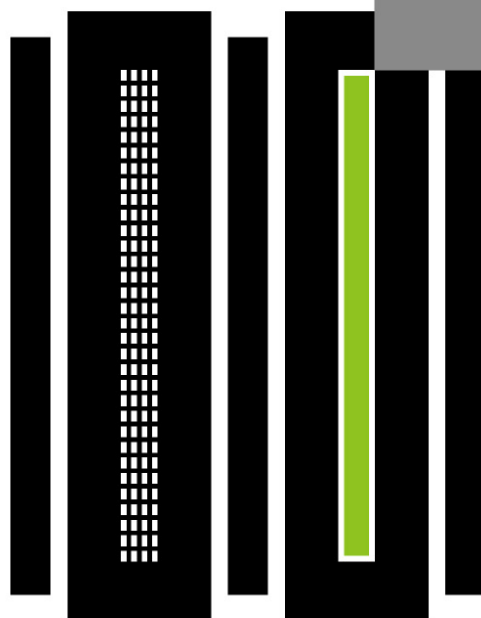


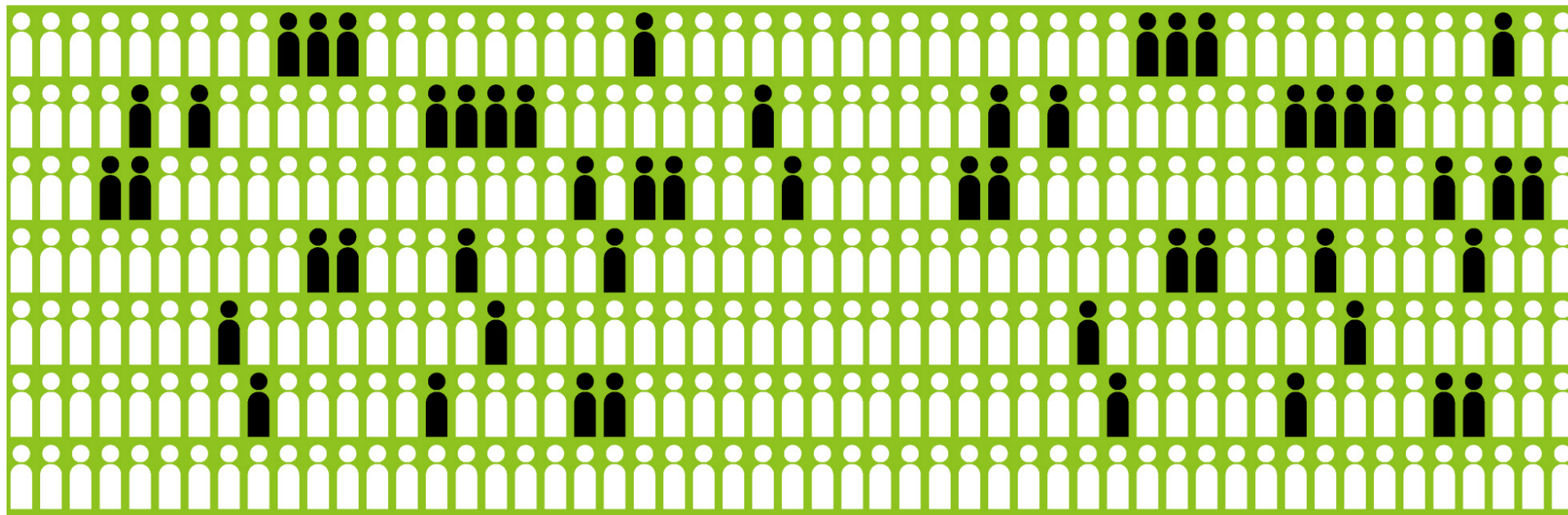
分析为道 **Z** 者见智

IBM主机商业分析(BA)高峰论坛



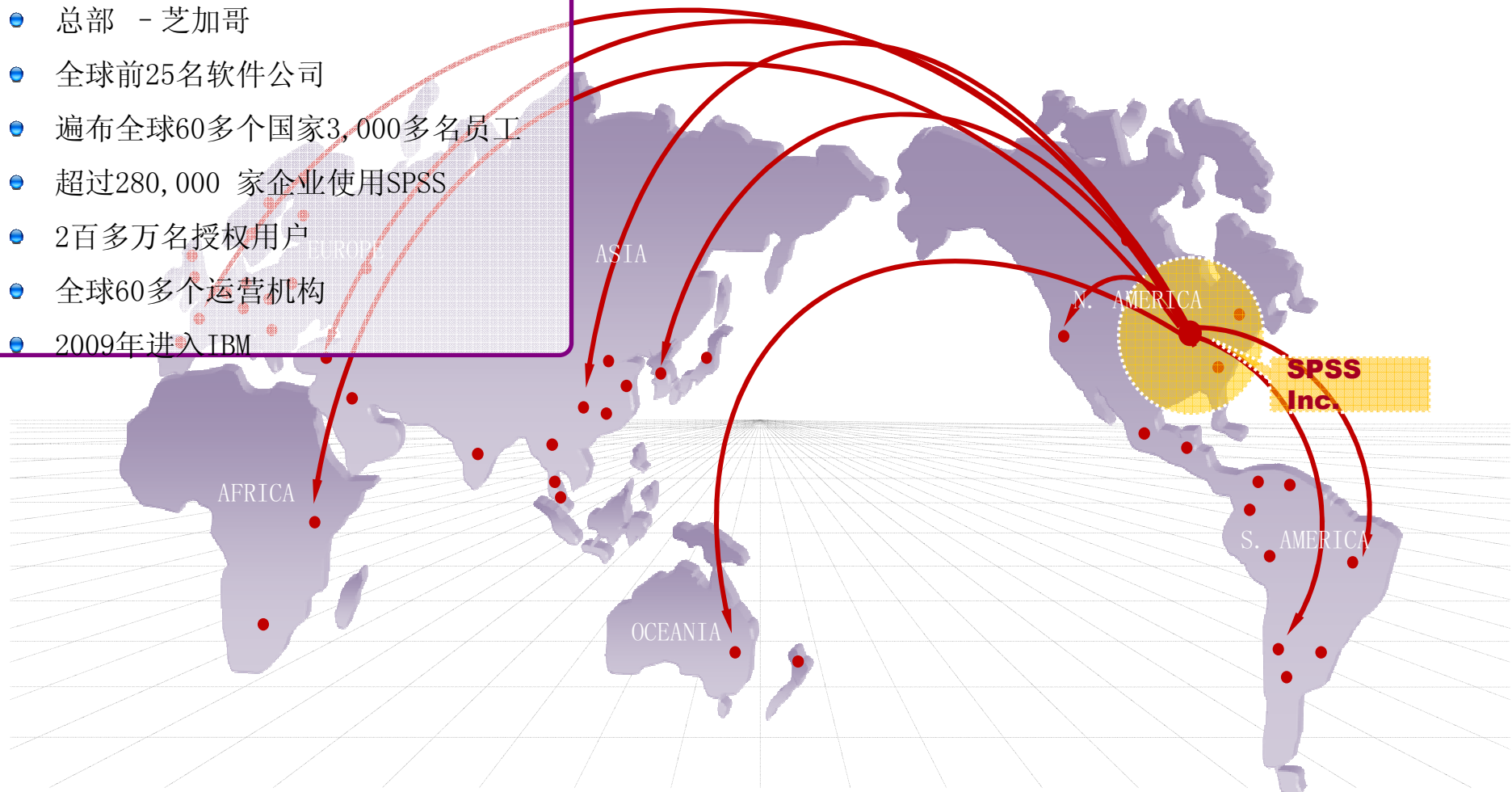
SPSS预测分析对主机用户的价值

王清华

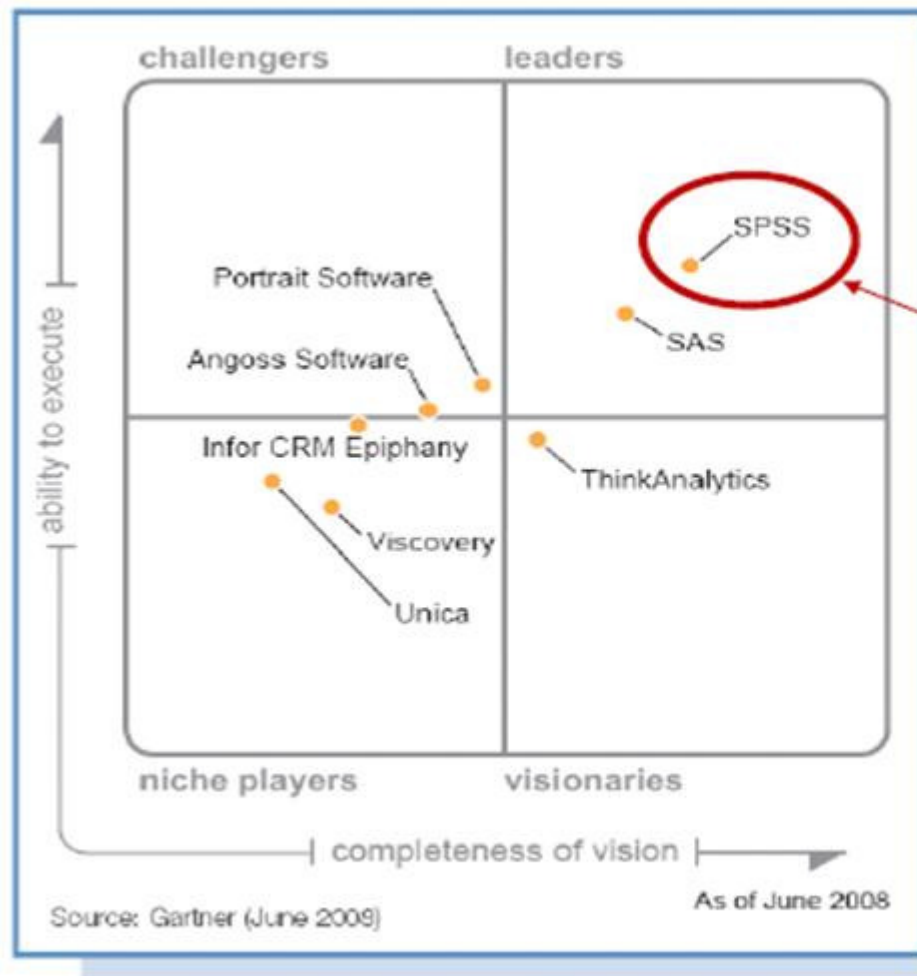


> SPSS

- 1968年成立
- 40多年的数据分析技术积累
- 1993年纳斯达克上市
- 总部 - 芝加哥
- 全球前25名软件公司
- 遍布全球60多个国家3,000多名员工
- 超过280,000 家企业使用SPSS
- 2百多万名授权用户
- 全球60多个运营机构
- 2009年进入IBM



Gartner Magic Quadrant for Customer Data Mining Applications



**SPSS is
THE leader**

SPSS 预测分析产品 (PASW)

4种分类产品

- ➔ 数据收集
为客户的态度和观点提供准确的视图
 - IBM SPSS Data Collection
- ➔ 统计分析
为分析人员提供专业的统计分析结果，提高决策可信度
 - IBM SPSS Statistics
- ➔ 预测模型
建立预测模型，提供直观的辅助决策信息
 - IBM SPSS Modeler
 - IBM SPSS Text Mining
- ➔ 部署
在企业日常运营中，降低分析维护成本，最大限度地提高数据分析为企业带来的价值。
 - IBM SPSS Decision Management
 - IBM SPSS Collaboration & Deployment Services



预测分析过程

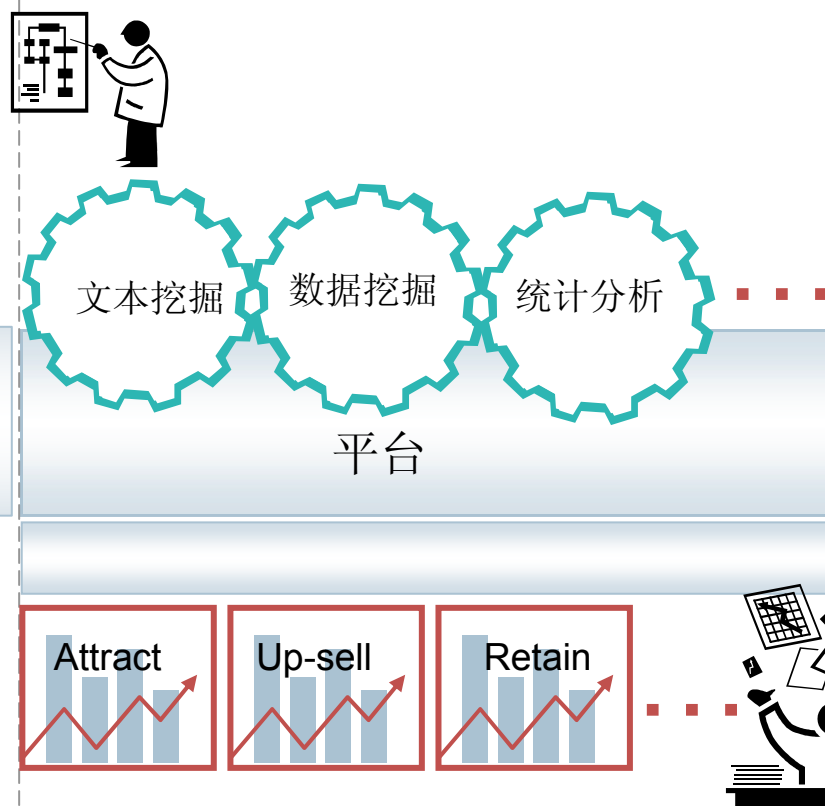
获取数据

为客户的态度和观点提供准确的视图



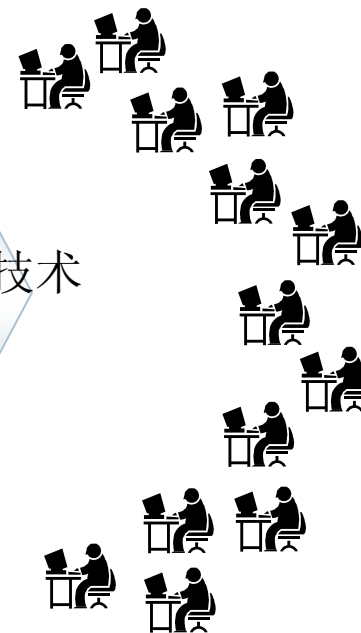
预测

为您的决策提供可信的结果，使您的决策更科学



行动

将结果有效地发布到您的业务系统，从而最大程度发挥数据挖掘的价值



什么是预测分析——一个市场营销的例子

Table Edit Generate Help			
id	age	gender	responder
ID12106	57	FEMALE	YES
ID12108	58	MALE	NO
ID12120	31	MALE	NO
ID12138	36	FEMALE	YES

女性对市场活动做出回应，男性对市场活动不做出回应，和年龄无关

在数据中发现有价值的规则或者模式

什么是预测分析——一个市场营销的例子

Table Edit Generate Help			
id	age	gender	responder
ID12106	57	FEMALE	YES
ID12108	58	MALE	NO
ID12120	31	MALE	NO
ID12138	36	FEMALE	YES

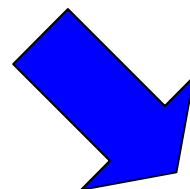


Table Edit Generate Help			
id	age	gender	responder
ID12106	57	FEMALE	YES
ID12108	58	MALE	NO
ID12120	31	MALE	NO
ID12121	61	MALE	YES
ID12138	36	FEMALE	YES

数据变的复杂会如何？

女性对市场活动做出回应，老年男性也可能对市场活动做出回应

什么是预测分析——一个市场营销的例子

通过数据挖掘发现回应的5条规则:

1、如果收入大于29622, 有孩子, 并且孩子的数量小于等于2, 那么对市场活动会回应

.....

通过数据挖掘发现不回应的5条规则:

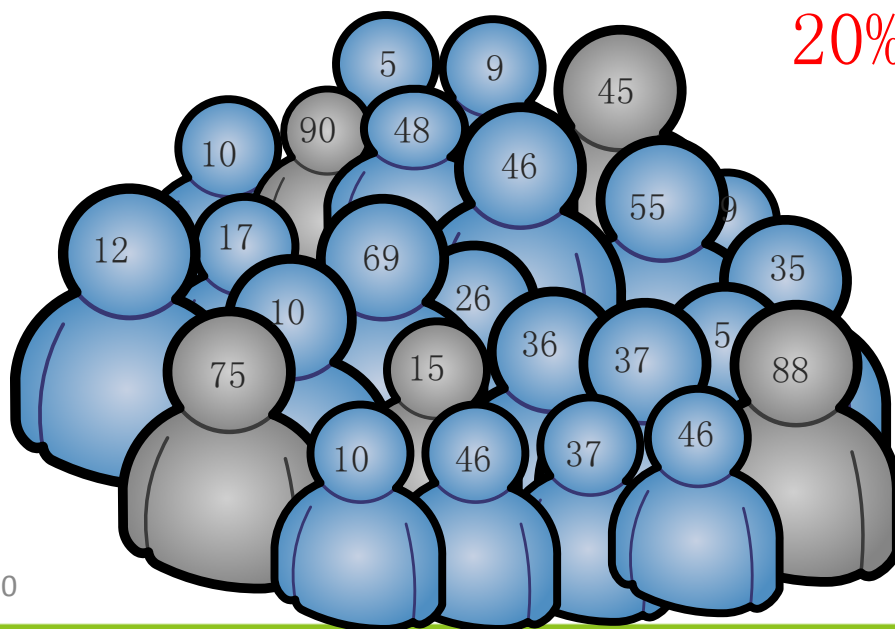
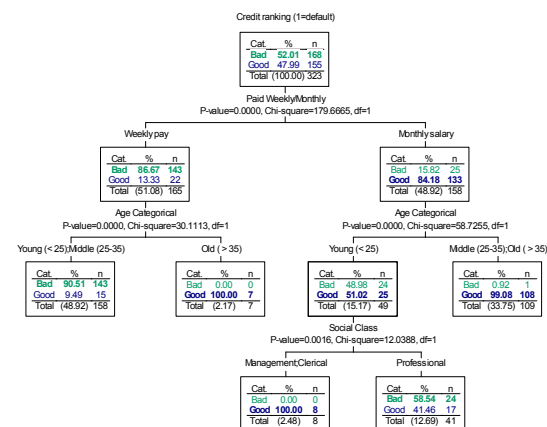
1、如果收入小于12640.3, 并且有一个孩子, 那么对市场活动不会回应

.....

table (12 fields, 300 records) #1												
	id	age	sex	region	income	married	children	car	save_act	current_act	mortgage	responder
1	ID12101	48	FEMALE	INNER_CITY	17546.000	NO		1 NO	NO	NO	NO	YES
2	ID12102	40	MALE	TOWN	30085.100	YES		3 YES	NO	YES	YES	NO
3	ID12103	51	FEMALE	INNER_CITY	16575.400	YES		0 YES	YES	YES	NO	NO
4	ID12104	23	FEMALE	TOWN	20375.400	YES		3 NO	NO	YES	NO	NO
5	ID12105	57	FEMALE	RURAL	50576.300	YES		0 NO	YES	NO	NO	NO
6	ID12106	57	FEMALE	TOWN	37869.600	YES		2 NO	YES	YES	NO	YES
7	ID12107	22	MALE	RURAL	8877.070	NO		0 NO	NO	YES	NO	YES
8	ID12108	58	MALE	TOWN	24946.600	YES		0 YES	YES	YES	NO	NO
9	ID12109	37	FEMALE	SUBURBAN	25304.300	YES		2 YES	NO	NO	NO	NO
10	ID12110	54	MALE	TOWN	24212.100	YES		2 YES	YES	YES	NO	NO
11	ID12111	66	FEMALE	TOWN	59803.900	YES		0 NO	YES	YES	NO	NO
12	ID12112	52	FEMALE	INNER_CITY	26658.800	NO		0 YES	YES	YES	YES	NO
13	ID12113	44	FEMALE	TOWN	15735.800	YES		1 NO	YES	YES	YES	YES
14	ID12114	66	FEMALE	TOWN	55204.700	YES		1 YES	YES	YES	YES	YES
15	ID12115	36	MALE	RURAL	19474.600	YES		0 NO	YES	YES	YES	NO
16	ID12116	38	FEMALE	INNER_CITY	22342.100	YES		0 YES	YES	YES	YES	NO
17	ID12117	37	FEMALE	TOWN	17729.800	YES		2 NO	NO	NO	YES	NO
18	ID12118	46	FEMALE	SUBURBAN	41016.000	YES		0 NO	YES	NO	YES	NO
19	ID12119	62	FEMALE	INNER_CITY	26909.200	YES		0 NO	YES	NO	NO	YES
20	ID12120	31	MALE	TOWN	22522.800	YES		0 YES	YES	YES	NO	NO
21	ID12121	61	MALE	INNER_CITY	57880.700	YES		2 NO	YES	NO	NO	YES
22	ID12122	50	MALE	TOWN	16497.300	YES		2 NO	YES	YES	NO	NO
23	ID12123	54	MALE	INNER_CITY	38446.600	YES		0 NO	YES	YES	NO	NO
24	ID12124	27	FEMALE	TOWN	15538.800	NO		0 YES	YES	YES	YES	NO
25	ID12125	22	MALE	INNER_CITY	12640.300	NO		2 YES	YES	YES	NO	NO
26	ID12126	56	MALE	INNER_CITY	41034.000	YES		0 YES	YES	YES	YES	NO
27	ID12127	45	MALE	INNER_CITY	20809.700	YES		0 NO	YES	YES	YES	NO
28	ID12128	39	FEMALE	TOWN	20114.000	YES		1 NO	NO	YES	NO	YES
29	ID12129	39	FEMALE	INNER_CITY	29359.100	NO		3 YES	NO	YES	YES	NO
30	ID12130	61	MALE	RURAL	24270.100	YES		1 NO	NO	YES	NO	YES
31	ID12131	61	FEMALE	RURAL	22942.900	YES		2 NO	YES	YES	NO	NO
32	ID12132	20	FEMALE	TOWN	16325.800	YES		2 NO	YES	NO	NO	NO
33	ID12133	45	MALE	SUBURBAN	23443.200	YES		1 YES	YES	YES	NO	YES
34	ID12134	33	FEMALE	INNER_CITY	29921.300	NO		3 YES	YES	NO	NO	NO
35	ID12135	43	MALE	SUBURBAN	37521.900	NO		0 NO	YES	YES	NO	YES
36	ID12136	27	FEMALE	INNER_CITY	19868.000	YES		2 NO	YES	YES	NO	NO
37	ID12137	19	MALE	RURAL	10953.000	YES		3 YES	YES	YES	NO	NO

数据挖掘可以从异常复杂的数据中发现规律

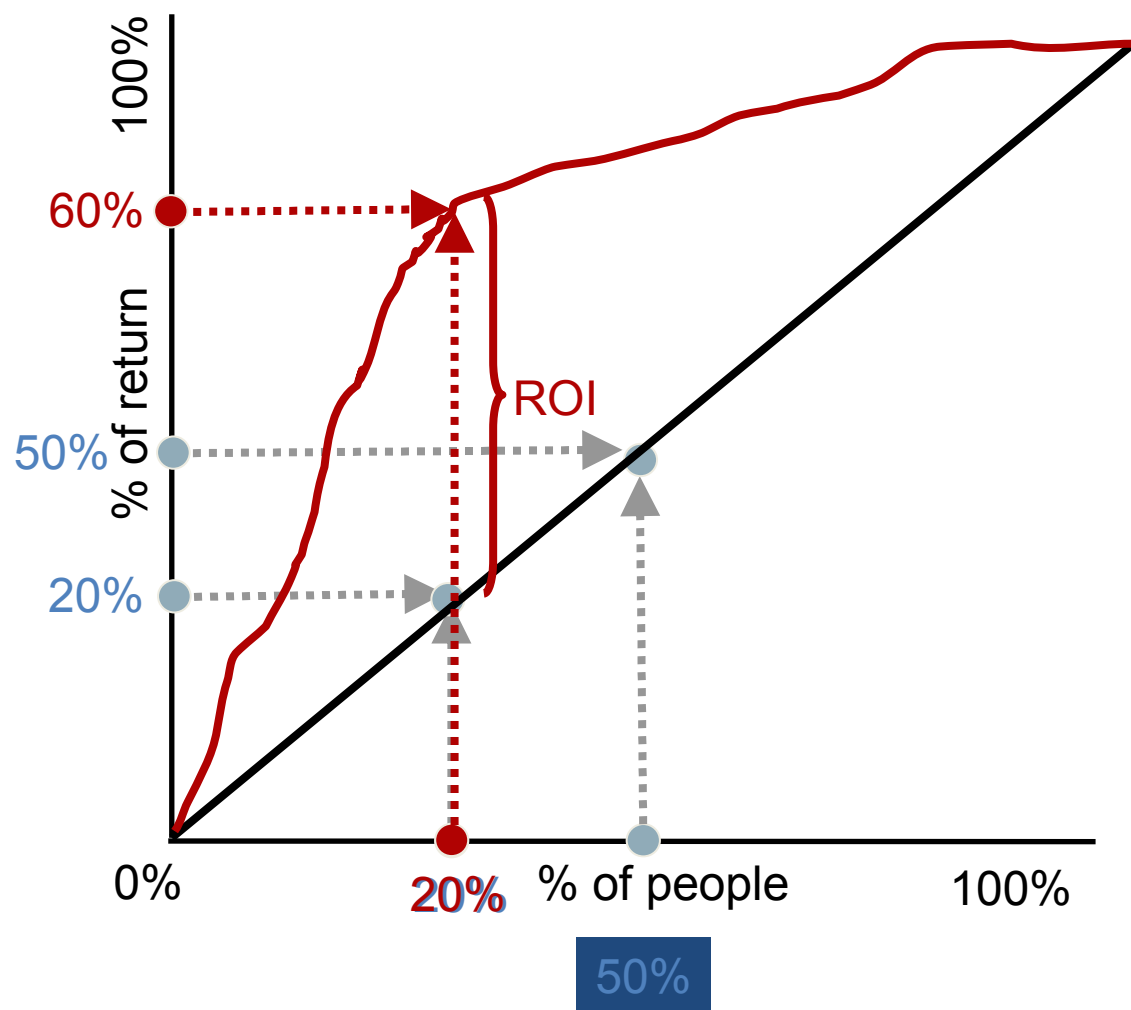
客户重点营销



20%

60%

数据挖掘的价值

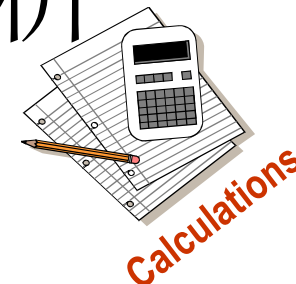


数据挖掘效果模拟分析

目的：发现新客户(使响应率从1%提高到1.2%)

	数据挖掘以前	数据挖掘以后	差别
发信的数量	1, 000, 000	750, 000	(250, 000)
成本	¥ 1, 000, 000	¥ 750, 000	(¥ 250, 000)
响应的数量	10, 000	9, 000	(1, 000)
每个响应的收入	¥ 125	¥ 125	¥ 0
总收入	¥ 1, 250, 000	¥ 1, 125, 000	(¥ 125, 000)
净利润	¥ 250, 000	¥ 375, 000	¥ 125, 000
建模的费用	0	40, 000	¥ 40, 000
最终的利润	¥ 250, 000	¥ 335, 000	¥ 85, 000

一些数据挖掘应用的ROI分析



交叉销售模型

假设:

- 帐户数量 = 200万
- 每一个帐户每年促销次数 = 12
- 每次促销成本 = 0.30

使用预测型模型减少20%的促销而使得总回应率基本保持不变

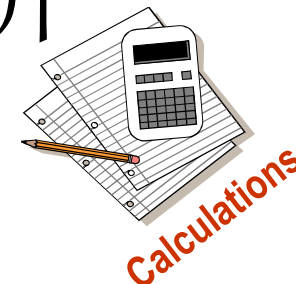
建模前成本= $200\text{万} \times 12 \times 0.30 = 720\text{万}$

建模后成本= $160\text{万} \times 12 \times 0.30 = 576\text{万}$

每年节约= 144万

一些数据挖掘应用的ROI分析

客户挽留



假设:

- 帐户数量= 200万
- 每个帐户每年利润 = 350元
- 每年帐户流失率 = 5%

通过挽留策略使流失率降低10%

每年由于5%的流失造成利润损失= $5.0\% * 200万 * 350 = 3500万$

每年由于4.5%的流失造成利润损失= $4.5\% * 200万 * 350 = 3150万$

每年节省= 350万

一些数据挖掘应用的ROI分析

市场细分

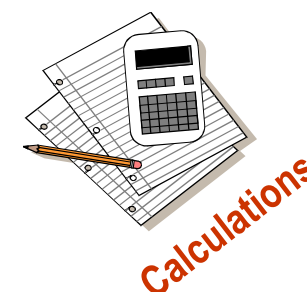
假设:

- 帐户数量= 200万
- 每一个帐户每年促销次数 = 12
- 平均响应率= 1%
- 每个响应带来利润 = 350元

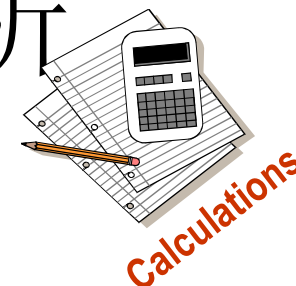
响应率提高5%(从1% 到 1.05%)

每年收益= $0.05\% * 200万 * 12 * 350$

每年收益= 420万



一些数据挖掘应用的ROI分析



信用评分

假设:

- 贷款笔数= 10000笔
- 平均贷款金额 = 20000
- 平均违约率 = 5%

违约率降低到4.5%

违约率5%的损耗= $10000 * 20000 * 5\% = \$1000\text{万/年}$

违约率4.5%的损耗= $10000 * 20000 * 4.5\% = 900\text{万/年}$

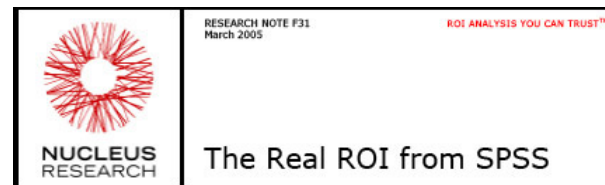
每年减少损耗= 100万

ROI回报

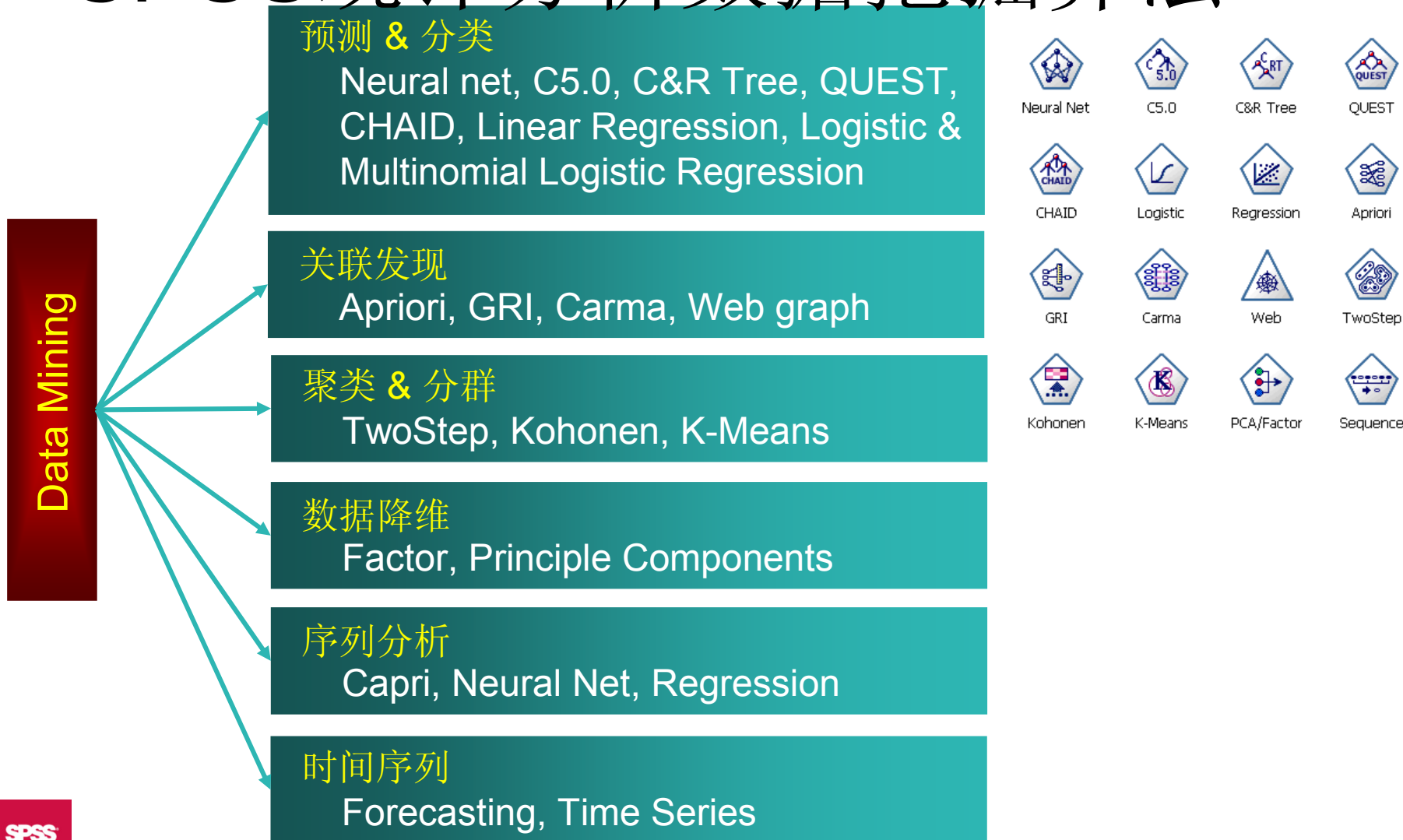
- 94%的客户通过实施预测分析项目得到了正的ROI，平均仅用10.7个月就可以收回投入
- 超过90%的客户通过预测技术提高的“生产力”
- 81%的项目按时发布，75%的项目小于等于预算

“这是迄今为止Nucleus看到的最好投资回报率（This is one of the highest ROI scores Nucleus has ever seen in its Real ROI series of research reports）”

Rebecca Wettemann, VP of Research, Nucleus Research



SPSS统计分析数据挖掘算法



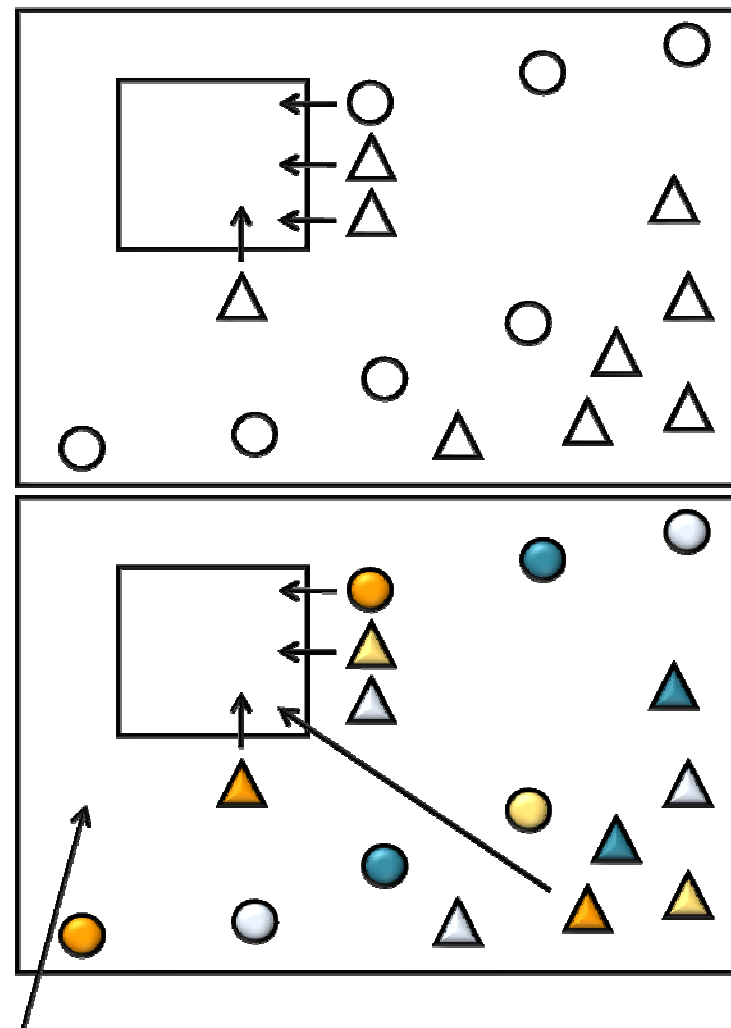
SPSS应用领域



价值体现：获取



- 通过对现有客户的研究，找到相似客户开展市场活动，减少市场活动成本，并提高响应率
- 找到潜在的最有价值客户
- 典型应用行业：
 - 银行
 - 保险
 - 电信
 - 零售





Standard Life Bank通过研究客户再抵押响应的历史数据研究，对客户聚类并发现潜在客户

- 产生了 **9倍的响应率提升**
- 单次市场活动获得了 **£33M的抵押申请收入**



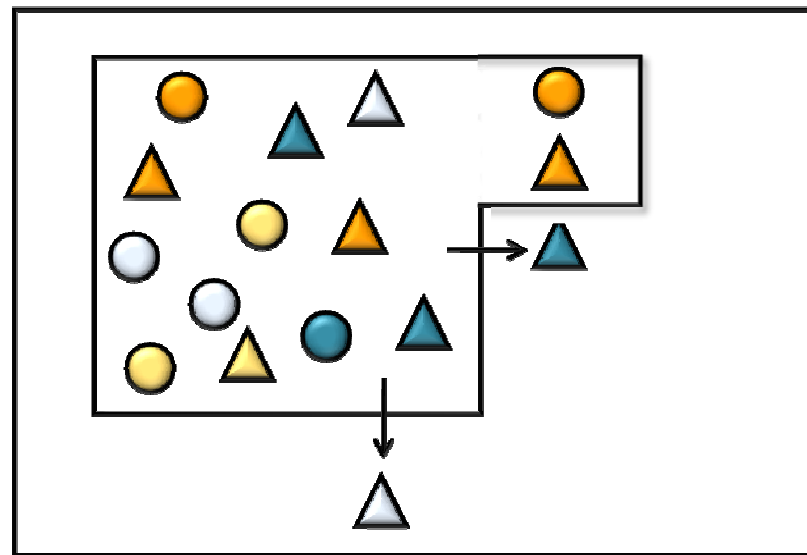
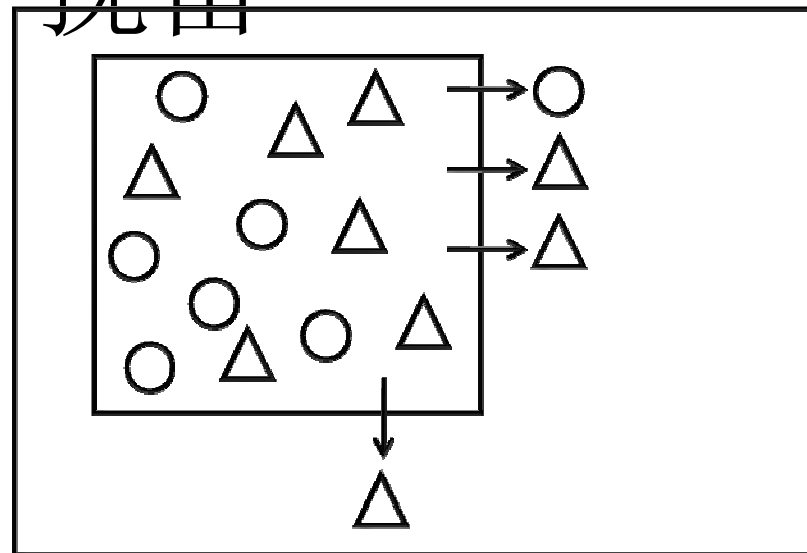
对市场获取建立预测模型，并依据响应率模型开展市场推广

- 响应率 **翻番**
- 获取到 **高价值** 客户

价值体现: 挽留



- 识别流失潜在客户
- 对具有流失高风险的高价值客户进行流失挽留
- 依据客户特征，选取最适合该客户的产品和服务进行推荐
- 典型应用行业：
 - 电信、保险、零售、银行



预测型呼叫中心示例

SPSS Deployment Demonstration - [SPSS]

Demos

Simulation Call Center

Customer: Julie Chesson Customer ID: ID12887

SPSS

Customer Search

First Name: Julie
Last Name: Chesson
Search

Results			
	id	lastname	firstname
	ID12895	Ford	Constance
▶	ID12887	Chesson	Julie
	ID12852	Drake	Sally
	ID12839	Davidson	Claudette

Select
Clear

Customer Profile

Married: NO
Income: 16583.8
Children: 3
Region: SUBURBAN
Loyalty: NO

Contact Notes

Type: Complaint
Channel: Phone
Time: 3:27:40 PM
Date: 7/29/2003
Note Concept(s): minute_charges, new_phone

Note Summary / Detail:
New plan request
Does not have enough minutes so is getting charged penalties. Also, phone is outdated. Would like a new phone asap.

Save Notes

Call Scripting

Script History Tasks Sales Products Service

Offer Name: RET -600 Midwest minutes - \$19/mo
Prediction: Voluntary Churn
Probability: 0.965 HIGH
3 Matching Offers

Script:
As a valued customer, we would like to offer you a special promotion of 600 Midwest minutes for only \$19/mo. Can I switch you to your personal plan now?

Accept
Reject

1. 坐席员输入客户投诉

2. 将投诉存入数据库，并结合其他客户数据建立预测性模型

3. 实时得到客户流失评分

4. 坐席员根据动态生成的脚本进行应答



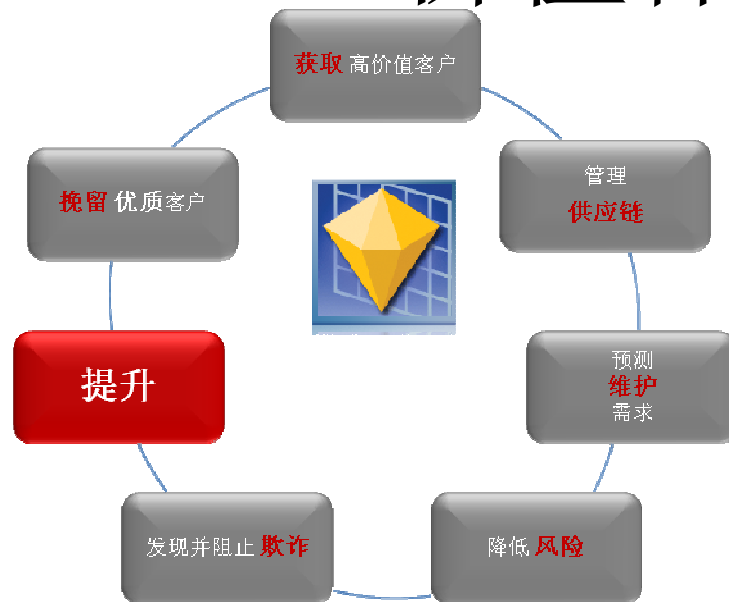
建立流失管理模型，对客户通话数据、特征数据、交互数据，及网页浏览数据进行分析，制定挽留方案

- 流失率由**大于2%减至低于1%**
- 市场挽留成本**减少60%**
- 通过**多渠道**进行挽留
- 每月刷新模型
 - 自动化更新300-400模型，基于最新数据选择最优方案和目标客户

运用预测模型识别高价值高风险客户，并选择最优方案进行客户挽留

- 通过挽留活动，一年内使银行利润**提升10-20%**

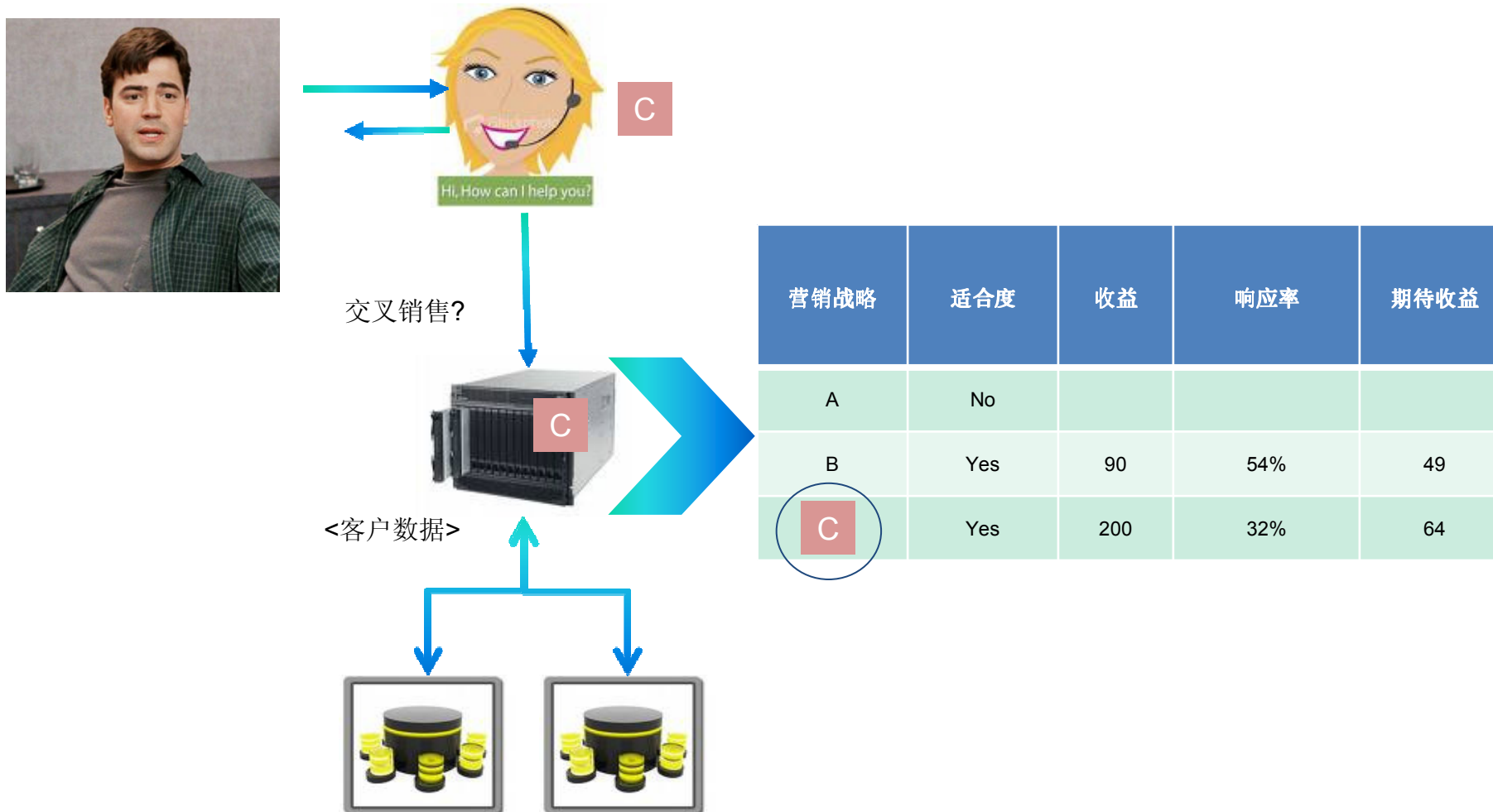
价值体现: 提升



- 通过客户购买习惯数据，分析产品间的购买关联，进行交叉销售和提升销售
- 典型应用行业：
 - 银行
 - 保险
 - 零售
 - 电信

		2010					
		A	B	C	D	E	F
2009	New						
	A						
	B						
	C						
	D						
	E						
	F						

通过呼叫中心帮助销售的决策管理过程

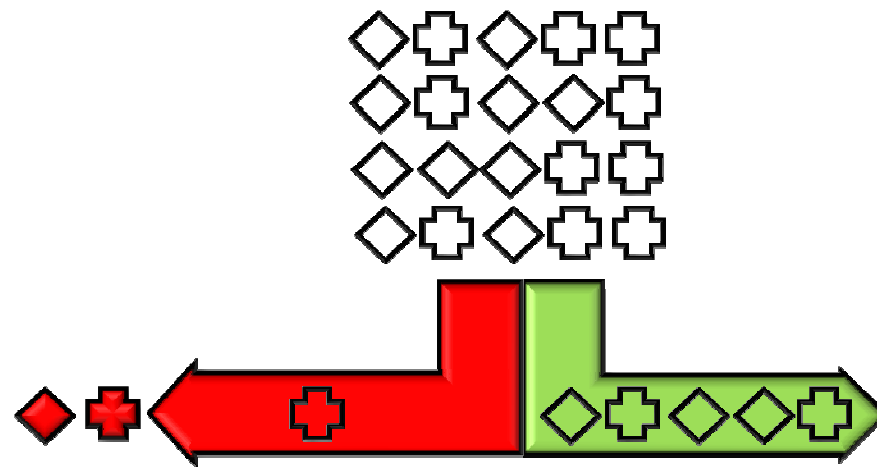
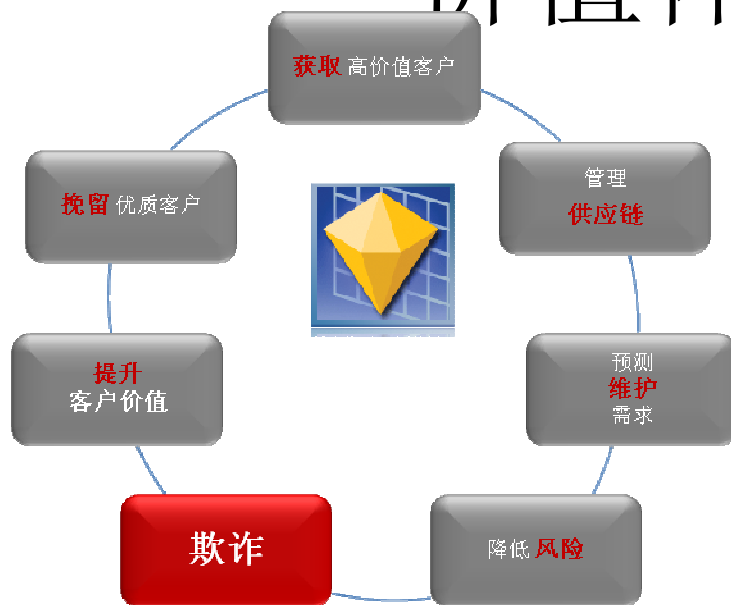




对Sofmap.com的一百多万注册用户进行分析和聚类，并建立实时的交叉销售推荐系统，给用户推荐最适合用户的产品和链接

- 网页点击率提升50%
- 销售提升18%
- 在线盈利能力提升三倍

价值体现：欺诈



- 识别欺诈模式（反复欺诈）和异常行为（一次性）
- 实时阻止潜在欺诈交易
- 标识风险原因，整合数据资源，提升调查效率
- 识别新的欺诈模式
- 典型应用行业：
 - 政府（税务、社保、交通.....）
 - 保险、银行、电信、零售



ATM欺诈识别

- 离线分析。发现欺诈模式
- 实时识别、干预

- 识别欺诈模式和异常模式
- 代理人欺诈、客户欺诈
- 确定重点调查方向
- 阻止欺诈、拒绝索赔

反洗钱



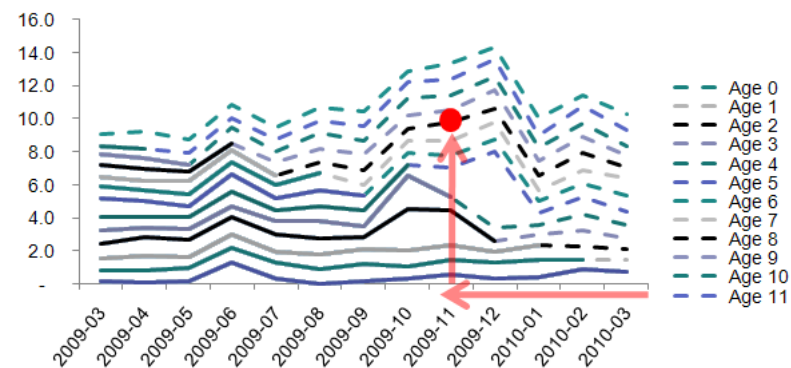
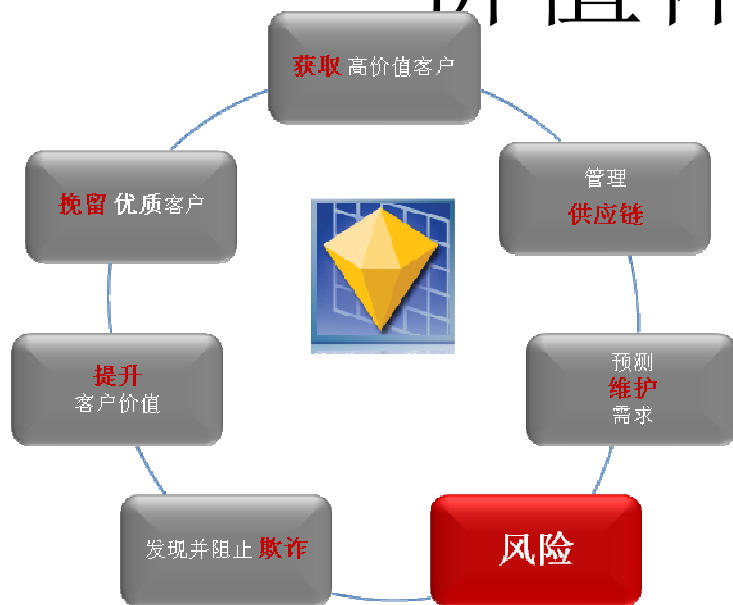
- 需要审核的可疑行为数量下降90%
- 提升识别洗钱行为的精确性，调整模型使之适应不断变化的洗钱策略
- 调查结案率由低于25%提升到60%
- 由撒网式调查变为针对性调查，节省调查成本达60%



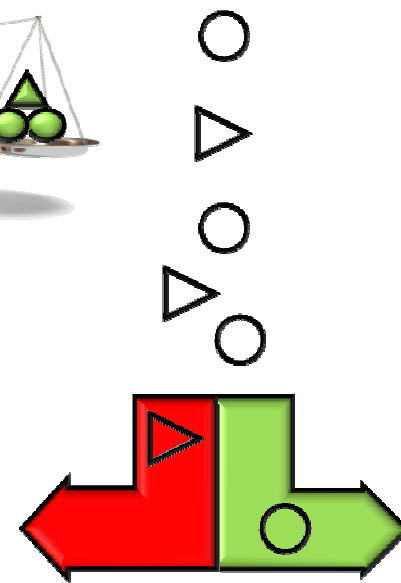
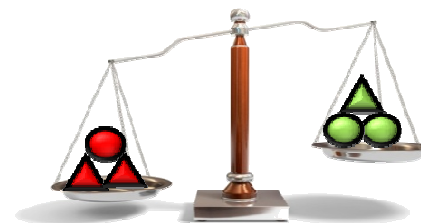
Australian Government
**Australian Transaction Reports
and Analysis Centre**

- AUSTRAC通过对跨行交易数据的分析，包含结构化和非结构化数据，来评估洗钱的相似性风险

价值体现：风险



- 基于精确的风险估计，接受或拒绝客户或交易
- 预判可能出现的风险，制定措施来降低损失
- 依据风险预测，合理进行资源、资本配置
- 典型应用行业
 - 银行
 - 其他大额交易、合同等



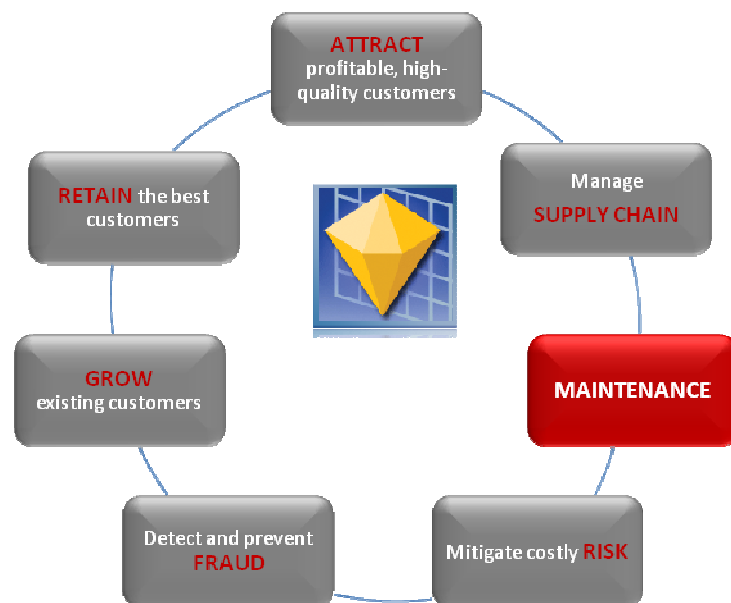
基于信用风险的操作决策



- 以前使用规则集来管理信用风险
 - 精确性和有效性不高
- 现采用IBM SPSS预测分析对信用风险进行管理，并且嵌入Cornercard的决策支持系统
- 每天对每客户进行信用风险评价
 - 制定限额，提供决策
- 坏账减少达15%以上

价值体现：维护

质量管理、失效率预测等，实现及时维护



要点:

- 根据历史的失效、故障、维护数据，建立预测模型

典型应用行业:

- 工业
- 工程
- 制造
- 电信 (网络)

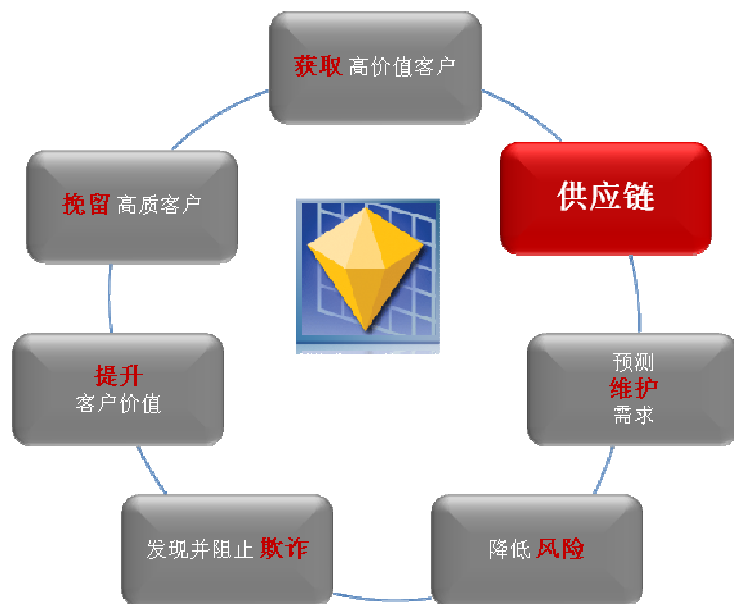
KOMATSU®



建立监控系统来识别和发现质量问题及危险信号，并且设定合理的技术方案

- 在问题出现前，进行预警并提供最佳方案。在最初的2周生产中，就节省维修费用达**\$1 million**

价值体现:供应链



针对各个商户、商店群体，精细
确立产品计划和分配

要点:

- 建立预测模型，实现对产品门类和产品的精确需求预测
 - 管理大量的模型，自动化流程

典型应用行业:

- 零售



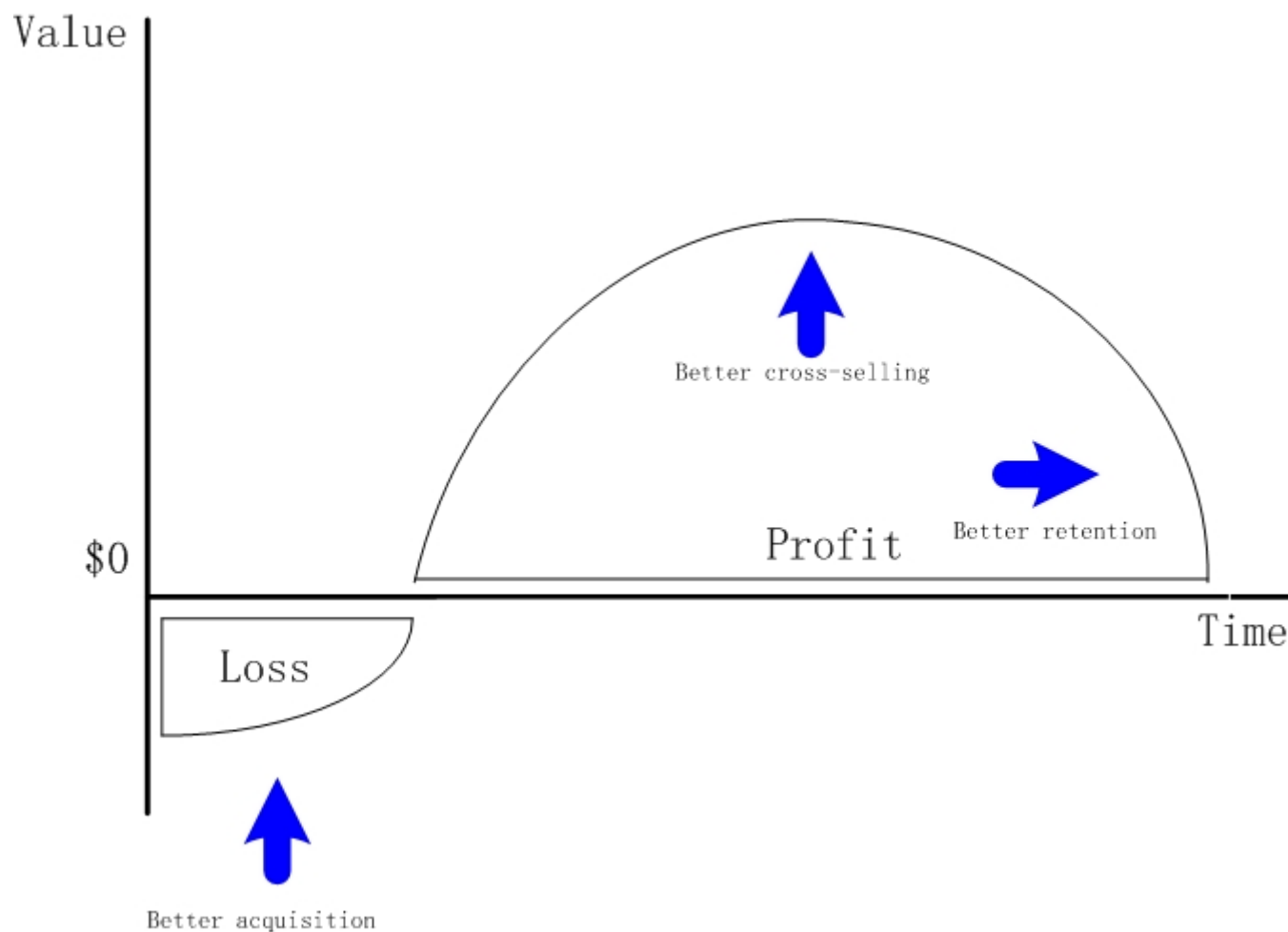
运用大量的预测模型，来优化产品配置

- 3200店铺，需要对每个店铺每种产品进行需求预测，达到准确的产品配置设计
- 市场购物篮分析通过一宗捆绑销售就实现 \$20M
- 在商店、重点商品上实现精确预测
- 在最初的90天中，新的资源配置方案实现了节省\$50M+ 库存和提升\$100M+ 销量

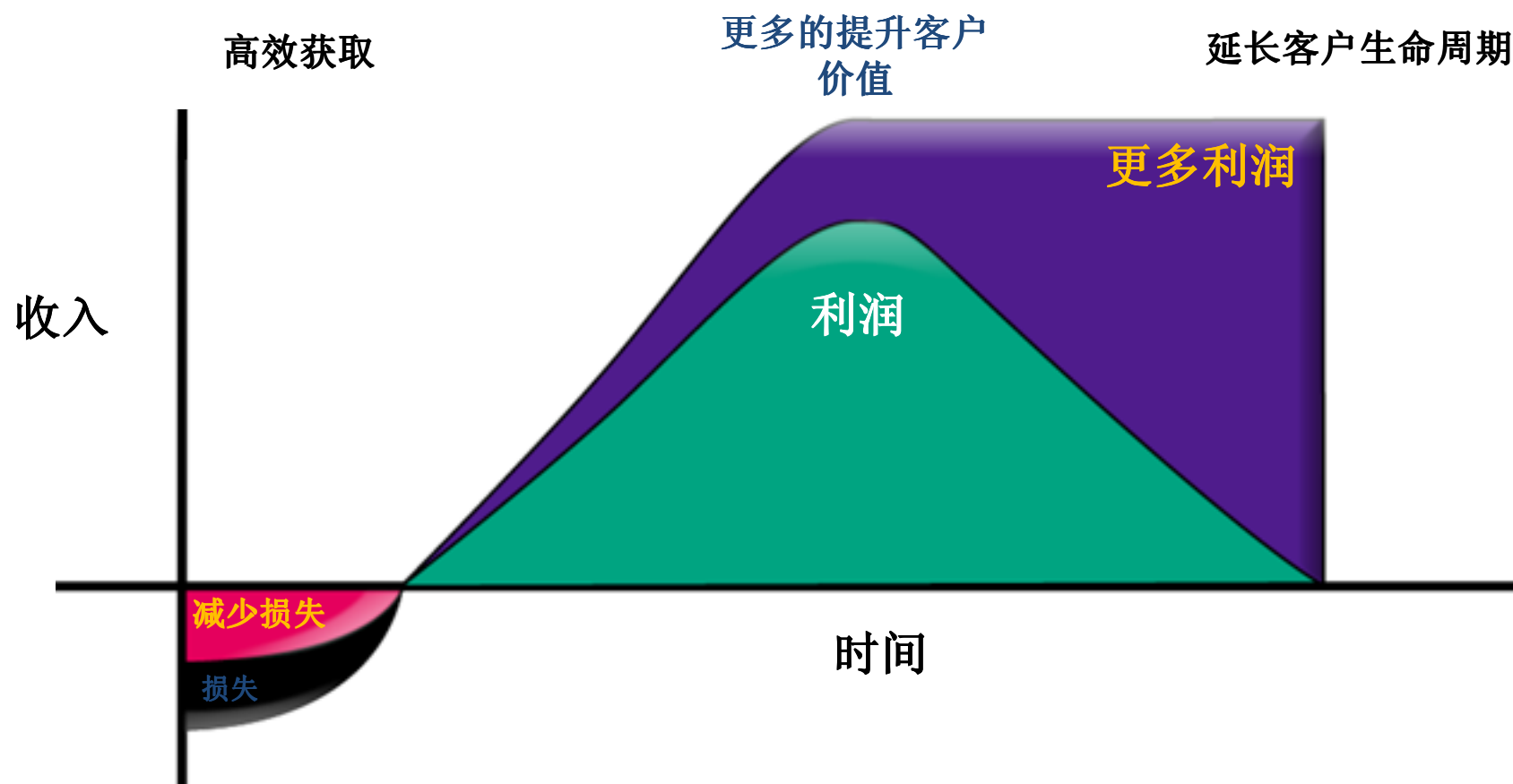
SPSS应用领域



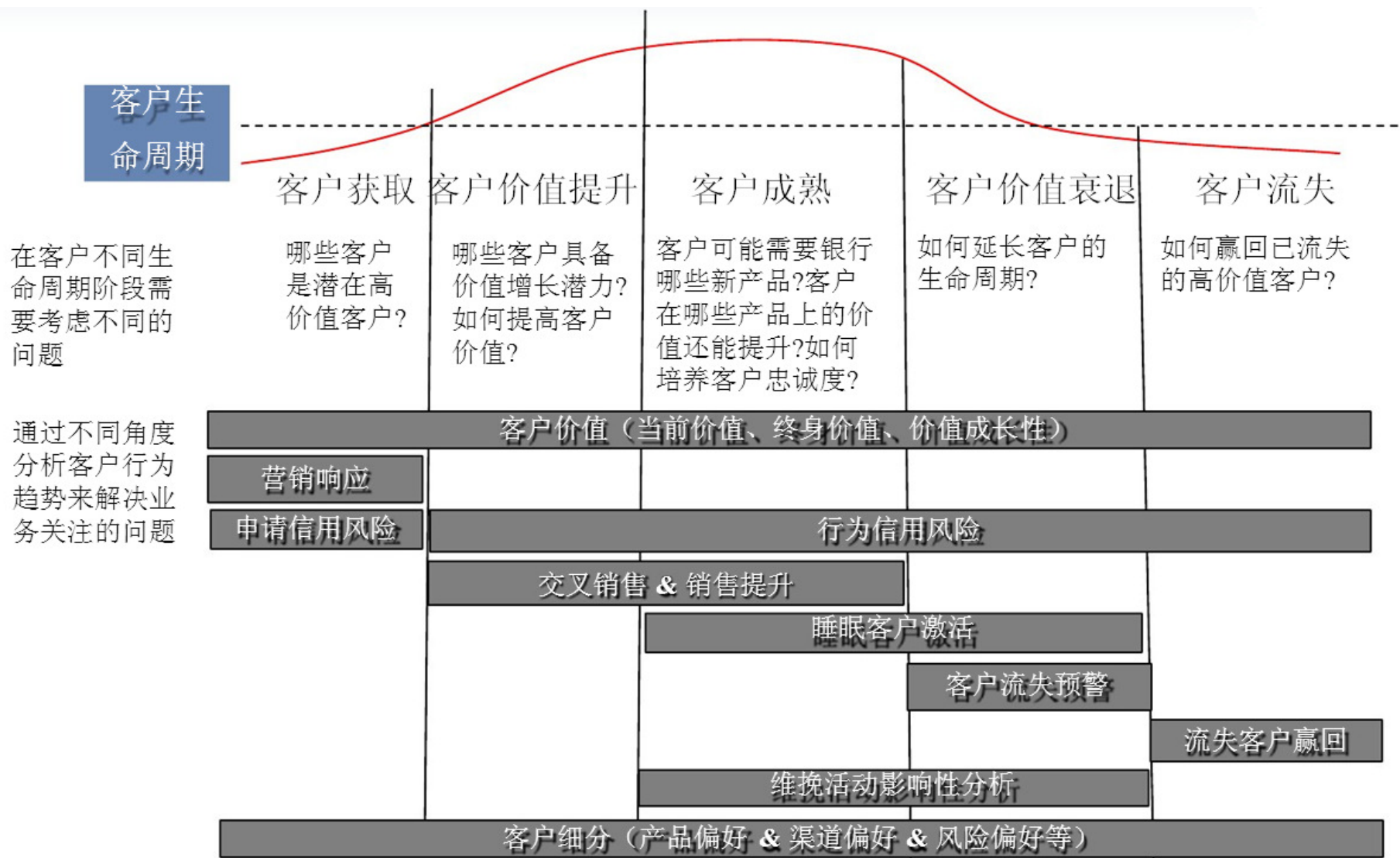
客户生命周期模型和客户关系管理



SPSS应用价值



基于银行客户生命周期设计数据挖掘应用



主要解决方案



银行/保险

客户细分: 针对不同客户提供不同的服务
客户流失: 挽留高价值客户
交叉销售: 提高收入和客户的收益率
客户满意度分析: 提高客户忠诚度
欺诈监测: 分析欺诈模式, 识别欺诈行为
信用风险管理: 计量和规避信用风险
降低索赔: 识别出具有较低索赔可能的客户.....



通讯

客户细分、客户流失、营销响应、满意度分析
销售提升: 预测特定消费群的行为特征, 设定有效的销售战略
客户市场推广: 模拟计费 and 模拟出账
客户欠费分析和动态防欺诈: 欠费行为的事前预测
呼叫中心: 话务量预测, 降低运营成本.....



政府及公共事业

辅助税收管理: 发现未申报的潜在税务责任人、税种; 识别容易偷漏税的纳税人、针对不同的征收案例制订正确的征收策略;
犯罪行为分析: 分析案件关联、串并案; 对高危人群及地区犯罪进行预警.....



教育/零售

教学评估: 成绩评测、教学方法有效率
门店选址: 找出最佳的新店铺位置
库存管理: 满足客户需求, 降低库存成本
配置、物流管理



制造业

产品设计: 在尽可能少的试验次数下确定影响质量的因素并选择合适的生产水平;
失效、故障预测
质量控制: 实现品管七大手法.....

使用CRISP-DM方法论进行数据挖掘

商业理解 (文档)

Health Churn Demo

Setup

First need to add an ODBC driver to connect to the Access database containing the data.

NT: Control Panel - Data Sources (ODBC)

2000: Control Panel - Administrative Tools - Data Sources (ODBC)

In the "ODBC Source Administrator" window,

- click "System DSN" tab

- click "Add" button

- select "Microsoft Access Driver (*.mdb)"

In the "ODBC Microsoft Access Setup" window:

- click Database "Select" button

- in the browser select the "data/health demo.mdb" file. This can be found in the directory

- where the demo was placed

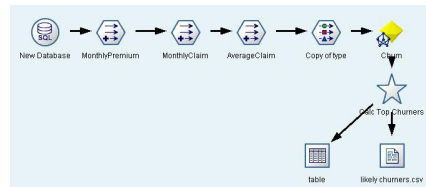
- In "Data Source Name" enter "Health Demo"

- Click "OK"

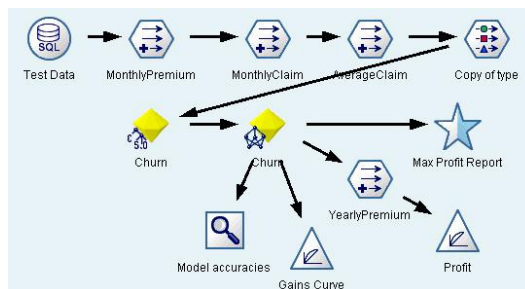
That's it. An ODBC driver called "Health Demo" should have appeared.

Check the streams load and execute. You may need to configure the ODBC input nodes in the Clementine streams.

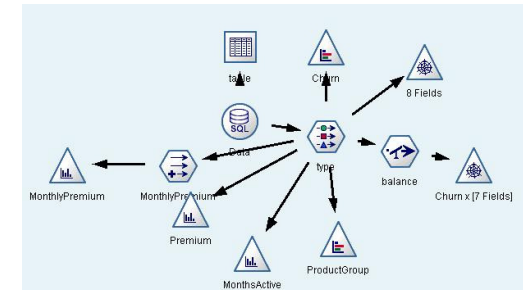
结果发布 (数据流)



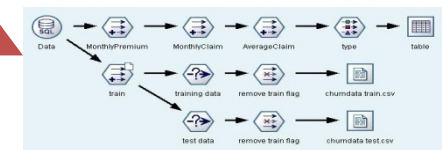
模型评估 (数据流)



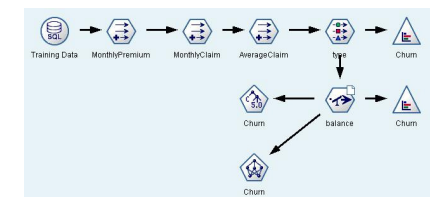
数据理解 (数据流)



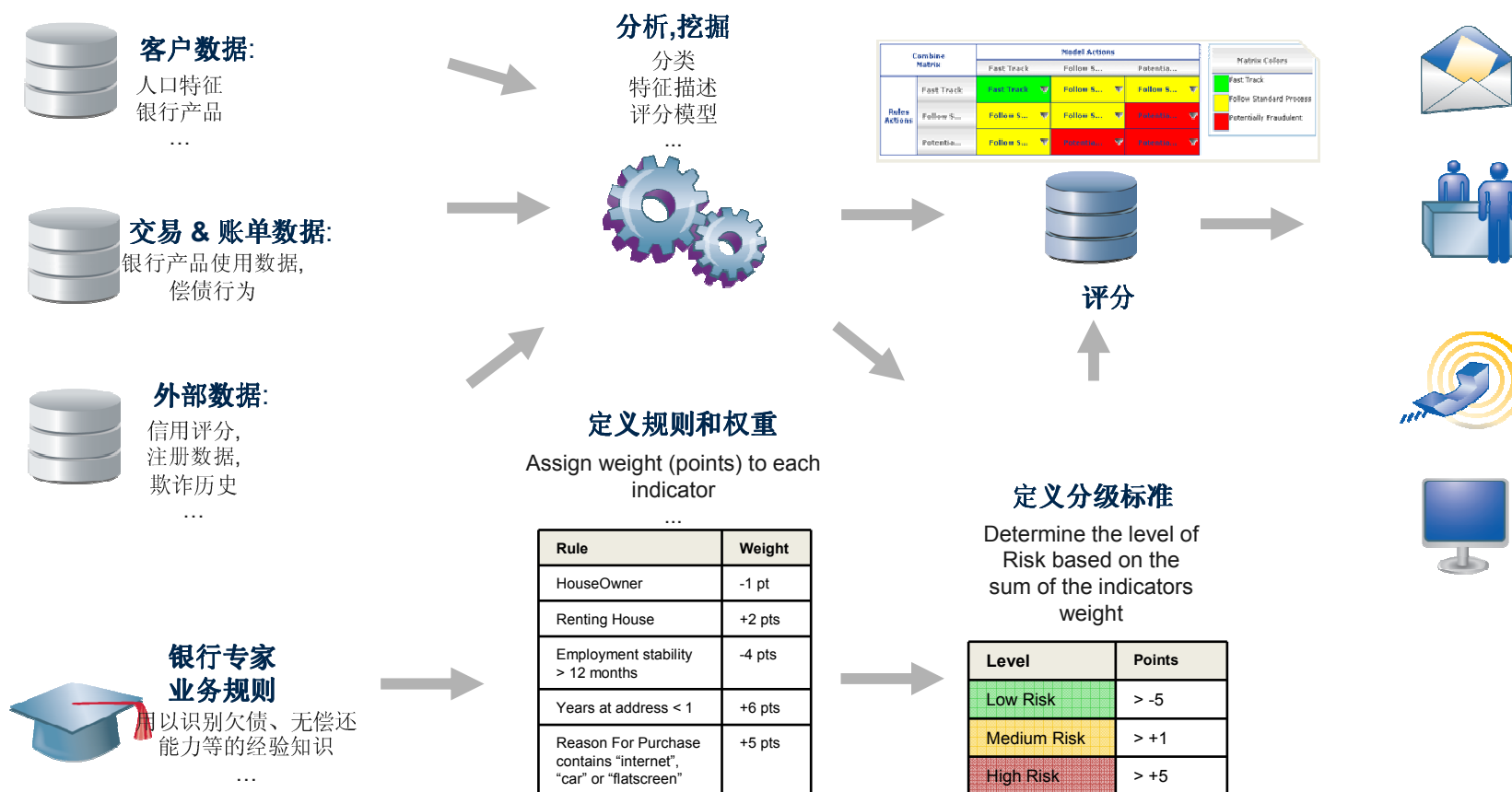
数据准备 (数据流)



建立模型 (数据流)



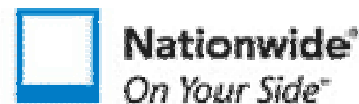
分析流程示例



SPSS客户——银行业



SPSS客户——保险业



SPSS客户——电信业



SPSS客户——零售

Wegmans
helping you make great meals easy

 **Apoteket**

OfficeMax

vision express
What would you like to see?

**Advance/
Auto Parts**
the best part is our people

DEBENHAMS

WAL★MART®
ALWAYS LOW PRICES. *Always*

Office DEPOT.

expert 

 **OHMV**.co.uk

halfords

ICA

 **Nestlé**
Good Food, Good Life

TESCO
Every little helps

B&Q

 **Boots**


PUMA®

**BEST
BUY**

Thousands of Possibilities |
GET YOURS

TOYS R US
www.toysrus.co.uk

 **STARBUCKS
COFFEE**

CJL1

ellos

MARY KAY®
*America's
Best-Selling Brand

@nifty

Sofmap.com へようこそ!

幻灯片 47

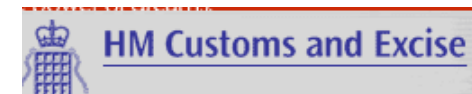
CJL1

Add Argos and other retailer logos
Colin Linsky, 2010-5-22

SPSS客户——能源、学术、政府



SAINT LOUIS UNIVERSITY



STANFORD UNIVERSITY



SPSS客户——制造

nationalgrid



Sikorsky

A United Technologies Company



KOMATSU



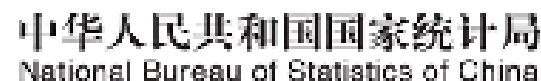
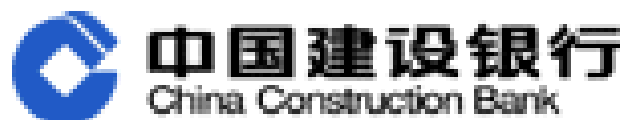
anglianwater



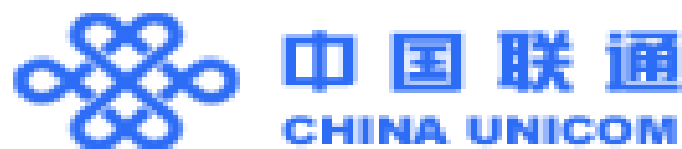
MOTOROLA



中国客户

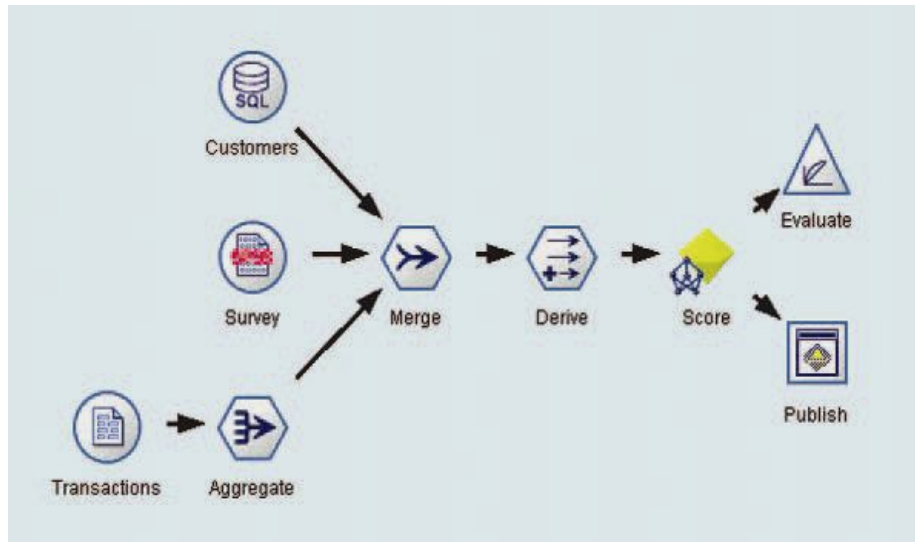


信息产业部
劳动保障部
国家信息中心
北京疾病控制中心
中国人口与计生委
北京经济信息中心
中国人口情报信息中心
巴中市卫生局
江西统计局
北京大学
北京科技大学。。。



SPSS产品优势

①简单直观的图形化数据挖掘



②丰富的数据挖掘算法

预测 & 分类

Neural net, C5.0, C&R Tree, QUEST, CHAID, Linear Regression, Logistic & Multinomial Logistic Regression

关联发现

Apriori, GRI, Carma, Web graph

聚类 & 分群

TwoStep, Kohonen, K-Means

数据降维

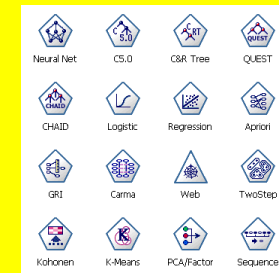
Factor, Principle Components

序列分析

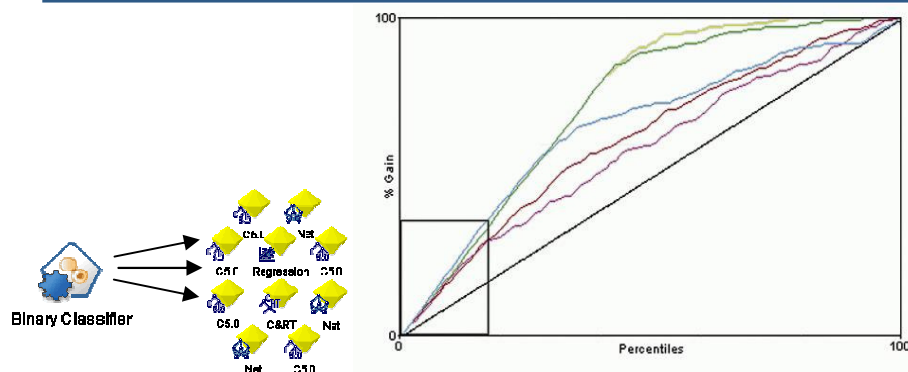
Capri, Neural Net, Regression

时间序列

Time Series Forecasting



③快速自动发现最优模型



④标准流程、整合性好

