



University of
Sistan and Baluchestan

Geography and Territorial Spatial Arrangement

Print ISSN: 2345 - 2277

Online ISSN: 2783 - 5278



Association of Geography
and Planning
of Border Areas of Iran

Predicting Residential Rents across Tehran's 22 Districts Using Supervised Machine-Learning Algorithms

Ebrahim Rezaei¹, Abolfazl Meshkini^{2✉}, Mahdi Pourtaheri³

1. PhD student in Geography and Urban Planning, Faculty of Humanities, Tarbiat Modares University, Tehran, Iran.
E-mail: e.rezaee@modares.ac.ir
2. Professor, Department of Geography and Urban Planning, Faculty of Humanities, Tarbiat Modares University, Tehran, Iran.
✉ E-mail: meshkini@modares.ac.ir
3. Professor, Department of Geography and Rural Planning, Faculty of Humanities, Tarbiat Modares University, Tehran, Iran.
E-mail: mahdit@modares.ac.ir

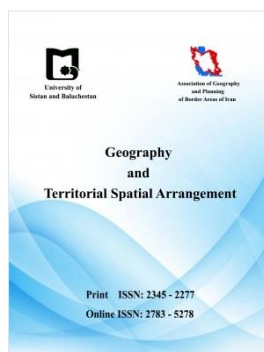


How to Cite: Rezaei, E; Meshkini, A & Pourtaheri, M. (2026). Predicting Residential Rents across Tehran's 22 Districts Using Supervised Machine-Learning Algorithms. *Geography and Territorial Spatial Arrangement*, 16 (60), 63-70.

DOI: <http://dx.doi.org/10.22111/gaij.2026.55808.3360>

Article type:
Research Article

Received:
15/06/2026
Received in revised form:
20/07/2026
Accepted:
26/07/2026
Publisher online:
04/08/2026



ABSTRACT

In a dynamic real estate market, accurate residential rent prediction is important for landlords, tenants, investors, and policymakers. This study aimed to develop an effective model for predicting residential rents in Tehran using supervised machine learning algorithms. The study used a dataset of 9,891 rental properties across Tehran's 22 districts, with 17 property-related features recorded for each observation. The algorithms were evaluated using accuracy and precision metrics. The results showed that XGBoost achieved the best predictive performance, with both accuracy and precision reaching 98.1%. The decision tree model also performed well, achieving an accuracy of 96.3% and a precision of 96.1%. The strong performance of these tree-based algorithms may be related to their ability to identify complex and nonlinear relationships in the data. The findings can help real estate professionals, investors, and policymakers make better-informed decisions and improve rental pricing and housing strategies in Tehran. Future studies could improve the predictive performance of these models by adding further variables, such as neighborhood characteristics, spatial factors, and economic indicators. The models could also be tested in other cities and regions to assess their wider use and improve understanding of rental market patterns.

Keywords:

Residential rent prediction;
supervised machine learning;
predictive modelling; Tehran.



© the Author(s).

Publisher: University of Sistan and Baluchestan

Extended Abstract

Introduction

Rapid urbanization has intensified economic, social, and spatial challenges, particularly access to adequate and affordable housing. Housing is not merely a shelter but a fundamental right encompassing security of tenure, affordability, and access to infrastructure. Rental housing plays an essential role in urban development, yet tenants may face rent increases, non-renewal of contracts, eviction, and repeated displacement. Moreover, the diversity of factors affecting residential rents reduces market transparency and complicates decision-making for tenants, landlords, investors, and policymakers.

These problems are especially important in Tehran, where chronic inflation, rising land and housing prices, higher construction costs, and increasing demand have intensified pressure on the rental market. Although machine-learning algorithms are increasingly used in property valuation, most previous studies have focused on housing sale prices, macroeconomic indicators, time-series models, or a limited number of urban districts. Property-level rent prediction across Tehran's 22 districts and comparison of several supervised algorithms within a common framework remain limited. This study therefore aimed to identify the housing attributes with the greatest explanatory power for rent prediction and to determine the optimal supervised machine-learning model for Tehran's 22 districts. Its main contributions were its focus on rents rather than sale prices, use of cross-sectional property-level data for all districts, uniform comparison of several algorithms, and district-scale examination of variables and model performance.

Study Area

The study area was Tehran, covering approximately 615 km² and containing a population of 8,693,706 with an average household size of 2.9 persons. The city is divided into 22 municipal districts and 370 neighborhoods and contains more than 2.6 million apartment units. Approximately 54% of residential units are owner-occupied and nearly 43% are occupied by tenants. Tehran displays substantial spatial differentiation: northern districts generally have higher land values, better services and infrastructure, and higher rents, whereas southern districts commonly record lower rents and different physical and accessibility conditions.

Material and Methods

This applied, quantitative, cross-sectional study used supervised machine learning to classify residential rent levels in Tehran. Rental-apartment advertisements were collected from the Divar online platform during a 30-day period from December 1 to December 31, 2023. Proportional stratified sampling distributed observations among the 22 districts according to their residential-unit stocks. From a total stock 9,891 listings were selected, and 17 property-related features were recorded. These features covered district and neighborhood, building and unit characteristics, amenities, geographical fields, floor area, building age, and total rent.

After missing observations and outliers were removed, 6,923 valid records remained. The data were divided into training and test sets at a ratio of 70:30. Grid search and K-fold cross-validation were used for hyperparameter optimization and model generalization. Ten classifiers were evaluated: decision tree, support vector classifier, Gaussian naive Bayes, K-nearest neighbors, logistic regression, multilayer perceptron, random forest, gradient boosting, AdaBoost, and XGBoost. Accuracy, precision, recall, F1-score, confusion matrices, ROC curves, and ROC-AUC were used for evaluation.

Result and Discussion

The findings revealed a strongly polarized rental geography. Districts 1, 2, and 3, together with parts of Districts 4 and 6, had rents above the city average, whereas southern districts generally recorded the lowest nominal values. Inequality was also evident in dwelling size and amenities. One-bedroom units represented approximately 66% of rentals in District 17, whereas more than 44% of units in affluent Districts 1 and 3 were larger and had more rooms. Elevator availability ranged from 72% to 85% in Districts 1-6 but was below 45% in Districts 11, 14, and 16.

Tree-based and ensemble models achieved the strongest performance. XGBoost ranked first, with accuracy of 0.981132, precision of 0.981250, recall of 0.979889, F1-score of 0.980569, and ROC-AUC of 0.998361. Decision tree achieved accuracy of 0.963275, precision of 0.961246, recall of 0.963245, F1-score of 0.962245, and ROC-AUC of 0.963274. Random forest also performed strongly, with accuracy of 0.953167, precision of 0.960424, recall of 0.942441, F1-score of 0.951348, and ROC-AUC of 0.993411. Gradient boosting and AdaBoost achieved accuracies of 0.941712 and 0.924191, respectively, while the remaining models produced lower accuracies. The superior performance of tree-based algorithms was attributed to their ability to identify

complex and nonlinear relationships among housing variables and their resistance to overfitting. These models can reduce pricing uncertainty, assist landlords and investors in determining reasonable rents, and support public institutions in market monitoring, taxation policy, and targeted rental-housing programs.

Conclusion

Supervised machine-learning algorithms, particularly XGBoost, showed strong capacity to classify residential rental conditions across Tehran's 22 districts. The spatial analysis also demonstrated that rental inequality extends beyond price: lower nominal rents in southern districts coexist with smaller dwellings, fewer amenities, lower physical quality, weaker accessibility, and greater tenure insecurity. The findings can support more informed decisions by tenants, landlords, real-estate professionals, investors, and policymakers.

The study was limited by the volume and diversity of available data. Future research can improve the models by using larger datasets, macroeconomic indicators, socioeconomic characteristics of neighborhoods, additional spatial and locational variables, and deep-learning approaches. Testing the models in other cities and regions can assess their generalizability. Integrating economic, spatial, and physical variables with machine learning can help identify districts exposed to rental pressure and support interventions directed toward affordability, security of tenure, and spatial justice.

Key words: Residential rent prediction; supervised machine learning; predictive modelling; Tehran.

References

- Ambrose, B. W., & Diop, M. (2018). Information Asymmetry, Regulations, and Equilibrium Outcomes: Theory and Evidence from the Housing Rental Market. *Microeconomics: Asymmetric & Private Information eJournal*. <https://doi.org/10.2139/ssrn.2851156>
- Bansal, M., Goyal, A., & Choudhary, (2022). A comparative analysis of K-Nearest Neighbor, Genetic, Support Vector Machine, Decision Tree, and Long Short Term Memory algorithms in machine learning. *Decision Analytics Journal*, 3, 100071. <https://doi.org/10.1016/j.dajour.2022.00071>
- Baur, K., Rosenfelder, M., & Lutz, B. (2023). Automated real estate valuation with machine learning models using property descriptions. *Expert Systems with Applications*, 213, 119147. <https://doi.org/10.1016/j.eswa.2022.119147>
- Bentéjac, C., Csörgő, A., & Martínez-Muñoz, G. (2021). A comparative analysis of gradient boosting algorithms. *Artificial Intelligence Review*, 54. <https://link.springer.com/article/10.1007/s10462-020-09896-5>
- Bisong, E. (2019). *Building Machine Learning and Deep Learning Models on Google Cloud Platform (A Comprehensive Guide for Beginners)* (1 ed.). Apress Berkeley, CA. <https://doi.org/10.1007/978-1-4842-4470-8>
- Blanco, A., & Razu, D. (2019). Rental Housing. In *The Wiley Blackwell Encyclopedia of Urban and Regional Studies* (pp. 1-6). Wiley. <https://doi.org/https://doi.org/10.1002/9781118568446.eurs0267>
- Bracke, P. (2013). How long do housing cycles last? A duration analysis for 19 OECD countries. *Journal of Housing Economics*, 22(3), 213-230. <https://doi.org/10.1016/j.jhe.2013.06.001>
- Braubach, M., Jacobs, D. E., & Ormandy, D. (2011). Environmental burden of disease associated with inadequate housing: a method guide to the quantification of health effects of selected housing risks in the WHO European Region (World Health Organization. Regional Office for Europe, Issue. <https://apps.who.int/iris/handle/10665/108587>
- Chen, Y., Liu, X., Li, X., Liu, Y., & Xu, X. (2016). Mapping the fine-scale spatial pattern of housing rent in the metropolitan area by using online rental listings and ensemble learning. *Applied Geography*, 75, 200-212. <https://doi.org/10.1016/j.apgeog.2016.08.011>
- Chicco, D., & Jurman, G. (2020). The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. *BMC genomics*, 21(1), 6.
- Clark, S. D., & Lomax, N. (2018). A mass-market appraisal of the English housing rental market using a diverse range of modelling techniques. *Journal of Big Data*, 5, Article 43. <https://doi.org/10.1186/s40537-018-0154-3>
- Cordero, E. A., & Gutiérrez, P. R. (2021). Social Housing in Guadalajara: A Viable Strategy to Mitigate the Negative Externalities of Metropolitan Urban Sprawl? *Applied Economics and Policy Studies*. https://doi.org/10.1007/978-981-16-5359-9_48
- David, W., Hosmer, J., Rodney, X., & Sturdivant, S. L. (2013). *Applied Logistic Regression* (3rd Edition ed.). Wiley & Sons. <https://www.wiley.com/en-us/Applied+Logistic+Regression%2C+3rd+Edition-p-9780470582473>
- De Giorgio, A., Cola, G., & Wang, L. (2023). Systematic review of class imbalance problems in manufacturing. *Journal of Manufacturing Systems*, 71, 620-644. <https://doi.org/10.1016/j.jmsy.2023.10.014>

- Diniz, P. S. R. (2023). *Signal Processing and Machine Learning Theory* (Academic Press Library in Signal Processing) (1st ed ed.). Academic Press. <https://doi.org/10.1016/B978-0-32-391772-8.00019-3>
- Divar Online Platform. (2023). Housing rental advertisements in Tehran (published from December 1, 2023, to December 31, 2023). Tehran, Iran. Available at <https://divar.ir/>
- Elariane, S. A. (2022). Location based services APIs for measuring the attractiveness of long-term rental apartment location using machine learning model. *Cities*, 103588, 122. <https://doi.org/10.1016/j.cities.2022.103588>
- Füss, R., & Koller, J. A. (2016). The role of spatial and temporal structure for residential rent predictions. *International Journal of Forecasting*, 32(4), 1352-1368. <https://doi.org/10.1016/j.ijforecast.2016.06.001>
- Gao, Q., Shi, V., Pettit, C., & Han, H. (2022). Property valuation using machine learning algorithms on statistical areas in Greater Sydney, Australia. *Land Use Policy*, 123, 106409. <https://doi.org/10.1016/j.landusepol.2022.106409>
- Garboden, P., & Rosen, E. (2019). Serial Filing: How Landlords Use the Threat of Eviction. *City & Community*, 18. <https://doi.org/10.1111/cico.12387>
- Géron, A. (2022). *Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow*. " O'Reilly Media, Inc".
- Ghaedrahmati, S., & Rezaei, E. (2024). Turkey, the second home for Iranians: push and pull motivations in Turkish housing market. *Journal of European Real Estate Research*, 17(1), 123-136. <https://doi.org/10.1108/JERER-06-2023-0019>
- Gopal, Patro, K., & Sahu, K. K. (2015). Normalization: A Preprocessing Stage. *ArXiv*, abs/1503.06462. <https://doi.org/10.48550/arXiv.1503.06462>
- Grybauskas, A., Pilinkienė, V., & Stundžienė, A. (2021). Predictive analytics using Big Data for the real estate market during the COVID-19 pandemic. *Journal of Big Data*, 8(1), 105. <https://doi.org/10.1186/s40537-021-00476-0>
- Gupta, S. C., & Goel, N. (2023). Predictive Modeling and Analytics for Diabetes using Hyperparameter tuned Machine Learning Techniques. *Procedia Computer Science*, 218, 1257-1269. <https://doi.org/10.1016/j.procs.2023.01.104>
- Haque, M. B., Choudhury, C., & Hess, S. (2020). Understanding differences in residential location preferences between ownership and renting: A case study of London. *Journal of Transport Geography*, 88, 102866. <https://doi.org/10.1016/j.jtrangeo.2020.102866>
- HasanGoodarzi, S and Armanmehr, M. (2019). Market analysis and forecasting of housing prices in Tehran. *Journal of Iranian Economic Issues*, 5(2), 79-103. Available at https://economics.ihcs.ac.ir/article_4310.html?lang=en (In Persian)
- Hosseini Ramandi, S and Kashani, H. (2023). Multiple Model for Estimating House Prices: Case Study Tehran City. *Journal of Structural and Construction Engineering*, 10(1), 172-189. <https://doi.org/10.22065/jsce.2022.294445.2496> (In Persian)
- Hu, J., & Szymczak, S. (2023). A review on longitudinal data analysis with random forest. *Brief Bioinform*, 24(2). <https://doi.org/10.1093/bib/bbad002>
- Hu, L., He, S., Han, Z., Xiao, H., Su, S., Weng, M., & Cai, Z. (2019). Monitoring housing rental prices based on social media: An integrated approach of machine-learning algorithms and hedonic modeling to inform equitable housing policies. *Land Use Policy*, 82, 657-673. <https://doi.org/10.1016/j.landusepol.2018.12.030>
- Huang, Y. (2019). Predicting Home Value in California, United States via Machine Learning Modeling. *Statistics, Optimization & Information Computing*, 7. <https://doi.org/10.19139/soic.v7i1.435>
- Iban, M. C. (2022). An explainable model for the mass appraisal of residences: The application of tree-based Machine Learning algorithms and interpretation of value determinants. *Habitat International*, 128, 102660. <https://doi.org/10.1016/j.habitatint.2022.102660>
- Jacobs, K., & Manzi, T. (2020). Conceptualising 'financialisation': governance, organisational behaviour and social interaction in UK housing. *International Journal of Housing Policy*, 20(2), 184-202. <https://doi.org/10.1080/19491247.2018.1540737>
- Joint Secretariat for the Preparation and Steering of Tehran Urban Development Plans. (2007). *Strategic-Structural (Master) Plan of Tehran City* (Approved by the Supreme Council for Urban Planning and Architecture of Iran in 2007). Tehran, Iran. Available at <https://shorayeali.mrud.ir/> (In Persian)
- Kang, J., Lee, H. J., Jeong, S. H., Lee, H. S., & Oh, K. J. (2020). Developing a Forecasting Model for Real Estate Auction Prices Using Artificial Intelligence. *Sustainability*, 12(7), 2899. <https://www.mdpi.com/2071-1050/12/7/2899>

- 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. <https://doi.org/10.1016/j.landusepol.2020.104919>
- Kim, T. Y., Park, E., & Ryu, D. (2025). Determinants of housing rental prices in Seoul: Applying explainable AI. *Spatial Economic Analysis*, 20(2), 312–332. <https://doi.org/10.1080/17421772.2024.2418906>
- Kishor, N. K., & Morley, J. (2015). What factors drive the price–rent ratio for the housing market? A modified present-value analysis. *Journal of Economic Dynamics and Control*, 58, 235–249. <https://doi.org/10.1016/j.jedc.2015.06.006>
- Kobzan, S., Ivakhnenko, A., & Tolsta, M. (2021). Research of features of rent market development. *Municipal economy of cities*, 1, 116–123. <https://doi.org/10.33042/2522-1809-2021-1-161-116-123>
- Lahmiri, S., Bekiros, S., & Avdoulas, C. (2023). A comparative assessment of machine learning methods for predicting housing prices using Bayesian optimization. *Decision Analytics Journal*, 6, 100166. <https://doi.org/10.1016/j.dajour.2023.100166>
- Lee, C., & Park, K. K.-H. (2022). Forecasting trading volume in local housing markets through a time-series model and a deep learning algorithm. *Engineering, Construction and Architectural Management*, 29(1), 16–5, 178. <https://doi.org/10.1108/ECAM-10-2020-0850>
- Lenaers, I., & De Moor, L. (2023). Exploring XAI techniques for enhancing model transparency and interpretability in real estate rent prediction: A comparative study. *Finance Research Letters*, 58, 104306. <https://doi.org/10.1016/j.frl.2023.104306>
- Leo Breiman, J. F., R.A. Olshen, Charles J. Stone. (2017). *Classification and Regression Trees* (1st ed ed.). Chapman and Hall/CRC. <https://doi.org/10.1201/9781315139470>
- Li, H. (2024). *Machine Learning Methods* Springer. <https://link.springer.com/book/10.1007/978-981-99-3917-6>
- Li, R. Y. M. (2022). Housing Real Estate Economics and Finance. *Journal of Risk and Financial Management*, 15(3), 121. <https://www.mdpi.com/1911-8074/15/3/121>
- Lipton, Z. C. (2018). The Mythos of Model Interpretability: In machine learning, the concept of interpretability is both important and slippery. *Queue*, 16(3), 31–57. <https://doi.org/10.1145/3236386.3241340>
- Liu, B. (2011). *Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data (Data-Centric Systems and Applications)* (2nd ed ed.). Springer. <https://link.springer.com/book/10.1007/978-3-642-19460-3>
- Mansour, A., Bentley, R., Baker, E., Li, A., Martino, E., Clair, A., Daniel, L., Mishra, S. R., Howard, N. J., Phibbs, P., Jacobs, D. E., Beer, A., Blakely, T., & Howden-Chapman, P. (2022). Housing and health: an updated glossary. *Journal of Epidemiology and Community Health*, 76(9), 833. <https://doi.org/10.1136/jech-2022-219085>
- Meyberg, C., Rendtel, U., & Leerhoff, H. (2024). Flat rent price prediction in Berlin with web scraping. *AStA Wirtschafts- und Sozialstatistisches Archiv*, 18, 245–278. <https://doi.org/10.1007/s11943-024-00340-6>
- Michael W. Berry, A. M., Bee Wah Yap. (2020). *Supervised and Unsupervised Learning for Data Science (Unsupervised and Semi-Supervised Learning)* (1st ed ed.). Springer. <https://www.springerprofessional.de/en/supervised-and-unsupervised-learning-for-data-science/17138882>
- Mienye, D., & Sun, Y. (2022). A Survey of Ensemble Learning: Concepts, Algorithms, Applications, and Prospects. *IEEE Access*, PP, 1–1. <https://doi.org/10.1109/ACCESS.2022.3207287>
- Mora-Garcia, R. T., Cespedes-Lopez, M. F., & Perez-Sanchez, V. R. (2022). Housing price prediction using machine learning algorithms in COVID-19 times. *Land*, 11(11), 2100. <https://doi.org/10.3390/land1112100>
- Murtagh, F. (1991). Multilayer perceptrons for classification and regression. *Neurocomputing*, 2(5), 183–197. [https://doi.org/10.1016/0925-2312\(91\)90023-5](https://doi.org/10.1016/0925-2312(91)90023-5)
- Najib, T., Muntasir, F., & Wasi, W. A. W. (2023). Transparency in House Rent of Dhaka: Explainable AI Based Predictive Framework. <https://doi.org/10.46254/BA06.20230146>
- Ogundunmade, T., Abidoye, M., & Olunfunbi, O. (2023). Modelling Residential Housing Rent Price Using Machine Learning Models. *Modern Economy and Management*, 4, 14. <https://doi.org/10.53964/mem.2023014>
- Pilehvar, A. A., & Ghasemi, A. (2024). Advanced modeling of housing locations in the city of Tehran using machine learning and data mining techniques. *Humanities and Social Sciences Communications*, 11, 804. <https://doi.org/10.1057/s41599-024-03244-6>
- Potrawa, T., & Tetereva, A. (2022). How much is the view from the window worth? Machine learning-driven hedonic pricing model of the real estate market. *Journal of Business Research*, 144, 50–65. <https://doi.org/10.1016/j.jbusres.2022.01.027>

- Radhoush, S., Whitaker, B. M., & Nehrir, H. (2023). An Overview of Supervised Machine Learning Approaches for Applications in Active Distribution Networks. *Energies*, 16(16), 5972. <https://www.mdpi.com/1996-1073/16/16/5972>
- Rafiei, M. H., & Adeli, H. (2018). Novel Machine-Learning Model for Estimating Construction Costs Considering Economic Variables and Indexes. *Journal of Construction Engineering and Management*, 144. [https://doi.org/10.1061/\(ASCE\)CO.1943-7862.0001570](https://doi.org/10.1061/(ASCE)CO.1943-7862.0001570)
- Renju, K., & Freni, S. (2024). Ensemble Approach for Predicting The Price of Residential Property. *International Journal of Information Technology, Research and Applications*, 3, 27-38. <https://doi.org/10.59461/ijitra.v3i2.99>
- Rezaeian S, Asgari H, Darvishi B. The Study of Determinants of Rent Housing in Ilam City Based on Hedonic Spatial Econometrics. *IUESA 2019*; 7 (26) :15-27. Available at <http://iueam.ir/article-1-1189-fa.html> (In Persian)
- Rizun, N., & Baj-Rogowska, A. (2021). Can Web Search Queries Predict Prices Change on the Real Estate Market? *IEEE Access*, 9, 70095-70117. <https://doi.org/10.1109/ACCESS.2021.3077860>
- Rokach, L. (2023). *Machine Learning for Data Science Handbook: Data Mining and Knowledge Discovery Handbook* (3rd ed ed.). Springer. <https://link.springer.com/book/10.1007/978-3-031-24628-9>
- Ryszard S, M., Jaime G, C., & Tom M, M. (2013). *Machine Learning: An Artificial Intelligence Approach* (1st ed ed., Vol. 1). Morgan Kaufmann. <https://www.amazon.com/Machine-Learning-Artificial-Intelligence-Approach/dp/0934613095>
- Sagi, O., & Rokach, L. (2018). Ensemble learning: A survey. *WIREs Data Mining and Knowledge Discovery*, 8(4), e1249. <https://doi.org/10.1002/widm.1249>
- Saremi, H. , Heydari, M. and Aghaei, F. (2018). Spatial analysis of housing price using geographically weighted regression (A case study in District 2 of Tehran Metropolitan City, Iran). *Urban Economics*, 3(2), 19-38. <https://doi.org/10.22108/ue.2018.109447.1056> (In Persian)
- Sharma, H., Harsora, H., & Ogunleye, B. (2024). An Optimal House Price Prediction Algorithm: XGBoost. *Analytics*, 3(1), 30-45. <https://www.mdpi.com/2813-2203/3/1/3>
- Shi, Y. (2022). *Advances in Big Data Analytics: Theory, Algorithms and Practices* (1st ed ed.). Springer. <https://www.amazon.com/Advances-Big-Data-Analytics-Algorithms/dp/9811636060>
- Shuzlina, R., Nor, Z., & Sofianita, M. (2021). Advanced Machine Learning Algorithms for House Price Prediction: Case Study in Kuala Lumpur. *International Journal of Advanced Computer Science and Applications*, 12. <https://doi.org/10.14569/IJACSA.2021.0121291>
- Soltani, A., Heydari, M., Aghaei, F., & Pettit, C. (2022). Housing price prediction incorporating spatio-temporal dependency into machine learning algorithms. *Cities*, 131, 103941. <https://doi.org/10.1016/j.cities.2022.103941>
- Statistical Center of Iran. (2016). *Population and Housing Census Results*. Tehran. Iran. Available at <https://amar.org.ir/>
- Sujata Dash, S. K. P., Joel J. P. C. Rodrigues, Babita Majhi. (2022). *Deep Learning, Machine Learning and IoT in Biomedical and Health Informatics: Techniques and Applications (Biomedical Engineering)* (1st Ed ed.). CRC Press. <https://doi.org/10.1201/9780367548445>
- Tajmiri, S., Azimi, E., Hosseini, M. R., & Azimi, Y. (2020). Evolving multilayer perceptron, and factorial design for modelling and optimization of dye decomposition by bio-synthesized nano CdS-diatomite composite. *Environmental Research*, 182, 108997. <https://doi.org/10.1016/j.envres.2019.108997>
- Tang, Y., Qiu, S., & Gui, P. (2018). Predicting Housing Price Based on Ensemble Learning Algorithm. <https://doi.org/10.1109/IDAP.2018.8620781>
- Tehran Municipality. (2022). *Statistical Report of Tehran Municipality*. Information and Communication Technology Organization of Tehran Municipality, Tehran, Iran. <https://data.tehran.ir/> (In Persian)
- Tekouabou, S. C. K., Gherghina, S. C., Kameni, E. D., Filali, Y., & Idrissi Gartoumi, K. (2024). AI-Based on Machine Learning Methods for Urban Real Estate Prediction: A Systematic Survey. *Archives of Computational Methods in Engineering*, 31(2), 1079-1095. <https://doi.org/10.1007/s11831-023-10010-5>
- Teremetskyi, V., Avramova, O., Svitlychnyy, O., Sloma, V., Bodnarchuk, O., Telestakova, A., & Kokhan, V. (2021). Housing Rights Protection in the Context of Legislation and Judicial Practice of Ukraine. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.4138075>
- Tharwat, A. (2021). Classification assessment methods. *Applied computing and informatics*, 17(1), 168-192.
- Trinka, O., & Kauermann, G. (2023). Can machine learning algorithms deliver superior models for rental guides? *AStA Wirtschafts- und Sozialstatistisches Archiv*, 17, 305–330. <https://doi.org/10.1007/s11943-023-00333-x>

- Truong, Q., Nguyen, M., Dang, H & ,Mei, B. (2020). Housing Price Prediction via Improved Machine Learning Techniques. *Procedia Computer Science*, 174, 433-442. <https://doi.org/10.1016/j.procs.2020.06.111>
- UN-Habitat. (2023). Executive Director's Thematic Paper: Achieving the Sustainable Development Goals in times of global crises. <https://unhabitat.org/executive-directors-thematic-paper-achieving-the-sustainable-development-goals-in-times-of-global>
- Wang, H., Yu, F., & Zhou, Y. (2018). Property Investment and Rental Rate Under Housing Price Uncertainty: A Real Options Approach (January 30, 2018). *Real Estate Economics*, Forthcoming, Claremont McKenna College Robert Day School of Economics and Finance Research Paper No. 3113143. <https://doi.org/10.1111/1540-6229.12235>
- Wang, K., Zhao, H., & Li, J. (2023). Machine Learning-Based House Rent Prediction Using Stacking Integration Method. *American Journal of Management Science and Engineering*, 8(2), 50-55. <https://doi.org/10.11648/j.ajmse.20230802.12>
- Wang, Y., & Zhao, Q. (2022). House Price Prediction Based on Machine Learning: A Case of King County. <https://doi.org/10.2991/aebmr.k.220307.253>
- Watson, T. J. (2001). An empirical study of the naive Bayes classifier. (22), 41-46. <https://doi.org/10.1039/b104835j>
- World_Bank. (2022). Poverty and shared prosperity 2022: Correcting course. World Bank. <https://doi.org/https://doi.org/10.1596/978-1-4648-1893-6>
- Xu, K. (2022). Predicting housing prices and analyzing real estate markets in the Chicago suburbs using machine learning *Journal of Student Research*, 11(3). <https://doi.org/10.47611/jsrhs.v11i3.3459>
- Xu, X., & Zhang, Y. (2023). A Gaussian process regression machine learning model for forecasting retail property prices with Bayesian optimizations and cross-validation. *Decision Analytics Journal*, 8, 100267. <https://doi.org/10.1016/j.dajour.2023.100267>
- Xu, Y., & Goodacre, R. (2018). On Splitting Training and Validation Set: A Comparative Study of Cross-Validation, Bootstrap and Systematic Sampling for Estimating the Generalization Performance of Supervised Learning. *Journal of Analysis and Testing*, 2(3), 249-262. <https://doi.org/10.1007/s41664-018-0068-2>
- Yoshida, T., Murakami, D., & Seya, H. (2024). Spatial prediction of apartment rent using regression-based and machine learning-based approaches with a large dataset. *The Journal of Real Estate Finance and Economics*, 69(1), 1–28. <https://doi.org/10.1007/s11146-022-09929-6>
- Zabor, E. C., Reddy, C. A., Tendulkar, R. D & ,Patil, S. (2022). Logistic Regression in Clinical Studies. *Int J Radiat Oncol Biol Phys*, 112(2), 271-277. <https://doi.org/10.1016/j.ijrobp.2021.08.007>
- Zali, S. , Pahlavani, P. and Bigdeli, B. (2023). A Spatial-Temporal Analysis of the Factors Effective on Housing Prices (Case study: District 5 of Tehran Municipality). *Town and Country Planning*, 15(1), 115-130. <https://doi.org/10.22059/jtcp.2022.341584.670318> (In Persian)
- Zarghamfard, M., & Meshkini, A. (2022). Analysis of factors affecting the realization of right to adequate housing in Iran: developing an interpretive-structural model. *International Journal of Housing Markets and Analysis*, 15(2), 411-428. <https://doi.org/10.1108/IJHMA-02-2021-0021>
- Zhang, L. (2023). Housing Price Prediction Using Machine Learning Algorithm *Journal of World Economy*, 2(3), 18-26. <https://www.pioneerpublisher.com/jwe/article/view/392>
- Zhou, J. (2023). The Private Sector in the Rental Housing Market. *BCP Business & Management*. <https://doi.org/10.54691/bcpbm.v43i.4631>
- Ziyadi, H. , Salavati, E. and Lotfi Heravi, M. M. (2023). Housing Price Forecasting Using AI (LSTM). *Financial Research Journal*, 25(4), 557-576. <https://doi.org/10.22059/frj.2023.349924.1007398> (In Persian)

پیش‌بینی اجاره‌بهای مسکن در مناطق ۲۲ گانه شهر تهران با استفاده از الگوریتم‌های یادگیری ماشین نظارت‌شده*

ابراهیم رضائی^۱، ابوالفضل مشکینی^۲، مهدی پورطاهری^۳

مقاله پژوهشی

چکیده

در چشم‌انداز پویای املاک و مستغلات، پیش‌بینی دقیق اجاره مسکن برای مالکان و مستأجران اهمیتی حیاتی دارد. پژوهش حاضر، بر توسعه مدلی کارآمد برای پیش‌بینی اجاره مسکن در شهر تهران با بهره‌گیری از الگوریتم‌های پیشرفته یادگیری ماشین نظارت‌شده تمرکز داشته که در خلال آن، از یک مجموعه داده غنی در ارتباط با املاک اجاره‌ای در مناطق ۲۲ گانه شهر تهران (شامل ۹۸۹۱ ملک اجاره‌ای، با ۱۷ ویژگی مرتبط با هر مشاهده)، استفاده شد. ارزیابی الگوریتم‌ها براساس معیارهای صحت و دقت نشان داد که الگوریتم «ایکس‌جی‌بوست» به صحت و دقت قابل توجه ۹۸/۱ درصدی دست یافت؛ در حالی که «درخت تصمیم» صحت ۹۶/۳ درصد و دقت ۹۶/۱ درصد را کسب کرد. عملکرد قوی این الگوریتم‌ها، از توانایی آن‌ها در درک روابط پیچیده و غیرخطی درون داده‌ها و مقاومتشان در برابر بیش‌برازش نشأت می‌گیرد. یافته‌های این مطالعه، بینش‌های ارزشمندی را به متخصصان حوزه املاک، سرمایه‌گذاران و سیاست‌گذاران در تهران ارائه می‌دهد تا تصمیمات کاملاً آگاهانه‌ای اتخاذ کنند و استراتژی‌های بازار اجاره مسکن را بهبود بخشند. مسیرهای بالقوه برای پژوهش‌های آینده، گنجانیدن ویژگی‌های بیشتری مانند مشخصات محله و شاخص‌های اقتصادی را در بر می‌گیرد تا قابلیت‌های پیش‌بینی الگوریتم‌ها را بیش‌ازپیش ارتقا دهد. علاوه بر این، بررسی قابلیت تعمیم این الگوریتم‌ها به سایر شهرها یا مناطق می‌تواند راه را برای کاربردهای گسترده‌تر و درک جامع‌تر از پویایی بازار اجاره هموار سازد.

جغرافیا و آمایش شهری - منطقه‌ای
پاییز ۱۴۰۵، سال ۱۶، شماره ۶۰
تاریخ دریافت: ۱۴۰۵/۰۳/۲۵
تاریخ بازنگری: ۱۴۰۵/۰۴/۲۹
تاریخ پذیرش: ۱۴۰۵/۰۵/۰۴
تاریخ انتشار آنلاین: ۱۴۰۵/۰۵/۱۳
صفحات: ۶۳-۹۸



واژه‌های کلیدی:

اجاره‌بهای مسکن، یادگیری ماشین، مدل‌سازی پیش‌بینانه، تهران.

مقدمه

جهان در حال تجربه روند شتابان شهرنشینی است؛ به گونه‌ای که طبق برآورد برنامه اسکان بشر سازمان ملل متحد^۱، این پدیده تا سال ۲۰۳۰ منجر به آن خواهد شد که بیش از شش نفر از هر ده نفر در سراسر جهان، در مناطق شهری سکونت داشته باشند (UN-Habitat, 2023)؛ همچنین براساس پیش‌بینی‌های بانک جهانی، جمعیت ساکن در مناطق شهری تا سال ۲۰۵۰ از مرز ۷۰ درصد خواهد گذشت (World Bank, 2022). این دگرگونی عظیم، در دهه‌های اخیر چالش‌های بی‌سابقه‌ای را در ابعاد اقتصادی، اجتماعی و فضایی ایجاد کرده است که در رأس آن‌ها، بحران دسترسی به مسکن مناسب و مدیریت فضاهای شهری قرار دارد (Tekouabou et al., 2024). مسکن نه تنها یک سرپناه، بلکه

1. UN-Habitat

* مقاله حاضر مستخرج از رساله دکتری آقای ابراهیم رضائی با عنوان "ارائه الگوی برنامه‌ریزی فضایی کاهش فقر مسکن در کلانشهرها (مورد مطالعه: شهر تهران)" از گروه جغرافیا و برنامه‌ریزی شهری دانشگاه تربیت مدرس با راهنمایی نویسنده دوم و مشاوره نویسنده سوم است.

e.rezaee@modares.ac.ir
meshkini@modares.ac.ir
mahdit@modares.ac.ir

۱- دانشجوی دکتری جغرافیا و برنامه‌ریزی شهری، دانشکده علوم انسانی، دانشگاه تربیت مدرس، تهران، ایران.
۲- استاد گروه جغرافیا و برنامه‌ریزی شهری، دانشکده علوم انسانی، دانشگاه تربیت مدرس، تهران، ایران. (نویسنده مسئول)
۳- استاد گروه جغرافیا و برنامه‌ریزی روستایی، دانشکده علوم انسانی، دانشگاه تربیت مدرس، تهران، ایران.

به عنوان حقی قانونی و بنیادی، دربرگیرنده مؤلفه‌هایی همچون امنیت مالکیت، استطاعت مالی و دسترسی به زیرساخت‌هاست (Braubach et al., 2011; Teremetskyi et al., 2021)؛ با این حال، در سایه نئولیبرالی شدن فضا، بخش وسیعی از جمعیت شهری از حق سکونت در مسکن مطلوب محروم شده‌اند (Jacobs & Manzi, 2020; Zarghamfard & Meshkini, 2022). امنیت تصدی، به خانواده این امکان را می‌دهد که بدون نگرانی از قراردادهای کوتاه‌مدت یا جابه‌جایی‌های مکرر، با آرامش در خانه خود زندگی کنند (Mansour et al., 2022). مالکیت مسکن، تصمیمی بلندمدت است که به سرمایه‌گذاری و هزینه‌های زیادی نیاز دارد. در مقابل، اجاره‌نشینی یکی از ناامن‌ترین شیوه‌های سکونت به‌شمار می‌رود؛ زیرا مستأجران همواره ممکن است با افزایش اجاره‌بها، تمدید نشدن قرارداد یا خطر تخلیه روبه‌رو شوند؛ با این حال، این ناامنی را می‌توان یکی از واقعیت‌های زندگی در نظام سرمایه‌داری دانست (Garboden & Rosen, 2019). اجاره‌بها همواره یکی از متغیرهای مهم در تحلیل نظری و تجربی پدیده‌های شهری بوده‌است (Chen et al., 2016) و به تبع آن، بازار اجاره نیز به دلیل فراهم آوردن امکان دسترسی افراد و خانواده‌ها به سرپناه و محل سکونت، نقشی اساسی در توسعه شهری ایفا می‌کند (Blanco & Razu, 2019; Zhou, 2023). هرچند این بازار به مثابه جایگزین مالکیت مسکن، به افزایش پایداری و انعطاف‌پذیری سکونتی کمک می‌کند (Cordero & Gutiérrez, 2021) اما تنوع و پیچیدگی عوامل مؤثر بر تعیین اجاره‌بها، شفافیت بازار را کاهش داده و تصمیم‌گیری را برای ذی‌نفعان با چالش جدی روبه‌رو کرده‌است (Ambrose & Diop, 2018).

در ایران، نرخ شهرنشینی با جهشی قابل توجه از ۳۱ درصد در سال ۱۳۳۵ به ۸۵ درصد در سال ۱۴۰۳ رسیده است (مرکز آمار ایران، ۱۴۰۳). تهران به عنوان پرجمعیت‌ترین کلان‌شهر غرب آسیا، با وضعیت حادثری روبه‌روست؛ به طوری که جمعیت این شهر در سال ۱۴۰۱ بیش از ۹ میلیون نفر برآورد شده که از این تعداد، حدود ۵۱ درصد مستأجر بوده‌اند (شهرداری تهران، ۱۴۰۱). در این کلان‌شهر، اجاره‌نشینی نه یک انتخاب، بلکه ضرورتی اجتناب‌ناپذیر ناشی از تورم مزمن و افزایش سرسام‌آور قیمت مسکن است (Ghaedrahmati & Rezaei, 2024). بی‌ثباتی اقتصادی، نرخ‌های بالای تورم، افزایش شدید قیمت زمین و مسکن، رشد هزینه‌های ساخت‌وساز و افزایش تقاضای مسکن از مهم‌ترین عوامل شکل‌دهنده این وضعیت به‌شمار می‌روند. افزایش سالانه قیمت اجاره که ناشی از تورم و رشد ارزش ملک است، به چرخه‌ای از قراردادهای کوتاه‌مدت (معمولاً یک‌ساله) و جابه‌جایی مکرر مستأجران منجر شده‌است. مالکان اغلب برای همگام‌شدن با افزایش ارزش ملک، قراردادها را تمدید یا مستأجران را جابه‌جا می‌کنند که در نهایت موجب تورم در بخش اجاره مسکن می‌شود. اغلب مالکان، اجاره را به عنوان منبعی برای کسب درآمد تلقی می‌کنند و همین امر بازاری فعال در این بخش ایجاد کرده‌است. بازار اجاره املاک هر ساله با افزایش قیمت مواجه است و سیاست‌های دولتی نظیر اعمال مالیات بر خانه‌های خالی نیز نتوانسته تغییراتی بنیادی در نرخ اجاره ایجاد کند. قیمت‌گذاری متناسب اجاره مسکن برای مستأجران، مالکان و حتی مدیران از اهمیت بسزایی برخوردار است؛ زیرا این مهم تأثیر قابل توجهی در تصمیم‌گیری‌های مالی و سرمایه‌گذاری آنان دارد (Li, 2022; Wang, Yu, & Zhou, 2018). اجاره‌بها نقشی مشابه سود در تعیین ارزش سهام در بازارهای مالی را ایفا می‌کند. نسبت اجاره به قیمت که گاه با عنوان «نرخ سرمایه» شناخته می‌شود، انتظارات بازار از بازدهی آتی مسکن یا رشد اجاره را منعکس می‌کند؛ همان گونه که نسبت سود سهام به قیمت، انتظارات مربوط به بازدهی آتی سهام یا رشد سود تقسیمی را نشان می‌دهد (Bracke, 2013; Kishor & Morley, 2015). در واقع، ذی‌نفعان برای اتخاذ تصمیم صحیح درباره اجاره یک

^۱ Cap Rate

واحد مسکونی، نیازمند استفاده از یک مدل پیش‌بینی مناسب هستند تا بتوانند ارزش دقیق اجاره مسکن را برآورد کنند (Renju & Freni, 2024)؛ برای مثال، مستأجر به یک مدل دقیق برای ارزیابی سرمایه‌گذاری، مالک برای ارزیابی میزان سود، بستانکار برای سنجش ریسک و مدیر ارشد و حسابرس برای مدیریت بهینه سبد دارایی‌ها به چنین مدلی اتکا می‌کند (Najib, Muntasir, & Wasi, 2023). افزون بر این، در کشورهایمانند: لوکزامبورگ، هلند یا سوئیس که میزان اجاره‌بها برای مالکین واحدهای مسکونی مشمول مالیات بر درآمد است، پیش‌بینی اجاره‌های محلی به‌عنوان مبنای مالیاتی مورد استفاده قرار می‌گیرد. افزون بر این، اجاره، بخش قابل‌توجهی از شاخص درآمد ناخالص داخلی را تشکیل می‌دهد؛ از همین‌رو است که آگاهی عمیق از ساختار و روند تحول اجاره‌بها در بازارهای مسکن محلی و ملی، برای سیاست‌گذاری عمومی در حوزه مسکن نیز از اهمیت بالایی برخوردار است (Füss & Koller, 2016). با توجه به این‌که اجاره‌بها بازتاب انتظارات بازار و مبنایی برای سیاست‌گذاری‌های مالی است، دستیابی به مدلی دقیق برای پیش‌بینی آن، نیازی مبرم برای سیاست‌گذاران، مستأجران و سرمایه‌گذاران محسوب می‌شود (Füss & Koller, 2016; Najib et al., 2023). درواقع، پرداختن به بازار اجاره مسکن، به‌عنوان موضوعی اساسی در حوزه جذاب ارزش‌گذاری و مدیریت املاک، اهمیت دارد؛ زیرا می‌تواند به شرکت‌های فعال در حوزه املاک و مستغلات، نهادهای مالی و سرمایه‌گذاران کمک کند تا درباره قیمت‌ها مذاکره کرده و اقدامات مناسب را به‌صورت پیش‌دستانه اتخاذ کنند (Lahmiri, Bekiros, & Avdoulas, 2023).

با عنایت به موارد مطرح‌شده فوق، پژوهش حاضر اهداف زیر را دنبال می‌کند:

(۱) تعیین مؤلفه‌هایی از مسکن که بیشترین قدرت تبیین‌گری را در پیش‌بینی اجاره‌بها در مناطق ۲۲گانه شهر تهران دارند.

(۲) ارائه مدل بهینه پیش‌بینی اجاره‌بهای مسکن در مناطق ۲۲گانه شهر تهران.

(۱) مرور ادبیات

(۱-۱) متغیرهای اثرگذار بر اجاره‌بهای مسکن

مسکن کالایی ناهمگن است و اجاره‌بهای آن صرفاً به ویژگی‌های فیزیکی بنا محدود نمی‌شود، بلکه از مجموعه عوامل ساختاری، مکانی، محله‌ای و اقتصادی تأثیر می‌پذیرد. در چارچوب قیمت‌گذاری هدانیک، اجاره‌بهای هر واحد مسکونی را می‌توان بازتاب ارزش ضمنی ویژگی‌هایی مانند: مساحت، تعداد اتاق، سن بنا، طبقه، کیفیت ساختمان، پارکینگ و آسانسور، در کنار موقعیت جغرافیایی، دسترسی به حمل‌ونقل عمومی، مراکز اشتغال و خدمات شهری دانست. ویژگی‌های اجتماعی و اقتصادی محله نیز در تعیین اجاره‌بها نقش بسزایی دارد. یافته‌های مطالعات اخیر نشان می‌دهد اثر این عوامل لزوماً خطی و در سراسر شهر یکسان نیست و ممکن است با توجه به موقعیت ملک و ساختار بازار محلی تغییر کند (Gao et al., 2022; Kim et al., 2025; Soltani et al., 2022). در ایران نیز نتایج مدل‌های هدانیک فضایی، تأثیر معنادار ویژگی‌های فیزیکی ملک، وضعیت اقتصادی-اجتماعی محله، مبلغ ودیعه و مجاورت فضایی را بر اجاره‌بها تأیید کرده‌است (رضائیان و همکاران، ۱۳۹۸). ماهیت فضایی بازار مسکن سبب می‌شود قیمت یا اجاره واحدهای نزدیک به یک‌دیگر تا حدی متأثر از شرایط مشترک محله و املاک مجاور باشد؛ در نتیجه، فرض استقلال کامل مشاهدات همواره برقرار نیست. نادیده گرفتن وابستگی و ناهمگونی فضایی ممکن است به برآورد نادرست اثر متغیرها و ارزیابی خوش‌بینانه عملکرد مدل منجر شود. پژوهش‌های اخیر نشان داده‌اند که وارد کردن

اطلاعات مکانی، متغیرهای دسترسی و مؤلفه‌های فضایی-زمانی، قدرت پیش‌بینی مدل‌های ارزش‌گذاری مسکن را افزایش می‌دهد (Gao et al., 2022; Soltani et al., 2022; Yoshida et al., 2024). مطالعات انجام‌شده در تهران نیز حاکی از آن است که مدل‌های رگرسیون وزنی جغرافیایی و جغرافیایی-زمانی، در مقایسه با رگرسیون معمولی، توان بیشتری در بازنمایی تفاوت‌های محلی بازار مسکن دارند (صارمی و همکاران، ۱۳۹۷؛ زالی و همکاران، ۱۴۰۲).

۱-۲) کاربرد الگوریتم‌های یادگیری ماشین در پیش‌بینی اجاره‌بها

بازیگران بازار و سیاست‌گذاران مایل هستند از الگوریتم‌های پیش‌بینی برای ارزیابی بخش املاک و مستغلات استفاده کنند (Xu & Zhang, 2023). مدل‌های اقتصادسنجی هدایک، به دلیل امکان تفسیر ضرایب، جایگاه مهمی در مطالعات مسکن دارد؛ با این حال، کاربست این مدل‌ها معمولاً مستلزم تعیین شکل تابع و تعریف قبلی روابط میان متغیرها است. در بازار مسکن، روابط میان ویژگی‌های ملک و اجاره‌بها غالباً پیچیده، غیرخطی و همراه با اثرات متقابل است. الگوریتم‌های یادگیری ماشین نظارت‌شده می‌تواند بدون تحمیل ساختار خطی ثابت، این الگوها را از داده استخراج کند؛ از این رو، پژوهشگران به منظور توسعه مدل‌های پیش‌بینی اجاره مسکن، به استفاده از الگوریتم‌های یادگیری ماشین روی آورده‌اند (Akinwamide & Hahn, 2021; Tekouabou et al., 2024). استفاده از این الگوریتم‌ها برای مدل‌سازی قیمت املاک در سال‌های اخیر به طور گسترده‌ای در ادبیات پژوهشی دنیا مورد بحث قرار گرفته و منجر به افزایش فراوانی اسناد منتشرشده در این حوزه شده است. البته باید اذعان نمود که در اغلب موارد، پژوهشگران قیمت خانه یا ملک را پیش‌بینی کرده‌اند (Sharma, Harsora, & Ogunleye, 2024; Shuzlina, Nor, 2021; Sofianita, 2021; Tang, Qiu, & Gui, 2018). متغیرهای مرتبط دیگر، از جمله شاخص بازار مسکن (Grybauskas, Pilinkienė, & Stundžienė, 2021; Lee & Park, 2022)، تغییرات قیمت مسکن (Rizun & Baj-Rogowska, 2021)، مکان‌یابی آپارتمان‌های اجاره‌ای بلندمدت (Elariane, 2022)، ارزش خانه/مسکن/زمین (Huang, 2019)، هزینه‌های ساخت و ساز (Rafiei & Adeli, 2018) و قیمت مزایده (Kang et al., 2020) توسعه داده‌اند. در این بین، مطالعات محدودی بر استفاده از الگوریتم‌های یادگیری ماشین برای پیش‌بینی اجاره مسکن تمرکز داشته‌اند (Chen et al., 2016; Hu et al., 2019; Ogundunmade, Abidoeye, & 2023; Wang, Zhao, & Li, 2023; Xu, 2022; Yoshida, Murakami, & Seya, 2024). هدف آن، شناسایی عوامل کلیدی مؤثر بر هزینه‌های اجاره و ارائه پیش‌بینی‌هایی دقیق‌تر و قابل‌اعتمادتر در این زمینه بوده است. مطالعات انجام‌شده در زمینه پیش‌بینی اجاره و ارزش‌گذاری انبوه املاک نشان داد که الگوریتم‌های مبتنی بر درخت و روش‌های تجمیعی، از جمله جنگل تصادفی، تقویت گرادیان، گت‌بوست و ایکس‌جی‌بوست، در بسیاری از موارد عملکرد بهتری از رگرسیون خطی داشته‌است (Clark & Lomax, 2018; Mora-Garcia et al., 2022; Soltani et al., 2022; Yoshida et al., 2024). با این حال، نمی‌توان یک الگوریتم را برای تمام بازارها و مجموعه داده‌ها برتر دانست. ساختار داده، حجم نمونه، کیفیت متغیرها، روش تنظیم ابرپارامترها و شیوه اعتبارسنجی بر نتیجه مقایسه مدل‌ها اثر می‌گذارند. برای مثال، ایکس‌جی‌بوست در پیش‌بینی اجاره مسکن ژاپن عملکرد مناسبی داشته است، اما پژوهش‌های دیگر نشان داده‌اند که مدل‌های تجمیعی نیز در صورت تنظیم نامناسب یا ارزیابی ناقص ممکن است دچار بیش‌برازش شوند (Mora-Garcia et al., 2022; Trinkaus & Kauermann, 2023). از این رو، ایکس‌جی‌بوست و درخت تصمیم باید در کنار مدل‌های پایه، مانند رگرسیون خطی و رگرسیون‌های منظم‌شده و با استفاده از داده آزمون مستقل ارزیابی شود. پژوهش‌ها همچنین نشان می‌دهد که افزایش دقت مدل فقط به انتخاب الگوریتم وابسته

نیست. افزودن داده‌های مربوط به کیفیت محله، دسترسی‌های شهری، توصیف متنی آگهی و تصاویر ملک می‌تواند بخش دیگری از ناهمگونی اجاره‌ها را توضیح دهد. استفاده از متن آگهی در بازار اجاره برلین و ترکیب داده‌های متنی و تصویری در مدل‌های هدانیک، خطای پیش‌بینی را کاهش داده است (Baur et al., 2023; Potrawa & Teterova, 2022)؛ با این حال، داده‌های استخراج شده از آگهی‌های اینترنتی با محدودیت‌هایی مانند: ثبت چندباره یک ملک، مقادیر گم شده، داده‌های پرت و تفاوت میان اجاره پیشنهادی و مبلغ نهایی قرارداد روبه‌رو هستند؛ از این‌رو، کیفیت داده و فرایند پیش‌پردازش باید بخشی اساسی از طراحی مدل تلقی شود (Meyberg et al., 2024). در جدول ۱، خلاصه‌ای از ادبیات اخیر درباره استفاده از الگوریتم‌های یادگیری ماشین برای پیش‌بینی بازار املاک و مستغلات شهری ارائه شده است.

جدول ۱. خلاصه ادبیات پژوهشی اخیر مرتبط با به‌کارگیری الگوریتم‌های یادگیری ماشین در پیش‌بینی بازار املاک و مستغلات شهری

الگوریتم	حوزه کاربرد و محدوده مورد مطالعه	یافته‌ها	مراجع
ایکس‌جی‌بوست XGBoost	پیش‌بینی قیمت یا اجاره در ترکیه، نیجریه، مالزی، چین، آمریکا و ژاپن	عموماً یکی از دقیق‌ترین و پایدارترین مدل‌ها، به‌ویژه در داده‌های بزرگ و روابط غیرخطی بود.	(Iban, 2022; Ogundunmade, Abidoye, & Olunfunbi, 2023; Sharma, Harsora, & Ogunleye, 2024; Shuzlina, Nor, & Sofianita, 2021; Truong et al., 2020; Wang & Zhao, 2022; Xu, 2022; Yoshida, Murakami, & Seya, 2024; Zhang, 2023)
تقویت گرادیان Gradient Boosting (GB)	ارزش‌گذاری مسکن در ترکیه، آمریکا و استرالیا	از مدل‌های خطی کارآمدتر و مناسب برای ثبت روابط غیرخطی و فضایی - زمانی بود.	(Iban, 2022; Kang et al., 2021; Soltani et al., 2022; Truong et al., 2020; Wang, Zhao, & Li, 2023)
طبقه‌بندی کننده پرسپترون چندلایه MLP Classifier (MLP)	پیش‌بینی قیمت مسکن در ایمز آمریکا	تنظیم ابرپارامترها عملکرد MLP را بهبود داد، اما از XGBoost ضعیف‌تر بود.	(Sharma, Harsora, & Ogunleye, 2024)
جنگل تصادفی Random Forest (RF)	پیش‌بینی قیمت در ترکیه، چین، آمریکا، استرالیا و ژاپن	عمدتاً دقیق‌تر از مدل‌های خطی؛ ولی در برخی مطالعات مستعد بیش‌برازش بود.	(Iban, 2022; Sharma, Harsora, & Ogunleye, 2024; Soltani et al., 2022; Truong et al., 2020; Xu, 2022; Yoshida, Murakami, & Seya, 2024; Zhang, 2023)
درخت تصمیم Decision Tree (DT)	بازار مسکن استرالیا و آمریکا	از رگرسیون خطی بهتر، اما از روش‌های تجمیعی درختی ضعیف‌تر بود.	(Soltani et al., 2022)
رگرسیون بردار پشتیبان Support Vector Regression (SVR)	پیش‌بینی قیمت مسکن در آمریکا	سطح عملکرد متوسط یا ضعیف‌تر از XGBoost و RF به‌ویژه در داده‌های پیچیده داشت.	(Sharma, Harsora, & Ogunleye, 2024; Wang, Zhao, & Li, 2023; Xu, 2022)
نزدیک‌ترین همسایه K-Nearest Neighbors (K-NN)	پیش‌بینی قیمت مسکن در ایمز آمریکا	بیشترین خطا را میان مدل‌های مقایسه شده داشت و به مقیاس ویژگی‌ها حساس بود.	(Wang, Zhao, & Li, 2023; Zhang, 2023)
رگرسیون ترکیبی Hybrid Regression (HR)	پیش‌بینی قیمت معاملات مسکن پکن	ترکیب مدل‌ها از مدل‌های منفرد دقیق‌تر بود.	(Truong et al., 2020)
رگرسیون تعمیم‌یافته انباشته Stacked Generalization Regression (SGR)	پیش‌بینی قیمت معاملات مسکن پکن	کمترین خطای آزمون و بهترین تعمیم‌پذیری را نشان داد.	(Truong et al., 2020)

الگوریتم	حوزه کاربرد و محدوده مورد مطالعه	یافته‌ها	مراجع
رگرسیون ریدج Ridge Regression (RR)	پیش‌بینی قیمت در مالزی و آمریکا	از مدل خطی ساده پایدارتر بود، اما معمولاً از مدل‌های تقویتی ضعیف‌تر عمل کرد.	(Zhang, 2023)
شبکه عصبی عمیق Deep Neural Network (DNN)	پیش‌بینی اجاره آپارتمان در داده‌های بزرگ ژاپن	عملکرد رقابتی داشت، اما XGBoost و RF در نمونه‌های بزرگ دقیق‌تر بودند.	(Yoshida, Murakami, & Seya, 2024)
لایت جی بی ام Light GBM (LG)	پیش‌بینی قیمت در ترکیه، مالزی، چین و آمریکا	سریع و دقیق‌تر از مدل‌های خطی، ولی در مطالعات بررسی شده غالباً مدل نهایی برتر نبود.	(Iban, 2022; Shuzlina, Nor, & Sofianita, 2021; Wang, Zhao, & Li, 2023; Wang & Zhao, 2022)
کتاب‌بوست Cat Boost (CB)	پیش‌بینی قیمت فروش مسکن در کینگ کانتری آمریکا	بهترین دقت و تعمیم‌پذیری را با بیش‌برازش کمتر از RF نشان داد.	(Wang & Zhao, 2022)
رگرسیون خطی Linear Regression (LR)	مدل پایه و تبیینی در مطالعات آمریکا، مالزی و استرالیا	تفسیرپذیر، اما در ثبت روابط پیچیده و غیرخطی ضعیف‌تر از مدل‌های تجمیعی بود.	(Kang et al., 2021; Sharma, Harsora, & Ogunleye, 2024; Soltani et al., 2022)
کمترین مربعات معمولی Ordinary Least Squares (OLS)	ارزش‌گذاری و پیش‌بینی اجاره در ترکیه و نیجریه	برای تبیین اثر متغیرها مناسب، اما دقت پیش‌بینی آن کمتر از مدل‌های یادگیری ماشین بود.	(Ogundunmade, Abidoye, & Olunfunbi, 2023)

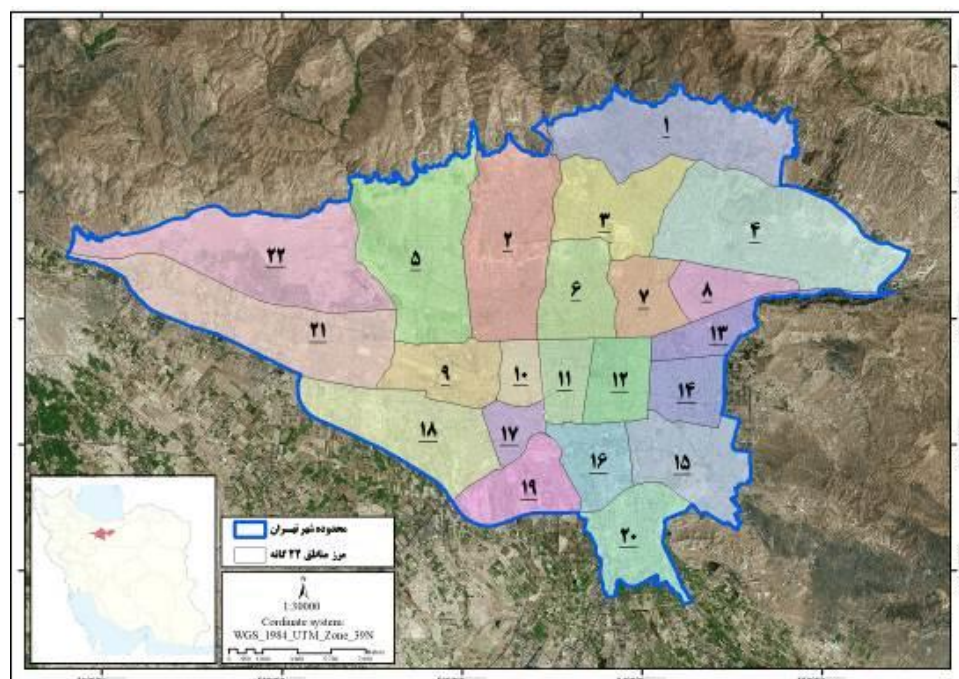
یکی از ویژگی‌های بازار اجاره ایران، ترکیب مبلغ ودیعه و اجاره ماهانه است. این ویژگی تعریف متغیر وابسته را پیچیده می‌کند؛ زیرا این دو جزء معمولاً جانشین یکدیگرند. استفاده از مبلغ ودیعه به عنوان متغیر ورودی برای پیش‌بینی اجاره ماهانه می‌تواند موجب نشت اطلاعات شود و دقت مدل را به طور مصنوعی افزایش دهد. برای جلوگیری از این مسئله، می‌توان اجاره معادل ماهانه را براساس نرخ تبدیل مشخص و مستند محاسبه کرد. راه دیگر آن است که در صورت انتخاب اجاره ماهانه به عنوان متغیر هدف، مبلغ ودیعه از مجموعه متغیرهای توضیحی حذف شود. این تصمیم باید برای تمام مشاهدات به صورت یکسان اعمال و با شرایط زمانی بازار هماهنگ شود. بخش عمده مطالعات داخلی در سال‌های اخیر بر پیش‌بینی قیمت فروش مسکن یا تحلیل متغیرهای کلان بازار تمرکز داشته است. در تهران، روش‌هایی مانند پویایی‌شناسی سیستم، مدل‌های سری زمانی، الگوریتم LSTM و روش‌های داده‌کاوی مکانی برای پیش‌بینی یا تحلیل قیمت مسکن استفاده شده‌اند (حسن‌گودرزی و آرمان‌مهر، ۱۳۹۷؛ حسینی‌رامندی و کاشانی، ۱۴۰۲؛ زیادی و همکاران، ۱۴۰۲؛ پیلهور و قاسمی، ۲۰۲۴). مطالعات فضایی نیز اغلب یک منطقه مشخص مانند مناطق ۲ یا ۵ تهران، را بررسی کرده و متغیر وابسته آن‌ها قیمت فروش بوده است (صارمی و همکاران، ۱۳۹۷؛ زالی و همکاران، ۱۴۰۲). در میان پژوهش‌های داخلی مرتبط، مطالعه رضائیان و همکاران (۱۳۹۸) مستقیماً عوامل تعیین‌کننده اجاره‌بها را با رویکرد هدائیک فضایی بررسی کرده است؛ با این حال، قلمرو آن شهر ایلام بوده و از الگوریتم‌های یادگیری ماشین استفاده نشده است. بررسی پیشینه نشان می‌دهد کانون اصلی اکثر مطالعات، پیش‌بینی "قیمت مسکن" بوده و بازار اجاره، علی‌رغم ماهیت پویا و حساسیت بالای آن به نوسانات اقتصادی، کمتر مورد توجه دقیق مدل‌های پیش‌بینی قرار گرفته است. مطالعات ملک‌محور درباره پیش‌بینی

اجاره‌بهای تهران، به‌ویژه در مقیاس مناطق ۲۲ گانه، همچنان محدود است. بیشتر پژوهش‌های داخلی به قیمت فروش، شاخص‌های کلان یا تعداد محدودی از مناطق شهری پرداخته‌است. افزون بر این، مقایسه الگوریتم‌های یادگیری ماشین در یک چارچوب ارزیابی یکسان، تحلیل تفاوت عملکرد مدل در مناطق مختلف و بررسی روابط غیرخطی میان ویژگی‌های ملک و اجاره‌بها کمتر مورد توجه قرار گرفته‌است. بر این مبنای نوآوری پژوهش حاضر در پنج‌محور قابل طرح است: الف- تمرکز بر اجاره‌بهای واحدهای مسکونی به‌جای قیمت فروش؛ ب- استفاده از داده‌های خرد مناطق ۲۲ گانه تهران؛ ج- بهره‌گیری از رویکرد داده‌محور مقطعی به‌جای رویکردهای رایج مبتنی بر داده‌های طولی؛ د- مقایسه الگوریتم‌های یادگیری ماشین نظارت‌شده در یک فرایند یکسان پیش‌پردازش و ارزیابی و ه- بررسی اهمیت متغیرها و خطای مدل در مقیاس منطقه‌ای.

۲) داده‌ها و روش

۲-۱) محدوده مورد مطالعه

محدوده مورد مطالعه پژوهش، شهر تهران است که با مساحتی در حدود ۶۱۵ کیلومتر مربع، جمعیتی بالغ بر ۸,۶۹۳,۷۰۶ نفر، متوسط بعد خانوار ۲,۹، از حیث تقسیمات کالبدی، به ۲۲ منطقه و ۳۷۰ محله تقسیم شده‌است. از نظر شاخص‌های مسکن، این محدوده شامل بیش از ۲,۶ میلیون واحد مسکونی آپارتمانی بوده و نرخ مالکیت واحد مسکونی در این شهر حدود ۵۴ درصد و سهم خانوارهای مستأجر نزدیک به ۴۳ درصد ثبت شده‌است. درصد باقیمانده شامل مصادیق متنوعی از تصرف است (مانند موقوفه و...) که در دو گروه ملکی و استیجاری قرار نمی‌گیرند (مرکز آمار ایران، ۱۳۹۵؛ شهرداری تهران، ۱۴۰۱).



شکل ۱. نقشه موقعیت جغرافیایی و تقسیمات کالبدی محدوده مورد مطالعه

(منبع: شهرداری تهران، ۱۴۰۱؛ ترسیم: نگارندگان، ۱۴۰۵)

۲-۲) روش انجام پژوهش

این پژوهش از نظر هدف، کاربردی، از منظر استراتژی، کمی و مبتنی بر داده‌های مقطعی است که با بهره‌گیری از رویکرد مدل‌سازی پیش‌بینانه مبتنی بر یادگیری ماشین نظارت‌شده^۱ جهت تحلیل و طبقه‌بندی نرخ اجاره‌بها در بازار مسکن شهر تهران تدوین شده‌است. هدف اصلی این مطالعه، توسعه و ارزیابی یک مدل هوشمند به‌منظور پیش‌بینی نرخ اجاره‌بها (بالا/پایین) براساس ویژگی‌های کالبدی واحدهای مسکونی در مناطق ۲۲گانه شهر تهران است؛ اقدامی که با شناسایی الگوهای پنهان در داده‌های بزرگ پلتفرم "دیوار"^۲، به کاهش عدم قطعیت در فرآیندهای تصمیم‌گیری بازار مسکن کمک می‌کند. در همین راستا، از میان رکوردهای اجاره آپارتمان واقع در محدوده مورد مطالعه که در این پلتفرم نمایه شده‌بود، تعداد ۹۸۹۱ رکورد به‌همراه ۱۷ ویژگی (متغیر) به‌صورت ثانویه و ازطریق استخراج مستقیم^۳ از پلتفرم آنلاین دیوار در بازه زمانی ۳۰ روزه (۱۴۰۲/۰۹/۱۰ تا ۱۴۰۲/۱۰/۱۰) گردآوری شد. جامعه آماری این مطالعه شامل تمامی آگهی‌های ثبت‌شده در شهر تهران طی بازه مذکور است که پس از طی مراحل پالایش داده‌ها، فیلترینگ و حذف رکوردهای مفقود^۴، مجموع ۶۹۲۳ رکورد جهت انجام تحلیل‌های نهایی برگزیده شد.

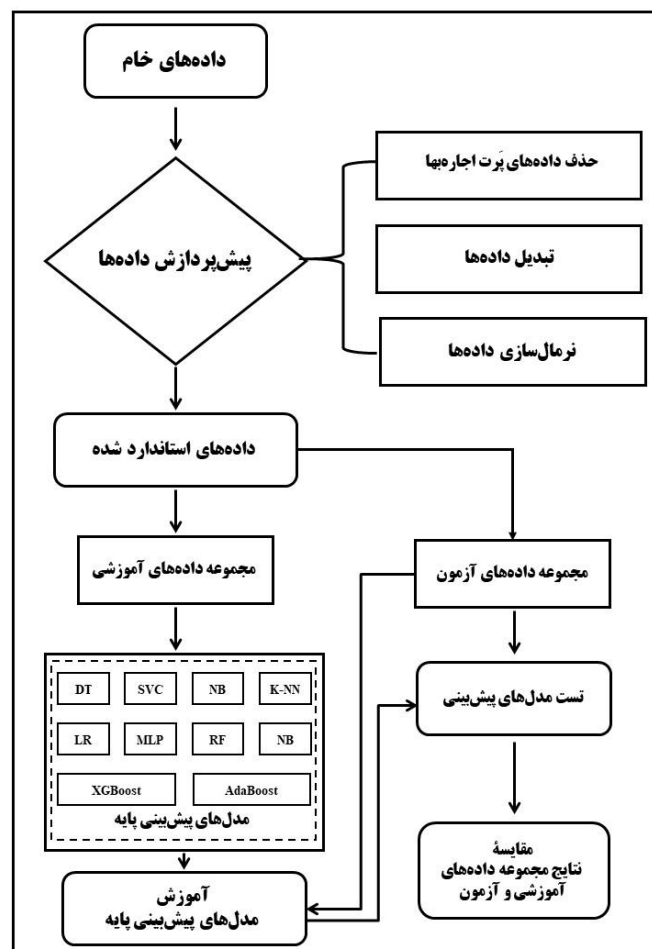
جدول ۲. توزیع حجم نمونه آماری در مناطق ۲۲گانه شهر تهران براساس روش طبقه‌بندی‌شده تناسبی

منطقه	تعداد کل واحد مسکونی	حجم نمونه	منطقه	تعداد کل واحد مسکونی	حجم نمونه
۱	۱۶۳،۷۴۰	۵۶۴	۱۲	۷۶،۴۸۰	۲۶۴
۲	۲۳۶،۹۶۶	۸۱۶	۱۳	۸۳،۲۴۷	۲۸۷
۳	۱۱۷،۲۳۲	۴۰۴	۱۴	۱۶۹،۰۳۶	۵۸۲
۴	۳۰۰،۸۴۲	۱،۰۳۷	۱۵	۲۰۱،۳۷۳	۶۹۴
۵	۲۸۹،۳۹۰	۹۹۷	۱۶	۸۵،۸۲۰	۲۹۶
۶	۸۳،۱۱۳	۲۸۶	۱۷	۸۶،۵۹۰	۲۹۸
۷	۱۱۴،۵۶۴	۳۹۵	۱۸	۱۲۸،۶۱۳	۴۴۳
۸	۱۴۷،۳۳۹	۵۰۸	۱۹	۷۸،۴۶۵	۲۷۰
۹	۵۶،۹۷۱	۱۹۶	۲۰	۱۱۴،۳۹۲	۳۹۴
۱۰	۱۱۶،۰۸۲	۴۰۰	۲۱	۵۹،۹۸۲	۲۰۷
۱۱	۱۰۶،۱۲۶	۳۶۶	۲۲	۵۴،۲۹۰	۱۸۷
جمع	۲،۸۷۰،۶۵۳	۹،۸۹۱	-	-	-

(منبع: مرکز آمار ایران، ۱۳۹۵ و محاسبات نگارندگان، ۱۴۰۵)

^۱Supervised Machine Learning^۲Big Data^۳ پلتفرم دیوار به‌عنوان یک بستر نیازمندی‌های آنلاین در ایران، آگهی‌های متنوعی را در دسته‌بندی‌های گوناگون از جمله املاک، خودرو، لوازم خانگی، خدمات کسب‌وکار و سرگرمی میزبانی می‌کند.^۴Web Scraping^۵Missing Values

با توجه به ماهیت هدفمند پژوهش، از روش نمونه‌گیری طبقه‌بندی‌شده تناسبی استفاده شد. برای تعیین حجم نمونه، با توجه به ابعاد کلان جامعه آماری (تعداد ۹۰,۰۰۰ آگهی ثبت شده در بازه زمانی مورد مطالعه) و با هدف کاهش خطای نمونه‌گیری، مجموع واحدهای مسکونی معادل ۲۸,۷۰,۶۵۳ در مناطق ۲۲ گانه شهر تهران (مرکز آمار ایران، ۱۳۹۵) مبنا قرار داده شد و تعداد ۹,۸۹۱ واحد مسکونی به عنوان حجم نمونه کل انتخاب گردید. این حجم نمونه معادل ۰,۳ درصد از مجموع واحدهای مسکونی است که ضمن رعایت صرفه اقتصادی، دقت آماری لازم برای تعمیم نتایج به پهنه‌های مختلف شهری را تضمین می‌کند. به منظور بازنمایی دقیق وضعیت سکونت در مناطق ۲۲ گانه، توزیع نمونه‌ها به صورت مستقیم با تعداد واحدهای مسکونی موجود در هر منطقه متناسب‌سازی شد (جدول ۲). با اتمام فرآیند گردآوری، داده‌های خام در محیط نرم‌افزار Visual Studio Code با پیش‌پردازش دقیق (حذف داده‌های پرت و نرمال‌سازی) به داده‌های استاندارد تبدیل گشته است. در ادامه، برای پیش‌بینی دوکلاسه با بهره‌گیری از چندین الگوریتم پیش‌بینی پایه نظارت‌شده، الگوریتم‌های مختص پیش‌بینی اجاره‌بها آموزش داده شده‌اند. ارزیابی نهایی الگوریتم‌ها از طریق مقایسه عملکرد بر روی مجموعه‌های آموزشی و آزمون صورت می‌پذیرد تا بهترین مدل برای پیش‌بینی دقیق اجاره‌بها در کلان‌شهر تهران انتخاب گردد (شکل ۲).



شکل ۲. مراحل کلیدی ارزیابی الگوریتم‌های یادگیری ماشین در پژوهش حاضر

(ترسیم: نگارندگان، ۱۴۰۵)

^۱ Proportional Stratified Sampling

به منظور ارزیابی عملکرد الگوریتم‌ها، مجموعه داده نهایی با نسبت ۷۰ درصد برای داده‌های آموزشی و ۳۰ درصد برای داده‌های آزمون تقسیم‌بندی شد (Lipton, 2018). فرآیند آموزش الگوریتم‌ها با به کارگیری الگوریتم‌های یادگیری نظارت‌شده و با بهره‌گیری از تکنیک‌های جستجوی شبکه‌ای^۱ برای بهینه‌سازی فرآیند پارامترها^۲ و اعتبارسنجی متقاطع K-لایه^۳ جهت اطمینان از تعمیم‌پذیری مدل انجام‌پذیرفت. در یادگیری نظارت‌شده، الگوریتم‌های طبقه‌بندی از طریق آموزش بر روی داده‌های برچسب‌دار^۴ یادمی‌گیرند تا داده‌های جدید و ناشناخته را به کلاس‌های هدف (در این پژوهش: کلاس ۰ و ۱) تخصیص دهند (Liu, 2011). این رویکرد به مدل اجازه می‌دهد تا با پالایش مداوم وزن‌ها و پارامترها، دقت پیش‌بینی خود را به حداکثر برساند. ماتریس درهم‌ریختگی (جدول ۳)، ابزاری تحلیلی برای ارزیابی عملکرد مدل‌های طبقه‌بندی است که امکان محاسبه شاخص‌های کلیدی ارزیابی را فراهم می‌آورد. این ماتریس از طریق مقایسه برچسب‌های پیش‌بینی‌شده توسط مدل با مقادیر واقعی در مجموعه داده‌های آزمون تشکیل می‌شود و چهار پارامتر اساسی زیر را تبیین می‌کند که شالوده اصلی برای محاسبه معیارهای ارزیابی عملکرد مدل هستند (de Giorgio, Cola, & Wang, 2023):

- مثبت واقعی (TP): مواردی که به درستی به عنوان کلاس «افزایش هزینه» (۱) شناسایی شده‌اند.
- منفی واقعی (TN): مواردی که به درستی به عنوان کلاس «سطح معمول» (۰) شناسایی شده‌اند.
- مثبت کاذب (FP): مواردی که مدل به اشتباه کلاس «سطح معمول» (۰) را «افزایش هزینه» (۱) پیش‌بینی کرده‌است.
- منفی کاذب (TN): مواردی که مدل به اشتباه کلاس «افزایش هزینه» (۱) را «سطح معمول» (۰) پیش‌بینی کرده‌است.

جدول ۳. ماتریس درهم‌ریختگی

کلاس واقعی (منفی/۰)	کلاس واقعی (مثبت/۱)	
TN	FN	پیش‌بینی منفی (۰)
FP	TP	پیش‌بینی مثبت (۱)

(منبع: Chicco and Jurman, 2020)

جدول ۴. معیارهای ارزیابی عملکرد مدل

فرمول محاسباتی	معیارهای ارزیابی
$(TP+TN)/(TP+TN+FP+FN)$	صحت ^۶
$TP/(TP+FP)$	دقت ^۷
$TP/(TP+FN)$	فراخوانی ^۸
$(Precision \times Recall) / (Precision + Recall) \times 2$	شاخص F1 ^۹

(منبع: Tharwat, 2021)

^۱ Grid Search^۲ Hyperparameters^۳ K-fold Cross Validation^۴ Labeled Data^۵ Test Set^۶ Accuracy^۷ Precision^۸ Recall^۹ F-Score

۳) یافته‌های پژوهش

در پلتفرم دیوار، تعداد ۹۸۹۱ رکور (داده اولیه) با ۱۷ ویژگی (متغیر) گردآوری شد. جدول ۵ هر یک از این ویژگی‌ها، نوع داده و توضیحات مربوط به آن‌ها را نمایش می‌دهد. متغیر هدف نیز به منظور تسهیل فرآیند پیش‌بینی در طول تحقیق، در این مجموعه داده گنجانده شده‌است.

جدول ۵. شرح ویژگی‌ها و ساختار داده

نام ویژگی (لاتین)	نام ویژگی (فارسی)	نوع داده	شرح ویژگی
OBJECT ID	شناسه	int64	کد منحصر به فرد هر رکورد
District	منطقه	int64	منطقه‌ای که ملک در آن واقع شده
Neighborhood	محله	int64	محله‌ای که ملک در آن واقع شده
Total Floor	کل طبقات	int64	تعداد کل طبقات ساختمان
Floor	طبقه	int64	طبقه‌ای که واحد مسکونی در آن قرار دارد
Units per floor	تعداد واحد در طبقه	int64	تعداد واحدهای مسکونی در هر طبقه
Parking	پارکینگ	int64	تعداد پارکینگ اختصاص یافته به واحد
Storage	انباری	int64	تعداد انباری اختصاص یافته به واحد
Rooms	اتاق‌ها	int64	تعداد اتاق‌های موجود در واحد
Elevator	آسانسور	int64	دارا بودن آسانسور
Natural lighting	نور طبیعی	int64	بهره‌مندی از نور طبیعی
Terrace	تراس	int64	دارا بودن تراس
X	مختصات X	object	مختصات جغرافیایی X
Y	مختصات Y	object	مختصات جغرافیایی Y
Area	مساحت	int64	مساحت زیربنا یا مساحت زمین ملک
Life of structure	عمر بنا	int64	سن ساختمان (عمر بنا)
Total Rent	مجموع اجاره‌بها	int64	مبلغ اجاره‌بها
Target	هدف (متغیر هدف)	int64	مقدار صفر یا یک براساس شرایط کارشناسی
Location	جهت (موقعیت)	int64	جهت قرارگیری ملک (شمالی، جنوبی و...)

(منبع: یافته‌های پژوهش، ۱۴۰۵)

ساختار داده فوق که به‌طور کلی شامل ویژگی‌های کالبدی^۲، ویژگی‌های مکانی/جغرافیایی^۳ و متغیر هدف است، نشان‌دهنده یک "داده ساختار متمرکز بر ویژگی" است. در پیاده‌سازی مدل‌های یادگیری ماشین، این یکپارچگی بین داده‌های کالبدی و مکانی^۵ یک نقطه قوت قابل تأمل در پژوهش است.

شایان ذکر است که ویژگی‌های مختصات X، مختصات Y و موقعیت از مجموعه داده نهایی حذف شده‌اند. دلیل حذف ویژگی‌های مختصات، ثابت بودن مقادیر آن‌ها (فاقد واریانس لازم برای آموزش مدل) بوده‌است. همچنین،

^۱ Target Feature

^۲ Physical Attributes

^۳ Spatial/Geospatial Attributes

^۴ Feature-based Data Structure

^۵ Spatial-Physical Integration

ویژگی موقعیت به دلیل نرخ بالای داده‌های مفقود که در حدود ۸۳ درصد از کل نمونه‌ها را شامل می‌شد، از فرآیند تحلیل کنار گذاشته شد تا از کاهش دقت و سوگیری احتمالی در عملکرد مدل‌های یادگیری ماشین جلوگیری شود. مجموعه داده اولیه شامل ۹۸۹۱ رکورد بود. در فرآیند پاک‌سازی داده‌ها، رکوردهایی که دارای مقادیر مفقود در ویژگی‌های کلیدی بودند، حذف شدند که این امر منجر به پالایش مجموعه داده به ۶۹۲۳ رکورد معتبر گردید. در گام بعدی، با بهره‌گیری از نظرات سه خبره حوزه برنامه‌ریزی مسکن، سه ویژگی کلیدی شامل: "مجموع اجاره‌بها"، "عمر بنا" و "مساحت" به عنوان متغیرهای مرجع جهت تعریف متغیر هدف شناسایی شدند. مبنای استدلال این است که در صورت احراز همزمان این سه شاخص، احتمال وقوع اجاره‌بهای بالاتر افزایش می‌یابد. پارامترها و میانگین‌های تعیین شده برای این سه شرط در هر یک از مناطق ۲۲ گانه تهران، در جدول ۶ تدوین و ارائه شده است.

جدول ۶. ویژگی‌های سه‌گانه محوری برای تعیین متغیر هدف

منطقه	اجاره‌بهای ماهانه (ریال)	مساحت (مترمربع)	عمر بنا (سال)	منطقه	اجاره‌بهای ماهانه (ریال)	مساحت (مترمربع)	عمر بنا (سال)
۱	۳۴۰,۰۰۰,۰۰۰	۸۰	۵	۱۲	۱۶۰,۰۰۰,۰۰۰	۸۰	۹
۲	۲۷۰,۰۰۰,۰۰۰	۷۲	۸	۱۳	۲۳۰,۰۰۰,۰۰۰	۷۵	۵
۳	۲۷۰,۰۰۰,۰۰۰	۷۷	۷	۱۴	۱۶۰,۰۰۰,۰۰۰	۹۰	۵
۴	۲۱۰,۰۰۰,۰۰۰	۸۷	۶	۱۵	۱۳۰,۰۰۰,۰۰۰	۹۰	۵
۵	۲۲۰,۰۰۰,۰۰۰	۷۵	۸	۱۶	۱۵۵,۰۰۰,۰۰۰	۷۵	۷
۶	۳۲۰,۰۰۰,۰۰۰	۹۰	۸	۱۷	۱۶۵,۰۰۰,۰۰۰	۸۵	۸
۷	۱۸۰,۰۰۰,۰۰۰	۷۵	۷	۱۸	۱۳۰,۰۰۰,۰۰۰	۹۰	۷
۸	۲۱۰,۰۰۰,۰۰۰	۸۲	۸	۱۹	۱۰۰,۰۰۰,۰۰۰	۸۲	۸
۹	۱۶۵,۰۰۰,۰۰۰	۷۵	۸	۲۰	۱۲۵,۰۰۰,۰۰۰	۷۵	۶
۱۰	۱۶۰,۰۰۰,۰۰۰	۸۵	۷	۲۱	۱۵۰,۰۰۰,۰۰۰	۷۵	۸
۱۱	۱۶۰,۰۰۰,۰۰۰	۷۵	۸	۲۲	۲۲۵,۰۰۰,۰۰۰	۹۰	۷

(منبع: یافته‌های پژوهش، ۱۴۰۵)

مجموعه داده نهایی شامل ۶۹۲۳ رکورد است که از این میان، ۳۵۸۴ نمونه متعلق به کلاس «اجاره‌بها در سطح معمول» و ۳۳۳۹ نمونه متعلق به کلاس «افزایش هزینه اجاره‌بها (Target=1)» می‌باشد. با توجه به تفاوت در توزیع این دو کلاس و به منظور پیشگیری از سوگیری مدل به نفع کلاس اکثریت، از تکنیک نمونه‌برداری بیش‌ازحد اقلیت مصنوعی^۱ استفاده شده است. این روش با تولید داده‌های مصنوعی برای کلاس اقلیت، توزیع داده‌ها را متوازن می‌سازد. در جدول ۷، مقایسه تعداد رکوردهای هر کلاس پیش و پس از اعمال این تکنیک ارائه شده است.

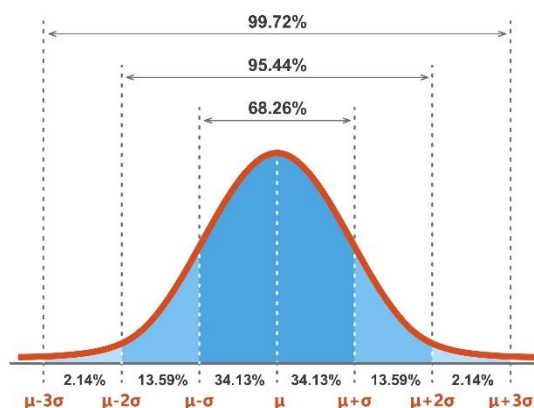
^۱ Missing Values^۲ Data Cleaning^۳ SMOTE

جدول ۷. روند تغییرات تعداد رکوردهای مجموعه داده در مراحل پیش پردازش و توازن سازی

وضعیت	مجموع کل رکوردهای معتبر	تعداد نمونه های کلاس "سطح معمول" (۰)	تعداد نمونه های کلاس "افزایش هزینه" (۱)
پیش از SMOTE	۶,۹۲۳	۳,۵۸۴	۳,۳۳۹
پس از SMOTE	۷,۱۶۸	۳,۵۸۴	۳,۵۸۴

(منبع: یافته های پژوهش، ۱۴۰۵)

در این پژوهش، متغیر هدف به صورت یک متغیر دودویی^۱ تعریف شده است؛ به طوری که مقدار (۰) نشان دهنده "سطح معمول اجاره بها (کلاس اکثریت)" و مقدار (۱) نشان دهنده "افزایش هزینه اجاره بها (کلاس اقلیت)" است. این کدگذاری عددی، زیربنای مدل سازی برای پیش بینی و تحلیل احتمال وقوع افزایش اجاره بها در مناطق مختلف تهران را فراهم آورده است. در پژوهش های تجربی، ویژگی ها اغلب دارای مقیاس ها و دامنه های عددی متفاوتی هستند که می تواند بر عملکرد مدل های یادگیری ماشین تأثیر منفی بگذارد؛ از این رو، به منظور هم مقیاس سازی، از روش "نمره استاندارد" برای استاندارد سازی داده ها استفاده شد. این تکنیک با انتقال توزیع داده ها به شکلی که میانگین آن ها برابر با صفر و انحراف معیار آن ها برابر با یک شود، مقیاس های متفاوت ویژگی ها را هم تراز می کند (Géron, 2022).



شکل ۳. فرآیند استاندارد سازی داده ها (Z-score) و نرمال سازی ویژگی ها جهت هم مقیاس سازی (منبع: Géron, 2022)

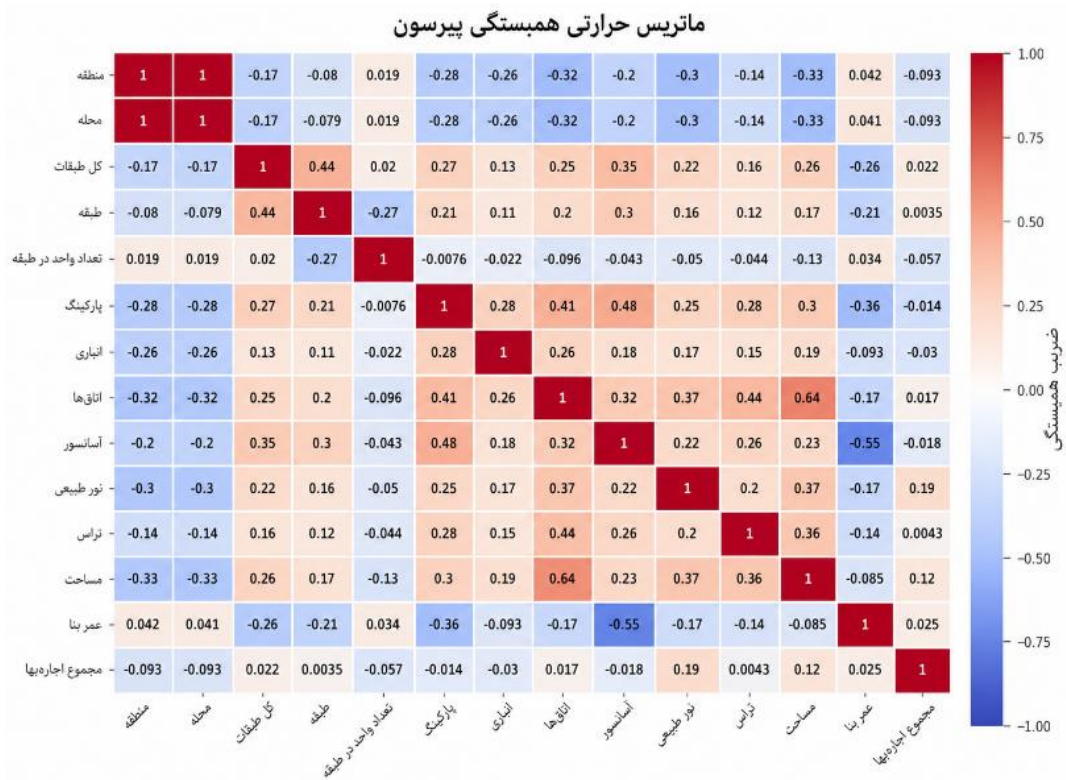
علاوه بر این، جهت تحلیل هم بستگی^۲ میان متغیرها، از "ضریب هم بستگی پیرسون" استفاده گردید. این شاخص بیانگر شدت و جهت رابطه خطی بین متغیرهای مستقل و وابسته است؛ به طوری که مقادیر نزدیک به (۱)، نشان دهنده هم بستگی قوی بین دو ویژگی است. نتایج حاصل از تحلیل هم بستگی داده ها در شکل ۴ ارائه شده است. در نهایت، به منظور بهینه سازی عملکرد مدل و کاهش پیچیدگی محاسباتی، از روش "انتخاب ویژگی بهینه" استفاده شد. در این فرآیند، پس از حذف متغیر هدف، ویژگی هایی که بیشترین تأثیر را در تبیین خروجی دارند، شناسایی شدند. براساس این تحلیل، ویژگی های نهایی منتخب شامل شناسه، منطقه، محله، تعداد اتاق ها، مساحت، عمر بنا و مجموع اجاره بها است.

¹ Binary

² Z-score

³ Correlation

⁴ Optimal Feature Selection



شکل ۴. نقشه حرارتی همبستگی پیرسون بین ویژگی‌ها

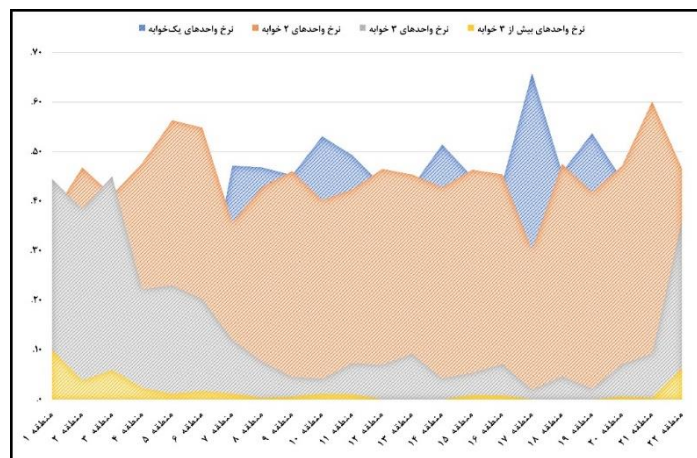
(منبع: یافته‌های پژوهش مبتنی بر داده‌های پلتفرم دیوار، ۱۴۰۲)

همان‌طور که در شکل ۴ مشاهده می‌شود، ماتریس همبستگی پیرسون برای شناسایی ارتباط خطی میان ویژگی‌های مجموعه داده رسم شده است. در این فرآیند، ویژگی‌هایی با بالاترین ضریب همبستگی با متغیر هدف و سایر متغیرهای مستقل، جهت بهبود دقت مدل انتخاب شدند؛ برای نمونه، همبستگی مثبت و قابل توجهی با ضریب ۰,۶۴ میان دو ویژگی "تعداد اتاق‌ها" و "مساحت" مشاهده شد که نشان‌دهنده هم‌خطی کم‌نادر میان این دو متغیر در بازار اجاره مسکن تهران است. این رویکرد به ما کمک کرد تا با حذف متغیرهای زائد، مدل‌سازی را بر ویژگی‌های اثرگذار متمرکز کنیم.

پردازش داده‌های احصاء شده از پلتفرم دیوار، اطلاعاتی پیرامون نرخ اتاق بر واحد مسکونی را در اختیار ما قرار داد که نشان‌دهنده یک شکاف ساختاری و قطب‌بندی فضایی ملموس در خطوط جغرافیایی شمال-جنوب و مرکز-پیرامون شهر تهران است. این تحلیل را می‌توان در سه سطح کلیدی تبیین نمود:

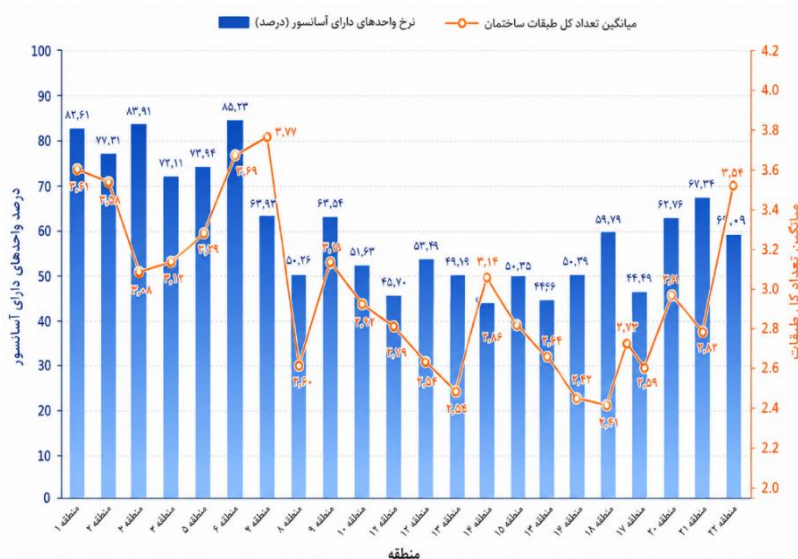
¹ Pearson Correlation Heatmap

² Multicollinearity



شکل ۵. نمودار نرخ اتاق بر واحد مسکونی در مناطق ۲۲ گانه شهر تهران
(منبع: یافته‌های پژوهش مبتنی بر داده‌های پلتفرم دیوار، ۱۴۰۲)

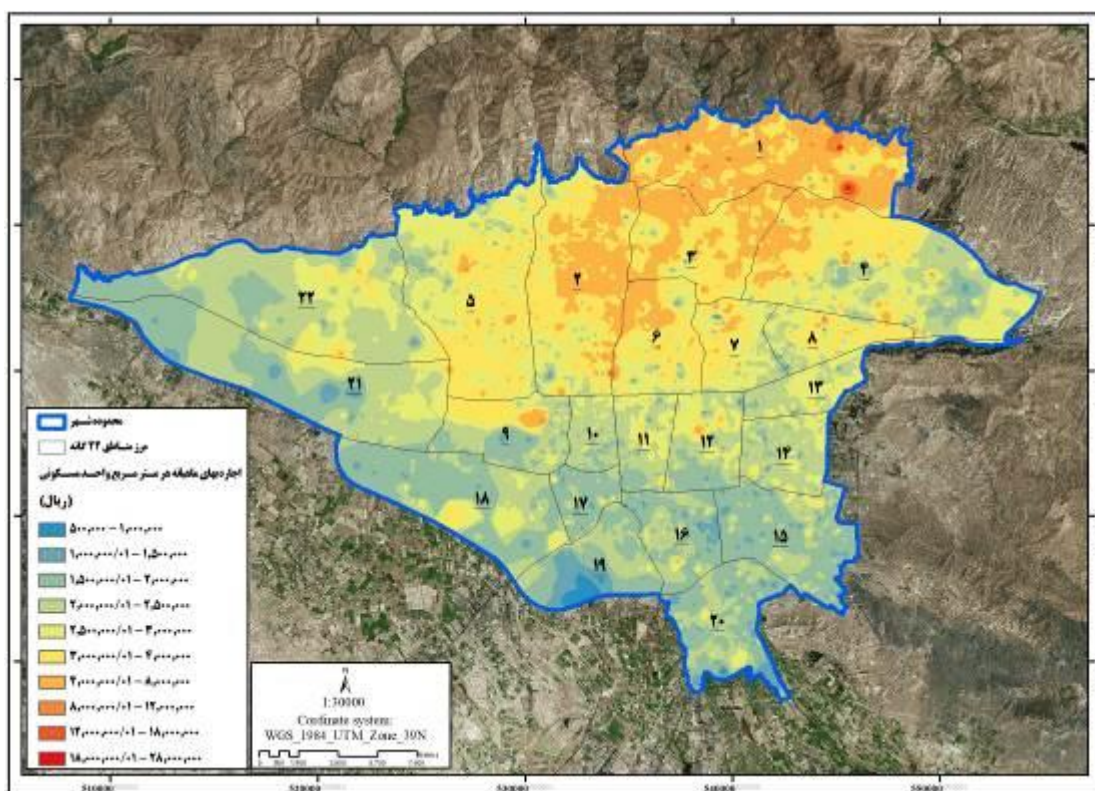
این نابرابری شدید فضایی، سیگنال‌های مستقیمی به بازار اجاره‌بها صادر می‌کند؛ جایی که مستأجران کم‌درآمد در بافت‌های متراکم مرکزی و جنوبی (مانند مناطق ۱۷، ۱۹، ۱۰ و ۱۴) با وجود پرداخت سهم بزرگی از درآمد خود، ناگزیر به پذیرش «تک‌اتاقه‌شدن» و سکونت در واحدهای یک‌خوابه هستند (مانند منطقه ۱۷ که نزدیک به ۶۶ درصد بازار اجاره آن‌را واحدهای یک‌خوابه تشکیل می‌دهند). تلاقی بُعد خانوار بالا و کمبود شدید اتاق خواب در این پهنه‌ها، ارزش اجاره‌ای «هر متر مربع حریم خصوصی» را به شدت بالا برده و خانوارها را دچار تراکم شدید هم‌سکونت‌ی کرده‌است؛ در مقابل، در مناطق مرقه (مانند ۱ و ۳)، بازار اجاره تحت سیطره واحدهای بزرگ‌مترای و چندخوابه (بالای ۴۴ درصد) است؛ موضوعی که نشان می‌دهد فرمول اجاره‌بها در تهران به جای قیمت‌گذاری براساس کیفیت زیست و کفایت فضایی، براساس قطب‌بندی طبقاتی شکل گرفته و طبقه فرودست را عملاً به پرداخت هزینه برای بقا در فقر مسکن محکوم کرده‌است.



شکل ۶. نمودار نرخ واحدهای دارای آسانسور و میانگین طبقات در مناطق ۲۲ گانه شهر تهران
(منبع: یافته‌های پژوهش مبتنی بر داده‌های پلتفرم دیوار، ۱۴۰۲)

فرمول تعیین اجاره‌بها در بازار مسکن تابع رفاه و کیفیت سازه است. در مناطق مرفه و گران قیمت شمال و مرکز شهر (مناطق ۱ تا ۶) وجود آسانسور با نرخ بالای ۷۲ تا ۸۵ درصد به یک استاندارد بدیهی و پیش فرض تبدیل شده و بهانه‌ای برای جهش اجاره‌بهاست، در مناطق جنوبی و کم‌درآمدتر (مانند ۱۱، ۱۴ و ۱۶) این امکان رفاهی با سقوط به زیر ۴۵ درصد همچنان یک کالای لوکس و نایاب به‌شمار می‌رود. اوج این نابرابری ساختاری در منطقه ۷ رخ می‌دهد؛ جایی که با وجود داشتن بلندترین ساختمان‌ها با میانگین ۳/۷۷ طبقه، نرخ آسانسور آن تنها ۶۳ درصد است که این یعنی مستأجران این منطقه بیشترین بهای آسیب جسمی را در قبال میانگین اجاره‌بهای خود پرداخت می‌کنند. در این میان، تنها مناطق ۲۰ و ۲۱ هستند که با تعداد طبقات کنترل شده (زیر ۳ طبقه) و نرخ آسانسور مناسب (بالای ۶۲ درصد)، یک پناهگاه اقتصادی و رفاهی پنهان را برای مستأجران هوشمند فراهم می‌کند تا بدون پرداخت هزینه‌های گزاف شمال شهر، از رفاه فیزیکی بهتری برخوردار شوند.

هزینه‌های اجاره‌بها یکی از دغدغه‌های اصلی افرادی است که به دلیل شرایط اقتصادی، توانایی خرید مسکن را ندارند. زندگی در تهران با هزینه‌های بالایی همراه است و انتخاب محله، تأثیر مستقیمی بر نرخ اجاره‌بها دارد؛ از این رو، بررسی دقیق میانگین کلی و جزئیات اجاره‌بها در هر منطقه ضروری است. تهران دارای ۲۲ منطقه است که نرخ اجاره‌بها در برخی از آن‌ها به دلایلی نظیر: برخورداری از امکانات رفاهی بهتر، دسترسی‌های مطلوب‌تر و زیرساخت‌های نوین، به مراتب بالاتر است؛ در نتیجه، واکاوی دقیق میانگین نرخ اجاره‌بها به صورت کلی و همچنین جزئیات هر منطقه امری اجتناب‌ناپذیر است. در شکل ۷، نقشه حرارتی قیمت اجاره‌بهای هر متر مربع آپارتمان‌های مسکونی در مناطق ۲۲ گانه تهران ترسیم شده است.



شکل ۷. نقشه حرارتی نرخ اجاره‌بهای مسکن در مناطق ۲۲ گانه تهران

(منبع: یافته‌های پژوهش مبتنی بر داده‌های پلتفرم دیوار، ۱۴۰۲)

تحلیل توزیع فضایی متوسط اجاره‌بهای هر مترمربع واحد مسکونی در مناطق ۲۲ گانه شهر تهران (براساس داده‌های زمستان ۱۴۰۲)، فراتر از یک نمایش کارتوگرافیک ساده از قیمت‌ها، به مثابه سندی از بازتولید ساختاری نابرابری و طرد اجتماعی است. در چارچوب اقتصاد سیاسی فضا، «نرخ اجاره‌بها» صرفاً هزینه‌بهای یک سرپناه نیست؛ بلکه جریان نقدی مستمری است که رانت زمین را از جیب طبقات حقوق‌بگیر و فرودست به سمت مالکان و سرمایه‌داران مستغلات کانال‌کشی می‌کند. تضاد فضایی شدیدی که در این نقشه میان شمال و جنوب تهران خودنمایی می‌کند، نشان‌دهنده یک گسست کالبدی-طبقاتی است که در آن، بازار مسکن به ابزار اصلی سلب استطاعت مالی و به تبع آن، فروپاشی امنیت تصدی برای دهک‌های پایین جامعه تبدیل شده است.

در بازه زمانی مورد مطالعه، میانگین اجاره ماهیانه یک آپارتمان مسکونی ۱۰۰ متری در تهران حدود ۲۵۰ تا ۳۰۰ میلیون ریال بود؛ این رقم حداقل دو برابر «حداقل حقوق مصوب در سال ۱۴۰۲ برای یک خانواده ۳ نفره» است که اکثر مستاجران توانایی پرداخت آن را ندارند. اگرچه مالکان برای تسهیل شرایط، دریافت «ودیعه بالا» را جایگزین بخشی از اجاره‌بها می‌کنند، در این پژوهش به منظور یکپارچه‌سازی داده‌ها، مبالغ و دیعه با استفاده از فرمول‌های استاندارد به اجاره‌بهای خالص تبدیل گردید. تحلیل توزیع فضایی (شکل ۷) نشان می‌دهد که مناطق ۱، ۲، ۳ و بخش‌هایی از مناطق ۴ و ۶، فراتر از میانگین اجاره‌بهای تهران قرار دارند؛ در مقابل، مناطق جنوبی با میانگین ۱۰۰ تا ۲۰۰ میلیون ریال برای مترآژ مشابه، کمترین نرخ اجاره‌بها را ثبت کرده‌اند. این رفتارهای فضایی نامتقارن را می‌توان در قالب سه پهنه ساختاری تبیین کرد:

پهنه انباشت، رانت فضایی و استثمار جغرافیایی (شمال و شمال غرب شهر): در این پهنه که عمدتاً مناطق ۱، ۲، ۳ و بخش‌هایی از ۵ و ۶ را دربرمی‌گیرد و در نقشه با لکه‌های قرمز، نارنجی و زرد تند مشخص شده است، مسکن از یک «کالای زیستی-مصرفی» به یک «کالای سرمایه‌ای خالص» دگرگون شده است. در این جغرافیا، اعداد شگفت‌انگیز ۱۲ تا ۲۸ میلیون ریال به‌ازای هر مترمربع اجاره‌بها، بازتاب‌دهنده حبابی نیست که تصادفی شکل گرفته باشد، بلکه نتیجه تمرکزگرایی شدید سرمایه، رانت توپوگرافیک و قطبی‌سازی عامدانه فضا است. مدیریت شهری در دهه‌های گذشته از طریق مکانیزم اعطای تراکم در این پهنه، عملاً نوعی ارزش‌افزوده فضایی ایجاد کرده که تنها در انحصار طبقات خاص است. این پهنه، با جذب لایه‌های اصلی سرمایه‌های سرگردان، «خط‌کش قیمتی» شهر تهران را تعیین می‌کند؛ یعنی گرانی لجام‌گسیخته در پهنه قرمز، ساختار قیمت را در پهنه‌های پایین‌تر به صورت زنجیره‌ای بالا می‌کشد بدون آن‌که کیفیت زیست در پهنه‌های پایینی افزایشی یافته باشد. این پدیده، استطاعت‌ناپذیری را به صورت عمودی به کل پیکره شهر تزریق می‌کند و اولین جرقه «فقدان امنیت تصدی» را از طریق رانش مستأجران لایه‌های میانی به لایه‌های فرودست می‌زند.

پهنه ضربه‌گیر، گذار و بورس‌بازی مدرن (مرکز، غرب و شبه جزیره‌های توسعه): این پهنه که لکه‌های زرد کم‌رنگ و سبز در مناطق ۴، ۶، ۷، ۲۱ و به‌ویژه ۲۲ را شامل می‌شود، جالب‌ترین رفتار فضایی را از خود نشان می‌دهد. منطقه ۲۲ به عنوان یک شبه جزیره توسعه یافته در غرب، با دسترسی به اتوبان‌ها و برج‌های نوین، به عنوان یک راهبرد برنامه‌ریزی برای سرریز جمعیت طراحی شده بود، اما نقشه نشان می‌دهد که این منطقه عملاً در بازه قیمتی متوسط (۳ تا ۶ میلیون ریال) تثبیت شده و نتوانسته به مسکنی استطاعت‌پذیر برای فقرای شهری تبدیل شود؛ چرا

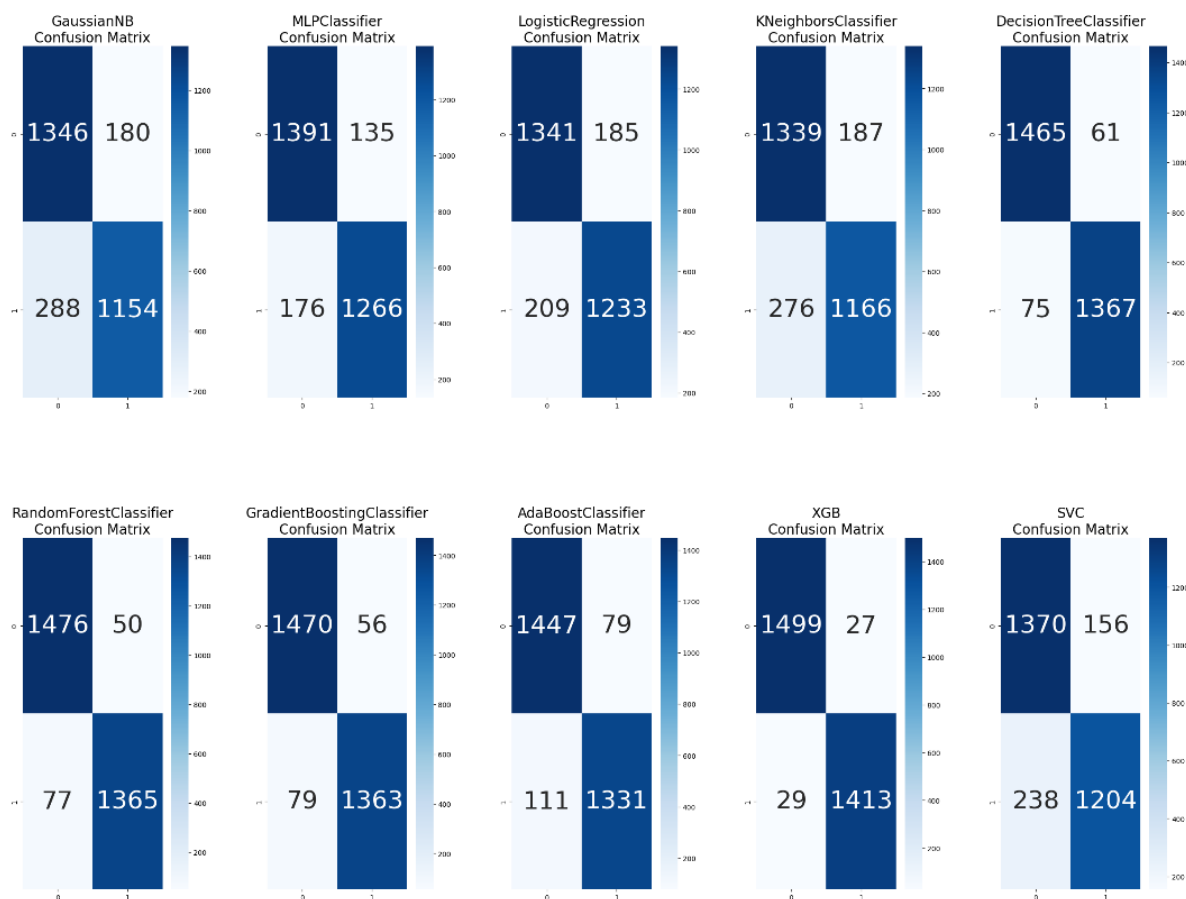
^۱ Spatial Surplus Value

که منطق حاکم بر این مناطق نیز منطق مالی سازی مسکن است. در پهنه‌های مرکزی مانند منطقه ۶ و ۷، داده‌ها نشان می‌دهد که الگوی غالب مسکن در این مناطق به سمت واحدهای متوسط و کوچک‌متراژ سوق یافته‌است. خانوار طبقه متوسط (کارمندان، متخصصان و بدنه دانشگاهی) برای حفظ پیوند جغرافیایی خود با مرکز شهر، بورس‌های شغلی و دسترسی به زیرساخت‌ها، حاضر است استانداردهای رفاهی فضا را سرکوب کرده و به واحدهای با متراژ بسیار کمتر تن دهد تا تعادل میان درآمد و اجاره‌بهای ۳ تا ۶ میلیون ریالی این پهنه را حفظ کند. براساس مستندات طرح جامع تهران (نهاد مشترک تهیه و راهبری طرح‌های توسعه شهری تهران، ۱۳۸۶) و همچنین داده‌های پلتفرم دیوار (۱۴۰۲)، مناطق ۶ و ۷ دارای تراکم بالای کاربری‌های اداری و تجاری پنهان (در قالب شرکت‌ها، کلینیک‌ها و دفاتر) در واحدهای مسکونی هستند. این هجوم سرمایه اداری-تجاری به بخش مسکونی، ارزش اجاره‌بهای زمین را به طور مصنوعی بالا می‌برد. در این بازار، مستاجر مسکونی در یک رقابت نابرابر با «سرمایه تجاری» قرار می‌گیرد که نتیجه آن، تهدید مداوم امنیت تصدی به دلیل ناتوانی در تمدید قراردادهای سالانه است.

پهنه تله فضایی، مادون استطاعت و سلب مالکیت زیستی (جنوب، جنوب شرق و جنوب غرب): این پهنه وسیع که با رنگ‌های آبی تند، آبی آسمانی و سبز تیره در مناطق ۱۵، ۱۶، ۱۷، ۱۸، ۱۹ و ۲۰ خودنمایی می‌کند، جغرافیای فقر مسکن در تهران است. قیمت‌های زیر ۲٫۵ میلیون ریال (و حتی پهنه‌های آبی تیره زیر ۱٫۵ میلیون ریال در منطقه ۱۹ و ۱۸) در نگاه نخست شاید «استطاعت‌پذیر» به نظر برسند؛ اما در تحلیل عمیق، این پهنه یک «تله فضایی فقر» است. نرخ پایین اجاره‌بها در این جا معلول یک سقوط ساختاری است: سقوط کیفیت کالبدی، مخاطرات بالای محیطی (فرونشست زمین و گسل‌های فعال جنوب)، آلودگی‌های زیست‌محیطی و انزوای جغرافیایی از فرصت‌های شغلی باکیفیت. در این پهنه، فقدان امنیت تصدی ابعاد وحشتناک‌تری به خود می‌گیرد. بازار اجاره در این مناطق به شدت غیررسمی، مبتنی بر قراردادهای شفاهی یا پشت‌نویسی‌های فاقد اعتبار قانونی است. مستأجرانی که در این جغرافیا ساکن هستند، کمترین ضریب تاب‌آوری اقتصادی را دارند؛ بنابراین، حتی یک نوسان جزئی در نرخ تورم، منجر به رانش نهایی آن‌ها به خارج از محدوده شهر (به سمت اسلامشهر، ملارد، پاکدشت و غیره) می‌شود. به منظور ارزیابی دقت پیش‌بینی و شاخص‌های عملکردی شامل دقت، صحت، فراخوانی، شاخص F و منحنی مشخصه عملکرد سیستم^۲ ده الگوریتم یادگیری ماشین مورد ارزیابی قرار گرفتند. مدل‌های آزمایش شده عبارت بودند از: درخت تصمیم (DT)، ماشین بردار پشتیبان (SVC)، بیزی ساده گاوسی (GNB)، K-نزدیک‌ترین همسایه (K-NN)، رگرسیون لجستیک (LR)، پرسپترون چندلایه (MLP)، جنگل تصادفی (RF) و الگوریتم‌های تقویت (Gradient Boosting, AdaBoost, XGBoost).

^۱ Financialization of Housing^۲ Receiver Operating Characteristics (ROC)

Confusion Matrices



شکل ۸. ماتریس درهم‌ریختگی مدل طبقه‌بندی اجاره‌بها

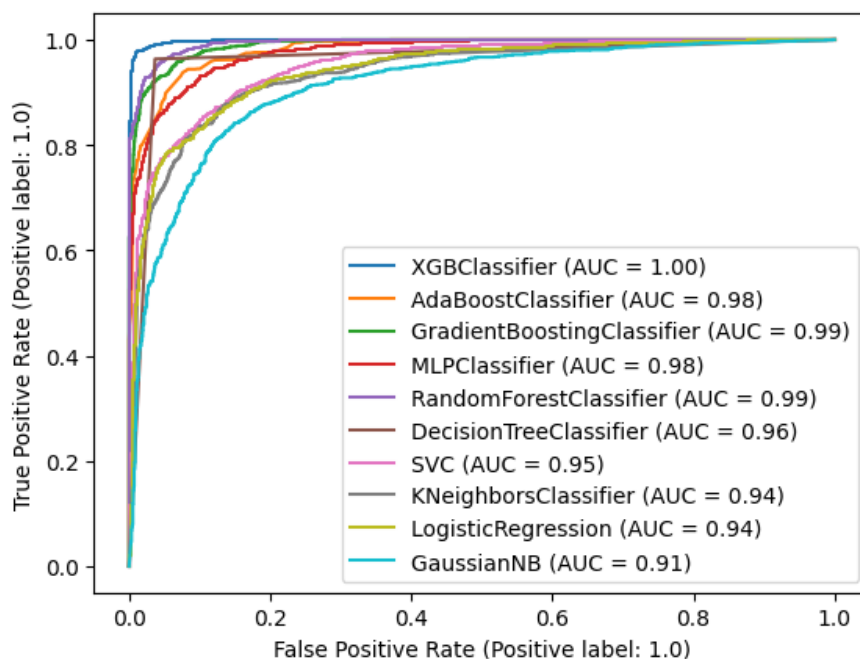
(منبع: یافته‌های پژوهش، ۱۴۰۵)

الگوریتم‌های یادگیری ماشین، ۱۵۲۶ رکورد را از میان ۶۹۲۳ رکوردی که به‌عنوان ورودی در نظر گرفته شده‌بود، به‌عنوان داده‌های آزمون انتخاب و پردازش کرد. ماتریس درهم‌ریختگی (شکل ۸) عملکرد طبقه‌بندی مدل را به‌صورت بصری نشان می‌دهد و تعداد موارد مثبت واقعی، مثبت کاذب، منفی واقعی و منفی کاذب را مشخص می‌کند؛ این ماتریس برای درک توانایی‌های پیش‌بینی مدل و شناسایی حوزه‌های نیازمند بهبود، امری حیاتی است. جدول ۸ خلاصه‌ای از نتایج ارزیابی عملکرد هر مدل را ارائه می‌دهد و شاخص‌های مربوط به هر یک را برجسته می‌سازد.

جدول ۸. نتایج ارزیابی عملکرد مدل‌های پیش‌بینی

مساحت زیر منحنی (AUC) ^۱	شاخص F1	فراخوانی	دقت	صحت	مدل‌های یادگیری ماشین
XGBoost	۰,۹۸۱۱۳۲	۰,۹۸۱۲۵۰	۰,۹۷۹۸۸۹	۰,۹۸۰۵۶۹	۰,۹۹۸۳۶۱
AdaBoost	۰,۹۲۴۱۹۱	۰,۹۴۱۹۰۳	۰,۸۹۹۴۴۵	۰,۹۲۰۱۸۴	۰,۹۸۱۴۲۷
Gradient Boosting (GB)	۰,۹۴۱۷۱۲	۰,۹۴۷۷۷۷	۰,۹۳۱۳۴۵	۰,۹۳۹۴۸۹	۰,۹۹۰۰۱۳
MLP Classifier	۰,۹۱۰۳۷۷	۰,۹۳۲۹۹۰	۰,۸۷۸۶۴۱	۰,۹۰۵۰۰۰	۰,۹۷۶۰۲۷
Random Forest (RF)	۰,۹۵۳۱۶۷	۰,۹۶۰۴۲۴	۰,۹۴۲۴۴۱	۰,۹۵۱۳۴۸	۰,۹۹۳۴۱۱
Decision Tree (DT)	۰,۹۶۳۲۷۵	۰,۹۶۱۲۴۶	۰,۹۶۳۲۴۵	۰,۹۶۲۲۴۵	۰,۹۶۳۲۷۴
SVC	۰,۸۷۲۳۰۵	۰,۸۹۱۰۹۶	۰,۸۳۹۸۰۶	۰,۸۶۴۶۹۱	۱,۰۰
K-Nearest Neighbors (K-NN)	۰,۸۶۵۹۰۳	۰,۸۷۹۹۱۳	۰,۸۳۸۴۱۹	۰,۸۵۸۶۶۵	۰,۹۳۵۳۹۰
Logistic Regression (LR)	۰,۸۶۷۲۵۱	۰,۸۶۹۵۳۵	۰,۸۵۵۰۶۲	۰,۸۶۲۳۳۸	۰,۹۴۱۲۷۱
Naive Bayes (NB)	۰,۸۴۲۳۱۸	۰,۸۶۵۰۶۷	۰,۸۰۰۲۷۷	۰,۸۳۱۴۱۲	۰,۹۱۲۷۰۵

(منبع: یافته‌های پژوهش، ۱۴۰۵)



شکل ۹. منحنی مشخصه عملکرد سیستم (ROC) برای مدل‌های پیش‌بینی

(منبع: یافته‌های پژوهش، ۱۴۰۵)

منحنی مشخصه عملکرد سیستم (شکل ۹)، بیانگر مبادله میان حساسیت (نرخ مثبت واقعی) و ویژگی (نرخ منفی واقعی) در آستانه‌های مختلف تنظیمات است و مساحت زیر منحنی (AUC) ROC نیز عملکرد کلی مدل را اندازه‌گیری می‌کند. با توجه به تحلیل نتایج، مدل ایکس‌جی‌بوست به دلیل برخورداری از بالاترین سطح صحت و دقت، بهترین گزینه برای پیش‌بینی اجاره‌بهای مسکن در تهران برای سال آینده شناخته می‌شود. این مدل با دستیابی به

^۱ Area Under the Curve (AUC)

صحت ۹۸/۱ درصد، دقت ۹۸/۱ درصد و شاخص F1 برابر با ۰/۹۸، عملکردی متمایز از خود نشان داد؛ همچنین، کسب امتیاز ۰/۹۹۸ در شاخص ROC-AUC حاکی از توانایی استثنایی این مدل در تفکیک دسته‌های مختلف اجاره‌بها است. مدل‌های درخت تصمیم (با صحت ۹۶/۳ و دقت ۹۶/۱ درصد) و جنگل تصادفی (با صحت ۹۵/۶ و دقت ۹۶ درصد) نیز در جایگاه‌های بعدی قرار گرفت.

عملکرد موفقیت‌آمیز مدل‌ها در انطباق برچسب‌های پیش‌بینی‌شده با مقادیر واقعی، مؤید دقت بالای این رویکرد است. این پژوهش با استفاده از داده‌های اجاره‌بهای مسکن در پلتفرم «دیوار» تهران انجام شد؛ هرچند حجم نمونه، بسیار بزرگ نبود، اما مدل‌ها نتایج قابل‌قبولی ارائه داد.

۴) بحث و نتیجه‌گیری

در این پژوهش، به پیش‌بینی اجاره‌بهای مسکن در مناطق ۲۲ گانه شهر تهران با استفاده از الگوریتم‌های یادگیری ماشین نظارت‌شده پرداخته شد؛ حوزه‌ای که در مقایسه با پیش‌بینی قیمت خرید و فروش مسکن، به‌طور قابل‌توجهی در پژوهش‌های منتشرشده نادیده گرفته شده بود. از منظر روش‌شناختی، یکی از تمایزات بارز پژوهش حاضر نسبت به مطالعات پیشین، استفاده از داده‌های مقطعی به‌جای رویکردهای رایج مبتنی بر داده‌های طولی است. این انتخاب روش‌شناختی این امکان را فراهم آورد تا یک "تصویر لحظه‌ای" و بسیار دقیق از وضعیت بازار اجاره مسکن تهران در یک مقطع زمانی خاص (آذر و دیماه ۱۴۰۲) تحلیل شود. علاوه بر این، اتکا بر نظر خبرگان مسکن برای تعریف شروط هدف (شامل: ترکیبی از میزان اجاره، متراژ و عمر بنا در هر منطقه) سبب شد تا مدل‌ها براساس واقعیات کالبدی و اقتصادی مختص به هر یک از مناطق ۲۲ گانه تهران آموزش ببینند و این امر به‌نوبه خود رویکردی نوآورانه و بومی‌سازی‌شده به‌شمار می‌رود.

نتایج پژوهش نشان داد که توزیع اجاره‌بهای مسکن در تهران از یک الگوی فضایی قطبی پیروی می‌کند؛ به‌گونه‌ای که مناطق شمالی و برخوردار، به‌ویژه مناطق ۱، ۲ و ۳، بالاترین نرخ اجاره‌بها و مناطق جنوبی، از جمله مناطق ۱۵ تا ۲۰، پایین‌ترین مقادیر را دارند؛ باین‌حال، پایین بودن اجاره‌بها در مناطق جنوبی الزاماً به‌معنای استطاعت‌پذیری بیشتر نیست؛ زیرا این مناطق غالباً با کیفیت کالبدی پایین‌تر، دسترسی محدودتر به فرصت‌های شغلی و خدمات شهری و هزینه‌های بیشتر رفت‌وآمد مواجه‌اند؛ بنابراین، تفاوت اجاره‌بها را باید بازتاب هم‌زمان ارزش زمین، دسترسی، کیفیت محیطی و موقعیت اجتماعی مناطق دانست، نه صرفاً تفاوت در ویژگی‌های واحد مسکونی؛ هرچند که این الگوی فضایی در کیفیت و کفایت فضای سکونت نیز بازتولید شده است. در مناطق مرکزی و جنوبی مانند ۱۰، ۱۴، ۱۷ و ۱۹، سهم واحدهای یک‌خوابه بالاست؛ به‌طوری‌که در منطقه ۱۷ حدود ۶۶ درصد واحدهای اجاره‌ای را این گروه تشکیل می‌دهد. در مقابل، در مناطق برخوردار ۱ و ۳، بیش از ۴۴ درصد واحدها دارای متراژ و تعداد اتاق بیشتری هستند. تفاوت در برخورداری از آسانسور نیز همین شکاف را تأیید می‌کند؛ به‌گونه‌ای که نرخ برخورداری در مناطق ۱ تا ۶ بین ۷۲ تا ۸۵ درصد و در مناطقی مانند ۱۱، ۱۴ و ۱۶ کمتر از ۴۵ درصد است؛ در نتیجه، نابرابری بازار اجاره تهران، تنها در سطح قیمت رخ نمی‌دهد، بلکه در قالب تفاوت در تعداد اتاق، حریم خصوصی، امکانات بنا و کیفیت سکونت نیز نمایان می‌شود. از منظر اقتصاد شهری و جغرافیای اجتماعی، یافته‌های توصیفی و تحلیلی این پژوهش بازتاب‌دهنده واقعیت‌های تلخ و چالش‌برانگیز در بازار مسکن تهران است. نتایج نشان داد که میانگین اجاره‌بهای ماهانه یک آپارتمان ۱۰۰ متری در تهران (حدود ۲۵۰ تا ۳۰۰ میلیون ریال)،

حداقل دو برابر حداقل دستمزد مصوب برای یک خانواده سه نفره در سال ۱۴۰۲ است. نقشه حرارتی استخراج شده نیز تمرکز شدید این فشار قیمتی را در مناطق ۱، ۲، ۳ و بخش هایی از مناطق ۴ و ۶ نشان می دهد که به مراتب بالاتر از میانگین شهر هستند. با توجه به این که حدود ۵۱ درصد از ساکنان تهران را مستأجران تشکیل می دهند، این شکاف عمیق میان درآمد و هزینه های مسکن، در کنار قراردادهای کوتاه مدت یک ساله و تورم بالا، پدیده اجاره نشینی در پایتخت را از یک «انتخاب» به یک «اجبار محض» تبدیل کرده است.

چند بُعدی بودن این روابط، ضرورت استفاده از روش هایی را نشان می دهد که بتوانند اثر هم زمان و غیر خطی متغیرهای مکانی و کالبدی را شناسایی کنند. در همین راستا، در این پژوهش الگوریتم های درخت تصمیم (DT)، ماشین بردار پشتیبان (SVC)، بیزی ساده گاوسی (GNB)، K-نزدیک ترین همسایه (K-NN)، رگرسیون لجستیک (LR)، پرسپترون چندلایه (MLP)، جنگل تصادفی (RF) و الگوریتم های تقویت (Gradient Boosting, AdaBoost, XGBoost) مورد ارزیابی گرفت. نتایج این مطالعه نشان داد که الگوریتم های پیشرفته یادگیری ماشین، به ویژه مدل های مبتنی بر تقویت گرادیان و یادگیری گروهی، ظرفیت بالایی در مدل سازی پویایی های بازار اجاره دارند. ثبت صحت ۹۸٫۱ درصد و دقت ۹۸/۱ درصدی برای مدل ایکس جی بوست و قرارگیری مدل های درخت تصمیم (با صحت ۹۶/۳ و دقت ۹۶/۱ درصد) و جنگل تصادفی (با صحت ۹۵/۶ و دقت ۹۶ درصد)، برتری مطلق این مدل ها را در پیش بینی طبقه بندی اجاره بها اثبات می کند. دلیل اصلی عملکرد خیره کننده این مدل ها، توانایی ذاتی آن ها در کشف و درک روابط پیچیده و غیر خطی میان متغیرهای متعدد بازار املاک (نظیر: متراژ، عمر بنا، امکانات و موقعیت مکانی) و همچنین مقاومت بالای آن ها در برابر بیش برآزش است. این یافته ها با روند روبه رشد مطالعات جهانی که کارایی بالای الگوریتم ایکس جی بوست را در ارزیابی انبوه املاک تأیید کرده اند، کاملاً مطابقت دارد (Iban, 2022; Sharma, Harsora, & Ogunleye, 2024). در چنین بازار پیچیده و پرنوسانی که ویژگی هایی نظیر: عدم شفافیت، بی ثباتی اقتصادی و عدم تقارن اطلاعاتی در آن موج می زند، کاربردی سازی مدلی نظیر ایکس جی بوست که در این مطالعه توسعه یافت، می تواند با ارائه پیش بینی های دقیق و داده محور، ابهامات قیمتی را برای مستأجرانی که به دنبال مسکن مقرون به صرفه هستند کاهش دهد و به مالکان و سرمایه گذاران در تعیین قیمت های منصفانه و منطقی کمک کند. از سوی دیگر، نهادهای حاکمیتی می توانند از خروجی های این مدل ها به عنوان ابزاری تحلیلی برای رصد بازار، سیاست گذاری های مالیاتی و اجرای طرح های حمایتی هدفمند در حوزه مسکن استیجاری بهره برداری نمایند.

با وجود نتایج بسیار مطلوب، این مطالعه با محدودیت هایی در حجم و تنوع داده ها روبه رو بود. همان گونه که یافته ها تأیید می کند، غنی سازی مدل ها با افزودن شاخص های کلان اقتصادی و ویژگی های اجتماعی-اقتصادی محلات، در کنار بهره گیری از حجم داده های کلان تر و استفاده از رویکردهای یادگیری عمیق می تواند افق های جدیدی را برای پژوهش های آتی در این حوزه بگشاید. برای پایش و تنظیم این بازار، باید متغیرهای اقتصادی، مکانی و کالبدی به صورت یکپارچه بررسی شوند؛ از این رو، ترکیب تحلیل فضایی و یادگیری ماشین می تواند مبنایی برای شناسایی مناطق در معرض فشار اجاره و هدفمند کردن مداخلات مدیریت شهری فراهم کند؛ مشروط بر آن که این ابزارها در خدمت ارتقای استطاعت پذیری، امنیت تصدی و عدالت فضایی قرار گیرد.

منابع

- پلتفرم آنلاین دیوار. (۱۴۰۲). آگهی‌های اجاره مسکن در تهران (منتشرشده از ۱۴۰۲/۰۹/۱۰ لغایت ۱۴۰۲/۱۰/۱۰). تهران: ایران. <https://divar.ir/>
- حسن‌گودرزی، سپیده و آرمان‌مهر، محمدرضا. (۱۳۹۷). تحلیل بازار مسکن و پیش‌بینی قیمت آن تا سال ۱۴۰۵؛ مطالعه موردی: شهر تهران. بررسی مسائل اقتصاد ایران، ۵(۲)، ۷۹-۱۰۳. https://economics.ihs.ac.ir/article_4310.html?lang=en
- حسینی‌رامندی، شیوا و کاشانی، حامد. (۱۴۰۲). مدلی ترکیبی برای تخمین قیمت مسکن؛ مطالعه موردی: شهر تهران. مهندسی سازه و ساخت، ۱۰(۱)، ۱۷۲-۱۸۹. <https://doi.org/10.22065/jsce.2022.294445.2496>
- شهرداری تهران. (۱۴۰۱). آمارنامه شهرداری تهران. انتشارات سازمان فناوری اطلاعات و ارتباطات شهرداری تهران، تهران: ایران. <https://data.tehran.ir/>
- رضائیان، سجاد؛ عسگری، حشمت‌الله و درویشی، باقر. (۱۳۹۸). بررسی عوامل تعیین‌کننده اجاره مسکن در شهر ایلام با رویکرد اقتصادسنجی فضایی هدائیک. اقتصاد و مدیریت شهری، ۷(۲)، ۱۵-۲۷. <https://iueam.ir/article-1-1189-fa.html>
- زالی، سعید؛ پهلوانی، پرهام و بیگدلی، بهناز. (۱۴۰۲). تحلیل فضایی-زمانی عوامل مؤثر بر قیمت مسکن؛ موردشناسی: منطقه ۵ شهرداری تهران. آمایش سرزمین، ۱۵(۱)، ۱۱۵-۱۳۰. <https://doi.org/10.22059/jtcp.2022.341584.670318>
- زیادی، حسین؛ صلواتی، عرفان و لطفی‌هروی، محمدمهدی. (۱۴۰۲). پیش‌بینی قیمت مسکن با استفاده از الگوریتم هوش مصنوعی LSTM. تحقیقات مالی، ۲۵(۴)، ۵۵۷-۵۷۶. <https://doi.org/10.22059/frj.2023.349924.1007398>
- صارمی، حمیدرضا؛ حیدری، محمد و آقایی، فاطمه. (۱۳۹۷). تحلیل فضایی قیمت مسکن با استفاده از تکنیک رگرسیون موزون جغرافیایی؛ مورد مطالعه: منطقه دو شهرداری تهران. اقتصاد شهری، ۳(۲)، ۱۹-۳۸. <https://doi.org/10.22108/ue.2018.109447.1056>
- مرکز آمار ایران. (۱۳۹۵). نتایج سرشماری نفوس و مسکن. تهران: ایران. <https://amar.org.ir/>
- نهاد مشترک تهیه و راهبری طرح‌های توسعه شهری تهران. (۱۳۸۶). طرح راهبردی-ساختاری (جامع) شهر تهران. مصوب شورای عالی شهرسازی و معماری ایران در سال ۱۳۸۶، تهران: ایران. <https://shorayeali.mrud.ir/>

References

- Ambrose, B. W., & Diop, M. (2018). Information Asymmetry, Regulations, and Equilibrium Outcomes: Theory and Evidence from the Housing Rental Market. *Microeconomics: Asymmetric & Private Information eJournal*. <https://doi.org/10.2139/ssrn.2851156>
- Bansal, M., Goyal, A., & Choudhary, (2022). A comparative analysis of K-Nearest Neighbor, Genetic, Support Vector Machine, Decision Tree, and Long Short Term Memory algorithms in machine learning. *Decision Analytics Journal*, 3, 100071. <https://doi.org/10.1016/j.dajour.2022.00071>
- Baur, K., Rosenfelder, M., & Lutz, B. (2023). Automated real estate valuation with machine learning models using property descriptions. *Expert Systems with Applications*, 213, 119147. <https://doi.org/10.1016/j.eswa.2022.119147>
- Bentéjac, C., Csörgő, A., & Martínez-Muñoz, G. (2021). A comparative analysis of gradient boosting algorithms. *Artificial Intelligence Review*, 54. <https://link.springer.com/article/10.1007/s10462-020-09896-5>
- Bisong, E. (2019). *Building Machine Learning and Deep Learning Models on Google Cloud Platform (A Comprehensive Guide for Beginners)* (1 ed.). Apress Berkeley, CA. <https://doi.org/10.1007/978-1-4842-4470-8>
- Blanco, A., & Razu, D. (2019). Rental Housing. In *The Wiley Blackwell Encyclopedia of Urban and Regional Studies* (pp. 1-6). Wiley. <https://doi.org/https://doi.org/10.1002/9781118568446.eurs0267>

- Bracke, P. (2013). How long do housing cycles last? A duration analysis for 19 OECD countries. *Journal of Housing Economics*, 22(3), 213-230. <https://doi.org/10.1016/j.jhe.2013.06.001>
- Braubach, M., Jacobs, D. E., & Ormandy, D. (2011). Environmental burden of disease associated with inadequate housing: a method guide to the quantification of health effects of selected housing risks in the WHO European Region (World Health Organization. Regional Office for Europe, Issue. <https://apps.who.int/iris/handle/10665/108587>
- Chen, Y., Liu, X., Li, X., Liu, Y., & Xu, X. (2016). Mapping the fine-scale spatial pattern of housing rent in the metropolitan area by using online rental listings and ensemble learning. *Applied Geography*, 75, 200-212. <https://doi.org/10.1016/j.apgeog.2016.08.011>
- Chicco, D., & Jurman, G. (2020). The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. *BMC genomics*, 21(1), 6.
- Clark, S. D., & Lomax, N. (2018). A mass-market appraisal of the English housing rental market using a diverse range of modelling techniques. *Journal of Big Data*, 5, Article 43. <https://doi.org/10.1186/s40537-018-0154-3>
- Cordero, E. A., & Gutiérrez, P. R. (2021). Social Housing in Guadalajara: A Viable Strategy to Mitigate the Negative Externalities of Metropolitan Urban Sprawl? *Applied Economics and Policy Studies*. https://doi.org/10.1007/978-981-16-5359-9_48
- David, W., Hosmer, J., Rodney, X., & Sturdivant, S. L. (۲۰۱۳). *Applied Logistic Regression* (3rd Edition ed.). Wiley & Sons. <https://www.wiley.com/en-us/Applied+Logistic+Regression%2C+3rd+Edition-p-9780470582473>
- De Giorgio, A., Cola, G., & Wang, L. (2023). Systematic review of class imbalance problems in manufacturing. *Journal of Manufacturing Systems*, 71, 620-644. <https://doi.org/10.1016/j.jmsy.2023.10.014>
- Diniz, P. S. R. (2023). *Signal Processing and Machine Learning Theory* (Academic Press Library in Signal Processing) (1st ed ed.). Academic Press. <https://doi.org/https://doi.org/10.1016/B978-0-32-391772-8.00019-3>
- Elariane, S. A. (2022). Location based services APIs for measuring the attractiveness of long-term rental apartment location using machine learning model. *Cities*, 103588, 122. <https://doi.org/10.1016/j.cities.2022.103588>
- Füss, R., & Koller, J. A. (2016). The role of spatial and temporal structure for residential rent predictions. *International Journal of Forecasting*, 32(4), 1352-1368. <https://doi.org/10.1016/j.ijforecast.2016.06.001>
- Gao, Q., Shi, V., Pettit, C., & Han, H. (2022). Property valuation using machine learning algorithms on statistical areas in Greater Sydney, Australia. *Land Use Policy*, 123, 106409. <https://doi.org/10.1016/j.landusepol.2022.106409>
- Garboden, P., & Rosen, E. (2019). Serial Filing: How Landlords Use the Threat of Eviction. *City & Community*, 18. <https://doi.org/10.1111/cico.12387>
- Géron, A. (2022). *Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow*. " O'Reilly Media, Inc".
- Ghaedrahmati, S., & Rezaei, E. (2024). Turkey, the second home for Iranians: push and pull motivations in Turkish housing market. *Journal of European Real Estate Research*, 17(1), 123-136. <https://doi.org/10.1108/JERER-06-2023-0019>
- Gopal, Patro, K., & Sahu, K. K. (۲۰۱۵). Normalization: A Preprocessing Stage. *ArXiv*, abs/1503.06462. <https://doi.org/10.48550/arXiv.1503.06462>
- Grybauskas, A., Pilinkienė, V., & Stundžienė, A. (2021). Predictive analytics using Big Data for the real estate market during the COVID-19 pandemic. *Journal of Big Data*, 8(1), 105. <https://doi.org/10.1186/s40537-021-00476-0>
- Gupta, S. C., & Goel, N. (2023). Predictive Modeling and Analytics for Diabetes using Hyperparameter tuned Machine Learning Techniques. *Procedia Computer Science*, 218, 1257-1269. <https://doi.org/10.1016/j.procs.2023.01.104>

- Haque, M. B., Choudhury, C., & Hess, S. (2020). Understanding differences in residential location preferences between ownership and renting: A case study of London. *Journal of Transport Geography*, 88, 102866. <https://doi.org/10.1016/j.jtrangeo.2020.102866>
- Hu, J., & Szymczak, S. (2023). A review on longitudinal data analysis with random forest. *Brief Bioinform*, 24(2). <https://doi.org/10.1093/bib/bbad002>
- Hu, L., He, S., Han, Z., Xiao, H., Su, S., Weng, M., & Cai, Z. (2019). Monitoring housing rental prices based on social media: An integrated approach of machine-learning algorithms and hedonic modeling to inform equitable housing policies. *Land Use Policy*, 82, 657-673. <https://doi.org/10.1016/j.landusepol.2018.12.030>
- Huang, Y. (2019). Predicting Home Value in California, United States via Machine Learning Modeling. *Statistics, Optimization & Information Computing*, 7. <https://doi.org/10.19139/soic.v7i1.435>
- Iban, M. C. (2022). An explainable model for the mass appraisal of residences: The application of tree-based Machine Learning algorithms and interpretation of value determinants. *Habitat International*, 128, 102660. <https://doi.org/10.1016/j.habitatint.2022.102660>
- Jacobs, K., & Manzi, T. (2020). Conceptualising 'financialisation': governance, organisational behaviour and social interaction in UK housing. *International Journal of Housing Policy*, 20(2), 184-202. <https://doi.org/10.1080/19491247.2018.1540737>
- Kang, J., Lee, H. J., Jeong, S. H., Lee, H. S., & Oh, K. J. (2020). Developing a Forecasting Model for Real Estate Auction Prices Using Artificial Intelligence. *Sustainability*, 12(7), 2899. <https://www.mdpi.com/2071-1050/12/7/2899>
- 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. <https://doi.org/10.1016/j.landusepol.2020.104919>
- Kim, T. Y., Park, E., & Ryu, D. (2025). Determinants of housing rental prices in Seoul: Applying explainable AI. *Spatial Economic Analysis*, 20(2), 312-332. <https://doi.org/10.1080/17421772.2024.2418906>
- Kishor, N. K., & Morley, J. (2015). What factors drive the price-rent ratio for the housing market? A modified present-value analysis. *Journal of Economic Dynamics and Control*, 58, 235-249. <https://doi.org/10.1016/j.jedc.2015.06.006>
- Kobzan, S., Ivakhnenko, A., & Tolsta, M. (2021). Research of features of rent market development. *Municipal economy of cities*, 1, 116-123. <https://doi.org/10.33042/2522-1809-2021-1-161-116-123>
- Lahmiri, S., Bekiros, S., & Avdoulas, C. (2023). A comparative assessment of machine learning methods for predicting housing prices using Bayesian optimization. *Decision Analytics Journal*, 6, 100166. <https://doi.org/10.1016/j.dajour.2023.100166>
- Lee, C., & Park, K. K.-H. (2022). Forecasting trading volume in local housing markets through a time-series model and a deep learning algorithm. *Engineering, Construction and Architectural Management*, 29(1), 16-30. <https://doi.org/10.1108/ECAM-10-2020-0850>
- Lenaers, I., & De Moor, L. (2023). Exploring XAI techniques for enhancing model transparency and interpretability in real estate rent prediction: A comparative study. *Finance Research Letters*, 58, 104306. <https://doi.org/10.1016/j.frl.2023.104306>
- Leo Breiman, J. F., R.A. Olshen, Charles J. Stone. (2017). *Classification and Regression Trees* (1st ed ed.). Chapman and Hall/CRC. <https://doi.org/10.1201/9781315139470>
- Li, H. (2024). *Machine Learning Methods* Springer. <https://link.springer.com/book/10.1007/978-981-99-3917-6>
- Li, R. Y. M. (2022). Housing Real Estate Economics and Finance. *Journal of Risk and Financial Management*, 15(3), 121. <https://www.mdpi.com/1911-8074/15/3/121>
- Lipton, Z. C. (2018). The Mythos of Model Interpretability: In machine learning, the concept of interpretability is both important and slippery. *Queue*, 16(3), 31-57. <https://doi.org/10.1145/3236386.3241340>

- Liu, B. (2011). Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data (Data-Centric Systems and Applications) (2nd ed ed.). Springer. <https://link.springer.com/book/10.1007/978-3-642-19460-3>
- Mansour, A., Bentley, R., Baker, E., Li, A., Martino, E., Clair, A., Daniel, L., Mishra, S. R., Howard, N. J., Phibbs, P., Jacobs, D. E., Beer, A., Blakely, T., & Howden-Chapman, P. (2022). Housing and health: an updated glossary. Journal of Epidemiology and Community Health, 76(9), 833. <https://doi.org/10.1136/jech-2022-219085>
- Meyberg, C., Rendtel, U., & Leerhoff, H. (2024). Flat rent price prediction in Berlin with web scraping. AStA Wirtschafts- und Sozialstatistisches Archiv, 18, 245–278. <https://doi.org/10.1007/s11943-024-00340-6>
- Michael W. Berry, A. M., Bee Wah Yap. (2020). Supervised and Unsupervised Learning for Data Science (Unsupervised and Semi-Supervised Learning) (1st ed ed.). Springer. <https://www.springerprofessional.de/en/supervised-and-unsupervised-learning-for-data-science/17138882>
- Mienye, D., & Sun, Y. (2022). A Survey of Ensemble Learning: Concepts, Algorithms, Applications, and Prospects. IEEE Access, PP, 1-1. <https://doi.org/10.1109/ACCESS.2022.3207287>
- Mora-Garcia, R. T., Cespedes-Lopez, M. F., & Perez-Sanchez, V. R. (2022). Housing price prediction using machine learning algorithms in COVID-19 times. Land, 11(11), 2100. <https://doi.org/10.3390/land11112100>
- Murtagh, F. (1991). Multilayer perceptrons for classification and regression. Neurocomputing, 2(5), 183-197. [https://doi.org/10.1016/0925-2312\(91\)90023-5](https://doi.org/10.1016/0925-2312(91)90023-5)
- Najib, T., Muntasir, F., & Wasi, W. A. W. (2023). Transparency in House Rent of Dhaka: Explainable AI Based Predictive Framework. <https://doi.org/10.46254/BA06.20230146>
- Ogundunmade, T., Abidoye, M., & Olunfunbi, O. (2023). Modelling Residential Housing Rent Price Using Machine Learning Models. Modern Economy and Management, 4, 14. <https://doi.org/10.53964/mem.2023014>
- Pilehvar, A. A., & Ghasemi, A. (2024). Advanced modeling of housing locations in the city of Tehran using machine learning and data mining techniques. Humanities and Social Sciences Communications, 11, 804. <https://doi.org/10.1057/s41599-024-03244-6>
- Potrawa, T., & Tetereva, A. (2022). How much is the view from the window worth? Machine learning-driven hedonic pricing model of the real estate market. Journal of Business Research, 144, 50–65. <https://doi.org/10.1016/j.jbusres.2022.01.027>
- Radhoush, S., Whitaker, B. M., & Nehrir, H. (2023). An Overview of Supervised Machine Learning Approaches for Applications in Active Distribution Networks. Energies, 16(16), 5972. <https://www.mdpi.com/1996-1073/16/16/5972>
- Rafiei, M. H., & Adeli, H. (2018). Novel Machine-Learning Model for Estimating Construction Costs Considering Economic Variables and Indexes. Journal of Construction Engineering and Management, 144. [https://doi.org/10.1061/\(ASCE\)CO.1943-7862.0001570](https://doi.org/10.1061/(ASCE)CO.1943-7862.0001570)
- Renju, K., & Freni, S. (2024). Ensemble Approach for Predicting The Price of Residential Property. International Journal of Information Technology, Research and Applications, 3, 27-38. <https://doi.org/10.59461/ijitra.v3i2.99>
- Rizun, N., & Baj-Rogowska, A. (2021). Can Web Search Queries Predict Prices Change on the Real Estate Market? IEEE Access, 9, 70095-70117. <https://doi.org/10.1109/ACCESS.2021.3077860>
- Rokach, L. (2023). Machine Learning for Data Science Handbook: Data Mining and Knowledge Discovery Handbook (3rd ed ed.). Springer. <https://link.springer.com/book/10.1007/978-3-031-24628-9>
- Ryszard S, M., Jaime G, C., & Tom M, M. (2013). Machine Learning: An Artificial Intelligence Approach (1st ed ed., Vol. 1). Morgan Kaufmann. <https://www.amazon.com/Machine-Learning-Artificial-Intelligence-Approach/dp/0934613095>
- Sagi, O., & Rokach, L. (2018). Ensemble learning: A survey. WIREs Data Mining and Knowledge Discovery, 8(4), e1249. <https://doi.org/10.1002/widm.1249>
- Sharma, H., Harsora, H., & Ogunleye, B. (2024). An Optimal House Price Prediction Algorithm: XGBoost. Analytics, 3(1), 30-45. <https://www.mdpi.com/2813-2203/3/1/3>

- Shi, Y. (2022). *Advances in Big Data Analytics: Theory, Algorithms and Practices* (1st ed ed.). Springer. <https://www.amazon.com/Advances-Big-Data-Analytics-Algorithms/dp/9811636060>
- Shuzlina, R., Nor, Z., & Sofianita, M. (2021). Advanced Machine Learning Algorithms for House Price Prediction: Case Study in Kuala Lumpur. *International Journal of Advanced Computer Science and Applications*, 12. <https://doi.org/10.14569/IJACSA.2021.0121291>
- Soltani, A., Heydari, M., Aghaei, F., & Pettit, C. (2022). Housing price prediction incorporating spatio-temporal dependency into machine learning algorithms. *Cities*, 131, 103941. <https://doi.org/10.1016/j.cities.2022.103941>
- Sujata Dash, S. K. P., Joel J. P. C. Rodrigues, Babita Majhi. (2022). *Deep Learning, Machine Learning and IoT in Biomedical and Health Informatics: Techniques and Applications (Biomedical Engineering)* (1st Ed ed.). CRC Press. <https://doi.org/10.1201/9780367548445>
- Tajmiri, S., Azimi, E., Hosseini, M. R., & Azimi, Y. (2020). Evolving multilayer perceptron, and factorial design for modelling and optimization of dye decomposition by bio-synthesized nano CdS-diatomite composite. *Environmental Research*, 182, 108997. <https://doi.org/10.1016/j.envres.2019.108997>
- Tang, Y., Qiu, S., & Gui, P. (2018). Predicting Housing Price Based on Ensemble Learning Algorithm. <https://doi.org/10.1109/IDAP.2018.8620781>
- Tekouabou, S. C. K., Gherghina, Ş. C., Kameni, E. D., Filali, Y., & Idrissi Gartoumi, K. (2024). AI-Based on Machine Learning Methods for Urban Real Estate Prediction: A Systematic Survey. *Archives of Computational Methods in Engineering*, 31(2), 1079-1095. <https://doi.org/10.1007/s11831-023-10010-5>
- Teremetskyi, V., Avramova, O., Svitlychnyy, O., Sloma, V., Bodnarchuk, O., Telestakova, A., & Kokhan, V. (2021). Housing Rights Protection in the Context of Legislation and Judicial Practice of Ukraine. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.4138075>
- Tharwat, A. (2021). Classification assessment methods. *Applied computing and informatics*, 17(1), 168-192.
- Trinka, O., & Kauermann, G. (2023). Can machine learning algorithms deliver superior models for rental guides? *AStA Wirtschafts- und Sozialstatistisches Archiv*, 17, 305-330. <https://doi.org/10.1007/s11943-023-00333-x>
- Truong, Q., Nguyen, M., Dang, H & Mei, B. (2020). Housing Price Prediction via Improved Machine Learning Techniques. *Procedia Computer Science*, 174, 433-442. <https://doi.org/10.1016/j.procs.2020.06.111>
- UN-Habitat. (2023). Executive Director's Thematic Paper: Achieving the Sustainable Development Goals in times of global crises. <https://unhabitat.org/executive-directors-thematic-paper-achieving-the-sustainable-development-goals-in-times-of-global>
- Wang, H., Yu, F., & Zhou, Y. (2018). Property Investment and Rental Rate Under Housing Price Uncertainty: A Real Options Approach (January 30, 2018). *Real Estate Economics*, Forthcoming, Claremont McKenna College Robert Day School of Economics and Finance Research Paper No. 3113143. <https://doi.org/10.1111/1540-6229.12235>
- Wang, K., Zhao, H., & Li, J. (2023). Machine Learning-Based House Rent Prediction Using Stacking Integration Method. *American Journal of Management Science and Engineering*, 8(2), 50-55. <https://doi.org/10.11648/j.ajmse.20230802.12>
- Wang, Y., & Zhao, Q. (2022). House Price Prediction Based on Machine Learning: A Case of King County. <https://doi.org/10.2991/aebmr.k.220307.253>
- Watson, T. J. (2001). An empirical study of the naive Bayes classifier. (22), 41-46. <https://doi.org/10.1039/b104835j>
- World Bank. (2022). *Poverty and shared prosperity 2022: Correcting course*. World Bank. <https://doi.org/10.1596/978-1-4648-1893-6>
- Xu, K. (2022). Predicting housing prices and analyzing real estate markets in the Chicago suburbs using machine learning *Journal of Student Research*, 11(3). <https://doi.org/10.47611/jsrhs.v11i3.3459>

- Xu, X., & Zhang, Y. (2023). A Gaussian process regression machine learning model for forecasting retail property prices with Bayesian optimizations and cross-validation. *Decision Analytics Journal*, 8, 100267. <https://doi.org/10.1016/j.dajour.2023.100267>
- Xu, Y., & Goodacre, R. (2018). On Splitting Training and Validation Set: A Comparative Study of Cross-Validation, Bootstrap and Systematic Sampling for Estimating the Generalization Performance of Supervised Learning. *Journal of Analysis and Testing*, 2(3), 249-262. <https://doi.org/10.1007/s41664-018-0068-2>
- Yoshida, T., Murakami, D., & Seya, H. (2024). Spatial prediction of apartment rent using regression-based and machine learning-based approaches with a large dataset. *The Journal of Real Estate Finance and Economics*, 69(1), 1-28. <https://doi.org/10.1007/s11146-022-09929-6>
- Zabor, E. C., Reddy, C. A., Tendulkar, R. D & ,Patil, S. (2022). Logistic Regression in Clinical Studies. *Int J Radiat Oncol Biol Phys*, 112(2), 271-277. <https://doi.org/10.1016/j.ijrobp.2021.08.007>
- Zarghamfard, M., & Meshkini, A. (2022). Analysis of factors affecting the realization of right to adequate housing in Iran: developing an interpretive-structural model. *International Journal of Housing Markets and Analysis*, 15(2), 411-428. <https://doi.org/10.1108/IJHMA-02-2021-0021>
- Zhang, L. (2023). Housing Price Prediction Using Machine Learning Algorithm *Journal of World Economy*, 2(3), 18-26. <https://www.pioneerpublisher.com/jwe/article/view/392>
- Zhou, J. (2023). The Private Sector in the Rental Housing Market. *BCP Business & Management*. <https://doi.org/10.54691/bcpbm.v43i.4631>