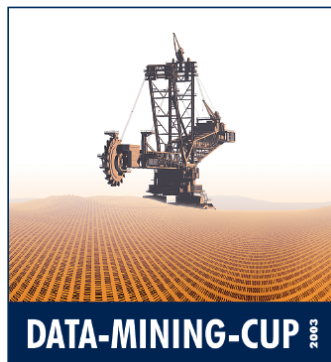

**Vortrag auf den 4. DATA-MINING-CUP Anwendertagen
(Chemnitz, 18.-20. Juni 2003)**



<http://www.data-mining-cup.de>

Copyright-Hinweis:

Das Urheberrecht des folgenden Vortrags liegt beim Author. Verbreitung, Vervielfältigung und Kopie, auch auszugsweise, ist nur mit schriftlicher Genehmigung des Authors erlaubt.

Modularer Aufbau von Kundenbiographien – Mustererkennung in Daten am Beispiel eines TK – Unternehmens

Andreas Gyenis, Walter Warmuth

DATA-MINING-CUP Anwendertage 2003 19.06.03 Chemnitz

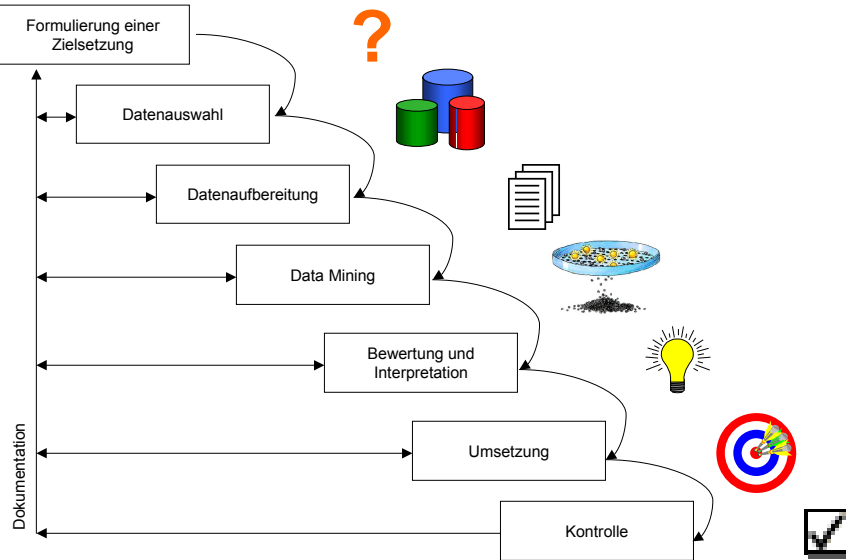
Ausgangspunkt

In Branchen mit einer langen Historie haben die Unternehmen über einen gewissen Zeitraum, mit einer Verkaufskultur, mit Vertriebswegen, mit Produkten, mit Dienstleistungen und mit Steuerungsmechanismen, mit Anreizsystemen einen spezifischen Kunden generiert.

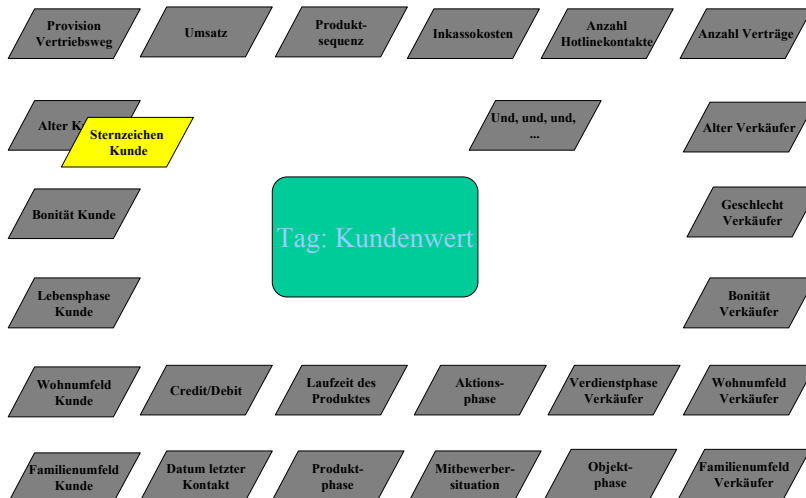
Dieses gilt für Versicherungen, für Banken, für Versandhäuser, für den Einzelhandel, ... unabhängig davon, ob der Vertrieb über Vertreter, den Schalter, über Kataloge, über den Ladentisch, das Call-Center oder via Internet erfolgt.

In der TK-Branche, wo die Dynamik dominiert, muß diese modellhaft mit erkundet werden, um die Baussteine des Wissens zu separieren.

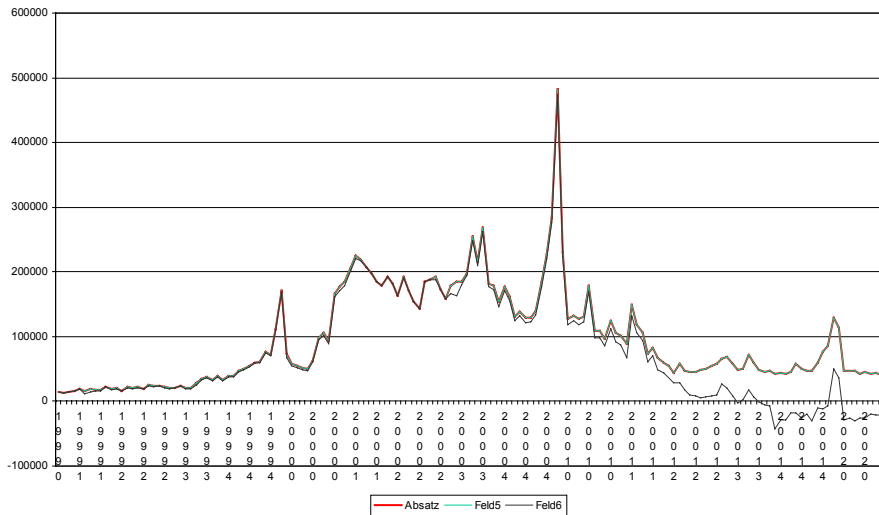
Der Prozeß der Wissensentdeckung



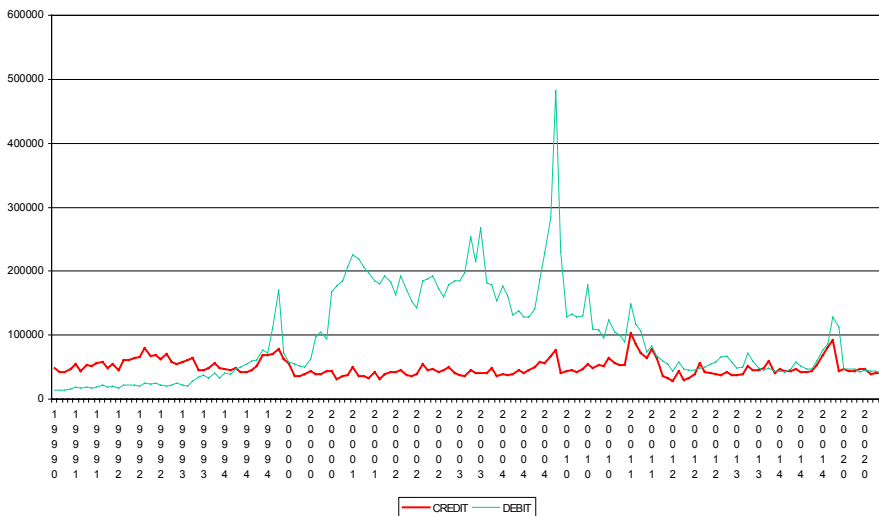
TK-Kunde



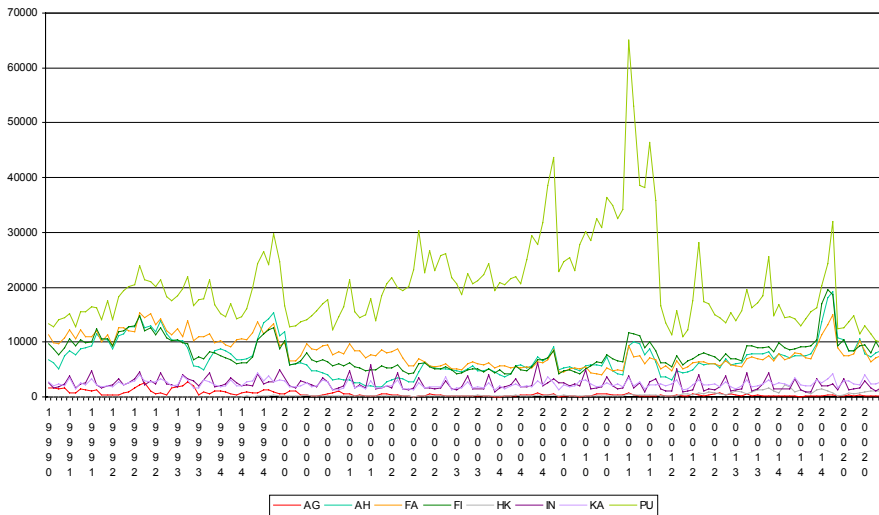
Debit



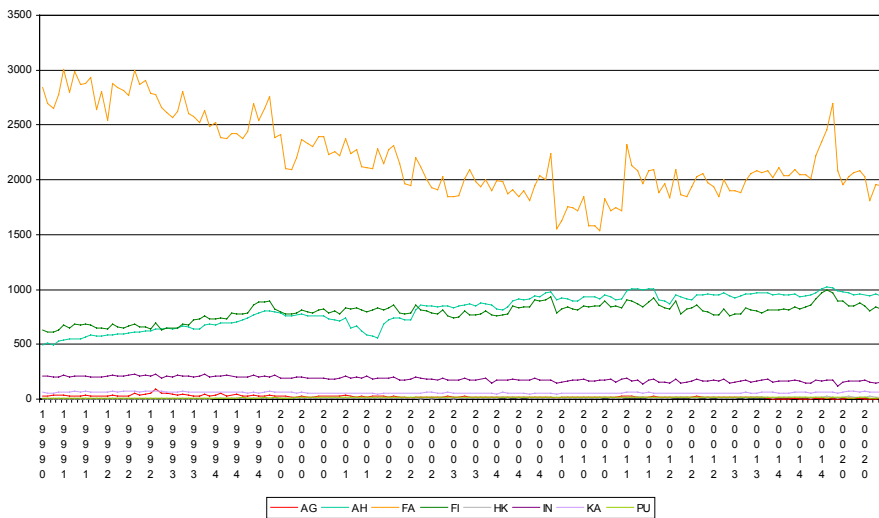
Vergleich Credit/Debit



Credit

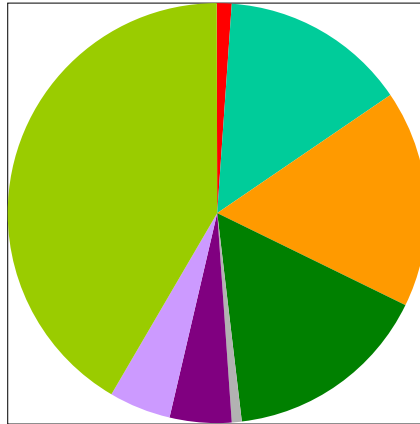


Debit



Einfluß der Vertriebswege (Umsatz)

Credit

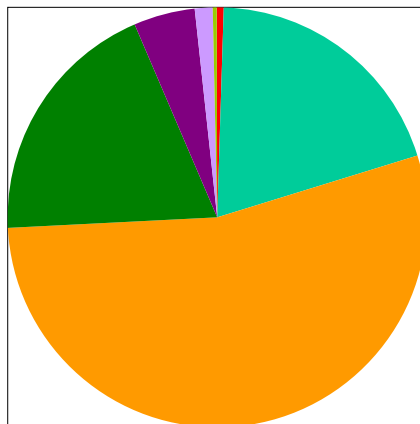


AG AH FA FI HK N KA PU

Modularer Aufbau von Kundenbiographien ...

Einfluß der Vertriebswege (Verkäufer)

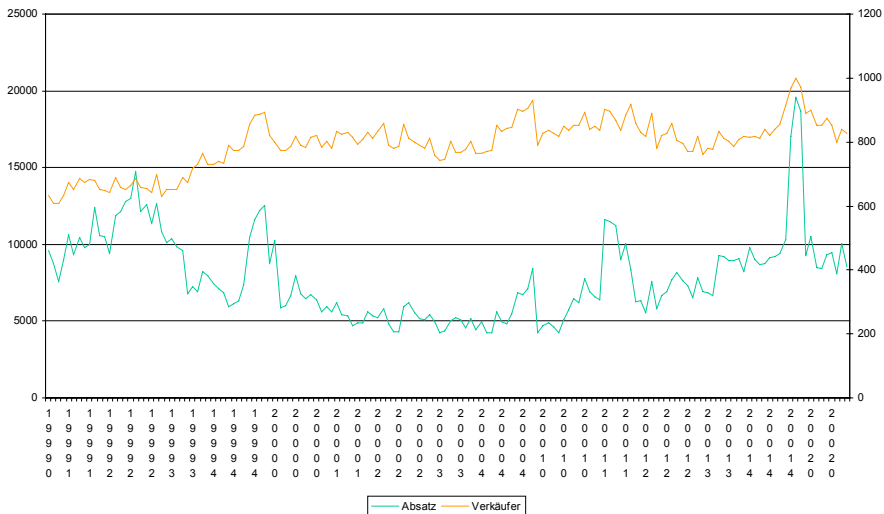
Credit



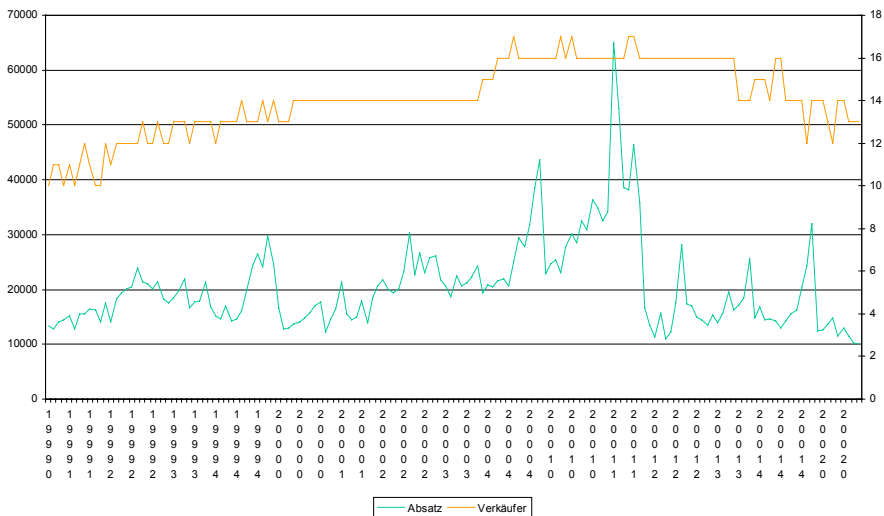
AG AH FA FI HK N KA PU

Modularer Aufbau von Kundenbiographien ...

Credit, Vertriebsweg FI (Filiale)

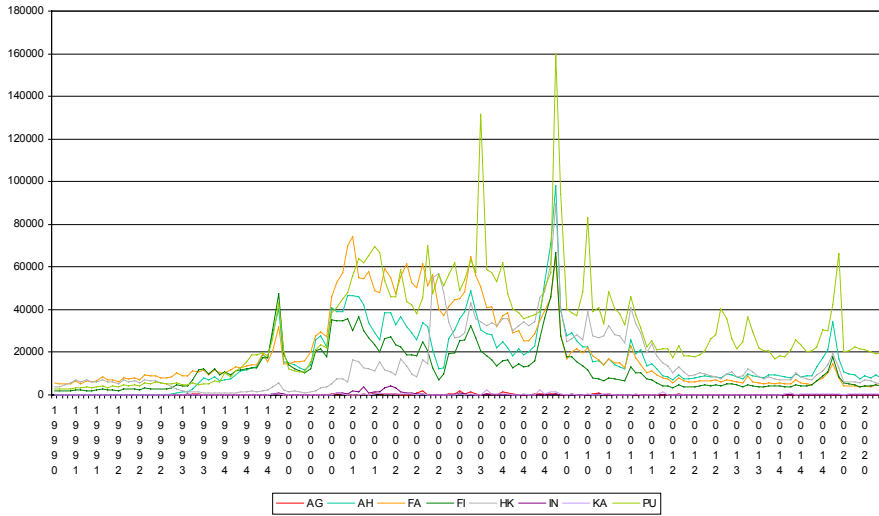


Credit, Vertriebsweg PU (Partnerunternehmen)



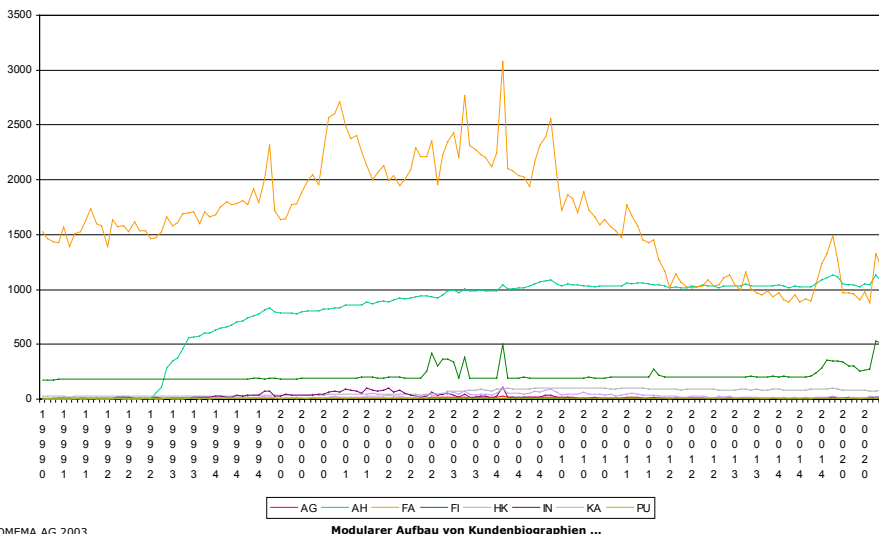
Absatz nach Vertriebsweg

Debit



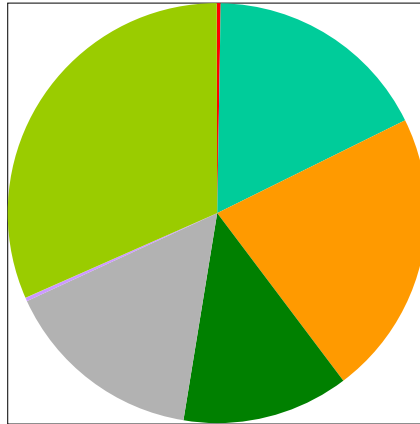
Verkäufer nach Vertriebsweg

Debit



Absatz nach Vertriebsweg

Debit

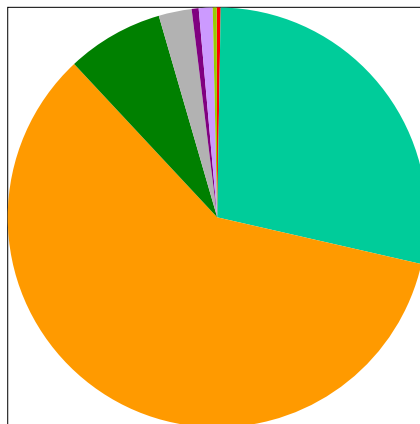


AG AH FA FI HK N KA PU

Modularer Aufbau von Kundenbiographien ...

Verkäufer nach Vertriebsweg

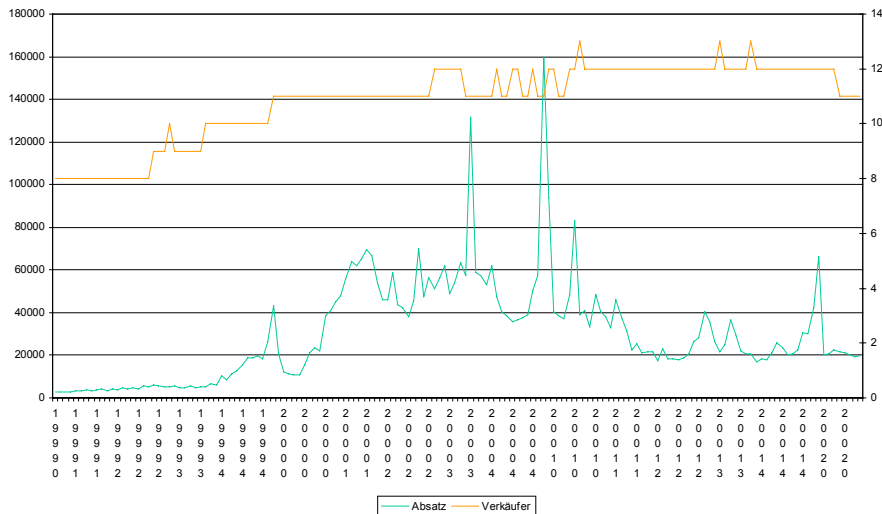
Debit



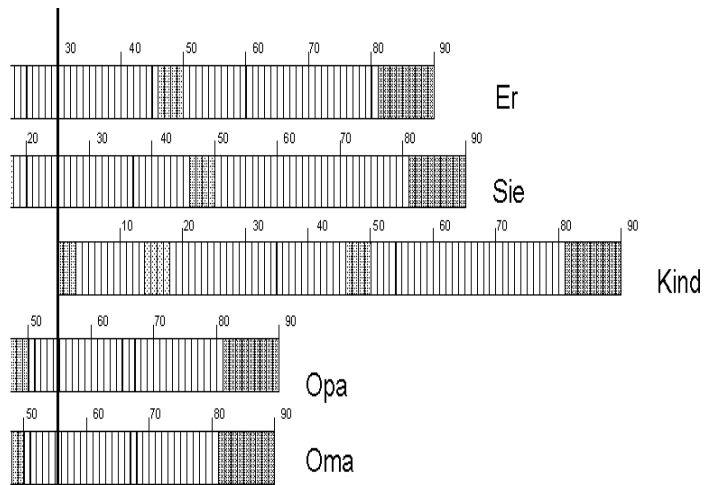
AG AH FA FI HK N KA PU

Modularer Aufbau von Kundenbiographien ...

Debit, Vertriebsweg PU (Partnerunternehmen)



- Der Absatz von Credit und Debit unterscheidet sich deutlich.
- Für Credit wird der Absatz von den Vertriebswegen AH, FA, FI und PU bestimmt, während die Verkäufer (Typologie) hauptsächlich in AH, FA und FI zu finden sind.
- Für Debit wird der Absatz von den Vertriebswegen AH, FA, FI, HK und PU bestimmt, während die Verkäufer hauptsächlich in AH und FI zu finden sind.
- „Karte“ der Botschaften der Kunden
- Soziodemographische Merkmale aus Kundenkarten, Stars-Programm, Payback, Credit Cards, Shop, ...: Familienstruktur, Ausländeranteil, Altersstruktur, Einkommensstruktur, Bebauung, ..., u. U. Indexwerte (z. B. Abweichung vom Durchschnitt) umrechnen, ...



- Wie wird ein Kunde ein Kunde?
- Symbolische (kategoriale) Daten; Maschinelles Lernen (Neuronale Netze, Entscheidungsbaum-Algorithmen, Assoziations-Algorithmen) und Modellierung stochastischer Modelle (Bedienungsmodelle, Zuverlässigkeitsmodelle, $X > Y$ -Modelle, Finanzmathematische Modelle)
- **Neuronale Netze** (allein Black Box), **Entscheidungsbaum-Algorithmen** bilden Regeln mit genau einem Schluß, **Assoziations-Algorithmen** suchen nach mehreren Regeln, jede von ihnen kann einen unterschiedlichen Schluß zulassen. Im deutschsprachigen Raum wird bei dieser Form der „graphischen Methoden“, Zusammenhänge oder Assoziationen zu ergründen, von der „**Warenkorbanalyse**“ gesprochen.

- Bei der Warenkorbanalyse geht es nicht nur um die konstatierende Beschreibung aller Warenkörbe, sondern um das Aufdecken und Identifizieren von gültigen, nützlichen, und wenn möglich neuen **Regeln** aus den gesammelten Daten. Unter allen möglichen Kombinationen werden Regeln für jene Produkte, Zeitpunkte, Vertriebswege etc. erstellt, die eine Abhängigkeit aufweisen, d. h. häufig gemeinsam vorkommen, sich gegenseitig bedingen oder hemmen.
- Verknüpft mit der zeitlichen Reihenfolge, geeignet gruppiert und fachlich bereinigt (einer Vertragskündigung geht immer der Vertragsabschluß voraus; die Kausal-Kette, die mit der Scheidung endet, beginnt immer mit der Hochzeit) setzen sich diese Regeln aus „**Modulen**“ zusammen.
- Module: **Portale, Drehtüren, finale Klassen**

- Die **Gültigkeit** drückt aus, wie sicher eine Aussage ist.
Beispiel: Die Regel „Wenn X eintritt, dann kaufen 45 von 50 Kunden“ ist zu 90 % sicher.
- Die **Nützlichkeit (Relevanz)** misst den Bereich, den die Aussage abdeckt.
Beispiel: Die oben aufgeführte Regel ist zwar gültig, wäre bei einer Gesamtanzahl von 50.000 Käufern aber nicht sehr nützlich, da sie lediglich 50 von 50.000 Käufern (= 0,1 %) betrifft.
- **Neuheit** drückt aus, inwieweit die Aussage neu oder unerwartet ist.
Beispiel: Die Regel „Alle Schwangeren sind Frauen“ ist gültig und nützlich aber nicht ganz unerwartet
- **Verständlichkeit** misst, inwieweit eine Aussage von einem Benutzer verstanden werden kann.
Beispiel: Die Aussage „Der Kunde Y hat einen Scorewert von 97 %“ ist weniger verständlich als eine Regel der Art „Wenn X eintritt, dann ist der Scorewert für Kunde Y hoch“

1 PALMOLIVE_ULTRA ___ 500_ML <- 10 DENTAGARD_ZAHNCREME_75ML
(0.4%/269, 3.0%/1974, 13.6%, 454.0%)

Palmolive: 1.633 oder 2,5 %

Dentagard: 1.974 oder 3,0 %

Nutzen(Palmolive und Dentagard): 269 oder 0,4 %

Gültigkeit(Palmolive wenn Dentagard): $\frac{269}{1.974} = 13,6 \%$

Steigerung(Palmolive wenn Dentagard): $\frac{13,6 \%}{2,5 \%} = 5,44$ oder 454 %

Clementine von SPSS hat in seiner Version 7.0 drei Assoziations-Algorithmen zur Auswahl:

GRI (Generalized Rule Induction)

Inputs: numeric, symbolic

Outputs: symbolic

Apriori

Inputs: symbolic

Outputs: symbolic

Sequence

Erkundet sequentielle Muster in zeitorientierten Daten.

Dr. Walter Warmuth:

Eine „Sequence“ ist eine Liste von Ereignis-Mengen, die dazu tendieren in vorhersagbarer Weise einzutreten.

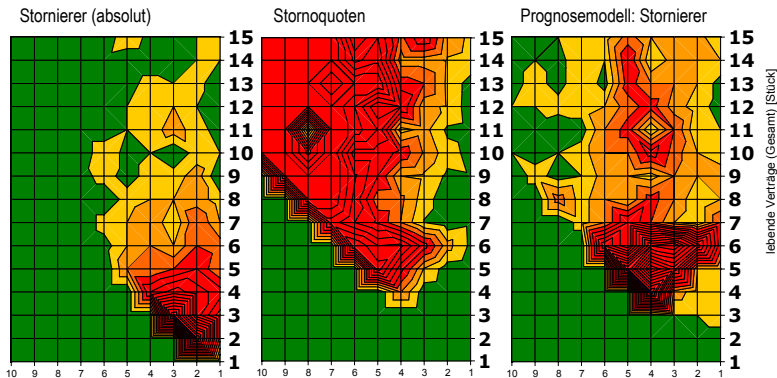
apriori [options] infile outfile [appfile]
version 4.0 (12.02.2002) (c) 1996-2002 Christian Borgelt

-t# target type (s: item sets, r: rules (default), h: hyperedges)	-l print a list of additional rule evaluation measures
-m# minimal number of items per set/rule/hyperedge (default: 1)	-d# minimal value of additional evaluation measure (default: 10%)
-n# maximal number of items per set/rule/hyperedge (default: 5)	-v print value of additional rule evaluation measure
-s# minimal support of a set/rule/hyperedge (default: 10%)	-g write output in scanable form (quote certain characters)
-c# minimal confidence of a rule/hyperedge (default: 80%)	-l do not load transactions into memory (work on input file)
-o use original definition of the support of a rule (body & head)	-q# sort items w.r.t. their frequency (default: 1) (1: ascending, -1: descending, 0: do not sort)
-x extended support output (print both rule support types)	-z minimize memory usage (default: maximize speed)
-a print absolute support (number of transactions)	-i# ignore records starting with characters in the given string
-p print support/confidence with high precision	-b/f/r# blank characters, field and record separators (default: " \t\r", " \t", "\n")
-e# additional rule evaluation measure (default: none)	

Infile-file to read transactions from, **outfile**-file to write item sets/association rules/hyperedges to, **appfile**-file stating item appearances (optional)

0384	AH	N_900	PLZ_	H_04	Tag_96	3
0385	AH	N_900	PLZ_	H_04	Tag_96	1
0388	AH	N_900	PLZ_	H_04	Tag_80	1
0392	AH	N_900	PLZ_	H_08	Tag_96	8
0393	AH	N_900	PLZ_	H_08	Tag_96	3
0395	AH	N_900	PLZ_	H_07	Tag_96	5
0399	AH	N_900	PLZ_	H_07	Tag_96	6
0401	AH	N_900	PLZ_	H_07	Tag_89	1
0402	AH	N_900	PLZ_	H_06	Tag_80	
0404	AH	N_900	PLZ_	H_06	Tag_96	7
0406	AH	N_900	PLZ_	H_06	Tag_80	
0409	AH	N_900	PLZ_	H_13	Tag_80	
0						
0413	AH	N_900	PLZ_	H_12	Tag_96	9
0415	AH	N_900	PLZ_	H_13	Tag_96	15
0420	AH	N_900	PLZ_	H_13	Tag_80	15
0421	AH	N_900	PLZ_	H_13	Tag_96	25
0427	AH	N_900	PLZ_	H_02	Tag_96	2
0429	AH	N_900	PLZ_	H_03	Tag_96	Fasching
0432	AH	N_900	PLZ_	H_05	Tag_96	1
0434	AH	N_900	PLZ_	H_01	Tag_96	4
0438	AH	N_900	PLZ_	H_02	Tag_96	2
0439	AH	N_900	PLZ_	H_01	Tag_96	8
0440	AH	N_900	PLZ_	H_01	Tag_96	3
0441	AH	N_900	PLZ_	H_03	Tag_96	1
0442	AH	N_900	PLZ_	H_01	Tag_93	9
0444	AH	N_900	PLZ_	H_02	Tag_96	3
0447	AH	N_900	PLZ_	H_11	Tag_96	8
0451	AH	N_900	PLZ_	H_12	Tag_96	6
0453	AH	N_900	PLZ_	H_09	Tag_96	4
0454	AH	N_900	PLZ_	H_11	Tag_96	10
0455	AH	N_900	PLZ_	H_09	Tag_96	15
0458	AH	N_900	PLZ_	H_08	Tag_96	7
0459	AH	N_900	PLZ_	H_12	Tag_95	3
0462	AH	N_900	PLZ_	H_12	Tag_95	15

Modellierungsproblem: Wo finden sich die Signalereignisse, die den Eintritt in ein Modul ankündigen?



Portale: Eintrittsprodukte (Credit, Debit)

Drehtüren: Cross- und Up-Selling in einer Produktgruppe („PartnerKarte“), Familienpackung

Finale Zustände: Folgestorni (Betreuerwechsel, Sperre), Datentarif („DataConnect XXL“), Credit nach „PartnerKarte“

Kundenbiographien sind Assoziationen – keine Kausalketten:
„Was war zuerst da? Die Glotze, die allen medieneilen Selbstdarstellern der Republik einen Auftritt ermöglicht? Oder die Medieneilen, die sich die Glotze als Ventil für ihre Profilneurosen schufen?“ (Oliver Kalkofe)

Bei Fragen wenden Sie sich bitte an:



COMEMA AG
Institut für Markt- und Unternehmensanalysen

Adresse	Hauptsitz Leipzig: Gottschedstr. 12 D- 04109 Leipzig	Niederlassung Köln: Berrenrather Str. 154-156 D- 50937 Köln
Tel.:	+49 (341) 124 55-80	+49 (221) 420 610
Fax:	+49 (341) 124 55-99	+49 (221) 420 61-100
eMail:	warmuth@COMEMA.de	
Internet:	www.COMEMA.de	