### 项目实施的进展情况及取得的成果

在 Python 方面，我们小组学习了一些基础知识和一些稍微高级的用法（正则表达式等）。这些知识对小组接下来的工作非常有帮助。此外，我们还学习了 Qt/C++，跟着教程完成了一些简单的项目。对于此次项目，小组写了一个简陋的前端，并留下了用 lambda 表达式完成的接口。

在前端方面，小组已经取得了一些进展。我们已经学习了 Qt/C++，并跟着教程完成了一些简单的项目。我们还写了一个简陋的前端，并留下了用 lambda 表达式完成的接口。这些工作为我们未来的项目奠定了坚实的基础。

在后端方面，我们小组学习了机器学习的传统算法，如 KNN、Kmeans、决策树等。还学习了深度学习的经典算法，如 CNN、RNN、LSTM。此外，我们还学习了 NLP 的经典算法及模型，如 Transformer、BERT、T5 等。我们还学习了一些技术，如迁移学习。目前，我们能够使用 hugging\_face 提供的预训练模型，并跟着 hugging\_face 官方的教程，完成了 Amazon 商品评论摘要的项目。

项目部署方面，小组成员在项目部署方面取得的进展。我们已经熟悉了 Linux（Arch）环境，并能够使用 docker-compose 等命令行工具完成项目的部署。我们还能够编写一些简单的 Shell 脚本。

### 遇到的的困难及下一步工作计划

不可否认，我们的项目确实与原定计划有所滞后。对此，我们深感抱歉并进行了反思和检讨。在此基础上，我们已经采取了一系列措施来加快项目进度。

首先，在前端方面，我们意识到基于 Qt/C++编写的前端过于粗糙，需要进行改进。因此，我们决定后续使用基于 flask 框架的 web 前端。这个决策是基于小组成员使用过 jupyter, alist 等 web 前端的启发，我们认为这些前端无论是审美还是效果都很好。

其次，在后端方面，我们遇到了一些困难。小组成员学习了相关的算法，但由于数学水平不够跟不上进度，导致进展缓慢。同时，数据集的制作也可能会比较繁琐，训练模型算力不够，互联网上的资料良莠不齐等问题也给我们带来了很多挑战。

针对这些问题，我们已经制定了解决方案。首先，小组成员将尽快提升自己的数学水平，确保能够应对项目中的相关算法。其次，我们将使用 python 的 scapy 爬虫来制作数据集，并使用 Doccano 标记工具或 ChatGPT 提供的 API 来实现数据的自动化标记。最后，在算力不够的情况下，我们将在本地机上执行小样本的训练，并在调试完毕以后，在算力服务器上执行大样本的训练。

在项目部署方面，由于后端还未成型，进度略有滞后，但并未遇到任何困难。我们的小组计划在暑期完成后端的开发工作。

深度学习开发大致分为以下几步：

首先，我们将使用scapy在计科飞书群和哈工大官网上爬取通知，并使用ChatGPT提供的API进行数据自动标注。接着，我们将使用dynamic padding和dynamic truncation技术将输入的通知转化为固定长度，并利用word2vec模型将文章转化为词向量（大概是300维的向量）。这两步都可以使用hugging\_face提供的封装好的包——tokenize来实现。

然后，我们将训练模型。得益于BERT（分类）预训练模型和T5（文本摘要）预训练模型，我们不必像GPT那样训练一个1750亿参数的大模型。我们只需在预训练模型的基础上进行微调。

最后，我们将使用pytorch封装好的metrics包进行模型评估。

以上是文本分类和文本摘要的大致步骤。此外，项目还可以使用正则表达式等技术匹配ddl日期，并投递到日历相应的日期中。

### 结题预期目标

实现通知内容管理 App，支持通知信息分类、信息精简、重点关注、模糊搜索、生成日 程安排五项功能。

### 经费使用情况

暂无经费使用