

# 1.1 一个Kaggle竞赛优胜解决方案

CSDN学院 2017年10月



### ►任务: Avazu 点击率预估竞赛



- Avazu Click-Through Rate Prediction是移动广告dsp公司 avazu在kaggle上举办的广告点击率预测的比赛
  - https://www.kaggle.com/c/avazu-ctr-prediction
- 根据一个事件的性质预测该广告是否回被点击

id		click	hour	C1	banner_pos	site_id	site_domain	site_categor
1000009418151	094273	0	14102100	1005	0	1fbe01fe	f3845767	28905ebd

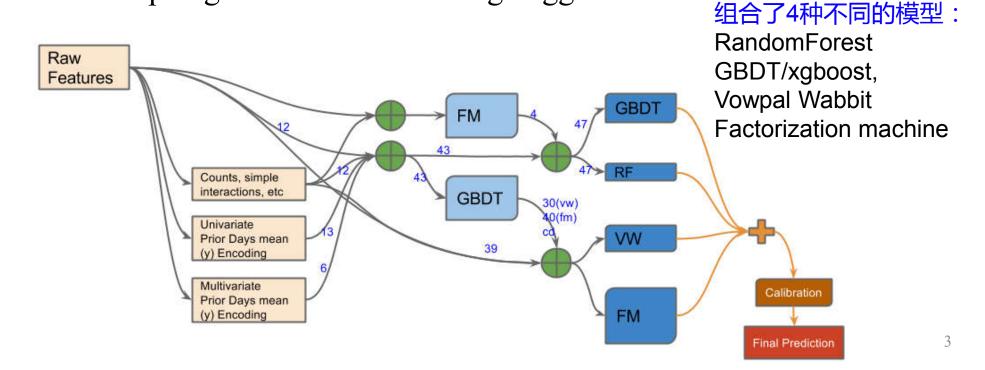
app_domain	app_category	device_id	device_ip	device_model	device_type	device_conn_type
7801e8d9	07d7df22	a99f214a	ddd2926e	44956a24	1	2

C17	C18	C19	C20	C21
1722	0	35	-1	79

## ► Rank 2nd Owen Zhang的解法



• https://github.com/owenzhang/kaggle-avazu



#### ▶优胜算法的特点



- 特征工程
- 融合大法
  - 多层
  - 多种不同类型模型的组合
- 所以:
  - 基础模型很重要(线性模型)
  - 集成学习模型单模型性能好(GBDT)
  - 特定问题的模型贡献大 (FM)
  - 模型融合很重要



#### ▶课程内容安排



- 基本模型
  - 线性模型:线性回归、logistic 回归、SVM
  - 非线性模型: (线性模型核化)、分类回归树
  - 集成学习模型(GBDT)
  - 数据预处理:数据清洗、特征提取、降维、聚类
- 模型融合
- 推荐系统 / 点击率预估问题特定解决方案





# THANK YOU



