设计说明

1. 采用LightGBM模型进行预测，选择五折交叉验证法作为判定标准进行验证。

2. 算法设计：

（1）预处理：

对缺失值进行填充，为需要转换的列设置新的值，即进行数值化处理。

（2）特征工程：

分为四个部分。

第一个部分为时间特征，把DOTTING\_TIME转换为一天的时间。

第二个部分为行统计特征，使用CPU\_USAGE, MEM\_USAGE, DISK\_USAGE, LAUNCHING\_JOB\_NUMS和RUNNING\_JOB\_NUMS这些特征进行聚合分组，并使用agg函数将min，max等值放进一张表，最后对数据的一些缺失值进行填充。

第三个部分为滑动窗口构造数据集，生成测试集与训练集的时间窗数据，其中训练集按QUEUE\_ID进行处理生成时间窗的值与需要预测的值，最后对这些数据进行reset\_index重新设置索引。

第四个部分为列统计特征。

最后一个部分为特征过滤，把一些无效列去除。

（3）训练与预测：

对lightGBM算法进行使用。先对之前处理好的特征进行选取，再对参数进行初始化。

设置完参数后，对训练集和验证集进行切割操作，再迭代地使用lightGBM模型对特征进行训练与验证，使用交叉验证方法去得到更加稳定的模型同时实时更新进度条。

最后对结果进行评估，并将结果输出到csv文件中上传到S3。