

2025-06-18

工作内容：（模型部署）

- 部署SARIMIA+随机森林（22特征）（未完成）
遇到的问题：
- ①SARIMIA模型存在局限性：
该模型的预测必须是连续的（基于历史数据，比如用6月6日前数据训练，那么模型的预测起点必须从6月6日开始，要预测之后的日期，必须先通过自回归的方式。比如要预测10月1日，那么必须先预测出6月6日~9月30日的值）
- 解决方案：
 - （1）每次调用都重新训练模型，保证模型的预测起点是最新的
 - （2）通过模型提供的更新功能，每次将新日期加入到模型中（目前采用该方案）
 - 延申问题：（数据如何存储？）
 - 解决方案：
 - （1）记录历史数据，保存到本地。每次调用就更新历史数据
 - （2）不保存，每次调用api来获取相差的天数的数据，更新模型（目前采用该方法）
- ②部分特征无法用于随机森林：
比如残差，是真实值-SARIMIA的预测值，但是真实值实际是未知的，所以这个特征无法用于真实的预测
将残差等特征去掉后，模型性能居然出现了提升。
- ③目前部署的SARIMIA预测实际效果不理想
原因：可能是基于的历史数据是5月1日-6月5日，而预测的是6月19日。中间差距过大。
解决方案：打算明天用最新的数据（-6月18日），重新训练下SARIMIA模型

结果展示：

删除残差等特征后的表现：

MAE: 1.8385

MSE: 10.2301

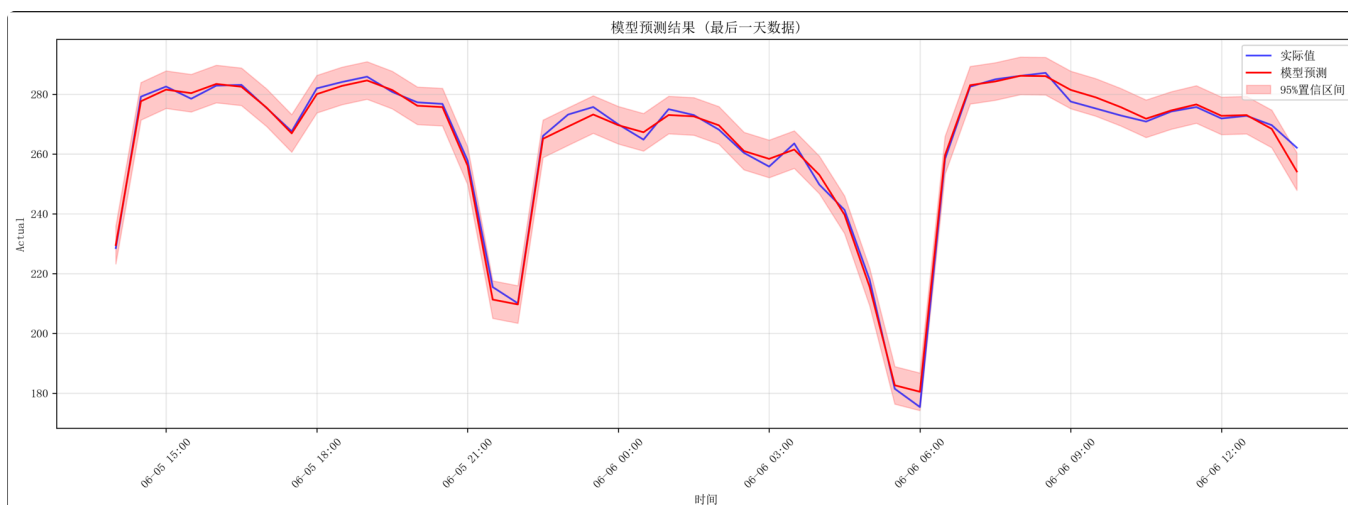
RMSE: 3.1985

R²: 0.9967

残差标准差: 3.1984

直接随机森林误差

mse = 100+



目前只部署了SARIMIA模型，实时预测效果如下：

不太理想，明天更新数据，重新训练下

Code

Details

200

Response body

```
{
  "status": "success",
  "prediction_date": "2025-06-19",
  "predictions": [
    {
      "timestamp": "2025-06-19T00:00:00",
      "value": 274.1957662413063
    },
    {
      "timestamp": "2025-06-19T00:30:00",
      "value": 269.8108766189585
    },
    {
      "timestamp": "2025-06-19T01:00:00",
      "value": 269.73741983013815
    },
    {
      "timestamp": "2025-06-19T01:30:00",
      "value": 273.65960658593184
    },
    {
      "timestamp": "2025-06-19T02:00:00",
      "value": 273.49132369705643
    }
  ]
}
```