



Prediktivni model segmentacije tržišta: Primena modela logističke regresije i CHAID procedure

Jasna Soldić-Aleksić*

Rezime: Segmentacija tržišta predstavlja jedan od najvažnijih koncepata savremenog marketinga. Cilj segmentacije jeste podela tržišta na grupe - segmente kupaca koji imaju slične karakteristike, potrebe, želje i/ili slično tržišno ponašanje. Na osnovu dobijenih segmenata kompanije mogu razvijati specifične marketing planove za svaki segment posebno, i na taj načini steći konkurentsku prednost, kako na kratak, tako i na dugi rok. U zavisnosti od cilja segmentacije mogu se primeniti različite procedure i tehnike. U ovom radu predstavljen je prediktivni model segmentacije tržišta koji se bazira na primeni modela logističke regresije i CHAID analize. Model logističke regresije je primenjen samo za potrebe izdvajanja iz datog skupa promenljivih onih promenljivih koje su statistički značajne u objašnjavanju zavisne promenljive. Zatim su tako izdvojene promenljive uključene u CHAID analizu kao prediktor promenljive. Na konkretnom empirijskom primeru prikazane su mogućnosti CHAID analize u generisanju prediktivnog modela segmentacije tržišta. Rezultati modeliranja predstavljani su u formi: tabele sumarnih rezultata modela, CHAID drveta, dijagrama dobiti (gain chart) i indeks dijagrama, kao i tabele rizika i klasifikacione tabele, koje otkrivaju prediktivnu moć modela.

Ključne reči: segmentacija tržišta, prediktivni modeli segmentacije, CHAID modeli, logistička regresija

1. UVOD

Segmentacija tržišta je jedan od ključnih elemenata savremenog marketinga. Danas je teško naći tržište koje nije segmentirano, jer moto „*jedan prouzvod za sve*“ jednostavno više ne funkcioniše. Kako se na tržištu pojavljuju kupci koji nisu homogena grupa, već se razlikuju po svojim preferencijama, željama, potrebama, karakteristikama, od segmentacije se očekuje da doprinese podeli kupaca na relativno homogene grupe, koje pokazuju slično tržišno ponašanje. U marketinškoj literaturi autori navode različite karakteristike (dimenzije) kupaca koje se koriste za potrebe segmentacije (Levin, Zahavi, 2001, str.2-22): socio-ekonomske, psihološke, geografske, nivo korišćenja proizvoda/usluga, koristi od proizvoda/usluga i dr. Stoga, segmentacija čini ključnu komponentu marketinškog procesa koja vodi ka targetiranju tržišta i pozicioniranju organizacija na njemu.

Segmentacija je proces koji se bazira na informacijama, tako da se može prihvatiti princip da, ukoliko se raspolaže sa većom količinom informacija utoliko su veća očekivanja da će rezultati procesa segmentacije biti precizniji. Ovo se posebno odnosi na slučaj marketinga zasnovanog na bazama podataka, gde se segmentacija tržišta koristi kao standardan marketinški alat. Da bi segmentacija bila uspešna potrebno je da su ispunjeni izvesni uslovi, kao što su: potrebe kupaca na tržištu su različite; informacije o kupcima (najčešće smeštene u marketinškim bazama podataka) su na raspolaganju; postoje načini da se izmere transakcije ili potencijalne transakcije pojedinih grupa kupaca; postoji ekonomski način da se dopre do segmenata kupaca sa određenim marketinškim programom.

2. TEHNIKE SEGMENTACIJE TRŽIŠTA

U zavisnosti od konkretnog cilja istraživanja mogu se primeniti različite šeme i tehnike segmentacije. *Wedel* i *Kamakura* navode četiri grupe metoda segmentacije tržišta (Galguera, Luna, Mendez, 2006, str. 459-479):

- *A priori* deskriptivne metode (npr. unakrsno tabeliranje);

* Rad je pripremljen pod pokroviteljstvom Ministarstva za nauku i tehnološki razvoj Republike Srbije u okviru projekta „Razvoj institucija i instrumenata hipotekarnog i finansijskog tržišta u Srbiji“.

- *A priori* prediktivne metode (npr. diskriminaciona analiza, logistička regresija);
- *Post hoc* deskriptivne metode (npr. klaster analiza);
- *Post hoc* prediktivne metode (npr. CHAID, CART analiza);

U okviru prethodne klasifikacije posebno je značajna podela na deskriptivne i prediktivne metode. Osnovna razlika između ove dve grupe metoda ogleda se u cilju istraživanja – da li je cilj istraživanja deskripcija pojava ili predviđanje jedne pojave na osnovu drugih pojava, odnosno karakteristika. Na metodološkom planu istraživanja ova se razlika odnosi na različiti tretman promenljivih koje predstavljaju pojedine pojave: kod deskriptivnih tehnika sve promenljive imaju isti status, dok se kod prediktivnih tehnika pravi razlika između zavisne promenljive (koja najčešće odražava pripadnost opservacija pojedinim grupama – segmentima) i nezavisnih – prediktor promenljivih.

Već smo ukazali da je pitanje segmentacije, i sa teorijskog i praktičnog aspekta, jedno od ključnih pitanja u domenu marketinga, a posebno direktnog marketinga ili marketinga zasnovanog na bazama podataka. U ovoj oblasti najčešće se koriste analiza frekvencija i analiza unakrsnog tabeliranja, koje su se pokazale kao efikasne tehnike segmentacije baze podataka kupaca. Takođe, primenjuju se i RFM (*Recency Frequency and Monetary values*), FRAC (*Frequency, Recency, Amount of money and Category of product*) analize, koje pripadaju grupi tehnika koje se označava kao metode zasnovane na proceni (*judgmentally based method*). Analiza na bazi unakrsnog tabeliranja (*cross-tabulation*) može se značajno unaprediti primenom CHAID analize u smislu određivanja karakteristika (atributa) koje najviše doprinose segmentaciji, kao i za određivanje granica podela između ćelija unakrsne tabele. Vredno je napomenuti da pored navedenih tehnika, koje se ubrajaju u klasične tehnike segmentacije, posebno za velike baze podataka (koje sadrže desetina miliona slogova) primenjuju se i mnoge druge sofisticirane tehnike segmentacije iz domena pretraživanja podataka (*data mining*) i mašinskog učenja.

3. PREDIKTIVNI MODELI: LOGISTIČKA REGRESIJA I CHAID MODEL

U ovom radu je prikazan jedan konkretan primer segmentacije zasnovan na modelima logističke regresije i CHAID analize. Metodološki posmatrano jedan od važnih ciljeva segmentacije jeste da se što preciznije

otkriju promenljive koje mogu da ukažu na različite segmente kupaca. U praksi se, pak, segmentacija može zasnivati na izuzetno velikom broju promenljivih, metričkih ili nemetričkih, koje mogu imati veliki broj mogućih modaliteta. Stoga se kao važno postavlja pitanje izdvajanja promenljivih (karakteristika) koje objektivno imaju najznačajniji uticaj na proces segmentacije. U tu svrhu smo u radu najpre iz niza od jedanaest nezavisnih promenljivih uz pomoć logističke regresije izdvojili samo one promenljive čiji je uticaj na zavisnu promenljivu statistički značajan, da bi zatim tako izdvojene promenljive uključili u CHAID analizu kao prediktor promenljive.

Pre nego što predstavimo rezultate empirijske analize, sažeto ćemo prikazati osnovne karakteristike navedenih modela.

3.1. Logistička regresija

Logistička regresija je poznata tehnika iz grupe prediktivnih tehnika, gde je zavisna promenljiva najčešće binarna (ima dva modaliteta, odnosno kategorije), a nezavisne promenljive (jedna ili više) mogu biti kategorijske i/ili metričke. Pored toga što ovaj model daje odgovor na pitanje pripadnosti svake opservacije jednoj od dve grupe, takođe, daje i odgovarajuće verovatnoće pripadnosti. Ovo je posebno interesantno u marketinškim istraživanjima kada kompanija sprovodi istraživanje sa ciljem da predvidi da li će konkretan potrošač kupiti ili neće proizvod, da li će ponovo koristiti našu uslugu ili ne i sl. U tom slučaju zavisna promenljiva je binarna (šifre 0 i 1), gde 1 predstavlja klasu koja nas interesuje (na primer, 1 predstavlja klasu potrošača koji će kupiti proizvod, ili ponovo koristiti našu uslugu, isl.), a 0 suprotnu klasu opservacija. Model logističke regresije predviđa verovatnoću da će zavisna promenljiva uzeti vrednost 1. U tu svrhu koristi se sledeća logistička funkcija (Myatt, Johnson, 2009, str.151):

$$P(y = 1) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \dots + \beta_k x_k)}}$$

gde je β_0 konstanta, a $\beta_1, \beta_2, \dots, \beta_k$ su koeficijenti za k nezavisnih promenljivih. Zapažamo da je veza između zavisne i nezavisnih promenljivih *nelinearna*. Ali ako formiramo količnik šanse (*odds ratio*) za pripadnost opservacije jednoj grupi, i to na sledeći način:

$$odds = \frac{P(y = 1)}{1 - P(y = 1)}$$

odnosno

$$odds = \frac{P(y=1)}{P(y=0)}$$

imamo mogućnost da logističku funkciju predstavimo na sledeći način:

$$odds = e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \dots + \beta_k x_k}$$

odnosno u linearnom obliku

$$\log(odds) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \dots + \beta_k x_k$$

Ocenjeni koeficijenti β_i imaju jednostavnu interpretaciju. Za jediničnu promenu nezavisne promenljive x_i pod pretpostavkom da se ostale nezavisne promenljive ne menjaju, šansa za pripadnost opservacije grupi koja nas interesuje povećava se za faktor e^{β_i} . U postupku ocenjivanja koeficijenata β_i vrši se testiranje statističke značajnosti ovih koeficijenata. Primenom tehnike uključenja nezavisnih promenljivih u model u koracima (*stepwise logistic regression*) u modelu se zadržavaju samo nezavisne promenljive čiji je uticaj na segmentaciju statistički značajan. U empirijskom delu rada primenili smo upravo ovu tehniku postepeno uključenja nezavisnih promenljivih u model logističke regresije, kako bismo otkrili značajne nezavisne promenljive.

3.2. CHAID analiza

CHAID (*Chi Square Automatic Interaction Detection*) analiza pripada grupi tehnika zavisnosti iz oblasti multivarijacione analize podataka. Zavisna promenljiva je kategorijska promenljiva, a nezavisne – prediktor promenljive mogu biti kategorijske i/ili metričke. U slučaju da su metričke promenljive vrši se njihovo automatsko prevođenje u kategorijske promenljive. Ova tehnika predstavlja *chi-square* verziju AID (*Automatic Interaction Detection*) tehnike, koja ima za cilj da automatski otkrije veze između promenljivih. Kao što to ime tehnike upućuje ovde se vrši izdvajanje nezavisnih – prediktor promenljivih na bazi *hi-kvadrat* testa.

U postupku CHAID analize vrši se rekurzivno deljenje datog skupa podataka na posebne podgrupe (segmente) uz primenu algoritma koji minimizira varijacije zavisne promenljive unutar segmenata, odnosno maksimizira između segmenata. Ovo rekurzivno deljenje se vrši kroz tri koraka: udruživanje, podela i zaustavljanje (*merging, splitting and stopping*), a krajnji

rezultat je CHAID drvo. Ono se razvija uzastopnim ponavljanjem navedenih koraka na svakom segmentu – čvoru drveta, počev od korenog čvora. Osnovni kriterijum koji se koristi u navedenoj proceduri je *hi-kvadrat* statistika i njoj pridružena *p*-vrednost koja se računa za svaku prediktor promenljivu, a kojom se testira (ne)zavisnost prediktor i zavisne promenljive. Promenljiva koja ima najnižu *p*-vrednost bira se kao prva prediktor promenljiva, jer se nalazi u najjačoj interakciji sa zavisnom promenljivom. Ako je data *p*-vrednost manja ili jednaka od unapred definisane α vrednosti (uobičajeno je $\alpha=0,5$), vrši se podela čvora (grupe) koristeći datu prediktor promenljivu, u protivnom se dati čvor ne deli, već se smatra da se radi o završnom čvoru (*terminal node*). Za svaku prediktor promenljivu vrši se udruživanje njenih kategorija ako njihova razlika nije statistički značajna s obzirom na zavisnu promenljivu. Kako se *hi-kvadrat* test primenjuje na kategorijske promenljive, jasno je da klasična CHAID analiza podrazumeva da su ulazne promenljive takođe kategorijske. Naravno, neprekidne promenljive mogu se pojaviti kao ulazne promenljive, pošto su prethodno prekodirane na kategorijske promenljive. Danas mnogi računarski programi (*SAS Enterprise Miner*, *SPSS* i dr.) koriste *hi-kvadrat* test za formiranje podela za kategorijske promenljive, a *F* test za kreiranje podela za neprekidne promenljive.

Osnovni rezultat ove tehnike je klasifikaciono drvo – CHAID drvo, koje je veoma jednostavno za interpretaciju. „Stablo“ ovog drveta predstavlja celinu skupa podataka, prvi sloj grana (čvorovi – podgrupe – segmenti) prikazuje kategorije nezavisne promenljive koja je najjači prediktor zavisne promenljive, odnosno pripadnosti određenoj grupi. CHAID automatski (kroz iterativnu proceduru) određuje kako će se grupisati vrednosti prediktor promenljivih, tako da se na drvetu pojavljuje onaj broj grupa – čvorova koji se statistički značajno razlikuju.

Zbog jednostavne vizuelne interpretacije rezultata CHAID tehnika je veoma zastupljena u oblasti segmentacije tržišta. U slučaju kada se koristi za generisanje prediktivnog modela segmentacije tržišta, finalni marketinški segmenti su određeni završnim (terminalnim) čvorovima CHAID drveta.

4. EMPIRIJSKA ANALIZA

Cilj analize – Kao što je u uvodnom delu rada naglašeno u fokusu našeg istraživanja je prediktivna segmentacija, što znači da postoji kategorija - grupa zavisne promenljive koja je od posebnog interesa za

marketinške istraživače, a u odnosu na koju se sprovodi segmentacija tržišta. Naš primer se odnosi na generisanje prediktivnog modela segmentacije tržišta posetilaca jednog restorana brze hrane (Izvor: www.mhhe.com.hair06). Menadžment restorana je sproveo istraživanje među posetiocima restorana u vezi njihovog zadovoljstva posetom restorana i elementima koji imaju uticaja na njihovu odluku da ponovo posete restoran. Kako su njihovi odgovori na pitanje da li će ponovo posetiti restoran u formi „da“ / „ne“, menadžment je zainteresovan da što detaljnije upozna svoje posetioce, posebno sa stanovišta njihove odluke da ponovo posete restoran. Za preduzimanje odgovarajućih marketinških odluka važni su i drugi segmenti – posetioци koji su se izjasnili da neće ponovo posetiti restoran, kao i njihove karakteristike.

Opis podataka – Istraživanje je obuhvatilo 400 posetilaca restorana. Pored pitanja da li će ponovo posetiti restoran, koje metodološki predstavlja zavisnu promenljivu, postavljena su i mnoga druga pitanja koja predstavljaju nezavisne promenljive. Zavisna promenljiva – „posetilac će ponovo posetiti restoran“ ima dva modaliteta: „da“ (šifra 1) i „ne“ (šifra 0). Za istraživanje je od interesa kategorija da će posetioци ponovo doći u restoran, sa oznakom 1. Nezavisne promenljive su mešavina kategorijskih i metričkih promenljivih:

- X8 – posetilac jede umereno i zdravu hranu (1-7 intervalna skala: 1 – izrazito neslaganje, 7- izrazito slaganje);
- X11 – posetilac rado proba nove brendove (1-7 intervalna skala: 1 – izrazito neslaganje, 7- izrazito slaganje);
- X14 – u restoranu se služe velike porcije (1-7 intervalna skala: 1 – izrazito neslaganje, 7- izrazito slaganje);
- X16 – u restoranu su razumne cene (1-7 intervalna skala: 1 – izrazito neslaganje, 7- izrazito slaganje);
- X17 – atraktivan prostor (1-7 intervalna skala: 1 – izrazito neslaganje, 7- izrazito slaganje);
- X18 – odličan ukus hrane (1-7 intervalna skala: 1 – izrazito neslaganje, 7- izrazito slaganje);
- X21 – u restoranu je brza usluga (1-7 intervalna skala: 1 – izrazito neslaganje, 7- izrazito slaganje);
- X32 – pol posetioca (0-muški, 1-ženski);
- X33 – broj dece posetioca (1: „bez dece“, 2: „1-2 deteta“, 3: „više od 2 deteta“);
- X34 – starost posetioca (1: „18-25“, 2: „26-34“, 3: „35-49“, 4: „50-59“, 5: „60 i više“);

- X35 – dohodak posetioca (1: „15000-30000“, 2: „30001-50000“, 3: „50001-75000“, 4: „75001-100000“, 5: „100001 i više“).

U nastavku su prikazani rezultati primene modela logističke regresije i CHAID modela.

4.1. Primena modela logističke regresije

Kako je naš cilj primene modela logističke regresije usmeren na izdvajanje statistički značajnih prediktor promenljivih, ovde ćemo prikazati samo za tu svrhu relevantne izlazne rezultate modeliranja. Naime, u sledećim tabelama dati su sumarni rezultati primene logističkog modela i procedura uključenja nezavisnih promenljivih u model u koracima (*stepwise logistic procedure*).

TABELA 1. Sumarni rezultati

Korak	-2 Log Verovatnoća	Cox & Snell R kvadrat	Nagelkerke R kvadrat
1	375,428(a)	,337	,455
2	327,925(a)	,411	,556
3	305,058(b)	,444	,600
4	287,970(b)	,467	,631
5	273,836(b)	,486	,656
6	267,391(b)	,494	,667
7	260,465(b)	,503	,679

Na osnovu podataka u tabeli 2 vidimo da je od ukupno 11 nezavisnih promenljivih koje su bile „kandidati“ za uključenje u model, u poslednjem, sedmom, koraku procedure u model uključeno sedam promenljivih: X8, X16, X17, X18, X33, X34 i X35. U nastavku analize koristićemo samo ove promenljive i to kao prediktor promenljive u CHAID modelu.

4.2. Primena CHAID modela

Rezultati primene CHAID modela dati su u formi: tabele sumarnih rezultata o modelu, CHAID drveta, dijagrama dobiti (*gains chart*) i indeks dijagrama, kao i tabele rizika i klasifikacione tabele, koje otkrivaju prediktorsku moć modela.

Sumarni elementi primenjenog CHAID modela prikazani su u tabeli 3. Zavisna promenljiva je kate-

TABELA 2. Postupak uključanja prediktor promenljivih u model

		B	S.E.	Wald	Df	Sig.	Exp(B)
Korak 1(a)	x33	1,738	,160	117,693	1	,000	5,686
	Konstanta	-3,711	,340	118,921	1	,000	,024
Korak 2(b)	x17	1,031	,164	39,410	1	,000	2,803
	x33	1,563	,173	81,890	1	,000	4,775
	Konstanta	-7,575	,778	94,846	1	,000	,001
Korak 3(c)	x8	-,614	,134	20,888	1	,000	,541
	x17	1,148	,175	43,014	1	,000	3,153
	x33	1,422	,179	63,154	1	,000	4,147
	Konstanta	-5,692	,847	45,212	1	,000	,003
Korak 4(d)	x8	-,681	,141	23,359	1	,000	,506
	x17	1,276	,186	47,130	1	,000	3,581
	x33	1,791	,216	68,615	1	,000	5,998
	x34	,704	,179	15,498	1	,000	2,021
	Konstanta	-8,860	1,235	51,439	1	,000	,000
Korak 5(e)	x8	-2,432	,531	20,944	1	,000	,088
	x17	1,498	,204	53,723	1	,000	4,472
	x18	-2,049	,587	12,196	1	,000	,129
	x33	1,895	,232	66,534	1	,000	6,650
	x34	,754	,191	15,646	1	,000	2,127
	Konstanta	7,066	4,690	2,270	1	,132	1171,263
Korak 6(f)	x8	-2,475	,554	19,967	1	,000	,084
	x16	,361	,145	6,217	1	,013	1,434
	x17	1,532	,214	51,315	1	,000	4,628
	x18	-2,139	,614	12,154	1	,000	,118
	x33	1,683	,242	48,456	1	,000	5,380
	x34	,742	,203	13,352	1	,000	2,101
	Konstanta	6,420	4,833	1,764	1	,184	613,719
Korak 7(g)	x8	-2,678	,584	21,065	1	,000	,069
	x16	,422	,154	7,478	1	,006	1,525
	x17	1,733	,248	48,682	1	,000	5,660
	x18	-2,173	,638	11,614	1	,001	,114
	x33	1,444	,254	32,223	1	,000	4,236
	x34	,525	,219	5,751	1	,016	1,690
	x35	,357	,140	6,483	1	,011	1,429
	Konstanta	6,332	4,991	1,610	1	,205	562,447

gorijska promenljiva – „posetilac će opet doći“, koja ima dve kategorije: „ne“ i „da“, pri čemu je naša ciljna kategorija - kategorija „da“. U prvom delu tabele data je specifikacija modela, a u drugom delu rezultati: prediktor promenljive koje su zadržane u modelu, broj ukupnih čvorova – segmenata, broj završnih (terminalnih) segmenata i dubina CHAID drveća.

Na slici 1 prikazano je CHAID drvo, koje je dobijeno na osnovu prethodno prikazanih parametara CHAID modela. U formiranju datog modela primenjena je tehnika validacije modela – *cross validation*, koja naizmenično izdvaja po 50 opservacija za testiranje modela, a 350 opservacija za obuku (*training*) modela. Od ukupno 14 segmenata posetilaca koje

TABELA 3. Sumarne karakteristike CHAID modela

Specifikacija	Metod rasta	CHAID
	Zavisne varijable	Doći će opet: ne, da
	Nezavisne varijable	X8 – jede umereno i zdravu hranu, X16 – razumne cene, X17 – atraktivan prostor, X18 – Odlican ukus hrane, X33 – Broj dece, X34 – Starost, X35 – Dohodak
	Validacija modela	Unakrsna ocena modela
	Maksimalna dubina drveta/ razgranatost?	7
	Minimum slučajeva u 'roditeljskim' čvorovima	40
	Minimum slučajeva u 'deci' čvorovima	10
Rezultati	Uključujući nezavisne varijable	X17 – atraktivan prostor, X33 – Broj dece, X8 – jede umereno i zdravu hranu, X35 – Dohodak, X18 – Odlican ukus hrane
	Broj čvorova	14
	Broj završnih (terminalnih) čvorova	9
	Dubina	3

formiraju date prediktor promenljive, izdvaja se 9 završnih -terminalnih segmenata, čije se karakteristike posebno analiziraju.

CHAID drvo pokazuje da se u prvom nivou nalazi promenljiva *atraktivan prostor*, što znači da ona ima najjaču moć u podeli posetilaca na dve grupe - one koji su se izjasnili da će ponovo posetiti restoran, odnosno one koji ne žele da ponovo posete restoran, a zatim u drugom nivou slede prediktor promenljive: *broj dece posetilaca, dohodak posetilaca i odličan ukus hrane u restoranu*. Promenljiva *atraktivan prostor* deli uzorak na 4 segmenta, gde je zastupljena izrazita razlika u procentu posetilaca koji su odgovorili da će opet doći u restoran: posetioци koji se izrazito ne slažu da je prostor u restoranu atraktivan (na skali 1-7 zaokružili su odgovore pod 1 ili 2) odgovorili su da će ponovo posetiti restoran u 24,8% slučajeva; oni koji su se opredelili za odgovore 3 i 4 ponovo će posetiti restoran u 12,2% slučajeva; oni koji su odgovorili sa 4 i 5 doći će ponovo u 81,2%, i najzad posetioци koji se izrazito slažu da je prostor u restoranu atraktivan (vrednosti odgovora 6 i 7) ponovo će doći u 100% slučajeva.

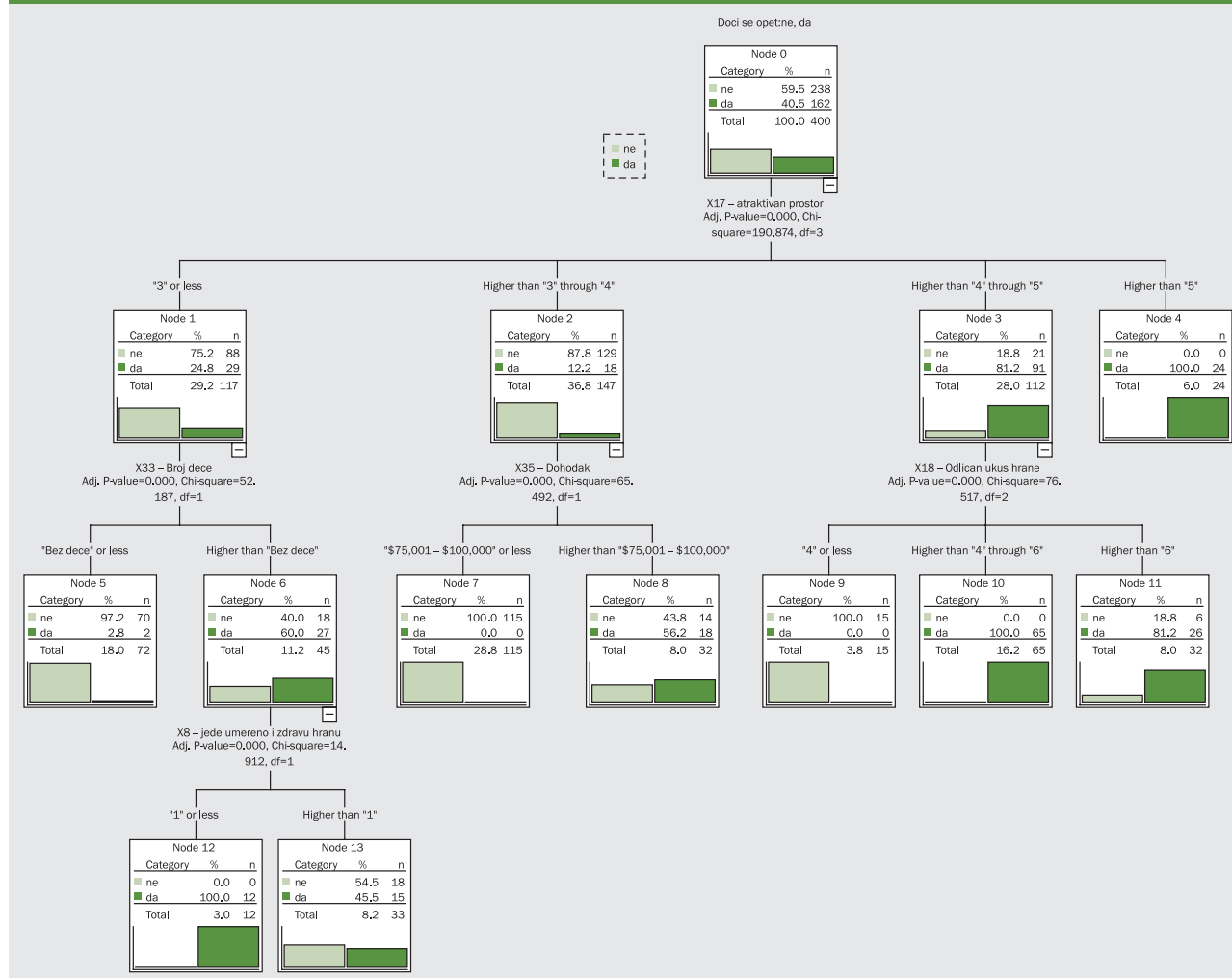
U drugom nivou segmentacije zapaža se da promenljiva *broj dece posetioca* formira dva segmenta kod posetilaca koji nisu visoko vrednovali atraktivnost prostora u restoranu: posetioци bez dece od kojih

će samo 2,8% procenata ponovo doći, dok će posetioци sa decom u 60% slučajeva ponovo doći. Druga značajna prediktor promenljiva je *dohodak posetioca*, gde se zapaža značajno veća zainteresovanost posetilaca sa dohotkom većim od \$100000 da ponovo posete restoran. Takođe, u drugom nivou kod grupe posetilaca za koje je važna atraktivnost prostora, izdvaja se značajan broj posetilaca za koje je važan i *ukus hrane* i oni su uglavnom opredeljeni da ponovo dođu u restoran. U trećem nivou CHAID drveta nalazi se promenljiva *posetilac jede umereno i zdravu hranu*. Na ovom nivou CHAID drveta posebno se izdvaja segment posetilaca koji ne vrednuju visoko atraktivnost prostora u restoranu, imaju jedno ili više dece i ne vode mnogo računa o umerenom i zdravom konzumiranju hrane. Oni su se u 100% slučajeva izjasnili da će ponovo posetiti restoran.

4.3. Performanse po završnim – terminalnim segmentima

U analizi CHAID drveta za potrebe segmentacije poseban značaj imaju završni – terminalni čvorovi, odnosno segmenti, koji predstavljaju krajnje rezultate segmentacije u skladu sa ciljnom kategorijom zavisne promenljive. U našem primeru imamo devet termi-

SLIKA 1. CHAID drvo



nalnih čvorova, odnosno segmenata, i to pod sledećim rednim brojevima: 4, 5, 7, 8, 9, 10, 11, 12 i 13. Tamno obojeni stubići na slici 1 pokazuju učešće nama interesantne kategorije posetilaca – oni koji su se izjasnili da će ponovo doći u restoran. Na osnovu visine stubića već na prvi pogled se vidi da su posebno važni segmenti pod rednim brojevima 4, 10, 11 i 12, jer je kod njih izrazito značajan procenat posetilaca koji su se izjasnili da će ponovo posetiti restoran: taj procenat je kod segmenata 4, 10 i 12 100%, a kod segmenta pod rednim brojem 11 je 81,2%. Ovi rezultati ukazuju marketinškim istraživačima da postoje značajni segmenti posetilaca restorana (segmenti 4, 10 i 11) koji visoko vrednuju atraktivnost prostora i odličan ukus hrane i koji se izjašnjavaju u značajnoj većini da će ponovo posetiti restoran. Segment 12 je karakterističan po tome što ukazuje na grupu posetilaca restorana kojima nije važan atraktivan prostor, dolaze u restoran sa

decom i nisu skloni da jedu umereno i zdravu hranu, a izjašnjavaju se da će ponovo posetiti restoran.

Izrazito je nepovoljan procenat posetilaca koji su se izjasnili da neće ponovo posetiti restoran kod segmenata pod rednim brojevima 5, 7 i 9. Koje su odlike ovih grupa posetilaca? Prvu grupu čine posetioci koji ne cene atraktivnost prostora i uglavnom dolaze sami bez dece; drugu grupu čine posetioci kojima je nešto važnija atraktivnost prostora u restoranu, ali pripadaju najnižoj dohodovnoj grupi posetilaca; i treću grupu (segment) čine posetioci kojima je važan atraktivan prostor restorana, ali ne vrednuju visoko ukus hrane u restoranu.

U svakom slučaju sagledavanje karakteristika završnih segmenata od izuzetnog je značaja za marketinške menadžere, jer im ovi segmenti pružaju informacije neophodne za formiranje adekvatnih marketinških kampanja usmerenih na posebne segmente potrošača.

TABELA 4. Tabela dobiti po segmentima

Čvor	Čvor		Porast		Odgovor N	Indeks Percent
	N	Procenat	N	Procenat		
10	65	16,3%	65	40,1%	100,0%	246,9%
4	24	6,0%	24	14,8%	100,0%	246,9%
12	12	3,0%	12	7,4%	100,0%	246,9%
11	32	8,0%	26	16,0%	81,3%	200,6%
8	32	8,0%	18	11,1%	56,3%	138,9%
13	33	8,3%	15	9,3%	45,5%	112,2%
5	72	18,0%	2	1,2%	2,8%	6,9%
7	115	28,8%	0	,0%	,0%	,0%
9	15	3,8%	0	,0%	,0%	,0%

4.4. Ocena validnosti procesa segmentacije

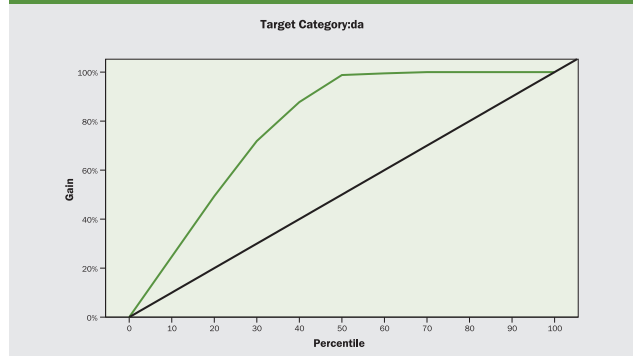
Pored CHAID drveća, koje vizuelno veoma jasno prikazuje rezultate segmentacije, nešto detaljnije informacije o konačnim rezultatima segmentacije dobijaju se u tabeli dobiti po segmentima (tabela 4), kao i na pratećim dijagramima (slika 2 i 3).

U tabeli dobiti za svaki završni segment data su tri indikatora: dobit (kolona *porast - procenat*), indikator odgovora ciljne kategorije (kolona *odgovor*) i kolona *Indeks*. Kolona dobiti pokazuje koliki se procenat posetilaca restorana koji su se izjasnili da će ponovo doći u restoran nalazi u svakom završnom segmentu. Kolona *Odgovor* pokazuje unutrašnju strukturu završnih segmenata, tj. koliko procenata posetilaca u svakom segmentu se izjasnilo da će ponovo posetiti restoran. Najzad, kolona *Indeks* na sumaran način prikazuje performanse segmentacije: indeks vrednost je indikator koji u formi količnika pokazuje koliki se procenat ciljne kategorije za konkretan segment razlikuje od

„očekivanog” procenta ciljne kategorije za ceo uzorak, tj. od prosečne vrednosti za ceo uzorak. Pod „očekivanim” procentom ciljne kategorije podrazumeva se procentualno učešće ciljne kategorije u korenom čvoru CHAID drveća. Indeks vrednosti koji prevazilazi 100% ukazuje da se u datom segmentu nalazi veći procenat opservacija u ciljnoj kategoriji u poređenju sa opštim procentom opservacija u ciljnoj kategoriji za ceo uzorak. Na taj način indeks vrednosti ukazuju na „lift” efekat za svaki čvor – segment. Kako je u našem primeru „očekivani” procenat za ceo uzorak 40,5%, za sve segmente čiji je „odgovor” (kolona *Odgovor*) veći od ove vrednosti, indeks vrednost je veća od 100%. (Na primer, za segment pod rednim brojem 10 Indeks vrednost je jednaka $100\%/40,5\% = 246,9\%$, itd.).

Dijagrami na prethodnim slikama daju veoma dobru vizuelnu prezentaciju prethodno objašnjenih indikatora. Oni nam ukazuju da naš model ima veoma dobre performanse: strmo kretanje linije dobiti koja je jasno udaljena od prosečne prave linije otkriva da CHAID model donosi dobre rezultate segmentacije. Potvrda za ovu činjenicu je data i na pratećim dijagramima kretanja vrednosti odgovora i indeks vrednosti (slika 3).

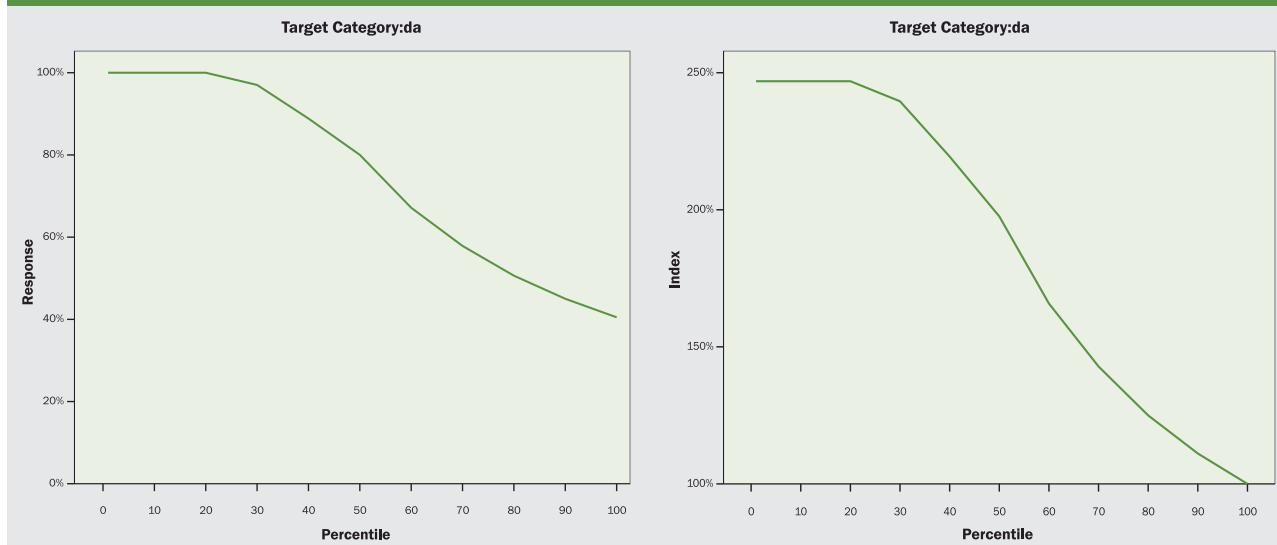
SLIKA 2. Dijagram dobiti po završnim segmentima



4.5. Prediktorska moć modela

Pored već analiziranih rezultata primene CHAID modela na polju segmentacije, posebnu pažnju zaslužuje uočavanje prediktorske moći dobijenog modela. Podatke o ovim karakteristikama našeg CHAID modela nalazimo u tabelama 5 i 6. U tabeli 5 je prikazan rizik predviđanja, koji predstavlja procenat pogrešne klasi-

SLIKA 3. Dijagrami odgovora i indeks vrednosti po završnim segmentima



fikacije opservacija. Naime, za naš primer segmentacije to znači da u slučaju kada su poznate karakteristike nekog posetioca restorana u smislu navedenih sedam prediktivnih promenljivih, rizik njegove pogrešne klasifikacije po pitanju ponovnog dolaska u restoran je 9,3% (na bazi celog uzorka), odnosno 13,3% (primenom test uzoraka u unakrsnoj oceni modela). Saglasno sa prethodnim, tabela klasifikacije otkriva da u 90,8% slučajeva CHAID model vrši tačno predviđanje namere posetioca da ponovo poseti restoran.

TABELA 5. Rizik

Metod	Procenjena vrednost	Standardna devijacija
Resupstitucija	,093	,014
Unakrsna ocena	,133	,017

TABELA 6. Klasifikacija

Posmatrano	Predviđeno		
	ne	da	Procenat
Ne	218	20	91,6%
Da	17	145	89,5%
Ukupni procenat	58,8%	41,3%	90,8%

5. ZAKLJUČAK

Segmentacija tržišta predstavlja jedan od najvažnijih koncepata savremenog marketinga. Ona se može re-

alizovati na bazi manjeg ili većeg broja promenljivih koje odražavaju karakteristike potrošača (demografske, ekonomske, socio-psihološke i sl.), ali i drugih promenljivih, kao što su stavovi ispitanika o različitim pitanjima, ili raznovrsnih podataka koji se odnose na konkretne ekonomske ili marketinške transakcije. Opšti cilj segmentacije jeste da se što preciznije dođe do karakteristika koje mogu da otkriju segmente potrošača kao relativno homogene grupe.

Posle opštih napomena u vezi tehnika segmentacije u radu je dat kratak opis primenjenih modela: logističke regresije i CHAID modela. Model logističke regresije primenjen je samo u delu izdvajanja statistički značajnih promenljivih iz početnog skupa od jedanaest nezavisnih promenljivih, da bi zatim bili razmatrani bitni rezultati segmentacije dobijeni primenom CHAID analize na tako izdvojenim promenljivama. Empirijski primer se odnosi na generisanje prediktivnog modela segmentacije tržišta posetilaca jednog restorana brze hrane. Menadžment restorana je sproveo istraživanje među posetiocima restorana u vezi njihovog zadovoljstva posetom restorana i elementima koji imaju uticaja na njihovu odluku da ponovo posete restoran. Rezultati procesa segmentacije pokazuju da se jasno izdvajaju grupe posetilaca koji visoko vrednuju atraktivnost prostora u restoranu, kao i ukus hrane koja se u njemu služi, i koji su se u velikom procentu izjasnili da će ponovo posetiti restoran. Takođe, u velikom broju slučajeva ponovo će doći u restoran i grupa posetilaca koji dolaze u restoran sa decom i ne pridaju poseban značaj umerenoj i zdravoj ishrani. Sa druge strane, jasno su izdvojena tri segmenta posetilaca koji su se izjasnili da neće ponovo posetiti resto-

ran. Njihove odlike su: prvu grupu čine posetioци koji ne cene atraktivnost prostora i uglavnom dolaze sami bez dece; drugu grupu čine posetioци kojima je nešto važnija atraktivnost prostora u restoranu, ali pripadaju najnižoj dohodovnoj grupi posetilaca; i treću grupu (segment) čine posetioци kojima je važan atraktivan prostor restorana, ali ne vrednuju visoko ukus hrane u restoranu.

Pored ocena o validnosti konkretnog CHAID modela u smislu segmentacije tržišta, koje otkrivaju dobre performanse modela, empirijska analiza je pokazala da ovaj model ima značajnu prediktorsku moć. Naime, tabele rizika i klasifikacije pokazuju da model generiše visok procenat (oko 90%) pouzdane klasifikacije posetilaca u navedene dve grupe: oni koji će ponovo posetiti restoran i one koji neće ponovo doći u restoran.

Reference:

1. Baron, S., Phillips, D. (1994) „Attitude Survey Data Reduction Using CHAID: An Example in Shopping Centre Market Research“, *Journal of Marketing Management*, No. 10, str. 75-88.
2. Berry, J.A.M., Linoff, S.G. (2004) *Data Mining Techniques – for Marketing, Sales and Customer Relationship Management*, John Wiley Publishing, Inc. Indianapolis
3. Drozdenko, G.R., Drake, D.P. (2002), *Optimal Database Marketing – Strategy, Development and Data Mining*, Sage Publications, Inc. Thousand Oaks
4. Galguera, L., Luna, D., Mendez, M.P. (2006) „Predictive Segmentation in Action“, *International Journal of Market Research*, Vol. 48. No. 4., str. 459-479.
5. Giudici, P., Figini, S. (2009) *Applied Data Mining*, 2. izdanje, John Wiley & Sons, Inc. Publication,
6. Hughe, M.A. *Database Marketing Institute*, www.dbmarketing.com/articles/Art160.htm
7. Kass, G. (1980), „An Exploratory Technique for Investigating Large Quantities of Categorical Data“, *Applied Statistics*, No. 29 (2), str. 119-127.
8. Kavak, B., Gumusluoglu, L. (2007) Segmenting Food Markets, *International Journal of Market Research*, Vol. 49, No. 1, str. 71-94.
9. Levin, N., Zahavi, J. (2001) „Predictive Modeling Using Segmentation“, *Journal of Interactive Marketing*, Vol. 15. No. 2, spring, str. 2-22.
10. Myatt, G.J., Johnson, W.P. (2009): *Making Sense of Data II*, John Wiley & Sons, INC. Publication
11. Ratner, B. (2003), *Statistical Modeling and Analysis for Database Marketing – Effective Techniques for Mining Big Data*, Chapman & Hall/CRC
12. www.mhhe.com.hair06.
13. SPSS 15.0. User guide, 2007.

Summary:

Predictive Market Segmentation Model: An Application of Logistic Regression Model and CHAID Procedure

Jasna Soldić-Aleksić

Market segmentation presents one of the key concepts of the modern marketing. The main goal of market segmentation is focused on creating groups (segments) of customers that have similar characteristics, needs, wishes and/or similar behavior regarding the purchase of concrete product/service. Companies can create specific marketing plan for each of these segments and therefore gain short or long term competitive advantage on the market. Depending on the concrete marketing goal, different segmentation schemes and techniques may be applied. This paper presents a predictive market segmentation model based on the application of logistic regression model and

CHAID analysis. The logistic regression model was used for the purpose of variables selection (from the initial pool of eleven variables) which are statistically significant for explaining the dependent variable. Selected variables were afterwards included in the CHAID procedure that generated the predictive market segmentation model. The model results are presented on the concrete empirical example in the following form: summary model results, CHAID tree, Gain chart, Index chart, risk and classification tables.

Key words: market segmentation, predictive segmentation models, CHAID models, logistic regression

Kontakt:

Prof. dr Jasna Soldić-Aleksić

Ekonomski fakultet, Beograd

e-mail: jasnasol@one.ekof.bg.ac.rs