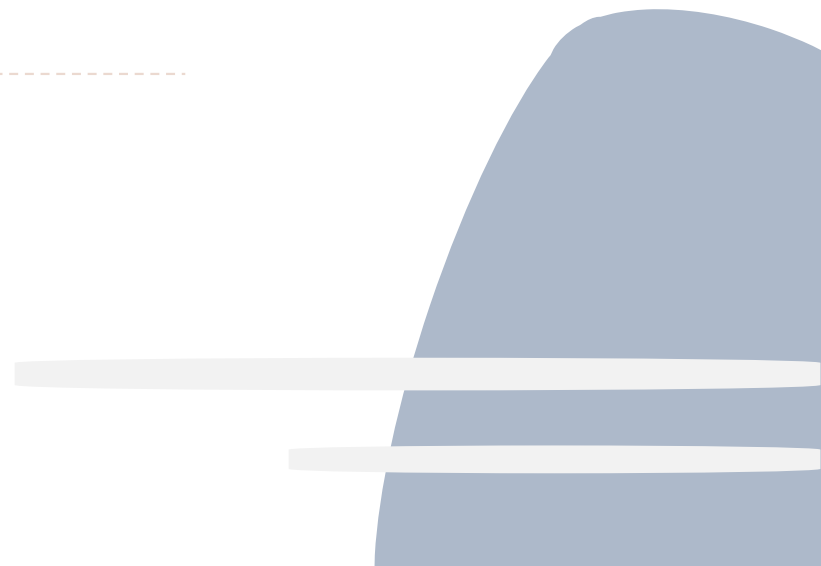
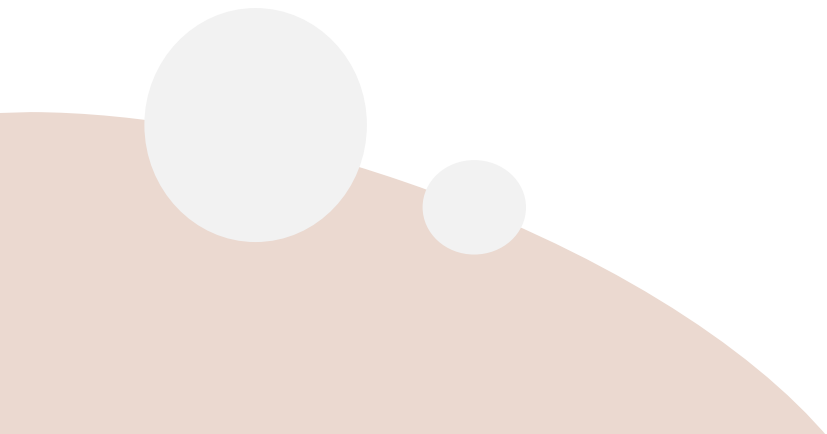


期末考试

2022经数 方国圳





01 特征工程

- ❑ 提取建筑面积数值，户型中室、厅、厨、卫的数量，总楼层数，梯户比例，交易时间和上次交易的间隔
- ❑ 提取是否有核心卖点、周边配套、交通出行的信息
- ❑ 对其余变量进行独热编码处理



02 数据预处理

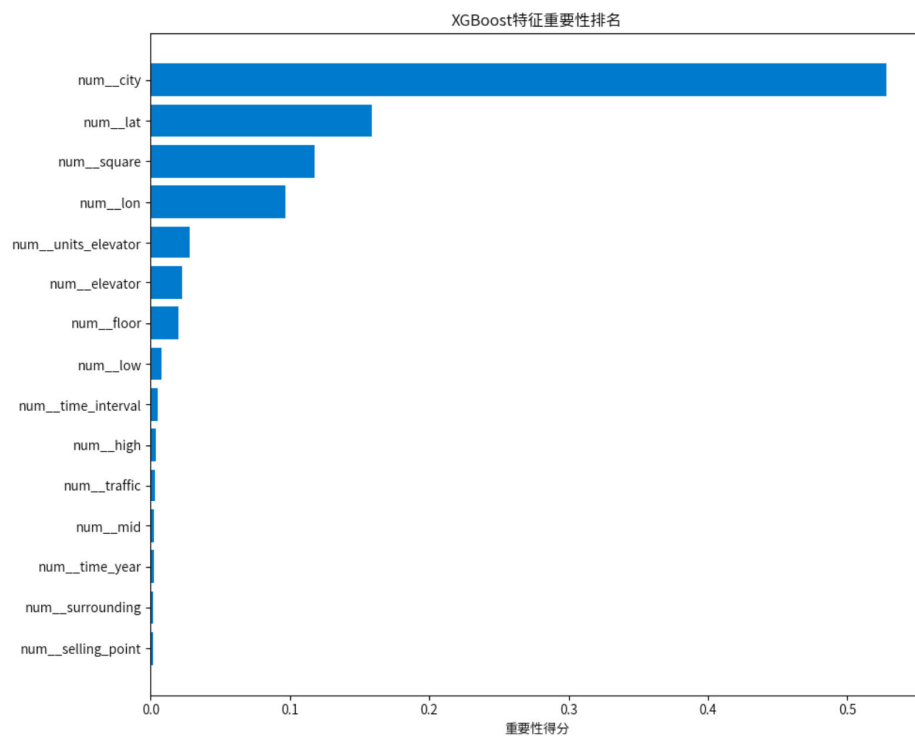
- 划分训练集和测试集，20%为测试集
- 在不同集合内部利用均值进行填充，防止数据泄露
- 利用孤立森林同时处理X和y在训练集上的异常值
- 通过预处理管道在异常值处理后进行数据的标准化

移除了 **3366** 个异常样本，占比：**5.0%**



03 建立模型

- 构建XGBoost模型
- 使用GridSearchCV网格搜索进行超参数调优





03 建立模型

- 最优超参数: {'xgbregressor__colsample_bytree': 1.0, 'xgbregressor__learning_rate': 0.1, 'xgbregressor__max_depth': 7, 'xgbregressor__min_child_weight': 1, 'xgbregressor__n_estimators': 2000, 'xgbregressor__subsample': 1.0}

Metrics	In sample		Cross-validation		Testing	Hackthon Score
MAE	194465.40	194465.40	194465.40	194465.40	81.50	
R ²	0.9422	0.9422	0.9422	0.9422		
Total N of testing set					16827	



感谢观看

