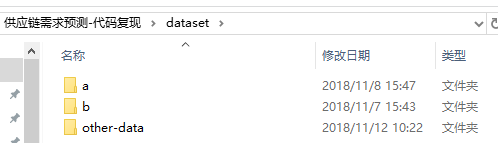
**代码复现文件说明**

# 1、数据来源说明

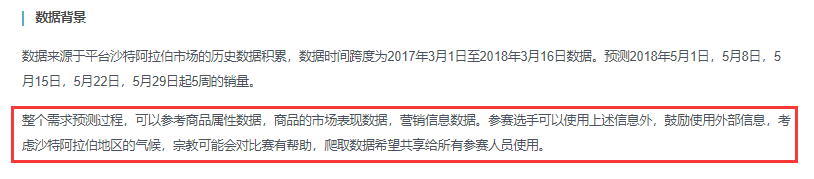
1、在dataset文件夹中，有三个文件夹：



2、**对于文件夹a**：存放了初赛数据集，因为我们有1个模型使用了初赛数据与复赛数据一起训练的。

3、**对于文件夹b**：存放了复赛数据集。

4、**对于文件夹other-data**，存放了其他数据集，有1个模型使用了外部数据。外部数据集，主要包括沙特阿拉伯的天气数据和节假日数据，因为题目要求是可以使用外部数据的：



我们通过下面的网址进行了数据的爬取与整理：

**天气数据来源**：

<https://www.wunderground.com/weather/eg/saudi-arabia>

**节假日数据来源**：

<http://shijian.cc/114/jieri_2017/>

# 2、运行环境说明

1、python版本：python3.6.0

2、xgb版本：0.8

3、lgb版本：

（1）zh开头py文件的lgb版本为：0.1

（2）yw开头py文件的lgb版本为：0.2

4、机器配置：处理器i7/内存16g及其以上

5、说明：因为队友自己用自己的电脑写的代码，因此模型版本有区别。相信大家也不会为了复现代码而让lgb和xgb模型版本号不同，生成的模型会有一定的区别，因此结果可能会有一定偏差。

6、为了方便运行就使用**python3.6.0-xgb0.8-lgb0.1-处理器i7-内存16g及以上**，我们已经用这个标红的运行环境执行过**run.py**文件，复现结果基本一致。

# 3、代码执行说明

## 3.1代码执行第1种方式

1、在我们文件解压的当前目录下，有这样一份python文件**run.py**：



2、run.py是我们写的python的执行脚本，**运行run.py，即可生成结果文件**，这份文件产生在当前目录下，**也就是我们最终提交的结果文件ans\_merge\_1111\_B\_lucky\_复现.csv：**



## 3.2代码执行第2种方式

1、如果不执行run.py，想分别查看结果，也可以按这个顺序运行我们的代码：

**yw-lgb.py**

**yw-rule.py**

**cy-lr.py**

**zh\_lgb\_v1.py**

**zh\_lgb\_v2.py**

**zh\_lgb\_v3.py**

**zh\_lgb\_v4.py**

**zh\_lr\_v1.py**

**zh\_lr\_v2.py**

**zh\_xgb\_v1.py**

**zh\_xgb\_v2.py**

**merge-1111-提交.py**

2、这些python文件存放在当前目录中，最终生成的**ans\_merge\_1111\_B\_lucky\_复现.csv，也就是我们的最终结果文件。**

3、其中以zh开头的py文件，来源于同一套特征，通过训练集扰动、特征扰动、参数扰动构造了有差异性的模型，如果是想学习源代码时，看前四份文件：yw-lgb.py、yw-rule.py、cy-lr.py、zh\_lgb\_v1.py即可。

## 3.3代码执行时间说明

由于复赛增加数据集明显，若使用xgb模型耗时比较明显，其余模型均比较高效。代码复现总的执行时间大致为12小时左右，挂着run.py睡一觉起来就ok啦。