری بیشتر از عوامل مکانی قیمت واحد مسکونی را تحت تأثیر قرار می‌دهند، اما به‌طور کل عوامل مکانی تأثیر قابل توجهی داشته‌اند. مساحت به‌عنوان یک ویژگی ساختاری، مهم‌ترین ویژگی تأثیرگذار بر قیمت است. در رتبه‌ی دوم و سوم ویژگی‌های تأثیرگذار، طول و عرض جغرافیایی قرار گرفته است که به‌عنوان مهم‌ترین ویژگی مکانی اثرگذار بر قیمت ملک قابل مشاهده‌اند.

اساساً طول و عرض جغرافیایی بیانگر معیار محلات و تیپ و پرستیژ مکان‌های مختلف شهر و کلاس اجتماعی خیابان‌ها و محله‌های مختلف شهری است که مشخصاً عامل تأثیرگذاری در قیمت محسوب می‌شود. همچنین در میان ویژگی‌های مکانی دسترسی به مترو رتبه‌ی سوم را دارا است. با توجه به نگاره ۴، دسترسی مترو در نواحی مرکزی و جنوبی منطقه بیشتر است، بنابراین انتظار می‌رود که ارزش املاک نیز در این نواحی نسبتاً بیشتر باشد.



نگاره ۱۰: نمودار رتبه‌بندی اهمیت ویژگی‌ها به روش TabNet

در نگاره ۱۱ نقشه‌ی منطقه‌ای املاک منطقه‌ی ۵ تهران تهیه شده است. قیمت‌ها در این نقشه نیز نتیجه قیمت‌گذاری املاک مسکونی توسط الگوریتم TabNet می‌باشد.

از جمله نمای ساختمان، تعداد واحدهای آپارتمانی موجود در هر طبقه و در کل ساختمان، سیستم‌های گرمایشی و سرمایشی و موضوع پرستیژ، زیبایی و کلاس اجتماعی کوچه‌ها و خیابان‌ها، قابل توجه است. شایان ذکر است که عواملی از قبیل نیاز فوری فروشنده به فروش ملک و یا نیاز خاص خریدار به خرید عجلولانه نیز می‌توانند سبب تغییرات شدید در قیمت هر ملک خاص شوند. در مجموع، عوامل ذکر شده در بالا منجر به کاهش دقت در نتایج همی الگوریتم‌ها می‌شود. به طور کلی عملکرد مناسب یادگیری عمیق TabNet در مقایسه با دیگر الگوریتم‌های یادگیری عمیق بیان‌کننده اهمیت توجه به ساختار داده‌ها و تعدادشان در انتخاب الگوریتم یادگیری عمیق مناسب است. در نهایت ارزیابی دو معیار RMSE و MAE ترکیب الگوریتم TabNet به روش انباشتن بیانگر بهبود و برتری این روش ترکیبی نسبت به الگوریتم پایه TabNet و دیگر روش‌ها بوده است. طبق نتایج RMSE این الگوریتم ترکیبی نسبت به الگوریتم پایه TabNet ۶/۶۵ درصد کاهش یافته است که بیان‌کننده بهبود خطای متوسط پیش‌بینی بوده است. با این حال ارزیابی معیار سوم نتیجه‌ی قابل قبولی نداشته است که نشان‌دهنده وجود اختلاف پیش‌بینی و انحراف از معیار تعدادی از نمونه‌هاست. با توجه به اینکه در این ترکیب سیستم استنتاج فازی و TabNet به صورت دو مدل مجزا در نظر گرفته شده است، دقت هر دو مدل پایه بر روی فرا مدل نهایی تأثیرگذار است. از آنجایی که ساختار فازی به تنهایی نتیجه‌ی بسیار دقیقی نداشته است، عدم بهبود معیار سوم ارزیابی قابل توجهی به نظر می‌رسد. با استفاده از روش ترکیب زیر کلاس‌بندی تا حدودی این معیار افزایش یافته است که موجب افزایش درجه‌ی تناسب مدل با تمامی داده‌ها شده است. با این حال به دلیل عدم بهبود دو معیار RMSE و MAE، متوسط خطای پیش‌بینی کاهش یافته است. اما با این حال معیار مورد نظر در مقایسه با الگوریتم پایه بهبودی نداشته است. از لحاظ دو معیار MAE، RMSE نیز بهبودی مشاهده نمی‌شود و متوسط خطای پیش‌بینی کاهش

لایه‌های فازی شامل چهار لایه فازی‌ساز، قوانین فازی، نرمال‌سازی و فازی‌زدایی می‌باشد. در نهایت بخش فازی در ابتدای الگوریتم خود رمزگذار به صورت دنباله‌ای قرار گرفته است. برای ترکیب TabNet و بخش فازی نیز از دو روش متفاوت انباشتن و زیر کلاس‌بندی استفاده شده است. به منظور ارزیابی نتایج الگوریتم‌های یادگیری عمیق، داده‌های ۲۷۹۷ ملک مسکونی تهران با یکدیگر مقایسه شد. برای مقایسه‌ی قابل اطمینان‌تر، نتایج الگوریتم XGBoost نیز به عنوان یکی از رایج‌ترین الگوریتم‌های یادگیری ماشین در مسائل رگرسیون، لحاظ شده است. در آخر برای ارزیابی روش ترکیبی یادگیری عمیق-فازی بر روی داده‌های تهران، به مقایسه‌ی نتایج این الگوریتم با شبکه‌های پایه پرداخته شده است. رتبه‌بندی تأثیرگذاری ویژگی‌ها در قیمت‌گذاری املاک مسکونی تهران نیز توسط الگوریتم TabNet که دقت بهتری داشته، مشخص شده است. نتایج رتبه‌بندی ویژگی‌ها حاکی از آن است که اکثر متغیرها معنادار بوده و به درستی انتخاب شده‌اند. البته با وجود اینکه بعضی از عوامل ساختاری بیشتر از عوامل مکانی، قیمت واحد مسکونی را تحت تأثیر قرار می‌دهند، اما به طور کلی عوامل مکانی تأثیر قابل توجهی داشته‌اند. رتبه‌ی دوم و سوم تأثیرگذاری ویژگی‌های طول و عرض جغرافیایی تأییدی بر این ادعاست. همچنین در میان ویژگی‌های مکانی دسترسی به مترو نیز تأثیر قابل توجهی داشته است. با توجه به جدول ۱ در مجموعه داده تهران کلیه معیارها نشان‌دهنده برتری مدل TabNet نسبت به دیگر مدل‌های یادگیری عمیق بوده است. این نتایج بیانگر کارایی بیشتر این الگوریتم در پیش‌بینی قیمت املاک و بهبود پنج درصدی R^2 است. با این حال الگوریتم یادگیری ماشین XGBoost بهتر از TabNet و دیگر الگوریتم‌های یادگیری عمیق عمل کرده است. این برتری بر این احتمال تأکید می‌کند که کم بودن تعداد داده‌ها مانع از دستیابی به نتیجه‌ی مطلوب و ایده آل روش‌های یادگیری عمیق شده است. همچنین با در نظر گرفتن نتایج کلیه روش‌ها به نظر می‌رسد که ارتباط بین قیمت املاک تهران و تعدادی پارامتر

ترکیبی مشاهده شود؛ بنابراین به منظور انجام مطالعه و تحقیقات آتی در راستای این مقاله می‌توان از ترکیب الگوریتم‌های یادگیری عمیق و فازی برای مجموعه داده‌ی بزرگ‌تر استفاده کرد. هم چنین در تحقیقات آتی می‌توان برای نقشه‌های دسترسی به امکانات در مقیاس منطقه‌ای، از فاصله‌ی تحت شبکه استفاده کرد. در نهایت پیشنهاد می‌شود که با افزایش بعد زمان به داده‌ها، پیش‌بینی مکانی-زمانی صورت گیرد تا ظرفیت الگوریتم یادگیری عمیق-فازی در داده‌های وابسته به زمان نیز بررسی شود. در نهایت این پیش‌بینی زمانی، تأثیر ویژگی ترافیک را به‌عنوان یک معیار متغیر، به شکل مناسب‌تری مدل‌سازی می‌کند.

منابع و مأخذ

- ۱ - عبده کلاهچی، م.، رفیعان نجف آبادی، م.، دهقانی، م.، & میرزاده، س. (۱۳۹۳). تحلیل عوامل مؤثر بر قیمت مسکن با استفاده از مدل تحلیل رگرسیون گام به گام (مطالعه موردی: محله فاطمی تهران). اقتصاد و مدیریت شهری، 2(7)، -. <https://www.sid.ir/fa/journal/ViewPaper.aspx?id=233672>
- 2 - Afonso, B., Melo, L., Oliveira, W., Sousa, S., & Berton, L. (2019). Housing prices prediction with a deep learning and random forest ensemble. *Anais do XVI Encontro Nacional de Inteligência Artificial e Computacional*,
- 3 - Arık, S. O., & Pfister, T. (2021). Tabnet: Attentive interpretable tabular learning. *AAAI*,
- 4 - Ayan, E., & Eken, S. (2021). Detection of price bubbles in Istanbul housing market using LSTM autoencoders: A district-based approach. *Soft Computing*, 25(12), 7957-7973.
- 5 - Bae, S., & Yu, J. (2017). Predicting the real estate price index using deep learning. *Korea Real Estate Review*, 27(3), 71-86.
- 6 - Benjamin, J., Guttery, R., & Sirmans, C. (2004). Mass appraisal: An introduction to multiple regression analysis for real estate valuation. *Journal of Real Estate Practice and Education*, 7(1), 65-77.
- 7 - Borst, R. A. (1991). Artificial neural networks: the

نداشته است. اما باین وجود از لحاظ هر سه معیار ارزیابی، این ترکیب در رقابت با سایر الگوریتم‌های یادگیری عمیق مقدار قابل قبولی داشته است. در آخر تفاوت در بهبود نتایج ترکیب TabNet و فازی به دو روش انباشتن و زیر کلاس‌بندی بیانگر تأثیر انتخاب روش ترکیب، در بهبود این الگوریتم‌های ترکیبی است. همچنین برای ترکیب دنباله‌ای خود رمزگذار و فازی نیز از لحاظ دو معیار RMSE و MAE بهبود قابل قبولی مشاهده شده است که بازهم تأییدی بر نتیجه‌ی قابل قبول این روش ترکیبی بوده است. به علاوه نتایج بیانگر برتری الگوریتم خود رمزگذار فازی نسبت به همه‌ی الگوریتم‌های یادگیری عمیق است. البته باید توجه داشت که به‌طور کل ترکیب TabNet و فازی از این ترکیب بهتر عمل کرده است. بنابراین با وجود اینکه ترکیب سیستم استنتاج فازی و الگوریتم‌های یادگیری عمیق بر اساس تمام معیارهای ارزیابی بهبود قابل توجهی نداشته است، باین حال منجر به بهبود قابل قبول دو معیار RMSE و MAE شده است. همان‌طور که اشاره شد در ارزیابی نهایی مسئله به دلیل اهمیت کاهش متوسط خطای پیش‌بینی، معیار خطای RMSE ملاک قرار گرفته است و معیار R^2 به دلیل تأثیرپذیری از انحراف معیار و توجه به نتیجه‌ی موردی کمتر مورد توجه است. در نتیجه با توجه به این ارزیابی در مسائلی که با عدم قطعیت و لزوم استفاده از منطق فازی همراه است، به نظر می‌رسد که بهتر است از الگوریتم ترکیبی یادگیری عمیق-فازی استفاده شود. از طرفی با توجه به اینکه روش مناسب برای دستیابی به مزایای هر دو روش و حل مشکلات مربوط به هر کدام، ترکیب آن‌ها در یک سیستم یکپارچه است، در پژوهش ترکیب یکپارچه‌ی مفاهیم فازی به یادگیری عمیق معرفی شده است تا ضمن یادگیری سریع و دقیق‌تر مسائل پیچیده، بر کاستی‌های قطعی بودن مدل‌های یادگیری عمیق و در نظر نگرفتن عدم قطعیت در این مدل‌ها غلبه شود.

با توجه به اینکه اضافه کردن لایه‌های فازی به افزایش پارامترهای قابل آموزش منجر می‌شود، انتظار می‌رود که با افزایش داده‌ها بهبود قابل توجه‌تری برای این الگوریتم

of house price: Incorporating a geographically weighted regression approach within the Belfast housing market. *Journal of Financial Management of Property and Construction*.

19 - Nghiep, N., & Al, C. (2001). Predicting housing value: A comparison of multiple regression analysis and artificial neural networks. *Journal of real estate research*, 22(3), 313-336.

20 - Rafiei, M. H., & Adeli, H. (2016). A novel machine learning model for estimation of sale prices of real estate units. *Journal of Construction Engineering and Management*, 142(2), 04015066.

21- Ribeiro, M. H. D. M., & dos Santos Coelho, L. (2020). Ensemble approach based on bagging, boosting and stacking for short-term prediction in agribusiness time series. *Applied Soft Computing*, 86, 105837.

22 - Seya, H., & Shiroi, D. (2019). A comparison of apartment rent price prediction using a large dataset: Kriging versus DNN. *arXiv preprint arXiv:1906.11099*.

23 - Shwartz-Ziv, R., & Armon, A. (2022). Tabular data: Deep learning is not all you need. *Information Fusion*, 81, 84-90.

24 - Souri, D., & MONIRI, J. S. (2011). Estate pricing model, an application of geographic balanced regression.

25 - Tabales, J. M. N., Caridad, J. M., & Carmona, F. J. R. (2013). Artificial neural networks for predicting real estate price. *Revista de Métodos Cuantitativos para la Economía y la Empresa*, 15, 29-44.

26 - Talpur, N., Abdulkadir, S. J., & Hasan, M. H. (2020). A deep learning based neuro-fuzzy approach for solving classification problems. *2020 International Conference on Computational Intelligence (ICCI)*,

27 - Victoria, A. H., & Maragatham, G. (2021). Automatic tuning of hyperparameters using Bayesian optimization. *Evolving Systems*, 12(1), 217-223.

28 - Xiaolong, H., & Ming, Z. (2010). Applied research on real estate price prediction by the neural network. *2010 The 2nd Conference on Environmental Science and Information Application Technology*,

29 - Xu, H., & Gade, A. (2017). Smart real estate assessments using structured deep neural networks. *2017 IEEE SmartWorld, Ubiquitous Intelligence & Computing*,

next modelling/calibration technology for the assessment community. *Property Tax Journal*, 10(1), 69-94.

8 - Chen, C. P., Zhang, C.-Y., Chen, L., & Gan, M. (2015). Fuzzy restricted Boltzmann machine for the enhancement of deep learning. *IEEE Transactions on Fuzzy Systems*, 23(6), 2163-2173.

9 - Chen, W., An, J., Li, R., Fu, L., Xie, G., Bhuiyan, M. Z. A., & Li, K. (2018). A novel fuzzy deep-learning approach to traffic flow prediction with uncertain spatial-temporal data features. *Future generation computer systems*, 89, 78-88.

10 - Chen, X., Wei, L., & Xu, J. (2017). House price prediction using lstm. *arXiv preprint arXiv:1709.08432*.

11 - Guan, J., Shi, D., Zurada, J. M., & Levitan, A. S. (2014). Analyzing massive data sets: an adaptive fuzzy neural approach for prediction, with a real estate illustration. *Journal of organizational computing and electronic commerce*, 24(1), 94-112.

12 - Guan, J., Zurada, J., & Levitan, A. (2008). An adaptive neuro-fuzzy inference system based approach to real estate property assessment. *Journal of real estate research*, 30(4), 395-422.

13 - Jang, J.-S. (1993). ANFIS: adaptive-network-based fuzzy inference system. *IEEE transactions on systems, man, and cybernetics*, 23(3), 665-685.

14 - Kang, Y., Zhang, F., Peng, W., Gao, S., Rao, J., Duarte, F., & Ratti, C. (2021). Understanding house price appreciation using multi-source big geo-data and machine learning. *Land Use Policy*, 111, 104919.

15 - Khalid, S., Khalil, T., & Nasreen, S. (2014). A survey of feature selection and feature extraction techniques in machine learning. *2014 science and information conference*,

16 - Kim, H., Kwon, Y., & Choi, Y. (2020). Assessing the impact of public rental housing on the housing prices in proximity: based on the regional and local level of price prediction models using long short-term memory (LSTM). *Sustainability*, 12(18), 7520.

17 - Li, Z. X. (2007). Using fuzzy neural network in real estate prices prediction. *2007 Chinese control conference*,

18 - McCord, M., Davis, P., Haran, M., McGreal, S., & McIlhatton, D. (2012). Spatial variation as a determinant

Advanced & Trusted Computed, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCOM/IOP/SCI),

30 - Yazdanbakhsh, O., & Dick, S. (2019). A deep neuro-fuzzy network for image classification. arXiv preprint arXiv:2001.01686.

31 - Yiorkas, C., & Dimopoulos, T. (2017). Implementing GIS in real estate price prediction and mass valuation: the case study of Nicosia District. Fifth International Conference on Remote Sensing and Geoinformation of the Environment (RSCy2017),

32 - Zaremba, W., Sutskever, I., & Vinyals, O. (2014). Recurrent neural network regularization. arXiv preprint arXiv:1409.2329.



COPYRIGHTS

©2023 by the authors. Published by National Geographical Organization. This article is an open access article distributed under the terms and conditions of the Creative Commons [Attribution-NoDerivs 3.0 Unported \(CC BY-ND 3.0\)](https://creativecommons.org/licenses/by-nd/3.0/)





پروپوزیشن گاہ علوم انسانی و مطالعات فرہنگی
پرتال جامع علوم انسانی