论文结构

- 1. Analysis of the Problem / Introduction 总结问题 (自己的给重描述) + 对问题的逻辑 和解决思路
- 2、Assumption & Nomenclature (Notion) 根据对问题证解的发展定样模型 并结出数学符号定义表
 - 3. Data processing & Analysis 数据处理方法,对处理后的数据建作标述。

因为这个任务可以凝处墙和数据。没有想好这个部分一与假决、定义为号的前后关系

4. Model Construction

根据这个好的符号描述模型; 介绍派模型怎样作用于当前任务 6. Answers for required questions
回答题目给出的具体问题 1-3C41

7. Conclusion

8. Report for HK Broker ig les

9. Appendice

提供伪数据 (xlsx形式)

一) Monohulled sailboats
2346个样本
② Catamarans

1146个样本

2个美型=>每个对应 excel 中 纳一张 Tab

表的结构:

在包括从下数据 的前提下可维定 进阶括元

Make	Variant	Length (fi)	Georgraphic Region	Country / Region / State	Listing Price (USD)	Year
str	str	float	str	str	int	str

国称: 解释 listing price

数据传光什么?

D str 名称转化为数字形式 (才能用数字方压进针训练) 例如美型从 ? 数字表示? 松然何量表示?

Year 的某型是 Str并平台区

- D 归一 化处型? 将数值区域限制在O-1内
- ② 锁法、不光整的数据。(次行物有色于是有布运行问题)

数据扩充什么? ① 样本数量扩充? • 继续我帆船 一岁的的产品处理? ~能否确定已给xlsx基本? 一个轮的的怎么办; 什么样的物品更粗做? 一急公康圣相似望, 我终年移时帆船上 ②扩充样本特征 beam, droft, displacement, rigging sail area, shall material, engine hours, sleeping capacity headroom, electronics 步船的特征(粉丝) ● 区域地区的经济特征

X 2 K til L E I I I I A D I X L A L E L M F

并个是处乎然好,我们是当为州后场从时已如一

解决方法用什么?

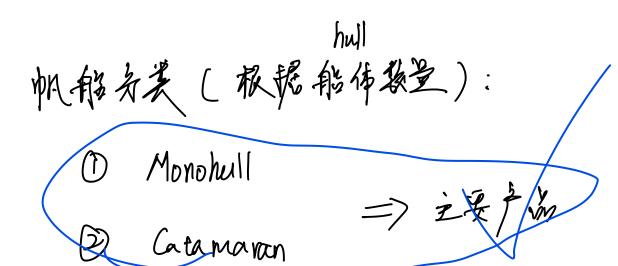
个人认为是一个典型的机器学习问题

国归间篇

数据特征一)高载十连续》价格 快度 地区 美型 午份?

4. 横度国升国归

- 5. XG Boost Wy 2
- 6. 类特何堂机 四归



正有 trimarans (五千hulls) 都 multi-hulls

当前问题。

1. 数据问题

殿园要求研究"二年帆船的价格"

一提供数据完全次有是否是一个的标记一一一个价格的预测器是历史数据? 从哪藏取数据?获取代数据?

Z. 将格强测问题

· 数据的价格街堂单位一)不同国家地区美国的多城? 当时这样的多城?

· 过去数据预测结果再处理? 一)通公的帐片问题之号与是考虑?

中任名的最终自由是对台港地区的价格进行预测。 比较强调地区、船型号的国际与价格的问题。 region variont listing price CUSD)

3. 采用什么方法?

. 上河山南至北层照色义 经计意义的解析

- (牙科解性要求)
 - 等低是否岩质浓化成参考?
 - 。 方独的建模友述是否统情哪当出?
 - ·我们的数据是否可以陷足训练买求?