**2024美赛C题参考翻译**

**题目：网球的气势**



**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年以来在温布尔登的首次失利，也终结了这位历史上最伟大的大满贯选手之一的辉煌战绩。**

**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要轻松取胜，因为他以6 - 1的优势赢了第一盘（7局中赢了6局）。然而，第二盘比赛气氛焦灼，最终****阿尔卡拉斯在抢七局中以7 - 6获胜。第三盘比赛与第一盘比赛结果相反，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开始似乎完全掌控了局面，但不知为何，比赛再次改变了方向，德约科维奇完全掌控了比赛以6 - 3赢得了比赛。第5盘也是最后一盘，开始德约科维奇连续的第四盘的优势，但比赛再次改变了方向，阿尔卡拉斯以6 – 4获胜。本次比赛的数据在“2023-wimbledon-1701”的“match\_id”中。您可以在“set\_no”第一列看到德约科维奇在第一盘领先时的所有得分。发生在似乎很有优势的球员上的令人难以置信的成绩波动，有时意味着可以赢得很多分数或者甚至比赛， 通常被归因于“气势”。**

**字典上对气势的定义是“通过运动或一系列事件获得的力量。”[2]在体育运动中，一个团队或运动员在比赛中可能会感到他们有气势，或“优势/力量”，但这种现象很难衡量。此外，如果它真实存在，我们也不清楚比赛中的各种事件是如何创造或改变气势的。**

**我们提供了2023年温布尔登男单前两轮之后的每一分数据。您可以选择使用其他玩家信息或者其他数据，但是你必须完整地记录这些信息的来源。使用信息来完成：**

* **建立一个模型，来描述得分发生时到地比赛流程，并且将其应用于一场或者多场的比赛中。您的模型应该可以确认出哪些球员在比赛的特定时间表现得更好，以及他们表现的有多好。基于您的模型，提供可视化来描述比赛流程。*注意：在网球的比赛中，发球的选手赢得得分/比赛的可能性要高很多，你可以将此因素以某种方式加入到您的模型中。***
* **一位网球教练对“气势“在比赛中发挥的作用保持怀疑态度。相反，他认为一个选手在比赛中的成绩的波动和如何取得成功是随机的。使用您的模型/指标来评估这种说法。**
* **教练们很想知道是否有哪些指标可以帮助确定，何时比赛过程会从偏向一个球员到偏向另一个球员。**
* **使用至少一场提供的比赛数据，模型来预测，比赛中的波动。哪些因素？似乎相关性最强（如果有的话）？**
* **考虑到过去比赛中“气势”波动的差异，您将被一名即将与另一名球员展开新的比赛的球员提供哪些建议？**
* **在一场或多场其他比赛中，测试你建立的模型对比赛结果的预测有多准确。如果模型有时表现不佳，您是否确定可能需要考虑模型外的其他因素？您建立的模型对其他比赛（如女子比赛），锦标赛，场地赛和其他运动如乒乓球的通用性如何？**
* **提交一份不超过25页的调查报告，并包括1-2页的备忘录。总结您的结果，并就“气势”的作用，向教练提出建议，以及如何让球员应对影响网球比赛过程中的事件。**

**您的总页数不超过25页的PDF解决方案应该包括：**

* **一页摘要**
* **目录**
* **您的完整的解决方案**
* **1-2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页的限制中）**

**注意：完整的MCM竞赛提交报告没有确定的最小页面长度要求。您可以最多使用总计25页的解决方案和任何您想包括的附加信息（例如：图示、图表、计算、表格）。我们允许谨慎地使用人工智能，如ChatGPT，尽管没有必要为这个问题创建一个解决方案。如果您选择使用生成AI，则必须遵守COMAP AI使用规则。这将产生一份额外的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总页数限制中。**

**提供的文件**

* ***Wimbledon\_featured\_matches.csv* – 2023年温布尔登男单第二轮后比赛数据集**
* ***data\_dictionary.csv* – 数据集的描述**
* ***data\_examples* – 为帮助理解所提供数据的实例**

**术语表**

**大满贯:网球的大满贯是在一个年内赢得一个项目的所有四个大满贯的成就。四项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，每项比赛都持续两周。**

**关键术语/概念词汇表**

* **计分：[3]**
* **整场比赛:五盘三胜制(温布尔登男子赛)**
* **盘：每盘比赛包括几局比赛，****赢得6局则拿下这盘比赛，但选手必须先赢两局，直到比赛打成6比6平局(见下文)**
* **局：每局比赛包括得分，赢得4分则拿下这局比赛，但必须有两分的优势。请参阅下面的“比赛得分”。**
* **比赛得分**
* **0分=love**
* **1分=15**
* **2分=30**
* **3分=45**
* **打成平手=All**
* **40-40=平局(玩家获得相同的分数，每人至少3分)**
* **发球者赢得了平局分=Ad-in（或者优势）**
* **接球者赢得了平局分=Ad-out**
* **发球:玩家在游戏中交替扮演“发球者”(最先击球的玩家)和“接发球者”。在职业网球比赛中，发球者往往有很大的优势。在每个点，球员有两次发球机会将球送入发球区。在比赛中连续两次发球失败是“双误”，回发球者得分。**
* **突破发球 - 当回发球者赢得一场比赛**
* **破发点 - 如果接发球者赢了，他们就会赢得比赛**
* **持球发球-当发球选手赢得比赛。**
* **决胜局:当一名球员赢了6局，只要他们领先至少2局(即6 - 4)，每一盘结束。如果没有，继续比赛，直到6 - 6打成平局。这时，决胜局开始了。在温布尔登决胜局中，先得7分(必须领先2分)，但在第5局比赛中先得10分时除外(必须领先2分)。**
* **休息时间/场边:球员在第一场比赛后更换场边，然后每两场比赛后更换场边。从第三场比赛开始，每次换防时，允许有90秒的休息时间。在决胜局中，球员每得6分换边。每盘比赛结束后，选手至少休息2分钟。医疗暂停和一次洗手间是允许的**

**参考文献**

**[1] Braidwood, J. (2023), Novak Djokovic has created a unique rival – is Wimbledon defeat the beginning of the end, The Independent, https://www.independent.co.uk/sport/tennis/novak-djokovic-wimbledon-final-carlos-alcarazb2376600.html.**

**[2]** [**https://www.merriam-webster.com/dictionary/momentum**](https://www.merriam-webster.com/dictionary/momentum)

**[3] Rivera, J. (2023), Tennis scoring, explained: A guide to understanding the rules terms & point system at Wimbledon, The Sporting News, https://www.sportingnews.com/us/tennis/news/tennis-scoring-explained-rules-system-pointsterms/7uzp2evdhbd11obdd59p3p1cx.**

**帮助理解数据集的例子**

**例子1：第5行**

|  |  |  |
| --- | --- | --- |
| **列** | **值** | **描述** |
| match\_id | “2023-wimbledon-1301” | “1301”中的3表示第3轮匹配  " 01 "表示该轮中列出的第一个匹配项。 |
| elapsed\_time | **“0:01:31”** | 在比赛的第一分开始后1分31秒，发球开始得分 |
| point\_no, game\_no, set\_no (“no” is an abbreviation for number) | **4,1,1** | 打的点是第一局的第4点  比赛第一局。 |
| p1\_sets, p2\_sets, p1\_games, p2\_games | **0,0,0,0** | 由于这是比赛的第一局，双方都还没有赢过一局。 |
| p1\_score, p2\_score | **15,30** | 玩家1的得分为15分，玩家2的得分为30分。因此，玩家1赢得1分，玩家2赢得2分。 |
| **srver** | **1** | 玩家1 (阿尔卡拉斯)在这一分上发球 |
| **Sever\_no** | **1** | 这一分是按第一发球的意思打的  阿尔卡拉斯在比赛中打出了他的第一个发球。 |
| point\_victor | **1** | 阿尔卡拉斯赢得这一点(玩家1) |
| p1\_points\_won,p2\_points\_won | **2,2** | 玩家1 (阿尔卡拉斯)是积分胜利者，所以他的总分现在是2(之前是1)。对于玩家2来说，价值仍然是2，因为玩家2失去了积分。 |
| game\_victor, set\_victor | **0,0** | 阿尔卡拉斯赢得这一分使比赛的比分变为30比30(各得2分)，因此双方在这一分上都没有获胜(均为0)。 |
| Columns U – AC |  | 让我们来看看这一分是怎么赢的: |
| p1\_winner | **1** | 阿尔卡拉斯以一记“不可触碰”的击球赢得了这一分 |
| p1\_ace | **0** | 这个球不是发球(since = 0) |
| winner\_shot\_type | **F** | 这是一个正手击球(而不是反手击球)。 |
| p2\_net\_pt | **1** | 玩家2 (杰瑞)在这个点的某个位置靠近球网。 |
| p2\_net\_pt\_won | **0** | 由于阿尔卡拉斯赢得了这一分，虽然在这一分中杰瑞在网前，但这个值是0。 |
| Columns AH – AM | **All=0** | 即使玩家2赢得了这个点，游戏也不会结束，所以这个点不是“断点”，这些都是0 |
| p1\_distance\_run, p2\_distance\_run | **51.108,75.631** | 每个运动员在这一点上跑的距离(以米为单位) |
| rally\_count | **13** | 两名球员在该点的总击球数。 |
| speed\_mph, serve\_width, serve\_depth, return\_depth | 130, BW, CTL, D | 阿尔卡拉斯发球方)打了一个130发的“身体/宽度”发球(我们之前看到过这是第一次发球)，并且接近表示比赛中或出局的线。雅里  (回击者)把球“深”回了场  (那么靠近球场的另一端) |

**例子2：第8-12行**

第一场比赛的最后四分说明了平局(deuce)和优势(ad)的概念。

每一行都是比赛中的一个后续时间点。

|  |  |  |  |
| --- | --- | --- | --- |
| **行** | **列** | **值** | **描述** |
| **第8行** | p1\_score, p2\_score | **40，40** | 比分是40比40，意味着每个玩家都赢得了3分  (这也被称为“deuce”)。 |
|  | point\_victor | **1** | 阿尔卡拉斯赢得第7分(第8连胜) |
| **第9行** | p1\_score, p2\_score | **AD，40** | 由于阿尔卡拉斯赢得了前一分(第7点)，因此第8点的得分现在是“AD”，而杰瑞的得分是“40”，这意味着阿尔卡拉斯又赢得了一分，可能会赢得下一分 |
|  | point\_victo | **2** | 杰瑞(玩家2)赢得第8点(第9行)。 |
| **第10行** | p1\_score, p2\_score | **40，40** | 分数回到40比40(“平分”)，这意味着每个玩家之前赢得的分数相同，尽管现在是每人4分 |
|  | point\_victo | **1** | 阿尔卡拉斯赢得第9分(第10行)。 |
| **第11行** | p1\_score, p2\_score | **AD,40** | 阿尔卡拉斯赢得了第9分，再次占据优势。 |
|  | point\_victo | **1** | 阿尔卡拉斯赢得第10点(第11行)，这意味着他赢得了比赛  (现在又得了两分)。 |
| **第12行** | game\_no | **2** | 这是第二局的第一分 |
|  | p1\_games | **1** | 阿尔卡拉斯赢了第一场。 |

**例子3：第51行**

比赛的第51分说明了“破发点”——在这个点上，没有发球的球员(正在回发球的球员)有机会赢得比赛。

|  |  |  |  |
| --- | --- | --- | --- |
| **行** | **列** | **值** | **描述** |
| **第51行** | p1\_score, p2\_score | **40，30** | 比分是40比30，意味着玩家1 (阿尔卡拉斯)领先 |
|  | **Sever** | **2** | 杰瑞(玩家2)发球。 |
|  | p1\_break\_pt | **1** | 如果阿尔卡拉斯赢得这一分，他将赢得比赛;由于他没有发球，这是一个“断点”。 |
|  | point\_victor | **1** | 阿尔卡拉斯赢得了分数(因此赢得了游戏)。 |
|  | p1\_break\_pt\_won | **1** | 阿尔卡拉斯赢得了这场比赛，并没有发球 |

**在COMAP竞赛中使用大型语言模型和生成式AI工具**

**这一政策是由大型语言模型(llm)和生成式人工智能辅助技术的兴起推动的。该政策旨在为团队、顾问和评委提供更大的透明度和指导。这项政策适用于学生工作的所有方面，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将适当地完善这一政策。**

**团队必须公开和诚实地使用人工智能工具。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、赞赏和正确使用。这些披露有助于理解智力工作的发展和正确承认贡献。如果没有对人工智能工具作用的公开和清晰的引用和参考，有问题的段落和工作更有可能被认定为抄袭并被取消资格。**

**解决这些问题不需要使用人工智能工具，尽管允许负责任地使用它们。COMAP认识到大型语言模型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;为一个结构产生最初的想法，例如，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是必不可少的，并且依赖于人工智能工具带来了风险。因此，我们建议在将这些技术用于模型选择和构建等任务时要谨慎。**

**值得注意的是，大型语言模型和生成式人工智能有局限性，无法取代人类的创造力和批判性思维。COMAP建议团队在选择大型语言模型时要意识到这些风险。**

* **客观性:** **大型语言模型生成的文本中可能出现先前发表的包含种族主义，性别歧视或其他偏见的内容，并且可能没有代表一些重要的观点。**
* **准确性:大型语言模型可以产生“幻觉”，即产生虚假内容，特别是在他们的领域之外使用或处理复杂或模棱两可的主题时。它们可以生成语言上可信但科学上不可信的内容，它们可以误解事实，而且它们已经被证明会生成不存在的引用。一些大型语言模型只接受特定日期之前发布的内容的培训，因此呈现的是不完整的画面。**
* **语境理解:** **大型语言模型不能将人类的理解应用到一篇文章的语境中，特别是在处理习惯表达、讽刺、幽默或隐喻语言时。这可能导致生成的内容出现错误或误解。**
* **训练数据:LLMS需要大量高质量的训练数据来达到最佳性能。但是，在某些领域或语言中，这些数据可能不容易获得，从而限制了任何输出的有用性。**

**参赛队伍须:**

**1. 在报告中明确指出大型语言模型或其他人工智能工具的使用情况，包括使用哪种模型以及用于什么目的。请使用内联引用和参考部分。在你的25页解决方案之后，还要附上人工智能使用报告(如下所述)。**

**2. 验证内容的准确性、有效性和适当性以及由语言模型生成的任何引用，并纠正任何错误或不一致之处。**

**3. 提供引用和参考，遵循这里提供的指导。仔细检查引用，确保它们是准确的，并被适当引用**

**引文及参考指引**

**仔细考虑如何记录和引用团队可能选择使用的工具。各种风格指南开始纳入人工智能工具的引用和参考政策。在25页的解决方案的参考部分，使用内联引用并列出所有使用的AI工具。**

**无论团队是否选择使用AI工具，主要解决方案报告仍然限制在25页。如果一个团队选择使用人工智能，在你的报告结束后，添加一个新的标题为人工智能使用报告。这个新的部分没有页数限制，也不会被计算在25页的限制中。**