关于行为评分模型的数据说明

在当前的数据信息社会中，个人信用已变得十分重要，其关乎着生活中的出行，消费，借贷等等各个方面。而商业银行和金融借贷公司为控制自身风险，往往需分析客户办理借贷后的各项行为，进而来判别客户是否有违约风险，以此作为公司的预警系统，为后续的针对性操作提供帮助。本次大作业的内容即为分析客户个人行为数据，建立适当的模型，判断客户在未来是否会出现逾期行为。

该数据为客户在某银行办理信用卡后一段时间内的个人行为数据，包括账户基本信息，交易行为，消费行为，信用卡还款行为，分期行为，和取现行为。目标变量**target**代表客户在未来是否出现逾期（1为有逾期，可称为坏客户；0为未逾期，可称为好客户），其他各个变量具体的名称和含义见**附件变量说明表**。数据中的缺失值已做过简单处理，处理方式见附件变量说明表，也可自行变更处理方式。数据并未提供测试集，可自行将数据分为训练集与测试集，给出模型分别在训练集和测试集上的表现即可，如给出ROC曲线，KS曲线，召回率，准确率，F1值等。建模方法可使用logistic regression, classification tree, random forest, svm, boosting(gbdt, xgboost , catboost etc.)等等。

大作业的要求如下：

1. 数据预处理，变量筛选，模型选择以及模型表现的具体过程和结果需整理成文档；
2. 代码能运行通畅，需要能在另一批不同期的数据上运行；
3. 编程语言采用R或Python。