



OCENA MODELI SCORINGOWYCH W SKOK STEFCZYKA

Katarzyna Cioch

Specjalista ds. Ryzyka, SKOK Stefczyka

Kamila Karnowska

Specjalista ds. Ryzyka, SKOK Stefczyka

Scoring SKOK Stefczyka

Scoring kredytowy w Spółdzielczej Kasie Oszczędnościowo Kredytowej im. F. Stefczyka funkcjonuje od 2006 roku i tak jak w innych instytucjach stosujących tego typu metodę oceny ryzyka kredytowego, ma na celu prognozowanie prawdopodobieństwa stania się „złym” klientem na podstawie porównania cech osób wnioskujących o kredyt w SKOK z cechami tych klientów, którzy korzystali w przeszłości z produktów w Kasie.

W SKOK Stefczyka klient podlegający ocenie scoringowej jest oceniany w oparciu o scoring aplikacyjny oraz o scoring behawioralny, w tym o ocenę Biura Informacji Kredytowej (BIK). Pierwszy rodzaj scoringu - aplikacyjny - bazuje na danych pozyskanych od klienta w momencie składania wniosku o kredyt i dotyczy informacji, które klient podaje w arkuszu aplikacyjnym, np.: stan cywilny, wykształcenie, adres, posiadany majątek itd. Scoring aplikacyjny w SKOK Stefczyka uzupełnia niezwykle ważny scoring behawioralny, który pozwala oceniać klienta w oparciu o jego historię kredytową. W tym wypadku bierze się pod uwagę informacje o długości tej historii w SKOK oraz o zachowaniu klienta podczas korzystania z produktu. Ten rodzaj oceny wzbogacony jest o scoring Biura Informacji Kredytowej, który umożliwia zweryfikowanie historii spłat w innych instytucjach oraz zapoznanie się z liczbą zapytań klienta o kredyt.

Z uwagi na dominujący typ kredytów detalicznych w SKOK Stefczyka przedmiotem tego wystąpienia będzie ocena modeli scoringowych zbudowanych dla klientów ubiegających się o produkt właśnie z tej grupy.

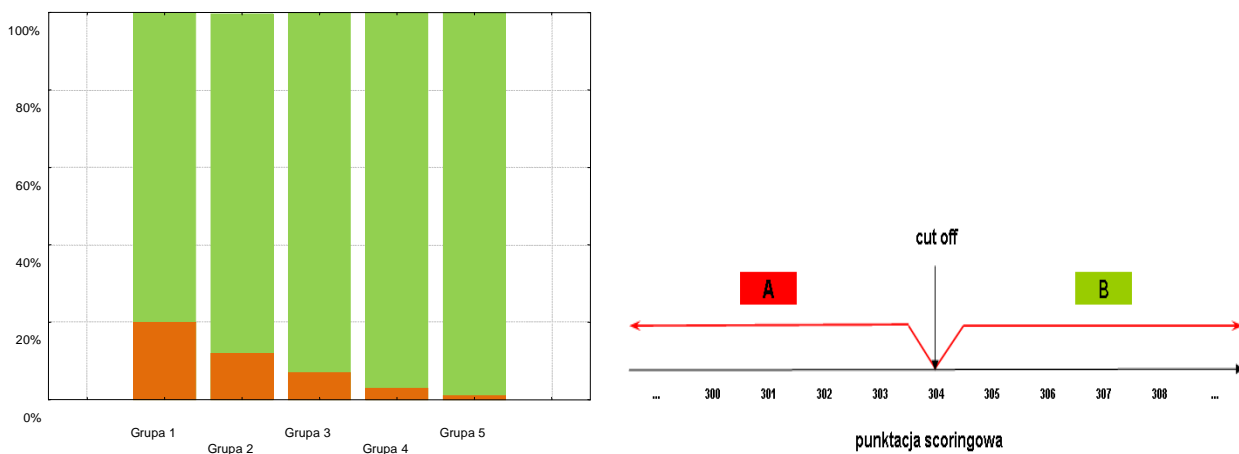
W ramach końcowej oceny klienta w SKOK Stefczyka funkcjonują tzw. grupy ryzyka. Grupa ryzyka jest kategorią, do której klasyfikowany jest potencjalny klient o ustalonym, na podstawie modelu scoringowego, prawdopodobieństwie niespłacalności zadłużenia. Poszczególne klasy ryzyka wyznacza się w oparciu o ocenę punktową (scoringową), którą klient uzyskał z modelu. Wyniki oceny punktowej zapisywane są w przedziałach, które ostatecznie wyznaczają poszczególne grupy.

Aktualnie w SKOK Stefczyka wykorzystywane są różne modele w zależności na przykład od kanału dystrybucji produktu - modele do oceny klientów ubiegających się o kredyt w oddziałach lub agencjach SKOK oraz modele do oceny klienta starającego się o pożyczkę internetową. Konieczność stosowania różnych typów ocen wiąże się z dużym zróżnicowaniem profilu i zachowania (spłacalności) klientów „anonimowych”, kontaktujących się poprzez Internet, a klientów, z którymi jest kontakt bezpośredni. Wszystkie sposoby oceny wykorzystują te same rodzaje scoringu, czyli aplikacyjny, behawioralny i ocenę BIK.

W SKOK Stefczyka obowiązuje pięć grup ryzyka dla klientów ubiegających się o kredyt oddziałowy. Przy czym klienci, w przypadku których istnieje duże prawdopodobieństwo trudności w obsłudze zadłużenia, są przypisywani do grupy o najwyższym ryzyku. Każda z grup charakteryzuje się pewnym poziomem ryzyka, który pokrywany jest w zależności od grupy ryzyka - ceną produktu i zabezpieczeniem. Poniżej przedstawiono interpretację poziomu zagrożenia w poszczególnych grupach ryzyka:

- ◆ Grupa 1 - klasa o najwyższym ryzyku;
- ◆ Grupa 2 - klasa o wysokim ryzyku;
- ◆ Grupa 3 - klasa o średnim ryzyku;
- ◆ Grupa 4 - klasa o niskim ryzyku;
- ◆ Grupa 5 - klasa o najniższym ryzyku.

Wśród klientów ubiegających się o pożyczkę drogą elektroniczną są wyznaczone dwie grupy: A i B, które oznaczają odpowiednio: grupa A - odmowa udzielenia pożyczki, grupa B - zgoda na udzielenie pożyczki.



Rys. 1. Przykładowy podział na grupy ryzyka.

Modele scoringowe mają bardzo duże znaczenie zarówno w kształtowaniu poziomu sprzedaży jak i ograniczaniu ryzyka kredytowego. W związku z tym istnieje potrzeba ciągłej weryfikacji i kontroli działania modeli w celu zapewnienia ich możliwie najwyższej jakości.



W celu właściwego, sprawnego i efektywnego wykorzystywania metod scoringowych przyjmuje się za konieczność, a nawet obowiązek systematycznego monitorowania systemu scoringowego. Skuteczność nawet najlepszych statystycznie modeli zależy w dużym stopniu od tego, w jaki sposób będą one monitorowane i zarządzane. Kontrola działania modeli odbywa się poprzez cykliczny comiesięczny monitoring oraz przeprowadzaną co najmniej raz na 12 miesięcy funkcjonowania modelu - walidację.

Monitoring modeli

Podstawowym założeniem regularnego procesu obserwacji modelu jest identyfikacja takich elementów, które mogą decydować o trwałej i znaczącej w porównaniu do sytuacji pierwotnej, zmianie działania modelu. W SKOK Stefczyka istotnym zadaniem cyklicznych analiz jest ocena poprawności działania modelu oraz wykrycie zmian i różnic w populacji.

Można wyróżnić kilka rodzajów czynników wpływających na działanie modelu scoringowego:

- ◆ **Zmiana struktury** obecnej populacji wnioskodawców w stosunku do populacji wejściowej, na podstawie której był budowany model - wymaga się stałego monitoringu struktury populacji klientów wnioskujących o kredyt, czyli analizy rozkładu wszystkich wniosków zarejestrowanych w systemie w poszczególnych klasach ryzyka oraz analizy rozkładu wniosków w grupie akceptowanej w zależności od klasy ryzyka;
- ◆ **Zmiana zachowania klientów** - należy obserwować poziom „złych” kredytów, zarówno w zależności od punktacji scoringowej, jak i z uwzględnieniem grup ryzyka w celu obserwowania zmian spłacalności kredytów udzielonych po wdrożeniu modelu; istotna jest również analiza przeterminowania portfela z podziałem na okres wypłaty kredytu;
- ◆ **Warunki ekonomiczne** - powinno się rejestrować momenty szczególnych zmian w otoczeniu firmy;
- ◆ **Działania marketingowe** - wymagane jest posiadanie informacji o akcjach marketingowych, które w znaczący sposób mogły wpłynąć na wzrost sprzedaży, na przykład w pewnej grupie wiekowej, produktowej lub klasie ryzyka;
- ◆ **Procedury windykacyjne** - istotnym elementem w udziale „złych” kredytów są zmiany w procesach windykacyjnych, stąd należy kontrolować rozwój procesów, których celem jest odzysk przeterminowanego zadłużenia;
- ◆ **Procesy udzielania kredytów** - zmiany w formułach produktowych, metodykach, polityce i strategii kredytowej.

Każdy z wymienionych czynników może powodować pojawianie się zmian w funkcjonowaniu modelu w różnym czasie, ich wpływ może następować bardzo szybko lub może być zauważalny nawet dopiero po kilku miesiącach. Stąd wymaga się stałego analizowania pewnych trendów i dynamiki, odzwierciedlających sposób udzielania kredytów oraz ich spłacalności. Ważnym elementem w procesie monitorowania jest

umiejętność przeprowadzenia niezbędnych analiz oraz właściwe odczytanie wyników, które mogą decydować o dalszym zastosowaniu modelu.

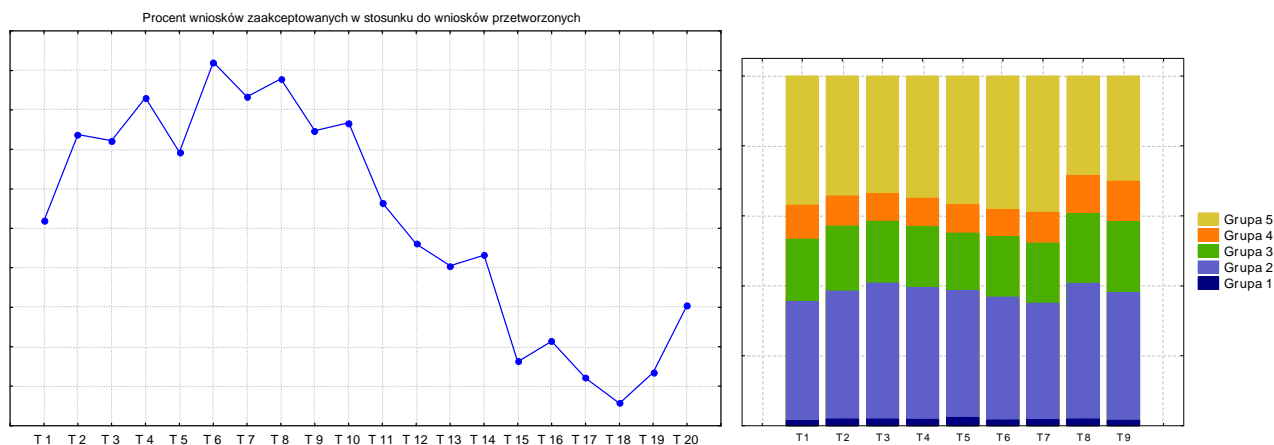
Skuteczne monitorowanie modelu scoringowego w SKOK Stefczyka wymaga analizowania raportów zarówno opartych na informacjach o kliencie (oznaczonych tutaj jako raporty typu A), jak i tych opartych na informacjach o spłacalności udzielonych kredytów w ujęciu portfelowym (raporty typu B). Uwzględniając różne czynniki, które mogą mieć wpływ na ocenę działania modelu, w SKOK Stefczyka wyróżnia się następujące raporty:

1. **Raporty typu A** - dostarczają wiedzę na temat populacji oraz informują o efektach polityki sprzedażowej.

- Liczba (lub wartość) wniosków przetworzonych i zaakceptowanych w poszczególnych miesiącach z uwzględnieniem punktacji scoringowej, grup ryzyka i całościowo;

Ten raport wykorzystuje informacje o wszystkich ocenianych przez model klientach. Poprzez generowanie raportu istnieje możliwość monitorowania trendu zmian poziomu wniosków zaakceptowanych w stosunku do wszystkich analizowanych wniosków w danym miesiącu. Pojawiające się odchylenia mogą świadczyć o istotnej zmianie w strukturze populacji klientów. Ponadto obserwowany jest rozkład wniosków zaakceptowanych w poszczególnych grupach ryzyka - tutaj należy monitorować, czy bieżąca struktura jest zgodna z pierwotnie przyjętą.

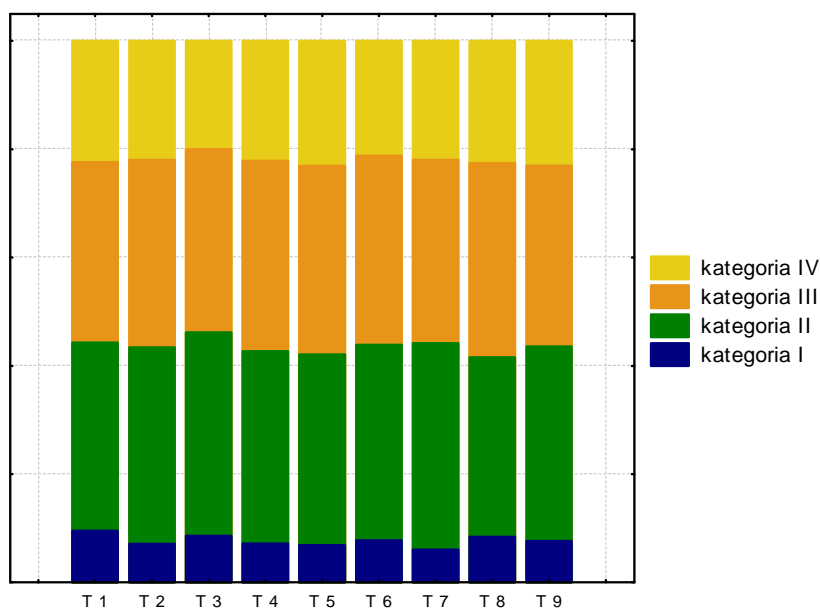
Poniższy przykład obrazuje duży spadek wniosków zaakceptowanych w stosunku do wszystkich analizowanych w okresie T10 - za co może odpowiadać np. zmiana polityki kredytowej, wprowadzenie zmian w algorytmie zdolności kredytowej. Natomiast weryfikując wykres po prawej stronie, zauważa się, że od okresu T8 nastąpiły pewne odchylenia w strukturze wniosków zaakceptowanych - nastąpił spadek w grupie o najniższym ryzyku (grupa 5) i wzrost w grupie o wysokim poziomie ryzyka (grupa 2). Ta sytuacja może świadczyć o zmianie profilu klienta.



Rys. 2. Udział wniosków zaakceptowanych wzg. wniosków przetworzonych (wykres po lewej) oraz rozkład wniosków zaakceptowanych w zależności od grupy ryzyka (wykres po prawej).

- Analiza wybranych cech użytych w modelu

Raport porównuje rozkład wyników uzyskanych dla poszczególnych kategorii pewnej cechy lub cech w każdym badanym miesiącu. Wówczas istnieje możliwość monitorowania zmian zachodzących w profilu klientów – np. można obserwować, czy z miesiąca na miesiąc rośnie udział klientów w określonej kategorii wiekowej lub z dochodem z określonego przedziału i czy taka zmiana wiąże się jednocześnie ze zmianą w rozkładzie wniosków w zależności od punktacji.



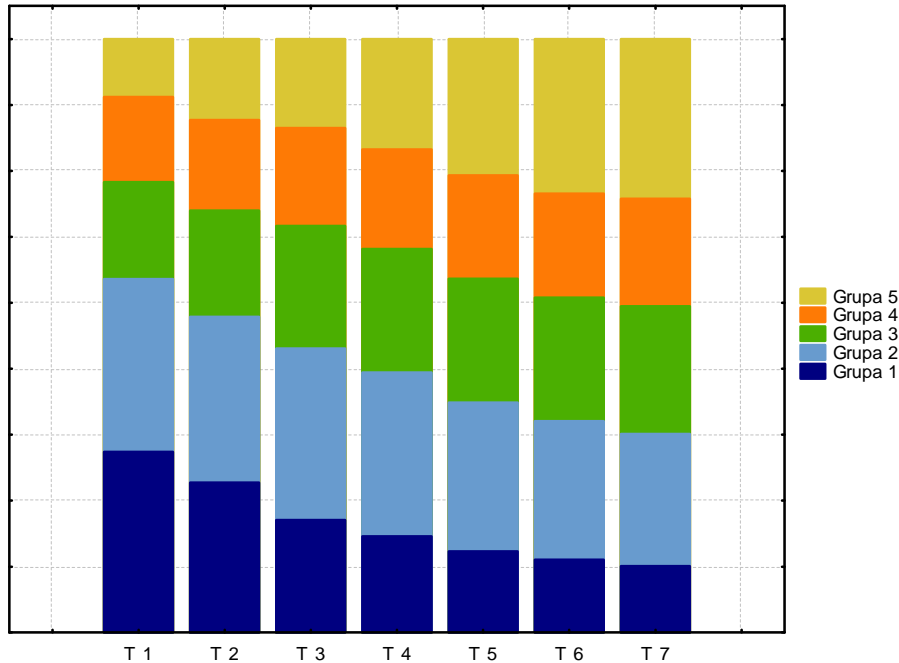
Rys. 3. Rozkład klientów w zależności od kategorii zmiennej wiek w poszczególnym okresie działania modelu.

Podany przykład, ilustrujący rozkład klientów w zależności od poszczególnych kategorii zmiennej *wiek*, wskazuje na utrzymującą się strukturę tej zmiennej w okresie T1-T9. Minimalne odchylenia w poszczególnych kategoriach wiekowych nie powodują istotnego zaburzenia, co może świadczyć, że profil klienta ze względu na wiek nie uległ zmianie.

2. **Raporty typu B** - służą weryfikacji jakości portfela kredytowego. Należy tutaj badać poziom niewypłacalności klientów. Celem tych analiz jest ocena, czy model prawidłowo klasyfikuje klientów w zależności od stopnia ryzyka.

- Struktura „złych” kredytów w zależności od punktacji scoringowej i w poszczególnych klasach ryzyka - porównanie do poprzedniego miesiąca, okresu.

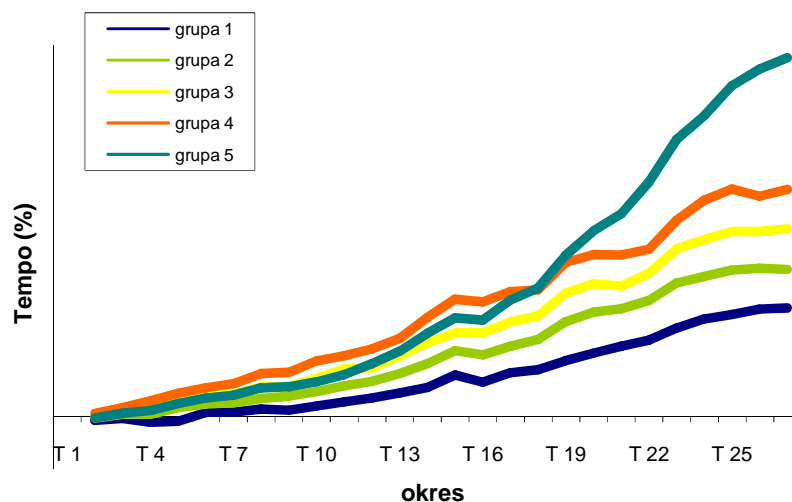
Należy obserwować, czy wnioski o wyższej punktacji lub te kwalifikujące się do grup o niższym ryzyku mają rzeczywiście lepsze wyniki. Powinno się badać, czy udział „złych” kredytów w tych kategoriach jest niższy i odpowiada pierwotnie przyjętej strukturze „złych” klientów. Poniższy rysunek przedstawia przykładowy rozkład „złych” klientów w grupach ryzyka.



Rys. 4. Rozkład „złych” kredytów w poszczególnej grupie ryzyka w okresie od T1 do T7.

Przykład wskazuje na niestabilny udział „złych” kredytów w poszczególnych grupach ryzyka. Występuje dość wyraźna tendencja spadku szkodowych produktów w grupie 1 i 2 oraz wzrost w grupach traktowanych jako najmniej ryzykowne. Zauważa się przesunięcie punktacji wśród „złych”- stąd należy rozważyć przesunięcie granic dla poszczególnych grup ryzyka - można to jednak stwierdzić dopiero po dalszych analizach jakości działania modelu.

- Analiza bieżącego przeterminowania kredytów w każdej grupie ryzyka (wskaźnik wyrażający saldo kredytów przeterminowanych do salda kredytów ogółem)



Rys. 5. Tempo zmian wskaźnika przeterminowania powyżej 30 dni dla poszczególnej grupy ryzyka w odniesieniu do okresu T1.



Tutaj proponowanym rozwiązaniem jest analiza tempa zmian wskaźnika przeterminowania w poszczególnych grupach ryzyka (zob. rys. 5).

W tym wypadku obserwuje się niebezpieczny wzrost wskaźnika wśród klientów zakwalifikowanych do grupy o najniższym ryzyku. Niepokojące jest tempo wzrostu tego wskaźnika w grupie 5 w porównaniu do pozostałych grup, w których przyjmuje się poziom szkodowości wyższy.

Sugeruje się zbadać strukturę „złych” kredytów w grupach ryzyka i w okresach, w których wykonano badanie tempa.

- Miary jakości modelu

Monitorowanie poprawności działania modelu scoringowego wykonuje się w SKOK Stefczyka w oparciu o wskaźniki wykorzystywane do badania zdolności modelu do rozróżniania „złych” i „dobrych” klientów (współczynnik dywergencji), a także miary informujące o zdolności modelu do oceniania klientów wg związanego z nimi poziomu ryzyka (współczynnik Giniego). Szczegółowy opis miar oraz sposoby ich interpretacji zostały zamieszczone w kolejnym rozdziale.

- Raport Vintage

Raport ten jest analizą stanu kredytów w kolejnych miesiącach trwania (spłaty) rozumianą jako dynamiczna analiza szkodowości kredytów w czasie. Weryfikacja dotyczy kredytów wypłaconych w poszczególnych miesiącach. Natomiast stan kredytów wyraża wskaźnik przeterminowania, który jest wyznaczony jako stosunek przeterminowanego salda (np. dni przeterminowania > 30) kredytów wypłaconych w danym miesiącu do kwoty wszystkich kredytów wypłaconych w tym okresie. Stan kredytów jest badany na koniec każdego miesiąca. Raport jest wykonywany cyklicznie, a wyniki porównywane w miesięcznych odstępach. Wskaźnik przeterminowania w raporcie Vintage jest liczony według następującego wzoru:

$$\text{Wskaźnik przeterminowania} = \frac{\text{Saldo kredytów przeterminowanych}}{\text{Kwota wypłacona}}$$

Analiza Vintage uwzględnia dwa raporty.

Raport 1 - pokazuje wskaźnik przeterminowania kredytów w kolejnym miesiącu i roku kalendarzowym.



Tabela 1. Przykład raportu Vintage.

	2006 10	2006 11	2006 12	2007 01	2007 02	2007 03	2007 04	2007 05	2007 06	2007 07	2007 08	2007 09	2007 10	2007 11	2007 12	2008 01	2008 02
2006-10	0,0%	0,0%	1,1%	0,0%	0,0%	1,1%	1,1%	1,1%	1,1%	1,1%	3,0%	1,9%	1,9%	3,9%	3,2%	3,2%	3,4%
2006-11		0,0%	0,0%	0,0%	0,4%	0,4%	1,2%	1,4%	1,4%	1,0%	1,3%	1,1%	1,2%	1,9%	1,8%	1,9%	1,6%
2006-12			0,0%	0,0%	0,1%	0,9%	0,9%	1,2%	0,9%	1,6%	1,5%	2,5%	2,0%	2,0%	2,0%	2,3%	2,4%
2007-01				0,0%	0,0%	0,0%	0,2%	0,7%	1,8%	1,8%	2,2%	2,4%	2,5%	2,5%	3,4%	3,6%	3,3%
2007-02					0,0%	0,1%	0,1%	0,5%	0,5%	0,6%	0,9%	1,8%	2,1%	2,0%	1,8%	1,7%	1,6%
2007-03						0,0%	0,1%	0,2%	0,2%	0,2%	0,6%	0,8%	1,0%	1,4%	1,9%	2,1%	2,0%
2007-04							0,0%	0,0%	0,3%	0,2%	0,2%	0,2%	0,2%	0,4%	0,9%	1,0%	0,8%
2007-05								0,0%	0,1%	0,1%	0,3%	0,4%	0,2%	0,3%	1,3%	0,9%	0,5%
2007-06									0,0%	0,0%	0,1%	0,0%	0,4%	0,8%	0,9%	1,7%	1,7%
2007-07										0,0%	0,0%	0,2%	0,3%	0,8%	1,9%	1,7%	1,2%
2007-08											0,0%	0,0%	0,9%	0,8%	1,1%	1,5%	1,6%
2007-09												0,0%	0,0%	0,0%	0,4%	0,4%	0,9%
2007-10													0,0%	0,0%	0,7%	0,7%	1,2%
2007-11														0,0%	0,2%	1,0%	0,3%
2007-12															0,0%	0,0%	0,1%
2008-01																0,0%	0,5%
2008-02																	0,0%

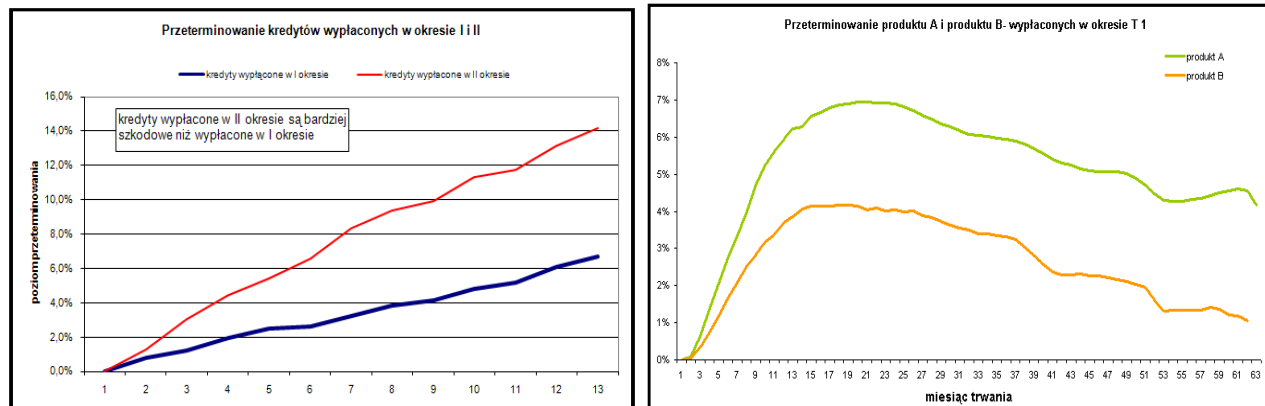
Raport 2 - pokazuje wskaźnik przeterminowania kredytów w poszczególnym miesiącu trwania.

Tabela 2. Przykład raportu Vintage.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
2006-10	0,0%	1,1%	0,0%	0,0%	1,1%	1,1%	1,1%	1,1%	1,1%	3,0%	1,9%	1,9%	3,9%	3,2%	3,2%	3,4%
2006-11	0,0%	0,0%	0,4%	0,4%	1,2%	1,4%	1,4%	1,0%	1,3%	1,1%	1,2%	1,9%	1,8%	1,9%	1,6%	
2006-12	0,0%	0,1%	0,9%	0,9%	1,2%	0,9%	1,6%	1,5%	2,5%	2,0%	2,0%	2,0%	2,3%	2,4%		
2007-01	0,0%	0,0%	0,2%	0,7%	1,8%	1,8%	2,2%	2,4%	2,5%	2,5%	3,4%	3,6%	3,3%			
2007-02	0,1%	0,1%	0,5%	0,5%	0,6%	0,9%	1,8%	2,1%	2,0%	1,8%	1,7%	1,6%				
2007-03	0,1%	0,2%	0,2%	0,2%	0,6%	0,8%	1,0%	1,4%	1,9%	2,1%	2,0%					
2007-04	0,0%	0,3%	0,2%	0,2%	0,2%	0,2%	0,4%	0,9%	1,0%	0,8%						
2007-05	0,1%	0,1%	0,3%	0,4%	0,2%	0,3%	1,3%	0,9%	0,5%							
2007-06	0,0%	0,1%	0,0%	0,4%	0,8%	0,9%	1,7%	1,7%								
2007-07	0,0%	0,2%	0,3%	0,8%	1,9%	1,7%	1,2%									
2007-08	0,0%	0,9%	0,8%	1,1%	1,5%	1,6%										
2007-09	0,0%	0,0%	0,4%	0,4%	0,9%											
2007-10	0,0%	0,7%	0,7%	1,2%												
2007-11	0,2%	1,0%	0,3%													
2007-12	0,0%	0,1%														
2008-01	0,5%															

Uwaga: oba raporty pokazują te same wartości wskaźnika, tylko różnią się sposobem ich prezentacji.

Raport Vintage pozwala weryfikować, np. czy kredyty udzielane w ostatnich miesiącach są mniej szkodowe niż te wypłacone w poprzednim okresie – na podstawie poziomu przeterminowania (lub poziomu „złych” kredytów). Ponadto ten typ raportu umożliwia doszukiwanie się pewnych analogii - badając jakość (spłacalność) kredytów wypłacanych w różnych terminach, ale np. na podobnych warunkach (np. czy produkt wypłacony na oświadczenie w styczniu 2008 roku spłaca się podobnie jak ten wypłacony w styczniu 2007). W analizie raportu Vintage pomocne są wykresy - przykładowe przedstawiono poniżej.



Rys. 6. Przykładowe wykresy do raportu Vintage.

- Raport o odstępstwach

Celem tego raportu jest analiza cykliczna poziomu akceptacji wniosków kredytowych klientów niespełniających wszystkich warunków formuły danego typu produktu. Jest to raport informujący o wielkości tzw. przełamań decyzyjnych, które stanowią odstępstwo od standardowej formuły produktowej. Raport zawiera liczbę oraz wartość kwotową kredytów udzielonych z odstępstwem w poszczególnych okresach. Analizie podlega również kategoria odstępstw oraz spłacalność tych kredytów z uwzględnieniem średniego poziomu przeterminowania w grupach produktowych.

Walidacja modeli

Walidacja modelu scoringowego jest całościowym, cyklicznym procesem oceny modelu. Stanowi ona istotny element rozwoju kart scoringowych. Celem walidacji jest dokonanie jakościowej i ilościowej oceny działania modelu. W skład jakościowej oceny wchodzi: merytoryczna ocena rozwiązań zastosowanych przy budowie i wdrożeniu modelu, ocena danych oraz ocena funkcjonalności operacyjnej zastosowanych rozwiązań. Ocena ilościowa obejmuje badanie mocy dyskryminacyjnej i stabilności modelu. Po wykonaniu analizy i ocenie modelu powstaje raport walidacyjny. Na jego podstawie należy podjąć decyzję o ewentualnym wprowadzeniu zmian w modelu lub wycofaniu go z użycia. Wyróżnia się następujące etapy walidacji:

Analiza danych pod kątem kompletności, aktualności i jakości

Podczas generowania danych do zbioru walidacyjnego dokonuje się analizy danych (statystyki opisowe, rozkłady zmiennych, badanie współzależności). Na podstawie analizy podejmuje się decyzje, które zmienne są niekompletne, nieaktualne lub wątpliwej jakości i rekomenduje odpowiednie zmiany w systemie. Ponadto zgłasza się błędy i nadużycia odpowiednim komórkom SKOK Stefczyka.

Badanie stabilności cech i populacji

Badanie stabilności sprawdza strukturę obecnej populacji w stosunku do populacji bazowej (populacji, na której zbudowano model). Analizie podlegają zmiany w rozkładzie kredytobiorców dla każdej cechy i całej populacji względem wartości punktów lub atrybutów w określonym czasie. W celu sprawdzenia różnic w strukturze zbioru walidacyjnego i zbioru uczącego przeprowadza się badanie stabilności cech i populacji, posługując się miarą WSP. WSP (współczynnik stabilności populacji) - porównuje rozkład badanej grupy kredytobiorców (populacja bieżąca) z rozkładem kredytobiorców z próby, która została wykorzystana do zbudowania tablicy scoringowej (populacja bazowa); określa zmiany zachodzące w populacji (rozkład punktacji lub rozkład atrybutów poszczególnych zmiennych) w czasie od momentu budowy modelu do jego walidacji.

Współczynnik stabilności populacji wyrażony jest wzorem:

$$WSP = \sum_{i=1}^n (\%W_i - \%U_i) * \ln\left(\frac{\%W_i}{\%U_i}\right)$$

gdzie:

n - liczba atrybutów danej cechy,

%W - udział procentowy cechy w populacji bieżącej,

%U - udział procentowy danej cechy w populacji bazowej.

Stosuje się następującą interpretację współczynnika WSP:

- ◆ $WSP \leq 0,1$ - obecna populacja zachowuje się podobnie jak populacja wyjściowa;
- ◆ $0,1 < WSP < 0,25$ - obecna populacja nieznacznie odbiega od populacji wyjściowej;
- ◆ $WSP \geq 0,25$ - obecna populacja znacznie odbiega od populacji bazowej.

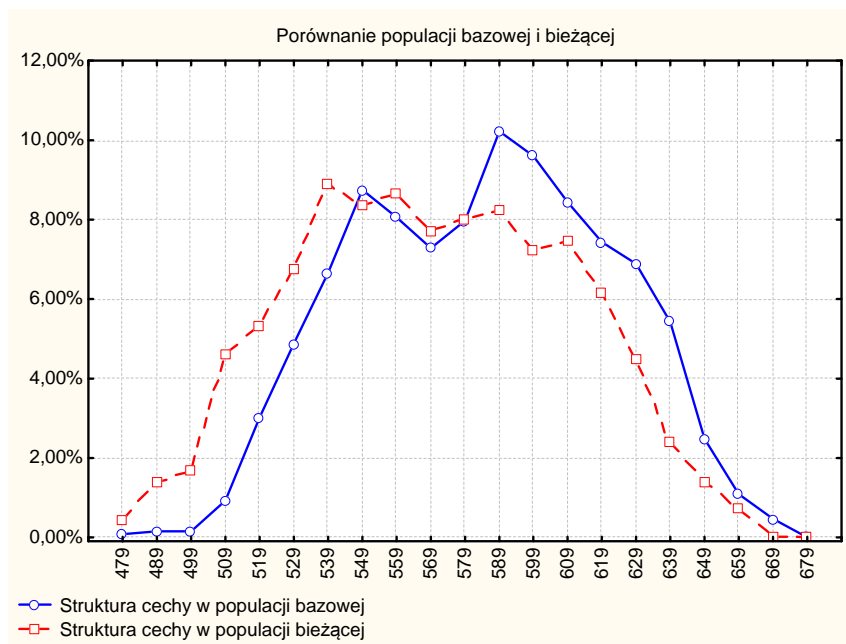
Poniższa tabela i wykres ilustrują przykładowy rozkład wniosków w poszczególnych przedziałach punktacji scoringowej w zbiorze walidacyjnym i uczącym. Przesunięcie wykresu dla populacji bieżącej w lewo w stosunku do wykresu dla populacji bazowej informuje, że nastąpiło nieznaczne zwiększenie udziału klientów o niższej punktacji, ocenianych przez model jako gorzej spłacający, i nieznaczne zmniejszenie udziału klientów o wyższej punktacji.

Tabela 3: Badanie stabilności populacji – przykładowa tabela danych.

	Liczba przypadków dla populacji bazowej	Liczba przypadków dla populacji bieżącej	Struktura cechy w populacji bazowej	Struktura cechy w populacji bieżącej	Zmiana struktury	Wskaźnik zmiany	Weight of evidence	Struktura wskaźnika stabilności populacji
$479 \leq x < 489$	1	11	0,08%	0,45%	0,369%	5,766	1,752	0,006
$489 \leq x < 499$	2	34	0,15%	1,38%	1,226%	8,911	2,187	0,027
$499 \leq x < 509$	2	41	0,15%	1,66%	1,51%	10,745	2,374	0,036
$509 \leq x < 519$	12	113	0,93%	4,59%	3,658%	4,936	1,597	0,058
$519 \leq x < 529$	39	131	3,02%	5,32%	2,298%	1,761	0,566	0,013



	Liczba przypadków dla populacji bazowej	Liczba przypadków dla populacji bieżącej	Struktura cechy w populacji bazowej	Struktura cechy w populacji bieżącej	Zmiana struktury	Wskaźnik zmiany	Weight of evidence	Struktura wskaźnika stabilności populacji
529<=x<539	63	166	4,88%	6,74%	1,86%	1,381	0,323	0,006
539<=x<549	86	220	6,66%	8,93%	2,271%	1,341	0,293	0,007
549<=x<559	113	206	8,75%	8,36%	-0,389%	0,956	-0,045	0,
559<=x<569	104	213	8,06%	8,65%	0,592%	1,074	0,071	0,
569<=x<579	94	190	7,28%	7,71%	0,433%	1,059	0,058	0,
579<=x<589	103	198	7,98%	8,04%	0,061%	1,008	0,008	0,
589<=x<599	132	203	10,22%	8,24%	-1,983%	0,806	-0,216	0,004
599<=x<609	124	178	9,60%	7,23%	-2,378%	0,752	-0,284	0,007
609<=x<619	109	184	8,44%	7,47%	-0,973%	0,885	-0,122	0,001
619<=x<629	96	152	7,44%	6,17%	-1,265%	0,83	-0,186	0,002
629<=x<639	89	111	6,89%	4,51%	-2,387%	0,654	-0,425	0,01
639<=x<649	70	59	5,42%	2,40%	-3,027%	0,442	-0,817	0,025
649<=x<659	32	34	2,48%	1,38%	-1,098%	0,557	-0,585	0,006
659<=x<669	14	18	1,08%	0,73%	-0,354%	0,674	-0,395	0,001
669<=x<679	6	1	0,46%	0,04%	-0,424%	0,087	-2,438	0,01
679<=x<689	0	0	0,00%	0,00%	0,%	0,	0,	0,
Suma	1291	2463	100,00%	100,00%				0,222

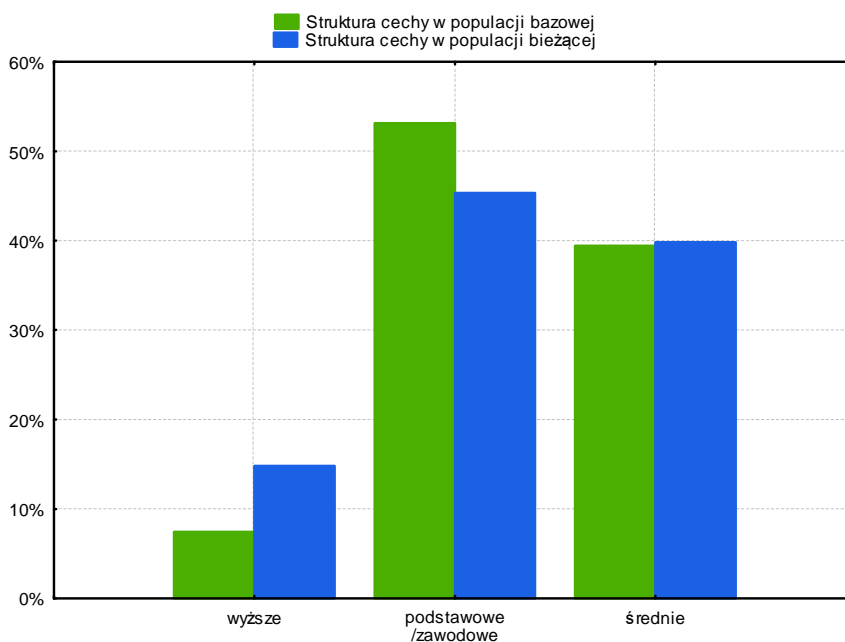


Rys. 7. Badanie stabilności populacji – przykładowy wykres.

W celu identyfikacji zmiennych, na których wystąpiły największe różnice należy przeprowadzić badanie stabilności cech. Poniższa tabela i wykres przedstawiają strukturę przykładowej zmiennej.

Tabela 4. Badanie stabilności cechy – przykładowa tabela danych.

	Liczba przypadków dla populacji bazowej	Struktura cechy w populacji i bazowej	Liczba przypadków dla populacji bieżącej	Struktura cechy w populacji bieżącej	Zmiana struktury	Punktacja	Różnica w punktacji
Wyższe	96	7,44%	365	14,82%	7,383%	63	4,651
Podstawowe/Zawodowe	686	53,14%	1117	45,35%	-7,786%	58	-4,516
Średnie	509	39,43%	981	39,83%	0,403%	60	0,242
Suma	1291	100,00%	2463	100,00%			0,377



Rys. 8. Badanie stabilności populacji – przykładowy wykres.

Brak stabilności cech lub całej populacji jest przesłanką do wprowadzenia zmian w modelu lub budowy nowego modelu. Brak stabilności oznacza bowiem, że populacja oceniana przez model dzisiaj nie jest podobna do tej, na której model był budowany, a założenie o podobieństwie teraźniejszości do przeszłości jest podstawą budowy i funkcjonowania modelu scoringowego.

Analiza miar jakościowych cech i modelu

Za pomocą miar jakościowych cech i modelu przeprowadza się analizy zdolności modelu do prawidłowej klasyfikacji klientów w zależności od związanego z nimi ryzyka oraz



analizy mocy dyskryminacyjnej (predykcijnej) modelu. Moc predykcijna modelu określa, jak dobrze model odróżnia klienta „złego” od „dobrego”.

Współczynnik IV

Współczynnik IV (Information Value – moc informacyjna zmiennej) odnosi się do oddzielenia rozkładu wyników dla kredytów „dobrych” i „złych”. Do obliczenia współczynnika stosuje się średni wynik dla rozkładu każdej grupy, jak również statystyczny rozrzut wyników w ramach każdego rozkładu. Im bardziej oddalone są od siebie rozkłady wyników dla kredytów „dobrych” i „złych” (tj. im większy jest współczynnik IV), tym większy jest potencjał prognostyczny modelu w zakresie odróżniania kredytów „dobrych” od „złych”.

Współczynnik IV wyrażony jest wzorem:

$$IV = \sum (P_g - P_b) \cdot \ln\left(\frac{P_g}{P_b}\right)$$

gdzie:

P_g - oznacza udział procentowy „dobrych” kredytów w każdym przedziale wyników,

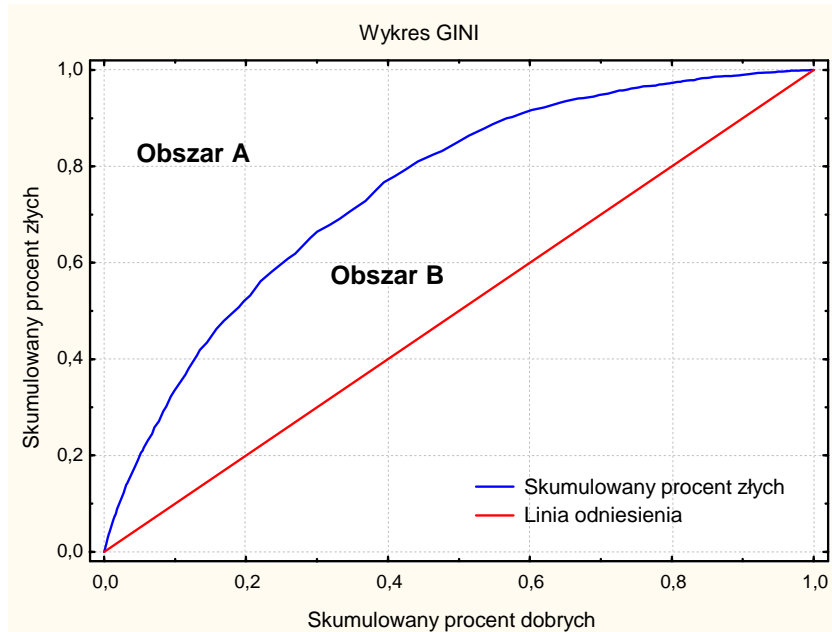
P_b - oznacza udział procentowy „złych” kredytów w każdym przedziale wyników.

Stosuje się następującą interpretację współczynnika IV:

- ◆ $IV < 0,02$ - brak mocy predykcijnej;
- ◆ $0,02 \leq IV < 0,1$ - słaba moc predykcijna;
- ◆ $0,1 \leq IV < 0,3$ - średnia moc predykcijna;
- ◆ $IV \geq 0,3$ - silna moc predykcijna.

Współczynnik Giniego

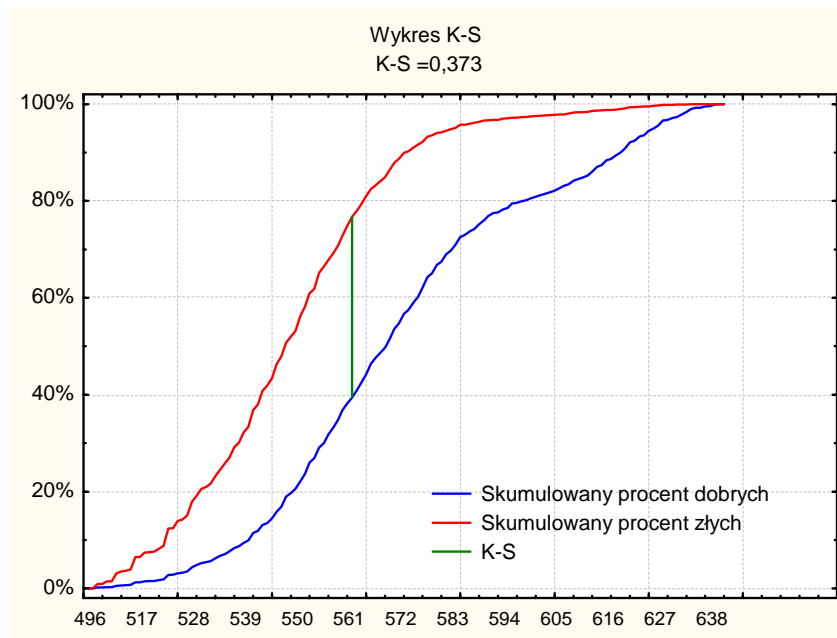
Współczynnik Giniego służy do pomiaru przewagi karty punktowej nad systemem losowo podejmowanych decyzji. W szczególności przedstawia on stosunek wielkości pola obszaru B - nad przekątną i pod krzywą (oznaczającego efektywność podejmowania decyzji w oparciu o wyniki oceny punktowej) - do wielkości pola obszaru A+B - nad przekątną (oznaczającego losowo podejmowane decyzje). Wartości współczynnika mieszczą się w przedziale $\langle -1, 1 \rangle$. Wyższa wartość współczynnika informuje o lepszym segregowaniu klientów. Spadek wartości współczynnika Giniego poniżej 0,35 oznacza, że model stracił zdolność poprawnej oceny. Poniższy rysunek przedstawia przykładowy wykres Giniego.



Rys. 9. Przykładowy wykres Giniego.

Statystyka Kołmogorowa-Smirnowa

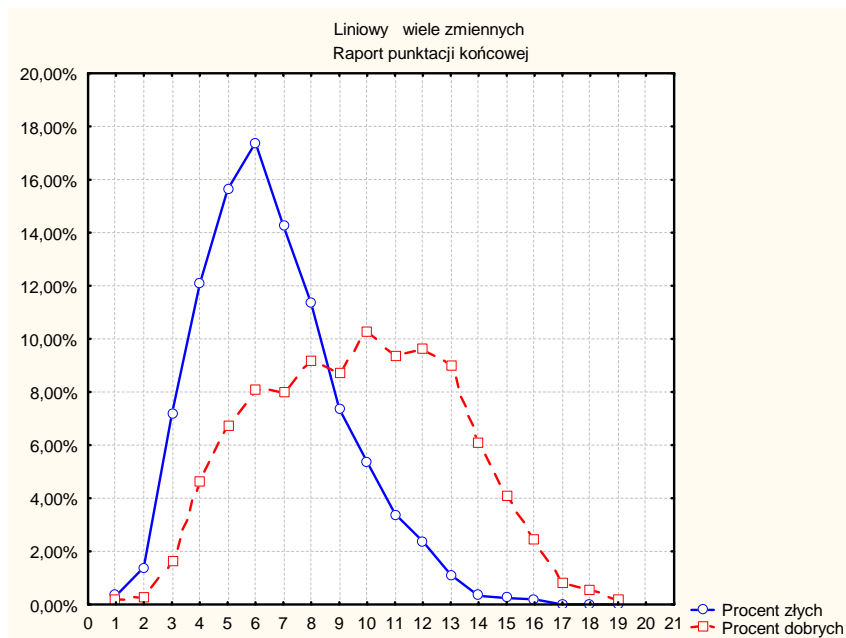
Statystyka Kołmogorowa-Smirnowa (K-S) służy do ustalania maksymalnej różnicy pomiędzy skumulowanym procentem „dobrych” i „złych” kredytów. Im wyższa wartość K-S, tym większy potencjał modelu. Jeżeli wartość K-S wynosi 0, model nie jest w stanie rozróżnić kredytów „dobrych” i „złych”, natomiast wartość K-S równa 100 oznacza zdolność do całkowitego rozróżnienia tych dwóch rozkładów. Poniższy rysunek przedstawia przykładowy wykres K-S.



Rys. 10. Przykładowy wykres K-S.

Dywergencja

Dywergencja jest miarą pokazującą zdolność modelu scoringowego do separowania „dobrych” i „złych” klientów. Bazuje ona na średnim wyniku dla rozkładu klientów „dobrych” i „złych” oraz zróżnicowaniu wyników w ramach każdego z tych rozkładów. Im bardziej oddalone od siebie są te rozkłady, tym większy jest potencjał prognostyczny modelu. W praktyce zakłada się, że dla poprawnie działającego modelu scoringowego wartość współczynnika dywergencji jest większa od 0,5. Poniższy rysunek przedstawia przykładowy wykres dywergencji.



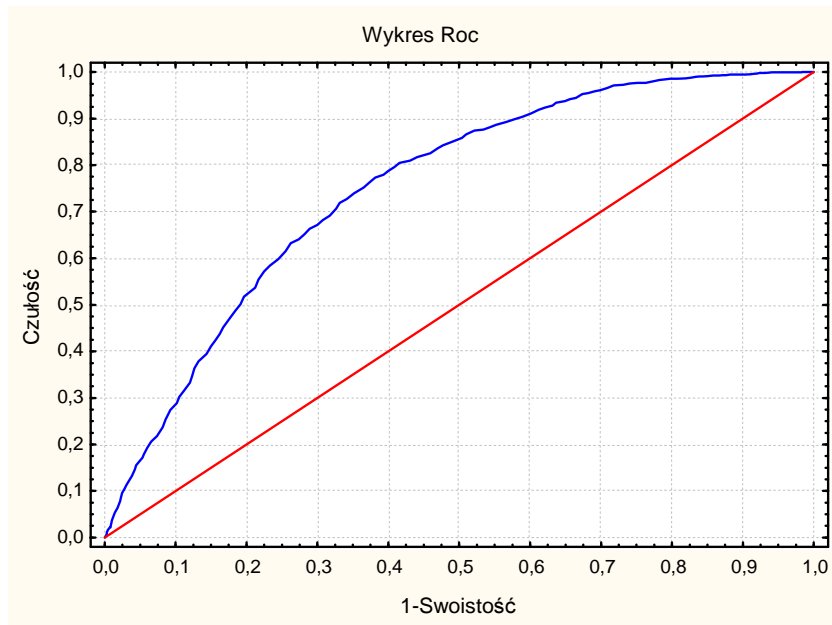
Rys. 11. Przykładowy wykres dywergencji.

Krzywa ROC

Krzywa ROC (*Receiver Operating Characteristic Curve*) umożliwia graficzne przedstawienie dokładności diagnostycznej modelu. Wykres krzywej ROC przedstawia zależność wskaźników odpowiadających odpowiednio za czułość i swoistość modelu. Czułość modelu określa zdolność modelu do właściwego klasyfikowania „złych” kredytów do grupy „złych”, natomiast swoistość określa zdolność modelu do wykrywania kredytów „dobrych” i przypisywania ich do tej właśnie grupy. Oceną modelu na podstawie krzywej ROC jest pole nad krzywą (im większe, tym lepsze). Stosuje się następującą interpretację:

- ◆ Pole = 1 – model doskonały;
- ◆ $0,9 \leq \text{pole} < 1$ – model bardzo dobry;
- ◆ $0,8 \leq \text{pole} < 0,9$ – model dobry;
- ◆ $0,7 \leq \text{pole} < 0,8$ – model słaby;
- ◆ pole = 0,5 – model losowy;
- ◆ pole < 0,5 – model nieprawidłowy (gorszy niż losowy).

Poniższy rysunek przedstawia przykładowy wykres krzywej ROC.

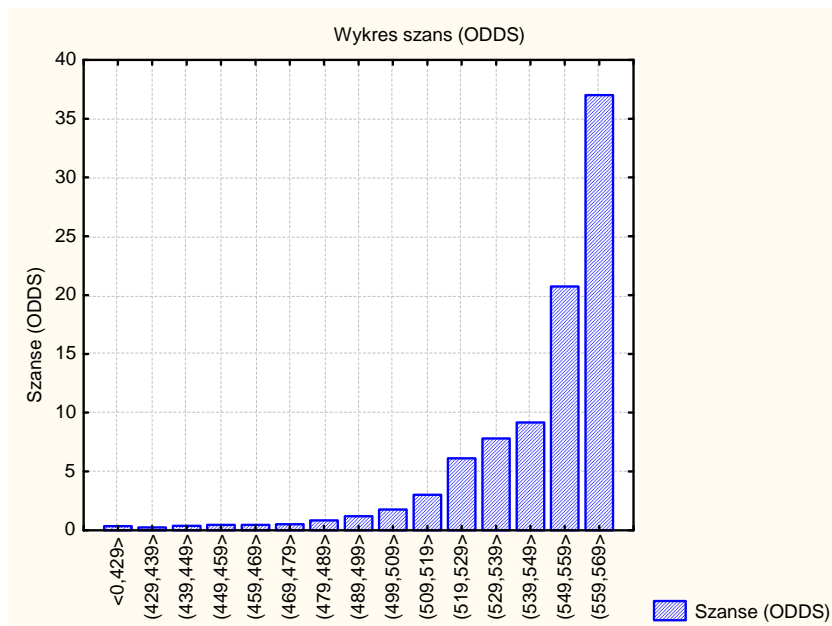


Rys. 12. Przykładowa krzywa ROC.

Po wygenerowaniu wszystkich miar należy porównać wartości poszczególnych współczynników dla zbioru uczącego i walidacyjnego oraz ocenić różnice pomiędzy wielkościami współczynników.

ODDS

ODDS (*odds ratio* – *iloraz szans*) jest to stosunek prawdopodobieństwa, że zdarzenie A wystąpi, do prawdopodobieństwa, że to zdarzenie nie wystąpi.



Rys. 13. Przykładowy wykres ODDS.

W praktyce wartość ODDS w danym przedziale wyraża stosunek liczby „dobrych” klientów do liczby „złych” klientów w tym przedziale. Powyższy rysunek przedstawia przykładowy wykres ODDS.

Wskaźnik ODDS wyraża się wzorem:

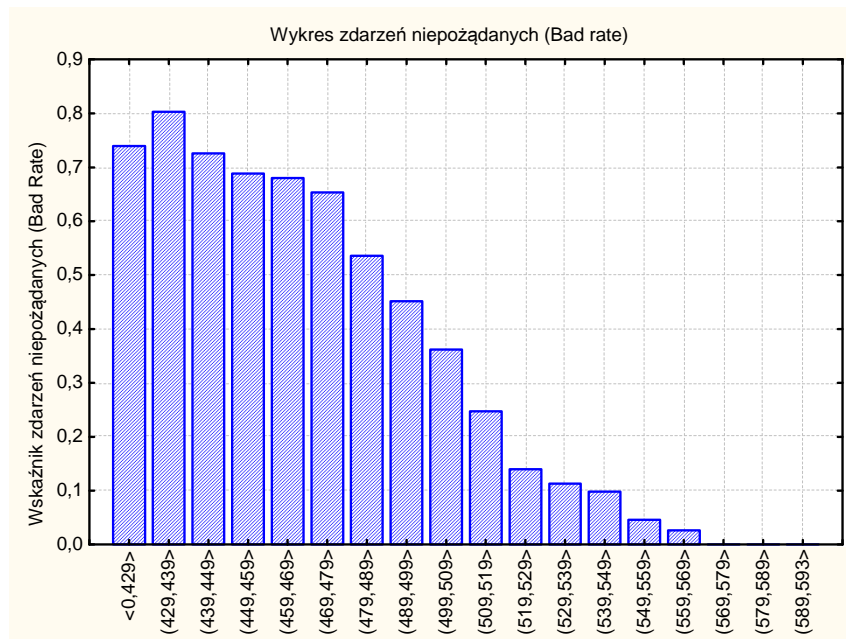
$$ODDS = \frac{P(A)}{1 - P(A)}$$

gdzie:

A - oznacza zdarzenie – „dobry” klient wystąpi w danym przedziale.

Bad Rate

Wskaźnik bad rate oznacza stosunek liczby „złych” klientów do liczby wszystkich klientów w danym przedziale. W praktyce oznacza udział procentowy „złych” klientów w danym przedziale. Poniższy rysunek przedstawia przykładowy wykres wskaźnika *Bad Rate*.



Rys. 14. Wykres Bad Rate dla zbioru uczącego.

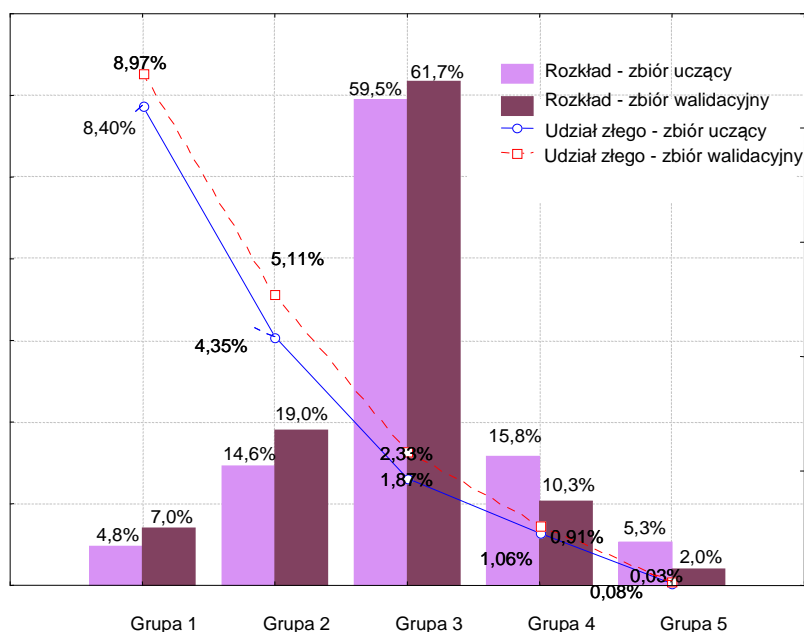
Podczas oceny modelu należy porównać rozkłady ODDS i *Bad Rate* w zbiorze uczącym i walidacyjnym. Wykresy ODDS i *Bad Rate* badanych zbiorów powinny wyglądać podobnie i układać się w logiczny trend. Jeśli wykresy nie są monotoniczne, oznacza to, że wraz ze spadkiem punktacji nie rośnie (choć powinno) prawdopodobieństwo zajścia złego zdarzenia.

Badanie stabilności rozkładu oceny scoringowej

Badanie stabilności rozkładu oceny scoringowej jest porównaniem rozkładu na grupy ryzyka w poszczególnych miesiącach trwania życia modelu. Badanie to zostało szczegółowo omówione w rozdziale 2.

Badanie procentów default

Dzięki badaniu procentów default można stwierdzić, jaki jest udział „złego” klienta w poszczególnych grupach ryzyka w zbiorze uczącym oraz w zbiorze walidacyjnym. Badanie polega na porównaniu udziału „złych” i „dobrych” klientów w poszczególnych grupach ryzyka zbioru uczącego i walidacyjnego. Należy sprawdzić, czy udział procentowy „złego” klienta oraz rozkład grup ryzyka różni się istotnie w badanych zbiorach. Poniższy rysunek przedstawia przykładowy rozkład „dobrych” i „złych” klientów.



Rys. 15. Rozkład „dobrych” i „złych” klientów w poszczególnych grupach ryzyka dla zbioru uczącego i walidacyjnego.

Dobre praktyki

Zasady raportowania

Stosowanie monitoringu modeli scoringowych jest konieczne, lecz aby było efektywne, musi być przeprowadzane mądrze i w oparciu o podstawowe zasady raportowania, do których należą:

- ◆ Cykliczność (regularność) - raporty powinny być wykonywane w równych odstępach czasu (miesiąc, kwartał), ale dopuszcza się sporządzenie dodatkowych raportów kontrolnych, np. w chwili zajścia awarii systemu.



- ◆ Porównywalność i czytelność - unifikacja typów raportów oraz sposobów prezentacji wyników. Należy przyjąć taki schemat przedstawiania wyników, aby był możliwy do odczytania i zaobserwowania zmian.
- ◆ Unifikacja interpretacji wyników - standaryzacja w rozumieniu pewnych zachodzących zjawisk czy miar powinna mieć miejsce z uwagi na konieczność zastosowania właściwych środków w chwili, kiedy zostanie stwierdzona nieprawidłowość w działaniu modelu. Jeżeli przyjmujemy, że ze wzrostem punktacji powinna maleć szkodowość kredytów, to przy zajściu sytuacji odwrotnej nie należy traktować takiego zjawiska jako sytuacji dopuszczalnej.
- ◆ Wnioski i rekomendacje - w chwili zaobserwowania istotnych zmian bądź różnic w uzyskanych wynikach należy, po przeanalizowaniu przyczyn, sformułować rekomendacje. Te z kolei pozwolą podjąć ostateczną decyzję w kwestii dalszych działań w procesie monitorowania modeli.

Dziennik zdarzeń

Programy monitorujące modele scoringowe często odnoszą się do tzw. *Portfolio Chronology Log*, czyli do listy zdarzeń, które mogą mieć wpływ na jakość portfela kredytowego lub modelu scoringowego. Prowadzenie takiego dziennika zmian i zdarzeń istotnych należy także do zasadniczych działań w procesie monitoringu. Oprócz tego, że taki rejestr jest podstawowym sposobem analizy wpływu różnych czynników na model, jest również jednym z kluczowych elementów świadomego kontrolowania działalności instytucji na tle zmian zachodzących wewnątrz niej oraz na rynku zewnętrznym. Poprzez tworzenie tego typu dziennika można skuteczniej ustalić przyczynę obserwowanych zjawisk mających wpływ na wynik portfela kredytowego lub na profil klientów, a ostatecznie na monitorowany model.

Decyzje

Proces monitoringu stosowany jest w celu gromadzenia różnego rodzaju danych, wyników, aby na podstawie tych informacji mieć możliwość podjęcia właściwych działań. Stąd, każdy proces monitorowania modeli scoringowych powinien mieć pewien schemat działania w chwili zaobserwowania zmian w funkcjonowaniu modelu. Należy przede wszystkim pamiętać, że model scoringowy powinien być monitorowany tuż po wdrożeniu, po to, aby wykluczyć wszelkie błędy programowania i kodowania modelu. Jeśli te pozostaną niezauważone, wówczas narażamy się na utratę jakości modelu, choć w sytuacji braku błędów kodu, mógłby działać bardzo dobrze.

Zmiany w funkcjonowaniu modelu nie zawsze oznaczają pogorszenie jego jakości, dlatego należy najpierw ocenić skalę występujących różnic oraz zidentyfikować ich źródło. To właśnie poprzez regularne gromadzenie informacji przez kilka miesięcy można rozróżnić zmiany tymczasowe od stałych. W chwili pojawienia się „niegroźnych” zmian często wystarczy wykonać raport powtórnie w kolejnym okresie, aby stwierdzić, że zjawisko się nie powtarza. Oczywiście ważne jest, aby umieć ocenić powód wystąpienia takiej sytuacji. Gdy np. różnice są spowodowane awarią systemu lub wynikają z jakiegoś błędu



kodowania modelu, nie ma potrzeby przebudowywać modelu. Natomiast, jeśli zmiany pojawiają się cyklicznie, należy przystąpić do wprowadzenia pewnych modyfikacji, np. korekty wartości wagi dla atrybutu cechy lub przesunięcia punktu odcięcia. O ile takie rozwiązanie nie jest wystarczające, należy natychmiast przeprowadzić raport walidacyjny modelu scoringowego, którego wyniki pomogą określić dalsze postępowanie, czyli przebudowę modelu lub budowę nowego.

Po wykonaniu raportu walidacyjnego, jeśli zachodzi taka potrzeba, należy podjąć decyzje o wprowadzeniu zmian w modelu lub wycofaniu go z użycia. W zależności od zdiagnozowanych błędów w funkcjonowaniu modelu, zmian w otoczeniu zewnętrznym lub zmian w polityce kredytowej, modyfikacje modelu mogą polegać na przykład na: wprowadzeniu innego rozkładu na grupy ryzyka, przesunięciu punktu odcięcia, obniżeniu lub podwyższeniu wagi niektórych cech, zmianie rozkładu atrybutów danej cechy. Przesłankami do rezygnacji z modelu mogą być na przykład: istotne obniżenie mocy predykcyjnej modelu lub niestabilna populacja.

Procedury

Odpowiednią jakość modelu scoringowego przez cały cykl jego życia od momentu budowy, poprzez wdrożenie, funkcjonowanie i ocenę zapewni dobrze zdefiniowany proces, procedury i instrukcje. Dokumenty te powinny w szczególności definiować:

- ◆ Zakres odpowiedzialności;
- ◆ Częstotliwość przeprowadzania oceny i raportowania;
- ◆ Wzory raportów;
- ◆ Definicje stosowanych mierników i ich wartości graniczne;

Opis postępowania w przypadku przekroczenia wartości granicznych.

Oprogramowanie

W procesie budowy i oceny funkcjonowania modelu scoringowego w SKOK Stefczyka istotną rolę odgrywa automatyzacja pewnych wyliczeń lub generowania wykresów - niezbędnych w procesie tworzenia raportów. W tym celu wykorzystywany jest program *STATISTICA 9*, ze szczególnym uwzględnieniem dodatku *Zestaw Scoringowy*.

Dzięki funkcjonalności tych produktów SKOK Stefczyka uzyskała następujące korzyści:

- ◆ Skrócony czas potrzebny na budowę i ocenę modeli - zwiększenie wydajności i skierowanie uwagi na aspekty merytoryczne, a nie techniczne.
- ◆ Standaryzacja postępowania przy budowie i monitorowaniu modeli.
- ◆ Możliwość prezentacji wyników w dowolnej formie.
- ◆ Czytelność i przejrzystość wyników.