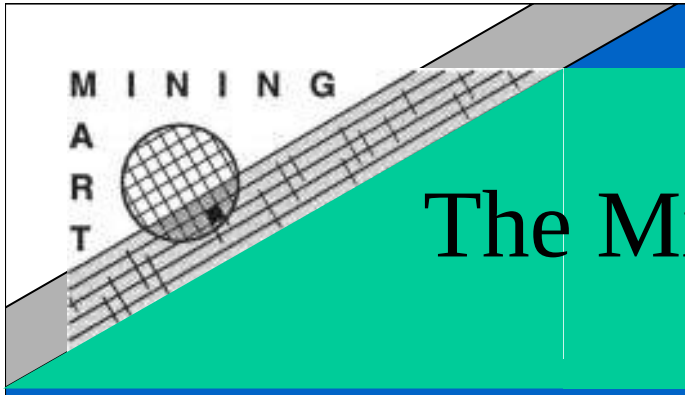
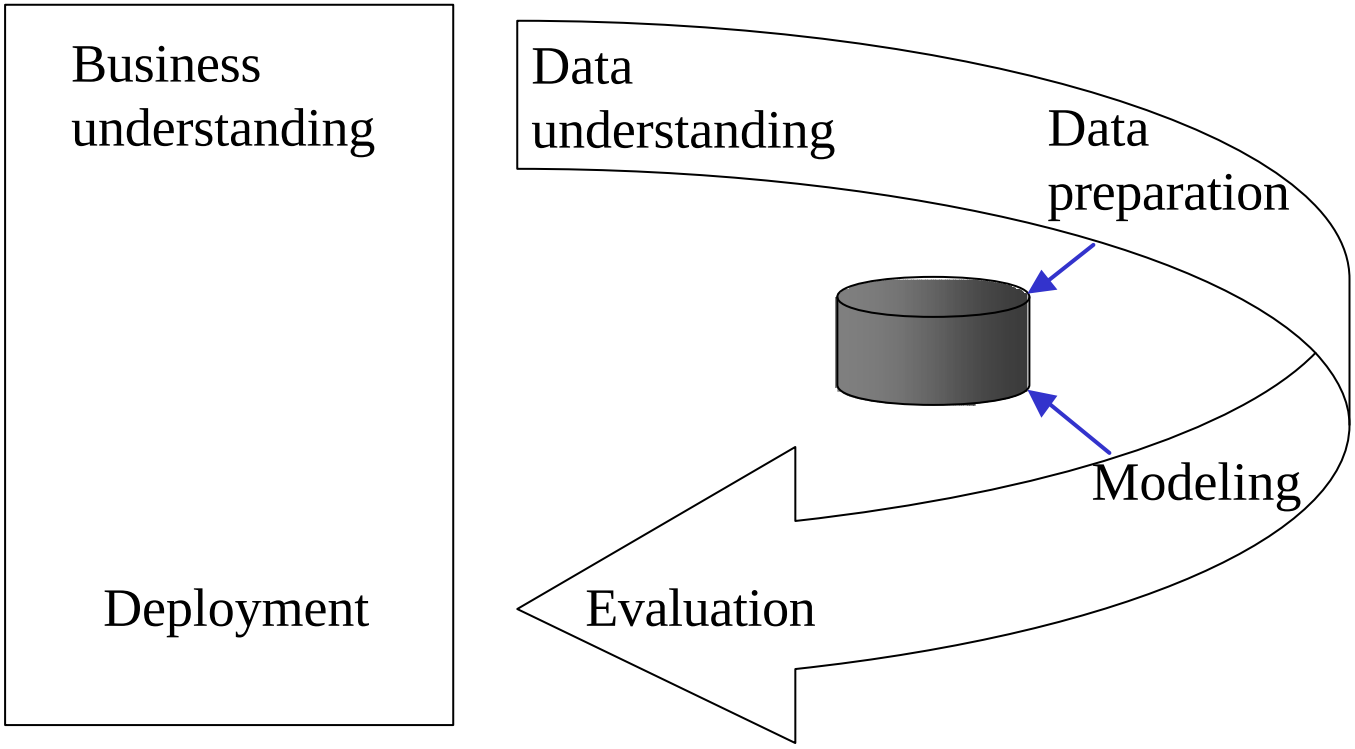


The Mining Mart Approach

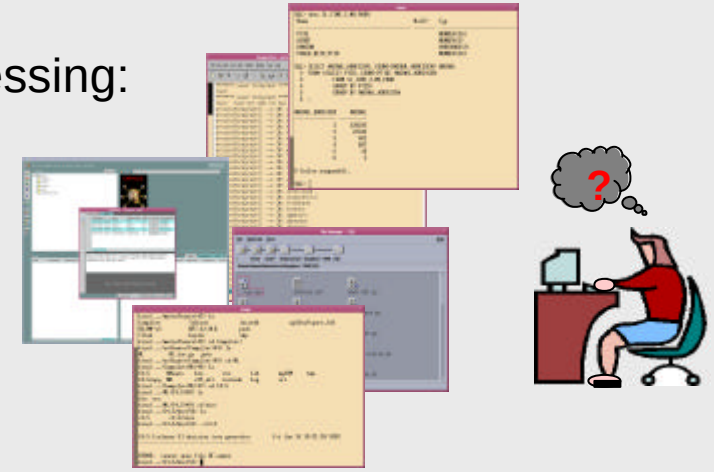
1. The process of knowledge discovery and its common practice
2. Supporting the re-use of successful knowledge discovery cases
   - Supporting pre-processing
   - Meta-data for concepts, data, and cases
   - Documenting and adapting a case
   - Compiling meta-data into SQL – executing a case
3. System demonstration
4. Summary

CRISP-DM Process Model



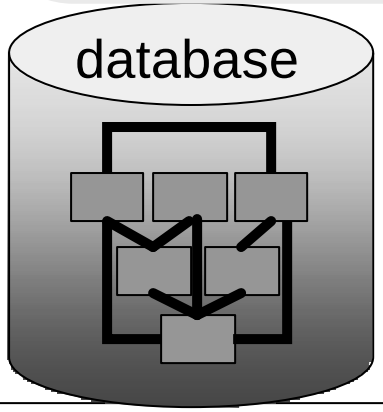
Common Practice

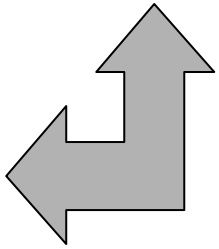
manual pre-processing:

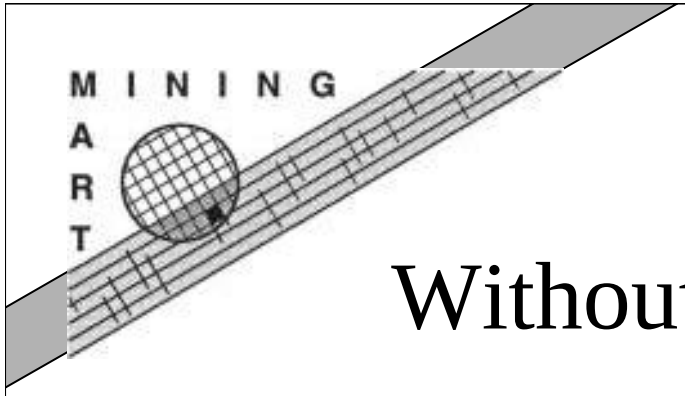


drawbacks

- tedious
- time consuming
- not re-useable
- no documentation
- low level operations







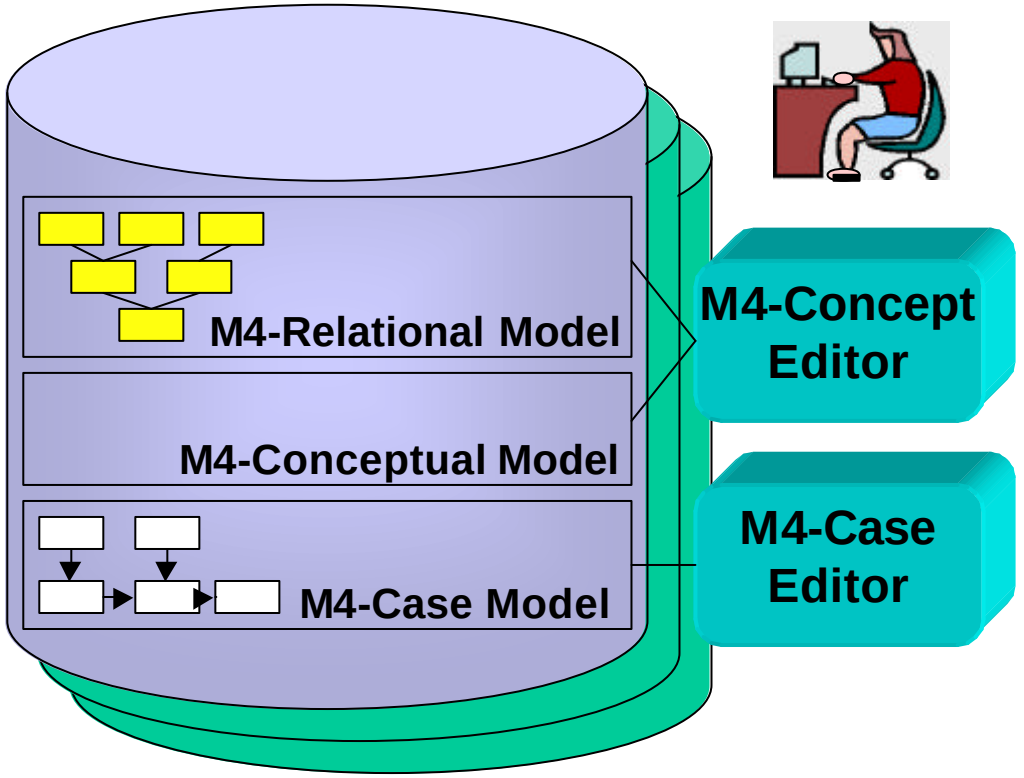
Without Mining Mart

- Pre-processing is not supported by the tools.
  - 80 % of the efforts in a knowledge discovery application are invested during pre-processing.
  - Pre-processing enhances data – better data deliver better data mining results.
- Documentation of pre-processing is missing.
  - Similar procedures are performed over and over again.
  - Experience is not passed over to new employees.
- Operators do not access the database directly, but can only handle an excerpt.

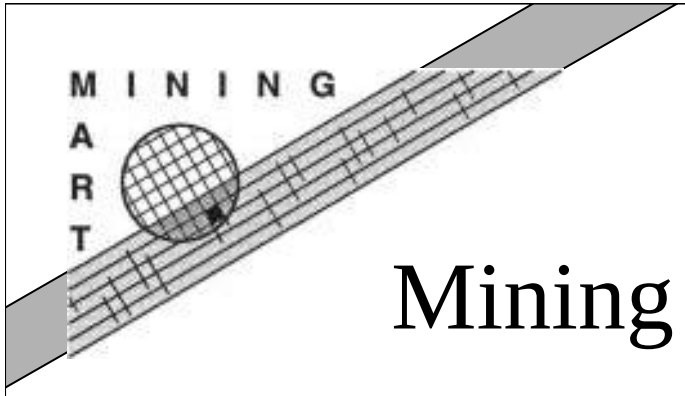
Using Mining Mart

Conceptual Model
(Shops, items, sales...)

Abstract case
(Selection of shops, items,
Running the support vector
machine...)

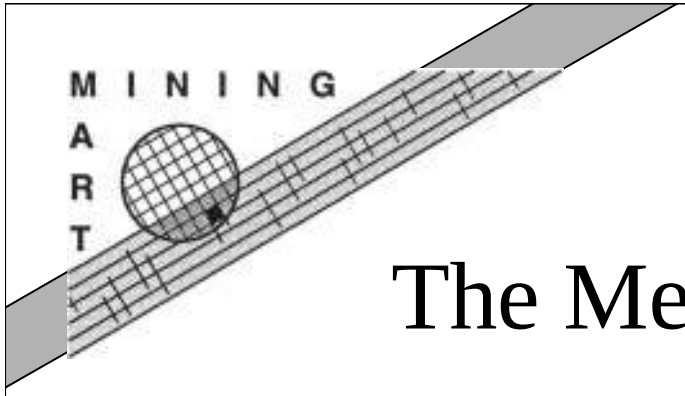


Linking business data and
conceptual model,
Compile the case and see the results!



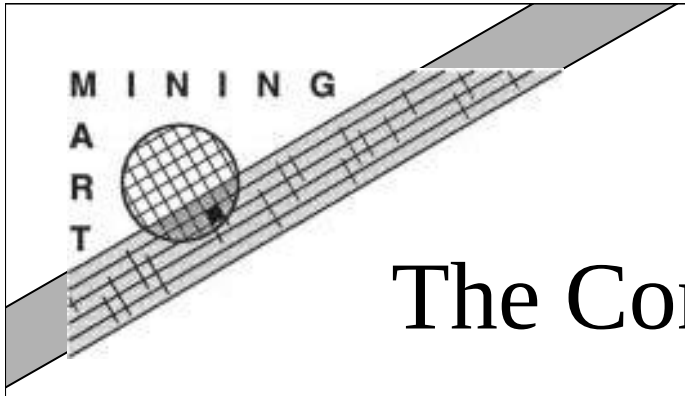
Mining Mart Users

- The database administrator delivers the relational data model.
- The data analyst
  - acquires the conceptual model from the end-user (decision maker),
  - develops (adapts) the case,
  - links relational and conceptual model,
  - runs the case and delivers the results to the end-user.



The Meta Model for Meta Data

The Relational Model describes the database	The Execution Model generates SQL statements or calls to external tools
The Conceptual Model describes the individuals and classes of the domain with their relations	The Case Model describes chains of preprocessing operators

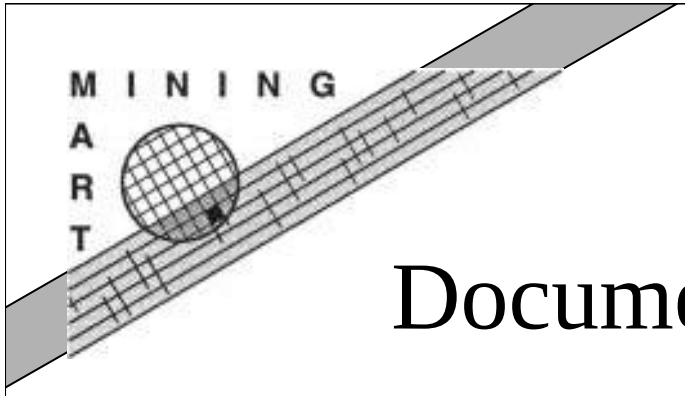


The Conceptual Model

- Concept
  - Attributes: name, subConceptRestriction
  - Associations: isA, correspondsToColumnSet, FromConcept, ToConcept, Constraints
- Relationship
- FeatureAttribute
- Value
- RoleRestriction
- DomainDataType

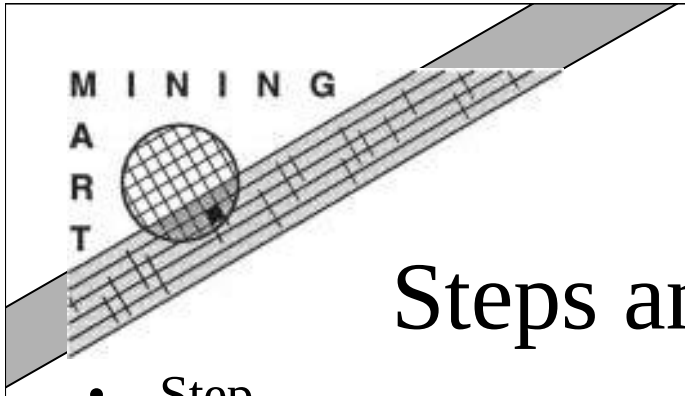
The Case Model

- Case
  - Attributes: name,
    - Case mode {test, final}
    - caseInput – list of entities from the conceptual model
    - caseOutput – concept, typically the input to data mining step
    - Documentation – free text
  - Associations: listOfSteps
    - Population – the concept of interest in this case
    - targetAttributes – FeatureAttribute to which the data analysis is applied



Documentation

- The case model documents the sequence of steps that have led to a good data mining result.
- For each step, the input, output, and parameter settings are stored.
- Since steps refer to concepts, the case model can be understood even by non-experts.

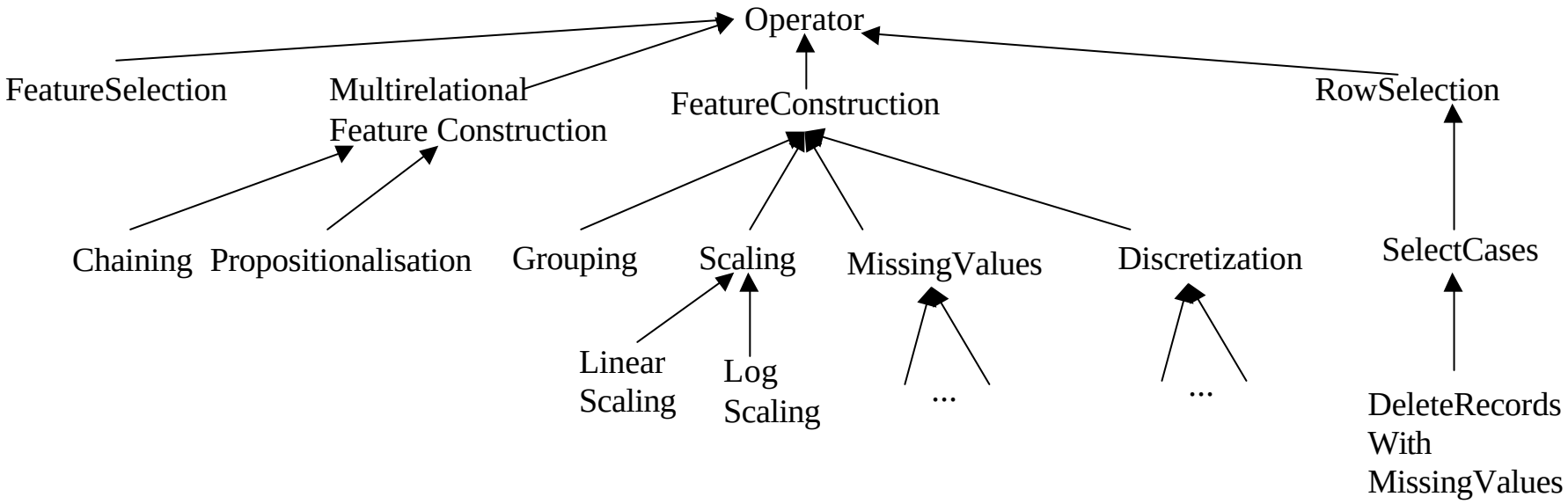


Steps and operators

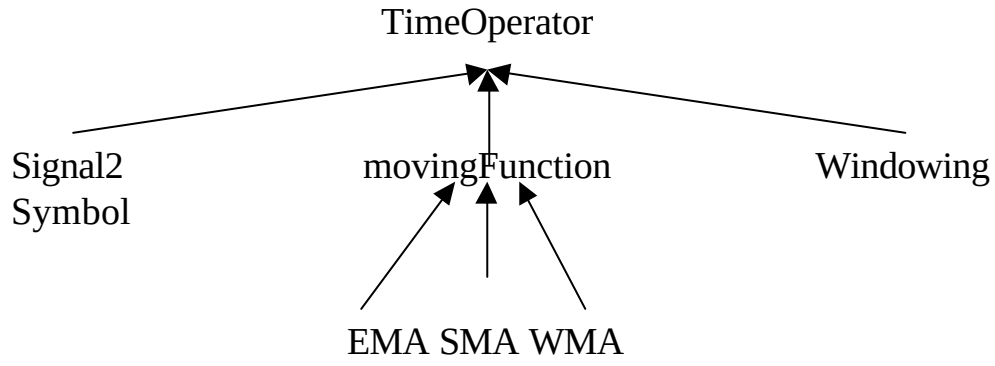
- Step
  - Attributes:name
  - Associations: belongsToCase, embedsOperator, predecessor, successor
- Operator
  - Attributes (binary): manual,
    - Loopable – apply operator several times with changed parameters
    - Multi-step – operator delivers several results which will be processed in parallel
  - Associations: all input to a step (parameters)
    - Conditions – to be checked given the data
    - Constraints – to be checked without access to data
    - Assertions – will be true after operator execution

Validity of operator chains are checked, unnecessary database scans are avoided!

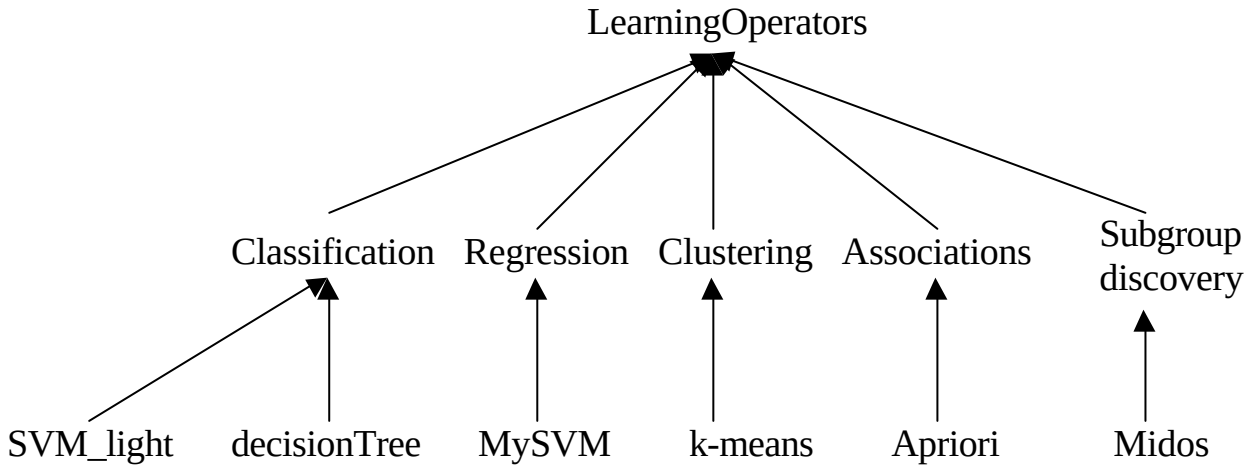
Manual Operators



Time Operators

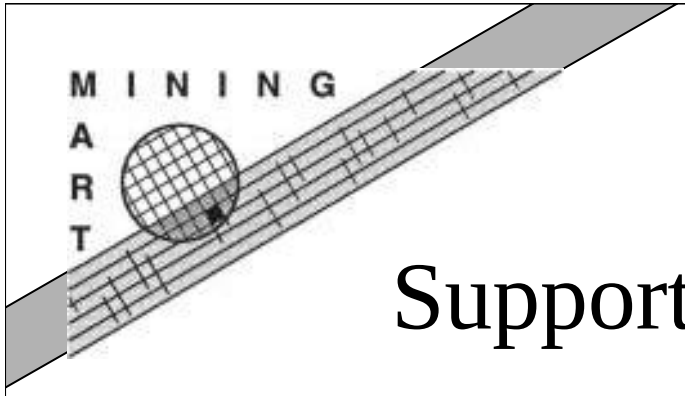


Learning Operators



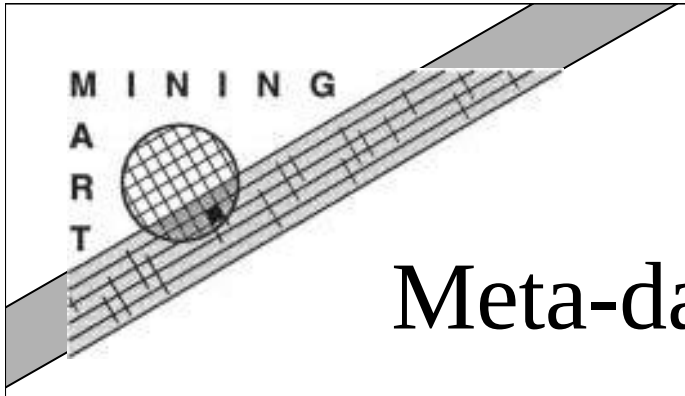
New

Learning operators are not only good for the data mining step!
Example: C4.5 for discretisation or prediction of missing values.



Supporting Pre-processing

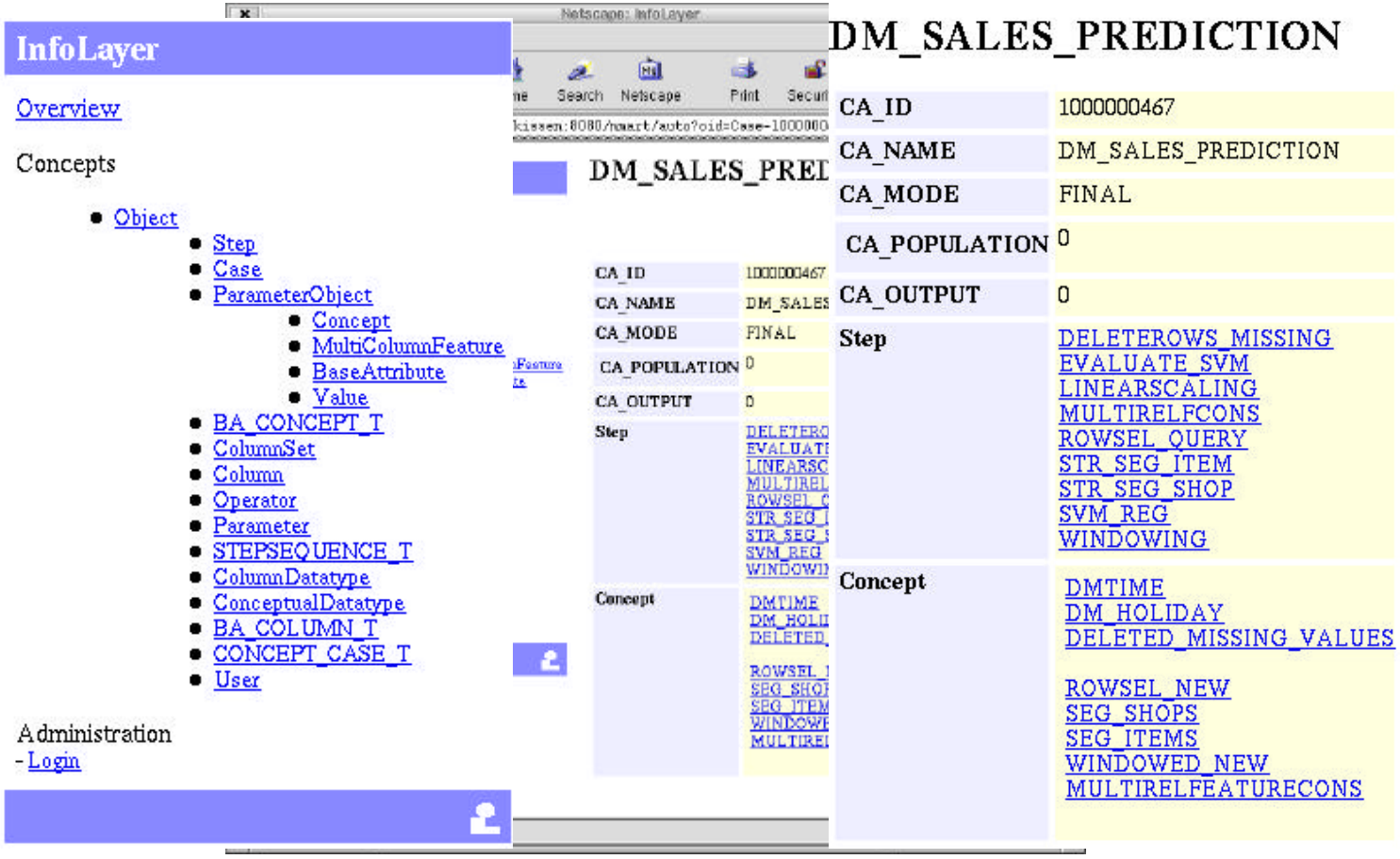
- The operators are implemented – users just select them.
- Most operators directly access the database.
- Intermediate results can be inspected.
- The system is open for the integration of further operators:
  - Store the SQL implementation
  - Store the meta-data within the M4 tables.

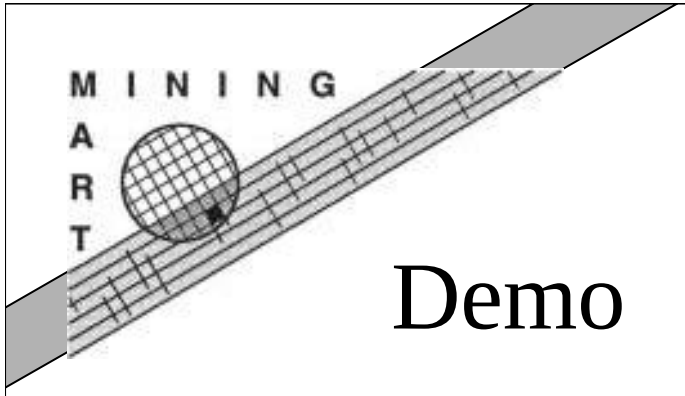


Meta-data

- Meta-model and meta-data are stored in the database.
- Used
  - in order to verify applicability conditions
  - in order to avoid unnecessary steps
  - by the compiler
  - by the GUI

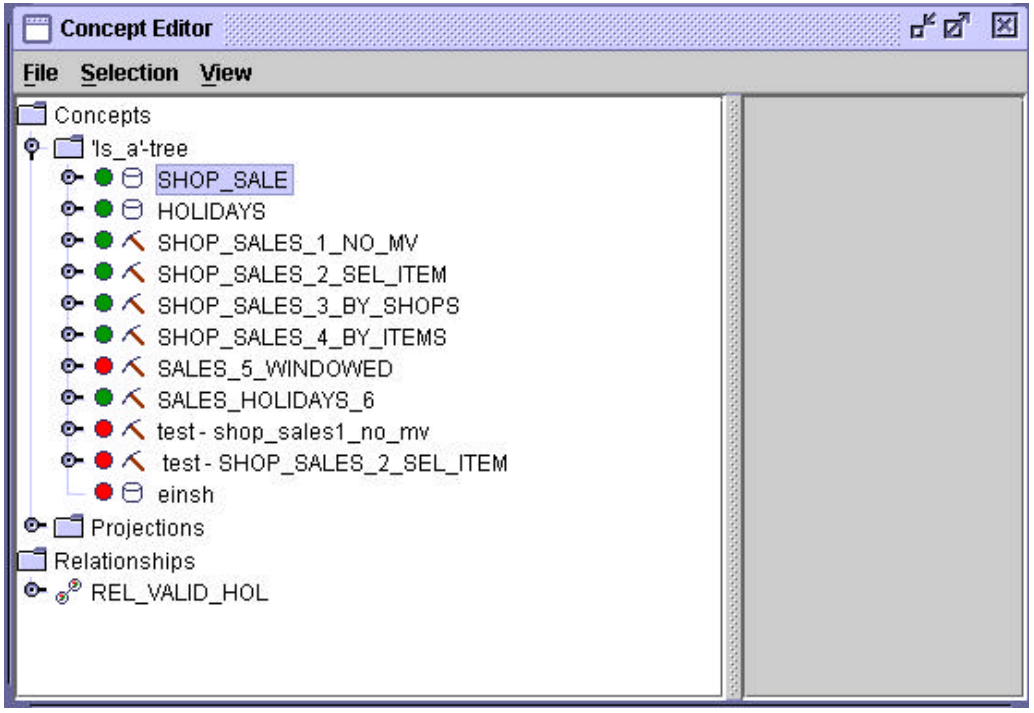
The Internet Case Base





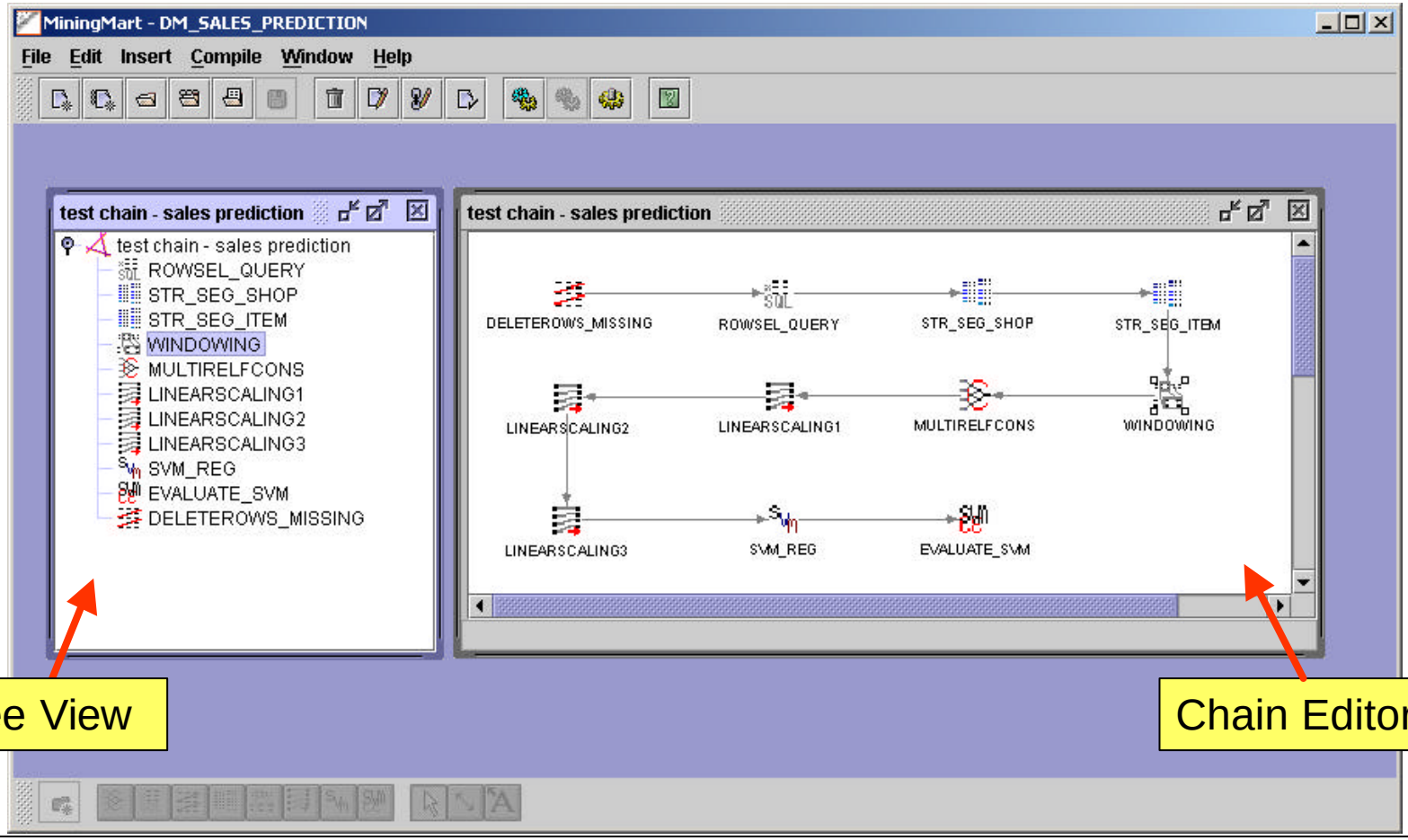
Demo

The Concept Editor

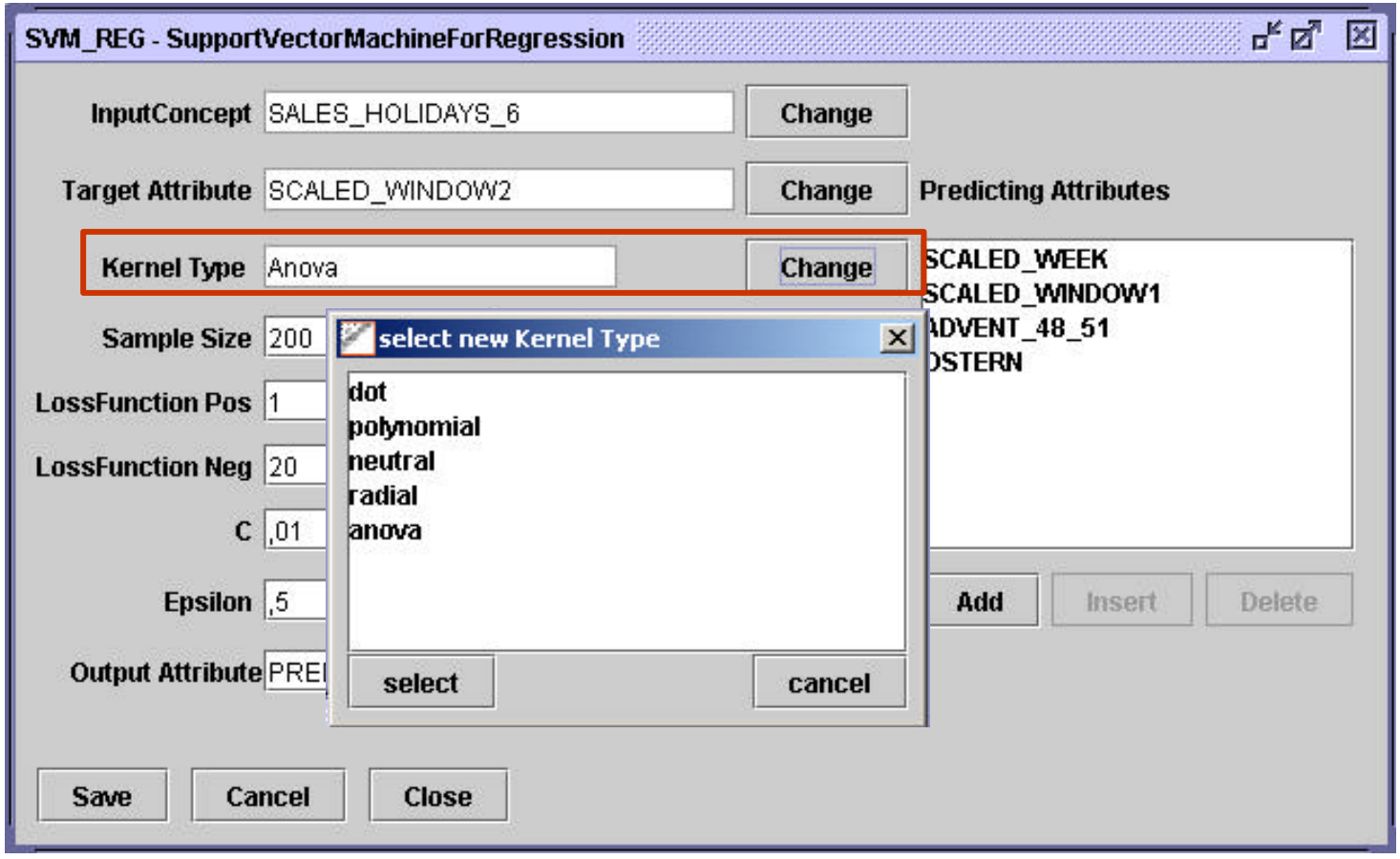


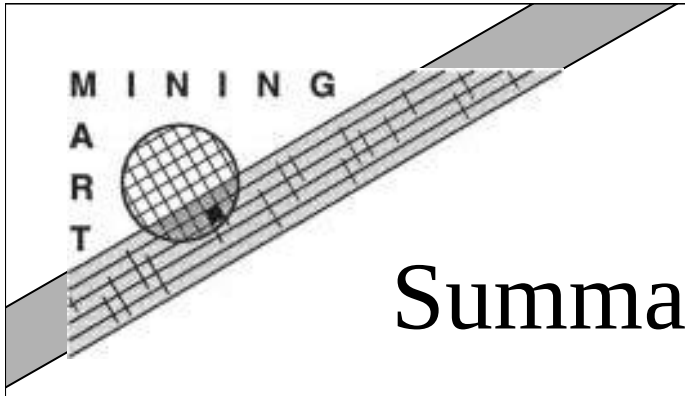
- Define and edit concepts and relations
- Mapping from concepts to relations of the database.

The Case Editor



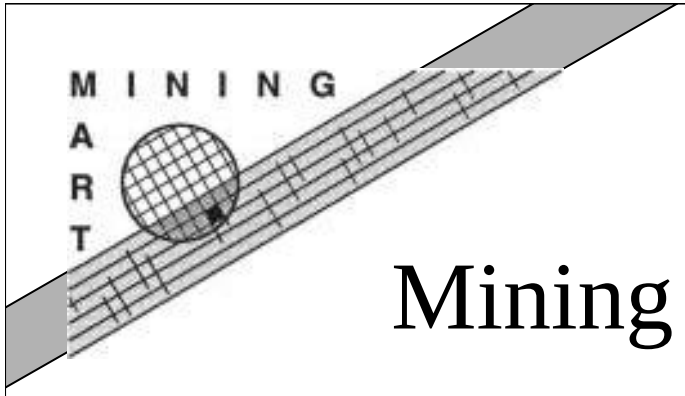
Setting up an SVM Step





Summary

- Mining Mart eases pre-processing:
  - Many operators are implemented in the database.
  - Validity and necessity of operator execution is checked.
- Mining Mart documents cases of successful data mining. These can be used as blueprints and easily be adapted to similar data.
- Meta-data are made operational by the compiler.



Mining Mart Partners

- Univ. Dortmund,
- Univ. Piemonte del Avogadro (DISTA),
- Univ. Economics Prague,
- Perot Systems Netherland,
- Fraunhofer Gesellschaft (AIS),
- SwissLife,
- Telecom Italia Laboratory,
- National Institute of Telecommunication Warsaw

You may use the Mining Mart system.

You may contribute to the public case base.

Only conceptual and case model, please.

www-ai.cs.uni-dortmund.de/MMWEB/