1. 模型算法

我采用了 Lightgbm 算法对《数据预处理说明文档》中构建的数据进行训练、验证和预测。

用 14 组 Lightgbm 模型、14 个独立的数据集来训练和预测未来 14 天的病例数量，采用第 1 组 Lightgbm 模型和第 1 个数据集训练和预测未来第 1 天的病例数量；采用第 2 组 Lightgbm 模型和第 2 个数据集训练和预测未来第 2 天的病例数量；依次类推。14 个独立数据集采样的时间段都是 2017.11.1 以后所有时间。每组 Lightgbm 模型采用七折交叉验证的方法，使用该组数据集七分之六的数据进行训练，七分之一的数据进行验证，最后将该组七个模型对该组测试集的预测结果取平均，即可得到当天的病例数量预测结果。

2. 后处理方法

2020 年由于疫情的侵袭，上半年很多人不敢去医院，因此医院的病例数据大幅减少，count 文件给定的就诊病例数据也佐证了这一观点。因此我对于模型预测结果 2020 年的部分乘上一个系数，通过在 A 榜上的多次尝试，我决定该系数为 0.92，此时得到预测结果 final1.csv。

其次，为了将参加比赛以来历次提交的预测结果利用上，我使用 nudge 的后处理方法对结果进一步进行缩放。预测结果总计 2520 行，对应 30 个病种在 84 天的病例数量预测值。对于 final1.csv 2520 行中的每个预测值 A，与 86 个过往提交文件该位置的预测结果逐一进行比较。若 A 大于这些值的数量大于 83 个，则将 A 乘上 1.005；若 A 小于这些值的数量大于 83 个，则将 A 乘上 0.995。