**计算机科学与工程学院 实验报告**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **实验课程名称** | | **Python编程与数据分析project4** | | | **实验总成绩** |  |
| **专业** |  | | **班级** |  | **指导教师签字** |  |
| **学号** |  | | **姓名** |  | **实验报告批改时间** |  |
| **基于Project3的二手房价格预测功能** | | | | | | |
| **项目背景及思路**  1.实验三的项目我是用Django框架搭建的服务器和客户端，所以实验四就在实验三的基础上新增了一个页面用来作为客户端用户输入房屋信息，在服务器通过数据处理、线性回顾分析后将结果在新的页面上展示出来。  2.项目环境：PyCharm2020.2 + python3.7 + Django3.2；界面使用html编写；css和js使用bootstrap-3.4.1及font-awesome-4.7.0；图表展示使用echarts.js。  额外导入的包都在项目文件夹下app01/static文件夹中。  3. Django 采用了 MVT 的软件设计模式，即模型，视图和模板。M是Model，模型，是用于完成操作数据库的，进行数据处理。V是View，视图，接收请求，进行业务处理，返回应答。T是Template，模板，负责封装构造要返回的html，也就是用于展示给客户端的页面效果。路由控制器（即项目中url.py文件中的url）将请求转发给对应的视图函数，完成业务逻辑，视图函数将从model中获取的数据嵌入到template中的模板文件（html）渲染成一个页面字符串，返回给客户端的流程。  4.由于在前两次实验中我将数据存入我本地的postgresql数据库中，而且还有许多不利于数据分析的数据信息（比如数据库中有两列是“标题”和“开发商信息”），数据的预处理就会比较麻烦，所以我在本次实验中我重新对二手房信息进行了爬取，并且在爬取阶段就对数据进行了预处理，按照不同的区分别进行爬取。本次实验的数据集来自于链家二手房网站上“沈阳”的二手房信息，用爬虫分别对沈阳市“浑南”，“于洪”, “和平”,“沈河”等区的二手房进行爬取，将爬取的结果保存在csv文件中；模型采用线性回归模型，将所获取到的数据分为训练集和测试集；之后通过多次调参最终取得了不错的预测结果。  **项目文件结构**  demo3:  ├─demo3  │ ├─\_\_init\_\_.py  │ ├─settings.py  │ ├─urls.py  │ ├─wsgi.py  │ └─asgi.py  ├─app01  │ ├─migrations  │ │ └─\_\_pycache\_\_  │ ├─static  │ │ ├─css  │ │ ├─img  │ │ ├─js  │ │ └─plugins  │ │ ├─bootstrap-3.4.1  │ │ │ ├─css  │ │ │ ├─fonts  │ │ │ └─js  │ │ └─font-awesome-4.7.0  │ │ ├─css  │ │ ├─fonts  │ │ ├─less  │ │ └─scss  │ ├─templates  │ ├─utils  │ └─\_\_pycache\_\_  └─  主要文件以及文件夹介绍：   1. \_\_init\_\_.py：inti文件，标识当前所在的项目目录是一个 Python 包。 2. settings.py：项目配置文件，用来进行app的注册以及数据库的连接。 3. urls.py：路由文件，即项目包括的网址。 4. manage.py：主文件，项目的入口，用于启动项目、创建应用和完成数据库的迁移。 5. static文件夹：是我自己创建的静态文件夹，里面是一些插件、css、js以及图片文件。 6. templates文件夹：保存自己写的html文件。 7. app01文件夹中的views.py：是视图函数文件，用来实现逻辑功能，比如数据库数据的增删改查、接受POST请求的内容等功能。 8. app01文件夹中的models.py：是模型文件，定义了所需要的表结构（如用户表、房屋信息表），通过Django的ORM来转换为SQL语句。   **运行结果展示**   1. 启动项目：     点击PyCharm控制台中生成的蓝色网址，127.0.0.1是本地环回地址表示本机，即主机号，8000是项目对应的端口号（与python Socket中的hostname和port相对应）  点击之后会自动打开浏览器并出现这个界面，1-8都可以分别打开。     1. 由于在实验三的报告中都已经展示过数据的获取、展示、图标展示、更新、注册、登录等功能，所以在这里就直接进行实验四新增房价预测功能的介绍：先进入主界面网址：<http://127.0.0.1:8000/mainscene/>     点击右上角的“预测”按钮，会跳转到下面这个界面：    这个界面是供用户输入想要预测的房屋的相关信息。  在这个界面中我们可以输入面积、户型（如三室一厅）、朝向（如南 北）、楼层（分为三类：高楼层、中楼层、低楼层）、装修（分为三类：毛坯、简装、精装）、位置（'于洪', '和平', '大东', '沈河', '浑南', '皇姑', '铁西'）以及新旧程度（分为三类：2005之前、2005~2015、2015之后），并且除了“面积”我为其他的输入都设置了下拉框，方便用户的输入。  在输入相关信息后，点击“确定按钮”，会跳转到“结果展示”界面：    在“结果展示”界面中首先在一个面板中显示出了在后端经过模型学习、预测之后给出的相应房屋信息的房价预测结果，按照我刚才的输入显示出房价大约是189万元。  在下方的面板中展示了“于洪”、“和平”、“大东”、“沈河”、“浑南”这5个区中与刚才输入的城市信息最相符的5个房屋，我在每个区中都选取相似度最高的房屋信息进行输出。  同时在控制台中输出了线性回归的散点图：    **代码展示**   1. 新增predict()函数，在函数中先判断是不是POST请求，若是就获取用户在页面上输入的数据：   （1）在predict对应的前端页面中我设置了form表单，方式设为“post”，使服务器端能获取到前端页面的数据；除了第一个面积是input输入框外，其他的输入全部都是<select>选择框，实现了下拉选择的功能。    （2）在predict函数中通过输入框对应的name来分别获取用户的输入，转化成string类型存入myhouse列表中：       1. 通过pandas库中处理csv文件的方法对爬取到的数据进行预处理：      1. 数据预处理的结果：（在vscode中新建ipynb文件运行的结果）     （5）使用train\_test\_split进行交叉验证，将数据拆分成训练集和测试集。    （6）模型训练：在创建数据集之后我们需要将训练集放入skleran中的线性回归模型LinearRegression()进行训练，使用.fit方法进行模型的训练操作。  在模型训练完成之后会得到所对应的方程式（线性回归方程式）需要利用函数中的intercept\_与coef\_来得到。    （7）把myhouse里的元素从原来的string类型变成float类型，转置之后通过model.predict()进行房屋价格的预测：    （8）通过减少特征向量、特征降维、调整测试机和训练集的数据比例来修正、改进模型，在这里不再一一列出。  （9）最终通过多次测试可以得到精度比较高的预测值。  **感悟与体会**  经过这四次的Python实验，我从零开始完成了一个数据分析、模型训练、机器学习的python项目，从一开始的python语法不熟练、写代码磕磕绊绊，到后来逐渐掌握了不同库中函数的用法，我感到了很大的成就感。爬虫在我上大一的时候就粗浅地接触过，但那时对爬虫地原理、为什么要用那些库等都一知半解，经过这四次的python实验以及老师学长学姐的教导，我认为我还是比较熟练地掌握了爬虫的方式；清楚的记得实验一要求写一个日志系统，我一开始是用写入文件的方式手动添加日志记录，后来有一次无意中在网上搜到python有自带的日志库叫做loggru，当时也算是第一次体会到“Python的创作者有一扇传送门，他们早就预知到你今后在开发中需要使用的功能，并且已经提前帮你们实现了，你需要做的就只是pip一下。”；在实验二中进行了pyqt的开发，还清楚的记得老师说：“不一定非得用海龟画图”的喜悦，因为我大一和大二的C++课设以及数据结构课设均是用Qt做的，对Qt也算是十分熟悉了。在查阅资料之后成功在pycharm上安装了Qt开发的相关插件，选择用pycharm写起Qt来也算是十分得心应手了，这是第二个收获；在实验三中，在实验课前本来我已经用socket模块写好了客户端和服务器端，也就是在实验二pyqt的基础上添加了一些代码就已经完成了题目要求，但上实验课时看到学长使用的Django框架，我当时都震惊了，竟然有这么吸引我的东西，于是我就用了好几天的时间从头到尾学习了Django框架的开发，在学习过程中有将早已遗忘的html、css等语法捡了回来，通过教程自己下载安装了许多前端网站开发需要使用的静态插件，如bootstrap.min.css等，这也算是我初探网站开发；在最后的实验四中听老师细细讲解了如何使用python中的sklearn库在人工智能方面的应用，这是我第一次真正接触到有关机器学习方面的代码，以前听到过的那些“训练集”、“主成分析”、“线性回归模型”等词语在实验四中进行了初次的理解与运用，让我提回到了人工智能的乐趣。  真的十分感谢老师、学长提供的这次Python实验课的学习材料，让我在多个方面都有所收获，不管是代码能力还是对与软件的应用都有了很大的提高，这次的python实验课给我的感觉超越了以往上过的所有课设，不管是从内容、题目设置还是其他方面都让我在打代码的过程中感到了真真切切的乐趣，在看到自己这四次实验课总共数千行的代码时也体会到了十足的成就感。当然，依然是是有许多不足的地方，还需要认真学习，多加努力。 | | | | | | |