



THE
POWER
TO KNOW.



**THE
POWER
TO KNOW.®**

Rapid Predictive Modeler

23.04.2010
Stefan Ahrens

Copyright © 2008, SAS Institute Inc. All rights reserved.



THE
POWER
TO KNOW.

Vorhersagemodellierung als Herausforderung

- Stichprobenziehung / Over-Sampling?
- Variablenelektion?
- Umgang mit fehlenden Werten und Ausreißern?
- Notwendigkeit von Variablentransformationen?
- Wahl des richtigen Modellansatzes?
- Optimierung der jeweiligen Modellparameter?
- Modellbewertung (Auswahl des besten Modells)?
- ...

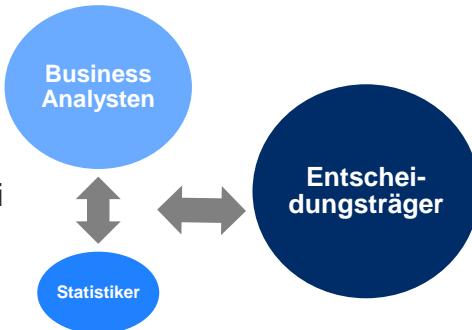
Copyright © 2008, SAS Institute Inc. All rights reserved.



Copyright © 2008, SAS Institute Inc. All rights reserved.

SAS Rapid Predictive Modeling

- **Automatisches** schnelles Erstellen von treffsicheren Vorhersagemodellen
- **Einfach** zu bedienende Engine für Modellierung und Berichtgenerierung
- Erhöht Autarkie von **Business Analysten** bei Modellierung
- Zugriff über **Enterprise Guide** oder **Microsoft Office Add-In**
- Kein Ersatz für **Enterprise Miner**, sondern sinnvolle Ergänzungskomponente!



Copyright © 2008, SAS Institute Inc. All rights reserved.

A screenshot of the SAS Enterprise Guide interface. The window title is "SAS Enterprise Guide - Projekt_1.egg". The menu bar shows "Extras", "Add-In", "Add-In Manager...", and "Rapid Predictive Model...". A tooltip in the bottom left corner says: "Aktuell aktive Datenquelle wird ausgewählt, ansonsten Aufforderung zur Auswahl der Datenquelle". Another tooltip in the bottom right corner says: "Zugriff über Menüleiste im Enterprise Guide oder im Microsoft Office Add-In (geplant als eigener Task in der Gruppe Data Mining)". The main pane displays a table with columns: VALUE, REASON, JOB, Y0J, DEROG, and DELINQ. The table contains several rows of data. The bottom status bar shows "Fertig" and "Verbindung: sasdemo, gersta.sas.com".

SAS Enterprise Guide - Projekt_1.egg

Projektbaum

- Prozessfluss
 - dm_hmeq_train
 - Einfache Häufigkeiten
 - Rapid Predictive Model

dm_hmeq_train

Eltern und sortieren Abfrage erstellen Daten > Beschreiben > Grafiken > Analysieren > Exportieren > Senden an

Rapid Predictive Model for: D:\Daten\Quellen\dm_hmeq_train.sas7bdat

Data Model Report Options Registration

Datenquelle: D:\Daten\Quellen\dm_hmeq_train.sas7bdat Anwendungsroutine: Ohne Bearbeiten...

Input Variables:

| | |
|---------|---------------------------------|
| Name | Modeling Roles: |
| CLAGE | Dependent Variable (Maximal: 1) |
| CLINO | BAD |
| DEBTINC | Frequency Count (Maximal: 1) |
| DELINQ | ID |
| DEROG | Excluded |
| JOB | |
| LOAN | |
| MORTDUE | |
| NINJ | |
| REASON | |
| VALUE | |
| Y0J | |

Run

Fertig Verbindung: sasdemo, gersta.ger.sas.com

Rollenzuweisung für Variablen:

- Abhängige Variable (Zielgröße)
- ID-Variable
- Variable mit Häufigkeitszählung
- Ausgeschlossene Variablen

→ alle anderen Variablen dienen als Kandidaten für Modell

SAS Enterprise Guide - Projekt_1.egg

Projektbaum

- Prozessfluss
 - dm_hmeq_train
 - Einfache Häufigkeiten
 - Rapid Predictive Model

dm_hmeq_train

Eltern und sortieren Abfrage erstellen Daten > Beschreiben > Grafiken > Analysieren > Exportieren > Senden an

Rapid Predictive Model for: D:\Daten\Quellen\dm_hmeq_train.sas7bdat

Data Model Report Options Registration

Model

Dependent variable: BAD Decisions and priors...

Modeling method: Basic

Advanced

Decision and Priors

Model > Decision and Priors

Event level: None

Prior probability: None

| Level | Data Count | Data Proportion | Priors |
|-------|------------|-----------------|--------|
| 0 | 0 | | |
| 1 | 0 | | |
| 2 | 0 | | |

Decision: None

| Actual Value | Decision Value |
|--------------|----------------|
| 0 | 0 |
| 1 | 0 |
| 2 | 0 |

Fertig

Je nach Wahl der Modellierungs-methode werden die Schritte zur Transformation und Selektion von Variablen und zur Modellfindung komplexer und intelligenter.

Optional können für Klassifikationsmodelle auch die A-priori-Wahr-scheinlichkeiten und Entschei-dungsvariablen spezifiziert werden.

SAS Enterprise Guide - Projekt_1.egg

Projektbaum

- Prozessfluss
- dm_hmeq_train
- Erstelle H
- Rapid Predictive Model

Rapid Predictive Model for: D:\Daten\Quellen\dm_hmeq_train.sas7bdat

Report

Report options

- Model Summarization
- Variable Ranking
- Cross Tabulations
- Classification Matrix
- Fit Statistics
- Lift Plot

JSON

| | JOB | Y0J | DEROG | DELINQ |
|---------|------|-----|-------|--------|
| Other | 10.5 | 0 | 0 | 0 |
| Other | 7 | 0 | 2 | |
| Other | 4 | 0 | 0 | |
| . | . | . | . | |
| Office | 3 | 0 | 0 | |
| Other | 9 | 0 | 0 | |
| Other | 5 | 3 | 2 | |
| Other | 11 | 0 | 0 | |
| Other | 3 | 0 | 2 | |
| Sales | 16 | 0 | 0 | |
| | 18 | . | . | |
| Office | 11 | 0 | 1 | |
| Other | 3 | 0 | 0 | |
| Mgr | 25 | 0 | 0 | |
| Other | 8 | 0 | 1 | |
| Other | . | 0 | 1 | |
| Mgr | 7 | 2 | 6 | |
| | 19 | . | . | |
| Other | 4.5 | 0 | 0 | |
| Office | 2 | 0 | 0 | |
| Other | 3 | 0 | 1 | |
| ProfExe | 5 | 1 | 0 | |
| Mgr | 12 | 0 | 0 | |
| Mgr | 22 | . | 2 | |
| Other | . | 0 | 0 | |
| Mgr | 12 | 0 | 0 | |
| Office | 4 | 0 | 0 | |
| | 18 | 0 | 0 | |
| Office | 10 | 1 | 2 | |

Run Save Abbrechen

Anwendungsroutinestatus

Anwendungsroutine Status

Hier wird der notwendige Umfang der zu generierenden Berichte festgelegt.

SAS Enterprise Guide - Projekt_1.egg

Projektbaum

- Prozessfluss
- dm_hmeq_train
- Erstelle H
- Rapid Predictive Model

Rapid Predictive Model for: D:\Daten\Quellen\dm_hmeq_train.sas7bdat

Options

Save Enterprise Miner Project Data

SAS App Server Folder

JSON

| | JOB | Y0J | DEROG | DELINQ |
|---------|------|-----|-------|--------|
| Other | 10.5 | 0 | 0 | 0 |
| Other | 7 | 0 | 2 | |
| Other | 4 | 0 | 0 | |
| . | . | . | . | |
| Office | 3 | 0 | 0 | |
| Other | 9 | 0 | 0 | |
| Other | 5 | 3 | 2 | |
| Other | 11 | 0 | 0 | |
| Other | 3 | 0 | 2 | |
| Sales | 16 | 0 | 0 | |
| | 18 | . | . | |
| Office | 11 | 0 | 1 | |
| Other | 3 | 0 | 0 | |
| Mgr | 25 | 0 | 0 | |
| Other | 8 | 0 | 1 | |
| Other | . | 0 | 1 | |
| Mgr | 7 | 2 | 6 | |
| | 19 | . | . | |
| Other | 4.5 | 0 | 0 | |
| Office | 2 | 0 | 0 | |
| Other | 3 | 0 | 1 | |
| ProfExe | 5 | 1 | 0 | |
| Mgr | 12 | 0 | 0 | |
| Mgr | 22 | . | 2 | |
| Other | . | 0 | 0 | |
| Mgr | 12 | 0 | 0 | |
| Office | 4 | 0 | 0 | |
| | 18 | 0 | 0 | |
| Office | 10 | 1 | 2 | |

Run Save Abbrechen Help

Anwendungsroutinestatus

Anwendungsroutine Status

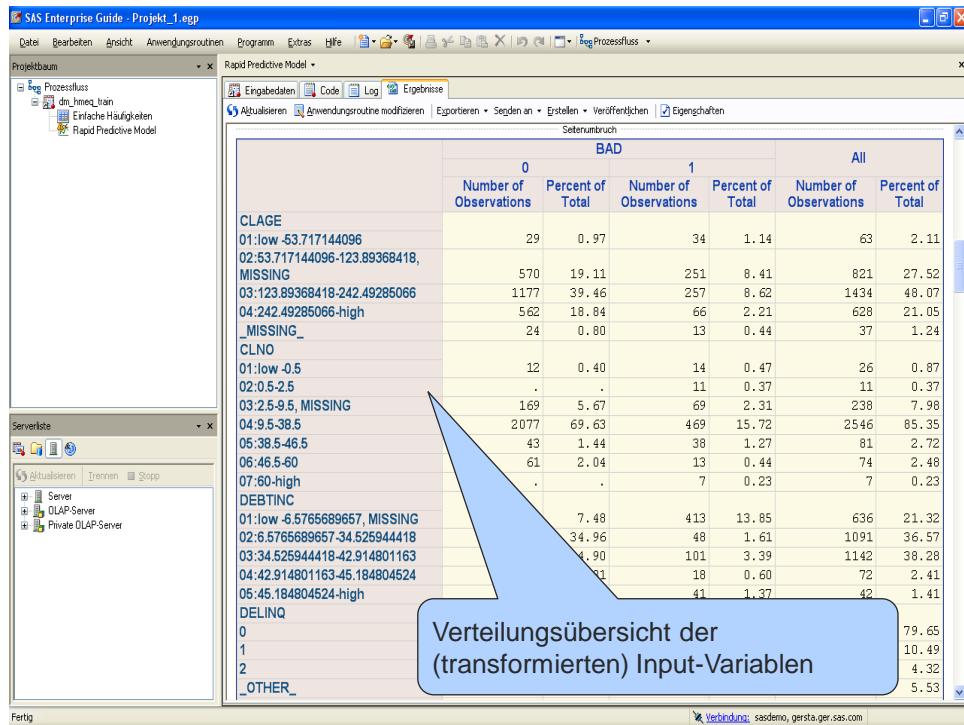
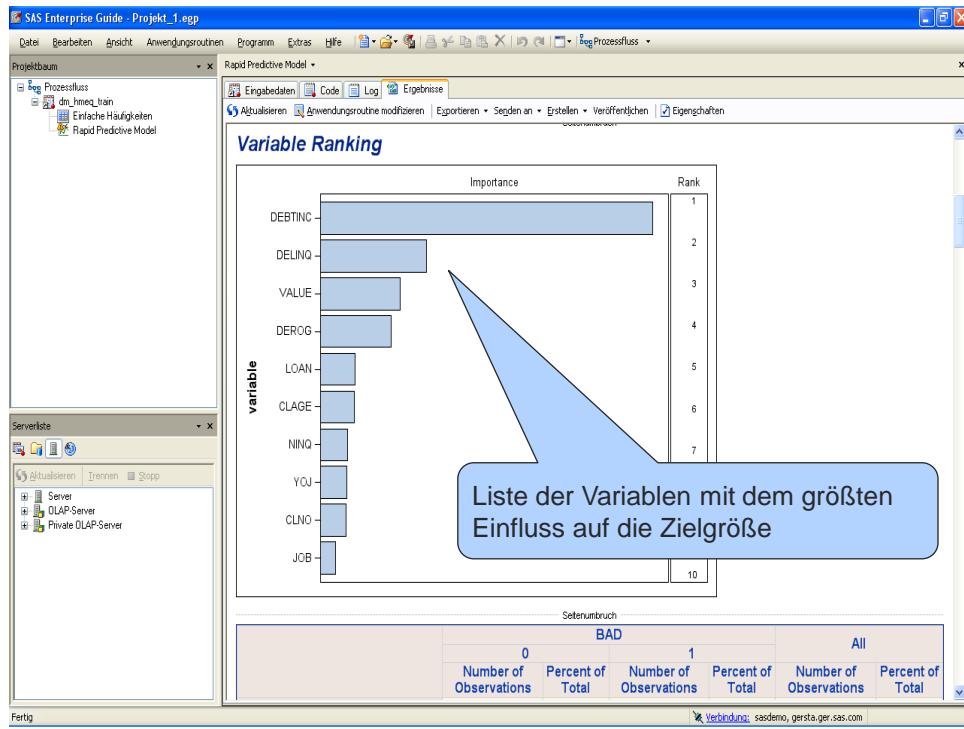
Warteschlange Server

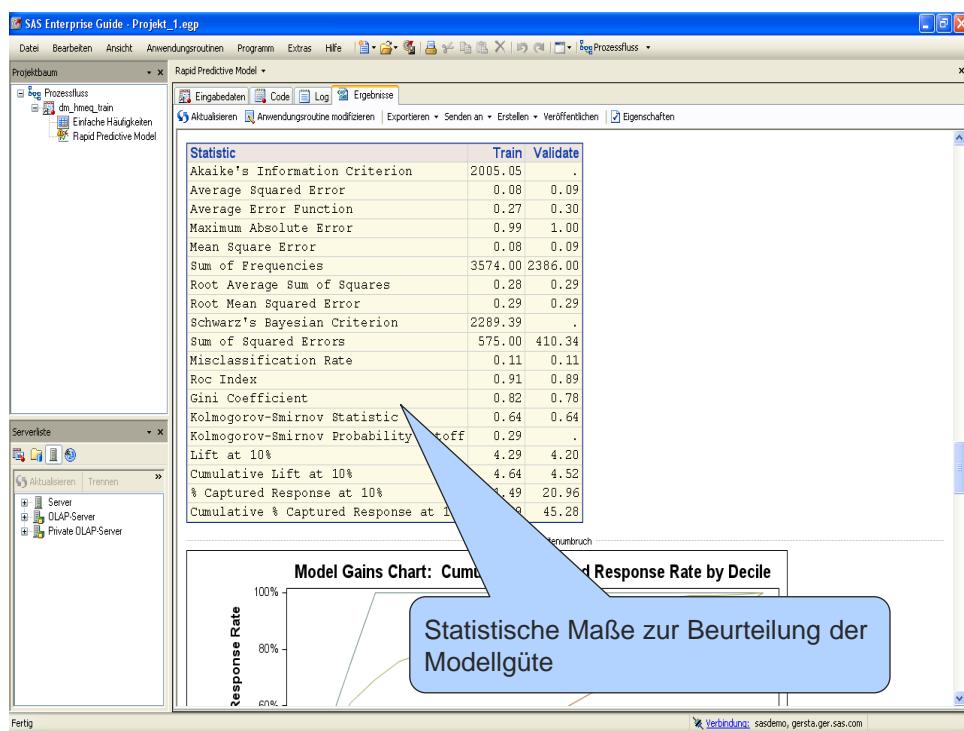
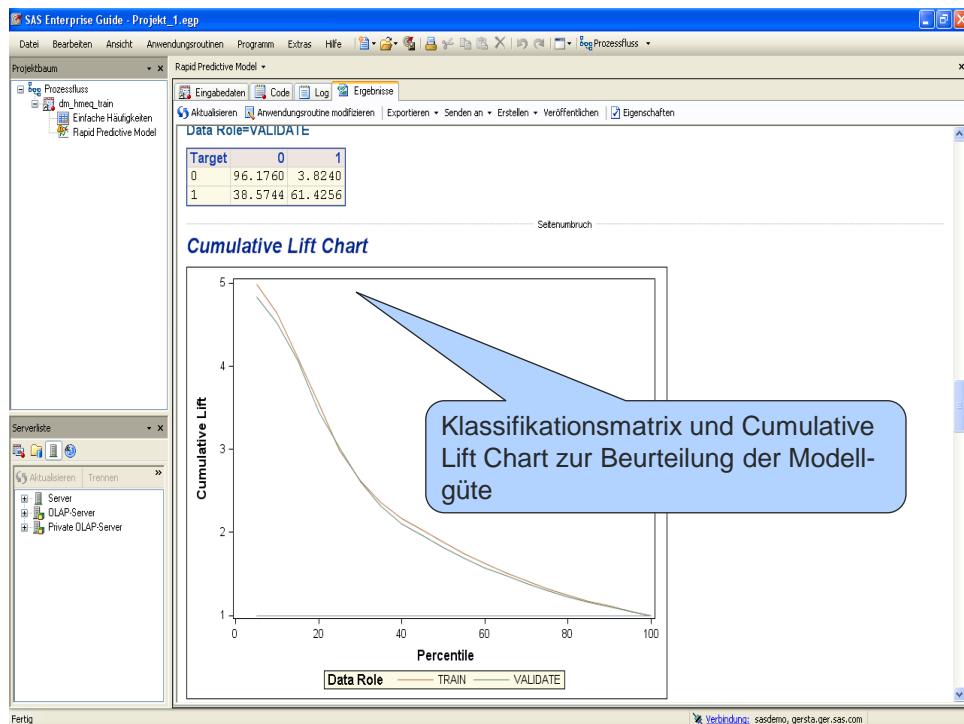
Hier kann ausgewählt werden, wo der im Hintergrund generierte Enterprise Miner Prozessfluss als Projekt gespeichert werden soll.

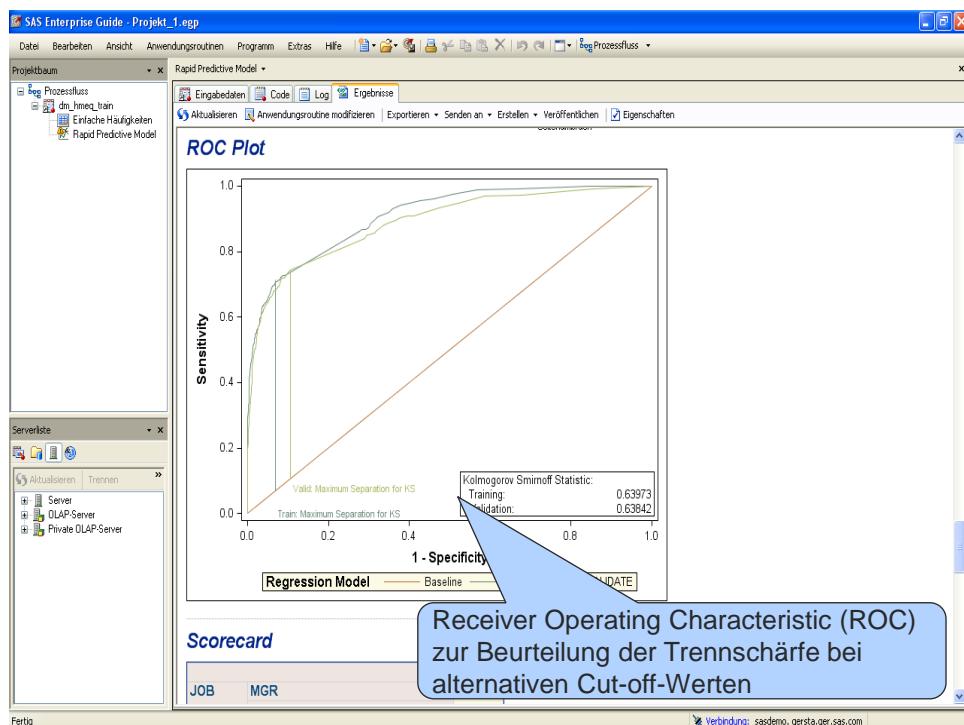
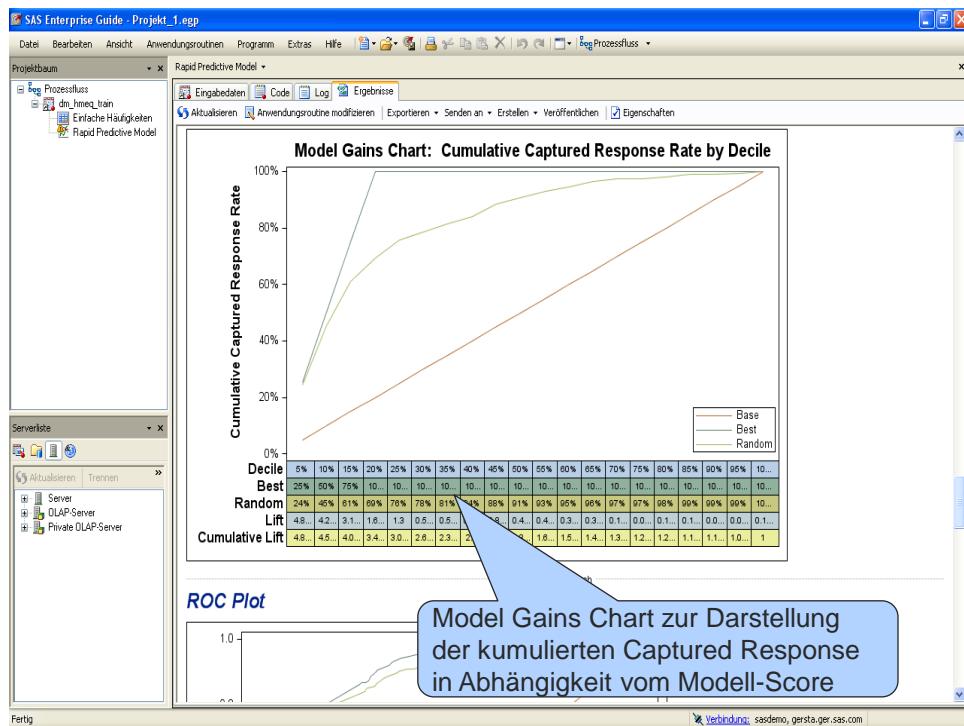
The screenshot shows the SAS Enterprise Guide interface with a project titled 'Rapid Predictive Model for: D:\Daten\Quellen\dm_hmeq_train.sas7bdat'. The 'Registration' dialog is open, displaying a table of data with columns JOB, YOJ, DEROG, and DELINQ. A callout bubble points to the table with the text: 'Das generierte Projekt kann zudem in den SAS Metadaten registriert werden.'

The screenshot shows the SAS Enterprise Guide interface with a project titled 'Rapid Predictive Model'. The 'Model Summarization' section displays a table of values for the target variable 'BAD'. A callout bubble points to this table with the text: 'Projektzusammenfassung und Werteverteilung für Zielgröße'. The 'Variable Summary' section shows a table of roles and levels for input variables.

| Role | Level | Original Count | Selected Count |
|--------|----------|----------------|----------------|
| INPUT | BINARY | 1 | . |
| INPUT | INTERVAL | 7 | 6 |
| INPUT | NOMINAL | 4 | 4 |
| TARGET | BINARY | 1 | . |







Aus dem Modell generierte Score-Karte mit Punktezahl (Scores) pro Attribut (Variablenausprägung)

| | | score |
|---------|----------------------------------|--------|
| JOB | MGR | 7.00 |
| | OFFICE | 0.00 |
| | OTHER | 17.00 |
| | PROFEXE | 10.00 |
| | SALES | 63.00 |
| | SELF | 22.00 |
| CLAGE | 01:LOW-53.717144096 | 111.00 |
| | 02:53.717144096-123.89368418, MI | 102.00 |
| | 03:123.89368418-242.49285066 | 77.00 |
| | 04:242.49285066-HIGH | 55.00 |
| | _MISSING_ | 0.00 |
| CLNO | _MISSING_ | 110.00 |
| | 01:LOW-0.5 | 182.00 |
| | 02:0.5-2.5 | 122.00 |
| | 03:2.5-9.5, MISSING | 58.00 |
| | 04:9.5-38.5 | 37.00 |
| | 05:38.5-46.5 | 78.00 |
| | 06:46.5-60 | 0.0 |
| | 07:60-HIGH | 110.00 |
| DEBTINC | 01:LOW-6.5765689657, MISSING | 95.00 |
| | 02:6.5765689657-34.525944418 | 0.00 |
| | 03:34.525944418-42.914801163 | 19.00 |
| | 04:42.914801163-45.184804524 | |
| | 05:45.184804524-HIGH | |
| LOAN | 01:LOW-5050 | |
| | 02:5050-15050 | |
| | 03:15050-37900, MISSING | |



Fazit - Rapid Predictive Modeler ...

- ... führt Anwender in wenigen Schritten zum fertigen Vorhersagemodell und verkürzt Entwicklungszeiten
- ... macht die im SAS Enterprise Miner vorhandene Funktionalität über einfache Oberfläche leicht zugänglich
- ... erschließt das Thema Data Mining auch Anwendern ohne umfassendes Statistik-Know-How
- ... bedeutet einen weiteren Meilenstein von SAS in Richtung Industrialisierung von Data Mining

Weitere Informationen:

<http://support.sas.com/resources/papers/proceedings10/113-2010.pdf>
(SAS Global Forum 2010 Paper von W. Thompson, D. Duling)