



中國人民大學

RENMIN UNIVERSITY OF CHINA

第12组模型展示

實事求是

日期：2025年10月29日



- 特征考量精细，而不是机械全用：合理量化非结构化文本数据（如对于物业费、暖气费等数据的处理）；
- 异常值判断精细，而不是直接套用（0-1判别、Z—score、IQR）；
- 异常值处理精细，而不是简单删除：将其转化为空白数据再填充，以便最大程度保留数据；
- 特征工程精细，而不是过度拟合：以可视化散点图观察特征与因变量之间的关系；
- 模型佐证：随机森林模型不在考核要求范围之内，仅用以佐证数据处理的质量——no garbage in！

买房定价模型	样本内	样本外	交叉验证	Kaggle分数
OLS	771463.93	782916.3	794998.24	23.19
Lasso	825517.91	829945.59	843514.06	23.64
Ridge	680123.66	711583.87	728210.24	24.58
Elastic Net	791054.74	801101.75	814352.38	23.76
Best Linear Model	680123.66	711583.87	728210.24	24.58
Random Forest	86240.40	224618.27	238263.29	59.74

租房定价模型	样本内	样本外	交叉验证	Kaggle分数
OLS	174068.79	177528.44	177674.71	23.19
Lasso	183733.83	185617.55	186701.92	23.64
Ridge	161233.89	166358.61	167990.17	24.58
Elastic Net	176395.01	179769.79	180610.62	23.76
Best Linear Model	161233.89	166358.61	167990.17	24.58
Random Forest	23700.18	63619.75	65165.23	59.74