

# 基于机器学习的成都市住房租金预测研究

汪宇, 陈美\*

成都信息工程大学计算机学院, 四川 成都

收稿日期: 2025年2月16日; 录用日期: 2025年3月14日; 发布日期: 2025年3月20日

## 摘要

我国城市现代化与住房商品化推进中, 租赁市场热度上升, 但租金定价普遍依赖房东主观判断, 导致信息不对称和资源浪费。科学定价分析与预测成为解决租赁市场健康发展问题的有效途径。基于此, 本文采用机器学习算法对成都市多个区域的住房租金进行了影响因素分析和预测研究。首先, 构建了包含丰富特征的数据集并进行了数据预处理和特征工程。接着, 分别建立了基于随机森林、XGBoost、BP神经网络算法的三种住房租金预测模型, 并采用网格寻优方法对模型参数进行了调优。最后, 采用多种评价指标对模型的预测性能进行了评估。结果表明, 房屋面积、所在行政区域、朝向和出租方式是影响租金的关键因素。XGBoost预测模型在租金预测中表现最佳, 具有更小的误差和更强的泛化能力。

## 关键词

租金预测, 机器学习, 随机森林, XGBoost, BP神经网络

# Research on Housing Rent Prediction in Chengdu Based on Machine Learning

Yu Wang, Mei Chen\*

School of Computer Science, Chengdu University of Information Technology, Chengdu Sichuan

Received: Feb. 16<sup>th</sup>, 2025; accepted: Mar. 14<sup>th</sup>, 2025; published: Mar. 20<sup>th</sup>, 2025

## Abstract

Amid the rapid urban modernization and housing commercialization in China, the rental market has witnessed increasing popularity. However, rent pricing largely relies on landlords' subjective judgments, leading to information asymmetry and resource wastage. Scientific pricing analysis and prediction emerge as effective approaches to addressing the healthy development of the rental market. Based on this background, this paper employs machine learning algorithms to conduct an

\*通讯作者。

analysis of influencing factors and predictive research on housing rents in multiple districts of Chengdu. Firstly, a dataset with rich features was constructed, followed by data preprocessing and feature engineering. Subsequently, three rent prediction models based on Random Forest, XGBoost, and BP Neural Network algorithms were established, respectively, and the model parameters were optimized using a grid search method. Finally, multiple evaluation metrics were adopted to assess the predictive performance of the models. The results indicate that house area, administrative district, orientation, and rental method are key factors influencing rent. The XGBoost prediction model performs best in rent prediction, with smaller errors and stronger generalization ability.

## Keywords

Rent Prediction, Machine Learning, Random Forest, XGBoost, BP Neural Network

Copyright © 2025 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

随着我国城市现代化和住房商品化进程的日益推进,房地产行业蓬勃发展。住房租赁市场,作为房地产市场中不可或缺的一部分,在政策的有力支持和年轻人租房观念转变等多重因素的推动下,热度持续上升。然而,当前租金定价大多依赖于房东主观判断,缺乏科学的定价机制,导致房东与租客之间的信息不对称问题。有时因定价过高致使房源难以出租,进而造成房源空置,资源浪费,长期制约着租赁市场的健康、稳定发展。因此,把控租赁市场的价格走向显得尤为关键,而对住房租金进行科学、合理的定价分析与预测,成为了解决当前问题的有效途径。

近年来,国内外学者采用时间序列分析等回归分析方法对住房租金影响因素进行了深入研究。相较于国内,国外住房租赁市场发展更为成熟,相关研究起步较早。Marks [1]利用价格特征指数法对温哥华租赁市场进行了分析,发现配备有停车位和洗衣店的住房其租金相对更高。Guntermann 和 Norrbin [2]则对凤凰城大都市区域的公寓租金进行了研究,发现住房附近设有健身房、游泳池等休闲设施也会提升租金水平。Sirmans [3]提出了一个评价住房提供特定设施和服务的成本与效益的模型,研究表明,住房内部配套设施的完善程度同样对租金有着明显的影响。

尽管我国关于住房租金的相关研究起步较晚,但近年来逐渐受到广泛关注。张若曦和贾士军[4]采用回归分析方法建立了最优特征价格模型,探讨了广州市住房与地铁站距离、周边学校等级等因素对租金的影响。陆军等[5]研究了北京市住房租金与公共设施服务之间的关系,发现租金与医院之间存在负相关关系,而与电影院等娱乐设施之间存在正相关关系。王世桢等[6]则通过建立房租的收益率预测模型与租售比预测模型,对上海市住房租金增长率的影响因素进行了详细的分析。

此外,住房价格预测方面的相关研究也成为了学者们关注的焦点。Antipov 和 Pokryshevskaya [7]采用随机森林、多元线性回归以及决策树模型对房价进行了预测,结果表明随机森林模型表现最优。Sarip 等[8]提出了基于模糊最小二乘回归的房价预测模型,并与自适应神经模糊推理系统和神经网络模型在预测精度和算法复杂度上进行了对比。Li [9]运用 LightGBM 模型对脱敏后的真实租赁市场数据进行建模,证实了其在租金预测中的优秀性能。谢勇等[10]也通过对比在不同数据集训练下的预测结果,发现 XGBoost 和 LightGBM 模型对住房月租金的预测效果优于传统 GBDT 模型。

综上所述,相较于对住房租金影响因素的定性探讨,当前利用机器学习算法和大数据分析技术对住

房租金进行定量预测的研究仍显缺乏。部分研究在收集住房租金的影响因素时, 存在考虑影响因素类别过少、收集数据的区域范围过小等问题, 造成数据集特征不全面, 进而影响预测性能。本文借助机器学习强大的特征表示及其在回归任务中的优秀表现, 拟提出一种基于机器学习算法的住房租金预测模型, 全面分析影响成都市住房租金的各类因素, 实现对成都市住房租金的高效、准确预测, 为房东和租客提供直观、可靠的租金信息参考。

2. 机器学习相关算法

随机森林(Random Forest)算法[11]与 XGBoost (Extreme Gradient Boosting)算法[12], 作为集成学习领域的典型代表方法, 通过结合多个弱学习器的预测结果, 利用平均或多数投票等方式, 显著提高了预测的准确性。此外, BP (Back Propagation)神经网络[13]以其显著的非线性拟合能力, 在解决回归任务时展现出了强大的潜力。鉴于上述算法的各自优势, 本文拟通过采用这些机器学习算法, 对住房租金进行影响因素分析和价格预测研究, 以更精准地捕捉住房特征与住房租金之间的复杂非线性映射关系, 从而提升预测精度。

2.1. 随机森林算法

随机森林算法的核心思想在于通过构建多个决策树, 并将这些决策树的预测结果进行整合, 以此来提高预测准确率并有效避免过拟合现象。这种集成学习方法利用了不同决策树之间的差异性来减少预测偏差, 同时, 由于每个决策树在构建过程中都引入了随机性, 使得它们之间的相关性较低, 进而降低了模型整体的方差, 增强了模型的泛化能力。因此, 随机森林算法被广泛认为是一种高效且鲁棒的机器学习算法, 算法流程如图 1 所示。

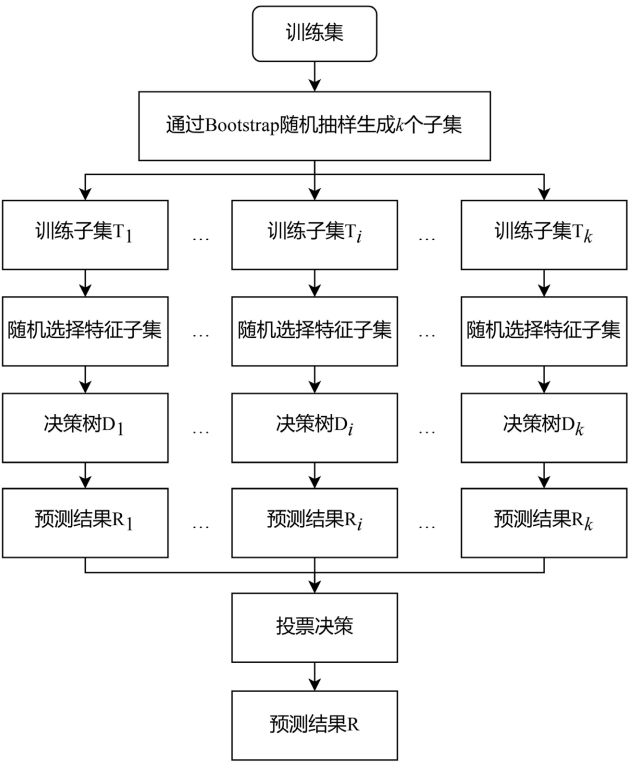


Figure 1. Flowchart of random forest algorithm  
图 1. 随机森林算法流程图

## 2.2. XGBoost 算法

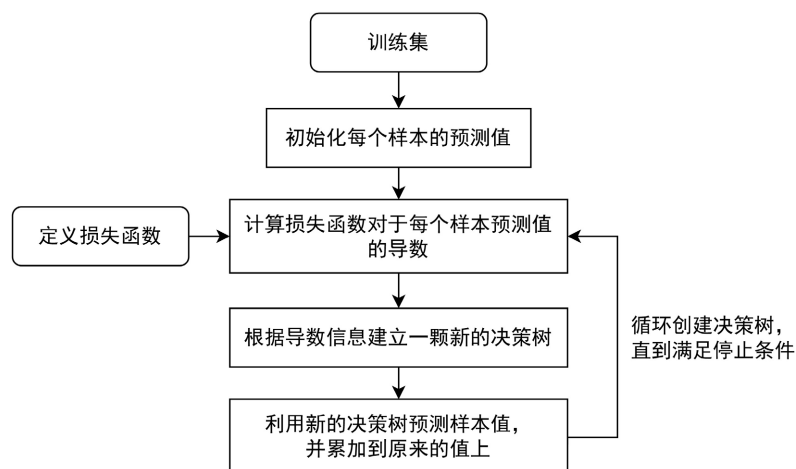


Figure 2. Flowchart of XGBoost algorithm

图 2. XGBoost 算法流程图

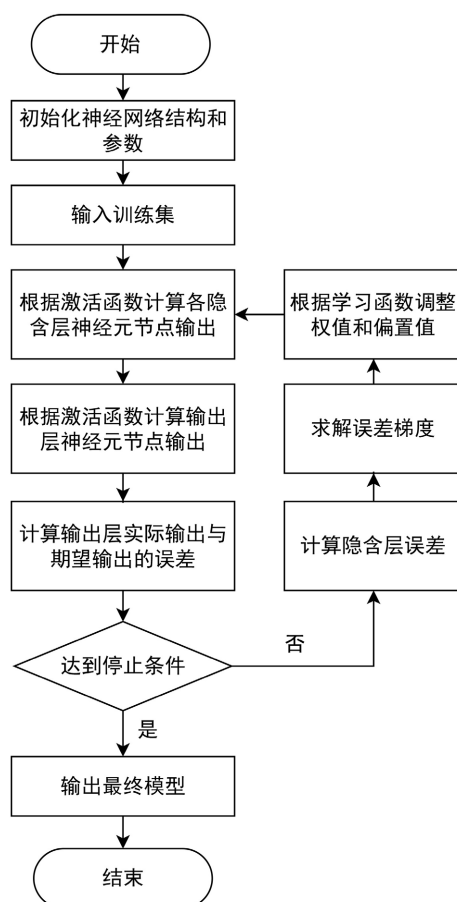


Figure 3. Flowchart of BP neural network algorithm

图 3. BP 神经网络算法流程图

XGBoost 算法的核心思想在于通过构建一系列增强的决策树(梯度提升树), 并以前一棵树的结果为

基础不断优化和构建后续树,以此来提升预测性能。这种梯度提升框架结合了多个弱学习器(决策树)的力量,通过迭代地减少预测值与真实值之间的残差来逐步逼近最优解。XGBoost 算法流程如图 2 所示。

2.3. BP 神经网络算法

BP 神经网络算法的核心思想在于通过构建一个多层前馈神经网络,并利用反向传播算法来训练这个网络,以实现输入数据的准确预测。该算法的优势在于强大的非线性映射能力和自学习能力,通过不断调整网络参数,网络能够捕捉到输入数据中的复杂模式和特征,并据此进行准确的预测。BP 神经网络算法流程如图 3 所示。

3. 住房租金预测模型

3.1. 分析流程

本文住房租金预测分析流程如下:

- 1) 住房数据的采集、预处理与特征工程。首先,采用网络爬虫技术,爬取某租售房网站近年来成都市的租房交易数据及房屋相关信息,将影响住房租金的各个因素视为特征,同时通过统计分析描述特征变量对住房租金分布情况的影响;随后,通过分析数据的来源与结构对数据进行预处理操作,例如对异常值进行剔除、对缺失值进行填充、对特征进行标准化处理等,以确保数据的准确性和完整性;最后,对数据集进行特征工程,以提升模型的学习效率和预测准确性。
- 2) 住房租金预测模型的建立。分别建立基于随机森林、XGBoost 以及 BP 神经网络三种机器学习算法的住房租金预测模型,并采用网格搜索方法对模型参数进行调优,随后对经过预处理后的训练特征集进行建模分析。
- 3) 住房租金预测模型的最优选取。采用平均绝对百分比误差(MAPE)、均方根误差(RMSE)以及决定系数( $R^2$ )作为模型评价标准,对建立的三种租金预测模型的性能进行综合评价,最终选取具有最优预测性能的住房租金预测模型。

3.2. 数据描述

3.2.1. 数据采集

本文针对成都市 7 个主城区的住房租金进行研究,所采用的数据主要爬取自某大型租售房网站,共获得 19,887 条样本。每条样本包含 26 个特征,具体如表 1 所示。

Table 1. Sample characteristics and explanation  
表 1. 样本特征及说明

序号	特征	说明
1	出租方式	整租、合租
2	区域	高新区、天府新区、锦江区、青羊区、武侯区、金牛区、成华区
3	户型	例如: 3 室 1 厅 1 卫
4	朝向	东、西、东南等
5	面积	房屋的总面积
6	楼层	低、中、高
7	总楼层	房屋所在楼栋的总楼层
8	用水	民水、商水

续表

9	用电	民电、商电
10	电梯	有为1，没有为0
11	车位	免费试用、租用车位
12	精装修	是为1，不是为0
13	近地铁	是为1，不是为0
14	洗衣机	有为1，没有为0
15	空调	有为1，没有为0
16	衣柜	有为1，没有为0
17	电视	有为1，没有为0
18	冰箱	有为1，没有为0
19	热水器	有为1，没有为0
20	床	有为1，没有为0
21	暖气	有为1，没有为0
22	宽带	有为1，没有为0
23	天然气	有为1，没有为0
24	月租金	上一次交易的月租金
25	付款方式	月付、季付、年付等
26	租期	月租、季租、年租等

通过对原始数据集进行统计分析可获得月租金分布情况，如图4所示。可以看出，月租金变化范围主要集中在1200至20,000元之间。

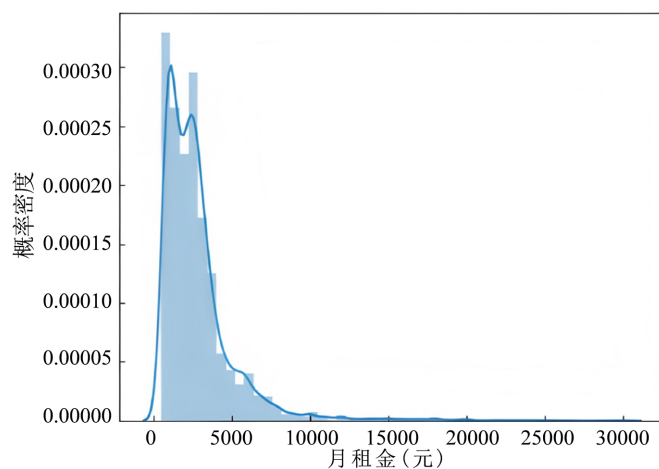


Figure 4. Monthly rent distribution

图4. 月租金分布图

### 3.2.2. 数据预处理

#### 1) 异常值处理

对原始数据集进行分析发现，某些出租房屋中，一个样本存在有多个朝向的问题。同一样本同一特

征存在多个特征值难以直接代入预测模型进行分析, 因此本文按照生活常识对房屋朝向优劣的顺序进行判断, 并选择最优的朝向作为该样本的朝向, 进而将房屋朝向依次归类为东, 南, 西, 北, 东南, 西南, 东北, 西北共 8 个方向。

### 2) 重复样本处理

考虑到网站可能设有反爬机制, 导致采集到的数据中可能包含部分重复的样本。经过分析, 共发现 285 条重复样本。为了保证数据的准确性和有效性, 本文直接将这些重复的样本进行删除。

### 3) 缺失值处理

通过分析可知, 原始样本中的某些特征值存在缺失, 缺失比例如表 2 所示。

**Table 2.** Proportion of missing features in the dataset

**表 2.** 数据集所含特征缺失值比例

特征	缺失比例
付款方式	1.70%
用电	10.80%
用水	11.20%

为保证数据特征的完整性和有效性, 本文对上述缺失值进行处理。由表 2 可知, 付款方式特征存在少量缺失值, 本文选择直接删除含有该缺失值的样本数据。而用电和用水这两个特征的缺失值比例相对较大, 且都为定性特征, 因此分别采用出现频率最高的用电和用水特征值对其进行填充。

### 4) 标准化处理

为了消除不同特征之间在数值范围上存在差异对模型性能带来的不利影响, 本文对面积、总楼层等数值型特征进行了标准化处理, 从而确保了模型能够更均衡地考虑各个特征, 提高其准确性和鲁棒性。本文采用 Z-score 标准化方法, 使得经过处理后的数据均服从均值为 0、方差为 1 的正态分布。Z-score 标准化方法如式(1)所示:

$$z = \frac{x - \mu}{\sigma} \quad (1)$$

式中,  $x$  为原始数据,  $z$  为处理后的数据,  $m$  为均值,  $s$  为标准差。

### 5) 分类变量独热编码

通过分析发现, 区域、户型、朝向、楼层、用水、用电, 车位、付款方式以及租期这些特征未进行编码, 不能代入模型直接进行计算, 因此对上述分类变量进行独热编码(One-Hot Encoding)处理[14]。以楼层特征为例, 对其进行独热编码后效果如表 3 所示。

**Table 3.** Floor feature unique hot coding and explanation

**表 3.** 楼层特征独热编码及说明

楼层	编码前	编码后		
		楼层_低	楼层_中	楼层_高
取值	低	1	0	0
	中	0	1	0
	高	0	0	1



通过对各分类变量进行独热编码, 最终将 26 个特征变量扩充为 51 个。

### 6) 描述性统计分析

经过上述处理后, 本文对数据特征进行统计分析, 获得不同特征对应的样本数量分布情况, 如图 5 所示。可以看出, 区域和楼层特征的样本数量分布较为均匀, 而朝向和租期特征的样本数量分布差异较大, 这也从侧面反映了市场中大多数人的租房偏好。

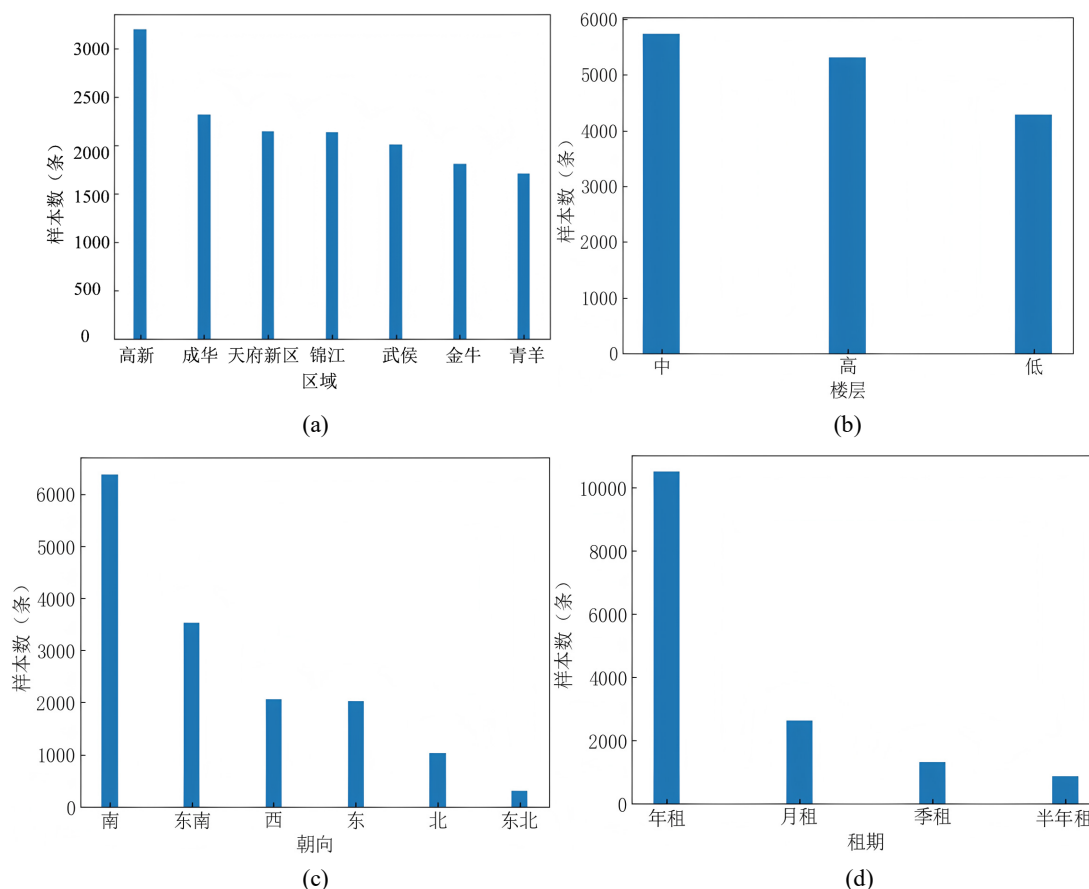


Figure 5. Distribution of sample size corresponding to some features

图 5. 部分特征对应的样本数量分布情况

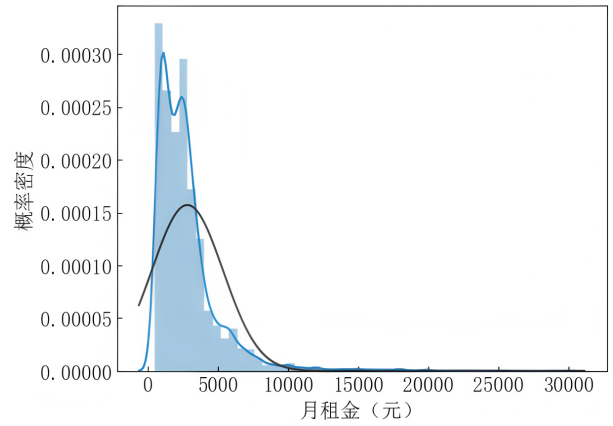
根据图 4 的月租金分布图可以观察到, 月租金数据可能呈现正态分布。对月租金数据进行正态分布曲线的拟合得到月租金核密度图, 如图 6 所示。可以看出, 月租金数据存在明显高峰左偏, 而长尾向右方延伸的情况。针对这一现象, 本文对月租金数据进行对数处理, 处理结果如图 7 所示。

由图 7 可以看出, 经过对数处理后的月租金数据的拟合曲线呈现出明显的双峰分布趋势, 这种趋势可能对模型的泛化能力产生不利影响。具体来说, 模型在训练过程中可能会过于关注某一类样本而忽略另一类样本, 进而引入潜在的偏见。因此在训练模型时, 需要特别关注并解决这种数据分布的不均匀性, 以确保模型能够全面、准确地捕捉数据的特征。

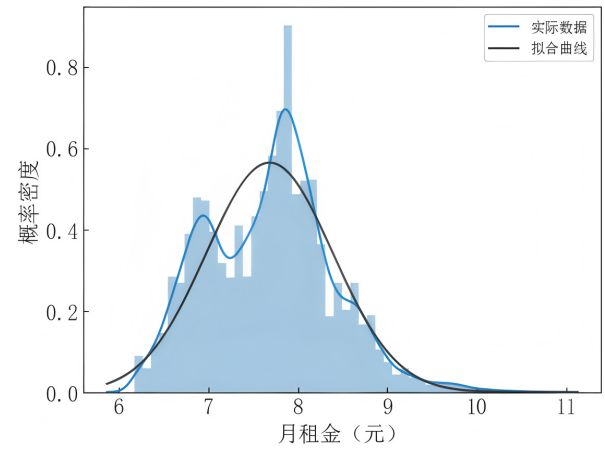
为解决上述问题, 本文采用高斯混合模型(Gaussian Mixture Model) [15]对数据进行拟合并将其分解成两个单峰曲线。如图 8 所示, 经过对数处理后的月租金数据通过高斯混合模型被有效地拆分成了两个数据集。图中分界线对应的对数值为 7.124, 将其还原为原始的月租金数据对应的租金约为 1240.4 元。因此, 根据这一标准, 数据集被明确地划分为了月租金低于 1240.4 元的数据集 1 和高于 1240.4 元的数据集



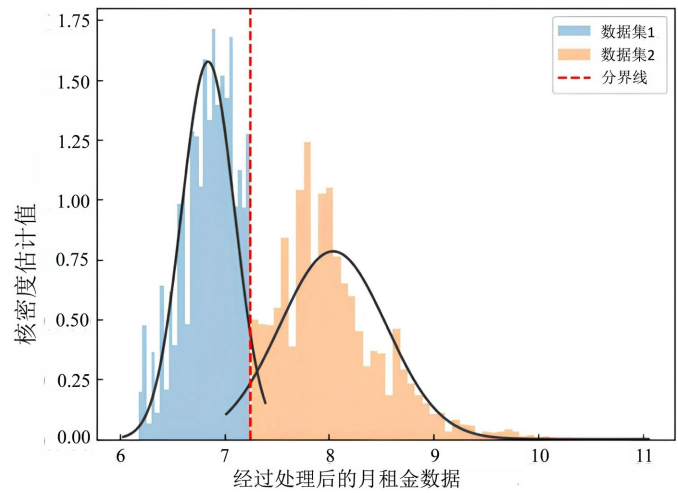
2. 这种分类有助于更精细地分析不同租金水平下的数据特性。



**Figure 6.** Kernel density plot for monthly rent  
**图 6.** 月租金核密度图



**Figure 7.** Kernel density plot of monthly rent after logarithmic transformation  
**图 7.** 对数处理后的月租金核密度图



**Figure 8.** Splitting the dataset into two Gaussian-distributed datasets using a Gaussian mixture model  
**图 8.** 使用高斯混合模型将数据集拆分为两个符合高斯分布的数据集

### 3.2.3. 特征工程

特征工程是机器学习领域中至关重要的一环, 通过构建新的特征或对现有特征进行改造和筛选, 使得机器学习算法能够更深入地理解数据的内在规律和结构, 从而更有效地提取出有价值的信息。

本文选取互信息法对影响租金的各个特征进行过滤筛选, 剔除与租金关联性较弱或完全无关的特征, 从而进一步增强模型的解释性, 并有效防止因模型复杂度过高而引发的过拟合现象。对于两个离散型随机变量  $X$  和  $Y$ , 其互信息值可由式(2)表示:

$$I(X, Y) = \sum_{y \in Y} \sum_{x \in X} p(x, y) \log \left( \frac{p(x, y)}{p(x)p(y)} \right) \quad (2)$$

式中,  $p(x, y)$  表示  $X$  和  $Y$  的联合概率分布,  $p(x)$  和  $p(y)$  分别是  $X$  和  $Y$  的边缘概率分布。

本文选取 0.1 作为互信息值阈值, 即保留互信息值大于 0.1 的特征。经筛选之后得到的特征共 16 个, 具体特征及其互信息值如表 4 所示。

**Table 4.** Selected features and mutual information values

**表 4.** 经筛选后的特征及其互信息值

特征名称	互信息值
面积	0.4256
区域_武侯区	0.4125
区域_成华区	0.3620
朝向_南	0.2569
区域_金牛区	0.2456
方式_合租	0.2412
方式_整租	0.2358
朝向_东南	0.2212
区域_天府新区	0.2132
用水_民水	0.1256
采暖_集中供暖	0.1136
用电_民电	0.1123
区域_高新区	0.1121
租期_年租	0.1120
租期_月租	0.1023
楼层_高	0.1001

由表 4 可知, 房屋面积、所在行政区域、朝向以及租期是影响租金的关键因素, 这些因素共同决定了租金水平, 为房租预测提供了重要参考。

### 3.3. 评价指标

本文采用平均绝对百分比误差(MAPE)、均方根误差(RMSE)以及决定系数( $R^2$ )来综合度量预测模型的性能。

1) MAPE 反映了模型预测值与真实值之间的平均相对误差大小, 计算如式(3)所示。

$$MAPE = \frac{1}{m} \sum_{i=0}^m \left| \frac{\hat{y}_i - y_i}{y_i} \right| \quad (3)$$

式中,  $y_i$  为真实值,  $\hat{y}_i$  为预测值,  $m$  为样本总数量。

2) RMSE 反映了模型预测值与真实值之间的偏差程度, 计算如式(4)所示。

$$RMSE = \sqrt{\frac{1}{m} \sum_{i=1}^m (y_i - \hat{y}_i)^2} \quad (4)$$

式中,  $y_i$  为真实值,  $\hat{y}_i$  为预测值,  $m$  为样本总数量。

3)  $R^2$  反映了模型预测值与真实值之间的相关程度, 计算如式(5)所示。

$$R^2 = 1 - \frac{\sum_{i=1}^m (\hat{y}_i - y_i)^2}{\sum_{i=1}^m (y_i - \bar{y})^2} \quad (5)$$

式中,  $y_i$  为真实值,  $\hat{y}_i$  为预测值,  $\bar{y}$  为样本真实值的平均值,  $m$  为样本总数量。

3.4. 模型的建立

Table 5. Optimal parameter combination for random forest algorithm  
表 5. 随机森林算法最优参数组合

参数名称	参数取值
n_estimators	1000
max_depth	26
min_samples_leaf	2
min_samples_split	2
max_feature	'auto'

Table 6. Optimal parameter combination for XGBoost algorithm  
表 6. XGBoost 算法最优参数组合

参数名称	参数取值
n_estimators	900
max_depth	8
min_child_weight	1
gamma	0.4
subsample	1
colsample_bytree	1
reg_lambda	1
reg_alpha	69
learning_rate	0.3

根据第 2 节介绍的三种机器学习算法的建模流程, 分别建立基于三种算法的住房租金预测模型。模型参数的选择对模型的预测性能有着较大的影响, 故需要对三种预测模型的参数进行调优。对于随机森林算法和 XGBoost 算法, 本文采用先粗调(Coarse to Fine)后细调(Fine-Tuning)的策略, 先大范围地搜索参

数空间找到大致的最佳参数,再通过网格搜索(Grid Search)来搜索参数空间中的最佳组合。具体地:首先,确定每个参数的候选取值范围;然后,对所有可能的参数组合进行交叉验证,找到在训练集上性能最优的参数组合。基于随机森林算法和 XGBoost 算法的两种租金预测模型的最优参数组合分别如表 5 和表 6 所示。

基于 BP 神经网络算法的租金预测模型,本文选取的超参数如下:

① 隐层神经元节点数,本文按照经验公式  $n_c = \frac{n_i n_o + \frac{1}{2} n_o (n_i^2 + n_i) - 1}{n_i + n_o}$  确定  $n_c = 3$ , 其中  $n_i$  为输入节点数,  $n_o$  为输出节点数。

② 激活函数,本文采用 Sigmoid 函数。

③ 训练函数,本文采用基于 Levenberg-Marquardt 算法的训练函数。

④ 期望误差,本文选取为 0.0001。

#### 4. 预测结果分析

通过上述的分析与处理,本文已将原始数据集拆分为两个独立且符合高斯分布的数据子集。其中,数据集 1 包含月租金在 1240.4 元以下的数据,共计 3929 条样本;数据集 2 则包含月租金在 1240.4 元以上的数据,共计 12,029 条样本。两个数据集的特征数量均为 16 个。随后,按照 8:2 的比例将这两个数据集分别随机划分为训练集和测试集,以便进行后续的模型训练和性能评估。

确定三种预测模型的相关参数后,采用两个数据集分别对模型进行 5 折交叉验证评估,评价指标结果如表 7 所示。

**Table 7.** Comparison of performance of three rent prediction models based on different algorithms

**表 7.** 基于不同算法的三种租金预测模型性能对比

		RMSE	MAPE	R <sup>2</sup>
随机森林算法	测试集 1	0.1141	2.2213	0.7654
	测试集 2	0.0956	1.9568	0.9568
XGBoost 算法	测试集 1	0.0958	0.5478	0.8864
	测试集 2	0.0356	0.7568	0.9963
BP 神经网络算法	测试集 1	0.1145	2.5684	0.7102
	测试集 2	0.0859	1.8759	0.7256

由表 7 分析可知,三种租金预测模型在数据集 2 上的表现均优于数据集 1,这可能是由于数据集 2 的样本数量远大于数据集 1 所导致。此外,基于 XGBoost 算法的租金预测模型在两个数据集上的表现均优于另外两个预测模型,具体为 RMSE、MAPE 数值相对较小,表明模型预测误差更小;R<sup>2</sup> 接近于 1,表明模型泛化能力更强。

#### 5. 结论

本文采用机器学习算法对成都市多个区域的住房租金进行了影响因素分析和预测研究。首先,通过爬虫技术获取了较长时间跨度和较广空间跨度的房屋出租数据,尽可能全面地考虑了对租金有影响的因素,丰富了数据集特征。其次,通过对原始数据集的预处理、特征工程的优化,构建了有效的数据集,为模型训练提供了可靠的数据基础;随后,分别建立了基于随机森林、XGBoost、BP 神经网络算法的三种住房租金预测模型,并采用网格寻优方法对模型参数进行了调优;最后,采用交叉验证法验证了预测模

型的准确性和泛化能力, 确保了模型在不同数据集上的鲁棒性。研究结果表明, 房屋面积、所在行政区域、朝向以及租期是影响租金的关键因素, 这些特征的识别为租金预测和定价决策提供了重要参考。此外, 基于 XGBoost 算法的预测模型的预测性能优于其他两种预测模型。

## 基金项目

四川省科技计划资助(2024NSFSC0940)和中央引导地方科技发展项目资助(2024ZYD0133)资助。

## 参考文献

- [1] Marks, D. (1984) The Effect of Rent Control on the Price of Rental Housing: An Hedonic Approach. *Land Economics*, **60**, 81-94. <https://doi.org/10.2307/3146095>
- [2] Guntermann, K.L. and Norrbin, S. (1987) Explaining the Variability of Apartment Rents. *Real Estate Economics*, **15**, 321-340. <https://doi.org/10.1111/1540-6229.00435>
- [3] Sirmans, S., Sirmans, C. and Benjamin, J. (1989) Determining Apartment Rent: The Value of Amenities, Services and External Factors. *Journal of Real Estate Research*, **4**, 33-43. <https://doi.org/10.1080/10835547.1989.12090581>
- [4] 张若曦, 贾士军. 广州市住宅租金影响因素的研究[J]. 工程管理学报, 2014, 28(6): 118-123.
- [5] 陆军, 王志文, 张骥. 公共服务设施对北京市城区租赁住宅租金的影响[J]. 价格月刊, 2017(7): 7-12.
- [6] 王世桢, 江莉, 李齐霖. 基于 ARDL 模型的上海市住房租金影响因素分析[J]. 上海房地, 2021(4): 22-25.
- [7] Antipov, E.A. and Pokryshevskaya, E.B. (2012) Mass Appraisal of Residential Apartments: An Application of Random Forest for Valuation and a Cart-Based Approach for Model Diagnostics. *Expert Systems with Applications*, **39**, 1772-1778. <https://doi.org/10.1016/j.eswa.2011.08.077>
- [8] Sarip, A.G., Hafez, M.B. and Daud, M.N. (2016) Application of Fuzzy Regression Model for Real Estate Price Prediction. *Malaysian Journal of Computer Science*, **29**, 15-27. <https://doi.org/10.22452/mjcs.vol29no1.2>
- [9] Li, J.Z. (2018) Monthly Housing Rent Forecast Based on LightGBM (Light Gradient Boosting) Model. *International Journal of Intelligent Information and Management Science*, **7**, 3-9.
- [10] 谢勇, 项薇, 季孟忠, 等. 基于 XGBoost 和 LightGBM 算法预测住房月租金的应用分析[J]. 计算机应用与软件, 2019, 36(9): 151-155, 191.
- [11] Breiman, L. (2001) Random Forests. *Machine Learning*, **45**, 5-32. <https://doi.org/10.1023/a:1010933404324>
- [12] Chen, T. and Guestrin, C. (2016) XGBoost: A Scalable Tree Boosting System. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, San Francisco, 13-17 August 2016, 785-794. <https://doi.org/10.1145/2939672.2939785>
- [13] Faraway, J. and Chatfield, C. (1998) Time Series Forecasting with Neural Networks: A Comparative Study Using the Air Line Data. *Journal of the Royal Statistical Society Series C: Applied Statistics*, **47**, 231-250. <https://doi.org/10.1111/1467-9876.00109>
- [14] 梁杰, 陈嘉豪, 张雪芹, 等. 基于独热编码和卷积神经网络的异常检测[J]. 清华大学学报(自然科学版). 2019, 59(7): 523-529.
- [15] 朱映秋, 黄丹阳, 张波. 基于高斯混合模型的分布因子聚类方法[J]. 统计研究, 2024, 41(6): 147-160.