- 基本建模流程
  - Stage 1.业务背景解读与数据探索
    - 业务背景探索
      - 大业务背景
      - 小建模目标
    - 字段含义解读
      - 根据数据字典解读
        - 竞赛/公司内部通常会提供
    - 数据分布检验/共线性检验
      - 正态分布/伯努利分布
      - 方差膨胀系数
      - 方差检验
      - 统计学特性的假设检验对机器学习建模来说非必要，但是是重要的了解数据的维度
    - 数据正确性校验
      - id是否重复检验
      - 字段是否和数据字典保持一致
      - 竞赛通常无误，但是是纪律性操作，以防万一
    - 数据质量检验
      - 缺失值
      - 异常值
      - 重复值
      - 简单情况下可以在探索时直接处理
    - 训练集/测试集规律一致性检验
      - 单变量分布一致性检验
      - 多维分布一致性检验
      - 一致则重建模，不一致则重trick
    - 标签的特征分布探索
      - 探索标签取值是否在不同特征上存在分布规律
  - Stage 2.数据预处理与特征工程
    - 字段类型调整
      - 转化成数值型
    - 重复值处理
      - 删除
      - 统计合并
    - 异常值处理
      - 修改
        - 天花板盖帽法
      - 标记
        - 可能是一类特殊情况
    - 缺失值处理
      - 填补
        - 统计指标填补
        - 模型结果
      - 删除
    - 需要结合业务情况进行判别
    - 特殊类型字段处理
      - 时序字段
      - 文本字段
    - 特征构建/衍生
    - 特征筛选
    - 特征工程部分重点内容
  - Stage 3.算法建模与模型调优
    - 模型选择与训练
      - 重可解释性
      - 重模型效果
      - 基本模型选择依据
    - 模型超参数调优
      - 网格搜索
        - 网格搜索
        - 随机网格搜索
        - Halving搜索策略
        - 权衡搜索效率与搜索精度
      - 贝叶斯优化器
    - 模型融合策略
      - 投票/权重法
      - Stacking
    - tricks
      - 大量依赖实战经验以及“灵感”