

Tomasz Ząbkowski

Szkoła Główna Gospodarstwa Wiejskiego w Warszawie

ZASTOSOWANIE SZTUCZNYCH SIECI NEURONOWYCH DO OCENY RYZYKA KREDYTOWEGO KLIENTA W TELEKOMUNIKACJI

1. Wstęp

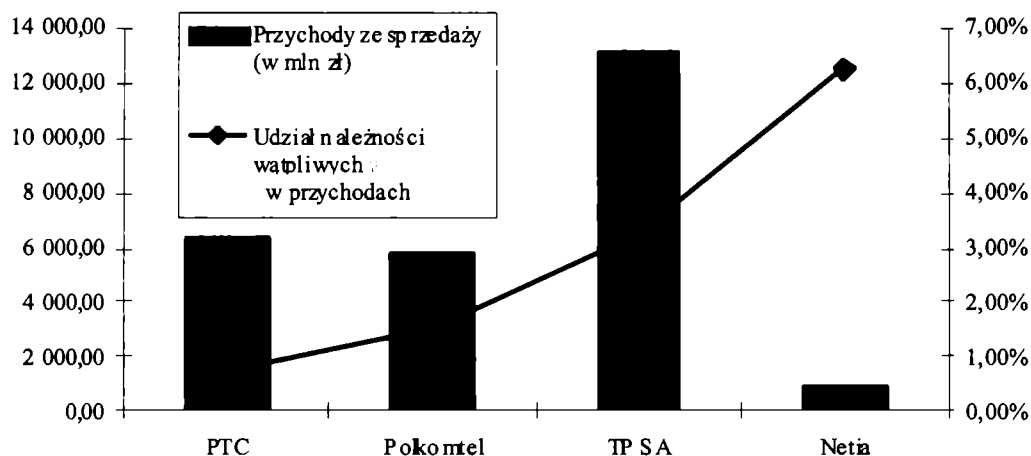
Celem prezentowanych badań jest zastosowanie sztucznych sieci neuronowych do klasyfikacji ryzyka kredytowego związanego z niewypłacalnością klienta w telekomunikacji. Praca obejmuje weryfikację następujących hipotez: (1) sieci neuronowe w powyższym zastosowaniu są narzędziem, które w istotny sposób pozwala na skuteczne ograniczenie ryzyka oraz (2) trafność modeli klasyfikacyjnych w poszczególnych percentylach jest miarą pozwalającą na wykazanie przewagi sieciowych modeli klasyfikacyjnych nad modelami logitowymi we wskazywaniu należności zagrożonych.

2. Ryzyko kredytowe

Termin ryzyka kredytowego pojawia się zwykle w kontekście związanym z działalnością bankową i wskazuje przede wszystkim na niebezpieczeństwo niewypłacalności kredytobiorcy bądź ryzyko bankructwa [Altman i in. 1994; Witkowska, Staniec 2003; Witkowska 1999; 2006]. Jednakże określenie to nie jest zarezerwowane wyłącznie do opisu charakterystyki sektora bankowego, gdyż ryzyko kredytowe istnieje wszędzie tam, gdzie odbiorcy dóbr bądź usług otrzymują towary, a moment płatności za nie zostaje odroczone w czasie. Przewidywanie ryzyka kredytowego klientów związanego z odchodzeniem od operatora, często przy jednoczesnym nieopłaceniu należności za wykorzystane usługi, jest jednym z najpoważniejszych wyzwań stojących przed operatorami telekomunikacyjnymi. W obliczu skali zjawiska związanego z niesolidnymi klientami oraz dużej liczby klientów obsługiwanych przez operatorów podstawą do oceny ryzyka nie mogą być stosowane metody traktujące każdy przypadek indywidualnie. W związku z tym do modelowania tego typu problemów wykorzystuje się metody pozwalające do-

konać odwzorowania bardzo złożonych funkcji. Do grona takich technik można zaliczyć sztuczne sieci neuronowe, będące ogólną nazwą struktur matematycznych i ich modeli, realizujących obliczenia poprzez szereg elementów wykonujących pewne operacje na swoim wejściu, zwanych neuronami.

Mimo coraz doskonalszych technik służących do rozpoznawania ryzyka i zarządzania nim, problem pozostaje nadal aktualny i jednocześnie przysparza największych kłopotów, czego przykładem może być rosnący udział należności wątpliwych w oficjalnie raportowanych wynikach firm telekomunikacyjnych. Skalę tego niekorzystnego zjawiska za rok 2004 dla wybranych firm w branży telekomunikacyjnej przedstawia rys. 1.



Rys. 1. Przychody ze sprzedaży a udział należności wątpliwych

Źródło: opracowanie własne.

Mniejszy udział należności wątpliwych w przychodach ze sprzedaży obserwujemy przede wszystkim wśród operatorów telefonii komórkowej, gdzie stale zwiększający się udział użytkowników systemu przedpłat (*pre-paid*) w ogólnej liczbie klientów pozwala na ograniczenie ryzyka kredytowego. W przypadku telefonii stacjonarnych odsetek należności wątpliwych sięga nawet kilku procent rocznych przychodów. Duże różnice przedstawione na wykresie mają także związek z przyjętymi metodami szacowania odpisów. Oszacowanie odpisów na należności wątpliwe następuje, gdy uzyskanie pełnej kwoty należności przestaje być prawdopodobne na podstawie analizy danych historycznych ściągłości należności od poszczególnych dłużników. Innym sposobem szacowania odpisów (Netia) mogą być wytyczne związane bezpośrednio z prowadzoną działalnością gospodarczą, gdzie należności niespłacone w przyjętym okresie rozliczeniowym (np. po 30 dniach) są automatycznie klasyfikowane jako zagrożone.

3. Opis danych do analizy

Dane wykorzystane w analizie stanowiły losową próbkę 51 733 klientów operatora telekomunikacyjnego. Dla każdego klienta zmienna o nazwie BP jest wykorzystywana do stwierdzenia rozwiązania umowy z powodu braku płatności za dostarczone przez operatora usługi. Zmienna ta ma postać zmiennej zero-jedynkowej i przyjmuje wartość 1, jeśli nastąpił brak płatności, w wyniku czego umowa została rozwiązana, oraz 0 w przypadku przeciwnym. W badanej próbie odsetek klientów, u których stwierdzono brak płatności skutkujący rozwiązaniem umowy wynosił 9,33%.

Jako dane charakteryzujące klienta posłużyły zmienne transakcyjne oraz demograficzne. Łączna liczba zmiennych objaśniających to 58. W celu uproszczenia eksperymentu praktycznego zdecydowano się na zastosowanie procedury wstępnej eliminacji cech. W związku z tym z puli dostępnych zmiennych zostały usunięte te zmienne, które nie objaśniały w istotnym stopniu (mierzonej za pomocą współczynnika R^2) zmienności zmiennej BP. Ponadto, jeśli współczynnik korelacji między dwiema cechami był większy niż 0,7, w dalszych badaniach uwzględniano tę cechę, dla której współczynnik R^2 z ze zmienną BP był większy. Wybrane w ten sposób zmienne to: X1 (kwota abonamentu), X2 (informacja o liczbie nałożonych blokad na połączenia głosowe z powodu nieopłacenia należności), X3 (wyznacznik mobilności użytkownika), X4 (wartość kwot po terminie wymagalności), X5 (informacja o liczbie wysłanych pism ponagających do zapłaty), X6 (liczba różnych modeli aparatów telefonicznych, z których korzystał użytkownik), X7 (kwota wpłat użytkownika z tytułu usług świadczonych przez operatora).

Statystyki opisowe dla zmiennych przedstawione w tab. 1. Wykazały m.in. braki danych (X4 i X7), ujemne wartości oznaczające nadpłatę na koncie użytkownika (X4). Zwraca uwagę fakt, że zmienne charakteryzują się dość dużą wartością współczynnika zmienności. Z danych kolumny zawierającej wartości maksymalne wynika, że najbardziej aktywni użytkownicy potrafią korzystać w ciągu badanego okresu z 70 różnych modeli aparatów telefonicznych, przemieszczać się bardzo intensywnie, a także posiadać zadłużenie wielkości 4965 zł.

Tabela 1. Statystyki opisowe zmiennych

| Zmienna | Średnia | Odch. std | Min | Max | Mediana | Brak danych |
|---------|---------|-----------|-------|---------|---------|-------------|
| X1 | 79,73 | 96,10 | 4,81 | 2102,65 | 156,46 | 0 |
| X2 | 0,30 | 0,78 | 0 | 7 | 0 | 0 |
| X3 | 12,20 | 10,74 | 0 | 265 | 9 | 0 |
| X4 | 79,56 | 124,01 | -1,68 | 4965,89 | 47,25 | 23 |
| X5 | 0,12 | 0,40 | 0 | 8 | 0 | 0 |
| X6 | 1,72 | 1,71 | 1 | 70 | 1 | 0 |
| X7 | 385,43 | 226,21 | 0 | 5587,63 | 327,38 | 21 |

Źródło: opracowanie własne.

4. Zastosowane techniki

Przedstawiona kwestia oceny ryzyka kredytowego użytkownika telefonii komórkowej sprowadza się do rozwiązania problemu klasyfikacyjnego, którego celem jest przypisanie obiektom przynależności do danej klasy(grupy). W związku, z tym, że znane są wartości klas, z których pochodzą obiekty, rozważany będzie przypadek tzw. uczenia z nauczycielem, zwany inaczej klasyfikacją wzorcową.

Do tworzenia modeli klasyfikacyjnych wykorzystano sztuczne sieci neuronowe, które dzięki takim cechom, jak zdolność do uczenia, umiejętność generalizacji oraz odporność na uszkodzenia są atrakcyjną techniką do tego typu zastosowań. Przykłady zastosowania sieci neuronowych do zagadnień klasyfikacyjnych można znaleźć w pracach [Altman i in. 1994; Witkowska 1999; Yang i in. 1999]. Z kolei prace [Canalli 2001; Daskalaki i in. 2003; Estevez i in. 2006; Ezawa, Norton 1996] przedstawiają zastosowania sieci do problemów rynku telekomunikacyjnego.

Podstawowym elementem każdej sieci neuronowej są neurony, czyli elementy przetwarzające sygnały. Neurony są zwykle łączone ze sobą, tworząc bardziej skomplikowane struktury zwane sieciami neuronowymi. Każdy neuron ma wiele wejść x_i , $i = 1, 2, \dots, n$ oraz jedno wyjście y , co można zapisać jako:

$$y = f(e) = f\left(\sum_{i=0}^n w_i x_i\right), \quad (1)$$

gdzie: x_i – wartości wejściowe neuronu dla $i = 1, 2, \dots, n$,
 w_i – wagi, czyli parametry modelu do oszacowania w trakcie uczenia,
 $f(e)$ – funkcja aktywacji neuronu,
 y – wyjście neuronu.

Funkcja aktywacji może mieć różne postaci, które determinują sposób przetwarzania sygnałów przez neuron. Zwykle pakiety statystyczne oferujące sieci neuronowe posiadają najbardziej popularne funkcje w postaci: liniowej, logistycznej (2) czy tangensa hiperbolicznego (3).

$$f(e) = \frac{1}{1 + \exp(-\beta e)}, \quad (2)$$

$$f(e) = \tanh(e) = \frac{1 - \exp(-\beta e)}{1 + \exp(-\beta e)}. \quad (3)$$

Neurony wchodzące w skład sieci tworzą zazwyczaj warstwy, z których pierwsza nosi nazwę warstwy wejściowej, ostatnia warstwy wyjściowej, wszystkie zaś warstwy znajdujące się pomiędzy nimi określane są jako warstwy ukryte. W procesie tzw. uczenia sieci istotną złożoność polega na określeniu prawidłowych wartości parametrów sieci (wag, struktury sieci). W trakcie uczenia na wejścia sieci wprowadzane są zgromadzone wartości i obserwowane jest towarzyszące tej ope-

racji zachowanie sieci. Jeżeli zachowanie sieci nie prowadzi do poprawy jej funkcjonowania, to przeprowadzana jest modyfikacja parametrów. Sposób przeprowadzenia modyfikacji określany jest przez zastosowany algorytm uczenia. Najczęściej wykorzystywane do tego celu algorytmy to: gradienty sprzężone, wsteczna propagacja, quasi-Newton, Levenberga-Marquardta [Rutkowski 2006]. Jakość wytrenowania sieci można oceniać za pomocą mierników dokładności prognoz. Gdy zbudowany model ma służyć do przewidywania przynależności do danej klasy, wówczas o jakości modelu decyduje skuteczność klasyfikacji na jego podstawie. Najczęściej wykorzystuje się tutaj miarę mówiącą o tym, jaki jest procent poprawnych klasyfikacji (PPK):

$$PPK = \frac{n_{00} + n_{11}}{n} \times 100, \quad (4)$$

gdzie: n – liczba obserwacji,

n_{00} – liczba obserwacji, dla których $\hat{y}_i = y_i = 0$,

n_{11} – liczba obserwacji, dla których $\hat{y}_i = y_i = 1$.

Miara ta w literaturze anglojęzycznej jest określana jako PCC (*percentage correctly classified*) i jest stosowana do określania zdolności klasyfikacyjnych modeli. Często jednak miara ta nie jest wystarczająca do stwierdzenia skuteczności modelu.

W zastosowaniach biznesowych, gdzie model ma być stosowany na całej populacji, istotną miarą mówiącą o jakości modelu jest trafność w zależności od wielkości bazy, do której możemy dotrzeć. Zwykle nie ma możliwości (np. z powodu kosztów), aby dotrzeć do całej populacji i tym samym ograniczyć ryzyko związane z niesolidnymi klientami, którzy nie uiszczą należności za dostarczone usługi. W związku z tym, że modele klasyfikacyjne grupują przypadki w zależności od prawdopodobieństwa przynależności do danej klasy, informacja ta może posłużyć do uporządkowania tychże przypadków według malejącego ryzyka zajścia przewidywanego zdarzenia. W ten sposób można dokonać wyliczenia trafności modelu w poszczególnych percentylach.

Ponadto w badaniach został wyznaczony model logitowy, który posłużył jako punkt odniesienia dla trafności klasyfikacji modeli sieci neuronowych. W przypadku modelu logitowego przyjmuje się, że funkcja F dla zmiennych X_1, \dots, X_n , jest dystrybuantą rozkładu logistycznego:

$$F(\beta_0 + \beta_1 x_{1i} + \dots + \beta_k x_{ki} + \varepsilon_i) = \frac{\exp(\beta_0 + \beta_1 x_{1i} + \dots + \beta_k x_{ki} + \varepsilon_i)}{1 + \exp(\beta_0 + \beta_1 x_{1i} + \dots + \beta_k x_{ki} + \varepsilon_i)}, \quad (5)$$

gdzie wartość funkcji odwrotnej do F jest tzw. logitem $\ln \frac{p}{1-p}$. Przekształcenie

to sprowadza zależność prawdopodobieństwa p od zmiennych niezależnych do postaci liniowej. Logit jest logarytmem ilorazu szans przyjęcia oraz nieprzyjęcia

wartości 1 przez zmienną Y . Jeżeli szanse są jednakowe, równe 0,5, to logit wynosi 0. W przypadku gdy $p > 0,5$ logit jest dodatni, gdy $p < 0,5$, to jest ujemny.

5. Eksperyment praktyczny

Przed przystąpieniem do budowy modeli dostępny zbiór danych został w sposób losowy podzielony na trzy części w następujących proporcjach: 40% obserwacji trafiło do zbioru uczącego, 30% do zbioru walidacyjnego, a pozostałe 30% do zbioru testowego, służącego do przeprowadzenia ostatecznej oceny.

Eksperyment został zapoczątkowany zbudowaniem modelu logitowego (LOGIT) ze wszystkimi dostępnymi zmiennymi, służącego jako punkt odniesienia dla trafności klasyfikacji modeli sieci neuronowych (tab. 2).

Tabela 2. Macierz klasyfikacji dla modelu logitowego

| Klasyfikacja obserwowana | Klasyfikacja uzyskana modelem LOGIT | | | PPK |
|--------------------------|-------------------------------------|-----|--------|-------|
| | | 1 | 0 | |
| 1 | | 293 | 1 146 | 20,4% |
| 0 | | 251 | 13 732 | 98,2% |
| Razem | | 544 | 14 878 | 90,9% |

Źródło: opracowanie własne.

W przypadku sieci neuronowych badanie zostało przeprowadzone w ten sposób, że dla każdej sieci typu perceptron wielowarstwowy były tworzone układy z różną liczbą neuronów w warstwie ukrytej. Eksperyment zapoczątkowało badanie sieci z 2 neuronami w warstwie ukrytej (struktura MLP 7:2:1), a następnie w każdym kolejnym modelu dodawano dodatkowe neurony. Proces został zakończony w momencie, gdy każde następne dodanie neuronów nie poprawiało wyników osiąganych przez ostatnią sieć bądź też pojawiał się efekt przeuczenia, czyli nadmiernego dopasowania sieci do danych. We wszystkich analizowanych przypadkach funkcją aktywacji neuronu był tangens hiperboliczny. Z kolei do uczenia wszystkich struktur była wykorzystywana metoda Quasi-Newtona, która polega na modyfikacji wag w modelu jednorazowo, w końcowej fazie każdej iteracji. W ten sposób w algorytmie tym wyznaczany jest średni gradient dla powierzchni błędu, który determinuje sposób modyfikacji wag.

Zestawienie wyników klasyfikacji dla różnych struktur sieci prezentują tab. 3-4.

Tabela 3. Macierz klasyfikacji dla modelu MLP 7:2:1

| Klasyfikacja obserwowana | Klasyfikacja uzyskana modelem MLP 7:2:1 | | | PPK |
|--------------------------|---|-----|--------|-------|
| | | 1 | 0 | |
| 1 | | 306 | 1 133 | 21,3% |
| 0 | | 229 | 13 754 | 98,4% |
| Razem | | 535 | 14 887 | 91,2% |

Źródło: opracowanie własne.

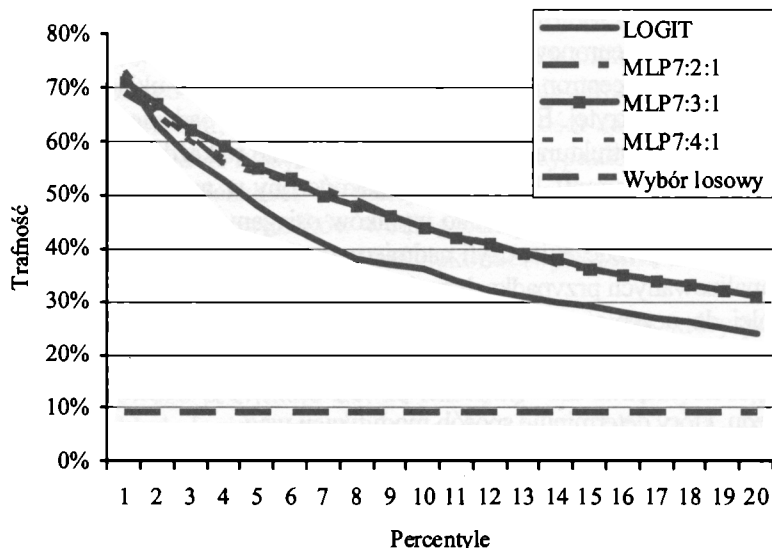
Tabela 4. Macierz klasyfikacji dla modelu MLP 7:3:1

| Klasyfikacja obserwowana | Klasyfikacja uzyskana modelem MLP 7:3:1 | | | PPK |
|--------------------------|---|--------|--|--------------|
| | 1 | 0 | | |
| 1 | 312 | 1 127 | | 21,7% |
| 0 | 199 | 13 784 | | 98,6% |
| Razem | 511 | 14 911 | | 91,4% |

Źródło: opracowanie własne.

Proces tworzenia modeli neuronowych został zakończony wraz z utworzeniem modelu MLP 7:3:1, gdyż każde następne dodanie neuronów (MLP 7:4:1) nie poprawiało ogólnego wyniku klasyfikacji osiągniętego przez ostatnią sieć. Stwierdzono także, że zbudowane modele neuronowe osiągały wyniki bardzo zbliżone do modelu logitowego, gdzie procent poprawnych klasyfikacji wynosił odpowiednio 91,2-91,4% (sieci) oraz 90,9% (LOGIT). Zwraca tu uwagę także to, że zbudowane modele znacznie trafniej przewidują zajście zdarzenia 0, czyli przeciwnego do rozwiązania umowy z powodu braku płatności.

O ile w przypadku miary PPK model logitowy nie odbiegał znacznie od modeli neuronowych, o tyle w przypadku trafności modeli w poszczególnych percentylach sieci neuronowe znacznie trafniej wskazują sytuacje, z którymi jest związane ryzyko kredytowe (rys. 2).



Rys. 2. Trafność modeli w percentylach

Źródło: opracowanie własne.

Dysponując nakładami pozwalającymi na monitoring 5% populacji użytkowników telefonii i posługując się modelem logitowym do wskazywania przypadków

ze statusem $BP = 1$, osiągniemy trafność na poziomie 48%. W przypadku sieci trafność klasyfikacji zwiększa się do 55%. Jeszcze większe różnice można zaobserwować, przesuując się na osi percentyli do poziomu 8-9% – różnica w zastosowaniu modelu logitowego wobec modelu neuronowego sięga 11 punktów procentowych (odpowiednio 38% wobec 49%).

6. Podsumowanie

Otrzymane w wyniku eksperymentu rezultaty pozwalają na sformułowanie następujących wniosków:

1. Sztuczne sieci neuronowe zastosowane do klasyfikacji ryzyka kredytowego klienta telefonii komórkowej wydają się być skutecznym narzędziem, które w istotny sposób pozwala na skuteczniejsze zabezpieczenie przychodów.

2. Porównując modele neuronowe z modelem logitowym, można otrzymać wynik klasyfikacji na zbliżonym poziomie. Procent poprawnych klasyfikacji wynosił odpowiednio 91,2-91,4% dla sieci oraz 90,9% dla modelu logitowego.

3. W przypadku trafności modeli, w poszczególnych percentylach, poszeregowanych według malejącego prawdopodobieństwa przyjęcia wartości 1 przez zmienną oznaczającą brak płatności (BP), sieciowe modele klasyfikacyjne wykazują przewagę nad modelem logitowym. W tym wypadku różnica w zastosowaniu modelu logitowego wobec modelu neuronowego może sięgać nawet 11 punktów procentowych (odpowiednio 38% wobec 49%).

4. Niewykluczone, że na poprawę trafności modeli neuronowych wpłynęłoby uwzględnienie dodatkowych cech, które na tym etapie nie zostały włączone do badań, bądź też podział zbiorowości na podgrupy (np. według lokalizacji, wielkości przychodów od użytkownika) i opracowanie dla nich oddzielnych modeli.

Literatura

- Altman E., Giancarlo M., Varetto F. (1994), *Corporate Distress Diagnostic: Comparison Using Linear Analysis and Neural Networks*, „Journal of Banking and Finance”, 18.
- Canalli E. (2001), *Experimenting Neural Networks to Forecast Business Insolvency*, „Neural Network World”, 11 (4), s. 349-361.
- Daskalaki S., Kopanas I., Goudara M., Avouris N. (2003), *Data Mining for Decision Support on Customer Insolvency in Telecommunications Business*, „European Journal of Operational Research”, 145, s. 239-255.
- Estevez P., Held C., Perez C. (2006), *Subscription Fraud Prevention in Telecommunications Using Fuzzy Rules and Neural Nets*, „Expert Systems with Applications”, 31, s. 337-344.
- Ezawa K.J., Norton S.W. (1996), *Constructing Bayesian Networks to Predict Uncollectable Telecommunication Accounts*, „Expert Systems with Applications”, 11, s. 45-51.
- Rutkowski L. (2006), *Metody i techniki sztucznej inteligencji*, Wydawnictwo Naukowe PWN, Warszawa.

- Yang Z.R., Platt M.B., Platt H.D. (1999), *Probabilistic Neural Networks in Bankruptcy Prediction*, „Journal of Business Research”, 44, s. 67-74.
- Witkowska D. (1999), *Application of Artificial Neural Networks to Bank Decision Simulations*, „International Advances in Economic Research”, 5, s. 350-368.
- Witkowska D. (2006), *Discrete Choice Model Application to the Credit Risk Evaluation*, „International Advances in Economic Research”, 12, s. 33-42.
- Witkowska D., Staniec I. (2003), *Credit Granting procedure: Multilayer Perceptron and Classification Tree*, [w:] *Neural Networks and Soft Computing*, red. L. Rutkowski, J. Kacprzyk, New York, Heidelberg, Physica-Verlag, s. 748-753.

APPLICATION OF ARTIFICIAL NEURAL NETWORKS FOR CUSTOMER CREDIT RISK CLASSIFICATION IN TELECOMMUNICATION

Summary

The article presents an application of artificial neural networks for customer credit risk classification in telecommunication. Based on characteristics of a customer, models that estimate customers' insolvency are proposed. The results confirm the usefulness and high performance of the method for more effective revenue assurance.