

数据挖掘与商务分析-第 1 次平时作业

姓名 + 学号

请于 10 月 22 日晚 10 点前提交

1 个贷违约预测

利用个人贷款违约记录数据 (data.csv, 字段描述见表1), 进行个贷违约预测, 主要进行数据预处理、数据集划分、模型方法实践、模型参数选择和模型效果比较, 重点实践决策树、逻辑回归和朴素贝叶斯三个算法。作业要求提交完整的实验报告, 包括文字叙述、代码和运行结果。请同时提交 *.ipynb 和 *.html 文件, 其中 *.html 是将 *.ipynb 导出后的版本。

表 1: 个人贷款违约记录数据字段描述

字段	描述
loan_id	贷款记录唯一标识
user_id	借款人唯一标识
total_loan	贷款数额
year_of_loan	贷款年限
interest	当前贷款利率
monthly_payment	分期付款金额
class	贷款级别
employer_type	所在公司类型 (世界五百强、国有企业、普通企业...)
industry	工作领域 (传统工业、商业、互联网、金融...)
work_year	工作年限
house_exist	是否有房
issue_date	贷款发放的月份
use	贷款用途类别
debt_loan_ratio	债务收入比
del_in_18month	借款人过去 18 个月逾期 30 天以上的违约事件数
known_outstanding_loan	借款人档案中未结信用额度的数量
known_dero	贬损公共记录的数量
pub_dero_bankrup	公开记录清除的数量
early_return	借款人提前还款次数
early_return_amount	贷款人提前还款累积金额
early_return_amount_3mon	近 3 个月内提前还款金额
isDefault	是否违约