



CZEGO MOŻNA DOWIEDZIEĆ SIĘ O KLIENTACH, WYKORZYSTUJĄC DATA MINING, CZYLI ANALIZA DANYCH W CRM

Rafał Wajda, StatSoft Polska Sp. z o.o.

Wstęp

Niezależnie od branży klienci wymagają w dzisiejszych czasach zarówno wysokiej jakości produktów (czy usług), jak i obsługi na najwyższym poziomie. Jeśli zaczną odnosić wrażenie, że są źle traktowani lub będą niezadowoleni z kontaktów z naszą firmą z jakichkolwiek powodów, to najprawdopodobniej skorzystają z oferty konkurencji. Należy pamiętać o tym, że zniechęcony klient może bardzo szybko podjąć decyzję o rezygnacji z naszych usług. Dlatego ważna jest „pielęgnacja” relacji z naszymi klientami.

Zmiany zachodzące od pewnego czasu w otoczeniu przedsiębiorców sprawiły, że obniża się poziom lojalności klientów, których dochód rozporządzalny wzrasta. Dodatkowo rynki, które były jeszcze jakiś czas temu lokalne, stają się dostępne dla coraz większej liczby dostawców, m.in. dzięki rozpowszechnianiu się i doskonaleniu nowych, elektronicznych kanałów kontaktowych i dystrybucyjnych.

Podejście, które jeszcze do niedawna było powszechnie stosowane, sprowadzające się często do minimalizacji kosztów lub maksymalizacji zysku (przy czym źródeł efektywności produkcji, alokacji zasobów czy kosztowej szukano najczęściej wewnątrz przedsiębiorstwa i wiązano ją głównie z procesami zachodzącymi wewnątrz organizacji), nie w każdych warunkach jest rozwiązaniem wystarczającym do uzyskania i utrzymania przewagi nad konkurencją. Bardzo rzadko rynek jest wystarczająco chłonny, aby można było sprzedać wszystko, co jesteśmy w stanie wytworzyć. Trzeba walczyć o klientów, skupiając jednocześnie uwagę na tych, którzy mogą przynieść nam najwięcej korzyści.

Identyfikacja klientów, którzy mogą przynieść nam największe korzyści jest bardzo ważna, jednak często bardzo trudna lub wręcz niemożliwa. Oprócz dochodów, jakie uzyskamy dzięki sprzedaży do konkretnego klienta, równie istotne (jeśli nie ważniejsze) okazać się mogą jego lojalność oraz skłonność do rekomendowania nas znajomym, czyli potencjalnym nowym klientom. Najprościej jest obliczyć wartość dotychczasowej sprzedaży dla konkretnego klienta i na tej podstawie prognozować przyszłe dochody, jednak taki wskaźnik może (i najczęściej będzie) nieprawidłowo wskazywać całościową wartość danego klienta dla nas. Jeden klient, który kupi u nas niewiele, może jednocześnie polecać nas dziesiątkom znajomym lub odwrotnie klient kupujący u nas dużo może nie polecać nas



innym. Dlatego niezależnie od wielkości sprzedaży dla poszczególnych odbiorców powinniśmy traktować wszystkich klientów możliwie najlepiej, starać się, aby byli oni zadowoleni z kontaktów z nami, budować z nimi stałe relacje. Takie podejście wymaga traktowania każdego klienta z osobna, indywidualnego reagowania na jego potrzeby. Jest to możliwe, jeśli liczba klientów jest stosunkowo niewielka. W przypadku masowej liczby klientów takie podejście jest niemożliwe, chyba że skorzystamy z odpowiedniego rozwiązania. Pomocne w osiągnięciu tego celu okaże się na pewno zastosowanie metodyki CRM, czyli systemu zarządzania relacjami z klientami. CRM jest realizacją marketingu relacyjnego, który w odróżnieniu od tradycyjnego podejścia stawia w centrum zainteresowania budowanie trwałych relacji z klientem (które jest traktowane jako podstawa osiągnięcia zysków w długim okresie), a nie tylko maksymalizację sprzedaży krótkoterminowej. Stosowanie podejścia skierowanego na klienta jest konieczne, aby móc utrzymać się na rynku, aby nasi klienci nie opuścili nas na rzecz konkurencji.

CRM, czyli zarządzanie relacjami z klientami

Kontaktując się każdego dnia z wieloma klientami, nie można pozwolić sobie na to, aby stracić informacje (często bardzo cenne), jakie możemy uzyskać w ten sposób. Jednocześnie nasze działania powinny zmierzać w kierunku zapewnienia możliwie jak największego zadowolenia naszych klientów. Jest to podstawą do tego, aby stali się oni lojalni względem naszego przedsiębiorstwa, a więc bardziej wartościowi. Zadowolenie klientów można osiągnąć, nawiązując i utrwalając relacje z nimi, które sprawiają, że poczują się oni wyjątkowo traktowani przez nas. W osiągnięciu tych celów pomocna jest metodyka CRM (*Customer Relationship Management*). Jest to podejście wspomagające i optymalizujące proces budowania i utrzymywania relacji z klientami. Na CRM składają się trzy podsystemy:

1. **CRM operacyjny** jest odpowiedzialny za zbieranie/gromadzenie informacji o klientach. Mogą one pochodzić z danych transakcyjnych, opinii klientów (o produktach, usługach, obsłudze itd.) lub innych dostępnych źródeł wewnętrznych i zewnętrznych.
2. **CRM analityczny** jest odpowiedzialny za analizy danych (pochodzących z CRM operacyjnego i innych źródeł). Ta część jest bardzo ważna, ponieważ prawidłowe jej działanie decyduje w znacznej mierze o tym, czy wdrożenie i stosowanie całości systemu CRM pozwoli tylko utrzymać się na rynku, czy także uzyskać przewagę nad konkurencją.
3. **CRM interakcyjny** jest odpowiedzialny za kontakty z klientami wynikające z opracowanych kampanii marketingowych, wyników analitycznego CRM, potrzeb klientów itd.

CRM analityczny

CRM analityczny odpowiedzialny jest za przetwarzanie danych za pomocą różnych narzędzi analizy danych – najczęściej data mining. Wyniki tych analiz stanowią podstawę podejmowania strategicznych decyzji związanych z działaniami przedsiębiorstwa. To ogniwo



CRM jest bardzo ważne, ponieważ umożliwia wykorzystanie wiedzy ukrytej w zgromadzonych danych, których jest każdego dnia coraz więcej. Przedsiębiorstwa zaniedbujące ten etap CRM są narażone na to, że o ile utrzymają się na rynku dzięki wprowadzeniu CRM, to nie będzie on w pełni wykorzystywał potencjału, jaki zgromadzony jest w przechowywanych danych. Ze względu na wagę decyzji, jakie podejmowane są na podstawie wyników analiz danych, należy upewnić się, że wyniki są wiarygodne i właściwe dla rozważanych zagadnień. Przestrzeganie pewnych zasad związanych z **danymi, analizami i organizacją** musi być zapewnione, aby możliwości, jakie daje CRM analityczny były optymalnie wykorzystywane.

Aspekt danych. Analizy można przeprowadzić na każdym zbiorze danych, jednak jeśli jakość danych wejściowych będzie wątpliwa, to na wynikach też nie będziemy mogli polegać. Bardzo ważne jest, by dane charakteryzowały się możliwie najwyższą jakością, czyli były zgodne z rzeczywistością, dokładne, kompletne itd. Spełnienie tych warunków leży głównie w gestii CRM operacyjnego, w zakresie którego zbierane są dane. Wszyscy pracownicy muszą mieć świadomość wagi, jaką należy przywiązywać do jakości danych, które za ich pośrednictwem trafiają do systemów bazodanowych, powinni wiedzieć, że każda informacja może być źródłem bardzo cennej dla przedsiębiorstwa wiedzy.

Dane przedsiębiorstwa przechowywane są zwykle w różnych systemach, różnych bazach danych. Do analiz związanych z CRM analitycznym dane powinny być odpowiednio dobrane i przygotowane. Do tego celu wykorzystuje się analityczną bazę danych (lub korporacyjną hurtownię danych), która integruje dane pochodzące z różnych źródeł (np. systemów transakcyjnych, księgowych, operacyjnego CRM itd.) po to, aby dostarczała dane potrzebne do wykonywanych analiz. Poprawnie przygotowana do tego celu baza danych charakteryzuje się m.in. tym, że:

- ◆ jej zawartość nie zmienia się podczas analizy (aktualizacja następuje w ściśle określonym czasie),
- ◆ na podstawie zgromadzonych w niej danych można generować wielowymiarowe zestawienia,
- ◆ „zwykły” użytkownik nie ma uprawnień do zmiany jej zawartości,
- ◆ jest odizolowana od innych systemów baz danych w organizacji,
- ◆ jednym z wymiarów danych jest czas.

Wyniki analiz (statystycznych, data mining) są tym lepsze, im bardziej dane mają charakter atomowy (nie są zagregowane) i dlatego taką też cechą powinny wykazywać przechowywane w analitycznej bazie dane. Jest to bardzo istotne, ponieważ w zależności od konkretnego zagadnienia wymagana postać danych może być różna. Dane surowe można przekształcić do odpowiednio zagregowanych, ale nie odwrotnie – agregacja i inne przekształcenia często uniemożliwiają powrót do pierwotnej postaci danych.

Aspekt analizy. Poprawnie przeprowadzona analiza danych powinna uwzględniać wszystkie dostępne dane, które mogą mieć znaczenie dla interesujących nas zjawisk. Odpowiednie dane należy analizować za pomocą odpowiednich narzędzi. Ze względu na charakter zagadnień związanych z analitycznym CRM właściwymi technikami są w tym



przypadku metody typu data mining, będące implementacjami algorytmów zoptymalizowanych pod kątem analizy dużych zbiorów danych i poszukiwania w nich ukrytych zależności. Oprócz tych narzędzi często używa się metod „tradycyjnej” statystyki. Narzędzia zgłębiania danych (data mining) powinny w przyjaznym dla użytkownika środowisku oferować wyczerpujący zestaw technik analitycznych potrzebnych na każdym etapie analizy: od przygotowywania danych, ich wstępnej analizy (*preprocessing*), poprzez główną część analiz, po generowanie raportów końcowych. Bardzo istotne jest, aby uzyskiwane wyniki (modele) można było łatwo zastosować dla nowych danych, których nie brano pod uwagę w trakcie analizy. Występowanie nietypowych problemów analitycznych wymaga od systemów data mining wykorzystywanych w analitycznym CRM otwartości – użytkownik musi mieć możliwość łatwego i wygodnego, ale jednocześnie efektywnego dostosowania systemu do swoich potrzeb.

Należy pamiętać o tym, że nawet najbardziej zaawansowane technologicznie narzędzia nie są w stanie odpowiedzieć na pytania, które nie zostały postawione, sprecyzowane. Nie wystarczy wskazanie dostępnych danych i nadzieja, że automat coś znajdzie, policzy, jeśli wcześniej nie określimy problemu, jaki chcemy rozwiązać. To od osób nadzorujących i przeprowadzających analizy zależy w dużej mierze to, czy uzyskiwane wyniki będą się dało wykorzystywać w optymalizacji działań marketingowych, czy właściwe analizy będą przeprowadzane na właściwych danych za pomocą właściwych narzędzi analitycznych, czy ukryte w zgromadzonych danych zależności, które pomogą optymalizować działania marketingowe, zostaną znalezione i poprawnie zidentyfikowane.

Aspekt organizacyjny. Stosowanie metodyki CRM daje największe korzyści, jeśli zgodnie z nią postępują wszystkie części organizmu, jakim jest przedsiębiorstwo, jeśli wprowadzone zostaną odpowiednie zmiany organizacyjne. Ani odpowiednie dane, ani wykorzystanie najlepszych narzędzi data mining i specjalistów w tej dziedzinie nie dadzą (optymalnych) rezultatów, jeśli na podstawie otrzymywanych wyników nie będą podejmowane odpowiednie decyzje i działania. Zamierzona optymalizacja działań marketingowych może być osiągnięta tylko wtedy, gdy CRM obejmie całą organizację, jej pracownicy będą wymieniali się istotnymi informacjami, wyniki analiz (lub decyzje podjęte na ich podstawie) będą docierały do wszystkich, którzy ich potrzebują, a te osoby będą postępowały zgodnie z ich zaleceniami.

Nie należy zapominać, że jest jeden czynnik, który łączy wszystkie te trzy aspekty analitycznego CRM – ludzie. To od nich zależy, czy dane gromadzone będą właściwie, czy podczas przeprowadzania analiz będą wykorzystywane najlepsze dostępne techniki i wreszcie czy otrzymywane wyniki, wiedza wydobyta z danych będą w praktyce stosowane.

Wartość klienta

Często w zagadnieniach związanych z CRM spotykamy się z wartością klienta. Przykładowo jedna z wielu definicji mówi: „*CRM jest strategią mającą na celu zoptymalizowanie całościowej wartości klientów.*”



Jak już wspomniano wcześniej, prawidłowa ocena wartości klienta nie jest łatwa. W najprostszym ujęciu klient jest dla nas tym wartościowszy, im więcej (za większą kwotę) u nas kupuje. Trzeba jednak pamiętać o tym, że przychody (bezpośrednie), jakie uzyskujemy z tytułu sprzedaży dla danego klienta są tylko jednym z wymiarów, jaki powinien być brany pod uwagę przy ocenie. Inne czynniki są związane z przychodami pośrednimi, czyli takimi, które są generowane przez sprzedaż, która została zrealizowana dzięki wpływowi (umyślnemu lub nie) tego klienta na inne jednostki (osoby czy organizacje). Prawie nigdy nie znamy wszystkich szczegółów dotyczących poczynań naszych klientów, czynników wpływających na ich decyzje (związane np. z zakupami). Niekiedy klienci sami nie są w stanie określić, dlaczego zdecydowali się na ten, a nie inny produkt, czy przekonała ich reklama, rekomendacje znajomych, zła opinia o konkurencji, analiza porównawcza z dostępnymi substytutami, chwilowy impuls czy może wszystkie te motywy po trosze. Lista potencjalnych powodów jest praktycznie nieograniczona.

Nie ma jednego i uniwersalnego sposobu oceny wartości klienta. Może okazać się, że wartość klienta zależeć będzie od polityki przedsiębiorstwa, opracowanej strategii działania lub od celów poszczególnych akcji marketingowych. Często stosowanym wskaźnikiem wartości klienta jest LTV (*Life-Time Value*; wartość życiowa klienta), który ma pokazywać łączne potencjalne przychody od danego klienta. LTV jest wyliczany w oparciu o ekstrapolację dotychczasowych przychodów od danego klienta na spodziewany okres jego aktywnego korzystania z naszych usług/produktów lub w oparciu o dane opisujące innych, podobnych klientów. Zarówno LTV, jak i inne tego typu wskaźniki możemy wyliczać za pomocą technik data mining lub innych metod analizy danych. W szczególnych przypadkach przedmiotem naszego zainteresowania może być nie konkretna wartość liczbowa wskaźnika wartości klienta, ale informacja o tym, czy z jakiegoś konkretnego powodu dany klient powinien być dla nas interesujący, jest potencjalnie wartościowy (np. czy istnieje duże prawdopodobieństwo, że skorzysta on z nowej oferty,5 jeśli ją do niego skierujemy). Do tego celu najczęściej wykorzystuje się narzędzia klasyfikacyjne.

Właściwe oszacowanie wartości klienta jest często bardzo istotne i ma wpływ na prawidłową ocenę powodzenia danego przedsięwzięcia marketingowego. Przykładowo, jeśli wytypujemy osoby, do których powinniśmy skierować naszą nową ofertę, ale nie określimy właściwie ich wartości, to może okazać się, że przychody z kampanii nie pokryją związanych z nią kosztów. Z drugiej strony może okazać się, że osoby zaklasyfikowane (prawidłowo) jako te, które nie skorzystają z naszej oferty, zostałyby zachęcane do skorzystania z naszej oferty w przyszłości.

Pomimo tego, że oszacowanie wartości klienta jest zagadnieniem nietrywialnym, to należy tego typu analizy przeprowadzać (wykorzystując dostępne narzędzia analityczne data mining oraz statystyczne), aby lepiej poznawać klientów i uwzględniać możliwie jak najwięcej informacji o nich w procesie podejmowania strategicznych decyzji dotyczących np. akcji marketingowych.



Analizy w CRM, czyli jaką wiedzę możemy wydobyć z danych

Profesjonalne narzędzia data mining dają użytkownikowi dostęp do szerokiego wachlarza technik drążenia danych i wydobywania z nich użytecznej wiedzy. Za ich pomocą możemy analizować zagadnienia, które są szczególnie interesujące, ponieważ pozwalają lepiej poznać naszych klientów i zrozumieć ich zachowania. Najważniejsze analizy związane są z:

- ◆ **segmentacją** klientów, czyli ich podziałem na jednorodne grupy,
- ◆ **cyklem życia klienta**,
- ◆ **koszykami zakupów**, czyli badaniem sekwencji i asocjacji, zgodnie z którymi klienci dokonują zakupów,
- ◆ **satysfakcją** (zadowoleniem) klienta.

Segmentacja

Podział populacji klientów na jednorodne grupy określane jest mianem segmentacji (ang. *clustering*). Wyodrębnienie segmentów pozwala dowiedzieć się, jakich klientów mamy, jakie są typy ich zachowań. Stosując CRM powinniśmy podchodzić do każdego klienta w miarę możliwości indywidualnie. Może to okazać się niemożliwe, jeśli liczba klientów jest zbyt duża. Segmentacja daje podstawy do opracowania planów działania skierowanych do poszczególnych, wyodrębnionych w ten sposób segmentów. Założenie jest takie, że elementy należące do tego samego skupienia (segmentu, klastra) są takie same lub bardzo do siebie podobne i dlatego możemy skutecznie opracować taką samą strategię skierowaną do wszystkich osób z danej grupy. Dobre zaplanowanie takich strategii może sprawić, że klienci będą mieli wrażenie indywidualnego traktowania, że skierujemy naszą ofertę do tych osób, które na nią odpowiedzą (z największym prawdopodobieństwem), że ograniczymy kontakty z klientami, którzy są z jakiegoś powodu mało interesujący.

W zależności od rodzaju analizowanych danych można wyróżnić segmentację demograficzną, behawioralną (dane o zachowaniach klientów), pod względem motywacji (przyczyny określonych działań). Zakres zmiennych, jakie bierzemy pod uwagę przy podziale populacji na jednorodne grupy, należy dobrać odpowiednio do kontekstu, w jakim chcemy ją przeprowadzić.

Metody segmentacyjne zalicza się do nieukierunkowanego data mining (bez nauczyciela). Oznacza to, że na początku analizy nie ma zmiennej wskazującej podział. Otrzymywane wyniki (liczba segmentów i przynależność do nich poszczególnych elementów) zależą od zestawu zmiennych oraz wybranej metody analitycznej.

Wyniki segmentacji można wykorzystać do dalszych analiz. Jeśli podzielimy populację na jednorodne podgrupy (zawierające podobne jednostki, przypadki), oszacowane modele dla każdej z nich z osobna dadzą znacznie lepsze dopasowanie niż jeden model, wspólny dla całej populacji. Dzięki temu stosowanie takiego rozwiązania (różnych modeli poszczególnych segmentów) da znacznie lepsze rezultaty.



Segmentacja pokazuje najczęściej występujące wśród naszych klientów profile. Możemy szybko przyrzeć się charakterystycznym cechom poszczególnych segmentów i dzięki temu określić te grupy, które są dla nas wartościowe lub nie.

Cykl życia klienta

Budując relację z klientem, powinniśmy pamiętać, że można na osi czasu wyróżnić etapy jego kontaktów z nami. Kolejne stadia to klient:

potencjalny → aktualny → zagrożony odejściem → były

Jest to tzw. cykl życia klienta. W zależności od tego, w jakim stadium znajduje się dany klient, należy wykonywać inne analizy prowadzące do maksymalizacji naszych korzyści. Działania przedsiębiorstwa powinny zdążać w tym kierunku, aby klienci potencjalni stali się aktualni (żeby skorzystali z naszej oferty), zagrożeni odejściem nie stali się byłymi, aktualni stali się klientami stałymi itd. Podejmując decyzje dotyczące działań skierowanych do klientów znajdujących się w różnych stadiach, należy pamiętać o tym, że:

- ◆ stali klienci są bardziej wartościowi niż incydentalni,
- ◆ łatwiej (i taniej) jest zatrzymać aktualnego klienta, niż odzyskać go od konkurencji lub pozyskać zupełnie nowego,
- ◆ łatwiej sprzedać nowy lub dodatkowy produkt/usługę naszemu klientowi niż osobie nieznaney.

Różne etapy cyklu życia klienta wiążą się z różnymi zagadnieniami, jakie można rozważać, z innymi analizami jakie mogą wspomagać podejmowanie decyzji.

Potencjalni klienci

Pozyskiwanie nowych klientów jest bardzo ważne dla organizacji. Są oni źródłem dodatkowego dochodu. W zależności od sytuacji rynkowej ich pozyskiwanie może być stosunkowo łatwe (np. jeśli rynek jest nienasycony, mamy do czynienia z wieloma potencjalnymi klientami) lub utrudnione (np. w przypadku nasyconego rynku, nasz nowy klient to często były klient konkurencji). Informacje o osobach (lub organizacjach), które mogą być dla nas potencjalnymi klientami pochodzą najczęściej z zewnętrznych źródeł, które są nieprecyzyjne, niekompletne itd. Prognozowanie na ich podstawie może być trudne i obarczone dużym błędem (m.in. dlatego w takich przypadkach kampanie marketingowe charakteryzują się małą skutecznością). Niemniej jednak można podejść do tego zagadnienia niejako z drugiej strony. Jeśli widzimy, że znalezienie osób, które będą skłonne odpowiedzieć na naszą ofertę jest mało prawdopodobne lub niemożliwe, to staramy się znaleźć tych, którzy na pewno nie skorzystają z naszej oferty. Taka wiedza byłaby bardzo cenna ze względu na możliwość ograniczenia kosztów związanych z dotarciem do dużej liczby osób (koszty wysyłki, materiałów itp.).



Aktualni klienci

Tę grupę klientów można podzielić na następujące podgrupy:

- ♦ **nowi** – wyrażają zainteresowanie naszą ofertą (odpowiedzieli na nią) lub zakupili nasz produkt/usługę po raz pierwszy,
- ♦ **incydentalni** – pierwszy kontakt z nimi mamy już za sobą, jednak nie możemy uznać ich za klientów stałych, kupują coś rzadko, najczęściej nie są przywiązani tylko do naszej marki,
- ♦ **stali** – ci klienci kupują u nas stosunkowo często,
- ♦ **lojalni** – są przywiązani do naszych produktów/usług.

O klientach z tej grupy mamy zazwyczaj pełniejsze informacje niż w przypadku potencjalnych klientów. Do danych demograficznych (zweryfikowanych i uzupełnionych podczas pierwszych kontaktów) dochodzą informacje o niektórych zachowaniach (np. co zostało zakupione, w jakiej kolejności). Możemy także zdobyć informacje o cechach osobowości klienta, które mogą mieć wpływ na podejście, jakie powinniśmy w stosunku do niego zastosować. Te informacje mogą także podpowiadać, jakiego rodzaju argumenty najbardziej przekonują danego klienta.

Na początku powinniśmy określić jaki, może dla nas być dany klient: wysoce dochodowy, przeciętny, mało opłacalny, czy też może będzie narażał nas na straty, czyli określić wartość klienta. Wyniki takiej analizy mogą mieć istotne znaczenie dla czynności, jakie powinniśmy podejmować w stosunku do nowych klientów. Należy pamiętać o tym, że każdy klient może być dla nas w przyszłości dochodowy i nie powinniśmy „ucinać” kontaktów z tymi, których ocenimy na początku jako bardzo nisko dochodowych lub przynoszących straty. Po pierwsze nasza ocena może być nieprecyzyjna lub po prostu błędna, a po drugie klient może z czasem zmienić „status” i nawet jeśli nie będzie nam na początku przynosił (dużych) zysków, to powinniśmy w miarę możliwości utrzymywać z nim dobry kontakt, po to by w przypadku zmian w przyszłości był on skłonny korzystać z naszej oferty, a nie oferty konkurencji.

W stosunku do klientów, z którymi udało nam się nawiązać relacje, możemy pójść dalej. Najlepiej pokazać, że rozumiemy naszego klienta i wychodzimy naprzeciw jego potrzebom i oczekiwaniom. Nie jest to łatwe, ale dysponując odpowiednimi danymi, możemy tego dokonać. Najlepiej jest zastosować tu techniki wspomagające sprzedaż krzyżową (cross-selling) i sprzedaż rozszerzoną (up-selling).

Cross-selling polega na proponowaniu naszym klientom usług lub produktów, które są często kupowane przez innych klientów w połączeniu z tym, co dany klient zakupił lub czym wyraża zainteresowanie. Jednym z wielu przykładów stosowania tej techniki są sklepy internetowe – kupując lub tylko przeglądając informacje o jakimś produkcie automatycznie prezentowana jest lista towarów, które zostały zakupione łącznie z tą pozycją przez innych klientów serwisu, lub które także zostały zakupione przez osoby o podobnej historii zakupów. W tym przypadku dane, jakie podlegają analizie, dotyczą dotychczasowych zakupów naszych klientów. Inna sytuacja to zakupy sekwencji produktów w odstępach



czasu. Klient, który zakupił profesjonalny sprzęt lub oprogramowanie, może za jakiś czas być zainteresowany szkoleniem (dla siebie lub pracowników), które pozwoli optymalnie wykorzystywać możliwości danego produktu. Po takim szkoleniu, za jakiś czas może uznać, że wykorzystuje już w pełni dany sprzęt i jest gotowy na rozszerzenie jego funkcjonalności itd. Taka ścieżka wydarzeń jest przewidywalna i analiza odpowiednich danych transakcyjnych wykaże opisaną prawidłowość, jeśli taka rzeczywiście występuje. Trzeba pamiętać o tym, że przeprowadzając tego typu analizę jest dla nas najistotniejsze, aby zweryfikować oczywiste reguły (stwierdzić czy rzeczywiście one występują, czy też tylko wydają się one być prawdziwe) oraz znaleźć reguły nieintuicyjne – takie których nie widać na pierwszy rzut oka, a które wyłaniają się dopiero po dogłębnej analizie zgromadzonych danych.

Up-selling polega na sprzedaży klientom rozszerzonych, lepszych wersji produktów lub usług, z których już korzystają. Przykładem może być klient firmy telekomunikacyjnej, można mu zaproponować przejście na wyższy abonament, a jeśli jest to firma, to także dodatkową linię telefoniczną, dodatkowy numer w sieci komórkowej. Analiza może wskazać klientów, którzy będą skłonni skorzystać z tego typu rozszerzeń i ulepszeń.

Kampanie marketingowe mające na celu cross-selling i up-selling, jeśli są prawidłowo planowane i przeprowadzane, są pomocne w budowaniu relacji z klientami. Korzyści wynikające z takiej kampanii są obopólne: oprócz oczywistych, wynikających ze zwiększonej sprzedaży, przedsiębiorstwo utrwala jednocześnie relacje z klientami (jeśli trafimy z proponowanymi produktami lub usługami w ich potrzeby, będą oni mieli wrażenie, że ich rozumiemy, wychodzimy im naprzeciw), którzy uzyskują informacje o produktach lub usługach, którymi mogą być zainteresowani i dzięki temu mogą zaoszczędzić wysiłek związany z poszukiwaniami odpowiednich informacji. Takie kampanie często wiążą się ze specjalnymi ofertami dla klientów, promocjami, upustami itp., co może dodatkowo zwiększyć zadowolenie klienta.

Za pomocą narzędzi analitycznych data mining można prognozować, którzy z naszych klientów będą skłonni skorzystać z oferty sprzedaży krzyżowej lub rozszerzonej. W poszukiwaniu reguł, zgodnie z którymi postępują klienci podczas zakupów, pomocne są narzędzia wykrywające asocjacje i sekwencje, jakie często występują w zgromadzonych danych transakcyjnych, a więc w historii zakupów. Planując kampanie marketingowe związane z cross-sellingiem, najpierw budujemy po jednym modelu dla każdego z oferowanych produktów, a następnie optymalizujemy wyniki w celu maksymalizacji łącznych korzyści (sprzedającego i kupującego), jakie mogą przynieść przygotowane oferty i związana z nimi akcja marketingowa. Nie zawsze należy oferować wszystko, co dany klient może chcieć kupić. Takie postępowanie może go zniechęcić, ponieważ zajmiemy mu zbyt dużo czasu, wyślemy zbyt dużo wiadomości itd. W przypadkach, w których analiza pokazuje nam, że dany klient może być zainteresowany wieloma różnymi produktami można na początku zaproponować mu te, którymi będzie najbardziej zainteresowany, a resztę zaoferować w późniejszym terminie lub po prostu spytać go, czy chce otrzymać także inne oferty. Prognozy należy sporządzać z odpowiednim wyprzedzeniem, aby oferta dotarła do klienta na czas, zanim nawiąże on kontakt z konkurencją i podejmie decyzję o zakupie.



Klienci zagrożeni odejściem

Jednym z zadań systemu CRM jest budowanie i utrwalanie relacji z klientami. Jeśli cel ten zostanie osiągnięty, to klient stanie się prawdopodobnie stałym i lojalnym klientem. Tacy klienci są bardzo cenni, ponieważ przy niewielkim wysiłku ze strony sprzedawcy przynoszą oni wiele korzyści (nie tylko w formie przychodów, są oni często także źródłem referencji). Dlatego należy dbać o nich, o ich zadowolenie i przez to zapewnić ich lojalność, szczególnie gdy przejście do konkurencji jest łatwe i nie wiąże się z dodatkowymi kosztami dla klienta. Trzeba pamiętać, że jeśli tylko będzie on niezadowolony z kontaktów z nami lub konkurencja zaproponuje lepsze rozwiązanie, to szybko możemy go stracić. Odejścia (ang. *churn*) klientów stanowią problem szczególnie wtedy, gdy rynek nasycy się i pozyskanie nowych odbiorców jest coraz trudniejsze.

Możemy wyróżnić dwa typy zerwania kontaktu z klientem. W pierwszym przypadku to klient decyduje o odejściu (przedsiębiorstwo chce nadal świadczyć dla niego usługi). Może to być wywołane zmianą warunków demograficznych (zmiana miejsca zamieszkania, stylu życia, upodobań itd.), rynkowych (np. lepsza oferta konkurencji) lub innych. Z drugiej strony to my możemy chcieć zakończyć naszą współpracę z klientem. Dzieje się tak najczęściej wtedy, gdy klient nie wywiązuje się ze swoich zobowiązań i naraża nas na straty (np. przestaje płacić abonament za telefon, spłacać raty kredytu itd.). W obu przypadkach przedsiębiorstwo powinno podjąć odpowiednie działania – im wcześniej, tym lepiej. Narzędzia data mining pomagają identyfikować klientów obu tych grup.

Działania zapobiegające odejściom wartościowych klientów mogą być wspierane różnymi analizami danych związanymi z pomiarem i monitorowaniem satysfakcji i zadowolenia. Za pomocą narzędzi statystycznych i data mining można szukać przyczyn wpływających na decyzje klientów o odejściu. Jeśli uda nam się taką wiedzę wydobyć ze zgromadzonych danych, możemy starać się zapobiegać tym zmianom, podejmować działania, które wpłyną na wzrost lojalności klientów wobec nas, przekonują ich, że powinni nadal korzystać z naszych usług.

Analizy związane z odejściami klientów można sprowadzić do szukania odpowiedzi na dwa pytania:

1. Jakie dla danego klienta jest prawdopodobieństwo odejścia (lub tego, że zacznie generować straty)?
2. Jakie są przyczyny odejść?

Modele generujące odpowiedzi na pierwsze pytanie są budowane w oparciu o dane dotyczące naszych byłych i aktualnych klientów, najczęściej za pomocą metod ukierunkowanego data mining. Takie metody wymagają wskazania zmiennej, jaką chcemy prognozować, czyli w tym przypadku – oszczędz. Tak/Nie. Model na podstawie danych o osobach, dla których możemy określić wartość takiej zmiennej identyfikuje cechy charakterystyczne rezygnujących. Te reguły są podstawą do oceny prawdopodobieństwa odejścia dla pozostałych klientów. Dzięki nim w pierwszej kolejności możemy skupić uwagę na klientach, dla których to prawdopodobieństwo jest największe, i widzimy jeszcze szansę ich zatrzymania. W niektórych przypadkach może okazać się, że danych o interesujących nas



z punktu widzenia analizy osobach (które zrezygnowały z naszych usług lub zaczęły generować straty) jest zbyt mało. Przykładem mogą być działy banków zajmujące się przyznawaniem kredytów. Bardzo cennym jest dla nich model wskazujący potencjalnych kredytobiorców, którzy nie spłacą lub będą mieli problemy ze spłaceniem zaciągniętego kredytu. Ze względu na ostrożność w udzielaniu ryzykownych kredytów może okazać się, że brakuje danych o „złych” (niespłaconych) kredytach, że ich liczba jest zbyt mała, aby zbudować dobrze działający model. W takich przypadkach możemy sztucznie dodać przypadki niespłaconych kredytów, np. „klonując” te dostępne (jednak taki sposób działania może przynieść niepożądane efekty), określić większy koszt błędnych klasyfikacji dla przypadków „złych” kredytów, zastosować v-krotny sprawdzian krzyżowy.

Szukając odpowiedzi na drugie pytanie, czyli o przyczyny odejść (lub cechy sygnalizujące, że mogą wystąpić problemy z wywiązywaniem się klienta ze zobowiązań), można wykorzystywać zarówno techniki nieukierunkowanego, jak i ukierunkowanego data mining. Wyniki takich analiz są kluczowe w podejmowaniu właściwych decyzji dotyczących kontaktów z klientami, którzy mogą nas w najbliższym czasie opuścić (lub narazić na straty).

W przypadku analiz związanych z klientami zagrożonymi odejściem (lub niewywiązywaniem się ze zobowiązań) prognozy muszą być sporządzane z odpowiednim wyprzedzeniem czasowym. Zbyt późny sygnał nie da możliwości podjęcia odpowiednich działań – część klientów, których moglibyśmy zatrzymać, stanie się byłymi klientami, a znacznie trudniej jest nakłonić naszego byłego klienta do ponownego skorzystania z naszych usług, niż zatrzymać zagrożonego odejściem.

Byli klienci

Analizy związane z grupą byłych klientów mają na celu wskazanie tych osób, do których powinniśmy ponowić ofertę, ponieważ mogą oni na nią odpowiedzieć. Klienci, którzy zrezygnowali z naszych usług, są grupą, o której mamy wiele informacji pochodzących z dotychczasowych kontaktów. Celem działań akcji skierowanych do tych klientów jest reaktywacja naszych kontaktów z nimi, czyli sprawienie, aby znów stali się naszymi klientami aktualnymi.

Analizy „koszyków zakupów” – sekwencje i asocjacje

Znajomość wzorców postępowania klientów podczas dokonywania zakupów jest bardzo cenna, ponieważ pozwala efektywnie planować działania mające na celu zwiększanie sprzedaży. Data mining może pomóc w wykrywaniu takich reguł. Najczęściej szukamy odpowiedzi na pytania dotyczące:

- ◆ łącznego zakupu poszczególnych produktów,
- ◆ sekwencji zakupów, czyli w jakiej kolejności po sobie kupowane są poszczególne produkty (i ewentualnie w jakich odstępach czasu).



Podstawą do tego typu analiz są historyczne dane transakcyjne dotyczące zakupów. W rezultacie otrzymujemy szereg reguł typu „klienci, którzy zamówili produkt A, często zamawiają produkt B lub C”, „pracownicy, którzy dobrze oceniają rozwiązanie X, również często narzekają na rozwiązanie Y, ale zadowolą ich rozwiązanie Z” itp. Znajomość takich reguł pozwala nie tylko poznać wzorce zachowań klientów podczas zakupów, ale także wyciągać dodatkowe wnioski, dotyczące np. motywów zakupów.

Tego rodzaju analizy często określa się mianem analiz koszykowych, ponieważ najpopularniejsze ich zastosowanie polega na analizie grup produktów, które trafiają do jednego koszyka podczas zakupów w sklepach samoobsługowych. Ten konkretny przypadek zastosowania technik analitycznych może dać podstawę do optymalnego rozmieszczenia towaru na półkach sklepowych. Można przykładowo często kupowane razem produkty umieścić obok siebie w celu zwiększenia ich łącznej sprzedaży lub wręcz przeciwnie – oddalić je od siebie licząc na to, że klient, szukając drugiego produktu z takiej pary, trafi przy okazji na inne produkty, które też kupi. Analizując dostępne dane, algorytm szuka najczęściej występujących kombinacji, ale także pojedynczych produktów. Przykładem może być pieczywo kupowane często niezależnie od reszty zawartości koszyka – na pewno każdy, będąc w sklepie samoobsługowym, zwrócił uwagę na to, że stoisko z chlebem usytuowane jest daleko od wejścia, a więc klient, chcąc kupić tylko pieczywo, jest zmuszony przejść przez cały sklep, obok wielu rzędów półek z produktami, które może wrzuci dodatkowo do koszyka.

Techniki określane jako analizy koszykowe znajdują wiele różnych zastosowań. Oprócz poszukiwania reguł, zgodnie z jakimi kupowane są razem produkty (analiza asocjacji), możemy także szukać zależności zakupowych w czasie (analiza sekwencji i połączeń). Metody te są zaliczane do nieukierunkowanego data mining (bez nauczyciela). Oprócz tego można także stosować techniki ukierunkowane, np. drzewa klasyfikacyjne. Wszystkie te analizy mają szereg zastosowań, np. we wspomnianych już przy okazji cyklu życia klienta i cross-sellingu sklepach internetowych i wielu innych przedsiębiorstwach w zagadnieniach sprzedażowych, np. bankowości i finansach, telekomunikacji, marketingu itd.

Pomiar satysfakcji klientów

Satysfakcja klienta to poziom zadowolenia z całokształtu kontaktów z przedsiębiorstwem. Ze względu na to, że jest to subiektywna ocena (świadoma lub nie) klienta o nas, naszej działalności i produktach, wpływać na nią może praktycznie wszystko, co tylko kojarzy się z produktem czy też jego dostawcą. Duże znaczenie mają pierwsze kontakty – wymiana informacji przed zakupem (precyzowanie potrzeb, oferta). Klient może oczekiwać z jednej strony dokładnych i prawdziwych informacji, a z drugiej miłej obsługi, zrozumienia jego potrzeb i wymagań. Gdy już uda nam się go przekonać do zakupu, na jego satysfakcję będzie miało wpływ na pewno zgodne z oczekiwaniami działanie zakupionego rozwiązania i związane z nim usługi dodatkowe (serwis, pomoc techniczna itp.), ale również prestiż, jaki wiąże się z posiadaniem danego produktu.



Im wyższy poziom satysfakcji klienta, tym częściej i chętniej korzysta on z naszych usług i odwrotnie, dlatego pomiar i monitorowanie poziomu satysfakcji klientów są bardzo ważne dla organizacji. Tego typu wskaźniki informują, jak jesteśmy postrzegani przez najbliższe otoczenie – osoby, które nam zaufały i wybrały nasze produkty. Przez cały czas powinno się dokładać wszelkich starań, aby satysfakcja była na możliwie najwyższym poziomie – oczywiście w granicach rozsądku i opłacalności. Takie wskaźniki dla poszczególnych klientów mówią o tym, jak postrzegają nas jednostki, natomiast zagregowane dają całokształt naszego obrazu w oczach innych. W zależności od sytuacji oba te rodzaje informacji mogą być bardzo cenne. Jeśli łącznie ocenimy poziom ogólnego zadowolenia naszych klientów jako zadowalający, ale jednocześnie zauważymy grupę pojedynczych klientów (których oceniliśmy jako wartościowych) o spadającym poziomie satysfakcji, to jest to sygnał do podjęcia odpowiednich działań do nich skierowanych. W sytuacji gdy ogólne zadowolenie naszych wszystkich klientów spada lub już jest niskie, należy zastanowić się nad przyczynami tego zjawiska i podjąć decyzje zmierzające do zmian w organizacji.

Dokonując pomiaru satysfakcji klienta, mamy przy okazji sposobność zdobycia o nim wielu cennych danych. Wynikiem takiej analizy może być nie tylko wskaźnik odzwierciedlający satysfakcję, ale także dodatkowe informacje o przyczynach zadowolenia lub niezadowolenia z naszych usług, a więc naszych silnych i słabych stronach. W celu pomiaru i monitorowania satysfakcji klientów przeprowadza się badania marketingowe, które pomagają określać przyczyny określonych postaw czy zachowań.

Statystyka dostarcza całej gamy narzędzi analitycznych, bez których badania marketingowe byłyby znacznie mniej efektywne. Najczęstsze analizy związane z badaniami ankietowymi to wyznaczenie optymalnej liczności próby (minimalnej liczby ankiet/respondentów, która zapewni statystyczną istotność otrzymanych wyników), opracowanie odpowiednich pytań. Otrzymywane wyniki są poddawane analizom pozwalającym uzyskać odpowiedzi na nurtujące nas pytania, odpowiedzi, które mogą być ukryte w zgromadzonych danych.

Pomiar satysfakcji daje największe korzyści, jeśli jest stale powtarzany procesem, a nie jednorazowym eksperymentem. Jest to konieczne, aby móc oceniać skuteczność budowania relacji z klientami przez naszych pracowników i w porę reagować na ewentualny spadek satysfakcji.

Data mining w CRM

Do powyżej opisanych analiz można stosować różne techniki analityczne, jednak najlepsze do tych celów są metody data mining. Wykorzystują one algorytmy zoptymalizowane pod kątem przetwarzania ogromnych zbiorów danych i wyszukiwania w nich ukrytych zależności, są nastawione na praktyczne stosowanie otrzymywanych wyników. Powyżej przedstawiono podział ze względu na wykorzystanie analiz w CRM w różnych zagadnieniach. Inny podział technik data mining obejmuje metody ukierunkowane (z nauczycielem) i nieukierunkowane (bez nauczyciela).



Ukierunkowany data mining wymaga określonej zmiennej objaśnianej, której wartości są nam znane dla przypadków, na podstawie których chcemy szacować model. Innymi słowy, jeśli chcemy uzyskać model za pomocą metod ukierunkowanego data mining, jego parametry szacujemy na podstawie już zaobserwowanych wartości zmiennej, którą chcemy prognozować. Aby przewidzieć, czy dany klient odpowie na daną ofertę, musimy mieć informacje o innych klientach (odpowiedzieli lub nie), do których wysłaliśmy już taką samą lub podobną ofertę. W celu określenia, czy osoba ubiegająca się o kredyt będzie z dużym prawdopodobieństwem planowo go spłacała, musimy mieć dane dotyczące spłacalności kredytów przez inne, podobne do niej osoby. Metody ukierunkowanego data mining wykorzystują informacje o obiektach, w przypadku których interesujące nas zjawisko zostało już zaobserwowane, i na ich podstawie uczą się (stąd nazwa – z nauczycielem) i ustalają reguły, które później możemy stosować dla nowych danych. W ukierunkowanym data mining wykorzystujemy takie techniki, jak: drzewa decyzyjne (klasyfikacyjne i regresyjne, np. C&RT, CHAID, drzewa wzmacniane, losowy las), sieci neuronowe, *MARSplines*, naiwną metodę Bayesa, metodę wektorów nośnych (SVM) i inne.

Nieukierunkowany data mining (bez nauczyciela) służy celom eksploracyjnym (poznawczym). Na wstępie analizy nie precyzujemy parametru, który nas szczególnie interesuje – przedmiotem analizy jest cała populacja i występujące w niej prawidłowości oraz podziały, które chcemy odnaleźć, aby lepiej zrozumieć zachodzące procesy, zachowania klientów itd. W omawianych w niniejszym artykule zagadnieniach najczęstszymi obszarami zastosowań technik data mining bez nauczyciela jest segmentacja oraz analizy związane z koszykami zakupów (asocjacji, sekwencji, połączeń). Przykład segmentacji przedstawiony jest poniżej.

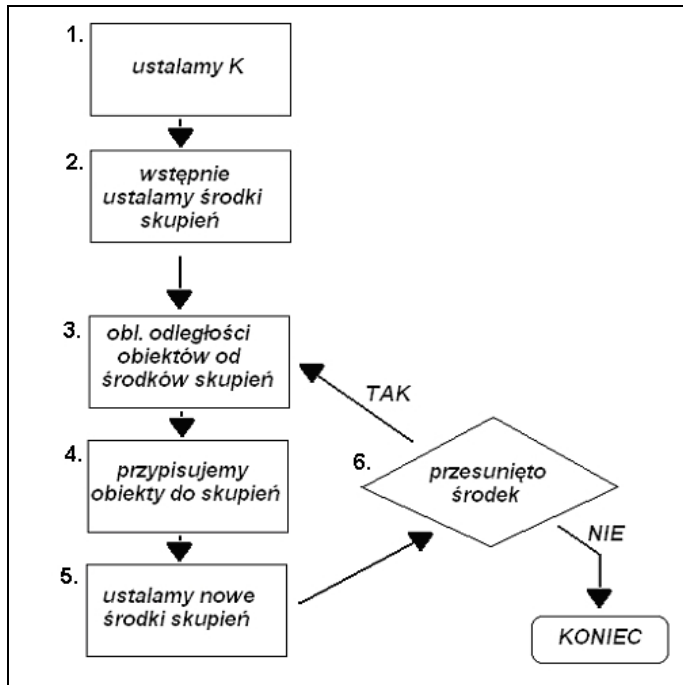
Segmentacyjne metody data mining

Przystępując do analizy segmentacyjnej, najczęściej nie dysponujemy informacją o tym, jakie segmenty występują w danych, ani też ile jest tych segmentów (choć możemy mieć co do tego pewne przeczucia). Wiedzę tę pragniemy dopiero wydobyć z danych. Jak już wcześniej wspomniano, segmentację przeprowadzamy, wykorzystując metody nieukierunkowanego data mining. Najczęściej stosowanymi tu metodami są analiza skupień – metody k-średnich i EM oraz sieci neuronowe Kohonena.

Analiza skupień – metoda k-średnich i EM

Analiza skupień (ang. *cluster analysis*) łączy elementy (obiekty, np. klientów) w grupy – elementy podobne do siebie (bliskie) trafiają do tych samych grup, różniące się do odrębnych. W wyniku takiego działania niejednorodny zbiór jest dzielony na grupy. Analiza skupień może być przeprowadzana za pomocą różnych algorytmów, najważniejsze metody to: k-średnich oraz EM. Algorytmy te wymagają podania na wstępie analizy liczby skupień (segmentów). Można ominąć ten wymóg, stosując tzw. v-krotny sprawdzian krzyżowy, dzięki któremu algorytm sam dobiera optymalną liczbę skupień (taka opcja jest dostępna w *STATISTICA Data Miner*).

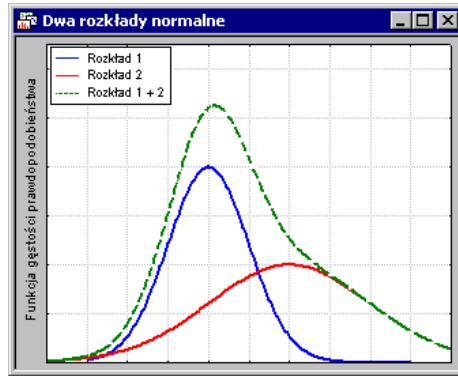
Algorytm **k-średnich** działa następująco:



1. Ustalana jest liczba skupień (może być dobierana automatycznie, jeśli wykorzystywane oprogramowanie analityczne daje taką możliwość).
2. Wybierane są wstępne środki skupień spośród obiektów analizowanego zbioru. Może to być realizowane na różne sposoby, np. losowo lub poprzez dobór elementów najbardziej od siebie oddalonych.
3. Obliczane są odległości pozostałych elementów od wytypowanych (w kolejnej iteracji) środków skupień. W zależności od analizowanych danych i innych przesłanek typy odległości tu obliczanych mogą być różne, np. euklidesowa, jej kwadrat, Manhattan (miejska).
4. Na podstawie odległości od środków każdy z obiektów jest przypisywany do grupy o najbliższym mu środku.
5. Oblicza się nowe środki każdego skupienia na podstawie średnich arytmetycznych ze współrzędnych należących do nich obiektów.
6. Algorytm sprawdza, czy środki ustalone w punkcie 5. pokrywają się z poprzednimi. Jeśli nie, to następuje powrót do punktu 3. i wykonywany jest kolejny przebieg algorytmu. W przeciwnym razie algorytm kończy pracę i mamy gotowy podział analizowanej populacji na segmenty.

Metoda EM (*Expectation Maximization*) jest także nazywana analizą skupień bazującą na prawdopodobieństwie lub statystyczną analizą skupień. Program wyznacza skupienia, zakładając różnorodne rozkłady prawdopodobieństwa zmiennych uwzględnianych

w analizie. Algorytm EM ma za zadanie określić parametry rozkładów segmentów na podstawie rozkładu całej grupy oraz określić prawdopodobieństwo przynależności poszczególnych obiektów do wyodrębnionych segmentów.



Na powyższym rysunku przedstawiono prosty przykład dla zobrazowania tej idei. Jeśli założymy, że chcemy wyodrębnić dwa segmenty ze względu na jedną cechę, której obserwowany rozkład jest przedstawiony przez linię przerywaną (1+2), to algorytm będzie szukał parametrów rozkładów 1 (bardziej wysmukły, o średniej zbliżonej do rozkładu 1+2) i 2 (w tym przykładzie założono, że badana cecha ma w obu segmentach rozkład normalny, co widać na rysunku). Algorytm EM może dokonywać klasyfikacji również przy założeniu innych rozkładów. Istotną różnicą wyników uzyskiwanych w przypadku zastosowania tego algorytmu w porównaniu z wynikami uzyskiwanymi metodą k-średnich jest to, że nie dostajemy jednoznacznego przypisania poszczególnych obiektów do segmentu. W tym przypadku generowana jest tabela, w której zawarte są prawdopodobieństwa przynależności każdego obiektu do każdej z wyodrębnionych grup. Jest to o tyle istotne, że w wątpliwych przypadkach mamy świadomość tego, które obiekty leżą niejako na granicy (lub w części wspólnej) dwóch segmentów, mogą należeć z dużym prawdopodobieństwem do obu z nich.

Segmentacja klientów banku – przykład analizy w CRM

Stosując CRM, należy do każdego klienta podchodzić indywidualnie. Jest to bardzo trudne, jeśli obsługujemy dużą liczbę odbiorców naszych produktów czy usług. Mając kontakty z wieloma tysiącami klientów, nie ma możliwości, aby dla każdego z nich z osobna opracowywać plan działania. W takich sytuacjach bardzo efektywne okazują się techniki segmentacji klientów. Poprawnie przeprowadzona analiza segmentacyjna daje możliwość lepszego poznania klientów i opracowania działań, dzięki którym będziemy umacniali nasze relacje z nimi.

Poniżej przedstawiony przykład segmentacji został przeprowadzony na podstawie danych dotyczących około 4,5 tys. klientów banku. Dostępny zbiór zawierał informacje demograficzne i transakcyjne, m.in. zapis wszystkich operacji (wpłaty, wypłaty, przelewy)



w pewnym okresie czasu, informacje dotyczące kredytów (historia spłat itp.). Do analizy wykorzystano część tego zbioru – dane transakcyjne, i na ich podstawie podzielono populację klientów banku na segmenty. Ze względu na charakter danych tę segmentację można określić jako behawioralną (dane transakcyjne opisują zachowania klientów). W analizie posłużono się analizą skupień uogólnioną metodą k-średnich. Obliczenia wykonano w *STATISTICA Data Miner*.

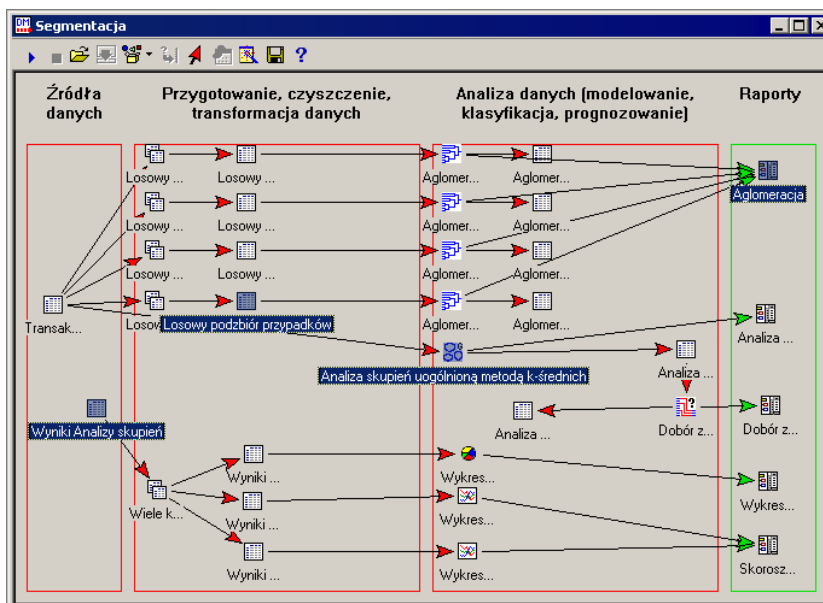
Celem analizy był podział klientów banku na jednorodne segmenty oraz określenie ich profili, ponieważ dzięki temu można doskonale uzupełnić wiedzę o klientach. Segmentacja pozwala spojrzeć na nich niejako z „lotu ptaka”, poznać ich strukturę, cechy poszczególnych grup, występujące różnice i podobieństwa. Na podstawie informacji o jednorodnych podgrupach klientów można lepiej planować działania marketingowe kierowane do poszczególnych segmentów, dostosowywać ofertę do modelu zachowania klienta.

Przygotowanie danych

Dostępne dane transakcyjne przetworzono (dokonano ich *preprocessingu*) tak, aby możliwie najłatwiej można było przeprowadzić analizę. W wyniku tego zabiegu otrzymano arkusz danych zawierający 4 500 przypadków (klientów) oraz 25 zmiennych: 12 określających sumy kwot, na jakie zrealizowano danego rodzaju transakcje, 12 opisujących liczbę transakcji danego rodzaju, 1 identyfikująca klienta (id).

Analiza

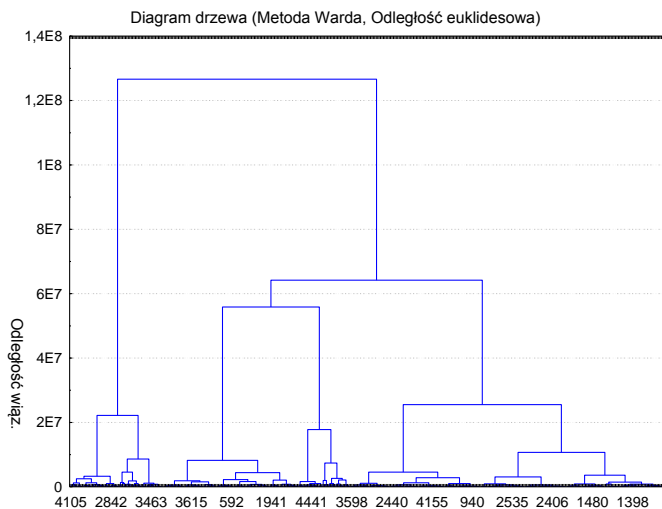
W celu przeprowadzenia analizy utworzono projekt w środowisku graficznym *STATISTICA Data Miner* (przedstawiony poniżej).





Zdecydowano, że segmentacja przeprowadzona zostanie za pomocą analizy skupień uogólnioną metodą k-średnich. Jak już wcześniej wspomniano, w *STATISTICA Data Miner* możemy użyć v-krotnego sprawdzianu krzyżowego, dzięki któremu wybrana zostaje optymalna liczba skupień (podajemy przedział liczby skupień, który będzie brany pod uwagę). Idea v-krotnego sprawdzianu krzyżowego polega na tym, że analizowane obiekty dzielone są na v podgrup. Na podstawie v-1 z nich szacujemy podział na skupienia (segmenty), natomiast podgrupę, która nie brała udziału w budowie modelu, traktujemy jako testową (służy ona ocenie trafności przyporządkowywania obiektów do wyznaczonych segmentów). Czynność ta jest powtarzana v-krotnie (stąd nazwa), za każdym razem inna spośród v podgrup jest testową. Wyniki z kolejnych v powtórzeń są uśredniane i dają jedną ocenę stabilności modelu, tzn. trafności przypisywania obserwacji do skupień. Jako jeden z parametrów dla działania algorytmu podajemy minimalny spadek procentowy funkcji błędów (dla metody k-średnich jest to średnia odległość przypadków od centrów skupień), na jego podstawie oceniana jest optymalna liczba skupień. W prezentowanym przypadku parametr ten ustawiono na 5%.

Przed analizą nie było wiadomo nic na temat ewentualnych podziałów w badanej populacji, więc w celach eksploracyjnych przeprowadzono analizę aglomeracyjną z różnymi ustawieniami, dla prób losowanych z dostępnych danych. Dzięki temu zidentyfikowany został przedział poszukiwań (od 2 do 8 skupień). Poniżej przedstawiony jest diagram drzewa (dendrogram) będący wynikiem podziału metodą aglomeracyjną (jednego z wariantów).



Po takich zabiegach przeprowadzono analizę, której wyniki przedstawiono poniżej.

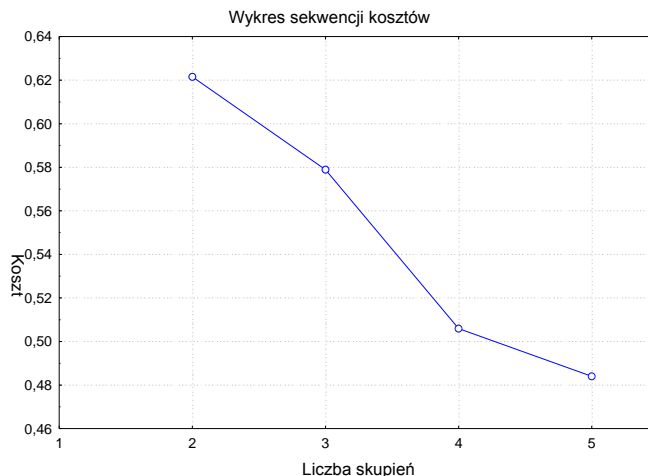
Wyniki analizy skupień

Podczas planowania analizy zdecydowano, że odległości pomiędzy obiektami i środkami skupień będą wyliczane za pomocą miary euklidesowej. Początkowe środki skupień zostały dobrane tak, aby były od siebie możliwie jak najdalsze (czyli najmniej podobne).

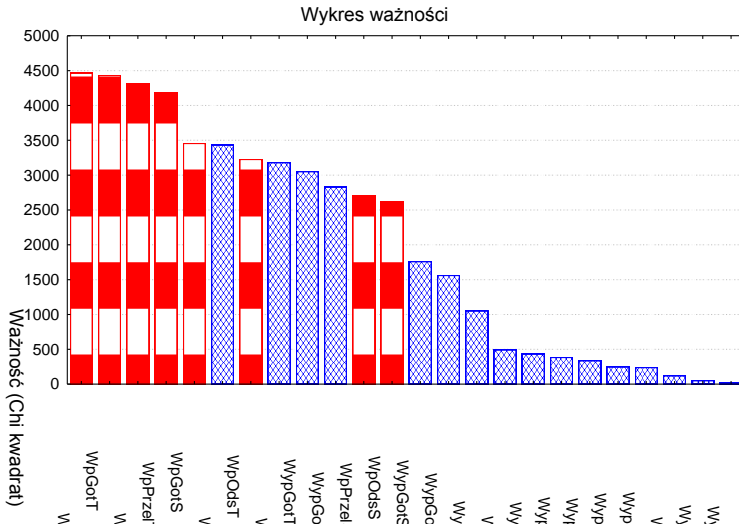


Podsumowanie analizy skupień metodą k-średnich	
Algorytm	k-średnich
Odległość	Euklidesowa
Wstępne środki	Maksymalizuj odległość skupień
BD usuwane przypadkami	Tak
Sprawdzian krzyżowy	3 podzbiory
Próba testowa	0
Próba ucząca	4500
Błąd w próbie uczącej	0,502183
Liczba skupień	4

Populacja klientów została podzielona na 4 segmenty. Dla mniejszej wartości minimalnego spadku procentowego funkcji błędu ta liczba mogłaby być większa. Im większa liczba skupień, tym wartość funkcji kosztu byłaby niższa. Nie należy jednak oceniać optymalnej liczby skupień na podstawie tylko tego parametru. Bardzo ważne jest, aby otrzymany podział na podgrupy umożliwiał skuteczne wykorzystanie go w praktyce. Dla przykładu weźmy skrajny przypadek podziału populacji na taką liczbę segmentów, ile jest w niej elementów. Skutkowałoby to na pewno bardzo niską wartością funkcji kosztu, jednak wyniki byłyby zupełnie bezużyteczne. Liczba segmentów powinna być na tyle duża, żeby odzwierciedlała rzeczywisty podział obiektów, a jednocześnie na tyle mała, by można było wyciągać wnioski dotyczące zróżnicowania obiektów w populacji. Poniższy wykres osypiska pokazuje, że zwiększenie liczby skupień z 3 do 4 skutkuje znacznym spadkiem wartości funkcji kosztu, natomiast pójście o krok dalej i podział na 5 podzbiorów daje już znacznie mniejszą poprawę.

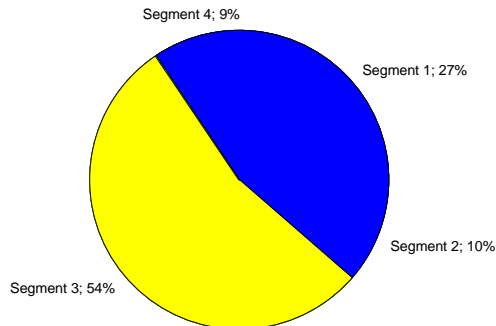


Znając przyporządkowanie obiektów do poszczególnych klas, możemy za pomocą modułu *Dobór zmiennych i analiza przyczyn* (dostępnego w *STATISTICA Data Miner*) sprawdzić, które zmienne miały największy wpływ na podział.



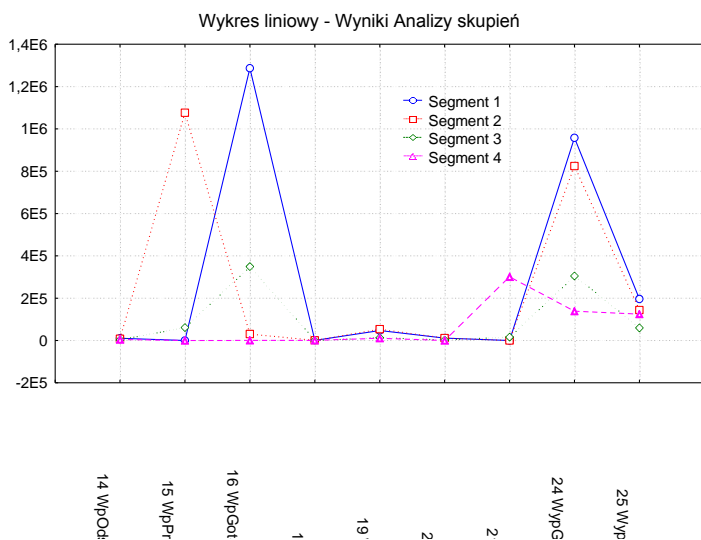
Na powyższym wykresie widać, które zmienne miały największy wpływ na uzyskany podział, a których wpływ był znikomy. Jak już wcześniej wspomniano, segmentację przeprowadzono na 24 zmiennych. Można wśród nich wyróżnić 12 par (liczba transakcji – na końcu nazwy zmiennej oznaczenie T, łączna kwota wszystkich transakcji – na końcu nazwy zmiennej oznaczenie S) dla każdego rodzaju transakcji. Na wykresie możemy zaobserwować, że największy wpływ na podział miały zmienne opisujące wpłaty przelewem (szczególnie emerytury) oraz wpłaty i wypłaty gotówki. Najmniejszy wpływ miały zmienne opisujące kształtowanie się wypłat: odsetek związanych z debetem na koncie, kartą kredytową, na opłaty ubezpieczeniowe, na ratę pożyczki lub kredytu. Podsumowując te wyniki, możemy stwierdzić, że na segmentację większy wpływ mają kwoty i liczba wpłat (na wykresie słupki z paskami poziomymi) niż wypłat (na wykresie słupki w kratkę), a więc głównie ta pierwsza grupa transakcji różnicuje klientów banku.

Wykres kołowy - Wyniki Analizy skupień



Powyższy wykres kołowy pokazuje procentowy udział poszczególnych segmentów w populacji klientów. Największy jest segment nr 3 – to ponad połowa wszystkich klientów. Skupienie 1 to jedna czwarta całej populacji, natomiast pozostałe skupienia: 2 i 4 są niemal tak samo liczne (około 10% i 9%).

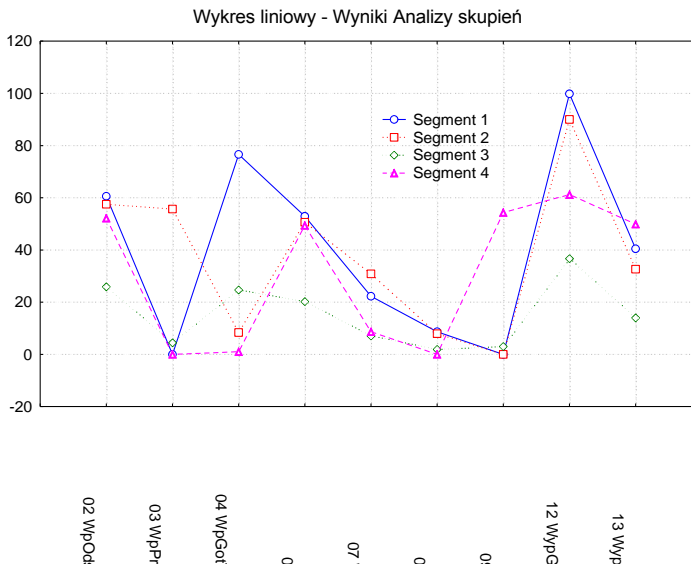
W raporcie z przeprowadzonej analizy wygenerowana została cała gama pomocniczych wyników, np. arkusz zawierający środki segmentów dla poszczególnych zmiennych. Jest to najciekawszy element wyników. Zawiera on informacje o najbardziej typowych przedstawicielach poszczególnych segmentów – określa profile otrzymanych grup klientów. Najłatwiej analizować zawarte w nim informacje po przeniesieniu ich na wykres. Zakres wartości zmiennych opisujących liczbę transakcji jest o wiele mniejszy od zakresu wartości zmiennych opisujących kwoty, na jakie dokonywano transakcji. Aby nie wpłynęło to na czytelność wyników, sporządzono dwa wykresy, oddzielne dla każdej z obu tych grup zmiennych. Na poniższych wykresach nie uwzględniono trzech typów transakcji, które wykazywały znikomy wpływ na przeprowadzoną segmentację.



Powyższy wykres przedstawia środki skupień dla zmiennych opisujących sumy kwot transakcji. Pierwsza grupa to osoby dokonujące wpłat i wypłat w gotówce na największą kwotę. Druga grupa to osoby korzystające przede wszystkim z przelewów, jeśli chodzi o wpłacanie środków, natomiast wypłacające przede wszystkim gotówkę, w ilościach zbliżonych do segmentu pierwszego. Trzecia grupa to osoby wpłacające i wypłacające gotówkę, jednak znacznie niższe kwoty od dwóch pierwszych grup. Czwarty segment jest podobny do trzeciego pod względem kwoty pieniędzy, którymi obraca, różni się od wszystkich źródłem wpłat – są to przelewy z tytułu renty lub emerytury. Analizując dane transakcyjne, można dowiedzieć się wielu ciekawych faktów o klientach banku. Segment 4 to emeryci i renciści (główne źródło wpływów identyfikuje ich jednoznacznie), operują oni małymi kwotami, a suma pieniędzy wpływających na konto jest porównywalna z wypłatami. Podobnie sytuacja ma się w przypadku segmentu 3. Najatrakcyjniejsze wydają się dla

banku pozostałe dwie grupy: 1 i 2. Są to klienci, którzy dużo wpłacają i dużo wypłacają, przy czym suma wpływów na ich konta jest większa niż wypłat. Może to sugerować, że dysponują oni środkami pieniężnymi, które będą chcieli przeznaczyć na dodatkowe usługi banku, jak np. lokaty, fundusze inwestycyjne.

Uzupełnieniem tych informacji jest kolejny wykres pokazujący środki tych samych czterech skupień, jednak tym razem ze względu na zmienne określające liczbę poszczególnych typów transakcji.



Wykres pokazuje, że najmniej aktywny jest trzeci segment. Najwięcej transakcji wykonują klienci z segmentów pierwszego i drugiego, nieco mniej z segmentu czwartego – czyli emeryci. Widzimy także, że wniosek wyciągnięty na podstawie poprzedniego wykresu dotyczący wolnych środków (nadwyżka wpływów nad wypłatami) segmentów 1 i 2 był uzasadniony – regularnie, co miesiąc na ich konto wpływają odsetki z tytułu dodatniego salda na koncie. Okazuje się, że w podobnym stopniu „na plusie” są emeryci, jednak należy spodziewać się, że średnie saldo ich kont jest niższe niż w pierwszych dwóch segmentach. Najmniej aktywna jest grupa najliczniejsza – segment 3. Pomimo tego, że obracają oni kwotami podobnej wielkości do emerytów i rencistów, to zdecydowanie częściej mają na koncie ujemne saldo (sugeruje to dwukrotnie mniejsza liczba wpłat z tytułu odsetek od środków zgromadzonych na koncie).

Najbardziej liczni są klienci najmniej atrakcyjni, czyli segment 3. Środki pieniężne, jakimi obracają, są stosunkowo niewielkie w stosunku do segmentów 1 i 2, często korzystają z debetu. Działania banku można zaplanować w zależności od segmentu. Pierwsza grupa, na którą należy zwrócić uwagę, to segmenty 1 i 2. Mogą oni łatwo dać się namówić na dodatkowe usługi, ponieważ wiemy, że dysponują oni często wolnymi środkami. Można także podjąć działania zachęcające klientów z grupy 3 (najliczniejszej) do utrzymywania dodatniego salda na koncie. Emeryci i renciści stanowią około 9% wszystkich klientów



banku. Utrzymują oni prawie cały czas dodatni stan konta, więc część z nich może chcieć zainwestować wolne środki.

Podsumowanie

Data mining jest bardzo przydatny, ponieważ pozwala wydobyć z danych (których gromadzimy z czasem coraz więcej) różne informacje, które mogą stać się bardzo użyteczną wiedzą o naszych klientach i nie tylko. Jest to o tyle istotne, że dane są często gromadzone z różnych powodów, jednak nikt nawet nie próbuje uzyskać na ich podstawie jakichkolwiek informacji, oprócz tych, które są wymagane. Analizy danych, zarówno data mining jak i statystyczne, pozwalają wykorzystywać dane do czegoś więcej niż tylko obligatoryjne raportowanie. Dzięki nim możemy dowiedzieć się wiele o klientach – ich zachowaniach, przyzwyczajeniach, preferencjach, i dzięki temu budować z nimi relacje i zapewniać ich zadowolenie oraz lojalność względem naszej marki, a co za tym idzie zwiększać ich wartość dla naszej działalności.

Literatura

1. *Analiza satysfakcji i lojalności klienta*, Materiały seminaryjne StatSoft, Kraków 2003.
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. Hill T., Lewicki P., *Statistics: Methods and Applications*, StatSoft, Tulsa 2006.