

# GhostMiner

## D A T A M I N I N G S O F T W A R E

### Informacje o produkcie

**GhostMiner** to zaawansowane narzędzie firmy Fujitsu do analitycznej eksploracji danych, które nie tylko wspiera rozmaite bazy danych (oraz arkusze kalkulacyjne), zaawansowane algorytmy uczenia maszynowego, ale także przygotowanie i selekcję danych, walidację modeli, multimodele takie jak komitety lub k-klasyfikatory i wreszcie wizualizację.

Wszystko to jest dostępne w jednym pakiecie – **szeroki zakres przygotowania danych, selekcji cech, algorytmów eksploracji danych oraz technik wizualizacyjnych jest w pełni zintegrowany**. Oznacza to że wymagany jest tylko jeden format danych (projekt) zaś stosowanie i porównywanie różnych metod jest wyjątkowo proste. Pakiet posiada również **intuicyjny interfejs**, pozwalający na łatwą obsługę programu nawet mało doświadczonemu użytkownikowi.

Dzięki umiejętności wydobywania wiedzy z nagromadzonych danych **GhostMiner** pomaga w podjęciu trafniejszych decyzji we właściwym czasie i we właściwym miejscu, zapewniając menadżerom uzyskanie przewagi nad konkurencją.

Przykładowo **GhostMiner** skutecznie pomaga w rozwiązywaniu następujących problemów:

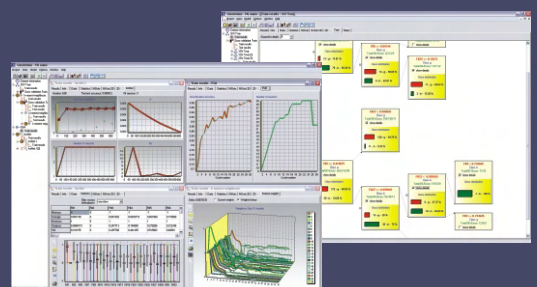
- Segmentacja klientów na grupy w celach marketingowych
- Tworzenie profilu klientów w celu indywidualnego zarządzania kontaktami
- Poprawa wskaźnika odpowiedzi na pocztowe kampanie reklamowe
- Wykrywanie „dezerterów” wśród klientów
- Wyszukiwanie klientów, którzy są dobrymi potencjalnymi nabywcami dodatkowych usług
- Analiza przyczyn odchodzenia klientów do konkurencji
- Analiza wniosków kredytowych
- Przewidywanie nieprawidłowych zachowań lub transakcji
- Wybór najlepszego kandydata do pracy

Program jest przeznaczony dla następujących sektorów gospodarki:

- Bankowość
- Telekomunikacja
- Handel elektroniczny
- Dystrybucja
- Usługi zarządzania zasobami

Dzięki swej uniwersalności **GhostMiner** jest również przydatny w wielu innych dziedzinach niezwiązanych z biznesem, takich jak:

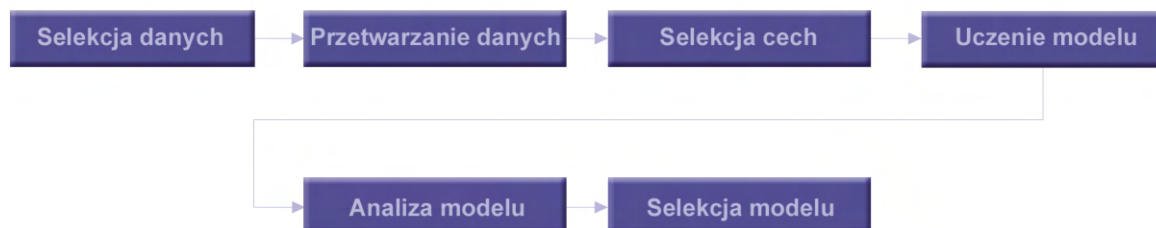
- Bioinformatyka - genetyka, proteomika, QSAR
- Medycyna - kliniczne i laboratoryjne systemy diagnostyczne, farmakologia, aktywność biologiczna leków, analiza sygnałów i obrazów (za pomocą dodatkowych filtrów do przetwarzania obrazów/sygnałów)
- Psychologia i psychiatria - obsługa analizy testów psychometrycznych i innych testów opierających się na kwestionariuszach
- Zastosowania techniczne - diagnostyka silników, analiza procesów technologicznych
- Szkolnictwo
- Zawsze i wszędzie tam, gdzie analizie trzeba poddać duże bazy danych



**GhostMiner** składa się z dwu narzędzi: **Developer** oraz **Analyzer**

**GhostMiner Developer** jest przeznaczony dla projektantów modeli danych, którzy korzystając z próbnych baz danych mogą wytrenować oraz przetestować i wybrać użyteczne modele. Do korzystania z tego programu potrzebna jest dobra znajomość analizy statystycznej i pewna wiedza z zakresu metod inteligencji obliczeniowej.

Modele danych są budowane jako projekty modułu **Developer**. Każdy projekt wiąże się z analizą danych typowych dla konkretnego problemu. Analiza ta rozpoczyna się od **danych w formie nieprzetworzonej** i obejmuje następujące etapy:



Końcowym produktem modułu **Developer** jest model wiedzy właściwy dla rozpatrywanych danych. Model ten można wykorzystać jako element pomocniczy w procesach decyzyjnych.

Z myślą o uproszczeniu tych procesów opracowano program **GhostMiner Analyzer**, który pozwala osobom podejmującym decyzje analizować nowe przypadki w miarę rozrastania się bazy danych.

**GhostMiner Analyzer** jest przeznaczony dla użytkowników, niebędących ekspertami w zakresie metod statystycznych, którzy do analizowania własnych danych używają modeli wykreowanych przez Developera i opracowanych przez ekspertów. Umożliwia on ocenę i wizualizację pojedynczych przypadków lub zbiorów przypadków w odniesieniu do innych, wzorcowych przypadków z bazy danych. Program Analyzer wyświetla informacje o modelu i o danych, dla których model został przeszkolony i udostępnia część narzędzi, służących w programie Developer do tworzenia modeli danych. Ponadto Analyzer dzięki swojej nieskomplikowanej strukturze może zostać łatwo dostosowany do potrzeb konkretnej aplikacji biznesowej.

## Funkcjonalność

- Algorytmy adaptacyjne
- **Drzewo decyzji SSV** – drzewo binarne oparte na Kryterium Separowalności które może być zastosowane zarówno do zmiennych numerycznych jak i kategoriowych
- **Inkrementacyjne sieci neuronowe** – zdolne do dopasowania swojej architektury do złożonej struktury danych, oparte na Radialnych Funkcjach Bazowych oraz z Gaussowską i Bi-centralną Funkcją Transferu
- **System neurorozmyty** – Odwzorowanie Przestrzeni Cech sieć neurorozmyta oparta o wielowymiarowe funkcje separowalne mogące efektywnie pracować z tysiącami cech
- **K najbliższych sąsiadów** – z różnymi metrykami, wylicza nie tylko optymalne “K” ale także wagi cech
- **Metoda wektorów wspierających** do klasyfikacji – z automatycznym (zoptymalizowanym) doбором stosunku C/Bias
- **Metoda wektorów wspierających dla klasteryzacji**
- **Dendrogramy** – wizualizują proces tworzenia klastru (klasteryzacja hierarchiczna) oraz przypisują poszczególne rekordy do odpowiednich klastrów
- Multimodele
- Model **Transformuj & Klasyfikuj** – kombinacja transformacji (selekcja cech) oraz klasyfikacji umożliwiająca walidację modelu na pełnym zakresie atrybutów
- **Komitet** modeli – wykorzystujący głosowanie pojedynczych modeli w podjęciu ostatecznej decyzji dotyczącej niejednoznacznych przypadków
- **K-klasyfikatory** – dzielą jedno zadanie klasyfikacyjne na K zadań klasyfikacyjnych dla każdej klasy zbioru treningowego w oparciu o kryterium przynależności
- Techniki wizualizacyjne
- **Skalowanie wielowymiarowe** – redukuje oryginalny rozmiar bazy danych z zachowaniem zarówno własności topologicznych jak i metrycznych danych
- **Analiza składowych głównych** – metoda liniowa pozwalająca na redukcję oryginalnego wymiaru bazy danych
- Metody testowania i walidacji
- Test na danych zewnętrznych
- Kroswalidacja z różnymi metodami próbkowania

- Metoda "Leave-one-out"
- Rozszerzona krosvalidacja (X-Test)
- Macierz konfuzji
- Metody selekcji cech
  - Ręczna selekcja cech umożliwiająca wykluczenie dowolnej liczby cech ze zbioru danych
  - Automatyczna metoda współczynników korelacji szybka i efektywna, niezależna od modelu klasyfikacyjnego
  - Automatyczna metoda nakładkowa związana z klasyfikatorami (dla dowolnego zaimplementowanego algorytmu adaptacyjnego o których mowa powyżej)
  - Komitety selektorów oparte o głosowanie pojedynczych selektorów
- Statystyka
  - Informacje statystyczne oraz wykresy wizualizujące dane
  - Narzędzia do numerycznej i graficznej prezentacji wartości średniej, odchylenia standardowego, minimalnej/maksymalnej wartości oraz wartości brakujących
  - Dwu- i trzowymiarowe wykresy do identyfikacji potencjalnie użytecznych cech
  - Dwuwymiarowe rzutowania do obserwacji danych dla dowolnej kombinacji cech

## Dodatkowe usługi

Wychodząc z założenia, iż dla klienta oprócz pozyskania wartościowego produktu ważna jest również pomoc techniczna oraz udzielenie fachowych konsultacji, FQS Poland oferuje wszechstronne wsparcie techniczne w formie dostosowanej do indywidualnych potrzeb klienta. Nasi specjaliści pomogą klientom począwszy od udzielenia niezbędnych informacji telefonicznie lub pocztą elektroniczną, poprzez prowadzenie kursów, aż do wykonania pełnej implementacji określonej aplikacji.

## Co potrafi GhostMiner

**GhostMiner** potrafi wydobyć z danych wiele przydatnych informacji niewidocznych w inny sposób, a ponadto umożliwia:

### GhostMiner Developer

- Przeglądanie liczbowych i numerycznych zbiorów danych, wstępne przetwarzanie danych (normalizacja lub standaryzacja), prezentacja danych w postaci wykresów i reasumowanie ich właściwości statystycznych
- Przypisywanie nowych przypadków do wcześniej zdefiniowanych kategorii z zastosowaniem kryteriów kowariancji wielu zmiennych, dla potrzeb wspierania procesów decyzyjnych, diagnostyki, klasyfikacji, oceny i heteroasocjacji dwóch zbiorów obiektów
- Tworzenie „przećwiczonych” predykcyjnych modeli typu neuronowego, „neuro-rozmytego”, drzewa decyzyjnego i najbliższego sąsiada za pomocą procedury krosvalidacji w celu uzyskania optymalnego uogólnienia
- Odkrywanie wiedzy w postaci drzew decyzyjnych lub reguł logicznych, zarówno klasycznych jak i rozmytych
- Poprawa predykcyjnej precyzji modeli danych z zastosowaniem „komisji” modeli i tworzenia K-klasyfikatorów, po jednym dla każdej klasy
- Wyszukiwanie brakujących wartości w danych i uzupełnianie wzorców z brakującymi elementami
- W przypadku reguł rozmytych automatyczna dyskretyzacja zmiennych wejściowych
- Dobór zmiennych istotnych dla danej decyzji lub przypisywanie wagi zmiennym wejściowym, odpowiednio do ich ważności
- Wprowadzanie ciągłych dyskretnych danych wejściowych z zastosowaniem przedziałów użytecznych przy symbolicznym opisie danych
- Wizualizacja istotnych zależności w danych, ukazująca te dane w odniesieniu do innych znanych przypadków zarejestrowanych w bazie danych
- Redukcja rozmiaru bazy danych, z zachowaniem ich właściwości predykcyjnych
- Znajdowanie klastrow danych o podobnych właściwościach i przypisywanie każdego rekordu do odpowiedniego klastra

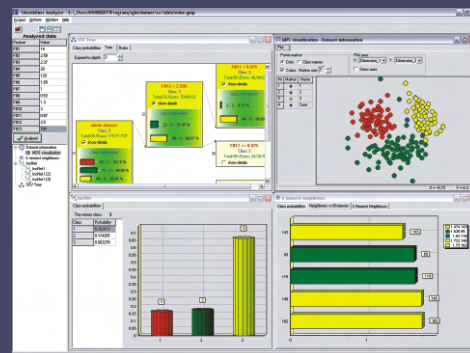
### GhostMiner Analyzer

- Obliczanie prawdopodobieństwa przyporządkowania nowych przypadków do istniejących kategorii





- Wyszukiwanie uzasadnień takich przyporządkowań z wykorzystaniem modeli zawierających reguły logiczne
  - Prezentacja podobnych przypadków z zastosowaniem interaktywnej wizualizacji, ułatwiająca wynajdywanie innych możliwości, które często mogłyby zostać pominięte
  - Wyszukiwanie przypadków, które są najbardziej do siebie podobne pod względem zmiennych istotnych w danej kategorii
  - Korzystanie z projektów bazujących na „Komitetach” modeli w celu analizy przypadków niejednoznacznych
- Klasyfikowanie nowych zbiorów danych w oparciu o wytrenowane modele



## Wymagania techniczne

Program **GhostMiner** pracuje na komputerach klasy PC pod następującymi systemami operacyjnymi:

- Windows 2000 / Windows XP

**Wymagania minimalne:** dowolny komputer z procesorem Pentium lub nowszym, pamięcią operacyjną 64 MB RAM oraz 50 MB wolnego miejsca na dysku.

**Zalecana konfiguracja:** zależnie od zastosowania; w przypadku dużych baz danych, pomimo zastosowania efektywnych algorytmów, mogą być wymagane szybkie komputery z dużą ilością pamięci RAM. Ilość wymaganej pamięci RAM jest wprost proporcjonalna do wielkości analizowanej bazy danych. Nie istnieje określone ograniczenie, co do liczby elementów lub przypadków w bazie danych. Oprogramowanie działa w trybie wielowątkowym, co umożliwia równoległe przetwarzanie wielu modeli.

**Dostęp do danych:** GhostMiner rozpoznaje i obsługuje następujące formaty danych:

- pliki tekstowe ASCII (w tym dowolne pliki typu CSV)
- arkusz kalkulacyjny Excel
- dowolne bazy danych zgodne ze standardem ODBC, w tym Microsoft Access, SQL Server oraz Oracle
- dowolne bazy danych zgodne ze standardem OLE DB, w tym SQL Server, Informix, Oracle oraz wiele innych

Aby uzyskać więcej informacji zapraszamy do odwiedzenia naszych stron internetowych pod adresem: <http://www.fqs.pl/GM>

FQS Poland Sp. z o.o.  
Fujitsu Company  
Pałac Pugetów, ul. Starowiślna 13-15  
31-038 Kraków

Tel. (+48 12) 429 43 45  
Fax (+48 12) 429 61 24  
Email: [gm@fqs.pl](mailto:gm@fqs.pl)  
<http://www.fqs.pl>