










A hybrid machine learning framework for forecasting house price

Choujun Zhan ^{a b} , Yonglin Liu ^b , Zeqiong Wu ^b , Mingbo Zhao ^{b c}  , Tommy W.S. Chow ^c 

[Show more](#) 

 Share  Cite

<https://doi.org/10.1016/j.eswa.2023.120981> 

[Get rights and content](#) 

Highlights

- A new house prices forecasting models are developed.
- Three types of hybrid models are established to improve generalization ability of performance.
- A hybrid Bayesian approach is developed to optimize the hyperparameter of the above models.
- Extensive simulations are used to evaluate the performance of the proposed work.
- Simulation results verify the effectiveness of the proposed work.

Abstract

House price prediction is one of the most important factors affecting national real estate policies. However, developing an accurate housing price prediction model is a significant challenge for the real estate market. This study presents a framework of house price prediction models that address this issue by improving forecasting performance, explicitly demonstrating the novelty in the Hybrid Bayesian Optimization (HBO) models combined with Stacking (HBOS), Bagging (HBOB), and Transformer (HBOT) techniques. These hybrid models employ Bayesian Optimization for hyperparameter tuning, leading to superior prediction accuracy and stability. Additionally, the

proposed framework can assess a statistical and accurate assessment of the predictive performance of house price forecasting models in different scenarios. Furthermore, we constructed a multi-source dataset containing 1,898,175 transactions of the Hong Kong real estate market covering a period from January 2, 1996, to May 13, 2021. This dataset, another major contribution to the field, enables comprehensive model testing and could be a valuable resource for future research. Then, the proposed hybrid models are compared with 18 benchmark models. Thirteen evaluation metrics are used to evaluate the predictive performance, while the non-parametric testing, including Friedman, Iman–Davenport, and Nemenyi post-hoc tests methods, are adopted to assess the significance of differences in the predictive performance of each model. The experimental results show that the HBOS models are superior to the other benchmark models for application in the house price prediction problem. The HBOS-CatBoost model showed superior performance in terms of RMSE compared to both the HBOB-XGBoost and HBOT-ConvLSTM models, with relative RMSE reductions of 5.11% and 25.56%, respectively. The main contributions of this work are the creation of a rich multi-source dataset, the proposal of novel hybrid models for improved house price prediction, and a comprehensive performance evaluation framework. These findings offer a significant step forward in the housing price prediction field.

Introduction

For many individuals and families, residential housing is one of the most important living resources and properties. While for cities, houses not only provide living space for people in society but also serve as an attractive market for investors, especially in international metropolis (Chen, Ong, Zheng, & Hsu, 2017). The development of the real estate market can highly influence the economic activities of a country or even the entire world. For instance, the US subprime mortgage crisis in 2007 triggered a severe global financial crisis, which caused huge asset losses to financial markets and organizations. This crisis brought about the shrinkage of bank credit and the depreciation of the real estate market, resulting in slowing down economic development all over the world (Hodson & Quaglia, 2009). Hence, the healthy development of a real estate market is one of the most important factors strongly associated with the sustainable development of the economy. The real estate market is not just an important issue for individuals and businesses but also for socioeconomic stability. Therefore, house price prediction has attracted significant attention from various fields, including economy, politics, computer science and etc. (Iacoviello and Minetti, 2008, Park and Bae, 2015, Selim, 2009, Tsai, 2013). However, the main challenge is that the real estate prices prediction is a nonlinear time series forecasting issue for complex systems, which is influenced by various factors, including economic factors (such as income ratio, housing vacancy rate, income inequality and etc.), house attributes (such as house age, the number of transactions, living area, parking spaces, number of rooms and geographic location) and crisis events (such as pandemic and economic crisis). Part of the studies focuses on determining factors related to house prices. Researchers investigate the relationship between house prices and various factors, including monetary policy (Tsai, 2013), labor market (Fischer, Huber, Pfarrhofer, & Stauffer-Steinnocher, 2021), investment and consumption (Benetton, Bracke, Cocco, & Garbarino, 2022), and etc. In this work, we aim to develop a state-of-art framework for house price prediction and conduct an extensive experimental comparison for assessing a statistical and accurate assessment of the predictive performance of house price forecasting models on different datasets.

Accurately predicting house prices can greatly assist real estate market analyses and place-based decisions. Hence, researchers have developed various housing price prediction models, roughly classified into three categories: (1) Classical linear regression model. The Auto Regressive Integrated Moving Average (ARIMA) is a time series prediction statistical modeling model that is frequently used as a baseline model for house price forecasting tasks. Research has demonstrated that in terms of forecasting, deep learning and machine learning models generally outperform ARIMA models (Abidoeye et al., 2019, Temur et al., 2019); (2) Hedonic price models. The hedonic price model can be used to determine the relationship between attributes and house prices (Goodman, 1978). Then, the hedonic pricing model has been applied to predict house prices (Harding et al., 2003, Maurer et al., 2004); (3) Machine learning models. Due to the inability of hedonic price models to capture non-linear relationships between features (Limsombunchai, 2004), machine learning (ML) methods are employed to discover potential non-linear relationships in time series. Researchers employ machine learning techniques to forecast property prices, which always show a better predictive performance (Abidoeye and Chan, 2018, Rahman et al., 2019). These techniques include artificial neural networks (ANN) (Abidoeye et al., 2019), ensemble learning models (Kang et al., 2020, Park and Bae, 2015), and deep learning models (Ge, 2019, Law et al., 2019).

Studies have discovered that hybrid models can combine the strengths of different models to obtain higher prediction accuracy (Polley & Van Der Laan, 2010). Different ensemble approaches, including Bagging, Boosting and Stacking, have been applied in the research. Bagging and Boosting focus on reducing variance and bias, respectively, while Stacking achieves the two goals of reducing variance and bias by finding the optimal way to combine algorithms (Large, Lines, & Bagnall, 2019). In addition, hyperparameters are essential for machine learning models, as they have a significant impact on the effectiveness of machine learning models and directly influence predictive performance. However, it is difficult to optimize hyperparameters using traditional methods such as grid search or random search (Bischl et al., 2021). Therefore, designing an effective strategy to determine the optimal hyperparameters of the predictive model is a crucial step in enhancing the predictive power. During the past decade, various techniques for hyperparameter optimization have been proposed, including Bayesian optimization algorithms (Wang, Wang, & Peng, 2021), genetic algorithms (Zhou, Huang, Wang, & Qiu, 2021), differential evolutionary algorithms (Wang, Zeng, & Chen, 2015) and particle swarm optimization algorithms (Zhang, Zhang, & Zhang, 2020). Therefore, our work will aim to employ hyperparameter optimization techniques and ensemble approaches to develop hybrid house price prediction models with high predictive performance.

Recently, various methods have been proposed for dealing with house price prediction problems. However, it is well known that the performance of a predictive model is heavily dependent on the available datasets. For evaluating the predictive performance of a model, researchers commonly adopt one or several evaluation metrics. However, one method may have a different evaluation as we adopt different evaluation metrics. The adoption of evaluation metrics has an impact on the evaluation of model performance. Improper or insufficient usage of evaluation metrics may lead to a biased assessment of the overall performance of a model. Hence, it is important to conduct statistical tests on the result (Demšar, 2006). Our work will also focus on deriving a statistical and accurate assessment of the predictive performance of house price forecasting models in various scenarios.

In conclusion, this research proposes a novel framework for predicting housing prices, consisting of three different hybrid models based on the stacking, bagging, and transformer approaches. Meanwhile, the hyperparameters of the models are globally optimized by the bayesian optimization algorithm to improve the forecasting accuracy of the models. Finally, we construct a framework for comparing the predictive performance of the 27 house price forecasting models through the Friedman, Iman–Davenport and Nemenyi post-hoc tests methods.

Our contributions are three-fold:

- We constructed a multi-source dataset containing 1,898,175 housing transactions across three regions of Hong Kong (Kowloon, Hong Kong Island, and the New Territories), which includes 17 housing properties and 10 economic factors and covers the period from 1996 to 2021. This dataset provides a valuable resource for researchers and practitioners interested in studying the property market in Hong Kong, which represents a significant contribution to the field. Furthermore, our work identifies the relationship between house prices and various influencing factors. For instance, we noted that monetary policies heavily influence house prices in affluent districts, while their impact is less pronounced in less wealthy areas.
- We found that the constraints of data (short time span, limited vector length, and high similarity among price trends across regions) often lead to reduced prediction accuracy. In response to the unique challenges posed by the limited length and high similarity of house price vectors, we innovatively leveraged a framework of hybrid models for house price prediction, namely, we proposed three types of hybrid models—Hybrid Bayesian Optimization with Stacking-based (HBOS), Bagging-based (HBOB), and Transformer-based (HBOT) models. The framework of house price prediction models integrates diverse machine learning methods and utilizes a Bayesian Optimization algorithm for hyperparameter optimization, thus addressing the data limitations and enhancing prediction accuracy and stability. We propose these hybrid models to provide more robust and reliable methods for analyzing and predicting house prices in real estate markets. This framework contributes a novel perspective for improving house price predictions, particularly in contexts with similar data constraints.
- We proposed a comparative framework for assessing a statistical and accurate assessment of the predictive performance of house price forecasting models on different datasets. This framework contains three non-parametric tests, i.e., Friedman, Iman–Davenport and Nemenyi post-hoc tests, for evaluating and ranking the predictive performance of 27 house price forecasting models on three Hong Kong house price datasets by 13 evaluation metrics. Our analysis demonstrated the effectiveness of the proposed hybrid models in predicting house prices, thereby demonstrating the practical implications of our work.

This research contributes to the field by elucidating the relationships between housing prices and various factors, providing new hybrid models to overcome data constraints, and offering a robust comparative framework for evaluating predictive performance. These advancements provide

valuable insights and tools for scholars and practitioners aiming to better understand and predict housing market trends.

The remainder of this paper is organized as follows. Section 2 reviews relevant research about house price forecasting. The data description and analysis are presented in Section 3. Section 4 introduces the details of our proposed house price forecasting framework. In Section 5, the evaluation measures, statistical evaluation and detailed experimental results are described, while Section 6 presents the conclusion of the entire study and the direction of future research.

Section snippets

Literature review

In this section, we discuss the existing work on the housing price prediction problem. Here, we mainly focus on works about hedonic price models and machine learning models.

Classical hedonic price models: Early house price prediction models mainly focused on hedonic price regression, which identified price determinants based on internal attributes of the good being sold as well as external factors influencing it (Rosen, 1974). For instance, Yang proved that the hedonic price models could be ...

Data description and analysis

Hong Kong is a metropolis located in the eastern Pearl River Delta in South China, which has one of the most active real estate markets in the world. In this section, we will introduce the data about the Hong Kong real estate in detail, which includes the data collection, processing and analysis. ...

Methodology

In this section, we first define the problem description for house price forecasting. Then, we propose a framework for house price prediction based on machine learning and Bayesian optimization. In this framework, we investigate the performance enhancement of three different type of hybrid models (HBOS, HBOB and HBOT) for house price forecasting, with a total of 9 hybrid models. In addition, the framework applies 18 benchmark models for comparison. 13 evaluation measures are adopted to evaluate ...

Experimental results

In this section, we present the results of the experimental evaluation. First, we compare the predictive performance of the various models across the three regional datasets. Then, the statistical differences between the models are analyzed and shown in critical diagrams. Finally, we conclude the assessment.

Table 4, Table 5, Table 6, Table 7, Table 8, Table 9 show the performance evaluation results, rankings and average ranks of the 27 models considering 13 evaluation metrics in the three ...

Conclusion

There is evidence that the issue of predicting housing prices plays a crucial role in the field of economics. Hence, a large number of studies on the application of machine learning to construct house price prediction models, but the current study still has certain limitations. First, different machine learning models have different strengths and weaknesses, and the fact that a model performs well on one metric or in one dataset does not imply that it performs well in all aspects. Second, ...

CRedit authorship contribution statement

Choujun Zhan: Conceptualization, Methodology, Writing – original draft, Funding acquisition. **Yonglin Liu:** Software, Visualization. **Zeqiong Wu:** Data curation, Resources. **Mingbo Zhao:** Validation, Investigation, Funding acquisition, Writing – review & editing. **Tommy W.S. Chow:** Formal analysis, Visualization. ...

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Mingbo Zhao reports was provided by City University of Hong Kong. Mingbo Zhao reports a relationship with City University of Hong Kong that includes: employment. We certify that there is no actual or potential conflict of interest in relation to this article. ...

Acknowledgments

This work was supported by, the Natural Science Foundation of Guangdong Province, China (2023A1515011618), Key Scientific Research Platform and Project of Guangdong Provincial Education Department (2022ZDZX1040, 2022ZDZX1039) and Educational Science Planning Topic of Guangdong province (2022GXJK382). This work was also partially supported by the National Natural Science Foundation of China (61971121, 62073272). ...

[Recommended articles](#)

References (40)

Adetunji A.B. *et al.*

[House price prediction using random forest machine learning technique](#)

Procedia Computer Science (2022)

Goodman A.C.

[Hedonic prices, price indices and housing markets](#)

Journal of urban economics (1978)

IacovielloM. *et al.*

The credit channel of monetary policy: Evidence from the housing market

Journal of Macroeconomics (2008)

ParkB. *et al.*

Using machine learning algorithms for housing price prediction: The case of fairfax county, virginia housing data

Expert Systems with Applications (2015)

SelimH.

Determinants of house prices in turkey: hedonic regression versus artificial neural network

Expert systems with Applications (2009)

TsaiI.-C.

The asymmetric impacts of monetary policy on housing prices: A viewpoint of housing price rigidity

Economic Modelling (2013)

WangY. *et al.*

Rice diseases detection and classification using attention based neural network and bayesian optimization

Expert Systems with Applications (2021)

WangL. *et al.*

Back propagation neural network with adaptive differential evolution algorithm for time series forecasting

Expert Systems with Applications (2015)

YangL. *et al.*

Place-varying impacts of urban rail transit on property prices in shenzhen, china: Insights for value capture

Sustainable Cities and Society (2020)

AbidoyeR.B. *et al.*

Improving property valuation accuracy: A comparison of hedonic pricing model and artificial neural network

Pacific Rim Property Research Journal (2018)



View more references

Cited by (29)

A multi-scale analysis method with multi-feature selection for house prices forecasting

2025, Applied Soft Computing

[Show abstract](#) ✓

GA-FCFNN: A new forecasting method combining feature selection methods and feedforward neural networks using genetic algorithms

2024, Information Sciences

Citation Excerpt :

...For instance, Huang et al. [3] successfully applied a forecasting model to predict interest rates, while Liu et al. [4] introduced an interpretable forecasting model to anticipate patient readmissions for liver disease. Zhan et al. [5] devised a novel forecasting framework and utilized it to predict housing prices. Ma et al. [6] innovated a new classification method to construct a forecasting model for anticipating stock trend changes....

[Show abstract](#) ✓

Real estate price estimation through a fuzzy partition-driven genetic algorithm

2024, Information Sciences

Citation Excerpt :

...Some approaches are based on multiple regression analysis to evaluate the value of a property by subjective features [40,49]. Machine Learning approaches are widely employed in real estate to predict property prices [28,43,45,47]. In [7] a single imputation Chained Paasche method for building real estate price indices is proposed....

[Show abstract](#) ✓

Optimal decisions for selling on an online group buying platform in a competitive fuzzy game environment

2024, Expert Systems with Applications

Citation Excerpt :

...Price prediction aims to forecast future price trends based on historical values. Zhao et al. (2023) have reviewed the existing price prediction methods into statistical analysis methods (Kuranga & Pillay, 2022; Wang et al., 2022b), machine learning methods (Chaudhari & Thakkar, 2023; Hossain et al., 2022; Zhan et al., 2023) and deep learning methods (Gülmez, 2023; Kaya et al., 2023; Xu & Tan, 2021; Yu et al., 2023). Furthermore, Triveni et al. (2022) have suggested a game theory model for price forecasting based on several players of various conflicts of interest and mutually interdependent situations....

[Show abstract](#) ✓

IDENTIFYING THE CURRENT STATUS OF REAL ESTATE APPRAISAL METHODS ↗

2024, Real Estate Management and Valuation

AI-Based on Machine Learning Methods for Urban Real Estate Prediction: A Systematic Survey ↗

2024, Archives of Computational Methods in Engineering



[View all citing articles on Scopus](#) ↗

[View full text](#)

© 2023 Elsevier Ltd. All rights reserved.



All content on this site: Copyright © 2025 Elsevier B.V., its licensors, and contributors. All rights are reserved, including those for text and data mining, AI training, and similar technologies. For all open access content, the relevant licensing terms apply.

