



腾讯社交广告  
Tencent Social Ads



`<=Date()`

`freeze()`

`for`

# 高校算法大赛

9

2017-07-06

8

`Your`

`var`

`if`

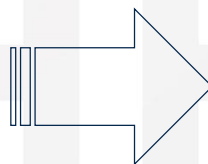
`time`

|        |           |
|--------|-----------|
| PART01 | 成员组成      |
| PART02 | 任务描述&赛题理解 |
| PART03 | 数据分析&处理   |
| PART04 | 样本构建      |
| PART05 | 特征工程      |
| PART06 | 模型构建      |
| PART07 | 经验总结      |



## 任务描述&赛题理解

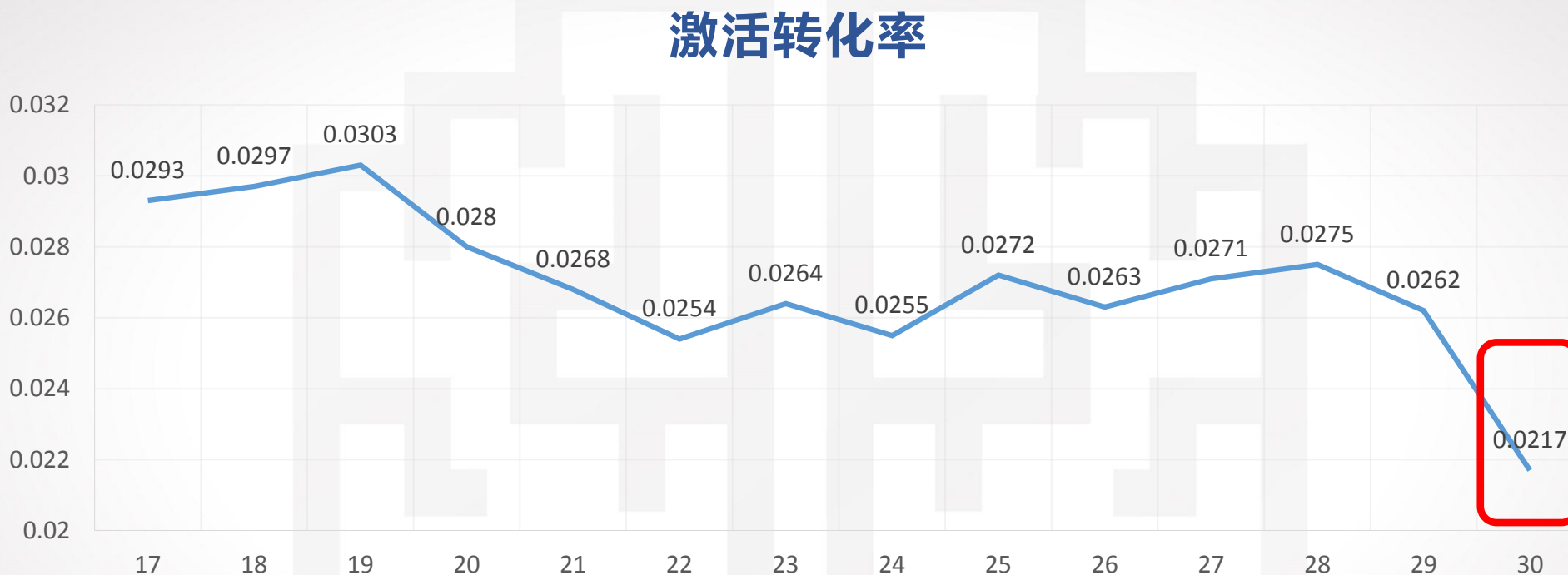
第17日到第30日的用户广告交互数据、用户APP交互信息、用户基本信息、广告信息、App信息等数据



预测第31日App广告点击后被激活的概率

## PART 03

# 数据分析&处理



主要由以下两方面原因导致：

- 1) 用户可能在下载之后过了很久才启动App
- 2) 某些广告主回流时间过长



**例如：广告主A的最后一次转化时间（label=1）是3:00PM，将该广告主在3:00PM之后的样本删除**

## PART 04

### 样本构建



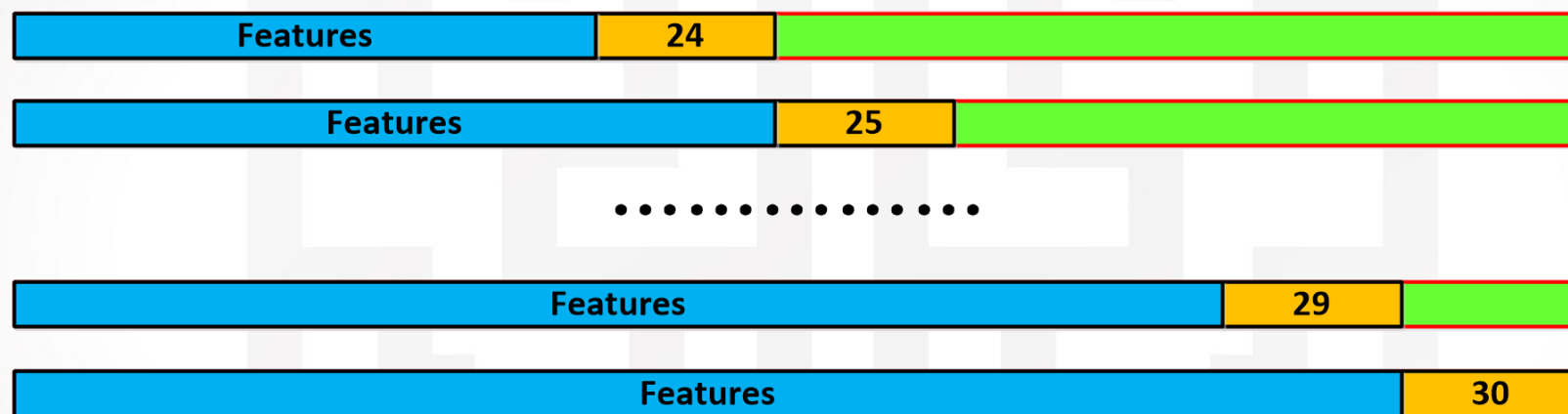
## PART 04

# 样本构建

线上训练集

第16天

第30天



线上







## 特征工程——基础特征

### 用户基础特征

age、gender、  
education、  
marriageStatus、  
haveBaby、  
hometown、  
residence

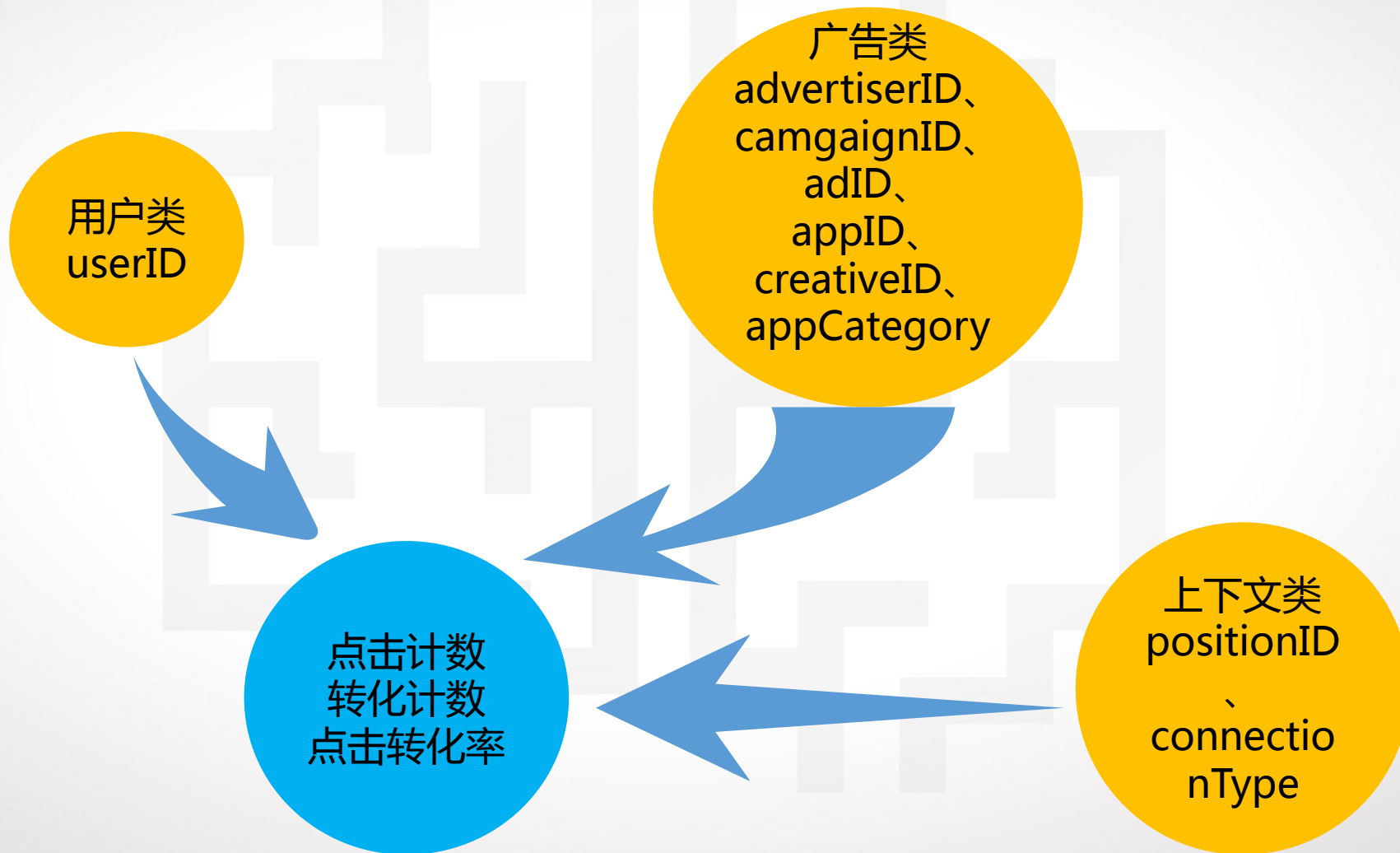
### 广告基础特征

advertiserID、  
campaignID、  
creativeID、  
appID、adID、  
appCategory、  
appPlatform

### 上下文基础特征

positionID、  
sitesetID、  
positionType、  
connectionType、  
telecomsOperator

## 特征工程——统计特征



## 特征工程——统计特征

### 用户App安装列表文件

➤ 描述了在第1天之前用户全部的App安装列表

◆ user安装计数

◆ app安装计数

### 用户App安装流水文件

➤ 训练数据开始时间之前16天开始连续30天的安装流水

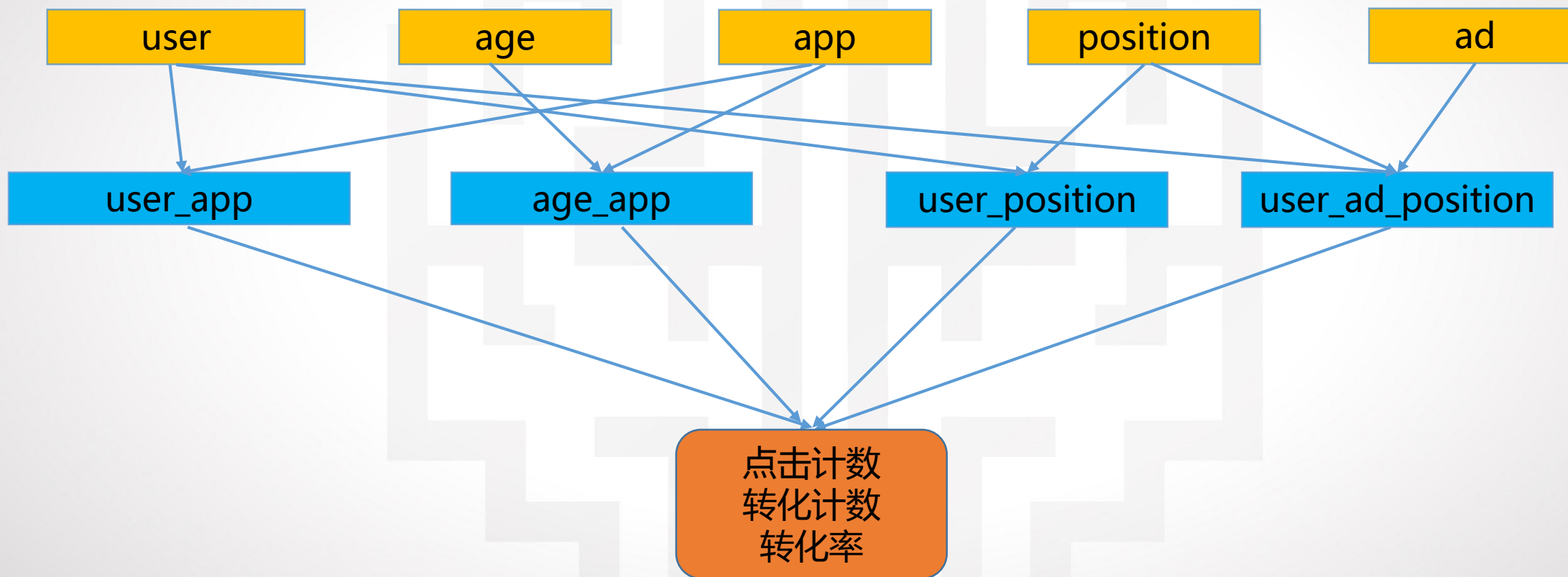
◆ user安装计数

◆ app安装计数

◆ userApp安装计数



## 特征工程——交互特征



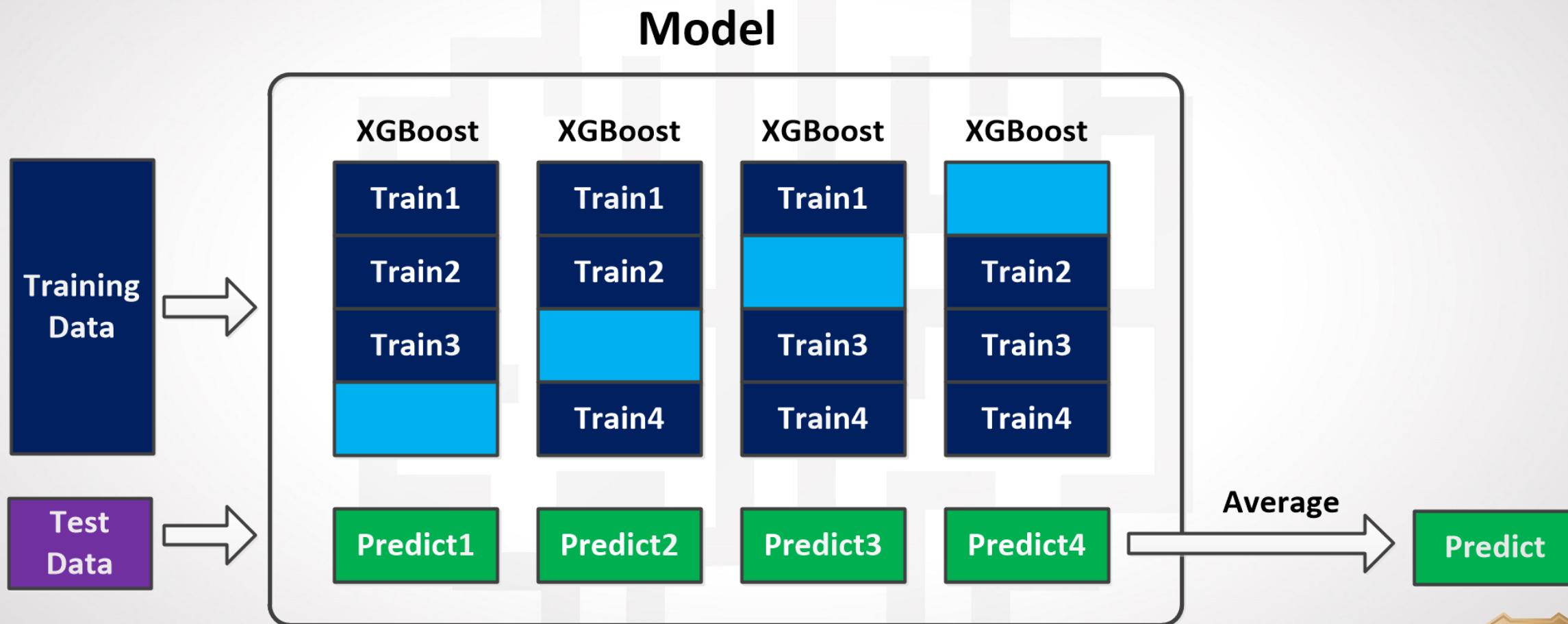
## 特征工程——其他特征

### Trick特征

- 预测日当天用户重复点击只计一次转化
- ◆ 预测日当天重复样本的第一条与最后一条的时间差、与第一条的时间差、与最后一条的时间差
- ◆ 预测日当天重复样本按时间排序

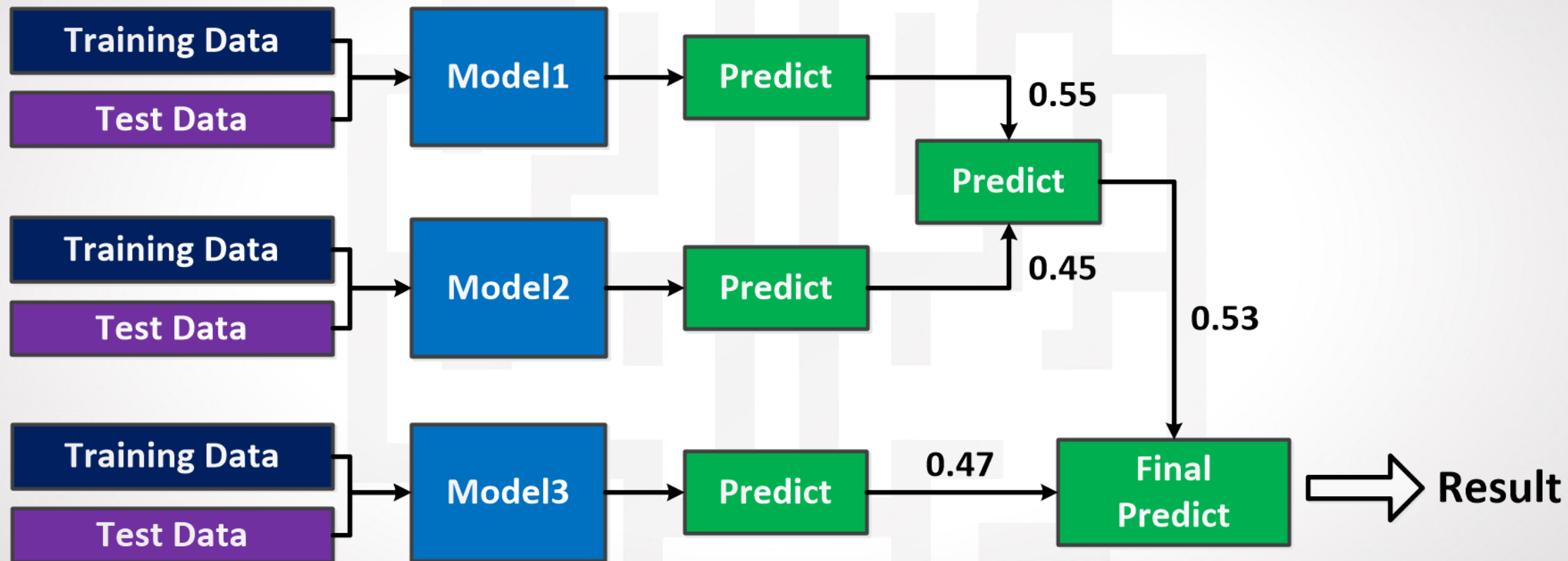
## PART 06

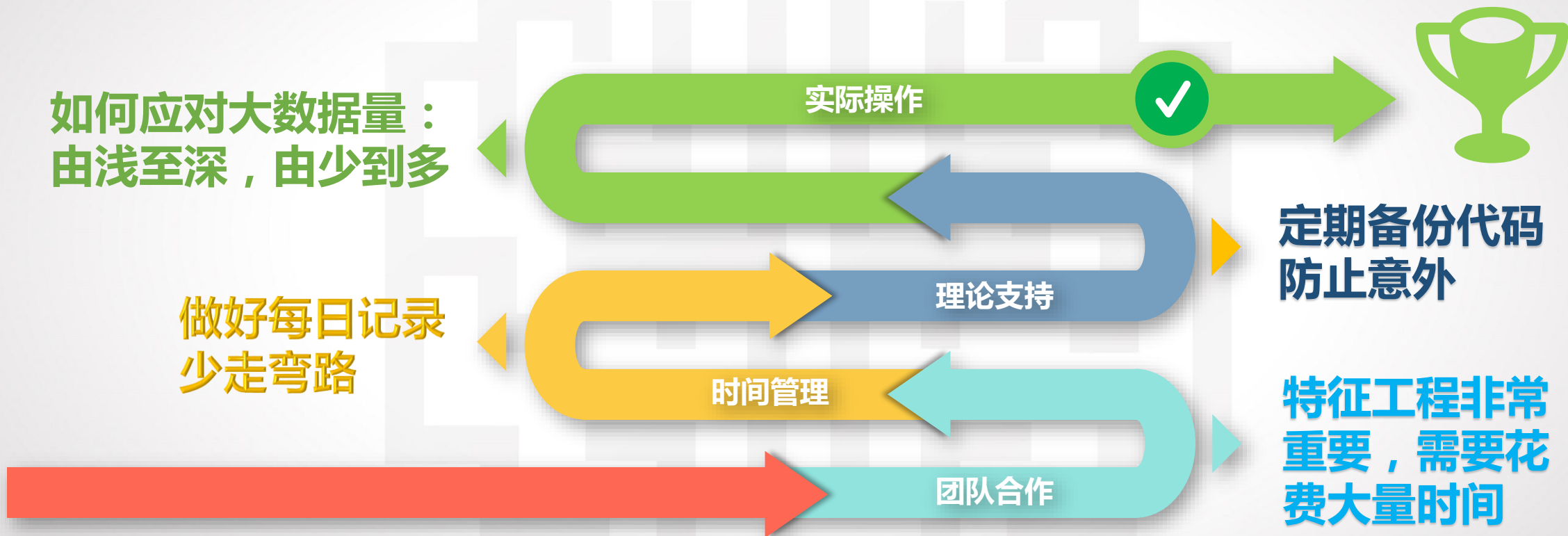
# 模型构建



## PART 06

# 模型构建







THANKS

freeze()

for

<=Date()

+

9

2017-07-06

8

#

W

3

if

Your

var

time