

UNIVERZA V LJUBLJANI
EKONOMSKA FAKULTETA

MAGISTRSKO DELO

**PRIMER INFORMACIJSKEGA SISTEMA ZA
MODELIRANJE FUNKCIONALNIH REGIJ V SLOVENIJI**

LJUBLJANA, OKTOBER 2010

LIDIJA BAJT

IZJAVA

Študentka Lidiya Bajt izjavljam, da sem avtorica tega magistrskega dela, ki sem ga napisala v soglasju s svetovalcem prof. dr. Ivom Lavračem, in da v skladu s 1. odstavkom 21. člena Zakona o avtorskih in sorodnih pravicah dovolim njegovo objavo na fakultetnih spletnih straneh.

V Ljubljani, dne 15.10.2010

Podpis: _____

KAZALO

UVOD	1
1 OPREDELITEV FUNKCIONALNIH REGIJ	3
1.1 Metodologije in pristopi pri zamejitvi funkcionalnih regij	4
1.1.1 Dvostopenjski model z opredelitvijo središča	5
1.1.2 Razdelitev na samozadostne regije.....	5
1.2 Velikost in združljivost administrativnih in funkcionalnih regij	6
1.3 Problem spremenljive prostorske enote.....	7
1.4 Uporaba in uporabnost opredelitve funkcionalnih regij	8
1.5 Administrativne in funkcionalne regije v Sloveniji.....	9
1.5.1 Projekt RePUS.....	9
1.5.2 ESPON 1.1.1 in ESPON 1.4.3	11
1.6 Opredelitev funkcionalnih regij z metodo Intramax	11
1.6.1 Metoda Intramax.....	11
1.6.2 Intramax analiza z uporabo programa Flowmap.....	13
1.6.3 Opredelitev funkcionalnih regij za Slovenijo z uporabo metode Intramax	14
2 ANALIZE IN MODELI PROSTORSKIH INTERAKCIJ	17
2.1 Izgradnja prostorskega modela	18
2.1.1 Namen modela.....	18
2.1.2 Opredelitev območja proučevanja in izbira sistema prostorskih enot	18
2.1.3 Podatki.....	19
2.1.3.1 Značilnosti prostorskih enot	19
2.1.3.2 Tokovi	20
2.1.3.3 Razdalje.....	21
2.2 Analiza dostopnosti	22
2.2.1 Kazalniki dostopnosti	22
2.2.1.1 Potencialna dostopnost	23
2.2.1.2 Multimodalna analiza dostopnosti.....	23
2.3 Modeli prostorskih interakcij	24
2.3.1 Gravitacijski model nakupovalnih središč	24
2.3.2 Gravitacijski model brez omejitev.....	25
2.3.3 Gravitacijski model z enojno omejitvijo	25
2.3.4 Garin-Lowryjev gravitacijski model.....	26
2.3.5 Razširjeni gravitacijski modeli	28
2.4 Regresijski modeli	28
2.4.1 Prostorski regresijski modeli.....	29
2.4.2 Geografsko tehtana regresija	30
2.5 Optimizacijski modeli	32
2.6 Scenariji	33
3 INFORMACIJSKI SISTEM ZA MODELIRANJE FUNKCIONALNIH REGIJ	34
3.1 SISTEMI ZA PODPORO ODLOČANJU	34
3.1.1 Funkcionalne zahteve sistema za podporo odločanju	35
3.1.2 Oris informacijske rešitve.....	36

3.2	Sistem za upravljanje podatkovnih baz	37
3.2.1	<i>Prostorske podatkovne baze</i>	37
3.2.2	<i>Upravljanje in analiza prostorskih podatkov</i>	38
3.2.3	<i>Shranjevanje prostorskih podatkov v bazi.....</i>	39
3.2.4	<i>Vizualizacija prostorskih podatkov in Oracle MapViewer.....</i>	39
3.2.5	<i>Načrtovanje in izgradnja podatkovne baze</i>	40
3.3	Podatkovno rudarjenje	42
3.3.1	<i>Podatkovno rudarjenje in odkrivanje znanja iz podatkov</i>	43
3.3.2	<i>Koncepti in metode podatkovnega rudarjenja.....</i>	45
3.3.2.1	<i>Opis podatkov in združevanje</i>	46
3.3.2.2	<i>Razvrščanje v skupine</i>	47
3.3.2.3	<i>Opis konceptov</i>	47
3.3.2.4	<i>Klasifikacija.....</i>	48
3.3.2.5	<i>Napovedovanje.....</i>	49
3.3.2.6	<i>Analiza odvisnosti</i>	49
3.3.3	<i>Proces podatkovnega rudarjenja</i>	49
3.3.4	<i>Oracle Data Mining</i>	51
4	PRIMER UPORABE INFORMACIJSKEGA SISTEMA	54
4.1	Podatki	54
4.2	Razširjeni gravitacijski model za napovedovanje dnevnih migracij med slovenskimi občinami	55
4.2.1	<i>Razumevanje vsebine in uporabljena metoda</i>	55
4.2.2	<i>Razumevanje in priprava podatkov</i>	56
4.2.3	<i>Modeliranje in vrednotenje</i>	59
4.2.3.1	<i>Ocena potenčne funkcije z metodo posplošenih najmanjših kvadratov</i>	59
4.2.3.2	<i>Ocena modela z metodo podpornih vektorjev</i>	63
4.2.4	<i>Uporaba.....</i>	64
4.3	Uporaba metode k-Means za razvrstitev občin v funkcionalne regije.....	66
4.3.1	<i>Razumevanje vsebine.....</i>	66
4.3.2	<i>Razumevanje podatkov</i>	66
4.3.3	<i>Priprava podatkov</i>	67
4.3.4	<i>Modeliranje</i>	68
4.3.5	<i>Uporaba in vrednotenje.....</i>	70
SKLEP.....		72
LITERATURA IN VIRI.....		75

KAZALO SLIK

Slika 1: Vpliv različnih razdelitev območij na zaznavanje pojava.....	7
Slika 2: Razporeditev urbanih središč nacionalnega, regionalnega in medobčinskega pomena v lokalnih zaposlitvenih sistemih glede na število prebivalcev	10
Slika 3: Izhodna datoteka analize Intramax.....	15
Slika 4: Dendrogram združevanja občin.....	15
Slika 5: Razmejitev občin po statističnih regijah (SURs) in 12 funkcionalnih regijah opredeljenih z metodo Intramax	16
Slika 6: Razmejitev občin na osem funkcionalnih regij opredeljenih z metodo Intramax	16
Slika 7: Razmejitev občin na 20 funkcionalnih regij opredeljenih z metodo Intramax	17
Slika 8: Sistem za podporo odločanju in arhitektura informacijske rešitve	36
Slika 9: E-R diagram	41
Slika 10: Koraki odkrivanja znanja v podatkovni bazi	45
Slika 11: Koncepti in metode podatkovnega rudarjenja.....	46
Slika 12: Proces podatkovnega rudarjenja	50
Slika 13: Oracle Data Miner – interaktivni grafični vmesnik	52
Slika 14: Krog podatkov.....	54
Slika 15: Prikaz podatkov za izdelavo modela v aplikaciji Oracle Data Miner	56
Slika 16: Razvrstitev atributov po napovedovalni moči.....	57
Slika 17: Razsevni diagram ostankov linearnega regresijskega modela	58
Slika 18: Razsevni diagram ostankov potenčnega regresijskega modela.....	59
Slika 19: Splošna statistika potenčnega regresijskega modela.....	60
Slika 20: Ocene parametrov potenčne regresijske funkcije.....	60
Slika 21: Napovedna zanesljivost potenčnega modela.....	61
Slika 22: Testna statistika potenčnega modela	62
Slika 23: Razsevni diagram ostankov potenčnega modela.....	62
Slika 24: Napovedna zanesljivost SVM modela	63
Slika 25: Testna statistika SVM modela	63
Slika 26: Razsevni diagram ostankov SVM modela	64
Slika 27: Napoved regresijskega modela, ocenjenega z metodo podpornih vektorjev	65
Slika 28: Opisna analiza toka dnevnih migrantov za posamezno občino.....	67
Slika 29: Vhodna tabela za razvrščanje v skupine	68
Slika 30: Model za razvrščanje skupin s k-Means metodo	69
Slika 31: Rezultat razvrščanja v skupine z metodo k-Means	70
Slika 32: Prikaz rezultata razvrščanja v skupine z metodo k-Means za 12 regij.....	71
Slika 33: Prikaz rezultata razvrščanja v skupine z metodo k-Means za 8 regij.....	71

KAZALO TABEL

Tabela 1: Matrika tokov z j regijami	12
Tabela 2: Metode in algoritmi, ki jih podpira Oracle Data Mining.....	53

UVOD

Problematika magistrskega dela

Razvoj Slovenije bo v prihodnosti odvisen od njene konkurenčnosti, ki je odvisna tudi od razvoja regij, njihove povezanosti in odprtosti za čezmejno sodelovanje. Zmanjšanje razvojnih razlik in aktiviranje notranjih potencialov države torej ni le politično, temveč povsem realno razvojno vprašanje konkurenčnosti države kot celote. Zato je treba preseči razprave o številu regij in jo nadgraditi z razpravo o učinkovitem razvoju Slovenije kot celote, pri katerem bodo gonilo razvoja regije, opredeljene na podlagi funkcionalnih značilnosti (Pogačnik et al., 2008).

Funkcionalne regije so regije, ki jih zaznamuje visoka frekvenca notranjih regionalnih gospodarskih interakcij, kot so regionalna trgovina dobrin in storitev, delovne migracije ter nakupovanje. Funkcionalno regijo zaznamuje tudi gostitev aktivnosti in notranja prometna infrastruktura, ki omogoča interakcije. Glavna značilnost funkcionalne regije je povezan trg dela, na katerem so dnevne migracije na delo, iskanje zaposlitev in delavcev bolj intenzivne kot med regijami, zato je lokalni trg dela dober približek področja funkcionalne regije (Johansson, 1998).

S pojavom globalnega gospodarstva, ki temelji na znanju, imajo funkcionalne regije vedno bolj vlogo neodvisnih, dinamičnih trgov, ki so z drugimi funkcionalnimi regijami povezane s tokovi informacij, znanja ter dobrin in storitev. Vsaka funkcionalna regija ima svojo bazo znanstvenega, tehnološkega in podjetniškega znanja v obliki znanja v podjetjih in drugih organizacijah, ki delujejo v regiji, ter v človeškem in družbenem kapitalu, ki je povezan s prebivalstvom regije. Funkcionalno regijo zaznamuje tudi njen izobraževalni sistem, nenehno pridobivanje novega znanja na univerzah, v raziskovalnih laboratorijih in v podjetjih, ter uvoz in izvoz znanja, vsi ti vidiki znanja pa so opredeljeni kot sestavni deli regionalnega inovacijskega sistema (Andersson, Grasjö & Karlsson, 2006).

Meje funkcionalne regije so torej meje regije, ki jo zaznamuje visoka frekvenca notranjih regionalnih gospodarskih interakcij. Funkcionalna regija je tako opredeljena z ekonomskimi in manj z administrativnimi lastnostmi in zato bolj primerna za primerjave in ekonomske analize. Ker je pomembna interakcija znotraj regij in med njimi, so za modeliranje najbolj primerni prostorski interakcijski modeli, ki poleg dejavnosti upoštevajo tudi tokove in razdalje med regijami (Foot, 1981). Z ocenjenimi modeli prostorskih interakcij lahko na podlagi različnih scenarijev razvoja napovedujemo, kakšne bodo interakcije v prihodnosti, in na podlagi tega, kakšne funkcionalne regije se bodo oblikovale. Z analizo dostopnosti lahko proučujemo, kakšna je dostopnost prebivalstva do določenih storitev in kakšna je optimalna razporeditev zmogljivosti, ki bo omogočala čim večjo konkurenčnost ob upoštevanju trajnostnega razvoja regij.

Namen in cilji magistrskega dela

Namen magistrskega dela je na podlagi poznavanja zahtev za modeliranje funkcionalnih regij oblikovati informacijski sistem za podporo odločanju, ki bo osnova za modeliranje funkcionalnih regij v Sloveniji. Osnova informacijskega sistema je prostorska podatkovna baza, ki omogoča integracijo z GIS orodji in v kateri bodo shranjeni podatki, ki so potrebni za modeliranje, modeli in ocene parametrov ter rezultati modeliranja. Modeliranje se izvaja s procesom podatkovnega rudarjenja neposredno na podatkovni bazi.

Z modeliranjem pridobljeni rezultati bodo izhodišče za analizo, kakšne bi bile funkcionalne regije v prihodnosti ob izpolnitvi določenih scenarijev. V magistrskem delu želim z uporabo informacijskega sistema na podatkih za Slovenijo preveriti hipotezo, ali se lahko tako oblikovan informacijski sistem uporabi za modeliranje funkcionalnih regij.

Cilj magistrskega dela je zasnova informacijskega sistema za podporo odločanju pri oblikovanju takšnih razvojnih programov, ki bodo omogočali večjo konkurenčnost, kohezijo in trajnostni razvoj regij in s tem države kot celote. Informacijski sistem bo informacijska podpora pri modeliranju regij na podlagi različnih scenarijev, katerega rezultati bodo podlaga za oblikovanje funkcionalnih regij. Za informacijski sistem se izdelava funkcionalna specifikacija zahtev, ki opisuje, kako naj bi informacijski sistem deloval, kako naj bi ga uporabniki uporabljali ter kakšni naj bi bili vhodni in izhodni podatki. Na podlagi zahtev so izbrane funkcionalnosti in orodja, s katerimi bo izgrajen informacijski sistem.

S tako zasnovanim informacijskim sistemom bom poskušala izvesti analizo na dejanskih podatkih in prikazati oblikovanje funkcionalnih regij na podlagi rezultatov modeliranja za možen scenarij razvoja. Pri tem bom uporabila dostopne podatke o tokovih (interakcijah) med slovenskimi občinami, dejavnostih in razdalje med občinami.

Metode dela

Magistrsko delo v prvem delu vsebuje teoretično-analitičen pregled strokovne literature, znanstvenih razprav in raziskav ter člankov o oblikovanju funkcionalnih regij, modelih prostorskih interakcij in oblikovanja scenarijev. Pregledane so tudi dosedanje raziskave za Slovenijo in izdelana analiza razvrščanja enot v skupine z metodo Intramax na podlagi dnevnih migracij delavcev na delo med slovenskimi občinami.

V drugem delu je s pomočjo oblikovanja funkcionalne specifikacije zahtev izdelana zasnova informacijskega sistema. Na podlagi zahtev je izdelan podatkovni model in diagram entitet in povezav (E-R model), ki prikazuje strukturo podatkov v podatkovni bazi. Opisane so metode podatkovnega rudarjenja, tj. procesa, ki se uporabi za modeliranje funkcionalnih regij. Na podlagi tako oblikovanega informacijskega sistema je izvedena analiza na podatkih o slovenskih regijah, s katero tudi preverim ustreznost tako zasnovanega informacijskega sistema.

Pri izdelavi magistrskega dela bom uporabila znanje, ki sem ga pridobila med študijem in znanje, ki sem ga pridobila s praktičnimi izkušnjami pri delu tehnične in funkcionalne svetovalke v poslovni informatiki.

Vsebinska zasnova magistrskega dela

Magistrsko delo je sestavljeno iz štirih poglavij, tematika pa je dodatno razdelana v podpoglavjih. V prvem poglavju je predstavljeno, kaj so funkcionalne regije, kako se razmejujejo, kako jih različne države opredeljujejo in uporabljajo v praksi ter kako so funkcionalne regije opredeljene v Sloveniji. V tem poglavju sem za Slovenijo tudi razvrstila regije v skupine z metodo Intramax na podlagi dnevnih migracij delavcev na delo.

V drugem poglavju so predstavljene analize in modeli prostorskih interakcij, kakšna je ustrezna razmejitev proučevanega območja na cone, podatki in njihova dostopnost za izvedbo analiz. V tem poglavju so na kratko predstavljeni tudi scenariji, na podlagi katerih se izvaja modeliranje.

Tretje poglavje vsebuje funkcionalno specifikacijo zahtev za izdelavo informacijskega sistema za modeliranje funkcionalnih regij. Funkcionalna specifikacija temelji na zahtevah uporabnikov in vsebuje izdelan podatkovni model, ki je uporabljen za hrambo podatkov v prostorski podatkovni bazi, in opis podatkovnega rudarjenja kot procesa ter metod podatkovnega rudarjenja za modeliranje.

V četrtem poglavju je izvedeno modeliranje na podlagi dejanskih podatkov za Slovenijo na prej zasnovanem informacijskem sistemu. Za modeliranje je uporabljena metoda napovedovanja, in sicer napovedovanje dnevnih migracij med slovenskimi občinami z uporabo razširjenega gravitacijskega modela. Na podlagi dnevnih migracij so nato s podatkovnim rudarjenjem občine razvrščene v funkcionalne regije z uporabo metode razvrščanja v skupine.

1 OPREDELITEV FUNKCIONALNIH REGIJ

Funkcionalne regije zaznamuje visoka frekvenca notranjih regionalnih gospodarskih interakcij, kot so regionalna trgovina dobrin in storitev, delovne migracije ter nakupovanje. Funkcionalno regijo zaznamuje tudi gostitev aktivnosti in notranja prometna infrastruktura, ki omogoča interakcije (Johansson, 1998).

Podobno opredelitev poda tudi Castells (1999), ki pa se nanaša na metropolitanske regije. Opredeli jih kot območja brez jasne prostorske hierarhije, s šibko politično in upravno organiziranostjo, vendar z razmeroma močno funkcionalno povezanostjo, skupnim tržiščem, trgom delovne sile, medijskim trgom in skupno infrastrukturo.

Obstaja veliko gospodarskih interakcij, kot so dnevne migracije prebivalcev, selitve, trgovina s storitvami in dobrinami, komunikacije, prometni tokovi, blagovni tokovi, storitvene povezave, cirkulacija časopisov in finančni tokovi. Kljub temu se za opredelitev funkcionalnih regij najpogosteje uporablja podatke o dnevni migraciji na delo, zlasti zato, ker praksa kaže, da je vzorec dnevni migraciji na delo dober približek za opredelitev funkcionalne regije. Funkcionalne regije so tisto, kar večina ljudi pojmuje kot njihove regije in so območja, na katerem podjetja najdejo večino svojih zaposlenih. Kakovost funkcionalnih regij ima pomemben vpliv na produktivnost in blaginjo. Dostop do želenih zaposlitev za prebivalstvo in podjetij do izobraženih delavcev ter do znanja v drugih podjetjih, je eden najpomembnejših elementov konkurenčnosti regije.

1.1 Metodologije in pristopi pri zamejitvi funkcionalnih regij

Da bi lahko ocenjevali uspešnost regij in jih primerjali med seboj, tudi na mednarodni ravni, je treba najti neko skupno opredelitev funkcionalnih regij. Pomanjkanje enotne opredelitve predstavlja veliko oviro pri analizi konkurenčnosti regij.

V grobem obstajajo trije različni pristopi pri opredelitvi regij:

- Administrativni pristop opredeljuje regije glede na administrativne meje občin ali drugih ustreznih enot, včasih tudi skupino občin pod določeno regionalno (ali pokrajinsko) upravo. Ta pristop uporablja predvsem državna uprava za strukturiranje, organizacijo in upravljanje države, na primer za analizo fiskalne politike.
- Morfološki pristop upošteva velikost in/ali sklenjenost naselja, število prebivalcev, delež poseljenega občinskega območja. Ta pristop, ki se navadno uporablja neodvisno od politično določenih mej, je učinkovit predvsem pri opredelitvi grafične predstavitve mest in naselij, še posebej iz satelitskih ali zračnih posnetkov. Morfološki pristop je tudi primernejši za analizo razdelitve v cone, okoljskih problemov in prihodnje stanovanjske politike. Ne upošteva pa ljudi, ki živijo zunaj mesta, a so od njega odvisni zaradi svojega dela, dostopnosti javnih storitev in prostočasni dejavnosti.
- Funkcionalni pristop opredeljuje regije glede na interakcije med jedrom – ki se lahko opredeli po morfoloških ali administrativnih kriterijih – in zajetjem okrog jedra. Najpogosteje se z analizo dnevni migraciji na delo določi, katera področja se vključijo v določeno regijo.

Interakcija z naseljenim jedrom, ki jo opredeljuje funkcionalna opredelitev, je primerna za analizo socialno-ekonomskih problemov, kot so infrastruktura in transport, zgostitve prometa, za analizo trga dela in povezave med podjetji.

Znotraj funkcionalnega pristopa obstajajo različne metode, kako opredeliti funkcionalno regijo, najpogosteje pa jih lahko razvrstimo v dve metodi: dvostopenjski model in razdelitveno metodo.

1.1.1 Dvostopenjski model z opredelitvijo središča

V dvostopenjskem modelu se regija oblikuje kot območje, ki je sestavljeno iz središča z veliko populacijo in sosednjih občin, za katere je značilna visoka stopnja integracije s središčem. Stopnja integracije se navadno meri z dnevnimi migracijami na delo med sosednjimi občinami in središčem. Prvi korak modela je torej izbira središča, tj. mestno središče ali jedro, ki se ga lahko opredeli po administrativnih, morfoloških, funkcionalnih kriterijih ali kombinaciji teh kriterijev. V drugem koraku se regija razširi na sosednja območja, ki imajo veliko stopnjo integracije s središčem. Velikost tega območja je odvisna od opredelitve. Stopnja integracije med sosednjimi območji se navadno meri z deležem prebivalstva, ki je zaposleno v središču. Če je delež nad neko določeno mejo, potem je območje del regije.

Metoda je zelo primerna za monocentrične strukture, ki jih zaznamujejo jedro in sosednja območja vpliva, ni pa primerna za opredelitev policentričnih regij, saj predpostavlja eno središče. Zaradi različnih kriterijev pri opredelitvi središča (Eurostat in ESPON sta opredelila središča z administrativnim kriterijem glede na spodnjo mejo števila prebivalstva, GEMACA raziskava pa je opredelila središča po morfološkem kriteriju (šest delovnih mest na hektar)), tokov (nekateri opredeljujejo zaledje na podlagi enosmernih, drugi dvosmernih povezav) in mej za vključitev pripadajočega območja – zaledja, območja po velikosti variirajo in so pogosto med sabo neprimerljiva.

1.1.2 Razdelitev na samozadostne regije

Razdelitvena metoda območje proučevanja (navadno državo) razdeli na izključujoče in obširne regionalne enote, ki so dovolj velike, da upravičijo izločitev; so samozadostne, tj. relativno avtonomne glede na interne tokove.

Za razliko od dvostopenjske metode razdelitvena metoda ne opredeli središča, temveč razdeli območje na samozadostne ekonomske regije na podlagi interakcij oziroma tokov med sosednjimi območji in nato določi, kakšne so te regije (tipologijo), ključen pa je kriterij za samozadostnost.

Ta metoda uporablja bolj sofisticirana orodja in se je pojavila z razvojem računalnikov, saj je bilo s temi mogoče izmeriti kompleksne interakcije v obliki tokov, tako v smislu fizične mobilnosti (migracije na delo, v šolo ipd.) in interakcij na podlagi sodobne tehnologije (npr. dela na daljavo). V praksi je kriterij, ki določa avtonomijo regije, tudi zaradi dostopnosti podatkov, najpogostejše osnovan na dnevnih tokovih na delo. Tako oblikovane regije so samozadostne in opredeljene kot območja trga dela. Če podatki o dnevnih migracijah niso dostopni, se lahko ocenijo na podlagi mikropodatkov o zaposlenih in zaposlitvah po metodi zajetja (angl. *catchment areas*), ki se uporabijo kot okvirne regije (Trapasso, 2007).

Večina algoritmov za regionalizacijo temelji na metodah razvrščanja enot v skupine, pri čemer se v homogene regije razvrščajo najmanjše enote, za katere obstajajo podatki. Prvi algoritmi so temeljili na nehierarhičnih metodah, kasneje hierarhičnih, s predhodno opredelitvijo števila skupin ali brez. Nekatere metode pri razdelitvi enot upoštevajo geografske informacije in druge podatke, da zagotovijo sosednost vključenih enot. V zadnjem času je prišlo tudi do razvoja optimizacijskih modelov, ki poskušajo agregirati enote ob upoštevanju problema spremenljivih prostorskih enot (Duque & Ramos, 2003).

Prednost razdelitvene metode je v tem, da vključuje fleksibilnost glede območij, ki sestavljajo regijo in ni odvisna od vnaprej opredeljenih središč. Zato je tudi bolj primerna za primerjavo: je prenosljiva in uporabna med državami z različnimi urbanimi sistemi, tokovi dnevnih migracij, podatki in območji (conami). Bolj ustrezna je za opredelitev regij s policentričnimi strukturami, saj dovoljuje enako pojavnost monocentričnih in policentričnih regij. Obstaja pa nevarnost, da se velika območja s preveč središči združijo v eno območje, še posebej če gre za krožne ali zaporedne migracije med verigo mest in če veliko dnevnih migracij poteka v vse smeri na gosto poseljenem območju. Glavna slabost te metode je bila odvisnost od računalniških algoritmov za razdelitev območja v samozadostne ekonomske regije, kar pa danes ni več problem. Za mednarodne primerjave je potrebno opredeliti tudi enotno mejo za samozadostnost (Trapasso, 2007).

1.2 Velikost in združljivost administrativnih in funkcionalnih regij

Združljivost funkcionalnih in administrativnih regij je pomembno vprašanje, saj odraža, kako se različne prostorske enote združujejo ter kako se delijo naloge in pristojnosti med njimi.

V večini držav so funkcionalne regije opredeljene tako, da so sestavljene iz celotnih občin in so tako združljive z osnovnimi teritorialnimi enotami, za katere se vodijo podatki. S tem so tudi socialno-ekonomske statistike in kazalniki dosegljivi na ravni funkcionalnih regij (OECD, 2002).

Ker je razmejitev funkcionalnih regij navadno vezana na dnevne migracije med občinami, te niso vedno združljive z administrativnimi regijami na višji ravni in v veliko državah se razmejitve ne prilagajajo tako, da bi funkcionalne regije ustrezale administrativnim (OECD, 2002). Čeprav ni enotnega mnenja, ali je uskladitev potrebna, pa vseeno olajša prostorske analize. Uskladitev omogoča bolj natančno analizo regionalnih razlik in lahko razkrije strukturne spremembe, ki jih prevelike regije in povprečne ocene kazalnikov zameglijo. Kljub temu je uporaba funkcionalnih regij za načrtovanje politike smiselna le, če taka razmejitev občutno prekaša administrativne razmejitve v smislu pomembnih kazalnikov povezanosti trga dela ali regionalnih razlik (Cörvers, Hensen & Bongaerts, 2009).

Povprečna velikost funkcionalnih regij glede na površino in prebivalstvo variira v OECD državah vsaj toliko kot povprečna velikost administrativnih regij. Velikost regij je pomembna za mednarodne primerjave, vendar je v primeru funkcionalnih regij ustrezna, ker so

funkcionalne regije v večini držav opredeljene po enakih principih, in tako odvisna od mobilnosti delavcev in velikosti trga dela.

Velikost funkcionalnih regij se spreminja v času, saj je odvisna od dejavnikov, ki vplivajo na interakcije, kot so razvoj prometne infrastrukture in komunikacij, poselitev in v primeru tokov dnevnih migrantov velikost lokalnega trga dela in pripravljenost delavcev za potovanja na delo. Spremembe v organizaciji dela, ki ponuja več fleksibilnosti v delovnem času, in možnost dela na daljavo sta nadaljnja dejavnika, ki večata funkcije in velikost regij, opredeljenih na podlagi lokalnega trga dela.

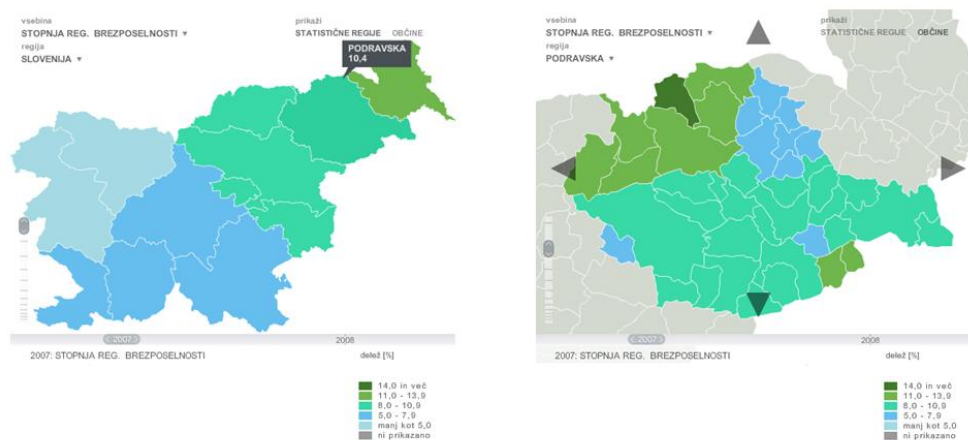
1.3 Problem spremenljive prostorske enote

Problem spremenljive prostorske enote je povezan z izborom oziroma določitvijo prostorske enote proučevanja in vpliva na rezultate prostorskih raziskav. Pojavi se, kadar analiziramo agregirane prostorske podatke, tj. podatke, ki se namesto na posameznike nanašajo na območja. Območja, na primer regije, so pogosto arbitrarno določene in drugačna agregacija območij na nižji ravni, na primer občin, bi lahko enako smiselno prikazala pojav, zato je potrebna pozorna izbira, da se prostorska variabilnost prikaže na smiseln način (Openshaw, 1984).

Problem spremenljivih prostorskih enot se izraža na dva načina:

- občutljivost rezultatov na raven in na število prostorskih enot, na katere je območje proučevanja razdeljeno (veliko variabilnosti med občinami se na primer izgubi, ko so podatki občin agregirani na ravni regij),
- občutljivost na način združevanja območij na prostorske enote tudi, če ravni ne spreminjamo (na primer združitev občin v regije administrativno ali po funkcionalnih kriterijih).

Slika 1: Vpliv različnih razdelitev območij na zaznavanje pojava



Vir: Interaktivni statistični atlas, Statistični urad Republike Slovenije, 2009.

Problem izbora prostorske enote je v tesni zvezi s t. i. ekološko napako (angl. *ecological fallacy*), ki pomeni nepravilno sklepanje o posameznikih na podlagi rezultatov analize agregiranih podatkov.

Na žalost lahko v zvezi s tem problemom storimo le malo – razen tega, da se zavedamo njegovega možnega obstoja in da proučimo občutljivost rezultatov na spreminjanje merila in morda (če podatki to omogočajo) spremenimo način delitve območja proučevanja na prostorske enote (ESPON 3.4.3).

Funkcionalne regije naj bi predstavljale gospodarski prostor, ki je hkrati samovzdržen in homogen in zato bolj primeren za strukturne analize (Trapasso). Tako je bila v projektu ESPON 3.4.3 kot ena od metodologij za zmanjšanje problema spremenljive prostorske enote predlagana in testirana razdelitev na funkcionalna urbana območja (FUA), saj naj bi se znotraj FUA prerazdelila večina BDP in dohodka. Pri raziskavi so upoštevali nedeljivost funkcionalnih območij in omejili velikost FUA glede najmanjšega števila prebivalcev ter upoštevali homogenost območij (v smislu BDP na prebivalca).

Rezultati so pokazali, da bi imeli s tako razdelitvijo manj regij na ravni NUTS 3, torej bi bile regije večje. Porazdelitev povprečnih vrednosti med državami se je zmanjšala, kar pomeni večjo mednarodno homogenost. Znotraj držav so se razlike povečale zaradi združitve centralnih in obrobni območij velikih funkcionalnih urbanih območij v enotna območja. Primerjava obstoječe razdelitve z novo razdelitvijo glede BDP na prebivalca na NUTS 2 ravni je pokazala, da so zelo visoke vrednosti, ki so bile posledica ozke razmejitve centralnih območij izginile; primerjava z obstoječimi NUTS 3 enotami pa je pokazala, da so se močno znižale najvišje vrednosti in standardni odkloni (ESPON 3.4.3.).

1.4 Uporaba in uporabnost opredelitve funkcionalnih regij

Razmejitev na podlagi ekonomskega delovanja osvetljuje ključna vprašanja uspešnosti regij, vključno z razlikami v brezposelnosti, vplivom industrije, prometne infrastrukture in sprememb v javnih izdatkih na lokalne trge dela. Poznavanje interakcij med sosednjimi regijami tudi olajša oceno primernosti trenutnih administrativnih razmejitev, ki jih uporablja statistika.

Ekonomske razlike znotraj administrativno določenih regij so lahko tako velike, da primerjava ni smiselna. Odločitve glede porazdelitve sredstev med regijami niso tako učinkovite in smiselne, kot če bi upoštevali dejansko stanje. Področja z visoko brezposelnostjo, ki so uvrščena v regije z nizko povprečno stopnjo nezaposlenosti, na primer ne morejo pridobiti pomoči nacionalnih vlad ali Evropske unije (Cörvers et al., 2009).

Zato so funkcionalne regije v veliko državah podlaga za družbeno-ekonomske analize, strukturne raziskave lokalnih trgov dela in ocenjevanje regionalnih razlik ter okvir za

izvajanje določenih politik, ki so vezane na trg delovne sile in promet. Pogosto so osnova za določanje območij, ki lahko zaprosijo za pomoč in podporo na nacionalnem ali evropskem nivoju (OECD, 2002).

1.5 Administrativne in funkcionalne regije v Sloveniji

Dokler ne bodo ustanovljene pokrajine, se v Sloveniji se za podporo regionalnemu razvoju, pri strokovnem načrtovanju in merjenju učinkov regionalne politike ter za družbenogospodarske analize uporabljajo t. i. statistične regije. Členitev na 12 funkcionalnih regij oziroma 12 območij medobčinskega sodelovanja je nastala v sedemdesetih letih prejšnjega stoletja in je bila izdelana na podlagi analize gravitacijskih območij (zaposlitev, oskrba, šolanje) (Statistični urad Republike Slovenije, 2009).

Ustanavljanje pokrajin je pomemben projekt, ki bo decentraliziral Slovenijo in pospešil njen enakomernejši razvoj, zato morajo biti območja pokrajin oblikovana tako, da bodo zagotavljala razvojno, socialno in ekonomsko močne pokrajine, ki bodo zagotavljale trajnostni razvoj svojega območja in države. Pri pripravi strokovnih podlag za uvedbo pokrajin pa se bo poleg teritorialnih, organizacijskih, finančnih komponent upoštevalo tudi funkcionalne (Svet za decentralizacijo in regionalizacijo, 2009).

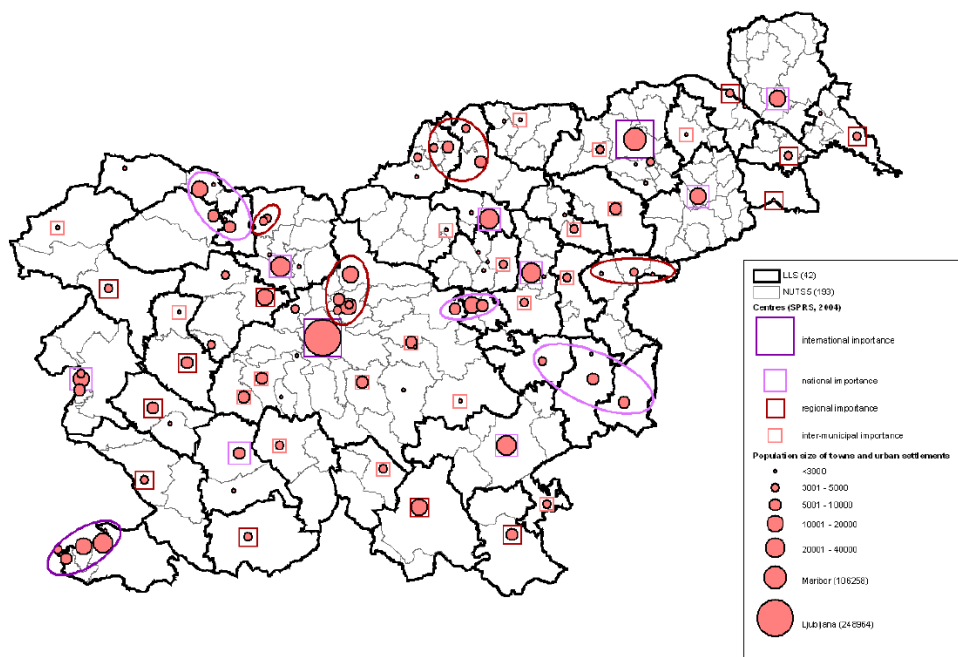
V preteklih letih so bile v okviru evropskih projektov RePUS, ESPON in v nekaterih slovenskih projektih že opredeljene funkcionalne regije za Slovenijo, predvsem po dvostopenjskem modelu zamejitve okoli središča.

1.5.1 Projekt RePUS

Strategija regionalnega policentričnega urbanega sistema v srednjem in vzhodno evropskem gospodarskem integracijskem območju, 2007, v nadaljevanju projekt RePUS, je opredelil širša mestna območja tj. lokalne zaposlitvene sisteme in funkcionalne urbane regije (angl. *local labour systems* (LLS) / *functional urban areas* (FUA)) velikih, srednje velikih in malih mest na podlagi dnevnih delovnih migracij ter indikatorjev in kriterijev razvoja na podlagi katerih so podatki analizirani in razvrščeni po tipologijah LLS/FUA vseh državah partnericah v okviru projekta RePUS.

V Sloveniji je zbiranje in analiza podatkov ter kartografski prikaz kazalnikov potekalo na prostorski ravni NUTS 5 (193 občin), NUTS 4 (58 upravnih enot), NUTS 3 (12 statističnih regij), LLS (42 mikroregije) in RLS (17 mezuregij) v Sloveniji. Primerljiva analiza podatkov in razvrščanje širših mestnih območij je potekala predvsem na ravni LLS/FUA v vseh državah partnericah v okviru projekta RePUS. Na podlagi različnih metod razvrščanja je bila opredeljena tipologija širših mestnih ter vloga in položaj velikih, srednje velikih in malih mest v (nad)nacionalnem urbanem sistemu v srednji Evropi (REPUS, 2007).

Slika 2: Razporeditev urbanih središč nacionalnega, regionalnega in medobčinskega pomena v lokalnih zaposlitvenih sistemih glede na število prebivalcev



Department of Geography, Faculty of Arts, University of Ljubljana, 2006

Vir: REPUS, Final report, 2007, str.45.

1.5.2 ESPON 1.1.1 in ESPON 1.4.3

S projektom ESPON 1.1.1 (Potenciali za policentrični razvoj v Evropi) je bilo kot funkcionalna urbana območja v Sloveniji opredeljenih šest območij. Na podlagi sedmih kriterijev (prebivalstvo, promet, turizem, industrija, znanje, odločanje in upravljanje) so bile funkcionalne urbane regije razdeljene v tri kategorije: Ljubljana je kot edino urbano območje v Sloveniji in tudi v njeni bližnji okolici prepoznana kot evropsko metropolitansko območje rasti (MEGA). Izmed ostalih slovenskih funkcionalnih urbanih območij je Koper opredeljen kot transnacionalno/nacionalno funkcionalno urbano območje, Nova Gorica, Celje, Maribor in Novo mesto pa kot regionalna/lokalna funkcionalna urbana območja.

Projekt je za analizo podatkov in vrednotenje uporabil predvsem kvantitativne metode primerjalne analize in nekatere oblike večkriterijske analize. Funkcionalno urbano območje je sestavljeno iz večjega osrednjega naselja (vsaj 15.000 prebivalcev) ter njegovega funkcionalnega zaledja, ki ima skupaj vsaj 50.000 prebivalcev (v državah z več kot 10 milijoni prebivalcev) oz. vsaj 0,5 % prebivalcev države (v manjših državah). Funkcionalna povezanost osrednjega naselja in zaledja temelji predvsem na lokalnem trgu dela oz. na dnevni migraciji.

S projektom ESPON 1.4.3 (Študija urbanih funkcij) so bili upoštevani tudi morfološki kriteriji za opredelitev funkcionalnih regij kot dopolnitev projektu ESPON 1.1.1, in sicer gostota in število prebivalcev, sklenjenost urbane strukture in identiteta. Tako so bile kot funkcionalne urbane regije opredeljene Ljubljana, Maribor, Celje, Kranj, Koper, Nova Gorica in Novo mesto. Kot čezmejna funkcionalna območja so opredeljena Gorica – Nova Gorica, Trst – Koper in Gradec – Maribor.

1.6 Opredelitev funkcionalnih regij z metodo Intramax

1.6.1 Metoda Intramax

Intramax metoda pojmuje velikost interakcije (na primer dnevne migracije na delo) kot ključne za regionalizacijo območja. Interakcijsko matriko, tj. matriko tokov, vzame kot obliko verjetnostne matrike in na njeni podlagi oblikuje ciljno funkcijo, katere podlaga so razlike med opaženimi in pričakovanimi verjetnostmi, ki so povezane z mejnimi vsotami.

Na podlagi Tabele 1 lahko pričakujemo, da bodo pričakovane vrednosti vsakega elementa izpeljane kot zmnožek vsote tokov stolpca z deležem seštevka tokov vrstice glede na skupno število interakcij.

Tabela 1: Matrika tokov z j regijami

Delo Prebivališče	Regija 1	Regija 2	...	Regija j	Skupaj
Regija 1	iz 1 v 1	iz 1 v 2	...	iz 1 v j	$\sum_j a_{1j}$
Regija 2	iz 2 v 1	iz 2 v 2	...	iz 2 v j	$\sum_j a_{2j}$
...			...	...	...
Regija j	iz j v 1	iz j v 2	...	iz j v j	$\sum_i a_{jj}$
Skupaj	$\sum_i a_{i1}$	$\sum_i a_{i2}$	...	$\sum_i a_{ij}$	$n = \sum_i \sum_j a_{ij}$

Vir: W. Mitchell et al., *Identifying functional regions in Australia using hierarchical aggregation techniques*, 2007.

Pričakovan tok iz regije 2 v regijo 1, je tako:

$$a_{21}^* = \sum_i a_{i1} \frac{\sum_j a_{2j}}{\sum_i \sum_j a_{ij}} = \sum_i a_{i1} \left(\frac{\sum_j a_{2j}}{n} \right) \quad (1)$$

To je tok, ki ga lahko pričakujemo samo na podlagi velikosti seštevnikov stolpcev in vrstic.

Vsota vrstic v matriki je:

$$a_{i*} = \sum_j a_{ij} \quad (2)$$

Vsota stolpcev je:

$$a_{j*} = \sum_i a_{ij} \quad (3)$$

Skupno število interakcij je:

$$n = \sum_i \sum_j a_{ij} \quad (4)$$

Ničelna hipoteza za neodvisnost mejnih seštevnikov vrstic in stolpcev je tako:

$$H_0: a_{ij}^* = \frac{\sum_j a_{ij} \sum_i a_{ij}}{n} = \frac{a_{i*} a_{j*}}{n} \quad (5)$$

Če normaliziramo tokove tako, da je $n=1$ in $a_{ij}^* = a_{i*} a_{j*}$, potem avtorja metode Intramax Masser in Brown dodajata, da je razlika med opazovanimi in pričakovanimi vrednostmi tokov

med območjem i in območjem j ($a_{ij} - a_{ij}^*$), ki je merilo, koliko opazovana vrednost presega tok, ki bi ga lahko pričakovali na podlagi skupnih tokov.

V teh okoliščinah je ciljna funkcija algoritma za hierarhično razvrščanje v skupine na podlagi nesimetrične matrike opredeljena kot:

$$\max I = (a_{ij} - a_{ij}^*) + (a_{ji} - a_{ji}^*), \quad i \neq j \quad (6)$$

Avtorja metode uporabita tudi omejitve za sosednost v postopku, da bi odstranila možnost združevanja nesosednjih enot v skupine (Mitchell et al., 2007).

1.6.2 Intramax analiza z uporabo programa Flowmap

Prostorsko grupiranje z uporabo postopka Intramax je vključeno v program Flowmap, ki ga je razvila Fakulteta za geografske znanosti Univerze v Utrechtu. Program je narejen za analizo interakcijskih podatkov kot so migracijski tokovi, za analizo omrežij in analizo interakcij, na primer analizo dostopnosti in interakcijsko modeliranje.

Flowmap uporablja podatke, ki jih lahko združimo v tri razrede: podatke o tokovih, tabele razdalj in zemljevide. Podatki o tokovih podajajo povezave kot migracije med območji izvora (območje prebivališča) in ciljnim območji oz. destinacijami (območje zaposlitve). Tabela razdalj je matrika, ki vsebuje razdalje za vse kombinacije interakcij med izvori in cilji (na primer v metrih ali minutah). Tabele so izračunane in zgrajene znotraj Flowmapa. Zemljevidi vključujejo osnovni zemljevid, ki je običajno zemljevid območij ali mrež, zemljevid točk izvorov in zemljevid točk ciljev. Te točke so centriidi območij osnovnega zemljevida.

V Flowmapu ni kriterija za sosednost. Rezultati na višjih ravneh so avtomatično skoraj vedno sosednji, večinoma pa lahko nesosednje združitve povežemo z neustreznimi podatki ali napako.

Interakcije med različnimi lokacijami lahko pojmuje kot stopnjo funkcionalne razdalje, tj. več ko je interakcij, manjša je razdalja. Flowmap uporablja postopek Intramax, ki izvede regionalizacijo interakcijske matrike. Cilj metode Intramax je maksimizirati delež interakcije znotraj skupine na vsaki stopnji grupiranja, z upoštevanjem razlik v seštevkih vrstic in stolpcev matrike. V vsakem koraku Flowmap pogleda vsak par območij, ki ju lahko poveže, in združi vsak par območij, za katerega ima ciljna funkcija najvišjo vrednost. To pomeni, da se združi tisti par, ki maksimizira ciljno funkcijo:

$$\frac{T_{ij}}{O_i \cdot D_j} = \frac{T_{ji}}{O_j \cdot D_i} \rightarrow \max, \quad (7)$$

kjer je T_{ij} interakcija med izvorno lokacijo i in ciljno lokacijo j , $O_i = \sum_j T_{ij}$, $D_j = \sum_i T_{ij}$. Ciljna funkcija se lahko izračuna za vse $D_j > 0$ in vse $O_i > 0$, tj. cona, ki nima prebivališč ali zaposlitev se ignorira.

Postopek Intramax je metoda postopnega združevanja. V vsakem koraku se združita po dve območji in interakcija teh dveh postane notranja interakcija novega območja. Proces se ponavlja dokler ni vseh N območij združenih v $N - 1$ korakih v eno območje in vse interakcije postanejo notranje.

1.6.3 Opredelitev funkcionalnih regij za Slovenijo z uporabo metode Intramax

Za opredelitev funkcionalnih regij Slovenije sem izvedla analizo z uporabo hierarhičnega algoritma, ki je že kodiran v programu Flowmap. Analiza funkcionalnih področij podaja katalog možnih sistemov območij, ki temelji na identifikaciji območij z visoko stopnjo samozadostnosti na podlagi dnevnih delovnih migracij med 210 slovenskimi občinami za leto 2007.

Slika 3 prikazuje del izhodne datoteke analize za Slovenijo. Celotno število interakcij je celotno število dnevnih tokov na delo, od tega je 52,19 odstotka tokov znotraj območij (znotraj občin). Proces na vsaki stopnji pridruži območje, ki najbolj poveča delež interakcij znotraj območij, v absolutnem smislu, ne glede na velikost območja. V vsakem koraku je območje iz prvega stolpca združeno z ustreznim območjem v drugem stolpcu. V tretjem stolpcu je prikazano skupno število interakcij znotraj območij (število potovanj znotraj območja). Četrty stolpec prikazuje odstotno povečanje po združitvi. Zadnji stolpec pa prikazuje kumulativni odstotek interakcij in se postopno povečuje. Po 209 korakih (210 občin) je kumulativen odstotek 100, vsa območja (občine) so združene v eno območje (Slovenija).

Pri razvrstitvi je prišlo do napake pri občini Ribnica na Pohorju, ki pa jo glede na ostale podatke lahko enostavno razvrstimo v skupino koroških občin (program jo je uvrstil med občine jugovzhodne regije). Napačna razvrstitev je bila posledica napake v podatkih Statističnega urada in ne napake programa.

Slika 3: Izhodna datoteka analize Intramax

IMFUS003.txt - Notepad

File Edit Format View Help

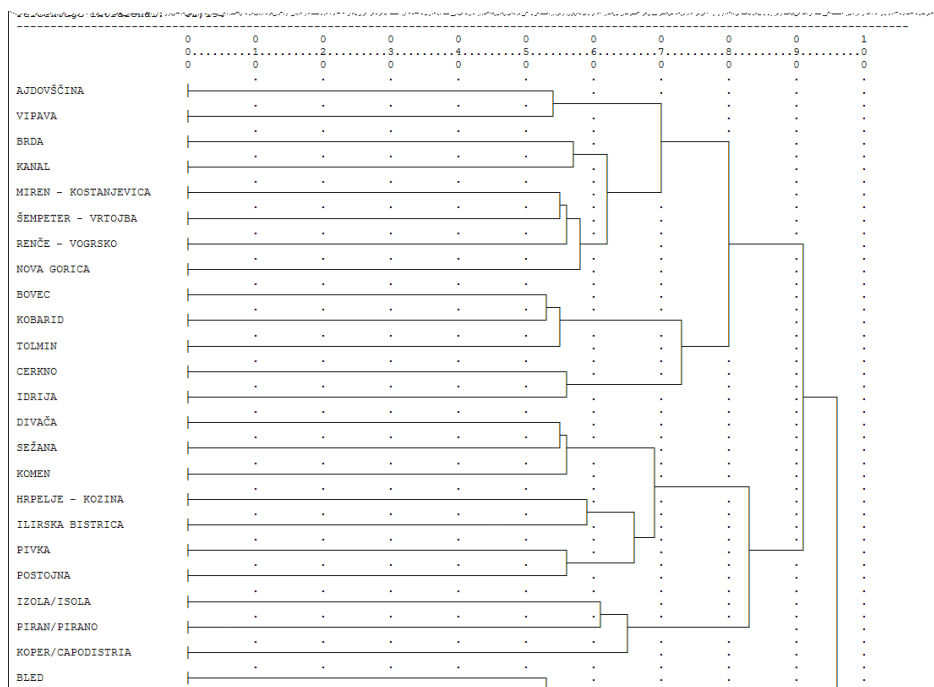
INTRAMAX ANALYSIS by Flowmap

Origin data from: C:\Program Files\FLOWMAP\FLOWMAP_LB\koordin1.dbf
 Destination data from: C:\Program Files\FLOWMAP\FLOWMAP_LB\koordin1.dbf
 Flow data from: C:\Program Files\FLOWMAP\FLOWMAP_LB\Flowobc2.dbf

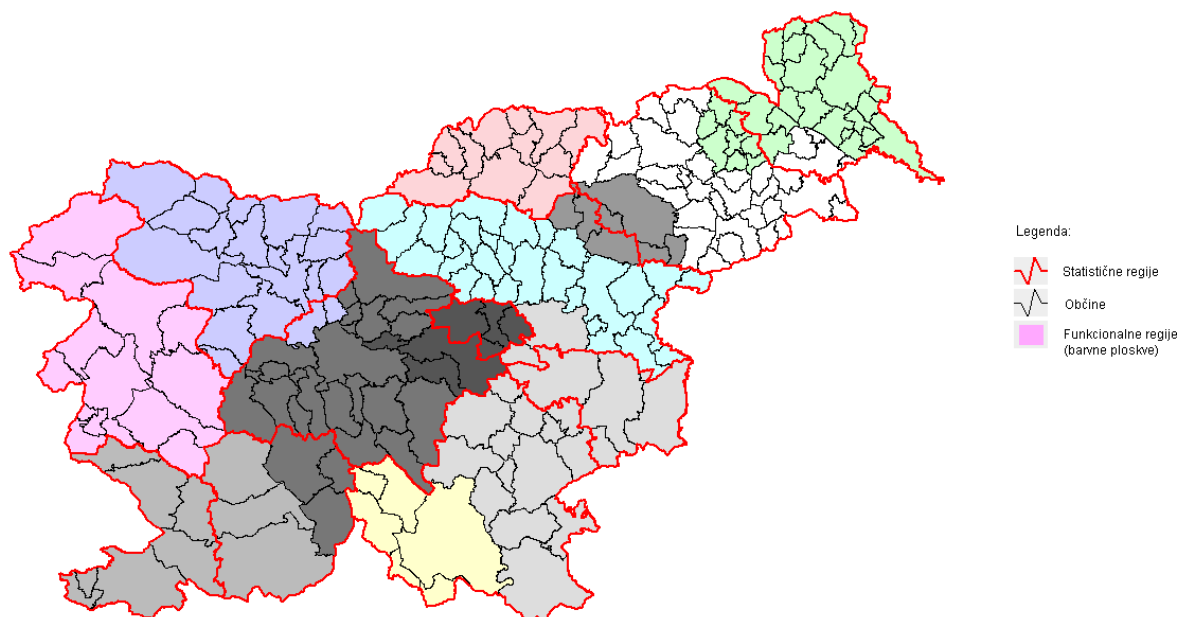
Total interaction: 825290
 Intrazonal interaction: 430730
 Percentage intrazonal: 52,19%

Step	Dissolved..... Area.....	Enlarged..... Area.....	Total Area.....	Percentage	Cumulative Interaction	Intrazonal Interaction	Intrazonal Interaction
1	RIBNICA NA POHORJU	-> RIBNICA	432737	0,24%	52,43%		
2	SVETI JURIJ V SLOV. GORICAH	-> SVETI JURIJ	432980	0,03%	52,46%	0,24%	52,43%
3	SALOVCI	-> HODOS/HODOS	432986	0,00%	52,46%	0,03%	52,46%
4	SOLČAVA	-> LUCE	433013	0,00%	52,47%	0,00%	52,47%
5	HODOS/HODOS	-> GORNJI PETROVCI	433111	0,01%	52,48%	0,01%	52,48%
6	ROGAŠOVCI	-> KUZMA	433148	0,00%	52,48%	0,00%	52,48%
7	MEZICA	-> CRNA NA KOROŠKEM	433654	0,06%	52,55%	0,06%	52,55%
8	SVETI ANDRAŽ V SLOV. GORICAH	-> CERKVENJAK	433678	0,00%	52,55%	0,00%	52,55%
9	VUZENICA	-> MUTA	434023	0,04%	52,59%	0,04%	52,59%
10	OSILNICA	-> KOSTEL	434028	0,00%	52,59%	0,00%	52,59%
11	KUZMA	-> GRAD	434083	0,01%	52,60%	0,01%	52,60%
12	KOBILJE	-> DOBROVNIK/DOBRONAK	434090	0,00%	52,60%	0,00%	52,60%
13	LJUBNO	-> GORNJI GRAD	434210	0,01%	52,61%	0,01%	52,61%
14	REČICA OB SAVINJI	-> NAZARJE	434514	0,04%	52,65%	0,04%	52,65%
15	KOZJE	-> BISTRICA OB SOTLI	434559	0,01%	52,66%	0,01%	52,66%
16	SENTRUPERT	-> MOKRONOG - TREBELNO	434703	0,02%	52,67%	0,02%	52,67%
17	ZAVRE	-> CERKULANE	434725	0,00%	52,68%	0,00%	52,68%
18	ŽETALE	-> PODLEHNİK	434749	0,00%	52,68%	0,00%	52,68%
19	VERŽEJ	-> KRIZEVCI	434815	0,01%	52,69%	0,01%	52,69%
20	RAZKRITZJE	-> LJUTOMER	435100	0,03%	52,72%	0,03%	52,72%
21	NAZARJE	-> MOZIRJE	435648	0,07%	52,79%	0,07%	52,79%
22	VELIKA POLANA	-> ČRENŠOVCI	435701	0,01%	52,79%	0,01%	52,79%
23	STRAŽA	-> DOLENJSKE TOPLICE	435965	0,03%	52,83%	0,03%	52,83%
24	RADLJE OB DRAVI	-> PODVELKA	436237	0,03%	52,86%	0,03%	52,86%
25	POLJČANE	-> MAKOLE	436343	0,01%	52,87%	0,01%	52,87%
26	SREDIŠČE OB DRAVI	-> ORMOŽ	436760	0,05%	52,92%	0,05%	52,92%
27	VRANSKO	-> TABOR	436803	0,01%	52,93%	0,01%	52,93%
28	PODČETRTEK	-> BISTRICA OB SOTLI	436941	0,02%	52,94%	0,02%	52,94%
29	TRNOVSKA VAS	-> CERKVENJAK	436956	0,00%	52,95%	0,00%	52,95%
30	SVETA ANA	-> BENEDIKT	436977	0,00%	52,95%	0,00%	52,95%
31	GORJE	-> BLEJ	437335	0,04%	52,99%	0,04%	52,99%
32	MOZIRJE	-> GORNJI GRAD	438123	0,10%	53,09%	0,10%	53,09%
33	SODRAŽICA	-> RIBNICA	438431	0,04%	53,12%	0,04%	53,12%
34	GRAD	-> CANKOVA	438489	0,01%	53,13%	0,01%	53,13%
35	ŠMARJEŠKE TOPLICE	-> ŠKOCJAN	438582	0,01%	53,14%	0,01%	53,14%
36	LOŠKA DOLINA	-> BLOKE	438712	0,02%	53,16%	0,02%	53,16%
37	GORNJA RADGONA	-> APAČE	439183	0,06%	53,22%	0,06%	53,22%
38	KOBARID	-> BOVEC	439374	0,02%	53,24%	0,02%	53,24%
39	SENTJUR	-> DOBJE	439534	0,02%	53,26%	0,02%	53,26%
40	PODVELKA	-> MUTA	440301	0,09%	53,35%	0,09%	53,35%
41	GORIŠNICA	-> DORNAVA	440408	0,01%	53,36%	0,01%	53,36%
42	SVETA TROJICA V SLOV. GORICAH	-> LENART	440724	0,04%	53,40%	0,04%	53,40%
43	LJUTOMER	-> KRIZEVCI	441516	0,10%	53,50%	0,10%	53,50%
44	TURNIŠČE	-> ČRENŠOVCI	441641	0,02%	53,51%	0,02%	53,51%

Slika 4: Dendrogram združevanja občin

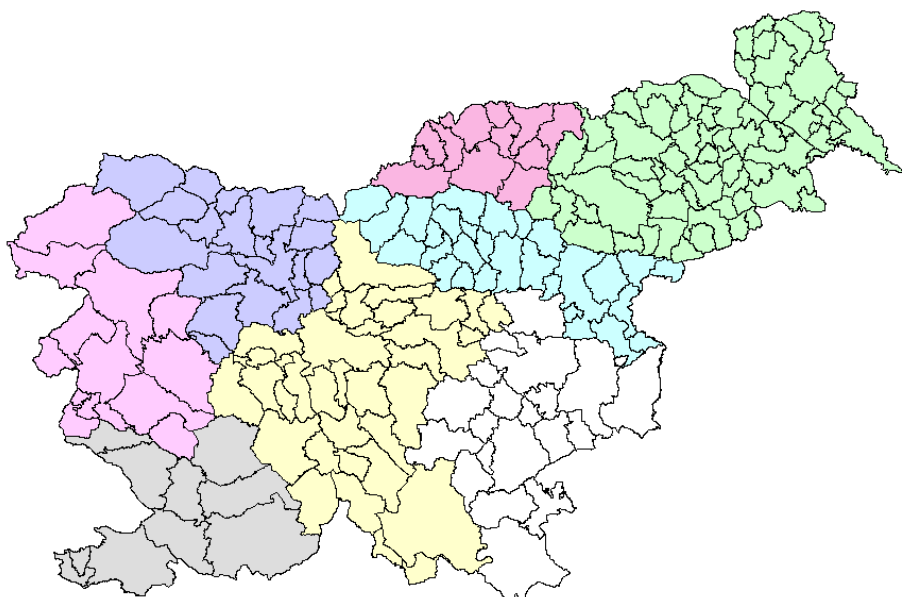


Slika 5: Razmejitev občin po statističnih regijah (SURS) in 12 funkcionalnih regijah opredeljenih z metodo Intramax

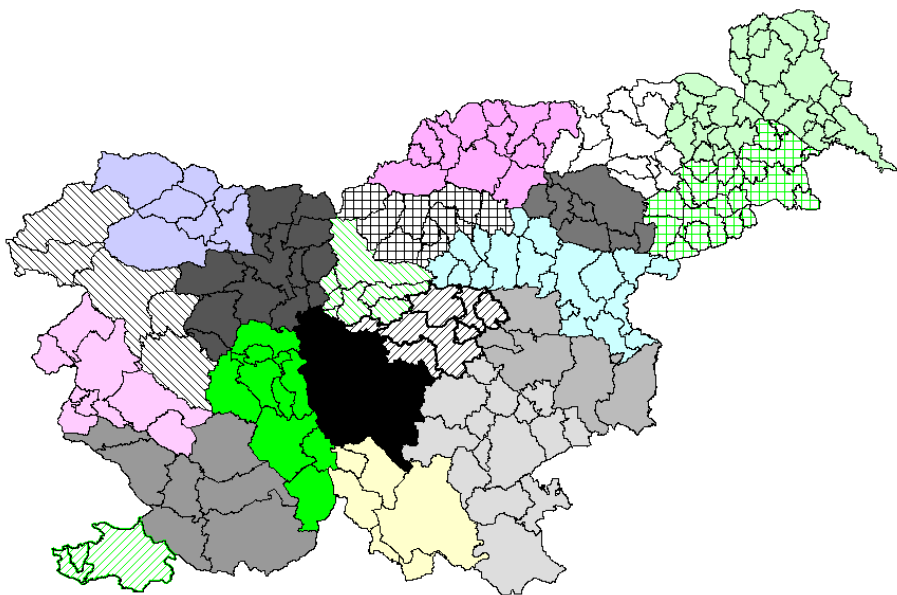


Z metodo je mogoče opredeliti katero koli število regij. Za lažjo primerjavo funkcionalnih z administrativno opredeljenimi regijami sem izbrala enako število regij, torej 12. Na spodnjih dveh slikah (Sliki 6 in 7) pa je prikazana razmejitev, pri kateri je 75 odstotkov interakcij znotraj regij (20 regij) in razmejitev, ki prikazuje razmejitev na 8 regij, ko je približno 90 odstotkov vseh interakcij znotraj tako opredeljenih regij.

Slika 6: Razmejitev občin na osem funkcionalnih regij opredeljenih z metodo Intramax



Slika 7: Razmejitev občin na 20 funkcionalnih regij opredeljenih z metodo Intramax



2 ANALIZE IN MODELI PROSTORSKIH INTERAKCIJ

Prostorska analiza oziroma prostorska statistika obsega metode, s katerimi se entitete proučujejo na podlagi njihovih topoloških, geometrijskih in geografskih značilnosti. V prostorski analizi se uporablja niz metod, ki so večinoma še v razvoju. Uporabljajo se različni analitični pristopi na zelo različnih področjih. Izraz se pogosto uporablja v bolj omejenem smislu, opisuje namreč metode, ki se uporabljajo pri analizi prostorskih oziroma geografskih podatkov (Wikipedia, 2009).

Prostorske analize so se pojavile z zgodnjimi kartografijami, anketami in geografijo, vendar pa so bile formalizirane šele v drugi polovici 20. stoletja. Moderna prostorska analiza zaradi velikih količin podatkov temelji na računalniških metodah, na uporabi modernih statističnih programov, GIS in kompleksnega računalniškega modeliranja. Metode prostorske analize so bile razvite v geografiji, biologiji, epidemiologiji, sociologiji, demografiji, statistiki, geoinformatiki, računalništvu, matematiki itd.

V prostorski analizi se pojavlja vrsta kompleksnih vprašanj, ki niso niti jasno opredeljena niti v celoti rešena in so podlaga za trenutne raziskave. Glavni problem je določitev prostorske lokacije entitet, ki se proučujejo, problematični pa so tudi omejeno matematično znanje, zahtevane predpostavke obstoječih statističnih metod in problemi računalniških izračunov.

2.1 Izgradnja prostorskega modela

Prostorsko modeliranje je praktični pristop k prostorski analizi, katerega cilj je najprej razumeti in opisati mehanizme, ki določajo strukturo in delovanje urbanega sistema in nato napovedati izide alternativnih političnih odločitev v prihodnosti. Njihov razvoj je posledica uporabe v prostorskem načrtovanju.

2.1.1 Namen modela

Razlogi za izgradnjo prostorskega modela morajo biti jasno opredeljeni na začetku projekta. Poleg splošnih ciljev je potrebno opredeliti tudi vprašanja, ki se zastavljajo ob problemu, na primer, kako bo izgradnja nove hitre ceste vplivala na vzorce dnevnih migracij in dostopnost odročne regije ipd. Izgradnja modela je samo del celotne naloge, zato mora biti vloga modela v analizi jasna. Ker se prostorski modeli ukvarjajo z rabo prostora, se razvijajo predvsem za analize umestitev poselitev, zaposlitvenih in storitvenih zmogljivosti ter interakcij med njimi v smislu prometnega sistema (Foot, 1981).

Načeloma se prostorski modeli uporabljajo za:

- analizo alternativnih (strukturnih, regionalnih) razvojnih načrtov in izbiro najboljšega načrta.
- analizo vplivov specifičnih ukrepov in politik na spremembe ter ovrednotenje teh sprememb.

2.1.2 Opredelitev območja proučevanja in izbira sistema prostorskih enot

Ko se odločimo za implementacijo prostorskega modela, je treba najprej opredeliti meje območja, ki ga bomo proučevali. Območje proučevanja je lahko država, regija, občina ali celo mesto oziroma naselje. Splošno pravilo je, da se izbere širše območje, ki je zaokroženo in kar se da samozadostno, tokovi iz tega območja pa minimalizirani. Mesto na robu območja na primer ni smiselno, saj tako ne zajamemo vseh interakcij.

Območje proučevanja nato razdelimo na enote, pri čemer velikost in narava teh enot lahko močno variira. Pri odločitvi, kakšne enote izbrati, je pomembno, da vzamemo take enote, za katere obstajajo podatki. Pri tem vzamemo najmanjše možne prostorske enote, ki jih po potrebi lahko tudi združimo (Masser & Scheurwater, 1980).

Prostorske enote naj bi si bile po velikosti čim bolj podobne. Na eni strani ne smejo biti prevelike, saj je v tem primeru delež notranjih tokov večji kot med enotami, veliko enot pa je tudi na obrobju. Na drugi strani pa so pri zelo majhnih enotah tokovi majhni, zato modeli težko opišejo vzorce migracij. Pri zelo velikem številu enot je tudi težko interpretirati in razumeti rezultat modela (Foot, 1981).

Za modeliranje funkcionalnih regij je v območje proučevanja smiselno vključiti tudi enote zunaj neke zaključene celote, kot je država, zlasti kadar je na robu območja mesto, ki je dostopno tudi iz druge države ali pa zunaj območja proučevanja obstaja mesto, ki pritegne tokove iz proučevanega območja.

Problemi, ki se lahko pojavijo z razširitvijo regij preko državne meje, so vezani na samo razdelitev proučevanega območja na enote: katero raven enot vzeti in kakšna je njihova opredelitev ter primerljivost opredelitve in velikosti enot z enotami znotraj države. Naslednji problem so podatki; njihova dosegljivost in opredelitev. Države Evropske unije na primer zbirajo podatke za Eurostat po istih definicijah, problem so podatki na nižjih ravneh, predvsem podatki o tokovih, ki bi jih bilo potrebno pridobiti.

2.1.3 Podatki

Zbiranje podatkov je tesno povezano s predhodnimi koraki v izgradnji modela. Razpoložljivost podatkov lahko močno vpliva na opredelitev proučevanega območja in njegovo razčlenitev, tip modela in obseg njegove aplikacije.

Prostorski podatki vsebujejo tako informacije o značilnostih nekega območja kot tudi informacijo o lokaciji. Razlikovanje med prostorskimi in neprostorskimi podatki je pomembno, ker veliko analitičnih metod za neprostorske podatke ni ustreznih za analizo prostorskih podatkov. Ti imajo vrsto značilnosti in problemov, zato zahtevajo posebne statistične metode in pristope pri modeliranju (Fortheringham, Brunson & Charlton, 2002).

Ena prvih nalog je določiti razpoložljivost podatkov iz vseh možnih virov. To vključuje tako objavljene statistične informacije, kot je na primer popis prebivalstva in medpopisne ocene ter druge uradne publikacije, dodatne informacije lokalnih oblasti in rezultate raziskav. Če podatkov za izgradnjo modela ni dovolj, jih je potrebno dopolniti z dodatnimi raziskavami in izračuni (Foot, 1981).

Ključni podatki za prostorske analize interakcij so podatki o značilnostih enot v proučevanem območju, tokovi med enotami in razdalje.

2.1.3.1 Značilnosti prostorskih enot

Podatki, ki opisujejo značilnosti prostorskih enot, opisujejo neko stanje ali gibanje na ekonomskem, demografskem in socialnem področju ter na področju okolja in naravnih virov, zbirajo pa se na različnih prostorskih ravneh, na državni, regionalni, občinski ali še nižjih ravneh. Prostorske ravni, na katerih se zbirajo podatki, so navadno določene na državni ravni, saj je le tako mogoče dobiti kar se da popolne in točne podatke.

2.1.3.2 Tokovi

Castells meni, da tokovi kapitala, informacij, tehnologije, organizacijskih interakcij, slik, zvoka, simbolov dominirajo ekonomsko, politično in simbolno življenje. Prostor vse bolj postaja prostor tokov, zato zahteva drugačne analize (Spiekermann & Wegener Urban and Regional Research, 2007). Čeprav so se tokovi med lokacijami zelo povečali, je lokacija še vedno pomembna, saj ljudje še vedno bivajo, delajo, preživljajo prosti čas in se družijo na neki lokaciji.

Tokovi so močno povezani s prostorsko politiko in cilji konkurenčnosti, teritorialno kohezijo in okoljsko trajnostjo. Dostopnost informacij ter razvoj komunikacij in tehnologije lahko ustvari koridorje med velikimi mesti in izboljša njihov položaj na globalnem trgu, zmanjša pa možnosti manjših mest in deluje proti policentričnosti. Zato je pomembno razumeti, kako tokovi vplivajo na prostor in prostorsko politiko. Spremljanje in napovedovanje tokov bi moralo biti zato eno ključnih področij zbiranja podatkov in raziskovanja pri oblikovanju regij (ESPON 1.4.4).

Tokovi dnevnih migracij: Zaradi boljših povezav med mesti se je povečalo zaledje dnevnih migrantov na delo, povečalo se je število migrantov iz bolj oddaljenih krajev, prometni zamaški ob prometnih konicah in okoljski problemi v mestih. To ima pomembno posledico za regionalno analizo, saj so regionalni gospodarski kazalniki, kot je BDP na prebivalca, pogosto izkrivljeni, ker se dohodek, ustvarjen v glavnih mestih, troši v gospodinjstvih v širšem zaledju teh mest. Analiza dnevnih migracij je zato zelo pomembna za prostorski razvoj in prostorsko politiko, saj lahko dopolnjuje analizo urbano-ruralnih odnosov in različnih prostorskih postavitev v urbanih sistemih.

Migracijski tokovi: Migracije so ena od dveh najpomembnejših komponent razvoja prebivalstva. Poleg naravne rasti ali upadanja zaradi spremenjenih trendov v rodnosti in naraščanja pričakovane življenjske dobe, je rast prebivalstva v državi, regijah, občinah in mestih odvisna od bilance migracijskih tokov; neto učinek rodnosti, smrtnosti in migracij določa razvoj prebivalstva na nekem ozemlju.

Medtem ko je migracije med državami zaradi političnih, jezikovnih in drugih ovir težko napovedovati, migracij med regijami in mesti ni težko modelirati. Navadno je ovira pri notranjih selitvah strošek selitve in zaposlitvene in/ali izobraževalne možnosti v regiji. Pomemben dejavnik so tudi dostopne stanovanjske zmogljivosti.

Transportni tokovi: Transportni tokovi so eden najbolj očitnih dokazov prostorske narave večine socialnih in ekonomskih dejavnosti. Zaradi večje integracije trgov so se povečali tokovi potnikov, še posebej pa tovara/blaga, zaradi česar je postal transport eden najpomembnejših pogojev za ekonomsko dejavnost, zato je tako pogosto na političnem dnevnem redu. Transport tudi pomembno prispeva k porabi energentov in proizvodnji toplogrednih plinov. Analiza in napovedovanje transportnih tokov ter s tem njihovih prostorskih in okoljskih vplivov sta zato pomembni, saj s pravimi in podrobnimi

informacijami oblikovalci politike uravnotežijo konflikte pri doseganju konkurenčnosti, kohezije in trajnosti.

Trgovinski tokovi: Na proces globalizacije je najbolj vplivala mednarodna trgovina, vendar pa so trgovinske povezave vedno manj stalne v prostoru in jih ni mogoče več analizirati brez nekaterih drugih povezav, še posebej kapitalskih interakcij (neposredne tuje investicije) in tudi družbenih interakcij (npr. migracij). Mednarodna trgovina je tudi vse manj izmenjava dobrin med državami in vse bolj izmenjava delov in vmesnih proizvodov čez državne meje znotraj mednarodnih korporacij.

V ekonomski geografiji se za opis tokov blaga in kapitala uporablja teorija gravitacije. Pri analizi gravitacijske interakcije igra pomembno vlogo funkcija oddaljenosti (angl. *distance decay*), kjer so interakcije navadno obratno sorazmerne z razdaljo med regijama. Danes je uporaba gravitacijskih modelov zaradi vse manjšega pomena fizične razdalje postavljena pod vprašaj. Gravitacijski modeli so bili pogosto kritizirani tudi zaradi predpostavke funkcijske komplementarnosti med državami in zaradi administrativnih ovir v trgovini.

Finančni tokovi: Neposredne tuje investicije igrajo pomembno vlogo v gospodarstvu vsake države. Podjetja, ki vlagajo v nove tehnologije, nove oblike vodenja in organizacijo proizvodnje prispevajo k modernizaciji podjetij in povečujejo zaposlenost. V zadnjem času so se povečale investicije v storitvenem sektorju, manj pa je vlaganj v proizvodne zmogljivosti.

Turistični tokovi: Turistični tokovi predstavljajo pomembno komponento potovanj in imajo velik pomen za ekonomski razvoj.

2.1.3.3 Razdalje

V prostorski analizi so opazovanja vezana na lokacijo, bližina med lokacijami pa se meri z razdaljo. Tobler v Prvem zakonu v geografiji ugotavlja, da je v primeru, da se interakcije med entitetami z bližino povečujejo, primerna predstavitev v geografskem prostoru in ocena z uporabo metod prostorske analize. Za modeliranje interakcij med aktivnostmi v regiji torej potrebujemo neko merilo razdalje.

Evklidska razdalja predstavlja geografsko bližino med lokacijami in jo izračunamo kot razdaljo med središčema (centroidoma) območij. Razdalja je lahko cestna razdalja, ki jo dobimo iz podrobnega zemljevida cestnega omrežja. Pri tem lahko upoštevamo tudi razdalje znotraj cone. Razdalja je lahko potovalni čas ali nek drug splošen strošek. V nekaterih državah so ti podatki del uradne popisne statistike o načinu potovanja na delo in porabljenem času, iz česar dobimo povprečni potovalni čas med območji.

Bogataj je v projektu Razvoj modela za potrebe regionalne politike (2008) ugotovil, da je najboljše merilo potovalni čas, saj so na različnih kategorijah cest različne potovalne hitrosti. Največkrat se potovalne razdalje računajo za potovanje z osebnim avtom, smiselno pa bi bilo

upoštevati tudi druga prevozna sredstva, npr. avtobus ali vlak. Potovalno razdaljo lahko tudi izboljšamo z vključitvijo časa za parkiranje in časom hoje do cilja.

Poleg razdalj lahko na interakcijami med območji vpliva tudi povezanost (angl. *connectivity*), oziroma sosednost, kot je na primer obstoj skupnih mej, in smer (angl. *direction*).

2.2 Analiza dostopnosti

Dostopnost je glavni 'proizvod' prometnega sistema. Naloga prometne infrastrukture je omogočiti prostorsko interakcijo, tj. mobilnost oseb in dobrin za družbene, kulturne in ekonomske dejavnosti. V okviru prostorskega razvoja kakovost prometne infrastrukture, v smislu zmogljivosti, povezanosti, prometnih hitrosti ipd. določa lokacijsko prednost območja (regije, občine, mesta) glede na druge lokacije in pomeni konkurenčno prednost.

Kazalniki dostopnosti merijo koristi, ki jih imajo gospodinjstva in podjetja od obstoja in uporabe prometne infrastrukture pomembne za njihovo območje (Spiekermann & Wegener Urban and Regional Research, 2007).

Pomen prometne infrastrukture za prostorski razvoj v najpreprostejši obliki je v tem, da bodo območja z boljšim dostopom do proizvodnih virov in trgov, ceteris paribus, produktivnejša, konkurenčnejša in zato uspešnejša kot bolj oddaljena in odmaknjena območja (Linneker, 1997). S tem vodijo investicije v prometno infrastrukturo v spreminjanje kakovosti lokacij in lahko spremenijo tok prostorskega razvoja.

2.2.1 Kazalniki dostopnosti

V splošnem je dostopnost zgrajena iz dveh funkcij; prve, ki predstavlja dejavnosti ali priložnosti do katerih se dostopa, in druge, ki predstavlja napor, čas, oddaljenost ali strošek dostopa do teh dejavnosti.

Kazalnike dostopnosti lahko razvrstimo glede na njihovo specifikacijo funkcij dejavnosti in oddaljenosti. Tako kazalniki stroškov potovanja merijo skupen ali povprečen strošek potovanja do vnaprej opredeljenega niza ciljev oziroma destinacij, na primer povprečen čas potovanja do vseh mest z več kot 50 tisoč prebivalcev.

Dnevna dostopnost temelji na določenem proračunu za potovanje, v okviru katerega je treba doseči neki zeleni cilj. Kazalnik izhaja iz primera poslovnega potnika, ki želi potovati v nek kraj in se zvečer vrača v kraj prebivališča.

2.2.1.1 Potencialna dostopnost

Potencialna dostopnost je eden najbolj splošnih in najbolj testiranih kazalnikov dostopnosti. Temelji na predpostavki, da se privlačnost cilja povečuje z njegovo velikostjo in upada z razdaljo. Velikost ciljnega območja navadno predstavljajo prebivalstvo ali ekonomski kazalniki, kot je BDP. Dostopnost glede na populacijo je za dobavitelje dobrin in storitev kazalnik velikosti trga, dostopnost glede na BDP pa kazalnik velikosti trga 'high-level'? poslovnih storitev. Potencialna dostopnost je oblikovana na jasnih vedenjskih osnovah, toda vsebuje parametre, ki jih je treba umeriti.

Potencialna dostopnost je sestavljena iz dveh funkcij, funkcije dejavnosti, ki predstavlja dejavnosti in priložnosti, do katerih se dostopa, in funkcije oddaljenosti, ki predstavlja napor, čas, razdaljo ali strošek dostopa. Za merjenje potencialne dostopnosti se ti dve funkciji dopolnjujeta multiplikativno, tj. sta uteži ena drugi in sta obe potrebna elementa dostopnosti:

$$A_i = \sum_j W_j^\alpha \exp(-\beta c_{ij}) \quad (8)$$

kjer je A_i dostopnost območja i , W_j je dejavnost W , do katere se dostopa v območju j , in c_{ij} strošek dostopanja iz območja i do območja j . Večje ko je število privlačnih ciljev v območju j in bolj ko so območja j dostopna iz območja i , večja je dostopnost območja i .

W_j je na primer število delovnih mest v območju j , govorimo pa o ciljno omejenem modelu, ko želimo izračunati dostopnost do storitev in zaposlitev. Možen pa je drug pristop, ko nas zanima dostopnost podjetij do trga dela in potrošnikov. V tem primeru uporabimo model, ki je omejen z izvirnim območjem, na primer prebivalstvom nekega območja.

Privlačnost dejavnosti W_j je lahko tehtana z eksponentom α , ki je večji od 1 in upošteva učinke zgostitev (aglomeracije). Funkcija oddaljenosti ni linearna. Običajno se uporablja negativna eksponentna funkcija, v kateri velik parameter β daje večjo težo bližnjim destinacijam kot bolj oddaljenim.

2.2.1.2 Multimodalna analiza dostopnosti

Vsak od različnih tipov dostopnosti ima svoje prednosti in slabosti. Kazalniki o potovalnem času in dnevni dostopnosti so lahko razumljivi in razložljivi, vendar jim manjka teoretična podlaga. Potencialna dostopnost je osnovana na jasnih vedenjskih principih, a vsebuje parametre, ki jih je treba oceniti in njihove vrednosti ni mogoče izraziti v poznanih enotah.

Iz treh osnovnih kazalnikov je mogoče izpeljati vrsto izvedenih kazalnikov, od katerih so najpomembnejši multimodalni, intermodalni in interoperabilna dostopnost, ki se razlikujejo po tem, kako se izračunavajo stroški dostopa.

Kazalniki modalne dostopnosti se navadno predstavljajo ločeno za prikaz razlik med posameznimi prevoznimi sredstvi. Lahko pa jih povežemo v en sam kazalnik, ki prikazuje sestavljeni učinek ali alternativne prevozne možnosti za neko lokacijo. Obstajata dve pomembni metodi integracije. Prva je izbira najhitrejšega prevoznega sredstva, ki je običajno zračni prevoz za daljše ter cestni ali železniški prevoz za krajše razdalje. Drug način pa je izračun agregatnega kazalca dostopnosti s kombiniranjem modalnih dostopnosti.

Intermodalni kazalniki dostopnosti upoštevajo intermodalna potovanja, ki so kombinacija dveh ali treh sredstev npr. prevoza po cesti, železnici in pomorskega prevoza. Ti kazalniki so najbolj pomembni za logistične verige in prevoz tovora z različnimi možnimi kombinacijami prevoznih sredstev, kot je kombinacija železniškega in ladijskega tovornega prevoza. V potniškem prometu se izračunavajo za kombinacije železniškega in letalskega prevoza, dostop z avtomobilom do železnice itd.

Z multimodalno analizo dostopnosti se analizira dostopnost glede za več prevoznih sredstev, in sicer se upošteva število ali odstotek uporabnikov in strošek potovanja za določeno prevozno sredstvo.

Dostopnost je tesno povezana z zgodnjimi gravitacijskimi modeli, ki so predstavljeni v nadaljevanju.

2.3 Modeli prostorskih interakcij

Prostorska interakcija je širok pojem in obsega gibanje v prostoru, ki je rezultat človeškega delovanja, in vključuje dnevne migracije na delo, migracije, tokove informacij in blaga, vpise študentov na univerze in udeležbo na konferencah, uporabo javnih in zasebnih objektov in storitev ter celo posredovanje znanja. Gravitacijski modeli so najbolj razširjeni modeli prostorskih interakcij, ki se uporabljajo v analizi in pri napovedovanju vzorcev prostorskih interakcij (Haynes & Fotheringham, 1984).

Gravitacijski modeli temeljijo na ideji Newtonovega zakona o gravitaciji in so se uporabljali že sredi 19. stoletja, ko je E. G. Ravenstein na podlagi popisa objavil zakon o migraciji, v katerem je poudaril, da večina ljudi potuje na kratke razdalje in da manj ljudi potuje na daljše razdalje.

2.3.1 Gravitacijski model nakupovalnih središč

William J. Reilly je v 30. letih prejšnjega stoletja predlagal dve preprosti pravili, da bi opisal nakupovanje v mestih, in sicer je opazil, da večje mesto pritegne več trgovine iz ostalih mest

v regiji. Mesto z dvakrat večjo populacijo pritegne približno dvakrat več trgovine. Hkrati mesto pritegne več trgovine iz manj oddaljenih mest kot iz bolj oddaljenih. Opazil je, da trgovina upada približno s kvadratom razdalje. Privlačnost nakupovalnega središča R_i v mestu i s populacijo P_i je tako za nakupovalce iz mesta k in razdaljo med mestoma d_{ij} sledeča:

$$R_i = \frac{P_i}{d_{ki}^2} \quad (9)$$

Reillyja je zanimala tudi relativna privlačnost dveh nakupovalnih središč za prebivalce locirane med središčema in razmejitvena točka med središčema. Razmejitvena točka je točka, do katere ima neko središče dominanten vpliv. Metoda za iskanje razmejitvene točke se lahko razširi na vse možne poti med dvema središčema, s čimer dobimo prodajno območje nekega nakupovalnega središča.

2.3.2 Gravitacijski model brez omejitev

Prve uporabe gravitacijskih modelov so temeljile na modelu brez omejitev. Newtonov zakon o gravitaciji je bil prilagojen za proučevanje prostorskih interakcij tako, da so interakcije med dvema mestoma premo sorazmerno povezane z velikostjo mest (merjene s številom prebivalcev prvega in drugega) in inverzno s kvadratom razdalje med njima:

$$T_{ij} = G \frac{P_i P_j}{d_{ij}^2} \quad (10)$$

Bolj splošna formulacija je:

$$T_{ij} = G O_i D_j d_{ij}^{-\lambda}, \quad (11)$$

saj je potrebno parameter zmanjševanja z razdaljo λ oceniti.

Seveda se ljudje ne gibljejo iz kraja v kraj zaradi razlik v številu prebivalcev. Število prebivalcev se uporablja kot merilo za priložnosti, na primer zaposlitve, storitve, bivanje itd. Predpostavljamo, da večja populacija pomeni več priložnosti za zaposlitev in ustvari več storitev ter s tem večje število za potovanje ob enaki verjetnosti, da bo oseba potovala.

Razdaja med krajema ne sme biti evklidska razdalja med dvema točkama, temveč stvarna razdalja, ki upošteva kakovost in način prevoza med tema točkama, na primer prisotnost ali odsotnost cest, železniške povezave in kakovost le-teh. Zgornja formula je samo groba ponazoritev odnosov med dvema naseljenima območjema in migracijo kot povezavo med njima, ki ne upošteva brezposelnosti, ravni plač, razlike v življenjskih stroških in kakovosti življenja.

2.3.3 Gravitacijski model z enojno omejitvijo

Gravitacijski modeli z enojno omejitvijo so omejeni ali z izvorom ali s ponorom in se uporabljajo predvsem za napovedovanje prihodnjih interakcij med izvornimi in ponornimi območji. Oba modela delujeta na enak način, različna je samo smer. V modelu z omejitvijo izvora je fiksno določeno število ljudi, ki izhajajo iz nekega območja; napovedujemo, kam bodo šli. V modelu, omejenem s ponorom, je fiksno določeno, kam bodo šli; napovedujemo, iz kje prihajajo.

Model z enojno omejitvijo ocenjuje najbolj verjetno porazdelitev tokov v matriki izvorov in ponorov in je sestavljen iz dveh enačb.

Za model omejen s ponorom:

$$T_{ij} = B_j W_i D_j f(c_{ij}, \lambda), \quad (12)$$

$$B_j = 1 / \sum_i O_i f(c_{ij}, \lambda), \quad (13)$$

$f(c_{ij}, \beta) = e^{-\lambda c_{ij}}$ v primeru eksponentne funkcije,

$f(c_{ij}, \beta) = c_{ij}^{-\lambda}$ v primeru potenčne funkcije, kjer je:

T_{ij} – interakcija med območjem izvora i in območjem ponora j ,

O_i – dejavnost, ki se razporeja iz območja i ,

D_j – indeks privlačnosti območja j ,

B_j – utež oziroma faktor konkurenčnosti,

c_{ij} – strošek potovanja iz i v j ,

$f(c_{ij}, \lambda)$ – funkcija stroška potovanja, ki je običajno inverzna potenčna ali negativna eksponentna funkcija.

2.3.4 Garin-Lowryjev gravitacijski model

Garin-Lowryjev gravitacijski model skuša prikazati stanovanjsko, zaposlitveno in storitveno strukturo območja in je sestavljen iz dveh modelov z omejitvijo, ki se izvajata zaporedno; iz 'stanovanjskega modela' in 'storitvenega modela'. 'Stanovanjski model' se uporablja za izračun števila prebivalcev, ki se bodo naselili v območjih regije, in sicer na osnovi zaposlitvenih priložnosti v bližnjih območjih in bivanjske privlačnosti območij.

$$T_{ij} = A_i E_i H_j f(c_{ij}, \lambda), \text{ kjer} \quad (14)$$

$$A_i = 1 / \sum_j H_j f(c_{ij}, \lambda) \text{ in} \quad (15)$$

T_{ij} – število ljudi, ki dela v območju i

A_i – utež privlačnosti

E_i – zaposlitev v nestoritvenem sektorju v območju i
 H_j – privlačnost območja j , na primer število stanovanj
 c_{ij} – strošek potovanja iz i v j ,
 $f(c_{ij}, \lambda)$ – funkcija stroška potovanja, λ se določi ob umeritvi modela.

'Storitveni model' izračuna število zaposlitev v storitvenem sektorju na podlagi števila prebivalcev v območjih po 'stanovanjskem modelu' in t.i. 'storitvene privlačnosti' območja, ki nakazuje privlačnost za podjetja, da na proučevanem območju ponudijo svoje storitve.

$$S_{ij} = B_j P_j F_i f(c_{ij}, \mu), \text{ kjer} \quad (16)$$

$$B_j = 1 / \sum_i F_i f(c_{ij}, \lambda) \text{ in} \quad (17)$$

S_{ij} – število ljudi iz območja j , ki uporablja storitve v območju i
 B_i – utež privlačnosti
 P_{ji} – porazdelitev prebivalstva v j , izračunana s prejšnjo enačbo
 F_j – privlačnost storitev območja i , na primer površina storitvenih objektov
 c_{ij} – strošek potovanja iz i v j ,
 $f(c_{ij}, \mu)$ – funkcija stroška potovanja, μ se določi ob umeritvi modela.

Ker se v regiji pojavi več zaposlitev, se priseli več ljudi. Zato se t. i. stanovanjski model izvede še enkrat in nato 'storitveni model'. Cikel se nadaljuje do točke, ko se nova delovna mesta skoraj ne ustvarjajo več in se priseljevanja ustavijo. Tako se oba modela izvedeta večkrat.

Glavna predpostavka modela je, da je regionalna rast funkcija razvoja osnovnega sektorja, ki je sektor, ki zadovoljuje povpraševanje zunaj regije (nestoritveni sektor). Zaposlenost v tem sektorju pa vpliva na druga dva sektorja, trgovino oziroma storitve in na stanovanjski sektor.

Garin-Lowryjev model poda vrsto rezultatov. Izračunana sta dva parametra zmanjševanja z razdaljo za vsak model svoj. Izračuna se tudi porazdelitev prebivalstva po območjih in porazdelitev delovnih mest v storitvenem sektorju. Rezultata se lahko dodata kot novi spremenljivki v datoteko območij. Parametra razdalje pa se lahko uporabita za izračun interakcijskih matrik.

Potem ko ocenimo Garin-Lowryjev model, ga lahko uporabimo za napovedovanje. Običajen postopek je ocena vhodnih spremenljivk v prihodnosti in nato napoved porazdelitve prebivalstva in delovnih mest v nekem prihodnem obdobju. Spremenljivke v modelu se lahko dodatno spremenijo, da bi lahko testirali vpliv sprememb v regiji in simulirali rezultate alternativnih razvojnih politik. Za simuliranje alternativnih zaposlitvenih politik spreminjamo E_i , za simuliranje stanovanjskih politik spreminjamo H_j , lahko simuliramo ponudbo storitev,

prometno politiko in dostopnost ter s postavitvijo omejitev glede poselitve in gradnje novih storitvenih kapacitet celo okoljske politike. Model se ne uporablja samo za boljše razumevanje planskih problemov na proučevanem območju, temveč tudi za ocenjevanje posledic nekaterih planskih politik.

Prihodnje prerazdelitve prebivalstva ali zaposlitev so lahko omejene z omejitvami glede prihodnje rasti območij. Lahko gre za fizične omejitve nekega območja ali pa so rezultat planskih politik, ko gre za zaščito kakovostnih kmetijskih zemljišč ali okoljevarstvene pomisleke. V tem primeru je potrebno v vsaki ponovitvi cikla testirati, ali je nova porazdelitev znotraj postavljenih omejitev, lahko pa omejitve vgradimo tudi v sam model, npr. z vpeljavo novih spremenljivk.

Sam model ima več omejitev. Je statičen in ne pove ničesar o razvoju prometne infrastrukture. Poleg tega je trenutno gonilo razvoja storitveni sektor in s tem osnova produktivnosti in dinamike, čemur pa bi se lahko izognili, če bi v model vključili zaposlitev v storitvenem sektorju kot osnovno zaposlitev (Rodrigue, Comtois & Slack, 2009).

2.3.5 Razširjeni gravitacijski modeli

Razširjeni gravitacijski modeli upoštevajo gravitacijsko načelo v prej obravnavanih modelih, zanje pa je značilno, da vanje vključimo dodatne spremenljivke, ki so vezane na ekonomsko dogajanje, na primer zaposlitvene priložnosti, dohodke, nezaposlenost, cene stanovanj, tveganje, geografske značilnosti, politične razmere itd.

Model lahko zapišemo linearno. Po logaritmiranju vrednosti dobimo linearen model, kjer so koeficienti neodvisnih spremenljivk kar elastičnosti. Model ocenimo z regresijsko analizo.

Razširjen gravitacijski model je bil na primer uporabljen za napovedovanje dnevnih migracij v projektu Razvoj modela za potrebe regionalne politike, pri čemer je bil gravitacijski model za dnevne migracije med slovenskimi statističnimi regijami dopolnjen z indeksom razvojne ogroženosti, osebnimi dohodki in zaposlenostjo (Bogataj et al., 2008):

$$DC_{ij} = 2P_i^{0.75} P_j^{0.126} d_{ij}^{-2.66} \frac{K_{BOD,i}^{4.33} K_{BOD,j}^{10.53}}{K_{ZAP,i}^{-4.15} K_{ZAP,j}^{-2.73} K_{OGR,i}^{-0.86} K_{OGR,j}^{-0.99}} \quad (18)$$

Iz ocenjene enačbe sledi, da raste število dnevnih migracij v regije z višjim indeksom ogroženosti v povprečju bolj kot v regije z nižjim indeksom ogroženosti.

Razširjeni gravitacijski model spada v regresijske modele, ki so obravnavani v nadaljevanju.

2.4 Regresijski modeli

Najbolj splošno uporabljena metoda za modeliranje v družbenih znanostih je regresija. Običajno se regresija uporablja tako, da je odvisna spremenljivka povezana z nizom neodvisnih spremenljivk, pri čemer so ocene parametrov posamezne neodvisne spremenljivke na odvisno spremenljivko eden glavnih rezultatov analize.

Za regresijske analize se običajno uporabljajo socialno-ekonomski podatki in podatki o prostorskih dejavnostih, ki so zbrani po območjih oziroma conah. Modeli so bili večinoma razviti za spremenljivke, ki predstavljajo spremembe v določenem obdobju, ne v nekem točno določenem trenutku. Kot odvisna spremenljivka največkrat nastopa prebivalstvo ali zaposlitev, neodvisne pa so izbrane glede na predvidevanja o vplivih na odvisno spremenljivko s ciljem čim bolj razložiti razlike med regijami in upoštevajo različne parametre, od dostopnosti do prebivalstva in zaposlitve do površine, ravni zaposlenosti in prebivalstva v prejšnjem obdobju, dostopnosti javnih storitev. Koeficiente ocenimo z umerjanjem, najpogosteje z uporabo metode najmanjših kvadratov.

Linearna regresija je zelo razvita metoda, ki se široko uporablja na različnih področjih. Omogoča dobre napovedi, kadar gre za neke splošne vzorce rasti. Obstaja tudi cela vrsta računalniških programov, ki omogočajo izračun. Kljub temu ima metoda pri modeliranju urbanega sistema vrsto pomanjkljivosti:

- Veliko povezav v prostoru ni linearnih, zato je determinacijski koeficient pogosto nizek.
- Pogosto se srečujemo s problemi multikolinearnosti in avtokorelacije.
- Linearni model deluje dobro, če porazdelitev aktivnosti med conami ustreza nekemu splošnemu vzorcu, medtem ko so koeficienti v enačbi zelo občutljivi za posamezna območja v regiji, posebej če ta zelo odstopajo. Linearni model je tudi občutljiv na probleme v zasnovi sistema območij.
- Predpostavljamo tudi, da koeficienti, ki jih dobimo z oceno tj. povezave med spremenljivkami, veljajo tudi v prihodnosti. To drži za regije, za katere je značilen nek splošen vzorec rasti, ne drži pa za regije s velikimi nihanji v spremembah, za katere zato lahko dobimo zelo nerealistične napovedi.

Linearni modeli so zgolj alokacijski modeli, ki prikažejo, kako se prostorske dejavnosti porazdeljujejo po območjih, ne prikažejo pa interakcije med temi dejavnostmi. To lahko dosežemo samo z razvojem novega modela ali uporabo rezultatov linearne regresije kot vhodnih podatkov v gravitacijski model (Foot, 1981).

Razširjeni gravitacijski modeli to pomanjkljivost odpravljajo, vendar razmerja med odvisno in neodvisnimi spremenljivkami niso linearni. Wang je testiral različne oblike funkcij in ugotovil, da eksponentna funkcija najbolje opiše odnose v modelih, ki temeljijo na gravitacijskih modelih (Wang & Guldman, 1996).

2.4.1 Prostorski regresijski modeli

S standardno linearno regresijo želimo najti linearen odnos med odvisno spremenljivko in nizom neodvisnih spremenljivk, pri tem pa predpostavljamo, da se slučajna spremenljivka porazdeljuje normalno okoli povprečne vrednosti 0 ter da napake oziroma ostanki med seboj niso odvisni (homoskedastičnost). Prostorski podatki v splošnem niso neodvisni, zato običajni linearni regresijski modeli niso najbolj primerni za analizo. Regresijski modeli, ki poskušajo vključiti prostorsko odvisnost, se imenujejo prostorski regresijski modeli.

Prvi prostorski regresijski modeli sicer niso bili lokalni, priznavajo pa lokalno naravo prostorskih podatkov s sprostitevjo predpostavke, da so napake neodvisne. Kadar so ta opazovanja povezana z lokacijami v prostoru, se predpostavlja, da so napake za opazovanja, ki so si v prostoru blizu, med seboj odvisne. Vektor ostankov ε naj bi se porazdeljeval z Gaussovo porazdelitvijo okoli ničelne povprečne vrednosti, variančno-kovariančna matrika pa naj bi vsebovala neničelne diagonalne vrednosti. To pomeni, da bo pogojna porazdelitev odvisna od vrednosti napak za opazovanja v bližini, čeprav ima posamezen ε_i porazdelitev centrirano na 0. Gre še vedno za globalne ocene parametrov, so pa lokalni odnosi vpeljeni v model preko kovariančne strukture napak (Fotheringham et al., 2002).

2.4.2 Geografsko tehtana regresija

Geografsko tehtana regresija je novejša metoda modeliranja za lokalne prostorske analize. V nasprotju z globalnimi omogoča lokalno obtežene variacije in povezave. Osnovana je na trditvi, da povezava med spremenljivkami, ki so merjene različnih lokacijah, ni konstantna po prostoru. Podatki so obteženi okoli točke v prostoru tako, da je sosednji podatek bolj obtežen kot ostali, bolj oddaljeni podatki. Na ta način se za vsako točko v prostoru naredijo različne ocene parametra.

Običajni regresijski model lahko zapišemo kot: $y_i = a_0 + \sum_k a_k x_{ik} + \varepsilon_i$. Parametri so ocenjeni za razmerje med odvisno in neodvisnimi spremenljivkami, razmerje pa je v proučevanem območju konstantno.

Geografsko tehtana regresija je relativno preprosta metoda, ki tradicionalno regresijsko zgradbo razširi, tako da dovoljuje oceno tako lokalnih kot globalnih parametrov. Model dobi naslednjo obliko:

$$y_i = a_0(u_i, v_i) + \sum_k a_k(u_i, v_i) x_{ik} + \varepsilon_i \quad (19)$$

Kjer (u_i, v_i) pomeni koordinate točke i v prostoru, $a_k(u_i, v_i)$ pa je vrednost zvezne funkcije $a_k(u, v)$ v točki i .

Funkcijo lahko ocenimo z metodo najmanjših kvadratov in tako dobimo ocene koeficientov na lokaciji (u, v) in napovedano vrednost ter tudi standardne ocene regresije, kot je

determinacijski koeficient. To dosežemo z implementacijo sheme geografskega tehtanja. Shema je postavljena tako, da se da podatkom, ki so bližje lokaciji (u,v), večja utež kot podatkom, ki so bolj oddaljeni. Geografsko tehtana regresija se razlikuje od metode tehtanih najmanjših kvadratov v tem, da je matrika uteži pri slednji konstantna, pri geografsko tehtani regresiji pa se uteži spreminjajo z i.

Uteži lahko opredelimo kot zvezno funkcijo d_{ij} , ki je razdalja med i in j. Diagonalne elemente matrike uteži tako opredelimo kot $w_{ij} = \exp(-d_{ij}^2/h^2)$, kjer se h označuje kot pasovna širina.

2.5 Optimizacijski modeli

Zgoraj opisani modeli opisujejo in napovedujejo razvoj regij in interakcije med njimi. Optimizacijski modeli obenem dosežejo še optimalno rešitev.

Razvoj neke regije v naslednjih desetih letih lahko ubere zelo različne poti. Optimizacijski modeli upoštevajo te možnosti in določijo optimalen potek ob upoštevanju določenih omejitev. Optimizacijski modeli dajo zelo dobre informacije o tem, katere politike je treba izvajati, da bi dosegli neko raven razvoja ob najnižjih možnih stroških. To ne pomeni, da se bo te politike tudi izvedlo, saj je potrebno v optimizacijski model vključiti celo vrsto dejavnikov, ki jih je treba upoštevati. So odlični za primerjavo različnih politik.

Optimizacijski modeli temeljijo na matematičnem, še posebej linearnem programiranju. Optimizacijski postopek vključuje izbiro ciljne funkcije, ki se jo maksimira glede na vrsto omejitev. Ciljna funkcija je lahko opredeljena tako, da poišče porazdelitev novih dejavnosti na proučevanem območju, tako da minimizira celotne stroške razvoja v regiji, minimizira celotne stroške potovanja na delo, maksimira družbeno korist itd.

Enačbe omejitev prikazujejo omejitve za razvoj v enotah in v regiji in lahko vključujejo minimalno in maksimalno stopnjo razvoja za vsako prostorsko enoto, okoljske omejitve za razvoj, omejitve glede javnih storitev, industrijskega in poselitvenega razvoja v enotah ter omejitve dovoljene interakcije med določenimi prostorskimi enotami. Za optimizacijske modele so potrebni zelo podrobni podatki, še posebej o stroških, ki jih je pogosto težko pridobiti.

Matematično programiranje je bilo uporabljeno pri proučevanju širokega spektra planskih problemov glede rabe prostora in razvoja transporta, pri proučevanju problemov ravnanja z odpadki, gospodarjenja z vodnimi viri ter umestitve bolnišnic in šol v prostor.

Optimizacijski modeli so bili uporabljeni za različne namene, npr. v transportnih modelih za primerjavo dejanske porazdelitve neke aktivnosti glede na optimalno, raziskave o obiskovanju šol in bolnišnic, kjer se je minimiziral potovalni čas. Modeli se uporabljajo za primerjavo razlik ali izbiro lokacij za nove zmogljivosti. Takšna uporaba modelov z eno spremenljivko, ki jo optimiziramo, je lahko zelo uspešna, če so predpostavke modela jasne.

Obsežni optimizacijski modeli, ki temeljijo na kompleksnem sistemu enačb, pa se ne morejo uporabljati za napovedovanje prihodnjega razvoja tako kot regresijski in gravitacijski modeli. Model tako ni namenjen ocenjevanju vzorcev v prihodnosti, je pa zelo primeren pri proučevanju razvojnih usmeritev, ki najbolje dosegajo neke cilje, npr. maksimirajo koristi regije. Modeli se lahko uporabljajo tudi za proučevanje alternativnih razvojnih strategij. Optimizacijski modeli so zahtevni tako glede same opredelitve modela in ciljne funkcije kot velikih zahtev po podatkih.

2.6 Scenariji

Metoda scenarijev je način za obravnavo različnih možnih prihodnosti na opisni način in ima značaj in obliko pripovedi. Osredotoča se na vzročne procese in točke odločitev, zato je še posebej uporabna za odločevalce na različnih ravneh. Različni scenariji pomagajo izbrati primerno pot tako, da predstavijo različna možna sosledja dogodkov. Možne prihodnosti opisujejo skozi razvoj vzročno-posledičnih korakov. Kadar se uporabljajo za potrebe planiranja ali politik so scenariji skice prihodnosti, ki temeljijo na shematskih opisih določenih ključnih spremenljivk.

Za razliko od drugih metod napovedovanja scenariji ne predvidevajo ene same prihodnosti, temveč predstavijo različne možne prihodnosti, ki so odvisne od tega, koliko lahko različne spremenljivke vplivajo druga na drugo. Prednost metode scenarijev je večdimenzionalnost, saj dopušča uporabo poljubnega števila spremenljivk. Metoda se tudi izogne problemu izrecnega napovedovanja prihodnosti, saj zanjo predvideva več možnih opcij. Zavedati pa se je treba, da scenariji niso objektivne narave ter da podlegajo omejitvam uporabljenih metod ali ekspertnih mnenj.

Za oblikovanje scenarijev je mogoče uporabiti različne metode. V grobem lahko razločimo dva osnovna tipa metod, to so:

- kvantitativne metode ter
- metode temelječe na ekspertnih mnenjih.

Prve uporabljajo statistiko in računalniške modele ter tudi kvantificirane reprezentacije kvalitativnih informacij, kakršna je npr. metoda Delphi vprašalnikov, pri katerih strokovnjaki z vrednostmi ocenijo različna dogajanja in pojave. Druge temeljijo izključno na poglobljenem ekspertnem mnenju in tehnikah izmenjave mnenj. V praksi je večina scenarijev kombinacija obeh osnovnih tipov metod.

Končni scenariji se po značaju lahko bistveno razlikujejo. V okviru projekta ESPON 3.2, ki je obravnaval različne prostorske vizije Evrope, je bila predlagana sledeča tipologija scenarijev:

- Vnaprej (iz sedanjosti v prihodnost):
 - Tip 1: temeljni ali trend scenariji; slonijo na predpostavki, da bodo strategije in politike javnih in zasebnih akterjev ostale bolj ali manj nespremenjene,
 - Tip 2: pričakovani scenariji; slonijo na predpostavki, da se bodo zgodile pomembne spremembe v obnašanju in aktivnostih vpletenih akterjev (posameznikov, gospodinjstev, podjetij), medtem ko do pomembnih sprememb javnih politik ne bo prišlo,
 - Tip 3: scenariji pričakovanih politik; slonijo na predpostavki, da se bodo glavne spremembe zgodile v polju javnih politik.
- Od zadaj (iz prihodnosti v sedanjost):

- Tip 4: proaktivni scenariji; izhajajo iz prihodnjih situacij in ugotavljajo, kako bi se lahko razvile iz sedanjega stanja in trendov, kaj bi se lahko doseglo in čemu bi se bilo dobro izogniti v okviru danih omejitev, virov ali tehnologij (pri čemer so strategije javnih in zasebnih akterjev precej neodvisne); ti scenariji običajno nastajajo na interaktivni način, kjer vsaj enakovredno sodelujejo tudi odločevalci in drugi deležniki.

3 INFORMACIJSKI SISTEM ZA MODELIRANJE FUNKCIONALNIH REGIJ

Cilj dela je zasnova informacijskega sistema za podporo odločanju pri oblikovanju funkcionalnih regij na različnih ravneh, s katerim bo mogoče na podlagi razpoložljivih podatkov in informacij, pridobljenih iz različnih virov, ovrednotiti razvojne potenciale regij. Informacijski sistem mora omogočiti modeliranje funkcionalnih regij na podlagi postavljenih ciljev (trajnostni razvoj, zmanjševanje razvojnih razlik med regijami, ...) in ugotovljenih razvojnih potencialov.

Načrtovanje in izgradnja informacijskih sistemov je praviloma zahteven proces, ki poteka skozi celo vrsto razvojnih korakov. Razdelitev in opredelitev teh korakov je v strokovni literaturi pogosto različno obravnavana. Ne glede na različne opredelitve pa lahko po začetni ideji ugotovimo naslednje korake (Davis & Yen, 1998):

- definicija naloge/opredelitev problema,
- analiza in opredelitev funkcionalnih zahtev,
- načrtovanje,
- izgradnja,
- uvedba,
- preverjanje rešitev,
- vzdrževanje.

Problem je bil vsebinsko opredeljen v prvih dveh poglavjih, kjer je opisano, kaj so funkcionalne regije in kako se opredeljujejo, ter nato obravnavani še interakcijski modeli, ki se lahko uporabljajo za modeliranje regij. V nadaljevanju je opisan sistem za podporo odločanju in funkcionalne zahteve sistema za modeliranje regij.

3.1 SISTEMI ZA PODPORO ODLOČANJU

Sistem za podporo odločanju je vrsta informacijskega sistema, ki podpira odločanje (lahko vključuje računalniški sistem, vendar ni nujno omejen nanj). Dobro zasnovan informacijski sistem je interaktiven, na računalniških programskih orodjih zasnovan sistem, ki na podlagi podatkov, dokumentov, znanja ter modelov pomaga pridobiti koristne informacije, s tem pa prepoznati probleme in vodi do učinkovitejšega odločanja.

Hkrati z razvojem sistemov za podporo odločanju so se pojavili t. i. prostorski sistemi za podporo odločanju, ki vključujejo geografski informacijski sistem (Wikipedia, 2010).

Ključne komponente sistema za odločanje so:

- sistem za upravljanje podatkovnih baz, ki omogoča tudi hrambo in obdelavo geografskih podatkov,
- knjižnica potencialnih modelov, ki se uporabljajo za napovedovanje možnih izidov odločitev,
- vmesnik, ki omogoča interakcijo uporabnika z računalniškim sistemom in analizo izidov.

3.1.1 Funkcionalne zahteve sistema za podporo odločanju

Namen specifikacije zahtev je jasna opredelitev zahtev oziroma potreb, in sicer tako, da je mogoče preveriti, ali so zahteve tudi izpolnjene. Običajno je specifikacija zahtev poseben dokument, ki zagotavlja, da so potrebe ustrezno opredeljene, zato da ne bi prišlo do vložkov v napačno rešitev. Običajno je specifikacija zahtev zadnji korak v fazi analize razvoja informacijskega sistema (Davis & Yen, 1998).

Zahteve za sistem za podporo odločanju za oblikovanje funkcionalnih regij so:

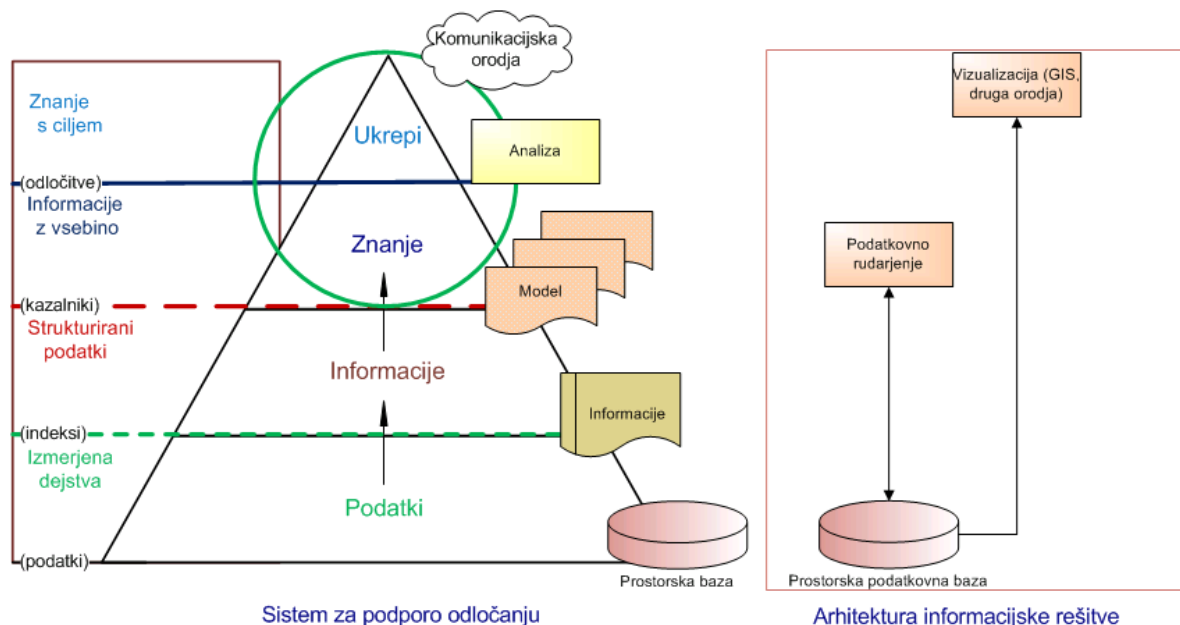
- Informacijski sistem mora zagotoviti hrambo razpoložljivih podatkov in informacij, pridobljenih iz različnih virov v različnih obdobjih, ter obdelavo in shranjevanje podatkov, ki se spreminjajo na podlagi različnih razvojnih scenarijev.
- Podatki in kazalniki morajo biti ločeni po posameznih vsebinskih sklopih, na primer gospodarstvo, infrastruktura, človeški in naravni viri.
- Informacijski sistem mora biti razširljiv v smislu novih podatkov in s tem uporaben tudi v prihodnosti, da bo vzdrževanje podatkovne baze neproblematično.
- Uporaben mora biti na način, da uporabniki lahko hitro vidijo, kateri podatki so na voljo, in da lahko enostavno dostopajo do njih. Hkrati mora biti struktura baze koherentna, dodajanje ali posodabljanje podatkov pa enostavno.
- Podatkovni model mora omogočati tudi spremljanje kakovosti podatkov, ki jih vsebuje.
- Izgradnja modelov, njihovo testiranje, dokumentiranje in uporaba morajo biti enostavni, stroški razvoja novih modelov pa čim nižji.
- Obstoječe modele, ki so bili že uporabljeni, naj bi bilo možno poizvedovati, pregledovati in ponovno uporabiti.
- Eksperimentiranje in prototipiranje mora biti enostavno in hitro, prav tako prehod v produkcijo in zaganjanje modelov za izdelavo napovedi.
- Komunikacija med uporabnikom in sistemom mora potekati preko interaktivnega uporabniškega vmesnika, ki omogoča pregledovanje podatkov, kode ipd.
- Informacijski sistem mora biti fleksibilen, tako da omogoča eksperimentiranje z različnimi podatki, algoritmi oziroma statističnimi metodami in predstavitvami ter vizualizacijo rezultatov.

- Omogočati mora enostaven pregled vhodnih, vmesnih podatkov in s tem že vnaprejšnje diagnosticiranje problemov.
- Biti mora razširljiv, tako da uporabnikom omogoča hitro spreminjanje obnašanja modelov bodisi z vključevanjem novih spremenljivk bodisi izgradnjo novih modelov iz obstoječih delov in uporabo/aplikacijo izdelanih modelov na podatkih različnih scenarijev.
- Integracija mora biti enostavna: informacijski sistem naj bo združljiv z drugimi sistemi, ki bi lahko dopolnili obstoječo rešitev, npr. ocenjevanje, prikazovanje podatkov, shranjevanje podatkov, integracija z GIS itd.
- Sistem mora biti sposoben procesirati različne podatke in velike količine podatkov.
- Informacijski sistem naj omogoči dopolnjevanje in sodelovanje uporabnikov, tako da lahko v vsakem trenutku vidijo delo sodelavcev in ga dopolnjujejo ali uporabijo izgrajene modele.
- Informacijski sistem mora zagotoviti grafični prikaz prostorskih podatkov za lažjo predstavitev geografskih vsebin.

3.1.2 Oris informacijske rešitve

Sistem za podporo odločanju sem zasnovala na podatkovnem rudarjenju, ki črpa podatke iz relacijske podatkovne baze, ki omogoča tudi hrambo prostorskih podatkov. Uporabnik lahko dostopa do podatkov in jih obdeluje neposredno preko orodja za poizvedovanje v bazi, lahko pa dostopa tudi preko grafičnega vmesnika za podatkovno rudarjenje, ki omogoča tudi interakcijo uporabnika s sistemom in jasn prikaz rezultatov. Prostorske podatke nato prikažemo z izdelavo zemljevidov v za to namenjenih orodjih ali pa podatkovno bazo povežemo z GIS orodji.

Slika 8: Sistem za podporo odločanju in arhitektura informacijske rešitve



Vir: Prirejeno po Mcrit, *What's an executive support system?*, 2009

V podpoglavjih, ki sledita, sta opisani ključni komponenti informacijskega sistema, tj. prostorska podatkovna baza in podatkovno rudarjenje.

3.2 Sistem za upravljanje podatkovnih baz

Sistem za upravljanje podatkovnih baz je sestavljen iz programov, ki upravljajo podatkovno bazo, zagotavljajo hrambo, dostop, varnost, varnostno kopiranje in druge funkcionalnosti. Lahko jih ločimo glede na podatkovni model, ki ga podpirajo, na primer relacijske ali XML, ali poizvedovalne jezike, s katerimi se dostopa do podatkovne baze itd.

Večina podatkovnih baz danes podpira relacijski podatkovni model. Koncept relacijske podatkovne baze je prvi predlagal Codd v 70. letih prejšnjega stoletja, da bi zmanjšal rigidnost hierarhičnih in mrežnih sistemov ter ločil uporabnike od fizične implementacije sistema. Osrednji del relacijskega sistema za upravljanje podatkovnih baz je relacija oziroma tabela z nizom podatkov, ki spadajo v določeno vsebinsko področje neke organizacije. Zaradi hitrega napredka koncepta in metod relacijskih podatkovnih baz so relacijske podatkovne baze hitro postale standard v svetu podatkovnih baz.

Relacijska podatkovna baza ima torej podatke shranjene v tabelah, ki so sestavljene iz vrstic in stolpcev. Vrstica ustreza zapisu, stolpec pa polju v zapisu. Vsak stolpec ima določen podatkovni tip. Nabor podatkovnih tipov je običajno zelo majhen, npr. celo število, decimalno število, niz, datum in znak. Vsako polje v zapisu lahko vsebuje eno samo vrednost. Povezave niso eksplicitne, ampak so določene z vrednostmi polj tujih ključev v prvi tabeli, ki ustrezajo zapisom v drugi tabeli. Povezave mnogo : mnogo zahtevajo vmesno tabelo s podatki, ki določajo povezavo (Ramakrishnan & Gehrke, 2002).

3.2.1 Prostorske podatkovne baze

Programska orodja za upravljanje prostorskih informacij so bila tradicionalno znana kot geografski informacijski sistemi (GIS). Gre za specializirane aplikacije za hrambo, obdelavo, analizo in prikaz prostorskih podatkov, ki so se uporabljale pri načrtovanju rabe tal, v geomarketingu, logistiki, prometu ipd. Do nedavno so GIS uporabljali specifične prostorske podatkovne modele in lastne razvojne jezike, kar jih je ločilo od večine (poslovnih) podatkovnih baz. To je bila ovira, saj ni bilo mogoče doseči popolnega izkoriščanja dodane vrednosti prostorskih podatkov v organizacijah.

Ko se je povečala uporaba GIS v podjetjih in javnem sektorju, so njihove omejitve postale še bolj očitne. Zaradi različnih, pogosto nekompatibilnih standardov za shranjevanje prostorskih podatkov in različnih programskih jezikov in vmesnikov so se problemi pogosto reševali kar z izvozi podatkov iz enega v drug sistem.

Koncepti in metode, na katerih temeljijo prostorske baze podatkov, so danes podobne kot podatkovne baze, ki so jih razvili v preteklem desetletju. Dodane so bile prostorske

informacije, ki so bile prej razumljene kot samostojna tehnologija. Iz GIS aplikacij, vezanih na specifične projekte, so se razvile podatkovne baze, ki so s pokrivanjem dnevnih potreb po informacijah dobra podpora pri odločanju.

Prostorski podatkovni sistemi se od običajnih podatkovnih baz razlikujejo v dveh pomembnih vidikih. Prvi vidik je zahteva po hrambi kompleksnih podatkovnih tipov, kot so točke, linije in poligoni, drugi pa je potrebna funkcionalnost za obdelavo teh kompleksnih podatkov, in sicer uporaba prostorskih operatorjev, ki so veliko bolj napredni kot operatorji za obdelavo alfanumeričnih podatkovnih tipov (Yeung & Hall, 2007).

Oracle Spatial ima pomembno vlogo pri uveljavitvi prostorskih podatkovnih baz, saj se podatki, ki so shranjeni v Oraclovi podatkovni bazi, lahko poizvedujejo, obdelujejo in povezujejo z ostalimi, neprostorskimi podatki. Prostorski podatek oziroma lokacija je samo eden od atributov nekega objekta. To odstrani potrebo po usklajevanju podatkov iz različnih virov in uporabo različnih poizvedbenih jezikov. Pomembna lastnost Oracle Spatial je, da podpira standarde za prostorske informacije (SQL in Open Geospatial). Oraclova baza je bila tudi prva, ki je bila integrirana s Spatial Database Engine (SDE) in ArcView GIS podjetja ESRI ter kasneje tudi z drugimi produkti kot sta Intergraph in MapInfo. Kasneje so temu zgledu sledili tudi drugi ponudniki podatkovnih baz (dopolni) (Kothuri, Godfried & Beinat, 2007).

Kot sistem za upravljanje baz podatkov in izgradnjo podatkovne baze sem izbrala Oraclovo bazo 11g, ki omogoča prostorske podatke oziroma Oracle Spatial.

3.2.2 Upravljanje in analiza prostorskih podatkov

Med prostorske funkcionalnosti med drugim spadajo:

- hramba prostorskih podatkov
 - hramba prostorskih podatkov, kot so točke, linije, poligoni ter druge oblike vektorskih podatkov v ustrezni obliki (tip Geometry) v podatkovni bazi;
 - vrivanje, brisanje in posodabljanje teh vrst prostorskih podatkov v bazi;
- analiza vektorskih prostorskih podatkov – običajno sem spadajo naslednje analitične funkcije:
 - Within-distance: poišče vse prostorske podatke znotraj določene razdalje do neke lokacije, ki jo poizvedujemo;
 - Contains: poišče vse prostorske podatke, ki vsebujejo določeno lokacijo (geometry) (funkcije, ki poiščejo druge vrste povezav se tudi lahko opredelijo);
 - Nearest-neighbor: poišče vse prostorske podatke, ki so najbližje neki lokaciji;
 - Distance: izračuna razdaljo med dvema prostorskima objektoma;
 - Buffer: zgradi območje okrog prostorskih podatkov;
 - Overlay: naloži/prekrije različne ravni prostorskih podatkov;
 - Visualization: predstavi prostorske podatke z uporabo zemljevidov.

- analiza omrežij: običajno so lahko prostorski podatki, kot je na primer cestno omrežje, predstavljeni kot mrežni podatki (poleg vektorskih podatkov); pri analizi teh podatkov potem raje uporabljamo razdaljo po omrežju kot prostorsko bližino.

Primer: Izračun centroidov in razdalje med centroidoma občin Kranj in Ljubljana s funkcijama SDO_CENTROID in SDO_DISTANCE:

```
SELECT ORIGIN.ob_ime  IZVOR
      , DESTINATION.ob_ime  PONOR
      , SDO_GEOM.SDO_DISTANCE(SDO_GEOM.SDO_CENTROID(ORIGIN.geom, M.diminfo)
                              , SDO_GEOM.SDO_CENTROID(DESTINATION.geom, M.diminfo)
                              , 0.5)  EVKLIDSKA_RAZDALJA

FROM OB ORIGIN
      , OB DESTINATION
      , USER_SDO_GEOM_METADATA M
WHERE ORIGIN.ob_uime = 'Kranj'
AND DESTINATION.ob_uime = 'Ljubljana'
AND M.table_name = 'OB';
```

IZVOR	PONOR	EVKLIDSKA_RAZDALJA
KRANJ	LJUBLJANA	28463.8605963489

3.2.3 Shranjevanje prostorskih podatkov v bazi

Če vzamemo vektorske podatke, navadno ločimo med:

- točkami , katerih opis zahteva le x in y koordinate (ali x,y, z v primeru 3D),
- linijami (na primer ceste), katerih opis zahteva začetno, končno in nekaj vmesnih koordinat;
- poligoni (na primer stanovanjsko naselje), ki ga opišemo s sklenjeno linijo.

Vektorski podatki se lahko shranjujejo v eni ali več tabelah. Najboljši način je opredelitev podatkovnih ravni, niza podatkov s skupnimi značilnostmi, ki se shranijo v tabeli. Večina prostorskih baz uporablja poseben tip podatka za shranjevanje prostorskih podatkov, ki ga označujejo z *geometry*. Uporabniki lahko dodajo polje tipa *geometry* v tabelo, da bi lahko shranili prostorske podatke (Kothuri et al., 2007).

3.2.4 Vizualizacija prostorskih podatkov in Oracle MapViewer

Pri prostorskih podatkih je pomembna tudi vizualizacija podatkov z uporabo zemljevidov. Vizualizacijo prostorskih podatkov je mogoča s standardnimi GIS-sistemi, ki jih povežemo s prostorsko podatkovno bazo.

Za prikaz podatkov na zemljevidu bom v magistrskem delu uporabila orodje Oracle MapViewer, ki zgradi zemljevide na podlagi podatkov, shranjenih v tabelah in pogledih v podatkovni bazi, in jih nato vrne v ustrezni obliki v odjemalcu. MapViewer dopolnjuje zmogljivosti upravljanja podatkov v Oraclovi bazi, saj lahko z njim preko spleta dostopamo in pregledujemo prostorske podatke. Omogoča analizo in razumevanje geografskih značilnosti

problema iz podatkov, ki so shranjeni v podatkovni bazi ali skladišču, saj jih naredi dostopne v aplikacijah, ki so enostavne za uporabo. (Oracle, 2009)

Vsak zemljevid v MapViewerju je opredeljen z eno ali več ravnmi oziroma temami (angl. *themes*). Vsaka tema predstavlja logično skupino prostorskih značilnosti, na primer cest, lokacij kupcev, rek ipd. Te značilnosti so prikazane s specifičnimi stili (angl. *styles*).

3.2.5 Načrtovanje in izgradnja podatkovne baze

Dobro zasnovana podatkovna baza omogoča dostop do ažurnih in točnih informacij. Ker je dober dizajn ključen za doseganje ciljev pri delu s podatkovno bazo, je treba v načrtovanje vložiti kar nekaj časa, s tem pa dobimo podatkovno bazo, ki ustreza našim potrebam.

Za dobro zasnovano podatkovno bazo velja, da:

- ima informacije vsebinsko ločene v tabele in zmanjšuje možnost podvajanja podatkov,
- omogoča dostop do želenih informacij in njihovo združevanje,
- podpira in zagotavlja točnost in integriteto informacij,
- olajša obdelavo podatkov in ustreza potrebam poročanja.

Pri izdelavi modela relacijske podatkovne baze sem upoštevala namen in cilje, ki naj jih doseže izgradnja baze oziroma informacijski sistem. Pri tem sem upoštevala, da mora biti podatkovni model zasnovan tako, da bo razširljiv v smislu novih podatkov in uporaben, torej v skladu s funkcionalnimi zahtevami.

Načrtovanje podatkovne baze, ki je znano tudi kot model entitet in povezav (angl. *Entity-Relationship Model*) omogoča postavitvev podatkovnega modela na organiziran in relativno enostaven način. Elementi modela so entiteta, atribut in relacija. Načrtovanje podatkovne baze po določitvi njenega namena obsega naslednje korake:

- izbor in organizacija vseh informacij, ki jih želimo hraniti v podatkovni bazi,
- razdelitev informacij v entitete oziroma tabele (vsaka entiteta postane tabela),
- določitev informacij po tabelah (vsaka informacija postane polje in je predstavljena s stolpcem v tabeli),
- postavitvev povezav med entitetami,
- izboljšave dizajna (pregled možnih napak/testiranje s testnimi podatki/popravki),
- uvedba pravil normalizacije za pregled pravilnosti strukture in izvedba popravkov, če je to potrebno. (Database design basics)

Rezultat uporabljene metode je diagram entitet in povezav, ki prikazuje model podatkov: povezane entitete, ki vsebujejo po več atributov.

Diagram je izdelan v orodju Toad Data Modeler podjetja Quest, ki je namenjeno načrtovanju podatkovnih baz in ki omogoča uporabnikom, da izdelajo, vzdržujejo in dokumentirajo nov ali obstoječ sistem za upravljanje podatkovnih baz.

Slika 9: E-R diagram



Osrednja tabela je tabela Spremenljivka, ki opisuje podatek. Spremenljivka ima šifro, ki je ID spremenljivke in je primarni ključ, naziv in opis ter naslednje attribute, ki so zaradi normalizacije v ločenih tabelah in sestavljajo primarni ključ: šifra prostorske ravni, šifra vira podatka, leto oziroma obdobje za katero je podatke zbran in šifra vsebinskega podpodročja. Dodatno jo opcijsko lahko opišemo s tipom podatka, ki je npr. grobi podatek ali izračunani kazalnik.

Kot primer sta navedeni tabeli POPULACIJA in DNEVNE MIGRACIJE. Dnevne migracije imajo ključ sestavljen iz enote izvora in enote ponora. Praviloma bi morali narediti vmesno

tabelo s kombinacijami mnogo:mного, vendar to tukaj ni potrebno, saj kombinacije niso stalne, ker se spreminjajo zaradi spreminjanja števila enot.

Prostorska raven

Prostorska raven je določena s šifro, nazivom in obdobjem veljavnosti ter nadrejeno ravno. Je atribut spremenljivke.

Podatki o posamezni prostorski ravni so shranjeni v posebni tabeli (primer: OB), katere ime je unikatno in zapisano v polju Ime tabele v tabeli Prostorske ravni. Primer prostorske ravni so Občine, ki imajo nadrejeno raven Statistične regije. Občine so omejene z veljavnostjo, saj smo priča nastajanju vedno novih občin. Enota je določena s prostorskimi podatki, tip podatka je SDO_Geometry, ki prostorsko določa posamezno občino (poligon).

Vir podatka nam pove, kje smo pridobili podatek, na primer SURS, GURS ali kakšen drug vir.

Obdobje določa časovni okvir podatkov.

Vsebina

Vsebinska področja in podpodročja se uporabljajo za ločevanje podatkov po vsebinskih sklopih. Tabeli vsebujeta šifro in naziv področja in podpodročja. Vsako podpodročje ima eno nadrejeno področje. Spremenljivko zato označuje podpodročje.

3.3 Podatkovno rudarjenje

Podatkovno rudarjenje je proces pridobivanja zanimivih in pogosto skritih informacij iz velike množice prej neznanih podatkov, ki so običajno shranjeni v podatkovni bazi ali podatkovnem skladišču in ki jih tradicionalne statistične metode zaradi velikega števila atributov, kompleksnosti povezav ali težav pri sami izvedbi analize težko zaobjamejo.

Podatkovno rudarjenje je bilo prvotno razvito za uporabo na podatkovnih skladiščih, danes pa je vključeno v vse več komercialnih produktov za upravljanje podatkovnih baz (kot so Oracle, IBM-ov BD2 in Microsoftov SQL Server). Podatkovno rudarjenje je tako danes sestavni del modernih sistemov za upravljanje podatkovnih baz in igra ključno vlogo pri razvoju podatkovnih baz od enostavnih orodij za upravljanje podatkov do podpore odločanju.

Nova generacija podatkovnih baz so integrirani in inteligentni sistemi za upravljanje podatkovnih baz, ki kombinirajo upravljanje podatkov in funkcije za podporo odločanju v istem sistemu (Yeung & Hall, 2007).

Podatkovno rudarjenje se kot nova metoda pridobivanja informacij razlikuje od običajnih SQL in OLAP metod v nekaj pomembnih točkah, med drugim:

- Oblikovano je za uporabo na zelo velikih podatkovnih bazah in podatkovnih skladiščih z milijoni zapisov in tisoči atributov, ki bi jih težko analizirali z običajnimi metodami.
- S sekundarno analizo velikih podatkovnih nizov poskuša odkriti prej neznano ali 'skrito' znanje, vključno z nepredvidljivimi vzorci, povezavami in trendi, ki jih je zaradi količine in kompleksnosti podatkov drugače težko odkriti.
- Sledi induktivni strategiji analize podatkov, pri kateri uporabniki za pridobivanje novega znanja uporabljajo algoritme strojnega učenja brez vnaprejšnjih predpostavk ali hipotez, kakšni bodo rezultati na koncu.
- Osredotoča se na odkrivanje značilnosti in povezav v velikem številu atributov, vendar pa dokončno ne identificira tudi vzrokov in posledic med posameznimi atributi. Z drugimi besedami, je podatkovno rudarjenje po svoji naravi raziskovalno in nedeterministično, ne pa pojasnjevalno in deterministično.

Podatkovno rudarjenje je tako več kot le razumevanje preteklosti na podlagi prvotnih podatkov. Z uporabo algoritmov strojnega učenja poskuša odkriti skrite informacije ali zamegljeno kompleksnost v podatkih. Rezultati podatkovnega rudarjenja niso samo odsev preteklosti, ampak so tudi vpogled v prihodnost. Sposobnost povezave preteklih scenarijev s tistim, kar se bo mogoče zgodilo v prihodnosti, je ključno pri podpori strateških odločitev v organizacijah v javnem in zasebnem sektorju (Yeung & Hall, 2007).

3.3.1 Podatkovno rudarjenje in odkrivanje znanja iz podatkov

Projekt podatkovnega rudarjenja navadno zahteva veliko dela pri zbiranju in obdelavi podatkov. Podatkovne tabele so sestavljene iz različnih tipov in virov podatkov. Podatki v 'resničnem svetu' so pogosto neprečiščeni, vsebujejo napačne ali manjkajoče vrednosti, zato je treba podatke pred uporabo prečistiti. Filtrirani, normalizirani, vzorčeni in transformirani podatki se nato uporabijo kot vhod za algoritme podatkovnega rudarjenja. Do 80 odstotkov npora pri projektih podatkovnega rudarjenja je namenjenega prav pripravi podatkov. Ko so podatki shranjeni v tabelah v podatkovni bazi, se podatki lahko obdelajo z uporabo orodij podatkovne baze.

Modeli podatkovnega rudarjenja morajo biti zgrajeni, testirani, uporabljeni in uporabljeni za posamezno vsebinsko področje, prav tako so rezultati pogosto naknadno obdelani glede na vsebino (npr. izračun ocene tveganja, pričakovana koristnost in verjetnost odziva) ter nato shranjeni v stalno podatkovno bazo ali podatkovno skladišče.

Da bi celoten proces podatkovnega rudarjenja deloval na ponovno uporabljiv in zanesljiv način, ni lahko doseči, saj lahko vsebuje avtomatizacijo in prenose med strežniki, podatkovnimi shrambami, aplikacijami in orodji, nekatera orodja za podatkovno rudarjenje na primer zahtevajo izvoz podatkov iz podatkovne baze in pretvorbo v obliko, ki se zahteva za podatkovno rudarjenje, rezultati rudarjenja pa so nato ponovno uvožijo nazaj v bazo.

Odstranitev ali zmanjšanje teh ovir lahko vodi v pogostejšo uporabo rudarjenja, bolj koristne informacije in lahko pomembno vpliva na odločitve. Z rudarjenjem neposredno na podatkovni bazi postane premeščanje podatkov nepotrebno, podatki so ažurni, skrajša pa se tudi čas, potreben za izvedbo rudarjenja. Premeščanje podatkov lahko zmanjša varnost. Če podatki ne zapuščajo baze, varnost, postavljena na podatkovni bazi, varuje podatke.

Prednosti rudarjenja na podatkovni bazi so (Oracle, 2007):

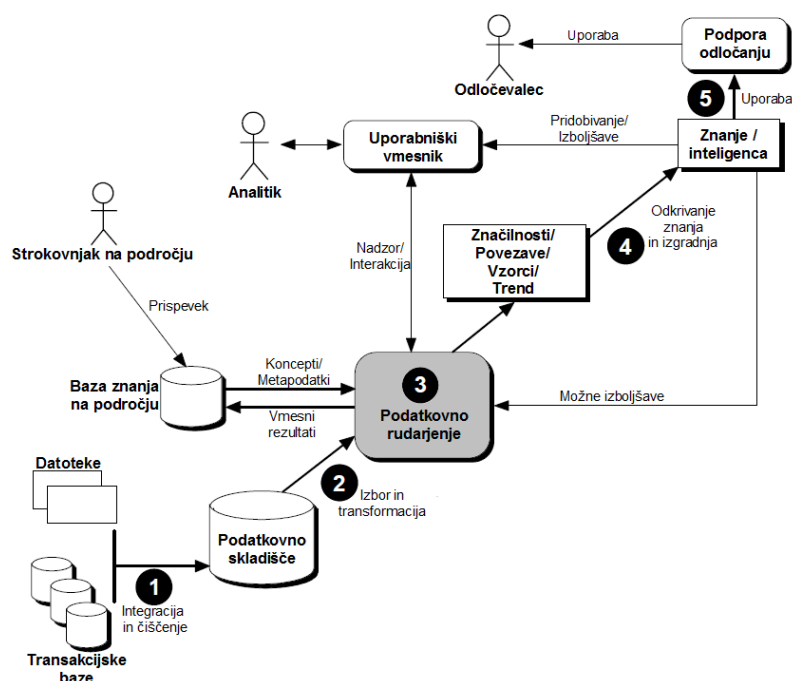
- manj premeščanja podatkov,
- večja varnost,
- ažurni podatki.

Podatkovno rudarjenje se pogosto uporablja kot sinonim za odkrivanje znanja iz podatkov (angl. *Knowledge Discovery in Databases* ali KDD), medtem ko nekateri menijo, da gre zgolj za korak v procesu odkrivanja znanja (Han & Kamber, 2006).

Odkrivanje znanja kot proces je prikazan na sliki 10 in je sestavljeno iz naslednjih korakov:

- **Čiščenje in integracija podatkov:** Odstranjevanje šuma in nekonsistentnih podatkov ter združevanje podatkov iz različnih podatkovnih virov.
- **Izbor in transformacija podatkov:** Iz podatkov v bazi se izberejo podatki, pomembni za analizo. Podatki se nato pretvorijo v obliko, ki je primerna za podatkovno rudarjenje.
- **Podatkovno rudarjenje:** Ključni proces, v katerem z uporabo inteligentnih metod dobimo vzorce iz podatkov.
- **Odkrivanje znanja in izgradnja:** Identifikacija pomembnih in resnično zanimivih vzorcev, ki predstavljajo novo znanje.
- **Uporaba:** Vizualizacija in predstavitvene metode za prikaz odkritega znanja uporabniku.

Slika 10: Koraki odkrivanja znanja v podatkovni bazi



Vir: A.Yeung et al., *Spatial database systems: design, implementation and project management*, 2007, str. 414.

Odkrivanje znanja je interaktiven in iterativen proces. Analitik lahko z uporabo ustreznega uporabniškega vmesnika in s spreminjanjem vhodnih parametrov za oblikovanje različnih scenarijev nadzoruje proces podatkovnega rudarjenja ter z uporabo različnih metod podatkovnega rudarjenja zagotavlja zanesljivost in integriteto rezultatov. Interaktiven in iterativen pristop omogoča neprestano izboljševanje ali optimizacijo pridobljenega znanja s kombiniranjem vsebinskega znanja in znanja pridobljenega iz zaporednih iteracij v procesu. Odkrivanje znanja iz podatkov se konča, ko je analitik zadovoljen in se s ponavljanjem ne da več pridobiti novega znanja.

Zaradi interaktivne in iterativne narave odkrivanja znanja iz podatkov ni vedno lahko določiti, kje se proces začne in kje konča. Han in Kamber (2006) sta predlagala, da se koraki odkrivanja znanja razdelijo v dve funkcionalni fazi. Prva faza vključuje čiščenje, izbor in transformacijo podatkov, ki se jo lahko izvede z izgradnjo podatkovnega skladišča. V drugi fazi se podatkovno rudarjenje in metode vrednotenja integrirajo v en, po možnosti iterativen proces, ki se imenuje podatkovno rudarjenje.

3.3.2 Koncepti in metode podatkovnega rudarjenja

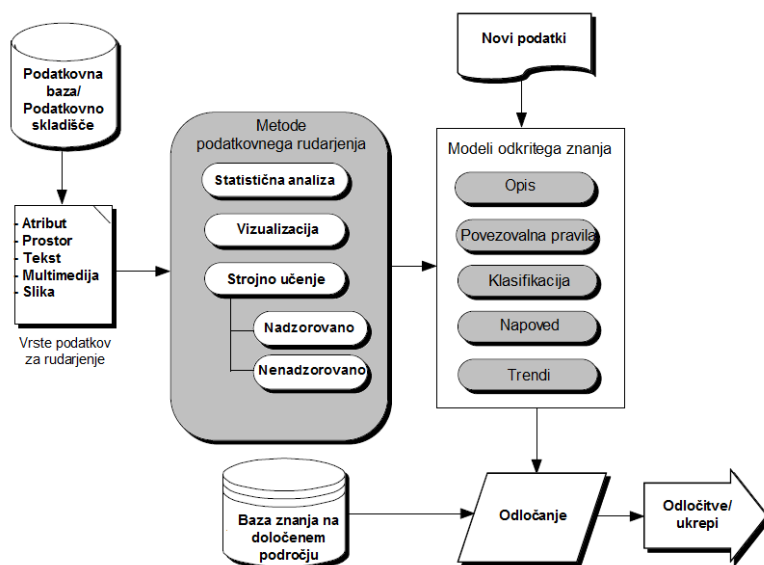
Slika 11 prikazuje podatkovno rudarjenje kot povezavo med podatki, analitičnimi metodami in iz tega izhajajočimi modeli. Med analitičnimi metodami ima še vedno velik pomen multivariantna statistična analiza, ki je tudi najbolj prispevala k razvoju in uporabi metod podatkovnega rudarjenja. Vizualizacija obsega vrsto dvodimenzionalnih, tridimenzionalnih in animiranih grafičnih metod za prikaz matematičnih povezav. Danes vse bolj prevladujejo

metode strojnega učenja, ki uporabljajo algoritme učenja tako, da najprej poiščejo značilnosti na nekem nizu podatkov ter oblikujejo model, s katerim na podlagi novih podatkov dobimo klasifikacijo, vzorec, napoved ali trend.

Strojno učenje lahko razvrstimo v usmerjeno in neusmerjeno strojno učenje, glede na to, koliko je treba posegati v proces. Usmerjeno strojno učenje, ali tudi napovedna analiza, je namenjeno reševanju problemov. Usmerjeno je zato, ker zahteva, da analitik opredeli ciljno spremenljivko in odvisne attribute v podatkovnem nizu, na katerem se izvaja podatkovno rudarjenje. Izbrani algoritem poizkuša iz podatkov izluščiti vzorce in povezave med neodvisno in odvisnimi spremenljivkami. Odkrite vzorce in odnose oblikuje v model odkritega znanja, ki se nato lahko uporabi za napovedovanje vedenja ali značilnosti novih objektov v podatkovnem nizu.

Neusmerjeno strojno učenje, ali tudi opisno podatkovno rudarjenje, je bolj raziskovalne narave in poizkuša odkriti značilnosti v podatkih. Algoritmi neusmerjenega učenja ne postavljajo predpostavk ali hipotez o ciljnem podatkovnem nizu in poizkušajo najti v podatkih povezave, skupine in trende brez vnaprej opredeljenega cilja.

Slika 11: Koncepti in metode podatkovnega rudarjenja



Vir: A. Yeung et al., *Spatial database systems: design, implementation and project management*, 2007, str. 419.

3.3.2.1 Opis podatkov in združevanje

Cilj opisa in združevanja kot metode podatkovnega rudarjenja je natančen opis značilnosti podatkov, običajno v osnovni ali agregirani obliki, s čimer uporabnik dobi vpogled v strukturo podatkov. Včasih je lahko že sam opis ali združevanje podatkov cilj podatkovnega rudarjenja, vendar se običajno uporablja v zgodnjih fazah procesa, ko še nista znana cilj

rudarjenja in narava podatkov. Preproste opisne in vizualne statistične metode dajo prvi vpogled v podatke, razumevanje podatkov in oblikovanje prvih hipotez.

Opis podatkov in združevanje se uporablja skupaj z drugimi vrstami problemov podatkovnega rudarjenja. Opis podatkov lahko na primer vodi do zanimivih segmentov v podatkih, zato se izvaja že v fazi razumevanja podatkov.

3.3.2.2 Razvrščanje v skupine

Pri razvrščanju v skupine gre za ločitev objektov v zanimive in smiselne skupine ali razrede, pri čemer je skupina opredeljena kot nabor podatkovnih objektov, ki so si med seboj podobni in so hkrati različni od objektov v drugih skupinah. Razvrščanje v skupine se uporablja na različnih področjih proučevanja in je pogosta metoda podatkovnega rudarjenja kot oblika neusmerjenega učenja, saj skupine niso opredeljene vnaprej. V tem se razvrščanje v skupine tudi razlikuje od klasifikacije.

Razvrščanje v skupine je računsko zahtevna metoda, rezultati pa so lahko odvisni tako od tipa podatkov kot od 'šuma', ki ga povzročajo napake ali manjkajoči podatki.

Razvrščanje v skupine je lahko samostojen problem, ko je odkrivanje skupin glavni namen podatkovnega rudarjenja, lahko pa je namen razvrščanja dobiti obvladljivo ali homogeno množico podatkov, ki jih je potem lažje analizirati. V velikih nizih podatkov namreč različni vplivi lahko prekrivajo drug drugega in zameglijo zanimive vzorce, zato nam v tem primeru razvrstitev olajša nalogo (CRISP, 2000).

Metode razvrščanja lahko razdelimo v naslednje skupine:

- Metode razvrščanj v skupine, ki so lahko nehierarhične, ko se podatki, ki jih analiziramo, razvrstijo v skupine tako, da so si objekti znotraj skupine bližje kot v odnosu do objektov v drugih skupinah, ali hierarhične, ko se ponavlja grupiranje ali ločevanje objektov, dokler ni dosežena neka meja. Razvrščanje na podlagi mreže in slikovnih elementov, ki se splošno uporablja v daljinskem zaznavanju, je v bistvu hierarhična metoda.
- Lokalne metode, ki grupirajo objekte v skupine glede na lokalne pogoje in kot osnovo za razvrščanje običajno uporabljajo gostitev.
- Nevronske mreže, ki izvajajo razvrščanje z uporabo ene od prej navedenih metod, potem ko izvedejo proces učenja na obstoječih podatkih (Han & Kamber, 2006).

3.3.2.3 Opis konceptov

Opis konceptov ali tudi simbolna analiza podatkov poda opis konceptov oziroma razredov. Koncepti predstavljajo razrede ali kategorije objektov, ki jih opišemo s simbolnimi podatki – običajno so to kategorični ali numerični podatki. Z metodo dobimo vpogled v določen koncept, vključuje pa statistike kot so seštevki, povprečje, kazalniki razpršenosti, kot je varianca ali primerjava splošnih značilnosti posameznega koncepta.

Primer opisa konceptov je vpogled v značilnosti zvestih in nezvestih kupcev. Iz opisa teh dveh konceptov (zvesti, nezvesti kupci) lahko ugotovimo, kako lahko podjetje ukrepa, da bo obdržalo zveste kupce ali spremenilo vedenje nezvestih kupcev.

Opis konceptov je povezan z razvrščanjem in klasifikacijo. Razvrščanje v skupine lahko vodi v razdelitev objektov v razrede brez razumljivega opisa, zato se običajno izvaja pred opisom konceptov. Nekatere metode podatkovnega rudarjenja pa razvrščanje v skupine in opis konceptov celo izvajajo istočasno.

Opis konceptov se lahko uporablja tudi za klasifikacijo, s katero se nato izdelajo razumljivi modeli, ki jih lahko jemljemo kot konceptualne opise. Razlika je v tem, da je klasifikacija sama zase zaključena in se nanaša na vse enote v izbrani populaciji. Za koncepte je dovolj, da opišejo pomembne dele konceptov ali razredov.

Metodi, s katerimi se izvaja opis konceptov, sta na primer učenje opisnih pravil (angl. *Rule Induction Methods*) in konceptualno razvrščanje v skupine (angl. *Conceptual Clustering*).

3.3.2.4 Klasifikacija

Objekte, ki jih ločijo določene značilnosti, s klasifikacijo razvrstimo v različne razrede. Razred ima diskretno (simbolično) vrednost, ki je znana za vsak objekt. Oznako razreda lahko vnaprej določi uporabnik ali pa se dobi z razvrščanjem v skupine. Klasifikacijski modeli se največ uporabljajo za napovedovalno modeliranje, zato je cilj izgradnja takšnega klasifikacijskega modela, ki bo razvrstil nove, prej neobravnavane objekte v prave razrede.

Klasifikacija je ena najpomembnejših metod podatkovnega rudarjenja, ki se uporablja na različnih področjih, hkrati lahko veliko problemov pretvorimo v problem klasifikacije. Na primer: s kreditnim točkovanjem želimo oceniti kreditno tveganje novega kupca. Problem preoblikujemo v klasifikacijski problem tako, da kreiramo dva razreda, dobre in slabe kupce. Klasifikacijski model lahko zgradimo na podlagi obstoječih podatkov o kupcih in njihovem kreditnem obnašanju, nato pa model uporabimo za razvrstitev novih, potencialnih kupcev v enega od razredov in ga na podlagi tega sprejmemo ali zavrnemo.

Metoda se povezuje z ostalimi metodami podatkovnega rudarjenja. Nekatere klasifikacijske metode najdejo razumljive razrede ali opišejo koncepte. Povezava obstaja tudi z analizo odvisnosti, saj klasifikacijski modeli odkrivajo odvisnosti med atributi. S klasifikacijo lahko dobimo oznake razredov ali pa se niz podatkov omeji, tako da lahko izdelamo ustrezen klasifikacijski model. Pred izgradnjo klasifikacijskega modela je pomembno odkriti odstopajoče vrednosti oziroma odklone, ki lahko zameglijo vzorce in s tem onemogočijo izgradnjo dobrega modela. Po drugi strani pa klasifikacijo lahko uporabimo za odkrivanje izjem v podatkih in druge probleme znotraj podatkov.

Metode, ki se uporabljajo za klasifikacijo, so diskriminacijska analiza, učenje opisnih pravil, odločitvena drevesa, nevronske mreže, k-najbližjih sosedov, sklepanje na osnovi primerov (angl. *Case-based reasoning*), genetski algoritmi.

3.3.2.5 Napovedovanje

Napovedovanje se uporablja na veliko področjih in je zelo podobno klasifikaciji, razlika je v tem, da ciljni atribut pri napovedovanju ni razred, torej diskretna vrednost, temveč zvezna. Cilj napovedi je poiskati numerično vrednost ciljnega atributa za neopazovane enote.

V literaturi se ta vrsta problema označuje s pojmom regresija, kadar pa se analizirajo časovne vrste se označuje s pojmom napoved.

Metode, ki se uporabljajo za napovedovanje, so: regresijska analiza, odločitvena drevesa, nevronske mreže, najbližji sosed, Box-Jenkinsove metode, genetski algoritmi.

3.3.2.6 Analiza odvisnosti

Analiza odvisnosti poišče model, ki opiše pomembne povezave med enotami. Odvisnost se uporablja za napovedno modeliranje, tj. napovedovanje vrednosti nekega atributa na podlagi vrednosti drugih atributov.

Asociacije (angl. *Associations*) so poseben primer odvisnosti, ki so v zadnjem času postale zelo popularne. Asociacije opisujejo povezanost med objekti (na primer enote, ki se pogosto pojavljajo skupaj). Tipičen primer je nakupovalna košarica. Pravilo 'v 30 odstotkih nakupov so bili oreški nabavljeni skupaj s pivom' je tipičen primer povezave. Algoritmi za odkrivanje asociacij so zelo hitri in najdejo veliko povezav.

Analiza odvisnosti je tesno povezana z napovedovanjem in klasifikacijami, pri čemer se odvisnost neposredno uporablja za oblikovanje napovedovalnih modelov. Povezujejo se tudi z opisi konceptov, ki pogosto že na začetku analize poudarijo odvisnosti.

Metode, ki se uporabljajo v analizi odvisnosti, so korelacijska analiza, regresijska analiza, asociacijska pravila, Bayesove mreže, induktivno logično programiranje, vizualizacijske metode.

3.3.3 Proces podatkovnega rudarjenja

CRISP-DM je splošno sprejeta metodologija v projektih podatkovnega rudarjenja (CRISP-DM). Koraki v procesu rudarjenja so:

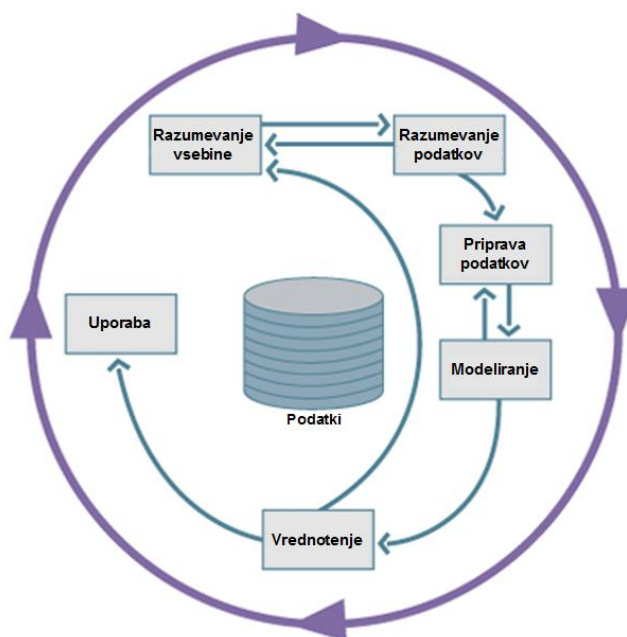
1. **Razumevanje vsebine:** Razumevanje ciljev in zahtev z vsebinskega vidika in pretvorba tega znanja v oblikovanje problema za podatkovno rudarjenje in začasni načrt, ki bo oblikovan za uresničitev postavljenih ciljev.
2. **Razumevanje podatkov:** Zbiranje podatkov, poznavanje, prepoznavanje problemov v kakovosti podatkov ali odkrivanje zanimivih vzorcev za oblikovanje hipotez o skritih informacijah.

3. **Priprava podatkov:** Vse aktivnosti, ki jih je potrebno izvesti, da bi zgradili končni niz podatkov (podatkov, ki bodo vhod v orodje za modeliranje) iz začetnih neobdelanih podatkov. Sem spada izbira tabel, primerov in atributov ter pretvorba in čiščenje podatkov za modeliranje.
4. **Modeliranje:** Izbira in uporaba različnih metod za modeliranje in najboljšo oceno parametrov. Običajno obstaja več metod za isti problem rudarjenja po podatkih. Nekatere metode imajo posebne zahteve glede oblike podatkov, zato se je pogosto treba vrniti nazaj na korak priprave podatkov.
5. **Ocenjevanje:** Temeljita ocena modela in analiza predhodnih postopkov pri izgradnji modela, s čimer se zagotovi, da ta ustreza želenim ciljem. Potrebno je pregledati, da ni bil izpuščen kakšen pomemben problem/smo izpustili pomembnega problema. Na koncu se je treba odločiti o uporabi rezultatov podatkovnega rudarjenja.
6. **Uporaba:** Ureditev in predstavitev rezultatov podatkovnega rudarjenja. Uporaba je lahko enostavna, kot je izdelava poročila, ali kompleksna, kot ponavljajoč se proces podatkovnega rudarjenja.

Podatkovno rudarjenje je ponavljajoč se proces, ki se nadaljuje z uporabo rešitve. Med procesom pridobljeno znanje lahko sproži nova vprašanja. Spreminjanje podatkov lahko zahteva nove modele. Nadaljnji proces podatkovnega rudarjenja pa gradi na preteklih izkušnjah.

Proces podatkovnega rudarjenja po metodi CRISP-DM je prikazan na Sliki 12.

Slika 12: Proces podatkovnega rudarjenja



Vir: CRISP-DM, CRISP-DM 1.0, Step-by-step datamining guide, 2000, str.13.

3.3.4 Oracle Data Mining

Oracle Data Mining (ODM) podpira zadnje tri korake procesa CRISP-DM. Prvi korak (poznavanje vsebine) je vezan na področje proučevanja. Poznavanje področja proučevanja in dostopnih podatkov je še vedno najpomembnejši dejavnik v procesu. Preostala koraka pa podpira kombinacija podatkovnega rudarjenja in podatkovne baze, zlasti v primeru podatkovnih skladišč. Oracleova podatkovna baza in njene funkcije so lahko zelo uporabne pri razumevanju in pripravi podatkov (Oracle, 2007).

Oracle Data Mining avtomatizira postopke izgradnje, testiranja in uporabe modela, tako da se lahko uporabnik osredotoči na poslovni vidik problema, ne na matematične in statistične podrobnosti (Oracle, 2008b).

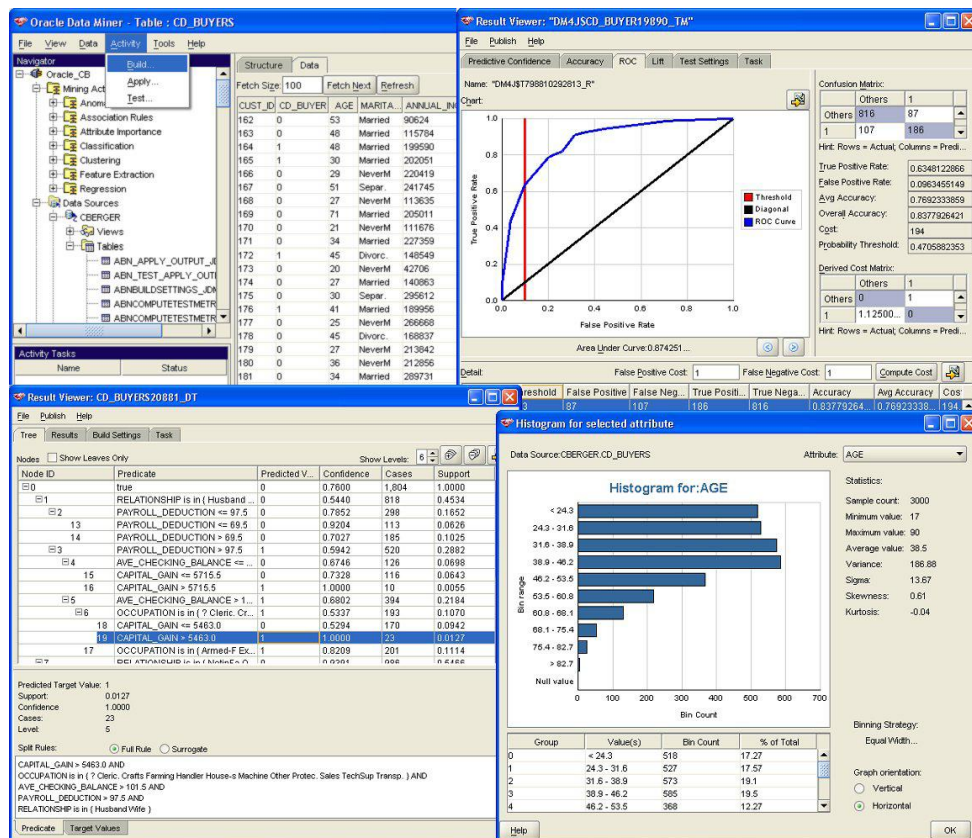
Oracle Data Mining ima podatkovno rudarjenje vključeno v samo podatkovno bazo. Algoritmi za podatkovno rudarjenje Oracle se uporabljajo na relacijskih tabelah in pogledih, zato ni treba izvoziti in prenašati podatke v samostojna orodja in specializirane analitične strežnike. ODM-jeva integrirana arhitektura vodi do preprostejšega, zanesljivejšega in učinkovitejšega upravljanja podatkov in izvedbe analize.

Izvajanje opravil podatkovnega rudarjenja je lahko asinhrono in neodvisno od kakršnega koli uporabniškega vmesnika kot del standardnih procesov na podatkovni bazi. Analitiki lahko rudarijo po podatkih v bazi, gradijo modele in metodologije ter nato rezultate in metodologije vrnejo kot sestavne dele aplikacij za uporabo v produkcijskem okolju. Koristi integracije s podatkovno bazo so neprimerljive, ko gre za uporabo v produkcijskem okolju. ODM omogoča izkoriščanje vseh prednosti Oracleove tehnologije. Manj gibljivih delov pomeni preprostejšo, zanesljivejšo in zmogljivejšo napredno aplikacijo poslovnega obveščanja. (Oracle, 2007).

Oracle Data Mining ima dva grafična vmesnika, ki omogočata lažjo uporabo podatkovnega rudarjenja, in sicer:

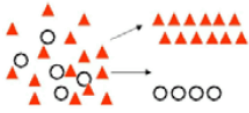
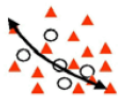
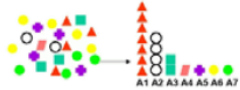
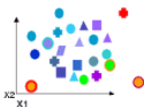
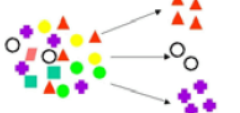
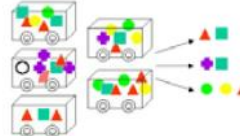
- Oracle Data Miner je uporabniški vmesnik za ODM in je v pomoč analitikom in razvijalcem pri izgradnji aplikacij za poslovno inteligenco.
- Oracle Spreadsheet Add-In for Predictive Analytics omogoča uporabnikom Microsoft Excel-a podatkovno rudarjenje v Excelovih preglednicah z uporabo PL/SQL paketa DBMS_PREDICTIVE_ANALYTICS.

Slika 13: Oracle Data Miner – interaktivni grafični vmesnik



V Tabeli 2 so predstavljene metode in algoritmi, ki jih podpira Oracle Data Mining.

Tabela 2: Metode in algoritmi, ki jih podpira Oracle Data Mining

Metoda	Algoritem
Klasifikacija 	• Logistična regresija (GLM) – klasična statistična metoda. • Naivni Bayes – hiter, preprost, splošno uporaben. • Metoda podpornih vektorjev – najnovejša tehnologija. Podpira veliko vhodnih podatkov, transakcijske podatke in besedila. • Odločitvena pravila – Priljubljen algoritem. Rezultat so berljiva "Če ..., potem..." pravila.
Regresija 	• Multipla linearna regresija (GLM) – klasična statistična metoda. • Metoda podpornih vektorjev – Najnovejša tehnologija. Podpira linearno in nelinearno regresijo (Gaussovo jedro).
Pomembnost atributov 	• Minimalna dolžina opisa – algoritem poišče tiste attribute, ki imajo največji vpliv na ciljni atribut.
Odkrivanje anomalij 	• Enorazredna metoda podpornih vektorjev – nenadzorovna metoda učenja se uči na 'normalnih primerih' in zgradi model. Neobičajne primere označi z verjetnostmi.
Razvrščanje v skupine 	• O-cluster (angl. <i>Orthogonal Partitioning Clustering</i>) – Oraclov zaščitene algoritem za hierarhično razvrščanje na osnovi gostote. • Izboljšani K-means – podpira razvrščanje besedil. Hierarhično razvrščanje na osnovi razdalj.
Asociacijska pravila 	• Apriori – standard za analizo nakupovalnih košaric in odkrivanje pravil, kako se posamezni predmeti pojavljajo skupaj.
Feature Extraction 	• Nenegativna matrična faktorizacija (angl. <i>Non-negative Matrix Factorization, NMF</i>) – kreiranje manjšega števila atributov na podlagi starih.

Vir: Oracle, Data mining, Data Sheet, 2009

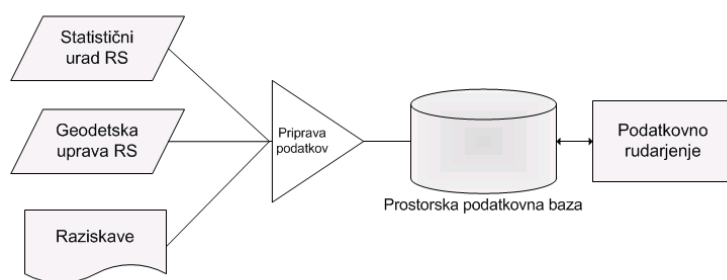
4 PRIMER UPORABE INFORMACIJSKEGA SISTEMA

4.1 Podatki

Iz poglavja o prostorskih analizah in modelih izhaja, da vstopajo v modele družbeno-ekonomski podatki, podatki o tokovih in podatki o razdaljah. Enote za katere se ti podatki zbirajo so prostorske enote, na določeni ravni.

Statistični urad republike Slovenije zbira, evidentira, obdeluje in posreduje podatke za analize na podlagi klasifikacije teritorialnih enot (SKTE). SKTE do tretje ravni temeljijo na klasifikaciji NUTS, za potrebe statistike pa je območje Slovenije razdeljeno na pet nižjih ravni. Večina podatkov je še razpoložljiva na ravni SKTE5, tj. na ravni občin. Družbenoekonomski podatki so relativno dobro pokriti, podatki o interakcijah oziroma tokovih pa manj.

Slika 14: Krog podatkov



Socialno-ekonomski in demografski in drugi podatki se zbirajo tudi na ravni občin so dostopni preko SI-STAT podatkovnega portala in elektronske publikacije Slovenske občine v številkah, ki je letno objavljena tudi kot tiskana publikacija.

Statistični urad pokriva naslednje podatke o tokovih:

- Slovenska statistika vodi tokove dnevnih migracij na ravni občin, podatki so dostopni na letni ravni. Večina dnevnih migracij poteka med strnjenimi naselji ali med glavnimi mesti in podeželsko okolico, zato so idealni podatki o dnevnih migracijah na ravni občin. Popisni podatki o dnevnih migracijah vsebujejo tudi podatke o pogostnosti in načinu potovanja na delo ter porabljenem času.
- Vodiijo se tudi evidence dijakov in študentov po občinah stalnega bivališča in občini šole oziroma visokošolskega zavoda.
- Na podlagi podatkov notranjega ministrstva statistični urad dobro pokriva selitveno gibanje prebivalstva, tako meddržavnih migracij kot notranjih migracij, kjer so podatki dostopni tako na ravni statističnih regij kot na ravni občin. Notranje selitve in skupni

prirast med občinami se spremlja po priseljenih in odseljenih v drugo občino. Nekateri podatki o notranjih selitvah so tudi razčlenjene po starosti in spolu.

- Podatki o cestnem javnem potniškem prevozu, cestnem blagovnem prevozu, železniškem blagovnem in potniškem prevozu, pomorskem in zračnem transportu so več ali manj in javni prevoz za nekatera mesta v državi. Cestni blagovni prevoz je voden kot tok tudi na ravni statističnih regij (regija nalaganja / razlaganja). Glede na način pridobivanja podatkov (poročila poslovnih subjektov), verjetno podatki obstajajo ali pa se jih da izračunati tudi na ravni občin.

Razdalje med občinami niso dostopne, lahko pa se izračunajo na podlagi prostorskih podatkov (na primer cestnega omrežja), ki jih vodi in posreduje Geodetska uprava Republike Slovenije. Zbirka prostorskih podatkov je tudi potrebna za prikazovanje podatkov na zemljevidih oziroma v geografskem informacijskem sistemu.

Podatki se po potrebi lahko dopolnijo s podatki drugih institucij in raznih raziskav.

V podatkovno bazo naj bi, če je le mogoče, vstopali izvorni podatki. Če ni mogoče, naj bi se navedla pot, kako smo dobili podatke. Priporočljivo je tudi, da se zbirajo surovi podatki, ne pa izračunani kazalniki. Ko ni mogoče pridobiti surovih podatkov, se navede utež uporabljena pri agregiranju. Kazalniki so pogosto izračunani na podlagi matematičnih kombinacij surovih podatkov (seštevki, ulomki, zmnožki), ki sicer niso neposredno uporabni, je pa mogoče iz niza surovih podatkov izpeljati niz kazalnikov za različne ravni prostorske agregacije in za različne opredelitve kazalnikov.

4.2 Razširjeni gravitacijski model za napovedovanje dnevnih migracij med slovenskimi občinami

4.2.1 Razumevanje vsebine in uporabljena metoda

Cilj je na podlagi podatkov o dnevni migraciji zaposlenih iz občine prebivališča v občino zaposlitve, podatkov o velikosti prebivalstva, brezposelnosti, plačah itd. prikazati izdelavo modela, s katerim je možno na podlagi prihodnjih scenarijev napovedati dnevne migracije za oblikovanje funkcionalnih regij. Za model sem izbrala razširjeni gravitacijski model.

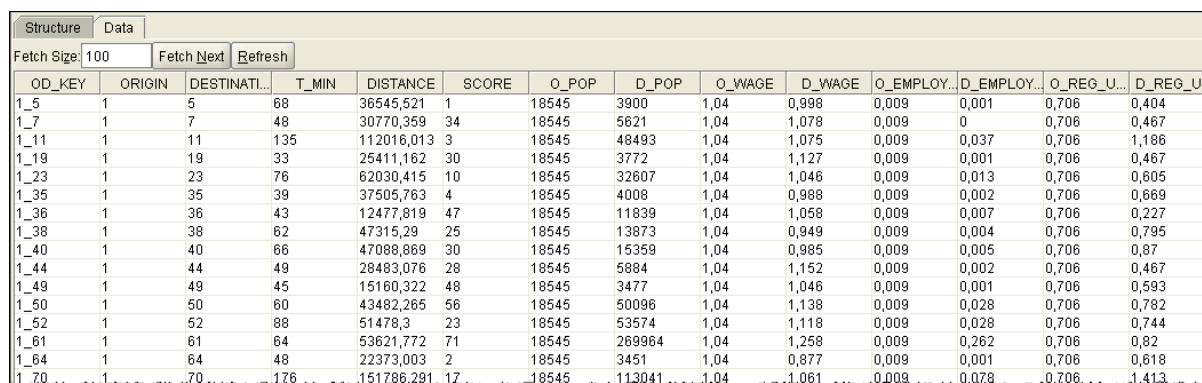
Za napovedovanje zveznih vrednosti ima Oracle Data Mining na voljo regresijo, in sicer sta na voljo dva algoritma za izdelavo regresijskih modelov, prvi je metoda posplošenih najmanjših kvadratov, drugi algoritem pa temelji na metodi podpornih vektorjev. Metoda posplošenih najmanjših kvadratov je splošna, zato ni potrebe po njeni razlagi. Najprej bom model ocenila z metodo najmanjših kvadratov, nato še z metodo podpornih vektorjev.

4.2.2 Razumevanje in priprava podatkov

Za izdelavo modela sem vzela podatke o občinah za leto 2007, in sicer število prebivalcev, podatke o neto plačah, stopnji nezaposlenosti, številu delovnih mest in številu visoko izobraženih, ki so dostopni preko Statističnega urada, tako kot podatki o aktivnih prebivalcih po občini prebivališča in občini delovnega mesta.

Kot merilo razdalje sem vzela čas potovanja z avtom iz občine izvora do občine ponora. Podatke o potovalnih časih mi je prijazno posredoval gospod Samo Drobne iz FG, saj izračun ni najlažje izvedljiv. Podatkov o razdaljah je okrog 7099 in so časovne razdalje izpred nekaj let. Predpostavljamo, da so podatki vseeno dovolj dobri za ponazoritev problema, saj se prometna infrastruktura v tem času ni toliko spremenila, zaposleni pa se ne odločajo za menjavo zaposlitev na kratek rok in čas potovanja na delo ni edini kriterij izbire.

Slika 15: Prikaz podatkov za izdelavo modela v aplikaciji Oracle Data Miner



OD_KEY	ORIGIN	DESTINATI...	T_MIN	DISTANCE	SCORE	O_POP	D_POP	O_WAGE	D_WAGE	O_EMPLOY...	D_EMPLOY...	O_REG_U...	D_REG_U...
1_5	1	5	68	36545,521	1	18545	3900	1,04	0,998	0,009	0,001	0,706	0,404
1_7	1	7	48	30770,359	34	18545	5621	1,04	1,078	0,009	0	0,706	0,467
1_11	1	11	135	112016,013	3	18545	48493	1,04	1,075	0,009	0,037	0,706	1,186
1_19	1	19	33	25411,162	30	18545	3772	1,04	1,127	0,009	0,001	0,706	0,467
1_23	1	23	76	62030,415	10	18545	32807	1,04	1,046	0,009	0,013	0,706	0,605
1_35	1	35	39	37505,763	4	18545	4008	1,04	0,988	0,009	0,002	0,706	0,669
1_36	1	36	43	12477,819	47	18545	11839	1,04	1,058	0,009	0,007	0,706	0,227
1_38	1	38	62	47315,29	25	18545	13873	1,04	0,949	0,009	0,004	0,706	0,795
1_40	1	40	66	47088,869	30	18545	15359	1,04	0,985	0,009	0,005	0,706	0,87
1_44	1	44	49	28483,076	28	18545	5884	1,04	1,152	0,009	0,002	0,706	0,467
1_49	1	49	45	15160,322	48	18545	3477	1,04	1,046	0,009	0,001	0,706	0,593
1_50	1	50	60	43482,265	56	18545	50096	1,04	1,138	0,009	0,028	0,706	0,782
1_52	1	52	88	51478,3	23	18545	53574	1,04	1,118	0,009	0,028	0,706	0,744
1_61	1	61	64	53621,772	71	18545	269964	1,04	1,258	0,009	0,262	0,706	0,82
1_64	1	64	48	22373,003	2	18545	3451	1,04	0,877	0,009	0,001	0,706	0,618
1_70	1	70	176	151786,291	17	18545	113041	1,04	1,061	0,009	0,078	0,706	1,413

Iz množice podatkov najprej poskusimo ugotoviti, kateri atributi so tisti, ki najbolj vplivajo na dnevni tok migrantov iz neke občine v drugo občino.

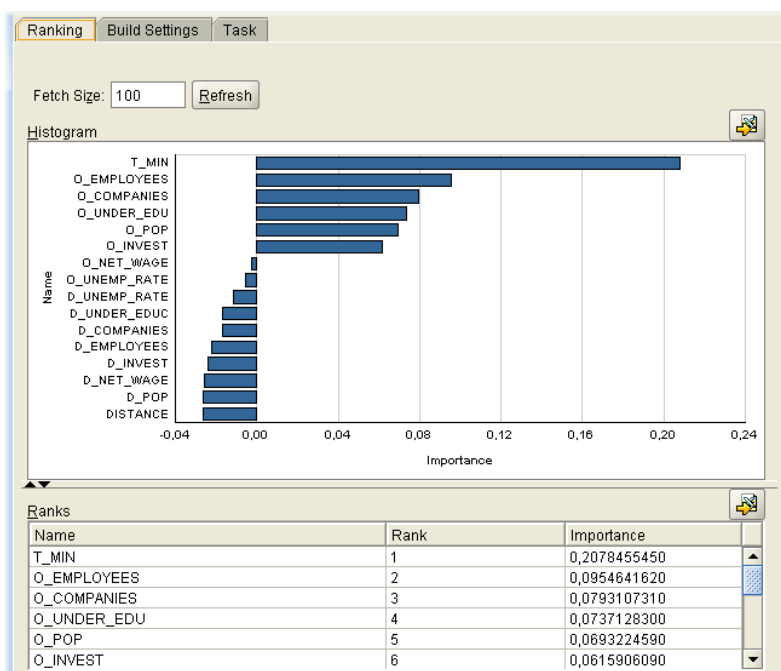
Včasih lahko preveč informacij zmanjša učinkovitost podatkovnega rudarjenja. Nepomembni atributi se označujejo tudi s 'šum' v podatkih. Z odstranitvijo šumov lahko izboljšamo točnost in kakovost modela ter skrajšamo čas izgradnje. Obstaja več pristopov, kako zmanjšamo število atributov. Lahko določimo relativno napovedno pomembnost atributov in odstranimo tiste pod neko mejo. Druga možnost je, da povezane attribute povežemo v pomembne značilke.

ODM podpira dve funkciji za zmanjšanje atributov: Pomembnost atributov, ki razvrsti attribute glede na njihovo napovedno moč oziroma pomembnost. Izločanje značilk (*Feature Extraction*) pa poišče pomembne značilnosti v podatkih tako, da kombinira povezane attribute v značilke.

Pomembnost atributov je nadzorovana funkcija podatkovnega rudarjenja, ki modelira rang atributov glede na njihovo napovedno moč in je pomemben predkorak zlasti za klasifikacijske modele, ki uporabljajo naivni Bayes ali metodo podpornih vektorjev.

Iz razvrstitve atributov po moči vpliva na tok dnevnih migrantov (slika 15) je razvidno, da potovalna razdalja najbolj vpliva na tok migrantov, ter da evklidska razdalja (spremenljivka DISTANCE) ni najbolj primerna za napovedovanje toka dnevnih migrantov. Pomembne pri napovedovanju so vse spremenljivke, ki so vezane na občino izvora in so povezane z velikostjo populacije v izvornem območju (število prebivalcev, število zaposlenih, število podjetij, število prebivalcev z višjo in visoko izobrazbo) ter da stopnja nezaposlenosti in višina neto plače nimata tolikšnega vpliva na dnevne migracije. Prav tako ima manjši vpliv velikost populacije v občini delovnega mesta.

Slika 16: Razvrstitev atributov po napovedovalni moči

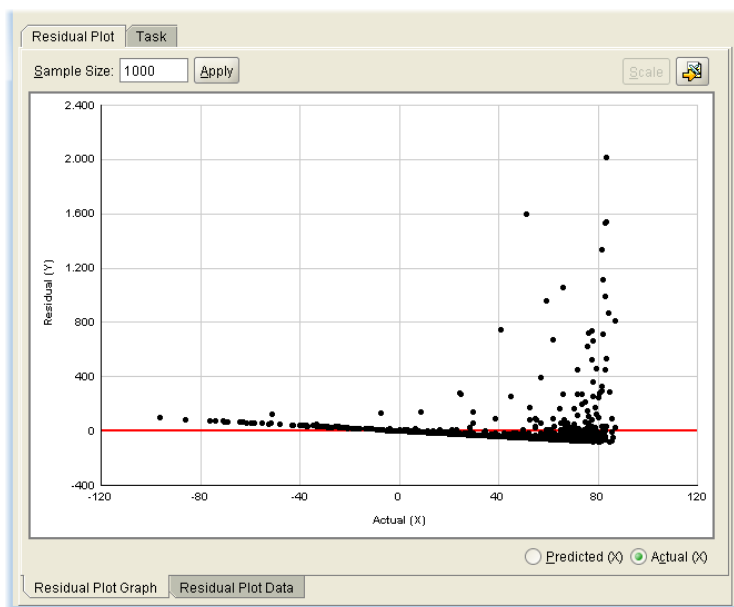


Ker so spremenljivke, ki določajo velikost populacije med seboj odvisne sem v model vključila samo velikost populacije, čeprav bi bilo verjetno bolj smiselno velikost aktivnega prebivalstva. Izločila sem tudi neto plačo v regiji ponora, saj je ta statistično neznačilna (kar izhaja iz prejšnjih oblikovanj modela), vseeno pa sem vključila velikost populacije v regiji ponora, podatke o plačah in nezaposlenosti.

Pred modeliranjem sem preverila, ali lahko uporabim linearno funkcijo, tako da sem preverila odvisnost dnevnih migracij od potovalne razdalje. Razsevni diagram ostankov prikazuje razlike med dejanskimi vrednostmi in napovedanimi vrednostmi. Možna sta dva prikaza, prikaz dejanskih vrednosti ali napovedanih vrednosti na X osi. V vsakem primeru pika na 0 črti pomeni natančno napoved, stran od črte pa napako.

Spodnji diagram prikazuje dejanske vrednosti na X osi in odgovori na vprašanje: 'Za katere nize dejanskih vrednosti je model točen?' Napake začnejo odstopati pri večjih vrednostih tokov, kjer so odstopanja ogromna. V vsakem primeru pomeni vrednost na premici 0 natančno napoved, točke stran od premice pa predstavljajo napako.

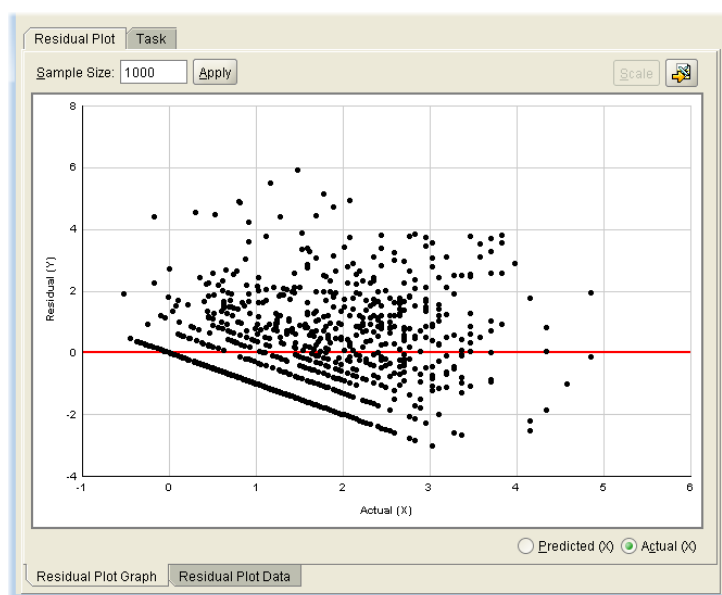
Slika 17: Razsevni diagram ostankov linearnega regresijskega modela



Vidimo, da linearna funkcija ni najbolj primerna. Poskus tudi pokaže, da model ocenjen z linearno funkcijo ni skoraj nič boljši kot naivni model. Vrednost determinacijskega koeficienta je blizu nič in linearni model kaže zelo slabo napovedno noč.

Preverimo še s potenčno funkcijo. Odvisno in neodvisno spremenljivko logaritmiramo ter ponovno ocenimo model. Napake se porazdeljujejo bolj normalno okoli nič.

Slika 18: Razsevni diagram ostankov potenčnega regresijskega modela



Za izvedbo modela je potrebno podatke pripraviti v eni tabeli ali pogledu v bazi. Podatki se lahko pripravijo z orodji za delo na bazi ali pa se izberejo in združijo v Oracle Data Minerju.

4.2.3 Modeliranje in vrednotenje

4.2.3.1 Ocena potenčne funkcije z metodo posplošenih najmanjših kvadratov

Model: Izberemo tabelo z logaritmiranimi podatki, izberemo unikatni ključ tabele in izberemo stolpce (spremenljivke) na katerih naj se izvede algoritem. Določimo odvisno spremenljivko.

Ocenjen model poda vrsto standardnih statistik, ki opisujejo model in so prikazane na spodnji sliki.

$$R^2 = 0,62 \quad (20)$$

Determinacijski koeficient kaže, da je 62 odstotkov variance toka dnevnih migrantov pojasnjene s potenčnim vplivom vključenih neodvisnih spremenljivk.

Slika 19: Splošna statistika potenčnega regresijskega modela

Global Statistics	
Target Attribute: SCORE	
Metric Name	Value
Model degrees of freedom	6
Model sum of squares	8904,3471841464
Model mean square	1484,0578640244
Model F value statistic	1190,3185894887
Model F value probability	0
Error degrees of freedom	4277
Error sum of squares	5332,450942532
Error mean square	1,2467736597
Corrected total degrees of freedom	4283
Corrected total sum of squares	14236,7981266785
Root mean square error	1,1165901933
Dependent mean	1,6061122462
Coefficient of variation	69,5213049955
R-Square	0,6254459117
Adjusted R-Square	0,6249204676
Akaike's information criterion	951,8696406939
Schwarz's Bayesian information criterion	996,4081377149
Estimated mean square error of the prediction	1,2488145469
Hocking Sp statistic	0,0002915748
JP statistic	1,2488108716
Number of parameters	7
Number of rows	4284
Model converged?	yes
Covariance matrix computed?	yes

Slika 20: Ocene parametrov potenčne regresijske funkcije

Global Statistics

Coefficients

Results

Build Settings

Task

Target Attribute: SCORE

☒ Show Intercept Row

Coefficients

Fetch Size: 100

Refresh

Filter

Attribute Name	Value	Coefficient	Standard Error	Wald Chi-Squa...	Pr > Chi-Square	Standardized ...	Lower Coeffici...	Upper Coeffici...	Exp
<< Intercept Ro...		-13,5832742614	1,180772	-11,50372	0	0	-15,8982	-11,268348	0
T_MIN		-1,6765699956	0,023774	-70,521979	0	-0,710178	-1,723179	-1,629961	0
O_NET_WAGE		1,3560437163	0,17768	7,631931	0	0,077757	1,007698	1,704389	0
O_POP		0,9077810984	0,01697	53,493135	0	0,542331	0,874511	0,941051	0
D_POP		0,4163162915	0,01605	25,938795	0	0,257624	0,38485	0,447783	0
D_UNEMP_RA...		0,2417253900	0,043168	5,59966	0	0,059184	0,157094	0,326357	0
O_UNEMP_RA...		0,1264307963	0,042961	2,942915	0,003269	0,031915	0,042205	0,210657	0

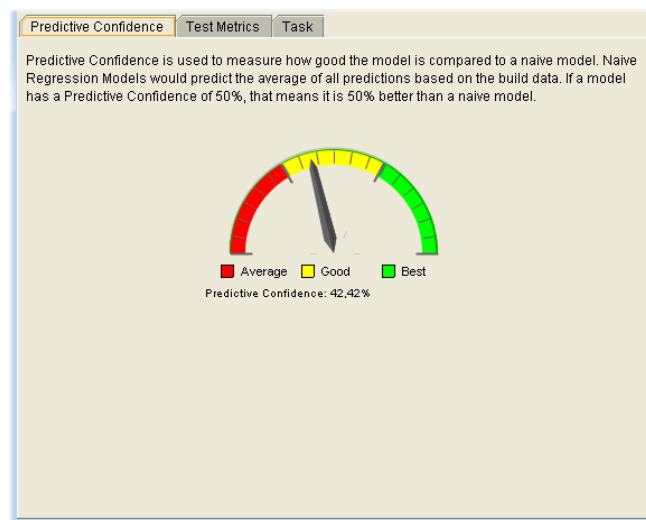
☒ Sort coefficients based on absolute values

Enačba modela:

$$SCORE = 1.26 \cdot 10^6 \frac{O_NET_WAGE^{1.35} O_POP^{0.90} D_POP^{0.41} O_UNEMP_RATE^{0.24} D_UNEMP_RATE}{T_MIN^{1.68}} \quad (21)$$

Oracle Data Miner izračuna napovedno zanesljivost regresijskega modela. Napovedna zanesljivost je merilo, ki meri koliko je model boljši kot 'naiven' model, ki ne bi izvedel analize in bi preprosto napovedal povprečje. Regresijski model ocenjen s potenčno funkcijo ima napovedno zanesljivost 42 odstotkov, kar pomeni, da je model za 42 odstotkov boljši kot naivni model.

Slika 21: Napovedna zanesljivost potenčnega modela



Koren povprečne kvadratne napake in povprečna absolutna napaka sta pogosto uporabljeni testni statistiki, ki ocenjejeta splošno kakovost regresijskega modela. Na voljo so različne statistike, katere pa je odvisno od regresijske metode, ki jo uporablja algoritem.

Koren povprečne kvadratne napake (angl. *Root Mean Squared Error*, RMSE) je kvadratni koren povprečne kvadratne razdalje od podatkovne točke do premice.

$$RMSE = \sqrt{\frac{1}{n} \sum_{j=1}^n (y_j - \hat{y})^2} = 1.06 \quad (22)$$

Povprečna absolutna napaka (angl. *Mean Absolute Error*, MAE) je povprečje absolutnih vrednosti napake. Mera je podobna RMSE, a je manj občutljiva na velike napake.

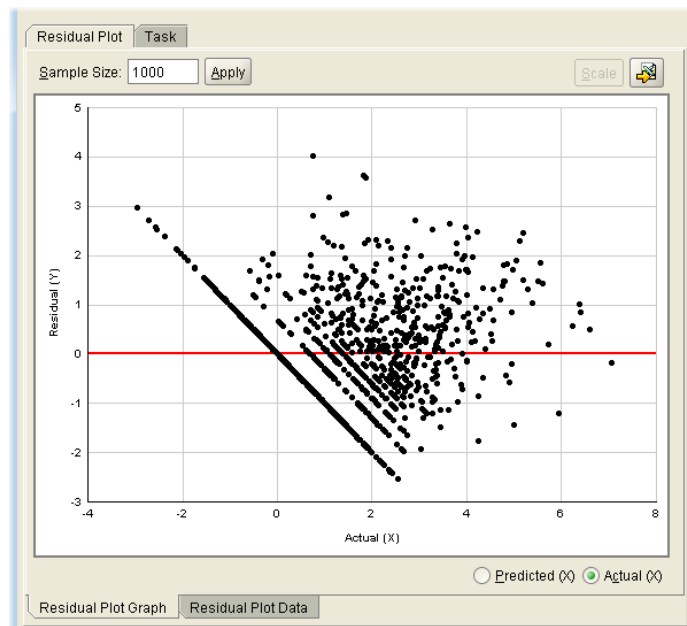
$$MAE = \frac{1}{n} \sum_{j=1}^n |y_j - \hat{y}| = 0.85 \quad (23)$$

Testna statistika za potenčni model je prikazana na Sliki 21. Na Sliki 22 pa je prikazan razsevni diagram ostankov potenčnega modela.

Slika 22: Testna statistika potenčnega modela

Predictive Confidence	Test Metrics	Task
Mean Absolute Error	0,8537733771	
Mean Actual Value	1,6544653	
Mean Predicted Value	1,6526082785	
Root Mean Square Error	1,0624115885	

Slika 23: Razsevni diagram ostankov potenčnega modela

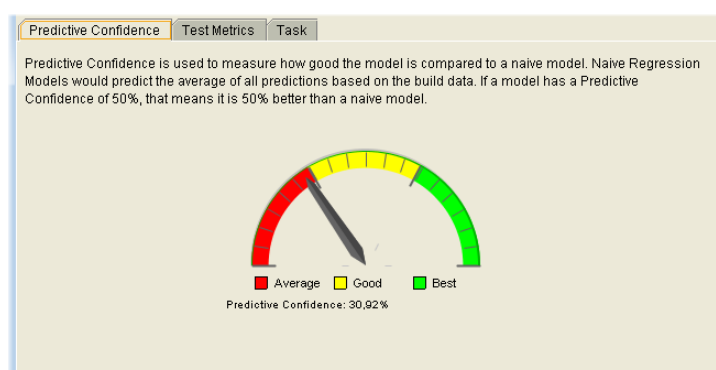


4.2.3.2 Ocena modela z metodo podpornih vektorjev

Izgradnjo modela zaženemo podobno kot smo to naredili v potenčnem modelu. Iz pregleda podatkov na začetku podatkovnega rudarjenja sledi, da odnosi niso linearni, zato kot parameter izberemo Gaussovo jedro.

Rezultati modela kažejo, da je napovedna zanesljivost modela za 31 odstotkov boljša kot bi bil naivni model. Napovedna zanesljivost in testna statistika ter razsevni diagram ostankov modela ocenjenega z metodo podpornih vektorjev so prikazani na Slikah 23, 24 in 25.

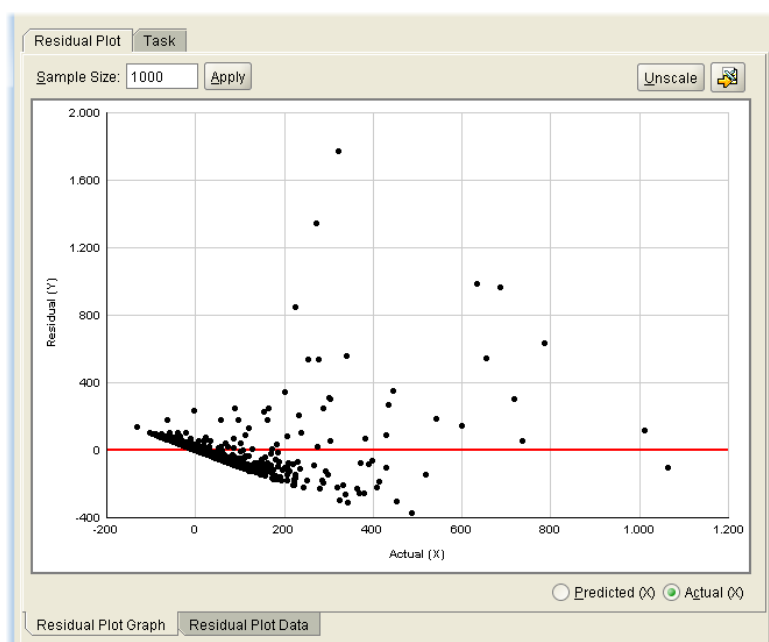
Slika 24: Napovedna zanesljivost SVM modela



Slika 25: Testna statistika SVM modela

Predictive Confidence	Test Metrics	Task
Mean Absolute Error	75,3886043243	
Mean Actual Value	55,9680284192	
Mean Predicted Value	59,3379278468	
Root Mean Square Error	180,4441664905	

Slika 26: Razsevni diagram ostankov SVM modela



4.2.4 Uporaba

Model, ki smo ga dobili z metodo posplošenih najmanjših kvadratov ali metodo podpornih vektorjev uporabimo tako, da z modelom dobimo napoved za nek prihodnji scenarij. Uporaba modela je zelo enostavna, saj samo izberemo model in izberemo tabelo s podatki na katerih bomo izdelali napoved. Tabela mora biti v enaki obliki, kot je bila vhodna tabela v model, podatki pa so seveda lahko drugi. Podatki za napoved lahko vsebujejo nove potovalne čase med občinami glede na planirano gradnjo infrastrukture, spremembe v številu prebivalcev ipd.

Na Sliki 26 je rezultat napovedi izdelane na osnovnih podatkih z modelom podpornih vektorjev, ki vsebuje ključ (sestavljen je iz šifre občine izvora in šifre občine ponora) ter napoved, ki pomeni število dnevnih migrantov iz občine prebivališča v občino zaposlitve. Za lažje razumevanje in kasnejšo obdelavo rezultatov sta dodani polji šifra občine izvora in šifra občine ponora.

Slika 27: Napoved regresijskega modela, ocenjenega z metodo podpornih vektorjev

Apply Output			
Apply Settings			
Task			
Apply Output Table: GRAVITY_EXT_LINE697564518_A			
Fetch Size: 5000 Refresh			
DMR\$CASE_ID	ORIGIN	DESTINATION	PREDICTION
52_61	52	61	779,2104
52_63	52	63	76,3482
52_64	52	64	134,9134
52_70	52	70	212,4001
52_71	52	71	611,108
52_72	52	72	418,8816
52_73	52	73	20,4998
52_80	52	80	74,3192
52_82	52	82	657,9232
52_84	52	84	-28,4349
52_85	52	85	10,0874
52_90	52	90	47,1534
52_94	52	94	70,6482
52_95	52	95	523,5847
52_98	52	98	39,6062
52_102	52	102	513,6804
52_106	52	106	49,2532
52_108	52	108	52,8533
52_110	52	110	19,171
52_111	52	111	-68,1265
52_112	52	112	42,5488
52_113	52	113	57,6282
52_114	52	114	79,8137

Tabela se avtomatično shrani na bazi, lahko pa rezultat tudi ročno shranimo pod nekim drugim imenom in ga uporabimo za kasnejše analize in/ali predstavitev rezultatov. Rezultat lahko izvozimo tudi v Excel.

4.3 Uporaba metode k-Means za razvrstitev občin v funkcionalne regije

4.3.1 Razumevanje vsebine

S podatkovnim rudarjenjem želimo na podlagi podatkov o dnevnih migracijah zaposlenih iz občine prebivališča v občino delovnega poiskati skupine občin, ki funkcionalno težijo skupaj. S podatkovnim rudarjenjem torej želimo poiskati funkcionalne regije.

Funkcionalne regije morajo biti oblikovane tako, da bodo občine, ki bodo uvrščene v posamezno funkcionalno regijo sosednje ali pa bodo odstopanja minimalna.

4.3.2 Razumevanje podatkov

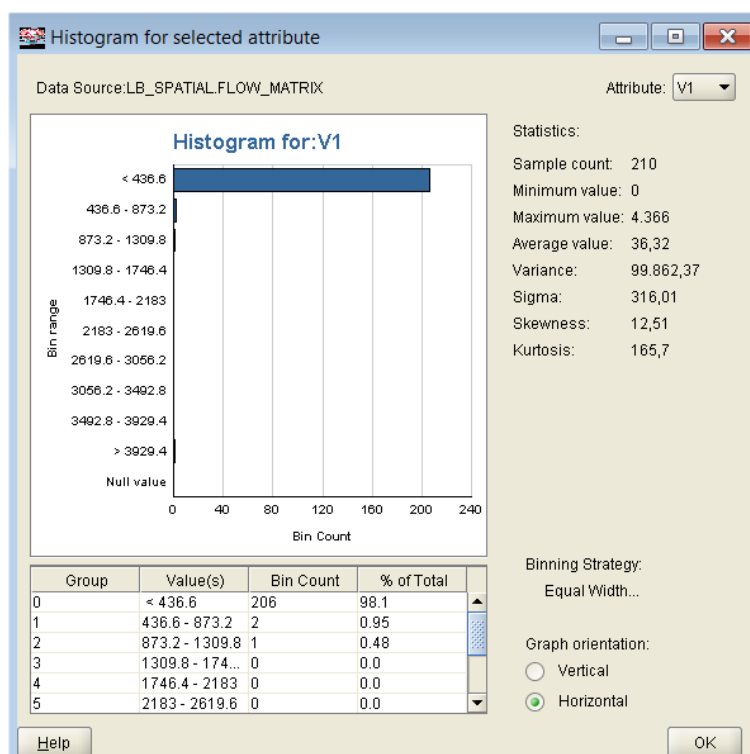
Za analizo sem vzela podatke o delovno aktivnem prebivalstvu (brez kmetov) po občini prebivališča in občini delovnega mesta za leto 2007.

Podatke sem izvozila iz SI-STAT podatkovnega portala v Excel datoteko. Podatki so v obliki matrike, kjer je občina prebivališča vrstični podatke in občina delovnega mesta stolpčni podatek. Podatke sem zaradi preglednosti transformirala v obliko za uvoz v podatkovno bazo Oracle 11g, in sicer tako kot je opisano v opredelitvi podatkovnega modela. Podatki se uvozijo v tabelo s tremi stolpci: občina izvora, občina ponora, število migrantov.

Podatki o dnevnih migrantih zelo nihajo po občinah, npr. število migrantov iz občine prebivališča v občino zaposlitve je odvisen od števila delovno aktivnih v tej občini, ki je odvisno od velikosti občine in zaposlitvenih možnosti v domači občini. V zadnjih dveh desetletjih je v Sloveniji nastala cela množica občin z majhnim številom prebivalcev. Nekatere so tudi slabo razvite in praktično nimajo nobene ekonomske baze. Poleg tega je število delovnih migrantov druge občine odvisno od zaposlitvenih priložnosti v drugi občini, torej od števila delovnih mest oziroma velikosti občine.

Če pogledamo razporeditev migrantov v občino Ajdovščina (Slika 27), lahko vidimo, da je iz večine občin tok zelo majhen ali ga celo ni, večji tok lahko opazimo iz sosednjih občin, vendar je ta (kot že povedano) odvisen od velikost občine. Kar 66 odstotkov občin z Ajdovščino ni povezanih na podlagi delovnih migracij in kar iz 90 odstotkov občin je ta ničeln ali pa zanemarljiv (manj kot 10 dnevnih migrantov). Kar 48 odstotkov Ajdovcev se zaposli v svoji občini.

Slika 28: Opisna analiza toka dnevnih migrantov za posamezno občino



4.3.3 Priprava podatkov

Podatki za rudarjenje v ODM morajo biti shranjeni v eni tabeli ali pogledu na Oraclevi bazi, ki mora biti standardna relacijska tabela, pri čemer morajo biti stolpci v tabeli takega tipa, ki ga podpira ODM.

Tabelo sem izdelala tako, prikazuje migracijsko matriko. Ideja je bila preveriti, kako se občine izvora oziroma občine prebivališča grupirajo skupaj po občinah ponora oziroma delovnega mesta. Občine izvora postanejo objekti predstavljeni z vrstico v tabeli, stolpci v matriki pa postanejo spremenljivke. Vrednosti posamezne spremenljivke so dnevni migranti iz različnih občin prebivališča v določeno občino delovnega mesta.

Da bi utežili vrednosti posamezne spremenljivke oziroma zmanjšali pomen majhnih oziroma ničelnih vrednosti, velike vrednosti pa stisnili skupaj (zmanjšamo variacijo), zato sem podatke najprej logaritmiral in jih potem transformiral v prej omenjeno matriko oziroma tabelo, kjer so vrstice občine prebivališča, stolpci pa spremenljivke po katerih se združujejo občine prebivališč. Tabela ima tako 210 vrstic in 210 stolpcev oziroma spremenljivk, na podlagi katerih izvedemo metodo razvrščanja v skupine.

Slika 29: Vhodna tabela za razvrščanje v skupine

Structure		Data																	
Fetch Size: 210		Fetch Next		Refresh															
OB	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	V11	V12	V13	V14	V15	V16	V17	V18	V19
1	4366	1	0	0	1	4	34	5	7	0	3	4	8	13	0	0	1	0	0
2	0	760	0	0	0	0	0	0	0	29	0	0	0	0	81	0	0	0	0
3	0	0	1273	272	0	1	0	2	2	1	2	5	0	1	0	0	1	0	0
4	1	1	38	1005	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
5	0	0	0	0	267	0	0	13	0	0	0	0	8	3	0	0	0	0	0
6	3	0	0	0	0	824	1	0	0	0	1	0	0	1	0	0	0	0	0
7	8	0	0	0	0	0	433	0	0	0	0	0	0	0	0	0	0	0	0
8	0	0	2	0	31	1	0	964	4	0	7	1	22	0	0	0	0	5	0
9	0	1	1	1	0	0	0	1	5300	0	14	0	0	0	1	1	12	0	0
10	0	9	0	0	0	0	0	1	0	215	0	0	0	0	1	0	0	0	0
11	17	7	5	2	3	1	1	21	55	3	13945	3	27	1	2	2	9	0	0
12	1	0	21	3	2	0	0	3	3	0	7	736	2	0	0	0	1	0	0
13	7	0	0	0	12	0	0	8	0	0	0	0	2294	6	0	0	1	0	0
14	0	0	0	0	0	2	0	0	0	0	0	0	1	1307	0	0	0	0	0
15	0	42	0	0	0	0	0	0	0	2	1	0	0	0	384	0	0	0	0
16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	405	0	0	0
17	1	0	0	0	0	0	0	0	6	0	1	0	0	0	0	0	3760	0	0
18	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
19	4	0	0	2	0	0	1	0	1	0	1	0	11	0	0	0	0	0	0
20	0	0	1	0	0	0	0	1	4	0	0	1	1	0	0	0	16	0	0
21	0	0	0	0	9	0	0	17	1	0	5	0	3	0	0	0	1	0	0
22	1	1	2	0	1	0	0	6	0	0	3	2	1	0	0	0	1	0	0
23	8	1	6	3	7	1	3	29	14	0	29	34	9	1	0	2	5	0	0
24	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

4.3.4 Modeliranje

Za grupiranje občin na podlagi dnevnih migracij izberemo metodo razvrščanja v skupine, algoritem k-voditeljev.

Metoda voditeljev k-means je nehierarhična ali optimizacijska metoda razvrščanja v skupine, ki v vsakem koraku se izračuna določeno kriterijsko funkcijo in poskuša z drugačno razvrstitvijo doseči, da se zmanjša vrednost kriterijske funkcije. Optimizacijske metode izhajajo iz vnaprej določenega števila skupin k , to število mora določiti uporabnik.

Računalnik izbere k točk slučajno, te točke so začetni voditelji. Vsako enoto doda najbližjemu voditelju. Ko so vse enote dodane in s tem narejene nove skupine, se izračuna težišča novonastalih skupin, ki postanejo novi voditelji. Ta postopek se ponavlja, dokler se voditelji ne premikajo več.

Dobljena razvrstitev je odvisna od začetne pozicije voditeljev, zato se ta postopek velikokrat ponavlja z različnimi začetnimi konfiguracijami voditeljev. Kot “najboljšo” razvrstitev se vzame tisto, ki ima najmanjšo vrednost kriterijske funkcije (Košmelj & Breskvar Žaucer, 2006).

Običajno se kot merilo razdalje uporablja kvadrat evklidske razdalje. ODM omogoča tudi kosinus razdaljo, ki podobnost med spremenljivkama meri tako, da izračuna kot med vektorjema.

$$Podobnost(A,B) = \cos(\theta) = \frac{A \cdot B}{|A| |B|} \quad (24)$$

Ko se kot manjša, se kosinus približuje 1, kar pomeni, da ko se vektorja približujeta, se podobnost povečuje.

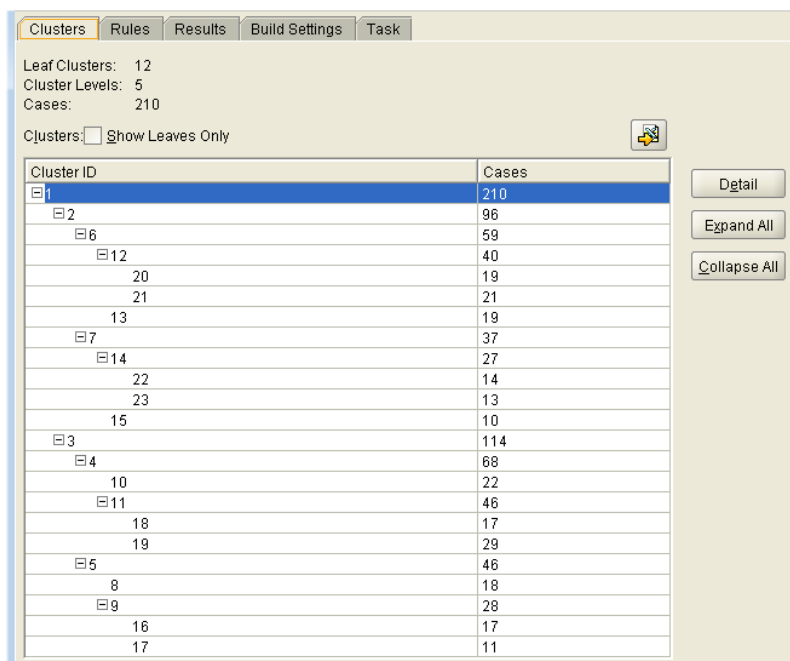
Algoritem k-Means v ODM ima izbiro dva kriterija razporejanja enot, varianco in velikost. Varianca meri razpršitev točk okrog centroida skupine (voditelja) in meri, kako dobro je voditelj kot težišče umeščen med te točke. Izračuna se tako, da izmeri razdaljo posamezne točke do težišča. Skupina z največjo razpršenostjo je kandidatka za razdelitev. Velikost skupine je preprosto število točk znotraj skupine, kandidatka za razdelitev pa je skupina, ki ima največ točk oziroma največ enot.

Za izdelavo modela, ki bo občine razvrstil v skupine izberemo tabelo s podatki, primarni ključ tabele, ki določa unikatnost posameznega zapisa in izberemo stolpce, ki predstavljajo spremenljivke, na katerih naj se izvede algoritem.

Poleg teh nastavitvev določimo še, koliko skupin želimo ter izberemo funkcijo razdalje in kriterij delitve v skupine. Določimo tudi vrsto normalizacije, ki se izvede pred izvedbo algoritma, saj je normalizacija za razvrščanje v skupine pomembna in je koristno, da se izvede pred samim razvrščanjem.

Z metodo se pred izgradnjo modela avtomatično izvede obdelava izjem oziroma odstopajočih vrednosti, obdelava manjkajočih vrednosti in normalizacija. Rezultat obdelave oziroma model je prikazan na sliki 29 in prikazuje razdelitve in končne skupine. Če kliknemo posamezno skupino, lahko vidimo tudi histogram njenih atributov. V zavihku pravil pa so prikazana pravila, na podlagi katerih je bila oblikovana posamezna skupina. Na prvi pogled lahko vidimo, da so se v občine v skupine porazdelile dokaj enakomerno glede na število občin v posamezni skupini, tj. med skupinami ni ogromnih razlik.

Slika 30: Model za razvrščanje skupin s k-Means metodo



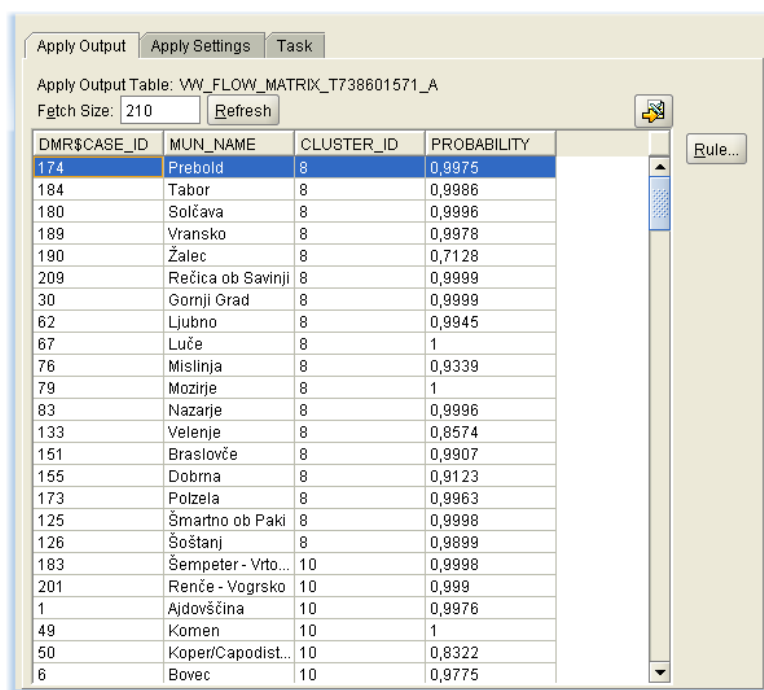
Cluster ID	Cases
1	210
2	96
6	59
12	40
20	19
21	21
13	19
7	37
14	27
22	14
23	13
15	10
3	114
4	68
10	22
11	46
18	17
19	29
5	46
8	18
9	28
16	17
17	11

4.3.5 Uporaba in vrednotenje

Sam model in pregled pravil, kako so se občine razvrstile v posamezno skupino nam zaradi množice atributov, ne povejo veliko, zato model uporabimo. Model uporabimo tako, da nanj apliciramo podatke na katerih smo zgradili model in po potrebi dodamo polja, ki označujejo objekte, na primer naziv občine.

Kot rezultat dobimo tabelo z naslednjimi stolpci: šifra enote, ime enote (občine), šifra skupine v katero je bila razvrščena občina ter s kakšno verjetnostjo je bila občina vključeno v skupino. Rezultat dobimo prikazan v Oracle Data Minerju in je prikazan na sliki 30. Rezultati se zapišejo v tabelo na bazi, lahko pa jih tudi izvozimo v Excel ali shranimo pod nekim drugim imenom za kasnejše obdelave.

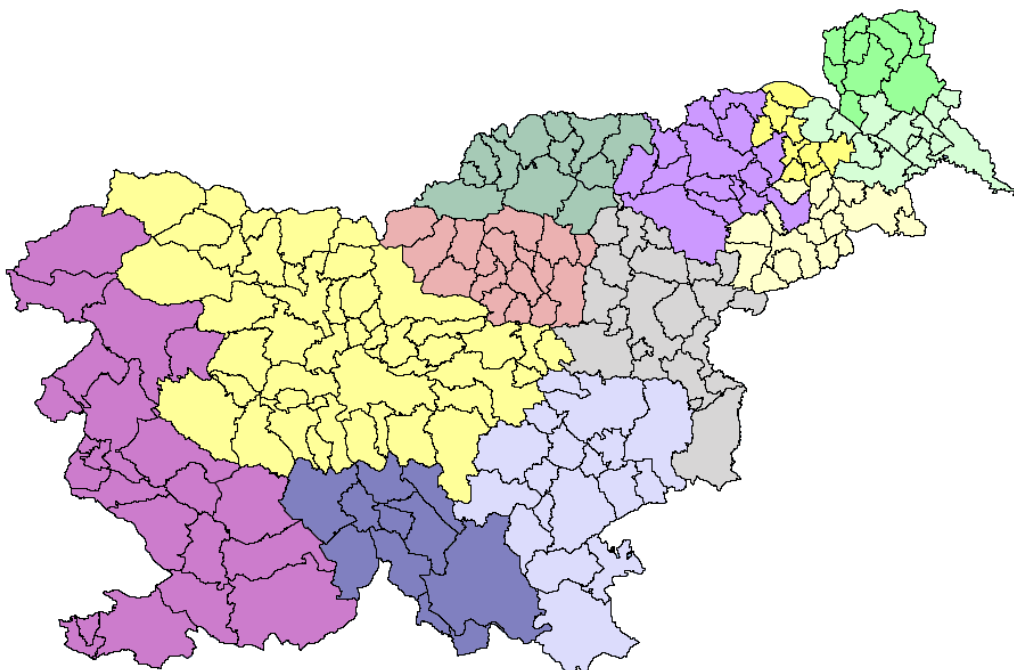
Slika 31: Rezultat razvrščanja v skupine z metodo k-Means



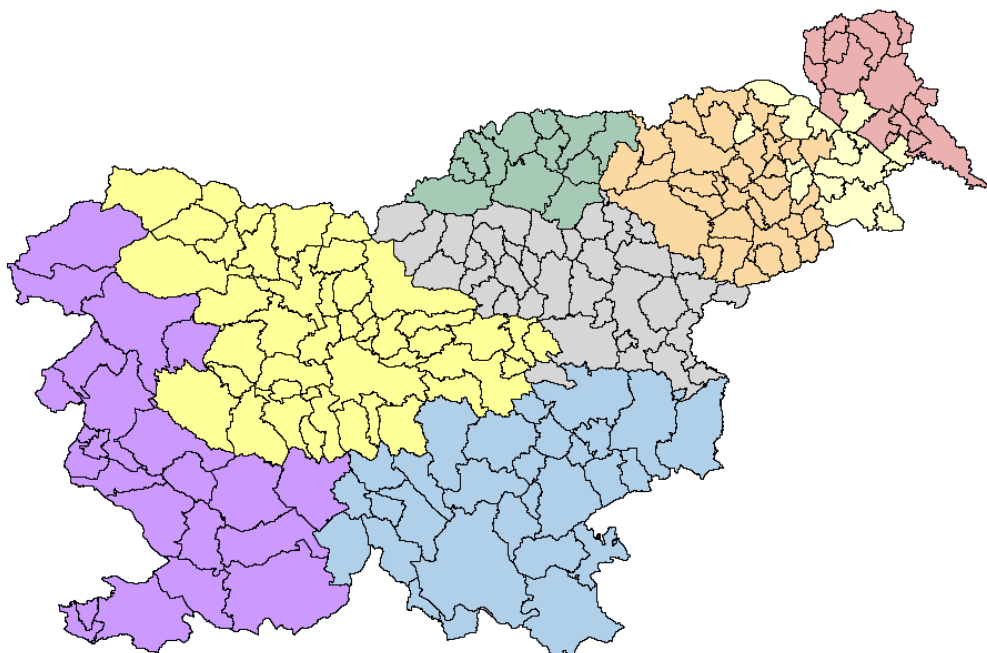
DMR\$CASE_ID	MUN_NAME	CLUSTER_ID	PROBABILITY
174	Prebold	8	0,9975
184	Tabor	8	0,9986
180	Solčava	8	0,9996
189	Vransko	8	0,9978
190	Žalec	8	0,7128
209	Rečica ob Savinji	8	0,9999
30	Gornji Grad	8	0,9999
62	Ljubno	8	0,9945
67	Luče	8	1
76	Mislinja	8	0,9339
79	Mozirje	8	1
83	Nazarje	8	0,9996
133	Velenje	8	0,8574
151	Braslovče	8	0,9907
155	Dobrna	8	0,9123
173	Polzela	8	0,9963
125	Šmartno ob Paki	8	0,9998
126	Šoštanj	8	0,9899
183	Šempeter - Vrto...	10	0,9998
201	Renče - Vogrsko	10	0,999
1	Ajdovščina	10	0,9976
49	Komen	10	1
50	Koper/Capodist...	10	0,8322
6	Bovec	10	0,9775

Rezultat obdelave je bolje viden na zemljevidu, ki ga izdelamo v Oracle Mapiewerju in je prikazan na sliki 31. Na spodnji sliki je prikazan rezultat obdelave, ko smo izbrali 12 skupin, min-max metodo normalizacije, kosinus razdaljo in varianco kot kriterij delitve skupin. Slika je izdelana na podlagi rezultata razvrščanja v skupine in prostorskih podatkov o občini, ki je shranjen v polju Geometrija v tabeli občin.

Slika 32: Prikaz rezultata razvrščanja v skupine z metodo k-Means za 12 regij



Slika 33: Prikaz rezultata razvrščanja v skupine z metodo k-Means za 8 regij



Z metodo k-Means smo dobili skupine občin, ki naj bi funkcionalno spadale skupaj. Verjetnost za vključitev občine v posamezno regijo so precej visoke, za vse občine razen dveh, je verjetnost za vključitev v posamezno regijo višja od 70 odstotkov.

Občine, ki so vključene v funkcionalne regije so sosednje, torej so regije prostorsko homogene, kar nakazuje, da je razvrstitev dovolj dobra.

Opazimo lahko, da so se občine okrog Murske Sobote razvrstile v dve regije. Razlogov za to je lahko več. Gre za majhne občine z zelo velikim deležem kmečkega prebivalstva, ki je iz podatkov Statističnega urada o dnevnih migrantih izvzeto. Te občine imajo od 20 do celo 70 odstotkov kmečkega prebivalstva, drugače pa v teh občinah ni zaposlitvenih možnosti in so zelo visoki deleži delovnih migrantov, ki se gibljejo okrog 80 odstotkov. Dve od občin, Moravske Toplice in Hodoš imata zelo velik delež zaposlenih v storitveni dejavnosti, prva je imela tudi najnižjo verjetnost za vključitev v svojo regijo.

Poleg tega so podatki za leto 2007, ko še ni bila izgrajena avtocesta, ki je povezala Prekmurje in je bilo do tedaj to območje prometno zelo slabo razvito. Rezultat je lahko tudi posledica tega, da je na razmeroma majhnem območju več manjših bolj razvitih občin: poleg Murske Sobote še Gornja Radgona, Lendava, Ljutomer in Ormož.

Občina Brežice, ki ima tudi zelo visok delež zaposlenih v storitveni dejavnosti je pri razmejitvi na 12 regij razvrščena v Celjsko, pri razdelitvi na 8 regij pa je povezana s Krškim. Edina regija, ki je v vseh razvrstitvah ostajala enaka je Koroška, ki je očitno toliko ločena od vseh ostalih območij, da bi morala biti v vsakem primeru ločena regija.

SKLEP

Spremembe v socialnih, ekonomskih in demografskih spremenljivkah so posledica vrste različnih interakcij na lokalni, regionalni, državni in meddržavni ravni. Razvojni mehanizmi so močno odvisni od lokalnega okolja. Kompleksna igra različnih dejavnikov, od podjetij, ki poganjajo gospodarsko rast do delovanja socialnih institucij, se odvija v prostoru, interakcije pa so omejene z razdaljo. To pomeni, da je izbor geografske enote za katero se oblikuje ali nameni določena razvojna politika ključna determinanta za njen uspeh. Zato da bi odsevale specifične lokalnega okolja, naj bi bile politike usmerjene na področja, ki so homogena.

V zadnjih nekaj desetletjih je postalo očitno, da administrativno določena območja, ki jih uporabljajo države za oblikovanje razvojnih politik, za alokacijo virov in raziskave, ne podajajo več najustreznejših informacij o dejanskih pogojih določenega območja ali regije. Zato je prišlo do določitve in razmejevanja funkcionalnih regij, ki so v praksi splošno opredeljene z lokalnimi zaposlitvenimi sistemi. Uporaba prostorske opredelitve lokalnih trgov dela kot funkcionalnih regij je posebnega pomena iz več razlogov, med drugim so sredstvo za predstavitev informacij in statistik o zaposlitvenih in socialno-ekonomskih strukturah ter so ustrezne za oceno učinkovitosti regionalnih politik in za reorganizacijo lokalne uprave.

Vzorci dnevnih migracij na delo so splošno uporabljani kot eden najpomembnejših kazalnikov velikosti lokalnega trga dela in zato so rezultati pogosto uporabljeni za

oblikovanje funkcionalnih regij. Kot taka je funkcionalna regija pogosto opredeljena kot regija v kateri se zaposli večina lokalnega prebivalstva in večina podjetij najde zaposlene.

Funkcionalne regije se z razvojem in pod vplivom različnih politik ter drugih dejavnikov spreminjajo, zato lahko s spreminjanjem scenarijev predvidimo, kako se bodo spreminjale interakcije v prostoru. Z modeliranjem na podlagi različnih scenarijev lahko ugotavljamo, kako se bodo spremembe v prometni infrastrukturi, zaposlovanju, investicijah ipd. odrazile na spreminjanje interakcij in s tem, kako se bodo oblikovale funkcionalne regije.

Cilj magistrskega dela je bila zasnova informacijskega sistema za podporo odločanju, ki bo omogočil modeliranje regij. Na podlagi poznavanja vsebine problema, ki je obravnavana v prvih dveh poglavjih magistrskega dela, je bila oblikovana specifikacija funkcionalnih zahtev, ki naj jih sistem za modeliranje funkcionalnih regij izpolni. Na podlagi zahtev je bila oblikovana arhitektura informacijskega sistema, ki jo sestavljajo trije gradniki: podatkovna baza, podatkovno rudarjenje in vizualizacija prostorskih podatkov.

Podatki se shranjujejo v prostorski podatkovni bazi v relacijskih tabelah. Strukturo podatkov določa izdelan model entitet in povezav, ki med drugim določa, da so podatki in kazalniki ločeni po vsebinskih področjih. Podatkovni model je zastavljen tako, da je razširljiv in omogoča enostavno dodajanje in shranjevanje novih podatkov. Podatke je mogoče enostavno poizvedovati, saj so poleg vsebinske členitve ločeni tudi po obdobjih, prostorskih ravneh ter po virih in tipih podatkov. Uporabniki do podatkov lahko dostopajo preko grafičnih orodij za delo s podatkovno bazo ali preko grafičnega vmesnika za podatkovno rudarjenje. Podatki se lahko uvozijo v bazo na različne načine, tudi iz datoteke Microsoft Excel. Podatkovna baza je zmogljiva in omogoča procesiranje velikih količin podatkov.

Modeliranje se izvaja s procesom podatkovnega rudarjenja, ki se izvaja neposredno na podatkovni bazi, zato so podatki vedno ažurni in jih ni potrebno premeščati, izvajanje podatkovnega rudarjenja pa je zato enostavno in časovno učinkovito. Podatkovno rudarjenje vsebuje zbirko algoritmov, s katerimi se izdelajo modeli, ki rešujejo različne probleme. Uporabnik komunicira z bazo in izvaja proces podatkovnega rudarjenja preko interaktivnega grafičnega vmesnika. Izdelava novih modelov in njihova uporaba sta enostavni, testna metrika pa omogoča ovrednotenje modelov. S spreminjanjem vhodnih podatkov ali parametrov modela, uporabnik izdeluje nove modele brez dodatnih stroškov. Modeli in rezultati se shranijo v podatkovni bazi, lahko pa se tudi izvozijo, na primer v Excel. Glede na dodeljene pravice, ki jih določa varnost, postavljena na bazi, uporabnik lahko tako vedno pregleduje svoje modele ali modele drugih uporabnikov in jih kadar koli uporabi za analizo.

Informacijski sistem omogoča integracijo z drugimi aplikacijami, saj gre dejansko za integracijo podatkovne baze z aplikacijo. Ker baza omogoča upravljanje prostorskih podatkov, jo je mogoče povezati s standardnimi GIS-orodji, ki omogočajo vizualizacijo prostorskih podatkov.

Tako zasnovan sistem je bil uporabljen za modeliranje na podatkih za Slovenijo. Kot prostorska podatkovna baza je bil uporabljen Oracle Spatial (11g). Modeliranje je bilo izvedeno preko grafičnega vmesnika Oracle Data Miner, ki uporabniku, ki ni večč dela s funkcijami na bazi, olajša podatkovno rudarjenje. Za prikaz prostorskih podatkov na zemljevidu je bilo uporabljeno orodje Oracle MapViewer. Vsa tri orodja so komercialna, vendar pa bi lahko enak informacijski sistem zasnovali tudi z brezplačnimi orodji.

Informacijski sistem je bil uporabljen za napovedovanje dnevnih migracij zaposlenih med slovenskimi občinami. Razširjeni gravitacijski model dnevnih migracij je bil ocenjen z metodo posplošenih najmanjših kvadratov in metodo podpornih vektorjev. Z metodo k-Means so bile občine nato na podlagi dnevnih migracij razvrščene v funkcionalne regije.

Tako oblikovan informacijski sistem omogoča modeliranje funkcionalnih regij na podlagi različnih scenarijev in zadošča vnaprej postavljenim zahtevam. Primernost posameznega algoritma, z njim dobljenega modela ter rezultatov je tudi vsebinski problem, zato ga je potrebno ovrednotiti.

LITERATURA IN VIRI

1. Andersson, M., & Ejermo, O. (2002). *Knowledge Production in Swedish Functional Regions 1993–1999*. Jönköping: Jönköping International Business School. Najdeno 1. junija 2009 na spletnem naslovu ftp://ftp.unibocconi.it/pub/RePEc/cri/papers/CESPRI_wp1391.pdf
2. Andersson, M., Gråsjö, U., & Karlsson, C. (2006). *Industry R&D Location – the role of accessibility to university R&D and institutions of higher education*. Stockholm: Royal Institute of Technology. Najdeno 1. junija 2009 na spletnem naslovu <http://www.infra.kth.se/cesis/documents/WP68.pdf>
3. Bogataj, L., Bogataj, M., Drobne, S., & Lisec, A. (2008). *Razvoj modela za potrebe regionalne politike* (zaključno poročilo). Projekt št. V5-0304. Ljubljana: CRP »Konkurenčnost Slovenije 2006–2013«.
4. Bogataj, M., & Drobne, S. (1997). The influence of investments in highways on gravity and interaction in Slovenia. V Rupnik V., Zadnik Stirn L., & Drobne S. (ur.). *Proceedings of the 4th International Symposium on Operational Research in Slovenia* (str. 55–60). Ljubljana: Slovenian Society Informatika.
5. Bogataj, M., & Drobne, S. (2005). Does the improvement of roads increase the daily commuting? Numerical analysis of Slovenian interregional flows, V Zadnik Stirn L., Indihar Štemberger M., Ferbar L., & Drobne S. (ur.). *Selected Decision Support Models for Production and Public Policy Problems* (str. 185–206). Ljubljana: Slovenian Society Informatika.
6. Bond, S., & Coombes, M. (2007). *2001-based Travel-To-Work Areas Methodology*. London: Office for national statistics. Najdeno 1. junija 2009 na spletnem naslovu <http://www.statistics.gov.uk/geography/ttwa.asp>
7. Castells, M. (1999). The Culture of Cities in Information Age, V Susser I. (ur.) (2002). *The Castells Reader on Cities and Social Theory* (str. 367–389). Oxford: Blackwell Publishers.
8. Cheshire, P. (1997). *Economic Indicators for European Cities and Regions: Why Boundaries Matter*. London: London School of Economics. Najdeno 7. junija 2009 na spletnem naslovu http://www.perfilciutat.net/fixters/Cheshire_1997.pdf
9. Cheshire P. (2007). *Optimal Areas for planning, local economic development and transportation*, London: London School of Economics. Najdeno 7. junija 2009 na spletnem naslovu http://www.cers.tuke.sk/cers2007/PDF/A_Cheshire.pdf
10. Cörvers, F., Hensen, M., & Bongaerts, D. (2009). The Delimitation and Coherence of Functional and Administrative Regions, *Regional Studies*, 43(1), 19–31.
11. CRISP-DM (2000). CRISP-DM 1.0, Step-by-step data mining guide. Najdeno 5. oktobra 2009 na spletnem naslovu <http://www.crisp-dm.org>
12. Davis, W. S., & Yen, D. C. (1999). *The Information System Consultant's Handbook, System Analysis and Design*. Boca Raton: CRC Press.

13. Database design basics (n.d.). Database design basics. Microsoft Corporation. Najdeno 1. oktobra 2009 na spletnem naslovu <http://office.microsoft.com/en-us/access/HA012242471033.aspx>
14. De Goei, Burger M., Van Oort F., & Kitson M. (2008). Testing the 'Super-Region', *Town & Country Planning*, 77 (11), 458–464.
15. Drobne, S., & Bogataj, M. (2005). Intermunicipal gravity model of Slovenia. V Zadnik Stirn, M., Drobne, S. (ur.). *Proceedings of the 8th International Symposium on Operational Research in Slovenia* (str. 207–212). Ljubljana: Slovenian Society Informatika.
16. Duque, J. C., & Ramos, R. (2004). *Design of homogenous territorial units: A methodological proposal* (Working Papers in Economics, 115). Barcelona: Universitat de Barcelona.
17. ESPON Project 1.2.3 (2007). *Identification of Spatially Relevant Functions of the Information Society*, Final Report. IGEAT – Institut de Gestion de l'Environnement et d'Aménagement du Territoire. Brussels: Université Libre de Bruxelles (Lead Partner). Najdeno 7. junija 2009 na spletnem naslovu <http://www.espon.lu>
18. ESPON Project 1.4.3 (2007). *Study on Urban Functions*, Final Report. IGEAT – Institut de Gestion de l'Environnement et d'Aménagement du Territoire. Brussels: Université Libre de Bruxelles (Lead Partner). Najdeno 7. junija 2009 na spletnem naslovu <http://www.espon.lu>
19. ESPON Project 1.4.4 (2007). *Preparatory Study on Feasibility of Flows Analysis*, Final Report. IGEAT – Institut de Gestion de l'Environnement et d'Aménagement du Territoire. Brussels: Université Libre de Bruxelles (Lead Partner). Najdeno 7. junija 2009 na spletnem naslovu <http://www.espon.lu>
20. ESPON Project 3.1 (2005). *Integrated Tools for European Spatial Development*, Final report. IGEAT – Institut de Gestion de l'Environnement et d'Aménagement du Territoire. Brussels: Université Libre de Bruxelles (Lead Partner). Najdeno 7. junija 2009 na spletnem naslovu <http://www.espon.lu>
21. ESPON Project 3.2 (2007). *Scenarios on the Territorial Future of Europe*, Final Report. IGEAT – Institut de Gestion de l'Environnement et d'Aménagement du Territoire. Brussels: Université Libre de Bruxelles (Lead Partner). Najdeno 20. junija 2009 na spletnem naslovu <http://www.espon.lu>
22. ESPON Project 3.4.3 (2006). *The Modifiable Areas Unit Problem*, Final Report. IGEAT – Institut de Gestion de l'Environnement et d'Aménagement du Territoire. Brussels: Université Libre de Bruxelles (Lead Partner). Najdeno 7. junija 2009 na spletnem naslovu <http://www.espon.lu>
23. Faggian, A., & McCann, P. (2008). *Human Capital, Graduate Migration and Innovation in British Regions*. Hamilton: University of Wakaito. Najdeno 2. julija 2009 na spletnem naslovu <http://www.nzae.org.nz/conferences/2008/110708/nr1215389794.pdf>
24. Feldman, O. , Simmonds, D., Troll, N., & Tsang, F. (2005). *Creation of a System of Functional Areas for England and Wales and for Scotland*. London: David Simmonds Consultancy. Najdeno 1. junija 2009 na spletnem naslovu <http://www.mvaconsultancy.com>
25. Foot, D. (1981). *Operational Urban Models, An Introduction*. London: Routledge.

26. Fotheringham, S.A., Brunson, C., & Charlton, M. (2002). *Geographically Weighted Regression: The Analysis of Spatially Varying Relationships*. Chichester: Wiley
27. GEMACA II (2001). *Science and Technology in the Functional Urban Regions in Northwest Europe*. Paris: IAURIF. Najdeno 7. julija 2009 na spletnem naslovu http://www.iau-idf.fr/fileadmin/Etudes/etude_495/science_and_techno_E2_01.pdf
28. Geodetska uprava Republike Slovenije. (2009). Register prostorskih enot [zgoščenka] Ljubljana: Geodetska uprava Republike Slovenije
29. Haining, R. (2003). *Spatial Data Analysis: Theory and Practice*. Cambridge: Cambridge University Press.
30. Han, J., & Kamber, M. (2006). *Data Mining Concepts and Techniques* (2nd ed.). San Francisco: Elsevier. The Morgan Kaufmann Series in Data Management Systems
31. Haynes, K. E., & Fotheringham, S. (1984). *Gravity and spatial interaction models*. Beverly Hills: Sage Publications. Scientific Geography Series.
32. Henry, M. S. (1999). Spatial labour markets, new economic geography, and urban-rural linkages: implications for the rural south, *Southern rural sociology* 15, 38–58. Najdeno 3. julija 2009 na spletnem naslovu <http://www.ag.auburn.edu/auxiliary/srsa/pages/Articles/SRS%201999%2015%2038-58.pdf>
33. Johansson, B. (2007). *Transport Infrastructure Inside and Across Urban Regions: Models and Assessment Methods*. OECD Working Papers. OECD/ITF Joint Transport Research Centre Discussion Papers No. 2007/12. Najdeno 5. junija 2009 na spletnem naslovu <http://puck.sourceoecd.org/vl=2788526/cl=20/nw=1/rpsv/cgibin/wppdf?file=5kzbxssxz4vh.pdf>
34. Johansson, B. (1998). *Infrastructure, Market Potential and Endogenous Growth*. Jönköping: Jönköping International Business School.
35. Karlsson, C., & Olsson, M. (2006). The Identification of Functional Regions: theory, methods, and applications. V: *The Annals of Regional Science*, 40 (1), 1–18.
36. Knapp, W. , & Schmitt, P. (2002). Re-structuring Competitive Metropolitan Regions: On Territory, Institutions and Governance, RheinRuhr compared with London, Paris and the Randstad Holland. *Paper no. 437 - ERSA Congress 2002*. Najdeno 1. junija 2009 na spletnem naslovu <http://www-sre.wu-wien.ac.at/ersa/ersaconfs/ersa02/cd-rom/papers/437.pdf>
37. Knapp, W., Scherhag, D., & Schmitt, P. (2005). *Commuting & the definition of functional urban regions*, POLINET Action 1.1, Interreg. Najdeno 1. junija 2009 na spletnem naslovu http://www.polynet.org.uk/docs/1_1_rruhr.pdf
38. Košmelj, K., Breskvar Žaucer, L. (2006). Metode za razvrščanje enot v skupine; osnove in primer. *Acta agriculturae Slovenica*, 87 (2), 299–310
39. Kothuri, R., Godfried, A., & Beinart, E. (2007). *Pro Oracle Spatial for Oracle Database 11g* (2nd ed.). New York: Apress.
40. Linneker, B. (1997). *Transport Infrastructure and Regional Economic Development in Europe: A Review of Theoretical and Methodological Approaches*, TRP 133. Sheffield: Department of Town and Regional Planning.
41. Masser, I., & Scheurwater, J. (1980). Functional regionalisation of spatial interaction data: an evaluation of some suggested strategies , *Environment and Planning A*, 12, 1357–1382.

42. Mcrit (2009). What's an executive support system? Najdeno 15. Septembra 2009 na spletnem naslovu http://www.mcrit.com/assembling/ASSEMB_CENTRAL/WhatESS.htm
43. Mitchell, W., Bill, A., & Watts, M. (2007). *Identifying functional regions in Australia using hierarchical aggregation techniques*. Working paper 07–06. Australia: The University of Newcastle. Najdeno 1. junija 2009 na spletnem naslovu <http://e1.newcastle.edu.au/coffee/pubs/wp/2007/07-06.pdf>
44. Noronha, V. T. (1986). *Functional Regions Reexamined: The Black/White Interaction Model*. Kanada: University of Western Ontario. Najdeno 4. junija 2009 na spletnem naslovu <http://digitalgeographic.com/resources/publications/divide/Noronha-Nystrom1986.pdf>
45. Noronha, V. T., & Goodchild, M. (1992). Modelling interregional interaction: implications for defining functional regions. *Annals of the Association of American Geographers*, 82(1), 86–102.
46. OECD (2002). *Redefining Territories: The Functional Regions*. Paris: OECD Publications.
47. Openshaw, S. (1984) The modifiable areal unit problem, *Concepts and Techniques in Modern Geography*, 38(41).
48. Oracle (2007). *Data mining, Concepts*, 11g Release 1. Redwood City: Oracle Corporation.
49. Oracle (2008a). *Data mining, Developer's Guide*, 11g Release 1. Redwood City: Oracle Corporation.
50. Oracle (2008b). *Data mining, Tutorial*, 11g Release 1. Redwood City: Oracle Corporation.
51. Oracle (2009). *Data mining, Data Sheet*, 11g Release 1. Redwood City: Oracle Corporation.
52. Pogačnik, A., Pichler-Milanović, N., Sitar, M., Lavrač, I., Kobal, J., Peterlin, M., Zavodnik Lamovšek, A., Drobne, S., Lisec, A., Soss, K. & Trobec, B. (2008). *Analiza razvojnih virov in scenarijev za modeliranje funkcionalnih regij* (prvo poročilo). Projekt št. V2-0507. Ljubljana: CRP »Konkurenčnost Slovenije 2006–2013«.
53. Pogačnik, A., Pichler-Milanović, N., Sitar, M., Lavrač, I., Kobal, J., Peterlin, M., Zavodnik Lamovšek, A., Drobne, S., Lisec, A., Soss, K. & Trobec, B. (2009). *Analiza razvojnih virov in scenarijev za modeliranje funkcionalnih regij* (drugo poročilo). Projekt št. V2-0507. Ljubljana: CRP »Konkurenčnost Slovenije 2006–2013«.
54. Prodromídis, P.-I. (2006). *Functional Economies or Administrative Units in Greece: What difference does it make for policy?*. Athens: Centre of Planning & Economic Research. Najdeno 1. junija 2009 na spletnem naslovu <http://www-sre.wu-wien.ac.at/ersa/ersaconfs/ersa06/papers/358.pdf>
55. Ramakrishnan, R., & Gehrke, J. (2002). *Database management systems*. New York: McGraw-Hill Science/Engineering/Math.
56. REPUS (2007). Strategy for a regional Polycentric Urban System in Central-Eastern Europe Economic Integration Zone, Final Report. Budimpešta. Interreg III B CADSES. Najdeno 1. junija 2009 na spletnem naslovu http://www.repus.it/repus-docs/repus_finalreport.pdf
57. Rodrigue, J.-P., Comtois, C., & Slack, B. (2009). *The geography of transport systems* (2nd ed.). New York: Routledge.

58. Study Programme on European Spatial Planning (SPESP) (2000). Final report. Najdeno 1. junija 2009 na spletnem naslovu <http://nordregio.shotcode.no/filer/spespn/welcome.htm>
59. Spiekermann & Wegener Urban and Regional Research (2007). *Update of selected accesibility indicators*. Final Report. Najdeno 6. junija 2009 na spletnem naslovu http://www.mdrl.ro/espon_cd2/Project_Reports/Scientific_briefing_and_networking/Map_Update_final_report.pdf
60. Statistični urad Republike Slovenije. (2009a). Interaktivni statistični atlas Slovenije. Najdeno 2. avgusta 2009 na spletni strani <http://www.stat.si/pxweb/Dialog/statfile2.asp>.
61. Statistični urad Republike Slovenije. (2009b). SI-STAT podatkovni portal. Najdeno 2. avgusta 2009 na spletni strani <http://www.stat.si/pxweb/Dialog/statfile2.asp>.
62. Statistični urad Republike Slovenije. (2010). Slovenske občine v številkah. Najdeno 5. maja 2010 na spletni strani <http://www.stat.si/obcinevstevilkah/>
63. Trapasso, R. (2007). *Complexity in Regional Economics. Theoretical Modelling and Empirical Applications* (doktorsko delo). Piacenza: Università Cattolica del Sacro Cuore Najdeno 3. junija 2009 na spletnem naslovu <http://tesionline.unicatt.it/bitstream/10280/230/7/Tesi%20Trapasso%20integrale.pdf>
64. Svet za decentralizacijo in regionalizacijo (2009). Prvo zasedanje sveta za decentralizacijo in regionalizacijo. Sporočilo za javnost. Najdeno 5. maja 2010 na spletnem naslovu <http://www.svlr.gov.si/nc/si/splosno/cns/novica/article//2022/>
65. Yeung, A. K. W., & Hall, G. B. (2007). *Spatial Database Systems: Design, Implementation and Project Management*. Dordrecht: Springer, The GeoJournal Library.
66. Van der Zwan, J., Van der Wel, R., Van Eck, R., De Jong, T., & Floor, H.(2005). *Flowmap 7.2 Manual*. Utrecht: Utrecht University, Faculty of Geographical Sciences. Najdeno 1. junija 2009 na spletnem naslovu <ftp://ftp.geog.uu.nl/pub/flowmap/FM72manual.pdf>
67. Wang, F., & Guldmann, J.-M. (1996). Simulating urban population density with a gravity based model. *Socio-Economic Planning Sciences*, 30 (4,) 245-256.
68. Wikipedia (b.l.). Spatial Analysis. Wikipedia, spletna enciklopedija. Najdeno na spletni strani 7. junija 2010 na spletnem naslovu http://en.wikipedia.org/wiki/Main_Page