

# 基于决策树模型的电动汽车目标客户销售策略

岳巧琳, 林旭旭\*, 杨菁岩

湖南科技学院理学院, 湖南 永州

收稿日期: 2024年7月19日; 录用日期: 2024年8月11日; 发布日期: 2024年8月20日

## 摘要

针对消费者对于购买新能源汽车的意愿存在很大的不确定性的问题, 基于顾客满意度以及个人特征信息为影响的主要因素进行分析, 采用决策树模型对顾客的购买意愿进行研究。通过案例分析, 得出影响各品牌顾客购买意愿的主要因素, 即对于品牌1和品牌3, 影响顾客购买意愿的主要因素为电动汽车的质量和驾驶感受, 而对于品牌2, 主要影响因素为经济问题。并由Kano模型满意度对各品牌营销给出了营销建议。

## 关键词

决策树模型, Kano模型, 满意度调查表

# Targeted Customer Selling Strategy for Electric Vehicles Based on Decision Tree Modeling

Qiaolin Yue, Xuxu Lin\*, Jingyan Yang

College of Science, Hunan University of Science and Engineering, Yongzhou Hunan

Received: Jul. 19<sup>th</sup>, 2024; accepted: Aug. 11<sup>th</sup>, 2024; published: Aug. 20<sup>th</sup>, 2024

## Abstract

There is great uncertainty about consumers' willingness to buy new energy vehicles. Based on customer satisfaction and personal characteristics as the main factors of the analysis. The decision tree model is used to study the purchase intention of customers. Through case analysis, we identify the main factors that affect customers' purchasing intentions for various brands. For brands 1

\*通讯作者。

文章引用: 岳巧琳, 林旭旭, 杨菁岩. 基于决策树模型的电动汽车目标客户销售策略[J]. 应用数学进展, 2024, 13(8): 3885-3891. DOI: 10.12677/aam.2024.138370

and 3, the main factors affecting customer purchase intention are the quality and driving experience of electric vehicles, while for brand 2, the main influencing factor is economic issues. And the satisfaction of the Kano model was used to provide marketing recommendations for each brand's marketing.

## Keywords

Decision Tree Model, Kano Model, Satisfaction Questionnaire

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

随着全球能源危机的加剧和国际局势的动荡，油价随之上涨，传统燃油汽车的销售也受到一定程度的影响。为了面对全球环境恶化的局面，世界各国将目光瞄准在电动汽车产业，也都制定了一些扶持电动汽车产业的政策。如今电动汽车最核心问题的电池问题也慢慢在解决，电动汽车产业已然成为战略新型产业，发展以电动汽车为代表的新能源汽车是解决能源环境问题的不错手段。但是，电动汽车毕竟是一个新兴的事物，与传统汽车相比，消费者在一些领域，如电池问题，还是存在着一些疑虑，仅仅依靠市场的自运行还很难达到产业化目标，其市场销售需要科学决策。

随着人们生活、生产方式与信息之间的联系愈发紧密，科学研究默默影响着人们的生活，国内外学者对顾客满意度指数模型进行了比较深入全面的研究。如 Olasanmi [1]等采用频率计数和回归分析的方法，对收集的结构化问卷数据进行分析，进一步对与顾客满意度存在显著关系的因素进行探索；章凯兵 [2]基于神经网络模型，提出从用户满意度的影响因素方面构建用户满意度预测模型，基于用户偏好挖掘模型的输出结果定义用户偏好特征；李燕仪 [3]采用聚类分析和关联分析的方法提炼出的购车客户群体特征、群体喜好和购车需求，为销售部门制定差异化营销策略提供依据。尽管目前已对顾客满意度的评估和预测进行了广泛探讨，但是对于顾客购买意愿的研究却相对较少，且其在实际操作中的适应性和效果尚未得到充分验证。

因此，本文基于顾客满意度以及个人特征信息为影响的主要因素进行分析，构建决策树模型以对顾客的购买意愿进行评估，便于决策者对产品做出精准定位。在此基础上，又结合 kano 模型对顾客满意调查表进行数据挖掘，对不同品牌的电动汽车给出不同的营销策略，将研究成果运用于实践。同时，本文的创新之处在于：模型是在对大量数据进行分析和处理后建立的，并且建模过程严格按照观察数据 - 发现问题 - 分析问题 - 提出假设 - 建立模型 - 验证模型 - 综合分析 - 规律总结与推广的科学探索过程来对问题进行研讨；并且通过对模型的层层检验和比较，所构建的模型简洁实用，可移植性强。

## 2. 相关理论概述

### 2.1. 决策树模型

决策树是一种基本的分类与回归方法 [4]，它可以认为是定义在特征空间与类空间的条件概率分布，决策树思想，实际上就是寻找最简便的划分方法。决策树与逻辑回归的不同之处在于，逻辑回归是根据所有特征求出概率，然后与某一阈值进行比较从而分类，而决策树每一步是通过最优特征进行划分，直

到叶节点。决策树的学习过程主要包括 3 个步骤：特征选择、决策树的生成和决策树的剪枝。常用的算法有 ID3 算法、C4.5 算法以及 CART 算法。本文主要使用 CART 算法。

## 2.2. 基尼系数

CART 分类树算法使用基尼系数选择特征，基尼系数代表了模型的不纯度，基尼系数越小，不纯度越低，特征越好。数据集  $D$  的纯度可用基尼值来度量

$$Gini(D) = \sum_{i=1}^n p(x_i) * (1 - p(x_i)) = 1 - \sum_{i=1}^n p(x_i)^2 \quad (1)$$

其中， $p(x_i)$  是分类  $x_i$  出现的概率， $n$  是分类的数目。 $Gini(D)$  反映了从数据集  $D$  中随机抽取两个样本，其类别标记不一致的概率。因此， $Gini(D)$  越小，则数据集  $D$  的纯度越高。

对于样本  $D$ ，个数为  $|D|$ ，根据特征  $A$  是否取某一可能值  $a$ ，把样本  $D$  分成两部分  $D_1$  和  $D_2$ 。所以 CART 分类树算法建立起来的是二叉树，而不是多叉树。

$$D_1 = (x, y) \in D | A(x) = a \quad (2)$$

$$D_2 = D - D_1 \quad (3)$$

在属性  $A$  的条件下，样本  $D$  的基尼系数定义为：

$$GiniIndex(D | A = a) = \frac{|D_1|}{|D|} Gini(D_1) + \frac{|D_2|}{|D|} Gini(D_2) \quad (4)$$

## 2.3. CART 分类树步骤

输入：训练集  $D$ ，基尼系数的阈值，切分的最少样本个数阈值。

输出：分类树  $T$ 。

算法从根节点开始，用训练集递归建立 CART 分类树。

Step1：对于当前节点的数据集为  $D$ ，如果样本个数小于阈值或没有特征，则返回决策子树，当前节点停止递归；

Step2：计算样本集  $D$  的基尼系数，如果基尼系数小于阈值，则返回决策子树，当前节点停止递归。

## 2.4. Kano 模型 Better-Worse 系数

Kano 模型是由 Noriaki Kano 教授于 1980 年代提出的产品开发和客户满意度理论，它将客户偏好分为五类。这五类偏好即五种需求类型，分别是，基本需求(Must-be)、期望需求(Satisfiers)、魅力需求(Exciters/Delighters)、无差异需求、反向需求(Dissatisfiers)。Better-Worse 系数，表示某功能可以增加满意或者消除不喜欢的影响程度。Better，简单说就是满意系数，代表如果产品提供某种功能或服务，用户满意度会提升。Better 值越接近 1，则表示用户满意度提升的效果会越强。Worse，不满意系数(通常为负)，代表产品如果不提供某种功能或服务，用户的满意度会降低。其绝对值越接近 1，则表示对用户不满意的影响最大，满意度降低的影响效果越强，下降的越快。其计算公式如下：

$$(\text{满意影响力}) SJ = (A + O) / (A + O + M + I); \quad (5)$$

$$(\text{非满意影响力}) DSJ = -1 * (O + M) / (A + O + M + I); \quad (6)$$

其中， $A$ ：魅力需求； $O$ ：期望型需求； $M$ ：基本需求； $I$ ：无差别需求。

### 3. 潜在客户挖掘的决策树模型

#### 3.1. 数据预处理

本文数据参照 2021 年华数杯数学建模竞赛 C 题的附件，附件中提供了 1964 位目标客户的分别对三种品牌电动汽车的 8 个体验指标数据(a1~a8)，以及目标客户个人特征数据(b1~b17)。

为比较目标客户对于汽车品牌的满意度差异[5]，本文对数据进行了三个品牌的体验满意度变量的小提琴图，如图 1 所示。小提琴图刻画了数据的分布特征，这种图用来显示数据的分布和概率密度，可以看成是箱线图和密度图的结合。小提琴图越宽的地方表示数据密度越大，可以展示出数据的多个峰值。

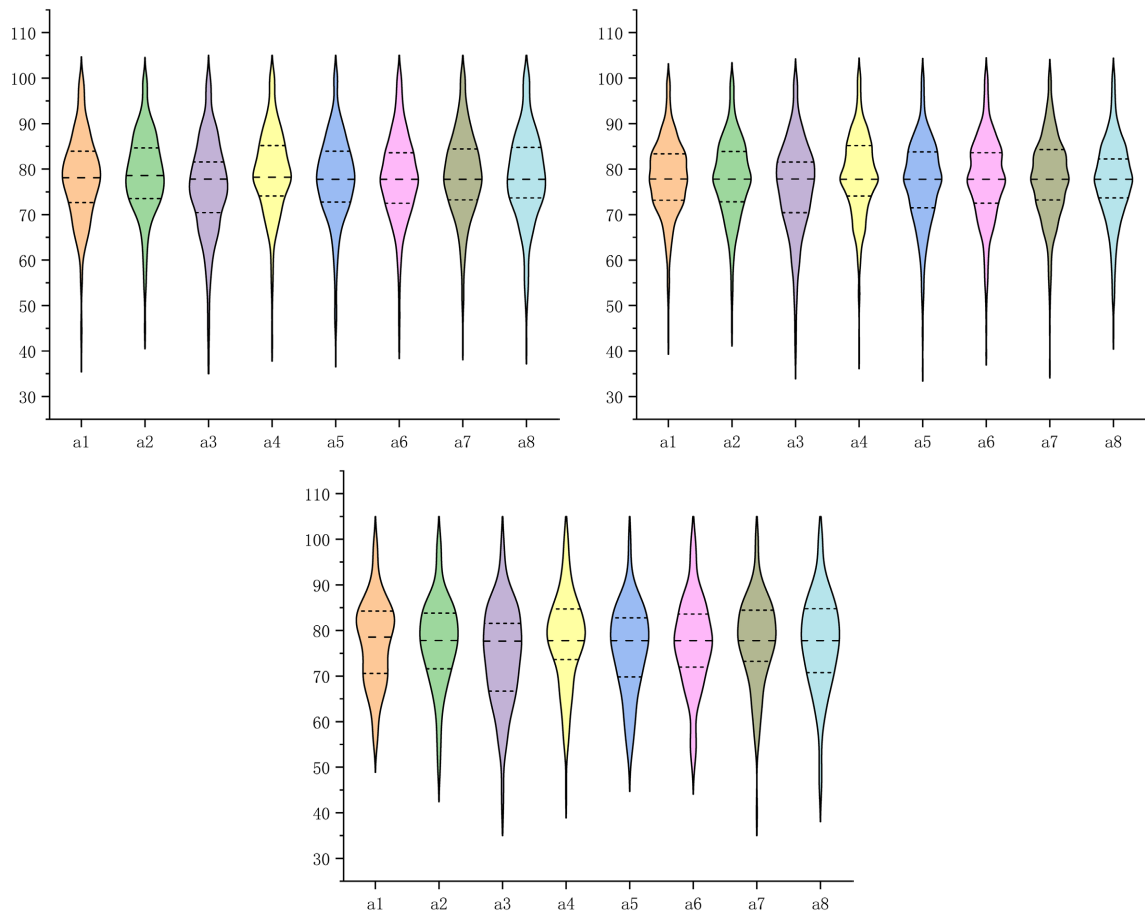


Figure 1. Three brands of violin charts

图 1. 三个品牌的小提琴图

从图 1 中可以看出，三个品牌对应的满意度得分的中位数都非常接近，而分布存在着差异。对于品牌类别 1，数据分布基本上只有一个明显的峰值，说明客户对于该种品牌的满意度较为一致。对于品牌类别 2，数据分布呈多峰形态，说明不同客户对该种品牌的满意度呈现多级分化。对于品牌类别 3，a1 (电池技术性能满意度)得分有两个分支，说明其满意度呈两极分布；a2~a8 大体上呈单峰分布，说明客户对于品牌 3 的 a2~a8 满意度较为统一。为了量化目标客户对三种电动汽车品牌的满意度，本文将电池技术性能满意度得分 a1、环保与空间座椅整体表现满意度得分 a2、耗能与保值率整体满意度得分 a3、刹车和行车视野整体满意度得分 a4、爬坡和加速整体满意度得分 a5、转弯和高速的稳定性整体满意度得分 a6、外观内饰整体表现满意度得分 a7、配置与质量品质整体满意度得分 a8 视为主要影响因素。

对全体目标客户评价的 a1~a8 分别求平均值, 然后再对这 8 项指标的平均值再次求平均, 并求出最后得分。运行结果显示, 合资品牌、自主品牌和新势力品牌的满意度最终得分分别为 78.16 分、77.54 分和 76.93 分, 表明目标客户普遍对合资品牌的满意度评价最高, 新势力品牌的满意度最低[6]。

### 3.2. 潜在客户挖掘模型构建

为了建立不同品牌电动汽车的客户挖掘模型, 首要条件是对每个影响因素做决策分析, 再用结果建立决策树模型。换句话说, 从满意度调查表中得出哪些因素会影响目标客户最终是否购买。下面本文以合资品牌为例, 利用 SPSS 中的决策树模块, 来分析影响合资品牌电动汽车销售的主要因素。

#### 1) 数据导入

由于是三种品牌的满意度调查表, 本文的思路是将三种品牌的电动汽车满意度调查表数据分开, 分别检验影响三种品牌的销售的主要因素。以品牌 1 为例, 由前面已经进行过数据清洗工作所得数据, 从中将 556 名目标客户的满意度调查导入 excel 文件, 其中 533 名客户没有选择购买, 23 名客户选择购买。

#### 2) 决策树的建立

为了更清楚的了解各个因素对顾客购买意愿的影响, 本部分将以 a1~a8, b1~b17 为自变量, 最终是否购买为因变量, 本文对是否购买这里做了 0~1 规划, 将最终购买视为 1, 没有购买视为 0, 建立相应的决策树[7]。在 SPSS 的分析选项卡下的分类中, 选择决策树并运行后, 所得结果如表 1 所示:

**Table 1.** Decision tree result diagram

**表 1.** 决策树结果图

		模型摘要
指定项	生长法	CHAID
	因变量	是否购买
	自变量	a1, a2, a3, a4, a5, a6, a7, a8, b1, b2, b3, b4, b5, b6, b7, b8, b9, b10, b11, b12, b13, b14, b15, b16, b17
	验证	无
	最大树深度	3
	父节点中的最小个案数	100
	子节点中的最小个案数	50
	包括的自变量	a6, a7, b17
结果	节点数	9
	终端节点数	6
	深度	2

不难发现对于合资品牌来说, 驾驶操控性表现(a6)、外观内饰整体表现(a7)和全年车贷的支出占家庭年总收入的比例(b17)是影响目标客户最终是否会购买主要原因[8]。同理, 用相同的操作可得, 自主品牌主要影响因素有电池技术性能(a1)、家庭年收入(b13)、全年房贷的支出占家庭年总收入的比例(b16)和全年车贷的支出占家庭年总收入的比例(b17)。对于新势力品牌, 目标客户购买需求是配置与质量品质整体满意度(a8)。总结对品牌 1 和品牌 3, 顾客购买主要看电动汽车的质量和驾驶感受, 而对于品牌 2 最大的

问题就是经济问题。

利用上述建立的决策树模型分别对品牌 1、2 和品牌 3 的 15 名潜在客户进行预测[9]，结果如表 2 所示：

**Table 2.** Shows the prediction results

**表 2.** 预测结果图

客户编号	品牌编号	预测结果	客户编号	品牌编号	预测结果	客户编号	品牌编号	预测结果
1	1	不购买	6	2	购买	11	3	不购买
2	1	不购买	7	2	不购买	12	3	不购买
3	1	不购买	8	2	不购买	13	3	购买
4	1	不购买	9	2	不购买	14	3	不购买
5	1	不购买	10	2	不购买	15	3	不购买

由结果可知，附件三中 15 名待预测的目标客户中，仅有编号为 6、13 的客户有较大概率会选购品牌 1、品牌 2、品牌 3 的电动汽车，其他客户选购电动车的可能性较低。

#### 4. 基于 Kano 模型的销售策略

对营销者加大服务力度，可以在短的时间内提高 a1~a8 五个百分点的满意度，且服务难度与提高的满意度百分点是成正比的假设下，本文在附录 3 每个品牌中各挑选编号为 2、8、15 的目标客户，通过其 a1~a8 的满意度调查表，通过 kano 模型中的满意度来分析品牌是否需要加大服务力度。编号 2、8、15 没有购买意愿的顾客 a1~a8 评分如表 3 所示。

**Table 3.** a1~a8 ratings for customers with numbers 2, 8, 15 and no intention to buy

**表 3.** 编号 2、8、15 且没有购买意愿的顾客 a1~a8 评分

品牌编号	a1	a2	a3	a4	a5	a6	a7	a8	是否购买
品牌 1	85.04	87.65	81.56	88.88	85.63	86.14	84.31	81.44	0
品牌 2	82.80	83.24	77.81	85.61	84.02	83.60	82.15	81.09	0
品牌 3	83.58	81.09	70.44	81.45	74.69	80.85	75.43	73.66	0

根据前面选出决定因素的 a3、a6、a7、a8，依次为 *M*、*A*、*O*、*I*，运算结果如表 4 所示：

**Table 4.** *SJ* and *DSJ* values of the three brands

**表 4.** 三个品牌的 *SJ* 值和 *DSJ*

	品牌 1	品牌 2	品牌 3
<i>SJ</i>	0.511	0.511	0.520
<i>DSJ</i>	-0.497	-0.503	0.486

当 *SJ* 值和 *DSJ* 绝对值均 >0.5 时判定为期望质量，当 *SJ* 值和 *DSJ* 绝对值均 <0.5 时为无关质量，当 *SJ* 值 >0.5、*DSJ* 绝对值均 <0.5 时为魅力质量，当 *SJ* 值 >0.5、*DSJ* 绝对值均 <0.5 时为魅力质量。一般来说，



基本型需求 > 期望型需求 > 魅力型需求, 要尽量避免做无差异需求和反向需求[10]。由上得出结论品牌 1、3 为魅力质量, 品牌 2 为期望质量。因此, 对于品牌 2 质量的提高 5% 对服务难度影响小, 品牌 1、3 质量的提高 5% 对服务难度影响大。所以要加大品牌 2 的服务力度。

## 5. 结论

本文针对消费者对于购买新能源汽车的意愿存在很大的不确定性的问题, 结合决策树模型构造了潜在客户挖掘模型, 最终研究结果为: 目标客户普遍对品牌 1 的购买意愿较高, 对品牌 3 的购买意愿较低; 并且对于品牌 1 和品牌 3, 影响顾客购买意愿的主要因素为电动汽车的质量和驾驶感受, 而对于品牌 2, 主要影响因素为经济问题。基于研究结果, 各品牌可以挖掘出潜在的客户, 并且不断有针对性的提高自身在市场竞争中的竞争力。再由 kano 模型满意度对各品牌营销给出了营销建议, 即品牌 1、3 需提高质量保障, 品牌 2 需提高服务力度, 以此进一步增加企业利润。

## 基金项目

湖南科技学院大学生创新创业训练计划项目(2023 年度); 湖南省教育厅科学研究一般项目(22C0542); 永州市指导性科技计划项目(2023 年度)。

## 参考文献

- [1] Olanmi, O.O. (2019) Online Shopping and Customers' Satisfaction in Lagos State, Nigeria. *American Journal of Industrial and Business Management*, **9**, 1446-1463. <https://doi.org/10.4236/ajibm.2019.96095>.
- [2] 章凯兵. 基于神经网络的用户满意度预测研究[D]: [硕士学位论文]. 赣州: 江西理工大学, 2020.
- [3] 李燕仪. 基于数据挖掘方法的汽车客户画像分析及流失客户预测[D]: [硕士学位论文]. 广州: 华南理工大学, 2017.
- [4] 谢兆贤, 邹兴敏, 张文静. 大型数据集的高效参数剪枝决策树算法研究[J]. *计算机工程*, 2024, 50(1): 156-165.
- [5] 黄诗瑶. 聚类分析在移动通信用户行为分析中的研究与应用[D]: [硕士学位论文]. 广州: 广东工业大学, 2013.
- [6] Leung, C.W., Chan, S.C., Chung, F. and Ngai, G. (2011) A Probabilistic Rating Inference Framework for Mining User Preferences from Reviews. *World Wide Web*, **14**, 187-215. <https://doi.org/10.1007/s11280-011-0117-5>
- [7] Malik, A.J. and Khan, F.A. (2017) A Hybrid Technique Using Binary Particle Swarm Optimization and Decision Tree Pruning for Network Intrusion Detection. *Cluster Computing*, **21**, 667-680. <https://doi.org/10.1007/s10586-017-0971-8>
- [8] Sawant, S.S., Wiedmann, M., Göb, S., Holzer, N., Lang, E.W. and Götz, T. (2022) Compression of Deep Convolutional Neural Network Using Additional Importance-Weight-Based Filter Pruning Approach. *Applied Sciences*, **12**, Article 11184. <https://doi.org/10.3390/app122111184>
- [9] 白一凡. 基于 SARIMA 和 BP 神经网络的新能源汽车销售预测[D]: [硕士学位论文]. 湘潭: 湘潭大学, 2019.
- [10] 陈建凯, 王鑫, 何强, 等. 区间值属性的单调决策树算法[J]. *模式识别与人工智能*, 2016, 29(1): 47-53.