

# 计划书修改版

	A	B	C
1	阶段	进度安排	具体安排
2	第一阶段： 2022年3月—7月	学习项目需求的预备知识	（1）基础的分析方法：可视化分析、相关分析、信息熵增益分析等。学习机器学习相关的统计学习模型与算法知识：逻辑回归算法、关联挖掘、聚类算法、决策树算法、随机森林算法、支持向量机模型、蒙特搜索算法、隐式马尔可夫模型、多示例学习算法、迁移学习算法、卷积网络算法等。（3）了解智能运维：在有编程语言、数学、机器学习的基础上，学习智能运维的相关知识，了解掌握智能运维体系的基本架构后面的实现打下基础。（4）学习Linux Shell脚本语言：掌握常用Linux Shell命令，熟练使用Linux操作系统。
3	第二阶段： 2022年7-9月	收集整理互联网公司的真实数据	收集部分互联网公司服务器的KPI数据（CPU利用率、每秒查询的数量、响应延迟、PV、GC等运维数据）。所收集的数据应包括已经标记了异常的数据和未标记的原始数据，然后将得到的数据进行进一步整理、筛选，为下一步的预处理操作做准备。
4	2022年10月—2022年12月	预处理、聚类、提取特征值	（1）对事先收集的不同类型的数据进行标准化，将它们的数据缩小到大致相同的范围，使其具有可比性。（2）通过聚类算法，将大量的数据分成几个集群，并分别处理每个集群，得到对应集群中心的KPI曲线。每种特征参数化，从已经标记过的数据的集群中心提取KPI曲线的特征值。对于新的数据，把该KPI曲线的特征值、所属集群中心的特征值和异常值作为训练集。
5	第四阶段： 2023年12月-1月	实现服务指标异常检测模型	尝试采用算法OmniAnomaly作为半监督学习的算法来检测KPI曲线的异常。基于已形成的服务异常指标检测模型，可以实现自动对一组新的KPI曲线进行异常检测。将各个特征量化，通过其对应的阈值来判断是否存在异常。
6	第五阶段： 2023年1月—2月	服务指标异常检测模型的训练及改进	将不断出现的新的KPI曲线分配到一个已经存在的集群中，合并新的KPI曲线中未标注的数据和聚类中心，最后通过半监督学习训练成一个新的KPI模型，实现服务指标异常检测模型的不断更新，确保其判断的准确率进一步的提升。
7	第六阶段： 2023年3月—4月	数据的再收集及正确性的验证，模型评估	重新收集在阶段二的采样企业近十个月的数据，并与服务指标异常检测得到的结果进行对比，测试模型的正确性。（1）真实数据采集：从互联网公司收集真实的原始运维数据。（2）对于每种特征，设定阈值，定义评估异常检测的度量方法。（3）将本项目所采取的方法实际的性能与监督学习方法和非监督学习方法的性能建立一定的度量比较方法进行对比，分别从准确率、数据需求量、人力劳动资源的需求量等角度进行评价。