会议记录-2024/11/10

1. **会议内容和目标**

关于数据竞赛具体分工和一些需要讨论的事项

1. **参会人员：**

李科、林国栋、刘养明、韩重华、范冰冰

1. **选择数据竞赛：**

Child Mind Institute — Problematic Internet Use

<https://www.kaggle.com/competitions/child-mind-institute-problematic-internet-use/overview>

**背景介绍：**

在当今的数字化时代，儿童和青少年的**问题性互联网使用**是一个日益增长的关注点。更好地理解这一问题对于解决诸如抑郁和焦虑等心理健康问题至关重要。目前用于测量儿童和青少年问题性互联网使用的方法通常较为复杂，需要专业评估。这为许多家庭创造了获取、文化和语言障碍。由于这些限制，问题性互联网使用通常不是直接测量的，而是与青少年的抑郁和焦虑等问题相关联。相反，身体和健身测量方法非常容易获得，广泛可用，且干预或临床专业知识最少。在过度使用技术的用户中，常见的身体习惯变化包括姿势不良、饮食不规律和体育活动减少。我们建议使用这些容易获得的身体健康指标作为识别问题性互联网使用的代理，特别是在缺乏临床专业知识或合适的评估工具的背景下。这项竞赛挑战你**开发一个预测模型，能够分析儿童的身体活动数据，以检测问题性互联网和技术使用的早期指标**。这将使得及时干预成为可能，旨在促进更健康的数字习惯。你的工作将有助于创造一个更健康、更快乐的未来，在这个未来中，儿童能够更好地负责任地导航数字领域。

**总结**：通过两类特征（状态特征和时序特征），来预测问题性互联网使用程度。

一类数据是`HBN Instruments`，共有81维（一维是id），描述相应的状态，如性别，升高体重等。

**特点：**The majority of measures are missing for most participants。

一类数据是运动记录的数据。（时间序列）

**特点：**也并非所有id都有该类数据。

要预测的使用程度用sii分数（0~100）来表示，将其量化为0，1，2，3

**特点：**并非所有id都有标签（无监督）

初步想法：

两种类型数据分别给出一个结果，然后综合（pass掉）

最终结果的类别分布不均衡，需要考虑

1. **分工**

主要在于数据预处理，模型设置之后做。

（1）、先加一下team

（2）、讨论一下怎么分工

分工结果：按照要处理的数据类型

HBN：（3人）李科、林国栋、刘养明

Series：（2人）韩重华、范冰冰

1. **时间节点**

12月9号展示，还有4周。

该周计划：对于数据如何处理，给出具体的解决方案。以及最后如何总结起来。