

2019 未来杯高校 AI 挑战赛 城市-房产租金预测

竞赛报名地址：https://ai.futurelab.tv/contest_detail/3#contest_index

数据下载链接：<https://pan.baidu.com/s/118NFPr1CPUx36elflLggKQ>(提取码: 66r8, 解压密码: Zp7Y)

智慧城市赛道——房产租金预测

近几年，国内住房租赁市场进入全新的发展阶段，长期公寓市场作为租赁市场的重要部分，越来越受到广泛的关注。但同时中国长期公寓市场也面临着企业市场进入、业务(门店)扩张、资本市场博弈、企业重组并购等多重挑战，其中，如何准确的预测租金便成为该行业发展进程中的一大难题。

本次赛题主要是通过房产市场、租赁市场、市场需求以及房屋配置来做出合理的房租预测，以应对市场变化对运营商和房产机构带来的影响。命题方向为运用机器学习、人工智能等模型算法，结合模型的创新能力，来实现准确预测的目的。

易居企业集团克而瑞作为中国房地产大数据应用服务商，成立于2006年，在房地产大数据领域深耕十载，建立起房地产数据库系统“CRIC”，克而瑞在大数据应用上，同时具备深度与广度的特色，远远走在行业前列。克而瑞此次将相关数据用于比赛，除了能让同学们更加真实的了解企业需求外，也是期望能够借此促进房地产行业的健康快速发展。


易居企业集团·克而瑞

赛题说明

规则概述

线上比赛要求参赛选手根据给定的数据集，建立模型，预测房屋租金。

数据集中的数据类别包括租赁房源、小区、二手房、配套、新房、土地、人口、客户、真实租金等。

平台提供的数据包括训练集、预测试集(Test A)、正式测试集(Test B)，详见数据集说明。

参赛选手需要将预测的结果按照测试集 ID 顺序输出，以便平台对提交模型的结果进行验证。每天每个参赛者最多可以提交 2 次。

模型评分标准

模型结果评分公式如下：

$$\text{score} = 1 - \frac{\sum_{i=1}^m (\hat{y}_i - y_i)^2}{\sum_{i=1}^m (y_i - \bar{y})^2}$$

其中 y_i 表示真实值, \hat{y}_i 表示预测值, \bar{y} 表示测试样本的均值, m 表示测试样本数。Score值越大，拟合效果越好。

线上赛数据发放与结果提交

训练数据Train_data和预测试集Test A将在赛事开始时开放，Train_data供参赛选手训练模型；预测试集Test A只供参赛选手熟悉提交流程，Test A的预测结果不计入正式成绩。

正式测试数据Test B在赛事规定的时间开放，仅供模型验证使用，禁止调整测试数据Test B的顺序及任何其他形式的修改。参赛选手需要在赛事结束前提交模型、代码以及预测结果。平台将及时更新结果排名，取最好成绩作为客观成绩，每人每天有2次提交机会。

比赛要求

- 赛选手可以根据模型需求，筛选适当的数据维度。
- 参赛选手可以根据模型需求，自行清洗训练数据。
- 不得使用提供的数据集之外的数据训练模型。
- 不得将自行手工标注的结果运用于训练和测试。
- 参赛选手需要同时提交说明文档、预测结果、训练和预测代码、模型（参数集），详见结果提交要求。
- 参赛选手所提程序应可在Linux/windows环境顺利运行。建议（不强制）使用Python 2.7或Python3.6作为首选编程环境，使用Anaconda 3 * 搭建软件环境。不限制第三方软件包使用，但须获得合法授权和软件拷贝，以确保主办方可以复现程序运行过程。
- 参赛选手的提交作品需要可完整复现训练和预测过程，主办方不会做任何代码修改，如无法复现的，视作无效提交。

复赛晋级与综合评审

根据模型有效提交的得分，前 10 名选手可以晋级复赛。晋级复赛的选手需要通过 PPT 展示进行线下答辩，PPT 展示主要包括数据清洗规则、模型创新、算法实现、结果分析等。

数据集字段说明

对于小区信息中，关于 city、region、plate 三者的关系：city>region>plate

土地数据中，土地楼板面积是指在土地上建筑的房屋总面积。

类别	字段	字段内容	说明
租赁房源	ID	房屋编号	每间房屋编号唯一
	area	房屋面积	租赁房间的面积，单位平方米
	rentType	出租方式：整租/合租/未知	房间的出租方式
	houseType	房型	房屋的形状
	houseFloor	所在楼层	房间所在的楼层(分为高、中、低三类)
	totalFloor	总楼层数	房间所在楼栋的总楼层数
	houseToward	房屋朝向	房间的朝向
	houseDecoration	房屋装修	房屋内的装修情况
小区信息	communityName	小区名称	房屋所在小区(XQ00001)
	city	城市	城市(SH)
	region	区域	城市行政区域(RG00001)
	plate	板块	区域板块(BK00001)
	buildYear	小区建筑年代	小区建筑年代
	saleSecHouseNum	该板块当月挂牌房源数	板块当月二手房挂牌房源数

配套设施	subwayStationNum	该板块地铁站数量	板块当前地铁站总数量
	busStationNum	该板块公交站数量	板块当前公交站总数量
	interSchoolNum	该板块国际学校数量	板块当前国际学校总数量
	schoolNum	该板块公立学校数量	板块当前公立学校总数量
	privateSchoolNum	该板块私立学校数量	板块当前私立学校总数量
	hospitalNum	该板块综合医院数量	板块当前综合医院总数量
	drugStoreNum	该板块药房数量	板块当前药房总数量
	gymNum	该板块健身中心数量	板块当前健身中心总数量
	bankNum	该板块银行数量	板块当前银行总数量
	shopNum	该板块便利店数量	板块当前便利店总数量
	parkNum	该板块公园数量	板块当前公园总数量
	mallNum	该板块购物中心数量	板块当前购物中心总数量
二手房	superMarketNum	该板块超市数量	板块当前超市总数量
	totalTradeMoney	该板块当月二手房成交总金额(元)	当月板块二手房成交总金额
	totalTradeArea	该板块当月二手房成交总面积(平米)	当月板块二手房成交总面积
	tradeMeanPrice	该板块当月二手房成交均价(元/平米)	当月板块二手房成交均价
	tradeSecNum	该板块当月二手房成交套数	当月板块二手房成交总套数

新房	totalNewTradeMoney	该板块当月新房成交总金额	当月板块新房成交总套数
	totalNewTradeArea	该板块当月新房成交总面积	当月板块新房成交总面积
	tradeNewMeanPrice	该板块当月新房成交均价	当月板块新房成交均价
	tradeNewNum	该板块当月新房成交套数	当月板块新房成交总套数
	remainNewNum	该板块当月新房剩余未成交套数	当月板块新房剩余未成交套数
	supplyNewNum	该板块当月新房供应套数	当月板块新房供应总套数
土地	supplyLandNum	该板块当月土地供应幅数	当月板块土地供应幅数
	supplyLandArea	该板块当月土地供应面积	当月板块土地供应面积
	tradeLandNum	该板块当月土地成交幅数	当月板块土地成交幅数
	tradeLandArea	该板块当月土地成交面积	当月板块土地板楼面积
	landTotalPrice	该板块当月土地成交总价	当月板块土地成交总价
	landMeanPrice	板块土地楼板价（元/㎡）	板块楼板均价(板块)
人口	totalWorkers	当前板块现有办公人数	当前板块现有办公人数(包括本月新增招聘)
	newWorkers	该板块当月流入人口(新招聘人数)	本月新增招聘人数
	residentPopulation	当前板块常住人口	当前板块常住人口
客户	pV	该板块当月租客浏览网页次数	当月板块租客浏览网页次数
	uv	该板块当月租客浏览网页总人数	当月板块租客浏览网页总人数
	lookNum	带看次数	线下看房次数
真实租金	tradeTime	成交日期	租房的日期
	tradeMoney	成交租金	成交房屋每月租金