**•开发一个模型，捕捉得分发生时的游戏流程，并将其应用于一场或多场比赛。你的模型应该识别出哪名球员在比赛的特定时间表现更好，以及他们的表现有多好。提供基于您的模型的可视化来描述match flow。注:在网球比赛中，发球的选手赢得得分/比赛的可能性要大得多。您可能希望以某种方式将此因素纳入您的模型中。**

（量化表现得分，可视化随match的发展，表现得分的变化）

详细：（chat笼统总结在下面，可以先看那个）

**特定时间的表现可以用此时选手的状态来表现，主要是对选手此时的评价，需要建立多级指标的评价体系，一锅烩容易使得评价分数存在不透明度，于是依靠具体数据、现有研究将状态表现分为四方面：心态分、技巧分、 体力分、发球分，还有一个其他项为扰动项。进行评分，由于数据中扰动项的缺失和模糊，比如：天气对每位球员的主观影响难以评价，具有很强的主观性，因此我们对此扰动项进行随机生成，提高在一定外界影响下对选手表现评价的客观性和容错率。**

**将题目所给变量数据按照论文分为以上四个方面**

**用随机森林模型得到每个方面的每个变量的权重，根据每一方面对一次game的结果的影响得到每个方面内每个变量的重要程度**

**（具体的数据处理：**

**由xx文献/资料我们知道影响较大的是得分点/失分点处之前几次得分点发生的事情。于是设置滑动窗口，步长为1，窗口大小为8，探究影响心态的具体指标，我们的想法是：因为现在有每个得分时间点的各种球（如制胜球、ace球、双球误发等）的数量，这些数量对比赛双方的心态影响是短时的，match1set1game1的得分智慧短暂影响到match1内的之后的那些比赛，一般不可能影响到下一个match2set1game1，**

**由于人的心态只跟最近的几场数据和长程的数据情况有关，比如一个match内match越高，match的得分率影响更高，一个set内set越高，set的得分率影响更高。**

**长期的选手set得分比例对信心（表现的一部分）的影响程度是长期的，变化较快的短期选手games得分的比例对选手的表现影响是短暂的，（处理数据处理成窗口数据）0 15 30 40 抢七1-9这个用不到，只在数据特征描述的时候提一下数据异常我们注意到了但是数据无误就行。**

**然后技巧方面由于发球深度、方向等变量是符号表示，实际意义、包括对结果的影响很难通过现有公开去表征一名专业运动员的水平，我们通过询问chatgpt专家得到其意见，给出符号变量的赋值，（怎么解释看chatgpt）**

**正反手攻球算个人习惯，不算技巧，且数据正反手的比例并不协调，容易引喻失义**

**由于休息机制到位，每场的match、set、game number\_no这些我们不考虑会影响强壮的男单决赛选手们的体力，甚至表现**

**每个数据都需要标准化和归一化，以消除量纲影响。）**

**然后根据权重值和原有指标，得出每条时间点数据的新的四个方面评分的值：****心态分、技巧分、 体力分、发球分和其他扰动项（比如外界环境）**

**然后继续用随机森林模型，通过新的评分对最终结果进行预测，得到每个评分的权重（相当于是把原数据变成这四个指标数据，对最终结果进行预测，这种观念也符合我们通过上升数据，总结归纳定性中间量，从而使预测更合理、准确的基本价值观巴拉巴拉）**

**最终得到一个预测结果的公式，其具有很好的可解释性**

**然后做交叉验证，用多个模型解释我们建模的合理性，比如我们在不分评价指标：心态分、技巧分、 体力分、发球分和其他扰动项的时候，使用随机森林得到的预测值明显低于我们分完类的预测值**

**确定所有权重之后，对预测的实时得分（表现）进行可视化即可：就是根据公式代入数据给出预测的实施得分，然后绘制折线图**

**评价体系建立：**

* **建立多级指标的评价体系，将特定时间的球员表现分为四个方面：心态、技巧、体力、发球，并加入一个扰动项。特定时间点的各个表现进行全面而细致的刻画。**
* **扰动项的引入通过随机生成，以提高在外界影响下对选手表现的客观性和容错率。**

**变量数据处理：**

* **利用滑动窗口和步长进行数据处理，探究影响心态的具体指标。**
* **通过心态的具体指标，考虑不同得分时间点的影响，考虑短时和长时的数据动态。**

**技巧方面的处理：**

* **技巧方面的变量包括发球深度、方向等，是符号表示。**
* **通过与ChatGPT专家讨论，为符号变量赋值，以获得实际意义和对比赛结果的影响。**

**数据标准化和归一化：**

* **对所有数据进行标准化和归一化，以消除量纲影响，确保各项指标的可比性。**

**权重计算和模型训练：**

* **利用随机森林模型，计算每个方面和每个变量的权重，以了解其对比赛结果的影响程度。**
* **通过模型训练，得到每个方面内每个变量的重要性。**

**预测公式的建立：**

* **根据权重值和原有指标，建立每个时间点数据的新的四个方面评分的值。**
* **使用随机森林模型预测最终比赛结果，得到每个评分的权重。**

**合理性解释和模型验证：**

* **在不分评价指标的情况下，使用随机森林得到的预测值与分完类的预测值进行对比，以验证模型合理性。**
* **利用多个模型对建模过程进行解释，确保模型的准确性和可解释性。**

**可视化展示：**

* **根据预测结果的公式，将实时得分（表现）进行可视化展示。**
* **利用折线图等方式展示不同评价指标对预测结果的影响，验证模型的可解释性和合理性。**