Analyzing Churn of Customers

Marco Richeldi
Alessandro Perrucci
TELECOM ITALIA LAB
Via G. Reiss Romoli 274, 10148 Torino – Italy
{Marco.Richeldi, Alessandro.Perrucci@tilab.com}

Agenda

- **Churn management in Telcos**
- **A Churn Analysis system for wireless network services**
- **The MiningMart solution**
- **Conclusions**

Business Scenario: Customer Orientation is key for Telcos

- **Most Telcos' products and services: <u>commodities</u> (no longer relevant for competitive advantage)**
- **Telcos: evolving a process-oriented organization (CRM, SCM)**
  - **CRM application architectures: integrate front-office / back-office applications**
  - **Through 2005, telcos: mktg automation applications + call centers => unified customer interaction frameworks**
- **Europe: Analytical CRM solutions market growing rapidly**
  - **CAGR: ~ 50% (from $0.5 billion in 1999 to $3.5 billion in 2004)**
- **Telco's investment in Analytical CRM moderate due to investments in 2.5G and 3G (UMTS) technology, but relevant**

Churn management: a bottom line issue

- **Attracting thousands of new subscribers is worthless if an equal number are leaving**
- **Minimizing customer churn provides a number of benefits, such as:**
  - **Minor investment in acquiring a new customer**
  - **Higher efficiency in network usage**
  - **Increase of added-value sales to long term customers**
  - **Decrease of expenditure on help desk**
  - **Decrease of exposure to frauds and bad debts**
  - **Higher confidence of investors**

Churn management: scooping the problem (1)

- **Churn can be defined and measured in different ways**
  - **"Absolute" Churn. number of subscribers disconnected, as a percentage of the subscriber base over a given period**
  - **"Line" or "Service" Churn. number of lines or services disconnected, as a percentage of the total amount of lines or services subscribed by the customers**
  - **"Primary Churn". number of defections**
  - **"Secondary Churn". drop in traffic volume, with respect to different typology of calls**

Churn management: scooping the problem (2)

- **Measuring churn is getting more and more difficult**
  - **Growing tendency for Business users to split their business between several competing fixed network operators**
  - **Carrier selection enables Residential customers to make different kind of calls with different operators**
  - **Carrier pre-selection and Unbundling of the Local Loop makes it very difficult to profile customers according to their “telecommunication needs”**
- **Other frequent questions for Fixed Network Services**
  - **What if a customer changes his type of subscription, but remains in the same telco? What if the name of a subscriber changes? What if he relocates?**

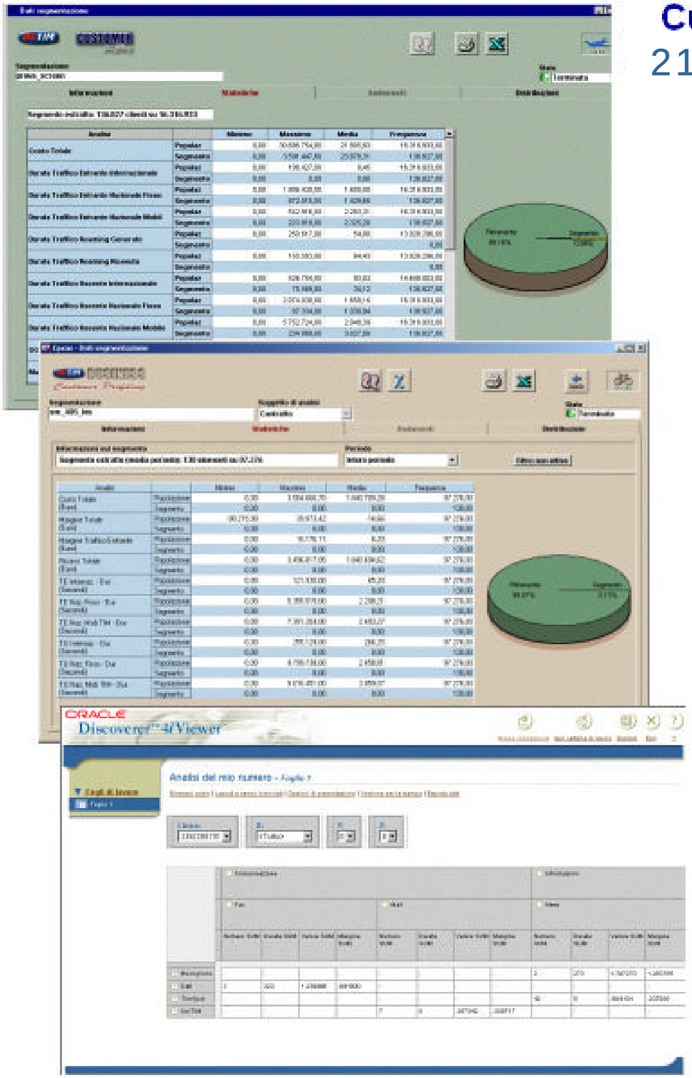
The case study: Churn Analysis for wireless services

- **The framework**
  - **A major Italian network operator willing to establish a more effective process for implementing and measuring the performance of loyalty schemes**
- **Objectives of the “churn management” project**
  - **Building a new corporate Customer Data Warehouse aimed to support Marketing and Customer Care areas in their initiatives**
  - **Developing a Churn Analysis system based upon data mining technology to analyze the customer database and predict churn**

Business understanding

- **Sponsors**
  - **Marketing dept., IT applications, IT operations**
- **Analysis target**
  - **Residential Customers, subscriptions**
- **Churn measurement**
  - **Absolute, primary churn**
- **Goal:**
  - **Predict churn/no churn situation of any particular customer given 5 months of historical data**

Solution scope



Customer Profiling Consumer:
21 millions of residential customers

Usage patterns analysis of Voice Services by single subscriber line

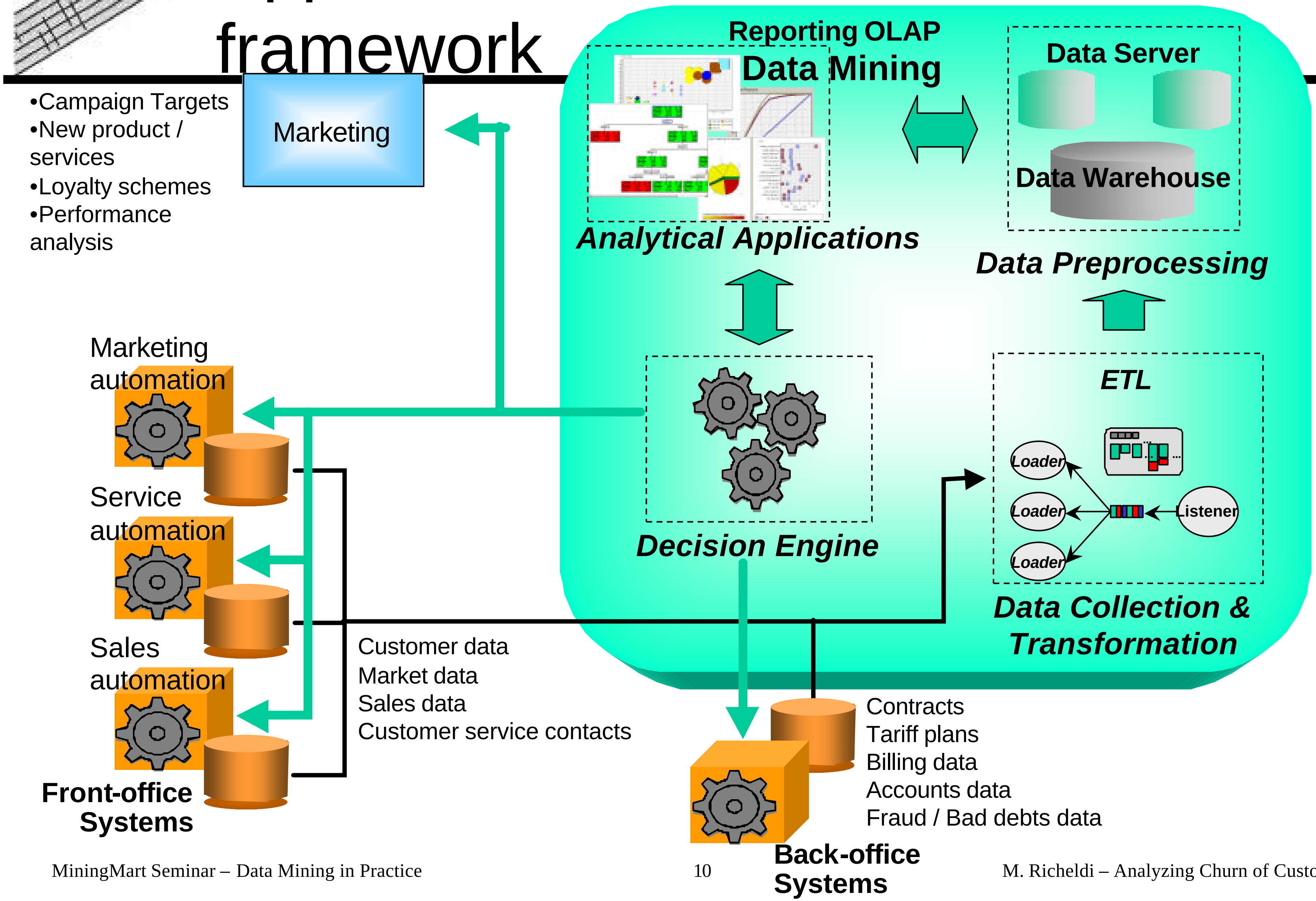
Customer Profiling Business:
2 millions of business customers

Usage patterns analysis of Voice Services by subscriber line, contract, company, etc.

Customer Profiling VAS:
23 millions of customers

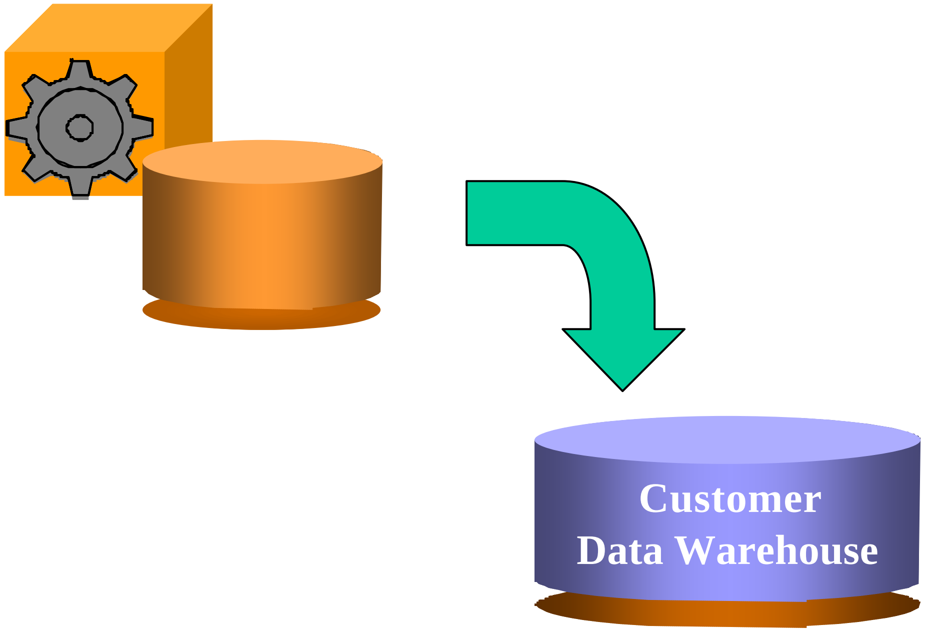
Usage patterns analysis of VAS by single subscriber line

Application framework



Data understanding

13 operational systems



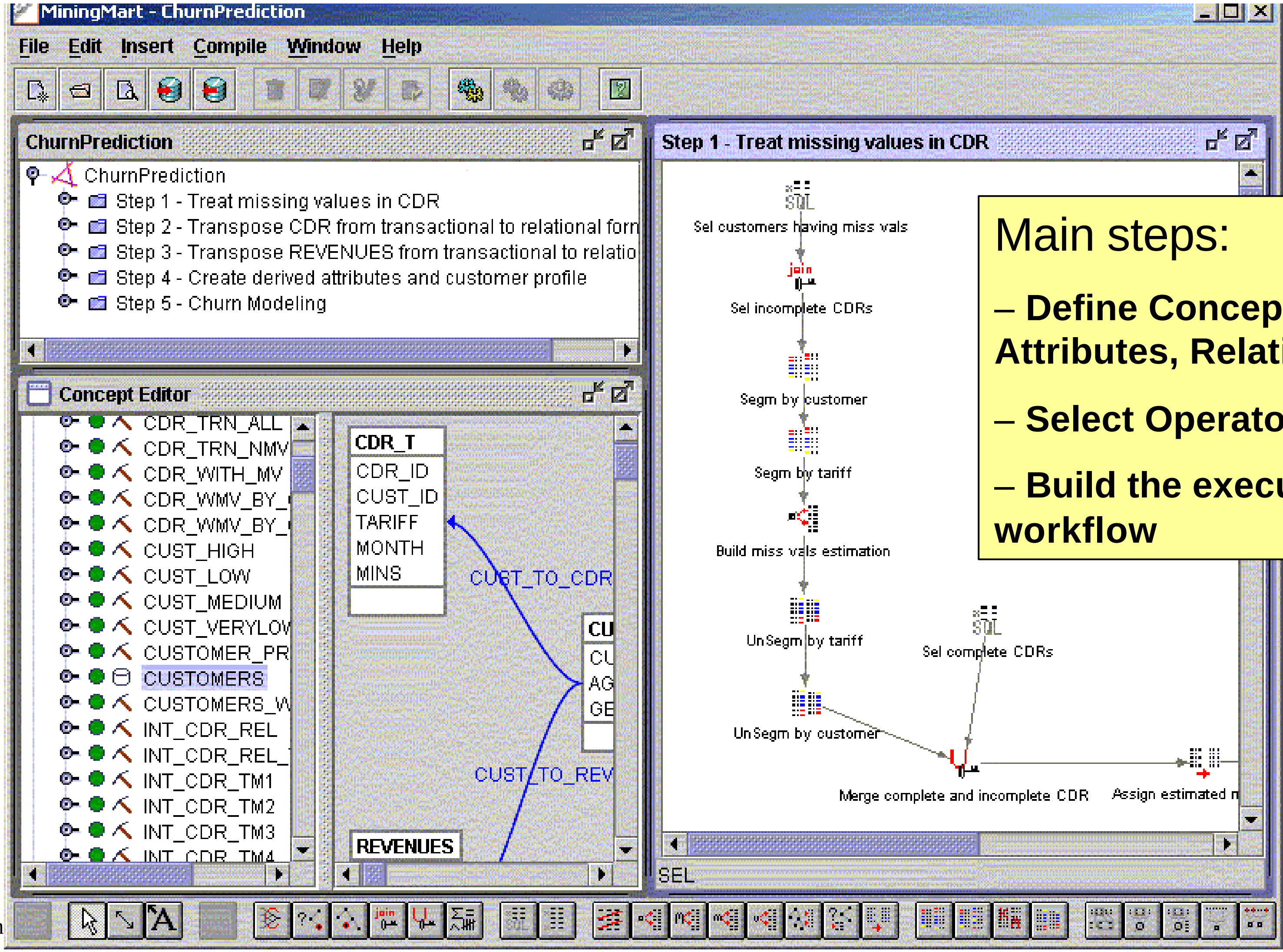
- More than 500 indicators per customer
- Extraction delay: 2 months
- Loading: on a monthly basis
- Size: 1.5 Tb

<u>Input Data</u>

- Customer demographics
  Basic customer information
- Service Profile
  Products/services purchased by each customer.
- Tariff plans
  Details of the tariff scheme in use
- Extra service information
  - Special plans / rates
  - Service bundles
- Call data aggregated by month
- Billing data aggregated by month

- *Complaint information*
- *Fraud and bad debts data*
- *Customer service contacts*
- *Sales force contacts*
- *Market data*

Modeling with Mining Mart



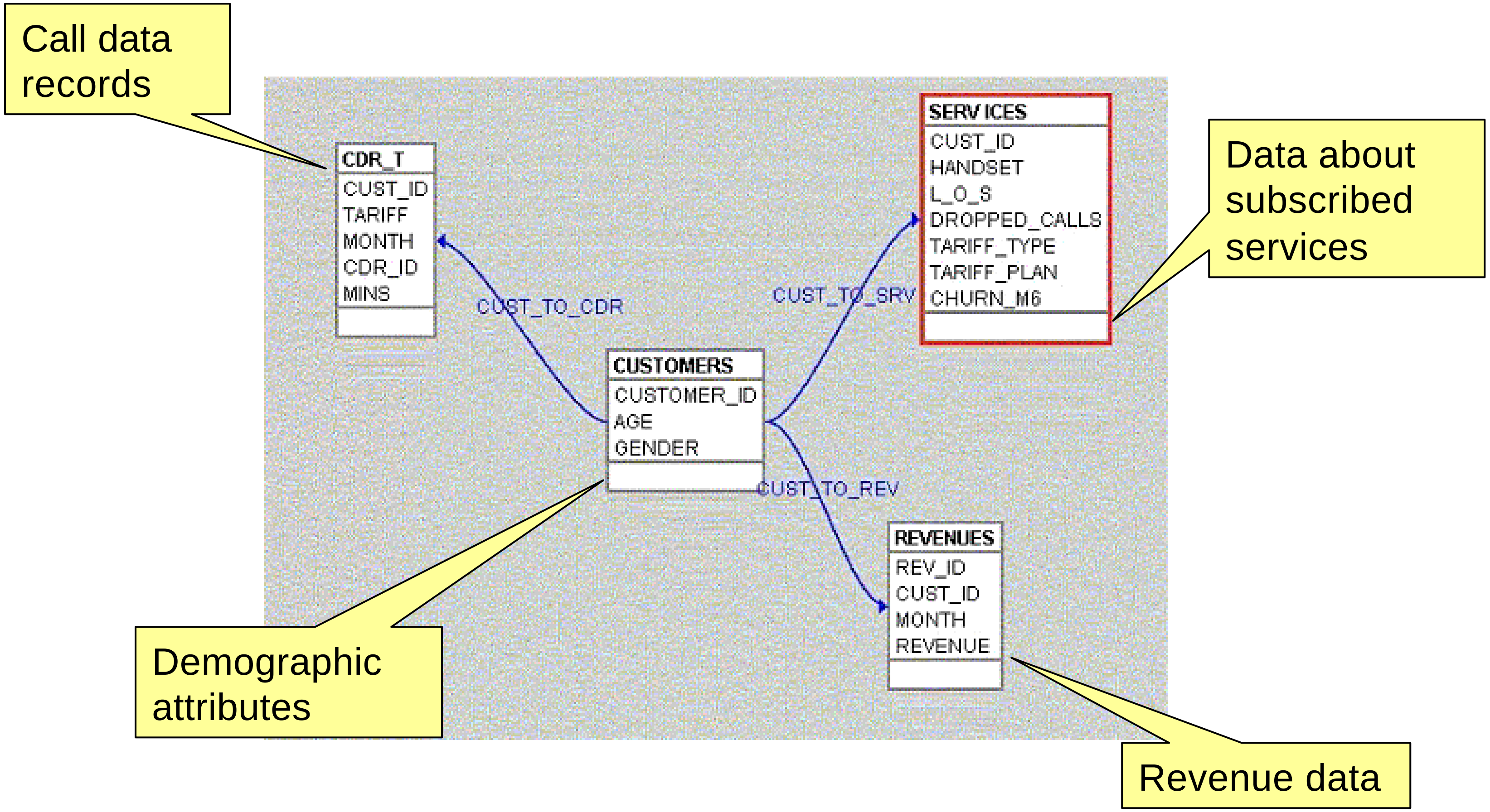
Main steps:

– **Define Concepts, Attributes, Relationships ...**

– **Select Operators**

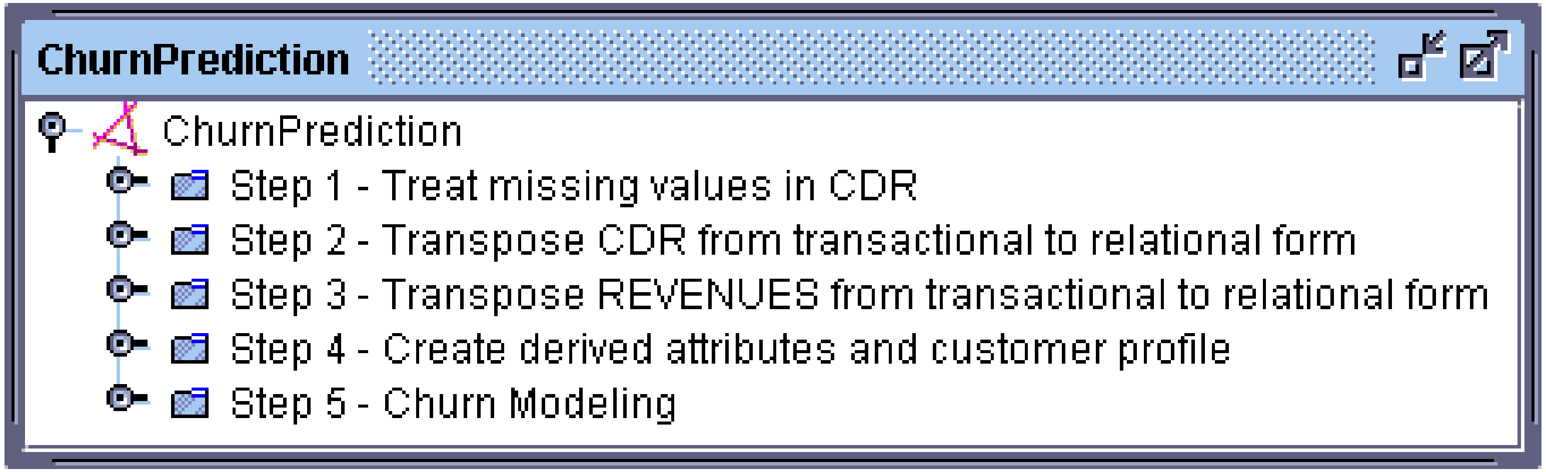
– **Build the execution workflow**

Concepts, Attributes, Relationships

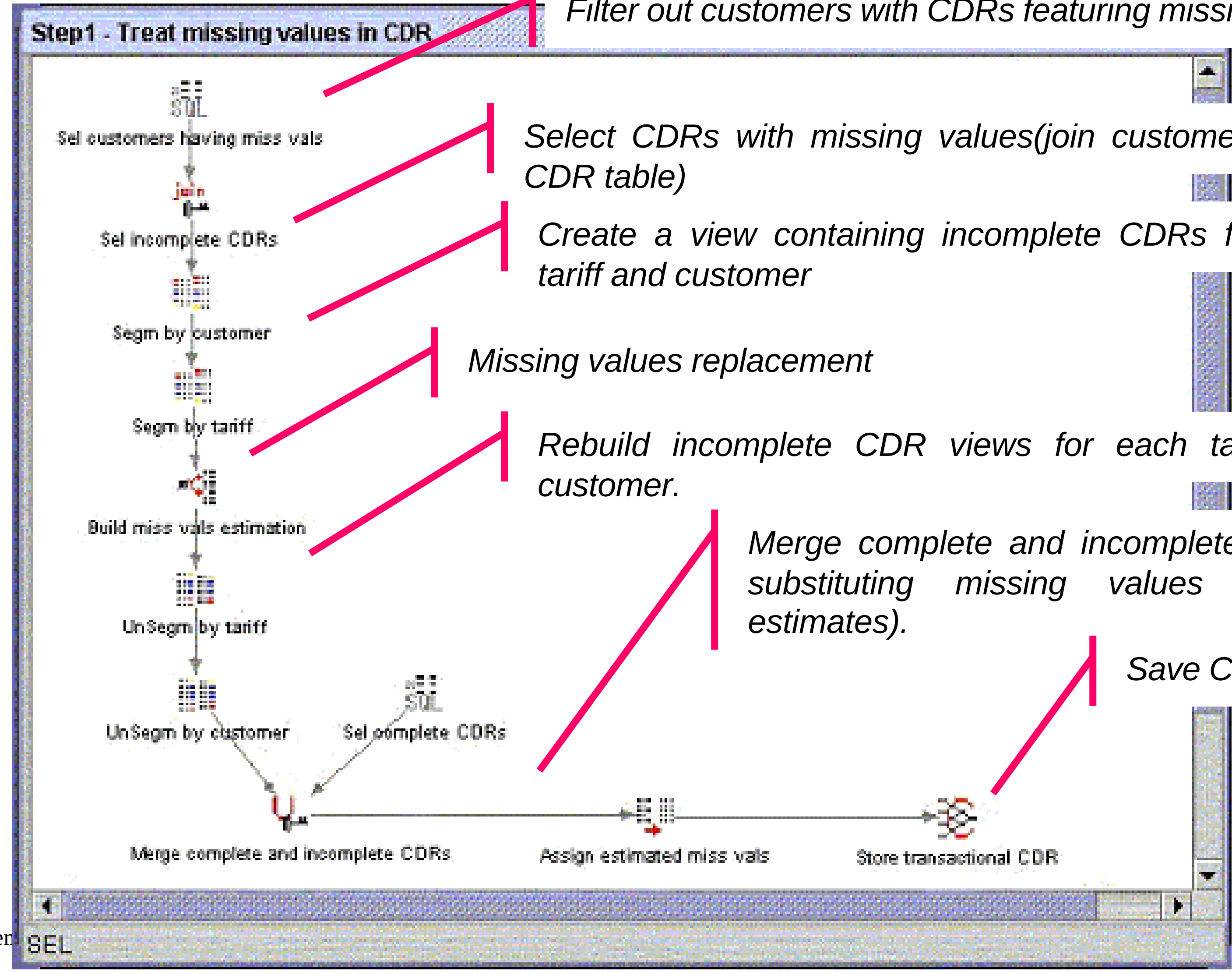


Pre-processing chains

The data mining process has been divided into five tasks as follows:



Handle missing values in CDRs



Filter out customers with CDRs featuring missing values

Select CDRs with missing values(join customers with CDR table)

Create a view containing incomplete CDRs for each tariff and customer

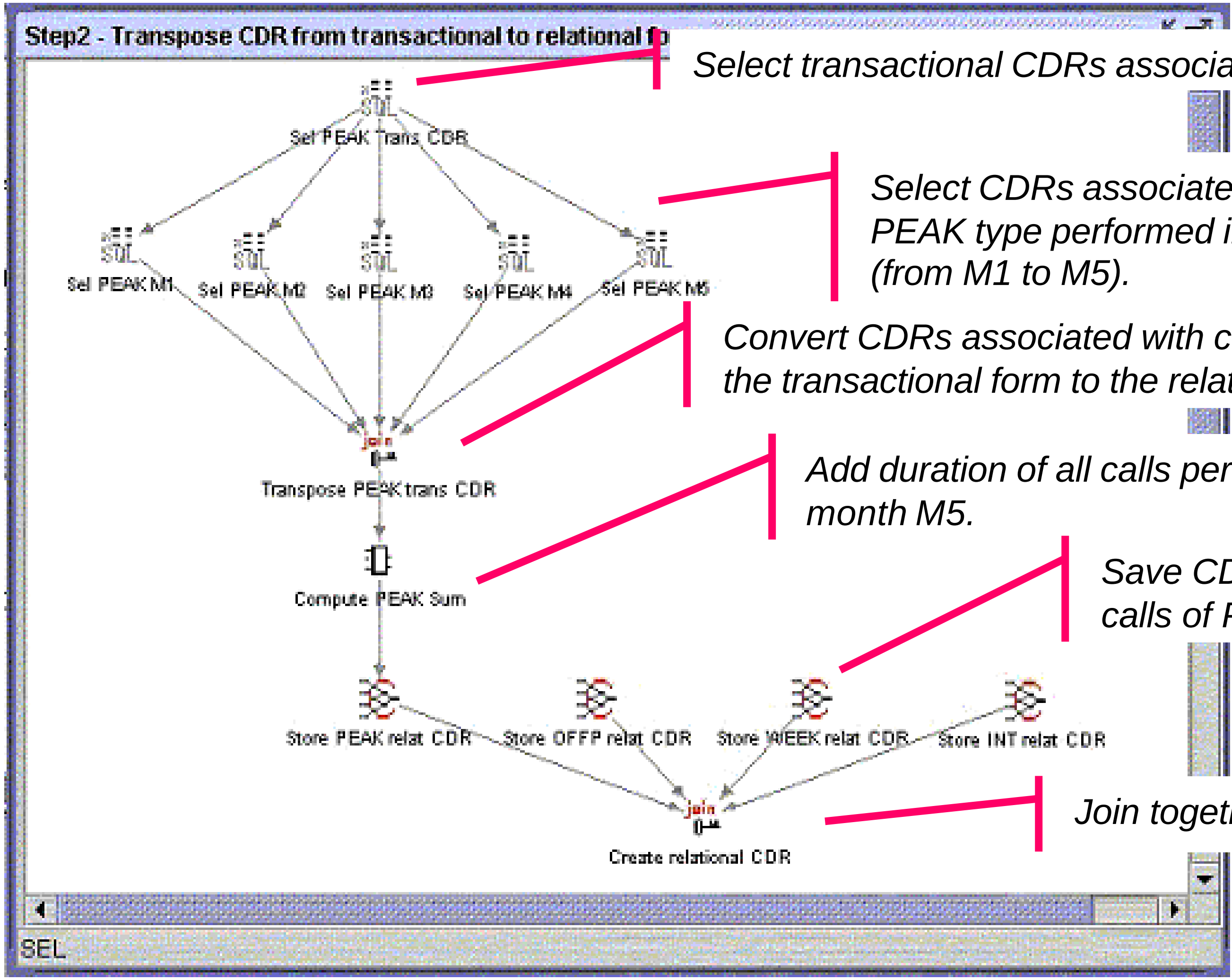
Missing values replacement

Rebuild incomplete CDR views for each tariff and customer.

Merge complete and incomplete CDRs (by substituting missing values with their estimates).

Save CDRs

Transpose CDR from transactional to relational form



Select transactional CDRs associated with calls of PEAK type

Select CDRs associated with calls of PEAK type performed in a specific month (from M1 to M5).

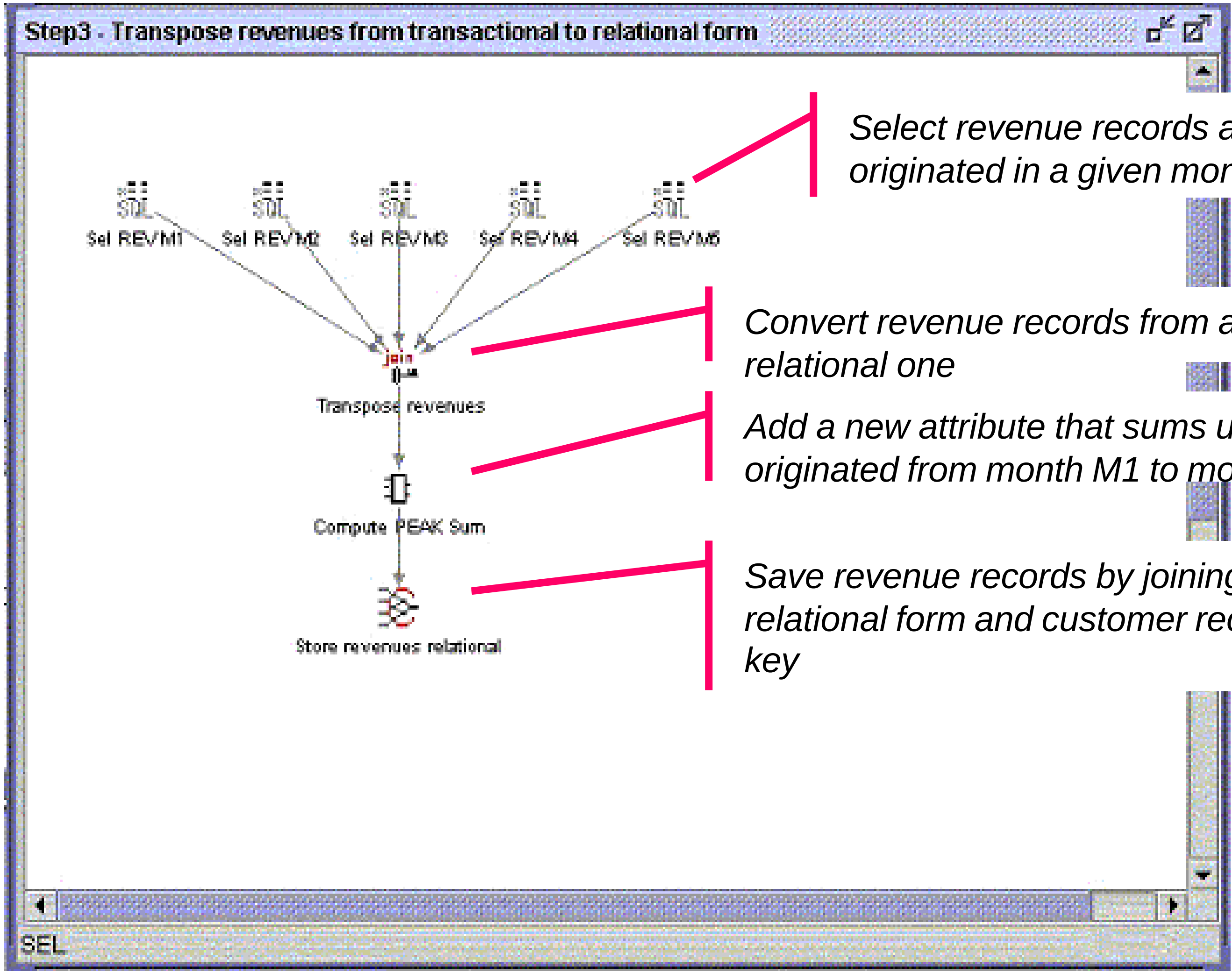
Convert CDRs associated with calls of PEAK type from the transactional form to the relational one

Add duration of all calls performed from month M1 to month M5.

Save CDRs associated with calls of PEAK type

Join together all CDRs

Transpose REVENUES from transactional to relational form



Select revenue records associated with calls originated in a given month (from M1 to M5)

Convert revenue records from a transactional form into a relational one

Add a new attribute that sums up the revenue of calls originated from month M1 to month M5

Save revenue records by joining revenue records in relational form and customer records by customer key

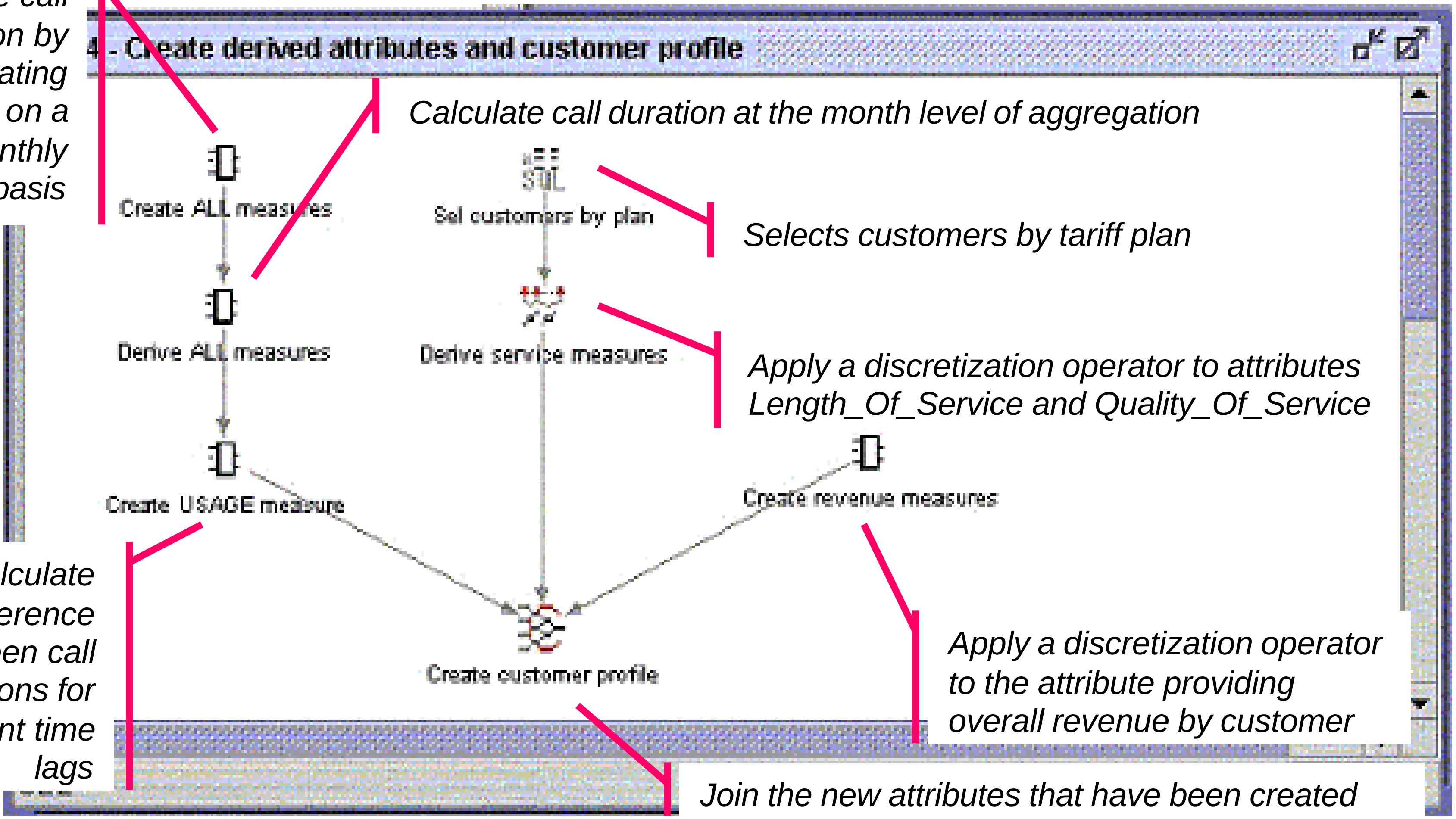
Create derived attributes and customer profile

Calculate call duration by aggregating CDRs on a monthly basis

Calculate call duration at the month level of aggregation

Selects customers by tariff plan

Apply a discretization operator to attributes Length_Of_Service and Quality_Of_Service

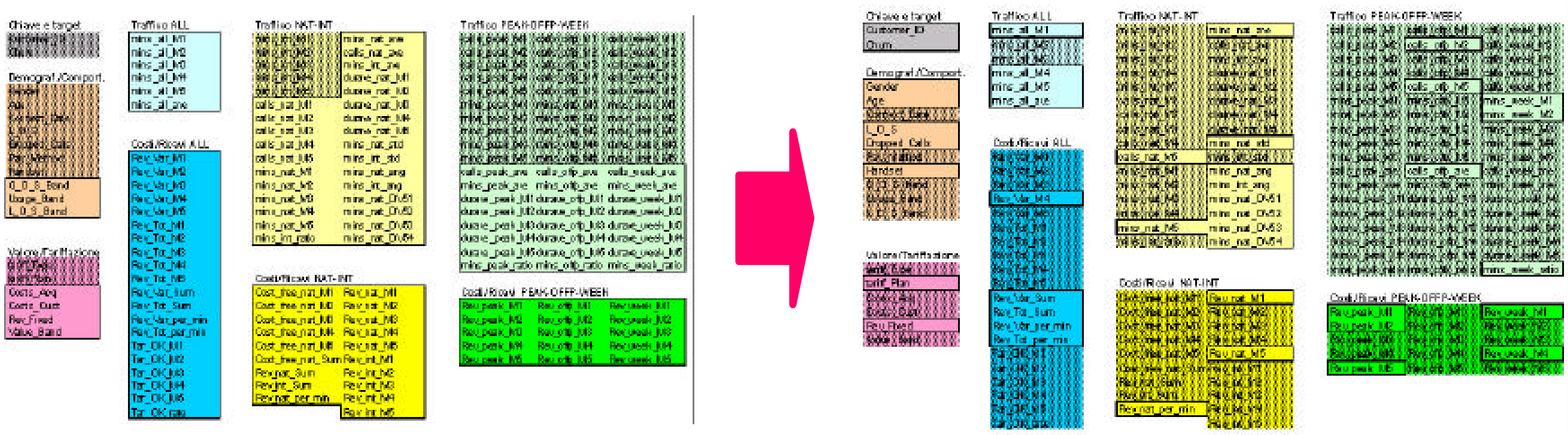


Calculate difference between call durations for different time lags

Apply a discretization operator to the attribute providing overall revenue by customer

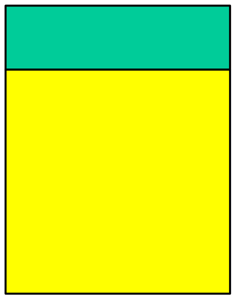
Join the new attributes that have been created

Construction stage output



Data Construction

Feature Selection

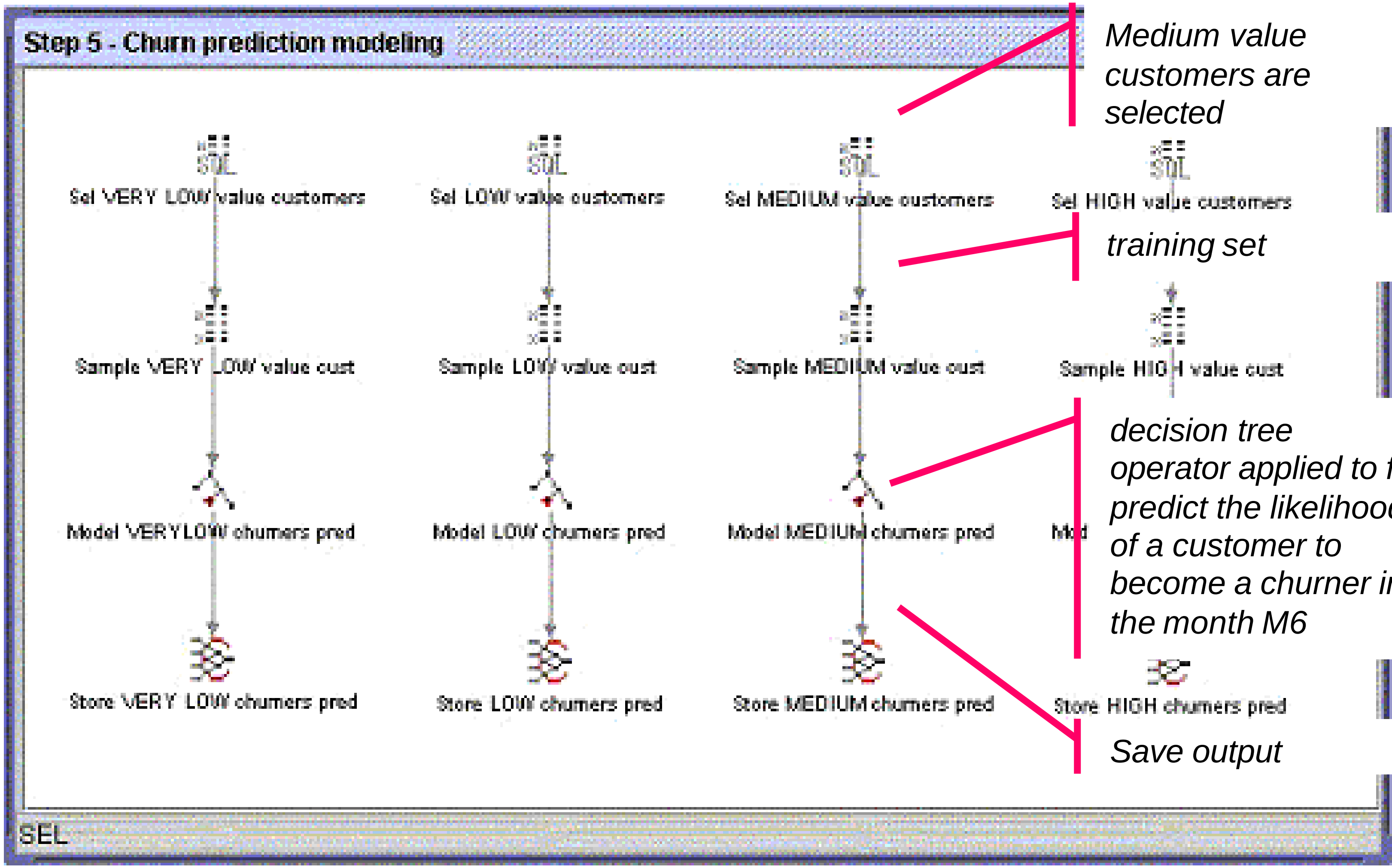


16 Raw attributes

45 Derived attributes

Churn modeling chain

4 Predictive models, one for each customer segment



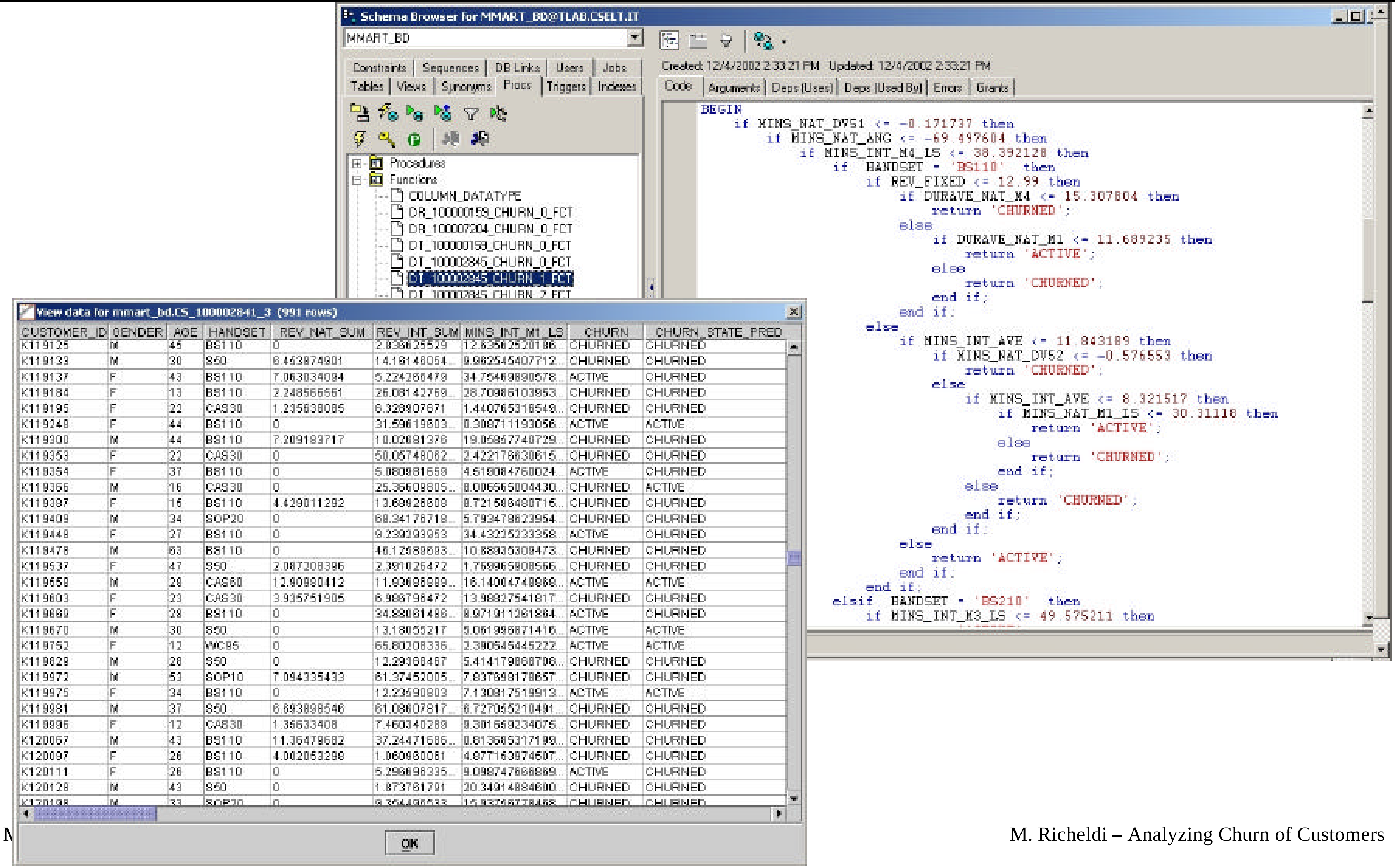
Medium value customers are selected

training set

decision tree operator applied to fit predict the likelihood of a customer to become a churner in the month M6

Save output

The resulting model

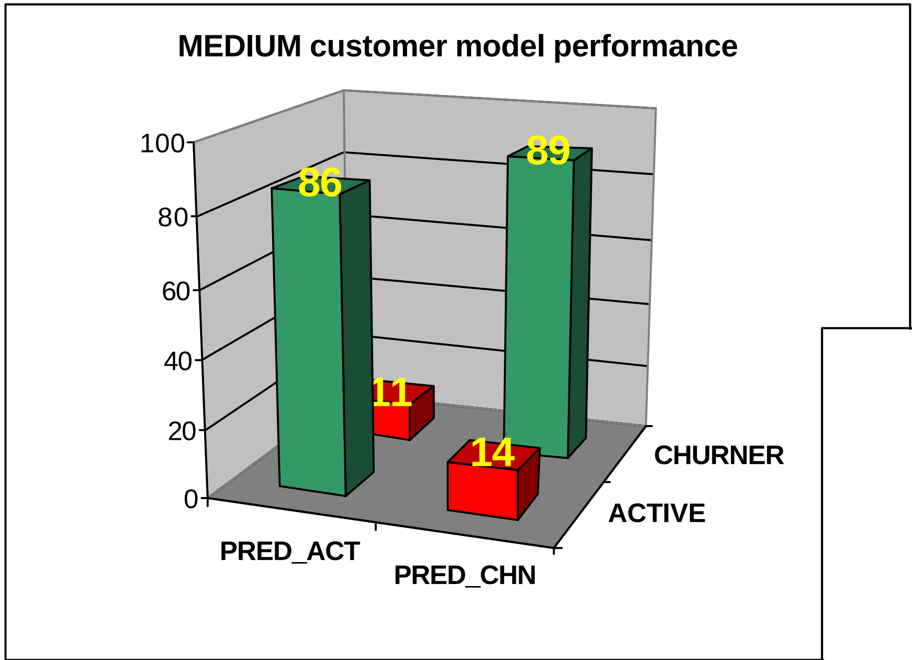


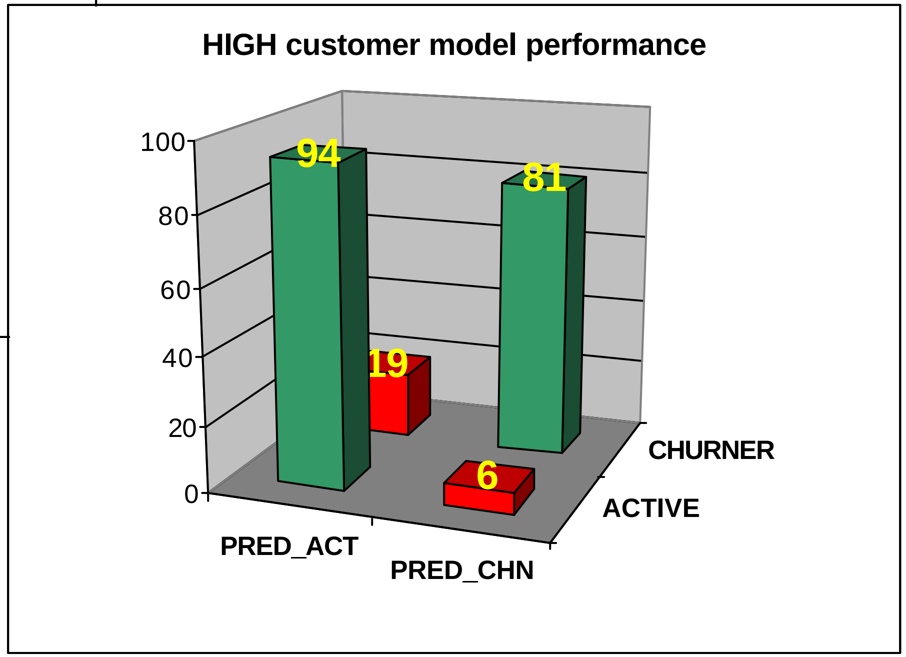
The decision tree - excerpt

```
BEGIN
        if ALL_M5 <= 483.526001 then
                if  HANDSET = 'ASAD1'  then
                        return 'ACTIVE';
                elsif  HANDSET = 'ASAD9'  then
                        if PEAK_M1 <= 139.363846 then
                                if OFFP_M3 <= 106.607796 then
                                        return 'ACTIVE';
                                else
                                        return 'CHURNED';
                                end if;
                        else
                                return 'CHURNED';
                        end if;
                elsif  HANDSET = 'S50'  then
                        if PEAK_M3 <= 144.418304 then
                                return 'CHURNED';
                        else
                                if REV_SUM <= 294.393341 then
                                        if  L_O_S_BAND = 'HIGH'  then
                                                return 'ACTIVE';
                                        elsif  L_O_S_BAND = 'MEDIUM'  then
                                                return 'ACTIVE';
```

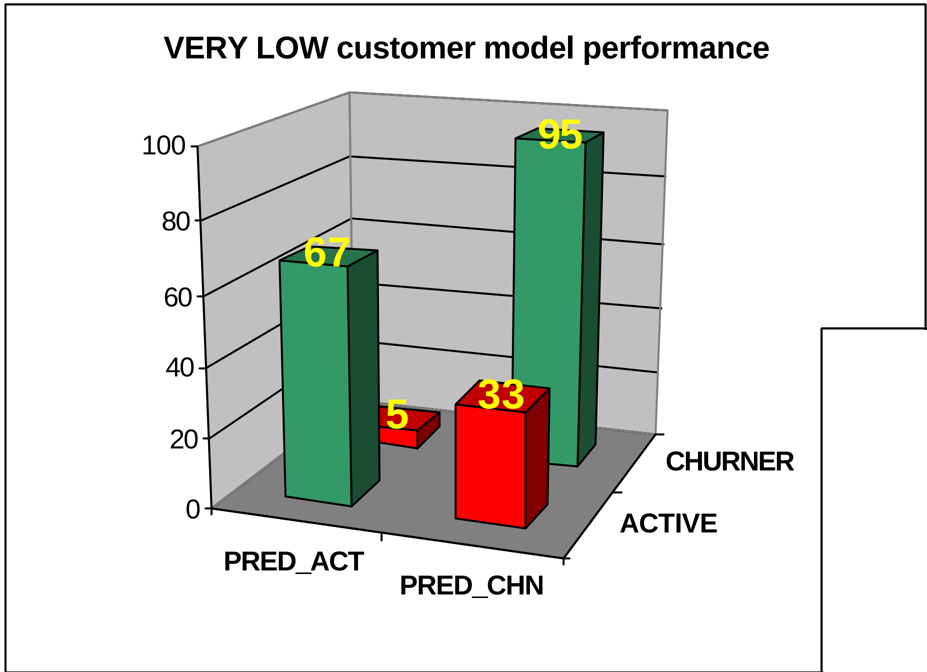
Predictive performance

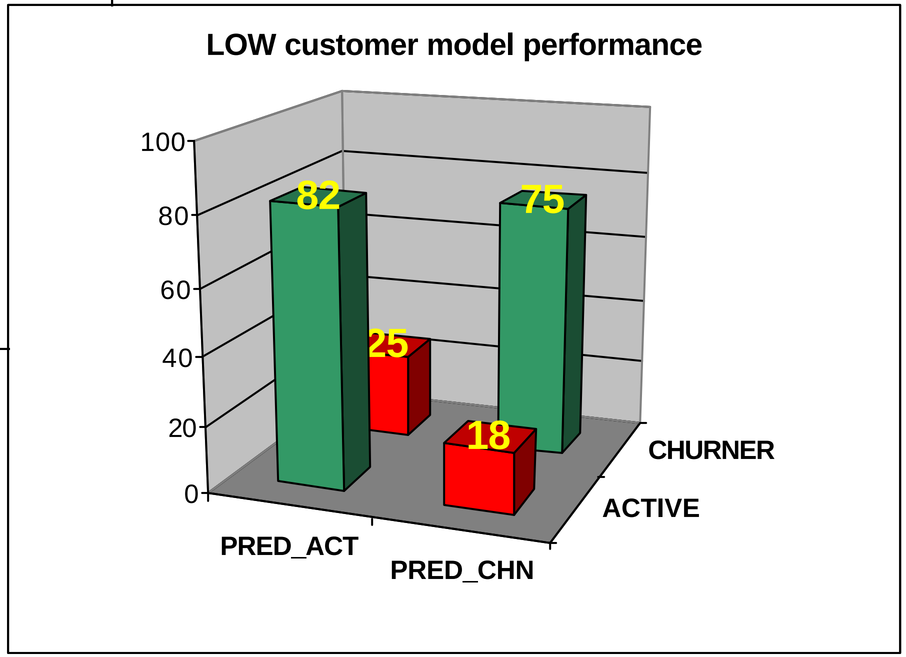
Training / test set: 70% / 30%





Predictive performance





Execution Time

Data Set Size (num. records)	Pre-processing Time (mins)	Modeling Time (hours)
8,000	17.3	4.3
800,000	27.8	13.5

Mining Mart evaluation

- **Usability**
- **Mining process speed-up**
- **Mining process quality**
- **Integration (into the business processes)**

Usability

- **Human Computer Interface is user-friendly and effective. Few steps required to implement any data mining process**
- **Interface quality compares to the ones of leading commercial tools (SPSS, SAS). Improves on IBM Intelligent Miner's interface with respect to a number of features**
- *Suggestions for future work*
  - **Definition of concepts can be further simplified (db attributes defined by directly editing table column names)**

Mining process speed-up

- **Preprocessing operators show quite good scalability on large data set:**
  - **MMart leverages Oracle scalability when carrying out preprocessing tasks. Overhead due to parsing of operators is negligible (unless for very small datasets)**
  - **Modeling operators are not optimized**
- **Processing chains can be quickly tested during chain set-up**
- **Multistep and loopable operators enable users to define parallel mining tasks consistently and effectively**
- **Processing chains can be saved an restored, allowing versioning**

Mining process speed-up

- **Less trials required to develop the data mining solution**
  - **Operator constraints drive unskilled users to build correct and effective analytical applications**
  - **Users achieve a better understanding of data structure by:**
    - Browsing source and processed data
    - Computing descriptive statistics
  - **Operator chains makes it possible to implement data mining best-practices**
- *Suggestions for future work*
  - **Improve graphical investigation features**
  - **Improve workgroup enabling features: multiple users capabilities, definition of user roles and access rights**

Mining process quality

- **Best practices may be easily pre-packaged**
- **Libraries of data mining applications may be developed and customized to satisfy new business requirements**
- **MMart framework ensures chain consistence and correctness, avoiding potential conceptual mistakes**
- **Users can focus their effort on modeling tasks rather than on preprocessing tasks**
- **Domain knowledge improves and extend usability of pre-packaged data mining applications**

Integration

- **The Mining Mart system may be integrated into the Analytical CRM platform as the analytical extension of either the enterprise data warehouse or the business-oriented data marts**

Conclusions

- **Speed up for some preprocessing tasks increased by 50% at least**
- **Power users may find Mining Mart as much easy to use as the leading commercial dm platforms**
- **It enables building libraries of predefined data mining applications that can be easily modified**
- **MMart guarantees the highest scalability, since it exploits leading commercial db tools features**
- **Quality of data mining output increases as the number of preprocessing trials decrease in number**
- **Bottom line: <u>Mining Mart supports efficiently and effectively the preprocessing stage of a data mining process</u>**