

marketing

Na potrzeby kampanii reklamowych marketerzy odnajdują potencjalnych klientów, którzy z największym prawdopodobieństwem odpowiedzą na ofertę i staną się klientami przynoszącymi zyski.

serwisy www

Personalizacja zawartości poprzez dostosowanie jej do oczekiwań odwiedzających. Prezentacja produktów i usług, które z największym prawdopodobieństwem zostaną zakupione, wyświetlanie informacji poszukiwanych przez odwiedzających strony.

wykrywanie nadużyć

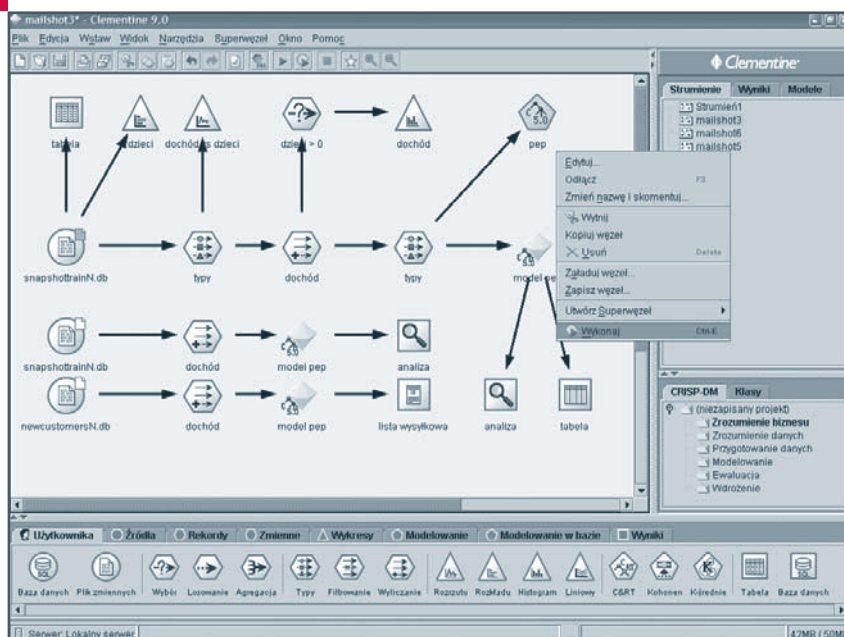
Klasyfikacja analizowanych przypadków i weryfikacja najbardziej „podejrzanych”, co gwarantuje najlepsze wykorzystanie dostępnych środków i zasobów.

nauka

Poszukiwanie odpowiedzi na pytanie o funkcję genów, ich wzajemne oddziaływanie i wpływ na występowanie chorób.

edukacja

Pogłębiona wiedza o wzorcach zachowań i tendencji wśród studentów. Pozwala stworzyć optymalną ofertę obejmującą doradztwo akademickie, pakiety pomocy finansowej i rozwój oferty programowej.



Rysunek 1.

Wizualny interfejs Clementine pozwala zobaczyć na ekranie ciąg podejmowanych działań: strumień danych i węzły.

Połączenie wiedzy biznesowej i posiadanych danych pozwala tworzyć i wdrażać skuteczne modele z pomocą Clementine

Dzięki rozwojowi technologii informatycznych mamy dostęp do ogromnych ilości danych. Paradoksalnie jednak, im więcej danych tym trudniej jest nam dotrzeć do istotnych informacji, których potrzebujemy. Rozwiązaniem tego paradoksu są dogłębna wiedza biznesowa i profesjonalne narzędzia analityczne, które pozwalają w oparciu o duże zbiory danych wydobywać informacje biznesowe, kluczowe dla sukcesu realizowanego przedsięwzięcia. Jednym z liderów wśród dostawców narzędzi drążenia danych (data mining) jest SPSS z popularnym wśród analityków na całym świecie systemem Clementine.

System w unikalny sposób wspiera cały proces drążenia danych. Umożliwia on odkrywanie wzorców występujących w danych, pozwalając następnie podejmować decyzje, które przyniosą mierzalne korzyści. Dzięki Clementine i wspierającym go innym programom SPSS, użytkownicy mogą w łatwy sposób używać danych przechowywanych w różnych źródłach i formatach. Posłużą one do budowania modeli, porównywania ich skuteczności, efektywnego wdrażania do pracy w czasie rzeczywistym oraz do tworzenia rekomendacji i wsparcia procesu podejmowania decyzji. Dzięki możliwości łatwej

integracji rezultatów drążenia danych z innymi wynikami, systemami i procesami biznesowymi, Clementine pomaga w podejmowaniu szybkich i lepszych decyzji w skali całego przedsiębiorstwa.

Praca z systemem Clementine jest oparta o przyjazny dla użytkownika interfejs. Dzięki niemu użytkownik może obserwować na ekranie ciąg podejmowanych działań: strumienie przepływu danych czy też mapy wizualne każdego etapu drążenia danych. Poprzez działania na strumieniach analityk ma wpływ na każdy etap procesu analitycznego, co pozwala w pełni wykorzystać wiedzę biznesową – klucz do sukcesu każdego projektu data mining. Przejrzysty interfejs Clementine pozwala skupić się bardziej na wiedzy uzyskiwanej dzięki wynikom, niż na technicznych aspektach projektu drążenia danych, dzięki czemu proces ten jest bardziej efektywny.

Współpracujący z programem Clementine system SPSS Model Manager pozwala podnieść poziom wiedzy biznesowej. Analitycy mogą zachować strumienie, modele i pliki wynikowe w centralnym repozytorium danych, gdzie dostęp do nich mają wszyscy uprawnieni użytkownicy wykorzystujący je w swojej pracy. Dzięki rozbudowanemu systemowi uprawnień administratorzy mogą w łatwy sposób nadawać i modyfikować uprawnienia dostępu do obiektów i wyników analiz.

Wybór spośród najszerszej liczby technik

System analityczny Clementine dysponuje szerokim spektrum technik umożliwiających uzyskanie użytecznych i rzetelnych rezultatów drążenia danych. Oferuje nieporównywalną z innymi platformami paletę technik statystycznych i sztucznej inteligencji. Dodatkowo trzydziestoletnie doświadczenie SPSS sprawia, iż algorytmy Clementine są dobrze zweryfikowane pod względem

przydatności do wspierania i tworzenia efektywnych modeli drążenia danych.

Integracja z istniejącymi systemami

Clementine jest otwartym, skalowalnym rozwiązaniem, które z łatwością integruje się z otoczeniem informatycznym przedsiębiorstw i instytucji. Dzięki temu może skutecznie dostarczać informacji wspierających proces decyzyjny na każdym szczeblu organizacji. Clementine, dzięki integracji z systemami IT firmy, umożliwia uzyskanie wyższego zwrotu z inwestycji w technologie, które już funkcjonują w firmie.

Skuteczne wdrażanie rezultatów data mining

SPSS oferuje wiele opcji wdrożeń pozwalających na przekazanie przewidywań modelu do systemów korporacyjnych czy umożliwiających scoring w czasie rzeczywistym. Clementine eksportuje nie tylko modele, ale także wszystkie kroki procesu drążenia danych – włączając w to dostęp do danych, modelowanie, post-processing. Oszczędza to czas potrzebny do zintegrowania z systemami przedsiębiorstwa. Dlatego pozytywne efekty wprowadzonych rozwiązań są widoczne szybciej.

Kluczowe cechy Clementine 9.0

Clementine w swojej najnowszej wersji ma do zaoferowania użytkownikom szeroką gamę technik analitycznych, działających w oparciu o otwartą architekturę, która pozwala na większą elastyczność w modelowaniu i drążeniu danych.

- Integracja z produktem SPSS Model Manager zapewnia scentralizowaną kontrolę nad efektami drążenia danych dla zwiększenia ich skuteczności;

- Ulepszenia w możliwościach drążenia baz danych i modelowaniu. Organizacja może szybko i skutecznie prowadzić projekt drążenia danych w oparciu o bazy danych IBM, Oracle czy Microsoft;
- Wsparcie procesu interaktywnego budowania i definiowania przez użytkownika kryteriów podziału do budowy drzew decyzyjnych. W systemie pojawiły się nowe algorytmy CHAID, Exhaustive CHAID, QUEST, CRT;
- Automatyczny podział zbioru na część ucząca, testową i walidacyjną. Istnieje też oczywiście możliwość podziału danych w sposób tradycyjny, poprzez selekcję rekordów;
- Zaawansowane opcje wizualizacji: zestaw węzłów CEMI dla obrazowania zbiorów danych i strumieni. Wiele opcji graficznych dostępnych w wersji 9 jest nowych i pozwala użytkownikom dostrzec takie aspekty danych, które są niedostępne za pomocą tabel i innych form prezentacji;
- Zintegrowane narzędzia text mining pozwalają organizacji docierać do wielu istotnych informacji zawartych w formie tekstowej.

Dzięki Clementine analitycy wykorzystujący techniki drążenia danych mogą skupić się na rozwiązywaniu problemów biznesowych, a nie na programowaniu i technicznych aspektach analizy. Na każdym kroku Clementine wspiera działania określone w metodologii CRISP-DM (CRoss-Industry Standard Process for Data Mining). Definiuje ona poszczególne etapy projektu data mining, których poprawne przeprowadzenie zapewni sukces wdrożenia. Projekty realizowane przy wsparciu Clementine mogą być efektywnie organizowane dzięki menedżerowi CRISP-DM dostępnemu w aplikacji, który umożliwia organizację danych, strumieni analitycznych i wyników według zasad tej metodologii. Po więcej szczegółowych informacji na temat metodologii CRISP-DM odsyłamy na nasze strony

internetowe http://www.spss.pl/konsulting/konsulting_datamining_metodologia.html.

Etapy drążenia danych w oparciu o metodologię CRISP-DM

Zrozumienie celów biznesowych

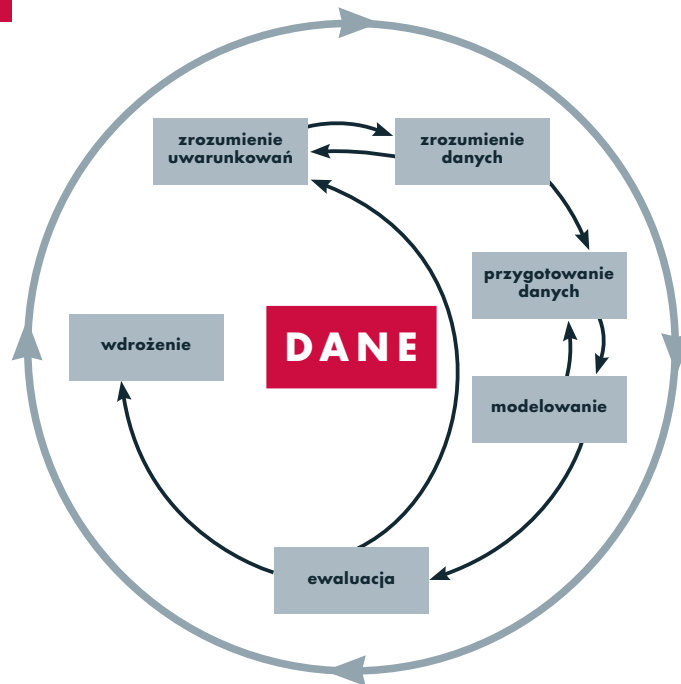
Pierwsza faza projektu data mining koncentruje się na zrozumieniu celów projektu i stawianych mu wymagań z perspektywy biznesowej, a następnie na przetworzeniu tej wiedzy w definicję problemu drążenia danych i stworzeniu planu działań, zmierzających do osiągnięcia celu.

Interfejs Clementine gwarantuje dobre wykorzystanie wiedzy biznesowej w projekcie drążenia danych. Użytkownik ma do dyspozycji szablony konkretnych zastosowań (Clementine Applications Template) – CAT, pozwalające na szybsze osiągnięcie celów biznesowych w konkretnych działaniach analitycznych. Wraz z Clementine 9.0 dostępne są następujące szablony:

- CRM CAT;
- Telco CAT;
- Fraud CAT;
- Microarray CAT;
- WebMining CAT (zastosowanie wymaga obecności WEB Mining for Clementine).

Zrozumienie danych

Faza zrozumienia danych rozpoczyna się od wstępnego zebrania danych i polega na zaznajomieniu się z danymi, rozpoznaniu problemów z jakością danych, dotarciu do pierwszych spostrzeżeń i wykryciu interesujących grup, pozwalających na skonstruowanie hipotez o kryjących się w danych informacjach. Na tym etapie Clementine 9.0 oferuje:



tez o kryjących się w danych informacjach. Na tym etapie Clementine 9.0 oferuje:

■ nowość!

Możliwość korzystania ze standardowych i z nowych typów wykresów, takich jak: wykresy słupkowe, wykres skrzynkowy, rozrzutu, wykresy panelowe, wykresy powiązań, mapy termiczne, wykresy współrzędnych równoległych, wykresy analizy powiązań. Poszerzają one funkcjonalność programu dzięki tzw. technologii CEMI (Clementine External Module Interface), która umożliwia podłączanie także zewnętrznych aplikacji, zawierających procedury przetwarzania danych, modelowania i prezentacji wyników;

- Klarowne „pierwsze” spojrzenie na dane dzięki węzłowi audytu danych;
- Wizualna interakcja z danymi: możliwość wyboru określonego obszaru

Rysunek 2.

Metodologia CRISP-DM umożliwia analitykom właściwe zaplanowanie procesu drążenia danych, dając pewność otrzymania planowanych efektów.

wykresu i dotarcia do wybranych informacji w tabeli lub użycia wybranych informacji do dalszych analiz;

- Tworzenie histogramów, wykresów dystrybucji, wykresów liniowych i punktowych;
- Wykrywanie asocjacji webowych;
- Wyświetlanie paneli 3D i wykresów animowanych;
- Szybki wgląd w dane dzięki wykresom, statystynom podsumowań i oszacowaniom jakości danych.

Przygotowanie danych

Faza ta obejmuje wszystkie czynności zmierzające do stworzenia ostatecznego zbioru danych (tj. danych, które będą wykorzystywane przez narzędzia modelujące) z danych surowych. Zadania te obejmują wybór źródeł danych, tabel

w bazie, cech, rekordów i atrybutów, jak również dokonanie przekształceń i wyczyszczenie danych przed ich wykorzystaniem w modelowaniu.

1. Dostęp do danych:

- Ustrukturalizowane dane w bazach danych i innych źródłach:
 - dostęp poprzez sterowniki ODBC, JDBC i inne zawarte w pakiecie SPSS Data Access Pack (dostępny razem z Clementine). Sterowniki dostępne dla IBM DB2, Oracle, Microsoft SQL Server, Informix, Sybase i innych źródeł danych;
 - import z plików tekstowych, plików w formacie SPSS i SAS;
- Nieustrukturalizowane dane tekstowe: automatyczna ekstrakcja pojęć z różnych dokumentów z użyciem modułu Text Mining for Clementine;
- Dane z WWW: automatyczna ekstrakcja danych z logów serwerów z użyciem modułu Web Mining for Clementine;
- Wyniki projektów data mining mogą być zapisywane również w różnych źródłach, takich jak bazy danych poprzez ODBC, pliki tekstowe, arkusze Microsoft Excel; pliki danych w formacie SPSS, SAS oraz we własnych plikach wyników systemu Clementine;

2. Wybór opcji czyszczenia danych, usuwanie lub zamiana niepoprawnych danych;

3. Automatyczne uzupełnianie braków danych;

4. Operacje na danych:

- **nowość!** podział danych na treningowe, testowe i walidacyjne, również praca z pełnym zakresem operacji na rekordach i polach;
- filtrowanie cech, rekodowanie, kategoryzacja wizualna, wyliczanie nowych zmiennych, zastępowanie wartości, wypełnianie

nowymi wartościami, reorganizacja danych i inne;

- selekcja rekordów, wybór próby, łączenie (poprzez sprzężenie wewnętrzne, pełne sprzężenie zewnętrzne, częściowe sprzężenie zewnętrzne i anty-sprzężenie), sortowanie, agregacja i ważenie;
- specjalne opcje ukazujące „historię” wartości, umożliwiające restrukturyzację i operacje przekształcające zmienne w zmienne typu flaga (np. binarne).

Modelowanie

W tej fazie następuje wybór i zastosowanie różnych technik, a także optymalizacja ich parametrów. Przeważnie z danym typem problemu data mining wiąże się kilka technik. Niektóre z nich mają specyficzne wymagania co do postaci danych. Clementine 9.0 pozwala na:

- Modelowanie bezpośrednie w bazach danych, dzięki integracji systemu Clementine z algorytmami zaimplementowanymi lub towarzyszącymi systemom baz danych:
 - IBM DB2 Enterprise Edition 8.2: DB2 IM Decision Tree, DB2 IM Association, DB2 IM Regression, DB2 IM Clustering;
 - Oracle 10g: Naive Bayes, Adaptive Bayes oraz Support Vector Machines (SVM);
 - Microsoft SQL Server & Analysis Services: MS Decision Tree;
- Techniki predykcyjne i klasyfikacyjne:
 - sieci neuronowe (wielowarstwowy perceptron, wsteczna propagacja błędu, RBF, sieci Kohonena);
- Przegląd ważności predyktorów:
 - drzewa decyzyjne i techniki indukcji reguł, w tym CHAID, Exhaustive CHAID, QUEST i C&RT;
 - Interaktywne tworzenie podziałów w drzewach decyzyjnych;
 - techniki indukcji reguł: wyszukiwanie, zwijanie i rozwijanie reguł decyzyjnych;

- regresja liniowa, logistyczna i wielomianowa regresja logistyczna: widok oszacowania modelu i zaawansowanych statystyk;
- Techniki skalowania i segmentacji:
 - sieci Kohonena, K-średnie, dwustopniowe grupowanie (TwoStep Cluster): charakterystyka skupień widoczna za pomocą przeglądarki graficznej;
- Wybór spośród kilku algorytmów asocjacyjnych:
 - algorytmy GRI, Apriori, sekwencyjne i CARMA:
 - scoring danych z użyciem modeli generowanych przez algorytmy wykrywania asocjacji;
 - filtrowanie, sortowanie i tworzenie podzbiorów modeli asocjacyjnych z poziomu podglądu modelu asocjacyjnego;
- Techniki redukcji danych:
 - analiza czynnikowa i analiza głównych składowych: widok równań modelu i zaawansowanych statystyk;
- Opcje łączenia modeli poprzez meta-modelowanie: modele mogą być łączone lub wykorzystywane w budowie następujących;
- Import modeli PMML, tworzonych za pomocą innych narzędzi, takich jak AnswerTree i SPSS;
- Użycie CEMI (Clementine External Module Interfaces) do podpięcia zewnętrznych algorytmów;
- Podręcznik użytkownika, który wyjaśnia teorię i metody stojące za algorytmami oferowanymi w Clementine.

Ewaluacja

Na tym etapie projektu mamy już stworzony model (lub wiele modeli), który posiada wysoką jakość z perspektywy analizy danych. Clementine 9.0 umożliwia :

- łatwą ewaluację modelu z użyciem wykresów korzyści, zysków i odpowiadających im kosztów;

- użycie jednostopniowego procesu, oszczędzającego czas w przypadku ewaluacji wielu modeli;
- definiowanie warunków trafień i wyrażen scoringowych dla oceny jakości modelu;
- Analizę całościowego dopasowania modelu poprzez macierz koincydencji i inne narzędzia automatycznej ewaluacji.

Wdrożenie

Stworzenie modelu zasadniczo nie kończy projektu. Zdobyta wiedza musi zostać uporządkowana i przedstawiona w taki sposób, aby można z niej było łatwo skorzystać. Polega to często na zastosowaniu „żywych” modeli w ramach procesu podejmowania decyzji w organizacji, np. na personalizowaniu stron WWW w czasie rzeczywistym (on-line) albo na tworzeniu powtarzanych scoringów marketingowych w bazach danych.

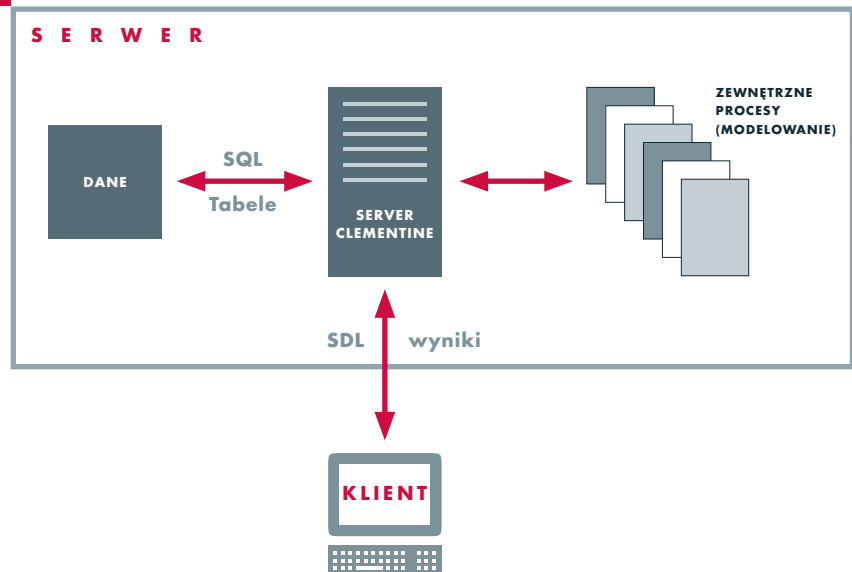
Clementine oferuje bogate możliwości wdrożeniowe, które odpowiadają na potrzeby każdej organizacji. Modele budowane w Clementine mogą być bezpośrednio stosowane w aplikacjach predykcyjnych SPSS, jak również w obrębie rozwiązań technologicznych innych dostawców.

Clementine Solution Publisher

Automatyczny eksport wszystkich operacji na danych, łącznie z dostępem do źródeł danych, przekształceniami danych, operacjami typu Text Mining, modelem scoringowym (w tym ich kombinacji) do wykonywalnego poza środowiskiem Clementine modelu.

PredictiveCallCenter, Predictive-Marketing

- Automatyczny eksport strumienia Clementine do użycia w środowisku Predictive Applications w celu dostarczania rekomendacji w czasie rzeczywistym;



Rysunek 3.

Skalowalność Clementine daje możliwość modelowania w bazie danych z wykorzystaniem technologii i algorytmów zaimplementowanych w zewnętrznych aplikacjach. Clementine przejmuje tylko potrzebne dane z burtowni danych i przesyła z powrotem jedynie istotne rezultaty.

- łączenie eksportowanych strumieni z modelami Predictive Applications (PredictiveCallCenter, PredictiveMarketing), regułami biznesowymi i regułami wykluczeń do optymalizacji interakcji z klientem.

Cleo

- Implementacja rozwiązań Webowych dla szybszego wdrożenia modelu;
- Umożliwienie wielu użytkownikom symultanicznego dostępu i ewaluacji pojedynczych rekordów lub całej bazy danych przez interfejs, oparty na bazie przeglądarki internetowej.

Clementine Batch

- Automatyczne realizowanie zadań obliczeniowych poza interfejsem aplikacji użytkownika;
- Automatyzacja procesów Clementine z wykorzystaniem innych aplikacji i systemów zadaniowych.

Skrypty

- Skrypty pomagają automatyzować powtarzalne działania, które musiałyby być w innym wypadku za każdym razem wykonywane przy pomocy klawiatury i myszki;
- Wykonanie wybranych linii poleceń z poziomu strumienia, Superwęzła lub pojedynczego skryptu za pomocą ikony na pasku zadań;
- Aktualizacja parametrów strumienia z pomocą skryptu;
- Export modeli jako PMML 2.1:
 - realizowanie scoringu wewnątrz baz danych, co eliminuje koszty ich transferu na maszynę klienta i konieczność przeprowadzania na niej analiz;
 - wdrożenie modeli Clementine PMML w narzędziach IBM DB2 Intelligent Miner Visualisation i Intelligent Miner Scoring.

SPSS Model Manager

■ nowość!

Centralizacja projektów drążenia danych pomaga wzbogacić wiedzę biznesową przedsiębiorstwa:

- składowanie strumieni, modeli i innych obiektów w centralnym repozytorium, które można następnie przeszukiwać;
- grupowanie strumieni analitycznych w folderach i ochrona folderów i strumieni;
- zapewnienie dostępu za zezwoleniem, aby chronić informacje poufne;
- wykorzystanie najbardziej efektywnych modeli i strumieni do ulepszenia procesów i zwiększenia dokładności rezultatów;
- przeszukiwanie zmiennych wejściowych, wyjściowych, typów modeli, notatek, słów kluczowych, autorów i innych typów metadanych;
- zapewnienie wiarygodnych wyników przez kontrolowanie wersji modeli predykcyjnych: automatyczne przypisanie wersji do strumieni i innych obiektów;
- ochrona strumieni przed nadpisaniem przez automatyczne przypisanie informacji o wersji.

Skalowalność

- Wykorzystanie kodu SQL w celu zlecenia transformacji w bazie danych;
- Możliwość modelowania w bazie danych z wykorzystaniem technologii i algorytmów zaimplementowanych w zewnętrznych aplikacjach bazodanowych;
- Minimalizowanie ruchu w sieci poprzez inteligentną współpracę z systemem bazodanowym. Clementine przejmuje z hurtowni danych tylko potrzebne dane i przesyła z powrotem jedynie istotne wyniki.

Aby uzyskać dodatkowe informacje na temat Clementine i drążenia danych zapraszamy na stronę www.webmining.pl oraz do bezpośredniego kontaktu z konsultantami SPSS.

SPSS Polska

ul. Raclawicka 58
30-017 Kraków
tel./faks 012.636.96.80
tel./faks 012.636.07.91
tel./faks 012.636.45.35
e-mail: info@spss.pl
www.spss.pl
www.analizadanych.pl
www.webmining.pl