Final exam

张雷臻

数据处理

- **异常值处理:**四分位距（IQR）方法
- **类别变量：**标签编码、独热编码、类别编码
- **正则表达式：**清洗标点符号、数字、提取文本（建筑面积、梯户、户型）
- **文本分析：**提取高频词，自定义词库，构建新的特征变量
- **聚类方法：**使用K Means聚类创建环线特征、使用KNN创建区域价值特征

机器学习模型与预测

模型选择

- **神经网络：**添加梯度裁剪、调整学习率调度器、添加完整监控
- XG Boost：扩大最大迭代次数、延长早停观察窗口、更频繁的进度输出

指标	值
MAE	536435.135823
RMSE	724914.122789
R^2	0.403804
MAPE	61.393556

指标	值
MAE	524094.435011
RMSE	744902.620562
R^2	0.370472
MAPE	51.768512