

## ■ Knowledge Discovery Services and Applications

kdlabs AG  
[www.kdlabs.com](http://www.kdlabs.com)  
Dr. Jörg-Uwe Kietz

## Content

### Knowledge Discovery @ kdlabs

#### Key features of Mining Mart for KD services and applications

- Clever processing is the key to successful knowledge discovery
- Re-use is the key to provide knowledge discovery services
  - Repeat a KD-process for the same customer
  - Adapt a KD-process to a new customer
  - Make a new KD-process for a known customer
- DB based (pre-) processing of the data is the key to handle large amounts of data
- Mining Mart as an open-system

## About kdlabs



- kdlabs AG was founded in July 2000 to deliver services and to develop applications in the area of Knowledge Discovery Services (KD) and Knowledge Discovery Application (KDA).
- kdlabs core competence is KD and KDA. In addition, kdlabs staff has extensive experience in complementary fields, such as Marketing and Marketing Research, CRM and e-CRM, Data Warehousing and Application Integration.
- While kdlabs is vendor-independent, it is part of a strong partner network when it comes to the implementation of complete KDA- and CRM-solutions.

© January 2003, [kdlabs ag](#), Knowledge Discovery Services and Applications

## Our Mission



- kdlabs is driven by the basic understanding that data are being accumulated at a dramatic pace - but there is a general lack of expertise in extracting useful information and knowledge from the rapidly growing volume of data.
- kdlabs is focused on the extraction of knowledge from different data sources and from large volumes of data to enable its customers to act smarter and faster in their markets and therefore to increase their profit.

© January 2003, [kdlabs ag](#), Knowledge Discovery Services and Applications

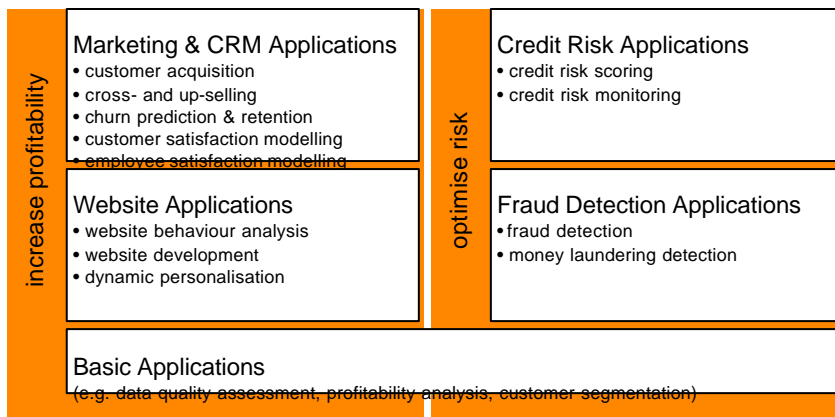
## Knowledge Discovery & Application



- Overall, the core competency of kclabs is Knowledge Discovery, Data Mining and analytical CRM.
- In short, kclabs
  - extracts available data from company's databases,
  - integrates and analyses these data and
  - delivers useful and valuable results back to the company.
- In addition, kclabs
  - implements analytical results into the company's business processes,
  - enables the company to act faster and more precise in the markets and
  - therefore optimizes the company's profitability.

© January 2003, [kclabs ag](#), Knowledge Discovery Services and Applications

## Focus on application fields



© January 2003, [kclabs ag](#), Knowledge Discovery Services and Applications

Knowledge Discovery @ kclabs

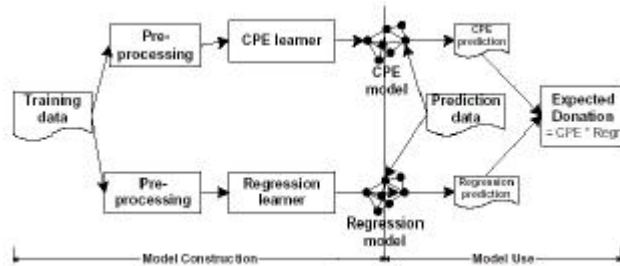
Key features of Mining Mart for KD services and applications

- 👉 Clever processing is the key to successful knowledge discovery
  - Re-use is the key to provide knowledge discovery services
    - Repeat a KD-process for the same customer
    - Adapt a KD-process to a new customer
    - Make a new KD-process for a known customer
  - DB based (pre-) processing of the data is the key to handle large amounts of data
  - Mining Mart as an open-system

© January 2003, kclabs ag, Knowledge Discovery Services and Applications

KDDCUP 98: Response Prediction

Taken from: Bernstein, Abraham, Shawndra Hill, and Foster Provost. 2002.  
[http://pages.stern.nyu.edu/~abernste/publ/IDEA\\_CeDR\\_0202.pdf](http://pages.stern.nyu.edu/~abernste/publ/IDEA_CeDR_0202.pdf)



Plan	Profit	%Gain
NN	\$6,919	-34.48
Lin	\$11,968	13.33
Log(CPE)	\$10,520	-0.37
Rule(CPE)	\$9,924	-6.02
NB(CPE)	\$9,538	-9.68
DT(CPE)	\$8,496	-19.54

Plan	Profit	%Gain
Log(CPE) + NN	\$14,914	41.23
Log(CPE) + Lin	\$14,778	39.95
Rule(CPE) + NN	\$13,672	29.47
Rule(CPE) + Lin	\$13,456	27.42
DT(CPE) + NN	\$11,055	4.69

© January 2003, kclabs ag, Knowledge Discovery Services and Applications

## Knowledge Discovery @ kclabs

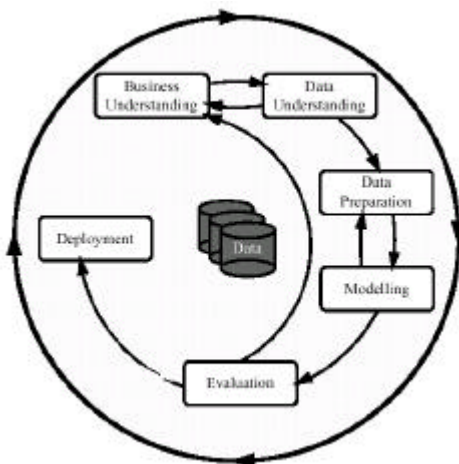
### Key features of Mining Mart for KD services and applications

- Clever preprocessing is the key to successful knowledge discovery
- 👉 Re-use is the key to provide knowledge discovery services
  - Repeat a KD-process for the same customer
  - Adapt a KD-process to a new customer
  - Make a new KD-process for a known customer
- DB based (pre-) processing of the data is the key to handle large amounts of data
- Mining Mart as an open-system

© January 2003, [kclabs ag](http://www.kclabs.ag), Knowledge Discovery Services and Applications

## The KD-Process

CRISP-DM <http://www.crisp-dm.org/>



Business Understanding	Data Understanding	Data Preparation	Modeling	Evaluation	Deployment
<ul style="list-style-type: none"> <li>Business Requirements</li> <li>Business Objectives</li> <li>Business Processes</li> <li>Business Data</li> <li>Business Constraints</li> <li>Business Risks</li> <li>Business Success Criteria</li> </ul>	<ul style="list-style-type: none"> <li>Collect Initial Data</li> <li>Initial Data Exploration</li> <li>Initial Data Cleaning</li> <li>Initial Data Profiling</li> <li>Initial Data Understanding</li> <li>Initial Data Assessment</li> <li>Initial Data Selection</li> <li>Initial Data Transformation</li> <li>Initial Data Integration</li> <li>Initial Data Storage</li> </ul>	<ul style="list-style-type: none"> <li>Final Data Selection</li> <li>Final Data Cleaning</li> <li>Final Data Profiling</li> <li>Final Data Understanding</li> <li>Final Data Assessment</li> <li>Final Data Transformation</li> <li>Final Data Integration</li> <li>Final Data Storage</li> </ul>	<ul style="list-style-type: none"> <li>Model Selection</li> <li>Model Development</li> <li>Model Evaluation</li> <li>Model Deployment</li> <li>Model Monitoring</li> <li>Model Maintenance</li> <li>Model Retirement</li> </ul>	<ul style="list-style-type: none"> <li>Model Evaluation</li> <li>Model Performance</li> <li>Model Accuracy</li> <li>Model Precision</li> <li>Model Recall</li> <li>Model F1 Score</li> <li>Model AUC</li> <li>Model ROC Curve</li> <li>Model Confusion Matrix</li> <li>Model Residuals</li> <li>Model Diagnostics</li> </ul>	<ul style="list-style-type: none"> <li>Model Deployment</li> <li>Model Monitoring</li> <li>Model Maintenance</li> <li>Model Retirement</li> <li>Model Evaluation</li> <li>Model Performance</li> <li>Model Accuracy</li> <li>Model Precision</li> <li>Model Recall</li> <li>Model F1 Score</li> <li>Model AUC</li> <li>Model ROC Curve</li> <li>Model Confusion Matrix</li> <li>Model Residuals</li> <li>Model Diagnostics</li> </ul>

© January 2003, [kclabs ag](http://www.kclabs.ag), Knowledge Discovery Services and Applications

## Process Step Duration and Importance



From D. Pyle:	Time	Importance
<b>Business understanding</b>	<b>20</b>	<b>80</b>
a) Exploring the problem	10	15
b) Exploring the solution	9	14
c) Implementation specification	1	51
<b>Data preparation &amp; mining</b>	<b>80</b>	<b>20</b>
d) Data exploration	15	3
e) Data preparation	60	15
f) Modeling (data mining)	5	2

⇒ The numbers are idealized, but reflect our experiences

⇒ Doing CRISP-DM each time from scratch is not cost-effective

© January 2003, [kclabs ag](#), Knowledge Discovery Services and Applications

## Content



### Knowledge Discovery @ kclabs

#### Key features of Mining Mart for KD services and applications

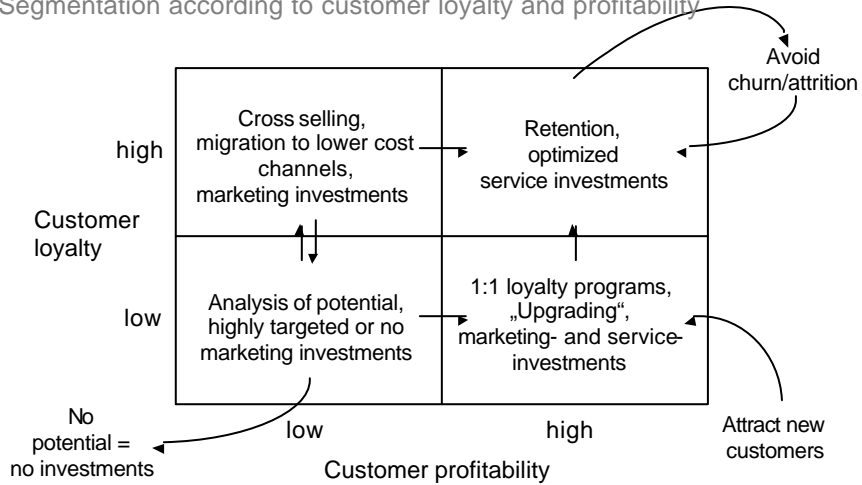
- Clever processing is the key to successful knowledge discovery
- Re-use is the key to provide knowledge discovery services
  - Repeat a KD-process for the same customer
  - Adapt a KD-process to a new customer
  - Make a new KD-process for a known customer
- DB based (pre-) processing of the data is the key to handle large amounts of data
- Mining Mart as an open-system

© January 2003, [kclabs ag](#), Knowledge Discovery Services and Applications

# CRM & Customer segmentation



Segmentation according to customer loyalty and profitability

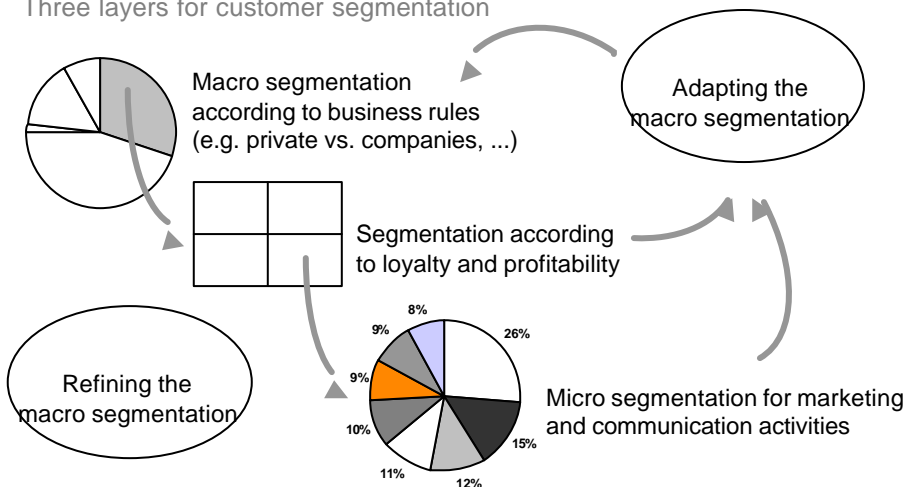


© January 2003, kclabs ag, Knowledge Discovery Services and Applications

# Customer segmentation



Three layers for customer segmentation

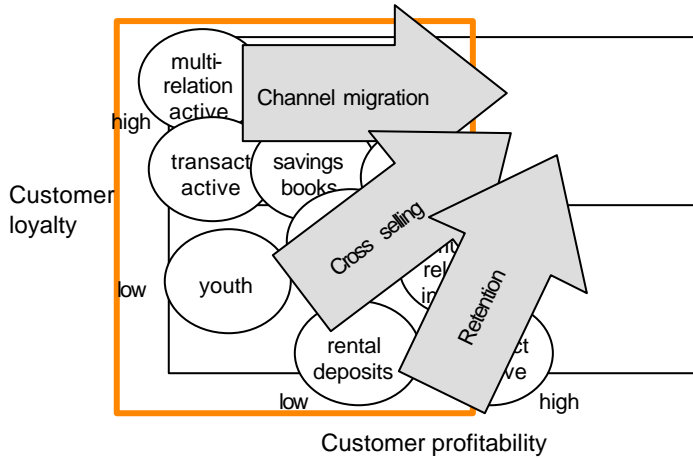


© January 2003, kclabs ag, Knowledge Discovery Services and Applications

## Segmented customer communication



Segmentation in lower retail banking: potential applications

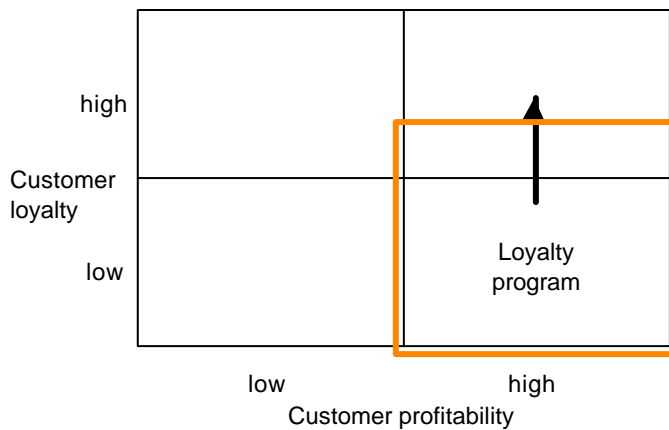


© January 2003, [kclabs ag](#), Knowledge Discovery Services and Applications

## Targeted marketing campaigns



Launching a loyalty program for customer retention



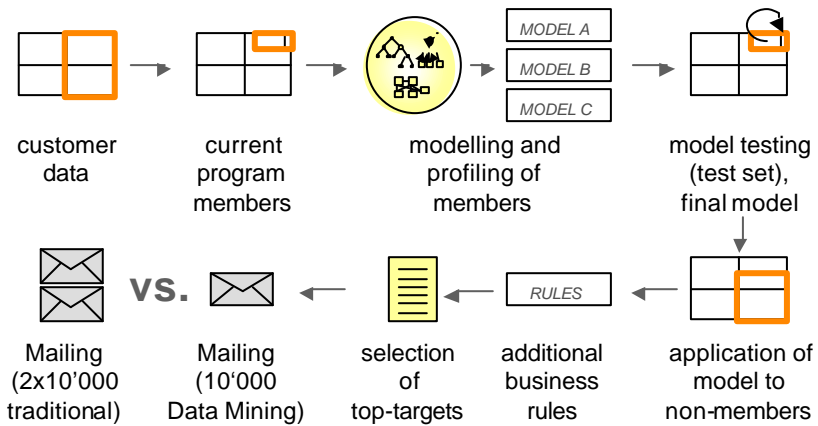
© January 2003, [kclabs ag](#), Knowledge Discovery Services and Applications



# Targeted marketing campaigns



## Process of KD-driven customer selection

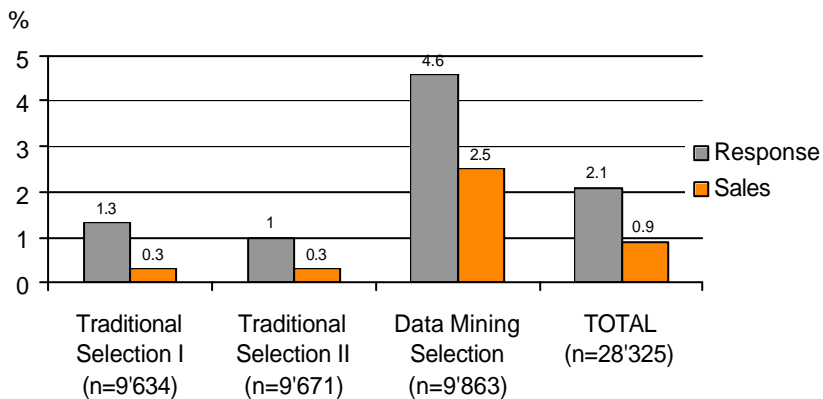


© January 2003, [kclabs ag](#), Knowledge Discovery Services and Applications

# Targeted marketing campaigns



## Mailing campaign for a loyalty program



© January 2003, [kclabs ag](#), Knowledge Discovery Services and Applications

### Re-use is the key to provide knowledge discovery services

- Repeat a KD-process for the same customer, e.g.:
  - KPI's, like customer and employee satisfaction, must be build every year
  - Marketing campaigns are repeated, e.g. for different segments or products
  - Risk assessment has to be updated
  - ...
- What can be reused
  - ⇒ same business problem
  - ⇒ same KD-process
  - ⇒ same data format
  - ⇒ most likely the same data quality problems
  - ⇒ different data content

### Knowledge Discovery @ kclabs

#### Key features of Mining Mart for KD services and applications

- Clever processing is the key to successful knowledge discovery
- Re-use is the key to provide knowledge discovery services
  - Repeat a KD-process for the same customer
  - Adapt a KD-process to a new customer
  - Make a new KD-process for a known customer
- DB based (pre-) processing of the data is the key to handle large amounts of data
- Mining Mart as an open-system

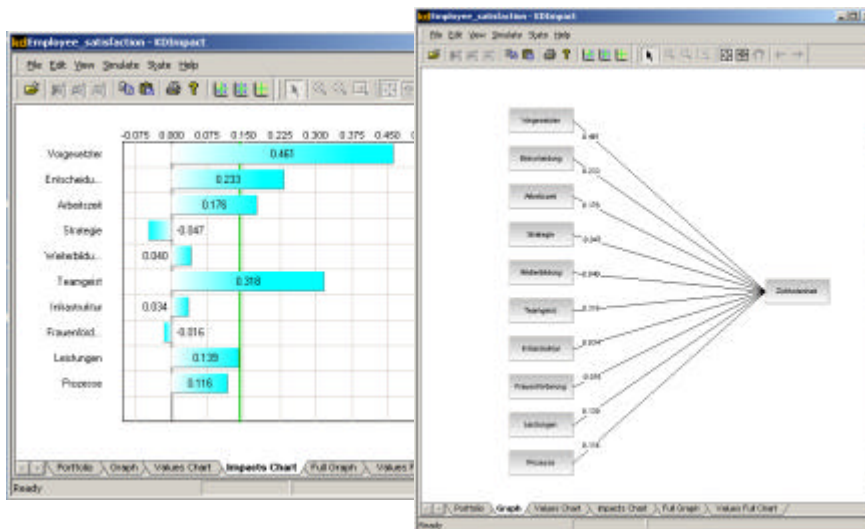
## Causal Modelling for Marketing Research



- 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)

© January 2003, [kclabs ag](#), Knowledge Discovery Services and Applications

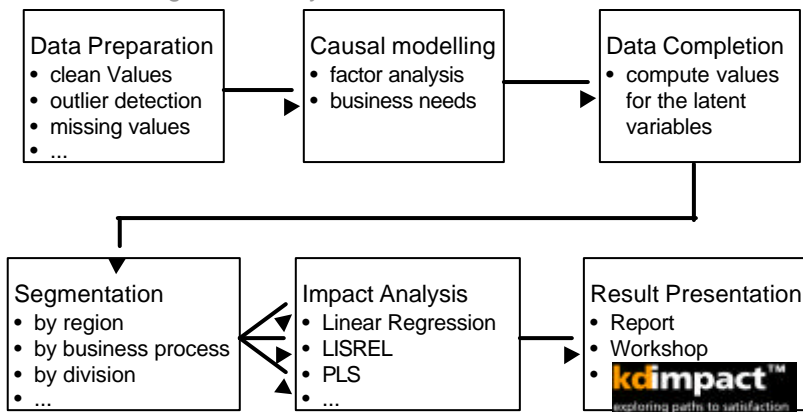
## Causal Modelling for Marketing Research



© January 2003, [kclabs ag](#), Knowledge Discovery Services and Applications

- **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:**
  - kdimapct

## The Knowledge Discovery Process



### Re-use is the key to provide knowledge discovery services

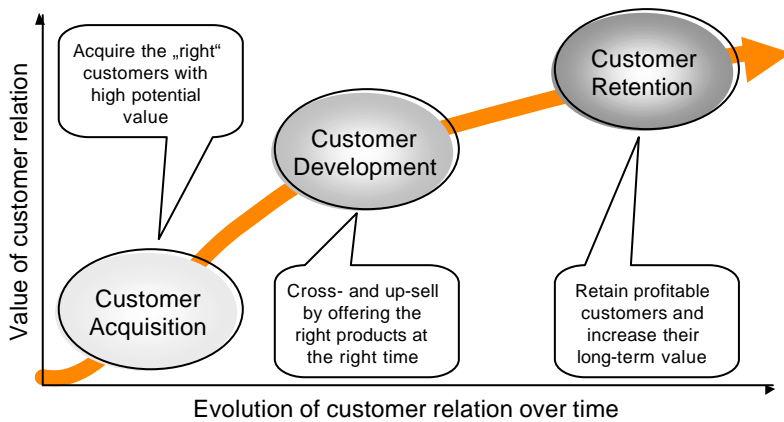
- **Adapt a KD-process to a new customer**
  - KPI's - and the methods to obtain them - should be comparable
  - CRM is a common methodology
  - ...
- **What can be reused**
  - ⇒ similar business problem
  - ⇒ similar KD-process
  - ⇒ different data format, but similar type of data
  - ⇒ similar types of data quality problems
  - ⇒ different data content

### Knowledge Discovery @ kclabs

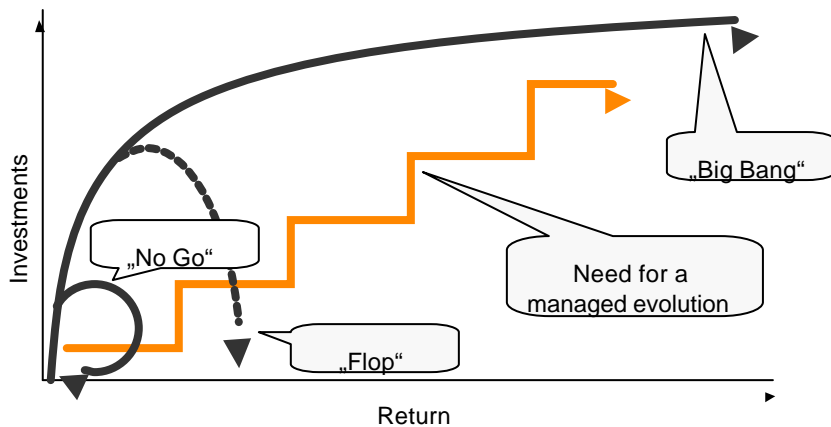
#### Key features of Mining Mart for KD services and applications

- **Clever processing is the key to successful knowledge discovery**
- **Re-use is the key to provide knowledge discovery services**
  - Repeat a KD-process for the same customer
  - Adapt a KD-process to a new customer
  - Make a new KD-process for a known customer
- **DB based (pre-) processing of the data is the key to handle large amounts of data**
- **Mining Mart as an open-system**

Three simple business goals of CRM



© January 2003, [kclabs ag](#), Knowledge Discovery Services and Applications




© January 2003, [kclabs ag](#), Knowledge Discovery Services and Applications

Re-use is the key to provide knowledge discovery services

- **Make a new KD-process for a known customer**
  - have an overall vision (as CRM)
  - introduce KD in small, realistic and controllable steps
  - priorities them according to business value and expected ROI
- **What can be reused**
  - ⇒ different business problem
  - ⇒ different KD-process
  - ⇒ partially the same data format
  - ⇒ partially the same data quality problems
  - ⇒ partially the same data content

Knowledge Discovery @ kclabs

Key features of Mining Mart for KD services and applications

- **Clever processing is the key to successful knowledge discovery**
- **Re-use is the key to provide knowledge discovery services**
  - Repeat a KD-process for the same customer
  - Adapt a KD-process to a new customer
  - Make a new KD-process for a known customer
-  **DB based (pre-) processing of the data is the key to handle large amounts of data**
- **Mining Mart as an open-system**

### 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
- Large banks have millions of transactions per day to check

© January 2003, [kd labs ag](#), Knowledge Discovery Services and Applications

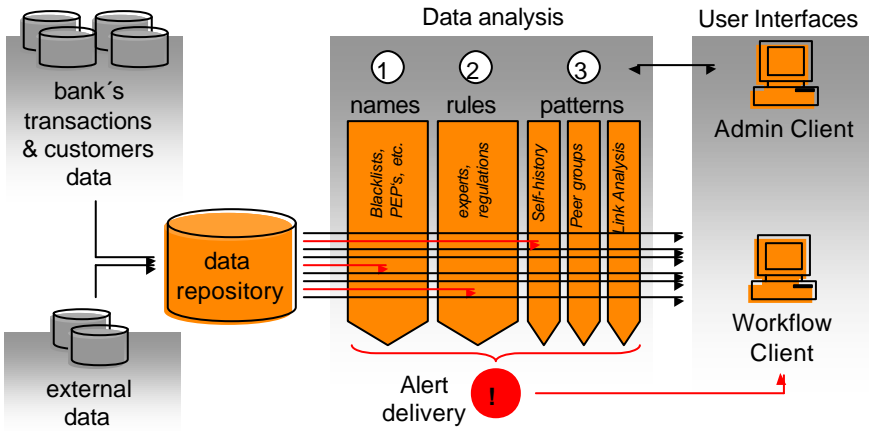
### Examples of what has to be detected

- 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](#)
  - [Yearly report of the Swiss MROS](#)

© January 2003, [kd labs ag](#), Knowledge Discovery Services and Applications

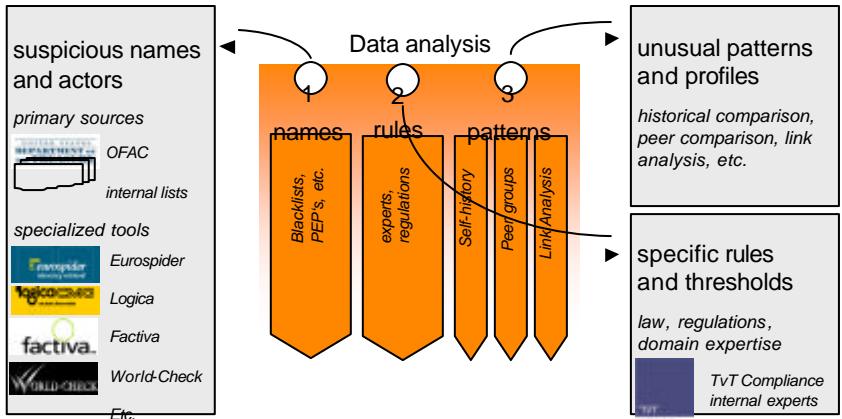


Overview



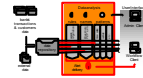
© January 2003, kclabs ag, Knowledge Discovery Services and Applications

Data analysis: three core detection techniques

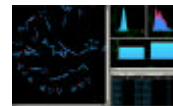
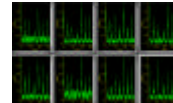


© January 2003, kclabs ag, Knowledge Discovery Services and Applications

## Data analysis: detecting unusual patterns / profiles



- Pattern discovery 1: self history
  - e.g. unusual activity in an account history based on multidimensional time series analysis and comparison
    - ⇒ time series analysis and comparison
- 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 outlier 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



© January 2003, kclabs ag, Knowledge Discovery Services and Applications

## Pre-processing in DMBS and DM-suite

- The raw data (transactions) have to be processed in several ways
  - Aggregations (e.g. total amount incoming cash per week)
  - Time-series (e.g. volume of the days of a month)
  - Customer profiles
  - ...
- E.g. the aggregation and time-series building
  - takes ~15min per 1 mio. transactions to process in a DBMS
  - it is not possible to (pre-) process them in current data mining workbenches
    - as they have only basic operations to be performed in the DB
    - any more complex operations tries (and fails) to load all data

© January 2003, kclabs ag, Knowledge Discovery Services and Applications

## Knowledge Discovery @ kclabs

### Key features of Mining Mart for KD services and applications

- Clever processing is the key to successful knowledge discovery
- Re-use is the key to provide knowledge discovery services
  - Repeat a KD-process for the same customer
  - Adapt a KD-process to a new customer
  - Make a new KD-process for a known customer
- DB based (pre-) processing of the data is the key to handle large amounts of data



### Mining Mart as an open-system

© January 2003, [kclabs ag](#), Knowledge Discovery Services and Applications

## Mining Mart as an open system

### Mining Mart under the GNU general public license?

⇒ The “Linux” of the Data Mining Workbenches?

What could that mean?

- Everyone can get, use and extend the software (e.g. operators)
- Successful extensions can be given back to public
- Everyone has access to successful KD-cases
- Successful KD-cases can be stored in the public case-base

Why could it be interesting to contribute to it, for

- the Data Mining Workbench providers
- the Data Mining Services and Application providers
- the (large scale) Data Mining Users
- the Consortium

© January 2003, [kclabs ag](#), Knowledge Discovery Services and Applications

## Mining Mart as an open system



Why could it be interesting to contribute to it, for the Data Mining Workbench (DMW) providers

The System, an unwanted concurrence? Not really!

- The Mining Mart strength is pre-processing
- The DMW's strength is mining and visualisation
- Pre-processing requirements are the limit of the DMW's usability

Why should they contribute to the case-base?

- It makes their main business (selling their tool) easier
- They can access public cases and adapt them for their system
- Deploying cases of successful usage of their system is a free advertisement of their system
- They gain from public improvements of their deployed cases

© January 2003, [kclabs ag](#), Knowledge Discovery Services and Applications

## Mining Mart as an open system



Why could it be interesting to contribute to it, for the Data Mining Services and Application providers

There is an open system alternative to DMW's

- The "base price" of deployed KD services is smaller
- The system can be much easier extended
- Giving extension back to public reduces "Maintenance-Costs"

The public case base is an unwanted concurrence? Not really!

- They don't have to publish the technical cases, and the system helps to maintain an internal case-base as well.
- Published "business parts" of cases can be an advertisement
- Public cases can serve for inspiration and training of staff.

© January 2003, [kclabs ag](#), Knowledge Discovery Services and Applications

## Mining Mart as an open system



Why could it be interesting to contribute to it, for the (large scale) Data Mining Users

It's not their business, but they can improve their business better with data mining

Advantages due to an open system

- They can access a free system
- They don't have to maintain the extensions they made

Advantages due to an public Case-base

- They can access a vendor independent reference of successful cases
- They can take the cases or contact the provider
- They get free improvements and maintenance on published cases

© January 2003, [kclabs ag](#), Knowledge Discovery Services and Applications

## Mining Mart as an open system



Why could it be interesting to contribute to it, for the Consortium

Selling Mining Mart is not an option as

- No one owns the whole system
  - No one is interested to provide support
  - Without active maintenance it stops running sooner or later
- ⇒ There is no market to sell it

Being a mayor provider of an open Mining Mart is advertisement

- Be a "Linus Thorwald"
- An successful open system Mining Mart can generate business for you around the system, if you use and keep your advantage of knowing the system

© January 2003, [kclabs ag](#), Knowledge Discovery Services and Applications

**Mining Mart can provide**

- **unique features that are**
- **urgently needed to do**
- **Knowledge Discovery Services & Applications**

- ⇒ **A system to support large scale data pre-processing in a DMBS**
- ⇒ **A public vendor independent reference of successful KD cases**
- ⇒ **Case re-use and adaptation for effective KD services**
- ⇒ **A open public software environment for expert users**