**1、主要用到的软件和库**

-- python 3.6

-- pandas

-- xgboost(gpu加速版本)

-- scikit-learn

-- numpy

-- matplotlib

**2、xgboost配置GPU加速（快速安装版）**

环境：anaconda（python3.6） + win10第一步：

下载xgboost源码：https://github.com/dmlc/xgboost第二步：

从http://ssl.picnet.com.au/xgboost/中下载支持GPU版已编译好的DLL文件第三步：

将xgboost.dll复制到xgboost-master/python-package/xgboost目录下，

打开控制台，进入到python-package目录下，执行python setup.py install命令，

至此xgboost GPU版安装成功。

GPU版使用方法：

param[‘gpu\_id’] = 0

param[‘tree\_method’] = ‘gpu\_hist’

**3、附加材料**

在submission文件夹中是六个模型预测结果的csv文件以及每个模型的特征重要性可视化图，主要区别是是否包含额外特征以及max\_depth的不同，在文件名中可以清晰区别

**4、运行时间**

先是在本地下载比赛提供的数据集，train.csv，test.csv和store.csv（下载地址：https://www.kaggle.com/c/rossmann-store-sales/data），已经打包至目录下，运行data\_processing.py，再运行run\_model.py。其中run\_model.py中的参数设置是是最好的模型运行结果，跑一次需要大概2个小时的时间。

