一、代码流程

1.数据处理

（提取基本特征、加入更多特征、缺失值处理、整体有序、剔除周末、时间滞后处理）

2.XGBoost模型预测

（当天前为训练集、当天为验证集、下一天为测试集；参数设置）

代码中含较详细注释

二、相关原理

1.时间滞后处理（还不是很清楚在代码中的运用）

时间滞后处理（Time Lagging）是一种在时间序列分析中常用的技术，用于将变量在时间上进行延迟或滞后。它通常用于时间序列预测和回归分析中，以捕捉变量之间的时序关系和滞后效应。

时间滞后处理的基本思想是使用当前和过去的变量值来预测未来的变量值。通过将输入变量在时间上向后移动若干步（滞后），可以创建新的特征，其中每个特征代表了过去一段时间内的变量信息。这些滞后特征可以提供更多有关变量之间的时序关系和趋势的信息。

在时间序列预测中，常见的时间滞后处理方法包括：

* 自回归滞后（Autoregressive Lagging，AR Lagging）：使用过去时刻的变量值作为特征，将其作为输入来预测当前时刻的变量值。
* 移动平均滞后（Moving Average Lagging）：使用过去时刻的变量值的滑动平均作为特征。
* 季节性滞后（Seasonal Lagging）：对于具有季节性变化的数据，使用同一季节上过去时刻的变量值作为特征。

通过时间滞后处理，我们可以考虑并利用变量之间的时序关系，从而改善预测模型的准确性和性能。它对于处理时间相关性强的数据具有重要意义，如气象数据、股票价格、销售数据等。

2.XGBoost模型

原理：[https://blog.csdn.net/weixin\_39910711/article/details/121210569](https://blog.csdn.net/weixin_39910711/article/details/121210569" \t "_blank)

<https://blog.csdn.net/kewei168/article/details/90375743>

参数：<https://blog.csdn.net/weixin_38003620/article/details/115556672>

https://xgboost.readthedocs.io/en/latest/parameter.html#general-parameters（官方指南）

https://xgboost.readthedocs.io/en/latest/python/python\_api.html（Python API参考）

三、使用方法

设置数据集路径及预测结果文件名

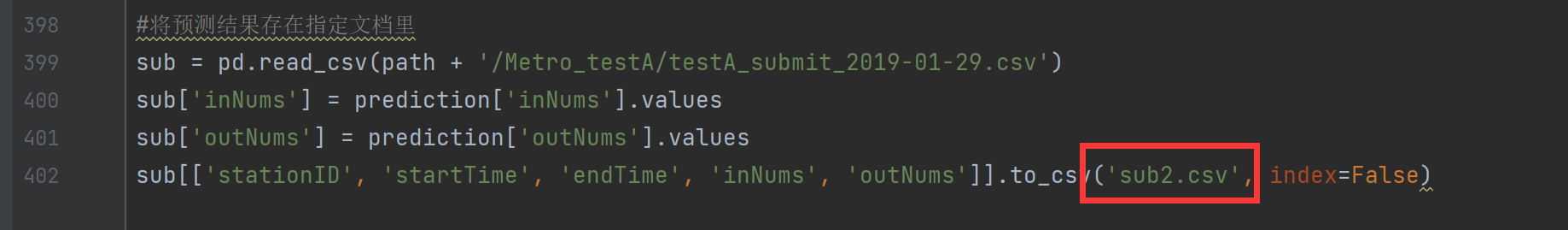
数据集路径设置：

数据集按照原先给出方式存储只需修改path即可，较为方便



预测结果文件设置：

预测文件如此设置默认和py文件同目录



以上述设置为例，文件夹目录如下：

