



AgEcon SEARCH
RESEARCH IN AGRICULTURAL & APPLIED ECONOMICS

The World's Largest Open Access Agricultural & Applied Economics Digital Library

This document is discoverable and free to researchers across the globe due to the work of AgEcon Search.

Help ensure our sustainability.

Give to AgEcon Search

AgEcon Search
<http://ageconsearch.umn.edu>
aesearch@umn.edu

*Papers downloaded from **AgEcon Search** may be used for non-commercial purposes and personal study only. No other use, including posting to another Internet site, is permitted without permission from the copyright owner (not AgEcon Search), or as allowed under the provisions of Fair Use, U.S. Copyright Act, Title 17 U.S.C.*

**ANNALS OF THE POLISH ASSOCIATION
OF AGRICULTURAL AND AGRIBUSINESS ECONOMISTS**

ROCZNIKI NAUKOWE
STOWARZYSZENIA EKONOMISTÓW ROLNICTWA I AGROBIZNESU

Received: 22.12.2022

Acceptance: 08.02.2023

Published: 22.03.2023

JEL codes: C36, D8, G21, H81

Annals PAAAE • 2023 • Vol. XXV • No. (1)

License: Attribution 3.0 Unported (CC BY 3.0)

DOI: 10.5604/01.3001.0016.2473

MARIUSZ DACKO^{1*}, IRENEUSZ KURCZYNA^{}**

^{*}Uniwersytet Rolniczy w Krakowie, Polska, ^{**}Agencja Ratingowa BS sp. z o.o., Polska

ZASTOSOWANIE MODELU DRZEWA KLASYFIKACYJNEGO W OCENIE RYZYKA KREDYTOWEGO

Słowa kluczowe: drzewo klasyfikacyjne, Data Mining, banki spółdzielcze,
spłaty kredytów, ryzyko kredytowe

ABSTRAKT. Celem pracy była ocena możliwości wykorzystania modelu drzewa klasyfikacyjnego, jako narzędzia wspierającego działalność banku przez ograniczenie ryzyka kredytowego. Omówiono kwestie ryzyka bankowego i kredytowego oraz przedstawiono kwestię możliwości wykorzystania zaawansowanych modeli statystycznych w charakterze narzędzi wspomagających działalność banków. Zaprezentowano model drzewa klasyfikacyjnego, pozwalającego ograniczyć ryzyko w podejmowaniu decyzji kredytowej. Baza danych obejmowała 1308 wniosków kredytowych, pozytywnie rozpatrzonych w latach 2015-2016, które opisano zestawem 27 cech potencjalnego kredytobiorcy. Skonstruowany na tej podstawie model z trafnością 93% wskazywał statystyczne prawidłowości, pozwalające ustalić *a priori* starających się o kredyt klientów, którzy wywiążą się z umowy oraz klientów, którzy prawdopodobnie będą mieli problemy ze spłatą kredytu. Pierwszym kryterium podziału zbiorowości klientów na „dobrych” i „złych” były zaległości w spłacie zobowiązań w okresie ostatnich trzech lat. Inne istotne czynniki ryzyka, to posiadanie limitu w ROR, posiadanie rachunku w banku udzielającym kredytu oraz deklarowana przez wnioskodawcę kwota jego obciążeń. Ważny był także wiek kredytobiorcy oraz jego wykonywany zawód. Na tle uzyskanych wyników przedyskutowano możliwości, jakie oferuje model drzewa klasyfikacyjnego, zwrócono uwagę na jego intuicyjność i łatwość jego interpretacji. Omówiono korzyści stosowania takich modeli w zarządzaniu ryzykiem towarzyszącym działalności kredytowej prowadzonej przez banki spółdzielcze.

¹ Corresponding author: m.dacko@ur.krakow.pl

WPROWADZENIE

W ostatnich dekadach rozwinięta technologia cyfrowa pozwoliła na przetwarzanie coraz większych zbiorów danych, a sektor bankowy zaczął wspomagać się zaawansowanymi narzędziami ograniczania ryzyka i utrzymywania równowagi między bezpieczeństwem a rozwojem. Do takich narzędzi coraz częściej należą zaawansowane modele Data Mining. Posługując się nimi, analityk bankowy może upewnić się co do słuszności podjętej decyzji kredytowej.

Kluczową pozycję w sprawozdaniu finansowym każdego banku zajmują należności od sektora niefinansowego, tj. wartość udzielonych kredytów. Sprzedaż kredytów przynosi działalności bankowej największe dochody, ale jednocześnie niesie ona za sobą duże ryzyko. Szybkie i trafne przewidywanie ryzyka kredytowego oraz jego kontrola są dla każdego banku poważnym wyzwaniem i aby było to możliwe, potrzebne są dane, wiedza i zaawansowane programy komputerowe.

Ryzyko kredytowe, jako jedna z determinant ryzyka działalności bankowej wymieniane jest zazwyczaj obok ryzyka płynności, ryzyka stopy procentowej i ryzyka dewizowego [Bereza 1995]. Jest więc ono jednym z wielu ryzyk, które ponosi bank. Ryzyko kredytowe określa się jako ryzyko nieoczekiwanego niewykonania zobowiązania przez klienta banku lub pogorszenia się jego zdolności kredytowej, zagrażającej wykonaniu takiego zobowiązania.

Kryzys finansowy trwający w latach 2007-2009, długość jego trwania, ogólnoswiatowy charakter i dotkliwość jego skutków wskazują, że ryzyko w działalności bankowej, w tym w szczególności ryzyko kredytowe, jest zjawiskiem coraz poważniejszym. Zarządzanie bankiem jest wyzwaniem nie tylko z tego powodu. Rozważając kwestie przyszłości zarządzania w biznesie, Piotr Płoszajski [2005] stwierdził, że obecny świat podlega zmianom tak głębokim i wielopostaciowym, że w ich efekcie wyłania się często rzeczywistość niepoddająca się analizie z wykorzystaniem pojęć istniejących w naukach o zarządzaniu.

Nie tylko aparat pojęciowy, ale i dotychczasowe koncepcje i instrumenty zarządzania zdają się ulegać przyspieszonej dezaktualizacji [Płoszajski 2005]. Już pod koniec lat 70. XX wieku Janusz Gościński [1977] przewidywał, że sukcesy w zarządzaniu w zmieniających się warunkach w coraz większym stopniu zależeć będą od nowoczesności rozwiązań w dziedzinie struktur i systemowych metod kierowania procesami. To spostrzeżenie jest także adekwatne do kredytowej działalności banków i stosowanych przez nie systemów oceny ryzyka. Stawia ono nowe wyzwania przed zarządami banków spółdzielczych i ich wyspecjalizowanymi komórkami, takimi jak działy kredytów (spełniające główne funkcje związane z udzielaniem kredytów i obsługą klientów w tym zakresie), działy informatyki (prowadzące nadzór nad systemami informatycznymi i oprogramowaniem), a w szczególności przed zespołami oceny ryzyka kredytowego (weryfikującymi propozycje

udzielenia kredytu przed podjęciem decyzji kredytowej, oceniającymi kompletność dokumentacji kredytowej, zdolność kredytową klienta i adekwatności zabezpieczeń i ryzyko transakcji).

Należy zauważyć, że wiele z czynników ryzyka kredytowego ma charakter synergiczny. Zważywszy na fakt wzajemnych powiązań czynników tego ryzyka, wydaje się oczywiste, że ich analiza statystyczna da lepsze rezultaty, gdy będzie ona miała charakter wielowymiarowy i zastosowane zostaną modele nieliniowe. Ponadto należy wziąć pod uwagę to, że w odniesieniu do determinant ryzyka kredytowego badane zależności nie muszą być monotoniczne. Jednoosobowe gospodarstwo domowe może np. lepiej radzić sobie ze spłatą kredytów niż gospodarstwo dwu- lub trzyosobowe, ale dalsze zwiększanie liczby osób może już stanowić czynnik zwiększonego ryzyka. Ponadto tam, gdzie gospodarstwo domowe tworzą dwie osoby lub więcej – pojawia się aspekt współkredytobiorcy i wspólnoty majątkowej. W takich okolicznościach zawodzić mogą proste metody badawcze, takie jak testy nieparametryczne i modele regresji.

Warto w próbach oceny ryzyka kredytowego rozważyć metodę drzew. Metoda ta zaliczana jest, podobnie jak sieci neuronowe, do metod Data Mining [Dacko, Sroka 2010]. Metoda ta jest szczególnie predestynowana do dużej liczby danych i jest wolna od ograniczeń, względem których klasyczne modele regresji zwykle zawodzą. Drzewa są szczególnie przydatne w rozmaitych zagadnieniach predykcyjnych i eksploracyjnych. Znalazły szerokie zastosowanie w badaniach społeczno-ekonomicznych [Łapczyński 2003, 2007]. Warto przypomnieć, że stosując drzewo klasyfikacyjne oparte na algorytmie C&RT, dwaj ekonomiści, Paolo Manasse i Nouriel Roubini [2009], zdobyli rozgłos międzynarodowy, przedstawiając algorytm oceny ryzyka upadłości i niewypłacalności państw.

W ograniczaniu ryzyka kredytowego metodę drzew zastosowały Dorota Witkowska i Mariola Chrzanowska [2006] na próbie 2272 klientów banków. Podobne zagadnienie rozpatrywał Grzegorz Migut [2003], prezentując podejście typowe dla filozofii Data Mining – w ramach jednego projektu dane o klientach banków analizowano w trzech modelach: sieć neuronowa, drzewa C&RT i drzewa wzmacniane. W badaniach Grzegorza Miguta (podobnie jak w części empirycznej tego opracowania) wszystkie modele częściej myliły się podczas rozpoznawania przypadków, dla których ocena klienta była negatywna. Warto też odnotować, że w przypadku oceny negatywnej lepiej od drzew sprawdziły się sieci neuronowe. Neuronowe sieci klasyfikacyjne mogą z powodzeniem (podobnie jak drzewa) przewidywać niebezpieczeństwo niedotrzymania warunków umowy przez kredytobiorcę, który nie spłaci całości lub części kwoty kredytu i odsetek. W odniesieniu do tego problemu metodę sieci neuronowych zastosowali Paweł Baster i Katarzyna Pocztońska [2011]. Autorzy wykorzystali duży zbiór obserwacji z polskiego banku komercyjnego – 3000 rachunków kredytowych otwartych w okresie od stycznia 2000 roku do września 2001 roku. Każdy przypadek był opisany zestawem 13 zmiennych objaśniających. Jednak, co ciekawe, w prezentowanych obliczeniach sieć neuronowa

wypadała gorzej od alternatywnych propozycji: modeli logitowych i probitowych. O znacznej popularności sieci neuronowych świadczy to, że w polskich badaniach nad ryzykiem bankowym tę metodę wykorzystywali także Piotr Mierzejewski [1998] oraz Aleksandra Wójcicka-Wójtowicz [2018].

Z punktu widzenia banku każdy niespłacony kredyt stanowi przykład nie tylko nietrafnej decyzji kredytowej, ale jest to przede wszystkim zmniejszenie potencjalnych zysków i przyczynę do konieczności tworzenia rezerw celowych. Niestety, równie szkodliwa dla banku byłaby także nadmierna ostrożność, skutkująca nieudzieleniem kredytu klientowi, który wywiązałby się ze swojego zobowiązania. Dlatego zarządzanie ryzykiem jest kluczową sferą funkcjonowania każdego banku. Istotą działań w tym obszarze jest identyfikacja i pomiar ryzyka [Baster, Pochtowska 2011]. W tej kwestii bardzo pomocne wydają się modele Data Mining, a w ich ramach model drzewa klasyfikacyjnego, stanowiący przedmiot poniższych badań empirycznych.

Celem pracy była ocena możliwości wykorzystania modelu drzewa klasyfikacyjnego, jako narzędzia wspierającego działalność banku przez ograniczenie ryzyka kredytowego.

MATERIAŁ I METODYKA BADAŃ

W pracy wykorzystano dane gromadzone przez pięć banków spółdzielczych na potrzeby opracowania i wdrożenia systemu scoringowego. Dane zostały udostępnione przez Agencję Ratingową BS sp. z o.o. z siedzibą w Dąbrowie Górniczej, na mocy umowy o współpracy podpisanej przez tę instytucję z Uniwersytetem Rolniczym im. H. Kołłątaja w Krakowie. Punktem wyjścia do budowy modelu było 1308 wniosków kredytowych rozpatrzonych pozytywnie w latach 2015-2016. Każdy przypadek opisano zestawem 27 charakterystyk, opisujących transakcję, profil socjodemograficzny, sytuację ekonomiczno-finansową wnioskodawcy oraz relacje klienta z bankiem i innymi instytucjami finansowymi (tabela 1). W maksymalnie dwuletnim okresie kredytowania w badanej grupie około 11% klientów nie wywiązało się z warunków umowy kredytowej. Problemy w spłacie kredytu, przesądzające o klasyfikacji klientów jako „złych”, były rozpatrywane pod kątem terminowości (opóźnienie w spłacie wynoszące co najmniej 30 dni) oraz istotności (kwota zaległości wyższa niż 100 zł).

W badaniach wykorzystano metodę binarnych drzew klasyfikacyjnych C&RT, dostępnych w programie Statistica wyposażonym w moduł Data Mining. Metoda wykorzystuje zasadę rekurencyjnego podziału zbioru obserwacji na podzbiory. Jej celem jest uzyskanie maksymalnej jednorodności podzbiorów ze względu na przynależność obiektów do klas [Łapczyński 2003, 2005]. Podzbiory tworzące strukturę hierarchiczną nadają kształt drzewu. W programie Statistica składa się ono z wierzchołka obrazującego pierwszą decyzję podziału, węzłów dzielonych oraz węzłów końcowych [Sokołowski 2002]. Drzewo nie musi, a nawet nie powinno mieć zbyt dużej liczby węzłów.

Tabela 1. Predyktory uwzględnione w modelu drzewa klasyfikacyjnego C&RT

Charakterystyka	Predyktor
Transakcja kredytowa	<ul style="list-style-type: none"> – kwota kredytu, – współkredytobiorca, – rodzaj zabezpieczenia kredytu, – wartość zabezpieczenia przekraczająca kwotę wnioskowanego kredytu, – deklarowana kwota obciążeń wnioskodawcy
Profil socjodemograficzny	<ul style="list-style-type: none"> – wiek, – płeć, – stan cywilny, – wspólnota majątkowa, – miejsce zamieszkania, – wykształcenie, – wykonywany zawód, – miejsce pracy
Sytuacja ekonomiczna	<ul style="list-style-type: none"> – miesięczny dochód netto gospodarstwa domowego, – główne źródło dochodu, – liczba osób w gospodarstwie domowym, – status mieszkaniowy
Relacje klienta z bankiem i innymi instytucjami finansowymi	<ul style="list-style-type: none"> – posiadanie rachunku w BS, – posiadanie innych produktów depozytowych, oszczędnościowych lub inwestycyjnych w BS, – obecne zaangażowanie kredytowe, – limit w ROR, – posiadanie limitu karty kredytowej, – inne zobowiązania, – wskaźnik DTI, – zaleganie w spłacie jakichkolwiek zobowiązań w okresie ostatnich 3 lat, – prowadzone działania windykacyjne, – liczba punktów w BIK

Źródło: opracowanie własne na podstawie danych udostępnionych przez ZRBS

Reguły klasyfikacyjne powinny być raczej proste [Misztal 2000]. Dlatego ustalając parametry modelu C&RT przyjęto, że:

- koszty błędnych klasyfikacji będą równe,
- dopasowanie modelu będzie oceniane według miary Giniego,
- proces poszukiwania i tworzenia podziałów zostanie zatrzymany przez przycinanie przy błędzie złej klasyfikacji,
- węzły końcowe będą składać się z co najmniej 5 obserwacji,
- kontrola jakości uzyskiwanych wyników nastąpi przy wykorzystaniu V-krotnego sprawdzianu krzyżowego dla $V = 10$.

Wyniki modelu zaprezentowano w formie tabelarycznej wraz z opisem słownym, koncentrując się na interpretacji diagramu drzewa i kryteriów jego podziału.

WYNIKI BADAŃ

Drzewa binarne C&RT to jedna z efektywniejszych metod Data Mining, umożliwiająca odkrywanie wzorców i zależności ukrytych w danych [Dacko i in. 2019]. Metoda ta została zaproponowana przez Leo Breimana i zespół [1998] już ponad dwie dekady temu. Wydaje się ona szczególnie ciekawą propozycją w bankowości, ponieważ pozwala na operowanie relatywnie dużą liczbą powiązanych ze sobą czynników sprawczych. Tak właśnie jest w istocie – ryzyko kredytowe zależy od wielu aspektów jednocześnie. Dlatego, o tym czy klient będzie „dobry”, czy „zły” wnioskowano na podstawie licznej grupy charakterystyk, w tym 8 predyktorów ilościowych i 19 predyktorów jakościowych (tabela 1).

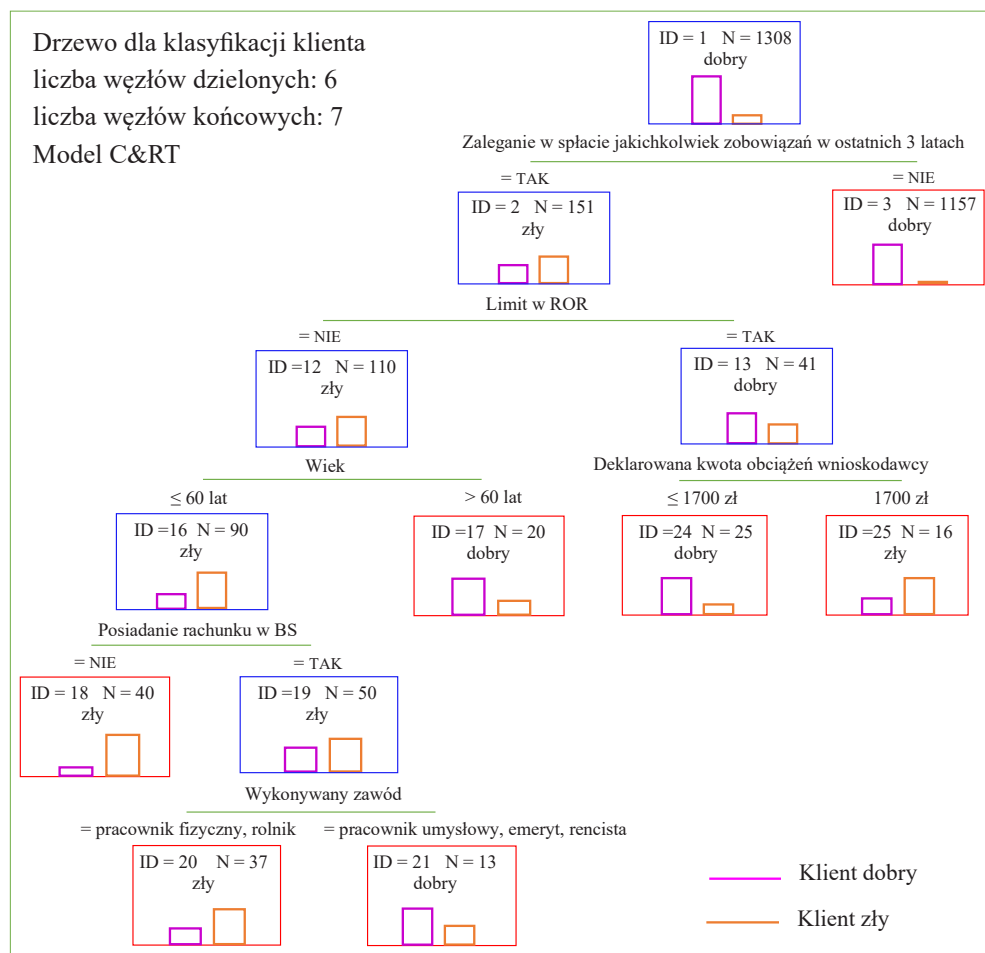
Utworzone drzewo składało się z 6 węzłów dzielonych i 7 końcowych (rysunek 1). Charakteryzowało się ono umiarkowaną złożonością, a poszczególne cechy i kryteria podziału można było uznać za przekonujące i zgodne z teorią. Prezentowany model w 98% (tj. 1144 z 1165) trafnie przewidywał klientów, którzy sumiennie spłacą swoje zobowiązania względem banku. W odniesieniu do klientów „złych” jego trafność była jednak znacząco niższa. Jak wynikało z macierzy klasyfikacji (tabela 2), odsetek poprawnych wskazań klientów, którzy mieli problemy w spłacie kredytu wyniósł 50% (tj. 72 ze 143).

Tabela 2. Macierz klasyfikacji dla modelu C&RT

	Obserwowana	Przewidywana dobry	Przewidywana zły	Łącznie w wierszu
Klasyfikacja	dobry	1144	21	1165
Procent z wiersza		98,20%	1,80%	-
Klasyfikacja	zły	71	72	143
Procent z wiersza		49,65%	50,35%	-

Źródło: opracowanie własne

Kluczowym czynnikiem podziału zbiorowości klientów na „dobrych” i „złych” były zaległości w spłacie zobowiązań w okresie ostatnich 3 lat (rysunek 1). Klientów niemających takich zaległości można było uznać za „dobrych”. Prawdopodobieństwo, że w ich gronie pojawi się kredytobiorca sprawiający problemy było mniejsze od 5%. Z diagramu wynika, że w gronie 151 klientów, którzy w momencie zaciągania kredytu mieli już jakieś zaległości w spłacie zobowiązań w okresie ostatnich 3 lat, większość stanowili klienci „źli”. Jednak ci, którzy posiadali limity w ROR, relatywnie rzadziej stawali się problematycznymi klientami. Na podstawie opracowanego modelu można wnioskować, że istotna w ich przypadku była deklarowana kwota obciążeń. Jeśli przekraczała ona 1700 zł/m-c, prawdopodobieństwo, że klient będzie „zły” znacząco zwiększało się (69%).



Rysunek 1. Diagram drzewa klasyfikacyjnego C&RT

Źródło: opracowanie własne

Jednak w gronie klientów mających limity w ROR i obciążenia do 1700 zł/miesiąc, już tylko średnio co czwarty był klientem „złym” (24%). Wykorzystując drzewo klasyfikacyjne C&RT można zatem zauważyć, że zaległości w spłacie jakichkolwiek zobowiązań w okresie ostatnich 3 lat i brak limitu w ROR, to czynniki wyraźnie zwiększające ryzyko kredytowe.

Jak wynikało z modelu, wśród determinant ryzyka kredytowego dosyć istotny okazał się także wiek. Klienci starsi (w wieku 60+) zdawali się być bardziej odpowiedzialni i rzadziej sprawiali problemy bankom spółdzielczym. Biorąc pod uwagę klientów w wieku poniżej 60 lat, model akcentował, że istotne jest to czy posiadali oni rachunek

w banku spółdzielczym. Jeśli nie – to zdecydowana większość z nich (90%) przysporzyła bankowi problemów. Kolejny, głębszy podział drzewa wskazywał, że ryzyko w odniesieniu do tych kredytobiorców, którzy mieli rachunki w bankach spółdzielczych, różnicował wykonywany zawód. Grupę większego ryzyka tworzyli rolnicy i pracownicy fizyczni – w ich gronie aż 68% stanowili klienci „źli”. W gronie pracowników umysłowych oraz emerytów i rencistów odsetek klientów „złych” był wyraźnie niższy (31%).

Mając na względzie nie tylko stosunkowo dobre dopasowanie modelu mierzone odsetkiem błędnych klasyfikacji, ale też możliwość wprowadzania do modelu zarówno liczb, jak i opisów słownych, a także przejrzystość i łatwość interpretacji diagramu wynikowego, drzewa klasyfikacyjne C&RT można potraktować jako godne uwagi narzędzie w działalności kredytowej banków.

PODSUMOWANIE

Modele C&RT są cenione ze względu na intuicyjność i łatwość interpretacji wyniku. W celu dokonania klasyfikacji klienta i ustalenia poziomu ryzyka kredytowego wystarczyłoby w zasadzie przestudiować diagram drzewa (rysunek 1), zwracając uwagę na cechy stanowiące kryteria podziału. Można więc na takie modele spoglądać także jako na narzędzie doskonalenia arkuszy służących do oceny ryzyka kredytowego.

Model trafnie wskazywał klienta, który wywiąże się z umowy, ale właściwe rozpoznanie „złego” klienta stanowiło już wyraźnie większe wyzwanie. Można stwierdzić, że w praktyce prezentowane drzewo klasyfikacyjne C&RT słabiej zabezpieczyłoby bank przed niespłaconymi kredytami, ale za to doskonale ograniczyłoby ryzyko nieudzielenia kredytu klientowi, który wywiązałby się ze swojego zobowiązania. Warto podkreślić, że także modele prezentowane w badaniach Grzegorza Miguta [2003] częściej myliły się podczas rozpoznawania przypadków, dla których ocena klienta była negatywna. To wskazuje, że problemy ze spłatą zdawały się być determinowane także innymi czynnikami, pozostającymi poza polem statystycznej obserwacji prowadzonej przez banki. Typowe zestawy predyktorów wykorzystywanych w celach scoringowych opisują zazwyczaj szeroko i precyzyjnie cechy socjoekonomiczne, historię klienta i jego relacje z bankiem. Nie biorą one jednak pod uwagę motywów i intencji klienta, jego psychiki, orientacji temporalnej oraz skłonności do prokrastynacji. Ich symptomami mogą być takie zjawiska, jak rosnące obciążenia wnioskodawców, powtarzające się problemy ze spłatą zobowiązań lub prowadzone wobec nich działania windykacyjne. Mimo że takie cechy bierze się już pod uwagę w modelach, to najwyraźniej nie jest to wystarczające.

Poszukując możliwości poprawy trafności modelu klasyfikacyjnego C&RT w rozpoznawaniu klientów „złych”, warto zauważyć, że danych było wystarczająco dużo, ale ich struktura z naturalnych przyczyn nie była optymalna – „zły” klient nie może

pojawiać się nazbyt często. Biorąc pod uwagę, że model bazował na zestawie 27 cech, zabrakło większej liczby przypadków właśnie takiego „złego” klienta. Warto przypomnieć, że dysponowano 143 takimi przykładami. Przeważał klient „dobry” (1165 przypadków) i model dobrze nauczył się rozpoznawać przede wszystkim takiego klienta.

W kontekście obliczeniowym problem został wzmocniony przez podział danych na podzbiory losowe – co jest typowe dla metod Data Mining. V-krotny sprawdzian krzyżowy implikował podziały na próby uczące i testowe w modelu drzewa – a zatem dostępna liczba przykładów w próbach była jeszcze mniejsza, choć nadal każda obserwacja była opisywana kombinacją 27 cech. Pewnym remedium mogłoby tu być ustalenie wyższego kosztu błędnych klasyfikacji dla „złego” klienta.

Warto zwrócić uwagę na to, że drzewa klasyfikacyjne C&RT mogą bazować na wielu predyktorach wyrażanych zarówno liczbami (np. deklarowana kwota obciążeń 1700 zł), jak i opisem słownym (np. wykonywany zawód – pracownik umysłowy). Co więcej, zmienne objaśniające mogą być nie tylko liczne i ze sobą skorelowane, ale nawet niekompletne (co w przypadku innych metod nie byłoby do przyjęcia).

Zaprezentowany model miał swoje ograniczenia, ale wydaje się, że nawet w obecnej postaci byłby przydatny w pracy analityka, ugruntowując jego werbalną wiedzę o źródłach i złożonej specyfice ryzyka kredytowego.

Żaden model nie zwalania z konieczności krytycznego spojrzenia na wynik. Nawet nowoczesne narzędzia Data Mining (tj. drzewa klasyfikacyjne i sieci neuronowe) kierują się jedynie kryteriami statystycznymi, a nie merytorycznymi. Modele drzew klasyfikacyjnych C&RT mogą być w działalności bankowej wartym rozważenia narzędziem. Nie zastąpią one jednak analityka, a co najwyżej poprawią skuteczność jego pracy.

BIBLIOGRAFIA

- Baster Paweł, Katarzyna Pochtowska. 2011. Sieci neuronowe i polichotomiczne modele zmiennych jakościowych w analizie ryzyka kredytowego (Neural networks and models for polychotomous ordered data in credit risk analysis). *Folia Oeconomica Cracoviensia*, LII: 5-17.
- Bereza Stanisław. 1995. *Zarządzanie ryzykiem bankowym* (Banking risk management). Warszawa: Związek Banków Polskich.
- Breiman Leo, Jerome Friedman, Richard A. Olshen, Charles J. Stone. 1998. *Classification and Regression Trees*. Boca Raton: Chapman & Hall/CRC.
- Dacko Mariusz, Jarosław Janus, Jacek Pijanowski, Jarosław Taszakowski, Tomasz Wojewodziec. 2019. *Efektywność ekonomiczna scaleń gruntów w Polsce* (Economic effectiveness of land consolidation in Poland). Kraków: Poligraficzny Zakład Usługowy Drukmar.

- Dacko Mariusz, Wojciech Sroka. 2010. Ocena czynników rozwoju przodujących gospodarstw rolniczych z wykorzystaniem drzew regresyjnych typu C&RT (Evaluation of development factors in the leading agricultural holdings using the method of regression trees C&RT). *Zagadnienia Ekonomiki Rolnej* 2: 100-113.
- Gościński Janusz. 1977. *Zarys teorii sterowania ekonomicznego* (Outline of economic control theory). Warszawa: PWN.
- Łapczyński Mariusz. 2003. *Drzewa klasyfikacyjne w badaniach satysfakcji i lojalności klientów* (Classification trees in customer satisfaction and loyalty research). StatSoft Polska, <http://www.statsoft.pl/czytelnia/marketing/drzewa.pdf>, access: 17.12.2022.
- Łapczyński Mariusz. 2005. Podejście regresyjne w budowie drzew klasyfikacyjnych CART (A regression approach in constructing CART classification trees). *Zeszyty Naukowe Akademii Ekonomicznej w Krakowie* 680: 135-151.
- Manasse Paolo, Nouriel Roubini. 2009. „Rules of thumb” for sovereign debt crises. *Journal of International Economics* 78 (2): 192-205.
- Mierzejewski Piotr. 1998. Wykorzystanie sieci neuronowych do oceny ryzyka kredytowego (Using neural networks to assess credit risk). *Prace Naukowe Akademii Ekonomicznej we Wrocławiu* 797: 165-175.
- Migut Grzegorz. 2003. *Modelowanie ryzyka kredytowego* (Credit risk modeling). Statsoft Polska, http://media.statsoft.nazwa.pl/_old_dnn/downloads/migut.pdf, access: 17.12.2022.
- Misztal Małgorzata. 2000. *Wykorzystanie drzew klasyfikacyjnych do wspomagania procesów podejmowania decyzji* (Using classification trees to support decision-making processes). Statsoft Polska, https://media.statsoft.pl/_old_dnn/downloads/nawykorzystanie5.pdf, access: 17.12.2022.
- Płoszajski Piotr. 2005. *Przerażony kameleon. Eseje o przyszłości zarządzania* (Scared chameleon. Essays on the future of management). Warszawa: Fundacja Rozwoju Edukacji Menadżerskiej SGH.
- Sokołowski Andrzej. 2002. *Metody stosowane w Data Mining* (Methods used in Data Mining). Statsoft Polska, https://media.statsoft.pl/_old_dnn/downloads/metody_stosowane_w_data_mining.pdf, access: 17.12.2022.
- Witkowska Dorota, Mariola Chrzanowska. 2006. Drzewa klasyfikacyjne jako metoda grupowania klientów banku (Classification trees as a method of grouping bank customers). *Prace Naukowe/Akademia Ekonomiczna w Katowicach: Modelowanie Preferencji a Ryzyko* 5: 485-496.
- Wójcicka-Wójtowicz Aleksandra. 2018. Sieci neuronowe a podejście oparte na ratingach wewnętrznych. Studium przypadku (Neural networks vs. internal-rating based approach. A case study). *Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu/Finanse Przedsiębiorstw i Gospodarstw Rolnych* 533: 250-258.

APPLYING THE CLASSIFICATION TREE MODEL IN CREDIT RISK ASSESSMENT

Key words: classification tree, Data Mining, cooperative banks, credit repayment, credit risk

ABSTRACT. The aim of the study was to assess the possibility of using the classification tree model as a tool supporting the bank's operations by reducing credit risk. The question of the possibility of using advanced statistical models as tools to support the activities of banks was presented. A classification tree model was demonstrated to reduce risk in credit decision-making. On the basis of a database of 1,308 successful credit applications (in 2015-2015), described by a set of 27 characteristics of a potential borrower. The model with 93% accuracy indicated statistical regularities allowing to identify *a priori* credit applicants customers who will fulfill the contract and customers who are likely to have problems with credit repayment. The first criterion for dividing the set of customers into "good" and "bad" was default in the last three years. Other important risk factors were having a limit in the ROR, having an account with the lending bank and the amount of debits declared by the applicant. The age of the borrower and his occupation were also important. Against the background of the results obtained, the possibilities offered by the classification tree model were discussed, and attention was paid to its intuitiveness and ease of interpretation. The benefits of using such models in managing the risks associated with the credit activities carried out by cooperative banks were analysed.

AUTHORS

MARIUSZ DACKO, PHD

ORCID: 0000-0001-8424-4720

University of Agriculture in Krakow

Faculty of Agriculture and Economics

Department of Economics and Food Economy

e-mail: m.dacko@ur.krakow.pl

IRENEUSZ KURCZYNA, AUDITOR

Expert of the Rating Agency BS limited liability company

Proposed citation of the article:

Dacko Mariusz, Ireneusz Kurczyna. 2023. Zastosowanie modelu drzewa klasyfikacyjnego w ocenie ryzyka kredytowego. *Annals PAAAE XXV* (1): 39-49.