

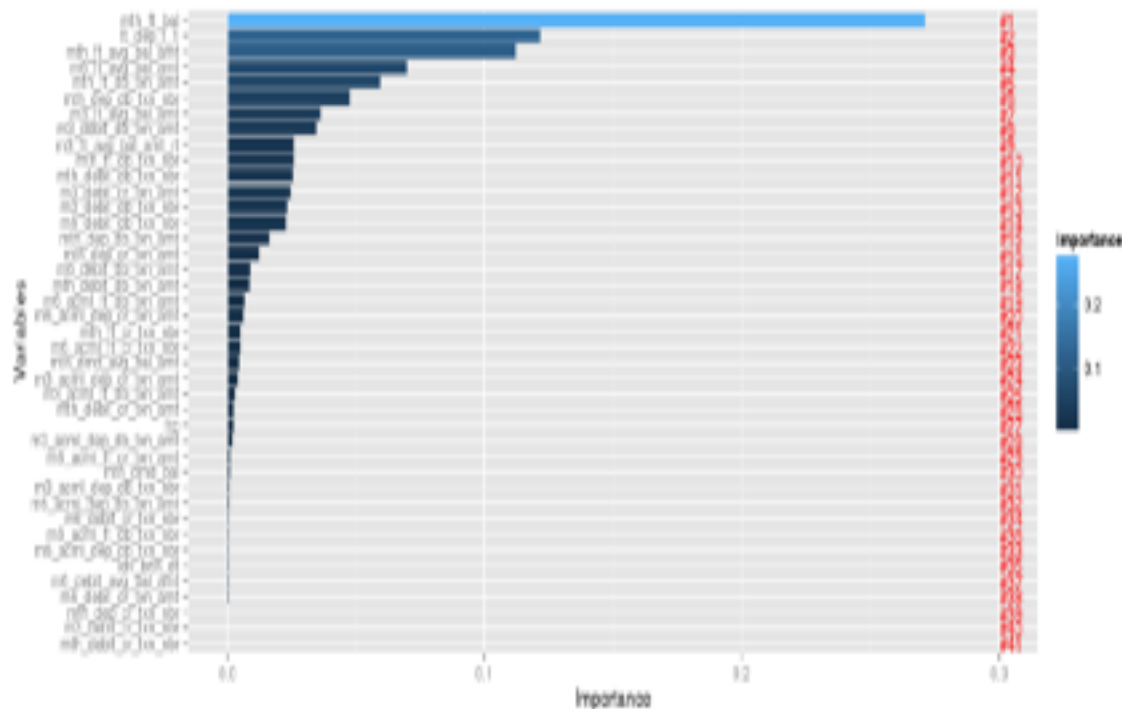
## □ 数据建模

对剩余的**73个自变量**，使用逻辑回归模型、决策树、随机森林和支持向量机这4种不同的模型，在训练集上进行统计建模，并使用验证集进行模型检验。

模型	AUC	召回率	精确率	F-测度	K-S值
随机森林	0.879	0.888	0.874	0.881	0.761
逻辑回归	0.763	0.676	0.749	0.71	0.449
决策树	0.778	0.695	0.795	0.741	0.7118

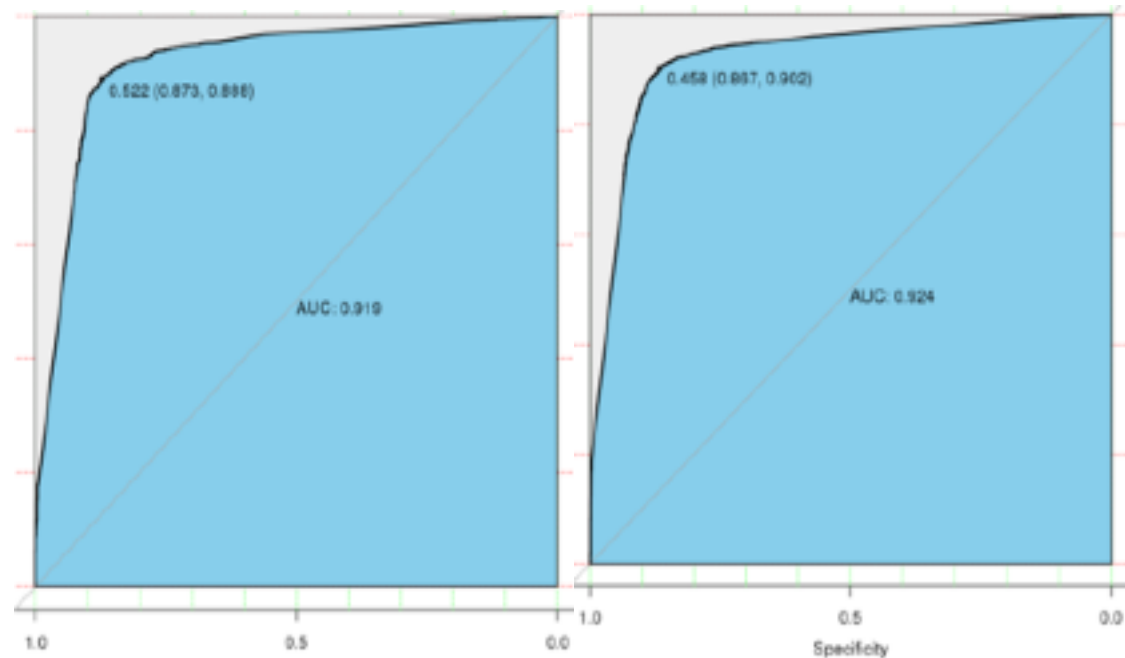
逻辑回归、随机森林和决策树这三个模型都可以得到较好的预测效果。其中随机森林效果最优，最终选择**随机森林**用于构建中高端客户流失预测模型。

模型中和定期相关的字段对模型的影响较大，其次是月累计活期借方交易笔数，近3月累计借记卡借方交易金额，它们都体现了客户在我行的资产减少的情况。另外，存款余额变化率逐渐负向增大，也预示了客户有流失的倾向。



选择前17个字段建模，模型有最高的AUC值指标,效果表现较好

模型的ROC曲线和AUC值表现效果图。模型在训练集和测试集的AUC值很接近,模型不存在过拟合现象 ,而且AUC值为0.9,模型的效果较好。



训练集和测试集指标图