



WYKORZYSTANIE MODELI SKORINGOWYCH W OBSŁUDZE PORTFELI WIERZYTELNOŚCI DŁUŻNIKÓW BIZNESOWYCH

Piotr Jabkowski
Instytut Socjologii UAM w Poznaniu

Piotr Karol Potempa
Prezes Zarządu GPPI S.A.

Choć dominującym obszarem zastosowań modeli skoringowych wydają się wciąż analizy rynku kredytowego oraz marketingowego, to jednym z coraz bardziej popularnych obszarów implementacji procedur skoringowych stają się również analizy ukierunkowane na obsługę portfeli wierzytelności. Z formalnego punktu widzenia modele skoringowe mieszczą się w obrębie statystycznych procedur klasyfikacyjnych, które pozwalają na oszacowanie prawdopodobieństwa przynależności określonych jednostek statystycznych do jednej z dwóch rozłącznych kategorii. Ujmując problem nieco inaczej, można powiedzieć, że modele skoringowe wykorzystuje się wtedy, gdy zmienna „zależna” przyjmuje charakter nominalny dwumianowy.

Konstrukcję modeli skoringowych podzielić można na dwa zasadnicze etapy. Pierwszym z nich jest wybór predyktorów oraz przygotowanie ich do analizy. Drugim natomiast budowa modelu skoringowego przy wykorzystaniu metod analiz regresyjnych, oszacowanie heurystycznej i statystycznej mocy modelu oraz jego implementacja praktyczna na nowe zbiory danych. Warto dodać, że powszechną miarą szacowania mocy predykcyjnej każdego z predyktorów jest tzw. wskaźnik *information value (IV)*, z kolei wyznacznikiem mocy predykcyjnej poszczególnych wartości lub przedziałów wartości zmiennych prognozujących jest miara *Weight of Evidence (WoE)*.

Poza statystycznymi podstawami budowy modeli skoringowych w referacie zaprezentowany został przykład implementacji modelu w analizie obsługi portfeli wierzytelności dłużników biznesowych jednej w firm usługowych. Zasadniczym celem tej implementacji praktycznej było oszacowanie przewidywanej kwoty, jaką w wyniku działań windykacyjnych uda się „odzyskać” z pakietu dłużników. W trakcie budowy modelu ujawniły się jednak wszystkie trudności, w jakie uwikłana jest próba wnioskowania statystycznego na podstawie „ubogiego” zestawu predyktorów. Wydaje się, że głównym ograniczeniem wykorzystywania modeli skoringowych w obsłudze portfeli wierzytelności jest właśnie bardzo ograniczony zasób wiedzy, jaką *a priori* dysponuje się o poszczególnych dłużnikach. Pakiety wierzytelności zawierają bowiem jedynie podstawowe informacje o dłużnikach, np. o typie działalności gospodarczej, rodzaju firmy, czy też wartości długu oraz



części, jaką stanowi dług przedawniony, nie zawierają natomiast informacji o takich cechach dłużnika, które z całą pewnością w większym stopniu decydują o prawdopodobieństwie spłaty (np. historia zadłużenia, struktura zatrudnienia w firmie, roczne obroty i zyski firmy, jej dotychczasowa kondycja finansowa i inne charakterystyki dłużników).

Schemat budowy modelu skoringowego oraz prezentacja głównego celu badawczego

Zbiór danych historycznych, na podstawie których budowano model skoringowy do wyceny portfeli wierzytelności zawierał informacje o ponad 2700 dłużnikach. Zmienną objaśnianą była wartość spłaty zadłużenia, z kolei zbiór predyktorów obejmował następujące informacje o dłużnikach: całkowita wartość długu oraz wartość długu przedawnionego, typ spółki i region działalności gospodarczej. Kolejne etapy budowy modelu skoringowego obejmowały: przygotowanie danych, konstrukcję tablicy skoringowej, oszacowanie modelu oraz jego implementację.

Etap 1: Przygotowanie danych:

- ◆ Przygotowanie danych historycznych potrzebnych do budowy modelu – zdefiniowanie zmiennych zależnych (prognozowanych) oraz niezależnych (predyktorów).
- ◆ Wybór predyktorów istotnych dla budowy modelu skoringowego, w oparciu o ranking predyktorów pokazujący, która zmienna wnosi największy wkład w predykcję spłaty długu.
- ◆ Dyskretyzacja zmiennych – przygotowanie danych do analizy poprzez dyskretyzację wartości zmiennych. W wyniku tej operacji uzyskuje się obraz wpływu poszczególnych wartości predyktorów na prawdopodobieństwo spłaty długu.

Etap 2: Budowa tablicy skoringowej:

- ◆ Najbardziej popularną metodą budowy modelu skoringowego jest regresja logistyczna. W efekcie otrzymuje się przejrzystą tablicę skoringową, w której poszczególnym poziomom predyktorów (zdefiniowanym na etapie dyskretyzacji) przypisywana jest odpowiednia liczba punktów oraz prawdopodobieństwo zaklasyfikowania dłużnika do kategorii pozytywnej, oznaczającej tu pełną lub częściową spłatę długu.

Etap 3. Ocena modelu:

- ◆ W tym etapie określa się zasady oceny mocy predykcyjnej modelu skoringowego.

Etap 4. Wyznaczenie prognozy wartości „odzyskanego” zadłużenia:

- ◆ W oparciu o zbudowaną tablicę skoringową oraz statystyczne modele szacowania prawdopodobieństw wyznaczono dla każdego dłużnika prawdopodobieństwo spłaty długu (oznaczane tutaj symbolem $\hat{\pi}$), a dokładniej prawdopodobieństwo zaklasyfikowania dłużnika do kategorii „częściowa lub pełna spłata długu” – zmienne zdefiniowane zostały w kolejnej części artykułu. Wartość prawdopodobieństwa mieści się oczywiście w granicach pomiędzy 0% a 100% i ukazuje ono przeciętną szansę, jaką



(w oparciu o liczbowe i nieliczbowe charakterystyki dłużnika) ma się na uzyskanie pełnej lub częściowej spłaty długu.

Ponieważ charakter zmiennej zależnej jest taki, że model prognozuje prawdopodobieństwo zaklasyfikowania jednostek do grupy dłużników o pełnej lub częściowej spłacie długu, to fakt ten należy uwzględnić przy ostatecznym szacowaniu prognozowanej wartości ściągniętego zadłużenia. Z tego względu matematyczny wzór na prognozowaną wartość odzyskanej kwoty długu zawiera informację o przeciętnym odsetku spłaty długu (symbol c_1) w grupie dłużników z kategorii „częściowa lub pełna spłata długu”. Gdyby taka poprawka nie została wprowadzona, model skoringowy przeszacowałby prognozę wartości odzyskanego zadłużenia, przypisując dłużnikom spłatę pełną. Owo oszacowanie przedstawia się zatem następująco:

$$V(\underline{x}) = \sum_{i=1}^N c_1 \hat{\pi}_i x_i,$$

gdzie:

- ◆ \underline{x} jest zbiorem wartości zadłużenia „po wycofaniach” dla N dłużników, poszczególny jednostkowy dług oznaczany jest jako x_i dla $i \in \{1, 2, \dots, N\}$,
- ◆ $\hat{\pi}_i$ dla $i \in \{1, 2, \dots, N\}$ jest wyznaczonym na podstawie tablicy skoringowej prawdopodobieństwem zaklasyfikowania i -tego dłużnika do kategorii „częściowa lub pełna spłata długu”,
- ◆ c_1 jest przeciętnym odsetkiem spłaty długu w grupie dłużników zaliczonych do kategorii „częściowa lub pełna spłata długu”.

Definicja zmiennej niezależnej oraz opis predyktorów

Wprawdzie do budowy tablicy skoringowej konieczne jest zdefiniowanie zmiennej zależnej o charakterze dychotomicznym (czyli takiej, w której występują dwie kategorie wartości zmiennej), to jednak z uwagi na główny cel modelu właściwą zmienną prognozowaną była wartość odzyskanego zadłużenia w wyniku działań windykacyjnych. Można tutaj wobec tego *de facto* mówić o dwóch zmiennych prognozowanych. Pierwszą jest dychotomiczna zmienna jakościowa definiowana na potrzeby implementacji modeli skoringowych, drugą natomiast właściwa analizie zmienna ilościowa. Dychotomizacja wartości zmiennej zależnej przebiegała w następujący sposób:

- ◆ Pierwszą kategorię (wartość 1) – przypisano tym wszystkim dłużnikom, którzy dokonali pełnej lub częściowej spłaty długu (spłata przynajmniej 25% wartości zadłużenia). Dłużnicy tacy to pożądane obiekty w bazie – stanowią oni pozytywną (dobrą) kategorię dłużników;
- ◆ Drugą kategorię (wartość 0) – przypisano tym dłużnikom, którzy nie dokonali żadnej spłaty swojej wierzytelności lub co najwyżej dokonali spłaty w wysokości nieprzekraczającej 25% kwoty zadłużenia. Ta kategorię dłużników stanowi niepożądane obiekty w bazie danych – negatywna (zła) kategoria dłużników.

Model skoringowy ma w swoim założeniu prognozować prawdopodobieństwo zaklasyfikowania dłużników do pierwszej kategorii, czyli prawdopodobieństwo pełnej lub przynajmniej częściowej spłaty długu.

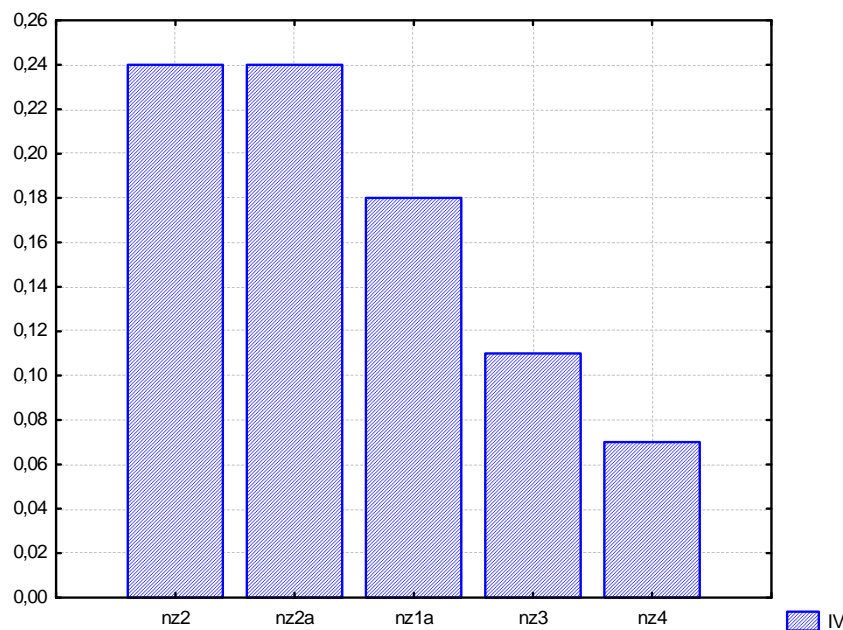
Zbiór predyktorów obejmował z kolei następujące dane o dłużnikach:

- ◆ Wartość zadłużenia po wycofaniach (dane liczbowe) – zmienna_1a;
- ◆ Wartość długu przedawnionego (dane liczbowe) – zmienna_2;
- ◆ Odsetek długu przedawnionego (dane liczbowe) – zmienna_2a;
- ◆ Typ dłużnika (dane jakościowe: typ1; typ2; typ3) – zmienna_3;
- ◆ Region działalności (dane jakościowe: R1; R2; R3; R4; R5; R6; R7; R8; R9; R10) – Zmienna_4.

Przygotowanie zbioru danych – dyskretyzacja zmiennych

Dane wykorzystywane w opisywanym przykładzie obejmowały zbiór 2779 dłużników biznesowych jednej z firm usługowych, którym przypisane zostały wartości odpowiednich zmiennych. Pierwszym etapem analizy było czyszczenie zbioru danych oraz przygotowanie zmiennych do procedury skoringowej. Wykorzystano w tym celu specjalnie dedykowany moduł programu *STATISTICA 8*.

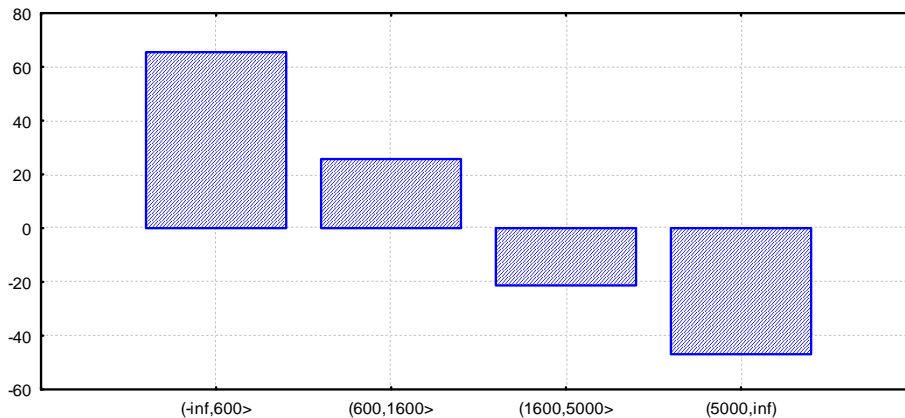
Wejściowy zbiór pięciu predyktorów został opisany pod względem wagi, jaką każda zmienna wnosi w predykcję prawdopodobieństwa spłaty długu. Miarą umożliwiającą wykonanie rankingu predyktorów jest miara *information value* w skrócie *IV*, na podstawie której ogranicza się wejściowy zbiór danych jedynie do zmiennych istotnie wpływających na badane zjawisko. Na rys. 1 przedstawiono ranking predyktorów w oparciu o miarę *IV*.



Rys. 1. Ranking predyktorów dla zdefiniowanego zbioru zmiennych.

Z powyższego wykresu można łatwo odczytać, że największą moc wyjaśniającą mają dwa predyktory opisujące dług przedawniony. Zarówno bowiem nominalna wartość długu (*zmienna_nz2*), jak i zmienna pokazująca odsetek długu przedawnionego (*zmienna_nz2a*) uzyskały identyczną wartość *IV* równą 0,24, świadczącą o średniej mocy predykcyjnej. Trzecią zmienną w rankingu ważności predyktorów jest wartość zadłużenia po wycofaniach (*zmienna_nz1a*), a otrzymana wartość *IV* również oznacza średnią moc predykcyjną. Stosunkowo najmniejszą mocą predykcyjną, choć co należy podkreślić istotną dla budowy modelu, cechują się dwie zmienne jakościowe – typ dłużnika (*zmienna_nz3*) oraz region działalności zadłużenia (*zmienna_nz4*). Jednocześnie w zbiorze predyktorów nie występuje taka zmienna, która dominowałaby swoją mocą predykcyjną.

Ponieważ tablica skoringowa jest narzędziem stosowanym dla danych skategoryzowanych, to każda ze zmiennych liczbowych musi być podzielona na przedziały (dyskretyzacja wartości), natomiast zmienne jakościowe muszą zostać przekodowane przez połączenie ze sobą podobnych klas przypadków – upraszczanie modelu. W tym celu oblicza się dwie miary, które pomagają zrozumieć zależności pomiędzy klasami wartości predyktorów, a prognozowaną zmienną. Miary te to tzw. *WoE* (*weight of evidence*), która ukazuje, jak dany poziom pomiaru zmiennej wpływa na prawdopodobieństwo wystąpienia niepożądanego klasy dłużnika, oraz omawiana już miara *IV*, która pozwala kontrolować, jak w wyniku kategoryzacji predyktorów zmienia się moc predykcyjna całej zmiennej. Na poniższych wykresach przedstawiono efekt dyskretyzacji wartości wszystkich zmiennych wejściowych.

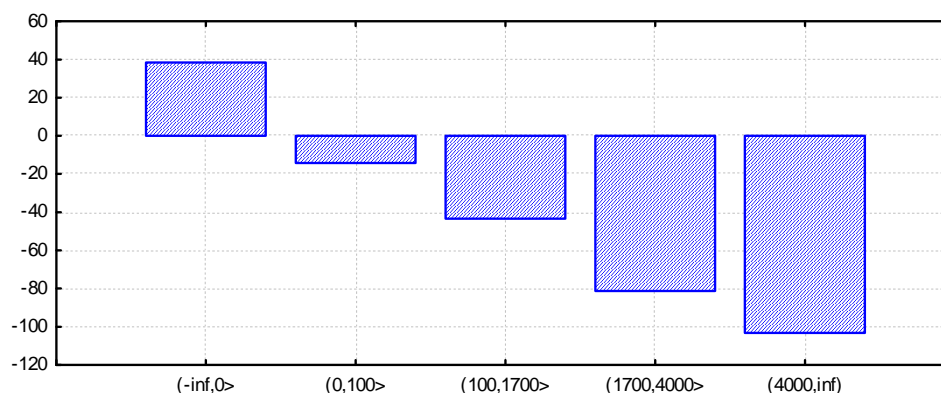


Rys. 2. Rozkład *weight of evidence* dla zmiennej wartości zadłużenia *nz1a*.

Tabela 1. Kategoryzacja zmiennej wartości zadłużenia *nz1a*.

nz1a_kat	Dobry	Zły	Portfel	IV	WoE weight of
(-inf,600>	234	287	55,09%	0,09	65,46
(600,1600>	205	374	64,59%	0,01	25,76
(1600,5000>	203	593	74,50%	0,01	-21,32
(5000,inf)	185	698	79,05%	0,06	-46,91
Ogół	827	1952	70,24%	0,18	

Dane zaprezentowane na powyższym rysunku oraz wykresie ukazują, w jaki sposób wartość zadłużenia wpływa na częstość występowania dłużników zaklasyfikowanych do kategorii negatywnej. Kolumna *portfel* zawiera informację o tym, jaką część w danym przedziale zadłużenia stanowiła negatywna kategoria dłużników. Analizując dane przedstawione na powyższym wykresie oraz w tabeli, można odczytać, że wraz ze wzrostem zadłużenia spada jakość portfela wierzycieli. Największą liczbę „dobrych” dłużników odnotowano w przedziale zadłużenia do 600PLN (w tej grupie proporcje wynoszą 45% dłużników „dobrych” oraz 55% „złych”), a także w przedziale od 600 do 1600 PLN (proporcje 35% - 65% z przewagą „złych” dłużników). Można również zauważyć, że w przedziale powyżej 1600PLN wartość *WoE* spada poniżej wartości 0, co oznacza, iż długi o wartości większej niż 1600PLN znacznie osłabiają jakość portfela wierzycieli.



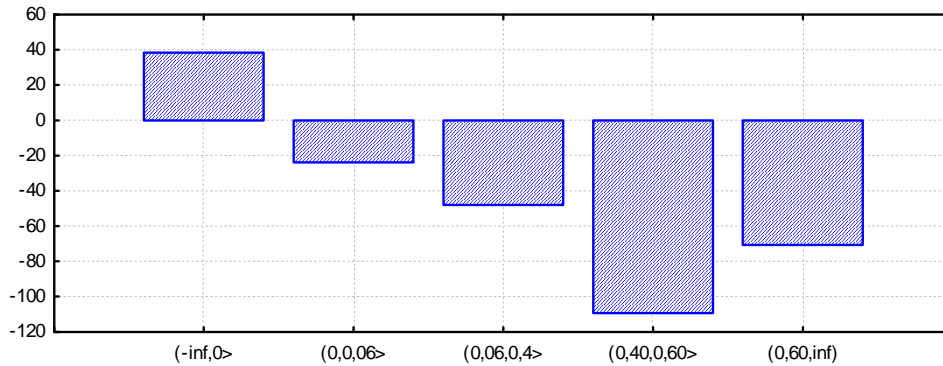
Rys. 3. Rozkład *weight of evidence* dla zmiennej wartość długu przedawnionego *nz2*.

Tabela 2. Kategoryzacja zmiennej wartość długu przedawnionego *nz2*.

Kategoryzowana zmienna: <i>nz2</i>					
<i>nz2_kat</i>	Dobry	Zły	Portfel	IV	WoE
(-inf,0>	596	958	61,65%	0,09	38,42
(0,100>	32	87	73,11%	0,00	-14,14
(100,1700>	118	430	78,47%	0,03	-43,43
(1700,4000>	46	245	84,19%	0,06	-81,38
(4000,inf)	35	232	86,89%	0,08	-103,26
Ogół	827	1952	70,24%	0,26	

Przyglądając się wynikom dla drugiego z predyktorów – wartość długu przedawnionego – wyraźnie widać, że jedynie pierwsza kategoria wartości zmiennej ma pozytywny wpływ na jakość portfeli wierzycieli. W tej pierwszej grupie, obejmującej dłużników nieposiadających żadnego długu przedawnionego, wartość *WoE* jest najwyższa, czyli zawiera najmniej niepożądanych dłużników (proporcja „dobrych” do „złych” dłużników wynosi 39 do 61 punktów procentowych). Zaobserwowana zależność przyjmuje charakter wprost proporcjonalny, czyli w każdej kolejnej grupie dłużników znacząco pogarsza się jakość portfela wierzycieli – występuje coraz mniej dłużników, którzy spłacają swoje zadłużenie. Dla przykładu najmniejsza wartość *WoE* występuje w grupie dłużników, których

wartość długu przedawnionego wynosi ponad 4000PLN, w tej bowiem grupie występuje jedynie 13% dłużników „dobrych” oraz aż 87% dłużników „złych”.

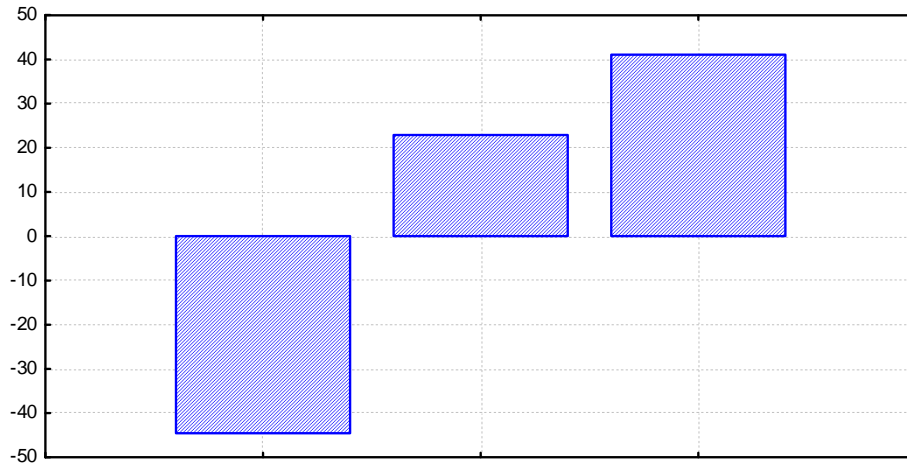


Rys. 4. Rozkład *weight of evidence* dla zmiennej odsetek długu przedawnionego *nz2a*.

Tabela 3. Kategoryzacja zmiennej odsetek długu przedawnionego *nz2a*.

nz2a_kat	Dobry	Zły	Portfel	IV	WoE
(-inf,0>	596	958	61,65%	0,09	38,42
(0,0,06>	30	90	75,00%	0,00	-23,98
(0,06,0,4>	120	458	79,24%	0,04	-48,06
(0,40,0,60>	26	183	87,56%	0,07	-109,26
(0,60,inf)	55	263	82,70%	0,05	-70,60
Ogół	827	1952	70,24%	0,25	

Podobnie jak w przypadku zmiennej opisującej wartość długu przedawnionego, tak i dla odsetka posiadanego długu przedawnionego najlepszą jakość portfela zanotowano dla dłużników nieposiadających żadnego przedawnionego zadłużenia. W tej bowiem grupie odsetek „dobrych” dłużników był relatywnie najwyższy w stosunku do pozostałych kategorii i wyniósł prawie 39 punktów procentowych. Analizując wartość wskaźnika *WoE*, widać, że posiadanie jakiegokolwiek części długu przedawnionego pogarsza już znacząco jakość portfela wiarygodności. Co ciekawe, odsetek „złych” dłużników wzrastał aż do poziomu 60% wartości długu przedawnionego (w kategorii od 40 do 60 procent przedawnionego zadłużenia odsetek „dobrych” dłużników wynosi niespełna 12%), by spaść w ostatniej kategorii – powyżej 60% zadłużenia przedawnionego (tu procentowy udział „dobrych” dłużników wyniósł prawie 18%). Jak się jednak okazuje, w tej ostatniej grupie przeważali dłużnicy, których spółki notowane były na indeksach giełdowych. Ten z kolei typ dłużnika biznesowego charakteryzował się (pomimo posiadania znacznej części długu przedawnionego) większą skłonnością do regulowania swoich należności.

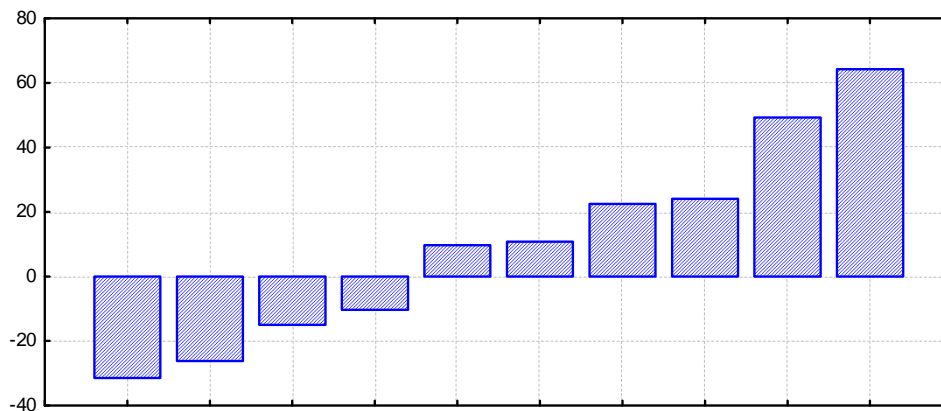


Rys. 5. Rozkład *weight of evidence* dla zmiennej typ dłużnika biznesowego *nz3*.

Tabela 4. Kategoryzacja zmiennej typ dłużnika biznesowego *nz3*.

nz3_kat	Dobry	Zły	Portfel	IV	WoE
typ 1	229	844	78,66%	0,07	-44,56
typ 2	552	1036	65,24%	0,03	22,92
typ 3	46	72	61,02%	0,01	41,08
Ogół	827	1952	70,24%	0,11	

Kolejnym predyktorem była jakościowa zmienna charakteryzująca dłużników ze względu na typ spółki. Z zaprezentowanych na wykresie oraz w tabeli danych widać, że największe prawdopodobieństwo wystąpienia dłużników, którzy dokonają pełnej lub przynajmniej częściowej spłaty swojego zadłużenia, występuje w kategorii typ 3. (w grupie tej występowało 39% dłużników „dobrych” oraz 61% dłużników „złych”). Dodatnia wartość współczynnika *WoE* występuje również w typie 2. Najgorszą grupę dłużników stanowią z kolei dłużnicy z kategorii typ 1 – prawdopodobieństwo wystąpienia dłużnika, który dokona spłaty swojego zadłużenia, wynosi niespełna 21 procent.



Rys. 6. Rozkład *weight of evidence* dla zmiennej region działalności dłużnika *nz4*.

Tabela 5. Kategoryzacja zmiennej region działalności dłużnika *nz4*.

nz4_kat	Dobry	Zły	Portfel	IV	WoE
R1	154	498	76,38%	0,02	-31,48
R2	74	227	75,42%	0,01	-26,21
R3	70	192	73,28%	0,00	-15,02
R4	68	178	72,36%	0,00	-10,35
R5	56	120	68,18%	0,00	9,67
R6	76	161	67,93%	0,00	10,81
R7	172	324	65,32%	0,01	22,56
R8	76	141	64,98%	0,00	24,08
R9	52	75	59,06%	0,01	49,26
R10	29	36	55,38%	0,01	64,26
Ogół	827	1952	70,24%	0,07	

Ostatnią ze zmiennych był region działalności dłużnika. Analizując jakość portfeli wierzytelności pod kątem wartości tej zmiennej, można zauważyć, że największe prawdopodobieństwo wystąpienia „dobrych” dłużników występuje w regionie 10 (45% dłużników, którzy dokonali spłaty pełnej lub częściowej) oraz w regionie 9 (odsetek dłużników w grupie pozytywnej wyniósł prawie 41 punktów procentowych). Dodatnia wartość współczynnika *WoE* występuje też w regionie 8 oraz 7 (prawie identyczne proporcje pozytywnej oraz negatywnej kategorii dłużników w obu regionach), a także w regionie 6 i 5 (w obu regionach również bardzo zbliżone odsetki „dobrych” dłużników). Dla wszystkich czterech pozostałych regionów działalności gospodarczej znacznie pogarsza się jakość portfeli wierzytelności – odsetki dłużników, którzy dokonali spłaty zadłużenia nie przekraczały 27 punktów procentowych.

Budowa tablicy skoringowej

Przedstawiona kategoryzacja predyktorów stanowi podstawę budowy tablicy skoringowej. W pierwszym etapie budowy tablicy skoringowej określa się w oparciu o analizy regresji, czy konieczne jest uwzględnianie wszystkich wybranych wcześniej zmiennych. W poniższej tabeli przedstawiono wynik procedury regresji z eliminacją krokową wsteczną dla pięciu wejściowych predyktorów:

Tabela 6. Wynik analizy regresji.

z1b - Wyniki budowy modelu (Arkusz3) Rozkład: DWUMIANOWY , F. wiążąca: LOGIT Modelowane prawdopodobieństwo z1b = Spłata					
Krok	Efekt	Stopnie Swobody	Walda Stat.	Walda p	War. Stan
Krok 1	nz1a_kat	1	34,20515	0,000000	W modelu
	nz2_kat	1	0,22009	0,638968	Usunięty
	nz2a_kat	1	6,16616	0,013022	W modelu
	nz3_kat	2	51,89509	0,000000	W modelu
	nz4_kat	9	27,01453	0,001391	W modelu
Krok 2	nz1a_kat	1	40,26801	0,000000	W modelu
	nz4_kat	9	26,98031	0,001409	W modelu
	nz2a_kat	1	67,59542	0,000000	W modelu
	nz3_kat	2	51,91450	0,000000	W modelu
	nz2_kat	1			Poza

Zaprezentowane wyniki pokazują, że wystarczy wybrać cztery z pięciu analizowanych wcześniej predyktorów. Z modelu skoringowego usunięty został predyktor *wartość długu przedawnionego* (*zmienna_nz2*), dający nadmiarową informację – pochodną względem odsetka długu przedawnionego. Do budowy tablicy skoringowej wykorzystane będą zatem następujące zmienne:

- ◆ Wartość zadłużenia po wycofaniach (dane liczbowe) – *zmienna_1a*;
- ◆ Odsetek długu przedawnionego (dane liczbowe) – *zmienna_2a*;
- ◆ Typ dłużnika (dane jakościowe: typ1; typ2; typ3) – *zmienna_3*;
- ◆ Region działalności (dane jakościowe: R1; R2; R3; R4; R5; R6; R7; R8; R9; R10) – *zmienna_4*.

Idea wykorzystania tablicy skoringowej jest taka, że każdemu dłużnikowi przyznawana jest odpowiednia liczba punktów dla każdej zmiennej predykcyjnej. Im wyższa liczba przyznanych punktów, tym większe prawdopodobieństwo, że analizowany dłużnik dokona pełnej lub częściowej spłaty długu. Parametry modelu skoringowego zostały tutaj określone tak, że liczba punktów równa 500 oznacza, iż dany dłużnik ma przypisane 50% prawdopodobieństwo spłaty swojej wierzytelności. Wszystkie wartości skoringu powyżej 500 oznaczają zatem większą (a wszystkie wartości skoringu do 500 punktów mniejszą), niż 50% szansę na dokonanie spłaty długu. Tablica skoringowa przedstawiona została w poniższej tabeli:

Tabela 7. Karta skoringowa.

	1 Nazwa zmienniej	2 Zakres wartości	3 WoE	4 Skoring cząstkowy
1	nz3	typ 1	22,923	125
2	nz3	typ 2	-44,563	29
3	nz3	typ 3	41,078	139
4	nz2a	0	38,42	142
5	nz2a	(0;0,06>	-23,981	70
6	nz2a	(0,06;0,4>	-48,057	42
7	nz2a	(0,40;0,60>	-109,258	-29
8	nz2a	(0,60;inf)	-70,602	16
9	nz4	R3	-15,02	49
10	nz4	r1	-31,484	62
11	nz4	R8	24,078	104
12	nz4	R5	9,667	82
13	nz4	R7	22,556	126
14	nz4	R10	64,258	173
15	nz4	R6	10,813	105
16	nz4	R9	49,256	146
17	nz4	R4	-10,347	70
18	nz4	R2	-26,208	60
19	nz1a	(0;600>	65,464	163
20	nz1a	(600;1600>	25,756	123
21	nz1a	(1600;5000>	-21,318	77
22	nz1a	(5000;inf)	-46,906	51

Ostateczny skoring ustalamy poprzez zsumowanie czterech cząstkowych skoringów. Dla przykładu dłużnik z zadłużeniem o wartości 1235PLN, odsetkiem spraw przedawnionych równym 0%, będący typem 3, z regionu R6 uzyskał następujące wartości skoringów cząstkowych:

- ◆ Wartość długu 1235PLN – skoring cząstkowy: 123 punkty (wiersz 20 w tabeli).
- ◆ Odsetek długu przedawnionego 0% – skoring cząstkowy: 142 punkty (wiersz 4 w tabeli).
- ◆ Typ działalności gospodarczej – typ 3 – skoring cząstkowy: 139 punktów (wiersz 3 tabeli).
- ◆ Region działalności gospodarczej R6 – skoring cząstkowy: 105 punktów (wiersz 15 tabeli).

Zatem łączny skoring dla takiego dłużnika wynosi: $123+142+139+105=509$ punktów. Moduł programu *STATISTICA Zestaw Skoringowy* dokonuje automatycznego przeliczenia wartości skoringu dla całego zbioru danych, podając dodatkowo dla każdego dłużnika (poza wartością punktową skoringu i skoringami cząstkowymi) również prawdopodobieństwo spłaty długu. To prawdopodobieństwo spłaty długu wykorzystywane jest następnie do szacowania możliwego zysku z całego portfela wierzytelności, zgodnie z zasadami opisanymi w pierwszej części artykułu.

Implementacja modelu skoringowego, ocena modelu, błędy prognozy

Ostatnim elementem wykorzystania zestawów skoringowych w obsłudze portfeli wierzytelności jest ustalenie prognozy wartości ściągniętego zadłużenia. Przypomnieć warto zatem, że przewidywaną wartość zadłużenia obliczano jako iloczyn prawdopodobieństwa spłaty, wartości zadłużenia oraz przeciętnego odsetka spłaty długu. Poniższa tabela zawiera zestawienie rzeczywistej spłaty długu w zbiorze uczącym oraz prognozowanej przez model skoringowy wartości spłaty.

Tabela 8. Błąd prognozy modelu.

	Wartość w PLN	Błąd szacunku w %
Spłata rzeczywista	3263200,25	-6,2%
Spłata prognozowana	3062149,70	

Porównanie rzeczywistej i prognozowanej spłaty długu pokazuje, że utworzony model jest dość dobrze dopasowany do danych. Błąd szacunku wyniósł -6,2%, co oznacza, że model miał tendencję do niedoszacowania wartości ściągniętego zadłużenia.

Warto ponadto przyjrzeć się zależności pomiędzy kategorią dłużników dokonujących pełnej lub częściowej spłaty swojego zadłużenia, a przyznawanymi tym dłużnikom przez model punktami skoringowymi.

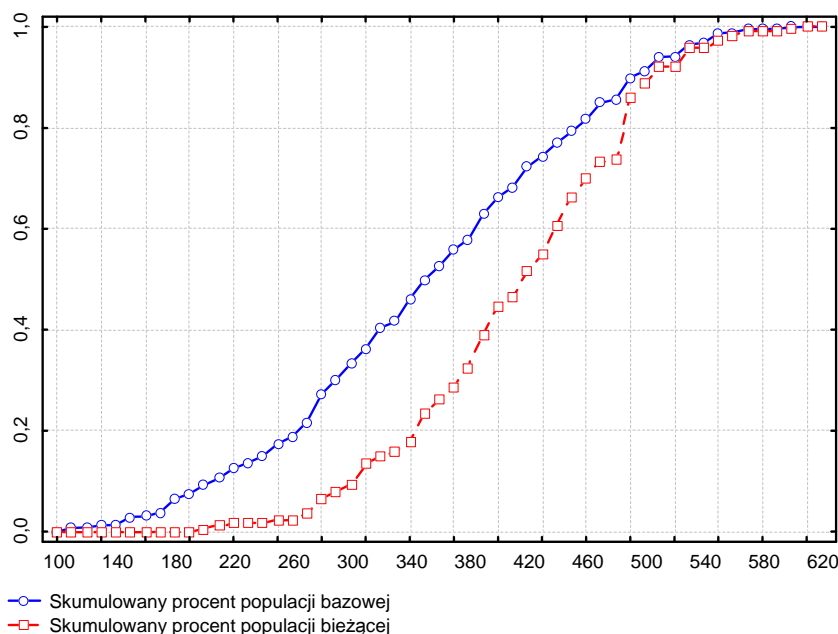
Tabela 9. Ocena modelu w oparciu o moc predykcyjną skoringu.

Wartość skoringu	Dobry	Zły	Suma	Portfel	Procent ogółem	IV	WoE
(-inf,200>	17	195	212	91,98%	7,63%	0,13	-158,10
(200,300>	103	539	642	83,96%	23,10%	0,12	-79,62
(300,450>	398	904	1302	69,43%	46,85%	0,00	3,84
(450,500>	157	206	363	56,75%	13,06%	0,05	58,72
(500,inf)	152	108	260	41,54%	9,36%	0,15	120,06
Ogół	827	1952	2779	70,24%	100,00%	0,45	

Kolumna *portfel* zawiera informację o procentowym udziale „złych” dłużników w każdej z pięciu klas wyodrębnionych ze względu na wyliczoną wartość skoringu, w kolumnie *procent ogółem* widać natomiast, jaki procent całości portfela stanowią dłużnicy przyporządkowani do jednej z tych pięciu klas. Powyższa tabela pokazuje też, że istnieje silne powiązanie pomiędzy przyznanymi przez model punktami skoringowymi a udziałem „złych” dłużników w analizowanym portfelu wierzytelności. Świadczy to o dobrym dopasowaniu do danych. Ponadto można zaobserwować, że najlepsza kategoria dłużników (o wartości skoringu powyżej 500 punktów) stanowi jedynie nieco ponad 9% całego portfela – dodatkowo najlepsza oznacza tyle, że znajduje się tam największa liczba „dobrych” dłużników, stanowiąca w tym przypadku prawie 59% wszystkich dłużników o skoringu

powyżej 500 punktów. Inaczej mówiąc, w grupie dłużników, którzy otrzymali powyżej 500 punktów, przeciętnie swoje zadłużenie spłaciło prawie 59% dłużników. W taki sam sposób analizuje się wszystkie pozostałe grupy ustalone na podstawie przedziałów skoringowych. Najbardziej liczną grupę stanowią przy tym dłużnicy o liczbie punktów od 300 do 450 – ponad 46% całości portfela, natomiast przeciętnie w tej grupie dług spłacił niespełna co trzeci dłużnik. Warto jeszcze dodać, że w najgorszej klasie dłużników – od 0 do 200 punktów skoringowych – przeciętnie swoje zadłużenie spłaciło niespełna 9% dłużników.

Oceniając rentowność nowego pakietu wierzytelności, uznamy go za potencjalnie lepszy od pakietu bazowego, wtedy gdy większy będzie udział procentowy dłużników o większej liczbie punktów skoringowych – oznacza to zwiększoną szansę na większy zysk w stosunku do pakietu bazowego. Jeżeli natomiast po wyznaczeniu skoringu okaże się, że w nowym portfelu występuje więcej dłużników o mniejszej liczbie punktów skoringowych, to pakiet ten uznany będzie za potencjalnie gorszy niż pakiet bazowy. Poniższy wykres ukazuje porównanie pakietu bazowego oraz pakietu „nowego” w zakresie skumulowanych rozkładów punktów skoringowych. Można z niego odczytać, że „nowy” portfel dłużników powinien dać większy zysk niż portfel bazowy. W „nowym” portfelu skumulowany rozkład punktów skoringowych „wolniej” zbliża się do jedności, zatem zawiera mniej dłużników o niskich wartościach skoringu.



Rys. 7. Ocena modelu w oparciu o moc predykcyjną skoringu.

Podsumowanie i wnioski

Przedstawione w artykule procedury analizy modeli skoringowych oraz uzyskane wyniki miały na celu zilustrowanie możliwości wykorzystania tablic skoringowych w obsłudze windykacyjnej portfeli wierzytelności. Wnioskowanie takie było, jak wielokrotnie



podkreślano, obarczone dużym marginesem niepewności, nie tyle z uwagi na złożoność procesu windykacji należności, co wnioskowania w oparciu o niewielki zasób wiedzy posiadanej o dłużnikach. Niemniej jednak utworzony model pozwolił uporządkować wiedzę o procesie windykacji, ukierunkować działania oraz wskazać potencjalną zyskowność portfeli dłużników. Warto jednocześnie podkreślić, iż zasadniczym celem budowy modelu było szacowanie ogólnej przewidywanej wartości ściągniętego zadłużenia, nie udało się natomiast w sposób sensowny przewidywać wartości spłaty pojedynczego dłużnika. Z uwagi na zadowalające parametry modelu, szczególnie w zakresie błędów szacunków modelu w stosunku do rzeczywistej globalnej wartości spłaty długu, można uznać, że narzędzia analizy danych w programie *STATISTICA 8* okazały się przydatne w parametrycznej ocenie portfeli wierzycelności.

Literatura

1. Agresti A., *Categorical Data Analysis*, Second Ed., Wiley Series in Probability and Statistics, 2002, ss. 165-210.
2. Anderson, R., *Theory and Practice for Retail Credit Risk Management and Decision Automation*, Oxford University Press, 2007, ss. 192-194.
3. Francuz, P., Mackiewicz, R., *Liczyby nie wiedzą skąd pochodzą. Przewodnik po metodologii I statystyce nie tylko dla psychologów*, Wydawnictwo KUL, 2005.
4. Migut G., *Wykorzystanie skoringu marketingowego do optymalizacji kampanii sprzedażowych*, w: *Zastosowania nowoczesnej analizy danych marketingowych w badaniach rynku*, StatSoft Polska Sp. z o.o., 2009, ss. 107-130.