



ROZWIĄZANIA *STATISTICA* DLA SKORINGU

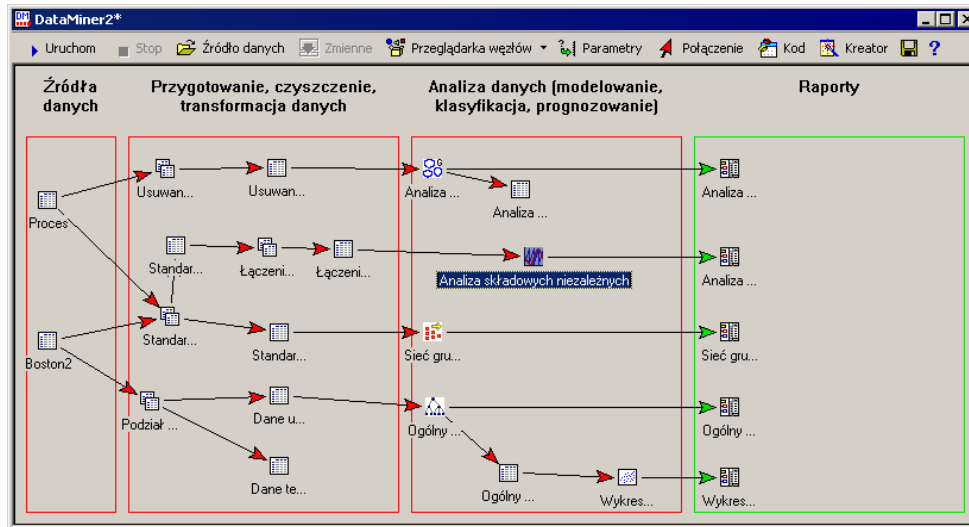
Piotr Wójtowicz
StatSoft Polska Sp. z o.o.

Skuteczność działania systemu skoringowego przede wszystkim zależy od dostępności, zakresu i jakości gromadzonych danych. Dużą rolę odgrywa także jakość i aktualność modeli skoringowych, czyli te aspekty, które zazwyczaj są najslabszą stroną generycznych modeli. Jest to ważny argument za korzystaniem z rozwiązań informatycznych, które zapewnią szybką aktualizację modeli z wykorzystaniem różnorodnych algorytmów analitycznych. Rozwiązań, które nie tylko zapewnią wysoką jakość otrzymanych modeli, ale również pozwolą na ich budowę w możliwie krótkim czasie, pozwalając na częstą aktualizację i budowę wielu modeli dla właściwie posegmentowanych portfeli. Inwestycja w odpowiednie rozwiązania analityczne oraz wiedzę pozwalającą na efektywne ich wykorzystanie to potencjał, który nierzadko pozwala na uzyskanie przewagi konkurencyjnej.

Systemy *STATISTICA*

Takim rozwiązaniem są systemy analityczne *STATISTICA*. Umożliwiają one sprawną i efektywną budowę, eksploatację i walidację modeli skoringowych (i nie tylko), w szczególności w postaci kart skoringowych. Analitycy korzystający w swojej pracy ze *STATISTICA* mają możliwość pełnej kontroli nad modelami (kartami) skoringowymi, odświeżania ich w dowolnym momencie i szybkiego reagowania na zmiany na rynku, a także budowy odrębnych modeli skoringowych dla poszczególnych grup klientów i produktów.

Zestaw Skoringowy STATISTICA pozwala łatwo zamienić dane w gotową kartę skoringową. Co ważne - do jej zbudowania nie jest wymagana specjalistyczna wiedza - interfejs został zaprojektowany tak, aby poprowadzić użytkownika krok po kroku przez poszczególne etapy, a same analizy mogą być wykonywane automatycznie. Dzięki temu analitycy mają możliwość pełnej kontroli nad kartami skoringowymi, odświeżania ich w dowolnym momencie i szybkiego reagowania na zmiany na rynku.



Projekt data mining w *STATISTICA*

W ramach jednego środowiska aplikacji analitycy mogą przeprowadzić cały proces budowy modeli – od pobrania danych historycznych i ich wstępną analizę, po właściwy model i końcowy raport. Za pomocą modułów dedykowanych do budowy modeli skoringowych analityk w wygodny sposób może dokonać dyskretyzacji bądź kategoryzacji predyktorów na wcześniej zdefiniowane przedziały lub z wykorzystaniem zaawansowanych metod analitycznych (np. algorytmu CHAID) oraz dokonać oceny jakości podziału i mocy predykcyjnej przekształconej zmiennej.

Budowa modeli

Budowa modelu skoringowego może przebiegać w sposób automatyczny, ale zaawansowani użytkownicy mają możliwość wyboru trybu eksperckiego, w którym mogą modyfikować wszystkie parametry modelu.

Oprócz najczęściej stosowanego podejścia – budowy tablic skoringowych na podstawie modelu regresji logistycznej – można także budować modele skoringowe typu SURVIVAL (z wykorzystaniem analizy przeżycia za pomocą proporcjonalnego hazardu Coxa), które pozwalają nie tylko określić prawdopodobieństwo zajścia danego zdarzenia, ale również czas, w jakim osiągnie ono dany (krytyczny) poziom – np. klient przestanie spłacać kredyt.

To najczęściej stosowane techniki budowy modeli skoringowych, ale to tylko fragment możliwości analitycznych – *STATISTICA* oferuje szeroki zakres zaawansowanych narzędzi do budowy modeli liniowych i nieliniowych, umożliwiających zarówno wykorzystanie predyktorów ciągłych, jak i jakościowych, także z uwzględnianiem interakcji. Ponadto zawiera bogaty zestaw technik eksploracji danych różnego typu, począwszy od analizy skupień, a skończywszy na metodach data mining, np. drzewach decyzyjnych czy sieciach neuronowych.



STATISTICA Data Miner oferuje następujące techniki analizy danych:

	Statystyki podstawowe i tabele
	Dopasowanie rozkładów
	Regresja wieloraka
	Analiza wariancji (ANOVA)
	Statystyki nieparametryczne
	Rozkłady i symulacje
	Ogólne modele liniowe
	Uogólnione modele liniowe i nieliniowe
	Ogólne modele regresji
	Modele cząstkowych najmniejszych kwadratów
	Komponenty wariacyjne
	Analiza przeżycia
	Estymacja nieliniowa
	Linearyzowana regresja nieliniowa
	Analiza log-liniowa tabel liczości
	Szeregi czasowe i prognozowanie
	Modelowanie równań strukturalnych
	Analiza skupień
	Analiza czynnikowa
	Składowe główne i klasyfikacja
	Analiza kanoniczna
	Analiza rzetelności i pozycji
	Drzewa klasyfikacyjne
	Analiza korespondencji
	Skalowanie wielowymiarowe
	Analiza dyskryminacyjna
	Ogólne modele analizy dyskryminacyjnej
	Analiza Mocy Testów
	Ogólne modele drzew klasyfikacyjnych i regresyjnych (<i>G-Trees</i>)
	Ogólne modele CHAID
	Drzewa interakcyjne (<i>C&RT</i> , <i>CHAID</i>)
	Drzewa wzmacniane (<i>Boosted Trees</i>)
	Automatyczne sieci neuronowe
	Losowy las
	Analiza składowych niezależnych
	Analiza skupień uogólnioną metodą EM i k-średnich
	Analiza koszykowa
	Uogólnione modele addytywne
	MARSplines
	Inne metody uczenia maszyn



Dane

Dla sprawnej realizacji zadań analitycznych niezbędny jest wygodny dostęp do danych zgromadzonych w różnego rodzaju źródłach: systemach CRM, systemach transakcyjnych czy plikach Excela. *STATISTICA Query* (komponent zawarty w każdym systemie *STATISTICA*) umożliwia dostęp do danych gromadzonych w różnego typu relacyjnych bazach danych za pomocą standardu OLE DB i standardu ODBC (np. Oracle czy MS SQL Serwer). Za pomocą *STATISTICA Query* w graficznym środowisku użytkownika można interakcyjnie budować zapytania do baz danych i określać dla nich kryteria.

Ponadto *STATISTICA* czyta i zapisuje dane w bardzo wielu formatach plików (poza własnym formatem plików danych i wyników także MS Excel, SAS, dBase, HTML, pliki tekstowe z separatorami i o stałej szerokości pola itp). Do wymiany danych można też wykorzystywać standardy OLE DB i ODBC, aby pobrać dane z dowolnej bazy danych. Można także przygotować własną procedurę *STATISTICA Visual Basic* i wykorzystać ją jako źródło danych dla systemu - w ten sposób można obsłużyć dowolny format plików.

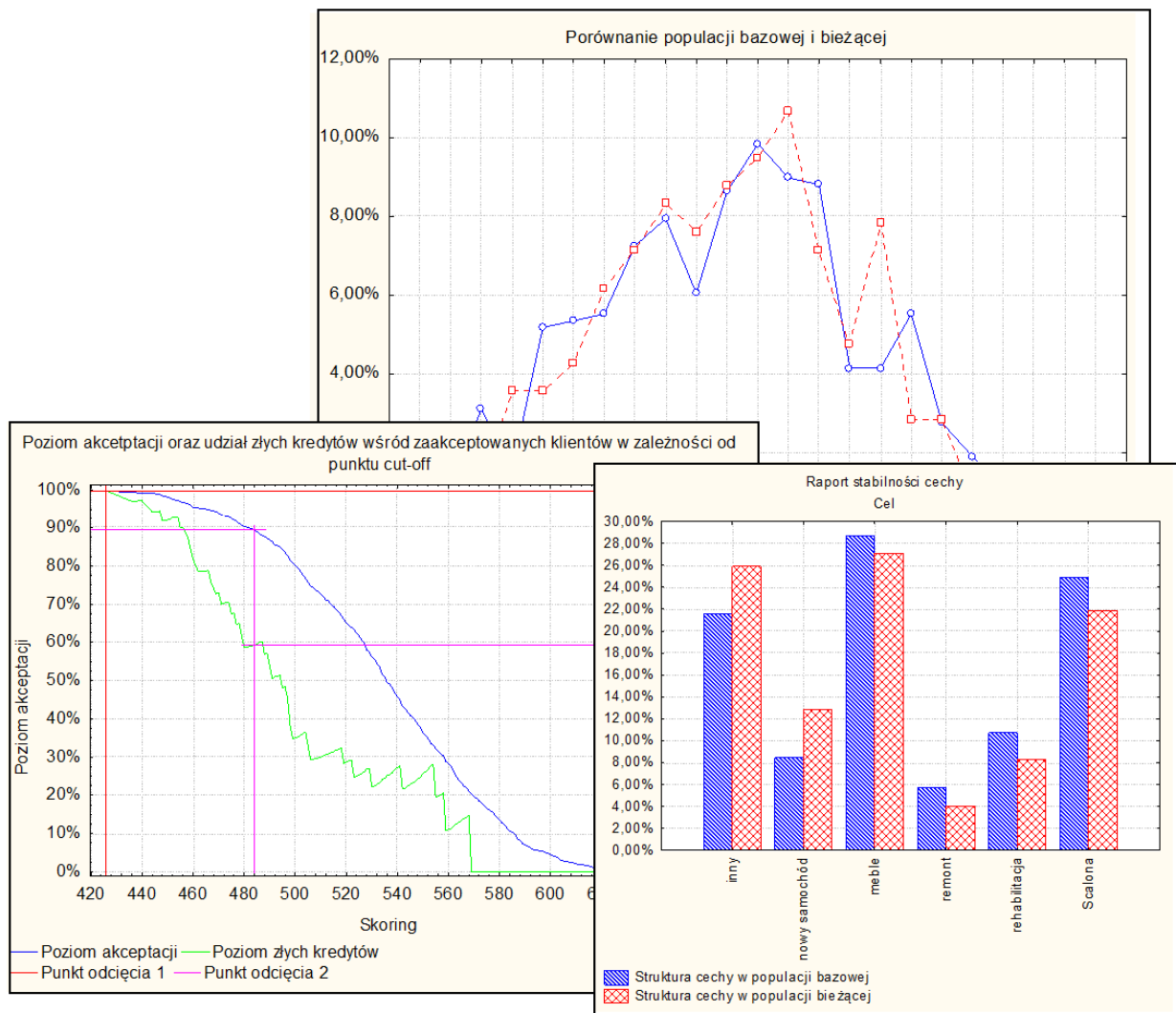
Integracja z innymi systemami

STATISTICA może pracować zarówno jako samodzielna aplikacja analityczna, jak i jako składnik większego systemu przetwarzania danych. Po opracowaniu gotowych modeli *STATISTICA Data Miner* oferuje elastyczne narzędzia do wdrażania i stosowania modeli, zintegrowane ze środowiskiem tworzenia projektów. Dzięki otwartej architekturze *STATISTICA* można bardzo łatwo wbudować w istniejącą strukturę informatyczną lub stworzyć nowy system informatyczny, którego jądrem będzie właśnie *STATISTICA*. *STATISTICA Data Miner* może, pracując w tle, generować prognozy i umieszczać je w bazie danych, w systemie CRM lub w Intranecie organizacji. *STATISTICA* oferuje dedykowane rozwiązania, które automatycznie wyznaczają skoringi dla modeli zbudowanych w *STATISTICA Data Minerze*, z wykorzystaniem np. webservice'ów. Pozwala to wykorzystać silnik skoringowy *STATISTICA* w ramach odpowiednich procesów biznesowych on-line.

Opcjonalnie można dla zbudowanych modeli analitycznych generować kody w językach C i Visual Basic, które pozwalają na korzystanie z tak utworzonych modeli niezależnie od platformy *STATISTICA*.

Wizualizacja

STATISTICA od lat jest bezkonkurencyjna wśród aplikacji do analizy danych pod względem narzędzi wizualizacji danych i możliwości tworzenia wykresów. Oferuje setki typów wykresów, które można osadzać w postaci aktywnych obiektów (edytowalnych i automatycznie aktualizowanych przy każdej modyfikacji danych) w innych aplikacjach np. w raportach MS Word. Praktycznie każdy szczegół wyglądu wykresu (a są to setki cech) może być ustawiany przez użytkownika.

Graficzna prezentacja stabilności i doboru *cut-off* w *STATISTICA*

Obszary wykorzystania *STATISTICA* w zarządzaniu ryzykiem

Działalność instytucji finansowych jest w szczególności związana z identyfikacją ryzyka i oceną zagrożeń, jakie ze sobą niesie. Wprowadzenie odpowiednich procedur zarządzania ryzykiem wymaga prawidłowej oceny jego wielkości i źródeł, a do tego potrzebne są precyzyjne informacje na temat zachowań klientów w przeszłości oraz prawdopodobnych scenariuszy w przyszłości. Informacji tego typu dostarcza analiza danych, w szczególności analiza statystyczna i data mining.

Karty skoringowe

Wykorzystanie modeli i kart skoringowych pozwala szybko przyznawać kredyty według obiektywnych kryteriów, zapewniając wyższe bezpieczeństwo. Stosowane jeszcze do



niedawna eksperckie i generyczne karty skoringowe są zastępowane przez nowoczesne systemy oparte na kartach skoringowych budowanych na podstawie własnego portfela kredytów.

STATISTICA Zestaw Skoringowy umożliwia szybką i wygodną budowę i walidację modeli skoringowych oceniających wiarygodność kredytową klientów, zarówno na podstawie danych z wniosku kredytowego (skoring aplikacyjny), jak również na podstawie analizy zachowań klienta (skoring behawioralny). *STATISTICA* wspiera proces konstrukcji kart skoringowych, począwszy od wyboru i przygotowania cech, aż po walidację zbudowanych kart. Karty te mogą być wykorzystane w ramach środowiska *STATISTICA*, jak również zewnętrznych silników skoringowych.

Wykorzystanie takiego podejścia pozwala na:

- ◆ zwiększenie skuteczności weryfikacji kredytobiorców, dzięki zastosowaniu modeli analitycznych,
- ◆ obniżenie kosztów weryfikacji wniosków kredytowych,
- ◆ zobiektywizowanie kryteriów przyznawania kredytów,
- ◆ wyeliminowanie błędów związanych z występowaniem czynnika ludzkiego.

Inne modele predykcyjne

STATISTICA pozwala na budowę różnorodnych modeli nie tylko w postaci kart skoringowych. Mogą to być modele predykcyjne zbudowane z wykorzystaniem całego spektrum metod analitycznych, np. drzewa decyzyjne, sieci neuronowe, modele regresji itp. Pozwalają one na wczesne wykrywanie i przeciwdziałanie nadużyciom finansowym, wykrycie podejrzanych operacji, zidentyfikowanie rachunków, które służą do „prania brudnych pieniędzy”, wyboru ścieżki windykacji, ale także wspierają ocenę ryzyka kredytowego. *STATISTICA* umożliwia wywoływanie tego typu modeli również spoza środowiska *STATISTICA*.

Analizy rozkładów

Ważnym elementem w analizie ryzyka jest identyfikacja jego składników. *STATISTICA* dostarcza wielu metod wizualizacji danych ułatwiających ekspercką ocenę rozkładów zmiennych i wektorów losowych, a wykresy skategoryzowane ułatwiają analizę tego samego zjawiska w podziale na klasy. *STATISTICA* oferuje także szereg bardziej formalnych narzędzi do badania rozkładów oraz ich wzajemnej odległości. Dzięki temu możliwa jest ocena ryzyka, jego źródeł, częstości występowania i wysokości potencjalnych szkód.

Symulacje i kalkulacja rezerw

Istotnym elementem każdej analizy ryzyka są również odpowiednie symulacje oraz opracowanie właściwych scenariuszy *what-if*. *STATISTICA* umożliwia przeprowadzenie wielu scenariuszy losowania i generowania losowych wartości o zadanych rozkładach. Dedykowany moduł do dopasowania rozkładów i symulacji umożliwia także zachowanie



odpowiedniej współzależności zmiennych (struktury korelacji) i uwzględnienie jej przy symulacjach. Generator liczb losowych *STATISTICA* należy do wąskiego grona tych, które zdają egzamin jakości w postaci zestawu testów DIEHARD.

Rodzaje zadań

Poniżej zostały zebrane zadania, do których najczęściej wykorzystywane są systemy analityczne *STATISTICA*:

- ◆ Analiza portfela kredytów
- ◆ Analiza sekwencji działań
- ◆ Dopasowanie rozkładów
- ◆ Dopasowanie składki do ryzyka
- ◆ Identyfikacja składników ryzyka
- ◆ Ocena wartości narażonej na ryzyko (VaR)
- ◆ Optymalizacja ścieżki windykacji
- ◆ Przewidywanie skuteczności windykacji
- ◆ Schematy losowania
- ◆ Segmentacja pod względem ryzyka
- ◆ Skoring kredytowy - aplikacyjny i behawioralny
- ◆ Skoring windykacyjny
- ◆ Skoring wyłudzeń
- ◆ Stress testy z wykorzystaniem scenariuszy *what-if*
- ◆ Symulacje
- ◆ Walidacja i raportowanie
- ◆ Wycena portfela należności
- ◆ Wykrywanie nadużyć (*fraud detection*)
- ◆ Wykrywanie nieuzasadnionych roszczeń
- ◆ Wykrywanie prania brudnych pieniędzy (AML)
- ◆ Wyszukiwanie wzorców nietypowych zachowań

Oczywiście nie wyczerpuje to katalogu zastosowań *STATISTICA* w obszarach zarządzania ryzykiem w instytucjach finansowych.

STATISTICA w ryzyku wśród praktyków

Powyższe cechy systemów *STATISTICA* powodują, że analitycy chętnie z nich korzystają, dzięki czemu szybciej i efektywniej mogą budować modele i realizować zadania analityczne, a *STATISTICA* jest powszechnie wykorzystywana w polskich instytucjach



finansowych. W gronie naszych klientów jest wiele polskich banków i instytucji finansowych: np. Komisja Nadzoru Finansowego, Narodowy Bank Polski, PKO BP, BRE Bank, GE Money Bank, Bank Millennium, Fortis Bank, Bank Pekao, Bank BPH, ING Bank Śląski, Volkswagen Bank Polska, Bank Ochrony Środowiska, Bank Gospodarstwa Krajowego, Bank Gospodarki Żywnościowej, Europejski Fundusz Leasingowy, Master-lease, Raiffeisen Bank, AIG Bank, Invest-Bank, Kredyt Bank, FM Bank, Bank Handlowy, Bank Polskiej Spółdzielczości, SKOK im. Stefczyka, SKOK im. Kopernika, SKOK Piast, SKOK Rafineria, Grupa PZU, Commercial Union Polska, TUiR Warta, UNIQA, STU ERGO HESTIA, AXA Życie, Generali, GPPI, Kredyt Inkaso, Krajowa Izba Rozliczeniowa, Krajowy Depozyt Papierów Wartościowych.

Poniżej znajduje się kilka wybranych opinii użytkowników *STATISTICA* o wykorzystaniu systemu w codziennej pracy:

Wysoko oceniamy użyteczność STATISTICA Zestaw Skoringowy w rozwijaniu metod modelowania w codziennym zarządzaniu ryzykiem kredytowym.

Bank Millennium

Korzystanie ze wspomnianego oprogramowania znacząco skróciło czas pracy nad tworzeniem kart scoringowych, okresowej ich walidacji i raportowania oraz podniosło wydatnie jakość tych prac.

SKOK Stefczyka

Zestaw Skoringowy STATISTICA jest programem, który pozwala nie tylko na automatyzację poszczególnych procedur prowadzących do zbudowania modelu skoringowego, ale również wspomaga raportowanie całego procesu, co jest szczególnie istotne w każdej instytucji finansowej.

SKOK im. M. Kopernika