摘要

奥运会是一项全球性的体育盛会，各国在这一赛事中展示其体育实力。奖牌榜的预测一直是体育研究领域和公众关注的热点之一，因为奖牌榜结果受到多种因素的影响，包括著名教练对国家获奖情况的影响。预测奖牌榜是一项具有挑战性的工作，为此我们建立了两个模型：奥运奖牌LightGBM预测模型和名帅效应多元线性回归模型。

对于任务1，基于LightGBM模型。首先，我们对题目提供的数据进行了处理，考虑了NOC、Sport、Event、athletes\_number、Year、Host等特征对预测结果的影响，并计算出每个特征与奖牌数量的相关性系数，筛选出用于模型训练的特征。模型的R^2值分别为金牌数0.9885、银牌数0.9872、铜牌数0.9804、总牌数0.9916。根据2028年奥运会项目的调整和东道主国家情况，调整了模型输入数据，并预测了所有国家的获奖情况.设定置信区间为90%，计算出预测区间的上下限。对不同奖牌赋予分数，计算国家2024年和2028年奖牌总得分，比较得出进退步国家。再根据2028年预测结果，统计出可能获得首枚奖牌的国家，并根据预测奖牌数量估计胜算概率。最后，我们使用AHP层次分析法对金牌数、银牌数、铜牌数计算权重，根据国家在某个项目上的奖牌数量计算加权得分，分析对于不同国家哪些体育项目最为重要。

对于任务2，基于多元线性回归模型。首先，我们筛选出最可能出现“名帅效应”的国家和对应Event，计算这些国家奥运会的得分。考虑教练执教的滞后性，我们加入了短期（1-2届）和长期（3-5届）的coach\_lag1-2和coach\_lag3-5作为特征，计算各个特征与得分的皮尔逊相关系数,共筛选出10个特征，使用多元线性回归模型训练了短期和长期模型,模型的R^2值为0.828。最后，我们选择了美国、日本和英国三个国家，通过对每个项目上历史平均得分和近三届奥运会得分的比较，得出在哪些体育项目上建议聘请名帅。

对于任务3，我们从两个模型中挖掘与奥运会奖牌榜相关的独特发现，分析了奖牌分布与项目参与度、奖牌榜与经济发展水平的关系，以及奖牌背后的训练体系与科研支持，并给出奥委会参考建议。最后，我们对奥运奖牌LightGBM预测模型的year和total\_event两个参数进行了敏感性分析。这些参数的高敏感性有助于捕捉细微变化对奖牌数量的潜在影响，从而更准确地进行预测。