**评分卡与决策树混合风控模型**

本项目旨在恢复

Huang Y, Zhang L, Li Z, et al. Fintech credit risk assessment for SMEs: evidence from China[J]. 2020.

文章的模型，算法以及实验结果。

Table

Description automatically generated

以下在三个数据集上对恢复出的模型进行了测试。

label = 1代表违约

评分卡原理：

* 对特征进行离散化分箱
* 逻辑回归得到回归系数
* 计算WOE(Weight of Evidence)后对一条记录的每个特征进行打分
* 根据评分表得到最终得分

Remark：WOE单变量某个取值的“含1率”

IV(Information Value)是WOE的加权平均，IV越大，变量的区分效果越好

数据集1：Ant信用卡违约数据集（真实数据）

训练集数据中，好客户数量为：15442,坏客户数量为：4558,坏客户所占比例为：22.79%



Chart

Description automatically generated

特征的IV值

1. 先对特征进行WOE编码，计算IV值
2. 剔除掉IV值过小的变量，进行逻辑回归，计算AUC

Chart, line chart

Description automatically generated

评分结果：分数越低，违约概率越高，基础分为92

Chart, histogram

Description automatically generated



pred为决策树预测的标签

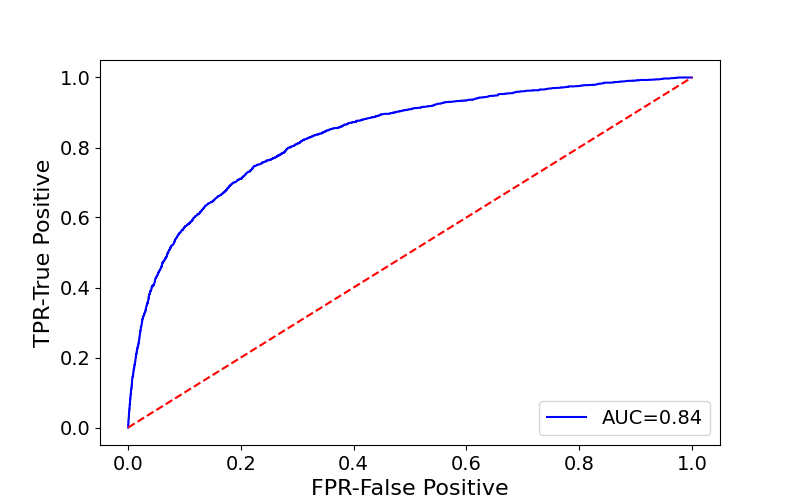
数据集2：贷款违约数据集（真实数据）

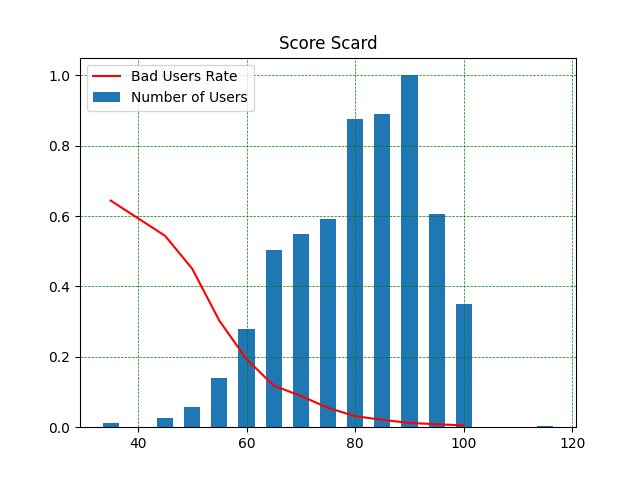
训练集数据中，好客户数量为：134065,坏客户数量为：8494,坏客户所占比例为：5.96%



Chart, bar chart

Description automatically generated







数据集3：IMF供应链贷款数据集（合成数据集）

训练集数据中，好客户数量为：35988,坏客户数量为：4012,坏客户所占比例为：10.03%

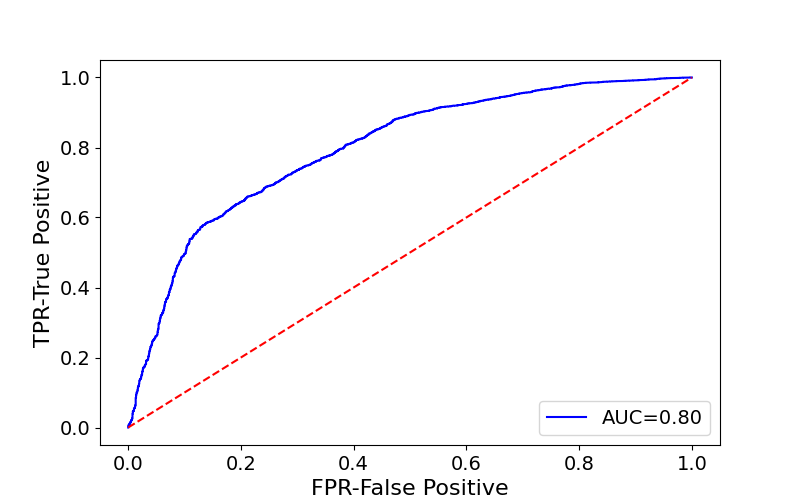


Table

Description automatically generated with medium confidence

A picture containing chart

Description automatically generated

 Chart

Description automatically generated

