

哈爾濱工業大學

实验报告

题 目：根据客户信息进行违约表现预测

专 业 大数据科学与技术

学 号 1170300916

姓 名 彭钰驯

课 程 量化策略开发与程序化交易

日 期 2020-6-7

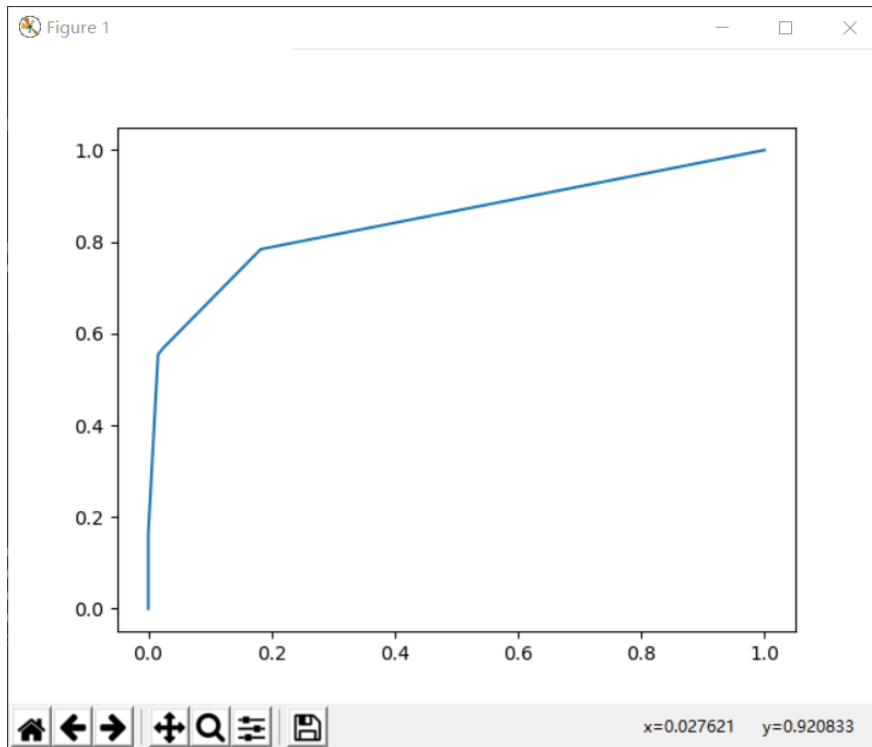
1.1 研究的问题

研究的问题是根据客户信息预测其是否违约。使用的模型为决策树模型，训练数据为收入，年龄，性别，历史授信额度和历史违约次数，测试数据为是否违约。

收入	年龄	性别	历史授信额度	历史违约次数	是否违约
503999	46	1	0	1	1
452766	36	0	13583	0	1
100000	33	1	0	1	1
100000	25	0	0	1	1
258000	35	1	0	0	1
933333	31	0	28000	3	1
665000	40	1	5000	1	1
291332	38	0	0	0	1
259000	45	1	0	1	1
3076666	39	1	71000	2	1
695000	40	1	5000	0	1
600000	35	1	18000	3	1
440000	36	1	0	2	1
511999	35	0	0	2	1
248000	31	0	0	2	1

1.2 结果分析

运行结果截图：



```

预测值  实际值
195      0      0
196      0      0
197      0      0
198      1      0
199      1      1
[[0.86004515 0.13995485]
 [0.86004515 0.13995485]
 [0.86004515 0.13995485]
 [0.86004515 0.13995485]
 [0.55782313 0.44217687]]
[0.86004515 0.86004515 0.86004515 0.86004515 0.55782313]
[0.13995485 0.13995485 0.13995485 0.13995485 0.44217687]
      阈值      假警报率      命中率
0  2.000000  0.000000  0.000000
1  1.000000  0.000000  0.162162
2  0.959350  0.015873  0.554054
3  0.750000  0.023810  0.567568
4  0.442177  0.182540  0.783784
5  0.139955  1.000000  1.000000
0.8464714714714715

```

结果显示，预测准确率为 0.846

