



## SKORING KREDYTOWY A MODELE DATA MINING

**Janusz Wątroba**

*StatSoft Polska Sp. z o.o.*

Przedmiotem rozważań w niniejszym artykule jest problematyka oceny ryzyka kredytowego oraz wybrane zagadnienia związane z wdrażaniem i monitorowaniem systemów skoringu kredytowego.

### **Wprowadzenie**

Jedną z najbardziej charakterystycznych cech otaczającej nas rzeczywistości jest niepewność. Dotyczy to zarówno zjawisk i procesów przyrodniczych, jak i tych, które są związane z szeroko pojętym otoczeniem gospodarczym i społecznym. Przyczyny występowania niepewności mogą być bardzo różne. Sama natura zjawisk i procesów implikuje bardzo często występowanie niepewności. Niekiedy jej przyczyną jest niepełna lub nieścisła informacja albo wręcz całkowity brak informacji.

Mimo iż potrzeba poskromienia niepewności pojawiła się w działalności człowieka bardzo dawno temu, to jednak skuteczne ilościowe sposoby jej oceny pojawiły się stosunkowo niedawno (Rao 1994). Zagadnienie to ma szczególne znaczenie w sytuacjach decyzyjnych. Podejmowanie decyzji w warunkach niepewności zawsze jest obarczone ryzykiem. Stąd bierze się powszechne zapotrzebowanie na narzędzia wspierające procesy decyzyjne. Wymagania stawiane tym narzędziom to przede wszystkim minimalizacja ryzyka błędnej decyzji (koszty błędnych decyzji są coraz wyższe), szybkość działania (warunki konkurencji) oraz możliwość uwzględniania różnych informacji jakościowych i ilościowych.

Wiele spośród współczesnych dziedzin działalności gospodarczej jest powszechnie kojarzonych z występowaniem niepewności. Wystarczy wspomnieć o finansach, bankowości i ubezpieczeniach.

### **Ryzyko kredytowe i sposoby jego pomiaru**

Działalność banku jest nierozdzielnie związana z występowaniem ryzyka. Można byłoby chyba zaryzykować stwierdzenie, że istotą działalności banku jest identyfikacja różnych rodzajów ryzyka, ocena jego wielkości oraz stosowanie odpowiednich procedur zarządzania ryzykiem. Biorąc pod uwagę specyfikę funkcjonowania banku można tutaj



wspomnieć o ryzyku związanym z płynnością finansową, poziomem stóp procentowych, działalnością kredytową czy też ryzykiem występującym na rynku kapitałowym (np. inwestycje). Konieczność odpowiedniego podejścia do zagadnienia ryzyka nie jest wyłącznie sprawą samego banku. Za zapewnienie bezpiecznego operowania powierzonymi środkami finansowymi oraz utrzymywanie odpowiednich standardów w zakresie ryzyka jest także odpowiedzialny bank centralny. Ważną rolę pełni też tutaj odpowiednie prawodawstwo (Janc i Kraska 2001).

Spośród różnych form swojej działalności banki szczególną wagę przywiązują do bezpieczeństwa działalności kredytowej. Efektem tego są formalne wymogi stawiane bankom przez ustawę *Prawo bankowe*. W jej świetle bank ma obowiązek zbadania zdolności kredytowej osoby lub jednostki gospodarczej starającej się o kredyt. Bankom pozostawiono natomiast dużą swobodę w zakresie wyboru odpowiednich metod oceny zdolności kredytowej. Z punktu widzenia potrzeb banku najważniejsze wymagania stawiane metodom wspomagającym ocenę zdolności kredytowej to:

- ◆ Łatwa dostępność informacji stanowiących podstawę oceny ryzyka,
- ◆ Sprawne przetwarzanie tych informacji, z uwzględnieniem czasu i kosztów,
- ◆ Możliwość łatwej interpretacji proponowanych reguł oraz ustalenia jednoznacznej decyzji dotyczącej przyznania kredytu.

W literaturze są omawiane różne podejścia stosowane przy ocenie zdolności kredytowej (Janc i Kraska 2001, Gruszczyński 2002). Mimo różnic, na jakie zwracają uwagę niektórzy autorzy, podejścia te są w zasadzie bardzo podobne. Chodzi generalnie o to, żeby ocenić, czy dana osoba starająca się o przyznanie kredytu daje gwarancję jego spłaty. W tym celu pracownik podejmujący decyzję o przyznaniu kredytu gromadzi pewne informacje charakteryzujące kredytobiorcę, a następnie w oparciu o opinie ekspertów, określone wskaźniki lub system skoringowy ocenia dany wniosek kredytowy. W dalszej części scharakteryzujemy krótko najczęściej stosowane podejścia (Gruszczyński 2002):

- ◆ Systemy eksperckie,
- ◆ Systemy wskaźnikowe,
- ◆ Systemy skoringowe (ocena punktowa).

W systemach eksperckich decyzja kredytowa jest podejmowana przez odpowiedniego urzędnika. Przy jej podejmowaniu bierze on pod uwagę różne informacje na temat kredytobiorcy. Kieruje się przy tym tzw. regułą 5C (Character, Capacity, Capital, Collateral, Conditions). Pierwszy element tej reguły polega na sprawdzeniu, czy firma lub osoba starająca się o kredyt cieszy się dobrą reputacją. Chodzi tutaj o takie cechy jak: solidność, stabilność czy uczciwość. Najłatwiej jest ją ocenić w przypadku, gdy kredytobiorca jest stałym klientem banku. Kolejny element oznacza zdolność spłacania zadłużenia (bierze się tu pod uwagę wyniki finansowe uzyskiwane w przeszłości, skład zarządu firmy, jego kompetencje itp.). Trzeci element to kapitał, którym dysponuje firma starająca się o kredyt. Brana jest pod uwagę wielkość kapitału, który zainwestowali w firmę jej właściciele oraz efektywność wykorzystania kapitału. Następne brane pod uwagę kryterium to zabezpieczenie spłaty kredytu, które może przyjmować różną postać. Ostatni



element to czynniki określające uwarunkowania zewnętrzne (chodzi tu m.in. o sytuację gospodarczą na rynku oraz kondycję branży, w której działa kredytobiorca). Na podstawie zebranych informacji pracownik banku odpowiedzialny za przyznawanie kredytu podejmuje odpowiednią decyzję. Jest ona w dużym stopniu subiektywna, chociaż przy jej podejmowaniu bierze się również pod uwagę pewne normy ustalone dla konkretnej branży.

Drugim z powszechnie stosowanych podejść jest wykorzystanie analizy wskaźnikowej. Jest ona przeprowadzana na podstawie sprawozdań finansowych firmy, która stara się o przyznanie kredytu. Bank analizuje informacje dotyczące takich wskaźników efektywności działania firmy jak: bilans, rachunek wyników, obroty itp. Lista branż pod uwagę wskaźników może być bardzo obszerna i zależy również od wysokości środków, o które stara się kredytobiorca, i deklarowanego okresu spłaty.

W systemach skoringowych przeprowadza się punktową ocenę ryzyka kredytowego (stąd zresztą bierze się nazwa tego podejścia) na podstawie różnych charakterystyk kredytobiorców. Mogą to być takie same lub podobne informacje, które są brane pod uwagę przy podejściu eksperckim, ale wyrażone w postaci odpowiednich czynników jakościowych lub ilościowych. Następnie w wyniku zastosowania odpowiednich metod analizy danych budowany jest model pozwalający obliczyć sumaryczną ocenę punktową dla każdego z kredytobiorców. Ocena ta jest z kolei wykorzystywana do klasyfikowania kredytobiorców na grupy o określonym poziomie ryzyka kredytowego.

W dalszej części artykułu zajmiemy się bliżej modelami skoringowymi, które są budowane dla indywidualnych wniosków kredytowych.

## Skoring kredytowy w przypadku indywidualnej umowy kredytowej

Z technicznego punktu widzenia ilościowa ocena ryzyka kredytowego polega na obliczeniu prognozy dla jakościowej zmiennej o rozkładzie dwumianowym (jest to zmienna objaśniana). Wartości zmiennej oznaczają przynależność osoby starającej się o kredyt do jednej z dwóch kategorii:

- ◆ Osoby, którym nie należy przyznawać kredytu (ze względu na duże ryzyko dla banku),
- ◆ Osoby, którym można kredyt przyznać (niewysokie ryzyko).

Lista potencjalnych zmiennych objaśniających (predyktorów) obejmuje różnego rodzaju informacje jakościowe i ilościowe, zarówno na temat samego klienta, jak i jego otoczenia. W praktyce punktowa ocena ryzyka oznacza wartość liczbową z pewnego przedziału (wygodnie jest przyjąć zakres od 0 do 100 punktów). Wartość prognozy obliczona dla konkretnego klienta oznacza szacunkowy poziom ryzyka danego wniosku kredytowego.

Ocenę punktową należy odróżniać od modelu skoringowego. Najpierw trzeba zbudować określony model skoringowy, tzn. ustalić wysokość punktacji dla przyjętych zmiennych objaśniających. Ustala się ją na podstawie doświadczeń banku z osobami, które poprzednio starały się o przyznanie kredytu. Inaczej mówiąc, są to umowne wartości przypisywane oryginalnym wartościom predyktorów na podstawie tzw. karty skoringowej. Przykładowo



dla predyktora *Wiek* można ustalić następującą punktację: (1) od 18 do 30 lat – 0 pkt, (2) od 31 do 44 lat – 3 pkt i (3) od 45 lat – 6 pkt. Utworzenie karty skoringowej wymaga użycia odpowiedniego modelu skoringowego (jednym z najważniejszych zadań jest wybranie pytań, które stanowią dobre predyktory ryzyka). Ocena punktowa jest natomiast sumą punktów, którą otrzymał dany kredytobiorca na podstawie odpowiedzi na zadane przez urzędnika kredytowego pytania.

## Zalety i wady punktowej oceny ryzyka kredytowego

Omawiana metoda oceny zdolności kredytowej ma zarówno swoje dobre strony, jak i pewne mankamenty (Janc i Kraska 2001). Główna jej zaleta to fakt, że jej przeprowadzenie jest bardzo łatwe, co pozwala zaoszczędzić czas niezbędny do analizy wniosku kredytowego. Jednocześnie jest to metoda umożliwiająca obiektywną ocenę zdolności kredytowej. Dobrze przygotowany model skoringowy pozwala zmniejszyć liczbę „złych” kredytów. Dla banku jej stosowanie umożliwi na ogół zwiększenie wydajności pracy urzędników kredytowych oraz obniżenie kosztów obsługi. Z kolei dla osoby starającej się o kredyt stosowanie przez bank oceny skoringowej oznacza zmniejszenie liczby wymaganych dokumentów niezbędnych do przeprowadzenia oceny zdolności kredytowej.

Chociaż generalna ocena skoringu kredytowego jest na ogół bardzo wysoka, to jednak nie sposób pominąć również pewnych zarzutów, które są jej stawiane. Czasami stosowane karty skoringowe dyskryminują pewne grupy społeczne. Podobnym zarzutem jest to, że przy budowie karty skoringowej wykorzystuje się jedynie informacje o kredytobiorcach, którym kredyt został przydzielony, a pomija się grupę klientów, których wnioski zostały odrzucone. Przynajmniej z częścią wad przytaczanych przez różnych autorów można chyba polemizować. Przykładowo można zapobiec możliwości szybkiej dezaktualizacji stosowanego przez dany bank systemu skoringowego poprzez bieżące wprowadzanie do systemu informacji o nowych kredytobiorcach i uaktualnianie wykorzystywanej karty skoringowej. Kolejny zarzut to klasyfikowanie kredytobiorców tylko do dwóch grup ryzyka (kredyt „dobry” i „zły”). W tym przypadku istnieje możliwość wprowadzenia trzeciej kategorii (np. „nieokreślony”). Wymaga to tylko zastosowania odpowiednich metod modelowania w przypadku zmiennych jakościowych (w takim przypadku zmienna objaśniana przyjmuje postać zmiennej jakościowej wielomianowej).

## Metody analizy danych stosowane w skoringu kredytowym

Skonstruowanie dobrego modelu, tzn. takiego, który pozwala trafnie zakwalifikować dany wniosek kredytowy do jednej z dwóch klas („zły” i „dobry”) nie jest zadaniem łatwym. Warunkiem koniecznym jest tutaj posiadanie takich informacji o osobie czy firmie starającej się o kredyt, które rzeczywiście pozwalają ocenić zdolność kredytową. Nawet najbardziej wyszukane metody modelowania nie zapewnią zbudowania dobrego modelu na postawie nieadekwatnych bądź nierzetelnych danych. Warto o tym pamiętać szczególnie przy tworzeniu koncepcji zbierania odpowiednich informacji o kredytobiorcach.



W tym miejscu wypada wreszcie wyjaśnić, dlaczego w tytule artykułu znalazło się określenie „modele data mining”. Najważniejszy powód jest taki, że przy konstruowaniu modeli skoringowych dobiera się odpowiednie informacje o kredytobiorcach w taki sposób, aby na ich podstawie móc trafnie oceniać, czy dany kredyt dobrze czy źle rokuje (tak więc bierze się pod uwagę własności predykcyjne tych informacji). Nie opieramy się zatem na jakiejś istniejącej teorii (z zakresu bankowości lub psychologii konsumenta) czy też postulowanych mechanizmach kształtujących ryzyko kredytowe, lecz poszukujemy pewnych wzorców lub prawidłowości w zebranych wcześniej danych. Właśnie dane stanowią punkt wyjścia przy budowie modelu, a nie a priori sformułowane hipotezy. Taki sposób prowadzenia analizy danych jest charakterystyczny dla dziedziny nazywanej *data mining*. Podejście to znajduje zastosowanie przede wszystkim w tych obszarach badań empirycznych, gdzie z różnych względów nie jest możliwe przeprowadzanie ściśle zaplanowanych i dobrze kontrolowanych eksperymentów (np. ze względu na koszty lub powody etyczne) lub brak wystarczająco uzasadnionych teorii, lub też złożoność zjawisk jest zbyt duża (np. w medycynie, naukach społecznych, ekonomii, finansach czy ubezpieczeniach).

Stosowanie tych metod nie wymaga co prawda spełnienia dość krępujących założeń, jakie są stawiane przy klasycznym (statystycznym) podejściu do analizy danych, ale niejako w zamian za to z reguły wymagana jest bardzo duża liczba obserwacji, a ponadto wyniki niełatwo dają się uogólniać na populację. Mimo tego można zaobserwować zarówno stały rozwój tego typu metod, jak i rosnący zakres praktycznych zastosowań.

Przy budowie modeli wykorzystywanych w skoringu kredytowym stosuje się metody uwzględniające fakt, że zmienna zależna (objaśniana) jest zmienną jakościową (najczęściej jest to zmienna dychotomiczna). Wśród powszechnie stosowanych technik analitycznych spotyka się zarówno bardziej klasyczne metody, np. regresję logistyczną czy analizę dyskryminacyjną, jak i techniki zaliczane często do grupy metod określanych terminem *machine learning* (wśród których wymienia się np. uogólnione modele addytywne, techniki drzew decyzyjnych, metodę Support Vector Machines, metody wykorzystujące sztuczne sieci neuronowe). Obecnie istnieje już dość bogata literatura na ten temat, zwłaszcza w języku angielskim (m. in. Hastie i wsp. [2001], Krawiec i Stefanowski [2003], Giudici [2003], Lasek [2002], Gruszczyński [2002]).

Praktyczne wykorzystanie tych metod jest coraz bardziej dostępne ze względu na szybki rozwój odpowiedniego oprogramowania (bardzo dobrym przykładem mogą być systemy i programy analizy danych z rodziny *STATISTICA* firmy StatSoft).

## Podsumowanie

Trzeba sobie zdawać sprawę, że zbudowanie dobrego modelu skoringowego nie jest zadaniem łatwym i wymaga zwykle bardzo przemyślanego zaprojektowania całego przedsięwzięcia. Proces budowy analitycznych modeli jest tylko jednym z etapów tego procesu, w dużym stopniu uzależnionym od jakości i rzetelności zebranych danych. Trafność oceny zdolności kredytowej zależy zatem od tego, co zostało w taki model wbudowane. Model skoringowy jest budowany w oparciu o dane dotyczące ubiegłych okresów,



a zatem musi on być umiejętnie „konserwowany” (tzn. co pewien okres czasu należy sprawdzać, czy jego własności prognostyczne są zachowane).

Na zakończenie wypada przypomnieć, że nawet najlepszy system skoringowy powinien być traktowany jako jeden z elementów wspierających podejmowanie rzeczywistych decyzji kredytowych, a nie jako narzędzie zastępujące decyzje urzędnika kredytowego odpowiedniego szczebla.

## Literatura

1. Giudici P., 2003, Applied Data Mining. Statistical Methods for Business and Industry, Wiley.
2. Gruszczyński M., 2002, Modele i prognozy zmiennych jakościowych w finansach i bankowości, Oficyna Wydawnicza Szkoły Głównej Handlowej, Warszawa.
3. Hastie T., Tibshirani R., Friedman J., 2001, The Elements of Statistical Learning, Springer.
4. Janc A., Kraska M., 2001, Credit-Scoring. Nowoczesna metoda oceny zdolności kredytowej, Biblioteka Menedżera i Bankowca, Zarządzanie i Finanse, Warszawa.
5. Krawiec K., Stefanowski J., 2003, Uczenie maszynowe i sieci neuronowe, Wydawnictwo Politechniki Poznańskiej.
6. Lasek M., 2002, Data Mining. Zastosowania w analizach i ocenach klientów bankowych, Biblioteka Menedżera i Bankowca, Zarządzanie i Finanse, Warszawa.
7. Rao C. R., 1994, Statystyka i prawda, PWN.