

A Comparison of Multiple Linear Regression and Random Forest Regression to Evaluate the Price of Residential Units (Case Study: North Valiasr, Tabriz)

Mohammad Ali Koushesh Vatan*

PhD. Candidate of Geography and Urban Planning at University of Tabriz

Akbar Asghari Zamani

Associate Professor of Geography and Urban Planning at University of Tabriz

Mohammad Nemati

PhD. Candidate of Geography and Urban Planning at University of Tabriz

Firuz Pourmohammad

M.Sc. of Geography and Urban Planning at University of Tabriz

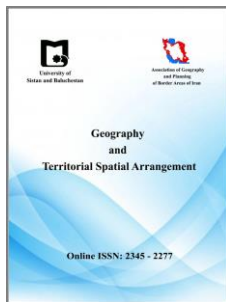


Koushesh Vatan, M.A & Asghari Zamani, A & Nemati, M & Pourmohammad, F (2021). [A Comparison of Multiple Linear Regression and Random Forest Regression to Evaluate the Price of Residential Units (Case Study: North Valiasr, Tabriz)]. *Geography and Territorial Spatial Arrangement*, 11 (40), 57-82.

doi <http://dx.doi.org/10.22111/GAIIJ.2021.6486>

Received:20/01/2021

Accepted:11/09/2021



ABSTRACT

One of the most fundamental aspects of any country is the housing economy. Because its price changes will cause numerous effects on the national economy in the short and long periods; therefore, it is essential to obtain a model which can assess housing prices. In this regard, the objectives of this research are to compare multiple linear regression and random forest regression to evaluate the price of residential units and extract the important factors in relation to the price of them. The statistical population included four north valiasr neighborhoods (n=30,272 units) and the sample size was estimated to be 379 units using the cochrane formula at 95% confidence level and with a 5% error. But 400 samples considered. To eliminate the effect of time, only the data of september, 2020 were used. Also, arcmap, spss and rstudio software were used to analyze the data. According to the results, area, apartment floors, construction year, proximity to health centers, urban facilities, green spaces, religious land-use, medical centers, military land-uses and floor level, are the ten most important variables in relation to housing prices in the north valiasr neighborhoods, respectively. Further, according to the findings, random forest regression has a superior capability in predicting housing prices in north valiasr of tabriz compared to multiple linear regression.

Keywords:

Housing Price,
Multiple Linear
Regression, Random
Forest Regression,
North Valiasr, Tabriz.

Copyright©2021, Geography and Territorial Spatial Arrangement. This is an open-access article distributed under the terms of the Creative Commons Attribution- noncommercial 4.0 International License which permits copy and redistribute the material just in noncommercial usages, provided the original work is properly cited.

Extended Abstract

1- Introduction



One of the most fundamental aspects of any country is housing economy. Because its price changes

will cause numerous effects on the national economy in the short and long periods, however, this trend is significantly fast in countries with inflationary economy. Housing is a multidimensional matter. It means, not only it has a consumption aspect but also capital. Moreover, significant amount of the capital of the countries is injected into the housing sector; therefore, it is essential to obtain a highly efficient model which can assess housing price. This study is based on lancaster's consumerism theory. It says. A

*Corresponding Author: Mohammad Ali Koushesh Vatan

E-mail: ali.kousheshvatan@tabrizu.ac.i

product has no value on its own. Whereas, its attributes make it advantageous. In other words, the way that housing gains value is based on its characteristics.

In this regard, the objectives of this research are as follows:

1. To compare multiple linear regression and random forest regression to evaluate the price of residential units.
2. To extract the important factors in relation to the price of residential units.

2- Study Area

North valiasr is located in the northeast of tabriz (region 1). It is one of the big and dense zones of this city. Also, it is located in the high elevation areas of tabriz. Further, economically and socially it is a stand-out zone in tabriz. In addition, it consists of four neighborhoods with the names of valiasr one, two, three, and four, which have an area of 150 ha, 61 ha, 152 ha, and 123 ha, respectively. According to the statistical blocks of tabriz, the population of the four mentioned neighborhoods is 45440.

3- Material and Methods

This study has made use of multiple linear regression and random forest regression to evaluate the price of residential units. The statistical population included four north valiasr neighborhoods (n=30,272 units) and the sample size was estimated to be 379 units using the cochrane formula at 95% confidence level and with a 5% error. But 400 samples considered. To eliminate the effect of time, only the data of september, 2020 were used. Also, arcmap, spss and rstudio softwares were used to analyze the data. Research variables are in two categories of physical and accessibility.

4- Result and Discussion

For prediction, with multiple linear regression and random forest regression we have divided our sample into two groups of experimenting (30 percent) and training (70 percent). In order to run random forest regression, we have selected three parameters including ntree, mtry, and max_depth. We have identified ntree to 4000, mtry by n/3 that for our study was equal to 7, and max_depth to 22. We have used coefficient of determination, root mean squared log error (rmsle), root mean squared error (rmse), and mean absolute error (mae) to assess the performance of two regressions. Based on these indices, we have found that random forest regression is superior to multiple linear regression. Further, we have interpolated observed and predicted results from random forest regression. They were highly correlated ($r = 0.98$). Furthermore, we have extracted the importance of physical and

accessibility variables. Results are as follows: area (24.78%), apartment floors (9.86%), construction year (7.9%), proximity to health centers (6.89%), urban facilities (6.84%), green spaces (5.03%), religious land-use (4.48%), medical centers (4.26%), military land-uses (3.99%) and floor level (3.92%).

5- Conclusion

According to the results, area, apartment floors, construction year, proximity to health centers, urban facilities, green spaces, religious land-use, medical centers, military land-uses and floor level, are the ten most important variables in relation to housing prices in the north valiasr neighborhoods, respectively. Further, according to the findings, random forest regression has a superior capability in predicting housing prices in north valiasr of tabriz compared to multiple linear regression.

Keywords: housing price, multiple linear regression, random forest regression, north valiasr, tabriz.

6- References (Persian)

Asghari Zamani, A., Rostaei, S., Koushesh Vatan, M. (2021). Evaluating the Land Subdivision of Residential and commercial Lands in terms of Land Subdivision Indicators and Land Stakeholders Case study: District 1 and 3 of Tabriz City. *Geography and Planning*, 24(74), 13-28. doi: 10.22034/gp.2021.10793.

https://geoplanning.tabrizu.ac.ir/article_10793.html

Akbari N, Khoshakhlaq R, Mardiha S. (2013). Measurement and Valuation of Factors Affecting Housing Choice Using a Choice Experiment Method: Viewpoints of Households Living at Old Urban Textures of Isfahan. *QJER*. 13 (3):19-47.

<http://ecor.modares.ac.ir/article-18-10602-fa.html>

Pour Mohammadi, M.R. (2000). Housing Planning, Tehran: Samat Publications.

<https://samta.samt.ac.ir/content/10806>

Rostaei, S., Teimory, I., & Nemati, M. (2020). Assessment of Effective Factors on Urban House Prices Using Artificial Neural Network (Case Study: District 2 of Tabriz. *Geography and Development Iranian Journal*, 18(59), 129-148. doi: 10.22111/gdij.2020.5464.

https://gdij.usb.ac.ir/article_5464.html

Amanpour S, Soleymani Rad E, Keshtkar L, Mokhtari Chelcheh S. (2015). Ahwaz Estimated Housing Prices Using Neural Networks. *IUESA*. 3 (9):45-57.

<http://iueam.ir/article-1-105-fa.html>

Teimoori, D., Soltan gheys, N., Gholizadeh, Y. (2017). Estimation of Urban housing Price by Using Hedonic and Artificial Neural Networks; (Case Study Koye Valiaser, Tabriz). *Geography and Territorial Spatial Arrangement*, 7(22), 41-56. doi: 10.22111/gaij.2017.2995.

https://gaij.usb.ac.ir/article_2995.html

Khalili Araghi M, Nobahar E. (2012). Predicting Housing Prices for the City of Tabriz: Application of the Hedonic Pricing and Artificial Neural Network Models. *qjerp*.19 (60):113-138.

<http://qjerp.ir/article-1-189-fa.html>

Sadat Miraei, N. (2012). Pathology and Capacity Assessment of Urban Road Network with Crisis Management Approach, Case Study: Valiasr, Tabriz. M.Sc. Thesis, Payame Noor University of Tehran, Faculty of Social Sciences. <https://ganj.irandoc.ac.ir/#/articles/f37a8cc7cb3ae98dc778e10abf45462>

Shaqaqi, Hassan (2014). An Analysis of the Role and Position of Urban Street Lighting in Urban Planning, Case Study: Valiasr Alley and Abbasi Neighborhood in Tabriz, M.Sc. Thesis, University of Tabriz, Aras Campus.

<https://ganj.irandoc.ac.ir/#/articles/6990d8bf49347c04a209b6947441fd6e/search/97eed76bdbb89481a2484fdcc25eb85d>

Ghorbani S, Afigheh S M. (2017). Forecasting the House Price for Ahvaz City: the Comparison of the Hedonic and Artificial Neural Network Models. *IUESA*. 5 (19):29-44.

<http://iueam.ir/article-1-738-fa.html>

Koushesh Vatan, MA. (2019). A Comparative Study of the Role of Land Subdivision Quality in Optimal Land Use Productivity in Iranian Capitals, Case Study: Districts One and Three of Tabriz, M.Sc. Thesis, Tabriz University.

<https://ganj.irandoc.ac.ir/#/articles/0da0e12bdf45730035babcac7d8a2f9f/search/75bb541ae1e937a22ae06535f0b38925>

Mokhber, Abbas (1984). *Social Dimensions of Housing*, translated by the Center for Socio-Economic Documents of the Program and Budget Organization, Tehran.

<https://www.gisoom.com/book/144209>

Naghshe Mohit The Consulting Engineers Company. (2012). *Tabriz city development plan, social studies report of the current stage*.

Mirhashemi H, Yarahmadi D, Sharifi S, Farzin S. (2019). Improvement of Support Vector Machine and Random Forest Algorithm in Predicting Khorramabad River Flow Using Non-uniform De-Noising of data and Simplex Algorithm. *jwmseir*. 13 (47):40-51.

<http://jwmseir.ir/article-1-846-fa.html>

Yazdani, Fardin; Eliassy, Tahmurth. (2001). Economic study of housing supply and demand in urban areas of Isfahan province, Management and Planning Organization of Isfahan province, Deputy of Planning and Coordination.

7- References (English)

Abbot, M. L. (2017). *Using Statistics in the Social and Health Sciences with SPSS and Excel*. John Wiley & Sons.

DOI: 10.1002/9781119121077.

Abrougui, K. & Gabsi, K. & Mercatoris, B. & Khemis, C. & Amami, R. & Chehaibi, Sa. (2019). Prediction of organic potato yield using tillage systems and soil properties by artificial neural network and multiple linear regressions. *Soil and Tillage Research*. Vol. 190. pp 202-208.

DOI: 10.1016/j.still.2019.01.011.

Ahmad, Muhammad & Mourshed, Monjur & Rezgui, Yacine. (2017). Trees vs Neurons: Comparison between random forest and ANN for high-resolution prediction of building energy consumption. *Energy and Buildings*. 147.

DOI: 10.1016/j.enbuild.2017.04.038.

Breiman, L. (2002). *Manual on setting up, using, and understanding random forests v3.1*.

https://www.stat.berkeley.edu/~breiman/Using_random_forests_V3.1.pdf

Caffo, B. (2019). *Regression Models for Data Science in R: A companion book for the Coursera Regression Models class*. Leanpub.

<https://leanpub.com/regmods>.

Case, B., Clapp, J., Dubin, R. Rodriguez, M. (2004). Modeling Spatial and Temporal House Price Patterns: A Comparison of Four Models. *The Journal of Real Estate Finance and Economics* 29, 167-191.

<https://doi.org/10.1023/B:REAL.0000035309.60607.53>

Case, K. & Quigley, J. & Shiller, R. (2005). «Comparing Wealth Effects: The Stock Market

- versus the Housing Market," *Advances in Macroeconomics*, Berkeley Electronic Press, vol. 5(1), pages 1-32. DOI: 10.3386/w8606.
- Cozmei, C., & Onofrei, M. (2012). Impact of Property Taxes on Commercial Real Estate Competition in Romania, *Journal of Procedia Economics and Finance*, Vol3. Pages 604-610.
[https://doi.org/10.1016/S2212-5671\(12\)00202-X](https://doi.org/10.1016/S2212-5671(12)00202-X).
- George, D. & Mallery, P. (2020). *IBM SPSS Statistics 26 Step by Step: A Simple Guide and Reference*. 16th Edition. Routledge: Taylor & Francis Group.
<https://www.routledge.com/IBM-SPSS-Statistics-26-Step-by-Step-A-Simple-Guide-and-Reference/George-Mallery/p/book/9780367174354>.
- Giussani, Andrea (2020), *Applied Machine Learning with Python*, Milano, EGEA S.p.A.
https://books.google.com/books/about/Applied_Machine_Learning_with_Python.html?id=IN5AygEACAAJ.
- Heyman, A., & Sommervoll, D.E. (2019). House prices and relative location. *Cities*, 95, 1-14.
<https://doi.org/10.1016/j.cities.2019.06.004>.
- Hong, Jengei & Choi, Heeyoul «Henry & Kim, Woo-sung. (2020). A house price valuation based on the random forest approach: the mass appraisal of residential property in South Korea. *International Journal of Strategic Property Management*. 24. 1-13.
DOI: 10.3846/ijspm.2020.11544.
- Jui J.J., Imran Molla M.M., Bari B.S., Rashid M., Hasan M.J. (2020) Flat Price Prediction Using Linear and Random Forest Regression Based on Machine Learning Techniques. In: Mohd Razman M., Mat Jizat J., Mat Yahya N., Myung H., Zainal Abidin A., Abdul Karim M. (eds) *Embracing Industry 4.0. Lecture Notes in Electrical Engineering*, vol 678. Springer, Singapore.
https://doi.org/10.1007/978-981-15-6025-5_19.
- Korosteleva, O. (2019). *Advanced Regression Models with SAS and R*. CRC Press.
DOI: 10.1201/9781315169828.
- Kotta H., Pardasani K., Pandya M., Ghosh R. (2021) Optimization of Loss Functions for Predictive Soil Mapping. In: Hassanien A., Bhatnagar R., Darwish A. (eds) *Advanced Machine Learning Technologies and Applications. AMLTA 2020. Advances in Intelligent Systems and Computing*, vol 1141. Springer, Singapore.
https://doi.org/10.1007/978-981-15-3383-9_9.
- Lancaster K.J. (1966). A New Approach to Consumer Theory, *Journal of Political Economy*, Vol. 74. No. 2. Pages 132-157.
doi.org/10.1086/259131.
- Leung, C. (2004). Macroeconomics and housing: a review of the literature, *Journal of Housing Economics*, Vol.13. Pages 249-267.
<https://doi.org/10.1016/j.jhe.2004.09.002>.
- Liaw, Andy & Wiener, Matthew. (2001). Classification and Regression by RandomForest. *RNews*. 23.
https://www.researchgate.net/publication/228451484_Classification_and_Regression_by_RandomForest.
- Mahajan U., Krishnan A., Malhotra V., Sharma D., Gore S. (2021). Predicting Competitive Weightlifting Performance Using Regression and Tree-Based Algorithms. In: Hassanien A., Bhatnagar R., Darwish A. (eds) *Advanced Machine Learning Technologies and Applications. AMLTA 2020. Advances in Intelligent Systems and Computing*, vol 1141. Springer, Singapore.
https://doi.org/10.1007/978-981-15-3383-9_36.
- Makhija R., Ali S., Jaya Krishna R. (2021) Detecting Influencers in Social Networks Through Machine Learning Techniques. In: Hassanien A., Bhatnagar R., Darwish A. (eds) *Advanced Machine Learning Technologies and Applications. AMLTA 2020. Advances in Intelligent Systems and Computing*, vol 1141. Springer, Singapore.
https://doi.org/10.1007/978-981-15-3383-9_23.
- Čeh Marjan & Kilibarda, Milan & Liseč, Anka & Bajat, Branislav. (2018). Estimating the Performance of Random Forest versus Multiple Regression for Predicting Prices of the Apartments. *ISPRS International Journal of Geo-Information*. 7. 168.
DOI: 10.3390/ijgi7050168.
- Masri, Hilmi & Nawawi, Abdul & Sipan, Ibrahim. (2016). Review of Building, Locational, Neighbourhood Qualities Affecting House Prices in Malaysia. *Procedia - Social and Behavioral Sciences*. 234. 452-460.
DOI: 10.1016/j.sbspro.2016.10.263.
- Mayer, A. (2013). *Introduction to Statistics and SPSS in Psychology*, Edinburgh: Pearson Higher Ed.
https://books.google.com/books/about/Introduction_to_Statistics_and_SPSS_in_P.html?id=SDHtMgEACAAJ.

- Nair S.N., Gopi E.S. (2020) Deep Learning Techniques for Crime Hotspot Detection. In: Kulkarni A., Satapathy S. (eds) Optimization in Machine Learning and Applications. Algorithms for Intelligent Systems. Springer, Singapore.
https://doi.org/10.1007/978-981-15-0994-0_2.
- Pal, M. & Bharati, P. (2019). Applications of Regression Techniques. Springer Singapore.
<https://www.springer.com/gp/book/9789811393136>.
- Richardson, R. (2015). Business Applications of Multiple Regression, Second Edition - Quantitative approaches to decision making collection. Business Expert Press.
<https://www.businessexpertpress.com/books/business-applications-multiple-regression-second-edition>.
- Schulz, R., Werwatz, A. (2004). A State Space Model for Berlin House Prices: Estimation and Economic Interpretation. The Journal of Real Estate Finance and Economics 28, 37-57.
<https://doi.org/10.1023/A:1026373523075>.
- Selim, H. (2009). Determinants of House Prices in Turkey: Hedonic Regression versus Artificial Neural Network, Journal of Expert Systems with Applications, Vol.36. Issue 2. Part 2. Pages 2843-2852.
doi.org/10.1016/j.eswa.2008.01.044.
- Sharma R., Rani S. (2021) A Novel Approach for Smart-Healthcare Recommender System. In: Hassanien A., Bhatnagar R., Darwish A. (eds) Advanced Machine Learning Technologies and Applications. AMLTA 2020. Advances in Intelligent Systems and Computing, vol 1141. Springer, Singapore.
https://doi.org/10.1007/978-981-15-3383-9_46.
- Sirmans, G., MacPherson, D., & Zietz, E. (2009). The Composition of Hedonic Pricing Models. Journal of Real Estate Literature, 13, 3-43.
<https://www.semanticscholar.org/paper/The-Composition-of-Hedonic-Pricing-Models-Sirmans-MacPherson/2fde22eba11fad9e671eacb588353f123df0d3ac>
- Sullivan, W. (2017). Machine learning Beginners Guide Algorithms: Supervised & Unsupervised learning, Decision Tree & Random Forest Introduction. Healthy Pragmatic Solutions Inc.
https://books.google.com/books/about/Machine_Learning_For_Beginners_Guide_Algorithm?id=v6saxAEACAAJ&source=kp_book_description
- Truong, Quang & Nguyen, Minh & Dang, Hy & Mei, B. (2020). Housing Price Prediction via Improved Machine Learning Techniques. Procedia Computer Science. 174. 433-442.
DOI: [org/10.1016/j.procs.2020.06.111](https://doi.org/10.1016/j.procs.2020.06.111).
- Varma, A., & Sarma, A., & Doshi, S., & Nair, R. (2018). House Price Prediction Using Machine Learning and Neural Networks, 2018 Second International Conference on Inventive Communication and Computational Technologies, Coimbatore, pp. 1936-1939.
DOI: [org/10.1109/ICICCT.2018.8473231](https://doi.org/10.1109/ICICCT.2018.8473231).
- Wang, C. & Wu, H. (2018). A new machine learning approach to house price estimation. New Trends in Mathematical Science. 4. 165-171.
DOI: [10.20852/ntmsci.2018.327](https://doi.org/10.20852/ntmsci.2018.327).
- Wlodarczak, Peter (2020). Machine Learning and its applications, First edition. Boca Raton, FL: CRC Press/Taylor & Francis Group.
DOI: [10.1201/9780429448782](https://doi.org/10.1201/9780429448782).
- Yeşilkanat, C. (2020). Spatio-temporal estimation of the daily cases of COVID-19 in worldwide using random forest machine learning algorithm. Chaos, Solitons & Fractals. 140. 110210.
DOI: [10.1016/j.chaos.2020.110210](https://doi.org/10.1016/j.chaos.2020.110210).

مقایسه عملکرد رگرسیون‌های خطی چندگانه و جنگل تصادفی در جهت ارزیابی قیمت واحدهای مسکونی (موردشناسی: ولیعصر شمالی، شهر تبریز)

محمدعلی کوشش وطن* (دانشجوی دکتری جغرافیا و برنامه‌ریزی شهری، دانشکده برنامه‌ریزی و علوم محیطی، دانشگاه تبریز)

اکبر اصغری زمانی (دانشیار گروه جغرافیا و برنامه‌ریزی شهری، دانشکده برنامه‌ریزی و علوم محیطی، دانشگاه تبریز)

محمد نعمتی (دانشجوی دکتری جغرافیا و برنامه‌ریزی شهری، دانشکده برنامه‌ریزی و علوم محیطی، دانشگاه تبریز)

فیروز پورمحمد (کارشناسی ارشد جغرافیا و برنامه‌ریزی شهری، دانشگاه تبریز)

چکیده

اقتصاد مسکن یکی از مهم‌ترین ابعاد اقتصادی هر کشوری می‌باشد. چرا که تغییرات قیمتی آن تبعات مختلفی را در کوتاه‌مدت و بلندمدت بر اقتصاد ملی در بر خواهد داشت؛ بنابراین، تعریف مدلی که بتواند قیمت مسکن را ارزیابی نماید بسیار ضروری است. در این راستا، اهداف پژوهش حاضر مقایسه رگرسیون‌های خطی چندگانه و جنگل تصادفی برای تعریف مدل ارزیابی قیمت واحدهای مسکونی و استخراج عوامل مؤثر بر قیمت واحدهای مسکونی می‌باشد. جامعه آماری پژوهش حاضر کلیه واحدهای مسکونی محله‌های یک الی چهار کوی ولیعصر واقع در منطقه یک شهر تبریز به تعداد ۳۰۲۷۲ واحد می‌باشد. با استفاده از فرمول کوکران ۳۷۹ نمونه در سطح اطمینان ۹۵ درصد و ضریب خطای ۵ درصد برآورد گردید. جهت مدل‌سازی مطلوب، اطلاعات ۴۰۰ واحد مسکونی به‌عنوان نمونه به‌صورت میدانی جمع‌آوری شد. به‌منظور حذف اثر زمان در این پژوهش، تنها از داده‌های شهریور سال ۱۳۹۹ استفاده شد. همچنین، جهت تجزیه و تحلیل داده‌ها از نرم‌افزارهای ArcMap، SPSS و RStudio استفاده شد.

بر اساس نتایج، متغیرهای مساحت با ۲۴/۷۸ درصد، کل طبقات ساختمان با ۹/۸۶ درصد، سال ساخت با ۷/۹ درصد، فاصله از مراکز بهداشتی با ۶/۸۹ درصد، فاصله از تأسیسات شهری با ۶/۸۴ درصد، فاصله از فضای سبز با ۵/۰۳ درصد، فاصله از کاربری مذهبی با ۴/۴۸ درصد، فاصله از مراکز درمانی با ۴/۲۶ درصد، فاصله از مراکز نظامی و انتظامی با ۳/۹۹ درصد و استقرار واحد در طبقه با ۳/۹۲ درصد به ترتیب ده متغیر مؤثر در رابطه با قیمت مسکن محله‌های ولیعصر شمالی می‌باشند.

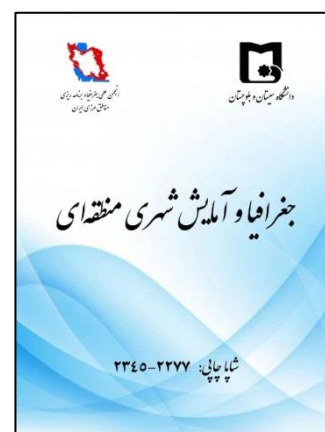
بر پایه یافته‌های پژوهش مشخص گردید که رگرسیون جنگل تصادفی نسبت به رگرسیون خطی چندگانه، رگرسیون جنگل تصادفی، ولیعصر شمالی، شهر تبریز.

مقاله پژوهشی

تاریخ دریافت: ۱۱ بهمن ۱۳۹۹

تاریخ پذیرش: ۲۰ شهریور ۱۴۰۰

صفحات: ۵۷-۸۲



کلید واژه‌ها:

قیمت مسکن، رگرسیون خطی چندگانه، رگرسیون جنگل تصادفی، ولیعصر شمالی، شهر تبریز.

* نویسنده مسئول: محمدعلی کوشش وطن

پست الکترونیک: ali.kousheshvatan@tabrizu.ac.ir

مقدمه

اقتصاد مسکن از متغیرترین و غیرشفاف‌ترین حوزه‌های برنامه‌ریزی می‌باشد. چرا که مشخصه‌های عرضه و تقاضای آن پویا بوده و روندهای کوتاه و بلندمدت قیمت مسکن را دستخوش تغییر قرار می‌دهد. البته این مسئله در کشورهایی با اقتصاد تورم‌زا روند شتاب‌زده‌ای دارد. مسکن دارای ماهیت چندبعدی است. یعنی علاوه بر اینکه جنبه مصرفی دارد، دارای قابلیت سرمایه‌ای نیز می‌باشد. همچنین، بخش قابل توجهی از سرمایه کشورها به بخش مسکن تزریق می‌گردد. چرا که مسکن به دلیل خودکفا بودن، قدرت بالای جذب نقدینگی و وجود تقاضای ثابت در خرید و اجاره به‌عنوان یکی از ارکان اصلی اقتصادی کشورها شناخته می‌شود. بخش مسکن دارای روابط پیچیده و متعددی با بخش‌های مختلف اقتصادی و تولیدی است. در حقیقت، قیمت مسکن به‌نوعی وضعیت اقتصادی کشور و نظام سیاست‌گذاری و مدیریت زمین و مسکن را نشان می‌دهد. به‌طوری‌که در اقتصادهای با تورم پایین و کنترل‌شده، بیشتر تقاضای مسکن بر اساس وجه مصرفی آن شکل می‌گیرد. در نظام‌های اقتصادی همچون ایران که نقدینگی و تورم بالا وجود دارد، وجه سرمایه‌ای مسکن است که بیشتر تقاضا را تشکیل می‌دهد. در این راستا، شناخت عوامل مؤثر بر قیمت مسکن و همچنین تعریف مدل ارزیابی قیمت مسکن مهم می‌باشد. چرا که، قیمت مسکن مسئله مهمی برای صاحبان حقیقی و بالقوه آن است و شناسایی عواملی که در قیمت مسکن تأثیرگذار هستند در سیاست‌گذاری و مدیریت قیمت مسکن حائز اهمیت است. به‌طوری‌که به هر میزان حوزه قیمتی مسکن شفاف و معلوم باشد، برنامه‌ریزی جهت کنترل نوسان قیمت مسکن نیز تسهیل می‌گردد. علاوه بر این، مقوله قیمت مسکن برای بنگاه‌ها، بانک‌ها و تصمیم‌گیران نیز از اهمیت بالایی برخوردار است (Schulz & Werwatz,

37: 2004). از طرفی، تغییر قیمت مسکن نگرانی بزرگی برای همه افراد جامعه و دولت‌ها است؛ زیرا وضعیت اجتماعی-اقتصادی را از خود متأثر ساخته و باعث ایجاد تأثیرات آتی در وضعیت اقتصاد ملی خواهد شد (Selim, 2009: 2843).

خانه‌های تحت مالکیت خانوارها نه تنها گزینه‌ای جهت مکانی برای زندگی در اختیار آنان قرار می‌دهد، بلکه به‌خاطر ناهمگن بودن، نقدینگی پایین، هزینه معاملات بالا و ثابت بودن در موقعیت مکانی خود، کالایی خاص در بازار تلقی می‌گردد. در حقیقت، مسکن نوعی پس‌انداز بوده و مهم‌ترین سهم را در سبد دارایی خانوار تشکیل می‌دهند. به‌طوری‌که اکثر مردم به آن به‌عنوان مهم‌ترین سرمایه‌گذاری نگاه می‌کنند. از این‌رو، توسعه و سرمایه‌گذاری در مسکن به‌عنوان یک فعالیت اقتصادی محلی محسوب می‌گردد (Cozmei & Masri et al., 2016: 453; Onofrei, 2012: 604). در واقع، در بیشتر کشورها املاک و مستغلات بزرگ‌ترین جزء دارایی خانوارها است. در این راستا، ارزش مسکن خانوارها تأثیر عمده‌ای در مصرف و فرصت‌های پس‌انداز آن‌ها دارد (Case et al., 2005: 7)؛ بنابراین تغییرات در قیمت مسکن موجب نوسانات قابل توجهی در دارایی خانوار خواهد شد (Leung, 2004: 251).

پیرو نظریه مصرف‌کننده لنکستر¹، کالا به‌خودی‌خود، هیچگاه سودمند نیست؛ بلکه ویژگی‌های آن است که باعث سودمندی آن کالا می‌شود. به عبارتی، نحوه ارزش‌گذاری مسکن براساس ویژگی‌های آن است (Lancster, 1966, 107). کاملاً مشخص است که ارزش املاک و مستغلات وابسته به ویژگی‌های موقعیتی و بازاری آن است (Case et al., 2004: 167). به‌طوری‌که برای هر خانوار مسکن مجموعه‌ای از خدمات را ارائه می‌دهد؛ برخی از این ویژگی‌ها در رابطه

1 Lancaster

عصبی مصنوعی پرداخته‌اند. متغیرهای استفاده‌شده در پژوهش ایشان در قالب متغیرهای کالبدی، اجتماعی، امکانات داخلی و خارجی بنا و دسترسی است. نتایج پژوهش ایشان نشان داده است که شبکه عصبی مصنوعی دارای دقت ۹۱ درصدی در تخمین قیمت واحدهای مسکونی شهر اهواز است. همچنین، از بین عوامل تأثیرگذار بر قیمت مسکن در این شهر، زیربنای ساختمان و دسترسی بیشترین سهم را به خود اختصاص داده‌اند.

تیموری و همکاران (۱۳۹۵) در مطالعه‌ای با عنوان «برآورد قیمت مسکن شهری با استفاده از تابع هدانیک و شبکه‌های عصبی مصنوعی» به پیش‌بینی قیمت مسکن با استفاده از تابع هدانیک و شبکه عصبی مصنوعی در کوی ولیعصر شهر تبریز پرداخته‌اند. متغیرهای استفاده‌شده در پژوهش ایشان در قالب متغیرهای کالبدی، امکانات داخلی بنا و فاصله از پارک، مسجد، مراکز خرید و ایستگاه حمل‌ونقل عمومی است. یافته‌های پژوهش ایشان نشان می‌دهد که تابع هدانیک در مقایسه با شبکه عصبی مصنوعی از دقت کمتری در برآورد و پیش‌بینی قیمت مسکن برخوردار است. همچنین، بین متغیرهای فضای و قیمت مسکن در محدوده مورد مطالعه همبستگی وجود دارد.

قربانی و افقه (۱۳۹۶) در پژوهشی با عنوان «پیش‌بینی قیمت مسکن برای شهر اهواز: مقایسه مدل هدانیک با مدل شبکه عصبی مصنوعی» به تخمین قیمت مسکن پرداخته‌اند. در این راستا مدل‌های هدانیک و شبکه عصبی مصنوعی را مورد مقایسه قرار داده‌اند. متغیرهای استفاده‌شده در پژوهش ایشان در قالب متغیرهای کالبدی، امکانات داخلی بنا و دسترسی به مراکز بهداشتی، مذهبی، آموزشی و مرکز شهر است. نتایج پژوهش ایشان نشان می‌دهد که قیمت مسکن در اهواز بیشتر از عوامل فیزیکی و ساختاری تأثیر می‌پذیرد. همچنین، تفاوت قدرت پیش‌بینی دو مدل به لحاظ

با خصوصیات فیزیکی مسکن مانند مساحت، تعداد اتاق، وجود تراس و ویژگی‌هایی از این دست است (Sirmans et al., 2009: 3). بخش دیگری از ویژگی‌ها نیز در رابطه با خصوصیات محیطی مسکن است (Heyman et al., 2019: 1). در راستای موارد مطرح‌شده، اهداف پژوهش حاضر به شرح زیر است:

۱. مقایسه رگرسیون‌های خطی چندگانه^۱ و جنگل تصادفی^۲ در جهت تعریف مدل ارزیابی قیمت واحدهای مسکونی؛
۲. استخراج عوامل مؤثر بر قیمت واحدهای مسکونی.

پیشینه پژوهش

در زمینه پیش‌بینی قیمت مسکن پژوهش‌های مختلفی انجام شده که برخی از آن‌ها به شرح زیر است.

پژوهش‌های داخلی

عراقی و نوبهار (۱۳۹۰) در مطالعه‌ای با عنوان «پیش‌بینی قیمت مسکن در شهر تبریز: کاربرد مدل‌های قیمت هدانیک و شبکه عصبی مصنوعی» به پیش‌بینی قیمت مسکن با استفاده از تابع هدانیک و شبکه عصبی مصنوعی پرداخته‌اند. متغیرهای استفاده‌شده در پژوهش ایشان در قالب متغیرهای کالبدی، امکانات داخلی بنا و همچنین متغیرهای دسترسی و فواصل است. نتایج پژوهش ایشان نشان می‌دهد که عوامل فیزیکی بیشتر از عوامل مکانی، قیمت واحدهای مسکونی را تحت تأثیر قرار می‌دهد. همچنین، نتایج مقایسه دو مدل نشان داده که شبکه عصبی مصنوعی دارای خطای کمتر و کارایی بیشتری نسبت به مدل هدانیک است.

امان‌پور و همکاران (۱۳۹۳) در پژوهشی با عنوان «تخمین قیمت مسکن شهر اهواز با استفاده از شبکه عصبی» به پیش‌بینی قیمت مسکن با استفاده از شبکه

1 Multiple Linear Regression
2 Random Forest Regression

مسکن دارند. همچنین، شبکه عصبی مصنوعی به دلیل دقت در تحلیل داده و برآوردهای آماری، می‌تواند به‌عنوان جایگزین مناسبی برای تابع هدانیک در ترکیه باشد.

وارما^۲ و همکاران (۲۰۱۸) در مطالعه‌ای با عنوان «پیش‌بینی قیمت مسکن با استفاده از یادگیری ماشینی و شبکه‌های عصبی» به تخمین قیمت مسکن در بمبئی هند پرداخته‌اند. متغیرهای استفاده‌شده در پژوهش ایشان در قالب متغیرهای کالبدی، مشخصات و امکانات داخلی بنا قرار دارد. ایشان در پژوهش خود مدل‌های مختلف رگرسیون خطی، رگرسیون جنگل تصادفی، رگرسیون تقویت‌شده و شبکه عصبی را با یکدیگر ترکیب کرده‌اند. در این راستا نتایج ایشان نشان می‌دهد که ترکیب مدل‌های مختلف خطی پایین و صحت بالایی را از خود نشان می‌دهد.

ونگ و وو^۳ (۲۰۱۸) در پژوهشی با عنوان «رویکرد یادگیری ماشینی جدید در جهت تخمین قیمت مسکن» به پیش‌بینی قیمت مسکن در ویرجینیای شمالی آمریکا پرداخته‌اند. بدین منظور از رگرسیون جنگل تصادفی و رگرسیون خطی استفاده کرده‌اند. متغیرهای استفاده‌شده در پژوهش ایشان در قالب متغیرهای کالبدی است. نتایج ایشان نشان می‌دهد که رگرسیون جنگل تصادفی عملکرد بهتری نسبت به رگرسیون خطی داشته است.

چنخ^۴ و همکاران (۲۰۱۸) در مطالعه‌ای با عنوان «برآورد عملکرد جنگل تصادفی در مقابل رگرسیون چندمتغیره برای پیش‌بینی قیمت آپارتمان‌ها» به تخمین قیمت آپارتمان‌ها در لیوبلیانا^۵ پایتخت اسلوانی پرداخته‌اند. متغیرهای استفاده‌شده در پژوهش ایشان در قالب متغیرهای کالبدی، مختصات جغرافیایی،

آماری نیز معنی‌دار است. در این راستا، شبکه عصبی مصنوعی با دقت ۹۸ درصد نسبت به تابع هدانیک با دقت ۸۸ درصد، کارایی بهتری از خود نشان داده است. روستایی و همکاران (۱۳۹۹) در مطالعه‌ای با عنوان «بررسی عوامل مؤثر بر قیمت مسکن شهری با استفاده از شبکه عصبی مصنوعی» پیش‌بینی عوامل مؤثر بر قیمت مسکن شهری با استفاده از شبکه عصبی مصنوعی در منطقه دو شهر تبریز پرداخته‌اند. متغیرهای استفاده‌شده در پژوهش ایشان در دسته‌بندی متغیرهای کالبدی و دسترسی‌های مختلف است. یافته‌های پژوهش ایشان نشان می‌دهد که سهم متغیرهای کالبدی در تعیین قیمت واحدهای مسکونی برابر با ۵۳/۸ درصد، سهم متغیرهای دسترسی برابر با ۳۹/۲ درصد و سهم متغیرهای محیطی برابر با ۷ درصد است. از بین کل متغیرها، مساحت زیربنا با ۲۸/۴ درصد، فاصله از مراکز درمانی با ۶/۴ درصد، فاصله از مراکز بهداشتی با ۵/۱ درصد و نمای ساختمان با ۴/۶ درصد بیشترین سهم متغیر وابسته قیمت مسکن را به خود اختصاص داده‌اند.

پژوهش‌های خارجی

سلیم^۱ (۲۰۰۹) در پژوهشی با عنوان «عوامل تعیین‌کننده قیمت مسکن در ترکیه: رگرسیون هدانیک در مقابل شبکه عصبی مصنوعی» به برآورد قیمت مسکن پرداخته است. در این راستا، قابلیت تابع هدانیک و شبکه عصبی را در پیش‌بینی قیمت مسکن مقایسه کرده است. متغیرهای استفاده‌شده در پژوهش ایشان در دسته‌بندی موقعیتی (شهری یا روستایی)، کالبدی، مشخصات و امکانات داخلی بنا قرار دارد. نتایج پژوهش وی نشان می‌دهد که داشتن آب لوله‌کشی، استخر، نوع واحد مسکونی، تعداد اتاق، مساحت، موقعیت و مصالح بیشترین تأثیر را بر قیمت

2 Varma

3 Wang & Wu

4 Čeh

5 Ljubljana

1 Selim

در پیش‌بینی قیمت مسکن نسبت به رگرسیون خطی چندگانه داشته است.

هونگ^۷ و همکاران (۲۰۲۰) در پژوهشی با عنوان «ارزیابی قیمت مسکن بر اساس رویکرد جنگل تصادفی» به پیش‌بینی قیمت مسکن در سؤال کره جنوبی پرداخته‌اند. در این راستا، از رگرسیون جنگل تصادفی و تابع هدانیک استفاده کرده‌اند. متغیرهای استفاده‌شده در پژوهش ایشان در قالب متغیرهای کالبدی، اقتصادی، فاصله از پارک، دبیرستان، دانشگاه، بیمارستان عمومی، موزه، مترو و نواحی توسعه مجدد است. نتایج پژوهش ایشان نشان می‌دهد که رگرسیون جنگل تصادفی عملکرد بهتری نسبت به تابع هدانیک از خود نشان داده است.

بررسی پژوهش‌های داخلی نشان می‌دهد که در رابطه با قیمت مسکن، تابع هدانیک و شبکه عصبی مصنوعی بیشتر استفاده شده و از رگرسیون جنگل تصادفی در این مقوله بهره گرفته نشده است. در پژوهش‌های خارجی نیز از رگرسیون‌های خطی چندگانه و جنگل تصادفی در جهت پیش‌بینی قیمت مسکن بسیار استفاده شده است. متغیرهای پژوهش حاضر در راستای پیشینه پژوهش بوده و بعد فراکتال شاخصی جدید در ادبیات پژوهش مربوط به پیش‌بینی قیمت مسکن است. همچنین، تیموری و همکاران در پژوهش خود کوی ولیعصر (محلّه ولیعصر چهار و بخشی از محلّه ولیعصر سه) واقع در منطقه یک شهر تبریز، عراقی و نوبهار کل شهر تبریز و روستایی و همکاران نیز منطقه دو شهر تبریز را مورد بررسی قرار داده‌اند. در پژوهش حاضر نیز محله‌های یک الی چهار ولیعصر شمالی تبریز مورد بررسی قرار گرفته است.

ویژگی‌های طبیعی محل استقرار ملک، موقعیت ملک، مشخصات امکانات داخلی بنا و فاصله از کاربری‌های مختلف است. نتایج ایشان نشان می‌دهد که رگرسیون جنگل تصادفی عملکرد بهتری نسبت به تابع هدانیک داشته است.

ترنگ^۱ و همکاران (۲۰۲۰) در پژوهشی با عنوان «پیش‌بینی قیمت مسکن از طریق روش‌های بهبودیافته یادگیری ماشینی» به برآورد قیمت مسکن در پکن، پایتخت چین پرداخته‌اند. در این راستا، از تکنیک بهبودیافته یادگیری ماشینی استفاده کرده‌اند. متغیرهای استفاده‌شده در پژوهش در قالب متغیرهای موقعیت جغرافیایی، کالبدی، فاصله از مرکز شهر و مترو و همچنین مشخصات و امکانات داخلی بنا است. در پژوهش ایشان روش‌های مختلف یادگیری ماشینی شامل رگرسیون جنگل تصادفی^۲، XGBoost، LightGB، رگرسیون ترکیبی^۳ و رگرسیون تعمیم انباشته^۴ در جهت دستیابی به راه‌حل بهینه با یکدیگر مقایسه شده است. نتایج پژوهش ایشان نشان می‌دهد که هر کدام از روش‌های مذکور دارای نقاط قوت و ضعف بوده و هر روش برای اهداف مختلف عملکرد بهتری از خود نشان می‌دهد.

جوی^۵ و همکاران (۲۰۲۰) در مطالعه‌ای با عنوان «پیش‌بینی قیمت مسکن با استفاده از رگرسیون خطی و جنگل تصادفی بر اساس تکنیک‌های یادگیری ماشینی» به برآورد قیمت مسکن در داکا پایتخت بنگلادش پرداخته‌اند. متغیرهای استفاده‌شده در پژوهش ایشان در دسته‌بندی امنیتی، کالبدی، جمعیت و مساحت منطقه است. نتایج پژوهش ایشان نشان می‌دهد که رگرسیون جنگل تصادفی عملکرد بهتری

1 Truong
2 Random Forest
3 Hybrid Regression
4 Stacked Generalization Regression
5 Jui
6 Dhaka

مبانی نظری

مسکن

در دومین اجلاس اسکان بشر سال ۱۹۹۶ که در استانبول برگزار شد، مسکن مناسب چنین تعریف شده است: سرپناه مناسب تنها به معنای وجود یک سقف بالای سر هر شخص نیست؛ سرپناه مناسب یعنی آسایش مناسب، فضای مناسب، دسترسی فیزیکی و امنیت مناسب، امنیت مالک، پایداری و دوام سازه‌ای، روشنایی، تهویه و سیستم گرمایی مناسب، زیرساخت‌های اولیه مناسب از قبیل آبرسانی، بهداشت و آموزش، دفع زباله، کیفیت مناسب زیست محیطی، عوامل بهداشتی مناسب، مکان مناسب و قابل دسترسی از نظر کار و تسهیلات اولیه است که همه این موارد باید با توجه به استطاعت مردم تأمین شود؛ بنابراین مسکن چیزی بیش از یک سرپناه فیزیکی است و همه خدمات و تسهیلات عمومی لازم برای بهتریستن انسان را شامل می‌شود (پورمحمدی، ۱۳۷۹: ۳-۴). در واقع تعریف و مفهوم عام مسکن یک واحد مسکونی نیست، بلکه کل محیط مسکونی را شامل می‌شود (مخبر، ۱۳۶۳: ۱۷).

نظریه مصرف کننده لنکستر

از دید لنکستر، نظریه‌های سنتی رفتار مصرف کننده تغییرات پویای بازار را در نظر نمی‌گیرند. نظریه وی در نتیجه تلاش‌های انجام شده برای توصیف این پویایی شکل گرفته است. همچنین، در رویکرد سنتی، کالاها به طور مستقیم مطلوبیت قلمداد می‌شوند؛ اما منظور از مطلوبیت در نظریه وی، مطلوبیت حاصل از ویژگی‌های کالا است. در این رویکرد فرض بر این است که مصرف یک فعالیت است که در آن کالاها به صورت جداگانه یا ترکیبی، ورودی بوده و در آن خروجی مجموعه‌ای از ویژگی‌ها است. همچنین، این رویکرد بین کالا و ترجیحات مصرف کننده قرار دارد. بدین معنی که ویژگی‌هایی که یک کالا یا ترکیبی از کالاها دارند برای

همه مصرف کننده‌ها یکسان است که این امر، برای کالاهایی که کیفیت مشابهی دارند یک واحد اندازه‌گیری در اختیار ما قرار می‌دهد؛ بنابراین عنصر شخصی در انتخاب مصرف کننده، ویژگی‌های کالاها نیست، بلکه انتخاب از بین ویژگی‌ها است (Lancaster, 1966: 56-57).

لنکستر اساس نظریه خود را به شرح زیر خلاصه می‌کند:

۱. کالا به خودی خود برای مصرف کننده مطلوبیت ایجاد نمی‌کند؛ کالا دارای ویژگی‌هایی است و این ویژگی‌هاست که مصرف کننده را به مطلوبیت می‌رساند.
۲. به طور کلی، یک کالا بیش از یک ویژگی خواهد داشت و این تنوع ویژگی‌ها در بیش از یک کالا عرضه خواهد شد.

۳. ممکن است ویژگی‌های کالاها در صورت ترکیب متفاوت از ویژگی‌های کالاها در حالت غیر ترکیبی و جداگانه باشد.

مسکن به عنوان کالایی چندبُعدی، سبدهی از کالا و خدمات است. این کالا و خدمات را می‌توان در سه دسته خدمات فیزیکی (تعداد اتاق، مصالح، قدمت و...) محیطی (موقعیت ملک، وضعیت همسایگی و...) و دسترسی (دسترسی به مراکز خرید، تسهیلات عمومی و...) دسته‌بندی کرد. این کالا و خدمات با یکدیگر ترکیب شده و در قالب واحد مسکونی تعریف می‌شود (اکبری و همکاران، ۱۳۹۲: ۲۱). بر این اساس، یک واحد مسکونی به وسیله برداری از n ویژگی قابل اندازه‌گیری توصیف می‌شود. قیمت واحد مسکونی در بازار برآیندی از ارزش ویژگی‌های آن است. در واقع هر یک از ویژگی‌ها از ارزش ضمنی برخوردارند و ترکیب این ویژگی‌ها با ارزش‌های متفاوت، قیمت مسکن در بازار را شکل می‌دهد (یزدانی و الیاسی، ۱۳۸۰: ۸).

روش پژوهش

مدل‌سازی مطلوب، اطلاعات ۴۰۰ واحد مسکونی به‌عنوان نمونه به‌صورت میدانی جمع‌آوری شد. به‌منظور حذف اثر زمان در این پژوهش، تنها از داده‌های شهریور سال ۱۳۹۹ استفاده شد. همچنین، برای تجزیه و تحلیل داده‌ها از نرم‌افزارهای ArcMap، SPSS و RStudio استفاده شد. جدول (۱) متغیرهای مستقل پژوهش را نمایش می‌دهد. قیمت مسکن نیز به‌عنوان متغیر وابسته پژوهش است.

پژوهش حاضر از نوع کاربردی بوده و روش آن توصیفی-تحلیلی است. داده‌ها از طریق مراجعه مستقیم به مشاوران املاک و برداشت میدانی جمع‌آوری شد. جامعه آماری کلیه واحدهای مسکونی محله‌های یک تا چهار کوی ولیعصر واقع در منطقه یک شهر تبریز به تعداد ۳۰۲۷۲ واحد است. با استفاده از فرمول کوکران، ۳۷۹ نمونه در سطح اطمینان ۹۵ درصد و ضریب خطای ۵ درصد برآورد شد. به‌منظور

جدول ۱. متغیرهای مستقل پژوهش

منبع	متغیرهای دسترسی	منبع	متغیرهای فیزیکی
روستایی و همکاران، ۱۳۹۹	فاصله از مراکز درمانی	Varma et al., 2018 Jui et al., 2020	مساحت زیربنا
عراقی و نوبهار، ۱۳۹۰ روستایی و همکاران، ۱۳۹۹	فاصله از مراکز بهداشتی	تیموری و همکاران، ۱۳۹۵ روستایی و همکاران، ۱۳۹۹	نمای ساختمان
روستایی و همکاران، ۱۳۹۹	فاصله از تأسیسات شهری	عراقی و نوبهار، ۱۳۹۰؛ Wang & Wu, 2018	سال ساخت (عمر بنا)
روستایی و همکاران، ۱۳۹۹	فاصله از تجهیزات شهری	امان‌پور و همکاران، ۱۳۹۳	کل طبقات ساختمان
تیموری و همکاران، ۱۳۹۵؛ روستایی و همکاران، ۱۳۹۹	فاصله از کاربری مذهبی	تیموری و همکاران، ۱۳۹۵ روستایی و همکاران، ۱۳۹۹	تعداد واحد در هر طبقه
عراقی و نوبهار، ۱۳۹۰ قربانی و افقه، ۱۳۹۶	فاصله از کاربری آموزشی	روستایی و همکاران، ۱۳۹۹ Truong et al., 2020	استقرار واحد در طبقه
روستایی و همکاران، ۱۳۹۹ Hong et al., 2020	فاصله از مراکز آموزش عالی	Selim, 2009 Čeh et al., 2018	مصالح عمده ساختمانی
تیموری و همکاران، ۱۳۹۵ Hong et al., 2020	فاصله از پارک و فضای سبز	کوشش وطن، ۱۳۹۷ اصغری زمانی، ۱۳۹۹	شاخص فراکتال (منظم بودن قطعه زمین)
روستایی و همکاران، ۱۳۹۹	فاصله از مراکز نظامی و انتظامی	عراقی و نوبهار، ۱۳۹۰ روستایی و همکاران، ۱۳۹۹	چند بری
		عراقی و نوبهار، ۱۳۹۰ تیموری و همکاران، ۱۳۹۵	جهت استقرار ساختمان
		عراقی و نوبهار، ۱۳۹۰	عرض معبر

است (Caffo, 2019: 3). رگرسیون خطی چندگانه، یکی از روش‌های آماری است که به مدل‌سازی رابطه بین متغیرهای مستقل و وابسته بر مبنای معادله خطی در داده‌های مشاهده‌شده می‌پردازد (Abrougui et al., 2019: 204). این رگرسیون مدل گسترش‌یافته

مدل پژوهش

رگرسیون خطی چندگانه

مدل‌های رگرسیونی به‌خاطر سادگی، کم‌هزینه‌بودن، تفسیرپذیری بالا و تعمیم دقیق آن‌ها، به یکی از اولین ابزارهای حل مسئله در رشته‌های گوناگون تبدیل شده

Yeşilkanat, 2020: 3;) ارائه شده است (۲۰۰۱ این روش از نوع نظارت‌شده است که برای مسائل رگرسیون و طبقه‌بندی براساس توسعه درخت تصمیم ارائه شده است (Makhija et al., 2021: 258; Sharma & Rani, 2021: 506). زمانی که متغیر وابسته به صورت نسبتی-فاصله‌ای باشد، از درخت‌های رگرسیونی و هنگامی که متغیر وابسته اسمی-رتبه‌ای باشد، از درخت‌های طبقه‌بندی استفاده می‌شود (Sullivan, 2017: 43; Mahajan et al., 2021: 400).

این نوع رگرسیون ترکیبی از چندین درخت تصمیم است که در ساخت آن نمونه‌های خودسازمان‌یافته از داده‌ها شرکت دارند؛ منظور از روش خودسازمان‌یافته، روش نمونه‌گیری با جای‌گذاری است که در آن هیچ‌یک از داده‌های انتخاب‌شده از نمونه‌های ورودی برای تولید زیرمجموعه بعدی حذف نمی‌شود (میرهاشمی و همکاران، ۱۳۹۸: ۴۳). این الگوریتم، نوعی از مدل‌های گروهی است که پیش‌بینی درخت‌های تصمیم‌گیری را برای دستیابی به پیش‌بینی دقیق ترکیب می‌کند (Truong et al., 2020: 437). این روش با استفاده از یک درخت تصمیم ضعیف، یک یادگیرنده قوی‌تر تشکیل می‌دهد. همچنین، رگرسیون جنگل تصادفی، توانایی مناسبی را در راستایی کنترل داده‌های از دست‌رفته و نامتوازن از خود نشان می‌دهد. علاوه بر این، آن‌ها در مقایسه با سایر الگوریتم‌های یادگیری ماشینی، کمتر مستعد بیش‌برازش^{۱۱} هستند (Włodarczak, 2020: 73-74). شکل (۱) یک جنگل تصادفی متشکل از سه درخت تصمیم را نمایش می‌دهد.

رگرسیون خطی ساده است؛ به طوری که رگرسیون خطی تأثیر یک متغیر بر روی یک متغیر وابسته را نشان می‌دهد و رگرسیون چندگانه تأثیر دو یا چند متغیر را بر یک متغیر وابسته نشان می‌دهد (Mayer, 2013: 398; Abbot, 2017: 418; George & Mallery, 2020: 208). متغیرهای رگرسیونی بسته به حوزه کاربرد، اصطلاحات مختلفی دارند؛ به طوری که از متغیرهای مستقل^۱، همبسته^۲، رگرسورها^۳ و توضیح‌کننده^۴ برای توصیف متغیرهای پیش‌بینی‌کننده^۵ استفاده می‌شود. همچنین، برای متغیر پاسخ^۶، عناوینی همچون متغیر وابسته^۷، تغییرکننده^۸ و نتیجه^۹ نیز استفاده می‌شود (Korosteleva, 2019: 1). معادله عمومی رگرسیون خطی چندگانه به شرح زیر است:

$$Y_i = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n + \varepsilon_i \quad (1)$$

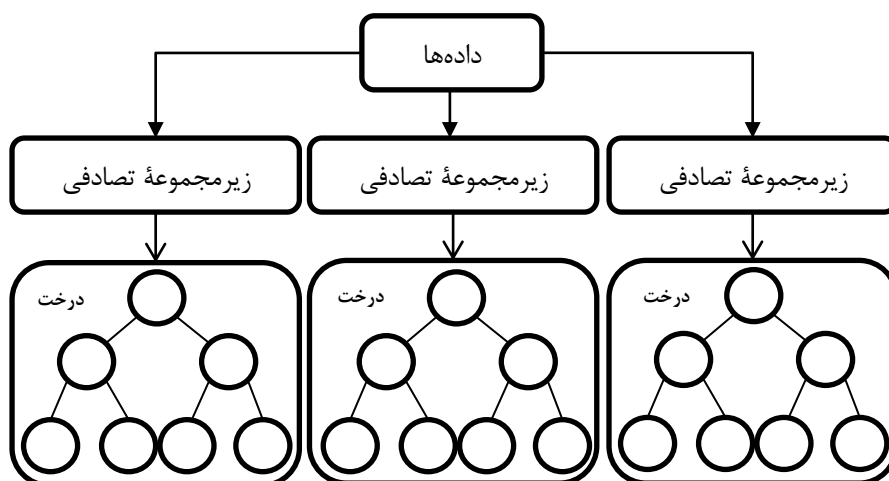
در این معادله Y_i معرف میزان پیش‌بینی‌شده برای متغیر وابسته، X معرف متغیرهای پیش‌بین، β_0 معرف عرض از مبدأ، β معرف شیب خطوط و ε_i معرف خطای مدل است (Mayer, 2013, 409; Richardson, 2015: 71; Pal & Bharati, 2019: 13).

رگرسیون جنگل تصادفی

رگرسیون جنگل تصادفی یکی از بهترین الگوریتم‌های یادگیری ماشینی است که توسط بریمن^{۱۰} در سال

- 1 Independent
- 2 Covariates
- 3 Regressors
- 4 Explanatory
- 5 Predictor
- 6 Response
- 7 Dependent
- 8 Variate
- 9 Outcome
- 10 Breiman

11 Overfitting



شکل ۱. یک جنگل تصادفی متشکل از سه درخت تصمیم (منبع: همان: ۷۴).

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (A_i - \hat{P}_i)^2} \quad (۴)$$

$$MAE = \frac{1}{N} \sum_{i=1}^N |A_i - \hat{P}_i| \quad (۵)$$

محدوده مورد مطالعه

شهرک ولیعصر در شمال شرقی تبریز و در حوزه استحفاظی منطقه یک قرار دارد. این شهرک در سال ۱۳۵۵ با حضور اعضای کمیسیون ماده ۵ به تصویب رسید. ساخت آن توسط شرکت عمران و آبادی فیروز در اراضی به مساحت ۱۶۰۰۰۰۰ مترمربع آغاز شد (السادات میرایی، ۱۳۹۱: ۹۱). این منطقه در ابتدای تأسیس با عنوان کوی ولیعهد شناخته می‌شد که پس از پیروزی انقلاب به کوی ولیعصر تغییر نام یافت (شقایق، ۱۳۹۳: ۵). شهرک ولیعصر یکی از مناطق بزرگ و پرجمعیت شهر تبریز است. این شهرک از جهت ارتفاع در مناطق مرتفع شهر قرار گرفته است. همچنین به لحاظ اقتصادی و اجتماعی نیز در سطح بالایی قرار دارد (مهندسین مشاور نقش محیط، ۱۳۹۱:

در پژوهش حاضر از ضریب تعیین^۱ (۲)، خطای لگاریتمی جذر میانگین مربع^۲ (۳)، خطای جذر میانگین مربع^۳ (۴) و خطای میانگین قدرمطلق^۴ (۵) برای ارزیابی عملکرد دو مدل رگرسیون خطی چندگانه و رگرسیون جنگل تصادفی بهره گرفته شد. در فرمول‌های زیر N بیانگر تعداد مشاهدات، A_i و \hat{P}_i و \bar{A} به ترتیب بیانگر اعداد مشاهده شده، اعداد پیش‌بینی شده و میانگین اعداد مشاهده شده است (Ahmad et al., 2017: 82; Nair & Gopi, 2020: 20; Yeşilkanat, 2020: 3; Sharma & Rani, 2021: 508; Kotta et al., 2021: 98).

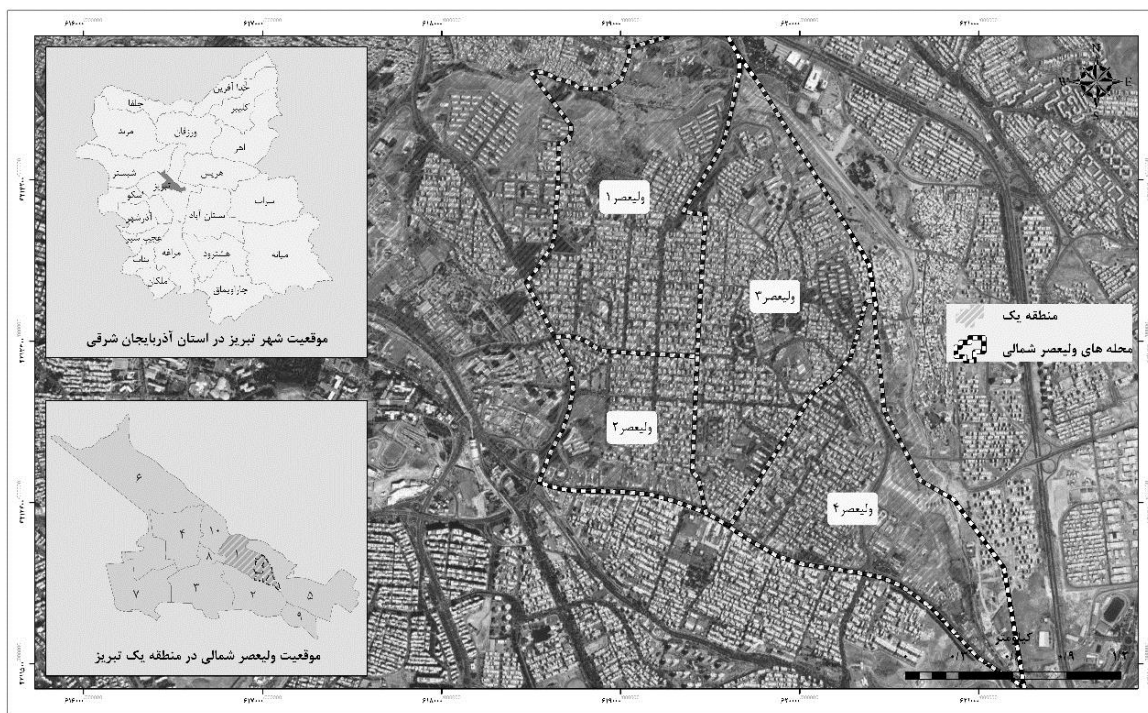
$$R^2 \left(1 - \frac{\sum_{i=1}^N (A_i - \hat{P}_i)^2}{\sum_{i=1}^N (A_i - \bar{A})^2} \right) \quad (۲)$$

$$RMSLE = \sqrt{\frac{1}{N} \sum_{i=1}^N (\log(A_i + 1) - \log(\hat{P}_i + 1))^2} \quad (۳)$$

- 1 Coefficient Of Determination
- 2 Root Mean Squared Log Error (RMSLE)
- 3 Root Mean Squared Error (RMSE)
- 4 Mean Absolute Error (MAE)

۱۵۰ هکتار، ۶۱ هکتار، ۱۵۲ هکتار و ۱۲۳ هکتار هستند. براساس بلوک های آماری شهر تبریز، جمعیت چهار محله مذکور بالغ بر ۴۵۴۴۰ نفر است.

همان طور که در شکل (۲) مشاهده می شود، این شهرک متشکل از چهار محله با نام های ولیعصر یک، دو، سه و چهار است که به ترتیب دارای مساحت



شکل ۲. موقعیت محدوده مورد مطالعه

(ترسیم: نگارندگان، ۱۳۹۹).

اسکلت بتنی است. همچنین ۵۶/۳ درصد از کل نمونه دارای مساحت زیربنای بین ۱۰۰ تا ۲۰۰ متر است. به لحاظ عرض معبر، ۷۹/۵ درصد مشرف به معبر ۶ تا ۱۲ متری است. همچنین ۵۹/۵ درصد از کل نمونه دارای قدمت ۵ ساله است.

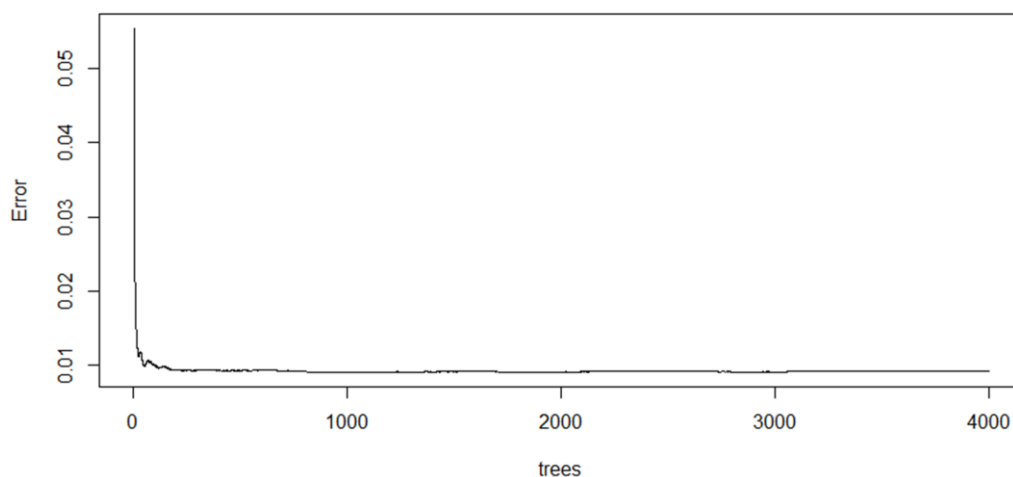
در جهت انجام پیش بینی با استفاده از رگرسیون خطی چندگانه، داده ها به دو دسته آزمایش (۳۰ درصد) و آموزش (۷۰ درصد) تقسیم شد. سپس مدل آموزش داده شده و تست شد. در جهت اجرای رگرسیون جنگل تصادفی نیز از بسته randomForest در نرم افزار RStudio استفاده شد. در این مدل نیز داده ها به دو دسته آزمایش (۳۰ درصد) و آموزش (۷۰ درصد)

یافته ها

یافته های توصیفی پژوهش نشان می دهد که از کل نمونه پژوهش، ۹۶/۵ درصد بین ۲ تا ۹ طبقه است. به لحاظ تعداد واحد در هر طبقه، ۱ واحد در هر طبقه با ۹۸ درصد دارای بیشترین سهم است. همچنین موقعیت جنوبی با ۵۷/۵ درصد دارای بیشترین سهم در جهت استقرار ساختمان است. به لحاظ چندبری، ۸۷/۳ درصد از کل نمونه پژوهش را ساختمان های یک بر تشکیل داده است. همچنین ۹۲/۵ درصد از ساختمان های نمونه نیز دارای نمای گرانیته است. به لحاظ مصالح عمده ساختمانی، ۶۷ درصد از کل نمونه دارای مصالح آجری - فلزی و ۳۲ درصد نیز دارای

بدین جهت میزان n_{tree} ۴۰۰۰ تعیین شد. میزان مناسب برای m_{try} از نسبت تعداد متغیرها بر سه حاصل می‌شود. البته عدد حاصل نقطه شروعی در جهت تنظیم این پارامتر است که می‌توان بهینه شود (Włodarczak, 2020: 73). در پژوهش حاضر میزان این پارامتر برابر با ۷ تعیین شد. برای max_depth نیز ۲۲ سطح تعیین شد. با تعیین اعداد مورد نظر برای سه پارامتر مذکور، خطا مدل دیگر کاهش پیدا نکرد.

تقسیم شد. پارامترهای مورد استفاده در رگرسیون جنگل تصادفی شامل n_{tree} ، m_{try} و max_depth است. در این مدل هرچه تعداد درخت بیشتر باشد، عملکرد آن بهتر بوده و همچنین تخمین مناسبی از اهمیت متغیرها حاصل خواهد شد (Breiman, 2002: 3; Liaw & Wiener, 2001: 21). همان‌طور که در شکل (۳) مشاهده می‌شود، در رگرسیون جنگل تصادفی پژوهش حاضر با افزایش تعداد درخت خطای مدل کاهش یافته است. این خطا بعد از ۳۰۰۰ درخت ثابت شده و دارای نوسان نیست.



شکل ۳. رابطه بین خطا و تعداد درخت در جنگل تصادفی (منبع: یافته‌های پژوهش، ۱۳۹۹).

۱ بیانگر تعداد درخت در جنگل تصادفی است.

۲ بیانگر تعداد متغیرهای موجود برای تقسیم در هر گره درخت است.

۳ بیانگر عمق و سطوح هر درخت در جنگل تصادفی است.

با ۰/۰۱۵۰ و ۰/۰۰۹۱ است. بر اساس خطای RMSE، خطای رگرسیون خطی چندگانه و رگرسیون جنگل تصادفی در داده های آموزش به ترتیب برابر با ۰/۰۷۵۰ و ۰/۰۳۰۹ است. همچنین، خطای مذکور این دو مدل در داده های آزمایش به ترتیب برابر با ۰/۰۶۹۴ و ۰/۰۴۰۶ است. بر اساس خطای MAE، خطای رگرسیون خطی چندگانه و رگرسیون جنگل تصادفی در داده های آموزش به ترتیب برابر با ۰/۰۵۷۵ و ۰/۰۲۱۹ است. همچنین خطای مذکور این دو مدل در داده های آزمایش به ترتیب برابر با ۰/۰۵۱۷ و ۰/۰۲۹۸ است. بنا بر نتایج فوق، مشخص می شود که رگرسیون جنگل تصادفی با خطای کمتر نسبت به رگرسیون خطی چندگانه عملکرد بهتری از خود نشان داده است.

همان طور که در جدول (۲) مشاهده می شود، نتایج حاصل از داده های آموزش و آزمایش رگرسیون خطی چندگانه و رگرسیون جنگل تصادفی آورده شده است. در این راستا، نتایج ضریب تعیین رگرسیون خطی چندگانه و جنگل تصادفی در داده های آموزش به ترتیب برابر با ۰/۸۹۰۱ و ۰/۹۸۳۸ است. همچنین، نتایج ضریب تعیین در داده های آزمایش برای رگرسیون خطی چندگانه و رگرسیون جنگل تصادفی به ترتیب برابر با ۰/۹۱۲۶ و ۰/۹۸۲۴ است. بر اساس خطای RMSLE، خطای رگرسیون خطی چندگانه و رگرسیون جنگل تصادفی در داده های آموزش به ترتیب برابر با ۰/۰۱۶۱ و ۰/۰۰۶۶ است. همچنین، خطای مذکور این دو مدل در داده های آزمایش به ترتیب برابر

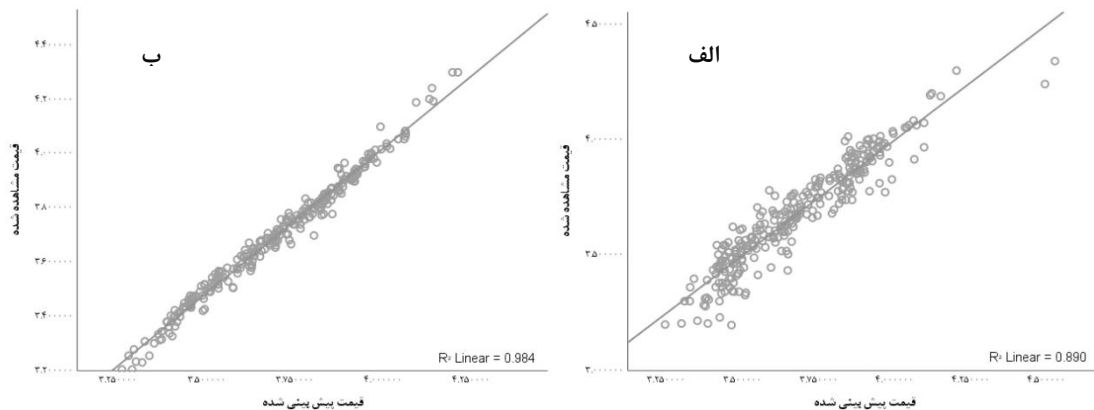
جدول ۲. عملکرد رگرسیون های خطی چندگانه و رگرسیون جنگل تصادفی

مدل	رگرسیون خطی چندگانه	رگرسیون جنگل تصادفی	رگرسیون خطی چندگانه	رگرسیون جنگل تصادفی
شاخص	آموزش		آزمایش	
R^2	۰/۸۹۰۱	۰/۹۸۳۸	۰/۹۸۲۴	۰/۹۸۳۸
RMSLE	۰/۰۱۶۱	۰/۰۰۶۶	۰/۰۰۹۱	۰/۰۰۹۱
RMSE	۰/۰۷۵۰	۰/۰۳۰۹	۰/۰۴۰۶	۰/۰۴۰۶
MAE	۰/۰۵۷۵	۰/۰۲۱۹	۰/۰۲۹۸	۰/۰۲۹۸

(منبع: یافته های پژوهش، ۱۳۹۹).

است. معادله خط داده های آموزش رگرسیون خطی چندگانه برابر با $y = 0/8901x + 0/4046$ است. همچنین، معادله خط داده های آموزش الگوریتم جنگل تصادفی نیز برابر با $y = 0/9215x + 0/2908$ است.

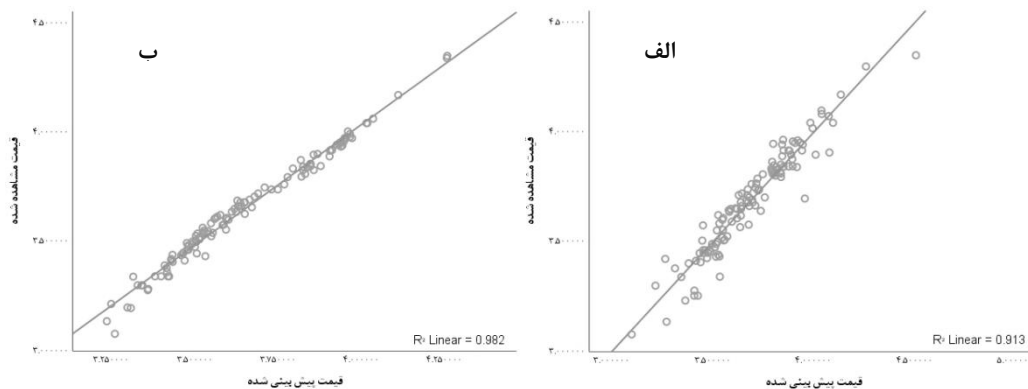
شکل (۴) نمودار ضریب تعیین رگرسیون خطی چندگانه و جنگل تصادفی را در داده های آموزش نشان می دهد. در این نمودارها محور X نشانگر قیمت پیش بینی شده و محور Y نشانگر قیمت مشاهده شده



شکل ۴. الف) نمودار ضریب تعیین داده‌های آموزش رگرسیون خطی؛ ب) نمودار ضریب تعیین داده‌های آموزش الگوریتم جنگل تصادفی (منبع: یافته‌های پژوهش، ۱۳۹۹).

است. معادله خط داده‌های آزمایش رگرسیون خطی، چندگانه برابر با $y = 0/9126x + 0/3227$ است. همچنین معادله خط داده‌های آزمایش الگوریتم جنگل تصادفی نیز برابر با $y = 0/8895x + 0/4028$ است.

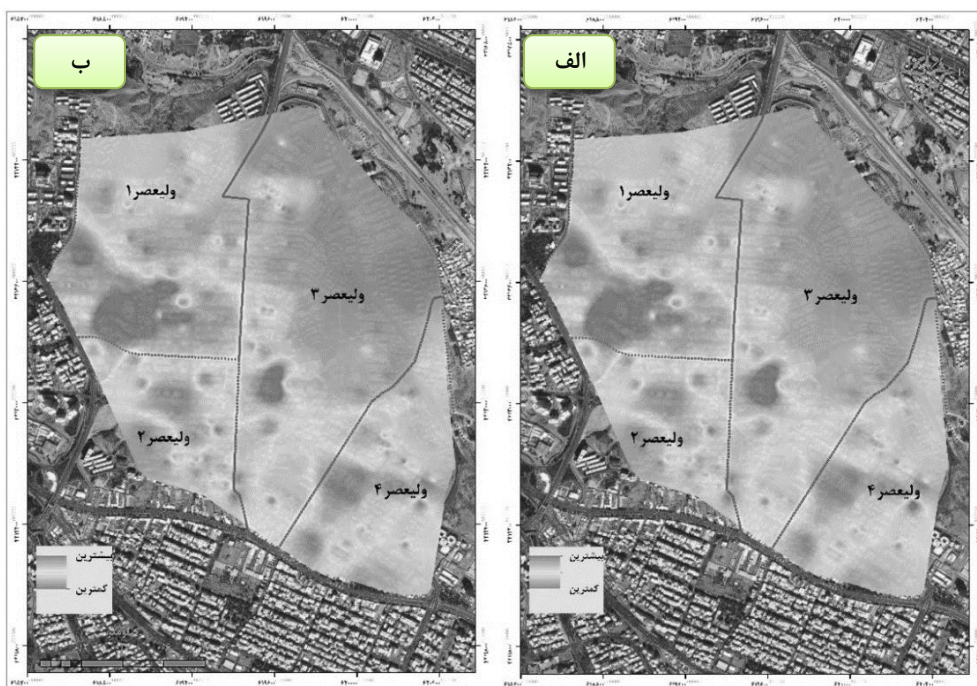
شکل (۵) نمودار ضریب تعیین رگرسیون خطی چندگانه و جنگل تصادفی را در داده‌های آزمایش نشان می‌دهد. در این نمودارها محور X نشانگر قیمت پیش‌بینی شده و محور Y نشانگر قیمت مشاهده‌شده



شکل ۵. الف) نمودار ضریب تعیین داده‌های آزمایش رگرسیون خطی؛ ب) نمودار ضریب تعیین داده‌های آزمایش الگوریتم جنگل تصادفی (منبع: یافته‌های پژوهش، ۱۳۹۹).

ولیعصر شمالی دارای قیمت بیشتری است؛ علاوه بر این، پیش‌بینی صورت گرفته با جنگل تصادفی نیز الگوی فضایی یکسانی را با قیمت مشاهده‌شده دنبال کرده است. همچنین، همبستگی این دو نقشه نیز در حد همبستگی قوی ۰/۹۸ است.

به جهت عملکرد بهتر رگرسیون جنگل تصادفی نسبت به رگرسیون خطی چندگانه، تنها نقشه حاصل از آن ارائه شد. شکل (۶) نقشه قیمت مشاهده‌شده و پیش‌بینی شده توسط رگرسیون جنگل تصادفی را نشان می‌دهد. همان‌طور که در این شکل مشاهده می‌شود، می‌توان گفت که حدود مرکز و قسمت‌های غربی



شکل ۶. الف) نقشه درون یابی قیمت مشاهده شده؛ ب) نقشه درون یابی قیمت پیش بینی شده (منبع: یافته های پژوهش، ۱۳۹۹).

درصد، فاصله از تأسیسات شهری با ۶/۸۴ درصد، فاصله از فضای سبز با ۵/۰۳ درصد، فاصله از کاربری مذهبی با ۴/۴۸ درصد، فاصله از مراکز درمانی با ۴/۲۶ درصد، فاصله از مراکز نظامی و انتظامی با ۳/۹۹ درصد و استقرار واحد در طبقه با ۳/۹۲ درصد به ترتیب مؤثرترین ده متغیر در رابطه با قیمت مسکن محله های ولیعصر شمالی هستند.

از آنجایی که رگرسیون جنگل تصادفی نسبت به رگرسیون خطی چندگانه عملکرد بهتری نشان داد، تنها متغیرهای مهم مستخرج از رگرسیون جنگل تصادفی ارائه می شود. همان طور که در جدول (۳) مشاهده می شود، متغیر مساحت با ۲۴/۷۸ درصد، متغیرهای کل طبقات ساختمان با ۹/۸۶ درصد، ساخت با ۷/۹ درصد، فاصله از مراکز بهداشتی با ۶/۸۹

جدول ۳. درصد اهمیت شاخص های پژوهش براساس رگرسیون جنگل تصادفی

درصد اهمیت	متغیرهای دسترسی	درصد اهمیت	متغیرهای فیزیکی
۴/۲۶	فاصله از مراکز درمانی	۲۴/۷۸	مساحت زیربنا
۶/۸۹	فاصله از مراکز بهداشتی	۰/۹۲	نمای ساختمان
۶/۸۴	فاصله از تأسیسات شهری	۷/۹	سال ساخت (عمر بنا)
۱/۷۹	فاصله از تجهیزات شهری	۹/۸۶	کل طبقات ساختمان
۴/۴۸	فاصله از کاربری مذهبی	۰/۷۶	تعداد واحد در هر طبقه
۳/۰۶	فاصله از کاربری آموزشی	۳/۹۲	استقرار واحد در طبقه
۵/۷۹	فاصله از مراکز آموزش عالی	۰/۵۴	مصالح عمده ساختمانی
۵/۰۳	فاصله از پارک و فضای سبز	۲/۱۶	شاخص فراکتال (منظم بودن قطعه زمین)
۳/۹۹	فاصله از مراکز نظامی و انتظامی	۰/۹۷	چندبری
		۲/۲	جهت استقرار ساختمان
		۳/۸۶	عرض معبر

(منبع: یافته های پژوهش، ۱۳۹۹)

منطقه ای بالایی داشته باشد؛ زیرا باید مسکن بتواند مرغوبیت کالبدی حاصل از نوسازی را در قیمت بازتاب دهد؛ از این رو، سال ساخت به عنوان سومین متغیر اثرگذار در قیمت مسکن شهرک ولیعصر تبریز حاصل شده است.

همچنین اکثر متغیرهایی که دارای پراکندگی کم‌تری در سطح نمونه هستند، توسط رگرسیون جنگل تصادفی دارای اهمیت کمتری شده‌اند. از جمله متغیر تعداد واحد در هر طبقه است. همان‌طور که مشخص شد، اکثر نمونه‌های پژوهش آپارتمان‌های مسکونی تک‌واحدی (۹۸ درصد نمونه) هستند که بدین امر این متغیر دارای اهمیت کمتری در رابطه با قیمت زمین در این محدوده مشخص شده است.

نتیجه‌گیری

پژوهش حاضر با دو هدف مقایسه رگرسیون خطی چندگانه و جنگل تصادفی در جهت تعریف مدل ارزیابی قیمت واحدهای مسکونی و استخراج عوامل مؤثر بر قیمت واحدهای مسکونی طراحی شد. در این راستا متغیرهای پژوهش از ادبیات پژوهش استخراج شد. متغیرهای مورداستفاده در دو دسته متغیرهای فیزیکی و دسترسی قرار داده شد.

داده‌های موردنیاز از طریق مراجعه مستقیم به مشاوران املاک و برداشت میدانی جمع‌آوری شد. در جهت مدل‌سازی مطلوب، اطلاعات ۴۰۰ واحد مسکونی به‌عنوان نمونه توسط پیمایش جمع‌آوری شد. به‌منظور حذف اثر زمان در این پژوهش، تنها از داده‌های شهریور ۱۳۹۹ استفاده شد.

در جهت انجام پیش‌بینی، نمونه‌های پژوهش در دودسته آزمایش (۳۰ درصد) و آموزش (۷۰ درصد) قرار داده شد. سپس هر دو مدل آموزش داده شده و آزمایش شدند. در این راستا، نتایج ضریب تعیین رگرسیون خطی چندگانه و جنگل تصادفی در داده‌های

ساختمان‌های شهرک ولیعصر با توجه به قدمت تأسیس و ساخت، از لحاظ فیزیکی مشابهت‌های زیادی به هم دارند، اما باین حال تفاوت قیمتی نیز در بین آن‌ها کاملاً مشهود است. عموماً قیمت مسکن در مفهوم بازاری آن، قیمت به ازای هر متر مورد بررسی و ارزیابی قرار می‌گیرد. به طوری که هرچقدر مساحت و زیربنای یک مسکن بیشتر باشد قیمت مسکن نیز بیشتر خواهد شد. از آنجایی که شهرک ولیعصر تبریز، به‌عنوان یک بافت برنامه‌ریزی شده است، فراوانی مساحت قطعات نیز در بازه معین قرار دارد؛ از این رو انتظار داشتن سهم بالای متغیر مساحت در تعیین قیمت مسکن، منطقی است. همان‌طور که مشاهده شد، متغیر مساحت در پژوهش‌های پیشین نیز در بین متغیرهای مهم در رابطه با قیمت مسکن مشخص شده است؛ علاوه بر این، شهرک ولیعصر به دلیل ارزش منطقه‌ای بالایی که دارد، همواره مورد توجه آپارتمان‌سازی بوده و هست؛ به طوری که الگوی آپارتمان‌سازی، الگوی غالب ساختمان‌های مسکونی در این شهرک است. به هر میزان که تعداد طبقات یک ساختمان مسکونی افزایش یابد، قیمت کل آن ساختمان نیز افزایش خواهد داشت؛ علاوه بر این، قیمت طبقات مختلف با یکدیگر متفاوت است، به طوری که قیمت طبقه اول به نسبت کمتر از طبقه‌های بالاتر است. همچنین، با افزایش موقعیت واحد در طبقه قیمت آن نیز با ضریب معینی افزایش پیدا می‌کند؛ از این رو متغیر کل طبقات ساختمان نیز در این پژوهش درجه اهمیت بالایی به خود اختصاص داده است. همچنین، به میزانی که مسکن نوساز باشد، عمر مفید آن نیز بیشتر است؛ بنابراین این معادله در قیمت مسکن نیز حائز اهمیت است. از آنجایی که عمر بنای ساختمان‌های مسکونی در این شهرک عمدتاً نوساز است (حدود ۶۰ درصد از نمونه‌ها دارای قدمت ۵ ساله است) و نوسازی نیز عمدتاً در محله‌هایی صورت می‌گیرد که ارزش

درصد، فاصله از مراکز درمانی با ۴/۲۶ درصد، فاصله از مراکز نظامی و انتظامی با ۳/۹۹ درصد و استقرار واحد در طبقه با ۳/۹۲ درصد به ترتیب مؤثرترین ده متغیر در رابطه با قیمت مسکن محله های ولیعصر شمالی هستند.

در نهایت می توان مشاهده کرد که نتایج فوق، هم راستا با نتایج امان پور و همکاران (۱۳۹۳)، قربانی و افقه (۱۳۹۶)، روستایی و همکاران (۱۳۹۹) و سلیم (۲۰۰۹) می باشد. در این پژوهش ها نیز مساحت زیربنا، فاصله از مراکز درمانی و بهداشتی به عنوان متغیرهای مهم در رابطه با قیمت مسکن گزارش شده اند.

منابع

اصغری زمانی، اکبر؛ روستایی، شهریور؛ کوشش وطن، محمدعلی. (۱۳۹۹). ارزیابی تفکیک اراضی مسکونی و تجاری از منظر شاخص های تفکیک زمین و ذی نفعان عرصه زمین؛ مطالعه موردی: منطقه یک و سه شهر تبریز، جغرافیا و برنامه ریزی، صاحب امتیاز: دانشگاه تبریز، دوره بیست و چهارم، شماره ۷۴، صص ۲۸-۱۳.

https://geoplanning.tabrizu.ac.ir/article_10793.html.

اکبری، نعمت اله؛ خوش اخلاق، رحمان؛ سارا مردیها، سارا (۱۳۹۲). سنجش و ارزش گذاری عوامل مؤثر بر انتخاب مسکن با استفاده از روش انتخاب تجربی از دیدگاه خانوارهای ساکن در بافت فرسوده شهر اصفهان، فصلنامه پژوهش های اقتصادی، صاحب امتیاز: دانشگاه علامه طباطبائی، دوره سیزدهم، شماره ۳، صص ۴۷-۱۹.

<https://ecor.modares.ac.ir/article-18-10602-fa.html>.

امان پور، سعید؛ سلیمانی راد، اسماعیل؛ کشتکار، لیلا؛ مختاری چلچله، صادق. (۱۳۹۳). تخمین قیمت مسکن شهر اهواز با استفاده از شبکه عصبی، فصلنامه علمی - پژوهشی اقتصاد و مدیریت شهری، صاحب امتیاز: انجمن علمی اقتصاد شهری ایران، دوره سوم، شماره ۹، صص ۵۷-۴۵.

https://iueam.ir/browse.php?a_id=105&sid=1&slc_1_ang=fa.

پورمحمدی، محمدرضا. (۱۳۷۹). برنامه ریزی مسکن، تهران: انتشارات سمت.

<https://samta.samt.ac.ir/content/10806>.

آموزش به ترتیب برابر با ۰/۸۹۰۱ و ۰/۹۸۳۸ است. همچنین، نتایج ضریب تعیین در داده های آزمایش برای رگرسیون خطی چندگانه و رگرسیون جنگل تصادفی به ترتیب برابر با ۰/۹۱۲۶ و ۰/۹۸۲۴ است. براساس خطای RMSLE، خطای رگرسیون خطی چندگانه و رگرسیون جنگل تصادفی در داده های آموزش به ترتیب برابر با ۰/۰۱۶۱ و ۰/۰۰۶۶ است. همچنین، خطای مذکور این دو مدل در داده های آموزش به ترتیب برابر با ۰/۰۱۵۰ و ۰/۰۰۹۱ است. براساس خطای RMSE، خطای رگرسیون خطی چندگانه و رگرسیون جنگل تصادفی در داده های آموزش به ترتیب برابر با ۰/۰۷۵۰ و ۰/۰۳۰۹ است. همچنین، خطای مذکور این دو مدل در داده های آموزش به ترتیب برابر با ۰/۰۶۹۴ و ۰/۰۴۰۶ است. براساس خطای MAE، خطای رگرسیون خطی چندگانه و رگرسیون جنگل تصادفی در داده های آموزش به ترتیب برابر با ۰/۰۵۷۵ و ۰/۰۲۱۹ است. همچنین، خطای مذکور این دو مدل در داده های آموزش به ترتیب برابر با ۰/۰۵۱۷ و ۰/۰۲۹۸ است. باتوجه به نتایج حاصل از خطاهای مذکور، مشخص شد که رگرسیون جنگل تصادفی با خطای کمتر نسبت به رگرسیون خطی چندگانه عملکرد بهتری از خود نشان داده است.

این نتیجه، هم راستا با نتایج ونگ و وو (۲۰۱۸)، چنخ و همکاران (۲۰۱۸) و جوی و همکاران (۲۰۲۰) است. در این پژوهش ها رگرسیون جنگل تصادفی عملکرد بهتری نسبت به رگرسیون خطی چندگانه داشته است. براساس نتایج رگرسیون جنگل تصادفی، متغیر مساحت با ۲۴/۷۸ درصد، متغیرهای کل طبقات ساختمان با ۹/۸۶ درصد، سال ساخت با ۷/۹ درصد، فاصله از مراکز بهداشتی با ۶/۸۹ درصد، فاصله از تأسیسات شهری با ۶/۸۴ درصد، فاصله از فضای سبز با ۵/۰۳ درصد، فاصله از کاربری مذهبی با ۴/۴۸

<https://www.sid.ir/fa/journal/ViewPaper.aspx?id=346973>.
کوشش وطن، محمدعلی. (۱۳۹۷). بررسی تطبیقی نقش کیفیت تفکیک زمین در بهره‌وری بهینه از کاربری اراضی در مادرشهرهای ایران، نمونه موردی: منطقه یک و سه مادرشهر تبریز، کارشناسی ارشد، دانشگاه تبریز.

<https://ganj.irandoc.ac.ir/#/articles/0da0e12bdf45730035babcac7d8a2f9f/search/75bb541ae1e937a22ae06535f0b38925>.

مخبر، عباس. (۱۳۶۳). ابعاد اجتماعی مسکن، ترجمه مرکز مدارک اقتصادی اجتماعی سازمان برنامه‌ریزی، تهران.

<https://www.gisoom.com/book/144209>.

مهندسين مشاور نقش محیط. (۱۳۹۱). طرح توسعه و عمران شهر تبریز، گزارش مطالعات اجتماعی مرحله موجود.

میرهاشمی، حمید؛ یاراحمدی، داریوش؛ شرفی، سیامک؛ فرزین، سعید. (۱۳۹۸). بهبود عملکرد ماشین بردار پشتیبان و الگوریتم جنگل تصادفی در پیش‌بینی جریان رودخانه خرم‌آباد با استفاده از نویززدایی غیریکساخت داده‌ها و الگوریتم سیمپلکس، علوم مهندسی آبخیزداری ایران، صاحب امتیاز: انجمن آبخیزداری ایران، دوره سیزدهم، شماره ۴۷، صص ۵۱-۴۰.

<http://jwmsei.ir/article-1-846-fa.html>.

یزدانی، فردین؛ الیاسی، طهمورث. (۱۳۸۰). بررسی اقتصادی عرضه و تقاضای مسکن در مناطق شهری استان اصفهان، سازمان مدیریت و برنامه‌ریزی استان اصفهان، معاونت برنامه‌ریزی و هماهنگی.

Abbot, M. L. (2017). Using Statistics in the Social and Health Sciences with SPSS and Excel. John Wiley & Sons.

DOI: 10.1002/9781119121077.

Abrougui, K. & Gabsi, K. & Mercatoris, B. & Khemis, C. & Amami, R. & Chehaibi, Sa. (2019). Prediction of organic potato yield using tillage systems and soil properties by artificial neural network and multiple linear regressions. Soil and Tillage Research. Vol. 190. pp 202-208.

DOI: 10.1016/j.still.2019.01.011.

Ahmad, Muhammad & Mourshed, Monjur & Rezgui, Yacine. (2017). Trees vs Neurons: Comparison between random forest and ANN for high-resolution prediction of building energy consumption. Energy and Buildings. 147.

تیموری، ایرج؛ سلطان قیس، نوید؛ قلی زاده؛ یاسر. (۱۳۹۶). برآورد قیمت مسکن شهری با استفاده از تابع هدانیک و شبکه‌های عصبی مصنوعی مورد شناسی: کوی ولیعصر شهر تبریز، فصلنامه جغرافیا و آمایش شهری- منطقه‌ای، صاحب امتیاز: دانشگاه سیستان و بلوچستان، دوره هفتم، شماره ۲۲، صص ۵۶-۴۱.

https://gajj.usb.ac.ir/article_2995.html.

خلیلی عراقی، منصور؛ نوبهار، الهام. (۱۳۹۰). پیش‌بینی قیمت مسکن در شهر تبریز: کاربرد مدل‌های قیمت هدانیک و شبکه عصبی مصنوعی، فصلنامه پژوهش‌ها و سیاست‌های اقتصادی، صاحب امتیاز: معاونت سیاست‌گذاری اقتصادی- وزارت امور اقتصادی و دارایی، دوره نوزدهم، شماره ۶۰، صص ۱۳۸-۱۱۳.

<http://qjerp.ir/article-1-189-fa.html>.

روستایی، شهرپور؛ تیموری، ایرج؛ نعمتی، محمد. (۱۳۹۹). بررسی عوامل مؤثر بر قیمت مسکن شهری با استفاده از شبکه عصبی مصنوعی مورد شناسی: منطقه دو تبریز، فصلنامه جغرافیا و توسعه، صاحب امتیاز: دانشگاه سیستان و بلوچستان، دوره هجدهم، شماره ۵۹، صص ۱۴۸-۱۲۹.

https://gdij.usb.ac.ir/article_5464.html.

السادات میرایی، نفیسه. (۱۳۹۱). آسیب‌شناسی و ظرفیت‌سنجی شبکه معابر شهری با رویکرد مدیریت بحران، مطالعه موردی: شهرک ولیعصر تبریز، پایان‌نامه ارشد، دانشگاه پیام‌نور تهران، دانشکده علوم اجتماعی.

<https://ganj.irandoc.ac.ir/#/articles/f37a8cc7cb3aae98dc778e10abf45462/search/fea4309690cb55816c5bc82ba8f61773>.

شقایق، حسن. (۱۳۹۳). تحلیلی بر نقش و جایگاه روشنایی معابر شهری در برنامه‌ریزی شهری، مطالعه موردی: کوی ولیعصر و محله عباسی شهر تبریز، پایان‌نامه ارشد، دانشگاه تبریز، پردیس ارس.

<https://ganj.irandoc.ac.ir/#/articles/6990d8bf49347c04a209b6947441fd6e/search/97eed76bdbb89481a2484fdcc25eb85d>.

قربانی، سالار؛ افقه، سید مرتضی. (۱۳۹۶). پیش‌بینی قیمت مسکن برای شهر اهواز: مقایسه مدل هدانیک با مدل شبکه عصبی مصنوعی، فصلنامه اقتصاد و مدیریت شهری، صاحب امتیاز: انجمن علمی اقتصاد شهری ایران، دوره پنجم، شماره ۱۹، صص ۴۴-۲۹.

DOI: 10.3846/ijspm.2020.11544.

Jui J.J., Imran Molla M.M., Bari B.S., Rashid M., Hasan M.J. (2020) Flat Price Prediction Using Linear and Random Forest Regression Based on Machine Learning Techniques. In: Mohd Razman M., Mat Jizat J., Mat Yahya N., Myung H., Zainal Abidin A., Abdul Karim M. (eds) Embracing Industry 4.0. Lecture Notes in Electrical Engineering, vol 678. Springer, Singapore.

https://doi.org/10.1007/978-981-15-6025-5_19.

Korosteleva, O. (2019). Advanced Regression Models with SAS and R. CRC Press.

DOI: 10.1201/9781315169828.

Kotta H., Pardasani K., Pandya M., Ghosh R. (2021) Optimization of Loss Functions for Predictive Soil Mapping. In: Hassanien A., Bhatnagar R., Darwish A. (eds) Advanced Machine Learning Technologies and Applications. AMLTA 2020. Advances in Intelligent Systems and Computing, vol 1141. Springer, Singapore.

https://doi.org/10.1007/978-981-15-3383-9_9.

Lancaster K.J. (1966). A New Approach to Consumer Theory, Journal of Political Economy, Vol. 74. No. 2. Pages 132-157.

doi.org/10.1086/259131.

Leung, C. (2004). Macroeconomics and housing: a review of the literature, Journal of Housing Economics, Vol.13. Pages 249-267.

<https://doi.org/10.1016/j.jhe.2004.09.002>.

Liaw, Andy & Wiener, Matthew. (2001). Classification and Regression by RandomForest. RNews. 23.

https://www.researchgate.net/publication/228451484_Classification_and_Regression_by_RandomForest.

Mahajan U., Krishnan A., Malhotra V., Sharma D., Gore S. (2021). Predicting Competitive Weightlifting Performance Using Regression and Tree-Based Algorithms. In: Hassanien A., Bhatnagar R., Darwish A. (eds) Advanced Machine Learning Technologies and Applications. AMLTA 2020. Advances in Intelligent Systems and Computing, vol 1141. Springer, Singapore.

https://doi.org/10.1007/978-981-15-3383-9_36.

Makhija R., Ali S., Jaya Krishna R. (2021) Detecting Influencers in Social Networks Through Machine Learning Techniques. In: Hassanien A., Bhatnagar R., Darwish A. (eds) Advanced Machine Learning Technologies and Applications. AMLTA 2020. Advances in Intelligent Systems and Computing, vol 1141. Springer, Singapore.

DOI: 10.1016/j.enbuild.2017.04.038.

Breiman, L. (2002). Manual on setting up, using, and understanding random forests v3.1.

https://www.stat.berkeley.edu/~breiman/Using_random_forests_V3.1.pdf

Caffo, B. (2019). Regression Models for Data Science in R: A companion book for the Coursera Regression Models class. Leanpub.

<https://leanpub.com/regmods>.

Case, B., Clapp, J., Dubin, R. Rodriguez, M. (2004). Modeling Spatial and Temporal House Price Patterns: A Comparison of Four Models. The Journal of Real Estate Finance and Economics 29, 167-191.

<https://doi.org/10.1023/B:REAL.0000035309.60607.53>.

Case, K. & Quigley, J. & Shiller, R. (2005). «Comparing Wealth Effects: The Stock Market versus the Housing Market," Advances in Macroeconomics, Berkeley Electronic Press, vol. 5(1), pages 1-32. DOI: 10.3386/w8606.

Cozmei, C., & Onofrei, M. (2012). Impact of Property Taxes on Commercial Real Estate Competition in Romania, Journal of Procedia Economics and Finance, Vol3. Pages 604-610.

[https://doi.org/10.1016/S2212-5671\(12\)00202-X](https://doi.org/10.1016/S2212-5671(12)00202-X).

George, D. & Mallery, P. (2020). IBM SPSS Statistics 26 Step by Step: A Simple Guide and Reference. 16th Edition. Routledge: Taylor & Francis Group.

<https://www.routledge.com/IBM-SPSS-Statistics-26-Step-by-Step-A-Simple-Guide-and-Reference/George-Mallery/p/book/9780367174354>.

Giussani, Andrea (2020), Applied Machine Learning with Python, Milano, EGEA S.p.A.

https://books.google.com/books/about/Applied_Machine_Learning_with_Python.html?id=IN5AygEACAAJ.

Heyman, A., & Sommervoll, D.E. (2019). House prices and relative location. Cities, 95, 1-14.

<https://doi.org/10.1016/j.cities.2019.06.004>.

Hong, Jengei & Choi, Heeyoul «Henry & Kim, Woo-sung. (2020). A house price valuation based on the random forest approach: the mass appraisal of residential property in South Korea. International Journal of Strategic Property Management. 24. 1-13.

- Advanced Machine Learning Technologies and Applications. AMLTA 2020. Advances in Intelligent Systems and Computing, vol 1141. Springer, Singapore.
- https://doi.org/10.1007/978-981-15-3383-9_46.
- Sirmans, G., MacPherson, D., & Zietz, E. (2009). The Composition of Hedonic Pricing Models. *Journal of Real Estate Literature*, 13, 3-43.
- <https://www.semanticscholar.org/paper/The-Composition-of-Hedonic-Pricing-Models-Sirmans-MacPherson/2fde22eba11fad9e671eacb588353f123df0d3ac>
- Sullivan, W. (2017). *Machine learning Beginners Guide Algorithms: Supervised & Unsupervised learning, Decision Tree & Random Forest Introduction*. Healthy Pragmatic Solutions Inc.
- https://books.google.com/books/about/Machine_Learning_For_Beginners_Guide_Algorithms.html?id=v6saxAEACAAJ&source=kp_book_description
- Truong, Quang & Nguyen, Minh & Dang, Hy & Mei, B. (2020). Housing Price Prediction via Improved Machine Learning Techniques. *Procedia Computer Science*. 174. 433-442.
- DOI: org/10.1016/j.procs.2020.06.111.
- Varma, A., & Sarma, A., & Doshi, S., & Nair, R. (2018). House Price Prediction Using Machine Learning and Neural Networks, 2018 Second International Conference on Inventive Communication and Computational Technologies, Coimbatore, pp. 1936-1939.
- DOI: org/10.1109/ICICCT.2018.8473231.
- Wang, C. & Wu, H. (2018). A new machine learning approach to house price estimation. *New Trends in Mathematical Science*. 4. 165-171.
- DOI: 10.20852/ntmsci.2018.327.
- Wlodarczak, Peter (2020). *Machine Learning and its applications*, First edition. Boca Raton, FL: CRC Press/Taylor & Francis Group.
- DOI: 10.1201/9780429448782.
- Yeşilkanat, C. (2020). Spatio-temporal estimation of the daily cases of COVID-19 in worldwide using random forest machine learning algorithm. *Chaos, Solitons & Fractals*. 140. 110210.
- DOI: 10.1016/j.chaos.2020.110210.
- https://doi.org/10.1007/978-981-15-3383-9_23.
- Čeh Marjan & Kilibarda, Milan & Liseć, Anka & Bajat, Branislav. (2018). Estimating the Performance of Random Forest versus Multiple Regression for Predicting Prices of the Apartments. *ISPRS International Journal of Geo-Information*. 7. 168.
- DOI: 10.3390/ijgi7050168.
- Masri, Hilmi & Nawawi, Abdul & Sipan, Ibrahim. (2016). Review of Building, Locational, Neighbourhood Qualities Affecting House Prices in Malaysia. *Procedia - Social and Behavioral Sciences*. 234. 452-460.
- DOI: 10.1016/j.sbspro.2016.10.263.
- Mayer, A. (2013). *Introduction to Statistics and SPSS in Psychology*, Edinburgh: Pearson Higher Ed.
- https://books.google.com/books/about/Introduction_to_Statistics_and_SPSS_in_P.html?id=SDHtMgEACAAJ.
- Nair S.N., Gopi E.S. (2020) Deep Learning Techniques for Crime Hotspot Detection. In: Kulkarni A., Satapathy S. (eds) *Optimization in Machine Learning and Applications. Algorithms for Intelligent Systems*. Springer, Singapore.
- https://doi.org/10.1007/978-981-15-0994-0_2.
- Pal, M. & Bharati, P. (2019). *Applications of Regression Techniques*. Springer Singapore.
- <https://www.springer.com/gp/book/9789811393136>.
- Richardson, R. (2015). *Business Applications of Multiple Regression, Second Edition - Quantitative approaches to decision making collection*. Business Expert Press.
- <https://www.busessexpertpress.com/books/business-applications-multiple-regression-second-edition>.
- Schulz, R., Werwatz, A. (2004). A State Space Model for Berlin House Prices: Estimation and Economic Interpretation. *The Journal of Real Estate Finance and Economics* 28, 37-57.
- <https://doi.org/10.1023/A:1026373523075>.
- Selim, H. (2009). Determinants of House Prices in Turkey: Hedonic Regression versus Artificial Neural Network, *Journal of Expert Systems with Applications*, Vol.36. Issue 2. Part 2. Pages 2843-2852.
- doi.org/10.1016/j.eswa.2008.01.044.
- Sharma R., Rani S. (2021) A Novel Approach for Smart-Healthcare Recommender System. In: Hassanien A., Bhatnagar R., Darwish A. (eds)

