**题目要求**

本次设计要求实现信贷用户逾期预测功能。具体要求如下：利用所学数据挖掘算法对给定数据进行训练得出信用评估模型，依据模型对1000个贷款申请人是否逾期做出预测（0-未逾期 1-逾期）。

**数据说明**

数据概述 ：

数据包括脱敏后的个人基本信息、持卡信息、消费信息及信贷信息等。详见字段解释.xlsx。

数据结构和数据规模 ：

数据分为训练集和验证集。 详见model.csv和test.csv

**报告要求**

1. 使用A4单面打印装订，将姓名、学号、班级写在封皮处；
2. 报告字体宋体5号，行距1.5倍，首行缩进2字符，代码部分只需写出主要部分。
3. 报告内容包含所用环境及编程语言说明、算法思想、算法伪码、数据预处理过程、预测结果、结果分析以及调式过程。
4. 程序部分在学习通提交，命名方式：学号+姓名+班级，提交时间为2019/5/23-2019/5/24,报告纸质部分由各班班长负责收齐送至212，提交时间为2019/5/24（附纸条：交了报告的学生名字及总人数）。

**给分准则**

本次大作业分数由程序执行结果的准确率和报告书写情况两部分构成，杜绝一切抄袭可能，若有雷同视为零分。

**对班长的要求**

负责将每个同学的预测结果统计好，将结果写入以下表格中，通过学习通单独提交给我，命名方式:班级+预测结果，提交时间2019/5/24。

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| UserId | 学号 | | | |
| 学号1 | 学号2 | ………… | 学号n |
| 预测数据1 |  |  |  |  |
| 预测数据m |  |  |  |  |