**基于机器学习的异常检测系统开发**

# 1 项目总体完成情况

项目目前已经实现了异常检测的基本功能，能够根据导入的csv格式数据进行训练与预测，完整的算法模块包括xgboost以及ESD方法，覆盖有监督与无监督。

能够得到检测到异常检测数据文件csv格式；

能够展示检测数据的异常图；

能够动态展示某一个时间段内的异常点位置；

能够检测单变量异常值；

能够在web端进行算法类别选择；

能够选择所需要检测或者训练的文件；

# 2 技术方案总结

技术方案分为系统开发方案以及算法开发方案；

系统开发使用的是java语言；

算法库调用方面使用Java调用Python使用控制台进行调用，

前端可视化方面需要保证使用Ajax与后台持续连接，同时要高效利用数据库连接池，减少不必要的资源浪费。

算法开发方案

使用python开发语言，主要使用的第三方库包括sklearn，numpy，pandas，还有统计类第三方库；

算法开发过程包括数据预处理，数据特征提取，模型训练以及模型调用，

检验流程包括选取算法库中的算法，选择待处理文件，进行测试。

# 3 项目质量总结

功能

系统功能仅仅实现简单的功能，还需要进一步的完善；

系统性能

系统性能，由于是通过读取文件的格式，并且用java调用python控制台的形式进行运行算法代码，所以性能较差，如果能够从数据库直接读取，将会极大增强系统性能；

易用性

易用性方面，由于我们提供的算法选项比较多，而且使用同意的输入和输出格式，所以对于用户还是非常友好的；

可靠性可维护性

可靠性方面，分模块开发，系统性能相对可靠，也比较方便进行维护；

# 4 亮点与缺陷

项目亮点

项目采取有监督和无监督算法，另外能够动态查看异常点；

项目缺陷

有监督和无监督算法还没完全实现完，web系统也有功能有待进一步的完善。

# 5 收获与建议

通过全小组一起写协作，背景调查，方案制定，工程实施，数据测试，让小组成员了解了一个完整项目的较为完善的流程，并且增强了大家协作开发的能力。

建议：

虽然系统基本上完成了，但是在每一个部分，功能实现都较为简单，而且在数据存储以及处理方面采取的方法和方案都有优化的空间，与实际生产环境下工程还有一定距离，建议以后在类似项目中一定要了解当前环境下其他人的解决方案。