



Data-Mining-Lösungen von Prudential Systems

Holm Sieber <sieber@prudsys.de>

- Vorstellung der Prudential Systems Software GmbH (demnächst prudsys AG)
- Überblick über unsere Produkte
- Beispiele für unsere Projekte
- kurze Demonstration des prudsys[®] DISCOVERER

- Vorstellung der Prudential Systems Software GmbH (demnächst prudsys AG)
- Überblick über unsere Produkte
- Beispiele für unsere Projekte
- kurze Demonstration des prudsys[®] DISCOVERER

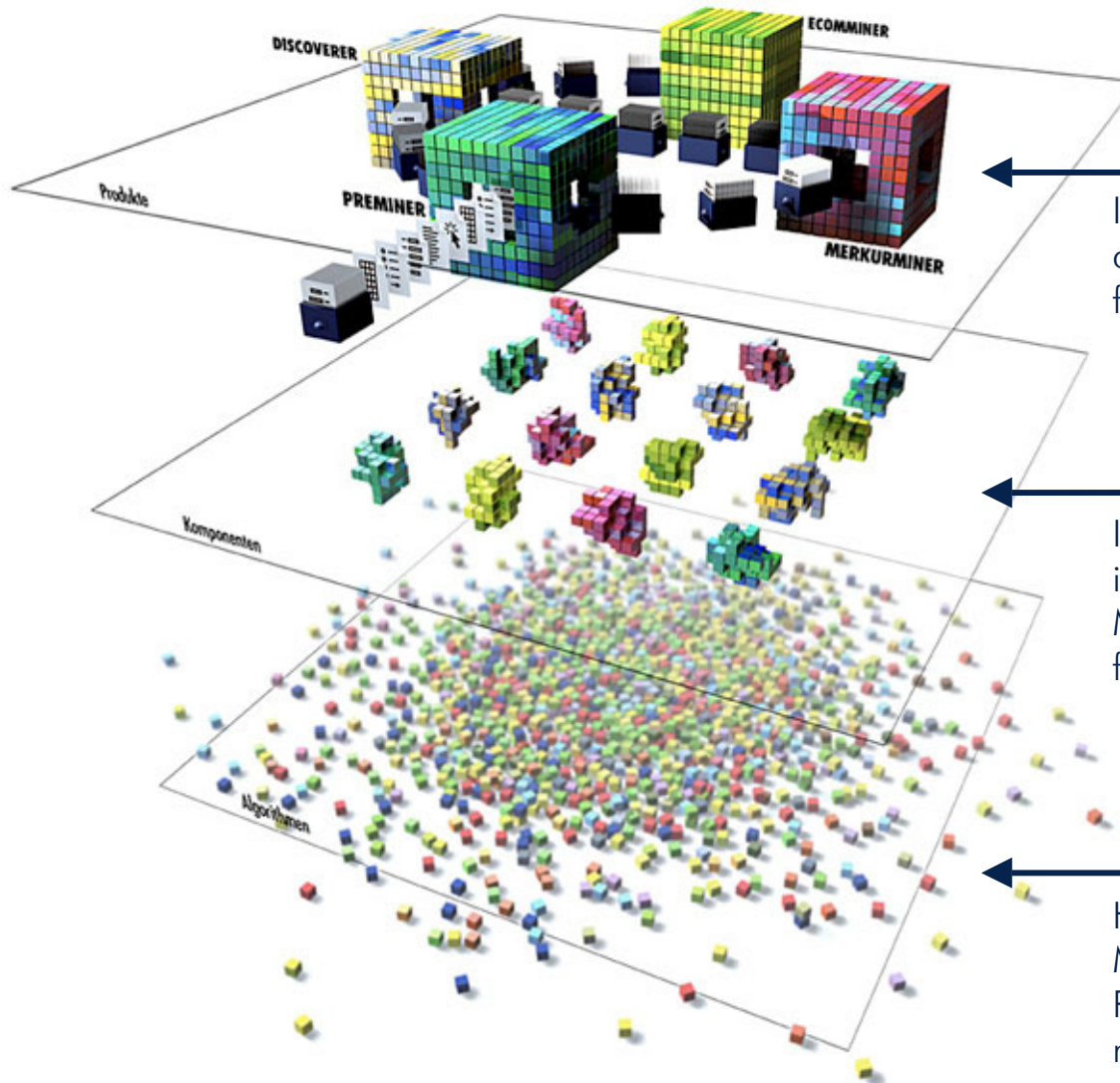
Unser Unternehmensprofil – Historie, Gegenwart und Zukunft



- außerordentlich innovatives Chemnitzer Software-Unternehmen auf stetigem Wachstumskurs, bereits 1998 als spin-off aus der TU Chemnitz ausgegründet
- interdisziplinäres Team (ca. 30 Mitarbeiter) aus hervorragenden Mathematikern und Informatikern (Patente, internationale Publikationen, Referate) sowie Marketingspezialisten und Vertriebsprofis
- Selbstverständnis: Software-Entwickler aber auch Berater und Dienstleister beim Analytischen CRM mit Data Mining (auf Kundenwunsch auch gern zusammen mit Partnern wie accenture, Mummert + Partner, KPMG, ...)
- seit 07/2001 ist die RWE-Tochter *envia* AG (2.700 Mitarbeiter / 2 Mrd. € Bilanzsumme) strategischer Partner, Mitgesellschafter von prudsys und „Multiplikator“ in der Branche Energieversorgung
- unsere Unternehmensphilosophie:
 - Wir verhelfen unseren Kunden zu einem deutlich besseren Verständnis ihrer Kunden (customer insight) und
 - unterstützen sie dabei, einen wirklichen, monetären Return-on-Investment aus ihren bisherigen oder zukünftigen operativen CRM-Investitionen zu ziehen.

- Vorstellung der Prudential Systems Software GmbH (demnächst prudsys AG)
- Überblick über unsere Produkte
- Beispiele für unsere Projekte
- kurze Demonstration des prudsys[®] DISCOVERER

Ergebnis unserer Forschung und Entwicklung: ein Baukastensystem für individuelle Lösungen



Software-Endprodukte

Intuitiv bedienbar und direkt von der Fachabteilung gewinnbringend für Kundenanalysen einzusetzen!

Software-Komponenten

In operative CRM- / ERP-Systeme integrierbar („Data Mining unter der Motorhaube“ – beliebige Ampel-funktionen am POS möglich!)

Data-Mining-Verfahren

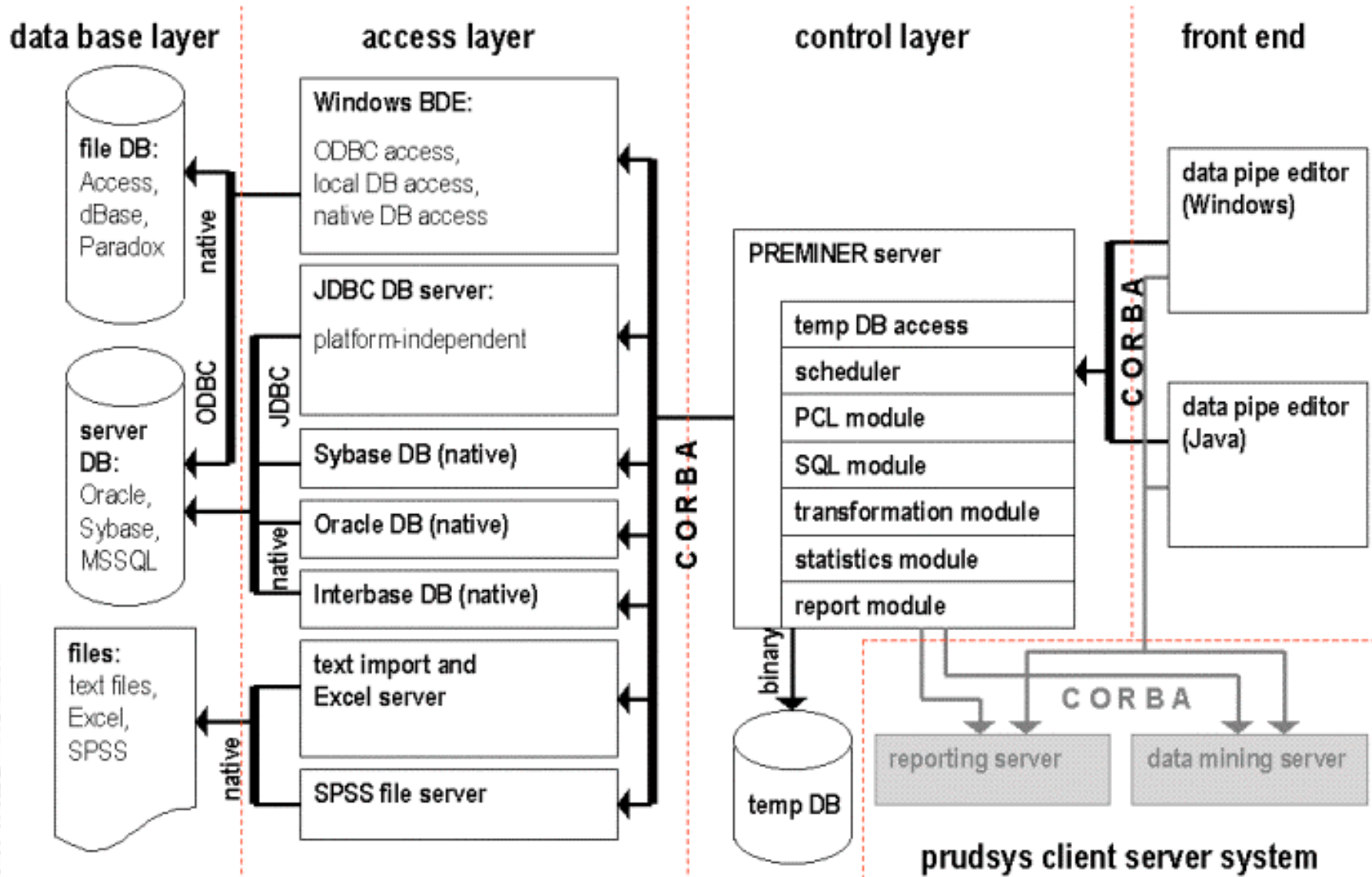
Höchster Wirkungsgrad unserer Data-Mining-Verfahren weltweit. Durch Patente und internationale Publikationen nachweis- und lizenzierbar!



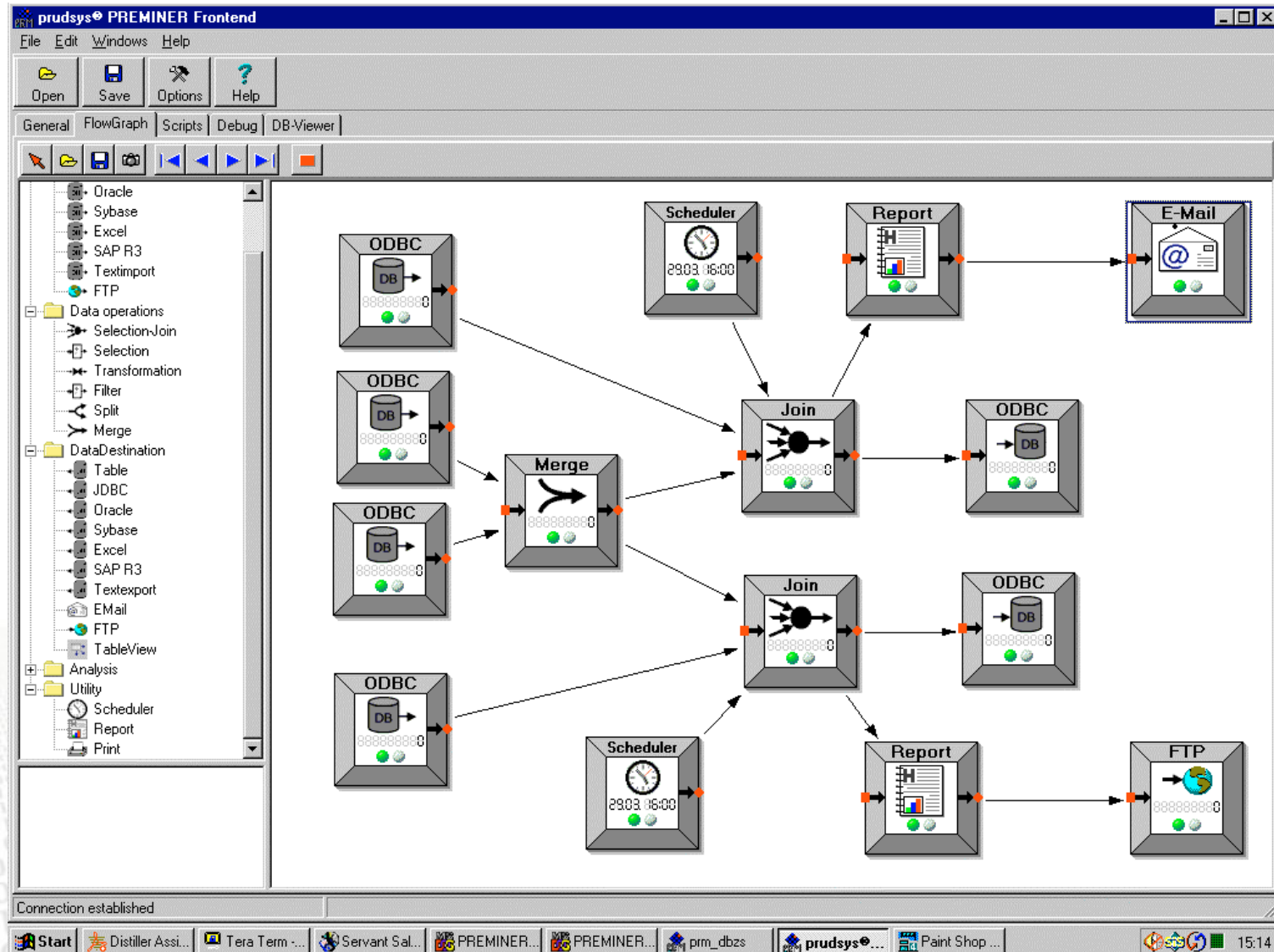
Vorverarbeitungswerkzeug für das
Zusammenführen und Bereinigen von
Daten aus unterschiedlichen Quellen

■ Merkmale

- plattformunabhängige Client/Server-Architektur
- komplexe Datenbereinigung (wichtig vor allem im Rahmen von Data-Mining-Analysen)
- aussagefähige Visualisierung komplexer Statistiken
- schnelle Verarbeitung von Daten aller gängigen Datenbanksysteme
- zeitgesteuerte Abarbeitung von komplexen Abläufen
- individuelles und automatisiertes Erstellen von Reports



prudsys® PREMINER: Screenshot des Frontends



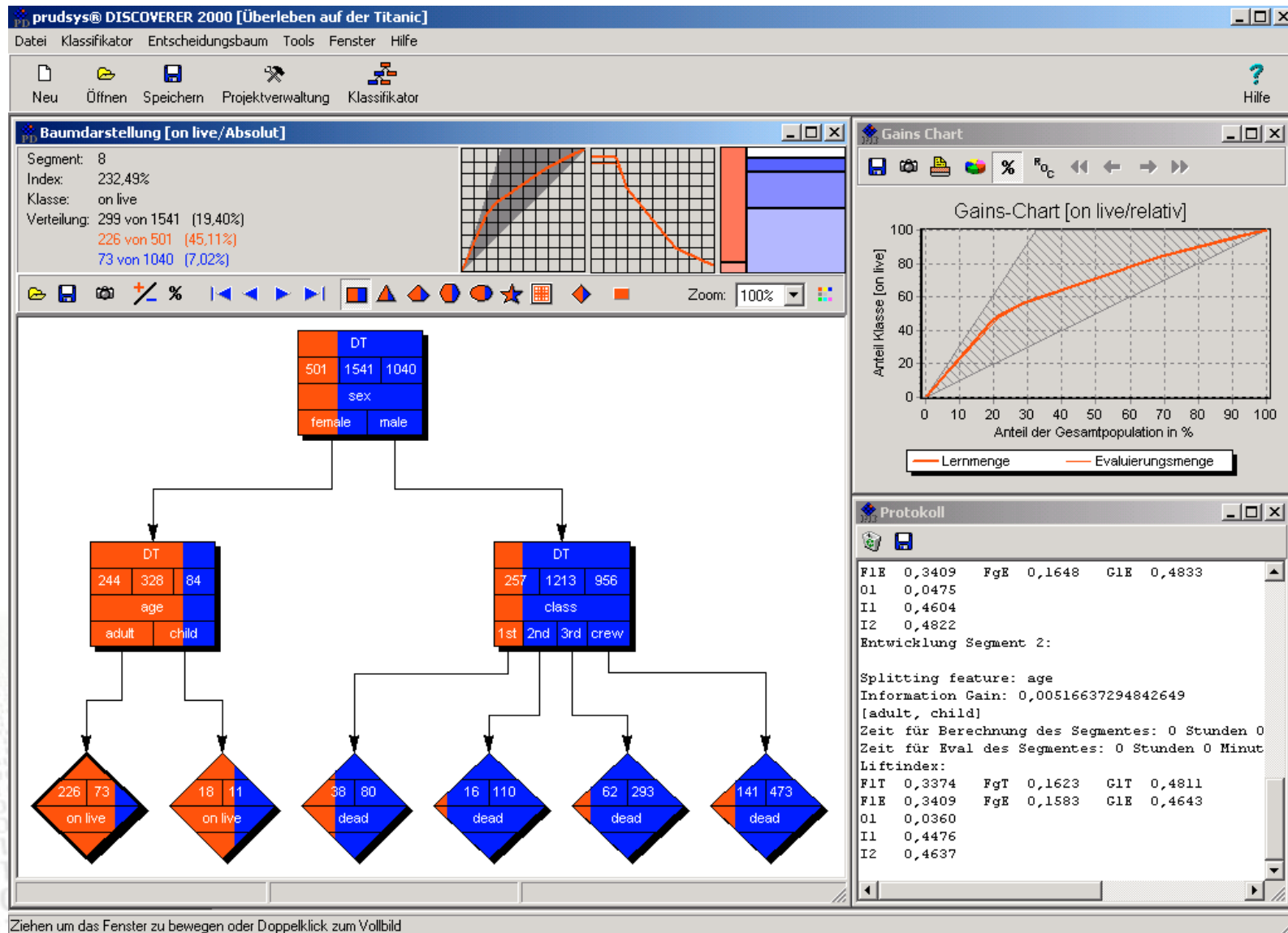


Data-Mining-Software zur Individualisierung des Massenmarketings

■ Merkmale

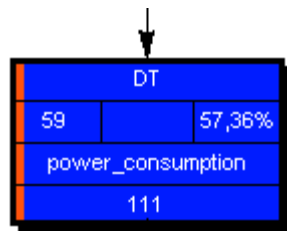
- große Auswahl leistungsfähiger Algorithmen für beste Analyseergebnisse bei spezifischen Voraussetzungen
- intuitive Programmführung und visuelle Kontrolle der Analyse
- wahlweise interaktive oder voll automatische Data-Mining-Analyse
- permanente Qualitätskontrolle beim Erstellen von Profilen und Vorhersagemodellen
- direktes ROI-Feedback durch Business-Charts (z.B. Gains, Profit, Lift)

prudsys® DISCOVERER: Screenshot

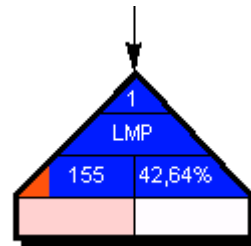


prudsys® DISCOVERER: Trennfunktionen im nichtlinearen Baum

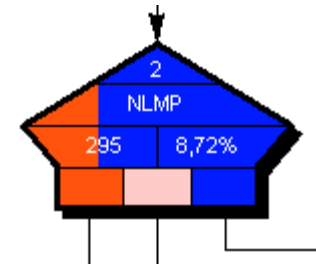
Der Funktionsumfang kombinierter Entscheidungsbäume reicht von einfachen Funktionen wie z.B. achsenparallelen Trennfunktionen bis hin zu komplexen nichtlinearen Funktionen.



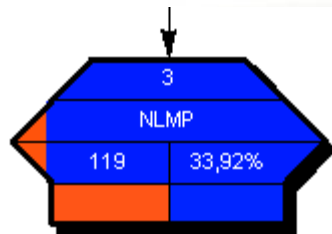
achsenparallel



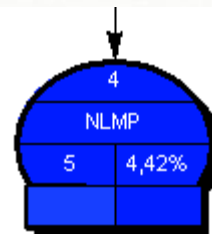
lineare Hyperebene



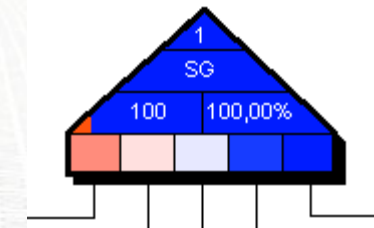
quadratische Hyperfläche



Hyperfläche 3. Grades



Hyperfläche 4. Grades



Sparse Grids Kurve

- Basiert auf BDE (Borland Database Engine),
- Zugriff auf:
 - Text- und Excel-Dateien,
 - dBase- und Paradox-Datenbankdateien,
 - SPSS-Dateien,
 - alle Datenbanken über ODBC,
 - ausgewählte Datenbanken über BDE-native Treiber
- Erfahrungen haben wir speziell mit:
 - Oracle 8i, Sybase, Interbase, Access, MySQL,
- Zugriff auf bestimmte weitere Datenbanken wird garantiert.

- Abhängig von der Anzahl der Datensätze und der Anzahl Merkmale, die analysiert werden sollen,
- Beispiel: Pentium III, 900 MHz, 512 MByte RAM ausreichend für 1 Mio. Datensätze und gleichzeitig 50 Merkmale (50 Mio. Zahlen),
- damit ist noch immer eine interaktive Arbeit möglich,
- Klassifikation praktisch beliebig großer Datenmengen möglich,

Betriebssysteme:

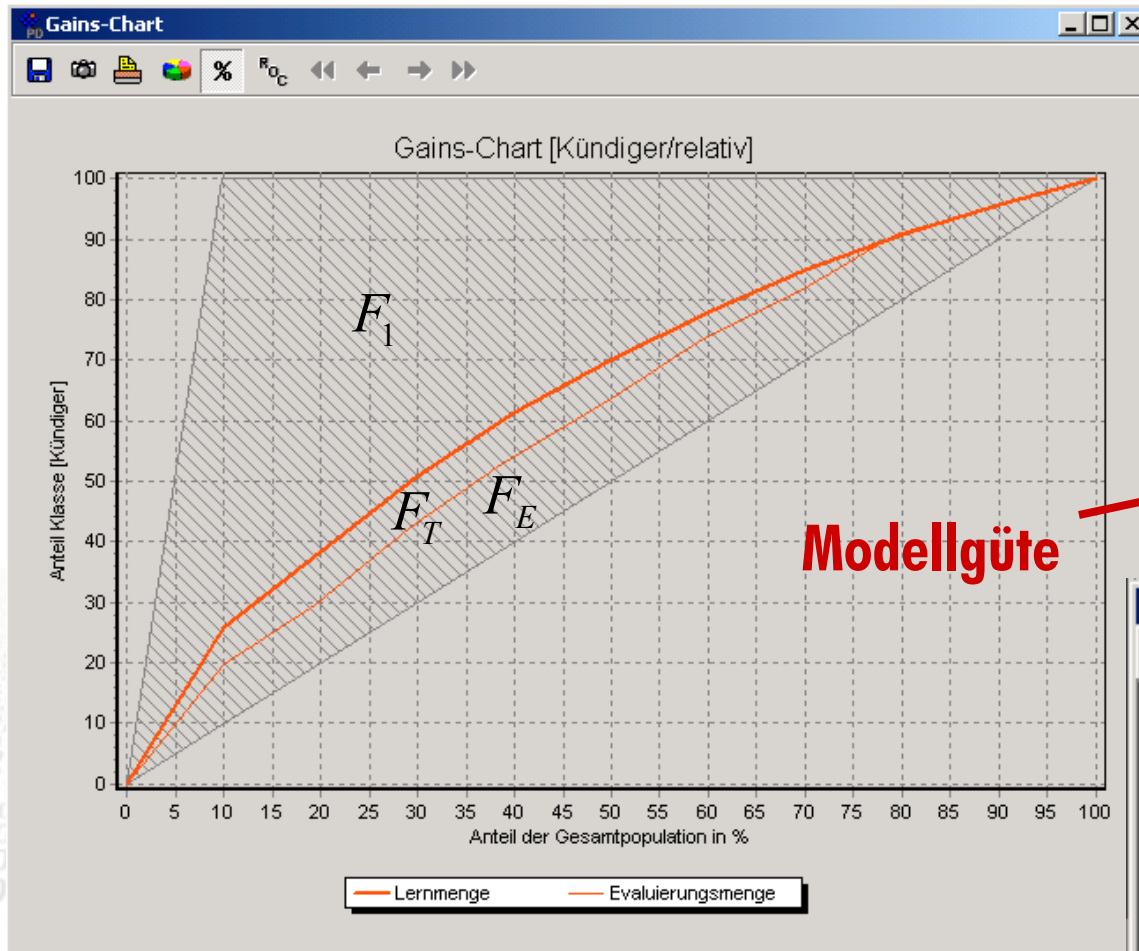
- Windows 95/98/ME,
- Windows NT/2000,
- Windows Terminalserver.

prudsys® DISCOVERER: Featurematrix DISCOVERER 3.1



Funktion	Version		
	Foundation	Standard	Professional
Data-Mining-Verfahren			
achsenparallele Bäume	X	X	X
lineare Bäume (Optimierer von prudsys)		X	X
lineare Bäume (schneller Optimierer)			X
Sparse-Grids-Bäume			X
Module			
Textimport	X	X	X
Analyse-Modul	X	X	X
Datenbank-Wizzard			X
Export von Klassifikatoren (SQL, PMML)			X
Datenbank Schnittstellen			
Import von Excel-, SPSS- und Textdateien	X	X	X
dBase, Paradox	X	X	X
Import aus ODBC-Datenbanken		X	X
native Datenbank-Import			X
schnelle binäre temp. Datenspeicherung	X	X	X
Legende: X - Feature vorhanden, O - als optionales Modul erhältlich			

Sicherung der Generalisierbarkeit des Ergebnisses durch die Verwendung einer unabhängigen Evaluierungsmenge



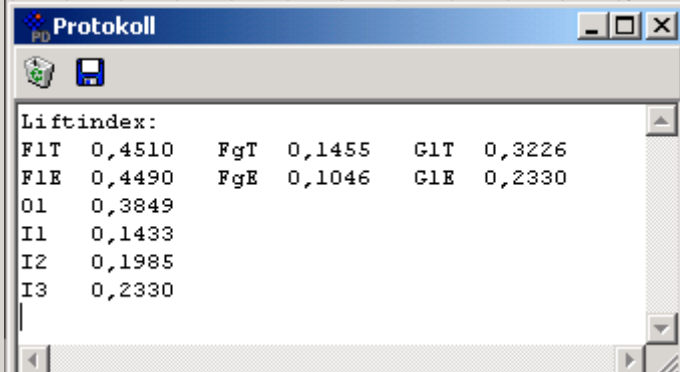
$$L_T = \frac{F_T}{F_1}$$

$$L_E = \frac{F_E}{F_1}$$

$$O = \frac{L_T - L_E}{L_E}$$

$$I_{p=1,2} = (1 - O^p) L_E$$

$$I_3 = L_E$$



Liftindex:			
F1T	0,4510	FgT 0,1455	G1T 0,3226
F1E	0,4490	FgE 0,1046	G1E 0,2330
O1	0,3849		
I1	0,1433		
I2	0,1985		
I3	0,2330		

prudsys® DISCOVERER: Optimierung mit dem Analysemodul

Analyse

Anteil Gesamt	Anteil Klasse [Reagierer]	Lift	Anzahl Klasse [Reagierer]	Anzahl Klasse [Nichtreagierer]
10%	69%	6,94	197	303
20%	89%	4,45	253	747
30%	94%	3,16	269	1231
40%	100%	2,50	284	1716
50%	100%	2,00	284	2216
60%	100%	1,67	284	2716
70%	100%	1,43	284	3216
80%	100%	1,25	284	3716
90%	100%	1,11	284	4216
100%	100%	1,00	284	4716

Verteilungsmatrix

Anteil: 15%
Reaktion (reale Klasse)

Aktion (vorhergesagte Klasse)	Reagierer	Nichtreagierer	Summe
Reagierer	245 (4,90%)	505 (10,10%)	750 (15,00%)
Nichtreagierer	39 (0,78%)	4211 (84,22%)	4250 (85,00%)
Summe	284 (5,68%)	4716 (94,32%)	5000 (100,00%)

Kostenmatrix

	Reagierer	Nichtreagierer	Währung bzw. Einheit:
Reagierer	50	-2,5	Euro
Nichtreagierer	0	0	Euro

Optimalwerte

Gewinn auf Evaluierungsmenge: 10987,50 Euro

OK Abbrechen Hilfe

• 50 EUR Gewinn je Reagierer

• 2,50 EUR Kosten je angeschriebenem Nichtreagierer

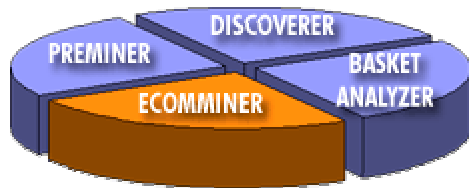
• Optimierung auf Knopf-Druck:

beste Ausstattungs-
dichte 15 %!

prudsys® DISCOVERER: Zusammenfassung der Vorteile



- Interaktive Generierung von Entscheidungsbäumen,
- Verarbeitung von großen Datenmengen (1 Million Datensätze und mehr),
- sehr leistungsfähige lineare und nichtlineare Trennfunktionen (Sparse Grids, entwickelt gemeinsam mit Uni Bonn, patentiert, weltweite Vorstellung auf der KDD im August),
- geringer Einarbeitungsaufwand,
- gutes Preis-Leistungs-Verhältnis.



Spezial-Data-Mining- Software für den Einsatz in E-Commerce-Systemen

■ Merkmale

- plattformunabhängige Client/Server-Architektur
- leistungsfähigste Warenkorbalgorithmen
- schnelle Formulierung von Abfragen durch Makro-Konzept (z.B. „Welche Produktkombinationen werden im Web-Shop von wem, wann, wie oft gekauft oder auch nur angesehen?“)
- individuelle Erstellung automatisierter Reports und Auswertungen
- umfangreiche Auswertungsmöglichkeiten durch Kopplung von Shop-Datenbank und Web-Server-Logfiles

- Klassische Warenkorbalgorithmen (sehr schnell für große Datenmengen 100.000-1.000.000 Datensätze),
- Verfahren zur Sequenzanalyse mit und ohne Zyklen (schnell für mittlere bis große Datenmengen, d. h. 20.000-1.000.000),
- sequentielle Warenkorbanalyse (schnell für große Datenmengen)

prudsys® ECOMMINER®: Screenshot Programm

Run Basket Analysis

Run | View Rules | View Baskets | Load / Save | Export


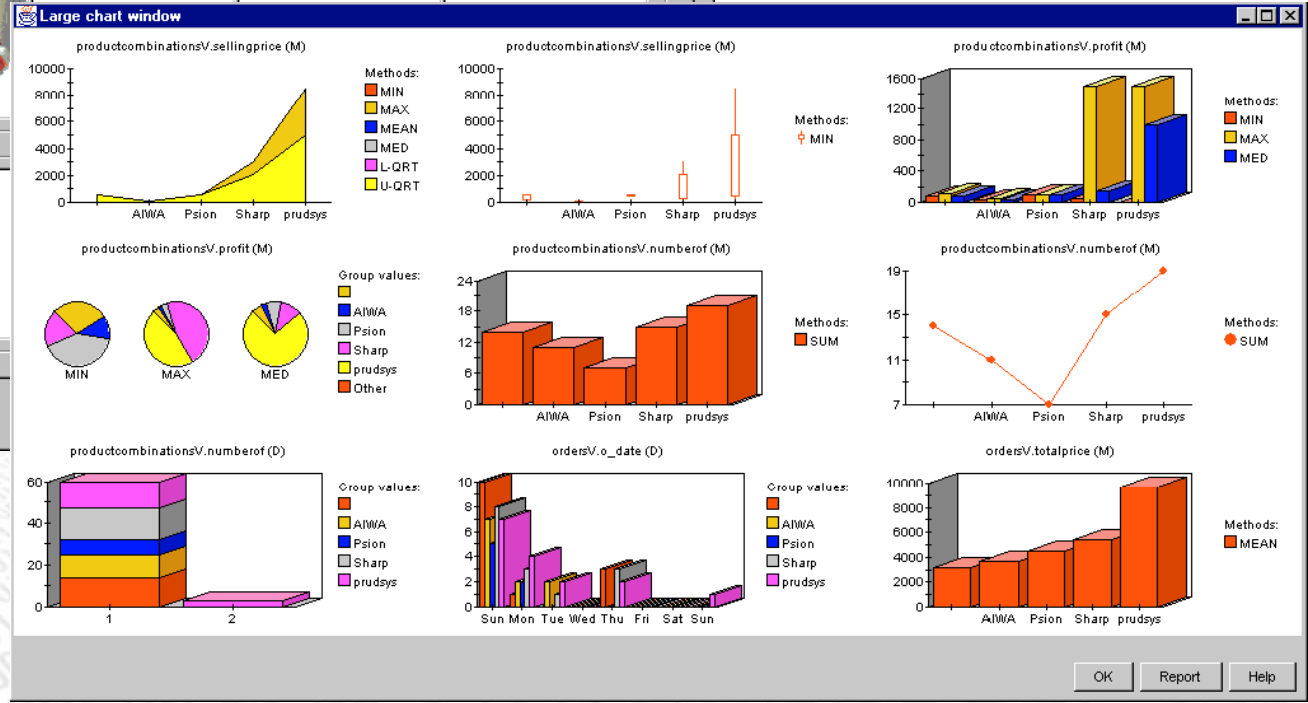
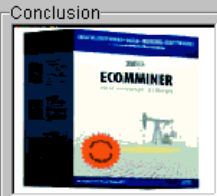
Selected association rule:

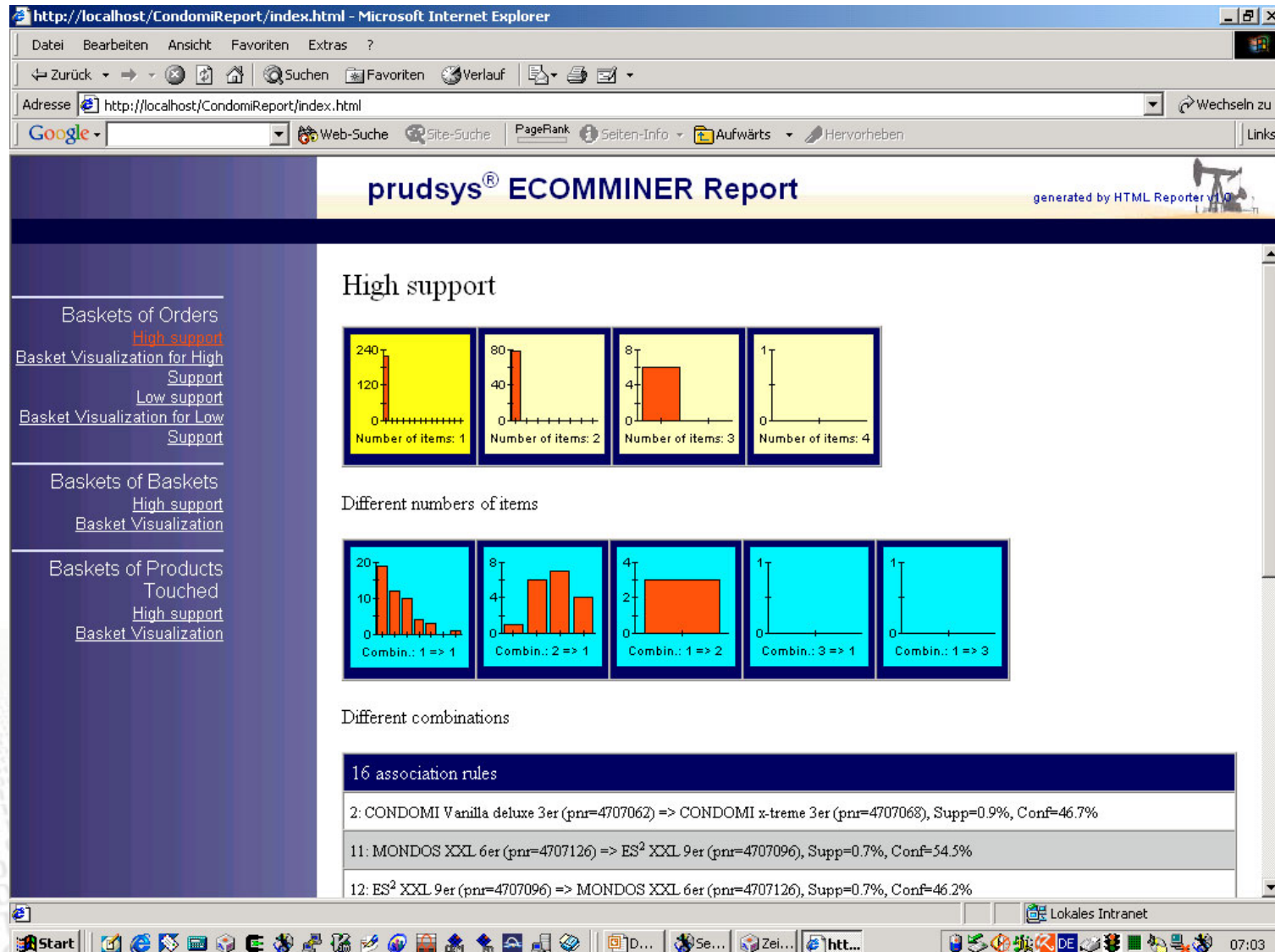
- 5: prudsys ECOMMINER (pnr=1001), Aiwa AM/FM Stereo (pnr=10328) => Body bridge (pnr=10335), Supp=5.2%, Conf=50.0%
- 6: Aiwa AM/FM Stereo (pnr=10328), Super Bundle Plus (pnr=10804) => prudsys ECOMMINER (pnr=1001), Supp=5.2%, Conf=60.0%**
- 7: Super Bundle Plus (pnr=10804) => prudsys ECOMMINER (pnr=1001), Aiwa AM/FM Stereo (pnr=10328), Supp=5.2%, Conf=37.5%

Load visualization: Picture sent successfully!

Report Status:

Premise





Plattformunabhängiges
Tool zur Durchführung
schneller und leistungsfähiger
Warenkorbanalysen

■ Merkmale

- plattformunabhängige Java-Implementierung
- leistungsfähige Umsetzung der wichtigsten Warenkorbalgorithmen für schnelle Analysen von großen Datenmengen
- intuitive Analyse mit Wizard-Unterstützung
- flexibler Import von Textdateien und Zugriff auf beliebige Datenbanksysteme

prudsys® Basket Analyzer: Screenshot



Wizard

- 1 Verfahren
- 2 Datenquelle
- 3 Spaltenauswahl
- 4 Analyseparameter**
- 5 Analyse

Startparameter

Minimaler Support:

Minimale Konfidenz:

Filter für Regellänge

Minimale Regellänge:

Maximale Regellänge:

Vorgabe für Anzahl der Regeln

Minimale Anzahl Regeln:

Maximale Anzahl Regeln:

Zurück Abbruch Weiter

Wizard

prudsys Basket Analyzer Ergebnisse

Zusammenfassung

Transaktionen: 35
Items: 10

Start Reading: Thu Jun 27 07:08:54 CEST 2002
Start Calculation: Thu Jun 27 07:08:54 CEST 2002
End Calculation: Thu Jun 27 07:08:55 CEST 2002

Assoziationsregeln

Anzahl gefundener Regeln: 151

Nr.	Premise		Conclusion	Support	Konfidenz
1	Milch, Käse	→	Wurst	8.57	100.00
2	Milch, Wurst	→	Käse	8.57	100.00
3	Chips	→	Windeln	5.71	100.00
4	Chips	→	Bier	5.71	100.00
5	Bier, Chips	→	Windeln	5.71	100.00
6	Windeln, Chips	→	Bier	5.71	100.00
7	Chips	→	Windeln, Bier	5.71	100.00
8	Brot, Milch, Käse	→	Wurst	5.71	100.00
9	Brot, Milch, Wurst	→	Käse	5.71	100.00
10	Eier	→	Butter	2.86	100.00
11	Butter, Milch	→	Brot	2.86	100.00

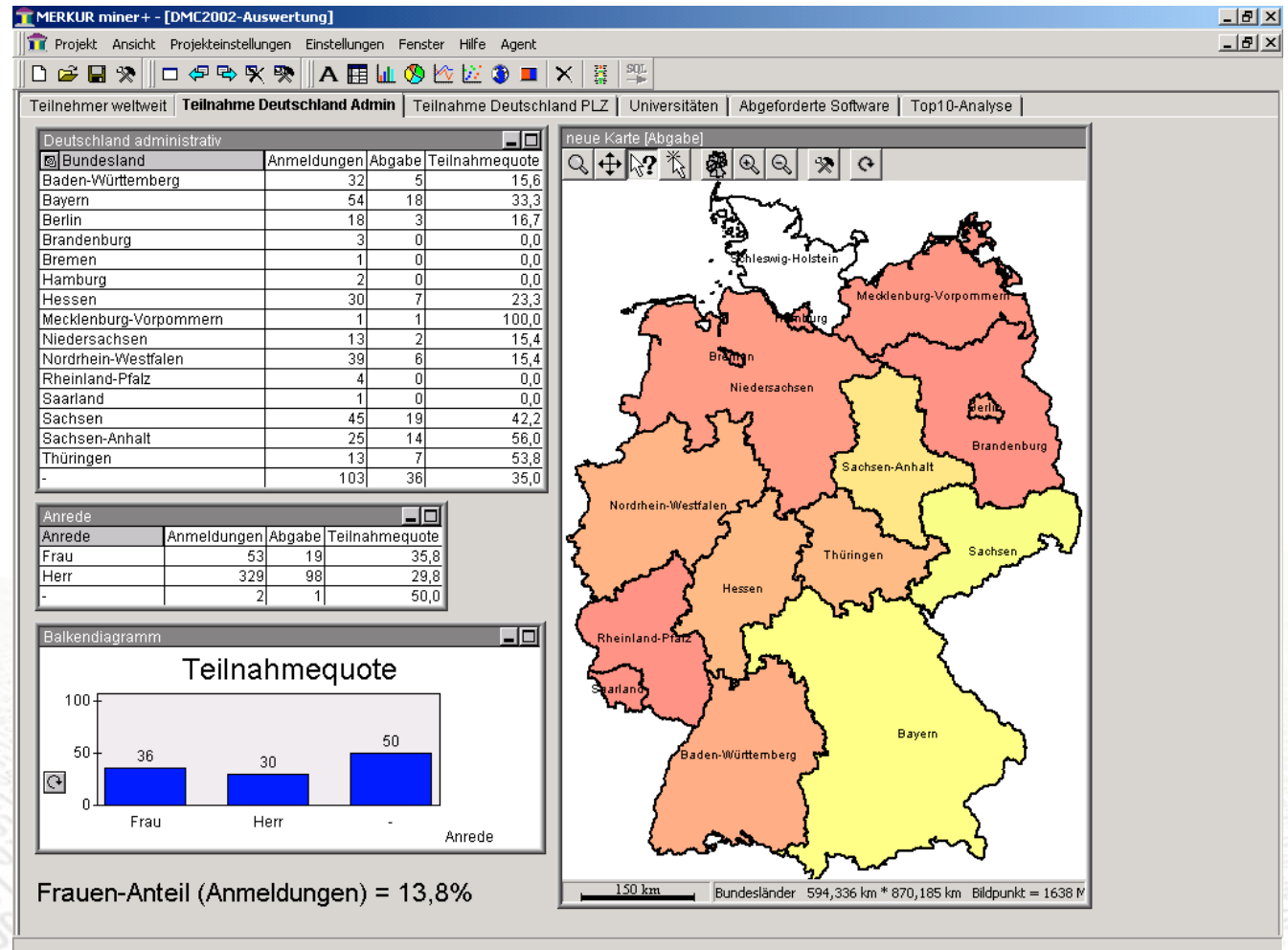
Sortieren nach: **Konfidenz** Drucken

Zurück Abbruch Beenden

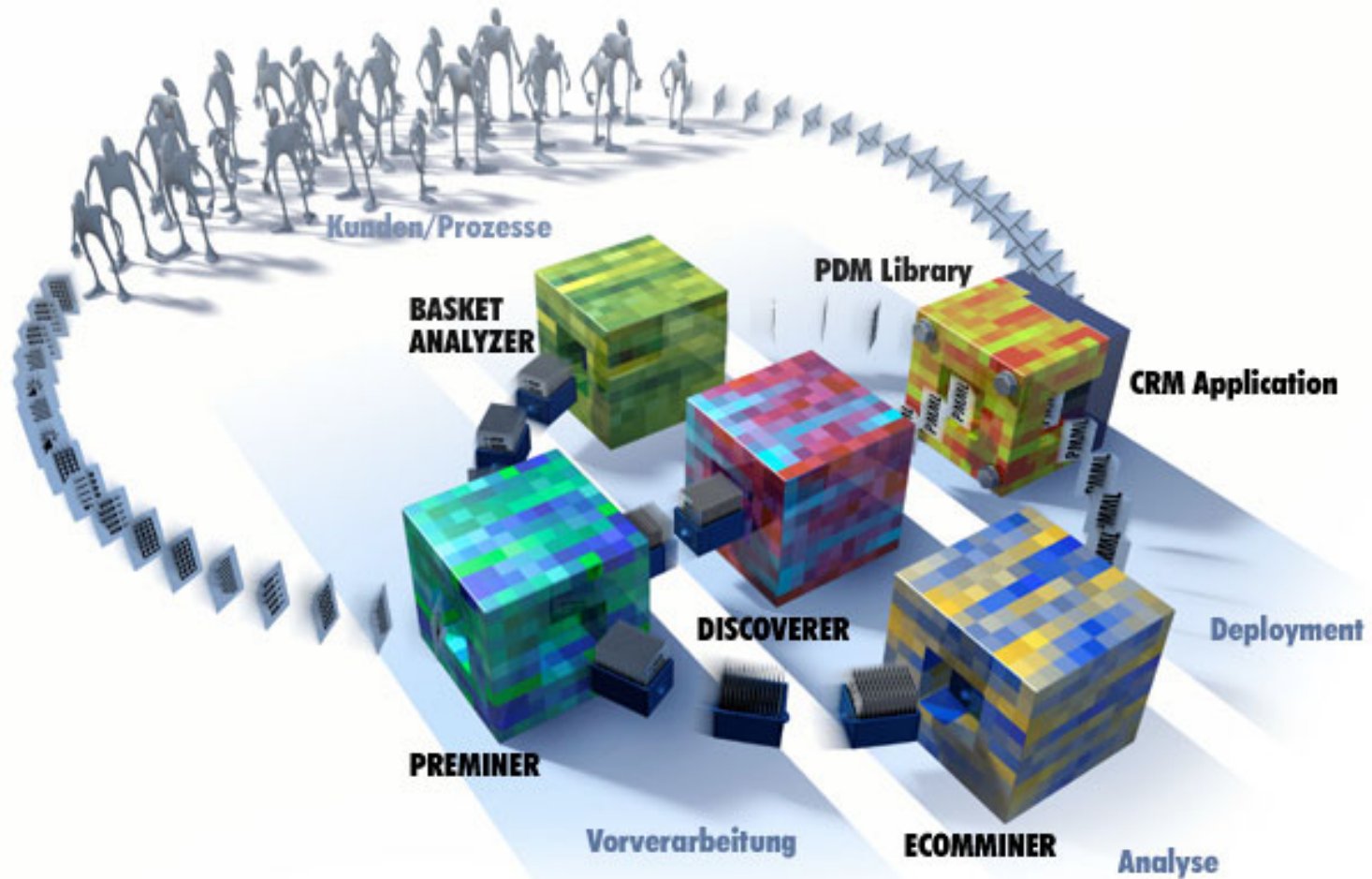
Unsere Produkte: neu ab Herbst -- prudsys® MERKURminer+

Kombination
von **OLAP**,
GIS und
Data Mining,

basierend
auf der
Technologie
von QlikTech
und prudsys

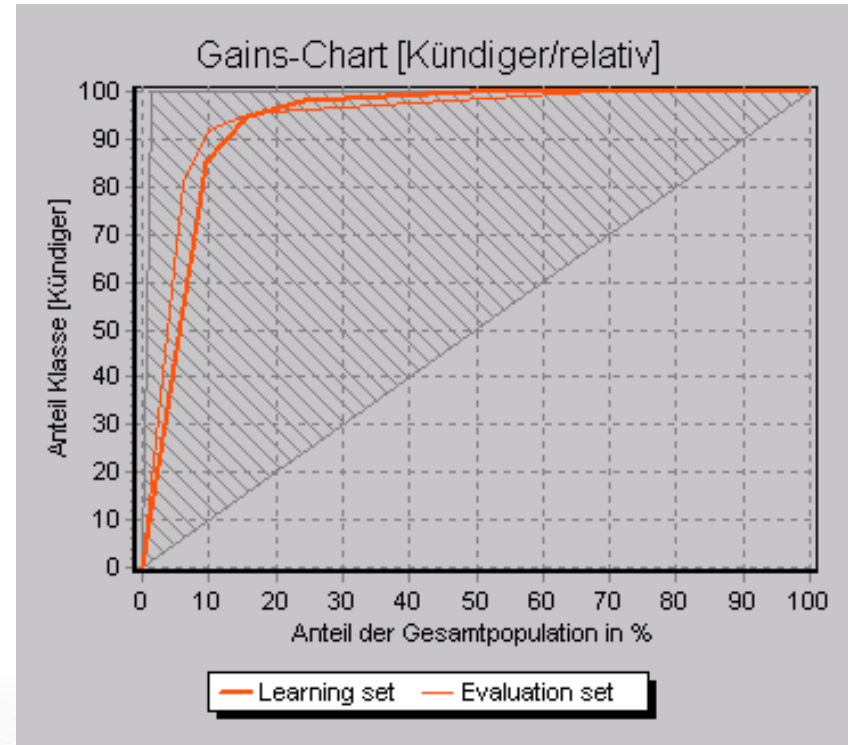
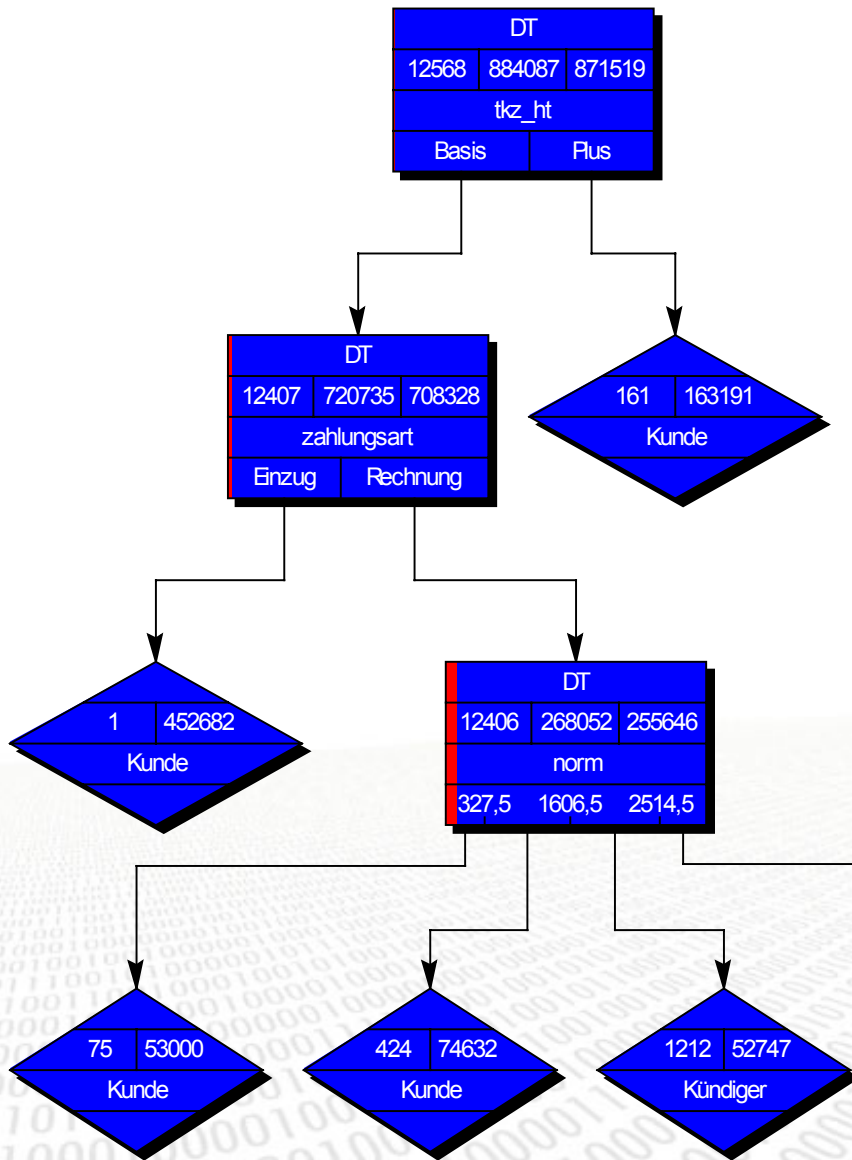


Datenaustausch über PMML: Verbindung der prudsys® Produktfamilie



- Vorstellung der Prudential Systems Software GmbH (demnächst prudsys AG)
- Überblick über unsere Produkte
- **Beispiele für unsere Projekte**
- kurze Demonstration des prudsys® DISCOVERER

Projekte: Kündigeranalysen bei verschiedenen EVUs



9% aller Kunden entsprechen dem Profil von 85% aller Kündiger

Projekte: Vergleich Mikrogeographie-Anbieter

- Vergleich von sechs Anbietern im Rahmen eines Marketing-Projektes bei envia
- Ziel: Finden der besten Daten für die Kündigerprävention und die Mailingoptimierung

Vergleichskriterien	Wichtung	AZB	Claritas	Consodata	GfK	Microm	Schober
Anteil zuordenbarer Kunden	15%	■ ■ ■	■ ■	■ ■	■ ■	■ ■	■ ■
Vorhersagegüte	25%	■ ■	■ ■	■ ■	■ ■	■ ■	■ ■
Individuelle Lösungsbeiträge	5%	■ ■	■ ■	■ ■	■ ■	■ ■	■ ■
Aussagekraft	15%	■ ■ ■	■ ■	■ ■	■ ■	■ ■	■ ■
Qualität der Zusammenarbeit	5%	■ ■	■ ■	■ ■	■ ■	■ ■	■ ■
Kosten der Datenanreicherung	35%	■ ■	■ ■	■ ■	■ ■	■ ■	■ ■
ERGEBNIS	100%	■ ■ ■	■ ■	■ ■	■ ■	■ ■	■ ■

Projekte: Integration der Data-Mining-Scores ins CRM-System

QlikView - Händlerexemplar - [E:\Projekte\E-World\eworld-0.3.qvw]

Datei Bearbeiten Ansicht Auswahl Layout Einstellungen Bookmarks Objekt Fenster ?

Call-Center View | Details zum Milieu | Grafische Auswertungen

Auswahl Kunden-Nummer: 702

Christian Scholz
Musterstraße 99
99999 Musterstadt

Verbrauch (kWh): 1934
Zahlungsart: Bankeinzug
Umsatz (in €): 380,44

Kundenkarte: Ja
Kundentyp: Strom & Netz
Strom-Produkt: Standard

Empfehlung: Multi-Utility-Produkte empfehlen!

Verhaltensprognose

Kündigungsrisiko: niedrig
Karten-Affinität: niedrig
Präferenz-Produkt: Rund-Um Sorglos

Sozio-Typen

Milieu-Schätzung: Status-orientiertes Arbeitermilieu
Dominantes Alter: 63 bis unter 70
Sekundäres Alter: 55 bis unter 63

Regio-Typen

Regio-Typ: -
Zentrums-Typ: -
Bebauungstyp: Wohngebiet
Straßentyp: 1 bis 4 WE/Gebäude


Bedürfnisprognose

Präferenz Kontaktstelle: Bevorzugt eine persönliche Beratung im Kundenzentrum

Präferenz Information: Wollen möglichst viel Informationen

Präferenz Multi Utility: Wollen Elektro-, Sanitär- und Heizungsinstallationsleistungen.

Präferenz Zusatzleistungen: Hoher Nutzen aus Service-Qualitätsgarantie und kostenloser Kundenkarte. Fast kein Bedarf nach Kundenzeitschrift.

Customer Insight Prototyp E-World Essen 2002 

Drücken Sie F1 für Hilfe 12.02.2002 15:22:46 D: 1/8530 F: 1/8530 13:35

Projekte: Analyse der Netzausfälle bei einem EVU

QlikView - Händlerexemplar - [E:\Projekte\envia Netze\Projekt-2001-002\Daten envia 2001-11-12\KombBesy-Analyse-0.4.1(hsi).qvw]

Störungen | **Analyse** | Standorte | Ableiter | E-Spulen | Hilfs-/Nebenanlagen | Leistungsschalter | Ölabscheider | Regelschaltwerke | Schutztechnik | Trafos | Trenner | Wandler | Strukturtabelle | Test

Ausfälle: 7

Status

Auswahlstatus	
Felder	Werte
UW_Name	● Altenburg

Datum

Jahr: 2000, 2001

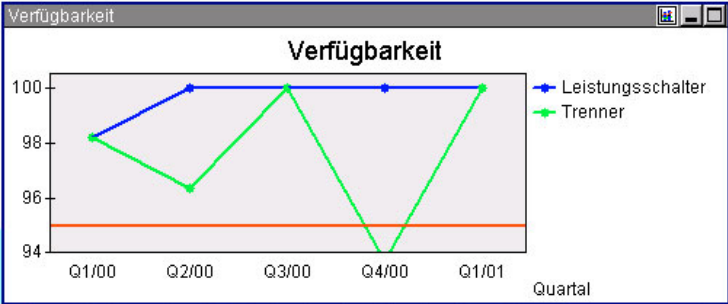
Quartal: Q1/01, Q3/00, Q1/00, Q2/00, Q4/00, Q2/01, Q4/01, Q3/01

Monat: Januar, April, Juni, November, Dezember, Februar, März, Mai, Juli, August, September, Oktober

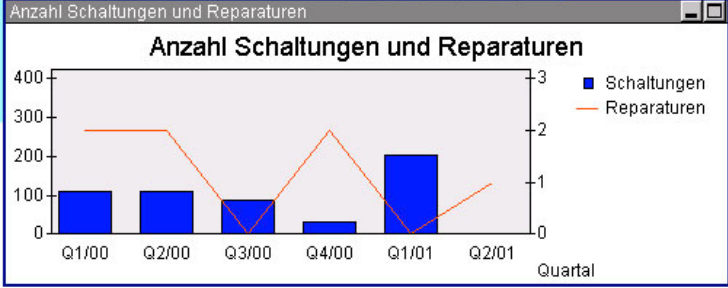
Tag: 1-31

Analysen

Verfügbarkeit



Anzahl Schaltungen und Reparaturen



Schaltungen						
Quartal	Schaltungen	Ausfälle	Rep.zeit	Fahrzeit	reg. AZ	Ber.zeit
Q1/00	110	2	17,5	3	20,5	0
Q2/00	110	2	25	2	27	0
Q3/00	86	0	0	0	0	0
Q4/00	32	2	51	5	56	0
Q1/01	202	0	0	0	0	0
Q2/01	0	1	1	1,5	0	2,5

Auswahl aufheben

Ort/Gerät

Umspannwerk (Geräte)

- Altenburg
- Bad Dübén
- Belgern
- Bennewitz
- Borna
- Böhllitz-Ehrenb
- Böhllitz-Ehrenberg
- Brandis
- Dellitzsch
- Dornreichenbach
- Dornreichenbach
- Döbeln

Geräteklasse

- Ableiter
- E-Spulen
- Hilfs- und Nebenanlagen
- Leistungsschalter
- Ölabscheider
- Regelschaltwerke
- Schutztechnik HS
- Schutztechnik MS
- Trafos
- Trenner
- Wandler

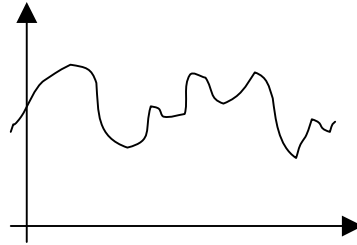
Arbeitszeit/Bereitschaft

- Bereitschaftszeit
- reguläre Arbeitszeit

Start | 13:26

Projekte: Prognose der Lastprofilverläufe

Testgruppe



Kunde	V ₁	V ₂	W ₁	W ₂
1	x	x	x	x
2	x	x	x	x
...

übrige Geschäftskunden

Kunde	V ₁	V ₂	W ₁	W ₂
1	x	x	x	x
2	x	x	x	x
...

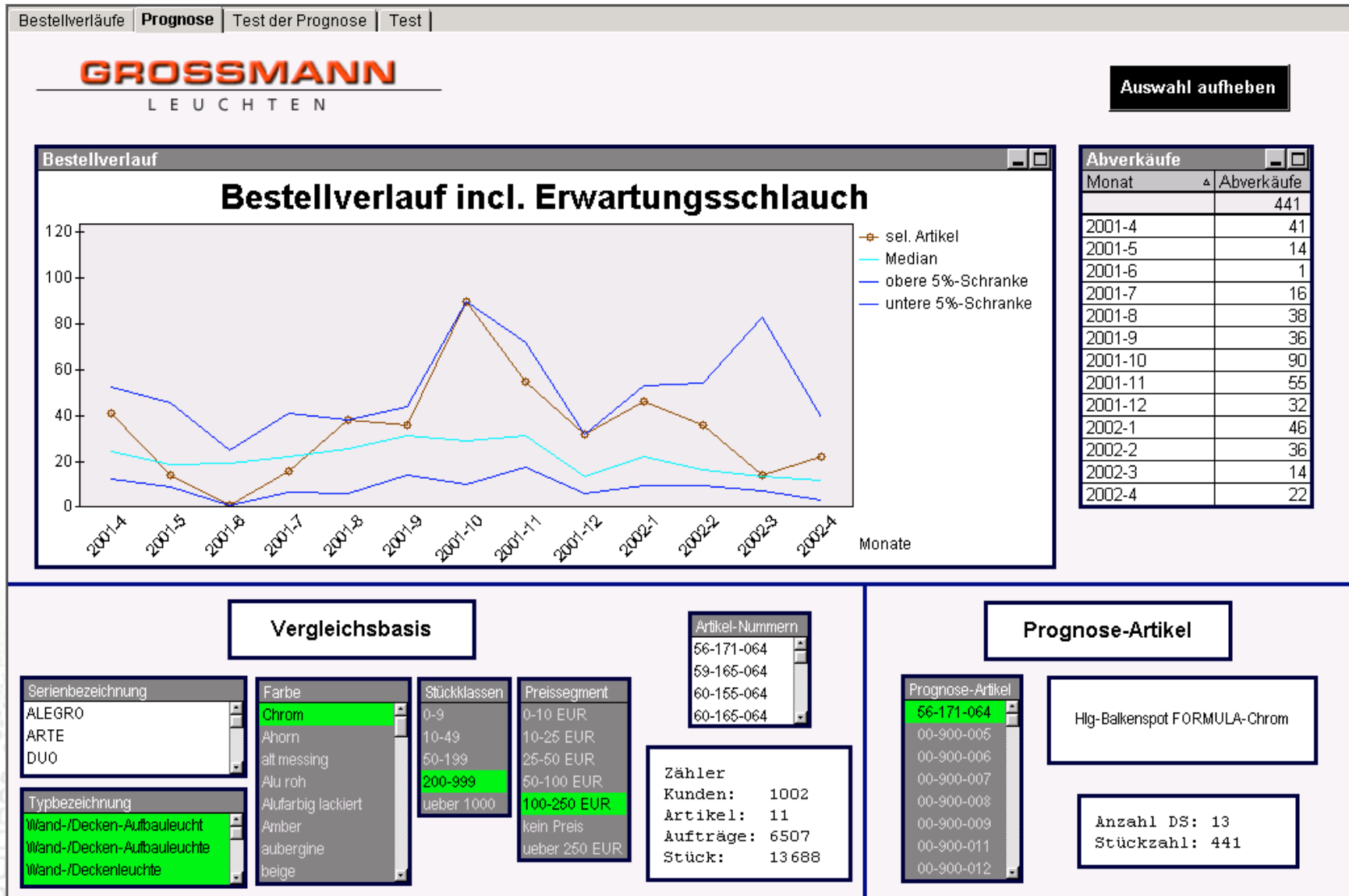
BTU Lastprofilsegmente:

- Segment 1
- Segment 2
- ...
- Segment 11

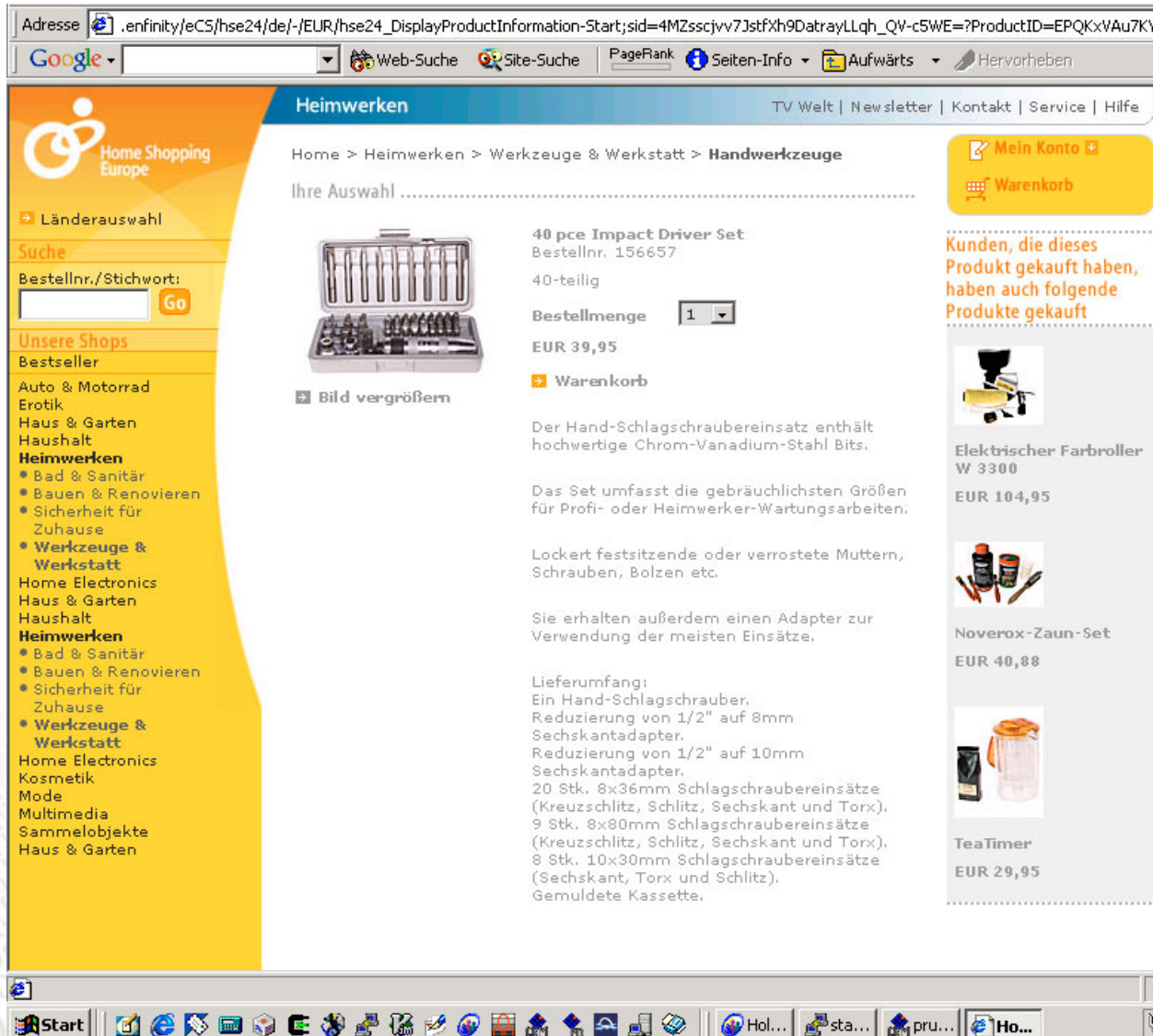
$F(V_i, W_i)$

Data Mining

Projekte: Prognose der Bestellzahlen bei einem Leuchtenhersteller



Projekte: Empfehlungs-Engine bei Home Shopping Europe



The screenshot shows a web browser window displaying the Home Shopping Europe website. The browser's address bar shows the URL: `.infinity/eCS/hse24/de/-/EUR/hse24_DisplayProductInformation-Start;sid=4MZsscjv7JstfXh9DtrayLLqh_QV-c5WE=?ProductID=EPQKxVAU7KY`. The website header includes the Home Shopping Europe logo and navigation links for 'Heimwerken', 'TV Welt', 'Newsletter', 'Kontakt', 'Service', and 'Hilfe'. The main content area is titled 'Heimwerken' and shows a breadcrumb trail: 'Home > Heimwerken > Werkzeuge & Werkstatt > Handwerkzeuge'. The product being viewed is a '40 pce Impact Driver Set' (Bestellnr. 156657), priced at EUR 39,95. A 'Warenkorb' button is visible. To the right, a recommendation section titled 'Kunden, die dieses Produkt gekauft haben, haben auch folgende Produkte gekauft' lists three items: 'Elektrischer Farbroller W 3300' (EUR 104,95), 'Novorex-Zaun-Set' (EUR 40,88), and 'TeaTimer' (EUR 29,95). The left sidebar contains a search bar, a 'Länderauswahl' dropdown, and a list of categories under 'Unsere Shops', including 'Bestseller', 'Auto & Motorrad', 'Haus & Garten', and 'Heimwerken'.

Projekt-Angebot: Kick-off-Workshop zum Data-Mining-Einsatz



- Sie erfahren alles Wissenswerte über den Data-Mining-Einsatz und wir untersuchen gemeinsam mit ihnen strategisch notwendige und sinnvolle Einsatzmöglichkeiten zur Kostenreduzierung, Kundengewinnung und Kundenbindung für ihr Unternehmen.
- Angebot* 1-Tages Kick-off-Workshop:
 - bei Ihnen vor Ort,
 - mit unbegrenzter Teilnehmeranzahl,
 - eine kostenfreie Test-Lizenz unserer Software „prudsys® DISCOVERER 2000“ mit umfassendem deutschen Handbuch,
 - zum Komplettpreis von 1.450 Euro zzgl. MWSt.
- *Es gelten die AGB der Prudential Systems Software GmbH.

- Vorstellung der Prudential Systems Software GmbH (demnächst prudsys AG)
- Überblick über unsere Produkte
- Beispiele für unsere Projekte
- **kurze Demonstration des prudsys® DISCOVERER**

Demonstration: prudsys® DISCOVERER 2000

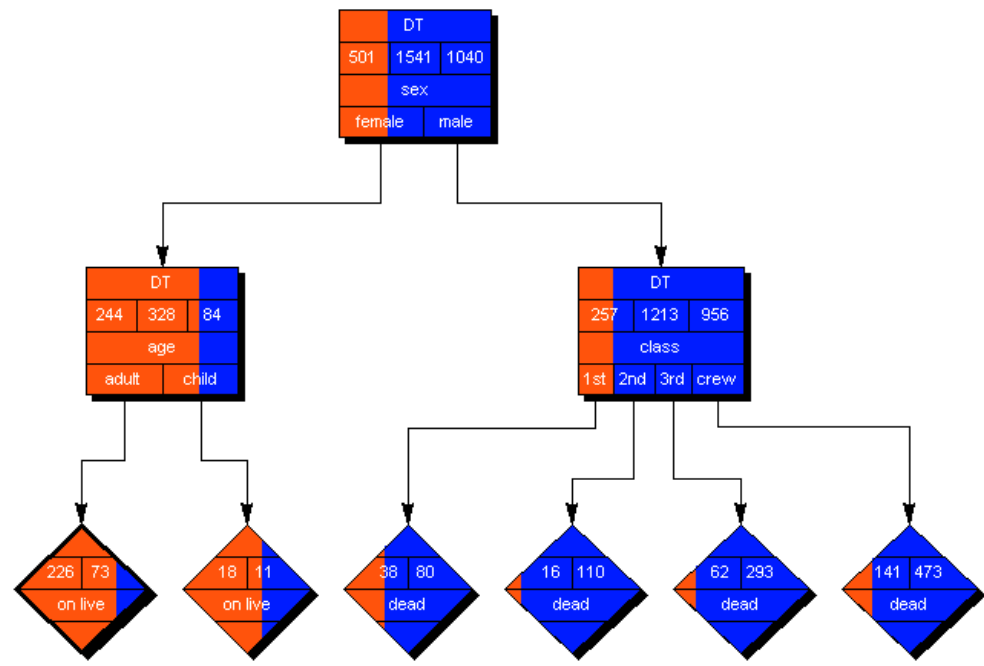
prudsys® DISCOVERER 2000 [Überleben auf der Titanic]

Datei Klassifikator Entscheidungsbaum Tools Fenster Hilfe

Neu Öffnen Speichern Projektverwaltung Klassifikator Hilfe

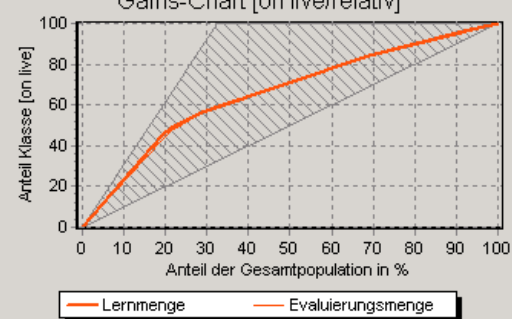
Baumdarstellung [on live/Absolut]

Segment: 8
Index: 232,49%
Klasse: on live
Verteilung: 299 von 1541 (19,40%)
226 von 501 (45,11%)
73 von 1040 (7,02%)



Gains Chart

Gains-Chart [on live/relativ]



Protokoll

```

FILE 0,3409  FgE 0,1648  G1E 0,4833
O1  0,0475
I1  0,4604
I2  0,4822
Entwicklung Segment 2:

Splitting feature: age
Information Gain: 0,00516637294842649
[adult, child]
Zeit für Berechnung des Segmentes: 0 Stunden 0
Zeit für Eval des Segmentes: 0 Stunden 0 Minut
Liftindex:
FIT 0,3374  FgT 0,1623  G1T 0,4811
FILE 0,3409  FgE 0,1583  G1E 0,4643
O1  0,0360
I1  0,4476
I2  0,4637
    
```

Ziehen um das Fenster zu bewegen oder Doppelklick zum Vollbild

■ Firmensitz

Prudential Systems Software GmbH
Bergstraße 61
09113 Chemnitz

■ Telefon/Fax

Tel.: (03 71) 2 70 93-0
Fax: (03 71) 2 70 93-90

■ Internet

info@prudsys.de
www.prudsys.de, www.ecomminer.com, www.data-mining-cup.de

■ Direkt

Holm Sieber <sieber@prudsys.de>
Tel.: (03 71) 2 70 93-61