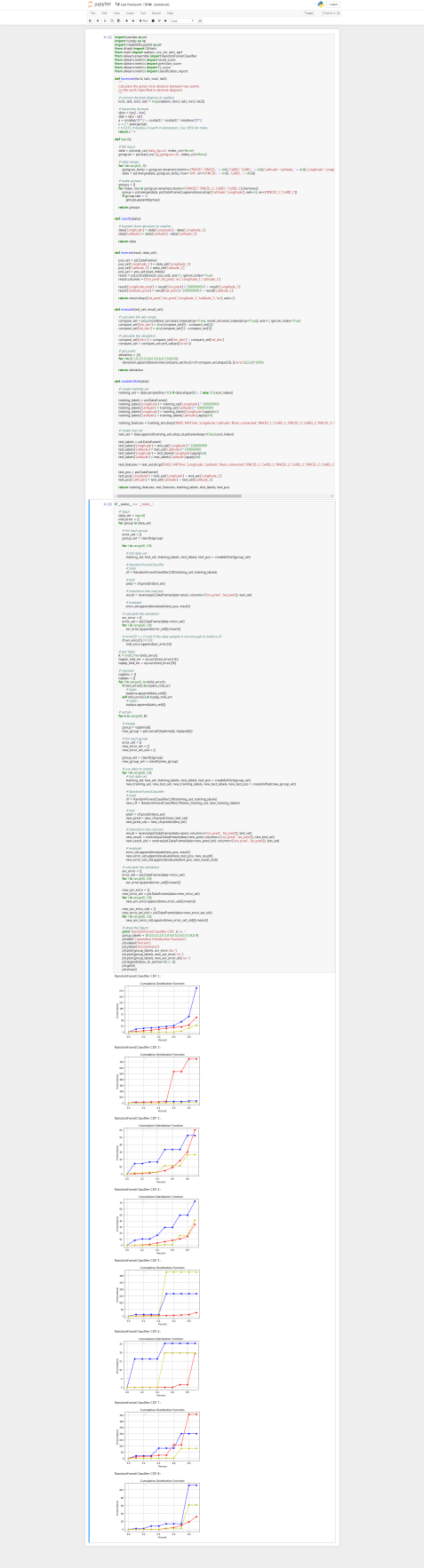
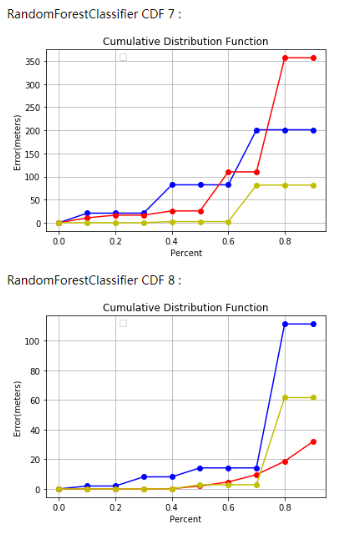
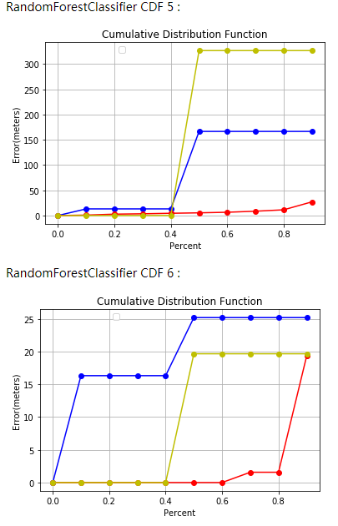
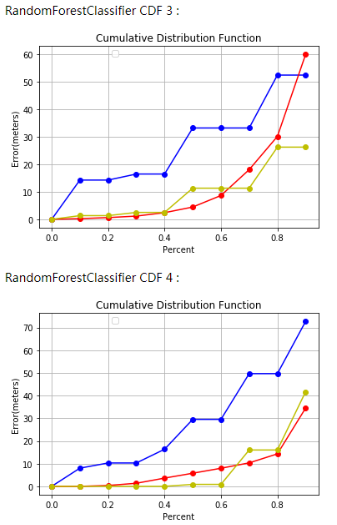
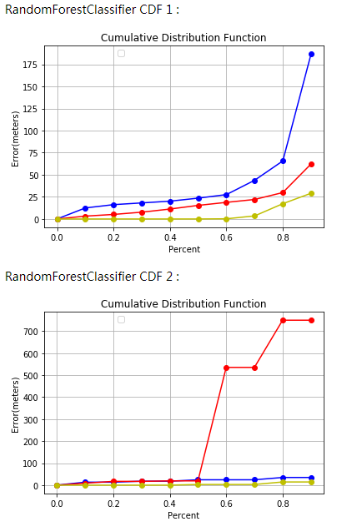
## 代码运行

[点击查看原图](d.png)



### 讨论分析

1. 本题中重复计算b中各分组的训练各10次并求取中位误差的平均值，进行比较，最后从中选取出了0.2的topk+最佳数据集模型以及0.2的topk-最差数据集模型，并将两者之中的各原数据集进行混合。之后利用新数据集作为训练集，新测试集和原测试集分别作为测试集，进行验证，以获得数据集混合对位置预测误差的影响。
2. 以下为结果的部分截图：



* 1. 以上为全部7组比较结果，蓝线为原始模型的误差曲线，红线为使用新数据集进行训练和测试获得的误差曲线，黄线为使用新数据集进行训练、旧数据集进行测试的误差曲线。
  2. 较为优秀的为模型1，在加入新数据后红黄两线的误差均大幅减小，而且黄线低于红线，表面新数据的加入确实改进了预测的结果，且这一点并不来自于加入数据的数量巨大。模型2、3、4、6、7、8等，同。
  3. 其他较为典型的为5，5是8个混合数据集中唯一黄线误差增大，而红线误差减小的。这表明在这一数据集中，首先旧数据量较少，极少，仅为2个，其次新数据量极大，最后在加入新数据之后由于新数据量远大于旧数据量，导致红线误差结果向加入的top+数据集的旧误差曲线接近，表现为误差变好，但是利用旧数据集进行验证之后，实际误差增大，表明数据混合起到了反作用。
  4. 这一点将在下一题中被改进。

### 性能比较

此处性能比较将比较新旧数据集的训练和验证耗时，主要自变量为数据集的大小。