

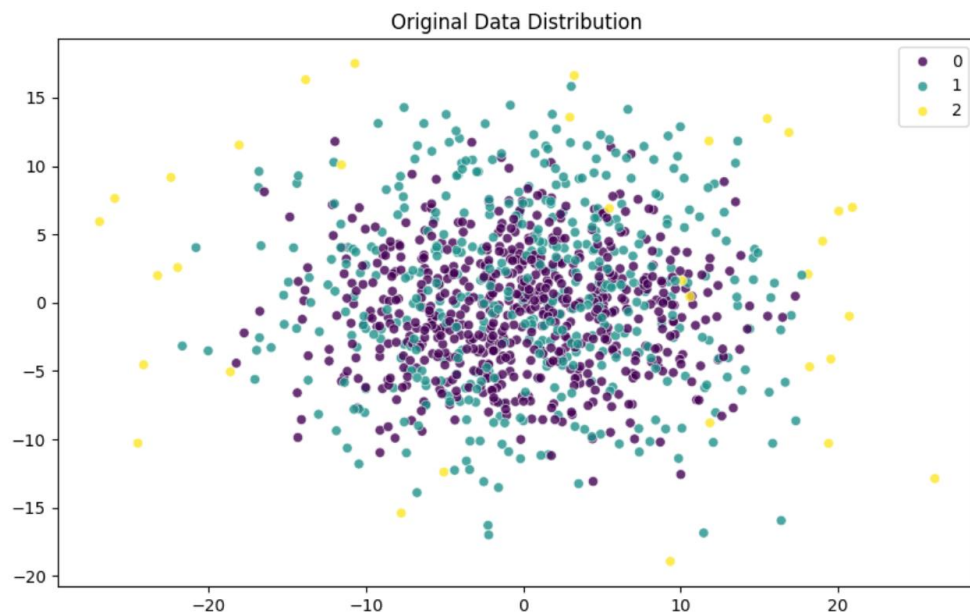
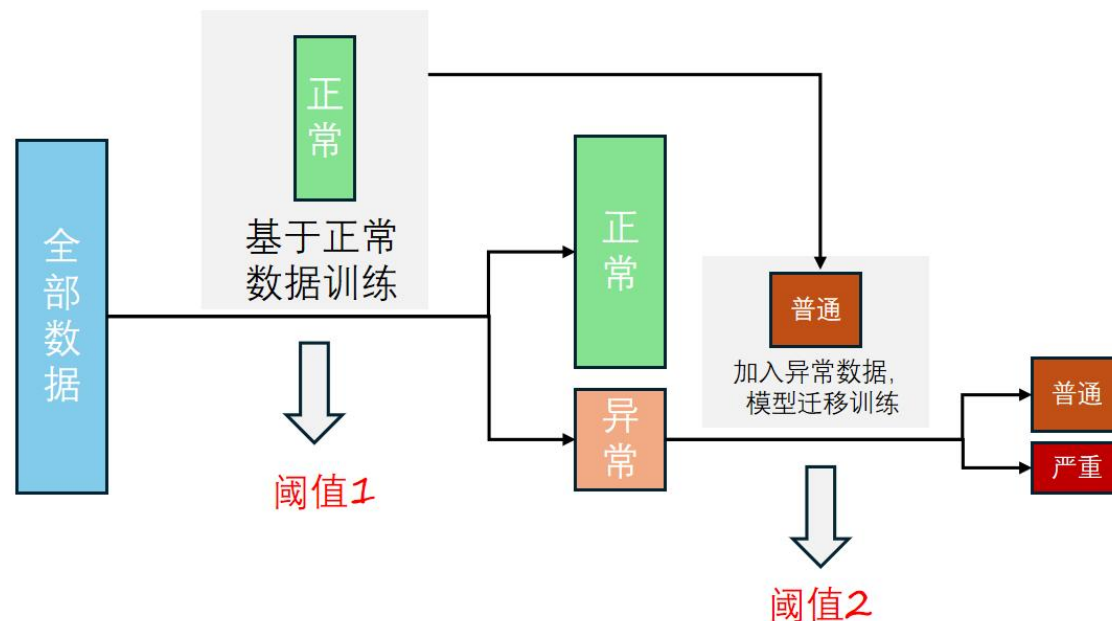
# 面向磁浮轨道异常检测的大数据分析框架研究

## 小组汇报

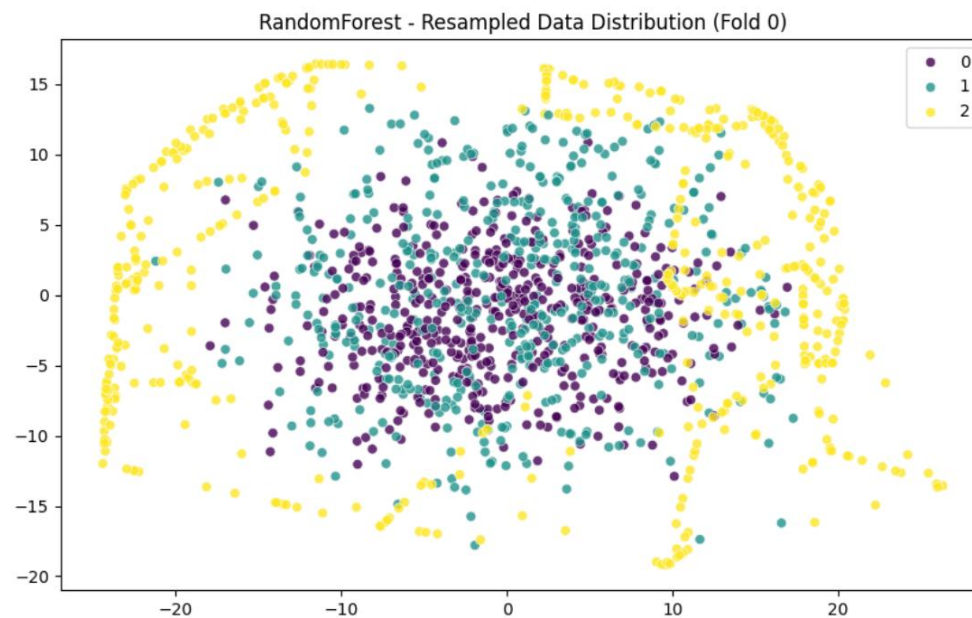
2025-2-22    By 刘震

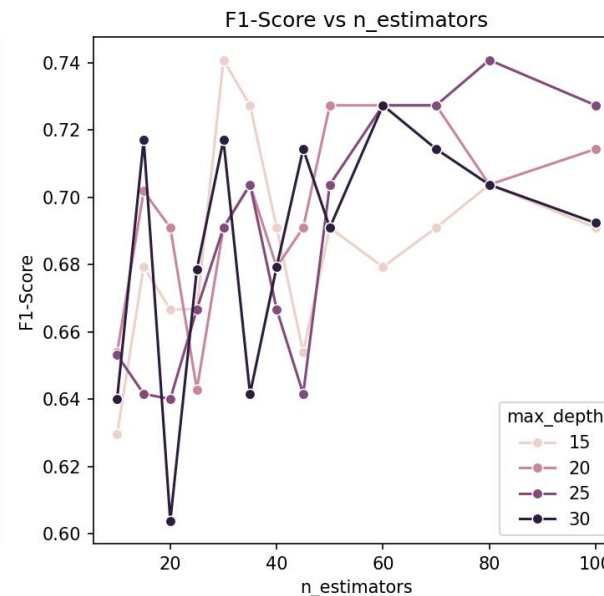
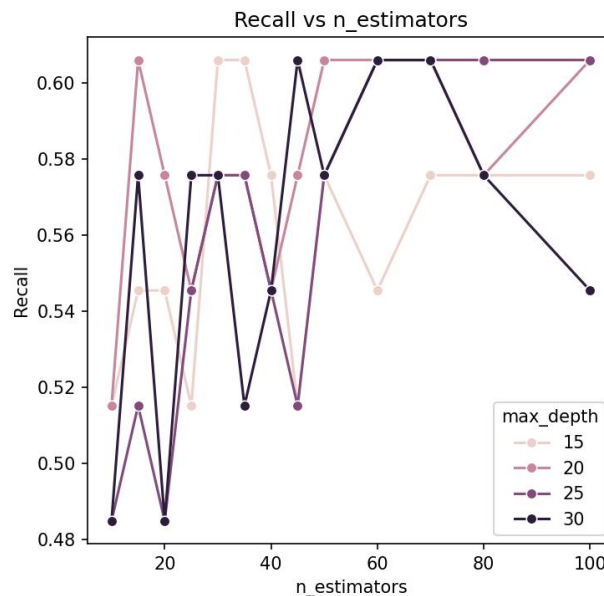
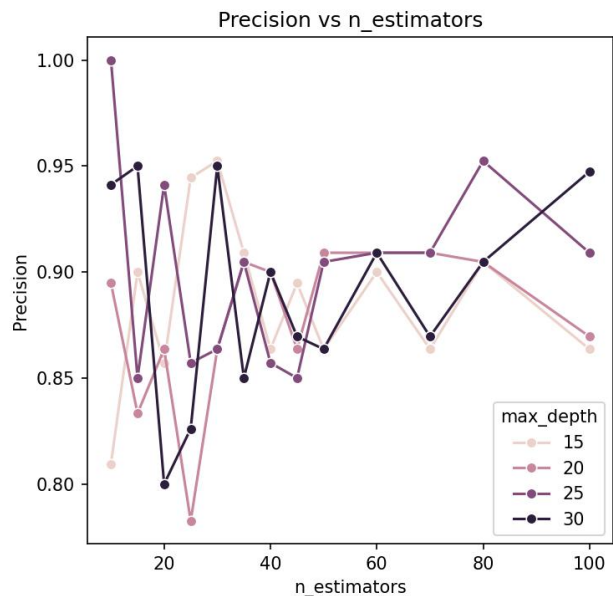
改进训练思路：如右图所示  
先区分正常和异常  
再从异常点中区分出普通异常和严重异常

并且结合之前的针对数据集样本不平衡的处理方法（SMOTE 等），以及样本数量较少、不够丰富的缺陷（利用 k 折交叉验证，取  $k = 5$ ）



SMOTE扩充后  
的数据分布

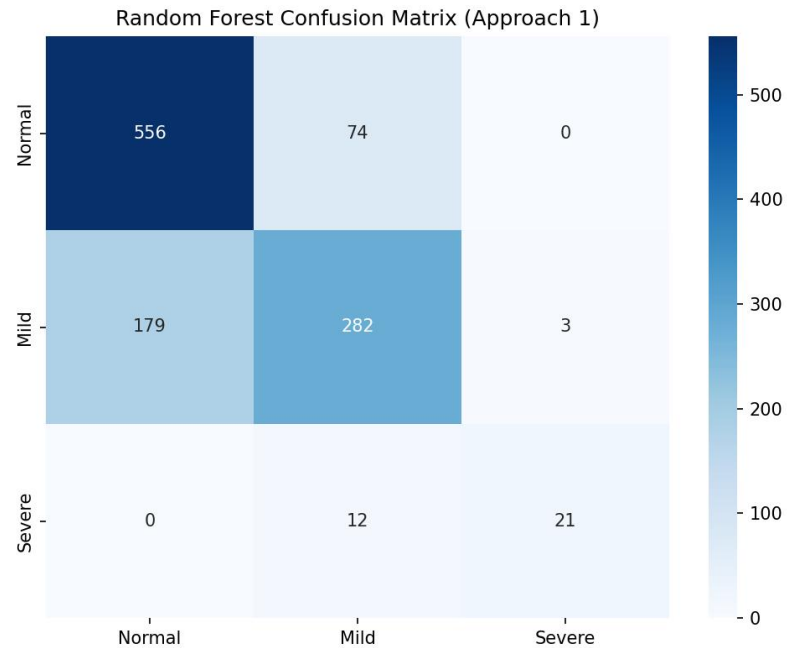




RF算法调参过程  
最终确定  
**n\_estimators=35**

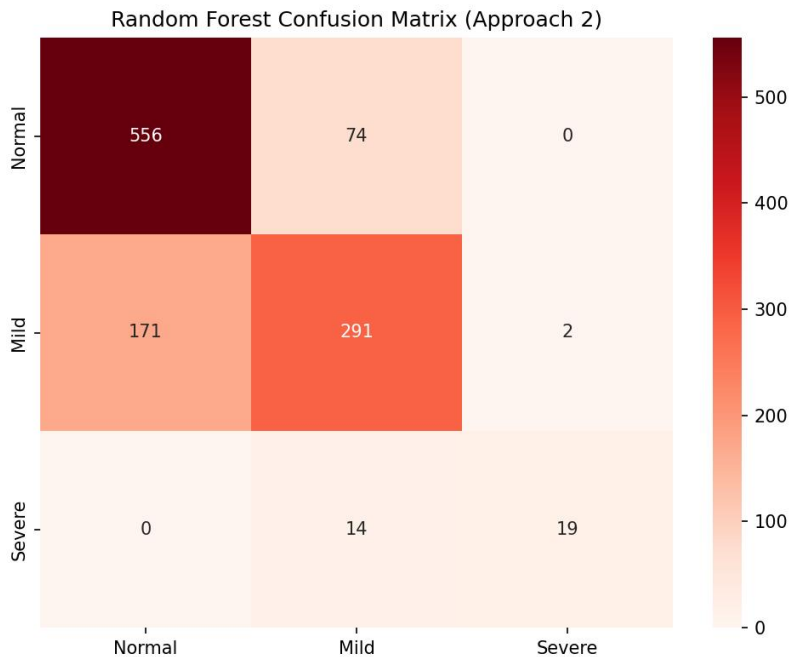
Algorithm	Task	Precision	Recall	F1-Score
KNN	正常点	0.648975	0.853968	0.737491
KNN	普通异常点	0.608997	0.37931	0.467463
KNN	严重异常点	0.666667	0.181818	0.285714
Random Forest	正常点	0.748982	0.87619	0.807608
Random Forest	普通异常点	0.749326	0.599138	0.665868
Random Forest	严重异常点	0.894737	0.515152	0.653846
MLP	正常点	0.705148	0.71746	0.711251
MLP	普通异常点	0.572632	0.586207	0.57934
MLP	严重异常点	0.272727	0.0909091	0.136364

可以看出：RF算法仍然明显优于KNN、MLP，哪怕是在已经利用SMOTE减轻了数据集的不平衡性之后；  
同时，利用这种处理方式可以略微提高对异常点中**严重异常点**的检测效果，但不明显



Algorithm	Precision	Recall	F1-Score
KNN	0.631219	0.635315	0.607178
Random Forest	0.777649	0.774623	0.768018
MLP	0.679133	0.686779	0.679455

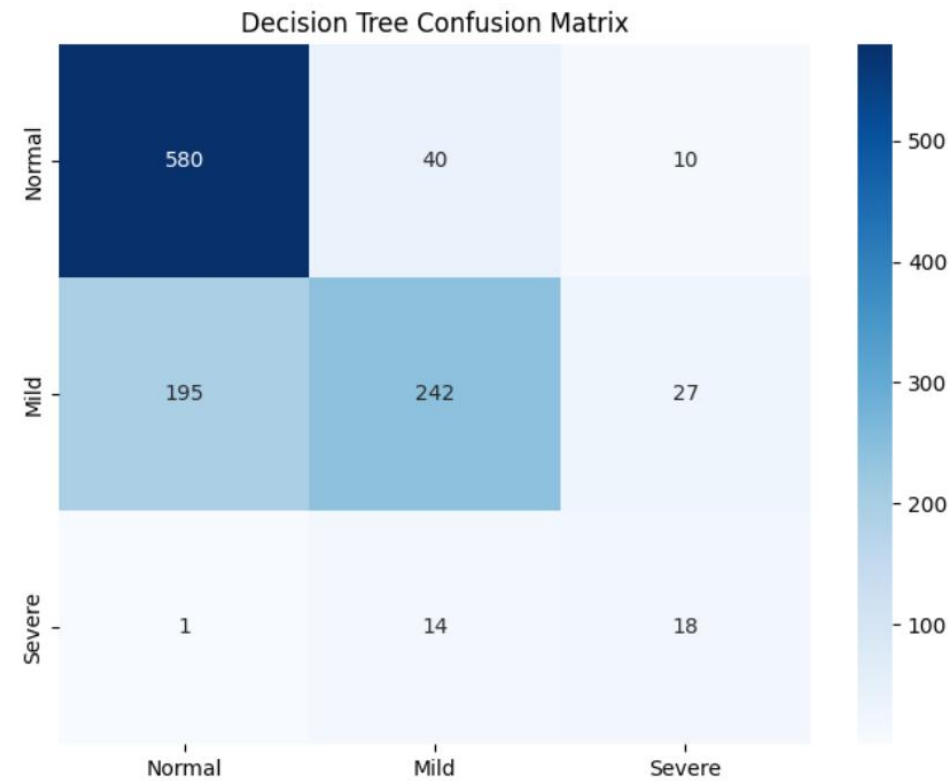
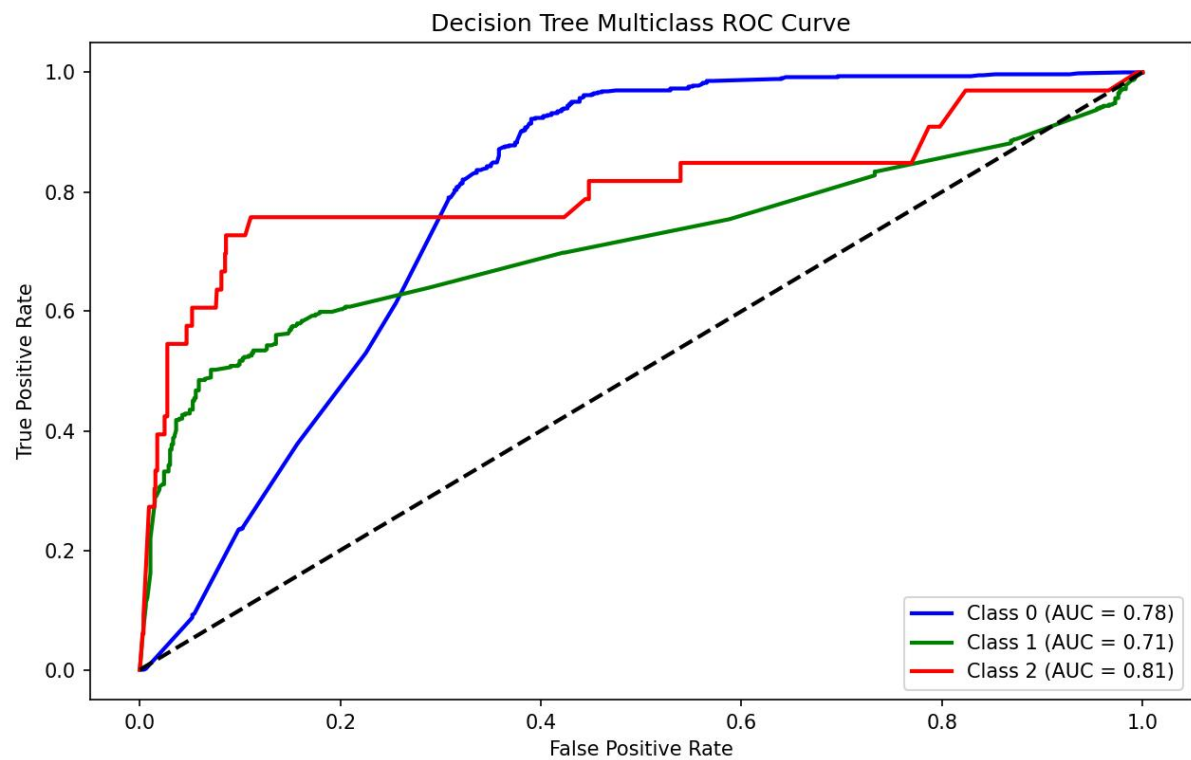
(改进后的检测方式)



Algorithm	Precision	Recall	F1-Score
KNN	0.607857	0.616681	0.598913
Random Forest	0.77013	0.768412	0.762928
MLP	0.684242	0.693878	0.685277

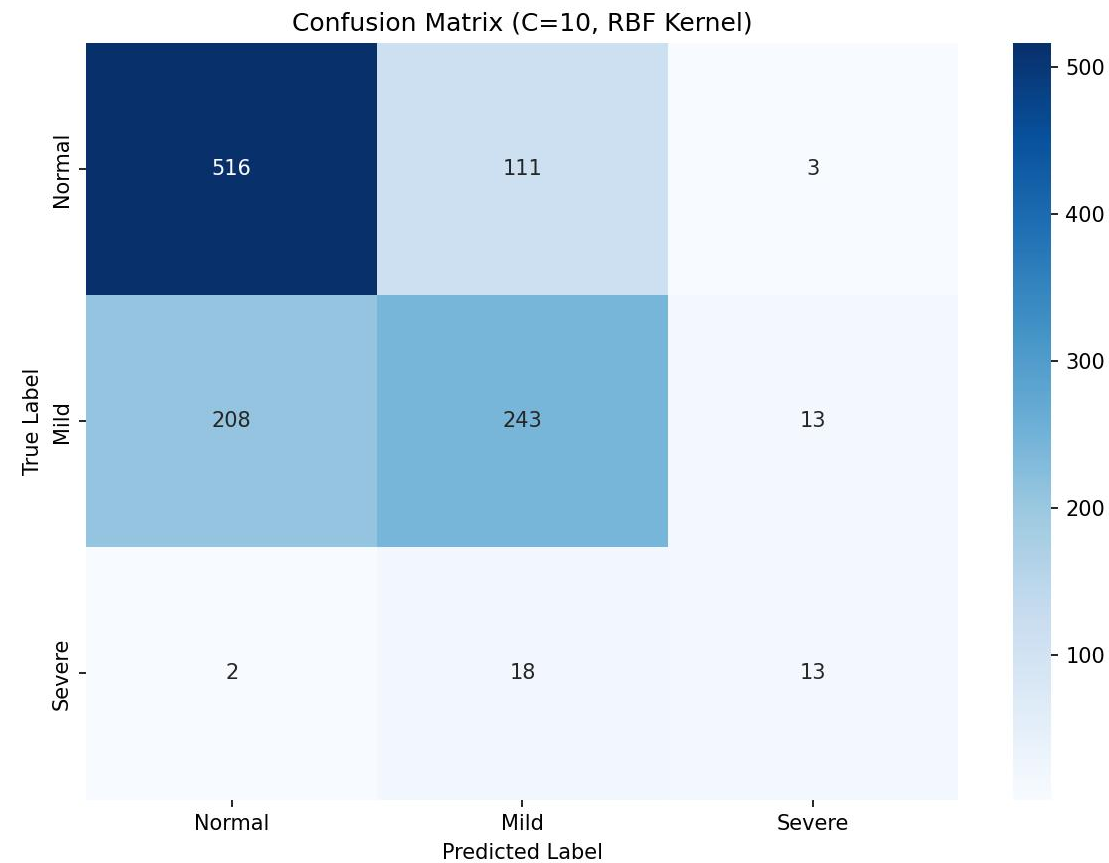
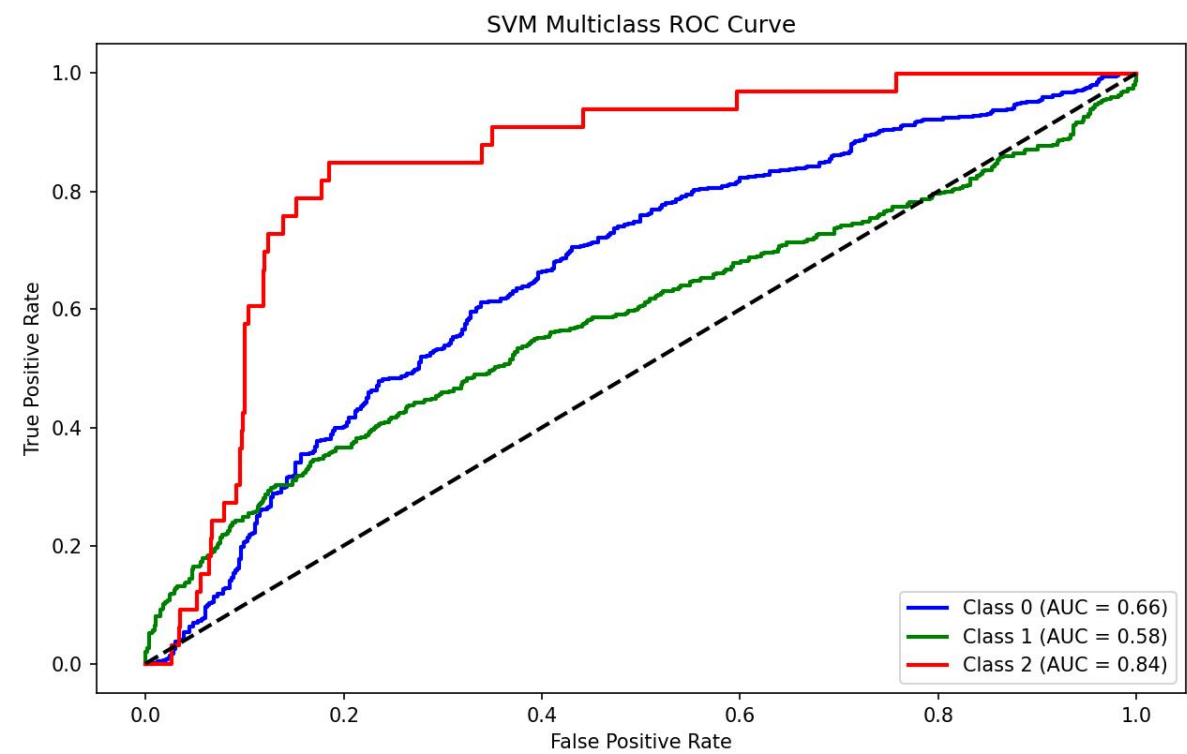
(改进前的检测方式)

# 新算法应用与尝试：决策树



该模型在区分正常点时效果较好，但是区分异常点仍然不如RF算法

# 新算法应用与尝试: SVM



尝试了不同的kernal，最终采用的 kernal 为 RBF

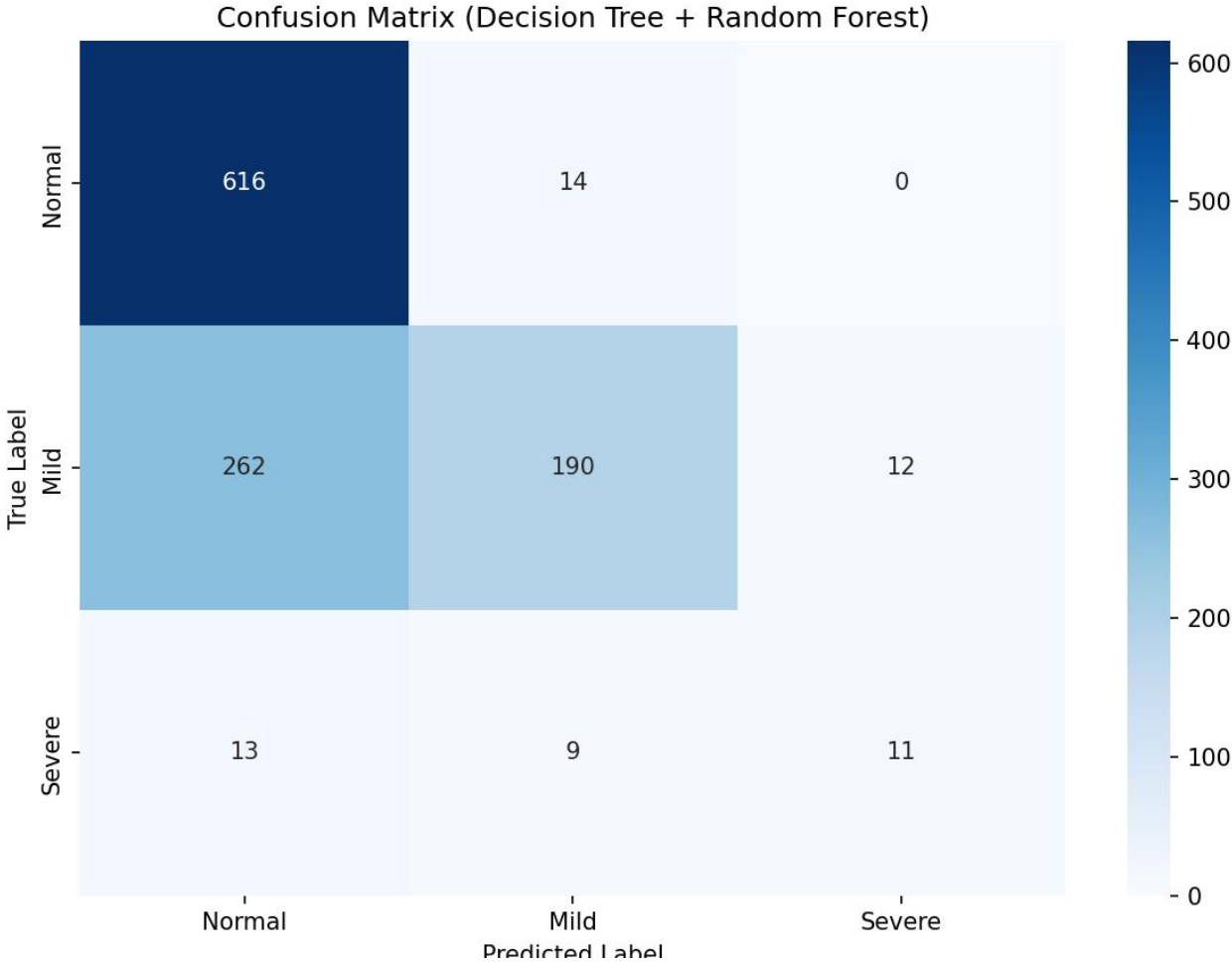
做一个组合：

前面提到，训练两个模型，其中一个区分正常点和异常点，另一个区分普通异常点和严重异常点

这两个模型可以用不同的算法

其中：区分普通异常和严重异常仍然使用RF算法，但是区分正常点和异常点时使用决策树算法，可以发挥其特长

结果如下：



## 总结

先区分正常和异常，再从异常点中区分出普通异常和严重异常的新思路，确实能提高对严重异常点的检测效果

决策树算法检测正常点的precision和recall都很高，但是对严重异常点的检测效果并不佳，在本数据集上的表现优于 KNN、MLP，但是逊于 RF

SVM 算法在本数据集上的表现不理想，当然，也可能是参数选择不合适

目前来看，就本数据集而言，使用 RF 算法进行双模型检测是最佳实践



# 问题

MLP算法经常出现如下情况，模型无法收敛，可能是导致结果不理想的原因之一

error information : ConvergenceWarning: Stochastic Optimizer: Maximum iterations (1000) reached and the optimization hasn't converged yet.

# 想法

对于每一个算法写一个程序，能自行调参，以 f1-score 为优化对象，找出最适合训练集的参数，提高准确率，同时加上一些正则化、交叉训练的手段，防止过拟合