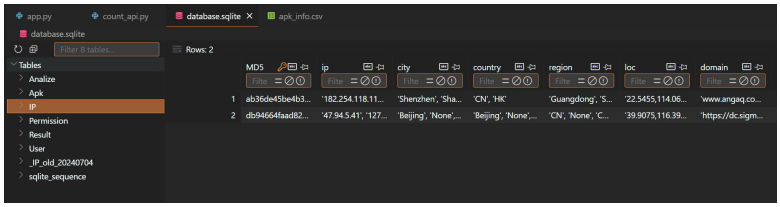
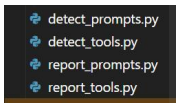
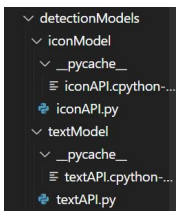
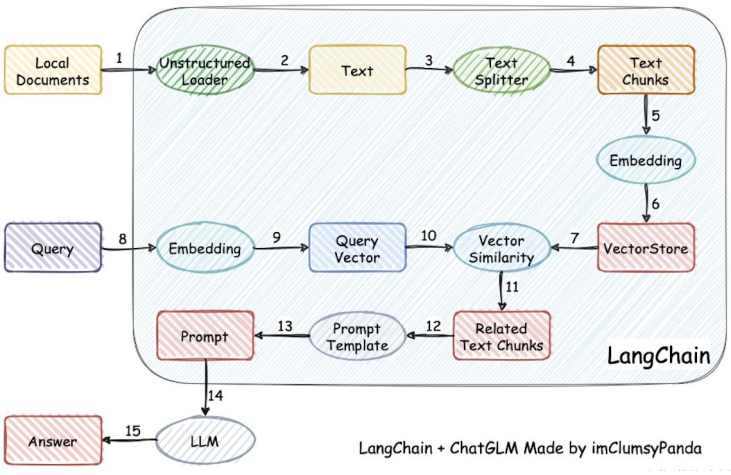
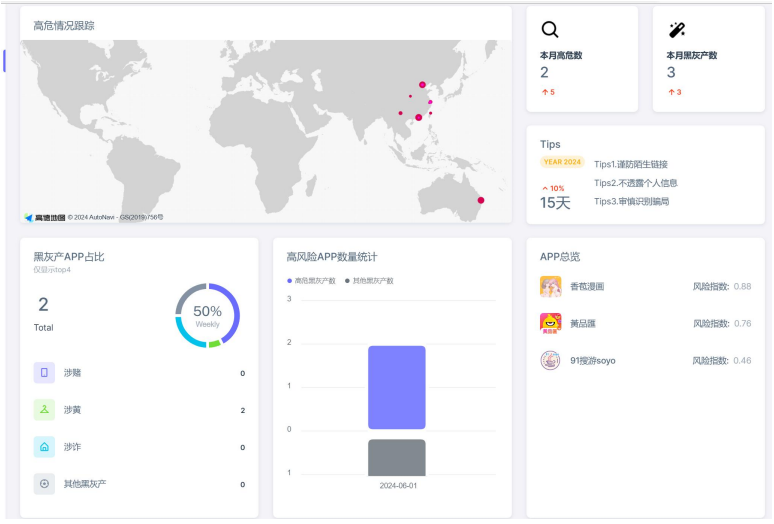
重庆大学大数据与软件学院

软件综合实践第二周进度报告

<table>
<tr><td>项目名称</td><td colspan="4">基于深度学习和大模型的 android 恶意软件检测系统</td></tr>
<tr><td>项目成员</td><td>学号</td><td>20221982</td><td>姓名</td><td>潘铀葳</td></tr>
<tr><td rowspan="3">组内其他成员</td><td>学号</td><td>姓名</td><td>学号</td><td>姓名</td></tr>
<tr><td></td><td></td><td></td><td></td></tr>
<tr><td></td><td></td><td></td><td></td></tr>
<tr><td>项目进展情况
（请认真思考填写，至少 100 字以上）</td><td colspan="4">1. 优化和扩展数据特征，显著提升了分析的准确性和深度。我不仅新增了应用的图标、域名、IP 地址、城市、国家及地理坐标等关键特征，还对现有特征如应用名称、包名、主要活动、权限及证书信息进行了深入优化和清洗，确保数据的完整性和一致性，并构造了新的数据库：
 
2. 训练了两个深度学习分类模型：一个针对文本特征，使用 BERT、LSTM 和 Transformer 进行对比，最终保留了准确率最高的 Transformer 模型；另一个针对图像特征，利用 CLIP 模型进行 Logo 图片的多分类。通过集成这两个模型，实现了多源数据的融合与分类：
 
 
3. 为了进一步提升 APK 分析和分类的能力，我们利用 LangChain 结合大模型构建了多模态模型，能够同时处理文本和图像特征，实现更全面和精确的分类。通过 LangChain，我们高效地集成和管理文本与图像数据，利用预训练模型和自定义算法进行深度特征提取，并采用 Transformer 和 CLIP 模型分别处理文本和图像特征。我们通过特征融合和深度神经网络训练，实现了对多模态数据的综合分析和分类：</td></tr>
</table>



4. 为了提升用户体验和展示效果，我对前端页面进行了全面优化。在首页中，我集成了高德地图 API，添加了高危地区 IP 的地图可视化功能，使用户能够直观地看到高风险 IP 地址的地理分布。在项目详情页面中，我实现了 IP 追踪定位和黑灰产联系图展示，提供了域名、IP 和地区名字的详细信息，集成了自训练多模态模型的分析报告，展示了应用的权限控制列表和关键观测指标。通过这些优化，用户能够清晰地了解黑灰产应用的来源和活动区域，查看应用之间的联系，并获取详细的分类报告和风险评估。通过这些前端页面的优化，我们不仅提升了系统的可用性，还提供了丰富的数据展示和分析功能，帮助用户更好地理解和应对 APK 相关的安全风险：



5. 后端逻辑的编写，包括以下内容：

（1）地图可视化数据处理

高危地区 IP 数据获取：从数据库中提取高危 IP 地址及其地理位置信息。

高德地图 API 集成：将高危 IP 地址数据与高德地图 API 进行整合，生成可视化地图数据。

（2）项目详情页面数据处理

IP 追踪定位：接收 IP 地址请求，调用高德地图 API 获取定位信息，并返回给前端。

黑灰产联系图数据生成：提取和处理黑灰产应用之间的关联数据，生成联系图所需的数据格式。

域名、IP 和地区名字信息获取：从数据库中获取并处理与高危 IP 相关的域名、IP 和地区名字信息。

多模态模型分析报告生成：调用自训练的多模态模型，对上传的 APK 进行分析，并生成分类报告和风险评估。

权限控制列表生成：提取和处理 APK 请求的权限信息，生成权限控制列表。

观测指标数据处理：收集和处理 APK 的关键行为和网络活动数据，生成观测指标。

（3）权限控制与用户管理

用户交互优化：优化用户交互相关的逻辑，提升系统的可用性和用户体验。

（4）项目列表管理

<table>
<tr><td></td><td>项目列表数据获取: 从数据库中提取已检测的 APK 列表，包括项目名称、检测时间、漏洞数量、修复进度和状态等信息。
项目状态更新: 实现项目状态更新逻辑，确保项目列表中的信息是最新的。
（5）项目上传功能
多种上传方式支持:
链接上传: 接收用户提供的链接，下载 APK 文件并保存到服务器。
二维码上传: 解析用户扫描的二维码，下载 APK 文件并保存到服务器。
本地上传: 处理用户通过本地文件上传的 APK 文件，并保存到服务器。
文件存储与管理: 实现 APK 文件的存储和管理逻辑，确保文件的安全性和可用性。
（6）项目配置管理
黑名单配置: 实现黑名单的添加、删除和修改功能，支持用户对 APK 进行分类和风险评级。
白名单配置: 实现白名单的添加、删除和修改功能，支持用户对 APK 进行分类和风险评级。
自定义名单配置: 提供自定义名单配置功能，增加管理的灵活性和精细度。</td></tr>
<tr><td>问题</td><td>多模态模型分析报告的实时性问题：在高负载时期，多模态模型分析 APK 的速度可能变慢，导致用户等待时间过长或无法及时获取分析报告。
可能的原因：模型计算资源不足或数据处理流程不够优化。</td></tr>
<tr><td>下一步工作任务安排</td><td>1. 解决多模态模型分析报告的实时性问题：预先优化模型的计算性能，考虑使用分布式计算或缓存机制加速分析过程，确保在不同负载情况下都能快速生成分析报告。
2. 建立关系网、实现可视化展示和训练节点检测模型：首先，我将从数据库中收集并整理数据，定义关系网中的节点和边，构建一个完整的关系网。然后，选择合适的可视化工具，将关系网的可视化功能集成到前端页面中，实现交互式图表的动态更新。最后，将对关系网中的节点进行标注和数据划分，选择适用于图结构数据的深度学习模型进行训练，并通过性能评估和参数优化，不断提升模型的检测能力和鲁棒性。</td></tr>
<tr><td rowspan="2">项目进度自评
（在相应栏目内打“√”）</td><td>正常</td><td>滞后延期</td><td>进展超前</td></tr>
<tr><td></td><td></td><td>√</td></tr>
<tr><td>教师意见</td><td colspan="3"></td></tr>
</table>

	教师签字： 年　　月　　日

大数据与软件学院制表