

**Synop Analyzer** ist eine modular aufgebaute Software für **interaktive Datenanalyse (Drill down), Daten-Visualisierung, Data Mining, Datenqualitätsüberwachung, Vorhersage** und **Mustererkennung**. Es umfasst also Funktionalitätsbereiche, die traditionell nur von mehreren unterschiedlichen Softwarepaketen aus den Bereichen Data Warehousing, Business Intelligence, Reporting, Ad-Hoc-Analyse, Data Mining und Visualisierung abgedeckt werden.

Im Vergleich zu etablierten Reporting-, BI-, OLAP-, Data-Mining- oder Datenvisualisierungs-Softwarepaketen bietet Synop Analyzer die folgenden Differenzierungsmerkmale:

- **Leichte Bedienbarkeit:** Synop Analyzer erfordert weder eine komplexe Datenvorverarbeitung oder Datenmodellierung noch den Aufbau eines ‚Cube Models‘ oder OLAP-Systems.
- **Geschwindigkeit, Skalierbarkeit, Interaktivität:** typischerweise liefert die Arbeit mit Synop Analyzer schon Minuten nach dem Beginn der Datenexploration interessante Resultate. Die Software nimmt automatisch die für jede Art der Datenexploration sinnvollen Datenvorverarbeitungsschritte wie Attributauswahl oder Klassenbildungen vor, und jeder einzelne ‚Drill-Down‘-Schritt liefert dank **In-Memory-Technologie** und parallelisierter Algorithmen auch auf den größten verfügbaren Daten sekundenschnell ein Ergebnis.
- **Einzigartige Kombination von interaktiven Explorations- und Data-Mining-Komponenten:** der Synop Analyzer verbindet die Vorteile von spezialisierten Business-Intelligence-Datenbanksystemen mit den flexible Drill-Down-Fähigkeiten einer OLAP-Anwendung und den hoch leistungsfähigen Algorithmen eines Data-Mining-Expertenwerkzeugs.
- **Integration in bestehende IT-Strukturen und Prozesse:** Synop Analyzer interagiert mit Datenbanken, BI-Systemen und Unternehmensapplikationen über standardisierte Schnittstellen wie JDBC, Web Services (SOA) und XML. Sie können Prozesse mit Echtzeit-Algorithmen optimieren oder regelmäßig Datensätze automatisiert im Batch-Job oder per Systemservice ausführen lassen.
- **Günstige ‚Total Cost of Ownership‘:** geringe Hardware-, Administrations- und Schulungskosten und das modulare Preismodell machen die Software für Unternehmen jeder Größe interessant.

## Modul-Beschreibungen

- **In-Memory-Datenbank**

Die eigenentwickelte In-Memory-Datenbank ist das Herzstück, das es dem Synop Analyzer ermöglicht sehr große Datenmengen in Echtzeit zu visualisieren und in Sekunden oder Minuten zu analysieren. Die Daten werden durch einen Ladeprozess in das spaltenbasierte Format gewandelt und komprimiert. Mittels API- oder Batch-Prozessen kann die In-Memory-Datenbank pro Analyseobjekt automatisch erzeugt und geladen werden. Aktuell können bis zu 128 Mrd. Datenzeilen verwaltet werden. Mit Multi-Threading, Multi-Core (bis 64 Cores) und parallelisierter Ausführung können auch rechenintensive Algorithmen in Minuten statt in Stunden oder Tagen ausgeführt werden. Data-Mining und Mustererkennung auf langen zeitlich zusammenhängenden Daten wird erstmals wirtschaftlich möglich.

- **Datenvorbereitung, ETL**

Einlesen von Daten aus beliebigen relationalen Datenbanksystemen, Excel-Spreadsheets, Textdateien oder Json- und XML-Dateien. Zusammenführen mehrerer Tabellen, Berechnung neuer Felder, Definition von Klarnamen und Hierarchien, Filtern nach Zeilen und/oder Spalten, Anonymisieren, Aggregieren (Pivot-Tabellen). Abspeichern und automatisches Starten von Datenvorbereitungsprozessen.

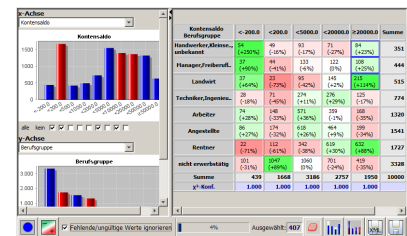
**• Werteverteilungen und Statistiken über Datenattribute**

In einer übersichtlichen Darstellung werden die Eigenschaften der Datenfelder in tabellarischer und grafischer Form angezeigt, wobei die Software automatisch eine geeignete Skalierung und Klasseneinteilung vornimmt. Datenfehler und Datenlücken werden klar sichtbar.



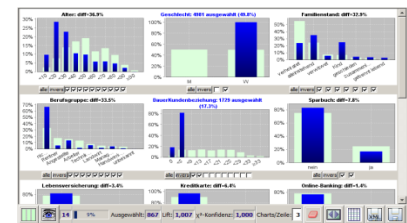
**• Korrelationsanalysen und Bivariate Analysen**

Berechnung von Korrelationen zwischen Datenfeldern; Detail-Darstellung der Zusammenhänge zwischen den Werten zweier Felder mit Hilfe einer Bivariaten Matrix; grafische Darstellung signifikant über- und unterrepräsentierter Wertekombinationen.



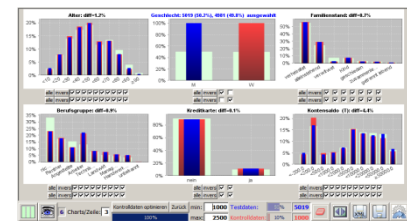
**• Multivariate Analysen (Ad-Hoc-Drill-Down)**

Die vorher-arbeiteten Daten werden in hoch optimierter Form im Arbeitsspeicher vorgehalten und ermöglichen einen flexiblen Echtzeit-Drilldown mit Wertebereichs-Eingrenzungen auf beliebigen Kombinationen von Datenfeldern. Anwendungsbereiche: Zielgruppenauswahl im CRM, Controlling, Qualitätssicherung und viele mehr.



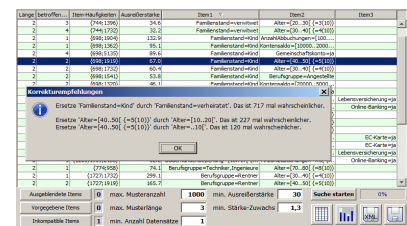
**• Test-/Kontrollgruppenanalysen und Hypothesentests**

Der statistische Vergleich zweier Daten-Teilmengen, einer Test- und einer Kontrollgruppe, eignet sich zum quantitativen Test von Hypothesen und Vermutungen sowie zur ROI-Berechnung von Vertriebs-, Kostensenkungs- oder Qualitätssteigerungs-Kampagnen.



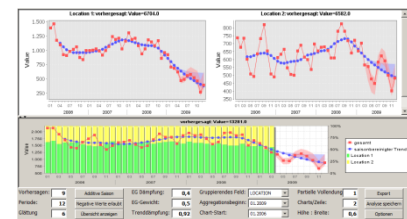
**• Ausreißer, Datenfehlern und Inkonsistenzen**

Traditionelle Methoden der Datenqualitätssicherung erfordern die vorherige Definition von Filter-Regeln und übersehen oft wichtige Inkonsistenzmuster. Synop Analyzer bietet dagegen eine hypothesenfreie Methode zum Auffinden von Inkonsistenzen und mutmaßlichen Datenfehlern.



**• Analyse und Prognose von Zeitreihen**

Zeitreihen über Umsätze, Kosten, Verkaufszahlen und vieles mehr lassen sich auf Trends und zyklische Muster hin analysieren. Die Software kann Prognosen für die zukünftige Entwicklung unter verschiedenen Annahmen bezügl. Trenderhaltung und Fortwirken vergangener Ereignisse berechnen.



**• Assoziations-Analysen und Vorhersage (Scoring)**

Mithilfe leistungsfähiger Mustererkennungs-Algorithmen lassen sich sowohl häufig oder auch sehr selten vorkommende Muster in Datensätzen erkennen. Hiermit lassen sich Warenkorbanalysen durchführen, typische Verkaufs- oder Verhaltensmuster von Kunden erkennen und verborgene Fehlerursachen für Produktionsprobleme oder gehäufte Garantiefälle aufdecken.

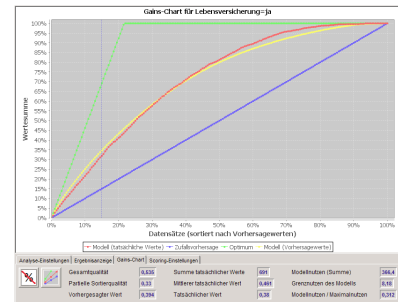


• **Sequenzmuster-Analysen und Vorhersage (Scoring)**

Mit der Sequenzmusteranalyse können Erfolg entscheidende Erkenntnisse aus Daten mit zeitlichen Informationen gewonnen werden. Typische Anwendungen sind z.B. die Analyse von Kauf- und Verhaltensmuster zur Optimierung (Conversion Rate) der Kundeninteraktionen, Aufdeckung von Fehlermustern in der Produktion, Frühwarnsysteme bei Maschinendaten oder auch Recommendation Engines für Webshops und Beratungsapplikationen am POS.

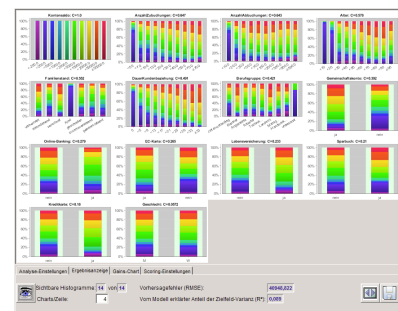
• **Regressions-Training und Vorhersage (Scoring)**

Die lineare Regression wird für die Vorhersage von Werten aus nummersichen Daten und die logistische Regression für die Vorhersage von zweiwertigen Informationen eingesetzt. So können z.B. Ausfallraten bei Krediten oder Teilnehmer für Kamapgnen gescort werden. Das Modul kann ohne spezielle Datenvorauswahl eine Modell trainieren, Konflikte durch hohe Korrelationen oder textuelle Informationen können bearbeitet die Software automatisch.



• **Naive-Bayes-Training und Vorhersage (Scoring)**

Ein Naive-Bayes-Klassifikationsmodell ist ein Data-Mining-Modell, das den (unbekannten) Wert eines bestimmten Datenmerkmals, des sogenannten Zielfeldes, aus den Werten mehrerer anderer Datenfelder, der sogenannten Prädiktorfelder, vorhersagt. Die Vorhersage-Formel wird während eines sogenannten Trainings auf Trainingsdaten ermittelt, bei denen sowohl das Zielfeld als auch die Prädiktorfelder mit Werten gefüllt sind. Das Klassifikationsmodell kann später auf neue Daten angewendet werden, in denen die Zielfeldwerte fehlen.



Auf diese Weise kann man Zielfeldwerte vorhersagen. Die spaltenorientierte In-Memory-Datenhaltung wie die interne Datenhaltung von Synop Analyzer ermöglicht es, dass Naive-Bayes-Modelle selbst auf extrem großen Daten äußerst schnell trainiert und anwendbar sind - meist sogar annähernd in Echtzeit.

• **Segmentierung (Clustering) und Vorhersage (Scoring)**

Verschiedene neuronale und nicht-neuronale Data-Mining-Verfahren ermöglichen die Segmentierung von Daten und das Berechnen von Vorhersagen und 'Score'-Werten. Es gibt zwei große praktische Anwendungsgebiete für SOMs: Datenvisualisierung und Datensegmentierung auf der einen Seite und 'Scoring', das heißt das Vorhersagen unbekannter Attributwerte auf der anderen Seite. In letzterem Fall wird die trainierte SOM auf eine neue Datenquelle angewendet, bei der einige der Attribute oder der Attributwerte, die in den ursprünglichen Trainingsdaten vorhanden waren, fehlen, z.B. Zahlungsausfallrisiken, zu erwartende Deckungsbeiträge, Kündigungsrisiken oder zukünftiges Kundenpotenzial.

