

Vorhersage mit CA CleverPath™ Predictive Analysis Server

Laszlo Mihalka
Senior Consultant
Laszlo.Mihalka@ca.com

Agenda

- Herausforderung
- CP PAS
- Neugents
- Modelle
- Mustererkennung
- Tools
- CA World



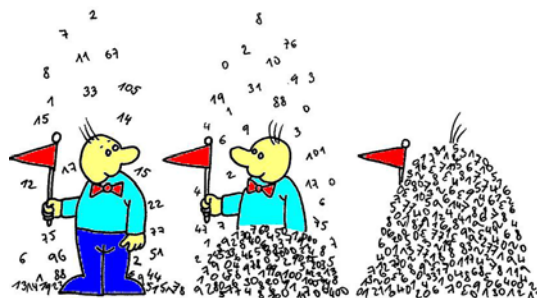
Herausforderung

- Aggressiver Wettbewerb
- Wichtigkeit von guten Kundenbeziehungen



Herausforderung

- Bedarf an personalisierten Diensten
- Große Datenbestände mit hohem Zuwachs



PAS: Predictive Analysis Server

Bietet eine Lösung mit Intelligenz auf Vorhersage um u.a.



- Systemausfälle zu minimieren
- Systeme gegen Mißbrauch zu schützen

PAS: Predictive Analysis Server

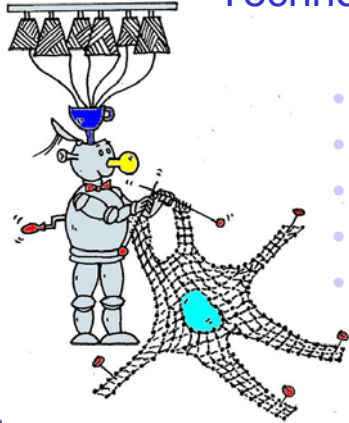
Bietet eine Lösung mit Intelligenz auf Vorhersage um u.a.



- Gewinn zu erhöhen
- Kundenbeziehungen zu verbessern
- Wettbewerbsvorteile zu erhöhen

Neugents

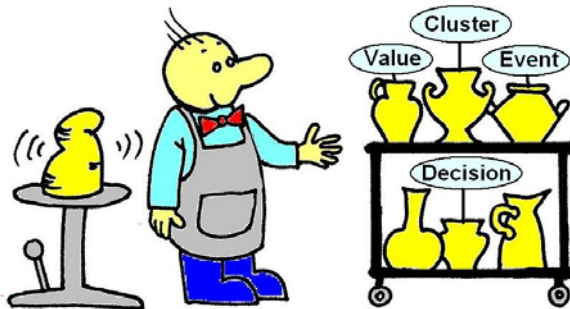
- PAS verwendet die patentierte Technologie „Neugents“ um



- Muster zu erkennen
- Zusammenhänge zu entdecken
- Regeln zu extrahieren
- Ereignisse zu prognostizieren
- Optimumwerte zu finden

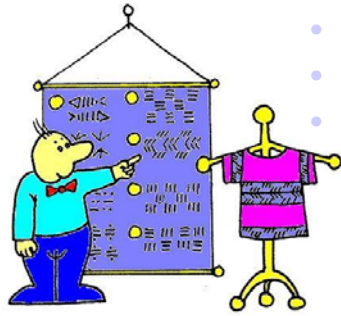
Modellbildung für Prognosen

- PAS entdeckt die Zusammenhänge zwischen Ursachen und Folgen, die
 - Für Experten zu subtil oder unbekannt sind
 - Zu komplex für andere statistischen Werkzeuge sind
- und bildet Modelle für Prognosen



Mustererkennung

- Identifizierung von bekannten und Entdeckung von neuen Mustern



- Muster: clusters / profiles gebildet.
- Übergänge zwischen Mustern
- Neue Muster:
 - Ausnahmefall
 - Modellbildung


Real-Time Intelligenz

- Verwendung von Modellen auch real-time für




- Betrugserkennung
- Personalisierung
- Cross-Selling
- Prozess-Steuerung
- Systemsicherheit

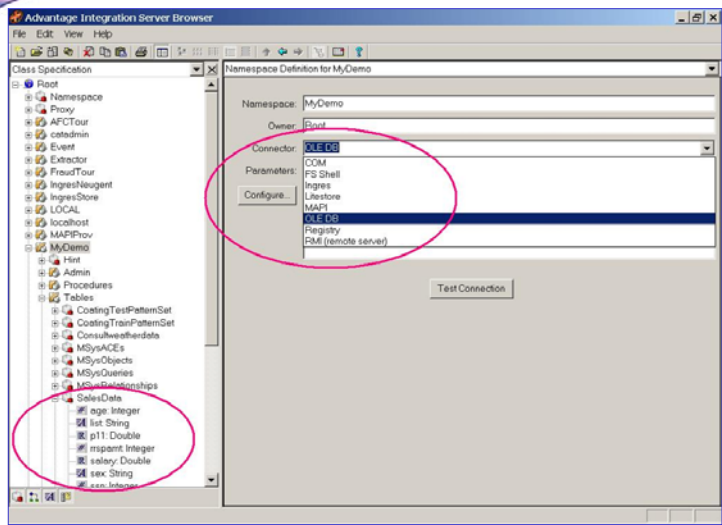
Tools




- Advantage Intergration Server
 - Einbindung von Datenquellen
- Neugents IDE
 - Selektion / Manipulation von Daten, Statistiken
- Provider Wizard
 - Modell-Bildung und Verifizierung
- Optimization Wizard
 - Suchen nach Optimumwerten
- Deployment Wizard
 - Web Services, JSP, ASP, Java, COM, C++, VB
- Interaktive Tutorials
 - Beispiele


 Computer Associates®
 ca.com

Tools: Advantage Integration Server



The screenshot displays the Advantage Integration Server Browser interface. On the left, a tree view shows a hierarchy of class specifications under a 'MyDemo' namespace, including 'SalesData' with fields like 'age: Integer', 'list: String', 'p11: Double', 'mspaint: Integer', 'sales: Double', 'sex: String', and 'ssn: Integer'. On the right, the 'Namespace Definition for MyDemo' window is open, showing a 'Connector' dropdown set to 'OLEDB' and a list of parameters including 'COM', 'FS Shell', 'Ingres', 'Liststore', 'MAPS', 'OLEDB', 'Registry', and 'RMI (remote server)'. A 'Test Connection' button is visible at the bottom of the right window.


 Computer Associates®
 ca.com

Tools: IDE

The screenshot displays the IDE interface with three overlapping dialog boxes:

- Clearse Rows:** Contains 'Analysis Options' (Range, Statistic, Constraint) and a 'Description' field.
- Sample Data:** Shows 'Options' (Random sample, First, Every other, Balance feature) and a list of features: ssn, sex, age, salary, rrspamt, p11.
- Historian:** Shows 'Normal Features' (same as Sample Data), 'Historical Features' (Log: 4, Delta: 2, and a list of historical expressions like ssn(t), sex(t), age(t), salary(t), rrspamt(t), list(t), p11(t)), and 'Forecasted Features' (From: 0, Delta: 0).

At the bottom left, there is a small data table:

24	552947146	F
25	551862737	F
26	547617388	F
27	545656359	F
28	536685688	F

ca.com logo is visible in the bottom left corner.

Tools: Provider Wizard

- Auswahl vom passenden Neugent
- Festlegen von inputs / outputs

The Provider Wizard dialog box contains the following fields and options:

- Choose a namespace containing Neugent classes:** Text field with value "Root/Neugent/CA/Neugent".
- Choose the type of Neugent to create:** Dropdown menu with options: ClusteringNeugent, ClusteringNeugent (selected), DecisionNeugent, EventPredictNeugent, TextClusteringNeugent, ValuePredictNeugent.
- Enter the name of your new Neugent:** Text field with value "rmyNeugent".
- Input Field Names:** Text area containing "sex", "age", "salary", "rrspamt", "list".
- Label Field Name:** Text field with value "ssn".
- Output Field Names: (Optional):** Text area with value "p11".

ca.com logo is visible in the bottom left corner.

Tools: Provider Wizard- Clustering

- Z.B Kunden werden nach ihren Eigenschaften gruppiert
- Die „beste“ Gruppe kann ermittelt werden

Configure ClusteringNeugent

Threshold population: 99

Max level: 2

Max clusters per level: 8

Tree View:

- RootCluster
 - 1
 - 2
 - 2_1
 - 2_2
 - 3
 - 4
 - 5
 - 6
 - 7
 - 8

Locate Pattern with label:

Search

Cluster Attributes

Attributes	Value
Population	99
Center	41.4444, 29213.1, 69.1515
Radius	6.31112, 17881.5, 1341.18
normalizedRadius	0.140247

Population: 99

Export Summary to File

Tools: Provider Wizard- Value

- Z.B. Verkaufszahlen, Produktionswerte werden prognostiziert
- Qualität des Modells wird ermittelt

ValuePredictNeugent@myNeugent No. of Patterns loaded: 84

Output	TrainScore	TestScore	TrainError	TestError
thickness	99.5188	98.5894	8.66728e-005	0.00025406

Combined trained model quality is: 99.5187590691962

Choose Validation Data Percentage:

Smart Selection % of the training data to validate the model

Choose Validation Data From Database:

Use query Use native query

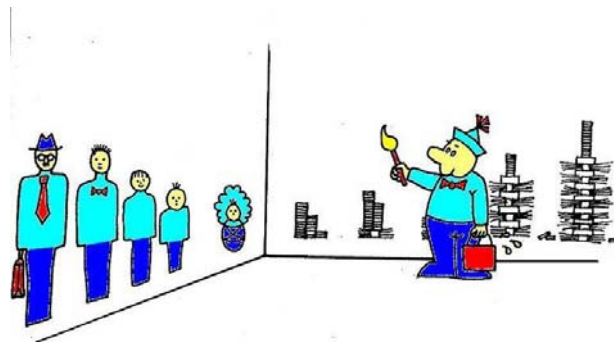
Combined tested model quality is: 98.5893605445324

Tools: Provider Wizard- Quality

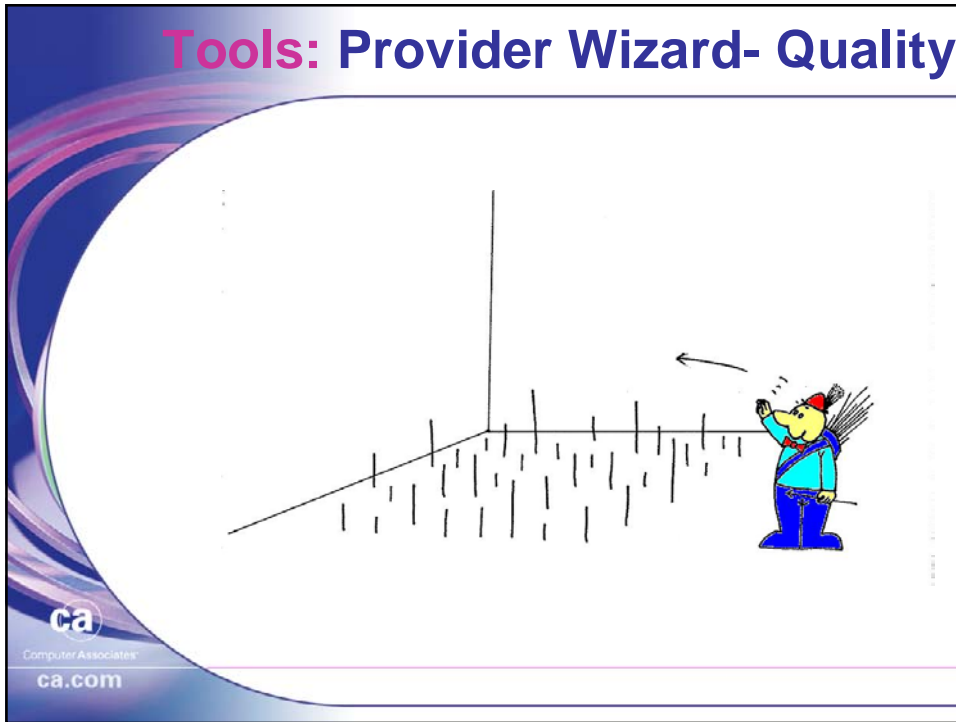
ANOVA RSQ

$$\left(1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}\right) * 100$$

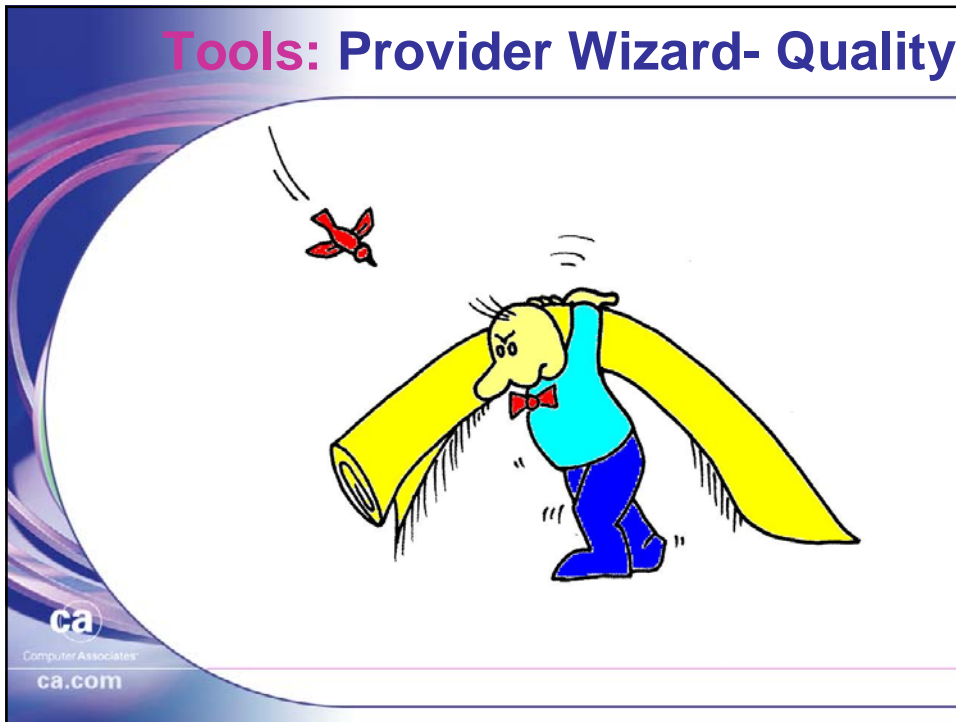
Tools: Provider Wizard- Quality



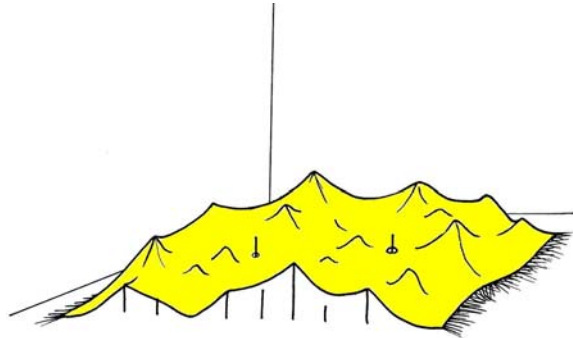
Tools: Provider Wizard- Quality




Tools: Provider Wizard- Quality

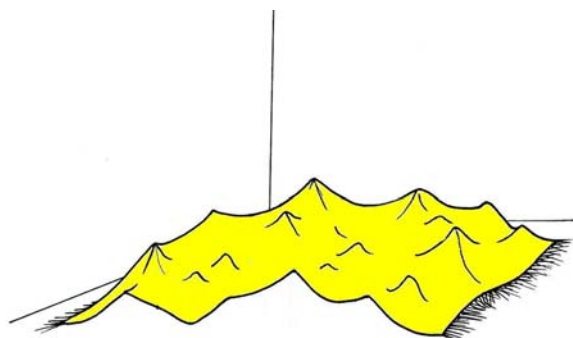



Tools: Provider Wizard- Quality



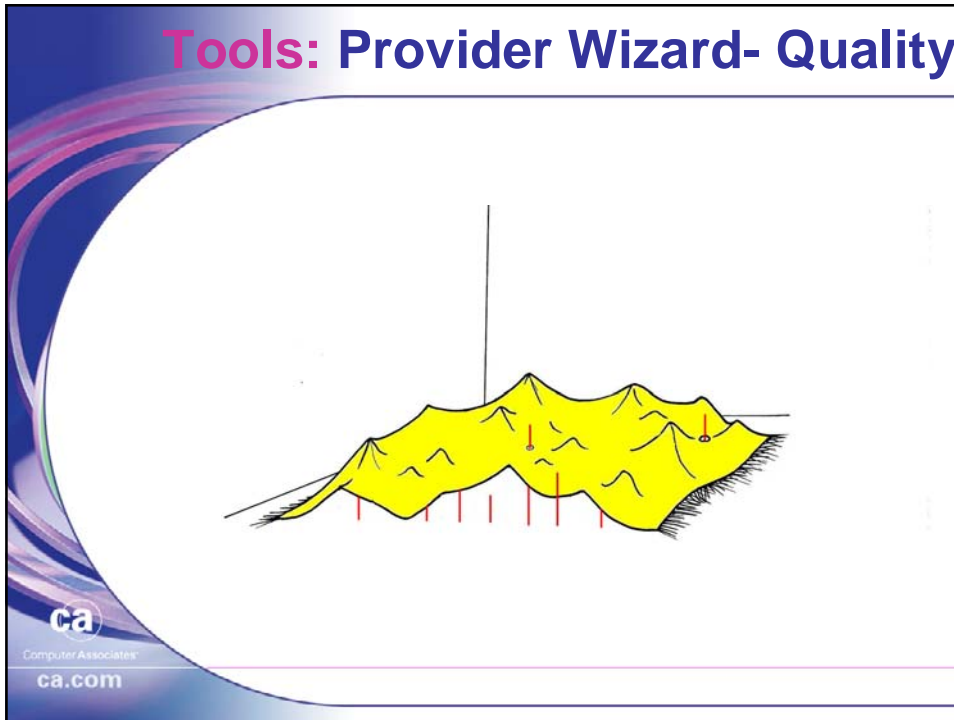

Computer Associates®
ca.com

Tools: Provider Wizard- Quality




Computer Associates®
ca.com

Tools: Provider Wizard- Quality



Tools: Provider Wizard- Decision

- Regel-Generierung
- Wird das Modell in Klartext übersetzt

Status: Input

of Bins:

Fit Type:

Bin Labels:

Bin Limits:

Sigma: 9.60821041525028
Mean: 45.5322419817012

Coverage: 65% (26/40)

```

p11 is 69% (550) 0.30% (241) 1
├─ if salary is low: p11 is 0.81% (30), 1.18% (7)
│ └─ if age is low: p11 is 0.55% (5), 1.44% (4)
│   └─ if rrsparmt is low: p11 is 0.55% (5), 1.44% (4)
│     └─ if age is medium: p11 is 0.90% (18), 1.10% (2)
│       └─ if rrsparmt is low: p11 is 0.90% (18), 1.10% (2)
├─ if age is high: p11 is 0.87% (7), 1.12% (1)
├─ if salary is medium: p11 is 0.72% (387), 1.27% (145)
│ └─ if age is low: p11 is 0.63% (61), 1.36% (35)
│   └─ if age is medium: p11 is 0.75% (270), 1.24% (87)
│     └─ if age is high: p11 is 0.70% (56), 1.29% (23)
└─ if salary is high: p11 is 0.59% (133), 1.40% (88)
                    
```

Prune Tree

Minimum Support:

Minimum Percentage:

Tools: Provider Wizard- Event

- Die Wahrscheinlichkeit von Ereignissen (z.B. Wartung, Regen, Systemausfall,...) werden ermittelt

Field Name List: day, evap, **prec**, rad, rh

Event Name List: rain

Timeframe: t

To include a field or event name in your rule, double click the name from the list box above. Add operators such as >, <, &, ||, and so on to create your rule.

Rule Name List:
if (prec(t) > 0.2) 'rain' = 1

Buttons: Edit Rule, Delete Rule, Load Rules, Save Rules

Consult Results:
 + R9905010000: It is unlikely to have event occur.
 + R9905020000: rain.
 + R9905030000: rain.
 + R9905040000: It is unlikely to have event occur.
 + R9905050000: rain.
 + R9905060000: It is unlikely to have event occur.
 + R9905070000: It is unlikely to have event occur.
 + R9905080000: rain.
 + R9905090000: It is unlikely to have event occur.
 + R9905100000: It is unlikely to have event occur.
 + R9905110000: It is unlikely to have event occur.
 + R9905120000: It is unlikely to have event occur.
 + R9905130000: It is unlikely to have event occur.
 + R9905140000: It is unlikely to have event occur.

Tools: Provider Wizard- Text

- Klassifizierung von Dokumenten nach Inhalt
- Z.B. automatische Weiterleitung von e-mails

Single Word (No Processing needed)

Formatted Text

Transfer to LowerCase

Delimiter: SPACE

Text Quote: EMPTY

Free Text

FreeText File Names: text

Word	Counter	Percentage
a	1	0.375940
access	4	1.503759
occur	2	0.751880
action	3	1.127820
activ	3	1.127820
adapt	1	0.375940
add	1	0.375940
addit	1	0.375940
advantage	2	0.751880
all	2	0.751880
an	3	1.127820
analyti	1	0.375940
analyt	1	0.375940

Word above threshold %: 0 Add

Word below threshold %: 0 Add

Ignore these words:

Buttons: Export, Import

Export Frequency List

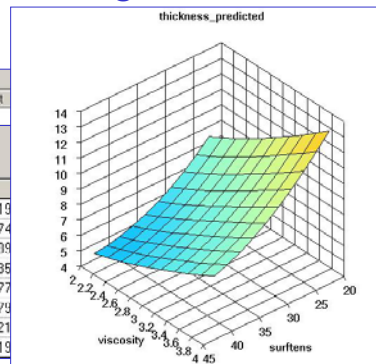
Tools: Optimization Wizard

- Sucht nach Optimumwerten (Value Neugents)
- Z.B. Entwicklung von Produkten, Produktionssteuerung

Field Name	Min	Max	Field Type
viscosity	2.0086	3.9798	Double
surfens	22	30	Integer
persolids	25	35	Integer
nipclear	29	39	Integer
ra	1	1	Integer
rp	4	4	Integer
rl	9	9	Integer

Outputs	Min	Mid	Max	Weight
thickness	2.977300	8.826300	14.675300	

Optimization Result	
Best output is: 0	
Field Name	Value
viscosity	3.10019
surfens	29.1874
persolids	44.1709
nipclear	29.835
ra	2.06677
rp	4.80979
rl	10.8621
thickness	0.95119



Tools: Deployment Wizard

- Generierung von Klassen
- Erstellen von kompletten Applikationen

Choose the platform to deploy:

- Aion 9.5
- ASP (Active Server Pages)
- C++
- COM
- Java
- Web Services

WSB Host Name: localhost

WSB Port Number: 7078

Enter a name for the service: VFMyNeugent

User Name (optional):

Password (optional):

Application Type:

Java (Native)

JSP (Java)

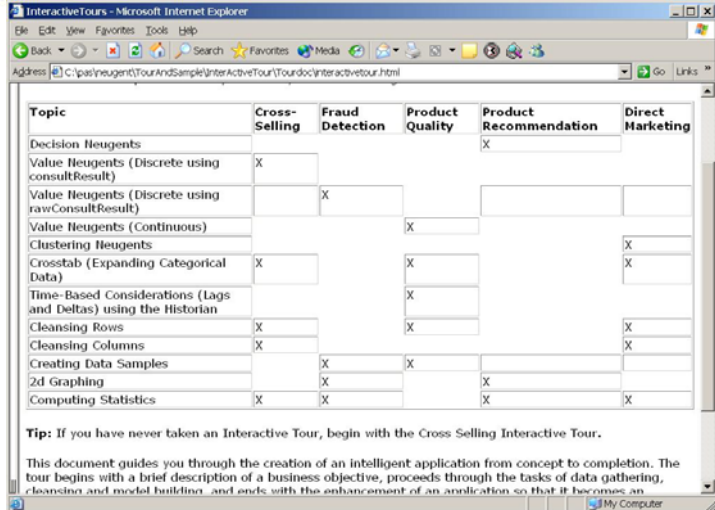
The following files have been created:

```

Broker/webapps/VFMyNeugent/WEB-INF/web.xml
Broker/webapps/VFMyNeugent/WEB-INF/classes/CreateService.bat
Broker/webapps/VFMyNeugent/VFMyNeugentClient.jsp
Broker/webapps/VFMyNeugent/JSP_Build_Error.html

```

Tools: Interaktive Tutorials



Topic	Cross-Selling	Fraud Detection	Product Quality	Product Recommendation	Direct Marketing
Decision Neugents				X	
Value Neugents (Discrete using consultResult)	X				
Value Neugents (Discrete using rawConsultResult)		X			
Value Neugents (Continuous)			X		
Clustering Neugents					X
Crosstab (Expanding Categorical Data)	X		X		X
Time-Based Considerations (Lags and Deltas) using the Historian			X		
Cleansing Rows	X		X		X
Cleansing Columns	X				X
Creating Data Samples		X	X		
2d Graphing		X		X	
Computing Statistics	X	X		X	X

Tip: If you have never taken an Interactive Tour, begin with the Cross Selling Interactive Tour.

This document guides you through the creation of an intelligent application from concept to completion. The tour begins with a brief description of a business objective, proceeds through the tasks of data gathering, cleansing and model building, and ends with the enhancement of an application so that it becomes an

CA World – Termin, Lokation

- **13. – 17. Juli 2003**
- **Mandalay Bay Convention Center**
- **Las Vegas, Nevada**
- **<http://ca.com/germany/caworld/>**

CA World - Schwerpunkte

- Informative Sessions zu neusten CA-Technologien
- Praktische Demos im Exhibition Center zu allen CA-Softwarelösungen
- Kostenlose CA-Schulungskurse
- Persönliche Gesprächstermine mit CA-Entwicklern und Support-Mitarbeitern im Technical Campground
- Austausch mit Fachkollegen und Branchenexperten

CleverPath™ Predictive Analysis Server

Laszlo Mihalka
Senior Consultant
Laszlo.Mihalka@ca.com