**房屋贷款顾客进件违约预测模型之建置及预测-  
模拟实操题**

一、题目说明

房屋贷款虽然它的利率较低，但因借贷金额高，它通常是一家银行的主要获利来源之一。为因应日益增长的暴险资产，因此银行管理阶层会针对房贷客群，订定风险管理目标。本考题要求考生应用既有房贷顾客基础，运用数据挖掘分析方法，建置房屋贷款顾客进件违约预测模型，以协助授信单位进行授信准驳、风险订价及资产组合风险监控，进而能在违约风险的情况下挖掘潜在的房贷商机，提升客户贡献度。

二、数据

数据集将包含126个字段(包含1个顾客编号(B0001)以及1个目标字段(Target)-违约注记(1代表违约，0代表正常还款))，详细的字段定义及字段值的编码可参考字段定义及字段值的编码表.xlsx 。

考生将取得训练数据及测试数据。训练数据包含24,000笔新闻数据。测试数据包含6,000笔新闻数据；输入字段与训练数据相同，但没有目标字段。

三、评估方式

考生以训练数据为基础，建立一个能判断房屋贷款客户是否会违约的分类预测模型，并输出一个测试结果的档案(考生姓名\_results.csv)。考生姓名\_results.csv中只有两个字段，分别是顾客编号id以及预测是否会违约(1/0)。考生姓名\_results.csv的形式如下：

id,Predicted\_Results  
M00001,1

M00002,0

M00003,0

...  
最后需将 考生姓名\_results.csv 连同代码及建模的相关档案 (未传相关档案者将酌情扣分)，一并压缩成.zip拷贝给监考老师。考生姓名\_results.csv的格式请务必正确，否则也将酌情扣分。

评分方式是以假新闻的F1值来评估预测结果的好坏。考生的成绩以F1值的结果排序，F1值越大者越好。

此部分，共30分。

评估的结果在所有考生排名的前5%， 得30分

评估的结果在所有考生排名的前10%，得24分

评估的结果在所有考生排名的前20%，得18分

评估的结果在所有考生排名的前50%，得12分

评估的结果在所有考生排名的前75%，得6分

评估的结果在所有考生排名的后25%，得0分