Applied Analytical Data Science

Teil 13: Data Mining Anwendungen

Dr. Jörg-Uwe Kietz,
Vorlesung an der Univ. Zürich,

Mittwoch, 14:00-15:45 Uhr Vorlesung,
16:00-17:30 Uhr Übung

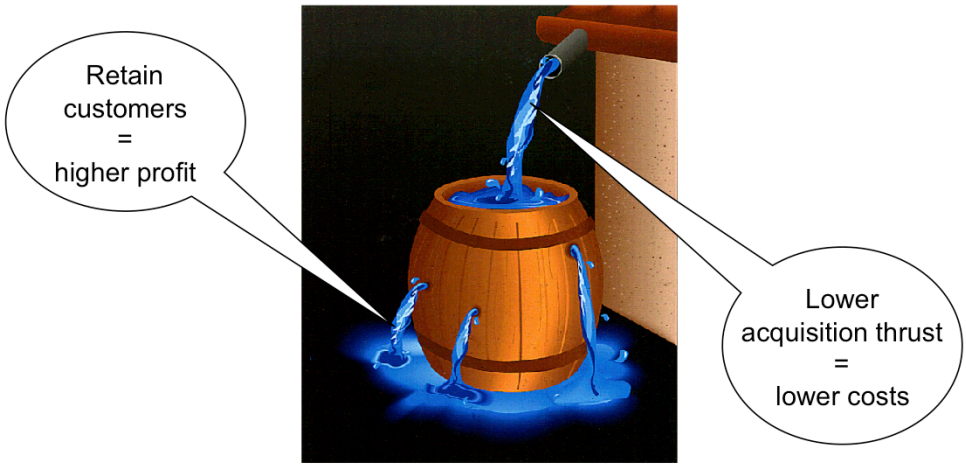
http://www.kietz.ch/AADS/

Content

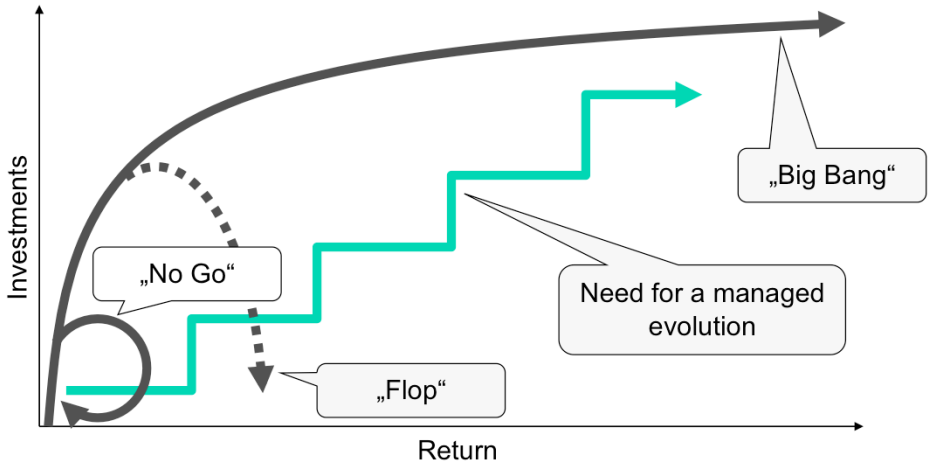
Data Mining Applications:

- Knowledge Discovery Services (KDS)
  for Customer Relationship Management (CRM):
  - Acquire the rigth Cutomers,
  - make them profitable, and
  - keep them
- Knowledge Discovery Application (KDA)
  Integrate Data Mining into an application
  - Detect Money Laundering
- Summary

A basic understanding of CRM

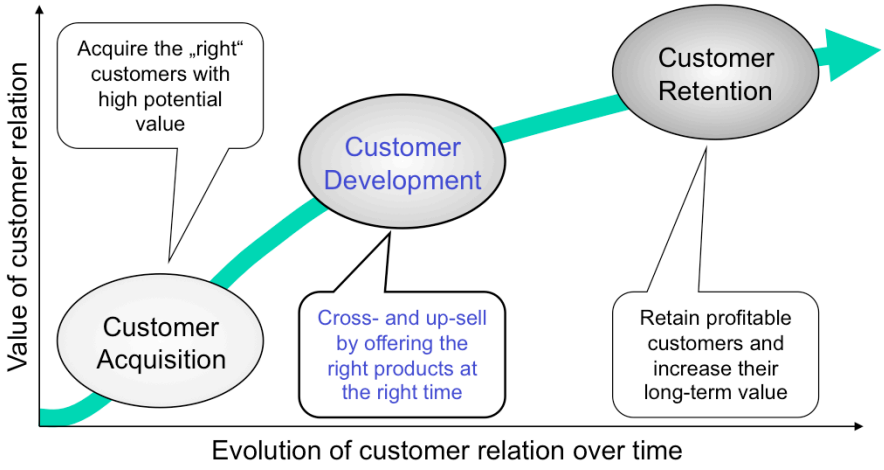


Doing KDS for CRM



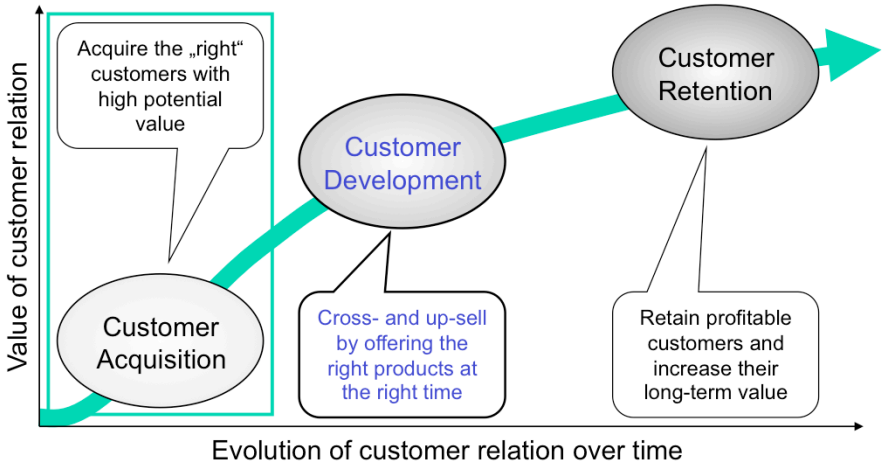
KD for CRM

Three simple business goals of CRM



KD for CRM

Three simple business goals of CRM



Data Mining for Customer Acquisition

Use Data Mining to:

- Predict who would like to be a customer
  - Example: Mailing Action
  - Main problem: Non-customer data are needed

- Predict who could be a **good** customer
  - Example: Credit Risk Scoring
  - Main problem: Online integration into the sales-process

Mailing Action for Customer Acquisition

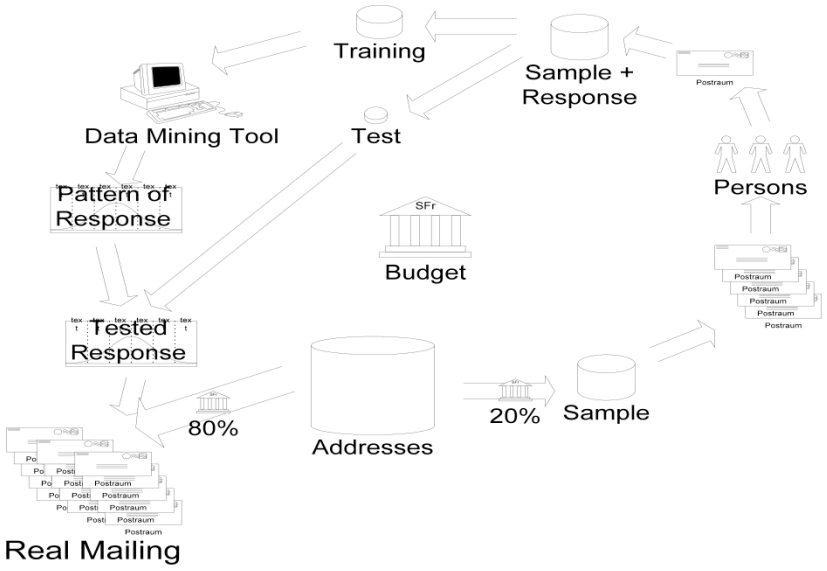
The goal of the mailing actions is

- to create a distribution channel for insurance that is not intense in terms of consultancy,
- to gain new clients to which the insurance consultant has no access and
- to allow for more inexpensive contract conclusions compared to those of insurance consultants.

The process is successful

- if new clients are gained and
- average customer acquisition costs of the mailing action are lower than of insurance consultants

The Data Mining Process



Data Mining Target

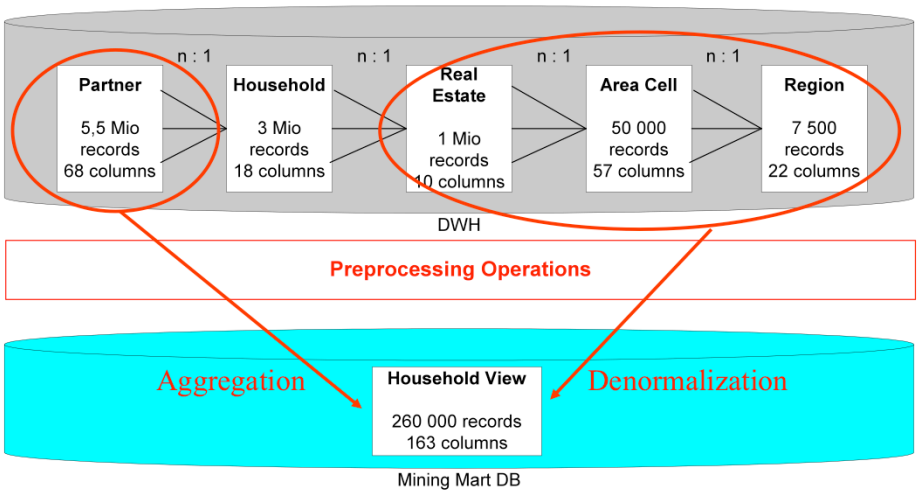
Determination of a group of persons in address data base

- based on a sample mailing and
- by using the given attributes

which is most likely to buy a contract.

Construct **household-view of all Swiss households**
is a first step for a direct-marketing data mart.

Data used to analyse the mailing campaign



Why one Household-view?

- most DM-Tools accept only one input-table.
- all relevant information in one table
- could be considered as a Data Mart for the Business Task „Direct Marketing“

DWH is no Data Mart!

DWH is integrated, normalized view on data from several operative systems

- DWH is problem-independent
- DWH is usable for several different data marts
- data from DWH cannot used directly for mining: attributes are not in a mining-relevant form, e.g data of birth is stored instead of age of the person when taking out a police.

Results

The integrated use of **Data Mining** in direct marketing **increases profitability**:

- Study showed **double response-rate**
  - using DWH data, dm-plus (BWV) data and responses from 260.000 direct mails in 1998-99

⇒ same result for less money, or

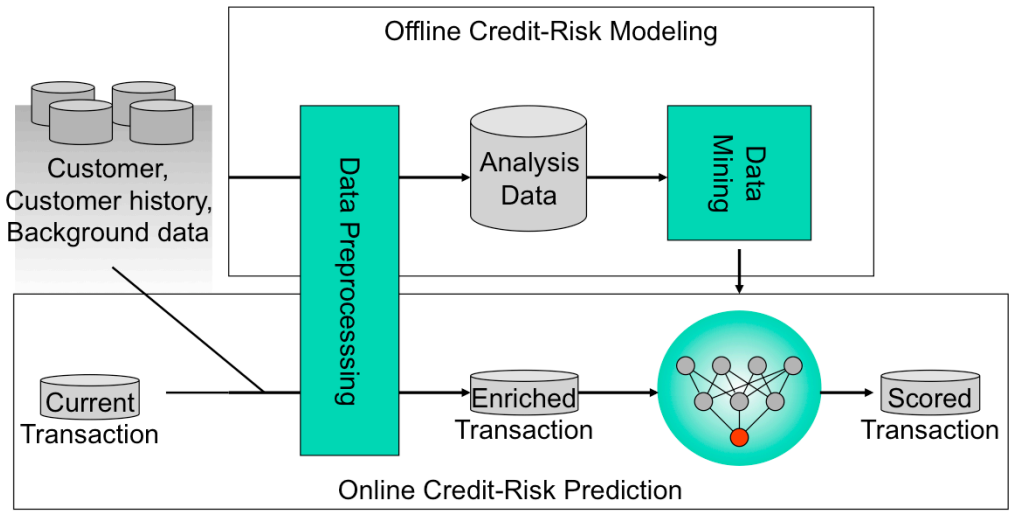
⇒ more results for same money

(campaign from former CH/VSLD)

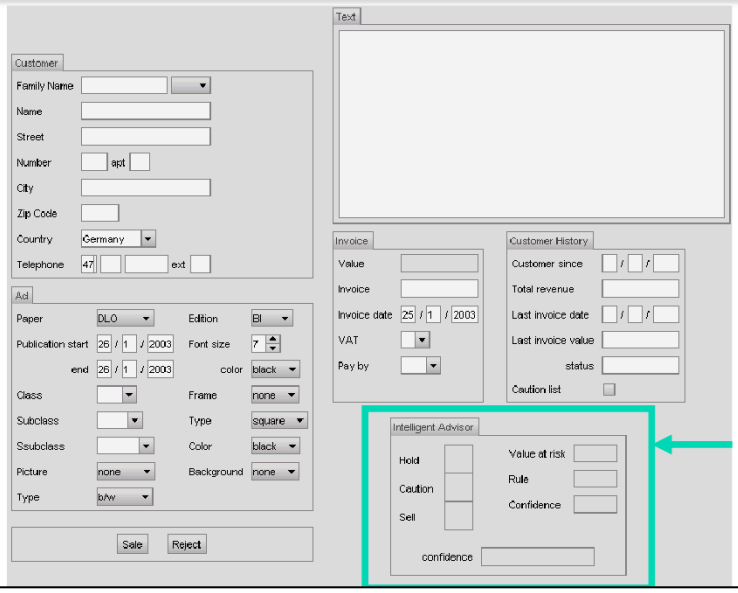
Credit-Risk Scoring

- Not only banks have a credit-risk, everyone who accepts to deliver services or goods before he receives payment has one.
- Models predicting credit-risk can be build from previous cases.
- These models have to be applied on-line at the point-of-sales
- They can be build off-line, but should be rebuild periodically
- kdlabs build a predictive model for an advertisement supplier predicting non-payers with accuracy of 79% on the test set
- This model is integrated into the customers point-of-sales application.

How does it work



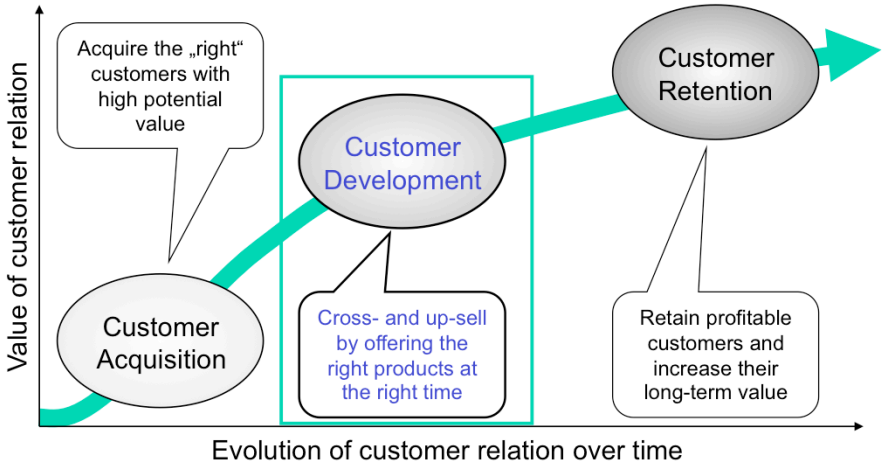
Point-of-Sales credit-risk-scoring



The model acquired by KD is used to provide immediate feedback into the sales process

KD for CRM

Three simple business goals of CRM



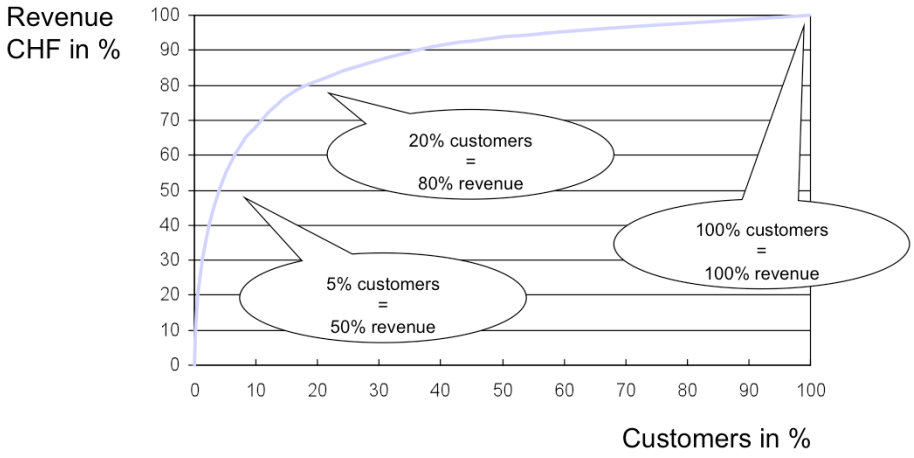
Data Mining for Customer Development

Use Data Mining to:

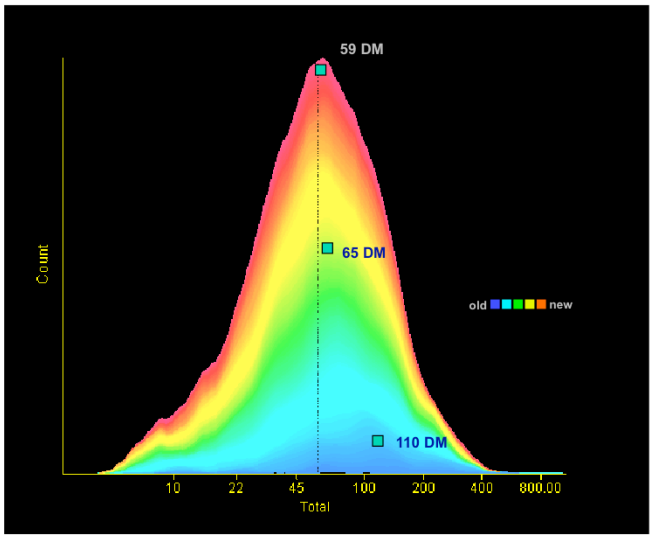
- Segment Customers
  - Channel Assignment
  - Targeted Actions
- Predict Cross-Selling affinities
  - Example: Mailing Action for frozen food
  - Example: Mailing Action for life insurances

Customer profitability

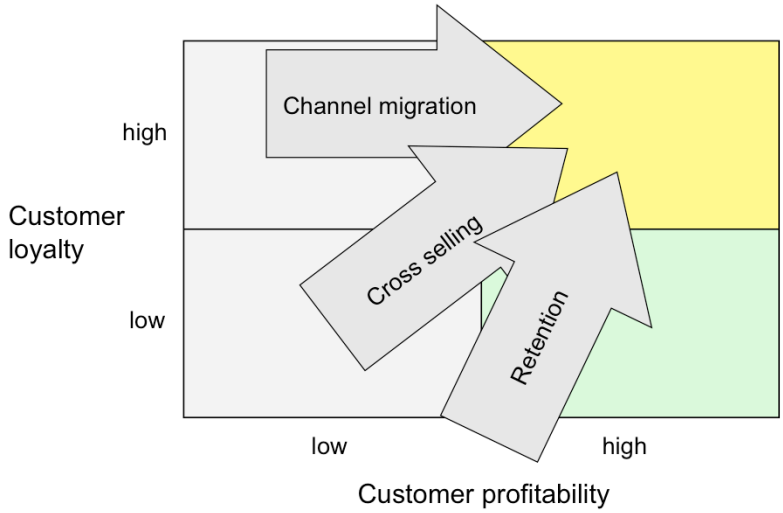
Basic profitability analysis



Customer profitability and loyalty

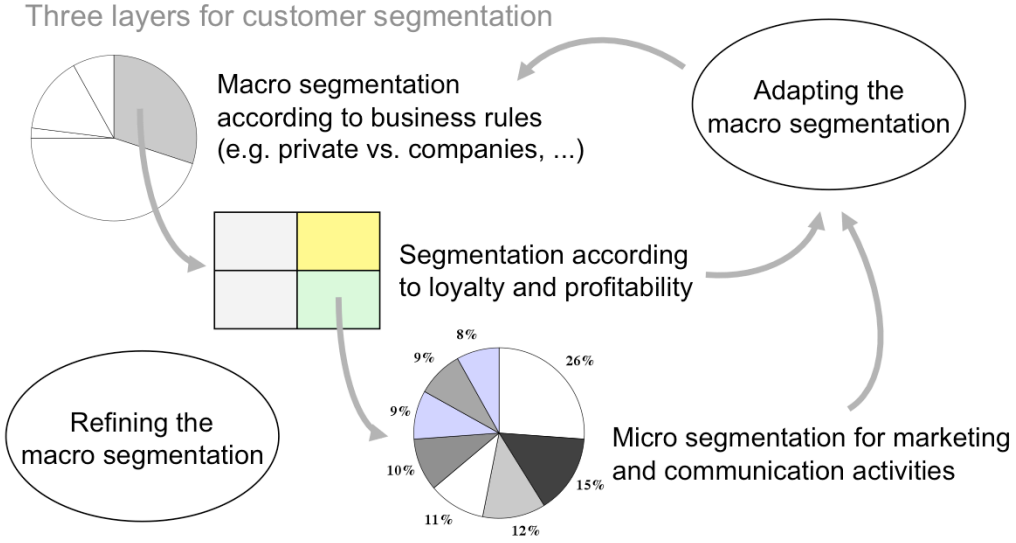


Customer segmentation



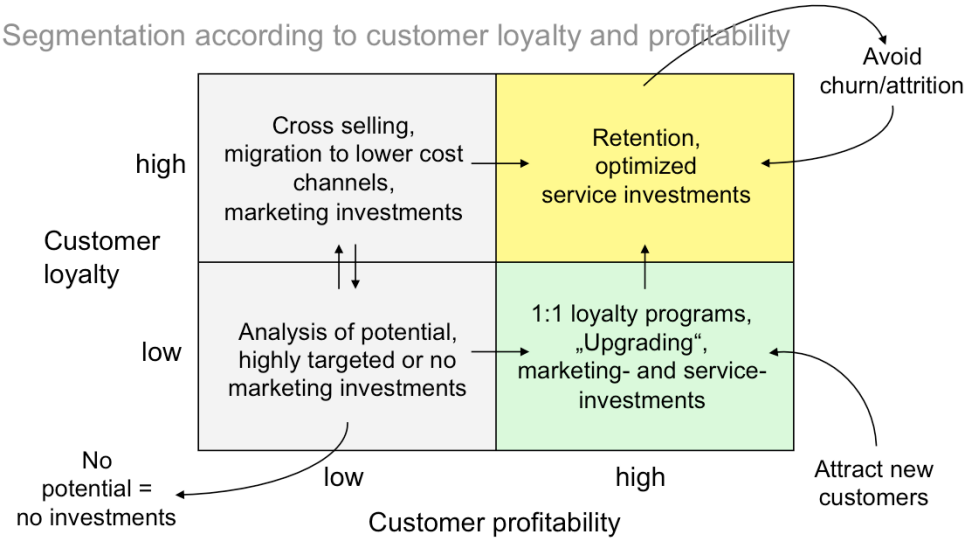
Customer segmentation

Three layers for customer segmentation



Customer segmentation

Segmentation according to customer loyalty and profitability



The business context and goals

Background

- privately held company
- frozen foods direct sales company
- about 4 mio customers of which about 3 mio are *active*
- about 3000 truckers/salesmen
- about 30% of truckers know their customers

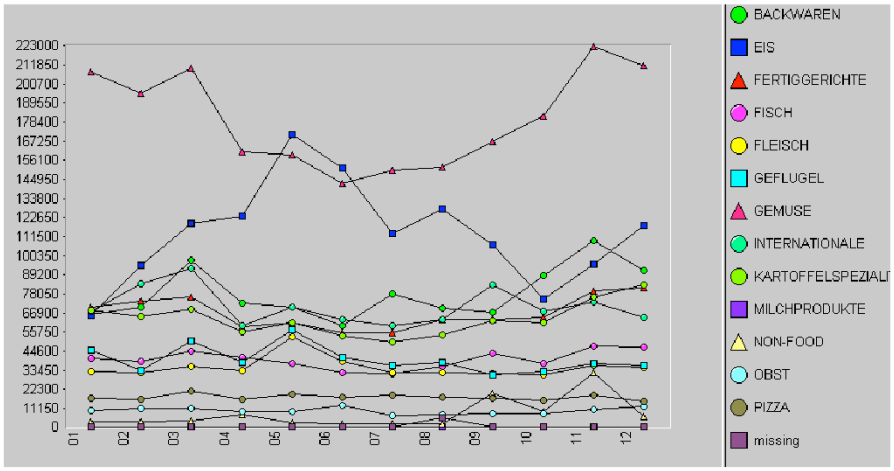
Basic marketing approach

- door-to-door customer acquisition
- catalog advertising
- monthly postcard
- driver *double punch*
- discounted products

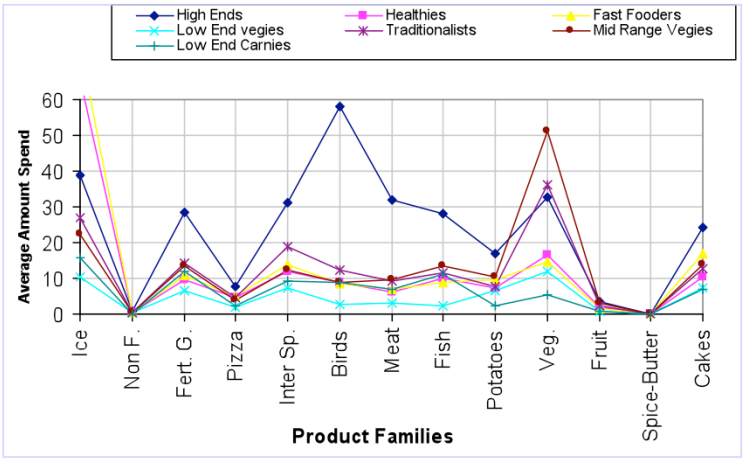
Business Goal for KD

- Improve the efficiency of the marketing methods
- Improve the customer-profitability by
  - Sell more per customer
  - Sell higher end products
- Improve the customer relation by
  - Propose customers the goods they want

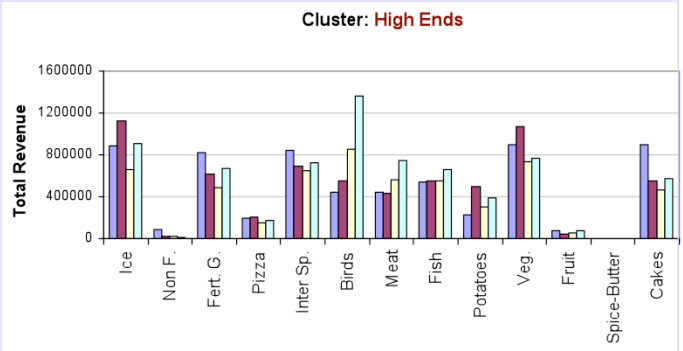
Prepare and Analyze the Data



Buying Behavious of clusters

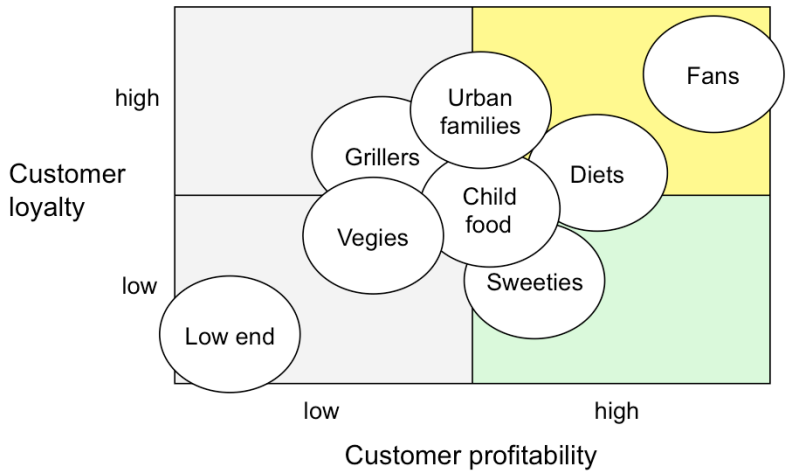


Characterize natural clusters



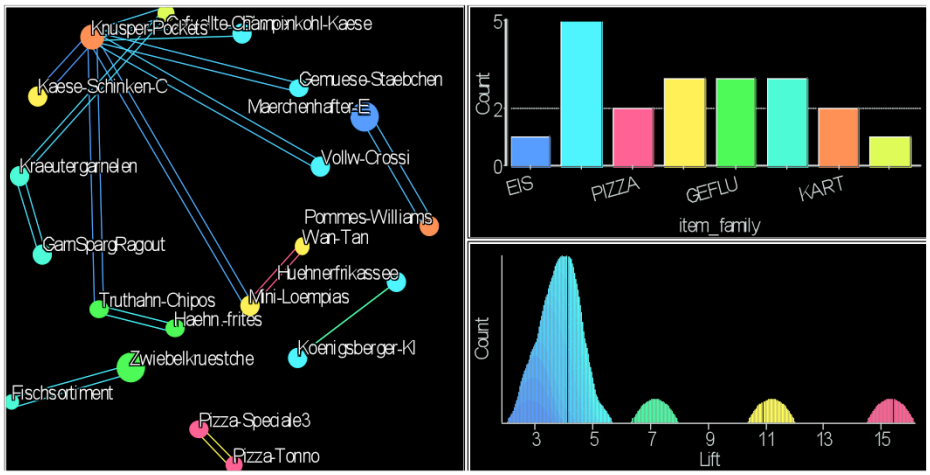
Customer segmentation

Micro segmentation in food retailing build with clustering



Segmented cross-selling

Market basket analysis: Association rules per micro segment



Choose best cluster affinity rules

Top10 2-prod inc_in_revenue

left_prod	left_fam	right_prod	right_fam	right_rev_incr	right_revenue
Rezeptkalen2000	NON-FOOD	Adventskalender	BACKWAREN	45,940.95	212,973.20
Rahmspinat	GEMUSE	Fixfertig-Apfel	GEMUSE	29,694.72	273,354.40
Valpolicella-0	NON-FOOD	Kaese-Piccos	INTERNATIONALE	29,523.39	210,024.80
Fixfertig-Apfel	GEMUSE	Premium-Auslese	BACKWAREN	29,270.91	473,281.70
Blumenkohl-12	GEMUSE	Broccoli-Roesche	GEMUSE	25,062.46	191,896.95
Creme-Vanille-B	EIS	Schokosplitter-	EIS	23,489.16	96,745.70
Broccoli-Roesche	GEMUSE	Blumenkohl-12	GEMUSE	23,272.12	139,419.15
Reibekuchen	KARTOFFELSPEZIA	Premium-Auslese	BACKWAREN	22,730.10	473,281.70
Rezeptkalen2000	NON-FOOD	Buttermuerbeteig	BACKWAREN	22,276.10	216,482.10
Rahmspinat	GEMUSE	Fischstaebchen	FISCH	22,058.16	157,770.20

Top 10 2-prod conf

left_prod	left_fam	right_prod	right_fam	right_rev_incr	right_revenue
Pinot-Grigio	NON-FOOD	Valpolicella-0	NON-FOOD	1,911.66	152,754.40
Putenrollbraten	FERTIGGERICHTE	Schweineschnitz	FERTIGGERICHTE	2,166.43	13,536.00
Haehnchenkeule-	FERTIGGERICHTE	Schweineschnitz	FERTIGGERICHTE	2,365.09	13,536.00
gefuelltes-Fisch	FERTIGGERICHTE	Schweineschnitz	FERTIGGERICHTE	2,261.59	13,536.00
Schnittlauch	GEMUSE	Petersilie	GEMUSE	6,594.40	48,005.10
Putenrollbraten	FERTIGGERICHTE	Haehnchenkeule-	FERTIGGERICHTE	2,120.54	9,480.00
Chardonnay-Wein	NON-FOOD	Valpolicella-0	NON-FOOD	435.02	152,754.40
Rindersaftgulas	FERTIGGERICHTE	Schweineschnitz	FERTIGGERICHTE	2,812.73	13,536.00
Putenrollbraten	FERTIGGERICHTE	Rindersaftgulas	FERTIGGERICHTE	2,363.34	13,113.00
Haehnchenkeule-	FERTIGGERICHTE	Putenrollbraten	FERTIGGERICHTE	2,240.47	8,720.00

Build models for chosen products

```
IF    Eierstich > 0  AND
      cluster not in { 5 7 }  AND
      Schwarzwaelder-K <= 2  AND
      Zwiebelwuerfel <= 4  AND
      Haehn.-frites = 0  AND
      EIS = 0  AND
      Maultaschen <= 1  AND
      Haehn.-Brustf.na <= 1  AND
      Piadina <= 1  AND
      Backofenkrokett <= 1  AND
      Creme-Berliner <= 1  AND
      Butterpfannenge = 0  AND
      Eierspaetzle = 0  AND
      OBST <= 1
 THEN  Markkloesschen = 1
```

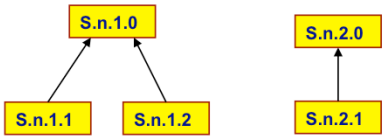
Evaluate models

Rule	Lift	Targets	Non-Targets	Target confidence	Prospects	Add_Rev
16	2.9881	740	10942	0.6495	7107	56,142.65
88	1.9842	272	7997	0.5431	4343	34,311.24
164	1.8391	298	8004	0.5234	4189	33,095.67
55	3.7647	372	5754	0.6805	3915	30,930.99
53	2.3214	349	6663	0.5850	3898	30,792.00
120	2.7171	236	4393	0.6461	2838	22,421.56
119	1.7952	203	5002	0.5192	2597	20,517.83
252	1.7269	151	5144	0.5035	2590	20,460.91
91	4.9226	371	2960	0.7816	2313	18,275.84
70	2.5472	147	3681	0.6071	2235	17,655.66
104	2.2473	96	2512	0.5833	1465	11,576.13
226	1.9348	66	2435	0.5263	1282	10,124.48
134	2.1971	98	2282	0.5595	1277	10,086.99
165	6.1468	518	1239	0.9432	1169	9,232.44
268	2.1517	53	1883	0.5536	1042	8,234.76
137	5.4010	442	961	0.9913	953	7,525.59
214	1.7813	76	1811	0.5075	919	7,260.22
178	2.1242	68	1620	0.5490	889	7,026.36
5	5.7922	275	990	0.8154	807	6,377.13

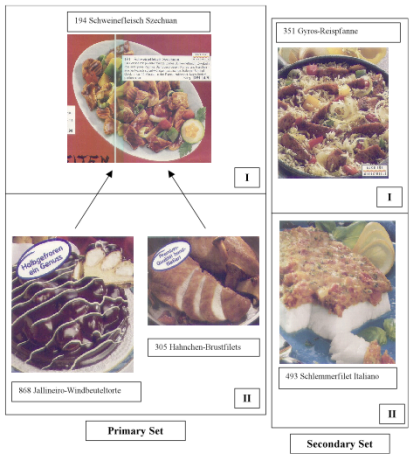
Produce message and lists

targeting

- choose the *best* two products (by predictability) in each cluster
- choose the *best* single and double associators



example message



Results

increase in purchased items (hit rate per buyer)

- of the top 10 winners, 8 were our products
- of the top 15 winners, 12 were our products

total increase in revenue per customer

- 7 of our products are in the top 10 winners
- 12 of our products are in the top 15 winners

revenue increase per purchase

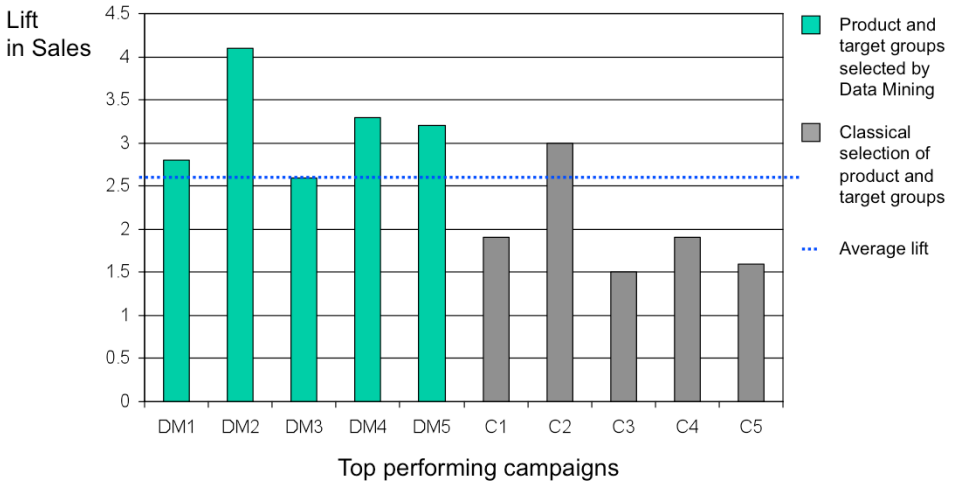
- the average hitrate per buyer (from advertised products) in clusters a, b, and c was, on average, about twice that in the control group x

most spectacular winners

- some of the most dramatic increases in purchase rates happened among low-buying customers (cluster c). In this cluster, for example, sales of one product increased by a factor of 4.15

Segmented cross-selling Results

Targeted vs. classical campaigns

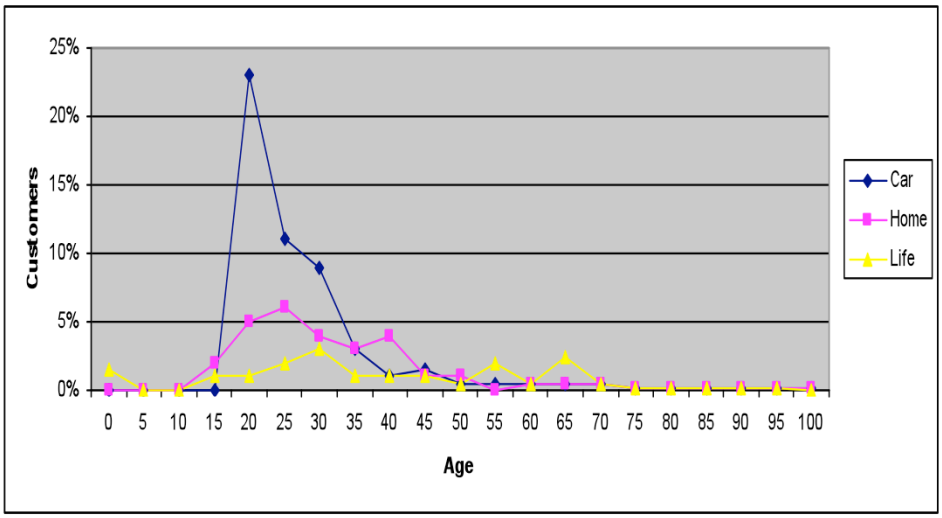


Customer Life-cycle Analysis

- The Business Problem:
  - Life insurances are profitable products of insurance-companies
  - Only a small percentage of all customers own one

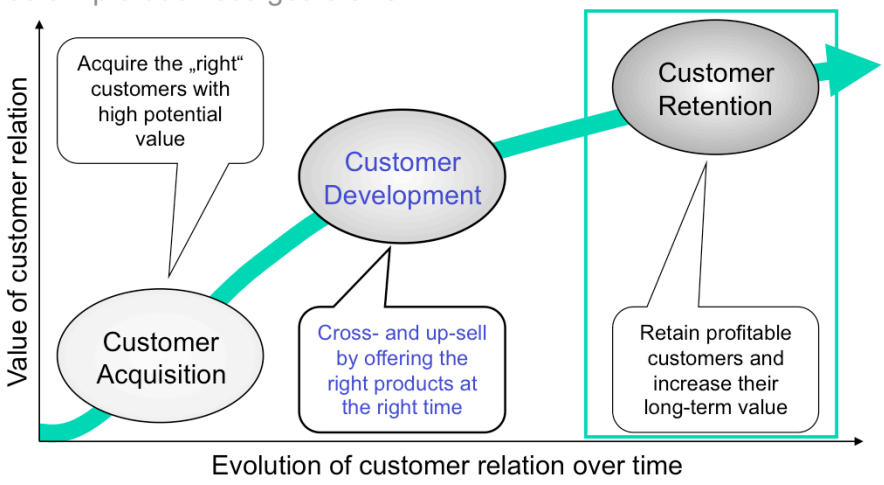
  ⇒ Cross-sell life-insurances to customers
- The Approach:
  - Transform the "state-description"
    Customer C (a ….) owns product P (with …) since X
  - into a life-cycle description
    A Customer C started with P1 at X1, X2 years later she bought P2, …
  - "Data Mine" typical states of customers when they bought a life-insurance
  - Rate customers based on their similarity with such prototypical states
  - Advertise a life insurance to the N highest rated customers
- Resulted in one of the most successful campaigns

Buying Age Distribution of Insurances



KD for CRM

Three simple business goals of CRM



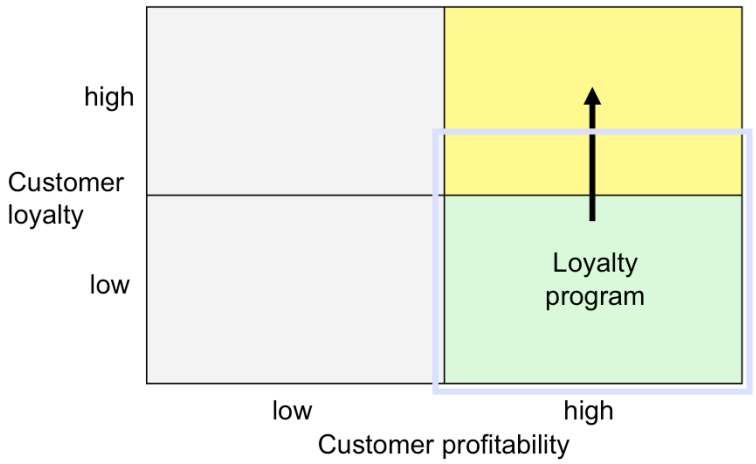
Data Mining for Customer Retention

Use Data Mining to:

- Predict affinity to a loyality program
  - Example: Targeted Action
- Analyse Customer Satisfaction
  - Example: Causal Modelling for Market Research
- Predict Churn
  - Example: Life Insurance Surrender

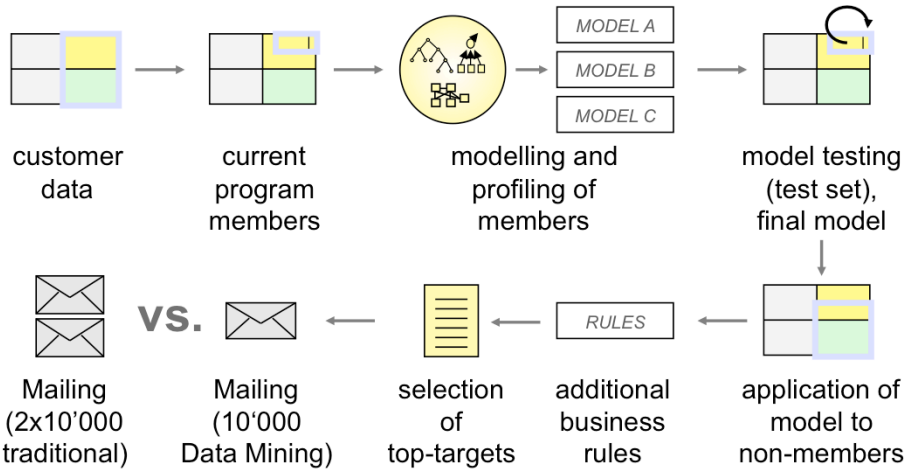
Targeted marketing campaigns

Launching a loyalty program for customer retention



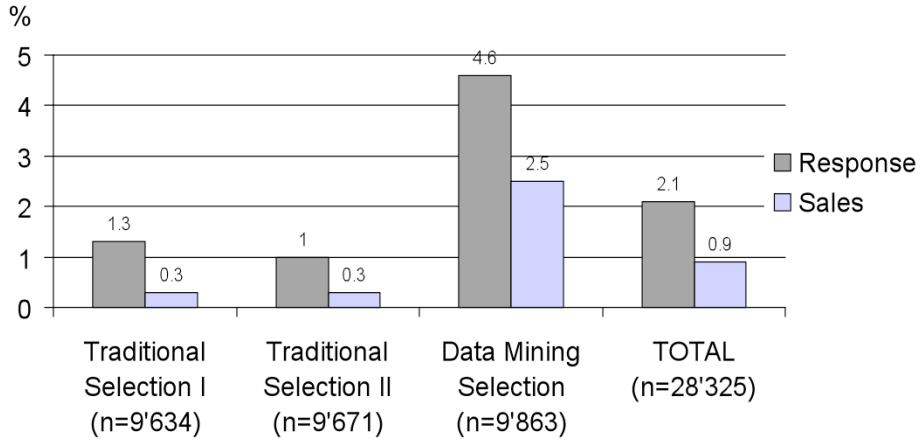
Targeted marketing campaigns

Process of KD-driven customer selection



Targeted marketing campaigns

Mailing campaign for a loyalty program



Customer & Employee Satisfaction

Business Goals

- Employee and customer satisfaction are prerequisites for loyal behavior and customer relation, and by that influence profitability.
- Only detailed knowledge about the influence-factors of customer and employee satisfaction enable effective actions to increase satisfaction, loyalty and therefore profitability.

Therefore:

- New Management Methods like Total Quality Management (TQM) or Balanced Scorecard (BSC) contain the assessment of customer and employee satisfaction.

Causal Modelling for Marketing Research

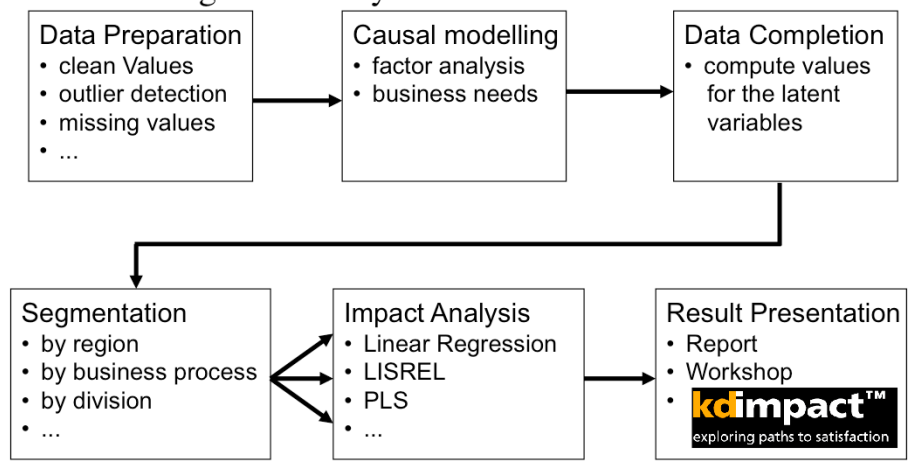
- Marketing Research starts with a questionnaire
- Results are analysed to build a causal model of
  - Customer satisfaction
  - Branding acceptance
  - Employee satisfaction
  - ….
- to determine the influence factors and their impacts
- Needed
  - to steer marketing actions,
  - to control their success, and
  - to report them to public (Key Performance Indicators)

Causal Modelling for Marketing Research

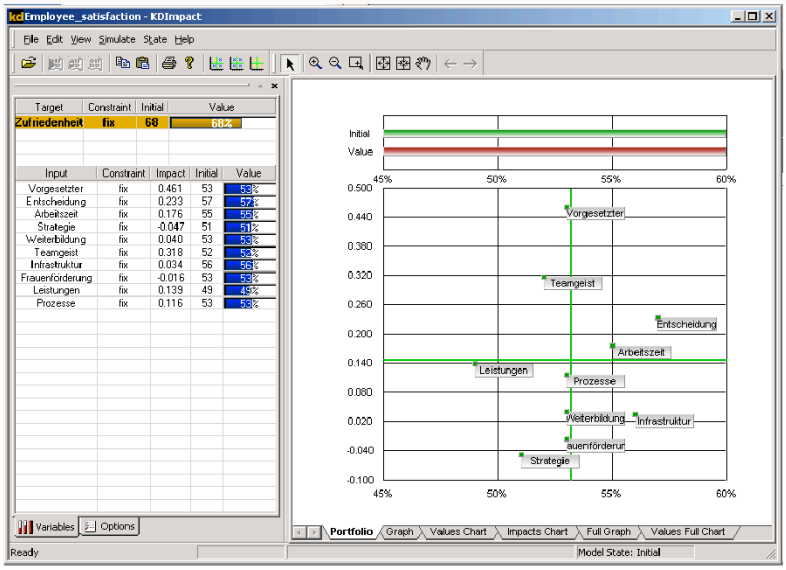
- kdlabs did causal modelling for several customers
  - Customer Satisfaction
    - Gastronomy group (repeated)
    - Insurance company (repeated)
    - Public transport
    - Large Bank
  - Branding acceptance
    - Soft drink company
  - Employee Satisfaction
    - Large Bank
    - University
- Causal modelling product:
  - kdimpact

Causal Modelling for Marketing Research

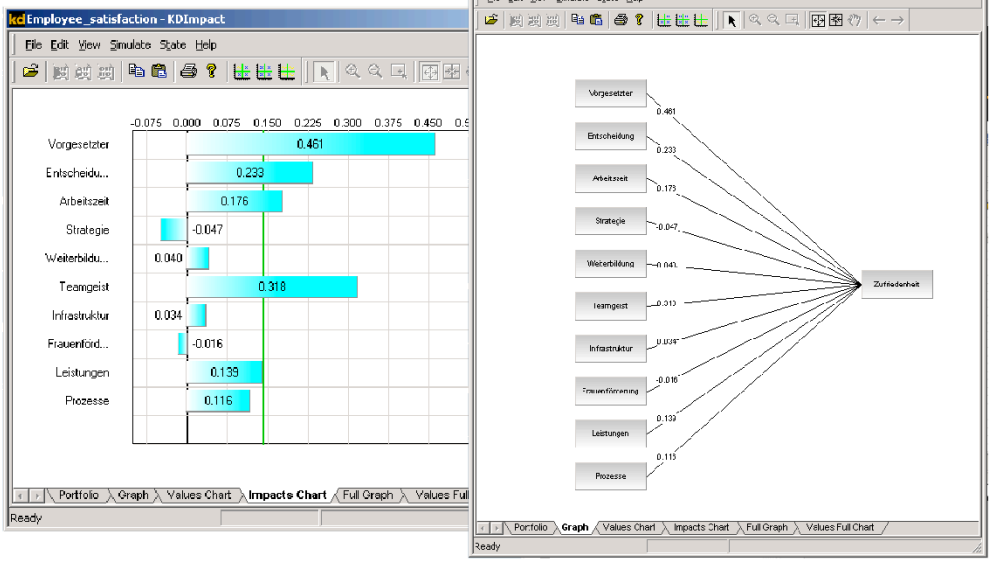
The Knowledge Discovery Process



Causal Modelling for Marketing Research



Causal Modelling for Marketing Research

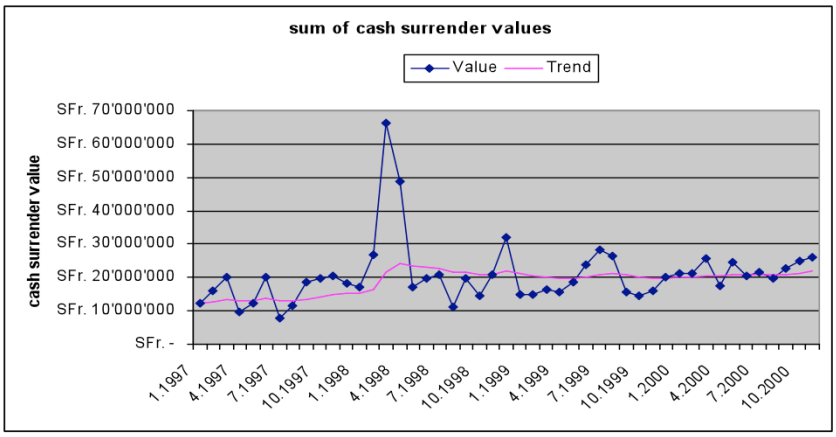


Customer Retention

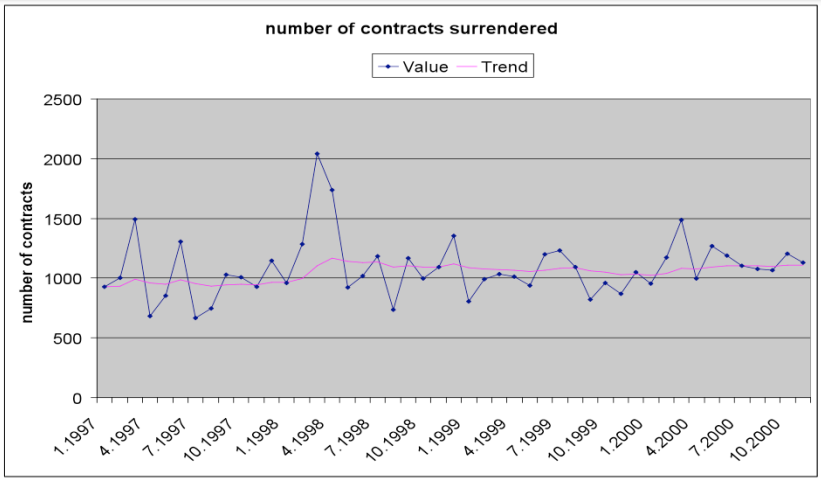
Life Insurance Surrender

- Definition:
  - Action by the owner of a cash value policy to relinquish it for its cash surrender value.
- The company pays the cash surrender value.
- The surrender option normally also extends to any dividends or bonuses as well.
- Contracts are at least three years old.

Life Insurance Surrender



Life Insurance Surrender



Life Insurance Surrender

Questions of business area “Private Insurance CH”

- insurance agents, GAs, RDZs with higher share
- substitution of contracts
- noticeable points in time w.r.t. product life, year, ...
- characterisation of subgroups concerning
  - age, profession, region, ...
  - single premiums vs. periodic

Goal of the Analysis

Problem

Detection of contracts threatened by surrender

- Which kinds of contracts do exist?
- What characterises the threat?

Solution

- Characterisation of the status of contracts at the beginning of the year
- Characterisation of surrendered contracts
- Data mining to identify groups of contracts at risk
- Predict risk of surrender acc. to data mining results

Analysis Model

start
31.12.1999

end
30.11.2000

1.284 contracts surrendered

117.216
contracts

128.121
contracts

Some Results

Default Risk	(7.2%)	Hidden Cause
• is new agent = no, is insured person = no, is premium payer = no	(19.6%)	• Social situation
• kind of premium financing = premium free, product = 50	(15.7%)	• Product
• kind of future provisions = 3b contract, age = up to 30 years	(14.9%)	• Customer
• start of contact = 1997, is new kind of premium financing = no, is new premium = yes	(10.3%)	• Business decision
• share of product life = greater 66 %, actual premium = up to 1000 SFr	(3.73%)	• Economic situation

Content

Data Mining Applications:

- Knowledge Discovery Services (KDS)
  Data Mining for Customer Relationship Management (CRM):
  - Acquire the rigth Cutomers,
  - make them profitable, and
  - keep them
- Knowledge Discovery Application (KDA)
  Integrate Data Mining into an application
  - Detect Money Laundering
- Summary

Detecting Money Laundering Activities

The Business Problem

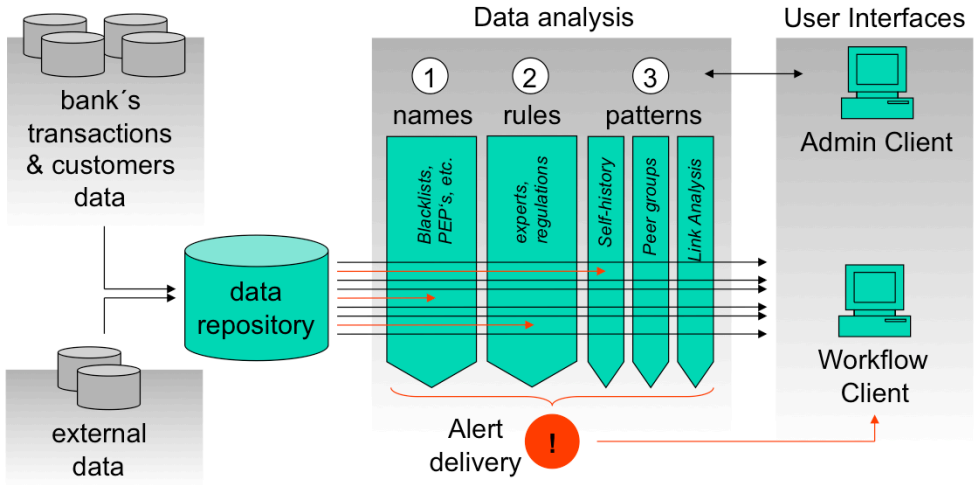
- Size of worldwide money laundering per year US$ 590-1‘500 billion
- Over 95% of delinquency sum still undiscovered
- Criminal potential obvious since September 11, 2001; top-priority for countering the financing of terrorism
- Significant damage of reputation and high fines for involved financial institutions and managers
- FATF (financial action task force) demands for stronger regulations in affiliated countries
- Governments strengthen anti-money laundering laws and regulations
- Effective Money Laundering detection by bank‘s helps to protect the secrecy of banking

Detecting Money Laundering Activities

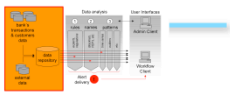
Examples of what has to be detected

- transactions from/to uncooperative countries or exposed persons
- unusual high cash deposits
- high level of activity on accounts that are generally little used
- withdrawal of assets shortly after they were credited to the account
- many payments from different persons to one account
- repeated credits just under the limit
- fast flow of a high volume of money through an account
- and many more ... e.g. have a look at:
  - FIU's in action: 100 cases from the Egmont Group
    http://www.gfsc.guernseyci.com/documents/fiu_in_action_full.pdf
  - Yearly report of the Swiss MROS
    http://internet.bap.admin.ch/d/aktuell/berichte/mros-2001_d.pdf

Overview of kdprevent™

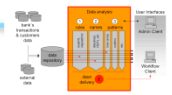


Data collection and integration in kdprevent™



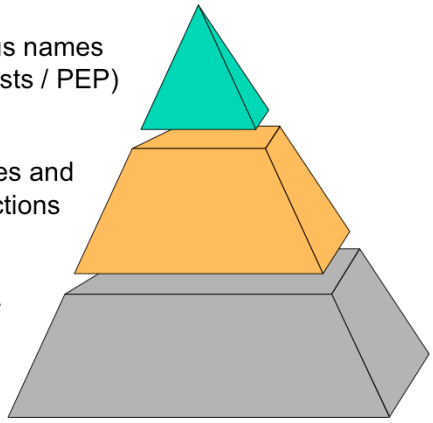
- data retrieval from any operational sources and transformation into consistent database
- basic data sources: customer profile (CIF and related data), customer view (linked CIF's), accounts, transactions and many other behaviour oriented data
- automatic and robust loading process with supportive error correction
- business-related description of data structure for the ease of use simultanously in other parts of the system
- statistical and quality measures for focused tuning of all technical processes

Data analysis in kdprevent™



How to detect suspicious activities?

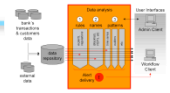
1. by looking for suspicious names and actors (*e.g.* black lists / PEP)
2. by applying specific rules and thresholds (*e.g.* transactions > CHF 200'000)
3. by detecting *suspicious patterns* through advanced analytics



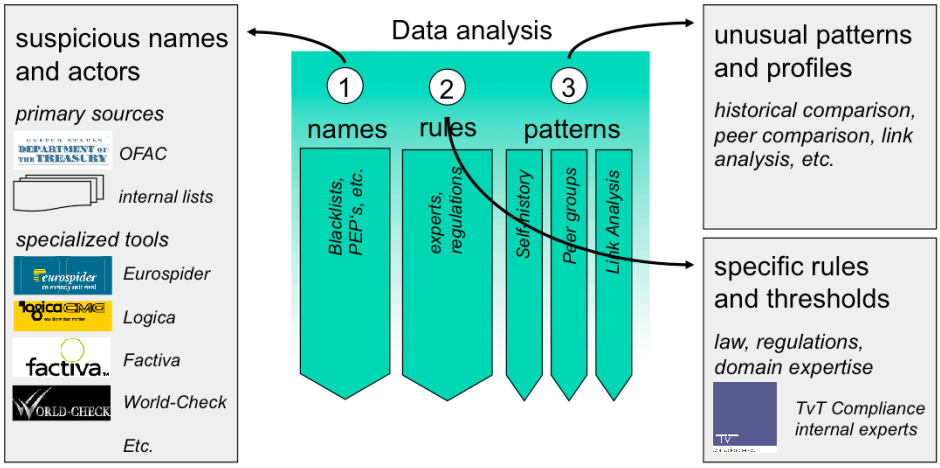
search criteria are known in advance

search criteria are unknown

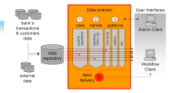
Data analysis in kdprevent™



- Data analysis: three core detection techniques



Data analysis in kdprevent™ ①

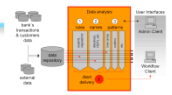


Detecting suspicious names / actors

- Is a must: not detecting exposed persons (PEP or officially black listed) is ignorant.
- Is useful: potential extension into account opening process.
- Is not sufficient: criminals know their registration, hide their identity and cover their personal trail.

- kdpreventTM provides matching of OFAC and internal lists as well as integration of third party products (e.g. Eurospider's Relevancy Compliance Support, World-Check, Factiva).
- kdpreventTM supports any solution appropriate to a banks current situation and strategy.

Data analysis in kdprevent™ ②

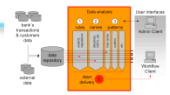


Applying specific rules / thresholds

- Is a must: regulations ask for rules and thresholds (e.g. non-cooperative countries) and rough filtering criteria can be applied.
- Is convenient: individual, more complex filtering criteria can be defined.
- Is not sufficient: criminals know published regulations, and investigators cannot fully predict upcoming fraudulent patterns.

- kdprevent™ provides a flexible and powerful rule-builder that allows the definition of arbitrary rules, individual queries, and reports.

Data analysis in kdprevent™ ③

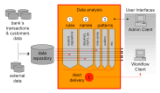


Detecting unusual patterns / profiles

- Is a must: regulations ask for identification of unusual and suspicious activities and patterns (without defining them).
- Is a must: previously unknown fraudulent patterns can only be detected by applying unbiased (objective) analytical methods.
- Is a must: pattern identification and profiling automatically adapt to changing customer (and money launderers) behaviour.

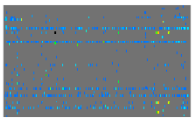
- kdprevent™ provides a powerful set of analytical methods to detect unusual, unexpected and suspicious behavioral patterns in customer activity profiles.

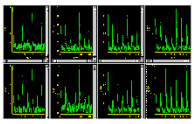
Data analysis in kdprevent™ ③

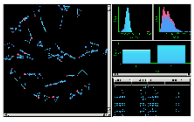


Detecting unusual patterns / profiles

- Pattern discovery 1: self history
  - e.g. unusual activity in an account history based on multidimensional time series analysis and comparision
    - time series analysis and comparision
- Pattern discovery 2: peer groups
  - e.g. unusual behaviour compared to peer group based on natural clusters and/or pre-defined segments
    - clustering, segmentation and outlyer detection
- Pattern discovery 3: link analysis
  - e.g. similarities in different accounts based on connected/linked transactions that are not otherwise expected to occur
    - Pattern detection and matching







Content

Data Mining Applications:

- Knowledge Discovery Services (KDS)
  Data Mining for Customer Relationship Management (CRM):
  - Acquire the rigth Cutomers,
  - make them profitable, and
  - keep them
- Knowledge Discovery Application (KDA)
  Integrate Data Mining into an application
  - Detect Money Laundering

Summary

Summary

- Knowledge Discovery Services
  - need a clear business value
  - adequate data preparation is the key success factor
  - reuse of KD-cases is the key to offer profitable services
  - think big, start small:
    - search for the most valuable and realistic sub-problems to solve
  - to sell and do Knowledge Discovery Services you have to be:
    - Business Consultant
    - Data Mining Expert
    - Project Manager
- Knowledge Discovery Applications
  - enable application, that are not possible without KD
  - The applications provide a fixed context that enables the automatisation of the KD-process

References

- Reported cases are cases done by either
  - kdlabs AG (projects for various customers)
  - Swiss Life (internal projects)
- kdprevent and kdimpact are products of the kdlabs AG http://www.kdlabs.com/

For further cases:

- Hippner; Küsters; Meyer; Wilde (Eds.): Handbuch Data Mining im Marketing,Vieweg, 2001.