

**目录**

[第一章 项目概述 1](#_Toc41424894)

[1.1项目背景 1](#_Toc41424895)

[1.2项目目标 1](#_Toc41424896)

[1.3平台特色 1](#_Toc41424897)

[1.4相关负责人 9](#_Toc41424898)

[第二章 解决思路 10](#_Toc41424899)

[2.1 问题描述 10](#_Toc41424900)

[2.2解决思路 11](#_Toc41424901)

[2.3相关负责人 13](#_Toc41424902)

[第三章 技术方案 14](#_Toc41424903)

[3.1 算法架构 14](#_Toc41424904)

[3.1.1 等级化算法 14](#_Toc41424905)

[3.1.2 模块分类数据降维 14](#_Toc41424906)

[3.1.3 加权无监督聚类算法 16](#_Toc41424907)

[3.1.4 智能推荐算法 18](#_Toc41424908)

[3.2系统技术架构 19](#_Toc41424909)

[3.2.2 Django架构的工程应用 20](#_Toc41424910)

[3.2.3数据交换共享接口设计 21](#_Toc41424911)

[3.2.4聚类模型设计接口 21](#_Toc41424912)

[3.3数据库方案 21](#_Toc41424913)

[3.4用户体验改善方案 22](#_Toc41424914)

[3.5系统使用技术与工具明细 23](#_Toc41424915)

[3.6相关负责人 23](#_Toc41424916)

[第四章 业务分析 25](#_Toc41424917)

[4.1服务对象 25](#_Toc41424918)

[4.2业务流程 25](#_Toc41424919)

[4.3业务分析 26](#_Toc41424920)

[4.3.1普通用户（金融信贷机构、企业） 26](#_Toc41424921)

[4.3.2数据管理员 34](#_Toc41424922)

[4.3.3系统管理员 36](#_Toc41424923)

[4.4 相关负责人 39](#_Toc41424924)

[第五章 组织结构 40](#_Toc41424925)

[5.1组织结构 40](#_Toc41424926)

[5.2角色与职责 40](#_Toc41424927)

[5.3协作与沟通 42](#_Toc41424928)

[5.3.1团队内部协作 42](#_Toc41424929)

[5.3.2项目接口人员 42](#_Toc41424930)

[5.4项目日程 43](#_Toc41424931)

[5.5相关负责人 44](#_Toc41424932)

[第六章 质量保证与测试计划 45](#_Toc41424933)

[6.1质量保证计划 45](#_Toc41424934)

[6.2单元测试计划 46](#_Toc41424935)

[6.2.1测试条件 46](#_Toc41424936)

[6.2.2测试方法 46](#_Toc41424937)

[6.2.3人员计划 46](#_Toc41424938)

[6.2.4专题测试 47](#_Toc41424939)

[6.2.5测试文档 49](#_Toc41424940)

[6.3系统测试计划 50](#_Toc41424941)

[6.3.1测试条件 50](#_Toc41424942)

[6.3.2测试方法 50](#_Toc41424943)

[6.3.3人员计划 50](#_Toc41424944)

[6.3.4专题测试计划 51](#_Toc41424945)

[6.4相关负责人 52](#_Toc41424946)

[第七章 商业计划 53](#_Toc41424947)

[7.1市场机遇 53](#_Toc41424948)

[7.2需求分析 54](#_Toc41424949)

[7.2.1市场环境 54](#_Toc41424950)

[7.2.2企业画像构建平台商业模式分析 56](#_Toc41424951)

[7.2.3竞争分析 59](#_Toc41424952)

[7.3管理体系 61](#_Toc41424953)

[7.4成本分析 61](#_Toc41424954)

[7.4.1人员成本 61](#_Toc41424955)

[7.4.2设备成本 62](#_Toc41424956)

[7.4.3营销推广成本 62](#_Toc41424957)

[7.4.4风险成本 62](#_Toc41424958)

[7.4.5学习成本 63](#_Toc41424959)

[7.4.6售后成本 63](#_Toc41424960)

[7.4.7其他成本 63](#_Toc41424961)

[7.4.8成本总计 63](#_Toc41424962)

[7.5成本控制体系 64](#_Toc41424963)

[7.5.1保盈利期 64](#_Toc41424964)

[7.5.2增收入期 64](#_Toc41424965)

[7.5.3精益化期 65](#_Toc41424966)

[7.6商业运行 65](#_Toc41424967)

[7.6.1客户划分 65](#_Toc41424968)

[7.6.2产品推广 65](#_Toc41424969)

[7.6.3营销模式 66](#_Toc41424970)

[7.7资源计划 66](#_Toc41424971)

[7.7.1软硬件资源计划 66](#_Toc41424972)

[7.7.2人力资源计划 66](#_Toc41424973)

[7.7.3培训计划 67](#_Toc41424974)

[7.8相关负责人 68](#_Toc41424975)

[第八章 可行性分析 69](#_Toc41424976)

[8.1技术可行性分析 69](#_Toc41424977)

[8.2运营可行性分析 69](#_Toc41424978)

[8.3组织可行性分析 69](#_Toc41424979)

[8.4市场可行性分析 69](#_Toc41424980)

[8.5投资可行性分析 70](#_Toc41424981)

[8.5.1经费预算 70](#_Toc41424982)

[8.5.2预期经济效益 70](#_Toc41424983)

[8.5.3相关参数计算 70](#_Toc41424984)

[8.6相关负责人 71](#_Toc41424985)

[第九章 风险计划 72](#_Toc41424986)

[9.1 项目风险分析 72](#_Toc41424987)

[9.1.1相关性风险 72](#_Toc41424988)

[9.1.2需求风险 72](#_Toc41424989)

[9.1.3市场风险 73](#_Toc41424990)

[9.1.4技术风险 73](#_Toc41424991)

[9.1.5开发风险 73](#_Toc41424992)

[9.1.6知识产权风险 74](#_Toc41424993)

[9.1.7管理风险 74](#_Toc41424994)

[9.1.8财务风险 74](#_Toc41424995)

[9.1.9风险因素评价 75](#_Toc41424996)

[9.2风险监控 76](#_Toc41424997)

[9.2.1管理体系 76](#_Toc41424998)

[9.2.2风险监控 77](#_Toc41424999)

[9.2.3规避措施 77](#_Toc41425000)

[9.3风险应对预案 78](#_Toc41425001)

[9.3.1技术风险预案 79](#_Toc41425002)

[9.3.2市场风险预案 80](#_Toc41425003)

[9.4退出机制 80](#_Toc41425004)

[9.4.1 IPO上市 80](#_Toc41425005)

[9.4.2兼并与收购 80](#_Toc41425006)

[9.4.3管理层收购 80](#_Toc41425007)

[9.4.4破产清算 81](#_Toc41425008)

[9.5相关负责人 81](#_Toc41425009)

[第十章 总结与展望 82](#_Toc41425010)

# 第一章 项目概述

## 1.1项目背景

金融科技为机器学习、大数据、人工智能等技术，它们的应用克服了传统金融信贷场景审核流程长、放贷慢的弊端，应用机器学习方法自动评估小微企业信用水平、企业还款能力等是金融科技在金融场景中的重要应用之一。

然而，金融科技场景中，企业这一信贷主体，数据量大且来源广泛，企业信息维度丰富，在分析企业还款能力、信用水平过程中，面临巨大的挑战。

* 如何对无标识的企业数据进行预处理、特征筛选、提取，形成有效的评估指标？
* 如何对提取的有效特征信息，进行无监督分类，对小微企业群体形成合适划分？
* 如何对分类的评估结果，构建企业画像和信用评分体系？
* 如何根据企业画像和信用评分，挖掘更多潜在的值得投资的企业？

## 1.2项目目标

在充分调研和评估企业特征信息的基础上，“e企查”公司尝试构建基于加权多层K-Means的企业分类系统“e企查”。

该平台旨在为企业提供精准的企业画像和信用评估，为企业生成明显的标签，并为金融机构信贷放款等投资企业的行为，提供客观准确的报告。同时根据无监督分类算法，挖掘潜在客群，精准获客，更好地解决“融资难、融资贵”的痛点问题，具有社会价值和商业价值。

## 1.3平台特色

* 算法特色

在构建企业画像的过程中，经过实验对比，选定K-Means作为无监督聚类模型，并在此基础上改进，提出了一种加权多层K-Means算法。其中加权多层K-Means中主要包含了等级化算法、模块分类降维、加权K-Means聚类。整体上核心指标领先于传统无监督聚类算法，并在最终分类打标签上，能够有很出色的表现。在此基础上，开发了相似企业推荐算法，助力拓展潜在客群。如图1-1所示，是整个算法的流程图，图1-2是算法数据流，图1-3是推荐算法流程图。

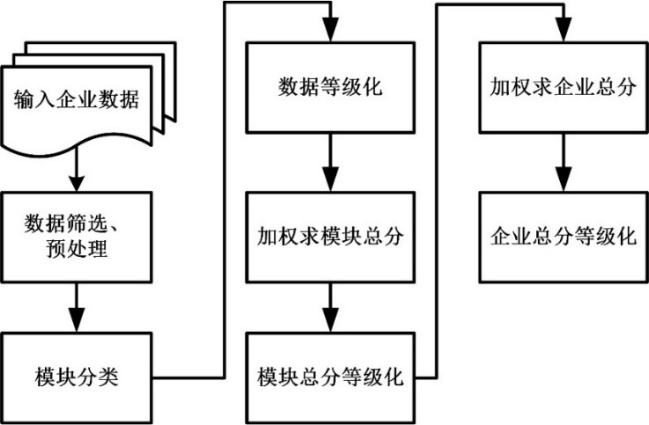


图1-1：算法流程图

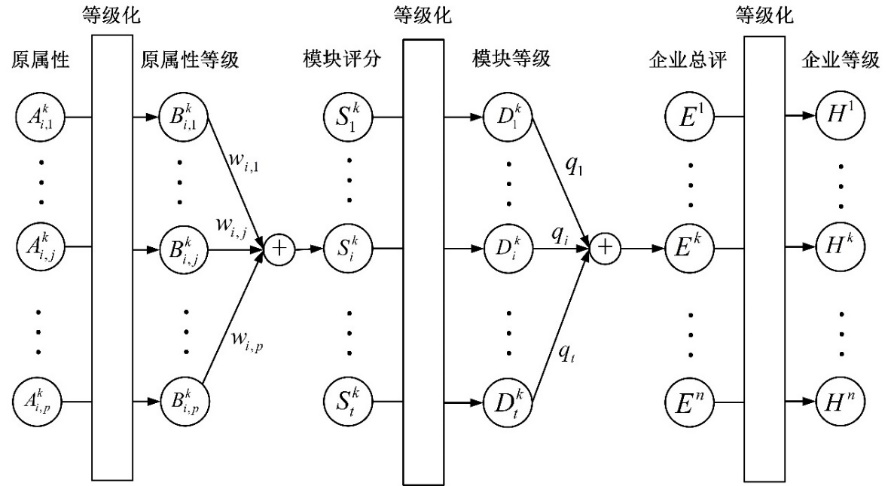


图1-2：算法数据流

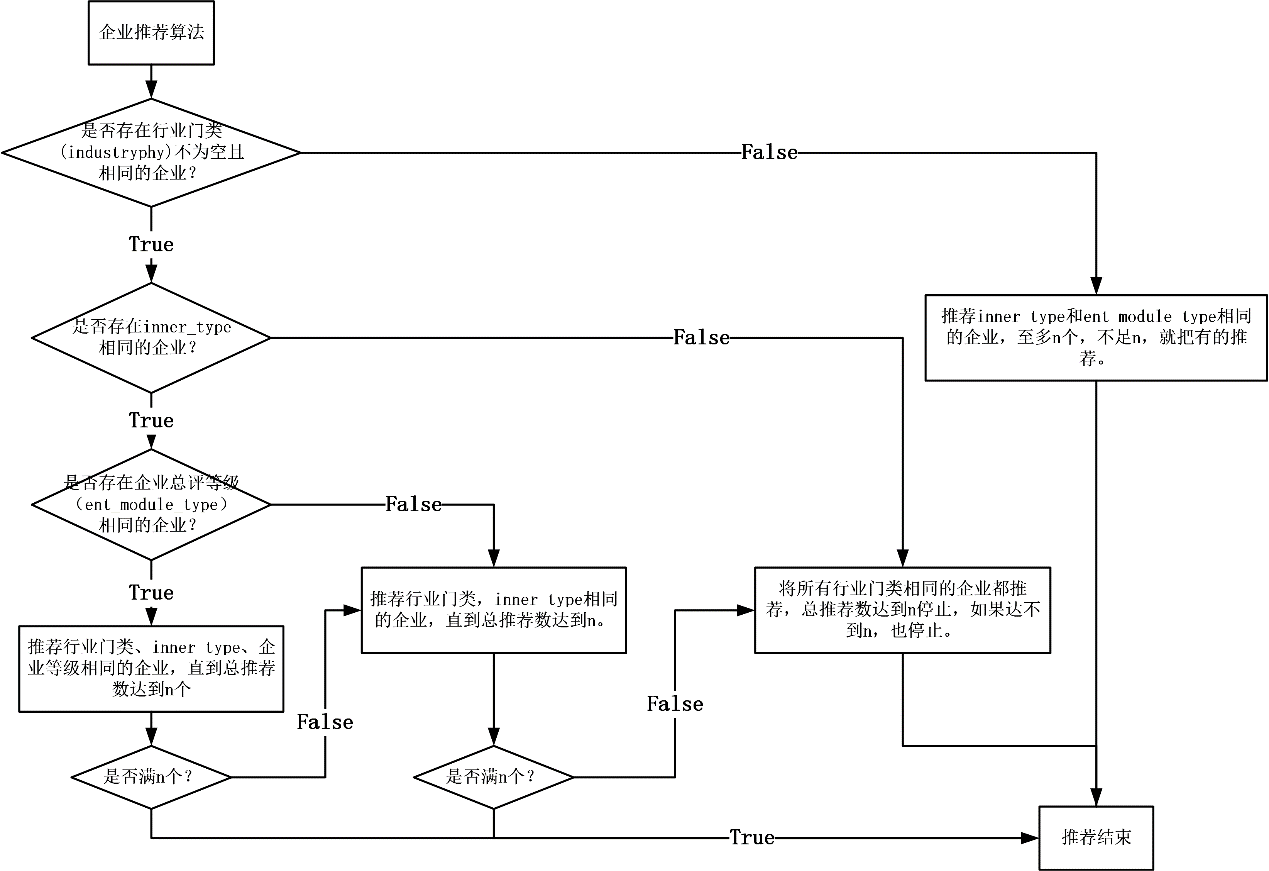


图1-3：智能推荐流程图

①算法在各项指标领先传统的无监督聚类模型，其中指标CH、SH越大表示聚类效果越优，DB越小表示聚类结果越优，测试数据是出题方提供的19w条企业记录，结果如表1-1所示，是无监督聚类模型在簇个数相同下得到的聚类结果：

表1-1：聚类指标比较

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 模型 | 加权多层  K-Means | K-Means  （K均值聚类） | Birch  （层次聚类） | GMM（高斯混合聚类） |
| CH | 136499.42 | 80648.91 | 28242.63 | 26089.34 |
| DB | 0.87 | 0.928 | 1.16 | 2.72 |
| SH | 0.56 | 0.55 | 0.42 | 0.26 |
| 训练时间(s) | 11.43 | 14.75 | 19.67 | 30.97 |
| 预测时间(s) | 0.23 | 0.25 | 0.26 | 0.29 |

②等级化算法：通过对一维属性使用K-Means，并根据簇中心进行排序，将数值划分成0-k个等级，建立统一度量衡的标准，效率高且鲁棒性强，可以实现等级随企业数据的动态变化。

③模块分类降维：将企业属性根据其特征，分入到不同的模块当中，并将模块内的属性综合加权，得到该模块的指标。该方法实现降维，解决数据缺失较大问题。

④加权K-Means聚类算法：对于聚类的建模指标赋予权重，调整不同建模指标对于最终聚类的影响程度。

⑤扩展性良好：用户可以根据自定义筛选条件，如行业信息或其他文字特征，结合聚簇分类结果，进行进一步分类

⑥智能推荐算法：综合根据行业门类、企业总分、加权K-Means聚类标签综合，综合起来推荐企业的潜在客群，如图1-3所示，相似企业推荐的网页界面，如图1-4所示。

* 系统特色

系统在技术实现上，采用Vue+Django，以restful接口的方式实现标准的前后端分离，保证系统运行稳定，调试和迭代方便。

①可视化功能丰富，便于直观反映企业数据特征、模型性能，如图1-4所示。

②单独测试界面，方便系统管理人员进行机器学习模型的训练和预测调试，如图1-5所示。

③查询平台（前台）和管理平台（后台）分离，提升使用者和管理者的使用体验。后台界面如1-5所示，前台界面如图1-6所示。

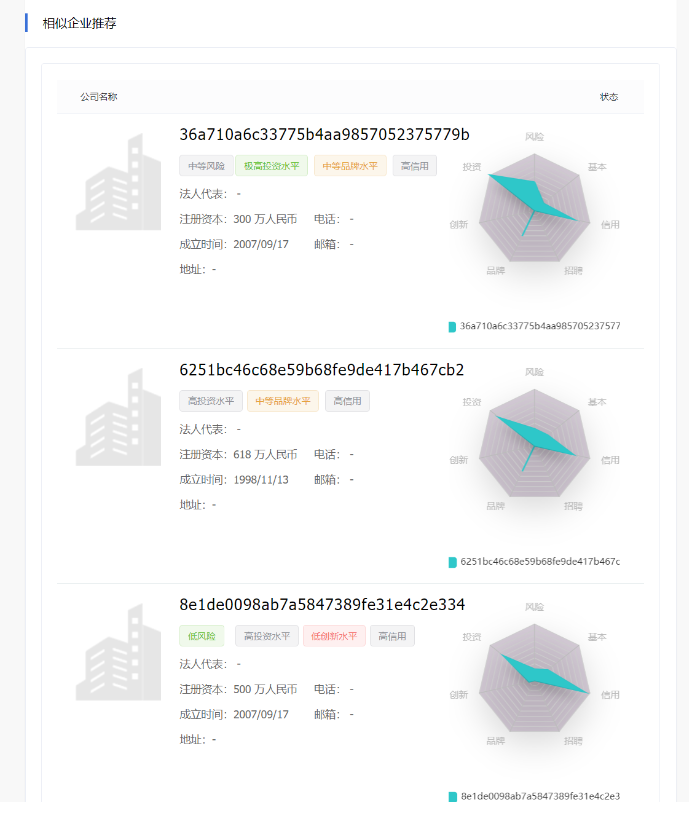


图1-4：企业雷达图可视化数据



图1-5：机器学习模型训练和预测界面



图1-6：前台查询界面

* 功能特色：

1. 将企业的多维特征划分成7个模块，对应风险、投资、创新、品牌、招聘、信誉、基本信息，使得用户能够非常直观的对企业有总体了解。
2. 评估手段多样。既可以在线同时选中多个企业，进行可视化对比，也可以批量导出企业评估报告，进行详细评估。

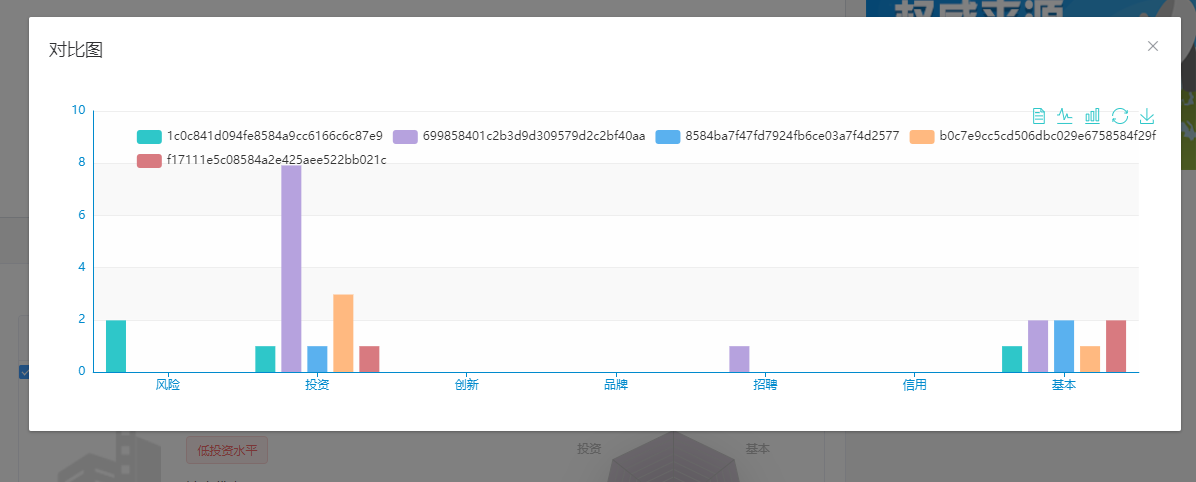


图1-7：企业维度信息对比

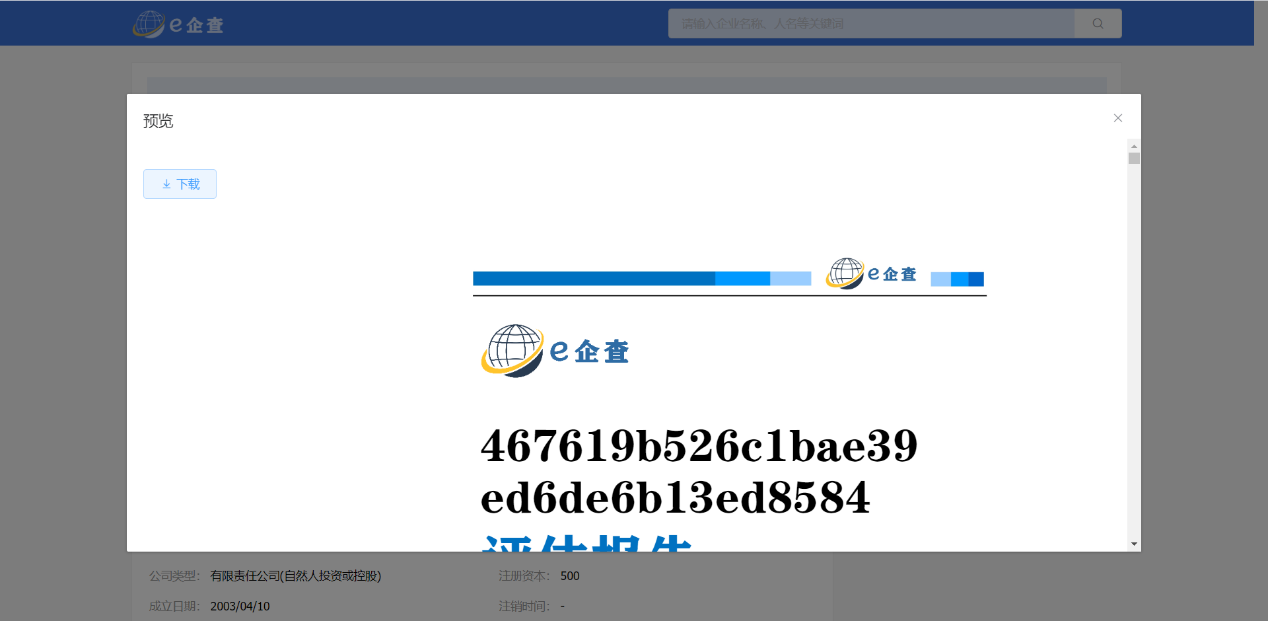


图1-8：企业评估报告

1. 偏好设置，用户可以根据自己的特定需求，过滤不感兴趣的数据。



图1-9：企业条件筛选

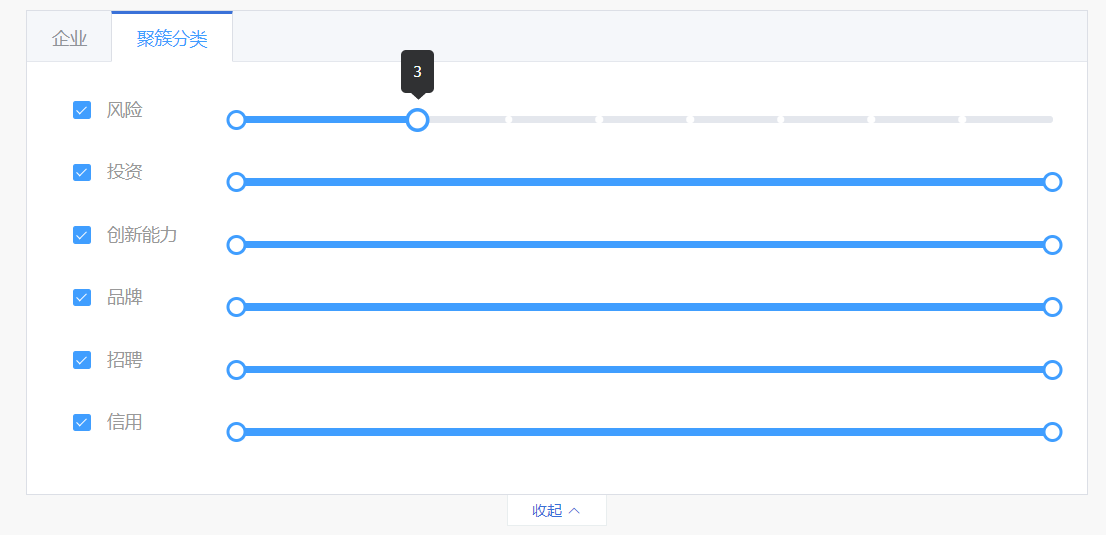


图1-10：模块指标筛选

* 工具特色：

1. 企业标签自动生成，导入数据后，自动生成企业评价，标签丰富。



图1-11：企业标签生成

1. 操作便利，支持多种批量化操作，比如批量上传数据、批量训练预测、批量查询等，这里以批量查询为例。详细网页页面，见第四章业务分析部分。



图1-12：批量查询

## 1.4相关负责人

组内人员项目概述分工如表1-2所示。

表1-2：项目概述分工表

|  |  |  |
| --- | --- | --- |
| 角色名称 | 角色 | 职责 |
| 小鲍 | 项目方案 | 确定项目整体方案和框架 |
| 小程 | 功能技术 | 协助确定项目解决思路、计划 |
| 小陈 | 算法与产品 | 文档整理与记录 |
| 小陆 | 项目经理助理 | 协助提炼项目亮点、需求 |
| 小阳 | 项目经理助理 | 协助梳理整体思路流程 |

# 第二章 解决思路

## 2.1 问题描述

目前市面上，评估企业基本信息的平台，如天眼查，尽管能提供全方面的企业风险、信誉、投资、背景等相关信息，并附上详细的企业报告，但是仍然存在以下一些问题：

**1、企业维度丰富，分析压力较大。**金融场景中企业这一信贷主体的数据覆盖互联网、政府、线上应用等来源的方方面面，数据量大，来源广泛、涉及企业的维度丰富。

**2、推荐模式单一，缺乏精准定制推荐。**市面上的企业信息平台，筛选模式只根据指定字段进行筛选，进入指定企业详细信息页面时，只支持查看当前企业的股东企业、投资企业、控制企业、竞品企业等。但对于金融信贷机构而言，最想要的是能够找到自己心仪的投资公司后，平台能够自动推荐与当前心仪公司特征相似的企业，进而拓展潜在业务。

**3、企业画像定位不清。**大多数企业信息平台，能够提供企业属性的信息，然而这些信息，只描述企业本身，无法评估该属性在整个行业、市场中的水平，很多描述，仅仅是对企业特征的一种定性，无法以一种定量的方式进行定位。此外平台，对企业进行的归类，都是根据本身企业提供的行业门类、规模等进行打标，忽视了企业多个属性之间的关联特性。

通过分析以上问题的原因，结合市场用户的需求，我们整理出的系统期望如表2-1所示：

表2-1：系统功能期望

|  |  |  |
| --- | --- | --- |
| 一级指标 | 二级指标 | 三级指标 |
| 界面 | 风格 | 整个系统各模块在风格上具有一致性 |
| 布局 | 页面简洁美观、操作人性化 |
| 导航 | 导航系统直观明确、简易便用，用于访问个模块  定位标记明确，标明在整个系统中的当前位置 |
| 运行环境与技术 | 基于Web | 获取在线生成企业评估文档和图表  保证数据随时更新 |
| 企业画像构建 | 算法指标 | 模型的训练和预测效率  无监督学习的CH、DB、SH(轮廓系数)等指标  生成标签是否显著 |
| 模型动态更新 | 根据企业数据变化更新模型  根据不同属性对企业评估的重要性进行权重更新 |
| 模块归类 | 将企业属性根据特征，分到7个模块下，即风险、投资、创新、品牌、招聘、信誉、背景，数据预处理实现数据降维。 |
| 智能推荐 | 从企业总评分和内部属性关联等多个角度综合评估，推荐与当前浏览企业和收藏企业的类似企业。 |
| 辅助分析 | 给出企业风险评估意见并在线生成文档图表  构建类似管理驾驶舱风格的系统，可以及时分析模型性能，企业数据整体分布形势等。 |
| 信息上传 | csv、excel等批量导入 |
| 系统功能 | 用户权限 | 权限设置，明确分工，便于平台维护 |
| 自动聚类 | 保证数据在导入时自动打标，并在数据变动和模型更新时，更新标签。 |
| 数据查询 | 实现批量查询导出，也会根据聚类算法推荐相似企业。 |
| 整体要求 | | 易于部署调试，前后端分离，便于二次开发，具有安全性。 |

## 2.2解决思路

基于智能企业分类推荐和直观评估企业行业水平的需求分析，提出信息一体化结局方案和思路，只在于快速构建能满足企业需求的无监督分类系统，为用户提供智能化、多样化、简洁化、人性化的全方位服务。其主要方案从系统结构和功能结构两部分考虑。

* **系统运作结构：**

如图2-1所示，本系统首先接收外部csv和sql等外部数据进入企业无监督分类系统，经过数据系统的清洗和预处理后，通过无监督学习模型，预测出对应的标签返回到数据系统中，进行分析，之后存入MySQL数据库中。

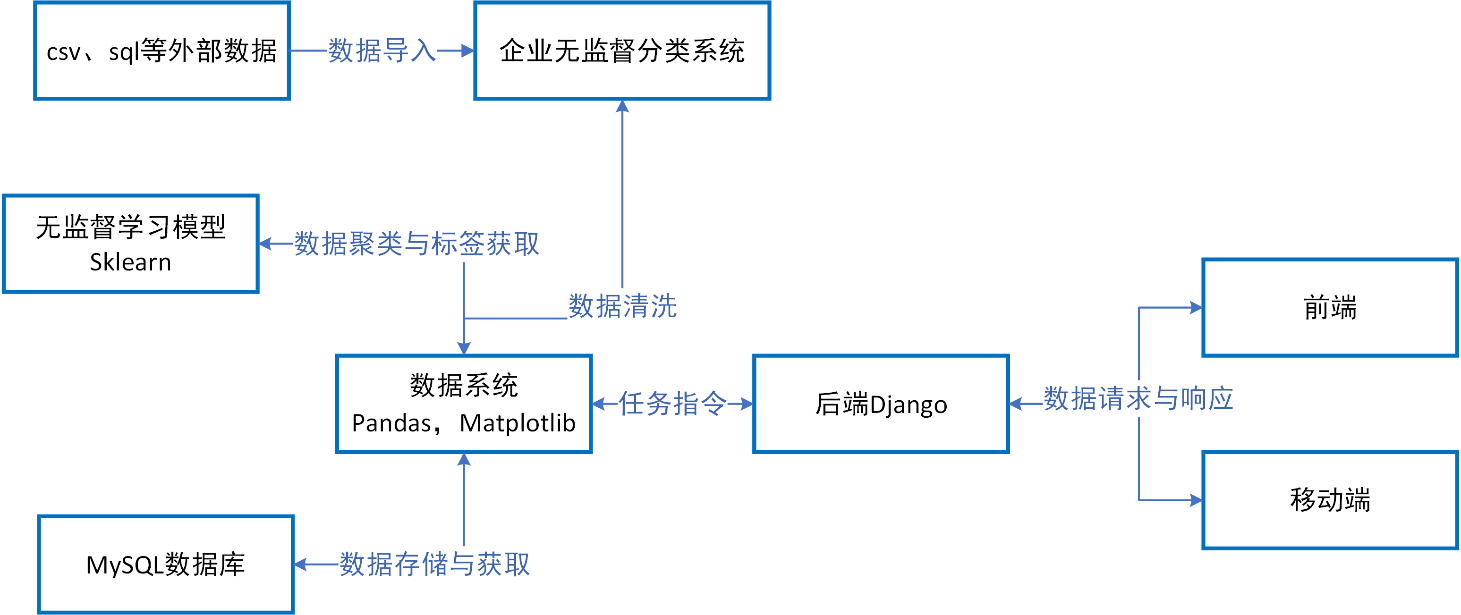


图2-1：系统运作结构图

系统后端采用Django架构，在接收前台传送的请求后，以restful风格的json将数据回传，结合ajax，提升系统响应性能。

* **功能结构：**

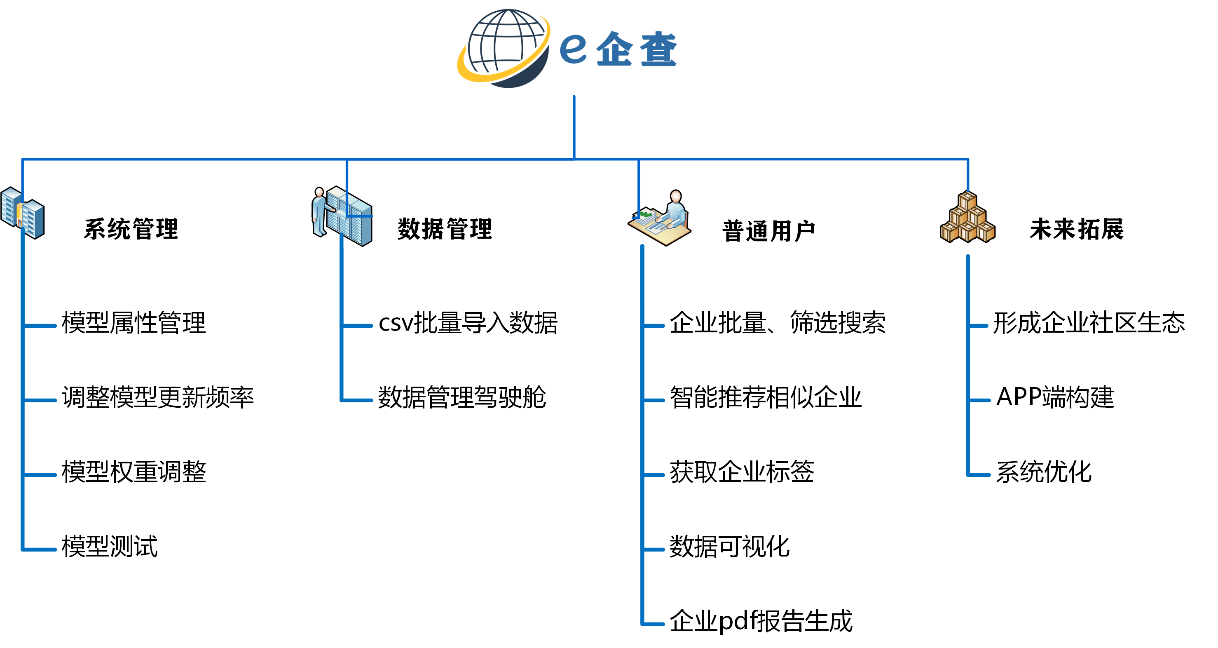


图2-2：系统功能结构图

如图2-2所示，阐述了系统每个模块所拥有的功能。

系统管理中，系统管理员根据每个模块对企业的最终评估影响不同，也可以调整每个属性的权重。

数据管理中，管理者可以对数据进行批量导入；系统内部拥有强大的数据可视化功能，可以轻易了解企业数据的分布趋势，以及每个标签内企业的各个维度属性的情况。

普通用户部分的功能属于前台功能，普通用户可以通过导入excel，csv的形式对企业进行批量查询，也可以通过聚类模型的聚类结果给出推荐企业，并获得当前企业与相似企业的企业标签。用户也可以通过可视化和pdf手段，分别以直观和详细的角度去评估一个企业。

本系统在未来，还打算完善平台的社区功能，并能够打通移动端，并在系统性能上进行优化，并在算法上，尝试使用更加高级的无监督聚类算法。

## 2.3相关负责人

解决方案的相关负责人列表如表2-2所示。

表2-2：相关负责人列表

|  |  |  |
| --- | --- | --- |
| 角色名称 | 角色 | 具体职责 |
| 小鲍 | 项目方案 | 分析问题、确定解决方案 |
| 小程 | 功能技术 | 协助分析问题，确定解决方案 |
| 小陈 | 模型算法 | 参与讨论，提出项目方案建议 |
| 小陆 | 界面UI | 参与讨论，提出解决方案建议 |
| 小阳 | 功能技术 | 协助项目经理分析问题、需求 |

# 第三章 技术方案

## 3.1 算法架构

由于主要涉及到无监督学习，不涉及损失函数和梯度计算来构建计算图，因此最终决定使用简易、高性能的sklearn作为机器学习的框架。

其技术亮点在于提出一种加权多层K-Means算法，解决企业数据量纲不一致、缺失严重的问题，并能够抽象出企业数据之间的内部关系，实现降维，并综合评估企业的水平，构建企业画像。在3.1.1至3.1.3中，概括了加权多层K-Means算法的思路。

### 3.1.1 等级化算法

该部分主要目的在于统一量纲，体现了加权多层的“K-Means”部分。

由于在企业画像构建过程中涉及到多个属性和中间属性，每个属性量纲不同。例如企业注册资本分布在0-20000万，而各类处罚，分布在0-20之间。需要通过预处理的方式，统一其最后对于企业画像构建的影响。

我们通过将某一属性的值，使用K-Means算法，分成k个簇。由于属性仅1维，可以个根据簇中心值，对标签进行重新排序分类。使得低的值，拥有较小的等级标签，较大的值，拥有较大的等级标签，使其最终分布在1-k的等级。对于那些空的或者为0的值，可以给其贴上0的标签，便于区分。

这种方法，相较于传统的归一化算法，有以下优点：

①可以避免极值对于整体分布造成的影响，具有更好的鲁棒性。

②同时也能使得等级的分布更加合理。

③更加利于贴上属性相应的标签。

详细过程可见算法部分的《企业画像实现技术》第三章方法第一节，等级化操作。

### 3.1.2 模块分类数据降维

不同的企业属性，可能是描述企业某一方面的具体特征，例如行政处罚、经营异常等，可以归入到风险模块当中；而智联招聘、中华英才、前程无忧招聘记录数等，可以归入到招聘模块当中。

根据所提供的属性表，本系统将企业模块分成风险、投资、知识产权、品牌、招聘、信用、基本信息7大模块。并将所提供的的企业数据归入到对应模块下。

**加权多层K-Means中，K-Means体现在等级化中，“加权多层”体现下计算过程：**

①对于每个模块中对应的属性，通过预处理的方式转化成可以建模的指标，再通过3.1.1中的等级化算法统一量纲，获得属性等级。即第一层，对应图3-1中的第一次等级化操作。

②再通过加权求和的方式，求出每个模块的评分。即第一次加权求和，对应图3-1中原属性等级等模块评分的部分。

③对于每个模块的评分，使用3.1.1中的等级化算法，求出每个模块评分的等级。即第二层，对应图3-1中模块评分到模块等级的部分。

④对于每个模块的等级，再次执行加权求和，获得企业的综合评分。即第二次加权求和，对应图3-1中的模块等级到企业总评的部分。

⑤对于最后企业的综合评分，再次使用等级化算法，获得企业综合等级。即第三层，对应图3-1中的第三次等级化操作。

整个算法的建模指标数据流如图3-1所示：

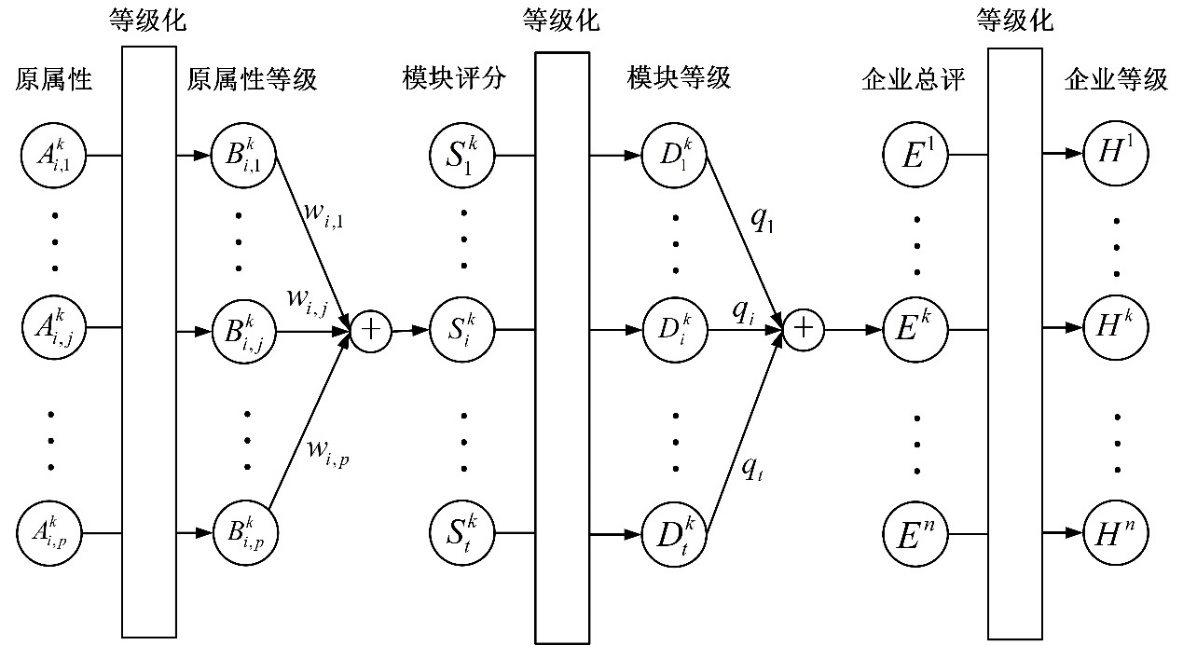


图3-1：建模指标数据流

如图3-1所示，代表企业的建模属性，表示企业编号，表示模块的编号，表示模块的第个属性编号，在后面的描述中同理。表示企业建模属性等级，即对应模块加权总分，对应模块等级，表示企业总分，即模块加权总分，表示企业等级。

等级化算法与模块分类的整合工作流程如图3-2所示：

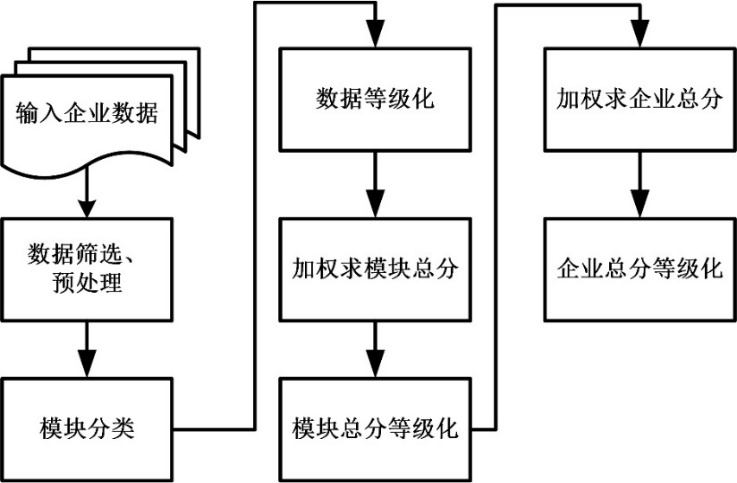


图3-2：等级化算法与模块分类降维整合工作流程

该方法的优点，相较于传统的降维方式，例如PCA，有以下优点：

①能够关注企业各个属性本身的内在关联，降维的角度更加合理。

②降维的过程中，能够关注到每个模块内属性的权重，方便算法设计者根据实际情况对模型进行调整。

③不同的企业，属性记录存在不同程度的缺失，通过这种降维压缩，加权求和的方式，减轻了这种数据缺失造成的影响，即便是一个企业只有1到2条记录，也能给出企业的属性评价。

④通过降维，能够给后续的加权K-Means算法，降低运算负担。

详细理论过程见算法部分，《企业画像实现技术》模块统计部分。

### 3.1.3 加权无监督聚类算法

该部分目的在于从多个无监督学习模型中，筛选出最适合这个体系的聚类模型，调节参数，并通过权重调整每个维度对最终聚类结果的影响。

聚类算法的目的，是希望能够找出企业属性间的内在关联，将相似的企业贴上相同的标签。然而模型的确定和簇个数需要由实验验证。

根据3.1.2中获得的7个模块的属性等级，赋予不同的权重。假设有n条记录，便可以通过对这个7×n的矩阵（其中n约为19万），对7个维度的属性先乘以权重使用加权聚类算法，将其分成指定个数个簇。

聚类模型，通过与K-Means（K均值聚类算法）， Birch（层次聚类模型）和GaussianMixture（概率聚类模型）对比实验，它们聚类的原理各自不同，但相同的是都需要预先设置簇个数。以CH（Calinski-Harabasz，值越大效果越好），DB（Davies-Bouldin，值越小效果越好），SH（Silhouette-Coefficient，值越大效果越好），即轮廓系数，作为评估指标。

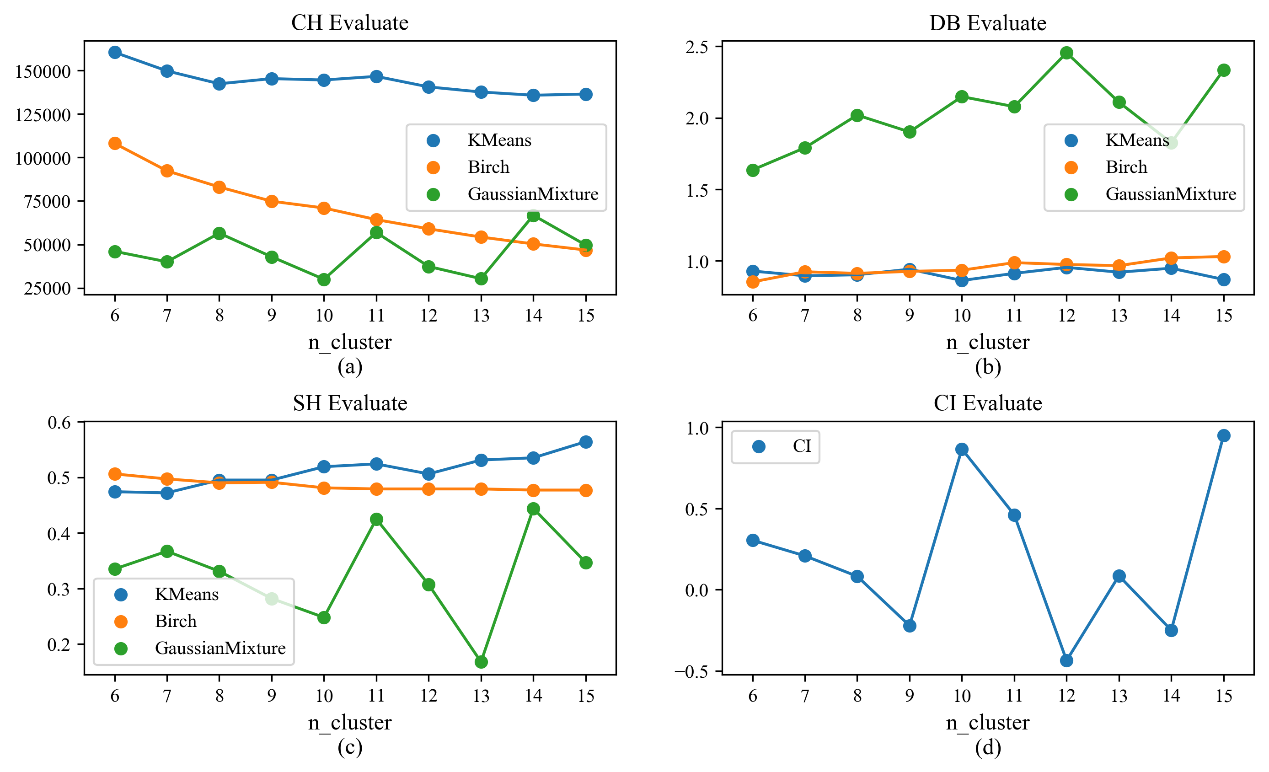


图3-3：K-Means，Birch，GaussianMixture及簇个数的评估.(a)三种聚类模型CH与簇个数关系性能的评估；（b）三种聚类模型DB与簇个数关系性能的评估；（c）三种模型SH与簇个数关系性能的评估；（d）K-Means与簇个数关系在自定义关系指数CI上的性能评估

图3-3中的四幅子图横坐标n\_cluster表示不同的簇个数。如图3-3（a）所示，显然K-Means曲线始终在另外两个模型之上，K-Means表现最优；如图3-3（b）DB分析中，DB越小越优，K-Means与Birch性能相当；如图3-3（c）中，K-Means总体曲线在另外两个模型之上。

因此综合考虑K-Means结合加权算法，即加权K-Means作为最终的聚类模型。

对于K-Means模型，最终簇的个数选择，综合考虑CH、DB、SH。由于三者值域并不同，因此对三者分别做归一化处理，得到CH’，DB’，SH’。根据单调性，CH、SH越大越优，DB越小越优，因此自定义指数CI如公式3-1所示：

 （3-1）

显然，由CH’，DB’，SH’的单调性可知，CI的值与聚类效果好坏呈单调递增关系。K-Means的CI值随簇个数的变化关系，如图3-3(d)所示，综合考虑到簇的个数与CI的值，选择n\_cluster=15作为企业总评聚类的簇个数。

系统未来簇的个数确定，也是采用这种方式来确定。通过训练与预测，给每一个企业赋予一个簇标签。详细部分见算法部分，《企业画像实现技术》实验部分。

经过3.1.1-3.1.3的企业画像指标构建和模型训练与预测，约19万条企业记录总的时间指标如表3-1所示：

表3-1：算法处理效率

|  |  |
| --- | --- |
| 事件 | 总时间（s） |
| 模型训练总时间 | 11.4 |
| 预测样本总时间 | 0.2 |
| 系统训练响应时间（预处理+训练） | 25.1 |
| 系统预测响应时间（预处理+预测） | 9.1 |

评估指标使用CH（Calinski-Harabasz）、DB（Davies-Bouldin）、SH（轮廓系数）。得到结果如下：CH = 136499.419；DB = 0.868；SH = 0.564。

### 3.1.4 智能推荐算法

实现聚类应用的具体场景主要在推荐。可以结合行业综合信息，加权K-Means聚类结果，以及企业等级，为用户当前查询企业，推荐相类似的企业，挖掘潜在客群。

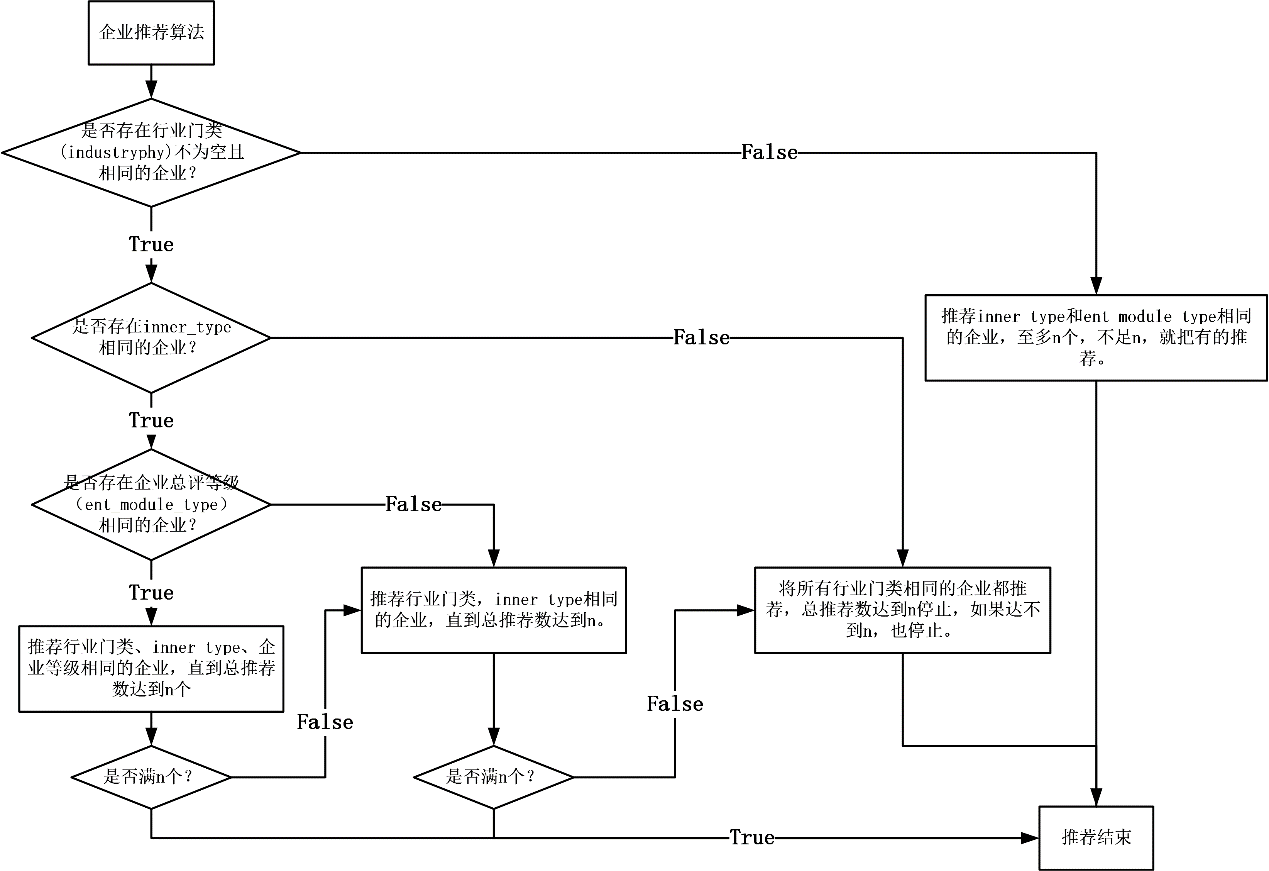


图3-4：企业推荐算法流程

如图3-4所示，3.1.3中通过最终聚类给企业贴上的簇标签为inner\_type，在图3-1中确定的总分等级，为ent\_module\_type，行业门类定义为industryphy。

该推荐算法流程中，优先考虑行业门类信息，结合inner\_type和企业综合等级结合推荐，确保推荐给用户的，都是与目标企业相似的企业。算法的伪代码实现如下：

* 步骤1：判断数据库中是否存在与当前企业行业门类(industryphy)相同的企业。如果存在，则跳转到步骤2；否则跳转到步骤5。
* 步骤2：推荐行业门类(industryphy)相同，聚类标签(inner\_type)相同，且企业等级(ent\_module

\_type)相同的企业，查询结果加入到查询集合中。如果查询集合中满n个企业，跳转到步骤6；否则，跳转到步骤3。

* 步骤3：推荐行业门类相同，聚类标签相同，企业等级不同的企业，查询结果加入到查询集合中。如果查询集合中满n个企业，跳转到步骤6；否则，跳转到步骤4。
* 步骤4：推荐行业门类不同，聚类标签不同的企业。查询结果加入到查询集合中。当查询集合满n个，或者行业门类相同企业推荐完毕，跳转到步骤6。
* 步骤5：推荐聚类标签和企业等级相同的企业。查询结果加入到查询集合中。当查询集合满n个，或者聚类标签和企业等级相同的企业推荐完毕，跳转至步骤6。
* 步骤6：推荐结束。

## 3.2系统技术架构

综合考虑企业维度特征的丰富性以及不断增加的数据量，以及使用平台良好的交互性，以及与后端机器学习、数据分析库的无缝适配，决定基于B/S模式，采用Django后端架构，选用MySQL数据库实现对数据的高效管理。本章将主要介绍用户体验、安全管理、数据库等方案。如图3-5所示，前端采用Element-UI提升UI体验，Vue和Ajax与后端交互信息，减少服务器负担。使用Echarts更好的支持数据可视化模块。

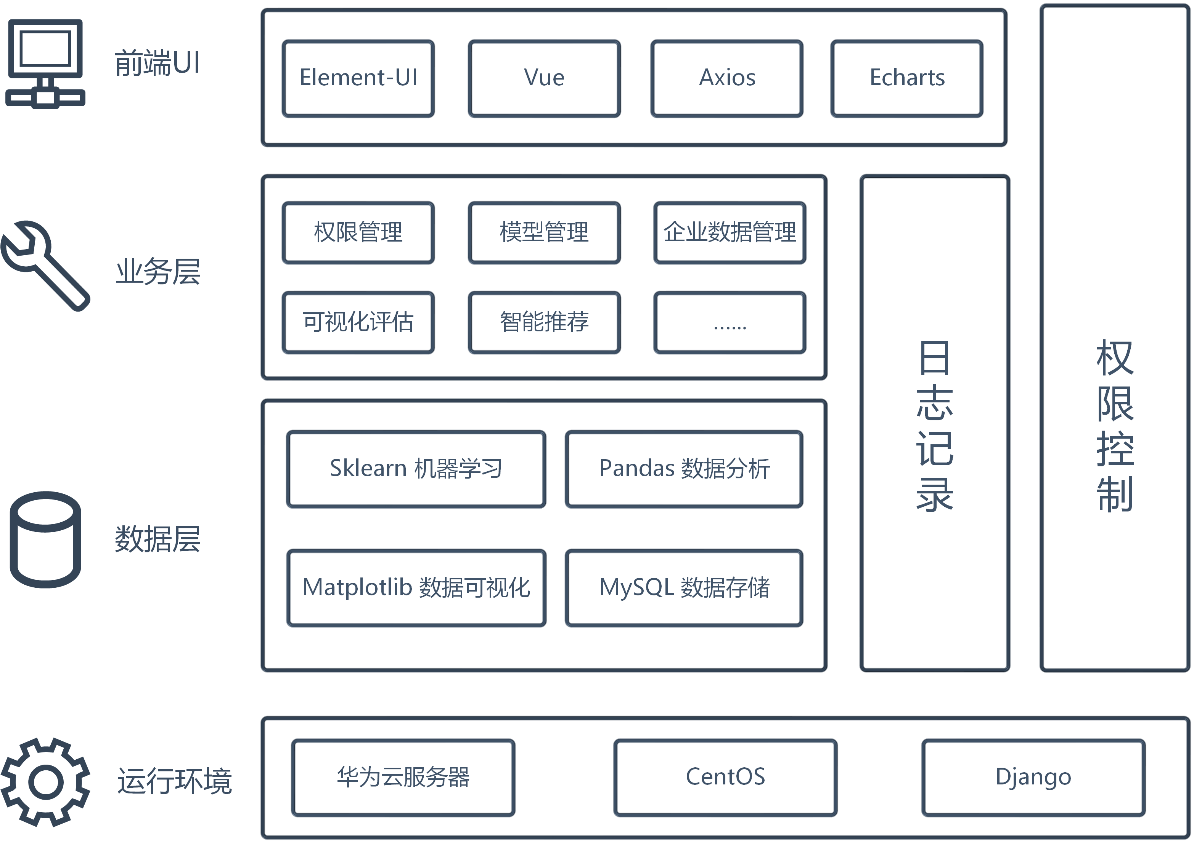


图3-5：系统技术架构图

### 3.2.2 Django架构的工程应用

采用先进的Django技术架构，基于MTV三层架构，即M（Model，模型层），T（Template，模板层），V（View，视图层）。

控制器即最后用户使用的浏览器，用于向服务器发起请求，接收信息。它不是MTV三层结构中的一部分，不在分析中列出。

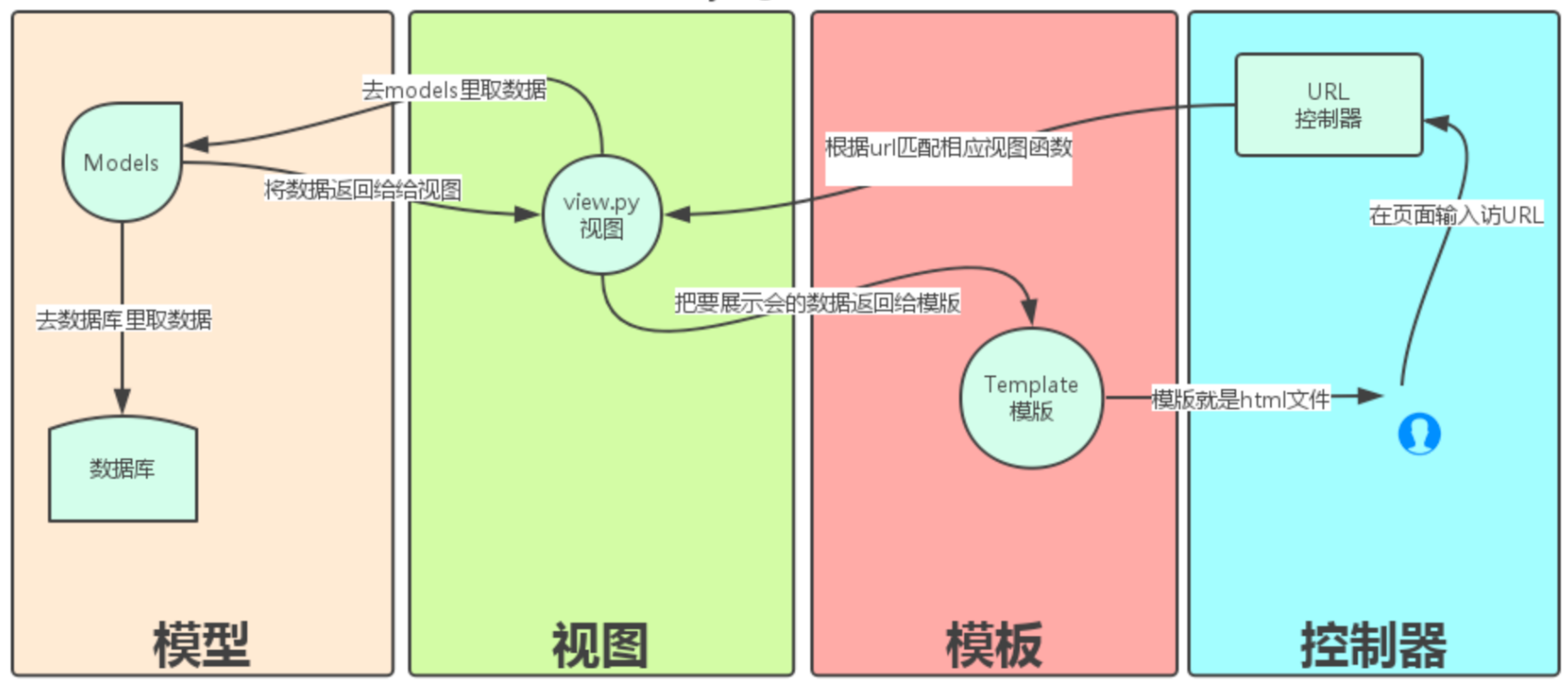


图3-6：MTV三层架构数据流向

如图3-6，展示了MTV三层架构的数据流向。

而Django作为一种后端架构，相比传统的Java系的框架，显得更加轻量易于管理。同时，由于机器学习采用的是Python的Sklearn库，能够更好的实现无缝衔接，易于数据交互和模型管理。

### 3.2.3数据交换共享接口设计

本系统采用前后端分离设计，故在数据的交互上，采用restful风格。restful是一种网络应用程序的设计风格和开发方式，基于http，可以使用xml格式定义或json格式定义。json具有易于读写、易于解析、占用带宽少和传输速度快等优点，所以本系统采用json结合restful接口设计规范进行前后端数据交互。

系统采用数据库为MySQL，为了确保数据库的操作效率，对于数据库的设计实现结合存储过程、触发器和Django中自带的ORM与数据库服务器进行交互，并将最后获取的数据用json封装，交付给前端。

### 3.2.4聚类模型设计接口

为方便后端人员的调用，将聚类函数的所有接口压缩到只有训练和预测两个模块。在数据的传递上，统一使用Numpy数组或矩阵，支持快速的批量训练和预测，确保效率。并在模型重新训练完成后，生成模型训练报告，及时让开发者了解聚类的结果。

## 3.3数据库方案

由于金融科技场景中，企业这一信贷主体，数据量大且来源广泛，企业信息维度丰富，为了便于企业各项属性的管理，以及统一每一种模块的机器学习训练场景，决定使用MySQL作为数据库存储。其主要优点有：

稳定性：坚持ACID原则，可以确保在高并发的环境下，事物的稳定性、安全性和可预测性。

灵活性：支持join等事物操作，满足一系列复杂查询的要求。

高效性：可以通过设计索引的方式，大大加速查询速度，提高系统的响应性能。

实现海量的数据库管理，数据库的设计优劣非常重要，在数据库设计时，遵循以下原则：

1、一致性：统一分析和设计，协调各个数据源之间的关系，对于相关联的数据库确保数据的同步更新。

2、完整性：确保数据的正确性和兼容性，相关联的数据库之间要通过外键来约束，还要有严格的审核和程序约束机制。

3、安全性：数据保护，防止数据泄密、篡改或破坏，加强数据库的权限控制，加以认证和授权机制。

4、扩展性：充分考虑到系统的升级和移植需要，确保良好的扩展性和可伸缩性。

5、规范化：确保数据库能够达到3NF规范，明确表之间的关系，尽可能的提升数据库的性能，减少资源浪费。

6、适当冗余：对于一些需要频繁关联来查询和操作的表项和属性，可以采用适当冗余，避免在外键关联时造成时间的浪费。

7、索引设计：根据实际应用场景，设计对应的索引，提升系统的响应性能。

## 3.4用户体验改善方案

由于用户角色分为普通的用户和管理员，因此根据使用习惯决定将查看页面和管理页面，做成两种布局风格。

由于用户的计算机水平可能参差不齐，良好的用户体验能帮助其快速适应系统，通过终端操作与服务器交互，获得自己目标企业的信息，和推荐企业的信息。由于使用本平台的用户，可能大多使用过类似天眼查、企查查等商业查询平台，因此在前端布局方面，一定程度上参考了它们的风格，能够更加亲切。

而后端的企业数据和模型数据管理平台，风格更加偏向传统的后台管理系统，方便操作人员更加快速上手。

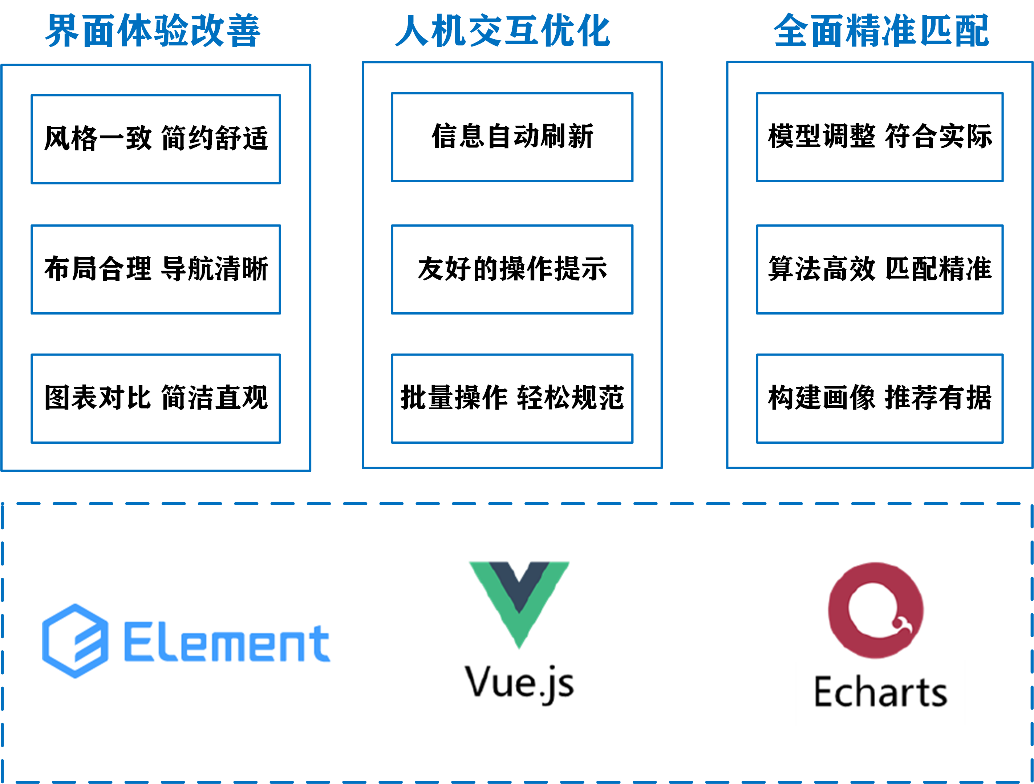


图3-7：用户体验改善

如图3-7所示，列出了前端在用户界面体验改善、人机交互优化、全面精准匹配所做的具体事项，以及使用到的技术。

主要思路与技术实现：

1、提供风格统一的简洁页面，参考用户的传统使用习惯，更好上手与操作。

2、统一css样式风格，易于页面渲染与维护。

3、使用js脚本确保页面交互，并使用Axio技术与后端服务器进行交互，减少数据加载负荷。

4、使用echarts等第三方库，将数据呈现更加科学和简洁。

5、前端只负责简单的交互和验证逻辑，后端负责复杂的运算操作。

## 3.5系统使用技术与工具明细

表3-2展示了系统使用的技术明细。

表3-2：系统使用技术与工具明细表

|  |  |  |
| --- | --- | --- |
| 名称 | 版本号 | 备注 |
| Django | 2.2.3 | 后端架构 |
| Python | 3.7.0 | 模型训练、数据分析、后端语言 |
| Sklearn | 0.0 | 机器学习框架 |
| Vue | Vue2.0 | JS前端框架 |
| MySQL | MySQL 5.1 | 数据库 |
| Pycharm | Pycharm 2017 | Python集成开发工具 |
| vscode | 1.45.1 | 前端开发工具 |
| Visio | Visio2019 | 流程图工具 |

## 3.6相关负责人

表3-3展示了系统技术责任的分配。

表3-3：项目技术责任分配表

|  |  |  |
| --- | --- | --- |
| 角色名称 | 角色 | 职责 |
| 小鲍 | 负责人 | 负责技术方案、标准制定 |
| 小程 | 参与人员 | 审核技术方案制定，制定技术标准 |
| 小陈 | 参与人员 | 技术方案文档撰写 |
| 小陆 | 参与人员 | 协助前端中可视化方面的标准制定 |
| 小阳 | 参与人员 | 汇总和整理前后端、算法思路，协助技术文档撰写 |

# 第四章 业务分析

## 4.1服务对象

“e企查”，作为基于无监督学习的企业分类系统，一方面致力于帮助企业数据的管理者，降低人力成本，做到有效地利用企业多个维度的特征信息，准确地对企业进行打标归类；另一方面，帮助金融信贷机构，评估企业综合水平、还贷能力，挖掘潜在业务对象，同时对于普通企业，也可以更好地找准自身行业定位。

表4-1：平台服务对象及内容

|  |  |
| --- | --- |
| 服务对象 | 主要服务内容 |
| 金融信贷机构、企业  （主要对象） | 1、搜索业务。提供目标企业的详细信息，以及经过评估的指标，簇标签。查询手段可以是精确、模糊、批量查询。  2、智能推荐。根据企业的当前查询企业等，依据模型对企业打上标签，依据综合评分、行业门类、模块属性等，按照匹配度由高到低推荐类似企业，挖掘潜在业务对象。  3、提供目标企业的评估文档与图表，对企业每个模块以及综合实力，分析优势和缺陷。  4、支持选中多个目标企业，生成对比图表，更加直观的比较不同企业间的优劣势，更好评估企业信贷能力。 |
| 企业数据管理平台 | 1、管理企业数据。其中支持以csv格式批量上传企业数据。  2、企业数据驾驶管理仓，丰富可视化体验。  3、提供单独测试界面，测试模型训练预测各种指标。  4、调整模型权重、更新模型等。 |

## 4.2业务流程

如图4-1所示是系统管理员业务流程。

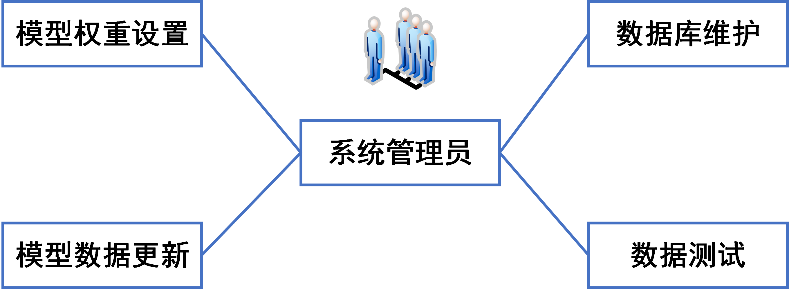


图4-1：系统管理员业务流程、

如图4-2所示是数据管理员业务流程。



图4-2：数据管理员业务流程

如图4-3所示，是普通用户的业务流程。

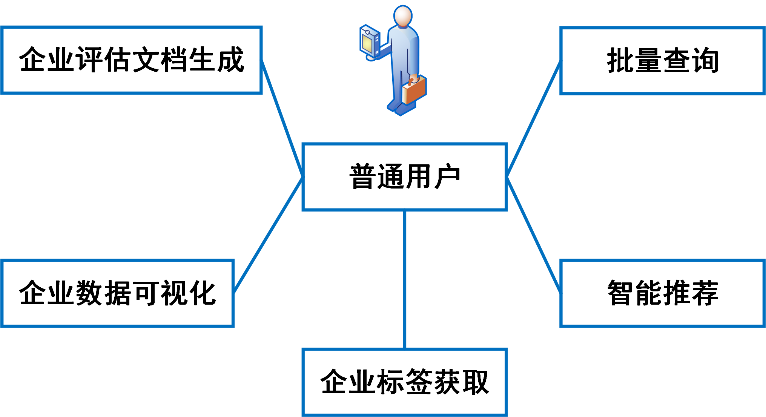


图4-3：普通用户业务流程

## 4.3业务分析

本系统分为前后台两个部分，前台主要用户查询、分析服务，后台用户系统、数据管理人员对数据、模型进行管理。

### 4.3.1普通用户（金融信贷机构、企业）

* **搜索业务**

用户在注册完登录后，就可以对自己感兴趣的目标企业进行搜索。用户首先可以根据自己想要查找的目标企业，或者是按照指定的标签范围进行查找，也可以通过csv导入的方式，批量返回查询结果。

如图4-4所示，是前台的首页。用户名可以通过输入企业名全称或部分的方式精确查询企业，见图4-5；也可以选择上传csv，批量对企业进行查询，见图4-6；在底下还有当前热搜企业。



图4-4：查询首页



图4-5：输入企业名查询企业



图4-6：批量查询文件上传界面

在精确查询或者批量查询文件上传后，跳转至查询结果界面。查询结果界面如图4-7所示，以输入“8564”为例，显示了所有企业命中含该字段的企业。这里粗略展示了企业的基本信息，企业标签，各个模块的评分，以雷达图的形式呈现，在上方，还可以通过筛选的方式，选择自己心仪的企业，如图4-8所示。



图4-7：查询结果界面

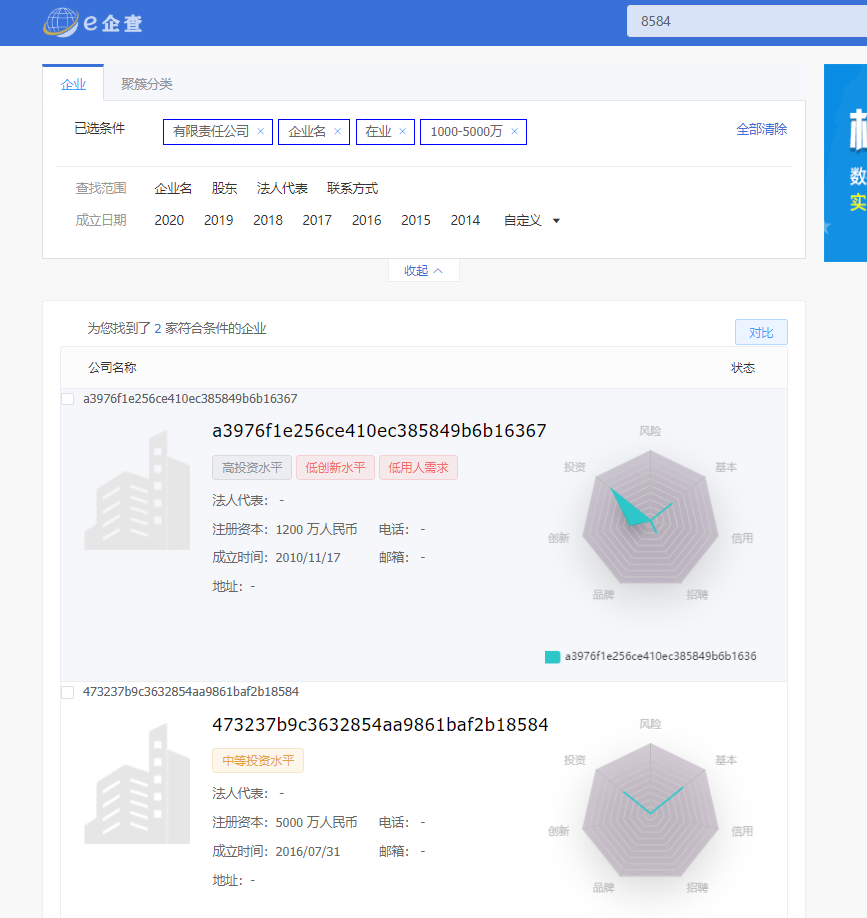


图4-8：设置条件筛选心仪企业

* **数据可视化**

为了更好的评估不同企业之间的差异，用户可以同时选中多个自己感兴趣的企业，可以生成对比图，更加直观的比较不同企业间的差异。每个企业都有对应的雷达图，直观反映各个维度的水平。

查询返回的企业，可以通过生成柱状图的方式，比较每个企业在每个维度上的评分的特征对比图，如图4-9所示：

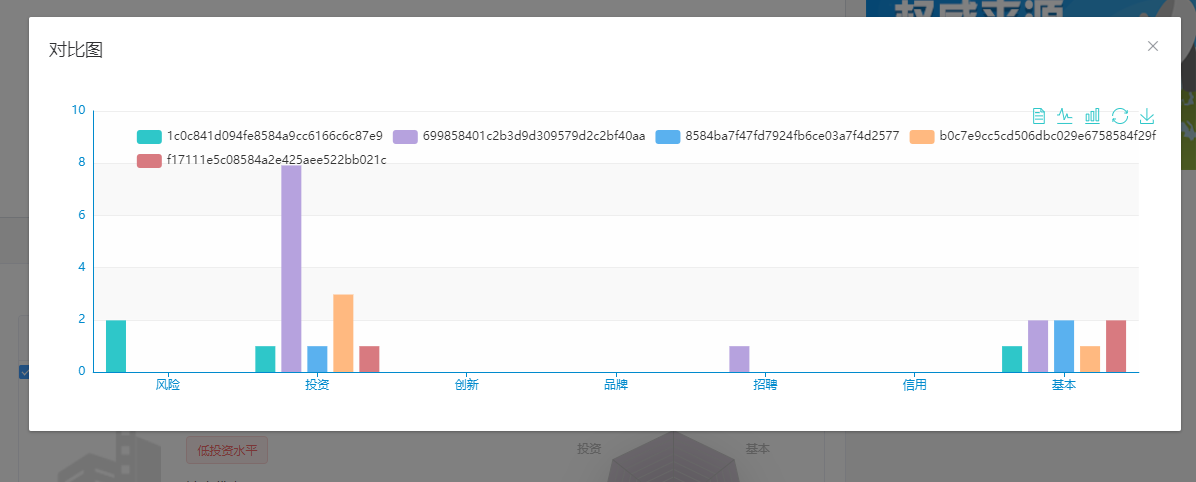


图4-9：企业数据对比

由于每个模块的属性指标可以量化，用户可以筛选每个模块指定的范围，来筛选心仪的企业，如图4-10所示：



图4-10：按模块指标筛选企业

用户通过点击企业信息，可以进入企业详细信息界面，如图4-11所示。这里显示了企业的基本经营信息，可以通过点击经营风险、投资信息、知识产权等，查看这些模块详细信息条目。

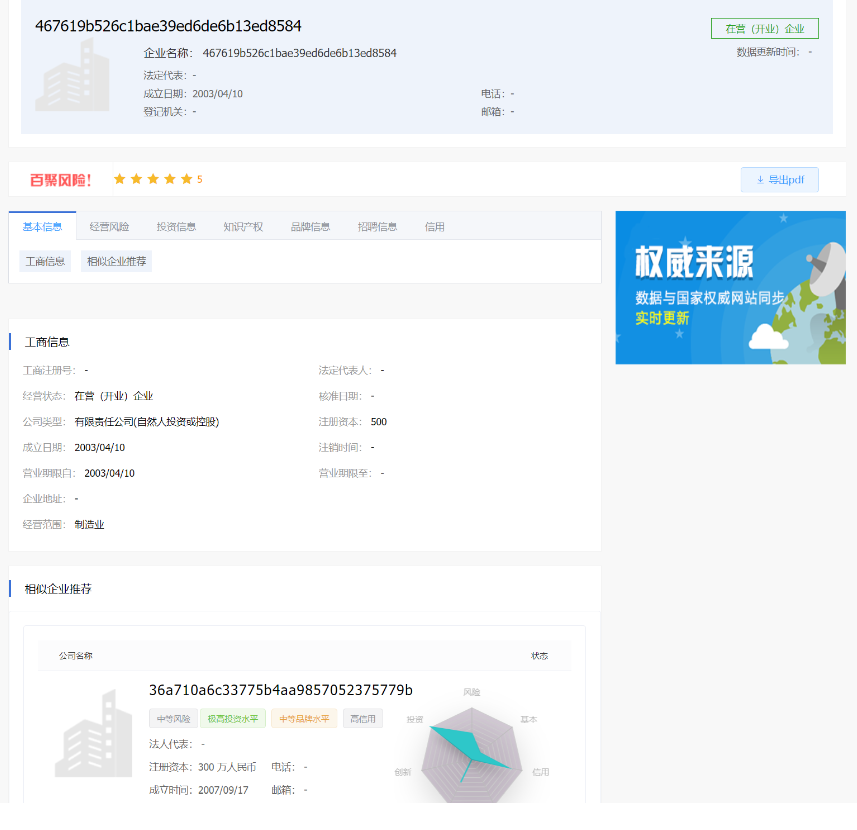


图4-11：企业详细信息界面

例如点击企业的知识产权，可以查看该企业知识产权相关记录，并查看该模块企业的评分，如图4-12所示，其他模块也类似，不再列举：



图4-12：企业知识产权模块信息

* **综合评估**

平台支持在线生成评估文档，整合企业详细信息，附加以属性在行业内水平的评分，以及图表分析、标签含义解释等，使得用户能够快速评估企业综合水平。

如果要查看整个企业的完整信息，也可以在图4-11的右上角，点击导出PDF，可以生成企业的PDF评估报告，生成PDF报告如图4-13所示，所包含的内容如图4-14的目录所示，以企业背景为例，信息如图4-15所示：

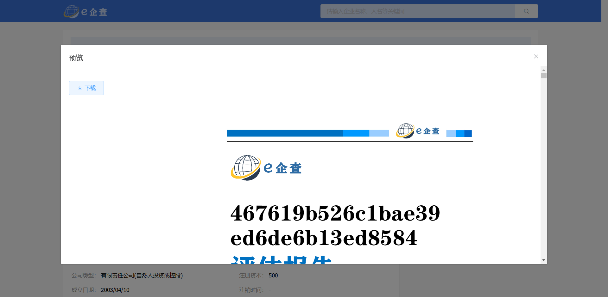


图4-13：企业PDF报告预览

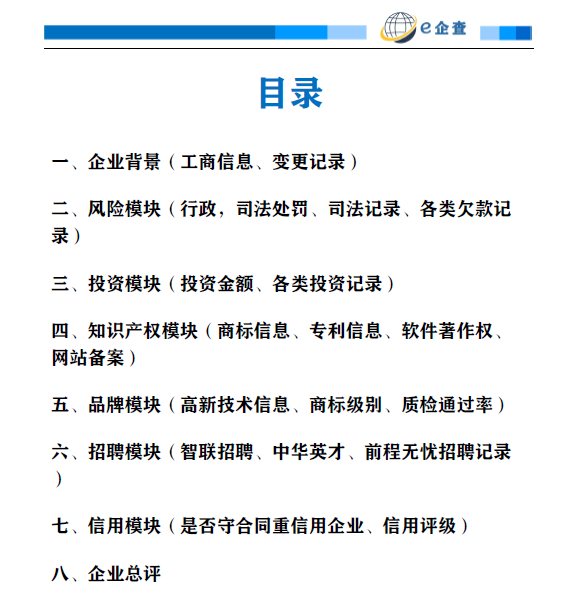


图4-14：企业PDF报告包含内容

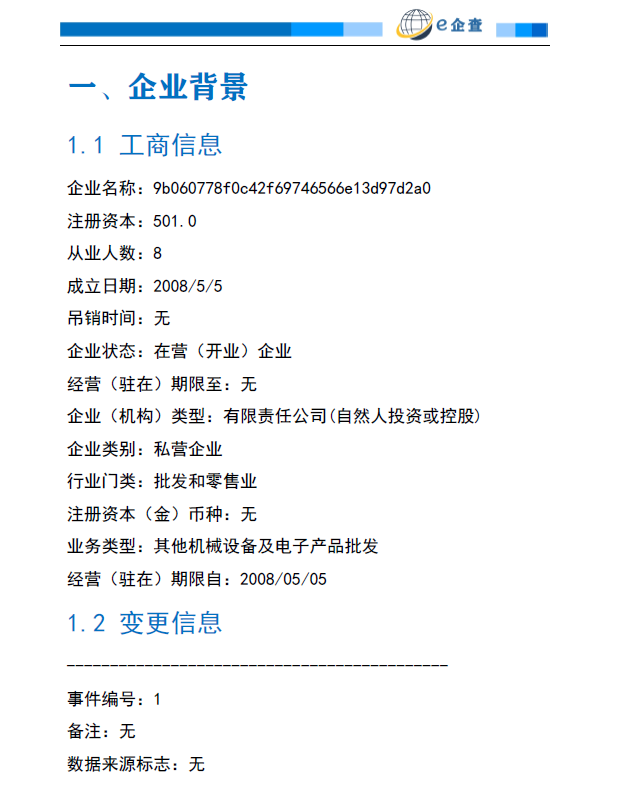


图4-15：PDF企业背景信息摘要

* **智能推荐**

用户查找的返回界面，找到平台推荐的与这些企业相类似的企业。推荐的依据是综合行业门类，企业总分，以及企业属性关联相似性，按照匹配度从高到低进行推荐。

在图4-11的下方，系统根据聚类算法得到的标签，为企业推荐当前与其类似的企业，如图4-16所示，原企业如图4-17所示，可以看出推荐企业与原企业有一定相似之处，所以被归到一类当中：

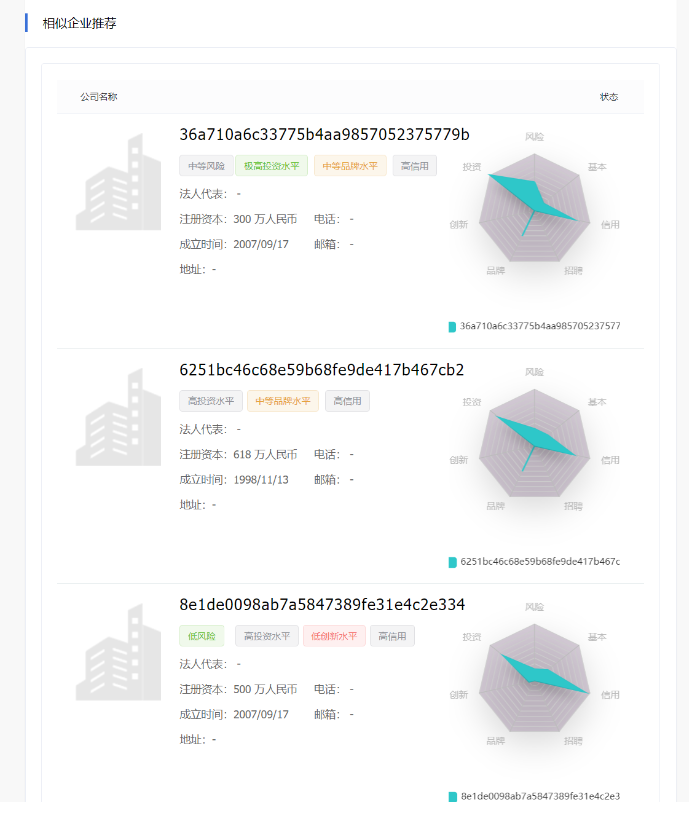


图4-16：推荐相似企业



图4-17：原企业

### 4.3.2数据管理员

管理员使用“e企查”的后台模块。进入后台首先需要登录，如图4-18所示。

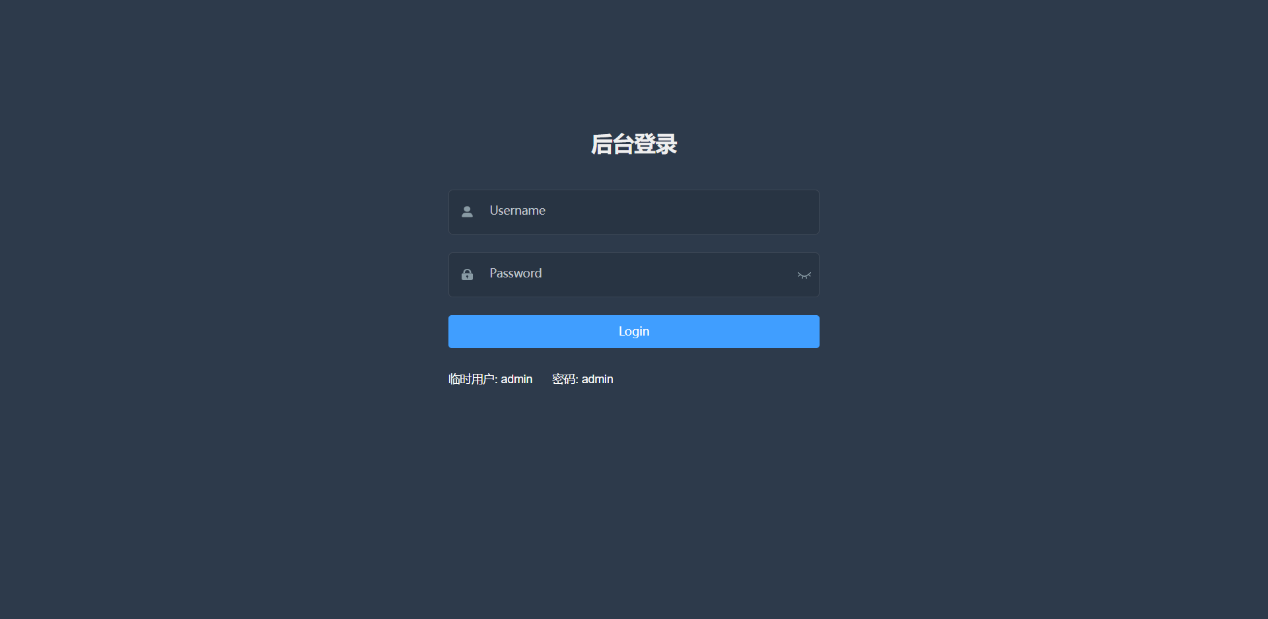


图4-18：e企查后台登录模块

* **数据管理**

数据管理员可以通过添加指定格式csv的形式，实现企业数据的批量导入。如图4-19所示：

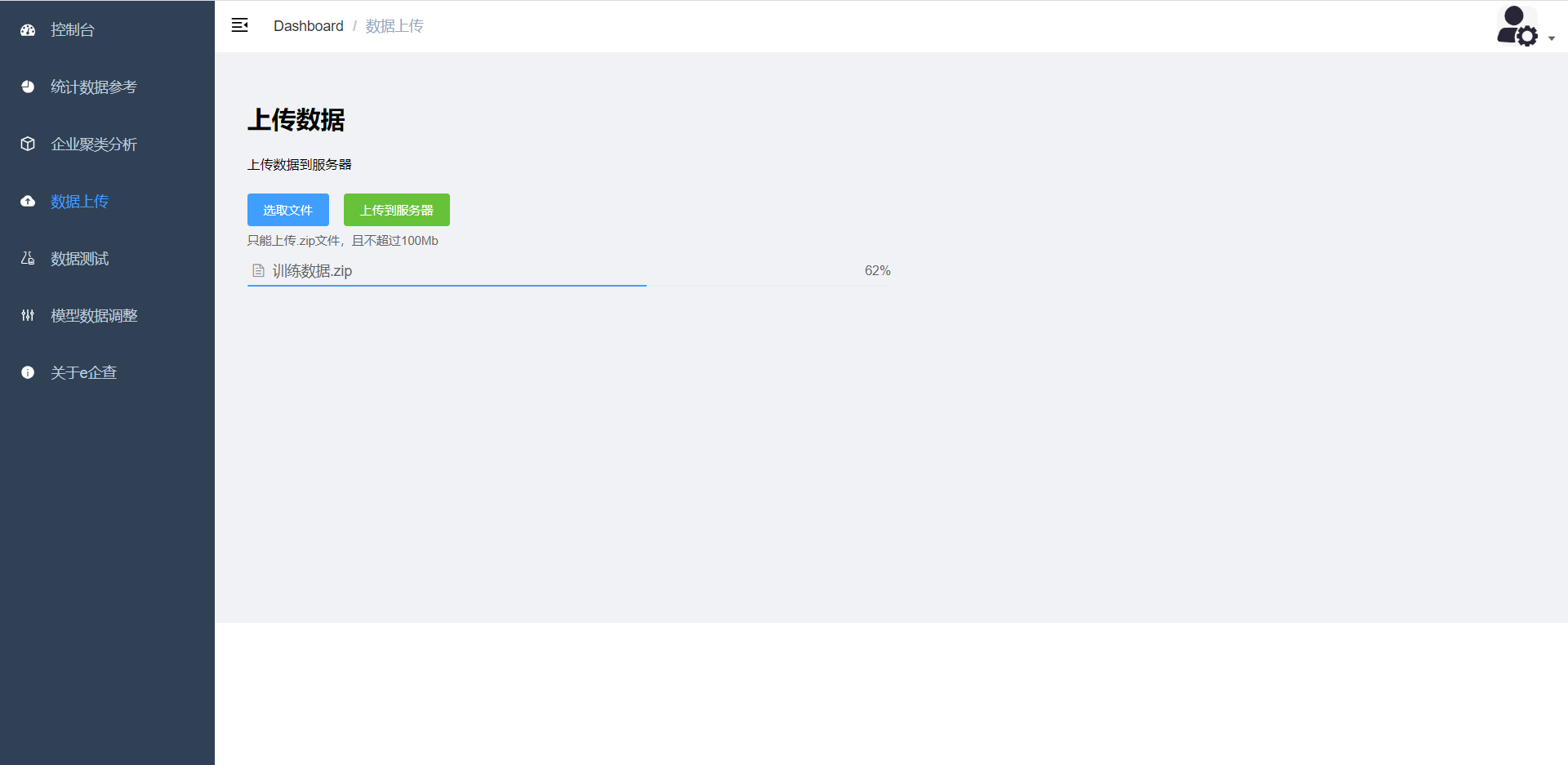


图4-18：导入数据至服务器

* **管理驾驶舱**

“e企查”提供了丰富的可视化效果，能够帮助使用者了解聚类的结果，以及每个簇内，企业在各个维度（模块）下数据的分布，以及聚类的结果。

如图4-20所示，点击统计数据参考，即可分析即可分析每个簇的情况。

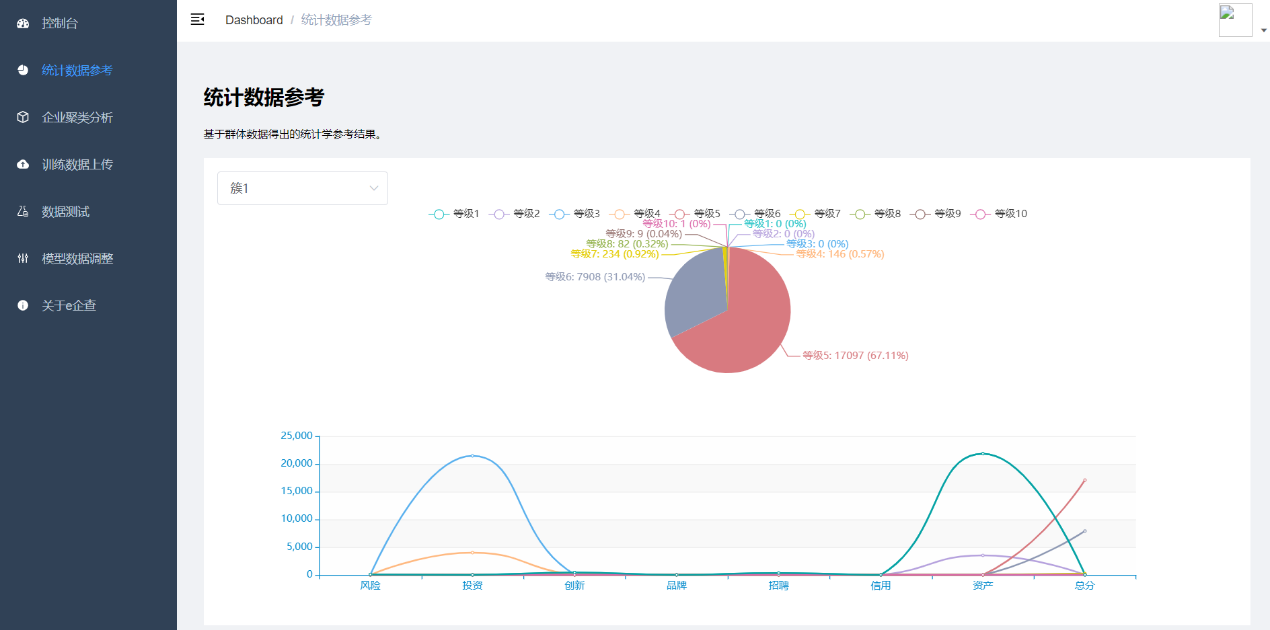


图4-20：统计数据参考——簇分析

可以在左上角点击自己想要了解的簇。进入指定簇后（以簇1为例），可以发现下方有风险、投资、创新、品牌、招聘、信用、资产、总分，共8个维度，每一次光标落到该区域，就可以分析对应模块内，企业等级占比。

如图4-20，当前滑动到总分模块，就可看到，簇1投资模块中，总分等级为5的企业占了67.11%，等级为6的企业占了31.04%，占了全部企业的98%。由于总分等级是0-10，我们就可以得出，这个簇的特征，在总评方面，属于综合水平中等的企业。剩余的模块，也是可以通过这个方法同理分析得出。

如图4-21所示，反映了企业数据的整体特征。由于通过等级化算法，将企业最终的等级划分成了10档，以及空数据等级为0，在左边的柱状图反映了企业11个等级（包括0），在总体企业中的分布，可以看到类似高斯分布，等级4-7的中等企业水平占最多。

图4-21的右侧，反映了我们通过加权K-Means算法，对n个企业的7个维度，即7×n的矩阵进行聚类，划分15个簇，在总体企业中的分布。可以看到每个簇内企业的分布相对均匀。

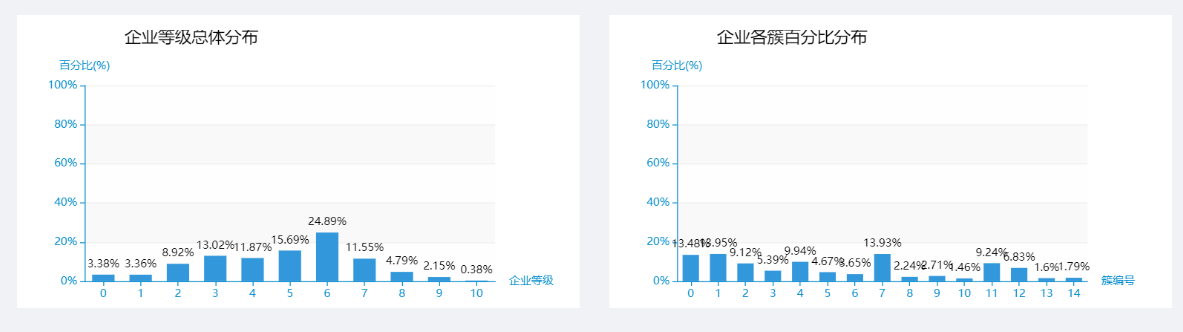


图4-21：企业等级分布与簇分布

由于最终在评估企业涉及到7个维度，无法在一个平面上直观了解最终的聚类结果。所以我们将企业数据的7×n的矩阵，通过PCA的方式，降到二维和三维，显示在网页中，如图4-22和4-23所示。

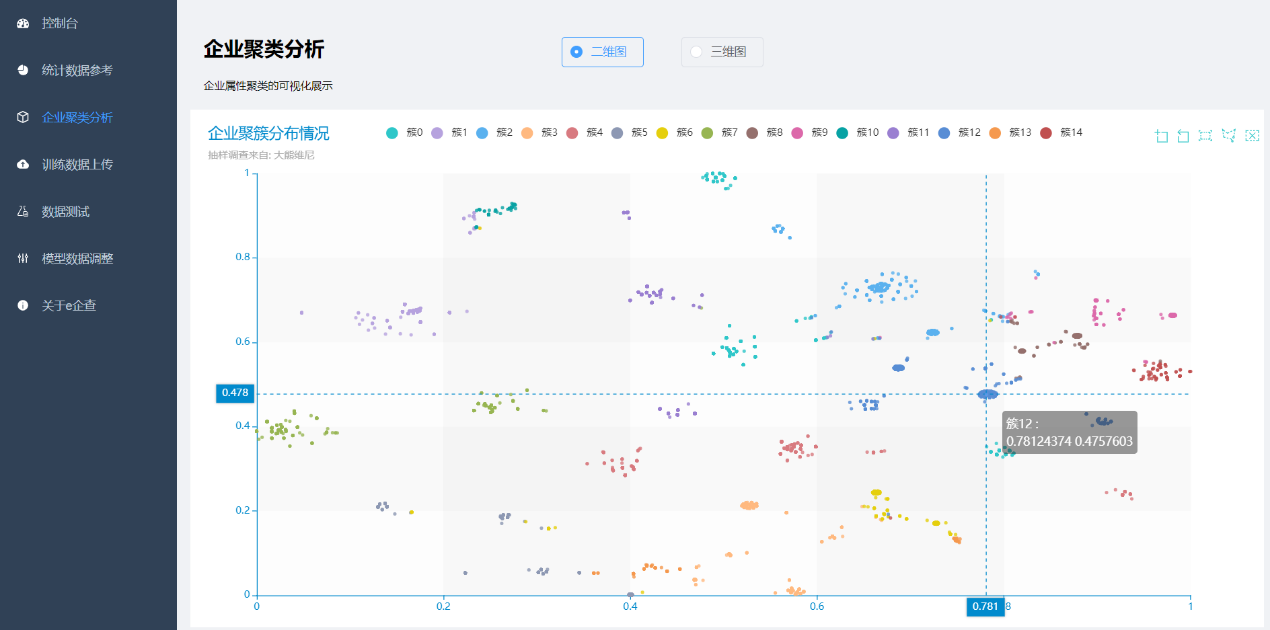


图4-22：PCA降维至二维可视化结果

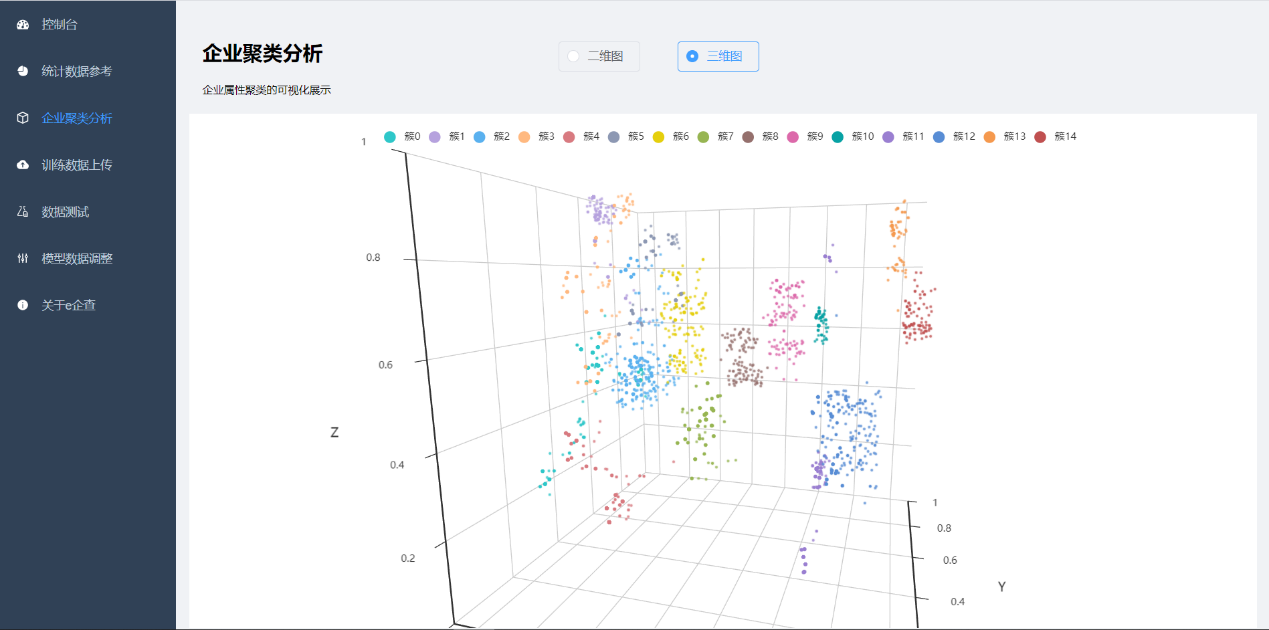


图4-23：PCA降维至三维可视化结果

二维和三维可视化结果，都可以通过拖动网页的形式，选择区域，放大观看。

### 4.3.3系统管理员

* **数据测试**

为了方便测试模型的效果，作为最终模型确定的参考，也为了最终能方便的呈现命题方所需的模型效率、标签描述等信息，单独开设一个测试界面，作为web接口，通过上传数据，返回训练或预测后得到的标签结果文件以及标签描述文件。测试界面如图4-24所示：



图4-24：模型测试界面

数据测试界面分为两部分。

第一部分是训练。用户可以通过上传自己的数据集，用于生成模型。在上传后，返回训练时间，系统响应时间（训练+数据预处理）时间，并在训练结束后，提供一个下载url，供用户下载训练结束后产生的带标签的csv文件，同时提供用于描述每个簇特征的图片。在后台会生成训练模型。

第二部分是预测，预测有两种选择。一种是选择系统默认的模型进行预测，还有一种是选择刚刚新训练出来的模型进行预测，前提是已经训练过。预测结束后，返回的东西和训练一样。

最终返回的csv，描述了每个模块内属性等级的描述，和最终总评等级的描述。如图4-25所示，是模块总评描述：

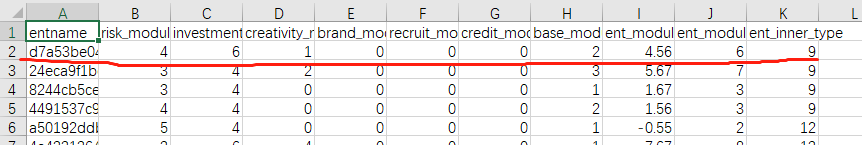


图4-25：总评等级描述

以第一个企业记录为例，分别展示了其风险、投资、知识产权、品牌、招聘、信用、基本信息（资产）的评分，其中0表示，该类记录为空，通过加权（其中风险权重为负）得到ent\_module，即企业总分，通过等级化算法，得到该企业等级为6，通过聚类，将企业分到9号簇中。根据以上描述，我们会在前端给出，中等风险水平，中等投资水平，低创新水平，低资产水平以及中等企业水平的标签描述。

由于9号簇这个描述很抽象，我们需要有一段描述，来概括9号簇的特征。我们在返回的压缩包中，提供了描述每个簇的图片，通过图片即可快速了解该簇特征，如图4-26所示：

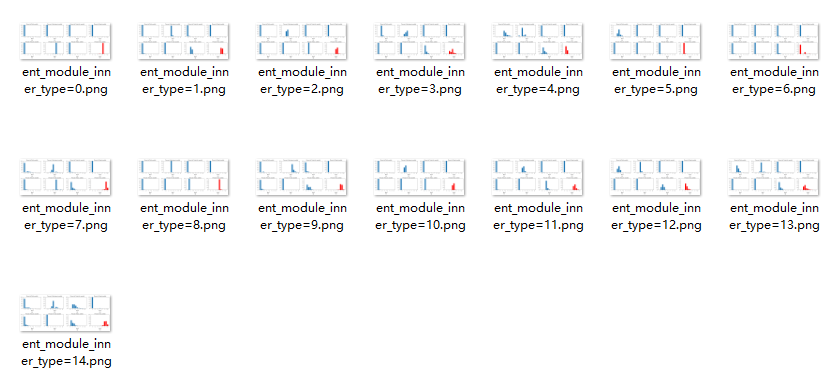


图4-26：簇描述图片

以9号簇为例，打开inner\_type为9的图片，如图4-27所示：

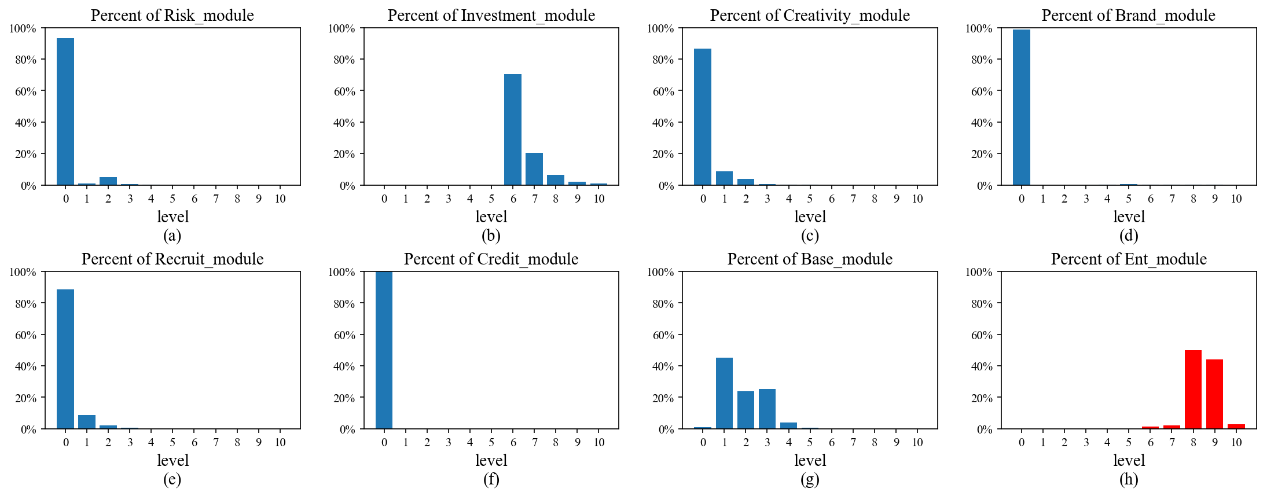


图4-27：9号簇企业各等级分布

在这张图中描述了9号簇的总体特征。其中大多数企业风险为0，投资等级6以上，有0-2的创新等级，资产等级在1-4，企业总体等级在6-10，以7和8为主。该企业归入该簇，总体特征是符合其描述的，可以体现聚类效果较好，簇描述准确。

* **模型管理**

由于模型需要适应数据的变化，管理员可以根据实际情况，调整模型参数的权重，使得聚类结果更加合理。如图4-28所示，是模型的权重调整界面。

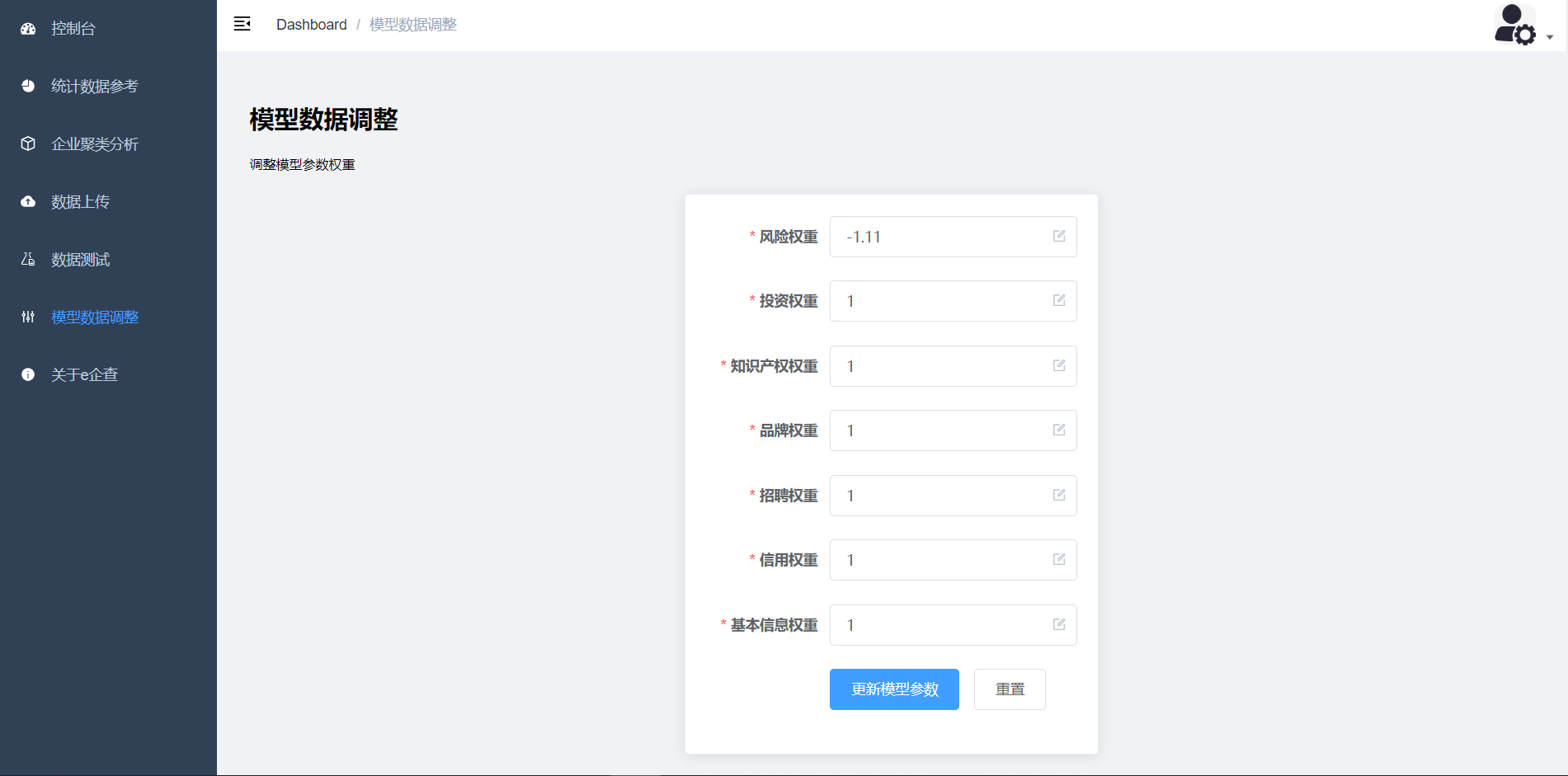


图4-28：权重调整

由于风险对于一个企业的综合总评，起负作用，所以对其权重设置为负。又为了区分空数据，即总分0，和由于正负相加得到的0得到的结果，将风险设置为-1.11。

### 4.4 相关负责人

业务分析人员责任分配如表4-2所示：

表4-2：业务分析人员工作分配

|  |  |  |
| --- | --- | --- |
| 角色名称 | 角色 | 职责 |
| 小鲍 | 负责人 | 确定业务方向，提供前端技术支持，并撰写使用说明。 |
| 小程 | 参与人员 | 为业务需求制定提供后端技术支持。 |
| 小陈 | 参与人员 | 为业务需求提供算法支持。 |
| 小陆 | 参与人员 | 根据具体业务方向，具体化功能。 |
| 小阳 | 参与人员 | 协助项目经理分析需求，提出业务方面意见 |

# 

# 第五章 组织结构

## 5.1组织结构

“e企查”平台拥有一支配合默契，各有所长的开发团队，能够确保项目有条不紊的开发和运转。如图5-1所示，是团队组织架构。

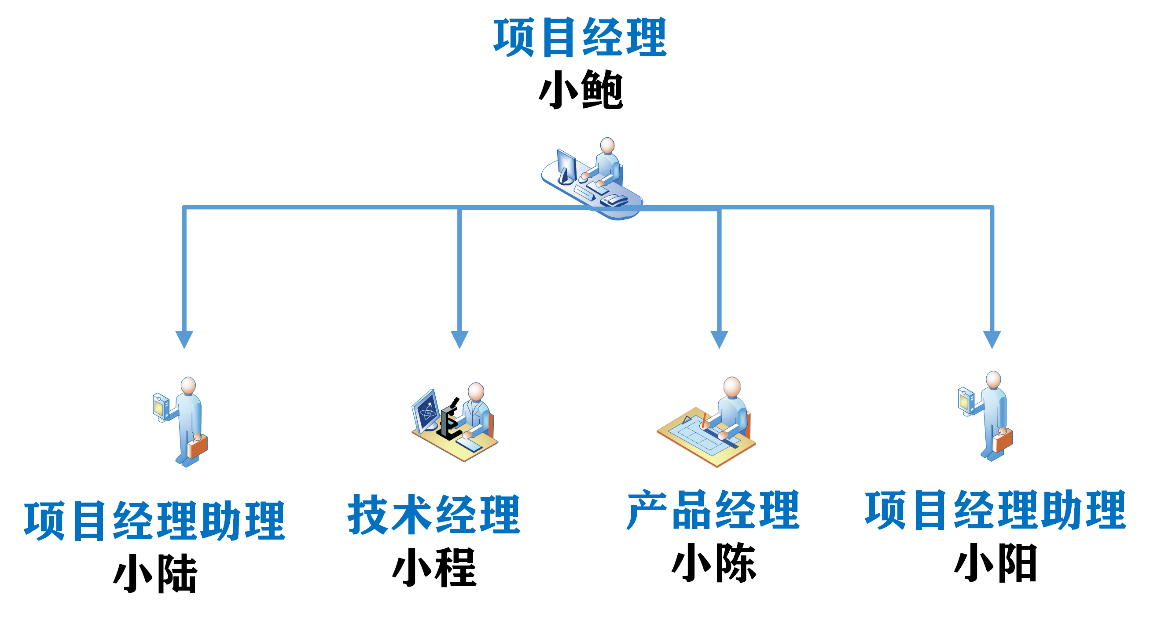


图5-1：团队组织结构

## 5.2角色与职责

团队内的职责与任务分配，见表5-1。

表5-1：团队职责

|  |  |  |
| --- | --- | --- |
| 成员 | 岗位 | 职责介绍 |
| 小鲍 | 项目经理 | 1、负责企业未来决策把握、企业发展目标制定，让企业能够顺应时代快速变革和发展的趋势。  2、确定项目整体设计方案，实现成员间合作调度，确定团队工作计划。  3、组织和协调工作团队，进行资源配置，定期组织团队成员开会，交流工作进展，并给出指导建议，团队建设等。  4、对项目进行质量评估、风险监控、成本控制、最终评估。  5、前端框架制定、业务逻辑实现、UI设计以及前端服务器维护、前后端整合调试。 |
| 小程 | 技术经理 | 1、共同参与制定平台技术标准，进行技术可行性评估与决策。  2、负责系统后端设计、数据库维护以及服务器运维，解决技术难题。  3、指导《项目详细方案》中技术方案部分。  4、编写系统开发过程中的技术文档和《用户手册》  5、负责系统首次安装调试、数据割接和项目测试，负责项目的质量保证工作。 |
| 小陈 | 产品经理 | 1、与项目经理共同制定经营战略，全面负责公司的企划工作。  2、进行市场调研，分析企业的业务需求，给出系统中相应的业务布置建议，撰写《项目详细方案》等文档。  3、负责根据企业的业务需求，进行相应的数据清洗、预处理，并进行数据挖掘、处理与分析以及算法设计。并协助技术经理共同完成对数据库部分的维护。  4、聚类模型的设计、调试以及相应接口API的编写。  5、根据数据分析的结果战略性地对企业未来的业务发展和运营提供响应的建议。 |
| 小陆 | 项目经理  助理 | 1、协助产品经理进行市场调研、完善商业计划和各项同舟计划  2、协助项目经理完善项目整体设计方案、商业计划等。  3、负责项目各项文件中的宣传部分，成功塑造企业形象。  4、构建公司良好的文化氛围和积极向上的精神面貌。 |
| 小阳 | 项目经理  助理 | 1、负责协调客户关系、企业的公共关系工作、增强公司信誉，美化公司形象等。  2、协助完成项目文件中的宣传工作，提供技术流程呈现思路指导。  3、协助项目经理完成《用户手册》编写。  4、协助项目经理和产品经理，完成项目可行性评估、风险部分。 |

## 5.3协作与沟通

### 5.3.1团队内部协作

团队内部协作见表5-2。

表5-2：团队内部协作

|  |  |  |
| --- | --- | --- |
| 方式 | 时间 | 内容 |
| 汇报今日工作计划 | 每天9:00 | 每位成员汇报自己今日的工作计划，对于涉及到人员交接的开发，制定好规范标准。 |
| 日常总结 | 每日23:00 | 每位成员汇报今日工作进度，总结开发遇到的问题，进行集中讨论解决。 |
| 阶段总结 | 3天/次 | 前端负责人将当前web页面开发进度进行汇报演示，后端负责人汇报当前工作模块完成情况，产品负责人汇报当前相关文档编写进度。 |
| 自由讨论 | 不定期 | 对于遇到的问题及时讨论解决。 |

### 5.3.2项目接口人员

项目接口内部协作见表5-3。

表5-3：项目内部协作

|  |  |  |
| --- | --- | --- |
| 接口名称 | 人员（职务） | 责任 |
| 项目总接口 | 小鲍 | 负责项目总体计划、实施。 |
| 技术接口 | 小鲍、小程、小陈 | 负责前后端、数据挖掘技术落实，实现产品经理提出的需求。 |
| 质量接口 | 小鲍、小程 | 系统调试、质量检测。 |
| 客户接口 | 小陈、小陆、小阳 | 对接用户提出的需求，具体化实现手段。 |
| 商务接口 | 小陈、小陆、小阳 | 完善商业计划，平台推广。 |

## 5.4项目日程

项目日程安排见表5-4。

表5-4：项目日程表

|  |  |
| --- | --- |
| 时间 | 任务描述 |
| 2020.1.25-2020.2.25 | 研究命题方给出的数据集，确定数据清洗、分类、挖掘策略，确定模型选择和评估的实验方案。 |
| 2020.2.26-2020.3.1 | 算法与实验说明文档撰写。 |
| 2020.3.1-2020.3.7 | 模型训练与实验部分完成。 |
| 2020.3.7-2020.3.10 | 筹备平台搭建，确定网页构建技术路线。 |
| 2020.3.10-2020.4.15 | 前后端同步开发，项目方案撰写。 |
| 2020.4.15-2020.5.1 | 项目集成与测试，完善项目策划书。 |
| 2020.5.1-2020.5.20 | 完成项目策划书、PPT、视频等辅助展示工具。 |

## 5.5相关负责人

项目组织责任分配表见表5-5。

表5-5：项目组织责任分配表

|  |  |  |
| --- | --- | --- |
| 角色名称 | 角色 | 具体职责 |
| 小鲍 | 主要负责人 | 根据需求确定平台业务方向、功能，并完成成员的任务分工 |
| 小陈 | 次要负责人 | 协助项目经理明确系统功能，提炼需求成对应开发任务，文档编写。 |
| 小程 | 参与人员 | 参与讨论、提出建议、会议记录 |
| 小陆 | 参与人员 | 参与讨论、提出建议、会议记录 |
| 小阳 | 参与人员 | 参与讨论、提出建议、会议记录 |

# 第六章 质量保证与测试计划

## 6.1质量保证计划

软件测试就是在软件投入运行前，对软件需求分析、设计规格说明和编码的最终审查，是软件质量保证的关键步骤。为保证软件的质量和可靠性，应力求在分析、设计等各个开发阶段结束前，对软件进行严格的技术评审。总体的测试流程如图6-1所示。

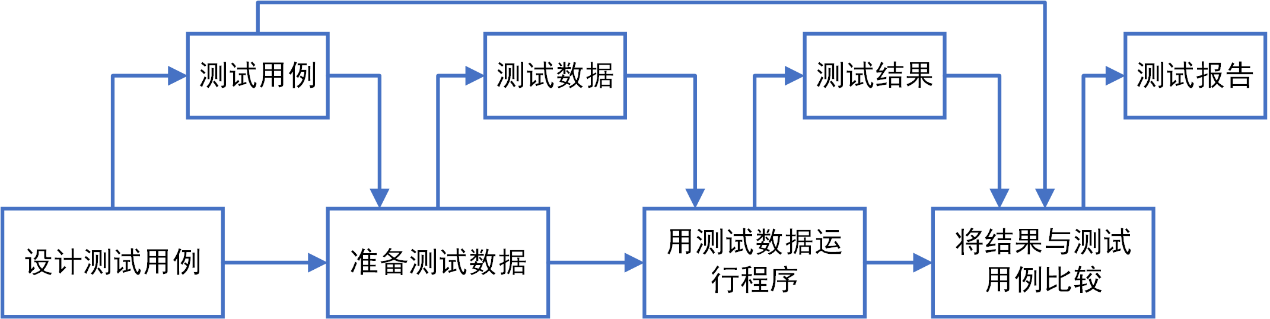


图6-1：质量测试流程

为了将团队内各个部门环节的质量监管活动严密组织起来，以及将产品设计、研制开发、销售服务和情报反馈的整个过程中影响产品质量的一切因素控制起来，形成有明确任务、职责、权限，相互协调的测试计划，制定质量保证计划如表6-1。

表6-1 质量保证计划

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 编号 | 质量控制活动 | 指标 | 计划值 | 参考值 | 方式 |
| 1 | 需求评审 | 评审缺陷率 | <5% | 1% | 组内讨论 |
| 2 | 设计评审 | 评审缺陷率 | <5% | 1% | 组内讨论 |
| 3 | 技术路线评审 | 评审缺陷率 | <5% | 1% | 组内讨论 |
| 4 | 代码评审 | 评审缺陷率 | <5% | 2% | 测试 |
| 5 | 单元测试 | 覆盖率 | 95% | 98% | 测试 |
| 6 | 系统测试用例评审 | 评审缺陷率 | <3% | 1% | 组内讨论 |
| 7 | 系统测试通过率 | 一轮通过率 | >80% | 90% | 测试 |
| 8 | 系统测试通过率 | 二轮通过率 | 95% | 98% | 测试 |
| 9 | 系统测试缺陷率 | 测试缺陷率 | <8% | 5% | 测试 |

## 6.2单元测试计划

### 6.2.1测试条件

1、服务器上的系统测试在华为云上进行：

（1）CPU：1核

（2）内存：2GB

（3）操作系统：CentOS 7.6 64位

2、本地测试环境：

（1）CPU：Intel(R) Core(TM) i7-8750H CPU @2.20GHz(12 CPUs),~2.2GHz

（2）内存：16GB

（3）操作系统：Win10专业版64位

### 6.2.2测试方法

采用白盒、黑盒相结合的测试方法。代码及代码编制规范的检查采用白盒测试方法，功能及性能测试采用黑盒测试方法。功能测试时，首先人工进行功能性测试，再利用测试工具进行回归验证测试。

### 6.2.3人员计划

人员安排如表6-2所示：

表6-2：测试人员安排表

|  |  |  |  |
| --- | --- | --- | --- |
| 编号 | 单元模块名称 | 人员安排 | 测试方法 |
| 1 | 用户基本信息管理 | 小鲍 | 白盒测试 |
| 2 | 企业画像构建 |
| 3 | 智能推荐逻辑 |
| 4 | 企业数据维护 | 小程 |
| 5 | 筛选逻辑 |
| 6 | 可视化分析 | 小陈 |
| 7 | 模型分析 |

### 6.2.4专题测试

1、数据和数据库完整测试

数据库需要作为一个单独的部分进行测试，因为其后期既要方便机器学习的模型批量进行处理，也要方便后端的Django通过模型的形式进行ORM映射。此外由于企业数据量巨大，表之间也存在较为复杂的依赖关系，因此外键、索引、存储过程、触发器的设计，都成为了数据库设计和测试中需要重点关注的点。数据库测试安排见表6-3。

表6-3：数据库测试安排

|  |  |
| --- | --- |
| 测试目标 | 确保数据库的CRUD过程能够按照标准进行 |
| 方法 | 1、数据库的相关缺失数据，是否符合缺失的原则，需要填充的数据是否填充了相应的缺省值。  2、对于频繁查询的表，通过MySQL的explain语句，查看索引的工作过程，并统计其响应时间是否达到要求。  3、数据表中的数据在增加和更新后，触发器能否正常工作；在删除时，级联数据是否能够正常删除。 |
| 完成标准 | 所有数据库的访问方法都可以按照设计正常工作 |
| 需要特殊考虑的事项 | 1、通过navicat数据库可视化工具设置好索引、触发器、存储过程等工作后，将sql文件导出部署到服务器上，加快配置速度。  2、测试要首先在本地进行测试，在成功达标后再部署到服务器上再进行一轮测试。  3、数据库在进行操作测试时，及时做好备份，避免不必要的后果出现。 |

2、后端功能测试

后端在测试时，可以采用postman等第三方插件，对请求进行模拟，测试其是否能按前端的json格式返回正确的数据；也需要确定django的model能否正确与数据库表关联，并用面向对象的方式对数据库进行操作，并保证数据库在更新时同步一致；最后还需要检测业务逻辑是否正确。后端功能测试安排见表6-4。

表6-4：后端功能测试安排

|  |  |
| --- | --- |
| 测试目标 | 确定后端业务逻辑、数据库关联正确 |
| 方法 | 使用postman模拟各种前端请求，关注返回类型和model的ORM能否正确的与数据库关联，进行正常的操作。 |
| 完成标准 | 正确按照前端请求返回对应格式json，数据库在django的model操作下，能够保证数据库更新同步一致，查询效率有保障。 |
| 需要特殊考虑的事项 | 需要判断数据的有效性 |

3、前端功能测试

包括前端UI布局以及常规的业务逻辑。能够确保向用户提供轻易上手的交互界面，符合预先设定的公司标准。也能够进行一部分的逻辑校验，减轻后端负担，前端功能测试安排见表6-5。

表6-5：前端功能测试安排

|  |  |
| --- | --- |
| 测试目标 | UI设计是否符合预先标准，业务逻辑正确 |
| 方法 | 采用Chrome开发者工具对数据的传输进行调试，验证js脚本的工作情况；使用极端测试样例对网站进行测试，检验其是否能正常工作。 |
| 完成标准 | UI设计符合用户使用体验，交互性良好，能够在前端准确进行数据校验，业务逻辑正确。 |
| 需要特殊考虑的事项 | 需要在IE、FireFox、Chrome等主流浏览器下检验效果。 |

4、机器学习模型测试

由于此时模型的种类以及相关参数已经确定，所以需要对模型的整体预处理、训练、预测效率进行统计。尤其需要的关注是训练对象指标的合理性，以及最后训练完的结果，是否能够将企业分成特征明显的簇群，机器学习测试安排见表6-6。

表6-6：机器学习测试安排

|  |  |
| --- | --- |
| 测试目标 | 模型的训练预测时间可接受，生成标签区分度高 |
| 方法 | 将数据以csv的形式存储，并使用pandas对数据进行读取和管理，并用matplotlib等可视化库，对聚类后的最终结果，以雷达、柱状图的形式，呈现企业各项指标，统计其分类效果。 |
| 完成标准 | 企业最终的分类效果明显，每个簇中的企业，都能拥有特征相似的多个相同标签。 |
| 需要特殊考虑的事项 | 企业的数据从不同网站获取而来，需要注意单位统一，度量转换。 |

### 6.2.5测试文档

所有测试的模块都需按照以下单元测试报告范本填写，如表6-7所示。

表6-7：单元测试报告范本

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 开发项目名称 | | | e企查企业分类系统 | | | | | | | 开发项目编号 | | | | TEST-20200301 | | | 第一责任人 | | | | | 小鲍 | |
| 单元名称 | | 企业画像构造 | | | | 责任人 | | | | 小鲍 | 单元所属子系统 | | | | | | 企业推荐 | | | | 开发周期 | | 1周 |
| 代码测试检查： | | | | | | | | | | | | | | | | | | | | | | | |
| 代码测试内容 | | | | 测试人员 | | | | 测试结果 | | | | | | | | | | | | 备注 | | | |
| 路径测试 | | | | 小程 | | | | 正确 | | | | | | | | | | | |  | | | |
| 声明测试 | | | | 小程 | | | | 正确 | | | | | | | | | | | |  | | | |
| 循环测试 | | | | 小程 | | | | 正确 | | | | | | | | | | | |  | | | |
| 边界测试 | | | | 小程 | | | | 未出现越界情况 | | | | | | | | | | | |  | | | |
| 接口测试 | | | | 小程 | | | | 接口运行未出现错误 | | | | | | | | | | | |  | | | |
| 界面测试 | | | | 小陈 | | | | 未出现错误 | | | | | | | | | | | |  | | | |
| 数据确认测试 | | | | 小陈 | | | | 模块工作正常 | | | | | | | | | | | |  | | | |
| 代码走查 | | | | 小陈 | | | | 正确 | | | | | | | | | | | |  | | | |
| 功能测试： | | | | | | | | | | | | | | | | | | | | | | | |
| 序号 | 功能名称 | | | | | | 操作方法 | | | | | | 结果 | | | 建议 | | | 测试人员 | | | | 备注 |
| 1 | 数据分析 | | | | | | 白盒测试 | | | | | | 功能正常 | | |  | | | 小陈 | | | |  |
| 测试结论 | 该功能运行正确，但反应速度还有待提高 | | | | | | | | | | | | | | | | | | | | | | |
| 责任人 | 小陈 | | | | | | 项目第一责任人 | | | | | | | | | 小鲍 | | | | | | | |
| 审核 | 合格 | | | | | | | | | | | | | | | | | | | | | | |
| 项目组 | 小鲍 | | | | 测试组 | | | | 小陈 | | | 总工办 | | | 小陈 | | | 总工程师 | | | | 小程 | |

## 6.3系统测试计划

### 6.3.1测试条件

同6.2.1

### 6.3.2测试方法

黑盒测试。

### 6.3.3人员计划

团队的5位成员，以及指导老师。

### 6.3.4专题测试计划

1、安全性和访问控制测试

安全性和访问控制测试侧重于安全性的两个关键方面：

（1）应用程序的安全性，包括对数据或业务功能的访问；

（2）系统的安全性，包括对系统的登录或远程访问。

应用程序的安全性可确保：在预期的安全性情况下，用户只能访问特定的功能，或者只能访问有限的数据。例如，用户对某模块的访问权限只有当超级管理员授予后才能访问，授予权限包括完全访问、只读权限、无权限。用户不能越权访问。

系统的安全性可确保只有具备系统访问权限的用户才能访问应用程序，而且只能通过相应的网关来访问，如表6-8所示。

表6-8 安全性和访问控制测试安排

|  |  |
| --- | --- |
| 测试目标 | 功能/数据的安全性：核实用户只能访问其所属用户类型已被授权访问的那些功能/数据。 |
| 方法 | 功能/数据的安全性：核实用户只能访问其所属用户类型已被授权访问的那些功能/数据。  系统安全性：核实只有具备系统和应用程序访问权限的那些用户才能访问系统和应用程序。 |
| 完成标准 | 各种已知的用户类型都可访问相应的功能或数据，而且所有事务都按照预期的方式运行，并在先前的应用程序功能测试中运行了所有的事务。 |
| 需考虑的特殊事项 | 必须与相应的网络或系统管理员一起对系统访问权进行检查和讨论由于此测试可能是网络管理或系统管理的职能，可能会不需要执行此测试。 |

2、负载测试

负载测试评测方法将使测试系统承担不同的工作量，以评估系统在不同工作量条件下的持续正常运行的能力。负载测试的目标是确定并确保系统在超出最大预期工作量的情况下仍能正常运行。此外，负载测试还要评估性能特征（响应时间、事务处理速率和其他与时间相关的问题），如表6-9所示。

表6-9 负载测试安排

|  |  |
| --- | --- |
| 测试目标 | 核实所指定的事务在不同的工作量条件下的系统响应时间。 |
| 方法 | 使用为业务周期测试制定的测试。  通过修改数据文件来增加事务数量，或通过修改测试来增加每项事务发生的次数。 |
| 完成标准 | 大量数据的处理：在可接受的时间范围内成功地完成测试，没有发生任何故障。 |
| 需考虑的特殊事项 | 负载测试应该在专用的计算机上或在专用的机时内执行。这样可以实现完全的控制和精确的评测。  负载测试所用的数据库应该是实际大小或相同缩放比例的数据库。 |

## 6.4相关负责人

表6-10项目组织责任分配表

|  |  |  |
| --- | --- | --- |
| 角色名称 | 角色 | 职责 |
| 小程 | 主要负责人 | 负责测试总体安排进度，整理测试文档标准。 |
| 小鲍 | 参与人员 | 负责前后端模块接通调试，白盒测试. |
| 小陈 | 参与人员 | 算法模型调试，以及系统和模块的黑盒测试。 |
| 小陆 | 参与人员 | 协助整理测试结果，并汇总文档。 |
| 小阳 | 参与人员 | 规范测试结果，撰写测试方案。 |

# 第七章 商业计划

## 7.1市场机遇

小微企业指的是雇佣人员较少，销售收入和资产总额较低的企业，和大中型企业相比，小微企业具有企业规模小，企业数量多，信贷需求量大，信贷需求增长快的特点，由于小微企业发展对于我国国民经济发展的重要推动作用，国务院在2008 年出台的“国九条”金融政策中要求商业银行的贷款业务向小微企业倾斜，增加小微企业的贷款比重。2012 年出台的《进一步支持小微企业健康发展指导意见》认为，小微企业的发展对于经济有重要的推动作用，并且细化了支持政策。财政部在2014 年联合国家税务总局发布了《小微企业税收减免的指导性措施》。

随着地方性商业银行的不断增加，金融市场之间的竞争日趋激烈，各金融信贷机构为了能够在市场竞争中占据主动，也都加大了小微企业信贷产品的开发力度。目前，各家金融信贷机构都推出了各具特色的小微企业信贷产品，信贷产品体系也得到了极大丰富。

市场上，存在例如天眼查、企查查等平台，金融信贷机构可以通过这些平台搜集的公开的企业信息，来挖掘企业数据，查企业、查老板、查关系，查征信、查信用、找客户、谈合作、做调研，企业信息维度丰富且全面，极大的降低了信息获取的压力。

然而，小微企业的信贷产品，特别是在企业信用审核评估仍然存在一些问题，主要表现在以下几个方面：

（1）金融信贷机构普遍缺乏动态的风险管理手段，在长期信贷管理中存在着重审批轻管理的问题，机构只重视申请阶段企业风险的筛查，而忽视了企业在得到信贷后的动态变化，导致小微企业信贷质量有所下降。

（2）金融信贷机构普遍采用传统的企业要素评价方法，缺乏针对小微企业综合信用评估体系。小微企业的信息，往往没有大企业全面和准确。小微企业信息获取后，如何针对一些缺失数据和异常数据的处理，并制定一种统一的标准，确保最终评估结果的合理性，存在较大压力。

（3）从平台上获取的企业属性维度较高，如何从中筛选，衡量权重，是一项工程较大的工作，这也导致了金融信贷审核流程过长，导致了放贷慢的弊端；而信贷机构本身也要面临着信用审查时巨大的人工压力，构建企业画像面临巨大挑战，还有可能存在一定出错率。

（4）金融信贷机构采用传统的企业要素评价方法，只能形成企业现阶段静态的综合水平评价，无法利用这些数据对企业进行有效合理的分类，实现后期的客群分析、精准获客等高级需求，挖掘潜在信贷客户，加速信贷流程和资金流通。

（5）企业的数据经过清洗和预处理后的数据，仍然可能较为抽象，此外对于企业属性而言，可能存在大量专业名词，对于非专业人士而言，仍然可能无法直观的评估企业的综合信贷水平，降低了企业水平评估的直观性和流通性。

基于以上问题，“e企查”企业无监督分类平台为用户提供智能、简洁、高效、直观的企业信用风险评估和分类的一站式服务，通过人工智能分类的方法，及时为用户推荐潜在客群；以辅助评估的方式，更好的比对企业间的优劣，更好的助力“一贷通”的实现。

## 7.2需求分析

### 7.2.1市场环境

#### 7.2.1.1现状及趋势

（1）小微企业融资现状：

A．融资难

①部分小微企业缺少符合商业银行要求的合格押品。一般而言，商业银行对押品有真实存在、权属清晰、符合法律规定、易变现、可设定抵押权的要求，针对不同押品设定不同抵押率。部分小微企业初始创业，难以提供商业银行认可的押品，导致融资难。

②部分小微企业未建立财务报表制度或财务报表不真实，商业银行难以根据小微企业财务状况作出信贷准入的客观判断。

B．融资贵

①调头资金高成本增加小微企业财务成本。小微企业贷款到期不能归还时，通过融入高成本调头资金归还到期贷款导致“融资贵”。以某正规金融服务公司提供调头资金为例，收取的费用包括两部分：调头服务费、咨询服务费。资金占用时间为5 天、10 天、15 天、20 天、25 天的收取费用分别是0.42%、0.66%、0.88%、1.07%、1.24%。据了解，民间借贷调头资金成本远远高于上述比例。

② 部分小微企业贷款到期不能偿还，或担心归还贷款后难以再取得贷款，导致短贷长用，以借新还旧方式继续使用贷款资金，商业银行贷款利率定价时上调一个限档次。

（2）金融信贷机构融资现状

A．中小企业征信体系不完善

从银行方面看,中小企业存在过高的经营风险, 这阻碍了银行支持中小企业信贷的加强。反而,金融企业对中小企业的贷款会十分谨慎。从企业方面来看,不少中小企业缺乏信用观念, 逃避银行等金融企业债务现象越来越严重。企业违约后, 金融企业很难收回贷款本息,因此不得不加强信贷管理, 提高放贷条件,致使中小企业贷款难的现象比较普遍。

B. 信用担保制度与补偿制度缺乏

担保制度与补偿制度缺乏成为中小企业信贷的主要障碍。中小企业信贷中存在的实物资产担保,抵押难是普遍现象。同时,目前面向中小企业的信用担保业务发展还难以满足广大中小企业提升信用能力的需要,由于担保的风险分散与损失分担及补偿制度尚未形成,使得担保资金的放大功能和担保机构的信用能力均受到较大制约。

C．中小企业信息数据失真

目前,我国全社会信用体系和征信管理还未形成。同时,中小企业经营随意性大,内部管理不健全,金融机构与企业信息不对称,金融企业很难获得中小企业真实的经营情况。中小企业信息披露不够和财务数据失真,使银行贷前调查困难,无法有效评估风险。

D．中小企业信贷收益和成本不相配

由于中小企业财务信息不可靠,企业素质不均衡,金融企业用于中小企业贷款的精力远大于大型企业;而由于中小企业单个贷款规模比较小,一笔贷款给金融企业带来的收益又比较少,造成金融企业的收益和成本不成正比,相对增加了金融企业的经营管理成本。

#### 7.2.1.2企业信息查询行业现状

在国家一系列大数据相关政策的指引下，近年来我国大数据产业实现了跨越式的增长与发展。2014年是我国企业信息查询平台的发展元年。用户对企业信息以及商业类大数据查询需求的增大，让这些企业信息查询平台快速崛起。

用户争夺方面，企查查在2015年率先宣布盈利，其月活跃用户量一直领先。直到2017年，天眼查和启信宝的月活跃用户开始反超企查查。如图7-1所示，反映了2017-2019年各企业信息查询平台月活跃用户变化情况。

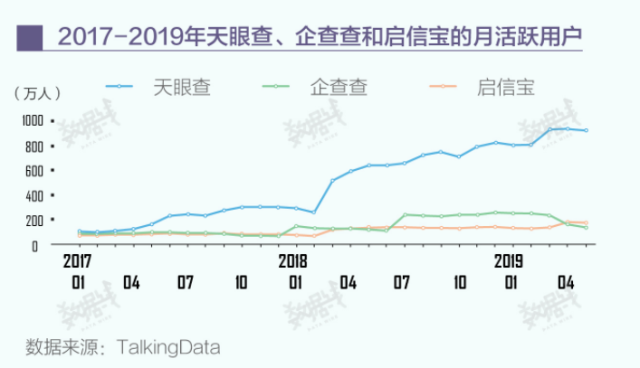


图7-1：2017-2019年各企业信息查询平台月活跃用户变化情况

截止2019年6月，天眼查月活跃用户已达900多万人，是企查查的9.55倍，启信宝的6.32倍。同时，也有数据显示自2019年4月起，天眼查人均单次使用时长赶超企查查。如图7-2，反映了2017-2019年各商业查询平台人均单次使用情况。

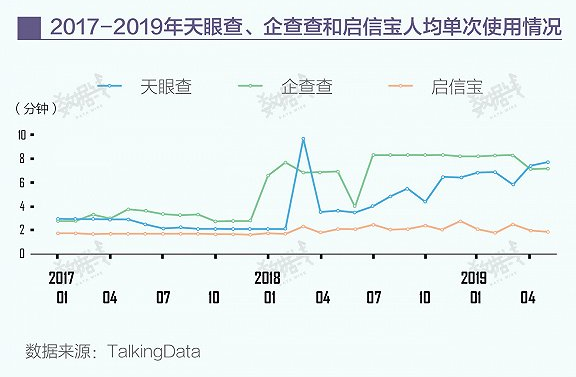


图7-2：2017-2019年各商业查询平台人均单次使用情况

在业务方面，以天眼查为例，天眼查的数据采集自开放和共享的政府公共数据覆盖2000多个数据源网站，经历数据清洗、数据聚合、数据建模等流程，包含数千项数据指标与上万个数据模型，最后以可视化形式呈现出来。



图7-3：三大商业查询平台

### 7.2.2企业画像构建平台商业模式分析

如表7-1所示，分析了三大商业查询平台的特点，异同。

表7-1：三大商业查询平台对比

|  |  |  |  |
| --- | --- | --- | --- |
| 竞品  维度 | 天眼查 | 企查查 | 启信宝 |
| 产品定位 | 查公司、查老板、查关系的**企业信息查询工具商业安全工具**：  企业信息查询、公司查询、工商查询 | 查企业就上企查查-查企业、查老板、查关系  查企业：覆盖全国1.8亿加企业信息，支持多维度查询，企业信息一览无余  查老板：精准获知老板投资，任职，风险信息，让投资、合作更放心  查关系：深度挖掘企业关联关系，错综关系更明了，让商业更透明 | 专业人士都在使用的智能**商业调查工具**  **AI商业调查工具** |
| 数据范围 | 全国1.8亿+社会实体  6000万+高管  全球企业：亚洲、大洋洲、南美洲、欧洲、非洲部分国家或地区（报告形式） | 全国1.8亿加企业信息  全球企业：亚洲、大洋洲、南美洲、欧洲、非洲部分国家或地区（即时+报告形式） | 1.8亿企业和组织机构  全球企业：亚洲、大洋洲、欧洲、美洲、非洲（即时+报告形式） |
| 用户群体 | 公众+机构2亿+，日活100万+ | 公众+机构2亿+，日活100万+ | 公众+机构2亿+，日活100万+ |
| 产品卖点 | **工商信息：**法定代表人、股东名录、股权结构、营业状态等全国各地工商信息实时同步  **知识产权：**企业的无形资产：商标、专利、软件著作、网站备案  **老板关系：**老板所有公司、风险人际间传递、商业六度人脉  **信用报告：**300种维度看清公司和人的信用状况。信用报告-基础班、信用报告-专业版、个人投资任职报告，三种信用报告任选  **企业认证：**一处认证，多出展示，花小钱就能做打广告。  **问大家：**到处打听，比如知情人一句话。对哪家公司感兴趣？来这里提问吧！ | **一站式查询：**全国工商信息、诉讼信息、知识产权信息等一站式免费查询  **企业认证：**完善企业信息，进行品牌推广  ，实时获取企业和人员的信息变动  **投融资：**30万+创业公司  **法律风险：**失信、被执行、裁判文书  **千寻地图：**找到目标区域企业名录 | **调查企业：**企业工商经营信息  **预防风险：**失信被执行人信息  **找新客户：**地址电话一键导出  **关联关系：**投资持股多层追溯 |
| 营销策略 | **网络推广：**百度等搜索引擎、一些主流网站  **线下推广：**地铁、商场等公共场所的墙体广告 | **网络推广：**百度等搜索引擎、一些主流网站  **线下推广：**地铁、商场等公共场所的墙体广告 | **网络推广：**百度等搜索引擎、一些主流网站  **线下推广：**地铁、商场等公共场所的墙体广告 |
| 盈利模式 | 针对C端用户开启会员服务  依靠B端盈利，一找客户需求将平台数据进行分类整合，随后售卖给投资机构、企业等 | 针对C端用户开启会员服务  依靠B端盈利，一找客户需求将平台数据进行分类整合，随后售卖给投资机构、企业等 | 针对C端用户开启会员服务  依靠B端盈利，一找客户需求将平台数据进行分类整合，随后售卖给投资机构、企业等 |
| 定价方式 | **非会员：**浏览基础信息，限制次数，免费  **一般会员：**部分功能，可查看可下载，限制次数，基础版报告10份/天，  **VIP会员：**部分限制，其他不限次，1年360元，2年720元，3年720元  **企业会员：**主从账户，权益同VIP，298元/人/年，至少10人 | **非会员：**浏览基础信息，限制次数，免费  **一般会员：**部分功能，可查看可下载，限制次数，基础版报告10份/天，信用决策报告15元/份  **VIP会员：**部分限制，其他不限次，1年360元，2年540元，3年720元  **企业会员：**主从账户，权益同VIP有差异，288元/人/年，至少10人 | **非会员：**浏览基础信息，限制次数，免费  **一般会员：**部分功能，可查看可下载，限制次数，基础版报告10份/天，增值版报告15元/份  **VIP会员：**部分限制，其他不限次，1年366元，支持3人以上的团购  **企业会员：**主从账户，权益与VIP有差异 |
| 成立时间 | 2014-10-30 | 2014-03-12 | 2014-05-28 |
| 平台背景 | 国家中小企业发展基金旗下  官方备案企业征信机构 | 官方备案企业查询机构 | 官方备案企业查询机构 |
| 运营主体 | 北京天眼查科技有限公司  北京金堤科技有限公司子公司，100%控股 | 苏州朗动网络科技公司 | 苏州贝尔塔数据有限公司  上海合合信息科技发展有限公司全资子公司（名片全能王、扫描全能王） |
| 产品优势 | 1、国家中小企业发展基金旗下，宣传效果好  2、产品定位清晰：商业安全工具  3、卖点定位清晰：有针对性  4、宏观、微观比较全面 | 1、定价比较合理  2、产品定位清晰：查企业、查老板、查关系  3、卖点定位清晰：有针对性  4、宏观、微观比较全面  5、界面体验较好 | 1、产品定位清晰：AI商业调查工具  2、宏观、微观比较全面  3、界面体验较好 |
| 产品不足 | 1、全球企业：无即即时查询功能  2、界面体验：有待优化和提升 | 1、功能聚合有待加强 | 1、定价策略不够清晰（企业会员）  2、产品卖点不够突出、差异化 |

### 7.2.3竞争分析

“e企查”不是普通的企业信息查询平台。目前商业查询市场上的网站或平台，都是通过汇总国家公开信息平台，例如签过企业信用信息公示系统、中国裁判文书网、中国执行信息公开网、国家知识产权局、版权局这些平台。这些商业查询平台，能够提供详尽的企业信息特征，并能够从多个角度深入挖掘内联关系。

然而企业维度的丰富，也给数据筛选、清洗造成了较大的压力。例如金融信贷机构利用这些信息对企业的信用状况、还贷能力进行评估的时候，需要衡量每个属性对于企业评估的影响程度，并给出一个简单直观的初步结果；此外，针对一个企业信息的查询，能够详细的呈现，然而企业间的相似性关联，传统的商业查询平台，表现仍有不足。通过无监督学习算法，挖掘企业间关联，对企业进行分类贴标签，利用相似性挖掘潜在客群，精准推荐，才能使企业大数据，发挥它更大的价值。

**针对上述分析，总结出传统商业查询平台的不足，“e企查”将主要经营目标立足于**：

* 分主题的对企业特征进行评分，最终进行综合评价，并根据主题评分、综合实力等多个要素，进行无监督聚类，构建企业画像
* 根据企业的相似性，实现企业智能推荐，挖掘潜在客群；
* 评估文档、图表等实现企业数据直观辅助性评价，上新手能快速上手，直接了解一个企业的情况。

以下是对“e企查”经营立足点进行的竞争力分析：

#### 7.2.3.1天眼查

作为首屈一指的商业查询网站的龙头企业，主要支持查公司、老板和关系。

天眼查的查询结果返回有企业的基本信息、品牌介绍、融资历程等，这些都是对企业信息的一种汇集，具有参考和推广意义的竞品信息部分，也只是根据产品标签，根据成立日期进行降序排列。

* 优点分析：能够较为全面的展示企业的基本信息、成员、新闻、社交服务等，对于大型企业，能够呈现其公司、人员之间的投资、权利关系，也可以生成报告，详细评估，适合用于针对性研究某个企业详细信息的人员使用。
* 缺点分析：这样的资讯更多属于普适性范围，更多的考虑是一种企业信息全面的呈现，由于信息过于丰富，缺少针对当前解决金融信贷难题中，“信用评估”这一项的解决方案。对于企业综合实力的评估，仍需要人工去淘到自己想要的信息。

#### 7.2.3.2企查查

通过查询和使用，相比天眼查，它还增加了查风险和查全球属性。但在实际使用后发现，查风险实际上在其他平台中，就是查询类似司法纠纷和行政处罚这块，只是在天眼查中都被合并了而已。

* 优点分析：能够提供详尽的企业数据，并能够针对人们最关心的风险模块单独整理出来，方便查询；此外也能将企业、人员之间的关系以各种关系图的形式列出，直观明了。
* 缺点分析：同天眼查一样，提供了详细丰富的数据，仍然需要人工的对数据进行加工，生成直观的风险评估；即使企查查单列出了风险信息，数据也具有广泛性和专业性，对于刚上手的分析者而言，具有较大距离感和阅读难度，需要更好的数据建模，以定量的方式呈现数据，评估风险。

#### 7.2.3.3 启信宝

启信宝的UI设计较为简洁大方，此外，单独列出一个1分钟知企业的模块，尽可能的做到将复杂的企业数据，以简洁的形式呈现。此外也和其他平台一样，将多种特征分出多个主题来进行分类。

* 优点分析：有“一分钟知企业”功能，可以简洁明了概括企业的经营、舆论、风险信息等，降低评估压力；UI简洁明了，具有辨识度，能够清楚用户需要关心企业，将报告和监控列在较为醒目的位置，更好的评估企业的变化。
* 缺点分析：呈现的企业数据不如前两个平台丰富，也没有在查询后的一级界面，显示各种企业关系等；使用体验相对较差，在点击拓展功能，例如快速了解，报告、图表等，都需要反复的输入验证信息；和之前平台一样，缺少企业内部属性的关联，没有做到相类似企业的推荐等。

## 7.3管理体系

“e企查”为了早日服务企业和金融信贷机构，切实为企业带来方便，公司成立了市场部专门负责产品商业运作，让更多的人了解产品、购买产品，为公司后续发展奠定良好基础，更好助力“一贷通”事业。

## 7.4成本分析

### 7.4.1人员成本

表7-2：系统研发人员成本

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 研发阶段 | 周期/天 | 人员/名 | 工作量/人月 | 月工资/元 | 总费用/元 |
| 需求调研 | 10 | 2 | 1 | 5,000.00 | 5,000.00 |
| 开发阶段 | 60 | 3 | 8 | 6,000.00 | 48,000.00 |
| 测试阶段 | 15 | 3 | 4 | 5,000.00 | 20,000.00 |
| 售后阶段 | 50,000.00 | | | | |
| 人员成本总计 | | 123,000.00 | | | |

如表7-2所示，系统开发周期为60天，需要开发人员3人，根据软件系统的规模估算（代码量大概在20000行左右），开发工作量约为8人月，月工资6000元计算，则开发费用为48000元;系统测试需要3名调试员，系统测试周期为15天，调试阶段的工作量为4人月，月工资5000元，若按进度完成项目则调试费用为20000元。系统研发总费用为68000元。

此外在产品售出后，需为客户提供售后服务，在此必须考虑售后人员的工资薪酬并将其计入总人员成本内。

### 7.4.2设备成本

表7-3：设备成本

|  |  |  |  |
| --- | --- | --- | --- |
| 设备名称 | 数量 | 单价/元 | 总计/元 |
| 服务器 | 1 | 30,000.00 | 30,000.00 |
| 交换机 | 1 | 15,000.00 | 15,000.00 |
| 研发电脑 | 5 | 5,000.00 | 25,000.00 |
| 测试设备 | 20,000.00 | | |
| 其他设备 | 5,000.00 | | |
| 硬件损耗 | 5,000.00 | | |
| 总计 | 100,000.00 | | |

在系统开发过程中，设备的购置、租用及硬件的损耗等的支出；且本系统涉及网络的构建和的硬件链接，因此在系统研发和测试还需购置相应的硬件设施，详见表7-3。

### 7.4.3营销推广成本

在系统成功开发后，将其推向市场，提高市场认可度，促使更多的潜在客户购买产品由此产生的费用，预计为80000元。

### 7.4.4风险成本

公司应对可能存在的风险的而在系统开发时预估的费用，此部分主要用来应对风险监管和风险发生后处理解决问题，预计50000元。

### 7.4.5学习成本

公司在进行项目开发前必须对所有参与研发的人员进行相关软硬件的培训，需要为他们购置必要的学习资料、技术参数，预计15000元。

### 7.4.6售后成本

在产品成功售出后，公司会对客户进行优质的售后服务，在售后过程中产生的费用，主要指系统的后续升级，跟踪服务等，预计50000元。

### 7.4.7其他成本

系统在研发，营销使用过程中差旅通信费用、耗材损失费用、水电费用和场地租用等费用，详见表7-4。

表7-4：项目其他成本

|  |  |
| --- | --- |
| 项目名称 | 费用金额 |
| 差旅通信成本 | 6,000.00 |
| 耗材成本 | 4,000.00 |
| 水电、场地成本 | 5,000.00 |
| 总计 | 15,000.00 |

### 7.4.8成本总计

根据以上成本分析并考虑适当的风险成本。公司市场部会同技术部制定了本次系统开发的成本预算（见表7-5）。各部门在进行系统研发和推广时需严格按照预算进行，如确实需要追加经费，必须向项目经理提出申请，待批准后方可实施。

表7-5：成本预算表

|  |  |
| --- | --- |
| 成本分类 | 预算金额/元 |
| 人员成本 | 123,000.00 |
| 设备成本 | 100,000.00 |
| 营销推广成本 | 80,000.00 |
| 学习成本 | 15,000.00 |
| 其他成本 | 15,000.00 |
| 售后成本 | 50,000.00 |
| 风险成本 | 50,000.00 |
| 成本总计 | 433,000.00 |

## 7.5成本控制体系

有效控制成本，是公司降低成本、提高经济效益的重要途径。经济效益是在市场经济条件下企业生产运营所围绕的中心点，而有效控制成本费用是能够提高经济效益的有力措施。公司成本控制拟分为三个时间段：成立初期——“保盈利”为主；发展期——“增收入”为主；成熟期——“精益化”为主。

### 7.5.1保盈利期

成立初期，公司固定成本较高，例如基本建设成本、技术成本，而同时所获得的等收入较少；另外，企业面临各方面的偿债压力。因此，公司应采用本量利成本分析法，对企业成本加以控制。

本量利分析法是在成本性态分析和变动成本法的基础上发展起来的，主要研究成本、销售数量、价格和利润之间的数量关系的方法。它是公司进行预测、决策、计划和控制等经营活动的重要工具，也是管理会计的一项基础内容。

### 7.5.2增收入期

公司进入发展期，生存压力缓解，但是面对管理的疲软，需要减少变动成本，提高毛利润。所以，对于此阶段的公司发展，应该采用标准成本法，加重对公司预算的管理，通过预算来发现制约公司发展的瓶颈，这点对于尚处在发展阶段未成熟期的信息技术公司尤其重要。

标准成本法是西方管理会计的重要组成部分。是指以预先制定的标准成本为基础，用标准成本与实际成本进行比较，核算和分析成本差异的一种产品成本计算方法，也是加强成本控制、评价经济业绩的一种成本控制制度。

### 7.5.3精益化期

成熟期的公司无论在技术方面还是客户群方面，都有自己良好的渠道，此时对公司而言，制约公司发展的是全方位的综合的成本控制。如何使公司在收入高速增长的同时仍保持净利润的增长，需要全面成本控制。

全面成本控制是指对企业生产经营所有过程中发生的全部成本、成本形成中的全过程、企业内所有员工参与的成本控制。公司围绕财富最大化这一目标，根据自身的具体实际和特点，建立管理信息系统和成本控制模式，确定以成本控制方法、管理重点、组织结构、管理风格、奖惩办法等相结合的全面成本控制体系，实施目标管理与科学管理相结合的全面成本控制制度。

## 7.6商业运行

“e企查”针对目前商业查询平台企业在企业画像构建方面能力的功能和相似企业挖掘与评估方面的缺失，搭建了这样的平台。该平台拥有核心无监督算法保证分类准确性，同时也能根据算法精准推荐类似企业。可以通过：

* 拓展搜索服务；
* 拓展可视化功能；
* 拓展推荐功能；
* 拓展辅助评估模块；

作为盈利重点，具有广阔的商业价值。

### 7.6.1客户划分

主要客户：金融信贷机构、企业用户。

### 7.6.2产品推广

为了在最快时间内让最多的人了解到产品，提高产品认知，促使潜在客户购买产品，公司市场部在对产品的目标受众进行了细致分析后制定了四大推广策略：

1. 人员推销；
2. 公关推广；
3. 广告推广；
4. 销售促进。

### 7.6.3营销模式

公司在经过调研之后，清楚地了解市场需求和现有市面产品的不足，科学开发了企业无监督分类系统“e企查”，轻松实现企业分类，准确构建企业画像。同时也可以利用聚类算法实现精准推荐，既降低企业大数据公司数据维护压力的同时，也降低金融信贷机构企业数据审核压力，同时挖掘潜在客群，加速金融市场资金流动，助力“一贷通”事业。

* 产品策略：差异化策略，个性化策略
* 价格策略：统一定价策略，剩余定价策略
* 其他营销策略：深度接触，缩短周期；免费体验，占领市场；先进文化，打造品牌。

## 7.7资源计划

### 7.7.1软硬件资源计划

软硬件资源需求见表7-6。

表7-6 软硬件资源计划表

|  |  |  |
| --- | --- | --- |
| 资源名称 | 需要数量 | 要求 |
| MySQL | 1 | MySQL 5.1 |
| Web服务器 | 1 | 华为云服务器，单核CPU，2G内存 |
| 数据库服务器 | 1 | 华为云服务器，单核CPU，2G内存 |

### 7.7.2人力资源计划

人力资源计划见表7-7。

表7-7：人力资源计划表

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 角色 | 职责 | 技能要求 | 数量 | 人员分配 |
| 项目经理 | 负责项目进度、成本、质量、风险，进行项目计划监控工作。 | 项目管理技能 | 3 | 小鲍  小陆  小阳 |
| 产品人员 | 负责调研市场，提出产品的需求，文档编写。 | 需求管理、市场调研 | 3 | 小陈  小陆  小阳 |
| 设计人员 | 负责按照需求制定平台架构、前端布局、数据库设计、算法设计 | 软件开发技能 | 3 | 小鲍  小程  小陈 |
| 编码人员 | 负责按照需求，设计进行代码的编写，进行单元测试、代码评审、集成测试。 | 软件开发技能 | 2 | 小程  小鲍 |
| 测试人员 | 负责按照需求和使用说明对系统进行黑白盒测试 | 软件测试用例编写 | 3 | 全部 |
| 实施人员 | 负责编写用户手册，将系统部署给用户，并指导用户使用。 | 软件实施技能 | 2 | 小鲍  小程 |
| 运维人员 | 负责项目的环境配置、制定配置管理计划、配置说明；负责系统集成、发布并保持记录。 | 服务器管理技能 | 1 | 小鲍 |

### 7.7.3培训计划

人员培训计划见表7-8。

表7-8：培训计划表

|  |  |  |
| --- | --- | --- |
| 培训课程 | 培训内容 | 培训时间 |
| 业务分析 | 分析命题，平台定位以及创新亮点 | 2020.1.18 |
| 机器学习 | 数据分析方法，无监督模型的原理及使用 | 2020.2.1 |
| 架构技术 | 本平台前后端技术的使用 | 2020.2.15 |
| 平台使用 | 学习平台各个功能的使用方法 | 2020.4.6 |

## 7.8相关负责人

商业计划负责人见表7-8。

表7-8：商业计划负责人

|  |  |  |
| --- | --- | --- |
| 角色名称 | 角色 | 职责 |
| 小陈 | 主要负责人 | 市场分析、预测、产品定位、商业运行制定 |
| 小鲍 | 参与人员 | 监督商业计划，成本体系构建、资源管理 |
| 小程 | 参与人员 | 提供技术参数支持，产品定位制定 |
| 小陆 | 参与人言 | 协助市场分析和竞品调研部分 |
| 小阳 | 参与人员 | 协助商业运行指定和产品定位部分 |

# 第八章 可行性分析

## 8.1技术可行性分析

公司聘请资深机器学习专家郑建炜教授作为公司顾问，在充分了解系统需求、机器学习模型的搭建方式以及平台实施方式后，郑教授认为公司实力完全可以独立完成系统开发，在理论和技术方面不成问题。

## 8.2运营可行性分析

郑教授认真分析了公司的商业运作模式，对其进行全面评估后，认为商业运作成熟管理体系完备，项目收益巨大，前景广阔。

## 8.3组织可行性分析

“e企查”制定了合理的项目实施进度，组织麻雀虽小，五脏俱全，且团队成员具有丰富的技术基础和合作经验，团队分工明确，也能够保证定期展开讨论，了解彼此工作进展和技术路线，能够保证项目按部就班进行。

## 8.4市场可行性分析

目前，市面上的商业查询系统，主要定位在于汇聚各大公开平台企业的数据信息，主要体现的是大而全，并且能够对企业、管理人员的关系进行详细地梳理。然而由于平台数据的专业性，维度丰富，对于金融信贷机构在评估企业信用水平、还贷能力时，需要人工筛选特征，并建模进行综合评估，需要大量的人力和时间投入，导致了金融放贷的效率降低。此外，传统的商业查询平台，对于企业的分类，大多只以行业门类进行区分，忽视了企业特征间内部的关联性，在企业分类和客群精准推荐方面，仍然表现不足。

在平台构建前期，“e企查”有限公司做了商业查询市场的充分调研，清楚地了解了市场情况和用户需求，为项目成功运行、快速打开市场奠定了坚实的基础。

综上所述，“e企查”开发此项目技术完备、成熟，经济效益巨大，直击金融信贷机构和企业在信用评估审核过程中的痛点，快速准确对企业主体进行分类，降低人工成本及出错率，对后期企业或金融信贷机构对企业进行客群分析、精准获客等应用。

## 8.5投资可行性分析

### 8.5.1经费预算

如7.4.8节所示，本项目申请资金总计为433,000元。

### 8.5.2预期经济效益

* 直接效益
* 通过聚类算法对企业进行分类，方便金融信贷机构对企业进行分类，更好实现客群分析，精准获客。
* 对企业数据进行特征分类，建立合适且灵活的评估模型，直观体现企业信贷水平，有助于降低金融信贷机构的审核压力，降低人工成本。
* 降低企业大数据管理者维护压力，提高公司服务响应速度与效率，在激烈的市场竞争中取得先机，推广公司文化品牌，建立公司市场形象。
* 间接效益
* 通过客户的口碑传播，推广平台品牌，挖掘出更多的潜在客户，创造更大的商业利润。
* 有效增强客户黏性，为平台带来更多客户。
* 建立一个以平台为核心的评估服务系统，为公司带来更大利益。

### 8.5.3相关参数计算

* 收益/投资比
* 每年内我们都为公司节约数据维护开支，以及用户使用本平台获取的利润，一年直接和间接收益为500,000元，而本项目申请资金为433,000元。因此收益投资比=收益/投资≈1.155
* 由此可得，资本周转速度较快，盈利可观
* 净现值（NPV）
* 假设本项目年收入为元，正常开销每年元，折现率为，初始投资元，则根据,得到NPV=2,867,000元，远大于0，计算期内盈利能力很好，投资方案可行。
* 根据现金流量表计算内含报酬率（IRR）如下：



当NPV等于0时，计算前5年的内部报酬率，此时的IRR即为内部报酬率。计算得到IRR=48%，高于资金成本率10%，因而该项目在5年内市场增长性很好。

* 回收期
* 通过净现金流量、折现率、投资额等数据用插值法计算，公式为：

回收期=累计净现值出现正值年数-1+（未回收现金/当年现值）

计算得投资回收期为2年。

## 8.6相关负责人

可行性分析负责人见表8-1。

表8-1：可行性分析负责人

|  |  |  |
| --- | --- | --- |
| 角色名称 | 角色 | 具体职责 |
| 小陈 | 主要负责人 | 市场分析，可行性分析撰写 |
| 小鲍 | 参与人员 | 投资可行性分析，撰写可行性分析部分 |
| 小程 | 参与人员 | 参与讨论，提供技术分析 |
| 小陆 | 参与人员 | 协助完成市场可行性分析 |
| 小阳 | 参与人员 | 协助完成投资可行性分析 |

# 第九章 风险计划

## 9.1 项目风险分析

能预见可能影响项目进度或正在开发的软件产品质量的风险，并采取行动避免这些风险，是项目管理者的一项重要计划。风险分析的结果以及对该风险发生时的后果分析都应该在项目计划中记录在案。有效的风险管理能使我们从容面对问题，避免这些风险带来无法承受的开支或进度失控。如表9-1所示，是本系统的风险管理流程。



图9-1：风险管理流程

### 9.1.1相关性风险

在软件开发的过程中，很多软件项目的风险都是与外部环境的相关。在项目开发的过程中，这种与环境的关联性一般都不能够得到有效的控制。软件项目与外部环境相关的主要风险如下：

* 队员生病，延误开发
* 队员间的沟通矛盾导致的凝聚力不足，延误开发
* 不可抗力伤害

### 9.1.2需求风险

绝大多数的软件项目在需求阶段都面临着许多不确定性和未知性，甚至会造成混乱。如果在软件项目开发的早期轻视了软件项目多方面的不确定性,在项目开发过程中得不到解决,这些问题就会对软件项目的结果产生巨大的危害，甚至会造成整个项目的失败。如果在需求分析阶段不控制与需求相关的风险因素,那么就极有可能性产生不合格软件产品，甚至可能产生错误的产品。

* 对赛题解读有误造成的需求分析偏差
* 对目标客群的实际需求理解存在分析偏差或者没能很好达到
* 平台方案设计不合理，不符合多数用户使用习惯
* 需求变化导致新的市场产生
* 需求分析过程中，缺乏有效的对需求不断变化的管理过程

### 9.1.3市场风险

* 与市场上的成熟软件存在定位重叠，导致的竞争力不足
* 商业计划的制定与实施存在不确定因素
* 市场用人、硬件、推广等成本的变化导致公司无法正常盈利

### 9.1.4技术风险

目前，计算机软件技术的发展是日新月异，在项目的开发过程中，可能会由于开发成员缺乏对所使用的软件技术深入的了解和使用，可能会导致整个项目的失败。引起软件技术风险的主要原因如下：

* 信息存储和传输风险
* 开发人员和企业人员缺乏培训
* 人力资源有限
* 项目组成员不能将全部时间投入开发
* 开发人员经验不足
* 部分特殊功能实现难度较大，无法及时交付
* 缺乏质量跟踪
* 测试方式不合理
* 低估系统缺陷修复率的风险
* 部门之间设计API对接部分，沟通不明确造成的开发效率低下

### 9.1.5开发风险

通常是指潜在的维护、验证、接口、实现以及设计等环节出现的问题，存在技术空白及未知领域，为软件开发工作带来较大的风险。

* 数据库结构设计不合理造成的开发问题
* 服务器环境不稳定
* 设计工具不合用
* 系统崩溃以及备份环境不稳定

### 9.1.6知识产权风险

* 由于网络环境的公开性、开放性，本系统开发的知识产权将会受到侵害。所以保障平台的知识产权将会是本公司必须要重视的环节。对于侵权事件，平台方会采用法律武器，维护自身合法权益。在平台发展大道一定规模时，会聘请专业的律师团队作为法律顾问，全权处理公司法律问题。

### 9.1.7管理风险

软件项目管理的任务是在整个软件设计和实现的过程中都能够在软件项目管理者的控制之下，使软件能够按照所预定的成本、在规定的时间内交付给用户使用。在软件项目开发的过程中最难预料和控制的是“人”的因素，包括开发人员和客户，与项目管理相关的风险因素有：

* 客户的需求经常发生变化
* 开发人员之间欠缺沟通
* 项目成员的责任心不强
* 计划时间线安排不当和不合理

### 9.1.8财务风险

软件开发行业涉及到各行各业的企业，我国软件开发企业也逐步呈现出信息化、集约化的发展趋势，并且软件开发企业属于技术密集型和资金密集型企业，软件开发企业前期需投入大量的研发资金，因此对其资金的流动性要求较高。

* 筹资风险：

由于平台受经营现状和科学技术条件的限制，平台若想实现持续健康发展，其务必对创新研发成果进行持续性的资金投入。软件研发投入需较长的周期才可收回资金，在此过程中很容易导致软件开发企业的资金链断裂，不利于其持续健康发展，进而引发其财务危机。

* 资金回收风险：

平台回收资金的渠道是将软件投入市场，逐步收回成本并赚取利润。倘若软件开发企业在软件研发过程中，无法科学控制其成本，盲目的加大资金投入力度，提高研发效率很容易导致资金链锻炼，最终导致其投入的项目无法实现资金回收。

* 财务管理风险：

随着研发项目的不断开展，软件研发企业的业务范围不断拓展，但是其财务风险管理方式较为单一，尚未对此实现及时更新。对于企业的内部员工而言，员工需对各个业务部门的工作内容和工作流程进行掌握，员工在观看操作手册时，发现操作手册和实际情况有所出入，员工很容易办错业务，进而引发企业的操作风险，加大财务风险。

### 9.1.9风险因素评价

利用风险坐标图法对本项目风险因素进行评估，结果见表9-1：

表9-1：风险因素评级表

|  |  |  |  |
| --- | --- | --- | --- |
| 编号 | 风险因素 | 可能性等级 | 影响程度 |
| 1 | 相关性风险 | 低 | 低 |
| 2 | 需求风险 | 低 | 高 |
| 3 | 市场风险 | 高 | 高 |
| 4 | 技术风险 | 高 | 高 |
| 5 | 开发风险 | 中等 | 中等 |
| 6 | 知识产权风险 | 低 | 低 |
| 7 | 管理风险 | 低 | 低 |
| 8 | 财务风险 | 中等 | 低 |

风险矩阵图译自Risk Matrix, 是一种有效的风险管理工具。可应用于分析项目的潜在风险，也可以分析采取某种方法的潜在风险。

如图9-2所示，风险矩阵可以划分成A、B、C、D、E五个区域。

* 如潜在问题在红色区域，则应该不惜成本阻止其发生，（如果成本大于可接受范围，则放弃该项目）。
* 如潜在问题在橘红色区域，应安排合理的费用来阻止其发生。
* 如潜在问题在黄色区域，应采取一些合理的步骤来阻止发生或尽可能降低其发生后造成的影响。
* 绿色区域，准备应急计划，该部分的问题是反应型，即发生后再采取措施，而前三类则是预防型。

如表9-2所示，是本系统的风险矩阵。

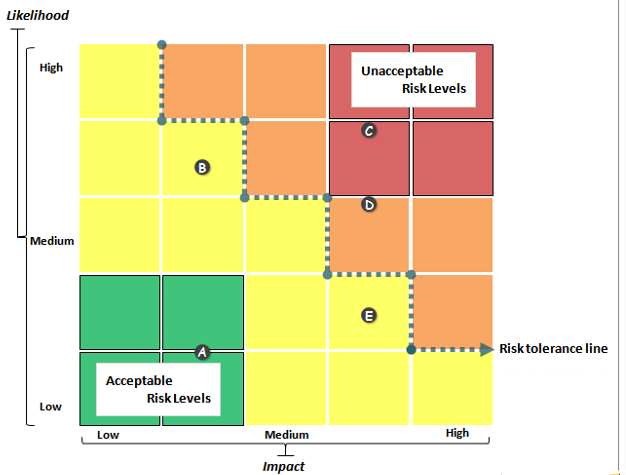


图9-2：风险矩阵图

## 9.2风险监控

### 9.2.1管理体系

公司积极应对可能的风险，力争将风险造成的损失降到最低，为此公司成立专门应对本次系统开发风险的风险监督组，分工及职责见表9-2：

表9-2：风险监督组分工表：

|  |  |  |
| --- | --- | --- |
| 监督组成员 | 角色 | 职责 |
| 小鲍 | 监督组组长 | 统筹项目风险管理，主持风险决策 |
| 小陈 | 组员 | 风险计划起草，提供建议支持 |
| 小程 | 组员 | 根据项目实际，提出项目可能面临的技术风险 |
| 小陆 | 组员 | 根据风险决策，制定相应规避措施 |
| 小阳 | 组员 | 根据风险决策，制定相应风险可能带来的结果 |

### 9.2.2风险监控

风险监控就是要对每一个识别的风险定期进行评估，从而确定风险出现的可能性是变大还是变小、风险的影响后果是否有所改变。当然，这些信息不可能直接通过观察获得，必须关注能提供有关风险可能性及其影响后果信息的其他因素。显然，具体是哪些因素取决于风险的类型。具体风险监控流程见表9-3。

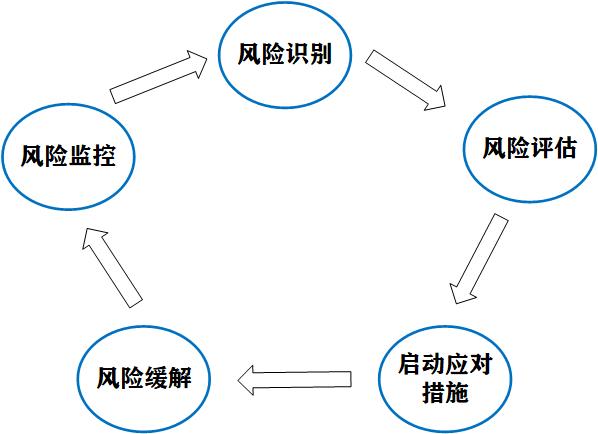


图9-3：风险监控流程

### 9.2.3规避措施

针对上述问题，公司制定了科学有效的风险规避措施，尽一切可能将风险降到最低。

* 相关性风险

由于开发时间正处于疫情期间，队员间无法直接在线下进行交流。所以团队通过定期组织视频会议的形式，了解开发成员的开发情况以及身心状况，做好技术指导的同时，也做好人文关怀，提升团队凝聚力。

* 需求风险

为了避免出现赛题解读失误而导致的需求严重偏差。成员反复组织多次讨论赛题含义，并请多位指导教师对此进行解读，深入调研市面上竞品的优劣势，把握“e企查”所应拥有的行业定位和业务范围。

在开发的过程中，也时刻与需求方保持联系，及时调整开发策略，紧跟市场需求，做最适合用户的企业无监督分类平台。

* 市场风险

由于本公司的业务是服务外包，所以在自己进行深入调研行业同类产品的同时，也及时与甲方公司沟通行业动向，了解行业前沿，根据市场实际情况，及时调整商业计划，并与行业内部人员多进行交流。

此外，市场风险的解决，涉及到多个部门，例如运营、广告、营销、管理等，公司在发展到一定规模时，也需要成立自己的相关团队，更好的解决上述风险。

* 技术风险

聘请有着丰富开发经验和算法理论研究的专家担任公司顾问，为公司提供必要的技术建议，此外部门内部也要及时的展开培训，进行内部学习，及时了解较为先进的开发技术，提高开发效率，降低技术风险。

* 开发风险

由于一个系统或者平台的开发涉及到多个部门间的合作，特别是前端、后端、算法、服务器在集成的过程中，有可能会因为环境、版本、沟通等，产生使用问题。因此公司内部制定严格的开发说明文档和开发标准，确保在跨部门合作时，能够保证高效，降低错误率和风险。

* 知识产权风险

为避免发生侵占知识产权的事情发生，公司将根据项目的实际开发状况适时申请知识产权专利，将该风险降到最低。

* 管理风险

在平台进行开发前，明确每位成员的责任与义务，并及时根据实际情况，调整开发进度。在开发之余，也要及时关心组内成员的身心状况，打造一支具有凝聚力的团队。

* 财务风险

为降低财务风险，公司随时关注财务状况，准确估计项目的投资规模和进度，及时筹备资金，并及时根据项目的所处阶段，制定相应的财务计划；同时向投资者展示公司的盈利能力，并向投资者说明公司未来的发展能力和预估的市场占有率，鼓励投资者投资；在进行利润分时，及时向同类的成熟企业借鉴经验，保证利益分配合理、公平，同时为未来公司的发展奠定基础。

## 9.3风险应对预案

本项目组采用SEI（软件工程研究所）的CRM（持续风险管理）模型来对风险进行分析和解决。CRM模型的特点是持续风险管理，强调对风险的沟通，复杂度低，核心活动是通讯。该模型有7个软件风险管理原则。

CRM软件风险管理原则具体内容，如图9-4所示：

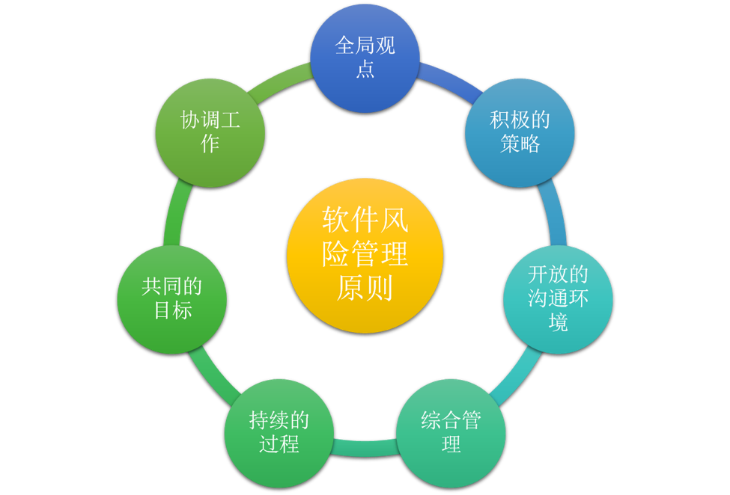


图9-4：CRM软件风险管理原则

CRM模型要求在项目生命周期的所有阶段都关注风险识别和管理，它将风险管理划分为5个步骤。风险管理模型CRM步骤，如图9-5所示：

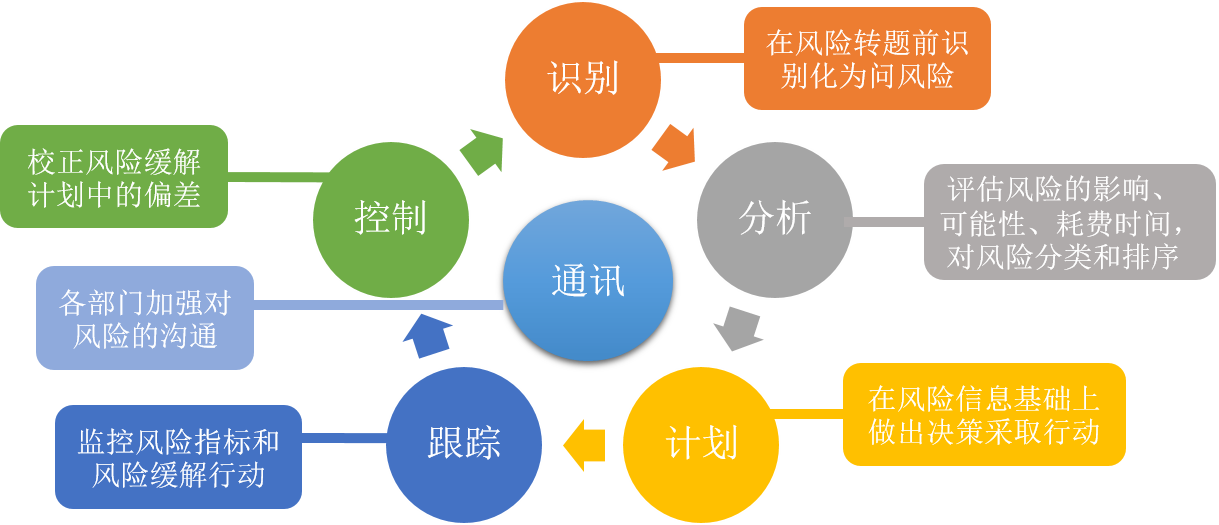


图9-5：风险预案CRM步骤

### 9.3.1技术风险预案

* 系统开发出现设计背离或是混乱时，首先停止工作进度进行内部自查，其次以最快的速度请顾问和公司工作人员进行项目评估，纠正错误。
* 项目进展缓慢，延误交付，公司全员加班；如确实有需要在征求客户意见后可聘请其他技术人员完成项目；进入项目违约处理机制。
* 产品在使用过程中发现未知错误，公司立即组建技术解决组，第一时间解决产品的使用问题，并考虑对客户进行合理赔偿，维护公司声誉。

### 9.3.2市场风险预案

* 市场进入障碍估计不足，首先进行未知障碍定位预估，对市场进行二次定位，调整公司营销侧重点；必要时重新制定《商业计划》。
* 客户分析不足，市场部客户关系组与客户积极进行沟通，深入了解客户需求及反馈；调整公司营销模式和推广计划，对系统进行适当调整升级。
* 商业计划执行不到位，首先进行内部自查，调离或解除部分工作人员，进行人员二次配置；其次对全体员工进行执行力培训，提高市场部全员的意识和执行力；与此同时市场部积极应对降低损失。

## 9.4退出机制

在我国由于实行市场经济时期较短，资本证券市场还处在较低的发展水平。考虑到我国目前的经济政策和资本市场的情况，以及估计公司未来可能出现的经营状况，风险投资退出公司的可能方式有以下几种：IPO上市、兼并与收购、管理层收购、破产清算。

### 9.4.1 IPO上市

当公司在未来的发展中，在公司得到了发展、壮大，产品占领了相当的市场份额，成长为一个相对成熟的公司时，在资金、管理层战略、公司规模、发展前景、市场形势、政策法规等方面都满足公开上市的要求时，本公司将公开上市，发行自己的股票，这将是一种收益可观的退出方式。此时，风险投资退出公司，采用首次公开上市方式，对于本公司而言，不仅可以保持公司的独立性，而且还可以获得在证券市场上持续融资的渠道；而对于风险投资者来说，则可以获得非常高的投资回报。

### 9.4.2兼并与收购

并购退出是指通过其他公司兼并或收购本公司从而使风险资本退出。在公司未来的发展过程中，由于股票上市及股票升值需要一定的时间，或者因为市场竞争的激烈使公司的市场占有率较低，公司难以达到首次公开上市的标准，那么风险投资者将采用股权转让的方式退出投资。

### 9.4.3管理层收购

在公司未来的发展过程中，当公司成功渡过了技术风险、经济风险和管理风险等，已经成长为一个有发展潜力的中型公司后（预估期为5年），风险投资以股份回购和二次收购的方式退出。主要是公司回购风险投资者的股份和风险公司回购风险投资者的股份，此时，公司的大部分股权将在管理层手中，利于公司的稳定发展。

### 9.4.4破产清算

当公司在未来的发展中，出现重大的经营决策失误或者市场竞争急剧恶化使得公司经营状况不好且难以扭转时，公司将申请破产清算，风险投资者撤回资金，以减少损失。清算的方式退出是企业建设者和经营者都不愿意的选择的，但是在企业经营状况恶化的情况下必须断然采用措施。

## 9.5相关负责人

风险计划相关责任人见表9-5。

表9-5 风险计划负责人

|  |  |  |
| --- | --- | --- |
| 角色名称 | 角色 | 职责 |
| 小陈 | 主要负责人 | 撰写风险计划，商业风险处理 |
| 小程 | 参与人员 | 负责提供技术参数支持，技术风险监控、预案 |
| 小鲍 | 参与人员 | 负责起草标准制定以及退出机制部分 |
| 小陆 | 参与人员 | 流程图制作，并撰写规避风险部分 |
| 小阳 | 参与人员 | 分析潜在风险，并列出具体可能结果 |

# 第十章 总结与展望

“e企查”系统，致力于解决金融机构在对中小企业无监督分类问题，人工分类企业劳动力成本高、出错率高的痛点问题，助力“一贷通”事业。

* 技术总结与展望：

在算法部分我们通过比较，得到在多种无监督聚类算法中K-Means最优的结论，并以K-Means算法为基础，进行改进得到基于加权多层的K-Means的算法。该算法整体包含了等级化算法、模块分类降维、加权K-Means聚类。能够有效的解决在企业分类过程中，面临的数据缺失、度量不统一、数据量大等问题，并在CH、DB、SH以及训练、预测时间上相对传统K-Means有一定提升。

在系统架构部分，我们使用Vue+Django的前后端架构，并在数据传输使用restful风格，符合目前主流网站的技术路线。这样的数据交互模式，有利于减少系统的请求负担。在前端数据可视化方面，还使用到了Echarts，有助于用户直观了解企业数据。

在未来，后端数据交互加上hadoop技术，并在部署时加上uwsgi，提高系统并发处理能力，并再次基础上引入基于MapReduce的K-Means技术，对现有的模型性能进行进一步改进。

* 业务总结与展望

在业务方面，我们针对前台的查询使用用户，和后台的系统数据、机器学习模型的管理，针对性的开发了前台和后台，提升了使用者和管理者的使用体验。在数据的插入和查询方面，均引入了批量操作，还提供了丰富的可视化服务，有助于我们了解企业能力和数据分布。

在未来，我们希望能够开发系统的APP端，提升用户使用体验。同时能够引入爬虫系统，自主更新系统数据，解决数据缺失问题。还可以开通社区服务，加强用户黏度。

* 商业计划总结与展望

在商业计划部分，我们详细的分析了市场需求，对市场上现有的商业查询平台的业务范围、优劣势进行了分析，并针对金融机构和微小企业的痛点，提出了我们的经营目标。

根据这个经营目标，对企业的运行，不同时期成本分析、商业运行、资源分配等展开进行了描述。

在未来，我们希望伴随着企业的发展，我们不断地扩充系统功能，针对竞品的劣势，发展自身优势，拓宽用户群体范围，并不断完善“e企查”的成本分析、资源分配、商业计划，使其不断成长。

* 可行性分析总结与展望

在可行性分析中，我们详细的分析了本系统的技术、组织、市场、投资可行性，在投资可行性里，分析了经费预算、预期经济效益、以及相关参数计算，最终得出结果本项目可行性强，可以投入运营。

* 风险总结与分析：

在本部分中，我们分析了项目风险中的相关性风险、需求风险、市场风险、技术风险、开发风险、知识产权风险、管理风险、财务风险、风险因素评价。并做了风险监控和应对预案，以及相应的退出机制。

对于在公司运营中所可能遇到的风险，都做到了尽可能的分析和准备方案，能够有效的面对各类风险。