

基于 Stacking 算法集成的资源型城市碳排放预测

张晓源¹, 金书亮²

1. 青海师范大学 数学与统计学院, 青海 西宁 810016

2. 郑州航空工业管理学院 经济学院, 河南 郑州 450046

DOI:10.61369/ASDS.2025080006

摘要 本文基于2006–2021年我国资源型城市的碳排放量及10个相关特征, 提出了一种融合CNN、RF和SVR的Stacking集成的碳排放达峰预测模型。首先, 采用随机森林筛选出影响碳排放的关键特征, 其次, 将关键特征输入至CNN、RF和SVR三种基学习器, 并通过交叉验证得到的各训练模型通过残差赋权方式组合起来; 最后, 将原始训练集与组合基学习器的预测结果共同输入元学习器进行训练与测试。实验表明, 与单一基学习器相比, 本文提出的Stacking集成模型在碳达峰预测中精度显著提高: 均方根误差和平均绝对误差分别降低9.39%和11.54%, 决定系数R²提升0.897个百分点, 有效验证了集成方法的优越性。

关键词 碳排放预测; 机器学习; Stacking 算法; 算法集成

Analysis and Prediction of Carbon Emission Influencing Factors in Resource-Based Cities Based on Stacking Algorithm Integration

Zhang Xiaoyuan¹, Jin Shuliang²

1. School of Mathematics and Statistics, Qinghai Normal University, Xining, Qinghai 810016

2. School of Economics, Zhengzhou University of Aeronautics, Zhengzhou, Henan 450046

Abstract This paper proposes a Stacking ensemble model for carbon emission peak prediction by integrating CNN, RF, and SVR, based on carbon emissions and 10 related features from China's resource-based cities between 2006 and 2021. First, a Random Forest is employed to screen key features influencing carbon emissions. Subsequently, these key features are input into three base learners—CNN, RF, and SVR—and the training models obtained through cross-validation are combined using residual weighting. Finally, both the original training set and the predictions from the combined base learners are input into a meta-learner for training and testing. Experiments demonstrate that compared to single base learners, the proposed Stacking ensemble model significantly improves prediction accuracy for carbon peak: the root mean square error and mean absolute error decrease by 9.39% and 11.54%, respectively, while the coefficient of determination R² increases by 0.897 percentage points, effectively validating the superiority of the ensemble approach.

Keywords carbon emissions prediction; machine learning; Stacking algorithms; algorithm integration

引言

碳排放是全球气候变化的主要诱因之一, 对生态系统和人类社会带来深远影响。《中国碳达峰碳中和进展报告(2023)》指出, 我国生态文明建设已进入以降碳为重点的关键时期, 但实现“双碳”目标仍面临新能源消纳困难、碳足迹制度不完善、国际规则约束等多重挑战。为推进碳市场建设, 2024年1月国务院通过《碳排放权交易管理条例》, 进一步强化制度保障。

我国各地区生态环境和经济发展差异显著, 城市碳排放水平不一。实现2030年碳达峰与2060年碳中和目标, 需所有城市协同减排。资源型城市作为国家能源与资源保障的核心主体, 长期支撑宏观经济平稳运行, 但也因持续资源开发面临产业路径锁定、高碳排放和高能耗等问题。尤其是矿业城市, 资源依赖易引发“资源诅咒”, 加剧生态退化, 形成经济与环境双重压力。因此, 识别资源型城市碳排放的影响因素并开展科学预测, 有助于从源头控制高碳排放, 缓解经济发展与减排之间的矛盾, 对落实“双碳”战略、应对气候变化、推动绿色低碳转型及经济高质量发展具有重要意义。

作者简介:

张晓源(2002-), 青海师范大学数学与统计学院, 硕士研究生, 研究方向为应用数理统计;

金书亮(2001-), 郑州航空工业管理学院经济学院, 硕士研究生, 研究方向为区域绿色创新。

在碳排放影响因素研究领域，已有诸多学者展开深入探讨。Yuan 等^[1]分析了人口、人均收入和能源结构等因素对碳排放的作用。穆佳音^[2]、施晓燕^[3]和崔艳芳^[4]则聚焦于资源型城市，揭示了其在碳排放方面的特殊性。李安和尹逊之^[5]研究发现，人口规模、人均GDP和能源强度与我国农村居民生活能源消费碳排放呈显著正相关。赵一航和赵会茹^[6]指出，碳市场与化石能源市场在试点阶段存在单向均值溢出效应，凸显能源结构调整的重要性。杨帆和甄江红^[7]、肖德和张媛^[8]探讨了经济增长与碳排放之间的互动机制。杨浩昌等^[9]发现发明型和改进型绿色技术创新显著促进碳减排。Khanna^[10]比较分析了低碳政策支持手段，强调政策对低碳发展的引导作用。要蓉蓉等^[11]基于“目标—资源—信息”传导机制，证实低碳试点政策有效降低了试点地区的碳排放强度。

近年来，学者们采用多种模型对碳排放趋势进行预测。ARIMA^[12]和LEAP^[13]模型被广泛用于相关研究。神经网络模型，尤其是卷积神经网络（CNN），也在该领域展现出较大潜力，如焦柳丹^[14]基于CNN构建了交通运输碳排放预测模型。随机森林（RF）同样受到关注，王伟同等^[15]利用RF有效预测了锅炉氮氧化物排放，刘慧甜和胡大伟^[16]则采用RF、XGBoost及KNN对比分析了30个省份的交通碳排放。在建筑碳排放方面，Chu^[17]采用支持向量回归（SVR）模型，其泛化能力和鲁棒性优于多元线性回归，取得了更高的决定系数和更低的误差。武冰等^[18]进一步提出了基于模型平均的随机森林模型（RF-MA），通过权重组优化对我国各省交通碳排放进行预测。碳排放预测问题复杂多变，单一模型往往难以应对。为此，部分学者开始采用组合或混合模型进行预测。例如，Liu等^[19]融合了自回归移动平均模型、二阶多项式回归和灰色模型；马良玉^[20]和杏艳^[21]则通过集成多个机器学习算法，分别实现了对风电机组异常温度和城市大气污染物排放的精准预测，不仅提升了模型性能，也增强了算法的鲁棒性和适应性。这些研究表明，融合多种方法有助于更准确预测碳排放趋势。

现有研究多聚焦于某一行业或区域，且常采用传统预测模型。本文以资源型城市为研究对象，这类城市以煤炭、钢铁、有色、建材等重工业为主，碳排放总量大、强度高，是实现“双碳”目标的关键区域。其碳排放受多种复杂因素影响，包括一次能源结构、产业结构、技术水平等社会经济维度，以及资源禀赋、地理区位、政策强度等空间异质因素。传统单一模型难以充分捕捉这些非线性、高维交互和区域差异特征，导致影响因素识别不全、预测精度有限，影响减排政策的科学制定。为此，本文基于2006—2021年我国112个资源型城市的碳排放数据及10个相关特征，构建了一种基于Stacking集成策略的碳排放预测模型，融合了CNN、RF和SVR三种算法，并与单一模型预测结果进行对比分析。

一、资源型城市碳排放指标体系构建

（一）指标体系设计

在全球气候变化背景下，资源型城市作为主要碳排放源，其减排对实现气候目标至关重要。本研究借鉴任宏洋^[22]等方法，基于随机森林模型，从劳动力、经济、技术、绿化、资源消费及产业结构6个维度选取10项指标，分析资源型城市碳排放影响因素并进行预测，以期为决策提供科学支撑，助力城市可持续发展。

（二）数据来源

根据《全国资源型城市可持续发展规划（2013—2020年）》界定的126个地级资源型城市，本文综合考虑数据连续性与完整性，最终选取2006—2021年间112个城市作为研究样本，其空间分布如图1所示。对个别缺失数据，采用Stata软件结合线性插值与移动平均法进行补全。所有原始数据来源于《中国统计年鉴》、各省（区、市）统计年鉴及EPS数据库。

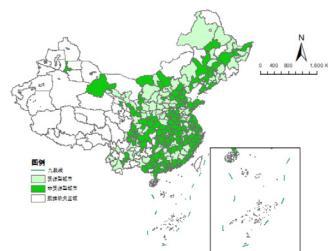


图1：研究区域分布图

Figure 1: Distribution Map of the Study Area

二、基于Stacking算法集成的预测模型

（一）机器学习算法

本文以双层Stacking框架为核心，首先从若干经典机器学习算法中择优几类基学习器，再以某一算法作为次级学习器对基学习器的输出进行二次集成，通过双层结构实现预测精度的提升，下面给出三类基本的机器学习算法。

1. 随机森林（RF）算法

随机森林是一种基于决策树集成的有监督机器学习算法^[23]，广泛应用于分类和回归任务，具有良好的可解释性，特别适用于变量重要性评估与特征选择。该算法通过在每棵树的生成过程中对数据和特征进行最优分割，有效降低过拟合风险，同时能够输出特征重要性排序，为高维特征筛选提供量化依据，有助于识别关键变量并剔除冗余信息。基于以上优点，本文采用随机森林算法评估资源型城市与非资源型城市碳排放影响因素的重要性，依据各变量在模型中的使用频率及对预测的贡献程度来确定其影响程度。其理论公式为：

$$H(x) = \frac{1}{T} \sum_{i=1}^T w_i h_i(x) \quad (1)$$

其中， w_i 是个体学习器 h_i 的权重。

2. 卷积神经网络（CNN）算法

卷积神经网络（CNN）是一种能够自动提取特征并进行模式识别的深度学习模型^[24]。它通过卷积运算和层次化结构自动学习

数据的抽象特征，无需人工设计特征提取器，并借助降维减少参数规模，扩展了深度学习在预测任务中的应用潜力。本文采用 CNN 构建预测模型，用于分析资源型与非资源型城市碳排放与其影响因素之间的关系，并对碳排放进行预测。模型构建步骤如下：

(1) 输入层：以 2006–2021 年城市碳排放影响因素数据作为输入；

(2) 卷积特征提取：通过卷积核在输入数据上滑动进行局部特征提取，计算如下：

$$C = f(X \otimes W_c + b_c) \quad (2)$$

其中， C 为卷积输出， f 为激活函数， X 为输入， \otimes 表示卷积操作， W_c 是权重矩阵， b_c 为偏置项；

(3) 池化降维：采用最大池化提取每个区域中的最大值，在保留关键特征的同时降低维度；

$$P = \text{maxpooling}(C) \quad (3)$$

(4) Dropout 正则化：在训练中随机丢弃部分神经元连接，以抑制过拟合；

(5) 全连接层：将特征展平并整合，完成高层特征映射；

(6) 输出层：生成最终预测结果。

3. 支持向量回归 (SVR) 算法

支持向量回归 (SVR) 由 Cortes^[25] 等人于 1995 年提出，在处理非线性回归问题中表现出良好性能，并已在多个领域获得广泛应用。SVR 建模过程中，样本通常划分为训练集与测试集：训练样本通过非线性映射转换到高维特征空间，以构建最优决策函数；测试样本则用于验证模型泛化能力。其回归函数可表示为：

$$f(\mu) = \langle E \cdot \Phi(\mu) \rangle + \theta \quad (4)$$

其中， μ 为输入变量； $\Phi(\mu)$ 表示从输入空间到高维特征空间的非线性映射； θ 为偏置项。

为估计回归函数，需引入 ϵ 不敏感损失函数，并将其转化为以下优化问题：

$$R_{\min} = \frac{1}{2} \|E\|^2 + \tau \sum_{i=1}^m (\xi_i + \xi_i^*) \quad (5)$$

式中： τ 为惩罚因子； ξ_i 和 ξ_i^* 为 l 个变量中第 i 个的松弛因子。

约束条件如下：

$$\begin{cases} E^T \Phi(\mu_i) + \theta_i - y_i \leq \epsilon + \xi_i \\ y_i - E^T \Phi(\mu_i) - \theta_i \leq \epsilon + \xi_i^* \\ \xi_i, \xi_i^* \geq 0, i = 1, \dots, l \end{cases} \quad (6)$$

式中： y_i 为 l 个变量中第 i 个的输出变量； θ_i 为 l 个变量中第 i 个的偏置项。

基于 Wolf 对偶理论，可引入拉格朗日乘子 α 将原问题转化为对偶形式，其约束为

$$\begin{cases} \sum_{i=1}^l (\alpha_i - \alpha_i^*) = 0 \\ 0 \leq \alpha_i, \alpha_i^* \leq \theta \end{cases} \quad (7)$$

SVR 预测模型可以通过二次规划求解：

$$f(\mu) = \sum_{i=1}^l (\alpha_i - \alpha_i^*) K(\mu_i, \mu) + \theta \quad (8)$$

式中 $K(\mu, \mu)$ 为 SVR 模型的核函数。

在支持向量机算法中，通常可选的核函数包括线性核、多项式核、径向基核函数 (RBF) 和 sigmoid 函数^[26]。由于 RBF 核函数具有广泛的收敛域，因此，RBF 被广泛作为 SVR 模型的核函数。

(二) 基于 Stacking 算法的集成预测模型

Stacking 集成学习^[27] 采用双层结构：第一层包含多个基学习器（不同算法），第二层为单个元学习器（结构见图 2）。训练时，通过 K 折交叉验证将训练集划分为多个子集，并将基学习器的交叉验证预测结果用于训练元学习器。测试时，各基学习器基于 K 折建立的模型对测试集进行预测，并将不同模型的预测结果取平均后输入元学习器，得到最终结果。吴佳润^[28] 对该方法进行了改进：将每个基学习器在 K 折交叉验证中对测试集的 K 个预测结果通过残差赋值进行加权组合，以此降低对噪声的敏感性，并将原始训练集与加权后的预测结果共同作为元学习器的输入，从而更好地保留原始数据信息。

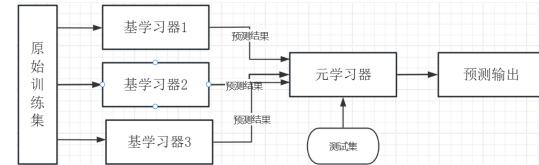


图 2: Stacking 集成算法基本框架

Figure 2: Basic Framework of the Stacking Integration Algorithm

本文提出的多算法集成碳排放预测模型采用两层 Stacking 结构：第一层选取 RF、CNN 和 SVR 作为基学习器，并从中选择性能最优者作为第二层的元学习器。训练阶段采用五折交叉验证，将各基学习器的验证预测结果输入元学习器进行训练。测试阶段中，各基学习器对测试集进行预测，基于预测值 \hat{y}_{base} 与真实值的残差 e_i 计算权重 w_i （残差与权重反比），通过加权融合得到 $\hat{y}_{combined}$ 。为保持数据完整性，将原始特征与加权预测结果共同输入元学习器，生成最终预测，流程详见图 3。

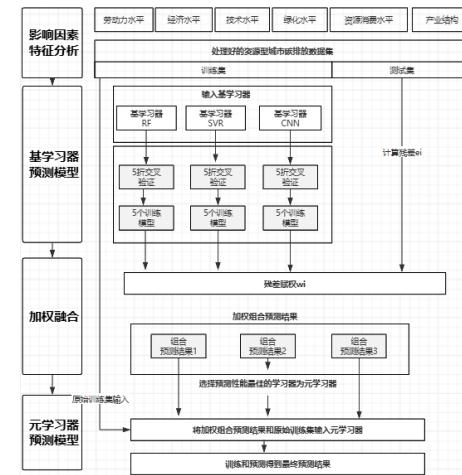


图 3: 多算法集成的碳排放预测流程图

Figure 3: Carbon Emission Prediction Flowchart Using Multi-Algorithm Integration

三、城市碳排放影响因素分析

(一) 影响因素分析

本实验基于 MATLAB 平台, 采用随机森林模型 (决策树数目 110, 最小叶子数为 1) 对资源型城市碳排放影响因素的重要性进行分析, 结果如图 4(a) 所示。产业结构高级化、城镇化率和常住人口的重要性最高, 表明这些因素与碳排放具有显著关联; 绿地面积、建成区覆盖率和固定资产投资的影响较小; 电力消费量、从业人员、能源效率及人均 GDP 的影响处于中等水平。

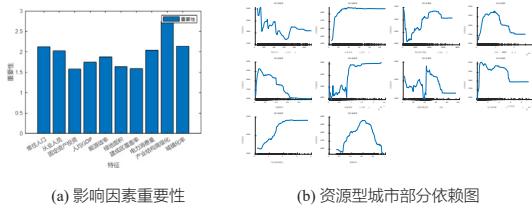


图4: 资源型城市影响因素重要性及其依赖图

Figure 4: Importance of Factors Affecting Resource-Based Cities and Their Dependency Diagram

部分依赖图 (图 4(b)) 揭示了各因素与碳排放之间的响应关系。产业结构高级化、城镇化率和常住人口是关键因素: 产业结构高级化与碳排放呈非线性正相关, 高技术产业占比在 10% - 60% 时碳排放增速显著, 超过 60% 后趋于平缓; 城镇化率低于 58% 时碳排放随之上升, 超过后影响减弱; 常住人口在达 0.52×10^7 人时抑制作用最强, 之后转为促进并渐趋平缓。能源效率、电力消费、从业人员和人均 GDP 具有中等影响; 绿地面积、建成区覆盖率和固定资产投资的影响较小。

基于随机森林特征重要性分析结果, 本文删去重要性最低的三个特征 (见表 1), 将剩余变量作为 Stacking 集成模型中基学习器的输入。该处理可降低模型维度与计算复杂度, 有助于提升预测准确性与鲁棒性。

Table 1: Indicators of Carbon Emission Influencing Factors in Resource-Based Cities

表1: 资源型城市的碳排放影响因素指标

指标分类	指标名称	资源型城市
劳动力水平	从业人员	✓
	常住人口	✓
经济水平	固定资产投资额	
	人均 GDP	✓
技术水平	城镇化率	✓
	能源效率	✓
绿化水平	绿地面积	
	建成区覆盖率	
资源消费水平	城市电力消费量	✓
产业结构	产业结构高级化	✓

(二) 评估指标

均方根误差 (E_{RMSE})、平均绝对误差 (E_{MAE}) 以及决定系数 (R^2) 是衡量模型的预测性能的三个关键指标。其中 E_{RMSE} 表示预测值与实际值之差的平方的期望值的平方根, 反映模型预测误差的标准偏差, E_{RMSE} 值越小, 表明模型预测结果越稳定; E_{MAE} 是预测误差的绝对值的平均数, 直接反映模型预测的精确度, E_{MAE} 越小, 说明模型的预测精确度越高; 决定系数 R^2 定义为回归平方和占总

平方和的比值, 反映回归模型中解释变量对因变量变化的解释程度, R^2 值越接近 1, 表示模型对数据的拟合效果越好。三者的计算公式 (9)、(10)、(11) 如下所示。

$$E_{RMSE} = \sqrt{1/n \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (9)$$

$$E_{MAE} = 1/n \sum_{i=1}^n |y_i - \hat{y}_i| \quad (10)$$

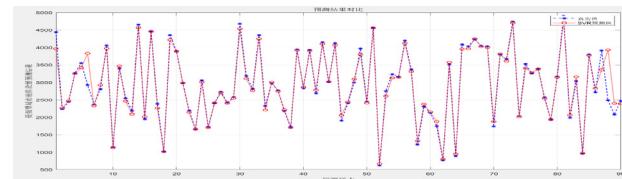
$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (11)$$

式中: y_i 为城市碳排放的实际值; \hat{y}_i 为城市碳排放的预测值; \bar{y} 为城市碳排放的平均值; n 为样本数量。

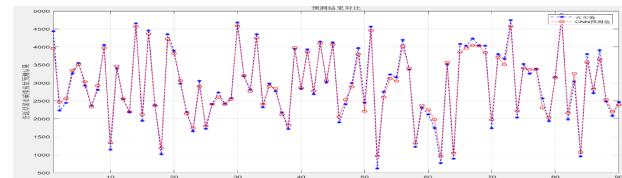
四、资源型城市碳排放量预测实证比较

(一) 基于 Stacking 算法集成的底层模型预测

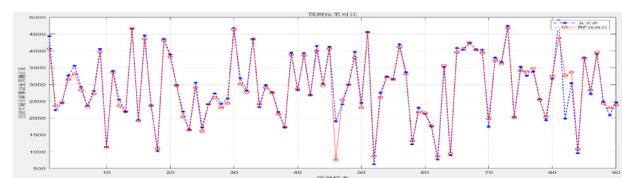
本文按照 9:1 的比例将资源型城市数据随机划分为原始训练集和测试集, 并分别输入 RF、CNN 和 SVR 三种基学习器进行训练。RF 设置决策树数目为 110, 最小叶子数为 1; CNN 输入数据规模为 $7 \times 1 \times 1$, 卷积核大小为 3×1 , 采用批归一化层、Relu 激活函数和 SGDM 梯度下降算法, 迭代 800 次, 初始学习率为 0.01。三种基学习器的预测结果如图 5 所示。测试集上的评估指标显示 (表 2), 所有单一模型的均方根误差 (ERMSE) 和平均绝对误差 (EMAE) 均分别低于 0.09466 和 0.09444, 决定系数 (R^2) 高于 0.95420, 表明各模型均能较准确预测碳排放量。其中 CNN 表现最优, 其 ERMSE、EMAE 和 R^2 分别为 0.09001、0.00615 和 0.97250。



(a) RF 模型预测值与真实值对比图



(b) SVR 模型预测值与真实值对比图



(c) CNN 模型预测值与真实值对比图

图5: 底层模型预测值与真实值对比图

Figure 5: Comparison of Predicted Values and Actual Values for the Underlying Model

Table 2 :Evaluation Metrics for the Underlying Base Learner
表2: 底层基学习器的评价指标

基学习器	R^2	E_{MAE}	E_{RMSE}
RF	0.95420	0.10966	0.01410
SVR	0.95983	0.10316	0.01044
CNN	0.97250	0.09001	0.00615

(二) 单一基学习器的残差赋值

集成学习方法通过融合多个基学习器的预测结果，并利用元学习器学习如何更优地组合这些预测，以提升模型性能。改进的 Stacking 算法^[28] 在 K 折交叉验证中，对每个基学习器的 K 个测试集预测结果进行残差加权融合，再将加权结果作为元学习器的输入，从而抑制误差较大的折对结果的影响，降低模型对噪声的敏感性。具体步骤如下：

1. 计算残差：设三个基模型 M_1, M_2, M_3 ，在五折交叉验证中，每个模型基于训练集生成预测值，并与真实值比较，得到每折的平方误差 e_i （其中 $i=1, 2, 3, 4, 5$ ）。

2. 残差转化为权重：归一化残差的倒数，即权重与残差成反比，具体公式如下：

$$w_i = \frac{1/e_i}{\sum_{i=1}^K (1/e_i)}, i=1, 2, 3, 4, 5 \quad (12)$$

3. 加权组合预测结果。对每个基学习器的 K 个模型预测结果按权重加权平均：

$$\hat{y}_{combined} = \sum_{i=1}^K w_i y_i \quad (13)$$

其中， \hat{y}_i 是第 i 个模型的预测值， w_i 是其权重。

该方法通过降低残差较大模型的权重，有效抑制噪声和异常值的影响，提升了集成模型的稳定性和鲁棒性。

(三) 基于 Stacking 算法集成的模型预测

在前述分析基础上，由于 CNN 在三种基学习器中表现出最小的碳排放预测误差，因此被选为 Stacking 集成模型的元学习器。以 RF、SVR 和 CNN 作为基学习器，利用原始训练集构建 Stacking 集成模型，并对测试集进行预测，结果如图6所示。进一步分析预测值与实际值的相关性可知（表3），Stacking 集成算

法的决定系数（ R^2 ）达到 0.98122，表明其预测精度优于任一单一基学习器。同时，与表现最优的基学习器 CNN 相比，该集成方法的均方根误差（ E_{RMSE} ）和平均绝对误差（ E_{MAE} ）分别降低了 9.39% 和 11.54%， R^2 提高了 0.897 个百分点。结果证明，本文所提出的 Stacking 集成算法在资源型城市碳排放预测中具有良好效果，更适合该类任务的建模与分析。

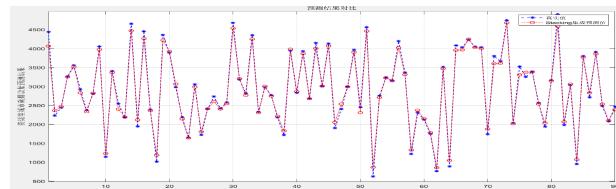


图6: Stacking 集成模型预测值与真实值对比图

Figure 6: Comparison of Stacking Ensemble Model Predictions and Actual Values

Table 3: Evaluation Metrics for Ensemble Models Based on the Stacking Algorithm

表3: 基于 Stacking 算法集成模型的评价指标

基学习器	R^2	E_{MAE}	E_{RMSE}
RF	0.95420	0.10966	0.01410
SVR	0.95983	0.10316	0.01044
CNN	0.97250	0.09001	0.00615
Stacking 集成	0.98122	0.08156	0.00544

五、结论

本文针对资源型城市碳排放预测问题，提出了一种基于 Stacking 算法的集成模型，通过融合 CNN、RF 和 SVR 构建双层预测框架，显著提升了预测准确性与鲁棒性。首先采用随机森林（RF）提取关键特征，识别出主要影响因素；进而将筛选特征输入三个基学习器，通过五折交叉验证训练并以残差赋权方式组合预测结果，最终与原始数据共同训练元学习器。实验表明，该方法不仅有效提高了预测精度、降低过拟合风险，且其 RMSE 和 MAE 较表现最优的 CNN 模型分别降低 9.39% 和 11.54%， R^2 提升 0.897%，显著优于单一模型。本研究为资源型城市碳排放预测提供了高效方法参考，未来可进一步探索更多算法与复杂集成策略以持续提升性能。

参考文献

- [1] Yuan R, Rodrigues J F D, Behrens P. Driving forces of household carbon emissions in China: A spatial decomposition analysis[J]. Journal of cleaner production, 2019, 233:932.
- [2] 穆佳音, 王金满, 刘彪, 等. 资源型城市碳排放演变及影响因素研究进展 [J/OL]. 煤炭学报 :1-15[2024-04-25].
- [3] 施晓燕, 赵有益, 王音璠, 等. 黄河流域资源型城市碳达峰情景模拟研究 [J]. 西北师范大学学报 (自然科学版), 2024, 60(01):106-114.
- [4] 崔艳芳, 张国兴. 黄河流域资源型城市碳排放影响因素与达峰预测研究 [J]. 人民黄河, 2023, 45(02):9-14.
- [5] 李安, 尹逊之. 我国农村居民生活能源消费碳排放影响因素分析 [J]. 山东社会科学, 2024(02):169-176.
- [6] 赵一航, 赵会茹. 我国碳市场与化石能源市场溢出效应研究——基于 VAR-GARCH-BEKK 模型的分析 [J]. 工业技术经济, 2024, 43(04):36-45.
- [7] 杨帆, 瓢江红. “双碳”目标下典型地区减污、降碳与经济增长关系研究: 以黄河流域城市群为例 [J]. 环境科学研究, 2023, 36(11):2050-2064.
- [8] 肖德, 张媛. 经济增长、能源消费与二氧化碳排放的互动关系——基于动态面板联立方程的估计 [J]. 经济问题探索, 2016(09):29-39.
- [9] 杨浩昌, 钟时权, 李廉水. 绿色技术创新与碳排放效率: 影响机制及回弹效应 [J]. 科技进步与对策, 2023, 40(08):99-107.

[10] Khanna N, Fridley D, Hong L. China's pilot low-carbon city initiative: A comparative assessment of national goals and local plans[J]. *Sustainable Cities & Society*, 2014, 12(7):110–121.

[11] 要蓉蓉, 郑石明, 邹克. 低碳试点政策传导机制及其效应——基于多维数据的实证研究 [J]. *公共管理与政策评论*, 2024, 13(02):131–149.

[12] 徐丽, 曲建升, 李恒吉, 等. 西北地区居民生活碳排放现状分析及预测 [J]. *干旱区地理*, 2019, 42(05):1166–1175.

[13] 苗安康, 袁越, 吴涵, 等. 中国省域碳达峰路径与政策 [J]. *环境科学*, 2023, 44(08):4623–4636.

[14] 焦柳丹, 刘莹, 吴雅, 等. 基于卷积神经网络的交通运输业碳排放预测研究 [J/OL]. *铁道运输与经济* :1–9[2024–04–29].

[15] 王伟同, 范海东, 梁成思, 等. 基于随机森林算法的对冲锅炉出口 NO_x 排放量预测模型研究 [J]. *热力发电*, 2022, 51(4):96–104.

[16] 刘慧甜, 胡大伟. 基于机器学习的交通碳排放预测模型构建与分析 [J/OL]. *环境科学* :1–17[2024–04–29].

[17] Chu, X. and Zhao, R. A building carbon emission prediction model by PSO-SVR method under multi-criteria evaluation. *Journal of Intelligent & Fuzzy Systems*, 2021, 41, 7473–7484.

[18] 武冰. 基于随机森林的模型平均预测方法研究. *北京大学学报(自然科学版)*, 2021, 57(06): 1133–1142.

[19] Liu, Zhong H, Zhao E, et al. Can China realize its carbon emission reduction goal in 2020: from the perspective of thermal power development[J]. *Applied Energy*, 2014, 124:199.

[20] 马良玉, 耿妍竹, 梁书源, 等. 基于 Stacking 多模型融合的风电机组齿轮箱油池温度异常预警 [J]. *中国电机工程学报*, 2023, 43(增刊1):242–251.

[21] 杏艳, 宋新宇, 李飞, 李茂刚, 郭琦, 张霖琳. (2025). 集成学习算法结合经验模型预测城市大气污染物排放. *中国环境监测*, 41(3), 14–23.

[22] 任宏洋, 杜若岚, 谢贵林, 等. 中国碳排放影响因素及识别方法研究现状 [J]. *环境工程*, 2023, 41(10):195–203+244.

[23] 曹桃云. 基于随机森林的变量重要性研究 [J]. *统计与决策*, 2022, 38(04):60–63.

[24] 谢煥增, 黄凌泽, 罗烨, 等. 基于多尺度卷积神经网络的 CRISPR/Cas9 脱靶预测方法 [J]. *生物工程学报*, 2024, 40(03):858–876.

[25] Cortes C, Vapnik V. Support-vector networks[J]. *Machine Learning*, 1995, 20(3):273–297.

[26] Elibey M S. Sea wave parameters prediction by support vector machine using a genetic algorithm[J]. *Journal of Coastal Research*, 2015, 31(4): 892–899.

[27] 李阳, 黄伟, 席建忠. 基于 Stacking 算法集成模型的电厂 NO_x 排放预测 [J]. *热能动力工程*, 2021, 36(5):73–81.

[28] 吴佳润, 林宇亮, 邢浩, 宁曦. 基于改进 Stacking 多模型融合的高速公路隧道建设碳排放预测模型. *湖南大学学报(自然科学版)*, 2025, 52(5), 57–65.