

面向电商领域的 AI 驱动 RPA：汽车价格预测与销售预测

徐如意¹ 袁冠聪² 谢金荣¹ 张秀婉¹ 尚玥¹ 陈兰兰¹

1. 广州城建职业学院, 中国·广东 广州 510925

2. 广州南方学院, 中国·广东 广州 510970

摘要：电子商务行业的迅猛发展已显著影响汽车市场，促使企业亟需精准的价格预测模型，以优化供应商的定价策略。本研究创新性地融合了人工智能（AI）与机器人流程自动化（RPA），旨在提升汽车价格预测的准确性和效率。其中，RPA 被用于自动从线上平台抓取车辆价格及销售数据，有效克服了反爬虫机制带来的挑战；与此同时，基于 AI 的回归模型则深入分析历史数据及发动机排量、整备质量、马力和车辆尺寸等关键特征，从而精准预测未来价格走势。通过对比决策树、支持向量机以及深度学习模型等多种机器学习算法，本研究重点突出了回归方法在兼顾预测精度与模型可解释性方面的独特优势。研究结果表明，AI 与 RPA 的协同应用不仅能够简化数据获取流程、提高预测准确性，还能为供应商提供基于数据驱动的定价建议。这一研究成果有助于推动汽车电商领域的智能化转型，为希望在动态市场中优化定价策略的企业提供了可扩展的解决方案。

关键词：人工智能；机器人处理；自动化；汽车价格预测；电子商务；机器学习；回归分析

AI-driven RPA for E-commerce: Car Price Prediction and Sales Forecasting

Xu Ruyi¹, Yuan Guancong², Xie Jinrong¹, Zhang Xiuwan¹, Shang Yue¹, Chen Lanlan¹

1. Guangzhou City Construction Vocational College, China Guangdong Guangzhou 510925

2. Guangzhou Nanfang College, China Guangdong Guangzhou 510970

Abstract: The rapid development of the e-commerce industry has significantly impacted the automotive market, compelling enterprises to urgently seek accurate price prediction models to optimize suppliers' pricing strategies. This study innovatively integrates artificial intelligence (AI) and robotic process automation (RPA) to enhance the accuracy and efficiency of automotive price prediction. Specifically, RPA is employed to automatically extract vehicle prices and sales data from online platforms, effectively overcoming the challenges posed by anti-crawling mechanisms. Meanwhile, an AI-based regression model is utilized to deeply analyze historical data and key features such as engine displacement, curb weight, horsepower, and vehicle size, thereby precisely forecasting future price trends. By comparing various machine learning algorithms including decision trees, support vector machines, and deep learning models, this study highlights the unique advantages of regression methods in balancing prediction accuracy and model interpretability. The research results demonstrate that the collaborative application of AI and RPA not only simplifies the data acquisition process and improves prediction accuracy but also provides data-driven pricing suggestions for suppliers. This research outcome contributes to the intelligent transformation of the automotive e-commerce sector and offers scalable solutions for enterprises aiming to optimize their pricing strategies in a dynamic market.

Keywords: Artificial intelligence; Robot processing; Automation; Car price prediction; E-commerce; Machine learning; Regression analysis

0 引言

电子商务的迅猛发展重塑了汽车销售模式，使线上平台成为消费者获取车辆信息、比价及完成购车决策的主要渠道。然而，受市场需求、品牌价值和折旧率等多种因素影响，定价仍是一项重大挑战。供应商亟须制定有效的定

价策略，以确保利润空间并保持市场竞争力。人工智能与 RPA 技术为此提供了极具前景的解决方案——通过自动化数据采集和实现预测分析，帮助供应商基于实时市场趋势优化定价决策。

人工智能和 RPA 在推动电子商务运营现代化方面发

挥着关键作用，它们通过增强数据收集与分析能力来实现这一目标。RPA 能够自动从各类平台中提取关键信息，如汽车价格、销售数据及市场趋势，从而简化数据收集流程并提升准确性。而 AI 则利用历史数据预测未来价格波动，为供应商提供有关潜在利润空间的宝贵洞察。此外，通过整合人工智能与 RPA 技术，供应商能够动态分析竞争对手的定价策略，从而做出更明智、更具竞争力的定价决策。

尽管取得这些进展，汽车电商行业仍面临诸多挑战，尤其是在数据获取方面——由于信息分散于多个平台，加之严格的反网络爬虫措施，企业难以高效采集所需数据。目前，各大网站纷纷采用 CAPTCHA 验证码、IP 封禁以及动态内容渲染等技术，以阻止自动化工具的非法数据提取。为突破这些障碍，本研究将创新性地引入动态 IP 轮换、User-Agent 切换及无头浏览器模拟等先进技术，确保数据采集过程既合法合规，又高效精准。同时，研究还将结合 RPA 实现自动化数据获取，并与 AI 驱动的预测分析深度融合，最终帮助供应商在不断变化的电商环境中，以更科学的方式优化价格策略，提升市场竞争力。

1 文献综述

人工智能 (AI) 通过利用回归分析、决策树及深度学习网络等机器学习模型，深入分析关键定价因素，已在提升汽车价格预测能力方面发挥了至关重要的作用。^[1] 研究表明，发动机尺寸、整备质量及马力等特征对价格预测具有显著影响^[2]。此外，将人工智能与大数据、云计算技术相结合，能够进一步提升预测精度，实现基于市场趋势的实时定价调整^[3]。与此同时，机器人流程自动化 (RPA) 可自动从电商平台采集数据，简化汽车价格数据的提取与处理流程^[4]。研究者已证实，RPA 在大规模数据提取任务中表现出色，并能无缝对接云端分析平台，从而支持实时价格预测^[5]。尽管 RPA 具备诸多优势，但针对电商平台的价格数据采集，其网络爬取工作仍面临重重挑战，这主要源于电商网站所采取的反爬虫措施，如 CAPTCHA 验证、IP 封禁以及动态内容渲染等^[6]。为突破这些障碍，研究人员已采用动态 IP 轮换等先进策略，浏览器自动化与基于 API 的数据检索^[7]。伦理与法律考量同样至关重要，因为未经授权的数据抓取可能违反平台政策^[8]。在预测模型方面，线性回归因其透明性和可解释性，仍然是最常用的技术之一^[9]。而其他模型，如决策树以及随机森林等集成学习方法，则能提供更高的预测精度，但需要仔细调参以避免过拟合^[10]。

人工智能与 RPA 的融合，为汽车价格的自动化预测提供了强大的框架，使企业能够实现数据提取与预测建模的全流程自动化^[11]。研究表明，RPA 能够高效地从多个平台收集数据，随后利用 AI 模型对这些数据进行分析，以预

测价格趋势^[12]。此外，有研究指出，将机器学习与 UiPath 和 Power Automate 等 RPA 工具相结合，能够进一步提升自动化水平，并优化决策流程^[13]。文献综述强调了人工智能驱动的预测分析和 RPA 驱动的数据采集的高效性，为本研究聚焦于基于回归的建模与自动化数据提取奠定了坚实基础^[14]。通过利用这些技术，供应商能够优化定价策略、提升利润率，并在汽车电商市场中赢得竞争优势^[15]。

表1 机器学习算法在价格预测中的对比分析

算法	优势	劣势	在研究中的用例
线性回归	简单、可解释性强、训练速度快	假设线性关系、存在过拟合风险	用作价格预测的基线模型
随机森林	处理非线性关系、对异常值鲁棒	参数调优复杂、计算成本较高	在价格预测中实现最佳性能
XGBoost	精度高、支持缺失值处理、高效	计算开销大、对噪声敏感	测试非线性特征交互的作用
支持向量机 (SVM)	适用于高维数据、泛化能力强	计算开销大、对噪声敏感	测试非线性特征交互的作用
神经网络 (DNN)	捕捉复杂模式、扩展性强	需要大量数据、可解释性差	探索基于深度学习的预测方案

2 方法论

本研究整合了 UiPath (RPA) 与机器学习 (ML) 技术，旨在为电子商务应用实现汽车价格预测的自动化。研究方法包含三个主要环节：利用 UiPath 进行数据采集、数据预处理及特征选择，以及基于机器学习的预测建模。此外，研究还通过结构化的工作流程，评估了 RPA 驱动的数据提取与机器学习辅助的价格预测效果。最后，本文从预测精度、数据一致性以及自动化效率三个方面，对所提出的系统性能进行了全面评估。

2.1 系统 workflow 设计

该系统的工作流程旨在实现数据提取、预处理及预测建模的自动化。通过 UiPath，RPA 组件可从多个电子商务网站抓取与汽车相关的信息，包括发动机尺寸、整备质量、马力以及燃油效率等。抓取后的数据将以结构化格式存储，确保后续分析的一致性。随后，机器学习模型将对这些数据进行处理，生成价格预测结果。整个工作流程完全自动运行：RPA 负责数据的获取与预处理，而机器学习模型则执行价格预测任务。这一自动化流水线最大限度地减少了人工干预，从而保障了预测过程的高效性和准确性。

2.2 数据集与数据源

本研究使用的数据集由 UiPath 从在线汽车销售平台采集而来，包含发动机尺寸、整备质量、马力、燃油效率、宽度和长度等结构化数据点。为全面捕捉各类车型的定价

趋势，数据集中同时涵盖了新车与二手车信息。此外，采集的数据经过严格验证，以消除不一致之处并确保其完整性。为进一步提升预测精度，研究还整合了历史销售数据及市场趋势信息。得益于数据集的多样性，模型能够有效泛化至不同类别和价格区间的汽车产品。

2.3 机器学习模型实现

该预测模型采用多种机器学习算法，以评估其在价格预测中的有效性。其中，线性回归因其易于解释且适用于数值预测，成为主要方法。此外，研究还测试了基于决策树的模型，如随机森林和梯度提升，以捕捉汽车属性与定价之间复杂的非线性关系。模型训练过程包括特征选择、超参数调优以及交叉验证，旨在优化整体性能。通过对比不同算法，本研究确定了最适合电商场景下价格预测的有效方法。

2.4 模型评估方法论

为了评估用于预测汽车价格的不同回归模型的性能，我们采用了若干关键的评价指标。这些指标通过将预测值与实际值进行比较，帮助确定预测模型的准确性和可靠性。本次评估所涉及的模型包括：简单线性回归，简单非线性回归，多元线性回归，多重非线性回归。这些模型的性能主要通过 R^2 （决定系数）作为评估指标进行衡量。

决定系数 (R^2 分数) R^2 分数用于衡量因变量（汽车价格）中，由自变量所解释的方差比例。自变量（如发动机尺寸、整备质量及马力等特征）。其计算公式为：

$$R^2 = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y})^2}$$

y_i = 实际购车价格

\hat{y}_i = 预测的汽车价格

\bar{y} = 实际汽车价格的平均值

更高的 R 数值越大，表明模型性能越好，因为自变量能够解释汽车价格中更多的方差。

而 R 提供了对所解释方差比例的洞察，但还需采用其他基于误差的指标来评估预测精度和模型可靠性。

平均绝对误差 (MAE)

MAE 衡量预测值与实际值之间误差的平均绝对值，其计算公式为：

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

较低的平均绝对误差 (MAE) 值表明模型的精度更高，因为这反映了平均预测误差更小。

均方根误差 (RMSE)。RMSE 对较大误差的惩罚力

度高于 MAE，因此它非常适合用于检测预测中的显著偏差。其计算公式为：

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

较低的 RMSE 值意味着更高的预测精度以及模型误差的更低方差。

模型性能比较方法：为确保客观的性能评估，所有模型均在相同条件下，使用同一数据集和输入特征进行评估。具体方法步骤如下：数据准备：对数据集进行预处理，确保输入变量（如发动机尺寸、整备质量、马力及其他相关特征）的一致性。模型训练：每个回归模型均基于数据集的一部分进行训练，同时在适用情况下保持统一的超参数设置。最后，通过比较各模型的评估得分，确定最有效的汽车价格预测。

2.5 方法论概述

本章概述了评估用于汽车价格预测的回归模型的系统化方法。该评估依赖于 R 为了衡量方差解释度，我们采用 MAE 和 RMSE 来量化预测误差。标准化的评估流程确保客观模型比较，指导选择最有效的预测模型。

3 原型开发

开发用于汽车价格预测的自动化 AI-RPA 系统，需要采用结构化方法，整合数据采集、机器学习模型以及基于 RPA 的自动化流程。该系统利用 UiPath 实现自动化的数据提取，同时通过线性与非线性回归模型开展预测分析，以准确预测汽车价格。本节将详细阐述系统架构，涵盖工作流自动化、数据预处理、模型集成，以及评估方法，确保预测结果可靠且具备良好的可扩展性。

3.1 系统架构与工作流程

该系统作为一个端到端的自动化工作流，能够自动采集汽车市场数据，处理关键车辆属性，应用预测模型，并以结构化格式输出预估的汽车价格。如图 X 所示，该工作流由多个顺序执行的模块组成：首先通过 RPA 自动化技术获取数据，接着进行数据预处理，随后开展模型预测，最后完成评估与报告生成。

该过程首先由 UiPath 执行网络爬取自动化，以提取汽车列表数据，包括发动机尺寸、整备质量、马力、高速公路燃油效率、宽度和长度等属性。这些提取的数值将作为汽车价格估算的输入变量。系统设计时充分考虑了多种数据格式的兼容性，如 CSV、JSON 和 Excel，从而确保能够顺畅对接多个数据源。

一旦原始数据被检索，预处理阶段包括：通过应用插补技术处理缺失值。对特征进行缩放和转换，以确保变量量级的一致性。数据归一化以提升模型效率。经过预处理

理后，数据被输入到预测模型中，包括简单线性回归、多元线性回归、简单非线性回归以及多元非线性回归。这些模型基于车辆属性与市场定价之间学习到的关系，生成价格估算结果。为提升预测可靠性，可采用诸如 R 等模型评估指 2，计算了平均绝对误差（MAE）和均方根误差（RMSE）。如果检测到任何数据不一致或极端异常值，系统会将其标记出来，以便进一步审查或调整。最后，预测的汽车价格与评估结果以 ExtractedDT 数据表的形式存储，并导出至 Microsoft Excel，以便进一步分析和支持供应商决策。所开发的工作流运行于 UiPath 框架内，确保价格预测过程的自动化执行、数据一致性和可重复性。

3.2 基于 UiPath 的实现

标题：所开发的原型利用 UiPath 作为机器人流程自动化（RPA）工具，以实现汽车价格预测的数据提取流程自动化。UiPath 能够无缝处理重复性的网络爬取任务，确保高效收集和存储汽车相关数据。选择 UiPath 作为主要自动化工具，是基于其强大的网页数据提取能力、结构化工作流自动化功能以及完善的错误处理机制，这些特性显著提升了系统的整体效率。

UiPath 工作流首先通过创建一个名为 ExtractedDT 的数据表来初始化，该表用于存储关键的车辆属性，包括发动机尺寸、整备质量、马力、高速公路燃油效率、宽度和长度。如图 X 所示，工作流首先识别并访问汽车数据源，例如电子商务网站和供应商列表。随后，程序会遍历这些数据源，利用网页抓取技术收集相关的车辆详细信息。

为确保数据质量，提取的原始信息会经过预处理，包括字符串操作、数据格式化以及缺失值处理。系统采用错误处理机制，以检测缺失或不一致的数据条目。当发现记录不完整时，自动化流程会为其分配占位符值（例如“N/A”），或尝试通过模式识别进行数据修正。

处理完成后，结构化的汽车数据集会被存储在 ExtractedDT 中，并导出至 Excel，以便进一步分析以及与机器学习模型的集成。

基于 UiPath 的标准化实施确保了可扩展性、最小化的人工干预以及更高的数据准确性，使其成为自动化汽车价格数据采集的有效工具。这种方法能够实现一致且可重复的流程，从而助力于模型的精准评估与预测。

4 结果与发现

本节展示了 UiPath 自动化数据提取与基于 AI 的价格预测模型的测试结果。分析内容包括数据提取性能（准确率、效率、数据质量）以及不同回归模型的预测能力。

4.1 数据提取性能

UiPath 自动化流程成功提取了关键的汽车属

性，包括发动机尺寸、整备质量、马力、高速公路油耗（L/100km）、宽度和长度。整体数据提取准确率达到 98.7%，确保了数据的完整性。此外，99.2% 的车辆记录被成功抓取，仅有个别数据缺失，主要源于网页结构差异或部分字段暂时无法获取。

此外，自动化工作流将数据提取时间从手动操作的 2 分钟大幅缩短至自动化的 10 秒，从而将人工投入减少了 85%。提取后的数据被格式化为结构化表格，确保了模型训练与分析的一致性。

表2 UiPath自动化性能对比

指标	UiPath自动化结果
数据提取准确率	
数据完整性	0.987
每条数据处理时间	0.992
手动工作量减少比例	10秒（自动化）vs2分钟（手动）
	0.85

4.2 预测模型性能

为了评估 AI 模型的预测能力，研究者采用了线性与非线性回归技术来预测车辆价格。同时，利用决定系数（ R^2 ）来衡量模型的准确性。

结果显示，多元线性回归（MLR）的精度最高（ $R^2=0.8385$ ），因此成为最适合价格预测的模型。相比之下，多元非线性回归（MNR）的 R^2 得分最低（0.6286），这表明可能存在过拟合现象，或特征交互作用不足。

表3 回归模型预测精度对比

回归模型	R^2 分数
简单线性回归（SLR）	0.6699
简单非线性回归（SNR）	0.6717
多元线性回归（MLR）	0.8385
多元非线性回归（MNR）	0.6286

4.3 研究结果摘要

评估表明，UiPath 自动化显著提升了数据收集效率，确保了数据的结构化与高质量。人工智能模型有效利用了这些提取的特征来预测车辆价格，其中线性回归模型被证明是最为可靠的。

对于汽车供应商而言，结合 UiPath 自动化数据提取技术应用 MLR 模型，可构建一个强大的框架，用于价格预测与利润率优化。这一方法不仅大幅减少了人工操作，还确保了以数据驱动的定价策略得以实施。

表4 关键发现及其对电商自动化的影响

关键发现	对电商自动化的影响
UiPath自动化准确率：98.7%	确保可靠的结构化数据
手动工作量减少：85%	提升运营效率
最佳回归模型：多元线性回归（MLR）($R^2=0.8385$)	提供准确的价格预测
表现最差的模型：多元非线性回归（MNR）($R^2=0.6286$)	表明模型存在不稳定性 and 过拟合问题

这些发现凸显，电子商务中基于人工智能的自动化能够简化定价决策，使供应商在市场预测和盈利能力方面占据竞争优势。

5 结语

5.1 结论

本研究评估了基于 UiPath 的自动化数据提取技术及 AI 驱动的回归模型对电子商务中车辆价格预测的影响。通过利用网络爬虫自动化与预测建模，研究展示了人工智能与 RPA 技术如何助力汽车供应商优化定价策略。

UiPath 自动化工作流程成功提取了关键车辆属性，准确率达 98.7%，显著将人工数据采集工作量降低了 85%。通过构建的结构化数据集，我们能够应用多种回归模型来预测车辆价格。其中，多元线性回归 (MLR) 模型的预测精度最高 ($R^2 = 0.8385$)，因此成为最适合用于价格预测的模型。相比之下，多元非线性回归 (MNR) 模型的预测性能则稍逊一筹 ($R^2 = 0.6286$)，建议在特征交互或模型复杂度方面加以限制。

这些发现证实，将 UiPath 自动化与基于 AI 的预测建模相结合，能够提升定价准确性、增强市场竞争力，并帮助汽车供应商优化利润空间。同时，研究结果也强调，必须根据模型的准确性和数据可靠性，选择合适的回归模型。

表5 关键发现及其对电商定价自动化的影响

关键发现	对电商定价自动化的影响
UiPath数据提取准确率: 98.7%	确保了结构化的高质量输入数据
手动工作量减少: 85%	最大限度地减少了时间和人力成本
最佳回归模型: 多元线性回归 (MLR) ($R^2=0.8385$)	提供了最精准的价格预测
表现最差的模型: 多元非线性回归 (MNR) ($R^2=0.6286$)	表明模型存在不稳定性和过拟合风险

5.2 未来工作

尽管本研究为基于人工智能的车辆价格预测构建了框架，但未来的研究应着重于加强数据收集、优化预测建模，并整合实时自动化技术。通过纳入更多车辆属性，如品牌声誉、燃油效率和安全评级，扩充数据集将有助于进一步

提升价格预测的精准度。此外，若能提升 UiPath 的数据提取能力，使其更高效地处理复杂且动态的网站内容，也将显著改善数据质量。从建模角度来看，应用深度学习技术，例如神经网络和集成模型，有望进一步提高预测精度；同时，结合先进的特征工程方法，则可优化模型的可解释性，从而确保预测结果更加可靠。

未来工作的另一个关键方向是实时价格监控以及大规模的企业级实施。开发一个集成实时市场趋势、竞争对手定价和供应链波动的实时仪表盘，将使企业能够动态调整价格。此外，基于人工智能的自动化决策系统可进一步优化定价策略，实时响应需求与供应的变化。最后，为满足企业级部署需求，对 UiPath-AI 框架进行扩展，确保其具备卓越的计算效率，并实现无缝的云端集成，从而使其适用于大型电子商务平台。通过聚焦这些领域，人工智能与 RPA 驱动的自动化技术将继续革新车辆定价策略，助力电商行业实现更高的精准度、效率与盈利能力。

参考文献:

- [1] E. Gegic, B. Isakovic, D. Keco, Z. Masetic, and J. Kevric, "Car priceprediction using machine learning techniques," TEM Journal, vol. 8, no. 1,p. 113, 2019.
- [2] M. Asghar, K. Mehmood, S. Yasin, and Z. M. Khan, "Used cars priceprediction using machine learning with optimal features," Pakistan Journalof Engineering and Technology, vol. 4, no. 2, pp. 113 - 119, 2021.
- [3] L. Bukvić, J. Pašagić Škrinjar, T. Fratrović, and B. Abramović, "Priceprediction and classification of used-vehicles using supervised machinelearning," Sustainability, vol. 14, no. 24, p. 17034, 2022.
- [4] C. Longani, S. Prasad Potharaju, and S. Deore, "Price prediction for pre-owned cars using ensemble machine learning techniques," in Recent Trendsin Intensive Computing, IOS Press, pp. 178 - 187, 2021.
- [5] F. Al-Turjman, A. A. Hussain, S. Alturjman, and C. Altrjman, "Vehicleprice classification and prediction using machine learning in the IoT smartmanufacturing era," Sustainability, vol. 14, no. 15, p. 9147, 2022.