Knowledge Discovery Services and Applications

kdlabs AG
www.kdlabs.com
Dr. Jörg-Uwe Kietz

Content

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.

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.
Overall, the core competency of kdlabs is Knowledge Discovery, Data Mining and analytical CRM.

In short, kdlabs
- extracts available data from company’s databases,
- integrates and analyses these data and
- delivers useful and valuable results back to the company.

In addition, kdlabs
- 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.

Focus on application fields

Marketing & CRM Applications
- customer acquisition
- cross- and up-selling
- churn prediction & retention
- customer satisfaction modelling
- employee satisfaction modelling

Website Applications
- website behaviour analysis
- website development
- dynamic personalisation

Basic Applications
(e.g. data quality assessment, profitability analysis, customer segmentation)

Credit Risk Applications
- credit risk scoring
- credit risk monitoring

Fraud Detection Applications
- fraud detection
- money laundering detection
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

KDDCUP 98: Response Prediction

Knowledge Discovery @ kdlabs

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

The KD-Process

CRISP-DM http://www.crisp-dm.org/
**Process Step Duration and Importance**

From D. Pyle:

<table>
<thead>
<tr>
<th>Business understanding</th>
<th>Time</th>
<th>Importance</th>
</tr>
</thead>
<tbody>
<tr>
<td>a) Exploring the problem</td>
<td>10</td>
<td>15</td>
</tr>
<tr>
<td>b) Exploring the solution</td>
<td>9</td>
<td>14</td>
</tr>
<tr>
<td>c) Implementation specification</td>
<td>1</td>
<td>51</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Data preparation &amp; mining</th>
<th>Time</th>
<th>Importance</th>
</tr>
</thead>
<tbody>
<tr>
<td>d) Data exploration</td>
<td>15</td>
<td>3</td>
</tr>
<tr>
<td>e) Data preparation</td>
<td>60</td>
<td>15</td>
</tr>
<tr>
<td>f) Modeling (data mining)</td>
<td>5</td>
<td>2</td>
</tr>
</tbody>
</table>

⇒ The numbers are idealized, but reflect our experiences
⇒ Doing CRISP-DM each time from scratch is not cost-effective

---

**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
Segmentation according to customer loyalty and profitability

<table>
<thead>
<tr>
<th>Customer loyalty</th>
<th>Customer profitability</th>
</tr>
</thead>
<tbody>
<tr>
<td>high</td>
<td>high</td>
</tr>
<tr>
<td>low</td>
<td>low</td>
</tr>
</tbody>
</table>

Avoid churn/attrition

- Cross selling, migration to lower cost channels, marketing investments
- Retention, optimized service investments

Analysis of potential, highly targeted or no marketing investments

- 1:1 loyalty programs, „Upgrading“, marketing and service investments
- Attract new customers

No potential = no investments

1:1 loyalty programs, „Upgrading“, marketing and service investments

Attract new customers

Three layers for customer segmentation

- Macro segmentation according to business rules (e.g. private vs. companies, ...)
- Adapting the macro segmentation
- Refining the macro segmentation
- Micro segmentation for marketing and communication activities
Segmented customer communication

Segmentation in lower retail banking: potential applications

Targeted marketing campaigns

Launching a loyalty program for customer retention
Targeted marketing campaigns

Process of KD-driven customer selection

- Customer data
- Current program members
- Modelling and profiling of members
- Model testing (test set), final model

VS.

- Mailing (2x10'000 traditional)
- Mailing (10'000 Data Mining)
- Selection of top-targets
- Additional business rules

Application of model to non-members

Targeted marketing campaigns

Mailing campaign for a loyalty program

<table>
<thead>
<tr>
<th></th>
<th>Response</th>
<th>Sales</th>
</tr>
</thead>
<tbody>
<tr>
<td>Traditional Selection I (n=9'634)</td>
<td>1.3%</td>
<td>0.3%</td>
</tr>
<tr>
<td>Traditional Selection II (n=9'671)</td>
<td>1%</td>
<td>0.3%</td>
</tr>
<tr>
<td>Data Mining Selection (n=9'863)</td>
<td>2.5%</td>
<td>2.1%</td>
</tr>
<tr>
<td>TOTAL (n=28'325)</td>
<td>2.1%</td>
<td>0.9%</td>
</tr>
</tbody>
</table>
Re-use of KD-processes

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

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
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

- 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

The Knowledge Discovery Process

Data Preparation
- clean Values
- outlier detection
- missing values
- ...

Causal modelling
- factor analysis
- business needs

Data Completion
- compute values for the latent variables

Segmentation
- by region
- by business process
- by division
- ...

Impact Analysis
- Linear Regression
- LISREL
- PLS
- ...

Result Presentation
- Report
- Workshop
Re-use of KD-processes

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

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
Three simple business goals of CRM

Customer Acquisition
- Acquire the "right" customers with high potential value

Customer Development
- Cross- and up-sell by offering the right products at the right time

Customer Retention
- Retain profitable customers and increase their long-term value

Doing KD for CRM

Investments
- "No Go"
- "Flop"

Return
- "Big Bang"
- Need for a managed evolution

© January 2003, kd labs ag. Knowledge Discovery Services and Applications
Re-use of KD-processes

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

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
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
- Large banks have millions of transactions per day to check

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
Data analysis: three core detection techniques

- **Suspicious names and actors**
  - Primary sources: OFAC, internal lists
  - Specialized tools: Eurospider, Logica, Factiva, World-Check, etc.

- **Data analysis**
  - Names, rules, patterns
  - Historical comparison, peer comparison, link analysis, etc.

- **Unusual patterns and profiles**
  - Specific rules and thresholds
    - Law, regulations, domain expertise
    - TzT Compliance, internal experts

---

Overview

- Bank’s transactions & customers data
- Data analysis: names, rules, patterns
- User Interfaces
  - Admin Client
  - Workflow Client

---

© January 2003, kd labs 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

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 (an fails) to load all data
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

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
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, kd labs 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!

- The don’t have to publish the technical cases, and the systems 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, kd labs 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

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
Summary

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