

基于XGBoost模型的二手房交易价格预测研究

莫丕昕, 林 耿

闽江学院计算机与大数据学院, 福建 福州

收稿日期: 2024年8月24日; 录用日期: 2024年9月18日; 发布日期: 2024年9月25日

摘 要

本文以2022年成都的二手房房价数据为研究对象, 构建随机森林模型和XGBoost模型来预测二手房价格。首先对数据集进行清洗并可视化处理, 构建虚拟变量, 接着绘制热力图并运用熵值法进行特征值筛选, 选取重要的特征进行训练模型。接着, 采用网格搜索技术分别开发了基于随机森林和XGBoost的预测模型, 并利用决定系数、均方误差和平均绝对误差这三个关键指标来衡量模型的预测准确性, 经过模型比较和结果分析, 发现优化后的XGBoost模型对二手房房价有良好的预测结果, 准确率达90.3%。

关键词

二手房价, 机器学习, XGBoost模型, 预测优化

Research on Predicting Second Hand Housing Transaction Prices Based on XGBoost Model

Pixin Mo, Geng Lin

School of Computer and Data Science, Minjiang University, Fuzhou Fujian

Received: Aug. 24th, 2024; accepted: Sep. 18th, 2024; published: Sep. 25th, 2024

Abstract

This article takes the second-hand housing price data of Chengdu in 2022 as the research object, and constructs a random forest model and XGBoost model to predict the second-hand housing price. Firstly, the dataset is cleaned and visualized to construct virtual variables. Then, a heat map is drawn and the entropy method is used for feature value screening to select important features for training the model. Subsequently, prediction systems based on random forest and

XGBoost were developed using grid search techniques, and the accuracy of the models was measured using three key indicators: coefficient of determination, mean square error, and mean absolute error. After model comparison and result analysis, it was found that the optimized XGBoost model had good prediction results for second-hand housing prices, with an accuracy rate of 90.3%.

Keywords

Second Hand Housing Prices, Machine Learning, XGBoost Models, Predictive Optimization

Copyright © 2024 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

随着中国城市化步伐的加快，城市居住人口持续攀升。受限于土地资源的有限性，城市房价呈现出逐步攀升的趋势。因此，配备设施较为完整且价格较划算的二手房成为了不少青年的选择。二手房周边往往拥有便利的交通以及丰富的商业资源和教育资源。但是，许多买家和卖家通常因为销售渠道复杂、住房情况条件差异大，二手房的价格评定没有统一标准，因此，购房者在二手房进行价格评估时，往往难以达到精确和客观的标准。在我国，房地产价格评估方法主要有市场法、收益法、成本法等，它们大多适用于不同的市场环境，受主观影响较大[1]。

2002 年，Lennart B [2]学者较早地使用了神经网络模型来替代传统的多元线性回归模型，以提高对房价预测的精确度，结果发现神经网络模型的精度明显高于传统模型。为克服传统模型预测存在的问题，2019 年，Liu R [3]首次基于改进遗传算法的长短期记忆预测房价，本研究提出了一种基于长短期记忆人工神经网络的房价预测模型，该模型通过一种改良的遗传算法进行特征选择和超参数优化，利用多级概率交叉方法来筛选出最优解。实验结果表明，该方法在房价预测方面表现出色，相较于传统的反向传播神经网络、支持向量回归以及其它进化算法，具有更显著的优越性。2022 年，Luca R [4]则采用了三种流行的机器学习技术来预测意大利两个城市的住宅价格，通过 API 收集的数据对 ElasticNet、XGBoost 和人工神经网络模型进行训练并使用平均绝对误差对三个模型的结果进行比对。

由于房价越炒越高，导致了大多数的购房者看向二手房市场，因此近几年出现了二手房市场蓬勃发展现象。在房地产市场分析的学术领域，众多学者通过创新的方法和技术，提供了对房价影响因素及其预测的深刻见解。陈世鹏等[5]人通过襄阳市 2012 年的房贷数据，巧妙地构建了随机森林模型，这一模型在预测房贷影响因素方面，相较于传统的多元线性回归和 ARMA 模型，展现出了更优异的预测能力。胡宇晨等[6]人则将目光投向海口市的商品房市场，他们利用自回归模型来洞悉未来房价的波动趋势。林新朗[7]的研究则更为宏观，他通过计算 1998 至 2010 年间中国 13 个省市住房价格的 Moran 指数，揭示了这些地区之间住房价格的正相关性，并进一步采用空间计量方法，深入分析了中国住房价格的区域性特征。李晓刚[8]在其硕士论文中，将集成方法与 XGboost 算法相结合，提出了一种创新的“拟袋装”方法，通过对样本数据的随机分配和子样本建模，实现了对个人信用风险的精准评估和分类。综上所述，大部分学者在预测二手房房价中往往选择机器学习模型进行预测，且在以往的研究成果中可以发现，机器学习预测结果优于传统模型，因此本文主要选择了 XGBoost 模型对成都市二手房价格进行预测。

2. 主要模型介绍

XGBoost 模型

XGBoost 是 Boosting 算法的一种。Boosting 算法的思想是将许多弱分类器集成在一起形成一个强大的分类器[9]。它是一种集成学习框架，通过结合众多 CART 回归树来构建一个强大的分类器，每个树模型都为整体的性能贡献力量。如图 1 为 XGBoost 模型的流程图。

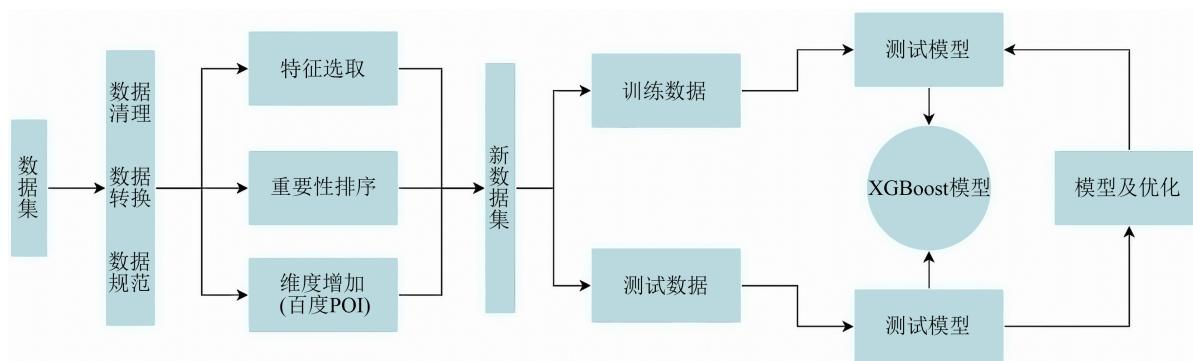


Figure 1. XGBoost model flowchart

图 1. XGBoost 模型流程图

XGBoost 的算法原理在于连续地增加新的决策树。在每次迭代中，新加入的树都会针对前一轮预测的误差进行拟合。这一过程重复进行，直至形成包含 K 棵树的完整模型。预测一个样本的分数，实际上是根据这个样本的特征，每棵树中会落到其对应的一个叶子节点，每个叶子节点对应一个分数，最后将每棵树对应的分数加起来就是该样本的预测值[10]。XGBoost 的目标函数由训练损失函数和正则化项两部分组成，目标函数定义如下：

$$Obj = \sum_{i=1}^N l(y_i, \hat{y}_i) + \sum_{k=1}^K \Omega(f_k),$$

其中 $\hat{y}_i$ 表示第 i 个样本的预测值， y_i 表示第 i 个真实值， $l(y_i, \hat{y}_i)$ 代表训练损失函数。常用的损失函数有以下两种：

- 1) 平方损失函数

$$l(y_i, \hat{y}_i) = (y_i - \hat{y}_i)^2$$

- 2) 逻辑回归损失函数

$$l(y_i, \hat{y}_i) = y_i \ln(1 + e^{-\hat{y}_i}) + (1 - y_i) \ln(1 + e^{\hat{y}_i})$$

XGBoost 作为一个加性模型，其最终的预测得分是通过将所有决策树的预测结果进行累加得到的。其中代表第 i 个样本的树， f_k 代表第 k 颗树的函数。 $\Omega(f_k)$ 代表树的复杂度，将全部 K 棵树的复杂度进行求和，添加到目标函数中作为正则化项，用于防止模型过度拟合[11]。

3. 基于 XGBoost 模型的房价预测

3.1. 数据预处理

3.1.1. 数据来源

本文的原始数据是从阿里云天池获取的开源数据——2022 年成都市二手房交易数据，共有 14 个指

标和 14,659 条数据。

3.1.2. 数据清洗及转换

首先对数据进行了预处理和清洗，并对数据进行了统一的单位处理，由于该数据比较完整，经检查没有缺失值和异常值，所以可以发现数据较准确，可信度高。此外，对分类变量进行了转换为虚拟变量处理，即对对象数据类型转化为模型可识别的类型。图 2 为建筑面积与总房价的散点图，可以看出房价与建筑面积之间大体上呈现出线性相关性，然而，建筑面积并非房价高低的唯一决定性因素。

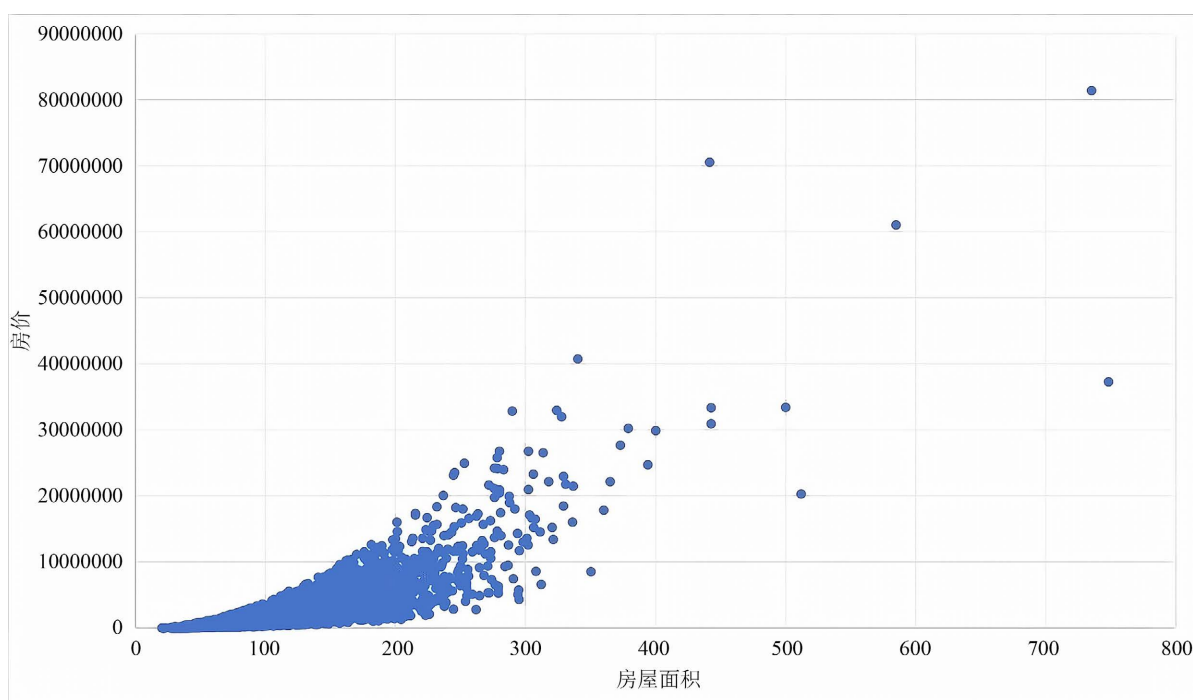


Figure 2. Scatter plot of building area and total housing price

图 2. 建筑面积与总房价的散点图

3.2. 特征选择

特征选取可以将高维数据转化为低维数据，从而达到数据降维的作用，通过删掉冗余和不相关的特征来进一步降维，以减少模型过度拟合，减少特征数量，从而提高泛化能力，还可以让模型得到更好的解释，通过增强特征值之间的联系加快模型的训练速度，从而获得更好的性能[12]。

本文中的特征经初步筛选共 11 个，分别是房屋所属市辖区、所在楼层、户型结构、建筑类型、房屋朝向、建筑结构、装修情况、配备电梯、交易权属、建筑面积和总房价。通过绘制热力图发现，有的特征与房价相关性强，有的比较弱。热力图展示了不同变量之间的相关性，热力图的行和列分别代表不同的变量，颜色的深浅表示相关性的强度，颜色越深表示相关性越强。从热力图中可以看出户型结构与建筑面积的相关性为 0.39，表示有中等程度的正相关性。户型结构与配备电梯的相关性为-0.21，表示有轻微的负相关性。户型结构与所在楼层的相关性为 0.076，表示有轻微的正相关性，意味着户型结构可能随着楼层的增加而变得更加复杂或多样化。户型结构与建筑类型的相关性为 0.13，表示有轻微的正相关性。建筑类型与配备电梯的相关性为-0.5，表示有强负相关性。建筑面积与所在楼层的相关性为 0.05，表示有轻微的正相关性。建筑面积与建筑类型的相关性为 0.077，表示有轻微的正相关性。从而看出户型结构、建筑面积等特征相关性较好，房屋所属市辖区、房屋朝向、装修情况、建筑类型等数据维度对房价影响较小，本文忽略这些特征。

4. 模型比较分析

4.1. 评价指标

回归模型的预测结果和真实值都是连续的, 常用的回归模型评价指标有 R^2 、 MSE 和 MAE , 假设有 n 个样本, $y_i, \hat{y}_i$ ($i=1,2,\dots,n$) 分别代表第 i 个样本的实际值和预测值, 用 $\bar{y}$ 表示 n 个实际值的平均值。

1) 决定系数(R -Squared, R^2)一般用在回归模型中, 用于评估预测值和实际值的符合程度, 该指标的取值介于 0 到 1 之间, 其值越接近 1, 表明模型对因变量方差的解释能力越强, 从而模型的准确性越高; 反之, 若值越小, 则意味着模型的解释力和准确性相对较低[13]。

R^2 定义如下:

$$R^2(y, \hat{y}_i) = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

2) 均方误差(MSE)是指预测值与真实值之差的平方的期望值。 MSE 是衡量数据波动性的一个指标, 它反映了预测模型与实际数据的接近程度, MSE 值的降低, 表明模型的预测精度得到了提升。

MSE 定义如下:

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

3) 平均绝对误差(MAE)用于评估预测结果和真实值的接近程度, 其值越小说明拟合效果越好[14]。

MAE 定义如下:

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|$$

4.2. 结果分析

本文使用了线性回归模型、随机森林模型和 XGBoost 模型对成都二手房房价进行预测, 统一采用了 Sklearn 包中的 `train_test_split()` 函数, 把原始数据按照比例划分为训练集和测试集, 由于训练集合过小的话会导致欠拟合, 所以将实验数据的 70% 作为训练集, 30% 作为测试集[15]。

图 3 和图 4 给出了线性回归模型的预测结果; 图 5 和图 6 给出了随机森林模型的预测结果; 图 7 和图 8 给出了 XGBoost 模型的预测结果; 具体如下。利用 Sklearn 库提供的线性回归工具来训练我们的线性回归模型, 得到的结果如图 3 所示, 可以发现, 预测值与实际值之间的差异显著, 仅在某些情况下两者接近。特别是在房价较高的预测中, 模型的表现并不尽如人意; 此外可以看到图 5 是采用随机森林进行预测得到的结果, 相较于线性模型而言, 效果会好一些。图 7 为采用 XGBoost 模型进行预测得到的结果, 与线性模型的拟合效果相似, 均无法拟合到较高房价, 但房价数值密集处拟合较好。但是需要更精准的指标和更好的模型来预测房价, 猜测是指标过多, 所以采用熵值法对指标数据进行筛选, 手动选筛选出指标重要性排名前 5 的指标, 得到的指标为“建筑面积(平方米)”, “房屋所属市辖区”, “房屋朝向”, “装修情况”, “建筑类型”。再分别通过三种方法进行分析, 得到的结果如图 4、图 6、图 8, 通过下面左右两幅图的比较可以看到在通过熵值法的筛选指标过后预测值与真实值的拟合程度更高了, 说明特征选择对模型优化具有一定的效果。

4.3. 模型优化

由于通过熵值法手动选择筛选指标个数的做法并不严谨, 所以将代码进行二次优化, 且由于第一次优化过后线性回归方程的拟合效果并不如其他两个模型, 所以在二次优化时, 直接剔除线性回归模型,

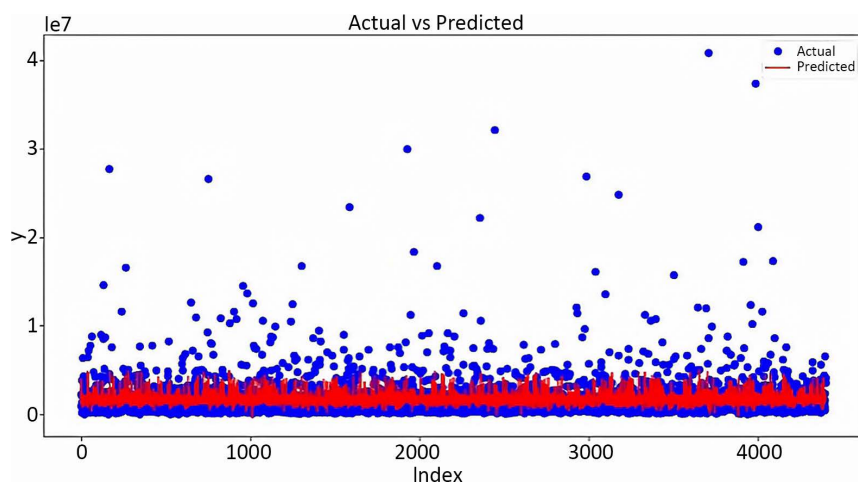


Figure 3. Linear model prediction chart (Original)

图 3. 线性模型预测图(原始)

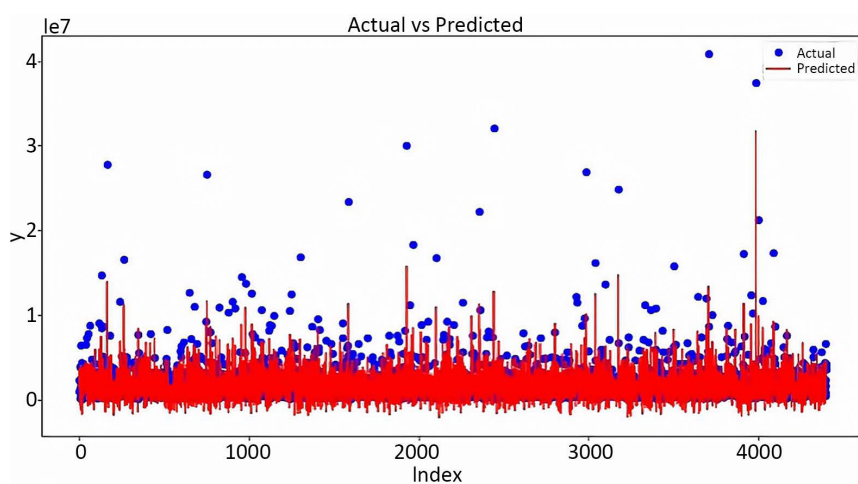


Figure 4. Linear model prediction chart (Once)

图 4. 线性模型预测图(一次)

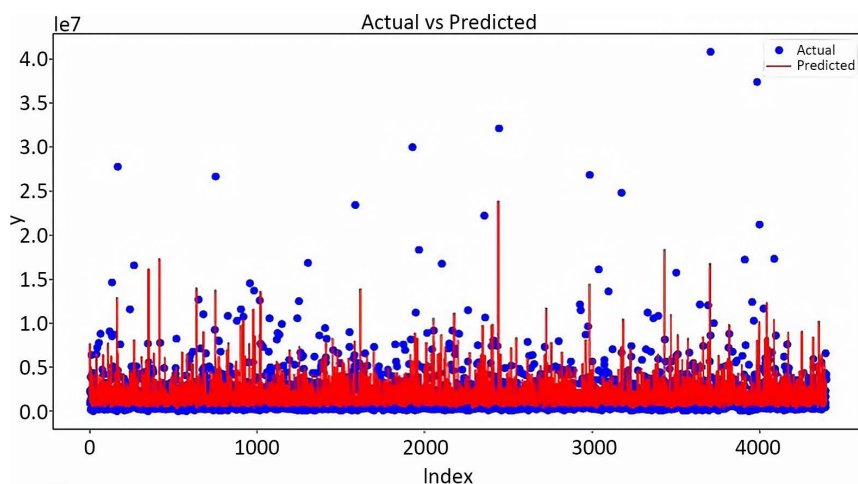


Figure 5. Random forest prediction map (Original)

图 5. 随机森林预测图(原始)

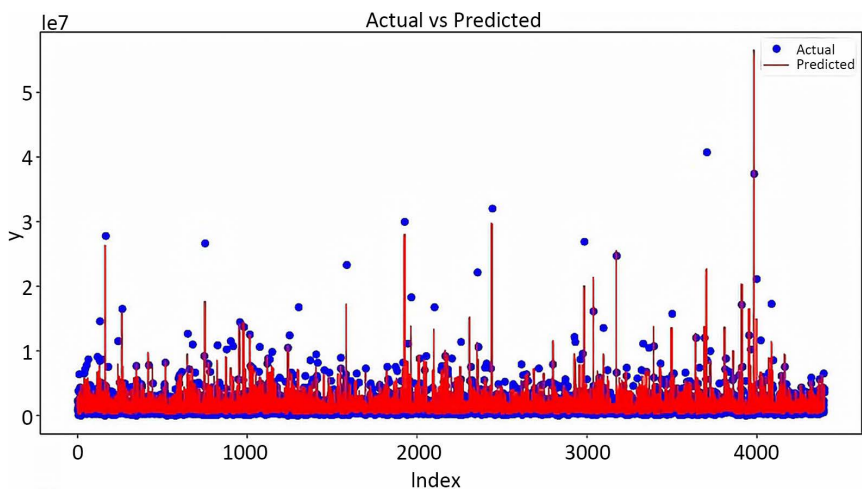


Figure 6. Random forest prediction map (Once)
图 6. 随机森林预测图(一次)

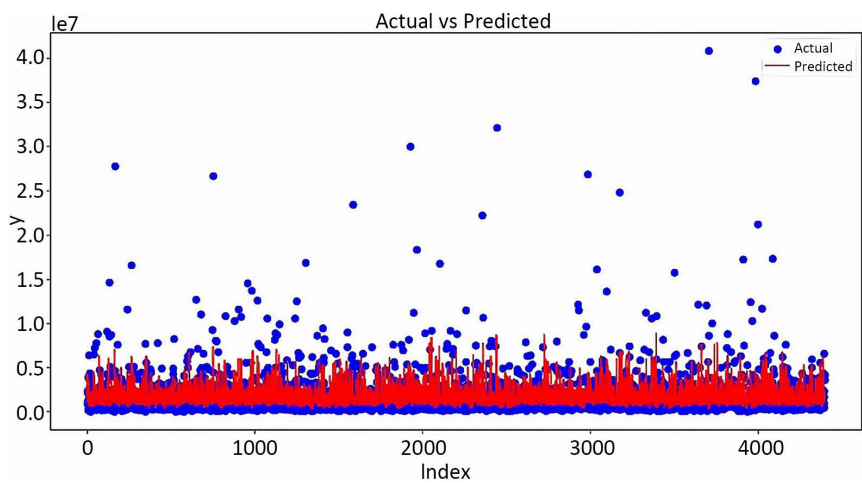


Figure 7. XGBoost prediction graph (Original)
图 7. XGBoost 预测图(原始)

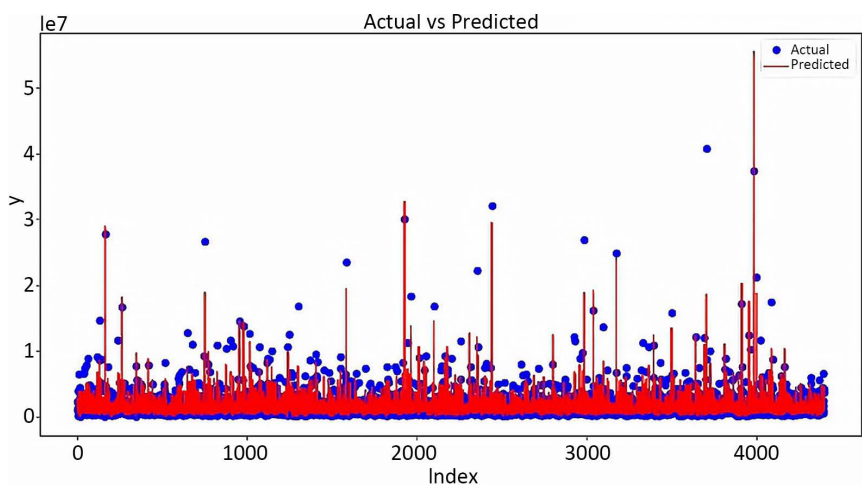


Figure 8. XGBoost prediction graph (Once)
图 8. XGBoost 预测图(一次)

只选取随机森林和 XGBoost 模型进行优化。将熵值法替换为 RFE 特征选择, 与熵值法相比, RFE 通过递归删除特征, 使得模型更加专注于重要的特征, 从而提高了模型的预测准确度。且 RFE 在每次迭代中只考虑剩余的特征, 因此计算量相对较小, 提高了计算效率。并且 RFE 可以输出被选择的特征集合, 就不需要手动筛选个数, 增加了准确性, 也减少了工作量。

此外, 通过网格搜索法(GridSearchCV)对随机森林模型的树的最大深度 `max_depth` 和森林中树的数量 `n_estimators` 做了参数调节, 对于树的最大深度, 测试了三个不同的深度值(例如 5, 10, 15)。对于森林中树的数量, 测试了三个不同的树数量(例如 100, 200, 300)。对 XGBoost 模型的学习率(`learning_rate`)、`max_depth` 和 `n_estimators` 叶进行参数调节[16], 使用交叉验证来评估不同参数组合的性能。对于学习率, 测试了三个不同的学习率(例如 0.05, 0.1, 0.3)。对于树的最大深度和树的数量, 与随机森林模型中的参数设置相同。具体而言, 网格搜索通过指定一个超参数的候选值列表, 对所有可能的参数组合进行穷举搜索, 并对每一组参数进行模型训练和评估, 最后选择具有最佳性能的参数组合作为最终模型的超参数[17]。通过以上优化操作, 图 9 和图 10 分别给出了随机森林模型的预测结果和残差图, 图 11 和图 12 分别给出了 XGBoost 模型的预测结果和残差图。由图 9 和图 11 的预测结果图可以看出, XGBoost 模型在房价预测上具有良好的效果。由图 10、图 12 的残差图可以看出, 真实值与预测值的重叠度较高, 且残差值也随机分布在 $y = 0$ 的上下随机波动, 虽然有个别值波动的情况较大, 但并不影响。

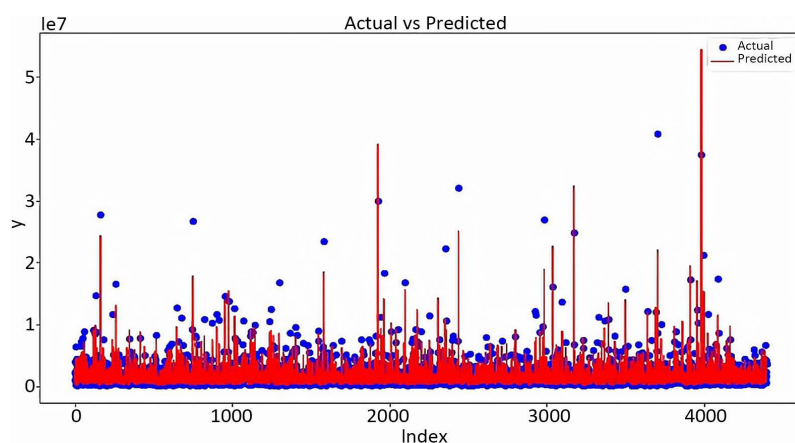


Figure 9. Random forest prediction chart (Secondary)

图 9. 随机森林预测图(二次)

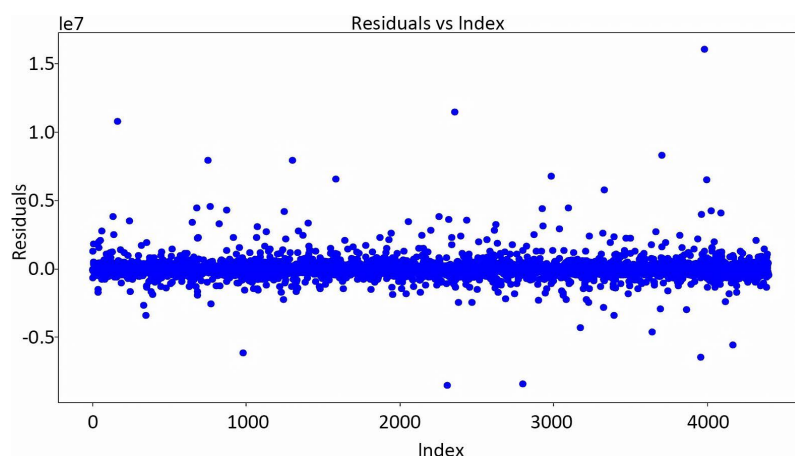


Figure 10. Random forest residual graph

图 10. 随机森林残差图

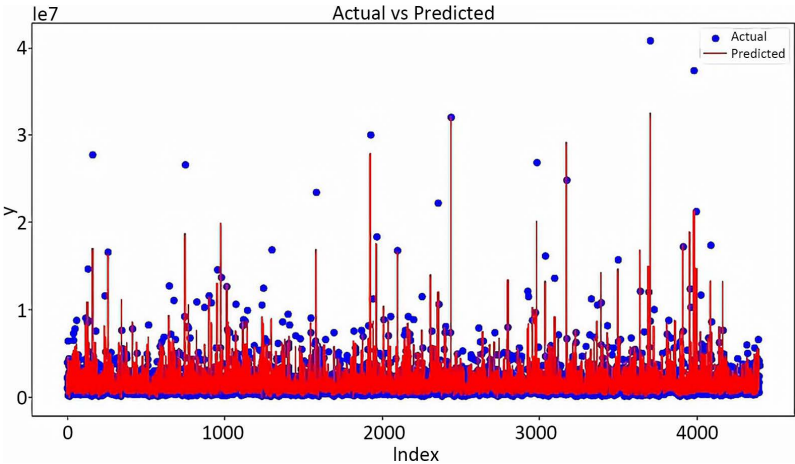


Figure 11. XGBoost prediction graph (Secondary)
图 11. XGBoost 预测图(二次)

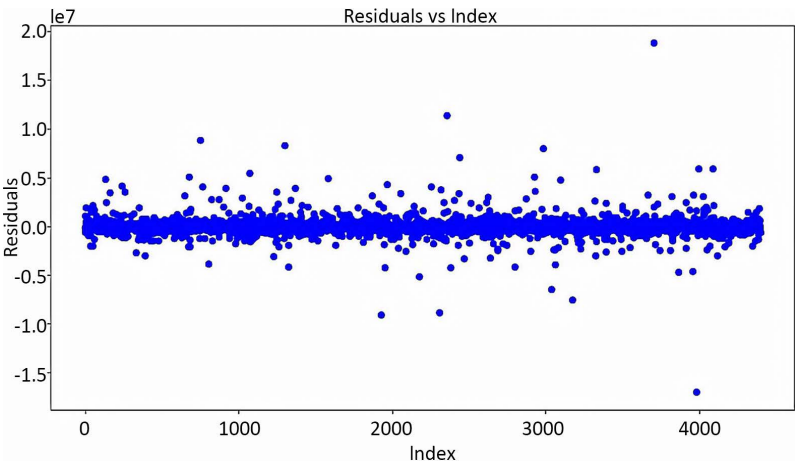


Figure 12. XGBoost residual graph
图 12. XGBoost 残差图

4.4. 比较分析

在本文中，从测试数据集中分别计算了线性回归、随机森林以及 XGBoost 三种模型的评价指标，结果如表 1。根据表 1 的数据显示，线性回归模型的拟合效果相对较弱，随机森林模型的性能优于线性回归模型，而 XGBoost 模型则展现出了最佳的拟合性能。在一次优化过后得到的各模型的拟合效果都有了显著的提升，如表 2 所示，线性回归模型由最初的 R^2 为 0.085 升至 0.603，随机森林的 R^2 值由 0.248 提升至 0.853，XGBoost 模型的 R^2 值由 0.256 提升至 0.834。通过一次优化，每个模型的表现都较好，但随机森林的表现最好。在二次优化后，我们可以看到表 3 的随机森林的初始模型在训练集上的 R^2 为 0.853，测试集上的 R^2 为 0.859，经过参数调优后，模型在训练集上的 R^2 略有提升，测试集上的 R^2 保持稳定。XGBoost 模型的初始模型在训练集上的 R^2 为 0.834，测试集上的 R^2 为 0.903。经过参数调优后，模型在训练集上的 R^2 进一步提升，测试集上的 R^2 略有下降。在随机森林模型中，增加树的数量 $n_estimators$ 和树的最大深度 max_depth 可以提高模型的复杂度，从而在一定程度上提高预测精度。然而，过多的树或过深的树可能导致模型训练时间增加和过拟合的风险。在 XGBoost 模型中，降低学习率 $learning_rate$ 可以使得每次迭代的步长变小，需要更多的迭代次数以达到相同的损失函数下降量，这有助于提高模型的

预测精度。同时，max_depth 和 n_estimators 的调整也需要考虑到模型的泛化能力和训练成本。对比两种模型，XGBoost 在参数调优后展现出更高的 R^2 值，尤其是在测试集上， R^2 值达到 0.903，表明其在未知数据上的预测性能更优。

Table 1. Comparison of initial model structures
表 1. 初始模型结构比较

	R^2	MSE	MAE
线性回归模型	0.085	4487390808171.969	3325.653
随机森林	0.248	3687939313299.952	7561.426
XGBoost 模型	0.256	3648063029321.772	8561.466

Table 2. Comparison of optimization results once
表 2. 一次优化结果比较

	R^2	MSE	MAE
线性回归模型	0.603	1944944939377.614	173152.504
随机森林	0.853	719498490957.419	224981.523
XGBoost 模型	0.834	812526188231.589	258817.536

Table 3. Comparison of secondary optimization results
表 3. 二次优化结果比较

	R^2	MSE	MAE
随机森林	0.859	690116324089.372	344325.531
XGBoost 模型	0.903	573083127172.155	340561.704

5. 研究结论、建议和不足

5.1. 研究结论

本文通过构建线性回归模型、随机森林和 XGBoost 模型对二手房房价进行预测，通过数据预处理、特征工程以及模型构建等环节的精心设计，我们优化了房价预测模型，使其预测结果更贴近实际市场价值。通过模型优化，将 XGBoost 模型的 R^2 提升至 0.903，如表 4 所示是部分真实值与预测值的数据对比，可以发现三种模型的预测误差中，XGBoost 模型的误差最小，但是在现实中影响房价的数据特征还有很多，如周边环境设施、交通便捷程度、小区是否有停车位等[18]因素，未来的研究可以在此基础上进一步优化模型参数，考虑引入更多影响房价的因素，以提高预测的准确性和可靠性。同时，也应关注房地产市场的最新动态，以便及时更新模型，减少预测滞后。

Table 4. Comparison table of real value and predicted value data
表 4. 真实值与预测值数据对比表

真实值	线性回归		随机森林		XGBoost	
	预测值	预测误差	预测值	预测误差	预测值	预测误差
949,198	1185111.631	24.85%	1018282.916	7.28%	943985.25	0.55%
2,306,570	2547989.367	10.47%	2349308.98	1.85%	2217678.5	3.85%

续表

3,904,560	3420350.717	12.40%	3885626.04	0.48%	3818799	2.20%
1,035,988	1960463.073	89.24%	1062816.651	2.59%	1015695.75	1.96%
2,102,430	2443966.966	16.24%	2315790.51	10.15%	2202411.5	4.76%
172,326	191030.1793	10.85%	187835.777	9.00%	171991.8438	0.19%
1,368,000	1167558.728	14.65%	1318147.476	3.64%	1311244.125	4.15%

5.2. 建议

成都市作为四川省省会，在全省经济社会发展中发挥着核心增长的作用，成都市商品住宅市场也真正占据四川商品住宅市场的核心地位。房地产调控政策不是一刀切，而是根据不同的时间、不同的市场、不同的问题需要通过贷款政策、土地政策、税收政策、人才落户政策、规定首付比例政策、出台二手房指导价等政策相互配合并及时做出调整。

政府通过巧妙地运用政策工具和市场自我调节机制，为二手房市场的稳定与繁荣提供支持，确保了市场的有序发展。政府应加强对二手房中介市场的监管，制定和完善相关法律法规，规范中介机构的经营行为，中介机构应公开房源信息、交易价格、服务费用等关键信息，提高市场透明度，保护消费者权益。也可以通过制定行业标准和规范，推动中介机构提高服务质量，维护市场秩序。中介机构应加强对从业人员的培训和管理，提高他们的专业素养和道德水平。同时，建立从业人员信用档案，记录其从业行为和信誉状况，以便消费者查询和选择。

购房者可以通过供需关系、政策变化、经济走势等方面了解市场动态，从而避免盲目跟风或恐慌性购买。中介机构在二手房交易中扮演着重要角色，购房者可以通过沟通，表达自身预期，得到提供相应的价格预测信息。购房者可以通过教育和培训活动，提高自身的市场敏感度和分析能力，使自身能够更好地理解和利用价格预测信息。

5.3. 不足

本研究依托真实的成都市二手房交易数据，从基于网络爬虫技术的数据获取到基于机器学习的价格预测，通过对比实验的方式实现了成都市二手房价格的有效评估。相较于对二手房均价和挂牌价的研究，预测二手房成交价具有更直接的现实意义。它能够在一定程度上指导购房者做出合理的购买决策，减少不必要的经济风险，但本文研究内容也有以下几点不足：

- 1、本文在模型构建的数据采集和加工过程中实施了多种策略，但仍可能存在一定的数据质量问题。例如，部分数据(如房屋内部装修情况、楼层高低、物业管理质量等)可能存在缺失、异常或噪声等问题，这可能会对预测结果产生一定的影响。未来研究可以进一步探索数据清洗和预处理技术，以提高数据质量。
- 2、本研究采用了线性回归和随机森林模型对成都市二手房成交价格的相关影响因素进行了变量筛选，对于模型可能难以完全考虑所有影响房价的外部因素，如自然灾害、社会事件等突发事件。在后续的学习过程中，还需深化对特征工程的理解，以期通过变量筛选获得更精确的预测模型基础变量组。
- 3、本研究主要基于历史数据进行预测，房地产市场变化迅速，模型可能无法及时反映最新的市场动态，导致预测滞后。未来研究可以探索将实时数据引入模型，以提高预测的实时性和准确性。

综合分析表明，应用 XGBoost 模型于成都市二手房市场的价格预测具有实际应用价值。在未来的研究工作中，我们将持续深化学习，针对当前研究的不足进行改进和优化。

基金项目

本研究受到闽江学院华纳数字孪生系统研发实验室(MJUGJSYS2022004)、虚拟仿真实验课程项目(MJUYLKC2022029)的支持。

参考文献

- [1] 李蕾. 基于一维卷积神经网络的住宅类二手房价格评估研究[D]: [硕士学位论文]. 重庆: 重庆理工大学, 2023.
- [2] Berg, L. (2002) Prices on the Second-Hand Market for Swedish Family Houses: Correlation, Causation and Determinants. *European Journal of Housing Policy*, 2, 1-24. <https://doi.org/10.1080/14616710110120568>
- [3] Liu, R. and Liu, L. (2019) Predicting Housing Price in China Based on Long Short-Term Memory Incorporating Modified Genetic Algorithm. *Soft Computing*, 23, 11829-11838. <https://doi.org/10.1007/s00500-018-03739-w>
- [4] Rampini, L. and Re Cecconi, F. (2021) Artificial Intelligence Algorithms to Predict Italian Real Estate Market Prices. *Journal of Property Investment & Finance*, 40, 588-611. <https://doi.org/10.1108/jpif-08-2021-0073>
- [5] 陈世鹏, 金升平. 基于随机森林模型的房价预测[J]. 科技创新与应用, 2016(4): 52.
- [6] 胡宇晨. 农民拆迁安置社区景观环境构建策略研究[D]: [硕士学位论文]. 北京: 中央美术学院, 2020.
- [7] 周文兴, 林新朗. 中国住房价格与城市化水平的关系研究——动态面板和空间计量的实证分析[J]. 重庆大学学报(社会科学版), 2012, 18(5): 1-7.
- [8] 李晓刚. 个人信用风险评估的一种基于 XGBoost 的集成学习方法[D]: [硕士学位论文]. 合肥: 中国科学技术大学, 2018.
- [9] 李天帅. 基于 XGBoost 算法的二手房价格批量评估研究[D]: [硕士学位论文]. 呼和浩特: 内蒙古财经大学, 2024.
- [10] 胡晓伟, 马春梅, 孔祥山, 等. 基于 XGBoost 的深圳二手房价格预测[J]. 曲阜师范大学学报(自然科学版), 2022, 48(1): 57-65.
- [11] 吴广玉. 基于 XGBoost 算法的在线医疗社区患者满意度预测研究[D]: [硕士学位论文]. 昆明: 昆明理工大学, 2021.
- [12] 吴云飞. 基于机器学习的合成致死基因预测研究[D]: [硕士学位论文]. 东莞: 江南大学, 2023.
- [13] 崔慧莹. 基于极端随机森林算法的改进与二手房价预测的研究[D]: [硕士学位论文]. 大连: 大连交通大学, 2023.
- [14] 向鹏成, 高天, 段旭, 等. 基于机器学习的“一带一路”投资国别风险预测研究[J]. 工业技术经济, 2024, 43(7): 150-160.
- [15] 陈庆辉. 面向时间序列预测的深度学习模型研究及应用[D]: [硕士学位论文]. 济南: 齐鲁工业大学, 2024.
- [16] 姚景瑜. 基于 BP 神经网络的轴流风扇气动噪声预测[D]: [硕士学位论文]. 南京: 南京大学, 2020.
- [17] 吕子璇. 基于机器学习方法的烟台市二手房价格预测研究[D]: [硕士学位论文]. 烟台: 鲁东大学, 2023.
- [18] 金长宏, 王晓晓. 基于特征价格模型的合肥市二手房价格影响因素剖析[J]. 湖北第二师范学院学报, 2023, 40(9): 53-61.