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15000条贷款数据，每条数据11个相关变量，分别是：

|  |  |  |
| --- | --- | --- |
| **Variable Name** | **Description** | **Type** |
| **SeriousDlqin2yrs** | **Person experienced 90 days past due delinquency or worse** | **Y/N** |
| RevolvingUtilizationOfUnsecuredLines | Total balance on credit cards and personal lines of credit except real estate and no installment debt like car loans divided by the sum of credit limits | percentage |
| age | Age of borrower in years | integer |
| NumberOfTime30-59DaysPastDueNotWorse | Number of times borrower has been 30-59 days past due but no worse in the last 2 years. | integer |
| DebtRatio | Monthly debt payments, alimony,living costs divided by monthy gross income | percentage |
| MonthlyIncome | Monthly income | real |
| NumberOfOpenCreditLinesAndLoans | Number of Open loans (installment like car loan or mortgage) and Lines of credit (e.g. credit cards) | integer |
| NumberOfTimes90DaysLate | Number of times borrower has been 90 days or more past due. | integer |
| NumberRealEstateLoansOrLines | Number of mortgage and real estate loans including home equity lines of credit | integer |
| NumberOfTime60-89DaysPastDueNotWorse | Number of times borrower has been 60-89 days past due but no worse in the last 2 years. | integer |
| NumberOfDependents | Number of dependents in family excluding themselves (spouse, children etc.) | integer |

**SeriousDlqin2yrs**变量作为y值，0表示没有违约、1表示违约。

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