

Web Mining – Analyse von E-Commerce-Kunden mit der Enterprise Miner™ Software

Ulrich Reincke / Stefan Weingärtner

SAS Institute / CSC Ploenzke

4. Konferenz der SAS®-Anwender
in Forschung und Entwicklung

Zitat

In 5 Jahren sind alle Unternehmen
“Internetunternehmen”,

oder sie existieren nicht mehr.

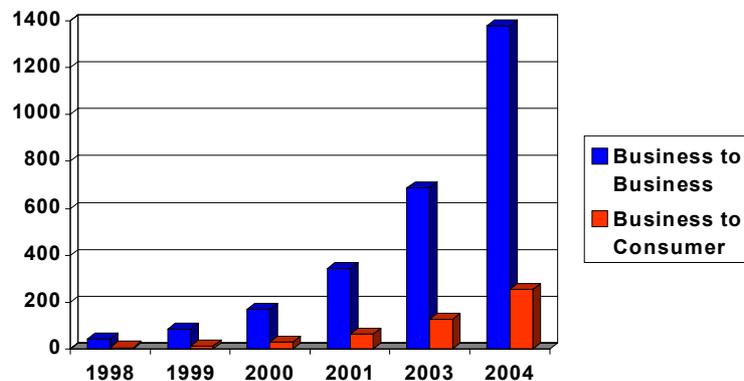
Andy Grove, Chairman of Intel



4. Konferenz der SAS®-Anwender
in Forschung und Entwicklung

US Umsatzzuwächse im E-Commerce

Mrd \$



Quelle: The Economist June 26 1999



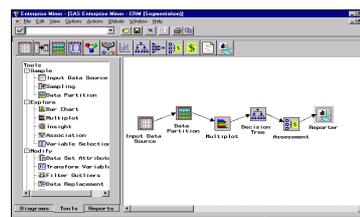
Warum Web Mining?

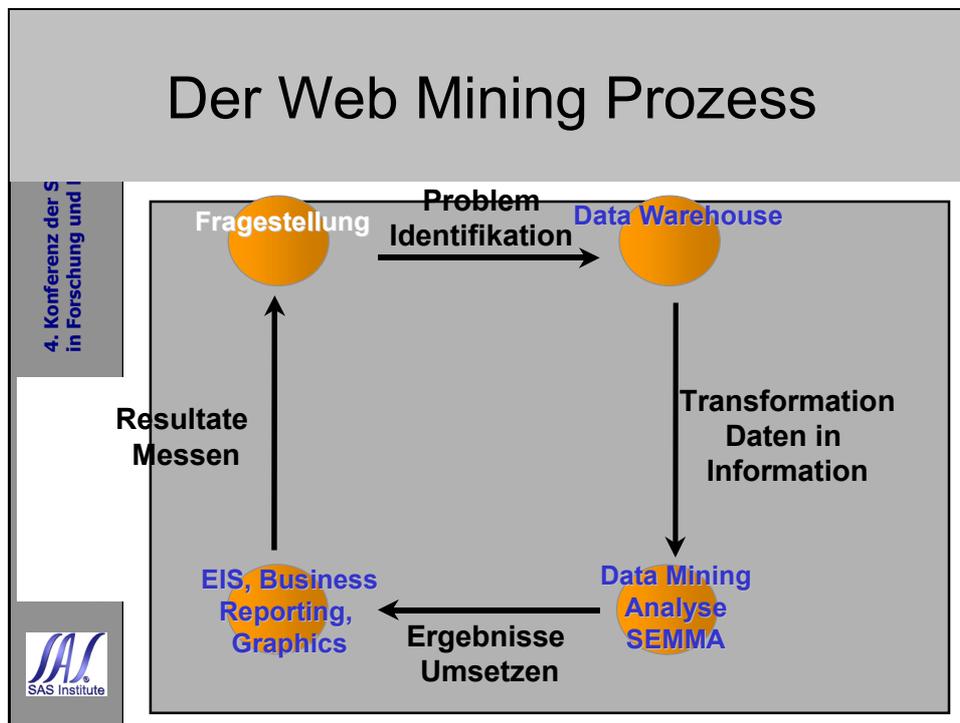
- Weil man im Web jede Transaktion erfassen, aggregieren und analysieren kann.
- Weil sich der Wettbewerb im Web zunehmend verschärft
- Nur eine kundenfokussierte Strategie führt im Web zum Erfolg und zur Marktführerschaft
- Weil es CRM und Data Mining Lösungen gibt, die wertvolle Informationen zum Thema Zielgruppenmanagement, Pricing und Positioning liefern können.



Was ist Web Mining?

Der Prozess von Selektion, Exploration und Modellierung von Web-Daten, um unbekannte Muster und Beziehungen zu entdecken und zum Vorteil des eigenen Unternehmens nutzbar zu machen.





Web-Business Fragen:

- Wie erkenne ich meine profitabelsten Besucher?
- Wer wird meine neuen Online Produkte und Dienstleistungen in Anspruch nehmen?
- Was sind die wichtigsten Trends der Besucher meiner Web Seiten?
- Was sind die Eigenschaften und Charaktereistiken meiner loyalsten Kunden?
- Wie sollte ich meine Online Lagerhaltung optimieren?

4. Konferenz der SAS®-Anwender in Forschung und Entwicklung



Erzeugung von Web Daten

Mit welchen Mechanismen kann man Kundendaten über Internettransaktionen erstellen?

Web Daten, können über

- Log Files,
- Cookies und
- Formulare erzeugt werden

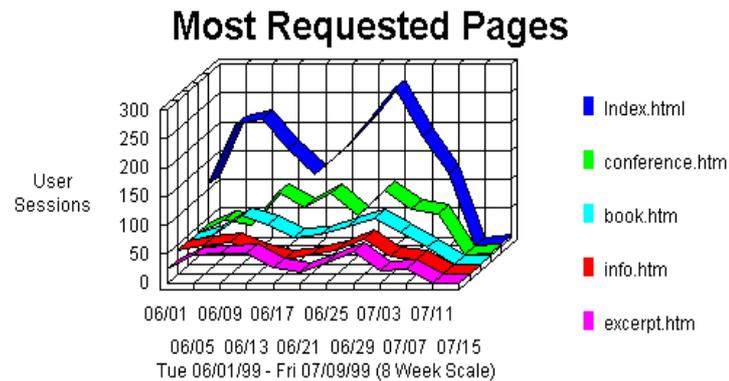


Was sind die Informationsträger?

- **Log Files** - geben an von, wo ein Besucher und wie er zu der Seite kam.
- **Cookies** - Numerieren die Besuche eines Kunden und registrieren welche Seiten besucht wurden.
- **Formulare** - Identifizieren Kunden.



Nutzungsprofile mit Log Files:



Wie arbeiten Cookies?

Cookies werden vom Server an den Browser gegeben, damit jedesmal wenn ein Besucher zum Server zurückkommt, er vom Server erkannt wird um eine komplette Historie abzulegen.



Formulare geben detaillierte Informationen über die Besucher

excite **sign up!**
3 simple steps gets you FREE...
 Excite Mail, Excite Communities, Message Boards, Chat, a Personal Front Page.

1. Choose Your Login Information - 6-20 characters; only letters, numbers, and dashes.

User Name:
 Password: [Outside the U.S.? Click here.](#)
 Re-enter Password:

2. Password Reminder Phrase - In case you forget your password.
 Enter a phrase that will remind you of your password. For example, if your password is the last six digits of your social security number, you might enter "the last six digits of my SSN".

Hint Phrase:

3. Personalization Information - Information to provide customized features.
 Get local weather reports and events, your horoscope, and other cool features.

First Name:
 Last Name:
 ZIP Code:
 Email Address:
 Birthdate: / / 19
 Gender: Female Male



Das Geheimnis zu besseren
Data Mining Modellen sind
bessere Daten.

Das Einbinden von
Microgeographie Daten:

Anwendung eines Formulars, das die PLZ aufnimmt.



Eine Frage nach der PLZ um mit den lokalen Wettervorhersagen, Nachrichten, usw. zu verbinden ermöglicht das Einspielen von Microgeografie Daten (Schober, Bertelsmann, Panadress ...)



The Demographics of ZIP Code 94501

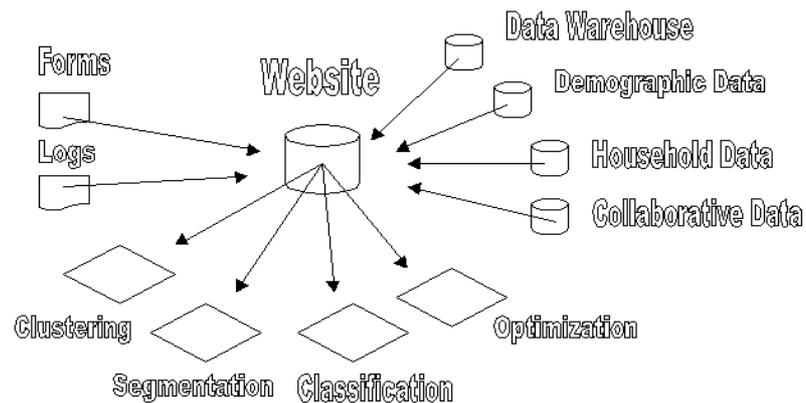
ALAMEDA Dominant ACORN: 3A (High Rise Renters)

This urban mix of single professionals with growing incomes and affluent tastes is well educated and well-employed, with their leisure time being singles and sports oriented. Ranking among the highest for this market is bicycling and fitness. They use credit cards extensively and are top-ranked for ATM use. These urban, mobile apartment dwellers favor investments over savings.

General Statistics for 1998	ZIP 94501	National Total Population
61677	270 Million	
Number of Households:	25968	101 Million
1998 Population by Race		
White:	61.6%	78.0%
Black:	6.8%	12.5%
Asian Pacific Islander: 25.6%	3.8%	
Other:	6.0%	5.7%
1998 Population by Gender		
Male: 48.5%	49.0%	
Female:	51.5%	51.0%
1998 Income Figures		
Median Household Income: \$ 46204		\$ 38135
HH Income Under 50K:	54.9%	64.2%
HH Income \$50K-\$100K:	34.2%	27.3%
HH Income Over \$100K:	10.9%	8.5%
1990 Housing Figures		
Median Home Value: \$ 258,956	\$ 79,098	
Median Rent:	\$ 626	\$ 374



Warehousing ist ein Meilenstein auf dem Weg zum Web Mining



Zusammenfassung

E-commerce Firmen, die ein Web Warehouse pflegen und kontinuierlich weiterentwickeln, können mit Web Mining in die besten Start Positionen gelangen für das erwartete "Hyperwachstum" des Marktplatzes Internet.

Upselling Cross Selling und Kundengewinnung /

-bindung ist das Ein Mal Eins eines Marktes, der sich alle 100 Tage verdoppelt.

Literatur:

Jesus Mena: "Data Mining Your Website" released by Digital Press
July 1999.

info@webminer.com



Data Mining im Internetshopping der Firma ask|net GmbH

Eine Diplomarbeit die im Rahmen des SAS Data Mining Fellowship gefördert wurde



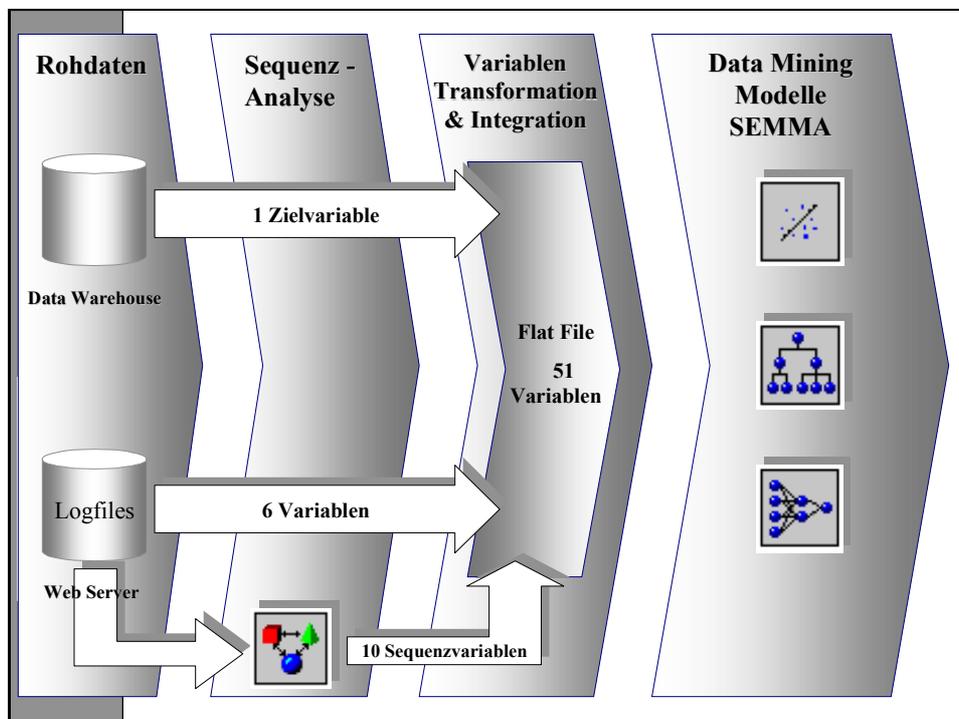
Von Stefan Weingärtner, CSC Plönzke

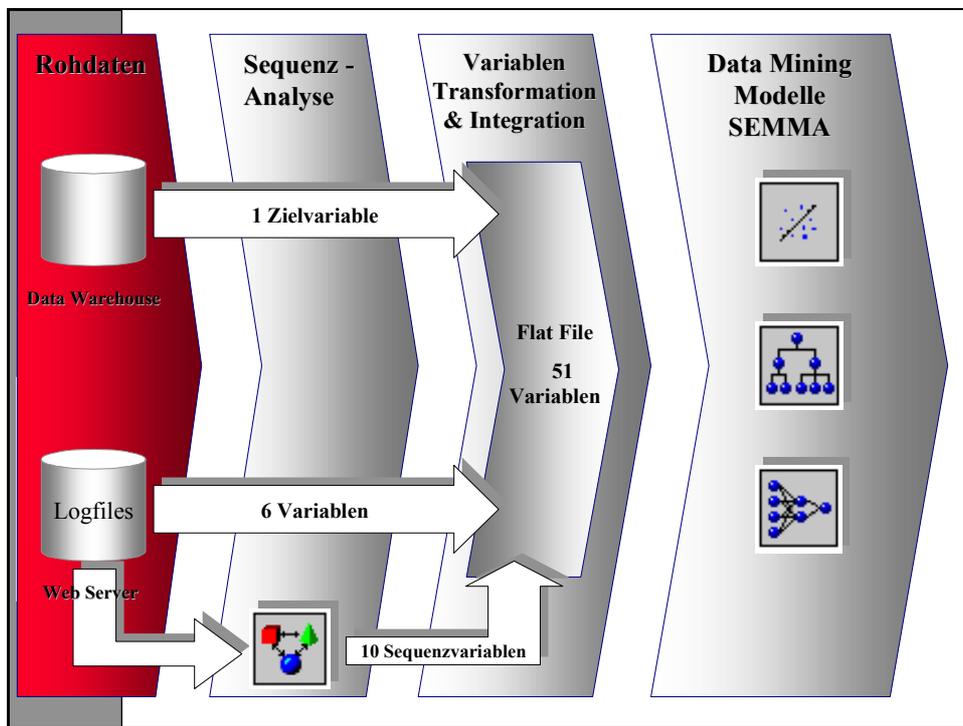
Data Mining Fragestellungen:

- Lassen sich aus den Logfiles konkrete Benutzerprofile ableiten ?
- Was sind die Einflußfaktoren auf einen Kauf im Online-Shop *SoftWarehouse* ?
- Kann Data Mining konkrete Hinweise für die Shop-Optimierung bieten ?

zudem...

Wie entwickelt man überhaupt ein Data Mining-Modell ?





4. Konferenz der SAS®-Anwender
in Forschung und Entwicklung

Wie werden die Rohdaten generiert ?



Beispielhafte "Click - Session" im Shop
www.SoftWarehouse.de ...







Produktübersicht - Netscape

File Edit View Go Communicator Help

Location: <https://www.softwarehouse.de/cgi-bin/product/P10375/ml=DE>

Instant Message WebMail Contact People Yellow Pages Download Channels

Back Forward Reload Home Search Netscape Print Security Stop

SoftWarehouse.de Ein Service der ask|net
The World of Downloads

Home Neues Feedback Hilfe & Info AGB
Registrierung Anmeldung Katalog Warenkorb Abholregal

Sie sind nicht angemeldet.

Duden Die deutsche Rechtschreibung

2 komplette Dudenbände auf einer CD-ROM

Jeweils rund 115 000 Stichwörter und über 500 000 Bedeutungserklärungen, Beispiele und Angaben zur Silbentrennung, Aussprache, Grammatik und Etymologie sowie Richtlinien für die Textverarbeitung.



- [Beschreibung](#)
- [Systemvoraussetzungen](#)
- [Duden Online](#)

zurück zu: [Duden Online](#)

Bitte wählen Sie die gewünschte Anzahl aus:

Anzahl	Name	Jahr 2000	Art	System	Nutzer	Preis
<input type="text" value="0"/>	Duden Die deutsche Rechtschreibung	k.A.	Paket Version Information	Mac OS Windows 3.1 Windows 95 Windows NT	Deutschland, Österreich, Schweiz	78.00 DM

Die angegebenen Preise verstehen sich inklusive der gesetzlichen Mehrwertsteuer!

Benutzeranmeldung - Netscape

File Edit View Go Communicator Help

Location: <https://www.softwarehouse.de/cgi-bin/login/ml=DE>

Instant Message WebMail Contact People Yellow Pages Download Channels

Back Forward Reload Home Search Netscape Print Security Stop

SoftWarehouse.de Ein Service der ask|net
The World of Downloads

Home Neues Feedback Hilfe & Info AGB
Registrierung Anmeldung Katalog Warenkorb Abholregal

Sie sind nicht angemeldet.

Melden Sie sich bitte mit Login-Namen und Paßwort an. Dann können Sie im Warenkorb Software bestellen, bezahlen und aus dem Abholregal direkt auf Ihren Rechner laden.

Haben Sie noch keinen Login-Namen und kein Paßwort? Dann [registrieren](#) Sie sich online. Nach Ihrer Registrierung erhalten Sie per Email sofort den Kundenschlüssel.

Login: Paßwort:

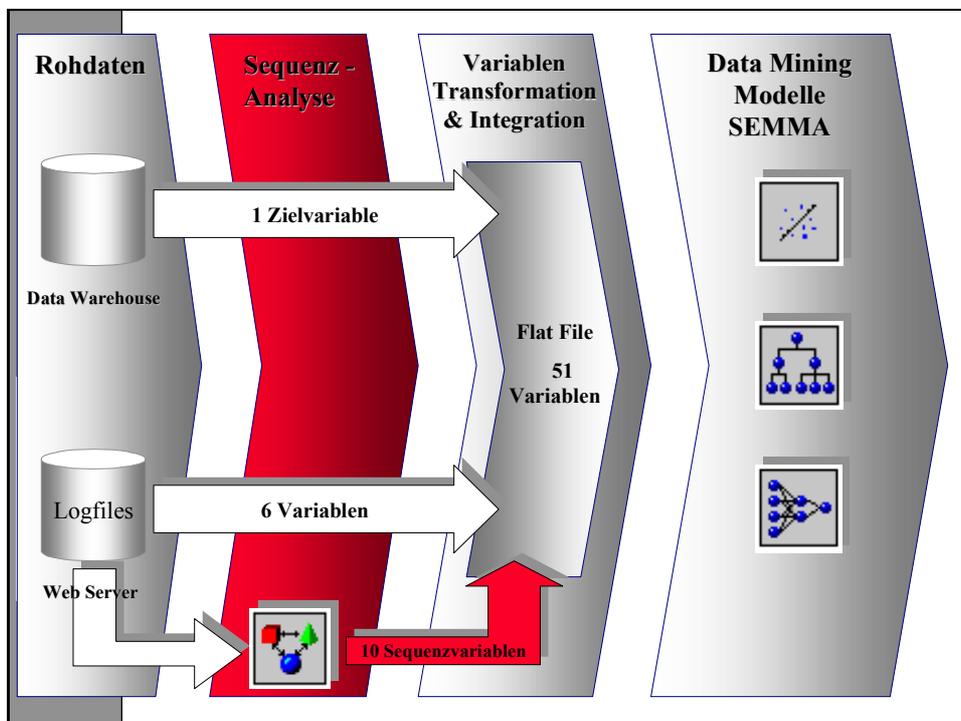
Hier können Sie sich vom System abmelden.

19/Aug/1999

4. Konferenz der SAS®-Anwender in Forschung und Entwicklung

Tabelleneintrag Oracle-Datenbank

C_CALLER	C_CALLER	C_FQDN	C_REFERERER	C_TIME	L_ID
...
a68adce1ab...	home	xyz.sas.com	www.softwarehouse.de	20-8-99:14:01	DE
a68adce1ab...	catalog	xyz.sas.com	home	20-8-99:14:02	DE
a68adce1ab...	program	xyz.sas.com	catalog	20-8-99:14:03	DE
a68adce1ab...	product	xyz.sas.com	program	20-8-99:14:05	DE
a68adce1ab...	login	xyz.sas.com	product	20-8-99:14:07	DE
...

Sequenzanalyse der Click-Streams

- ① Beantwortung konkreter Fragestellungen
- ② Generierung weiterer Variablen für spätere Data Mining - Modelle



Ergebnisse Sequenzanalyse

①

	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>

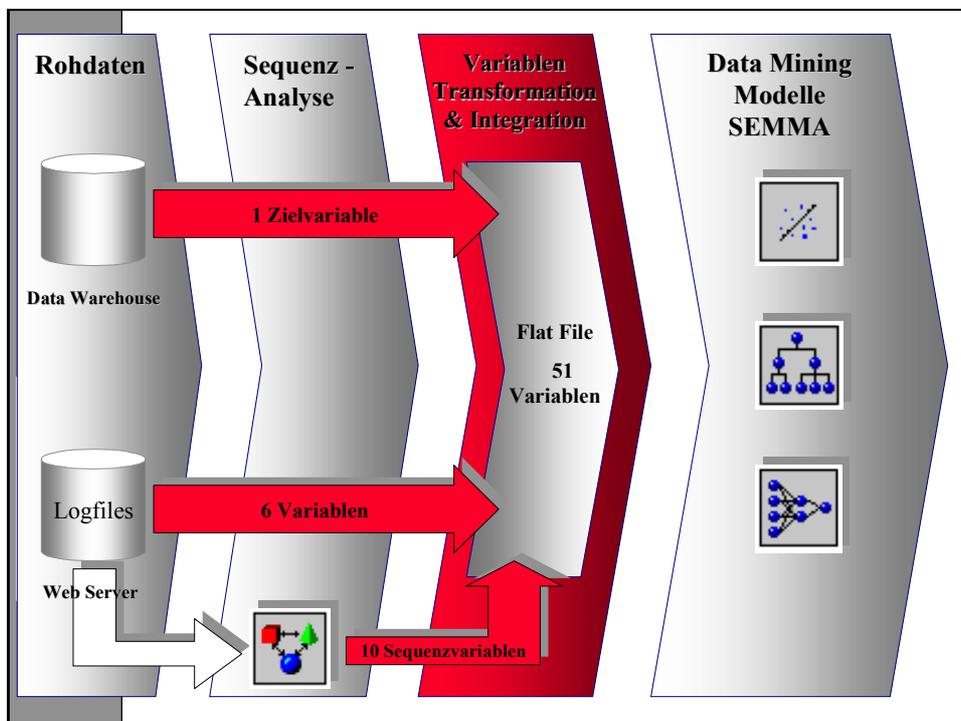


Konferenz der SAS®-Anwender in Forschung und Entwicklung

Ergebnisse Sequenzanalyse

②

#	Support (%)	Confidence (%)	Rule
seq_0	28.7	74.6	<i>program⇒product⇒p_info⇒product</i>
seq_1	26.4	60.2	<i>program⇒product⇒product⇒product</i>
seq_2	17.5	80.3	<i>program⇒product⇒addcart⇒freeze</i>
seq_3	17.1	82.6	<i>home⇒catalog⇒program⇒product</i>
seq_4	13.5	95.6	<i>logpost⇒catalog⇒program⇒product</i>
seq_5	12.7	92.2	<i>login⇒catalog⇒program⇒product</i>
seq_6	12.6	93.0	<i>logpost⇒product⇒addcart⇒freeze</i>
...
seq_9	6.7	67.3	<i>product⇒login⇒register⇒regpost</i>



Integration von 'Sequenzvariablen'

4. Konferenz der SAS®-Anwender
in Forschung und Entwicklung

C_VALUE	C_CALLER
...	...
a68adce1ab...	home
a68adce1ab...	catalog
a68adce1ab...	program
a68adce1ab...	product
a68adce1ab...	login
...	...

➔

c_value	col1	col2	col3	col4	col5
...
a68adcelab...	home	catalog	program	product	login
...

↓

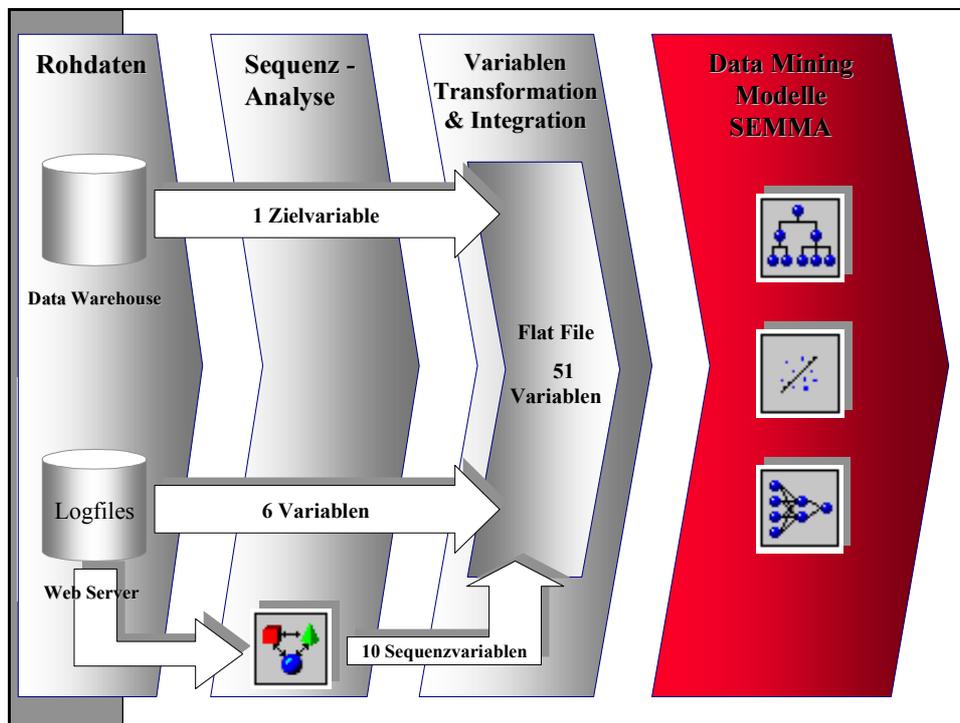
c_value	concat_column
...	...
a68adcelab...	homecatalogprogramproductlogin
...	...

↓

program⇒product⇒p_info⇒product
transformiert zu
'%program%product%p_info%product%'

Integration der Transaktionstabelle

C_VALUE	C_CALLER	C_FQDN	C_REFERERER	C_TIME	L_ID
...
a68adce1ab...	home	xyz.sas.com	www.softwarehouse.de	20-8-99:14:01	DE
a68adce1ab...	catalog	xyz.sas.com	home	20-8-99:14:02	DE
a68adce1ab...	program	xyz.sas.com	catalog	20-8-99:14:03	DE
a68adce1ab...	product	xyz.sas.com	program	20-8-99:14:05	DE
a68adce1ab...	login	xyz.sas.com	product	20-8-99:14:07	DE
...



Fragestellung

Mit Hilfe verschiedener Data Mining-Modelle sollen die Auswirkungen und Interaktionen von verschiedenen Inputvariablen, wie Klicksequenzen, Referern, Anzahl und Art der besuchten Seiten usw. auf einen Kaufvorgang im Shop (Zielvariable) untersucht werden.

Entwicklung verschiedener Prognosemodelle

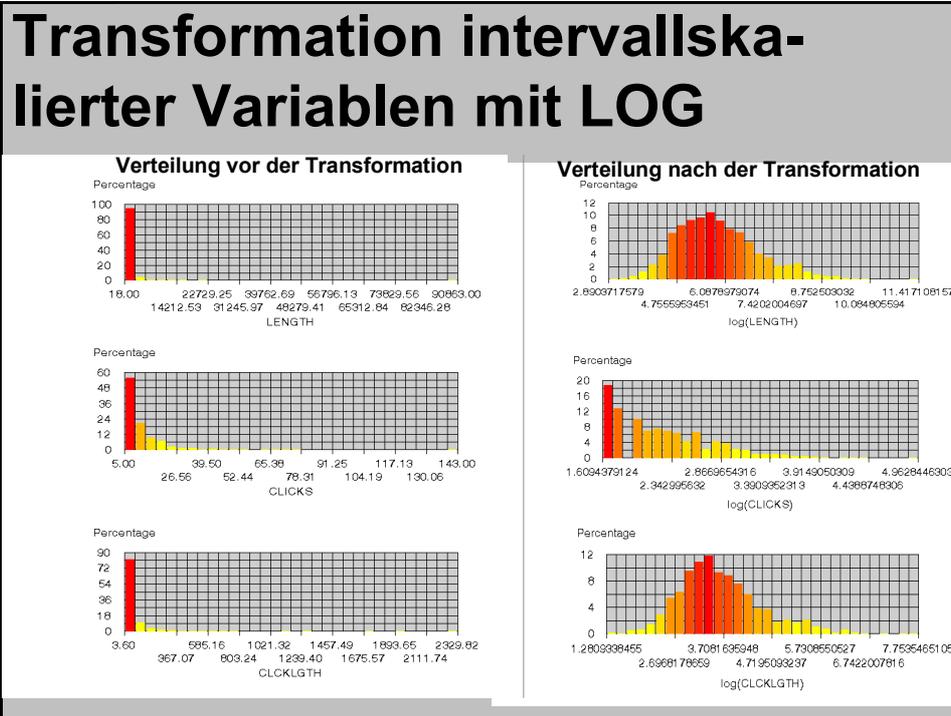
- Entscheidungsbaum
- Regressionsanalyse
- Neuronales Netz



Voraussetzungen aller Datenmodelle

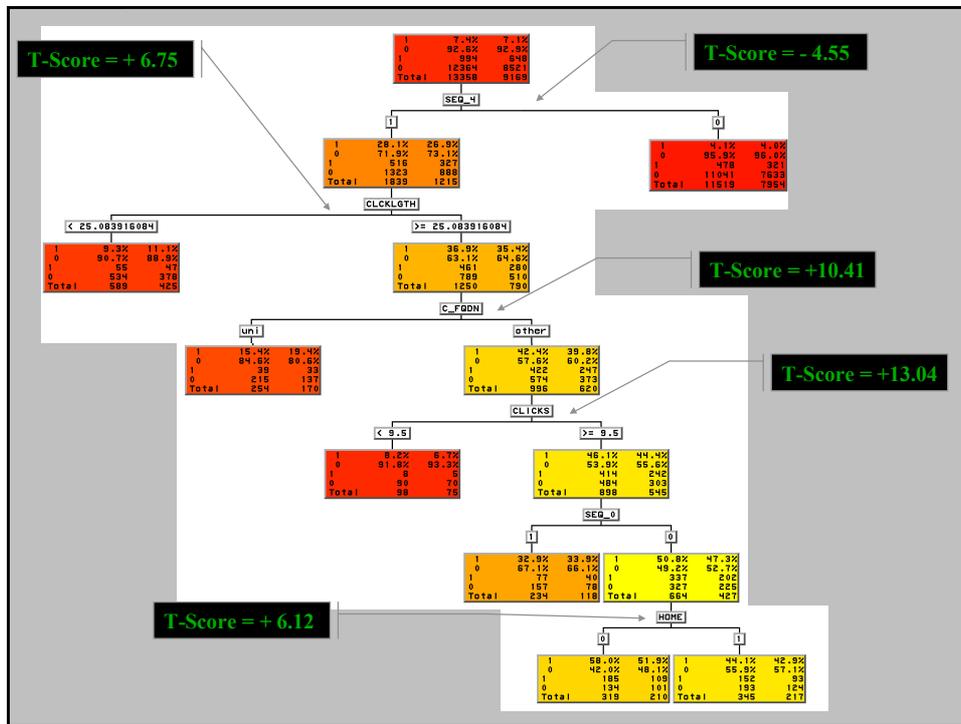
- Eingang fanden alle Sessions mit Anzahl
besuchter Seiten ≥ 5
- Anzahl der Sessions: 22527
- Anzahl an Käufen: 1642
- Relativer Kaufanteil: 7.29%
- Anteil der Trainingsdaten: 60%
- Anteil der Validationsdaten: 40%





Zusammenfassende Ergebnisse der Analysen

- Neuronales Netz ist bestes Prognosemodell
- Navigationsprofile aus dem SoftWarehouse lassen sich am besten aus dem Entscheidungsbaum ablesen
- Mit Hilfe der Regressionsanalyse lassen sich wichtige Abhängigkeiten erkennen und die Ergebnisse des Entscheidungsbaumes verifizieren



Erkenntnisse aus Data Mining

- Bestellwege zu lang und umständlich
 - ⇒ Sequenzanalyse
 - z.B. 'addcart⇒login' oder 'addcart⇒register'
 - ⇒ Entscheidungsbaum
 - z.B. Splits bei Variablen 'clcklgth' und 'clicks'
 - ⇒ Regressionsanalyse
 - z.B. T-Score von 'clicks' und 'clcklgth'



Erkenntnisse aus Data Mining

- Verhalten bei Registrierung/Login weiter beobachten...
 - ⇒ Sequenzanalyse
 - z.B. 'login⇒register' oder 'login⇒login'
 - ... denn Mehrfachregistrierungen erschweren eine Warenkorbanalyse



Fazit

- Data Mining mit dem SAS Enterprise Miner bereitete eine wertvolle Basis für Kundenanalysen im e-Commerce Bereich
- Verdeckte Unzulänglichkeiten der Internetseiten kommen ans Tageslicht
- Im Licht der gewonnenen DM Erkenntnisse überarbeitete die Firma ask|net ihre Internetseiten

