Zawansowana Eksploracja Danych: Przegląd systemów ich rola we wspomaganiu decyzji, podsumowanie

Jerzy Stefanowski

Wykład TPD

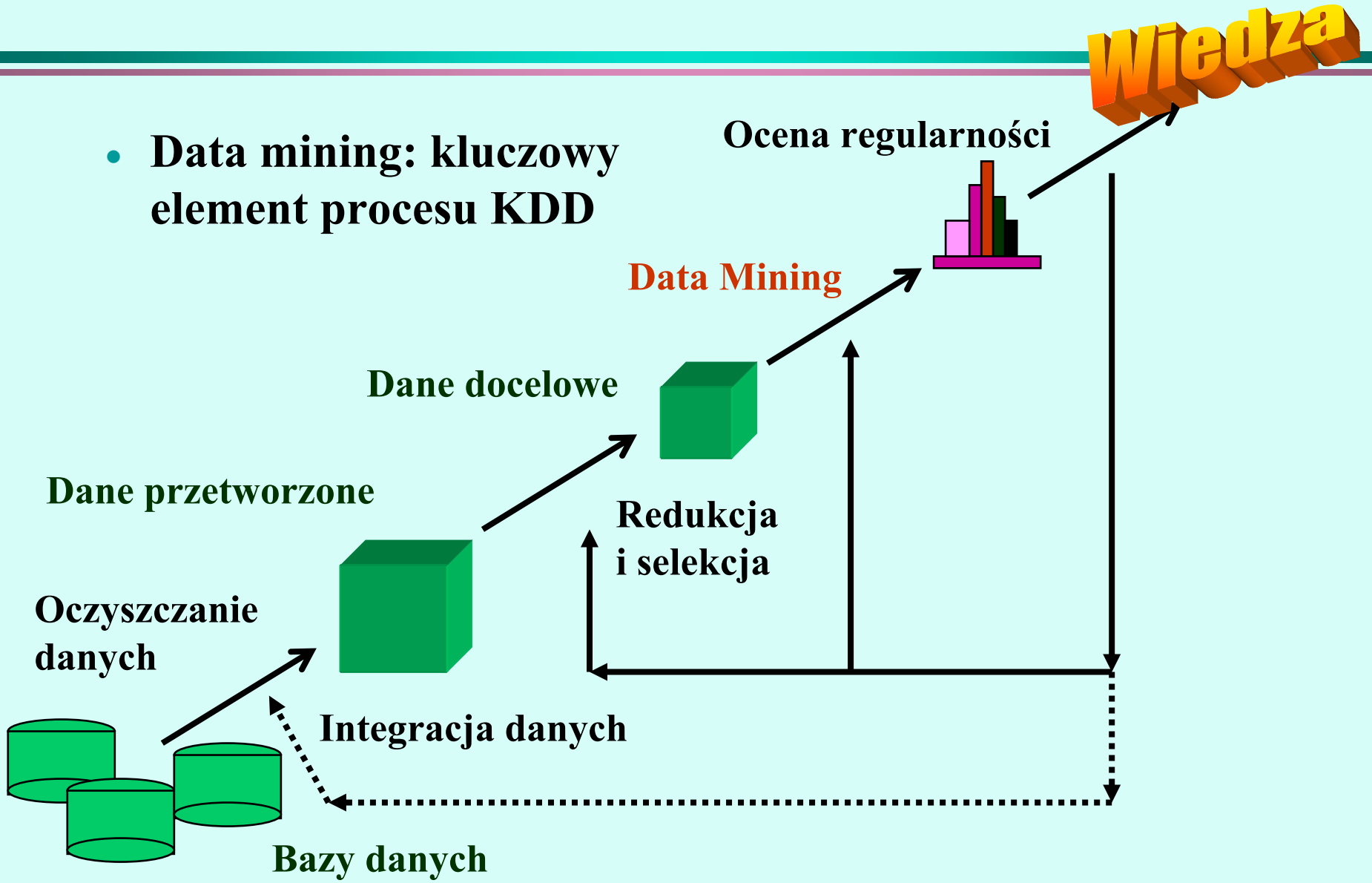
Poznan 2008/2009 – uzupełnienie 2010

Plan prezentacji

- **Proces Odkrywania Wiedzy w Bazach Danych**
- **Rola użytkownika**
- **Podstawowe metody eksploracji danych**
- **Istniejące systemy KDD – stan aktualny**
- **Zastosowania w przedsiębiorstwach**
- **Perspektywy rozwoju**

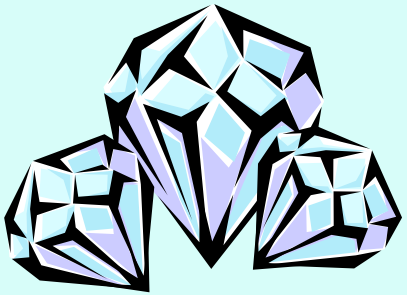
Proces Odkrywania Wiedzy - KDD

- **Data mining: kluczowy element procesu KDD**



Etapy procesu odkrywania wiedzy

- **Analiza i poznanie dziedziny zastosowania, identyfikacja dostępnej wiedzy i celów użytkownika,**
- **Wybór danych związanych z celami procesu,**
- **Czyszczenie i wstępne przetwarzanie danych oraz ich redukcja,**
- **Wybór zadań i algorytmów eksploracji danych,**
- **Pozyskiwanie wiedzy z danych (krok eksploracji danych),**
- **Interpretacja i ocena odkrytej wiedzy,**
- **Przygotowanie wiedzy do użycia.**



Interaktywność Procesu KDD

- **Znacząca rola użytkownika w odkrywaniu wiedzy z danych**
- **Podejmuje decyzje w zakresie np.**
  - **wyboru podzbioru danych, reprezentacji wiedzy, algorytmów eksploracji**
- **Rola użytkownika w interpretacji i ocenie wiedzy**
- **Proces odkrywania wiedzy wymaga intensywnego współdziałania człowieka z systemem**
- **Proces ten jest z definicji interaktywny i iteracyjny**

Typowe zadania

- **Podsumowywanie danych**
  - **Statystyki opisowe, charakteryzowanie danych, porównywanie**
- **Klasyfikowanie**
- **Regresja i predykcja**
- **Asocjacje i powiązania**
  - **znajdź reguły typu:**
    80% klientów którzy kupują towary A i B kupuje także towar C
    (75%wsparcie w bazie danych)
- **Grupowanie (analiza skupień) i poszukiwanie taksonomii**
  - **Tworzenie klas podobnych obserwacji**

Typowe zadania (cd.)

- **Modelowanie zależności funkcyjnych i praw lub równań**
- **Analiza przebiegów czasowych:**
  - **trendy, poszukiwanie prognoz, poszukiwanie anomalii,**
- **Analiza następstw zdarzeń**
- **Nowe kierunki w zakresie eksploracji danych:**
  - **Text mining, Web mining, visual and multimedia mining, analiza zaawansowanych i rozproszonych baz danych (hurtowanie danych, OLAP), systemy rozproszone, ...**

Systemy dla eksploracji danych

- **IBM: QUEST and Intelligent Miner**
- **Oracle Miner**
- **SAS Institute: Enterprise Miner**
- **Statsoft: Statistica Data Miner**
- **Integral Solutions Ltd.: Clementine /SPSS**
- **Silicon Graphics: MineSet**
- **SFU: DBMiner, GeoMiner, MultiMediaMiner**
- **Inne systemy**
  - **Rutger Univ.: DataMine**
  - **GMD: Explora**
  - **Univ. Munich: VisDB**

Systemy dla eksploracji danych

- **Zorientowane na bazy danych**
  - **IBM:Intelligent Miner**
  - **DBMiner (OLAP i magazyny danych)**
  - **Oracle i9 - i11 Miner**
  - **Silicon Graphics: MineSet (wizualizacja danych)**
- **Statystyczne**
  - **SAS Institute: Enterprise Miner (dobra integracja danych)**
  - **Także - SPSS, Statistica**
- **Uczenie Maszynowe**
  - **WEKA, YALE, INLEN, 49ner**

Dla prostych zadań można także używać bardziej typowych narzędzi

Statistica – Statsoft (www.statsoft.pl)

- **Stworzony jako „przyjazne dla użytkownika” oprogramowanie podstawowych metod statystycznej analizy danych. Środowisko systemu operacyjnego Ms Windows.**
- **Bardzo liczna biblioteka prostych i zaawnsowanych metod analizy danych.**
- **Szybkość wykonywania obliczeń; elastyczne zarządzanie wynikami.**
- **„Łatwość” obsługi; bardzo dobry „help”; proste skróty i dostęp do narzędzi.**
- **Wygodny intuicyjny interfejs graficzny. Wysoka jakość wykresów prezentacyjnych i analitycznych.**
- **Profesjonalny system raportów.**
- **W pełni zintegrowany z Visual Basic (możliwość budowania własnych modułów).**
- **Umożliwia także dostęp do różnego rodzaju danych (także baz danych).**

Inne powiązane systemy:

- **Oddzielna aplikacja zawierająca implementacje wielu sztucznych sieci neuronowych – Statistica Neural Networks.**
- **Systemy korporacyjne, rozwiązania dla przemysłu oraz …**

Rodzina systemów Statsoft

- **Pakiety statystyczne Statistica**
- **Six-Sigma – rozwiązania dla kontroli i sterowania procesów przemysłowych:**
  - **Karty Kontrolne**
  - **Analiza Procesu**
  - **Planowanie Doświadczeń**
- **Systemy Korporacyjne**

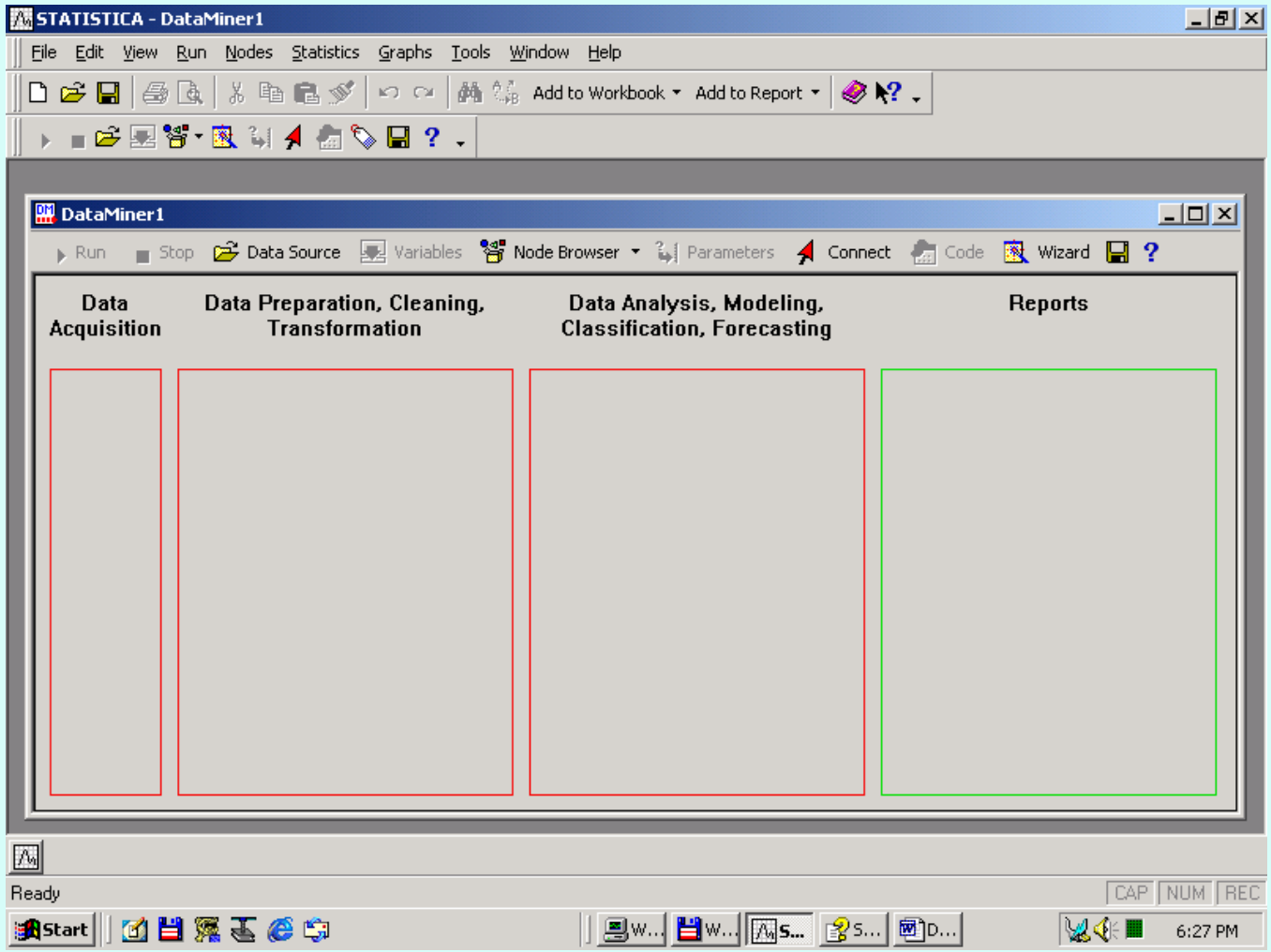
(dodatkowo zawierają narzędzia pracy zespołowej, dostęp z przeglądarek internetowych, serwer usług sieciowych Statistica, narzędzia dostępu do hurtowni i baz danych, interaktywne tworzenie zapytań do baz danych, integracja z hurtowniami danych, eksploracja przekrojów i kostek OLAP), np.

- **Enterprise-wide Data Miner**
- **SPC System (Six-Sigma + kontrola jakości)**
- **… inne**

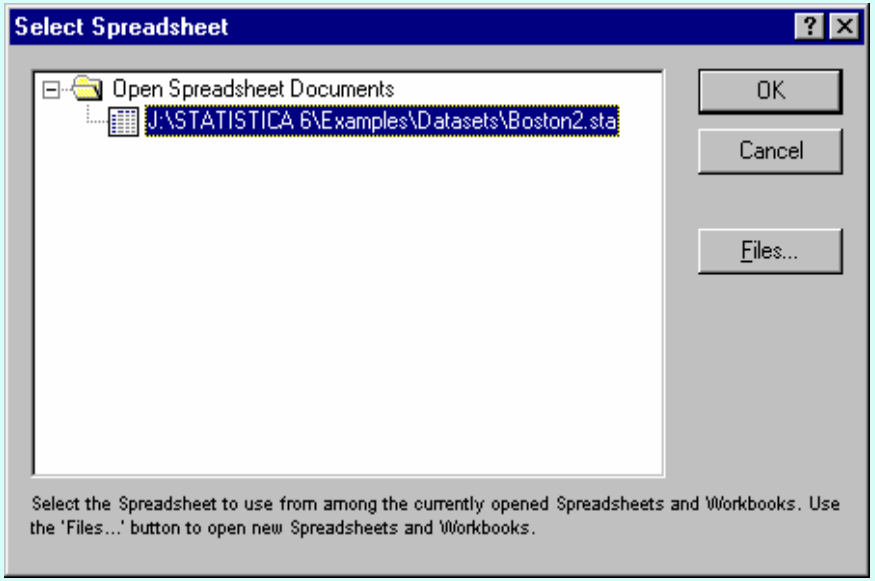
Statistica Data Miner

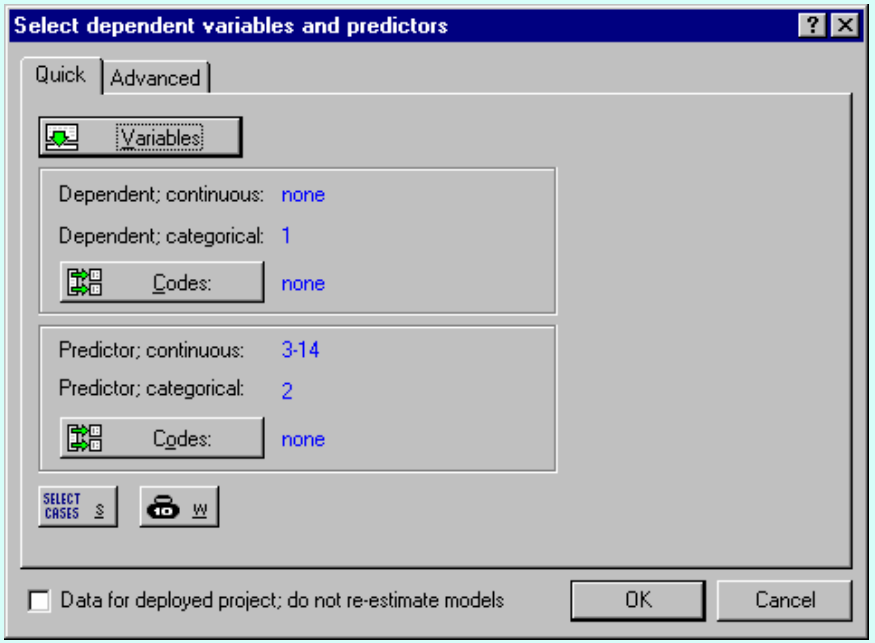
- **Przygotowanie nowych programów, w tym Statistica DataMiner, specjalizowanej aplikacji do realizacji procesu odkrywania wiedzy / eksploracji danych.**
- **Udostępnia w formie zintegrowanej wiele metod zarówno statystycznych jak i innych metod eksploracji danych.**
- **Możliwość dostępu do dużych repozytoriów danych bez tworzenia ich lokalnych kopii.**
- **Szablony tzw. projektów data mining – proste w użyciu.**
  - **Interfejs oparty na ikonach i technice typu „klikaj i przeciągaj”.**
  - **Proste zarządzanie danymi.**
  - **Łatwość aktualizacji wyników przy modyfikacji danych.**
  - **Możliwość stosowania różnorodnych metod w jednym projekcie.**
  - **Zapisanie projektu do kodu Visual Basic**
- **System o otwartej architekturze; możliwość rozbudowy o własne algorytmy.**
- **W wersji korporacyjnej dostęp za pośrednictwem Internetu (WebSTATISTICA Server).**

Arkusz projektu DataMiner

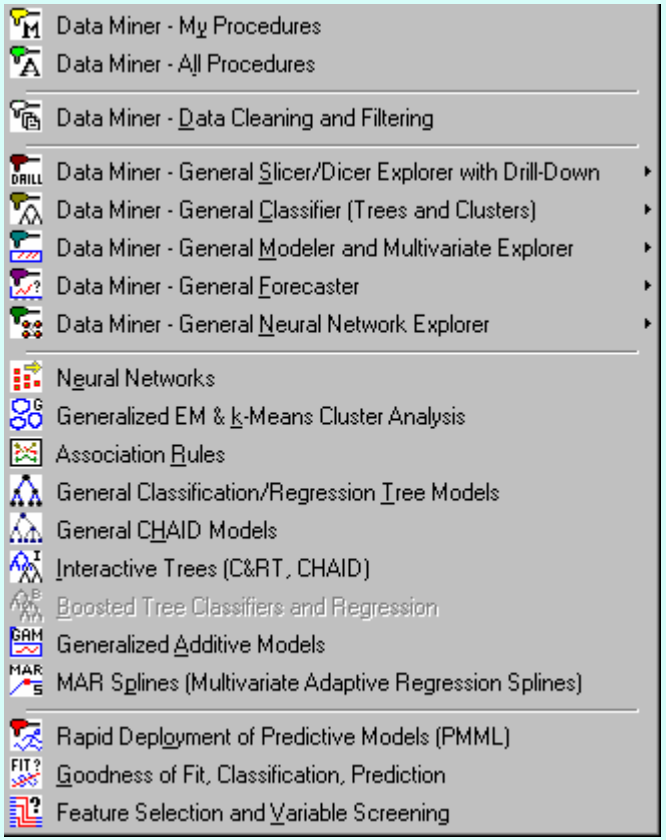


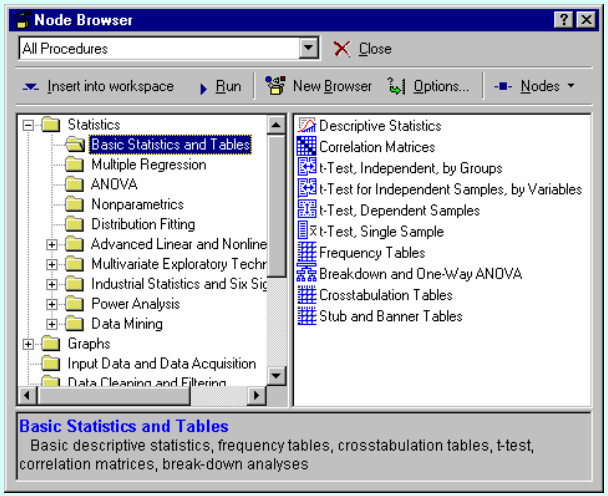
Data Miner – dostęp do danych, wybór zmiennych.

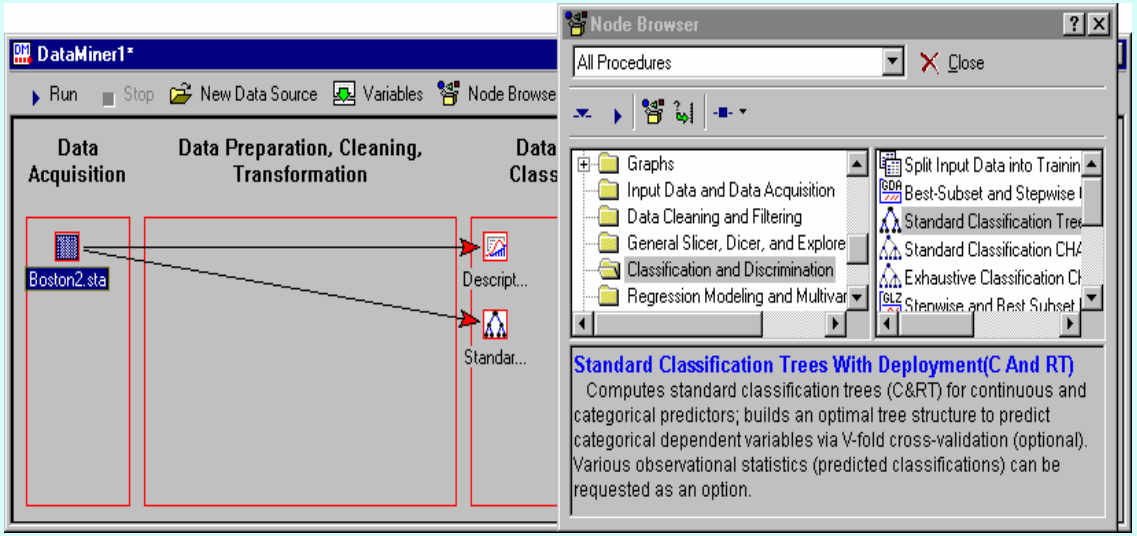




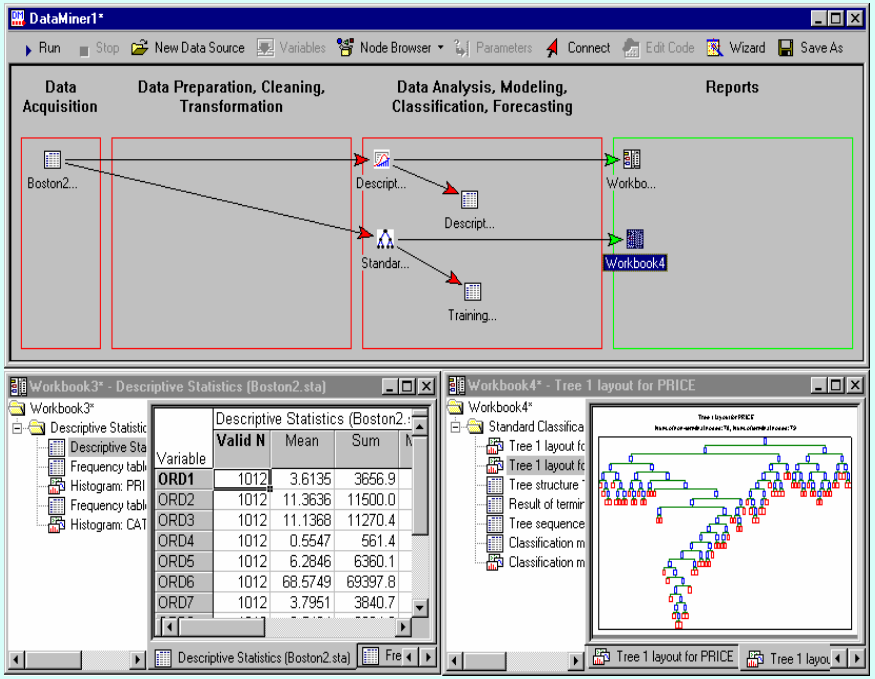
Data Miner – wybór metod

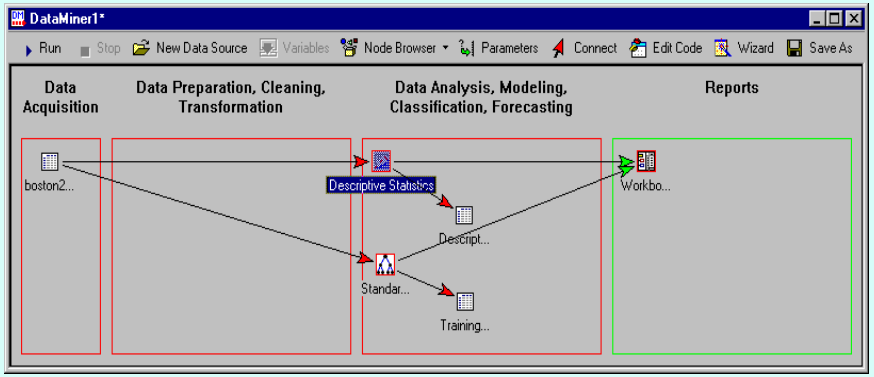


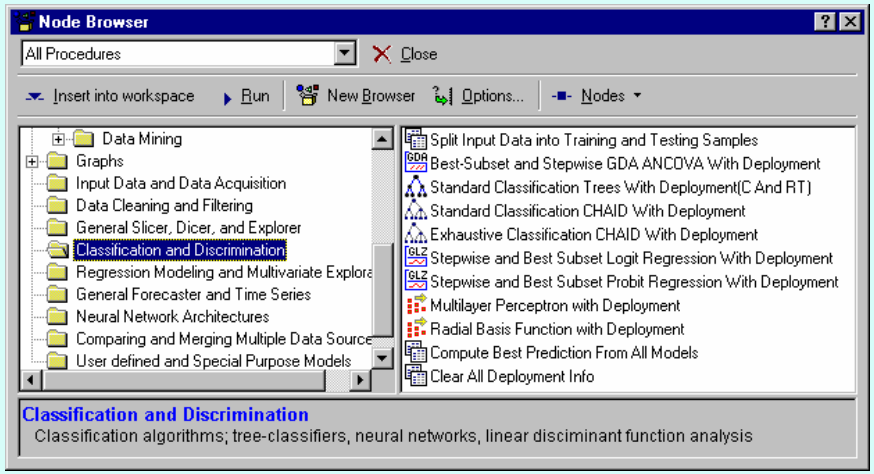




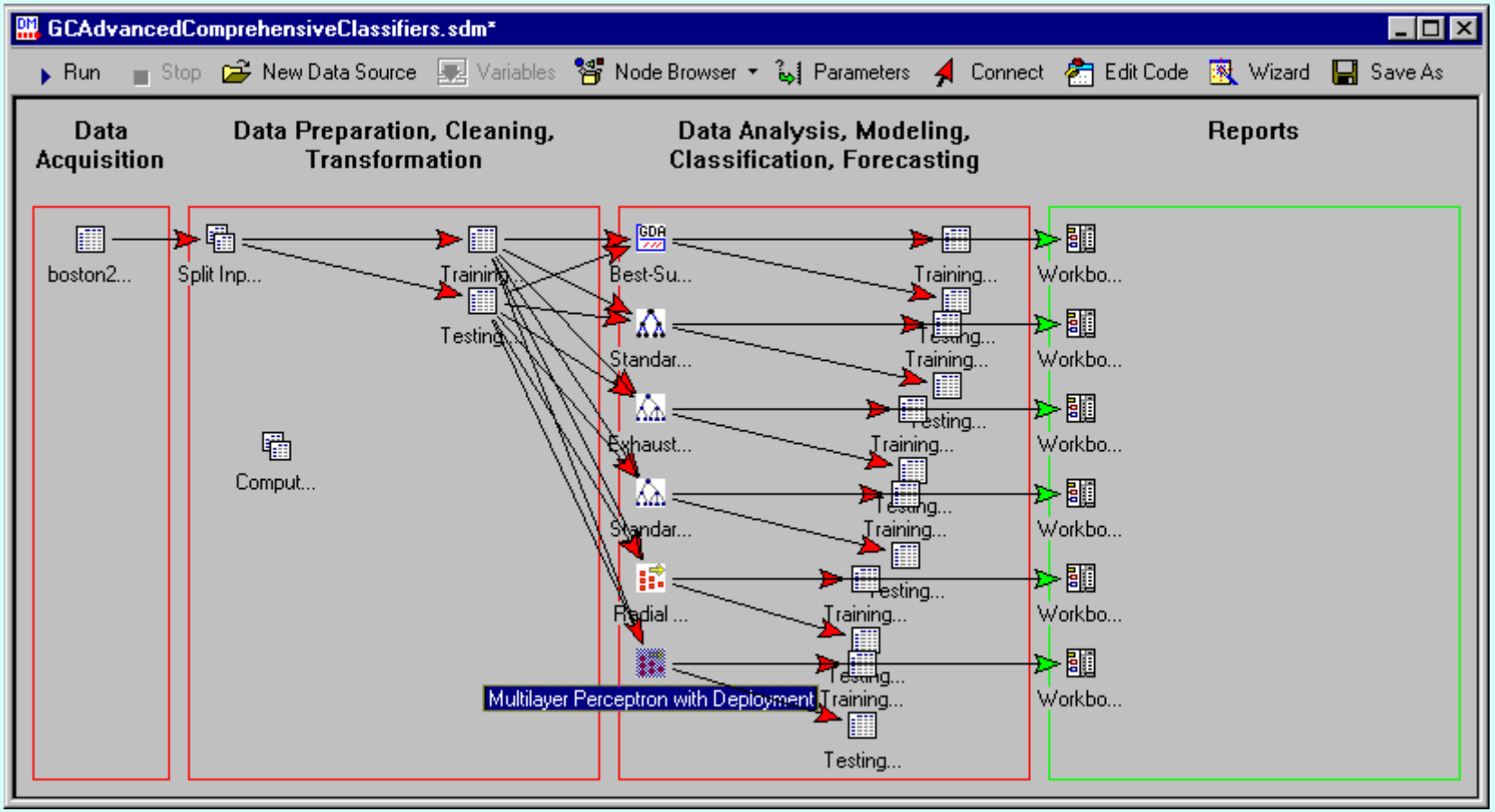
Definiowanie projektu data mining







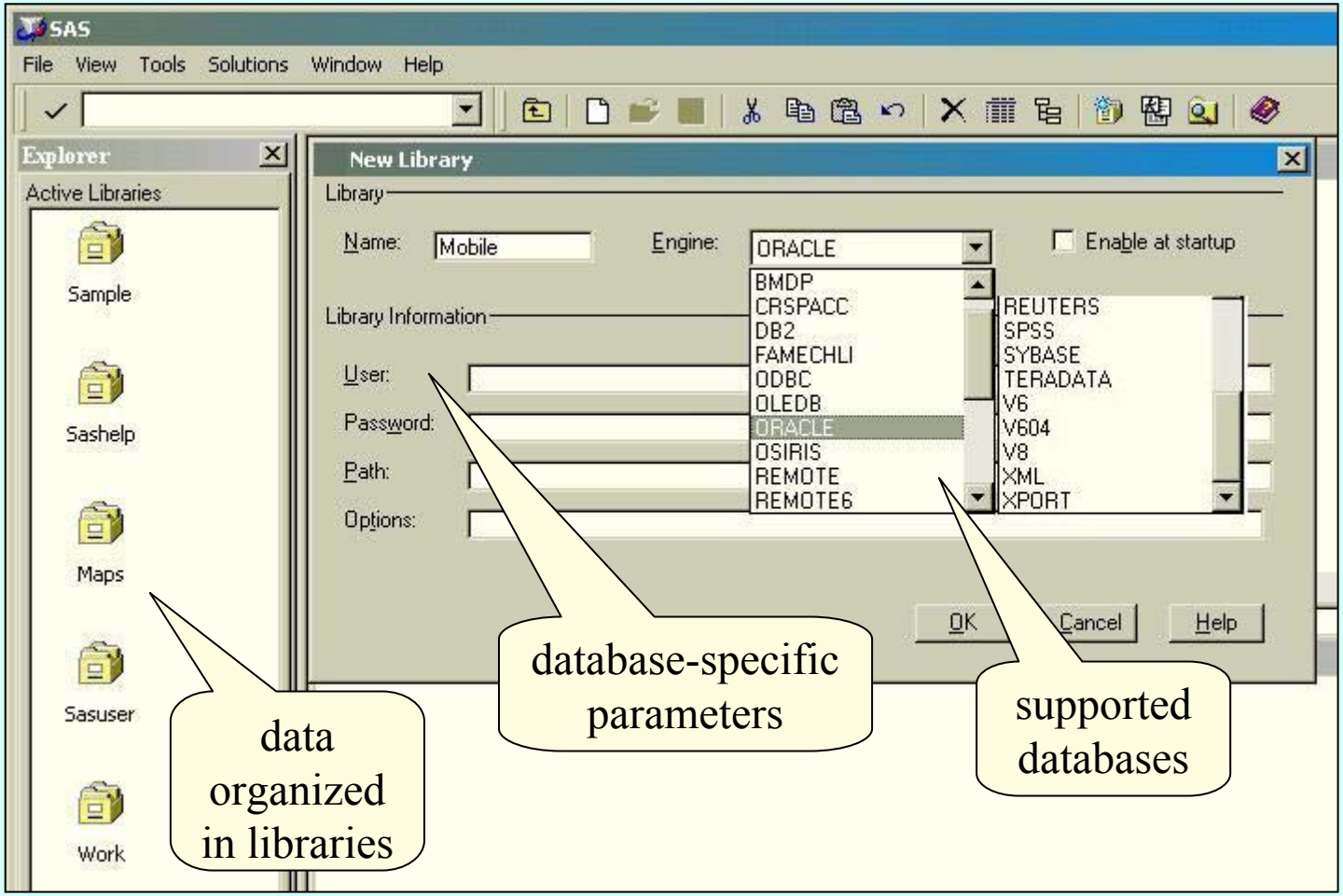
Sprawdzenie użycia wielu metod



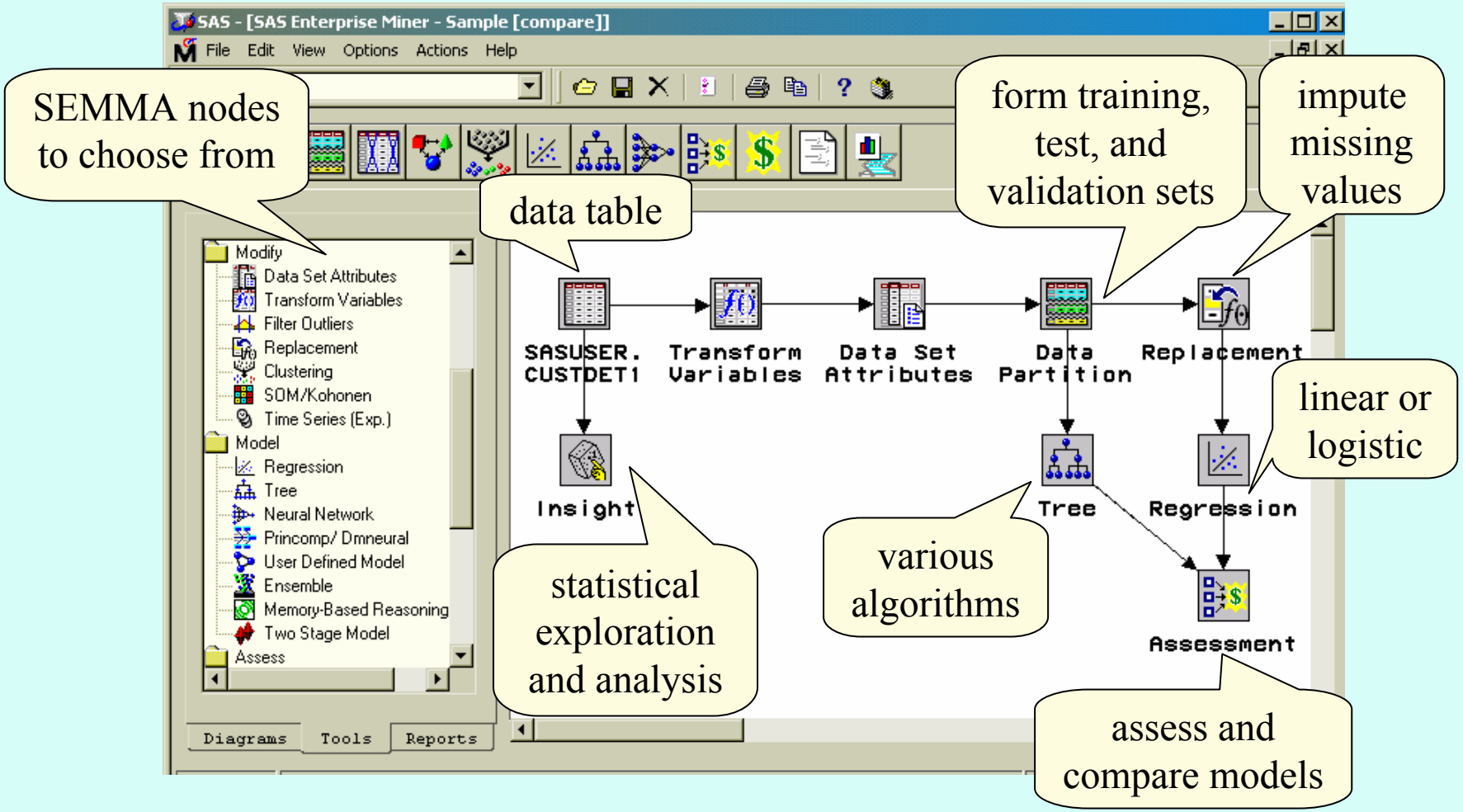
SAS

- **Historycznie system statystycznej analizy danych**
- **Rozszerzony o bardzo zaawansowane narzędzia dostępu do różnego rodzaju danych i ich integracji**
  - **przykłady: ADABAS, OpenIngres, DB2, Informix, Microsoft SQL server, ORACLE, SYBASE, Teradata, ODBC, OLE DB, różne formaty PC.**
- **Oferuje przetwarzanie danych za pomocą specjalnego języka oraz interfejsy graficznego**
- **Udostępnia w formie zintegrowanej wiele metod zarówno statystycznych jak i innych metod eksploracji danych**

Dostęp do danych w systemie SAS



Enterprise miner SAS



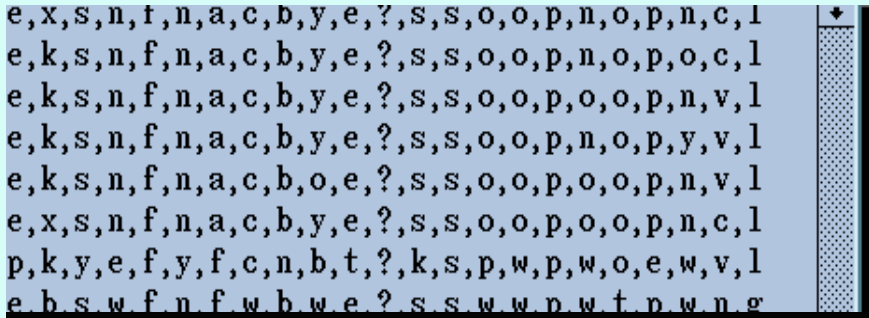
SAS

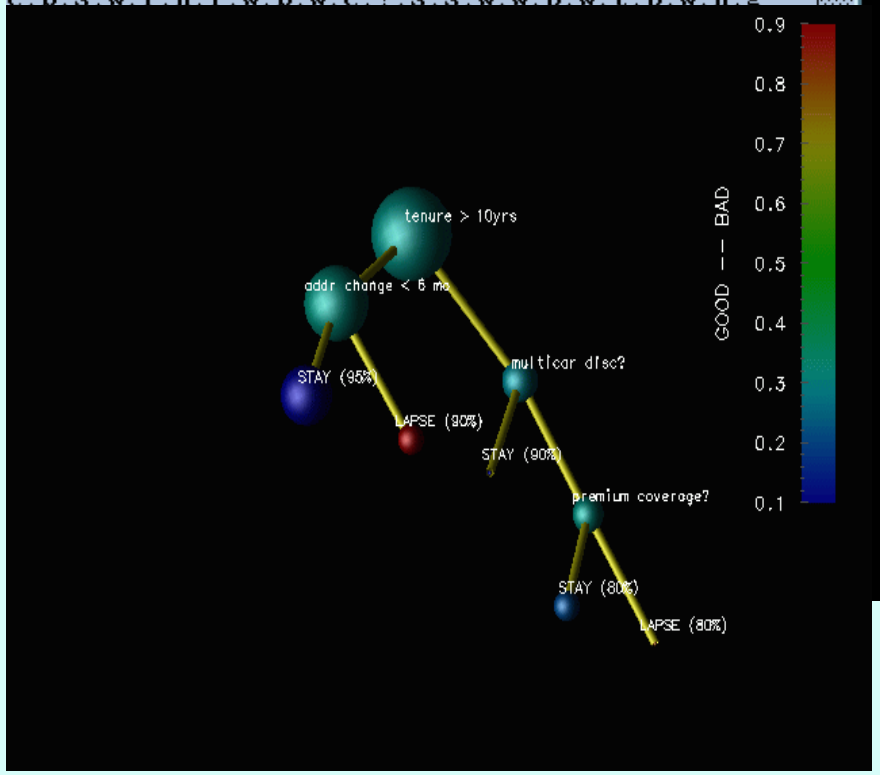
- Przykładowe algorytmy eksploracji danych (dostępne w tzw. węzłach SAS Enterprise Miner)
  - wiele metod statystki opisowej,
  - metody przekształceń danych (przeskalowania, uwzględnianie nieznanych wartości, wykrywanie nietypowych obserwacji),
  - poszukiwanie reguł asocjacyjnych,
  - analiza skupień (k-średnich, sieci SOM Kohonen'a)
  - modele predykcyjne (liniowa, nieliniowa, logistyczna regresja, drzewa regresji)
  - drzewa klasyfikacyjne (CART, CHAID, C4.5like)
  - sztuczne sieci neuronowe (liniowe/nieliniowe sieci wielowarstowe, różne wersje RBF).
  - modele złożonych klasyfikatorów (bagging, boosting, combiners,...)
  - modele k-NN
  - modele szeregów czasowych.
- Oferuje przetwarzanie danych za pomocą specjalnego języka oraz interfejsy graficznego

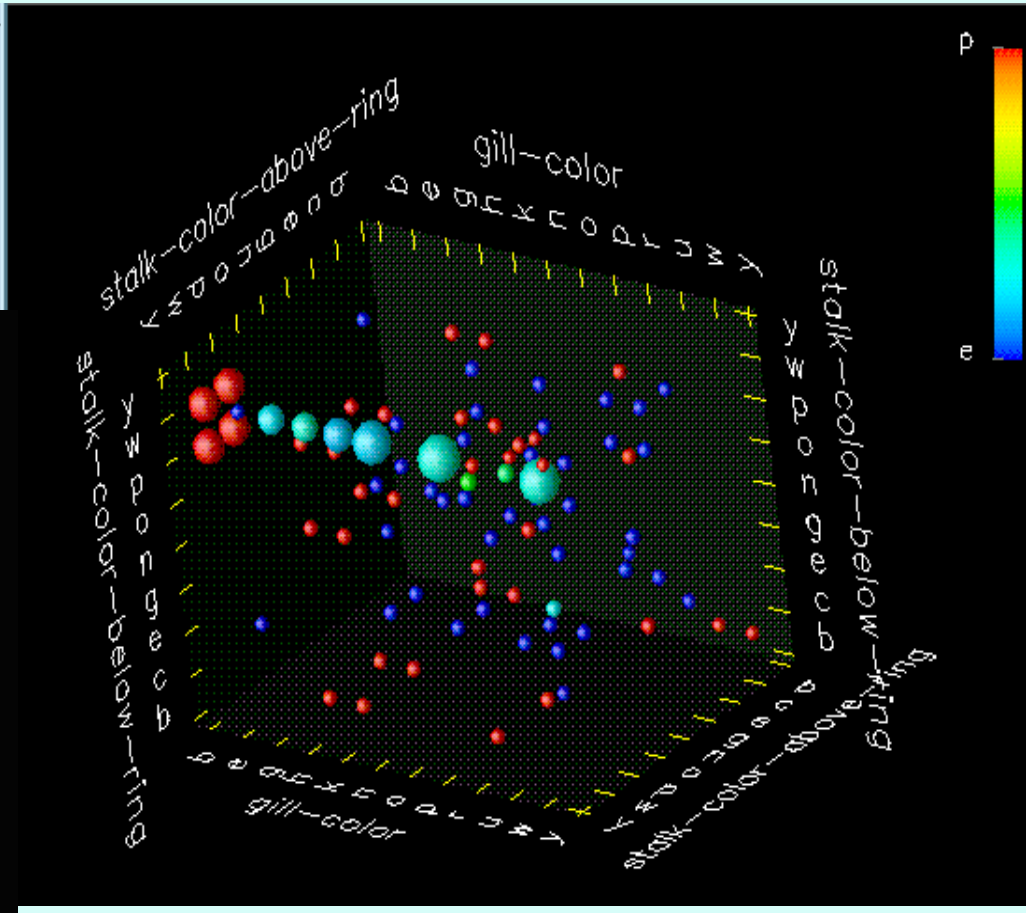
IBM Intelligent Miner

- **Skalowalny, ukierunkowany na przetwarzanie baz danych o dużych rozmiarach**
- **Oferuje wiele metod ekploracji danych**
  - **Asocjacje**
  - **Drzewa klasyfikujące**
  - **Analiza sekwencji**
  - **Grupowania**
- **Narzędzia do wizualizacji**
- **Znacząca inspiracja badawcza dla środowiska baz danych**

IBM Miner - wizualizacja danych







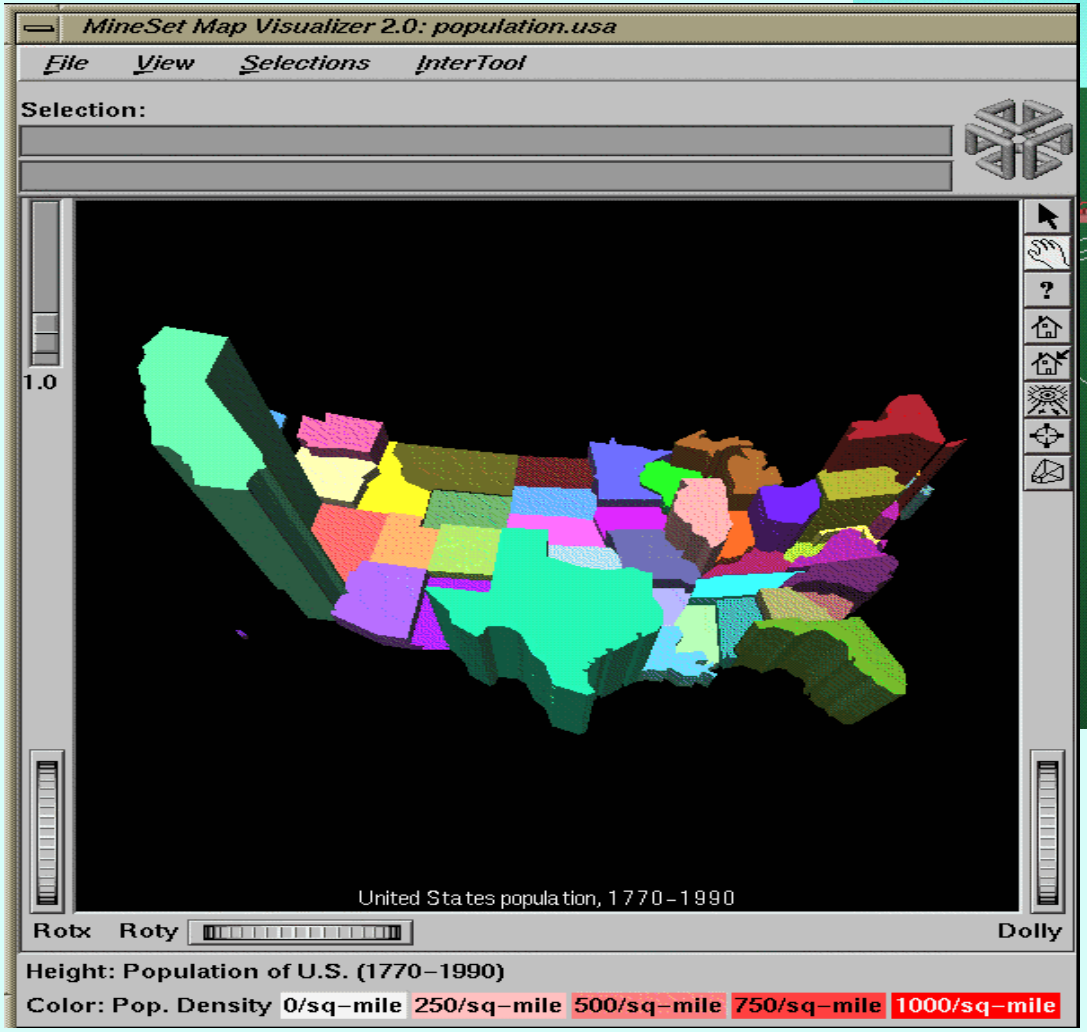
MineSet

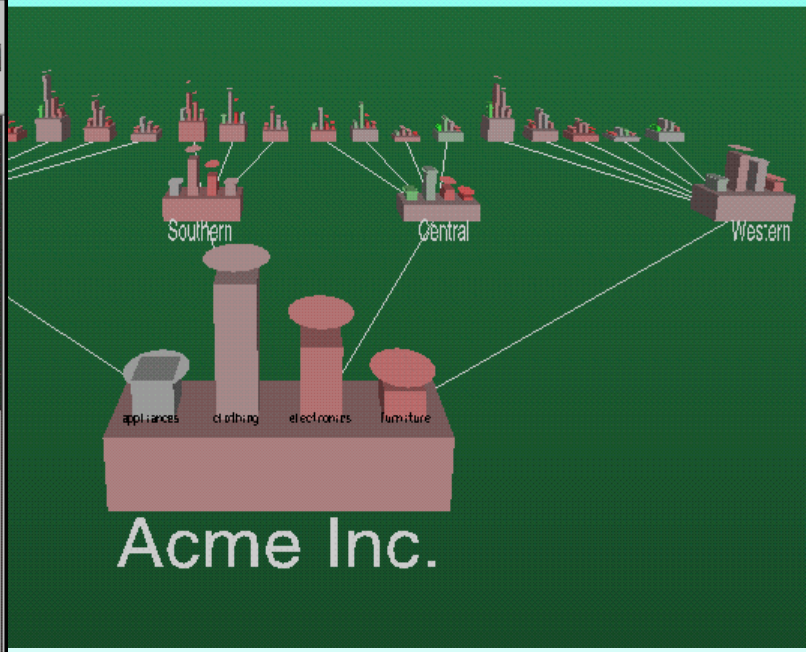
- MineSet – komercyjny system oferowany przez Silicon Graphics
- Interakcyjne środowisko integrujące: dostęp do baz danych i plików, algorytmy eksploracji danych, wizualizaje danych.
- Architektura klient-serwer „skalowalna” dla obliczeń na dużych bazach danych.
- Algorytmy analityczne wywodzące się z projektu MLC++ (Kohavi *et al.*) – poszukaj na WWW
- „Succesfull stories” – znaczące wdrożenia komercyjne.

MineSet (SGI)

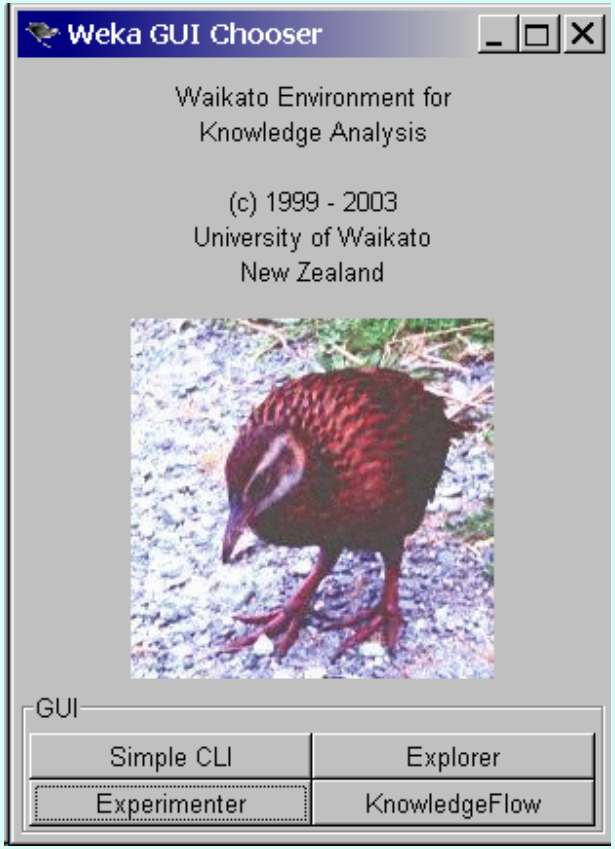
- Przykładowe algorytmy eksploracji danych:
  - naive bayesian,
  - drzewa decyzyjne (C4.5like),
  - drzewa regresji,
  - analiza skupień (k-średnich),
  - poszukiwanie reguł asocjacyjnych,
  - tablice decyzyjne,
  - metody selekcji zmiennych.
- Wizualizacja danych (oparty na „statistical”, „cluster”, „tree” „visualizer”.
- Narzędzia wspomagające wstępne przetwarzanie danych.

SGI – wizualizacja danych





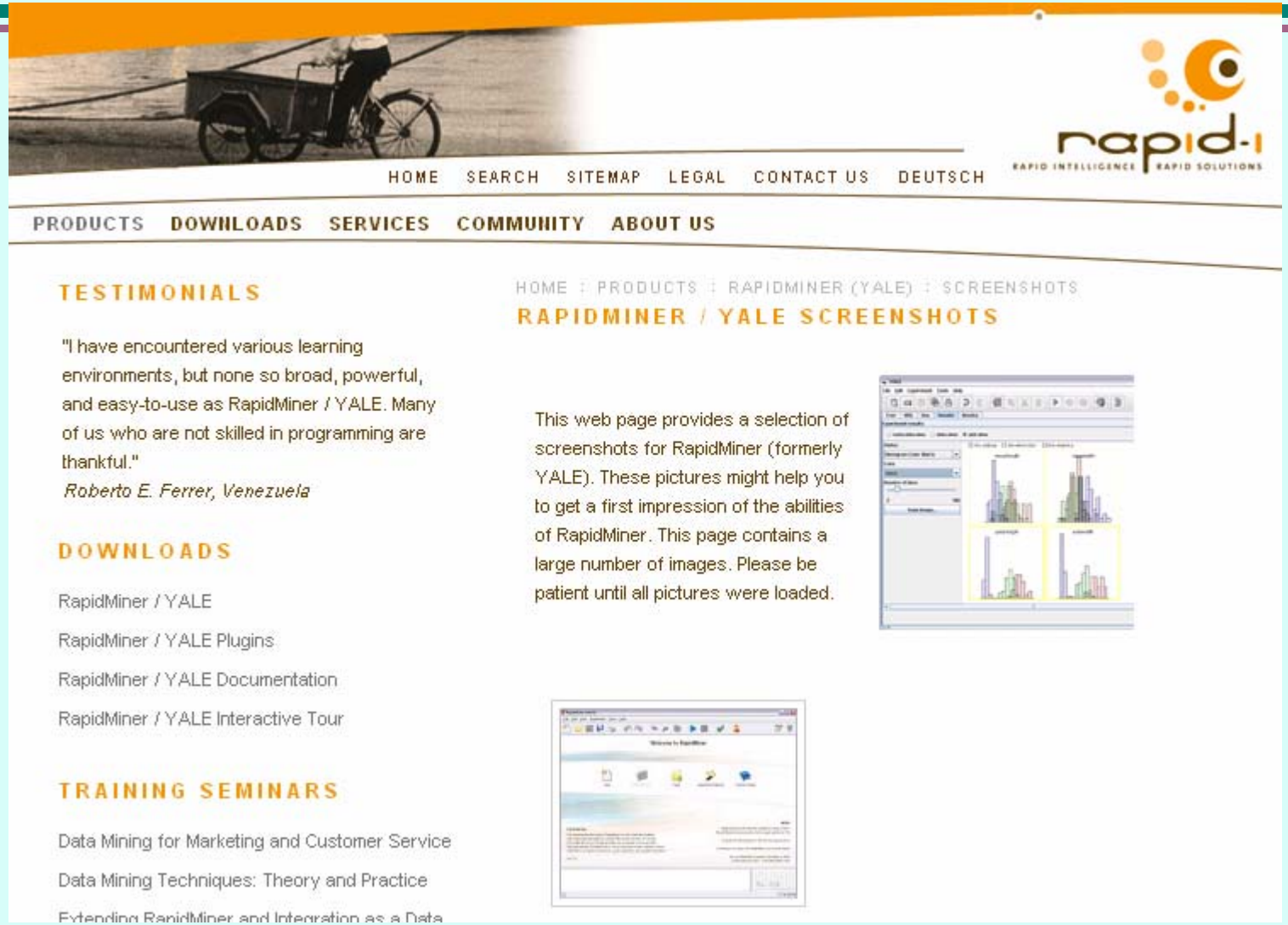
WEKA – Machine Learning and Data Mining



Implementacja w Java
wielu algorytmów

Nie jest to idealny projekt → lecz …

RapidMiner (YALE)



Orange (Slovenia)

orange

DATA MINING
FRUITFUL&FUN

- Home
- Screenshots
- Contact & Support
- Acknowledgements
- Download
- Forum (RSS)
- Documentation
- Search
- Visual Programming
- Catalog of Widgets
- Scripting for Beginners
- Class Reference
- Modules
- Example Scripts
- Data Sets

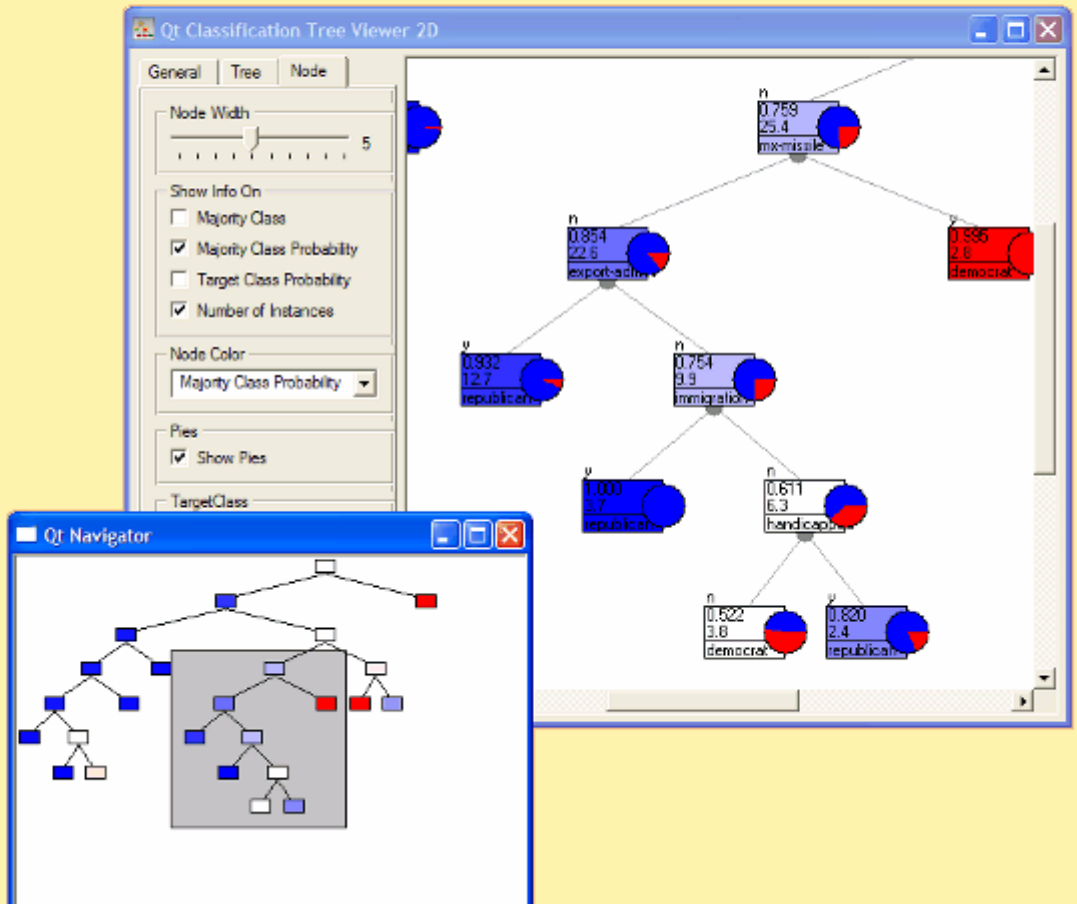
Latest News

Oct 31: The list of example scripts from documentation works again. For instance, you want to know how to induce random forests in

Orange Screenshots

Following are screenshots of Orange Widgets and Orange's visual programming interface for data mi

Classification tree viewer with a navigator.



Industries/fields where you currently apply data mining [KDD Pool - 216 votes total]

Banking (29) 13%
Bioinformatics/Biotech (18) 8%
Direct Marketing/Fundraising (19) 9%
eCommerce/Web (12) 6%
Entertainment/News (1) 0%
Fraud Detection (19) 9%
Insurance (15) 7%
Investment/Stocks (9) 4%
Manufacturing (9) 4%
Medical/Pharma (15) 7%
Retail (9) 4%
Scientific data (20) 9%
Security (8) 4%
Telecommunications (12) 6%
Travel (2) 1%
Other (19) 9%

Przykłady zastosowań eksploracji danych

- **Marketing**
  - **„Target marketing”, identyfikacja profilu klientów, ocena lojalności klientów, problem koszyka zakupów - asocjacje produktów w sieciach sprzedaży, segmentacja rynków, klientów, itp.**
- **Analizy finansowe**
  - **Analiza ryzyka kredytowego, rekomendacje produktów, przewidywanie trendów i przebiegów czasowych,…**
- **Wykrywanie nieprawidłowości i anomalii**
  - **Analiza defraudacji i nieprawidłowości kart kredytowych, systemy telekomunikacyjne, towarzystwa ubezpieczeniowe, systemy opieki medycznej.**
- **Text mining oraz Web mining (zachowania użytkowników w e-serwisach, wspomaganie wyszukiwania informacji), ...**
- **Wiele innych (przemysł, nauka, administracja),…**

SI w przedsiębiorstwach i Data Mining

- Systemy Business Intelligence →” Buzz word” ?
- Nowa kategoria informatycznych systemów zarządzania!
- Business intelligence — [H.Luhn 1958] — „aims to support better business decision-making”.
- Howard Dresner → an umbrella term to describe "concepts and methods to improve business decision making by using fact-based support systems.„
- A broad category of applications and technologies for gathering, storing, analyzing, and providing access to data to help enterprise users make better business decisions.

SI w zarządzaniu

- **Tradycyjne rozwiązania → systemy ewidencyjno-operacyjne (transakcyjne),**
  - **Ewidencja zdarzeń gospodarczych i obsługa bieżącej działalności.**
  - **Klasyczne systemy rachunkowości finansowej, ewidencja księgowa, obsługa sprzedaży, stanu magazynów,..**
- **Zintegrowane systemy zarządzania dla dużych przedsiębiorstw**
  - **MRP II, ERP, SCM, CRM**
- **Nowe rozwiązania - systemy analityczno-decyzyjne**
  - **Zaawansowane raportowanie, analiza danych i przetwarzanie informacji we wskazania przydatne do podejmowania decyzji biznesowych.**

Systemy transakcyjne w przedsiębiorstwie

- **Wykorzystywane na najniższym operacyjnym szczeblu zarządzanie.**
- **Wspomagają podejmowanie decyzji dobrze ustrukturalizownych.**
- **Automatyzują rytunowe sytuacje i procedury działania.**
- **Ukierunkowane na ewidencje faktów ...**
  - **pomimo złożoności procesów -> podstawowe dane dobrze ustrukturalizowane.**
- **Technologia – relacyjne bazy danych**

Cechy charakterystyczne systemów transkacyjnych

- **Duże ilości danych wejściowych.**
- **Duża ilość “wyjść – rezultatów”, dokumentów, raportów, itp.**
- **Efektywność przetwarzania dużych wolumenów danych.**
  - **Wydajność (czas, zasoby pamięciowe)**
  - **Wymagania wobec pamięci dyskowej**
- **Proste operacje przetwarzania.**
- **Wysoki stopień powtarzalności operacji.**
- **Edycja – aktualizacja danych.**

Prezentowanie danych wyjściowych

- **Dokumenty**
  - **Zapis transakcji lub innych danych org.**
  - **Rachunki, faktury, itp.**
  - **Drukowane i elektroniczne (standardy EDI)**
- **Raporty**
  - **Szczegółowa lub zaagregowana informacja operacyjna**
  - **Raporty periodyczne (np. lista kosztów tygodniowych, dobory raport produkcji)**
  - **Raporty na żądanie**
  - **Raporty wyjątków**
- **Wyniki zapytań (formularze lub swobod. SQL)**

Typowe operacje przetwarzania danych

- **Obliczenia – operacje artm. log.**
- **Porównywanie zestawów danych**
- **Agregacja**
  - **Połączenia (join) danych**
  - **Podsumowania**
- **Filtrowanie – usuwanie niepotrzebnych danych z dalszego przetwarzania**
  - **Selekcja i projekcja**
- **Wyszukiwanie**

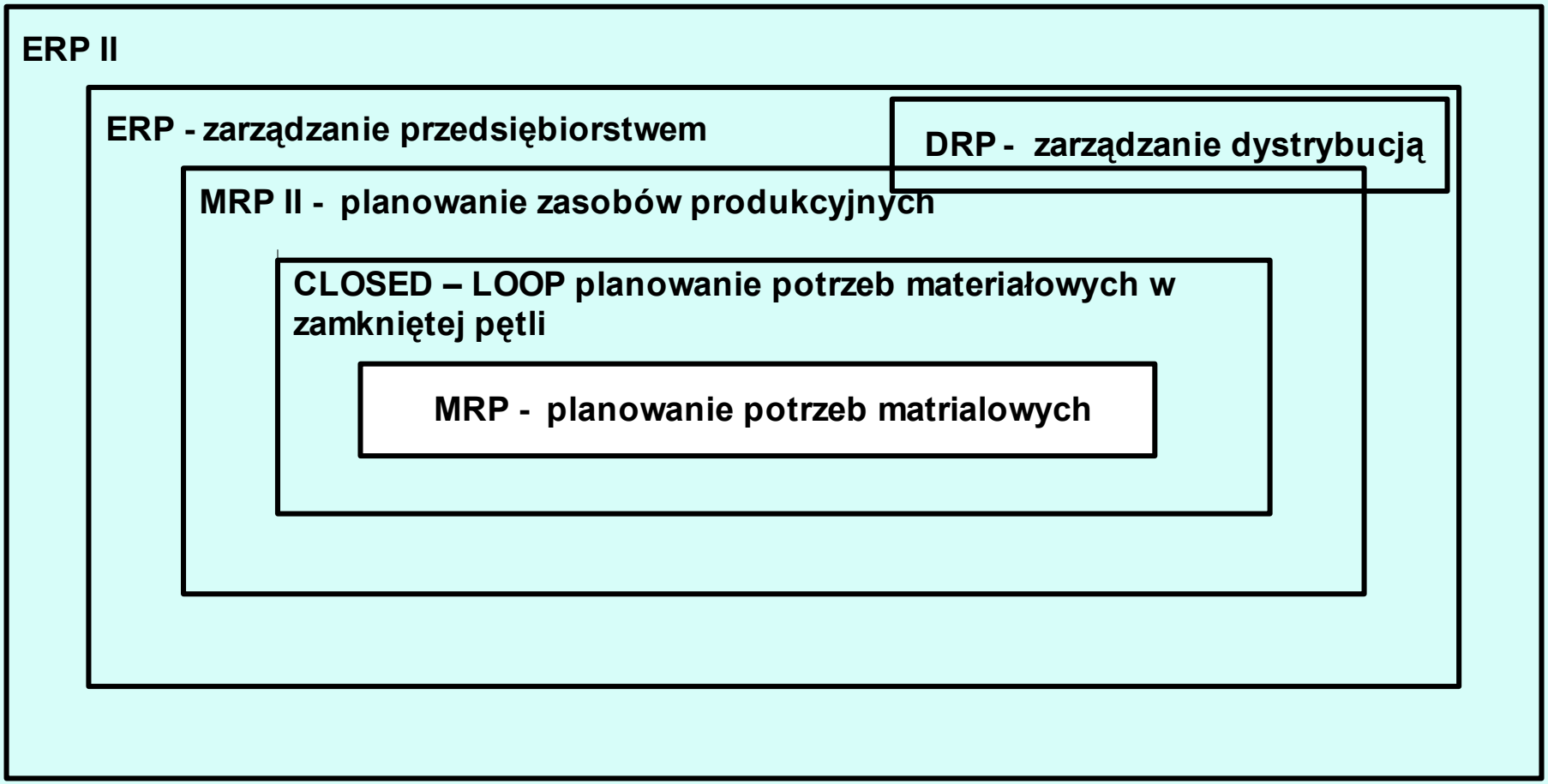
Typowe moduły

- **Sprzedaż**
- **Zakupy**
- **Podsystem wytwarzanie**
  - **Planowanie i harmonogramowanie produkcji**
  - **Operacje produkcyjne**
  - **Koordynowanie magazynów**
- **Zarządzanie zasobami ludzkimi**
- **Finanse i księgowość**

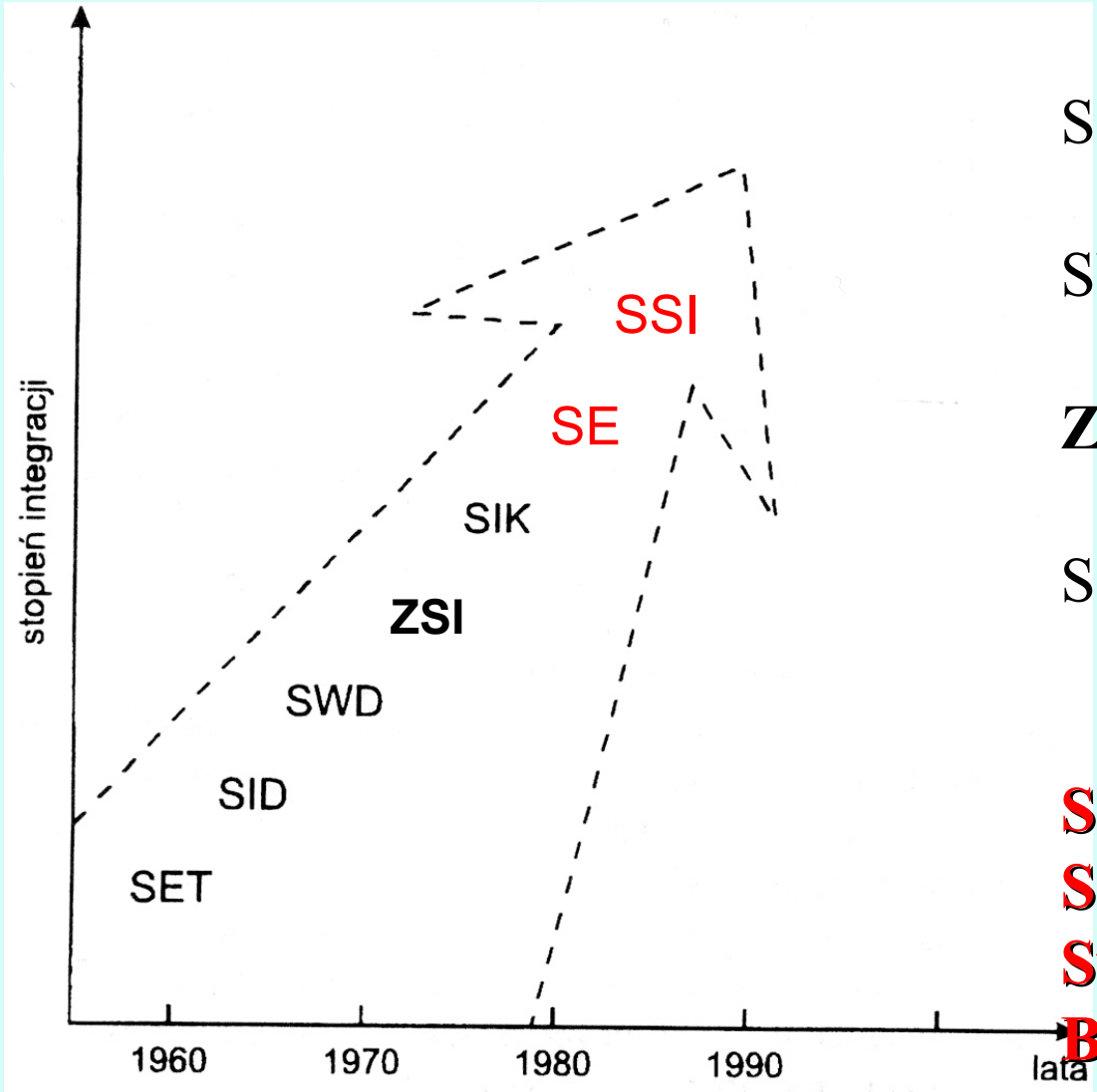
Systemy ZSI klasy **ERP**

- Do Zintegrowanych Systemów Informatycznych cieszących się ogromnym powodzeniem zalicza się systemy klasy ERP (Enterprise Resource Planning-Planowanie Zasobów Przedsiębiorstwa).
- Definiuje się je, jako systemy optymalizujące procesy biznesowe zarówno wewnętrzne w firmie (banku), jak i zachodzące w najbliższym jego otoczeniu, dzięki zastosowaniu gotowych narzędzi pozwalających automatyzować wymianę danych z kooperantami w całym łańcuchu logistycznym.

Klasy systemów informacyjnych zarządzania przedsiębiorstwem



Ewolucja systemów informatycznych do wspomagania zarządzania



SET - Systemy ewidencyjno-transakcyjne

SID - Systemy informacyjno-decyzyjne

SWD - Systemy wspomagania decyzji

ZSI- Zintegrowane systemy informatyczne

SIK - Systemy informowania kierownictwa

SE - Systemy eksperckie

SSI – Zaawansowane Systemy Sztucz. Int.

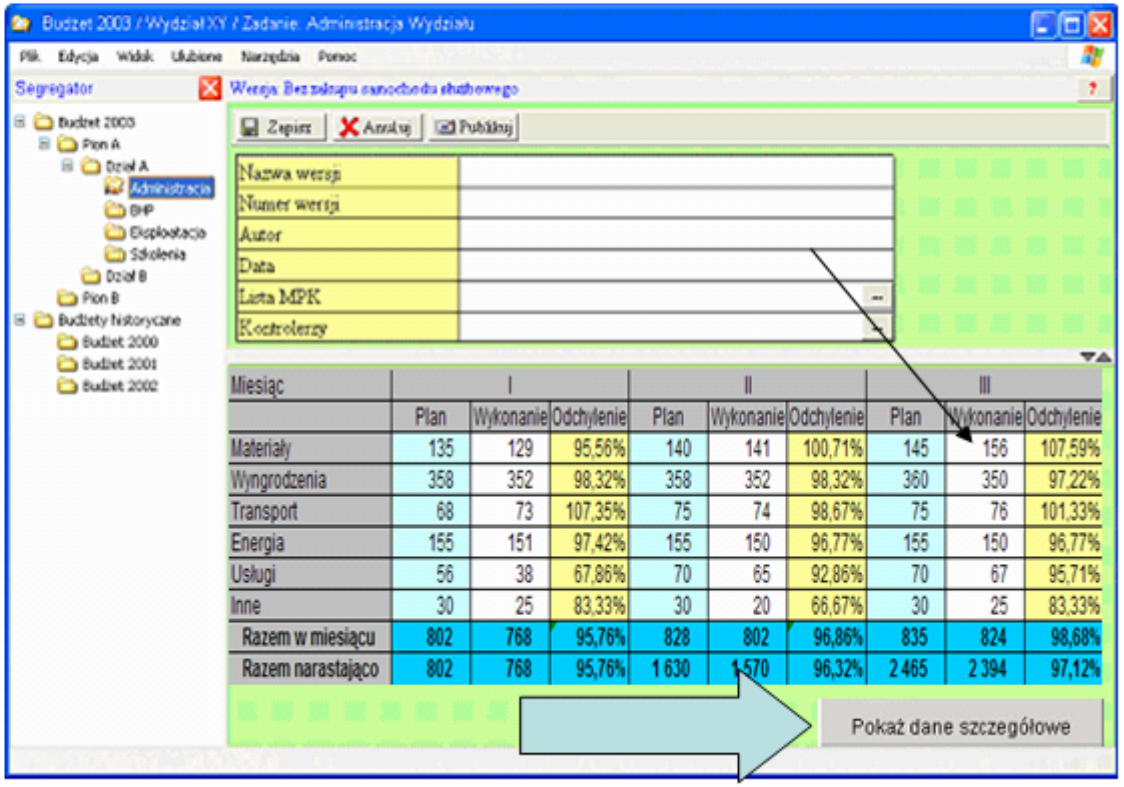
BI – Inteligencja biznesowa

Systemy informowania kierownictwa – SIK (EIS – Executive Information Systems)

- SIK są wyposażone w odpowiedni interfejs i język użytkownika umożliwiający wyszukiwanie i generowanie zbiorów danych wynikowych o swobodnie definiowanej strukturze i zakresie.
  - Opierają się na zasobach systemów ewidencyjno-sprawozdawczych
  - Są to systemy pozwalające skupić uwagę raczej na ogólnym, sprawnym działaniu firmy, niż na optymalizacji decyzji.
  - Służą temu rozbudowane systemy zapytań oraz indywidualizacja przedstawionych raportów i narzędzi komunikacji z systemem.
- Dostarczają informacji głównie kierownictwu najwyższego szczebla.

Przykłady raportowania finansowego

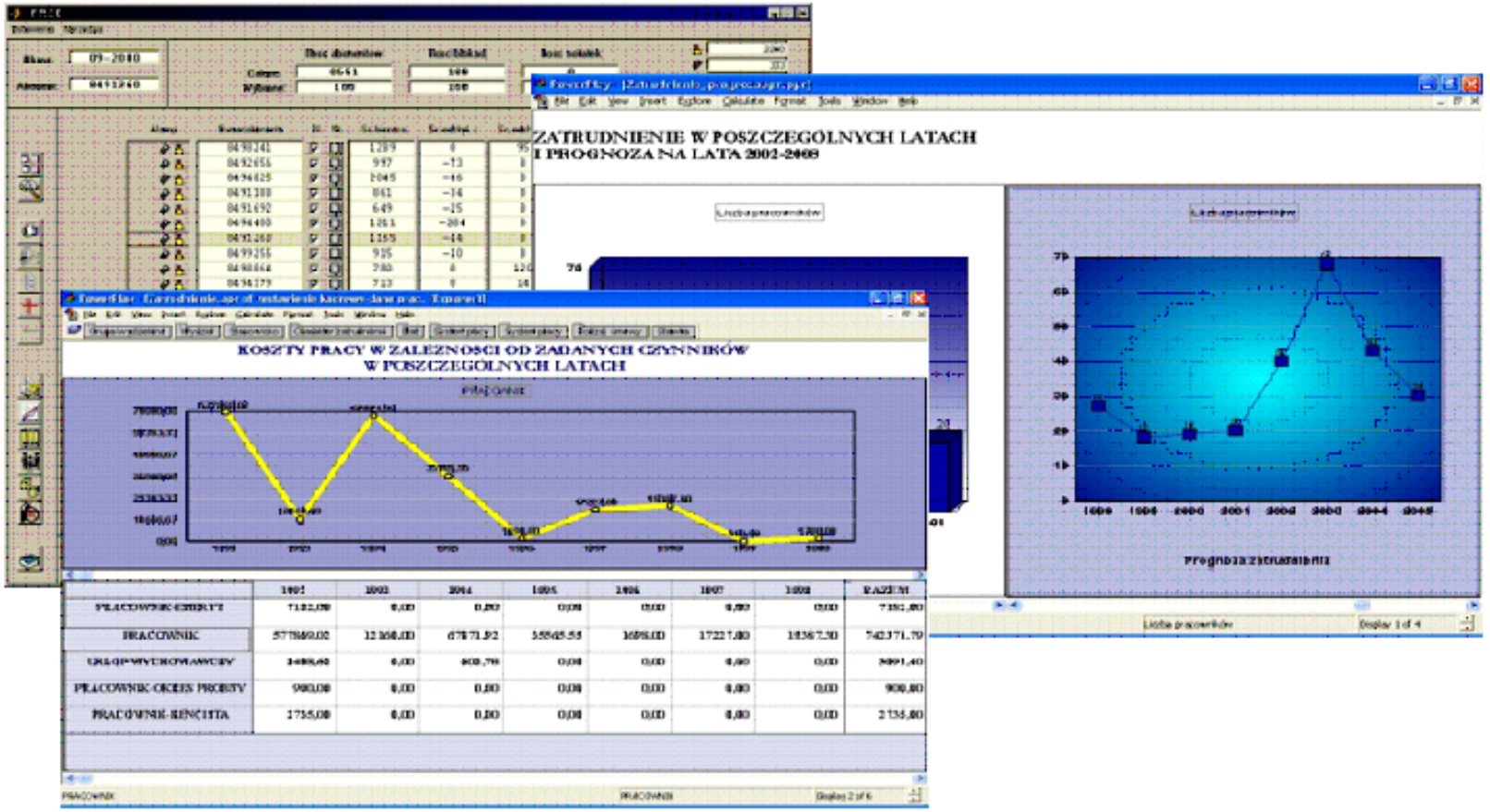
Kontroling operacyjny



Miesiąc	I			II			III		
	Plan	Wykonanie	Odchylenie	Plan	Wykonanie	Odchylenie	Plan	Wykonanie	Odchylenie
Materiały	135	129	95,56%	140	141	100,71%	145	156	107,59%
Wyngrodzenia	358	352	98,32%	358	352	98,32%	360	350	97,22%
Transport	68	73	107,35%	75	74	98,67%	75	76	101,33%
Energia	155	151	97,42%	155	150	96,77%	155	150	96,77%
Usługi	56	38	67,86%	70	65	92,86%	70	67	95,71%
Inne	30	25	83,33%	30	20	66,67%	30	25	83,33%
Razem w miesiącu	802	768	95,76%	828	802	96,86%	835	824	98,68%
Razem narastająco	802	768	95,76%	1 630	1 570	96,32%	2 465	2 394	97,12%

Przykład SIK

Raportowanie w Systemie Informowania Kierownictwa



Systemy wspomagania decyzji (DSS)

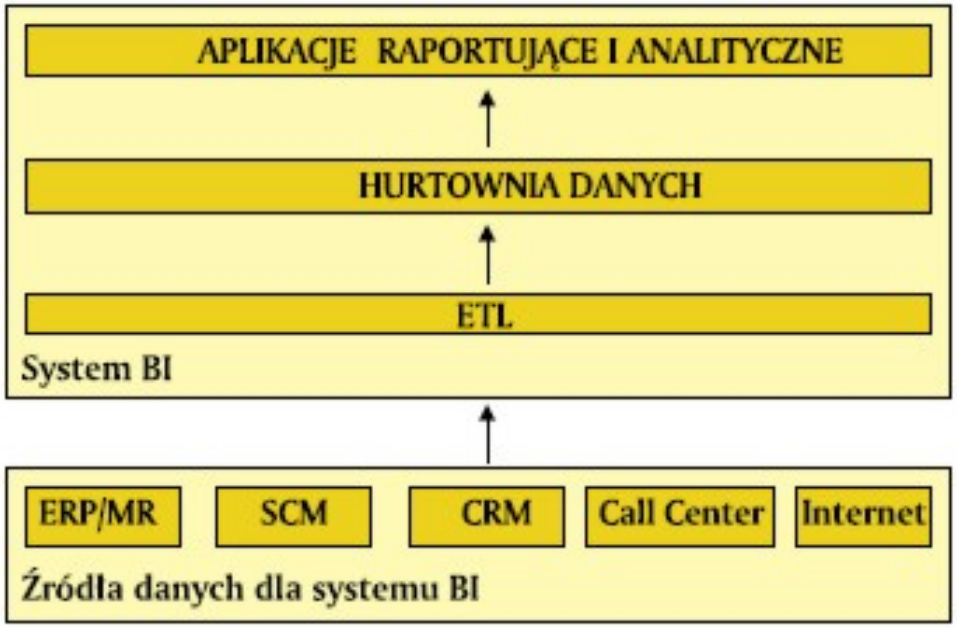
- Termin SWD – Decision Support Systems - powstał na początku lat siedemdziesiątych i został rozwinięty na początku lat osiemdziesiątych.
- Większość bardziej zaawansowanych systemów typu SES i SIK realizuje rutynowe procesy decyzyjne dla standardowych sytuacji decyzyjnych.
- SWD $\rightarrow$ bardziej zaawansowane modele, prognozy i symulacje.
- SWD cechuje wydzielenie bazy procedur (modeli) decyzyjnych z oprogramowania użytkowego oraz możliwość symulowania różnych sytuacji decyzyjnych.
- Użytkownik może dzięki temu analizować (śledzić) proces wyboru modelu i generowania projektów decyzji oraz generowania przez system objaśnień i uzasadnień realizowanego procesu decyzyjnego.

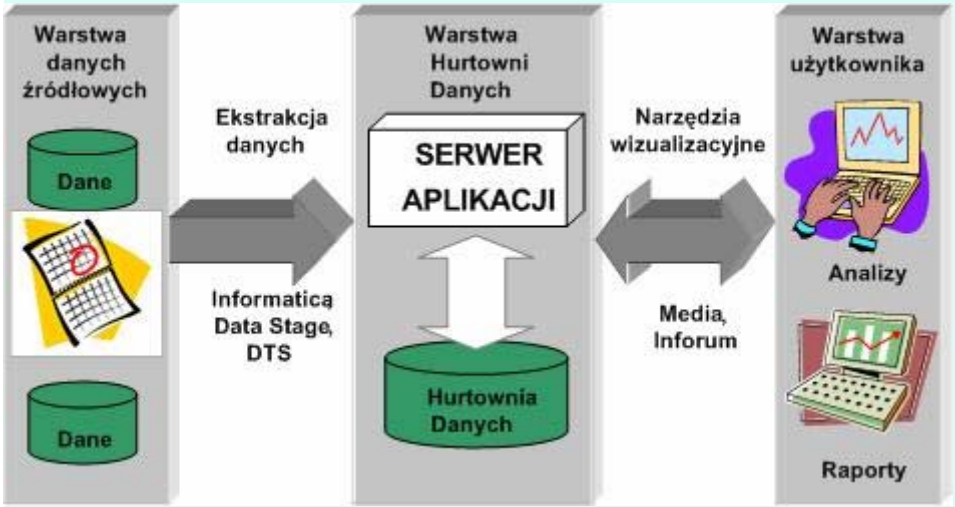
BI definicje [za C.Olszak]

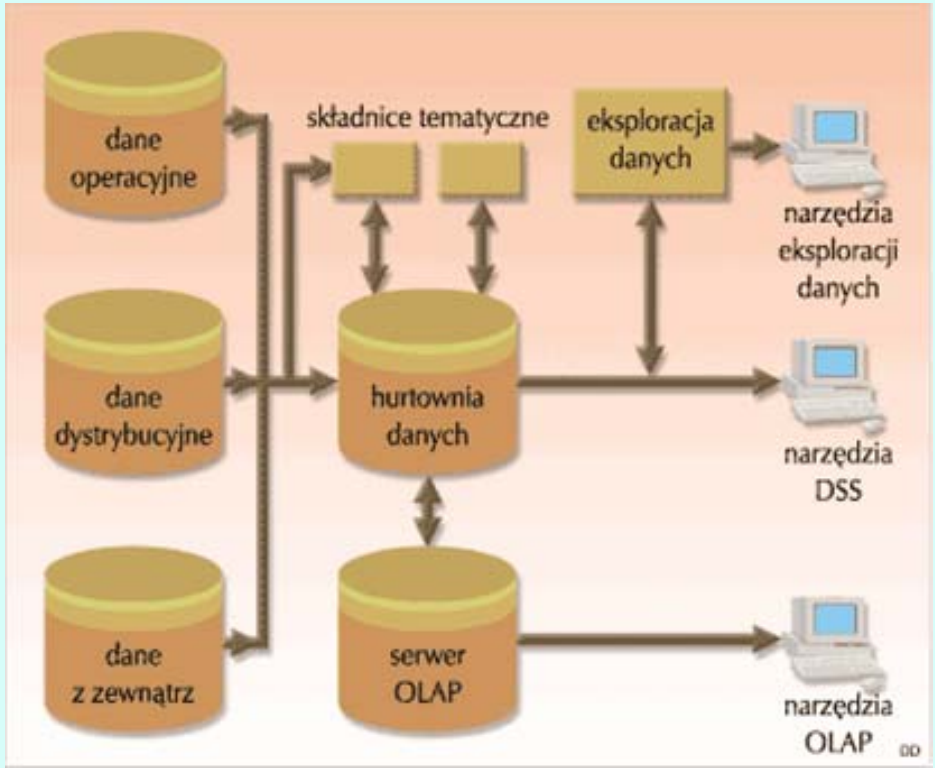
- Systemy BI określane są jako zintegrowany zestaw narzędzi, technologii oraz produktów programowych do zbierania, integrowania, analizowania i udostępniania danych, służący podejmowaniu decyzji na wszystkich szczeblach zarządzania.
- Adresowana do pracowników szczebla kierowniczego oraz analityków.
- Od tradycyjnych SIZ odróżniają je przede wszystkim:
  - szerszy zakres przedmiotowy,
  - wielowariantowa analiza słabo ustrukturalizowanych danych, pochodzących z różnych źródeł oraz ich wielowymiarowa prezentacja.

BI – wyzwania technologiczne

- Drastyczne wymagania wydajnościowe, przede wszystkim z powodu ogromnych rozmiarów danych, które podlegają przetwarzaniu.
- Skupiona wokół technologii hurtowni danych, będących tematycznymi bazami danych, gromadzącymi historyczne dane o działalności przedsiębiorstwa


APLIKACJE RAPORTUJĄCE I ANALITYCZNE
HURTOWNIA DANYCH
ETL
System BI
ERP/MR
SCM
CRM
Call Center
Internet
Źródła danych dla systemu BI


Warstwa danych źródłowych
Dane
Dane
Ekstrakcja danych
Informatica Data Stage, DTS
Warstwa Hurtowni Danych
SERWER APLIKACJI
Hurtownia Danych
Narzędzia wizualizacyjne
Media, Inforum
Warstwa użytkownika
Analizy
Raporty


dane operacyjne
składnice tematyczne
eksploracja danych
narzędzia eksploracji danych
dane dystrybucyjne
hurtownia danych
narzędzia DSS
dane z zewnątrz
serwer OLAP
narzędzia OLAP

Elementy składowe systemów klasy Business Intelligence

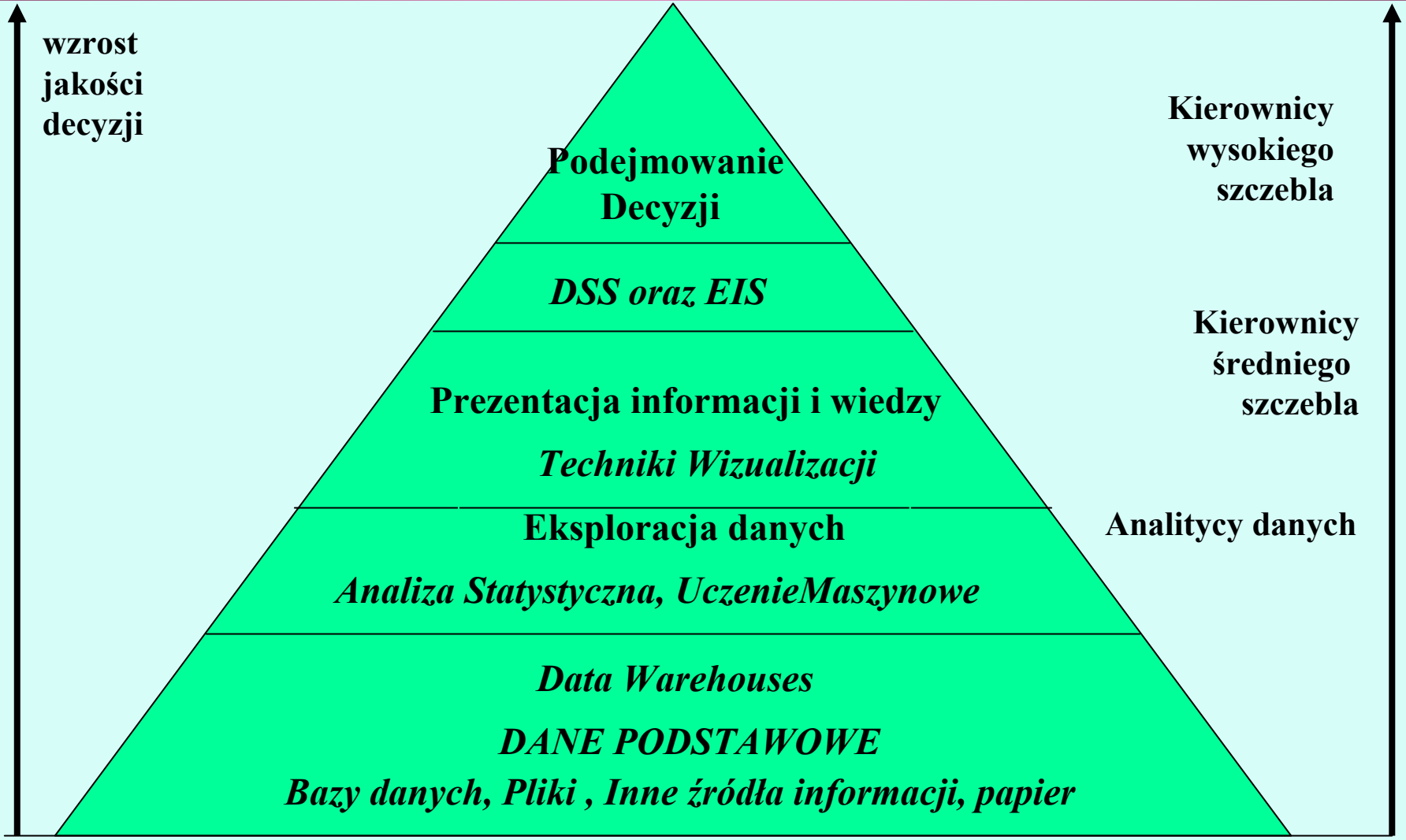
- **Technologie pozyskiwania i transformacji danych (ETL),**
- **Hurtownie danych, w których pozyskane dane są umieszczane,**
- **Aplikacje raportujące i analityczne (OLAP, data mining),**
- **także:**
  - **Systemy informowania kierownictwa (kokpity menadżerskie)**
  - **Systemy udostępniania wiedzy (Portale korporacyjne, e-Business, KM)**
  - **Systemy wspomagania decyzji**

Powiązane komponenty

- ***Aplikacje BI obejmują systemy wspomagania decyzji (DSS - Decision Support Systems), systemy pytająco-raportujące (Q&R - query and reporting), systemy analizy i przetwarzania danych online (OLAP - Online analitical processing), systemy analizy statystycznej, prognozowania i eksploracji danych (Data mining)***

wg. portalu searchCRM.com.

Data Mining oraz Business Intelligence



Przykłady analityki bizensowej

- **Marketing**
  - „Target marketing”, identyfikacja profilu klientów, ocena lojalności klientów, problem koszyka zakupów - asocjacje produktów w sieciach sprzedaży, segmentacja rynków, klientów, itp.
- **Analizy finansowe**
  - Analiza ryzyka kredytowego, rekomendacje produktów, przewidywanie trendów i przebiegów czasowych,…
- **Wykrywanie nieprawidłowości i anomalii**
  - Analiza defraudacji i nieprawidłowości kart kredytowych, systemy telekomunikacyjne, towarzystwa ubezpieczeniowe, systemy opieki medycznej.
- **Text mining** oraz **Web mining (**zachowania użytkowników w e-serwisach, wspomaganie wyszukiwania informacji**), ...**

Studium przypadku sieci sklepów [JSurma, 2009]

Sprzedaż	Analizy ilościowo-wartościowe w podziale na czas, obszar sprzedaży, rodzaj produktu Analizy porównawcze Analiza rentowności Porównanie ze sprzedażą planowaną Ranking, np.. 25 top products Badanie sezonowości indeksów towarowych Analiza wpływu ceny na sprzedaż
Logistyka	Analiza rotacji produktów w centrum logistycznych Zarządzanie zapasami
Finanse	Analiza kontroling planów i ich wykonania Analiza marży Rachunek ABC
Ekspansja	Poszukiwanie najlepszych nowych lokalizacji

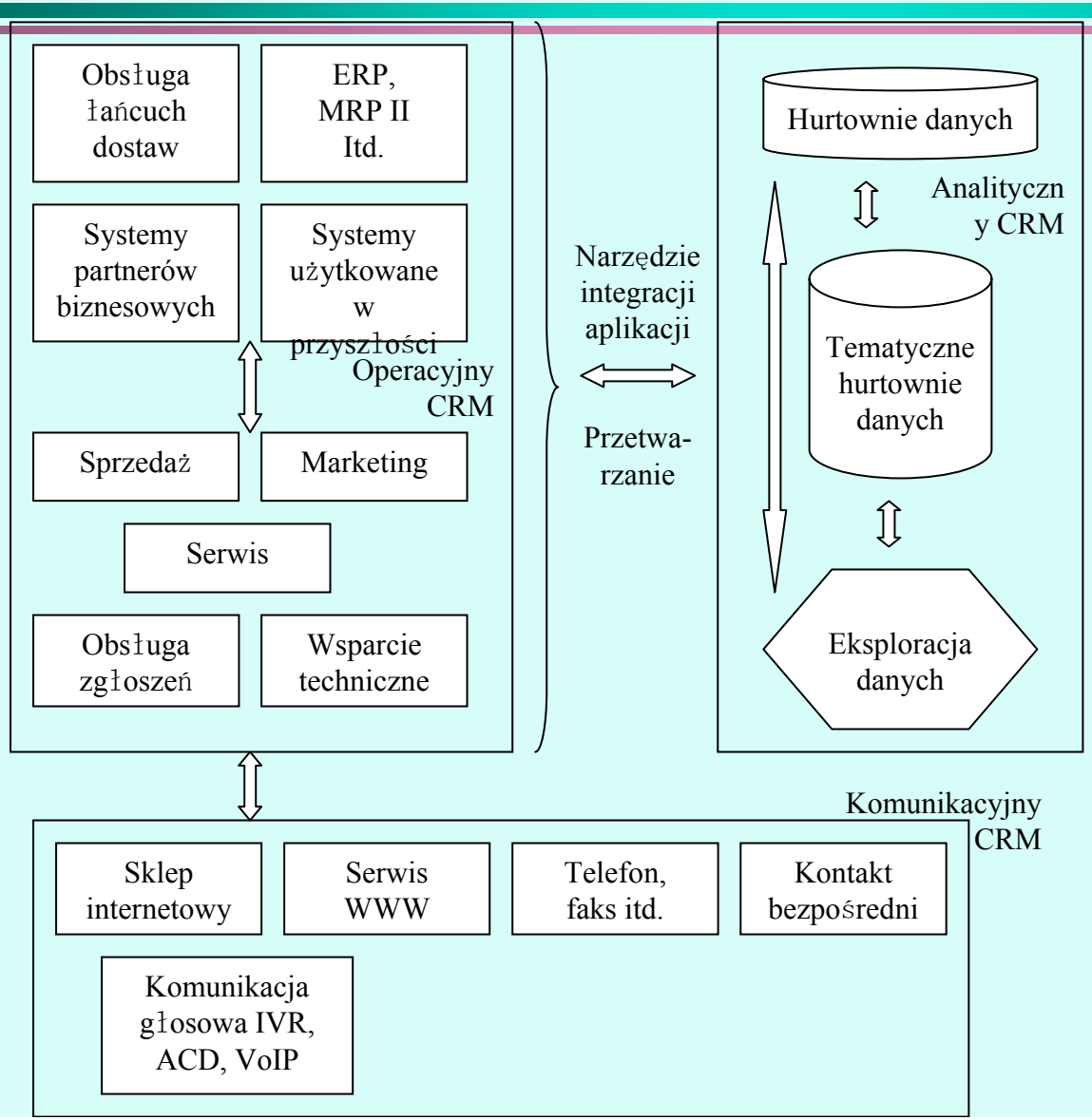
Studium przypadku sieci sklepów [JSurma, 2009]

Zarządzenie sklepami	Badanie koszyka zakupów: Nowe rozmieszczenia produktów Pakiety dla akcji promocyjnych Identyfikacji towarów impulsowych
Zarządzanie strategiczne	Kokpity menadżerskie Analizy wskaźnikowe Porównanie z branżą Poszukiwanie przewagi strategicznej.

Analityka wokół klienta

- **Rozwój systemów CRM**
  - Wzrost zainteresowania pojedynczym klientem
  - Utrzymanie („przywiązanie klienta do firma”) i zwiększenie jego lojalności.
    - „W celu zapewnienia bezpieczeństwo swojej płynności finansowej firma powinna skupić swoją uwagę na 20% swoich najlepszych klientów”
  - Zaoferować im produkty o wysokiej jakości, dobre usługi serwisowe, personalizacja kontaktów oraz ofert, zapewnienie wygody, bezpieczeństwa i zadowolenia ze współpracy.
  - **„Sztuka budowania trwałych związków z klientami”**

CRM jako system informatyczny



Trzy podstawowe części

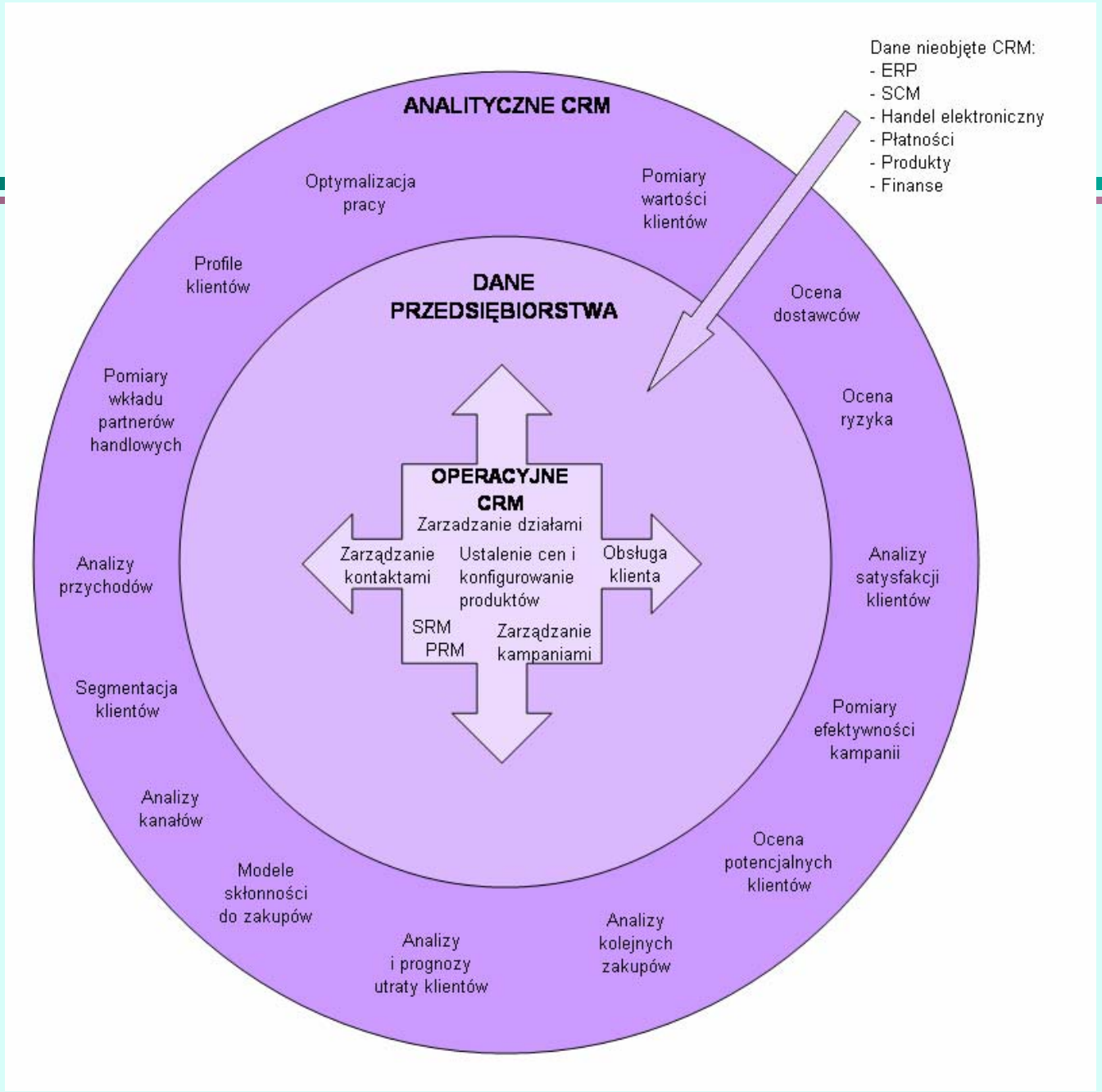
- Operacyjny CRM,
- Analityczny CRM,
- Komunikacyjny CRM.

Analityczne CRM

- Zadanie - przetwarzanie i analiza danych, data mining w celu planowania marketingowego, segmentacji i strategii instrumentalnych.
- Kluczowe pojęciem - wartość klienta.
  - np. dochody, jakie przynosi organizacji, a także jego lojalność oraz skłonność do polecania innym.
- Data mining np. wskaźnik wartości życiowej klienta LTV-Life-time value
  - LTV to przewidywana suma wydatków danego klienta odniesiona do kosztów wytworzenia produktu i kosztów związanych z pozyskaniem i obsługą klienta

Typowe zadania analityczne w CRM

- segmentacja,
- analizy związane z cyklem życia klienta,
  - Identyfikacja potencjalnych klientów (akcje reklamowe / minimalizacja kosztów)
  - Zwiększenie sprzedaży pozyskanym lub aktualnym klientom.
  - Analiza czasu „przetrwania” klienta.
  - Możliwości odejścia klienta (churn/retention)
- analizy dotyczące sekwencji zakupów oraz podobieństwa i powiązań między produktami,
- analiza satysfakcji klientów.


Dane nieobjęte CRM:
- ERP
- SCM
- Handel elektroniczny
- Płatności
- Produkty
- Finanse
ANALITYCZNE CRM
Optymalizacja pracy
Pomiary wartości klientów
Profile klientów
DANE PRZEDSIĘBIORSTWA
Ocena dostawców
Pomiary wkładu partnerów handlowych
Ocena ryzyka
OPERACYJNE CRM
Zarzadzanie działami
Zarządzanie kontaktami
Ustalenie cen i konfigurowanie produktów
Obsługa klienta
SRM
PRM
Zarządzanie kampaniami
Analizy przychodów
Analizy satysfakcji klientów
Segmentacja klientów
Pomiary efektywności kampanii
Analizy kanałów
Ocena potencjalnych klientów
Modele skłonności do zakupów
Analizy kolejnych zakupów
Analizy i prognozy utraty klientów

Metody Data Mining w aCRM

sowymi oraz wykresem histogramu dla aktualnie podświetlonej na liście zmiennej.

Na podstawie informacji zawartych w metadanych, użytkownik dokonuje wstępnej selekcji wielkości, które chce poddać analizie. Może także tworzyć nowe kolumny oraz wykonywać na nich operacje analogiczne jak w przypadku widoku danych.

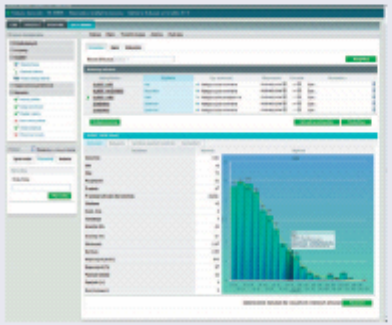
Budowa modeli opartych o Data Mining

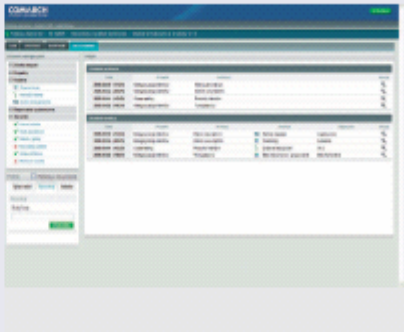
Aplikacja wyposażona jest w szereg sprawdzonych narzędzi analitycznych oraz nowoczesnych technik Data Mining. Narzędzia te wspierają następujące obszary analityczne:

- deduplikacja danych,
- analiza brakujących wartości,
- wykrywanie nietypowych wartości,
- redukcja wymiaru analizy,
- analiza korelacji dwóch i wielu zmiennych,
- regresja liniowa, multiplikatywna, eksponencjalna i logistyczna,
- tworzenie grup kategorycznych w oparciu o reguły narzucone przez użytkownika,
- testowanie statystyczne,
- grupowanie w oparciu o samouczące się algorytmy bez nauczyciela,
- klasyfikacja przy pomocy samouczących się algorytmów drzew decyzyjnych,
- analiza koszykowa.

Każde z tych narzędzi może być stosowane samodzielnie na zbiorze danych zawartych w arkuszu kalkulacyjnym aplikacji. Wstępna, optymalna dla większości zastosowań, parametryzacja udostępnionych narzędzi gwarantuje uzyskanie poprawnych i gotowych do wykorzystania w dalszych obliczeniach wyników, nawet w sytuacji, gdy aplikacją posługuje się osoba nie posiadająca wykształcenia statystycznego. Wszystkie kluczowe dla danej analizy parametry oraz zakres otrzymanych w ramach analizy wyników mogą być konfigurowane według preferencji użytkownika w celu dostosowania się do specyficznych potrzeb analizy oraz specyfiki danych, na których analiza ta jest przeprowadzana.

Budowanie efektywnych modeli Data Mining oznacza połączenie wszystkich tych narzędzi w jeden cykl rozpoczynający się wyborem cech, które mogą mieć wpływ na wielkość wynikową, następnie przechodzący poprzez detekcję i eliminację duplikatów, analizę brakujących wartości itd., a kończący się budową modelu klasyfikacyjnego, który można zapisać i wykorzystać do klasyfikowania nowych obiektów (klientów) zapisywanych w bazie danych.



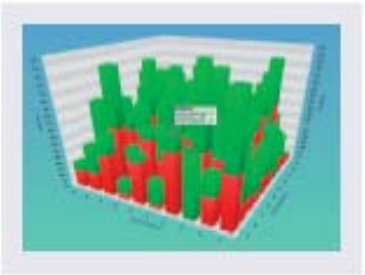


Algorytmy grupowania:

- **K-means, k-medoids, PAM, CLARA.**
- **DBSCAN**
- **AHC, BIRCH**

Drzewa klasyfikacyjne:

- **CART**
- **C4.5**
- **CHAID**
- **QUEST**



Histogram 3D

Drzewo decyzyjne ilustrujące zależność decyzji od numeru klastra oraz wartości Cechy1

Inne zastosowania BI oraz Data Mining

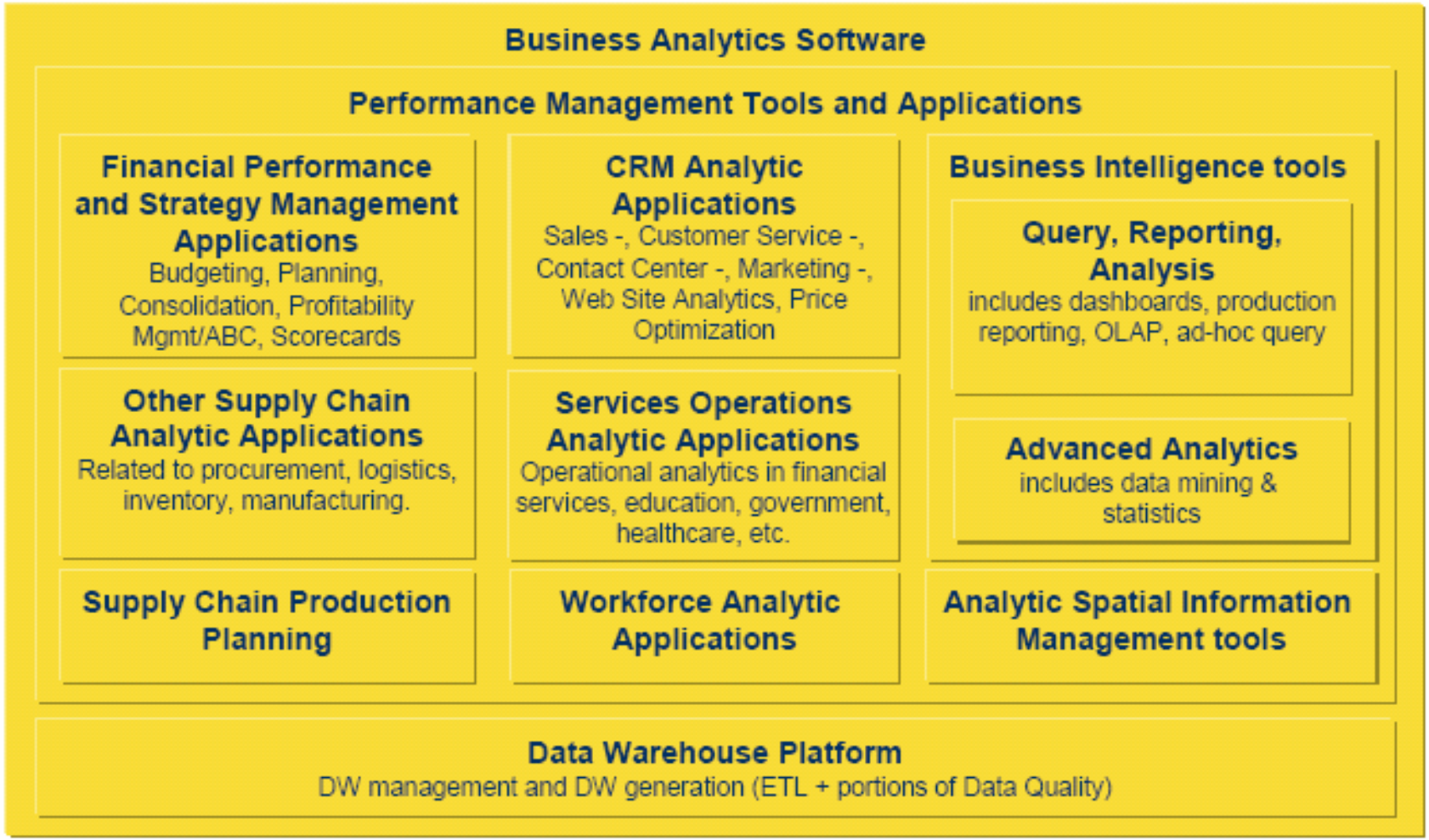
- **Logistyka**

  Badanie popytu na podstawie danych historycznych (drzewa regresji, prognozowanie dla szeregów czasowych)

- Finansowy kontroling
  - Wykrywanie zagrożeń i przewidywanie bankructw
  - Analiza ryzyka finansowego (np. zdolności kredytowej)
    - Modele analizy dyskyrminacyjnej, sieci neuronowe, specjalizowane metody klasyfikacyjnej)
  - Wykrywanie nadużyć (fraud detection)

BI vs. Business Analytics Tools

IDC's Business Analytics Software Taxonomy, 2008



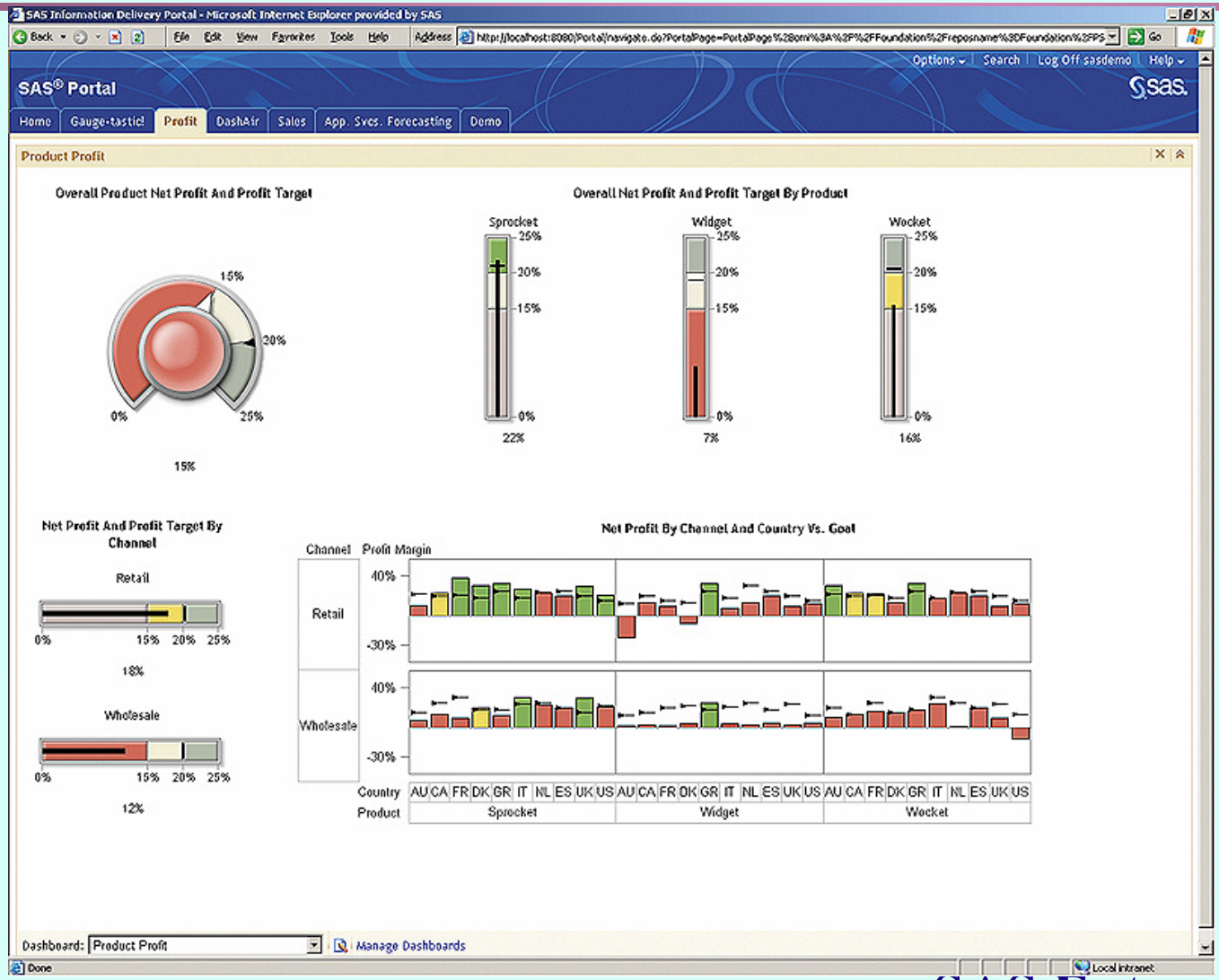
Source: IDC, 2008

Nowe narzędzia IT + nowe koncepcje zarządzania + właściwi specjaliści z danego przedsiębiorstwa

Kokpity menadżerskie (management dashboard**)**

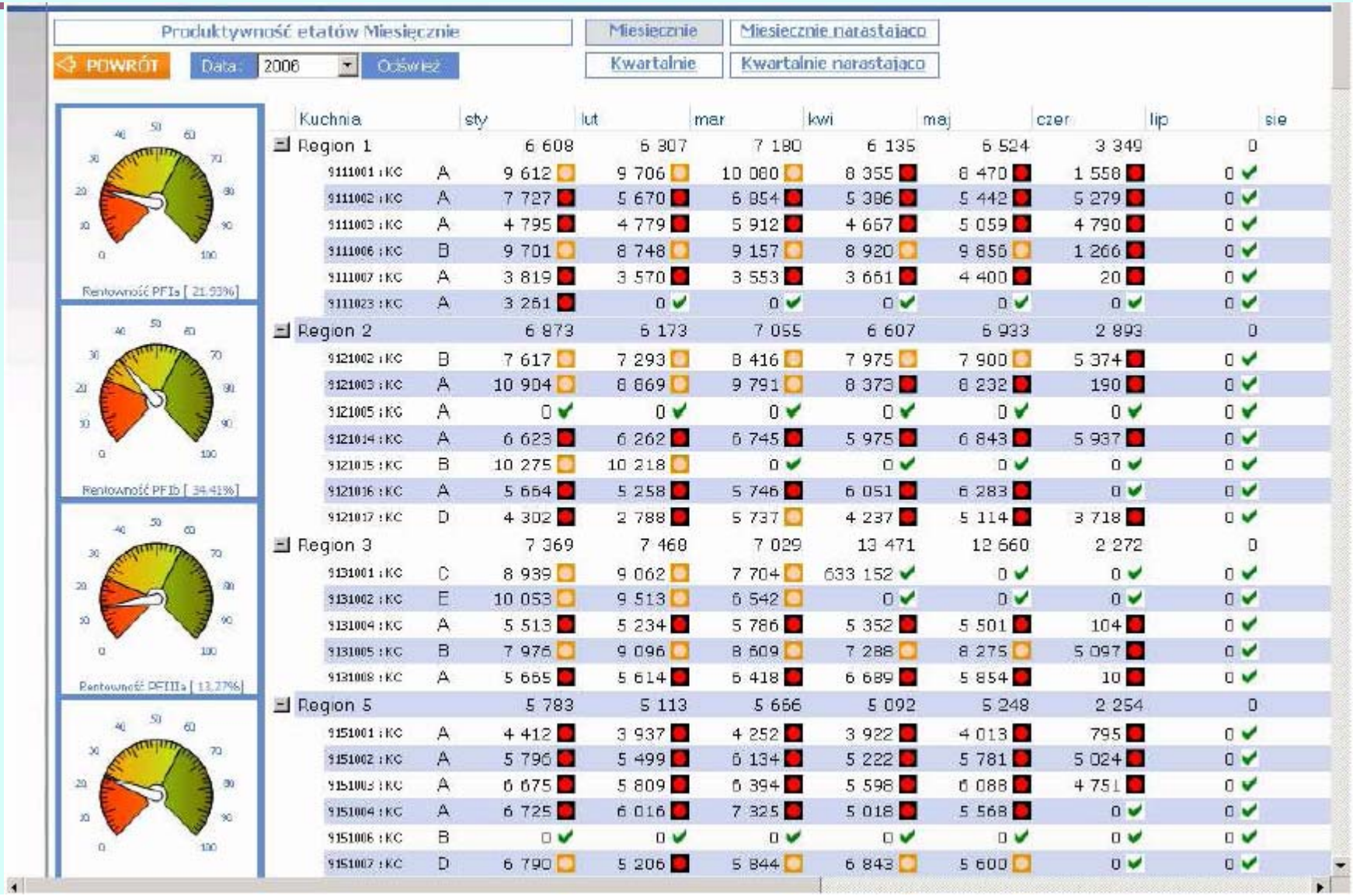
- Aplikacje analityczne, dające naczelnemu kierownictwu narzędzia do dopasowywania czynności operacyjnych do strategii firmy, monitorowania metryk biznesowych oraz zarządzania wydajnością przedsiębiorstwa
- Kokpity prezentują kluczowe wskaźniki biznesowe z punktu widzenia strategii całej organizacji, dzięki czemu pozwalają użytkownikom skoncentrować się na działaniach, które mają największy wpływ na strategię firmy.
- Kokpity wspierają koncepcje zarządzania, takie jak: Zrównoważona Karta Wyników (BSC - *Balanced Scorecard*), Six Sigma czy TQM.

Narzędzia prezentacyjne - kokpity



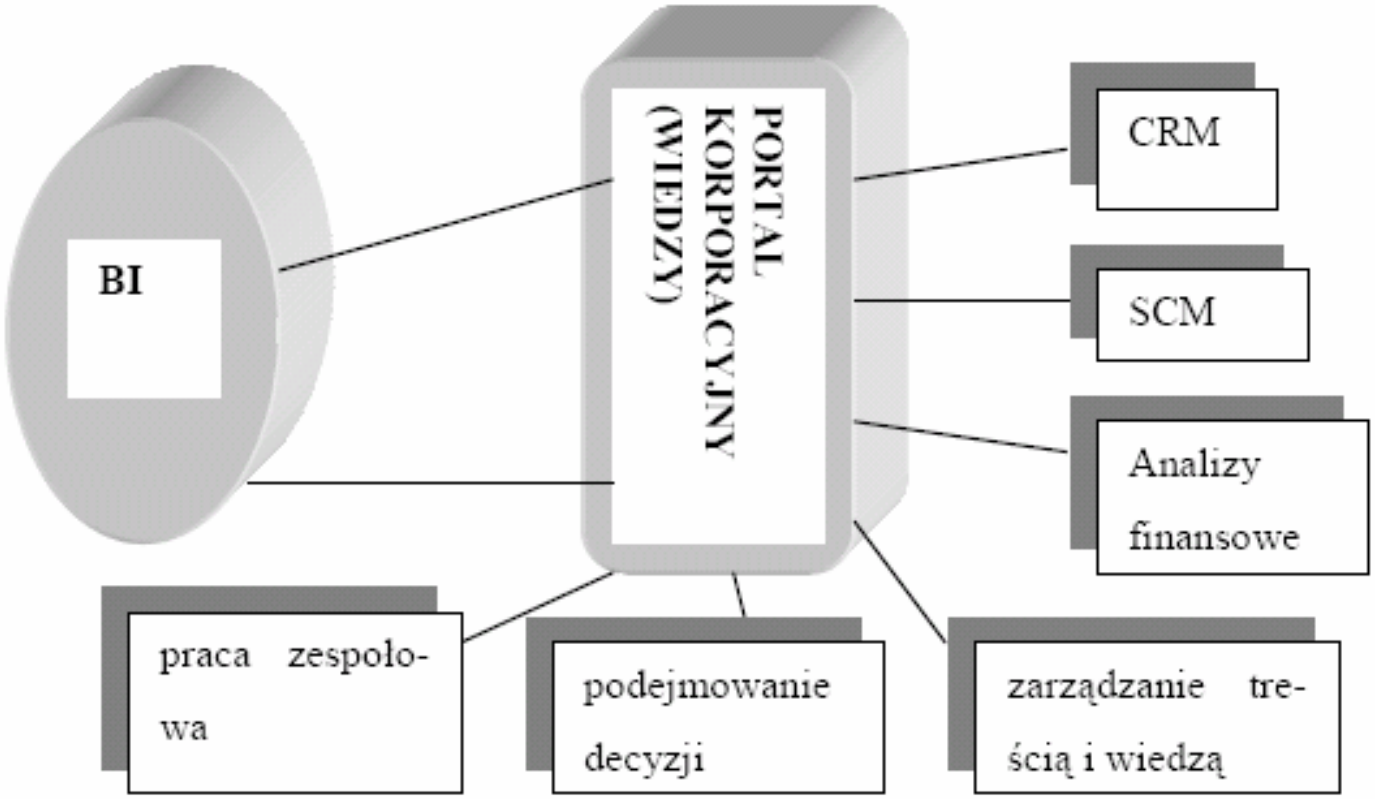
☜ **SAS Enterprise BI**

Narzędzia śledzenia wskaźników (Inforum CPM)



Kuchnia		sty	lut	mar	kwi	maj	czer	lip	sie
Region 1		6 608	6 307	7 180	6 135	6 524	3 349	0	
9111001 : KC	A	9 612	9 706	10 080	8 355	8 470	1 558	0	
9111002 : KC	A	7 727	5 670	6 854	5 386	5 442	5 279	0	
9111003 : KC	A	4 795	4 779	5 912	4 667	5 059	4 790	0	
9111006 : KC	B	9 701	8 748	9 157	8 920	9 856	1 266	0	
9111007 : KC	A	3 819	3 570	3 553	3 661	4 400	20	0	
9111023 : KC	A	3 261	0	0	0	0	0	0	
Region 2		6 873	6 173	7 055	6 607	6 933	2 893	0	
9121002 : KC	B	7 617	7 293	8 416	7 975	7 900	5 374	0	
9121003 : KC	A	10 904	8 869	9 791	8 373	8 232	190	0	
9121005 : KC	A	0	0	0	0	0	0	0	
9121014 : KC	A	6 623	6 262	6 745	5 975	6 843	5 937	0	
9121015 : KC	B	10 275	10 218	0	0	0	0	0	
9121016 : KC	A	5 664	5 258	5 746	6 051	6 283	0	0	
9121017 : KC	D	4 302	2 788	5 737	4 237	5 114	3 718	0	
Region 3		7 369	7 468	7 029	13 471	12 660	2 272	0	
9131001 : KC	C	8 939	9 062	7 704	633 152	0	0	0	
9131002 : KC	E	10 053	9 513	6 542	0	0	0	0	
9131004 : KC	A	5 513	5 234	5 786	5 352	5 501	104	0	
9131005 : KC	B	7 976	9 096	8 609	7 288	8 275	5 097	0	
9131008 : KC	A	5 665	5 614	6 418	6 689	5 854	10	0	
Region 5		5 783	5 113	5 666	5 092	5 248	2 254	0	
9151001 : KC	A	4 412	3 937	4 252	3 922	4 013	795	0	
9151002 : KC	A	5 796	5 499	6 134	5 222	5 781	5 024	0	
9151003 : KC	A	6 675	5 809	6 394	5 598	6 088	4 751	0	
9151004 : KC	A	6 725	6 016	7 325	5 018	5 568	0	0	
9151006 : KC	B	0	0	0	0	0	0	0	
9151007 : KC	D	6 790	5 206	5 844	6 843	5 600	0	0	

Wyniki BI – co dalej?



Rys. 3. Powiązanie BI z portalem korporacyjnym

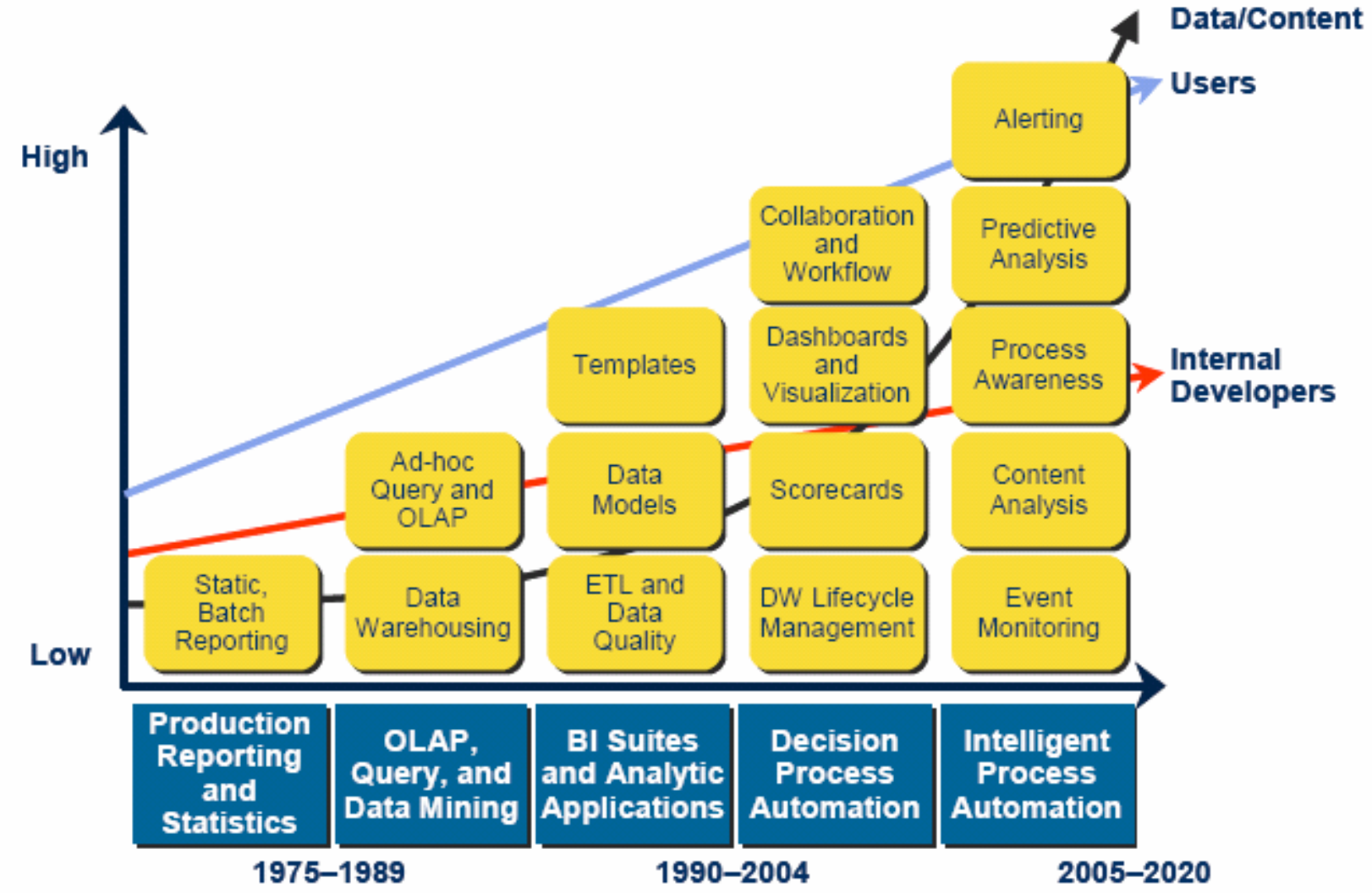
Portale korporacyjne

- Przeniesienie narzędzi BI na poziom procesów operacyjnych organizacji
- Rozszerzenie grupy użytkowników
  - z kierownictwa na pracowników podejmujących decyzje na niższych szczeblach
  - Możliwość udostępnienia części analiz także partnerom zewnętrznym (łańcuchy dostaw)
- Wykorzystanie technologii Intranetu organizacji
- Pobieranie danych z portali internetowych oraz udostępnianie wyników analizy za pośrednictwem przeglądarek
  - Możliwości formatowanie wyniku na urządzenia mobilne, telefony, emajle i inne ..
- „Inteligentne” wyszukiwanie w portalu

Wymagania stawiane systemom zarządzania wiedzą

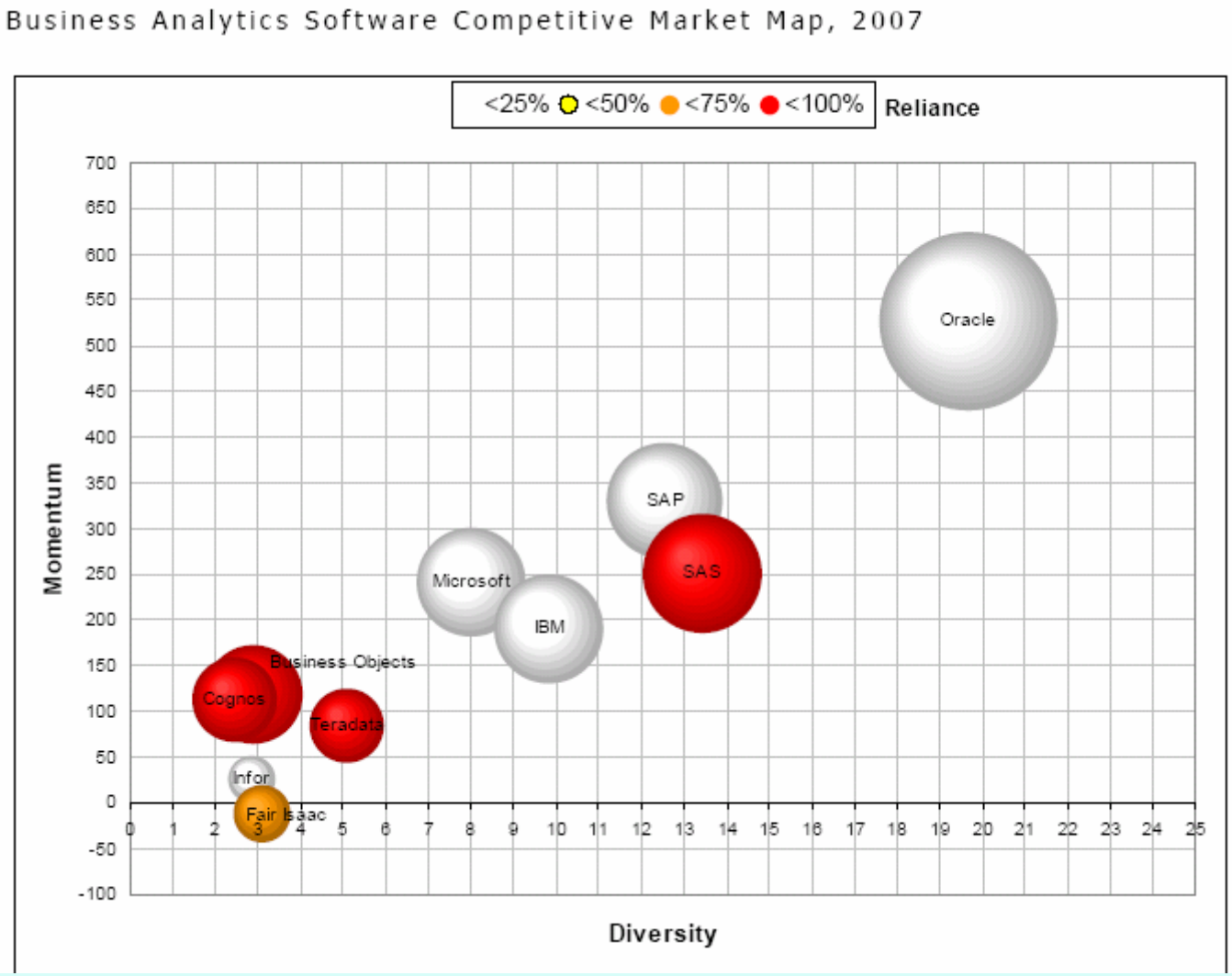
- Zapewnienie mechanizmów dostępu do wspólnych dla całej organizacji danych i dokumentów - istotne zwłaszcza w dużych instytucjach. "Lustrzana" funkcjonalność powinna zostać zapewniona klientom i partnerom biznesowym dla różnych zbiorów informacji.
- Dostarczenie danych menedżerom różnych szczebli.
- Obieg informacji dotyczących samej organizacji i procesów w niej się odbywających - ważne głównie dla dużych organizacji, w których przekazywanie informacji jest procesem złożonym. Podobna funkcjonalność, lecz w nieco innej formie, byłaby przydatna dla części klientów oraz inwestorów - właścicieli instytucji.
- Dostęp do "żywej wiedzy" - umożliwienie kontaktu z pracownikami-ekspertami, na przykład poprzez grupy dyskusyjne.

Business Analytics Market Trends



Source: IDC, 2008

Zestawienie dostawców oprogramowania BI i BA



Przykłady dostawców

- Oracle, SAP, Microsoft, Hyperion Solutions, Comshare, Adaytum, Seagate Software, Pilot Software, Gentia Software, SAS Institute, WhiteLight, Speedware, Kenan oraz Information Builders

Wyniki badań ankietowych w Polsce

- Raport H.Dyducz „Wstępna analiza istniejących rozwiązań informatycznych w obiektach gospodarczych w kontekście przetwarzania analitycznego” 2007
- Hurtownie → 60% (jest lub będzie w b. dużych przedsiębiorstwach, bankowość, telekomy, itp..)
- W 46% dużych firm korzysta się ze specjalizowanych systemów raportowania.
- Prawie 35% firm wykorzystuje specjalizowane pakiety statystyczne oraz systemy wspomagania decyzji
- w prawie 20% również data mining oraz specjalizowane pakiety wizualizacyjne.
- Trochę niższe wyniki z raportu w firmie KPMG z 2004r.

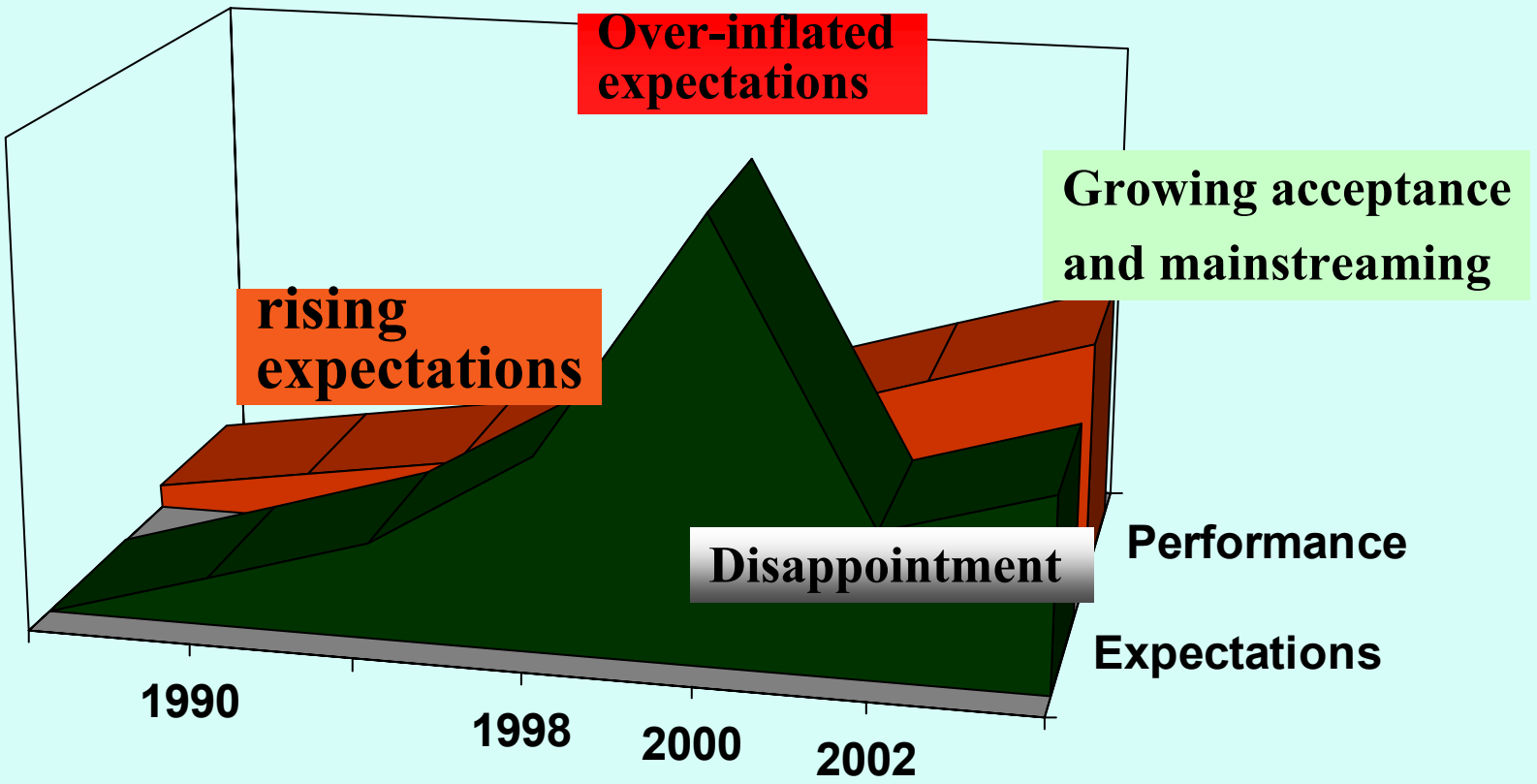
Perspektywy i kierunki rozwoju Data Mining

- **Większa integracja z technologią relacyjnych baz danych, magazynami danych, OLAP.**
- **Skalowalność obliczeń i przetwarzanie danych o wielkich rozmiarach**
  - **efektywne algorytmy dla wielu funkcji eksploracji danych, przetwarzanie równoległe i rozproszone, przetwarzanie przyrostowe**
- **Udostępnianie systemów oferujących wiele zintegrowanych metod**
- **Nowe języki zapytań – zwłaszcza do wiedzy**
- **„Visual data mining" i poszukiwanie nowych propozycji oceny i wizualizacji odkrytej wiedzy**
- **Nowe środowiska dla interaktywnej eksploracji danych**
- **Eksploracja złożonych i nowych typów danych, np. czasowe, tekstowe, multimedialne, rozproszone, ...**
- **Nowe propozycje analizy „niedoskonałych" danych**
- **Zastosowania – nowe wyzwania i podsumowanie doświadczeń**

Podsumowanie

- **Problemem nie jest elektroniczne gromadzenie danych ale ich właściwa analiza i wyciąganie użytecznych wniosków**
- **Metody statystyczne i uczenia maszynowego mogą być podstawą do eksploracji danych**
- **Metody klasyfikacji są najlepiej rozwinięte w ramach eksploracji danych; można je stosować do rozwiązanie szeregu problemów praktycznych**
- **Eksploracja danych jest wraz z metodami badań operacyjnych istotną częścią „biznesowych” systemów wspomagania decyzji**

The Hype Curve for Data Mining and Knowledge Discovery



Literatura:

- J.Zieliński (red.): Inteligentne systemy w zarządzaniu. Wydawnictwo PWN, Warszawa 2000
- M.Nycz (red.): Generowanie wiedzy dla przedsiębiorstwa: metody i techniki. Wydawnictwo AE we Wrocławiu, Wrocław, 2004.
- L.Owoc: Elementy systemów ekspertowych, Wydawnictwo AE we Wrocławiu, Wrocław, 2006.
- Larose D., Odkrywanie wiedzy z danych. Wprowadzanie do eskploracji danych, PWN, 2006.
- Larose D., Metody i modele eksploracji danych, PWN 2008.
- Hand D., Mannila H., Smyth P. Eksploracja danych, WNT, 2005.
- Krawiec K, Stefanowski J., Uczenie maszynowe i sieci neuronowe, Wyd. PP, 2003.
- Cichosz P., Systemy uczące się. WNT, 2000.
- Lasek M., Data mining: Zastosowanie w ocenach i analizach klientów bankowych. Biblioteka Menadżera, 2003.

Jerzy.Stefanowski@cs.put.poznan.pl
http://www.cs.put.poznan.pl/jstefanowski

Dziękuję !