INFRASTRUKTURA I EKOLOGIA TERENÓW WIEJSKICH
INFRASTRUCTURE AND ECOLOGY OF RURAL AREAS

Nr 4/2011, POLSKA AKADEMIA NAUK, Oddział w Krakowie, s. 163–172
Komisja Technicznej Infrastruktury Wsi

Mariusz Frukacz, Elżbieta Jasińska, Edward Preweda

SEGMENTACJA RYNKU NIERUCHOMOŚCI GRUNTOWYCH DLA WYBRANYCH GMIN MIEJSKO-WIEJSKICH POWIATU KRAKOWSKIEGO

REAL ESTATE MARKET SEGMENTATION FOR CHOSEN GROUND IMMOBILITY OF URBAN-RURAL MUNICIPALITIES DISTRICT OF KRAKOW

Streszczenie

Próby analizowania rynków nieruchomości klasycznymi metodami wymuszają zawężanie obszaru badań oraz standaryzację zmiennych, co między innymi ogranicza zasięg terytorialny. W artykule przedstawiono segmentację rynku nieruchomości położonych na terenach wybranych gmin miejsko-wiejskich, które celowo tak dobrano, aby nie stanowiły jednorodnego aglomeracji. W tym celu wybrano model drzew regresyjnych (C&RT), który pozwala łączyć atrybuty jakościowe i ilościowe, a co za tym idzie- włączyć cechy dotychczas pomijane, bez konieczności przypisywania im wartości liczbowych. W efekcie powstał schemat, który pozwala ocenić wartość ceny transakcyjnej nieruchomości gruntowej włączając takiej atrybuty, jak: „Gmina", czy „Miejscowość".

Słowa kluczowe: rynek nieruchomości, cena transakcyjna, drzewa klasyfikacyjne, nieruchomości gruntowe, gminy miejsko-wiejskie

Summary

Attempts to analyze real estate markets forcing classical methods, which narrow the area of research and standardization of variables, inter alia, limited territorial coverage. The paper presents a segmentation of the land immobility market at selected areas of urban-rural, that deliberately was chosen so, as not to create homogeneous cluster. To this end, the model regression trees (C&RT) was chosen, which allows you to combine qualitative and quantitative attributes, and thus-far overlooked features include, without assigning them values. The result

a scheme that allows to assess the value of real estate transaction price including attributes such as "Community" or "City" was prepared.

***Key words: t**he real estate market, the transaction price, classification trees, the real estate, urban-rural commune*

WSTĘP

Analizując rynki nieruchomości, dużo uwagi poświęca się terenom silnie zurbanizowanym, oczekując ich dalszej ekspansji. Niemniej jednak, taki rozwój nie spotyka się z aprobatą władz jednostek terytorialnych. Możliwość samodzielnego decydowania np. o sposobach rozdzielania środków finansowych, jest korzystniejsza od podporządkowania się władzą miast, przez które tereny te miałyby zostać pochłonięte. Formą pośrednią pomiędzy terenami rozwiniętymi gospodarczo, a terenami wiejskimi są gminy miejsko-wiejskie, czyli takie, na terenie których jedna z miejscowości posiada status miasta (Rozporządzenie Rady Ministrów z dnia 15 grudnia 1998 r.). To właśnie wokół niego skupia się aktywność mieszkańców gminy. W odróżnieniu od dużych aglomeracji, nie wyga to jednak migracji zarobkowej. Chcąc analizować preferencje nabywców nieruchomości gruntowych w wybranych gminach, wykorzystując np. współczynniki korelacji, należałoby przypisać wartości liczbowe lub w inny sposób uszeregować przede wszystkim nazwy miejscowości, co jest procesem trudnym oraz subiektywnym i tendencyjnym. Dlatego znacząca większość badań pomija te cechy, których nie można przekształcić do skali liczbowej. W opracowaniu zaproponowano podział rynku nieruchomości gruntowych wykonany przy pomocy drzew regresyjnych (C&RT), które pozwalają połączyć atrybuty jakościowe i ilościowe, a co za tym idzie - włączyć cechę „Miejscowość", czy „Gmina", bez przypisywania im wartości liczbowych [Jasińska 2009]

CHARAKTERYSTYKA MATERIALU BADAWCZEGO I ZASTOSOWANEGO ALGORYTMU

Za materiał badawczy posłużyły informacje o transakcjach sprzedaży nieruchomości gruntowych zawarte w Rejestrze Cen i Wartości, przeznaczonych na cele budownictwa jednorodzinnego dla wybranych gmin: Słomniki, Świątniki Górne, Skawina. Aby poszerzyć zebrane informacje przeprowadzono również wywiad terenowy. Łącznie zgromadzono dane o 147 sprzedanych nieruchomościach, które opisano wg kryteriów przedstawionych w tabeli 1. Zakres badań obejmuje transakcje z okresu styczeń - wrzesień 2010 r. Cechą wspólną zebranych danych jest miejsko-wiejski charakter gmin, w których się znajdują. Położenie geograficznie Słomnik pozwala podjąć próbę ustalenia cech uniwersalnych dla takiej struktury administracyjnej, bez obawy o wpływ czynników charakterystycznych dla gmin południowych, jakimi są Świątniki Górne i Skawina.

Segmentacja tak opisanego zbioru nie może zostać wykonana klasycznymi metodami, ponieważ wymagają one przekształcenia skal porządkowych na wartości liczbowe, co uniemożliwia włączenie atrybutów takich jak: „Miejscowość" i „Gmina", bez rozdzielenia analizowanej próby na mniejsze jednostki i powtórzenie badania w każdej z nich.

Tabela 1. Opis charakterystyk zbioru danych nieruchomości gruntowych
Table 1. Description of the characteristics of a data set of land immobility

Cecha	Kategoria/skala	Opis	Uwagi
Gmina	Słomki, Świątniki, Skawina,	Położenie w granicach administracyjnych jednej z gmin	Cecha wyrażona w skali jakościowej
Miejscowość	26 miejscowości	Położenie w granicach administracyjnych danej miejscowości	Cecha wyrażona w skali jakościowej
Powierzchnia	38 ÷ 8438	Powierzchnia nieruchomości wyrażona w [m^2]	Cecha wyrażona w skali ilościowej ciągłej
Położenie	Przeciętne	Lokalizacja działki w odniesieniu do ulicy, osiedla, dzielnicy.	Cecha wyrażona w skali porządkowej
	Modne		
	Bardzo modne		
	Najmodniejsze w okolicy		
Otoczenie	Przeciętne	Zagęszczenie budowy i wpływ otoczenia działki	Cecha wyrażona w skali porządkowej
	Korzystne		
	Bardzo Korzystne		
Kształt	Niekorzystny	Kształt nieruchomości pozwalający na posadowienie budynku mieszkalnego, wynikający z długości i szerokości działki	Cecha wyrażona w skali porządkowej
	Przeciętny		
	Korzystny		
Sieć wodociągowa	Brak	Posiadanie przyłącza sieci wodociągowej na działce	Cecha o charakterze binarnym
	Jest		
Sieć elektryczna	Brak	Posiadanie przyłącza sieci elektrycznej na działce	Cecha o charakterze binarnym, ponieważ nie odnotowano transakcji ją różnicującą, została pominięta w analizach
	Jest		
Sieć gazowa	Brak	Posiadanie przyłącza sieci gazowej na działce	Cecha o charakterze binarnym
	Jest		
Sieć kanalizacyjna	Brak	Posiadanie przyłącza sieci kanalizacyjnej na działce	Cecha o charakterze binarnym
	Jest		
Droga	Brak	Możliwość dojazdu do nieruchomości drogą utwardzoną	Cecha o charakterze binarnym
	Jest		
Cena jednostkowa	0,67 ÷ 202,02	Cena za 1[m^2] sprzedanej działki	Cecha o charakterze ilościowym ciągłym

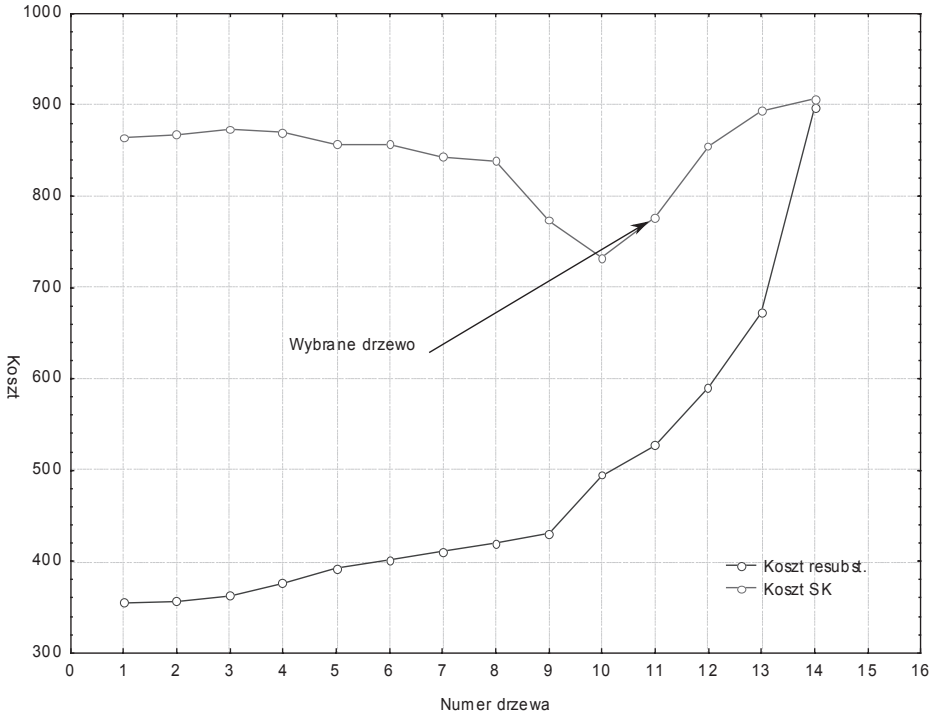
Źródło: Opracowanie własne.
Source: Own calculations

Metoda C&RT (*Classification and Regression Tree*) umożliwia włączenie cech jakościowych do badań, co w tym przypadku ogranicza się do predyktorów, które pozostawiono w postaci skali porządkowej, bez przypisywania im wartości liczbowych. Budowa drzewa wymaga rekurencyjnego podziału zbioru obserwacji na rozłączne podzbiory, przy czym poziom pomiaru zmiennej zależnej (wartości nieruchomości) jest co najmniej przedziałowy, a jej wariancja w węzłach końcowych zmierza do zera. Wspomniany proces polega na rekurencyjnym podziale obiektów na dwa podzbiory, przy czy każdy z nich opiera się na wartościach jednej zmiennej [Hand i in, 2001]. Istotą tej metody jest dokonanie takich podziałów, które pozwolą uzyskać największe polepszenie predykcji, dlatego zbiór danych w sposób jednoznaczny rozdzielany jest na kolejne podzbiory, co przedstawione jest na diagramie w postaci ścieżki od korzenia do liścia. Znalezienie kompromisu pomiędzy wielkością drzewa, a możliwie najmniejszym przyroście błędu predykcji w drzewach regresyjnych, jest możliwe dzięki zasadzie jednego odchylenia standardowego (*1 SE rule*). Autorzy algorytmu C&RT proponują, aby jako drzewo właściwej wielkości wybrać to, którego koszty sprawdzianu krzyżowego nie przekraczają minimalnych kosztów sprawdzianu krzyżowego powiększonych o wartość odchylenia standardowego tych kosztów. W konsekwencji tworzenia modelu usystematyzowaniu podlegają wszystkie predykatory charakteryzujące badane zjawisko, uwzględniając moc predykcyjną względem zmiennej zależnej. Ważność predyktorów wyraża się w skali 0-1 (przeskalowana tak, aby jej maksimum wynosiło 1), co stanowić może analogię do współczynników korelacji, jednak w rankingu ważności nie można ustalić czy cecha korzystnie, czy negatywnie wpływa na wartość zmiennej objaśnianej. Kontrowersje może budzić fakt, że zmienna, która nie wystąpiła, jako kryterium ostatecznego podziału drzewa, mimo to uzyska wysoką lokatę w rankingu. Możliwe jest to, gdy dany atrybut podczas większości podziałów, był drugim, co do możliwości zmniejszania wariancji w węzłach-potomkach [Jasińska E., 2009]. Pomimo, że ostatecznie nie został umieszczony na grafie, jego zdolność predykcyjna jest wyższa od takiego atrybutu, który „wykorzystał moc predykcyjną" w pierwszych podziałach drzewa, a następnie nie był istotny dla dalszej segmentacji.

OTRZYMANE WYNIKI

Pierwszym etapem badań było utworzenie drzewa maksymalnych wymiarów, które składało się z 20 węzłów dzielonych, przy 21 końcowych. Taki schemat powala dokładnie opisać wzajemne relacje pomiędzy cechami opisującymi sprzedane nieruchomości, jednak jest zbyt szczegółowy. Dlatego w celu wyłonienia najlepszego schematu posłużono się sekwencją kosztów zastąpienia (resubstytucji) i kosztów sprawdzianu krzyżowego, co przedstawiono na rys. 1. Wykres ten pozwala porównać ze sobą koszt sprawdzianu krzyżowego (SK)

oraz koszt resubstytucji dla każdego z utworzonych modeli. Koszt zastąpienia rośnie wraz z przycinaniem drzewa, natomiast wartość kosztu SK wyraźnie przemawia za wybranym modelem.



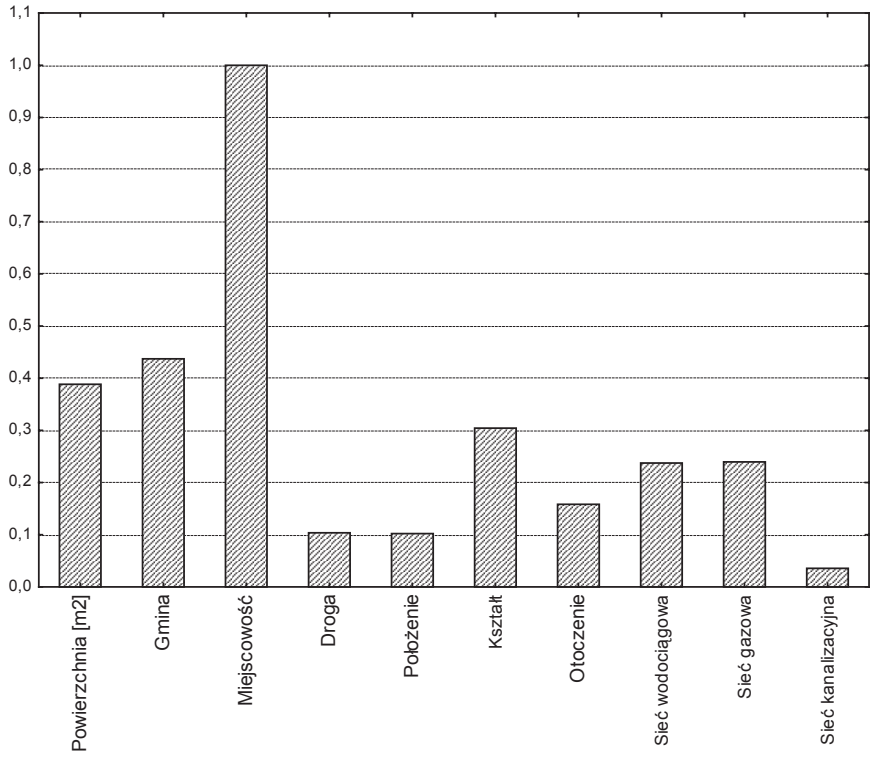
Źródło: Opracowanie własne
Source: Own calculations

Rysunek 1. Sekwencja kosztów drzew regresyjnych
Figure 1. Sequence of the cost of regression trees

Wybrane drzewo nr 11 spełnia zasadę 1 SE (autorzy algorytmu C&RT proponują, aby jako drzewo właściwej wielkości wybrać to, którego koszty sprawdzianu krzyżowego nie przekraczają minimalnych kosztów sprawdzianu krzyżowego powiększonych o wartość odchylenia standardowego tych kosztów).

Przedstawiony na rys. 2, wykres ważności cech objaśniających, umożliwia uszeregować atrybuty według ich możliwości predykcyjnych. Cecha najtrudniejsza do wyrażenia w sposób ilościowy, stała się najistotniejszą w procesie segmentacji rynku, niezależnie od gminy, w jakiej nieruchomość się znajduje. Kolejne pozycje zajęły: „Powierzchnia", „Gmina" oraz „Kształt działki". Wysoka lokata przynależności nieruchomości do jednej z gmin, potwierdza słuszność

wybranej metody, skoro ta cecha jest istotna dla kształtowania się średniej jednostkowej ceny nieruchomości, to pominięcie jej mogłoby przynieść zafałszowane wyniki. Kolejne pozycje w rankingowaniu predyktorów zajęły przyłącza sieci gazowej i wodociągowej, jako niezbędne dla gospodarstw domowych i co ważne, ich doprowadzenie wymaga znaczących środków finansowych. Niska ranga pozostałych cech sugeruje powszechność ich występowania.

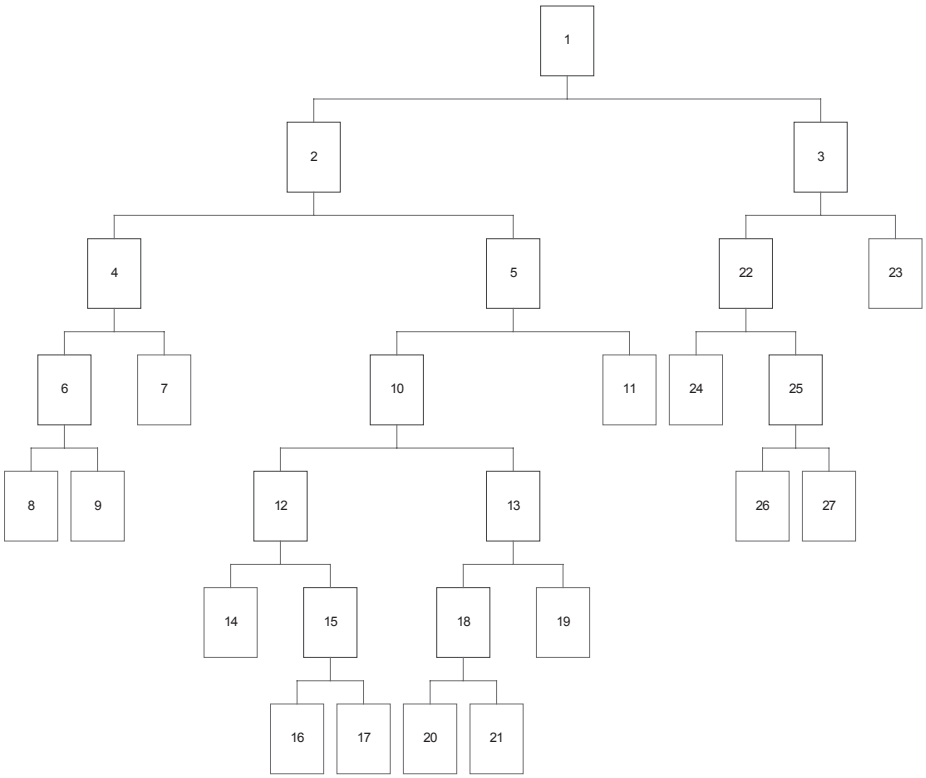


Źródło: Opracowanie własne
Source: Own calculations

Rysunek 2. Ranking predyktorów wykonany w oparciu o ich wpływu na kształtowanie się ceny 1 m^2 nieruchomości gruntowej
Figure 2. Predictors ranking made on the basis of their impact on the price of 1 m^2 of land

Skoro już możliwe jest usystematyzowanie wpływu poszczególnych cech, kolejnym krokiem było wyznaczenie miejscowości, które w badanym okresie były najatrakcyjniejsze dla nabywców oraz sprawdzenie wpływu pozostałych cech na kształtowanie się średniej ceny jednostkowej nieruchomości gruntowej. W tym celu przedstawiono układ wybranego drzewa regresyjnego, które cechuje

najlepsza zdolność predykcyjna oraz najmniejsze wymiary, jego schemat przedstawia rys 3.



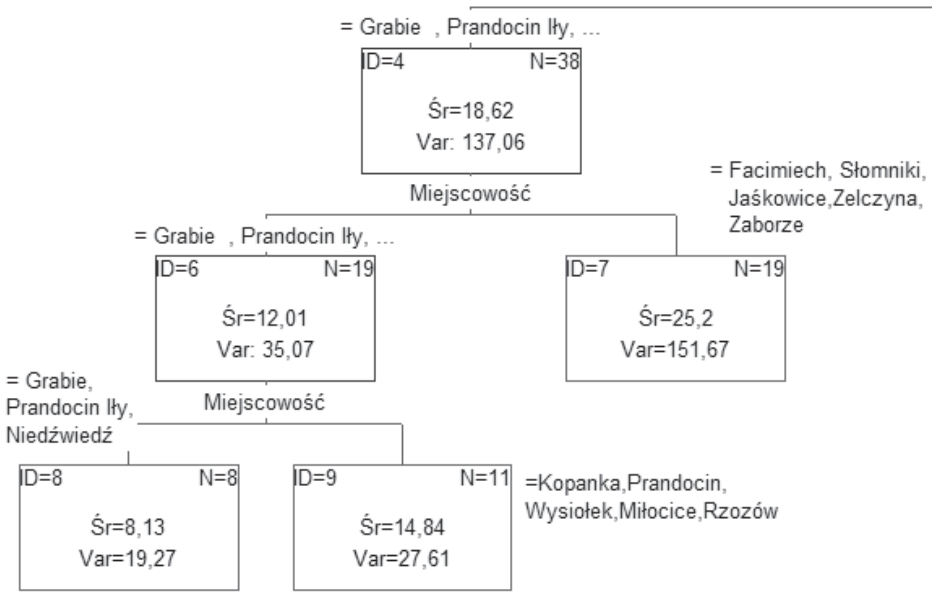
Źródło: Opracowanie własne.
Source: Own calculations.

Rysunek 3. Układ utworzonego drzewa regresyjnego
Figure 3. Diagram of created regression trees

Wadą używanego modelu, jest fakt, że każdy węzeł dzielony jest tylko na dwa podzbiory, co w efekcie wydłuża cały diagram, dlatego omówione zostaną wybrane jego fragmenty. Gałąź zakończona węzłami 8-9-7 pozwoliła oddzielić te nieruchomości, dla których jedynym kryterium kształtowania się ceny 1m^2 było położenie geograficzne. Wykres przedstawiony na rys. 4 zapisać można by w postaci następujących zdań warunkowych:

– jeżeli nieruchomość gruntowa położona jest w miejscowościach: Grabie (gmina Skawina), Prandocin Iły (gmina Słomniki) lub Niedźwiedź (gmina Słomniki) to średnia cena 1m^2 takiej nieruchomości wynosi 8,13 zł ± 4,39zł (węzeł nr 8)

– jeżeli nieruchomość gruntowa położona jest w miejscowościach: Facimiech, Jaśkowice, Zelczyna (gm. Skawina), lub: Słomniki, Zaborze (gm. Słomniki), to średnia cena 1m^2 takiej nieruchomości wynosi 25,20 zł ± 12,31zł



Źródło: Opracowanie własne
Source: Own calculations

Rysunek 4. Fragment utworzonego modelu, węzły 7-8-9
Figure 4. Piece of formed model, nodes 7-8-9

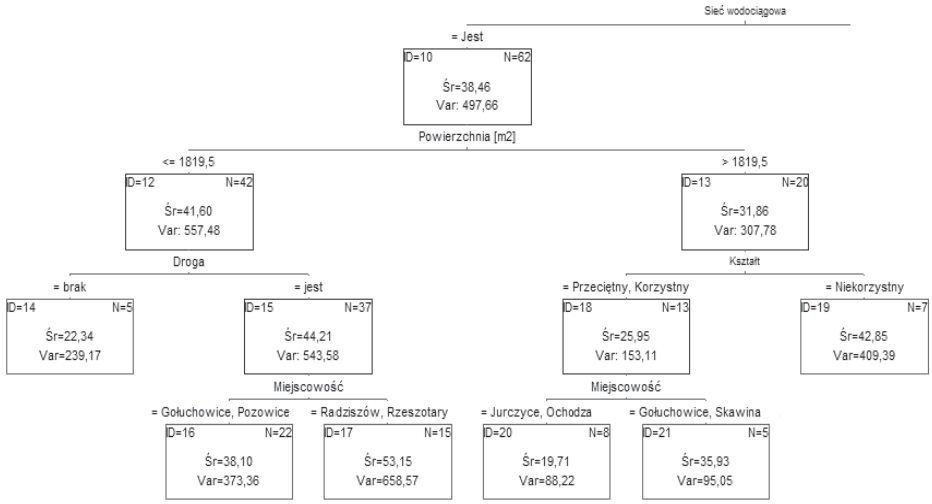
Analizując kolejne gałęzie możemy w jasny sposób ustalić kryteria podziału , pozwalające na zmniejszenie wariacji w kolejnych podzbiorach. Taki model uwzględnia zarówno cechy o charakterze nominalnym, jak porządkowym, czy liniowym. Sama forma schematu drzewkowego jest bardziej klarowna niż konstruowanie kolejnych zdań warunkowych. Na Rys. 5 przedstawiono kolejną gałąź schematu, dla nieruchomości gruntowych wyposażonych w przyłącze sieci wodociągowej, w miejscowościach: Pozowice, Jurczyce,

Gołuchowice, Ochodza, Skawina, Radziszów oraz Rzeszotary, w zależności od powierzchni, kształtu oraz utwardzonej drogi dojazdowej, prowadzącej do nieruchomości. Wśród nich znajdują się dwa rodzaje nieruchomości, warte dodatkowego komentarza:

– 5 nieruchomości gruntowych, o powierzchni mniejszej niż 18 arów, do których nie prowadzi utwardzona droga dojazdowa (co oznaczać może albo drogę gruntową, albo służebność przejazdu) - średnia 1 m^2 wynosiła 22 zł ± 15zł,

- 7 nieruchomości, dla których cena transakcyjna 1m^2 kształtuje się na poziomie 44zł ± 20zł, pomimo niekorzystnego kształtu działki.

Oba te przypadki zostały zgrupowane w osobnych węzłach, więc nie zaburzają pozostałych analiz.



Źródło: Opracowanie własne
Source: Own calculations

Rysunek 5. Fragment utworzonego modelu, węzły od 10 do 21
Figure 5. Piece of formed model, nodes from 10 to 21

PODSUMOWANIE

Wykorzystywanie drzew regresyjnych w analizie rynku nieruchomości przynosi wiele korzyści. Umożliwia włączenie cech jakościowych, takich jak „Gmina", czy „Miejscowość", a przez to pozwala na porównywanie rynków geograficznie oddalonych od siebie, ale posiadających cechy wspólne, jak w analizowanym przypadku gmin miejsko-wiejskich. Wynikiem analizy są jasne, przejrzyste reguły podziału próby, dzięki którym można wnioskować o kształtowaniu się cen transakcyjnych dla innych nieruchomości, nieujętych w badanym zbiorze. W tym przypadku możliwe jest zaprezentowanie potencjalnemu kupującemu np. nieruchomości położonych w innych obszarach, a jednocześnie spełniających wymogi, co do preferencji kupującego. Godnym zauważenia jest fakt tworzenia skupisk łączących transakcje o wątpliwych relacjach wartości atrybutów do ceny transakcyjnej, które jako obserwacje odstające mogłyby zanieczyszczając pozostałe węzły, obniżając zdolność predykcyjną całego modelu.

REFERENCES

[1] Hand D., Mannila H., Smyth P. Principles of Data Mining, MIT Press, 2001.

[2] Jasińska E. Ocena przydatności wybranych metod statystycznych w analizie rynku nieruchomości. Rozprawa doktorska, AGH, Kraków, 2009.

[3] Rozporządzenie rady ministrów z dnia 15 grudnia 1998 r. W sprawie szczegółowych zasad prowadzenia, stosowania i udostępniania krajowego rejestru urzędowego podziału terytorialnego kraju oraz związanych z tym obowiązków organów administracji rządowej i jednostek samorządu terytorialnego.

Artykuł powstał w ramach badań statutowych Katedry Geomatyki

Dr inż. Mariusz Frukacz
Dr inż. Elżbieta Jasińska
Dr hab. inż. Edward Preweda, prof. AGH
Akademia Górniczo – Hutnicza im. St. Staszica w Krakowie
Wydział Geodezji Górniczej i Inżynierii Środowiska
Katedra Geomatyki
30-059 Kraków, al. A. Mickiewicza 30, paw. C-4, pokój 406
e-mail: frukacz@agh.edu.pl
jasinska@agh.edu.pl
preweda@agh.edu.pl

Recenzent: *Prof. dr hab. inż. Urszula Litwin*