

Qualitätskontrolle industrieller Prozesse mit dem SAS Enterprise Miner

Claudia Tückmantel
INFORMATION WORKS GmbH



INFORMATION WORKS Dienstleistungsprofil im Überblick

- § 100 % spezialisiert auf die individuelle Konzeption, Realisierung und Einführung von Business Intelligence Systemen einschließlich Beratung und Projektmanagement
- § Standorte: Köln, Frankfurt, Stuttgart, Hamburg, Essen, Barcelona



Leistungen

- § Maßgeschneiderte Konzeption und Entwicklung von Data Warehouse- sowie analytischen Lösungen einschließlich Beratung, Projektmanagement und Coaching
- § Erfahrungen in Projekten der verschiedenen Gattungen und Branchen, z.B.:
 - Data Mining, Analytische Statistik
 - Automatisierung und Monitoring statistischer Verfahren
 - Credit Scoring
 - Balanced Scorecard
 - Analytisches CRM, eCRM/Web Mining
 - DWH-Nutzung im Intranet/Extranet

INFORMATION WORKS Dienstleistungsprofil im Überblick

- § Projektgesamtverantwortung
- § Projektbegleitung / Coaching
- § Intensiv-Seminare für IT und Fachbereiche über BI-Themen wie:
 - Projektmanagement
 - Software Werkzeuge
 - Data Mining
 - ...

Kooperationen mit Universitäten



Universität Hamburg

UNIVERSITÄT DORTMUND



Universität Stuttgart



Universität Köln

Auszug aus der Kundenliste

Banken

DaimlerChrysler Bank

Deutsche Bank

Dresdner Bank
Die Beraterbank

Frankfurter Sparkasse 1822

Versicherungen

GERLING
Wir machen alles mit Sicherheit

Hamburg-Mannheimer

ERGO
Die neue Größe.

VICTORIA



SIGNAL IDUNA
Versicherungen und Finanzen

PROVINZIAL
Westfälische Provinzial
Versicherung der Sparkassen



Handel / Logistik

KARSTADT

Frankfurt Airport
managed by fraport

Lufthansa

VORWERK



Industrie / Telco



INTER SNACK

ARCOR

BASF

Data Mining in der Industrie - Ausgangssituation

- § Umfangreiche Datenbestände aus Fertigungsprozessen
- § Automatisierte Prozesse sammeln online Produktionsdaten in großem Umfang (z.B. zahlreiche real-time Sensordaten)
- § Effizienz der Produktionsprozesse und Qualität der Produkte hängt von vielen Variablen ab
- § Bisher erfolgt häufig nur eine stichprobenartige Kontrolle einzelner Werte

Data Mining in der Industrie - Ausgangssituation

- § Ressource Information wird nicht in vollem Umfang genutzt
- § Extraktion von zusätzlichem Wissen über die Fertigungsprozesse mit Hilfe von Data Mining Methoden und anderen statistischen Verfahren ist möglich
- § Systematische Analysen liefern Wettbewerbsvorteile

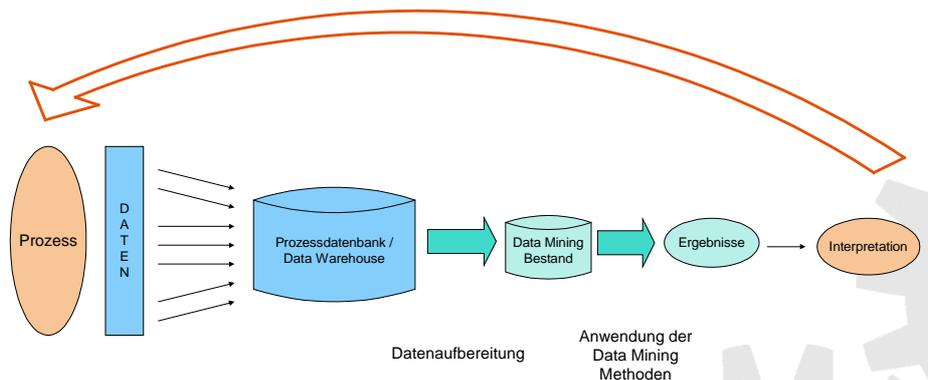
Was leistet Business Intelligence für Kontrolle, Monitoring und den laufenden Betrieb komplexer industrieller Prozess?

- § Informationen des momentanen Zustands, in dem sich der Fertigungsprozess befindet
- § Informationen zu Möglichkeiten einer langfristigen Performanzbeibehaltung oder -steigerung

Möglichkeiten durch Data Mining

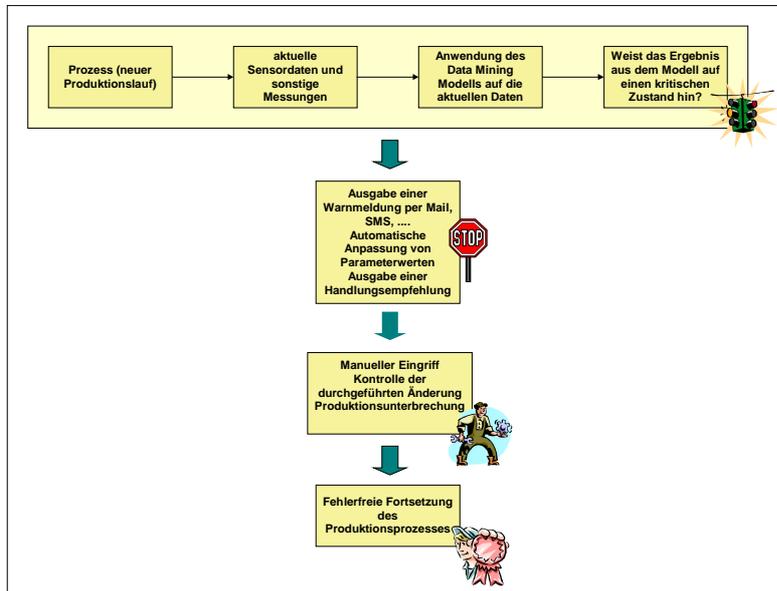
- § Qualitätskontrolle und -verbesserung
- § Verkürzung von Entwicklungszeiten
- § Produktivitätsverbesserung
- § Produktverbesserung
- § Kostenreduzierung
- § Vermeidung von Störungen
- § Überwachung der Prozess-Sicherheit
- § Einhaltung von Umweltschutzbestimmungen

Data Mining im Fertigungsprozess



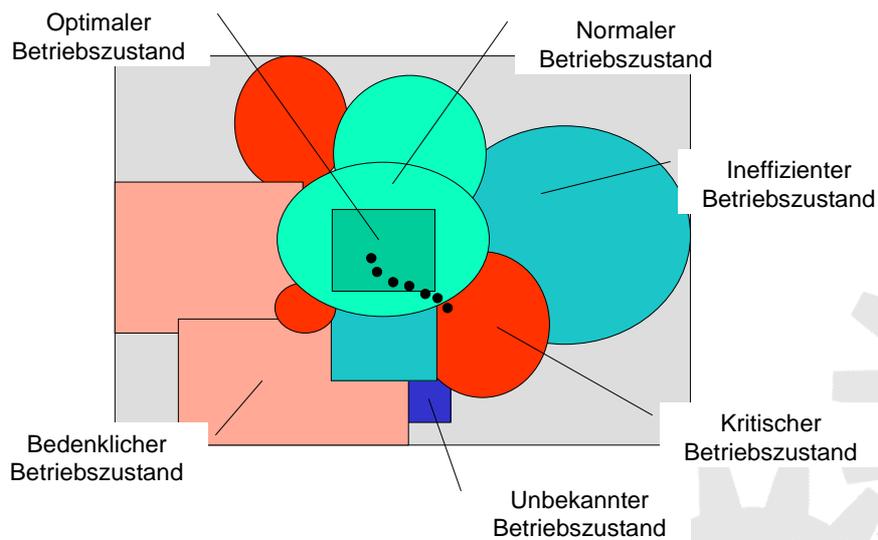
Ziele beim Data Mining

- ⌘ Ableitung von Steuergrößen und optimalen Betriebsparametern
- ⌘ Rechtzeitige Ausgabe von Warnmeldungen
- ⌘ Schätzung schwierig messbarer Parameter
- ⌘ Entdeckung von Regeln und Zusammenhängen zur Verbesserung des Produktionsprozesses
- ⌘ Identifikation unterschiedlicher Zustände des Prozesses



Claudia Tückmantel

KSFE Potsdam, 20.-21.02.2003



Claudia Tückmantel

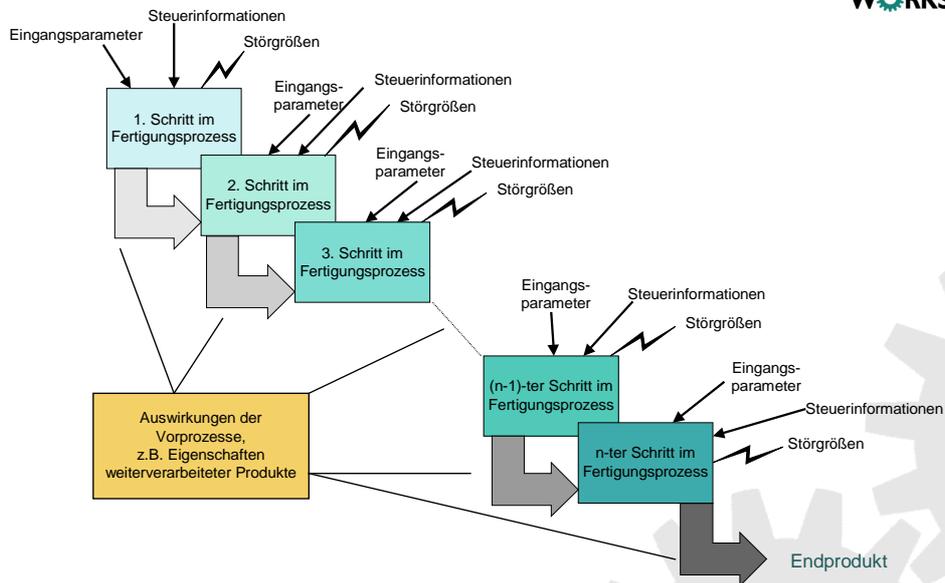
KSFE Potsdam, 20.-21.02.2003

Quality Correlation Analysis

- ☞ Datenbasierte Suche nach Ursache-Wirkungs-Zusammenhängen und Fehlerursachen
- ☞ Einsatz bei Prozessketten sensibler Erzeugnisse mit vielen Einflussfaktoren und komplexen Fertigungssystemen
- ☞ Fehlerursache für nicht reproduzierbare Fehler bzw. Qualitätsmängel

Komplexe Prozessketten

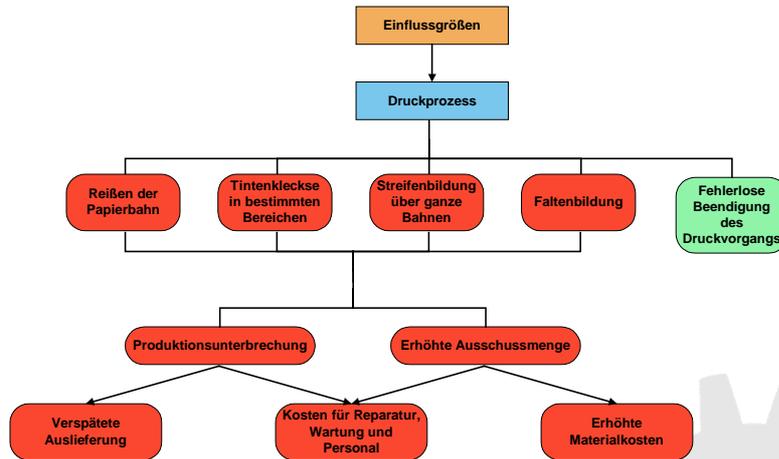
- ☞ Häufiges Problem in der Fertigung: Nicht geklärtes Auftreten von Qualitätsmängeln, -schwankungen und Fehlern
- ☞ Ursache liegt häufig in der Prozesskette und nicht in dem fehlerhaften Prozessschritt
- ☞ Extrem große Anzahl möglicher Einflussgrößen
- ☞ Keine Abhilfe durch Stichproben und statistische Versuchsplanung möglich



Beispiele

- ⊗ Auswertung von Motorentests/Reifentests
- ⊗ Verbesserte Einstellung auf unterschiedliche Zulieferer
- ⊗ Erkennung von Grenzwertüberschreitungen im Vorfeld (Kernkraftwerke, Kläranlagen)
- ⊗ Viskosität und Farbbeständigkeit von Lacken
- ⊗ Qualitätsmängel in der Weißblechherstellung
- ⊗ Qualitätssteigerung in der Chip-Produktion
- ⊗ Ausschussreduzierung in der Druckindustrie

Probleme in der Druckindustrie



Claudia Tückmantel

KSFE Potsdam, 20.-21.02.2003

Die Ziele

- Unterbrechungen vermeiden
- Bedingungen, unter denen Unterbrechungen auftreten, isolieren
- Ausschussmenge reduzieren
- Gründe für Unterschiede zwischen Druckereien aufdecken und nutzen

Claudia Tückmantel

KSFE Potsdam, 20.-21.02.2003

Konkrete Fragestellung

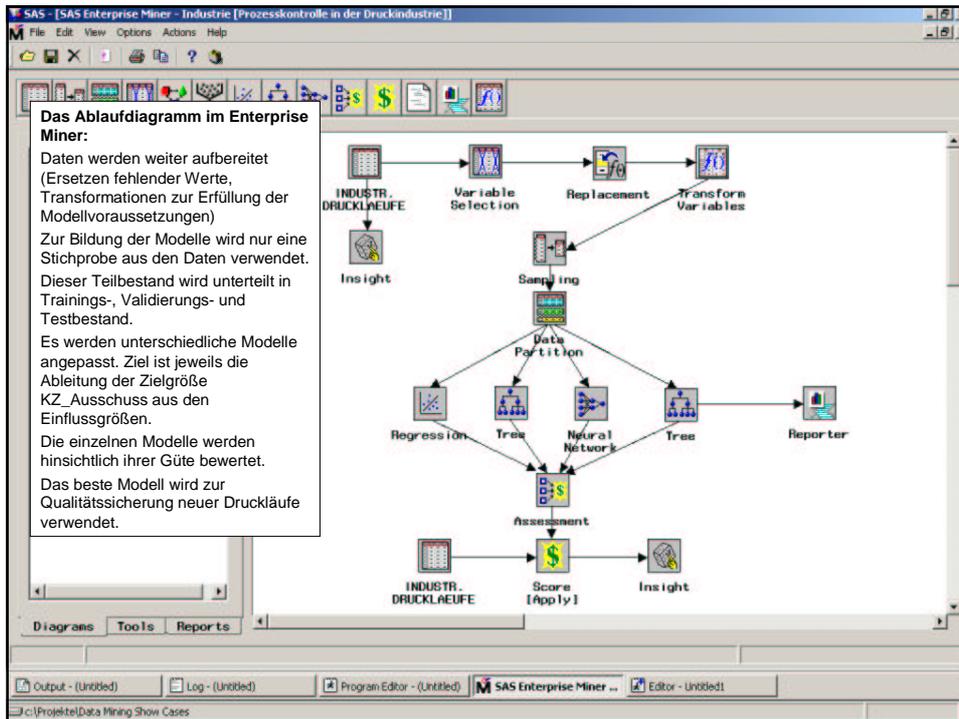
- § Welche Parameter haben Einfluss auf die Abfallmenge?
- § Regeln für Verbesserungen des Produktionsablaufes im Hinblick auf die Ausschussmenge



Demonstration im SAS Enterprise Miner



Data Mining Showcases.Ink



SAS - [EMPROJ.SMP_V122]

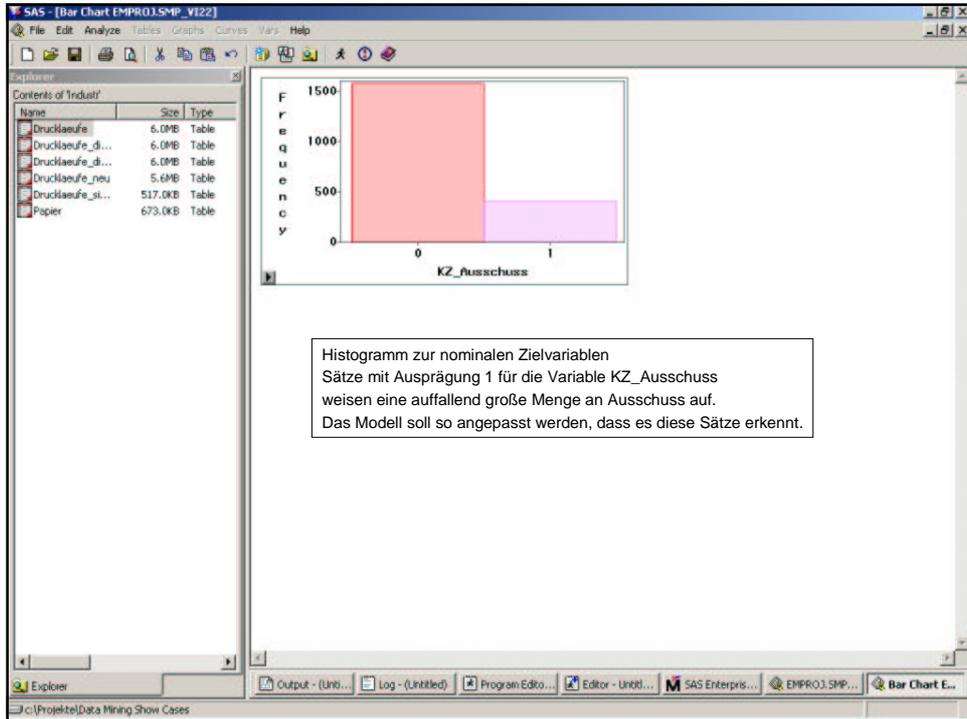
2000	25	Int	Int	Int	Int	Int	Int	Int	Int	Int
	Maschinen_Par	Maschinen_Par	Viskosität_Tinte	KZ_Beinschuss	Temperatur_Tinte	Zeit_Zylinderreinigung	M_Rollensinsp			
1	7.6636	42.7504	1.0816	1	25.7536	63.563	Thomas, Andres			
2	1.0038	35.0038	0.1004	0	25.6101	68.438	Newman, Al			
3	-2.3357	32.1308	0.0000	0	25.3393	66.875	Davis, Alan			
4	10.1265	39.4527	1.2755	0	26.3179	64.188	Dawson, Andre			
5	-3.1707	47.7815	0.0000	0	27.2453	51.250	Tranelli, Alan			
6	4.5851	38.9041	0.0000	1	27.2234	58.625	Thornton, Andre			
7	1.4184	35.4184	0.1418	1	26.6099	70.500	Trevino, Alex			
8	4.3668	44.3987	0.2293	1	28.2379	229.000	Trevino, Alex			
9	-5.8937	28.1063	0.2003	0	26.4858	324.500	Thomas, Andres			
10	4.0366	40.9724	0.0367	0	27.2026	34.063	Thomas, Andres			
11	-6.6908	27.3092	-0.6691	1	26.1790	36.438	Dawson, Andre			
12	0.0000	31.8541	-0.2788	1	26.8748	29.125	Dawson, Andre			
13	-2.1459	31.8541	-0.2146	0	28.1394	29.125	Almon, Bill			
14	-11.9558	23.7535	-0.8540	0	29.0425	32.938	Dawson, Andre			
15	3.5290	37.5290	-0.2817	1	26.6275	26.625	Newman, Al			
16	3.1051	38.5732	-0.0678	0	26.7145	110.688	Thomas, Andres			
17	5.0158	40.5751	-0.3092	0	26.0892	240.500	Trevino, Alex			
18	-11.1326	-35.9659	0.0000	0	26.8795	258.250	Newman, Al			
19	-11.1326	-35.9659	0.0000	0	27.9057	258.250	Trevino, Alex			
20	1.4123	35.4123	0.1412	1	25.9850	6.656	Trevino, Alex			
21	4.5294	37.8853	0.2597	0	26.2994	4.813	Davis, Alan			
22	-3.2078	22.9598	-0.3070	1	26.5108	136.438	Trevino, Alex			
23	-1.6629	31.8331	0.2218	0	27.2426	124.000	Tranelli, Alan			
24	-4.2895	29.7105	-0.3796	0	28.4524	126.750	Almon, Bill			
25	-21.4565	9.3444	-1.1749	0	26.3524	113.313	Almon, Bill			
26	5.0682	35.3645	0.2534	1	27.3822	64.125	Dawson, Andre			
27	-3.1194	30.8806	-0.3119	0	27.8194	26.063	Almon, Bill			
28	-4.0059	28.9448	0.0226	0	27.6266	27.875	Newman, Al			
29	-10.8527	23.1473	-0.7364	0	27.6929	32.250	Davis, Alan			
30				1	27.5441		Thomas, Andres			
31	1.8692	38.2056	0.2804	1	27.2907	26.750	Trevino, Alex			
32	5.2917	39.2917	0.0407	1	26.2500	184.250	Davis, Alan			
33						37.125	Newman, Al			
34						62.875	Davis, Alan			
35						63.563	Dawson, Andre			
36						66.125	Tranelli, Alan			
37						65.000	Davis, Alan			
38						46.250	Newman, Al			
39						49.750	Trevino, Alex			
40	13.8875	53.6128	0.0774	0	27.0719	161.563	Bell, Buddy			
41	14.4831	43.2896	0.7650	0	26.2690	22.875	Dawson, Andre			

Auszug aus dem verwendeten Datenbestand.

23 potentielle Einflussgrößen (Maschinenparameter, Reinigungsintervall,...)

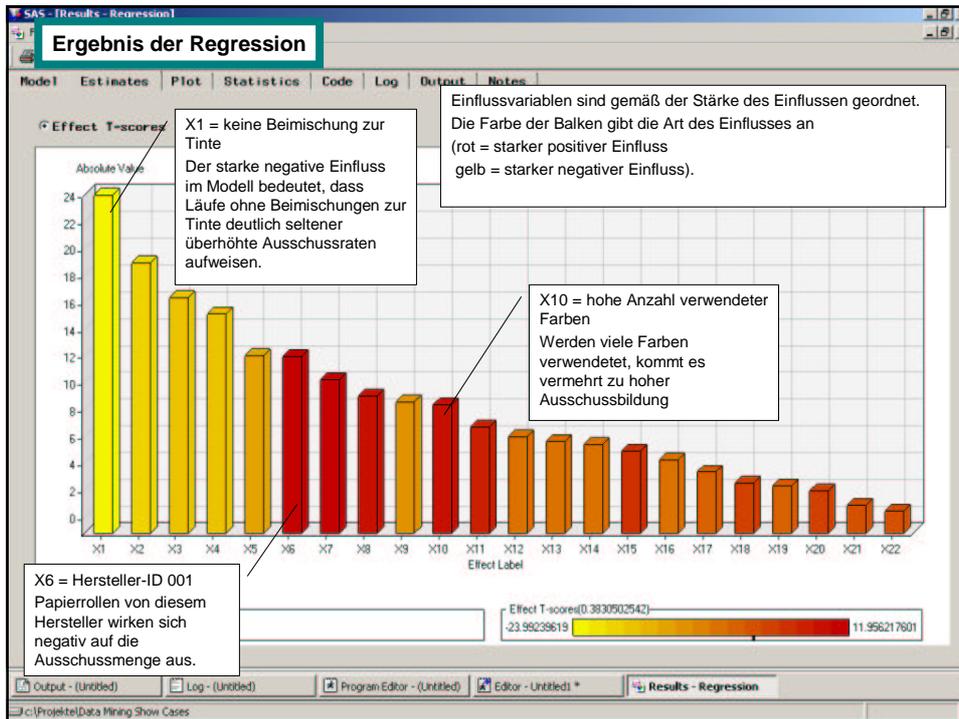
Abgeleitete Zielgröße: Kennzeichen für Läufe mit auffallend hohem Ausschuss

27.000 Beobachtungssätze



Name	Model Role	Measurement	Type	Format	Informat	Variable Label
LAUF_ID	id	interval	num	BEST12.	12.	
PAPIER_SORTE_ID	input	interval	num	BEST12.	12.	
SORTENNAME	rejected	nominal	char	\$26.	\$26.	
MASCH_ID	input	interval	num	BEST12.	12.	
HERST_ID	input	nominal	num	Z3.	3.	
HERST_DAT	rejected	interval	date	DATE7.	DATE7.	
EINSATZ_DAT	rejected	interval	date	DATE7.	DATE7.	
ALTER	input	interval	num	BEST12.	12.	
GEWICHT	input	interval	num	BEST12.	12.	
GROESSE	input	interval	num	BEST12.	12.	
PAPIERDICK	input	interval	num	BEST12.	12.	
ROLLENLAUF	input	ordinal	num	BEST12.	12.	
ANZFARBEN	input	ordinal	num	BEST12.	12.	
MASCHINEN_PAR1	input	interval	num	BEST12.	12.	
MASCHINEN_PAR2	input	interval	num	BEST12.	12.	
MASCHINEN_PAR3	input	interval	num	BEST12.	12.	
MASCHINEN_PAR4	input	interval	num	BEST12.	12.	
MASCHINEN_PAR5	input	interval	num	BEST12.	12.	
VISKOSITAET_INTE	input	interval	num	BEST12.	12.	
KZ_BEIMISCHUNG	input	binary	num	BEST12.	12.	
TEMPERATUR_INTE	input	interval	num	BEST12.	12.	
ZEIT_ZYKLINDERINIGUNG	input	interval	num	BEST12.	12.	
AUSCHUSS	rejected	interval	num	BEST12.	12.	
KZ_AUSCHUSS	target	binary	num	BEST12.	12.	
NA_ROLLENSPANNUNG	input	nominal	char	\$18.	\$18.	NA_Rolleinspannung

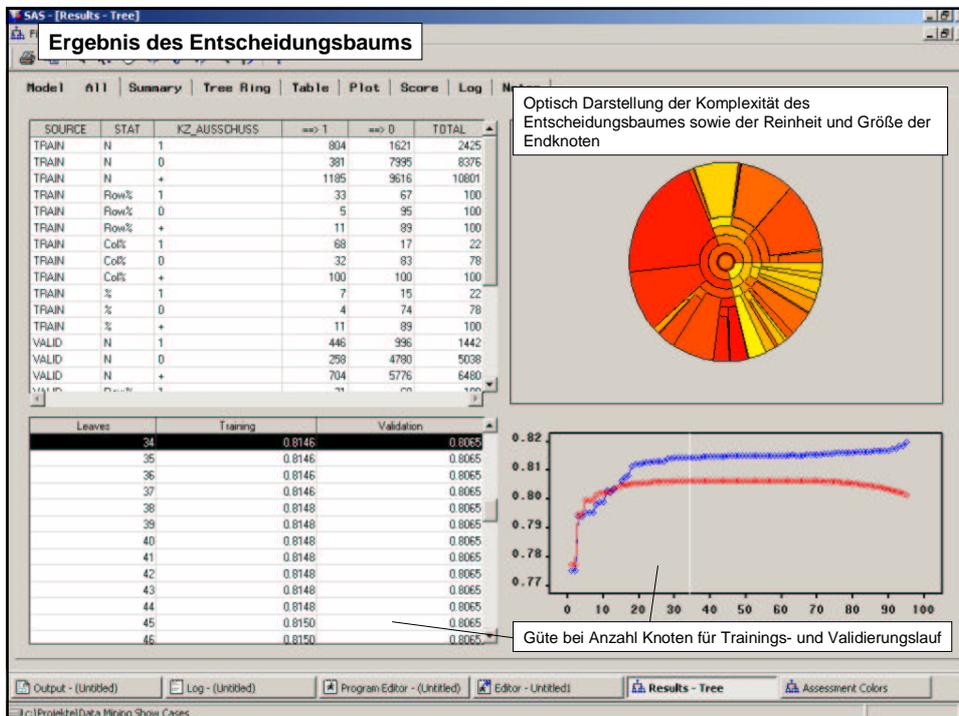
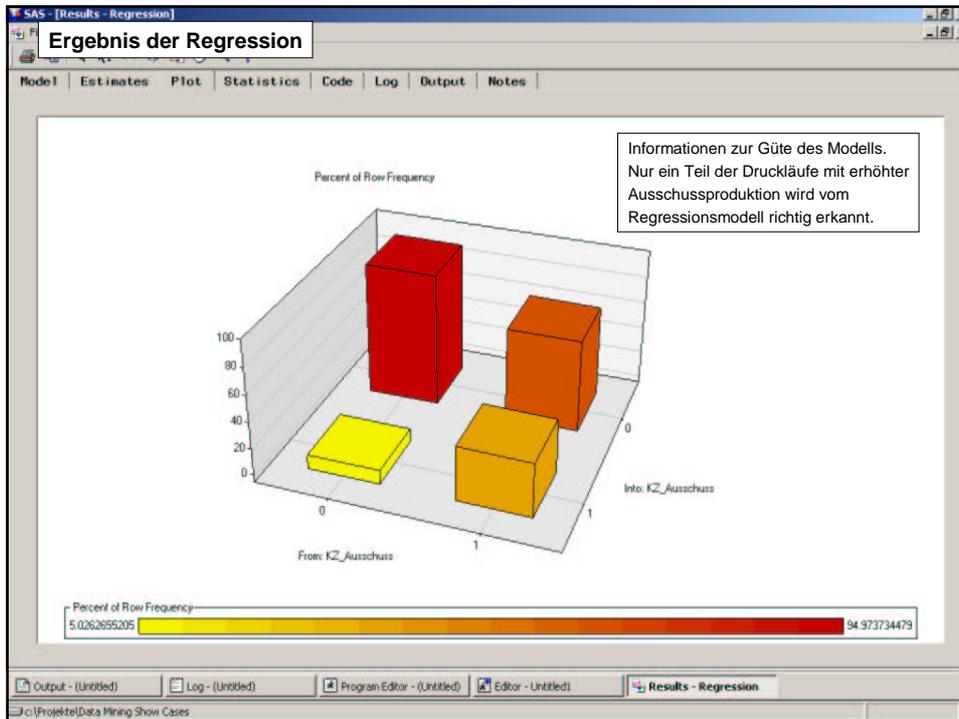
Zuweisung der Variablenrollen: (Einflussgrößen, Zielgröße, nicht verwendete Variablen)

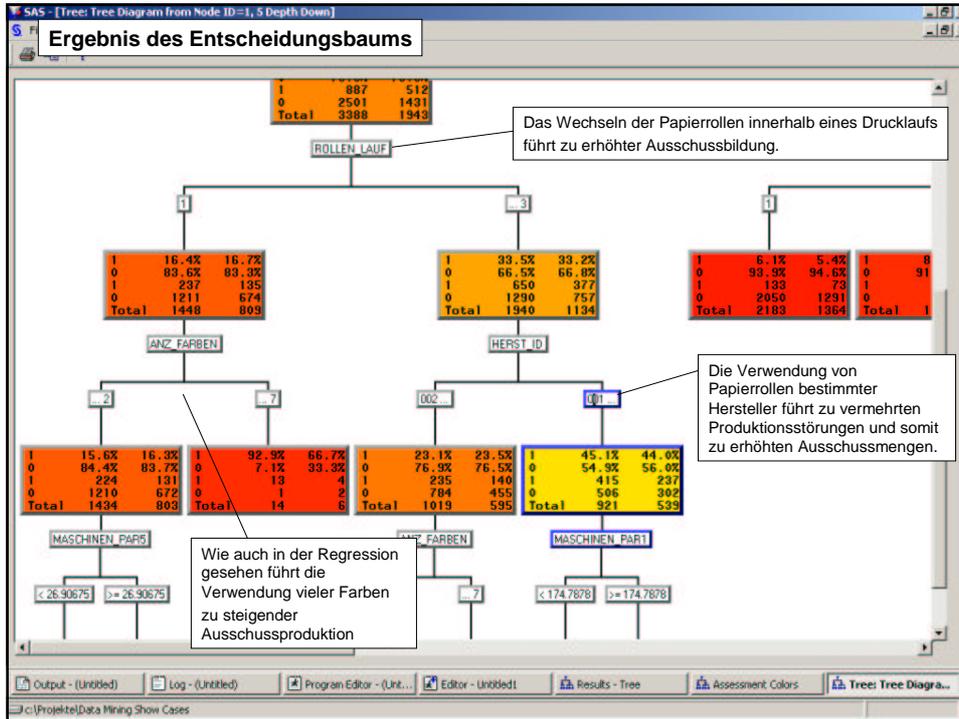


Ergebnis der Regression

Übersicht über die in das Modell einfließenden Variablen und die zugehörigen Parameterwerte.

Effect Name	Effect Label	Parameter Estimate	Effect T-score
Alter	Alter	0.0018365959	6.8267149746
Anz_Farben1	Anz_Farben 1	-1.187612424	-16.232847523
Anz_Farben2	Anz_Farben 2	-1.092601509	-15.03160075
Anz_Farben6	Anz_Farben 6	1.15710021	8.2862766968
Gewicht	Gewicht	.	.
Groesse	Groesse	.	.
Herst_ID 001	Herst_ID 001	0.5663253276	11.996229315
Herst_ID 002	Herst_ID 002	-0.250723928	-4.238847523
Herst_ID 003	Herst_ID 003	-0.545739063	-8.59512139
Herst_ID 004	Herst_ID 004	-0.359590073	-5.874315734
Intercept	Intercept KZ_Ausschuss=1	-2.030596835	-19.20977937
KZ_Beimischung0	KZ_Beimischung 0	-0.695148878	-24.00013316
MA_RolleneinspannungALMON_BILL	MA_Rolleneinspannung ALMON, BILL	-0.179298713	-2.283958222
MA_RolleneinspannungBELL_BUDDY	MA_Rolleneinspannung BELL, BUDDY	0.2144951048	1.8475139908
MA_RolleneinspannungDAVIS_ALAN	MA_Rolleneinspannung DAVIS, ALAN	0.5840432879	8.9746695112
MA_RolleneinspannungDAWSON_ANDR	MA_Rolleneinspannung DAWSON, ANDRE	0.678963177	10.161523891
MA_RolleneinspannungNEWMAN_AL	MA_Rolleneinspannung NEWMAN, AL	-0.39751087	-5.954357829
MA_RolleneinspannungTHOMAS_ANDR	MA_Rolleneinspannung THOMAS, ANDRES	-0.278542614	-3.313748932
MA_RolleneinspannungTHORNTON_AN	MA_Rolleneinspannung THORNTON, ANDRE	-0.380384635	-5.317065297
MA_RolleneinspannungTRAMMELL_AL	MA_Rolleneinspannung TRAMMELL, ALAN	-0.032282949	-0.843696383
Masch_ID	Masch_ID	.	.
Maschinen_Par1	Maschinen_Par1	.	.
Maschinen_Par2	Maschinen_Par2	0.0095106032	28.968237788
Maschinen_Par3	Maschinen_Par3	.	.
Maschinen_Par4	Maschinen_Par4	.	.
Maschinen_Par5	Maschinen_Par5	.	.
Papier_Sorte_ID	Papier_Sorte_ID	.	.
Papierdicke	Papierdicke	.	.
Rollen_Lauf1	Rollen_Lauf 1	-0.460429034	-11.96573012

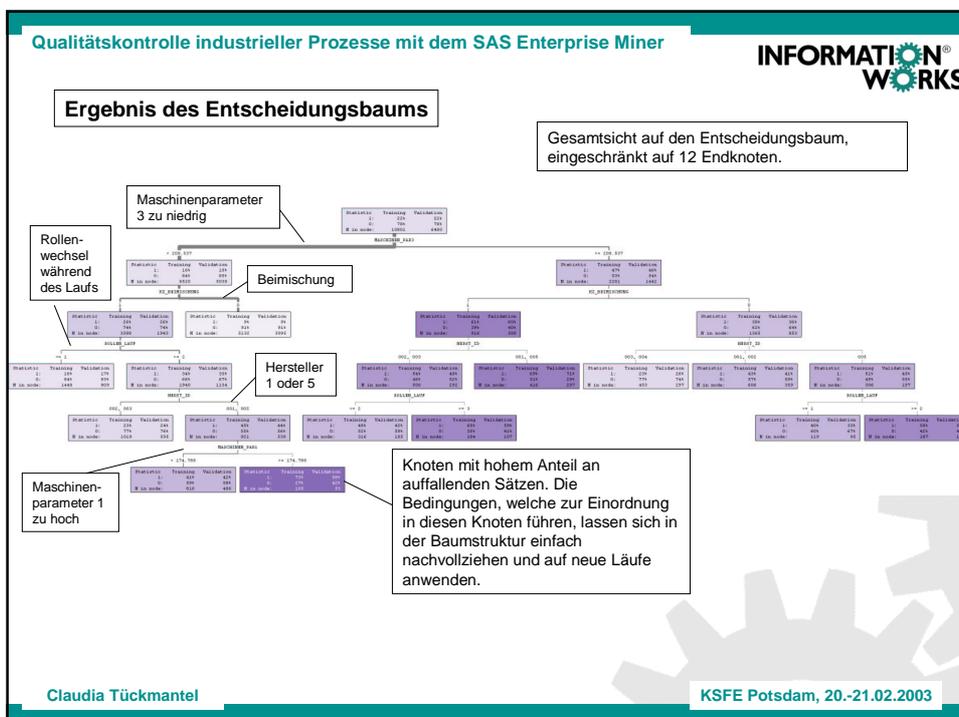
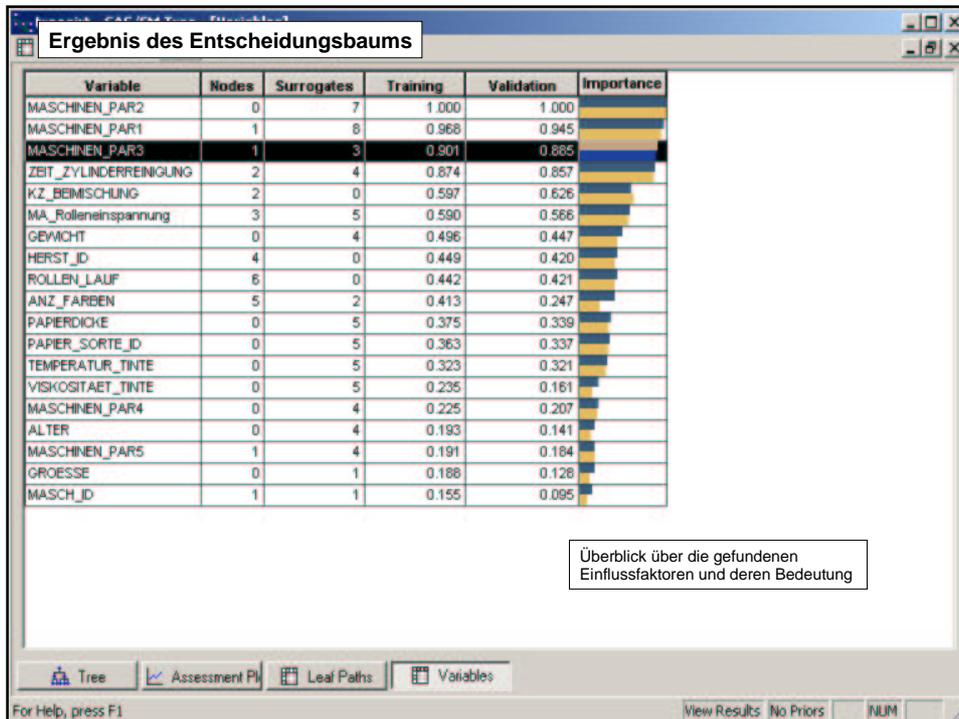


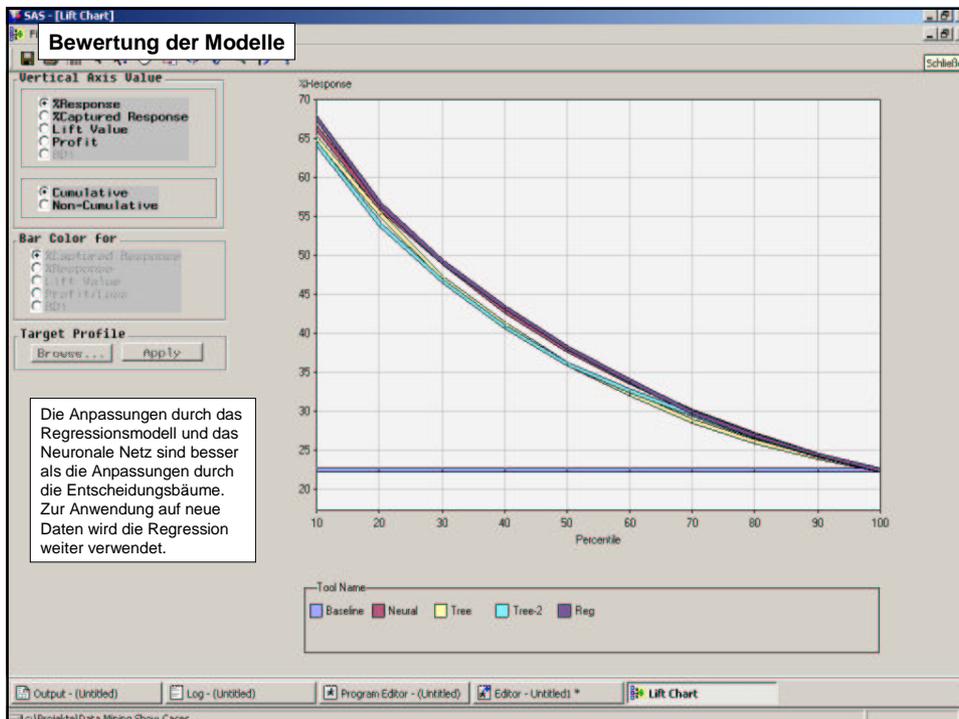
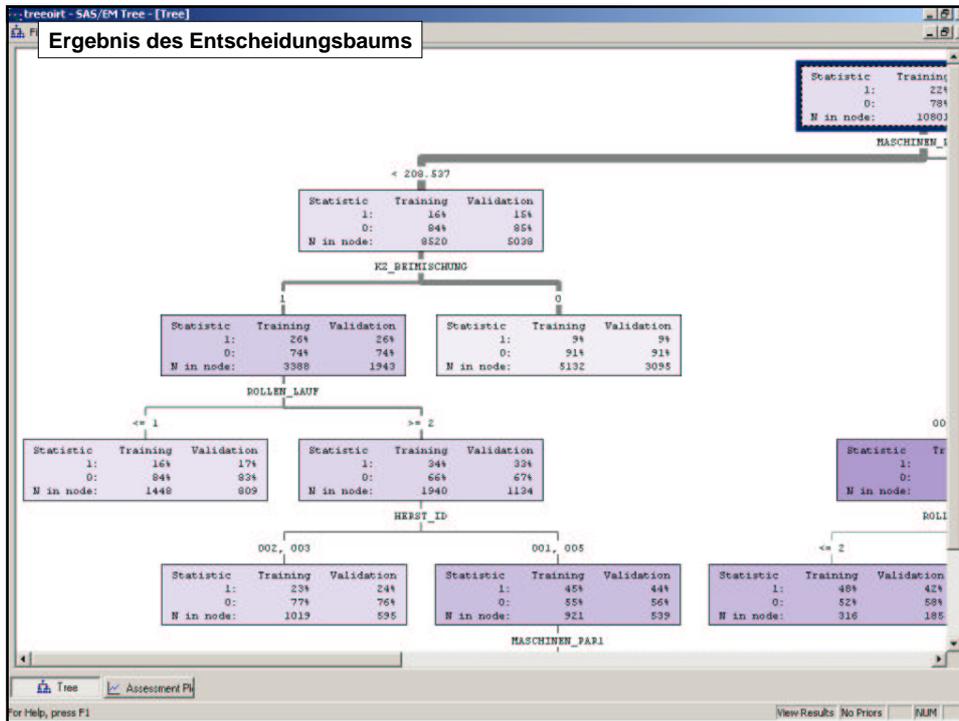


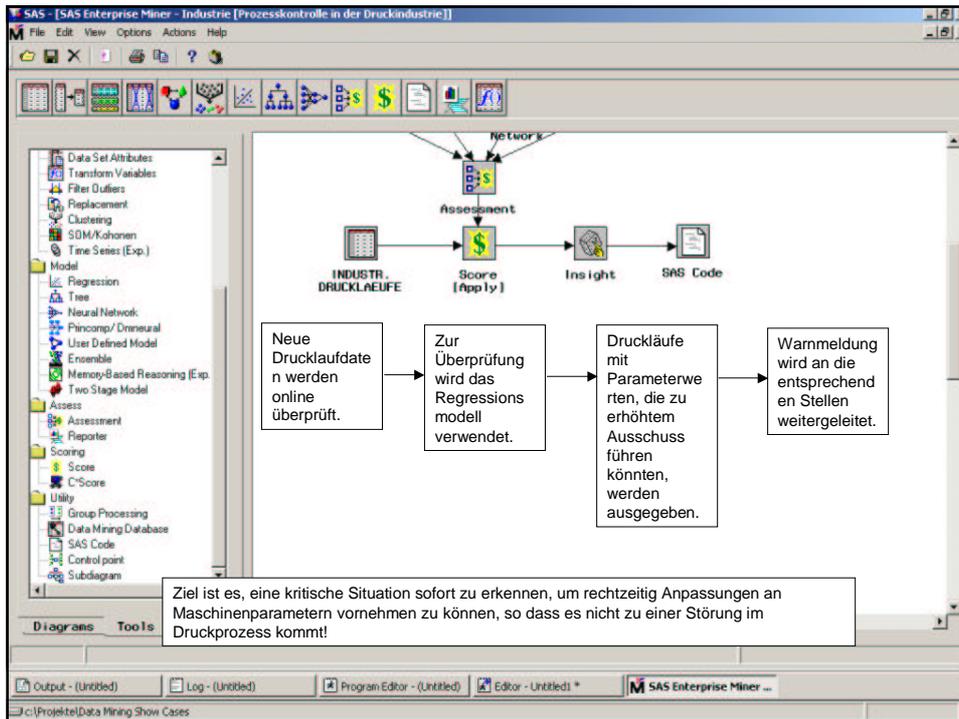
```

Tree_rules.txt - Editor
Ergebnis des Entscheidungsbaums
IF AND ZEIT_ZYLINDERREINIGUNG < 127.0315
THEN
  NODE : 6
  N : 4643
  1 : 8.1%
  0 : 91.9%
IF HERST_ID EQUALS 004
AND K2_BEIMISCHUNG EQUALS 1
AND ZEIT_ZYLINDERREINIGUNG < 127.0315
THEN
  NODE : 13
  N : 523
  1 : 17.6%
  0 : 82.4%
IF HERST_ID IS ONE OF: 002 003
AND 127.0315 <= ZEIT_ZYLINDERREINIGUNG < 149.8125
THEN
  NODE : 20
  N : 125
  1 : 20.8%
  0 : 79.2%
IF ANZ_FARBEN EQUALS 1
AND K2_BEIMISCHUNG EQUALS 0
AND 127.0315 <= ZEIT_ZYLINDERREINIGUNG < 149.8125
THEN
  NODE : 21
  N : 232
  1 : 9.5%
  0 : 90.5%
IF ANZ_FARBEN EQUALS 2
AND K2_BEIMISCHUNG EQUALS 0
AND 127.0315 <= ZEIT_ZYLINDERREINIGUNG < 149.8125
THEN
  NODE : 22
  N : 221
  1 : 15.4%
  0 : 84.6%
IF ANZ_FARBEN IS ONE OF: 6 7
AND K2_BEIMISCHUNG EQUALS 0
AND 127.0315 <= ZEIT_ZYLINDERREINIGUNG < 149.8125
THEN
  NODE : 23
  N : 19
  1 : 62.2%
  0 : 37.8%
IF K2_BEIMISCHUNG EQUALS 0
AND ROLLEN_LAUF EQUALS 1
AND 149.8125 <= ZEIT_ZYLINDERREINIGUNG
THEN
  NODE : 25
  N : 558
  1 : 26.2%
  0 : 73.8%
Tree_rules.txt - Editor
Ergebnis des Entscheidungsbaums
AND K2_BEIMISCHUNG EQUALS 0
AND HERST_ID IS ONE OF: 001 002
AND ROLLEN_LAUF EQUALS 2
AND 149.8125 <= ZEIT_ZYLINDERREINIGUNG
THEN
  NODE : 392
  N : 16
  1 : 50.0%
  0 : 50.0%
IF 225 <= ALTER
AND MA_Rolleinspannung IS ONE OF: THOMAS, ANDRES THORNTON, ANDRE
TREVINO, ALEX
AND K2_BEIMISCHUNG EQUALS 0
AND HERST_ID IS ONE OF: 001 002
AND ROLLEN_LAUF EQUALS 2
AND 149.8125 <= ZEIT_ZYLINDERREINIGUNG
THEN
  NODE : 393
  N : 33
  1 : 12.1%
  0 : 87.9%
IF MASCH_ID < 9.5
AND 31.77001 <= MASCHINEN_PARS < 40.08824
AND GEWICHT < 59.75
AND HERST_ID EQUALS 005
AND ROLLEN_LAUF EQUALS 2
AND 149.8125 <= ZEIT_ZYLINDERREINIGUNG
THEN
  NODE : 400
  N : 3
  1 : 0.0%
  0 : 100.0%
IF 9.5 <= MASCH_ID < 11.5
AND 31.77001 <= MASCHINEN_PARS < 40.08824
AND GEWICHT < 59.75
AND HERST_ID EQUALS 005
AND ROLLEN_LAUF EQUALS 2
AND 149.8125 <= ZEIT_ZYLINDERREINIGUNG
THEN
  NODE : 401
  N : 8
  1 : 87.5%
  0 : 12.5%
IF 11.5 <= MASCH_ID
AND 31.77001 <= MASCHINEN_PARS < 40.08824
AND GEWICHT < 59.75
AND HERST_ID EQUALS 005
AND ROLLEN_LAUF EQUALS 2
AND 149.8125 <= ZEIT_ZYLINDERREINIGUNG
THEN
  NODE : 402
  N : 13
  1 : 15.4%
  0 : 84.6%
  
```

Regeln aus dem Entscheidungsbaummodell können zur weiteren Verwendung exportiert werden.







Qualitätskontrolle industrieller Prozesse mit dem SAS Enterprise Miner

INFORMATION WORKS

Fragen?

Claudia Tückmantel

KSFE Potsdam, 20.-21.02.2003

Posterpräsentation: Automatisierung statistischer Verfahren in der Pharmaindustrie am Beispiel bioWORKS™

- § Generische Lösung zur Berechnung von Bioäquivalenzen in Konformität mit der Richtlinien der FDA
- § Entwickelt in Kooperation mit Prof. Dr. J. Hartung, der Universität Dortmund
- § Programmiert in SAS
- § Leicht Implementierbar in bereits vorhandene biometrische SAS Infrastrukturen

