关键词：网球比赛、势头预测、多层次指标、随机森林模型、置换检验、统计显著性分析

摘要：

在2023年温布尔登男单决赛中，20岁的西班牙新星卡洛斯·阿尔卡拉兹击败了36岁的大满贯选手诺瓦克·德约科维奇，引发了对比赛中貌似可以用来预测比赛的方向和预测胜利者的重要因素——势头的研究。本研究以2023年温网男单第二轮以后数据为基础，构建了一个综合的网球比赛势头评价和预测模型，通过多层次指标分析球员状态，准确评价选手表现，预测比赛势头，并评估其比赛结果的影响。

在第一问中，通过Kappa检验确认了发球与得分之间的中等相关性，建立了多维指标评价体系，将球员状态和表现分为心态、技巧、体力和发球和其他五个方面进行评价，其中：其他部分是由于天气、场地、个人特征等未知扰动项的存在。然后通过对已知数据归一化和标准化，心态方面的定性变量在滑动窗口和全局两种注意力下的好坏球数累计、技巧方面的符号变量赋值等等，将数据转换成更好的形式，接着，采用随机森林模型，分析两球员各方面及其内部变量的权重，构建了球员表现的预测模型，成功预测并可视化展示了1701场5个set比赛中两名选手各自的表现状况。

第二问中，针对网球教练对“势头”是否起作用的疑虑，首先我们通过随机森林对每一个得分点的势头数据进行预测，准确率为71.36%，然后通过置换检验和统计显著性分析，证明了在0.05的显著性水平下势头在对比赛结果的预测上有用。

在第三问中，针对教练们关心的比赛势头变化，我们用lightGBM, SVM, ARIMA, XGboost四种模型预测球员们的势头波动，选择RMSE、MAE最小，即准确率最高的模型lightGBM得出预测的势头值，并找出与比赛波动最相关的因素：发球的掌握权。

考虑到**历史比赛中，每名选手的“动量”波动情况，**对每个选手提出新的建议，其中对阿尔卡拉兹的一条建议为：发挥体力优势，注重发球时速度和技巧，增加对手接发球的难度，提高一发的成功率。

在第四问中， 通过在2023年温网女单比赛首场1301场和2023年温网女单比赛1/4决赛2503场中测试模型，评估了模型的适用性和通用性，在进行针对性的个例分析时，我们通过比赛势头波动变化的可视化，为教练们提供了更全面的球员表现评估。研究发现模型对于温网规则下男单、女单比赛具有极高的通用性，但在其他特定情境下仍需进一步优化。

最后，基于模型评价和敏感性分析的结果，本研究撰写了一份备忘录，总结了模型的优势、可能的改进，为教练和运动员的决策提供更科学、实用的支持，帮助他们更好地理解和应对影响网球比赛流程的各种因素。