**测试结果**

**1. 测试执行情况**

本系统的测试内容依据测试方案文档进行。根据小组实际开发进度，在第二周周末（9.4-9.5）对模型表现进行了测试，在第三周周中（9.7-9.8）对程序功能及界面进行了测试。截止目前（9.10），所有测试内容均已完成并通过。

**2. 测试结果**

1. 初次测试用例执行结果

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 测试用例编号 | 测试对象 | 测试输入 | 预期输出 | 实际输出 | 评价 |
| 1.1 | 未训练的分类器 | 所有训练样本 | 训练后的分类器及模型的表现参数 | 模型文件(.pkl)  模型参数json文件和混淆矩阵、ROC曲线图片 | **通过** |
| 1.2 | 训练后的分类器 | 分类器的表现参数 | 评价模型是否达到可用要求 | 根据参数进行人工评价，最低要求达到课题规定的75%准确率 | **通过**，训练的APT组别分类模型准确率达到90%左右，国别分类模型准确率达到95%左右，超过课题要求的75%准确率 |
| 2.1 | APT预测程序 | 点击界面“浏览”按钮，选择单/多个病毒报告文件；在下拉框选择模型 | 对话框中显示选择的模型和所有待测文件路径，为预测做好准备 | 与预期输出相同 | **通过** |
| 2.2 | APT预测程序 | 在2.1的基础上，点击开始预测按钮 | 结果显示区域展示使用的模型，预测文件名，预测概率分布，预测分类 | APT组别预测显示正常，国别分类时显示预测结果为数字而非可理解的国名字符串 | **不通过** |
| 2.3 | APT预测程序 | 在2.1的基础上，调整阈值滑动条，点击开始预测按钮 | 观察到预测分类与2.2有所变化，更多/更少的样本被划分到其他分类 | 调高阈值，部分被分为某APT的样本分类结果变为其他分类；调低阈值时相反 | **通过**，阈值参数正确影响了模型的预测结果 |
| 2.4 | APT预测程序 | 选中某个模型，点击现实模型参数按钮 | 弹出新窗口，显示该模型的算法、采样策略信息和精确率、召回率等表现参数，以及混淆矩阵和ROC曲线两张图片 | 与预期输出相同 | **通过** |
| 2.5 | APT预测程序 | 输入云数据库地址，点击连接 | 左下角状态栏刷新，显示连接到数据库 | 与预期输出相同 | **通过** |
| 2.6 | APT预测程序 | 选择数据库已有文件，点击查询数据库 | 显示数据库中对应文件的预测结果 | 与预期输出相同 | **通过** |
| 2.7 | APT预测程序 | 在2.2的基础上，点击上传预测结果，并点击刷新 | 在数据库已有文件下拉框中出现刚上传的文件，可以查询其预测结果 | 预测单文件上传正常，但多文件时只会上传预测的第一个文件 | **不通过** |

2. 测试总结

初次测试中发现分类模型导出和预测功能正常，输出的表现参数显示模型可以以较高准确率完成预测工作，达到课题要求。程序功能上出现了一些问题，将模型隔离到单独项目中测试后发现模型所有输出均正常，说明是界面逻辑中存在问题。另外UI设计上测试人员发现一些标签和按钮的文字表述存在含糊不清，撰写了修改意见。以上测试结果于9月7日向界面开发人员进行了反馈。

开发人员检查后确认了问题存在，并于当日进行了修改，提交第二版程序。经测试所有问题均得到解决。测试负责人于9月8日确认本系统通过测试。