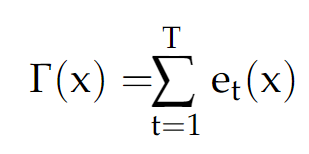
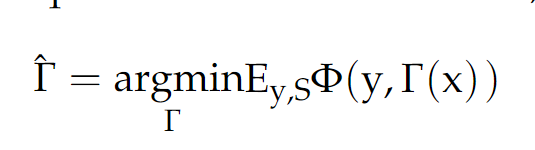
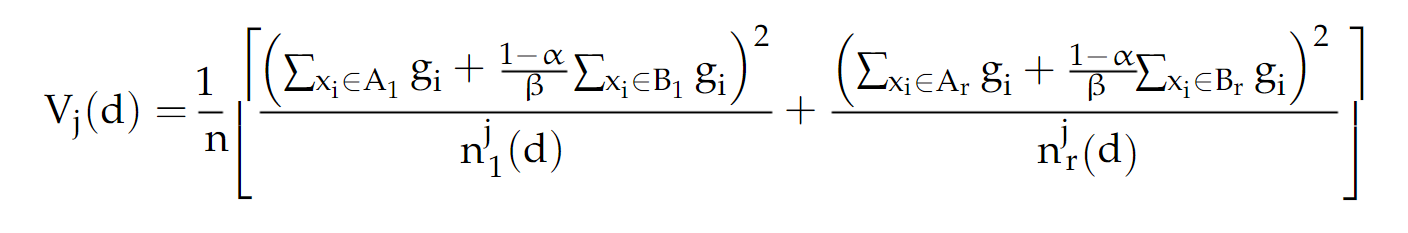
LightGBM, a variant of the gradient boosting decision tree (GBDT), builds upon decision tree algorithms with a leaf-wise splitting strategy. Unlike the traditional level-wise approach, the leaf-wise method in LightGBM efficiently minimizes loss during tree growth, leading to significantly enhanced classification accuracy compared to other established boosting algorithms. LightGBM introduces two innovative techniques: Exclusive Feature Bundling (EFB) and Gradient-based One-Side Sampling (GOSS).

The prediction of a Gradient Boosting Decision Tree (GBDT), denoted as T(x), is obtained by summing the outputs of a set of decision tree models, expressed as e(x), according to Equation (6).

To construct a GBDT model that effectively fits the given loss function Ф(y, T(x)), the aim is to determine the approximate function Γ that minimizes the loss.

LightGBM diverges from the traditional GBDT approach by employing the Gradient-based One-Side Sampling (GOSS) method to split internal nodes instead of relying on information gain. The specific formula for this splitting strategy is provided as follows.



For a detailed explanation, please refer to the citation [1] in the paper."

卡洛斯·阿尔卡拉兹 诺瓦克·德约科维奇2

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岁的诺瓦克·德约科维奇2。这是德约科维奇2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。S

这场比赛本身就是一场非凡的战斗。[1]德约科维奇2似乎注定要轻松获胜，因为他以6比1控制了第一局(7局中赢了6局)。然而，第二盘比赛气氛紧张，最终阿尔卡雷斯在抢七局中以7 - 6获胜。第三局与第一局相反，阿尔卡拉斯1以6 - 1轻松获胜。这位年轻的西班牙人在第四盘开始时似乎完全控制了比赛，但不知怎么的，比赛再次改变了方向，德约科维奇2完全控制了比赛，以6比3赢得了比赛。第五盘也是最后一盘，德约科维奇2从第四盘开始保持优势，但再次改变方向，阿尔卡拉兹获得控制并以6比4获胜。本次比赛的数据在提供的数据集中，“match\_id”为“2023-wimbledon-1701”。你可以用“set\_no”列= 1看到德约科维奇2在第一盘领先时的所有得分。令人难以置信的挥杆，有时是在许多分甚至是比赛中，发生在似乎有优势的球员身上，通常被归因于“势头”。字典上对动量的一个定义是“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动力，或“力量/力量”，但这种现象很难衡量。此外，我们也不清楚比赛中的各种事件是如何创造或改变动量的。

2023年温布尔登男单前两轮之后的每一分数据。你可以选择包含额外的玩家信息或其他数据，但你必须完整地记录这些来源。

1. **•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匹配流。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。**
2. **•一位网球教练怀疑“势头”在比赛中是否起作用。相反，他假设一个玩家在游戏中的波动和成功的运行是随机的。使用你的模型/指标来评估这种说法。**
3. **•教练们很想知道是否有一些指标可以帮助确定比赛流程何时会从一个球员转变为另一个球员。**
   1. **利用至少一场比赛提供的数据，建立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)?**
   2. **问:考虑到****过去比赛中“动量”波动的不同，你如何建议一名球员与另一名球员进行新的比赛?**
4. **•在一个或多个其他比赛中测试你开发的模型。你对比赛结果的预测有多准确?如果模型有时表现不佳，您能否确定可能需要包括在未来模型中的任何因素?您的模型对其他比赛(如女子比赛)、锦标赛、球场表面和其他运动(如乒乓球)的通用性如何?**
5. **Develop a model that captures the flow of play as points occur and apply it to one or more of the matches. Your model should identify which player is performing better at a given time in the match, as well as how much better they are performing. Provide a visualization based on your model to depict the match flow. Note: in tennis, the player serving has a much higher probability of winning the point/game. You may wish to factor this into your model in some way.**
6. **• A tennis coach is skeptical that “momentum” plays any role in the match. Instead, he postulates that swings in play and runs of success by one player are random. Use your model/metric to assess this claim.**
7. **• Coaches would love to know if there are indicators that can help determine when the flow of play is about to change from favoring one player to the other.** 
   1. **Using the data provided for at least one match, develop a model that predicts these swings in the match. What factors seem most related (if any)?**
   2. **Given the differential in past match “momentum” swings how do you advise a player going into a new match against a different player?**
8. **• Test the model you developed on one or more of the other matches. How well do you predict the swings in the match? If the model performs poorly at times, can you identify any factors that might need to be included in future models? How generalizable is your model to other matches (such as Women’s matches), tournaments, court surfaces, and other sports such as table tennis.**

**•提交一份不超过25页的调查报告，并包括一到两页的备忘录，总结你的结果，并就“势头”的作用向教练提出建议，以及如何让球员准备好应对影响网球比赛过程的事件。**

总页数不超过25页的PDF解决方案应包括:•一页摘要表。

•目录。

•完整的解决方案。

•一到两页的备忘录。

•参考文献列表。

•AI使用报告(如果使用不计入25页限制)注意:对于完整的MCM提交，没有特定的最小页面长度要求。您可以使用最多25页的总解决方案工作和任何额外的信息，你想包括(例如:图纸，图表，计算，表格)。部分解被接受。

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，则必须遵循COMAP AI使用策略。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总页数限制。

•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—帮助理解所提供数据的示例。

大满贯:网球的大满贯是指在一年中赢得一个项目的所有四项主要冠军。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:-得分:[3]o比赛:五局两胜(温布尔登男人赛)o局:比赛的集合;6局赢一局，但选手必须先赢两局，直到双方以6比6打平(见下文)。当达到4分时获胜，但必须赢得2分。请参阅下面的“为游戏得分”。

——得分游戏:[3]o =爱o 1点= 0点15 o 2分= 30 o 3分= 40 o系分数=所有(例如,“30”)o 40 - 40 =平手(玩家获得相同数量的点,至少3分)o服务器赢得一个平手点=广告(或“优势”)o接收机赢得一个平手点=了广告-服务:球员交替游戏的“服务器”(球员达到最初的拍摄点)和“回击者”。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。

o破发发球——当回击球员赢得一场比赛时。

o破发点——在这个点上，如果回击者赢了，他们就赢得了比赛。

o握发球——当发球选手赢得比赛时。

-决胜局:当一方赢得6局时，只要他们领先至少2局(即6 - 4)，每一局结束。如果没有，继续比赛，直到6 - 6打成平局。这时，决胜局开始了。在温网比赛中，抢七是先赢7分(必须赢2分)，但在比赛的第5局，先赢10分(必须赢2分)除外。

-休息时间/场边:球员在第1场比赛后更换场边，然后每两场比赛更换一次。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)，诺瓦克·德约科维奇2创造了一个独特的对手——是温布尔登击败结束的开始，《独立报》，https://www.independent.co.uk/sport/tennis/novak-djokovic-wimbledon-final-carlos-alcarazb2376600.html。

[2] https://www.merriam-webster.com/dictionary/momentum [3] Rivera, J.(2023)，网球计分，解释:在温布尔登理解规则条款和计分系统的指南，体育新闻，https://www.sportingnews.com/us/tennis/news/tennis-scoring-explained-rules-system-pointsterms/7uzp2evdhbd11obdd59p3p1cx。