个人房产信用评估模型

**背景**

国家政策调控趋严，房地产行业贷款额度趋紧。由于房地产是一个资金密集型的产业，投资规模大、周期长，其融资渠道相对单一，对银行贷款依赖过高成为资金筹措的瓶颈也是风险来源的渠道。在上述背景下，房地产行业信用愈发重要，在防范其风险上考虑需要结合房地产全生命周期过程中所有行为，深度分析个人的房产信用。

杭州中房信息科技有限公司作为杭州房地产大数据的运营者，拥有房地产交易的大量、多维数据。为了助力市场监督与检查，建设领域内信用，赋能行业发展，特设定本题。本题针对房产交易双方，利用真实成交数据，从房地产交易维度对交易参与人进行评价，计算个人房产信用评分。

**任务描述：**

1. **数据说明**

本赛题提供完整的房产交易数据与名下房屋查封情况数据，包括如下字段：

|  |  |  |
| --- | --- | --- |
| USER\_ID | 购房人ID | 抽样&脱敏 |
| Age | 年龄 | 枚举值，按照0-5表示：0:20岁以下、1:20-30岁、2:30-40岁、3:40-50岁、4:50-60岁、5:60岁以上 |
| Sex | 性别 | 枚举值，按照0-1表示：0:男、1:女 |
| TEL\_ID | 电话归属地 | 枚举值，按照0-1表示：0:浙江省、1:其他 |
| Province | 省份 | 请下载省份字典进行参照 |
| NATIONALITY | 国籍 | 请下载国籍字典进行参照 |
| HOU\_ID | 房屋编号 | 整数型 |
| Property\_ID | 房屋属性 | 枚举值，按照0-1表示： 0：存量房、1：商品房、 |
| Property\_Usage\_Type | 房屋用途 | 枚举值，按照0-1表示： 0；住宅、1：非住宅 |
| Property\_Architectural\_Area | 建筑面积 | 枚举值，按照0-4表示： 0:小于60平方米、1:60-90平方米、2:90-120平方米、3:120-140平方米、4:大于140平方米 |
| Property\_Sign\_Date | 网签时间 | 以YYYY/MM表示 |
| Property\_Record\_Date | 备案时间 | 以YYYY/MM表示 |
| Property\_Price | 房屋成交总价 | 枚举值，按照0-5表示： 0:100万以下、1:100-200万、2:200-300万、3:300-400万、4:400-500万、5:500万以上 |
| Property\_Loan\_Way | 贷款方式 | 枚举值，按照1-5表示： 1:组合贷款、2公积金贷款、3:商业贷款、4、银行贷款5、其他 |
| Property\_Payment | 付款方式 | 枚举值，按照0-3表示： 1:一次性付款、2:分期付款、3:抵押贷款、4：其他方式、5：其他情况 |
| CFCS | 查封次数 | 整数型 |
| JFCS | 解封次数 | 整数型 |
| Purchase\_houcs | 历史买房次数 | 整数型 |
| Sale\_houcs | 历史卖房次数 | 整数型 |
| Mortgage\_starttime | 抵押开始时间 | 以YYYY/MM表示 |
| Mortgage\_endtime | 抵押结束时间 | 以YYYY/MM表示 |

1. **任务描述**

本赛题提供2017年1月-2018年5月购房者的房产交易数据与房屋状态数据。

选手结合房地产行业知识通过对数据的处理加工，自行制定评价方案对用户房产信用进行评分（0-100分）。

数据中部分列存在空值或NULL，请参赛者自行处理。为保护用户的隐私和数据安全，所有数据均已进行了采样和脱敏。

**评分数据格式：**

参赛者完成预测后，需要将结果放入指定格式的数据表（非分区表）中，要求结果表名为：User\_Credit\_Predict，格式同如下，要求去除重复。

|  |  |
| --- | --- |
| 用户ID | 房产信用评分 |
| 如10001 | 90分 |

**初赛阶段**

初赛阶段提供10万条房产交易记录。

初赛阶段选手需提供方案文档说明书（格式是ppt或pdf），完整的程序代码（支持python和Java，下同）以及User\_Credit\_Predict结果表。

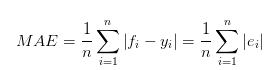
**决赛数据**

决赛阶段提供40多万条房产交易记录。

决赛阶段选手需提供数据量增大，针对模型的一些调参内容（pdf），完整的程序代码以及User\_Credit\_Predict结果表。

1. **评估指标**

本赛题用实际个人房产信用得分（风控模型评分）先行评价，指标维度如下：



其中，fi是预测值，yi是真实值，ei=|fi-yi|即是绝对误差。

后对最终结果通过专家组评定后给出成绩。