

# 高考志愿查询预测与推荐系统

项目需求分析文档

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 2 日

# 目 录

1. 引言.....	1
1.1 文档编写目的.....	1
1.2 项目开发背景.....	1
1.3 项目开发意义.....	1
2. 项目概述.....	2
2.1 项目标识.....	2
2.2 数据来源.....	2
2.3 用户特点.....	2
2.4 开发时限.....	2
2.5 开发工具.....	2
2.6 编程语言.....	2
3. 开发配置.....	3
3.1 硬件需求.....	3
3.2 软件需求.....	3
3.3 开发环境.....	3
4. 系统需求.....	3
4.1 功能需求.....	3
4.1.1 高校信息展示.....	3
4.1.2 分数线查询.....	3
4.1.3 专业推荐.....	3
4.1.4 模拟填报志愿.....	3
4.2性能需求.....	4
4.3接口需求.....	4
4.3.1用户界面.....	4
4.3.2硬件接口.....	4
4.3.3软件接口.....	4
4.4界面需求.....	4
5. 项目开发计划.....	4
5.1 需求确定阶段.....	4
5.2 数据爬取阶段.....	4
5.3 数据分析阶段.....	5
5.4 数据可视化阶段.....	5
6. 项目开发人员及分工.....	5
6.1 陈思翰，项目经理，数据分析师.....	5
6.2 李响，爬虫工程师.....	5
6.3 田如召，数据库工程师.....	5
6.4 薛安宇，数据库工程师.....	5
6.5 胡周峰，前端工程师.....	5

文档信息	
标题	项目需求分析文档 Project Requirements Analysis Document
作者	陈思翰
创建日期	2021.05.17
更新日期	2021.06.02
版本	V1.5
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	敲定项目框架，确定项目需求	2021.05.18	陈思翰
1.1	敲定了需要实现的功能	2021.05.19	陈思翰
1.2	完善了引言、概述、开发背景等文字	2021.05.21	陈思翰
1.3	更新文档框架，使其视觉效果更加美观	2021.05.22	陈思翰
1.4	更新第 6 部分项目开发人员及分工	2021.06.01	陈思翰
1.5	更新 2.2 节 数据来源 更新第 3 部分 开发配置	2021.06.02	陈思翰

# 项目需求分析文档

## Project Requirements Analysis Document

### 1. 引言

#### 1.1 文档编写目的

本文档是“高考志愿查询预测与推荐系统”开发过程中的重要文献资料，是工程的起点、软件设计的依据、用户需求的真实反映、开发人员进行系统总体按需设计内容及具体实施开发工程提供必要的参考资料。通过本文档，应能明确软件需求、安排项目规划与进度、组织软件开发与测试，明确平台后期运维目标。预期读者是从事“高考志愿查询预测与推荐系统”开发的相关人员。

#### 1.2 项目开发背景

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。

#### 1.3 项目开发意义

本项目集爬虫、数据存储、数据预处理、数据清洗、数据分析、机器学习和数据展示于一体，涵盖 Python、JavaScript、PyCharm、Anaconda、Visual Studio Code、Spark、MongoDB、MySQL、H5C3、ECharts 等编程语言、软件 and 平台的安装和使用。这次项目开发旨在帮助大数据专业的大学生掌握和熟练运用课程理论知识和相关开发工具，并结合具体案例开发一个独立的大数据展示系统，以了解行业相关信息，为将来就业打好基础，同时锻炼学生的工作能力和团队协作意识。

## 2. 项目概述

### 2.1 项目标识

项目全名：高考志愿查询预测与推荐系统

### 2.2 数据来源

本项目的数据来源为中国教育在线网[www.gkcx.eol.cn](http://www.gkcx.eol.cn)，所有数据均从该网站中爬取。预计共得到三个数据集：

①专业总表.csv，该数据集收录了中国所有高校开设的共 1519 个专业的简单信息；

②专业明细总表.csv，该数据集收录了上述 1519 个专业的详细信息，包括所属专业门类、所属专业大类、专业层次、修读年限、专业简介等；

③高校总表.csv，该数据集收录了中国共 2827 所高校的简略信息，包括高校名称、所在地、建校时间、占地面积、官网网址、联系电话等；

④高校分数线.csv，该数据集收录了上述 2827 所高校在全国 31 个地区，从 2015 年至今每年的分数线；

⑤专业分数线.csv，该数据集收录了上述 2827 所高校的各专业在全国 31 个地区，从 2015 年至今每年的分数线。

注：具体字段和数据在后续可能会有进一步更改。

### 2.3 用户特点

- 1、高考应届生及其家长，或相关领域的老师、学生和相关工作人士；
- 2、渴望得到详尽的高校相关信息及其录取分数线等相关数据；
- 3、有基本的网络运用能力。

### 2.4 开发时限

2021 年 5 月 17 日至 2021 年 6 月 4 日共 19 天。

### 2.5 开发工具

PyCharm、Visual Studio Code、Spark、MongoDB、MySQL、ECharts

### 2.6 编程语言

Python、JavaScript、H5C3

## 3. 开发配置

### 3.1 硬件需求

本项目可以在单机或集群上完成，没有特殊的硬件配置需求。

### 3.2 软件需求

- ①Windows10 操作系统；
- ②基于分布式文件存储的数据库 MongoDB；
- ③Python 开发工具 PyCharm 和 Visual Studio Code；
- ④Spark；
- ⑤Flask。

### 3.3 开发环境

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。

## 4. 系统需求

### 4.1 功能需求

#### 4.1.1 高校信息展示

用户可以在页面中查看中国各地区开设的高校数量（精确到市级行政区），并能够查询到各高校的具体信息，包括高校名称、地址、建校时间、占地面积、官网网址、联系电话等。

#### 4.1.2 分数线查询

用户能够根据自己所在省份、高校类型、高校名称、专业名称等关键字查询到历年来该省的省控线、以及该专业在该省录取的最低分数线和最低排名。同时系统也会展示出当年省控线和分数线的预测结果供用户参考。

#### 4.1.3 专业推荐

用户可以在页面中查看中国所有高校开设的所有专业的相关信息，包括专业门类、专业大类、专业层次、专业简介、就业情况等。同时系统也会推荐部分热门专业。

#### 4.1.4 模拟填报志愿

用户需填入自己所在省份、文理科和高考分数。系统会根据用户填入的信息预测该省当年的高考批次线，并向用户推荐 30 所高校及每所高校对应的若干专业作为高考志愿。其中 10 所高校为“录取难度较大，可以尝试冲击”、另 10 所高校为“录取难度不大，基本可以录取”、后 10 所高校为“没有录取难度，可以作为保底”。

## 4.2 性能需求

本项目对数据处理能力无要求，数据查询响应速度快即可。

## 4.3 接口需求

### 4.3.1 用户界面

用户在 Web 界面上点击特定区域时能切换到对应的其他页面。

### 4.3.2 硬件接口

无。

### 4.3.3 软件接口

通过 Flask 连接数据库与前端。

## 4.4 界面需求

①输入设备：鼠标；

②输出设备：PC 显示器；

③分辨率：满足 Web 网页的分辨率；

④提示风格：弹出窗口、跳转其他页面；

⑤其他：页面风格简约美观，能让用户直观地查询和浏览各数据、便捷地使用各项功能。

# 5. 项目开发计划

## 5.1 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。

## 5.2 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。

### 5.3 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。

### 5.4 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 4.4 中的界面需求，达到良好的用户交互。

## 6. 项目开发人员及分工

### 6.1 陈思翰，项目经理，数据分析师

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。

负责采用机器学习相关算法进行数据预测。

### 6.2 李响，爬虫工程师

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。

### 6.3 田如召，数据库工程师

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。

### 6.4 薛安宇，数据库工程师

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，将项目的各个部分连接起来。

### 6.5 胡周峰，前端工程师

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。