

ZASTOSOWANIE SZTUCZNYCH SIECI NEURONOWYCH DO OCENY ZDOLNOŚCI KREDYTOWYCH ROLNIKÓW – KLIENTÓW FIRMY LEASINGOWEJ

Marta Kiljańska, Marek Klimkiewicz

Katedra Organizacji i Inżynierii Produkcji, Szkoła Główna Gospodarstwa Wiejskiego w Warszawie

Streszczenie. W artykule przedstawiono badania dotyczące wstępnej oceny zdolności kredytowych klientów firmy leasingowej, która świadczy swoje usługi w sektorze rolniczym. Badania wykonano przy użyciu programu STATISTICA Sieci Neuronowe firmy StatSoft. W wyniku przeprowadzonych badań, utworzono model perceptronu trójwarstwowego, który posiada zdolność zakwalifikowania klientów firmy leasingowej do jednej z trzech grup ryzyka kredytowego: klient wiarygodny, klient mało wiarygodny, klient niewiarygodny.

Słowa kluczowe: sieci neuronowe, perceptron, zdolność kredytowa

Wprowadzenie i cel pracy

W ramach działalności firm udzielających kredyty, jedną z najważniejszych i podstawowych czynności jest ocena zdolności klientów do spłaty zaciągniętego kredytu w terminach i wysokościach określonych w umowie. Takie oceny przeprowadzane są między innymi w bankach i firmach leasingowych.

Celem badań przedstawionych w artykule jest budowa modelu opartego na sztucznych sieciach neuronowych, pozwalającego wstępnie ocenić zdolności kredytowe rolników - klientów firmy leasingowej.

Wybrano tę metodę sztucznej inteligencji, ponieważ sieci neuronowe są stosowane do rozwiązywania szerokiego zakresu problemów decyzyjnych, także z dziedziny finansów, które do niedawna badano tylko za pomocą modeli statystycznych, ekonomicznych i optymalizacyjnych. Sieci neuronowe mają możliwość równoległego przesyłania impulsów, posiadają umiejętność radzenia sobie z informacją niepełną i niespójną, mają zdolność dostosowania się do zmiennych sytuacji bez potrzeby dodatkowego oprogramowania oraz charakteryzują się dużą odpornością na uszkodzenia [Hel i in. 2000]. Na podstawie pokazanych przykładów sieć potrafi uogólniać zdobytą wiedzę (nabywa wiedzę o naturze i właściwościach rozwiązywanego zadania).

Sztuczna sieć neuronowa jest systemem prostych elementów zwanych neuronami, które mają zdolność przetwarzania informacji. Neurony połączone są współczynnikami wagowymi, które określają siłę powiązań i tworzą zbiór parametrów modelu. Poszczególne jednostki sieci grupowane są w tak zwane warstwy [Zieliński 2000]. Sieci neuronowe wykorzystywane są w przypadkach, gdy nie są znane zależności między danymi wejściowymi

i wyjściowymi. W sieciach neuronowych powstaje model na podstawie faktów i danych. Opierając się na samodzielnie stworzonej strukturze danych, sieć realizuje potem wszystkie funkcje związane z eksploatacją utworzonego modelu [StatSoft 2001].

Opis metody badawczej

Do budowy modelu neuronowego zgromadzono 145 przypadków dotyczących rolników prowadzących gospodarstwa rolne lub przedsiębiorstwa rolnicze, którzy ubiegali się o zawarcie umowy leasingu. Każdy z klientów opisany jest za pomocą dwunastu atrybutów q_1 - q_{12} , które będą zmiennymi niezależnymi budowanego modelu neuronowego:

- q_1 - *doświadczenie* - doświadczenie w prowadzeniu gospodarstwa rolnego (ile lat?),
- q_2 - *własne gospodarstwo* - okres prowadzenia własnego gospodarstwa (ile lat?),
- q_3 - *wykształcenie* - wykształcenie właściciela gospodarstwa (jakie?),
- q_4 - *rodzaj produkcji* - rodzaj produkcji (jaki rodzaj?),
- q_5 - *plony* - uzyskiwane plony powyżej średniej (czy są powyżej średniej krajowej?),
- q_6 - *dzierżawa* - dzierżawione grunty (jaki stanowią procent wszystkich gruntów?),
- q_7 - *melioracja* - zmeliorowanie gruntów (czy jest?),
- q_8 - *sprzęt* - sprzęt na umowie leasingowej (kombajn? ciągnik rol.? inny?),
- q_9 - *dealer* - dealer (jak układa się współpraca z dealerem?),
- q_{10} - *opłaty* - sposób spłaty kredytu (jaki okres finansowania i rodzaj rat?),
- q_{11} - *opóźnienia* - ewentualne opóźnienie w spłaceniu rat (ile tygodni?),
- q_{12} - *dodatkowa działalność* - dodatkowa działalność (czy jest?).

Każda z cech opisująca konkretną osobę, jest punktowana według systemu stosowanego w wybranej firmie leasingowej, co pozwala na określenie statusu klienta, który w modelu będzie zmienną wyjściową (zależną) - q_{13} - *status*.

Zmiennym niezależnym można przypisać następujące wartości: $V_{q_1} = \{0_1, 1_5, 5_10, >10\}$, $V_{q_2} = \{0_1, 1_5, 5_10, >10\}$, $V_{q_3} = \{\text{podstawowe, zawodowe, zawodowe rol., średnie, średnie rol., wyższe, wyższe rol.}\}$, $V_{q_4} = \{\text{uboga, przeciętna, urozmaicona, bardzo urozmaicona}\}$, $V_{q_5} = \{\text{tak, nie}\}$, $V_{q_6} = \{0_25, 25_50, 50_75, >75\}$, $V_{q_7} = \{\text{tak, nie}\}$, $V_{q_8} = \{\text{ciągnik, kombajn, pozostałe}\}$, $V_{q_9} = \{\text{bardzo dobra, akceptowalna, umiarkowana}\}$

$V_{q_{10}} = \{3_1, 3_3, 3_12, 3_3_1, 4_1, 4_3, 4_12, 4_3_1, 5_1, 5_3, 5_12, 5_3_1\}$ ¹, $V_{q_{11}} = \{<2, <3, <4, >4, >6\}$, $V_{q_{12}} = \{\text{tak, nie}\}$. Atrybut decyzyjny (zmienna zależna) może przyjmować następujące wartości $V_{q_{13}} = \{\text{wiarygodny, mało wiarygodny, niewiarygodny}\}$.

¹ oznaczenie okresu finansowania:

- 3 lata, raty spłacane co 1 miesiąc - 3_1,
- 3 lata, raty spłacane co 3 miesiące - 3_3,
- 3 lata, raty spłacane co rok - 3_12,
- 3 lata, raty spłacane przez pierwszy rok co 3 miesiące, przez kolejne lata co 1 miesiąc 3_3_1; w ten sam sposób oznaczono okres finansowania dla dłuższych okresów spłat.

Wszystkie atrybuty są zmiennymi jakościowymi. Zbiór wszystkich przypadków podzielono na podzbiory: uczący, walidacyjny i testowy w stosunku 2:1:1.

Wyniki badań

Analiza zmiennych wejściowych przeprowadzona za pomocą narzędzi doboru cech, jakimi dysponuje program STATISTICA Sieci Neuronowe wykazała, że zmienne wejściowe jak: q_2 - okres prowadzenia własnego gospodarstwa, q_3 - wykształcenie właściciela gospodarstwa, q_5 - uzyskiwane plony powyżej średnie, q_7 - zmeliorowanie gruntów i q_{12} - dodatkowa działalność nie wnoszą istotnych informacji do modelu, dlatego odrzucono te zmienne.

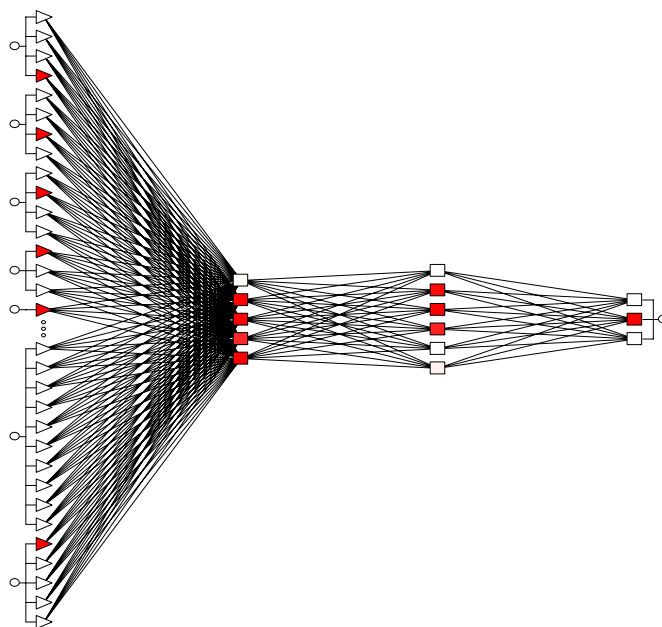
Do budowy modelu wybrano opcję automatycznego określenia złożoności sieci. Zbadano jakość klasyfikacji klientów przez sieci neuronowe różnego rodzaju i o różnych architekturach (sieci liniowe, perceptron wielowarstwowy, sieci o radialnych funkcjach bazowych, probabilistyczne sieci neuronowe). Okazało się, iż najlepsze wyniki osiągane są przez perceptrony trój- i czterowarstwowe, dlatego też ostateczne badania wykonano ograniczając się tylko do perceptronów wielowarstwowych. Spośród analizowanych sieci najlepszą jakością charakteryzował się perceptron trójwarstwowy, którego schemat przedstawiono na rysunku 1. Sieć ta uczona była za pomocą algorytmu wstecznej propagacji błędu przez 50 epok, a następnie, aby uniknąć przeuczenia, algorytmem gradientów sprzężonych przez 10 epok.

Warstwę wejściową sieci tworzy siedem zmiennych: doświadczenie klienta, rodzaj produkcji, ilość dzierzawionych gruntów, rodzaj sprzętu na umowie leasingu, dealer, system opłat i opóźnienia w płatnościach rat. Pierwsza warstwa ukryta ma 5 neuronów, a druga – 6. Trzecią warstwę tworzy warstwa wyjściowa, określająca status klienta firmy.

Wybrana sieć ma efektywność znakomitą, ponieważ współczynnik poprawnych klasyfikacji wynosi 1 zarówno dla zbioru uczącego, jak i dla walidacyjnego. Ponieważ błąd dla zbioru uczącego i walidacyjnego jest mały sieć ma dobre właściwości uogólniające.

Analiza wrażliwości sieci pozwoliła określić użyteczność poszczególnych zmiennych wejściowych. Wskaźnikiem przydatności poszczególnych zmiennych do klasyfikacji wykonywanej przez sieć jest iloraz błędu sieci otrzymanej bez użycia danej zmiennej i błędu tzw. bazowego, to jest błędu sieci otrzymanego z użyciem wszystkich zmiennych, który dla tej sieci wynosi 0,0029.

Analiza wrażliwości przedstawiona jest w tabeli 1. Zmiennym nadano rangi według malejącego ilorazu błędu. W zbiorze uczącym najważniejszymi zmiennymi są: q_1 - doświadczenie w prowadzeniu gospodarstwa rolnego, q_{10} - sposób spłaty kredytu i q_{11} - ewentualne opóźnienie w spłaceniu rat.



Rys. 1. Schemat sieci za pomocą, której uzyskano najlepsze wyniki klasyfikacji

Fig. 1. Diagram of the network, which allowed to obtain best classification results

Tabela 1. Analiza wrażliwości zmiennych modelu

Table 1. Sensitivity analysis for model variables

Zmienne niezależne	q ₁ doświadczenie	q ₄ rodzaj produkcji	q ₆ dzierżawa	q ₈ sprzęt	q ₉ dealer	q ₁₀ opłaty	q ₁₁ opóźnienia
Ranga	1	5	4	7	6	3	2
Błąd	0,303	0,161	0,259	0,009	0,089	0,266	0,278
Iloraz błędu	104,74	55,56	89,48	3,29	30,82	91,84	96,22

Źródło: obliczenia własne autorów na podstawie: STATISTICA Sieci Neuronowe

W tabeli 2 przedstawiono statystyki klasyfikacyjne wyznaczone oddzielnie dla zbioru uczącego, walidacyjnego i testowego. Można zauważyć, że w zbiorze testowym jeden przypadek, określony jako „mało wiarygodny”, został przyporządkowany do klasy „wiarygodny” oraz dwa przypadki zdefiniowane jako „wiarygodny” do klasy „mało wiarygodny”. Można przypuszczać, iż przypadki te znajdowały się na granicy decyzyjności.

Program STATISTICA Sieci Neuronowe umożliwia przedstawienie błędów jednostkowych dla poszczególnych przypadków na wykresie słupkowym, co pozwala na lokalizację przypadków, które są szczególnie trudne do przetworzenia przez sieć. Z wykresu odczytano, że istnieje co najmniej 7 przypadków, które utrudniają prawidłowy proces uczenia

sieci. Ilość tych przypadków jest mała (około 5% zbioru uczącego), więc nie wpływa to w znaczący sposób na prawidłowy przebieg procesu uczenia sieci neuronowej.

Tabela 2. Ocena klasyfikacji wykonanej przez sieć
Table 2. The evaluation of classification performed by the network

Zbiory	Zbiór uczący			Zbiór walidacyjny			Zbiór testowy		
	mało wiarygodny	wiarygodny	niewiarygodny	mało wiarygodny	wiarygodny	niewiarygodny	mało wiarygodny	wiarygodny	niewiarygodny
Razem	25	18	30	11	9	16	11	8	17
Poprawnie	25	18	30	11	9	16	10	6	17
Błędnie	0	0	0	0	0	0	1	2	0
Nieokreślone	0	0	0	0	0	0	0	0	0
Mało wiarygodny	25	0	0	11	0	0	10	2	0
Wiarygodny	0	18	0	0	9	0	1	6	0
Niewiarygodny	0	0	30	0	0	16	0	0	17

Źródło: obliczenia własne autorów na podstawie: STATISTICA Sieci Neuronowe

Podsumowanie

Przeprowadzone badania z wykorzystaniem sieci neuronowej do klasyfikacji zdolności kredytowej klienta wskazują, że nie wszystkie atrybuty opisujące klienta są przydatne do budowy modelu, dlatego można ograniczyć te dane, które wnoszą mało ważnych informacji.

Można też stwierdzić, iż sztuczna sieć neuronowa nie jest doskonałym narzędziem, które może klasyfikować klientów firmy leasingowej bez nadzoru pracowników. Przeprowadzone eksperymenty pokazały, że sieć neuronowa niepoprawnie klasyfikowała pojedyncze przypadki znajdujące się na granicy decyzyjności. Doświadczony analityk jest w stanie rozważyć głębsze aspekty problemu i zaklasyfikować dany przypadek do właściwej grupy kredytowej. Sieć może więc służyć jako narzędzie pomocnicze, a ostateczną decyzję powinien podejmować specjalista.

Bibliografia

- Helt P., Parol M., Piotrowski P. 2000. Metody sztucznej inteligencji w elektroenergetyce. Oficyna Wydawnicza Politechniki Warszawskiej. Warszawa. ISBN 83-7207-194-2.
- Zieliński J.S. (red.) 2000. Inteligentne systemy w zarządzaniu. Teoria i praktyka. Wydawnictwo Naukowe PWN. Warszawa. ISBN 83-01-12968-9.
- StatSoft. 2001. STATISTICA Neural Networks PL. Wprowadzenie do sieci neuronowych. Kraków. ISBN 83-88724-01-0

USING OF ARTIFICIAL NEURAL NETWORKS TO ASSESS CREDIT RATINGS OF FARMERS – CLIENTS OF A LEASING COMPANY

Abstract. The paper presents results of the research on preliminary assessment of credit ratings for clients of a leasing company, which offers its services in agricultural sector. The research was carried out using the STATISTICA Neural Networks application from StatSoft. Completed research allowed to develop a model of a three-layer perceptron able to classify the leasing company clients in one of three credit risk groups: reliable client, hardly reliable client, unreliable client.

Key words: neural networks, perceptron, credit rating

Adres do korespondencji:

Marek Klimkiewicz; e-mail: marek_klimkiewicz@sggw.pl
Katedra Organizacji i Inżynierii Rolnictwa
Szkoła Główna Gospodarstwa Wiejskiego w Warszawie
ul. Nowoursynowska 166
02-787 Warszawa