



上海大学

SHANGHAI UNIVERSITY

# 利用神经网络预测上海租房价格

李瑞林 人工智能研究院

2023年3月30日

本项目利用Python的Scrapy框架爬取链家网的上海市租房信息，利用**pandas**、**numpy**、**matplotlib**、**seaborn**、**folium**、**wordcloud**等库进行数据分析和可视化，通过one-hot编码和文本特征提取出120个训练特征，搭建3层神经网络对上海市租房价格进行预测。

**Pandas:**数据分析库，它提供了高效的数据结构和数据分析算法接口

**Numpy:**提供多维数组对象,各种派生对象(如掩码数组和矩阵),以及用于数组快速操作接口

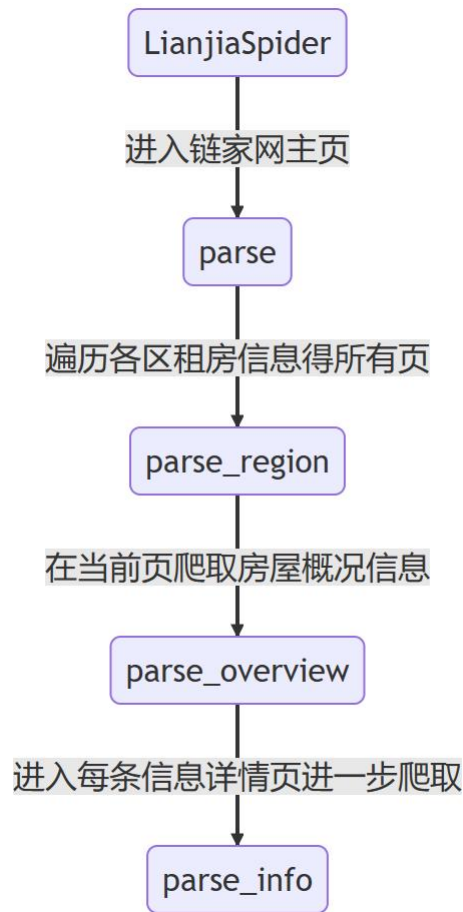
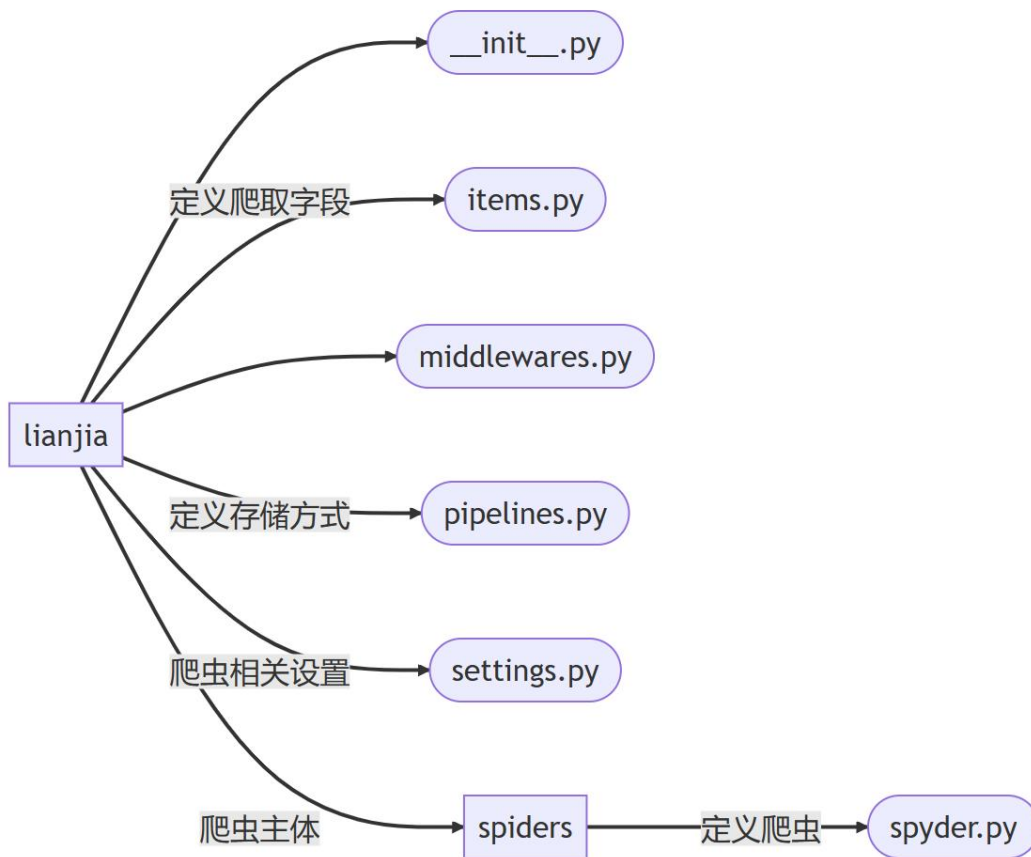
**Matplotlib:**数据绘图库,以各种硬拷贝格式和跨平台的交互式环境生成图形

**Seaborn:**数据分析可视化库,是对matplotlib进行二次封装而成

**Folium:**可视化地理空间数据

**Wordcloud:**词云展示库

Ref : <https://github.com/MaiMeng1204/Predict-house-rent-with-neutral-network>



字段名	字段中文	数据类型
title	标题	str
location	地点	str
house_type	房屋类型	str
house_code	房源编号	str
price	价格	float
tags	房源标签	str
lease	租赁方式	str
area	面积	str
orientation	朝向	str
floor	楼层	int
elevator	电梯	str
stall	车位	str
water	用水	str
electricity	用电	str
fuel_gas	燃气	str
heating	采暖	str
facility	配套设施	str
description	房源描述	str

area	descripti	electric	elevat	facility	floor	fuel_g	heatir	puse	house_ty	lease	location	orientatic	price	stall	tags	title	wate
72.38m		民电	无	洗衣机,空调,电视,床	高楼层/6层	有	自采暖		2室1厅1卫	整租	浦东-航头-鹤沙航城航武嘉园	南	4200元/月 (季付价)	租用车位	精装,动迁安置房	鹤沙航城航武嘉园 2室1	民水
42.83m		民电	无	洗衣机,空调,热水器,床	高楼层/6层	有	暂无数据		1室1厅1卫	整租	宝山-张庙-通河一村	南	3600元/月 (季付价)	暂无数据	官方核,近地铁,随时看	通河一村 1室1厅	民水
74.06m		民电	有		中楼层/25层	有	自采暖		2室1厅1卫	整租	杨浦-控江路-新凤城	南	8000元/月 (季付价)	暂无数据	业主自荐,新上,近地	整租-新凤城 2室1厅	民水
43.99m	中国银行、	民电	无		低楼层/6层	有	自采暖		1室1厅1卫	整租	闵行-老闵行-华江小区	南	3100元/月 (季付价)	暂无数据	官方核,随时看房	整租-华江小区 1室1厅	民水
108.38m		民电	有	洗衣机,电视,冰箱,热水器	高楼层/24层	有	自采暖		2室1厅1卫	整租	静安-江宁路-静安兴海城	南	13800元/月 (季付价)	租用车位	官方核,精装,随时看	租-静安兴海城 2室1厅	民水
41.90m		民电	无		高楼层/6层	有	自采暖		1室1厅1卫	整租	静安-南京西路-茂名公寓	南	7200元/月 (季付价)	暂无数据	官方核,近地铁,精装	整租-茂名公寓 1室1厅	民水
107.59m	绿地公园	民电	有		低楼层/28层	有	自采暖		2室2厅1卫	整租	徐汇-田林-华利公寓	南	10600元/月 (季付价)	租用车位	官方核,近地铁,精装	租-华利公寓 2室2厅	民水
38.84m		民电	无	洗衣机,电视,冰箱,热水器	中楼层/6层	有	暂无数据		1室1厅1卫	整租	长宁-仙霞-天山五村	南	5000元/月 (季付价)	暂无数据	官方核,近地铁,精装	租-天山五村 1室1厅	民水
123.35m		民电	有		中楼层/11层	有	自采暖		2室2厅1卫	整租	普陀-长征-金沙雅苑(二期)	南北	9000元/月 (季付价)	暂无数据	新上,近地铁,精装	动迁-金沙雅苑(二期) 2室2厅	民水
32.79m		商电	有	洗衣机,空调,热水器	中楼层/18层	有	自采暖		1室1厅1卫	整租	宝山-共康-卓越时代广场	北	3500元/月 (季付价)	暂无数据	精装,随时看房,非居	租-卓越时代广场 1室1厅	商水
114.91m		民电	无		低楼层/6层	有	自采暖		2室2厅1卫	整租	宝山-大场-梧桐城邦(二期)	南北	7200元/月 (季付价)	租用车位	官方核,精装	梧桐城邦(二期) 2室2厅	民水
45.67m		民电	有	洗衣机,空调,电视,床	低楼层/17层	有	自采暖		1室1厅1卫	整租	宝山-高境-美岸华庭(公寓)	南北	5800元/月 (季付价)	暂无数据	官方核,近地铁,精装	随-美岸华庭(公寓) 1室1厅	民水
61.00m		民电	有	洗衣机,空调,电视,床	低楼层/11层	无	暂无数据		1室1厅1卫	整租	宝山-罗店-美罗家园美平苑	南	3000元/月 (季付价)	租用车位	新上,精装,动迁安置房	美罗家园美平苑 1室1	民水
160.00m		商电	有	洗衣机,空调,电视	中楼层/31层	无	自采暖		2室2厅0卫	整租	宝山-上大-绿地领海(商业类)	南北	6000元/月 (季付价)	暂无数据	非居住房屋	绿地领海(商业类) 2室2厅	商水
150.00m		民电	无	空调,床	低楼层/6层	有	自采暖		3室1厅1卫	整租	浦东-金杨-金杨新村十一街坊	南	12000元/月 (季付价)	租用车位	随时看房,动迁安置	金杨新村十一街坊 3室1	民水
42.77m		民电	无	洗衣机,电视,冰箱,热水器	高楼层/6层	有	暂无数据		1室1厅1卫	整租	静安-彭浦-闻喜路1223弄	南北	4300元/月 (季付价)	暂无数据	官方核,近地铁,精装	随-闻喜路1223弄 1室1厅	民水
90.75m	场500米以	民电	有	洗衣机,电视,冰箱,热水器	低楼层/11层	有	自采暖		2室2厅1卫	整租	浦东-曹路-河滨城市花园	南	6000元/月 (季付价)	暂无数据	官方核,近地铁	租-河滨城市花园 2室2厅	民水
78.02m		民电	有	洗衣机,空调,热水器,床	低楼层/26层	有	自采暖		2室1厅1卫	整租	虹口-临平路-大连小区	东	7800元/月 (季付价)	暂无数据	官方核,近地铁	租-大连小区 2室1厅	民水
41.33m		民电	无		高楼层/6层	有	自采暖		1室1厅1卫	整租	浦东-塘桥-文兰小区	南	5000元/月	暂无数据	官方核,近地铁	租-文兰小区 1室1厅	民水
47.11m		民电	无	洗衣机,电视,冰箱,热水器	中楼层/6层	无	暂无数据		1室1厅1卫	整租	浦东-金桥-东陆新村四街坊	南	5000元/月 (季付价)	暂无数据	官方核,精装,随时看	房,东陆新村四街坊 1室1	民水

2023年3月24日：共爬取29983条租房信息

## 提取面积，楼层数，价格

```
row['面积'] = int(area[0]) if area else np.nan  
floor = re.findall(r'\d+', row['楼层'])  
row['楼层'] = int(floor[0]) if floor else np.nan  
price = re.findall(r'\d+', row['价格'])  
row['价格'] = int(price[0]) if price else np.nan
```

## 提取房型，即几室几厅几卫

```
room = re.findall(r'\d*室', row['房屋类型'])  
room_num = re.findall(r'\d+', room[0])  
row['室'] = int(room_num[0]) if room_num else np.nan  
hall = re.findall(r'\d+厅', row['房屋类型'])  
hall_num = re.findall(r'\d+', hall[0])  
row['厅'] = int(hall_num[0]) if hall_num else np.nan  
rest_room = re.findall(r'\d+卫', row['房屋类型'])  
rest_room_num = re.findall(r'\d+', rest_room[0])  
row['卫'] = int(rest_room_num[0]) if rest_room_num else np.nan
```

## 提取所属区

```
data['区'] = data['地点'].map(lambda x: x.split('-')[0])
```

## 提取配套设施

```
if row['配套设施'] is not np.nan:  
    facility = row['配套设施'].split(',')  
    row[facility] = 1  
    return row  
else:  
    return row
```

谢谢，再见

