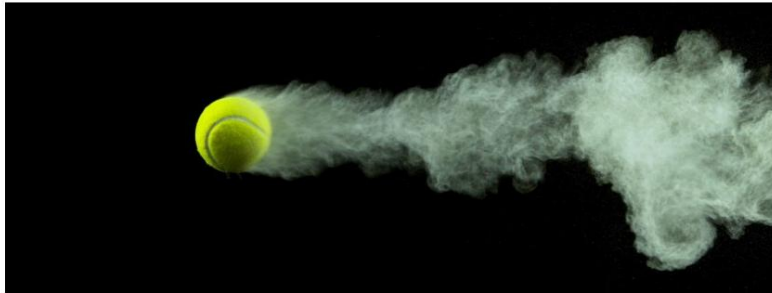


2024年MCM

问题 C:网球的势头



在2023年温网绅士决赛中,20岁的西班牙新星卡洛斯·阿尔卡拉斯击败了36岁的诺瓦克·德约科维奇。这是德约科维奇自2013年以来首次在温布尔登失利,并结束了这位历史上最伟大的大满贯球员之一的非凡表现。

这场比赛本身就是一场非凡的战斗。[1]德约科维奇似乎注定会轻松获胜,他以 6-1 统治了第一盘(7 场比赛中赢了 6 场)。然而第二盘气氛紧张,最终阿尔卡雷斯以7-6的决胜局获胜。第三盘与第一盘相反,阿尔卡拉斯以6-1轻松获胜。这位年轻的西班牙人在第四盘似乎完全掌控了局面。比赛开始后,比赛再次发生了变化,德约科维奇完全控制了局面,以 6-3 赢得了比赛。第五盘也是最后一盘,德约科维奇从第四盘开始就取得了优势,但方向再次发生了变化,阿尔卡拉兹控制了局面, 6-4 获胜。本场比赛的数据在提供的数据集中,“match_id”为“2023-wimbledon-1701”。当德约科维奇拥有优势时,您可以使用等于 1 的“set_no”列查看第一盘的所有得分。看似有优势的选手经常会出现令人难以置信的波动,有时是很多分甚至是比赛。归因于“势头”。

字典中对动量的定义是“通过运动或一系列事件获得的力量或力量。”[2]在体育运动中,一支球队或运动员可能会在一场比赛/比赛中感觉自己拥有动量或“力量/力量”,但这种现象很难衡量。此外,如果存在的话,比赛期间的各种事件如何产生或改变动力也不是很明显。

提供 2023 年温布尔登男子网球公开赛前 2 轮比赛后所有得分的数据。您可以自行决定选择包含其他玩家信息或其他数据,但您必须完整记录来源。使用数据来:

- 开发一个模型,捕捉得分发生时的比赛流程,并将其应用于一场或多场比赛。您的模型应该确定哪位球员在比赛的给定时间表现更好,以及他们的表现有多好。根据您的模型提供可视化来描述比赛流程。注意:在网球比赛中,发球的球员赢得分数/比赛的概率要高得多。您可能希望以某种方式将其纳入您的模型中。
- 一位网球教练对“势头”在比赛中发挥的作用表示怀疑。相反,他假设一名球员的比赛波动和成功是随机的。使用您的模型/指标来评估此声明。

- 教练们很想知道是否有指标可以帮助确定比赛流程何时会从有利于一名球员转向另一名球员。
 - o 使用至少一场比赛提供的数据,开发一个模型来预测比赛中的这些波动。哪些因素似乎最相关(如果有的话)?
 - o 考虑到过去比赛“势头”波动的差异,您如何建议一名球员与另一名球员进行一场新比赛? · 在一场或多场其他比赛中测试您开发的模型。您对比赛走势的预测能力如何?如果模型有时表现不佳,您能否确定未来模型中可能需要包含的任何因素?您的模型对其他比赛(例如女子比赛)、锦标赛、球场和其他运动(例如乒乓球)的通用性如何。

- 制作一份不超过 25 页的报告,其中包含您的发现,并包括一份一到两页的备忘录,总结您的结果,并为教练提供有关“势头”作用的建议,以及如何让玩家做好应对影响比赛的事件的建议。网球比赛期间的比赛流程。

您的 PDF 解决方案总页数不超过 25 页,应包括: · 一页摘要表。

- 目录。
- 您的完整解决方案。 · 一到两页的备忘录。
- 参考文献列表。
- [AI 使用报告](#) (如果使用,则不计入 25 页限制。)

注意:完整的 MCM 提交没有具体要求的最小页面长度。您最多可以使用 25 页来完成您的所有解决方案工作以及您想要包含的任何附加信息(例如:绘图、图表、计算、表格)。部分解决方案被接受。

我们允许谨慎使用 ChatGPT 等人工智能,尽管没有必要为此问题创建解决方案。如果您选择使用生成式 AI,则必须遵循[COMAP AI 使用政策](#)。这将导致额外的 AI 使用报告,您必须将其添加到 PDF 解决方案文件的末尾,并且不计入解决方案的 25 页总页数限制。

提供的文件:

- [Wimbledon_featured_matches.csv](#) – 2023 年温布尔登第二轮之后绅士单打比赛的数据集。
- [data_dictionary.csv](#) – 数据集的描述。
- [data_examples](#) – 帮助理解所提供数据的示例。

词汇表

大满贯:网球大满贯是指在一个日历年内赢得一个项目的全部四个大满贯冠军的成就。四大满贯赛事分别是澳大利亚网球公开赛、法国网球公开赛、温布尔登网球公开赛和美国网球公开赛,每场比赛都为期两周。

关键术语/概念词汇表:

-得分: [3]

o 比赛: 五盘三胜 (温布尔登的绅士比赛) o 盘数: 比赛集合; 6局获胜一盘,但在进行决胜局时,选手必须以两局获胜,直到盘数为6-6平 (见下文)

o 游戏: 收集积分; 玩家达到 4 分时获胜,但必须以 2 分获胜。请参阅下面的“比赛得分”。

-比赛得分: [3]

o 0 分 = 爱

o 1 分 = 15

o 2 分 = 30

o 3 分 = 40

o 并列分数 = 全部 (例如,“全部 30 分”)

o 40 - 40 = Deuce (玩家赢得相同数量的分数,至少 3 分每个)

o 服务器赢得平分 = Ad-in (或“优势”)

o 接收者赢得平分 = 广告输出

-发球: 球员作为“发球员”交替比赛 (击中第一球的球员)

点)和“返回者”。在职业网球比赛中,发球者往往具有很大的优势。球员在每个点上有两次发球机会将球投入比赛 (进入“发球区”)。

在比赛中两次尝试未能击中发球将被视为“双误”,返回的球员将获得该分。

o 破发 - 当返回的球员赢得比赛时。

o 破发点 如果回攻者获胜,他们将赢得比赛。

o 保持发球 当发球选手赢得比赛时。

-决胜局: 当一方赢得 6 局比赛且至少领先两局 (即 6 - 4) 时,每盘结束。如果没有,比赛将继续进行,直至打成 6-6 平局。此时进行决胜局。在温布尔登,决胜局是先得 7 分 (必须以 2 分获胜),但在比赛的第 5 盘中,先得 10 分 (必须以 2 分获胜)。

-休息时间/球场两侧: 球员在第一场比赛后以及每两场比赛后交换球场两侧。从第三场比赛开始,每次换边都允许有 90 秒的休息时间。在决胜局中,球员每六分就换边。每局比赛结束后,选手们也休息至少 2 分钟。允许医疗暂停和一次上厕所。

参考：

[1] Braidwood, J. (2023), 诺瓦克·德约科维奇创造了一个独特的对手 温网失利是结束的开始,《独立报》, <https://www.independent.co.uk/sport/tennis/novak-djokovic-温布尔登决赛卡洛斯-阿尔卡拉斯-b2376600.html>.

[2] <https://www.merriam-webster.com/dictionary/momentum>

[3] Rivera, J. (2023), 网球计分, 解释: 温布尔登规则术语和积分系统理解指南, 体育新闻, <https://www.sportingnews.com/us/tennis/news/tennis-评分解释规则系统积分-terms/7uzp2evdhbd11obdd59p3p1cx>.

帮助理解数据集的示例

示例 1:第 5 行

列)	值 “2023-	描述
比赛编号	wimbledon-1301”	“1301”中的 3 表示第 3 轮比赛，“01”表示该轮列出的第一场比赛。
经过时间	“0:01:31”	该分从比赛第一分开始后 1 分 31 秒后的发球开始。
点号、游戏号、设置号 （“不”是缩写 数字）	4, 1, 1	所打的分是本场比赛第一盘第一局的第4分。
p1_集、 p2_集、 p1_游戏、 p2_游 戏	0, 0, 0, 0	由于这是比赛的第一局,双方都还没有赢得一局或一盘。
p1_分数、 p2_分数	15, 30	打出这一点时的得分为 15（玩家 1)到 30（玩家 2）。因此,玩家 1 赢得了之前的一分,玩家 2 赢得了两分。
服务器	1	选手 1（阿尔卡拉斯)在这一点上发球。
服务号	1	这一分是在一发发球时打出的,这意味着阿尔卡拉兹在比 赛中击中了他的第一个发球局。
point_victor		Alcaraz 赢得这一分（玩家 1）。
p1_points_won, p2_points_won	1 2, 2	玩家 1（阿尔卡拉斯)是得分获胜者,因此他在比赛中的总得分现在 为 2（之前为 1）。对于玩家 2,该值仍为 2,因为玩家 2 失去了分 数。
游戏胜利者,设置胜利者	0, 0	阿尔卡拉斯赢得这一分后,整场比赛的比分变为 30 - 30（每人 2 分）,因此任何一位选手在这一分上都没有赢得一局或一盘（均 = 0）。
U – AC 列p1_winner		让我们来确定这一点是如何赢得的:
	1	阿尔卡拉兹凭借一记 “不可触碰”的击球赢得了这一分。
p1_ace	0	这次击球不是发球（因为 = 0）。
Winner_shot_type	F	这次击球是正手击球（而不是正手击球） 反手）。
p2_net_pt	1	球员 2（贾里)在这一点中将自己放置在靠近球网的位置。
p2_net_pt_won	0	由于阿尔卡拉兹赢得了这一分,尽管贾里在这一分期间在网前,但该 值仍为 0。
AH – AM 列	全部 = 0	即使选手 2 赢得了这一分,比赛也不会结束,因此该分不是 “破发点” , 这些都是 0。
p1_距离_运行、 p2_ 距离_运行	51.108, 75.631	每个玩家在该点跑动的距离（以米为单位）。
集会计数	13	双方球员在该得分期间击球的次数合计。
速度每小时、服务宽度、服务深度、 返回深度	130、BW、CTL、D	阿尔卡拉斯（发球手)在接发球球员的 “身体/边线”发球中打出了 130 分（我们之前看到这是一发）,并且接近表示比赛中或比赛外的线。贾 里（接球手)将球回击到球场 “深处”（靠近球场的另一端）。

示例 2:第 8 – 12 行

第一局比赛的最后四分说明了平局（“deuce”）和优势（“ad”）的概念。
每行都是比赛中的后续时间点。

排	列	值	描述
第 8 排	p1_score、p2_score	40, 40	分数为 40 – 40,表示每位玩家之前已赢得 3 分（也称为“平分”）。
	point_victor	1	阿尔卡拉兹赢得第 7 分（第 8 排）。
第 9 排	p1_score, p2_score	公元 40 年	由于阿尔卡拉兹赢得了上一分（第 7 分）,因此现在第 8 分的分数对于阿尔卡拉兹来说是“AD”,对于贾里来说是“40”,这意味着阿尔卡拉兹又赢得了一分,并且可以在下一分上赢得比赛。
	胜利点		Jarry（玩家 2）赢得第 8 分（第 9 行）。
第 10 行	p1_score、p2_score	2 40, 40	比分回到 40 - 40（“平分”）,这意味着每位玩家之前赢得的分数相同,但现在每人 4 分。
	胜利点	1	阿尔卡拉兹赢得第 9 分（第 10 排）。
第 11 行	p1_score、p2_score	AD,40	阿尔卡拉兹再次占据优势,赢得了第9分。
	胜利点	1	Alcaraz 赢得了第 10 分（第 11 行）,这意味着他赢得了比赛（现在又获得了 2 分）。
第 12 行	game_no	2	这是第二场比赛的第一分。
	p1_games	1	阿尔卡拉兹赢得了第一场比赛。

示例 3:第 51 行

比赛的第51分说明了“破发点”,即未发球的球员（接发球的球员）有机会赢得比赛的点。

排	列	值	描述
第 51 行	p1_score、p2_score	40,30	比分是 40 - 30,这意味着玩家 1 (Alcaraz) 领先。
	服务器	2	贾里（选手 2）正在发球。
	p1_break_pt	1	如果阿尔卡拉兹赢得一分,他将赢得比赛;因为他没有发球,所以这是一个“断点”。
	point_victor	1	阿尔卡拉兹赢得了这一分（因此也赢得了比赛）。
	p1_break_pt_won	1	阿尔卡拉兹赢得了比赛并且没有在发球点上发球。

COMAP 竞赛中大型语言模型和生成式 AI 工具的使用

这项政策的推动因素是大型语言模型 (LLM) 和生成式人工智能辅助技术的兴起。该政策旨在为团队、顾问和法官提供更大的透明度和指导。该政策适用于学生工作的各个方面,从模型的研究和开发 (包括代码创建)到书面报告。由于这些新兴技术正在迅速发展,COMAP 将酌情完善这一政策。

团队必须对人工智能工具的所有使用保持开放和诚实。团队及其提交的内容越透明,他们的工作就越有可能被其他人完全信任、欣赏和正确使用。这些披露有助于理解智力工作的发展并适当承认贡献。如果没有公开、明确地引用和参考人工智能工具的作用,有问题的段落和作品更有可能被认定为抄袭并被取消资格。

解决问题并不需要使用人工智能工具,尽管负责任地使用它们是允许的。COMAP 认识到法学硕士和生成式人工智能作为生产力工具的价值,可以帮助团队准备提交的材料;例如,生成结构的初步想法,或者在总结、释义、语言润色等时。在模型开发的许多任务中,人类的创造力和团队合作至关重要,而对人工智能工具的依赖会带来风险。因此,我们建议在使用这些技术执行模型选择和构建、协助创建代码、解释模型数据和结果以及得出科学结论等任务时务必谨慎。

值得注意的是,法学硕士和生成式人工智能都有局限性,无法取代人类的创造力和批判性思维。COMAP 建议团队如果选择使用法学硕士,请注意这些风险:

- 客观性:之前发布的包含种族主义、性别歧视或其他偏见的内容可能会受到影响。
出现在法学硕士生成的文本中,一些重要的观点可能没有得到体现。
- 准确性:法学硕士可以“产生幻觉”,即生成虚假内容,尤其是在使用时在他们的领域之外或处理复杂或模糊的主题时。他们能生成在语言上合理但在科学上不合理的内容,他们可能会弄错事实,并且已被证明会生成不存在的引用。一些法学硕士仅接受特定日期之前发布的内容的培训,因此呈现的情况不完整。
- 语境理解:法学硕士无法将人类的理解应用到特定的语境中。
一段文字,尤其是在处理惯用表达、讽刺、幽默或隐喻语言。这可能会导致生成的结果出现错误或误解内容。
- 训练数据:法学硕士需要大量高质量的训练数据才能实现最佳性能。然而,在某些领域或语言中,此类数据可能不容易获得,从而限制了任何输出的有用性。

团队指导

团队必须：

- 1.在报告中明确指出法学硕士或其他人工智能工具的使用情况,包括哪些使用模型以及用于什么目的。请使用内联引用和参考文献部分。还要在 25 页的解决方案之后附上人工智能使用报告（如下所述）。
- 2.验证内容和任何引用的准确性、有效性和适当性
由语言模型生成并纠正任何错误或不一致。
- 3.按照此处提供的指导提供引文和参考文献。再检查一遍
引文以确保它们准确且被正确引用。
- 4.意识到抄袭的可能性,因为法学硕士可能会复制大量文本
来自其他来源。检查原始来源以确保您没有抄袭别人的工作。

当我们发现提交的材料可能未公开使用此类工具时，
COMAP 将采取适当的行动。

引文和参考方向

仔细考虑如何记录和引用团队可能选择使用的任何工具。各种风格指南开始纳入人工智能工具的引用和参考政策。使用内联引用并列 出 25 页解决方案的参考部分中使用的所有 AI 工具。

无论团队是否选择使用人工智能工具,主要解决方案报告仍然限于25页。如果团队选择利用人工智能,请在报告结束后添加一个名为“人工智能使用报告”的新部分。这个新部分没有页数限制,不会计入 25 页解决方案的一部分。

示例（这并不详尽- 请根据您的情况调整这些示例）：

人工智能使用报告

- 1.OpenAI ChatGPT （2023年11月5日版本,ChatGPT-4）
 查询1： <插入您在AI工具中输入的确切措辞>
 输出： <插入 AI 工具的完整输出>
- 2.OpenAI Ernie （2023年11月5日版本,Ernie 4.0）
 查询1： <将任何后续输入的准确措辞插入AI工具>
 输出： <插入第二个查询的完整输出>
- 3.Github CoPilot （2024年2月3日版本）
 查询1： <插入您在AI工具中输入的确切措辞>
 输出： <插入 AI 工具的完整输出>
- 4. Google Bard （2024年2月2日版本）
 查询： <插入查询的确切措辞>
 输出： <插入 AI 工具的完整输出>