# 易访库系统开发文档

## 1. 开发环境及工具

系统：Windows 10 64bit

编程语言：Python 3.7

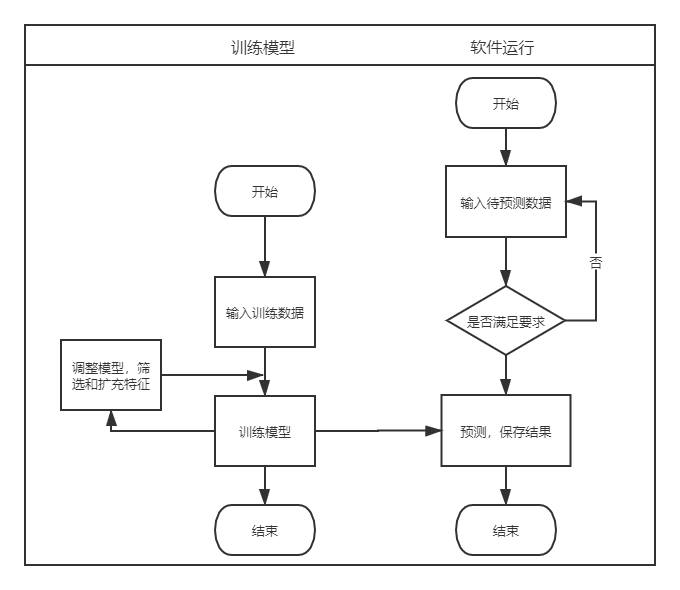
编辑器：VSCode

输入输出文件编辑工具：Excel 2019

辅助开发工具：Jupyter Notebook

## 2. 项目需求描述

基于PyQt5实现易访用户识别可视化系统，系统的主要功能包括基于给定的用户信息识别其是否属于易访用户，以及识别结果的保存和展示。对易访用户的识别主要采用决策树或随机森林等机器学习算法，基于用户的基本信息和通信行为数据训练模型来实现。用户的基本信息包括性别、年龄、是否宽带用户等；通信行为数据包括近三月主叫时长，近三月国内流量等。基本功能实现后，可进一步采用特征工程方法进行特征的筛选和扩充，研究特征工程对分类器性能的影响。



## 3. 模块与关系

#### 3.1 训练模块

核心技术：决策树和随机森林算法，Pandas库，Numpy库

模块功能：基于输入数据训练用于预测易访用户的机器学习模型，提供给预测模块，同时基于模型优化模块进行调整。

#### 3.2 预测模块

核心技术：PyQt5，Pandas库，Tkinter库

模块功能：提供程序的主体执行界面和预测功能，依赖训练模块的模型。

#### 3.3 模型优化模块

核心技术：特征工程，Pandas库，MatPlotLib库，Numpy库

模块功能：筛选和扩充特征，用于改进预测模型的性能。

## 4. 术语解释

|  |  |  |
| --- | --- | --- |
| **序号** | **术语** | **说明** |
| 1 | 易访用户 | 指愿意接听电信回访电话且配合受访的电信用户 |
| 2 | 特征工程 | 特征工程是将原始数据转化成更好的表达问题本质的特征的过程，使得将这些特征运用到预测模型中能提高对不可见数据的模型预测精度。 |
| 3 |  |  |

## 5. 用户界面

