

信用评分11

上次课程内容

信用评分

信誉

信誉模型

信任

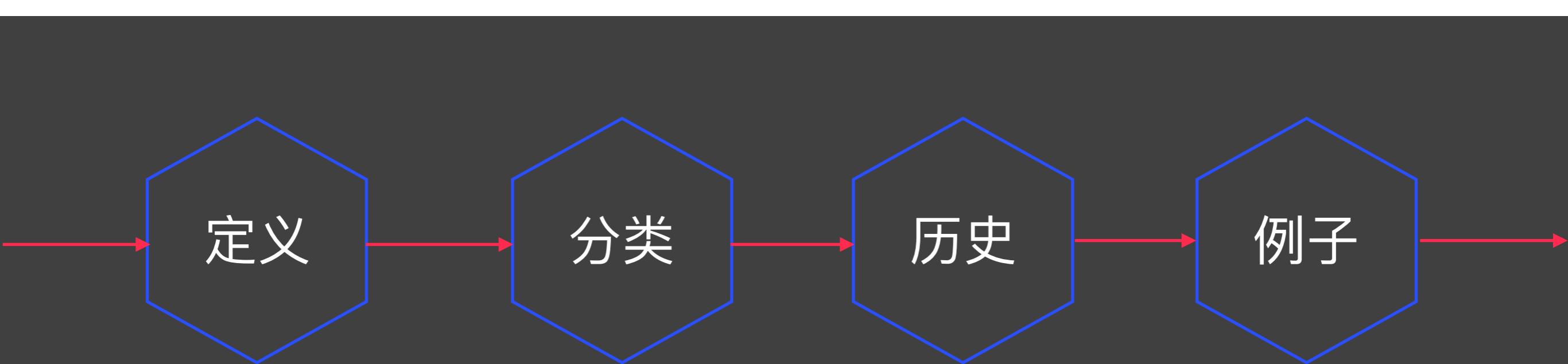
- 定义
- 分类
- 历史
- 例子

- 定义
- 例子作用
- Web信誉
- 要素

- 构成
- 计算
- 可视化
- 系统

- 定义
- 模型
- 描述
- 攻击

信用评分简介



- Credit Scoring is decision support systems used in consumer credit aims at assessment of potential borrowers and existing borrowers.
- Default probability is predicted from observed borrowers characteristics on the basis of the analysis of known performance pf previous customers.
- Risk / creditworthiness is usually measured by default probability.

信用评分起源



Character

Capacity

Coateral

Capital

Conditions

自动化

快速

一致

客观

不安全

信用评分需求

个人贷款

贷款额度小

贷款客户多

主要是预测

研究较少

教科书少

管理 + 数据

企业贷款

贷款额度大

贷款客户少

主要是因果

大量研究

大量教科书

金融 + 会计

中小微企业贷款

贷款额度?

贷款客户?

关注?

模型?

研究?

学科?

信用评分目的

申请客户

债务违约

产品使用

用户流失

已有客户

信用更新

交叉销售

再次申请

问题客户

预警

催收

坏账

风险定价

抵押担保

利润评分

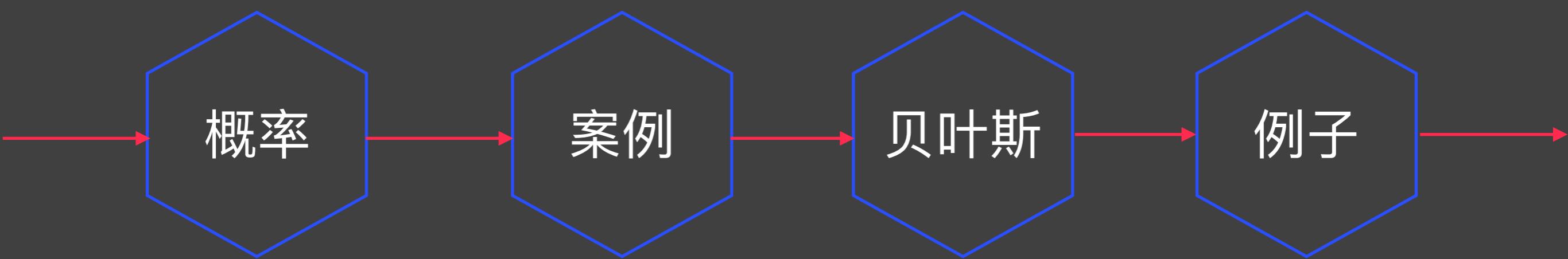
客户评价

资本充足

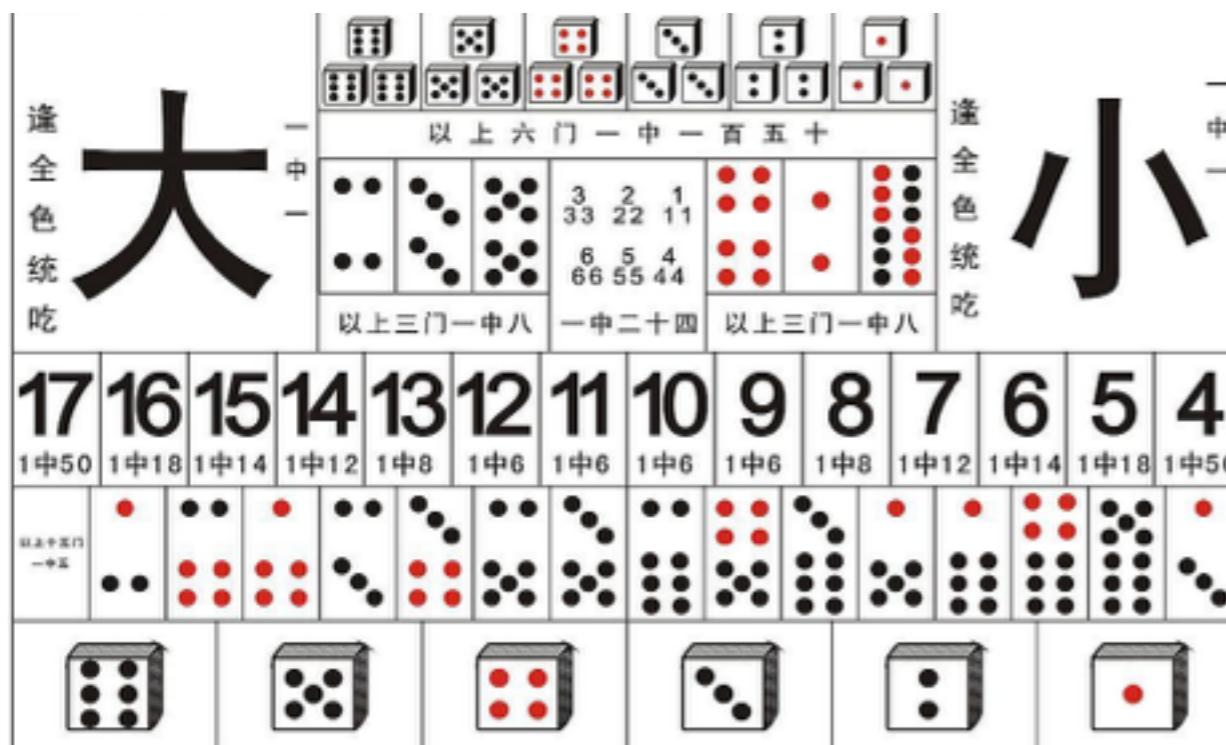
风险度量

IFRS9

评分卡



概率和信用评分



重复实验N次

R次结果是Good

P(G)、P(B)

$$P = G/(G+B)$$

$$G: B = P/(1 - P)$$

$$O(G) = P(G):P(B)$$

信用评分案例

银行有8000客户申请贷款，一年后7000是好的，1000是坏的，好客户平均收益一千，坏客户平均损失一万

好客户弥补坏客户、
损益平衡点：10:1

群体好坏概率：
7:1

Marital Status:

	Good	P(x G)	Bad	P(x B)	Marginal Odds
Married	4900	0.7	400	0.4	49 : 4
Not married	2100	0.3	600	0.6	21 : 6
Total	7000	1	1000	1	

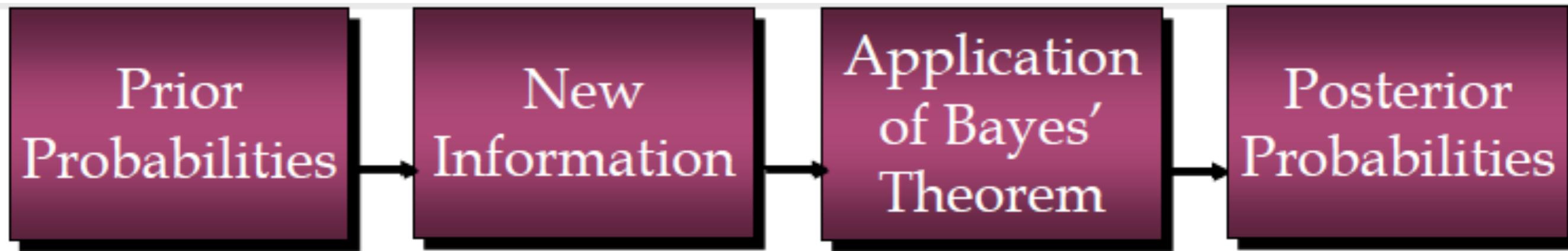
$$\text{Marginal Odds of Married} = \frac{0.7}{0.4} \times 7 : 1 = 12.25$$

$$\text{Marginal Odds of NM} = \frac{0.3}{0.6} \times 7 : 1 = 3.5$$

Information Odds

NB: Marginal Odds = Information Odds × Population Odds

信用评分中的贝叶斯理论



Let $\mathbf{X} = (X_1, X_2, \dots, X_m)$ be characteristics (variables) of the borrower such as age, marital status, etc.

$\mathbf{x} = (x_1, x_2, \dots, x_m)$ be outcomes/ attributes of characteristics.

$P(G)$ and $P(B)$ are prior probabilities.

Posterior probabilities:

$P(G|\mathbf{x})$ is the probability of being Good given certain attributes

$P(B|\mathbf{x})$ is the probability of being a Bad customer given certain attributes

工作时间

Marital Status:

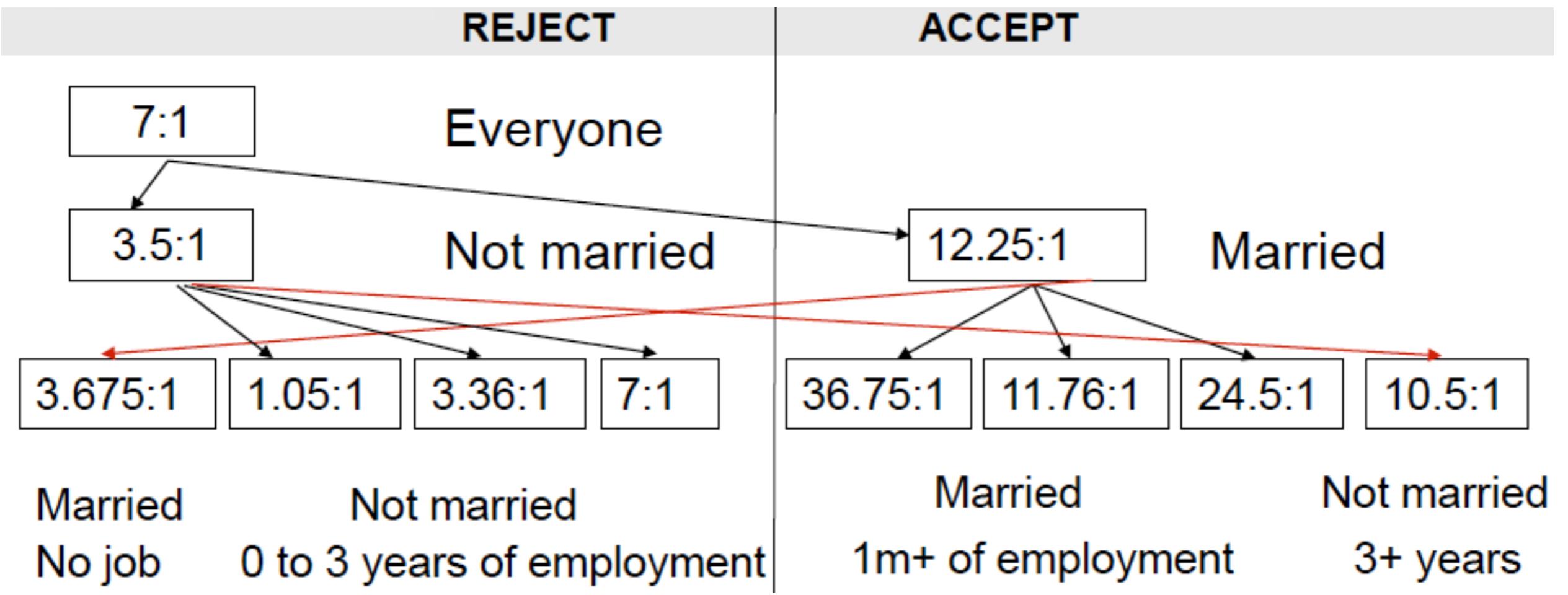
	Good	$P(x G)$	Bad	$P(x B)$	Marginal Odds, $O(G x)$
Married	4900	0.7	400	0.4	49 : 4 12.25:1
Not married	2100	0.3	600	0.6	21 : 6 3.5:1

Time in Employment :

0	1050	0.15	500	0.5	105 : 50 2.1:1
up to 6 m	1680	0.24	250	0.25	168 : 25 6.72:1
6m - 3y	1960	0.28	140	0.14	196 : 14 14:1
3y+	2310	0.33	110	0.11	231 : 11 21:1
Total	7000		1000		

贝叶斯评分卡

$\text{Pop Odds} \times \text{Info Odds}(\text{Char 1}) \times \dots \times \text{Info Odds}(\text{Char } n)$



$$\begin{aligned}\text{Odds of Married and No Job} &= 7/1 \times 0.7/0.4 \times 0.15/0.5 = \\ &= 7 \times 1.75 \times 0.3 = 3.675\end{aligned}$$

独立性
假设

$$\begin{aligned}\text{Odds of Not Married and 3+ years of employment} &= ? \\ &= 7/1 \times 0.3/0.6 \times 0.33/0.11 = 7 \times 0.5 \times 3 = 10.5\end{aligned}$$

Credit Scoring II

贝叶斯评分卡

Time at current address	Less than 6 months	6m – 2 years	2 – 6 years	6 - 10 years	10 + years	Unknown		
	0	3	6	13	25	0		
Residential Status	Owner	Tenant	With parents	Unknown		属性	性能	风险
	15	5	2	0				
Banking	Current account	Saving account	Current and saving	No account	Unknown	历史		评分
	5	10	14	0	0			
Occupation	Retired	Full-time	Part-time	Self-employed	Student	Other	Un-known	
	21	16	7	6	5	10	0	
Age	18-25	26-31	32-40	41-54	55+	Unknown		
	5	10	15	20	25	0		

Credit Scoring II

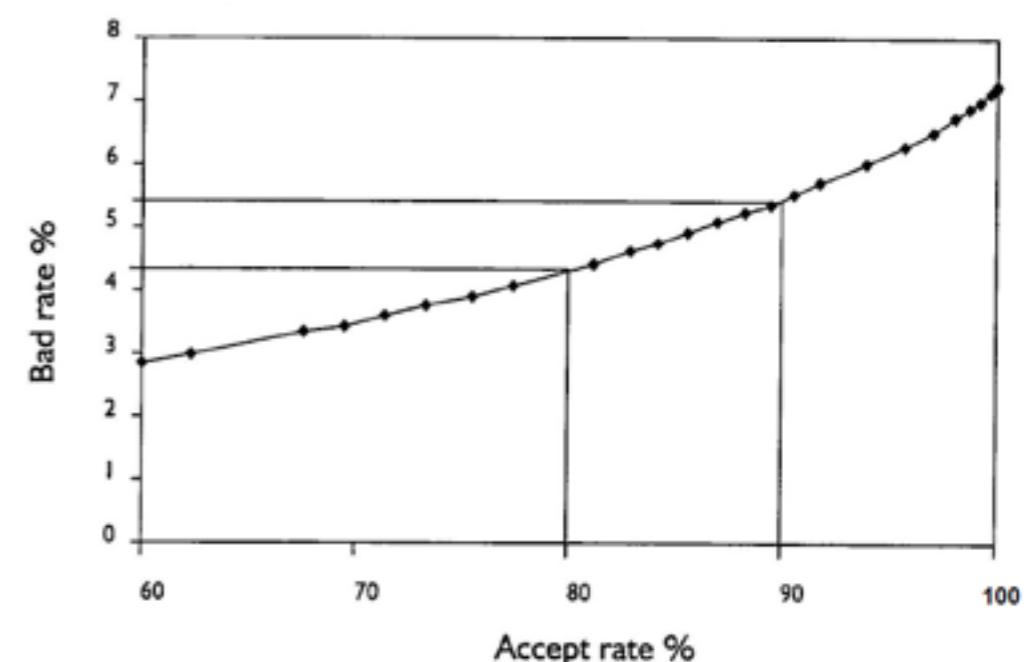
策略

36.75:1	Married	3+ years of employment
24.5:1	Married	6m – 3y of employment
11.76:1	Married	up to 6m of employment
10.5:1	Not married	3+ years of employment
7:1	Not married	6m – 3y of employment
3.675:1	Married	No job
3.36:1	Not married	up to 6m of employment
1.05:1	Not married	No job

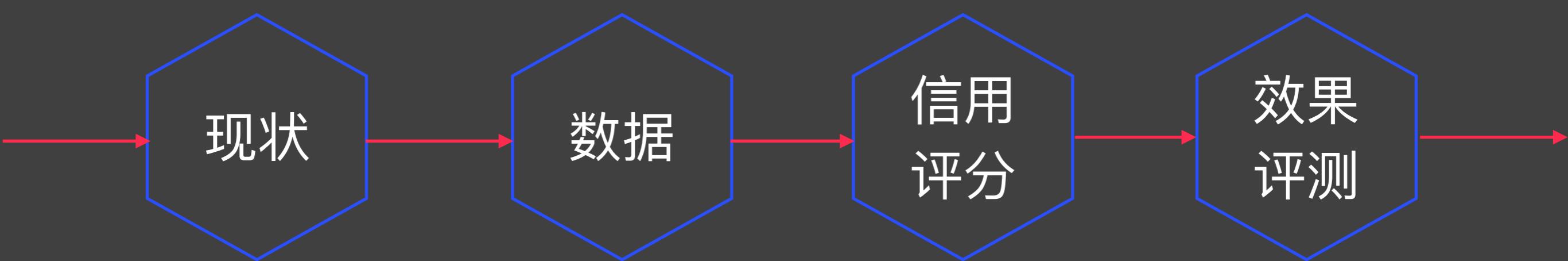
数据
易获取
自动化
廉价

假设：未来和过去相似

信用评分是预测，不是可解释的



评分卡基础



好坏样本

Good

各自1500

Bad

GD?

最好20000-50000, 所有的坏用户

随机性

正确性

实时性

完整性

合法性

6 -12 months

9 – 24 months

季节

Acceptance/ Sample
period

Outcome/ Performance
period

长度

数据来源

申请表

年龄

婚姻

地址

工作

社交媒体

手机使用

短信使用

地理位置

账户

征信报告

银行纪录

抵押

司法

借贷

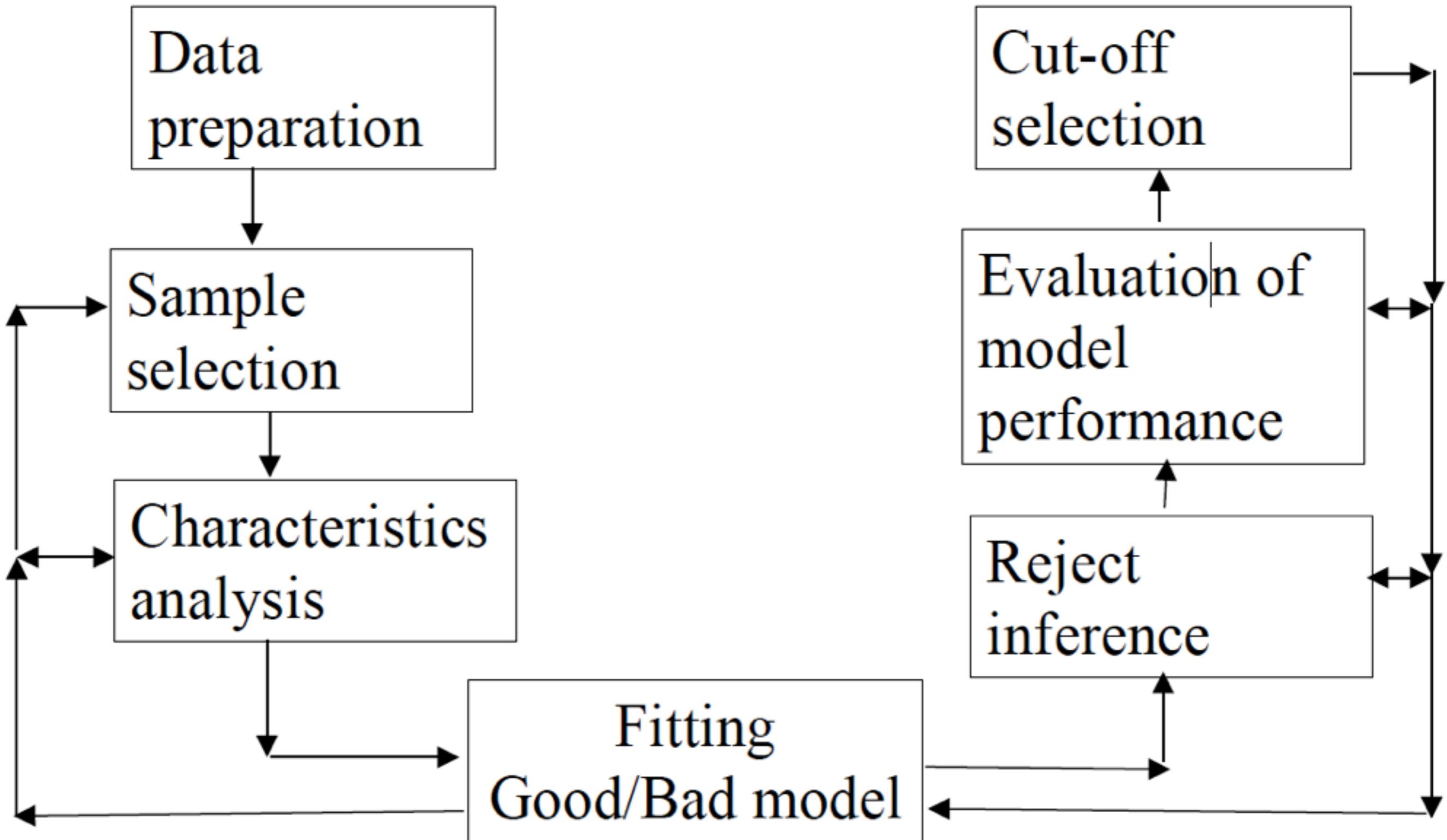
账户

流水

贷款

新数据

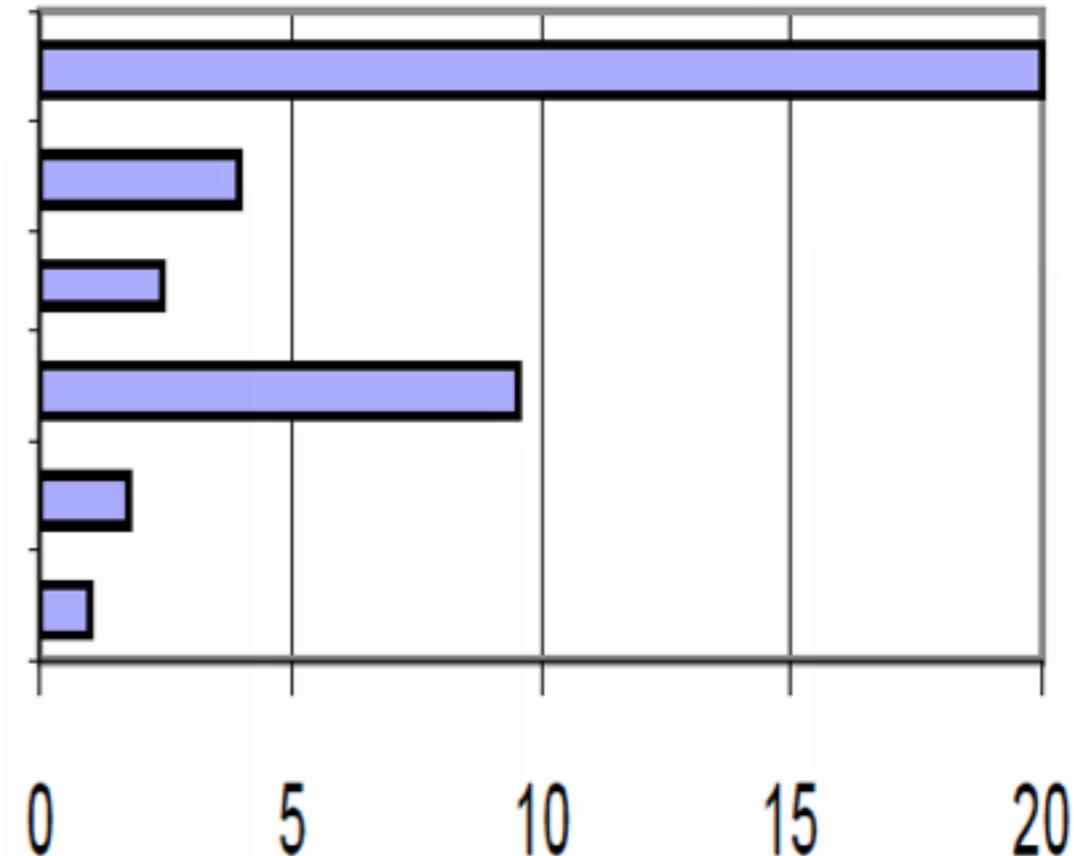
开发过程



Credit Scoring II

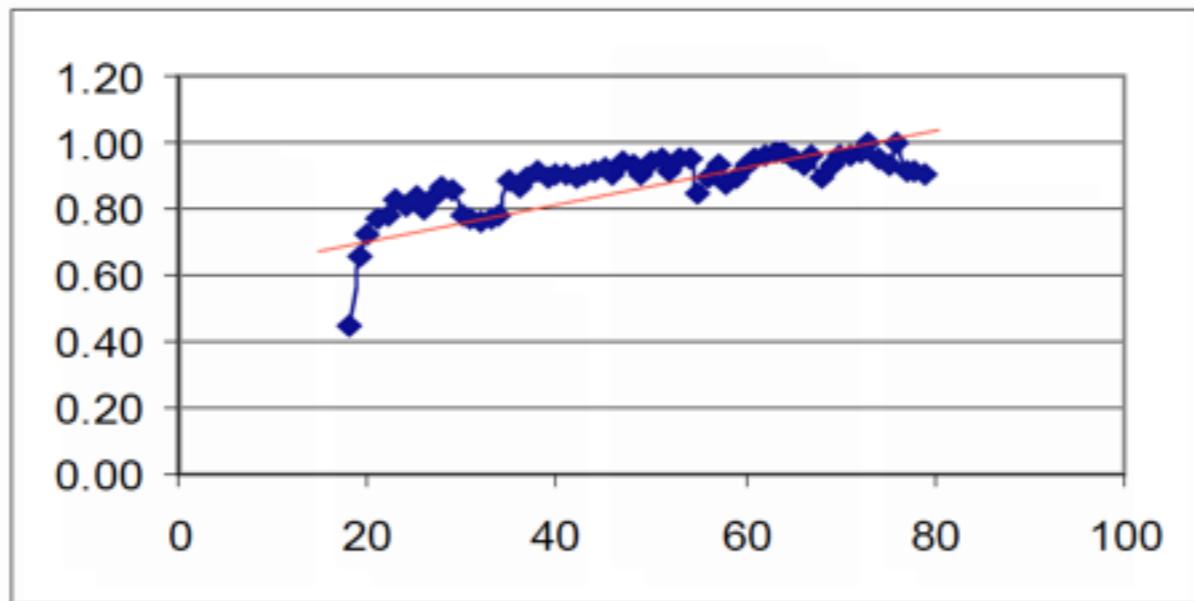
特征分析

Residential Status	Goods	Bads	Odds
Owner	6000	300	20 : 1
Rent unfurnished	1600	400	4 : 1
Rent furnished	350	140	2.5:1
With parents	950	100	9.5 : 1
Other	90	50	1.8 : 1
No answer	10	10	1 : 1

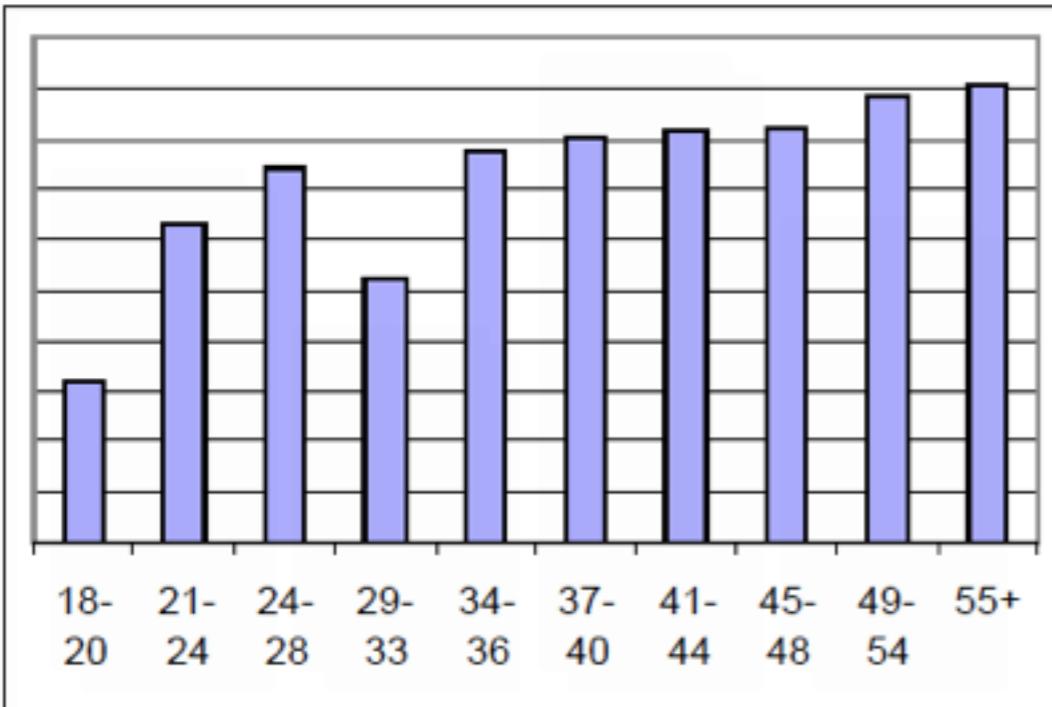


特征分析

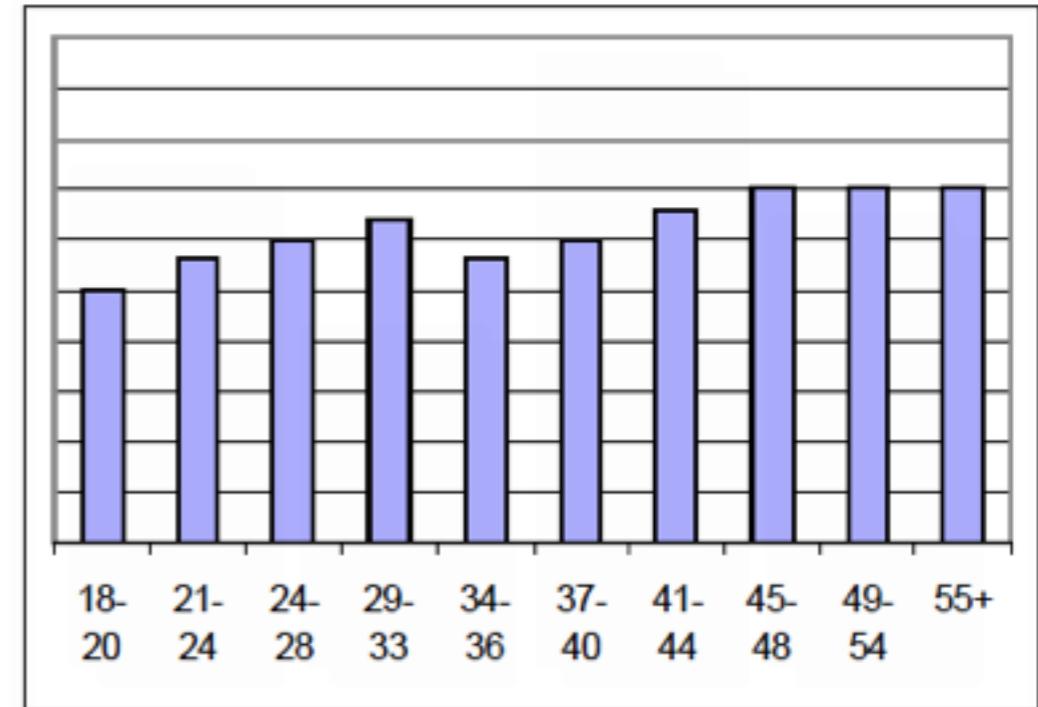
Good Rate by Applicant's Age



Good rate by Applicant's Age



Good rate by Spouse's Age



Statistical

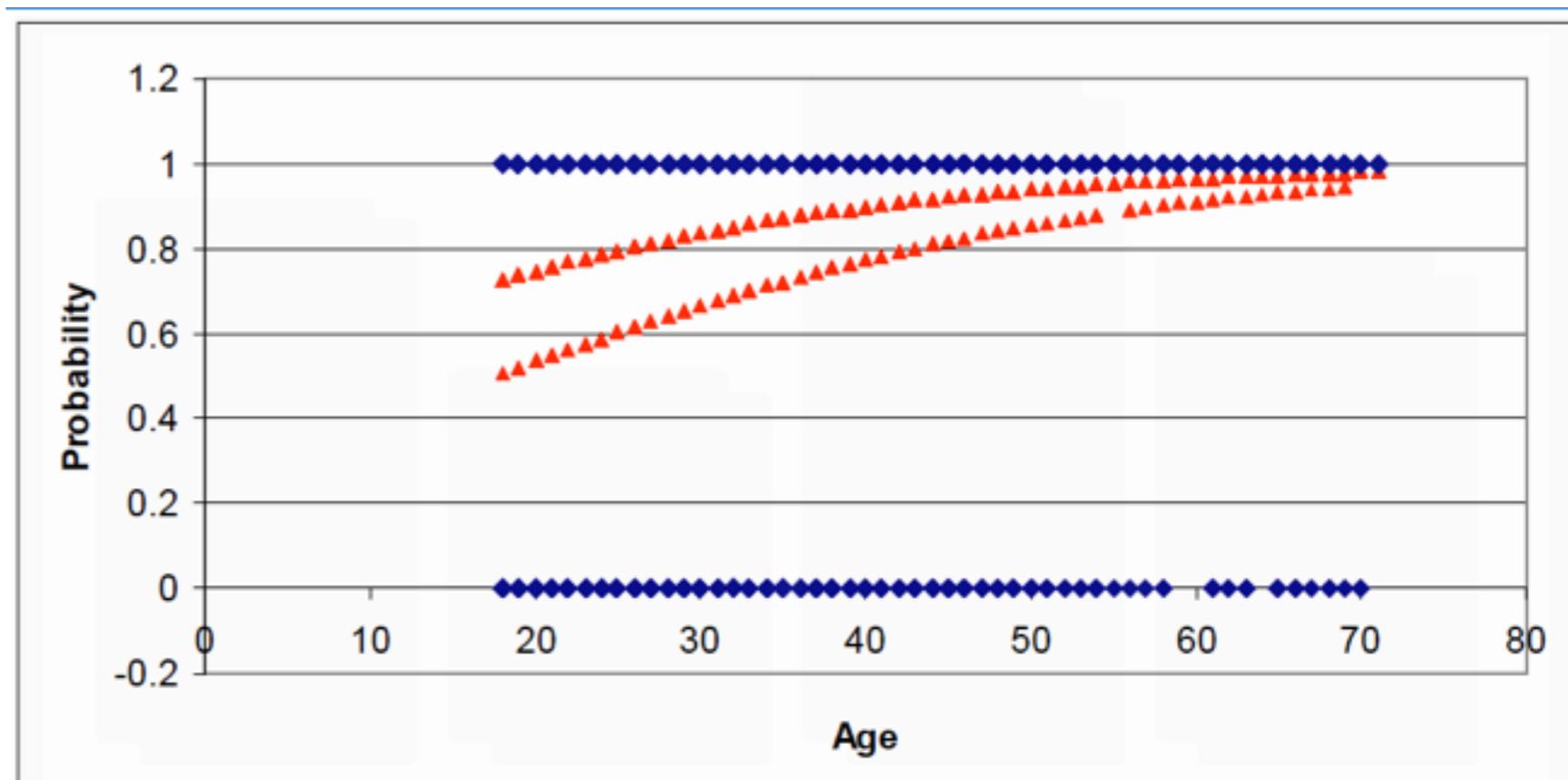
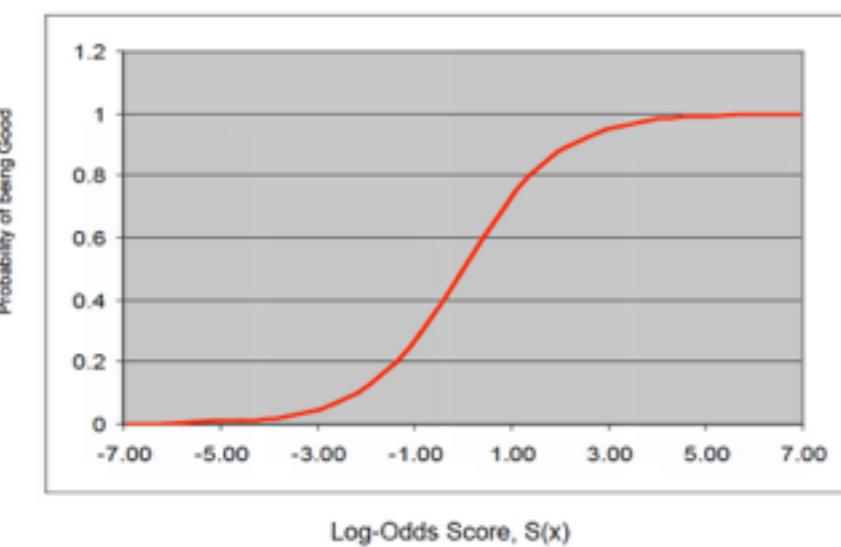
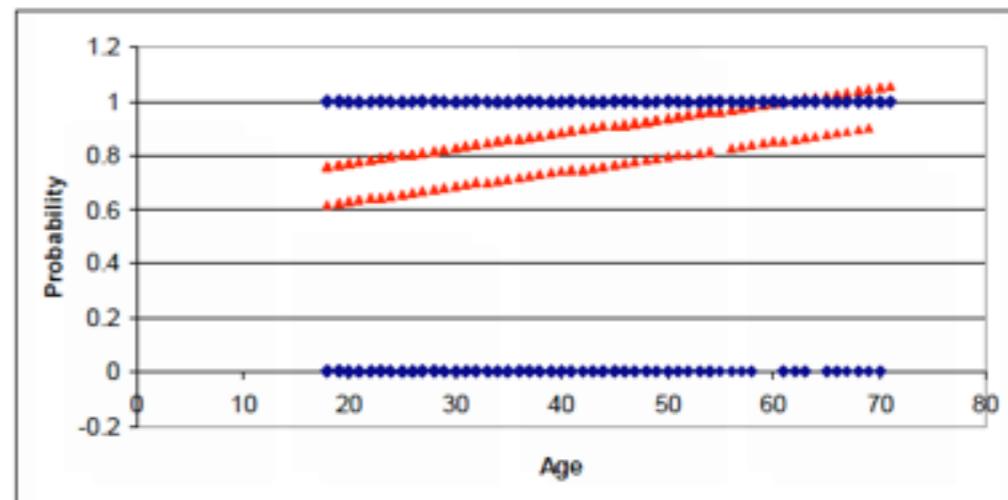
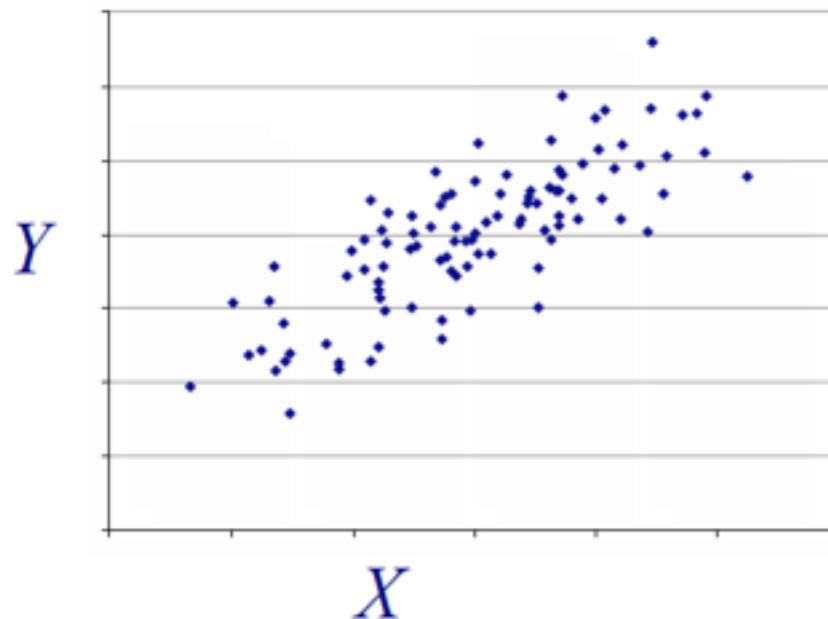
- Discriminant analysis (DA or LDA),
- Linear regression (LR),
- Logistic regression (LogR),
- Nearest-neighbours approach (k-NN),
- Classification trees.

Machine-Learning

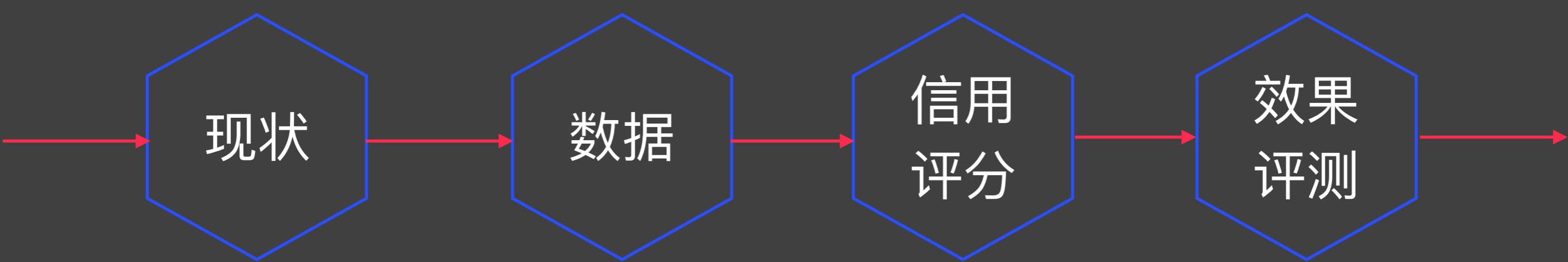
- Support Vector Machines (SVM),
- Neural networks (NN),
- Genetic algorithms,
- Random Forests,

Credit Scoring II

逻辑回归

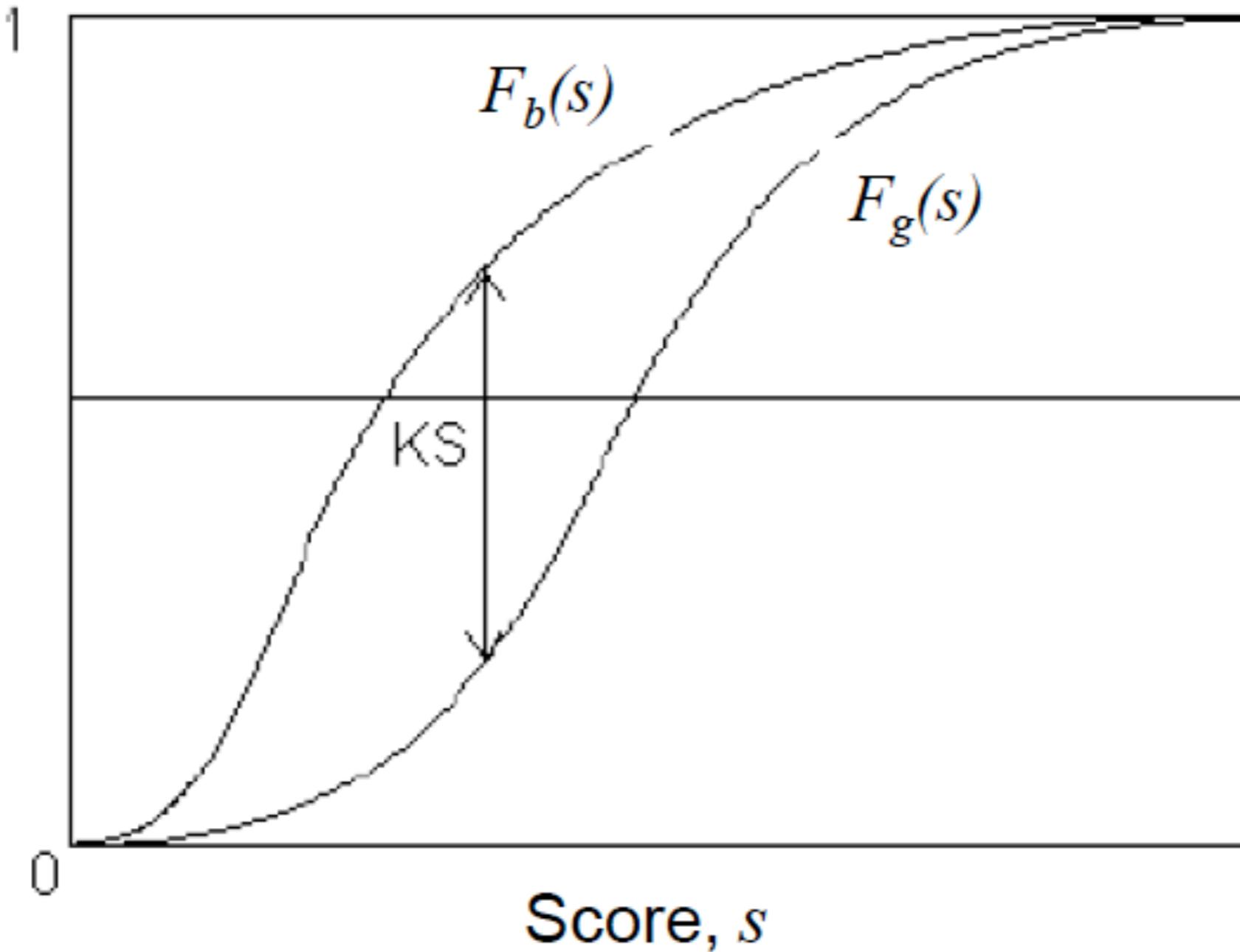


评价



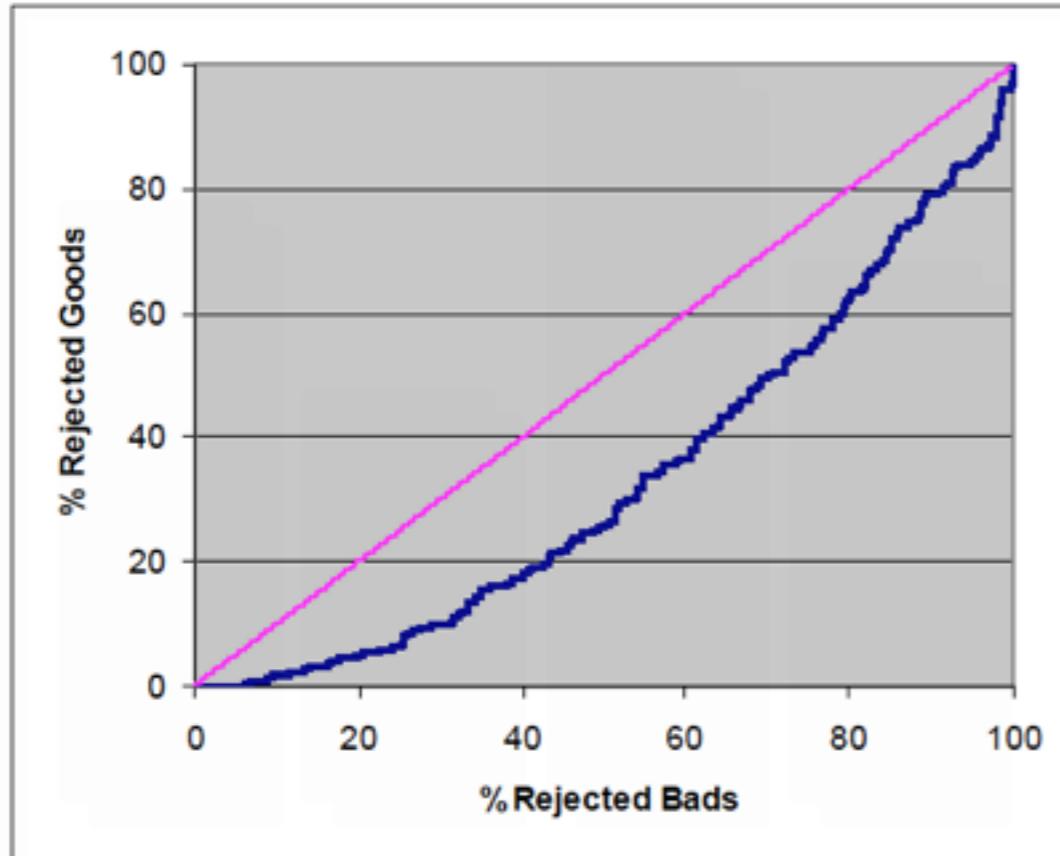
Credit Scoring II

KS



Credit Scoring II

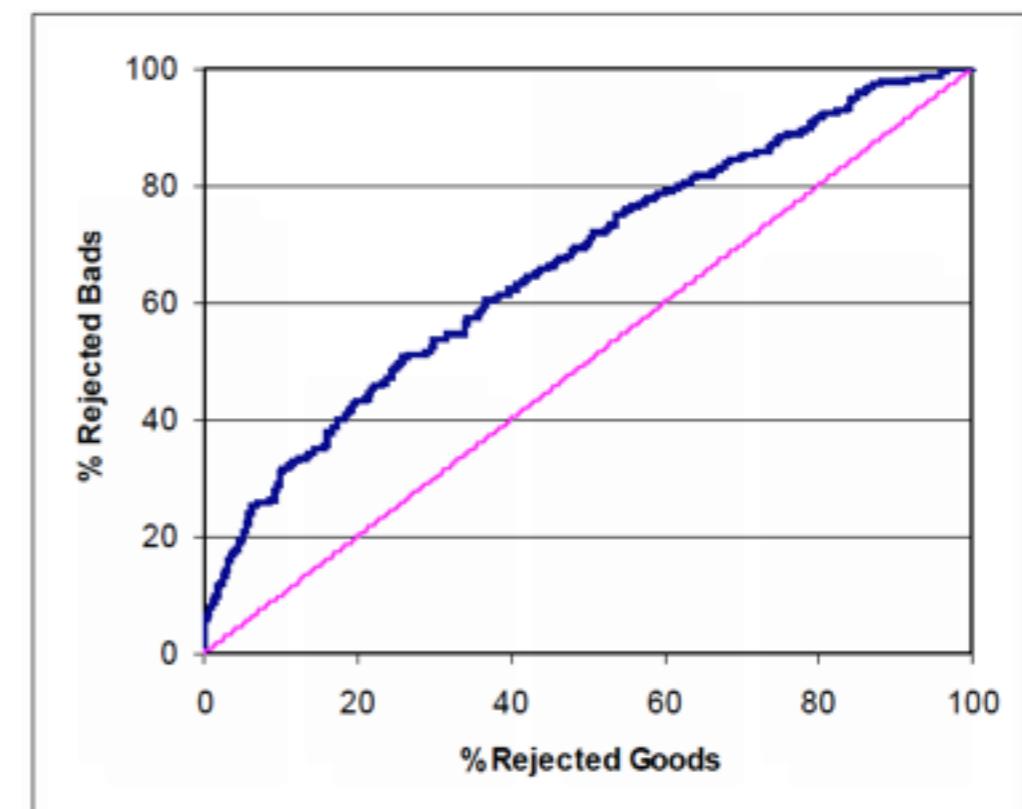
ROC



or Lorenz diagram

$F_b(s)$

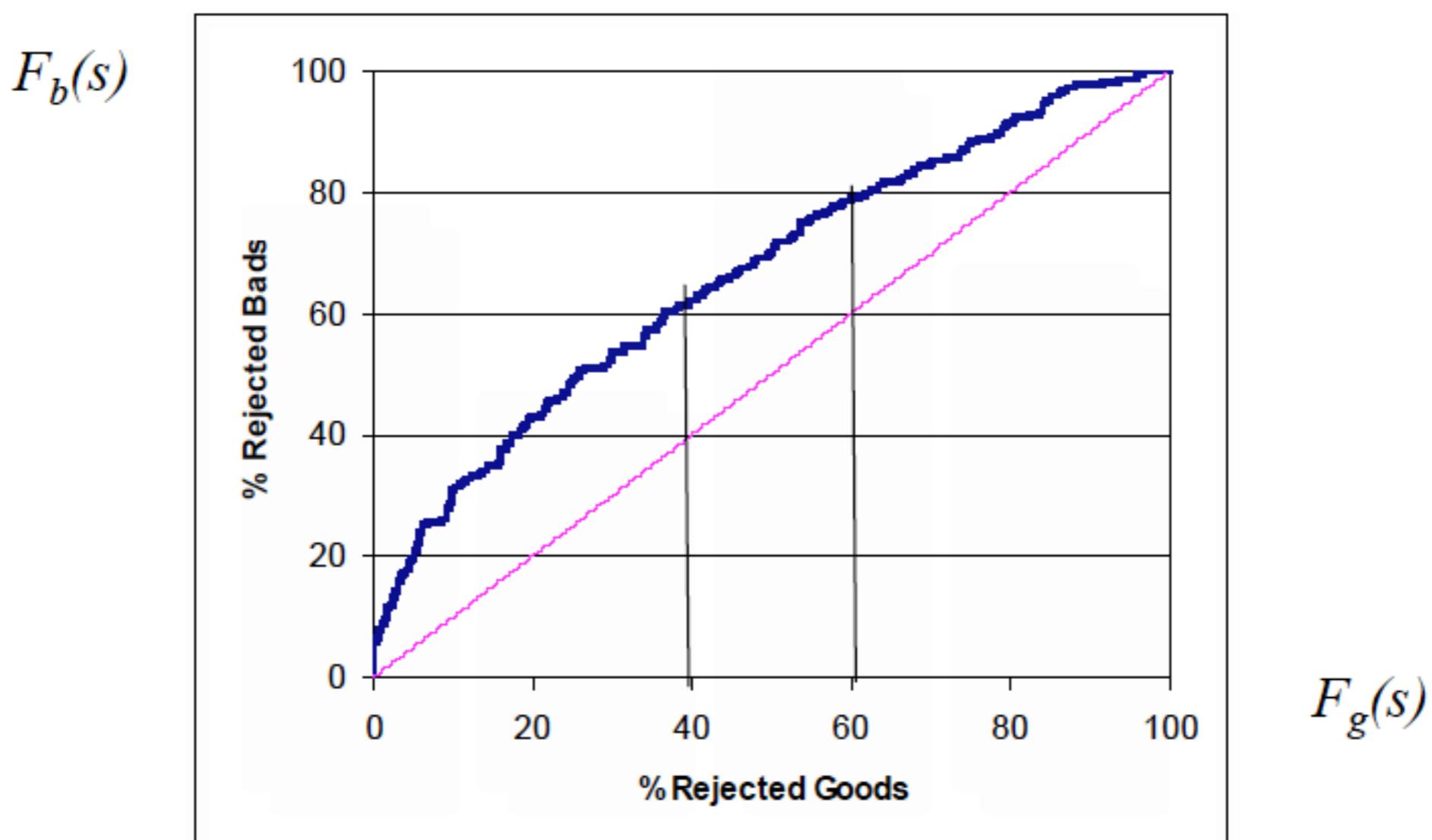
$F_g(s)$



$F_g(s)$

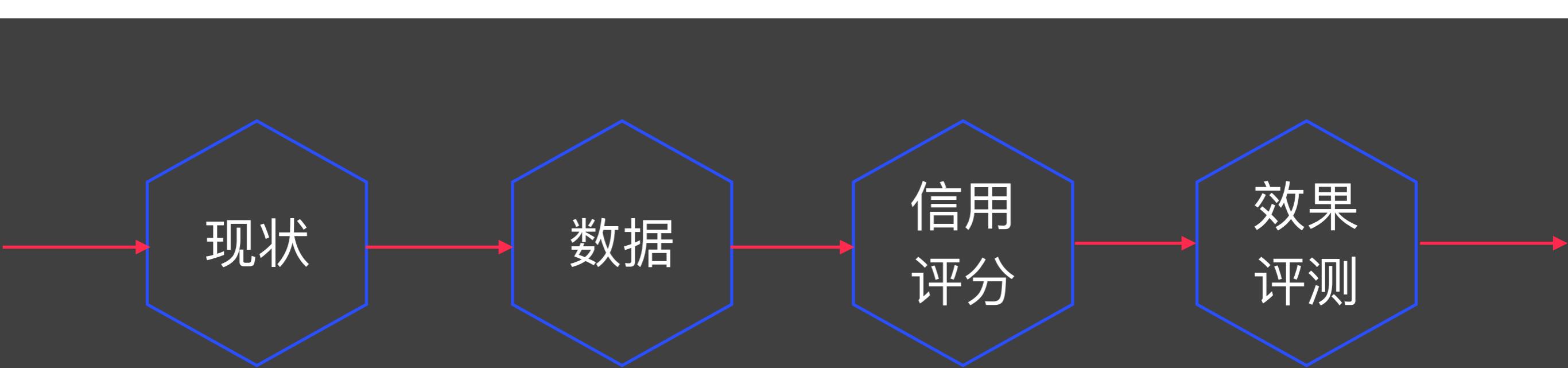
Credit Scoring II

基尼系数



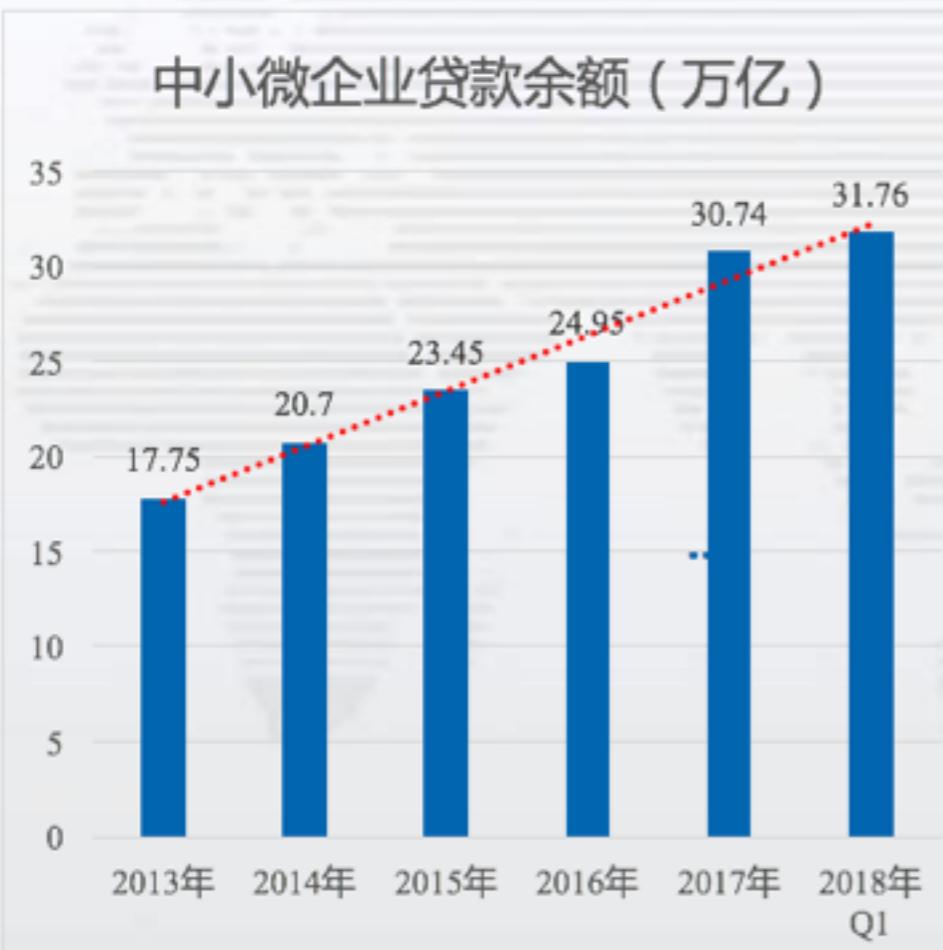
Classification algorithm	Hold-out AUC
LogR	0.779
LogR with interaction variables	0.777
SVM: linear	0.783
SVM: polynomial	0.755
SVM: Gaussian RBF	0.783
LDA	0.781
kNN	0.756

中小微企业 信用评分



中小微企业

中国中小微企业贷款规模快速增长

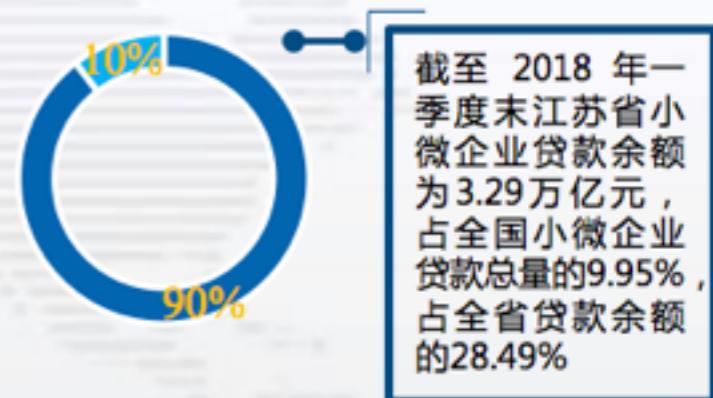


中小微企业贷款需求高



2017年7月小微
企业数量
7328.1万户，
小微企业贷款户
数约1545万户，
约占21%

江苏中小微企业贷款余额占比高



截至2018年一
季度末江苏省小
微企业贷款余额
为3.29万亿元，
占全国小微企业
贷款总量的9.95%，
占全省贷款余额
的28.49%

国家频繁发布政策，加大对中小微企业的扶持和支持力度

2018年11月9日主持召开国务 院常务会议

- 国务院总理李克强要求加大金融支持缓解民营企业特别是小微企业融资难融资贵。
- 从大型企业授信规模中拿出一部分，用于增加小微企业贷款。

2018年6月25日，五部委联合 发布银发〔2018〕162号)

- 加大金融科技等产品服务创新。银行业金融机构要加强对互联网、大数据、云计算等信息技术的运用，改造信贷流程和信用评价模型，降低运营管理成本，提高贷款发放效率和服务便利度。

中小微企业数据

- **企业信息类8个特征变量**

企业经营年限、企业注册资本、企业所在区县、一般纳税人资质、企业类型、所属行业、法人持股比例、商变更情况

- **实际控制人6个特征变量**

法人年龄、婚姻状况、子女情况、户籍种类、住房情况、申请人本行业从业年限

- **经营发票数据14个特征变量**

销项发票计算的销售额、主要销售地区省内、主要下游客户经营年限、主要下游客户企业类型、主要下游客户行业、红冲发票比例、无效发票比例、专票占比、近24个月月波动率、近24个月季度波动率、近24个月交易方一致性、近24个月集中度、销售额全国企业中排名、销售额行业排名

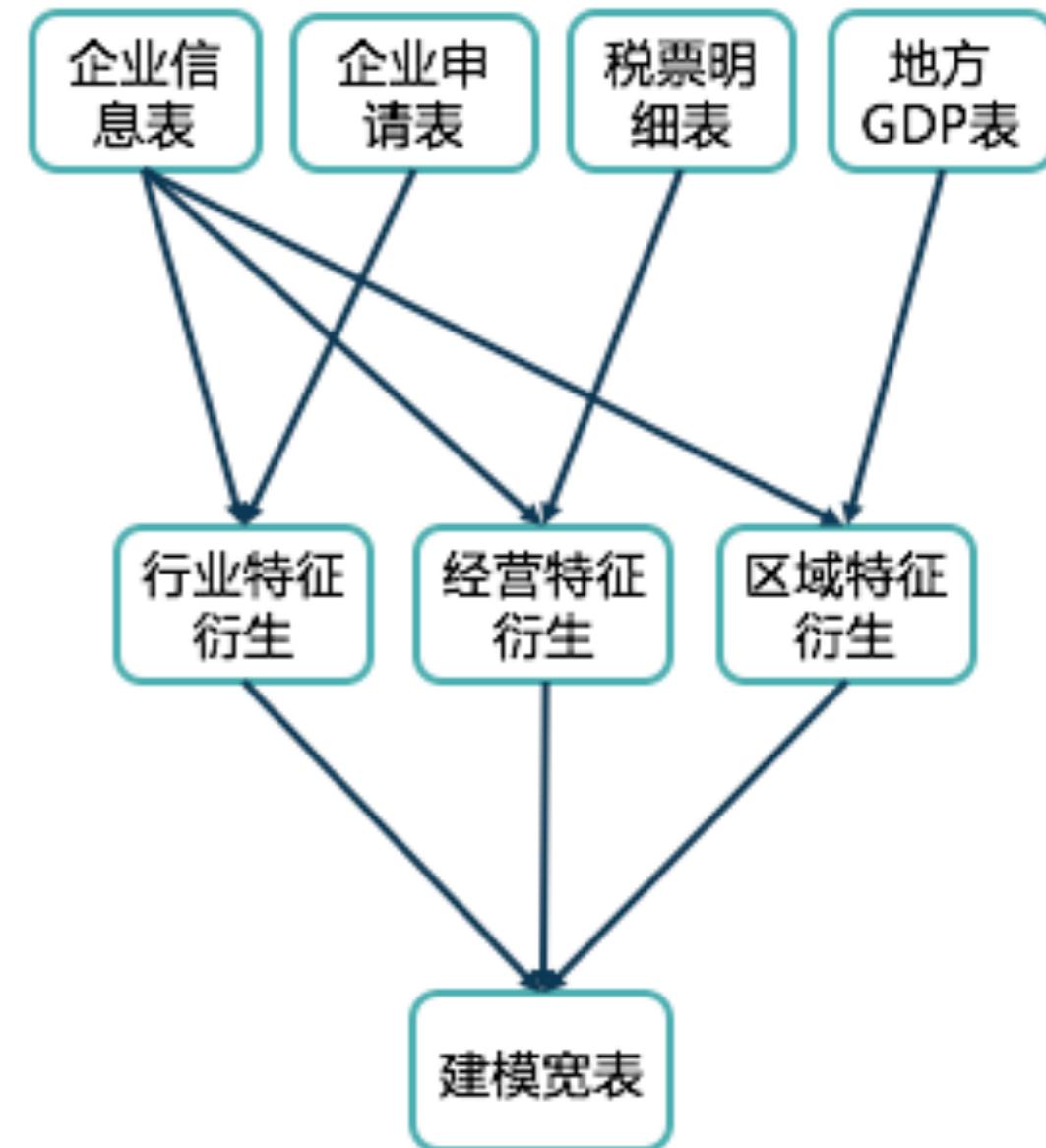
- **企业风险信息类2个特征变量**

企业有不良声誉记录的（如上过有关机构或部门的黑名单），

企业实际控制人有赌博、吸毒等不良嗜好的

- **综合评价类2个特征变量**

企业一致性指数(企业销售客户稳定性(下游))、企业授信倾向分。



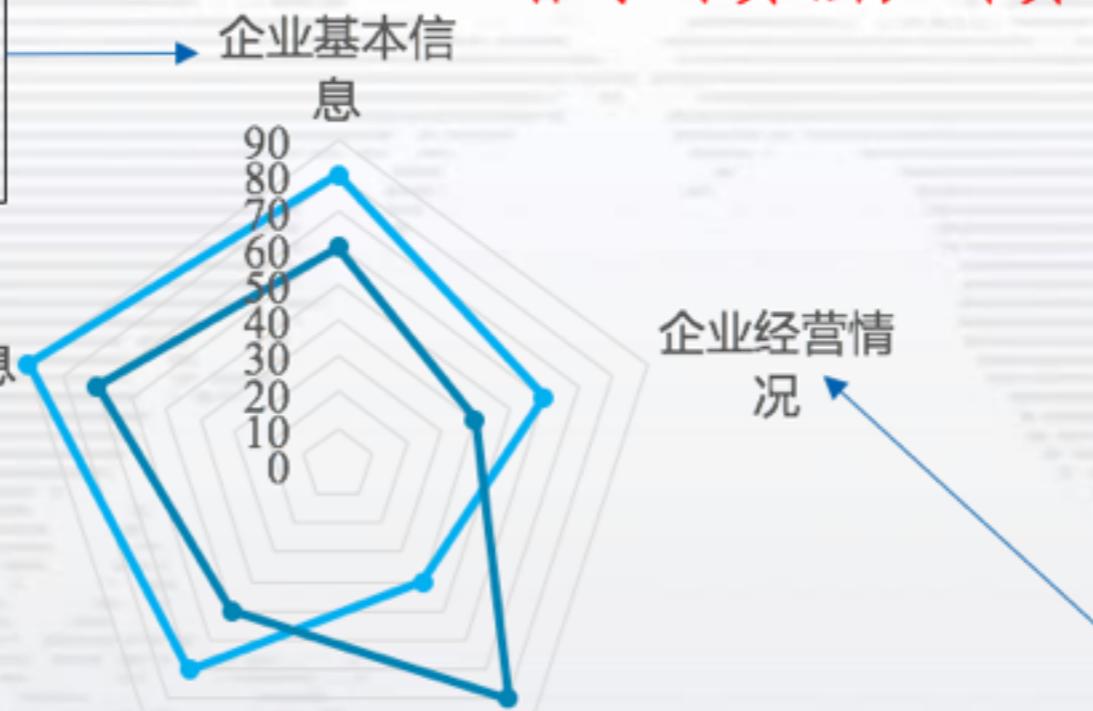
信用评分

行业
规模
注册年份
股东及出资信息
地区
企业性质
法定代表人
.....

诉讼信息
失信被执行人
行政处罚
环保处罚
税收违法
统计失信
环保失信
食品药品抽检不合格
环保违法
税票机开机情况
.....

行业排名
规模排名
经营年限排名
地区排名
增长率排名
.....

基于计量经济分析，通过大数据以及机器学习算法，计算中小微企业信用评分



行业增长率
地区GDP
地区经济增长率
地区PPI
行业利润率
地区人口
地区CPI
.....

年税票总额
主营商品
平均月税票额
年度同比税票额增长
季度数票额方差
最大季度税票额
最小季度税票额
最大月度税票额
最小月度税票额
有税票月份数
无税票月份数
专用增值税票金额
普票金额
红冲税票金额
无效税票金额
下游企业家数
.....

效果评测

- **模型超参数：**

学习速率：0.05

决策树最大深度：3

gbdt中决策树的棵数：76

- **模型效果：**

准确率：0.664

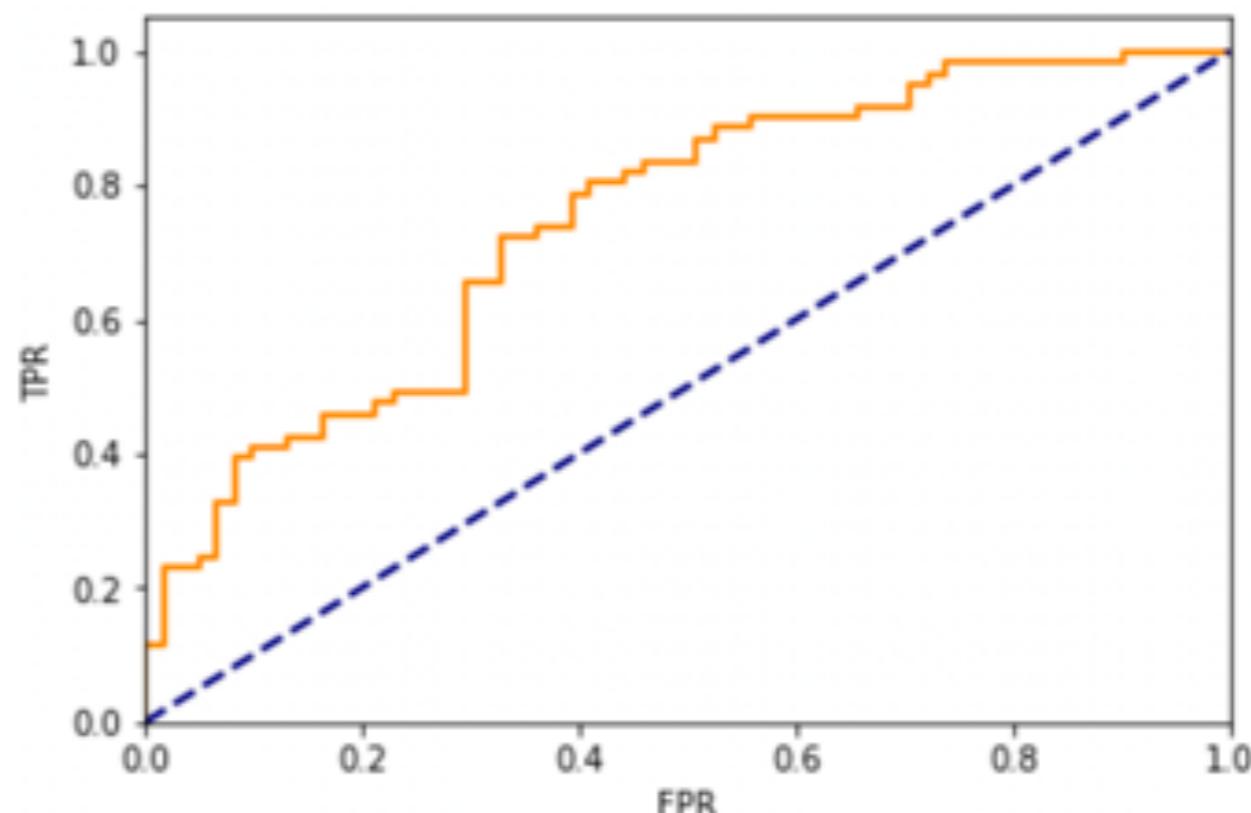
AUC：0.747

正样本精确率：0.61,

正样本召回率：0.90,

负样本精确率：0.81,

负样本召回率：0.43

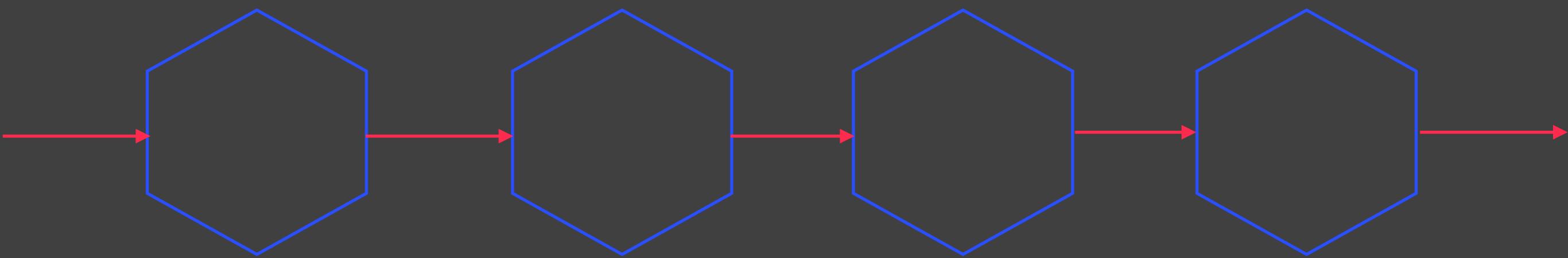


Credit Scoring II

未来



课后作业



课后作业

各组选择的课程项目

1、项目报告

3、项目汇报PPT

2、相关参考文献

4、代码可以演示

12月30日课堂上课程汇报

謝謝 !

Huijing Sun

sunhp@ss.pku.edu.cn

<https://huijingsun.github.io>