

# 推荐比赛总结

# 数据预处理

- 程序脚本化
- --数据清洗、抽样
- 构建离线评测体系
- --训练集、测试集、验证集

# 特征工程

- 个性化推荐系统比其他的机器学习方法更依赖特征工程，特征比模型要对推荐效果的提升占的比重大。
- 利用领域知识生成和提取特征

如何描述个性化并且用变量表示成特征，一般方法是，想想你就是该商品的目标用户，你会想要什么样的个性化。

- 特征总的来说可以分为三个大类：
  - 用户特征
  - 品牌特征
  - 用户-品牌特征

# 模型

- 基于规则：
- 聚类
- 基本的分类器：
- SVM、LR(哑编码)、RF
- 组合模型/模型融合
- GBDT(Gradient Boosting Decision Tree)
- 长尾效应

- 流行月份的限定（近期）
- --每个商品的流行分数
- 购物狂
- --用户的消费指数
- 爱买的用户，流行的商品建立关联
- --用户-品牌特征：
- 用户品牌特征
- --用户所买的商品的流行度\*用户所买的商品+该商品在所有购买商品的比例
- 使用购买率在用户品牌列表TopK中筛选
- 局限性：买过的用户会再买，不会买的用户一次都不会去买。

- 知识管理文档化