

UNIVERZA V LJUBLJANI
EKONOMSKA FAKULTETA

MAGISTRSKO DELO

**UPORABA METOD RUDARJENJA PODATKOV ZA
ANALIZO NEPREMIČNINSKIH TRANSAKCIJ V
REPUBLIKI SLOVENIJI IN IZGRADNJO MODELA
ZA TRŽNO VREDNOTENJE NEPREMIČNIN**

Ljubljana, januar 2005

JERNEJ KLEMENC

IZJAVA

Študent Jernej Klemenc izjavljam, da sem avtor tega magistrskega dela, ki sem ga napisal pod mentorstvom prof. dr. Iva Lavrača in somentorstvom izr. prof. dr. Nade Lavrač, in skladno s 1. odstavkom 21. člena Zakona o avtorskih in sorodnih pravicah dovolim objavo magistrskega dela na fakultetnih spletnih straneh

V Ljubljani, dne 3. 1. 2005

Podpis:

ZAHVALA

Najprej bi se rad zahvalil mentorju prof. dr. Ivu Lavraču in somentorici izr. prof. dr. Nadi Lavrač za njune tehtne komentarje, ki so mi bili v veliko pomoč pri izdelavi magistrskega dela.

Zahvalil bi se rad tudi Petru Ljubiču za pomoč pri postavitvi modelov za tržno vrednotenje stanovanjskih nepremičnin z metodo CIPER.

Na koncu bi se rad zahvalil tudi vsem ostalim, ki so mi na kakršen koli način pomagali pri izdelavi magistrskega dela.

Kazalo

1	Uvod.....	1
1.1	Problematika in namen magistrskega dela.....	1
1.2	Cilj magistrskega dela.....	3
1.3	Metode dela.....	3
1.4	Razlaga osnovnih pojmov.....	5
2	Trg nepremičnin v Sloveniji.....	6
2.1	Značilnosti trga nepremičnin.....	6
2.2	Zakonska ureditev razmer na trgu nepremičnin v Sloveniji.....	7
2.3	Metode vrednotenja nepremičnin.....	12
2.3.1	Uvod.....	12
2.3.2	Posamično vrednotenje nepremičnin.....	12
2.3.3	Množično vrednotenje nepremičnin.....	17
3	Uporabljene metode za pripravo podatkov in tržno vrednotenje nepremičnin.....	20
3.1	Uvod.....	20
3.2	Združevanje v skupine.....	22
3.3	Povezovalna pravila.....	23
3.4	Metoda linearne regresije.....	24
3.5	Lokalne metode za napoved.....	26
3.5.1	Uvod.....	26
3.5.2	Metoda najbližjih sosedov.....	26
3.5.3	Regresijski model z lokalnimi utežmi.....	27
3.6	Drevesne metode.....	28
3.6.1	Uvod.....	28
3.6.2	Modelno drevo.....	29
3.6.3	Regresijsko drevo.....	31
3.7	Nevronske mreže.....	31
3.8	Metode za vrednotenje kakovosti in primerjavo modelov.....	33
3.9	Predstavitev uporabniškega programa WEKA.....	34
4	Baza podatkov nepremičninskih transakcij DURS.....	37
4.1	Osnovne značilnosti baze nepremičninskih transakcij.....	37
4.2	Ocena kvalitete baze nepremičninskih transakcij.....	40

4.2.1	Celovitost in kakovost podatkov v bazi.....	40
4.2.2	Ravnanje z manjkajočimi in nepravilnimi podatki.....	43
4.3	Analiza značilnosti podatkov v bazi nepremičninskih transakcij.....	44
4.3.1	Izračun osnovnih statistik za bazo nepremičninskih transakcij.....	44
4.3.2	Značilnosti in vzorci, odkriti z uporabo povezovalnih pravil.....	52
4.4	Dopolnitev baze s podatki iz drugih virov.....	57
5	Primerjava modelov za vrednotenje stanovanj.....	60
5.1	Priprava podatkov za izgradnjo in preverjanje modelov za vrednotenje.....	60
5.1.1	Določitev spremenljivk za izgradnjo modelov za vrednotenje	60
5.1.2	Priprava podatkov za izgradnjo modelov za vrednotenje.....	64
5.2	Model za vrednotenje na osnovi linearne regresije	65
5.3	Model za vrednotenje na osnovi metode najbližjih sosedov	67
5.4	Model za vrednotenje na osnovi regresije z lokalnimi utežmi	68
5.5	Model za vrednotenje na osnovi regresijskega drevesa.....	70
5.6	Model za vrednotenje na osnovi modelnega drevesa	72
5.7	Model za vrednotenje na osnovi nevronske mreže	73
5.8	Primerjava modelov za vrednotenje in izbira najprimernejšega	75
5.9	Ločeni modeli za vrednotenje za posamezna območja transakcij.....	79
5.9.1	Uspešnost napovedi z ločenimi modeli za posamezna območja transakcij.....	79
5.9.2	Študij inflacijskega vpliva in časovnega trenda gibanja cen	81
6	Zaključek	82
7	Literatura in viri.....	85
7.1	Literatura	85
7.2	Viri.....	88
Priloga 1.....		I
Priloga 2.....		VII
Priloga 3.....		XI
Priloga 4.....		XIII
Priloga 5.....		XV
Priloga 6.....		XVIII
Priloga 7.....		XXX

1 Uvod

1.1 Problematika in namen magistrskega dela

Trg nepremičnin definiramo kot stičišče ljudi, ki trgujejo z lastninskimi pravicami nepremičnin. Osnovne funkcije, ki jih opravlja trg nepremičnin, so: menjava lastništva, informiranje o cenah in vrednostih nepremičnin ter prilagajanje kvalitet in količine prostora spremembam socialnih in ekonomskih potreb (Cirman et al. 2000, str. 8). Na ponudbo in povpraševanje na trgu nepremičnin vplivajo različni faktorji, kot so: želje, potrebe in zmožnosti kupcev in prodajalcev, življenjski slog, porazdelitev socialnih kategorij prebivalstva, vplivi okolja, značilnosti nepremičnin in njihova likvidnost, vplivi državnih in lokalnih predpisov, itn. Zato je vrednost nepremičnine odvisna od značilnosti same nepremičnine ter vplivov makro- in mikrookolja, v katerem se nepremičnina nahaja.

Vrednost nepremičnine ocenijo strokovnjaki (cenilci) v postopku vrednotenja nepremičnine. Pri vrednotenju nepremičnin se najpogosteje uporabljajo tržni, dohodkovni ali stroškovni pristop (Cirman et al. 2000, str. 19-32, Nagode 1999, str. 11-29, Eckert 1990). Metoda vrednotenja po tržnem pristopu je ob zadostni količini podatkov večinoma najprimernejša za določanje tržne cene nepremičnin (Nagode 1999, str. 23, Eckert 1990, str. 153). Vrednotenje nepremičnin po vsakem izmed teh pristopov lahko izvajamo za posamične nepremičnine ali za celo množico nepremičnin. Pri posamičnem vrednotenju nepremičnin se oceni vrednost ene nepremičnine, medtem ko se pri množičnem vrednotenju z modelom vrednotenja oceni več nepremičnin. Na množično vrednotenje nepremičnin lahko torej gledamo kot na sistem podatkov, metod in dejavnosti, ki se vzpostavi z namenom ocenjevanja tržnih vrednosti določenih vrst nepremičnin na osnovi razpoložljivih vzorcev in modeliranja tržišča. Vzpostavitev množičnega vrednotenja nepremičnin je potrebna iz več razlogov (Cirman et al. 2000, str. 32, Rotovnik 1997, str. 34-35, Zakrajšek 1993):

- zaradi boljše informiranosti udeležencev na nepremičninskem trgu;
- množica zbranih podatkov lahko služi kot kakovostna podlaga za individualno vrednotenje nepremičnin, ki se pogosto pojavljajo na trgu (npr. stanovanja ali hiše);
- za pomoč cenilcem pri zagotavljanju točnosti podatkov, ocenjevanju različnih modelov vrednotenja in preverjanju dobljenih cen;
- za potrebe davčnega sistema;
- za spremljanje in analiziranje trga nepremičnin, preprečevanje in zaviranje špekulacij;
- za lažje upravljanje nepremičnin v javni lasti;
- za pomoč pri sistemski opredelitvi posameznih sektorjev (npr. stanovanjsko gospodarstvo, urejanje prostora itn.).

Zaradi zgoraj navedenih razlogov smo se odločili, da bomo v sklopu magistrske naloge za delno bazo nepremičninskih transakcij v Republiki Sloveniji poskusili razviti in testirati različne modele za množično vrednotenje nepremičnin po tržni metodi. Za uporabo tržne metode vrednotenja nepremičnin mora obstajati baza izvedenih nepremičninskih transakcij. Na Davčni upravi Republike Slovenije (DURS) že obstaja baza nepremičninskih transakcij od leta 1999 do leta 2002, ki je bila vzpostavljena za namen obračuna davka od nepremičninskih

transakcij. Baza DURS-a je zbirna baza vseh uradno prijavljenih nepremičninskih transakcij, ki so bile izvedene v posameznih občinah. Obsega zapise o preko 100000 nepremičninskih transakcijah, zato je lahko dobra osnova za izgradnjo modelov za tržno vrednotenje nepremičnin. Ker je obstoječa baza nepremičninskih transakcij DURS-a izredno obsežna, je uporaba kakršnih koli ročnih metod za pripravo in analizo podatkov praktično izključena. Poleg tega se lahko zgodi, da zaradi velike časovne oddaljenosti posameznih transakcij in nepopolnosti trga nepremičnin tudi najbližji substituti na trgu nepremičnin dosežejo močno različne cene. Zaradi omenjenih značilnosti smo se odločili, da bomo pristopili k tržnemu vrednotenju nepremičnin z metodami, ki omogočajo obdelavo velike količine podatkov v realnem času. Preden smo se lotili izgradnje različnih modelov za tržno vrednotenje nepremičnin smo poskusili tudi ugotoviti ali nepremičninske transakcije v bazi skrivajo kakšne značilnosti, ki bi lahko bile koristne pri izgradnji modelov za tržno vrednotenje.

V praksi se za množično vrednotenje nepremičnin po tržni metodi uporabljajo najrazličnejše statistične metode. Najbolj se uporabljajo različni regresijski modeli. Njihov spekter se razteza od uporabe linearnih aditivnih regresijskih modelov (Nose 2003) do bolj zapletenih regresijskih modelov, ki poleg stvarnih lastnosti nepremičnin upoštevajo tudi časovno komponento transakcij (Ben-Shahar 2002, Clapp in Giaccotto 1992, Janssen in Söderberg 1999, Pace et al. 2000). V splošnem regresijski pristopi zagotavljajo zadovoljivo natančnost napovedi tržne cene nepremičnine. Vendar je pri uporabi regresijskega pristopa nujno vnaprej specificirati obliko cenovnega modela. Omenjeno slabost je mogoče omiliti z uporabo metod rudarjenja podatkov¹ (Han in Kamber 2001, Hand et al. 2001, Hastie et al. 2001, Lavrač in Bohanec 2003, Lavrač in Grobelnik 2003, Witten in Frank 2000), ki naj bi bile same sposobne izluščiti model za vrednotenje iz dane množice podatkov o nepremičninskih transakcijah. Različne metode rudarjenja podatkov so bile v svetu že uspešno uporabljene za tržno vrednotenje nepremičnin (Pavlov 2000, Iversen 2001).

V Sloveniji se z množičnim vrednotenjem nepremičnin ukvarja Geodetska uprava Slovenije, ki razpolaga z naborom različnih podatkovnih baz, ki omogočajo postavitev celovitih modelov za množično vrednotenje nepremičnin. Tako je bilo v preteklosti že izvedenih nekaj poskusov tržnega vrednotenja nepremičnin z metodami rudarjenja podatkov, vendar so bili narejeni le za relativno majhno število transakcij (Zakrajšek 1993, Zakrajšek 1996). To je morda presenetljivo, saj so bile v preteklosti metode rudarjenja podatkov tudi v Sloveniji večkrat uspešno uporabljene na področju ekonomsko-poslovnih znanosti (Cestnik et al. 2003).

Glavni namen predlaganega magistrskega dela je bil analizirati delno bazo nepremičninskih transakcij v Sloveniji, zgraditi različne modele za tržno vrednotenje nepremičnin in preveriti njihovo uporabnost s podatki o nepremičninskih transakcijah na slovenskem trgu nepremičnin. Pri tem nismo želeli konkurirati prizadevanjem pristojnih služb za postavitev modela tržnega vrednotenja, ampak le predstaviti nekatere alternativne poglede na to problematiko. Ker smo podatke za postavitev cenovnih modelov pridobili iz baze nepremičninskih transakcij DURS-a, smo v sklopu magistrske naloge skušali oceniti njeno primernost za tržno vrednotenje nepremičnin v Sloveniji in predlagali nekatere dopolnitve

¹ Pojem rudarjenje podatkov je prevod angleškega termina "*Data Mining*", ki ga v splošnem uporabljamo za označevanje metod, ki jih je mogoče uporabljati za analizo velikih količin podatkov.

baze, ki so se izkazale za potrebne. Z enostavnimi statističnimi metodami in naprednejšimi metodami rudarjenja podatkov smo analizirali obstoječo delno bazo nepremičninskih transakcij ter skušali ugotoviti, kateri faktorji vplivajo na tržno ceno nepremičnin in kateri ne. Za dano bazo smo poskušali izluščiti značilnosti transakcij, ki bi lahko pomagale pri izgradnji različnih cenovnih modelov. Končno smo zgradili nekaj alternativnih modelov za tržno vrednotenje nepremičnin, jih primerjali med seboj in izbrali najboljšega.

1.2 Cilj magistrskega dela

Cilj magistrskega dela je bil analizirati značilnosti delne baze nepremičninskih transakcij v Sloveniji, poiskati zanimive vzorce v tej bazi in postaviti uporaben model za tržno vrednotenje nepremičnin, ki upošteva vpliv stvarnih lastnosti nepremičnin in časovne oddaljenosti predhodnih transakcij na ceno nepremičnine. Ker niti najboljši model ne more biti kvalitetnejši od podatkov, ki so botrovali njegovemu nastanku, smo ocenili kvaliteto in uporabnost podatkov iz baze nepremičninskih transakcij DURS-a za tržno vrednotenje nepremičnin v Sloveniji. Skušali smo ugotoviti, ali je z uporabo povezovalnih pravil iz obstoječe baze mogoče izluščiti značilnosti transakcij, ki bi lahko koristile pri izgradnji cenovnih modelov. Želeli smo tudi na enem mestu predstaviti uporabnost različnih metod rudarjenja podatkov za vrednotenje nepremičnin po tržni metodi, jih primerjati med seboj in s klasičnim regresijskim modelom ter izbrati najprimernejšo metodo za vrednotenje. Analizo podatkov smo izvedli za celotno bazo, modele za tržno vrednotenje nepremičnin pa smo zgradili le za segment nepremičninskih transakcij s stanovanji. Ta tržni segment glede na število opravljenih transakcij predstavlja relativno največji delež trga nepremičnin v Sloveniji, poleg tega pa je bilo zaradi nekvalitetnih in manjkajočih podatkov v bazi uporabne modele za vrednotenje nepremičnin mogoče zgraditi le za stanovanja.

1.3 Metode dela

Metode dela, ki smo jih uporabili pri izdelavi magistrskega dela, so temeljile na značilnostih pojavov na trgu nepremičnin v Sloveniji, spoznavanju posebnosti tega trga in na proučevanju teoretične podlage, ki nam je omogočila postavitev primernega modela za tržno vrednotenje nepremičnin v Sloveniji. Pri izbiri metodologije smo se naslonili na strokovno literaturo domačih in tujih avtorjev, vire, prispevke in članke z najnovejšimi spoznanji s področja tržnega vrednotenja nepremičnin ter na lastno znanje, pridobljeno med študijem na podiplomskem magistrskem programu. V magistrskem delu smo na podatkih za Slovenijo preverili nekatera spoznanja, ki so se že uveljavila v praksi, hkrati pa smo skušali razširiti metodološki inštrumentarij za vrednotenje nepremičnin z nekaterimi novimi pristopi.

Analizo značilnosti podatkov v delni bazi nepremičninskih transakcij v Republiki Sloveniji smo izvedli za vse vrste nepremičninskih transakcij v obstoječi bazi podatkov. Različne modele za tržno vrednotenje nepremičnin smo zgradili le za segment nepremičninskih transakcij s stanovanji. Pri analizi podatkov v bazi in izgradnji modelov za tržno vrednotenje stanovanj smo se držali CRISP-DM² pristopa, ki opiše postopek reševanje nekega problema v šestih korakih (Chapman et al. 2004, Jorge et al. 2003, str. 63-78):

² CRISP-DM je angleška kratica za pojem *Cross Industry Standard Process for Data Mining*.

- razumevanje problema³,
- razumevanje podatkov,
- priprava podatkov za uporabo,
- postavitve modela,
- ovrednotenje kvalitete modela,
- vključitev modela v poslovno okolje.

Magistrsko nalogo smo razdelili na šest poglavij. Uvodnemu poglavju, v katerem smo predstavili za panogo značilno izrazoslovje, sledi poglavje, v katerem smo opredelili značilnosti trga nepremičnin in njegove posebnosti. Dotaknili smo se zakonov, ki urejajo nepremičninske transakcije v Sloveniji, in zaključili poglavje s pregledom značilnosti metod za ocenjevanje vrednosti nepremičnin v primeru posamičnega in množičnega vrednotenja. V tretjem poglavju smo predstavili osnovne značilnosti statističnih metod in metod rudarjenja podatkov, ki smo jih uporabili za urejanje in pripravo podatkov v bazi, iskanje značilnosti in zanimivih vzorcev v transakcijah ter izgradnjo modelov za tržno vrednotenje nepremičnin.⁴ Spekter predstavljenih metod obsega klasične (linearna regresija) in novejšje pristope (povezovalna pravila, metoda lokalnih napovedi, regresijska drevesa, nevronske mreže). Pri tem smo se naslonili na najnovejšo strokovno literaturo (Pindyck in Rubinfeld 1998, Han in Kamber 2001, Hand et al. 2001, Hastie et al. 2001, Lavrač in Bohanec 2003, Lavrač in Grobelnik 2003, Sharma 1996, Witten in Frank 2000). V četrtem poglavju smo natančneje predstavili bazo nepremičninskih transakcij DURS-a. Ocenili smo kvaliteto baze v smislu celovitosti podatkov in njene uporabnosti za tržno vrednotenje nepremičnin. Z enostavnimi statističnimi metodami in s povezovalnimi pravili smo skušali izluščiti značilnosti transakcij, ki bi lahko pomagale pri izgradnji cenovnih modelov. Zapise z manjkajočimi in nedoslednimi podatki smo pred nadaljnjo uporabo ustrezno preoblikovali. Bazo podatkov smo dopolnili s faktorji, ki bi lahko vplivali na kakovost zgrajenih modelov za vrednotenje: območja transakcij v smislu regij, razvitost občin, avtocestne povezave z Ljubljano, okvirne geografske koordinate občin. V petem poglavju smo zgradili različne modele za množično vrednotenje stanovanj v Republiki Sloveniji po tržni metodi. Pri tem smo si pomagali s programskim vmesnikom za uporabo metod rudarjenja podatkov WEKA (WEKA 2003). Model linearne regresije je predstavljal referenčni model za vrednotenje enote stanovanjske površine. Z njim smo primerjali modele za vrednotenje, ki smo jih zgradili z različnimi metodami rudarjenja podatkov: metoda najbližjih sosedov, regresijski model z lokalnimi utežmi, regresijsko drevo, modelno drevo, nevronska mreža. Za vsak model smo z metodo prečnega preverjanja⁵ ocenili

³ ang. *business understanding*

⁴ Metode rudarjenja podatkov v splošnem delimo na: a) metode napovedne indukcije (ang. *predictive induction*) uporabne za klasifikacijo in predikcijo ter b) metode opisne indukcije (ang. *descriptive induction*) uporabne za opis lastnosti zanimivih skupin podatkov. V magistrskem delu smo večinoma uporabili le napovedne metode. Od uporabljenih metod spadajo med opisne metode le povezovalna pravila in metoda združevanja v skupine.

⁵ ang. *cross-validation*

njegovo primernost in uporabnost ter v duhu obravnavanega problema izbrali najboljšega. V šestem poglavju smo povzeli sklepe, do katerih smo prišli pri izdelavi magistrske naloge.

1.4 Razlaga osnovnih pojmov

Nepremičnina (ang. *real-estate*) je zemljišče skupaj z njegovo izboljšavo.

Zemljišče (ang. *land*) je parcela ali določljiv del zemljiške oziroma katastrske parcele.

Izboljšava (ang. *improvement*) je vse, kar je zgrajeno na zemljišču. Izboljšava je lahko objekt, stavba, zunanja ureditev itn.

Objekt je lahko zgradba – stavba, v kateri živimo ali delamo, lahko je gradbeni objekt kot sta na primer most ali cesta, lahko pa so tudi druge vrste objektov, kot so na primer spomeniki.

Metode vrednotenja nepremičnin so metode, s katerimi ocenimo vrednost nepremičnin. V praksi se uporabljajo tri glavne metode za ocenjevanje vrednosti nepremičnin: metoda prodajnih cen, metoda donosov in metoda nabavnih cen. Metode vrednotenja delimo še na metode za posamično vrednotenje nepremičnin in metode za množično vrednotenje nepremičnin.

Tržna metoda (ang. *market approach, sales-comparison approach*) temelji na primerjavi ocenjevane nepremičnine s cenami in značilnostmi že prodanih nepremičnin. Za to metodo uporabljamo tudi izraza metoda prodajnih cen in pristop primerljivih prodaj.

Dohodkovna metoda (ang. *income approach, income capitalization method, discounted cash flow method*) temelji na izračunu kapitaliziranega donosa iz poslovanja z nepremičnino. Za to metodo uporabljamo tudi izraza metoda prodajnih cen in prihodkovni pristop.

Stroškovna metoda (ang. *cost approach*) temelji na amortizirani nabavni ceni objekta in tržni vrednosti praznega zemljišča. Za to metodo uporabljamo tudi izraz metoda nabavnih cen.

Posamično vrednotenje nepremičnin se uporablja za ocenjevanje vrednosti ene same nepremičnine.

Množično vrednotenje nepremičnin se uporablja za hkratno ocenjevanje vrednosti več nepremičnin. Množično vrednotenje nepremičnin definiramo kot sistem podatkov, metod in dejavnosti, ki se vzpostavi z namenom ocenjevanja tržnih vrednosti določenih vrst nepremičnin na osnovi razpoložljivih vzorcev in modeliranja tržišča.

Transakcijska cena je cena, za katero se dogovorita kupec in prodajalec. Lahko se razlikuje od vrednosti nepremičnine, ki jo ocenimo z eno izmed metod za ocenjevanje vrednosti nepremičnin.

Rudarjenje podatkov (ang. *data mining*) je termin, ki ga v splošnem uporabljamo za označevanje vseh metod, s katerimi je mogoče odkrivati zanimive modele ali zakonitosti v podatkih in ki izvirajo iz raziskav strojnega učenja na področju umetne inteligence.

Model za vrednotenje nepremičnin predstavlja statistične zakonitosti, ki povezujejo ceno nepremičnine z njenimi značilnostmi kot so na primer komunalna opremljenost, starost, lokacija, lega itn. Je rezultat z matematičnega modeliranja ali rudarjenja podatkov.

Metoda prečnega preverjanja je metoda, s katero skušamo oceniti povprečno kvaliteto modela za vrednotenje nepremičnin, ki ga lahko zgradimo iz danih podatkov.

2 Trg nepremičnin v Sloveniji

2.1 Značilnosti trga nepremičnin

Že v uvodu smo napisali, da definiramo trg nepremičnin kot stičišče ljudi, ki trgujejo z lastninskimi pravicami nepremičnin. Osnovne funkcije, ki jih opravlja trg nepremičnin, so menjava lastništva, informiranje o cenah in vrednostih nepremičnin ter prilagajanje kvalitete in količine prostora spremembam socialnih in ekonomskih potreb (Cirman et al. 2000, str. 8). Na ponudbo in povpraševanje na trgu nepremičnin vplivajo različni faktorji: želje, potrebe in zmožnosti kupcev in prodajalcev, življenjski slog, porazdelitev socialnih kategorij prebivalstva, vplivi okolja, značilnosti nepremičnin in njihova likvidnost, vplivi državnih in lokalnih predpisov itn.

Vplive različnih faktorjev na ceno stanovanjskih objektov je analiziralo že veliko avtorjev. Tako je na primer Ioannides (2003) analiziral vpliv soseščine in socialnih interakcij na ceno stanovanj, Forgey (1996) je analiziral vpliv likvidnosti posameznih nepremičnin na njeno ceno, Plaut (1998) pa je postavil model, s katerim je mogoče na osnovi lastnosti in cene nepremičnin identificirati enega ali več centrov v urbanih središčih. V slovenski literaturi sta zaradi relativno majhne časovne oddaljenosti morda zanimivi dve študiji vpliva različnih faktorjev na ceno stanovanj: opisna analiza vpliva velikosti in lokacije nepremičnin v Ljubljani na njihovo ceno (Štefančič 2003) ter statistična analiza vpliva lokacije, velikosti in starosti na ceno stanovanj v Sloveniji (Nose 2003).

Če opišemo ideal trga popolne konkurence z naslednjimi značilnostmi (Prašnikar in Debeljak 1998, str. 242-243, Cirman et al. 2000, str. 9):

- homogenost proizvodov,
- veliko število kupcev in prodajalcev, ki na ceno tržnih dobrin ne morejo vplivati,
- odsotnost umetnih ovir,
- popolna mobilnost proizvodnih dejavnikov, proizvodov in kupcev,
- popolna informiranost kupcev in prodajalcev,

vidimo, da se trgi nepremičnin močno razlikujejo od ideala trga čiste in popolne konkurence, saj zanje v splošnem veljajo naslednje lastnosti (Cirman et al. 2000, str. 9,10):

- nehomogenost proizvodov – vsaka nepremičnina je edinstvena, zato ji ni mogoče najti popolnega substituta,
- v določenem cenovnem razredu je le omejeno število kupcev in prodajalcev,
- prisotnost umetnih ovir – vplivi državnih in lokalnih predpisov, nepredvidljivost vladnih programov itn.,
- nemobilnost nepremičnin in njihova vpetost v obstoječe okolje,
- slaba informiranost na trgu, saj ne gre za redne kupce.

Zato trgi nepremičnin včasih dajejo vtis neurejenosti in neučinkovitosti. Čeprav na njih ne vlada popolna konkurenca, so konkurenčne sile kljub temu pomembne pri določanju cen nepremičnin. Trg nepremičnin se na spremembe povpraševanja in ponudbe odziva podobno

kakor uči ekonomska teorija. Tako na primer primerljive nepremičnine dosegajo podobno ceno, v primeru presežka povpraševanja nad ponudbo se vzpostavijo visoke cene nepremičnin, v obratnem primeru se cene nepremičnin nižajo, nižji stroški posojil navadno zvišujejo povpraševanje po nepremičninah, davčni predpisi in regulacije lahko spodbujajo ali zavirajo razvoj na področju nepremičnin itn. Odstopke trga nepremičnin od ekonomske teorije navadno pojasnjujemo z njegovo nepopolnostjo (Cirman et al. 2000, str. 12).

Trg nepremičnin lahko delimo glede na različne dejavnike. Glede na trgovano nepremičnino razdelimo trg nepremičnin na različne segmente: trg stanovanjskih nepremičnin, trg poslovnih nepremičnin, trg industrijskih nepremičnin, trg kmetijskih nepremičnin in trg nepremičnin s posebnim pomenom. Če se osredotočimo na lastninske ali uporabniške pravice, razdelimo trg na najemniški trg in na lastniški trg (Cirman et al. 2000, str. 11). V magistrski nalogi smo se v celoti posvetili le lastniškemu trgu nepremičnin, ker smo obravnavali samo transakcije z lastniškimi pravicami nepremičnin.

Zaradi množice vplivnih faktorjev in značilnosti trga nepremičnin se lastniška struktura na trgih nepremičnin spreminja od države do države. Velik vpliv na lastniško strukturo imata zagotovo tudi država, ki na trg nepremičnin vpliva z različnimi davčnimi politikami ali s politikami subvencij, ter finančno okolje v določeni sredini. Tako je že v nam bližnjih deželah lastniška struktura nepremičnin zelo različna: v Nemčiji je delež gospodinjstev, ki je imelo leta 1988 v lasti neko nepremičnino, okoli 46%, v Italiji je bilo leta 1993 v privatnih rokah 75% stanovanj in v Franciji leta 1988 54% (Cirman et al. 2000, Zakrajšek 1998).

Razmere na slovenskem trgu nepremičnin se od teh močno razlikujejo, saj je po nekaterih podatkih pri nas skoraj 90% stanovanj v lasti stanovalcev (Nose 2003, str. 1). Tako visok odstotek lastniških stanovanj je predvsem posledica Stanovanjskega zakona iz leta 1991 oziroma tako imenovanega Jazbinškovega zakona (SZ – Uradni list RS, št. 18/91), ki je omogočil ljudem pridobitev lastninske pravice za družbena stanovanja po izredno ugodnih cenah. Poleg tega je obseg nepremičninskega trga v Sloveniji mnogo manjši kakor v prej omenjenih državah. V Sloveniji je bilo leta 2002 okoli 460.000 stavb s stanovanji. Od tega je bilo največ samostojnih hiš, dvojčkov ali vrstnih hiš (preko 442.000). Večstanovanjskih stavb je bilo okoli 18.000. Stanovanj je bilo okoli 778.000, med njimi pa so prevladovala srednje velika stanovanja s površino med 61 in 80 kvadratnimi metri. Največ je bilo dvosobnih stanovanj. Po številu stanovanj je prednjačila osrednjeslovenska statistična regija, sledile so ji podravska in savinjska statistična regija. Najmanj stanovanj je bilo v notranjsko-kraški in Zasavski statistični regiji (Matejčič 2003).

2.2 Zakonska ureditev razmer na trgu nepremičnin v Sloveniji

Razmere na trgu nepremičnin v Sloveniji ureja več zakonov in drugih dokumentov, ki veljajo bodisi striktno za področje urejanja razmer na trgu nepremičnin bodisi urejajo splošna razmerja, ki se pojavljajo v razmerjih med kupci in prodajalci. Najpomembnejši zakoni in dokumenti, ki urejajo razmerja na trgu nepremičnin so naslednji:

- **Obligacijski zakonik** (Uradni list RS, št. 83/2001), ki vsebuje temeljna načela in splošna pravila za vsa obligacijska razmerja. Za obligacijska razmerja, ki jih urejajo drugi zakoni, se uporabljajo določbe obligacijskega zakonika glede vprašanj, ki niso urejena v teh zakonih. Za področje nepremičninskega trga je pomemben predvsem tisti

del obligacijskega zakonika, ki obravnava posredniške pogodbe ter opredeljuje pravice naročnika in posrednika, ki se nahajata v obligacijskem razmerju. Tako je posrednik dolžan iskati priložnost, obveščati naročnika ter ravnati odgovorno. Pod določenimi pogoji je upravičen do plačila za posredovanje.

- **Zakon o varstvu potrošnikov** (Uradni list RS, št. 14/2003), ki ureja pravice potrošnikov pri ponujanju, prodajanju in drugih oblikah trženja blaga in storitev s strani pravnih ali fizičnih oseb in določa dolžnosti državnih organov in drugih subjektov, da te pravice zagotavljajo. Poseben pomen tega zakona ima določilo, da pravic, ki pripadajo potrošniku v skladu s tem zakonom, s pogodbo ni mogoče omejiti ali izključiti. Glavni nosilci varstva potrošnikov so Urad RS za varstvo potrošnikov in organizacije potrošnikov. Pri varovanju potrošnikov sodelujejo tudi drugi (na primer pristojni inšpektorati in nepremičninska združenja).
- **Stanovanjski zakon** (Uradni list RS, št. 69/2003), ki ureja vrste stanovanjskih stavb, pogoje za njihovo vzdrževanje, pogoje za načrtovanje stanovanj, lastninskopravna razmerja in upravljanje v večstanovanjskih stavbah, stanovanjska najemna razmerja, gradnjo in prodajo novih stanovanj, pomoč pri pridobitvi in uporabi stanovanja, pristojnosti in naloge države, občin ter drugih organov in organizacij, ki delujejo na stanovanjskem področju, registre in ustrezno inšpekcijsko nadzorstvo. Zakon tudi določa, da se za lastninskopravna razmerja in za upravljanje v večstanovanjskih stavbah uporabljajo določbe stvarnopravnega zakonika (Uradni list RS, št. 87/2002).
- **Zakon o graditvi objektov** (Uradni list RS, št. 110/2000), ki ureja pogoje za graditev vseh objektov, določa bistvene zahteve in njihovo izpolnjevanje glede lastnosti objektov, predpisuje način in pogoje za opravljanje dejavnosti, ki so v zvezi z graditvijo objektov, ureja organizacijo in delovno področje Zbornice za arhitekturo in prostor Slovenije ter Inženirske zbornice Slovenije, ureja inšpekcijski nadzor ter ureja druga vprašanja, ki so povezana z graditvijo objektov.
- **Predvideni zakon o davku na nepremičnine**, ki naj bi stopil v veljavo leta 2005. Novi zakon naj bi uvedel proporcionalni davek na nepremičnine, ki naj bi nadomestil davek na posest stavb, ki ga določa zakon o davkih občanov (Uradni list RS, št. 8/1991, 7/1993), in nadomestilo za uporabo stavbnega zemljišča, ki ga določa zakon o stavbnih zemljiščih (Uradni list RS, št. 44/1997). Cilj davka naj bi bila racionalizacija uporabe stanovanj. Večina sredstev, zbranih z novim davkom, naj bi pripadla občinam. Davčna stopnja še ni določena. Zanimivost predvidenega zakona je v tem, da naj bi davčno osnovo predstavljala tržna vrednost nepremičnine, ki bi bila določena z metodo množičnega vrednotenja nepremičnin. Pri tem naj bi bila najpomembnejša dejavnika leto izgradnje in velikost nepremičnine (Nose 2003, str. 15-16).
- **Zakon o davku na promet nepremičnin** (Uradni list RS, št. 57/99), ki ureja promet nepremičnin z izjemo novogradenj. Promet nepremičnin je v skladu s tem zakonom definiran kot: odplačni prenos lastninske pravice, zamenjava nepremičnin, nadaljnja prodaja nepremičnin, oddaja nepremičnine v najem za najmanj 61 dni, oddaja kmetijskega zemljišča v zakup za najmanj leto dni (Jošt 2002, str. 25).
- **Zakon o zemljiški knjigi** (Uradni list RS, št. 58/2003), ki določa, da je zemljiška knjiga javna knjiga, namenjena vpisu in objavi podatkov o pravicah na nepremičninah

in pravnih dejstvih v zvezi z nepremičninami. Ta zakon temelji na naslednjih temeljnih načelih: načelo javnosti, začetek učinkovanja vpisov, načelo publicitetnih in oblikovalnih učinkov vpisa v zemljiško knjigo, načelo zaupanja v zemljiško knjigo, načelo pravnega prednika in načelo vrstnega reda (Pirih 2004, str. 8-10).

- **Kodeks o dobrih poslovnih običajih pri poslovanju z nepremičninami** in **Katalog nalog, opravil in storitev pri prometu z nepremičninami**. Kodeks o dobrih poslovnih običajih pri poslovanju z nepremičninami so novembra 1996 leta sprejeli poslovni subjekti združenja za poslovanje z nepremičninami pri Gospodarski zbornici Slovenije in z njim uredili pravila obnašanja nepremičninskih družb, ki niso bila regulirana s takratnim zakonom o obligacijskih razmerjih kot predhodnikom današnjega obligacijskega zakonika. Kodeks o dobrih poslovnih običajih pri poslovanju z nepremičninami zavezuje podpisnike k varovanju lastnih interesov in interesov naročnikov, k lojalni konkurenci med podpisniki in medsebojni pomoči v primerih nelojalnega ravnanja strank, k strokovnemu, poštenemu in vestnemu ravnanju, spoštovanju dobrih poslovnih običajev, k upoštevanju kataloga storitev ter k urejanju medsebojnih sporov. Med leti 1996 in 2002 naj bi h kodeksu pristopilo 328 gospodarskih združb. Poleg kodeksa o dobrih poslovnih običajih pri poslovanju z nepremičninami imajo svoja kodeksa tudi v Nepremičninski zbornici Slovenije in Društvu nepremičninskih posrednikov Slovenije. Namen vseh treh kodeksov je podoben (Nose 2003, str. 11-12, Kožar in Marinšek 2003, str. 19).
- **Zakon o nepremičninskem posredovanju** (Uradni list RS, št. 42/2003) je zagotovo najpomembnejši zakon s področja zakonske ureditve transakcij z lastninskimi pravicami nepremičnin, zato bomo njegovo vsebino obdelali bolj podrobno.

Zakon o nepremičninskem posredovanju je stopil v veljavo 24. maja 2003 in je prvi zakon, ki opredeljuje pravno podlago za posredovanje pri izvajanju transakcij z lastninskimi pravicami nepremičnin. Najpomembnejša področja, ki jih ureja zakon o nepremičninskem posredovanju so naslednja (Uradni list RS, št. 42/2003, Kožar in Marinšek 2003, str. 41-57):

- *Nepremičninskim družbam in nepremičninskim posrednikom določa pogoje za opravljanje posredovanja v prometu z nepremičninami*. Vstopni prag za gospodarske družbe in podjetnike posameznike v posredniško dejavnost načelno ureja 4. člen zakona o gospodarskih družbah (Uradni list RS, št. 30/1993, glej tudi Kožar in Marinšek 2003, str. 54-56). Zakon o nepremičninskem posredovanju pa zvišuje pogoje za vstop v posredovanje v prometu z nepremičninami zaradi:
 - zagotovitve ustreznih prostorov za opravljanje dejavnosti – potreben je vsaj en ločen prostor za pogovore s strankami;
 - redne ali pogodbene zaposlitve osebe s statusom nepremičninskega posrednika z licenco – pridobitev licence za nepremičninsko posredovanje ni vezana na določeno nepremičninsko družbo;
 - obveznega zavarovanja odgovornosti.

Sam postopek ustanavljanja nepremičninske družbe se bistveno ne razlikuje od ustanavljanja drugih kapitalskih ali osebnih družb. S podjetniškega vidika je ceneje poiskati nepremičninskega posrednika na trgu dela kot ustanoviti nov poslovni

subjekt. Če želi obstoječe podjetje opravljati dejavnost nepremičninskega posredovanja kot stransko dejavnost, mora doregistrirati dejavnost s šifro 70.31 po Standardni klasifikaciji dejavnosti, pridobiti nepremičninskega posrednika z licenco, urediti ustrezen poslovni prostor, urediti bistvene evidence za poročanje v skupno bazo na Ministrstvu za okolje in prostor, urediti zavarovanje odgovornosti, pripraviti splošne pogoje in redno ali po pogodbi zaposliti diplomiranega pravnika za sestavo listin (Kožar in Marinšek 2003, str. 41-43).

- *Določa zgornjo mejo plačila za posredovanje pri izvedbi transakcije z nepremičnino.* Zakon o nepremičninskem posredovanju omejuje provizijo na 4% od vrednosti posredovane nepremičnine, kadar gre za nakup. Ta omejitev ne velja pri kupoprodajah nepremičnin, kjer je cena nižja od 10.000 EUR. Omejitve pri posredovanju drugih nepremičnin (na primer najem ali lizing) niso predvidene. Od 1. januarja 2002 se na 4% plačilo pri posredovanju obračunava še 20% DDV, kar pomeni, da znaša celoten znesek plačila 4.8%. Dejanska provizija na trgu se giblje med 2% in 3%, odvisno od velikosti nepremičnine. To je predvsem posledica velike konkurence pri nujenju storitev nepremičninskega posredovanja, saj je bilo ob koncu leta 2002 za to dejavnost registriranih 474 gospodarskih subjektov. Konkurenco posrednikom predstavljajo tudi naročitelji sami, saj le-ti kar 70% kupoprodajnih poslov opravijo brez posrednikov (Kožar in Marinšek 2003, str. 43).
- *Določa pravila za varno in skrbno poslovanje pri nepremičninskem posredovanju, ki so jih nepremičninske družbe in nepremičninski posredniki dolžni upoštevati pri svojem delu.* Med pravila o varnem poslovanju sodijo člani, ki zakonsko obvezujejo nepremičninske posrednike k vrsti dejanj, s katerimi zagotavljajo varnost naročitelja:
 - nepristranost pri posredovanju,
 - razkritje nasprotja interesov,
 - preverjanje stanja nepremičnine,
 - razumljivo pojasnjevanje večjega števila okoliščin, kot so na primer: tržne razmere, predpisi, davki in morebitna tveganja.

Zakon o nepremičninskem posredovanju tako izrecno določa, da mora biti pogodba o posredovanju v pisni obliki. Med pomembnejšimi določili zakona sta tudi ureditev o strokovnem izpitu in pogojih za podelitev licence za nepremičninske posrednike ter obvezno zavarovanje za odškodninsko odgovornost v višini najmanj 40 milijonov SIT za posamezen primer (Kožar in Marinšek 2003, str. 44-45).

- *Določa javne evidence o posredovanju v prometu z nepremičninami.* V skladu s tem zakonom se morata vzpostaviti dve bazi podatkov o transakcijah na nepremičninskem trgu. Prva baza izvira iz sistema davka na promet nepremičnin in naj bi temeljila na podlagi podatkov Davčne uprave republike Slovenije (DURS). Datum za začetek oblikovanja te baze ni predpisan, minister za okolje in prostor pa mora v soglasju z ministrom za finance določiti način in roke za posredovanje takšnih podatkov. Na DURS-u že obstaja delna baza nepremičninskih transakcij od leta 1999 do leta 2002, ki je bila vzpostavljena za namen obračuna davka od nepremičninskih transakcij. Baza DURS-a predstavlja zbirno bazo vseh uradno prijavljenih nepremičninskih transakcij,

ki so se izvedle v posameznih občinah, in obsega zapise o preko 100000 nepremičninskih transakcij (Baza podatkov DURS 2003). To bazo smo uporabili tudi mi za postavitev modelov za tržno vrednotenje nepremičnin. Druga baza, ki jo zahteva zakon o nepremičninskem posredovanju, je skupna baza nepremičninskih družb, ki naj bi izvirala iz poročil samih nepremičninskih družb. Na ta način naj bi imeli v prihodnosti javni dostop do podatkov o realiziranih cenah za okoli 30% kupoprodajnih nepremičninskih transakcij. Približno takšen je namreč obseg kupoprodaj nepremičnin, pri katerih posredujejo nepremičninske družbe (Kožar in Marinšek 2003, str. 45).

- *Določa ekskluzivno posredovanje pri izvedbi transakcije z nepremičnino.* Novost, ki jo uvaja zakon o nepremičninskem posredovanju, je ekskluzivno posredovanje nepremičninske družbe, razen če ni posebej drugače dogovorjeno. Če ni drugače dogovorjeno, bo trajalo pogodbeno razmerje o nepremičninskem posredovanju devet mesecev od sklenitve posredniške pogodbe. S to določbo naj bi se zmanjšalo beganje potrošnika od enega posrednika do drugega, s čimer naj bi se trg izčistil zaradi manjše migracije potrošnikov (Kožar in Marinšek 2003, str. 46).
- *Opređeljuje strokovne podlage za posamično ocenjevanje tržne vrednosti nepremičnine.* Z dnem uveljavitve zakona o nepremičninskem posredovanju je prenehal veljati več kakor 16 let star Pravilnik o enotni metodologiji (PEM) (Uradni list SRS, št. 8/87), ki se je uporabljal za izračun prometne vrednosti hiš in stanovanj ter drugih nepremičnin. Minister, ki je pristojen za okolje in prostor, ima možnost predpisati administrativno metodologijo za ocenjevanje nepremičnin, kar podpirajo tudi nekateri sodni cenilci. Po drugi strani pa tisti, ki so bližje tržnim načelom oziroma mednarodnim standardom ocenjevanja nepremičnin, menijo, da takšne uradne metode ne potrebujemo, ker zadostuje Slovenski poslovno finančni standard (SPS2), ki ga je konec leta 1997 sprejel Slovenski inštitut za revizijo (Kožar in Marinšek 2003, str. 46, 214). Slovenski inštitut za revizijo je tudi objavil štiri vsebinska pojasnila, ki se nanašajo na ocenjevanje vrednosti podjetij strojev in nepremičnin. Poleg tega imajo slovenski ocenjevalci nepremičnin svoj Kodeks poklicne etike, ki se prav tako uporablja za ocenjevanje vrednosti podjetij, strojev opreme in nepremičnin. Tovrstni kodeks zavezuje tiste ocenjevalce nepremičnin, ki so vpisani v register pri Slovenskem inštitutu za revizijo. Svoj Kodeks ocenjevanja imajo tudi sodni izvedenci in cenilci gradbene stroke, ki so člani Društva sodnih izvedencev in cenilcev gradbene stroke v Sloveniji (SICGRAS) (Kožar in Marinšek 2003 str. 215-216). Minister, ki je pristojen za okolje in prostor, naj bi v soglasju s ministrom za finance predpisal navodilo o usposabljanju cenilcev in priznavanju znanj s področja uporabe standardov, skladno z Zakonom o revidiranju (Uradni list RS, št. 26/2003). Navodilo naj bi poenotilo izobraževalne vsebine za cenilce, ki ocenjujejo vrednost nepremičnin za potrebe sodišč in za komercialne potrebe na slovenskem trgu nepremičnin (Kožar in Marinšek 2003 str. 47).
- *Opređeljuje pogoje za opravljanje poklica nepremičninskega posrednika.* V skladu s tem zakonom pridobi poklic nepremičninskega posrednika samo tisti, ki opravi izpit na Ministrstvu za okolje in prostor in je vpisan v imenik. Na podlagi 31. člena Zakona o nepremičninskem posredovanju bo izdan podzakonski akt z opisom znanj, ki jih

mora kandidat obvladati, če želi pridobiti licenco za opravljanje posredovanja v prometu z nepremičninami.

- *Ureja inšpekcijsko nadzorstvo nad izvajanjem zakona.*

2.3 Metode vrednotenja nepremičnin

2.3.1 Uvod

Vrednost nepremičnine je odvisna od značilnosti same nepremičnine ter vplivov ožjega in širšega okolja, v katerem se nepremičnina nahaja. Značilnosti nepremičnine so njena velikost, izgled, kakovost, lokacija, lega, komunalna opremljenost itn. Ožje okolje vpliva na vrednost nepremičnine preko ponudbe različne infrastrukture, ekoloških pogojev ekonomske razvitosti področja, značilnosti ljudi, ki tam živijo itn. Širše okolje predstavljata politična in zakonska ureditev, ki na vrednost nepremičnine vplivata preko pravic in obveznosti prebivalcev, davčne politike, zagotavljanja varnosti s strani državnih organov itn. (Cirman et al. 2000, str. 15).

Vrednost nepremičnine ocenijo strokovnjaki v postopku vrednotenja nepremičnine. Vrednotenje torej pomeni ustvarjanje mnenja o vrednosti nepremičnine s strani strokovnjaka. Tovrstne strokovnjake imenujemo cenilci in njihova naloga je, da pri ocenjevanju vrednosti nepremičnine upoštevajo vse razpoložljive informacije, ki lahko vplivajo na vrednost proučevane nepremičnine (Cirman et al. 2000, str. 15).

Ocenjevanje vrednosti nepremičnin služi različnim namenom (Cirman et al. 2000, str. 15-16):

- kot pomoč prodajalcem in kupcem pri transakcijah z nepremičninami,
- kot osnova za določitev višine posojila, če je posojilo zavarovano z nepremičnino,
- za določitev zavarovalne vsote oziroma pri ocenitvi višine nastale škode,
- za oceno vrednosti podjetij v primerih spojitve podjetij, stečaja ali likvidacije podjetja,
- za davčne namene itn.

Ocenjevanje vrednosti nepremičnin je urejen proces, ki poteka v naslednjih korakih: opredelitev problema, izdelava načrta vrednotenja, zbiranje in analiza potrebnih podatkov, izbira ustrezne metode vrednotenja, uskladitev različnih ocen med seboj in izdelava končnega poročila (Cirman et al. 2000, str. 16, Eckert 1990, str. 88-109).

V splošnem razdelimo metode vrednotenja nepremičnin na metode posamičnega vrednotenja nepremičnin in na metode množičnega vrednotenja nepremičnin. V prvem primeru ocenjujemo vrednost posamične nepremičnine, za katero ocenimo njeno vrednost z uporabo ustreznih metod na osnovi zbranih podatkov, ki so povezani s to nepremičnino. Drugačen pristop uberemo pri množičnem vrednotenju nepremičnin, kjer vrednost nepremičnin ugotavljamo s pomočjo modeliranja trga in uporabe statističnih metod. Posamezne metode vrednotenja nepremičnin so opisane v naslednjih dveh podpoglavjih.

2.3.2 Posamično vrednotenje nepremičnin

V tržnem gospodarstvu je namen posamičnega vrednotenja nepremičnin običajno oskrbeti stranko z realnimi in celovitimi informacijami o vrednosti nepremičnine. S tem naj bi stranki olajšali delovanje v skladu z njenimi interesi. Pri posamičnem vrednotenju nepremičnin se

vrednost nepremičnine v splošnem ocenjuje s tremi pristopi: s tržno metodo, z dohodkovno metodo ali s stroškovno metodo. Izbira posamezne metode vrednotenja je odvisna od namena vrednotenja (Cirman et al. 2000, str. 19-30, Eckert 1990, str. 113-303, Nagode 1999, str. 11-26). V nadaljevanju bomo opisali omenjene tri metode.

Tržna metoda

Za tržno metodo se v literaturi uporabljata tudi izraza metoda primerljivih prodaj in metoda prodajnih cen. Osnovne značilnosti te metode bomo povzeli po delih Cirmanove in soavtorjev (2000, str. 19-22), Eckerta (1990, str. 153-176) in Nagodetove (1999, str. 23-25).

Pristop k vrednotenju nepremičnin s tržno metodo je sorazmerno enostaven. Večan je na tržne podatke, njegova logika pa temelji na principu substitucije. Racionalen kupec naj bi za neko nepremičnino plačal največ toliko, kolikor znašajo stroški pridobitve podobne nepremičnine z enako koristnostjo. Cenilec oceni vrednost ocenjevane nepremičnine na osnovi prodajnih cen primerljivih nepremičnin z upoštevanjem razlik med njimi in različnih pogojev prodaje. Ob zadostni količini tržnih podatkov je ta metoda večinoma najprimernejša za ocenjevanje tržne vrednosti nepremičnin.

Pri uporabi te metode mora cenilec poiskati in preveriti pred kratkim opravljene prodaje primerljivih nepremičnin ter prilagoditi prodajno ceno zaradi razlik med primerljivimi nepremičninami in ocenjevano nepremičnino. Dolžina obdobja, v katerem zajema podatke, je odvisna od stabilnosti gospodarskih in tržnih razmer. Podatke o primerljivih nepremičninah poišče cenilec v svoji bazi podatkov, v sodnih zapiskih, pri posrednikih z nepremičninami, pri ustanovah, ki se ukvarjajo s financiranjem nakupa nepremičnin itn. Cenilec primerjave navadno izvaja tako, da oceni vrednost na enoto nepremičnine, ki je lahko m² površine, število sob, število stanovanj itn. Pri tem mora paziti na to, da je bila transakcija s primerljivo nepremičnino opravljena med neodvisnim kupcem in prodajalcem, brez pritiskov na katero izmed obeh strank, da sta bili stranki povprečno obveščeni in da sta ravnali ekonomsko racionalno. Pomembne značilnosti nepremičnin, ki določajo njihovo vrednost so: starost, velikost, število spalnic, število kopalnic, kakovost gradnje itn. Vsaka značilnost je ovrednotena na trgu.

Za oceno vrednosti navadno zadostujejo tri v preteklosti prodane nepremičnine. Če je primerljivih prodaj več, je zanesljivost ocenjene vrednosti večja. Na podlagi ugotovljenih podobnosti in odstopanj cenilec prilagaja prodajne cene primerljivih nepremičnin ocenjevani nepremičnini. Prilagoditve cen so lahko nominalne, odstotkovne, združene ali posamezne. Cenilec jih določi z analizo prodaje po parih, z multiplo regresijsko analizo ali s pomočjo nabavnih vrednosti. Z *analizo prodaje po parih* cenilec med seboj primerja cene, dosežene pri prodaji nepremičnin, ki so med seboj enake v vseh značilnosti razen ene. Na ta način oceni tržno vrednost te značilnosti. Tako cenilec navadno vrednoti prilagoditve zaradi velikosti, načina gradnje, opremljenosti itn. Pri *multipli regresijski analizi* identičnost primerljivih nepremičnin ni pogoj, vendar je iz analize treba izključiti prodaje nepremičnin, ki so bile opravljene v drugačnih tržnih razmerah, kakor veljajo na dan ocenjevanja. S postavitvijo regresijskega modela za primerljive nepremičnine cenilec ovrednoti vpliv posameznih značilnosti na ceno nepremičnine. Prilagoditve na osnovi *nabavnih vrednosti* cenilec uporabi, če ni dovolj tržnih podatkov, s katerimi bi bilo mogoče oceniti vrednost posameznih značilnosti nepremičnine. S to metodo cenilec oceni vrednost neke značilnosti na osnovi

reprodukcijske ali nadomestitvene vrednosti za to značilnost (Nagode 1999, str. 24, Eckert 1990, str. 156-160).

Rezultat teh primerjav in popravkov je ocena intervala, v katerem naj bi se gibala tržna vrednost ocenjevana nepremičnine. Čim več podatkov ima ocenjevalec na voljo in čim bolj so si primerljive in ocenjevana nepremičnina podobne, tem bolj zanesljiva je ocena pričakovane tržne vrednosti proučevane nepremičnine.

Dohodkovna metoda

Za dohodkovno metodo se v literaturi uporabljata tudi izraza metoda donosov in prihodkovna metoda vrednotenja nepremičnin. Osnovne značilnosti te metode bomo povzeli po delih Brigham in Davesa (2002), Cirmanove in soavtorjev (2000, str. 22-26), Eckerta (1990, str. 231-299) in Nagodetove (1999, str. 14-22).

Dohodkovna metoda se navadno uporablja za ocenjevanje vrednosti nepremičnin, ki prinašajo stalni dohodek. To so najpogosteje nepremičnine, ki jih dajemo v najem. V skladu s to metodo predstavlja vrednost nepremičnine znesek, ki bi ga bil racionalen kupec pripravljen plačati na dan ovrednotenja za pravice do sprejema vseh prihodnjih čistih donosov od nepremičnine.

Osnovo za ocenjevanje vrednosti nepremičnin z dohodkovno metodo predstavlja neto sedanja vrednost prihodnjih denarnih tokov, ki jih ustvarimo z ocenjevano nepremičnino (Brigham in Daves 2002, str. 379):

$$NSV = \sum_{t=1}^T \frac{DT_t}{(1+r)^t} \quad (1),$$

kjer so:

NSV ... neto sedanja vrednost,

DT_t ... denarni tok v časovnem obdobju t ,

r ... kapitalizacijska stopnja,

T ... število časovnih obdobj, ki je merodajno za oceno neto sedanje vrednosti.

Enačbo (1) lahko transformiramo tako, da z njo ocenimo vrednost nepremičnine kot vsoto rednih donosov, ki jih ustvarimo z nepremičnino v T časovnih obdobjih, in preostale vrednosti nepremičnine na koncu časovnega obdobja T (glej tudi Brigham in Daves 2002, str. 107):

$$NSV = \sum_{t=1}^T \frac{Don_t}{(1+r)^t} + \frac{PV_T}{(1+r)^T} \quad (2),$$

kjer sta:

Don_t ... redni donos v časovnem obdobju t ,

PV_T ... preostala vrednost nepremičnine na koncu časovnega obdobja T .

Ob predpostavkah, da donosi naraščajo z enakomerno stopnjo rasti v vsakem časovnem obdobju, in da pridobivanje donosov traja zelo dolgo ($T \rightarrow \infty$), je neto sedanja vrednost nepremičnine enaka (Brigham in Daves 2002, str. 148):

$$NSV = Don_0 \cdot \sum_{t=1}^{\infty} \frac{(1+g)^t}{(1+r)^t} = \left[Don_0 \cdot \frac{1+g}{r-g} ; \text{če je } r > g \right] \quad (3),$$

kjer sta:

Don_0 ... donos na začetku opazovanega obdobja,

g ... konstantna stopnja rasti donosa.

Drugi del enačbe (3) velja le pri pogoju, da je $r > g$.

Iz enačb (1) do (3) sledi, da mora cenilec oceniti pričakovane denarne tokove oziroma donose, ki jih bo nepremičnina ustvarjala, ter kapitalizacijsko stopnjo, s katero prevede bodoče denarne tokove na sedanjo vrednost, če želi oceniti vrednost nepremičnine z metodo donosov.

Napoved *prihodnjih denarnih tokov* je zaradi negotove prihodnosti zahtevna naloga. Podatke in informacije za oceno donosov poišče cenilec navadno na trgu nepremičnin (podlaga za oceno je ekonomska najemnina, ki prevladuje na trgu za določeno vrsto objektov) ali pa si pomaga s pogodbeno najemnino, če so najemne pogodbe sklenjene za daljši čas. Za izračun denarnih tokov mora cenilec potencialni prihodek od pričakovanih najemnin korigirati s pričakovanim odstotkom zasedenosti nepremičnine in upoštevati še druge morebitne prilive, ki izhajajo iz lastništva nepremičnine (na primer oddajanje parkirnih prostorov, plačilo različnih storitev itn.). Od pričakovanih prihodkov mora cenilec odšteti tekoče poslovne odhodke: stroški zavarovanja nepremičnine, stroški upravljanja, davki na nepremičnine, nadomestitvene rezerve, stroški porabljene vode, elektrike, odvoza smeti itn. Na koncu mora cenilec upoštevati še denarni tok, ki naj bi ga lastnik nepremičnine prejel ob njeni prodaji.

Pri ugotavljanju sedanje vrednosti prihodnjih denarnih tokov, ki jih ustvarja nepremičnina, mora cenilec upoštevati takšno *kapitalizacijsko stopnjo*, ki posredno ali neposredno vključuje naslednje elemente (Cirman et al. 2000, str. 24):

- nadomestilo za časovno vrednost denarja, ki je navadno netvegana obrestna mera,
- premijo za tveganje,
- premijo za nelikvidnost,
- nadomestilo za upravljanje z naložbo.

Ocenjevanje vrednosti nepremičnine z metodo donosov je zahtevno, ker mora cenilec predvidevati razmere v prihodnosti. Ta pristop je zelo priljubljen v bankah. Cenilci pogosto uporabljajo njegovo poenostavljeno različico, to je **multiplikator bruto najemnine**, ki predstavlja razmerje med doseženimi prodajnimi cenami in bruto najemninami primerljivih nepremičnin. Vrednost ocenjevane nepremičnine je nato enaka zmnožku tega multiplikatorja in dosežene bruto najemnine za ocenjevano nepremičnino (Cirman et al. 2000, str. 25).

Stroškovna metoda

Za stroškovno metodo se v literaturi uporablja tudi izraz metoda nabavnih cen. Osnovne značilnosti te metode bomo povzeli po delih Cirmanove in soavtorjev (2000, str. 27-30), Eckerta (1990, str. 205-230) in Nagodetove (1999, str. 11-14).

Stroškovna metoda temelji na logiki nadomestitvene vrednosti. Dobro informiran racionalni kupec naj bi bil pripravljen za nepremičnino plačati največ toliko, kolikor bi znašali

minimalni stroški izgradnje nove nepremičnine z enako uporabno vrednostjo. Vrednost nepremičnine je zato enaka vrednosti nadomestitvenih stroškov.

Postopek ocenjevanja vrednosti po tej metodi poteka v štirih korakih:

- določitev vrednosti zemljišča,
- določitev stroškov nadomestitve sedanjega objekta,
- ocena izgube vrednosti obstoječega objekta,
- združitve vseh ugotovitev v oceno vrednosti.

Vrednost zemljišča cenilec oceni z uporabo tržne metode na osnovi podatkov o primerljivih prodajah. Pri tem cenilec ocenjuje vrednost za prazno in komunalno opremljeno zemljišče, ki je pripravljeno za gradnjo. Značilnost zemljišč je, da se ne starajo, zato zanje ni treba upoštevati tovrstnih popravkov vrednosti.

Stroške nadomestitve sedanjega objekta lahko cenilec določi z dvema pristopoma. Pri prvem pristopu cenilec ugotavlja višino reprodukcijskih stroškov. Reprodukcijski stroški so v trenutnih cenah izraženi stroški izgradnje identičnega objekta iz enakih materialov z uporabo enakih metod dela. Pristop z reprodukcijskimi stroški je mogoče uporabiti le za relativno nove stavbe, kjer so pri gradnji uporabljeni materiali še vedno na trgu. Z drugim pristopom cenilec ugotavlja višino nadomestitvenih stroškov, ki so enaki trenutnim stroškom postavitve objekta z enako velikostjo in uporabnostjo, vendar z uporabo sodobnih materialov, konstrukcijskih rešitev in metod dela. Ne glede na uporabljen pristop ugotavljanja stroškov se pri ocenjevanju nepremičnin nabavne vrednosti lahko določijo na štiri načine (Nagode 1999, str. 11-12):

- z natančnim popisom del in kalkulacijo, to je z množenjem enotnih cen in količin vgrajenega materiala in storitev;
- s seštevkom horizontalnih (nabavne vrednosti temeljev, strehe, elektrike itn.) in vertikalnih vrednosti (nabavne vrednosti sten, oken, vrat itn.);
- z množenjem enote mere objekta (na primer m²) in povprečnih nabavnih vrednosti enote mere;
- z revalorizacijo stroškov gradnje ocenjevanega objekta.

Način določanja nabavne vrednosti je odvisen predvsem od značilnosti ocenjevanega objekta in razpoložljivih podatkov.

Tako ugotovljena vrednost stroškov predstavlja vrednost novega objekta, zato morajo cenilci pri starejših objektih od nadomestitvene vrednosti odšteti *amortizacijo* oziroma *izgubo vrednosti* ocenjevanega objekta. Dejanska vrednost objekta je manjša od nadomestitvene vrednosti zaradi treh razlogov (Cirman et al. 2000, str. 28-29, Nagode 1999, str. 12-13):

- fizičnega zastarevanja, ki je posledica izrabe in poškodb;
- funkcionalnega zastarevanja, ki pomeni, da del objekta ni več primeren za uporabo (na primer neustrezno razmerje med površino objekta in zemljišča);
- ekonomskega zastarevanja, ki je povzročeno s strani dejavnikov v okolju (na primer zgraditev avtoceste v bližini, bližina hrupnih nočnih klubov itn.).

Zmanjšanje vrednosti objekta je mogoče meriti na štiri načine (Cirman et al. 2000, str. 29):

- s pomočjo ekonomske življenjske dobe objekta preko količnika med dejansko starostjo in predvideno ekonomsko življenjsko dobo;
- z izračunavanjem stroškov popravila;
- s tržnim načinom, kjer najprej analiziramo dosežene cene v prodajah podobnih nepremičnin in od njih odštejemo vrednosti zemljišča;
- s pomočjo kapitalizacije dobička, kjer skušamo ugotoviti razliko med sedaj doseženim donosom in donosom, ki bi ga dosegli, če ne bi bilo izgube vrednosti.

Vse ugotovitve združimo v oceno vrednosti tako, da vrednosti zemljišča prištejemo nadomestitveno vrednost novega objekta in od te vsote odštejemo ocenjeno izgubo vrednosti ocenjevanega objekta.

Stroškovno metodo cenilci najpogosteje uporabljajo, če imajo opravka z nerazvitim trgom nepremičnin, za ocenjevanje vrednosti javnih zgradb, za presojo različnih investicijskih možnosti v povezavi z izboljšavami in obnovitvami objektov itn. Glavna omejitev stroškovne metode je, da cenilci težko cenijo vrednosti starejših stavb in izračunajo izgubo vrednosti zaradi zastarelega stila, neustrezne kakovosti materialov in izdelave. Sam pristop tudi ne vključuje načela najboljše uporabe zemljišča (Cirman et al. 2000, str. 29).

Izbira metode za vrednotenje nepremičnin

V Sloveniji je bila od leta 1992 do 2003 za posamično vrednotenje nepremičnin v veljavi uradna administrativna metoda za ocenjevanje nepremičnin PEM. Zakon o nepremičninskem posredovanju zahteva za vrednotenje nepremičnin uporabo strokovnih standardov, ki so v skladu z Zakonom o revidiranju – glej poglavje 2.2. Vse tri zgoraj opisane metode vrednotenja so v skladu z Zakonom o revidiranju (Kožar in Marinšek 2003, str. 217).

Izbira metode vrednotenja posamezne nepremičnine je odvisna od vrste ocenjevané nepremičnine. Ob zadostni količini podatkov je tržna metoda oziroma metoda prodajnih cen večinoma najprimernejša za ocenjevanje tržne vrednosti nepremičnine (Nagode 1999, str. 23, Eckert 1990, str. 153). Seveda te metode ni mogoče uporabiti za ocenjevanje vrednosti nepremičnin, ki se redko znajdejo na trgu, na primer živalski vrtovi ali razne javne ustanove, kot so knjižnice, šole itn. Za ocenjevanje takšnih nepremičnin je najprimernejša stroškovna metoda (Cirman et al. 2000, str. 29), ki je po drugi strani popolnoma neprimerna za ocenjevanje zemljišč. Uporaba dohodkovne metode je vezana predvsem na ocenjevanje vrednosti nepremičnin, ki svojim lastnikom prinašajo donos v obliki denarnih tokov v različnih časovnih obdobjih. Čeprav lahko to metodo uporabimo za vse vrste nepremičnin, ki jih lastniki dajejo v najem (tudi na primer stanovanja ali hiše), je smiselnost uporabe te metode vprašljiva v primerih, ko je lastniki kupijo nepremičnino z namenom lastnega uživanja in ne z namenom dajanja le-te v najem (Eckert 1990, str. 108).

2.3.3 Množično vrednotenje nepremičnin

Pri posamičnem vrednotenju nepremičnin ocenimo vrednost ene nepremičnine, medtem ko pri množičnem vrednotenju z modelom vrednotenja ocenimo vrednosti več nepremičnin. Množično vrednotenje nepremičnin definiramo kot sistem podatkov, metod in dejavnosti, ki se vzpostavi z namenom ocenjevanja tržnih vrednosti določenih vrst nepremičnin na osnovi razpoložljivih vzorcev in modeliranja tržišča. Pri tej vrsti vrednotenja skušamo v model zajeti

dejavnike, ki pojasnjujejo vrednost posameznih nepremičnin, z modelom simulirati dejavnike trga in ovrednotiti prispevke posameznih dejavnikov (Cirman et al. 2000, str. 31-32). Pogoji za vzpostavitev množičnega vrednotenja nepremičnin je razpoložljivost ustreznih podatkov, ki se nanašajo na trg nepremičnin. Vzpostavitev množičnega vrednotenja nepremičnin je potrebna iz več razlogov (Cirman et al. 2000, str. 32, Rotovnik 1997, str. 34-35, Zakrajšek 1993):

- zaradi boljše informiranosti udeležencev na nepremičninskem trgu;
- množica zbranih podatkov lahko služi kot kakovostna podlaga za individualno vrednotenje nepremičnin, ki se pogosto pojavljajo na trgu (npr. stanovanja ali hiše);
- za pomoč cenilcem pri zagotavljanju točnosti podatkov, ocenjevanju različnih modelov vrednotenja in preverjanju dobljenih cen;
- za potrebe davčnega sistema;
- za spremljanje in analiziranje trga nepremičnin ter preprečevanje in zaviranje špekulacij;
- za lažje upravljanje nepremičnin v javni lasti;
- za pomoč pri sistemski opredelitvi posameznih sektorjev (npr. stanovanjsko gospodarstvo, urejanje prostora itn.).

Množično vrednotenje nepremičnin je sestavljeno iz dveh korakov: izdelave modela in kalibracije modela.

Izdelava modela

Splošni model za množično vrednotenje nepremičnin lahko zapišemo v naslednji obliki (Eckert 1990, str. 319, Nagode 1999, str. 27):

$$MV = \prod GQ \cdot [\prod BQ \cdot \sum BA + \prod LQ \cdot \sum LA + \sum OA] \quad (4),$$

kjer so:

MV	... tržna vrednost nepremičnine,
$\prod GQ$... produkt splošnih faktorjev kakovosti,
$\prod BQ$... produkt faktorjev kakovosti objekta,
$\prod LQ$... produkt faktorjev kakovosti zemljišča,
$\sum BA$... vsota cen posameznih elementov objekta v odvisnosti od velikosti,
$\sum LA$... vsota cen posameznih elementov zemljišča v odvisnosti od velikosti,
$\sum OA$... vsota cen drugih elementov v odvisnosti od velikosti.

Splošni faktorji kakovosti GQ , ki zadevajo celotno nepremičnino, so na primer lokacija, čas itn. Faktorji kakovosti objekta BQ so na primer stanje objekta, njegova starost itn. Faktorji kakovosti zemljišča LQ so na primer oblika zemljišča, dostop, razgled itn. Načeloma se faktorji, ki se nanašajo na kakovost, med seboj množijo, faktorji, ki se nanašajo na velikost, pa se seštevajo. Podatki o posamezni nepremičnini so lahko ovrednoteni kvantitativno ali kvalitativno. Kvantitativni podatki temeljijo na meritvah ali izmerah, kvalitativni podatki temeljijo na opisu lastnosti nepremičnine.

Splošni model iz enačbe (4) je mogoče poenostaviti v čisti aditivni ali čisti multiplikativni model, ki pogosto zadostujeta za namen množičnega vrednotenja (Zakrajšek 1996, str. 74, Zakrajšek 1993, str. 160, Eckert 1990, str. 339). Aditivni model za množično vrednotenje nepremičnin ima naslednjo obliko (Eckert 1990, str. 339):

$$MV = \beta_0 + \sum_{i=1}^p \beta_i \cdot X_i \quad (5),$$

kjer so:

- β_0 ... konstantni člen,
- X_i ... neodvisne spremenljivke v modelu,
- β_i ... koeficienti, ki ustrezajo posameznim neodvisnim spremenljivkam,
- p ... število neodvisnih spremenljivk.

Neodvisne spremenljivke X_i lahko predstavljajo kvantitativne in kvalitativne faktorje, ki vplivajo na vrednost nepremičnine. Multiplikativni model za množično vrednotenje nepremičnin ima naslednjo obliko (Eckert 1990, str. 339):

$$MV = \beta_0 \cdot \prod \beta_i^{X_i} \cdot \prod X_j^{\beta_j} \quad (6),$$

kjer so:

- β_0 ... konstantni člen,
- X_i ... binarne kvalitativne neodvisne spremenljivke z zalogo vrednosti 0 in 1,
- X_j ... kvantitativne neodvisne spremenljivke v modelu,
- β_i ... koeficienti, ki ustrezajo kvalitativnim neodvisnim spremenljivkam,
- β_j ... koeficienti, ki ustrezajo kvantitativnim neodvisnim spremenljivkam.

V primeru poenostavljenih aditivnih ali multiplikativnih modelov ne ločimo spremenljivk, ki so vezane na zemljišče in stavbo. Prednost poenostavljenih modelov (5) in (6) je, da je mogoče z ustrežno transformacijo osnovne oblike modelov neznanne koeficiente v vseh treh modelih oceniti z metodo multiple regresije (Eckert 1990, str. 319-350).

Kalibracija modela

Izbira oblike modela za množično vrednotenje nepremičnin je odvisna tudi od uporabljene kalibracijske metode. Kalibracija modela je postopek, s katerim cenilec priredi izbrani model za množično vrednotenje nepremičnin podatkom, ki jih ima na voljo. S postopkom kalibracije cenilec oceni koeficiente, ki ustrezajo posameznim neodvisnim spremenljivkam v modelu. Kalibracijo modela lahko cenilec izvaja z vsemi tremi metodami, ki smo jih že opisali pri posamičnem vrednotenju nepremičnin: s tržno metodo, z dohodkovno metodo ali s stroškovno metodo (Eckert 1990, str. 311-312, Nagode 1999, str. 28-29).

Pri *množičnem vrednotenju nepremičnin s tržno metodo* lahko vrednosti nepremičnin navadno določimo na podlagi (Nagode 1999, str. 28-29):

- zemljepisno zaključenih območij, ki so opredeljena z naravnimi ovirami: takšna razdelitev je primerna, če so značilnosti objektov na posameznem območju podobne, cene nepremičnin pa se močno spreminjajo od območja do območja;

- delitve nepremičnin glede na njihove značilnosti: ta način uporabljamo pri težko določljivi podobnosti nepremičnin v soseski in brez izrazite lokacije;
- enotnega modela, ko vrednotimo nepremičnine za celotno območje.

Za množično vrednotenje nepremičnin po tržni metodi pogosto uporabljamo poenostavljena modela, katerih koeficienti se ocenijo z metodo multiple regresije (Eckert 1990, str. 338-346).

Pri *množičnem vrednotenju z dohodkovno metodo* uporabljamo za oceno prihodka ali dobička modele z različnimi spremenljivkami. Tako lahko z modeli ocenjujemo bruto dobiček, neto dobiček ali multiplikator bruto dobička. Stopnjo kapitalizacije navadno izračunamo s pomočjo statističnih podatkov (Eckert 1990, str. 346-349, Nagode 1999, str. 28).

Pri *množičnem vrednotenju s stroškovno metodo* poteka kalibracija modela v naslednjih korakih (Nagode 1999, str. 28):

- razdelitev posameznih območij glede na vrsto nepremičnin;
- določitev merodajnih spremenljivk nepremičnine, kot so na primer število etaž objekta, vrsta konstrukcije itn.;
- določitev parametrov za vsako spremenljivko nepremičnine (število etaž ima lahko naslednje parametre: klet, pritličje, prvo do tretje nadstropje, itn.);
- določitev števila modelov s kombiniranjem možnih parametrov vseh spremenljivk;
- ocena nadomestitvene vrednosti za vsak model: vrednosti se ocenijo kot vsota vertikalnih, horizontalnih in konstantnih stroškov, ki so korigirani v skladu s kvalitativnimi stroškovnimi komponentami;
- ocena preostale stroškovne vrednosti za vsak model v odvisnosti od amortizacijskih faktorjev glede na dejansko starost objekta.

V tem poglavju smo predstavili osnovne modele za množično vrednotenje nepremičnin. V magistrskem delu smo prikazali tudi nekaj alternativnih modelov za množično vrednotenje nepremičnin, ki temeljijo na metodah rudarjenja podatkov. Uporabljene alternativne modele smo bolj podrobno predstavili v poglavju 3. Rezultate vrednotenij z alternativnimi modeli smo primerjali z rezultati vrednotenij, ki smo jih dobili z osnovnim modelom. Za kalibracijo modelov smo uporabili tržni pristop. Podatke za analizo smo črpali iz delne baze nepremičninskih transakcij Davčne uprave Republike Slovenije (Baza podatkov DURS 2003).

3 Uporabljene metode za pripravo podatkov in tržno vrednotenje nepremičnin

3.1 Uvod

Cilj magistrskega dela je bil analizirati obstoječo bazo podatkov DURS-a o transakcijah z lastninskimi pravicami na nepremičninah, ki so bile izvedene na področju Slovenije v letih od 1999 do 2002, ter zgraditi model, ki bi omogočal množično vrednotenje nepremičnin po tržni metodi. Baza nepremičninskih transakcij DURS-a je sorazmerno obsežna, saj vsebuje podatke o preko sto tisoč nepremičninskih transakcijah. Zaradi obsežnosti baze smo se odločili, da bomo naš cilj skušali doseči z uporabo metod rudarjenja podatkov. S pojmom rudarjenja

podatkov označujemo nabor najrazličnejših metod, ki služijo odkrivanju zanimivih modelov ali značilnosti, ki se skrivajo v bazah podatkov. Pri tem z "zanimivim modelom ali značilnostjo" označujemo vsako netrivialno, implicitno, do sedaj neznano in seveda uporabno informacijo, ki se skriva v podatkih iz proučevane baze (Lavrač in Grobelnik 2003, str. 5). Rudarjenje podatkov je višji nivo od t.i. sprotnega analitičnega procesiranja podatkov⁶, s katerim navadno opisujemo le osnovne statistične analize in ki daje uporabniku le omejene informacije o opazovanem pojavu. Zaradi tega so lahko metode rudarjenja podatkov močno orodje v ekonomskih vedah in pri podpori odločanju (Lavrač in Bohanec 2003). Kljub zelo raznovrstnim oblikam metod rudarjenja podatkov jih lahko razdelimo na dve področji:

- metode za opisno indukcijo (povezovalna pravila, združevanje v skupine itn.) in
- metode za napovedovanje in klasifikacijo (odločitvena drevesa, modelna drevesa, nevronske mreže itn.).

Pri analizi baze nepremičninskih transakcij smo poleg osnovnih statističnih analiz uporabili metode rudarjenja podatkov za iskanje povezovalnih pravil. Pri izgradnji modelov za vrednotenje nepremičnin smo uporabili metode za numerično napovedovanje.

Preden se lotimo opisa metod, ki smo jih uporabili v magistrskem delu, na kratko opredelimo nekaj osnovnih pojmov, ki jih uporabljamo v nadaljevanju magistrskega dela.

Poljubni pojav, ki se izraža z več lastnostmi, lahko opišemo z vektorjem \mathbf{X} :

$$\mathbf{X} = (X_1, X_2, \dots, X_p) \quad (7),$$

kjer so:

- \mathbf{X} ... skupek podatkov o proučevanem pojavu,
- X_i ... spremenljivke vektorja \mathbf{X} ,
- p ... dimenzija vektorja \mathbf{X} .

Vektor \mathbf{X} v našem primeru predstavlja skupek podatkov, ki so na voljo za opravljeno transakcijo z nepremičninami. Spremenljivke X_i predstavljajo različne lastnosti nepremičninske transakcije, kot so na primer cena, velikost, lega nepremičnine itn. V splošnem ločimo štiri vrste spremenljivk X_i (Sharma 1996, str. 2-3):

- *nominalne spremenljivke*, ki so predstavljene s kvalitativnimi vrednostmi, ki jih ni mogoče urediti po velikosti;
- *ordinalne spremenljivke*, ki so predstavljene s kvalitativnimi vrednostmi, ki jih je mogoče urediti po velikosti;
- *intervalne spremenljivke*, ki že imajo za osnovo numerično skalo;
- *racionalne spremenljivke*, s katerimi je mogoče tvoriti ulomke.

Intervalne in racionalne spremenljivke označujemo tudi s pojmom numerične spremenljivke, nominalne in ordinalne spremenljivke pa s pojmom kvalitativne spremenljivke.

⁶ Ang. *On-Line Analytical Processing* ali skrajšano OLAP.

Realizacijo vektorja \mathbf{X} imenujemo vzorec⁷ ali objekt in ga označimo z \mathbf{x}_i . Množico n vzorcev \mathbf{x}_i zapišemo z matriko podatkov:

$$\mathbf{X} \equiv (X_1, X_2, \dots, X_p)$$

$$\begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \\ \vdots \\ \mathbf{x}_n \end{bmatrix} = \begin{bmatrix} x_{11}, x_{12}, \dots, x_{1p} \\ x_{21}, x_{22}, \dots, x_{2p} \\ \vdots & \vdots & \vdots \\ x_{n1}, x_{n2}, \dots, x_{np} \end{bmatrix} \quad (8).$$

3.2 Združevanje v skupine

Metoda združevanja v skupine je tehnika, ki jo uporabljamo za združevanje vzorcev v skupine. Skupine so oblikovane tako, da so si vzorci v posamezni skupini čim bolj podobni glede na lastnosti, ki smo jih upoštevali za združevanje v skupine, pri tem pa so vzorci v vsaki skupini čim bolj različni od vzorcev v drugi skupini. Opredelitev podobnosti med vzorci se razlikuje od primera do primera. Metodo združevanja v skupine smo v magistrskem delu uporabili zato, da smo slovenske občine združili v skupine glede na realizirano povprečno ceno kvadratnega metra stanovanja z namenom oceniti območja podobnih cenovnih razredov.

Metode združevanja v skupine delimo na hierarhične in nehierarhične metode. V magistrskem delu smo uporabili samo hierarhične metode združevanja v skupine, zato je na kratko prikazano le teoretično ozadje teh metod. Povzeli smo ga po Sharmi (1996, str. 185-232). Opise metod hierarhičnega združevanja je mogoče najti tudi drugje (Han in Kamber 2001, str. 354-362, Hand et al. 2001, 308-314, Hastie et al. 2001, str. 472-479).

Za združevanje objektov v skupine potrebujemo mero podobnosti med objekti. Ker smo v našem primeru pri razvrščanju v skupine imeli opravka z numeričnimi spremenljivkami (cena), izberemo za mero podobnosti Evklidsko razdaljo. Če sta \mathbf{x}_i in \mathbf{x}_j dva opazovana vzorca, tedaj je evklidska razdalja med njima enaka:

$$d_{ij} = d(\mathbf{x}_i, \mathbf{x}_j) = \sqrt{\sum_{k=1}^p (x_{ik} - x_{jk})^2} \quad (9).$$

Razdalje med vsemi možnimi pari vzorcev \mathbf{x}_i in \mathbf{x}_j tvorijo matriko podobnosti $\mathbf{D} = \{d_{ij}\}$.

Za postopek hierarhičnega združevanja vzorcev v skupine je značilno, da število skupin ni znano vnaprej, ampak se spreminja. Začetno število skupin je enako številu vseh vzorcev, končno število skupin je enako ena. Hierarhično združevanje v skupine poteka takole:

1. Izhodišče združevanja je n skupin-vzorcev s pripadajočo matriko podobnosti \mathbf{D} , ki vsebuje mere podobnosti med pari obstoječih skupin-vzorcev.
2. Na osnovi elementov matrike podobnosti poiščemo najbližji par skupin.
3. Najbližji par skupin združimo v novo skupino in obnovimo matriko podobnosti \mathbf{D} . Razdalje med skupinami lahko izračunamo na načine, ki so razloženi v nadaljevanju.

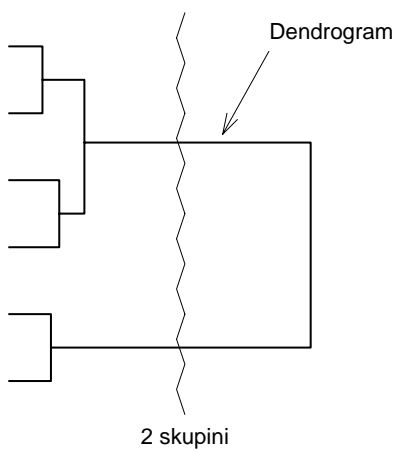
⁷ Pojem "vzorec" lahko uporabimo bodisi za opis lastnosti enega zapisa v bazi podatkov bodisi za opis ponavljajoče značilnosti nekega pojava. V našem primeru bomo pojem "vzorec" uporabljali v prvem pomenu.

4. Točki 2. in 3. ponavljamo, dokler ne dobimo ene same skupine z vsemi izhodiščnimi vzorci.

Metode hierarhičnega združevanja v skupine se razlikujejo glede na način izračuna razdalj oziroma podobnosti med obstoječimi skupinami (Sharma 1996, str. 188-194):

- *Metoda centroidov*: razdalja med dvema skupinama je enaka razdalji med centroidoma teh dveh skupin.
- *Metoda enojne povezanosti*: razdalja med dvema skupinama je enaka minimalni možni razdalji med pari vzorcev v dveh skupinah.
- *Metoda polne povezanosti*: razdalja med dvema skupinama je enaka maksimalni možni razdalji med pari vzorcev v dveh skupinah.
- *Metoda povprečne povezanosti*: razdalja med dvema skupinama je definirana s povprečno vrednostjo razdalj med vsemi možnimi pari vzorcev v dveh skupinah.
- *Wardova metoda*: pri tej metodi ne računamo razdalj med skupinami, ampak skupine formiramo tako, da maksimiramo homogenost znotraj skupin. Za mero homogenosti vzamemo vsoto kvadratov znotraj skupin.

Rezultate hierarhičnega združevanja v skupine prikažemo z drevesom združevanja ali dendrogramom. Višina točke združevanja v dendrogramu ustreza meri podobnosti med združenima izvornima skupinama. Število skupin ocenimo tako, da dendrogram prerežemo na mestu, kjer postanejo razdalje med skupinami, ki so se združevale, velike - glej Sl. 1.



Sl. 1: Primer ocene števila skupin iz dendrograma

3.3 Povezovalna pravila

Za začetek predpostavimo, da nek pojav opišemo z vektorjem $\mathbf{X} = (X_1, X_2, \dots, X_p)$. Ker povezovalna pravila spadajo med metode za opisno indukcijo, spremenljivk ne delimo na odvisne in neodvisne. Naj n realizacij vektorja \mathbf{X} sestavlja bazo podatkov, ki jo predstavimo z matriko podatkov iz enačbe (8). S povezovalnim pravilom opišemo značilno relacijo med dvema ali več spremenljivkami X_i vektorja \mathbf{X} . Primer povezovalnega pravila je lahko takšen:

$$\text{če je } X_2 = 3 \text{ in če je } X_5 = 2 \text{ tedaj sledi, da je } X_{p-1} = \text{zelena in } X_p = 5 \quad (10).$$

Zaradi značilne oblike povezovalnih pravil jih navadno tvorimo le za kvalitativne spremenljivke (Witten in Frank 2000, str. 294). Tipični primer uporabe povezovalnih pravil

na področju poslovnih in ekonomskih ved je na primer analiza košarice dobrin, ki jo potrošniki sestavijo v trgovinah. S to analizo je mogoče ugotoviti, katere izdelke potrošniki navadno kupujejo skupaj, kar lahko pomaga pri ureditvi prodajnih prostorov (Witten in Frank 2000, str. 25-26). Za uporabnika so seveda zanimiva le smiselna pravila, ki veljajo sorazmerno zanesljivo za dovolj veliko število vzorcev v bazi. Medtem ko mora smiselnost pravila oceniti vsak uporabnik sam, pa lahko za vsako povezovalno pravilo definiramo njegovo *zaupanje*⁸ in *podporo*⁹. Podpora povezovalnemu pravilu je definirana kot delež vseh vzorcev v bazi, za katere povezovalno pravilo drži. Zaupanje povezovalnemu pravilu predstavlja delež vzorcev, za katere je ob izpolnjenem pogoju povezava pravilna glede na vse vzorce, za katere je ta isti pogoj izpolnjen. Uporabno vrednost imajo navadno le povezovalna pravila z dovolj veliko podporo in zaupanjem (Witten in Frank 2000, str. 63-64).

Iskanje povezovalnih pravil v bazi podatkov je tipičen primer rudarjenja podatkov. Problem pri iskanju povezovalnih pravil je, da že za zelo omejeno bazo podatkov s samo nekaj spremenljivkami lahko tvorimo zelo veliko povezovalnih pravil. Tako je že za srednje veliko bazo praktično nemogoče poiskati vsa povezovalna pravila s kombiniranjem vseh možnih vrednosti spremenljivk med seboj. Zato je bil razvit *APRIORI algoritem*, ki omogoča učinkovito iskanje povezovalnih pravil v velikih bazah podatkov, ki so sestavljene le iz kvalitativnih spremenljivk. Podrobnosti APRIORI algoritma ne bomo navajali, ker jih je mogoče najti v literaturi (Han in Kamber 2001, str. 230-235, Hand et al. 2001, 157-160, Hastie et al. 2001, str. 441-447, Witten in Frank 2000, str. 108-110). Na tem mestu je naveden le osnovni princip delovanja tega algoritma.

Za uporabo APRIORI algoritma je najprej potrebno definirati minimalno podporo in zaupanje za iskanje povezovalnih pravil. APRIORI algoritem poteka v dveh korakih in izkorišča naslednjo lastnost povezovalnih pravil: če želimo, da ima množica realizacij k različnih spremenljivk zadovoljivo podporo, tedaj morajo imeti zadovoljivo podporo vse podmnožice realizacij $k-1$ spremenljivk, ki jih lahko tvorimo iz množice realizacij k spremenljivk. V prvem koraku algoritma tako naprej tvorimo vse množice realizacij dveh spremenljivk, ki imajo dovolj veliko podporo. Z ustreznimi kombinacijami tako dobljenih množic realizacij dveh spremenljivk tvorimo vse možne kombinacije množic realizacij treh spremenljivk in obdržimo le tiste, ki imajo dovolj veliko podporo. Postopek ponavljamo, dokler dobivamo množice realizacij za vse večje število spremenljivk, ki še imajo dovolj veliko podporo. Ko odkrijemo vse možne množice realizacij spremenljivk z dovolj veliko podporo, sledi drugi korak algoritma: za dane množice realizacij spremenljivk tvorimo vsa možna povezovalna pravila in preverimo njihovo zaupanje. Obdržimo le tista povezovalna pravila, ki imajo dovolj veliko zaupanje (Witten in Frank 2000, str. 108-110).

3.4 Metoda linearne regresije

Metoda linearne regresije je na področju ekonometrije ena izmed najbolj razširjenih metod za modeliranje odvisnosti med eno ali več neodvisnimi spremenljivkami ter eno odvisno spremenljivko (Pindyck in Rubinfeld 1998). V primeru, ko v modelu nastopa več neodvisnih

⁸ Ang. *coverage* ali *support*.

⁹ Ang. *accuracy* ali *confidence*.

spremenljivk, govorimo o multipli regresiji. Metoda multiple regresije omogoča poleg modeliranja odvisnosti med neodvisnimi in odvisno spremenljivko tudi sklepanje o pomembnosti posameznih neodvisnih spremenljivk v modelu na osnovi statistične značilnosti regresijskih koeficientov. Kljub temu, da je metoda multiple regresije klasična statistična tehnika, je pogosto omenjena v literaturi s področja rudarjenja podatkov (Han in Kamber 2001, str. 319-322, Hand et al. 2001, 368-384, Hastie et al. 2001, str. 41-75, Witten in Frank 2000, str. 112-114). V našem primeru je model za množično vrednotenje nepremičnin po tržni metodi, ki smo ga dobili z metodo linearne regresije, referenčni model, s katerim smo primerjali modele, dobljene z alternativnimi metodami rudarjenja podatkov. V nadaljevanju je na kratko opisan model multiple regresije, ki je povzet po delu Pindycka in Rubinfelda (1998, str. 85-113). Podrobnejši opis je mogoče najti v literaturi.

Naj Y predstavlja odvisno spremenljivko in naj $\mathbf{X} = (X_1, X_2, \dots, X_p)$ predstavlja vektor p neodvisnih spremenljivk. Multipli regresijski model ima tedaj naslednjo obliko:

$$Y = \beta_0 + \sum_{i=1}^p \beta_i \cdot X_i \quad (11),$$

kjer so:

β_0 ... konstanti člen,

β_i ... regresijski koeficienti.

Za uporabo modela multiple regresije morajo biti izpolnjeni naslednji pogoji (Pindyck in Rubinfeld 1998, str. 86):

- odvisnost med spremenljivko Y in spremenljivkami X_i mora biti linearna,
- spremenljivke X_i naj bi bile med seboj linearno neodvisne,
- napake napovedi $\varepsilon_j = \hat{y}_j - y_j$ naj bi bile med seboj neodvisne in normalno porazdeljene s srednjo vrednostjo enako nič in konstantno varianco; \hat{y}_j je napoved odvisne spremenljivke za j -ti vzorec v množici $\{(\mathbf{x}_j, y_j) ; j = 1, \dots, n\}$.

Konstantni člen in regresijski koeficienti so prosti parametri modela multiple regresije, ki jih ocenimo z metodo najmanjših kvadratov na osnovi parov realiziranih vrednosti neodvisnih in odvisnih spremenljivk $\{(\mathbf{x}_j, y_j) ; j = 1, \dots, n\}$ (Pindyck in Rubinfeld 1998, str. 107-111).

Statistično značilnost ocenjenih regresijskih koeficientov preverimo s t in F testom (Pindyck in Rubinfeld 1998, str. 87-95). Če odvisno in vse neodvisne spremenljivke standardiziramo preden zgradimo model multiple regresije, tedaj regresijski koeficienti standardiziranega modela predstavljajo relativno pomembnost ustrezne neodvisne spremenljivke. Te regresijske koeficiente imenujemo standardizirani regresijski koeficienti.

Če so neodvisne spremenljivke kvalitativne, jih v modelu multiple regresije upoštevamo s pomočjo slamnatih spremenljivk¹⁰. Slamnate spremenljivke so binarne spremenljivke, ki imajo le dve različni vrednosti: nič in ena. Za opis kvalitativne spremenljivke s k vrednostmi moramo v regresijski model vključiti $k-1$ binarnih slamnatih spremenljivk.

¹⁰ Ang. *dummy variables*.

Z uporabo različnih transformacij osnovnih spremenljivk lahko multiplo regresijo uporabimo tudi za modeliranje nelinearnih odvisnosti med neodvisnimi in odvisno spremenljivko.

3.5 Lokalne metode za napoved

3.5.1 Uvod

Lokalne metode za napoved so posebna skupina metod rudarjenja podatkov, pri katerih model za napoved sestavljajo vsi vzorci v bazi podatkov. Naj Y spet predstavlja odvisno spremenljivko in naj $\mathbf{X} = (X_1, X_2, \dots, X_p)$ predstavlja vektor p neodvisnih spremenljivk. Model za napoved je v tem primeru sestavljen iz vseh n parov realiziranih vrednosti neodvisnih in odvisnih spremenljivk $\{(\mathbf{x}_j, y_j) ; j = 1, \dots, n\}$ v bazi podatkov.

Lokalne metode za napoved se od ostalih metod rudarjenja podatkov ločijo po tem, da za izgradnjo modela ne porabimo nič časa, ker model predstavljajo že sami podatki v bazi. Vse delo, ki je potrebno za napoved vrednosti odvisne spremenljivke Y , opravimo šele v trenutku, ko izvajamo napoved. Napoved vrednosti odvisne spremenljivke Y za dano realizacijo \mathbf{x} neodvisnega vektorja \mathbf{X} naredimo tako, da v bazi podatkov $\{(\mathbf{x}_j, y_j) ; j = 1, \dots, n\}$ najprej poiščemo nekaj vzorcev, ki so najbolj podobni našemu vektorju \mathbf{x} . Ko poznamo množico parov $\{(\mathbf{x}_k, y_k)\}$, ki so najbolj podobni vektorju \mathbf{x} , s primerno metodo napovemo vrednost spremenljivke Y , ki bi najbolj ustrezala vektorju \mathbf{x} glede na podobne vzorce v bazi podatkov. Za uporabo lokalnih metod napovedi moramo definirati mero podobnosti med vzorci in način za izračun napovedi spremenljivke Y . Odločiti se moramo tudi, ali bomo število K najbolj podobnih vzorcev, ki jih bomo upoštevali za izračun napovedane vrednosti spremenljivke Y , določili vnaprej ali bomo število najbolj podobnih vzorcev določili avtomatsko.

Prednost lokalnih metod napovedi je, da podatki sami po sebi tvorijo model za napoved. Njihova slabost je, da je potrebno poiskati najbolj podobne vzorce in izgraditi nov lokalni model za napoved za vsak nov vektor \mathbf{x} , kar je lahko zelo zamudno pri večjih bazah podatkov (Witten in Frank 2000, str. 72, 114).

Za izgradnjo modela za tržno vrednotenje nepremičnin smo uporabili dve lokalni metodi za napoved: metodo najbližjih sosedov in regresijski model z lokalnimi utežmi¹¹. Oba modela sta na kratko predstavljena v naslednjih dveh podpoglavjih.

3.5.2 Metoda najbližjih sosedov

Metoda najbližjih sosedov je najenostavnejša lokalna metoda za napoved. Njen opis smo povzeli po delu Wittna in Franka (2000, str. 114-116, 193-201). Opis te metode je mogoče najti tudi v drugi literaturi s področja rudarjenja podatkov (Han in Kamber 2001, str. 314-315, Hand et al. 2001, 347-352, Hastie et al. 2001, str. 415-426).

Metoda najbližjih sosedov deluje tako, da najprej izberemo število najbolj podobnih vzorcev K za lokalno napoved. Nato poiščemo v bazi najbolj podobne vzorce glede na vektor \mathbf{x} ter izračunamo napovedano vrednost odvisne spremenljivke \hat{y} za dani vektor \mathbf{x} kot povprečje vrednosti odvisnih spremenljivk za K najbolj podobnih vzorcev v bazi:

¹¹ Ang. *locally weighted regression*.

$$\hat{y}(\mathbf{x}) = \frac{1}{K} \cdot \sum_{k=1}^K y_k(\mathbf{x}_k) \quad (12).$$

Pri tem vrednosti y_k ustrezajo tistim vektorjem \mathbf{x}_k iz obstoječe baze podatkov, ki so najbolj podobni vektorju \mathbf{x} glede na izbrano mero podobnosti. Za mero podobnosti pogosto vzamemo kar evklidsko razdaljo med vektorjema \mathbf{x} in \mathbf{x}_k (Witten in Frank 2000, str. 114):

$$d(\mathbf{x}, \mathbf{x}_k) = \|\mathbf{x} - \mathbf{x}_k\| = \sqrt{(x_1 - x_{k;1})^2 + \dots + (x_p - x_{k;p})^2} \quad (13).$$

Z x_i smo v enačbi (13) označili posamezne komponente vektorjev \mathbf{x} in \mathbf{x}_k .

V enačbi (12) vidimo, da imajo vsi vzorci, ki sodelujejo pri napovedi vrednosti odvisne spremenljivke Y , enak vpliv na napoved ne glede na oddaljenost vektorja \mathbf{x}_k od vektorja \mathbf{x} . To pomeni, da imajo manj podobni vzorci enak vpliv kakor najbolj podobni vzorci, kar lahko slabo vpliva na napoved spremenljivke Y . Ta pojav omilimo tako, da pri izračunu povprečja vrednosti odvisnih spremenljivk y_k za K najbolj podobnih vzorcev v bazi upoštevamo tudi oddaljenost teh vzorcev od vektorja \mathbf{x} . To naredimo z uteženim povprečjem:

$$\hat{y}(\mathbf{x}) = \sum_{k=1}^K w_k \cdot y_k(\mathbf{x}_k) \quad (14),$$

kjer je vrednost posameznih uteži w_k odvisna od oddaljenosti vektorja \mathbf{x}_k od vektorja \mathbf{x} . Navadno vrednost uteži pada obratno sorazmerno z razdaljo $d(\mathbf{x}, \mathbf{x}_k)$. Pri tem pazimo, da je vsota vseh K uteži $\sum_{k=1}^K w_k$ enaka 1.

3.5.3 Regresijski model z lokalnimi utežmi

Regresijski model z lokalnimi utežmi je lokalna metoda za napoved, ki je naprednejša od metode najbližjih sosedov. Njen opis smo povzeli po delu Wittna in Franka (2000, str. 208-209). Podrobnejši opis te metode je mogoče najti v drugi literaturi s področja rudarjenja podatkov (Hand et al. 2001, 175-176, Hastie et al. 2001, str. 168-171).

Pri regresijskem modelu z lokalnimi utežmi za vsak par odvisne spremenljivke Y in neodvisnega vektorja \mathbf{X} iz množice $\{(\mathbf{x}_j, y_j) ; j = 1, \dots, n\}$ definiramo ustrezno utež w_n , ki je odvisna od podobnosti vektorja \mathbf{x}_j in vektorja \mathbf{x} , za katerega želimo napovedati vrednost spremenljivke Y . Za mero podobnosti med vektorjema \mathbf{x}_j in \mathbf{x} lahko spet vzamemo evklidsko razdaljo – glej enačbo (13). Uteži so lahko obratno sorazmerne razdalji med vektorjema \mathbf{x}_j in \mathbf{x} ali jih določimo z metodo jeder oziroma jedrnih funkcij¹². Pristop za določitev uteži, ki smo ga ubrali tudi v našem primeru, je podoben kakor pri metodi najbližjih sosedov: število K najbližjih vzorcev, ki jih upoštevamo pri izgradnji regresijskega modela, določimo vnaprej. Ti najbližji vzorci imajo pri izgradnji regresijskega modela enake ali različne uteži, uteži vseh preostalih vzorcev v bazi so enake nič.

Ko smo določili uteži vsem vzorcem v bazi podatkov, zgradimo model multiple regresije, pri katerem poleg parov odvisnih in neodvisnih spremenljivk $\{(\mathbf{x}_j, y_j) ; j = 1, \dots, n\}$ upoštevamo tudi pripadajoče uteži w_n . S tem dosežemo, da se regresijski model bolj prilagodi tistim vzorcem iz baze podatkov, ki so najbližji vektorju \mathbf{x} . Za določitev konstantnega člena in

¹² Za podrobnejšo razlago pojma jeder in jedrnih funkcij glej literaturo (Hastie et al. 2001, str. 165-190).

regresijskih koeficientov, ki ustrezajo vektorju \mathbf{x} , uporabimo metodo uteženih najmanjših kvadratov (Hastie et al. 2001, str. 169).

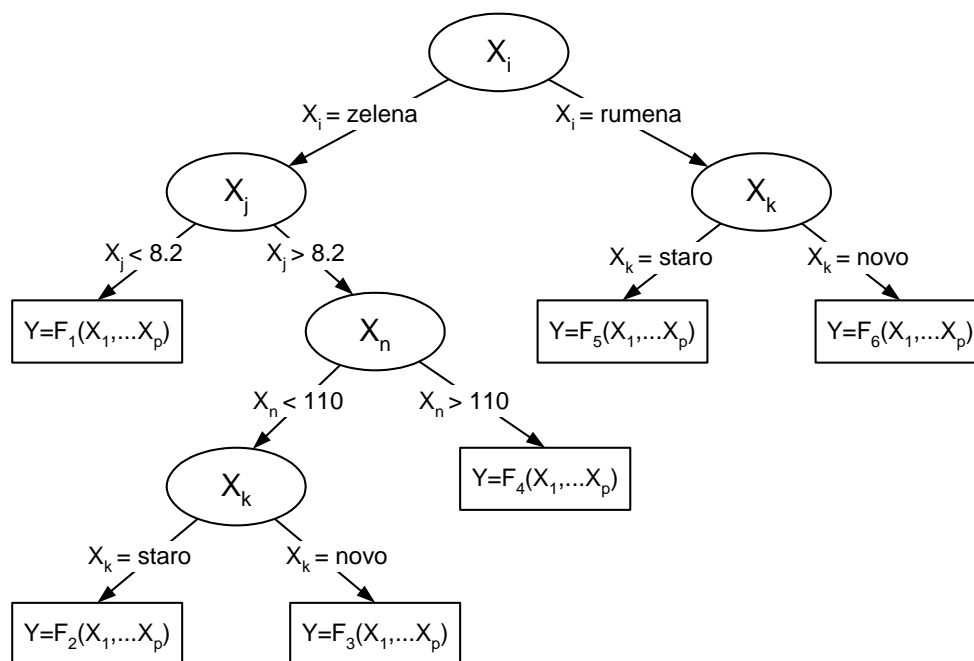
Regresijski model z lokalnimi utežmi je v splošnem boljši od metode najbližjih sosedov, vendar je napoved z njim tudi bistveno počasnejša, saj je potrebno poleg iskanja najbližjih vzorcev v bazi za vsak vektor \mathbf{x} zgraditi še regresijski model.

3.6 Drevesne metode

3.6.1 Uvod

Drevesne metode spadajo med tipične metode rudarjenja podatkov, ki jih uporabljamo za napoved vrednosti odvisne spremenljivke Y na osnovi realiziranih vrednosti neodvisnih spremenljivk $\mathbf{X} = (X_1, X_2, \dots, X_p)$. Ime te zvrsti metod rudarjenja podatkov izvira iz tega, da lahko matematični model za napoved predstavimo z drevesno strukturo. Opis drevesnih metod smo povzeli po Wittnu in Franku (2000, str. 70-72, 202-208). Bolj ali manj podrobne opise teh metod je mogoče najti tudi drugje (Han in Kamber 2001, str. 284-295, Hand et al. 2001, 343-347, Hastie et al. 2001, str. 266-278).

Primer modela z drevesno strukturo je prikazan na Sl. 2.



Sl. 2: Model za napoved, ki ima drevesno strukturo

Drevesni model je sestavljen iz vozlišč in listov. Vozlišča cepijo drevesni model v veje, vsaka veja pa se konča z listom. Vsako vozlišče ustreza eni neodvisni spremenljivki X_i in predstavlja odločitev za nadaljevanje poti po drevesu navzdol. Če je neodvisna spremenljivka v vozlišču kvalitativna, ima pripadajoče odločitveno pravilo naslednjo obliko:

$$\begin{aligned} &\text{če je } (X_i = \text{zelena}) \text{ nadaljuj v smeri prve veje} \\ &\text{če je } (X_i = \text{rumena}) \text{ nadaljuj v smeri druge veje} \end{aligned} \quad (15).$$

Če je neodvisna spremenljivka v vozlišču numerična, ima pripadajoče odločitveno pravilo navadno naslednjo obliko:

$$\begin{aligned}
&\text{če je } (X_j \leq 8.2) \text{ nadaljuj v smeri prve veje} \\
&\text{če je } (X_j > 8.2) \text{ nadaljuj v smeri druge veje}
\end{aligned}
\tag{16}.$$

Vozlišča v modelu z drevesno strukturo razbijejo celoten prostor vrednosti neodvisnih spremenljivk $\mathbf{X} = (X_1, X_2, \dots, X_p)$ na zaključene podmnožice, ki imajo obliko pravokotnikov (Hastie et al. 2001, str. 268). Pot do napovedi odvisne spremenljivke Y začnemo v prvem vozlišču na vrhu drevesa. Nato se v vsakem vozlišču odločimo za nadaljevanje po naslednji veji v skladu z vrednostjo neodvisne spremenljivke, ki sestavlja odločitveno pravilo v tem vozlišču. Ko pridemo do lista drevesa, napovemo vrednost spremenljivke Y z ustreznim pravilom $Y = F_i(X_1, X_2, \dots, X_p)$. Vsak list drevesa vsebuje eno pravilo F_i za napoved vrednosti odvisne spremenljivke Y , ki velja le za tisto podmnožico prostora vrednosti neodvisnih spremenljivk, ki je definirana z zaporedjem odločitev, ki pripeljejo do tega lista.

Če je Y kvalitativna spremenljivka, imenujemo drevesni model odločitveno drevo. Če je Y numerična spremenljivka, lahko drevesna struktura predstavlja bodisi modelno bodisi regresijsko drevo. Pri modelnem drevesu vsak list predstavlja en multipli regresijski model, ki velja za področje vzorčnega prostora, ki ga pokriva ta list. Pri regresijskem drevesu predstavlja vsak list povprečno vrednost odvisne spremenljivke Y za vzorce iz izvorne baze podatkov, ki pripadajo področju vzorčnega prostora, ki ga pokriva ta list. Ker smo z modeli za tržno vrednotenje nepremičnin napovedovali vrednost nepremičnine, ki je numerična spremenljivka, sta v nadaljevanju opisana le modelno in regresijsko drevo.

3.6.2 Modelno drevo

Modelno drevo je drevesna struktura, katere listi predstavljajo multiple regresijske modele (glej Sl. 2):

$$Y = F_i(X_1, X_2, \dots, X_p) = \beta_{i,0} + \sum_{j=1}^p \beta_{i,p} \cdot X_p
\tag{17}.$$

Opis procesa izgradnje modelnega drevesa na osnovi zaključene baze podatkov bomo povzeli po Wittnu in Franku (2000, str. 202-208). Proces izgradnje modelnega drevesa poteka v dveh korakih: *izgradnja drevesa* in *rezanje drevesa*¹³. V procesu izgradnje drevesa najprej razdelimo vzorčni prostor neodvisnih spremenljivk na podmnožice, za katere potem tvorimo liste – pravila za napoved spremenljivke Y . Naj T predstavlja množico $T \equiv \{(\mathbf{x}_t, y_t)\}$ vseh tistih vzorčnih parov za izgradnjo drevesa, ki so dosegli določeno vozlišče v drevesu. Z odločitvenim pravilom v tem vozlišču, ki se nanaša na eno izmed spremenljivk X_1, X_2, \dots, X_p , razdelimo množico T na dve ali več podmnožic T_j . Spremenljivko X_i in njene mejne vrednosti (za primer glej enačbi (15) in (16)), s katerimi razdelimo množico T in ki tvorijo odločitveno pravilo v danem vozlišču, izberemo tako, da maksimiramo pričakovano znižanje napake ob razbitju množice T v tem vozlišču. Napako v vozlišču merimo s standardno deviacijo napovedi za vse vzorce v množici T (Witten in Frank 2000, str. 203):

$$SDR = S(T) - \sum_j \frac{|T_j|}{|T|} \cdot S(T_j)
\tag{18},$$

¹³ Ang. *pruning the tree*.

kjer so:

SDR ... znižanje standardne deviacije,

$S(T)$... standardna deviacija množice T ,

$S(T_j)$... standardna deviacija množic T_j , ki jih dobimo z razbitjem množice T .

Absolutna vrednost v enačbi (18) predstavlja velikost množic oziroma število vzorcev v njih. Izgradnjo drevesa začnemo tako, da množico T sestavljajo vsi vzorci, ki jih uporabimo za izgradnjo drevesa. Nato razbijamo vse manjše množice T v posameznih vozliščih tako dolgo, dokler si niso vsi vzorci v posameznih vozliščih dovolj podobni med seboj, dokler standardna deviacija $S(T)$ ni dovolj nizka, ali dokler nam ne preostane le še nekaj vzorcev v vsakem vozlišču. Izgradnjo drevesa zaključimo s tem, da za vsako vozlišče neobrezanega drevesa določimo multipli regresijski model. Pri izračunu regresijskega modela za določeno vozlišče ne upoštevamo spremenljivk, ki so nastopale v odločitvenih pravilih predhodnih vozlišč, ki so pripeljala do tega vozlišča (Witten in Frank 2000, str. 203).

Izgradnji modelnega drevesa sledi njegovo rezanje, saj je neobrezano drevo navadno preveč prilagojeno naključnim odstopanjem v podatkih, ki smo jih uporabili za njegovo izgradnjo. Preden začnemo z rezanjem drevesa, v vseh vozliščih drevesa ocenimo, kakšna bi bila napaka za množico testnih vzorcev, ki niso sodelovali pri izgradnji drevesa. To napako ocenimo tako, da najprej izračunamo povprečno napako napovedi za vse vzorce, ki smo jih uporabili za izgradnjo drevesa in ki so dosegli posamezna vozlišča. Napaka za množico testnih vzorcev je navadno večja od te napake, zato jo ocenimo tako, da napako za učne vzorce pomnožimo s faktorjem $(J + pf \cdot v)/(J - pf \cdot v)$. J predstavlja število vzorcev, ki so dosegli dano vozlišče, v je število regresijskih koeficientov v pripadajočem regresijskem modelu, pf je faktor rezanja¹⁴, ki ga izbere uporabnik in se navadno giblje med vrednostima 1 in 10 (M5' 2003). Rezanje drevesa začnemo pri vsakem listu drevesa in odstranjujemo posamezne veje drevesa, dokler ocenjena napaka napovedi za testne vzorce pada, ko se vzpenjamo proti izhodišču drevesa.

Tako zgradimo modelno drevo, ki ga lahko uporabimo za napoved odvisnih spremenljivk. Vendar se je v praksi izkazalo, da so napovedi boljše, če tako dobljeno drevo zgladimo. To naredimo tako, da v vsakem vozlišču drevesa od njegovega izhodišča do listov izračunamo zglajeno napoved odvisne spremenljivke (Witten in Frank 2000, str. 202):

$$r' = \frac{J \cdot r + K \cdot q}{J + K} \quad (19),$$

kjer so:

r' ... zglajena napoved v danem vozlišču,

r ... zglajena napoved v predhodnem vozlišču,

q ... napoved z linearnim modelom v danem vozlišču,

J ... število objektov, ki so dosegli predhodno vozlišče,

K ... konstanta glajenja.

¹⁴ Ang. *pruning factor*. Z večanjem vrednosti faktorja pf dobimo enostavnejša drevesa z manj listi.

3.6.3 Regresijsko drevo

Regresijsko drevo je tudi drevesna struktura. Vsak njegov list je konstantna vrednost, ki predstavlja povprečje realizacij odvisnih spremenljivk Y za vse vzorce, ki so bili uporabljeni za izgradnjo drevesa in ki so dosegli ta list drevesa (glej Sl. 2):

$$Y = F_i(X_1, X_2, \dots, X_p) = C_i = \frac{1}{J} \sum_{j=1}^J y_j ; \text{ za vsak vzorec } j, \text{ ki je dosegel list } i \quad (20).$$

Regresijsko drevo je v bistvu modelno drevo, ki ima v posameznih listih namesto linearnih modelov konstantne vrednosti (linearni modeli, ki vsebujejo le konstantne člene). Izgradnja regresijskega drevesa poteka na enak način kakor izgradnja modelnega drevesa, s to razliko, da pri regresijskem drevesu napovedi ne zgladimo (Witten in Frank 2000, str. 201).

3.7 Nevronske mreže

S pojmom nevronska mreža označujemo nabor različnih matematičnih modelov, ki imajo mrežno strukturo jih lahko uporabljamo za napovedovanje ali razvrščanje. V magistrski nalogi smo uporabili posebno vrsto nevronskih mrež imenovano večnivojski perceptron.¹⁵ Opis večnivojskega perceptrona smo povzeli po Hanu in Kamberju (2001, str. 303-311) ter Handu in soavtorjih (2001, str. 391-393). Opis nevronskih mrež s statističnega vidika se nahaja v delu Hastie-ja in soavtorjev (2001, str. 347-369).

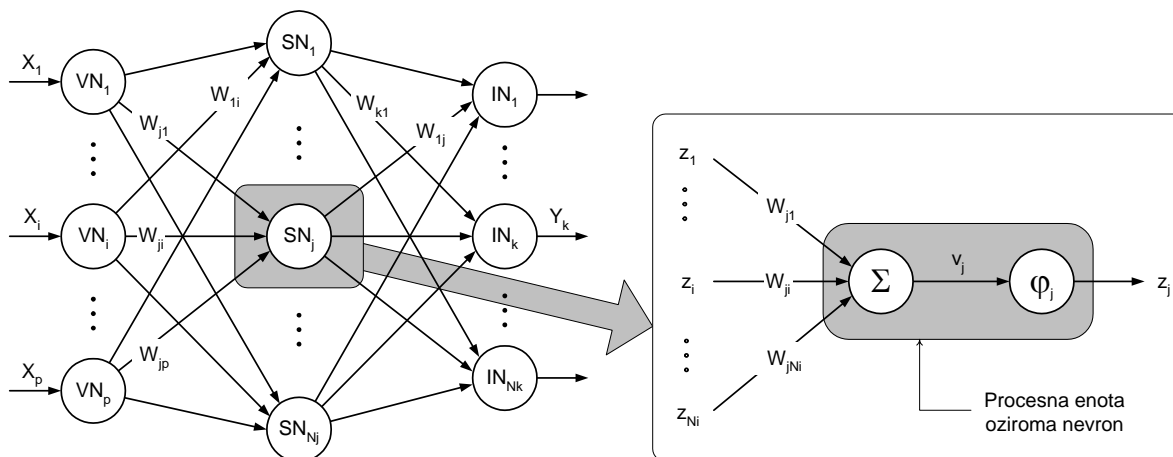
Večnivojski perceptron je mrežna matematična struktura, ki je sestavljena iz procesorskih enot – nevronov, ki so razvrščeni v posamezne plasti – nivoje. Nevroni na posameznih nivojih so med seboj povezani s povezavami, ki imajo različne uteži. Te povezave imenujemo sinapse. Nevroni v prvem nivoju služijo za vnos vrednosti neodvisnih spremenljivk X_i v nevronske mreže in navadno niso procesorske enote. Nevroni v zadnjem nivoju so izhodi iz nevronske mreže. Izhodne vrednosti teh nevronov predstavljajo napovedane vrednosti ene ali več odvisnih spremenljivk Y . Nivoje nevronov, ki se nahajajo med prvim in zadnjim nivojem nevronov imenujemo skrite nivoje, ker uporabnik nima dostopa do nevronov na njih. Primer večnivojskega perceptrona z enim skritim nivojem je shematsko prikazan na Sl. 3.

Oznake na Sl. 3 pomenijo naslednje:

X_i	... neodvisna spremenljivka,
Y_k	... odvisna spremenljivka,
VN	... nevron na vhodnem nivoju,
SN	... nevron na skritem nivoju,
IN	... nevron na izhodnem nivoju,
W	... utež sinaptične povezave, ki povezuje nevrona na dveh sosednjih nivojih,
z_j	... izhod iz j -tega nevrna,
v_j	... vzbuditev j -tega nevrna,

¹⁵ Ang. *multilayer perceptron* ali *multilayer feed-forward neural network*. To je najbolj pogosto uporabljena vrsta nevronskih mrež, zato jo uporabniki navadno imenujejo kar s splošnim izrazom nevronska mreža.

φ_j ... sigmoidna funkcija j -tega nevrona (na primer logistična funkcija).



Sl. 3: Večnivojski perceptron z eno skrito ravnino

Vrednost odvisnih spremenljivk za dane vrednosti neodvisnih spremenljivk $\mathbf{x} = (x_1, x_2, \dots, x_p)$ izračunamo tako, da najprej določimo izhodne vrednosti za nevrone na prvem skitem nivoju:

$$z_j = \varphi_j(v_j) = \varphi_j\left(\sum_{i=1}^{N_j} W_{ji} \cdot z_i + \theta_j\right) \quad (21).$$

z_i so realizirane vrednosti neodvisnih spremenljivk x_i . N_j je število vhodov v nevron, θ_j je prag vzbuditve nevrona. Tako izračunane vrednosti z_j nato preko sinaptičnih povezav z utežmi W pripeljemo do vhodov nevronov na naslednjem skitem nivoju ali na izhodnem nivoju, če ima nevronska mreža en sam skiti nivo. Za vsak nevron na naslednjem nivoju ponovno izračunamo njegovo izhodno vrednost z enačbo (21). Ta vrednost predstavlja bodisi napoved odvisne spremenljivke Y_k , če je ta nevron na izhodnem nivoju, bodisi vhod za naslednji nivo nevronov, če je ta nevron na skitem nivoju. Postopek ponavljamo, dokler ne pridemo do izhodnega nivoja. V našem primeru smo nevronska mrežo uporabili za izgradnjo modela za tržno vrednotenje nepremičnin. Neodvisne spremenljivke so predstavljale lastnosti nepremičnin. Odvisna spremenljivka, ki je predstavljala tržno vrednost nepremičnin, je bila le ena, zato je imela nevronska mreža en sam nevron na izhodnem nivoju.

Uteži W predstavljajo proste parametre matematičnega modela nevronske mreže, podobno kakor regresijski koeficienti β predstavljajo proste parametre modela multiple regresije. Vrednosti uteži W ocenimo na osnovi parov realiziranih vrednosti neodvisnih in odvisnih spremenljivk $\{(\mathbf{x}_j, y_j); j=1, \dots, n\}$ z algoritmom povratnega širjenja napak.¹⁶ Postopek določevanja vrednosti uteži W imenujemo učenje nevronske mreže. Hitrost učenja nevronske mreže reguliramo s parametrom učenja in s parametrom vztrajnosti. Podrobnosti algoritma za učenje nevronske mreže nismo navajali, ker presegajo obseg tega dela. Mogoče jih je najti v ustrezni literaturi (Han in Kamber 2001, str. 305-311).

Z razvojem vse hitrejših računalnikov v zadnjih desetih letih je prišlo do široke uporabe nevronske mreže na različnih področjih. Njihova prednost je, da omogočajo modeliranje zelo zapletenih in močno nelinearnih povezav med odvisnimi in neodvisnimi spremenljivkami.

¹⁶ Ang. *back-propagation algorithm*.

Glavna slabost, ki omejuje njihovo uporabo v ekonomsko-poslovnih vedah, je dejstvo, da je zaradi zapletene matematične strukture napovedane vrednosti odvisnih spremenljivk Y praktično nemogoče interpretirati v povezavi z vrednostmi neodvisnih spremenljivk X .

3.8 Metode za vrednotenje kakovosti in primerjavo modelov

Kot smo omenili že v uvodu, je del bil naše naloge vezan na izgradnjo alternativnih modelov za tržno vrednotenje nepremičnin in njihovo medsebojno primerjavo. Če smo želeli oceniti primernost postavljenega modela za opis določenega pojava in če smo ta model želeli primerjati z alternativnimi modeli, smo morali definirati mero za oceno uspešnosti napovedi ter predpisati postopek za njen izračun.

V literaturi s področja rudarjenja podatkov je mogoče najti najrazličnejše mere za oceno uspešnosti napovedi \hat{y} odvisne spremenljivke Y za dane realizacije neodvisnih spremenljivk $\mathbf{x}_n = (x_{n,1}, \dots, x_{n,p})$; $n = 1, \dots, N$ (Han in Kamber 2001, Hand et al. 2001, Hastie et al. 2001), (Witten in Frank 2000). Vse mere temeljijo na odstopanju dejanske vrednosti odvisne spremenljivke y_n in njene napovedi \hat{y}_n . V magistrskem delu smo uporabili naslednje mere za oceno uspešnosti napovedi (Witten in Frank 2000, str. 148):

- srednja vrednost absolutnih napak:
$$SAN = \frac{1}{N} \sum_{n=1}^N |\hat{y}_n - y_n| \quad (22),$$

- relativna absolutna napaka:
$$RAN = \frac{1}{N} \sum_{n=1}^N \frac{|\hat{y}_n - y_n|}{|y_n - \bar{y}|} \quad (23),$$

- koren srednjega kvadratičnega odstopanja:
$$KSKO = \sqrt{\frac{1}{N} \sum_{n=1}^N (\hat{y}_n - y_n)^2} \quad (24),$$

- koren relativnega kvadratičnega odstopanja:
$$KRKO = \sqrt{\frac{1}{N} \sum_{n=1}^N \frac{(\hat{y}_n - y_n)^2}{(y_n - \bar{y})^2}} \quad (25),$$

- korelacijski koeficient:

$$\rho = \frac{\frac{1}{N-1} \sum_{n=1}^N (\hat{y}_n - \bar{\hat{y}}) \cdot (y_n - \bar{y})}{\sqrt{\frac{1}{N-1} \sum_{n=1}^N (\hat{y}_n - \bar{\hat{y}})^2 \cdot \frac{1}{N-1} \sum_{n=1}^N (y_n - \bar{y})^2}} \quad (26).$$

Z merama SAN in $KSKO$ za testno množico vzorcev ocenjujemo nominalno odstopanje napovedi od dejanskih vrednosti. Z merama RAN in $KRKO$ pa ocenjujemo relativno uspešnost napovedi v primerjavi s povprečno vrednostjo odvisne spremenljivke za testno množico.

Oznake \bar{y} predstavljajo srednje vrednosti ustreznih količin. V praksi za mero odstopanja pogosto uporabljamo tudi determinacijski koeficient R^2 , ki je enak kvadratu korelacijskega koeficienta ρ iz enačbe (26).

Pri ovrednotenju kakovosti modela moramo biti previdni. Če kakovost modela ocenjujemo na osnovi podatkov, ki smo jih uporabili že pri izgradnji modela, so napovedi odvisne spremenljivke Y v splošnem boljše kakor v primeru, če kakovost modela ocenjujemo na osnovi podatkov, ki niso sodelovali pri izgradnji modela. Zato navadno množico vseh podatkov, ki so na voljo za izgradnjo modela, razdelimo na dva dela:

- na množico podatkov $\{(\mathbf{x}_j, y_j) ; j = 1, \dots, n_{uc}\}$, ki jo uporabimo le za izgradnjo modela
– to množico imenujemo *učna množica*,
- na množico podatkov $\{(\mathbf{x}_j, y_j) ; j = 1, \dots, n_{rest}\}$, ki jo uporabimo le za testiranje modela
– to množico imenujemo *testna množica*.

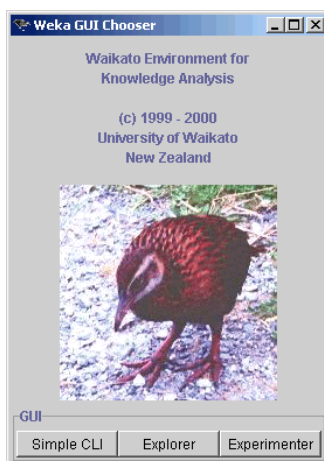
Kakovost modela preizkusimo tako, da ga zgradimo s podatki iz učne množice, nato zgrajen model preizkusimo s podatki iz testne množice. Na ta način dobimo boljšo informacijo o tem, kako se zgrajeni model obnaša pri napovedi vrednosti odvisne spremenljivke Y za vzorce, ki jih še nikoli "ni videl". Za oceno kakovosti modela na osnovi testne množice lahko uporabimo vse zgoraj omenjene mere uspešnosti napovedi vrednosti odvisne spremenljivke.

Z delitvijo modela na testno in učno množico se pojavi nov problem. Modeli za napoved, ki jih zgradimo na osnovi različnih delitev podatkov na testno in učno množico, so lahko različno kvalitetni. Za uporabnika je zelo ugodno, če lahko oceni povprečno kakovost napovedi, ki jo omogoča izbrani model za dano množico podatkov, ki so na voljo. Povprečno kakovost izbranega modela za napoved pri dani množici podatkov je mogoče izračunati s postopkom prečnega preverjanja (Witten in Frank 2000, str. 123-125, Hand et al. 2001, str. 148-149, Hastie et al. 2001, str. 214-217). Pri postopku prečnega preverjanja razdelimo množico vseh podatkov $\{(\mathbf{x}_j, y_j) ; j = 1, \dots, n\}$, ki so na voljo za izgradnjo modela, na G disjunktnih podmnožic. Nato zgradimo G modelov za napoved neodvisne spremenljivke Y tako, da vsakič uporabimo eno izmed G podmnožic za testiranje modela, vse preostale podmnožice pa za izgradnjo modela. Vsak zgrajeni model testiramo s pripadajočo testno množico vzorcev in izračunamo mero uspešnosti napovedi za to testno množico. Povprečno kakovost napovedi za izbrani model ocenimo tako, da izračunamo povprečno vrednost mer uspešnosti napovedi, ki ustrezajo posameznim zgrajenim modelom. V magistrskem delu smo s postopkom prečnega preverjanja ocenili kakovost modelov za množično vrednotenje nepremičnin po tržni metodi.

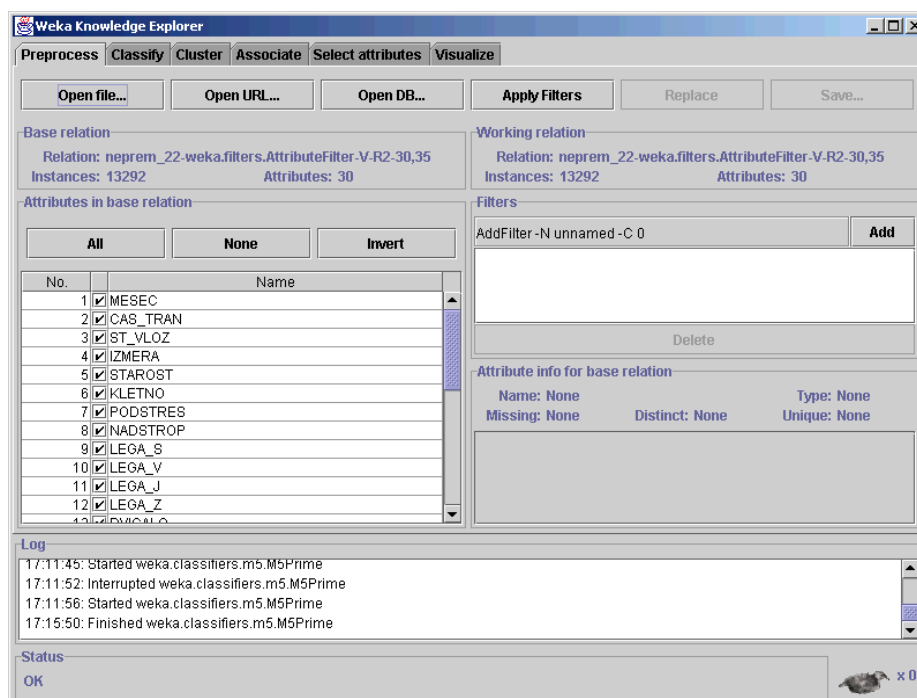
3.9 Predstavitev uporabniškega programa WEKA

Vse zgoraj opisane analize smo izvajali s programskima paketoma *SPSS 10.0* in *WEKA 3-2*. S programom *SPSS 10.0* smo opravljali osnovne statistične analize, za analize z metodami rudarjenja podatkov pa smo uporabili program *WEKA 3-2*. Program *SPSS 10.0* je sorazmerno znan med ekonomskimi analitiki, zato ga ne bomo posebej opisovali, ker lahko morebitni uporabnik najde njegov opis tudi v slovenski literaturi (Rovan in Turk 2001).

V nadaljevanju tega podpoglavja sta opisana uporabniški vmesnik in delo s programom *WEKA 3-2*. Podrobnejši opis programa *WEKA 3-2* je mogoče najti v knjigi Wittna in Franka (2000). *WEKA 3-2* je programsko okolje, ki združuje knjižnice za analizo podatkov z metodami rudarjenja podatkov. Program *WEKA 3-2* je napisan v programskem jeziku JAVA, zato ga lahko namestimo na poljubni operacijski sistem, torej tudi v okolje *WINDOWS*. Program je prosto dostopen na internetu (*WEKA 3* 2003). V *WINDOWS* okolju program zaženemo s pritiskom na ustrezno ikono v meniju START. S tem odpremo osnovno izbirno okno, ki je prikazano na Sl. 4. Nadaljujemo tako, da s pritiskom na gumb *EXPLORER* odpremo uporabniški vmesnik za analizo in obdelavo podatkov, ki je prikazan na Sl. 5.



Sl. 4: Vhodno okno uporabniškega programa WEKA 3-2



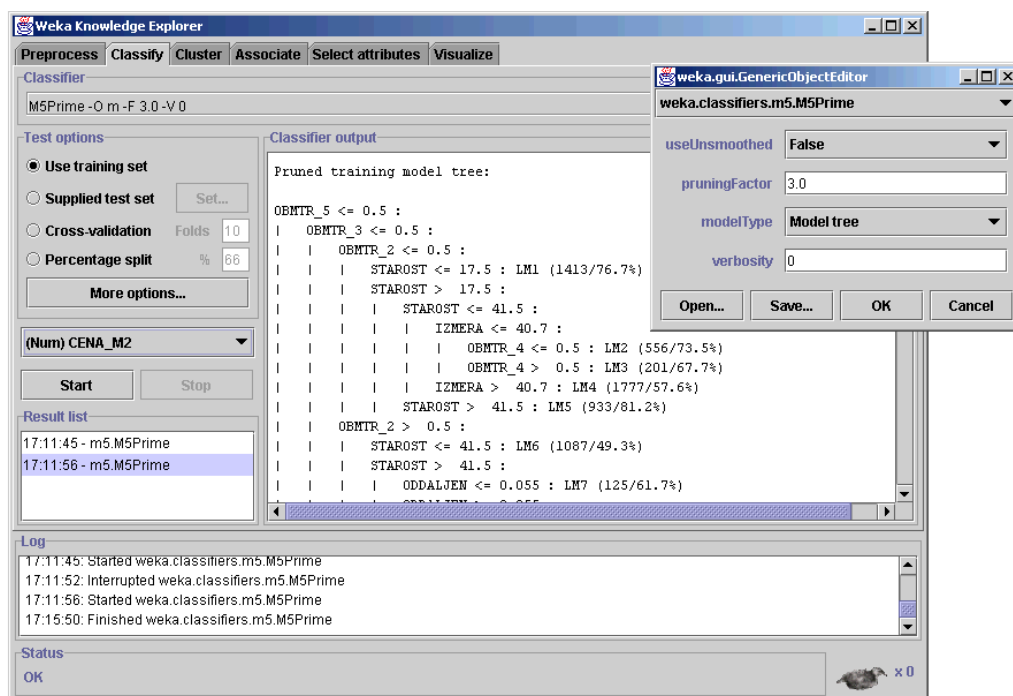
Sl. 5: Okno uporabniškega programa WEKA 3-2 za predprocesiranje podatkov

V uporabniškem vmesniku za analizo in obdelavo podatkov je mogoče izbrati več opcij (glej Sl. 5): predprocesiranje podatkov (mapa *Preprocess*), izgradnja modelov (mapa *Classify*), razvrščanje objektov v skupine (mapa *Cluster*), iskanje povezovalnih pravil (mapa *Associate*), izbira spremenljivk (mapa *Select attributes*) in prikaz podatkov (mapa *Visualize*). Pri izdelavi magistrskega dela smo uporabljali prvo, drugo in četrto opcijo.

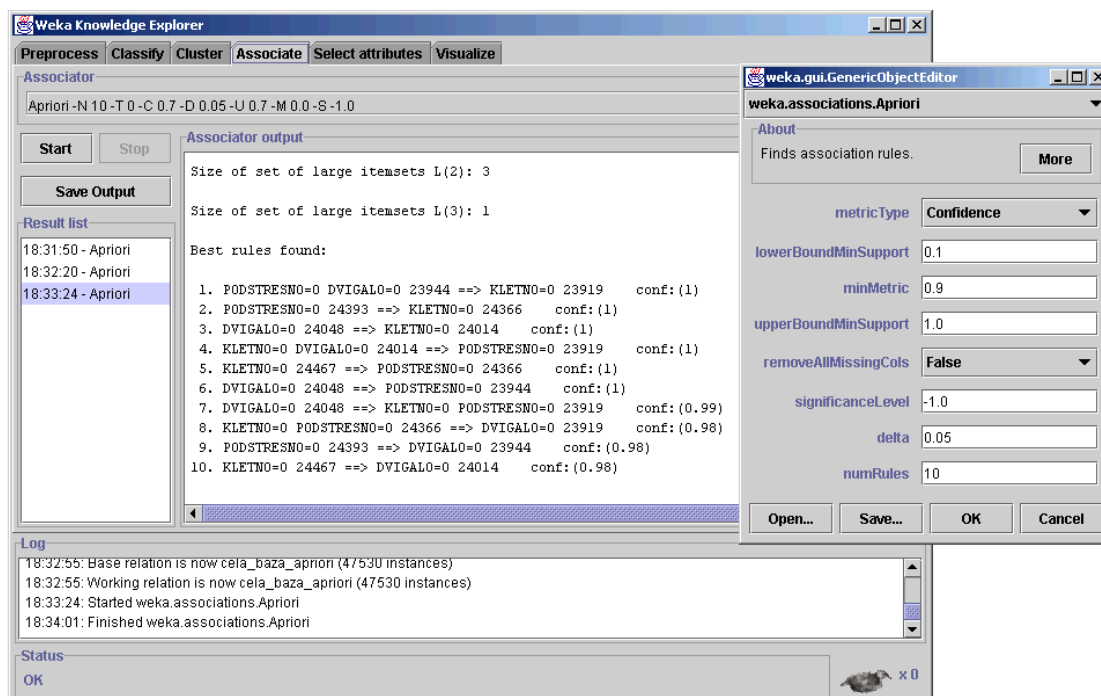
Mapa za predprocesiranje podatkov je prikazana na Sl. 5 in omogoča izbiro vira podatkov (tabelarična datoteka, baza podatkov ali internet). Podatki morajo biti v ARFF formatu. Iz izbranega vira se podatki prenesejo v program *WEKA 3-2*. V oknu na Sl. 5 se prikažejo osnovne informacije o spremenljivkah za podatke iz izbranega vira. Te spremenljivke je mogoče vključiti v analizo ali izključiti iz nje z ukazom *Apply filters* – glej sliko Sl. 5.

Z izbiro mape *Classify* odpremo okno za izgradnjo različnih modelov za napoved – glej Sl. 6. Želeni matematični model izberemo v okvirju *Classifier* (glej Sl. 6). Njegove parametre nastavimo v oknu za nastavitve parametrov (glej manjše okno na Sl. 6). Različnim modelom

ustrezajo različni parametri. Izbrani matematični model je uporabljen na podatkih, ki smo jih izbrali v oknu na Sl. 5. Izgradnjo modela poženemo tako, da v oknu na Sl. 6 izberemo način testiranja modela (okvir *Test options*) in odvisno spremenljivko (primer *CENA_M2*) ter poženemo program s pritiskom na gumb *START*. Zgrajeni model in parametri za oceno njegove kakovosti (glej poglavje 3.8) se izpišejo v okvirju *Classifier output*.



Sl. 6: Okno uporabniškega programa WEKA 3-2 za izgradnjo modela podatkov



Sl. 7: Primer okna uporabniškega programa WEKA 3-2 za iskanje povezovalnih pravil

Z izbiro mape *Associate* odpremo okno za iskanje povezovalnih pravil v bazi podatkov – glej Sl. 7. Iskanje povezovalnih pravil se izvaja z *APRIORI* algoritmom – glej poglavje 3.3. Preden začnemo z iskanjem povezovalnih pravil, v oknu za nastavitve parametrov (glej

manjše okno na Sl. 7) nastavimo parametra zaupanja (*confidence*) in podpore (*support*) ter želeno število najdenih povezovalnih pravil. Ko je postopek iskanja povezovalnih pravil končan, se najdena povezovalna pravila izpišejo v okvirju *Associator output* skupaj s pripadajočima vrednostima parametrov zaupanja in podpore.

4 Baza podatkov nepremičninskih transakcij DURS

4.1 Osnovne značilnosti baze nepremičninskih transakcij

Kot osnovo za analizo nepremičninskih transakcij v Republiki Sloveniji in izgradnjo modelov za množično vrednotenje stanovanjskih nepremičnin po tržni metodi smo uporabili delno bazo nepremičninskih transakcij DURS-a (Baza podatkov DURS 2003). Baza DURS-a je zbirna baza podatkov, ki je bila vzpostavljena za namen obračuna davka od nepremičninskih transakcij. Vsebuje zapise o uradno prijavljenih nepremičninskih transakcijah, ki so bile izvedene v posameznih občinah na območju Republike Slovenije od leta 1999 do leta 2002.

V bazi so zapisi o 104607 opravljenih transakcijah z nepremičninami v Republiki Sloveniji in sicer 17549 zapisov za leto 1999, 28117 zapisov za leto 2000, 28583 zapisov za leto 2001 in 30358 zapisov za leto 2002. V zadnjih treh letih se število opravljenih transakcij giblje okoli števila 30000, vendar to ne pomeni, da je bilo leta 1999 opravljenih bistveno manj transakcij. Glavni razlog za manjše število zapisov v bazi je, da je bila ta baza podatkov vzpostavljena šele leta 1999. To potrjuje tudi dejstvo, da je bila leta 1999 kvaliteta vnosov v bazi slabša v primerjavi z ostalimi tremi leti, kot bomo videli kasneje.

V bazi nepremičninskih transakcij DURS-a je ločenih 11 vrst nepremičnin, ki jih lahko v grobem razdelimo na zemljišča in stavbe. Podatki o vrstah nepremičnin in številu registriranih transakcij z njimi v letih od 1999 do 2002 so podani v tabeli 1. Številka pred imenom vrste nepremičnine v tabeli 1 je šifra nepremičnine po šifrantu DURS-a.

Tabela 1: Vrste nepremičnin in število opravljenih transakcij z njimi

Vrsta nepremičnine		Število transakcij
Zemljišča	11 – Komunalno opremljena zemljišča	8940
	12 – Kmetijska zemljišča	21605
	13 – Komunalno neopremljena zemljišča	12162
	14 – Funkcionalna zemljišča	4823
Stavbe oz. objekti	21 – Poslovni prostori	8206
	22 – Stanovanja	29972
	23 – Hiše	8895
	24 – Garaže	3732
	25 – Vikendi	951
	26 – Objekti z zemljiščem	3028
	27 – Ostalo	1650
Število vseh upoštevanih transakcij		103964
Število vseh transakcij v bazi		104607
Število neupoštevanih transakcij iz baze		643

V tabeli 1 nismo upoštevali zapisov, pri katerih ni bilo mogoče ugotoviti leta transakcije ali vrste nepremičnine. Teh nepremičnin prav tako nismo upoštevali v nadaljnjih analizah. V tabeli 1 je mogoče videti, da je bilo največ (kar 28,8%) nepremičninskih transakcij v Sloveniji

opravljenih s stanovanji. To je tudi eden izmed razlogov, da smo v nadaljevanju zgradili modele za množično vrednotenje nepremičnin po tržni metodi prav za stanovanja.

Za posamezne zapise oziroma nepremičninske transakcije se v bazi nahaja 29 podatkov, ki predstavljajo spremenljivke v naših analizah (Šifrant podatkov 2003):

- *LETO*: leto opravljene transakcije.
- *ZAP_POG*: zaporedna številka pogodbe.
- *ZAP_NEP*: zaporedna številka nepremičnine v pogodbi.
- *DATUM_PRIJAVE*: datum prijave pogodbe.
- *VRSTA_NEPR*: vrsta nepremičnine – glej tabelo 1.
- *SIF_KO*: šifra katastrske občine zemljišča.
- *SIF_OBC*: šifra občine po šifrantu o novih občinah, ki velja od leta 1995 do leta 2002 (Občine RS 2003a, 2003b, 2003c). Sredi leta 2002 se je šifrant občin spremenil tako, da je sedaj v Sloveniji 193 namesto prejšnjih 192 občin. Vendar v bazi nepremičninskih transakcij DURS-a ni nobenega vnosa za novo nastalo občino, zato smo v magistrskem delu upoštevali šifrant o novih občinah iz leta 1995 do konca leta 2002.
- *ST_VLOZ*: vložna številka zemljišča ali stavbe.
- *PARCELA*: številka parcele.
- *KK_ZEM*: katastrska kultura zemljišča, ki ima 13 vrednosti:
 - A – njiva,
 - B – plantažni sadovnjak,
 - C – ekstenzivni sadovnjak,
 - E – vinograd
 - F – hmeljišče,
 - G – travnik,
 - H – barjanski travnik,
 - I – pašnik,
 - J – trsišče,
 - K – gozdna plantaža,
 - L – gozd,
 - M – ostalo (pot, neplodno zemljišče itn.).
- *RAZRED*: bonitetni razred zemljišča, ki ima vrednosti od 1 do 9.
- *IZMERA*: izmera nepremičnine v m².
- *VREDNOST_POG*: pogodbena vrednost nepremičnine v SIT.
- *VREDNOST_ODM*: vrednost nepremičnine v SIT, od katere se bo odmeril davek (le v primeru naknadne cenitve nepremičnine).
- *LETO_IZGR*: leto izgradnje objekta.
- *LETO_PREN*: leto obnove objekta.

- *KLETNO*: indikator kletnega stanovanja, ki ima 2 vrednosti:
 - 1 – stanovanje je kletno,
 - 0 – stanovanje ni kletno.
- *PODSTRESNO*: indikator podstrešnega stanovanja, ki ima 2 vrednosti:
 - 1 – stanovanje je podstrešno,
 - 0 – stanovanje ni podstrešno.
- *NADSTROPJE*: številka nadstropja v katerem nepremičnina leži (šifra pritličja je 99).
- *LEGA*: lega nepremičnine, ki ima 4 vrednosti:
 - S – sever,
 - V – vzhod,
 - J – jug,
 - Z – zahod.
- *DVIGALO*: indikator dvigala, ki ima 2 vrednosti:
 - 1 – objekt ima dvigalo,
 - 0 – objekt nima dvigala.
- *ODDALJENOST*: oddaljenost nepremičnine od središča mesta v km.
- *ELEKTRIKA*: indikator električne napeljave, ki ima 2 vrednosti:
 - 1 – nepremičnina ima električno napeljavo,
 - 0 – nepremičnina nima električne napeljave.
- *VODOVOD*: indikator vodovoda, ki ima 2 vrednosti:
 - 1 – nepremičnina ima vodovod,
 - 0 – nepremičnina nima vodovoda.
- *OGREVANJE*: indikator ogrevanja, ki ima 2 vrednosti:
 - 1 – nepremičnina ima ogrevanje,
 - 0 – nepremičnina nima ogrevanja.
- *TELEFON*: indikator telefona, ki ima 2 vrednosti:
 - 1 – nepremičnina ima telefon,
 - 0 – nepremičnina nima telefona.
- *KANALIZACIJA*: indikator kanalizacije, ki ima 2 vrednosti:
 - 1 – nepremičnina ima kanalizacijo,
 - 0 – nepremičnina nima kanalizacije.
- *LETO_NAKUPA*: leto nakupa nepremičnine prodajalca.
- *NACIN_PRIDOB*: način pridobitve nepremičnine, ki ima 8 vrednosti:
 - 1 – nakup,
 - 2 – darilo,
 - 3 – dedovanje,
 - 4 – menjava,
 - 5 – denacionalizacija,
 - 6 – razlastitev,
 - 7 – nova gradnja,

- 9 – ostalo.

Na DURS-u seveda niso zbirali vseh podatkov za vse vrste nepremičnin. Tako je na primer podatek o letu obnove objekta nesmiseln za zemljišča, podatek o katastrski kulturi zemljišča nesmiseln za stanovanja itn. Informacije o tem, katere spremenljivke smo upoštevali pri posameznih analizah in katerih nismo upoštevali, so podane sproti.

4.2 Ocena kvalitete baze nepremičninskih transakcij

4.2.1 Celovitost in kakovost podatkov v bazi

V splošnem je vsaka nepremičninska transakcija opisana z več podatki ali spremenljivkami – glej poglavje 4.1. Celovitost in kakovost podatkov v bazi nepremičninskih transakcij je imela bistven vpliv na kvaliteto rezultatov, ki smo jih dobili pri analizi nepremičninskih transakcij in izgradnji modelov za vrednotenje nepremičnin. Splošna značilnost podatkovnih baz o nepremičninskih transakcijah je, da so podatki v njih močno pomanjkljivi, njihova kakovost pa je slaba (Knight et al. 1998). Knight in sodelavci (1998) ločijo tri vrste pomanjkljivosti, ki se nanašajo na podatke o nepremičninskih transakcijah:

- *Manjkajoči podatki*: manjkajoči podatki so podatki o transakciji, ki jih je načelno mogoče pridobiti, a iz kakršnega koli razloga niso bili vneseni v bazo podatkov.
- *Napačno izmerjeni podatki*: napačno izmerjeni podatki se pojavijo takrat, ko prave vrednosti za določeno spremenljivko bodisi ni mogoče izmeriti, ali pa jo je mogoče izmeriti le z napako.
- *Pogojno vključeni podatki*¹⁷: pogojno vključeni podatki so podatki, ki so na voljo le, če je izpolnjen določen pogoj – na primer: če neka spremenljivka pade v določeno kategorijo, tedaj so na voljo tudi vrednosti za ostale spremenljivke, sicer pa ne.

Baza DURS-a je tipična baza nepremičninskih transakcij, zato se v njej pojavijo vse tri vrste pomanjkljivosti.

Manjkajoči podatki

V bazi nepremičninskih transakcij DURS-a je veliko zapisov z manjkajočimi podatki. Ker je takšne podatke sorazmerno enostavno odkriti, smo v sklopu analize baze nepremičninskih transakcij DURS-a skušali ugotoviti uspešnost posameznih občin pri zbiranju podatkov o nepremičninskih transakcijah. Zato smo za vsako občino izračunali povprečni delež nepraznih vnosov na transakcijo¹⁸ za vsako leto posebej in nato še povprečni delež nepraznih vnosov za vsa leta skupaj. Pri tem smo ločeno obravnavali zemljišča (nepremičnine od 11 do 14) in stavbe (nepremičnine od 21 do 27).

Povprečni delež nepraznih vnosov na transakcijo z zemljiščem oziroma stavbo za posamezno občino v enem letu smo dobili tako, da smo najprej poiskali vse zapise v bazi podatkov za zemljišča oziroma stavbe za to občino v tem letu. Nato smo na osnovi najdenih zapisov za vse

¹⁷ Za pogojno vključene podatke se v angleški literaturi uporablja izraz “*censored data*”.

¹⁸ Povprečni delež nepraznih vnosov je dober indikator kakovosti podatkov v bazi, ker prikazuje povprečni delež vnesenih podatkov za eno transakcijo.

izbrane spremenljivke od *ST_VLOZ* do *NACIN_PRIDOB* izračunali deleže nepraznih vnosov za posamezne spremenljivke.¹⁹ Na koncu smo izračunali srednjo vrednost deležev nepraznih vnosov po vseh izbranih spremenljivkah. Pri tem nismo upoštevali istih spremenljivk za zemljišča in stavbe. Ker so pri zemljiščih podatki o spremenljivkah *KLETNO*, *PODSTRESNO*, *NADSTROPJE* in *DVIGALO* nepomembni, jih pri analizi nismo upoštevali. Pri stavbah nismo upoštevali spremenljivk *PARCELA*, *RAZRED* in *KK_ZEM*. Povprečni deleži nepraznih vnosov na transakcijo za zemljišča so prikazani v tabeli P1-1 v prilogi 1. Povprečni deleži nepraznih vnosov na transakcijo za stavbe so prikazani v tabeli P1-2 v prilogi 1. Poleg tega smo izračunali tudi delež nepraznih vnosov na transakcijo za vse vrste nepremičnin skupaj za vsa leta skupaj. Delež nepraznih vnosov za vse vrste nepremičnin skupaj smo izračunali z uteženim povprečjem odstotkov vnosov za zemljišča in stavbe. Uteži so bile določene glede na število opravljenih transakcij v štirih letih za posamezno vrsto nepremičnin v posamezni občini. Izračunani deleži nepraznih vnosov na transakcijo za vse vrste nepremičnin so prikazani v tabeli P1-3 v prilogi 1.

Iz tabele P1-1 v prilogi 1 je razvidno, da so pri zemljiščih deleži nepraznih vnosov v letih 2001 in 2002 zelo slabi, saj se gibljejo okoli vrednosti 0,6. To je posledica dejstva, da se v letih 2001 in 2002 komunalna opremljenost zemljišč ni več vpisovala v bazo nepremičninskih transakcij. Zaradi tega razloga so deleži nepraznih vnosov za stavbe v splošnem boljši kakor za zemljišča – glej tabeli P1-2 in P1-3. Če naredimo primerjavo med vsemi tremi pristopi (deleži nepraznih vnosov za zemljišča, stavbe in vse vrste nepremičnin), vidimo, da na skupni delež nepraznih vnosov vplivajo bodisi zemljišča bodisi stavbe, odvisno od tega, katerih transakcij je več. Nekaj je primerov, kjer vplivajo vse vrste nepremičnin. Pri nekaterih občinah so se pri stavbnih nepremičninah pojavile anomalije zaradi izredno slabe kakovosti vnosov v nekaterih letih (Krško in Gornji grad v letu 1999, Metlika v letu 2002), kar lahko pokvari indikator skupne kakovosti vnosov za dotično občino – glej tabelo P1-2.

Deleže nepraznih vnosov za posamezne občine za vsa leta smo tudi rangirali po velikosti, in sicer posebej za zemljišča, stavbe in vse nepremičnine skupaj – glej tabelo P1-3 v prilogi 1. Za tako dobljene range smo izračunali medsebojne korelacijske koeficiente – glej tabelo 2.

Tabela 2: Korelacijski koeficienti med rangi deležev nepraznih vnosov

	Rang deleža nepraznih vnosov za zemljišča	Rang deleža nepraznih vnosov za stavbe	Rang deleža nepraznih vnosov za vse neprem.
Rang deleža nepraznih vnosov za zemljišča	1,000	-0,214**	0,527**
Rang deleža nepraznih vnosov za stavbe	-0,214**	1,000	0,305**
Rang deleža nepraznih vnosov za vse neprem.	0,527**	0,305**	1,000

Opomba: ** ... Korelacijski koeficient je značilen pri stopnji značilnosti 0,01.

Zanimivo je, da je korelacijski koeficient med rangi deležev nepraznih vnosov za zemljišča in stavbe negativen. To pomeni, da je imelo kar nekaj občin z dobrimi vnosi za stavbe slabe

¹⁹ Spremenljivk od *LETO* do *SIF_OBC* pri tej analizi nismo upoštevali, ker so bile vnesene pri vseh transakcijah, zato bi njihovo upoštevanje lahko izkrivilo oceno dejanske kvalitete vnosov v bazo podatkov.

vnose za zemljišča in obratno. Takšna primera sta na primer Bistrica ob Sotli (dobra pri zemljiščih, slaba pri stavbah) in Vrhnika (dobra pri stavbah, slaba pri zemljiščih). V teh dveh primerih je to lahko tudi posledica majhnega števila transakcij za določene vrste nepremičnin, ki so bile izvedene v opazovanem obdobju. Občina Prebold je primer občine, ki je bila dobra pri zemljiščih in pri stavbah. Če gledamo kakovost vnosov za vse vrste nepremičnin, so najboljše občine: Celje, Ljubljana, Jesenice, Štore in Hodoš. V tem pogledu so najslabše občine: Jezersko, Odranci, Krško, Loški Potok in Dol pri Ljubljani – glej tabelo P1-3.

Napačno izmerjeni podatki

Napačno izmerjenih podatkov je sicer bistveno manj kot manjkajočih podatkov, a je bilo nekatere zelo težko odkriti. Tipični primeri napačno izmerjenih podatkov so bili naslednji:

- Zelo velika (nekaj tisoč m²) ali zelo majhna (1 m²) stanovanjska površina.
- Zelo visok delež stanovanj (60%), ki naj bi se nahajala v 2. nadstropju. Iz tega podatka sledi, da naj bi bilo največ na trgu ponujenih stanovanj v 2. nadstropju ali da naj bi bila stanovanja v 2. nadstropju najbolj zaželena. Vendar za ta pojav ni racionalne ekonomske razlage, zato sklepamo, da gre za posebno anomalijo podatkov v bazi. Zaradi tako velikega deleža stanovanj v 2. nadstropju je tudi težko sklepati na napake pri vnosu. Do te napake bi lahko prišlo, če je privzeta vrednost spremenljivke *NADSTROPJE* v uporabniškem vmesniku za vnašanje podatkov v bazo enaka 2.
- Zelo nizka cena kvadratnega metra stavbnih nepremičnin, ki ne doseže niti 1 SIT/m².
- Ekstremno visoko nadstropje nepremičnine, ki doseže vrednost tudi več deset nadstropij (50). Čeprav tako visoke stavbe obstajajo, se nobena od njih ne nahaja v Sloveniji. Tipična višina petdeset-nadstropne stavbe je med 150 in 200 m, najvišja stavba v Sloveniji (stolpnica WTC v Ljubljani) pa je visoka malo več kakor 70 m.
- Ekstremno velika oddaljenost nepremičnine od centra mesta, ki znaša tudi več sto kilometrov. Tako velikih občin namreč v Sloveniji ni, poleg tega bi se pri takšni oddaljenosti od centra mesta nepremičnina lahko nahajala že v sosednji državi.
- Pri občini Koper je vrednost spremenljivke *LETO_IZGR* pogosto enaka 9999, vrednost spremenljivke *LETO_NAKUPA* za nepremičninske transakcije v letu 1999 pa je pogosto enaka 1900.

Do večine teh napak je prišlo zaradi površnih vnosov podatkov v bazo. Nekatere napake so bile tudi posledica uporabe neprimerne informacijske tehnologije za zbiranje podatkov. Tako so se podatki o letih, ki imajo vrednosti 9999 ali 1900, verjetno pojavili zaradi tega, ker uporabljeni računalniški sistemi niso bili pripravljene za vstop v leto 2000.

Pogojno vključeni podatki

V bazi se nahajajo tudi pogojno vključeni podatki, ki pa jih sorazmerno enostavno odkriti. To so podatki, ki se nanašajo le na določeno vrsto nepremičnin, zato so neprazni vnosi le pri tistih zapisih v bazi, ki se nanašajo na to vrsto nepremičnin. Nekaj primerov pogojno vključenih podatkov smo omenili že na koncu poglavja 4.1, zato jih tukaj ne bomo ponavljali.

Nedoslednost podatkov v bazi

Baza podatkov DURS-a ima tudi to značilnost, da je zbirna baza nepremičninskih transakcij za celo Slovenijo. Značilnost zbirnih baz podatkov je, da imajo poleg treh zgoraj omenjenih pomanjkljivosti še dodatno pomanjkljivost, to je nedoslednost podatkov v njih. To pomeni, da imajo pri različnih zapisih podatki o isti spremenljivki, ki so bili vneseni v bazo podatkov iz različnih virov, različne vrednosti za opis enakega pojava (Inmon 1996, str. 73-76).

Nedoslednost podatkov je bila prisotna tudi v bazi nepremičninskih transakcij DURS-a. Tako so se na primer pri spremenljivkah *OGREVANJE* in *PODSTRESNO*, ki imata lahko le vrednosti 0 ali 1, pogosto pojavile tudi vrednosti 3 ali 4. Podobna nedoslednost se je pojavila pri spremenljivki *LEGA*, kjer so bile včasih namesto vrednosti *S*, *V*, *J* in *Z* zapisane vrednosti od 1 do 4. Nedoslednosti so se pojavile tudi pri spremenljivki *ST_VLOZ*, kjer so bile vrednosti spremenljivke za opis stanja "ni vložka" različne: "-", "0", "NI", "JE NI" ali prazno mesto. Poleg tega je bil v nekaterih zapisih pri več spremenljivkah nastavljen datumski tip spremenljivke, kar je povzročilo izgubo informacij o vrednosti teh spremenljivk za dani zapis.

Zaradi vseh zgoraj omenjenih napak v podatkih smo se odločili, da bomo podatke v bazi nepremičninskih transakcij pred nadaljnjo analizo prečistili. Seveda smo iz baze podatkov odstranili le tiste nepravilnosti, ki smo jih uspeli odkriti.

4.2.2 Ravnanje z manjkajočimi in nepravilnimi podatki

Problem manjkajočih podatkov je, da lahko zaradi njih izgubimo pomembno informacijo o opravljeni transakciji z nepremičnino. Pri nepravilnih podatkih je problem v tem, da je informacija o opravljeni nepremičnini bodisi napačna, bodisi neuporabna. V literaturi (Hastie et al. 2001, str. 293-294, Knight et al. 1998) je mogoče najti različne prijeme za ravnanje z manjkajočimi in nepravilnimi podatki, od najenostavnejših (neupoštevanje zapisov z manjkajočimi podatki, nadomeščanje manjkajočih podatkov pri posameznih spremenljivkah s povprečjem te spremenljivke) do bolj zapletenih (izračun vrednosti za manjkajoče podatke z regresijsko enačbo ali z Gibbsovim vzorčenjem).

V našem primeru smo uporabili enostavnejše prijeme ravnanja z manjkajočimi in nepravilnimi podatki, in sicer:

- Manjkajočih podatkov nismo nadomeščali, razen pri spremenljivki *NADSTROPJE*. Pri tej spremenljivki smo manjkajoče podatke nadomestili z vrednostjo 2, ki predstavlja povprečje in mediano te spremenljivke (glej tudi komentar v poglavju 4.2.1). S tem nismo bistveno vplivali na dobljene rezultate, saj spremenljivka *NADSTROPJE* ni bila značilna pri nobenem referenčnem modelu za vrednotenje stanovanj – glej poglavje 5. To smo storili zato, da nismo po nepotrebnem zmanjševali števila vzorcev za izgradnjo modelov zaradi neznačilne spremenljivke.
- Pri spremenljivkah *OGREVANJE* in *PODSTRESNO* smo privzeli vrednost 1 v vseh primerih, ko je bila vrednost teh dveh spremenljivk v bazi različna od 0.
- Pri spremenljivki *ST_VLOZ* smo vse vrednosti, ki predstavljajo opis stanja "ni vložka", nadomestili z vrednostjo 0.
- Pri spremenljivki *LEGA* smo vse vnose, ki niso bili enaki *S*, *V*, *J* ali *Z*, izbrisali, ker iz njih ni bilo mogoče razbrati v kateri smeri neba leži nepremičnina.

- Pri spremenljivki *NADSTROPJE* smo izbrisali vse vnose, ki so bili večji od 19. Razlog zato je, da so bili napačni vnosi navadno večkratnik števila 10 in so se začeli pri 20. nadstropju. Tako je bilo v bazi podatkov zapisov z 20. nadstropjem bistveno več kakor za zapisov z nadstropji od 15 do 19. To seveda ni logično, ker je v Sloveniji zelo malo tako visokih stavb, poleg tega ni racionalnega ekonomskega razloga zato, da bi ljudje kupovali več stanovanj v 20. nadstropju kakor v nadstropjih od 15 do 19.
- Pri spremenljivki *ODDALJENOST* smo izbrisali vse vnose, ki so bili večji od 49. Razlog zato je, da so bili napačni vnosi navadno večkratnik števila 100 in so se začeli nekako pri oddaljenosti 50 km. Tako je bilo podobno kakor pri spremenljivki *NADSTROPJE* tudi pri tej spremenljivki v bazi podatkov število zapisov z oddaljenostjo 50 km bistveno več kakor zapisov z oddaljenostjo od 20 do 49 km.
- Pri občini Koper smo pri spremenljivki *LETO_IZGR* izbrisali vse vnose, ki so imeli vrednost 9999, pri spremenljivki *LETO_NAKUPA* v letu 1999 pa smo izbrisali vse vnose, ki so imeli vrednost 1900.
- Pri izgradnji modelov za množično vrednotenje stanovanj po tržni metodi, smo pri stanovanjskih nepremičninah izbrisali vse vnose, ki so imeli nenormalno veliko (več kot 1000 m²) ali nenormalno majhno (manj kot 5 m²) stanovanjsko površino ter vse vnose, pri katerih je bila cena kvadratnega metra stanovanja ekstremno nizka (manj kakor 1000 SIT/m²) ali ekstremno visoka (več kakor 700000 SIT/m²).

4.3 Analiza značilnosti podatkov v bazi nepremičninskih transakcij

4.3.1 Izračun osnovnih statistik za bazo nepremičninskih transakcij

Najbolj pogosti informaciji, ki nas zanimata v zvezi z določeno nepremičnino, sta navadno njena izmera in cena za kvadratni meter. Izmera nepremičnine je bila ena izmed osnovnih spremenljivk v bazi nepremičninskih transakcij DURS-a z istoimensko oznako: *IZMERA*. Ceno na kvadratni meter smo izračunali kot kvocient vrednosti nepremičnine in njene izmere:

$$CENA_NA_m^2 [SIT/m^2] = \frac{vrednost\ nepremicnine}{IZMERA} \quad (27).$$

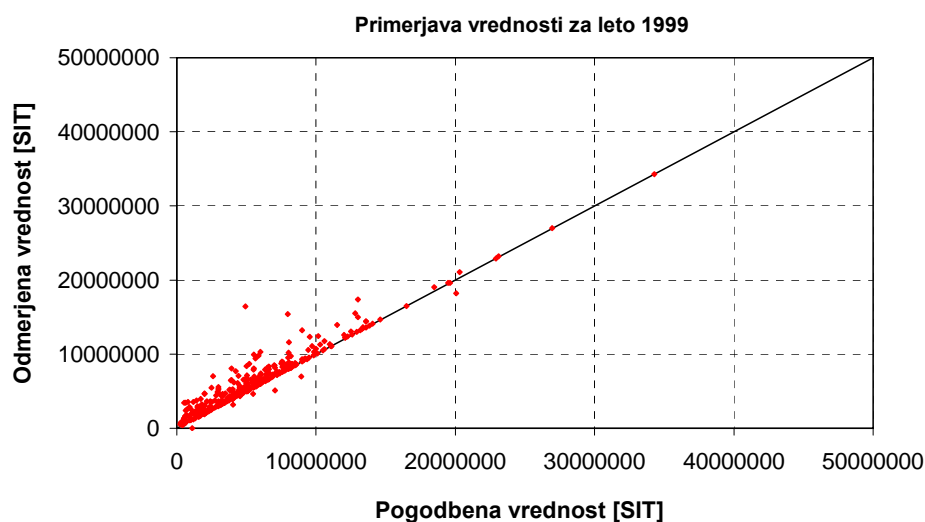
Pri upoštevanju vrednosti nepremičnine v zgornji enačbi smo naleteli na problem. V bazi DURS-a sta se namreč nahajala dva podatka o vrednosti nepremičnine: pogodbeni vrednost (*VREDNOST_POG*) in odmerjena vrednost (*VREDNOST_ODM*) – glej poglavje 4.1. Opazili smo, da sta se ti dve vrednosti pogosto razlikovali, predvsem pri nepremičninah z relativno nižjo vrednostjo, ki jih je bilo v tej bazi podatkov največ. Na Sl. 8 do Sl. 11 so prikazane primerjave pogodbenih in odmerjenih vrednosti za stanovanja (šifra nepremičnine = 22) v letih od 1999 do 2002. Razmerja so zelo podobna tudi pri drugih vrstah nepremičnin, vrednosti se razlikujejo le glede na red velikosti, značilen za posamezne vrste nepremičnin.

Ker je bila z dodatno cenitvijo odmerjena cena v veliki večini primerov višja od pogodbene vrednosti smo sklepali, da je bila prijavljena pogodbeni vrednost teh nepremičnin prenizka. Eden izmed razlogov zato je bil lahko tudi ta, da sta se pogodbeni stranki želeli izogniti prevelikemu plačilu davka od prometa z nepremičninami. Povezave med razliko pogodbene in odmerjene vrednosti ter načinom pridobitve nepremičnine (*NACIN_PRIDOB*) nismo opazili. Ker smo bili mnenja, da je v našem primeru najprimernejši indikator tržne vrednosti

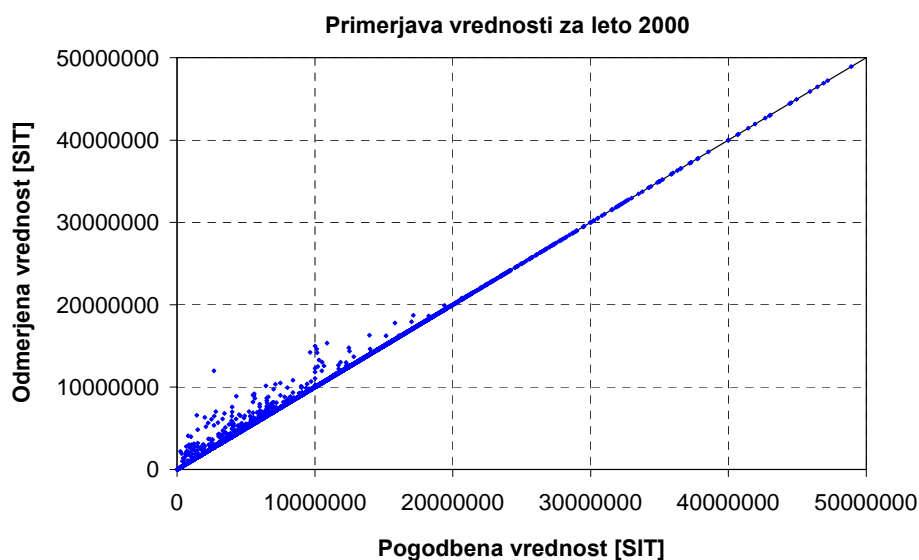
nepremičnine višja od obeh vrednosti, smo pri izračunu cene kvadratnega metra nepremičnine v enačbi (27) vedno upoštevali višjo od obeh vrednosti:

$$\text{vrednost nepremicnine} = \max\{VREDNOST_POG, VREDNOST_ODM\} \quad (28).$$

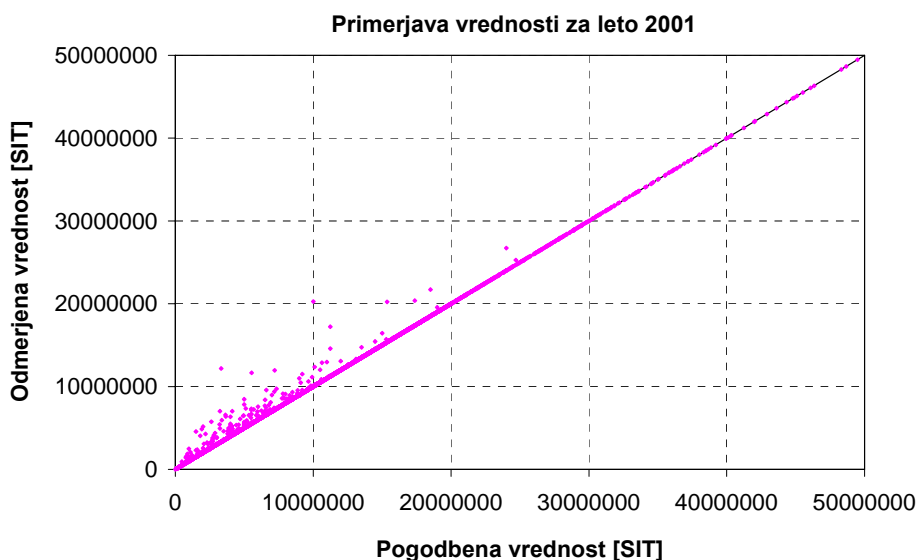
Za spremenljivki *IZMERA* in *CENA_NA_m²* smo izračunali srednjo vrednost in mediano. Obe značilki sta meri centralne tendence porazdelitve vzorčnih točk za določeno spremenljivko. Kljub temu, da za prikaz centralne tendence največkrat uporabljamo kar vzorčno srednjo vrednost, smo se odločili, da prikažemo tudi mediane. Razlog za to je, da je bilo v bazi nepremičninskih transakcij DURS-a nekaj zapisov, pri katerih so se pojavile izredno velike vrednosti za spremenljivki izmera in cena za kvadratni meter – glej tabeli 3 in 4 na strani 52. Ker bi lahko te vrednosti premočno vplivale na vzorčno povprečje, je v tem primeru mediana primernejša mera centralne tendence dotičnih spremenljivk.



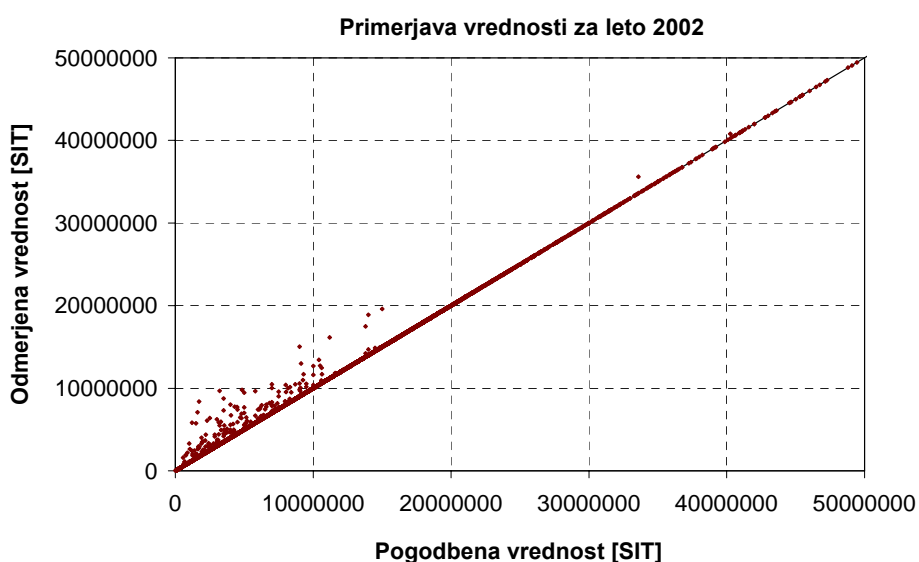
Sl. 8: Primerjava pogodbenih in odmerjenih vrednosti stanovanj za leto 1999



Sl. 9: Primerjava pogodbenih in odmerjenih vrednosti stanovanj za leto 2000



Sl. 10: Primerjava pogodbenih in odmerjenih vrednosti stanovanj za leto 2001



Sl. 11: Primerjava pogodbenih in odmerjenih vrednosti stanovanj za leto 2002

Če bi želeli grafično prikazati povprečja in mediane spremenljivk *IZMERA* in *CENA_NA_m²* za vsako občino posebej, bi bili diagrami zelo nepregledni, izračunani značilki pa ne bi bili zanesljivi, ker je bilo v opazovanem obdobju v nekaterih manjših občinah opravljenih zelo malo transakcij z nepremičninami. Poleg tega obstajajo v Sloveniji sorazmerno zaključena področja, kjer prevladujejo podobne cene nepremičnin. Zaradi teh razlogov smo se odločili, da za izračun povprečja in mediane spremenljivk *IZMERA* in *CENA_NA_m²* ter za izgradnjo modelov za vrednotenje stanovanj v poglavju 5 razporedimo vse občine v deset zaključenih območij. Pri tem smo si pomagali z razdelitvijo slovenskih občin v statistične regije in z razdelitvijo občin v skupine glede na vrednost spremenljivke *CENA_NA_m²* za stanovanja. Kombinirani pristop za razdelitev občin v zaključena območja smo uporabili zaradi naslednjih dveh razlogov:

- Če bi upoštevali le razdelitev občin v statistične regije, bi s tem v ista območja združili občine, ki z ekonomskega stališča ne spadajo skupaj. Tako bi na primer Ljubljana kot občina z najbolj likvidnim trgom nepremičnin padla v isto območje (osrednje-

slovenska regija) kakor na primer Borovnica ali Horjul, kjer je nepremičninski trg zelo nerazvit. Podobnih primerov bi lahko našli še nekaj (obalno-kraška regija na primer združuje tako s stališča kupcev nepremičnin zelo zelene obalne občine, kakor tudi tržno manj zanimive kraške občine).

- Če bi upoštevali le razdelitev občin glede na vrednosti spremenljivke $CENA_NA_m^2$, bi bil problem v tem, da bi imeli nekaj območij z visokimi cenami, ki bi bila podobna otokom v velikem območju, kjer prevladujejo nižje cene. Pri tem bi bile v isto območje združene občine iz geografsko medsebojno zelo oddaljenih območij.

Za razdelitev občin v statistične regije smo upoštevali standardno statistično razvrstitev občin, ki je veljala do leta 2002 (Občine RS 2003a, 2003b, 2003c, Statistične regije 2003a, 2003b, 2003c). Razdelitev občin v cenovne skupine smo izvedli s pomočjo hierarhičnega razvrščanja občin v skupine glede na povprečno realizirano vrednost spremenljivke $CENA_NA_m^2$ za stanovanja (šifra nepremičnin = 22). Za razvrščanje občin glede na ceno stanovanj smo se odločili predvsem zato, ker je bilo v opazovanem obdobju največ nepremičninskih transakcij izvedenih s stanovanji, in ker smo isto razdelitev občin v območja uporabili tudi v poglavju 5 za izgradnjo modelov za množično vrednotenje stanovanjskih nepremičnin po tržni metodi.

V tabeli P2-1 v prilogi 2 so prikazane povprečne vrednosti spremenljivke $CENA_NA_m^2$ za posamezne občine za leta od 1999 do 2002 in še posebej za vsa leta skupaj. Pri izračunu povprečnih vrednosti spremenljivke $CENA_NA_m^2$ nismo upoštevali vnosov, ki so imeli nenormalno veliko (več kot 1000 m²) ali nenormalno majhno (manj kot 5 m²) stanovanjsko površino, ter vnosov, pri katerih je bila cena kvadratnega metra stanovanja ekstremno nizka (manj kot 1000 SIT/m²) – glej tudi poglavje 4.2.2.

Podatki v tabeli P2-1 kažejo, da v štirinajstih občinah ni bilo opravljene nobene nepremičninske transakcije s stanovanji med letoma 1999 in 2002: 10 – Tišina, 18 – Destričnik, 47 – Kobilje, 56 – Kuzma, 143 – Zavrč, 149 – Bistrica ob Sotli, 150 – Bloke, 154 – Dobje, 161 – Hodoš, 163 – Jezersko, 170 – Mirna Peč, 176 – Razkrižje, 187 – Velika Polana in 191 – Žetale. Številka pred imenom občine pomeni šifro občine (Občine RS 2003a). To so večinoma manjše občine, ki so z izjemo občine Jezersko tudi manj razvite po razvrstitvi Rovana in Sambta (2003). Te občine smo razdelili v posamezna območja le z upoštevanjem razdelitve občin v statistične regije (Statistične regije 2003a). V tabeli P2-1 je mogoče videti, da je kar nekaj občin, v katerih nepremičninskih transakcij s stanovanji ni bilo v vseh letih. Zato smo za združevanje občin v cenovne skupine upoštevali le povprečne vrednosti spremenljivke $CENA_NA_m^2$ za vsa leta skupaj. Občine smo razdelili v cenovne skupine z Wardovo metodo hierarhičnega združevanja v skupine (glej poglavje 3.2). Za združevanje občin v skupine smo uporabili programski paket *SPSS 10.0*.

Dendrogram združevanja občin v skupine je prikazan na Sl. P2-1 v prilogi 2. Število cenovnih skupin, ocenjeno na osnovi dendrograma, je enako 5. Osnovne statistične značilke spremenljivke $CENA_NA_m^2$ za vsa leta skupaj in za posamezne skupine so prikazane v tabeli P2-2 v prilogi 2. Iz podatkov v tabeli P2-2 vidimo, da so skupine sorazmerno homogene, saj je standardna deviacija za posamezne skupine majhna v primerjavi z njihovim povprečjem. Iz dendrograma na Sl. P2-1 in tabele P2-1 v prilogi 2 vidimo, da v najvišjo cenovno skupino po pričakovanju sodijo Ljubljana (šifra občine 61) z nekaterimi bližnjimi gorenjskimi občinami (23 – Domžale, 72 – Mengeš in 186 – Trzin) ter priljubljeni letoviški kraji (40 – Izola, 53 –

Kranjska gora in 90 – Piran). V najnižji cenovni skupini se nahajajo manj razvite občine po razvrstitvi Rovana in Sambta (2003).

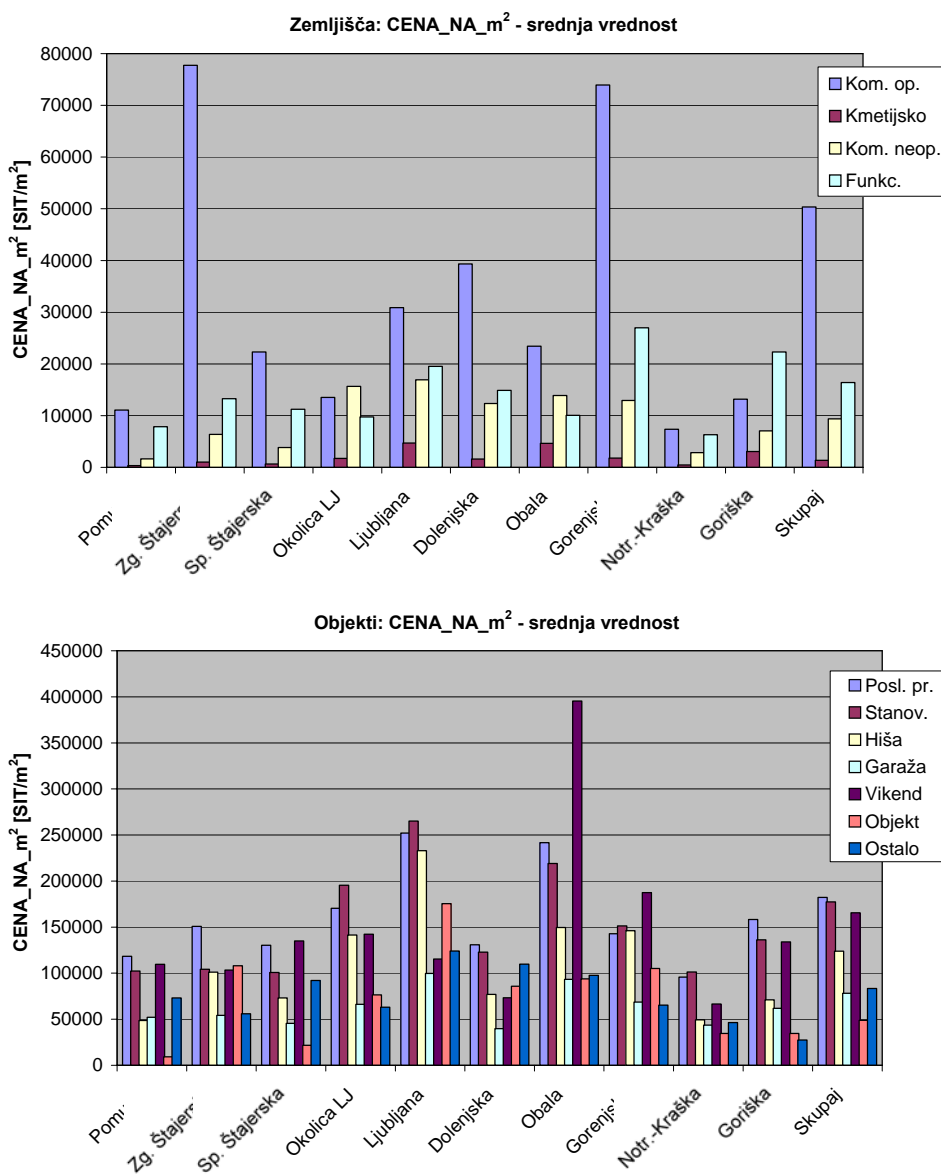
Glede na razdelitev občin v statistične regije (Statistične regije 2003a) in glede na razdelitev občin v cenovne skupine, ki smo jo dobili s hierarhičnim razvrščanjem (glej prilogo 2), smo občine razdelili v deset zaključenih in medsebojno neprekrivajočih se območij:

1. **Območje pomurske regije**, v katerem so občine, ki spadajo v pomursko regijo tudi po statistični delitvi.
2. **Območje zgornje Štajerske in Koroške**, v katerem so občine, ki spadajo po statistični delitvi v podravsko in koroško statistično regijo.
3. **Območje spodnje Štajerske in Posavja**, v katerem so občine, ki spadajo po statistični delitvi v savinjsko, zasavsko in spodnjeposavsko statistično regijo.
4. **Območje okolice Ljubljane**, v katerem so vse občine, ki spadajo po statistični delitvi v osrednjeslovensko statistično regijo razen Ljubljane.
5. **Območje mesta Ljubljane**, v katerega spada le mestna občina Ljubljana.
6. **Območje dolenske regije**, v katerem so vse občine, ki spadajo po statistični delitvi v jugovzhodno Slovenijo.
7. **Območje slovenske obale**, v katerem so občine: Koper, Izola in Piran.
8. **Območje Gorenjske**, v katerem so vse občine, ki spadajo po statistični delitvi v gorenjsko statistično regijo.
9. **Območje Notranjske in Krasa**, v katerem so vse občine, ki spadajo po statistični delitvi v notranjsko-kraško ali obalno-kraško statistično regijo razen Kopa, Izole in Pirana.
10. **Območje goriške regije**, v katerem so vse občine, ki spadajo po statistični delitvi v goriško statistično regijo.

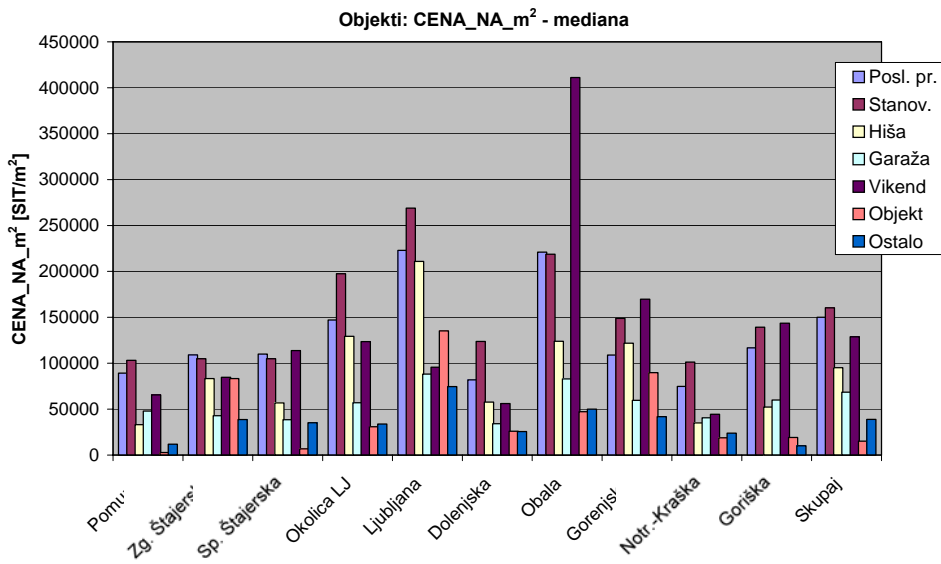
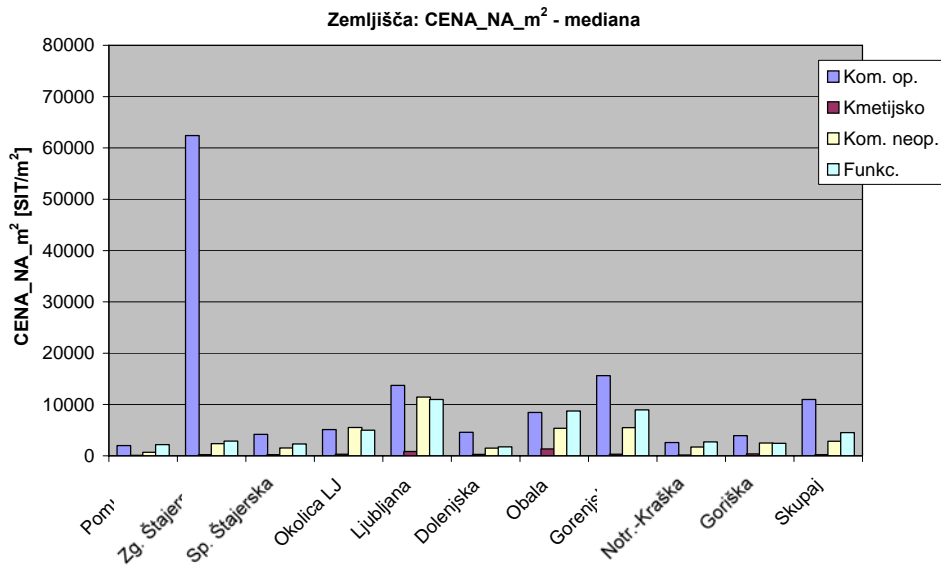
V grafih na Sl. 12 in Sl. 13 so za posamezne vrste nepremičnin prikazane srednje vrednosti in mediane za povprečne vrednosti spremenljivke $CENA_NA_m^2$. Povprečne vrednosti spremenljivke $CENA_NA_m^2$ smo izračunali posebej za vsako izmed zgoraj naštetih območij za vsa leta od 1999 do 2002. V grafih na slikah Sl. 14 in Sl. 15 so za posamezne vrste nepremičnin prikazane srednje vrednosti in mediane za povprečne vrednosti spremenljivke $IZMERA$. Tudi srednje vrednosti spremenljivke $IZMERA$ smo izračunali posebej za vsako izmed zgoraj naštetih območij in za vsa leta štiri leta skupaj. Za primerjavo so v tabelah 3 in 4 prikazane maksimalne vrednosti spremenljivk $CENA_NA_m^2$ in $IZMERA$ za posamezna območja in posamezne vrste nepremičnin.

Na Sl. 12 do Sl. 15 ter v tabelah 3 in 4 vidimo, da so imele ekstremne vrednosti spremenljivk $CENA_NA_m^2$ in $IZMERA$ velik vpliv na izračun povprečja, ker se mediane včasih močno razlikujejo od povprečja. Na Sl. 14 in Sl. 15 vidimo, da so bili vnosi za spremenljivko $IZMERA$ nedosledni, saj je bila pri vikendih in ostalih objektih za izmero včasih upoštevana velikost objekta, včasih pa velikost zemljišča. Na Sl. 12 in Sl. 13 je mogoče ugotoviti območja, kjer je večje povpraševanje po nepremičninah povzročilo višje cene kvadratnega metra nepremičnin: slovenska obala ter Ljubljana z nekaterimi sosednjimi občinami. Cene komunalno opremljenih zemljišč so bile v skladu s pričakovanji bistveno višje kakor cene

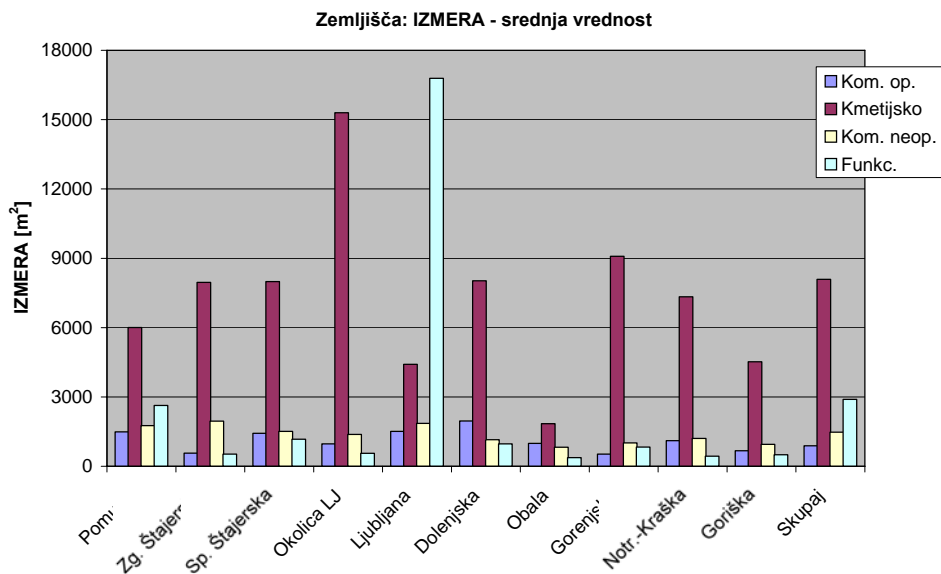
ostalnih zemljišč. Tudi sicer so bile cene zemljišč na bolj zelenih lokacijah višje. Poleg tega vidimo na Sl. 14 in Sl. 15, da so bila zemljišča načeloma manjša v območjih z dražjimi nepremičninami. Izjema je sicer okolica Ljubljane, vendar v to okolico spadajo tudi občine z velikimi kmetijskimi površinami, ki zvišujejo vrednosti povprečja in mediane. Zato lahko sklepamo, da so zemljišča na zelenih lokacijah omejena dobrina, kar potrjuje splošno načelo, da velik del razlike v cenah stavbnih nepremičnin predstavlja prav lokacija same nepremičnine. Očitno je tudi, da je bila v letih od 1999 do 2002 slovenska obala za marsikoga priljubljena počitniška destinacija, saj so bile cene vikendov na tem območju najvišje.



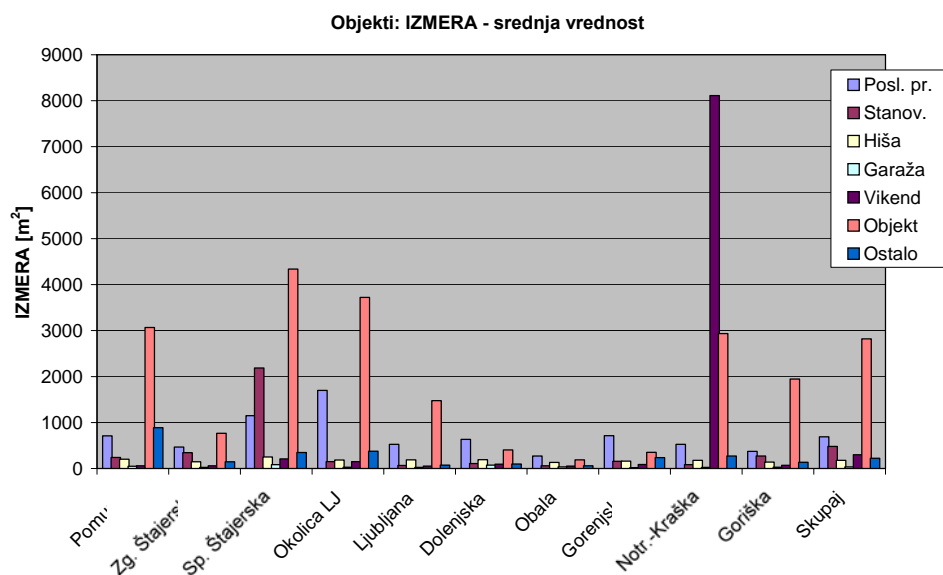
Sl. 12: Srednja vrednost spremenljivke $CENA_NA_m^2$ v $[SIT/m^2]$



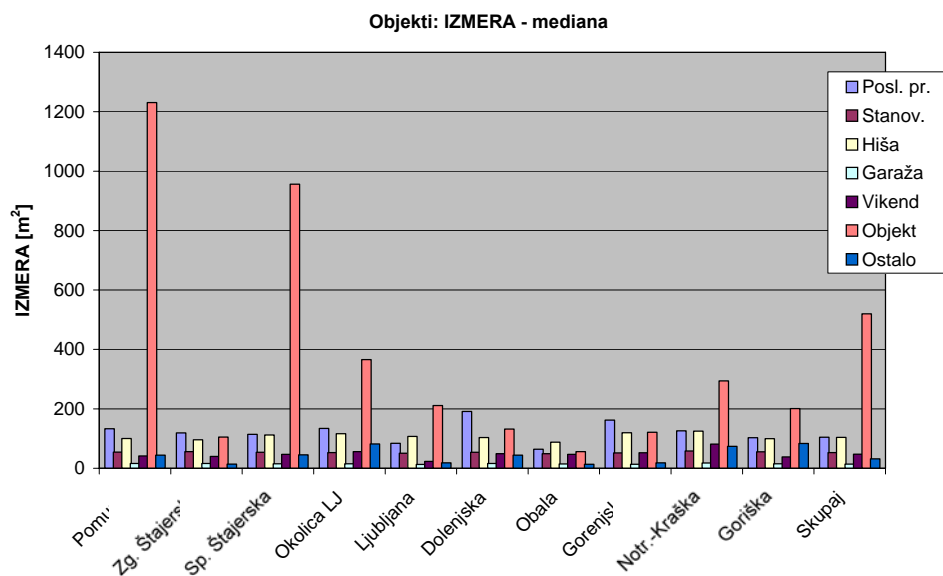
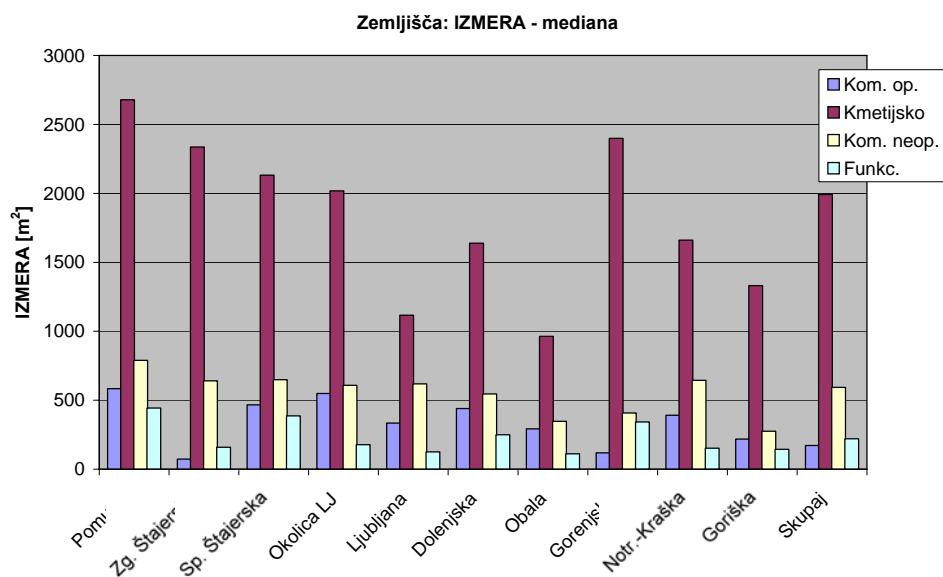
Sl. 13: Mediana spremenljivke CENA_NA_m² v [SIT/m²]



Sl. 14: Srednja vrednost spremenljivke IZMERA v [m²]



Sl. 14: Srednja vrednost spremenljivke IZMERA v $[m^2]$ (nadaljevanje)



Sl. 15: Mediana spremenljivke IZMERA v $[m^2]$

Tabela 3: Maksimalne vrednosti spremenljivke $CENA_NA_m^2$ v [SIT/m^2]

	Kom. op. zemljišče	Kmet. zemljišče	Kom. neop. zemljišče	Funkc. zemljišče	Poslovni prostor	Stanovanja	Hiša	Garaža	Vikend	Objekt z zemlj.	Ostalo
Pomurje	628141	131086	108889	97107	557621	334835	344710	113636	350871	232794	314376
Zg. Štaj.	2868009	158107	502900	973005	10393939	808459	3075356	1159205	539319	2691833	2361018
Sp. Štaj.	6076890	151163	537523	642962	1678600	1457590	1440000	199641	436933	421053	1092050
Okolica LJ	805000	330022	9280340	166019	1945778	5434783	982609	389093	732794	1569458	718750
Ljubljana	747826	321715	239544	803801	2186354	3608563	1500000	660776	308043	717949	1826771
Dolenjska	9200000	316667	4104086	1850000	3198003	512630	459030	125373	291667	3186180	3225000
Obala	450000	389764	334545	111360	4362416	1435302	1732323	582666	2803588	549294	643767
Gorenjska	1874157	306451	603529	485437	1270072	621483	937310	266667	732407	578763	580122
Notr./Kras	333611	25000	100714	73965	427132	311361	343348	207088	252452	289734	187316
Goriška	230545	1638500	268757	3810409	4075000	445618	401416	148197	230280	260674	338234
Skupaj	9200000	1638500	9280340	3810409	10393939	5434783	3075356	1159205	2803588	3186180	3225000

Tabela 4: Maksimalne vrednosti spremenljivke $IZMERA$ v [m^2]

	Kom. op. zemljišče	Kmet. zemljišče	Kom. neop. zemljišče	Funkc. zemljišče	Poslovni prostor	Stanovanja	Hiša	Garaža	Vikend	Objekt z zemlj.	Ostalo
Pomurje	33869	928454	74735	249089	19727	43009	3000	1097	296	63657	20173
Zg. Štaj.	205030	2216002	970552	19874	43448	1237531	16471	1605	1422	26738	7155
Sp. Štaj.	114968	1934601	93807	40595	214764	7991000	10343	14025	6366	170182	11872
Okolica LJ	54879	20000000	82999	25686	593992	110000	31741	1609	8471	131855	8000
Ljubljana	131519	76382	75571	9922000	197060	20616	50492	11312	202	37594	5943
Dolenjska	283368	4073889	52067	70097	32415	7638	15182	4197	956	6729	1455
Obala	26355	30000	25743	12580	9242	7405	7567	913	490	2525	2186
Gorenjska	29641	525295	39382	43549	51031	299842	2126	187	2584	18254	3543
Notr./Kras	25361	300000	55440	15593	14022	7428	2107	150	174361	75954	2373
Goriška	21536	738890	32650	24489	14740	66200	2884	345	1272	94837	1776
Skupaj	283368	20000000	970552	9922000	593992	7991000	50492	14025	174361	170182	20173

4.3.2 Značilnosti in vzorci, odkriti z uporabo povezovalnih pravil

Povezovalna pravila med spremenljivkami, s katerimi so opisane nepremičninske transakcije v bazi podatkov DURS-a, smo iskali z apriori algoritmom, ki je vgrajen v programsko okolje *WEKA 3-2*. Uporabljeni apriori algoritem lahko išče le povezovalna pravila za kvalitativne spremenljivke, v bazi nepremičninskih transakcij pa je kar nekaj spremenljivk numeričnih (na primer: *VREDNOST_POG*, *LETO_IZGR* itn.) ali tekstovnih (na primer: *ST_VLOZ*, *PARCELA* itn.). Zato smo za iskanje povezovalnih pravil vse numerične in tekstovne spremenljivke spremenili v kvalitativne. Kvalitativne spremenljivke, ki so ekvivalentne originalnim numeričnim in tekstovnim spremenljivkam, so opisane v tabeli P3-1 v prilogi P3. Povezovalna pravila smo iskali ločeno za transakcije z zemljišči (nepremičnine od 11 do 14) in za transakcije z objekti (nepremičnine od 21 do 27), ker so imele nekatere spremenljivke drugačen velikostni razred za zemljišča in objekte – glej tabelo P3-1 v prilogi 3. V nadaljevanju tega poglavja so prikazana tipična povezovalna pravila za zemljišča in objekte.

Povezovalna pravila za zemljišča (nepremičnine od 11 do 14)

Zgornja meja za parameter podpore: $s_{max} = 1,0$

Če smo izbrali zgornjo mejo za parameter podpore s enako 1,0, nismo dobili uporabnih rezultatov, ker je apriori algoritem izvedel pravila le za spremenljivki *ZAP_POG* in

ZAP_NEP, ki imata zalogo vrednosti {0,1} – glej tabelo P3-1 v prilogi 3. Povezave med tema dvema spremenljivkama so trivialne in enake praktično za vse nepremičninske transakcije, ker pri teh dveh spremenljivkah ni manjkajočih podatkov. Da bi našli netrivialna povezovalna pravila, smo znižali zgornjo mejo za parameter podpore.

Zgornja meja za parameter podpore: $0,7 \leq s_{max} \leq 0,9$

V tem primeru je apriori algoritem našel netrivialna povezovalna pravila, ki se v veliki meri nanašajo na komunalno neopremljena zemljišča. Vsa najdena povezovalna pravila imajo zaupanje *c* enako 1.0 in imajo v splošnem naslednjo obliko.²⁰

ELEKTRIKA=0 VODOVOD=0 KANALIZACIJA=0 ==> TELEFON=0 - supp:(0,456) conf:(1)

PODSTRESNO=0 ELEKTRIKA=0 VODOVOD=0 OGREVANJE=0 KANALIZACIJA=0 ==> TELEFON=0
- supp:(0,456) conf:(1)

KLETNO=0 DVIGALO=0 VODOVOD=0 OGREVANJE=0 ==> TELEFON=0 -
- supp:(0,459) conf:(1)

Zgornja pravila nam povedo, da na nepremičnini, ki ni komunalno opremljena, ni telefonskega priključka. Ta pravila so sicer logična, a z uporabniškega vidika nepomembna, ker je jasno, da na nepremičnini brez vode, elektrike in kanalizacije tudi ni telefona. Zadnji dve pravili sta praktično nesmiselni, saj na zemljiščih brez objektov ne more biti podstrešja, kleti ali dvigala. Podobnih pravil je bilo še več in so posledica tega, da pri spremenljivkah *DVIGALO*, *KLETNO* in *PODSTRESNO* niso bila blokirana vnosna polja, ko so bili vneseni podatki o transakcijah z zemljišči. Manjkajoči vnosi bi bili namreč v tem primeru bolj smiselni od vnosov z vrednostjo "0". Čeprav so bila najdena nekatera netrivialna povezovalna pravila, smo se odločili, da bomo zgornjo mejo za parameter podpore še znižali.

Zgornja meja za parameter podpore: $s_{max} = 0,4$

V tem primeru se praktično vsa pravila, ki jih je odkril apriori algoritem, nanašajo na komunalno opremljena zemljišča (vrsta nepremičnine 11). Podpora tem povezovalnim pravilom je sorazmerno nizka in se v vseh primerih nahaja okoli vrednosti 0,05. Vendar to velja, če smo jo izračunali glede na celotno število transakcij z zemljišči, ki je enako 47530. Če bi parameter podpore omejili le na komunalno urejena zemljišča, bi njegova vrednost narasla na 0,28, kar je že vredno upoštevanja. Tipična pravila iz te množice imajo naslednjo obliko (parameter podpore je izračunan glede na število transakcij z vsemi vrstami zemljišč):

VRSTA_NEPR=11 VODOVOD=1 ==> ELEKTRIKA=1 - supp:(0,053) conf:(0,99)

VRSTA_NEPR=11 ELEKTRIKA=1 VODOVOD=1 ==> LETO_IZGR=2 -
- supp:(0,050) conf:(0,96)

ELEKTRIKA=1 VODOVOD=1 ==> VRSTA_NEPR=11 - supp:(0,053) conf:(0,95)

LETO_IZGR=2 VODOVOD=1 ==> VRSTA_NEPR=11 - supp:(0,051) conf:(0,95)

Zgornja pravila nam povedo, da v primerih komunalno opremljenih zemljišč nepremičnina ima elektriko, oziroma če zemljišče ima elektriko, potem je to zemljišče komunalno

²⁰ Oznaka *supp* pomeni podporo povezovalnemu pravilu za transakcije z zemljišči, oznaka *conf* pa njegovo zaupanje – za definiciji teh dveh parametrov glej poglavje 3.3.

opremljeno. Spremenljivka *LETO_IZG* verjetno predstavlja čas izvedbe izboljšav. Vrednost 2 spremenljivke *LETO_IZG* namreč pomeni, da je bil vnos neprazen in hkrati realna letnica.

Povezovalna pravila za objekte (nepremičnine od 21 do 27)

Če smo izbrali zgornjo mejo za parameter podpore enako 1,0, je apriori algoritem, podobno kakor pri transakcijah z zemljišči, spet odkril le trivialna pravila. Zato smo tudi v tem primeru znižali zgornjo mejo za podporo. Ker je apriori algoritem našel v tem primeru veliko različnih pravil, smo našo pozornost namenili le najznačilnejšim in najpomembnejšim.

Zgornja meja za parameter podpore: $s_{max} = 0,9$

V tem primeru je apriori algoritem odkril netrivialna pravila, ki so v splošnem značilna za komunalno opremljene objekte. Večina pravil v tem primeru govori o stopnji medsebojne povezanosti med različno komunalno opremo: če nepremičnina ima vodovod, ogrevanje in kanalizacijo, ima tudi elektriko itn. Pri nekaterih pravilih so se pojavili tudi podatki o lastnosti nepremičnine (kletno, dvigalo). Iz teh pravil je mogoče sklepati, da glede na sorazmerno visoko vrednost parametra podpore večina nepremičnin nima dvigala ali ne leži v kleti. Nekaj tipičnih pravil z zelo visoko stopnjo podpore za stavbe je naslednjih:

VODOVOD=1 OGREVANJE=1 KANALIZACIJA=1 ==> ELEKTRIKA=1 - supp:(0,684) conf:(1)

VODOVOD=1 ==> ELEKTRIKA=1 - supp:(0,800) conf:(1)

DVIGALO=0 VODOVOD=1 ==> ELEKTRIKA=1 - supp:(0,681) conf:(1)

KLETNO=0 OGREVANJE=1 ==> ELEKTRIKA=1 VODOVOD=1 - supp:(0,656) conf:(0,99)

Kljub netrivialnim pravilom smo zgornjo mejo za podporo še znižali, ker ta povezovalna pravila niso izražala nekih posebnosti pri opravljenih transakcijah z nepremičninami.

Zgornja meja za parameter podpore: $s_{max} = 0,8$

V tem primeru se je podobno kakor pri zgornji meji podpore 0,9 veliko pravil nanašalo na komunalno opremljenost objektov. Teh pravil nismo še enkrat pisali, ker so podobna kot prej. Kar nekaj pa je tudi bolj specifičnih pravil, ki imajo naslednjo obliko:

VRSTA_NEPR=22 TELEFON=1 ==> KANALIZACIJA=1 - supp:(0,406) conf:(0,99)

LETO_IZGR=4 TELEFON=1 ==> OGREVANJE=1 - supp:(0,429) conf:(0,98)

TELEFON=1 NACIN_PRIDOB=1 ==> KANALIZACIJA=1 - supp:(0,426) conf:(0,98)

Ta pravila beremo takole: če stanovanje ima telefon, ima tudi kanalizacijo, ali če je bila nepremičnina zgrajena med letoma 1960 in 2002 (*LETO_IZGR=4*) in ima telefon, ima tudi ogrevanje. Zadnje povezovalno pravilo nas lahko zavede v sklep, da se kupci nepremičnin najraje odločajo za komunalno opremljene objekte. Vendar takšno sklepanje nima realne osnove, saj je to pravilo predvsem posledica dejstva, da je nakup (*NACIN_PRIDO=1*) najpogostejši način pridobitve lastninskih pravic na objektih, ki so večinoma komunalno opremljene nepremičnine. Ker smo pričakovali, da bomo z zmanjševanjem zgornje meje za parameter podpore dobili še bolj zanimiva pravila, smo zgornjo mejo za podporo še znižali.

Zgornja meja za parameter podpore: $s_{max} = 0,7$

V tem primeru je apriori algoritem odkril dve vrsti povezovalnih pravil. Prva vrsta pravil se nanaša na komunalno opremljenost objektov in predstavlja nasprotje pravil, ki smo jih odkrili,

če je bila zgornja meja podpore enaka 0,9 ali 0,8. Tipična pravila glede komunalne opremljenosti iz te skupine so naslednja:

VODOVOD=0 OGREVANJE=0 KANALIZACIJA=0 ==> TELEFON=0 - supp:(0,177) conf:(1)

VODOVOD=0 KANALIZACIJA=0 ==> OGREVANJE=0 TELEFON=0 -
- supp:(0,177) conf:(0,97)

NADSTROPJE=1 VODOVOD=0 TELEFON=0 ==> KANALIZACIJA=0 -
- supp:(0,175) conf:(0,97)

Iz teh pravil na primer sledi, da nepremičnina nima telefona, če nima vodovoda, ogrevanja in kanalizacije. Tretje povezovalno pravilo iz te skupine kaže na to, da so vnosi z vrednostjo 0 za komunalno opremljenost dostikrat povezani s točno določenimi vrednostmi vnosov za ostale spremenljivke, pri katerih je sicer relativno manj nepraznih vnosov. Vendar je težko reči, ali je šlo v teh primerih za resnično značilnost nepremičnin ali le za slabe vnose komunalne opremljenosti stavbnih nepremičnin. V nadaljevanju bomo namreč videli, da je v primeru zmanjšanja zgornje meje za podporo na vrednost 0,3 ali 0,2 velik delež podobnih pravil vezan na leto 1999, ki je bilo glede kvalitete in popolnosti vnosov zelo slabo.

Druga vrsta povezovalnih pravil povzema specifične značilnosti stanovanjskih nepremičnin. Ta povezovalna pravila govorijo o lastnostih tipičnega stanovanja, ki je bilo predmet nepremičninske transakcije med letoma 1999 in 2002. Značilne oblike teh pravila so:²¹

IZMERA=1 NADSTROPJE=2 ==> VRSTA_NEPR=22 -
- supp:(0,253) supp₂₂:(0,477) conf:(0,93)

LETO_IZGR=4 NADSTROPJE=2 ==> VRSTA_NEPR=22 -
- supp:(0,227) supp₂₂:(0,428) conf:(0,9)

NADSTROPJE=2 ==> VRSTA_NEPR=22 - supp:(0,321) supp₂₂:(0,605) conf:(0,89)

VRSTA_NEPR=22 LETO_IZGR=4 ==> IZMERA=1 -
- supp:(0,295) supp₂₂:(0,558) conf:(0,81)

VRSTA_NEPR=22 NADSTROPJE=2 ==> IZMERA=1 -
- supp:(0,253) supp₂₂:(0,477) conf:(0,79)

Ta pravila nam povedo, da je tipična nepremičnina, ki je v drugem nadstropju in ima izmero manjšo od 70 m² ali je bila zgrajena med leti 1960 in 2002 (glej prilogo P3), stanovanje, in obratno, da ima tipično stanovanje izmero manjšo od 70 m². Čeprav je podpora pri teh pravilih glede na vse objekte razmeroma nizka, se bistveno zviša, če se omejimo le na stanovanja – glej podporo *supp₂₂*. Vendar je treba parameter *supp₂₂* jemati z rezervo, saj ni nujno, da bi apriori algoritem našel ista pravila, če bi iskanje omejili le na stanovanja. Posebej zanimive so tri značilnosti najdenih povezovalnih pravil. Prva je ta, da iz drugega nadstropja večinoma sledi nepremičnina stanovanje. Samo iz teh pravil še ne bi mogli sklepati, da se največji delež stanovanj nahaja ravno v drugem nadstropju. Če pa obrnemo pravilo *nadstrop=2 ==> vrst_nep=22*, tako da bi iz stanovanja sledilo nadstropje, potem *supp₂₂* v bistvu predstavlja zaupanje za obrnjeno pravilo. Iz tega bi bilo mogoče sklepati, da se večina

²¹ Oznaka *supp₂₂* pomeni podporo glede na število nepremičninskih transakcij s stanovanji.

stanovanj nahaja v drugem nadstropju. Problem tega sklepanja je, da ne obstaja racionalna ekonomska razlaga, zakaj je bilo največ transakcij opravljenih s stanovanji v drugem nadstropju. To je verjetno tudi posledica napačnih ali privzetih vnosov pri polnjenju baze nepremičninskih transakcij. Druga značilnost najdenih povezovalnih pravil je, da je večina stanovanj manjših od 70 m². Ta podatek je logičen, saj je velik delež stanovanj srednje velikih (glej Sl. 15 in članek Matejčičeve (2003)), poleg tega so manjša stanovanja v splošnem tudi bolj likvidna zaradi nižjih stroškov nakupa. Tretja značilnost najdenih povezovalnih pravil je, da je bilo veliko stanovanj izgrajenih med letoma 1960 in 2002. Tudi to je logično, saj je bila v tem obdobju gradnja blokovskih stanovanjskih naselij zelo množična.

Zgornja meja za parameter podpore: $0,2 \leq s_{max} \leq 0,3$

Tudi v tem primeru je apriori algoritem odkril dve vrsti pravil. Prva vrsta pravil se nanaša na komunalno opremljenost in je nasprotje pravil, ki smo jih odkrili, ko je bila zgornja meja parametra podpore enaka 0,9 ali 0,8. Iz najdenih pravil sledi, da odsotnost dveh ali več komunalnih izboljšav pogojuje tudi odsotnost ostalih komunalnih izboljšav. Ker je mnogo teh pravil vezanih na leto 1999 sklepamo, da je šlo v tem primeru zaradi na splošno slabe kvalitete vnosov v letu 1999 (glej poglavje 4.2.1) prej za slabe vnose, kakor pa za dejansko komunalno neopremljenost nepremičnin. Tipična pravila za komunalno (ne)opremljenost so:

LETO=1999 ELEKTRIKA=0 VODOVOD=0 KANALIZACIJA=0 ==> OGREVANJE=0 -
- supp: (0,079) conf: (1)

LETO=1999 ELEKTRIKA=0 OGREVANJE=0 KANALIZACIJA=0 ==> VODOVOD=0 -
- supp: (0,079) conf: (1)

ELEKTRIKA=0 OGREVANJE=0 KANALIZACIJA=0 ==> VODOVOD=0 -
- supp: (0,142) conf: (0,99)

Druga vrsta pravil se nanaša na komunalno opremljenost garaž (vrsta nepremičnin 24). Iz teh pravil je mogoče razbrati, da tipična garaža ni opremljena z vodovodom ali ogrevanjem.²²

VRSTA_NEPR=24 VODOVOD=0 KANALIZACIJA=0 ==> OGREVANJE=0 -
- supp: (0,057) supp₂₄: (0,856) conf: (1)

VRSTA_NEPR=24 KANALIZACIJA=0 ==> OGREVANJE=0 -
- supp: (0,059) supp₂₄: (0,891) conf: (1)

VRSTA_NEPR=24 ==> VODOVOD=0 - supp: (0,059) supp₂₄: (0,886) conf: (0,89)

Čeprav je podpora tem pravilom glede na celotno bazo nizka, se bistveno zviša, če se omejimo le na transakcije z garažami – glej podporo *supp*₂₄. V tem primeru sklepamo, da pri komunalni opremljenosti garaž ne gre za slabe vnose, ampak za tipično značilnost garaž, to je pokrit prostor za vozilo brez vodovoda ali ogrevanja.

Povzetek povezovalnih pravil

Največ najdenih povezovalnih pravil se nanaša na komunalno opremljenost. V splošnem imajo naslednjo obliko: prisotnost dveh ali več vrst komunalnih izboljšav pogojuje prisotnost tudi ostalih komunalnih izboljšav. In obratno: odsotnost dveh ali več vrst komunalnih

²² Oznaka *supp*₂₄ pomeni podporo glede na število nepremičninskih transakcij z garažami.

izboljšav pogojuje odsotnost tudi ostalih komunalnih izboljšav. Nekaj povezovalnih pravil pa se nanaša na specifične vrste nepremičnin: stanovanja in garaže. Povezovalna pravila za stanovanja kažejo na to, da je bila večina stanovanj zgrajenih med letoma 1960 in 2002 in da so stanovanja večinoma manjša od 70 m². Povezovalna pravila za garaže kažejo na to, da večina garaž ni opremljenih z vodovodom ali ogrevanjem.

Za razliko od povezovalnih pravil za zemljišča lahko na osnovi povezovalnih pravil za objekte ugotovimo, da apriori algoritem ne najde pravil z zelo nizko stopnjo podpore *s*. Obstaja torej določen delež specifičnih nepremičnin (stanovanja, garaže), ki imajo nekaj skupnih lastnosti. Ostale nepremičnine se razen popolnosti komunalne opremljenosti močno razlikujejo med seboj, kar pomeni, da med njimi ni bližnjih substitutov.

4.4 Dopolnitev baze s podatki iz drugih virov

Cilj magistrskega dela je izgradnja modelov za množično vrednotenje stanovanjskih nepremičnin po tržni metodi. Kot smo ponazorili že v poglavjih 1 in 2, na vrednost nepremičnin vplivajo tako lastnosti same nepremičnine kakor tudi mikro in makrookolje, v katerem se nepremičnine nahajajo. V bazi nepremičninskih transakcij DURS-a so podatki o lastnostih nepremičnin, ki so bile predmet transakcij, sorazmerno obsežni in popolni. Podatki o makrolokaciji nepremičnine so ponazorjeni s šifro občine (glej Občine RS 2003a), v kateri se nepremičnina nahaja. Ti podatki so na voljo za vsako nepremičninsko transakcijo. Podatki o mikrolokaciji nepremičnine so ponazorjeni s spremenljivko *PARCELA* in delno tudi s spremenljivko *ODDALJENOST*. Problem podatkov o parceli je ta, da so zelo nepopolni, ker manjkajo pri veliko transakcijah. Poleg tega so bili za našo analizo popolnoma neprimerni, saj nam posamezne parcelne številke ne izdajajo dejanske mikrolokacije, ker pri izdelavi magistrskega dela nismo imeli vpogleda v ustrezno geodetsko bazo podatkov, ki bi omogočala določitev mikrolokacije nepremičnin na osnovi njihove parcelne številke. Spremenljivka *ODDALJENOST* ne daje vedno prave informacije o mikrolokaciji, saj meri le oddaljenost od centra naselja. Med dejavnike okolja, ki bi tudi lahko vplivajo na ceno nepremičnin, sodijo še razvitost infrastrukture, prisotnost javnih ustanov, ki so pomembne za blaginjo življenja, itn. Teh podatkov v bazi nepremičninskih transakcij DURS-a ni na voljo.

Da bi lahko zgradili čim boljše modele za vrednotenje stanovanjskih nepremičnin ob hkratnem upoštevanju čim večjega števila podatkov o opravljenih nepremičninskih transakcijah, smo se odločili za dopolnitev baze podatkov s šestimi spremenljivkami, ki so predstavljale vplive makrookolja. Zaradi zgoraj opisane nepopolnosti in nedostopnosti podatkov smo pri naši analizi vplive mikrookolja žal upoštevali le v zelo omejenem obsegu s spremenljivko *ODDALJENOST*, čeprav imajo ti dejavniki včasih zelo velik vpliv na vrednost nepremičnine (Štefančič 2003).

Vpliv makrolokacije nepremičnine

Vpliv lokacije na vrednost nepremičnin je v splošnem mogoče upoštevati bodisi s kvalitativnimi spremenljivkami, ki predstavljajo zaključena geografska območja, kot so: regije, mesta itn. (Jansen in Söderberg 1999, Nose 2003, Pavlov 2000), bodisi z numeričnimi spremenljivkami, ki predstavljajo geografske koordinate nepremičnin (Pace et al. 2000), (Plaut in Plaut 1998). Zanimiv je tudi pristop, ki ga je ubral Iversen (2001), ko je kvalitativni pristop dopolnil z informacijo o sosednjih območjih.

Za opis makrolokacije nepremičnin smo izbrali pristop s kombinacijo ene kvalitativne in dveh numeričnih spremenljivk. V bazi podatkov DURS-a je makrolokacija nepremičnine opisana s kvalitativno spremenljivko *SIF_OBC*, ki je bila zaradi obsežne zaloge vrednosti neprimerna za izgradnjo modelov za vrednotenje nepremičnin. To spremenljivko smo nadomestili s kvalitativno spremenljivko *OBM_TRANS*, ki ponazarja 10 območij, v katere smo razdelili 192 slovenskih občin, v katerih so bile izvedene transakcije z nepremičninami v letih od 1999 do 2002. Razdelitev občin v 10 območij nepremičninskih transakcij smo opisali že v poglavju 4.3.1. Na ta način smo močno zmanjšali število območij za nepremičninske transakcije in definirali lokalna, geografsko zaključena območja, za katera smo lahko spremljali spreminjanje vrednosti nepremičnin. Z združenjem posameznih občin v območja nepremičninskih transakcij bi se lahko zgodilo, da se cene kvadratnega metra nepremičnin opazno razlikujejo znotraj enega območja. Primer: občini Horjul in Trzin smo uvrstili v območje okolice Ljubljane, vendar so povprečne cene kvadratnega metra stanovanj v Horjulu bistveno nižje kakor v Trzinu. Zato smo bazo dopolnili še z dvema numeričnima spremenljivkama *X_KOORD* in *Y_KOORD*, ki predstavljata koordinate občin v smeri vzhod-zahod in sever-jug. Na ta način bi lahko upoštevali spreminjanje cen znotraj manjšega okolja in lokalno povezanost sosednjih občin, ki so zaradi statistične delitve padle v različna območja pri spremenljivki *OBM_TRANS*. Koordinate občin v smeri vzhod-zahod in sever-jug smo določili s pomočjo Sl. P4-1 v prilogi 4. Izhodišče koordinatnega sistema je predstavljalo središče občine Ljubljana (šifra občine je 61). Koordinate posameznih občin so bile določene glede na približne centre občin. Merilo za koordinate občin je navedeno na Sl. P4-1.

Drugi dejavniki makrookolja

Druge že prej navedene dejavnike makrookolja smo upoštevali s tremi spremenljivkami. Vpliv infrastrukture smo upoštevali s kvalitativnima spremenljivkama *AC* in *AC_LJ*. Ti dve indikatorski spremenljivki pojasnjujeta, ali avtocesta poteka skozi določeno občino (spremenljivka *AC*), oziroma ali določena občina ima neposredno avtocestno povezavo z Ljubljano (spremenljivka *AC_LJ*). Prisotnosti železniške infrastrukture pri naši analizi nismo upoštevali, ker smo predpostavili, da večina dnevnih migrantov, ki predstavljajo potencialne igralce na trgu nepremičnin, uporablja osebna prevozna sredstva. Vrednost 0 spremenljivke *AC* (*AC_LJ*) pomeni, da v določeni občini ni avtoceste (avtocestne povezave z Ljubljano), vrednost 1 pomeni, da ta povezava obstaja. Vrednosti spremenljivk *AC* in *AC_LJ* smo določili s pomočjo zemljevida avtocest v Republiki Sloveniji (DARS 2003) – glej Sl. P4-2 v prilogi 4. Poleg vpliva infrastrukture smo skušali upoštevati tudi dejavnike, ki so pogojeni s prisotnostjo določenih javnih ustanov (bolnice, šole) in ki bi lahko vplivali na kvaliteto življenja na posameznih območjih. Boljše življenjsko okolje naj bi bilo bolj zeleno, kar naj bi se odražalo tudi v večjem povpraševanju po nepremičninah in bi s tem zviševalo vrednost stanovanjskih nepremičnin v takšnih okoljih. Te vplive smo združili v kvalitativno spremenljivko *RAZV_OBC*, ki ima 4 vrednosti in predstavlja splošno razvitost občine. Vrednost 1 ustreza najmanj razvitim občinam, vrednost 4 ustreza najbolj razvitim občinam. Spremenljivka *RAZV_OBC* je bila povzeta po delu Rovana in Sambta (2003).

Analiza značilnosti vpliva dodatnih spremenljivk

Upravičenost dopolnitve baze z novimi spremenljivkami smo utemeljili tako, da smo z analizo variance prikazali statistično značilnost vpliva kvalitativnih spremenljivk

OBM_TRANS, *AC*, *AC_LJ* in *RAZV_OBC* na spremenljivko *CENA_NA_m²*, ki je bila definirana v poglavju 4.3.1. Rezultati analize variance so prikazani v tabeli 5. Odvisnosti spremenljivke *CENA_NA_m²* od spremenljivk *OBM_TRANS*, *AC*, *AC_LJ* in *RAZV_OBC* so prikazane na slikah Sl. 16 do Sl. 19. Analizo variance smo opravili s programom *SPSS 10.0*.

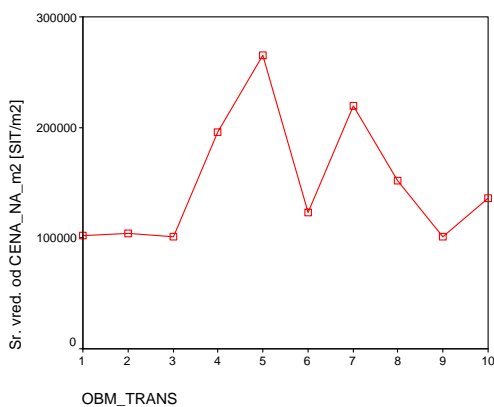
Tabela 5: Značilnost vpliva spremenljivk *OBM_TRANS*, *AC*, *AC_LJ* in *RAZV_OBC* na spremenljivko *CENA_NA_m²*

Neodv. sprem.	<i>OBM_TRANS</i>	<i>AC</i>	<i>AC_LJ</i>	<i>RAZV_OBC</i>
Število skupin	10	2	2	4
VKMS	$1,4 \cdot 10^{14}$	$4,9 \cdot 10^{13}$	$9,5 \cdot 10^{13}$	$1,1 \cdot 10^{14}$
Št. prost. stopenj	9	1	1	3
Sr. vred. VKMS	$1,6 \cdot 10^{13}$	$4,9 \cdot 10^{13}$	$9,5 \cdot 10^{13}$	$3,8 \cdot 10^{13}$
VKZS	$2,3 \cdot 10^{14}$	$3,2 \cdot 10^{14}$	$2,7 \cdot 10^{14}$	$2,5 \cdot 10^{14}$
Št. prost. stopenj	29009	29017	29017	29015
Sr. vred. VKZS	$7,8 \cdot 10^9$	$1,1 \cdot 10^{10}$	$9,3 \cdot 10^9$	$8,7 \cdot 10^9$
SVK	$3,7 \cdot 10^{14}$	$3,7 \cdot 10^{14}$	$3,7 \cdot 10^{14}$	$3,7 \cdot 10^{14}$
Št. prost. stopenj	29018	29018	29018	29018
F statistika	1994,4	4477,6	10234	4345,2
Značilnost F st.	0,000	0,000	0,000	0,000

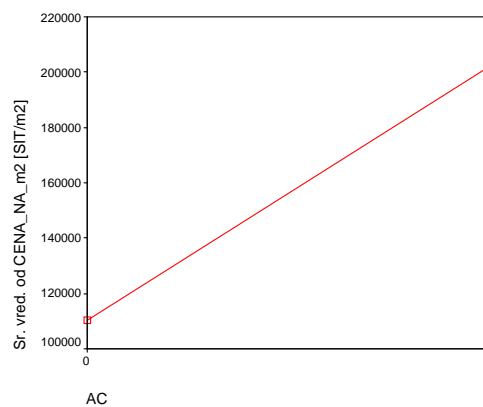
Opombe: VKMS ... Vsota kvadratov med skupinami

VKZS ... Vsota kvadratov znotraj skupin

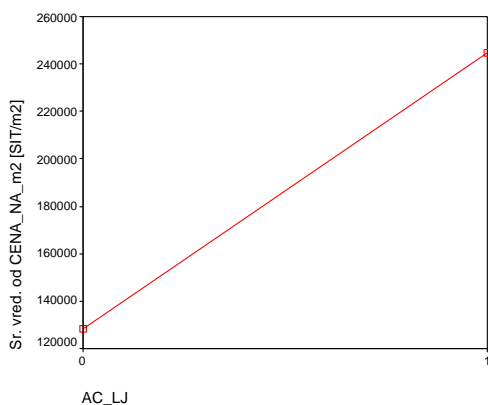
SVK ... Skupna vsota kvadratov



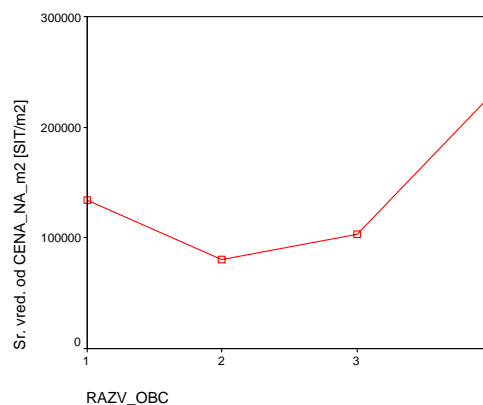
Sl. 16: Odvisnost cene/m² od *OBM_TRANS*



Sl. 17: Odvisnost cene/m² od *AC*



Sl. 18: Odvisnost cene/m² od *AC_LJ*



Sl. 19: Odvisnost cene/m² od *RAZV_OBC*

V tabeli 5 vidimo, da so testi značilnosti za spremenljivko $CENA_NA_m^2$ značilni glede na spremenljivke OBM_TRANS , AC , AC_LJ in $RAZV_OBC$ pri stopnji značilnosti 0,001, kar pomeni, da je vključitev dodatnih spremenljivk v analizo smiselna.

5 Primerjava modelov za vrednotenje stanovanj

5.1 Priprava podatkov za izgradnjo in preverjanje modelov za vrednotenje

5.1.1 Določitev spremenljivk za izgradnjo modelov za vrednotenje

Kot smo povedali že v uvodnem poglavju, smo v sklopu tega magistrskega dela skušali zgraditi nekaj alternativnih modelov za množično vrednotenje stanovanjskih nepremičnin po tržni metodi in jih primerjati med seboj. Za vrednotenje stanovanjskih nepremičnin smo se odločili zato, ker predstavljajo transakcije s stanovanji največji delež transakcij v bazi podatkov DURS-a, ker je tržna metoda za vrednotenje nepremičnin (ki jo edino omogoča obstoječa baza podatkov DURS-a) najbolj primerna prav za vrednotenje stanovanjskih nepremičnin in ker so podatki o gibanju vrednosti stanovanjskih nepremičnin v trenutku sproščanja nacionalne varčevalne sheme verjetno najbolj zanimivi za varčevalce iz te sheme.

Splošni model za množično vrednotenje nepremičnin po tržni metodi ima naslednjo obliko:

$$CENA_NA_m^2 = f(X_1, X_2, \dots, X_p) \quad (29).$$

Vrednost stanovanjske nepremičnine predstavimo z odvisno spremenljivko $CENA_NA_m^2$, to je cena kvadratnega metra stanovanj. V literaturi (glej na primer Eckert 1990) se za vrednost nepremičnin sicer pogosto uporablja kar vrednost celotne nepremičnine, ki pa ima to pomanjkljivost, da onemogoča primerjavo med sicer podobnimi, a različno velikimi nepremičninami. Zato smo se odločili za napoved cene na enoto stanovanjske površine. Spremenljivko $CENA_NA_m^2$ smo natančno definirali v poglavju 4.3.1 in te definicije smo se strogo držali pri celotni analizi.

Spremenljivke X_1, X_2, \dots, X_p predstavljajo bodisi kvalitativne bodisi numerične neodvisne spremenljivke v modelu za vrednotenje nepremičnin, ki opisujejo lastnosti nepremičnine, njeno lokacijo, mikro in makrookolje ter časovno komponento, ki lahko vpliva na vrednost nepremičnine. Neodvisne spremenljivke v modelih vrednotenja smo povzeli v nespremenjeni ali rahlo modificirani obliki po bazi nepremičninskih transakcij DURS-a, ki smo jim dodali še šest neodvisnih spremenljivk, ki smo jih definirali v poglavju 4.4. V modelih za vrednotenje nepremičnin smo upoštevali naslednje neodvisne spremenljivke²³:

- CAS_TRAN je numerična spremenljivka, ki je predstavljala čas transakcije na linearni skali. Vrednosti te spremenljivke smo dobili s transformacijo originalne spremenljivke $DATUM_PRIJAVE$. Izhodišče za spremenljivko CAS_TRAN je 1. 1. 1999, ko ima spremenljivka vrednost 0. Njena enota je eno tromesečje – to pomeni, da ima ta spremenljivka vrednost 1 na dan 1. 4. 1999. Čas transakcije bi lahko vplival na ceno kvadratnega metra stanovanj preko inflacije (multiplikativni efekt z ostalimi

²³ Imena neodvisnih spremenljivk so dolga največ osem črk zaradi omejitve v programu *SPSS 10.0*.

neodvisnimi spremenljivkami) ali preko časovne dinamike trga (efekt bi lahko bil aditiven, multiplikativen, kombiniran, itn.).

- *MESEC* je numerična spremenljivka, ki je predstavljala mesec v letu. Njene vrednosti so od 1 do 12 in so bile določene glede na originalno spremenljivko *DATUM_PRIJAVE*. Spremenljivko *MESEC* smo vključili v analizo zato, da bi lahko ovrednotili sezonski vpliv na ceno kvadratnega metra stanovanj, kljub temu da nismo pričakovali večjega vpliva te spremenljivke.
- *ST_VLOZ* je kvalitativna spremenljivka z vrednostima 0 in 1, ki sta bili definirani v skladu s tabelo P3-1 v prilogi 3. Spremenljivka ima vrednost 1, če je bil za trgovano nepremičnino podan vložek, in vrednost 0, če vložek ni bil podan oziroma ta informacija v bazi podatkov ni bila na voljo.
- *IZMERA* je numerična spremenljivka, ki je predstavljala velikost trgovane nepremičnine. Pričakovan vpliv velikosti stanovanja na ceno kvadratnega metra ni bil enoličen. Po eni strani smo pričakovali, da cena kvadratnega metra z velikostjo stanovanja pada zaradi manjšega povpraševanja po velikih stanovanjih, po drugi strani pa bi lahko cena kvadratnega metra z velikostjo stanovanja tudi naraščala, saj so luksuzna stanovanja navadno večja in dražja.
- *STAROST* je numerična spremenljivka, ki je predstavljala starost stanovanja. Dobili smo jo s transformacijo originalnih spremenljivk *LETO_IZGR* ter *LETO_PREN* in sicer tako, da smo od letnice 2002 odšteli vrednost spremenljivke *LETO_PREN*, če je bilo stanovanje obnovljeno, oziroma *LETO_IZGR*, če stanovanje ni bilo prenovljeno. Pričakovani vpliv starosti stanovanja na ceno kvadratnega metra je bil negativen, vendar je dejanski vpliv odvisen od mikrookolja, v katerem je stanovanje.
- *KLETNO* je kvalitativna spremenljivka z vrednostjo 1, če je stanovanje kletno, in vrednostjo 0, če stanovanje ni kletno. Pričakovani učinek kletnega stanovanja na ceno kvadratnega metra je bil negativen.
- *PODSTRES* je kvalitativna spremenljivka z vrednostjo 1, če je stanovanje podstrešno, in vrednostjo 0, če stanovanje ni podstrešno. Pričakovani učinek podstrešnega stanovanja na ceno je bil sicer v splošnem negativen, vendar bi lahko bila moderna mansardna stanovanja včasih tudi dražja od stanovanj pod njimi.
- *NADSTROP* je numerična spremenljivka, ki je predstavljala številko nadstropja, v katerem se nahaja stanovanje. Originalno spremenljivko smo spremenili tako, da je številka pritličja enaka 0. Pričakovan vpliv nadstropja na ceno kvadratnega metra stanovanja je bil nejasen. Višje nadstropje v splošnem pomeni boljši razgled in s tem večjo vrednost stanovanja, vendar se nekateri ljudje bojijo višine ali ne želijo stanovati visoko zaradi stopnic, kar bi lahko zmanjšalo vrednost visoko ležečih stanovanj.
- *LEGA* je kvalitativna spremenljivka s štirimi vrednostmi, ki je predstavljala lego stanovanja v smeri zemljepisnih smeri. Vpliv te spremenljivke na ceno kvadratnega metra stanovanja je različen. Sončna lega (jug) je navadno ugodna, čeprav je na morju lahko bolj zelena severna lega (stran od sonca). Sem gotovo spada tudi pogled: v Ljubljani je recimo lep pogled na Alpe na severni strani, kar bi lahko kompenziralo vrednost stanovanja za nejužno lego. Pri multiplih regresijskih modelih, regresijskih in

modelnih drevesih smo spremenljivko *LEGA* nadomestili s tremi binarnimi slamnatimi spremenljivkami (referenca je severna lega), pri ostalih modelih pa s štirimi.²⁴

- *DVIGALO* je kvalitativna spremenljivka z vrednostjo 1, če je dostop do stanovanja mogoč z dvigalom, in vrednostjo 0, če dostop do stanovanja ni mogoč z dvigalom. Pričakovani učinek prisotnosti dvigala na ceno kvadratnega metra je bil pozitiven.
- *ODDALJEN* je numerična spremenljivka, ki je predstavljala oddaljenost stanovanja od centra naselja. Navadno so stanovanja na periferiji cenejša kakor v centru, vendar je to spet odvisno od uglednosti posameznih sosesk (glej tudi Plaut in Plaut 1998).
- *ELEKTR* je kvalitativna spremenljivka z vrednostjo 1, če stanovanje ima elektriko in vrednostjo 0, če stanovanje nima elektrike. Pričakovani učinek elektrike na ceno kvadratnega metra je bil pozitiven.
- *VODOVOD* je kvalitativna spremenljivka z vrednostjo 1, če stanovanje ima vodovod, in vrednostjo 0, če stanovanje nima vodovoda. Pričakovani učinek vodovoda na ceno kvadratnega metra je bil pozitiven.
- *TELEFON* je kvalitativna spremenljivka z vrednostjo 1, če stanovanje ima telefon, in vrednostjo 0, če stanovanje nima telefona. Pričakovani učinek telefona na ceno kvadratnega metra je bil pozitiven.
- *KANALIZ* je kvalitativna spremenljivka z vrednostjo 1, če stanovanje je priključeno na kanalizacijo, in vrednostjo 0, če stanovanje nanjo ni priključeno. Pričakovani učinek priključka na kanalizacijo na ceno kvadratnega metra je bil pozitiven.
- *OBM_TRAN* je kvalitativna spremenljivka z desetimi vrednostmi, ki je predstavljala zaključeno geografsko območje, v katerem se nahaja stanovanje – glej poglavje 4.3.1. Pri multiplih regresijskih modelih, regresijskih in modelnih drevesih smo spremenljivko *OBM_TRAN* nadomestili z devetimi binarnimi slamnatimi spremenljivkami (referenca je 1. območje transakcije), pri ostalih modelih pa z desetimi. Razlog je enak kakor pri spremenljivki *LEGA*. Na osnovi rezultatov v poglavju 4.3 smo upravičeno pričakovali pomemben vpliv te spremenljivke na ceno kvadratnega metra stanovanja.
- *X_KOORD* in *Y_KOORD* sta numerični spremenljivki, ki sta predstavljali geografski položaj občine, v kateri se nahaja stanovanje glede na Ljubljano – glej poglavje 4.4. Glede na razmere v Sloveniji smo načeloma pričakovali padanje vrednosti kvadratnega metra stanovanj od zahoda proti vzhodu.
- *RAZV_OBC* je urejena kvalitativna spremenljivka s štirimi vrednostmi, ki je predstavljala razvitost občine (glej poglavje 4.4). Zato smo jo za izgradnjo modelov za vrednotenje spremenili v numerično spremenljivko, saj smo pričakovali, da so stanovanja v bolj razvitih občinah dražja.
- *AC* je kvalitativna spremenljivka z vrednostjo 1, če skozi občino, v kateri se nahaja stanovanje, poteka avtocesta, in vrednostjo 0 v nasprotnem primeru. Pričakovani učinek prisotnosti avtoceste na ceno kvadratnega metra stanovanja je bil pozitiven.

²⁴ Binarne slamnate spremenljivke imajo vrednost 0 ali 1 (Pindyck in Rubinfeld 1998). Različna števila binarnih slamnatih spremenljivk smo uporabili zaradi specifičnega načina upoštevanja podobnosti pri različnih modelih.

- *AC_LJ* je kvalitativna spremenljivka z vrednostjo 1, če ima občina, v kateri se nahaja stanovanje, avtocestno povezavo z Ljubljano, in vrednostjo 0 v nasprotnem primeru. Pričakovan učinek prisotnosti avtocestne povezave z Ljubljano na ceno kvadratnega metra stanovanja je bil pozitiven.

Celotno število registriranih nepremičninskih transakcij s stanovanji je 29972. Vendar vseh transakcij nismo upoštevali za izgradnjo modelov zaradi nepravilnosti v podatkih – glej poglavje 4.2. Ko smo podatke prečistili v skladu s poglavjem 4.2.2 in izločili vse preostale transakcije z manjkajočimi podatki, smo dobili množico 19938 prečiščenih nepremičninskih transakcij, ki niso vsebovale manjkajočih podatkov. To množico smo nato uporabili za izgradnjo modelov za množično vrednotenje nepremičnin po tržni metodi.

Originalnih spremenljivk *ZAP_POG*, *ZAP_NEP*, *VRSTA_NEP*, *SIF_KO*, *PARCELA*, *KK_ZEM*, *RAZRED* in *LETO_NAKUPA* pri izgradnji modelov za vrednotenje nismo upoštevali, ker so ali nesmiselne za stanovanja ali pa niso imele vpliva na ceno kvadratnega metra stanovanj. Spremenljivke *NACIN_PRIDOB* nismo upoštevali zaradi velikega števila manjkajočih vnosov, ki bi še dodatno zmanjšali število transakcij za izgradnjo modelov.

Glede na tip posameznih neodvisnih spremenljivk smo imeli v našem primeru opravka s tipičnim problemom prostorsko-časovnega modeliranja vrednosti stanovanjskih nepremičnin. Ta problem so raziskovalci navadno reševali z različnimi regresijskimi modeli, kjer so upoštevali čas s kvalitativnimi ali z numeričnimi spremenljivkami (Ben-Shahar 2002, Clapp in Giacotto 1992, Eckert 1990, Iversen 2001, Nose 2003, Pace et al. 2000, Plaut 1998).

V magistrskem delu smo predstavili nekatere alternativne metode, ki jih je mogoče uporabiti za reševanje tega problema in jih primerjali z regresijskim pristopom. Preden smo se lotili izdelave modelov za množično vrednotenje nepremičnin po tržni metodi, smo skušali ugotoviti, kakšne bi bile najprimernejše oblike odvisne in neodvisnih spremenljivk za izgradnjo naših modelov. Pri tem smo si pomagali z multipli regresijskimi modeli.

Najprej smo se odločili, ali bomo napovedovali realne ali nominalne vrednosti spremenljivke *CENA_NA_m²* za stanovanjske nepremičnine. Z napovedjo realnih cen kvadratnega metra stanovanj bi apriori izločili vpliv inflacije na napovedano vrednost. Zato smo zgradili dva aditivna regresijska modela (glej enačbo (11)) za celotno množico podatkov: s prvim smo napovedovali nominalne, z drugim pa realne cene kvadratnega metra stanovanj v odvisnosti od zgoraj opisanih neodvisnih spremenljivk. Realne cene kvadratnega metra stanovanj v letih 2000, 2001 in 2002 smo izračunali glede na raven cen iz leta 1999 s pomočjo indeksov rasti drobnoprodajnih cen za posamezna leta (Statistični letopis RS 2002). Za izgradnjo regresijskih modelov smo uporabili program *SPSS 10.0*. Vse neodvisne spremenljivke smo v modela vključili naenkrat. Regresijska modela za napoved nominalnih in realnih cen kvadratnega metra stanovanj sta prikazana v tabeli P5-1 v prilogi 5.

Zaradi majhne razlike med obema variantama, in ker je bil povrh tega model za napoved realnih cen slabši od modela za napoved nominalnih cen smo se odločili, da bomo z modeli za vrednotenje nepremičnin napovedovali nominalne cene kvadratnega metra stanovanj. Takšna izbira je primernejša tudi s stališča uporabnikov modelov, saj napovedanih cen ni treba dodatno transformirati za časovni trenutek, za katerega bi želeli izvedeti vrednost kvadratnega metra stanovanja. Če primerjamo standardizirane regresijske koeficiente vidimo, da je čas transakcije bistveno bolj pomemben pri napovedi nominalne cene kvadratnega metra. Iz tega

lahko sklepamo, da v povprečju čas transakcije največ doprinese k ceni kvadratnega metra stanovanja preko inflacijskega vpliva.

Vpliv posameznih neodvisnih spremenljivk je lahko aditiven, multiplikativen ali kombiniran, zato smo skušali oceniti, v kakšni obliki naj bi posamezne neodvisne spremenljivke nastopale v naših modelih. To smo naredili tako, da smo za celotno množico podatkov zgradili štiri različne regresijske modele za napoved nominalne cene kvadratnega metra stanovanj:

- *Aditivni model* – glej enačbo (11). Neodvisne spremenljivke X_i so bile numerične ali kvalitativne. Rezultati regresijske analize za ta model so prikazani že v tabeli P5-1.
- *Multiplikativni model*:

$$CENA_NA_m^2 = \beta_0 \cdot \prod_{i=1}^{p_i} X_i^{\beta_i} \cdot \prod_{j=1}^{p_j} \beta_j^{X_j} \quad (30).$$

X_i so bile numerične X_j pa kvalitativne neodvisne spremenljivke. Enačbo (30) smo z logaritmiranjem transformirali v linearen regresijski model za logaritme numeričnih spremenljivk in netransformirane kvalitativne spremenljivke. Rezultati regresijske analize za multiplikativni model so prikazani v tabeli P5-2.

- *Kombinirani model*, v katerem smo spremenljivko CAS_TRAN pomnožili z ostalimi neodvisnimi spremenljivkami, nakar smo tvorili aditivni regresijski model:

$$CENA_NA_m^2 = \beta_0 + \beta_1 \cdot CAS_TRAN + \sum_{i=2}^p \beta_i \cdot (CAS_TRAN \cdot X_i) \quad (31).$$

X_i so bile numerične ali kvalitativne neodvisne spremenljivke razen časa transakcije. Rezultati regresijske analize za kombinirani model so prikazani v tabeli P5-2.

- *Združeni model*, v katerem smo združili člene iz aditivnega in kombiniranega modela:

$$CENA_NA_m^2 = \beta_0 + \sum_{i=1}^p \beta_i \cdot X_i + \sum_{j=2}^p \gamma_j \cdot (CAS_TRAN \cdot X_j) \quad (32).$$

X_i in X_j so bile numerične ali kvalitativne neodvisne spremenljivke. Rezultati regresijske analize za združeni model so prikazani v tabeli P5-2.

Glede na rezultate regresijske analize v tabelah P5-1 in P5-2 smo se odločili, da za izgradnjo modelov za vrednotenje stanovanjskih nepremičnin posameznih spremenljivk ne bomo transformirali. Z izjemo združenega modela je dal regresijski model z netransformiranimi spremenljivkami najboljše napovedi cene kvadratnega metra stanovanj, poleg tega je bil ta model bistveno enostavnejši od združenega, ki je bil glede kvalitete napovedi le malo boljši.

Če v tabelah P5-1 in P5-2 v prilogi 5 ovrednotimo vpliv neodvisnih spremenljivk, katerih regresijski koeficienti imajo značilnost večjo od 0,001 vidimo, da se vplivi posameznih spremenljivk ujemajo z našimi predvidevanji. Vplive neznačilnih neodvisnih spremenljivk bi težko komentirali, ker so lahko izračunani regresijski koeficienti tudi posledica naključnih variacij v bazi podatkov.

5.1.2 Priprava podatkov za izgradnjo modelov za vrednotenje

Priprava podatkov pogosto predstavlja največji delež časa, ki ga porabimo za izvajanje rudarjenja podatkov (Štěpánková et al. 2003). V predhodnih poglavjih smo že dokaj podrobno

opisali pomanjkljivosti podatkov v bazi nepremičninskih transakcij in načine popraviljanja ali odpravljanja teh pomanjkljivosti, zato tega ne bomo še enkrat ponavljali. V tem podpoglavju je opisan postopek tvorjenja učne in testne množice vzorcev za izgradnjo in testiranje modelov za množično vrednotenje stanovanjskih nepremičnin po tržni metodi.

Kot smo povedali že v uvodu, smo za izgradnjo, testiranje in primerjavo modelov za množično vrednotenje stanovanj uporabili metodo prečnega preverjanja, ki je opisana v poglavju 3.8. Odločili smo se za razbitje množice 19938 prečiščenih nepremičninskih transakcij s stanovanji na tri pare učnih in testnih množic vzorcev za izgradnjo in testiranje modelov. To smo storili tako, da smo vsako izmed 19938 transakcij razvrstili v eno od treh podmnožic s postopkom stratificiranega vzorčenja po kombinaciji spremenljivk *CAS_TRAN* in *OBM_TRAN*. S tem smo zagotovili, da se je v vsaki izmed treh podmnožic nahajalo približno enako število nepremičninskih transakcij za vsako leto in za vsako območje transakcije. Vsaka izmed podmnožic je bila enkrat uporabljena za testiranje modela in dvakrat (v kombinaciji z eno izmed preostalih dveh podmnožic) za izgradnjo modela za vrednotenje nepremičnin - glej poglavje 3.8. V tabeli 6 so prikazane osnovne značilnosti učnih in testnih množic za izgradnjo in testiranje modelov za množično vrednotenje stanovanjskih nepremičnin. Vidimo, da razdelitev transakcij v testne množice ni bila natančno po tretjinah, vendar so razlike med posameznimi množicami zanemarljive glede na število vseh transakcij.

Tabela 6: Osnovne značilnosti učnih in testnih množic za izgradnjo in testiranje modelov za množično vrednotenje stanovanjskih nepremičnin

Kombinacija	K1	K2	K3
Testna množica	1	2	3
Št. testnih transakcij	6646	6651	6641
Učna množica	2, 3	1, 3	1, 2
Št. učnih transakcij	13292	13287	13297

Učne in testne množice podatkov za izgradnjo in testiranje modelov za množično vrednotenje stanovanjskih nepremičnin ter vsi zgrajeni modeli se nahajajo na priloženi zgoščenki. Struktura map na zgoščenki je opisana v prilogi 7.

5.2 Model za vrednotenje na osnovi linearne regresije

Referenčni model za množično vrednotenje stanovanjskih nepremičnin po tržni metodi smo zgradili z metodo multiple regresije. Uporabljene spremenljivke so bile opisane v poglavju 5.1.1. V poglavju 5.1.1 smo za izračun regresijskih koeficientov uporabili program *SPSS 10.0*, ker smo želeli oceniti tudi značilnost in relativni vpliv posameznih spremenljivk. Za izgradnjo regresijskega modela za napoved vrednosti spremenljivke *CENA_NA_m²* smo na tem mestu uporabili program *WEKA 3.2*, ker smo ta program uporabili tudi pri vseh drugih metodah, saj *SPSS 10.0* ne vsebuje podpore za alternativne metode rudarjenja podatkov.

V programu *WEKA 3.2* se metoda multiple regresije nahaja v knjižnici *weka.classifiers.LinearRegression*. Edina opcija, ki jo programski paket *WEKA 3.2* ponuja za to metodo, je način izbire neodvisnih spremenljivk v model. Za vključitev neodvisnih spremenljivk v multipli regresijski model smo uporabili dve metodi: v model smo vključili bodisi vse neodvisne spremenljivke naenkrat bodisi smo jih vključevali v model postopoma (metoda *greedy* v programskem paketu *WEKA 3.2*). V tabeli 7 je prikazana kvaliteta modelov, ki so

bili zgrajeni na osnovi učnih množic vzorcev za posamezne učno-testne kombinacije iz poglavja 5.1.2. Kvaliteta zgrajenih modelov je bila ocenjena z merami iz poglavja 3.8 za učno in testno množico vzorcev. V skladu s postopkom prečnega preverjanja so bile izračunane tudi povprečne mere uspešnosti glede na tri kombinacije učnih in testnih množic vzorcev.

Tabela 7: Kvaliteta multiplih regresijskih modelov

Učno-testna kombinacija	Mere ujemanja	Postopno vključevanje neodvisnih sprem.		Vključene vse neodvisne spremenljivke	
		Učna množica	Testna množica	Učna množica	Testna množica
K1	Število vzorcev	13292	6646	13292	6646
	Čas učenja	176 s	170 s	1,6 s	1,6 s
	Čas testiranja	~2 s	~1 s	~2 s	~1 s
	Korel. koeficient	0,771	0,768	0,771	0,768
	<i>SAN</i> [SIT]	44967,02	44847,78	44957,62	44830,82
	<i>KSKO</i> [SIT]	64417,23	64034,18	64411,79	64020,59
	<i>RAN</i> [%]	53,325	53,683	53,314	53,663
	<i>KRKO</i> [%]	63,691	64,114	63,685	64,101
K2	Število vzorcev	13287	6651	13287	6651
	Čas učenja	170 s	175 s	1,6 s	1,6 s
	Čas testiranja	~2 s	~1 s	~2 s	~1 s
	Korel. koeficient	0,768	0,773	0,768	0,774
	<i>SAN</i> [SIT]	44779,69	45163,11	44778,05	45149,30
	<i>KSKO</i> [SIT]	64398,12	64079,98	64391,96	64063,88
	<i>RAN</i> [%]	53,354	53,666	53,352	53,651
	<i>KRKO</i> [%]	64,043	63,412	64,037	63,396
K3	Število vzorcev	13297	6641	13297	6641
	Čas učenja	189 s	195 s	1,6 s	1,4 s
	Čas testiranja	~2 s	~1 s	~2 s	~1 s
	Korel. koeficient	0,772	0,766	0,772	0,766
	<i>SAN</i> [SIT]	44839,89	44978,70	44813,50	44938,48
	<i>KSKO</i> [SIT]	63887,50	65126,39	63878,42	65093,64
	<i>RAN</i> [%]	53,485	53,291	53,454	53,243
	<i>KRKO</i> [%]	63,594	64,332	63,585	64,299
Prečno preverjanje	Korel. koeficient	0,770	0,769	0,770	0,769
	Determ. koeficient	0,593	0,591	0,593	0,591
	<i>SAN</i> [SIT]	44862,20	44996,53	44849,73	44972,87
	<i>KSKO</i> [SIT]	64234,28	64413,52	64227,39	64392,70
	<i>RAN</i> [%]	53,388	53,547	53,373	53,519
	<i>KRKO</i> [%]	63,776	63,953	63,769	63,932

Opombe: *SAN* ... Srednja vrednost absolutne napake – glej poglavje 3.8
KSKO ... Koren srednjega kvadratičnega odstopanja – glej poglavje 3.8
RAN ... Relativna absolutna napaka – glej poglavje 3.8
KRKO ... Koren relativnega kvadratičnega odstopanja – glej poglavje 3.8

Iz primerjave korelacijskih koeficientov v tabeli 7 vidimo, da smo najboljše napovedi za testno množico vzorcev dosegli pri učno-testni kombinaciji K2. Dejansko so bile napovedi za testno množico vzorcev pri vseh modelih najboljše pri učno-testni kombinaciji K2, četudi napovedi za učno množico vzorcev niso bile najboljše. V prilogi 6 je prikazan multipli

regresijski model za napoved cene kvadratnega metra stanovanj, ki smo ga dobili z učno-testno kombinacijo vzorcev K2 z metodo postopnega vključevanja neodvisnih spremenljivk.

Če primerjamo povprečna determinacijska koeficienta za učno množico vzorcev iz tabele 7 z determinacijskim koeficientom iz tabele P5-1 v prilogi 5, vidimo, da so vsi trije determinacijski koeficienti enaki 0,593. To pomeni, da sta metodi, ki ju uporabljata programa *SPSS 10.0* in *WEKA 3.2*, enakovredni in da smo iz regresijskega modela ob ohranitvi kvalitete napovedi lahko izpustili manj značilne neodvisne spremenljivke. Vendar je bil čas izgradnje modela pri postopnem vključevanju neodvisnih spremenljivk bistveno daljši, kakor pri hkratni vključitvi neodvisnih spremenljivk v model. V tabeli 7 vidimo tudi, da so bile napovedi za testno množico vzorcev slabše kakor za učno množico vzorcev, kar je logično, saj zgrajen regresijski model ni bil prilagojen testnim podatkom.

5.3 Model za vrednotenje na osnovi metode najbližjih sosedov

Metoda najbližjih sosedov je lokalna metoda za napoved – glej poglavje 3.5.2. Uporabljene spremenljivke so bile opisane v poglavju 5.1.1. Model za vrednotenje stanovanjskih nepremičnin je bil sestavljen iz vseh nepremičninskih transakcij iz učne množice. V programu *WEKA 3.2* se metoda najbližjih sosedov nahaja v knjižnici *weka.classifiers.IBk*. Glavne opcije, ki jih program *WEKA 3.2* ponuja za to metodo, so naslednje:

- opcija za vključitev prečnega preverjanja,
- opcija za normalizacijo neodvisnih spremenljivk,
- uteženo upoštevanje najbližjih sosedov glede na razdalje do danega vzorca,
- število najbližjih sosedov ali velikost okna v katerem iščemo najbližje sosedu.

Za napoved cen kvadratnega metra stanovanj smo neodvisne spremenljivke normalizirali, da so imele enak vpliv pri določitvi najbližjih sosedov. Pomembnost posameznih sosedov smo upoštevali inverzno proporcionalno razdaljam med transakcijo, za katero smo napovedovali ceno kvadratnega metra, in relevantnimi najbližjimi sosedi. Modele za vrednotenje smo zgradili za različna števila najbližjih sosedov. Najboljše rezultate za testne množice smo dobili, če smo izbrali 50 najbližjih sosedov, zato so podani le rezultati za to izbiro. V tabeli 8 je prikazana kvaliteta modelov za posamezne učno-testne kombinacije iz poglavja 5.1.2. Kvaliteta modelov je bila ocenjena z merami iz poglavja 3.8 za učno in testno množico vzorcev. V tabeli so prikazane tudi povprečne mere uspešnosti za učne in testne množice vzorcev. Metoda najbližjih sosedov spada med metode za lokalno napoved, zato zgrajenega modela ne moremo prikazati, ker je bil le-ta različen za vsak testni vzorec.

V tabeli 8 vidimo, da so bile napovedi za testno množico vzorcev bistveno slabše kakor za učno množico vzorcev. To je posledica tega, da so bile transakcije iz učne množice vzorcev enake tistim v modelu. Zaradi tega se je z večanjem števila najbližjih sosedov kvaliteta napovedi za učno množico slabšala (najboljša napoved je bila pri enem najbližjem sosedu, ko je bil korelacijski koeficient enak ena), kvaliteta napovedi za testno množico pa se je izboljševala, dokler števila najbližjih sosedov nismo povečevali preveč. Čas testiranja modela je bil bistveno daljši, kakor čas učenja modela, kar je značilnost lokalnih metod za napoved. Iz primerjave korelacijskih koeficientov v tabeli 8 vidimo, da smo najboljše napovedi za testno množico vzorcev spet dobili pri učno-testni kombinaciji K2. Če primerjamo tabeli 7 in

8 vidimo, da so bile napovedi cen kvadratnega metra stanovanj za testno množico vzorcev boljše pri metodi multiple regresije kakor pri metodi najbližjih sosedov. Razmerje je bilo sicer obratno pri kakovosti napovedi za učno množico vzorcev, vendar je bil naš cilj doseči kakovostno vrednotenje za tiste nepremičnine, ki niso sodelovale pri izgradnji modela.

Tabela 8: Kvaliteta modelov 50 najbližjih sosedov

Učno-testna kombinacija	Mere ujemanja	50 najbližjih sosedov	
		Učna množica	Testna množica
K1	Število vzorcev	13292	6646
	Čas učenja	~1 s	~1 s
	Čas testiranja	19 min	9 min
	Korel. koeficient	0,958	0,754
	<i>SAN</i> [SIT]	20345,97	46714,24
	<i>KSKO</i> [SIT]	31333,51	65670,59
	<i>RAN</i> [%]	24,128	55,917
	<i>KRKO</i> [%]	30,980	65,753
K2	Število vzorcev	13287	6651
	Čas učenja	~1 s	~1 s
	Čas testiranja	20 min	9 min
	Korel. koeficient	0,958	0,758
	<i>SAN</i> [SIT]	20303,11	47208,57
	<i>KSKO</i> [SIT]	31286,76	65991,52
	<i>RAN</i> [%]	24,191	56,098
	<i>KRKO</i> [%]	31,114	65,304
K3	Število vzorcev	13297	6641
	Čas učenja	~1 s	~1 s
	Čas testiranja	19 min	10 min
	Korel. koeficient	0,958	0,755
	<i>SAN</i> [SIT]	20362,60	46916,20
	<i>KSKO</i> [SIT]	30989,61	66404,44
	<i>RAN</i> [%]	24,289	55,586
	<i>KRKO</i> [%]	30,847	65,594
Prečno preverjanje	Korel. koeficient	0,958	0,756
	Determ. koeficient	0,918	0,571
	<i>SAN</i> [SIT]	20337,23	46946,33
	<i>KSKO</i> [SIT]	31203,29	66022,18
	<i>RAN</i> [%]	24,202	55,867
	<i>KRKO</i> [%]	30,981	65,550

Opomba: *SAN*, *KSKO*, *RAN*, *KRKO* – glej tabelo 7 in poglavje 3.8.

5.4 Model za vrednotenje na osnovi regresije z lokalnimi utežmi

Regresija z lokalnimi utežmi je lokalna metoda za napoved – glej poglavje 3.5.3. Uporabljene spremenljivke so bile opisane v poglavju 5.1.1. Model za vrednotenje stanovanjskih nepremičnin so sestavljale vse nepremičninske transakcije iz učne množice. Le-te so dobile različne uteži, ki so bile pogojene s podobnostjo transakcij in testnega vzorca, za katerega smo zgradili lokalni regresijski model. V programu *WEKA* 3.2 se ta metoda nahaja v knjižnici *weka.classifiers.LWR*. Program *WEKA* 3.2 ponuja dve opciji za to metodo:

- izbira jedrne funkcije za lokalno regresijo (linearna, inverzna ali Gaussova funkcija),
- izbira števila najbolj podobnih vzorcev v bazi, ki sodelujejo pri izgradnji modela.

Za določitev uteži, ki pripadajo vzorcem v učni množici, smo izbrali linearne jedrne funkcije zaradi večje hitrosti računanja. Modele za vrednotenje smo spet zgradili za različna števila najbližjih sosedov. Najboljše rezultate za testne množice smo dobili, če smo lokalne regresijske modele tvorili za 2000 najbližjih sosedov, zato so podani le rezultati za to izbiro. V tabeli 9 je prikazana kvaliteta lokalnih regresijskih modelov, ki temeljijo na učnih množicah vzorcev za posamezne učno-testne kombinacije iz poglavja 5.1.2. Kvaliteta modelov je bila ocenjena z merami iz poglavja 3.8 za učno in testno množico vzorcev. V tabeli so prikazane tudi povprečne mere uspešnosti za učne in testne množice vzorcev. Regresija z lokalnimi utežmi je metoda za lokalno napoved, zato zgrajenega modela ne moremo prikazati, saj je bil različen za vsak testni vzorec.

Tabela 9: Kvaliteta regresijskih modelov z lokalnimi utežmi za 2000 najbližjih sosedov

Učno-testna kombinacija	Mere ujemanja	2000 najbližjih sosedov	
		Učna množica	Testna množica
K1	Število vzorcev	13292	6646
	Čas učenja	~1 s	~1 s
	Čas testiranja	183 min	93 min
	Korel. koeficient	0,815	0,788
	<i>SAN</i> [SIT]	39969,23	41871,48
	<i>KSKO</i> [SIT]	58661,75	61503,94
	<i>RAN</i> [%]	47,398	50,121
	<i>KRKO</i> [%]	58,000	61,581
K2	Število vzorcev	13287	6651
	Čas učenja	~1 s	~1 s
	Čas testiranja	186 min	95 min
	Korel. koeficient	0,813	0,795
	<i>SAN</i> [SIT]	39698,59	42414,43
	<i>KSKO</i> [SIT]	58571,82	61341,16
	<i>RAN</i> [%]	47,300	50,401
	<i>KRKO</i> [%]	58,249	60,702
K3	Število vzorcev	13297	6641
	Čas učenja	~1 s	~1 s
	Čas testiranja	183 min	94 min
	Korel. koeficient	0,816	0,789
	<i>SAN</i> [SIT]	39807,94	41855,23
	<i>KSKO</i> [SIT]	58157,07	62228,85
	<i>RAN</i> [%]	47,483	49,590
	<i>KRKO</i> [%]	57,890	61,470
Prečno preverjanje	Korel. koeficient	0,815	0,791
	Determ. koeficient	0,664	0,625
	<i>SAN</i> [SIT]	39825,25	42047,05
	<i>KSKO</i> [SIT]	58463,55	61691,32
	<i>RAN</i> [%]	47,394	50,037
	<i>KRKO</i> [%]	58,046	61,251

Opomba: *SAN*, *KSKO*, *RAN*, *KRKO* – glej tabelo 7 in poglavje 3.8.

V tabeli 9 vidimo, da so bile napovedi za testno množico vzorcev sicer slabše kakor za učno množico vzorcev, vendar ta razlika ni tako velika, kakor je bila v primeru metode najbližjih sosedov (glej tabelo 8). To je zaradi tega, ker je bil vpliv podobnosti med posameznim testnim vzorcem in njegovim ekvivalentom v učni množici manjši pri regresiji z lokalnimi utežmi kakor pri metodi najbližjih sosedov. Kljub temu je tudi pri regresiji z lokalnimi utežmi mogoče opaziti značilnost vseh lokalnih metod: to je, da se je z večanjem števila najbližjih sosedov kvaliteta napovedi za učno množico slabšala, kvaliteta napovedi za testno množico pa izboljševala. Čas testiranja modela je bil spet bistveno daljši kakor pri učenju modela in tudi bistveno daljši kakor pri metodi najbližjih sosedov. To je posledica tega, da smo pri regresiji z lokalnimi utežmi zgradili poseben regresijski model za vsak testni vzorec.

Iz primerjave korelacijskih koeficientov v tabeli 9 vidimo, da smo najboljše napovedi za testno množico vzorcev dobili pri učno-testni kombinaciji K2. Če primerjamo tabeli 7 in 9, vidimo, da so bile napovedi cen kvadratnega metra stanovanj za učno in testno množico vzorcev boljše pri regresiji z lokalnimi utežmi kakor pri metodi multiple regresije. To je logično, ker smo pri regresiji z lokalnimi utežmi priredili regresijski model testnemu vzorcu, saj smo pri izračunu regresijskih koeficientov upoštevali le tiste nepremičninske transakcije iz učne množice, ki so bile po značilnosti nepremičnine najbolj podobne testni nepremičnini. Ta pristop je tudi ekonomsko bolj upravičen, saj smo vrednotili nepremičnino le na osnovi njej podobnih nepremičnin. Problem regresije z lokalnimi utežmi pa je, da ne vemo, kateri učni vzorci so bili najbolj podobni testnemu vzorcu, kar otežuje izvajanje logičnih sklepov.

5.5 Model za vrednotenje na osnovi regresijskega drevesa

Regresijsko drevo je drevesna metoda za napoved – glej poglavje 3.6.3. Uporabljene spremenljivke so bile v poglavju 5.1.1. V programu *WEKA 3.2* se regresijsko drevo nahaja v knjižnici *weka.classifiers.m5.M5Prime*, če izberemo za opcijo "tip modela" (ang. *model type*) regresijsko drevo. Program *WEKA 3.2* ponuja dve opciji za to metodo:

- faktor rezanja, ki ima vrednosti od 1 do 10,
- velikost izpisa, ki ima vrednosti od 0 do 3.

Za velikost izpisa smo izbrali vrednost 0, pri katerem se izpiše največ podatkov o zgrajenem regresijskem drevesu. S faktorjem rezanja vplivamo na velikost zgrajenega regresijskega drevesa. Nizek faktor rezanja pomeni zelo razvejano drevo, katerega uporabnost je lahko vprašljiva zaradi prevelike prilagoditve učni množici vzorcev. Z večanjem faktorja rezanja se regresijsko drevo zmanjšuje, pri čemer se njegova sposobnost posploševanja načeloma povečuje. V našem primeru smo zgradili regresijska drevesa za različne faktorje rezanja. Najboljši kompromis med velikostjo drevesa in njegovo sposobnostjo za napoved smo dobili, če smo za faktor rezanja izbrali vrednost 3. V tabeli 10 je prikazana kvaliteta regresijskih dreves, ki so bila zgrajena iz učnih množic vzorcev za posamezne učno-testne kombinacije iz poglavja 5.1.2. Kvaliteta modelov je bila ocenjena z merami iz poglavja 3.8 za učno in testno množico vzorcev. V tabeli so prikazane tudi povprečne mere uspešnosti za učne in testne množice vzorcev.

Iz primerjave korelacijskih koeficientov v tabeli 10 vidimo, da smo najboljše napovedi za testno množico vzorcev dobili pri učno-testni kombinaciji K2. Zato je v prilogi 6 prikazan izsek iz regresijskega drevesa za napoved cene kvadratnega metra stanovanj, ki smo ga dobili

z učno-testno kombinacijo vzorcev K2. Na osnovi zgrajenega regresijskega drevesa lahko ocenimo vplivnost posameznih neodvisnih spremenljivk za napoved odvisne spremenljivke. Najvplivnejše so tiste neodvisne spremenljivke, ki nastopajo na vrhu regresijskega drevesa. V našem primeru so bile za napoved cene kvadratnega metra stanovanj najvplivnejše spremenljivke *RAZV_OBC*, *STAROST* in *OBMTR_5*, ki so bile na vrhu regresijskih dreves tudi pri učno-testnih kombinacijah K1 in K3. Napovedi so bile boljše za učno kakor za testno množico, ker so bila zgrajena regresijska drevesa prilagojena podatkom iz učne množice.

Tabela 10: Kvaliteta regresijskih dreves, zgrajenih s programom WEKA 3.2

Učno-testna kombinacija	Mere ujemanja	Obrezovalni faktor = 3	
		Učna množica	Testna množica
K1	Število vzorcev	13292	6646
	Čas učenja	34 s	34 s
	Čas testiranja	~4 s	~2 s
	Korel. koeficient	0,804	0,778
	<i>SAN</i> [SIT]	41224,89	43265,51
	<i>KSKO</i> [SIT]	60103,77	62869,28
	<i>RAN</i> [%]	48,887	51,789
	<i>KRKO</i> [%]	59,426	62,948
K2	Število vzorcev	13287	6651
	Čas učenja	34 s	34 s
	Čas testiranja	~4 s	~2 s
	Korel. koeficient	0,795	0,783
	<i>SAN</i> [SIT]	41830,87	43520,42
	<i>KSKO</i> [SIT]	60988,76	62826,70
	<i>RAN</i> [%]	49,840	51,716
	<i>KRKO</i> [%]	60,653	62,172
K3	Število vzorcev	13297	6641
	Čas učenja	34 s	34 s
	Čas testiranja	~4 s	~2 s
	Korel. koeficient	0,790	0,779
	<i>SAN</i> [SIT]	42459,17	43655,41
	<i>KSKO</i> [SIT]	61552,88	63468,62
	<i>RAN</i> [%]	50,645	51,723
	<i>KRKO</i> [%]	61,270	62,694
Prečno preverjanje	Korel. koeficient	0,797	0,780
	Determ. koeficient	0,635	0,609
	<i>SAN</i> [SIT]	41838,31	43480,45
	<i>KSKO</i> [SIT]	60881,80	63054,86
	<i>RAN</i> [%]	49,791	51,743
	<i>KRKO</i> [%]	60,450	62,605

Opomba: *SAN*, *KSKO*, *RAN*, *KRKO* – glej tabelo 7 in poglavje 3.8.

Če primerjamo tabeli 7 in 10 vidimo, da so bile napovedi cen kvadratnega metra stanovanj za učno in testno množico vzorcev malo boljše pri regresijskem drevesu kakor pri metodi multiple regresije. To je posledica tega, da smo pri izgradnji regresijskega drevesa učno množico razdelili na podmnožice podobnih nepremičnin, ki so imele podobne cene kvadratnega metra. Te podmnožice so tvorile liste regresijskega drevesa. S tem smo odpravili

posplošitev, ki je prisotna pri multiplem regresijskem modelu. Zato je ta pristop tudi ekonomsko upravičen, ker smo vrednotili nepremičnino na osnovi podobnih nepremičnin.

5.6 Model za vrednotenje na osnovi modelnega drevesa

Modelno drevo je drevesna metoda za napoved – glej poglavje 3.6.2. Uporabljene spremenljivke so bile v poglavju 5.1.1. V programu *WEKA 3.2* najdemo regresijsko drevo v knjižnici *weka.classifiers.m5.M5Prime*, če izberemo za opcijo "tip modela" (ang. *model type*) modelno drevo. Program *WEKA 3.2* ponuja tri opcije za to metodo:

- izgradnja zglajenega ali nezglajenega drevesa,
- faktor rezanja, ki ima vrednosti od 1 do 10,
- velikost izpisa, ki ima vrednosti od 0 do 3.

Velikost izpisa in faktor rezanja imata enak pomen kakor pri regresijskem drevesu. Za velikost izpisa smo izbrali vrednost 0. Tudi v tem primeru smo zgradili regresijska drevesa za različne obrezovalne faktorje. Najboljši kompromis med velikostjo drevesa in njegovo sposobnostjo za napoved smo dobili, če smo za faktor rezanja izbrali vrednost 3. Zaradi boljših napovedi smo se odločili za zglajeno modelno drevo. V tabeli 11 je prikazana kvaliteta modelnih dreves, ki so bila zgrajena iz učnih množic vzorcev za posamezne učno-testne kombinacije iz poglavja 5.1.2. Kvaliteta modelov je bil ocenjena z merami iz poglavja 3.8 za učno in testno množico vzorcev. V tabeli so prikazane tudi povprečne mere uspešnosti za učne in testne množice vzorcev.

Iz primerjave korelacijskih koeficientov v tabeli 11 vidimo, da smo najboljše napovedi za testno množico vzorcev dobili pri učno-testni kombinaciji K2. Zato je v prilogi 6 prikazan izsek iz modelnega drevesa za napoved cene kvadratnega metra stanovanj, ki smo ga dobili z učno-testno kombinacijo vzorcev K2. Na osnovi zgrajenega modelnega drevesa je mogoče oceniti pomembnost posameznih neodvisnih spremenljivk za napoved odvisne spremenljivke, podobno kakor pri regresijskem drevesu. Pri napovedi cene kvadratnega metra stanovanj so spet najvplivnejše spremenljivke *RAZV_OBC*, *STAROST* in *OBMTR_5*, ki so na vrhu modelnih dreves v vseh primerih učno-testnih kombinacij. Napovedi so bile boljše za učno kakor za testno množico, ker so bila modelna drevesa prilagojena podatkom iz učne množice.

Če primerjamo tabeli 7 in 11 vidimo, da so bile napovedi cen kvadratnega metra stanovanj za učno in testno množico vzorcev bistveno boljše pri modelnem drevesu kakor pri metodi multiple regresije. Če modelno drevo primerjamo še z ostalimi metodami vidimo, da smo z njim dobili najboljše napovedi cen kvadratnega metra stanovanj za testno množico vzorcev. Vzrok zato je, da je modelno drevo v bistvu model, ki je sestavljen iz več delnih regresijskih modelov za napoved, ki predstavljajo njegove liste. Vsak od teh delnih regresijskih modelov velja le za nepremičnine, ki so si podobne glede na določene značilnosti (na primer, da se nepremičnina nahaja v Ljubljani, ali da je starejša od 40 let itn.). Prednost tega je, da lahko z delnimi modeli bolj natančno opišemo posebnosti, ki jih z enovitim modelom ni mogoče dobro zajeti. Pri tem seveda ni nepomembno, da se podmnožice podobnih nepremičnin med izgradnjo modelnega drevesa tvorijo avtomatsko v procesu minimiziranja ciljne funkcije, ki naj bi ponazarjala napako napovedi – glej poglavje 3.6.2. Tudi v ekonomskem smislu je modelno drevo najbolj logičen pristop izmed vseh predstavljenih, ker z njim vrednotimo

nepremičnino na osnovi podobnih nepremičnin, poleg tega pa pri izračunu cene kvadratnega metra upoštevamo tudi variacije med vrednoteno in podobnimi nepremičninami.

Tabela 11: Kvaliteta modelnih dreves, zgrajenih s programom WEKA 3.2

Učno-testna kombinacija	Mere ujemanja	Obrezovalni faktor = 3	
		Učna množica	Testna množica
K1	Število vzorcev	13292	6646
	Čas učenja	38 s	38 s
	Čas testiranja	~6 s	~3 s
	Korel. koeficient	0,811	0,796
	<i>SAN</i> [SIT]	40494,37	41489,07
	<i>KSKO</i> [SIT]	59189,09	60478,43
	<i>RAN</i> [%]	48,021	49,663
	<i>KRKO</i> [%]	58,521	60,554
K2	Število vzorcev	13287	6651
	Čas učenja	37 s	37 s
	Čas testiranja	~6 s	~3 s
	Korel. koeficient	0,811	0,806
	<i>SAN</i> [SIT]	40050,38	41331,22
	<i>KSKO</i> [SIT]	58890,51	59870,64
	<i>RAN</i> [%]	47,719	49,114
	<i>KRKO</i> [%]	58,566	59,247
K3	Število vzorcev	13297	6641
	Čas učenja	39 s	39 s
	Čas testiranja	~6 s	~3 s
	Korel. koeficient	0,805	0,796
	<i>SAN</i> [SIT]	41178,29	41976,77
	<i>KSKO</i> [SIT]	59611,65	61300,11
	<i>RAN</i> [%]	49,117	49,734
	<i>KRKO</i> [%]	59,338	60,552
Prečno preverjanje	Korel. koeficient	0,809	0,799
	Determ. koeficient	0,654	0,639
	<i>SAN</i> [SIT]	40574,34	41599,02
	<i>KSKO</i> [SIT]	59230,42	60549,72
	<i>RAN</i> [%]	48,286	49,504
	<i>KRKO</i> [%]	58,808	60,118

Opomba: *SAN*, *KSKO*, *RAN*, *KRKO* – glej tabelo 7 in poglavje 3.8.

5.7 Model za vrednotenje na osnovi nevronske mreže

Nevronska mreža je matematična struktura za modeliranje nelinearnih zvez med neodvisnimi in odvisnimi spremenljivkami - glej poglavje 3.6.2. Uporabljene spremenljivke so bile opisane v poglavju 5.1.1. V programu WEKA 3.2 najdemo nevronske mreže v knjižnici *weka.classifiers.neural.neuralNetwork*. Za postavitve nevronske mreže je treba definirati več parametrov. Pomembnejše opcije, ki so na voljo v programu WEKA 3.2, so:

- izbor števila skritih nivojev in nevronov na njih,
- nastavitve spremembe kvalitativnih spremenljivk v binarne numerične spremenljivke,

- nastavitve normalizacije neodvisnih spremenljivk,
- nastavitve parametrov učenja in vztrajnosti,
- nastavitve normalizacije odvisne spremenljivke,
- nastavitve avtomatskega zmanjševanja parametra učenja,
- nastavitve avtomatske vzpostavitve povezav med nevroni,
- morebitna nastavitve velikosti testne množice,
- nastavitve števila iteracij pri učenju (ang. *training time*),
- nastavitve semena generatorja naključnih števil, ki se uporabi za definicijo začetnih uteži v nevronske mreži in za mešanje učnih vzorcev med učenjem.

Nevronske mreže, ki smo jih zgradili za vrednotenje stanovanjskih nepremičnin po tržni metodi, so vsebovale en skriti nivo z različnim številom nevronov na njem. Nevroni na skitem nivoju so imeli logistične aktivacijske funkcije. Najboljše rezultate smo dobili, če je bil skriti nivo sestavljen iz 25 nevronov. Število vhodnih nevronov je bilo enako številu vseh numeričnih in binarnih slamatih neodvisnih spremenljivk. Izhodni nevron, ki ustreza spremenljivki $CENA_NA_m^2$ je bil en sam in je imel linearno aktivacijsko funkcijo. Vzpostavitev povezav med nevroni je bila izvedena avtomatsko. S pomočjo preliminarnih testiranj smo izbrali vrednost 0,05 za parameter učenja in vrednost 0,15 za parameter vztrajnosti. Zaradi hitrejšega učenja nevronske mreže smo odvisno in neodvisne spremenljivke normalizirali. Število iteracij za učenje nevronske mreže, pri katerem smo dobili najboljše napovedi cen kvadratnega metra stanovanj, je bilo enako 200. Za seme generatorja naključnih števil smo poljubno izbrali vrednost 123. Parametra učenja med učenjem nevronske mreže nismo spreminjali. V tabeli 12 je prikazana kvaliteta nevronske mreže s 25 nevroni na enem skitem nivoju, ki so bile zgrajene z 200 iteracijami na osnovi učnih množic vzorcev za posamezne učno-testne kombinacije iz poglavja 5.1.2. Kvaliteta modelov je bila ocenjena z merami iz poglavja 3.8 za učno in testno množico vzorcev. V tabeli so prikazane tudi povprečne mere uspešnosti za učne in testne množice vzorcev.

Iz primerjave korelacijskih koeficientov v tabeli 12 vidimo, da smo najboljše napovedi za testno množico vzorcev dobili pri učno-testni kombinaciji K2. Model nevronske mreže za ta primer ni prikazan v prilogi 6, ker je preobsežen in popolnoma nepregleden, saj vsebuje $35 \times 25 + 25 = 900$ vrednosti za uteži, ki ustrezajo posameznim povezavam med nevroni. Poleg tega na osnovi prikazanih vrednosti za uteži ni mogoče izvajati sklepov o logičnosti povezav med napovedano vrednostjo odvisne spremenljivke in vrednostmi neodvisnih spremenljivk. V tabeli 12 vidimo, da so tudi v tem primeru napovedi boljše za učno kakor za testno množico. Razlog zato je prilagoditev naučene nevronske mreže podatkom iz učne množice.

Če primerjamo tabele 7, 11 in 12 vidimo, da so bile napovedi cen kvadratnega metra stanovanj za učno in testno množico vzorcev boljše pri nevronske mreži kakor pri metodi multiple regresije. Napovedi so bile skoraj tako dobre, kakor pri modelnem drevesu. Ti rezultati potrjujejo znano dejstvo, da so nevronske mreže zelo primerne za modeliranje zahtevnih povezav med odvisnimi in neodvisnimi spremenljivkami, tudi če uporabnik ne pozna dejanske funkcijske oblike teh povezav. Problem uporabe nevronske mreže z ekonomskega stališča je, da napovedano vrednost težko utemeljimo, ker eksplicitno ne

poznamo vplivnosti posameznih neodvisnih spremenljivk. Če bi na primer model nevronske mreže uporabljali kot osnovo za odmero davka od premoženja, se verjetno večina ljudi ne bi strinjala z odmerjeno višino davka le zato, ker je takšno vsoto izračunal računalnik.

Tabela 12: Kvaliteta nevronske mreže, zgrajene s programom WEKA 3.2

Učno-testna kombinacija	Mere ujemanja	25 skr. nevr.; 200 iteracij	
		Učna množica	Testna množica
K1	Število vzorcev	13292	6646
	Čas učenja	610 s	610 s
	Čas testiranja	~6 s	~3 s
	Korel. koeficient	0,812	0,794
	<i>SAN</i> [SIT]	45273,34	46100,40
	<i>KSKO</i> [SIT]	62169,48	63287,26
	<i>RAN</i> [%]	53,688	55,183
	<i>KRKO</i> [%]	61,468	63,366
K2	Število vzorcev	13287	6651
	Čas učenja	585 s	585 s
	Čas testiranja	~6 s	~3 s
	Korel. koeficient	0,809	0,799
	<i>SAN</i> [SIT]	48851,75	50673,30
	<i>KSKO</i> [SIT]	64624,81	66789,92
	<i>RAN</i> [%]	58,206	60,215
	<i>KRKO</i> [%]	64,269	66,094
K3	Število vzorcev	13297	6641
	Čas učenja	575 s	575 s
	Čas testiranja	~6 s	~3 s
	Korel. koeficient	0,812	0,796
	<i>SAN</i> [SIT]	44372,19	45452,46
	<i>KSKO</i> [SIT]	61182,03	63605,78
	<i>RAN</i> [%]	52,927	53,852
	<i>KRKO</i> [%]	60,901	62,830
Prečno preverjanje	Korel. koeficient	0,811	0,796
	Determ. koeficient	0,657	0,634
	<i>SAN</i> [SIT]	46165,76	47408,72
	<i>KSKO</i> [SIT]	62658,78	64560,99
	<i>RAN</i> [%]	54,940	56,417
	<i>KRKO</i> [%]	62,213	64,097

Opomba: *SAN*, *KSKO*, *RAN*, *KRKO* – glej tabelo 7 in poglavje 3.8.

5.8 Primerjava modelov za vrednotenje in izbira najprimernejšega

V poglavjih od 5.2 do 5.7 smo predstavili različne modele za množično vrednotenje stanovanjskih nepremičnin po tržni metodi, ki smo jih zgradili z metodami rudarjenja podatkov. V tabeli 13 so povzete povprečne mere kakovosti napovedi cene kvadratnega metra stanovanj za učne množice vzorcev, v tabeli 14 pa za testne množice vzorcev. K rezultatom metod, ki so opisane v poglavju 3, smo dodali še rezultate, ki smo jih dobili z metodo CIPER. CIPER je kratica za angleški izraz "*Constrained Induction of Polynomial Equations for Regression*". S to metodo je mogoče avtomatsko tvoriti regresijske enačbe, ki vsebujejo tudi

člene višjega reda, brez da bi te člene posebej definirali (kakor na primer pri linearni regresiji). Polinomske enačbe za napoved vrednosti neodvisnih spremenljivk se gradijo postopoma v več korakih. Končni model za napoved vrednosti odvisne spremenljivke predstavlja kompromis med kvaliteto napovedi in kompleksnostjo zgrajene polinomske enačbe. Metoda CIPER je bila razvita na Inštitutu Jožefa Stefana v Ljubljani. Podrobnosti metode CIPER na tem mestu niso navedene, ker jih je mogoče najti v ustrezni literaturi (Todorovski et al. 2004).

Tabela 13: Primerjava modelov za vrednotenje stanovanj za učne množice vzorcev

Model	Mult. regres.	Najbl. sosedi	Regres. z lok. ut.	Regres. drevo	Mod. drevo	Nevr. mreža	CIPER ²⁵
Kor. koef.	0,770	0,958	0,815	0,797	0,809	0,811	0,790
Det. koef.	0,593	0,918	0,664	0,635	0,654	0,657	0,624
SAN [SIT]	44862,2	20337,2	39825,3	41838,3	40574,3	46165,8	-
KSKO [SIT]	64234,3	31203,3	58463,6	60881,8	59230,4	62658,8	-
RAN [%]	53,39	24,20	47,39	49,79	48,29	54,94	-
KRKO [%]	63,78	30,98	58,05	60,45	58,81	62,21	-

Opomba: SAN, KSKO, RAN, KRKO – glej tabelo 7 in poglavje 3.8.

Tabela 14: Primerjava modelov za vrednotenje stanovanj za testne množice vzorcev

Model	Mult. regres.	Najbl. sosedi	Regres. z lok. ut.	Regres. drevo	Mod. drevo	Nevr. mreža	CIPER ²⁵
Kor. koef.	0,769	0,756	0,791	0,780	0,799	0,796	0,787
Det. koef.	0,591	0,571	0,625	0,609	0,639	0,634	0,619
SAN [SIT]	44996,5	46946,3	42047,1	43480,5	41599,0	47408,7	-
KSKO [SIT]	64413,5	66022,2	61691,3	63054,9	60549,7	64561,0	-
RAN [%]	53,55	55,87	50,04	51,74	49,50	56,42	-
KRKO [%]	63,95	65,55	61,25	62,61	60,12	64,10	-

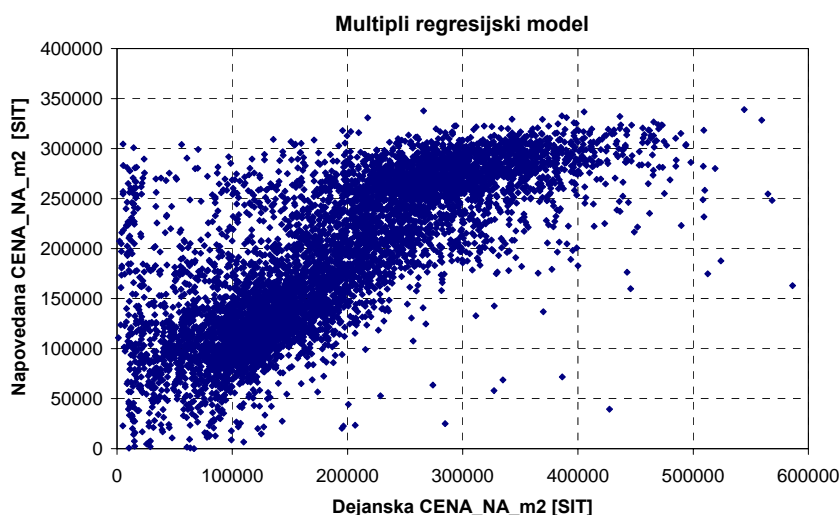
Opomba: SAN, KSKO, RAN, KRKO – glej tabelo 7 in poglavje 3.8.

Iz rezultatov v tabelah 13 in 14 vidimo, da so napovedi cene kvadratnega metra stanovanj boljše za učne, kakor za testne množice vzorcev. To je logično, ker so zgrajeni modeli za vrednotenje stanovanjskih nepremičnin prilagojeni podatkom iz učnih množic. Izkazalo se je, da se je najboljši model po kvaliteti napovedi za učno množico odrezal najslabše pri testiranju s testnimi množicami. Ker naj bi bili modeli za množično vrednotenje nepremičnin zanesljivi

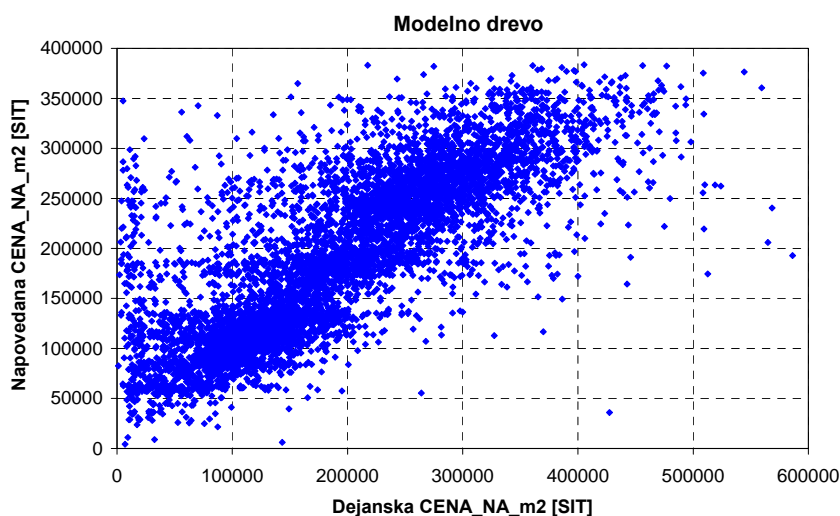
²⁵ Program za uporabo metode CIPER ni del programskega paketa WEKA 3-2, njegovo uporabo nam je omogočil njegov avtor Peter Ljubič. Za napoved cen kvadratnega metra stanovanj smo uporabili osnovno verzijo programa CIPER, ki je omogočala le izpis korelacijskega koeficienta kot mere ujemanja, ne pa tudi mer ujemanja SAN, KSKO, RAN in KRKO. V tabelah 13 in 14 so prikazani rezultati prečnega preverjanja za iste učno-testne množice, kot smo jih uporabili pri drugih metodah. Kakor pri ostalih metodah, smo tudi pri metodi CIPER zgradili modele z različnimi nastavitvami parametrov modela. Najboljše rezultate smo dobili, če smo za vse razen enega parametra modela uporabili privzete (ang. *default*) nastavitve. Le za nastavev števila delnih modelov, ki jih računalnik upošteva pri indukciji polinomov (ang. *beam size*), smo uporabili nastavev 5 namesto privzete nastavitve 1. Opisi posameznih nastavitvev v tekstu niso posebej omenjeni, ker jih je mogoče najti v ustrezni literaturi (Todorovski et al. 2004). Primer polinoma za napoved cene kvadratnega metra stanovanj, ki je bil dobljen z metodo CIPER za učno množico podatkov K2, je prikazan v prilogi 6.

predvsem za vrednotenje nepremičnin, ki niso sodelovale pri njihovi izgradnji, je za oceno kakovosti modelov primernejša medsebojna primerjava rezultatov za testne množice vzorcev.

Če primerjamo med seboj različne mere kakovosti modelov za napoved, vidimo, da imajo dobri modeli v splošnem boljše vrednosti pri vseh merah, čeprav razmerja posameznih mer med različnimi modeli včasih niso dosledna. Zato smo se odločili, da kakovost modelov ocenjujemo na osnovi korelacijskega koeficienta in njegovega ekvivalenta determinacijskega koeficienta. Iz tabele 14 tedaj sledi, da je najboljši model za množično vrednotenje stanovanjskih nepremičnin po tržni metodi modelno drevo, sledita mu nevronska mreža in regresija z lokalnimi utežmi. Vendar imata zadnja dva modela z ekonomskega stališča bistveno pomanjkljivost: njune napovedi je praktično nemogoče logično interpretirati v smislu lastnosti nepremičnin, kar močno zmanjšuje njuno uporabnost. Referenčni multipli regresijski model je sicer po kvaliteti predzadnji, vendar omogoča študij vpliva posameznih lastnosti stanovanja na njegovo vrednost, kar povečuje njegovo uporabnost. Napovedi z modelom CIPER so boljše kakor pri multipli regresiji, kar je tudi logično, saj je polinomski model kompleksnejši od linearnega. Slabše napovedi od modelnega drevesa so posledice tega, da pri metodi CIPER opišemo značilnosti cele množice podatkov z eno samo enačbo, modelno drevo pa je sestav več enačb, ki veljajo za posamezne podmnožice podatkov.



Sl. 20: Uspešnost napovedi cene kvadratnega metra stanovanj za multipli regresijski model



Sl. 21: Uspešnost napovedi cene kvadratnega metra stanovanj za modelno drevo

Na Sl. 20 in Sl. 21 sta prikazani uspešnosti napovedi cene kvadratnega metra stanovanj za referenčni multipli regresijski model in za najboljši model, to je modelno drevo. Oba modela sta bila zgrajena in preskušena z učno-testno kombinacijo K2. Na teh dveh slikah lahko vidimo, da so napovedi z modelnim drevesom dejansko boljše kakor napovedi z multiplim regresijskim modelom, saj je pri modelnem drevesu kar nekaj napovedi nad vrednostjo 350000 SIT/m². V obeh primerih je jasno viden trend med dejansko in napovedano ceno kvadratnega metra stanovanja. Na slikah je mogoče opaziti dve posebnosti: 1.) v višjih cenovnih razredih je kar nekaj napovedi tudi, če je dejanska cena na kvadratni meter zelo nizka, in 2.) raztros napovedi v primerjavi z dejanskimi cenami je sorazmerno velik. Poleg tega je treba še omeniti, da je v primerjavi z modelom linearne regresije pri modelnem drevesu manj napovedi, ki imajo negativno vrednost. Teh napovedi v diagrame nismo vrisali, ker negativna prodajna cena ni smiselna. Vzroki za te posebnosti so naslednji:

- Negativne napovedi cene kvadratnega metra stanovanja bi lahko dobili zaradi slabih, netočnih ali nelogičnih podatkov o lastnostih stanovanj. Ekonomsko gledano je negativna vrednost stanovanja nesmiselna, saj je nerealno pričakovati, da bi racionalni prodajalec plačal kupcu za prenos lastninske pravice na stanovanju. Drugi vzrok za ta pojav gre iskati v specifičnosti uporabljenih matematičnih modelov. Pri metodi najbližjih sosedov ali pri regresijskem drevesu namreč do negativnih napovedi ne more priti, če v bazi podatkov ni stanovanj z negativno transakcijsko ceno.
- Napovedim, ki so bistveno višje od dejanskih cen kvadratnega metra stanovanj v levem delu grafov na Sl. 20 in Sl. 21, so najverjetneje botrovali nenatančni podatki o izmeri stanovanja ali realizirani transakcijski vrednosti. Posledica tega je bila navidezno nizka dejanska cena kvadratnega metra stanovanj tudi v primeru, ko je stanovanje kvalitetno in na zaželeni lokaciji.
- Vzrokov za sorazmerno veliko odstopanje napovedanih in dejanskih cen kvadratnega metra stanovanj je več. Mednje zagotovo sodijo pomanjkanje podatkov o mikrolokaciji nepremičnine, ki bi lahko imeli velik vpliv na njeno vrednost (Štefančič 2003), vendar v bazi DURS-a niso bili na voljo, že prej omenjeni nenatančni podatki o lastnostih nepremičnin v bazi DURS-a in nenazadnje tudi nepopolnost trga nepremičnin, saj se lahko zelo podobne nepremičnine prodajajo za različne cene.

Kljub temu smo mnenja, da so glede na dostopne podatke zgrajeni modeli izkazali določeni potencial za množično vrednotenje stanovanjskih nepremičnin po tržni metodi. Njihova kakovost je primerljiva z modelom za Slovenijo, ki ga je postavil Nose (2003, str. 41-44) in je imel determinacijski koeficient enak 0,667. Zaradi mnogo večjega števila upoštevanih transakcij je bil v našem primeru raztros napovedanih cen kvadratnega metra stanovanj pričakovano večji, kar vpliva tudi na vrednosti determinacijskih koeficientov.

Ker smo zgoraj predstavljene modele zgradili za območje celotne Slovenije, bi se lahko zgodilo, da s tem nismo dobro opisali vpliv lokalnih posebnosti na ceno kvadratnega metra stanovanja. Zato smo z najbolj perspektivno metodo (modelno drevo) in z referenčno metodo (linearna regresija) zgradili ločene modele za vsako območje posebej. Pri tem deljenem pristopu smo pričakovali, da bomo za posamezna območja dobili boljše napovedi, če je na teh območjih cena kvadratnega metra stanovanj odvisna od lokalnih posebnosti.

5.9 Ločeni modeli za vrednotenje za posamezna območja transakcij

5.9.1 Uspešnost napovedi z ločenimi modeli za posamezna območja transakcij

Zaradi razlogov, ki so opisani v poglavju 5.8, smo se odločili, da z metodo multiple regresije in z modelnim drevesom zgradimo ločene modele za tržno vrednotenje stanovanj za vseh deset območij transakcij, ki so opisana v poglavju 4.3.1. Zato smo celotno množico 19938 prečiščenih nepremičninskih transakcij (glej poglavje 5.1) razdelili na deset ločenih podmnožic, ki predstavljajo nabor nepremičninskih transakcij s stanovanji za posamezna območja. Število nepremičninskih transakcij v posameznih območjih je prikazano v tabeli 15. V analizo smo vključili enake spremenljivke kakor pri celovitih modelih z izjemo spremenljivke *OBM_TRAN*, ki je v tem primeru nismo upoštevali.

Tabela 15: Število nepremičninskih transakcij s stanovanji po posameznih območjih

Območje transakcije	Št. trans.	Območje transakcije	Št. trans.
1 – Pomurska regija	412	6 – Dolenjska regija	732
2 – Zg. Štajerska in Koroška	2102	7 – Slovenska obala	1114
3 – Sp. Štajerska in Posavje	3507	8 – Gorenjska regija	1974
4 – Okolica Ljubljane	1848	9 – Notranjska in Kras	618
5 – Ljubljana	7009	10 – Goriška regija	622

Kvaliteto zgrajenih modelov smo ocenili s prečnim preverjanjem, tako da smo množico transakcij za vsako območje s stratificiranim vzorčenjem po spremenljivki *CAS_TRAN* razdelili na tri približno enako velike disjunktno podmnožice. S tem smo zagotovili, da se v vsaki izmed teh podmnožic nahaja približno enako število nepremičninskih transakcij za vsako leto. Iz podmnožic smo sestavili tri kombinacije učnih in testnih množic za izgradnjo in testiranje modelov. Ker je v tabeli 15 mogoče videti, da je v nekaterih območjih zelo malo transakcij, bi se lahko zgodilo, da bi se v teh primerih pri različnih učno-testnih kombinacijah uspešnost napovedi močno spremenjala. Zato smo poleg modelov, ki smo jih dobili s postopkom prečnega preverjanja, za vsako območje zgradili tudi modele za napoved cene kvadratnega metra stanovanj na osnovi vseh nepremičninskih transakcij.

Rezultati napovedi za posamezna območja, ki smo jih dobili s prečnim preverjanjem, in rezultati napovedi, ki smo jih dobili na osnovi vseh transakcij, so prikazani v tabeli 16. Za modele, dobljene s prečnim preverjanjem, so prikazani le povprečni rezultati napovedi za testne množice. Vsi modeli so bili zgrajeni z enakimi nastavitvami kakor prej. Regresijski modeli in modelna drevesa za posamezna območja, ki smo jih dobili na osnovi vseh transakcij na dotičnem območju, so prikazani v prilogi 6.

Rezultati v tabeli 16 kažejo, da je velikost nominalne napake napovedi (*SAN*, *KSKO*) odvisna od območja transakcije. Načeloma je napaka napovedi večja v območjih, kjer so cene stanovanj višje. To je za samo napoved v redu, ker so relativne napake napovedi (*RAN*, *KRKO*) za posamezna območja primerljive. V tabeli 17 so prikazane povprečne vrednosti napak iz tabele 16. Povprečno napako smo izračunali z uteženim povprečjem glede na število transakcij v posameznih območjih. Za primerjavo so v tabeli prikazane tudi napake celovitih modelov za vso Slovenijo.

Tabela 16: Uspešnost napovedi z modeli, ki so bili zgrajeni za posamezna območja

Vrsta modela	OT 1	OT 2	OT 3	OT 4	OT 5	OT 6	OT 7	OT 8	OT 9	OT 10
Multipla regresija – prečno preverjanje na testni množici										
Det. koef. [-]	0,3464	0,2552	0,2906	0,3818	0,2009	0,4768	0,2921	0,3616	0,4465	0,4515
Kor. koef. [-]	0,5886	0,5052	0,5390	0,6179	0,4482	0,6905	0,5405	0,6013	0,6682	0,6720
<i>SAN</i> [SIT]	25130,79	31678,58	24651,04	42391,36	57155,99	26763,97	57616,43	45344,34	28808,24	31435,53
<i>KSKO</i> [SIT]	36543,40	43015,80	33754,01	59277,47	80525,85	41620,24	76732,99	59963,72	36896,80	42198,27
<i>RAN</i> [%]	85,20	81,27	80,40	72,59	86,37	59,19	82,99	74,93	75,51	68,53
<i>KRKO</i> [%]	81,67	86,44	84,11	78,85	89,43	72,77	84,24	80,07	75,85	73,89
Multipla regresija – učenje z vsemi transakcijami za posamezno območje										
Det. koef. [-]	0,5071	0,2875	0,2999	0,4064	0,2065	0,5367	0,3142	0,3950	0,4976	0,5069
Kor. koef. [-]	0,7121	0,5362	0,5476	0,6375	0,4544	0,7326	0,5605	0,6285	0,7054	0,7120
<i>SAN</i> [SIT]	22728,38	30996,14	24520,65	41441,00	56974,96	25484,06	56160,40	44289,53	27396,11	29330,14
<i>KSKO</i> [SIT]	31422,90	42003,54	33561,74	57926,31	80201,40	38888,93	75430,13	58155,72	34583,95	40036,58
<i>RAN</i> [%]	77,42	79,55	79,99	70,97	86,10	56,52	80,95	73,32	71,69	63,91
<i>KRKO</i> [%]	70,21	84,41	83,68	77,05	89,08	68,06	82,81	77,78	70,88	70,22
Modelno drevo – prečno preverjanje na testni množici										
Det. koef. [-]	0,5188	0,2811	0,3055	0,3665	0,2610	0,5167	0,2835	0,3908	0,5316	0,4534
Kor. koef. [-]	0,7203	0,5302	0,5527	0,6054	0,5109	0,7188	0,5325	0,6251	0,7291	0,6734
<i>SAN</i> [SIT]	22505,98	30961,02	23770,35	43119,59	54720,76	26917,66	57858,68	43807,95	25789,45	31053,10
<i>KSKO</i> [SIT]	31400,52	42397,87	33548,30	59967,98	77436,75	39942,56	77222,33	58676,09	33606,77	42261,04
<i>RAN</i> [%]	76,35	79,43	77,53	73,84	82,69	59,55	83,32	72,40	67,61	67,68
<i>KRKO</i> [%]	70,36	85,19	83,60	79,77	85,99	69,79	84,76	78,38	69,07	74,02
Modelno drevo – učenje z vsemi transakcijami za posamezno območje										
Det. koef. [-]	0,6288	0,3399	0,4177	0,5139	0,2779	0,5634	0,2996	0,4776	0,6076	0,5458
Kor. koef. [-]	0,7930	0,5830	0,6463	0,7169	0,5272	0,7506	0,5474	0,6911	0,7795	0,7388
<i>SAN</i> [SIT]	20029,97	29412,86	22141,40	38705,07	54125,20	25827,83	57147,13	40497,10	24243,75	28179,64
<i>KSKO</i> [SIT]	27473,06	40431,83	30621,93	52462,53	76507,22	37761,01	76224,49	54054,40	30585,05	38426,46
<i>RAN</i> [%]	68,23	75,48	72,23	66,28	81,79	57,28	82,37	67,04	63,44	61,41
<i>KRKO</i> [%]	61,38	81,25	76,35	69,78	84,98	66,09	83,69	72,30	62,69	67,39

Opombe: OT ... Območje transakcije
SAN, *KSKO*, *RAN*, *KRKO* – glej tabelo 7 in poglavje 3.8.

Tabela 17: Povprečna uspešnost napovedi z modeli, ki so bili zgrajeni za posamezna območja

Povprečna napaka za vsa območja	Ločeni modeli za posamezna območja; prečno preverjanje na testnih množicah		Ločeni modeli za posamezna območja; vse transakcije		Celoviti modeli za vso Slovenijo; prečno preverjanje na testnih množicah	
	Mult. reg.	Mod. dr.	Mult. reg.	Mod. dr.	Mult. reg.	Mod. dr.
Det. koef. [-]	0,289	0,324	0,311	0,388	0,591	0,639
Kor. koef. [-]	0,532	0,565	0,551	0,617	0,769	0,799
<i>SAN</i> [SIT]	42780,1	41468,3	42143,2	39805,3	44996,5	41599,0
<i>KSKO</i> [SIT]	59239,7	57748,5	58265,1	55118,3	64413,5	60549,7
<i>RAN</i> [%]	80,27	77,72	78,98	74,20	53,55	49,50
<i>KRKO</i> [%]	84,30	82,27	82,74	77,91	63,95	60,12

Opombe: *SAN*, *KSKO*, *RAN*, *KRKO* – glej tabelo 7 in poglavje 3.8.

Ko primerjamo rezultate v tabeli 17, se moramo omejiti le na primerjavo napak *SAN* in *KSKO*. Korelacijski in determinacijski koeficient za samo primerjavo nista merodajna. Razlog zato je naslednji: ko smo zgradili modele za posamezna območja, je bil lahko korelacijski (determinacijski) koeficient za testno množico mnogo slabši, kakor v primeru, ko smo zgradili

enoten model za celo Slovenijo. Zaradi neenakomernih cenovnih razmer v Sloveniji se je namreč lahko zgodilo, da je bila korelacija dejanskih in s celovitim modelom napovedanih cen dobra, čeprav je celoviti model zajel posebnosti posameznih območij mnogo slabše kakor lokalizirani modeli. Za primerjavo kakovosti napovedi je pomembnejše, da poznamo dejanska odstopanja napovedanih in dejanskih cen, kakor pa, da poznamo korelacijske koeficiente.

Iz rezultatov v tabeli 17 vidimo, da so bile napovedi z ločenimi modeli v povprečju boljše od napovedi s celovitimi modeli. Čeprav odstopanje ni veliko, je bila izboljšava dosledna: tako pri modelu linearne regresije, kakor pri modelnem drevesu. To pomeni, da smo z ločenimi modeli uspeli boljše modelirati nekatere posebnosti lokalnih nepremičninskih trgov.

Analizo stanovanjskega trga nepremičnin smo zaključili s študijem inflacijskega vpliva in realnega časovnega trenda na gibanja cen stanovanjskih nepremičnin.

5.9.2 Študij inflacijskega vpliva in časovnega trenda gibanja cen

Problema časovno spremenljivih vrednosti stanovanjskih nepremičnin smo se dotaknili že v poglavju 5.1.1, ko smo izbirali obliko odvisne spremenljivke $CENA_NA_m^2$ za nadaljnje analize. Že takrat smo opazili, da je največji delež k časovni spremenljivosti vrednosti prispeval inflacijski vpliv. Na tem mestu pa smo vpliv inflacije in časovnega trenda na spreminjanje vrednosti stanovanj opisali bolj sistematično. To smo storili tako, da smo najprej za vsako izmed desetih območij nepremičninskih transakcij izračunali povprečno vrednost spremenljivke $CENA_NA_m^2$ za vsako izmed let od 1999 do 2002. Pri tem smo upoštevali le transakcije, ki smo jih uporabili že pri izgradnji modelov za vrednotenje. Tako smo dobili 40 povprečnih vrednosti transakcij (10 območij \times 4 leta). Nato smo vse povprečne vrednosti, ki so se nanašale na leta 2000, 2001 in 2002, spremenili v realne vrednosti cen glede na leto 1999 s pomočjo indeksov rasti drobnoprodajnih cen (Statistični letopis RS 2002). Na ta način smo iz vseh povprečnih vrednosti odpravili vpliv inflacije. S pomočjo tako dobljenih podatkov smo zgradili multipli regresijski model, ki je imel naslednjo obliko:

$$RCENA_NA_m^2 = \beta_0 + \beta_1 \cdot LETO + \sum_{i=2}^{10} \beta_i \cdot OBMTR_i + \sum_{j=2}^{10} \gamma_j \cdot (LETO \cdot OBMTR_j) \quad (33).$$

$RCENA_NA_m^2$ je bila realna cena stanovanja za kvadratni meter, $OBMTR_i$ je bila indikatorska spremenljivka za posamezna območja transakcij: $i=2, \dots, 10$. Indikatorska spremenljivka $OBMTR_1$ je služila za referenco. Če je pri spreminjanju povprečnih cen prevladujoč inflacijski vpliv, bi morali biti regresijski koeficienti pri spremenljivkah $LETO$ in $LETO \times OBMTR_i$ neznačilno različni od nič. Če pa je prevladujoč vpliv časovnega trenda, bi morali biti ti koeficienti značilno različni od nič. Na osnovi enačbe (33) smo s programom *SPSS 10.0* zgradili tri regresijske modele za tri množice podatkov:

- za osnovo smo uporabili dejanske vrednosti cen iz baze podatkov DURS-a;
- za osnovo smo uporabili vrednosti cen, ki smo jih v poglavju 5.9.1 napovedali z regresijskimi modeli, dobljenimi z upoštevanjem vseh transakcij;
- za osnovo smo uporabili vrednosti cen, ki smo jih v poglavju 5.9.1 napovedali z modelnimi drevesi, dobljenimi z upoštevanjem vseh transakcij.

V tabeli 18 so prikazani regresijski koeficienti in njihove stopnje značilnosti za vse tri modele.

Tabela 18: Regresijski koeficienti v modelih za študij inflacijskega vpliva

Spremenljivka	Originalni podatki o cenah na m ²		Cene napovedane z multipl. regresijo		Cene napovedane z modelnim drevesom	
	popravljeni $R^2=0,967$		popravljeni $R^2=0,981$		popravljeni $R^2=0,983$	
	Regr. koef.	Znač.	Regr. koef.	Znač.	Regr. koef.	Znač.
<i>konstanta</i>	117699,24	0,000	123383,49	0,000	119214,77	0,000
<i>LETO</i>	-8019,06	0,068	-9753,66	0,005	-8446,86	0,010
<i>OBMTR 2</i>	-14085,03	0,392	-33012,24	0,013	-23249,34	0,056
<i>OBMTR 3</i>	-20114,39	0,226	-23589,42	0,066	-22810,77	0,060
<i>OBMTR 4</i>	64257,89	0,001	49835,53	0,001	56915,85	0,000
<i>OBMTR 5</i>	129067,25	0,000	120715,42	0,000	121197,24	0,000
<i>OBMTR 6</i>	6786,33	0,678	11348,94	0,360	16181,84	0,173
<i>OBMTR 7</i>	102115,38	0,000	84570,67	0,000	87305,72	0,000
<i>OBMTR 8</i>	22185,24	0,184	20343,96	0,109	27249,66	0,027
<i>OBMTR 9</i>	-33340,12	0,052	-43253,51	0,002	-27957,13	0,024
<i>OBMTR 10</i>	37913,60	0,029	27876,10	0,032	36370,50	0,005
<i>LETO × OBMTR 2</i>	6172,73	0,306	11851,65	0,014	9018,06	0,043
<i>LETO × OBMTR 3</i>	5313,72	0,377	6377,28	0,165	6112,03	0,159
<i>LETO × OBMTR 4</i>	3958,23	0,509	8232,68	0,077	6063,44	0,163
<i>LETO × OBMTR 5</i>	3480,94	0,561	6013,16	0,189	5883,94	0,175
<i>LETO × OBMTR 6</i>	3396,10	0,570	2054,41	0,647	520,86	0,902
<i>LETO × OBMTR 7</i>	4693,65	0,434	10149,81	0,033	9324,79	0,037
<i>LETO × OBMTR 8</i>	10778,67	0,082	11330,67	0,019	9029,80	0,043
<i>LETO × OBMTR 9</i>	9173,89	0,134	12148,86	0,012	7401,84	0,092
<i>LETO × OBMTR 10</i>	-3495,83	0,559	-949,23	0,832	-3776,54	0,377

Opombe: Regr. koef. ... Vrednost koeficienta pripadajoče regresijske spremenljivke
 Znač. ... Značilnost regresijskega koeficienta

V tabeli 18 vidimo, da so regresijski koeficienti pri spremenljivkah *LETO* in *LETO × OBMTR_i* neznačilni (vsaj pri stopnji 0,5 %) v vseh treh primerih, medtem ko je kar nekaj indikatorskih spremenljivk za območja transakcij močno značilnih. Med leti 1999 in 2002 je bil torej v povprečju prevladujoči vpliv časa na spremembo vrednosti stanovanj inflacijski, kar smo ugotovili tudi v poglavju 5.1.1. Vpliv časovnega trenda je bil manj izrazit.

6 Zaključek

Cilj magistrskega dela je bila analiza značilnosti delne baze nepremičninskih transakcij v Sloveniji, iskanje zanimivih vzorcev v tej bazi in postavitev modela za množično vrednotenje stanovanjskih nepremičnin po tržni metodi, ki bi upošteval vpliv stvarnih lastnosti nepremičnin in časovne oddaljenosti predhodnih transakcij na ceno nepremičnine. Reševanja danega problema smo se lotili v skladu s CRISP-DM metodologijo, ki je sestavljena iz šestih korakov: razumevanje problema, razumevanja podatkov, priprave podatkov za uporabo, postavitev modelov, ovrednotenje kvalitete modelov in vključitev modelov v poslovno okolje (Chapman et al. 2004, Jorge et al. 2003, str. 63-78). Pri izdelavi magistrskega dela posameznih korakov nismo ločili, ampak smo jih med seboj prepletali.

V poglavju 2 smo se najprej seznanili z značilnostmi trga nepremičnin. Trg nepremičnin se razlikuje od ideala trga čiste in popolne konkurence zaradi dejavnikov, kot so nehomogenost proizvodov, nemobilnost nepremičnin, izrazite informacijske asimetrije na trgu itn., čeprav zanj veljajo ekonomske zakonitosti, ki so povezane s ponudbo in pričakovanjem. V nadaljevanju smo pregledali zakone, ki urejajo razmere na trgu nepremičnin v Sloveniji, in vzeli pod drobnogled metode, ki se uporabljajo za vrednotenje nepremičnin. V splošnem se vrednosti nepremičnin ocenjuje s tremi metodami (tržna metoda, dohodkovna metoda in stroškovna metoda), ki jih lahko uporabimo za posamično in množično vrednotenje nepremičnin. Za množično vrednotenje stanovanjskih nepremičnin smo izbrali tržno metodo, saj so v Sloveniji stanovanja kot najbolj trgovane nepremičnine najbolj podvržena delovanju tržnih zakonitosti, zato pričakujemo, da na trgu dosežena cena odraža njihovo dejansko vrednost. Razumevanje problema smo v poglavju 3 sklenili s pregledom statističnih metod in metod rudarjenja podatkov, ki smo jih uporabili za analizo in pripravo podatkov ter za postavitev modelov za vrednotenje nepremičnin.

Osnovo za izdelavo našega magistrskega dela so predstavljali podatki o izvršenih transakcijah z nepremičninami v Republiki Sloveniji v letih 1999 do 2000, ki se nahajajo v delni bazi nepremičninskih transakcij Davčne uprave Republike Slovenije. Baza nepremičninskih transakcij DURS-a je zbirna baza, v katero se vnašajo podatki iz različnih virov – občin. Posledica tega sta nepopolnost in nedoslednost podatkov v bazi. V poglavju 4.2 smo rangirali občine glede na popolnost vnosov in ugotovili, da so občine Celje, Ljubljana, Jesenice, Štore in Hodoš posredovale najbolj popolne, občine Jezersko, Odranci, Krško, Loški Potok in Dol pri Ljubljani pa najmanj popolne podatke o izvršenih nepremičninskih transakcijah. Preden smo se lotili analize podatkov, smo skušali odpraviti čim več nedoslednosti pri podatkih v bazi, ki so povezane s komunalno opremljenostjo, starostjo, lego nepremičnin itn. Iz podatkov v bazi nepremičninskih transakcij smo ugotovili, da so bile prijavljene vrednosti v transakcijah udeleženih nepremičnin za potrebe odmere davka na promet nepremičnin kar nekajkrat popravljene navzgor (glej poglavje 4.3.1). To nam daje slutiti, da prijavljena vrednost ni vedno dobra indikacija tržne vrednosti. Zato smo za vrednost nepremičnine upoštevali višjo od dveh vrednosti: prijavljene ali odmerjene vrednosti nepremičnine.

Najpomembnejša podatka o nepremičnini sta njena izmera in vrednost. Zato smo v poglavju 4.3.1 primerjali srednje vrednosti in mediane za izmero in ceno kvadratnega metra nepremičnin za deset območij v Sloveniji. Ugotovili smo, da je za medsebojno primerjavo izmer in cen kvadratnega metra nepremičnin zaradi nedoslednosti vnosov mediana boljše merilo od srednje vrednosti. Glede na ceno kvadratnega metra izstopajo območja, kjer je večje povpraševanje po nepremičninah povzročilo višje cene: slovenska obala ter Ljubljana z nekaterimi sosednjimi občinami. Cene zemljišč na bolj zelenih lokacijah so višje, velikosti zemljišč na teh lokacijah (z izjemo okolice Ljubljane) so načeloma manjše. Zato lahko sklepamo, da so zemljišča omejena dobrina na teh lokacijah, kar potrjuje splošno načelo, da je vzrok za velik delež razlike v cenah stavbnih nepremičnin prav lokacija. Cene komunalno opremljenih zemljišč so v skladu s pričakovanji bistveno višje kot cene ostalih zemljišč. Očitno je tudi, da je slovenska obala za marsikoga še vedno priljubljena počitniška destinacija, saj so cene vikendov na tem območju najvišje.

Proces spoznavanja podatkov v bazi nepremičninskih transakcij smo zaključili s tem, da smo z uporabo apriori algoritma skušali ugotoviti, ali obstajajo med zapisi v bazi kakšna

povezovalna pravila. Dobljena povezovalna pravila lahko razdelimo v dve skupini. V prvi skupini so bolj splošna povezovalna pravila, ki se nanašajo na komunalno opremljenost nepremičnin. Le-ta niso močno zanimiva, saj je jasno, da predstavlja komunalna opremljenost skupek izboljšav, kot so vodovod, elektrika, kanalizacija itn. V drugi skupini so pravila, ki se nanašajo na lastnosti specifičnih nepremičnin. Tako smo na primer ugotovili, da je večina stanovanj manjših od 70 m², in da so bila zgrajena po letu 1960, ter da večina garaž ni opremljenih z vodovodom in ogrevanjem – glej poglavje 4.3.2.

Pripravo podatkov za izvajanje analiz smo v veliki meri izvršili vzporedno s spoznavanjem podatkov. Preden smo postavili različne modele za množično vrednotenje stanovanjskih nepremičnin po tržni metodi, smo obstoječe transakcije v bazi podatkov dopolnili še z nekaterimi podatki o geografskih in splošnih značilnostih makrookolja, ki omogočajo postavitev boljših modelov za vrednotenje nepremičnin. Tako smo dodali geografske koordinate občin, razvitost občin ter prisotnost avtocestnih povezav in njihovo povezavo z Ljubljano – glej poglavje 4.4.

Ko smo prečistili bazo podatkov in vnesli vanjo nekatere dopolnitve, smo se v poglavju 5 lotili izdelave modelov za množično vrednotenje stanovanjskih nepremičnin po tržni metodi. Z modeli smo napovedovali ceno kvadratnega metra stanovanja v odvisnosti od njegovih lastnosti. Modele za vrednotenje smo zgradili s šestimi metodami: metoda multiple regresije, metoda najbližjih sosedov, regresija z lokalnimi utežmi, regresijsko drevo, modelno drevo in nevronska mreža. Za primerjavo smo dodali tudi modele, zgrajene z metodo CIPER. Vendar popolne primerjave z drugimi modeli nismo mogli izvesti, ker program *WEKA 3-2* ne vsebuje knjižnice za metodo CIPER, uporabljeni program za metodo CIPER pa ne izpisuje enakih mer podobnosti kakor program *WEKA 3-2*.

Model, dobljen z metodo multiple regresije, je služil kot referenca za primerjavo. Kvaliteto zgrajenih modelov in primerjavo med njimi smo izvršili s pomočjo prečnega preverjanja. S primerjavo determinacijskih koeficientov smo ugotovili, da je najboljši model za množično vrednotenje stanovanj po tržni metodi modelno drevo, ki je uporaben tudi z ekonomskega stališča, ker omogoča študij vpliva posameznih lastnosti stanovanja na njegovo vrednost. Sledita mu nevronska mreža in regresija z lokalnimi utežmi. Pomanjkljivost teh dveh modelov je, da je z njima dobljene vrednosti nepremičnin nemogoče logično interpretirati v smislu lastnosti nepremičnin. Referenčni multipli regresijski model je sicer po kvaliteti predzadnji, vendar podobno kakor modelno drevo omogoča študij vpliva posameznih lastnosti stanovanja na njegovo vrednost, kar povečuje njegovo uporabnost. Model, dobljen z metodo CIPER je boljši od enostavnega regresijskega modela. S stališča pojasnjevalne vrednosti je metoda CIPER podobna metodi multiple regresije, s to prednostjo, da je z metodo CIPER mogoče zgraditi polinomske enačbe brez vnaprejšnje definicije členov z višjimi potencami. Pri ločenih modelih za vsako območje posebej smo dobili manjšo napako napovedi kakor pri enotnih modelih za celo Slovenijo.

S primerjavo napovedanih in dejanskih cen kvadratnega metra stanovanj v poglavju 5.8 smo ugotovili, da je raztros napovedanih cen sorazmerno velik. Takšen raztros je posledica netočnih in nelogičnih podatkov o lastnostih stanovanj v bazi nepremičninskih transakcij, specifičnosti uporabljenih modelov za vrednotenje stanovanj, pomanjkanja podatkov o mikrolokaciji nepremičnin in nepopolnosti trga nepremičnin. Kljub temu je kakovost

postavljenih modelov za napoved cen kvadratnega metra stanovanj primerljiva z modeli drugih avtorjev (Nose 2003), čeprav smo v našem primeru zgradili modele za bistveno več nepremičninskih transakcij, kar lahko poveča raztros napovedanih cen kvadratnega metra stanovanj. S študijem časovnega vpliva na vrednost stanovanjskih nepremičnin smo ugotovili, da je bil med leti 1999 in 2002 prevladujoči vpliv časa na vrednost nepremičnin inflacijski.

Za konec bomo podali še nekaj besed o primernosti naših modelov za vključitev v poslovno okolje. Namen tega magistrskega dela ni bil konkurirati dejavnosti Geodetske uprave Republike Slovenije, ki se že kar nekaj let ukvarja z vrednotenjem nepremičnin in ima dostop tudi do drugih baz podatkov, ki so koristne pri vrednotenju nepremičnin. Zaradi omejitev uporabljene baze nepremičninskih transakcij bi bilo po našem prepričanju tudi preveč smelo predlagati zgrajene modele za uporabo v davčne namene, kot pomoč cenilcem, za spremljanje in analiziranje trga nepremičnin ali za izboljševanje informiranosti na trgu nepremičnin. Na predstavljene modele za vrednotenje nepremičnin je treba gledati predvsem kot na primerno alternativo obstoječim pristopom za množično vrednotenje nepremičnin po tržni metodi. Seveda pa smo prepričani, da je na osnovi primernejše baze podatkov predstavljene modele mogoče enostavno dopolniti tudi za prej omenjene namene. Uporaba opisanih in novih modelov (CIPER) na dopolnjeni bazi bi lahko predstavljala prihodnje usmeritev raziskav.

7 Literatura in viri

7.1 Literatura

1. Ben-Shahar Danny: Theoretical and Empirical Analysis of the Multiperiod Pricing Pattern in the Real Estate Market. *Journal of Housing Economics*, 11 (2002), str. 95-107.
2. Brigham Eugene F., Daves Phillip R.: *Intermediate Financial Management*, seventh edition. Fort Worth: South-Western/Thomson Learning, 2002, 988 str.
3. Cestnik Bojan, Lavrač Nada, Flach Peter, Gamberger Dragan, Kline Mihael: Data mining for decision support: supporting marketing decisions through subgroup discovery. Poglavje v knjigi: *Data mining and decision support, Integration and collaboration* (ur. Mladenčić D., Lavrač N., Bohanec M., Moyle S.). Norwell, Massachusetts: Kluwer Academic Publishers, 2003, str. 91-106.
4. Cirman Andreja, Čok Mitja, Lavrač Ivo, Zakrajšek Petra: *Poslovanje z nepremičninami, zapiski predavanj*. Ljubljana: Ekonomska fakulteta, 2000, 128 str.
5. Clapp John M., Giaccotto Carmelo: Estimating Price Indices for Residential Property: A Comparison of Repeat Sales and Assessed Value Methods. *Journal of the American Statistical Association*, 87 (1992), 418, str. 300-306.
6. Chapman Peter, Clinton Julian, Kerber Randy, Khabaza Thomas, Reinartz Thomas, Shearer Colin, Wirth Rüdiger: *CRISP-DM 1.0: Step-by-step data mining guide*. CRISP-DM consortium. [URL: <http://www.crisp-dm.org>], 16. 1. 2004.
7. Eckert Joseph K.: *Property Appraisal and Assessment Administration*. Chicago: International Association of Assessing Officers, 1990, 716 str.
8. Forgey Fred A., Rutherford Ronald C., Springer Thomas M.: Search and Liquidity in Single-Family Housing. *Real Estate Economics*, 24 (1996), 3, str. 273-291.

9. Han Jiawei, Kamber Micheline: Data Mining: Concepts and Techniques. San Francisco: Morgan Kaufman Publishers, 2001, 550 str.
10. Hand David, Mannila Heikki, Smyth Padharic: Principles of Data Mining. Cambridge, Massachusetts: A Bradford Book - Massachusetts Institute of Technology, 2001, 546 str.
11. Hastie Trevor, Tibshirani Robert, Friedman Jerome: The Elements of Statistical Learning, Data Mining, Inference, and Prediction. New York: Springer-Verlag, 2001, 533 str.
12. Inmon William H.: Building the Data Warehouse, second edition. New York: John Wiley & Sons, 1996, 401 str.
13. Ioannides Yannis M.: Interactive property valuation. Journal of Urban Economics, 53 (2003), str. 145-170.
14. Iversen Edwin S. Jr.: Spatially Disaggregated Real Estate Indices. Journal of Business & Economic Statistics, 19 (2001), 3, str. 341-357.
15. Jansen Christian, Söderberg Bo: Estimating Market Prices and Assessed Value for Income Properties. Urban studies, 36 (1999), 2, str. 359-367.
16. Jorge Alípio, Moyle Steve, Blockeel Hendrik, Voss Angi: Data processes and collaboration principles. Poglavlje v knjigi: Data mining and decision support, Integration and collaboration (ur. Mladenčić D., Lavrač N., Bohanec M., Moyle S.). Norwell, Massachusetts: Kluwer Academic Publishers, 2003, str. 63-78.
17. Jošt Barbara: Obdavčenje pri prenosu nepremičnin. Diplomsko delo. Ljubljana: Ekonomska fakulteta, 2002, 47 str.
18. Knight John R., Sirmans C. F., Gelfand Alan E., Ghosh Sujit K.: Analyzing Real Estate Data Problems Using the Gibbs Sampler. Real Estate Economics, 26 (1998), 3, str. 469-492.
19. Kožar Anton, Marinšek Nives: Zakon o nepremičninskem posredovanju s komentarjem. Ljubljana: Center za poslovno usposabljanje - GZS, 2003, 259 str.
20. Lavrač Nada, Bohanec Marko: Integration of data mining and decision support. Poglavlje v knjigi: Data mining and decision support, Integration and collaboration (ur. Mladenčić D., Lavrač N., Bohanec M., Moyle S.). Norwell, Massachusetts: Kluwer Academic Publishers, 2003, str. 37-48.
21. Lavrač Nada, Grobelnik Marko: Data mining. Poglavlje v knjigi: Data mining and decision support, Integration and collaboration (ur. Mladenčić D., Lavrač N., Bohanec M., Moyle S.). Norwell, Massachusetts: Kluwer Academic Publishers, 2003, str. 3-14.
22. M5' (M5 Prime). [URL: http://grb.mnsu.edu/grbts/doc/manual/M5_M5_Prime.html], 5. 12. 2003.
23. Matejčič Katarina: Prevladujejo dvosobna stanovanja. Finance, 4. 9. 2003, internetna izdaja. [URL: <http://www.finance-on.net/danes>], 5. 9. 2003.
24. Nose Peter: Cene rabljenih stanovanj v Sloveniji. Diplomsko delo. Ljubljana: Ekonomska fakulteta, 2003, 61 str.

25. Nagode Petra: Vrednotenje nepremičnin za potrebe obdavčenja v Sloveniji. Magistrsko delo. Ljubljana: Ekonomska fakulteta, 1999, 99 str.
26. Pace Kelley R., Barry Ronald, Gilley Otis W., Sirmans C. F.: A method for spatial-temporal forecasting with an application to real estate prices. *International Journal of Forecasting*, 16 (2000), str. 229-246.
27. Pavlov Andrey D.: Space-Varying Regression Coefficients: A Semi-parametric Approach Applied to Real Estate Markets. *Real Estate Economics*, 28 (2000), 2, str. 249-283.
28. Pindyck Robert S., Rubinfeld Daniel L.: *Econometric models and economic forecasts*. Singapore: Irwin/McGraw-Hill, 1998, 635 str.
29. Pirih Majda: Finančna analiza panoge posredovanja z nepremičninami. Diplomsko delo. Ljubljana: Ekonomska fakulteta, 2004, 50 str.
30. Plaut Pnina O., Plaut Steven E.: Endogenous Identification of Multiple Housing Price Centers in Metropolitan Areas. *Journal of Housing Economics*, 7 (1998), str. 193-217.
31. Prašnikar Janez, Debeljak Žiga: *Ekonomski modeli za poslovno odločanje*. Ljubljana: Gospodarski vestnik, 1998, 435 str.
32. Rotovnik Klavdija: Načini vrednotenja nepremičnin v Sloveniji. Diplomsko delo. Ljubljana: Ekonomska fakulteta, 1997, 41 str.
33. Rovan Jože, Sambt Jože: Socio-economic Differences Among Slovenian Municipalities: A Cluster Analysis Approach. *Developments in Applied Statistics, Metodološki zvezki št. 19*, Ur. Ferligoj Anuška in Mrvar Andrej. Ljubljana: Fakulteta za družbene vede, 2003, str. 265-278.
34. Rovan Jože, Turk Tomaž: *Analiza podatkov z SPSS za Windows*. Ljubljana: Ekonomska fakulteta, 2001, 262 str.
35. Sharma Subhash: *Applied Multivariate Techniques*. New York: John Willey & Sons, 1996. 493 str.
36. Štefančič Mateja: Analiza trga nepremičnin v središču Ljubljane. Diplomsko delo. Ljubljana: Ekonomska fakulteta, 2003, 39 str.
37. Štěpánková Olga, Aubrecht Peter, Kouba Zdenek, Mikšovský Petr: Preprocessing for data mining and decision support. Poglavlje v knjigi: *Data mining and decision support, Integration and collaboration* (ur. Mladenec D., Lavrač N., Bohanec M., Moyle S.). Norwell, Massachusetts: Kluwer Academic Publishers, 2003, str. 107-118.
38. Todorovski Ljupčo, Ljubič Peter, Džeroski Sašo: Inducing polynomial equations for regression. Poglavlje v knjigi: *Machine learning: ECML 2004: 15th European Conference on Machine Learning, Pisa, Italy, September 20-24, 2004: proceedings*, (Lecture notes in computer science, Lecture notes in artificial intelligence, vol. 3201). Berlin, Heidelberg, New York: 2004, str. 441-452.
39. Witten Ian H., Frank Eibe: *Data Mining, Practical Machine Learning Tools and Techniques with Java Implementations*. San Francisco: Morgan Kaufman Publishers, 2000, 371 str.

40. Zakrajšek Franc J.: Ekspertni sistemi in vrednotenje nepremičnin. Zbornik referatov Nacionalna strategija razvoja Slovenije / 12. Sedlarjevo srečanje, Postojna 28. in 29. maj 1992, Ur. Stančič Ivan. Postojna: Zveza društev urbanistov Slovenije, 1993, str. 159-165.
41. Zakrajšek Franc J.: Razvoj množičnega vrednotenja nepremičnin: izhodišča in koncept informacijskega sistema. Ljubljana: Urbanistični inštitut RS, 1996, 134 str.
42. Zakrajšek Petra: Trg nepremičnin v petih evropskih državah. Diplomsko delo. Ljubljana: Ekonomska fakulteta, 1998, 57 str.

7.2 Viri

1. Baza podatkov DURS-a o nepremičninskih transakcijah v Republiki Sloveniji od leta 1999 do leta 2002. Ljubljana: Davčna uprava Republike Slovenije, 2003.
2. Občine Republike Slovenije – preglednica. Statistični urad Republike Slovenije. [URL: <http://www.stat.si/katalogrds/podstrani/tabele/tabela1.html>], 6. 11. 2003.
3. Občine Republike Slovenije – komentar. Statistični urad Republike Slovenije. [URL: <http://www.stat.si/katalogrds/podstrani/komentar1.html>], 6. 11. 2003.
4. Občine Republike Slovenije – zemljevid. Statistični urad Republike Slovenije. [URL: <http://www.stat.si/katalogrds/podstrani/karta1.html>], 6. 11. 2003.
5. Obligacijski zakonik – OZ (Uradni list RS, št. 83/2001).
6. Shema avtocest v Republiki Sloveniji. DARS d.d. [URL: <http://www.dars.si>], 6. 11. 2003.
7. Stanovanjski zakon – SZ-1 (Uradni list RS, št. 69/2003).
8. Statistične regije – preglednica. Statistični urad Republike Slovenije. [URL: <http://www.stat.si/katalogrds/podstrani/tabele/tabela26.html>], 6. 11. 2003.
9. Statistične regije – komentar. Statistični urad Republike Slovenije. [URL: <http://www.stat.si/katalogrds/podstrani/dodatek2.html>], 6. 11. 2003.
10. Statistične regije – zemljevid. Statistični urad Republike Slovenije. [URL: <http://www.stat.si/katalogrds/podstrani/karte/karta26print.jpg>], 6. 11. 2003.
11. Statistični letopis Republike Slovenije 2002. Statistični urad Republike Slovenije. [URL: http://www.stat.si/letopis_n.htm], 6. 11. 2003.
12. Šifrant podatkov za bazo nepremičninskih transakcij DURS. Ljubljana: Ekonomska fakulteta, 2003.
13. WEKA 3: Machine Learning Software in Java. Hamilton, New Zealand: University of Waikato. [URL: <http://www.cs.waikato.ac.nz/ml/weka>], 11. 12. 2003.
14. Zakon o davku na promet nepremičnin – ZDPN-1 (Uradni list RS, št. 57/99).
15. Zakon o graditvi objektov – ZGO-1 (Uradni list RS, št. 110/2000).
16. Zakon o nepremičninskem posredovanju – ZNPosr (Uradni list RS, št. 42/2003).
17. Zakon o varstvu potrošnikov – ZVPot-UBP1 (Uradni list RS, št. 14/2003).
18. Zakon o zemljiški knjigi – ZZK-1 (Uradni list RS, št. 58/2003).

Priloga 1

Tabela P1-1: Deleži nepraznih vnosov za zemljišča za posamezne občine v letih 1999 do 2002

Občina	1999	2000	2001	2002	Skupaj	Občina	1999	2000	2001	2002	Skupaj
Ajdovščina	0,772	0,905	0,662	0,626	0,740	Ljubljana	0,765	0,888	0,652	0,589	0,726
Beltinci	0,872	0,976	0,704	0,665	0,790	Ljubno	0,806	0,978	0,663	0,654	0,779
Bled	0,811	0,938	0,684	0,619	0,750	Ljutomer	0,813	0,945	0,694	0,640	0,764
Bohinj	0,773	0,936	0,724	0,642	0,792	Logatec	0,826	0,911	0,622	0,615	0,758
Borovnica	0,788	0,951	0,607	0,630	0,754	Loška dolina	0,789	0,912	0,656	0,643	0,738
Bovec	0,838	0,923	0,612	0,625	0,744	Loški Potok	0,550	0,913	0,638	0,638	0,664
Brda	0,808	0,958	0,649	0,641	0,758	Luče	0,854	0,938	0,708	0,646	0,787
Brezovica	0,795	0,903	0,678	0,588	0,739	Lukovica	0,865	0,938	0,658	0,630	0,783
Brežice	0,885	0,946	0,676	0,655	0,779	Majšperk	0,844	0,983	0,683	0,659	0,796
Tišina	0,872	0,976	0,688	0,655	0,790	Maribor	0,800	0,852	0,604	0,583	0,781
Celje	0,836	0,930	0,685	0,608	0,756	Medvode	0,774	0,890	0,661	0,597	0,739
Cerklje na Gor.	0,878	0,936	0,623	0,614	0,755	Mengeš	0,820	0,928	0,595	0,593	0,692
Cerknica	0,790	0,883	0,667	0,612	0,724	Metlika	0,812	0,914	0,631	0,631	0,750
Cerkno	0,831	0,958	0,663	0,632	0,725	Mežica	0,750	0,825	0,670	0,578	0,657
Črenšovci	0,866	0,972	0,688	0,665	0,797	Miren – Kostan.	0,763	0,905	0,635	0,631	0,725
Črna na Kor.	0,750	0,850	0,609	0,563	0,685	Mislinja	0,822	0,934	0,625	0,612	0,728
Črnomelj	0,885	0,952	0,677	0,657	0,804	Moravče	0,894	0,931	0,650	0,610	0,758
Destrižnik	0,836	0,977	0,718	0,674	0,775	Moravske Toplice	0,855	0,959	0,695	0,661	0,797
Divača	0,852	0,932	0,645	0,616	0,742	Mozirje	0,806	0,979	0,698	0,649	0,794
Dobrepolje	0,795	0,949	0,679	0,640	0,785	Murska Sobota	0,846	0,950	0,664	0,638	0,753
Dobrova – P.G.	0,802	0,925	0,704	0,607	0,786	Muta	0,830	0,892	0,611	0,616	0,748
Dol pri Ljubljani	0,800	0,942	0,640	0,579	0,695	Naklo	0,845	0,942	0,586	0,610	0,733
Domžale	0,845	0,902	0,598	0,585	0,719	Nazarje	0,842	0,949	0,707	0,658	0,783
Dornava	0,819	1,000	0,688	0,674	0,781	Nova Gorica	0,833	0,927	0,660	0,622	0,748
Dravograd	0,856	0,903	0,750	0,597	0,790	Novo mesto	0,846	0,930	0,685	0,640	0,773
Duplek	0,818	0,890	0,642	0,609	0,734	Odranci	0,831	0,950	0,605	0,629	0,680
Gor. vas – Pol.	0,857	0,899	0,605	0,596	0,719	Ormož	0,890	0,964	0,678	0,653	0,808
Gorišnica	0,824	0,972	0,684	0,651	0,779	Osilnica	0,813	0,917	0,663	0,646	0,776
Gornja Radgona	0,816	0,941	0,722	0,650	0,785	Pesnica	0,853	0,877	0,651	0,627	0,755
Gornji Grad	0,883	0,962	0,677	0,639	0,752	Piran	0,823	0,923	0,616	0,610	0,714
Gornji Petrovci	0,851	0,961	0,740	0,656	0,810	Pivka	0,933	0,938	0,639	0,625	0,746
Grosuplje	0,830	0,931	0,661	0,605	0,749	Podčetrtek	0,853	0,985	0,723	0,688	0,836
Šalovci	0,848	1,000	0,804	0,660	0,810	Podvelka	0,750	0,938	0,615	0,606	0,684
Hrastnik	0,834	0,931	0,626	0,634	0,766	Postojna	0,849	0,924	0,623	0,625	0,731
Hrpelje - Kozina	0,799	0,912	0,653	0,601	0,748	Preddvor	0,844	0,930	0,621	0,613	0,716
Idrija	0,826	0,953	0,651	0,605	0,758	Ptuj	0,824	0,959	0,669	0,614	0,780
Ig	0,804	0,925	0,685	0,628	0,784	Puconci	0,839	0,958	0,703	0,648	0,781
Ilirska Bistrica	0,869	0,936	0,620	0,637	0,765	Rače - Fram	0,821	0,871	0,606	0,615	0,763
Ivančna Gorica	0,867	0,946	0,677	0,625	0,765	Radeče	0,857	0,980	0,640	0,613	0,787
Izola	0,831	0,916	0,618	0,610	0,721	Radenci	0,824	0,929	0,736	0,640	0,786
Jesenice	0,785	0,888	0,691	0,587	0,718	Radlje ob Dravi	0,813	0,852	0,623	0,608	0,727
Juršinci	0,784	0,993	0,714	0,651	0,781	Radovljica	0,795	0,925	0,642	0,621	0,759
Kamnik	0,907	0,953	0,638	0,616	0,776	Ravne na Kor.	0,768	0,875	0,660	0,569	0,654
Kanal	0,858	0,952	0,675	0,629	0,791	Ribnica	0,797	0,928	0,611	0,618	0,719
Kidričevo	0,829	0,963	0,698	0,655	0,762	Rogašovci	0,835	0,981	0,723	0,688	0,829
Kobarid	0,857	0,949	0,654	0,623	0,776	Rogaška Slatina	0,817	0,951	0,727	0,641	0,792
Kobilje	0,875	0,985	0,673	0,684	0,802	Rogatec	0,796	0,884	0,694	0,639	0,778
Kočevje	0,781	0,899	0,625	0,601	0,712	Ruše	0,808	0,854	0,614	0,597	0,769
Komen	0,792	0,920	0,650	0,633	0,741	Semič	0,879	0,963	0,638	0,642	0,746
Koper	0,799	0,917	0,619	0,586	0,693	Sevnica	0,831	0,941	0,669	0,666	0,777
Kozje	0,917	0,969	0,788	0,688	0,860	Sežana	0,802	0,924	0,632	0,598	0,719
Kranj	0,818	0,882	0,590	0,589	0,766	Slovenj Gradec	0,799	0,927	0,680	0,635	0,775
Kranjska Gora	0,750	0,910	0,701	0,586	0,760	Slov. Bistrica	0,921	0,974	0,631	0,628	0,769
Krško	0,644	0,923	0,670	0,659	0,733	Slov. Konjice	0,811	0,949	0,661	0,632	0,738
Kungota	0,814	0,873	0,635	0,639	0,743	Starše	0,823	0,913	0,712	0,638	0,774
Kuzma	0,646	0,977	0,792	0,648	0,797	Sveti Jurij	0,828	0,948	0,757	0,654	0,767

Tabela P1-1: (nadaljevanje)

Občina	1999	2000	2001	2002	Skupaj	Občina	1999	2000	2001	2002	Skupaj
Laško	0,864	0,927	0,696	0,648	0,767	Šenčur	0,865	0,913	0,609	0,589	0,733
Lenart	0,778	0,932	0,684	0,645	0,739	Šentilj	0,826	0,878	0,643	0,634	0,755
Lendava	0,863	0,965	0,699	0,651	0,784	Šentjernej	0,877	0,955	0,679	0,641	0,750
Litija	0,849	0,941	0,641	0,625	0,764	Šentjur pri Celju	0,876	0,970	0,682	0,639	0,810
Škocjan	0,850	0,940	0,696	0,668	0,776	Grad	0,844	0,925	0,652	0,688	0,762
Škofja Loka	0,859	0,918	0,636	0,608	0,720	Hajdina	0,833	0,978	0,673	0,630	0,779
Škofljica	0,814	0,907	0,663	0,600	0,738	Hoče - Slivnica	0,830	0,876	0,612	0,586	0,736
Šmarje pri Jel.	0,844	0,973	0,720	0,652	0,804	Hodoš	0,875	0,969	0,965	0,688	0,942
Šmartno ob Paki	0,895	0,936	0,653	0,582	0,776	Horjul	0,797	0,930	0,629	0,585	0,738
Šoštanj	0,885	0,932	0,689	0,606	0,776	Jezerško	0,000	0,000	0,688	0,563	0,592
Štore	0,830	0,926	0,646	0,640	0,784	Komenda	0,893	0,952	0,627	0,618	0,760
Tolmin	0,843	0,916	0,694	0,642	0,765	Kostel	0,788	0,985	0,677	0,671	0,793
Trbovlje	0,786	0,882	0,609	0,591	0,710	Križevci	0,796	0,965	0,798	0,661	0,765
Trebnje	0,847	0,951	0,672	0,681	0,789	Lovrenc na Poh.	0,806	0,839	0,568	0,563	0,706
Trzič	0,800	0,901	0,625	0,598	0,731	Markovci	0,823	1,000	0,712	0,667	0,748
Turnišče	0,905	0,969	0,667	0,675	0,769	Miklavž na D.P.	0,836	0,860	0,615	0,578	0,729
Velenje	0,867	0,918	0,635	0,596	0,743	Mirna Peč	0,830	0,946	0,696	0,647	0,766
Velike Lašče	0,804	0,958	0,657	0,650	0,750	Oplostnica	0,903	0,995	0,642	0,643	0,790
Videm	0,824	0,981	0,694	0,647	0,790	Podlehnik	0,774	0,997	0,682	0,661	0,776
Vipava	0,813	0,913	0,704	0,619	0,735	Polzela	0,832	0,915	0,699	0,608	0,770
Vitanje	0,775	0,950	0,656	0,653	0,818	Prebold	0,875	0,951	0,746	0,632	0,830
Vodice	0,769	0,898	0,677	0,619	0,735	Prevalje	0,740	0,828	0,663	0,569	0,702
Vojnik	0,833	0,951	0,723	0,640	0,777	Razkrižje	0,851	1,000	0,896	0,677	0,857
Vrhnika	0,792	0,899	0,593	0,596	0,703	Ribnica na Poh.	0,875	0,866	0,634	0,632	0,711
Vuzenica	0,828	0,894	0,618	0,638	0,750	Selnica ob Dravi	0,824	0,870	0,672	0,615	0,753
Zagorje ob Savi	0,872	0,947	0,678	0,616	0,764	Sodražica	0,817	0,893	0,628	0,638	0,715
Zavrč	0,861	1,000	0,659	0,646	0,784	Solčava	0,896	1,000	0,688	0,563	0,848
Zreče	0,790	0,947	0,663	0,603	0,765	Sveta Ana	0,799	0,946	0,675	0,666	0,765
Železniki	0,853	0,920	0,631	0,603	0,728	Sveti Andraž	0,906	1,000	0,778	0,669	0,786
Žiri	0,856	0,886	0,643	0,595	0,704	Šemp. - Vrtojba	0,857	0,914	0,640	0,606	0,750
Benedikt	0,813	0,927	0,656	0,650	0,772	Tabor	0,871	0,950	0,648	0,616	0,780
Bistrica ob Sotli	0,823	0,986	0,844	0,688	0,865	Trnovska vas	0,825	0,979	0,676	0,680	0,769
Bloke	0,802	0,920	0,636	0,659	0,751	Trzin	0,863	0,884	0,635	0,586	0,758
Braslovče	0,873	0,923	0,686	0,629	0,791	Velika Polana	0,906	0,944	0,646	0,652	0,788
Cankova	0,763	0,981	0,849	0,673	0,836	Veržej	0,754	0,975	0,625	0,639	0,776
Cerkvenjak	0,830	0,988	0,683	0,664	0,787	Vransko	0,853	0,952	0,681	0,620	0,748
Dobje	0,875	1,000	1,000	0,625	0,883	Žalec	0,867	0,926	0,683	0,618	0,783
Dobrna	0,854	0,992	0,701	0,672	0,824	Žetale	0,938	0,969	0,688	0,646	0,826
Dobrovnik	0,854	0,986	0,680	0,680	0,798	Žirovnica	0,803	0,899	0,710	0,594	0,774
Dol. Toplice	0,861	0,947	0,685	0,656	0,788	Žužemberk	0,866	0,966	0,731	0,671	0,788
Uteženo povprečje odstotkov vnosov (*)							0,823	0,919	0,658	0,623	0,758

Opombi: **Skupaj** ... povprečni delež nepraznih vnosov za vsa štiri leta

(*) ... uteži so določene glede na število opravljenih transakcij v občinah

Tabela P1-2: Deleži nepraznih vnosov za stavbe za posamezne občine v letih 1999 do 2002

Občina	1999	2000	2001	2002	Skupaj	Občina	1999	2000	2001	2002	Skupaj
Ajdovščina	0,836	0,946	0,942	0,935	0,927	Litija	0,867	0,978	0,913	0,922	0,923
Beltinci	0,000	0,960	0,912	0,861	0,901	Ljubljana	0,853	0,975	0,958	0,954	0,945
Bled	0,889	0,965	0,933	0,931	0,943	Ljubno	0,000	0,960	0,867	0,882	0,898
Bohinj	0,667	0,969	0,952	0,942	0,947	Ljutomer	0,830	0,957	0,888	0,883	0,893
Borovnica	0,852	0,975	0,974	0,994	0,958	Logatec	0,889	0,974	0,951	0,948	0,950
Bovec	0,905	0,949	0,913	0,908	0,922	Loška dolina	0,819	0,914	0,917	0,865	0,884
Brda	0,889	0,939	0,871	0,889	0,892	Loški Potok	0,889	0,944	0,889	0,914	0,912
Brezovica	0,872	0,948	0,883	0,862	0,903	Luče	0,000	0,958	0,865	0,833	0,881
Brežice	0,000	0,967	0,936	0,913	0,930	Lukovica	0,889	0,958	0,928	0,915	0,925
Tišina	0,944	0,944	0,847	0,884	0,894	Majšperk	0,711	0,985	0,931	0,899	0,914
Celje	0,877	0,976	0,975	0,965	0,949	Maribor	0,907	0,976	0,932	0,938	0,940
Cerklje na Gor.	0,000	0,939	0,895	0,912	0,911	Medvode	0,864	0,964	0,940	0,931	0,931
Cerknica	0,780	0,907	0,940	0,945	0,908	Mengeš	0,905	0,972	0,970	0,964	0,959
Cerkno	0,872	0,944	0,979	0,935	0,930	Metlika	0,837	0,954	0,862	0,629	0,817
Črenšovci	0,861	0,944	0,000	0,922	0,910	Mežica	0,831	0,963	0,929	0,917	0,909
Črna na Kor.	0,833	0,956	0,927	0,930	0,924	Miren – Kostan.	0,846	0,954	0,875	0,891	0,897
Črnomelj	0,871	0,961	0,959	0,971	0,942	Mislinja	0,857	0,956	0,907	0,924	0,922
Destričnik	0,681	0,944	0,900	0,880	0,846	Moravče	0,889	0,933	0,902	0,903	0,907
Divača	0,896	0,979	0,952	0,917	0,946	Mor. Toplice	0,872	0,962	0,873	0,862	0,888
Dobrepolje	0,833	0,958	0,873	0,927	0,907	Mozirje	0,000	0,964	0,900	0,859	0,905
Dobrova – P.G.	0,875	0,948	0,885	0,833	0,892	Murska Sobota	0,921	0,973	0,945	0,951	0,952
Dol pri Ljubljani	0,878	0,962	0,917	0,873	0,911	Muta	0,873	0,984	0,994	0,972	0,962
Domžale	0,883	0,971	0,960	0,958	0,952	Naklo	0,000	0,956	0,889	0,883	0,900
Dornava	0,889	1,000	0,907	0,922	0,923	Nazarje	0,000	0,960	0,876	0,926	0,913
Dravograd	0,883	0,959	0,955	0,940	0,943	Nova Gorica	0,896	0,969	0,938	0,923	0,935
Duplek	0,000	0,944	0,856	0,854	0,869	Novo mesto	0,899	0,966	0,946	0,936	0,942
Gor. vas – Polj.	0,917	0,956	0,877	0,886	0,917	Odranci	0,944	0,000	0,000	0,944	0,944
Gorišnica	0,733	0,939	0,924	0,881	0,893	Ormož	0,860	0,963	0,950	0,882	0,908
Gornja Radgona	0,000	0,964	0,937	0,938	0,948	Oslinica	0,833	0,972	0,861	0,000	0,922
Gornji Grad	0,222	0,960	0,907	0,870	0,886	Pesnica	0,000	0,958	0,846	0,877	0,867
Gornji Petrovci	0,833	0,951	0,873	0,842	0,881	Piran	0,892	0,971	0,953	0,962	0,950
Grosuplje	0,868	0,973	0,953	0,958	0,951	Pivka	0,931	0,955	0,946	0,935	0,941
Šalovci	0,926	0,954	0,884	0,833	0,891	Podčetrtek	0,789	0,940	0,871	0,856	0,857
Hrastnik	0,902	0,963	0,751	0,804	0,831	Podvelka	0,778	0,989	0,967	0,913	0,929
Hrpelje - Kozina	0,907	0,955	0,876	0,911	0,915	Postojna	0,932	0,969	0,977	0,969	0,968
Idrija	0,866	0,959	0,980	0,930	0,942	Preddvor	0,000	0,972	0,921	0,923	0,937
Ig	0,885	0,962	0,895	0,856	0,910	Ptuj	0,801	0,972	0,957	0,953	0,938
Ilirska Bistrica	0,956	0,963	0,938	0,928	0,943	Puconci	0,938	0,940	0,853	0,848	0,887
Ivančna Gorica	0,873	0,963	0,906	0,925	0,922	Rače - Fram	0,000	0,955	0,897	0,899	0,907
Izola	0,900	0,970	0,971	0,974	0,961	Radeče	0,931	0,966	0,917	0,936	0,936
Jesenice	0,832	0,963	0,949	0,941	0,933	Radenci	0,889	0,954	0,925	0,936	0,937
Juršinci	0,685	0,944	0,889	0,875	0,859	Radlje ob Dravi	0,801	0,965	0,953	0,942	0,911
Kamnik	0,895	0,974	0,963	0,970	0,960	Radovljica	0,671	0,966	0,953	0,964	0,953
Kanal	0,907	0,958	0,920	0,905	0,921	Ravne na Kor.	0,838	0,951	0,935	0,938	0,932
Kidričevo	0,821	0,972	0,932	0,923	0,924	Ribnica	0,838	0,964	0,915	0,922	0,921
Kobarid	0,889	0,951	0,925	0,906	0,921	Rogašovci	0,944	0,944	0,852	0,921	0,910
Kobilje	0,000	0,944	0,889	0,889	0,907	Rogaška Slatina	0,805	0,957	0,910	0,904	0,894
Kočevje	0,832	0,957	0,958	0,961	0,943	Rogatec	0,849	0,967	0,962	0,883	0,916
Komen	0,896	0,947	0,895	0,864	0,901	Ruše	0,000	0,978	0,936	0,925	0,935
Koper	0,834	0,926	0,952	0,923	0,911	Semič	0,856	0,956	0,938	0,944	0,919
Kozje	0,854	0,951	0,855	0,842	0,880	Sevnica	0,694	0,975	0,911	0,908	0,926
Kranj	0,747	0,982	0,976	0,974	0,973	Sežana	0,860	0,963	0,933	0,926	0,929
Kranjska Gora	0,832	0,959	0,963	0,939	0,930	Slovenj Gradec	0,857	0,979	0,941	0,941	0,940
Krško	0,429	0,936	0,935	0,940	0,754	Slov. Bistrica	0,892	0,957	0,925	0,918	0,928
Kungota	0,000	0,931	0,893	0,878	0,893	Slov. Konjice	0,877	0,975	0,936	0,943	0,940
Kuzma	0,917	0,944	0,870	0,833	0,897	Starše	0,000	0,000	0,874	0,903	0,881
Laško	0,899	0,965	0,892	0,922	0,918	Sveti Jurij	0,000	0,951	0,911	0,897	0,918
Lenart	0,769	0,954	0,953	0,941	0,933	Šenčur	0,000	0,944	0,914	0,944	0,927
Lendava	0,903	0,961	0,947	0,950	0,949	Šentilj	0,000	0,969	0,902	0,920	0,923

Tabela P1-2: (nadaljevanje)

Občina	1999	2000	2001	2002	Skupaj	Občina	1999	2000	2001	2002	Skupaj
Šentjernej	0,881	0,955	0,894	0,887	0,904	Dol. Toplice	0,896	0,944	0,878	0,865	0,896
Šentjur pri Celju	0,871	0,957	0,925	0,929	0,923	Grad	0,944	0,944	0,889	0,852	0,899
Škocjan	0,881	0,944	0,889	0,857	0,893	Hajdina	0,694	0,939	0,917	0,873	0,881
Škofja Loka	0,909	0,979	0,960	0,941	0,951	Hoče - Slivnica	0,000	0,957	0,896	0,894	0,901
Škofljica	0,892	0,960	0,886	0,871	0,894	Hodoš	0,000	0,000	0,917	0,000	0,917
Šmarje pri Jel.	0,833	0,962	0,890	0,866	0,878	Horjul	0,889	0,944	0,856	0,867	0,886
Šmartno ob Paki	0,944	0,965	0,944	0,905	0,943	Jezerško	0,000	0,000	0,926	0,000	0,926
Šoštanj	0,928	0,985	0,963	0,932	0,955	Komenda	0,889	0,948	0,925	0,930	0,929
Štore	0,868	0,958	0,977	0,988	0,974	Kostel	0,833	0,933	0,889	0,926	0,914
Tolmin	0,889	0,964	0,923	0,923	0,933	Križevci	0,889	0,944	0,900	0,839	0,891
Trbovlje	0,888	0,971	0,876	0,892	0,908	Lovrenc na Poh.	0,000	0,944	0,963	0,881	0,933
Trebnje	0,000	0,956	0,942	0,913	0,932	Markovci	0,786	0,944	0,910	0,865	0,872
Trzič	0,882	0,977	0,958	0,961	0,953	Miklavž na D.P.	0,000	0,944	0,858	0,857	0,863
Turnišče	0,889	0,944	0,000	0,926	0,926	Mirna Peč	0,880	0,944	0,839	0,838	0,859
Velenje	0,917	0,979	0,985	0,978	0,973	Oplotnica	0,889	0,944	0,889	0,958	0,913
Velike Lašče	0,907	0,944	0,878	0,864	0,903	Podlehnik	0,605	0,921	0,921	0,870	0,818
Videm	0,733	0,936	0,896	0,895	0,862	Polzela	0,881	0,965	0,984	0,952	0,959
Vipava	0,861	0,942	0,922	0,892	0,913	Prebold	0,903	0,977	0,961	0,958	0,958
Vitanje	0,917	0,969	0,956	0,852	0,942	Prevalje	0,848	0,946	0,938	0,927	0,924
Vodice	0,759	0,927	0,889	0,944	0,911	Razkrižje	0,889	0,944	0,000	0,833	0,894
Vojnik	0,883	0,967	0,936	0,901	0,923	Ribnica na Poh.	0,889	0,972	0,900	0,856	0,893
Vrhnika	0,861	0,972	0,971	0,976	0,958	Selnica ob Dravi	0,000	0,958	0,896	0,894	0,903
Vuzenica	0,000	0,937	0,939	0,994	0,957	Sodražica	0,878	0,944	0,877	0,800	0,870
Zagorje ob Savi	0,906	0,965	0,855	0,834	0,887	Solčava	0,000	0,944	0,907	0,000	0,917
Zavrč	0,569	0,944	0,889	0,883	0,822	Sveti Ana	0,000	0,000	0,000	0,926	0,926
Zreče	0,883	0,967	0,926	0,937	0,935	Sveti Andraž	0,833	0,926	0,875	0,972	0,906
Železniki	0,901	0,962	0,929	0,899	0,924	Šemp. - Vrtojba	0,896	0,957	0,939	0,920	0,933
Žiri	0,889	0,960	0,922	0,910	0,929	Tabor	0,789	0,963	1,000	0,952	0,889
Benedikt	0,889	0,981	0,944	0,889	0,931	Trnovska vas	0,806	0,917	0,911	1,000	0,909
Bistrica ob Sotli	0,815	0,944	0,870	0,833	0,863	Trzin	0,896	0,958	0,951	0,956	0,949
Bloke	0,819	0,917	0,833	0,833	0,843	Velika Polana	0,000	0,000	0,889	0,944	0,903
Braslovče	0,889	0,939	0,911	0,897	0,906	Verzej	0,759	0,972	0,933	0,900	0,881
Cankova	0,944	0,944	0,913	0,861	0,906	Vransko	0,883	0,956	0,931	0,907	0,913
Cerkvenjak	0,000	0,944	0,944	0,917	0,935	Žalec	0,897	0,974	0,965	0,940	0,949
Dobje	0,889	0,944	0,889	0,889	0,897	Žetale	0,000	0,944	0,000	0,917	0,926
Dobrna	0,900	0,963	0,913	0,833	0,913	Žirovnica	0,838	0,928	0,944	0,900	0,900
Dobrovnik	0,815	0,956	0,944	0,910	0,910	Žužemberk	0,901	0,944	0,878	0,867	0,891
Uteženo povprečje odstotkov vnosov (*)							0,848	0,967	0,944	0,939	0,935

Opombi: **Skupaj** ... povprečni delež nepraznih vnosov za vsa štiri leta



(*) ... uteži so določene glede na število opravljenih transakcij v občinah

Tabela P1-3: Skupni deleži nepraznih vnosov za zemljišča in stavbe za posamezne občine

Občina	stavbe	zem	skupaj	R_st	R_zem	R_sk	Občina	stavbe	zem	skupaj	R_st	R_zem	R_sk
Ajdovščina	0,927	0,740	0,800	122	52	58	Litija	0,923	0,764	0,829	107	93	114
Beltinci	0,901	0,790	0,807	54	157	74	Ljubljana	0,945	0,726	0,918	163	31	189
Bled	0,943	0,750	0,863	158	67	163	Ljubno	0,898	0,779	0,810	50	128	80
Bohinj	0,947	0,792	0,882	165	165	180	Ljutomer	0,893	0,764	0,821	38	94	99
Borovnica	0,958	0,754	0,861	181	76	161	Logatec	0,950	0,758	0,846	172	81	137
Bovec	0,922	0,744	0,814	106	57	86	Loška dolina	0,884	0,738	0,763	25	45	12
Brdra	0,892	0,758	0,803	35	85	62	Loški Potok	0,912	0,664	0,745	84	4	4
Brezovica	0,903	0,739	0,785	59	51	34	Luce	0,881	0,787	0,821	23	151	100
Brežice	0,930	0,779	0,827	130	130	111	Lukovica	0,925	0,783	0,816	116	139	94
Tišina	0,894	0,790	0,805	42	158	70	Majšperk	0,914	0,796	0,840	91	168	131
Celje	0,949	0,756	0,911	169	80	188	Maribor	0,940	0,781	0,864	149	135	167
Cerklje na Gor.	0,911	0,755	0,783	82	79	30	Medvode	0,931	0,739	0,822	133	49	102
Cerknica	0,908	0,724	0,798	72	28	53	Mengeš	0,959	0,692	0,836	184	8	127
Cerkno	0,930	0,725	0,781	131	29	28	Metlika	0,817	0,750	0,765	2	66	13
Crenšovci	0,910	0,797	0,804	76	171	67	Mežica	0,909	0,657	0,838	74	3	130
Crna na Kor.	0,924	0,685	0,848	112	7	144	Miren – Kostan.	0,897	0,725	0,805	47	30	71
Crnomelj	0,942	0,804	0,842	154	175	133	Mislinja	0,922	0,728	0,781	103	33	27
Destričnik	0,846	0,775	0,785	7	115	32	Moravce	0,907	0,758	0,782	66	83	29
Divaca	0,946	0,742	0,850	164	54	146	Mor. Toplice	0,888	0,797	0,817	30	170	96
Dobrepolje	0,907	0,785	0,800	69	145	55	Mozirje	0,905	0,794	0,825	62	167	109
Dobrova – P.G.	0,892	0,786	0,811	36	149	82	Murska Sobota	0,952	0,753	0,867	176	75	171
Dol pri LJ	0,911	0,695	0,751	83	10	5	Muta	0,962	0,748	0,813	188	60	84
Domžale	0,952	0,719	0,843	175	23	134	Naklo	0,900	0,733	0,762	52	39	11
Dornava	0,923	0,781	0,810	108	137	79	Nazarje	0,913	0,783	0,825	87	138	108
Dravograd	0,943	0,790	0,860	161	159	158	Nova Gorica	0,935	0,748	0,863	142	64	165
Duplek	0,869	0,734	0,759	15	41	8	Novo mesto	0,942	0,773	0,853	156	112	153
Gor. vas – Polj.	0,917	0,719	0,787	95	25	36	Odranci	0,944	0,680	0,693	162	5	2
Gorišnica	0,893	0,779	0,803	37	129	63	Ormož	0,908	0,808	0,830	71	176	117
Gor. Radgona	0,948	0,785	0,831	166	146	119	Osilnica	0,922	0,776	0,830	104	119	116
Gornji Grad	0,886	0,752	0,785	27	73	33	Pesnica	0,867	0,755	0,778	14	78	23
Gornji Petrovci	0,881	0,810	0,828	23	177	112	Piran	0,950	0,714	0,901	171	18	184
Grosuplje	0,951	0,749	0,827	174	65	110	Pivka	0,941	0,746	0,817	152	59	97
Šalovci	0,891	0,810	0,850	32	178	145	Podcetrtek	0,857	0,836	0,846	8	186	136
Hrastnik	0,831	0,766	0,804	5	103	65	Podvelka	0,929	0,684	0,830	126	6	115
Hrpelje-Kozina	0,915	0,748	0,798	92	62	52	Postojna	0,968	0,731	0,874	189	37	176
Idrija	0,942	0,758	0,860	155	84	159	Preddvor	0,937	0,716	0,769	147	20	16
Ig	0,910	0,784	0,816	75	144	92	Ptuj	0,938	0,780	0,878	148	132	177
Ilirska Bistrica	0,943	0,765	0,833	157	97	123	Puonci	0,887	0,781	0,806	29	134	72
Ivančna Gorica	0,922	0,765	0,801	105	99	60	Race - Fram	0,907	0,763	0,796	67	91	48
Izola	0,961	0,721	0,901	187	27	185	Radece	0,936	0,787	0,879	145	152	179
Jesenice	0,933	0,718	0,919	136	21	190	Radenci	0,937	0,786	0,853	146	148	154
Juršinci	0,859	0,781	0,797	10	137	50	Radlje ob Dravi	0,911	0,727	0,828	79	32	113
Kamnik	0,960	0,776	0,908	186	118	187	Radovljica	0,953	0,759	0,863	178	86	164
Kanal	0,921	0,791	0,847	101	163	138	Ravne na Kor.	0,932	0,654	0,867	135	2	172
Kidricevo	0,924	0,762	0,808	115	89	76	Ribnica	0,921	0,719	0,797	100	23	49
Kobarid	0,921	0,776	0,811	102	121	81	Rogašovci	0,910	0,829	0,851	77	183	147
Kobilje	0,907	0,802	0,805	69	173	69	Rogaška Slatina	0,894	0,792	0,847	43	164	140
Kocevje	0,943	0,712	0,848	160	17	142	Rogatec	0,916	0,778	0,837	93	127	128
Komen	0,901	0,741	0,794	55	53	46	Ruše	0,935	0,769	0,821	141	106	101
Koper	0,911	0,693	0,831	81	9	120	Semic	0,919	0,746	0,775	99	58	19
Kozje	0,880	0,860	0,869	19	189	173	Sevnica	0,926	0,777	0,823	121	125	105
Kranj	0,973	0,766	0,870	190	101	174	Sežana	0,929	0,719	0,817	128	23	95
Kranjska Gora	0,930	0,760	0,878	129	87	178	Slovenj Gradec	0,940	0,775	0,861	150	116	160
Krško	0,754	0,733	0,742	1	38	3	Slov. Bistrica	0,928	0,769	0,838	124	108	129
Kungota	0,893	0,743	0,767	39	55	15	Slov. Konjice	0,940	0,738	0,833	151	46	124
Kuzma	0,897	0,797	0,818	49	169	98	Starše	0,881	0,774	0,792	23	113	42
Laško	0,918	0,767	0,848	98	105	143	Sveti Jurij	0,918	0,767	0,794	97	104	45
Lenart	0,933	0,739	0,788	138	50	39	Šencur	0,927	0,733	0,756	123	40	7
Lendava	0,949	0,784	0,824	167	142	106	Šentilj	0,923	0,755	0,813	109	77	83

Tabela P1-3: (nadaljevanje)

Občina	stavbe	zem	skupaj	R_st	R_zem	R_sk	Občina	stavbe	zem	skupaj	R_st	R_zem	R_sk	
Šentjernej	0,904	0,750	0,787	61	70	35	Dol. Toplice	0,896	0,788	0,814	46	154	87	
Šentj. pri Celju	0,923	0,810	0,851	111	179	148	Grad	0,899	0,762	0,823	51	90	103	
Škocjan	0,893	0,776	0,800	40	122	56	Hajdina	0,881	0,779	0,800	21	131	57	
Škofja Loka	0,951	0,720	0,865	173	26	169	Hoce - Slivnica	0,901	0,736	0,779	56	44	25	
Škofljica	0,894	0,738	0,788	45	48	38	Hodoš	0,917	0,942	0,939	95	192	192	
Šmarje pri Jel.	0,878	0,804	0,834	18	174	126	Horjul	0,886	0,738	0,776	26	47	21	
Šmar. ob Paki	0,943	0,776	0,814	159	124	88	Jezerško	0,926	0,592	0,642	117	1	1	
Šoštanj	0,955	0,776	0,847	179	117	139	Komenda	0,929	0,760	0,807	125	88	75	
Štore	0,974	0,784	0,920	192	141	191	Kostel	0,914	0,793	0,816	90	166	91	
Tolmin	0,933	0,765	0,841	139	98	132	Križevci	0,891	0,765	0,789	34	95	40	
Trbovlje	0,908	0,710	0,866	70	15	170	Lovrenc na Poh.	0,933	0,706	0,752	140	14	6	
Trebnje	0,932	0,789	0,814	134	156	85	Markovci	0,872	0,748	0,770	17	61	17	
Trzic	0,953	0,731	0,862	177	36	162	Miklavž na D.P.	0,863	0,729	0,766	12	35	14	
Turnišče	0,926	0,769	0,778	119	109	24	Mirna Pec	0,859	0,766	0,784	9	102	31	
Velenje	0,973	0,743	0,908	191	56	186	Oplotnica	0,913	0,790	0,832	88	161	121	
Velike Lašce	0,903	0,750	0,792	58	70	43	Podlehnik	0,818	0,776	0,788	3	120	37	
Videm	0,862	0,790	0,809	11	160	78	Polzela	0,959	0,770	0,844	185	110	135	
Vipava	0,913	0,735	0,790	85	42	41	Prebold	0,958	0,830	0,900	182	184	183	
Vitanje	0,942	0,818	0,852	154	180	151	Prevalje	0,924	0,702	0,830	113	11	118	
Vodice	0,911	0,735	0,779	80	43	26	Razkrižje	0,894	0,857	0,863	44	188	166	
Vojnik	0,923	0,777	0,815	110	126	90	Ribnica na Poh.	0,893	0,711	0,775	41	16	20	
Vrhnika	0,958	0,703	0,833	183	12	122	Selnica ob Dravi	0,903	0,753	0,793	60	74	44	
Vuzenica	0,957	0,750	0,852	180	70	150	Sodražica	0,870	0,715	0,760	16	19	9	
Zagorje ob Savi	0,887	0,764	0,834	28	92	125	Solcava	0,917	0,848	0,873	95	187	175	
Zavrc	0,822	0,784	0,798	4	143	51	Sveti Ana	0,926	0,765	0,772	119	100	18	
Zrece	0,935	0,765	0,859	143	96	156	Sveti Andraž	0,906	0,786	0,809	63	147	77	
Železniki	0,924	0,728	0,805	114	34	68	Šemp. - Vrtojba	0,933	0,750	0,852	137	70	152	
Žiri	0,929	0,704	0,776	127	13	22	Tabor	0,889	0,780	0,825	31	133	107	
Benedikt	0,931	0,772	0,798	132	111	54	Trnovska vas	0,909	0,769	0,796	73	107	47	
Bistrica ob Sotli	0,863	0,865	0,865	13	190	168	Trzin	0,949	0,758	0,893	170	82	182	
Bloke	0,843	0,751	0,760	6	72	10	Velika Polana	0,903	0,788	0,803	58	153	64	
Braslovce	0,906	0,791	0,816	65	162	93	Verzej	0,881	0,776	0,807	20	123	73	
Cankova	0,906	0,836	0,855	64	185	155	Vransko	0,913	0,748	0,800	89	63	59	
Cerkvenjak	0,935	0,787	0,802	144	150	61	Žalec	0,949	0,783	0,860	168	140	157	
Dobje	0,897	0,883	0,888	48	191	181	Žetale	0,926	0,826	0,851	119	182	149	
Dobrna	0,913	0,824	0,847	87	181	141	Žirovnica	0,900	0,774	0,823	53	114	104	
Dobrovnik	0,910	0,798	0,804	78	172	66	Žužemberk	0,891	0,788	0,814	33	155	89	
Uteženo povprečje odstotkov vnosov (*)									0,935	0,758	0,854			

- Opombe:
- stavbe** ... delež nepraznih vnosov za stavbe za vsa štiri leta
 - zem** ... delež nepraznih vnosov za zemljišča za vsa štiri leta
 - skupaj** ... delež nepraznih vnosov za vse nepremičnine za vsa štiri leta
 - R_st** ... rang deleža nepraznih vnosov za stavbe za vsa štiri leta
 - R_zem** ... rang deleža nepraznih vnosov za zemljišča za vsa štiri leta
 - R_sk** ... rang deleža nepraznih vnosov za vse nepremičnine za vsa štiri leta
 - (*) ... uteži so določene glede na število opravljenih transakcij v občinah
 -  ... najslabših pet občin glede na skupno kakovost vnosov
 -  ... najboljših pet občin glede na skupno kakovost vnosov

Priloga 2

Tabela P2-1: Povprečna cena kvadratnega metra stanovanj v posameznih občinah v letih od 1999 do 2002; vse vrednosti so v SIT/m²

Šif.	Občina	1999	2000	2001	2002	Skupaj	Šif.	Občina	1999	2000	2001	2002	Skupaj
1	Ajdovščina	125147	124970	136750	155645	137429	55	Kungota	0	0	20353	90897	48570
2	Beltinci	0	64021	88645	116069	86842	56	Kuzma	0	0	0	0	0
3	Bled	202931	176098	210567	218752	200271	57	Laško	58459	97692	98311	73049	86112
4	Bohinj	0	196414	216785	201083	207333	58	Lenart	64866	96976	121114	126955	112188
5	Borovnica	149782	148525	164186	127953	141070	59	Lendava	102154	73011	62290	70400	71924
6	Bovec	201089	108632	81116	120895	111596	60	Litija	130328	127068	141473	145530	136581
7	Brda	90664	61467	92363	30563	70049	61	Ljubljana	224397	244310	268184	301919	265217
8	Brezovica	121658	174387	201018	194249	186500	62	Ljubno	0	59237	67405	107623	73376
9	Brežice	0	80356	117035	98707	103738	63	Ljutomer	78197	72576	81766	82691	79093
10	Tišina	0	0	0	0	0	64	Logatec	164907	182802	193353	202862	189737
11	Celje	86941	103214	113326	115128	107975	65	Loška dolina	85217	112399	116261	0	103071
12	Cerklje na Gor.	0	0	132013	188684	172493	66	Loški Potok	0	0	0	89164	89164
13	Cerkno	67689	104233	98119	115432	104117	67	Luče	0	87835	27997	0	57916
14	Cerkno	81939	87889	114498	113478	103501	68	Lukovica	156267	164088	258227	226281	203109
15	Črenšovci	0	0	0	36471	36471	69	Majšperk	64779	73827	94415	81691	79531
16	Črna na Kor.	64203	54490	26835	46606	37817	70	Maribor	152447	121065	122831	135958	128245
17	Črnomelj	91321	88146	80646	99396	90594	71	Medvode	167591	194107	235013	217248	211176
18	Destričnik	0	0	0	0	0	72	Mengeš	204942	224965	222778	229575	223411
19	Divača	114854	82803	90191	106063	92458	73	Metlika	92995	89469	98216	96174	94146
20	Dobrepolje	0	55814	0	116985	96595	74	Mežica	84400	63044	86288	67392	75321
21	Dobrova – P.G.	133698	185195	72399	0	124810	75	Miren – Kost.	136432	143939	37188	142645	125351
22	Dol pri LJ	224749	119433	130565	0	137659	76	Mislinja	41367	38355	12802	63903	43868
23	Domžale	187655	198356	241114	289663	241907	77	Moravče	0	0	143883	89654	116768
24	Dornava	76301	49953	105180	92812	77616	78	Mor. Toplice	145415	271249	193922	183288	215542
25	Dravograd	69239	94492	82239	77039	81599	79	Mozirje	0	87501	65385	86700	75572
26	Duplek	0	0	83333	0	83333	80	Mur. Sobota	97505	111910	121676	121047	115360
27	Gor. vas – Pol.	53840	116952	153732	108368	113849	81	Muta	89503	66269	72242	65715	71284
28	Gorišnica	0	118006	126588	152138	129012	82	Naklo	0	63544	0	0	63544
29	Gor. Radgona	0	84808	96805	93597	91014	83	Nazarje	0	100523	104939	68654	78657
30	Gornji Grad	90450	69572	83031	66932	77098	84	Nova Gorica	155909	155753	168771	180843	166641
31	Gornji Petrovci	0	60938	83795	0	72367	85	Novo mesto	136404	153936	163223	168911	158885
32	Grosuplje	151260	183728	203092	205204	193018	86	Odranci	0	0	0	69149	69149
33	Šalovci	0	100000	59432	0	72955	87	Ormož	62868	83678	69075	76420	75793
34	Hrastnik	96743	95706	78146	82183	85554	88	Osišnica	0	101208	0	0	101208
35	Hrpelje-Kozina	101255	104473	111596	87338	98044	89	Pesnica	0	12727	4167	35390	21919
36	Idrija	83605	74940	84009	102175	86974	90	Piran	217453	206864	264463	271368	243761
37	Ig	141555	177111	185761	255132	182639	91	Pivka	69734	77699	83565	75394	77467
38	Ilirska Bistrica	76263	78381	77463	96848	83954	92	Podčetrtek	106020	7892	28070	123086	66008
39	Ivančna Gorica	66080	153777	133361	157084	141523	93	Podvelka	0	61160	63840	48671	58098
40	Izola	195878	209477	238917	253755	231676	94	Postojna	93268	92894	93059	100050	95479
41	Jesenice	96512	102635	91598	77137	85828	95	Preddvor	0	99249	120400	218118	133188
42	Juršinci	71544	88549	0	0	80046	96	Ptuj	93789	102130	104621	119895	106808
43	Kamnik	176902	161322	206740	221912	194752	97	Puconci	51123	88094	59316	70324	65051
44	Kanal	77469	80582	95813	85995	85838	98	Rače - Fram	0	108852	83232	134691	101749
45	Kidričevo	68350	80454	75104	78630	77404	99	Radeče	84056	66106	65229	80044	71195
46	Kobarid	0	160714	125528	1095	115611	100	Radenci	0	106203	97228	68302	74984
47	Kobilje	0	0	0	0	0	101	Rad. ob Dravi	60590	59387	49646	59829	55642
48	Kočevje	72975	74079	77272	88685	79599	102	Radovljica	101157	161142	168439	186175	169995
49	Komen	15150	101119	23328	87529	67031	103	Ravne na Kor.	71801	86975	64895	56140	59637
50	Koper	179336	191447	205071	212534	198856	104	Ribnica	82525	81218	98287	115423	95377
51	Kozje	85223	58994	88347	43600	67032	105	Rogašovci	77472	0	0	69088	71184
52	Kranj	153822	177807	182563	210665	192174	106	Rog. Slatina	120747	108174	107290	115995	112445
53	Kranjska Gora	213711	242100	277623	245471	244546	107	Rogatec	63106	83895	80353	73358	77442
54	Krško	63757	73981	104540	85258	77585	108	Ruše	0	97212	135089	102815	117034

Tabela P2-1: (nadaljevanje)

Šif.	Občina	1999	2000	2001	2002	Skupaj	Šif.	Občina	1999	2000	2001	2002	Skupaj
109	Semič	47005	87272	84940	67559	70242	152	Cankova	0	76937	63587	76937	72487
110	Sevnica	82336	95947	72428	78039	83905	153	Cerkvenjak	0	0	0	151327	151327
111	Sežana	120032	125765	133889	171271	142600	154	Dobje	0	0	0	0	0
112	Slovenj Gradec	86729	103969	97928	108987	101576	155	Dobrna	57764	80884	98175	0	72613
113	Slov. Bistrica	79114	90035	95534	100216	92686	156	Dobrovnik	0	12717	39724	86364	46268
114	Slov. Konjice	87273	100632	104746	112332	102848	157	Dol. Toplice	105539	0	33166	0	69353
115	Starše	0	0	0	52740	52740	158	Grad	0	70083	66800	0	68441
116	Sveti Jurij	0	99194	0	15570	57382	159	Hajdina	91996	0	79479	0	85737
117	Šenčur	0	0	142976	195356	177896	160	Hoče - Slivnica	0	25715	53299	98882	73026
118	Šentilj	0	82278	92366	81138	85780	161	Hodoš	0	0	0	0	0
119	Šentjernej	0	106443	110944	153565	115797	162	Horjul	0	0	0	68283	68283
120	Šentjur pri CE	95121	101846	99163	94328	97487	163	Jezersko	0	0	0	0	0
121	Škocjan	0	0	47372	0	47372	164	Komenda	0	156000	196080	173027	186504
122	Škofja Loka	147808	161638	191296	205077	180856	165	Kostel	0	27805	0	0	27805
123	Škofljica	190432	193416	224716	188938	199040	166	Križevci	0	59442	0	0	59442
124	Šmarje pri Jel.	94982	112517	83433	100163	98530	167	Lovr. na Poh.	0	0	67833	135652	82904
125	Šmar. ob Paki	92849	51813	115415	96726	87554	168	Markovci	0	0	23333	0	23333
126	Šoštanj	94159	94771	107620	119195	103034	169	Mikl. na D.P.	0	0	72809	70041	71425
127	Štore	82625	109842	60678	48775	54529	170	Mirna Peč	0	0	0	0	0
128	Tolmin	96932	95321	83388	112061	97740	171	Oplotnica	145419	0	57788	159926	112008
129	Trbovlje	77847	83794	94233	100457	91494	172	Podlehnik	28614	55380	103566	0	62520
130	Trebnje	0	97082	129253	148245	134050	173	Polzela	111084	111049	68688	87349	81650
131	Trzič	121546	124002	130513	126812	126533	174	Prebold	71340	89598	98879	116700	102668
132	Turnišče	73712	0	0	40000	56856	175	Prevalje	37166	50029	66685	61620	58302
133	Velenje	96614	107175	122077	145619	121252	176	Razkrižje	0	0	0	0	0
134	Velike Lašče	0	0	164489	99600	138533	177	Rib. na Poh.	0	86879	20219	0	53549
135	Videm	0	0	126904	85063	99010	178	Sel. ob Dravi	0	117578	110272	43422	86116
136	Vipava	142030	158800	95796	98752	114671	179	Sodražica	0	0	110183	0	110183
137	Vitanje	74328	117739	87364	0	100922	180	Solčava	0	0	67438	0	67438
138	Vodice	0	0	158397	144239	145526	181	Sveta Ana	0	0	0	24268	24268
139	Vojnik	47258	100994	0	36220	65938	182	Sveti Andraž	0	0	0	45958	45958
140	Vrhnika	157066	174585	181300	184048	177159	183	Šemp. - Vrt.	104940	137087	176199	183319	161226
141	Vuzenica	0	0	67953	65516	66561	184	Tabor	0	52948	72759	96942	80918
142	Zag. ob Savi	82309	92624	91105	100653	94696	185	Trnovska vas	0	0	93378	86970	89106
143	Zavrč	0	0	0	0	0	186	Trzin	229467	252155	259429	305749	269681
144	Zreče	105337	109296	121602	138745	123753	187	Velika Polana	0	0	0	0	0
146	Železniki	132908	125666	151577	167913	141181	188	Veržej	0	181378	166099	164742	169042
147	Žiri	93181	69646	112932	116270	97258	189	Vransko	13765	36855	44657	90465	40694
148	Benedikt	92560	80574	92742	0	87838	190	Žalec	104990	105827	108226	117588	109133
149	Bistr. ob Sotli	0	0	0	0	0	191	Žetale	0	0	0	0	0
150	Bloke	0	0	0	0	0	192	Žirovnica	120700	108684	54710	0	102690
151	Braslovče	59184	0	57594	100281	79335	193	Žužemberk	97222	0	138281	69942	93847
Uteženo povprečje spremenljivke CENA_NA_m² (*)									163315	173403	178298	184312	177411

Opombe: Šif. ... šifra občine (Občine RS 2003a)

Skupaj ... povprečje spremenljivke CENA_NA_m² za vsa štiri leta skupaj

Vnos 0 ... v določeni občini v določenem letu ni bilo transakcij s stanovanji

(*) ... uteži so določene glede na število opravljenih transakcij v občinah

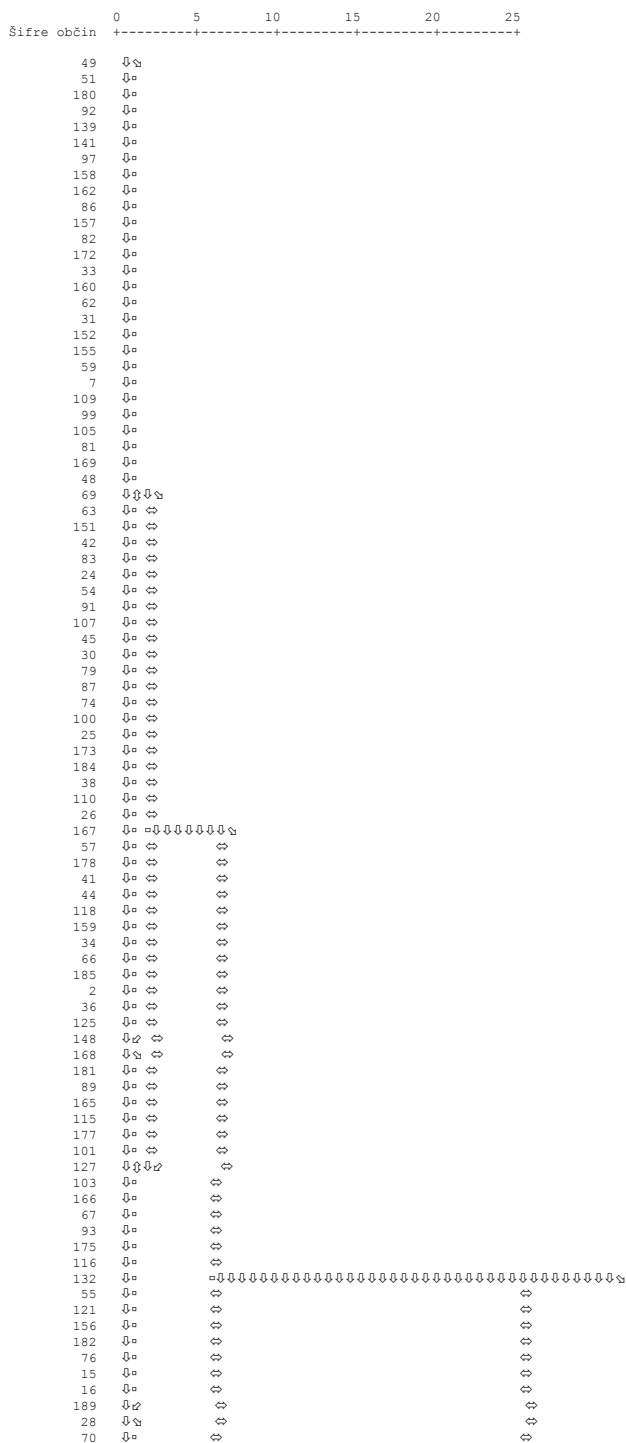
--

 ... vrednosti, ki smo jih upoštevali pri združevanju v skupine

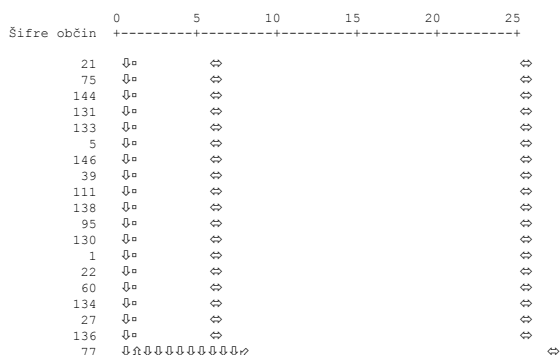
--

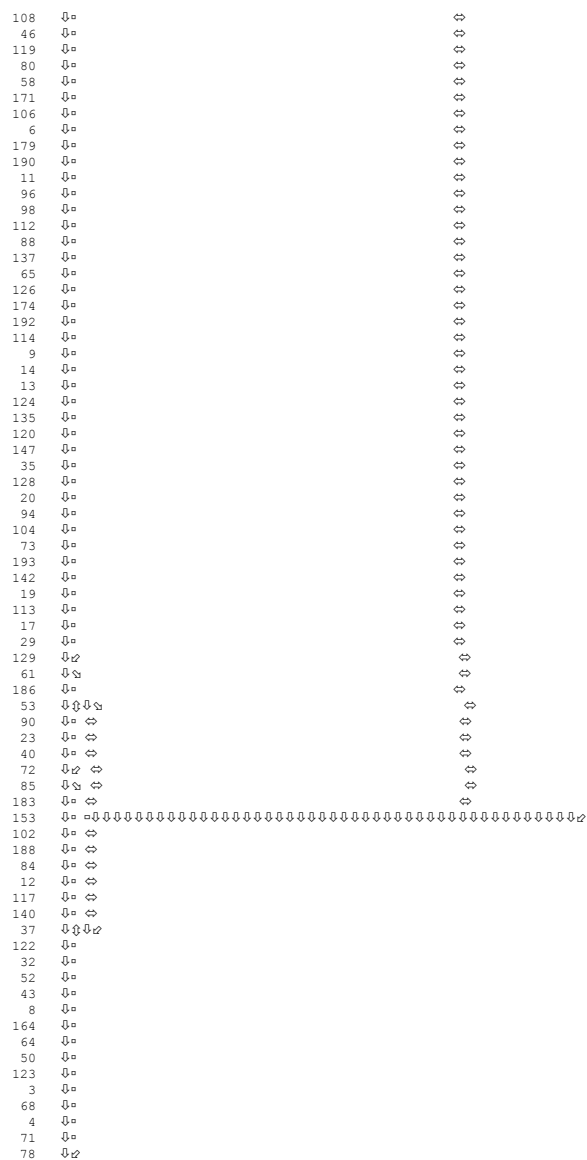
 ... občine, ki jih nismo upoštevali pri združevanju v skupine

Dendrogram, dobljen z Wardovo metodo - normalizirane razdalje med objekti



Sl. P2-1: Dendrogram razvrščanja občin v skupine glede na povprečno vrednost spremenljivke CENA_NA_m² za vsa leta skupa





Sl. P2-1: (nadaljevanje)

Tabela P2-2: Osnovne statistične značilke spremenljivke CENA_NA_m² za 5 cenovnih skupin, dobljenih na osnovi dendrograma na Sl. P2-1

Št. Skupine	Št. občin v skupini	Delež občin [%]	Sr. vrednost [SIT/m ²]	Mediana [SIT/m ²]	St. deviacija [SIT/m ²]
1	23	12,9	46454	48570	12456
2	62	34,8	76609	77252	7397
3	62	34,8	112053	108554	16052
4	24	13,5	185257	186502	17185
5	7	4,0	245743	243761	16678
Skupaj	178	100,0	106359	92572	50743

Priloga 3

Tabela P3-1: Kvalitativne spremenljivke za iskanje povezovalnih pravil, ki so ekvivalentne originalnim numeričnim in tekstovnim spremenljivkam v bazi nepremičninskih transakcij

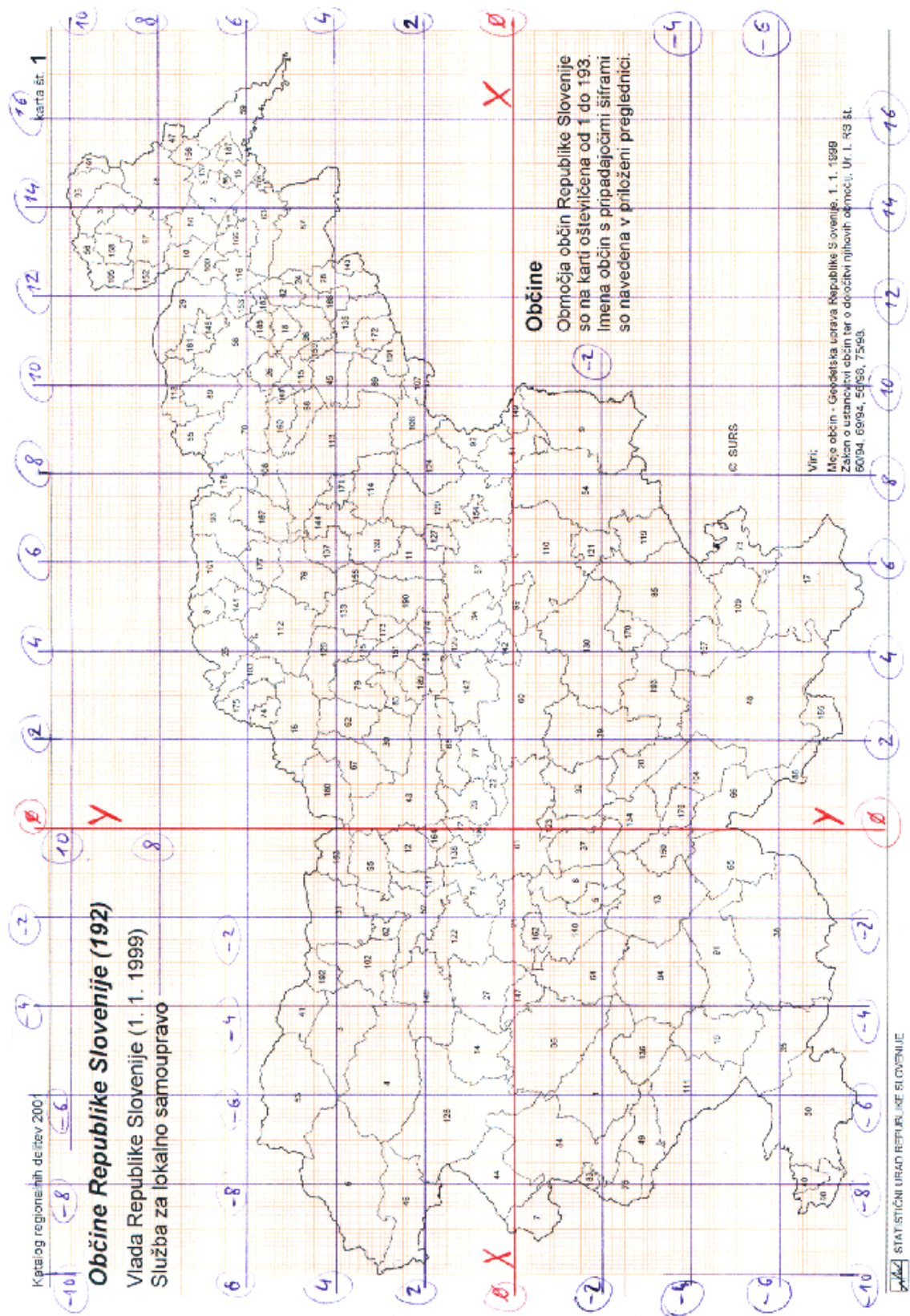
Zap. št.	Spremenljivka	Zaloga vrednosti za zemljišča (neprem. od 11 do 14)	Zaloga vrednosti za objekte (neprem. od 21 do 27)
1	LETO (*)	vrednosti od 1999 do 2002	vrednosti od 1999 do 2002
2	MESEC (*)	vrednosti od 1 do 12	vrednosti od 1 do 12
3	DAN (*)	vrednosti od 1 do 31	vrednosti od 1 do 31
4	ZAP_POG (T)	0 – vnos "0" 1 – ostalo	0 – vnos "0" 1 – ostalo
5	ZAP_NEP (T)	0 – vnos "0" 1 – ostalo	0 – vnos "0" 1 – ostalo
6	VRSTA_NEPREM	vrednosti od 11 do 14	vrednosti od 21 do 27
7	SIF_KO (T)	0 – vnos "0" 1 – ostalo	0 – vnos "0" 1 – ostalo
8	ST_VLOZ (T)	0 – brez vnosa 1 – ostalo	0 – brez vnosa 1 – ostalo
9	PARCELA (T)	0 – brez vnosa 1 – ostalo	0 – brez vnosa 1 – ostalo
10	KK_ZEM	A, B, C, D, E, F, G, H, I, J, K, L, M	A, B, C, D, E, F, G, H, I, J, K, L, M
11	RAZRED	vrednosti od 1 do 9	vrednosti od 1 do 9
12	IZMERA (N)	1 – manj kot 2000 m ² 2 – od 2000 m ² do 20000 m ² 3 – od 20000 m ² do 50000 m ² 4 – več kot 50000 m ²	1 – manj kot 70 m ² 2 – od 70 m ² do 300 m ² 3 – od 300 m ² do 800 m ² 4 – več kot 800 m ²
13	LETO_IZG (N)	1 – vnos "0" 2 – od 1 do 2002 3 – več kot 2002	1 – vnos "0" 2 – od 1 do 1900 3 – od 1900 do 1960 4 – od 1960 do 2002 5 – več kot 2002
14	LETO_PREN (N)	1 – vnos "0" 2 – od 1 do 2002 3 – več kot 2002	1 – vnos "0" 2 – od 1 do 1900 3 – od 1900 do 1960 4 – od 1960 do 2002 5 – več kot 2002
15	KLETNO	0, 1	0, 1
16	PODSTRESNO	0, 1	0, 1
17	NADSTROPJE (N)	1 – pritličje 2 – od 1 do 4 3 – več kot 4	1 – pritličje 2 – od 1 do 4 3 – več kot 4
18	LEGA	S, V, J, Z	S, V, J, Z

Tabela P3-1: (nadaljevanje)

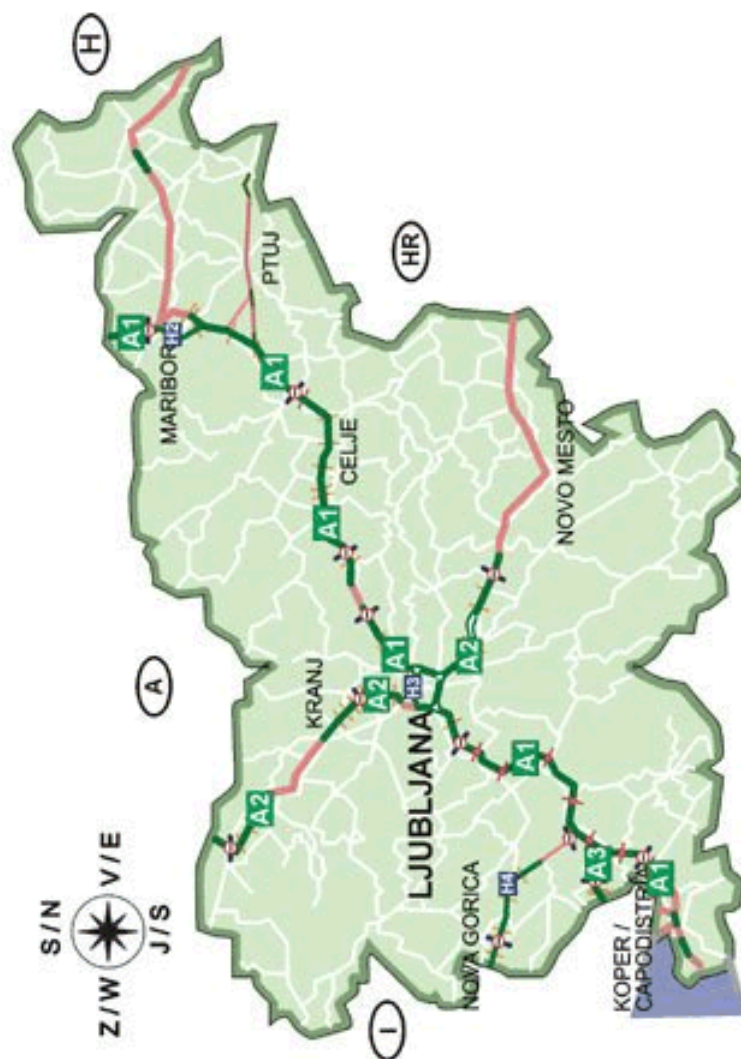
Zap. št.	Spremenljivka	Zaloga vrednosti za zemljišča (neprem. od 11 do 14)	Zaloga vrednosti za objekte (neprem. od 21 do 27)
19	<i>DVIGALO</i>	0, 1	0, 1
20	<i>ODDALJENOST</i> (N)	1 – manj kot 0,5 km 2 – od 0,5 km do 5 km 3 – več kot 5 km	1 – manj kot 0,5 km 2 – od 0,5 km do 5 km 3 – več kot 5 km
21	<i>ELEKTRIKA</i>	0, 1	0, 1
22	<i>VODOVOD</i>	0, 1	0, 1
23	<i>OGREVANJE</i>	0, 1	0, 1
24	<i>TELEFON</i>	0, 1	0, 1
25	<i>KANALIZACIJA</i>	0, 1	0, 1
26	<i>LETO_NAK</i> (N)	1 – vnos "0" 2 – od 1 do 2002 3 – več kot 2002	1 – vnos "0" 2 – od 1 do 2002 3 – več kot 2002
27	<i>NACIN_PRID</i>	1, 2, 3, 4, 5, 6, 7, 9	1, 2, 3, 4, 5, 6, 7, 9
28	<i>OBM_TRANS</i> (**)	od 1 do 10	od 1 do 10
29	<i>CENA_NA_m²</i> (***) (N)	1 – manj kot 10000 SIT/m ² 2 – od 10000 SIT/m ² do 50000 SIT/m ² 3 – od 50000 SIT/m ² do 100000 SIT/m ² 4 – več kot 100000 SIT/m ²	1 – manj kot 100000 SIT/m ² 2 – od 100000 SIT/m ² do 250000 SIT/m ² 3 – od 250000 SIT/m ² do 500000 SIT/m ² 4 – več kot 500000 SIT/m ²

Opombe: (*) ... te spremenljivke so nadomestile spremenljivko *DATUM_PRIJAVE*
 (T) ... v originalni bazi podatkov so te spremenljivke tekstovne
 (N) ... v originalni bazi podatkov so te spremenljivke numerične
 (**) ... razvrstitev občin v območja transakcij je povzeta po poglavju 4.3.1
 (***) ... cena na kvadratni meter je definirana v poglavju 4.3.1
 Razen pri spremenljivkah *ST_VLOZ* in *PARCELA* manjkajočih vnosov nismo nadomestili z nadomestnimi vrednostmi.

Priloga 4



Sl. P4-1: Koordinate občin v republiki Sloveniji (Vir: Občine RS 2003c)



Sl. P4-2: Avtoceste v Republiki Sloveniji (Vir: DARS 2003)

Priloga 5

Tabela P5-1: Primerjava regresijskih modelov za napoved nominalne oziroma realne cene kvadratnega metra stanovanj

Odpisna sprem.: <i>CENA NA m²</i>				Odpisna sprem.: <i>REALNA CENA NA m²</i>			
$R^2 = 0,593$; popravljeni $R^2 = 0,592$				$R^2 = 0,588$; popravljeni $R^2 = 0,588$			
Neodvisna sprem.	Regres. koef.	Stand. koef.	Znač. koef.	Neodvisna sprem.	Regres. koef.	Stand. koef.	Znač. koef.
<i>konstanta</i>	-30826,93	-	0,032	<i>konstanta</i>	-5144,70	-	0,672
<i>CAS TRAN</i>	4068,32	0,143	0,000	<i>CAS TRAN</i>	366,66	0,015	0,001
<i>MESEC</i>	360,20	0,013	0,007	<i>MESEC</i>	1296,17	0,054	0,000
<i>ST VLOZ</i>	-10788,89	-0,053	0,000	<i>ST VLOZ</i>	-9597,44	-0,056	0,000
<i>IZMERA</i>	-376,90	-0,116	0,000	<i>IZMERA</i>	-320,48	-0,117	0,000
<i>STAROST</i>	-608,05	-0,192	0,000	<i>STAROST</i>	-515,23	-0,193	0,000
<i>KLETNO</i>	-26365,54	-0,039	0,000	<i>KLETNO</i>	-22731,19	-0,040	0,000
<i>PODSTRES</i>	-16288,88	-0,033	0,000	<i>PODSTRES</i>	-13416,64	-0,032	0,000
<i>NADSTROP</i>	-40,95	-0,001	0,856	<i>NADSTROP</i>	-46,31	-0,001	0,808
<i>LEGA V</i>	3869,99	0,018	0,002	<i>LEGA V</i>	3111,06	0,017	0,003
<i>LEGA J</i>	6788,08	0,029	0,000	<i>LEGA J</i>	5705,26	0,029	0,000
<i>LEGA Z</i>	2478,74	0,010	0,068	<i>LEGA Z</i>	1892,41	0,009	0,101
<i>DVIGALO</i>	11444,10	0,049	0,000	<i>DVIGALO</i>	9692,67	0,049	0,000
<i>ODDALJEN</i>	-1721,65	-0,041	0,000	<i>ODDALJEN</i>	-1419,54	-0,040	0,000
<i>ELEKTR</i>	-22907,35	-0,011	0,040	<i>ELEKTR</i>	-20436,35	-0,012	0,031
<i>VODOVOD</i>	19309,70	0,013	0,020	<i>VODOVOD</i>	18230,42	0,014	0,009
<i>OGREVAN</i>	6555,50	0,014	0,005	<i>OGREVAN</i>	5954,06	0,016	0,003
<i>TELEFON</i>	8981,76	0,032	0,000	<i>TELEFON</i>	7413,70	0,031	0,000
<i>KANALIZ</i>	15368,46	0,022	0,000	<i>KANALIZ</i>	12867,49	0,022	0,000
<i>OBMTR 2</i>	6041,09	0,018	0,148	<i>OBMTR 2</i>	5509,35	0,020	0,120
<i>OBMTR 3</i>	9686,36	0,037	0,072	<i>OBMTR 3</i>	8846,71	0,040	0,052
<i>OBMTR 4</i>	69843,08	0,201	0,000	<i>OBMTR 4</i>	60199,61	0,206	0,000
<i>OBMTR 5</i>	139149,66	0,660	0,000	<i>OBMTR 5</i>	119199,74	0,670	0,000
<i>OBMTR 6</i>	35360,73	0,066	0,000	<i>OBMTR 6</i>	30720,29	0,068	0,000
<i>OBMTR 7</i>	145601,09	0,332	0,000	<i>OBMTR 7</i>	125609,31	0,340	0,000
<i>OBMTR 8</i>	55651,20	0,165	0,000	<i>OBMTR 8</i>	48261,24	0,170	0,000
<i>OBMTR 9</i>	-82,48	0,000	0,994	<i>OBMTR 9</i>	1966,23	0,004	0,820
<i>OBMTR 10</i>	28647,24	0,049	0,005	<i>OBMTR 10</i>	25343,64	0,052	0,003
<i>X KOORD</i>	340,75	0,017	0,388	<i>X KOORD</i>	383,69	0,022	0,251
<i>Y KOORD</i>	2908,29	0,091	0,000	<i>Y KOORD</i>	2507,81	0,093	0,000
<i>RAZV OBC</i>	31878,85	0,164	0,000	<i>RAZV OBC</i>	27664,28	0,169	0,000
<i>AC</i>	6849,36	0,030	0,000	<i>AC</i>	5658,72	0,029	0,000
<i>AC LJ</i>	2231,70	0,011	0,401	<i>AC LJ</i>	1827,07	0,011	0,417

Opomba: *LEGA_(i)* ... Binarne spremenljivke, ki nadomestijo sprem. *LEGA*
OBMTR_(i) ... Binarne spremenljivke, ki nadomestijo sprem. *OBM_TRAN*

Tabela P5-2: Primerjava regresijskih modelov za napoved cene kvadratnega metra stanovanj za različne oblike regresijskih modelov

Odvisna spremenljivka: $\ln(CENA_NA_m^2)$ multiplikativni model				Odvisna spremenljivka: $CENA_NA_m^2$ kombinirani model			
$R^2 = 0,451$; popravljeni $R^2 = 0,451$				$R^2 = 0,564$; popravljeni $R^2 = 0,564$			
Neodvisna sprem.	Regres. koef.	Stand. koef.	Znač. koef.	Neodvisna sprem.	Regres. koef.	Stand. koef.	Znač. koef.
<i>konstanta</i>	11,055	-	0,000	<i>konstanta</i>	149665,94	-	0,000
$\ln(CAS_TRAN)$	0,153	0,087	0,000	$CAS_TRAN \times$			
$\ln(MESEC)$	0,015	0,015	0,005	$\times MESEC$	4,97	0,002	0,709
$\ln(IZMERA)$	-0,190	-0,121	0,000	$\times ST_VLOZ$	-1034,01	-0,058	0,000
$\ln(STAROST)$	-0,239	-0,256	0,000	$\times IZMERA$	-33,95	-0,132	0,000
$\ln(NADSTROP)$	0,100	0,062	0,000	$\times STAROST$	-52,01	-0,186	0,000
$\ln(ODDALJEN)$	-0,064	-0,041	0,000	$\times KLETNO$	-2563,53	-0,038	0,000
$\ln(Y_KOORD)$	-0,064	-0,050	0,035	$\times PODSTRES$	-1230,04	-0,027	0,000
$\ln(X_KOORD)$	0,260	0,132	0,000	$\times NADSTROP$	-32,82	-0,009	0,127
$\ln(RAZV_OBC)$	0,496	0,110	0,000	$\times LEGA_V$	302,83	0,015	0,012
<i>ST VLOZ</i>	-0,097	-0,066	0,000	$\times LEGA_J$	753,12	0,034	0,000
<i>KLETNO</i>	-0,160	-0,033	0,000	$\times LEGA_Z$	210,87	0,009	0,120
<i>PODSTRES</i>	-0,243	-0,069	0,000	$\times DVIGALO$	1272,58	0,059	0,000
<i>LEGA V</i>	0,036	0,022	0,001	$\times ODDALJEN$	-140,84	-0,037	0,000
<i>LEGA J</i>	0,036	0,021	0,001	$\times ELEKTR$	-2982,02	-0,106	0,013
<i>LEGA Z</i>	-0,003	-0,002	0,776	$\times VODOVOD$	3830,99	0,137	0,000
<i>DVIGALO</i>	0,036	0,021	0,001	$\times OGREVA$	788,19	0,032	0,001
<i>ELEKTR</i>	-0,145	-0,010	0,121	$\times TELEFON$	721,72	0,035	0,000
<i>VODOVOD</i>	0,275	0,025	0,000	$\times KANALIZ$	1247,02	0,047	0,000
<i>OGREVAN</i>	0,084	0,026	0,000	$\times OBMTR_2$	1678,16	0,056	0,000
<i>TELEFON</i>	0,110	0,054	0,000	$\times OBMTR_3$	2259,30	0,091	0,000
<i>KANALIZ</i>	0,259	0,052	0,000	$\times OBMTR_4$	7958,42	0,247	0,000
<i>OBMTR 2</i>	-0,044	-0,018	0,154	$\times OBMTR_5$	14363,18	0,735	0,000
<i>OBMTR 3</i>	-0,015	-0,008	0,654	$\times OBMTR_6$	5051,89	0,100	0,000
<i>OBMTR 4</i>	0,357	0,143	0,000	$\times OBMTR_7$	15876,61	0,389	0,000
<i>OBMTR 5</i>	0,685	0,451	0,000	$\times OBMTR_8$	6743,52	0,215	0,000
<i>OBMTR 6</i>	0,256	0,066	0,000	$\times OBMTR_9$	2344,39	0,045	0,019
<i>OBMTR 7</i>	0,804	0,255	0,000	$\times OBMTR_10$	3760,49	0,060	0,000
<i>OBMTR 8</i>	0,279	0,115	0,000	$\times X_KOORD$	82,97	0,043	0,036
<i>OBMTR 9</i>	-0,079	-0,019	0,252	$\times Y_KOORD$	362,61	0,121	0,000
<i>OBMTR 10</i>	0,070	0,017	0,338	$\times RAZV_OBC$	3348,95	0,463	0,000
<i>AC</i>	0,068	0,041	0,000	$\times AC$	653,60	0,035	0,000
<i>AC LJ</i>	0,012	0,009	0,568	$\times AC_LJ$	168,59	0,009	0,513
				<i>CAS TRAN</i>	-16787,66	-0,590	0,000

Tabela P5-2: (nadaljevanje)

Odpisna spremenljivka: $CENA_NA_m^2$							
združeni model							
$R^2 = 0,600$; popravljene $R^2 = 0,599$							
Neodvisna sprem.	Regres. koef.	Stand. koef.	Znač. koef.	Neodvisna sprem.	Regres. koef.	Stand. koef.	Znač. koef.
<i>konstanta</i>	105540,5	-	0,017				
<i>CAS TRAN</i>	-10968,11	-0,386	0,012	<i>CAS TRAN</i> ×			
<i>MESEC</i>	874,18	0,030	0,029	× <i>MESEC</i>	-65,08	-0,031	0,090
<i>ST VLOZ</i>	-15458,20	-0,076	0,000	× <i>ST VLOZ</i>	429,39	0,024	0,137
<i>IZMERA</i>	-286,67	-0,088	0,000	× <i>IZMERA</i>	-8,72	-0,034	0,046
<i>STAROST</i>	-616,34	-0,194	0,000	× <i>STAROST</i>	1,16	0,004	0,789
<i>KLETNO</i>	-18803,11	-0,028	0,041	× <i>KLETNO</i>	-715,12	-0,010	0,442
<i>PODSTRES</i>	-13243,59	-0,027	0,053	× <i>PODSTRES</i>	-196,76	-0,004	0,759
<i>NADSTROP</i>	944,53	0,022	0,175	× <i>NADSTROP</i>	-104,74	-0,028	0,104
<i>LEGA V</i>	-737,64	-0,003	0,845	× <i>LEGA V</i>	415,63	0,021	0,239
<i>LEGA J</i>	-1916,94	-0,008	0,619	× <i>LEGA J</i>	897,51	0,040	0,015
<i>LEGA Z</i>	523,18	0,002	0,897	× <i>LEGA Z</i>	172,27	0,007	0,658
<i>DVIGALO</i>	-3556,19	-0,015	0,339	× <i>DVIGALO</i>	1531,15	0,071	0,000
<i>ODDALJEN</i>	-986,17	-0,023	0,124	× <i>ODDALJEN</i>	-70,38	-0,018	0,237
<i>ELEKTR</i>	-10191,60	-0,005	0,756	× <i>ELEKTR</i>	-186,64	-0,066	0,583
<i>VODOVOD</i>	-17484,25	-0,011	0,378	× <i>VODOVOD</i>	4764,74	0,170	0,027
<i>OGREVAN</i>	2045,01	0,005	0,762	× <i>OGREVA</i>	525,02	0,021	0,413
<i>TELEFON</i>	6327,82	0,022	0,118	× <i>TELEFON</i>	229,82	0,011	0,554
<i>KANALIZ</i>	15202,39	0,022	0,114	× <i>KANALIZ</i>	-53,76	-0,002	0,953
<i>OBMTR 2</i>	-30597,76	-0,093	0,013	× <i>OBMTR 2</i>	4027,77	0,134	0,001
<i>OBMTR 3</i>	-29488,22	-0,111	0,059	× <i>OBMTR 3</i>	4364,60	0,176	0,004
<i>OBMTR 4</i>	18612,75	0,054	0,411	× <i>OBMTR 4</i>	5723,41	0,178	0,008
<i>OBMTR 5</i>	67160,44	0,318	0,004	× <i>OBMTR 5</i>	7849,93	0,401	0,000
<i>OBMTR 6</i>	-24399,58	-0,046	0,296	× <i>OBMTR 6</i>	6638,49	0,131	0,003
<i>OBMTR 7</i>	51916,87	0,118	0,155	× <i>OBMTR 7</i>	10197,12	0,250	0,004
<i>OBMTR 8</i>	-1982,53	-0,006	0,932	× <i>OBMTR 8</i>	6251,83	0,199	0,005
<i>OBMTR 9</i>	-44232,67	-0,076	0,146	× <i>OBMTR 9</i>	5246,27	0,100	0,069
<i>OBMTR 10</i>	-4367,84	-0,008	0,885	× <i>OBMTR 10</i>	3472,52	0,056	0,234
<i>X KOORD</i>	83,57	0,004	0,943	× <i>X KOORD</i>	38,58	0,020	0,730
<i>Y KOORD</i>	40,35	0,001	0,978	× <i>Y KOORD</i>	324,65	0,108	0,021
<i>RAZV OBC</i>	20279,61	0,105	0,000	× <i>RAZV OBC</i>	1293,03	0,179	0,021
<i>AC</i>	2608,81	0,011	0,560	× <i>AC</i>	454,03	0,024	0,292
<i>AC LJ</i>	-5583,80	-0,028	0,517	× <i>AC LJ</i>	653,37	0,036	0,413

Opomba: *LEGA*_(i) ... Binarne spremenljivke, ki nadomestijo sprem. *LEGA*
*OBMTR*_(i) ... Binarne spremenljivke, ki nadomestijo sprem. *OBM_TRAN*

Priloga 6

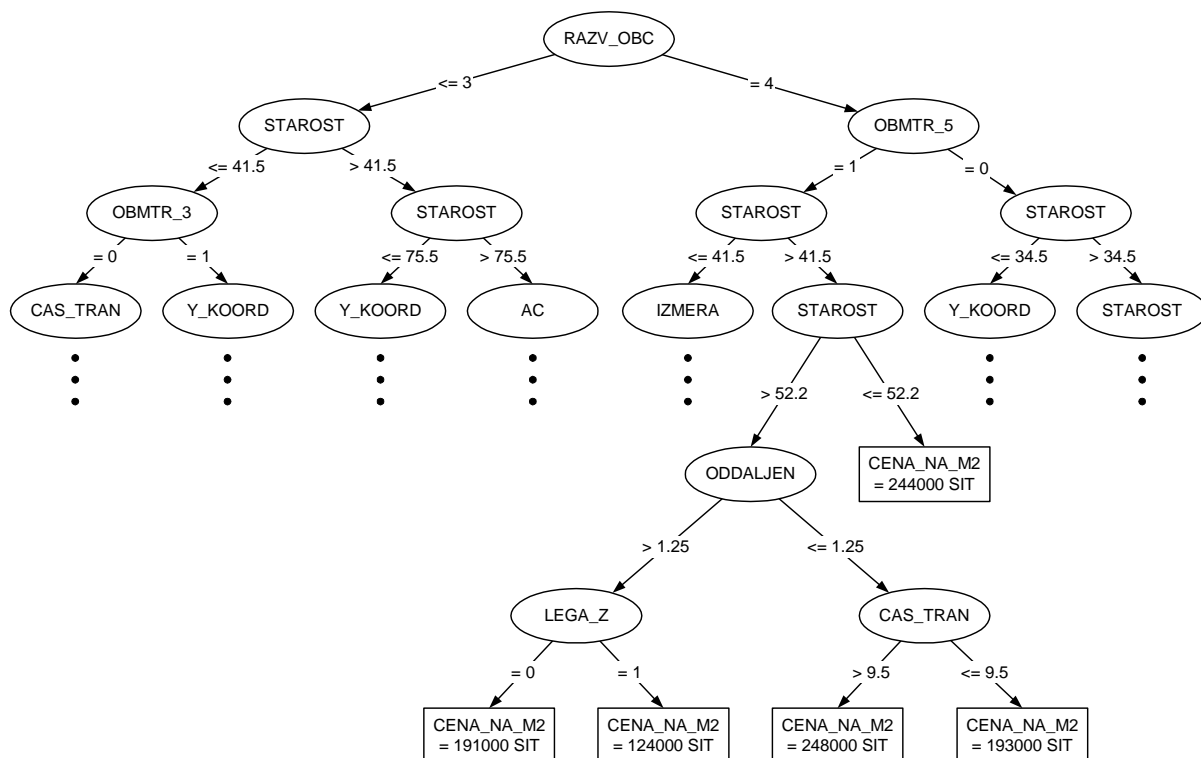
Multipli regresijski model za napoved cene kvadratnega metra stanovanj, zgrajen z učno-testno kombinacijo vzorcev K2 (izpis iz programa WEKA 3.2)

Scheme: weka.classifiers.LinearRegression -S 2
Instances: 13287
Attributes: 33

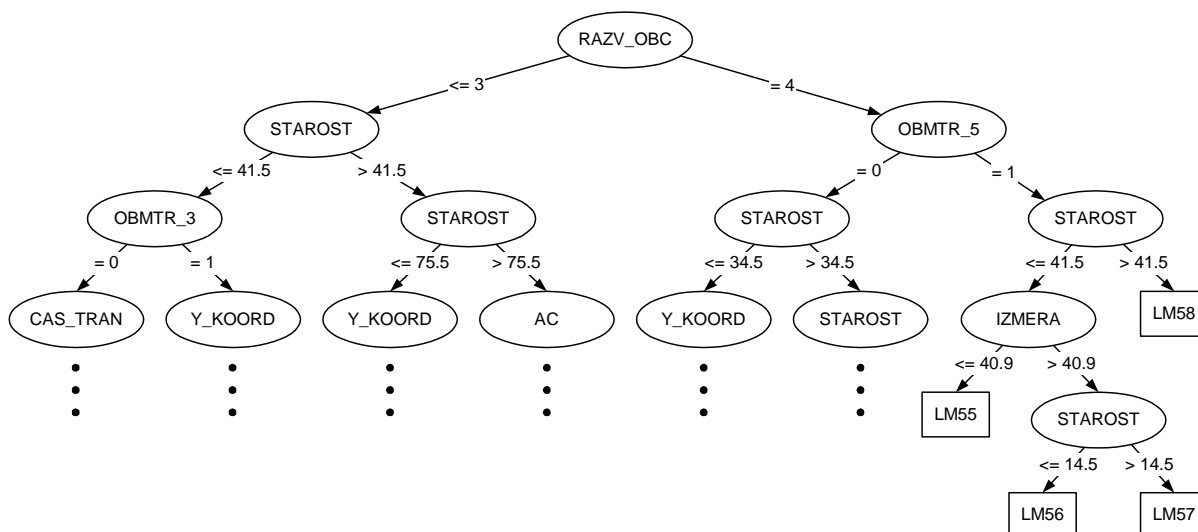
Linear Regression Model:

```
CENA_NA_M2 =  
  4158.9621 * CAS_TRAN +  
  351.5757 * MESEC +  
 -9331.9183 * ST_VLOZ +  
 -363.6702 * IZMERA +  
 -599.4726 * STAROST +  
 -24893.1765 * KLETNO +  
 -19066.0561 * PODSTRES +  
  3895.7879 * LEGA_V +  
  7166.173 * LEGA_J +  
  3336.4015 * LEGA_Z +  
 11140.9723 * DVIGALO +  
 -1787.5169 * ODDALJEN +  
  9451.7816 * OGREVAN +  
  8995.2552 * TELEFON +  
 18167.3534 * KANALIZ +  
  5844.6913 * OBMTR_2 +  
 10356.7229 * OBMTR_3 +  
  70718.8708 * OBMTR_4 +  
 138243.8664 * OBMTR_5 +  
  32710.1087 * OBMTR_6 +  
 141481.6044 * OBMTR_7 +  
  57485.7378 * OBMTR_8 +  
 25860.7915 * OBMTR_10 +  
   494.5459 * X_KOORD +  
  2498.3345 * Y_KOORD +  
 34319.999 * RAZV_OBC +  
  8560.9196 * AC +  
 -51856.2549
```


Izsek (za en del mesta Ljubljane) iz regresijskega drevesa za napoved cene kvadratnega metra stanovanj, zgrajenega z učno-testno kombinacijo vzorcev K2 (celotno regresijsko drevo, ki ima 110 pravil-listov, se nahaja na priloženi zgoščenki)



Izsek (za mesto Ljubljana) iz modelnega drevca za napoved cene kvadratnega metra stanovanj, zgrajenega z učno-testno kombinacijo vzorcev K2 (celotno modelno drevo, ki ima 58 pravil-listov, se nahaja na priloženi zgoščenki)



Zglajeni regresijski modeli v listih drevesa:

LM55: $CENA_NA_M2 = 236000 + 61.5*RAZV_OBC + 7280*CAS_TRAN - 105*ST_VLOZ - 16.2*IZMERA - 1010*STAROST - 328*KLETNO - 468*PODSTRES + 28.9*LEGA_J + 118*DVIGALO - 78.3*ODDALJEN + 16.6*OGREVAN + 30000*TELEFON + 114*KANALIZ + 12.9*OBMTR_3 + 207*OBMTR_4 + 553*OBMTR_5 + 134*OBMTR_6 + 377*OBMTR_7 + 96.4*OBMTR_8 - 200*OBMTR_9 - 169*OBMTR_10 + 123*AC - 88.1*AC_LJ - 13.4*X_KOORD + 19.3*Y_KOORD$

LM56: $CENA_NA_M2 = 231000 + 61.5*RAZV_OBC + 8000*CAS_TRAN - 354*ST_VLOZ - 24.2*IZMERA - 65.7*STAROST - 328*KLETNO - 1110*PODSTRES + 28.9*LEGA_J + 32400*DVIGALO - 7660*ODDALJEN + 16.6*OGREVAN + 287*TELEFON + 114*KANALIZ + 12.9*OBMTR_3 + 207*OBMTR_4 + 553*OBMTR_5 + 134*OBMTR_6 + 377*OBMTR_7 + 96.4*OBMTR_8 - 200*OBMTR_9 - 169*OBMTR_10 + 123*AC - 88.1*AC_LJ - 13.4*X_KOORD + 19.3*Y_KOORD$

LM57: $CENA_NA_M2 = 275000 + 61.5*RAZV_OBC + 5050*CAS_TRAN - 9890*ST_VLOZ - 733*IZMERA - 991*STAROST - 328*KLETNO - 572*PODSTRES + 28.9*LEGA_J + 118*DVIGALO - 4810*ODDALJEN + 16.6*OGREVAN + 23800*TELEFON + 114*KANALIZ + 12.9*OBMTR_3 + 207*OBMTR_4 + 553*OBMTR_5 + 134*OBMTR_6 + 377*OBMTR_7 + 96.4*OBMTR_8 - 200*OBMTR_9 - 169*OBMTR_10 + 123*AC - 88.1*AC_LJ - 13.4*X_KOORD + 19.3*Y_KOORD$

LM58: $CENA_NA_M2 = 181000 + 61.5*RAZV_OBC + 7080*CAS_TRAN - 25900*ST_VLOZ - 381*IZMERA - 10.9*STAROST - 777*KLETNO - 39500*PODSTRES + 10800NADSTROP + 28.9*LEGA_J - 30600*LEGA_Z + 247*DVIGALO - 62.5*ODDALJEN + 16.6*OGREVAN + 339*TELEFON + 114*KANALIZ + 12.9*OBMTR_3 + 207*OBMTR_4 + 553*OBMTR_5 + 134*OBMTR_6 + 377*OBMTR_7 + 96.4*OBMTR_8 - 200*OBMTR_9 - 169*OBMTR_10 + 123*AC - 88.1*AC_LJ - 13.4*X_KOORD + 19.3*Y_KOORD$

Model za napoved cene kvadratnega metra stanovanj, zgrajen z metodo CIPER in učno-testno kombinacijo vzorcev K2 (izpis iz tekstovnega okna operacijskega sistema Windows)

```
D:\USER\JernejK\CIPER>ciper -t K2_Trn.arff -T K2_Tst.arff -b 5 -p 0.5
```

```
Reading training file K2_Trn.arff...done
```

```
Beam size: 5  
Depth: 2147483647
```

```
CENA_NA_M2 =  
+ 865.541 * RAZV_OBC * CAS_TRAN  
- 58.3083 * STAROST * RAZV_OBC * RAZV_OBC  
+ 89343.2 * OBMTR_5  
- 382.058 * IZMERA * RAZV_OBC  
+ 0.123385 * CAS_TRAN * OBMTR_5 * STAROST * STAROST  
+ 2104.99 * OBMTR_7 * X_KOORD * Y_KOORD  
+ 45414 * OBMTR_4  
- 23746.4 * OBMTR_9  
+ 38516.6 * OBMTR_8  
- 7097.9 * ST_VLOZ  
+ 11125.5 * DVIGALO  
+ 10658 * TELEFON  
- 1294.38 * ODDALJEN * RAZV_OBC  
+ 0.0520256 * IZMERA * IZMERA * RAZV_OBC * RAZV_OBC  
+ 35.9586 * AC * X_KOORD * STAROST  
- 26051.4 * KLETNO  
- 19511.8 * PODSTRES  
+ 149.282 * ODDALJEN * ODDALJEN  
+ 24313.5 * RAZV_OBC * RAZV_OBC  
+ 2774.99 * CAS_TRAN * OBMTR_5  
+ 22241.7 * KANALIZ  
- 91143.6 * RAZV_OBC  
+ 1728.17 * Y_KOORD * OGREVAN  
+ 10516.4 * OGREVAN  
- 15060 * OBMTR_1  
- 4338.66 * LEGA_S  
+ 748.529 * IZMERA  
+ 134134
```

```
Reading testing file K2_Tst.arff...done
```

```
RE: 0.373362  
r: 0.791829
```

Regressijski modeli za napoved cene kvadratnega metra stanovanj za posamezna območja transakcij; učne množice so bile sestavljene iz vseh transakcij na posameznih območjih (izpisi iz programa WEKA 3.2)

Območje transakcij 1

Scheme: weka.classifiers.LinearRegression -S 2
Instances: 412
Attributes: 24

Linear Regression Model:

```
CENA_NA_M2 =  
-33541.2741 * RAZV_OBC +  
729.1189 * MESEC +  
-934.8336 * STAROST +  
-757.0925 * ODDALJEN +  
-128833.6296 * ELEKTR +  
56684.6726 * KANALIZ +  
31516.9339 * AC +  
280230.5303
```

Območje transakcij 2

Scheme: weka.classifiers.LinearRegression -S 2
Instances: 2102
Attributes: 24

Linear Regression Model:

```
CENA_NA_M2 =  
9600.7689 * RAZV_OBC +  
1983.6882 * CAS_TRAN +  
-4611.3665 * ST_VLOZ +  
-91.1976 * IZMERA +  
-393.5054 * STAROST +  
-20333.7186 * KLETNO +  
906.4008 * NADSTROP +  
11526.2222 * LEGA_V +  
9714.244 * LEGA_J +  
14650.7984 * LEGA_Z +  
6062.1499 * DVIGALO +  
-1818.4892 * ODDALJEN +  
10660.1014 * OGREVAN +  
7669.7923 * TELEFON +  
23981.3502 * KANALIZ +  
23325.7051 * AC +  
2892.7689 * X_KOORD +  
966.0915
```

Območje transakcij 3

Scheme: weka.classifiers.LinearRegression -S 2
Instances: 3507
Attributes: 24

Linear Regression Model:

```
CENA_NA_M2 =  
7301.9835 * RAZV_OBC +  
650.6879 * MESEC +  
1778.6831 * CAS_TRAN +  
-6943.6381 * ST_VLOZ +  
-142.8545 * IZMERA +  
-405.5805 * STAROST +  
-16588.2256 * PODSTRES +  
-5098.5447 * LEGA_V +  
-5870.9009 * LEGA_Z +  
5050.6965 * DVIGALO +  
-2084.1349 * ODDALJEN +  
12751.4276 * OGREVAN +  
9413.7346 * TELEFON +  
13495.9233 * KANALIZ +  
4837.8517 * AC +  
5410.4009 * Y_KOORD +  
47134.2536
```

Območje transakcij 4

Scheme: weka.classifiers.LinearRegression -S 2
Instances: 1848
Attributes: 24

Linear Regression Model:

```
CENA_NA_M2 =  
78269.2548 * RAZV_OBC +  
3827.3345 * CAS_TRAN +  
-18796.4738 * ST_VLOZ +  
-607.1978 * IZMERA +  
-961.3988 * STAROST +  
-52891.9674 * KLETNO +  
-20858.2491 * PODSTRES +  
23711.5155 * DVIGALO +  
-3094.5509 * ODDALJEN +  
-60681.6285 * ELEKTR +  
44459.0751 * VODOVOD +  
37325.7597 * OGREVAN +  
25034.2034 * KANALIZ +  
-11237.6335 * AC_LJ +  
5015.0817 * X_KOORD +  
-107632.5031
```

Območje transakcij 5

Scheme: weka.classifiers.LinearRegression -S 2
Instances: 7009
Attributes: 24

Linear Regression Model:

```
CENA_NA_M2 =  
457.1037 * MESEC +  
6401.4815 * CAS_TRAN +  
-15713.0396 * ST_VLOZ +  
-665.342 * IZMERA +  
-553.4086 * STAROST +  
-42256.3352 * KLETNO +  
-18802.9025 * PODSTRES +  
-619.4491 * NADSTROP +  
4150.4251 * LEGA_V +
```

Območje transakcij 6

Scheme: weka.classifiers.LinearRegression -S 2
Instances: 732
Attributes: 24

Linear Regression Model:

```
CENA_NA_M2 =  
15064.8697 * RAZV_OBC +  
2797.319 * CAS_TRAN +  
-240.6951 * IZMERA +  
-670.7449 * STAROST +  
-27010.1291 * KLETNO +  
-24737.6498 * PODSTRES +  
1849.5033 * NADSTROP +  
-2449.8686 * ODDALJEN +  
45791.0412 * VODOVOD +
```

9402.1952 * LEGA_J +
14843.6149 * DVIGALO +
-3462.8308 * ODDALJEN +
-67156.2693 * ELEKTR +
35132.5514 * VODOVOD +
17739.5977 * TELEFON +
284853.8236

Območje transakcij 7

Scheme: weka.classifiers.LinearRegression -S 2
Instances: 1114
Attributes: 24

Linear Regression Model:

CENA_NA_M2 =
5276.7672 * CAS_TRAN +
-481.5878 * IZMERA +
-1171.7971 * STAROST +
-13084.7041 * LEGA_V +
11073.0492 * LEGA_J +
-95928.1552 * ELEKTR +
107994.4705 * VODOVOD +
37843.3626 * KANALIZ +
-10419.34 * X_KOORD +
-26075.1296 * Y_KOORD +
-51036.7397

Območje transakcij 9

Scheme: weka.classifiers.LinearRegression -S 2
Instances: 618
Attributes: 24

Linear Regression Model:

CENA_NA_M2 =
-22340.8516 * RAZV_OBC +
2754.2759 * CAS_TRAN +
5409.6681 * ST_VLOZ +
-119.0483 * IZMERA +
-927.215 * STAROST +
-27250.2485 * PODSTRES +
4457.8906 * NADSTROP +
6995.4128 * LEGA_V +
-66769.05 * DVIGALO +
-3475.6374 * ODDALJEN +
11686.6054 * TELEFON +
15252.8313 * KANALIZ +
-14367.9823 * AC_LJ +
-12466.4303 * X_KOORD +
11851.6872 * Y_KOORD +
182153.8041

-26182.1397 * OGREVAN +
17561.5814 * KANALIZ +
3333.8563 * X_KOORD +
17252.1439 * Y_KOORD +
101987.907

Območje transakcij 8

Scheme: weka.classifiers.LinearRegression -S 2
Instances: 1974
Attributes: 24

Linear Regression Model:

CENA_NA_M2 =
65029.7621 * RAZV_OBC +
-1056.9829 * MESEC +
4793.0041 * CAS_TRAN +
-5528.4117 * ST_VLOZ +
-322.9963 * IZMERA +
-551.8574 * STAROST +
-25924.4524 * KLETNO +
-31185.6643 * PODSTRES +
18049.2997 * LEGA_V +
17056.4681 * LEGA_J +
17904.4425 * LEGA_Z +
13383.7711 * DVIGALO +
-1574.0952 * ODDALJEN +
39532.6349 * OGREVAN +
13042.5817 * TELEFON +
26979.1954 * KANALIZ +
-51553.1394 * AC +
51645.8381 * AC_LJ +
-6219.6179 * X_KOORD +
14555.1357 * Y_KOORD +
-199812.134

Območje transakcij 10

Scheme: weka.classifiers.LinearRegression -S 2
Instances: 622
Attributes: 24

Linear Regression Model:

CENA_NA_M2 =
1334.4201 * CAS_TRAN +
-324.0918 * IZMERA +
-558.3977 * STAROST +
2300.7291 * NADSTROP +
10642.4405 * LEGA_V +
7721.9041 * LEGA_Z +
16508.7408 * OGREVAN +
26511.4295 * TELEFON +
46606.1815 * AC +
-33441.3358 * AC_LJ +
-7916.7875 * X_KOORD +
34827.3568

Modelna drevesa za napoved cene kvadratnega metra stanovanj za posamezna območja transakcij; učne množice so bile sestavljene iz vseh transakcij na posameznih območjih (izpisi iz programa WEKA 3.2)

Območje transakcij 1

Scheme: weka.classifiers.m5.M5Prime -O m -F 3.0 -V 0
Instances: 412
Attributes: 24

Pruned training model tree (number of rules = 3):

```
X_KOORD <= 13.8 : LM1
X_KOORD > 13.8 :
| STAROST <= 12.5 : LM2
| STAROST > 12.5 : LM3
```

Models at the leaves:

LM1: CENA_NA_M2 = 100000 - 1570*RAZV_OBC - 638*STAROST + 11800*TELEFON+ 3010*KANALIZ + 28100*AC
LM2: CENA_NA_M2 = 153000 - 23800*RAZV_OBC - 385*STAROST + 38200*KANALIZ + 71300*AC
LM3: CENA_NA_M2 = 73600 - 11000*RAZV_OBC - 206*STAROST + 46200KANALIZ + 3760*AC

Območje transakcij 2

Scheme: weka.classifiers.m5.M5Prime -O m -F 3.0 -V 0
Instances: 2102
Attributes: 24

Pruned training model tree (number of rules = 5):

```
AC <= 0.5 :
| STAROST <= 39.5 : LM1
| STAROST > 39.5 :
| | STAROST <= 64.5 :
| | | Y_KOORD <= 5.55 : LM2
| | | Y_KOORD > 5.55 : LM3
| | STAROST > 64.5 : LM4
AC > 0.5 : LM5
```

Models at the leaves:

LM1: CENA_NA_M2 = 19800 + 25600*RAZV_OBC + 1680*CAS_TRAN - 3.55*IZMERA - 997*STAROST - 385*KLETNO + 203*LEGA_V + 177*LEGA_J + 261*LEGA_Z + 144*DVIGALO - 33.4*ODDALJEN + 397*TELEFON + 481*KANALIZ + 41*9AC + 2340*X_KOORD - 131*Y_KOORD
LM2: CENA_NA_M2 = 100000 + 1120*RAZV_OBC + 1990*CAS_TRAN - 244*IZMERA - 207*STAROST - 385*KLETNO - 3510*PODSTRES + 203*LEGA_V + 177*LEGA_J + 261*LEGA_Z + 144*DVIGALO - 33.4*ODDALJEN+ 810*TELEFON + 481*KANALIZ + 419*AC + 157*X_KOORD - 2830*Y_KOORD
LM3: CENA_NA_M2 = 89900 + 1120*RAZV_OBC + 103*CAS_TRAN - 6.91*IZMERA - 265*STAROST - 4860*PODSTRES + 203*LEGA_V + 177*LEGA_J + 261*LEGA_Z + 144*DVIGALO - 33.4*ODDALJEN + 810*TELEFON + 481*KANALIZ + 419*AC + 157*X_KOORD - 3400*Y_KOORD
LM4: CENA_NA_M2 = 56200 + 1120*RAZV_OBC + 103*CAS_TRAN - 6.91*IZMERA - 78.8*STAROST - 385*KLETNO + 203*LEGA_V + 177*LEGA_J + 261*LEGA_Z + 144*DVIGALO - 33.4*ODDALJEN + 810*TELEFON + 481*KANALIZ + 419*AC + 157*X_KOORD - 2460*Y_KOORD
LM5: CENA_NA_M2 = 370000 + 2870*CAS_TRAN - 1.17*IZMERA - 349*STAROST - 283*KLETNO + 14500*LEGA_V + 16100*LEGA_J + 20200*LEGA_Z + 106*DVIGALO - 3020*ODDALJEN + 15700OGREVAN + 115*TELEFON + 39100*KANALIZ + 308*AC - 47400*X_KOORD + 18200*Y_KOORD

Območje transakcij 3

Scheme: weka.classifiers.m5.M5Prime -O m -F 3.0 -V 0
Instances: 3507
Attributes: 24

Pruned training model tree (number of models = 17):

```
STAROST <= 41.5 :
| Y_KOORD <= 2.25 :
| | X_KOORD <= 6.45 : LM1
| | X_KOORD > 6.45 : LM2
| Y_KOORD > 2.25 :
| | CAS_TRAN <= 9.83 :
| | | Y_KOORD <= 3.3 :
| | | | CAS_TRAN <= 5.17 : LM3
| | | | CAS_TRAN > 5.17 :
| | | | | X_KOORD <= 5.65 :
| | | | | *ODDALJEN <= 2.5 : LM4
| | | | | *ODDALJEN > 2.5 : LM5
| | | | | X_KOORD > 5.65 : LM6
| | | | Y_KOORD > 3.3 : LM7
| | CAS_TRAN > 9.83 :
| | | IZMERA <= 38.8 :
| | | | Y_KOORD <= 3.75 : LM8
| | | | Y_KOORD > 3.75 : LM9
| | | | IZMERA > 38.8 :
| | | | | STAROST <= 32.5 : LM10
| | | | | STAROST > 32.5 :
| | | | | | TELEFON <= 0.5 :
| | | | | | CAS_TRAN <= 13.8 : LM11
| | | | | | CAS_TRAN > 13.8 : LM12
| | | | | TELEFON > 0.5 : LM13
STAROST > 41.5 :
| STAROST <= 49.5 :
| | Y_KOORD <= 1.95 :
| | | Y_KOORD <= 1.75 : LM14
| | | Y_KOORD > 1.75 : LM15
| | Y_KOORD > 1.95 : LM16
| STAROST > 49.5 : LM17
```

Models at the leaves:

```
LM1: CENA_NA_M2 = 105000 + 14.5*MESEC + 2120*CAS_TRAN - 214*ST_VLOZ - 150*IZMERA
      - 633*STAROST - 492*PODSTRES - 236*LEGA_V - 130*LEGA_Z + 115*DVIGALO
      - 1340*ODDALJEN + 70.7*OGREVAN + 223*TELEFON + 407*KANALIZ
      + 28.6*AC - 114*X_KOORD + 4970*Y_KOORD

LM2: CENA_NA_M2 = -82600 + 16800*RAZV_OBC + 14.5*MESEC + 118*CAS_TRAN - 214*ST_VLOZ
      - 12.5*IZMERA - 45.9*STAROST - 492*PODSTRES - 317*LEGA_V - 130*LEGA_Z
      + 115*DVIGALO - 56.5*ODDALJEN + 70.7*OGREVAN + 223*TELEFON + 407*KANALIZ
      + 16400*AC + 15800*X_KOORD - 3.68*Y_KOORD

LM3: CENA_NA_M2 = 32000 + 7.85*MESEC + 351*CAS_TRAN - 19600*ST_VLOZ - 9.81*IZMERA
      - 70.2*STAROST - 4890*PODSTRES - 28.8*LEGA_V - 18900*LEGA_Z + 62.3*DVIGALO
      - 500*ODDALJEN + 364*OGREVAN + 19800*TELEFON + 2390*KANALIZ - 254*AC
      + 11900*X_KOORD - 1210*Y_KOORD

LM4: CENA_NA_M2 = 126000 + 7.85*MESEC - 1080*CAS_TRAN - 390*ST_VLOZ - 28.7*IZMERA
      - 315*STAROST - 8260*PODSTRES - 28.8*LEGA_V - 296*LEGA_Z + 62.3*DVIGALO
      - 784*ODDALJEN + 364*OGREVAN + 443*TELEFON + 11300*KANALIZ - 254*AC
      + 66.8*X_KOORD - 2240*Y_KOORD

LM5: CENA_NA_M2 = 153000 + 7.85*MESEC - 9180*CAS_TRAN - 390*ST_VLOZ - 28.7*IZMERA
      - 296*STAROST - 8260*PODSTRES - 28.8*LEGA_V - 296*LEGA_Z + 62.3*DVIGALO
      - 784*ODDALJEN + 364*OGREVAN + 443*TELEFON + 32800*KANALIZ - 254*AC
      + 66.8*X_KOORD - 2240*Y_KOORD

LM6: CENA_NA_M2 = 141000 + 7.85*MESEC + 229*CAS_TRAN - 390*ST_VLOZ - 197*IZMERA
      - 436*STAROST - 51800*PODSTRES - 28.8*LEGA_V - 296*LEGA_Z + 62.3*DVIGALO
      - 504*ODDALJEN + 364*OGREVAN + 443*TELEFON + 2600*KANALIZ - 254*AC
      + 66.8*X_KOORD - 1220*Y_KOORD

LM7: CENA_NA_M2 = 111000 + 7.85*MESEC + 3670*CAS_TRAN - 553*ST_VLOZ - 220*IZMERA
      - 450*STAROST - 2740*PODSTRES - 28.8*LEGA_V - 431*LEGA_Z + 62.3*DVIGALO
      - 3560*ODDALJEN + 364*OGREVAN + 611*TELEFON + 1110*KANALIZ - 254*AC
      + 137*X_KOORD - 76.2*Y_KOORD

LM8: CENA_NA_M2 = 222000 + 7.85*MESEC + 732*CAS_TRAN - 833*ST_VLOZ - 19.9*IZMERA
      - 974*STAROST - 635*PODSTRES - 28.8*LEGA_V - 2320*LEGA_Z + 2270*DVIGALO
      - 229*ODDALJEN + 358*OGREVAN + 21600*TELEFON + 205*KANALIZ - 1700*AC
      - 35400*Y_KOORD

LM9: CENA_NA_M2 = 137000 + 7.85*MESEC + 1120*CAS_TRAN - 833*ST_VLOZ - 19.9*IZMERA
      - 41.3*STAROST - 635*PODSTRES - 28.8*LEGA_V - 3730*LEGA_Z + 3710*DVIGALO
      - 229*ODDALJEN + 358*OGREVAN + 1160*TELEFON + 205*KANALIZ - 1700*AC
      + 2380*Y_KOORD
```

LM10: $CENA_NA_M2 = 141000 + 7.85*MESEC + 2210*CAS_TRAN - 15100*ST_VLOZ - 10.5*IZMERA$
 $- 36.8*STAROST - 26800*PODSTRES - 28.8*LEGA_V + 8850*LEGA_J - 325*LEGA_Z$
 $+ 224*DVIGALO - 2560*ODDALJEN + 358*OGREVAN + 1000*TELEFON + 205*KANALIZ$
 $- 14600*AC - 7700*Y_KOORD$

LM11: $CENA_NA_M2 = 139000 + 7.85*MESEC + 1530*CAS_TRAN - 1470*ST_VLOZ - 184*IZMERA$
 $- 864*STAROST - 2980*PODSTRES - 28.8*LEGA_V - 325*LEGA_Z + 224*DVIGALO$
 $- 346*ODDALJEN + 358*OGREVAN + 13000*TELEFON + 205*KANALIZ - 19400*AC$
 $- 8790*Y_KOORD$

LM12: $CENA_NA_M2 = 161000 + 7.85*MESEC + 1530*CAS_TRAN - 1470*ST_VLOZ - 184*IZMERA$
 $- 864*STAROST - 2980*PODSTRES - 28.8*LEGA_V - 325*LEGA_Z + 224*DVIGALO$
 $- 346*ODDALJEN + 358*OGREVAN + 13000*TELEFON + 205*KANALIZ - 19400*AC$
 $- 8790*Y_KOORD$

LM13: $CENA_NA_M2 = 301000 + 7.85*MESEC + 890*CAS_TRAN - 18100*ST_VLOZ - 661*IZMERA$
 $- 438*STAROST - 2980*PODSTRES - 28.8*LEGA_V - 325*LEGA_Z + 224*DVIGALO$
 $- 346*ODDALJEN + 358*OGREVAN + 7160*TELEFON + 205*KANALIZ - 62400*AC$
 $- 30000*Y_KOORD$

LM14: $CENA_NA_M2 = 61500 + 254*MESEC + 2250*CAS_TRAN - 117*ST_VLOZ - 34.7*IZMERA$
 $- 17.6*STAROST - 912*PODSTRES + 142*NADSTROP - 84.9*LEGA_V$
 $- 97.5*LEGA_Z + 86*DVIGALO - 112*ODDALJEN + 2460*OGREVAN$
 $+ 561*TELEFON + 229*KANALIZ + 392*AC + 1070*Y_KOORD$

LM15: $CENA_NA_M2 = 61200 + 658*MESEC + 261*CAS_TRAN - 117*ST_VLOZ - 34.7*IZMERA$
 $- 17.6*STAROST - 912*PODSTRES + 142*NADSTROP - 84.9*LEGA_V$
 $- 97.5*LEGA_Z + 86*DVIGALO - 112*ODDALJEN + 2460*OGREVAN$
 $+ 561*TELEFON + 229*KANALIZ + 392*AC + 1070*Y_KOORD$

LM16: $CENA_NA_M2 = 114000 + 10.8*MESEC + 2350*CAS_TRAN - 117*ST_VLOZ - 254*IZMERA$
 $- 17.6*STAROST - 912*PODSTRES + 142*NADSTROP - 84.9*LEGA_V$
 $- 97.5*LEGA_Z + 86*DVIGALO - 112*ODDALJEN + 21300*OGREVAN$
 $+ 561*TELEFON + 229*KANALIZ - 15400*AC - 7420*Y_KOORD$

LM17: $CENA_NA_M2 = 76300 + 10.8*MESEC + 30*CAS_TRAN - 117*ST_VLOZ - 4.77*IZMERA$
 $- 190*STAROST - 19100*PODSTRES + 117*NADSTROP - 84.9*LEGA_V$
 $- 97.5*LEGA_Z + 86*DVIGALO - 99.1*ODDALJEN + 512*OGREVAN$
 $+ 491*TELEFON + 229*KANALIZ + 13800*AC + 270*Y_KOORD$

Območje transakcij 4

Scheme: weka.classifiers.m5.M5Prime -O m -F 3.0 -V 0
Instances: 1848
Attributes: 24

Pruned training model tree (number of models = 10):

```

ST_VLOZ <= 0.5 :
| IZMERA <= 42.6 :
| | X_KOORD <= 0.6 : LM1
| | X_KOORD > 0.6 : LM2
| IZMERA > 42.6 :
| | STAROST <= 26.5 : LM3
| | STAROST > 26.5 :
| | | X_KOORD <= 0.6 : LM4
| | | X_KOORD > 0.6 :
| | | | MESEC <= 10.5 : LM5
| | | | MESEC > 10.5 : LM6
ST_VLOZ > 0.5 :
| STAROST <= 39.5 : LM7
| STAROST > 39.5 :
| | STAROST <= 52.5 : LM8
| | STAROST > 52.5 :
| | | CAS_TRAN <= 6.17 : LM9
| | | CAS_TRAN > 6.17 : LM10

```

Models at the leaves:

LM1: $CENA_NA_M2 = 274000 + 7790*RAZV_OBC + 5400*CAS_TRAN - 305*ST_VLOZ - 1710*IZMERA$
 $- 1550*STAROST - 3280*KLETNO - 42300*PODSTRES + 314*NADSTROP$
 $+ 1420*DVIGALO - 701*ODDALJEN + 642*OGREVAN + 424*KANALIZ$
 $- 183*AC + 11200*X_KOORD$

LM2: $CENA_NA_M2 = -24200 + 49100*RAZV_OBC + 6200*CAS_TRAN - 305*ST_VLOZ - 45.5*IZMERA$
 $- 136*STAROST - 3280*KLETNO - 335*PODSTRES + 540*NADSTROP + 1420*DVIGALO$
 $- 1010*ODDALJEN + 642*OGREVAN + 424*KANALIZ - 183*AC + 79.7*X_KOORD$

LM3: $CENA_NA_M2 = -51800 + 72600*RAZV_OBC - 1780*MESEC + 5620*CAS_TRAN - 305*ST_VLOZ$
 $- 565*IZMERA - 1850*STAROST - 43000*KLETNO - 335*PODSTRES$
 $+ 26100*DVIGALO - 6630*ODDALJEN + 642*OGREVAN + 424*KANALIZ - 183*AC$
 $+ 79.7*X_KOORD$

LM4: $CENA_NA_M2 = 238000 + 9300*RAZV_OBC + 3190*CAS_TRAN - 305*ST_VLOZ - 1350*IZMERA - 779*STAROST - 7020*KLETNO - 335*PODSTRES + 3120*DVIGALO - 520*ODDALJEN + 642*OGREVAN + 424*KANALIZ - 183*AC + 9680*X_KOORD$

LM5: $CENA_NA_M2 = 21500 + 43000*RAZV_OBC + 526*CAS_TRAN - 305*ST_VLOZ - 431*IZMERA - 112*STAROST - 7020*KLETNO - 335*PODSTRES + 3120*DVIGALO - 520*ODDALJEN + 642*OGREVAN + 424*KANALIZ - 183*AC + 79.7*X_KOORD$

LM6: $CENA_NA_M2 = 22900 + 9300*RAZV_OBC + 526*CAS_TRAN - 305*ST_VLOZ + 1850*IZMERA - 112*STAROST - 7020*KLETNO - 335*PODSTRES + 3120*DVIGALO - 520*ODDALJEN + 642*OGREVAN + 424*KANALIZ - 183*AC + 79.7*X_KOORD$

LM7: $CENA_NA_M2 = -157000 + 72200*RAZV_OBC + 4480*CAS_TRAN - 305*ST_VLOZ - 678*IZMERA - 941*STAROST - 60800*KLETNO - 31400*PODSTRES + 384*DVIGALO - 49*ODDALJEN + 88900*OGREVAN + 423*KANALIZ - 182*AC + 5550*X_KOORD$

LM8: $CENA_NA_M2 = 117000 + 5950*RAZV_OBC + 264*CAS_TRAN - 305*ST_VLOZ - 46.7*IZMERA - 83.9*STAROST - 4380*KLETNO - 2230*PODSTRES + 384*DVIGALO - 49*ODDALJEN + 5240*OGREVAN + 423*KANALIZ - 182*AC + 463*X_KOORD$

LM9: $CENA_NA_M2 = 80300 + 5950*RAZV_OBC + 264*CAS_TRAN - 305*ST_VLOZ - 46.7*IZMERA - 83.9*STAROST - 4380*KLETNO - 2230*PODSTRES + 155000*LEGA_Z + 384*DVIGALO - 49*ODDALJEN + 5240*OGREVAN + 423*KANALIZ - 182*AC + 463*X_KOORD + 20700*Y_KOORD$

LM10: $CENA_NA_M2 = 77200 + 5950*RAZV_OBC + 264*CAS_TRAN - 305*ST_VLOZ - 46.7*IZMERA - 83.9*STAROST - 4380*KLETNO - 2230*PODSTRES + 8300*LEGA_Z + 384*DVIGALO - 49*ODDALJEN + 5240*OGREVAN + 423*KANALIZ - 182*AC + 463*X_KOORD$

Območje transakcij 5

Scheme: weka.classifiers.m5.M5Prime -O m -F 3.0 -V 0
 Instances: 7009
 Attributes: 24

Pruned training model tree (number of rules 5):

```

STAROST <= 41.5 :
| IZMERA <= 41 : LM1
| IZMERA > 41 :
| | STAROST <= 14.5 : LM2
| | STAROST > 14.5 : LM3
STAROST > 41.5 :
| STAROST <= 52.5 : LM4
| STAROST > 52.5 : LM5

```

Models at the leaves:

LM1: $CENA_NA_M2 = 241000 + 7580*CAS_TRAN - 106*ST_VLOZ - 9.06*IZMERA - 1060*STAROST - 114*KLETNO - 198*PODSTRES + 20.9*LEGA_J + 79.9*DVIGALO - 42.1*ODDALJEN + 25800*TELEFON$

LM2: $CENA_NA_M2 = 225000 + 8230*CAS_TRAN - 254*ST_VLOZ - 14.9*IZMERA - 41.5*STAROST - 114*KLETNO - 586*PODSTRES - 21.2*NADSTROP + 20.9*LEGA_J + 30600*DVIGALO - 6290*ODDALJEN + 122*TELEFON$

LM3: $CENA_NA_M2 = 286000 + 5150*CAS_TRAN - 11000*ST_VLOZ - 763*IZMERA - 1130*STAROST - 114*KLETNO - 266*PODSTRES - 6.26*NADSTROP + 20.9*LEGA_J + 97.7*DVIGALO - 4500*ODDALJEN + 18300*TELEFON$

LM4: $CENA_NA_M2 = 219000 + 5020*CAS_TRAN - 822*ST_VLOZ - 763*IZMERA - 5.31*STAROST - 401*KLETNO - 952*PODSTRES + 7590*NADSTROP + 73.7*LEGA_J - 678*LEGA_Z + 128*DVIGALO - 33.8*ODDALJEN + 731*TELEFON$

LM5: $CENA_NA_M2 = 161000 + 6810*CAS_TRAN - 34100*ST_VLOZ - 12*IZMERA - 5.31*STAROST - 401*KLETNO - 670*PODSTRES + 9630*NADSTROP + 73.7*LEGA_J - 41800*LEGA_Z + 128*DVIGALO - 33.8*ODDALJEN + 527*TELEFON$

Območje transakcij 6

Scheme: weka.classifiers.m5.M5Prime -O m -F 3.0 -V 0
 Instances: 732
 Attributes: 24

Pruned training model tree (number of rules 3):

```

Y_KOORD <= -3.65 :
| STAROST <= 44.5 : LM1
| STAROST > 44.5 : LM2
Y_KOORD > -3.65 : LM3

```

Models at the leaves:

LM1: $CENA_NA_M2 = 137000 + 1020*RAZV_OBC + 2460*CAS_TRAN - 373*IZMERA - 619*STAROST - 1130*PODSTRES + 5040*NADSTROP + 14900*LEGA_Z - 120*ODDALJEN + 7640*Y_KOORD$

LM2: $CENA_NA_M2 = 166000 - 32800*RAZV_OBC + 585*CAS_TRAN - 69.9*IZMERA - 117*STAROST - 1130*PODSTRES + 1340*NADSTROP - 120*ODDALJEN + 727*Y_KOORD$

LM3: $CENA_NA_M2 = 175000 + 749*RAZV_OBC + 2960*CAS_TRAN - 249*IZMERA - 1220*STAROST - 828*PODSTRES + 103*NADSTROP - 2290*ODDALJEN + 535*Y_KOORD$

Območje transakcij 7

Scheme: weka.classifiers.m5.M5Prime -O m -F 3.0 -V 0
Instances: 1114
Attributes: 24

Pruned training model tree (number of rules = 1):

LM1: $CENA_NA_M2 = 36900 + 5530*CAS_TRAN - 477*IZMERA - 1180*STAROST + 15600*LEGA_J - 31800*Y_KOORD$

Območje transakcij 8

Scheme: weka.classifiers.m5.M5Prime -O m -F 3.0 -V 0
Instances: 1974
Attributes: 24

Pruned training model tree (number of rules = 8):

STAROST <= 41.5 :
| RAZV_OBC <= 3.5 : LM1
| RAZV_OBC > 3.5 :
| | IZMERA <= 57 :
| | | X_KOORD <= -4.15 :
| | | | STAROST <= 9.5 : LM2
| | | | STAROST > 9.5 : LM3
| | | X_KOORD > -4.15 : LM4
| | IZMERA > 57 : LM5
STAROST > 41.5 :
| AC_LJ <= 0.5 :
| | STAROST <= 67.5 : LM6
| | STAROST > 67.5 : LM7
| AC_LJ > 0.5 : LM8

Models at the leaves:

LM1: $CENA_NA_M2 = 26500 + 2920*RAZV_OBC + 3060*CAS_TRAN - 269*IZMERA - 34.9*STAROST - 40200*KLETNO - 1430*PODSTRES + 699*LEGA_V + 568*LEGA_J + 696*LEGA_Z + 141*DVIGALO - 7110*ODDALJEN + 404*OGREVAN + 36700*TELEFON + 1210*KANALIZ - 109000*AC + 1780*AC_LJ - 35900*X_KOORD + 722*Y_KOORD$

LM2: $CENA_NA_M2 = 247000 + 1630*RAZV_OBC + 1420*CAS_TRAN - 19.3*IZMERA - 532*STAROST - 576*KLETNO - 5190*PODSTRES + 2220*NADSTROP + 735*LEGA_V + 639*LEGA_J + 6170*LEGA_Z + 1710*DVIGALO - 357*ODDALJEN + 404*OGREVAN + 6000*TELEFON + 692*KANALIZ - 65100*AC + 1470*AC_LJ + 692*X_KOORD + 7230*Y_KOORD$

LM3: $CENA_NA_M2 = 123000 + 1630*RAZV_OBC + 1000*CAS_TRAN - 19.3*IZMERA - 319*STAROST - 576*KLETNO - 5190*PODSTRES + 1090*NADSTROP + 735*LEGA_V - 29200*LEGA_J + 3430*LEGA_Z + 1710*DVIGALO - 357*ODDALJEN + 404*OGREVAN + 3100*TELEFON + 692*KANALIZ - 3010*AC + 1470*AC_LJ + 692*X_KOORD + 27300*Y_KOORD$

LM4: $CENA_NA_M2 = 132000 + 1630*RAZV_OBC + 5960*CAS_TRAN - 19.3*IZMERA - 68.7*STAROST - 576*KLETNO - 80500*PODSTRES + 735*LEGA_V + 639*LEGA_J + 801*LEGA_Z + 18700*DVIGALO - 195*ODDALJEN + 404*OGREVAN + 321*TELEFON + 692*KANALIZ - 2280*AC + 1470*AC_LJ + 318*X_KOORD + 1790*Y_KOORD$

LM5: $CENA_NA_M2 = 179000 + 1630*RAZV_OBC + 5460*CAS_TRAN - 747*IZMERA - 995*STAROST - 576*KLETNO - 2350*PODSTRES + 984*LEGA_V + 862*LEGA_J + 1100*LEGA_Z + 656*DVIGALO - 136*ODDALJEN + 404*OGREVAN + 321*TELEFON + 692*KANALIZ - 2310*AC + 1780*AC_LJ + 200*X_KOORD + 11100*Y_KOORD$

LM6: $CENA_NA_M2 = -34600 + 31400*RAZV_OBC + 5020*CAS_TRAN - 1820*ST_VLOZ - 21.2*IZMERA - 44.2*STAROST - 1050*PODSTRES + 634*LEGA_V + 602*LEGA_J + 612*LEGA_Z + 496*DVIGALO + 5850*OGREVAN + 466*TELEFON + 969*KANALIZ - 1800*AC + 2870*AC_LJ - 248*X_KOORD + 464*Y_KOORD$

LM7: $CENA_NA_M2 = 37800 + 3180*RAZV_OBC + 729*CAS_TRAN - 1820*ST_VLOZ - 21.2*IZMERA - 50.7*STAROST - 1050*PODSTRES + 634*LEGA_V + 602*LEGA_J + 612*LEGA_Z + 496*DVIGALO + 7060*OGREVAN + 466*TELEFON + 969*KANALIZ - 1800*AC + 2870*AC_LJ - 7400*X_KOORD + 464*Y_KOORD$

LM8: $CENA_NA_M2 = 118000 + 4080*RAZV_OBC + 519*CAS_TRAN - 3310*ST_VLOZ - 29.3*IZMERA - 38.9*STAROST - 1050*PODSTRES + 634*LEGA_V + 66200*LEGA_J + 612*LEGA_Z + 496*DVIGALO + 4570*OGREVAN + 466*TELEFON + 969*KANALIZ - 1800*AC + 3770*AC_LJ - 248*X_KOORD + 464*Y_KOORD$

Območje transakcij 9

Scheme: weka.classifiers.m5.M5Prime -O m -F 3.0 -V 0
 Instances: 618
 Attributes: 24

Pruned training model tree (number of rules = 5):

```
STAROST <= 42.5 :
| X_KOORD <= -5.35 : LM1
| X_KOORD > -5.35 :
| | Y_KOORD <= -3.9 : LM2
| | Y_KOORD > -3.9 : LM3
STAROST > 42.5 :
| STAROST <= 48.5 : LM4
| STAROST > 48.5 : LM5
```

Models at the leaves:

LM1: $CENA_NA_M2 = 75800 - 1120*RAZV_OBC + 4970*CAS_TRAN - 2200*STAROST + 132*NADSTROP - 17400*ODDALJEN + 1700*TELEFON + 77000*KANALIZ - 2640*AC - 1580*X_KOORD + 1410*Y_KOORD$

LM2: $CENA_NA_M2 = 97900 - 1120*RAZV_OBC + 361*CAS_TRAN - 41.6*IZMERA - 130*STAROST + 507*NADSTROP - 523*ODDALJEN + 13800*TELEFON + 1410*KANALIZ - 2840*AC - 1400*X_KOORD + 1670*Y_KOORD$

LM3: $CENA_NA_M2 = 139000 - 1120*RAZV_OBC + 2880*CAS_TRAN - 425*IZMERA - 799*STAROST + 345*NADSTROP - 5860*ODDALJEN + 1440*TELEFON + 1410*KANALIZ - 1990*AC - 1110*X_KOORD + 1270*Y_KOORD$

LM4: $CENA_NA_M2 = 99800 - 4080*RAZV_OBC + 304*CAS_TRAN - 212*STAROST + 477*NADSTROP - 395*ODDALJEN + 1350*TELEFON - 1120*X_KOORD + 1500*Y_KOORD$

LM5: $CENA_NA_M2 = 73000 - 4080*RAZV_OBC + 304*CAS_TRAN - 149*STAROST + 477*NADSTROP - 395*ODDALJEN + 1350*TELEFON - 1120*X_KOORD + 1500*Y_KOORD$

Območje transakcij 10

Scheme: weka.classifiers.m5.M5Prime -O m -F 3.0 -V 0
 Instances: 622
 Attributes: 24

Pruned training model tree (number of rules = 2):

```
Y_KOORD <= -0.95 : LM1
Y_KOORD > -0.95 : LM2
```

Models at the leaves:

LM1: $CENA_NA_M2 = 122000 + 2500*CAS_TRAN - 327*IZMERA - 1240*STAROST + 94*NADSTROP + 1080*TELEFON + 1460*AC - 14700*X_KOORD + 25200*Y_KOORD$

LM2: $CENA_NA_M2 = 156000 - 22600*RAZV_OBC - 21*IZMERA - 285*STAROST + 6130*NADSTROP + 31600*TELEFON + 2770*AC - 613*X_KOORD$

Priloga 7

Opis strukture podatkov na priloženi zgoščenki

Na priloženi zgoščenki se nahaja sedem osnovnih map, ki lahko vsebujejo več podmap. Struktura map na zgoščenki je naslednja:

- *Baza_Obdelana*: prečiščena baza podatkov, ki je zaščiten z geslom.
- *Baza_Original*: originalna neprečiščena baza podatkov DURS-a, ki je zaščiten z geslom.
- *Dodatni_Podatki*: podatki o občinah, statističnih regijah in avtocestnih povezavah, s pomočjo katerih smo dopolnili originalno bazo podatkov.
- *Rezultati_Apriori*: rezultati analize prečiščene baze podatkov z različnimi nastavitvami za apriori algoritem. Pri analizi z apriori algoritmom smo spreminjali podporo (s), zaupanje (c) in število najdenih pravil (p). Ta mapa vsebuje dve podmapi:
 - *1114_V3*: povezovalna pravila za zemljišča (nepremičnine od 11 do 14).
 - *2127_V3*: povezovalna pravila za objekte (nepremičnine od 21 do 27).

Poleg dveh podmap vsebuje mapa *Rezultati_Apriori* še pet datotek. V njih se nahajajo podatki za izvajanje APRIORI analize, opis spremenljivk in diagrami povprečnih vrednosti ter median za izmere in cene nepremičnin v bazi podatkov.

- *Rezultati_Neprem_22*: zbirka vseh modelov za napoved cen kvadratnega metra stanovanj, ki so opisani v magistrskem delu. Ta mapa vsebuje tri podmape:
 - *Clustering_Cena*: rezultati združevanja v skupine glede na povprečno ceno stanovanj v posameznih občinah. Rezultati te analize so bili uporabljeni za razdelitev občin v posamezna območja transakcij.
 - *Napoved_Cen*: modeli za napoved cene kvadratnega metra stanovanja, ki so bili zgrajeni za celo Slovenijo. Ta mapa vsebuje štiri podmape:
 - *LinReg_Primerjava*: regresijski modeli, s pomočjo katerih smo izbrali oblike odvisne in neodvisnih spremenljivk.
 - *Modeli_Baza_Originalna*: modeli za napoved cen, ki so bili zgrajeni za nedopolnjeno bazo podatkov.
 - *Modeli_Baza_Razsirjena*: modeli za napoved cen, ki so bili zgrajeni za dopolnjeno bazo podatkov (glej poglavje 4.4).
 - *Podatki*: učno-testne kombinacije podatkov za izgradnjo modelov. Vsebuje razdelitev za nedopolnjeno in za dopolnjeno bazo podatkov.
 - *Napoved_Cen_ObmTr*: modeli za napoved cene kvadratnega metra stanovanja, ki so bili zgrajeni za vsako območje transakcij posebej. Ta mapa vsebuje enajst podmap:
 - *Modeli_ObmTr(i)*: modeli za napoved cene kvadratnega metra stanovanja, zgrajeni za vsako območje transakcij ($i=1,\dots,10$) posebej. S

pomočjo teh modelov smo izvedli prečno preverjanje za vsako območje transakcij posebej.

- *Podatki*: podatki za posamezna območja transakcij. Ta podmapa vsebuje tudi modele za napoved cene kvadratnega metra stanovanja, ki so bili zgrajeni na osnovi vseh transakcij v določenem območju.

Poleg enajstih podmap vsebuje mapa *Napoved_Cen_ObmTr* še pet datotek. Datoteki *ObmTr_CasovniVpliv_Podatki.sav* in *ObmTr_CasovniVpliv_Rezultati.spo* sta SPSS podatkovna in tekstovna datoteka. Vsebudeta podatke in rezultate analize časovnega vpliva na cene stanovanj. Ostale tri datoteke vsebujejo primerjave različnih modelov za napoved cen, ki so bili zgrajeni za posamezna območja transakcij.

- *Rezultati_OdstVnos*: rezultati analize manjkajočih vnosov za posamezne občine.
- *XLS_Makroji*: makroji za obdelavo podatkov v programu Microsoft EXCEL. Če želi uporabnik direktno uporabiti priložene makroje, mora v njih popraviti absolutne poti do datotek, ki ustrezajo njegovi strukturi map. Absolutne poti do datotek, ki so napisane v makrojih, namreč veljajo le za računalnik, ki je bil uporabljen za izračun na zgoščenki predstavljenih rezultatov.

OPOMBA:

Baza podatkov DURS-a o opravljenih nepremičninskih transakcijah v Republiki Sloveniji je bila dana na voljo avtorju tega magistrskega dela le za raziskovalne namene. Zato je na priloženi zgoščenki zaščitena z geslom. Dostop do te baze podatkov je mogoč le v soglasju z mentorjem prof. dr. Ivom Lavračem.