

高考志愿查询预测与推荐系统

项目开发计划文档

项目名称：高考志愿查询预测与推荐系统

项目组：高考志愿查询预测与推荐系统开发组

项目成员：陈思翰（项目经理）、李响、薛安宇、胡周峰、田如召

2021 年 6 月 1 日

目 录

1. 引言.....	1
1.1 文档编写目的.....	1
1.2 术语与缩略语.....	1
2. 项目概述.....	1
2.1 项目背景.....	1
2.2 项目范围.....	2
2.3 交付的产品	2
3. 开发成员.....	2
4. 项目开发计划.....	2
4.1 需求确定阶段.....	2
4.2 数据爬取阶段.....	3
4.3 数据分析阶段.....	3
4.4 数据可视化阶段	3
5. 项目所需技能和培训计划	3
5.1 项目所需技能.....	3
5.2 项目培训计划.....	3
6. 开发计划.....	4
6.1 项目监控记录.....	4
6.2 风险管理计划.....	4
6.3 度量与分析计划	4
6.4 质量保证计划.....	4
6.5 配置管理计划.....	4
6.6 系统测试计划.....	4
7. 工作环境.....	5
7.1 开发环境.....	5
7.1.1 硬件设备.....	5
7.1.2 支持工具和软件环境	5
7.2 测试环境.....	5
7.2.1 硬件设备.....	5
7.2.2 支持工具和软件环境	5

文档信息	
标题	项目开发计划文档 Project Development Plan Document
作者	田如召
创建日期	2021.05.20
更新日期	2021.06.01
版本	V1.2
部门名称	高考志愿查询预测与推荐系统项目开发组-华中农业大学实训第五组

文档更新记录			
版本号	更新内容	操作日期	操作人员
1.0	初步拟定文档，并填入相关信息	2021.05.20	田如召
1.1	更新文档框架，使其视觉效果更加美观	2021.05.22	田如召、陈思翰
1.2	更新第 3 部分 开发成员	2021.06.01	陈思翰

项目开发计划文档

Project Development Plan Document

1. 引言

1.1 文档编写目的

制定本文档的目的，包括工作任务分解、估算工作产品和活动的规模、工作量、成本以及所需资源；制定开发计划、风险管理计划、质量保证计划、配置管理计划、度量计划、测试计划等，经评审、批准后作为项目实施和完成的依据，项目跟踪和监控的基础。

1.2 术语与缩略语

①PyCharm：是一种 Python IDE，带有一整套可以帮助用户在使用 Python 语言开发时提高其效率的工具；

②Requests：用Python语言编写，基于 urllib，采用 Apache2 Licensed 开源协议的 HTTP 库；

③MongoDB：一个基于分布式文件存储的数据库，由 C++语言编写。旨在为 WEB 应用提供可扩展性的高性能数据存储解决方案；

④Apache Spark：是专为大规模数据处理而设计的快速通用的计算引擎。

2. 项目概述

本项目为基于 spark 计算引擎的高考志愿查询预测与推荐系统，百万级数据经由框架存储、整合、清洗并上传前端界面实现数据可视化的目标，通过前端可视化给考生一个良好的志愿查询平台，根据考生分数以及位次给出一个合理的预测。

2.1 项目背景

高考一直以来都是我国传统应试教育中不可或缺的一部分。这场考试是对每个学生 12 年学习生涯的一次检验，是最后一次，也是最重要的一次测试。因此，无论是学生还是家长都对高考有着极高的关注度。本次开发的软件旨在基于当前网络中各式各样的高考数据网站，对各高校相关信息及其录取分数线等各项数据进行汇总，并将相关功能精简化，同时引入机器学习相关算法对数据进行预测和分类。达成以上目标后，用户就能轻便地达到数据查询的目标，并在合适的范围内参考本系统预测和推荐的结果。

2.2 项目概况

项目名称	高考志愿查询预测与推荐系统
任务提出者	华迪公司
用户	华迪公司
开发者	高考志愿查询预测与推荐系统开发组

2.3 项目范围

统计全国各高校和各专业的基本信息，以及各高校开设的各专业在全国各地区的历年最低录取分数和最低排名，随后基于这些信息对今年的各分数线与排名进行预测，并反馈给用户作为参考信息。

2.4 交付的产品

序号	产品名称（包括服务）	介质	交付日期	接收方	接受准则
1	高考志愿查询预测与推荐系统	前端网站	2021.06.05	华迪公司	

3. 开发成员

序号	角色	成员姓名	职责
1	项目经理、数据分析师	陈思翰	负责整个项目中大部分文档的撰写和核查，协调各个成员之间的工作。同时负责采用机器学习相关算法进行数据预测。
2	爬虫工程师	李响	负责整个项目的文档中，涉及爬虫部分的编写。负责编写爬虫程序获取数据。
3	数据库工程师	田如召	负责整个项目的文档中，涉及数据清洗和数据库部分的编写。负责进行数据清洗和数据存储。负责建立和维护数据库。
4	前端工程师	薛安宇	负责整个项目的文档中，涉及前后端接口部分的编写。负责建立前后端相关接口，将项目的各个部分结合起来。
5	前端工程师	胡周峰	负责整个项目的文档中，涉及数据可视化和前端架构部分的编写。负责前端部分的架构，并实现数据可视化功能。

4. 项目开发计划

4.1 需求确定阶段

对高考志愿查询预测与推荐系统进行需求分析，确定该项目需要实现的所有具体功能及其实现方法。随后对整个流程进行分工，撰写项目需求分析报告。

4.2 数据爬取阶段

根据上一阶段敲定的功能确定需要爬取的数据内容，编写爬虫程序，进入目标网站进行数据爬取。由于数据量非常大，数据爬取过程可能会在整个开发过程中持续。

4.3 数据分析阶段

将已爬取的数据进行分析。首先进行数据清洗，清除无关数据与各种符号，并将互相有关联的数据整合，得到若干表。随后采用机器学习相关算法对数据进行预测或分类。

4.4 数据可视化阶段

根据具体功能和 Web 页面的布局，将各数据、图像等展示在合适的位置，满足《项目需求分析文档》4.4 节中的界面需求，达到良好的用户交互。

5. 项目所需技能和培训计划

5.1 项目所需技能

管理技能：项目负责人对整体进度的规划调整，和对人力资源进行合理分配的技能；

技术技能：网站爬虫、前端后端实现、数据库交互、分布式集群计算等技能；

业务技能：团队成员相互沟通的能力，解决开发过程中遇到的各种问题的能力等。

5.2 项目培训计划

- ①Python 语言基础运用
- ②Requests、Scrapy、Selenium 等爬虫框架实例教学
- ③虚拟机分布式集群环境搭建教程
- ④前端框架构建教程
- ⑤MySQL、HDFS、MongoDB 等各数据库的原理与运用

6. 开发计划

6.1 项目监控记录

跟踪活动	日期	形式或工具	结果或记录
进度跟踪	05.21	人工检查	①爬虫部分几乎完成，部分细节需要调整； ②数据清洗部分完成约40%； ③机器学习部分因数据受限，线性回归无法实现，需要重做； ④数据库部分已搭建约60%； ⑤前端部分已有大致框架，因工作量大，进度可以适当放缓。

6.2 风险管理计划

- ①人员流动与请假风险：无流动可能，事假和病假每人每周控制在 8 小时；
- ②项目要求变更风险：采取敏捷开发模式，将产品功能模块化，高内聚低耦合，方便追加功能；
- ③技术风险计划：尽可能选用开发社区活跃、开发体系完整成熟的语言进行项目开发，并统一组内成员的开发环境

6.3 度量与分析计划

度量周期	度量任务描述	负责人	参与成员
每周	描述一周内工作执行情况与完成度	项目经理	全部小组成员
每周	各组员按照自己工作内容完成周报	各位成员	全部小组成员
系统测试完成后一天	测试人员生成《系统测试报告》进行反馈	项目经理	测试人员

6.4 质量保证计划

- ①每阶段任务结束前进行质量保证测试，根据测试结果优化软件产品；
- ②质量保障内容：可修改性、易用性、可移植性。

6.5 配置管理计划

- ①配置管理内容：源代码版本控制、开发环境配置、数据整合控制；
- ②配置管理方案：
 - 定义各类配置项入库、出库的准则和操作流程；
 - 定义基线变更的准则和操作流程；
 - 明确配置库的备份及维护方法，当出现异常后如何恢复的预案等；
 - 版本发布的准则、发布流程及发布计划，如测试版本、β 版本等；
- ③管理报告内容：基线库的基线项的清单、基线项的名称、版本、存放位置等。

6.6 系统测试计划

见测试计划文档

7. 工作环境

7.1 开发环境

7.1.1 硬件设备

本项目在 PC 环境下开发，要求各计算机配备足够的存储空间。

7.1.2 支持工具和软件环境

软件 / 工具	版本	获取方式	负责人	到位时间	费用预算（万元）	备注
IntelliJ IDEA	2018 版	破解版	无	/	0	学生资格认证
JetBrains PyCharm	2018	破解版	无	/	0	学生资格认证
CentOS7 虚拟机系统	7	国内镜像	无	/	0	
MongoDB 数据库	6.5	官方下载配置	无	/	0	

7.2 测试环境

7.2.1 硬件设备

个人计算机

7.2.2 支持工具和软件环境

软件 / 工具	版本	获取方式	负责人	到位时间	费用预算（万元）	备注
IntelliJ IDEA	2018 版	破解版	无	2/26	0	学生资格认证
JetBrains PyCharm	2018	破解版	无	2/27	0	学生资格认证
谷歌浏览器	72.0	官方下载	无	2/25	0	