1. 数据获取
2. 从MySQL获取数据
3. 从MongoDB获取数据
4. 从本地读取
5. 数据预处理和探索
6. 标准化
7. 标签编码
8. One-hot
9. 缺失率统计
10. 特征选择
11. 缺失值填充
12. 选择模型和训练
13. 数据划分

2、根据任务目标选择合适的模型范围

3、多模型训练和评测

4、参数调节

1. 模型评估
2. 模型结果报告（表格）
3. 图表

ROC，特征重要性

3、

1. 模型应用

Problems：

1.分类变量和连续变量需要分开处理，最后合成为单个数据集

2.分类和回归任务的差异

评估标准不一样

分类任务：分类报告，roc曲线

回归任务：

mean\_squared\_error, mean\_absolute\_error, explained\_variance\_score and r2\_score.

3.数据预处理应该在数据切分之后再进行