

ANALIZA DRZEW DECYZYJNYCH

PAWEŁ GRYGIEL
GRZEGORZ HUMENNY

Wzmacnianie polityki opartej na faktach na rzecz poprawy umiejętności podstawowych osób dorosłych

Projekt współfinansowany w ramach programu Unii Europejskiej Erasmus+
Warszawa, 20 listopada 2015 r.

Analiza drzew klasyfikacyjnych - założenia

Metoda wzrostu drzewa: wyczerpujący CHAID

Maksymalna głębokość drzewa: 4

Minimalna liczba obserwacji węzła nadrzędnego: 200

Minimalna liczba obserwacji węzła podrzędnego: 100

Poziom istotności testu ch-kwadrat: 0,05 skorygowany metodą Benferoniego

Możliwość ponownego podziału połączonych kategorii w węzłach

Przynależność do grupy wykluczonej ze względu na niskie umiejętności jako kategoria docelowa analizy

Zmienne wykorzystane w analizach

Wiek (w latach)

Płeć

Stan zdrowia (deklaracja)

Liczba osób w gospodarstwie domowym

Zamieszkiwanie z:

Liczba dzieci (w tym brak)

Liczba książek w domu

Poziom wykształcenia respondenta (tylko przy analizie uwarunkowań niskich osiągnięć)

Wykształcenia matki respondenta

Wykształcenia ojca respondenta

Status społeczno-ekonomiczny respondenta (bazujący na ISCO)

Roczne dochody (w podziale na grupy centylowe)

Status na rynku pracy (deklaracja)

Doświadczenie zawodowe

Generalna klasyfikacja zawodu respondenta



Rodzaj gminy

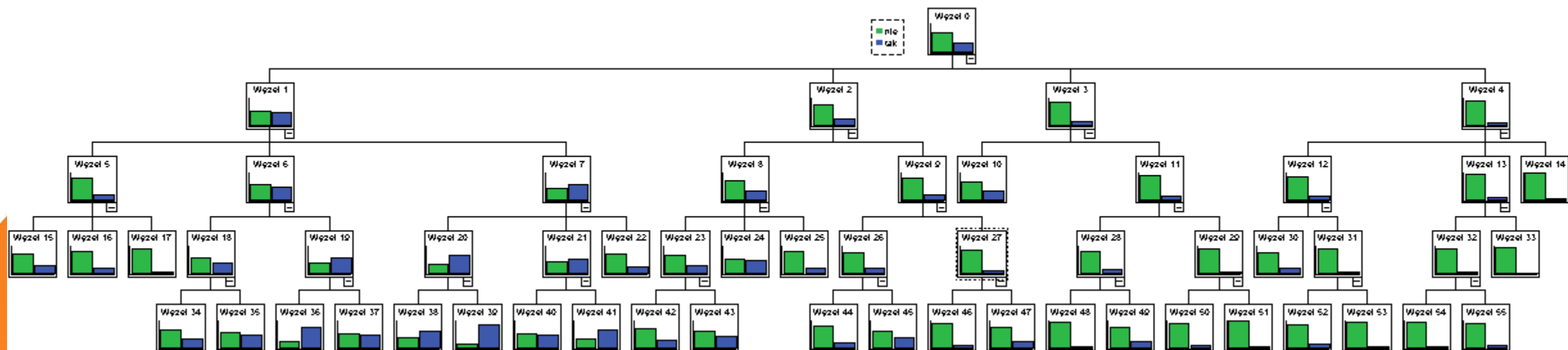
Dział gospodarki miejsca obecnej lub ostatniej pracy respondenta

Zmienna zależna:

Niskie osiągnięcia z matematyki lub języka (poniżej 226 punktów)

 nie
 tak

Węzeł 0		
Kategoria	%	n
 nie	70,5	5650
 tak	29,5	2369
Ogółem	100,0	8019



Korzyści dla poszczególnych węzłów						
Węzeł	Węzeł		Korzyść		Odpowiedzi	Indeks
	N	Procent	N	Procent		
39	120	1,5%	101	4,3%	84,2%	284,9%
36	132	1,6%	99	4,2%	75,0%	253,9%
41	195	2,4%	133	5,6%	68,2%	230,9%
38	515	6,4%	321	13,6%	62,3%	211,0%
40	639	8,0%	312	13,2%	48,8%	165,3%
37	198	2,5%	95	4,0%	48,0%	162,4%
24	158	2,0%	75	3,2%	47,5%	160,7%
35	535	6,7%	243	10,3%	45,4%	153,7%

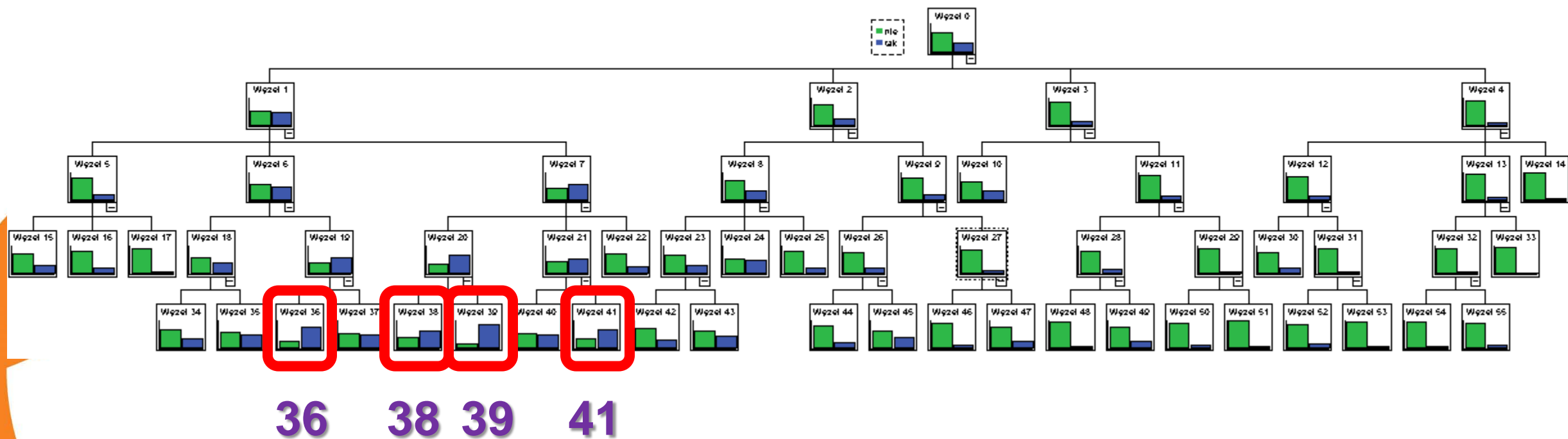
Algorytm budowy drzewa: Wyczerpujący CHAID
Zmienna zależna: Niskie osiągnięcia z matematyki lub języka (poniżej 226 punktów)

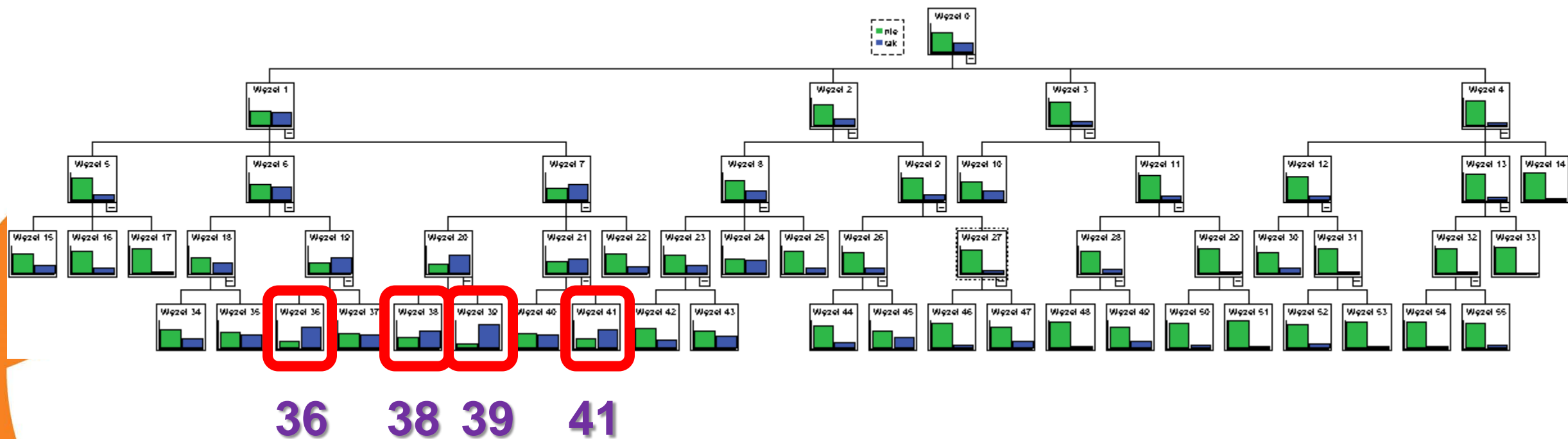
Korzyści dla poszczególnych węzłów

Węzeł	Węzeł		Korzyść		Odpowiedzi	Indeks
	N	Procent	N	Procent		
39	120	1,5%	101	4,3%	84,2%	284,9%
36	132	1,6%	99	4,2%	75,0%	253,9%
41	195	2,4%	133	5,6%	68,2%	230,9%
38	515	6,4%	321	13,6%	62,3%	211,0%
40	639	8,0%	312	13,2%	48,8%	165,3%
37	198	2,5%	95	4,0%	48,0%	162,4%
24	158	2,0%	75	3,2%	47,5%	160,7%
35	535	6,7%	243	10,3%	45,4%	153,7%

Algorytm budowy drzewa: Wyczerpujący CHAID

**Zmienna zależna: Niskie osiągnięcia z matematyki lub języka
(poniżej 226 punktów)**



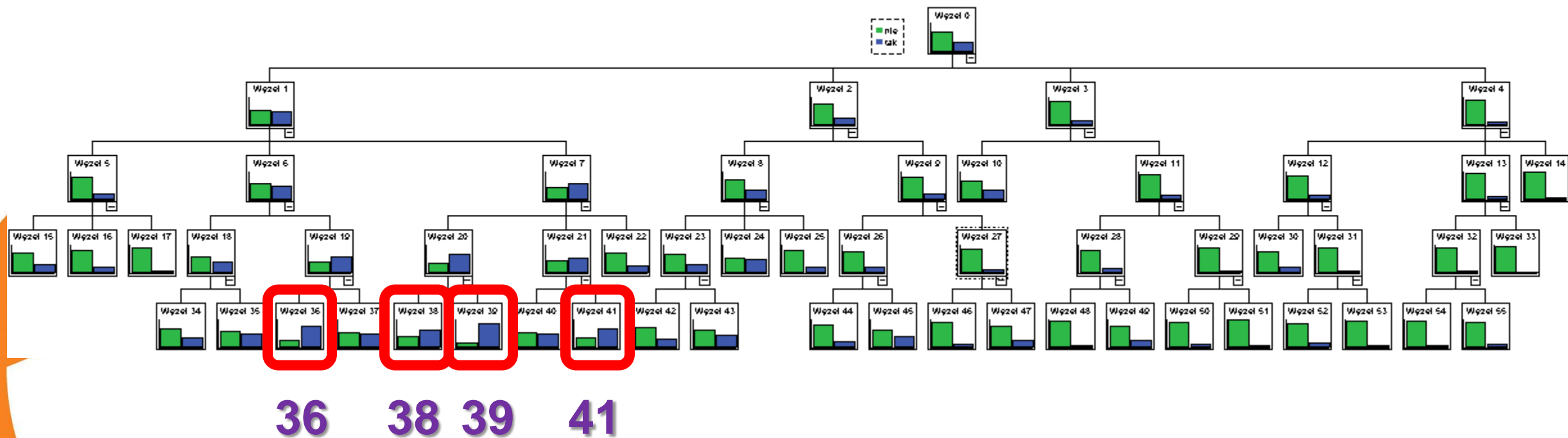


Zależna: niskie osiągnięcia z matematyki lub polskiego

- Zasadnicze zawodowe, gimnazjalne lub bez wykształcenia
- Starsi niż 47 lat
- Posiadający w domu 10 książek lub mniej
- Bez małżonka lub partnera → **84,2%**

Węzeł numer 39

Węzeł 39		
Kategoria	%	n
nie	15,8	19
tak	84,2	101
Ogółem	1,5	120

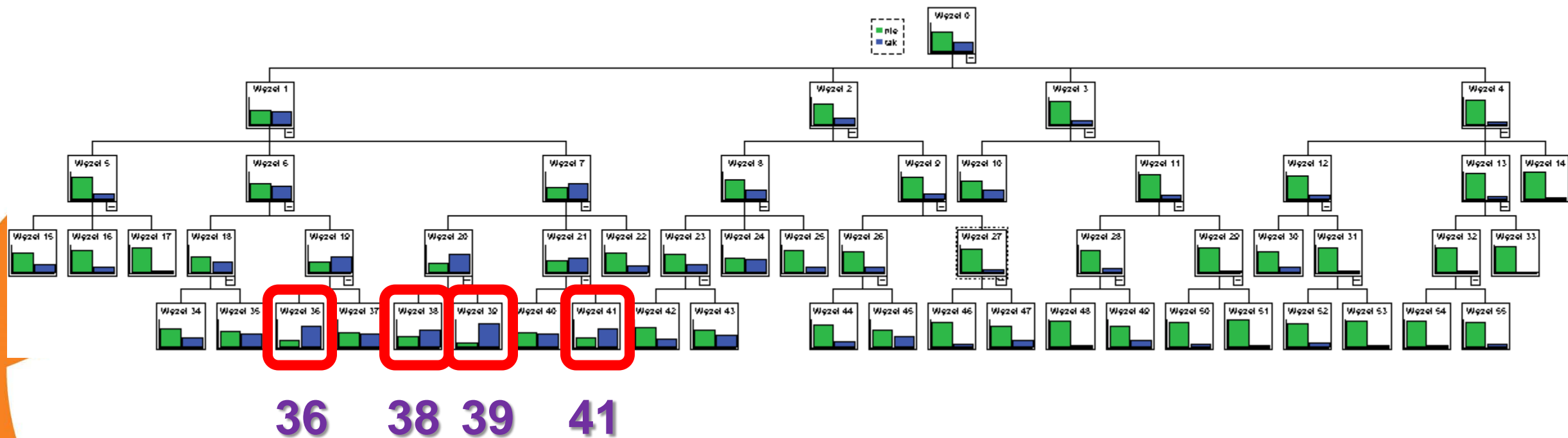


Zależna: niskie osiągnięcia z matematyki lub polskiego

- Zasadnicze zawodowe, gimnazjalne lub bez wykształcenia
- Starsi niż 47 lat
- Posiadający w domu 10 książek lub mniej
- Z małżonkiem lub partnerem; samotnie → **62,3%**

Węzeł numer 38

Węzeł 38		
Kategoria	%	n
nie	37,7	194
tak	62,3	321
Ogółem	6,4	515

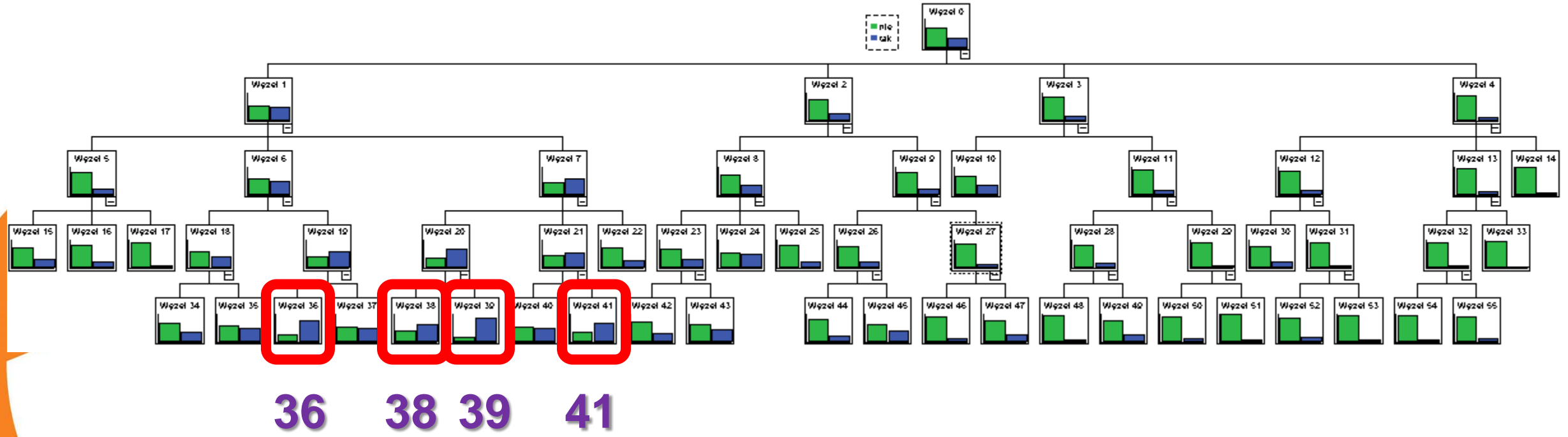


Zależna: niskie osiągnięcia z matematyki lub polskiego

- Zasadnicze zawodowe, gimnazjalne lub bez wykształcenia
- W wieku 26-49 lat
- Mieszkający bez małżonka/partnera; samotnie
- Pracujący w rolnictwie → **75%**

Węzeł numer 36

Węzeł 36		
Kategoria	%	n
■ nie	25,0	33
■ tak	75,0	99
Ogółem	1,6	132

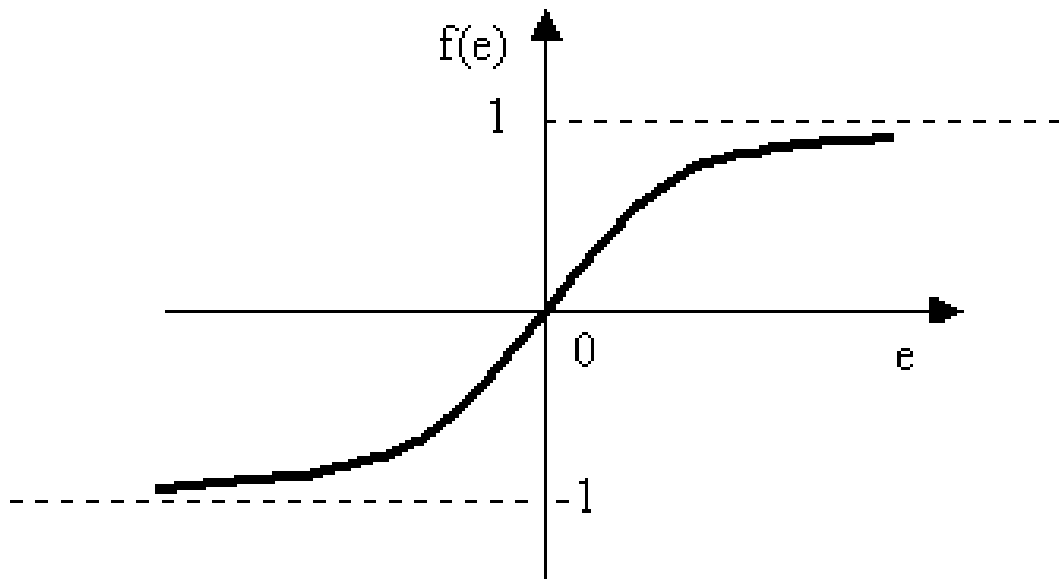


Zależna: niskie osiągnięcia z matematyki lub polskiego

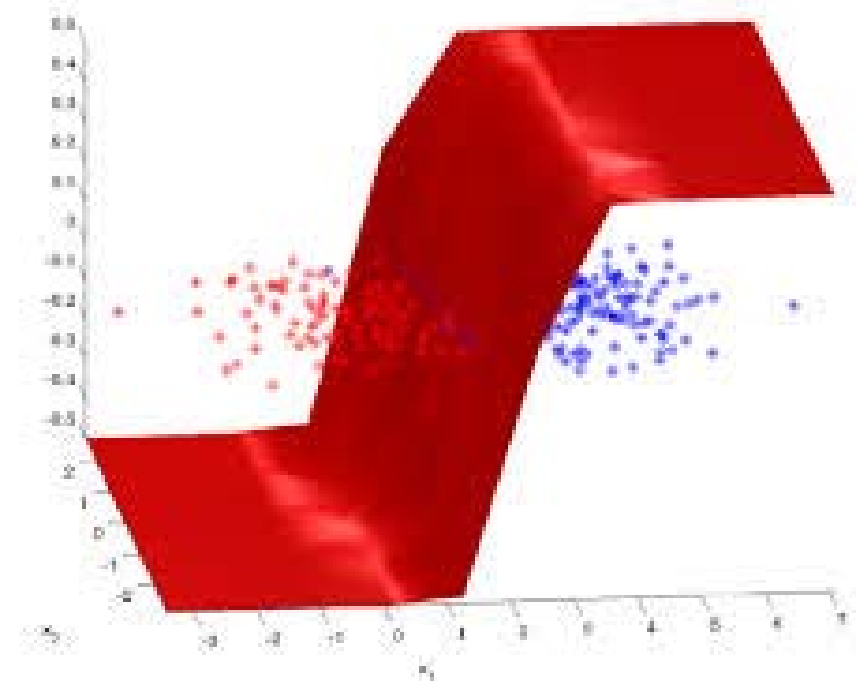
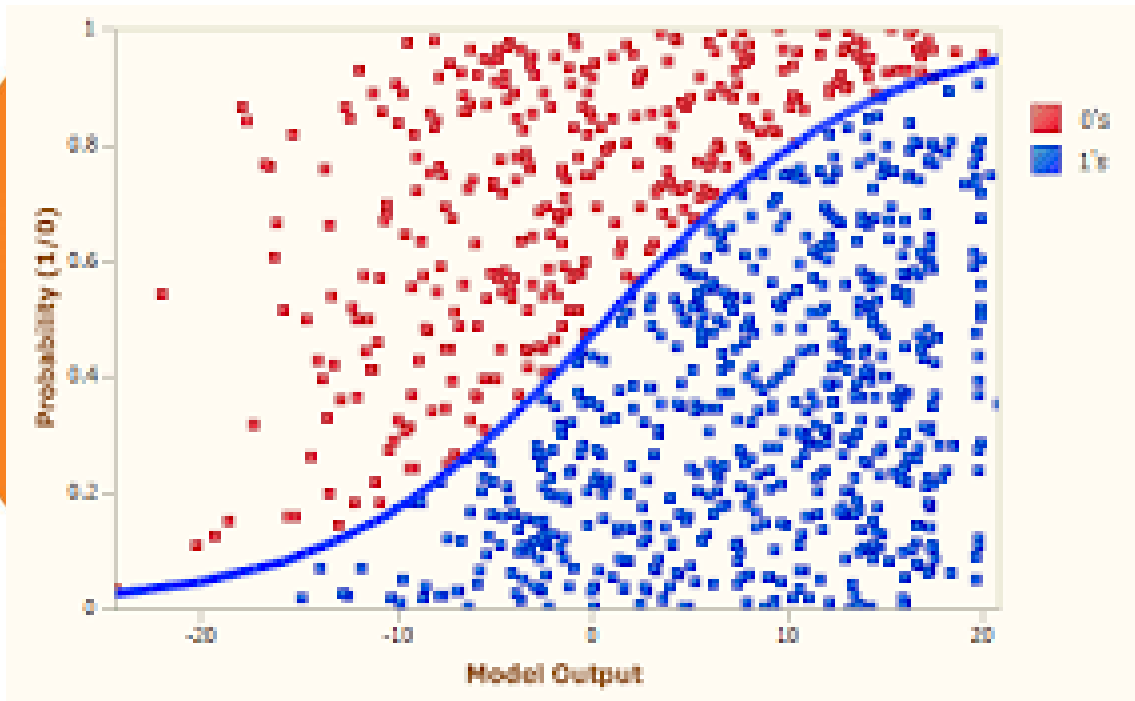
- Zasadnicze zawodowe, gimnazjalne lub bez wykształcenia
- Starsi niż 47 lat
- Posiadający od 10 do 100 książek
- O wykształceniu gimnazjalnym → **68,2%**

Węzeł numer 41

Węzeł 41		
Kategoria	%	n
nie	31,8	62
tak	68,2	133
Ogółem	2,4	195



REGRESJA LOGISTYCZNA



Opis analizy

Regresja logistyczna (STATA, logit)

Przeprowadzona z uwzględnieniem złożonego charakteru próby i wag replikacyjnych (STATA, svy)

Zmienna zależna: Niskie osiągnięcia z matematyki lub języka (poniżej 226 punktów);
wykluczony (1), nie wykluczony (0)



Zmienne niezależne:

- poziom wykształcenia respondenta,
- wykształcenie ojca respondenta,
- liczba książek,
- generalna klasyfikacja zawodu respondenta,
- płeć,
- dział gospodarki miejsca obecnej lub ostatniej pracy respondenta,
- stan zdrowia,
- wiek,
- wykształcenie matki,
- status na rynku pracy,
- roczne dochody,
- status społeczno-ekonomiczny respondenta,
- zamieszkiwanie (samotne, bez partnera, z partnerem)

Wyniki analizy regresji

IBE 

zmienna	oddziaływanie
poziom wykształcenia respondenta	⊗
wykształcenie ojca respondenta	⊗
liczba książek	√ —
generalna klasyfikacja zawodu respondenta	√
płeć	⊗
dział gospodarki miejsca obecnej lub ostatniej pracy respondenta	⊗
stan zdrowia	⊗
wiek	⊗
wykształcenie matki	⊗
status na rynku pracy	√
roczne dochody	√ —
status społeczno-ekonomiczny respondenta	⊗
zamieszkiwanie (samotne, bez partnera, z partnerem)	⊗

Wyniki analizy regresji

IBE 

zmienna	oddziaływanie
poziom wykształcenia respondenta	⊗
wykształcenie ojca respondenta	⊗
liczba książek	√ —
generalna klasyfikacja zawodu respondenta	√
płeć	⊗
dział gospodarki miejsca obecnej lub ostatniej pracy respondenta	⊗
stan zdrowia	⊗
wiek	⊗
wykształcenie matki	⊗
status na rynku pracy	√
roczne dochody	√ —
status społeczno-ekonomiczny respondenta	⊗
zamieszkiwanie (samotne, bez partnera, z partnerem)	⊗

Zmienna niezależna: liczba książek

Liczba książek (ref. „do 10”)	b	P>t
11 do 25 książek	-.13	0.556
26 do 100 książek	-.42	0.030
101 do 200 książek	-.87	0.000
201 do 500 książek	-1.07	0.000
Ponad 500 książek	-.58	0.134

Wyniki analizy regresji

IBE 

zmienna	oddziaływanie
poziom wykształcenia respondenta	⊗
wykształcenie ojca respondenta	⊗
liczba książek	√ —
generalna klasyfikacja zawodu respondenta	√
płeć	⊗
dział gospodarki miejsca obecnej lub ostatniej pracy respondenta	⊗
stan zdrowia	⊗
wiek	⊗
wykształcenie matki	⊗
status na rynku pracy	√
roczne dochody	√ —
status społeczno-ekonomiczny respondenta	⊗
zamieszkiwanie (samotne, bez partnera, z partnerem)	⊗

Zmienna niezależna: Generalna klasyfikacja zawodu respondenta

Generalna klasyfikacja zawodu respondenta (ref. „specjaliści”)	b	P>t
wykwalifikowani pracownicy niefizyczni	.51	0.002
wykwalifikowani pracownicy fizyczni	.44	0.035
pracownicy niewykwalifikowani	.39	0.107

Wyniki analizy regresji

IBE 

zmienna	oddziaływanie
poziom wykształcenia respondenta	⊗
wykształcenie ojca respondenta	⊗
liczba książek	√ —
generalna klasyfikacja zawodu respondenta	√
płeć	⊗
dział gospodarki miejsca obecnej lub ostatniej pracy respondenta	⊗
stan zdrowia	⊗
wiek	⊗
wykształcenie matki	⊗
status na rynku pracy	√
roczne dochody	√ —
status społeczno-ekonomiczny respondenta	⊗
zamieszkiwanie (samotne, bez partnera, z partnerem)	⊗

status na rynku pracy (ref. „pracujący lub uczący się”)	b	P>t
bezrobotny	-.09	0.805
emerytura	-.15	0.736
niepełnosprawny	1.81	0.063
zajmuje się domem/dziećmi	1.44	0.013
Inna sytuacja	.32	0.542

Wyniki analizy regresji

IBE 

zmienna	oddziaływanie
poziom wykształcenia respondenta	⊗
wykształcenie ojca respondenta	⊗
liczba książek	√ —
generalna klasyfikacja zawodu respondenta	√
płeć	⊗
dział gospodarki miejsca obecnej lub ostatniej pracy respondenta	⊗
stan zdrowia	⊗
wiek	⊗
wykształcenie matki	⊗
status na rynku pracy	√
roczne dochody	√ —
status społeczno-ekonomiczny respondenta	⊗
zamieszkiwanie (samotne, bez partnera, z partnerem)	⊗



centyl rocznych dochodów (ref. „do 10 centyla”)	b	P>t
od 10 do mniej niż 25	.46	0.045
od 25 do mniej niż 50	.45	0.031
od 50 do mniej niż 75	.09	0.721
od 75 do mniej niż 90	-.12	0.781
90 lub wyższy	-.21	0.542

Zasadnicze zawodowe, gimnazjalne lub bez wykształcenia

Starsi niż 47 lat		W wieku 26-49 lat		
Posiadający w domu 10 książek lub mniej		Posiadający od 10 do 100 książek		Mieszkający bez małżonka/partnera; samotnie
Bez małżonka lub partnera 84,2%		O wykształceniu gimnazjalnym 68,2%		Pracujący w rolnictwie 75%
Z małżonkiem lub partnerem; samotnie 62,3%				
1,5% populacji	6,4% populacji	2,4% populacji	1,6% populacji	Łącznie 12% osób w wieku 16-65 lat
4,3% grupy wykluczonej	4,2% grupy wykluczonej	5,6% grupy wykluczonej	13,6% grupy wykluczonej	Łącznie 27,6 % osób wykluczonych w wieku 16-65 lat

Ponad 2,3 razy większa szansa na znalezienie w tych grupach osób wykluczonych ze względu na niskie umiejętności, niż w całej populacji

Wnioski

Poza czynnikami związanymi z:

- wykształceniem
- wiekiem
- tym z kim się zamieszkuje
- liczbą książek w domu oraz
- działem gospodarki, w którym się pracuje

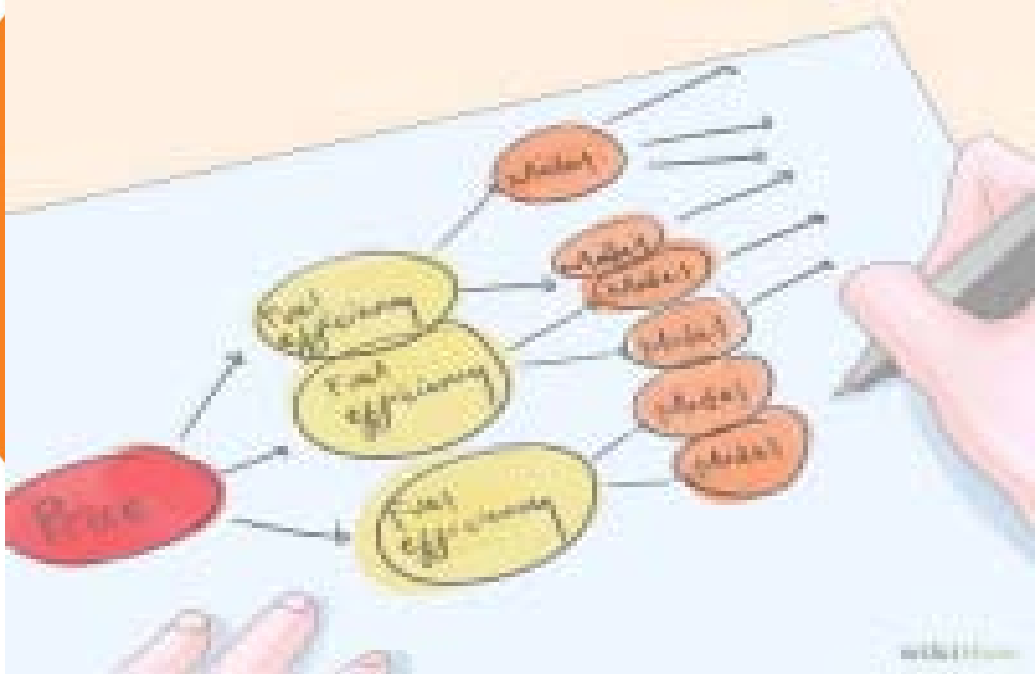
Modele regresji logistycznej wskazują na istotne znaczenie również:

- generalna klasyfikacja zawodu respondenta
- status na rynku pracy oraz
- roczne dochody

IBI



ANALIZA DRZEW DECYZYJNYCH



Analiza drzew klasyfikacyjnych - założenia

Metoda wzrostu drzewa: wyczerpujący CHAID

Maksymalna głębokość drzewa: 4

Minimalna liczba obserwacji węzła nadrzędnego: 200

Minimalna liczba obserwacji węzła podrzędnego: 100

Poziom istotności testu ch-kwadrat: 0,05 skorygowany metodą Benferoniego

Możliwość ponownego podziału połączonych kategorii w węzłach

Przynależność do grupy wykluczonej ze względu na niskie wykształcenie jako kategoria docelowa analizy



Zmienne wykorzystane w analizach

- Wiek (w latach)
- Płeć
- Stan zdrowia (deklaracja)
- Liczba osób w gospodarstwie domowym
- Zamieszkiwanie z:
- Liczba dzieci (w tym brak)
- Liczba książek w domu
- Poziom wykształcenia respondenta (tylko przy analizie uwarunkowań niskich osiągnięć)
- Wykształcenia matki respondenta
- Wykształcenia ojca respondenta
- Status społeczno-ekonomiczny respondenta (bazujący na ISCO)
- Roczne dochody (w podziale na grupy centylowe)
- Status na rynku pracy (deklaracja)
- Doświadczenie zawodowe
- Generalna klasyfikacja zawodu respondenta
- Rodzaj gminy
- Dział gospodarki miejsca obecnej lub ostatniej pracy respondenta

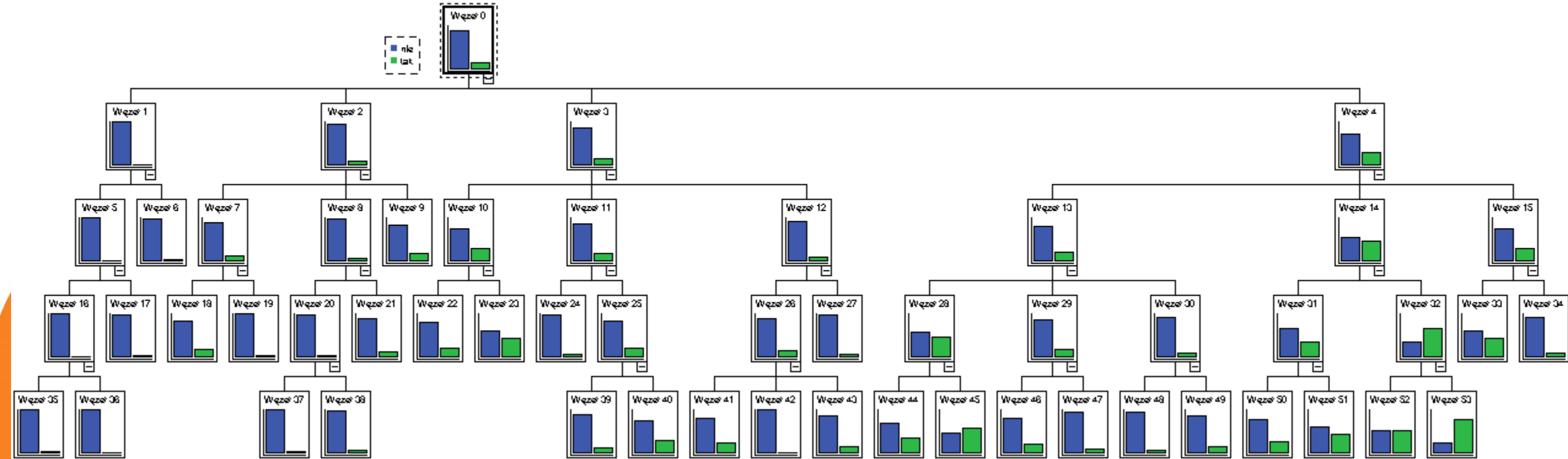
Zmienna zależna:

Niskie osiągnięte wykształcenie
(gimnazjalne lub niższe)

 nie
 tak

Węzeł 0		
Kategoria	%	n
 nie	87,2	6724
 tak	12,8	988
Ogółem	100,0	7712

IBE 54 węzły



Korzyści dla poszczególnych węzłów

Węzeł	Węzeł		Korzyść		Odpowiedzi	Indeks
	N	Procent	N	Procent		
53	128	1,7%	99	10,0%	77,3%	603,7%
45	158	2,0%	89	9,0%	56,3%	439,7%
52	101	1,3%	51	5,2%	50,5%	394,1%
33	151	2,0%	63	6,4%	41,7%	325,7%
51	209	2,7%	87	8,8%	41,6%	324,9%
23	168	2,2%	69	7,0%	41,1%	320,6%
44	148	1,9%	48	4,9%	32,4%	253,2%
40	203	2,6%	56	5,7%	27,6%	215,3%
50	166	2,2%	40	4,0%	24,1%	188,1%

Algorytm budowy drzewa: Wyczerpujący CHAID

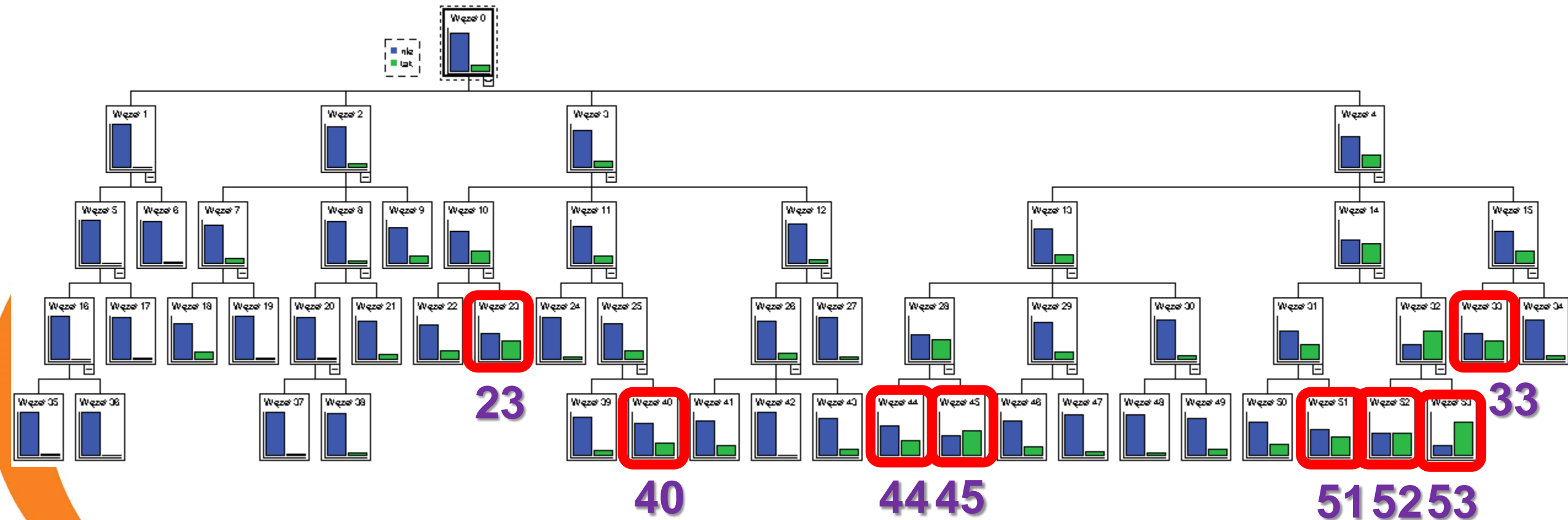
Zmienna zależna: Niskie osiągnięte wykształcenie (gimnazjalne lub niższe)

Korzyści dla poszczególnych węzłów

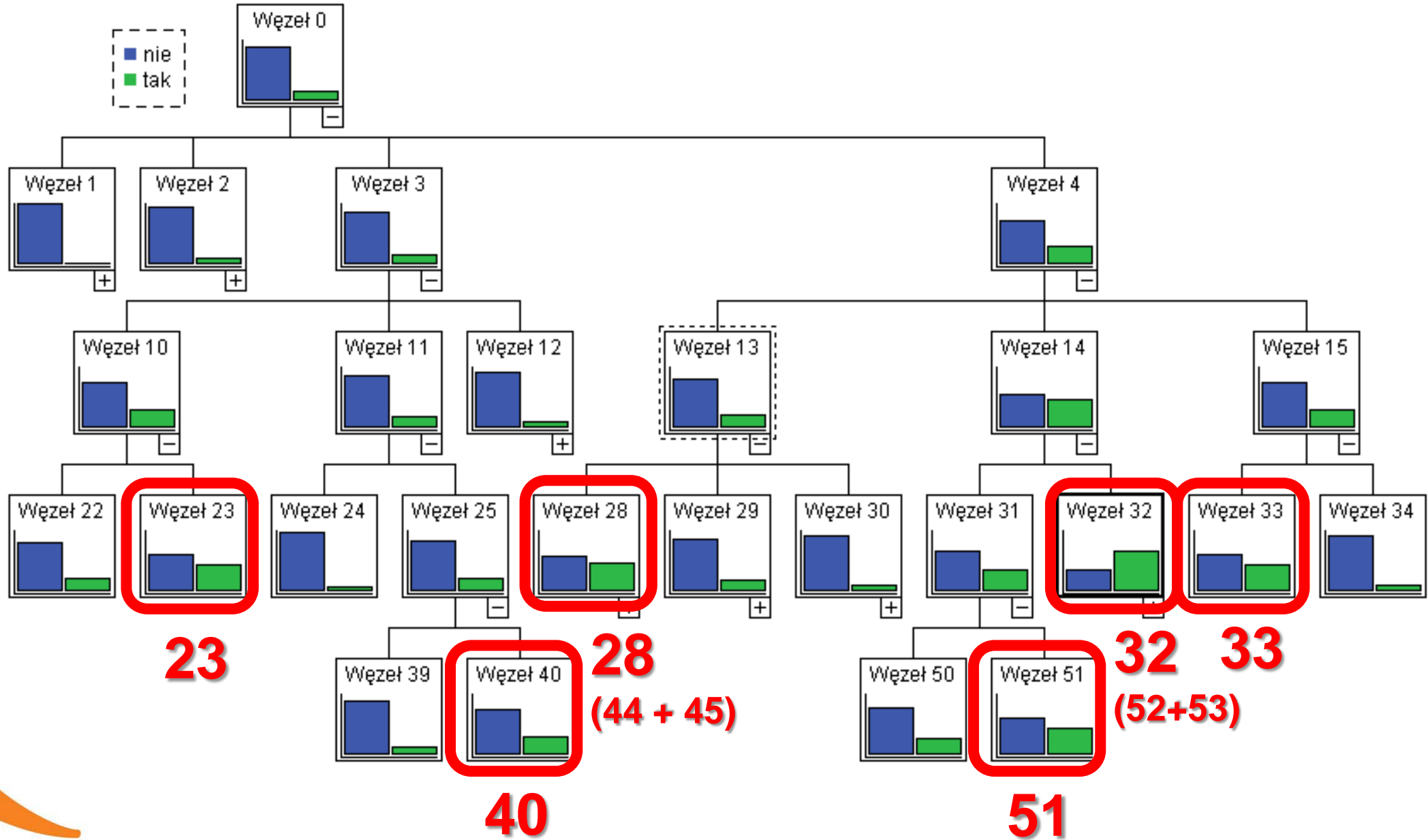
Węzeł	Węzeł		Korzyść		Odpowiedzi	Indeks
	N	Procent	N	Procent		
53	128	1,7%	99	10,0%	77,3%	603,7%
45	158	2,0%	89	9,0%	56,3%	439,7%
52	101	1,3%	51	5,2%	50,5%	394,1%
33	151	2,0%	63	6,4%	41,7%	325,7%
51	209	2,7%	87	8,8%	41,6%	324,9%
23	168	2,2%	69	7,0%	41,1%	320,6%
44	148	1,9%	48	4,9%	32,4%	253,2%
40	203	2,6%	56	5,7%	27,6%	215,3%
50	166	2,2%	40	4,0%	24,1%	188,1%

Algorytm budowy drzewa: Wyczerpujący CHAID

Zmienna zależna: Niskie osiągnięte wykształcenie (gimnazjalne lub niższe)







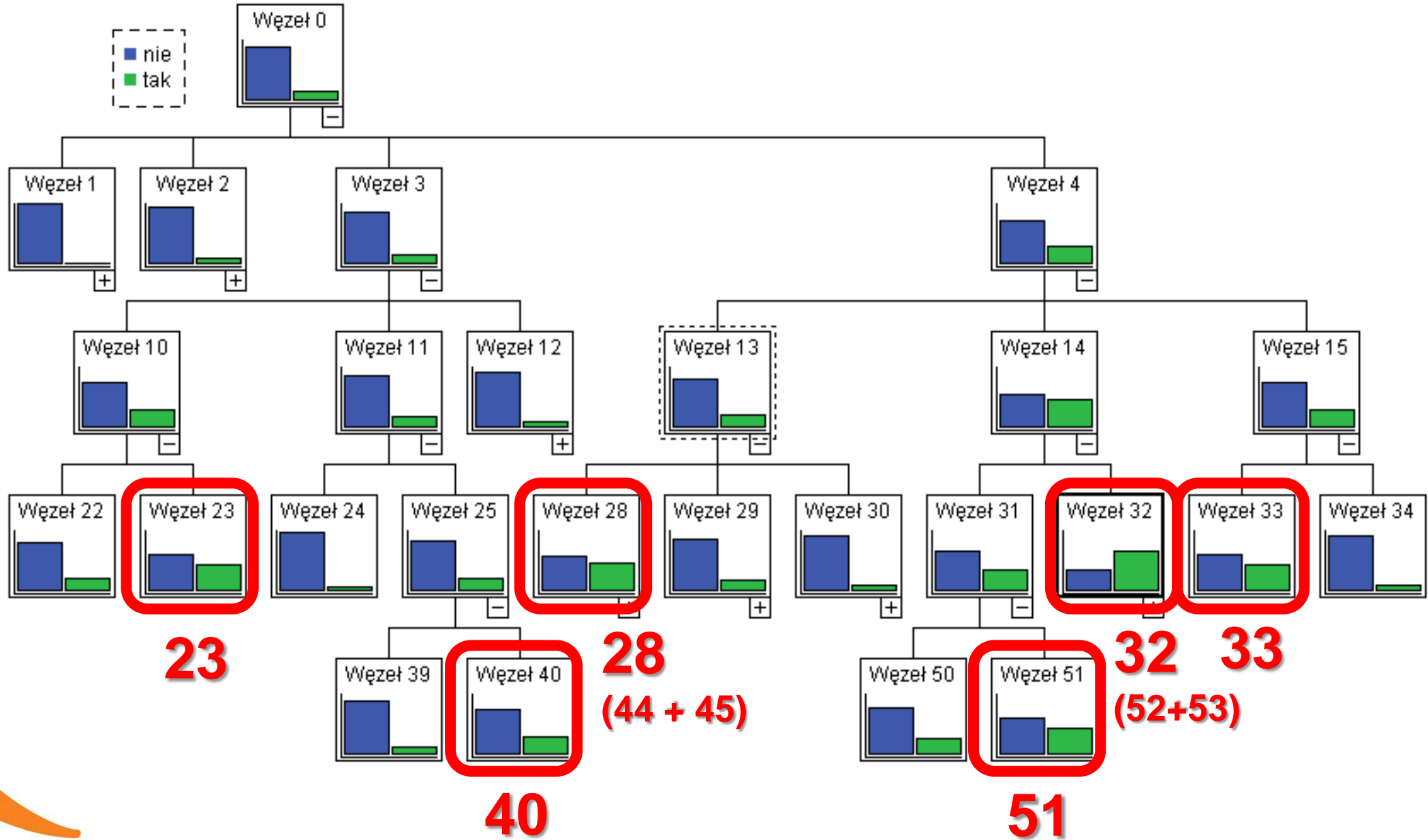
Zależna: niskie osiągnięte wykształcenie

- Pracownicy niewykwalifikowani
- Bez małżonka lub partnera
- Bez doświadczenia w pracy → **65,5%**

Węzeł numer 32

Węzeł 32		
Kategoria	%	n
nie	34,5	79
tak	65,5	150
Ogółem	3,0	229

Węzeł 52			Węzeł 53		
Kategoria	%	n	Kategoria	%	n
nie	49,5	50	nie	22,7	29
tak	50,5	51	tak	77,3	99
Ogółem	1,3	101	Ogółem	1,7	128

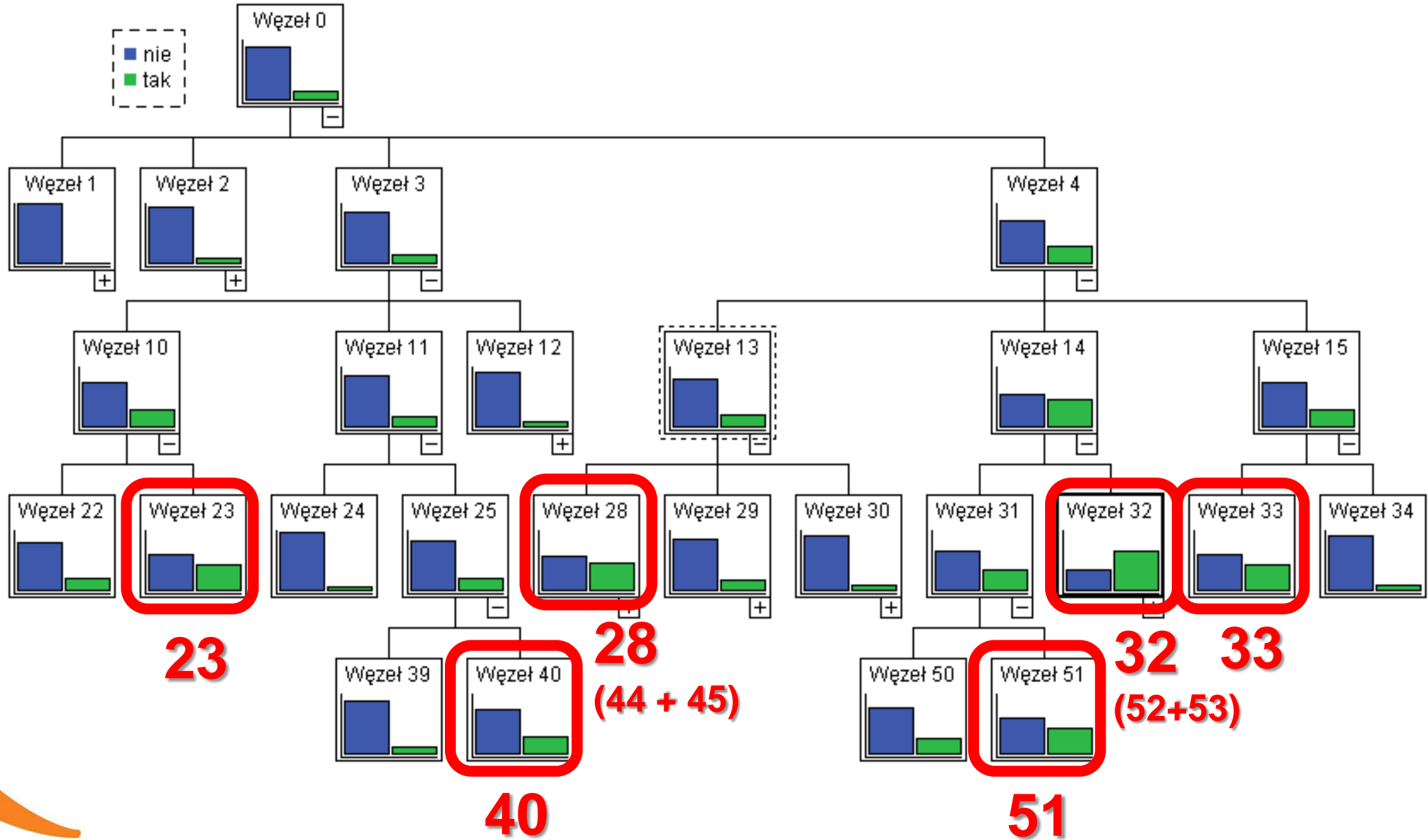


Zależna: niskie osiągnięte wykształcenie

- Pracownicy niewykwalifikowani
- Mieszkający samotnie
- Posiadający do 25 książek → **65,5%**

Węzeł numer 33

Węzeł 33		
Kategoria	%	n
nie	58,3	88
tak	41,7	63
Ogółem	2,0	151



Zależna: niskie osiągnięte wykształcenie

- Pracownicy niewykwalifikowani
- Bez pracy przez okres ostatnich 12 miesięcy lub pracujący obecnie (lub w ostatnich 12 miesiącach)
- Mieszkający samotnie
- Mężczyzna → **41,6%**

Węzeł numer 51

Węzeł 51		
Kategoria	%	n
nie	58,4	122
tak	41,6	87
Ogółem	2,7	209



Zależna: niskie osiągnięte wykształcenie

- Pracownicy niewykwalifikowani
- Mieszkający z małżonkiem lub partnerem
- Posiadający 10 książek lub mniej → **44,8%**

Węzeł numer 28

Węzeł 28		
Kategoria	%	n
nie	55,2	169
tak	44,8	137
Ogółem	4,0	306

Węzeł 44		
Kategoria	%	n
nie	67,6	100
tak	32,4	48
Ogółem	1,9	148

Węzeł 45		
Kategoria	%	n
nie	43,7	69
tak	56,3	89
Ogółem	2,0	158

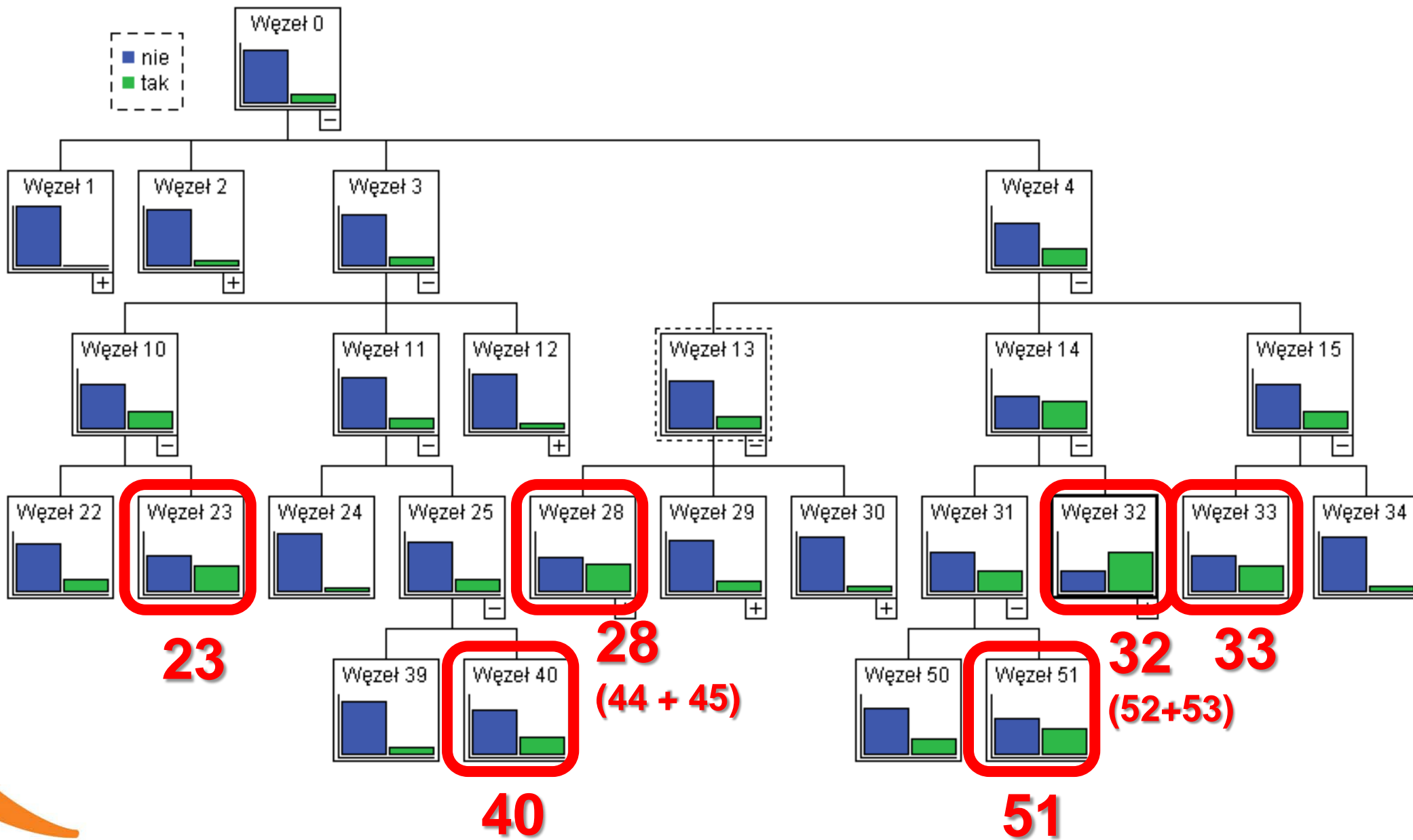


Zależna: niskie osiągnięte wykształcenie

- Wykwalifikowani pracownicy fizyczni
- Posiadający między 10 a 25 książek
- Dobry, przeciętny lub kiepski stan zdrowia
- Dochody roczne powyżej 50 centyla → **27,6%**

Węzeł numer 40

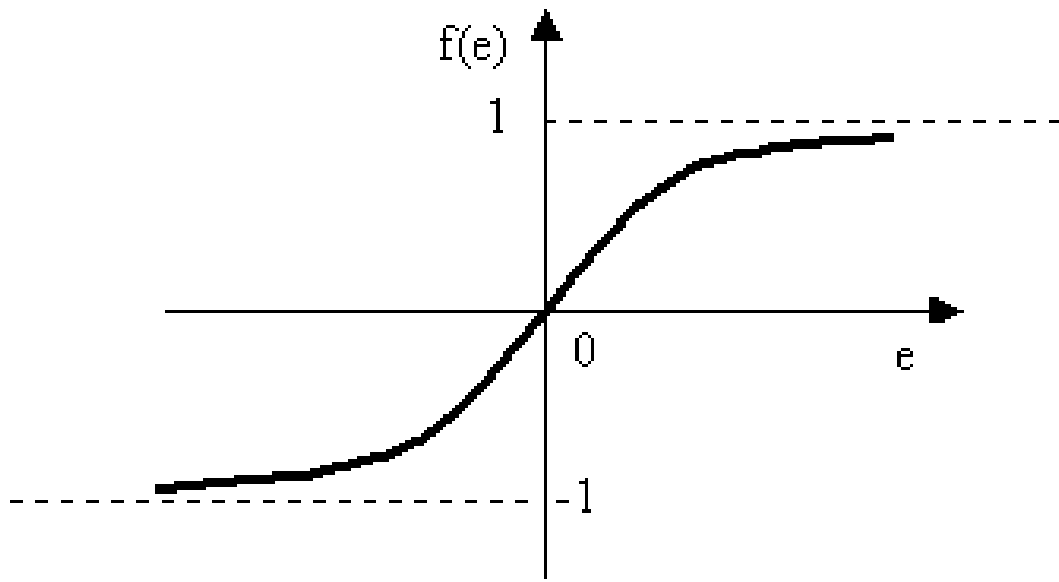
Węzeł 40		
Kategoria	%	n
nie	72,4	147
tak	27,6	56
Ogółem	2,6	203



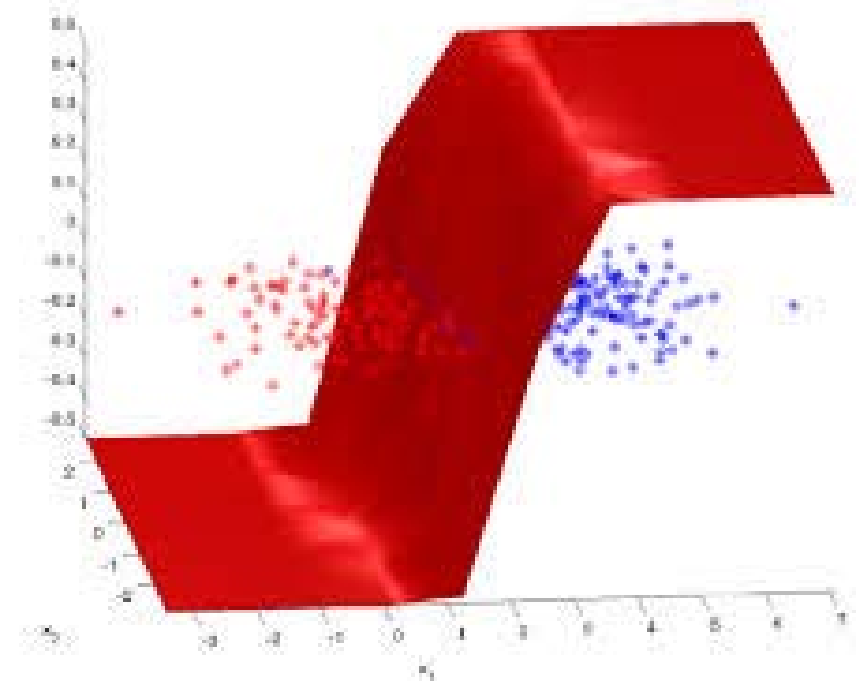
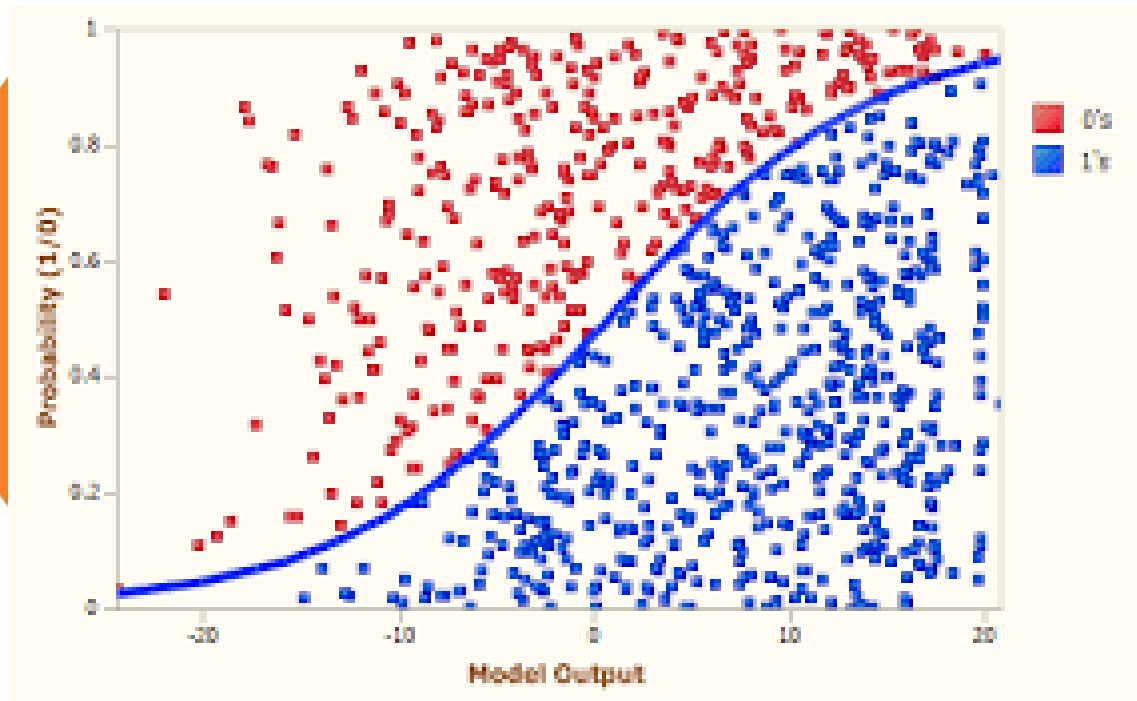
Zależna: niskie osiągnięte wykształcenie

- Wykwalifikowani pracownicy fizyczni
- Posiadający 10 książek lub mniej
- W wieku powyżej 52 roku życia → **41,1%**

Węzeł 23		
Kategoria	%	n
nie	58,9	99
tak	41,1	69
Ogółem	2,2	168



REGRESJA LOGISTYCZNA



Opis analizy

Regresja logistyczna (STATA, logit)

Przeprowadzona z uwzględnieniem złożonego charakteru próby i wag replikacyjnych (STATA, svy)

Zmienna zależna: Niskie osiągnięte wykształcenie (gimnazjalne lub niższe); wykluczony (1), nie wykluczony (0)



Zmienne niezależne:

- generalna klasyfikacja zawodu respondenta,
- zamieszkiwanie (samotne, bez partnera, z partnerem),
- typ gminy – miejsca zamieszkania respondenta,
- liczba dzieci,
- stan zdrowia,
- roczne dochody,
- dział gospodarki miejsca obecnej lub ostatniej pracy respondenta,
- płeć,
- liczba książek,
- wiek,
- status społeczno-ekonomiczny respondenta,
- wykształcenie ojca respondenta,
- status na rynku pracy

Zmienne niezależne

Z powodu współliniowości z modelu zostały usunięte dwie zmienne, które na etapie analizy drzew decyzyjnych okazały się istotne przy kształtowaniu :

- doświadczenie zawodowe,

wykształcenie matki respondenta



Wyniki analizy regresji

IBE 

zmienna	oddziaływanie
generalna klasyfikacja zawodu respondenta	✓
zamieszkiwanie (samotne, bez partnera, z partnerem)	✓
typ gminy – miejsca zamieszkania respondenta	⊗
liczba dzieci	⊗
stan zdrowia	⊗
roczne dochody	✓
dział gospodarki miejsca obecnej lub ostatniej pracy respondenta	✓
płeć	✓
liczba książek	✓
wiek	⊗
status społeczno-ekonomiczny respondenta	⊗
wykształcenie ojca respondenta	✓
status na rynku pracy	✓

Wyniki analizy regresji

IBE 

zmienna	oddziaływanie
generalna klasyfikacja zawodu respondenta	✓
zamieszkiwanie (samotne, bez partnera, z partnerem)	✓
typ gminy – miejsca zamieszkania respondenta	⊗
liczba dzieci	⊗
stan zdrowia	⊗
roczne dochody	✓
dział gospodarki miejsca obecnej lub ostatniej pracy respondenta	✓
płeć	✓
liczba książek	✓
wiek	⊗
status społeczno-ekonomiczny respondenta	⊗
wykształcenie ojca respondenta	✓
status na rynku pracy	✓



generalna klasyfikacja zawodu respondenta (ref. „specjalista”)	b	P>t
wykwalifikowani pracownicy niefizyczni	1.24	0.006
wykwalifikowani pracownicy fizyczni	1.68	0.000
pracownicy niewykwalifikowani	2.25	0.000

Wyniki analizy regresji

zmienna	oddziaływanie
generalna klasyfikacja zawodu respondenta	✓
zamieszkiwanie (samotne, bez partnera, z partnerem)	✓
typ gminy – miejsca zamieszkania respondenta	⊗
liczba dzieci	⊗
stan zdrowia	⊗
roczne dochody	✓
dział gospodarki miejsca obecnej lub ostatniej pracy respondenta	✓
płeć	✓
liczba książek	✓
wiek	⊗
status społeczno-ekonomiczny respondenta	⊗
wykształcenie ojca respondenta	✓
status na rynku pracy	✓



zamieszkiwanie z: (ref. „samotne”)	b	P>t
bez małżonka lub partnera	-.43	0.161
z małżonkiem lub partnerem	-1.05	0.001

Wyniki analizy regresji

zmienna	oddziaływanie
generalna klasyfikacja zawodu respondenta	✓
zamieszkiwanie (samotne, bez partnera, z partnerem)	✓
typ gminy – miejsca zamieszkania respondenta	⊗
liczba dzieci	⊗
stan zdrowia	⊗
roczne dochody	✓
dział gospodarki miejsca obecnej lub ostatniej pracy respondenta	✓
płeć	✓
liczba książek	✓
wiek	⊗
status społeczno-ekonomiczny respondenta	⊗
wykształcenie ojca respondenta	✓
status na rynku pracy	✓



centyl rocznych dochodów (ref. „do 10 centyla”)	b	P>t
od 10 do mniej niż 25	-.55	0.036
od 25 do mniej niż 50	-.90	0.001
od 50 do mniej niż 75	-1.18	0.010
od 75 do mniej niż 90	-1.80	0.156
90 lub wyższy	-.37	0.535

Wyniki analizy regresji

zmienna	oddziaływanie
generalna klasyfikacja zawodu respondenta	✓
zamieszkiwanie (samotne, bez partnera, z partnerem)	✓
typ gminy – miejsca zamieszkania respondenta	⊗
liczba dzieci	⊗
stan zdrowia	⊗
roczne dochody	✓
dział gospodarki miejsca obecnej lub ostatniej pracy respondenta	✓
płeć	✓
liczba książek	✓
wiek	⊗
status społeczno-ekonomiczny respondenta	⊗
wykształcenie ojca respondenta	✓
status na rynku pracy	✓



dział gospodarki miejsca obecnej lub ostatniej pracy respondenta (ref. „Rolnictwo”)	b	P>t
Przemysł	.47	0.259
Usługi tradycyjne	-.01	0.979
Usługi nowoczesne	.03	0.968
Administracja publiczna	.31	0.590
Edukacja	.09	0.880
Zdrowie	1.13	0.090
Inne	1.34	0.010

Wyniki analizy regresji

zmienna	oddziaływanie
generalna klasyfikacja zawodu respondenta	✓
zamieszkiwanie (samotne, bez partnera, z partnerem)	✓
typ gminy – miejsca zamieszkania respondenta	⊗
liczba dzieci	⊗
stan zdrowia	⊗
roczne dochody	✓
dział gospodarki miejsca obecnej lub ostatniej pracy respondenta	✓
płeć	✓
liczba książek	✓
wiek	⊗
status społeczno-ekonomiczny respondenta	⊗
wykształcenie ojca respondenta	✓
status na rynku pracy	✓



płeć: (ref. „mężczyzna”)	b	P>t
kobieta	-.64	0.015

Wyniki analizy regresji

zmienna	oddziaływanie
generalna klasyfikacja zawodu respondenta	✓
zamieszkiwanie (samotne, bez partnera, z partnerem)	✓
typ gminy – miejsca zamieszkania respondenta	⊗
liczba dzieci	⊗
stan zdrowia	⊗
roczne dochody	✓
dział gospodarki miejsca obecnej lub ostatniej pracy respondenta	✓
płeć	✓
liczba książek	✓
wiek	⊗
status społeczno-ekonomiczny respondenta	⊗
wykształcenie ojca respondenta	✓
status na rynku pracy	✓



liczba książek (ref. „do 10”)	b	P>t
11 do 25 książek	-.70	0.003
26 od 100 książek	-.80	0.001
101 do 200 książek	-1.12	0.005
201 do 500 książek	-1.34	0.029
Ponad 500 książek	-2.57	0.000

Wyniki analizy regresji

zmienna	oddziaływanie
generalna klasyfikacja zawodu respondenta	✓
zamieszkiwanie (samotne, bez partnera, z partnerem)	✓
typ gminy – miejsca zamieszkania respondenta	⊗
liczba dzieci	⊗
stan zdrowia	⊗
roczne dochody	✓
dział gospodarki miejsca obecnej lub ostatniej pracy respondenta	✓
płeć	✓
liczba książek	✓
wiek	⊗
status społeczno-ekonomiczny respondenta	⊗
wykształcenie ojca respondenta	✓
status na rynku pracy	✓

wykształcenie ojca respondenta (ref. „niepełne zawodowe lub niższe”)	b	P>t
Zasadnicze zawodowe	-.62	0.009
Średnie maturalne lub pomaturalne	-.51	0.138
Wyższe	-.39	0.516

Wyniki analizy regresji

zmienna	oddziaływanie
generalna klasyfikacja zawodu respondenta	✓
zamieszkiwanie (samotne, bez partnera, z partnerem)	✓
typ gminy – miejsca zamieszkania respondenta	⊗
liczba dzieci	⊗
stan zdrowia	⊗
roczne dochody	✓
dział gospodarki miejsca obecnej lub ostatniej pracy respondenta	✓
płeć	✓
liczba książek	✓
wiek	⊗
status społeczno-ekonomiczny respondenta	⊗
wykształcenie ojca respondenta	✓
status na rynku pracy	✓



status na rynku pracy (ref. „pracujący lub uczący się”)	b	P>t
bezrobotny	.75	0.120
emerytura	-.19	0.785
niepełnosprawny	-.39	0.634
zajmuje się domem/dziećmi	1.73	0.003
inny	-.28	0.722

Pracownicy produkcyjni lub administracyjni niższego szczebla			wykwalifikowani pracownicy fizyczni			
Bez małżonka lub partnera	Mieszkający samotnie	z małżonkiem lub partnerem	Bez pracy przez ostatni rok lub pracujący obecnie (lub w ostatnim roku)	Posiadający 10 książek lub mniej	Posiadający między 10 a 25 książek	
Bez doświadczenia w pracy → 65,5%	Posiadający do 25 książek → 65,5%	Posiadający 10 książek lub mniej → 44,8%	Mieszkający samotnie	W wieku powyżej 52 roku życia → 41,1%	Dobry, przeciętny lub kiepski stan zdrowia	
			Mężczyzna → 41,6%		Dochody roczne powyżej 50 centyla → 27,6%	
3% populacji	2% populacji	2,7% populacji	3,9% populacji	2,6% populacji	2,2% populacji	Łącznie 16% osób w wieku 18-65 lat
15,2% grupy wykluczonej	6,4% grupy wykluczonej	8,8% grupy wykluczonej	13,9% grupy wykluczonej	5,7% grupy wykluczonej	7,0% grupy wykluczonej	Łącznie 57 % osób wykluczonych w wieku 18-65 lat

Ponad 3,4 razy większa szansa na znalezienie w tych grupach osób wykluczonych ze względu na niskie umiejętności, niż w całej populacji

Wnioski

Poza czynnikami związanymi z:

- generalną klasyfikacją zawodu respondenta
- z kim się zamieszkuje
- liczbą książek w domu
- płcią
- doświadczeniem zawodowym oraz
- wiekiem,
- zdrowiem i
- rocznymi dochody

Wnioski

Modele regresji logistycznej wskazują na istotne znaczenie również:

- działu gospodarki miejsca obecnej lub ostatniej pracy respondenta
- wykształcenie ojca respondenta oraz
- statusu na rynku pracy

