

MARIUSZ DACKO¹

SIEĆ NATURA 2000 A STAN LOKALNEJ GOSPODARKI OBSZARÓW WIEJSKICH ZIELONYCH PŁUC POLSKI – UJĘCIE MODELOWE

Abstrakt. W opracowaniu przedstawiono modele drzew regresyjnych objaśniających kształtowanie się stanu sektora rolniczego oraz aktywności ekonomicznej mieszkańców obszarów wiejskich Zielonych Płuc Polski. Prezentowane modele stanowiły próbę empirycznej weryfikacji pesymistycznych poglądów o stanie lokalnej gospodarki gmin szeroko objętych specjalną ochroną ptaków i siedlisk w ramach sieci Natura 2000 na terenach ZPP. Wyniki modeli przedstawiały proste kryteria wyodrębniania gmin o zróżnicowanym poziomie rozwoju aktywności ekonomicznej i rolnictwa. Umożliwiły także ocenę ważności wykorzystanych predyktorów, które stanowiły wyróżnione w projekcie badawczym subkomponenty rozwoju zrównoważonego. Wyniki przewidywań modeli drzew regresyjnych przedstawiono na mapach, co umożliwiło ocenę przestrzennego kształtowania się lokalnej gospodarki na obszarach ZPP.

Słowa kluczowe: Natura 2000, rolnictwo, aktywność ekonomiczna, drzewa regresyjne

WPROWADZENIE

Podstawę delimitacji obszaru funkcjonalnego Zielonych Płuc Polski (ZPP) stanowiły granice systemów ekologicznych ocenianych jako jedne z najcenniejszych w kraju i Europie. Dlatego też niemal 2/3 z 341 zlokalizowanych tu gmin wiejskich i obszarów wiejskich gmin miejsko-wiejskich objęto obszarami specjalnej ochrony ptaków i siedlisk w ramach sieci Natura 2000. Poddając obserwacji skalę działań ochronnych, można stwierdzić, że ochrona ptaków i siedlisk w zdecydowanej większości gmin „naturowych” obejmuje niewielką ich część i nie przekracza 20% ich powierzchni. Omawianą formą ochrony objęto w spo-

¹ Autor jest pracownikiem naukowym Uniwersytetu Rolniczego w Krakowie (e-mail: m.dacko@ur.krakow.pl).

sób znaczny jedynie około 10% gmin ZPP (tereny sieci Natura 2000 stanowią tu powyżej 60% ich powierzchni ogółem).

Ze zgromadzonych w ramach realizacji projektu danych wynikało, że w przypadku 14 jednostek samorządowych, należących do obszaru ZPP, sieć Natura 2000 zajmuje powyżej 80% ich powierzchni. Wyniki pogłębionych wywiadów przeprowadzonych w tych gminach wskazywały, że wójtowie i przedstawiciele lokalnych społeczności oceniają tę sieć ekologiczną jako źródło licznych problemów o charakterze inwestycyjnym, ogólnorozwojowym oraz informacyjnym. Negatywnego wizerunku zwykle dopełniały sygnalizowane problemy gospodarki rolnej oraz utrudnienia, na jakie regularnie napotykali lokalni przedsiębiorcy. Słusznie stwierdzają Kłodziński i Bołtmiuk [2009], że tych negatywnych odczuć i ocen nie można pomijać milczeniem, nawet jeśli w jakimś stopniu różniły się one z rzeczywistością.

W niniejszym opracowaniu podjęto próbę empirycznej weryfikacji poglądów o negatywnym związku obszarów sieci Natura 2000 z lokalnym stanem gospodarki i rolnictwa.

W tym celu zbudowano modele drzew regresyjnych, poddając szczególnej obserwacji wyniki, jakie uzyskano dla gmin w sposób znaczny objętych tą siecią. Modele te umożliwiły rozpoznanie i syntetyczny opis podstawowych współzależności, występujących między aktywnością ekonomiczną² i rozwojem sektora rolnego³ a pozostałymi wyróżnionymi w projekcie subkomponentami rozwoju⁴. W pracy postawiono dwa pytania badawcze:

1. Czy tereny wiejskie ZPP objęte specjalną ochroną ptaków i siedlisk w ramach sieci Natura 2000 charakteryzuje ograniczona aktywność ekonomiczna mieszkańców?

2. Czy tereny wiejskie ZPP objęte specjalną ochroną ptaków i siedlisk w ramach sieci Natura 2000 wyróżnia niski poziom rozwoju rolnictwa?

METODA BADAWCZA

Drzewa regresyjne zastosowano do zaklasyfikowania gmin ZPP do grupy jednostek o zbliżonych stanach ocenianego subkomponentu rozwoju. W ramach przeprowadzonych badań dokonano wyboru dwóch istotnych subkomponentów

² Na zmienną tę składały się cztery wskaźniki zagregowane metodą sum standaryzowanych: bezrobocie rejestrowane na 100 osób w wieku produkcyjnym (w%) w 2006 roku; pracujący wyłącznie lub głównie w rolnictwie w wieku produkcyjnym na 100 ha użytków rolnych (w gospodarstwach indywidualnych) w 2002 roku; pracujący poza rolnictwem na 100 osób pracujących w rolnictwie w 2002 roku; wskaźnik zatrudnienia w wieku produkcyjnym (w%) w 2002 roku.

³ Na zmienną tę składało się pięć wskaźników zagregowanych tą samą metodą: odsetek gospodarstw indywidualnych powyżej 1 ha produkujących głównie na rynek w 2002 roku; odsetek gospodarstw kierowanych przez osobę z wykształceniem rolniczym średnim, policealnym i wyższym rolniczym w 2002 roku; średni obszar gospodarstwa indywidualnego powyżej 1 ha użytków rolnych (w ha) w 2002 roku; odsetek indywidualnych gospodarstw rolnych prowadzących działalność pozarolniczą w 2002 roku; liczba wspieranych gospodarstw rolnych z działań PROW (II + III + IX) na 100 gospodarstw rolnych powyżej 1 ha w latach 2004–2006.

⁴ Były to: atrakcyjność środowiska, presja na środowisko i jego ochrona, sektor pozarolniczy, finanse gmin, demografia, edukacja, warunki życia oraz aktywność społeczna i samorządowa.

rozwoju gospodarczego. Były to: aktywność ekonomiczna i sektor rolniczy. Zagregowane zmienne, składające się na te subkomponenty, najściślej wiązały się ze spektrum problemów sygnalizowanych przez lokalne społeczności gmin „naturalnych” [Boltromiuk i Kłodziński 2009]. Wykorzystując program STATISTICA wyposażony w moduł Data Mining, zbudowano modele drzew regresyjnych obu tych subkomponentów⁵.

Drzewa klasyfikacyjne i regresyjne C&RT (Classification and Regression Trees) są zaliczane do eksploracyjnych metod analizy danych. Zostały one rozpropagowane w latach osiemdziesiątych przez Breimana i innych [1993]. Stanowią narzędzie analityczne, umożliwiające budowę modeli predykcyjnych lub deskryptywnych [Łapczyński 2003]. Modele predykcyjne umożliwiają klasyfikację lub predykcję ilościową zjawiska, modele deskryptywne zaś opis i prezentację wzorców odkrytych w analizowanej zbiorowości. W niniejszych badaniach wykorzystano obie funkcje modeli drzew.

Istotą działania algorytmu C&RT jest poszukiwanie zbioru logicznych warunków podziału typu „jeżeli – to”⁶. Z literatury przedmiotu wynika, że w naukach ekonomicznych drzewa C&RT znajdują coraz szersze zastosowanie, dlatego właśnie z ich wykorzystaniem oceniano wpływ aktywności mieszkańców na poziom życia w gminach [Łapczyński 2007]. Drzewa te stosowano też w ocenie czynników rozwoju przodujących gospodarstw rolniczych [Sroka i Dacko 2010] oraz w badaniach popytu na lokalnych rynkach nieruchomości [Dacko 2007]. Szczególny rozgłos i zainteresowanie międzynarodowe zdobyli ekonomiści P. Manasse i N. Roubini, którzy stosując drzewo oparte na algorytmie C&RT, przedstawili prosty i przekonujący sposób oceny ryzyka upadłości i niewypłacalności państw w dobie obecnego kryzysu globalnego [Manasse i Roubini 2009].

W metodzie drzew regresyjnych podstawowym narzędziem analizy jest diagram, umożliwiający czytelny (w porównaniu do innych narzędzi badawczych, takich jak: regresja, modele strukturalne czy sieci neuronowe) wgląd w wyniki. Istotne zalety drzew regresyjnych, jako metody badawczej, stanowią:

- nieparametryczność i nieliniowość, które sprawiają, że nie trzeba dokonywać założeń co do natury związku między zmiennymi,
- łatwość analizowania logicznych warunków podziału drzewa,
- uwzględnianie zależności niemonotonicznych poprzez kolejne podziały względem tej samej zmiennej w sytuacji, gdy zmienna objaśniająca jest dodatnio skorelowana z objaśnianą przy jej mniejszych wartościach, ujemnie zaś, gdy przekroczy ona pewien poziom krytyczny,

⁵ Dany subkomponent traktowano jako zmienną zależną objaśnianą zestawem pozostałych subkomponentów za pomocą modeli drzew regresyjnych. Były one oparte na algorytmie C&RT, polegającym na poszukiwaniu podziałów jednowymiarowych.

⁶ Drzewo regresyjne powstaje w wyniku rekurencyjnych podziałów, których istotą jest poszukiwanie najlepszego predyktora przyczyniającego się do różnicowania stanu badanej zmiennej. Kolejne podzbiory o zróżnicowanych wartościach modelowanej zmiennej są wyodrębniane tak, aby jej wariancja wewnątrz każdego podzbioru była możliwie najmniejsza. Podzbiory te tworzą strukturę hierarchiczną i są nazywane gałęziami (węzłami dzielonymi), jeśli ulegają dalszym podziałom, bądź liśćmi (węzłami końcowymi), gdy podziały na nich się kończą.

- przydatność do zadań, gdzie wiedza a priori o tym, które zmienne są ze sobą powiązane i w jaki sposób, jest niepewna i jedynie intuicyjna,
- duża przydatność w zagadnieniach predykcyjnych,
- możliwości oszacowania i utworzenia rankingu ważności predyktorów (w tym przypadku pozostałych subkomponentów) w procesie kształtowania się wartości badanego subkomponentu.

Przymioty te sprawiają, że drzewa regresyjne w badaniach zależności między zmiennymi dają z reguły znacznie lepsze rezultaty niż ekonometryczne modele regresji.

W drzewach regresyjnych C&RT szczególnie ważny jest wybór parametrów oceny jakości modeli oraz parametrów pozwalających zapobiec ich nadmierne- mu rozrostowi [Sokołowski 2002]. Kontrolę jakości uzyskiwanych wyników przeprowadzono przy zastosowaniu V -krotnego sprawdzianu krzyżowego⁷. Przerwanie procesu tworzenia nowych węzłów odbywało się zaś każdorazowo przy zastosowaniu przycinania według wariancji oraz kryterium minimalnej liczności obserwacji w węźle dzielonym ($n \geq 30$).

Wybór drzewa o optymalnej strukturze jest dokonywany na podstawie oceny kosztów sprawdzianu krzyżowego oraz stopy błędnych klasyfikacji w próbie uczącej, czyli tzw. kosztów resubstytucji. Dlatego obie te kategorie zestawiane są na tzw. wykresach sekwencji kosztów. Dla ułatwienia, optymalne drzewo regresyjne jest wskazywane automatycznie użytkownikowi przy zastosowaniu zasady jednego odchylenia standardowego. Jest to drzewo o możliwie najmniejszej liczbie węzłów końcowych. Koszty sprawdzianu krzyżowego oszacowane dla tego drzewa nie powinny różnić się od kosztów tego sprawdzianu w całej sekwencji drzew o więcej niż jedno odchylenie standardowe. W przypadku modelu aktywności ekonomicznej dokonano wyboru drzewa innego niż sugerowane przez STATISTICA (nieznacznie bardziej skomplikowanego). Wybór taki poprzedziła analiza wykresu sekwencji kosztów, gdyż w budowie drzew regresyjnych szczególnie istotne jest zachowanie kompromisu między potrzebą budowy możliwie prostych modeli a ich dokładnością mierzoną kosztami sprawdzianu krzyżowego i resubstytucji.

WYNIKI BADAŃ

Zbudowane modele drzew wyjaśniały zróżnicowanie stanu sektora rolniczego oraz aktywności ekonomicznej mieszkańców poszczególnych gmin terenu Zielonych Płuc Polski. W procesie modelowania obu tych subkomponentów wygenerowana została sekwencja drzew o różnym stopniu skomplikowania. Infor-

⁷ V -krotny sprawdzian krzyżowy polega na wyodrębnieniu z próby uczącej wielu podprób losowych (tutaj dziesięciu). Drzewa regresyjne określonej wielkości były obliczane $V = 10$ razy, przy czym kolejno opuszczano w obliczeniach jedną z podprób i wykorzystywano jako próbę testową w sprawdzianie krzyżowym. Każda podpróba była zatem użyta $V - 1 = 9$ razy w próbie uczącej i jeden raz w charakterze próby testowej. Koszty sprawdzianu krzyżowego obliczone dla prób testowych są następnie uśredniane w celu obliczenia V -krotnej oceny kosztów sprawdzianu krzyżowego dla kolejnych drzew o złożoności ograniczonej minimalną liczbą obserwacji w węźle dzielonym ($n \geq 30$).

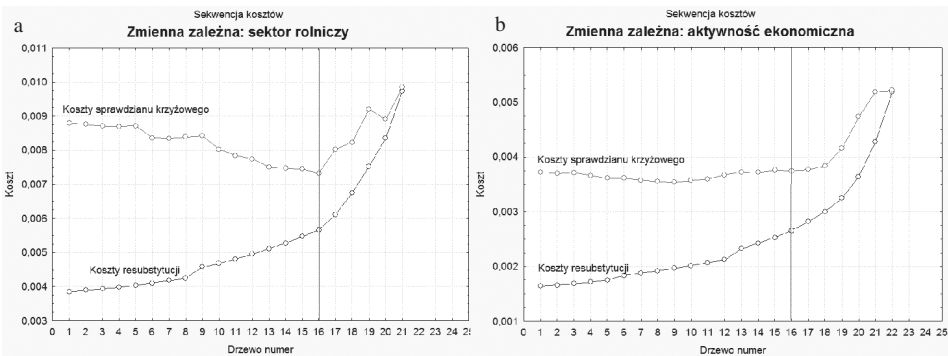
TABELA 1. Sekwencja drzew i parametry optymalnych modeli drzew subkomponentów

Komponent	Subkomponent	L_d	D_o	S_s	W_k
Gospodarka	aktywność ekonomiczna	22	16	6	0,70
	sektor rolniczy	21	16	5	0,65

Oznaczenia: L_d – liczba wygenerowanych drzew regresyjnych, D_o – numer drzewa regresyjnego uznanego za optymalne, S_s – stopień skomplikowania drzewa optymalnego, W_k – współczynnik korelacji Pearsona między faktycznymi wartościami subkomponentu a wartościami oszacowanymi przez drzewo optymalne.

macje zawarte w tabeli 1 pozwalają ocenić stopień skomplikowania i jakość drzew optymalnych.

Z wykresów sekwencji kosztów (rysunek 1) wynikało, że w przypadku obu subkomponentów za optymalne można było uznać drzewa regresyjne o numerze 16. Drzewa te wskazano na wykresach ciągłymi pionowymi liniami. Wraz z dalszym rozrostem drzew (na lewo od pionowych linii) koszty sprawdzianu krzyżowego zmieniły się nieznacznie (w modelu aktywności ekonomicznej) bądź wyraźnie rosły (w modelu sektora rolniczego). Pożądany dynamiczny spadek owych kosztów można było zaobserwować jedynie po prawej stronie pionowych linii i kończył się on właśnie na drzewach optymalnych.



RYSUNEK 1. Sekwencja kosztów modeli: a – sektora rolniczego, b – aktywności ekonomicznej

W zapisie stosowanym przez program STATISTICA drzewo o najwyższym numerze jest drzewem najprostszym. Składa się ono z jednej tylko gałęzi, w której obliczona zostaje średnia i wariancja modelowanego subkomponentu. Nie odbywają się tu zatem żadne podziały i nie są typowane predyktory. Nie jest ono w związku z tym interpretowane.

W najbardziej skomplikowanych drzewach (oznaczonych nr 1) występowało kilkanaście podziałów hierarchicznych. Koszty sprawdzianu krzyżowego wskazywały, że drzewa te mimo swej złożoności miałyby podobną (w przypadku aktywności ekonomicznej) bądź wyraźnie mniejszą (w przypadku sektora rolniczego) zdolność przewidywania niż drzewa optymalne (rysunek 1).

Wyniki zaprezentowane na rysunku 1b w szczególności potwierdzały pogląd, że proces podziałów powinien być w pewnym momencie przerwany, ponieważ postępującemu skomplikowaniu drzew regresyjnych może towarzyszyć niepożądany wzrost błędów w próbie testowej, określanymi kosztami sprawdzianu krzyżowego⁸.

Optymalne drzewa regresyjne, stanowiące podstawę dalszych badań, charakteryzował umiarkowany stopień skomplikowania. Współczynniki korelacji Pearsona, jakie obliczono między faktycznymi wartościami subkomponentów a przewidywaniami zbudowanych modeli, były zadowalające i osiągały wartości 0,70 i 0,65 (tabela 1). Na tej podstawie uznano, że oba modele charakteryzuje stosunkowo dobra zdolność predykcyjna i poddano szczegółowej interpretacji ich wyniki.

MODEL AKTYWNOŚCI EKONOMICZNEJ MIESZKAŃCÓW

Na obszarach wiejskich Zielonych Płuc Polski wartości subkomponentu aktywność ekonomiczna osiągały średnią arytmetyczną równą 0,46 i zawierały się w przedziale od 0,27 do 0,76. Ich rozkład był jednomodalny, symetryczny i zbliżony do normalnego. Wartości skrajne dawały rozstęp równy 0,49. Był on stosunkowo niewielki na tle pozostałych 10 wyróżnionych w projekcie subkomponentów rozwoju zrównoważonego.

Z diagramu drzewa wynikało, że z aktywnością ekonomiczną obszarów wiejskich ZPP w szczególności wiązały się warunki życia⁹ (podział gałęzi 1 i 23) oraz uwarunkowania demograficzne¹⁰ (podział gałęzi 2, 3 i 20) – rysunek 2. Czynniki te wystąpiły w podziałach drzewa odpowiednio dwu- i trzykrotnie. Istotne znaczenie miał także poziom rozwoju sektora pozarolniczego¹¹ (podział gałęzi 5). Wymienione subkomponenty zajęły najwyższą pozycję w rankingu ważności użytych zmiennych objaśniających.

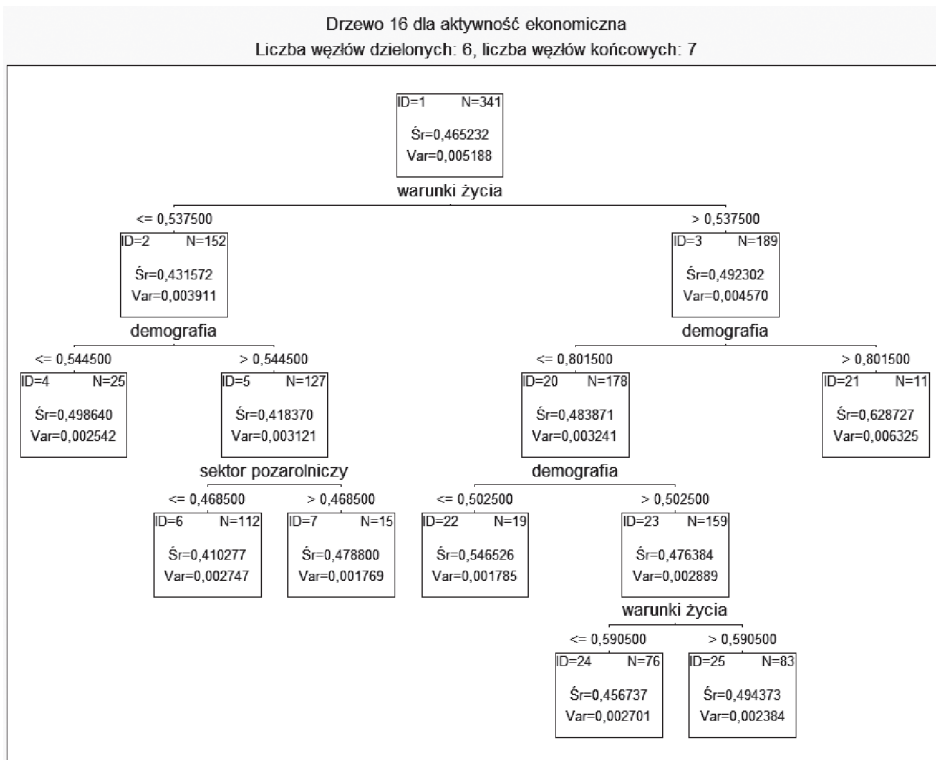
Przy siedmiu węzłach końcowych omawiany model w wysokim stopniu wyjaśnił zróżnicowanie badanego subkomponentu. Współczynnik korelacji Pearsona, obliczony między faktycznymi wartościami aktywności ekonomicznej w poszczególnych gminach a odpowiadającymi im przewidywaniami drzewa, wyniósł bowiem 0,70.

Jak wynika z pierwszego podziału drzewa, niższy poziom aktywności ekonomicznej mieszkańców ($\bar{S}r = 0,43$) odnotowano w 152 jednostkach samorządu terytorialnego o niższej ocenie warunków życia (ID 2). Tutaj badana zmienna pod-

⁸ Zagadnienie to jest szeroko komentowane w literaturze przedmiotu i zwane przeuczeniem lub nadmiernym dopasowaniem. Zgodnie z ogólnym podejściem do owego problemu zaleca się, by przerwać proces tworzenia nowych węzłów z podziałami, gdy kolejne dają niewielki przyrost trafności przewidywań drzewa.

⁹ Na zmienną tę składało się pięć wskaźników zagregowanych metodą sum standaryzowanych: odsetek mieszkań zamieszkałych stale w budynkach wybudowanych w latach 1989–2002 lub będących w budowie w ogólnej liczbie mieszkań zamieszkałych w 2002 roku; wydatki budżetu gminy na zasiłki i pomoc w naturze oraz składki na ubezpieczenia emerytalne i rentowe na 1 mieszkańca w 2006 roku (średnia 3-letnia); odsetek ludności utrzymującej się z niezarobkowych źródeł utrzymania w 2002 roku; zagęszczenie mieszkań (powierzchnia użytkowa mieszkań w m² na osobę) w 2006 roku; odsetek mieszkań wyposażonych w wodociąg z sieci w 2006 roku.

¹⁰ Na zmienną tę składało się pięć wskaźników zagregowanych tą samą metodą: odsetek ludności w wieku produkcyjnym w 2006 roku; wskaźnik obciążenia demograficznego w 2006 roku (stosunek ludności poprodukcyjnej do przedprodukcyjnej); współczynnik feminizacji (kobiety na 100 mężczyzn w wieku rozrodczym (20–34 lata) w 2006 roku; współczynnik atrakcyjności migracyjnej dla migracji wewnętrznych w 2006 roku (średnia 3-letnia); przyrost naturalny na 1000 mieszkańców w 2006 roku (średnia 3-letnia).



RYSUNEK 2. Diagram drzewa regresyjnego przedstawiający model aktywności ekonomicznej

legała dalszemu zróżnicowaniu według stanu demografii (ID 4 i 5), a także rozwoju sektora pozarolniczego (ID 6 i 7).

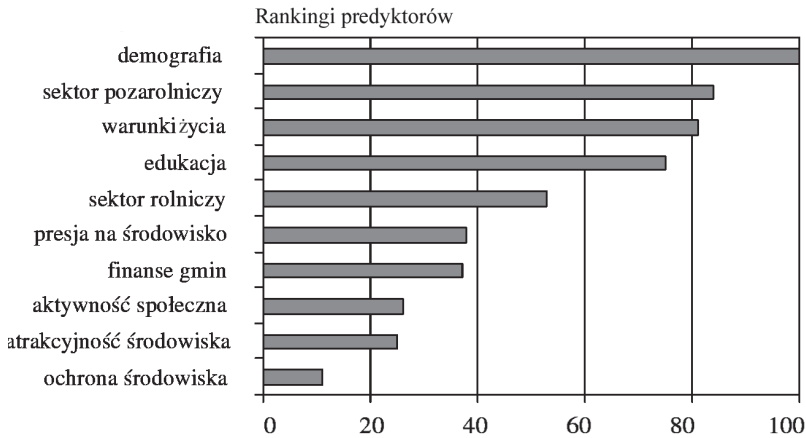
Diagram drzewa wskazywał, że w dużej grupie 189 gmin (ID 3) ocena warunków życia była wyższa od średniej i wynosiła 0,54. Tutaj poziom aktywności ekonomicznej również był wyższy i osiągał średnią wartość 0,49. Różnicował ją dalej stan demografii (ID 20 i 21), co m.in. pozwoliło wyróżnić grupę 178 gmin o niższej ocenie potencjału demograficznego, gdzie średnia aktywność ekonomiczna osiągnęła poziom 0,48.

W omawianym modelu uwarunkowania demograficzne były predyktorem istotnym, lecz powiązaniem z aktywnością ekonomiczną w sposób niejednoznaczny. W przypadku podziału gałęzi 2 i 20 niższym ocenom stanu demografii towarzyszył wyższy poziom aktywności ekonomicznej. W podziale gałęzi 3 zależność ta uległa odwróceniu: w grupie 11 gmin o korzystnych warunkach życia (> 0,54) i wysokim potencjale demograficznym (> 0,80) odnotowano najwyższy

¹¹ Stanowiły go zagregowane metodą sum standaryzowanych wskaźniki: odsetek pracujących w sekcjach usługowych (G, H, I, J, K, O, P, Q – serwicyzacja gospodarki) w 2002 roku; odsetek gospodarstw domowych bezrolnych + działki w 2002 roku; liczba zarejestrowanych podmiotów w REGON (sektor prywatny) na 1000 ludności w wieku produkcyjnym w 2006 roku; stosunek podmiotów prywatnych świadczących usługi publiczne do podmiotów publicznych świadczących takie usługi w 2006 roku.

poziom aktywności ekonomicznej. Wszystkie te gminy łączyła nader istotna cecha – bardzo bliskie lub bezpośrednie sąsiedztwo dużych miast regionu: Olsztyna (Barczewo, Dywity, Gietrzwałd, Jonkowo, Purda i Stawiguda), Torunia (Lubicz, Obrowo) oraz Białegostoku (Juchnowiec Kościelny, Supraśl, Wasilków).

W metodzie drzew istnieje także możliwość obiektywnej oceny ważności poszczególnych predyktorów aktywności ekonomicznej. Tworząc ranking zmiennych objaśniających, można automatycznie wyłonić z ich grona te, które były najistotniejsze (rysunek 3).



RYСУNEK 3. Ważność predyktorów aktywności ekonomicznej

Ranking predyktorów wskazywał, że z aktywnością ekonomiczną na obszarach wiejskich ZPP szczególnie silnie powiązany był stan demografii oraz sektora pozarolniczego, a także poziom warunków życia i edukacji¹².

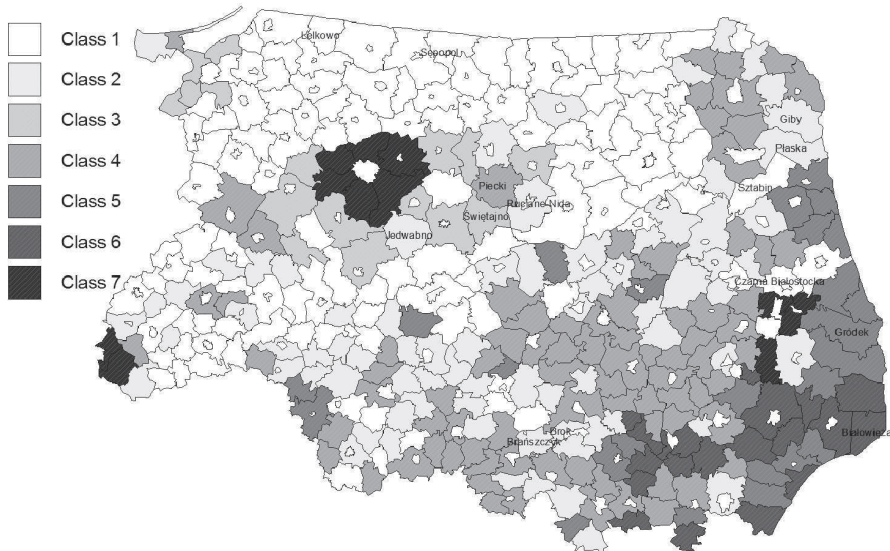
Interpretując wyniki drzewa, zestawiono charakterystykę węzłów końcowych. Utworzyły one 7 zbiorowości gmin o różnych poziomach aktywności ekonomicznej (tabela 2). Obliczona wariancja świadczyła o tym, że wszystkie węzły drzewa charakteryzowała stosunkowo duża jednorodność. Największe zróżnicowanie obserwacji wystąpiło w węzle oznaczonym nr 21. Był on najmniej liczny i skupiał gminy o najwyższym poziomie aktywności ekonomicznej.

¹² Z obserwacji ważności kluczowych predyktorów nasuwać może się pytanie: dlaczego warunki życia nie były predyktorem najważniejszym, skoro stanowiły pierwsze kryterium podziału drzewa? Sektor pozarolniczy i demografia również zyskały inną kolejność i wydają się niejako zaprzeczać diagramowi drzewa. Rozbieżność ta wynika z faktu, że ważność w STATISTICA obliczana jest jako całkowita poprawa oceny resubstytucji uzyskiwana dla danej zmiennej we wszystkich możliwych podziałach w całej sekwencji drzew (a nie tylko w modelu finalnym) i wyrażana jako ułamek w odniesieniu do wszystkich predyktorów. Nawet jeśli na analizowanym diagramie pewna zmienna „przegrywa” ze względu na jakość podziału z innymi, to jej wkład w ocenie ważności jest zawsze brany pod uwagę i traktowany całościowo. W drzewie uznanym za optymalne dana zmienna wystąpi na diagramie dlatego, że była choćby minimalnie lepszym predyktorem od innej. Wziąwszy zaś pod uwagę całą sekwencję drzew, inne zmienne mogą zyskać ważność podobną, a czasem nawet większą, mimo iż nie pojawiły się w drzewie finalnym.

TABELA 2. Wyniki w węzłach końcowych

Zmienna zależna: aktywność ekonomiczna					
Opcje: ilościowa zmienna zależna, drzewo nr 16					
numer węzła	klasa	średnia arytmetyczna	wariancja	liczebność obserwacji	
6	1	0,41	0,003	112	
24	2	0,46	0,003	76	
7	3	0,48	0,002	15	
25	4	0,49	0,002	83	
4	5	0,50	0,002	25	
22	6	0,55	0,002	19	
21	7	0,63	0,006	11	

Węzły końcowe drzewa zostały posortowane według wartości średniej arytmetycznej modelowanego subkomponentu. Przypisano im klasy, które posłużyły do wygenerowania mapy predykcji aktywności ekonomicznej (rysunek 4). Ów obraz przestrzennego kształtowania się przewidywań modelu drzewa pozwalał dostrzec stosunkowo wyraźne prawidłowości. Poza niewielkimi enklawami drzewo zgeneralizowało całą północno-zachodnią część ZPP jako tereny o niskim poziomie aktywności ekonomicznej. Uwagę zwracała liczna grupa gmin zlokalizowanych w obszarze przygranicznym południowo-wschodniej części ZPP (w tym Gródek i Białowieża). Po gminach stanowiących sąsiedztwo dużych miast regionu (Olsztyna, Torunia i Białegostoku) to właśnie tutaj odpowiedzi drzewa przewidywały wysoki poziom aktywności ekonomicznej.



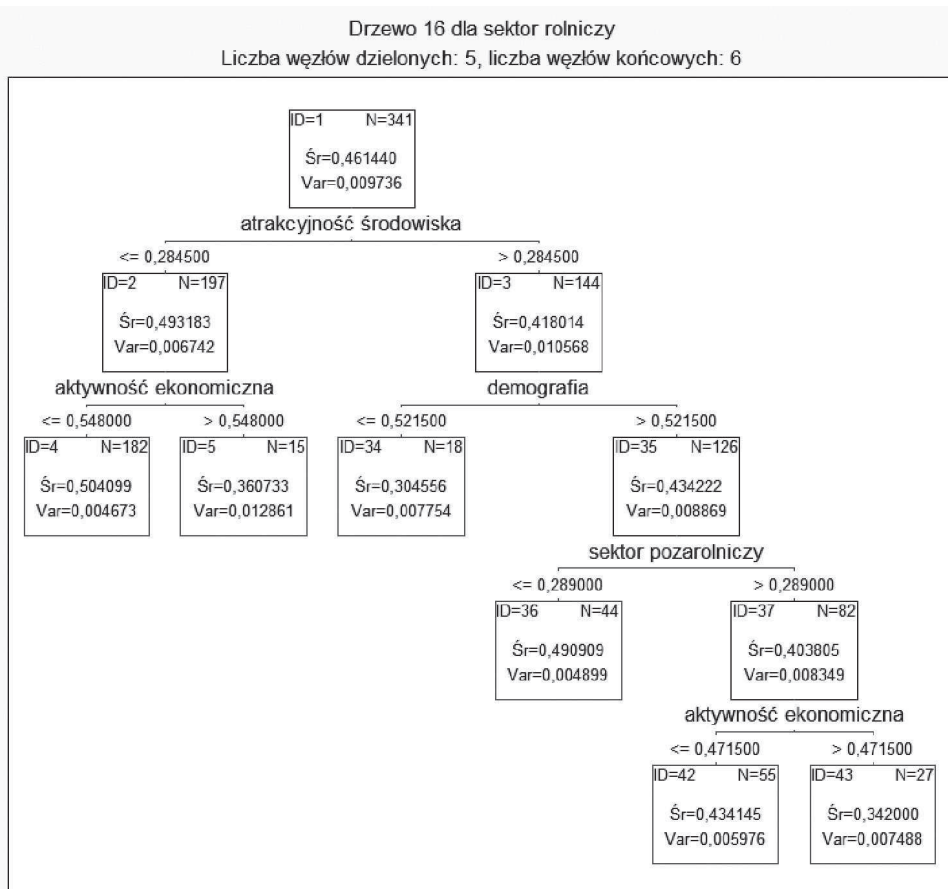
RYSUNEK 4. Model przestrzennego kształtowania się aktywności ekonomicznej

Należy nadmienić, że drzewo przewidywało niski lub bardzo niski poziom aktywności ekonomicznej aż w 11 gminach „naturowych” (Lelkowo, Sępoleń, Jedwabno, Świętajno, Ruciane Nida, Brańszczyk, Brok, Giby, Płaska, Sztabin i Czarna Białostocka). Jedynie w 3 gminach prognozowana była aktywność przeciętna lub ponadprzeciętna (Piecki, Gródek i Białowieża).

MODEL ROZWOJU SEKTORA ROLNICZEGO

Średnia arytmetyczna subkomponentu sektor rolniczy przyjmowała na obszarach wiejskich ZPP wartość 0,46. Wartości skrajne zawierały się w przedziale od 0,14 do 0,72. Rolnictwo charakteryzował zatem nieznacznie większy rozstęp o wartości 0,58. Rozkład tej zmiennej cechowała niewielka asymetria lewostronna, co świadczyło o przewadze obserwacji o wartościach większych od średniej. Był jednomodalny i zbliżony do normalnego.

Według diagramu, z rozwojem sektora rolniczego była silnie związana atrakcyjność środowiska¹³ (podział gałęzi 1), aktywność ekonomiczna (podział gałęzi 2 i 37), a także demografia (podział gałęzi 3) i stan sektora pozarolniczego (podział gałęzi 35). Były to zarazem cztery pierwsze zmienne w rankingu ważności zmiennych objaśniających sektor rolniczy (rysunek 5). Mimo że w drzewie finalnym powstało jedynie 6 węzłów końcowych, obliczony współczynnik korelacji Pearsona wyniósł 0,65, co można uznać za wynik dobry.



RYСУNEK 5. Diagram drzewa regresyjnego przedstawiający model sektora rolniczego

Z modelu wynikało, że sektor rolniczy był lepiej rozwinięty w gminach o mniej atrakcyjnym środowisku (gdzie jego zagregowane oceny nie przekraczały wartości 0,28). Zależność tę tłumaczy fakt, że atrakcyjność środowiska rozpatrywano głównie pod kątem turystyki. Składały się na nią m.in.: lesistość, atrakcyjność rzeźby terenu i udzielone noclegi.

Diagram drzewa wskazywał na odwrotnie proporcjonalną zależność między stanem sektora rolniczego a aktywnością ekonomiczną (rysunek 5). Poziom tej zmiennej przesądził o podziale gałęzi ze zbiorowością 197 gmin o mniejszej atrakcyjności środowiska (ID 2). Pozwoliło to wyróżnić grupę 182 gmin (ID 4) o aktywności nie większej niż 0,55.

W tym węźle końcowym oceny sektora rolniczego miały największą średnią wartość równą 0,50. W zbiorowości 15 gmin o zagregowanych ocenach aktywności większych niż 0,55 sektor rolniczy był już zdecydowanie gorzej rozwinięty ($\bar{S}r = 0,36$).

Najmniejsza średnia wartość oceny sektora rolniczego (0,30) charakteryzowała 18 gmin o większej atrakcyjności środowiska ($> 0,28$) i niższym potencjale demograficznym ($\leq 0,52$).

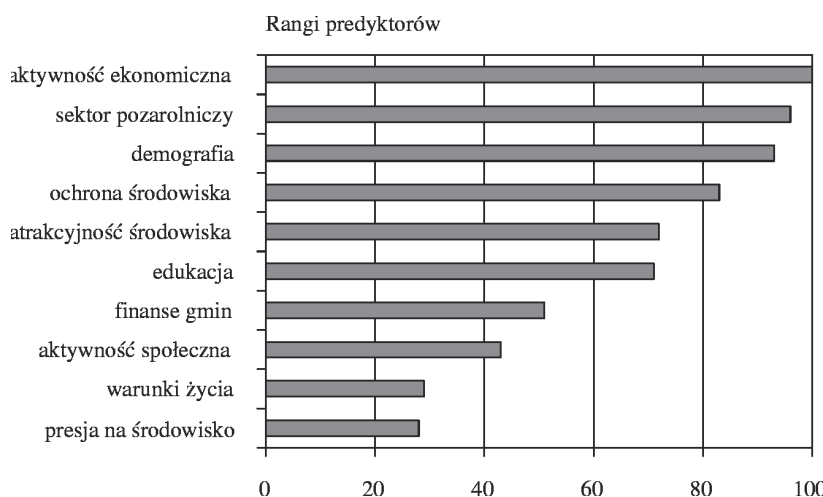
W grupie 126 jednostek samorządowych o atrakcyjniejszym środowisku oraz lepszym potencjale demograficznym (ID 35) stan sektora rolniczego był najbardziej zbliżony do średniej i wynosił 0,43. Podlegał on dalszemu różnicowaniu według stanu sektora pozarolniczego. W ten sposób wyodrębniono grupę 44 gmin (ID 36), które charakteryzowała większa atrakcyjność środowiska, wyższy potencjał demograficzny i niższy poziom rozwoju sektora pozarolniczego ($\leq 0,28$). Tu również rolnictwo stało na wysokim poziomie.

W 82 gminach ZPP o wyższym poziomie rozwoju sektora pozarolniczego (ID 37) ponownie istotnym czynnikiem podziału stała się aktywność ekonomiczna ludności. Zależność pomiędzy tą zmienną a stanem sektora rolniczego była również odwrotnie proporcjonalna.

W porównaniu z poprzednim modelem w modelu rozwoju sektora rolniczego wyraźnie więcej zmiennych osiągnęło stosunkowo dużą wagę (rysunek 6). Zgodnie z rysunkiem stan sektora rolniczego pozostawał w ścisłym związku z aktywnością ekonomiczną, poziomem rozwoju sektora pozarolniczego oraz potencjałem demograficznym. Wyraźnie ważniejsza była tu już atrakcyjność środowiska i intensywność jego ochrony, jak również poziom edukacji w poszczególnych gminach.

Węzły końcowe modelu drzewa utworzyły 6 kategorii gmin o różnych poziomach rozwoju sektora rolniczego (tabela 3). Zostały one posortowane według wartości średniej, a następnie przypisano im klasy, które posłużyły do wygenerowania mapy predykcji stanu sektora rolniczego.

¹³ Stanowiły ją zagregowane metodą sum standaryzowanych: udział gruntów leśnych w powierzchni ogółem w 2006 roku; udział trwałych użytków zielonych w ogóle użytków rolnych w 2006 roku; atrakcyjność rzeźby terenu w punktach jakości rolniczej przestrzeni produkcyjnej; korzystający z noclegów na 1000 mieszkańców w 2006 roku (w %).



RYSUNEK 6. Ważność predyktorów sektora rolniczego

TABELA 3. Wyniki w węzłach końcowych

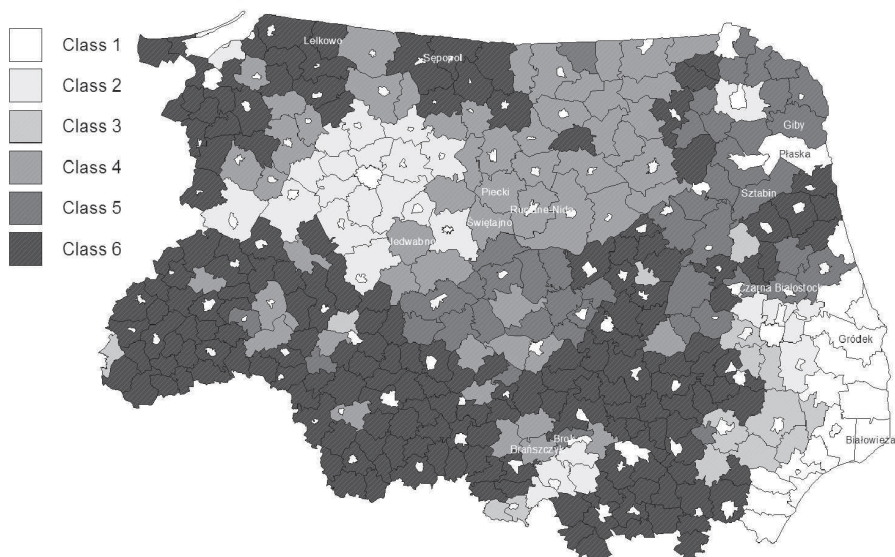
Zmienna zależna: aktywność ekonomiczna				
Opcje: ilościowa zmienna zależna, drzewo nr 16				
numer węzła	klasa	średnia arytmetyczna	wariancja	liczebność obserwacji
34	1	0,30	0,008	18
43	2	0,34	0,007	27
5	3	0,36	0,013	15
42	4	0,43	0,006	55
36	5	0,49	0,005	44
4	6	0,50	0,005	182

Wariancja obliczona w węzłach modelu sektora rolniczego świadczyła o tym, że obliczone średnie arytmetyczne charakteryzowała mniejsza jednorodność. Tutaj również największe zróżnicowanie obserwacji wystąpiło w najmniej liczonym węźle nr 5. Skupiał on kilkanaście gmin o przewidywanym przeciętnym poziomie rozwoju sektora rolniczego.

Przewidywania oparte na modelu drzewa wskazywały najslabiej rozwinięte rolnictwo w aktywnych ekonomicznie gminach zlokalizowanych w obszarze przygranicznym południowo-wschodniej części ZPP (w tym w gminach „naturalnych”, takich jak Płaska, Gródek i Białowieża).

W ujęciu przestrzennym prezentowany model wykazywał marginalizację rolnictwa w gminach stanowiących bezpośrednie lub bliskie sąsiedztwo dwóch dużych miast ZPP: Olsztyna i Białegostoku. Jednakże w sąsiedztwie innych ośrodków miejskich: Suwałk, Elbląga i Torunia, tendencja ta była już praktycznie niewidoczna.

Znaczny odsetek (ponad 66%) badanych gmin został zaklasyfikowany do klasy 5 i 6, tj. obszarów o wysokim poziomie rozwoju rolnictwa (rysunek 7). Poza



RYСУNEK 7. Model przestrzennego kształtowania się stanu sektora rolniczego

kilkoma mniejszymi skupieniami na północy (okolice Olecka, Bartoszyc, Korsz, Braniewa, Elbląga, Ornety i Pasłęka), tereny te obejmowały południową i południowo-zachodnią część Zielonych Płuc Polski.

W 6 gminach „naturowych” (Lelkowo, Sępoleń, Giby, Sztubin, Czarna Białostocka i Brok) przewidywania oparte na modelu drzewa wskazywały na wysoki poziom rozwoju rolnictwa. W 5 jednostkach samorządowych przewidywano umiarkowany poziom rozwoju tego sektora (Jedwabno, Piecki, Świętajno, Ruciane Nida i Brańszczyk). Najniższe predykcje odnotowano w przypadku 3 gmin „naturowych” (Płaska, Grodek i Białowieża).

DYSKUSJA WYNIKÓW, PODSUMOWANIE I WNIOSKI

Zastosowane w niniejszej pracy drzewa regresyjne oparte na algorytmie C&RT umożliwiły przeprowadzenie analizy statystycznej i udzielenie odpowiedzi na pytanie, jakie cechy wyróżniały gminy aktywne ekonomicznie i gminy o dobrze rozwiniętym sektorze rolniczym na terenach ZPP.

Na tle całego badanego obszaru ZPP stan gospodarki i rolnictwa gmin objętych w szerokim zakresie specjalną ochroną ptaków i siedlisk w ramach sieci Natura 2000 można było uznać za przeciętny lub gorszy od przeciętnego. Wskazywały na to przewidywania obu zbudowanych modeli drzew regresyjnych (tabela 4). Podsumowując wyniki uzyskane dla gmin „naturowych”, stwierdzono też, że w przypadku predykcji obu modeli widoczna była tendencja do przeszacowywania predykcji badanych subkomponentów w stosunku do ich wartości faktycznych (tabela 4). Faktyczna średnia wartość aktywności ekonomicznej gmin „naturowych” kształtowała się na poziomie 0,42 i została zawyżona o 9%. Średnia ocena poziomu rozwoju sektora rolniczego w gminach „naturowych”

wynosiła w rzeczywistości 0,38, a predykcje drzewa zawyżyły ją o 14%. Owe umiarkowane przewidywania obu modeli były zatem i tak nieco zbyt optymistyczne, gdyż faktyczne oceny aktywności ekonomicznej ludności i stanu sektora rolniczego były średnio o kilka, kilkanaście procent niższe.

TABELA 4. Przewidywania modeli i błędy względne predykcji

Badane gminy	Model aktywności ekonomicznej			Model sektora rolniczego		
	przewidywania modelu	błąd względny predykcji [%]	węzły końcowe	przewidywania modelu	błąd względny predykcji [%]	węzły końcowe
Brańszczyk	0,46	22	24	0,43	31	42
Płaska	0,46	23	24	0,30	-16	34
Sztabin	0,41	-17	6	0,49	-3	36
Czarna Białostocka	0,41	-18	6	0,49	41	36
Gródek	0,50	2	4	0,30	44	34
Białowieża	0,55	-14	22	0,30	49	34
Giby	0,46	18	24	0,49	-7	36
Sepopol	0,41	14	6	0,50	-15	4
Lelkowo	0,41	50	6	0,50	43	4
Piecki	0,49	8	25	0,43	-7	42
Ruciane-Nida	0,46	26	24	0,43	102	42
Jedwabno	0,46	2	24	0,43	21	42
Świątajno	0,48	20	7	0,43	-7	42
Brok	0,46	35	24	0,49	42	36
Wartość średnia	0,46	9	x	0,43	14	x

Na podstawie wyników zestawionych w tabeli 4 uznano, że zaprezentowane modele drzew regresyjnych charakteryzowały średnie błędy względne¹⁴ predykcji o akceptowalnej wielkości. W istotę modelowania wpisuje się wszak zwykle poszukiwanie najważniejszych zależności, co wymusza generalizację i upraszczanie rzeczywistości.

Wobec braku danych o stanie gospodarki i rolnictwa przed ustanowieniem sieci Natura 2000 należy nader ostrożnie wnioskować o jej szkodliwym wpływie na gospodarkę i rolnictwo badanych gmin. Nie można jednoznacznie stwierdzić kierunku przyczynowości: czy powoływanie i rozszerzanie sieci Natura 2000 przyczynia się do niższego poziomu rolnictwa i gospodarki, czy raczej umiarkowany poziom lokalnej gospodarki warunkuje powoływanie i poszerzanie tej sieci. Zastosowane metody nie pozwalają stwierdzić jednoznacznie kierunku owej przyczynowości. Tym bardziej że po przeanalizowaniu macierzy korelacji stwierdzono brak istotnego statystycznie związku liniowego między procentem pokrycia obszaru gminy siecią Natura 2000 a poziomem aktywności ekonomicznej jej mieszkańców. Związek ten uwiłdocił się jedynie w przypadku rolnictwa i choć miał oczekiwany odwrotnie proporcjonalny charakter, wcale nie był tak duży jak podejrzewano ($W_k = -0,31$).

¹⁴ Błąd względny predykcji to różnica między wartością obserwowaną i przewidywaną odniesiona do wartości obserwowanej.

Przedstawione powyżej fakty potwierdzały i uściślały wyniki obu drzew. W świetle przeprowadzonych badań stwierdzono bowiem, że aktywności ekonomicznej nie określały subkomponenty środowiskowe (presja, atrakcyjność czy ochrona), lecz społeczno-gospodarcze (demografia, stan rozwoju sektora pozarolniczego i warunki życia).

Model sektora rolniczego ujawniał bardziej widoczny związek korelacyjny między intensywnością ochrony środowiska a stanem rolnictwa. Z rankingu ważności predyktorów (rysunek 6) wynikało, że ochrona środowiska miała tu już wyraźnie większe znaczenie, choć nie była ona predyktorem kluczowym.

W modelu sektora rolniczego ważniejsze były również uwarunkowania o charakterze społeczno-gospodarczym. Rolnictwo rozwijało się wszak tam, gdzie odnotowano niższy poziom aktywności ekonomicznej, gdzie konkurujący z rolnictwem sektor pozarolniczy był słabiej rozwinięty, a atrakcyjność środowiska dla rozwoju turystyki niewielka.

Pozytywne odpowiedzi na oba postawione we wstępie pytania badawcze potwierdzał fakt, że ani predykcje aktywności ekonomicznej, ani predykcje stanu rolnictwa nie osiągały dużych wartości w objętych badaniami gminach „naturowych”.

Sieć Natura 2000 ustanawia się zwykle na cennych przyrodniczo obszarach, z reguły mniej zurbanizowanych i zaludnionych, a zatem i mniej aktywnych gospodarczo (swoiste wyjątki to m.in. objęcie tą siecią obszaru miasta, np. Broku). Aby więc mówić z całym przekonaniem o negatywnym wpływie działań związanych z ochroną środowiska na gospodarkę i rolnictwo gmin wiejskich ZPP, należałoby podjąć dodatkowe badania na podstawie danych obejmujących wieloletnia, o przekroju czasowym, uwzględniającym fakt powołania i/lub powiększenia obszaru chronionego na terenie wybranej jednostki (bądź kilku) jednostek samorządu terytorialnego.

W podjętych badaniach walory poznawcze obu zbudowanych modeli podniosł fakt oszacowania i utworzenia rankingów ważności predyktorów. Świadczyły one o drugorzędnej niejako roli czynników środowiskowych, ujawniały zaś silne powiązanie aktywności ekonomicznej i rolnictwa z rozwojowymi czynnikami o charakterze społeczno-gospodarczym.

BIBLIOGRAFIA

- Bołtromiuk A., Kłodziński M., 2009: *Konflikt sieci Natura 2000 z procesem rozwoju gminy*. „Czas Morza” 3 (39).
- Breiman L., Friedman J., Olshen R.A., Stone Ch. J., 1993: *Classification and Regression Trees*. Chapman and Hall, New York.
- Dacko M., 2007: *Badanie rynku mieszkań w Olsztynie z zastosowaniem metod statystycznych*. Studia i materiały Towarzystwa Naukowego Nieruchomości. „Rynek nieruchomości – analizy, modelowanie, inwestowanie” 15, 3–4.
- Łapczyński M., 2003: *Drzewa klasyfikacyjne w badaniach satysfakcji i lojalności klientów* (<http://www.statsoft.pl/czytelnia/marketing/drzewa.pdf>).
- Łapczyński M., 2007: *Wpływ aktywności mieszkańców na poziom życia w gminach woj. małopolskiego* (http://www.statsoft.pl/czytelnia/8_2007/Lapczynski05.pdf).

- Manasse P., Roubini N., 2009: „*Rules of thumb*” for sovereign debt crises. „*Journal of International Economics*” 78, 2: 192–205
- Sokołowski A., 2002: *Metody stosowane w Data mining*. Materiały seminaryjne „Data mining – metody i przykłady”. Statsoft, Kraków.
- Sroka W., Dacko M., 2010: *Ocena czynników rozwoju przodujących gospodarstw rolniczych z wykorzystaniem drzew regresyjnych typu C & RT*. „Zagadnienia Ekonomiki Rolnej” 2.

NATURA 2000 NETWORK AND THE STATE OF LOCAL ECONOMY OF THE RURAL AREAS OF THE GREEN LUNGS OF POLAND REGION – THE APPLICATION OF MODELS

Abstract. In the paper models of regressive trees were used to illustrate the shaping of the agricultural sector’s condition and economic activity of the residents of rural areas in the Green Lungs of Poland region. These models were prepared in an effort to empirically check the pessimistic assessments of the state of the local economy of communes situated in the Green Lungs of Poland region, which are broadly covered by special protection of birds and habitats within the Natura 2000 network. The models’ results presented simple criteria for distinguishing communes with different levels of the development of economic activity and agriculture. They also made it possible to assess the validity of the used predictors that constituted the sub-components of sustainable development distinguished in the research project. The results of predictions of the models of regressive trees were shown on maps, which permitted to make an assessment of the spatial shaping of local economy in the areas of the Green Lungs of Poland region.

Key words: Natura 2000, agriculture, economic activity, regression trees