## **设计方案**

1引言

面向使用该系统的人员，介绍该项目的功能，数据库设计，界面设计以及模型设计。

2 功能设计

功能包括创建数据库表，将数据集导入数据库，通过数据集训练模型，根据输入预测结果。

3 数据库设计

数据库的设计包括如下两张表的设计如表1所示：

|  |  |  |
| --- | --- | --- |
| 序号 | 表名（英文） | 说明 |
| 1 | dataset | 波士顿房价情况表 |
| 2 | predict | 预测房价结果表 |

表1 系统数据库表

1. 波士顿房价情况表（dataset）

用来存储原始的关于波士顿房价的基本情况，如表2所示。

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 属性（英文） | 类型 | 主键 | 外键 | 空值 | 属性（中文） |
| id | int | yes | no | no | 序号 |
| CRIM | float | no | no | no | 城镇人均犯罪率 |
| ZN | float | no | no | no | 住宅用地超过25000英尺的比例 |
| INDUS | float | no | no | no | 城镇非零售商用土地比例 |
| CHAS | float | no | no | no | 边界是否有河流 |
| NOX | float | no | no | no | 一氧化碳浓度 |
| RM | float | no | no | no | 住宅平均房间数 |
| AGE | float | no | no | no | 1940年之前建成的自用房比例 |
| DIS | float | no | no | no | 到波士顿五个中心区域的加权距离 |
| RAD | float | no | no | no | 辐射性公路接近指数 |
| TAX | float | no | no | no | 每10000美元的全值财产税率 |
| PTRATIO | float | no | no | no | 城镇师生比例 |
| B | float | no | no | no | 城镇中黑人的比例 |
| LSTAT | float | no | no | no | 人口中地位低下者比例 |
| MEDEV | float | no | no | no | 自住房的平均房价 |

表2 波士顿房价情况表

1. 预测房价结果表（predict）

用户使用界面进行预测时，由于模型训练时筛选了数据集中关键的五个变量作为特征。所以用户的预测记录存放在Predict表中，其中字段包括输入数据ZN，RM，DIS，PTRATIO，LSTAT及预测结果MEDEV，如表3所示。

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 属性（英文） | 类型 | 主键 | 外键 | 空值 | 属性（中文） |
| id | int | yes | no | no | 序号 |
| zn | float | no | no | no | 住宅用地超过25000英尺的比例 |
| rm | float | no | no | no | 住宅平均房间数 |
| dis | float | no | no | no | 到波士顿五个中心区域的加权距离 |
| ptratio | float | no | no | no | 城镇师生比例 |
| lstat | float | no | no | no | 人口中地位低下者比例 |
| medev | float | no | no | no | 自住房的平均房价 |

表3 预测房价结果表

4 界面设计

界面使用Bootstrap框架布局，Bootstrap是Twitter推出的一个开源的用于[web前端开发](https://baike.sogou.com/lemma/ShowInnerLink.htm?lemmaId=7678968&ss_c=ssc.citiao.link" \t "https://baike.sogou.com/_blank)的[工具包](https://baike.sogou.com/lemma/ShowInnerLink.htm?lemmaId=57704855&ss_c=ssc.citiao.link" \t "https://baike.sogou.com/_blank)。它由Twitter的设计师Mark Otto和Jacob Thornton合作开发，是一个CSS/HTML框架。Bootstrap提供了优雅的HTML和CSS规范，它即是由动态[CSS语言](https://baike.sogou.com/lemma/ShowInnerLink.htm?lemmaId=10581680&ss_c=ssc.citiao.link" \t "https://baike.sogou.com/_blank)Less写成。通过界面中五个输入框输入需要预测的波士顿房价的特征，点击预测按钮，将输入的特征数据传入后台模型计算，返回预测房价结果使用echart饼图显示在页面中。设计如图1所示

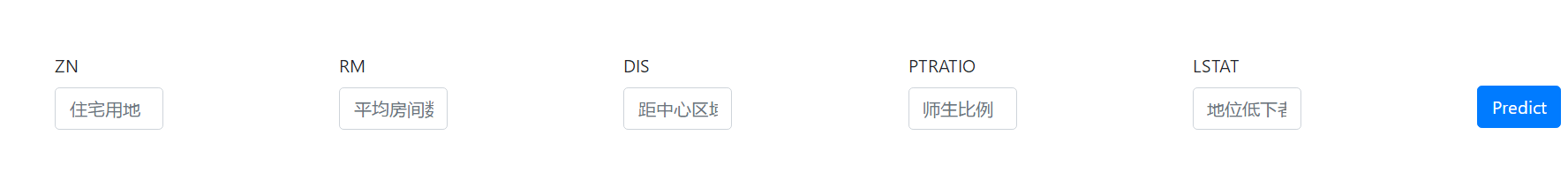


图1 界面设计图

5 模型设计

使用sklearn提供的数据集，将原始数据集存放进数据库中，然后经过分析，筛选出五个最为关键的特征作为模型训练的输入，将对应的平均房价作为目标值，训练模型。在模型训练之前，我们需要对筛选的特征进行归一化处理，消除各个特征之间由于度量单位的不同带来的影响。模型选用SVR，作为SVM用来解决回归问题的算法，相较于其他回归模型，SVR拟合效果更好，同时可以通过核函数和加入惩罚项等方式，优化模型或减少离群点的影响。模型训练完成后，将其序列化为模型文件，后续预测可直接反序列化调用。