

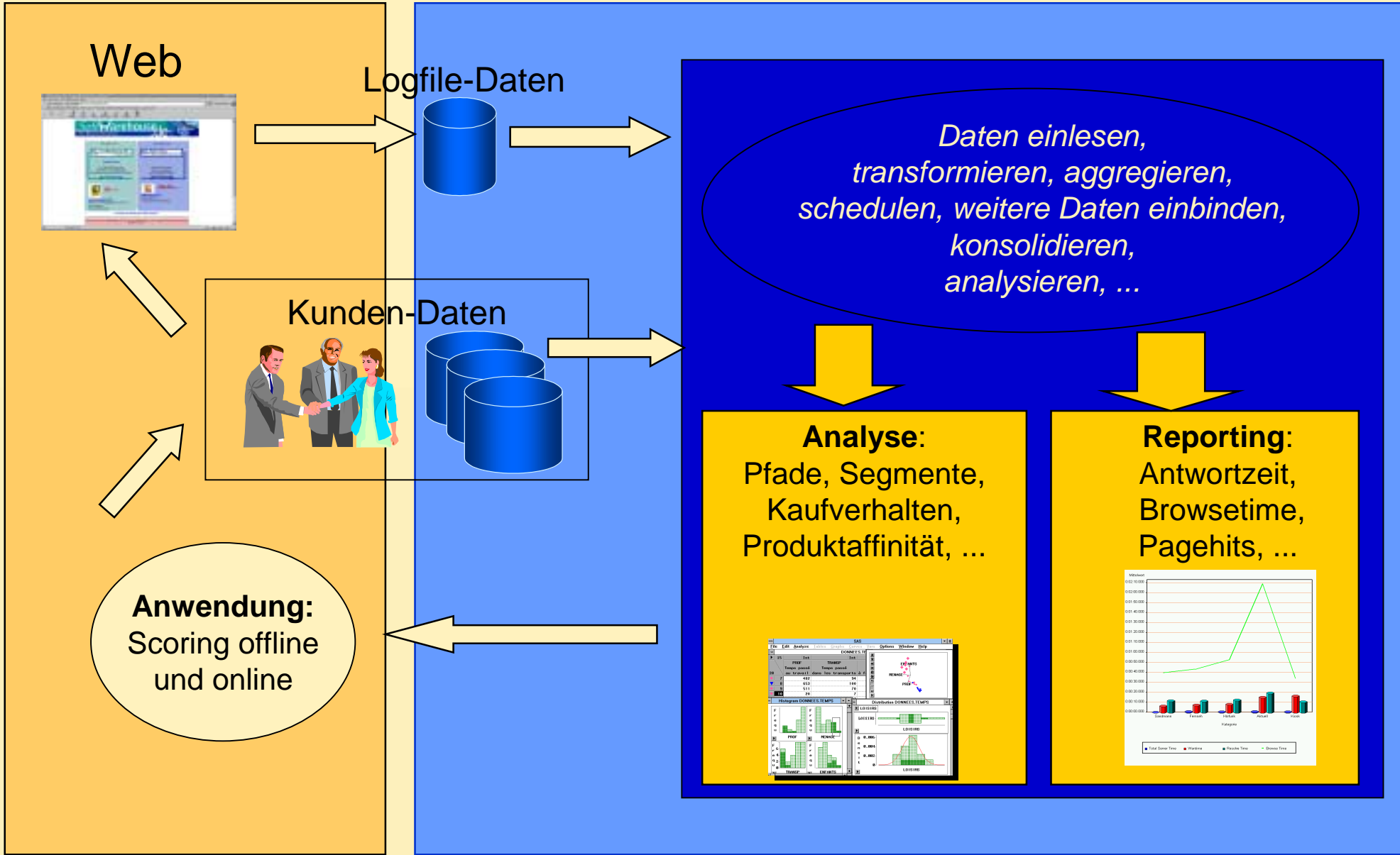


# **Web Mining – effektive Analyse des Nutzer- Verhaltens im Internet**

Dr. Frank Säuberlich  
Business Unit e-Intelligence  
SAS Deutschland

# Agenda

1. Begriffsdefinition e-Intelligence:
  - Systemdimension
  - Angebotsdimension
  - Kundendimension
2. Web Mining
3. Praxisbeispiele mit dem SAS Enterprise Miner™:
  - a) Web Mining bei einem Online-Shop
  - b) Web Enabled Credit Scoring



# 1. Zum Begriff „e-Intelligence“

Das Kunstwort „e-Intelligence“ bezeichnet Maßnahmen und Verfahren, mit deren Hilfe Wissen über e-Business Aktivitäten aufgebaut werden kann. Es lassen sich drei Wissensdimensionen unterscheiden:

- Systemdimension,
- Angebotsdimension,
- Kundendimension.

# Die Systemdimension

Messung der Performance und Verfügbarkeit des Web-Angebotes mit dem Ziel:

- Analyse und Gewährleistung von Antwortzeiten,
- Hohe Web Server Verfügbarkeit durch gezieltes Performance Management und Kapazitätsplanung,
- Minimierung von Fehlern:
  - z.B. Page not found - 404 error.

# Die Angebotsdimension

Optimierung des Web-Auftrittes durch Erfassen des Zugriffsverhaltens der Web-Besucher:

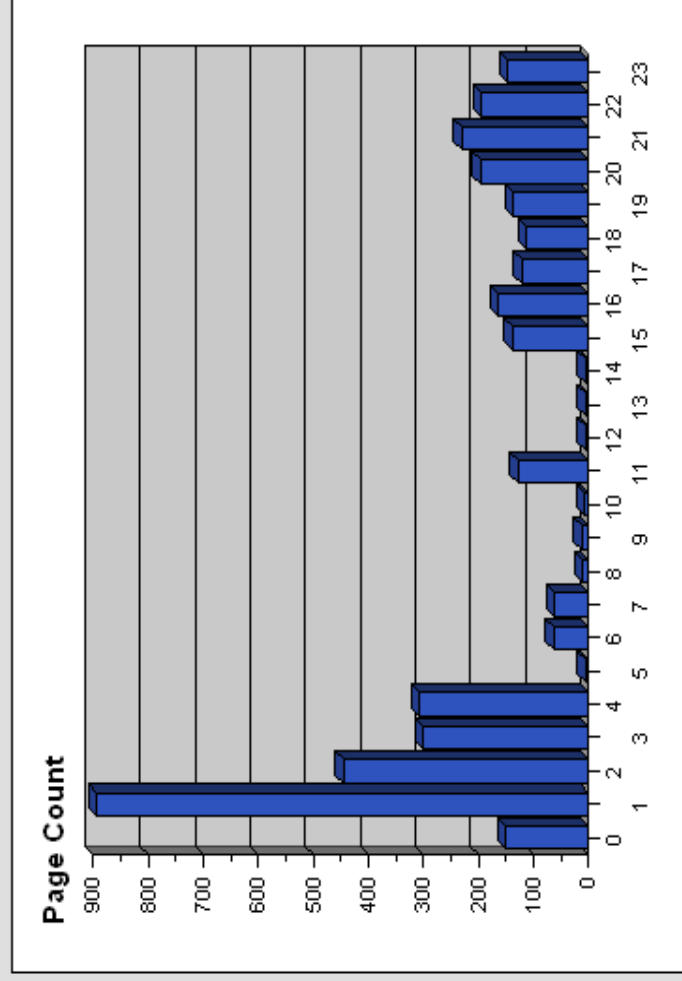
- Seitenzugriff (Hits, Pageviews),
- Ein- und Ausstiegsseiten,
- Verweildauer,
- Konversionsrate,
- ...



### Web Traffic Analysis

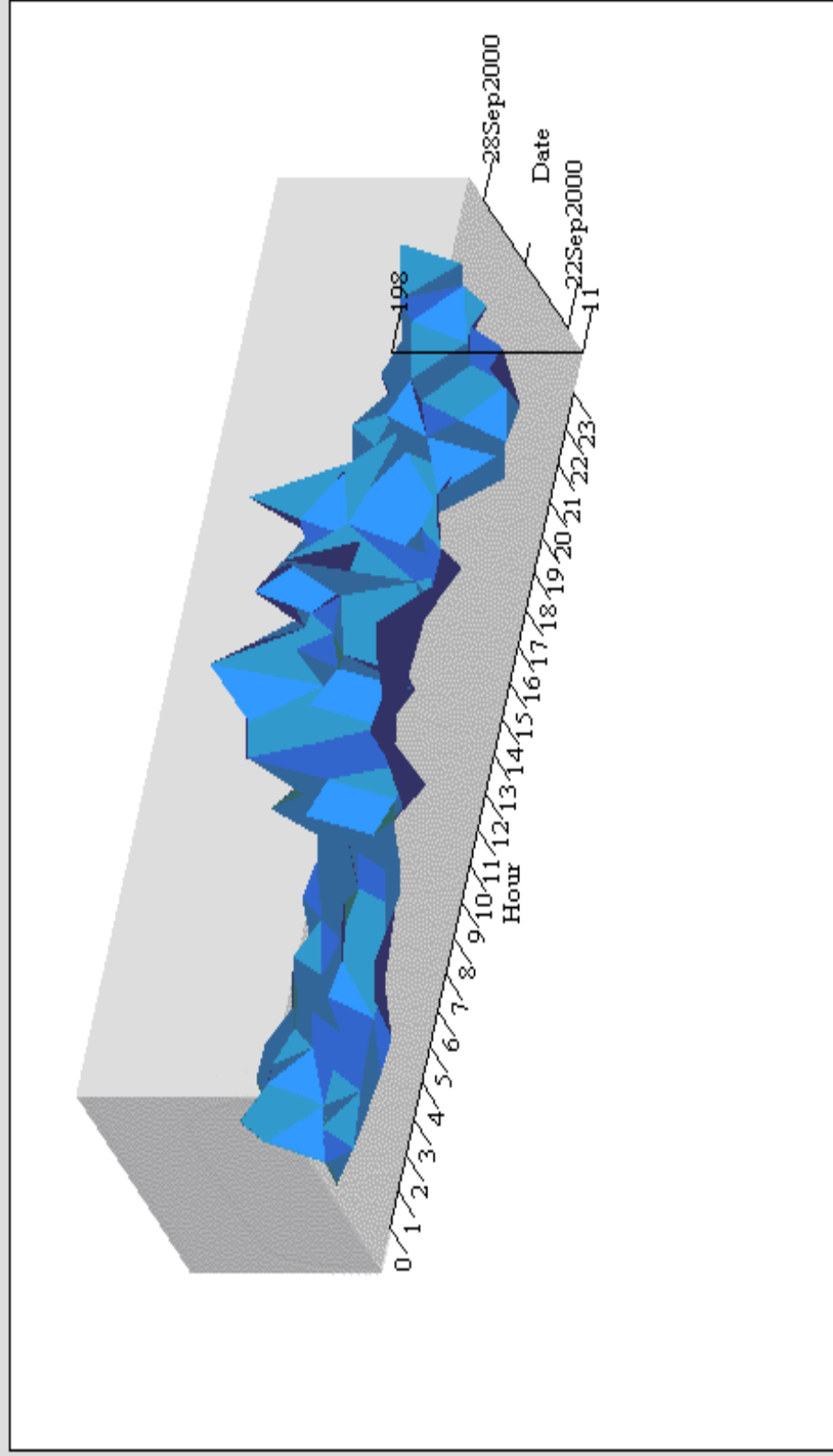
- [Daily Overview](#)
- [Weekly Overview](#)
- [Pages Top 10](#)
- [Pages Top 10 with Bytes](#)
- [Pages by Day](#)
- [Pages by Week](#)
- [Pages by Month](#)
- [Pages Hourly Profile](#)
- [Pages Hourly by Site](#)
- [Pages Hourly by Server](#)
- [Pages Daily/Hourly Profile](#)
- [Pages by Bad Status Code](#)
- [Pages Not Found](#)
- [URL Level 1 Overview](#)
- [Hits by Day](#)
- [Hits by Week](#)
- [Hits by Month](#)
- [Hits Hourly Profile](#)
- [Hits Daily/Hourly Profile](#)
- [Hits per Page](#)
- [Bytes Sent by Day](#)
- [Bytes Sent & Received by Day](#)
- [Bytes Sent by Week](#)
- [Bytes Sent Hourly Profile](#)
- [Bytes Sent Daily/Hourly Profile](#)
- [Bytes Sent per Page](#)
- [Status Codes](#)
- [Bad Status Codes](#)

## Total Page Count by Hour for January 10, 2000



Generated by Webhournd on 26SEP2000 Report SHTPVB

## Sessions by Date and Hour







- Bytes Sent by Day
- Bytes Sent & Received by Day
- Bytes Sent by Week
- Bytes Sent Hourly Profile
- Bytes Sent Daily/Hourly Profile
- Bytes Sent per Page

- Status Codes
- Bad Status Codes

- Explore Top Pages
- Explore Top Entry Points
- Explore Top Exit Points
- Explore Top Directories
- Explore Top File Types
- Explore Top Status Codes
- Explore Site & Server Trends

**Visitor Analysis**

- Explore Top Visitors
- Explore Top Referring Sites

**Browsers & Platforms**

- Platform Overview
- Platform Breakdown

- Explore Top Browsers
- Explore Top Platforms

**Visual Exploration Demo**

- Web Site Hits
- Search Term Hits

## Site and Server Trends

Site Name	W3SVC5				
	Unique Sessions	Bytes Sent	File Count	Page Count	Total Bytes
<b>Date</b>	<b>Unique</b>	<b>Sum</b>	<b>Sum</b>	<b>Sum</b>	<b>Sum</b>
06JAN2000	368	52,289,795	11,205	4,104	52,269,795
07JAN2000	246	56,108,427	10,323	3,629	56,108,427
08JAN2000	273	46,123,621	7,745	2,653	46,123,621
09JAN2000	305	60,317,777	8,797	3,328	60,317,777
10JAN2000	225	39,314,755	8,832	3,847	39,314,755
<b>TOTAL</b>	<b>1410</b>	<b>254,134,375</b>	<b>46,902</b>	<b>17,561</b>	<b>254,134,375</b>

**Filter by**

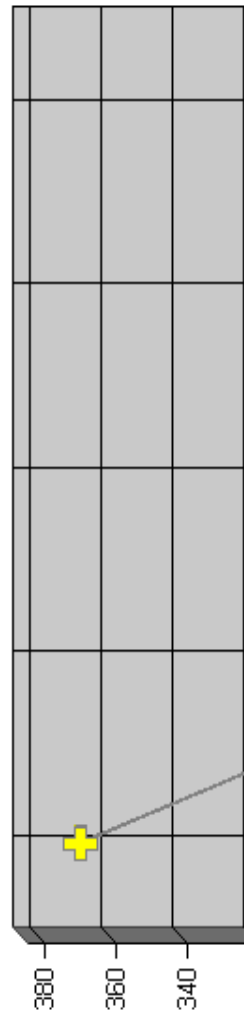
**Date:** 06JAN2000 07JAN2000

**Day of Week:** Sunday Monday

**Hour:** 0 1

Apply Filter

W3SVC5, Unique Sessions, Nunique



# Die Kundendimension

Analyse und Prognose des Besucherverhaltens:

- Welche Pfade werden häufig durchlaufen?
- Was sind die Eigenschaften und Charakteristika von Käufern einer bestimmten Produktkategorie?
- Existieren Kundensegmente mit unterschiedlichem Kaufverhalten?

→ Anwendung von Data Mining-Verfahren:  
„Web Mining“!

## 2. Web Mining

- Web-Daten sind nur eine weitere Datenquelle für Data Mining-Analysen.
- Daten sind sehr reich an Informationen.
- Pfad- und Navigationsanalysen können durchgeführt werden.
- Folgende Data Mining-Verfahren kommen zum Einsatz:
  - Assoziations- und Sequenzanalysen,
  - Segmentierungsverfahren,
  - Vorhersagemodelle.

# Datenvorbereitungsmaßnahmen

Datenvorbereitungsmaßnahme	Problemstellung
Seiten-Identifizierung (Data Cleaning)	<ul style="list-style-type: none"> <li>• Nur direkte Requests / Aktionen der Benutzer sind von Interesse</li> <li>• Sogenannte Auxiliary Resources (z.B. automatischer Aufruf von Bild-Dateien) müssen identifiziert und entfernt werden</li> </ul>
Benutzer-Identifikation	<ul style="list-style-type: none"> <li>• IP-Adressen identifizieren Benutzer nicht eindeutig (z.B. mehrere Benutzer greifen von der gleichen Maschine aus zu)</li> <li>• Proxy-Rechner</li> <li>• Dynamische IP-Nummern</li> </ul>
Pfadvervollständigung	<ul style="list-style-type: none"> <li>• Anfragen des Benutzers werden aus dem lokalen Cache des Browsers bedient</li> <li>• Verwendung von Bookmarks</li> </ul>

# Web Mining-Verfahren

Aufgabenstellung	Data Mining-Verfahren
Analyse von Nutzerpfaden	<ul style="list-style-type: none"> <li>• Assoziationsanalyse</li> <li>• Sequenzanalyse</li> </ul>
Erkennen von Nutzertypen	<ul style="list-style-type: none"> <li>• Clusteranalyse</li> <li>• Kohonen SOM</li> </ul>
Beschreiben / Vorhersage von Benutzerverhalten	<ul style="list-style-type: none"> <li>• Entscheidungsbaumverfahren</li> <li>• Regression</li> <li>• Neuronale Netze</li> </ul>

# Analyse von Nutzerpfaden

- Was macht einen „erfolgversprechenden“ Pfad aus?
- Welches Navigationsverhalten führt zu einem Kauf?
- Welche Navigationspfade führen häufig zum Verlassen des Web-Angebotes?
- Gehen Besucher Umwege auf der Site oder werden wichtige Bereiche zu wenig besucht?

## Erkennen von Benutzertypen

- Gibt es Kundensegmente mit unterschiedlichen Interessen bzw. Kaufverhalten?
- Sollte ich mein Angebot speziell auf die einzelnen Kundensegmente anpassen (Personalisierung)?
- Welche Kundengruppe bringt mir den größten Profit?

# Beschreiben / Vorhersage von Benutzerverhalten

- Was unterscheidet einen Besucher von einem Käufer?
- Wie erkenne ich die profitabelsten Kunden?
- Welche Produkt(kategorie) werden diese als nächstes kaufen?
- Wie kann ich die profitabelsten Kunden langfristig an mich binden?



## 3. Praxisbeispiele mit dem Enterprise Miner™

a) Web Mining bei einem Online-Shop:

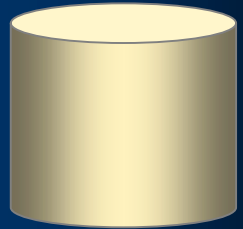
- Lassen sich aus den Logfiles konkrete Benutzerprofile ableiten?
  - Was sind die Einflußfaktoren auf einen Kauf im Online-Shop?
  - Kann Data Mining konkrete Hinweise für die Shop-Optimierung bieten?
- Einsatz von Data Mining-Verfahren mit dem Enterprise Miner™

Rohdaten

Sequenz -  
Analyse

Variablen  
Transformation  
& Integration

Data Mining  
Modelle  
SEMMA



Data Warehouse

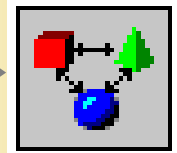
1 Zielvariable



Logfiles

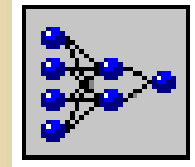
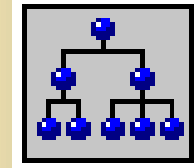
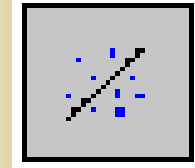
6 Variablen

Web Server

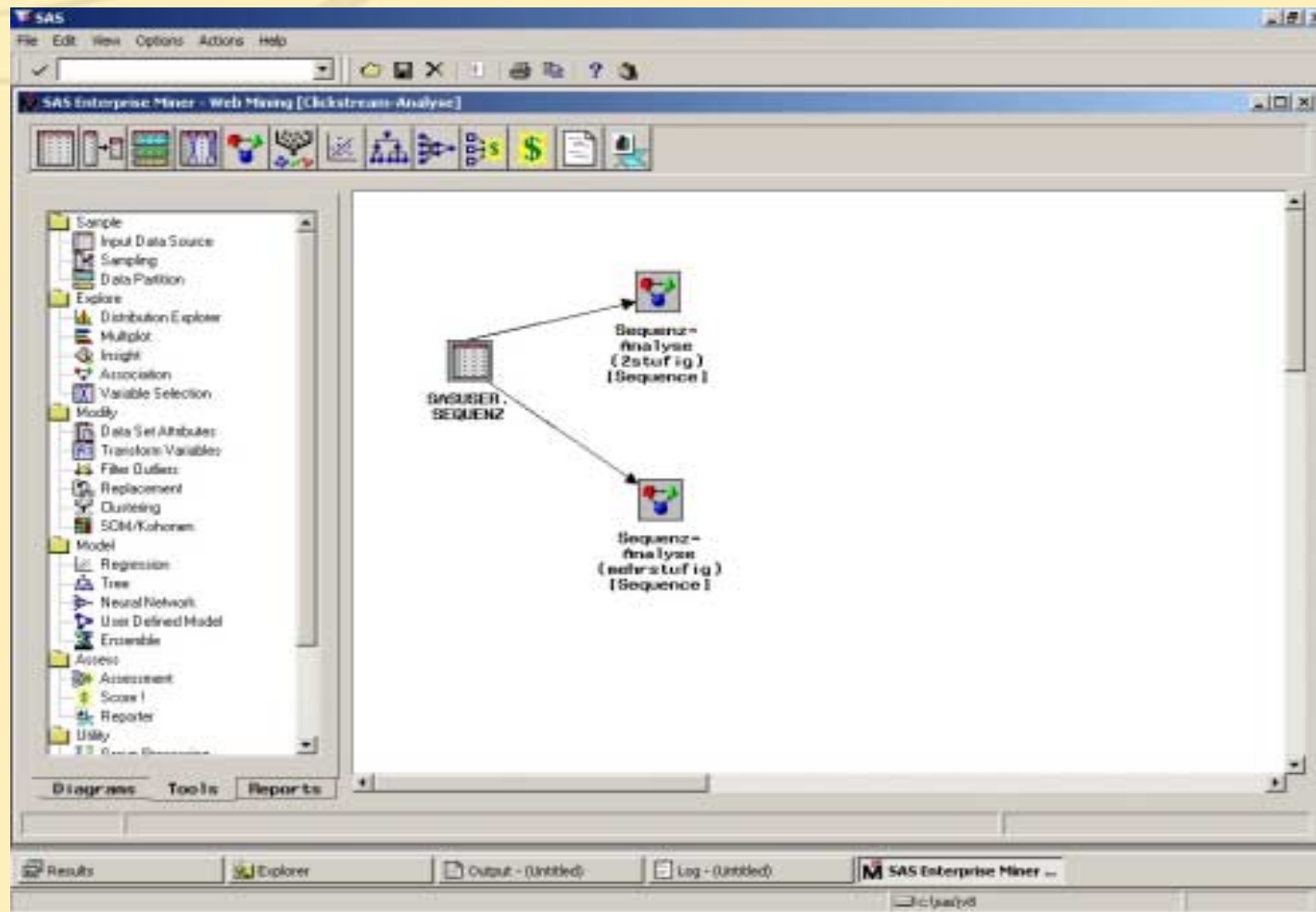


10 Sequenzvariablen

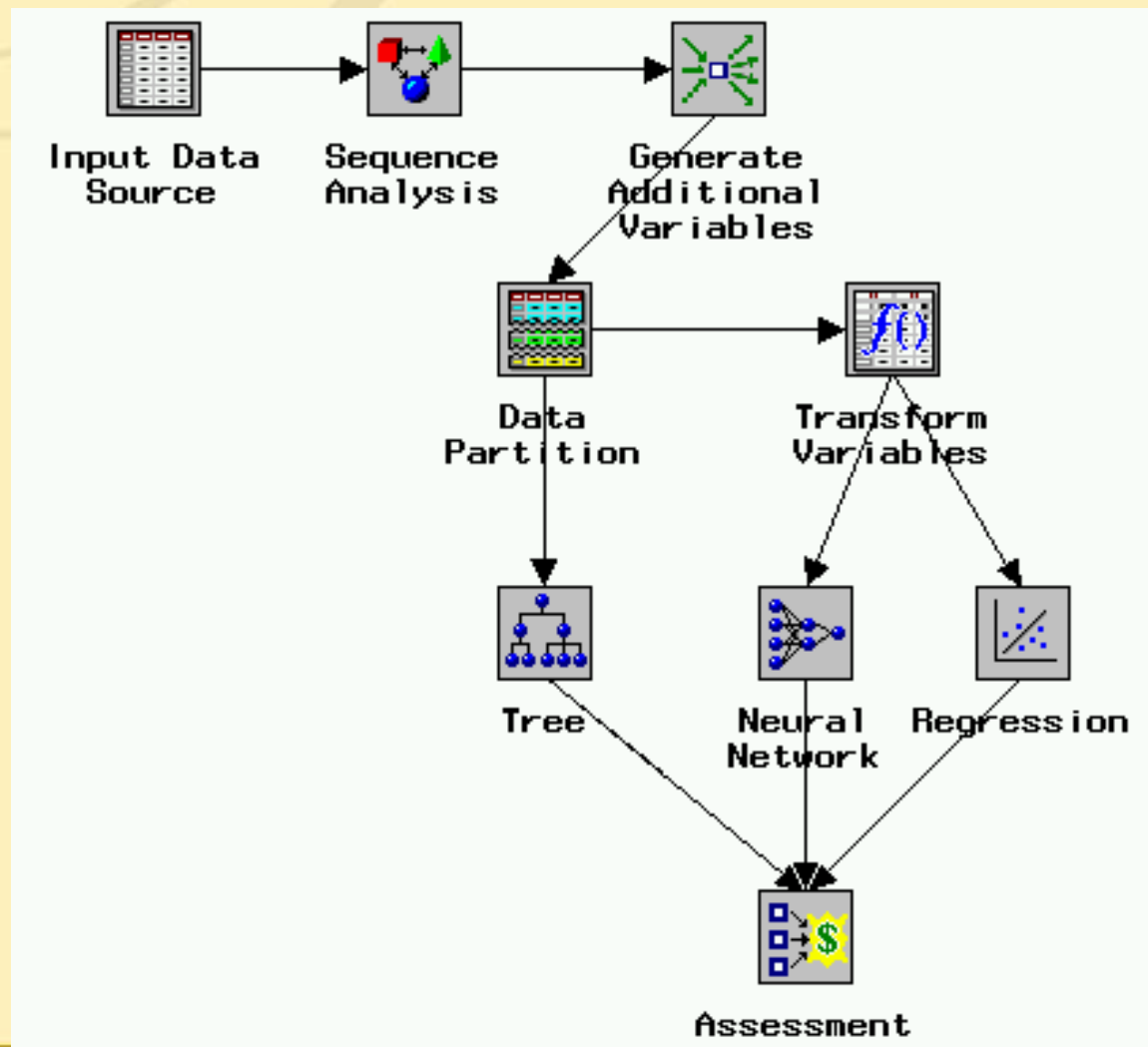
Flat File  
51  
Variablen



# Der SAS Enterprise Miner™



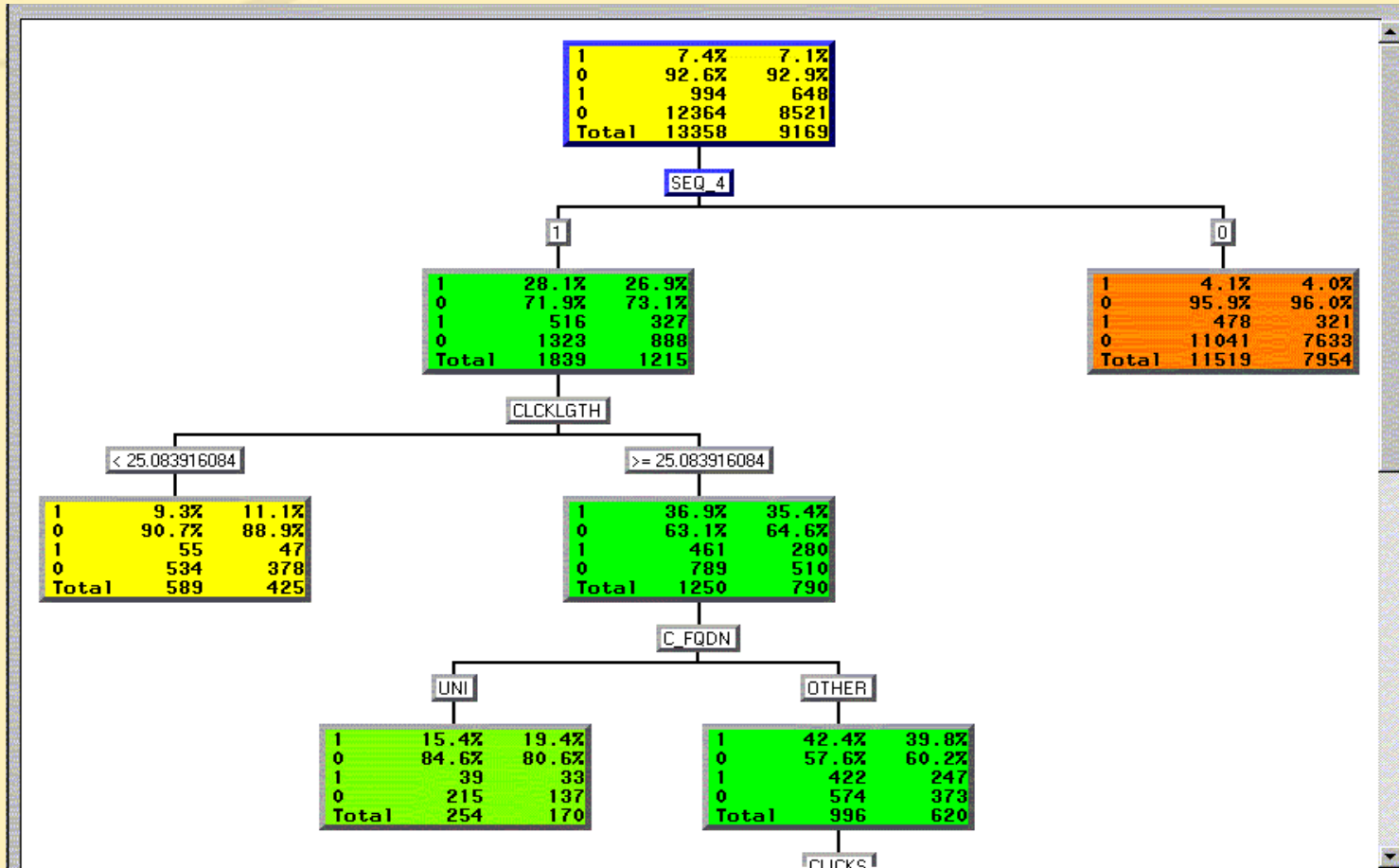
# Prozessfluss-Diagramm



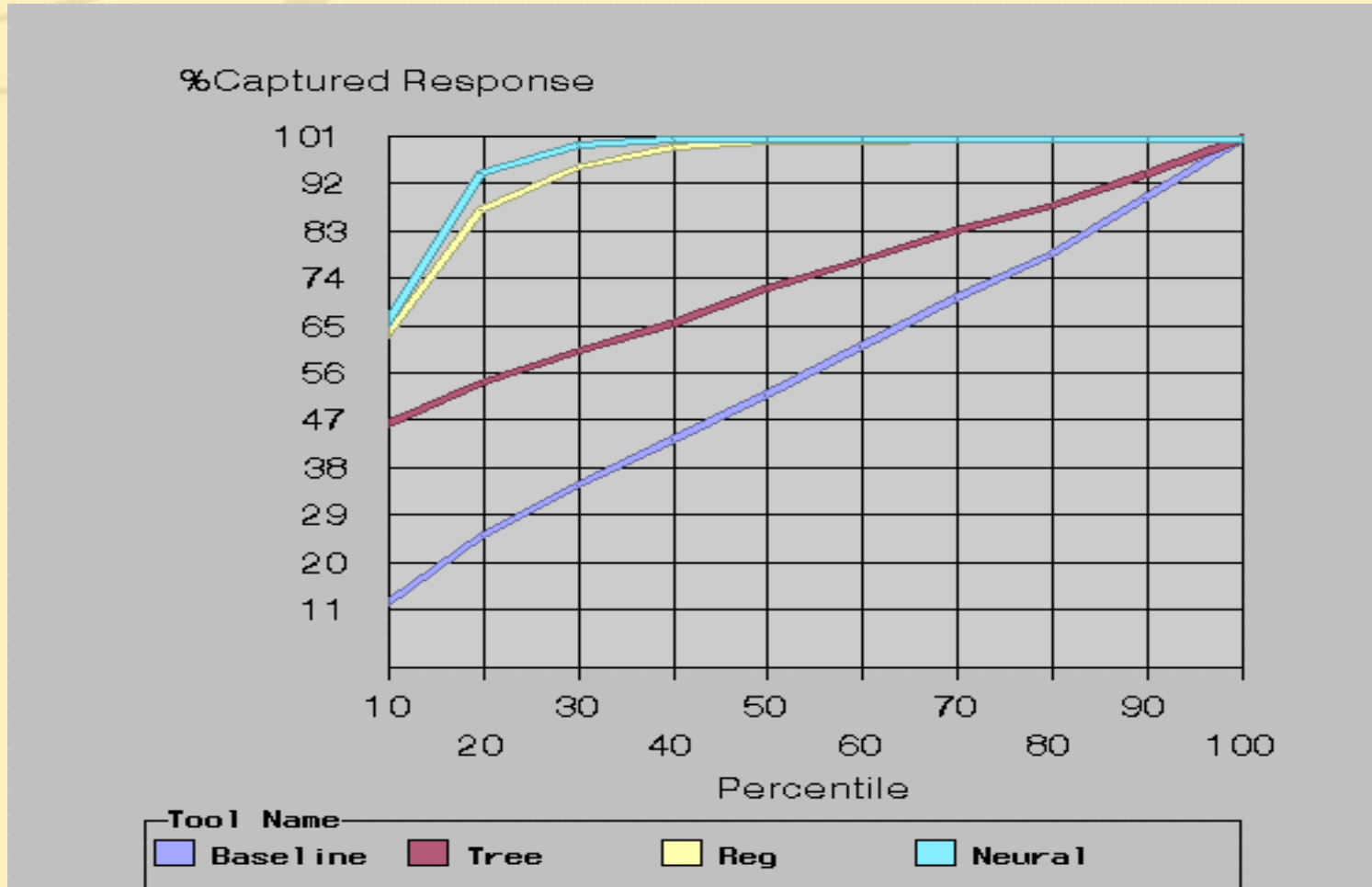
# Ausgewählte Ergebnisse der Sequenzanalyse

#	Support (%)	Confidence (%)	Rule
1	15.5	36.8	<i>login ⇒ register</i>
2	13.4	31.9	<i>login ⇒ login</i>
3	12.3	38.5	<i>addcart ⇒ login</i>
4	11.2	28.1	<i>addcart ⇒ register</i>
5	0.7	4.6	<i>pay_req ⇒ help</i>
6	0.3	3.6	<i>news ⇒ pay_res</i>

# Entscheidungsbaum



# Vorhersagequalität



## b) Anwendung von Data Mining Modellen am Beispiel „Webenabled Credit Scoring“

Die Situation im Kreditgewerbe hat sich grundlegend gewandelt:

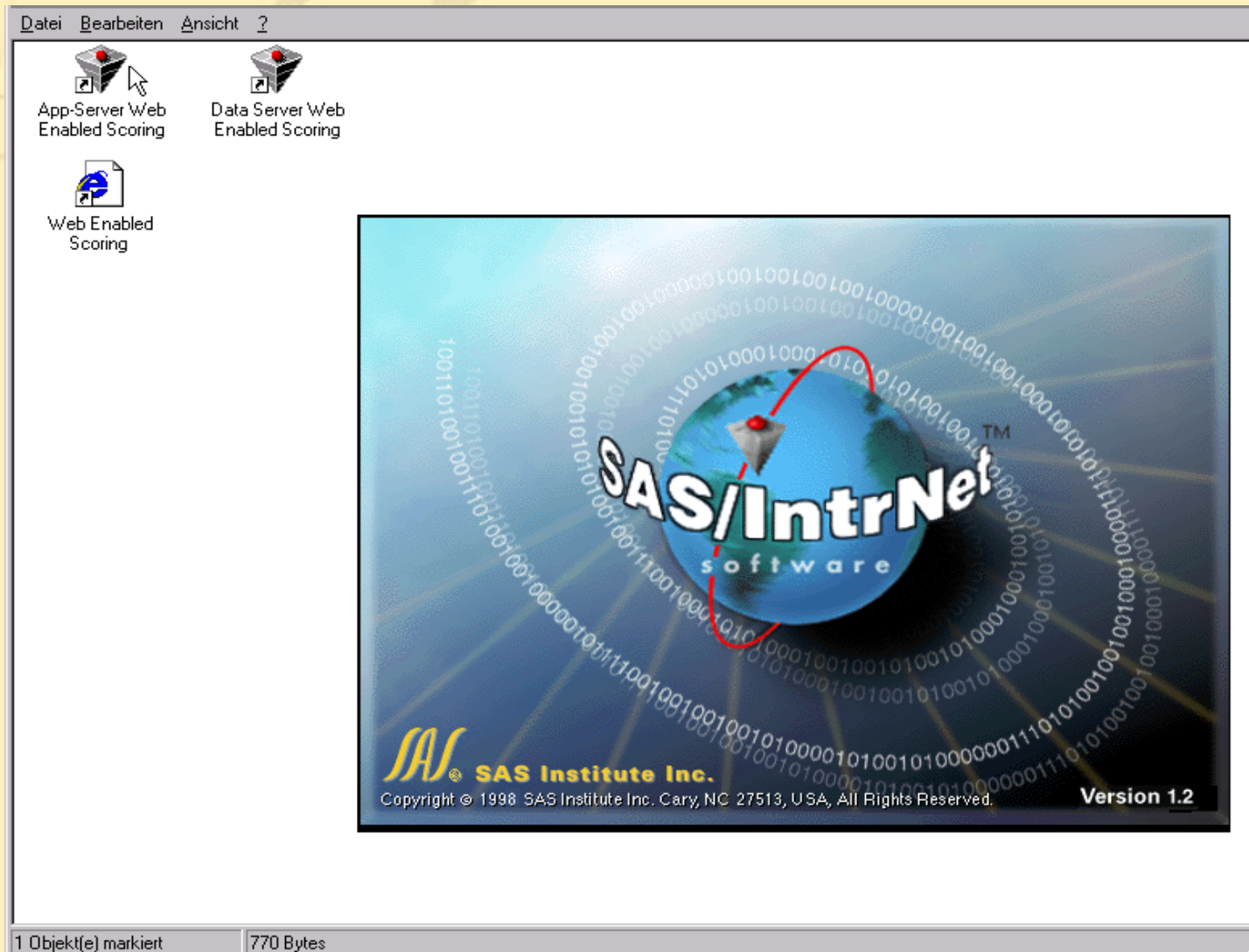
- Neben den Banken als den klassischen Kreditgebern agieren zahlreiche andere Unternehmen am Markt wie z.B. Leasing-Gesellschaften, Versandhandelsketten, Service-Provider im Mobilfunkgeschäft etc.
- Darum ist heute jeder Kreditgeber aufgrund der verschärften Konkurrenzsituation und der neuen Absatzkanäle (Internet) gezwungen, über die Annahme oder Ablehnung eines Kreditgesuchs in kürzester Zeit bei gleichzeitiger Risikominimierung zu entscheiden.



## Scorekartenentwicklung mit dem SAS Enterprise Miner™

- Einsatz alternativer Modellierungstechniken, z.B.
  - Diskriminanzanalyse,
  - Entscheidungsbäume,
  - Neuronale Netze.
- Vergleich Modellgüte traditionelle Scorekarte und andere Modellierungstechniken.
- Berücksichtigung von Interaktionseffekte.
- Online Scoring über Internet, WAP / Mobilfunk.

# Online-Scoring zur Kreditbewertung



# Online-Scoring zur Kreditbewertung

On-line scoring - Microsoft Internet Explorer von SAS Institute GmbH

Datei Bearbeiten Ansicht Favoriten Extras 2

Zurück Vorwärts Abbrechen Aktualisieren Startseite Suchen Favoriten Verlauf E-Mail Drucken

Adresse [http://localhost/sas\\_web/online\\_scoring.htm](http://localhost/sas_web/online_scoring.htm) Wechseln zu Links

---

**Details of the Customer**

Age:

Gender:  Male  Female Telephone:  Yes  No

Marital:  Time in current job:

Copyright information goes here.  
Last revised: November 04, 1998

---

**Details of the Loan:**

Purpose of the Loan:

Amount of the Loan:  Duration of the Loan (months):

---

**Account History**

Balance of checking:  Other installments:

Balance of savings:  Credit History:

Fertig Local intranet

Start Start Spawner Webenabled credit scoring SAS SAS/SHARE Data Server On-line\_scoring - Micr... 14:49

# Online-Scoring zur Kreditbewertung

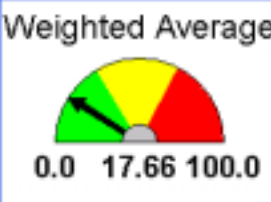
Scoring Output - Microsoft Internet Explorer von SAS Institute GmbH

Adresse: http://localhost/og-bin/broker.exe

by Enterprise Miner: **Probability of Default for Applicant (%)** *Live*

**Application approved !!**

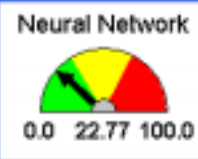
**Weighted Average**



0.0 17.66 100.0

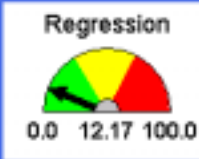
Weighted Average=(0\*NN+2\*REG+0T)/6

**Neural Network**



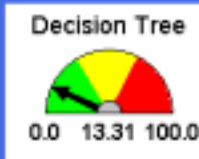
0.0 22.77 100.0

**Regression**



0.0 12.17 100.0

**Decision Tree**



0.0 13.31 100.0

Click on a dial for more detail

**Customer details:**

Age: 30 Gender: Female Marital Status: Married Employment status: < 7 years

Windows taskbar: Start, Start Spawner, Webenabled credit oc..., SAS, SAS/SHARE, Lotus ScreenCam, Scoring Output - ... 14:50

# Was bringt Web Mining?

- Erhöhung der Konversionsrate von anonymen Besuchern zu Kunden,
- Umsatzerhöhung durch Cross- und up-Selling im e-Shop,
- Erhöhung der Bannerwirkung durch personalisierte Werbung,
- Verbesserte Kundenbindung ==> Kundenloyalität,
- Multichannel-Management durch optimierte Kundenansprache,
- Optimierung des Web-Auftritts (Angebot, Seiten),
- Personalisierung von Produkten, Angeboten und Werbung.



Dr. Frank Säuberlich  
Business Unit e-Intelligence  
SAS Deutschland  
[Frank.Saeuberlich@ger.sas.com](mailto:Frank.Saeuberlich@ger.sas.com)