



Mining the Web

**Analyse von Benutzerpfaden und Nutzertypen
im Internet**

Dr. Frank Säuberlich
Business Unit CRM Solutions
SAS Deutschland

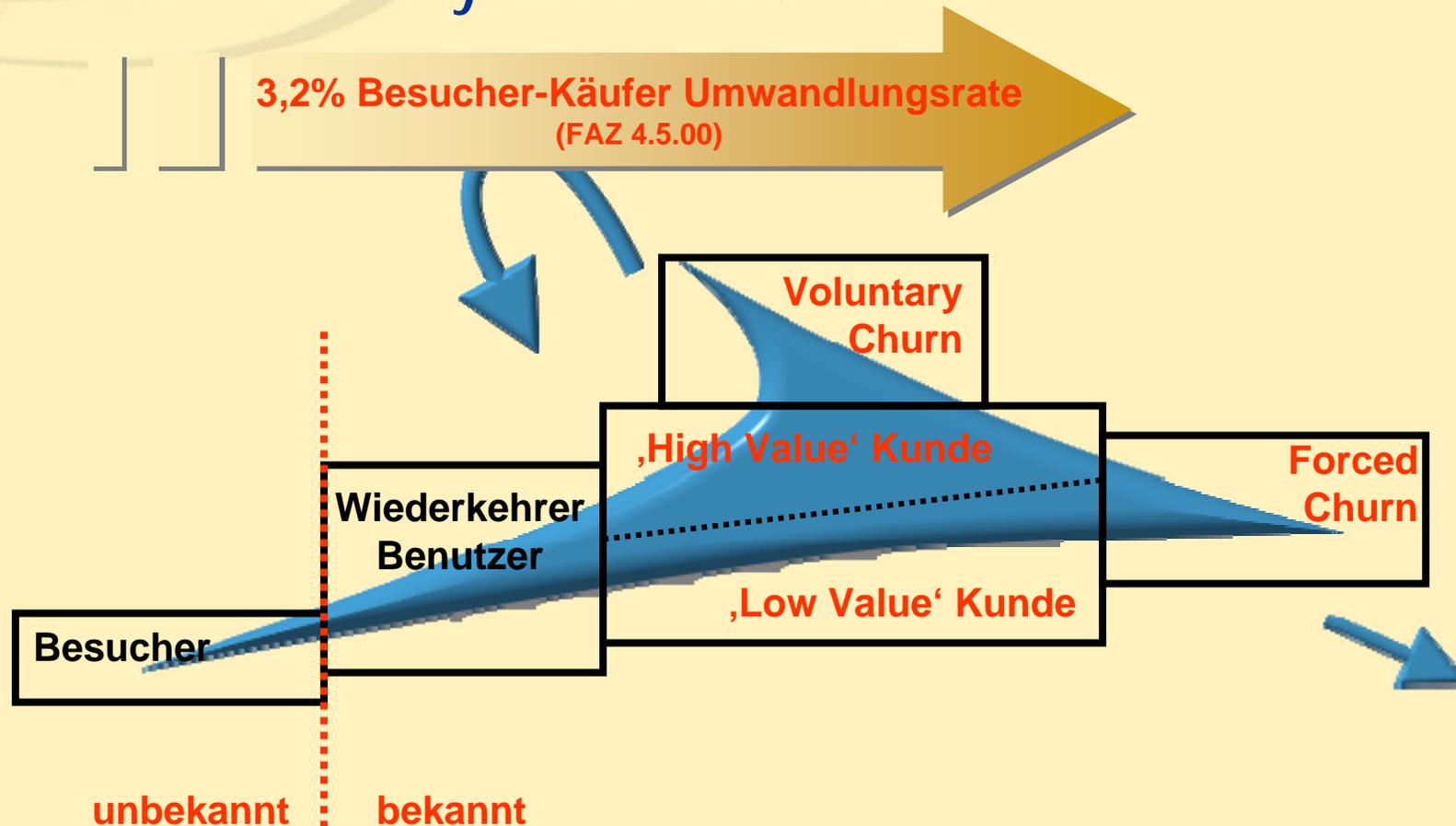
The Power to Know.

Agenda

1. Einleitung:
 - Der Lebenszyklus eines e-Kunden
 - Begriffsdefinition e-Intelligence
2. Web Mining
 - Datenvorbereitungsmaßnahmen
 - Web Mining-Verfahren
3. Praxisbeispiele
 - Web Mining bei einem Online-Shop
 - Vorhersage von Click-Aktionen

1. Einleitung

Der Lebenszyklus eines *e*-Kunden



Aus Daten entsteht *e*-Wissen



e-Intelligence und die drei Dimensionen

- ▶ Systemdimension
- ▶ Angebotsdimension
- ▶ Kundendimension



e- Intelligence und die Systemdimension



Systemdimension

- ▶ Performance
- ▶ Verfügbarkeit
- ▶ Systemmanagement

WEBOPTIMIERUNG

- ▶ Analyse und Gewährleistung von Antwortzeiten
- ▶ Hohe Webserver-Verfügbarkeit durch gezieltes Performance Management und Kapazitätsplanung
- ▶ Minimierung von Fehlern
 - ▶ z.B. Page not found - 404 error

e- Intelligence und die Angebotsdimension

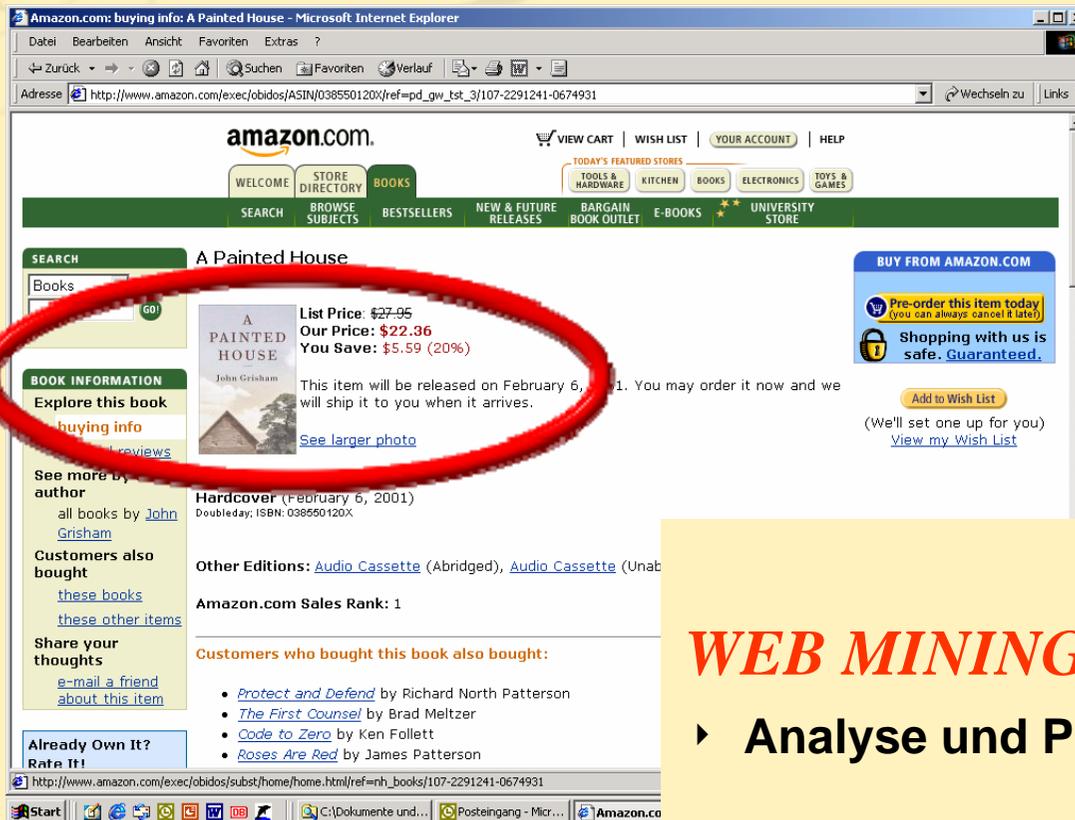


- ▶ **Systemdimension**
- ▼ **Angebotsdimension**
 - ▶ **Seitenzugriff**
 - ▶ **Ein- und Ausstieg**
 - ▶ **Verweilzeit**
 - ▶ **Konversionsrate**

WEB REPORTING

Erfassen und Darstellung des Zugriffsverhaltens im Web

e- Intelligence und die Kundendimension



▼ Kundendimension

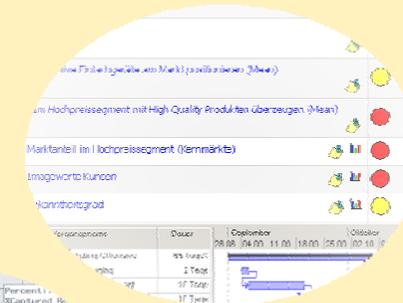
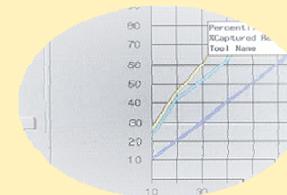
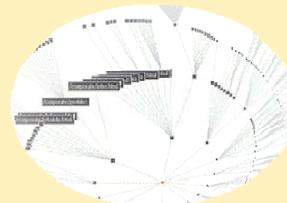
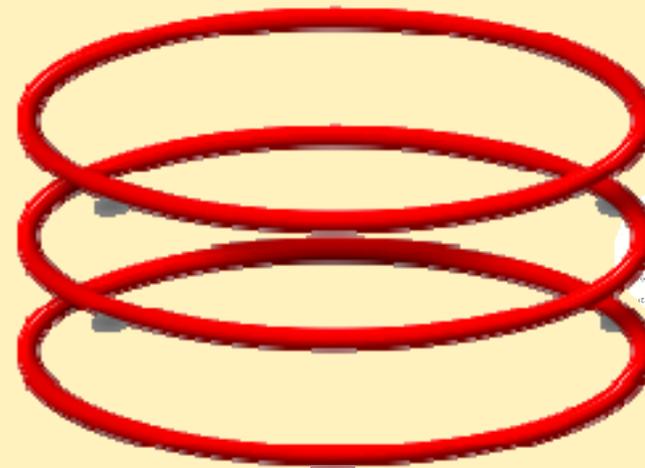
- ▶ Besucherpfade
- ▶ Besucher -> Nutzer
- ▶ Verhaltensprofile
- ▶ Kundensegmente
- ▶ Personalisierung
- ▶ Individuelle Angebote

WEB MINING

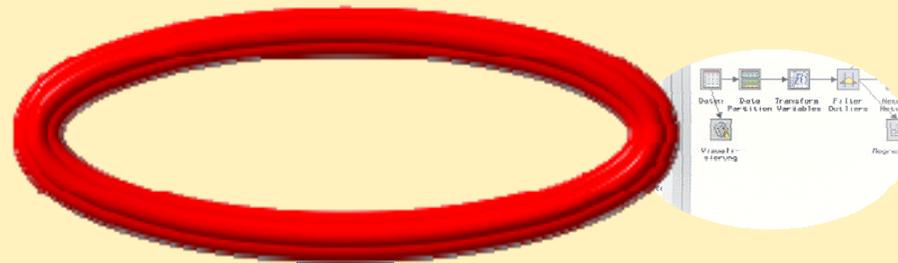
- ▶ Analyse und Prognose des Besucherverhaltens

e-Business und die drei Dimensionen

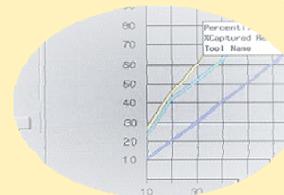
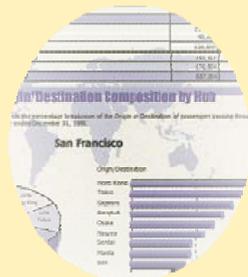
- ▶ **Systemdimension**
- ▶ **Angebotsdimension**
- ▶ **Kundendimension**



e-Intelligence



... Data in – Knowledge out



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

Web Mining-Verfahren

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

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 SAS Enterprise Miner™



- Web Mining bei einem Online-Shop für Zauberartikel:
 - Welche Produktkategorien werden häufig gemeinsam gekauft? Wie bewegen sich die Besucher im Shop?
 - Gibt es Besuchersegmente mit unterschiedlichen Interessen?
 - Was sind die Einflussfaktoren für das Interesse/den Kauf von Produkten einer bestimmten Produktkategorie (Zauberartikel)?
- Einsatz von Data Mining-Verfahren mit dem Enterprise Miner™



Praxisbeispiel – Analyse von Nutzerpfaden

SAS - [SAS Enterprise Miner - Jonglier [1] Analyse von Nutzerpfaden]

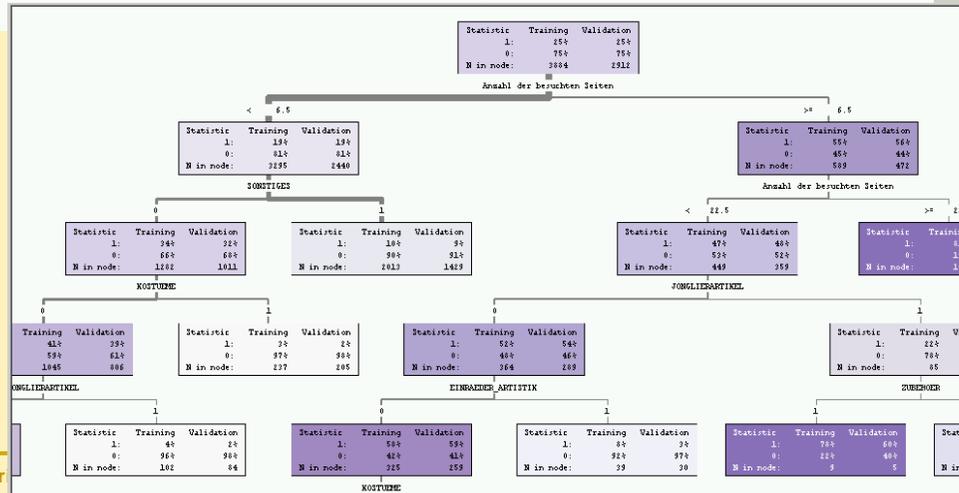
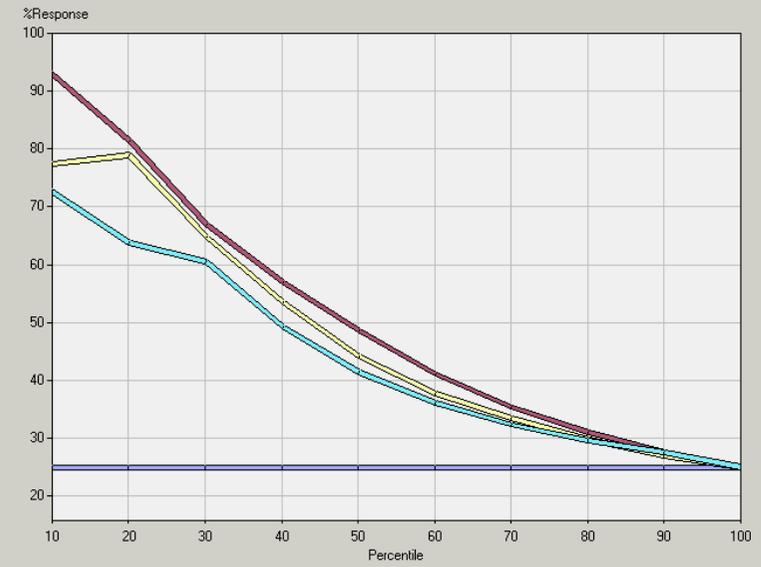
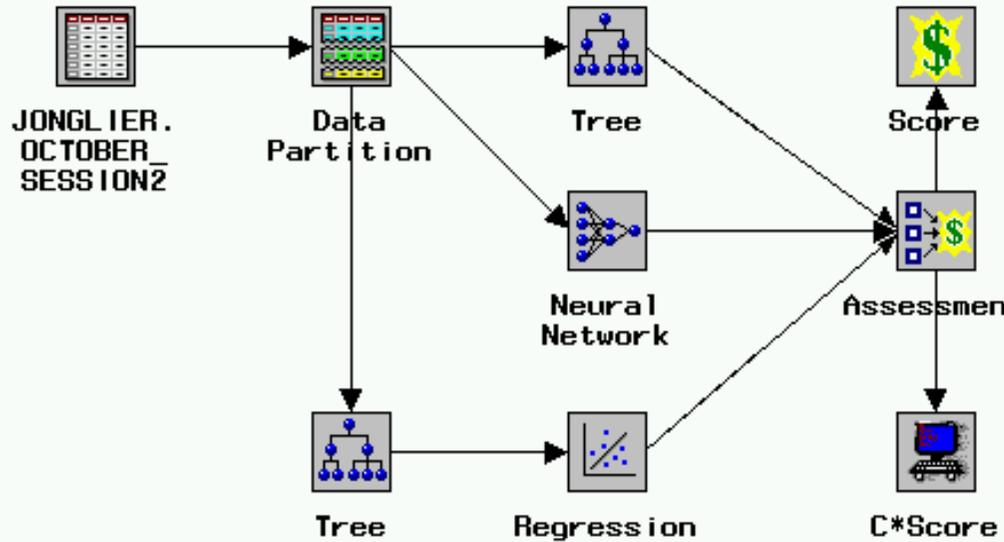
File Edit View Options Actions Help

Jonglier
 1) Analyse von Nutzerpfaden
 2) Erkennen von Nutzertypen
 3) Vorhersage von Nutzerverhalten

	Chain Length	Support(%)	Confidence(%)	Transaction Count	Rule
62	3	2.31	7.84	224	Sonstiges ==> Sonstiges ==> Pyroartikel
63	3	2.14	65.00	208	Neu ==> Sonstiges ==> Sonstiges
64	3	2.10	41.21	204	Schnaepchen ==> Schnaepchen ==> Zauberartikel
65	3	2.10	66.02	204	Buecher_Zeitschriften ==> Buecher_Zeitschriften ==> Buecher_Zeitschriften
66	3	2.07	73.09	201	Einraeder_Artistik ==> Einraeder_Artistik ==> Einraeder_Artistik
67	3	2.05	6.96	199	Sonstiges ==> Sonstiges ==> Buecher_Zeitschriften
68	3	2.04	36.26	198	Zauberartikel ==> Sonstiges ==> Zauberartikel
69	4	8.88	62.28	862	Sonstiges ==> Sonstiges ==> Sonstiges ==> Sonstiges
70	4	7.25	76.27	704	Zauberartikel ==> Zauberartikel ==> Zauberartikel ==> Zauberartikel
71	4	6.66	83.48	647	Sonstiges ==> Zauberartikel ==> Zauberartikel ==> Zauberartikel
72	4	6.21	80.61	603	Sonstiges ==> Sonstiges ==> Zauberartikel ==> Zauberartikel
73	4	4.14	29.05	402	Sonstiges ==> Sonstiges ==> Sonstiges ==> Zauberartikel
74	4	3.49	36.73	339	Zauberartikel ==> Zauberartikel ==> Zauberartikel ==> Sonstiges
75	4	3.16	41.04	307	Sonstiges ==> Sonstiges ==> Zauberartikel ==> Sonstiges
76	4	3.11	38.97	302	Sonstiges ==> Zauberartikel ==> Zauberartikel ==> Sonstiges
77	4	3.04	70.57	295	Sonstiges ==> Sonstiges ==> Schnaepchen ==> Schnaepchen
78	4	3.00	69.62	291	Sonstiges ==> Sonstiges ==> Schnaepchen ==> Sonstiges
79	4	2.82	78.51	274	Sonstiges ==> Schnaepchen ==> Schnaepchen ==> Schnaepchen
80	4	2.80	74.73	272	Schnaepchen ==> Schnaepchen ==> Schnaepchen ==> Sonstiges
81	4	2.70	75.07	262	Sonstiges ==> Schnaepchen ==> Schnaepchen ==> Sonstiges
82	4	2.62	62.56	254	Zauberartikel ==> Zauberartikel ==> Sonstiges ==> Sonstiges
83	4	2.54	63.01	247	Sonstiges ==> Sonstiges ==> Neu ==> Neu
84	4	2.49	65.76	242	Sonstiges ==> Zauberartikel ==> Sonstiges ==> Sonstiges
85	4	2.40	72.59	233	Jonglierartikel ==> Jonglierartikel ==> Jonglierartikel ==> Jonglierartikel
86	4	2.33	85.61	226	Sonstiges ==> Jonglierartikel ==> Jonglierartikel ==> Jonglierartikel
87	4	2.30	65.20	223	Sonstiges ==> Schnaepchen ==> Sonstiges ==> Sonstiges
88	4	2.15	53.32	209	Sonstiges ==> Sonstiges ==> Neu ==> Sonstiges
89	4	2.14	15.03	208	Sonstiges ==> Sonstiges ==> Sonstiges ==> Schnaepchen
90	4	2.11	64.06	205	Zauberartikel ==> Sonstiges ==> Sonstiges ==> Sonstiges

Diagrams Tools Reports

Praxisbeispiel – Vorhersage von Nutzerverhalten



Vorhersage von Click-Aktionen

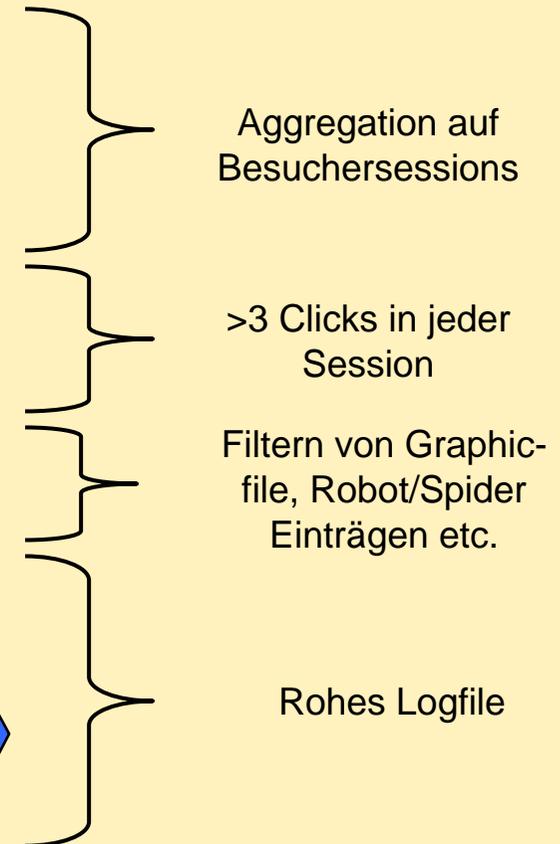
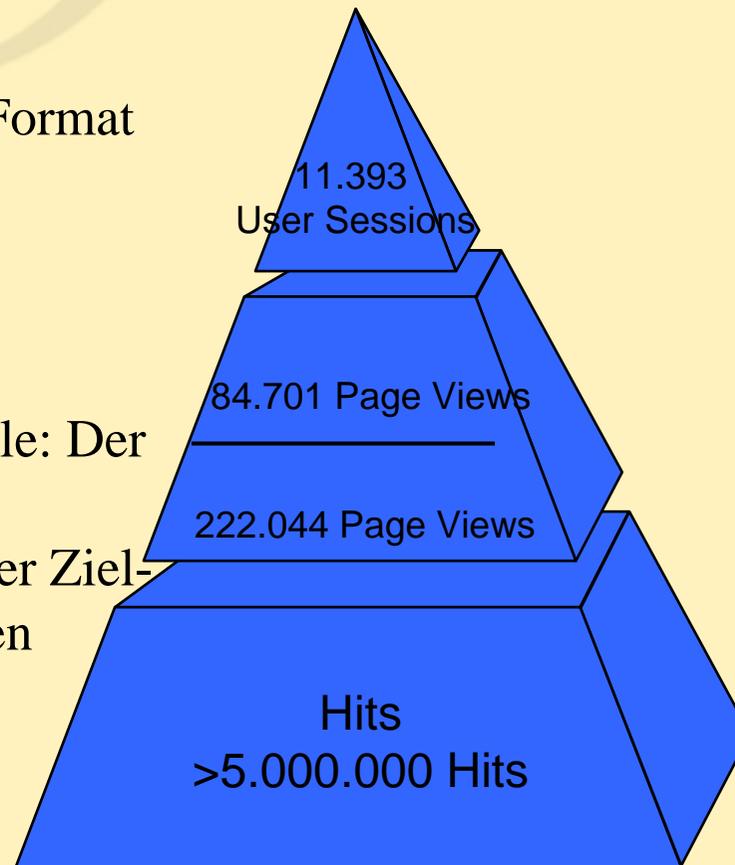
- Anwendung von Web Mining-Prognosemodellen.
- Schon auf anonymen Clickstreams von Besuchern können valide Vorhersagen ihres zukünftigen Clickverhaltens durchgeführt werden.
- Praxis-Beispiel: Analyse der ersten drei Clicks von Besuchern zur Vorhersage des vierten Clicks → Anwendung von Prognoseverfahren im Enterprise Miner™.

Anwendung auf reale Daten

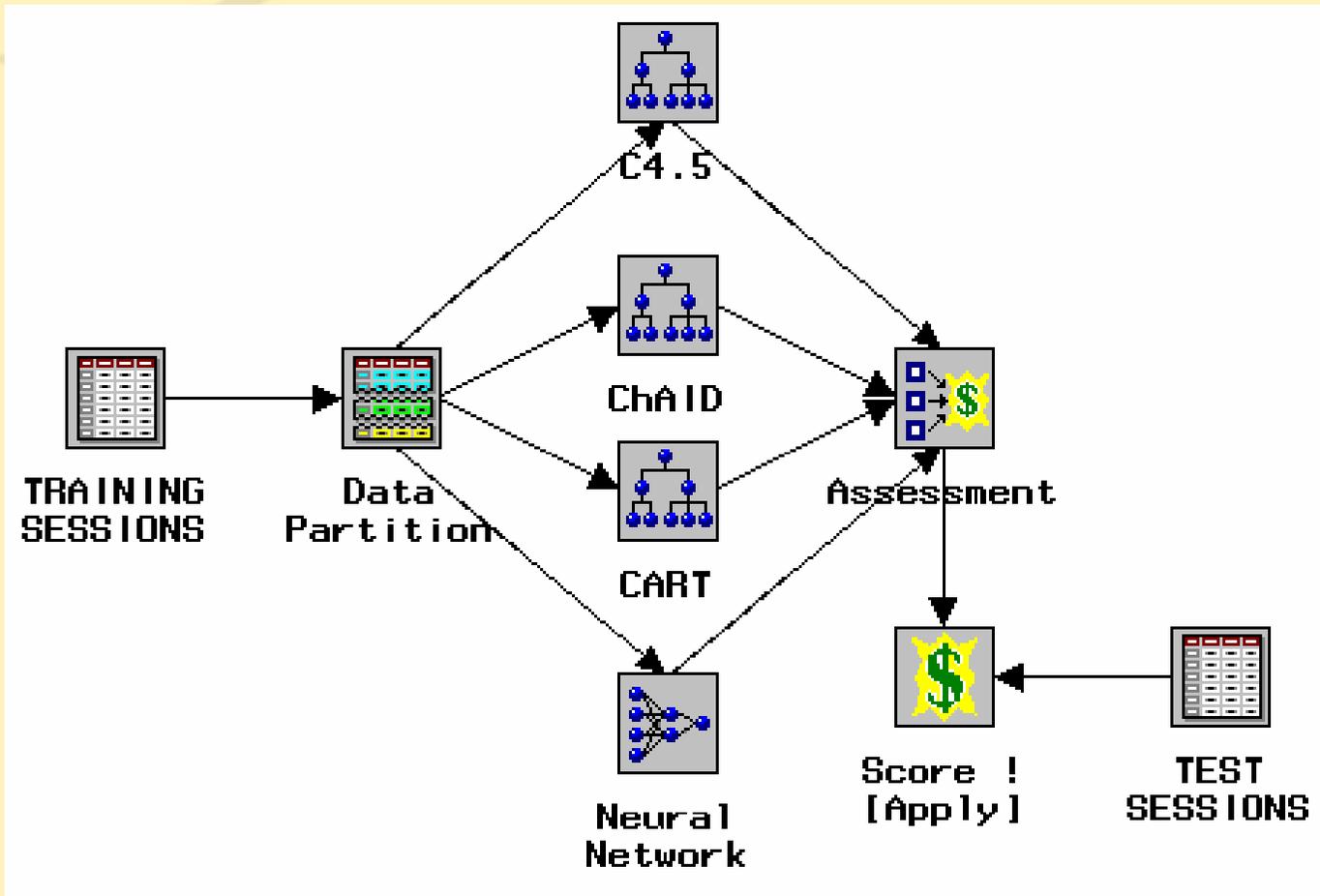
- Vorgehensweise der Analyse:
 - Ausschließliche Verwendung von Sessions mit mehr als 4 Clicks.
 - Trainingsdaten:
 - „Häufige Pfade“ der Länge 3 → binäre Codierung.
 - Generieren zusätzlicher Variablen.
 - Ziel-Variable: Besuch in Ziel-Kategorie (Ja/Nein).
 - Prognosemodell → Score Code (e.g. C-Code).
 - Testdaten:
 - Anwendung des Score Code auf die ersten drei Clicks einer Besuchersession.
 - Ziel-Variable: Besuch in der Ziel-Kategorie in den darauffolgenden Clicks (4,5,...).

Der Praxisdatensatz

- Extended Logfile Format (ELF).
- Session-IDs waren verfügbar.
- Binäre Ziel-Variable: Der Besucher hat Informationen in der Ziel-Kategorie abgerufen (Ja/Nein).



Modellierung im Enterprise Miner™



Ergebnisse auf den Testdaten

Model	Accuracy	Accuracy Target=1	Accuracy Target=0
Neural Network	78,24	72,54	79,31
CART	78,03	53,97	82,56
ChAID	80,08	70,26	81,93
C4.5	80,31	67,70	82,69

Vorhersagegüte auf den Testdaten; die Modelle wurden dabei nur auf die ersten drei Clicks einer jeden Besuchersession angewendet.

→ Unterbreiten von Angeboten/Inhalten anhand der Besucherinteressen!

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 CRM Solutions
SAS Deutschland
Frank.Saeuberlich@ger.sas.com

The Power to Know.