音乐网站流量建模

您好！非常高兴您能参与本次建模考试，本次考试旨在能够对您的建模能力进行更好的了解，希望您能认真阅读本次考试说明，完成模型构建，将结果反馈给我们，祝您取得优异的考试成绩，加油~

本次考试需要构建的模型为：基于音乐网站的用户行为数据（用户试听，付费，充值等行为特征），预测该用户在未来一段时间内是否会发生流失（即不再使用该网站）。

以下分为三个部分：数据样本说明、考试题目及要求、模型结果交付内容及格式。

1. **数据样本说明**

1、用户行为表（附件1 CustBehavior）数据字典

此表内的account列，为用户标识列，每一行为该用户在当月的行为特征表现，具体时间可通过最后一列（数据更新日期）判断，请基于此表数据进行特征构造。



2、用户流失结果表（附件2 TrainFLag）数据字典

此表内的account列，为用户标识列，用于与用户行为表（附件1 CustBehavior）做关联，每一行为该用户在未来一段时间内是否流失，即Y标签（模型的预测目标）



3、模型预测结果表（附件3 TestPRED）数据字典

此表内的account列，为用户标识列，用于与用户行为表（附件1 CustBehavior）做关联，第3列，为需要您基于模型预测结果，填写进去的预测流失概率/评分



4、样本整体情况

样本数据时间范围：



样本数据情况及说明：



注：部分数据做了脱敏处理，消费金额类和频次类的数据部分会映射到0~1范围之内，经过修改之后数值保留了相对性，其数值大小缺乏具体含义。

1. **交付结果及格式**

上述建模考试题目完成后，请提供以下两项交付物：

1. 测试集预测结果,填写进《附件3 TestPRED》表内

非评分卡模型，包含ACCOUNT,、PRED两列，评分卡模型需包含ACCOUNT,、PRED 、SCORE三列

模型预测结果（PRED列）请以概率值（0~1之间的小数）反馈

1. 建模代码