|  |  |  |
| --- | --- | --- |
|  | **Cite** | **Abstract** |
| 1 | @article{Zong2023HousePP,  title={House Prices Prediction Advanced Regression Techniques},  author={Yue Zong},  journal={Advances in Economics, Management and Political Sciences},  year={2023},  url={https://api.semanticscholar.org/CorpusID:265629120}  } | **[House Prices Prediction Advanced Regression Techniques](https://www.semanticscholar.org/paper/House-Prices-Prediction-Advanced-Regression-Zong/30ee417f35fae286e354b8696a870dca54aee886)**  A neural network model was built, parameters were adjusted, and the trained network was used to predict housing prices, positioning the model among the top performers among all machine learning methods.  Abstract  In recent years, with the development of the real estate industry, housing prices have continued to rise. The nation, society, and individuals are all concerned about these prices. For commodity housing prices, there are many factors that influence the housing prices. Apart from national regulations, factors such as lighting, layout, and environment of the houses themselves also have a certain impact on the prices, leading to significant flvuctuations in the real estate market. Therefore, researching an accurate model for predicting housing prices has practical significance. It can guide residents in housing consumption and provide policy recommendations for government price regulation. Machine learning methods have become a new type of prediction method in this regard. Based on the theories of data analysis and machine learning, a dataset consisting of 2920 data points with 81 attributes was selected from the publicly available Kaggle housing dataset. The data was normalized and analyzed for feature selection. The ranking of attributes most correlated with housing prices was obtained. Subsequently, a neural network model was built, parameters were adjusted, and the trained network was used to predict housing prices. On the Kaggle leaderboard, the RMSE test result stands at 0.1198, positioning our model among the top performers among all machine learning methods. |
| 2 | @article{KHARB2023ForestallingHP,  title={Forestalling House Prices Using Linear Regression in Machine Learning},  author={LATIKA KHARB and Deepak Chahal},  journal={International Research Journal of Modernization in Engineering Technology and Science},  year={2023},  url={https://api.semanticscholar.org/CorpusID:260941791}  } | **[Forestalling House Prices Using Linear Regression in Machine Learning](https://www.semanticscholar.org/paper/Forestalling-House-Prices-Using-Linear-Regression-KHARB-Chahal/087e2c4888cf8abeb34c745891220c0cbe793278)**  This paper focuses on using linear regression to predict house prices, discussing the dataset, preprocessing steps, model training, and evaluation and highlights the strengths, limitations, and potential improvements of the linear regression approach in the context of house price prediction.  Abstract  Predicting house prices is a fundamental problem in the field of real estate and economics. The accurate estimation of house prices is crucial for various stakeholders including homeowners, buyers, sellers, and policymakers. In this paper, we explore the application of linear regression, a simple yet effective machine learning technique, to predict house prices. We discuss the methodology, data preprocessing, feature selection, model training, and evaluation metrics. We also highlight the strengths, limitations, and potential improvements of the linear regression approach in the context of house price prediction. In this paper, we focus on using linear regression to predict house prices, discussing the dataset, preprocessing steps, model training, and evaluation. |
|  | @article{Nwankwo2023PredictionOH,  title={Prediction of House Prices in Lagos-Nigeria Using Machine Learning Models},  author={Mmesoma Peace Nwankwo and Ndukaku Macdonald Onyeizu and Emmanuel Chibuogu Asogwa and Chukwuogo Okwuchukwu Ejike and Okechukwu J. Obulezi},  journal={European Journal of Theoretical and Applied Sciences},  year={2023},  url={https://api.semanticscholar.org/CorpusID:262169298}  } | **[Prediction of House Prices in Lagos-Nigeria Using Machine Learning Models](https://www.semanticscholar.org/paper/Prediction-of-House-Prices-in-Lagos-Nigeria-Using-Nwankwo-Onyeizu/a5479e6447253f914c85e0d7a2e986699903eecd)**  The model shows that the number of bedrooms, parking spaces, and house types play an important role in determining the price of houses.  Abstract  This paper considers the relationship between the price of houses and the features namely the number of bedrooms, parking space, and different house types. In this study, a machine learning approach was used to develop prediction models that predicted house prices in Lagos. Different machine learning techniques were used, train-test split to split the data into training sets for training and building the model and test data to test the accuracy of the model, performance metric mean absolute error to set the baseline for the model, Variance Inflation Factor (VIF) to help remove multicollinearity between features and Streamlit interactive dashboards to communicate with the model. Correlation and regression methods were used to examine the relationship and build the model. It is observed that there is a strong positive correlation between the number of bedrooms and the number of toilets, likewise the number of bedrooms and the number of bathrooms. It also shows that there is a moderate positive correlation between the number of bedrooms and price. The model shows that the number of bedrooms, parking spaces, and house types play an important role in determining the price of houses. |
| 4 | @article{Sundari2023OptimizationHP,  title={Optimization house price prediction model using gradient boosted regression trees (GBRT) and xgboost algorithm},  author={Putri Susi Sundari and Mahardika Khafidz Putra},  journal={Journal of Student Research Exploration},  year={2023},  url={https://api.semanticscholar.org/CorpusID:263632752}  } | **[Optimization house price prediction model using gradient boosted regression trees (GBRT) and xgboost algorithm](https://www.semanticscholar.org/paper/Optimization-house-price-prediction-model-using-and-Sundari-Putra/f8fecd11695987c1e0f29361763d65871a1272eb)**  The researcher utilizes ensemble learning techniques, combining the Gradient Boosted Regression Trees (GBRT) and XGBoost algorithms, to develop a predictive model that can accurately estimate house prices based on relevant features or variables.  Abstract  In this rapidly advancing technological era, the demand for the real estate industry has also increased, including in the field of house price prediction. House prices fluctuate every year due to several factors such as changes in land prices, location, year of construction, infrastructure developments, and other factors. Numerous studies have been conducted on this issue. However, the challenge lies in building a proven accurate and effective model for predicting house prices with the abundance of features present in the dataset. The objective of this research is to develop a predictive model that can accurately estimate house prices based on relevant features or variables. The researcher utilizes ensemble learning techniques, combining the Gradient Boosted Regression Trees (GBRT) and XGBoost algorithms. The dataset used in this article is titled "Ames Housing dataset" obtained from Kaggle. The predictive model is then evaluated using the Root Mean Squared Error (RMSE) method. The RMSE result from a previous study that used the combination of Lasso and XGBoost was 0.11260, while the RMSE result from this research is 0.00480. This indicates a decrease in the RMSE value, indicating a lower level of error in the model. It also means that the combination of GBRT and XGBoost algorithms successfully improves the prediction accuracy of the previous research model. |
| 5 | @article{Jayadharshini2023MachineLT,  title={Machine learning techniques for predicting home rental prices in India},  author={P Jayadharshini and S. Santhiya and S Keerthika and N Abinaya and S Priyanka},  journal={Applied and Computational Engineering},  year={2023},  url={https://api.semanticscholar.org/CorpusID:266556584}  } | **[Machine learning techniques for predicting home rental prices in India](https://www.semanticscholar.org/paper/Machine-learning-techniques-for-predicting-home-in-Jayadharshini-Santhiya/ec7b52328bae99bbb22d4e8a7dd5b9a8c52e95ad)**  Predicting the selling price of houses has become increasingly crucial as land and housing prices rise annually. This task is particularly challenging for metropolitan areas like Chennai and Bangalore. Therefore, there is a growing demand for an easier and more effective approach to forecast house rental prices, ensuring future generations have access to reliable predictions. Several key factors, such as the house's location and area, significantly influence rental prices. In this paper, a dataset comprising ten similar crucial features is utilized. The model is developed using a Python library, where the data is preprocessed and prepared to ensure cleanliness for constructing the model. Various machine learning algorithms, including Random Forest, Linear Regression, Decision Tree Regression, and Gradient Boosting, are employed. Through feature extraction, it is determined that area and property type are the most important features that significantly impact rental prices. Among the techniques used, gradient boosting yields the most satisfactory predictive results for rent based on evaluation metrics like Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and R-Squared Metric (R2). |
| 6 | @article{Yamur2023HousePP,  title={House price prediction modeling using machine learning techniques: a comparative study},  author={Ayten Yağmur and Mehmet Kayakuş and Mustafa Ender Terzioğlu},  journal={Aestimum},  year={2023},  url={https://api.semanticscholar.org/CorpusID:257470072}  } | **[House price prediction modeling using machine learning techniques: a comparative study](https://www.semanticscholar.org/paper/House-price-prediction-modeling-using-machine-a-Ya%C4%9Fmur-Kayaku%C5%9F/ed42fa40fbc0a004411002765140cc1b55117d56)**  The artificial neural network (ANN) method made predictions with more meaningful results compared to support vector regression (SVR) and multiple linear regression (MLR) and appears to be a viable model for institutions that supply housing, mediate housing sales, and provide housing financing and valuation.  Abstract  In the literature, there are two basic approaches regarding the determination of house prices. One of them is the prediction of house price using macroeconomic variables in the country where the house is produced, and another one is the price prediction models, which we can express as micro-variables, by considering the features of the house. In this study, the price of the house was attempted to be predicted using machine learning methods by establishing a model with micro variables that reveal the features of the house. The study was conducted in Turkey’ Antalya province, where household housing demand of foreigners is also high. The house advertisements in locations belonging to the lower, middle- and upper-income groups were selected as the sample. In the results, it was observed that the artificial neural network (ANN) method made predictions with more meaningful results compared to support vector regression (SVR) and multiple linear regression (MLR). These results appear to be a viable model for institutions that supply housing, mediate housing sales, and provide housing financing and valuation. It is considered that this model, which can be used to predict fluctuating house prices, especially in developing countries, will regulate the housing market. |
| 7 | @article{Basha2023HarnessingDI,  title={Harnessing Data-Driven Insights: Predictive Modeling for Diamond Price Forecasting using Regression and Classification Techniques},  author={Md. Shaik Amzad Basha and Peerzadah Mohammad Oveis},  journal={International Journal on Recent and Innovation Trends in Computing and Communication},  year={2023},  url={https://api.semanticscholar.org/CorpusID:265713400}  } | **Harnessing Data-Driven Insights: Predictive Modeling for Diamond Price Forecasting using Regression and Classification Techniques**  This study delves deep into predicting diamond prices in terms of exact monetary values and broader price categories, and highlights the superior performance of the Random Forest regressor in predicting exact prices with an R2 value of approximately 0.975.  Abstract  In the multi-faceted world of gemology, understanding diamond valuations plays a pivotal role for traders, customers, and researchers alike. This study delves deep into predicting diamond prices in terms of exact monetary values and broader price categories. The purpose was to harness advanced machine learning techniques to achieve precise estimations and categorisations, thereby assisting stakeholders in informed decision-making. The research methodology adopted comprised a rigorous data preprocessing phase, ensuring the data's readiness for model training. A range of sophisticated machine learning models were employed, from traditional linear regression to more advanced   ensemble methods like Random Forest and Gradient Boosting. The dataset was also transformed to facilitate classification into predefined price tiers, exploring the viability of models like Logistic Regression and Support Vector Machines in this context. The conceptual model encompasses a systematic flow, beginning with data acquisition, transitioning through preprocessing, regression, and classification analyses, and culminating in a comparative study of the performance metrics. This structured approach underscores the originality and value of our research, offering a holistic view of diamond price prediction from both regression and classification lenses. Findings from the analysis highlighted the superior performance of the Random Forest regressor in predicting exact prices with an R2 value of approximately 0.975. In contrast, for classification into price tiers, both Logistic Regression and Support Vector Machines emerged as frontrunners with an accuracy exceeding 95%. These results provide invaluable insights for stakeholders in the diamond industry, emphasising the potential of machine learning in refining valuation processes. |
| 8 | @article{Chen2022PredictionAA,  title={Prediction and Analysis of House Prices in Boston Based on Regression Model},  author={Zuohang Chen},  journal={Proceedings of the 2022 4th International Conference on E-Business and E-Commerce Engineering},  year={2022},  url={https://api.semanticscholar.org/CorpusID:261583319}  } | **[Prediction and Analysis of House Prices in Boston Based on Regression Model](https://www.semanticscholar.org/paper/Prediction-and-Analysis-of-House-Prices-in-Boston-Chen/6c26a3f40ccaee5e15fb0e565c9bfda58b146b8e)**  The housing price data set of Boston is divided and the regression model linear regression, decision tree regression and support vector machine regression SVR is built so as to obtain the relationship between different data related to Boston house price, and this relationship can finally predict the future house price trend in Boston and display it through visual operation.  Abstract  In artificial intelligence learning, Boston housing price forecast analysis problem is a classic regression problem. Based on the housing price information collected by the U.S. Census Bureau in Boston, Massachusetts. this paper divides the housing price data set of Boston and builds the regression model linear regression, decision tree regression and support vector machine regression SVR and trains the data set, so as to obtain the relationship between different data related to Boston house price, and use this relationship to connect all data, it can finally predict the future house price trend in Boston and display it through visual operation. Through three regression model prediction value, respectively compared with the actual value, the trend of overall and actual and estimated values of the same, but there is a certain error, especially when spot prices higher or lower, often cannot get accurate forecast, so the data for the selection of the characteristic value still exists space for improvement, future study needs to get more data and the characteristics of abundant data. |
| 9 | @article{Chu2023CovariatedistanceWR,  title={Covariate-distance Weighted Regression (CWR): A Case Study for Estimation of House Prices},  author={Hone‐Jay Chu and Po-Hung Chen and Sheng-Mao Chang and Muhammad Zeeshan Ali and Sumriti Ranjan Patra},  journal={ArXiv},  year={2023},  volume={abs/2305.08887},  url={https://api.semanticscholar.org/CorpusID:258714640}  } | **[Covariate-distance Weighted Regression (CWR): A Case Study for Estimation of House Prices](https://www.semanticscholar.org/paper/Covariate-distance-Weighted-Regression-(CWR)%3A-A-for-Chu-Chen/21c336f4a650a4a612747694c4605d293d7f040a)**  Geographically weighted regression (GWR) is a popular tool for modeling spatial heterogeneity in a regression model. However, the current weighting function used in GWR only considers the geographical distance, while the attribute similarity is totally ignored. In this study, we proposed a covariate weighting function that combines the geographical distance and attribute distance. The covariate-distance weighted regression (CWR) is the extension of GWR including geographical distance and attribute distance. House prices are affected by numerous factors, such as house age, floor area, and land use. Prediction model is used to help understand the characteristics of regional house prices. The CWR was used to understand the relationship between the house price and controlling factors. The CWR can consider the geological and attribute distances, and produce accurate estimates of house price that preserve the weight matrix for geological and attribute distance functions. Results show that the house attributes/conditions and the characteristics of the house, such as floor area and house age, might affect the house price. After factor selection, in which only house age and floor area of a building are considered, the RMSE of the CWR model can be improved by 2.9%-26.3% for skyscrapers when compared to the GWR. CWR can effectively reduce estimation errors from traditional spatial regression models and provide novel and feasible models for spatial estimation. |
| 10 | @article{Saeed2023PredictiveMO,  title={Predictive Modeling of House Prices; A Regression Approach},  author={Natasha Saeed and Ahsan Saeed and Maria Bashir and Lismer Andres Caceres-Najarro},  journal={2023 18th International Conference on Emerging Technologies (ICET)},  year={2023},  pages={159-164},  url={https://api.semanticscholar.org/CorpusID:266738871}  } | **[Predictive Modeling of House Prices; A Regression Approach](https://www.semanticscholar.org/paper/Predictive-Modeling-of-House-Prices%3B-A-Regression-Saeed-Saeed/38adcc87b867b6537719185fd94cb53c5ceeca55)**  This study compares Random Forest Regressor and the Decision Tree Regressor using an extensive dataset of Bangalore that includes nine property features and shows that the models achieved an optimal accuracy level and provide insights into the variances in model performance.  Abstract  Housing prices have always been an important issue among the business community. The sudden increase in property prices has an impact on economic cycles as well as our means of living. However, understanding several features that affect residential real estate prices is still challenging. Traditional techniques have been unable to choose essential features for precise predictions, which frequently leads to reduced prediction accuracy. In this research work, we utilized machine learning techniques to increase prediction accuracy and get data-driven insights into property evaluation. This study compares Random Forest Regressor and the Decision Tree Regressor using an extensive dataset of Bangalore that includes nine property features. The findings show that our models achieved an optimal accuracy level and provide insights into the variances in model performance, assisting policymakers in choosing the most suitable approaches to get accurate house prices across various market circumstances. |
| 11 | @article{Li2023ForecastingHR,  title={Forecasting House Resale Prices using Ensemble learning},  author={Zhenyu Li and Zhaoji Li},  journal={2023 2nd International Conference on Big Data, Information and Computer Network (BDICN)},  year={2023},  pages={107-111},  url={https://api.semanticscholar.org/CorpusID:258379049}  } | **[Forecasting House Resale Prices using Ensemble learning](https://www.semanticscholar.org/paper/Forecasting-House-Resale-Prices-using-Ensemble-Li-Li/b6c37cc2179253126198a3a00409f3c3ac156ea2)**  This paper proposes a blending model based on Python and its xgboost, DF21, and Geatpy packages to predict resale prices in Singapore that can obtain a mean absolute percentage error (MAPE) of 7.36% in a stationary overall trend of house prices trend.  Abstract  Forecasting House sales prices is a critical issue for individual investors and government. prediction system has high accuracy and flexibility and plays an important role in decision making. However, previous studies have focused only on limited datasets, as well as nontime series or feature prediction. This paper proposes a blending model based on Python and its xgboost, DF21, and Geatpy packages to predict resale prices in Singapore. First, the high-cardinality categorical attributes are preprocessed by means of mean coding. Then, we propose a linear blending method that includes GA-HL-Xg-Boost, GARandom Forest (GA-RF), deep-Random Forest (DRF) and lightGBM, with Gini impurity to determine the importance of features. Finally, the result shows that it can obtain a mean absolute percentage error (MAPE) of 7.36% in a stationary overall trend of house prices trend. Our research can provide a powerful predictor of house resale prices in different economic environments. |
| 12 | @article{Miao2021SupportVR,  title={Support Vector Regression with Gaussian kernel for Housing Prices Prediction},  author={Dingyang Miao and Hongru Tang and Boshen Wang},  journal={Journal of Physics: Conference Series},  year={2021},  volume={1994},  url={https://api.semanticscholar.org/CorpusID:236974103}  } | **[Support Vector Regression with Gaussian kernel for Housing Prices Prediction](https://www.semanticscholar.org/paper/Support-Vector-Regression-with-Gaussian-kernel-for-Miao-Tang/539c6241d6f7d7a09aa2c3739422b58b23325dc4)**  The support vector regression (SVR) with Gaussian kernel is developed to predict housing prices and demonstrates superior accuracy and effectiveness compared with the SVR with linear kernel, KNN, and decision tree.  Abstract  The housing sector is one of the main sources of economic growth in both developing and developed countries. It is reported that nearly half of people buy or sell houses at an inappropriate price. Based on the public data set of Boston housing prices, this essay analyzed the factors affecting house prices and selected the five most important factors based on the decision tree with the ID3 algorithm. Then, this essay developed the support vector regression (SVR) with Gaussian kernel to predict housing prices. Experimental results showed that our method achieves superior accuracy and effectiveness compared with the SVR with linear kernel, KNN, and decision tree. To verify the applicability of our model, this research applied this model in Beijing housing price data, and it also achieved satisfactory fitting results. |
| 13 | @article{Lv2022AnalysisAF,  title={Analysis and Forecast of Influencing Factors on House Prices Based on Machine Learning},  author={Chenyue Lv and Yingxin Liu and Lidong Wang},  journal={2022 Global Conference on Robotics, Artificial Intelligence and Information Technology (GCRAIT)},  year={2022},  pages={97-101},  url={https://api.semanticscholar.org/CorpusID:252720096}  } | **[Analysis and Forecast of Influencing Factors on House Prices Based on Machine Learning](https://www.semanticscholar.org/paper/Analysis-and-Forecast-of-Influencing-Factors-on-on-Lv-Liu/a27d4329b03c114711979e68c74dad7eb3ddd84d)**  The analysis of the influencing factors and trends of Shenzhen's commercial house prices can provide reasonable policy suggestions for the development of the real estate industry and construction of long-term regulation of real estate in the future.  Abstract  Taking Shenzhen as an example, based on the relevant quantitative data affecting commercial house prices in Shenzhen from 2001 to 2020, the main factors affecting commercial house prices in Shenzhen were screened out by using Lasso regression, and grey prediction model was constructed by using the data passing the grade ratio test to predict the data of each variable in the next 5 years. Afterwards, a Support Vector Regression model was constructed to predict the trend of commercial house prices of Shenzhen based on the values of each variable from 2001 to 2025. The results of the study show that the main influencing factors of Shenzhen commercial house prices are the sales area of residential commercial housing, the purchase and sale prices of second-hand houses, etc. The growth of commercial house prices in Shenzhen from 2021 to 2025 is greater and faster compared to previous years. The analysis of the influencing factors and trends of Shenzhen's commercial house prices can provide reasonable policy suggestions for the development of the real estate industry and construction of long-term regulation of real estate in the future. |
| 14 | @article{Zhang2023ResearchOT,  title={Research on the Influencing Factors and Prediction of Housing Prices Based on Regression Analysis --- Taking California as an Example},  author={Kunlong Zhang},  journal={Advances in Economics, Management and Political Sciences},  year={2023},  url={https://api.semanticscholar.org/CorpusID:261784296}  } | **[Research on the Influencing Factors and Prediction of Housing Prices Based on Regression Analysis --- Taking California as an Example](https://www.semanticscholar.org/paper/Research-on-the-Influencing-Factors-and-Prediction-Zhang/8a163c9eded939c047511cb8811bb2c54f9defb2)**  A house is a necessity for everyone's life. But in today's world, with the continuous reduction of land area, the continuous increase of population, and the continuous maturity of the real estate industry, people have to consider more carefully what factors are the most important in the house. Based on the background of housing prices in California, this paper uses linear regression, random forest and principal component analysis to determine which variables have the greatest impact on housing prices. The reason for using three methods is to obtain more accurate results. According to the results, linear regression shows that income is the most relevant variable to housing prices. Random Forest shows an R square of 75.7%, meaning the predictions fit the data fairly well. Principal component analysis also shows that income is the most important variable. At the same time, house prices will be predicted based on the data obtained, which is about 116598.48588. Some suggestions for those who want to buy a house - which factors of the house are more important. Second, this article will also analyse how to face the rising housing prices. |
| 15 | @article{Afonso2019HousingPP,  title={Housing Prices Prediction with a Deep Learning and Random Forest Ensemble},  author={Bruno Klaus de Aquino Afonso and Luckeciano Carvalho Melo and Willian Oliveira and Samuel Bruno da Silva Sousa and Lilian Berton},  journal={Anais do Encontro Nacional de Intelig{\^e}ncia Artificial e Computacional (ENIAC 2019)},  year={2019},  url={https://api.semanticscholar.org/CorpusID:214699727}  } | **[Housing Prices Prediction with a Deep Learning and Random Forest Ensemble](https://www.semanticscholar.org/paper/Housing-Prices-Prediction-with-a-Deep-Learning-and-Afonso-Melo/82b1a7bbbca62c379a144ada2c95b782c6d6aa48)**  This study analyzes a dataset composed of 12,223,582 housing advertisements, collected from Brazilian websites from 2015 to 2018, and demonstrates that enriching the dataset and combining different ML approaches can be a better alternative for prediction of housing prices in Brazil.  Abstract  The development of a housing prices prediction model can assist a house seller or a real estate agent to make better-informed decisions based on house price valuation. Only a few works report the use of machine learning (ML) algorithms to predict the values of properties in Brazil. This study analyzes a dataset composed of 12,223,582 housing advertisements, collected from Brazilian websites from 2015 to 2018. Each instance comprises twenty-four features of five different data types: integer, date, string, float, and image. To predict the property prices, we ensemble two different ML architectures, based on Random Forest (RF) and Recurrent Neural Networks (RNN). This study demonstrates that enriching the dataset and combining different ML approaches can be a better alternative for prediction of housing prices in Brazil. |
| 16 | @article{Zong2023HousePP,    title={House Prices Prediction Advanced Regression Techniques},    author={Yue Zong},    journal={Advances in Economics, Management and Political Sciences},    year={2023},    url={https://api.semanticscholar.org/CorpusID:265629120}  } | [**House Prices Prediction Advanced Regression Techniques**](https://www.semanticscholar.org/paper/House-Prices-Prediction-Advanced-Regression-Zong/30ee417f35fae286e354b8696a870dca54aee886)  A neural network model was built, parameters were adjusted, and the trained network was used to predict housing prices, positioning the model among the top performers among all machine learning methods.  Abstract  In recent years, with the development of the real estate industry, housing prices have continued to rise. The nation, society, and individuals are all concerned about these prices. For commodity housing prices, there are many factors that influence the housing prices. Apart from national regulations, factors such as lighting, layout, and environment of the houses themselves also have a certain impact on the prices, leading to significant fluctuations in the real estate market. Therefore, researching an accurate model for predicting housing prices has practical significance. It can guide residents in housing consumption and provide policy recommendations for government price regulation. Machine learning methods have become a new type of prediction method in this regard. Based on the theories of data analysis and machine learning, a dataset consisting of 2920 data points with 81 attributes was selected from the publicly available Kaggle housing dataset. The data was normalized and analyzed for feature selection. The ranking of attributes most correlated with housing prices was obtained. Subsequently, a neural network model was built, parameters were adjusted, and the trained network was used to predict housing prices. On the Kaggle leaderboard, the RMSE test result stands at 0.1198, positioning our model among the top performers among all machine learning methods. |
| 17 | title={Forestalling House Prices Using Linear Regression in Machine Learning},    author={LATIKA KHARB and Deepak Chahal},    journal={International Research Journal of Modernization in Engineering Technology and Science},    year={2023},    url={https://api.semanticscholar.org/CorpusID:260941791}  } | [**Forestalling House Prices Using Linear Regression in Machine Learning**](https://www.semanticscholar.org/paper/Forestalling-House-Prices-Using-Linear-Regression-KHARB-Chahal/087e2c4888cf8abeb34c745891220c0cbe793278)  This paper focuses on using linear regression to predict house prices, discussing the dataset, preprocessing steps, model training, and evaluation and highlights the strengths, limitations, and potential improvements of the linear regression approach in the context of house price prediction.  Abstract  Predicting house prices is a fundamental problem in the field of real estate and economics. The accurate estimation of house prices is crucial for various stakeholders including homeowners, buyers, sellers, and policymakers. In this paper, we explore the application of linear regression, a simple yet effective machine learning technique, to predict house prices. We discuss the methodology, data preprocessing, feature selection, model training, and evaluation metrics. We also highlight the strengths, limitations, and potential improvements of the linear regression approach in the context of house price prediction. In this paper, we focus on using linear regression to predict house prices, discussing the dataset, preprocessing steps, model training, and evaluation. |
| 18 | @article{Mao2022ASO,    title={A Study on the Prediction of House Price Index in First-Tier Cities in China Based on Heterogeneous Integrated Learning Model},    author={Yaqi Mao and Yonghui Duan and Yibin Guo and Xiang Wang and Shen Gao},    journal={Journal of Mathematics},    year={2022},    url={https://api.semanticscholar.org/CorpusID:252522835}  } | [**A Study on the Prediction of House Price Index in First-Tier Cities in China Based on Heterogeneous Integrated Learning Model**](https://www.semanticscholar.org/paper/A-Study-on-the-Prediction-of-House-Price-Index-in-Mao-Duan/512e17b033dd039935f8ea577c1ce66a87cb5d8b)  The state-of-the-art interpretation method SHAP combined with the LightGBM algorithm is used to interpret the model, and the result can be used as a basis for real estate policy planning and adjustment and even guide the demand of home buyers, thus improving the efficiency and effectiveness of government policy making.  AbstractTo address the difficulty of low prediction accuracy, insufficient model stability, and certain lag associated with a single machine learning model in the prediction of house price, this paper proposes a multimodel fusion house price prediction model based on stacking integrated learning. Firstly, web search data affecting house prices were collected by web crawler technology, and Spearman correlation analysis was performed on the attribute set to reduce its complexity and establish a prediction index system for four first-tier cities in China. Secondly, with the goal of improving accuracy, diversity, and generalization ability, the types of base learners as well as metalearners are determined, and the parameters of the base learners are optimized using the grey wolf optimization algorithm to produce the GWO-stacking model, and the experimental results from four datasets demonstrate that the model has high prediction accuracy. Finally, to solve the issue of unintelligible black boxes in machine learning models, we have used the state-of-the-art interpretation method SHAP combined with the LightGBM algorithm to interpret the model, and the result can be used as a basis for real estate policy planning and adjustment and even guide the demand of home buyers, thus improving the efficiency and effectiveness of government policy making. |
| 19 | @inproceedings{Prakash2020GUIBH,    title={GUI Based Housing Price Prediction Using Machine Learning Approach},    author={T. Jaya Prakash and S. Ashok Kumar},    year={2020},    url={https://api.semanticscholar.org/CorpusID:219676309}  } | [GUI Based Housing Price Prediction Using Machine Learning Approach](https://www.semanticscholar.org/paper/GUI-Based-Housing-Price-Prediction-Using-Machine-Prakash-Kumar/3b8476c78884b59361b6de87dd3b60ea9465e8e1) The object is to break down gadget reading based absolutely techniques for Housing value Prediction guaging through forecast impacts in top notch precision with a far reaching manual affectability assessment of rendition parameters regarding by and large execution in expectation of air pleasant poisons by utilizing precision count.  AbstractLiving arrangement costs development consistently, it's miles a need for a machine to are expecting habitation charges inside what's to come. Habitation rate expectation can help the designer choose the advancing cost of a house and may assist the supporter with arranging the opportune time to purchase a home. To spare you this difficulty in affect the cost of a living arrangement which comprise of physical circumstances, thought and region. The object is to break down gadget reading based absolutely techniques for Housing value Prediction guaging through forecast impacts in top notch precision. The assessment of dataset with the guide of managed Machine Learning acing approach (SMLT) to catch a few measurements resembles, variable ID, uni-variate assessment, bi-variate and multi-variate examination, missing cost medicines and break down the information approval, realities cleaning/making prepared and records representation should be possible on the total given dataset. Our examination shows a far reaching manual affectability assessment of rendition parameters regarding by and large execution in expectation of air pleasant poisons by utilizing precision count. To prescribe a gadget picking up information on based technique to precisely expect the house charge forecast outcomes inside the state of palatable exactness from assessing oversee classification framework acing calculations. Besides, to assess and talk about the general execution of assorted Machine Learning calculations GUI based absolutely UI house value forecast by utilizing traits. |
| 20 | @article{Wang2021DeepLM,    title={Deep Learning Model for House Price Prediction Using Heterogeneous Data Analysis Along With Joint Self-Attention Mechanism},    author={Pei-Ying Wang and Chiao-Ting Chen and Jain-Wun Su and Tingwei Wang and Szu-Hao Huang},    journal={IEEE Access},    year={2021},    volume={9},    pages={55244-55259},    url={https://api.semanticscholar.org/CorpusID:233262582}  } | [Deep Learning Model for House Price Prediction Using Heterogeneous Data Analysis Along With Joint Self-Attention Mechanism](https://www.semanticscholar.org/paper/Deep-Learning-Model-for-House-Price-Prediction-Data-Wang-Chen/1b30433b8928ba3de190c1ab8ade1d974af466b2) This work proposes an end to end joint self-attention model for house price prediction that incorporates attention mechanism and STN network and achieves a low prediction error and outperforms the other models.  AbstractHouse price prediction is a popular topic, and research teams are increasingly performing related studies by using deep learning or machine learning models. However, because some studies have not considered comprehensive information that affects house prices, prediction results are not always sufficiently precise. Therefore, we propose an end to end joint self-attention model for house prediction. In this model, we import data on public facilities such as parks, schools, and mass rapid transit stations to represent the availability of amenities, and we use satellite maps to analyze the environment surrounding houses. We adopt attention mechanisms, which are widely used in image, speech, and translation tasks, to identify crucial features that are considered by prospective house buyers. The model can automatically assign weights when given transaction data. Our proposed model differs from self-attention models because it considers the interaction between two different features to learn the complicated relationship between features in order to increase prediction precision. We conduct experiments to demonstrate the performance of the model. Experimental data include actual selling prices in real estate transaction data for the period from 2017 to 2018, public facility data acquired from the Taipei and New Taipei governments, and satellite maps crawled using the Google Maps application programming interface. We utilize these datasets to train our proposed and compare its performance with that of other machine learning-based models such as Extreme Gradient Boosting and Light Gradient Boosted Machine, deep learning, and several attention models. The experimental results indicate that the proposed model achieves a low prediction error and outperforms the other models. To the best of our knowledge, we are the first research to incorporate attention mechanism and STN network to conduct house price prediction |
| 21 | @article{Zulkifley2020HousePP,    title={House Price Prediction using a Machine Learning Model: A Survey of Literature},    author={Nor Hamizah Zulkifley and Shuzlina Abdul Rahman and Nor Hasbiah Ubaidullah and Ismail Ibrahim},    journal={International Journal of Modern Education and Computer Science},    year={2020},    volume={12},    pages={46-54},    url={https://api.semanticscholar.org/CorpusID:230528609}  } | [House Price Prediction using a Machine Learning Model: A Survey of Literature](https://www.semanticscholar.org/paper/House-Price-Prediction-using-a-Machine-Learning-A-Zulkifley-Rahman/3f833b220ee504e341720a456f2e8e006e869c0a) The findings of this analysis verified the use of the Artificial Neural Network, Support Vector Regression and XGBoost as the most efficient models compared to others and suggest that locational attributes and structural attributes are prominent factors in predicting house prices.  AbstractData mining is now commonly applied in the real estate market. Data mining's ability to extract relevant knowledge from raw data makes it very useful to predict house prices, key housing attributes, and many more. Research has stated that the fluctuations in house prices are often a concern for house owners and the real estate market. A survey of literature is carried out to analyze the relevant attributes and the most efficient models to forecast the house prices. The findings of this analysis verified the use of the Artificial Neural Network, Support Vector Regression and XGBoost as the most efficient models compared to others. Moreover, our findings also suggest that locational attributes and structural attributes are prominent factors in predicting house prices. This study will be of tremendous benefit, especially to housing developers and researchers, to ascertain the most significant attributes to determine house prices and to acknowledge the best machine learning model to be used to conduct a study in this field. |
| 22 | @article{Thamarai2020HousePP,    title={House Price Prediction Modeling Using Machine Learning},    author={M. Thamarai and S. Malarvizhi},    journal={International Journal of Information Engineering and Electronic Business},    year={2020},    url={https://api.semanticscholar.org/CorpusID:216654981}  } | [House Price Prediction Modeling Using Machine Learning](https://www.semanticscholar.org/paper/House-Price-Prediction-Modeling-Using-Machine-Thamarai-Malarvizhi/701c45f31cecbf5d47133602b1615224e335f272) House availability based on desired features of the house and house price prediction are modeled in the proposed work and the model is constructed for a small town in West Godavari district of Andhrapradesh.  Abstract—Machine Learning is seeing its growth more rapidly in this decade. Many applications and algorithms evolve in Machine Learning day to day. One such application found in journals is house price prediction. House prices are increasing every year which has necessitated the modeling of house price prediction. These models constructed, help the customers to purchase a house suitable for their need. Proposed work makes use of the attributes or features of the houses such as number of bedrooms available in the house, age of the house, travelling facility from the location, school facility available nearby the houses and Shopping malls available nearby the house location. House availability based on desired features of the house and house price prediction are modeled in the proposed work and the model is constructed for a small town in West Godavari district of Andhrapradesh. The work involves decision tree classification, decision tree regression and multiple linear regression and is implemented using Scikit-Learn Machine Learning Tool. |
| 23 | @article{Das2020BoostingHP,    title={Boosting house price predictions using geo-spatial network embedding},    author={Sarkar Snigdha Sarathi Das and Mohammed Eunus Ali and Yuan-Fang Li and Yong-Bin Kang and Timos K. Sellis},    journal={Data Mining and Knowledge Discovery},    year={2020},    volume={35},    pages={2221 - 2250},    url={https://api.semanticscholar.org/CorpusID:221397667}  } | [Boosting house price predictions using geo-spatial network embedding](https://www.semanticscholar.org/paper/Boosting-house-price-predictions-using-geo-spatial-Das-Ali/d388466b2b927ed1ed1644529b5e2103d9f659ae) A novel method is presented, the geo-spatial network embedding (GSNE), that learns the embeddings of houses and various types of points of interest in the form of multipartite networks, where the houses and the POIs are represented as attributed nodes and the relationships between them as edges.  AbstractReal estate contributes significantly to all major economies around the world. In particular, house prices have a direct impact on stakeholders, ranging from house buyers to financing companies. Thus, a plethora of techniques have been developed for real estate price prediction. Most of the existing techniques rely on different house features to build a variety of prediction models to predict house prices. Perceiving the effect of spatial dependence on house prices, some later works focused on introducing spatial regression models for improving prediction performance. However, they fail to take into account the geo-spatial context of the neighborhood amenities such as how close a house is to a train station, or a highly-ranked school, or a shopping center. Such contextual information may play a vital role in users’ interests in a house and thereby has a direct influence on its price. In this paper, we propose to leverage the concept of graph neural networks to capture the geo-spatial context of the neighborhood of a house. In particular, we present a novel method, the geo-spatial network embedding (GSNE), that learns the embeddings of houses and various types of points of interest (POIs) in the form of multipartite networks, where the houses and the POIs are represented as attributed nodes and the relationships between them as edges. Extensive experiments with a large number of regression techniques show that the embeddings produced by our proposed GSNE technique consistently and significantly improve the performance of the house price prediction task regardless of the downstream regression model. Relevant source code for GSNE is available at: https://github.com/sarathismg/gsne. |
| 24 | @article{Ahtesham2020HousePP,    title={House Price Prediction using Machine Learning Algorithm - The Case of Karachi City, Pakistan},    author={Maida Ahtesham and Narmeen Zakaria Bawany and Kiran Fatima},    journal={2020 21st International Arab Conference on Information Technology (ACIT)},    year={2020},    pages={1-5},    url={https://api.semanticscholar.org/CorpusID:230999527}  } | [House Price Prediction using Machine Learning Algorithm - The Case of Karachi City, Pakistan](https://www.semanticscholar.org/paper/House-Price-Prediction-using-Machine-Learning-The-Ahtesham-Bawany/35335d6e7f690b2dc2dd3ab4bfbb3e36b0f3a18b) The proposed house price prediction model is able to predict 98% accuracy and Gradient Boosting Model XGBoost is utilized to predict housing prices in Pakistan.  AbstractHouse prices are a significant impression of the economy, and its value ranges are of great concerns for the clients and property dealers. Housing price escalate every year that eventually reinforced the need of strategy or technique that could predict house prices in future. There are certain factors that influence house prices including physical conditions, locations, number of bedrooms and others. Traditionally predictions are made on the basis of these factors. However such prediction methods require an appropriate knowledge and experience regarding this domain. Machine Learning techniques have been a significant source of advanced opportunities to analyze, predict and visualize housing prices. In this paper, Gradient Boosting Model XGBoost is utilized to predict housing prices. Publicly available dataset containing 38,961 records of Karachi city is attained from an Open Real Estate Portal of Pakistan. Lot of work has been done in predicting house prices across many countries, however very limited amount of work has been done for predicting house prices in Pakistan. Our proposed house price prediction model is able to predict 98% accuracy. |
| 25 | @article{MoraGarca2022HousingPP,    title={Housing Price Prediction Using Machine Learning Algorithms in COVID-19 Times},    author={Raul Tomas Mora-Garc{\'i}a},    journal={Land},    year={2022},    url={https://api.semanticscholar.org/CorpusID:253824372}  } | [Housing Price Prediction Using Machine Learning Algorithms in COVID-19 Times](https://www.semanticscholar.org/paper/Housing-Price-Prediction-Using-Machine-Learning-in-Mora-Garc%C3%ADa/45efb25c539abcec425466dd3f205a8d35bce89c) This research is one of the first studies to use machine learning and microdata to explore the incidence of the COVID-19 pandemic on house prices, and shows that machine learning algorithms perform better than traditional linear models because they are better adapted to the nonlinearities of complex data such as real estate market data.  AbstractMachine learning algorithms are being used for multiple real-life applications and in research. As a consequence of digital technology, large structured and georeferenced datasets are now more widely available, facilitating the use of these algorithms to analyze and identify patterns, as well as to make predictions that help users in decision making. This research aims to identify the best machine learning algorithms to predict house prices, and to quantify the impact of the COVID-19 pandemic on house prices in a Spanish city. The methodology addresses the phases of data preparation, feature engineering, hyperparameter training and optimization, model evaluation and selection, and finally model interpretation. Ensemble learning algorithms based on boosting (Gradient Boosting Regressor, Extreme Gradient Boosting, and Light Gradient Boosting Machine) and bagging (random forest and extra-trees regressor) are used and compared with a linear regression model. A case study is developed with georeferenced microdata of the real estate market in Alicante (Spain), before and after the pandemic declaration derived from COVID-19, together with information from other complementary sources such as the cadastre, socio-demographic and economic indicators, and satellite images. The results show that machine learning algorithms perform better than traditional linear models because they are better adapted to the nonlinearities of complex data such as real estate market data. Algorithms based on bagging show overfitting problems (random forest and extra-trees regressor) and those based on boosting have better performance and lower overfitting. This research contributes to the literature on the Spanish real estate market by being one of the first studies to use machine learning and microdata to explore the incidence of the COVID-19 pandemic on house prices. |
| 26 | @article{Guo2020CANML,    title={CAN MACHINE LEARNING ALGORITHMS ASSOCIATED WITH TEXT MINING FROM INTERNET DATA IMPROVE HOUSING PRICE PREDICTION PERFORMANCE?},    author={Jian-qiang Guo and Shu-hen Chiang and Min Liu and Chien-Chun Yang and Kai-yi Guo},    journal={International Journal of Strategic Property Management},    year={2020},    url={https://api.semanticscholar.org/CorpusID:225762221}  } | [CAN MACHINE LEARNING ALGORITHMS ASSOCIATED WITH TEXT MINING FROM INTERNET DATA IMPROVE HOUSING PRICE PREDICTION PERFORMANCE?](https://www.semanticscholar.org/paper/CAN-MACHINE-LEARNING-ALGORITHMS-ASSOCIATED-WITH-Guo-Chiang/627848628da8709abcbb84a4ef7502a61f755539) This research adopts a broader version of text mining to search for keywords in relation to housing prices and then evaluates the predictive abilities using machine learning algorithms to better understand the trends of house prices in China.  AbstractHousing frenzies in China have attracted widespread global attention over the past few years, but the key is how to more accurately forecast housing prices in order to establish an effective real estate policy. Based on the ubiquitousness and immediacy of Internet data, this research adopts a broader version of text mining to search for keywords in relation to housing prices and then evaluates the predictive abilities using machine learning algorithms. Our findings indicate that this new method, especially random forest, not only detects turning points, but also offers prediction ability that clearly outperforms traditional regression analysis. Overall, the prediction based on online search data through a machine learning mechanism helps us better understand the trends of house prices in China. |
| 26 | @article{Wang2021PredictionOH,    title={Prediction of House Price Index Based on Bagging Integrated WOA-SVR Model},    author={Xiang Wang and Shen Gao and Shiyu Zhou and Yibin Guo and Yonghui Duan and Daqing Wu},    journal={Mathematical Problems in Engineering},    year={2021},    url={https://api.semanticscholar.org/CorpusID:240259377}  } | [Prediction of House Price Index Based on Bagging Integrated WOA-SVR Model](https://www.semanticscholar.org/paper/Prediction-of-House-Price-Index-Based-on-Bagging-Wang-Gao/f6e066d10d0fef1481900621dddaf23a37f8179e) A whale algorithm optimized support vector regression model based on bagging ensemble learning method is proposed, and the results show that the prediction accuracy of the proposed integrated model is better than the comparison model in all cases.  AbstractAiming at the shortcomings of a single machine learning model with low model prediction accuracy and insufficient generalization ability in house price index prediction, a whale algorithm optimized support vector regression model based on bagging ensemble learning method is proposed. Firstly, gray correlation analysis is used to obtain the main influencing factors of house prices, and the segmentation forecasting method is used to divide the data set and forecast the house prices in the coming year using the data of the past ten years. Secondly, the whale optimization algorithm is used to find the optimal parameters of the penalty factor and kernel function in the SVR model, and then, the WOA-SVR model is established. Finally, in order to further improve the model generalization capability, a bagging integration strategy is used to further integrate and optimize the WOA-SVR model. The experiments are conducted to forecast the house price indices of four regions, Beijing, Shanghai, Tianjin, and Chongqing, respectively, and the results show that the prediction accuracy of the proposed integrated model is better than the comparison model in all cases. |

**HOUSE PRICES PREDICTION**

**(DỰ ĐOÁN GIÁ NHÀ)**

1. **Giới thiệu**

Những năm gần đây, cùng với sự phát triển của ngành bất động sản, giá nhà đất liên tục tăng cao. Cả nước, xã hội và các cá nhân đều quan tâm đến mức giá này. Đối với giá nhà ở hàng hóa, có rất nhiều yếu tố ảnh hưởng đến giá nhà ở. Ngoài các quy định quốc gia, bản thân các yếu tố như ánh sáng, cách bố trí, môi trường của ngôi nhà cũng có tác động nhất định đến giá cả, dẫn đến thị trường bất động sản có những biến động đáng kể. Vì vậy, việc nghiên cứu một mô hình dự báo giá nhà đất chính xác có ý nghĩa thực tiễn. Nó có thể hướng dẫn người dân trong việc tiêu dùng nhà ở và đưa ra các khuyến nghị chính sách về điều tiết giá của chính phủ [1] [2]. Dự đoán giá nhà là một bài toán cơ bản trong lĩnh vực bất động sản và kinh tế. Việc ước tính chính xác giá nhà là rất quan trọng đối với các bên liên quan khác nhau bao gồm chủ sở hữu nhà, người mua, người bán và các nhà hoạch định chính sách. Trong bài viết này, chúng tôi khám phá ứng dụng hồi quy tuyến tính, một kỹ thuật học máy đơn giản nhưng hiệu quả để dự đoán giá nhà. Chúng tôi thảo luận về phương pháp, tiền xử lý dữ liệu, lựa chọn tính năng, đào tạo mô hình và số liệu đánh giá. Chúng tôi cũng nêu bật những điểm mạnh, hạn chế và những cải tiến tiềm năng của phương pháp hồi quy tuyến tính trong bối cảnh dự đoán giá nhà [2]. Nhà nghiên cứu sử dụng các kỹ thuật học tập tổng hợp, kết hợp Cây hồi quy tăng cường độ dốc (GBRT) và thuật toán XGBoost, để phát triển mô hình dự đoán có thể ước tính chính xác giá nhà dựa trên các tính năng hoặc biến số có liên quan [4]. Trong tài liệu, có hai cách tiếp cận cơ bản liên quan đến việc xác định giá nhà. Một trong số đó là dự đoán giá nhà bằng cách sử dụng các biến số kinh tế vĩ mô ở quốc gia sản xuất ra ngôi nhà, và một trong số đó là mô hình dự đoán giá, mà chúng ta có thể biểu thị dưới dạng các biến vi mô, bằng cách xem xét các đặc điểm của ngôi nhà. Trong nghiên cứu này, giá của ngôi nhà đã được cố gắng dự đoán bằng phương pháp học máy bằng cách thiết lập một mô hình với các biến vi mô tiết lộ các đặc điểm của ngôi nhà. Nghiên cứu được thực hiện tại tỉnh Antalya của Thổ Nhĩ Kỳ, nơi nhu cầu nhà ở hộ gia đình của người nước ngoài cũng rất cao. Quảng cáo nhà ở những địa điểm thuộc nhóm thu nhập thấp, trung bình và cao được chọn làm mẫu [6]. Trong học tập trí tuệ nhân tạo, bài toán phân tích dự báo giá nhà ở Boston là một bài toán hồi quy cổ điển. Dựa trên thông tin về giá nhà đất do Cục Điều tra Dân số Hoa Kỳ tại Boston, Massachusetts thu thập. Bài viết này chia tập dữ liệu giá nhà ở Boston và xây dựng mô hình hồi quy hồi quy tuyến tính, hồi quy cây quyết định và hồi quy máy vectơ hỗ trợ SVR và huấn luyện tập dữ liệu để thu được mối quan hệ giữa các dữ liệu khác nhau liên quan đến giá nhà ở Boston và sử dụng Mối quan hệ này để kết nối tất cả dữ liệu, cuối cùng nó có thể dự đoán xu hướng giá nhà trong tương lai ở Boston và hiển thị nó thông qua hoạt động trực quan. Thông qua ba giá trị dự đoán của mô hình hồi quy, lần lượt được so sánh với giá trị thực tế, xu hướng của giá trị tổng thể, giá trị thực tế và giá trị ước tính là như nhau, nhưng có một sai số nhất định, đặc biệt là khi giá giao ngay cao hơn hoặc thấp hơn, thường không thể dự báo chính xác, do đó, Dữ liệu cho việc lựa chọn giá trị đặc trưng vẫn còn chỗ cần cải thiện, nghiên cứu trong tương lai cần lấy thêm dữ liệu và đặc điểm có dữ liệu phong phú [8]. Hồi quy trọng số địa lý (GWR) là một công cụ phổ biến để mô hình hóa tính không đồng nhất về không gian trong mô hình hồi quy. Tuy nhiên, hàm trọng số hiện tại được sử dụng trong GWR chỉ xem xét khoảng cách địa lý, trong khi tính tương tự thuộc tính hoàn toàn bị bỏ qua. Trong nghiên cứu này, chúng tôi đã đề xuất hàm trọng số đồng biến kết hợp khoảng cách địa lý và khoảng cách thuộc tính. Hồi quy có trọng số khoảng cách đồng biến (CWR) là phần mở rộng của GWR bao gồm khoảng cách địa lý và khoảng cách thuộc tính. Giá nhà bị ảnh hưởng bởi nhiều yếu tố, chẳng hạn như tuổi nhà, diện tích sàn và mục đích sử dụng đất. Mô hình dự đoán được sử dụng để giúp hiểu được đặc điểm của giá nhà khu vực. CWR được sử dụng để hiểu mối quan hệ giữa giá nhà và các yếu tố kiểm soát. CWR có thể xem xét khoảng cách địa chất và thuộc tính, đồng thời đưa ra ước tính chính xác về giá nhà để duy trì ma trận trọng số cho các hàm khoảng cách địa chất và thuộc tính. Kết quả cho thấy các thuộc tính/điều kiện của ngôi nhà và đặc điểm của ngôi nhà, chẳng hạn như diện tích sàn và tuổi nhà, có thể ảnh hưởng đến giá nhà. Sau khi lựa chọn yếu tố, trong đó chỉ xem xét tuổi nhà và diện tích sàn của tòa nhà, RMSE của mô hình CWR có thể được cải thiện từ 2,9% -26,3% cho các tòa nhà chọc trời khi so sánh với GWR. CWR có thể giảm các lỗi ước tính một cách hiệu quả từ các mô hình hồi quy không gian truyền thống và cung cấp các mô hình mới và khả thi để ước tính không gian [9].

1. **Nghiên cứu liên quan**

Trước thách thức khó khăn này, một loạt các nghiên cứu đã xuất hiện, mỗi nghiên cứu đều tìm cách giải mã những bí mật nằm trong lĩnh vực dự đoán giá nhà. Từ các kỹ thuật hồi quy nâng cao được khám phá trong "Kỹ thuật hồi quy nâng cao dự đoán giá nhà [1]" đến phương pháp hồi quy theo trọng số hiệp phương sai được đề cao trong "Hồi quy trọng số hiệp phương sai (CWR).): Một nghiên cứu điển hình về ước tính giá nhà [9]" , các nhà nghiên cứu đã không ngừng theo đuổi những con đường hiểu biết mới. Những tiếng nói đáng chú ý khác trong bản giao hưởng này bao gồm "Dự báo giá bán lại nhà bằng cách sử dụng phương pháp học tập hợp [10]" và "Dự đoán giá nhà bằng phương pháp học tập sâu và rừng ngẫu nhiên [15, 20]". Những tác phẩm này, cùng với hàng triệu tác phẩm khác, làm sáng tỏ các cách tiếp cận đa dạng và các phương pháp đang phát triển được triển khai nhằm mục đích dự đoán giá nhà chính xác. . Sau đó, mô hình dự đoán được đánh giá bằng phương pháp Lỗi bình phương gốc (RMSE). Kết quả RMSE từ một nghiên cứu trước đây sử dụng kết hợp Lasso và XGBoost là 0,11260, trong khi kết quả RMSE từ nghiên cứu này là 0,00480. Điều này cho thấy giá trị RMSE giảm, cho thấy mức độ lỗi trong mô hình thấp hơn. Điều đó cũng có nghĩa là sự kết hợp giữa thuật toán GBRT và XGBoost đã cải thiện thành công độ chính xác dự đoán của mô hình nghiên cứu trước đó [4]. Bài viết này đề xuất một mô hình kết hợp dựa trên Python và các gói xgboost, DF21 và Geatpy của nó để dự đoán giá bán lại ở Singapore có thể đạt được sai số phần trăm tuyệt đối trung bình (MAPE) là 7,36% trong xu hướng tổng thể cố định của xu hướng giá nhà, bài viết này đề xuất một mô hình kết hợp dựa trên Python và các gói xgboost, DF21 và Geatpy của nó để dự đoán giá bán lại ở Singapore. Đầu tiên, các thuộc tính phân loại có lượng số cao được xử lý trước bằng phương pháp mã hóa trung bình. Sau đó, chúng tôi đề xuất phương pháp trộn tuyến tính bao gồm GA-HL-Xg-Boost, GARandom Forest (GA-RF), deep-Random Forest (DRF) và lightGBM, với tạp chất Gini để xác định tầm quan trọng của các tính năng. Cuối cùng, kết quả cho thấy rằng nó có thể đạt được sai số phần trăm tuyệt đối trung bình (MAPE) là 7,36% trong xu hướng chung ổn định của xu hướng giá nhà. Nghiên cứu của chúng tôi có thể cung cấp một công cụ dự báo mạnh mẽ về giá bán lại nhà ở các môi trường kinh tế khác nhau [11]. Hồi quy vectơ hỗ trợ (SVR) với nhân Gaussian được phát triển để dự đoán giá nhà đất và chứng minh độ chính xác và hiệu quả vượt trội so với SVR với nhân tuyến tính, KNN và cây quyết định [12].