Computer Science and Application 计算机科学与应用, 2025, 15(5), 766-770
Published Online May 2025 in Hans. https://www.hanspub.org/journal/csa
https://doi.org/10.12677/csa.2025.155149

Hans汉斯

机器学习驱动的二手车价格预测方法研究

黄子敬

宁波工程学院统计与数据科学学院，浙江 宁波

收稿日期：2025年4月25日；录用日期：2025年5月23日；发布日期：2025年5月30日

摘　要

本文研究了利用机器学习方法预测二手车价格。作者使用了来自和鲸社区的二手车交易数据集，并构建了Adaboost、Bagging和LightGBM三种模型进行预测。通过MAPE和R^2两个指标评估模型性能，结果显示Bagging模型表现最佳，MAPE为0.1180，R^2为0.9027。

关键词

机器学习，二手车价格，Bagging

Research on Machine Learning-Driven Used Car Price Prediction Method

Zijing Huang

College of Statistics and Data Science, Ningbo University of Technology, Ningbo Zhejiang

Received: Apr. 25th, 2025; accepted: May 23rd, 2025; published: May 30th, 2025

Abstract

This paper studies the prediction of second-hand car prices using machine learning methods. The author utilized the used car transaction dataset from the Hejing community and constructed three models, namely Adaboost, Bagging, and LightGBM, for prediction. The performance of the model was evaluated through two indicators, MAPE and R^2. The results showed that the Bagging model performed the best, with an MAPE of 0.1180 and a R^2 of 0.9027.

Keywords

Machine Learning, Second-Hand Car Price, Bagging

文章引用: 黄子敬. 机器学习驱动的二手车价格预测方法研究[J]. 计算机科学与应用, 2025, 15(5): 766-770.
DOI: 10.12677/csa.2025.155149

1. 引言

随着我国经济社会的快速发展和消费能力的提高，消费结构也呈现出明显的转变和提高趋势。这种结构变化不仅反映在对物质需求的增加上，也反映在消费者概念的更新上。在汽车销售方面，消费者比起一辆新车的价值，更注重成本效率和车辆使用。这种心态的变化使二手车市场打破了传统知识的局限性，在汽车消费选择中变得越来越重要。与此同时，随着城市化进程的加快和人民生活质量的提高，公众对舒适旅行的期待也在不断提高。对汽车需求和消费水平提高的影响，最初通过打破边际状态，逐渐变成了二手车市场值得进一步研究的经济现象。在这种背景下，二手车市场逐渐成为学术界和业界的焦点。

跟随人工智能技术的蓬勃发展，二手车价格预测领域也深度融入了诸多机器学习技术。廖亚茹[1]结合特征加权利用 Stacking 集成模型做了二手车价格预测研究。任武彬[2]依托公开数据集构建了二手车价格预测模型，创新性地采用 MAE (平均绝对误差)作为模型性能评估的核心指标，该研究通过量化预测误差为二手车定价机制提供了更具解释力的评估维度。为提升二手车交易平台的运营效率，代金辉等[3]研究者采用 LightGBM 算法构建二手车价格预测模型，同步运用随机森林算法对成交周期的关键影响因素进行重要性排序，通过双模型协同机制为交易决策提供支持。胡诣文等[4]聚焦二手车价值评估这一核心议题，创新性地开发了基于机器学习的二手车价格预测模型，为车辆残值评估提供了量化决策工具。杨致远[5]依托二手车交易平台采集的真实交易数据，创新性地融合 PCA-DNN 深度神经网络与 LightGBM 集成学习算法，构建了双模型耦合的二手车成交价格预测系统，实现了对车辆最终成交价格的精准预测。

2. 数据介绍

本研究采用的实验数据集源自和鲸社区发布的二手车价格数据库，该数据集共包含 205 个有效样本，覆盖了各种二手车交易中的关键信息。其特征矩阵由 25 列变量构成，包含 symboling、CarName、fueltype 等特征，其中既包含连续型数值特征，也涉及需要通过编码转换处理的 object 类型分类特征。另外数据还包含 1 列预测标签 price。并且所有样本中均不存在缺失值或异常值，为后续的模型训练提供了干净规范的数据基础。

3. 算法简介

3.1. Adaboost 算法

AdaBoost 回归算法提出了解决回归问题的一个综合性的学习方法，基本思想是多次训练基本回归模型，然后根据预测错误的模型结果调节样本，最后将各个模型得到的预测结果综合得到最终的模型预测结果。该回归算法通过引入大量的预测误差样本，逐步改变模型的聚焦方向，提高准确度。实践表明，Adaboost 回归预测算法可以有效地捕获非线性关系和复杂的数据分布，为回归预测提供了新的方法。

3.2. Bagging 算法

Bagging 回归算法利用自助取样技术构建了一个集成学习框架，其基本概念是通过重复取样创建不同的训练子集，并在每个子集上独立地对基础模型进行训练，使之最终回归到一个模型上。在最终预测

时，Bagging 回归算法将所有模型的输出结果进行平均。其能够在提高预测结果稳定性和可靠性的同时，得到降低单一模型预测方差的最终预测结果。该算法凭借其出色的抗过拟合能力和强大的泛化性能，在各种复杂的回归预测任务中显示出巨大的优势，尤其适用于处理具有高维度的数据空间和包含较多噪声干扰的应用场景，成为增强回归预测模型稳定性的重要技术手段之一。

3.3. LightGBM 算法

LightGBM 是基于梯度提升决策树(Gridge Development Tree, GBDT)框架构建的高效回归模型，这种算法通过改善决策树的构建和分割，有效提高计算效率，降低内存消耗，确保预测的准确性，特别是对大规模数据处理场景的预测更为适用。LightGBM 算法在其训练模式下反复生成多棵树，每棵树都专注于学习之前模型无法捕捉到的数据，最终将所有树的预测结果整合在一起，然后再进行精准预测。LightGBM 算法在复杂的回归预测领域，如金融领域，已经成为解决非线性回归问题的强大工具之一。

4. 机器学习驱动的二手车价格预测

4.1. 模型构建

在本文里，首先对数据进行了预处理，删除了无关建模列 car_ID，对存在的 object 数据进行编码处理，将预测标签 price 与建模特征进行数据分离，将数据按 7:3 的比例分成了训练集和测试集，训练集用来训练模型，测试集用来评估模型的好坏。随后构建了 Adaboost、Bagging 和 LightGBM 二手车价格预测模型。Adaboost 模型设置了 random_state = 2025、n_estimators = 100、learning_rate = 0.1 这三个参数。Bagging 模型设置了 random_state = 2025、n_estimators = 100 这两个参数。LightGBM 设置了 random_state = 2025、learning_rate = 0.35、n_estimators = 100、max_depth = 8 这四个参数。最后输入划分好的训练集数据进行模型训练，利用测试集数据进行模型比较。

4.2. 模型评估

本研究采用 MAPE (平均绝对百分比误差)和 R^2 (决定系数)两种模型评估指标，对构建的 Adaboost 模型、Bagging 模型和 LightGBM 模型进行对比分析。三种二手车价格预测模型的两种指标如表 1 所示。

Table 1. Comparison of results of used car price prediction models
表 1. 二手车价格预测模型结果比较

	MAPE	R^2
Adaboost	0.1299	0.8782
Bagging	0.1180	0.9027
LightGBM	0.1507	0.8571

通过系统对比三个候选模型在测试集上的多维度性能指标，可以清晰看出，Bagging 集成模型凭借其优异的预测性能脱颖而出，被确立为本研究的最优预测方案。该模型在 MAPE (平均绝对百分比误差)关键指标上以 0.1180 的优异值位居首位，充分展现了其在误差控制方面的卓越表现；同时在 R^2 (决定系数)指标上以 0.8571 的高分领先其他模型，表明其对二手车价格变动的解释能力达到 90.27%的强解释力水平。综合各项指标评估结果，Bagging 模型不仅在本研究构建的评估体系中全面领先，更在实际预测场景中展现出卓越的实用价值，成为二手车价格预测领域具有推广意义的技术方案。Adaboost 模型、Bagging 模型和 LightGBM 模型的预测结果可视化如图 1，图 2，图 3 所示。

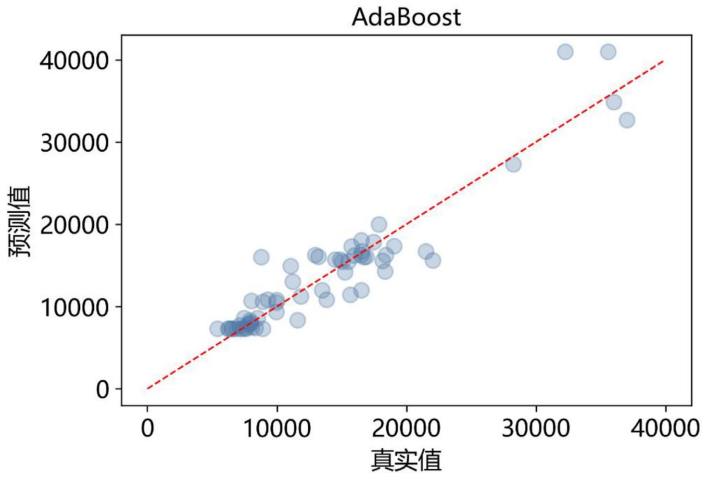

Figure 1. Visualization of the prediction results of the Adaboost model test set
图 1. Adaboost 模型测试集预测结果可视化

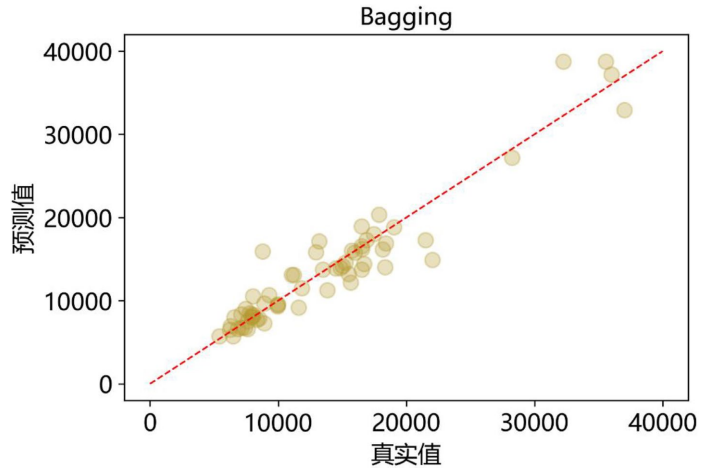

Figure 2. Visualization of the prediction results of the Bagging model test set
图 2. Bagging 模型测试集预测结果可视化

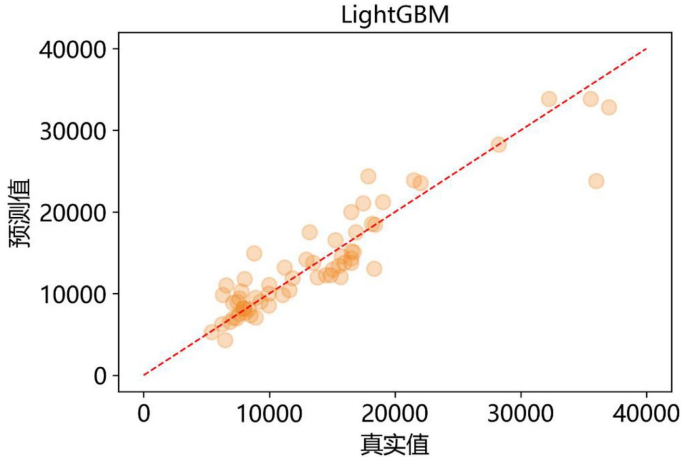

Figure 3. Visualization of the prediction results of the LightGBM model test set
图 3. LightGBM 模型测试集预测结果可视化

5. 结语

本研究围绕二手车价格预测这一核心问题，创新性地进行了一项系统性研究。首先，将二手车数据中与建模无关的列进行删除，并对数据中的类别型特征进行编码处理，从而实现了对二手车价格交易数据的预处理。随后，分别构建了 Adaboost 预测模型、Bagging 预测模型和 LightGBM 预测模型，并对三种模型的预测结果进行了比较分析。最后使用 MAPE (平均绝对百分比误差)和 R^2 (决定系数)两种评估指标，对三种二手车价格预测模型的预测结果进行比较。实验结果表明，Bagging 集成模型是本文最好的二手车价格预测模型，其 MAPE 值低至 0.1180，R^2 高达 0.9027。该成果不仅为二手车定价方案的制定提供了有效的参考依据，同时也为高效整合机器学习技术与二手车销售行业提供了新的思路，具有前瞻性的研究价值和实践指导意义。

参考文献

[1] 廖亚茹. 基于特征加权 Stacking 集成模型的二手车交易价格预测研究[D]: [硕士学位论文]. 武汉: 华中师范大学, 2024.

[2] 任武彬. 基于机器学习方法的二手车价格预测[D]: [硕士学位论文]. 湘潭: 湘潭大学, 2023.

[3] 代金辉, 仲璇, 王梦恩. 基于 LightGBM 和随机森林算法的二手车估价[J]. 高师理科学刊, 2022, 42(12): 15-22.

[4] 胡诣文, 张天佑, 张旭, 等. 基于机器学习的二手车价格预测算法研究[J]. 信息技术与信息化, 2022(10): 52-55.

[5] 杨致远. 基于 PCA-DNN 和 LightGBM 的二手车价格预测[J]. 信息与电脑(理论版), 2022, 34(21): 73-75.