俞乐楠 徐文彬 李皓然

一、问题重述

二、数据预处理（清洗+集成{规约+变换}+可视化？）

1.统一使用三个字母国家代码

2.缺失值/特殊符号处理

3.算历史奖牌数

4.创建主办国变量

5.整合运动员数据倒每一届里面

三、第一大问题（2028奖牌榜预测）

思路 （分类讨论：挂壁/稳定/记录多不稳定/新国家/没奖牌国家）

探究其他因素的影响（**东道主/新增项目/运动员**）

比赛项目的分析（某些项目统治/竞争激烈）

{

奖牌数量预测问题：负二项回归模型（不是基于历史奖牌数）

（建模、训练、评估模型、评估特征重要性、结果）

具体操作： 建模方法： 分类的时候（用聚类算法进行划分KMEANS/DBSCAN）

某些类型代表队用时间序列模型（ARIMA/LSTM）（稳定的）

对运动员：转化为评价类问题VIKOR

识别特色项目（某项目占该国家奖牌一大部分/某项目奖牌某国家占一大部分）

探究东道主影响（显著性检验）（主办方是否高于其他代表队/高于自己其他时间）相互效应模型/差分法

探究新项目设计的影响（某项目项目数影响该国家奖牌数）

}

四、评估因素对奖牌的影响

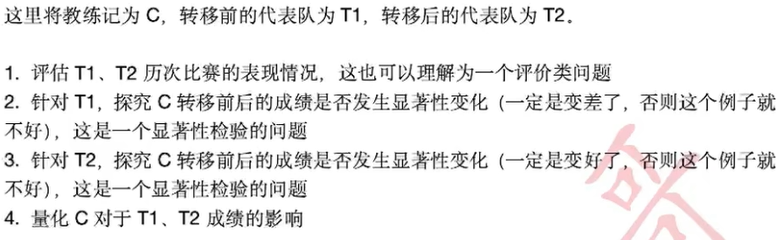
4.1特色项目：某项目奖牌某国家占一大部分

4.2新项目设计的影响

4.3东道主

五、伟大教练效应

（他没给数据，是开放问题，自己找数据来分析教练转移的影响）

数据量不大，找典型的例子来分析  


选三个国家来投资伟大教练+项目

1. 有潜力2.之前教练转移导致下滑的）

伟大教练的影响（意义类题目，就是随便·谈谈就行，不用定量建模分析）

六、您的模型还揭示了有关奥运奖牌数量的哪些其他独到见解？