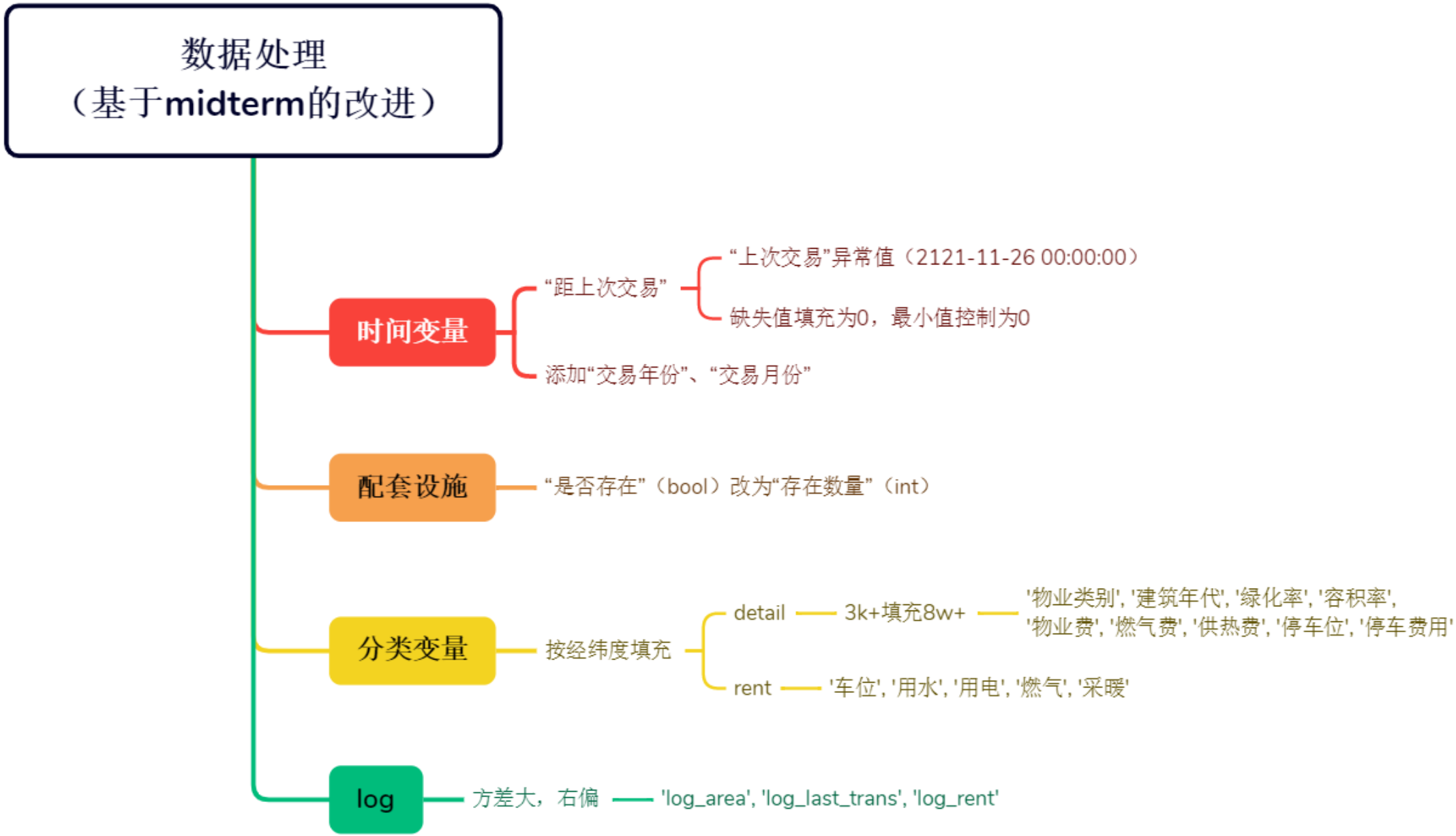
中國人民大學
RENMIN UNIVERSITY OF CHINA

AI & Python Final Pre

汇报人：梁艺高

2025-06-04



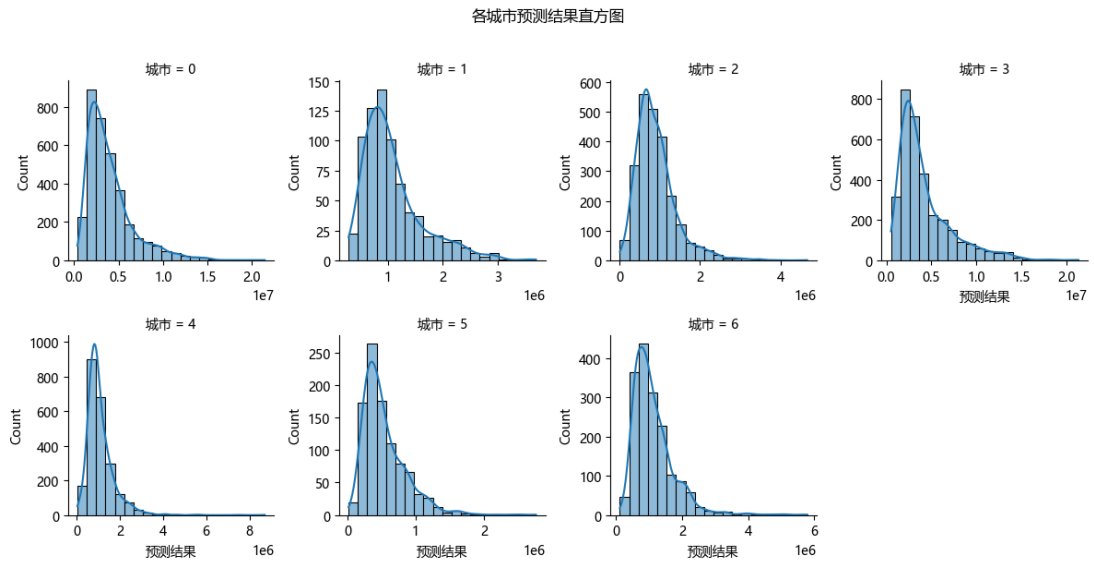
ANN

模型

- wide
  - 'leaky_relu' — l1(0.001)
  - 尝试 — wide增加重要特征（'使用面积'，'距上次交易'，'平均租价'，'室'）权重 — datahub分数↓
- deep
  - 三层“残差连接”（residual_block） — 'swish'（'silu'） — BatchNormalization
  - l2(0.01) — Dropout(0.3)
- loss function — Huber Loss（δ=95%分位数）

训练

- 早停策略
  - 监控MAE
  - patience=15
- 学习率动态调整
  - 初始学习率0.001
  - 学习率衰减比例0.5
  - patience=5 — 比早停策略更敏感
  - 最小学习率下限1e-5
- Test MAE — 15w左右



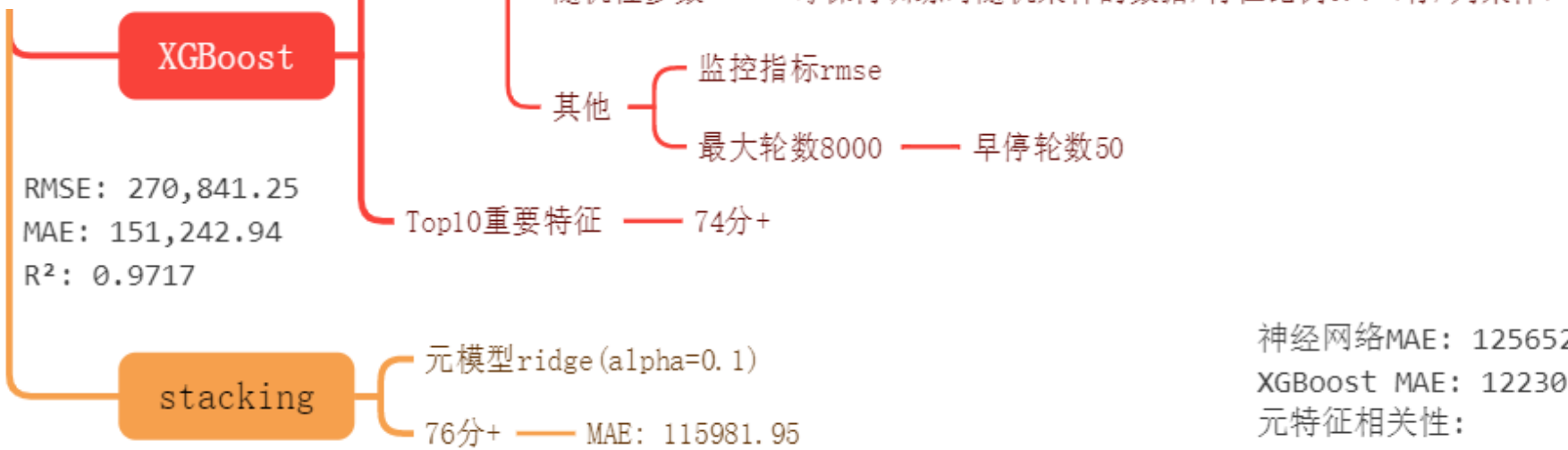
预测结果

- 未处理 — 78分+（最高78.703分）
- 按“经纬度”（半径20km）缩尾 — datahub分数 < 未处理
- 按“城市”缩尾 — datahub分数 < 按“经纬度”缩尾

XGBoost & stacking

Top 10重要特征:

	feature	importance
9	使用面积	52948.0
20	距上次交易	28151.0
1	交易月份	19855.0
8	总楼层	17330.0
35	log_area	15022.0
4	室	13726.0
2	lon	12764.0
34	平均租价	12161.0
18	户数	11124.0
19	elevators_per_household	11089.0



神经网络MAE: 125652.67401437927

XGBoost MAE: 122307.03279879095

元特征相关性:

	NN	XGB
NN	1.000000	0.994181
XGB	0.994181	1.000000


感谢观看
Thanks for watching
汇报人：梁艺高