



## **BAZY DANYCH I NARZĘDZIA ANALITYCZNE W BADANIACH SATYSFAKCJI I LOJALNOŚCI**

**Grzegorz Migut**

*StatSoft Polska Sp. z o.o.*

### **Bazy danych w strategiach lojalnościowych**

Podtrzymanie więzi z klientem, traktowanie go w sposób indywidualny wymaga coraz lepszego poznawania jego preferencji, upodobań, przyzwyczajzeń. To sprawia, że konieczne jest, by napływające o kliencie informacje były w odpowiedni sposób gromadzone, przechowywane i łatwo dostępne. Rejestrowanie wszystkich związanych z klientem zdarzeń rynkowych jest możliwe dzięki wykorzystaniu zaawansowanych technik informacyjnych opierających się na bazach danych.

By baza mogła skutecznie wspomagać rozwijanie więzi z klientem, powinna zawierać nie tylko informacje dotyczące danych personalnych (imię, nazwisko, wiek, adres zamieszkania), ale również przechowywać dane odnośnie zakupionych przez niego towarów, sumy wydanych pieniędzy, preferencji, przyzwyczajzeń, zamożności, stanu rodziny i szeregu innych cech. Im szersza gama zgromadzonych informacji na temat klienta, tym bardziej miarodajne będą analizy przeprowadzone na ich podstawie.

Ze strategiami lojalnościowymi oraz z bazami danych ściśle łączy się termin marketing partnerski, którego głównym założeniem jest budowanie związków organizacji z klientem. Można wyróżnić w nim trzy rodzaje działań mających na celu zwiększenie lojalności klienta [6].

- ◆ Pierwszy polega na stosowaniu programów lojalnościowych skierowanych do wszystkich klientów. Jako narzędzia oddziaływania można tu wymienić: karty stałego klienta, zbieranie punktów itp.
- ◆ Drugi poziom opiera się na kierowaniu specjalnych ofert jedynie do wąskiej grupy klientów. Oferty te mają na celu zmotywowanie klientów do bliższego kontaktu z firmą.
- ◆ Trzeci opiera się na kompleksowej obsłudze klienta w każdej fazie kontaktu z nim. To podejście zakłada wdrożenie w organizacji systemu informatycznego klasy CRM (*Customer Relationship Management*), który pozwala na integrację procesów związanych z klientem, tj.: sprzedaży, marketingu oraz obsługi po sprzedaży. Pozwala on także na połączenie tych procesów z pozostałymi procesami firmy (poprzez integrację z modułami odpowiedzialnymi za inne aspekty jej działalności (np. ERP i SCM)).



Na wszystkich tych poziomach dysponowanie kompletną informacją o kliencie jest kluczowym czynnikiem wpływającym na skuteczność pogłębiania relacji z nim. Dlatego bardzo istotna jest dbałość o informację i pamięć o tym, że każde zdarzenie polegające na kontakcie klienta z firmą powinno być rejestrowane przez system informatyczny.

Do gromadzenia informacji na temat codziennych operacji (w tym także interesujących nas informacji na temat klienta) mających miejsce w organizacji używane są transakcyjne bazy danych. Organizacja nie posiada zwykle jednej bazy danych przechowującej informacje na temat wszystkich aspektów jej działalności. Różnego typu wydarzenia są rejestrowane przez odrębne bazy, gromadzące informacje dotyczące poszczególnych działów firmy. Na przykład rejestrowanie sprzedaży odbywa się przy udziale bazy danych marketingu, a informacje finansowe są przechowywane w bazie rachunkowości. Dodatkowy podział może wynikać również z rozproszenia geograficznego naszej organizacji.

Informacje dotyczące naszego klienta mogą być również rozrzucone po różnych bazach, uniemożliwiając spójną ich analizę. Korzystanie z takich baz danych może okazać się bardzo niewygodne lub wręcz może uniemożliwić wyciągnięcie wartościowych wniosków. Dopiero integracja wszystkich danych zgromadzonych w organizacji pozwala na przeprowadzenie miarodajnej, rzetelnej analizy. Można tu zastosować podejście polegające na zintegrowaniu baz przy pomocy oprogramowania typu „gateway”, powodującym, że fakt istnienia kilku baz danych jest dla końcowego użytkownika niewidoczny, jednak lepszym rozwiązaniem jest zbudowanie bazy danych przeznaczonej specjalnie do celów analitycznych, tak zwanej hurtowni danych. Tego typu narzędzia są szczególnie pomocne, jeśli dysponujemy dużą ilością zgromadzonych danych i są one rozproszone w wielu oddziałach organizacji.

Hurtownie danych są używane w celu integracji często olbrzymiej ilości danych, pochodzących z wielu odrębnych baz danych, działających w różnych częściach organizacyjnych lub geograficznych przedsiębiorstwa. Filozofia działania hurtowni danych może być opisana przez cztery poniższe charakterystyki:

- ◆ Jest zorganizowana wokół pewnych głównych tematów, w naszym przypadku najbardziej interesującym tematem jest klient.
- ◆ Jest ona skonstruowana na podstawie wielu różnorodnych źródeł. Patrząc w aspekcie technicznym, może zawierać dane pochodzące z baz danych działających w oparciu o różne podejścia (relacyjne, hierarchiczne itp.), a także rozwiązania różnych firm. W aspekcie organizacyjnym scala dane pochodzące z różnych działów (finansowy, sprzedaży, techniczny itp.).
- ◆ Gromadzi dane historyczne, informacje w niej zawarte odnoszą się do konkretnych punktów w czasie (nazywanych również migawką danych [4]). Analityczna baza danych może przechowywać na przykład migawkę informacji dotyczących sprzedaży o północy każdego dnia. Na tej podstawie można dokonywać analiz porównawczych dla różnych okresów i w różnych odstępach czasu.
- ◆ Jest odizolowana od systemu transakcyjnego organizacji, znajduje się jakby na boku w stosunku do baz danych gromadzących informacje operacyjne.



Hurtownie danych buduje się w celu dostarczenia organizacji danych, które będą wykorzystane do analiz. Dlatego też struktura zgromadzonych danych powinna umożliwiać wykonywanie na nich operacji mających na celu uzyskanie porównań, analizę wzorców czy trendów. Z założenia jest ona zaprojektowana tak, by przechowywać oraz obsługiwać duże ilości danych, ponieważ takie wymogi stawiają zwykle przeprowadzane analizy.

Analityczną bazę danych, w przeciwieństwie do baz transakcyjnych, projektuje się jako stałą bazę danych (tylko do odczytu). Wynika to z różnych zadań stawianych przed tymi systemami. Systemy transakcyjne działają zarówno w trybie zapisu, jak i odczytu, przy czym główny nacisk położony jest na rejestrowanie zdarzeń dotyczących codziennej działalności firmy, a co za tym idzie zawartość bazy ulega nieustannym zmianom. Inaczej jest w przypadku hurtowni, której użytkownicy mogą jedynie odczytywać zawarte w niej informacje, dokonywać na nich potrzebnych analiz, ale nie mają możliwości zmiany danych przechowywanych w bazie. Uaktualnienie zawartości hurtowni danych wykonywane jest jedynie w zaplanowany, kontrolowany sposób. To ograniczenie jest zrozumiałe, ponieważ do przeprowadzania analiz potrzebne są stałe dane, niezmiennające się w trakcie wykonywania obliczeń.

Kolejną różnicą pomiędzy operacyjną bazą danych a hurtownią danych jest sposób przechowywania informacji. W hurtowniach są one przechowywane w formie tak zwanych schematów gwiazdy. Centralnym elementem tego schematu jest tablica faktów (*fact table*) (takim faktem może być na przykład sprzedaż), do niej dołączone są tak zwane tablice wymiarów (*dimension table*) opisujące dany fakt w poszczególnych aspektach (na przykład czas, miejsce, rodzaj towaru).

Z hurtowniami danych związane jest ściśle pojęcie OLAP (*On-Line Analytical Processing*), czyli analiza informacji w czasie rzeczywistym na potrzeby wspomagania decyzji. Największą siłą OLAP jest jego zdolność do wizualizacji danych w celu przedstawienia ich w sposób zrozumiały dla użytkownika. Jest to szczególnie przydatne w wypadku, gdy dane zawierają łatwo identyfikowalne wymiary, a także jeśli poszczególne wymiary dają się przedstawić na różnym stopniu szczegółowości (wymiar czasu może być przedstawiany w postaci dni, tygodni, miesięcy, kwartałów itp.). Dane przedstawiane są w formie wielowymiarowej kostki danych. Na przykład wartość sprzedaży przedstawiana jest równocześnie w trzech wymiarach: okresie, w jakim zostały dokonane zakupy, miejscu oraz w zależności od rodzaju zakupionych przedmiotów (jest możliwe obliczanie kostki dla większej ilości wymiarów, trudno jednak wyobrazić sobie jej wizualizację). OLAP oferuje wiele operacji na takich kostkach, między innymi *drill-down* oraz *roll-up*, które polegają na zmianie stopnia szczegółowości przedstawionych danych. Załóżmy, że wymiar miejsca zakupu określony jest dla poszczególnych miast, czas przedstawiany jest w miesiącach, natomiast produkty dzielą się na spożywcze i pozostałe. Pojedynczy wycinek kostki informuje nas o wielkości sprzedaży danego typu artykułu w danym miesiącu, w danym mieście. Operacja *drill-down* dla czasu powoduje, że zwiększa się stopień szczegółowości w tym wymiarze, analizujemy teraz sprzedaż w wymiarze tygodniowym. Operacją odwrotną do *drill-down* jest *roll-up*, polega ona na zmniejszeniu szczegółowości przedstawianych danych (zamiast miesięcy mamy na



przykład kwartały). Ogromną zaletą OLAP jest szybkość działania. Przy pomocy SQL wykonanie podobnych obliczeń było by bardzo długotrwałe, a czasami wręcz niemożliwe.

Przydatność hurtowni danych zaznacza się przede wszystkim we wspieraniu procesu decyzyjnego. Mogą być one szczególnie przydatne, jeśli mamy za zadanie ocenić potrzeby i preferencje klientów, co jest podstawową rzeczą dla organizacji myślącej o zatrzymaniu klienta przy sobie. Dysponowanie zintegrowanymi informacjami pozwala na łączną analizę wszystkich zgromadzonych informacji i skuteczne identyfikowanie grup o podobnych cechach psychograficznych i, co za tym idzie, na indywidualne dostosowanie ofert [6]. Hurtownia danych ułatwia również analizę programów marketingowych i planowanie strategiczne, dzięki szybkiemu dostępowi do kluczowych informacji.

Podczas budowania hurtowni danych ważnym problemem jest olbrzymia ilość danych, która powoduje rozrastanie się bazy do dużych rozmiarów. Dlatego też często stosuje się metody mające na celu redukcję ilości danych. Jedną z takich metod jest nieuwzględnianie w bazie niektórych wymiarów nieistotnych z punktu widzenia zapotrzebowania na informację. Inną częstą praktyką jest sumowanie poszczególnych transakcji i wprowadzanie ich w formie zagregowanej, na przykład danymi o najmniejszej ziarnistości są sumy transakcji z całego dnia. Możliwe jest także wprowadzanie jedynie próbek danych reprezentujących całość. Taki kształt bazy jest uzasadniony, gdy podstawowym zadaniem hurtowni jest pomoc w sporządzaniu raportów. Jednak jeśli hurtownia danych jest tworzona do celów analitycznych, przechowywanie zagregowanych danych jest zwykle niewystarczające, ponieważ wszystkie te zabiegi usprawniają co prawda działanie bazy, powodują natomiast utratę informacji przydatnych w analizie.

Przykładowo, nieistotny podczas projektowania bazy wymiar może okazać się kluczowy po pewnym czasie, a agregacja danych może uniemożliwić przeprowadzenie miarodajnych analiz. Przykładem mogą tu być opisane poniżej metody data mining (np. analiza asocjacji), które powinny być przeprowadzone na pojedynczych transakcjach, a nie na danych zagregowanych. Innym problemem występującym w przypadku, gdy dane w bazie przechowywane są w tradycyjnych strukturach relacyjnych, jest nadmierna normalizacja danych, która może sprawić, że przeprowadzenie analizy na ogromnej ilości rekordów może trwać bardzo długo. Tak więc zabiegi mające na celu optymalizację działania bazy, a także obniżkę kosztów, mogą znacząco utrudnić przeprowadzanie analiz.

Planując hurtownię danych użyteczną dla analiz statystycznych, należy pamiętać, że idealna do tych celów jest sytuacja, gdy dysponujemy szczegółowymi danymi na poziomie transakcji. Uwzględnienie tego wymagania jest konieczne, jeśli myślimy o przeprowadzeniu rzetelnych, miarodajnych analiz.

## **Narzędzia analityczne data mining w ocenie lojalności konsumentów**

Zalecaną praktyką stosowaną w analizie danych marketingowych jest tak zwane wsłuchiwanie się w dane [6], to znaczy analiza danych, która nie jest uwarunkowana naszymi wcześniejszymi przeświadczeniami o kształcie zależności występujących na rynku.



Otrzymane w ten sposób informacje odpowiadają rzeczywistości, a nie naszym odgórnym wyobrażeniom o niej. Mówiąc w skrócie, proces interpretacji powinien być poprzedzony procesem analizy, a nie na odwrót.

### *Narzędzia data mining*

Hurtownie danych, a także bazy danych mogą zostać połączone ze specjalistycznymi narzędziami analitycznymi służącymi do wyszukiwania w ogromnej ilości danych ukrytych zależności i powiązań czy anomalii (można inaczej powiedzieć, że poszukują wiedzy zawartej w danych). Tego typu analizy noszą łączne miano data mining.

Trudno jest jednoznacznie określić zbiór metod, które można zaklasyfikować do data mining, ponieważ w tego typu analizach można wykorzystywać zarówno tradycyjne metody statystyczne, jak i nowoczesne metody obliczeniowe, na przykład sieci neuronowe.

Poniżej zaprezentowano cztery tego typu metody, często pojawiające się w różnego rodzaju klasyfikacjach.

**Sieć neuronowa** jest narzędziem analizy danych, którego budowa i działanie zainspirowane zostało wynikami badań nad ludzkim mózgiem. Sieć składa się z:

- ◆ wejść, gdzie wprowadzane zostają dane,
- ◆ warstw połączonych ze sobą neuronów, w których przebiega proces analizy,
- ◆ wyjścia, gdzie pojawia się sygnał będący wynikiem analizy.

Docierające do neuronów sygnały są w nich przekształcane przez odpowiednią funkcję. Ważnym elementem struktury sieci są wagi, osłabiające lub wzmacniające poszczególne sygnały docierające do neuronów. To właśnie od rodzajów funkcji oraz wag zależą wartości, jakie wygeneruje sieć na wyjściu. Proces ustalania wag sieci jest nazywany uczeniem sieci i przeprowadzany jest przy pomocy danych będących przykładowymi realizacjami badanego zjawiska. Wyróżniamy dwa podstawowe tryby uczenia sieci:

- ◆ z nauczycielem,
- ◆ bez nauczyciela.

Uczenie z nauczycielem polega na tym, że sieci podaje się przykłady poprawnego działania, które powinna potem naśladować w swoim bieżącym działaniu. Na wejścia sieci podaje się konkretne sygnały, pokazując, jaka jest wymagana dla nich odpowiedź. Na tej podstawie tych danych sieć uczy się prawidłowego działania. Sieć uczona w tym trybie może być wykorzystana zarówno do zadań dyskryminacyjnych jak i predykcji, a także do klasyfikacji wzorcowej.

W trybie bez nauczyciela sieć ustala swoje wagi jedynie na podstawie danych wejściowych. Sieć uczona w tym trybie może być używana w analizie skupień lub do analizy głównych składowych.

Poprawnie nauczona sieć posiada umiejętność uogólnienia wiedzy zdobytej na podstawie historycznych obserwacji i udzielania trafnych odpowiedzi dla nowych danych. Dużą zaletą sieci neuronowych jest jej zdolność radzenia sobie z modelowaniem zależności



o charakterze nieliniowym, pewną ich wadą jest działanie na zasadzie czarnej skrzynki: nie jesteśmy w stanie podać reguły i zasad, na podstawie których otrzymano dany wynik.

Kolejną metodą, jakiej możemy użyć, są **drzewa klasyfikacyjne**. Proces budowy drzewa opiera się na zasadzie rekurencyjnego podziału. Zasada ta polega na przeszukiwaniu w przestrzeni cech wszystkich możliwych podziałów zbioru danych na dwie części, tak by dwa otrzymane podzbiory maksymalnie się między sobą różniły ze względu na zmienną zależną. Podział ten jest kontynuowany, aż do całkowitego podziału przypadków na jednorodne grupy lub spełnienia ustalonych warunków zatrzymania. Reguły, względem których dokonano podziału przestrzeni cech, można w łatwy sposób przedstawić w formie drzewa. Tego typu grafy składają się z wierzchołków i krawędzi. Każdy wierzchołek reprezentuje decyzję o podziale zbioru obiektów na dwa podzbiory ze względu na jedną z cech objaśniających. Z drzewami klasyfikacyjnymi związane jest też pojęcie przycinania. Jest ono użyteczne do zoptymalizowania struktury drzewa. Polega ono na upraszczaniu struktury drzewa (usuwaniu węzłów) przy równoczesnym zachowaniu (ewentualnie dopuszczalnym przez nas pogorszeniu) jego zdolności klasyfikacyjnych. Ważną zaletą drzew jest zrozumiała dla człowieka sekwencja reguł decyzyjnych, pozwalająca klasyfikować nowe obiekty na podstawie wartości zmiennych. Atrakcyjną jest również możliwość graficznej prezentacji procesu klasyfikacji. Dodatkową zaletą drzew klasyfikacyjnych jest ich odporność na obserwacje odstające.

**Analiza asocjacji** zwana jest również analizą koszykową, ponieważ najpopularniejszym jej wykorzystaniem jest analiza pod kątem znalezienia odpowiedzi na pytanie: jakie produkty kupowane są najczęściej razem. Ogólnie mówiąc, celem tej analizy jest znalezienie w dużym zbiorze danych pewnych ukrytych powiązań, które mogłyby zostać przedstawione w postaci reguł. Przeszukując zbiór danych, moglibyśmy na przykład poszukać powiązań dotyczących zachowań klientów i otrzymać regułę, brzmiącą na przykład: „Jeśli klient kupił telewizor, to kupił też magnetowid”.

W analizie asocjacji tworzone są reguły o postaci: JEŻELI [*Poprzednik*], TO [*Następnik*], gdzie *Poprzednik* i *Następnik* zawierają proste warunki logiczne w postaci kodów, wartości tekstowych, zestawów, wyborów użytkownika, fraz itp. Wszystkie przypadki analizowane pod kątem reguł asocjacji nazywane są transakcjami. Jeżeli jakaś transakcja jest zgodna z przyjętą regułą, to mówimy, że pasuje lub wspiera regułę asocjacji. Bardzo ważne jest, by odnajdywane reguły były możliwie jak najkrótsze, długość danej reguły określana jest przez termin *itemset*. Potencjalna ilość reguł, jakie mogą zostać odnalezione w zbiorze danych, jest ogromna i wzrasta w tempie  $n!$  w stosunku do ilości analizowanych zmiennych. Dlatego też wprowadza się pewne oceny, które mają na celu ograniczenie ilości odnalezionych reguł. Do ocen takich należą *support* oraz *confidence*.

*Support* (wsparcie lub pokrycie) oznacza prawdopodobieństwo jednoczesnego wystąpienia *Poprzednika* i *Następnika*. Innymi słowy, jest to miara, jak często wspólnie wystąpiły oba zdarzenia.

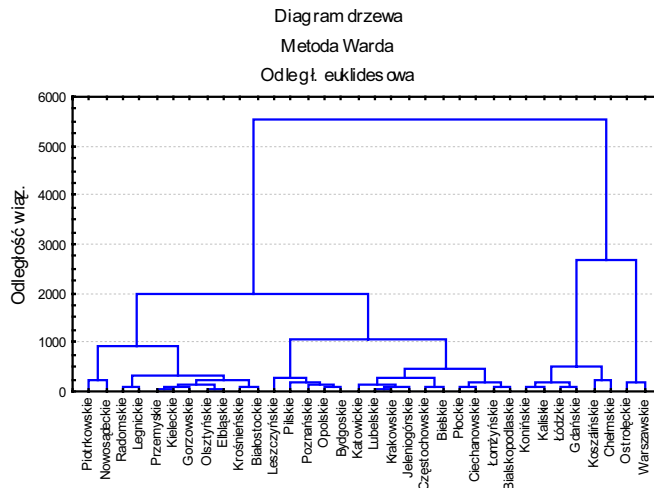
*Confidence* określa warunkowe prawdopodobieństwo, że w regule asocjacji *Poprzednik*, pod warunkiem, że *Następnik*.



Poszukiwanie reguł asocjacji jest bardzo użyteczną metodą, stosowaną często jako pierwsza w procesie analizy danych.

Celem **analizy skupień** (*cluster analysis*) jest wyodrębnienie ze zbioru danych obiektów, które byłyby podobne do siebie, i łączenie ich w grupy. W wyniku działania tej analizy z jednego niejednorodnego zbioru danych otrzymujemy grupę kilku jednorodnych zbiorów. Obiekty znajdujące się w tym samym zbiorze uznawane są za „podobne do siebie”, obiekty z różnych zbiorów traktowane są jako „niepodobne”. Pojęcie analizy skupień obejmuje faktycznie kilka różnych algorytmów klasyfikacji. Do najważniejszych należy zaliczyć metodę aglomeracji oraz k-średnich.

Algorytm aglomeracji służy do grupowania obiektów (np. klientów) w coraz to większe zbiory (skupienia), z zastosowaniem pewnej miary podobieństwa lub odległości. Typowym wynikiem tego typu grupowania jest hierarchiczne drzewo. Na początku tej analizy uznajemy, że każdy element zbioru stanowi oddzielną grupę. Następnie stopniowo osłabiamy kryterium uznawania obiektów za takie same, co powoduje grupowanie się obiektów podobnych. W miarę dalszego osłabiania kryterium wiążemy ze sobą coraz więcej obiektów i agregujemy je w coraz większe skupienia elementów, coraz bardziej różniących się od siebie. W końcu, na ostatnim etapie, wszystkie obiekty zostają ze sobą połączone. Efekty działania tego algorytmu można przedstawić w formie hierarchicznego drzewa, które przedstawia kolejne kroki działania analizy. Tego typu analizę możemy przeprowadzić nie tylko dla przypadków, ale również dla zmiennych, co polega na łączeniu najbardziej podobnych (w sensie odległości, a nie korelacji) w grupy, podobnie jak przedstawiono powyżej.



Odmiernym podejściem jest metoda k-średnich. Stosowanie jej wymaga od nas podania liczby grup, na które zostanie podzielony wejściowy zbiór danych. Jedną z wersji tej metody polega na losowym wyborze  $k$  obiektów z analizowanego zbioru i uznania ich za środki  $k$  grup. Każdy z pozostałych obiektów jest przypisywany do grupy o najbliższym mu środku. Następnie oblicza się nowe środki każdej podgrupy na podstawie średnich arytmetycznych ze



współrzędnych zawartych w nich obiektów. W kolejnym kroku następuje przegrupowanie elementów grup, każdy obiekt jest przesuwany do tej grupy, do której środka ma najbliżej. Procedurę tę powtarzamy do momentu, gdy w danej iteracji żaden z obiektów nie zmieni swojej podgrupy. Pewną wadą tej metody jest konieczność odgórnego określenia liczby skupień występujących w danych, dlatego też zaleca się powtórzenie procedury dla różnych wartości  $k$  i wybranie tej, dla której zbiór danych jest podzielony najlepiej.

Warto dodać, że opisane powyżej OLAP oraz metody data mining w wielu przypadkach są ze sobą komplementarne [1]. Analiza data mining może ułatwić proces budowania kostki danych poprzez uprawnienie procesu definiowania wymiarów, natomiast OLAP dostarcza bardzo silnych narzędzi do wizualizacji danych, które pozwalają lepiej zrozumieć rezultaty procesu data mining, uzyskanych w wyniku działania na przykład sieci neuronowych czy analizy skupień.

### ***Data mining w ocenie lojalności klientów***

Na podstawie informacji zawartych w bazach danych, przy pomocy odpowiednich technik data mining możemy określić szczegółowe profile klientów umożliwiające ich segmentację, czyli podział na rozłączne i jednorodne pod względem badanych cech grupy.

Możemy dokonywać segmentacji klientów pod kątem różnych cech. Segmentacja oparta o profil psychologiczny klienta, jego preferencje, zamożność itp. pozwala nam w większym stopniu rozpoznać jego potrzeby i traktować go w sposób bardziej indywidualny. Takie traktowanie klienta jest konieczne, jeżeli chcemy myśleć o zwiększeniu jego przywiązania do naszej firmy.

Naszych klientów możemy również podzielić na grupy w zależności od poziomu ich lojalności. Przykład tego typu podziału jest przedstawiony w [7]. W przykładzie tym lojalność klienta została zdefiniowana jako poziom konsekwencji w dokonywaniu zakupów i wyboru usług u konkretnego sprzedawcy. Czynnikiem, jaki brano pod uwagę była też ilość wydawanych pieniędzy oraz liczba dokonanych zakupów w ciągu roku. W wyniku analizy klienci zostali podzieleni na cztery poziomy lojalności. Najbardziej lojalni klienci dokonywali zakupów najczęściej, za największe sumy oraz kupowali konkretne, stałe grupy produktów. Podobnie zachowywała się druga grupa klientów, różniąc się od pierwszej tym, że kupowała różne kategorie produktów. Trzecia grupa klientów odwiedzała sklep kilka razy do roku, pozostawiając znacznie mniej pieniędzy niż klienci z pierwszych dwóch grup, dodatkowo cechowała się niewielką konsekwencją podczas zakupów. Ostatnia grupa reprezentowała klientów, którzy odwiedzili sklep bardzo rzadko lub dokonali jednorazowego zakupu.

Wiedza, jaką można uzyskać, prowadząc segmentację klientów ze względu na lojalność, może pozwolić na zidentyfikowanie czynników mających wpływ na ich zachowanie, a także określić, z jakich wyników działalności są oni najmniej zadowoleni. Ta wiedza może znacznie poprawić skuteczność działań mających na celu zatrzymanie naszych klientów, a także prób zdobycia klientów konkurencji. Ponadto można zasymulować reakcję klientów na odpowiednich poziomach lojalności po wprowadzeniu określonych działań czy produktów i przewidzieć w ten sposób ich zachowania, a przez to określić potencjalne korzyści.



W wyniku takiej analizy może się okazać, że nie warto przykładać zbyt dużej wagi do utrzymania własnych klientów „do wzięcia” (podobnych do grupy trzeciej i czwartej z przykładu), jeśli ich głównym priorytetem są niskie ceny, a zamiast tego podjąć próbę przechwycenia określonych segmentów klientów rywala, których potrzeba lepszej obsługi nie została zaspokojona.

Mając do dyspozycji dane dotyczące zakupów, możemy przeprowadzić analizę asocjacji, by zidentyfikować reguły, według których klienci kupują nasze towary lub korzystają z naszych usług. Jeżeli na przykład przypadkowy klient kupił u nas telewizor, a odkryta reguła wskazuje, że tego typu klienci w późniejszym okresie zwykle kupują też magnetowid, warto skierować do niego ofertę dotyczącą tego typu urządzeń.

Inną przydatną analizą, jaką można przeprowadzić przy pomocy data mining, jest określenie poziomu dochodowości klienta (obszerny przykład zawarty jest w [3]) oraz modelowanie wzrostu poziomu tej dochodowości pod wpływem akcji marketingowych. Jeśli dochód pochodzący od poszczególnego klienta wzrośnie o większą kwotę niż wynosi nakład środków na promocję wydaną na niego, możemy wtedy mówić o sukcesie. Na tej podstawie można określić, czy nasza kampania ma sens ekonomiczny.

W wielu gałęziach przemysłu koszt pozyskania nowego klienta jest o wiele wyższy od nakładów, jakie należy ponieść w celu zatrzymania osoby będącej już naszym klientem. Dlatego też warto przedsięwziąć pewne kroki mające na celu zwiększenie poziomu lojalności u grupy klientów najbardziej zagrożonych odejściem. Dzięki data mining możemy podzielić klientów na grupy w zależności od poziomu ryzyka ich odejścia i na tej podstawie przyjąć odpowiednią politykę marketingową dla poszczególnych grup. Tego typu analiza została zaprezentowana w [3] na przykładzie operatora telefonów komórkowych. Analizę przeprowadzono w oparciu o drzewa decyzyjne. Miała ona na celu podział klientów w zależności od ryzyka, z jakim mogą oni zrezygnować z usług operatora. Powstałe w wyniku analizy drzewo dzieliło klientów na 29 grup, z których każda cechowała się innym wskaźnikiem odejścia. Ponieważ analiza drzew decyzyjnych pozwala na prześledzenie reguł, na podstawie których zostały zbudowane, dla każdej z grup ustalono konkretne charakterystyki, takie jak rodzaj umowy czy długość korzystania z usług firmy.

W wyniku analizy stwierdzono, że 5,2% klientów zgromadzonych w bazie zawiera 27,7% ogólnej liczby spodziewanych odejść klientów, natomiast 19,7% klientów zawiera 55,8% tej liczby. Stąd kampania marketingowa skierowana jedynie do około 20% klientów może dotrzeć do ponad 55% osób zagrożonych odejściem, co może zmniejszyć koszty przeprowadzanej przez nas kampanii mającej na celu zwiększenie poziomu ich lojalności. Dodatkowo znajomość profili poszczególnych klientów może pozwolić na bardziej indywidualne sformułowanie oferty.

Przedstawione powyżej przykłady nie wyczerpują oczywiście możliwości wykorzystania metod data mining. Możemy je wykorzystać wszędzie tam, gdzie dysponujemy dużymi ilościami danych, często wzajemnie ze sobą powiązаныmi i pragniemy wydobyć z nich ukryte prawidłowości i reguły.



## Literatura

1. Berry M., Gordon L., *Data Mining Techniques for Marketing, Sales and Customer Support*, John Wiley & Sons, Inc, New York 1998.
2. Berry M., Gordon L., *Mastering Data Mining. The Art and Science of Customer Relationship Management*, John Wiley & Sons, Inc, New York 2000.
3. Berson A., Smith S., Thearling K., *Building Data Mining Applications for CRM*, McGraw-Hill, New York 2000.
4. Poe V., Klauer P., Brobst S., *Tworzenie hurtowni danych*, Wydawnictwo Naukowo-Techniczne, Warszawa 2000.
5. Sokołowski A., Demski T., *Analizy statystyczne i data mining z zastosowaniem oprogramowania StatSoft*, StatSoft Polska, Kraków 2003.
6. Unold J., *Systemy Informacyjne Marketingu*, Wydawnictwo Akademii Ekonomicznej im. Oskara Langego we Wrocławiu, Wrocław 2001.
7. Westphal Ch., Blaxton T., *Data Mining Solutions, Methods and Tools for Solving Real- World Problems*, John Wiley & Sons, Inc, New York 1998.