

Tomasz Bartłomowicz

Uniwersytet Ekonomiczny we Wrocławiu

PROGNOZOWANIE PREFERENCJI WYRAŻONYCH Z WYKORZYSTANIEM METODY k -NN*

Streszczenie: Klasyczne zastosowanie metody k -najbliższych sąsiadów (*k-Nearest Neighbours*, k -NN) ma miejsce w klasyfikacji, tj. w przydzielaniu obserwacji do klas na bazie atrybutów tych obserwacji. Ponadto, z racji przynależności metody do grupy tzw. metod uczących się, zastosowanie metody obejmuje także prognozowanie.

W artykule zaprezentowano prognozowanie preferencji wyrażonych konsumentów jednego z najszybciej rozwijających się obecnie rynków – rynku nieruchomości. Metodologiczną podstawę prognozowania stanowi metoda k -NN, przykładami zaś potwierdzającymi możliwość prognozowania wybranych aspektów tego rynku są m.in. prognozy atrakcyjności oraz aktualności ofert sprzedaży nieruchomości.

Słowa kluczowe: prognozowanie, metoda najmniejszych kwadratów, k -NN, preferencje.

1. Wstęp

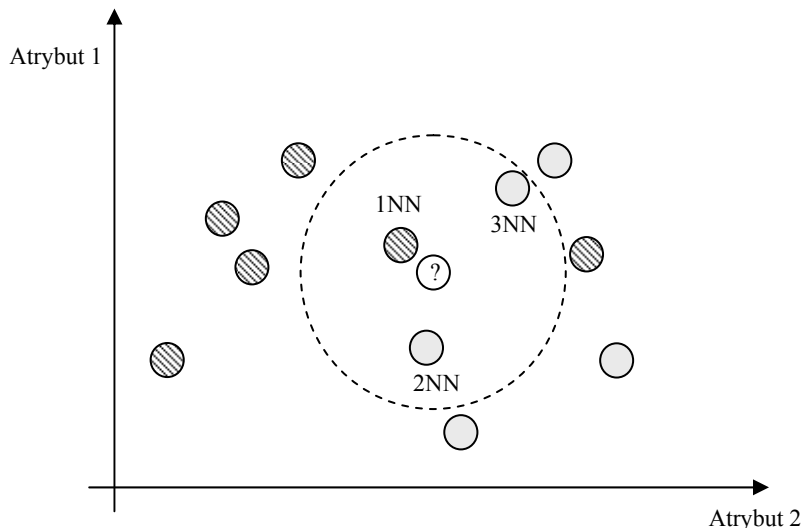
Według jednej z definicji termin preferencja oznacza pierwszeństwo, uprzywilejowanie, górowanie czegoś nad czymś, przedkładanie czegoś nad coś, tudzież akt wyboru polegający na uznaniu czegoś za lepsze (por. [*Internetowy słownik...*]). Wykorzystując tę ogólnikową charakterystykę i przenosząc ją na grunt teorii ekonomii, przez preferencje konsumenckie należy rozumieć zdolność konsumentów do porządkowania i wyboru produktów lub usług oferowanych na rynku na określonych warunkach. Jeśli dodatkowo zauważy się, iż wybór ten umożliwia subiektywny system oceny, którym kieruje się każdy z konsumentów, dokonując określonego zakupu, a wymierna ocena dokonywana przez nabywców oznacza możliwość odzwierciedlenia i formalizacji gustów, upodobań bądź potrzeb, możliwe staje się zrozumienie zachowań konsumentów, a w dalszej perspektywie ich prognozowanie.

* Praca naukowa finansowana ze środków na naukę w latach 2009-2012 jako projekt badawczy nr N N111 446037.

W artykule zaprezentowano prognozowanie preferencji wyrażonych konsumentów jednego z najszybciej rozwijających się obecnie rynków – rynku nieruchomości, którymi w opracowaniu są nabywcy niezabudowanych nieruchomości gruntowych. Metodologiczną podstawę prognozowania stanowi metoda k -najbliższych sąsiadów, a przykładami potwierdzającymi możliwość prognozowania wybranych aspektów tego rynku są m.in. prognozy atrakcyjności oraz aktualności ofert sprzedaży nieruchomości. Zakładając, iż głównymi odbiorcami ww. prognoz będą pośrednicy w obrocie nieruchomościami oraz ich klienci, głównie po stronie podażowej, można dowiedzieć, iż w artykule podejmuje się próbę zastosowania naukowej metody prognozowania w jej czysto praktycznym charakterze.

2. Metoda k -najbliższych sąsiadów jako metoda prognozowania

Metoda k -najbliższych sąsiadów (ang. *k-Nearest Neighbours*, k -NN) została zaproponowana w pracy [Fix, Hodges 1951, s. 261-279] jako jeden z najlepszych klasyfikatorów minimalnoodległościowych dla danych w ujęciu klasycznym. Oznacza to, że typowym zastosowaniem metody jest jej wykorzystanie do problemu klasyfikacji, przy czym metoda z powodzeniem jest stosowana także w prognozowaniu wartości zmiennej losowej.



Rys. 1. Graficzna prezentacja działania metody k -najbliższych sąsiadów

Źródło: opracowanie własne na podstawie [Classification: *k Nearest...*; *k-Nearest Neighbor...*].

W najprostszym ujęciu zastosowanie metody k -NN polega na wyborze k obserwacji leżących najbliżej nieznannej obserwacji i przypisaniu jej klasy, do której

należy najwięcej spośród jej k „sąsiadów” (por. [Tadeusiewicz, Lasiński 1991, s. 43-53]). W sposób graficzny, w ograniczeniu do dwóch atrybutów, działanie metody k -NN przedstawia rys. 1. Obiekt ze znakiem zapytania reprezentuje obserwację klasyfikowaną do pierwszej klasy – klasy obiektów zakreskowanych, lub drugiej klasy – klasy obiektów zacieniowanych. W sytuacji gdy $k = 1$, obiekt jest klasyfikowany do pierwszej klasy z uwagi na najbliższą odległość obiektu zakreskowanego. W sytuacji gdy $k = 3$, obiekt jest klasyfikowany do drugiej klasy, ponieważ wewnątrz okręgu zawierającego 3 najbliższe obserwacje znajdują się 2 zacieniowane kółka i tylko 1 kółko zakreskowane.

Należy w tym miejscu zauważyć, iż definicja „najbliższej obserwacji” sprowadza się do minimalizacji pewnej metryki mierzącej odległość między wektorami zmiennych dwóch obserwacji. Zwykle stosowana jest tu odległość euklidesowa:

$$d_E(x_i, x_k) = \sqrt{\sum_{j=1}^m (x_{ij} - x_{kj})^2}, \quad (1)$$

gdzie: $d_E(x_i, x_k)$ – odległość euklidesowa pomiędzy i -tym wektorem obserwacji ze zbioru uczącego a k -tym wektorem testowym ze zbioru rozpoznawanego¹,

$i = 1, \dots, k$ – numer profilu,

$j = 1, \dots, m$ – numer zmiennej,

m – liczba zmiennych,

lub odległość Mahalanobisa:

$$d_M(x_i, x_k) = \left[\sum_{j=1}^m \sum_{l=1}^m s_{jl} (x_{ij} - x_{kj})(x_{il} - x_{kl}) \right]^{0,5}, \quad (2)$$

gdzie: $d_M(x_i, x_k)$ – odległość Mahalanobisa pomiędzy i -tym wektorem obserwacji ze zbioru uczącego a k -tym wektorem testowym ze zbioru rozpoznawanego,

s_{jl} – element odwrotny do macierzy kowariancji zbioru obserwacji,

pozostałe oznaczenia jak we wzorze (1).

Znane są także przypadki stosowania innych, w tym prostszych miar, np. metryki miejskiej (por. [*k-Nearest Neighbor...*]).

W sytuacji gdy wykorzystanie metody k -NN ogranicza się do pomiaru odległości oraz wyboru najbliższych sąsiadów, oznacza to klasyfikację obserwacji do k licznych klas. Należy zauważyć, iż z racji przynależności metody do grupy tzw.

¹ Zbiór rozpoznawany to zbiór zawierający obserwacje podlegające klasyfikacji, czyli tzw. obserwacje testowe.

metod uczących się w metodzie nie korzysta się przy tym z jakiegokolwiek modelu, a klasyfikacja odbywa się na podstawie zbioru uczącego.

W przypadku zastosowania metody na potrzeby prognozowania należy zrealizować następujące założenia metody k -NN w postaci (por. [*k-Nearest Neighbor...*]):

- danego zbioru uczącego zawierającego obserwacje, z których każda ma przypisany wektor zmiennych objaśniających oraz wartość zmiennej objaśnianej,
- danej obserwacji C z przypisanym wektorem zmiennych objaśniających, dla której istnieje możliwość prognozowania wartości zmiennej objaśnianej.

Algorytm prognozowania w metodzie polega wówczas na: porównaniu wartości zmiennych objaśniających dla obserwacji C z wartościami tych zmiennych dla każdej obserwacji w zbiorze uczącym, wyborze k najbliższych do C obserwacji ze zbioru uczącego, a następnie wyznaczeniu prognozy, zgodnie z przyjętą regułą prognostyczną, na podstawie wartości zmiennej objaśnianej dla wybranych obserwacji. Najczęściej wyznaczenie prognozy polega na uśrednieniu wartości zmiennej objaśnianej (reguła prognozy nieobciążonej), wytypowaniu dominanty wartości zmiennej prognozowanej (reguła według największego prawdopodobieństwa) lub ich mediany (reguła według mediany).

Przykład obrazujący sposób prognozowania za pomocą metody k -najbliższych sąsiadów zaprezentowano w dalszej części artykułu (por. pkt 4).

3. Zbiór uczący metody k -NN jako źródło informacji o preferencjach wyrażonych

W badaniach preferencji wykorzystuje się obserwacje historyczne (wtórne źródła danych), a także informacje o charakterze antycypacyjnym, wyrażające intencje nabywców (pierwotne źródła danych). Oznacza to podział preferencji na dwie kategorie – preferencje ujawnione oraz preferencje wyrażone (por. [Bąk 2004, s. 42-43; Gatnar, Walesiak 2004, s. 139-140]).

Przez preferencje ujawnione należy rozumieć postępowanie rynkowe nabywców, które stanowi odbicie ich rzeczywistych decyzji rynkowych. Podstawą analizy preferencji jest w tym przypadku materiał statystyczny zgromadzony w wyniku rejestracji danych o rzeczywistych wyborach rynkowych. Uzupełniającym źródłem danych mogą być przeprowadzone *a posteriori* sondaże dotyczące wyborów rynkowych wcześniej dokonanych przez nabywców (por. [Bąk 2004, s. 42-43; Gatnar, Walesiak 2004, s. 139]).

W przypadku preferencji wyrażonych, odpowiadających deklarowanym preferencjom nabywców, źródła danych są gromadzone najczęściej *a priori* z wykorzystaniem sondaży umożliwiających rejestrację wyrażonych intencji nabywców (por. [Bąk 2004, s. 43; Gatnar, Walesiak 2004, s. 139-140]). Przykładem zbioru danych o preferencjach wyrażonych jest wykorzystywany na potrzeby prognozowania zbiór uczący metody k -najbliższych sąsiadów.

W opracowaniu zbioru uczący stanowi zestaw 14 hipotetycznych profili (wariantów) niezabudowanych nieruchomości gruntowych (P01, P02, ..., P14) opisanych zestawem 4 cech (atrybutów, zmiennych objaśniających) w postaci lokalizacji, uzbrojenia, powierzchni oraz kształtu nieruchomości². Każda ze zmiennych realizowana jest przez odpowiadające jej poziomy. W przypadku lokalizacji nieruchomości jest to odpowiednio lokalizacja: „korzystna”, „przeciętna” i „niekorzystna”, w przypadku pozostałych cech odpowiednio: uzbrojenie – „kompletne”, „podstawowe” i „niekompletne”, powierzchnia – „poniżej 1000 m²” oraz „1000 m² i więcej”, kształt – „typowy” oraz „nietypowy”³ (por. tab. 1).

Tabela 1. Zbiór uczący – 14 hipotetycznych nieruchomości oraz ich atrakcyjność

Profil	Cechy nieruchomości				Atrakcyjność
	lokalizacja	uzbrojenie	powierzchnia	kształt	
P01	korzystna	kompletne	poniżej 1000 m ²	nietypowy	duża
P02	korzystna	podstawowe	1000 m ² i więcej	typowy	duża
P03	przeciętna	niekompletne	1000 m ² i więcej	nietypowy	mała
P04	niekorzystna	kompletne	1000 m ² i więcej	typowy	mała
P05	korzystna	podstawowe	poniżej 1000 m ²	nietypowy	mała
P06	niekorzystna	podstawowe	poniżej 1000 m ²	nietypowy	mała
P07	przeciętna	niekompletne	1000 m ² i więcej	typowy	duża
P08	niekorzystna	niekompletne	1000 m ² i więcej	typowy	mała
P09	przeciętna	podstawowe	poniżej 1000 m ²	nietypowy	mała
P10	korzystna	podstawowe	poniżej 1000 m ²	typowy	duża
P11	niekorzystna	kompletne	1000 m ² i więcej	nietypowy	duża
P12	korzystna	kompletne	1000 m ² i więcej	typowy	duża
P13	przeciętna	niekompletne	poniżej 1000 m ²	nietypowy	mała
P14	niekorzystna	podstawowe	poniżej 1000 m ²	nietypowy	mała

Źródło: opracowanie własne.

Jeśli przyjmiemy, że przedstawione warianty nieruchomości odzwierciedlają oferty przykładowego biura pośrednictwa w obrocie nieruchomościami, kluczową dla metody prognozowania rolę zmiennej objaśnianej będzie odgrywać atrakcyjność danej oferty. Zakłada się, iż atrakcyjność ofert nieruchomości w prezentowanym przykładzie mierzona jest liczbą odwiedzin poszczególnych wariantów na stronie internetowej biura pośrednictwa w obrocie nieruchomościami, którego ofer-

² Mimo że przykład ma charakter teoretyczny, wybór cech opisujących nieruchomości bazuje na przeprowadzonych przez autora badaniach relatywnej ważności cech nieruchomości i odzwierciedla najważniejsze cechy niezabudowanych nieruchomości gruntowych (por. [Bartłomowicz 2002, s. 272-280; Pawlukowicz, Bartłomowicz 2005, s. 128-139]).

³ Zaprezentowane, niepowtarzające się warianty nieruchomości nie wyczerpują wszystkich możliwych kombinacji. Ze względu na liczbę poziomów realizacji poszczególnych zmiennych całkowita możliwa liczba wariantów nieruchomości zamyka się liczbą 36 profili ($3 \times 3 \times 2 \times 2 = 36$). W praktyce oznacza to możliwość pojawienia się wariantu wcześniej niesklasyfikowanego, dla którego istnieje potrzeba wyznaczenia prognozy.

ty w zakresie podstawowych cech odzwierciedlają warianty ze zbioru uczącego. Na tej podstawie atrakcyjność każdego z wariantów identyfikowana jest na poziomie „mała”, gdy dany profil charakteryzuje się zbyt małą liczbą odwiedzin, tym samym nie cieszy się wystarczającym zainteresowaniem, lub „duża” w przeciwnym wypadku.

Warto zauważyć, że znajomość atrakcyjności poszczególnych wariantów nieruchomości w postaci wspomnianego zbioru uczącego, w połączeniu z omawianą metodą, umożliwia prognozowanie atrakcyjności dowolnego wariantu nieruchomości. Odzwierciedla to sytuację, w której pośrednik w obrocie nieruchomościami, przyjmując kolejną ofertę nieruchomości, jest w stanie „naukowo” ocenić jej przewidywaną atrakcyjność.

Z punktu widzenia zaprezentowanego podziału preferencji na dwa rodzaje, wykorzystywany w badaniu zbiór danych, z racji wspomnianego internetowego sposobu pomiaru atrakcyjności poszczególnych ofert nieruchomości poprzez większe lub mniejsze deklaracje zainteresowania daną ofertą, w pełni odpowiada hipotetycznym preferencjom nabywców w postaci preferencji wyrażonych.

4. Przykład prognozowania preferencji wyrażonych nabywców nieruchomości z wykorzystaniem metody k -NN

Dokonując próby sformułowania zadania prognostycznego, należałoby zauważyć, iż w prezentowanym przykładzie prognozowanym zjawiskiem jest zainteresowanie ofertami nieruchomości, prognozowanymi obiektami są warianty nieruchomości, prognozowaną zmienną zaś atrakcyjność poszczególnych wariantów, co podkreśla fakt budowania prognoz dla rynku nieruchomości. W tej sytuacji wybór metody prognozowania w postaci metody k -najbliższych sąsiadów wydaje się uzasadniony. Wynika to z właściwości metody, która daje potencjalnie lepsze rezultaty wszędzie tam, gdzie zależność między zmiennymi objaśniającymi a objaśnianymi jest złożona lub nietypowa (np. niemonotoniczna), w praktyce trudna do modelowania w sposób klasyczny⁴. Wydaje się, że rynek nieruchomości spełnia ten warunek, gdyż na rynku tym istnieje możliwość traktowania przeciwstawnych poziomów poszczególnych cech jako najbardziej pożądanых, a więc najlepiej ocenianych⁵.

⁴ W przypadku gdy zależność ta jest łatwa do interpretacji (np. liniowa), a zbiór nie zawiera obserwacji odstających, metody klasyczne (np. regresja liniowa) dają zwykle dokładniejsze wyniki (por. [Tadeusiewicz, Lasiński 1991]).

⁵ Na przykład lokalizacji „peryferyjnej” nieruchomości może być przyporządkowana ocena najwyższa, jak również, w przypadku innych nabywców – ocena najniższa. Ponieważ najlepiej oceniany, w zależności od indywidualnych preferencji nabywców, może być na przykładzie lokalizacji dowolny jej wariant, nie istnieje ściśle rosnący lub malejący kierunek preferencji względem tych atrybutów.

Ponadto należy zauważyć, iż w przypadku rynku nieruchomości ilość oraz jakość dostępnych danych może się okazać niewystarczająca do budowy odpowiedniego modelu, co dodatkowo przemawia za wykorzystaniem metody k -NN.

W celu prezentacji sposobu prognozowania preferencji wyrażonych nabywców nieruchomości w przykładzie zakłada się pojawienie na rynku, tj. w ofercie biura pośrednictwa nieruchomości, nowego wariantu nieruchomości w postaci profilu P15 o cechach: „przeciętna” lokalizacja, „niekompletne” uzbrojenie, „powierzchnia 1000 m² i więcej” oraz „nietypowy” kształt. Przyjmując, iż celem prognozowania jest określenie atrakcyjności sprzedaży nowego wariantu nieruchomości (nowej oferty) zgodnie z algorytmem metody k -NN, w kroku pierwszym procedury prognozowania, dla nowego wariantu wybiera się k wariantów nieruchomości ze zbioru uczącego, położonych najbliżej względem wariantu testowego. Po zakodowaniu danych niemetrycznych, opisujących wszystkie 14 profili nieruchomości, pomiar odległości odbywa się z wykorzystaniem m.in. wzoru (1), przy czym możliwe jest wyznaczenie odległości nie tylko między poszczególnymi wariantami nieruchomości, ale także między wariantami niekoniecznie występującymi w zbiorze uczącym.

Przy założeniu⁶ $k = 3$, wykorzystując metrykę miejską (por. tab. 2) oraz metrykę euklidesową (por. tab. 3), wyznaczono odległości między profilem testowym (P15) a każdym z profili ze zbioru uczącego. Porównanie wyników pomiaru pozwala jednoznacznie wyróżnić 3 „najbliższych sąsiadów” dla profilu testowego. W przykładzie są to profile: P03, P07 oraz P13.

Występowanie najmniejszych odległości pomiędzy tymi profilami a wariantem testowym wynika bezpośrednio z ich wzajemnego podobieństwa. Można zauważyć, iż wariant P03 ma identyczne charakterystyki co testowy wariant P15, stąd zerowa odległość między poszczególnymi cechami tych wariantów, a co za tym idzie – samymi wariantami. W przypadku pozostałych 2 „najbliższych sąsiadów” profil P07 różni się „typowym” kształtem w miejsce „nietypowego” dla wariantu testowego; w przypadku wariantu P13 powierzchnia jest „poniżej 1000 m²”, podczas gdy wariant P15 charakteryzuje się powierzchnią „1000 m² i więcej”. Należy zauważyć, iż w przypadku obu pomiarów (por. tab. 2 i 3) wyróżniona klasa zawiera te same profile nieruchomości⁷.

Wyróżnienie „najbliższych sąsiadów” wariantu testowego, a tym samym rozpoznanie atrakcyjności tych wariantów oznacza, w kolejnym kroku, możliwość określenia przyszłej atrakcyjności testowego wariantu nieruchomości. Ponieważ dwa z wariantów „sąsiednich” (P03, P13) mają „małą” atrakcyjność (kod 0), a tylko

⁶ Wielkość klasy w metodzie k -najbliższych sąsiadów jest definiowana przez użytkownika.

⁷ W przypadku $k = 4$ do trzech ww. wariantów, zgodnie z ideą poszukiwania najmniejszej odległości między wariantem testowym a wariantami ze zbioru uczącego, należałoby dodać wariant P08, w przypadku $k = 5$ wariant P09 itd.

Tabela 2. Odległości miejskie między profilem P15 a profilami zbioru uczącego

Profil	Odległości miejskie cech nieruchomości				Odległość miejska
	lokalizacja	uzbrojenie	powierzchnia	kształt	
(1)	(2)	(3)	(4)	(5)	(6)
P01	1,00	2,00	1,00	0,00	4,00
P02	1,00	1,00	0,00	1,00	3,00
P03	0,00	0,00	0,00	0,00	0,00
P04	1,00	2,00	0,00	1,00	4,00
P05	1,00	1,00	1,00	0,00	3,00
P06	1,00	1,00	1,00	0,00	3,00
P07	0,00	0,00	0,00	1,00	1,00
P08	1,00	0,00	0,00	1,00	2,00
P09	0,00	1,00	1,00	0,00	2,00
P10	1,00	1,00	1,00	1,00	4,00
P11	1,00	2,00	0,00	0,00	3,00
P12	1,00	2,00	0,00	1,00	4,00
P13	0,00	0,00	1,00	0,00	1,00
P14	1,00	1,00	1,00	0,00	3,00
P15	0,00	0,00	0,00	0,00	0,00

Źródło: obliczenia własne.

Tabela 3. Odległości euklidesowe między profilem P15 a profilami zbioru uczącego

Profil	Odległości euklidesowe cech nieruchomości				Odległość euklidesowa
	lokalizacja	uzbrojenie	powierzchnia	kształt	
(1)	(2)	(3)	(4)	(5)	(6)
P01	1,00	4,00	1,00	0,00	2,45
P02	1,00	1,00	0,00	1,00	1,73
P03	0,00	0,00	0,00	0,00	0,00
P04	1,00	4,00	0,00	1,00	2,45
P05	1,00	1,00	1,00	0,00	1,73
P06	1,00	1,00	1,00	0,00	1,73
P07	0,00	0,00	0,00	1,00	1,00
P08	1,00	0,00	0,00	1,00	1,41
P09	0,00	1,00	1,00	0,00	1,41
P10	1,00	1,00	1,00	1,00	2,00
P11	1,00	4,00	0,00	0,00	2,24
P12	1,00	4,00	0,00	1,00	2,45
P13	0,00	0,00	1,00	0,00	1,00
P14	1,00	1,00	1,00	0,00	1,73
P15	0,00	0,00	0,00	0,00	0,00

Źródło: obliczenia własne.

dla jednego wariantu (P07) jest to poziom „duża” (kod 1), prognozę atrakcyjności wariantu pojawiającego się na rynku (P15) należy określić zgodnie z formułą:

$$P15_a = \frac{1}{m} \sum_{i=1}^m \{0,1,0\} = \frac{1}{3}.$$

Ze względu na to, że w przykładzie występują tylko dwa możliwe poziomy cechy atrakcyjność, punktem odniesienia w prognozowaniu atrakcyjności wariantu testowego jest poziom $1/2$. Ponieważ:

$$P15_a = \frac{1}{3} < \frac{1}{2},$$

tym samym należy oczekiwać „małej” atrakcyjności wariantu testowego. W uproszczony sposób, odpowiadający budowie prognozy w oparciu o regułę według największego prawdopodobieństwa, a więc dominantę wartości zmiennej prognozowanej, powyższą zależność można dodatkowo zapisać następująco:

$$\text{„mała”} + \text{„duża”} + \text{„mała”} \rightarrow \text{„mała”}.$$

W praktyce rynku nieruchomości oznacza to prognozę małego zainteresowania wariantem nieruchomości, który pomimo stosunkowo dużej powierzchni („1000 m² i więcej”) charakteryzuje się „przeciętną” lokalizacją, „niekompletnym” uzbrojeniem oraz „nietypowym” kształtem.

Należy w tym miejscu dodać, iż przykład prognozy atrakcyjności nieruchomości, w praktyce atrakcyjności oferty sprzedaży nieruchomości metodą k -NN, może być wykorzystany do prognozowania innych wielkości charakteryzujących oferty biura pośrednictwa w obrocie nieruchomościami. W opinii autora, w sposób analogiczny do sposobu budowy przedstawionej prognozy można prognozować np. aktualność oferty sprzedaży nieruchomości. Wówczas zmienną objaśnianą powinna być aktualność oferty sprzedaży nieruchomości identyfikowana np. w miesiącach jej trwania (z przykładowymi poziomami: do 1 miesiąca, 1-2 miesiące, powyżej 2 miesięcy), natomiast zmiennymi objaśniającymi, przy założeniu zgodności odpowiednich danych, mogą być zmienne występujące w zaprezentowanym przykładzie.

5. Uwagi końcowe

Unikalny charakter nieruchomości jako przedmiotu wymiany powoduje, iż rynek nieruchomości różni się znacznie od pozostałych rynków. Mówi się wręcz o rynku niejednorodnym, niedoskonałym, o małej elastyczności podaży i popytu, który wymaga fachowej obsługi. Wreszcie, ze względu na specyficzne cechy nieruchomości, jak m.in. ich stałość w miejscu, rynek ten ogranicza się do obszaru rynku lokalnego (por. [Kucharska-Stasiak 1999, s. 10]).

Wspomniana specyfika rynku nieruchomości nie oznacza jednak braku możliwości prognozowania wybranych jego aspektów. Co więcej, istnieje wręcz potrzeba prognozowania wybranych wielkości i wyraźne zapotrzebowanie na prognozy z rynku nieruchomości. W sposób jednoznaczny zauważa to np. L. Kałkowski: „[...] w dostępnych badaniach i publikacjach dominują opisy i analizy przeszłości, bardzo brakuje prób prognozowania spodziewanych trendów i koniunktury w poszczególnych segmentach nieruchomości” (por. [Kałkowski 2001, s. 134]).

Wobec powyższego, poszukiwanie alternatywnych metod prognozowania dla rynku nieruchomości, w tym metod prognozowania preferencji nabywców nieruchomości, co jest przedmiotem tego artykułu, wydaje się być w pełni uzasadnione. W opinii autora, zaprezentowany przykład potwierdza przydatność metody k -najbliższych sąsiadów do prognozowania wybranych aspektów tego rynku i może stanowić przykład kolejnego narzędzia jego modelowania.

Podsumowując, należy przyjąć, iż w praktyce rynku nieruchomości umiejętność prognozowania np. atrakcyjności rozmaitych wariantów nieruchomości może oznaczać dla dewelopera możliwość wyjścia na rynek z odpowiednią, tj. oczekiwaną przez klientów, ofertą. Dysponowanie konkretną ofertą nieruchomości może dać zarazem wyobrażenie o jej przyszłym potencjalnym zainteresowaniu ze strony inwestorów, co może oznaczać nawet zmianę jej wartości.

Literatura

- Bartłomowicz T., *Zastosowanie metody conjoint analysis do pomiaru preferencji potencjalnych nabywców nieruchomości*, [w:] K. Jajuga, M. Walesiak (red.), *Taksonomia 9. Klasyfikacja i analiza danych – teoria i zastosowania*, Prace Naukowe Akademii Ekonomicznej we Wrocławiu nr 942, Wydawnictwo AE, Wrocław 2002.
- Bąk A., *Dekompozycyjne metody pomiaru preferencji w badaniach marketingowych*, Wydawnictwo AE, Wrocław 2004.
- Classification: k Nearest Neighbours*, <http://www.cs.ucc.ie/~dgb/courses/tai/notes/handout4.pdf>.
- Fix E., Hodges J.L., *Discriminatory analysis – nonparametric discrimination: consistency properties*, USAF School of Aviation Medicine, Randolph Field 1951.
- Gatnar E., Walesiak M. (red.), *Metody statystycznej analizy wielowymiarowej w badaniach marketingowych*, Wydawnictwo AE, Wrocław 2004.
- Internetowy słownik języka polskiego*, PWN, Warszawa, <http://sjp.pwn.pl>.
- Kałkowski L. (red.), *Rynek nieruchomości w Polsce*, Twigger, Warszawa 2001.
- Kucharska-Stasiak E. (red.), *Rynek nieruchomości. Wybrane problemy*, Wydawnictwo Uniwersytetu Łódzkiego, Łódź 1999.
- k-Nearest Neighbor algorithm*, http://en.wikipedia.org/wiki/K-nearest_neighbor_algorithm.
- Pawlukowicz R., Bartłomowicz T., *Conjoint analysis jako sposób wyznaczania wag cech rynkowych w wycenie rynkowej nieruchomości za pomocą podejścia porównawczego*, [w:] J. Dziechciarz (red.), *Ekonometria 15*, Prace Naukowe Akademii Ekonomicznej we Wrocławiu nr 1096, Wydawnictwo AE, Wrocław 2005.
- Tadeusiewicz R., Lasiński M., *Rozpoznawanie obrazów*, PWN, Warszawa 1991.

FORECASTING OF STATED PREFERENCES USING K -NN METHOD

Summary: In classical implementation k -Nearest Neighbours algorithm (k -NN) is a method for classifying objects based on closest training examples in the feature space. K -NN is a type of instance-based learning, where the function is only approximated locally and all computation is deferred until classification. It can also be used for forecasting.

The paper presents implementation of k -NN method by way of forecasting for real estate market. Among examples of k -NN forecasting, in author's opinion, there can be prognosis of relevance and timeliness of the real estate offers.