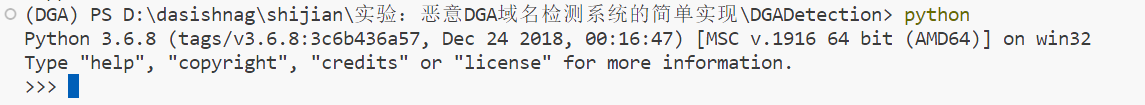
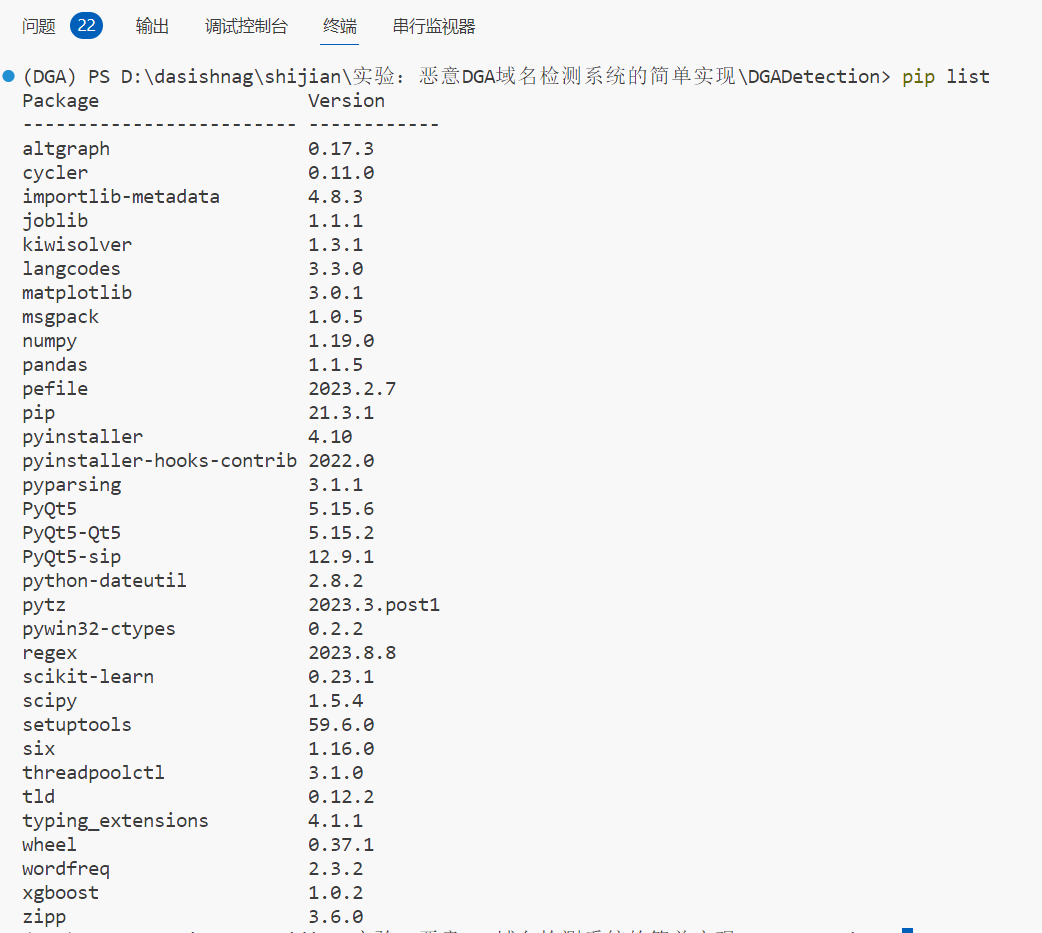
1. **实验环境**

操作系统：windows

语言：Python

编译器：Python3.6.8





1. **实验内容**
   1. **理解代码结构，配置环境（10分）**
2. 代码结构

|\_\_\_\_\_data

|\_\_\_\_\_sample——样本文件夹

|\_\_\_\_\_sample\_black.csv——恶意域名样本

|\_\_\_\_\_sample\_white.csv——正常域名样本

|\_\_\_\_\_modle——模型文件夹（存放训练好的模型及归一化尺度）

**|\_\_\_\_\_**feature——特征文件夹（存放已提取特征的训练集和测试集）

|\_\_\_\_\_static  
 |\_\_\_\_\_\_gib\_model.pki——Gib成文检测依赖文件

|\_\_\_\_\_feature\_extraction.py——特征提取代码

|\_\_\_\_\_detection.py——模型训练、测试及域名检测代码

|\_\_\_\_\_requirements——依赖包及版本

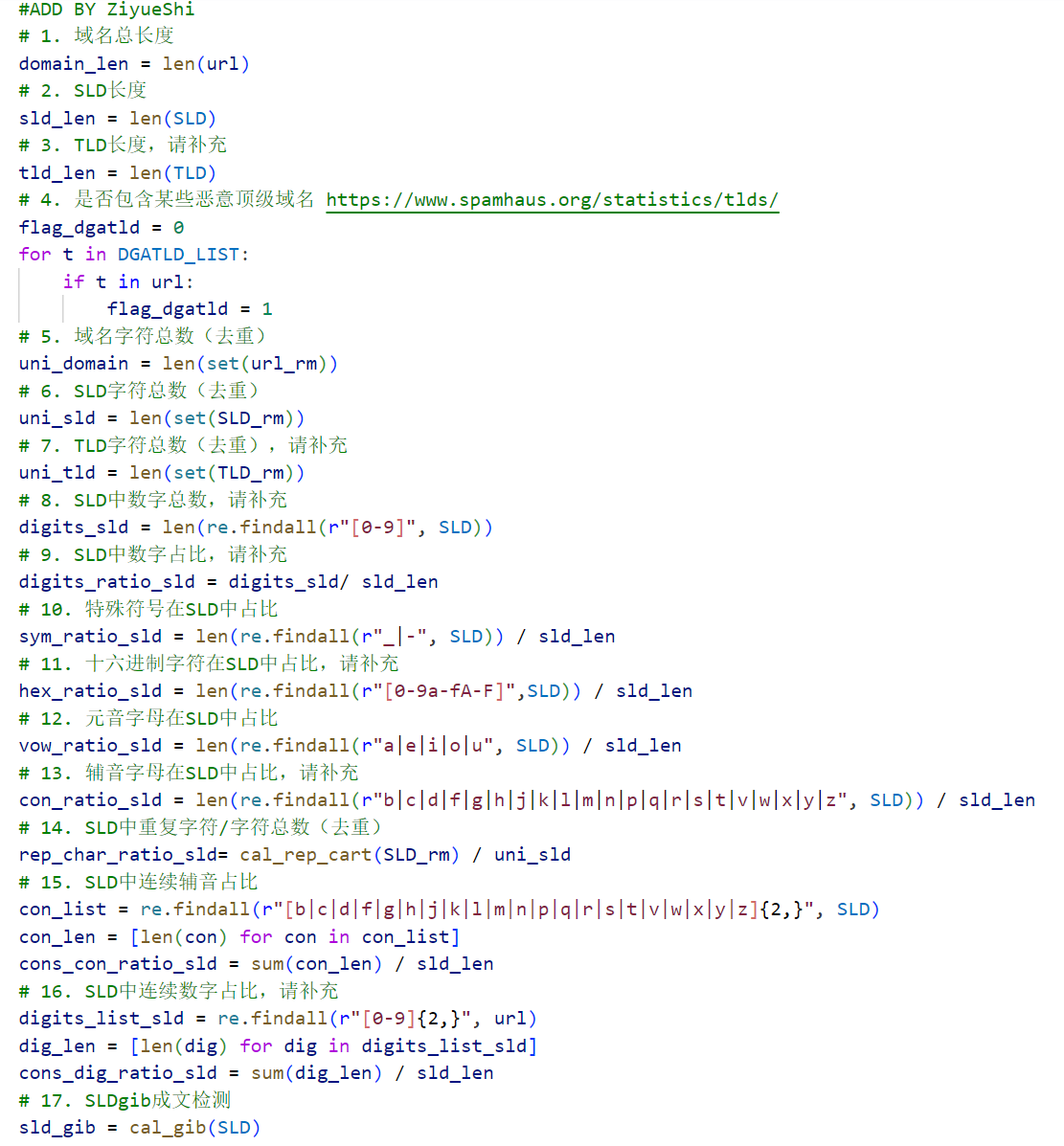
1. 环境配置

* 安装Python3.6（Ubuntu18.04系统已内嵌Python3.6，Windows系统需自行下载安装<https://www.python.org/downloads/>）
* 使用pip安装requirements中的依赖包



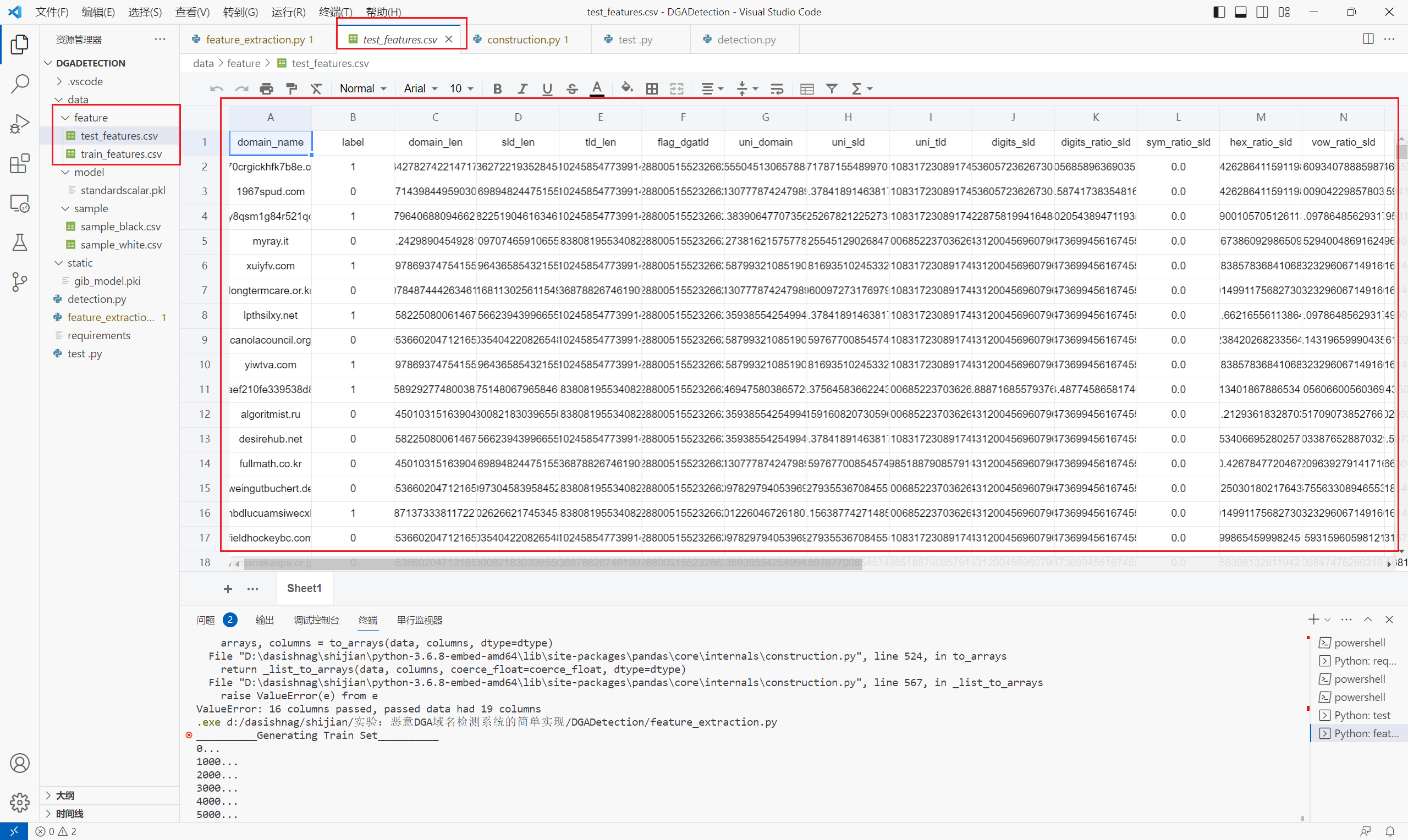
* 1. **补充feature\_extraction.py文件（30分）**

1. 结合注释学习并理解feature\_extraction.py文件中的各功能函数
2. 在已给代码的基础上补充函数get\_feature()，提取域名的至少17维特征



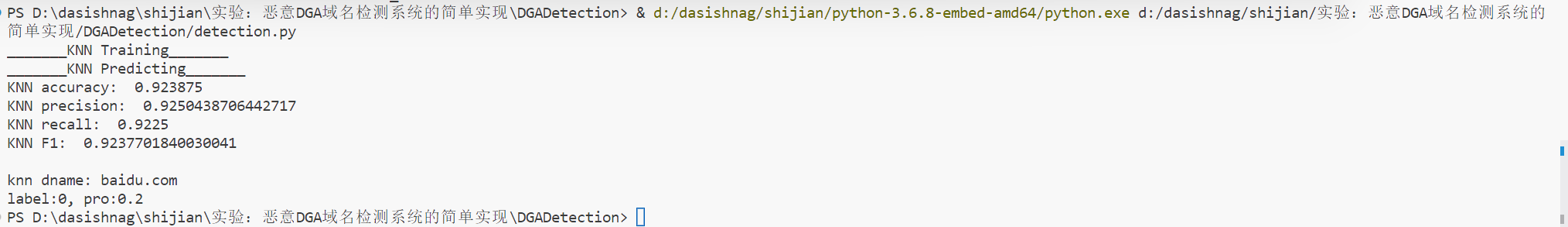
1. 在主函数中运行数据集生成函数dataset\_generation()，成功生成训练数据及测试数据



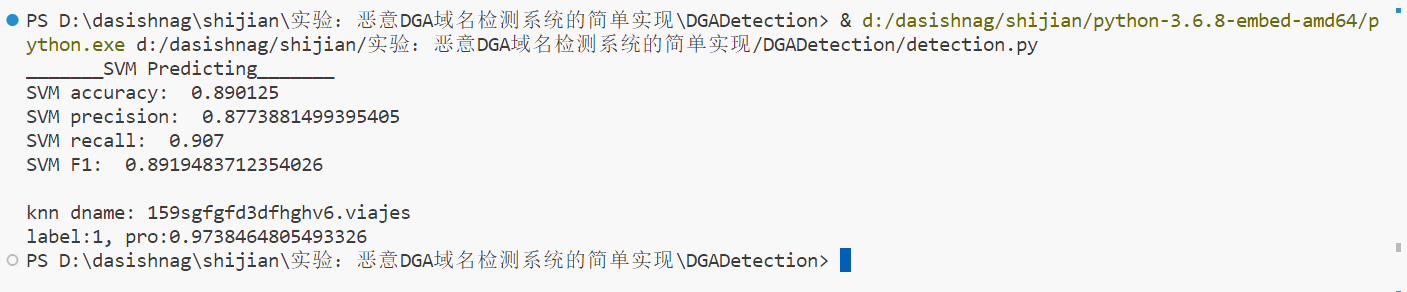


* 1. **补充detection.py文件（30分）**

1. 结合注释及KNeighborsClassifier官方文档（https://scikit-learn.org/stable/modules/generated/sklearn.neighbors.KNeighborsClassifier.html#sklearn.neighbors.KNeighborsClassifier）学习并理解自定义类KNNClassifier的功能，完成KNN分类器的训练，得到测试集在该模型上的表现。



1. 仿照自定义类KNNClassifier，参考SVC官方文档（https://scikit-learn.org/stable/modules/generated/sklearn.svm.SVC.html）编写类SVMClassifier，支持向量机分类器的训练、测试和其他功能。在此基础上，完成SVM分类器的训练，得到测试集在该模型上的表现。



1. 思考：模型如何调优？

超参数调优，特征工程优化，

* 1. **简单可视化（20分）**

**这里使用PyQt5进行可视化，在源代码目录下提供了DGA.exe可供使用。**

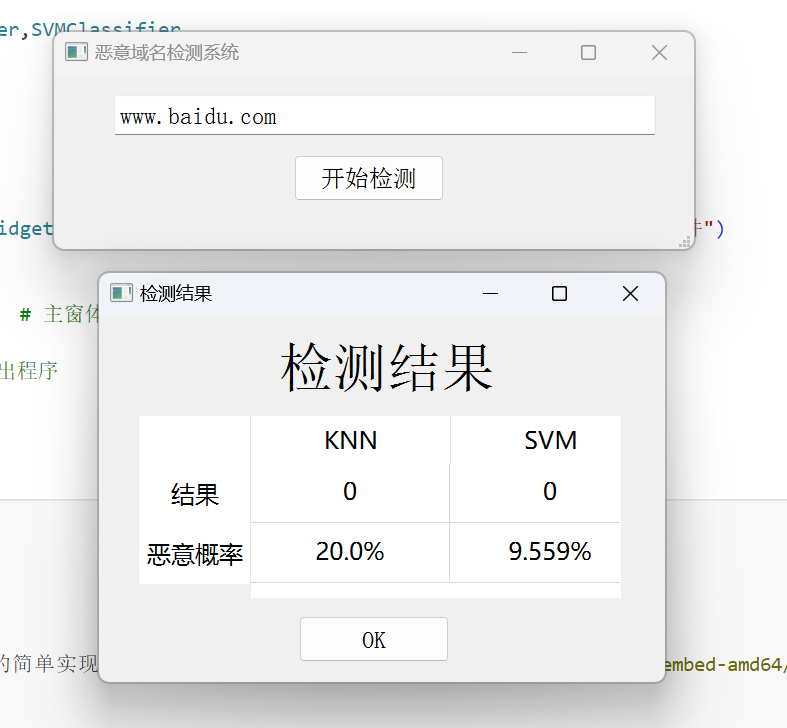
* 可输入域名字符串（对字符串有初步的合法性检查，如输入有误应给出提示）

合法性检查：



* 返回检测结果及恶意概率

这里将小数变成了百分数，感觉更易于理解



实验心得  
 通过这次实验进一步加强了对KNN与SVM的理解，同时熟悉了其在Python下的应用。对DGA域名有了更深刻的了解。