



復旦大學 管理学院
SCHOOL OF MANAGEMENT
FUDAN UNIVERSITY

数据智能与商业决策

——双美用户导入挑战赛

唐百川团队



目录

CONTENTS

01

【业务梳理】

02

【特征工程】

03

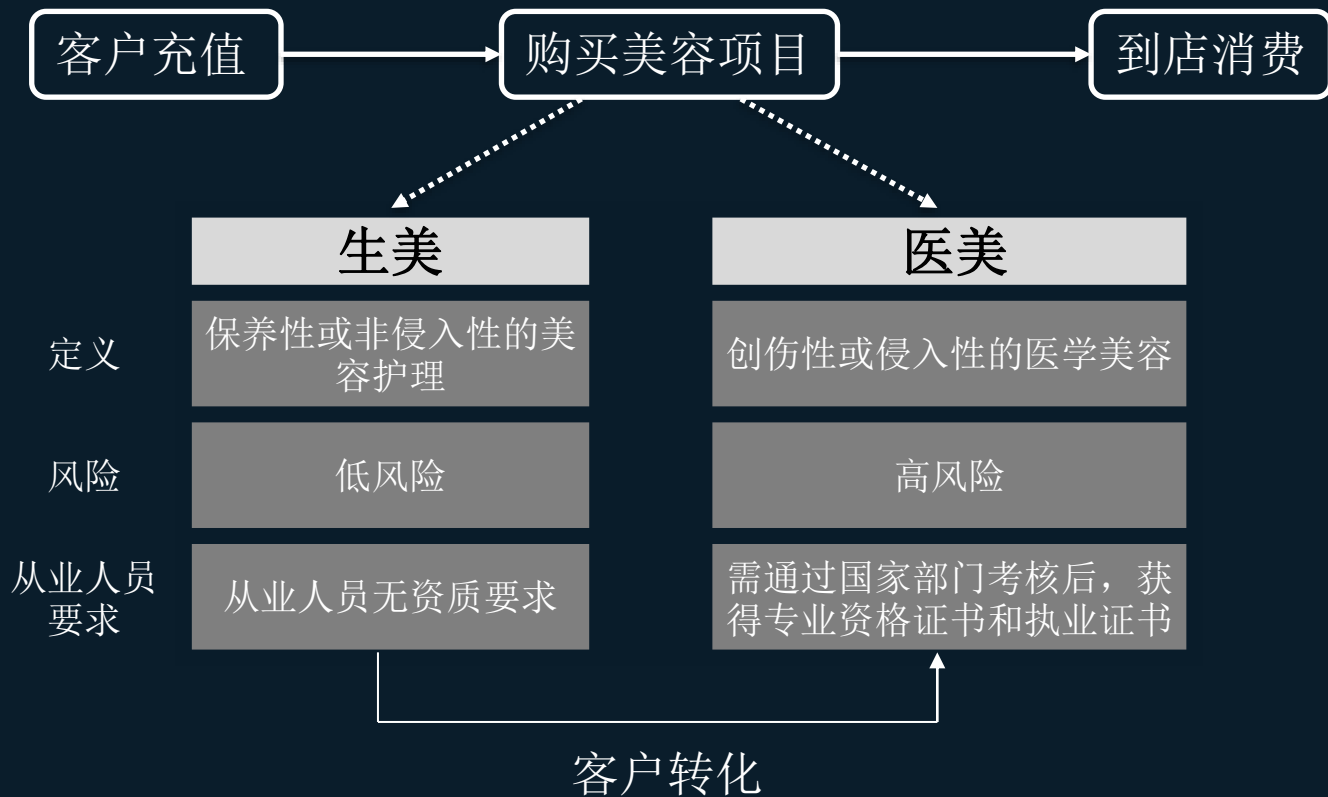
【模型训练】

04

【项目展望】

01

业务梳理

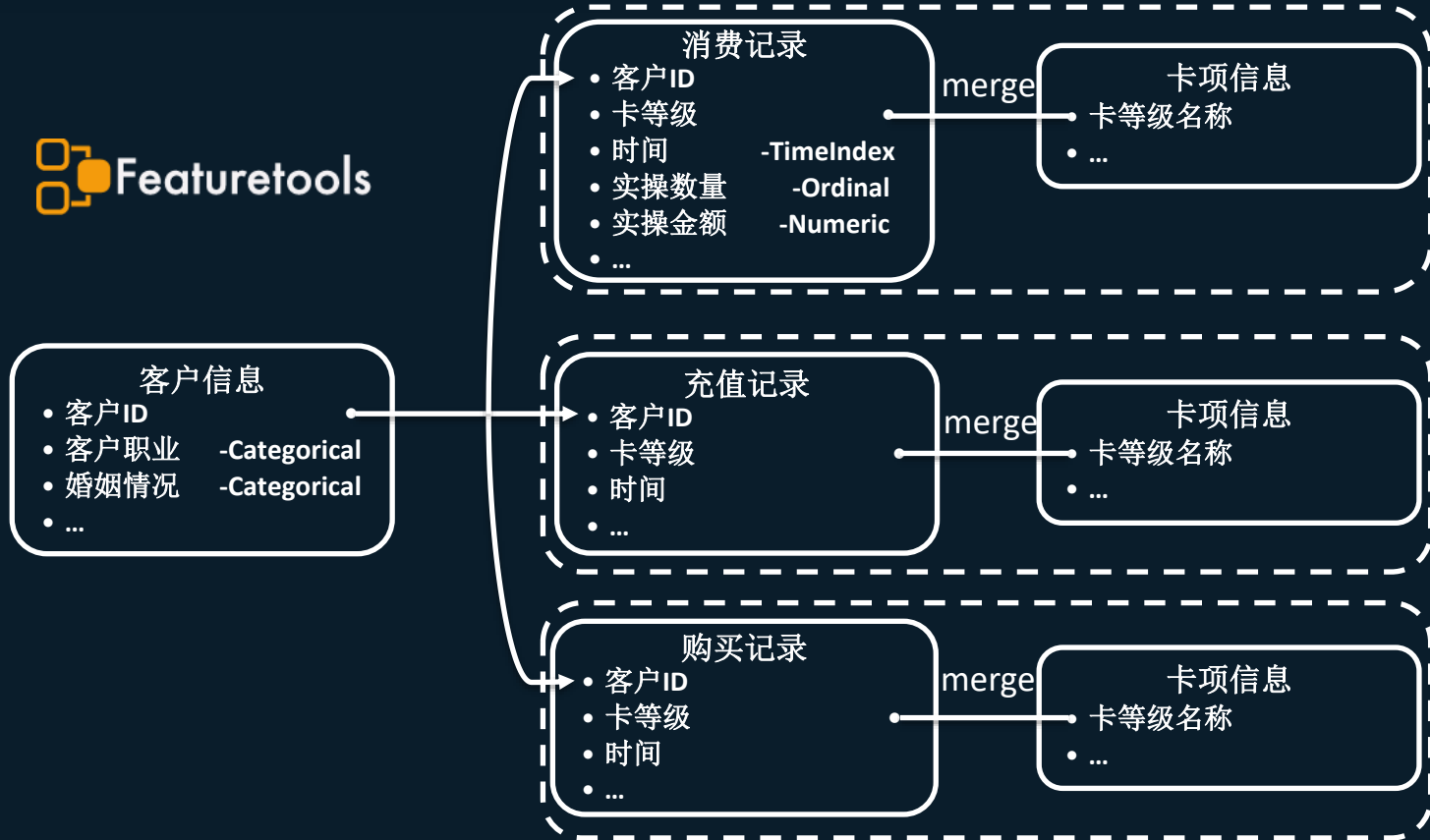


02

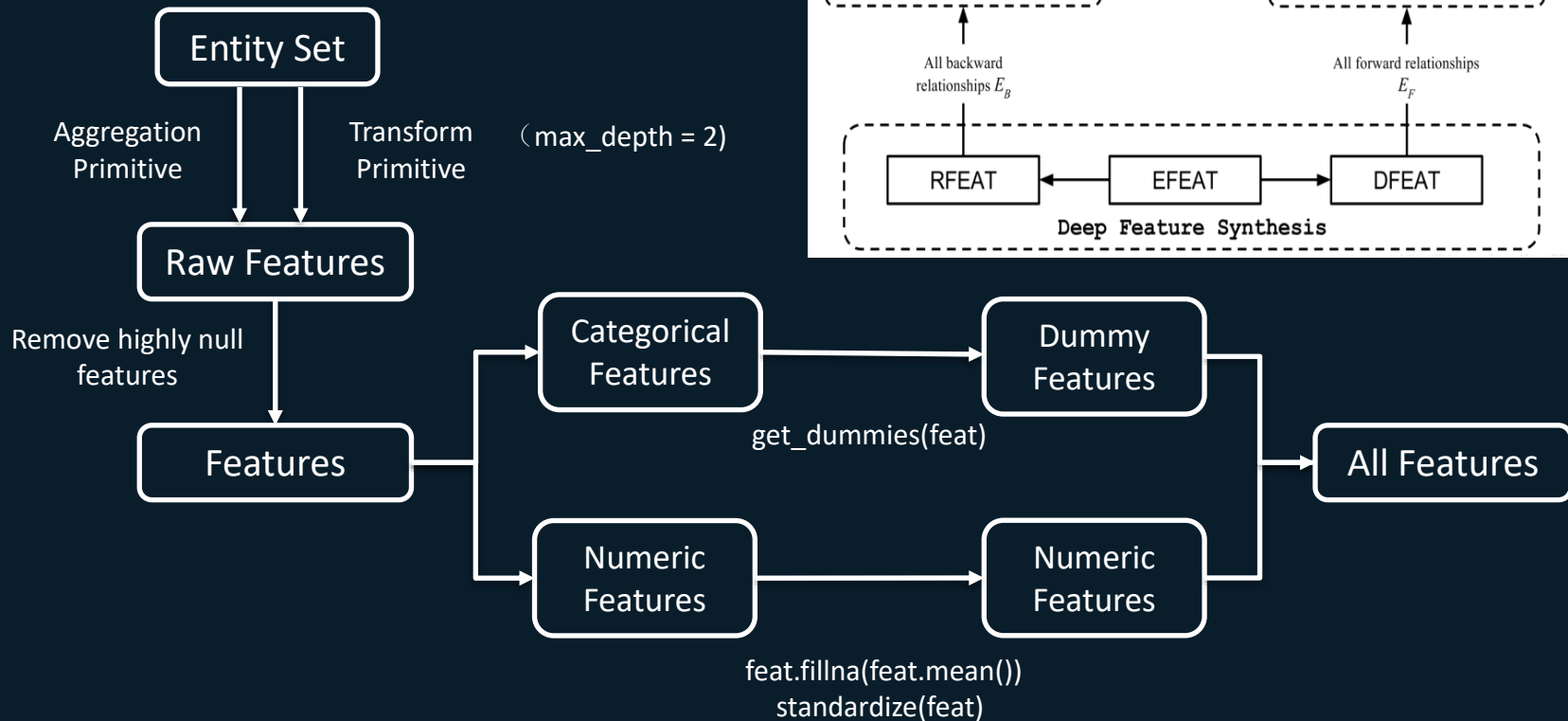
特征工程

- 搭建主附表关系
- 构造特征
- 基元选择
- 初步筛选特征

特征工程——搭建主附表关系



特征工程——构造特征



特征工程—基元选择

Aggregation Primitives:

‘sum’, ‘mean’, ‘std’, ‘mode’, ‘median’, ‘entropy’, ‘first’,
‘last’, ‘max’, ‘min’, ‘num_unique’, ‘time_since_first’,
‘time_since_last’, ‘n_most_common’,

.....

Transform Primitives:

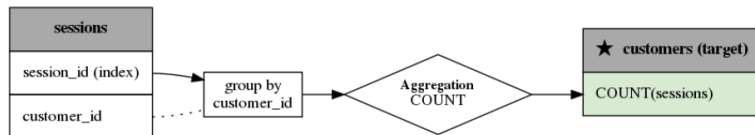
‘month’, ‘hour’, ‘weekday’, ‘is_weekend’, ‘percentile’,
‘diff’, ‘cum_count’, ‘cum_min’, ‘cum_max’,
‘cum_mean’, ‘cum_sum’, ‘time_since_previous’,

.....

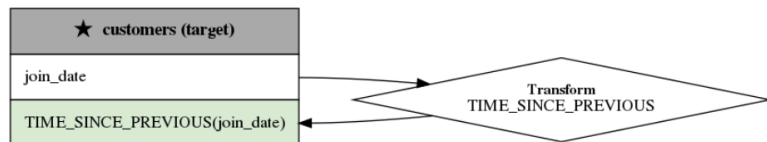
Aggregation vs Transform Primitive

In the example above, we use two types of primitives.

Aggregation primitives: These primitives take related instances as an input and output a single value. They are applied across a parent-child relationship in an entity set. E.g: "count", "sum", "avg_time_between".

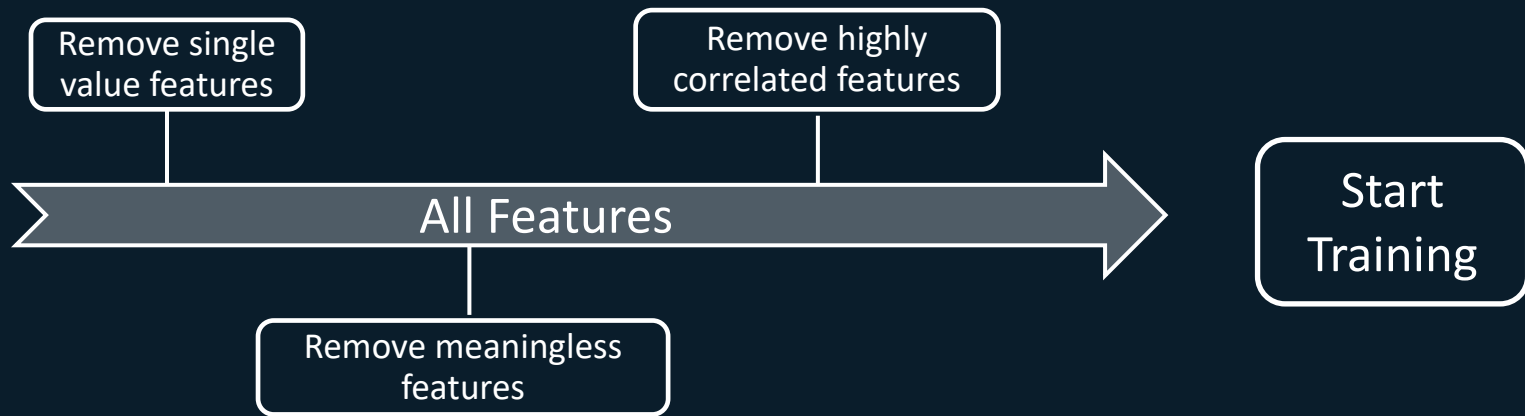


Transform primitives: These primitives take one or more variables from an entity as an input and output a new variable for that entity. They are applied to a single entity. E.g: "hour", "time_since_previous", "absolute".



The above graphs were generated using the `graph_feature` function. These feature lineage graphs help to visually show how primitives were stacked to generate a feature.

特征工程——初步筛选特征

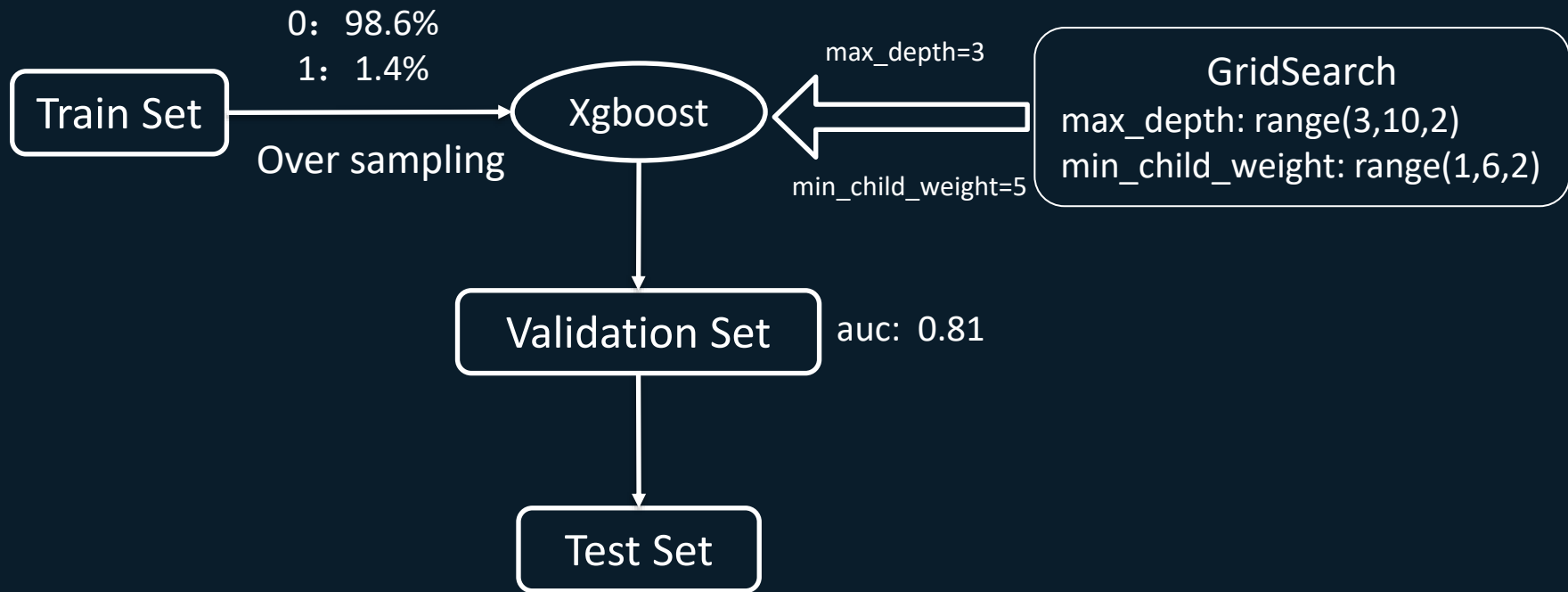


03

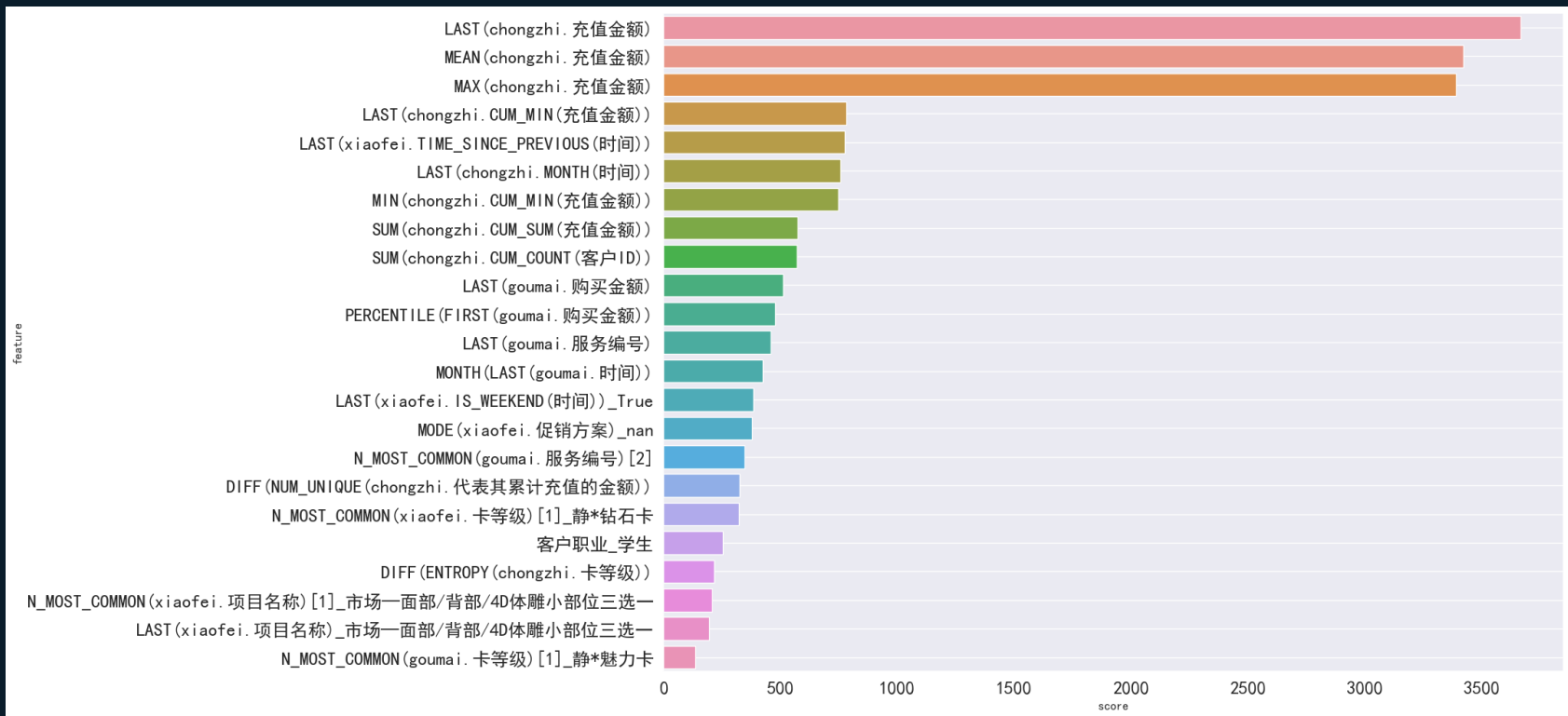
模型训练

- 模型处理
- 特征评分及解释

模型训练——模型处理



模型训练——特征评分及解释



04

项目展望

本次比赛

截面数据

2021.6.1

客户 1	1
客户 2	0
...	...
客户 i	1
...	...

数据类型

任务目标

在当前时间点用**70%**用户
预测剩下**30%**用户的行为

项目评价

无法真正上线



实际业务逻辑

面板数据

2020.1.1

2020.1.2

2021.1.1

2021.6.1

客户 1	1	1	...	1	...	1
客户 2	0	0	...	1	...	1
...			...			
客户 i	0	0	...	0	...	1
...			...			

用所有用户历史切片数据**对时间轴划分**训练测试集，完成模型调参与训练
用所有历史数据下一个月的用户行为

历史数据回流→模型自学习→形成闭环



復旦大學 管理学院
SCHOOL OF MANAGEMENT
FUDAN UNIVERSITY

数据智能与商业决策

——双美用户导入挑战赛

唐百川团队