任务说明

1. 任务目标

根据现有数据进行信用评分。

1. 数据说明

现有743条贷款数据，每条数据17个相关变量，分别是status、paid.ratio、group和14个模型变量，列表如下：

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 模型变量 | | | | |
| 年龄(age) | 薪水(salary) | 性别(gender) | 教育程度(education) | 毕业年份(graduated.years) |
| 婚姻状态(marriageStatus) | 有无小孩(hasChild) | 有无房子(hasHouse) | 有无房贷(houseLoan) | 有无车贷(carLoan) |
| 有无车子(hasCar) | 工作年限(workYears) | 公司规模(officeScale) | 有无信用卡(credit) |  |

其中1.教育程度一项给高中、大专、本科、研究生分别打分3、6、8、10。未填此项打分为0。2.对给出数额范围的项目做了取边界和平均的平滑处理。例如，薪水小于1000算作1000；1000-2000算作1500；2000-5000算作3500等等。

分类使用status变量作为y值，0表示没有违约、1表示违约。

回归使用paid.ratio变量，变量值表示已付的比例。

1. 数据预处理

对需要进行预处理的数据字段进行处理，比如缺失值处理、异常值处理等。对处理之后的数据提取特征。

1. 相关性分析

对特征的相关性进行分析。

1. 模型选择

利用提取的特征进行模型训练。比如逻辑回归、线性回归、决策树、随机森林、Adaboost、GBDT等。

1. 模型评价

使用常用的评分卡模型评价指标对训练的模型进行评价。比如KS、PSI、AUC等。

1. 模型解释

对训练出的模型进行解释。

1. 生成报告

任务完成对任务完成情况进行文档说明。

补充说明：

1. 所有代码均用python编写