**表结构与属性说明**

数据预处理遇到缺失数据，处理方法

1. 删除该部分数据
2. 数据补齐(手动添加，映射到高维，添加特殊值【均值、众数、中位数、分位数】，插值法【数据模型处理:回归、贝叶斯】)

目前数据存储在如下表中，表结构具体信息在每个部分有具体说明。

1. 投资机构表(agency\_invest)---存储投资机构基本信息，以下表格是属性说明

投资机构编号没有添加应该没用,标注红色的为需要讨论的，蓝色是变动字段

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 计数序号 | 字段名称 | 数据类型 | 初始值(范围) | 字段描述 |
| 0 | id | Varchar(20) | 1 | 标识 |
| 1 | name | Varchar(100) | null | 投资机构名称 |
| 2 | telphone | Varchar(20) | null | 联系电话 |
| 3 | address | Varchar(150) | null | 投资机构办公地址 |
| 4 | website | Varchar(100) | null | 投资机构官方网址 |
| 5 | agency\_des | Varchar(500) | null | 投资机构简介 |
| 6 | inf\_from | Varchar(100) | 桔子链接 | 信息来源网址 |
| 7 | fund\_count | INT | 0 | 投资次数 |
| 8 | round\_rank | Varchar(150) | 天使轮… | 投资轮次 |
| 9 | invest\_area | Varchar(200) | 互联网 | 投资领域 |
| 10 | register\_capital | Varchar(100) | 0.0 | 注册资本 |
| 11 | create\_time | Date | 1970 | 创立时间 |
| 12 | register\_place | Varchar(100) | 北京 | 注册地址 |
| 13 | manager | Varchar(150) | null | 投资机构管理者 |
| 14 | stockholder | Varchar(200) | null | 主要股东 |

2.公司表(company)---存储需要融资的公司的基本信息，以下表格是属性说明

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 计数序号 | 字段名称 | 数据类型 | 初始值(范围) | 字段描述 |
| 0 | id | Varchar(20) | 1 | 标识 |
| 1 | name | Varchar(100) | null | 融资公司名称 |
| 2 | field\_type | Varchar(50) | null | 融资公司的类型(领域) |
| 3 | address | Varchar(150) | null | 融资公司办公地址 |
| 4 | found\_time | Date | 1970 | 公司成立时间 |
| 5 | register\_capital | Varchar(100) | 0.0 | 注册资本 |
| 6 | stockholder | Varchar(200) | null | 主要股东 |
| 7 | team | Varchar(400) | null | 团队情况 |
| 8 | round\_rank | Varchar(150) | 天使轮… | 融资轮次 |

3.融资信息表(fund\_inf)---存储融资项目、融资公司和投资机构等基本信息，以下表格是属性说明

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 计数序号 | 字段名称 | 数据类型 | 初始值(范围) | 字段描述 |
| 0 | id | Varchar(20) | 1 | 标识 |
| 1 | agency\_name | Varchar(100) | 1 | (投资方)投资机构表id |
| 2 | leaderorfollow | int | 0(领投)，1(跟投),2(混合数据,分不清谁是领投) | 投资机构方是否领投 |
| 3 | company\_name | Varchar(100) | 1 | (融资公司)融资公司表id |
| 4 | fund\_amount | Varchar(100) | 0.0 | 融资金额 |
| 5 | round\_rank | Varchar(50) | 天使轮 | 融资轮次 |
| 6 | fund\_time | Date | 1970-- | 融资时间 |
| 7 | inf\_from | Varchar(100) | 桔子链接 | 信息来源网址 |
| 8 | project | Varchar(200) | null | 融资项目 |

4.投资人表(investor):投资机构有多个投资人，投资人负责具体项目的投资，以下表格是属性说明

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 计数序号 | 字段名称 | 数据类型 | 初始值(范围) | 字段描述 |
| 0 | id | Varchar(20) | 1 | 标识 |
| 1 | agency\_name | Varchar(100) | 1 | (投资方)投资机构表id |
| 2 | education\_back | Varchar(50) | 初中毕业 | 投资人教育背景 |
| 3 | sex | tinyint | 0(女)，1(男) | 投资人性别 |
| 4 | work\_experience | Varchar(400) | xx年-xx年在xx公司担任xx | 工作经历 |
| 5 | age | int | 18--65 | 年龄 |
| 6 | position | Varchar(80) | 职位 | 投资人职位 |
| 7 | invest\_experience | Varchar(300) | 投资经历(xx年xx月投资项目金额) | 投资人的投资经历 |

5.创业者表(entrepreneur)，创业公司的主要发起者中的leader，以下表格是属性说明

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 计数序号 | 字段名称 | 数据类型 | 初始值(范围) | 字段描述 |
| 0 | id | Varchar(20) | 1 | 标识 |
| 1 | company\_name | Varchar(100) | 1 | (创业公司与项目)融资公司表id |
| 2 | education\_back | Varchar(50) | 小学毕业 | 创业者教育背景 |
| 3 | work\_experience | Varchar(400) | xx年-xx年在xx公司担任xx | 工作经历 |
| 4 | sex | tinyint | 0(女)，1(男) | 创业者性别 |
| 5 | age | int | 18--80 | 年龄 |
| 6 | venture\_experience | Varchar(300) | 创业经历(xx年xx月创建以xx项目为主导的公司) | 创业者的创业经历 |
| 7 | family\_back | Varchar(150) | 创业者家庭背景 | 创业者家庭背景 |

6.投资机构评分表(score)---存储根据fund\_in表中投资记录，以融资公司的角度给曾经给自己投资的投资机构进行评分信息，以下表格是属性说明

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 计数序号 | 字段名称 | 数据类型 | 初始值(范围) | 字段描述 |
| 0 | id | Varchar(20) | 1 | 标识 |
| 1 | agency\_id | int | 1 | (投资方)fund\_inf  表中的id或者  投资机构表id |
| 2 | company\_id | int | 1 | fund\_inf表中的融资公司 |
| 3 | score | double | 0.0 | 每个投资机构获得评分 |

7.投资机构投资次数统计表(agencyinvestcount)---存储根据fund\_in表中投资记录，每个投资机构的总投资次数统计信息，以下表格是属性说明

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 计数序号 | 字段名称 | 数据类型 | 初始值(范围) | 字段描述 |
| 0 | id | Varchar(20) | 1 | 标识 |
| 1 | count\_invest | int | 1 | 每个投资机构的总投资次数 |
| 2 | name\_agency | varchar(100) | not null | 投资机构名称 |

8.给每个融资公司推荐的投资机构表(recommresult)---存储根据fund\_in表中投资记录，同过建模型给每个融资公司推荐投资机构信息，以下表格是属性说明

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 计数序号 | 字段名称 | 数据类型 | 初始值(范围) | 字段描述 |
| 0 | id | Varchar(20) | 1 | 标识 |
| 1 | company\_id | int | 1 | fund\_inf表涉及的融资公司的id一般对应在company表中id |
| 2 | agency\_id | int | 1 | 对每个融资公司推荐的  fund\_inf表涉及的投资机构对应id |
| 3 | score | double | 0.0 | 对投资机构的预测分数 |