

基于多源数据融合与Prophet-XGBOOST混合模型的新能源汽车渗透率预测

罗 婕, 艾美好, 徐 毅

长沙理工大学数学与统计学院, 湖南 长沙

收稿日期: 2026年3月22日; 录用日期: 2026年4月22日; 发布日期: 2026年5月6日

摘 要

在能源安全问题日益凸显与碳达峰、碳中和目标的催化下, 新能源汽车成为缓解能源压力、推动绿色转型的重要力量。因此, 准确预测新能源汽车渗透率对产业可持续发展具有重要意义。本文运用多源数据融合技术, 选取2023年1月到2025年12月新能源汽车月度销量数据、充电桩数据、GDP、能源价格及双积分政策文本作为基础数据, 构建Prophet-XGBoost模型进行预测。利用Prophet模型提取渗透率趋势项和季节项作为时序特征, 通过XGBoost模型融合时序特征和多源外部因素进行训练与预测。对比分析表明, 混合模型在测试集上的MAE和RMSE较单一Prophet模型分别降低0.12%和0.15%, 为未来新能源汽车渗透率的预测提供了一个精度较高的有效模型。结果显示2026年新能源汽车渗透率有望在年底达到54.20%。

关键词

新能源汽车, 渗透率预测, Prophet模型, XGBoost模型, 多源数据融合

Prediction of New Energy Vehicle Penetration Rate Based on Multi-Source Data Fusion and Prophet-XGBOOST Hybrid Model

Jie Luo, Meihao Ai, Yi Xu

School of Mathematics and Statistics, Changsha University of Science and Technology, Changsha Hunan

Received: March 22, 2026; accepted: April 22, 2026; published: May 6, 2026

Abstract

Under the catalysis of the increasing prominence of energy security issues and the goal of carbon

文章引用: 罗婕, 艾美好, 徐毅. 基于多源数据融合与 Prophet-XGBOOST 混合模型的新能源汽车渗透率预测[J]. 可持续发展, 2026, 16(5): 1-9. DOI: 10.12677/sd.2026.165181

peaking and carbon neutrality, new energy vehicles have become an important force to alleviate energy pressure and promote green transformation. Therefore, accurately predicting the penetration rate of new energy vehicles is of great significance to the sustainable development of the industry. This paper uses multi-source data fusion technology to select the monthly sales data, charging pile data, GDP, energy price and double-point policy text of new energy vehicles from January 2023 to December 2025 as the basic data, and construct the Prophet-XGBoost model for prediction. Use the Prophet model to extract the penetration rate trend items and seasonal items as timing characteristics, and use the XGBoost model to integrate timing characteristics and multi-source external factors for training and prediction. Comparative analysis shows that the MAE and RMSE on the test set of the hybrid model are 0.12% and 0.15% lower than the single Prophet model, respectively, providing an effective model with high accuracy for predicting the penetration rate of new energy vehicles in the future. The results show that the penetration rate of new energy vehicles in 2026 is expected to reach 54.20% by the end of the year.

Keywords

New Energy Vehicles, Penetration Rate Prediction, Prophet Model, XGBoost Model, Multi-Source Data Fusion

Copyright © 2026 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

发展新能源汽车被普遍认为是提升汽车产业国际竞争力、推动能源结构转型以及促进全球可持续发展的重要路径。由此可见，新能源汽车对我国及世界的重要性。此外，在当今这个交通发达的时代，随着居民出行需求增长，国内汽车消费市场持续扩大。而在能源安全问题日益凸显和碳达峰、碳中和目标的背景下，作为第三大碳排放源的道路交通领域承担的减碳任务之重不言而喻，新能源汽车以其低排放、高效能的特点，被视为交通领域缓解能源压力和减少环境污染的重要手段。中国政府高度重视新能源汽车产业的发展，出台了一系列政策来助力新能源汽车的发展和普及，从购车补贴、税收优惠到充电桩等基础设施建设的大力投入，为新能源汽车的发展创造了良好的政策环境，于是新能源汽车产业迅速发展壮大，成为国家实现绿色低碳发展、构建现代化产业体系的重要力量之一。

2024 年中共中央国务院发布了关于加快经济社会发展全面绿色转型的意见，意见中指出，到 2035 年新能源汽车成为销售车辆的主流。然而，新能源汽车的发展面临着许多重大挑战。毛华硕[1]指出政策因素、购买价格和成本、能源价格、配套基础设施与续航能力等多种因素共同影响着新能源汽车销量增长。通过系统地分析这些具体影响因素，能够更准确地预测新能源汽车未来的销量走向，为产业发展提供支撑。

综观国内外新能源汽车领域的研究进展，预测方法正从单一模型向混合建模演进，多源数据融合已成为提升预测精度的共识方向。在模型架构层面，国内学者刘连义等[2]将灰色系统理论与离散时间建模相结合，证明了在处理小样本、非线性数据时灰色模型的独特优势，高章等[3]采用 PSO-BP 神经网络对纯电动重型车能耗进行预测，表明改进型神经网络能够有效处理非线性、多因素耦合的预测问题，毛华硕[1]的综述进一步指出，组合模型与多因素融合模型已成为预测研究的主流趋势。国际研究则展现出更前沿的技术路径，Alatawneh 和 Torok [4]提出的修正高斯与 Logistic 混合建模框架，Cavus 等[5]的 Deep Charge-Fusion Model，以及 Dursun [6]针对机动化率上升国家提出的渗透率预测新方法，共同证明了混合模型架构在捕捉市场演化阶段性特征、融合多源异构数据方面的显著优势。在多源数据融合方面，史一龙等[7]在国内研究

中明确提出多源特征融合框架,将车辆运行、环境、驾驶行为等多维度信息整合建模,张志淦等[8]采用DTM主题模型和Bi-LSTM情感分类,将社交媒体数据纳入分析框架,为融合非结构化数据提供了方法论参考;国际研究则展现出更精细化的特征工程思路,Chandra等[9]在全球电动汽车需求预测中将人口增长与经济发展作为核心外生变量,Jain等[10]通过对主流媒体话语的批判性分析证明非结构化数据可作为政策效果评估的重要补充,Ramadoss等[11]通过电动汽车采用者分类为渗透率的细分市场预测提供了方法论支撑。值得注意的是,学者日益重视区域异质性与政策变量的影响。翁剑成等[12]揭示不同城市的激励政策效果存在显著差异,贾文娟等[13]聚焦农村市场指出城乡消费心理与基础设施条件的差异,王静[14]对中国新能源汽车产业政策变迁的系统梳理揭示了政策演进的阶段性特征,Hu[15]从协同创新视角识别了创新要素配置模式对产业发展的影响,进一步印证了政策引导与创新网络的关键作用。

综合而言,现有研究在预测方面已取得了一定的成就,但在数据融合方面,数据维度可以进一步拓宽;在模型构建方面,传统模型的预测精度和抗干扰性有待提高。

相较于目前研究成果,本文将Prophet-XGBoost混合模型与中国国情相结合,融合本土市场的多源特征体系,做出了如下贡献:

- 1) 相较传统模型,本模型进一步提高了预测精度;
- 2) 本文选取了影响新能源汽车渗透率的五大主要影响因素进行系统融合。

2. 研究设计与方法

本文整体研究思路基本遵循“数据融合-时序分解-特征增强-模型预测”的逻辑框架。首先构建影响新能源汽车渗透率的多源因素体系并进行预处理,其次运用Prophet模型对渗透率进行时序分解,然后将时序基准成分与外部驱动因素进行特征融合,最后通过XGBoost模型实现结果预测,并对模型进行评估,以验证模型的有效性。

2.1. 数据来源与预处理

2.1.1. 数据来源

基于多源数据融合预测的理论框架,及历史规律与外部驱动相结合的分析框架,将影响新能源汽车渗透率的核心因素划分为市场惯性、政策环境、经济条件与基础设施四个维度,由此系统选取具有代表性的五个指标,初步构建了如图1所示的新能源汽车渗透率影响因素体系。



Figure 1. The effect of penetration rate on the system
图 1. 渗透率影响体系

于是,从国家数据网站选取了 2023 年 1 月至 2025 年 12 月的新能源汽车历史销量数据、能源价格指数、GDP、充电桩数据及双积分政策作为基础数据。

2.1.2. 数据预处理

本节对数据进行预处理,处理流程为:数据合并与缺失值处理→政策文本量化→特征指标构造→目标变量处理。

1) 数据合并与缺失值处理

首先根据公式新能源汽车渗透率 = (新能源汽车销量/汽车总销量) × 100% 计算了月度渗透率,然后将五类数据按月份进行左合并,生成完整的时间序列表。最后进行缺失值处理,对于渗透率、充电桩数据、GDP、能源价格指数等四个连续型变量采用线性插值法填充缺失值,而对于具有低频突变特征的政策变量采用前向填充法再向后填充法处理缺失值。

2) 政策文本量化

本文基于 2017 年国家工信部等五部门联合发布的《乘用车企业平均燃料消耗量与新能源汽车积分并行管理办法》(简称“双积分政策”),将双积分政策转化为一套可计算的数值指标体系。政策文件规定了新能源汽车积分比例要求,包含精细的积分计算规则,并对续航里程、能量密度、能耗水平设定了多个门槛,由此可将计算规则中的各项技术参数和调整系数提取为结构化数据,并构建反映积分交易、结转与惩罚等行为的代理变量。

3) 特征指标构造

为了充分挖掘多源数据中的预测信息,本文构建了包含衍生比率特征、滞后特征、滚动统计特征、时间特征、Prophet 预测特征、政策变量等的多维度特征指标体系。其中衍生比率特征涵盖充电桩月度环比增速、充电桩年度同比增速、GDP 同比增速,用以反映各因素的动态变化趋势;滞后特征中引入滞后 1 期渗透率、滞后 2 期渗透率和滞后 1 期能源价格,用以刻画市场惯性和延迟效应;滚动统计特征包括过去 3 个月渗透率的滚动均值,用以平滑短期波动;时间特征中提取了月份和季度,以此捕捉周期性波动;Prophet 预测特征则是利用 Prophet 模型对历史渗透率进行分解,将其预测值作为时序特征输入 XGBoost,以此实现时序和特征的协同;而政策变量则是直接使用原始政策文本中的新能源积分比例、低油耗乘用车核算倍数、积分结转规则、积分池开放虚拟变量、政策修订虚拟变量。

4) 目标变量处理

由于渗透率序列具有较强的自相关性,直接用其进行预测可能导致模型过度依赖滞后项而忽略外部驱动因素的影响,因此本文构造渗透率的一阶差分作为 XGBoost 模型的预测目标以此降低渗透率序列的自相关性,具体如下:

$$\Delta y_t = y_t - y_{t-1}, t = 2, 3, \dots, T$$

其中 y_t 为第 t 月的渗透率, Δy_t 为月度变化量。

模型输出一阶差分预测值后,通过如下公式累加还原渗透率水平值,其中初始值取最后一月真实渗透率:

$$\hat{y}_t = \hat{y}_{t-1} + \hat{\Delta}_t$$

2.2. Prophet-XGBoost 混合预测模型

2.2.1. Prophet 模型

Prophet 模型[16] [17]是由 Facebook 在 2017 年开源的一种预测方法,其适用于具有强季节性和趋势特征的时间序列。而对于本文所研究的新能源汽车渗透率,其时间序列具有两个显著特征,一是随着产

业逐渐成熟，渗透率整体呈现上升的长期趋势；二是受促销季等因素的影响，渗透率存在年度的周期性波动。这使得 Prophet 模型非常适用。

该模型的核心思路是将时间序列拆解成趋势项、季节项、节假日项等，从而分别捕捉数据的长期变化规律和周期性波动。考虑到中国新能源汽车市场不存在显著的节假日效应，本文在研究时仅保留趋势项、季节项等，具体数学表达式如下：

$$y(t) = g(t) + s(t) + \varepsilon_t$$

其中 $y(t)$ 表示本研究中第 t 月的新能源汽车渗透率； ε_t 为误差项，代表模型未涵盖的其他成分影响； $g(t)$ 为趋势项，用于描述序列的长期变化方向，采用逻辑回归函数进行拟合：

$$g(t) = \frac{C}{1 + \exp(-k(t-m))}$$

其中 C 表示市场饱和下渗透率上限， k 为增长速度参数， m 为趋势拐点位置； $s(t)$ 为季节项，用于刻画周期性的波动，通过傅里叶级数展开来逼近：

$$s(t) = \sum_{n=1}^N \left(a_n \cos\left(\frac{2\pi nt}{P}\right) + b_n \sin\left(\frac{2\pi nt}{P}\right) \right)$$

其中 P 为周期长度，在本文中，取 12 用以捕捉年度周期性。

2.2.2. XGBoost 模型

XGBoost [17][18] 是由 Chen T 等人在 2016 年提出的一种基于决策树集成的机器学习算法，该算法通过迭代构建多个回归树，并将它们的结果加权组合，从而获得比单一模型更优的预测性能。且能够有效处理多源异构数据、捕捉复杂非线性关系和评估特征重要性，适用于本研究。

本文需要解决的问题可以简单描述为：给定包括充电桩数据、人均 GDP、能源价格、Prophet 时序特征及政策变量等的第 t 个月的特征向量 \bar{x}_t ，希望预测该月渗透率相较于上月的变化量 Δy_t 。XGBoost 的预测输出可表示为所有决策树的累加和：

$$\hat{y}_i = \sum_{k=1}^K f_k(\bar{x}_i)$$

其中 f_k 表示第 k 棵决策树， K 为树的总数， \bar{x}_i 表示第 i 个样本的特征向量。

模型训练的目标是最小化如下目标函数：

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

其中 $l(y_i, \hat{y}_i)$ 为损失函数，用于刻画预测值和真实值之间的差距； $\sum_{k=1}^K \Omega(f_k)$ 为正则化项，用于控制模型复杂度，防止过拟合。 $\Omega(f_k)$ 具体形式如下：

$$\Omega(f_k) = \gamma T + \frac{1}{2} \lambda \sum_{j=1}^T \omega_j^2$$

其中 T 表示决策树的叶子节点数， ω_j 为第 j 个叶子节点的权重， γ 和 λ 均为正则化参数，其中 γ 用于控制叶子节点数量， λ 用于控制权重幅度，从而保证模型在拟合训练数据时有良好的泛化能力。

2.2.3. Prophet-XGBoost 混合模型的构建

基于 Prophet 模型和 XGBoost 模型各自在新能源汽车渗透率预测中展现的优势，构建 Prophet-

XGBoost 混合模型，以综合 Prophet 模型在时序分解方面的优势和 XGBoost 模型在外部驱动因素上的灵敏反应，提高预测的准确性。如图 2 为研究流程图。

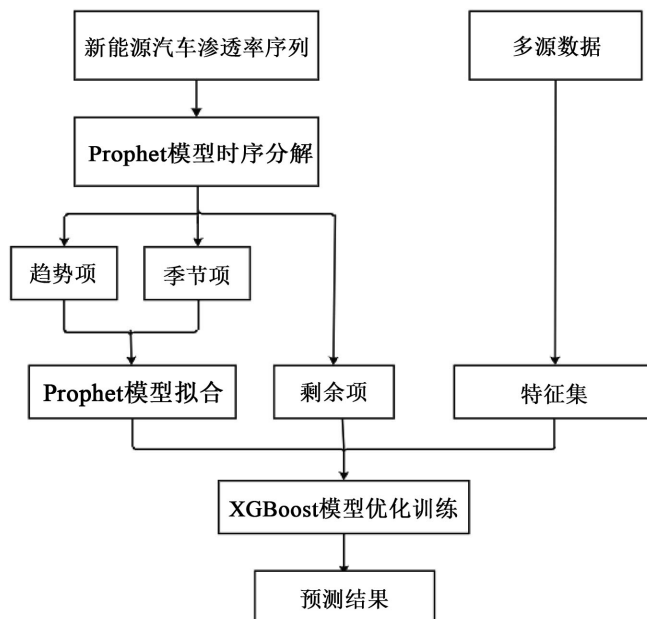


Figure 2. Research flowchart
图 2. 研究流程图

Prophet-XGBoost 混合模型实现流程如下：

- 1) 首先利用新能源汽车渗透率序列训练 Prophet 模型，然后利用 Prophet 模型将渗透率时序分解为趋势项、季节项和剩余项，并输出对整个历史时期的拟合值，作为后续 XGBoost 模型的 Prophet 预测特征。
- 2) 将 Prophet 预测特征与其他五个特征进行整合，形成混合特征集。
- 3) 以渗透率的一阶差分作为预测目标，采用 XGBoost 模型进行优化训练，确定最优参数组合，同时控制模型复杂度。
- 4) 对于未来 $t+1$ 月，将混合特征集输入训练好的 XGBoost 模型，得到渗透率一阶差分的预测值 $\hat{\Delta}_{t+1}$ ，然后通过累加的方式将差分预测值还原成渗透率水平值：

$$\hat{y}_{t+1} = \hat{y}_t + \hat{\Delta}_{t+1}$$

这一过程逐月推进，直至完成全部预测值的输出。

2.2.4. 模型评估指标

本文选取两个常用的模型预测精度评价指标：平均绝对误差(MAE, %)、均方根误差(RMSE, %)。

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

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

其中 y_i 为第 i 个样本的真实渗透率， \hat{y}_i 为对应的模型预测值， n 为测试集样本总数。其中 MAE、RMSE 值越小表示模型精度越好，越适用于本项研究，反之则越不适用。

3. 模型预测结果与分析

3.1. 实验设置

本文将收集的中国新能源汽车市场的月度数据分为两部分，其中 2023 年 1 月至 2025 年 6 月的数据作为训练集，2025 年 7 月至 12 月的数据作为测试集，并对 2026 年全年进行外推预测。本文选用单一 Prophet 模型作为基准参照，通过其与 Prophet-XGBoost 混合模型在预测性能上的对比，来突出混合模型的优越性。

3.2. 模型预测效果对比

两模型在测试集上的预测性能对比如表 1 所示。

Table 1. Comparison of prediction accuracy evaluation indicators

表 1. 预测精度评价指标对比

模型	<i>MAE</i>	<i>RMSE</i>
单一 Prophet	0.5%	0.55%
Prophet-XGBoost	0.38%	0.40%

由表结果可知 Prophet-XGBoost 的 *MAE* 和 *RMSE* 数值较单一 Prophet 均有降低，其中 *MAE* 降低 0.12%，*RMSE* 降低 0.15%，这表明混合模型有效提高了预测精度。

两模型对测试集的预测对比图如图 3 所示。

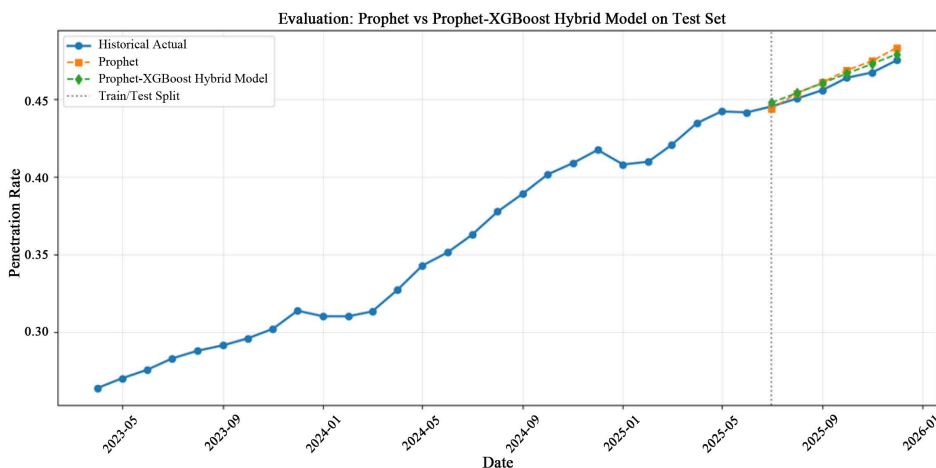


Figure 3. Comparison of predictions from the two models

图 3. 两模型预测对比图

图形直观地展示了两种模型预测值与真实值的对比，可以看出，单一 Prophet 模型整体拟合趋势较好，但对局部波动的捕捉能力有限，而 Prophet-XGBoost 模型的预测曲线能够较好地跟随渗透率实际变化曲线。

3.3. 影响因素重要性分析

本文利用 XGBoost 模型内置的重要性评估功能，基于增益输出了历史渗透率、时间特征、能源价格、充电桩数量、GDP 和政策变量等对预测结果的相对贡献度，结果如图 4 所示。

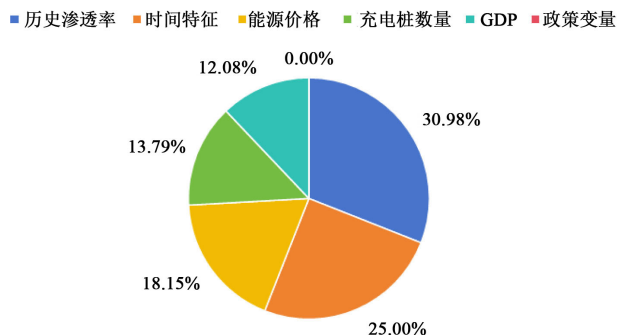


Figure 4. Factor importance analysis
图 4. 因素重要性分析

重要性分析显示：历史渗透率贡献最高，这表明市场具有显著的惯性，前期渗透率对后期渗透率走向具有重要影响；时间特征位居第二，表明渗透率存在较强的季节性和趋势性规律，Prophet 预测特征为 XGBoost 模型的预测提供了较为可靠的时序基准；能源价格、充电桩数量、GDP 贡献基本相近，体现了这些因素的基础性影响；双积分政策影响最小，可能源于政策变化频率低而 XGBoost 的增益评估倾向于优先选择变化频繁的特征。

3.4. 2026 年渗透率预测

基于训练好的 Prophet-XGBoost 混合模型，本文对 2026 年全年的新能源汽车渗透率预测结果如表 2 所示。

Table 2. 2026 forecast results
表 2. 2026 年预测结果

1 月	2 月	3 月	4 月	5 月	6 月
47.86%	48.27%	48.83%	49.42%	50.01%	50.60%
7 月	8 月	9 月	10 月	11 月	12 月
51.20%	51.79%	52.39%	52.99%	53.59%	54.20%

由结果可知，2026 年中国新能源汽车月度渗透率将持续攀升，到年底有望达到 54% 左右，新能源汽车市场主导地位将进一步巩固。

4. 研究不足与改进方向

尽管本文取得了较好的预测结果，但仍存在一些不足：1) 未考虑区域异质性对新能源汽车渗透率的影响；2) 对政策变量的处理较为粗糙，未能反映政策的动态调整过程。

针对以上不足未来可以从如下几个方面深入：1) 研究对象可转向区域层面，考虑地方政策等差异性，提供更具区域说服力的结果；2) 可采用时变参数模型等，将政策年度调整为动态输入，或结合政策文本量化政策强度指数，使模型能够动态响应政策变化；3) 基于该混合模型的特性，可将其推广至交通流量预测等领域，有望进一步增强该模型的适用性。

5. 结论

本文以中国新能源汽车渗透率为研究对象，构建了融合历史销量、能源价格、GDP、充电桩数据和

双积分政策数据的多源数据框架，并基于 Prophet 模型和 XGBoost 模型提出了一种综合两种单一模型优势的 Prophet-XGBoost 混合模型来对 2026 年全年中国新能源汽车渗透率进行预测。在数据处理阶段，先利用 Prophet 模型对渗透率这一时间序列进行分解，分解为趋势项、季节项和剩余项，有利于 XGBoost 模型在预测时提取时间序列中不同部分的特征，同时更好地融合其他四项外部特征以进行更高精度的预测。实验结果表明混合模型较单一 Prophet 模型有更好的预测精度，验证了混合模型的有效性。

参考文献

- [1] 毛华硕. 新能源汽车发展现状及销量预测方法研究综述[J]. 中国经贸导刊, 2025(4): 88-90.
- [2] 刘连义, 刘思峰, 吴利丰. 基于离散时间灰色幂模型的新能源汽车销售量预测[J]. 中国管理科学, 2024, 32(1): 106-114.
- [3] 高章, 刘杭, 赵凯. 纯电动重型车整车能耗影响因素及预测研究[J]. 兵器装备工程学报, 2025, 46(10): 89-97.
- [4] Alatawneh, A. and Torok, A. (2024) A Predictive Modeling Framework for Forecasting Cumulative Sales of Euro-Compliant, Battery-Electric and Autonomous Vehicles. *Decision Analytics Journal*, **11**, Article ID: 100483. <https://doi.org/10.1016/j.dajour.2024.100483>
- [5] Cavus, M., Ayan, H., Bell, M., et al. (2025) Deep Charge-Fusion Model: Advanced Hybrid Modelling for Predicting Electric Vehicle Charging Patterns with Socio-Demographic Considerations. *International Journal of Transportation Science and Technology*. <https://doi.org/10.1016/j.ijst.2025.03.002>
- [6] Dursun, N. (2025) Electric Vehicle (EV) Market Penetration in Countries with Rising Motorization Rates. *International Journal of Civil Engineering*, **23**, 461-480. <https://doi.org/10.1007/s40999-024-01039-z>
- [7] 史一龙, 黄淑兵, 赵北辰. 基于多源特征融合的新能源汽车运行实时风险预测研究[J]. 世界电动汽车期刊, 2025, 16(11): 626.
- [8] 张志淦, 申敏. 基于主题和情感的交叉网络舆情演化分析——以新能源汽车为例[J]. 情报探索, 2025(12): 90-97.
- [9] Chandra, M., Busch, P., Olguín, F.P. and Tal, G. (2025) Paths of Progress: Forecasting Global Electric Vehicle Demand Amidst Demographic and Economic Growth. *Transportation Research Part D: Transport and Environment*, **147**, Article ID: 104928. <https://doi.org/10.1016/j.trd.2025.104928>
- [10] Jain, M., Talwar, S., Rastogi, R. and Dhir, A. (2024) Policy Stimulation for the Electric Vehicle Industry: An Analysis of Mainstream Media Discourse. *Business Strategy and the Environment*, **33**, 5303-5324. <https://doi.org/10.1002/bse.3750>
- [11] Ramadoss, T.V., Lee, J.H., Davis, A.W., et al. (2025) Classifying Electric Vehicle Adopters and Forecasting Progress to Full Adoption. *npj Sustainable Mobility and Transport*, **2**, Article No. 32. <https://doi.org/10.1038/s44333-025-00049-1>
- [12] 翁剑成, 周慧缘, 张梦媛. 考虑城市与群体异质的新能源车激励策略有效性研究[J]. 交通运输系统工程与信息, 2025, 25(1): 2-14.
- [13] 贾文娟, 张卫林, 马自立. 基于助推理论的农村新能源汽车购买意愿研究[J]. 长安大学学报(自然科学版), 2025(4): 166-176.
- [14] 王静. 中国新能源汽车产业政策变迁及阶段特征研究[J]. 科研管理, 2025, 46(6): 146-156.
- [15] Hu, W., Liu, Y., Li, Y. and Chen, D. (2026) Behind the Wheel of Progress: Innovation Factor Allocation Patterns of Collaborative Innovation in Electric Vehicles. *Technological Forecasting and Social Change*, **223**, Article ID: 124412. <https://doi.org/10.1016/j.techfore.2025.124412>
- [16] Taylor, S.J. and Letham, B. (2018) Forecasting at Scale. *The American Statistician*, 2018, 72(1): 37-45. <https://doi.org/10.1080/00031305.2017.1380080>
- [17] 鲁铁定, 李祯. 基于 Prophet-XGBoost 模型的 GNSS 高程时间序列预测[J]. 大地测量与地球动力学, 2022, 42(9): 898-903.
- [18] 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>