

## Part1

### 任务 1:

按照 requirement 文档配置虚拟环境

### 任务 2:

读入数据，查看数据异常值情况，删除缺失值所在行，对数据完成清洗工作

### 任务 3:

对数据中一些关键信息，画图查看关键信息的分布。

根据数据类型不同，选择合适的画图类型例如想要观察某一数据取值在不同区间的比例可对其数据值按区间不同进行分类并对分类的数据的不同类别进行统计并画图，对所有数据分别进行分析。

### 任务 4:

将画图结果以 html 文件保存，方便之后进行查看

### 任务 5:

开始进行本项目中需要使用的几类算法的系统学习(逻辑回归、GBDT、mlp 模型)

## Part2

### 任务 1:

导入数据集，对数据集进行训练集测试集分割

### 任务 2:

学习 sklearn 包相关内容，分别建立逻辑回归、GBDT、mlp 模型

### 任务 3:

利用三种模型对训练集进行训练。根据前两周所学内容，调整超参数，对算法效果进行提升，对每一种算法显示其对应 precision score、recall score、f1 score 等参数，设置输出分别对应以上三种算法的 testdata 矩阵

### 任务 4:

根据算法运行情况以及上一步打印结果及数据集特点，调整模型超参数，针对以上情况

的不同多次调整模型参数到模型运行结果满意为止

任务 5:

学习 flask 框架原理及实现，学习 html 相关内容，利用 pickle 保存训练好的模型。为下一部分的工作做准备。

任务 6:

利用全体数据集进行训练，将训练好的模型进行保存

\* 选择性任务:

通过 Python 已有包中的部分函数，手写推导经典算法（逻辑回归），并创建简易一元数据集进行测试，并将算法打包成函数

## Part3

任务 1:

将效果最好的模型利用全体数据集进行训练，且将训练好的模型进行储存

任务 2:

利用 flask 框架在 Python 中编写算法，设置 flask 框架

任务 3:

编写 HTML 文件，按照自己喜欢的渲染方式构建算法配置的平台

任务 4:

在 python 中调整 html 语言下数据输入格式，并载入之前训练的最优模型，完成 Flask 框架