

 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.

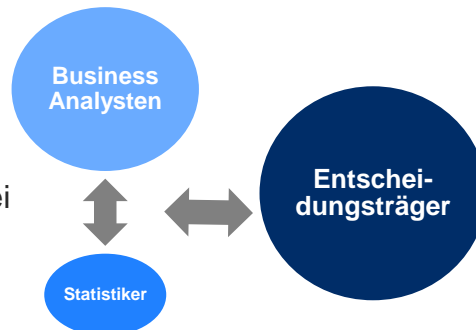
Vorhersagemodellierung als Herausforderung

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

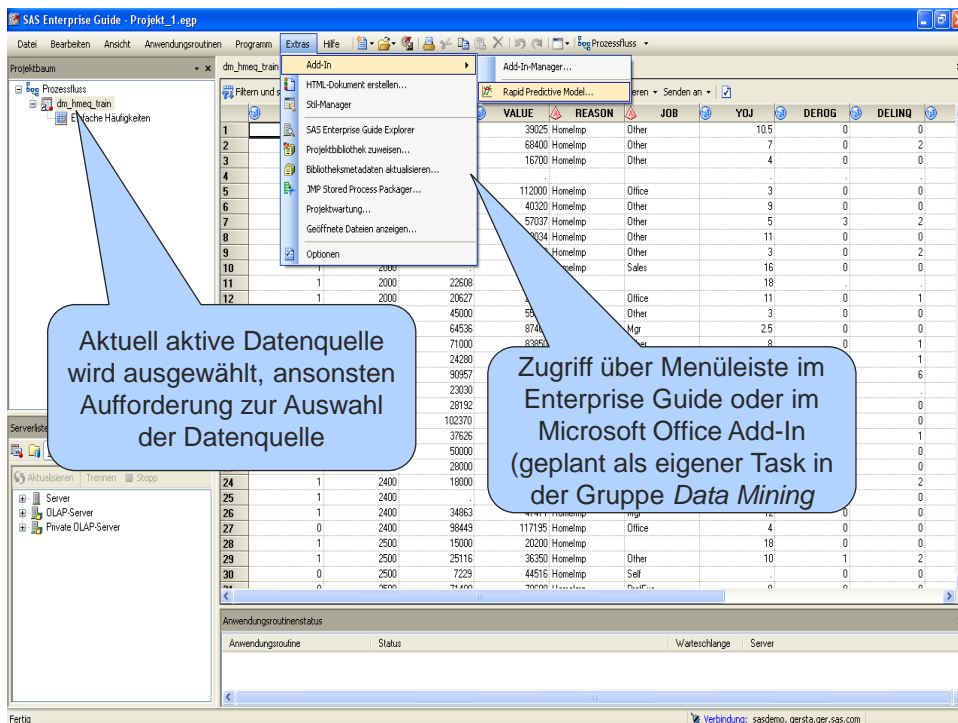
Copyright © 2008, SAS Institute Inc. All rights reserved.

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.



Rollenzuweisung für Variablen:

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

→ alle anderen Variablen dienen als Kandidaten für Modell

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

Optional können für Klassifikationsmodelle auch die A-priori-Wahrscheinlichkeiten und Entscheidungsvariablen spezifiziert werden.

SAS Enterprise Guide - Projekt_1.egg

Projektbaum: dm_hmreg_train

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

Report options:

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

Buttons: Run, Save, Abbrechen

Exportieren: Speichern an

ASON	JOB	YOJ	DEROG	DELINQ
Other		10.5	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
Office		11	0	1
Other		3	0	0
Mgr		2.5	0	0
Other		8	0	1
Other		7	2	6
Office		4.5	0	0
Office		2	0	0
Other		3	0	1
ProdExe		5	1	0
Mgr		12	0	0
Mgr		22	0	2
Other		12	0	0
Mgr		4	0	0
Office		18	0	0
Office		10	1	2
Office		0	0	0

Hier wird der notwendige Umfang der zu generierenden Berichte festgelegt.

Anwendungsroutinenstatus

Anwendungsroutine: Status

Fertig

Verbindung: sasdemo, gersta.ger.sas.com

SAS Enterprise Guide - Projekt_1.egg

Projektbaum: dm_hmreg_train

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

Options:

☐ Save Enterprise Miner Project Data

SAS App Server Folder: Browse...

Buttons: Run, Save, Abbrechen, Help

Exportieren: Speichern an

ASON	JOB	YOJ	DEROG	DELINQ
Other		10.5	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
Office		11	0	1
Other		3	0	0
Mgr		2.5	0	0
Other		8	0	1
Other		7	2	6
Office		4.5	0	0
Office		2	0	0
Other		3	0	1
ProdExe		5	1	0
Mgr		12	0	0
Mgr		22	0	2
Other		12	0	0
Mgr		4	0	0
Office		18	0	0
Office		10	1	2
Office		0	0	0

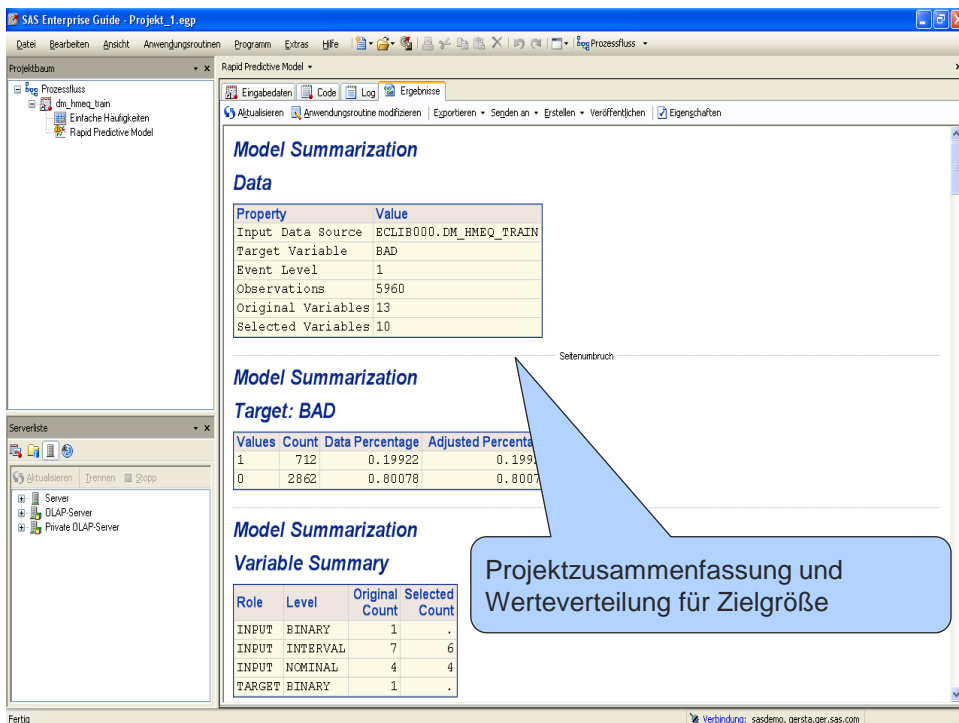
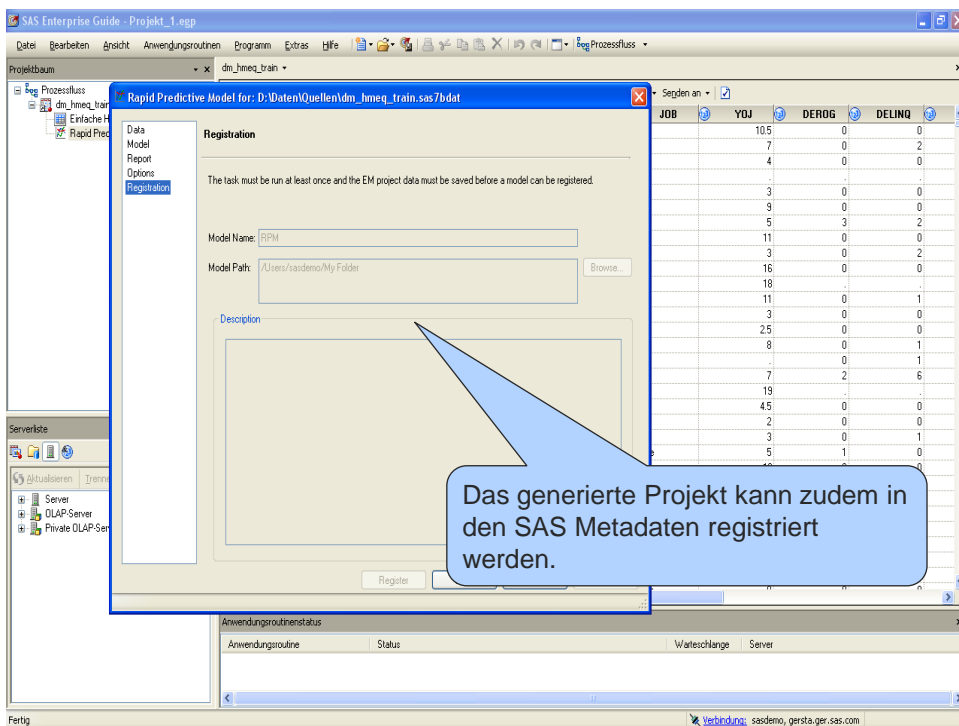
Hier kann ausgewählt werden, wo der im Hintergrund generierte Enterprise Miner Prozessfluss als Projekt gespeichert werden soll.

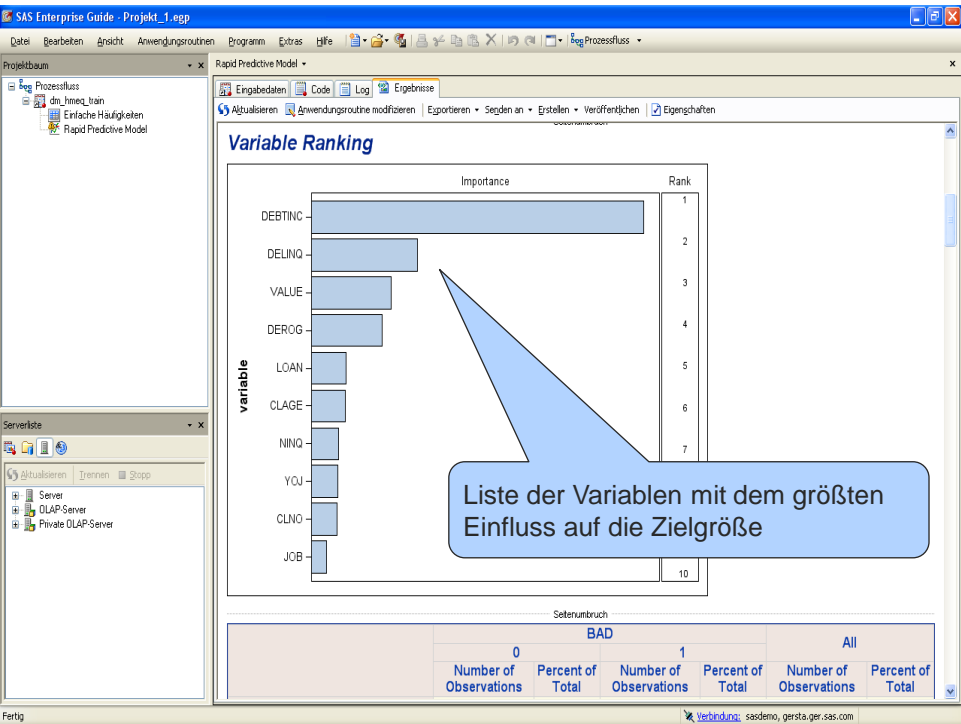
Anwendungsroutinenstatus

Anwendungsroutine: Status Warteschlange Server

Fertig

Verbindung: sasdemo, gersta.ger.sas.com





SAS Enterprise Guide - Projekt_1.egg

Projektbaum

Serverliste

Rapid Predictive Model

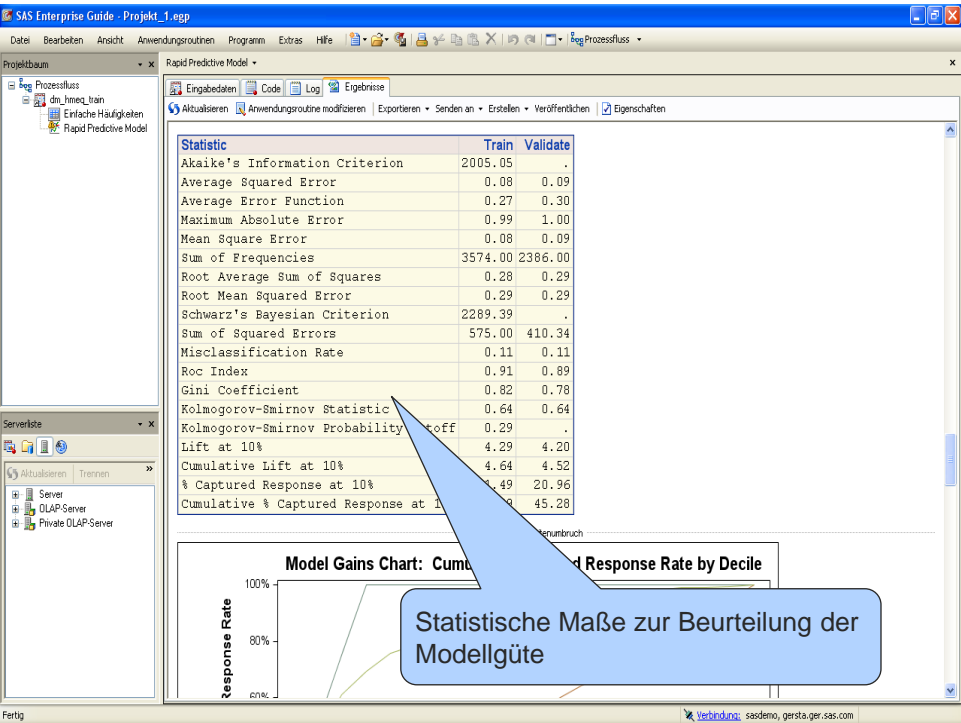
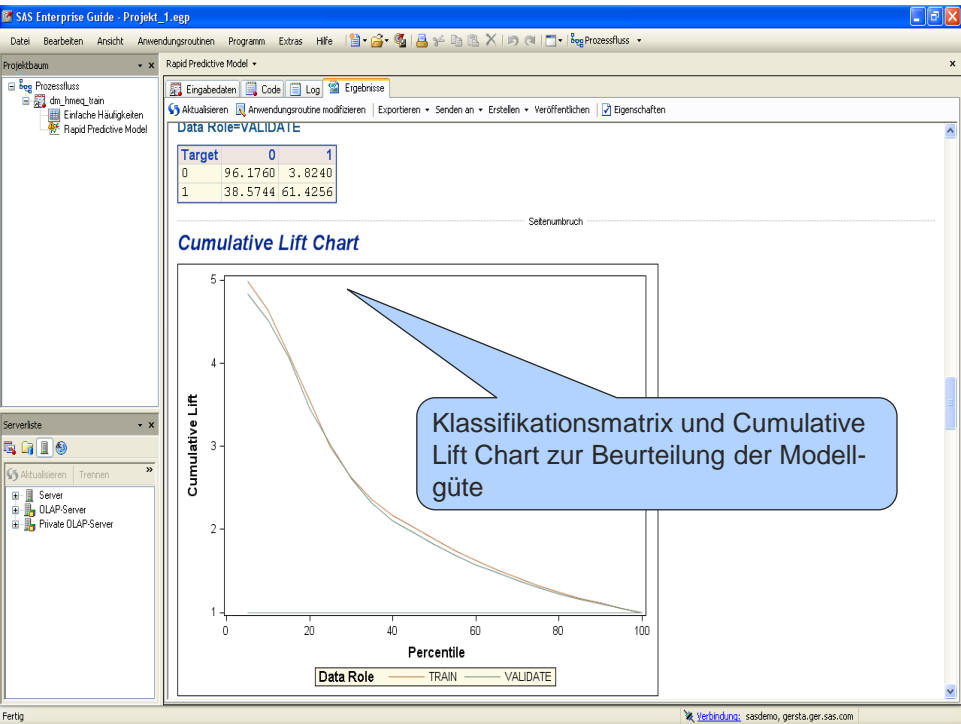
Variable Ranking

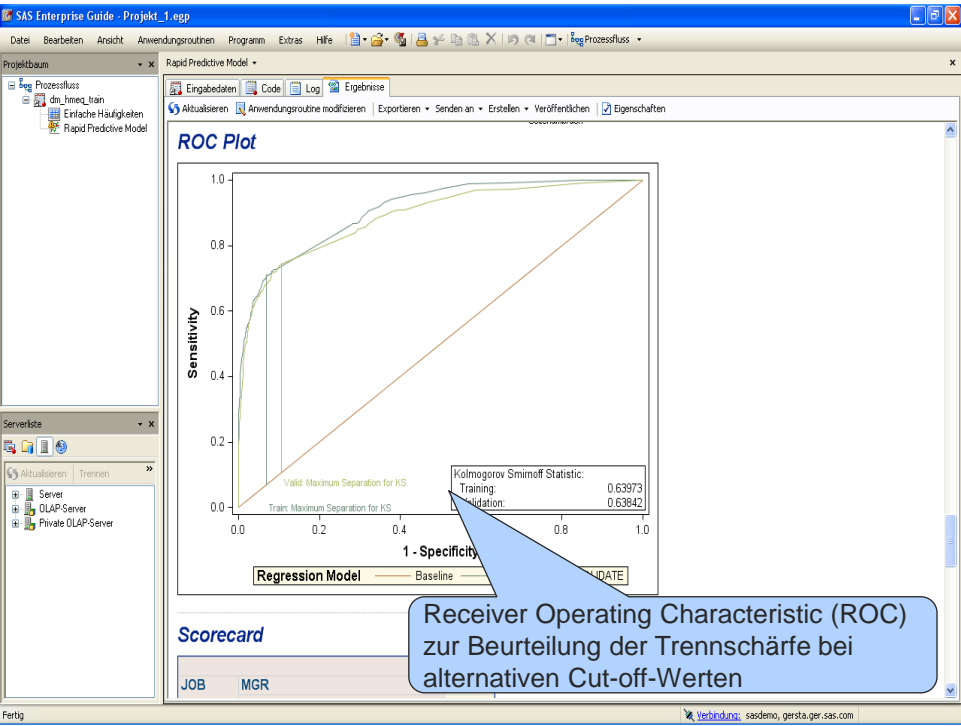
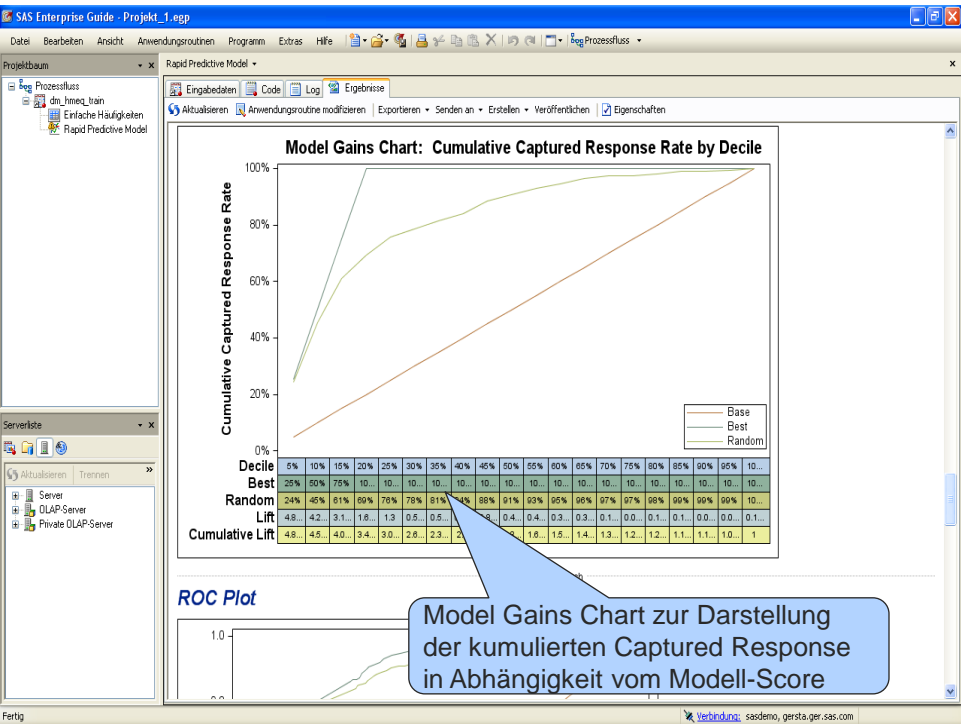
Selerumbruch						
0		1		All		
Number of Observations	Percent of Total	Number of Observations	Percent of Total	Number of Observations	Percent of Total	
CLAGE						
01:low -53.717144096	29	0.97	34	1.14	63	2.11
02:53.717144096-123.89368418, MISSING	570	19.11	251	8.41	821	27.52
03:123.89368418-242.49285066	1177	39.46	257	8.62	1434	48.07
04:242.49285066-high	562	18.84	66	2.21	628	21.05
MISSING	24	0.80	13	0.44	37	1.24
CLNO						
01:low -0.5	12	0.40	14	0.47	26	0.87
02:0.5-2.5	.	.	11	0.37	11	0.37
03:2.5-9.5, MISSING	169	5.67	69	2.31	238	7.98
04:9.5-38.5	2077	69.63	469	15.72	2546	85.35
05:38.5-46.5	43	1.44	38	1.27	81	2.72
06:46.5-60	61	2.04	13	0.44	74	2.48
07:60-high	.	.	7	0.23	7	0.23
DEBTINC						
01:low -6.5765689657, MISSING	7.48		413	13.85	636	21.32
02:6.5765689657-34.525944418	34.96		48	1.61	1091	36.57
03:34.525944418-42.914801163	4.90		101	3.39	1142	38.28
04:42.914801163-45.184804524	11		18	0.60	72	2.41
05:45.184804524-high			41	1.37	42	1.41
DELINQ						
0						79.65
1						10.49
2						4.32
OTHER						5.53

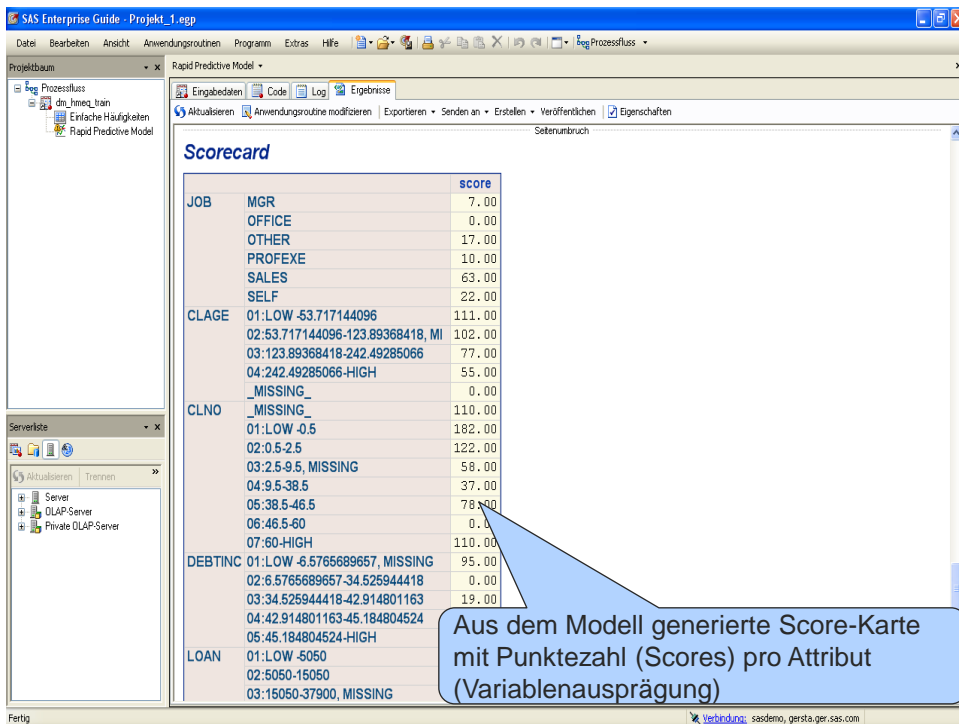
Fertig

Verteilungsübersicht der (transformierten) Input-Variablen

Verbindung: sasdemo, gersta.ger.sas.com







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)

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