厦门国际银行客户逾期风险通用分方案框架

背景：

我行目前需要构建自主风控体系，需要的有一套行之有效的评估方式来识别我行各个联合贷款的不良客户。

目标：

构建一套行之有效的通用分进行使用，从而满足我行在各个联合贷款渠道上使用。

核心需求：

* 模型在使用上具备泛用性，能够适用于多个渠道的联合贷场景。
* 但也需要能够有效识别出逾期客户，有效减少联合贷款的不良客户。
* 同时需要有相应的监测方案，后续在模型使用的过程中能够监测我行各个渠道的资产质量。

实施内容：

1. 部署机器学习模型上线（灰度上线）
2. 需要协调慧决策（外部系统）的协同
3. 结果/效果监控机制



整体上模型流程如上图所示，其中该模型部署在MLP平台上，该系统目前主要信息承接的通道为慧决策。

这里有部分信息问题亟需解决/确定：

1. 目前各个渠道的请求报文不一致，从而导致在加工上方式不一致以及基础信息不一致
2. 通用分在决策中调用的批次问题（灰度上线或由模型效果决定的，目前灰度上线，也需要使用外部的数据，需要慧决策调用外部数据一并推送）
3. 在灰度上线期的监测方案，效果统计（模型的效果-指标效果，业务效果）

模型开发过程中需要进行的工作：

准备阶段：

* 标签、窗口确认：虽然在数据验证过程中使用的标签和窗口已经确认，但还是需要和业务确认相应口径。主要是在后续通用分针对多个渠道的时候观察口径。
* 数据准备：目前仅有客户标签情况 和 外部数据评分。因此在数据准备上变化不大，只需要抽取目前有经过外部数据的客户信息（基本上卡着月份进行获取）。需要了解业务流程，进行抽取数据。目前主要使用的数据为。整合多个场景数据进行整合。以及数据的抽样问题。
* 数据质量分析：

开发阶段：

* 特征工程
* 模型开发：在Sophon上进行模型开发，需要注意训练态和运行态的状态转化。
* 模型效果验证

上线部署阶段：

* 模型上线以及和慧决策的协同工作

在开发过程中主要的问题：

1. 数据来源及整合的问题：如何有效的整合3个渠道的客户信息进行建模，抽样方式如何（默认先按权重进行抽取），以及是否3个渠道均抽取等？
2. 实时运行数据和训练数据或存在不一致的情况。训练数据是二次加工的数据，而原始数据信息（进件数据）匹配相对较少。需要考虑一下数据加工方面的事项。