



Zaawansowane techniki informatyczne wspomagające podejmowanie decyzji na podstawie analizy zgromadzonych danych

Grzegorz Migut

Analiza danych i CRM – przegląd zastosowań

Każda organizacja jest w olbrzymim stopniu zależna od oddziałujących na nią czynników zewnętrznych. Od umiejętności radzenia sobie z tymi czynnikami zależy, czy odniesie sukces czy też poniesie porażkę. Do czynników mających największy wpływ na obraz rynku należy zaliczyć ogromne zróżnicowanie produktów oraz „szum informacyjny” wywołany niespotykanym wcześniej natężeniem aktywności reklamowej.

Dodatkowo zaobserwować można wzrost przeciętnego rozporządzalnego dochodu klientów. Czynniki te powodują obniżenie poziomu lojalności klientów oraz występowanie efektów częstego przesuwania preferencji. Zjawisko to nasila się jeszcze pod wpływem rewolucji technologicznej, której owocami są nowe kanały komunikacji i dystrybucji, a co za tym idzie globalizacja konkurencji.

W tej sytuacji stosowanie starych metod marketingowych powoduje ryzyko szybkiej utraty klientów i pogorszenia wyników finansowych. Dlatego też bardzo popularne stało się nowe podejście nazywane marketingiem relacyjnym. Podejście to zakłada konieczność budowania indywidualnych, trwałych relacji pomiędzy dostawcą a klientem – to klient i jego potrzeby znajdują się w centrum zainteresowania organizacji. Kontakty z klientem traktuje się nie jako zbiór odrębnych epizodów, lecz jako proces, na który można oddziaływać.

W przypadku niewielkich organizacji budowanie indywidualnych relacji z klientem jest zadaniem stosunkowo prostym i naturalnym. Trudniej jest osiągnąć ten cel firmom mającym tysiące czy miliony klientów. W takim przypadku konieczne jest korzystanie z zaawansowanych technik informatycznych i podejmowanie działań na podstawie uważnej analizy gromadzonych danych. Jeśli będziemy umieli pokazać klientom, że rozumiemy ich potrzeby i potrafimy je zaspokoić, mamy szansę na nawiązanie trwałych i owocnych relacji.

CRM – Podstawowe pojęcia

Odpowiedzią na potrzebę zarządzania relacjami z klientami stała się metodyka CRM będąca praktyczną realizacją filozofii marketingu relacyjnego. Często CRM jest traktowany jako pewien rodzaj systemu informatycznego, co znacznie zawęża znaczenie tego terminu. W rzeczywistości CRM jest pewną strategią działania, a systemy informatyczne są jedynie narzędziami wspierającymi tę strategię. Samo wdrożenie takiego systemu nie wystarczy – by poprawnie zarządzać relacjami z klientem, konieczne są jeszcze radykalne zmiany w organizacji i kulturze pracy.

W CRM możemy wyróżnić trzy podstawowe poziomy (podsystemy): operacyjny, analityczny oraz interakcyjny.

- **CRM operacyjny** ma na celu zbieranie danych transakcyjnych i danych o klientach, ich opinii o produktach, sprzedawcach i komunikacji. Zadaniem operacyjnego CRM jest również wsparcie telemarketingu.
- Zadaniem **CRM analitycznego** jest przetwarzanie i analiza danych, data mining w celu planowania marketingowego, segmentacji i strategii instrumentalnych.

ZASOBY

Dodatkowe informacje na temat opisanych w artykule i innych metod data mining dostępne są pod adresem www.StatSoft.pl/dataminer2.html.

- **CRM interakcyjny** ma na celu kształtowanie bezpośrednich kontaktów z klientem i bieżące reagowanie, za pomocą różnorodnych kanałów dystrybucji i komunikacji, na pojawiające się wymagania.

CRM analityczny

Podsystemem, którego działanie ma największy wpływ na podejmowanie strategicznych decyzji odnośnie różnego rodzaju akcji i kampanii marketingowych jest CRM analityczny. Zadaniem tego podsystemu jest dostarczenie informacji niezbędnych do podejmowania efektywnych działań skierowanych na klienta. Informacje te są często efektem bardzo skomplikowanych analiz data mining, tak więc informacja docierająca do osób korzystających z tego podsystemu jest z reguły wysoce przetworzona. By móc uzyskać poprawne, wcześniej nieznanne oraz pełne informacje, a także by móc mądrze je wykorzystać, konieczne jest:

- zdobycie i integracja zarówno wewnętrznych, jak i zewnętrznych informacji oraz zapewnienie ich kompletności w skali całej organizacji (*aspekt danych*),
- wydobywanie z przygotowanych danych ukrytej w nich wiedzy oraz przedstawienie wyników w sposób umożliwiający ich praktyczne wykorzystanie (*aspekt analizy*),
- zapewnienie odpowiedniej kultury organizacyjnej umożliwiającej korzystanie z wyników analiz (*aspekt organizacji*).

Aspekt danych – stosowanie tego typu systemu ma sens jedynie w przypadku, gdy jesteśmy w stanie zapewnić odpowiednią jakość danych – ich dokładność, zgodność, kompletność itp. W organizacji powinna obowiązywać wysoka kultura gromadzenia i przechowywania danych. Wszystkie informacje, jakie docierają do organizacji, powinny być traktowane jako potencjalne źródło użytecznych informacji mogących pomóc w zrozumieniu klientów i rejestrowane w systemie.

Z punktu widzenia zapotrzebowania organizacji na dane do analizy kluczową rolę odgrywa wdrożenie bazy danych służącej celom analitycznym, tak zwanej korporacyjnej hurtowni danych. Hurtownia danych integruje różnego rodzaju dane mogące pochodzić z CRM operacyjnego, z różnego rodzaju systemów transakcyjnych, księgowych, produkcyjnych itp. Innymi słowy, hurtownia integruje dane pochodzące z całej organizacji oraz jej otoczenia i umożliwia ich wykorzystanie do celów analitycznych. Tego typu system powinien spełniać szereg wymagań, z których najważniejsze to:

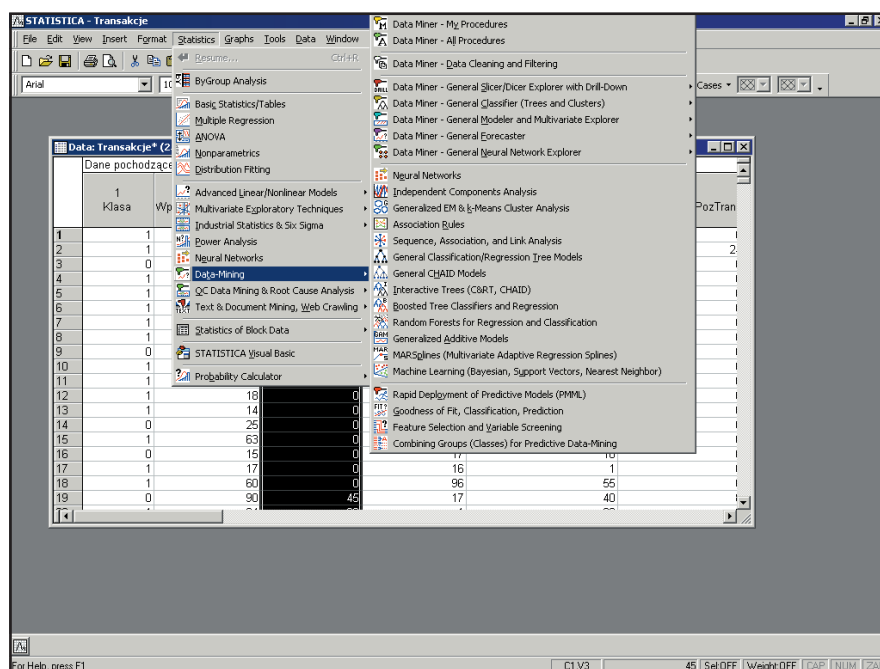
- struktura danych umożliwia generowanie wielowymiarowych zestawień i podsumowań,
- jednym z wymiarów danych jest czas,
- hurtownia jest odizolowana od innych systemów bazodanowych w przedsiębiorstwie,
- baza działa w trybie do odczytu – typowy użytkownik nie ma możliwości jej modyfikacji.

Bardzo ważne jest, by utworzona baza analityczna przechowywała dane atomowe, niezagregowane, ponieważ tylko na

podstawie takich danych możemy przeprowadzać dogłębne analizy statystyczne i data mining.

Aspekt analizy – zgromadzone przez nas dane zawierają w sobie wiele ukrytych informacji, niedostępnych przy pobieżnej analizie. Ukryte zależności, trendy i wzorce, innymi słowy ukryta wiedza, która może w znacznym stopniu zoptymalizować nasze wysiłki marketingowe, może zostać odkryta dzięki zastosowaniu odpowiednich narzędzi do analizy danych. Narzędzia analityczne tego typu noszą wspólną nazwę narzędzi do zgłębiania danych (data mining). Tego typu narzędzia powinny:

- udostępniać wszystkie metody analityczne potrzebne podczas analizy;
- umożliwiać przeprowadzenie pełnego cyklu analizy danych, począwszy od preprocessingu, czyli wstępnej analizy, przez zasadniczą część analizy, aż po generowanie raportów;
- umożliwiać łatwe stosowanie zbudowanych modeli;
- oferować intuicyjny, prosty w obsłudze interfejs;
- pozwalać na szerokie dostosowanie systemu oraz budowanie własnych procedur analitycznych.



Rysunek 1. Środowisko STATISTICA Data Miner

BUSINESS INTELLIGENCE

Przykładowym narzędziem posiadającym tę funkcjonalność jest program *STATISTICA Data Miner*.

Aspekt organizacyjny – nawet najlepiej rozwiązany problem gromadzenia i analizy danych nie da efektu, jeśli równocześnie z nimi nie nastąpią zmiany organizacyjne w przedsiębiorstwie. Dobrze skonstruowany model zbudowany na danych o dobrej jakości nie będzie przydatny, jeśli jego wyniki nie będą dostępne dla osób, które ich potrzebują. Dlatego też w parze ze zmianami technologicznymi muszą iść zmiany w kulturze organizacyjnej. Tak naprawdę kultura CRM powinna obejmować całą firmę. Do osiągnięcia sukcesu konieczna jest współpraca różnych części i działów przedsiębiorstwa, które dostarczają informacji potrzebnych do lepszego zrozumienia naszych klientów oraz w zapewnieniu procesów umożliwiających kontaktowanie się z naszymi klientami w odpowiedni sposób.

Wartość klienta

Kluczowym pojęciem występującym w CRM jest wartość klienta. Wartość

klienta nie powinna być jednak utożsamiana z kwotą, jaką u nas wydał – równie istotne są korzyści, jakie dany klient może przynieść w przyszłości. Najistotniejszymi czynnikami wpływającymi na wartość klienta są na pewno dochody, jakie przynosi organizacji, a także jego lojalność oraz skłonność do polecenia nas znajomym.

Znajomość wartości klienta jest bardzo istotną miarą pozwalającą ocenić skuteczność planowanych akcji marketingowych. Bez niej nie jesteśmy w stanie stwierdzić, czy planowana przez nas akcja przyniesie wymierny efekt.

Wykorzystując wskaźnik określający wartość klienta, przy pomocy odpowiednio zaplanowanego badania rynkowego możemy oszacować końcowy efekt ekonomiczny planowanej kampanii. Jeśli nie jesteśmy w stanie określić wartości klienta, korzyści z analiz data mining będą w dużym stopniu ograniczone. W takim przypadku możemy jedynie określić charakterystyki klientów lub określić, jakie grupy klientów będą zachowywać się podobnie w danej sytuacji. Jeśli jednak pragniemy zoptymalizować kampanię mającą określić,

którzy z naszych klientów są skłonni odpowiedzieć na naszą ofertę, to bez prognozy opartej na wartości klienta nie będziemy w stanie stwierdzić, czy tego typu akcja w ogóle ma sens ekonomiczny.

Rodzaje zadań analitycznych

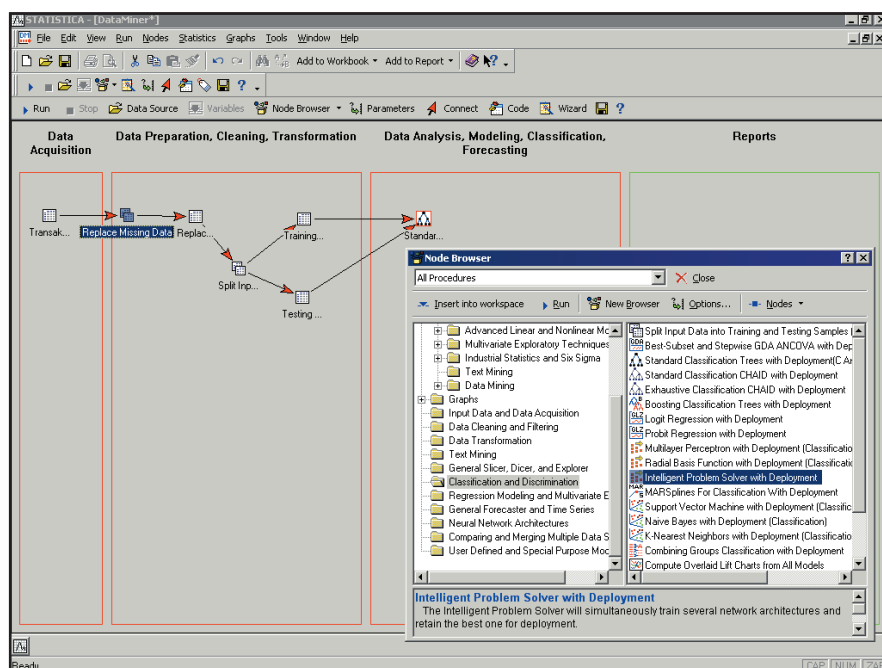
Metody statystyczne i data mining mogą być ogromnie użyteczne w doskonaleniu relacji z klientem. Do najczęstszych obszarów ich zastosowań należy zaliczyć:

- segmentację,
- analizy związane z cyklem życia klienta,
- analizy dotyczące sekwencji zakupów oraz podobieństwa i powiązań między produktami,
- analizę satysfakcji klientów.

Segmentacja. Analiza data mining może pomóc zidentyfikować grupy podobnych do siebie klientów, innymi słowy dokonać ich segmentacji. Tego typu analiza może być przydatna do optymalizowania naszych działań i bardziej indywidualnego traktowania naszych klientów, ponadto może być wstępem do dalszych analiz. Dalsze analizy przeprowadzamy już jedynie na wyodrębnionych segmentach, dzięki czemu możemy uzyskać bardziej precyzyjne modele, uzyskać bardziej szczegółowe informacje. Możemy wyróżnić kilka rodzajów segmentacji:

- segmentację demograficzną – jest przeprowadzana na podstawie danych, takich jak dochód klienta, wiek, płeć, wykształcenie, stan cywilny, ilość osób w rodzinie, status mieszkaniowy, typ mieszkania, grupa etniczna, wyznaniowa itp.,
- segmentację behawioralną – jest przeprowadzana w oparciu o dane reprezentujące zachowanie klientów,
- segmentację pod względem motywacji – opiera się na zmiennych opisujących przyczyny, z powodu których klient dokonał zakupu (dane ankietowe).

Cykl życia klienta. Wartościowy klient rzadko jest klientem statycznym. Relacja z nim ewoluuje i zmienia się w cza-



Rysunek 2. Umieszczanie w przestrzeni roboczej STATISTICA Data Miner modułów analitycznych

sie. Zrozumienie tej relacji jest krytyczne dla zarządzania kontaktami i podejmowania odpowiednich działań. Różne poziomy w relacjach z klientami można opisać za pomocą modelu cyklu życia klienta. Zakłada on, że niektórzy z potencjalnych klientów, dokonując zakupu, zmieniają się w naszych nowych klientów, ci zaś, jeśli przez pewien czas korzystają z naszych usług, mogą być postrzegani jako nasi stali klienci. Następnie pewna ich część może zrezygnować, stając się naszymi byłymi klientami. Budowanie modeli predykcyjnych związanych z cyklem życia klienta jest najbardziej typowym przykładem wykorzystywania metod data mining w zarządzaniu relacjami z klientem. Oczywiście w zależności od etapu cyklu zmieniają się typy analiz, jakie można przeprowadzić.

W odniesieniu do naszych potencjalnych klientów, czyli do osób, które znajdują się na rynku docelowym, lecz nie są naszymi klientami, możemy budować modele służące do optymalizowania kampanii polegającej na wysyłaniu ofert. Model wskazuje osoby, do których warto skierować ofertę. Budowa tego typu modeli jest dość trudna, ponieważ z reguły opieramy się na danych zewnętrznych (osoby nie są naszymi klientami).

Kolejna grupa (faza cyklu życia klienta) to osoby, które odpowiedziały na naszą ofertę i są nią zainteresowane, ewentualnie dokonały zakupu po raz pierwszy. Modele budowane dla tej grupy klientów mają na celu wskazanie, czy nowy klient będzie wysoce dochodowy, przeciętny czy też będzie przynosił niski dochód. Tego typu analiza jest możliwa z tego względu, że bardzo dobrym predyktorem dalszego zachowania klienta jest jego pierwszy zakup, niejako inicjujący kontakty z nami.

Kolejny etap to osoby, które są naszymi stałymi klientami. Naszym celem jest maksymalizacja korzyści wynikającej z nawiązanej relacji z klientem. Cel ten możemy starać się osiągnąć na wiele różnych sposobów, w szczególności możemy sprawić, by klienci:

- kupowali więcej towarów, usług – można zbudować model wskazujący, jacy klienci byłiby skłonni w większym stopniu korzystać z naszej oferty, ewentualnie którym warto zaproponować rozszerzoną wersję produktu (*up-selling*),
- kupowali szerszy asortyment – budowane są modele wskazujące, którym z klientów można zaproponować dodatkowe towary lub usługi (*cross-selling*),
- dłużej byli naszymi klientami – przeprowadzane są analizy mające na celu wskazanie, którzy klienci mogą odejść, co może dać szansę dostawcy na podjęcie odpowiednich kroków, mających na celu zatrzymanie ich (*churn*).

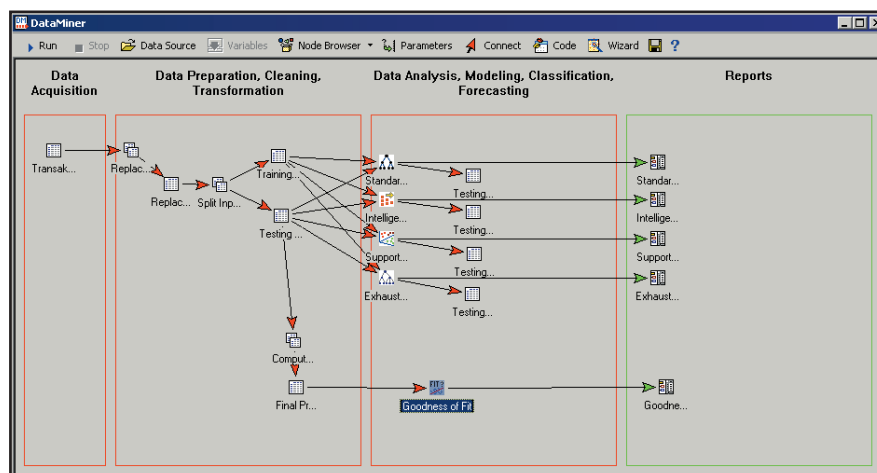
Ostatni etap życia klienta to grupa naszych byłych klientów, w odniesieniu do których możemy stosować działania mające ich skłonić, by ponownie zostali naszymi klientami. Budowane modele mogą nam pomóc w tego typu działaniach (*winback campaign*).

Warto zauważyć, że im późniejszy etap cyklu życia klienta, tym bardziej pełne i rzetelne są analizy dla niego przeprowadzane. Wynika to z faktu, iż dysponujemy coraz większym zestawem informacji o kliencie. W przypadku potencjalnych klientów w najlepszym wypadku dysponujemy danymi demo-

graficznymi, w późniejszych fazach dochodzi cały szereg informacji związanych z zachowaniem klienta, które zwykle są znacznie lepszymi predyktorami (jeśli podczas budowy modelu dysponujemy zarówno zmiennymi demograficznymi jak i opisującymi zachowanie klientów, to zwykle zmienne demograficzne rzadziej są uznawane za istotne).

Analizy dotyczące sekwencji zakupów oraz podobieństwa i powiązań między produktami. Kolejne zadanie wykonywane dzięki technikom data mining polega na określeniu powiązań między poszczególnymi produktami, a dokładniej na próbie odpowiedzi na pytanie *Jakie produkty są zwykle kupowane razem lub w pewnych odstępach czasu oraz w jakiej kolejności?* Odpowiedzią na to pytanie jest zwykle zestaw pewnych reguł zbudowany na podstawie historycznych danych (transakcyjnych), zawierających zapis dokonywanych zakupów. Znajomość takich reguł może być bardzo przydatna w różnego rodzaju działaniach skierowanych na klienta. Wiedza na temat kupowanych produktów umożliwia także uzyskanie informacji odnośnie naszych klientów oraz motywów, jakimi się kierują, dokonując określonych zakupów.

Bardzo często wyniki takich analiz stanowią wstęp do dalszych dociekań. Tego typu analizy poprzedzają często



Rysunek 3. Przykładowy projekt data mining w środowisku STATISTICA Data Miner

opisywane powyżej kampanie sprzedaży krzyżowej. Znajomość reguł postępowania klientów pozwala nam określić, jaki następny produkt najprawdopodobniej zakupi dany klient, i na tej podstawie podejmować działania odnośnie kierowania ofert.

Znajomość tego typu reguł jest również bardzo przydatna w handlu internetowym. Podczas zakupów w sklepie internetowym wszystkie transakcje zapisywane są w bazie danych, a klient, który ich dokonywał, jest jednoznacznie identyfikowany w systemie, tak więc przechowywana jest cała historia jego zakupów. Na podstawie tych danych generowane są wzorce, dzięki którym program z dużym prawdopodobieństwem „wie”, jakimi produktami dany klient może być zainteresowany i już na wstępie może zaproponować odwiedzającemu określony asortyment.

Oczywiście nie wszystkie odnalezione reguły muszą być użyteczne. Część z nich na pewno będzie przedstawiała zależności znane analizującym, część reguł może być niemożliwa do interpretacji. Jednak pozostałe mogą znacznie poszerzyć wiedzę na temat badanego zjawiska, sugerując określone działania lub wskazując obszary dalszej analizy.

Bardzo przydatne w zrozumieniu badanego zjawiska są narzędzia wizualizacji, dzięki którym możemy analizować całą siatkę połączeń pomiędzy poszczególnymi obiektami, co umożliwia głębsze zrozumienie wzajemnych zależności i relacji.

Pomiar satysfakcji klientów. Satysfakcja klientów określana jest jako poziom zaspokojenia potrzeb klienta przez całkowity produkt, jaki nabył u dostawcy. Pod pojęciem produktu całkowitego rozumiemy również obsługę klienta oraz wszystkie usługi późniejsze (przeglądy serwisowe itp.). Pomiar satysfakcji klienta jest kluczowy w zrozumieniu jego rzeczywistych potrzeb i oczekiwań. Dzięki niemu jesteśmy w stanie powiedzieć, co jest naszą mocną stroną, a jakie aspekty naszej działalności powinniśmy zmienić. Satysfakcja klienta bezpośrednio wpły-

wa na jego stosunek do nas. Im bardziej nasz produkt zaspokaja potrzeby klienta, tym chętniej powtarza zakupy, natomiast przykre incydenty zwykle skłaniają go do zmiany dostawcy. Dlatego też kluczową sprawą w zarządzaniu relacjami z klientem jest pomiar, a następnie monitorowanie poziomu satysfakcji klientów.

W celu mierzenia i monitorowania satysfakcji wśród klientów (rzeczywistych i potencjalnych) przeprowadza się odpowiednie badania marketingowe i na tej podstawie określa się przyczyny ich konkretnych postaw i zachowań. Podczas planowania badań oraz analizy ich wyników niezbędne jest wykorzystanie technik statystycznych. Głównymi zadaniami, jakie stoją przed analitykiem, jest wyznaczenie odpowiedniej liczności (reprezentatywności) próby, na której ma być prowadzone badanie, określenie skali, na której będzie odbywał się pomiar satysfakcji, opracowanie odpowiednich pytań, a następnie analiza wyników, podczas której obok całej gamy testów statystycznych można używać innych, często bardzo zaawansowanych metod, jak np. karty kontrolne, analiza czynnikowa czy głównych składowych.

Bardzo ważne jest, by pomiary satysfakcji klientów nie były przedsięwzięciem jednorazowym. Aby określić efekty naszych wysiłków, bądź w porę wychwycić niekorzystne zjawiska, konieczne jest regularne monitorowanie poziomu satysfakcji.

Techniki data mining w analitycznym CRM

Wszystkie przedstawione powyżej rodzaje analiz w praktyce wykonuje się za pomocą technik data mining. Pod tą nazwą kryje się w rzeczywistości szereg różnego rodzaju metod analitycznych. Ich zastosowanie zwykle się różni, dlatego też wprowadzono szereg klasyfikacji porządkujących te metody. Najbardziej odpowiedni z punktu widzenia przedstawionych zastosowań wydaje się podział na ukierunkowany data mining (uczenie z nauczycielem) oraz

nieukierunkowany data mining (uczenie bez nauczyciela).

Metody ukierunkowanego data mining są wykorzystywane w przypadkach, gdy interesujący nas podział (na przykład na klientów, którzy kupili i nie kupili danego produktu) został już zaobserwowany i zapisany w próbie uczącej. Wykorzystując te dane, budujemy model, który ma na celu przewidywanie tego podziału dla nowych danych. Tego typu modele mogą znacznie obniżyć koszty naszych działań, nie obniżając naszej skuteczności, ponieważ nasze wysiłki kierujemy jedynie do osób, wobec których istnieje szansa, że na nie odpowiedzą. Techniki ukierunkowanego data mining stosujemy we wszystkich analizach odnoszących się do cyklu życia klienta. Do metod ukierunkowanego data mining zalicza się: drzewa decyzyjne, sieci neuronowe, MARSplines, Naive Bayes oraz Support Vector Machine.

W sytuacji nieukierunkowanego data mining sytuacja jest inna. Posiadamy pewne dane o kliencie, które jednak nie zawierają żadnego podziału narzuconego z góry. Naszym celem jest wyjaśnienie struktury tych danych, na przykład określenie skupisk czy zależności niewidocznych na pierwszy rzut oka. Możemy przykładowo dysponować bazą danych klientów zawierającą różnego typu zmienne, które potencjalnie mogą mieć duży wpływ na zachowanie klientów. Celem analizy może być odnalezienie segmentów rynku, czyli grup osób podobnych do siebie, bądź odnalezienie związków pomiędzy kupowanymi towarami. Do technik nieukierunkowanego data mining zaliczyć można: analizę skupień, analizę koszykową (asocjacji), analizę sekwencji i połączeń, sieci neuronowe oraz analizę czynnikową i składowych głównych. ■

Grzegorz Migut pełni funkcję konsultanta data mining w firmie StatSoft Polska. Zajmuje się tematyką analitycznego CRM i wykrywania nadużyć. Kontakt: g.migut@statsoft.pl.