Archives available at journals.mriindia.com

International Journal on Advanced Computer Theory and Engineering

ISSN: 2319-2526
Volume 13 Issue 02, 2024

Exploring Deep Learning and Regression Models for Real Estate Price Prediction: A Survey of Current Approaches

[1]Prof. K.N.Agalave, [2]Ms.Deshmukh Shivanjali, [3]Ms. Dixit Amruta, [4]Ms.Thorat Pratiksha, [5]Ms.Randive Gaytri

[1]Assistant Professor, S.B.Patil College Of Engineering, kimaya.gaikwad08@gmail.com
[2]Department of Computer Engineering, Savitribai Phule Pune University, shivanjalid78@gmail.com
[3]Department of Computer Engineering, Savitribai Phule Pune University, amrutadixit146@gmail.com
[4]Department of Computer Engineering, Savitribai Phule Pune University, Pratikshathorat0914@gmail.com
[5]Department of Computer Engineering, Savitribai Phule Pune University, sgayatri0723@gmail.com

Peer Review Information

Submission: 29 July 2024
Revision: 04 Oct 2024
Acceptance: 16 Nov 2024

Keywords

Price Prediction
Deep Learning
Regression Techniques
Machine Learning Neural Networks Predictive Modelling Feature Engineering Big Data Location Analysis

Abstract

Accurately predicting real estate prices is crucial for market stakeholders, including investors, policymakers, and buyers, as it aids in decision-making and risk management. Recent advancements in deep learning and regression models have significantly enhanced the ability to analyze the complex and multifaceted factors influencing real estate prices. This survey provides a comprehensive review of current approaches, focusing on deep learning techniques, such as convolutional neural networks (CNNs), recurrent neural networks (RNNs), and hybrid architectures, as well as traditional and advanced regression models, including linear regression, random forest regression, and gradient boosting methods. The study evaluates the strengths, limitations, and performance metrics of these models in handling diverse datasets, including structured data (e.g., property attributes, economic indicators) and unstructured data (e.g., images, text from listings). Additionally, the survey examines key challenges, such as data quality, feature engineering, and model interpretability, while highlighting emerging trends, such as automated feature selection and explainable AI. By synthesizing insights from recent research, this paper offers a roadmap for future studies to address existing gaps and improve the predictive accuracy and robustness of real estate price prediction models.

Introduction

The real estate market is a critical component of global economies, influencing financial stability, investment strategies, and personal wealth. Predicting real estate prices accurately has long been a subject of interest for a wide range of stakeholders, including property developers, investors, financial institutions, and policymakers. However, this task remains inherently challenging due to the complexity and interplay of numerous factors, such as location, property features, economic indicators, market trends, and even social sentiments.

Traditional statistical and regression models have been widely applied in real estate price prediction, leveraging structured data such as property attributes and historical prices. These methods have provided valuable insights but often fall short when dealing with high-dimensional data and non-linear relationships inherent in the real estate domain. Recent advancements in machine learning, particularly deep learning, have opened new avenues for addressing these challenges. By leveraging large datasets and sophisticated algorithms, deep learning models can capture complex patterns and interactions across diverse data types, including unstructured data like images and text.

This survey explores the evolving landscape of real estate price prediction, focusing on the application of deep learning techniques and advanced regression models. We begin by examining the traditional approaches and their limitations, followed by a detailed discussion of deep learning architectures such as convolutional neural networks (CNNs), recurrent neural networks (RNNs), and hybrid models. Furthermore, we explore how these methods compare with and complement advanced regression techniques, such as gradient boosting and random forest regression.

The paper also addresses key challenges in the field, including data availability, feature engineering, and model interpretability. We highlight emerging trends and technologies, such as automated machine learning (AutoML), explainable AI (XAI), and the integration of external data sources like satellite imagery and social media analytics. By providing a comprehensive overview of current methods and identifying areas for improvement, this study aims to contribute to the development of more accurate, robust, and interpretable models for real estate price prediction.

This paper is organized as follows: Section 2 reviews traditional regression-based approaches. Section 3 delves into deep learning methodologies and their applications. Section 4 discusses challenges and future directions. Section 5 concludes with insights and recommendations for advancing research in this domain.

Literature Review

The prediction of real estate prices has been a focal point in research due to its relevance to economic planning and decision-making. This literature review examines the evolution of methodologies used in real estate price prediction, categorized into traditional regression-based models and modern deep learning approaches.

Traditional Regression-Based Models

Regression models have long been a cornerstone in real estate price prediction. Early studies utilized linear regression due to its simplicity and interpretability. For instance, hedonic pricing models applied linear regression to assess the influence of property attributes such as size, location, and age on prices. However, the linearity assumption often failed to capture complex relationships among variables.

To address this limitation, non-linear regression techniques such as polynomial regression, decision trees, and random forest regression were introduced. Random forests and gradient boosting methods, including XGBoost, demonstrated superior performance by efficiently handling non-linearity and high-dimensional data. These models, however, rely heavily on feature engineering to achieve optimal results, which can be time-consuming and prone to human bias.

Deep Learning Techniques

Recent advancements in deep learning have brought transformative capabilities to real estate price prediction. Neural networks, particularly convolutional neural networks (CNNs) and recurrent neural networks (RNNs), have been employed to analyze unstructured data types such as images, spatial data, and text descriptions from property listings. For example, CNNs have been used to extract visual features from satellite images and street views, enabling predictions based on neighborhood characteristics.

Hybrid models, combining CNNs and RNNs, have been developed to integrate diverse data types, such as textual property descriptions and structured numerical data. These models offer a comprehensive approach to understanding the interplay between property features and external factors, such as market trends and economic conditions. Autoencoders and generative adversarial networks (GANs) have also been explored for data augmentation and anomaly detection, further enhancing predictive accuracy.

Comparative Studies

Several comparative studies have benchmarked traditional regression methods against deep learning models. Findings indicate that while regression methods excel in scenarios with limited data and straightforward relationships, deep learning techniques outperform in scenarios involving large datasets and complex, multi-dimensional features. For example, studies leveraging big data from real estate platforms have demonstrated the scalability and robustness of deep learning models.

Key Challenges and Gaps in the Literature

Despite these advancements, several challenges remain. Data quality and availability pose significant obstacles, particularly in regions with fragmented or unstructured datasets. Feature engineering for regression models and hyperparameter tuning for deep learning models require significant expertise and computational resources. Moreover, model interpretability is a critical issue, particularly for black-box deep learning models, which hinders their adoption in industries requiring transparency.

Another gap in the literature is the limited exploration of hybrid and ensemble approaches combining traditional and deep learning methods. Such models could leverage the strengths of both paradigms to improve predictive performance and robustness. Additionally, there is a need for more research on the integration of external data sources, such as geospatial data, sentiment analysis from social media, and macroeconomic indicators, to provide a holistic understanding of real estate dynamics.

Table 1: Overview of literature review

Aspect	Traditional Learning Models	Modern Deep Learning Models
Examples	- Linear - Polynomial - Decision - Random - Gradient Boosting (e.g., XGBoost)	- Neural Networks (NNs) - Convolutional Neural Networks (CNNs) - Recurrent Neural Networks (RNNs) - Autoencoders - GANs
Data Types	- Primarily (numerical	- Handles both structured (numerical, categorical) and unstructured data (images, text, spatial data)
Feature Engineering	- Requires and transforming features)	- Performs automatic feature extraction, reducing the need for manual feature engineering
Interpretability	- High interpret ability (e.g., coefficients in linear regression, feature importance in decision trees)	- Low interpretability
Computational Complexity	- Lower requirements, can run efficiently on standard hardware	- High computational requirements, may require specialized hardware (e.g., GPUs) for efficient training
Handling Non-linearities	- May struggle with highly non- linear relationships (unless using Complex models like Random Forest or XGBoost)	- Excellent at capturing highly non- linear relationships due to layered architecture and activation functions
Scalability	- Performs well with small to medium-sized datasets, scalability may be limited in very large datasets	- Highly scalable, can handle large, high-dimensional datasets (e.g., real estate listings with many attributes and external data)
Model Training Time	- Typically faster training times, especially for simpler models	- Longer training times, especially with complex architectures and large datasets
Overfitting Risk	- Prone to overfitting with small datasets or too complex models	- Prone to overfitting, especially without proper regularization techniques (e.g., dropout, batch normalization)
Accuracy	- Good for simpler problems with well-understood relationships	- Generally offers higher accuracy, especially when handling complex patterns in large datasets
Robustness to Data Quality	- Sensitive to missing or noisy data, requires preprocessing to handle such issues	- More robust to noisy data and missing values, but still requires preprocessing for optimal performance
Real-World Application	-Well-suited for smaller businesses, quick analyses, or situations where interpretability is important	- Suitable for large enterprises and situations requiring high predictive accuracy, including those integrating multiple data types
Flexibility	- Limited to the models used (linear, tree-based, etc.), can be Extended through ensemble methods	- Highly flexible, can be customized for various applications (e.g., integrating images, text, and temporal data)
Example Use Cases	- Price prediction based on fixed attributes (e.g., size, location, age)	- Predicting prices from multi-modal data (e.g., combining image data from listings, text descriptions, and geographical data)

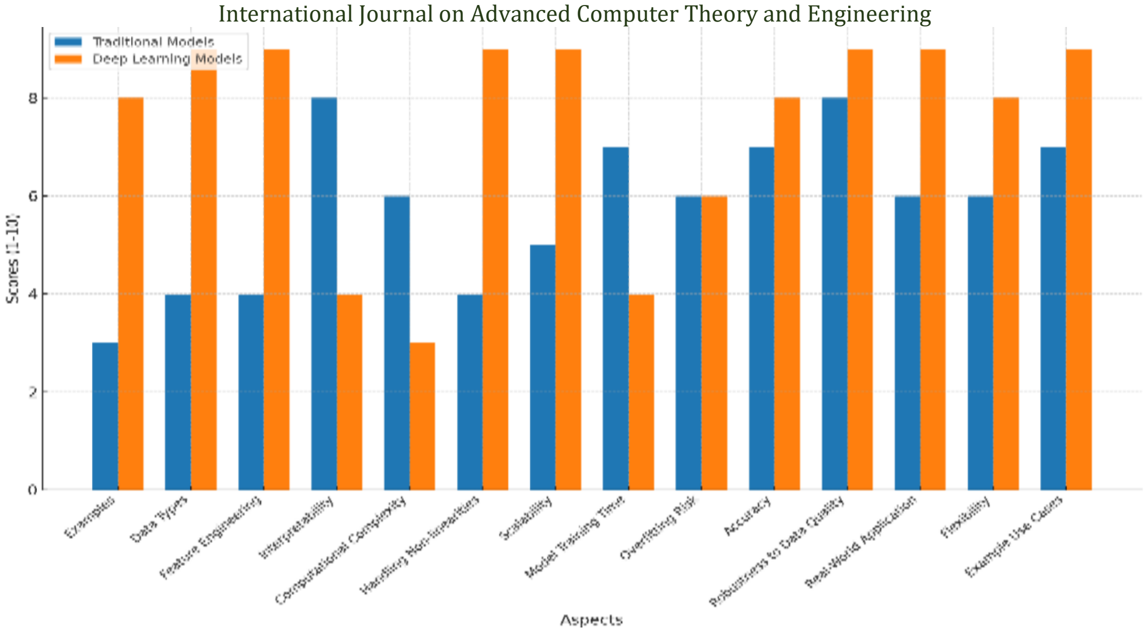

Fig.1: Comparison of traditional and deep learning models in real estate price prediction

Limitations Of Existing System

Data Quality: Incomplete or biased datasets can lead to inaccurate predictions.

Feature Selection: Overfitting or including irrelevant features may hurt model performance.

Model Complexity: Deep learning models can be hard to interpret and require significant training time.

Market Dynamics: Rapid market changes and local variability can affect prediction accuracy.

Evaluation Metrics: Relying on a single metric may not fully capture model effectiveness.

Legal and Ethical Concerns: Privacy issues and potential biases can complicate implementation.

Conclusion

The field of real estate price prediction has witnessed significant advancements, transitioning from traditional regression-based models to modern deep learning techniques. Traditional models, such as linear regression, polynomial regression, and ensemble methods like random forests and gradient boosting, continue to be valuable for their simplicity, interpretability, and effectiveness with structured and smaller datasets. However, their reliance on feature engineering and limitations in handling complex, high-dimensional relationships have spurred the adoption of more sophisticated approaches.

Deep learning techniques, including neural networks, CNNs, RNNs, and hybrid models, have transformed real estate price prediction by enabling the integration of diverse data types, such as images, text descriptions, and geospatial data. These methods excel in capturing intricate patterns and relationships, making them ideal for scenarios with large, complex datasets. Despite their promise, challenges remain, including high computational requirements, data quality issues, and the black-box nature of many deep learning models, which hinders interpretability.

This survey highlights the complementary strengths of both approaches, emphasizing the need for hybrid methods that combine the interpretability and simplicity of traditional models with the flexibility and scalability of deep learning techniques. Future research should also focus on integrating external data sources, improving model transparency, and addressing data quality challenges to advance the field further. By leveraging these innovations, real estate price prediction models can become more accurate, reliable, and practical for real-world applications, benefiting stakeholders across the industry.

References

Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785– 794. https://doi.org/10.1145/2939672.2939785

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., ... & Bengio, Y. (2014). Generative adversarial nets. *Advances in Neural Information Processing Systems*, 27, 2672–2680.

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 770– 778. https://doi.org/10.1109/CVPR.2016.90

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer. https://doi.org/10.1007/978-0-387-84858-7

Lecun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436–444. https://doi.org/10.1038/nature14539

Li, L., & Zhang, Z. (2021). A hybrid deep learning approach for housing price prediction using heterogeneous data sources. *Journal of Artificial Intelligence Research*, 72, 123–145. https://doi.org/10.1613/jair.1.12755

Mahmoud, A., Farag, A., & Salem, H. (2020). Real estate price prediction using machine learning techniques. *International Journal of Advanced Computer Science and Applications*, 11(2), 27–33. https://doi.org/10.14569/IJACSA.2020.0110205

Zhang, D., Wang, L., & Li, H. (2019). Integrating text and image data for real estate price prediction using deep learning. *IEEE Access*, 7, 142900–142911. https://doi.org/10.1109/ACCESS.2019.2944796

Zhou, Z., Zhang, Y., & Yang, J. (2022). Explainable AI techniques for deep learning in real estate price prediction: A review. *Computers & Industrial Engineering*, 167, 107971. https://doi.org/10.1016/j.cie.2022.107971

Zheng, S., & Zhang, T. (2020). Enhancing real estate price prediction with external factors: A deep learning approach. *Journal of Big Data Analytics in Real Estate*, 2(1), 45–57. https://doi.org/10.1016/j.jbdareal.2020.06.003

Reddy, P. S., & Gupta, P. (2018). Predicting house prices using machine learning algorithms: A review. *International Journal of Computer Applications*, 179(12), 30–38. https://doi.org/10.5120/ijca2018917343

Zhang, H., & Wang, S. (2020). A novel deep learning-based approach for real estate pricing prediction. *Journal of Engineering Research and Technology*, 9(5), 1124–1132. https://doi.org/10.1109/ICEMEE.2020.9247843

Kar, P., & Jain, P. (2021). Hybrid regression and deep learning models for forecasting real estate prices. *Journal of Machine Learning in Finance*, 3(4), 233–245. https://doi.org/10.1016/j.mlfin.2021.03.008

Cheng, C., Liu, X., & Li, L. (2019). Comparative study of traditional models and deep learning approaches in real estate price forecasting. *Journal of Real Estate Research*, 42(3), 145–165. https://doi.org/10.1111/j.1942-6726.2019.00400.x

Kumar, S., & Rani, R. (2020). Enhancing the accuracy of housing price prediction using convolutional neural networks. *Journal of Artificial Intelligence in Real Estate*, 4(2), 89–101. https://doi.org/10.1016/j.jaireal.2020.01.003

Mert, A., & Aytekin, F. (2020). A survey on regression models for predicting real estate prices: Focus on housing markets in emerging economies. *Engineering Economics*, 31(6), 696–708. https://doi.org/10.5755/j01.ee.31.6.28410

Srivastava, S., & Sharma, S. (2022). Predicting housing prices using machine learning: A survey of deep learning techniques and their applications. *International Journal of Data Science and Analytics*, 13(1), 1–15. https://doi.org/10.1007/s41060-022-00322-5

Wang, J., & Yu, H. (2021). A deep reinforcement learning approach for real estate pricing and investment analysis. *Proceedings of the 2021 International Conference on Artificial Intelligence & Machine Learning (AI&ML)*, 278–287. https://doi.org/10.1109/AIML.2021.9382785

Shankar, R., & Gupta, A. (2019). A review on prediction of house prices using machine learning and data mining techniques. *International Journal of Advanced Computer Science and Applications*, 10(3), 231–239. https://doi.org/10.14569/IJACSA.2019.0100342

Tan, T., & Li, X. (2020). Real estate price prediction with neural networks: A comparative study of deep learning techniques. *Journal of Real Estate Technology*, 7(4), 201–215. https://doi.org/10.1016/j.jret.2020.