

评分卡系统构建概述





课程目录

ONTENTS

1 概述

2 确定目标产品调研

3 建模流程

了解产品基本信息，现有流程，放款量，产品期限等产品要素，历史政策变化情况，确定入模的样本

根据原始数据生成衍生变量，分箱并计算每个变量的区分能力

在系统里配置模型上线，结合业务实际流程和风险偏好制定模型应用策略

确定目标
产品调研

变量衍生
及处理

模型上线
及应用

标签定义

模型构建
及检验

对客户贷款表现进行vintage分析，roll rate分析，根据结果以及产品期限在结合行方产品实际情况确定，观察期，好坏客户的定义

模型训练，并测试模型的稳定性，区分度。



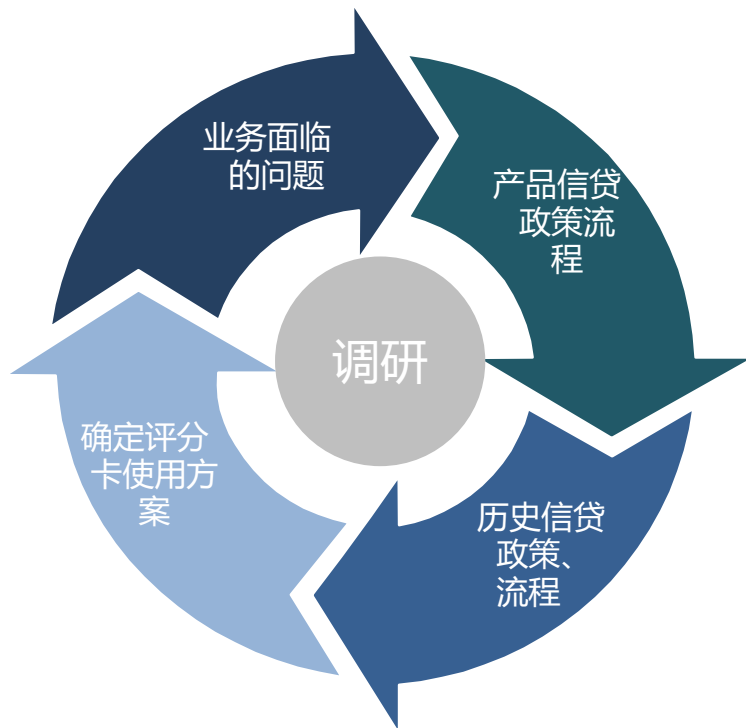
课程目录

ONTENTS

1 概述

2 确定目标产品调研

3 建模流程



业务面临的问题，业务上目前要解决什么问题：

- 1.流程：简化流程？提高时效？节省人力？
- 2.风控：风险指标未达标？
- 3.利润：利润未达标？

产品、信贷政策流程

产品：面向人群，获客渠道，产品金额，期限，定价
信贷政策：了解客户的准入，授信政策
流程：了解贷款进件，审批流程，收集了什么资料等

历史信贷政策流程

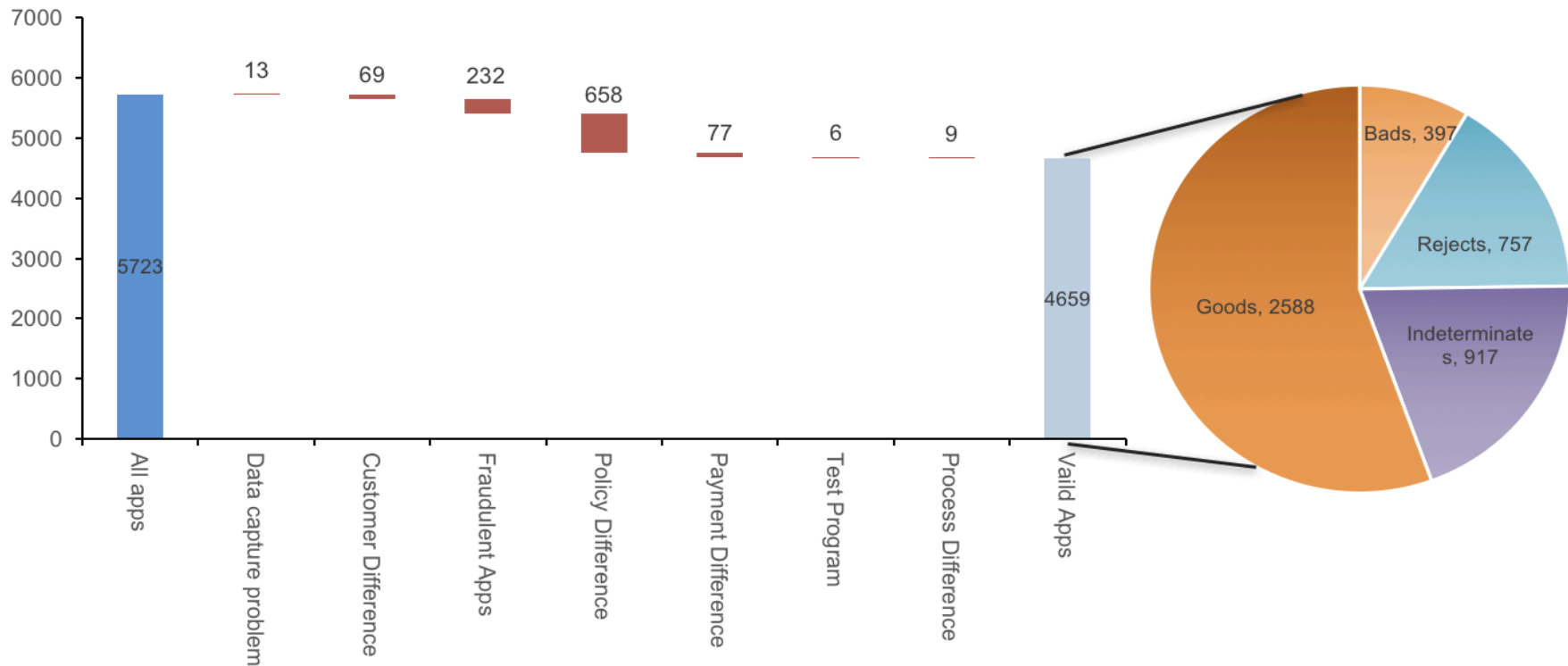
历史上信贷政策、审批流程、收集材料的重大变化
获客方式，获客渠道的变化

评分卡的使用方案

确定评分卡的在流程中的哪个阶段用,怎么用，预测使用效果。

入模样本筛选瀑布图：

从以下角度筛选样本：数据问题，客户群体差异，政策差异，产品要素差异，流程差异等筛选出和预测群体相似的变量。





课程目录

ONTENTS

1 概述

2 确定目标产品调研

3 建模流程

标签定义相关概念



目的：定义想要的客户为好客户，不想再要的客户为坏客户。



量化分析：利用成熟度分析（Vintage），滚动率（RolloverRate)确定好坏定义



专家分析：好坏定义是否能被业务接受，是否与目标相关。

观察窗口 (Observe Window) : 收集建模自变量的时间段。

表现窗口 (Performance Window) : 判断贷款好坏的表现时间段 , 模型应变量的要素。



滚动率分析 (Rollover Rate Analysis)

		status at 18 MOBS							Grand Total	Roll into	%WOFF	
		Paid-Off	Current	1-29DPD	30-59DPD	60-89DPD	90-119DPD	Write_off				
		3195	0	69	57	45	36	684	4086			
Worse Delinquency @6 MOB	Paid-Off	399							399			
	Current	2216		42	42	33	21	301	2655	11.3%	44.0%	
	1-29DPD	566		27	15	12	15	231	866	26.7%	33.8%	
	30-59DPD	13						98	111	88.3%	14.3%	Low Coverage 22.2% (152, 684)
	60-89DPD							27	27	100.0%	3.9%	
	90-119DPD	1						15	16	93.8%	2.2%	
	Write_off							12	12	100.0%	1.8%	
		3195	0	69	57	45	36	684	4086			
Worse Delinquency @9 MOB	Paid-Off	471							471			
	Current	2019		36	30	18		200	2303	8.7%	29.2%	
	1-29DPD	667		33	27	27	36	294	1084	27.1%	43.0%	
	30-59DPD	36						114	150	76.0%	16.7%	Low Coverage 27.8% (190, 684)
	60-89DPD	2						24	26	92.3%	3.5%	
	90-119DPD							19	19	100.0%	2.8%	
	Write_off							33	33	100.0%	4.8%	
		3195	0	69	57	45	36	684	4086			
Worse Delinquency @12 MOB	Paid-Off	597							597			
	Current	1698		21	21			58	1798	3.2%	8.5%	
	1-29DPD	857		48	36	45	30	78	1094	7.1%	11.4%	
	30-59DPD	35					6	197	238	82.8%	28.8%	Good Conversion 91.7% (548, 597) Good Coverage 80.1% (548, 684)
	60-89DPD	5						138	143	96.5%	20.2%	
	90-119DPD	3						103	106	97.2%	15.1%	
	Write_off							110	110	100.0%	16.1%	
		3195	0	69	57	45	36	684	4086			



成熟度分析

- 看不同程度逾期客户表现周期在什么时候成熟



滚动率分析

- 看在不同表现期内不同程度逾期的客户滚动到不良或者核销的概率



业务含义

- 从业务上检查量化分析出的好坏定义是否符合业务要求

根据上述分析定义：

好客户：在放款后12个月内从未发生过逾期的客户。

坏客户：在放款后12个月内曾经发生过30天以上逾期的客户。

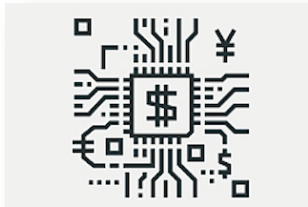


数据收集

收集历史上已收集的数据，或请三方数据公司回溯。

行方内部数据：申请表，PBOC数据等。

外部数据：运营商、多头等。



数据清理

质量和准确性（业务上有没有验证是否客观）

二义性（有二义性的建立数据标准）

覆盖率（过低的去除）

极值处理（Capping and Flooring）

缺失值处理（mean，mode等）



变量衍生

定量：暴力探索+深度学习；

定性：专家从各风控维度列出变量。

变量类型处理

01

连续变量

分箱，计算WOE、IV，特征编码

离散变量--无序

聚类按好坏比分类，特征编码

02

03

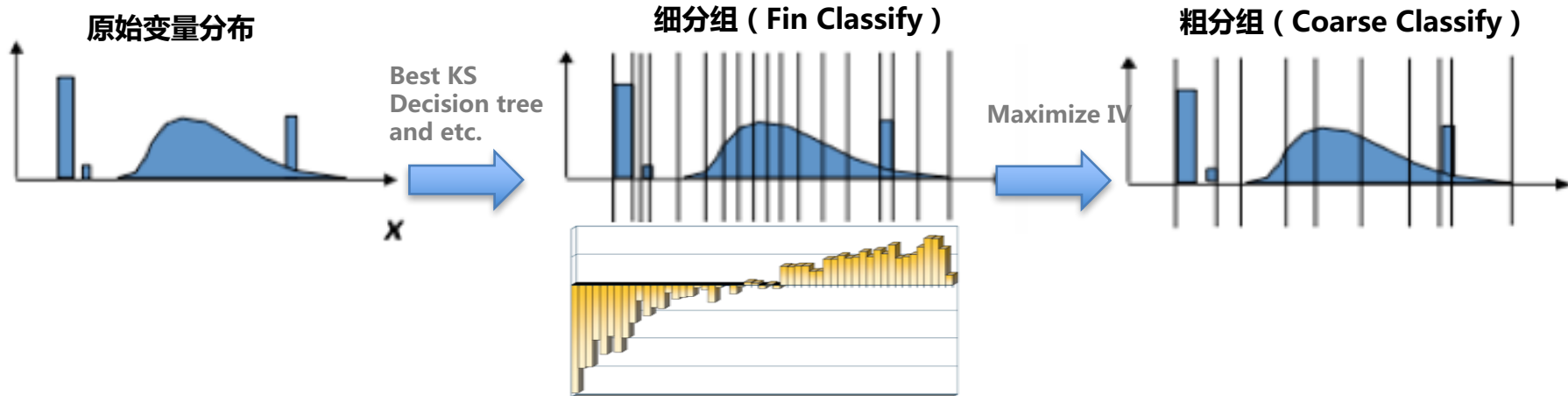
离散变量--有序

排序后同连续变量

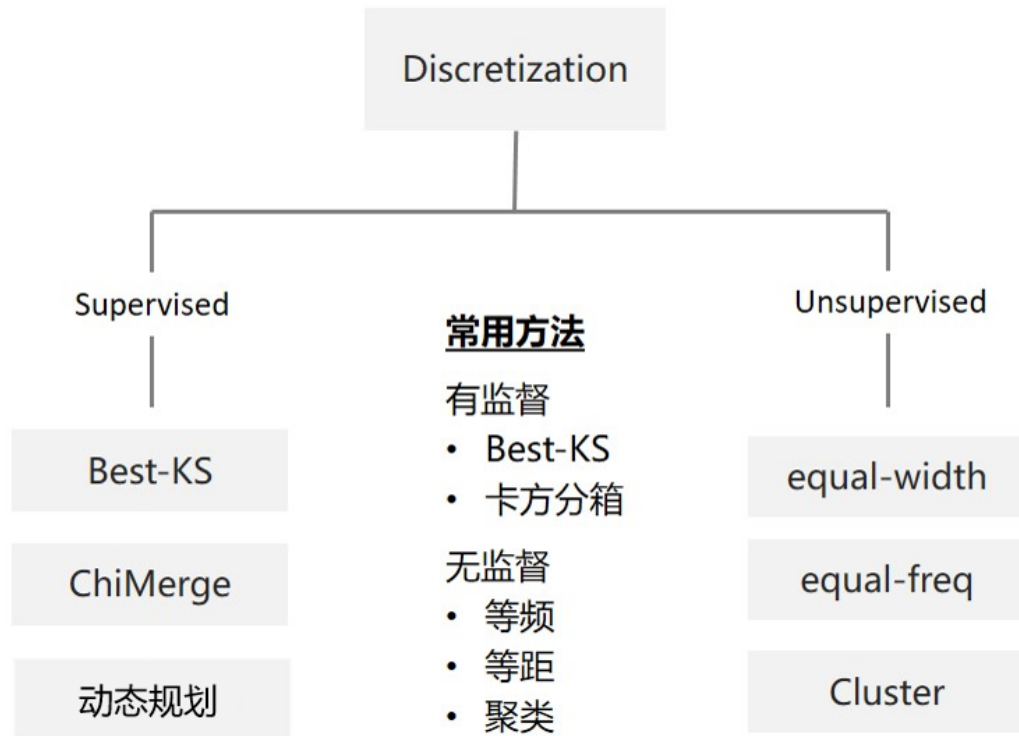
分箱

定义：将连续变量离散化，找到几个节点分成几个组。

目的：通过分箱可以增强模型的稳定性、健壮性，增强模型的解释性。



分箱常用方法：



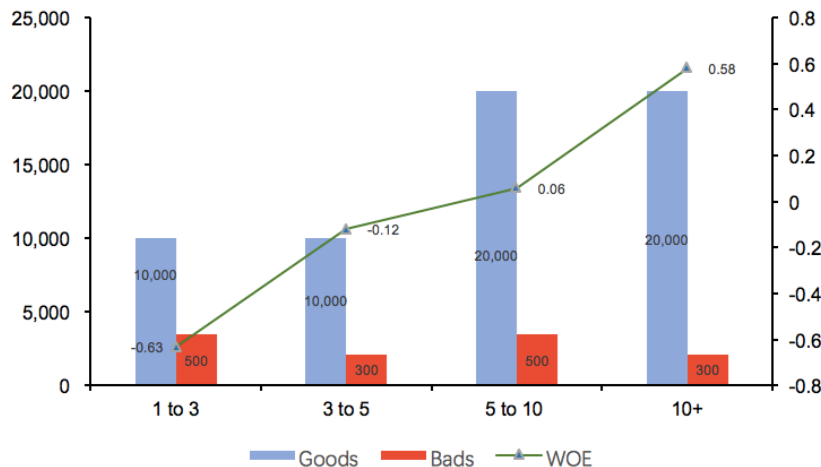
WOE和IV计算

$$woe = \ln\left(\frac{\text{good ratio}}{\text{bad ratio}}\right) = \ln\left(\frac{\frac{\#good\ in\ S}{\#good\ in\ total}}{\frac{\#bad\ in\ S}{\#bad\ in\ total}}\right)$$

$$IV = \sum_{i=1}^n (\text{good ratio} - \text{bad ratio}) \times \ln\left(\frac{\frac{\#good\ in\ S}{\#good\ in\ total}}{\frac{\#bad\ in\ S}{\#bad\ in\ total}}\right)$$

Example

EXAMPLE CHARACTERISTIC BINNING FOR VARIABLE "WORK EXPERIENCE"					
Attribute	Goods	Bads	Odds	WOE	IV
1 to 3	10,000	500	20.0	-0.63	0.092
3 to 5	10,000	300	33.3	-0.12	0.003
5 to 10	20,000	500	40.0	0.06	0.001
10+	20,000	300	66.7	0.58	0.085
Total	60,000	1,600			0.180



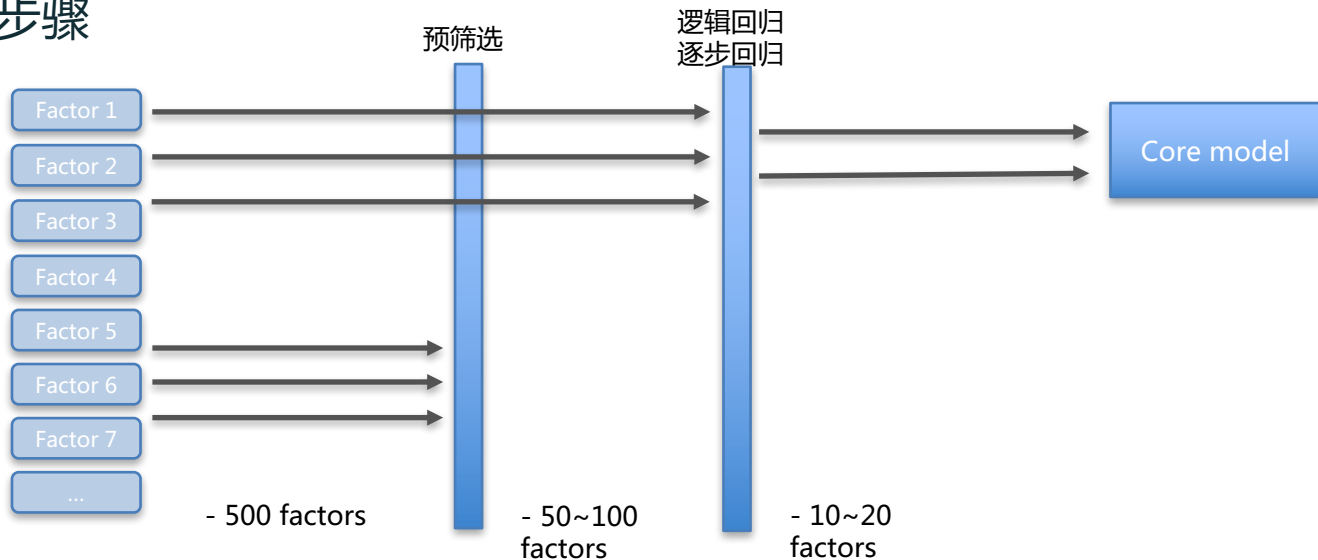
变量预筛选

规则名称	规则详细逻辑	规则参数设置	备注
单一阈值筛选	1、计算每个变量的每种取值占比 2、设定单一阈值阈值 3、遍历每个变量： 如果 $\max(\text{变量的不同取值占比}) > \text{单一阈值阈值}$ ，则剔除变量；否则保留变量	单一阈值=0.9	每一次筛选都是承接上一次筛选结果
IV筛选	1、计算每个变量的最优分组 2、设定IV阈值 3、遍历每个变量： 如果 $\text{变量IV} < \text{单一阈值阈值}$ ，则剔除变量；否则保留变量	IV阈值=0.02	
PSI筛选	1. 如果建模样本的时间点跨度几个月，甚至更长，可以按月切分或者更长步长切分 2. 选择时间最早的一组作为基本样本，剩下的作为验证样本，计算组间的PSI，剔除PSI大于"PSI阈值"的变量	PSI阈值=0.05	该部分视样本时间点实际情况而定
相关系数筛选	1、计算所有变量之间的相关性 2、设定相关系数阈值 3、遍历每个变量： 选择出与该变量相关系数大于相关系数阈值的所有变量，只保留这些变量中IV最大额那个变量，其余变量删除	相关系数阈值=0.8	
专家筛选	1. 专家结合实际业务场景，分析指标实际含义 2. 对于表现不错但是实际含义无法解释，或者分组无法解释的变量，剔除		该部分视实际业务场景和指标含义而定
...

模型搭建:

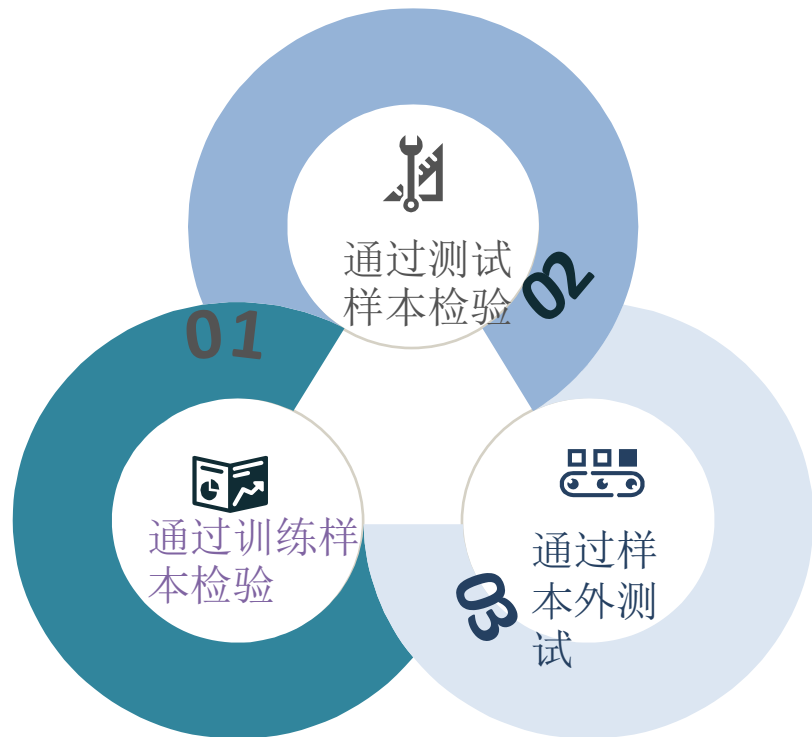
- 样本：训练样本（Training Sample）占比60%-70%，测试样本（Testing Sample）占比40%-30%，样本外测试样本（oot）

➤ 步骤



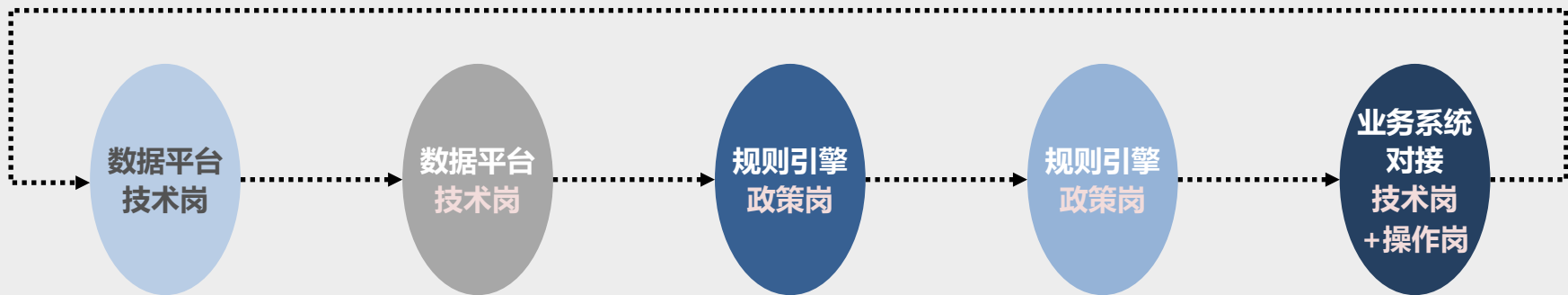


评分卡验收数据集



评分卡验收标准





- 1.三方数据接入
- 2.模型变量开发
- 3.模型配置与测试
- 4.流程配置
- 5.操作岗培训



T H A N K S