

HYBRYDOWY MODEL SYSTEMU EKSPERTOWEGO DO OCENY PODATNIKÓW

Ryszard Budziński, Leszek Misztal

Katedra Inżynierii Systemów Informatycznych
Zachodniopomorski Uniwersytet Technologiczny w Szczecinie
e-mails: R.Budzinski@man.szczecin.pl; lmisztal@wi.zut.edu.pl

Streszczenie: Proponowany model identyfikuje podatników na podstawie ich cech i właściwości, które wskazują na większą możliwość występowania problemów z przestrzeganiem prawa podatkowego. Eliminuje słabości występujące w znanych algorytmach zaliczających się do klasyfikatorów, jak również systemów wnioskujących oraz wspomagających typowanie stosowanych w administracji podatkowej. Jest to możliwe dzięki utworzeniu hybrydowego modelu, który dobrze odzwierciedla zachowania podatników. Model dzięki zastosowaniu nowoczesnych rozwiązań predysponuje do przyszłego utworzenia i uruchomienia w administracji bazującego na nim systemu ekspertowego.

Słowa kluczowe: klasyfikacja podatników, teoria zbiorów przybliżonych, rozmyta analiza skupień, metoda AHP, eksploracja danych

WPROWADZENIE

Techniki eksploracji danych oraz cały proces odkrywania wiedzy mają zastosowanie w wielu dziedzinach i branżach takich jak medycyna, biologia, biotechnologia, produkcja, telekomunikacja, ekonomia, finanse, handel oraz wielu innych sferach życia [Triantaphyllou 2006]. Zgodnie z pozycją [Maulik 2005] odmienna specyfika problemów występujących w różnych branżach powoduje, że proponowane modele rozwiązań są mocno dopasowane do konkretnego zastosowania.

Dlatego też w niniejszym artykule zostało zaproponowane dedykowane rozwiązanie dla dziedziny podatkowej. Jest to model klasyfikacji regulowej podatników w urzędach skarbowych, którzy potencjalnie mogą posiadać problemy podatkowe. Zaproponowany model odpowiada specyfice występującej w obszarze

podatków oraz umożliwia zwiększenie skuteczności kontroli podatkowych. Zadanie jest istotne z punktu widzenia uzyskiwanych dochodów do budżetu, ponieważ odpowiednia kwota pieniędzy, która zasila państwowe konta umożliwia wykonanie zadań zgodnie z założeniami rządu. Jednym z ważnych filarów mających wpływ na realizację wpływów budżetowych jest skuteczna kontrola podatkowa, zgodna z założeniami dyscypliny podatkowej [MF 2010], której celem jest osiągnięcie wysokiego poziomu dobrowolności i jakości wypełniania obowiązków podatkowych poprzez wykorzystanie działań urzędów skarbowych w zakresie kontroli podatkowej.

Obecnie występują w praktyce systemy identyfikacji podatników stosowane w administracji podatkowej, które za pomocą własnych modeli rozwiązują postawiony problem. Zaliczają się do nich systemy KONTROLA i ISKOS. Tworzą modele bazujące na regułach zbudowanych w oparciu o kryteria związane z atrybutami opisującymi dane podatników [ZPK 2009][Zabicka 2003][Goch 2006][Rogacka 2009], model zbudowany na podstawie podziału podatników oraz klasyfikacji za pomocą drzew decyzyjnych [Com.2004][SPSS 2004]. Posiadają wiele słabości związanych z wpływem subiektywnych ocen ludzkich, ograniczonym zakresem informacyjnym oraz rodzajem podatników objętych modelem, jak również częściowym lub całkowitym brakiem weryfikacji dokładności identyfikacji. Do rozwiązania problemu badawczego nadają się również metody z zakresu technik nadzorowanych zaliczających się do klasyfikatorów. Należą do nich maszyna wektorów wspierających [Bezdek 1999][Budziński 2009], zbiory przybliżone [Pawlak 1991][Piegat 2006][Duntsch 2000], drzewa decyzyjne [Miyamoto 2008][Han 2006] oraz naiwny algorytm Bayes'a [Hand 2006] [Witten 2005]. Posiadają liczne słabości, które dotyczą przede wszystkim niewystarczającej dokładności uzyskiwanych wyników.

Dlatego też w celu eliminacji występujących problemów i słabości został zaprojektowany i przedstawiony hybrydowy model systemu ekspertowego oparty na rozmytym algorytmie klastrującym, teorii zbiorów przybliżonych i metodzie AHP do klasyfikacji regułowej podatników. Umożliwi on identyfikację podatników, czyli wyznaczenie podatników posiadających i nieposiadających problemy podatkowe z większą dokładnością od opisanych i występujących w praktyce metod identyfikacji.

OPIS ISTNIEJĄCYCH ROZWIĄZAŃ

Szczególnie ważne dla prowadzonych badań w zakresie identyfikacji podatników są metody klasyfikujące zaliczane do metod nadzorowanych, ponieważ umożliwiają rozwiązania polegające na zakwalifikowaniu podatników na mogących lub niemogących posiadać problemy podatkowe. Do tego typu metod zaliczają się algorytmy takie jak: maszyna wektorów wspierających (ang. SVM) [Schoelkopf 2002], zbiory przybliżone [Bazan 2000][Drzymala 2006], drzewa decyzyjne [Berry 2008], naiwny algorytm Bayes'a [Jain 1988]. Obecnie istnieją

również metody identyfikacyjne podatników wykorzystywane w praktyce jako rozwiązania informatyczne w administracji. Umożliwiają selekcję i wybór podmiotów mogących posiadać określone problemy. Należą do nich systemy KONTROLA, ISKOS. Są szczególnie istotne w kontekście poprawy dokładności uzyskanych rezultatów przez zaproponowaną metodę hybrydową. Poniżej znajduje się krótki opis wymienionych systemów.

W systemie KONTROLA budowa modelu polega na wyborze kryteriów oraz tworzeniu selekcji na podstawie subiektywnej oceny osoby merytorycznej. Utworzone w ten sposób reguły nie przedstawiają obiektywnej oceny, ponieważ model jest w dużym stopniu uzależniony od preferencji użytkownika. Sam system nie posiada również oceny dokładności typowań. Kolejnym systemem identyfikacyjnym jest ISKOS, który tworzy model na podstawie zwykłego podziału podatników oraz klasyfikacji z wykorzystaniem drzew decyzyjnych. Zwykły podział nie umożliwia dokładnego wyodrębnienia podobnych do siebie grup, co może negatywnie wpływać na końcową dokładność wyników. Przeprowadzone badania [Budziński 2009][Becker 2010][Misztal 2009] wykazały również, że klasyfikacja z wykorzystaniem drzew decyzyjnych w dziedzinie podatkowej wykazuje się mniejszą dokładnością w porównaniu z innymi metodami na przykład teorią zbiorów przybliżonych.

PRZEDSTAWIENIE DOSTĘPNYCH DANYCH

Dane użyte w zaproponowanym modelu pochodzą z systemów administracji podatkowej. Zawierają zakres informacyjny dotyczący składanych deklaracji podatkowych, zarówno przez przedsiębiorstwa jak i podatników indywidualnych, dane rejestracyjne dotyczące na przykład adresu czy też formy prawnej, informacje związane z przeprowadzonymi postępowaniami kontrolnymi, egzekucyjnymi, mandatowymi oraz istniejącymi zaległościami podatkowymi. W tablicy numer 1 został przedstawiony zakres informacyjny uwzględniony w modelu.

Na potrzeby modelu została utworzona tabela z 127 atrybutami, które opisują podatnika oraz zawierają wyniki kontroli podatkowej. Dane dotyczące podatników pochodzą z dwudziestu aplikacji podatkowych urzędów administracji podatkowej województwa zachodniopomorskiego. Dotyczą ponad dwudziestu tysięcy kontroli podatkowych oraz zawierają informacje o deklaracjach i pozostałym zakresie dla okresu pięciu lat począwszy od 1 stycznia 2005 roku. Ilość zebranych deklaracji przekracza jeden milion sto tysięcy, natomiast ilość najczęściej występującej deklaracji VAT-7 oscyluje w pobliżu dziewięćset tysięcy.

Tabela 1. Zawartość informacyjna uwzględniona w modelu systemu

Nazwa tablicy	Opis
DANEGL	Podstawowe dane dotyczące podatnika
PIT	Informacje pochodzące z deklaracji PIT-36, PIT-37, PIT-38, PIT/D, PIT/O
PPE	Informacje pochodzące z deklaracji PIT-28, PIT/O związane z ryczałtem ewidencjonowanym
PPL	Informacje pochodzące z deklaracji PIT-36L, podatek dochodowy liniowy
CIT	Informacje pochodzące z deklaracji CIT-2, CIT-8, podatek dochodowy od osób prawnych
PCC	Informacje pochodzące z deklaracji PCC-3, PCC-1, podatek od czynności cywilno prawnych
VZM	Informacje pochodzące z deklaracji VZM-1, odliczenia z tytułu niektórych wydatków budowlanych
VAT	Informacje pochodzące z deklaracji VAT-7, VAT-7K, VAT-7D
PIT4R	Informacje pochodzące z deklaracji PIT-4R, dotycząca pobranych zaliczek na podatek dochodowy (ilość zatrudnionych osób)
POZOSTALE	Informacje pochodzące z pozostałych zakresów związanych z egzekucją administracyjną, mandatami, oraz zaległościami podatkowymi

Źródło: opracowanie własne

KONCEPCJA HYBRYDOWEGO MODELU SYSTEMU

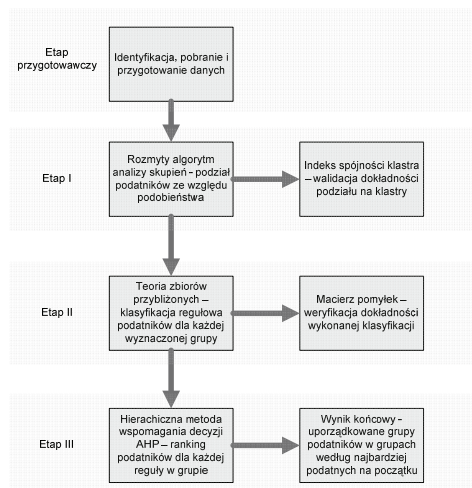
Hybrydowy model systemu ekspertowego ma na celu eliminację słabości związanych z ograniczeniem zakresu podatników objętych modelem jak ma to miejsce w istniejących systemach, eliminację wpływu subiektywnych ocen jak w przypadku KONTROLA, wykorzystanie w modelu większego zakresu informacyjnego dostępnego w aplikacjach, zastosowanie metod umożliwiających weryfikację dokładności, co dotyczy wszystkich wymienionych systemów włącznie z ISKOS. Eliminacja tych słabości umożliwia utworzenie klasyfikacji regułowej i rankingu z większą dokładnością od wszystkich wymienionych i funkcjonujących w praktyce systemów jak również przedstawionych technik klasyfikacyjnych. Proponowany model ma być również zgodny z przyjętymi założeniami CRISP-DM [CRISP-DM 2005], który dzieli proces odkrywania wiedzy na sześć etapów. Model składa się z trzech głównych etapów, w której wynik zakończenia poprzedniego kroku jest wejściem do kolejnego. Natomiast końcowym rezultatem po zakończeniu ostatniego etapu jest ranking podatników ze względu na podatność na problemy z prawem podatkowym. Ogólna postać całościowej metody została przedstawiona na rysunku numer 1. Szczegółowy opis kroków wykonywanych w poszczególnych etapach został ujęty w kolejnych akapitach.

W pierwszym etapie wyznaczenie grup podatników o podobnych cechach zostało podzielone na dwa główne kroki. W pierwszej części następuje pobranie, agregacja oraz kodowanie zidentyfikowanych danych opisujących cechy grupowe podatników, które wskazują na podobieństwa w zachowaniach podatkowych. W drugim etapie wykonane zostają eksperymenty z użyciem wspomnianego

rozmytego algorytmu c-modów. Wyznaczona zostaje ilość klastrów na podstawie metody opartej na klastrowaniu bazującym na entropii [Yao 2000]. Walidacja modelu zostaje wykonana na podstawie indeksu spójności [Xie 1991]. Sprawdzanie spójności zostaje powtórzone dla wielu wartości współczynnika m oraz różnych ilości wyznaczonych klastrów. Ostateczne wyniki przypisania przynależności każdego rekordu danych do określonego klastra zostają przeprowadzone dla ilości klastrów, które posiadają największą spójność. Na rysunku 2 została przedstawiona cała procedura badawcza dla tego kroku.

W drugim etapie przedstawionej metody zostanie przeprowadzona klasyfikacja regułowa podatników. Wykonywany eksperyment będzie polegał na ekstrakcji wiedzy zawartej w zgromadzonych atrybutach zależnych do postaci reguł decyzyjnych na podstawie wyników przeprowadzonych postępowań kontrolnych zapisanych w atrybucie decyzyjnym. Przed przeprowadzeniem właściwego procesu zgromadzone dane należy poddać procesowi dyskretyzacji, który zamieni atrybuty kategoriowe na wartości liczbowe oraz atrybuty ciągłe również na wartości liczbowe, ale odpowiadające określonym zakresom wartości.

Rysunek 1 Hybrydowy model systemu ekspertowego do klasyfikacji regułowej



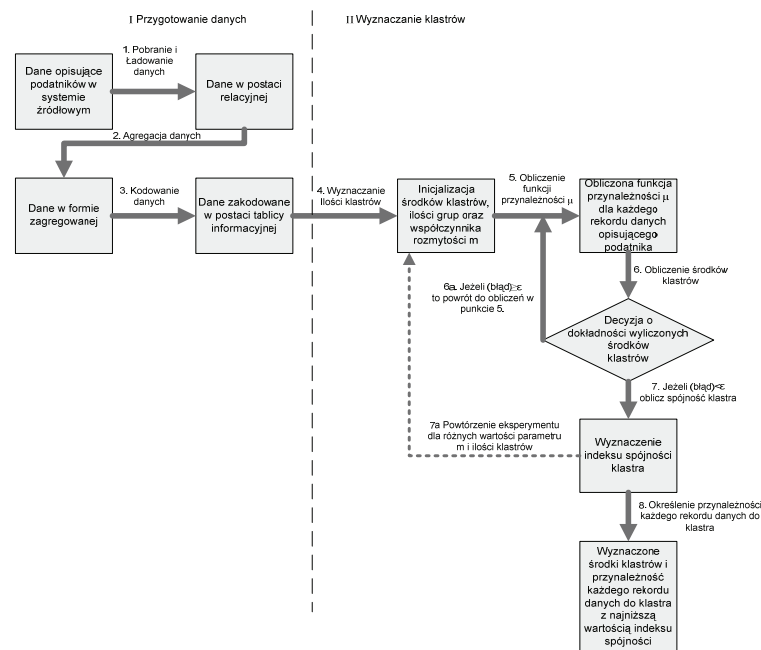
Źródło: opracowanie własne

Kolejnym krokiem wstępnym jest przeprowadzenie pomiarów względnej istotności atrybutów σ , który umożliwi redukcję wymiaru poprzez odrzucenie atrybutów nieistotnych lub o bardzo małym znaczeniu. Sam proces klasyfikacji z wykorzystaniem teorii zbiorów przybliżonych opiera się na definicji dolnego oraz górnego przybliżenia [Pawlak 1991]. Na bazie wspomnianych przybliżeń, dostępnej tablicy informacyjnej wraz z zdyskretyzowanymi danymi występuje zdolność do scharakteryzowania wszystkich klas poprzez utworzenie tablicy decyzyjnej, w której przypisany zostaje każdy z przypadków do określonej klasy

wraz z określeniem siły przynależności. Wynikiem końcowym takiej operacji jest zbiór reguł decyzyjnych, które klasyfikują podatników dla każdej z wyznaczonych grup utworzonych w etapie pierwszym. Procedura badawcza zastosowana w tym etapie została przedstawiona na rysunku nr 3.

Trzecim i końcowym wynikiem realizowanym w trzecim etapie metody jest ranking podatników, który zostaje wyznaczony za pomocą wielokryterialnej, hierarchicznej metody wspomaganie decyzji, jaką jest AHP. W wyniku działania metody uzyskujemy uporządkowaną listę podatników z uwzględnieniem każdego kryterium dla każdej wyznaczonej grupy oddzielnie. Uporządkowanie następuje według najistotniejszych podmiotów, które znajdują się na początku. Są najbardziej podatne na negatywne zachowania podatkowe. Jest to istotny krok, ponieważ do jednej reguły w grupie może być przyporządkowanych od kilkaset do wielu tysięcy podatników. W praktyce uniemożliwiłoby to wybór najbardziej interesujących podatników ze względu na kryterium poszukiwania.

Rysunek 2 Procedura badawcza wyznaczania klastrów – I etap metody

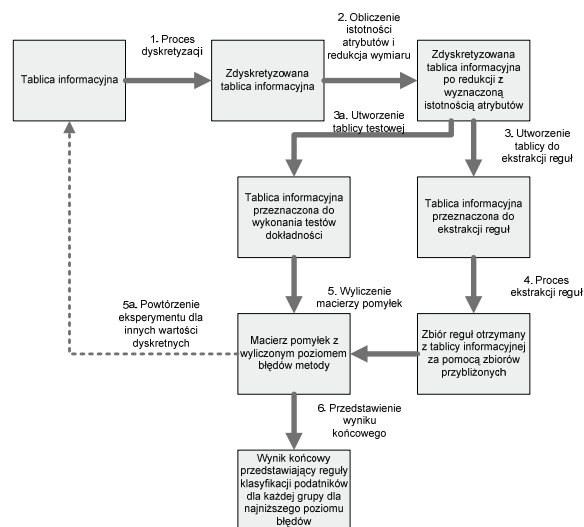


Źródło: opracowanie własne

Dla modelu zastosowano ocenę ilościową kryteriów, która jest możliwa do zrealizowania na podstawie atrybutów o największej istotności opisujących podatników dla klasyfikacji regułowej uzyskanych jako wynik końcowy etapu drugiego. Należy również zwrócić uwagę, że użyty ilościowy model jest w praktyce jedynym możliwym do zastosowania, ponieważ w innym przypadku

osoba merytoryczna musiałaby porównać od prawie dziesięciu do kilkudziesięciu różnych kryteriów. Zgodnie z dostępną wiedzą porównanie przez człowieka większej ilości kryteriów niż od pięciu do dziewięciu zgodnie ze współczesną wiedzą nie jest możliwe [Saaty 1994]. Ponadto porównywanie wielu kryteriów o dosyć podobnym znaczeniu, na przykład paru różnych rodzajów przychodu czy też kilka typów sprzedaży stwarzałyby również duże trudności i zaistniała by możliwość zniekształcenia wyniku. Dlatego też oparcie rankingu na obiektywnie obliczonych istotnościach atrybutów jest w tym przypadku właściwe. Procedura badawcza zastosowana w etapie trzecim została przedstawiona na rysunku 4.

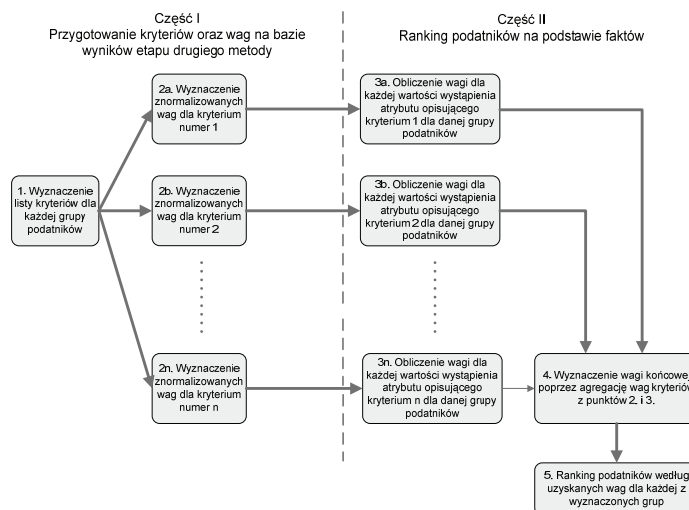
Rysunek 3 Procedura badawcza dla ekstrakcji reguł decyzyjnych – II etap metody



Źródło: opracowanie własne

Sprawdzanie dokładności metody odbywa się dwuetapowo. W pierwszym etapie metody wyznaczony zostaje indeks spójności klastra, który wyznacza jakość klastra na podstawie określenia skupiania obiektów wokół wyznaczonych punktów centralnych. W drugim etapie na podstawie otrzymanej macierzy pomyłek wyznaczone zostają następujące współczynniki: pozytywny współczynnik predykcji (ang. positive predictive value - PPV), negatywny współczynnik predykcji (ang. negative predictive values - NPV), czułość (ang. sensitivity - SE), specyficzność (ang. specificity - SP), całkowita dokładność (ang. accuracy - ACC), całkowity poziom błędów (ang. error rate level - ERR), współczynnik F (ang. F-Measure), współczynnik jakości przewidywań FOM [Lewis 1994] [Vercellis 2009]. Na potrzeby przeprowadzenia eksperymentów oraz uzyskania wyników zostało wytworzone oprogramowanie w języku PL/SQL dla bazy danych Oracle postaci pakietów.

Rysunek 4 Schemat blokowy procedury badawczej wyznaczania rankingu – III etap metody



Źródło: opracowanie własne

WYNIKI BADAŃ TESTOWYCH

Porównanie dokładności autorskiej metody z istniejącymi metodami oraz systemami zostało przeprowadzone z wykorzystaniem parametrów PPV, NPV, SE, SP, ACC, ERR, F, FOM uzyskanych na podstawie macierzy pomyłek. Obliczenia dokładności poszczególnych metod i systemów bazują na tych samych danych, które zostały użyte do uzyskania wyników zaproponowanej hybrydowej metody.

W tabeli 2 zostały przedstawione sumaryczne wyniki dokładności uzyskanej za pomocą zaproponowanej metody w porównaniu do istniejących metod. Wartość dodatnia oznacza większą dokładność przedstawionej hybrydowej metody. Sposób porównania został wykonany w następujący sposób:

$$\Delta KRYT_{sum} = KRYT_{\acute{s}red} - KRYT_{met} \quad (1)$$

,gdzie $\Delta KRYT_{sum}$ – wynik kryterium sumarycznego
 $KRYT_{\acute{s}red}$ – średnia wartość kryterium hybrydowej metody dla wszystkich klastrów, dotyczy kryteriów PPV, NPV, SE, SP, ACC, F, FOM,
 $KRYT_{met}$ – wartość kryteriów istniejących metod i systemów.

Dla kryterium ERR sposób wyliczenia jest następujący:

$$\Delta KRYT_{sum} = KRYT_{met} - KRYT_{\acute{s}red} \quad (2)$$

,gdzie oznaczenia mają takie same znaczenie jak powyżej, natomiast $KRYT$ dotyczy tylko kryterium całkowitego poziomu błęd.

Uzyskane sumaryczne wyniki porównania dokładności zaproponowanej autorskiej metody w porównaniu z istniejącymi metodami są lepsze we wszystkich mierzonych parametrach.

Tabela 2. Sumaryczne porównanie wyników metody z dokładnością istniejących metod

Metoda lub system	ΔPPV_{sum} [%]	ΔNPV_{sum} [%]	ΔSE_{sum} [%]	ΔSP_{sum} [%]	ΔACC_{sum} [%]	ΔERR_{sum} [%]	ΔF_{sum} [%]	ΔFOM_{sum} [%]
Drzewa decyzyjne	31,2	9,5	12,9	28,4	24,6	24,6	25,5	20,6
Naive Bayes	17,5	30,3	19,9	30,1	24,0	24,0	17,4	24,9
SVM	19,6	20,8	15,5	27,3	21,7	21,7	17,2	21,3
Zbiory przybliżone	17,1	20,6	35,3	5,7	17,4	17,4	24,1	20,3
System ISKOS*	15,1	25,9	18,8	23,7	22,2	22,2	17,5	21,1
System KONTROLA*	6,9	48,0	42,7	13,0	27,0	27,0	27,8	27,5

Źródło: opracowanie własne (* - symulowane wyniki)

Podsumowując uzyskane wyniki można stwierdzić, że w przypadku osób fizycznych najważniejsze są parametry zbyt wysokich lub zbyt niskich dochodów powiązane z istniejącymi zaległościami, co może oznaczać ukrywanie lub niewykazywanie dochodów, przy jednocześnie występujących problemach z różnego rodzaju zaległościami. Dla przedsiębiorstw do najważniejszych cech wskazujących na problemy to występowanie spadków płaconych podatków VAT z okresu na okres, występowanie nadwyżki podatku VAT naliczonego nad należnym do przesunięcia na następny miesiąc czy występowanie zaległości egzekucyjnych. Może to oznaczać próby generowania nieprawdziwych kosztów związanych z zakupami, inwestycjami lub uczestnictwem w mechanizmie karuzelowym. Czynnikiem zwiększającym ryzyko jest występowanie zaległości, które wskazują na wcześniejsze problemy z prawem oraz brak rzetelności wobec partnerów gospodarczych.

PODSUMOWANIE

Zgodnie z założeniami został przedstawiony hybrydowy model systemu ekspertowego umożliwiający wykonanie oceny podatników z większą dokładnością od występujących w praktyce metod. Realizacja postawiona w badaniach była możliwa poprzez eliminację słabości i ograniczeń występujących w istniejących metodach i technikach. Istotne było również poznanie i zrozumienie dziedziny badań, co umożliwiło wybór i przygotowanie odpowiednich danych. W wyniku powstała procedura, która składa się z trzech głównych etapów: rozmytej analizy skupień, klasyfikacji regułowej w poszczególnych grupach oraz rankingu podatników. Dla wyznaczenia dokładności została zaproponowana dwuetapowa kontrola uzyskanych wyników. Efektem końcowym jest osiągnięcie większej precyzji selekcji podatników od istniejących rozwiązań. Realizacja podjętych badań jest istotna z punktu widzenia pracy administracji skarbowej, która działa w interesie i na rzecz całego społeczeństwa. Wykorzystanie modelu

umożliwi bowiem między innymi niwelowanie działań w tak zwanej „szarej strefie” i gwarancję wpływów do budżetu na określonym poziomie. Z punktu widzenia praktyki gospodarczej będzie zapobiegać występowaniu nieuczciwej konkurencji pomiędzy przedsiębiorcami tej samej branży.

BIBLIOGRAFIA

- Bandyopadhyay S., Maulik U., Holder L. B, Cook D. J. (2005) *Advanced Methods for Knowledge Discovery from Complex Data*, Springer.
- Bazan J., Synak P., Wroblewski J. (2000) *Rough Set Algorithms in Classification Problem*, Springer.
- Becker J., Misztal L. (2010) Wielokryterialny model oceny podatników indywidualnych, Seria: *Studia i Materiały Polskiego Stowarzyszenia Zarządzania Wiedzą* nr 28, Bydgoszcz.
- Bezdek J.C., Dubois D., Prade H. (1999) *Fuzzy sets In apprioximate resoning and information systems*, Kluwer Academic.
- Budziński R., Misztal L. (2009) Wykorzystanie drzew decyzyjnych oraz ekstrakcji reguł w zadaniu klasyfikacji podatników, *Polskie Stowarzyszenie Zarządzania Wiedzą*, Bydgoszcz.
- Budziński R., Misztal L. (2009) Zastosowanie algorytmu maszyny wektorów wspierających do klasyfikacji podatników z wykorzystaniem bazy danych Oracle 11g, *Polskie Stowarzyszenie Zarządzania Wiedzą*, Bydgoszcz.
- COMARCH (2004) *Wizja systemu ISKOS*.
- Cross Industry Standard Process for Data Mining (2005), <http://www.crisp-dm.org>
- Drzymala-Busse J. (2006) *Rough Set Strategies to Data with Missing Attribute Values*, Springer.
- Duntsch I., Gediga G. (2000) *Rough set data analysis*, Methodos Publisher.
- Goch W. (2006) *Typowanie podmiotów do kontroli w podsystemie KONTROLA*, Białobrzegi.
- Han J., Kamber M. (2006) *Data Mining: Concepts and Techniques second edition*, Morgan Kaufmann Publishers.
- Hand D., Mannila H., Smith P. (2001) *Principles of Data Mining*, Massachusetts Institute of Technology.
- Jain K.A., Dubes C.R. (1988) *Algorithms for clustering data*, Prentice Hall.
- Lewis D., Gale W. (1994) Training text classifiers by uncertainty sampling, *ACM SIGIR Conference*.
- Michael W.Berry, Matu Castellanos (2008) *Survey of text mining II clustering, classification and retrieval*, Springer.
- Ministerstwo Finansów (2010) *Krajowy Plan Dyscypliny Podatkowej na 2010 rok*.
- Misztal L. (2009) *Applying Rough Sets for the Task of Rule Classification of Tax Payers*, *Advanced Computer Systems (PAK)*, Szczecin (Gliwice).
- Misztal L. (2009) Wykorzystanie naiwnego algorytmu Bayes'a w zadaniu klasyfikacji podatników, *Kwartalnik Informatyki Stosowanej*, Szczecin.
- Miyamoto S., Ichihashi H., Honda K. (2008) *Algorithms for fuzzy clustering*, Springer.

- Pawlak Z. (1991) *Rough Sets – Theoretical Aspects of Reasoning about Data*, Kluwer Academic Publishers.
- Piegat A. (2006) *Zbiory przybliżone – wykłady*, Szczecin.
- Rogacka E., Janicki T. (2009) *Opis tabel podsystemu KONTROLA wydanie 1.6.1*, IS Wrocław.
- Saaty T.L. (1994) *Fundamentals of Decision Making and Priority Theory with the analytic hierarchy process*, Pittsburgh, PA RWS Publications.
- Schoelkopf B., Smola A.J. (2002) *Learning with kernels, Support Vector Machines, Regularization, Optimization, and Beyond*, The MIT Press.
- SPSS (2004) *Clementine Data Mining Project*.
- Triantaphyllou E., Felici G. (2006) *Data Mining & Knowledge Discovery based in Rule Induction*, Springer Science.
- Vercellis C. (2009) *Business intelligence – data mining and optimization for decision making*, Wiley.
- Witten I.H., Eibe Frank (2005) *Data Mining Practical Machine Learning Tools and Techniques*, Morgan Kaufmann.
- Xie L.X., Beni G. (1991) A validity measure for fuzzy clustering, *IEEE Transactions*.
- Yao J., Dash M., Tan S.T., Liu H. (2000) Entropy-based fuzzy clustering and fuzzy modeling, *Fuzzy Sets and Systems*.
- Zabicka K. (2003) *Poradnik użytkownika aplikacji KONTROLA*, IS Wrocław OZ Bielsko-Biała.
- Zespół projektowy KONTROLA (2009) *Podręcznik użytkownika aplikacji POLTAX KONTROLA*, IS Wrocław.

HYBRID MODEL OF EXPERT SYSTEM FOR ESTIMATION OF TAXPAYERS

Abstract: Proposed model identifies taxpayers on the basis of their features and properties that point to bigger possibility of taxation law observance problems. Model eliminates weaknesses of well known classification algorithms, as well as expert systems and taxpayers typing assists applications used wildly in tax offices. It is possible because of designing hybrid model that reflects well behavior of payers. Applying modern concepts in model predisposes it for future implementation of software solution that can be used in taxation administration.

Key words: taxpayers classification, rough set theory, fuzzy clustering, AHP method, data mining