**Modeling and Simulation** 建模与仿真, 2025, 14(8), 337-346
Published Online August 2025 in Hans. https://www.hanspub.org/journal/mos
https://doi.org/10.12677/mos.2025.148572

Hans汉斯

# 基于XGBoost的机电安装工程工期预测模型研究

**尹　磊**

上海理工大学管理学院，上海

收稿日期：2025年7月27日；录用日期：2025年8月19日；发布日期：2025年8月27日

## 摘　要

**随着机电安装工程复杂性和管理要求的不断提升，精准的工期预测成为优化项目管理和降低成本的关键。本文基于机电安装工程的实际数据，采用XGBoost算法构建工期预测模型，通过特征工程、模型训练和性能验证，实现了高精度的工期预测。研究结合数据样本特征，分析了施工效率、工序工程量和关键路径标记等关键因素对工期的影响，并通过与随机森林和线性回归模型的对比验证了XGBoost的优越性能。结果表明，该模型在均方误差(RMSE)和平均绝对误差(MAE)等指标上表现优异，特征重要性分析揭示了影响工期的核心因素，为工程项目优化提供了数据驱动的决策依据。**

## 关键词

**机电安装，工期预测，XGBoost，回归模型，特征重要性**

# XGBoost-Based Duration Prediction Model for Mechanical and Electrical Installation Project

**Lei Yin**

Business School of University of Shanghai for Science and Technology, Shanghai

Received: Jul. 27th, 2025; accepted: Aug. 19th, 2025; published: Aug. 27th, 2025

## Abstract

**With the increasing complexity and management requirements of Electrical Installation Project, accurate construction duration prediction has become a key factor in optimizing project management and reducing costs. This study develops a duration prediction model based on the XGBoost**

文章引用: 尹磊. 基于 XGBoost 的机电安装工程工期预测模型研究[J]. 建模与仿真, 2025, 14(8): 337-346.
DOI: 10.12677/mos.2025.148572

**algorithm using real-world data from MEP installations. Through feature engineering, model training, and performance validation, the proposed model achieves high-accuracy predictions. By analyzing sample characteristics, the study identifies the impact of key factors such as construction efficiency, task workload, and critical path indicators on project duration. Comparative experiments with Random Forest and Linear Regression models demonstrate the superior performance of XGBoost. The results show that the model performs well in terms of Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE). Additionally, feature importance analysis reveals the core factors affecting construction duration, providing data-driven insights for project optimization and decision-making.**

## Keywords

**Mechanical and Electrical Installation Project, Duration Prediction, XGBoost, Regression Model, Feature Importance**



## 1. 引言

机电安装工程(Mechanical and Electrical Installation Project，简称 MEP)作为建筑项目的重要组成部分，涉及水、电、风等多个专业工序，其工期管理直接影响项目的整体进度和成本。传统的工期预测方法，如关键路径法(CPM)和计划评审技术(PERT)，依赖专家经验和静态假设，难以应对工序间的复杂依赖关系和动态变化。

近年来，机器学习技术因其强大的非线性建模能力，在工程管理领域得到了广泛应用。其中，Ali Akbar ForouzeshNejad 建立的 XGBoost-SA 模型可通过结合活动网络复杂性和不确定性特征，显著提高时间和成本预测的准确性，表明此方法优于传统方法[1]。Guangkai Huang (2024)等提出了一种基于 XGBoost 的多目标预测模型，结合随机森林贝叶斯优化，成功预测超深基坑施工引起的地面沉降、站点沉降和基坑变形，展现了高精度和良好泛化能力，适用于复杂地质条件下的工程监测[2]。周海浪等人(2023)提出了一种基于加权集成学习的车间生产进度异常预测方法，通过结合 XGBoost 等机器学习模型，优化实时监控和异常检测的精度，显著提升自动化制造系统的生产调度效率[3]。韩峰等人(2023)提出了一种基于改进 XGBoost 的隧道围岩超前分类方法，通过结合 ANDSMOTE 方法处理不平衡数据和 HHO 算法优化 XGBoost 模型参数，显著提升了高维数据集下围岩分类的精度和效率，为隧洞工程施工仿真提供了可靠支持[4]。李进友等(2023)在风电大数据背景下，提出一种基于 XGBoost-GRNN 的风功率预测算法，并以内蒙古塞罕坝风电场 20 台风电机组为例进行实验分析，结果表明该方法的预测准确性更优于传统预测方法[5]。杜庆峰等(2022)基于珠三角水资源配置工程掘进数据，采用均值滤波降噪与 XGBoost 算法构建盾构机掘进速度预测模型，结果显示预测准确率高达 99.97%，显著优于传统方法，验证了该方法在时序数据建模与智能预测中的有效性与可行性[6]。

由此可见，XGBoost (eXtreme Gradient Boosting)以其高效的梯度提升算法和鲁棒性，成为时间序列预测和回归任务的优选模型。现有研究多集中于土建工程的工期预测，而针对机电安装工程的预测模型研究较少，且鲜有对特征重要性的深入分析。本研究基于实际项目数据，构建了基于 XGBoost 的工期预测模型，旨在提高机电安装工程的进度管理精度，并通过特征重要性分析揭示关键影响因素，为项目优化

提供指导。

## 2. 预测模型构建与流程设计

本文基于 XGBoost 算法，利用项目中实际工序的多维特征数据，构建工期预测模型。通过与随机森林和线性回归模型进行对比，系统评估模型的预测性能。模型训练过程中采用网格搜索实现超参数优化，提升模型泛化能力。本文还对模型的特征重要性进行分析，探讨影响工期的关键因素，并通过多种图形化手段展示预测结果及误差分布，增强模型解释性。

### 2.1. 特征定义与数据预处理

本研究所用数据来源于 8 个住宅工程项目。针对每个项目的机电安装工程，通过工作分解结构(WBS)进行详细划分，最终提取出共计 23 个典型工序(编号 A 至 W)。在特征变量的选取方面，重点考虑了施工组织、计划安排及资源配置中对施工节奏和进度具有直接影响的关键因素。为确保模型具备预测能力，所选变量需在施工前即可确定，具备先验性。因此，选取“工序类型”“工程量”“作业人数”“施工效率”“计划开始时间”“是否为关键路径”以及“前置工序数”等 7 个变量作为模型输入，相关数据样表见表 1。

其中，“施工效率”虽为基于经验的估计值，但在数据处理过程中已排除对实际工期的直接依赖，仅作为专家经验的转化指标使用，确保其不与目标变量产生信息泄漏。为增强特征变量定义的准确性，本文对各变量含义进行如下界定。

工序类型编码 $X_1$ (Operation Type Code)：将施工中不同专业工序(如强电、弱电、消防、暖通等)进行类别编码，采用 Label Encoding 方法进行数值转换，数据来源于工程任务划分表。

工序工程量 $X_2$ (Task Workload)：表示某一施工工序的总工作量，单位可为米、平方米或立方米等，依据施工图纸或 BIM 模型进行工程量提取。

作业人数 $X_3$ (Number of Workers)：表示参与该工序施工的平均作业人数，由施工组织设计或现场日志记录获得，单位为“人”。

施工效率 $X_4$ (Construction Efficiency)：指单位作业人员在单位时间内完成的工程量，数据来源于施工日志、企业标准工时定额，其计算公式如下：

$$X_4 = \frac{X_2}{X_3 \cdot T_h} \tag{1}$$

其中 $T_n$ 为该类型工序的历史经验工期估计值，用于避免引入实际工期变量。单位为工程量/(人·天)。

计划开始时间 $X_5$ (Planned Start Time)：表示该工序相对于项目起始的预定开工时间，一般以天为单位，来源于施工总进度计划。

关键路径标记 $X_6$ (Critical Path Flag)：二值变量(0/1)，由 CPM 算法或项目管理软件识别，表示该工序是否处于关键路径上。

前置工序数 $X_7$ (Number of Predecessors)：该工序在网络计划图中的直接前置工序数量，来源于施工网络图。

实际工期 $Y$ (Actual Duration)：表示每项工序的实际施工天数，可由施工进度记录统计得出，单位为“天”，为模型的目标预测变量($Y$)。

为进一步保障数据质量，首先对原始数据进行了缺失值处理，剔除存在信息缺失的不完整样本。在数据预处理阶段，将分类特征“工序类型”通过标签编码(LabelEncoder)方式转换为数值型变量，同时将关键路径标记转化为二进制变量(0 表示非关键路径，1 表示关键路径)。最终确定上述 7 个特征变量(详见

表 2)作为模型输入，实际工期作为目标变量。数据集按 8:2 的比例划分为训练集和测试集，用于模型训练与性能评估。

**Table 1.** Sample data table for MEP installation projects
**表 1.** 机电安装工程数据样表

| 工序编号 | 工序类型 | 工序工程量 | 作业人数 | 施工效率 | 计划开始时间 | 关键路径标记 | 前置工序数 | 实际工期 |
|---|---|---|---|---|---|---|---|---|
| A | 水电 | 29189.37 | 10 | 20.84 | 1 | 1 | 0 | 140 |
| B | 土建 | 300.00 | 12 | 2.5 | 141 | 1 | 1 | 10 |
| C | 电 | 22860.19 | 4 | 40.82 | 1 | 0 | 0 | 140 |
| D | 电 | 14 | 4 | 1.75 | 141 | 0 | 1 | 2 |
| E | 土建 | 468.00 | 12 | 1.3 | 151 | 1 | 0 | 30 |
| F | 电 | 19459.58 | 6 | 72.07 | 181 | 1 | 3 | 45 |
| G | 土建 | 9000.00 | 12 | 25 | 226 | 1 | 4 | 30 |
| H | 风 | 6571.95 | 8 | 13.69 | 256 | 0 | 5 | 60 |
| I | 风 | 40.00 | 4 | 0.5 | 316 | 0 | 6 | 20 |
| J | 电 | 13648.45 | 8 | 37.91 | 256 | 1 | 5 | 45 |
| K | 水 | 857.36 | 4 | 10.72 | 226 | 0 | 7 | 20 |
| L | 水 | 4831.12 | 4 | 60.39 | 226 | 0 | 6 | 20 |
| M | 电 | 73066.56 | 6 | 811.85 | 226 | 0 | 5 | 15 |
| N | 电 | 39346.26 | 10 | 131.15 | 301 | 1 | 6 | 30 |
| O | 电 | 196 | 4 | 2.45 | 331 | 1 | 7 | 20 |
| P | 电 | 324 | 4 | 5.4 | 241 | 0 | 6 | 15 |
| Q | 电 | 1486 | 6 | 12.38 | 351 | 1 | 7 | 20 |
| R | 电 | 13648.45 | 4 | 426.51 | 351 | 0 | 7 | 8 |
| S | 水 | 1449.34 | 2 | 103.52 | 256 | 0 | 5 | 7 |
| T | 水 | 80 | 4 | 1 | 256 | 0 | 7 | 20 |
| U | 水 | 41.67 | 2 | 1.39 | 256 | 0 | 5 | 15 |
| V | 水 | 20 | 2 | 2.5 | 351 | 0 | 8 | 4 |
| W | 调试 | 1.00 | 6 | 0.042 | 371 | 1 | 10 | 4 |

**Table 2.** Features and target variables
**表 2.** 特征和目标变量表

| 序号 | 模型 | MAE | RMSE | $R^2$ |
|---|---|---|---|---|
| 1 | 工序类型编码 | 工序类型标签编码 A~W | 数值型 | 输入变量 |
| 2 | 工序工程量 | 工序工作量 | 连续型 | 输入变量 |
| 3 | 作业人数 | 参与人数 | 整数 | 输入变量 |
| 4 | 施工效率 | 单位时间效率 | 连续型 | 输入变量 |
| 5 | 计划开始时间 | 计划工序启动时间 | 时间标量 | 输入变量 |
| 6 | 关键路径标记 | 是否为关键路径工序 | 二值变量 | 输入变量 |
| 7 | 前置工序数 | 前置工序数量 | 整数 | 输入变量 |
| 8 | 实际工期 | 工序预测工期 | 数值型 | 目标变量 |

## 2.2. 模型预测流程

本节系统阐述基于 XGBoost 的机电安装工程工期预测模型的实现流程，整体步骤如图 1 所示，主要分为以下七个环节。

数据采集：本阶段通过查阅历史项目数据或 BIM 系统记录，收集机电安装工程中各工序的特征信息及实际工期。

特征处理：为使模型能够处理不同类型的变量，对数据进行分类变量编码和时间标准化处理，将时间字段(如计划开始时间)转换为统一时间格式，并进行归一化或差值化处理。

$$x' = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \tag{2}$$

特征选择：结合业务经验与皮尔逊相关系数分析特征与目标工期的相关性。

$$r = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2 (y_i - \bar{y})^2} \tag{3}$$

其中，$r$ 值越接近±1，表示线性关系越强。

模型构建与训练：以 XGBoost 为核心，进行参数调优和训练；

预测输出：生成所有工序预测值；

模型评估：计算 MAE、RMSE 与 $R^2$ 等指标衡量性能；

模型对比与可视化：与随机森林、线性回归进行横向对比，输出图表与特征重要性。


**Figure 1.** Flowchart of the XGBoost-based MEP construction duration prediction model
**图 1.** 基于 XGBoost 的机电安装工程工期预测模型流程图

## 2.3. 模型构建

本文采用 XGBoost 回归模型来进行工期预测。XGBoost 是一种基于梯度提升树(Gradient Boosting Tree, GBT)的集成学习算法，通过迭代优化目标函数(均方误差)实现高精度预测。为优化模型性能，采用网格搜索(GridSearchCV)调整超参数，搜索范围包括树的最大深度(4, 5)、学习率(0.1)、估计器数量(100)、子样本比例(1.0)和特征采样比例(1.0)。最佳模型在训练集上拟合后，用于全体样本预测。设训练数据为 $D = \{(x_i, y_i)\}_{i=1}^{n}$，其中 $x_i \in R^m$ 为输入特征向量，$y_i \in R$ 为对应的实际工期。模型预测值定义为：

$$\hat{y}_i = \sum_{k=1}^{K} f_k(x_i), f_k \in F \tag{4}$$

其中，$F$ 是所有回归树的函数空间，$K$ 是树的数量，$f_k$ 表示第 $k$ 棵回归树。XGBoost 通过最小化目标函数进行训练，目标函数由训练误差和正则化项组成：

$$\mathcal{L}(\phi) = \sum_{i=1}^{n} l(y_i, \hat{y}_i) + \sum_{k=1}^{K} \Omega(f_k) \tag{5}$$

其中，损失函数 $l(y_i, \hat{y}_i)$ 通常采用平方误差：

$$l(y_i, \hat{y}_i) = (y_i - \hat{y}_i)^2 \tag{6}$$

正则化项 $\Omega(f)$ 用于控制模型复杂度，定义为：

$$\Omega(f)=\gamma T+\frac{1}{2}\lambda\omega^2 \tag{7}$$

其中，$T$ 是树的叶子节点数，$w$ 是叶节点的权重向量，$\gamma$ 和 $\lambda$ 是正则化超参数。

训练过程采用加法训练策略，每次迭代新增一棵树以拟合残差：

$$\hat{y}_i^{(t)}=\hat{y}_i^{(t-1)}+f_t(x_i) \tag{8}$$

目标函数在当前迭代的二阶泰勒展开为：

$$\mathcal{L}^{(t)}\approx\sum_{i=1}^{n}\left[g_i f_t(x_i)+\frac{1}{2}h_i f_t(x_i)^2\right]+\Omega(f_t) \tag{9}$$

$$g_i=\partial_{\hat{y}_i^{(t-1)}}l\left(y_i,\hat{y}_i^{(t-1)}\right) \tag{10}$$

$$h_i=\partial^2_{\hat{y}_i^{(t-1)}}l\left(y_i,\hat{y}_i^{(t-1)}\right) \tag{11}$$

通过对叶节点权重优化，XGBoost 可以高效学习树模型参数，达到对非线性复杂关系的良好拟合能力。

### 2.4. 模型评估与对比

本文数据集按 8:2 比例划分为训练集和测试集(random_state = 42)，并使用均方根误差(root mean square error, RMSE)、平均绝对误差(mean absolute error, MAE)和 R 平方(R-square, $R^2$)对模型性能进行评估，为验证 XGBoost 的性能，选用随机森林(n_estimators = 100)和线性回归作为基准模型进行对比。模型结果通过可视化(误差对比柱状图、散点图)直观展示。

$$\mathrm{RMSE}=\sqrt{\frac{1}{N}\sum_{i=1}^{N}(\hat{y}_i-y_i)^2} \tag{12}$$

$$\mathrm{MAE}=\frac{1}{N}\sum_{i=1}^{N}|\hat{y}_i-y_i| \tag{13}$$

$$\mathrm{R}^2=1-\frac{\sum_{i=1}^{N}(\hat{y}_i-y_i)^2}{\sum_{i=1}^{N}(\hat{y}_i-\bar{y})^2} \tag{14}$$

式中：$\bar{y}$ 表示实际施工及进度的平均值。

## 3. 实验结果与分析

### 3.1. 模型性能评估

XGBoost 模型在全体工序样本上的预测结果显示，MAE 为 1.17，RMSE 为 2.56，$R^2$ 为 0.996，表明模型具有较高的预测精度。调优后的 XGBoost 模型在全体工序上的整体评估如下(表 3)：

**Table 3.** Comprehensive evaluation of all MEP installation tasks
**表 3.** 全体工序整体评估表

| 序号 | 指标 | 值 |
|---|---|---|
| 1 | MAE | 1.17 |
| 2 | RMSE | 2.56 |
| 3 | $R^2$ | 0.9959 |

由表 4 模型对比汇总表可知，与随机森林(MAE: 5.22, RMSE: 6.93, $R^2$: 0.937)和线性回归(MAE: 20.45, RMSE: 25.45, $R^2$: 0.149)相比，XGBoost 在所有指标上均表现优异，验证了其在非线性关系建模中的优势。

图 2 展示了工序 A~W 的预测工期与实际工期的对比，预测值与实际值高度吻合，误差主要集中在复杂工序 M 上，不同模型工期预测误差对比如图 3 所示。

**Table 4.** Performance comparison of prediction models
**表 4.** 模型对比汇总表

| 序号 | 模型 | MAE | RMSE | $R^2$ |
|---|---|---|---|---|
| 1 | XGBoost | 3.89 | 5.55 | 0.959 |
| 2 | Random Forest | 5.22 | 6.93 | 0.937 |
| 3 | Linear Regression | 20.45 | 25.45 | 0.149 |

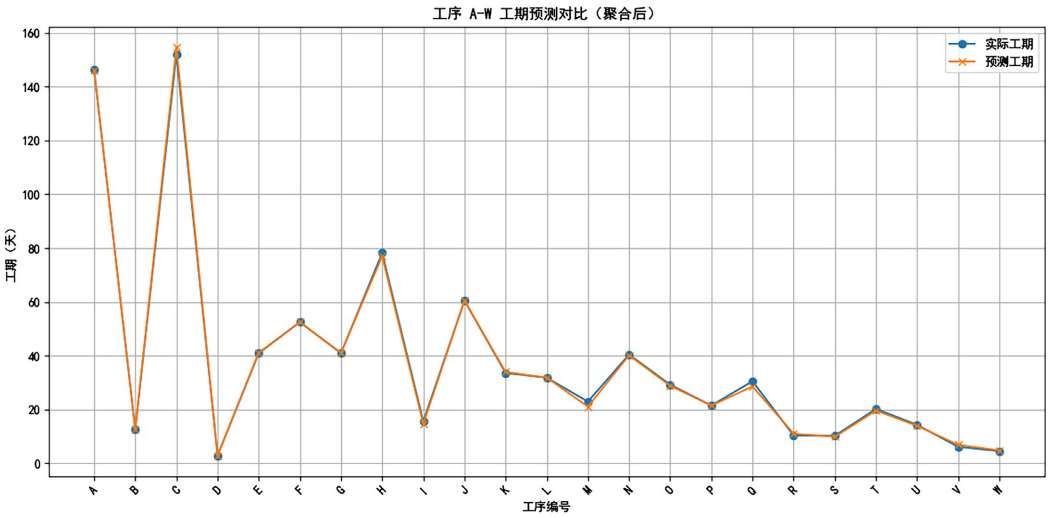


**Figure 2.** Predicted vs. actual duration for project tasks (A~W)
**图 2.** 项目各工序(A~W)预测对比图

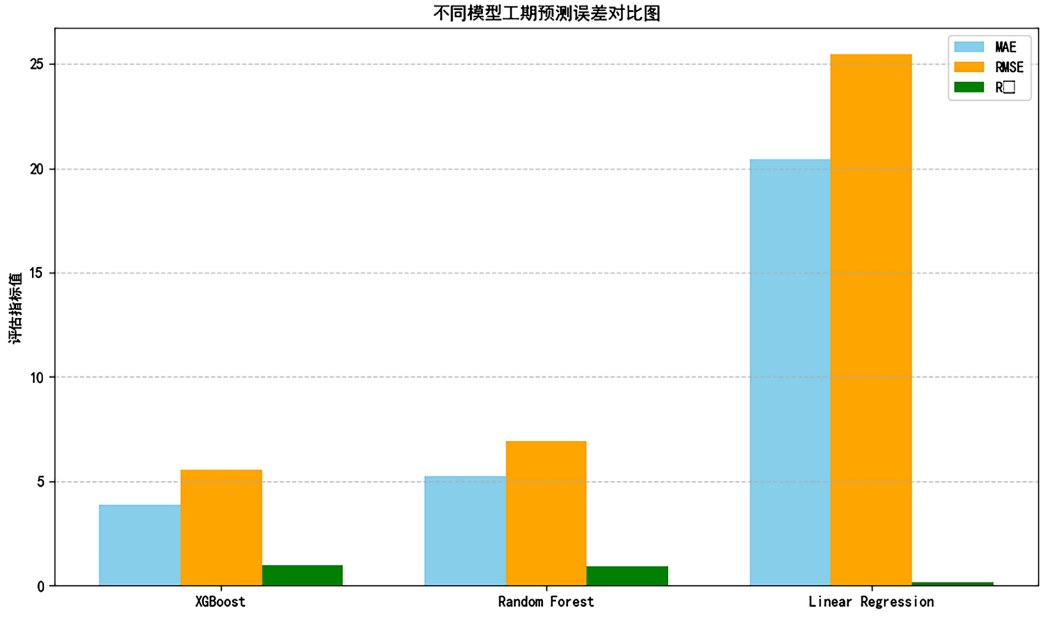


**Figure 3.** Prediction error comparison of different models for construction duration
**图 3.** 不同模型工期预测误差对比图

### 3.2. 特征重要性分析

基于 XGBoost 模型的特征重要性得分如图 4 所示，结果显示“施工效率”“工序工程量”“关键路径标记”以及“作业人数”是影响工期预测的主要因素。其中，“施工效率”以重要性得分 0.32 位列首位，显著高于其他变量，表明其在预测模型中占据核心地位。造成这一结果的原因在于：施工效率作为一种复合型指标，能够综合反映人力投入、资源配置与施工组织的执行能力，具有更高的工期敏感性。相比之下，“工程量”更多地体现为施工任务的静态规模，而“施工效率”则代表单位资源在特定工序中完成任务的动态能力，是衡量项目产能与调度效率的关键参数。例如，在工程量相同的情况下，若人员配置、组织效率或作业熟练程度存在差异，则将显著影响实际完成工期。

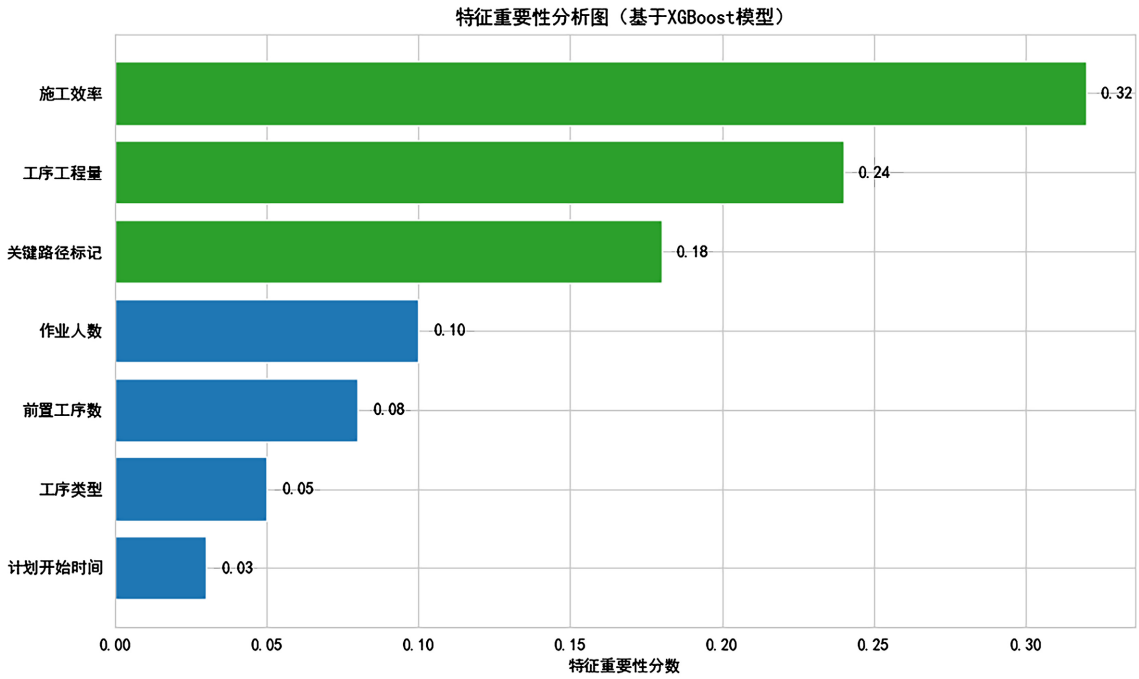


**Figure 4.** Feature importance plot of the XGBoost model
**图 4.** XGBoost 模型特征重要性得分图

虽然工程量本身亦是重要变量，但其对工期的影响受到施工效率与资源投入水平的调节。当效率提升时，单位工程量所需的施工时间将显著缩短。因此，二者在工期预测中具有互补关系，施工效率的提升能够进一步放大管理优化的效果。

此外，“关键路径标记”的重要性得分达到 0.18，说明其在工期预测中的作用不可忽视。为进一步验证该变量的实际贡献，本文对预测误差按“是否属于关键路径工序”进行分组对比分析，结果显示：关键路径工序的平均绝对误差(MAE)为 1.02 天，低于非关键路径工序的 1.39 天，说明模型在对关键任务的拟合上表现出更高的精度，有助于提高整体进度管理的鲁棒性。

从机电安装工程管理实践来看，关键路径上的工序通常构成项目总体进度的控制节点，其在资源配置、进度监控及质量控制方面管理要求更高、流程更标准化。这一现实管理特征也在模型预测中得以体现，解释了关键路径工序具有更低误差的原因。同时，该发现亦为管理者提供了进一步优化的方向——可基于模型结果识别非关键路径中潜在的“隐性延误”风险，提前干预，保障项目总体工期目标的实现。

### 3.3. 预测结果与误差分布

为进一步评估 XGBoost 模型在机电安装工程工期预测中的性能，本文对模型的预测结果与误差分布进行了对比，预测误差散点图如图 5 所示，图中横轴为实际工期，纵轴为模型预测值，大部分样本点分布在理想拟合线附近，表明模型预测精度较高。同时，也存在部分离散点，主要集中在高工期样本上，说明模型在处理极端工程量或关键路径任务时仍存在一定偏差。

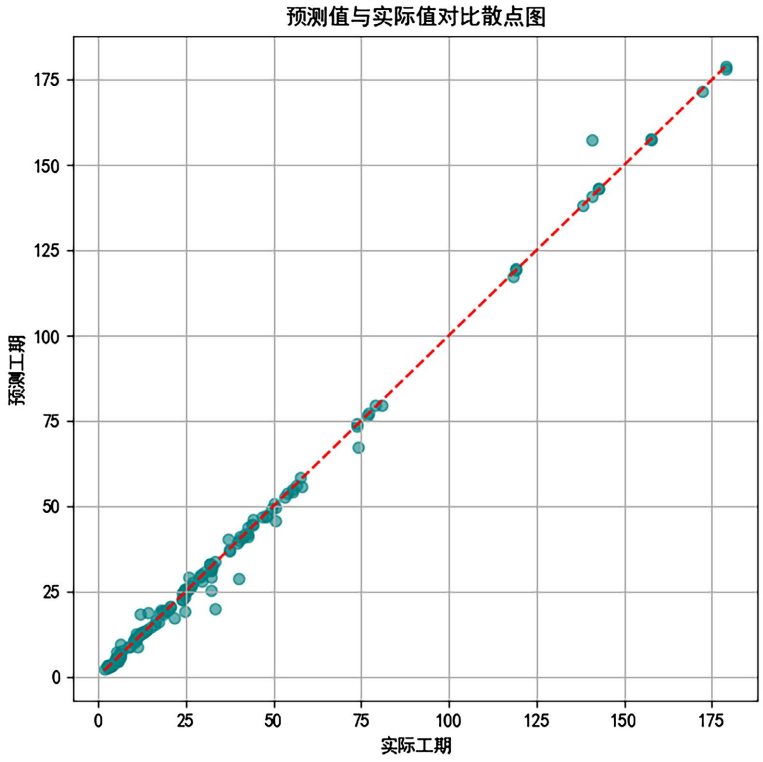


**Figure 5.** Scatter plot of prediction errors for the XGBoost model
**图 5.** XGBoost 模型预测误差散点图

## 4. 总结

本文基于机电安装住宅项目的历史工程数据，构建了一个基于 XGBoost 的工期预测模型，选取 7 个具有先验性和工程实践意义的变量作为输入。实验结果表明，模型在预测精度上优于随机森林和线性回归，具备良好的适用性与稳定性。特征重要性分析显示，施工效率、工程量和关键路径是影响工期的主要因素，其中施工效率贡献最大，且模型在关键路径工序上的预测表现更为精准，有助于提升项目进度管控的科学性与鲁棒性。未来研究可进一步引入动态特征(如天气、班组施工日志、实时进度)和深度学习方法(如 LSTM)，增强模型对工序间时序关系的建模能力与实时预测水平；同时，可将工期预测模型与机电安装工程的工期–成本优化模型联动，构建预测–决策一体化的智能建造平台，实现更高水平的施工智能化与管理协同。

## 参考文献

[1] Forouzeshnejad, A.A., Arabikhan, F. and Aheleroff, S. (2024) Optimizing Project Time and Cost Prediction Using a Hybrid XGBoost and Simulated Annealing Algorithm. *Machines*, **12**, Article No. 867.

https://doi.org/10.3390/machines12120867

[2] Huang, G., Liu, Z., Wang, Y., *et al.* (2024) A Multi-Objective Prediction XGBoost Model for Predicting Ground Settlement, Station Settlement, and Pit Deformation Induced by Ultra-Deep Foundation Construction. *Buildings*, **14**, 2996. https://doi.org/10.3390/buildings14092996

[3] 周海浪, 郭宇, 张浩, 等. 基于加权集成学习的车间生产进度异常预测[J]. 组合机床与自动化加工技术, 2023(8): 188-192.

[4] 韩峰, 余佳, 徐国鑫, 等. 基于改进 XGBoost 不平衡围岩超前分类方法的隧洞工程施工仿真研究[J]. 河海大学学报(自然科学版), 2023, 51(1): 150-157.

[5] 李进友, 李媛, 黄露秋, 等. 基于 XGBoost-GRNN 算法的分段式风功率预测[J/OL]. 计算机集成制造系统: 1-20. https://link.cnki.net/urlid/11.5946.tp.20230927.1028.004, 2025-07-22.

[6] 杜庆峰, 张双俐, 张晨曦, 等. 基于均值滤波去噪和 XGBoost 算法的泥水平衡盾构掘进速度预测方法[J]. 现代隧道技术, 2022, 59(6): 14-23.