



## 竞赛：房价预测总结



# 结果



- 172只队伍，2288次提交
- 前5位同学将获得奖品

#	△pub	Team Name	Notebook	Team Members	Score ?	Entries	Last
1	—	fxzero			0.11922	5	5d
2	▲1	sxjscience			0.12063	2	2mo
3	▲1	wuwawa			0.12283	77	5d
4	▲5	Leon			0.12283	163	5d
5	▲6	jack			0.12455	21	5d
6	▲2	totoro			0.12485	6	5d
7	▲8	automl (baseline)			0.12502	1	2mo
8	▼2	pullpullyes			0.12518	123	5d
9	▲5	Shaoqing			0.12559	84	5d

# 方法总结



- 第二和第七: autogluon

<https://www.bilibili.com/video/BV1rh411m7Hb/>

- 第三: h2o

<https://www.kaggle.com/wuwawa/automl-using-h2o>

- 第四: 随机森林

<https://www.kaggle.com/jackzh/the-4th-place-approach-random-forest>



# 一些分析

- 已知的排名靠前的4个成绩均使用了集成学习
- 目前不知道是否有使用书中的mlp取得好成绩
  - 特征预处理和超参数是取得好成绩的基础
- 这个数据的一些难点
  - 数值较大
  - 有文本特征（地址，介绍）
  - 训练数据是前6个月，公榜是后3个月，私榜是再后三个月

# 关于automl



- 数据科学家80%时间在处理数据，20%调模型
- Automl现在能处理一些基础的情况
  - 目前节省10%时间，未来节省20%时间
- 为什么还要学习模型
  - （我初中时说：会加减乘除就可以买菜了，为什么还要学三角函数）
  - 当人人都会玩游戏时，你需要成为玩得最好的，或者甚至去制作游戏