**附件1**

入选编号： （学生不填）

**东华大学数学与统计学院“树苗”计划**

**——创新实践育人项目申请表**

负责人姓名： 陈依林

负责人班级： 数学2301

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 东华大学数学与统计学院制表 | | | | | | |
| 填表日期： | 2025 | 年 | 5 | 月 | 21 | 日 |

1. **项目团队成员信息**

*队伍人数3~8人，第一栏填写负责人信息*

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **序号** | **姓名** | **班级** | **学号** | **联系电话** | **邮箱** | **团队分工** |
| 1 | 陈依林 | 数学2301 | 2301  10101 | 1843910  1006 | 406511381  @qq.com | 1.思路分析  2.NLP模型  3.推荐系统建立 |
| 2 | 徐文睿 | 统计2301 | 2301  10104 | 1587066  1092 | Xwr271744  @outlook.com | 1.数据收集  与处理  2. 推荐系统建立 |
| 3 | 刘炫阳 | 统计2301 | 2301  10103 | 1587309  1943 | Cjwdkalxy  @qq.com | 1.资料收集  2.效果指标对比  3.论文写作 |
| 4 | 叶琳婕 | 统计2301 | 2301  10106 | 1530085  4363 | yelj12304  @163.com | 1.文献查找  2.可视化展示  3.论文写作 |
| 5 | 范森林 | 数学2301 | 2304  00518 | 1753015  2423 | 321348477  @qq.com | 1.数据收集  与处理  2. NLP模型 |
| 6 | 彭邱翌 | 统计2301 | 2301  10118 | 1767188  4233 | 230110118  @mail.dhu.edu.cn | 1.可视化展示  2.效果指标对比 |
| 7 |  |  |  |  |  |  |
| 8 |  |  |  |  |  |  |

**注：队伍中必须包含大一或大二学生**

1. **项目志愿情况**

*请填写“序号+项目名称”，最后列“是”请打勾,“否”则留空。*

|  |  |  |
| --- | --- | --- |
| **第一志愿** | 16.基于大语言模型微调的零售优惠信息提取 | * **是否已和老师初步沟通** |
| **第二志愿** | 13.图结构数据预处理算法及其应用 | * **是否已和老师初步沟通** |
| **第三志愿** |  | * **是否已和老师初步沟通** |
| **R是否服从调剂** | |  |

1. **对意向项目的初步设想与理解**

*本栏未展示完全者可另附页*

|  |  |
| --- | --- |
| **第一志愿** | **一、项目背景：**  在当前数字零售环境中，商家通过多种渠道（如电商APP、短信、推送、公众号、海报等）发布大量优惠信息，内容表达复杂、多变，往往包含促销形式、金额、有效期、适用商品等多个维度。然而这些信息大多以非结构化文本存在，不便于直接被计算机识别或用于精准推荐。  传统的关键词匹配或规则抽取方法难以应对表达方式多样、上下文依赖强的促销语句（如“第二件半价”“买一送一”“满199减50”等）。同时，如何结合用户行为和兴趣，将匹配的优惠精准推送到目标用户，也是提升用户体验和转化率的关键环节。  因此，本项目拟结合大语言模型的文本理解能力与推荐系统的个性化建模技术，实现从“识别优惠信息”到“精准推荐优惠”的完整智能链路。  **二、项目目标与任务划分**   1. 构建一个能够从非结构化文本中提取优惠信息的系统，提取字段包括但不限于：优惠类型（满减、折扣、赠品等）、数值参数、适用商品、时间限制等。 2. 微调中文BERT模型以提升对中文优惠句子的理解能力，实现命名实体识别（NER）任务。 3. 构建基于用户行为（浏览、点击、购买等）和消费偏好的用户画像。 4. 设计并实现结合用户画像和实时上下文的个性化推荐算法。 5. 整合系统流程，展示从文本输入到推荐输出的完整路径，并完成可视化界面或PPT展示。    1. **关键技术与方法**   1.文本信息抽取：使用正则表达式与规则方法作为基础，结合HuggingFace的Transformers库微调BERT模型，进行优惠字段的命名实体识别（NER）。  2.数据处理：使用Python、pandas等工具对原始优惠文本、用户 行为数据进行清洗与结构化。  3.用户画像建模：根据用户行为日志提取特征，构建用户兴趣标签体系。  4.推荐算法：采用协同过滤（UserCF/ItemCF）与内容推荐相结合 的方法，推荐与用户兴趣匹配的优惠。  5.展示与集成：可使用Streamlit进行推荐系统演示界面构建，支持交互查询与推荐结果展示。  **四、项目亮点与应用前景**   * 利用大语言模型的语义理解能力，显著提升优惠信息提取准确率。 * 结合推荐系统技术，提升优惠触达效率与用户满意度。 * 可广泛应用于电商平台、线下门店APP、营销短信自动推荐等场景。 |
| **第二志愿** | **一、项目背景与研究意义**  在现实世界中，许多复杂系统可以抽象为图结构，例如：**社交网络**：用户节点、关系边；**生物分子结构**：原子为节点，化学键为边；**交通网络**：路口为节点，道路为边。  图数据因其**非欧几里得结构**和**高维、非结构化特征**，难以直接应用传统机器学习方法。对图数据进行有效的预处理和分类，是实现智能化分析（如社交影响预测、药物发现、路径优化）的基础。  因此，图数据的分类、特征提取与预测算法成为当前图学习领域的核心任务。  **二、研究目标与方向设想**  本课题旨在**选取一种代表性图分类算法方向**（如图模糊系统、图核、图嵌入或图神经网络），围绕其在图数据分类中的应用进行深入研究，包括：  了解该方法的基本原理与实现机制；  设计适合的数据预处理策略；  结合实际图数据集进行实验验证；  探索算法性能优化的潜力与适用边界。  **三、算法方向初步设想**  1. **基于模糊系统的图分类算法**；  利用模糊规则处理图全局特征，实现不确定性建模；  适合结构多样、标签模糊的图分类场景（如化学分子分类）。  2. **图核方法（Graph Kernel）**  原理：通过定义图之间的“相似度核函数”，将图映射到高维空间；  优点：无需训练深度网络；  缺点：计算复杂度高，扩展性有限。  3. **图嵌入（Graph Embedding）**  将图结构编码为低维向量，便于下游分类任务；  方法包括：DeepWalk、Node2Vec、Graph2Vec 等；  适合节点丰富、图量大的任务。  4. **图神经网络（Graph Neural Networks, GNN）**  利用消息传递机制对图结构建模；  方法包括：GCN, GAT, GIN 等；  应用广泛，性能优异，但训练复杂。  **四、数据集与工具**  1. 推荐数据集：   | 数据集 | 应用领域 | 特点 | | --- | --- | --- | | MUTAG | 化学分子 | 二分类，小图 | | PROTEINS | 生物结构 | 多分类 | | IMDB-B | 社交网络 | 图标签分类 | | Reddit-B / D&D | 大规模图 | 高维图特征 |   2. 工具与框架：  Python + PyTorch；  图学习库：**PyTorch Geometric (PyG)** / **DGL**；  可视化：TSNE、matplotlib；  模糊系统模拟：scikit-fuzzy / 自定义模块。  **五、预期研究内容与技术路线**  **文献综述与算法选择：**  阅读图分类相关经典及前沿论文，确定选题方向；  学习 Graph Fuzzy System 或 GNN/GK 等算法结构。  **算法实现与数据预处理：**  构建标准图分类数据预处理流程；  实现选定的图分类模型，设计特征提取流程。  **模型实验与评估分析：**  在多个数据集上进行对比实验；  使用精度、召回率、F1-score、ROC等评价指标。  **优化探索与应用分析：**  对模型结构或模糊规则进行简化或改进；  探索实际应用场景中的部署可能性（如药物预测）。  **六、预期成果**  一套完整的图数据分类与预处理流程；  对比实验与图分类模型实现代码；  提出至少一种改进策略并进行验证；  撰写论文/报告，总结方法、实验与应用前景。 |
| **第三志愿** |  |

1. **本项目团队优势**

*团队成员学生工作经历、科研经历、获奖情况、论文发表情况等*

|  |
| --- |
| 本项目团队由数学与统计学院具有多元背景的本科生组成，成员在学生工作、科研项目及竞赛中均展现出突出的组织能力与科研潜力，具体如下：  1.学生工作经历丰富，具备优秀的组织与沟通能力  团队成员曾担任统计2301班心理委员、数学与统计学院团委新媒体部部长、实践部部长、学生会文体部部长、纺织2305团支部书记等职务，具备较强的项目组织、协调与宣传推广能力，有助于保障项目高效运行与成果展示。  2.科研背景扎实，具备项目研发与论文撰写经验  四位成员参与2024年“树苗计划”等多项科研课题，如“多智能体编队的实时避障与自适应路径规划的协同理论研究”“高维模糊神经网络设计及其在图像识别的应用”等，具备良好的项目协作经验与算法理解能力。  其中范森林已参与撰写电场仿真与复合纳米纤维材料相关英文论文，具备初步的科研成果产出能力。  3.竞赛成绩优秀，具备实践能力与创新意识  团队成员徐文睿、叶琳婕曾获“正大杯”全国大学生市场调研大赛上海市二等奖，徐文睿同学获全国大学生英语竞赛国家级三等奖，范森林同学获东华大学高等数学竞赛二等奖，陈依林同学获东华大学奖学金等，体现出良好的学习能力、数据分析能力、语言表达能力和数学建模功底。  综上所述，团队成员在组织管理、科研能力与综合素质方面优势互补，能够胜任本项目中涉及的数据处理、语言模型微调与系统集成等多项挑战，确保项目在时间范围内高质量完成。 |

1. **项目开展计划**

*项目开展计划时间表、预计项目成果等*

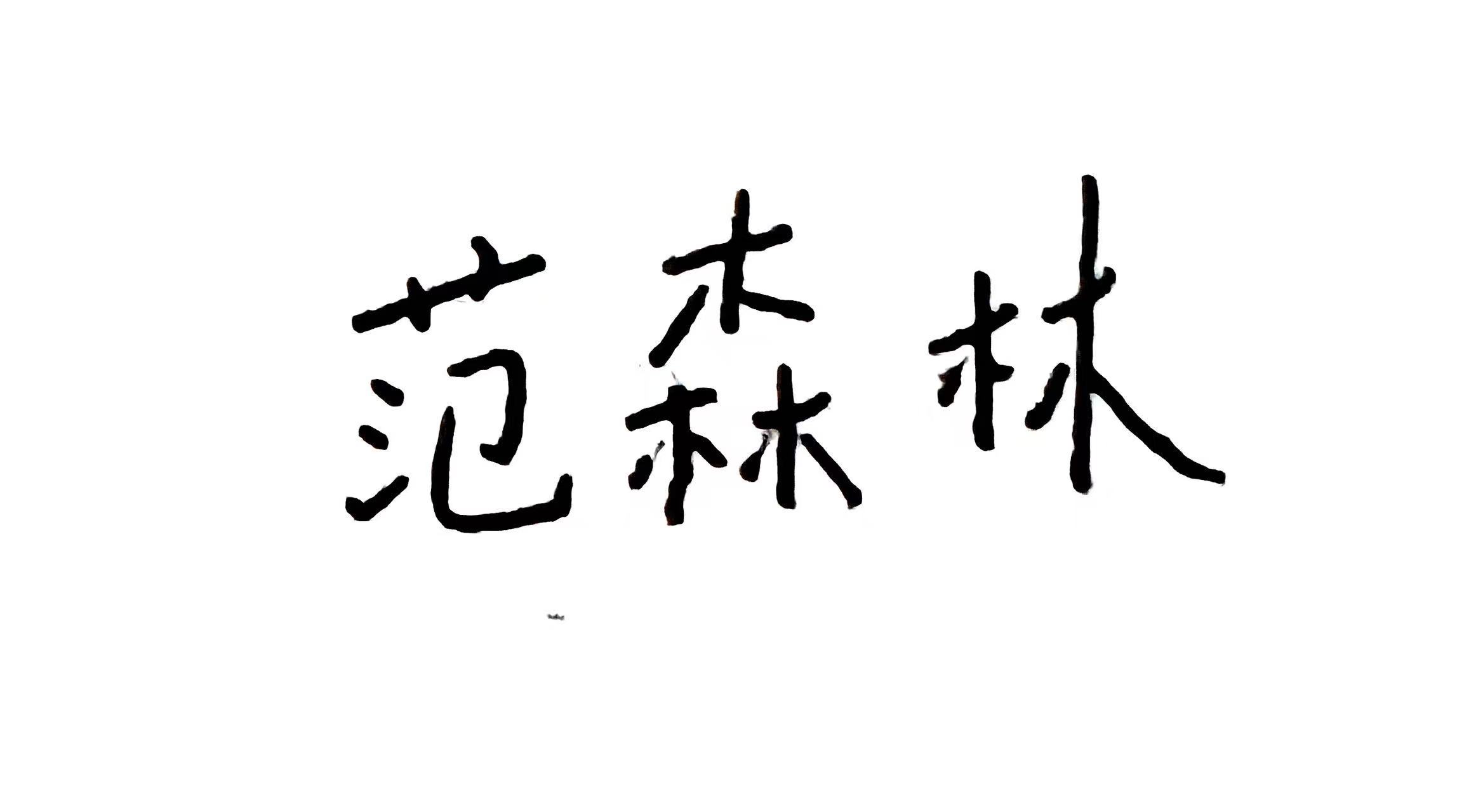
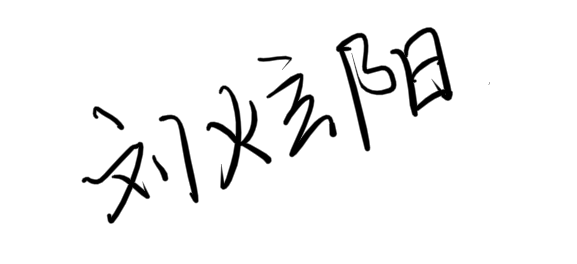
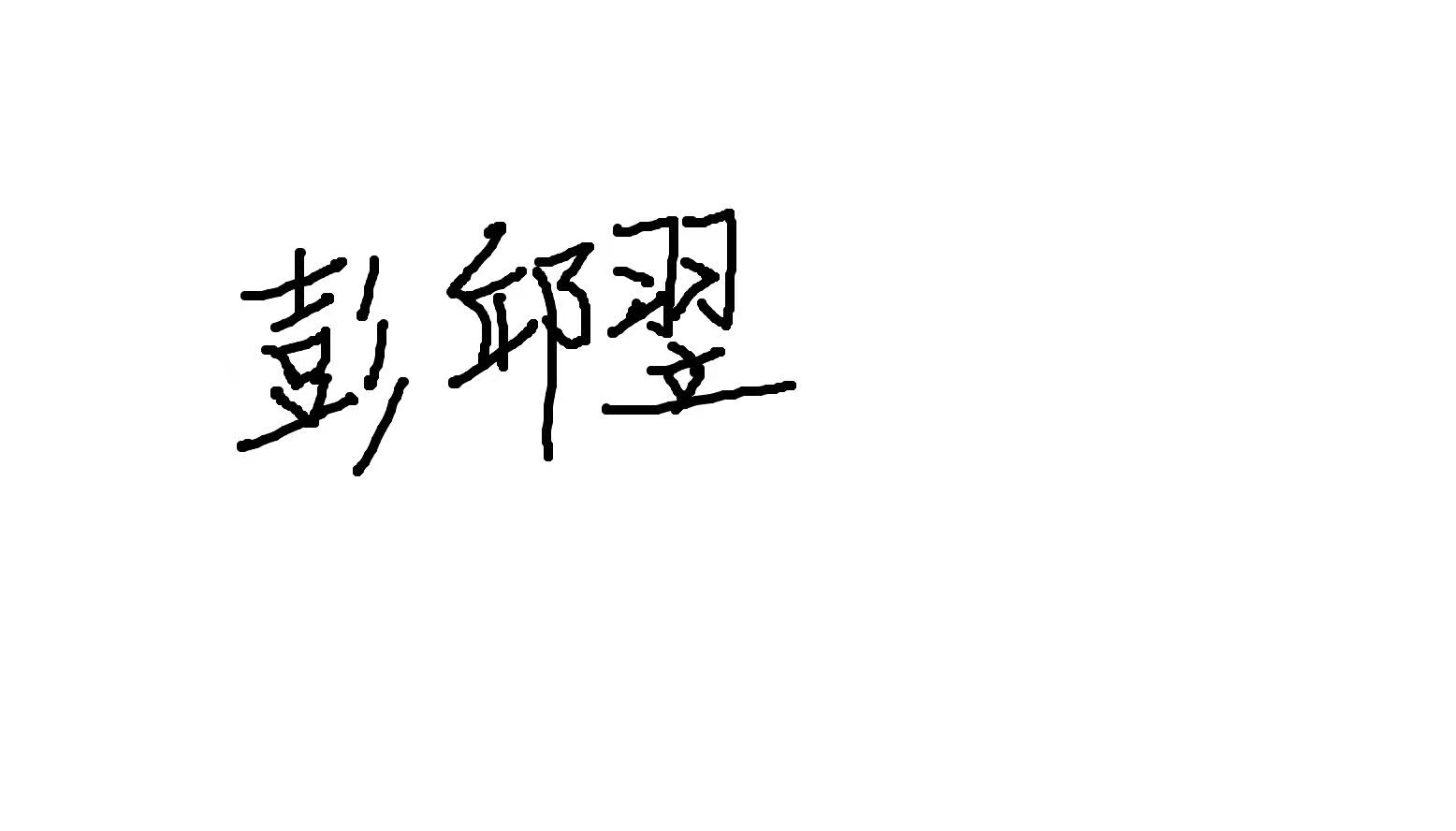
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **一、进度安排（可灵活调整）**   |  |  |  | | --- | --- | --- | | **月份** | **阶段名称** | **主要任务** | | **5月** | 立项与准备 | 明确项目目标与技术路线，组建团队并分工，立项答辩 | | **6-7月** | 数据准备  与预研 | 收集优惠文案样本，初步NER规则抽取尝试，学习BERT、Transformers库 | | **8月** | 模型微调 | 标注训练数据，使用Transformers微调BERT做NER，模型测试与评估 | | **9月** | 用户画像  与推荐开发 | 构造用户行为数据，构建用户画像标签体系，初步实现推荐策略（如协同过滤） | | **10月** | 推荐优化  与联调 | 调整推荐排序逻辑，联调NER输出与用户画像结果，开发个性化推荐流程 | | **11月** | 系统集成  与界面 | 使用Streamlit等工具搭建展示界面，集成抽取与推荐模块，进行组内测试 | | **12月** | 总结与交付 | 撰写结题报告、PPT展示材料，完成最终项目演示、结项答辩 |   **二、预期成果**   1. 优惠文案数据集及NER标注样本 2. 微调后的BERT模型及字段抽取系统 3. 用户画像构建与推荐模型实现 4. 系统原型界面或可视化报告 5. 项目文档（开题报告、结题报告、PPT） |

1. **承诺**

我们承诺：若成功入选，我们一定根据项目要求完成相关文献综述、论文、专利等，并参加国家级、上海市级大学生创新创业项目申报，且参加当年中国国际大学生创新大赛、“挑战杯”创新创业大赛等双创相关赛事。

**团队全体成员签字：（电子版请插入电子签名，纸质版请手签）**

**

**

*注：上表所填资料必须真实、完整、合法。*