

ISKANJE CILJNIH SEGMENTOV ZA OZAVEŠČANJE O DEPRESIJI S POMOČJO ODLOČITVENIH DREVES**

Povzetek. Namen prispevka je preveriti, ali lahko s pomočjo metode odločitvenih dreves na podlagi demografskih in nekaterih drugih določljivih značilnosti posameznika ločujemo med tistimi, ki so bolj, in tistimi, ki so manj nagnjeni k depresiji oziroma neugodnemu duševnemu počutju. Na ta način smo želeli poiskati ciljne segmente za ozaveščanje o depresiji in preventivnih programih. Študija je temeljila na analizi podatkov ESS za Slovenijo (leto 2012). Na podlagi analize smo identificirali štiri ciljne segmente, ki so v večji meri nagnjeni k depresiji: mladi brezposelni, brezposelni pasivni, starejši in ženske nad 52. letom starosti. Metoda odločitvenih dreves se je izkazala kot uporabna za iskanje ciljnih segmentov tudi na področju duševnega zdravja, rezultati pa potrjujejo izsledke drugih družboslovnih raziskav.

Ključni pojmi: neparametrične metode, klasifikacija, preventivne dejavnosti, depresija

Uvod

V prispevku se ne ukvarjamo primarno z vprašanjem, kateri dejavniki vplivajo na pojav depresije, pač pa ugotavljamo, ali lahko z demografskimi in nekaterimi drugimi lahko določljivimi značilnostmi posameznikov učinkovito ločujemo med osebami z večjim nagnjenem k depresiji oziroma k slabšemu duševnemu počutju in ostalimi. Ugotovitve so posebej uporabne pri načrtovanju oglaševanja preventivnih programov za ozaveščanje in preprečevanje depresije, saj poskušamo z metodo odločitvenih dreves oblikovati več ciljnih segmentov, ki bi jim bilo potrebno nameniti posebno pozornost.

* Marjan Cugmas, diplomirani družboslovni informatik (VS), Fakulteta za družbene vede; dr. Aleš Žiberna, izredni profesor, Fakulteta za družbene vede, Univerza v Ljubljani.

** Izvirni znanstveni članek.

Depresija

Po podatkih Inštituta za varovanje zdravja (Nacionalni inštitut za javno zdravje, 2014) naj bi 10 odstotkov prebivalcev Slovenije, starih od 55 do 64 let, poročalo o depresiji, ki jo je ugotovil zdravnik. Bolezen je poleg tesnobe najpogosteje zabeležena duševna motnja v ambulantah osnovnega varstva (Juričič Konec, 2014).

O depresiji govorimo, ko je človekovo razpoloženje več kot dva tedna zelo negativno (žalost, potrto, obup) ali če je v tem času opazen upad življenjske energije in interesa ter zadovoljstva ob vsakdanjih aktivnostih (Derogatis in Melisaratos, 1983; Juričič Konec, 2014). Osebe s simptomi depresije izbirajo vedenje, ki v večini primerov vodi v umik, samoto in/ali brezdelje, pogosto čutijo bolečine v mišicah, v križu, glavobol, utrujenost tudi po počitku, mravljinčenje po telesu, soočajo pa se tudi s pomanjkanjem energije, pretiranim tekom ali izgubo teka, nespečnostjo, gastrointestinalnimi težavami, povečano ali zmanjšano psihomotorično aktivnostjo in drugimi (Ohayon, 2004). Ni nujno, da se pojavijo vsi omenjeni simptomi, in vsi omenjeni simptomi niso enako izraziti (Juričič Konec, 2014), hkrati pa pojav nekaterih simptomov ne pomeni nujno, da oseba trpi za depresijo.

Različni strokovnjaki navajajo širok spekter dejavnikov, ki lahko pri določeni osebi vodijo v pojav depresije. Delijo jih v vsaj štiri skupine, in sicer na biološke, psihološke, socialne in druge dejavnike (Grum, 2010). Tako imajo kajenje, uživanje alkohola, viri socialne opore, brezposelnost (predvsem dolgotrajna), invalidnost in kronične bolezni, pomanjkanje fizične aktivnosti ter drugi dejavniki pomemben vpliv pri pojavu depresije (Turk, 2005; Planinšek et al., 2014; Vogrin, 2002). Depresija je v zahodnem svetu skoraj dvakrat pogostejša pri ženskah kakor pri moških (Pivec et al., 2014; Malnar in Hafner-Fink, 2013; Van de Velde et al., 2010). Nekateri dodajajo še dejavnike, ki "vzdržujejo" stanje depresije, na primer osamljenost (Juričič Konec, 2014). Pri duševnih boleznih, kot je depresija (in tudi izgorelost), je prepoznavanje simptomov zelo pomembno, saj ima zgodnje odkritje in iskanje strokovne pomoči pomemben vpliv na potek in uspešnost zdravljenja. Ljudje, ki ne poiščejo strokovne pomoči, se pogosto zatečejo k samomoru (Bajt et al., 2009). Po podatkih iz dokumenta Evropske komisije naj bi bilo 90 odstotkov samomorov storjenih zaradi depresije ali drugih duševnih motenj (Wahlbeck in Mäkinen, 2008).

Na nacionalni ravni potekajo številne preventivne akcije z namenom preprečevanja in spodbujanja zdravljenja depresije in samomorilnosti, kot je ozaveščanje o simptomih depresije in zdravem življenjskem slogu, delujejo pa tudi svetovalnice za prvo psihološko pomoč.

Ker je področje prepoznavanja in zdravljenja depresije dobro raziskano, obstajajo testi, ki posamezniku izračunajo nagnjenost k depresiji oziroma morebitno prisotnost depresije.

Podatki

Podatki, ki jih uporabljamo za izgradnjo odločitvenega drevesa, so bili zbrani v sklopu Evropske družboslovne raziskave, izvedene leta 2012. Anketna vprašanja se osredotočajo predvsem na merjenje stališč, prepričan in vedenjskih vzorcev. Populacijo predstavljajo vsi prebivalci Republike Slovenije, ki so bili na dan začetka raziskave stari najmanj 15 let. V vzorec je bilo zajetih 2250 oseb, realizirani vzorec je bil 1257, vzorčni okvir pa je predstavljal Register prebivalcev Slovenije z izvzetimi institucionaliziranimi prebivalci. Izvedeno je bilo dvostopenjsko stratificirano slučajno vzorčenje (European Social Survey, 2014a). V raziskavi so vključena tako vprašanja, na osnovi katerih lahko ocenimo pogostnost pojavljanja simptomov depresije, kot vprašanja, ki jih lahko uporabimo za prepoznavo ciljnih segmentov.

Opis spremenljivk

Skladno s ciljem prispevka analiziramo spremenljivke demografskega tipa in nekatere druge lahko določljive značilnosti posameznika. Te so: spol, starost, najvišja dokončana stopnja izobrazbe, delovna aktivnost in vernost, število oseb v gospodinjstvu, fizična aktivnost in družabnost oziroma pogostnost druženja s prijatelji.

Odvisna spremenljivka je "*indeks depresivnosti*", oblikovan kot vsota osmih trditev in v nadaljevanju pretvorjen v spremenljivko z dvema možnima vrednostma. Cilj je namreč razlikovati med osebami z močnejše izraženimi simptomi depresije in ostalimi. Tudi v okviru akcij, ki lahko izhajajo iz rezultatov analize, se moramo običajno odločati med dvema možnostma. Pri načrtovanju oglaševanja preventivnih dejavnosti se moramo na primer odločiti, ali osebo z določenimi lastnostmi povabiti k preventivnim dejavnostim ali ne.

Odvisna spremenljivka: indeks depresivnosti

Evropska družboslovna raziskava vsebuje skrajšano lestvico za merjenje simptomov depresije "Center for Epidemiological Studies Depression Scale" (Radloff, 1977), z osmimi indikatorji. Besedilo vsake se prične s "*Prosimo, povejte, koliko časa prejšnji teden /.../*" in nadaljuje z opisom določenega občutka (tabela 1). Anketiranec izbere enega izmed štirih ponujenih odgovorov, ki opisujejo pogostnost doživljanja navedenega občutka (0 - "nikoli ali skoraj nikoli", 1 - "le malo časa", 2 - "večino časa", 3 - "ves čas ali skoraj ves čas") (Radloff, 1977; Van de Velde et al., 2010).

Lestvica ni namenjena kliničnemu diagnosticiranju oziroma interpretiranju rezultatov na ravni posameznika, prav tako rezultatov ne moremo interpretirati v smislu prisotnosti bolezni, pač pa le v smislu povečanega tveganja

ali izrazitejših simptomov (Radloff, 1977). Veljavnost in zanesljivost merskega instrumenta sta ustrezni (Bracke et al., 2008; Karim et al., 2014). Cronbachov alfa, s katerim preverjamo zanesljivost testa, na naših podatkih znaša 0,85.

V vzorcu je med 1248 in 1256 veljavnih odgovorov različnih indikatorjev (tabela 1). Indikatorja *občutek depresije* in *občutek osamljenosti* imata najnižji povprečji (0,3), najvišji (2,1) pa imata indikatorja *občutek sreče* in *občutek uživanja življenja*. Poudariti velja, da večja vrednost omenjenih indikatorjev nakazuje odsotnost depresivnih občutij, medtem ko za ostale indikatorje velja obratno, zato smo jima obrnili merski lestvici po obrazcu $4 - (x - 1)$, kjer x pomeni vrednost indikatorja.

Tabela 1: OSNOVNE OPISNE STATISTIKE SPREMENLJIVKE "INDEKS DEPRESIVNOSTI" IN NJENIH INDIKATORJEV

	n	Povprečje	Standardni odklon	Mediana	Koeficient asimetrije	Koeficient sploščenosti
Indeks depresivnosti	1231	4,7	3,8	4,0	1,4	2,4
Občutek depresije	1254	0,3	0,6	1,0	1,7	2,6
Občutek opravljanja stvari s težavo	1253	0,5	0,7	1,0	1,4	1,7
Nemiren spanec	1256	0,7	0,8	1,0	1,1	0,4
Občutek sreče	1249	2,1	0,7	3,0	-0,4	0,2
Občutek osamljenosti	1255	0,3	0,6	1,0	2,3	5,4
Občutek uživanja življenja	1248	2,1	0,7	3,0	-0,4	0,1
Občutek žalosti	1253	0,5	0,7	1,0	1,3	1,8
Občutek nezmožnosti spraviti se v pogon	1250	0,6	0,7	1,0	1,2	1,1

Vir podatkov: Euro Evropska družboslovna raziskava (European Social Survey, 2014b).

Za izračun "*indeksa depresivnosti*" je ustreznih 1231 enot. Najmanjša možna vrednost spremenljivke je 0, največja pa 24. Najmanjša vrednost spremenljivke na analiziranih podatkih je 0, največja pa 23. Povprečje indeksa, ki se pričakovano ne porazdeljuje normalno (Radloff, 1977), je 4,7. Spremenljivko pretvorimo v spremenljivko nominalne merske lestvice, z vrednostma "šibki ali neizraženi simptomi depresije" in "močno izraženi simptomi depresije". Osebe, ki imajo vrednost "*indeks depresivnosti*" večji od ali enak 12 (simptome v povprečju doživljajo večino časa ali ves čas ali skoraj ves čas), razvrstimo v skupino z močno izraženimi simptomi depresije, v katero spada 7 odstotkov oziroma 84 enot.

Kriterij ≥ 12 je določen precej arbitrarno. Za originalno lestvico z 20 indikatorji se je pri uporabi lestvice kot začetnega presejalnega testa najprej priporočal kriterij ≥ 16 (Radloff, 1977), ki ostaja tradicionalni kriterij in je bil tudi večkrat preverjen (npr. Beekman et al., 1997). Ker je bila v Evropski družboslovni raziskavi uporabljena skrajšana lestvica z 8 indikatorji, je potrebno ta kriterij prilagoditi. Enostavna prilagoditev bi bila, da bi vzeli sorazmerni del ($16/20 * 8 = 6,4$), ker bi ustrezalo mejni vrednosti 6 ali 7. Vendar pa so Andresen et al. (1994) že pri lestvici z 10 indikatorji ugotovili, da tak pristop ni ustrezen in da pri lestvici z 10 indikatorji mejni vrednosti 16 najbolj ustreza mejna vrednost 10. Po analogiji s tem bi bil ustrezni kriterij za lestvico z 8 indikatorji ≥ 8 . Vendar pa novejša raziskave ugotavljajo, da mora biti kriterij prilagojen tudi populaciji in namenu uporabe ter da so pogosto primerne precej višje mejne vrednosti, lahko tudi za 60 (kadar je večji poudarek na občutljivosti) do 112 odstotkov (kadar je večji poudarek na specifičnosti) višji od te priporočene mejne vrednosti (Farukawa et al., 1997). Tudi mi smo za 50 odstotkov povečali prej izračunano mejno vrednost (8) in uporabili kriterij ≥ 12 . Ker pa se zavedamo precejšnje nedorečenosti na tem področju, smo preverili tudi druge možne kriterije (kriterij ≥ 8 in kriterij, da je anketiranec na vsaj 4 vprašanja odgovoril z "večino časa" ali "ves čas ali skoraj ves čas"). V večini primerov ob primernem uteževanju napak (glej poglavje o odločitvenih drevesih) dobimo vsebinsko zelo podobno drevo, le da so nekatere kategorije še bolj razčlenjene in tako manj zanimive za naš namen, kjer iščemo lahko določljive segmente. Poleg tega s to mejno vrednostjo dobimo tudi dobre rezultate pri merah kakovosti modela.

Neodvisne spremenljivke

Neodvisne spremenljivke so različnih merskih lestvic. Začnimo s spremenljivkami nominalne in ordinalne merske lestvice. Mednje sodijo: *spol*, *izobrazba* ("Katera je zadnja šola, ki ste jo končali, redno ali izredno?"), *fizična aktivnost* ("S pomočjo kartice prosimo povejte, kolikokrat v zadnjih 7 dneh ste bili neprekinjeno fizično aktivni vsaj 20 minut ali več."), *družabnost* ("Ocenite, kako pogosto se iz družabnih razlogov dobivate s prijatelji, sorodniki ali kolegi z dela."), *vernost* ("Ali se imate za pripadnika kakšne religije ali veroizpovedi?") in *delovna aktivnost* ("In kaj od naštetega najbolje opisuje vašo delovno aktivnost v zadnjem tednu dni? Prosimo, navedite le en odgovor.") (tabela 2).

V analizo vključujemo tudi spremenljivki intervalne merske lestvice: *število oseb v gospodinjstvu* ("Koliko oseb živi v vašem gospodinjstvu, vključno z otroki in vami?") ter *starost*, ki je preračunana iz leta rojstva (tabela 3).

Tabela 2: OPISNE STATISTIKE SPREMENLJIVK NOMINALNE IN ORDINALNE MERSKE LESTVICE

Spremenljivka	Faktor	Frekvenca	Delež (%)
Spol	moški	577	45,9
	ženski	680	54,1
	skupaj	1257	100
Izobrazba	poklicna izobrazba ali manj	408	32,5
	srednje- ali višješolska izobrazba	638	50,9
	visokošolska izobrazba ali več	208	16,6
	skupaj	1254	100,0
Fizična aktivnost	do dva dni	209	16,7
	od tri do pet dni	254	20,3
	več kot pet dni	790	63,0
	skupaj	1253	100,0
Družabnost	enkrat na mesec ali manj	297	23,7
	nekajkrat na mesec	510	40,7
	večkrat na teden	447	35,6
	skupaj	1254	100,0
Vernost	da	694	55,6
	ne	555	44,4
	skupaj	1249	100,0
Delovna aktivnost	bolan ali upokojen	394	32,0
	ne opravlja plačanega dela	94	7,6
	opravlja plačano delo ali gospodinji	591	48,0
	se izobražuje	153	12,4
	skupaj	1232	100,0

Vir podatkov: Evropska družboslovna raziskava (European Social Survey, 2014b).

Tabela 3: OPISNE STATISTIKE SPREMENLJIVK ŠTEVILO OSEB V GOSPODINJSTVU IN STAROST

	n	Povprečje	Standardni odklon	Mediana	Minimum	Maksimum	Koeficient asimetrije	Koeficient sploščenosti
Število oseb v gospodinjstvu	1255	3	1,5	3	1	12	0,7	1,1
Starost	1257	48	18,8	48	15	102	0,1	-0,9

Vir podatkov: Evropska družboslovna raziskava (European Social Survey, 2014b).

Odločitvena drevesa

Z metodo odločitvenih dreves iščemo "pravila", s katerimi lahko kar najbolje ločimo tiste s šibko izraženimi simptomi depresije od tistih z močno izraženimi simptomi depresije. Ugotavljamo torej, na podlagi katerih spremenljivk lahko najbolje napovemo stopnjo izražanja simptomov in kolikšne vrednosti teh spremenljivk kažejo na posamezno skupino (močno/šibko izraženi simptomi).

Odločitvena drevesa kot klasifikacijski model prištevamo v skupino metod nadzorovanega avtomatskega učenja na podlagi rešenih primerov (Štraus, 2012). Nadalje se delijo na regresijska in klasifikacijska drevesa, razlika med njimi pa je v tem, da je pri regresijskih drevesih odvisna spremenljivka številska, pri klasifikacijskih pa kategorialna.

Uporaba klasifikacijskih modelov je zelo raznolika. Bistvo njihove uporabe je segmentacija in oblikovanje čim boljše napovedi (Bramer, 2013) oziroma razlaga odvisne spremenljivke. Z odločitvenimi drevesi lahko ugotavljamo tudi interakcije (Camp in Slattery, 2002) in v procesu izgradnje zmanjšamo število neodvisnih spremenljivk. V družboslovju se najpogosteje uporabljajo za segmentacijo trga in v okviru tega za načrtovanje usmerjenega oglaševanja. Seveda pa so možnosti uporabe metode bistveno širše in bolj raznolike (Gorunescu, 2011).

Preden si pogledamo proces izvedbe metode, naštejmo še nekaj bistvenih prednosti: metoda je zelo enostavna v smislu razumevanja njenega delovanja in interpretacije rezultatov, poleg tega pa odlično deluje tudi na velikih podatkih s kompleksno strukturo, kot so interakcije in nelinearne povezave. Metoda ne predpostavlja določenih (na primer intervalnih) merskih lestvic in je torej primerna tudi za kombinacijo spremenljivk različnih merskih lestvic, hkrati pa je omejena z relativno nezahtevnimi predpostavkami (sodi med neparometrične metode) (Williams, 2011). Nekatere različice metode dovoljujejo tudi manjkajoče vrednosti pri neodvisnih spremenljivkah. Med slabosti odločitvenih dreves pa štejejo predvsem pristranskost (spremenljivke z veliko kategorijami lahko dobijo prevelik pomen) in prekomerno prilagajanje podatkom, ki nastopi, ko imamo v vzorcu malo enot. Prvi potencialni problem lahko rešimo z ustreznimi transformacijami neodvisnih spremenljivk, drugega pa vsaj delno ublažimo s tako imenovanim obrezovanjem (oziroma klestenjem) drevesa. Ideja slednjega je, poleg odpravljanja posledic prekomernega prilagajanja učnim objektom, omejiti velikost odločitvenega drevesa in tako doseči kar se da enostavno rešitev (Štraus, 2012).

Obrezovanje lahko opravimo *predhodno* (nastavimo prag v odstotkih ali številu enot, ki nam pove, kolikšnega deleža enot, ki odstopajo od večine, ne bomo upoštevali, se pravi ustvarili z njimi novega lista - algoritem za

gradnjo drevesa se predhodno ustavi) ali naknadno oziroma *od spodaj*. V tem primeru po kočani gradnji drevesa režemo njegove veje in jih nadomeščamo z vozlišči (Gorunescu, 2011). Obstaja več metod obrezovanja, vendar raziskovalci do sedaj še niso izpostavili nobene, ki bi v svoji učinkovitosti posebej izstopala (Rokach, 2007).

Obstaja več metod gradnje odločitvenih dreves.¹ Začnemo s praznim drevesom (korenino, angl. *root*) in celotno množico podatkov. S pomočjo hevrstične mere (ali metrik čistosti, npr. informacijski prirastek oziroma informacijski prispevek, Ginijev koeficient, hi-kvadrat in druge) izberemo spremenljivko (in mejno vrednost), ki zopet najbolj izboljša mero, ki jo optimiziramo, nato izberemo naslednjo spremenljivko, ki najbolj izboljša to isto mero, in nadaljujemo, dokler ne zmanjka spremenljivk (in mejnih vrednosti) oziroma dokler je vrednost izboljšanja metrike čistosti še nad predpostavljeno mejo.

Po gradnji nadaljujemo z obrezovanjem in preverjanjem modela (dobljenega drevesa). Najbolj znana metoda preverjanja odločitvenih dreves poteka tako, da pred izgradnjo slučajno izberemo polovico (ali nek drug delež) enot, ki jih ne uporabimo kot učno množico, pač pa na njih samo preverimo kakovost razvrstitve (Bramer, 2013). Obstajajo tudi mnoge druge metode. Ena takih je postopek, pri katerem ustvarimo drevo na vseh podatkih, nato pa preverimo, kolikšen delež enot je bil razvrščen pravilno. Ker je ocena kakovosti modela izračunana na podatkih, na katerih je bil zgrajen model, so ocene kakovosti praviloma preveč optimistične (metoda model oceni bolje, kot je v resnici) (Schunn, 2005).

Med bolj znane metode preverjanja modela sodi tudi metoda navzkrižnega preverjanja. Algoritem deluje tako, da podatke razdeli na določeno število (k) delov (običajno 5 ali 10), nato na $k - 1$ delih zgradi model (oziroma drevo), ki ga zatem testira na preostalem delu podatkov. Postopek ponovi k -krat, tako da so modeli zgrajeni in testirani na vseh podatkih (Bramer, 2013).

Analiza

Pri izgradnji klasifikacijskega drevesa upoštevamo vse predstavljene spremenljivke in vse enote, kot metriko čistosti pa uporabimo metodo informacijski prirastek. Drevo gradimo, dokler se ta metrika izboljšuje ob upoštevanju omejitev predhodnega obrezovanja. Parametre nastavimo tako, da se rast drevesa ustavi, če v naslednjem listu ni vsaj 30 enot². Glede

¹ CART, CHAID, Exhaustive, QUEST idr. (Rokach, 2007; Gorunescu, 2011).

² Meja 30 enot v celicah kontingenčnih tabel je del mnogih konkretnih priporočil k splošnim etičnim načelom statističnega poročanja (Vehovar in Kalton, 2001).

na to, da po izgradnji odločitvenega drevesa izvedemo še naknadno obrezovanje, odločitev o vrednosti 30 enot niti ni pomembna, saj bi po naknadnem obrezovanju dobili enake rezultate, kot če meje ne bi nastavili. Po pregledu nastalega drevesa ga naknadno obrežemo tako, da dobimo čim manjšo napako (z upoštevanjem uteževanja, več o tem v nadaljevanju) po navzkrižnem preverjanju.

Pri metodah klasifikacije lahko storimo dve vrsti napak: da osebo s šibko izraženimi simptomi depresije uvrstimo v skupino z močno izraženimi simptomi depresije in da osebo z močno izraženimi simptomi depresije uvrstimo v skupino s šibko izraženimi simptomi depresije. Pogosto je mogoče opredeliti resnost posledic ene vrste napake v razmerju z drugo (Hastie et al., 2009): glede na cilj raziskave je posledica prve napake le višji strošek oglaševanja, medtem ko je posledica druge napake zmanjšan doseg oglaševanja preventivne akcije in posledično nedosežen ciljni segment. Metoda odločitvenih dreves omogoča definiranje "škoda" pri posamezni napaki.

Predpostavljamo, da je "škoda", ko nekomu z močno izraženimi simptomi depresije napovemo šibko izražene simptome depresije, 15-krat večja kot obratno (torej razmerje 15 : 1). Tudi ta vrednost je določena precej arbitrarno. Pri njenem določanju po eni strani stremimo k ustrezni občutljivosti in specifičnosti, po drugi strani pa k ustrezno velikim dobljenim ciljnim segmentom (ne želimo premajhnih, pa tudi ne takih, ki bi vključevali preveč enot - večino populacije). Tako določimo čim manjšo vrednost (manjša vrednost zagotavlja večjo specifičnost in običajno večje segmente), ki še zagotavlja primerno občutljivost, obenem pa v ciljnih segmentih ni prevlekega deleža populacije. Če bi ciljni segmenti vključevali velik del populacije, ne bi bili primerni za ciljno oglaševanje. Poleg razmerja 15 : 1 preizkusimo tudi druge možnosti. Manjše vrednosti niso primerne, saj ne zagotavljajo primerne občutljivosti (že pri razmerju 13 : 1 občutljivost pade na 0). Pri povečevanju razmerja se rezultati do razmerja 19 : 1 ne spremenijo, kasneje pa občutljivost narašča na račun specifičnosti (in večjega dela populacije v "ciljnih segmentih" ter razdrobljenosti segmentov). Pri razmerju 70 : 1 bi tako občutljivost dosegla 100 odstotkov, specifičnost pa bi padla na 42 odstotkov (ciljni segmenti bi obsegali več kot 60 odstotkov).

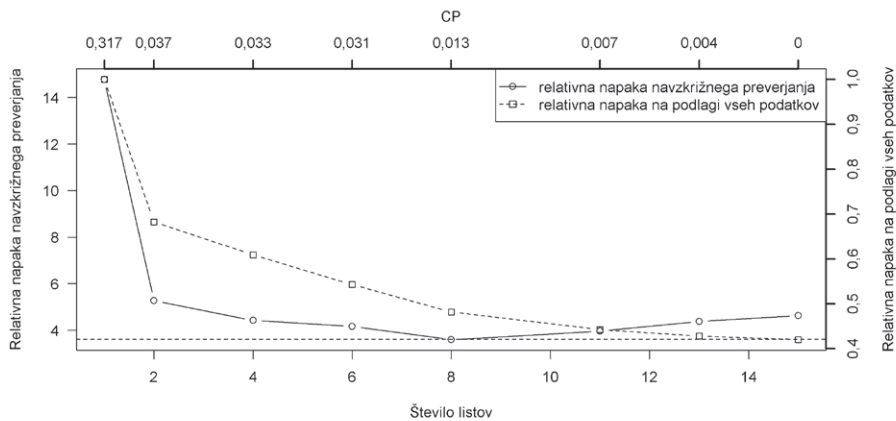
Neobrezano drevo (ni v prispevku) je še vedno precej kompleksno, čeprav že odraža predvidene ciljne segmente. V tem koraku ga podrobneje ne interpretiramo, saj opravimo še t. i. naknadno obrezovanje. Pri slednjem smo pozorni predvsem na relativno napako (angl. *cross validation error*) po navzkrižnem preverjanju. To so vrednosti navzkrižnega preverjanja kakovosti modela. Algoritem ponovimo stokrat, podatke pa vsakič razdelimo na deset delov. Tako dobimo bolj stabilne vrednosti, ki so v našem primeru povprečja napak navzkrižnega preverjanja za več ponovitev. Vrednosti praviloma padajo, saj z večanjem drevesa (dodajanjem vozlišč) model

načeloma izboljšujemo. Lahko pa nastopi tudi točka preloma, ko začnejo napake ponovno naraščati. To je lahko indikator, da se je drevo na neki točki rasti začelo preveč prilegati podatkom, in znak za obrezovanje (Everitt in Hothorn, 2010). Drevo obrežemo pri številu vozlišč, kjer je napaka navzkrižnega preverjanja najmanjša, kar nakazuje najbolj optimalno rešitev. V našem primeru je to pri sedmih vozliščih oziroma pri osmih listih (skupinah, segmentih).

Ko si pri določitvi števila vozlišč oziroma listov pomagamo tudi z grafično predstavitvijo napak (slika 1), najbolj optimalno število listov označuje mesto, kjer se "črta" najbolj lomi oziroma kjer začnejo napake navzkrižnega preverjanja naraščati. Tudi po tem kriteriju je to pri sedmih vozliščih oziroma osmih listih (skupinah, segmentih).

V primeru, ko napake razvrstitve preverjamo na istih podatkih, kot smo jih uporabili pri izgradnji modela (torej ko za oba koraka uporabimo vse podatke), vrednosti relativnih napak ves čas padajo, vendar pa je ocena kakovosti modela v tem primeru običajno pristranska (Schunn, 2005).

Slika 1: GRAFIKON ZA DOLOČANJE ŠTEVILA VOZLIŠČ, KI PRIKAZUJE VPLIV ŠTEVILA VOZLIŠČ NA VREDNOSTI NAPAK NAVZKRIŽNEGA (IN "NAVADNEGA") PREVERJANJA.



Končna rešitev

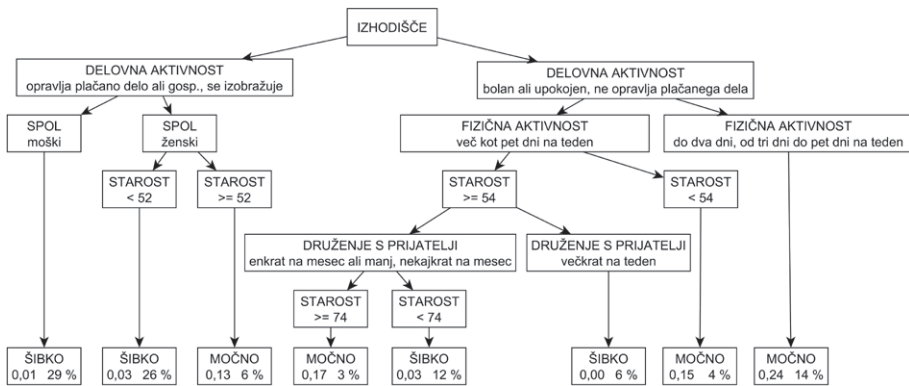
V tako obrezanem modelu ostanejo naslednje neodvisne spremenljivke: delovna aktivnost, fizična aktivnost, družabnost, starost in spol. V listih obrezanega klasifikacijskega drevesa (slika 2) so označene napovedi (šibko ali močno izraženi simptomi depresije), ki so določene na podlagi verjetnosti za pojav močno izraženih simptomov depresije (vrednost levo pod oznako

izida). Verjetnosti predstavljajo deleže enot z močno izraženimi simptomi depresije, ki so razvrščene v določene liste (skupine, segmente) klasifikacijskega drevesa. Vrednosti desno pod oznako napovedi označujejo delež enot (izmed vseh enot), razvrščenih v dani segment.

Na podlagi slike 2 lahko identificiramo štiri ciljne segmente za usmerjeno oglaševanje preventivnega programa za preprečevanje depresije, ki predstavljajo glavni rezultat naše analize. Delovna aktivnost je spremenljivka, ki najbolje ločuje med tistimi, ki pogosteje doživljajo simptome depresije, in ostalimi. Status bolan, ali upokojen, ali ne opravlja nobenega plačanega dela je značilen za tri od štirih segmentov, za katere velja večje tveganje za pojav pogostejšega izražanja simptomov depresije. Za tiste, ki opravljajo plačano delo, ali gospodinjijo, ali se izobražujejo, je verjetnost, da bodo izkazovali simptome depresije, manjša. Tako lahko večini tistih, ki delovno niso aktivni, napovemo močno izražene simptome, razen če imajo kakšne druge pozitivne okoliščine (starost vsaj 54 let skupaj z večjo fizično aktivnostjo, pri mlajših od 74 let pa še skupaj s pogostejšim druženjem s prijatelji). Po drugi strani pa na podlagi dobljenega modela, razen starejšim ženskam, vsem delovno aktivnim ni mogoče napovedati močnejše izraženih simptomov depresije. Vplive posameznih dejavnikov bomo dodatno osvetlili še v poglavju Razpravljanje.

Oglejmo si ciljne segmente za ozaveščanje o depresiji (tiste, ki jim lahko napovemo močnejše izražene simptome depresije) ($n = 1231$):

1. Bolni ali upokojeni in ostali, ki ne opravljajo nobenega plačanega dela ter so vsaj 20 minut fizično aktivni petkrat tedensko ali manj pogosto. Verjetnost za izkazovanje simptomov depresije je 24 odstotkov, segment zajema 14 odstotkov vseh enot.
2. Bolni ali upokojeni in ostali, ki ne opravljajo nobenega plačanega dela ter so za razliko od prvega segmenta fizično aktivni vsaj 20 minut dnevno več kot petkrat na teden in so mlajši od 54 let. Verjetnost za izkazovanje simptomov depresije je 15 odstotkov, segment pa zajema 4 odstotke vseh enot.
3. Bolni ali upokojeni in ostali, ki ne opravljajo nobenega plačanega dela ter so fizično aktivni vsaj 20 minut dnevno več kot petkrat tedensko. Stari so nad 74 let in se s prijatelji iz družabnih razlogov srečujejo nekajkrat na mesec ali redkeje. Verjetnost za izkazovanje simptomov depresije je 17 odstotkov, segment pa zajema 3 odstotke vseh enot.
4. Ženske, ki opravljajo plačano delo, ali gospodinjijo, ali se izobražujejo ter so stare 52 let ali več. Verjetnost za izkazovanje simptomov depresije je 13 odstotkov, segment pa predstavlja 6 odstotkov vseh enot.

Slika 2: OBREZANO KLASIFIKACIJSKO DREVO ($n = 1231$).

Diagnostika kakovosti klasifikacijskega drevesa

Ocenjevanja kakovosti klasifikacijskega drevesa, v smislu deleža pravilno napovedanih izidov, smo se že dotaknili. V nadaljevanju omenjamo še dva pristopa, ki ta delež, z vidika neposredne možnosti interpretacije, ocenjujeta najboljše. Pogledali si bomo pojma občutljivost in specifičnost ter krivuljo ROC (*Receiver Operating Characteristic*).

Občutljivost in specifičnost

Klasifikacijska tabela je kontingenčna tabela (tabela 4), kjer so v vrsticah izmerjene vrednosti, vrednosti v stolpcih pa na podlagi modela izračunane napovedi odvisne spremenljivke. V primeru, da je izračunana verjetnost dogodka (močno tveganje za pojav simptomov depresije) višja od 6,25 odstotka (tako imenovana *cut value*, vrednost izhaja iz uteži škode napak), se enoto razvrsti v stolpec "močno", v nasprotnem primeru pa v stolpec "šibko". Na podlagi modela pravilno razvrstimo 76,7 odstotka enot. V primeru, ko bi vse enote razvrstili v skupino šibko izraženih simptomov depresije, bi pravilno razvrstili 93,2 odstotka enot. Kljub višjemu deležu pa bi bila takšna razvrstitev neprimerna, saj bi napačno razvrstili vse enote z močno izraženimi simptomi depresije. Zaradi tega navadno navedemo še dve pomembni meri, ki ju izpeljemo iz klasifikacijske tabele:

1. *Občutljivost* je verjetnost napovedi dogodka, ko se je ta dejansko zgodil (Hastie et al., 2009). Interpretiramo torej delež oseb, ki jim na podlagi modela pravilno pripišemo močno izražene simptome depresije.
2. *Specifičnost* pa je verjetnost napovedi, da se dogodek ni zgodil, ko se dejansko ni zgodil (Hastie et al., 2009). Interpretiramo torej delež oseb,

ki jim na podlagi modela pravilno pripišemo šibko izražene simptome depresije.

Tabela 4: KLASIFIKACIJSKA TABELA SPREMENLJIVKE "INDEKS DEPRESIVNOSTI"

Izmerjene vrednosti	Napovedane vrednosti		Delež pravilno razvrščenih enot
	šibko	močno	
šibko	879	268	76,6
močno	19	65	77,4

Vir podatkov: Evropska družboslovna raziskava (European Social Survey, 2014b).

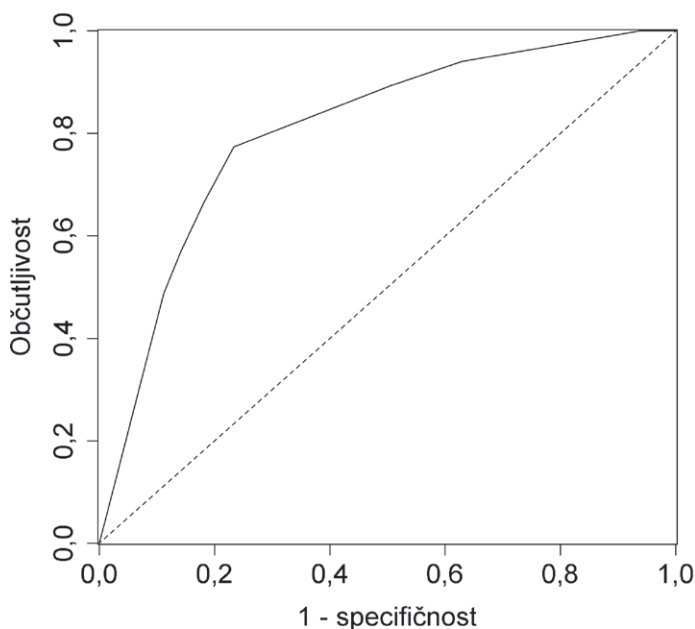
Stremimo k temu, da bi bili obe vrednosti enaki 1, saj bi to pomenilo, da smo z modelom pravilno razvrstili vse enote. Ker je v našem primeru pomembno doseči predvsem osebe z močno izraženimi simptomi depresije, smo pozornejši na občutljivost. Z algoritmom pravilno razvrstimo 77,4 odstotka enot z močno izraženimi simptomi depresije. Specifičnost je 76,6 odstotka. Če bi enote razvrščali popolnoma slučajno (z enako verjetnostjo za obe skupini), bi bila tako občutljivost 50 odstotkov. Če bi vse enote razvrstili v največjo skupino, bi bila specifičnost sicer 1, občutljivost pa 0.

Krivulja ROC

Ker ocene, kot so občutljivost, specifičnost in skupna uspešnost, zaradi odvisnosti od sestave skupin niso dovolj za oceno diskriminantne moči modela, je potrebna statistika, ki je neodvisna od sestave skupin. Ustrezna alternativa je krivulja ROC (Fawcett, 2006), ki jo izračunamo na podlagi podatkov odvisne spremenljivke in izračunanih izidov. Ta kaže tudi, kakšne kombinacije občutljivosti in specifičnosti lahko dosežemo z nastavljanjem različnih presečnih vrednosti.

Na sliki 3 površina desno pod diagonalo (črtkano) predstavlja kakovost naključnega modela (če bi enote razvrščali naključno), medtem ko površina pod krivuljo (polno) predstavlja kakovost izbranega modela. Površina med obema črtama predstavlja izboljšanje razvrstitve z modelom (Kaftandjieva, 2010) in lahko zavzame vrednosti med 0,5 in 1. V konkretnem primeru znaša 0,75, kar interpretiramo kot zadostno (Tape, 2014).

Slika 3: KRIVULJA ROC



Razprava in sklep

Metoda odločitvenih dreves predstavlja pomembno neparametrično alternativo obstoječim metodam. Rezultati analize kažejo, da je primerna za oblikovanje in prepoznavanje ciljnih segmentov na področju depresije oziroma slabega duševnega počutja, saj rezultati potrjujejo in dodatno osvetljujejo rezultate že opravljenih analiz.

Ena izmed pomembnih prednosti metode je enostavna interpretacija dobljenih segmentov in pogojnih verjetnosti pozitivnih izidov (pojava močnejše izražene simptome depresije). V razpravi se osredotočamo predvsem na razlago vplivov posameznih spremenljivk, ki najbolj ločujejo med tistimi, ki pogosteje doživljajo simptome depresije, in ostalimi. Kljub temu pa pomembnosti interpretacije interakcij med spremenljivkami nikakor ne smemo zanemariti. Kot zanimiv primer lahko navedemo spremenljivke delovna aktivnost, fizična aktivnost in starost. Osebe, ki so bolne ali upokojene oziroma ne opravljajo nobenega plačanega dela in so fizično aktivne več kot pet dni na teden, so razvrščene v skupino z višjo verjetnostjo pogostejšega doživljanja simptomov depresije, če so mlajše od 54 let, sicer pa je to odvisno še od drugih dejavnikov (torej starost na nek način zmanjšuje verjetnost za močno izražene simptome depresije). Za ostale (torej stare 54

let ali več), ki se redkeje kot večkrat na teden družijo s prijatelji, pa starost vsaj 74 let zvišuje verjetnost za močnejše izražene simptome depresije.

Dejavnik, ki najbolj diskriminira med tistimi z večjo verjetnostjo pojava depresije in ostalimi, je delovna aktivnost. Predhodne raziskave (Stankunas et al., 2006; Pharr et al., 2011; Lorant et al., 2007) potrjujejo, da lahko brezposelnost privede do težav v psihičnem in fizičnem počutju. Študije brezposelnosti kot stresorja ugotavljajo številne neposredne in posredne povezave med ekonomskimi spremembami in posledično stresnimi življenjskimi dogodki. Epidemiološke raziskave kažejo, da pri nezaposlenih osebah obstaja večja verjetnost pojava simptomov depresije (Catalano, 1991; Dooley et al., 1994). Raziskovalci izpostavljajo, da se simptomi depresije pojavljajo bolj kot posledica brezposelnosti kot pa vzrok, zaradi katerega so ljudje ostali brez dela (Patton in Patricia, 1990). Obstaja več teorij, ki poskušajo razložiti povezavo med depresivnostjo in brezposelnostjo. Tradicionalni model izhaja iz predpostavke, da delovna deprivacija predstavlja deprivacijo psiholoških potreb, zaradi česar prihaja do načenjanja mentalnega zdravja in depresivnosti brezposelnih oseb. Po tej teoriji ima delo latentne psihološke funkcije, saj časovno strukturira dan, omogoča dnevne stike in izmenjave izkušenj, omogoča postavitev ciljev, definira vidike osebnega statusa in identitete, spodbuja aktivnost ter ne nazadnje omogoča finančno varnost in stabilnost. Podaljšana brezposelnost privede do podaljšane deprivacije omenjenih psiholoških potreb in poslabšanja splošnega mentalnega zdravja, kar se manifestira v pasivizaciji, izolaciji, izgubi ciljev in socialnega statusa ter negotovosti na področju osebne identitete (Jahoda, 1982).

V segmentu delovno neaktivnih je pomembna spremenljivka, ki ločuje med tistimi, ki so bolj nagnjeni k slabšemu duševnemu počutju, in ostalimi – fizična aktivnost. Pri ljudeh, ki so fizično aktivni več kot pet dni na teden, obstaja manjša verjetnost za pojav slabega duševnega počutja kakor pri ostalih. Raziskave izpostavljajo razlike v stopnji in razširjenosti slabega duševnega počutja med fizično aktivnimi in fizično neaktivnimi (Planinšek et al., 2014) ter celo poudarjajo pomen fizične aktivnosti pri zdravljenju depresije (Plesničar in Plesničar, 2009). Lawlor in Hopker (2001) opozarjata na pomen izvajanja fizične aktivnosti v skupinah.

Pri bolnih ali upokojenih, bolj fizično aktivnih in starejših od 53 let se kot pomemben dejavnik pojavlja tudi družabnost. Pri tistih, ki se večkrat na teden družijo s prijatelji, obstaja manjše tveganje za močnejše izražene simptome depresije.

Starost, kot spremenljivka z visoko diskriminatorno močjo, v odločitvenem drevesu nastopa zelo pogosto. Pri starejših obstaja višja stopnja tveganja za pojav slabega duševnega počutja, raziskave, osredotočene na starejše, pa kažejo, da je depresija najbolj prisotna pri najstarejših starih (Touzery, 2041). Blazer (2003) ugotavlja, da so klinično pomembni simptomi

depresije prisotni pri približno 15 odstotkih starejše populacije. Prav tako je stopnja depresivnosti višja pri starejših ženskah kot pri starejših moških. Še večja je prevalenca simptomov depresije, ki ne dosegajo kriterijev za postavitev diagnoze, posebej med starejšimi od 75 let, kar avtorji pripisujejo večji prisotnosti telesnih bolezni, pogostejši prisotnosti kognitivnih težav in večjemu deležu žensk med najstarejšim delom populacije (prav tam). Podobno razmerje med spoloma je prisotno tudi v primeru depresivnostne simptomatike, saj so v študiji EURODEP (Braam et al., 2005) ugotovili, da je depresivnostna simptomatika v povprečju prisotna pri 14,1 odstotka starejših žensk in pri 8,6 odstotka starejših moških. Razlogi za višjo stopnjo tveganja depresivnih motenj pri starejši populaciji so predvsem v socialni izolaciji in osamljenosti ter nezmožnosti aktivnega udejstvovanja v lokalnem okolju, kar nakazujejo tudi rezultati naše analize.

Pomembno vlogo pri nastanku depresije v starosti imajo tudi socialno-ekonomski dejavniki. Starostniki se pogosto srečajo z znižanjem dohodka, kar poveča verjetnost za nastanek depresivne motnje, še posebej ker imajo nizki prihodki kronične posledice, kot sta stalni finančni pritiski in izpostavljenost nestabilnemu finančnemu položaju (Mojtabai in Olfson, 2004). Poleg tega lahko slab finančni položaj pomembno vpliva na dostopnost in uspešnost zdravljenja depresije.

Izsledki predstavljene analize potrjujejo vpliv socio-ekonomskega statusa oziroma delovne aktivnosti in starosti na pogostnost pojavljanja simptomov depresije. Rezultati naše raziskave kažejo, da lahko nižja starost v kombinaciji z drugimi dejavniki povečuje verjetnost močnejše izraženih simptomov, kar se kaže v ciljnem segmentu 2 (delovno neaktivni, fizično aktivni, pod 54 let). To pojasnujemo s tem, je v tem starostnem razredu delovna neaktivnost večkrat posledica brezposelnosti (in ne npr. upokojitve), ki pa ima, kot smo navedli že zgoraj, negativne učinke na duševno počutje.

Znotraj skupine delovno aktivnih obstaja večje tveganje za pojav slabega duševnega počutja pri ženskah, starejših od 52 let. Da se depresija pojavlja približno dvakrat pogosteje pri ženskah kakor pri moških, je eden najbolj konsistentnih rezultatov, ki so ga potrdile različne raziskave (Piccinelli in Wilkinson, 2000). Vzorec se pojavlja tako v klinični kot tudi v splošni populaciji ter je neodvisen od lokacije, metode ocenjevanja in diagnostičnih sredstev (Kessler et al., 1993). Teorije, ki poskušajo razložiti razlike med spoloma, se v večji meri opirajo na model občutljivosti na stres, ki opisuje odnos med stresorji in posameznikovim odzivom nanj (Pearlin, 1989). Zdi se, da so ženske bolj dovzetne za omejitve, povezane z njihovo vlogo in zmanjšano možnostjo izbire, kar privede do povečane ravni stresa. Večja prisotnost simptomov stresa pri ženskah je prav tako lahko posledica pritiska različnih družbenih vlog. Rezultati raziskave, ki sta jo v Sloveniji izvedla Malnar in Fink (2013), kažejo, da so simptomi stresa pogostejši pri ženskah

kot pri moških, še posebej ranljive pa so ženske z nižjo stopnjo izobrazbe in ženske, zaposlene na stresnem delovnem mestu, ki so hkrati odgovorne za hišna opravila in vzgojo otrok, kar še dodatno poveča raven stresa. Ženski socialni položaj tako zaznamujeta večja nemoč upravljanja z življenjem in nižji socialni status.

Namen prispevka je bil preveriti, ali lahko z metodo odločitvenih dreves na podlagi demografskih in nekaterih drugih lahko določljivih značilnosti posameznika učinkovito ločujemo med osebami, ki so bolj nagnjene k depresiji oziroma slabšemu duševnemu počutju, in ostalimi. Rezultate smo želeli uporabiti za iskanje ciljnih segmentov za ozaveščanje o depresiji oziroma za oglaševanje preventivnih programov za preprečevanje depresije. Cilj prispevka ni bil pojasniti izvorov pojava depresije.

Uporabili smo metodo klasifikacijskih dreves na podatkih iz Evropske družboslovne raziskave (za Slovenijo) iz leta 2012. Neodvisne spremenljivke bolj kakor na podlagi izsledkov drugih znanstvenih raziskav s področja duševnega zdravja izbiramo na podlagi teoretičnih predpostavk ciljnega oglaševanja ter enostavnosti izbire in določitve spremenljivk.

Identificirali smo štiri ciljne segmente. Prvim trem je skupen status zaposlitve: gre za bolne, upokojene ali brezposelne. V prvem segmentu so tisti, ki bi jih lahko na kratko opisali kot pasivne brezposelne. Poleg tega, da niso delovno aktivni, so tudi sicer manj fizično aktivni kakor ostali. V drugem segmentu so bolj fizično aktivni posamezniki, ki se s prijatelji iz družabnih razlogov srečujejo najredkeje in so starejši od 74 let. Tretji segment pa predstavljajo mlajši posamezniki do 54 let, ki so fizično aktivni več kot petkrat tedensko. Ženske, starejše od 52 let, so edini segment (s povečanim tveganjem za nastanek depresije), ki je delovno aktiven.

Glede na področje in namen uporabe kakovost drevesa označujemo kot primerno. Na podlagi modela smo uspeli pravilno razvrstiti 77 odstotkov oseb, kot take z višjim tveganjem za pojav depresije pa smo pravilno prepoznali 77 odstotkov teh oseb.

V prispevku smo pokazali, da je metoda odločitvenih dreves primerna za iskanje ciljnih segmentov za ozaveščanje o depresiji, in predstavili 4 segmente, ki bi jih bilo smiselno vabiti k preventivnim programom s področja duševnega zdravja in depresije.

LITERATURA

- Andresen, M. Elena, Judith A. Malmgren, William B. Carter in Donald L. Patrick (1994): Screening for Depression in Well Older Adults: Evaluation of a Short Form of the CES-D (Center for Epidemiologic Studies Depression Scale). *American Journal of Preventive Medicine* 10 (2): 77–84.
- Bajt, Maja, Mojca Gabrijelčič Blenkuš, Helena Jeriček Klanšček, Katja Kovše, Tatjana Kofol Bric, Radivoje Pribaković Brinovec, Saška Roškar, Sonja Tomšič, Romana

- Štokelj, Maja Zorko (2009): *Duševno zdravje v Sloveniji*. Ljubljana: Inštitut za varovanje zdravja.
- Blazer, G. Dan (2003): Depression in late life: review and commentary. *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences* 58 (3): M249–65.
- Braam, W. Arjan, Martin J. Prince, Aartjan T. F. Beekman, Philippe Delespaul, Michael E. Dewey, S. W. Geerlings, S.-L. Kivela, Brian A. Lawlor, H. Magnusson, Isac Meller, Karine Pérès, F. M. Reischies, Marc Roelands, Robert A. Schoevers, Pieró Saz, Ingmar Skoog, Cesare Turrina, Ann Versporten in John R. M. Copelan (2005): Physical health and depressive symptoms in older Europeans Results from EURODEP. *The British Journal of Psychiatry* 187 (1): 35–42.
- Beekman, T. F. Aartjan, Dorly J. Deeg, Jacques Van Limbeek, Arjan W. Braam, MZ De Vries in Williem Van Tilburg (1997): Criterion validity of the Center for Epidemiologic Studies Depression scale (CES-D): results from a community-based sample of older subjects in the Netherlands. *Psychological medicine* 27 (1): 231–35.
- Bracke, Piet F., Katia Levecque in Sarah Van de Velde (2008): The psychometric properties of the CES-D 8 depression inventory and the estimation of cross-national differences in the true prevalence of depression. *Proceedings of the International Conference on Survey Methods in Multinational, Multiregional, and Multicultural Contexts (3MC)*. Berlin, Nemčija.
- Bramer, Max (2013): *Principles of data mining*. London: Springer.
- Camp, Nicola J. in Martha L. Slattery (2002): Classification tree analysis: a statistical tool to investigate risk factor interactions with an example for colon cancer (United States). *Cancer Causes & Control* 13 (9): 813–23.
- Catalano, Ralph (1991): The health effects of economic insecurity. *American Journal of Public Health* 81 (9): 1148–52.
- Derogatis, R. Leonard in Nick Melisaratos (1983): The Brief Symptom Inventory: an introductory report. *Psychological Medicine* 13 (3): 595–605.
- Dooley, David, Ralph Catalano in Georjeanna Wilson (1994): Depression and unemployment: panel findings from the Epidemiologic Catchment Area study. *American journal of community psychology* 22 (6): 745–65.
- Everitt, Brian in Torsten Hothorn (2010): *A handbook of statistical analyses using R*. Boca Raton, Fla: CRC Press.
- Fawcett, Tom (2006): An Introduction to ROC Analysis. *Pattern Recognition Letters* 27 (8): 861–74.
- Fiske, Amy, Julie Loebach Wetherell in Margaret Gatz (2009): Depression in older adults. *Annual review of clinical psychology* 5: 363–89.
- Furukawa, Toshiaki, Toshiyuki Hirai, Toshinori Kitamura in Kiyohisa Takahashi (1997): Application of the Center for Epidemiologic Studies Depression Scale among first-visit psychiatric patients: a new approach to improve its performance. *Journal of Affective Disorders* 46 (1): 1–13.
- Gorunescu, Florin (2011): *Data Mining: Concepts, Models and Techniques*. Berlin: Springer.
- Grum Tančič, Alenka (2010): *Dejavniki tveganja za razvoj depresije in samo-*

- morilnega vedenja med starejšimi. V Alenka Grum Tančič (ur.), Vita Poštuvan (ur.) in Saška Roškar (ur.), Spregovorimo o depresiji in samomoru med starejšimi, 33–46. Ljubljana: Inštitut za varovanje zdravja RS.
- Hastie, Trevor, Robert Tibshirani in Jerome Friedman (2009): *The elements of statistical learning*. New York: Springer.
- Jahoda, Marie (1982): *Employment and unemployment: A social-psychological analysis* (Vol. 1). CUP Archive.
- Juričič Konec, Nuša (2014): Depresija. Dostopno preko <http://www.zzv-ce.si/depresija>, 1. 1. 2013.
- Kaftandjieva, Felianka (2010): *Methods for Setting Cut Scores in Criterion-referenced Achievement Tests. A Comparative Analysis of Six Recent Methods with an Application to Tests of Reading in EFL*. EALTA publication.
- Kalton, Graham in Vasja Vehovar (2001): *Vzorčenje v anketah*. Ljubljana: Fakulteta za družbene vede.
- Karim, Jahanvash, Robert Weisz, Zainab Bibi in Shafiq Rehman (2014): Validation of the Eight-Item Center for Epidemiologic Studies Depression Scale (CES-D) Among Older Adults. *Current Psychology*, 1–20.
- Kessler, C. Ronald, Katherine A. McGonagle, Marvin Swartz, Dan G. Blazer in Christopher B. Nelson (1993): Sex and depression in the National Comorbidity Survey I: Lifetime prevalence, chronicity and recurrence. *Journal of affective disorders* 29 (2): 85–96.
- Lawlor, Debbie A. in Stephen W. Hopker (2001): The effectiveness of exercise as an intervention in the management of depression: systematic review and meta-regression analysis of randomised controlled trials. *Bmj* 322 (7289): 763.
- Lorant, Vincent, Christophe Croux, Scott Weich, Denise Deliège, Johan Mackenbach in Marc Ansseau (2007): Depression and socio-economic risk factors: 7-year longitudinal population study. *The British journal of psychiatry* 190 (4): 293–98.
- Malnar, Brina in Mitja Hafner-Fink (2013): Thirty years of gender differences in self-assessed health: the case of Slovenia / Trideset let razlik v samoocenah zdravja med spoloma: primer Slovenije. *Zdravstveno varstvo* 52 (2): 99–107.
- Mojtabai, Ramin in Mark Olfson (2004): Major depression in community-dwelling middle-aged and older adults: prevalence and 2-and 4-year follow-up symptoms. *Psychological medicine* 34 (4): 623–34.
- Nacionalni inštitut za javno zdravje (2014): Duševno zdravje in starejši. Dostopno preko www.ivz.si/mediji/sporocila_za_javnost?pi=21&_21_view=item&_21_newsId=2364&pl=85-21.0, 1. 1. 2014.
- Ohayon, M. Maurice (2004): Specific characteristics of the pain/depression association in the general population. *The Journal of clinical psychiatry. Supplement* 65 (12): 5–9.
- Patton, Wendy in Patricia Noller (1990): Adolescent self-concept: Effects of being employed, unemployed or returning to school. *Australian Journal of Psychology* 42 (3): 247–59.
- Pearlin, I. Leonard (1989): The sociological study of stress. *Journal of health and social behavior* 3 (3): 241–56.

- Pharr, R. Jennifer, Sheinz Moonie in Timothy J. Bungum (2011): The impact of unemployment on mental and physical health, access to health care and health risk behaviors. *International Scholarly Research Notices* 2012: 1-7.
- Piccinelli, Marco in Greg Wilkinson (2000): Gender differences in depression Critical review. *The British Journal of Psychiatry* 177 (6): 486-92.
- Pivec, Nina, Tamara Sredinšek, Zalika Klemenc-Ketiš in Janko Kersnik (2014): Prevalence of Disease Symptoms in Slovenian Adult Population and Factors Associated with their Prevalence. *Slovenian Journal of Public Health* 53 (3): 262-9.
- Planinšek, Sandra, Branko Škof, Bojan Leskošek, Martina Tomori Žmuc in Maja Pori (2014): Correlation of sports activity with stress and satisfaction with life among adult Slovenians. *Zdravstveno varstvo* 53 (1): 1-10.
- Plesničar Kores, Branka in Andrej Plesničar (2009): Depression and physical activity. *Zdravstveno varstvo* 48 (1): 40-4.
- Radloff, Leonore Sawyer (1977): The CES-D Scale: A Self-Report Depression Scale for Research in the General Population. *Applied Psychological Measurement* 1 (3): 385-401.
- Rokach, Lior (2007): Data mining with decision trees: theory and applications (Vol. 69). London: World Scientific.
- Schunn, D. Christian in Dieter Wallach (2005): Evaluating goodness-of-fit in comparison of models to data. *Psychologie der Kognition: Reden and vorträge anlässlich der emeritierung von Werner Tack*. University of Saarland Press Saarbrueken, Germany 115-54.
- Stankunas, Mindaugas, Ramune Kalediene, Skirmante Starkuviene in Violeta Kapustinskiene (2006): Duration of unemployment and depression: a cross-sectional survey in Lithuania. *BMC Public Health* 6 (1): 174.
- Štraus, Sebastijan (2012): Podatkovno rudarjenje na primeru Zavarovalnice Maribor. *Diplomsko delo visokošolskega strokovnega študijskega programa*. Maribor: Fakulteta za elektrotehniko, računalništvo in informatiko.
- European Social Survey (2014a): ESS06 - Documentation Report. The ESS Data Archive. Dostopno preko http://www.europeansocialsurvey.org/docs/round6/survey/ESS6_data_documentation_report_e02_1.pdf, 29. 12. 2014.
- European Social Survey (2014b): Slovenia - Documents and Data Files. Dostopno preko <http://www.europeansocialsurvey.org/data/country.html?c=slovenia>, 29. 12. 2014.
- Tape, G. Thomas (2014): Interpreting Diagnostic Tests. Dostopno preko <http://gim.unmc.edu/dxtests/>, 29. 12. 2014.
- Touzery, S. Hvalič (2014): Zdravje, počutje in zadovoljstvo z življenjem najstarejših starih v Sloveniji. *Teorija in praksa* 51 (2-3): 458-74.
- Turk, Urška (2005): Čustva, zdravje in bolezni. Ljubljana: Univerza v Ljubljani, Fakulteta za družbene vede.
- Van de Velde, Sarah, Piet Bracke in Katia Levecque (2010): Gender differences in depression in 23 European countries. Cross-national variation in the gender gap in depression. *71 (2): 305-13*.
- Vogrin Hudopisk, Kristina (2002): Depresija pri bolnikih po možganski kapi: ocenjevalne lestvice in vpliv na izid rehabilitacije. *Rehabilitacija* 9 (2): 53-9.

Wahlbeck, Kristian in Mia Mäkinen (2008): Prevention of depression and suicide. Consensus paper. Luksemburg: European Communities.

Williams, Graham (2011): Use R: Data Mining with Rattle and R: the Art of Excavating Data for Knowledge Discovery. New York: Springer.