

UNIVERZA V LJUBLJANI
EKONOMSKA FAKULTETA

MAGISTRSKO DELO

**OCENJEVANJE TRGA GOSPODARSKIH VOZIL V SLOVENIJI Z
UPORABO NAPOVEDI EKONOMSKIH SPREMENLJIVK**

Ljubljana, avgust 2016

DAMJAN CIRMAN

IZJAVA O AVTORSTVU

Podpisani Damjan Cirman, študent Ekonomske fakultete Univerze v Ljubljani, avtor predloženega dela z naslovom Ocenjevanje trga gospodarskih vozil v Sloveniji z uporabo napovedi ekonomskih spremenljivk, pripravljenega v sodelovanju s svetovalcem prof. dr. Markom Pahorjem

IZJAVLJAM

1. da sem predloženo delo pripravil samostojno;
2. da je tiskana oblika predloženega dela istovetna njegovi elektronski obliki;
3. da je besedilo predloženega dela jezikovno korektno in tehnično pripravljeno v skladu z Navodili za izdelavo zaključnih nalog Ekonomske fakultete Univerze v Ljubljani, kar pomeni, da sem poskrbel, da so dela in mnenja drugih avtorjev oziroma avtoric, ki jih uporabljam oziroma navajam v besedilu, citirana oziroma povzeta v skladu z Navodili za izdelavo zaključnih nalog Ekonomske fakultete Univerze v Ljubljani;
4. da se zavedam, da je plagiatorstvo – predstavljanje tujih del (v pisni ali grafični obliki) kot mojih lastnih – kaznivo po Kazenskem zakoniku Republike Slovenije;
5. da se zavedam posledic, ki bi jih na osnovi predloženega dela dokazano plagiatorstvo lahko predstavljalo za moj status na Ekonomski fakulteti Univerze v Ljubljani v skladu z relevantnim pravilnikom;
6. da sem pridobil vsa potrebna dovoljenja za uporabo podatkov in avtorskih del v predloženem delu in jih v njem jasno označil;
7. da sem pri pripravi predloženega dela ravnal v skladu z etičnimi načeli in, kjer je to potrebno, za raziskavo pridobil soglasje etične komisije;
8. da soglašam, da se elektronska oblika predloženega dela uporabi za preverjanje podobnosti vsebine z drugimi deli s programsko opremo za preverjanje podobnosti vsebine, ki je povezana s študijskim informacijskim sistemom članice;
9. da na Univerzo v Ljubljani neodplačno, neizključno, prostorsko in časovno neomejeno prenašam pravico shranitve predloženega dela v elektronski obliki, pravico reproduciranja ter pravico dajanja predloženega dela na voljo javnosti na svetovnem spletu preko Repozitorija Univerze v Ljubljani;
10. da hkrati z objavo predloženega dela dovoljujem objavo svojih osebnih podatkov, ki so navedeni v njem in v tej izjavi.

V Ljubljani, dne _____

Podpis študenta: _____

KAZALO

UVOD.....	1
1 ZNAČILNOSTI TRGA GOSPODARSKIH VOZIL V SLOVENIJI	3
1.1 Predstavitev segmentacije trga gospodarskih vozil	5
1.2 Predstavitev segmenta lahkih dostavnih vozil.....	6
1.3 Predstavitev segmenta težkih tovornih vozil.....	8
1.4 Predstavitev segmenta avtobusov	11
2 PREDSTAVITEV PODATKOVNEGA RUDARJENJA IN NAPOVEDNE ANALITIKE	13
2.1 Proces podatkovnega rudarjenja.....	17
2.1.1 Predhodno znanje	19
2.1.2 Priprava	19
2.1.3 Modeliranje	20
2.1.4 Uporaba	22
2.1.5 Znanje.....	22
2.2 Pregled algoritmov napovedne analitike	23
2.2.1 Klasifikacijski algoritmi.....	23
2.2.1.1 Odločitvena drevesa.....	23
2.2.1.2 Učenje pravil.....	24
2.2.1.3 k-Najbližji sosed	24
2.2.1.4 Naivna Bayesova metoda.....	24
2.2.1.5 Umetne nevronske mreže.....	24
2.2.1.6 Učenje napovednih ansamblov	24
2.2.1.7 Metoda podpornih vektorjev	25
2.2.2 Regresijski algoritmi	25
2.2.2.1 Linearna regresija.....	26
2.2.2.2 Logistična regresija.....	29
2.2.3 Napovedovanje časovnih vrst	29
2.2.3.1 Metode napovedovanja, osnovane na podatkih	29
2.2.3.2 Metode osnovane na modelih	30
2.2.4 Metode izbora vhodnih spremenljivk.....	32
2.2.4.1 Metode filtrirnega tipa	33
2.2.4.2 Metode ovojnega tipa.....	33
2.3 Uporaba napovedne analitike za napovedovanje prodaje v avtomobilski industriji.....	34
2.3.1 Uporaba napovedi na osnovi analize časovnih vrst v avtomobilski industriji	35
2.3.2 Ekonometrični modeli.....	35
2.3.3 Sprejemljivostni modeli	35
2.3.4 Dolgoročne napovedi	36

3	PREDSTAVITEV UPORABLJENIH PODATKOV IN METODOLOGIJE	37
3.1	Predstavitev podatkov	37
3.1.1	Podatki registriranih vozilih	37
3.1.2	Nabor vhodnih spremenljivk	37
3.1.3	Napovedi vhodnih spremenljivk	41
3.2	Kratek opis uporabljene metodologije	42
3.2.1	Izbor vhodnih spremenljivk	42
3.2.2	Kratek opis uporabljenih procesov modeliranja	42
3.2.3	Vrednotenje dobljenih rezultatov posameznih regresijskih modelov	45
3.2.3.1	R^2 – determinacijski koeficient	45
3.2.3.2	RMSE – kvadratni koren povprečne kvadratne napake	46
3.2.3.3	Razpon napake	46
3.2.3.4	Povprečje napake	46
3.2.3.5	Histogram porazdelitve napake e_i	46
3.2.3.6	MAPE – Povprečje absolutne vrednosti relativne napake	47
3.2.3.7	Matrika korelacijskih koeficientov med vhodnimi spremenljivkami	47
3.2.3.8	VIF – faktor napihnjenosti variance	47
4	REZULTATI REGRESIJSKIH MODELOV ZA POSAMEZNE SEGMENTE	47
4.1	Lahka dostavna vozila	47
4.1.1	Rezultati z uporabo letnega nabora podatkov	47
4.1.2	Rezultati z uporabo četrletnega nabora podatkov	49
4.1.3	Rezultati z uporabo četrletnega nabora podatkov na letni ravni	53
4.1.4	Povzetek ugotovitev modeliranja segmenta lahkih dostavnih vozil	55
4.2	Težka tovorna vozila	56
4.2.1	Rezultati z uporabo letnega nabora podatkov	56
4.2.2	Rezultati z uporabo četrletnega nabora podatkov	58
4.2.3	Rezultati z uporabo četrletnega nabora podatkov na letni ravni	61
4.2.4	Povzetek ugotovitev modeliranja segmenta težkih tovornih vozil	63
4.3	Avtobusi	64
4.3.1	Rezultati z uporabo letnega nabora podatkov	64
4.3.2	Rezultati z uporabo četrletnega nabora podatkov	67
4.3.3	Rezultati z uporabo četrletnega nabora podatkov na letni ravni	70
4.3.4	Povzetek ugotovitev modeliranja segmenta avtobusov	71
4.4	Povzetek rezultatov in ugotovitev	72
	SKLEP	74
	LITERATURA IN VIRI	75
	PRILOGE	

KAZALO TABEL

Tabela 1: Nabor uporabljenih vhodnih spremenljivk	38
Tabela 2: Pregled nabora vhodnih spremenljivk in uporabljenih enot	39
Tabela 3: Slamnate sezonske spremenljivke in njihove vrednosti	40
Tabela 4: Primerjava rezultatov regresijskih modelov za segment lahkih dostavnih vozil	55
Tabela 5: Primerjava rezultatov regresijskih modelov za segment težkih tovornih vozil	64
Tabela 6: Primerjava rezultatov regresijskih modelov za segment avtobusov	72
Tabela 7: Primerjava rezultatov modelov za posamične segmente trga gospodarskih vozil	73

KAZALO SLIK

Slika 1: Deleži registracije gospodarskih vozil v letu 2014 po regijah (v tisočih)	4
Slika 2: Prve registracije gospodarskih vozil znotraj regij (v tisočih)	4
Slika 3: Stanje registriranih in prvič registriranih lahkih dostavnih vozil	7
Slika 4: Pregled prvič registriranih lahkih dostavnih vozil	8
Slika 5: Stanje registriranih in prvič registriranih težkih tovornih vozil	10
Slika 6: Pregled prvič registriranih težkih tovornih vozil v obdobju 1995–2015	11
Slika 7: Stanje registriranih in prvič registriranih avtobusov v obdobju 1995–2015	12
Slika 8: Pregled prvič registriranih avtobusov v obdobju 1995–2015	13
Slika 9: Področja uporabe podatkovnega rudarjenja	14
Slika 10: Viri podatkov za podatkovno rudarjenje	15
Slika 11: Napovedna analitika in modeliranje	15
Slika 12: Metodološko drevo metod napovedovanja	16
Slika 13: Faze CRISP-DM referenčnega modelov podatkovnega rudarjenja	18
Slika 14: Proces podatkovnega rudarjenja	19
Slika 15: Model z uporabo učenja napovednih ansamblov	25
Slika 16: Anscombejev kvartet	28
Slika 17: Primer uporabe različnih metod napovedovanja osnovanih na podatkih	30
Slika 18: Primer uporabe metod osnovanih na modelih	31
Slika 19: Splošni pregled metod izbora vhodni spremenljivk	32
Slika 20: Komponente napovedi prodaje	34
Slika 21: Št. prebivalcev, BDP na prebivalca in stopnja motoriziranosti v letu 2011	36
Slika 22: Uporabljen proces v RapidMiner-ju za modeliranja regresijskega modela	43
Slika 23: Shematska ponazoritev operatorja »Backward Elimination« in njegovih podprocesov	43
Slika 24: Podproces »Training & Testing« izračun in preverjanje regresijskega modela	44
Slika 25: Izpis regresijskega modela iz programa RapidMiner	45

Slika 26: Regresijski model lahkih dostavnih vozil z uporabo letnega nabora podatkov.....	48
Slika 27: Grafični prikaz rezultatov regresijskega modela lahkih dostavnih vozil z uporabo letnega nabora podatkov	48
Slika 28: Grafični prikaz korelacij med vhodnimi in izhodnimi spremenljivkami letnega modela segmenta lahkih dostavnih vozil	49
Slika 29: Regresijski model segmenta lahkih dostavnih vozil z uporabo četrtnetnega nabora podatkov.....	50
Slika 30: Regresijski model segmenta lahkih dostavnih vozil z uporabo četrtnetnega nabora podatkov in slamnatimi spremenljivkami	51
Slika 31: Kazalniki regresijskega modela segmenta lahkih dostavnih vozil z uporabo četrtnetnega nabora podatkov in slamnatimi spremenljivkami	51
Slika 32: Grafični prikaz rezultatov regresijskega modela segmenta lahkih dostavnih vozil z uporabo četrtnetnega nabora podatkov in slamnatih spremenljivk.....	52
Slika 33: Grafični prikaz korelacij med vhodnimi spremenljivkami četrtnetnega modela segmenta lahkih dostavnih vozil	53
Slika 34: Model segmenta lahkih dostavnih vozil z uporabo četrtnetnih napovedi na letni ravni	54
Slika 35: Grafični prikaz rezultatov modela segmenta lahkih dostavnih vozil z uporabo četrtnetnega nabora na letni ravni v primerjavi z letnim naborom podatkov.....	54
Slika 36: Regresijski model segmenta težkih tovornih vozil z uporabo	56
Slika 37: Grafični prikaz rezultatov regresijskega modela težkih tovornih vozil z uporabo letnega nabora podatkov	57
Slika 38: Grafični prikaz korelacij med vhodnimi spremenljivkami letnega modela	57
Slika 39: Grafični prikaz korelacij med vhodnimi spremenljivkami letnega modela	58
Slika 40: Regresijski model segmenta težkih tovornih vozil z uporabo četrtnetnega nabora podatkov.....	58
Slika 41: Analiza napake regresijskega modela segmenta težkih tovornih vozil z uporabo četrtnetnega nabora podatkov	59
Slika 42: Grafični prikaz korelacij med vhodnimi in izhodnimi spremenljivkami četrtnetnega modela segmenta težkih tovornih vozil.....	59
Slika 43: Regresijski model segmenta težkih tovornih vozil z uporabo četrtnetnega nabora podatkov in slamnatimi spremenljivkami	60
Slika 44: Analiza napake regresijskega modela segmenta težkih tovornih vozil z uporabo četrtnetnega nabora podatkov	61
Slika 45: Grafični prikaz rezultatov regresijskega modela segmenta težkih tovornih vozil z uporabo četrtnetnega nabora podatkov in slamnatih spremenljivk.....	61
Slika 46: Model segmenta težkih tovornih vozil z uporabo četrtnetnih napovedi na letni ravni	62

Slika 47: Grafični prikaz regresijskega modela segmenta težkih tovornih vozil z uporabo četrtnega nabora na letni ravni v primerjavi z letnim naborom podatkov.....	63
Slika 48: Regresijski model segmenta avtobusov z uporabo letnega nabora podatkov	65
Slika 49: Grafični prikaz rezultatov regresijskega modela segmenta avtobusov z uporabo letnega nabora podatkov	66
Slika 50: Prikaz korelacij med vhodnimi spremenljivkami letnega modela segmenta avtobusov	66
Slika 51: Regresijski model segmenta avtobusov z uporabo četrtnega nabora podatkov.....	67
Slika 52: Grafični prikaz korelacij med vhodnimi in izhodnimi spremenljivkami četrtnega modela segmenta avtobusov	68
Slika 53: Regresijski model segmenta avtobusov z uporabo četrtnega nabora podatkov in slamnatimi spremenljivkami.....	68
Slika 54: Analiza napake regresijskega modela segmenta težkih tovornih vozil z uporabo četrtnega nabora podatkov	69
Slika 55: Grafični prikaz rezultatov regresijskega modela segmenta avtobusov z uporabo četrtnega nabora podatkov in slamnatih spremenljivk	69
Slika 56: Model segmenta avtobusov z uporabo četrtnega nabora podatkov na letnem nivoju v primerjavi z letnim naborom podatkov.....	70
Slika 57: Grafični prikaz rezultatov modela avtobusov z uporabo četrtnega nabora na letni ravni v primerjavi z letnim naborom podatkov.....	71

UVOD

Problematika in namen magistrskega dela. Eden od ciljev vsakega uspešnega podjetja je čim bolj predvidljivo poslovanje v prihodnosti. S tem namenom večina podjetji v procesu načrtovanja poslovanja pripravlja operativne, taktične in strateške načrte, s katerimi poskušajo opredeliti svoje poslovanje v prihodnosti. Osnova za pripravo dobrega načrta za poslovanje v prihodnosti so napovedi gibanja ključnih dejavnikov poslovanja podjetja, med katere spada tudi načrt prodaje.

V avtomobilski industriji je zaradi hude konkurence in dolgih razvojnih ciklov veliko pozornost posvečeno pravilnemu načrtovanju proizvodne palete in ustreznih proizvodnih zmogljivosti. Učinkovito načrtovanje proizvodne funkcije zahteva natančne 6–24 mesečne napovedi prodaje in povpraševanja. Te napovedi so vhodni podatek za več ključnih poslovnih in proizvodnih odločitev, ki seveda bistveno vplivajo na dobičkonosnost poslovanja proizvajalcev vozil. V procesu načrtovanja proizvodnje se te napovedi uporabljajo za določitev potrebnega obsega delovne sile in prilagoditev obsega proizvodnje na obstoječih proizvodnih linijah. Napake pri napovedi povpraševanja pogosto vodijo k velikim dodatnim stroškom ali pa izpadu potencialnih prihodkov (Sa-ngasoongsong, Bukkapatnam, Kim, Iyer & Suresh, 2012, str. 875).

V delu avtomobilske industrije, ki se ukvarja s proizvodnjo in prodajo gospodarskih vozil, je ta proces še toliko pomembnejši, saj iz lastnih izkušenj vem, da so proizvodni cikli v tem segmentu daljši (10–15 let v primerjavi s 7. leti pri osebnih vozilih), dobičkonosnost prihodkov v povprečju nižja kot pri osebnih avtomobilih, nihanja in občutljivost trgov na spremembe v ekonomskem okolju pa bistveno višja.

Obseg trga v avtomobilski industriji merimo na osnovi podatkov o registraciji vozil na določenem geografsko-administrativnem območju (občina, država, skupina držav, gospodarski prostor). Z napovedovanjem obsega trga se je v zadnjih 15. letih ukvarjalo kar nekaj avtorjev, ki so se reševanja tega problema lotili z različnimi metodami. Od napovedovanja obsega trga na osnovi nakupnih namer (Armstrong, Morwitz & Kumar, 2000) do napovedovanja s pomočjo algoritmov rudarjenja s podatki (Hülsmann, Borscheid, Freidrich & Reith, 2012). Strnjen pregled prispevkov in avtorjev so pripravili Sa-ngasoongsong et al. (2012) ter Fantanzzini in Toktamysova (2015). Zadnja sta spremenljivke, ki vplivajo na prodajo vozil, razdelila v tri skupine:

- vpliv tehnoloških karakteristik vozil, kot so: kakovost, inovacije in tehnologija, zmogljivosti in ekonomičnost motorjev, funkcionalnost, varnost, prostornost, oblikovanje in estetika,
- vpliv trženjskega spleta, kot so na primer veleprodajne in maloprodajne cene, prodajne in poprodajne storitve, trženjske aktivnosti, ugled blagovne znamke,

- vpliv okolja (političnega, ekonomskega, socialnega), kot so organizacijske in politične teme, globalna ekonomska rast, ekološke in naravne sile, socialno-kulturni vplivi in obnašanje potrošnikov.

Pri tem ugotavljata, da sta prvi dve skupini spremenljivk pomembni, vendar nimata dolgoročnega vpliva na prodajo in jih avtomobilski proizvajalci v večini primerov obvladujejo. Zadnjo skupino predstavljajo vplivi, na katere avtomobilski trgi na splošno nimajo vpliva. Uporaba teh spremenljivk za napovedovanje obsega trga je bila do sedaj precej omejena.

Do podobnih ugotovitev so prišli tudi Sa-ngasoongsong et al. (2012, str. 875–876), ki prav tako ugotavljajo, da je bilo v tem času zelo malo prispevkov narejenih na temo napovedovanja prodaje avtomobilov z uporabo ekonomskih kazalcev in še ti se nanašajo pretežno na napovedovanje obsega trga osebnih vozil. Shahabudin (2009) je za napovedovanje obsega trga uporabil regresijsko analizo z 12. makroekonomskimi spremenljivkami in ugotovil, da je za trg v ZDA z njimi mogoče napovedati prodajo uvoženih osebnih avtomobilov in lahkih gospodarskih vozil (angl. *trucks*). Brühl, Hülsmann, Borscheid, Freidrich in Reith (2009) so s pomočjo različnih ekonomskih kazalcev, kot so bruto domači proizvod (v nadaljevanju BDP), inflacija, obrestna mera, nezaposlenost in cena goriva, testirali uporabnost le-teh za napovedovanje obsega trga vozil v Nemčiji. Ugotovitve omenjene raziskave so nadgradili še z uporabo metod podatkovnega rudarjenja in razširitvijo uporabe na ameriški avtomobilski trg (Hülsmann, Borscheid, Freidrich & Reith, 2012). Wang, Chang in Tzeng (2011) so nekaj podobnega pripravili na Tajvanu, kjer prav tako ugotavljajo, da so določeni ekonomski kazalci dobri napovedovalci gibanja obsega trga in da dajejo ti podatki v povezavi z ustreznimi modeli boljše rezultate kot metode analize časovnih vrst (npr. ARIMA). Zanimiv koncept sta predstavila Fantazzini in Toktamysova (2015), ki predlagata uporabo multivariantnega modela z uporabo ekonomskih spremenljivk in podatkov o iskanjih na Googlovem spletnem iskalniku.

Žal je navedenim prispevkom skupno tudi to, da posebej ne izpostavljajo trga gospodarskih vozil. To je do določene mere razumljivo glede na dejstvo, da so po podatkih združenja evropskih avtomobilskih proizvajalcev ACEA (v nadaljevanju ACEA) v letu 2014 registracije novih osebnih vozil predstavljale nekaj manj kot 80 % vseh na novo registriranih vozil na svetu (ACEA, 2015, str. 10). Po drugi strani pa lahko že z enostavno primerjavo podatkov o prvih registracijah osebnih in gospodarskih vozil z gibanjem bruto domačega proizvoda v EU ugotovimo, da gibanje prvih registracij gospodarskih vozil bistveno bolj odseva nihanje te makroekonomske spremenljivke, kot pa je to pri osebnih vozilih ACEA (2015, str. 33). Iz izkušenj vemo, da je trg gospodarskih vozil zelo segmentiran in podvržen cikličnim in drugim nihanjem, zato je načrtovanje prodaje temu primerno zahtevno.

V magistrskem delu bom s pomočjo izbranih metod analize podatkov, razpoložljivih napovedi gibanj ekonomskih spremenljivk in analize časovnih vrst določenih ekonomskih spremenljivk, z uporabo programskega paketa RapidMiner, poskušal razviti model za napovedovanje gibanja obsega slovenskega trga gospodarskih vozil, ki bo predstavljal osnovo za načrtovanje prodaje.

Cilj magistrskega dela je:

- predstaviti segmente in značilnosti posameznih segmentov trga gospodarskih vozil v Sloveniji,
- pripraviti kratek pregled metod napovedovalne analitike, ki so jih različni avtorji uporabili pri napovedovanju prodaje v avtomobilski industriji, in ugotoviti njihovo uporabnost za napovedovanje obsega trga gospodarskih vozil v Sloveniji,
- s pomočjo regresijske metode razviti in ovrednotiti model za napovedovanje posameznih segmentov trga gospodarskih vozil v Sloveniji.

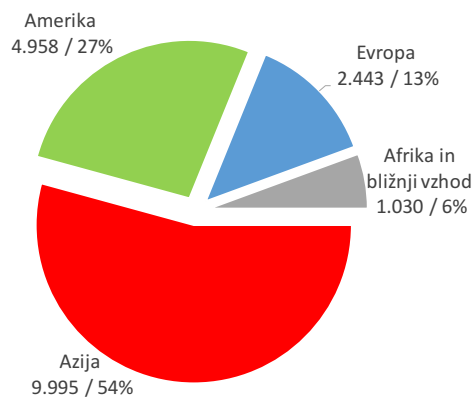
Metode dela. V prvem teoretičnem delu bom najprej z metodo analize segmentiral trg gospodarskih vozil in predstavil značilnosti njihovih segmentov. V drugem delu teoretičnega dela bom z uporabo metod zbiranja, analize razpoložljivega gradiva naredil kratek pregled metod podatkovnega rudarjenja in preveril možnost njihove uporabe na področju napovedovanja obsega trga v avtomobilski industriji.

V drugem praktičnem delu bom s pomočjo programskega paketa RapidMiner in metod podatkovnega rudarjenja razvil model za ocenjevanje obsega za vsak segment trga gospodarskih vozi na osnovi izbranih ekonomskih spremenljivk. Omejil se bom na tiste ekonomske spremenljivke, za katere so na voljo javno dostopne napovedi za zahtevan časovni horizont (tekoče leto + 2 leti), ki jih pripravljajo priznane inštitucije.

1 ZNAČILNOSTI TRGA GOSPODARSKIH VOZIL V SLOVENIJI

Gospodarsko vozilo (angl. *commercial vehicle*, nem. *Nutzfahrzeug*) je kot že ime pove namenjeno v prvi vrsti opravljanju gospodarske dejavnosti. Evropska unija definira gospodarsko motorno vozilo (angl. *commercial motor vehicle*) kot motorizirano cestno vozilo, ki je po svoji konstrukciji in opremljenosti oblikovano in ustrezno za plačani ali brezplačni prevoz več kot devet oseb, vključno z voznikom, ali blaga (Direktiva Sveta 85/347/EGS o spremembi Direktive 68/297/EGS o standardizaciji predpisov o brezcarinskem vnosu goriva, vsebovanega v rezervoarjih za gorivo gospodarskih motornih vozil, Uradni list L 183). ACEA med gospodarska vozila šteje avtobuse, tovornjake in dostavna vozila (ACEA, 2015). Prve registracije gospodarskih vozil po tej klasifikaciji razdeljene po svetovnih regijah prikazuje Slika 1. Kot je razvidno več kot polovico trga gospodarskih vozil predstavlja azijski trg.

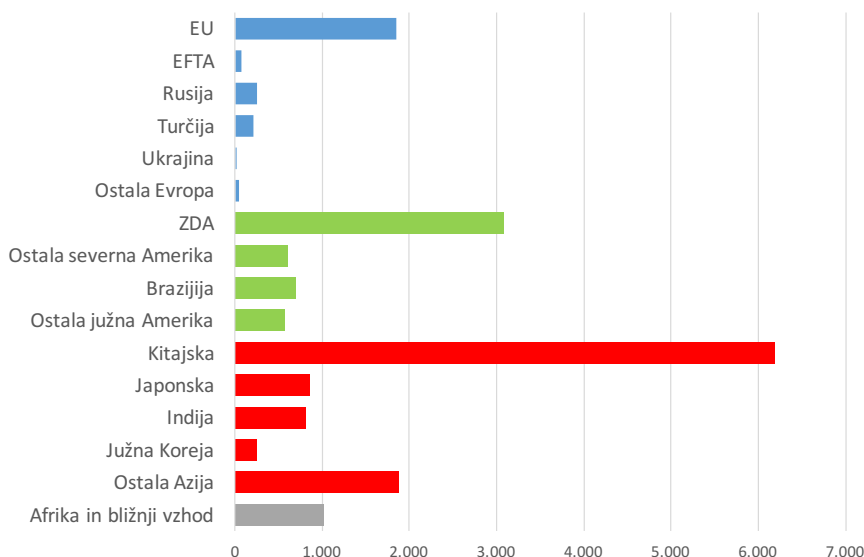
Slika 1: Deleži registracije gospodarskih vozil v letu 2014 po regijah (v tisočih)



Vir: Združenje evropskih avtomobilskih proizvajalcev ACEA, ACEA Pocket Guide 2015–2016, 2015, str. 31.

Podrobnejše podatke o velikosti trgov gospodarskih vozil znotraj posamičnih svetovnih regij prikazuje Slika 2 iz katere je razvidno, da Kitajska predstavlja daleč največje svetovno tržišče.

Slika 2: Prve registracije gospodarskih vozil znotraj regij (v tisočih)



Vir: Združenje evropskih avtomobilskih proizvajalcev ACEA, ACEA Pocket Guide 2015–2016, 2015, str. 31.

V Sloveniji Zakon o motornih vozilih (Ur.l. RS, št. 106/2010, 23/2015, v nadaljevanju ZMV) ne določa definicije gospodarskih vozil, pač pa so v zakonu navedene definicije vozil, ki jih prištevamo k gospodarskim vozilom (npr. avtobus, cestni vlačilec, delovno vozilo, tovorno vozilo..).

1.1 Predstavitev segmentacije trga gospodarskih vozil

Že iz uvodnega dela tega poglavja lahko razberemo, da je pojem gospodarsko vozilo precej širok, zato je temu primerno veliko tudi različnih segmentacij trga in klasifikacij vozil. V uradnih dokumentih in podatkovnih zbirkah se pogosto pojavlja klasifikacija organizacije Združenih narodov oziroma njene komisije United Nations Economic Commission for Europe (v nadaljevanju UNECE), ki v svojem dokumentu UNECE (2016, str. 6–13) klasificira vse vrste vozil. To klasifikacijo lahko najdemo tako v slovenski zbirki podatkov o registriranih motornih vozilih, objavljeni na spletnem portalu nio.gov.si, kakor tudi v uradni statistiki Statističnega urada Republike Slovenije (v nadaljevanju SURS). Objave SURS ravno tako temeljijo na tej klasifikaciji, kot je razvidno iz SURS (2016, str. 3). Ta klasifikacija v grobem deli motorna vozila na naslednje kategorije:

- Kategorija M – vozila s pogonom, z najmanj štirimi kolesi za prevoz oseb, ki se deli še na naslednje kategorije:
 - Kategorija M1 – vozila za prevoz potnikov, ki poleg voznikovega sedeža nimajo več kot 8 sedežev.
 - Kategorija M2 – vozila za prevoz potnikov, ki imajo poleg voznikovega sedeža več kot 8 sedežev in njihova skupna dovoljena masa ne presega 5 t.
 - Kategorija M3 – vozila za prevoz potnikov, ki imajo poleg voznikovega sedeža več kot 8 sedežev in njihova skupna dovoljena masa presega 5 t.
- Kategorija N – vozila s pogonom, z najmanj štirimi kolesi, za prevoz tovora, ki se deli še na naslednje kategorije:
 - Kategorija N1 – vozila za prevoz tovora, katerih skupna dovoljena masa ne presega 3,5 t.
 - Kategorija N2 – vozila za prevoz tovora, katerih skupna dovoljena masa presega 3,5 t vendar ne presega 12 t.
 - Kategorija N3 – vozila namenjena za prevoz tovora, katerih skupna dovoljena masa presega 12 t.

Nekoliko bolj enostavno segmentacijo uporablja združenje evropskih avtomobilskih proizvajalcev ACEA (2016, str. 2), ki v svojih objavah deli gospodarska vozila na naslednje segmente:

- lahka gospodarska vozila (angl. *light commercial vehicles*) do 3,5 t skupne mase, ki vključuje tudi lahke avtobuse do 3,5 t,
- srednja gospodarska vozila (angl. *medium commercial vehicles*) med 3,5 t in 16 t skupne mase, brez avtobusov,
- težka gospodarska vozila (angl. *heavy commercial vehicles*) nad 16 t skupne mase, brez avtobusov,

- srednje in težke avtobuse (angl. *medium and heavy buses & coaches*) nad 3,5 t skupne mase.

Glede na razpoložljive podatke o registriranih vozilih v Sloveniji sem v sklopu tega magistrskega dela uporabil 3 glavne segmente gospodarskih vozil:

- segment lahkih dostavnih vozil,
- segment težkih tovornih vozil,
- segment avtobusov.

Te segmente je mogoče dovolj natančno identificirati na podlagi razpoložljivih podatkov o registriranih vozilih in da imajo dovolj jasno profilirane kupce. Zaradi celovite obravnave trga gospodarskih vozil navajam še segmente gospodarskih vozil, ki niso predmet obravnave v tem magistrskem delu :

- segment srednjih tovornih vozil – srednja gospodarska vozila namenjena za prevoz tovora med 3,5 t in 16 t, brez avtobusov,
- segment lahkih avtobusov – vozila do 3,5 t skupne mase namenjena za prevoz oseb.

Teh dveh segmentov zaradi zasnove baze podatkov o presekih stanja registriranih vozil za obdobje 1995–2005 (Ministrstvo za infrastrukturo, 2015) žal ni mogoče nedvoumno identificirati, poleg tega pa predstavljajo relativno majhen podsegment gospodarskih vozil. Po podatkih Ministrstva za infrastrukturo (Ministrstvo za infrastrukturo, 2016a) je bilo od vseh gospodarskih vozil, namenjenih za prevoz tovora (kategorije N), na dan 31. 12. 2015 le 8,4 % vozil, ki spadajo v segment srednjih tovornih vozil. Na drugi strani pa segment lahkih avtobusov po namenu uporabe sovпада s segmentom osebnih vozil, zato v tem segmentu ne moremo govoriti o povsem gospodarskih vozilih, saj se uporabljajo tako za osebni prevoz kot za gospodarsko dejavnost.

1.2 Predstavitev segmenta lahkih dostavnih vozil

Segment trga lahkih dostavnih vozil je identičen segmentu lahkih gospodarskih vozil, kot ga navaja ACEA (2016, str. 2), vendar brez lahkih avtobusov do 3,5 t. V ta segment torej spadajo vsa vozila, namenjena prevozu tovora do skupne mase 3,5 t, kar ustreza vozilom kategorije N1 po klasifikaciji UNECE (2016, str. 6–13).

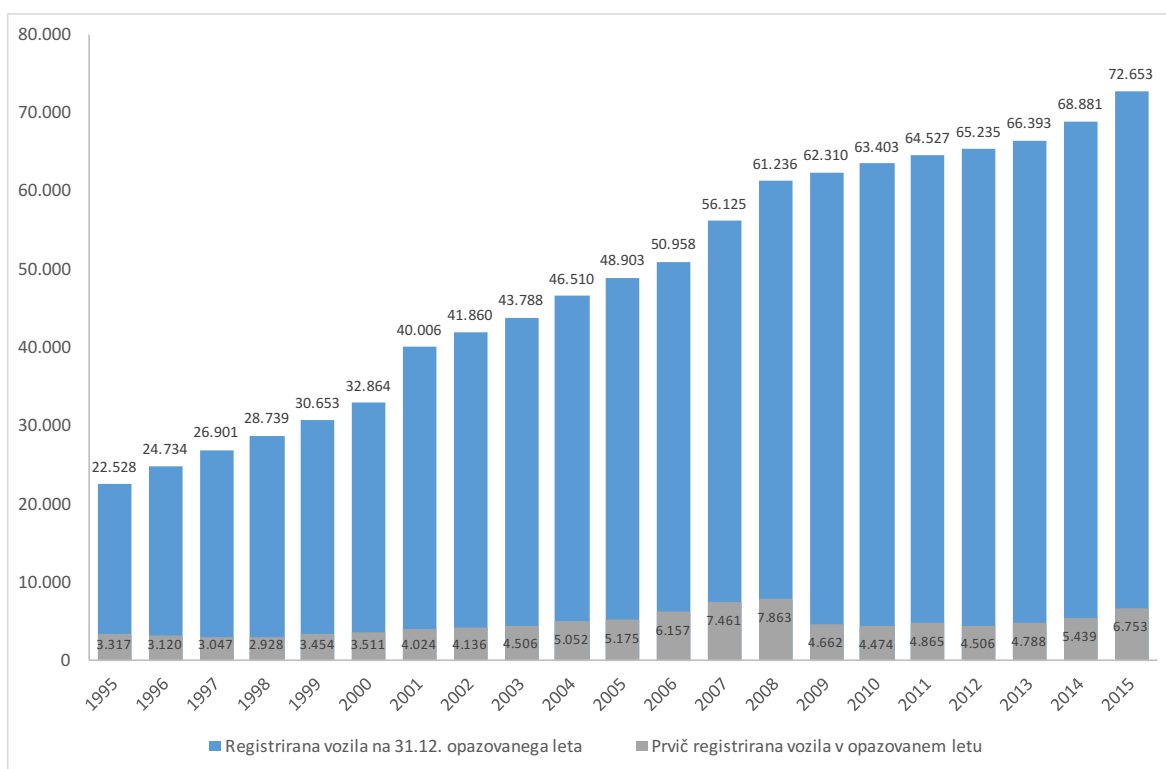
Vozila v tem segmentu se uporabljajo pretežno za prevoz tovora praviloma na krajših relacijah. Značilnost tega segmenta je, da osnovna dejavnost kupcev v večini primerov ni neposredno povezana s prevozom blaga, temveč prevoz predstavlja dopolnilno storitev, ki jo ponujajo na trgu skupaj s svojim osnovnim izdelkom ali storitvijo. Kupci teh vozil so večinoma manjši podjetniki in obrtniki iz proizvodne, trgovinske in storitvene dejavnosti.

Kupci lahkih gospodarskih vozil prihajajo pretežno iz naslednjih dejavnosti:

- trgovine,
- gradbeništva,
- storitvenih dejavnosti,
- pošte in paketne dostave,
- javnih služb (policije, vojske, gasilcev, reševalcev, zaščite in reševanja),
- mednarodnega transporta,
- prevoza oseb znotraj države,
- kmetijstva,
- industrije.

Slika 3 nam kaže, da se je število registriranih lahkih dostavnih vozil v zadnjih 20 letih več kot potrojilo, pri čemer so k rasti prispevale predvsem registracije novih vozil, saj le-te predstavljajo v povprečju kar 91 % prvih registracij, preostalih 8 % pa predstavljajo rabljena vozila iz uvoza, prvič registrirana v Republiki Sloveniji.

Slika 3: Stanje registriranih in prvič registriranih lahkih dostavnih vozil v obdobju 1995–2015 v Sloveniji

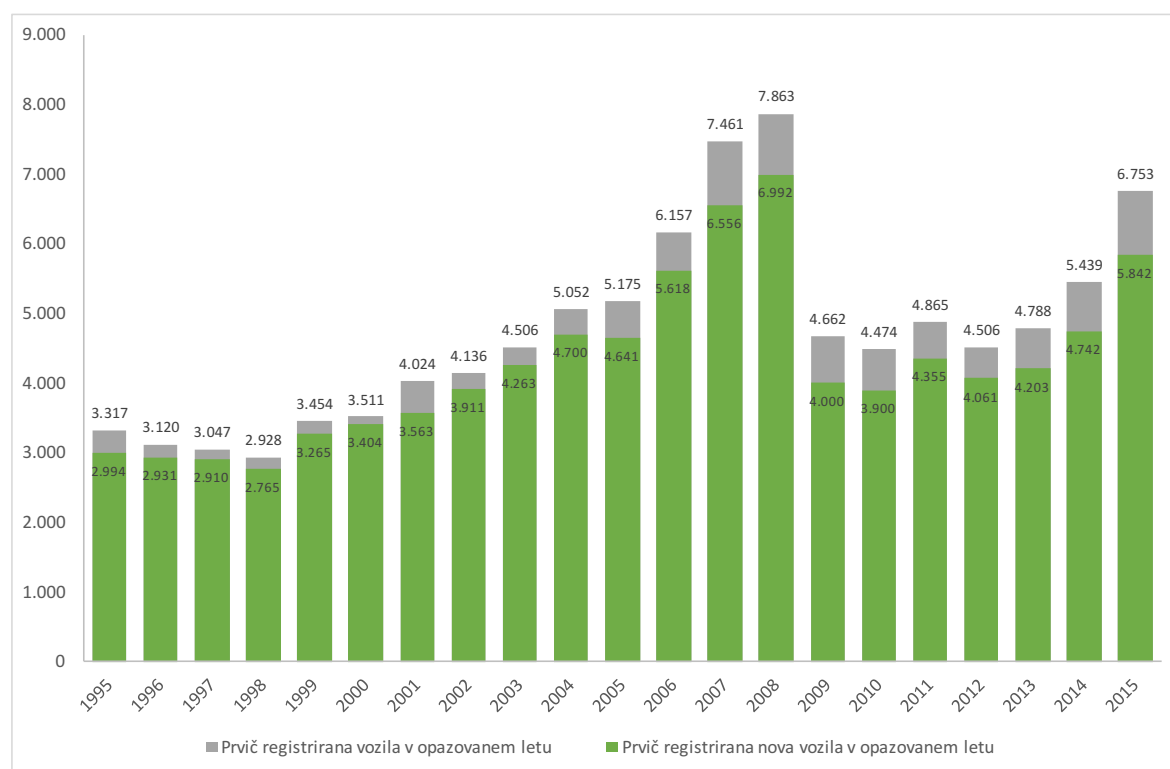


*Vir: Ministrstvo za infrastrukturo, Evidenca registriranih vozil – preseki stanja 1992–2005, 2015;
Ministrstvo za infrastrukturo, Evidenca registriranih vozil – preseki stanja 2006–2015, 2016a;
lastni izračuni.*

Pričakovano je segment v letih 2005–2008 močno rasel, nato pa v letu 2009 zaradi gospodarske krize padel za 41 %. Zanimivo je, da se je rast v segment vrnila že v letu 2011, nato pa je v letu 2012, zaradi uvedbe varčevalnih ukrepov, ponovno začela padati. Glede na navedeno in na značilnosti kupcev lahko predvidevamo, da je ta segment trga odvisen predvsem od domačega povpraševanja; pretežen del kupcev tega segmenta gospodarskih vozil je namreč odvisen od razmer na domačem trgu.

Dinamiko in delitev prvih registracij lahkih dostavnih na nova in rabljena vozila prikazuje Slika 4.

*Slika 4: Pregled prvič registriranih lahkih dostavnih vozil
v obdobju 1995–2015 v Sloveniji*



*Vir: Ministrstvo za infrastrukturo, Evidenca registriranih vozil – preseki stanja 1992–2005, 2015;
Ministrstvo za infrastrukturo, Evidenca registriranih vozil – preseki stanja 2006–2015, 2016a;
lastni izračuni.*

Tudi znotraj segmenta lahkih dostavnih vozil različni proizvajalci delijo segment še na podsegmente, glede na nosilnost in dimenzije posameznih modelov (npr. mikro, mini, midi in velika lahka dostavna vozila). Vendar pa ta delitev ni standardizirana, saj vsi proizvajalci niso zastopani v vseh podsegmentih.

1.3 Predstavitev segmenta težkih tovornih vozil

Segment težkih tovornih vozila je identičen segmentu težkih gospodarskih vozil, kot ga definira ACEA (ACEA, 2016, str. 2). V ta segment vozil uvrščamo vozila skupne

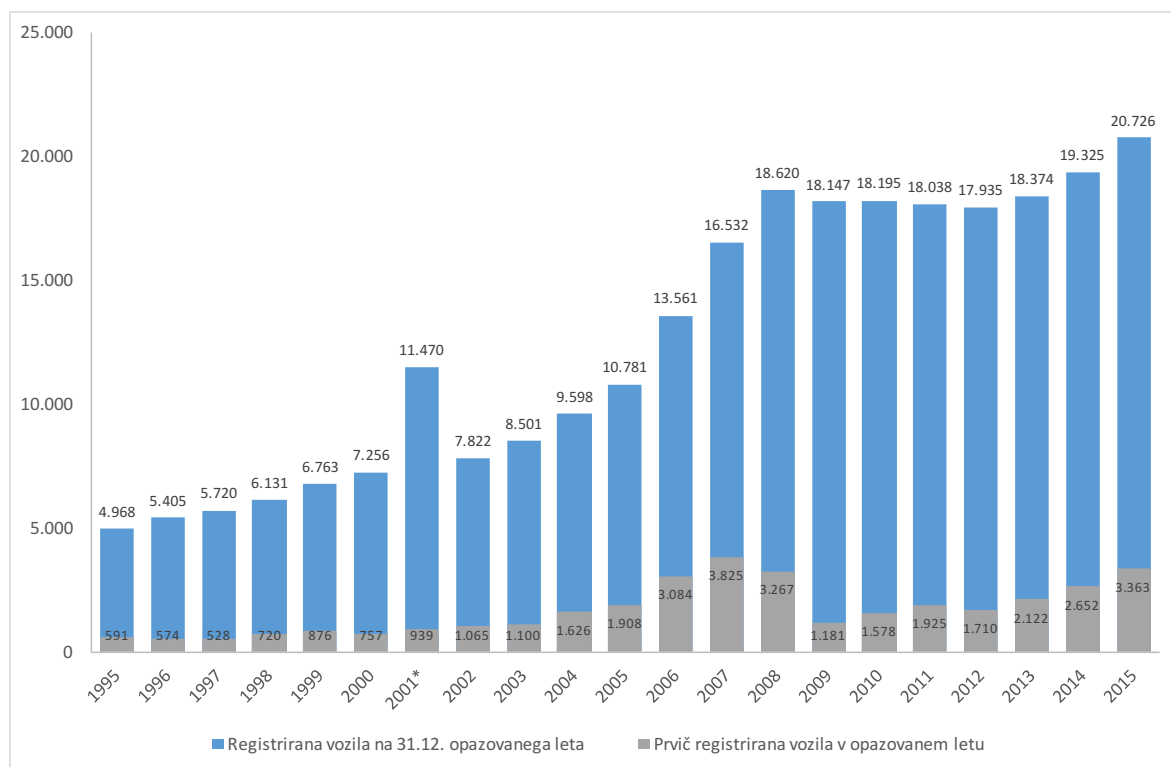
dovoljene mase nad 16 t in so namenjena prevozu tovora (brez avtobusov). Ta segment ustreza vozilom kategorije N3 po UNECE klasifikaciji (UNECE, 2016, str. 6–13). Večini kupcev vozil v tem segmentu transport in logistika predstavlja osnovno dejavnost. Poleg te so kupci teh vozil prihajajo še iz gradbeništva, komunale, gozdarstva, storitvenih dejavnosti ter zaščite in reševanja (gasilci, vojska, civilna zaščita). Segment težkih tovornih vozil lahko razdelimo v dva večja podsegmenta, in sicer podsegment vlačilcev in podsegment šasij.

Segment vlačilcev predstavlja standardno gospodarsko vozilo za prevoz blaga na daljše razdalje. Lastne analize podatkov o prvih registracijah težkih tovornih vozil v letu 2015 (Ministrstvo za infrastrukturo, 2016b), da ta podsegment trga obsega kar 75 % segmenta težkih tovornih vozil. To je mogoče razložiti z dejstvom, da večina kupcev ta vozila uporablja v mednarodnem transportu, kjer so, zaradi svoje univerzalnosti in enostavnosti uporabe vlačilci s polprikloniki, najbolj pogosto vozilo.

Preostalih 25 % segmenta predstavljajo šasije, ki so nadgrajene z različnimi vrstami nadgradenj, od kesonov s ceradami in prikolic za prevoz volumenskih tovorov, nadgradenj za prevoz avtomobilov do komunalnih in gasilskih nadgradenj ter nadgradenj za potrebe gradbeništva (prekucniki, mešalci betona, črpalke za beton ipd.). Slika 5 prikazuje razvoj števila registriranih vozil v Republiki Sloveniji po letih. Iz prikaza lahko ugotovimo strm porast števila registriranih vozil od leta 2003 do leta 2008, ki se nato ponovi v letih 2013 do 2015. Na to vplivajo naslednji dejavniki:

- vstop Slovenije v EU – s čimer se je slovenskim prevoznikom odprla možnost, da brez do tedaj veljavnih omejitev (dovolilnic) opravljajo prevoze znotraj EU. Slovenija je bila poleg Malte edina od novih članic, ki se je v pristopnih pogajanjih uspela izogniti prehodnemu obdobju, v katerem prevozniška podjetja niso smela opravljati kabotaže v drugih državah članicah EU,
- zakonska regulativa: uvedba emisijskih standardov EURO IV v letu 2005, uvedba obveznih digitalnih tahografov v letu 2006, uvedba emisijskih standardov EURO V v letu 2008, uvedba emisijskih standardov EURO VI v letu 2014,
- davčna regulativa: davčne olajšave za nakup ekoloških vozil do leta 2004 za vozila z EURO IV motorji, do leta 2012 za vozila z Euro V in EURO V EEV motorji, po letu 2012 za vozila z EURO VI motorji, subvencije za vozila z EURO V EEV in EURO VI motorji v letu 2013,
- rast izvoznega dela slovenskega gospodarstva, ki se je tudi po krizi leta 2009 hitro opomoglo in katero predstavlja pomemben trg za slovenska prevozniška podjetja.

*Slika 5: Stanje registriranih in prvič registriranih težkih tovornih vozil
v obdobju 1995–2015 v Sloveniji*



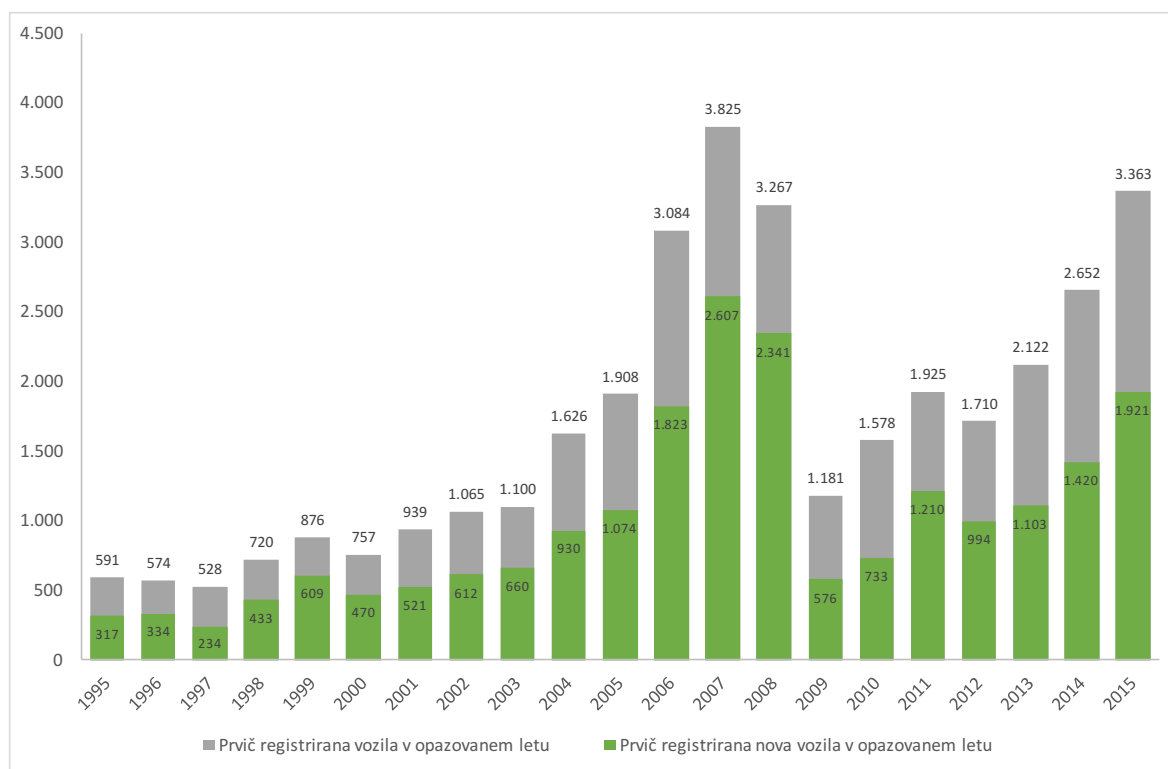
*podatek o številu vseh registriranih težkih tovornih vozil za leto 2001 odstopa zaradi nekonistentne spremembe baze podatkov

*Vir: Ministrstvo za infrastrukturo, Evidenca registriranih vozil – preseki stanja 1992–2005, 2015;
Ministrstvo za infrastrukturo, Evidenca registriranih vozil – preseki stanja 2006–2015, 2016a;
lastni izračuni.*

Časovni razvoj prvič registriranih vozil, ki ga prikazuje Slika 6 kaže, da je v nasprotju z razmerami v Sloveniji prišlo do gospodarske krize v tej panogi dejansko že v letu 2008, saj tu že beležimo padec prvih registracij v tem segmentu za 15 % v primerjavi z letom 2007, ki mu je sledil še dodaten padec v letu 2009 za dodatnih 64 % (glede na 2008) oziroma kumulativno v letih 2008 in 2009 za 70 %. Še večji upad zaznamo pri prvih registracijah novih težkih tovornih vozil, kjer je skupni padec trga v letih 2008 in 2009 znašal kar 78 %.

Delež prvih registracij novih vozil v obsegu vseh prvih registracij skozi opazovano obdobje je z izjemo v letih 1999 in 2008 stabilen in se giblje okoli 60 %, kar pomeni, da je delež rabljenih vozil med prvimi registracijami skoraj 5-krat večji kot v segmentu lahkih dostavnih vozil, kjer je v povprečju okoli 8 %. Razlog za to je v bistveno večji orientiranosti v mednarodno poslovanje in mobilnosti kupcev v tem segmentu gospodarskih vozil, ki rabljena vozila kupujejo pretežno v zahodni Evropi (Nemčija, Beneluks).

Slika 6: Pregled prvič registriranih težkih tovornih vozil v obdobju 1995–2015



*Vir: Ministrstvo za infrastrukturo, Evidenca registriranih vozil – preseki stanja 1992–2005, 2015;
Ministrstvo za infrastrukturo, Evidenca registriranih vozil – preseki stanja 2006–2015, 2016a;
lastni izračuni.*

1.4 Predstavitev segmenta avtobusov

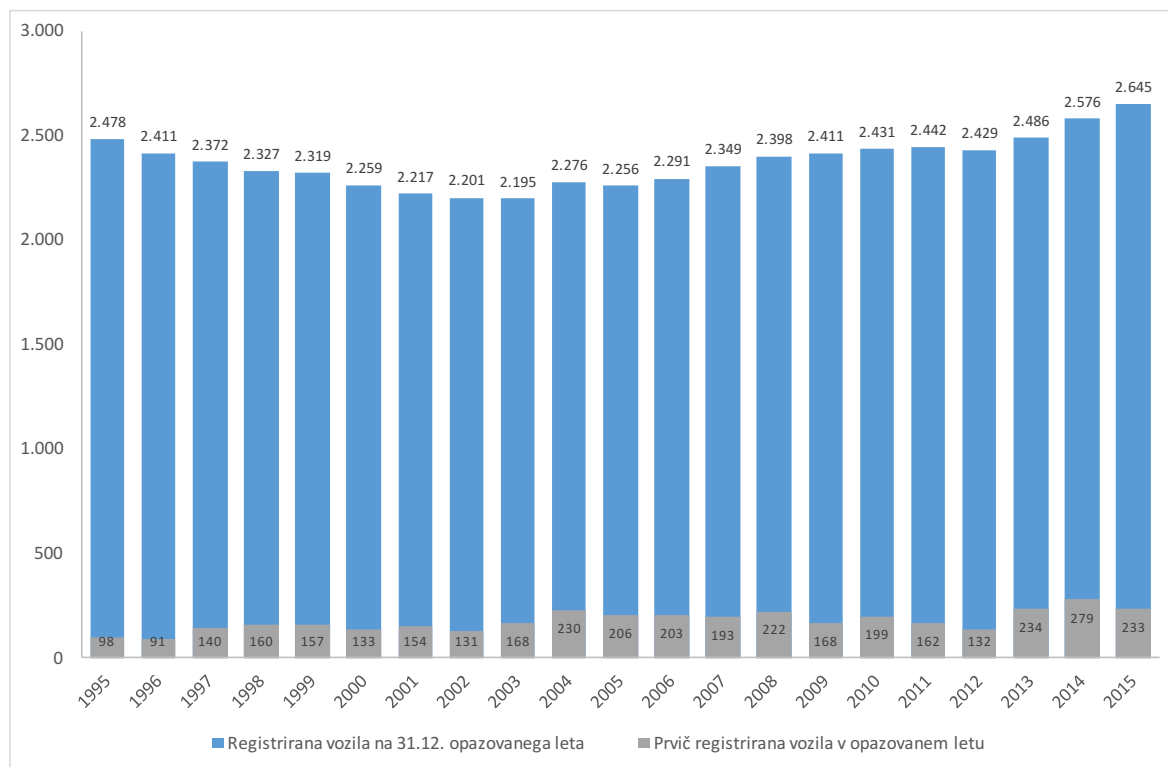
Ta segment trga ustreza segmentu srednjih in težkih avtobusov po ACEA segmentaciji (ACEA, 2016, str. 2). V ta segment se uvrščajo gospodarska vozila skupne dovoljene mase, večje od 3,5 t, in so namenjena prevozu oseb. Ta segment ustreza vozilom kategorije M2 in M3 po UNECE klasifikaciji (UNECE, 2016, str. 6–13).

Osnovna dejavnost kupcev tega segmenta gospodarskih vozil je prevoz osebi in to storitev prodajajo na trgu. Ta segment trga lahko glede na namen delimo na naslednje glavne podsegmente:

- mestni avtobusi – so namenjeni prevozu potnikov v mestnem prometu, kjer potniki med vožnjo lahko tudi stojijo,
- primestni avtobusi – so namenjeni prevozu potnikov v medkrajevem prometu, kjer potniki sedijo na sedežih, stojišča pa so omejena na prostor ob vstopu in izstopu,
- turistični avtobusi – so namenjeni prevozu potnikov na daljše razdalje, pri čemer potniki med vožnjo sedijo.

Kot kaže Slika 7 se je število registriranih avtobusov v Sloveniji do leta 2003 rahlo zniževalo, nato pa je z izjemo leta 2012 počasi raslo, vendar pa je dinamika v primerjavi z ostalima preučevanima segmentoma tako majhna, da bi lahko ta segment označili za relativno stabilen.

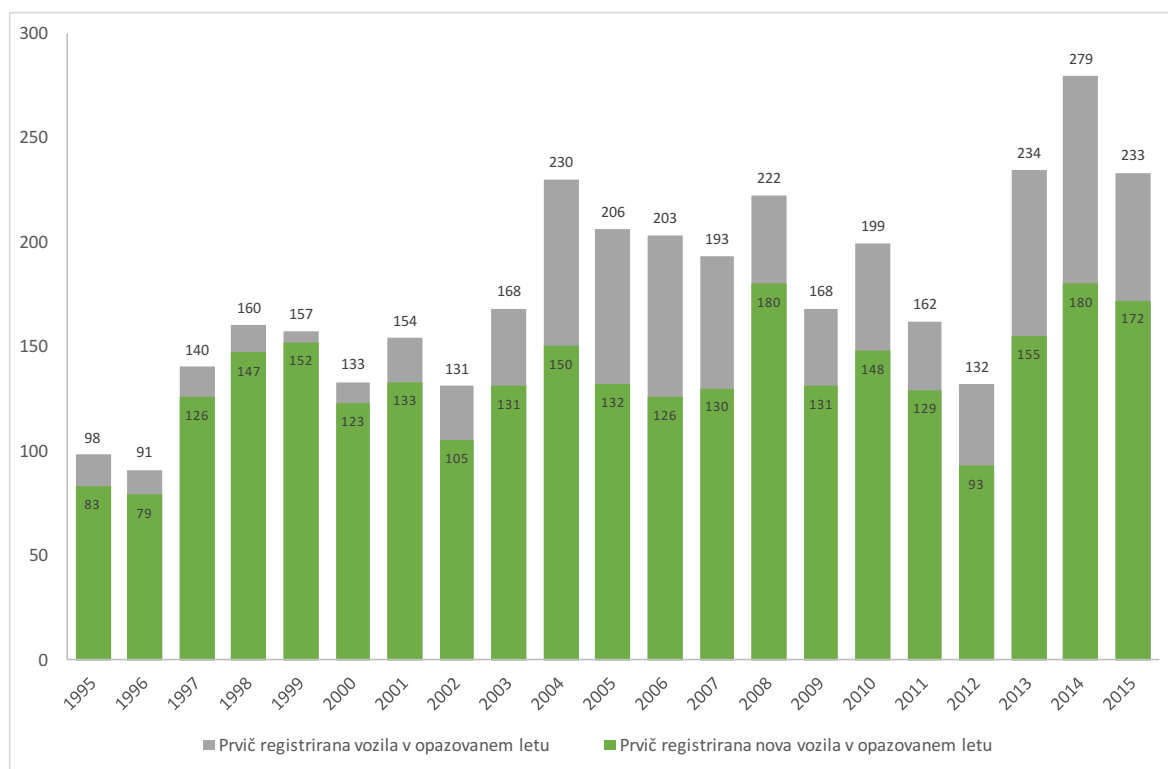
Slika 7: Stanje registriranih in prvič registriranih avtobusov v obdobju 1995–2015



*Vir: Ministrstvo za infrastrukturo, Evidenca registriranih vozil – preseki stanja 1992–2005, 2015;
Ministrstvo za infrastrukturo, Evidenca registriranih vozil – preseki stanja 2006–2015, 2016a;
lastni izračuni.*

Tudi število prvih registracij v obdobju 1995 do 2015 ne kaže nekega izrazitega trenda. Slika 8 kaže, da je tudi tu delež novih vozil v prvih registracijah zelo visok (v povprečju 78 %), kar pomeni, da ta segment trga, podobno kot segment lahkih dostavnih vozil, temelji predvsem na prodaji novih vozil. Vzroke za to lahko najdemo v strukturi kupcev tega segmenta gospodarskih vozil. Razdelimo jih lahko v dve skupini. Prvo skupino sestavlja majhno število kupcev z velikimi voznimi parki, ki vozila uporabljajo 10 in več let in so izključno kupci novih vozil. Drugo skupino kupcev predstavlja večje število manjših hitro rastočih prevoznih podjetij, ki svoje nakupe, predvsem zaradi finančnih omejitev, delno usmerjajo tudi na področje rabljenih avtobusov. Vendar pa, zaradi večje stroškovne učinkovitosti, glavnino njihovega voznega parka predstavljajo novi avtobusi.

Slika 8: Pregled prvič registriranih avtobusov v obdobju 1995–2015



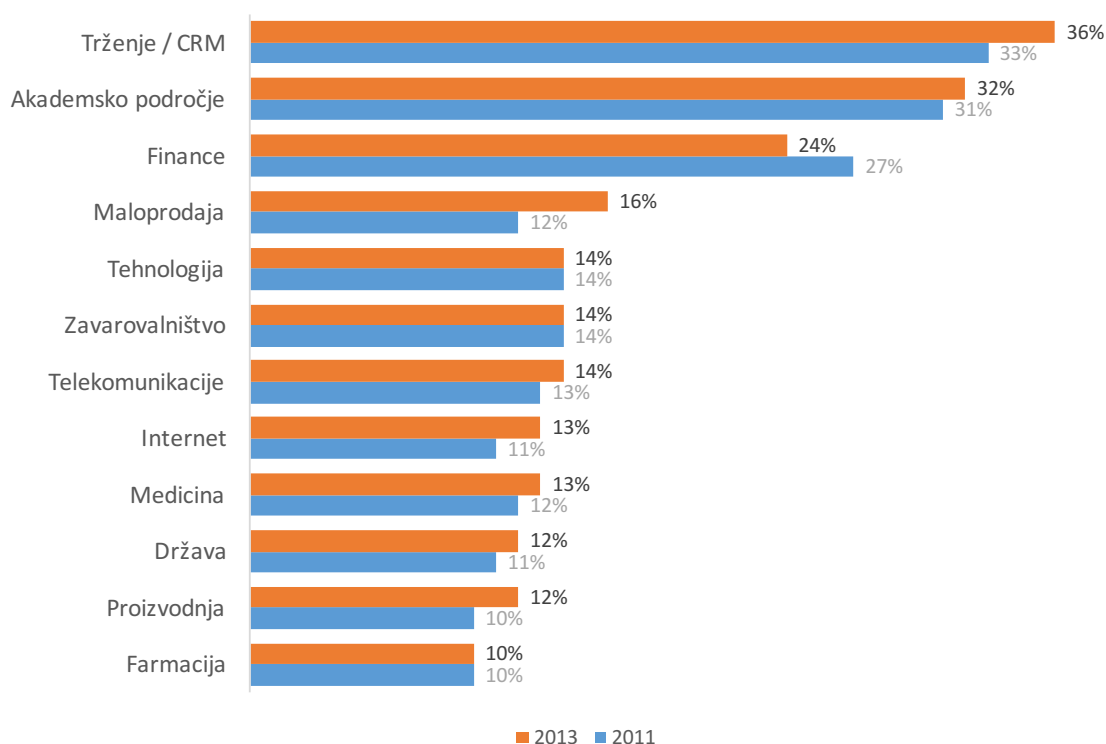
Vir: Ministrstvo za infrastrukturo, Evidenca registriranih vozil – preseki stanja 1992–2005, 2015;
Ministrstvo za infrastrukturo, Evidenca registriranih vozil – preseki stanja 2006–2015, 2016a;
lastni izračuni.

2 PREDSTAVITEV PODATKOVNEGA RUDARJENJA IN NAPOVEDNE ANALITIKE

Ljudje so že od nekdaj bili usmerjeni v prihodnost in napovedna analitika (angl. *predictive analytics*) je odsev težnje po napovedovanju dogodkov v prihodnosti. Napovedna analitika je del področja podatkovnega rudarjenja (angl. *data mining*), ki se osredotoča na iskanje uporabnih vzorcev v podatkih in je šele v zadnjih letih močno pridobila na popularnosti, kljub temu da so znanstvene podlage, na katerih temeljijo metode napovedne analitike stare že vsaj 40 let (Kotu & Deshpande, 2015, str. 1).

Po raziskavi Rexer Analytics (Rexer, 2013, str. 8) že kar 36 % uporabe metod podatkovnega rudarjenja pripada trženju, nekaj manj kot tretjina akademskemu področju, preostala tretjina pa ostalim področjem, kot prikazuje Slika 9.

Slika 9: Področja uporabe podatkovnega rudarjenja



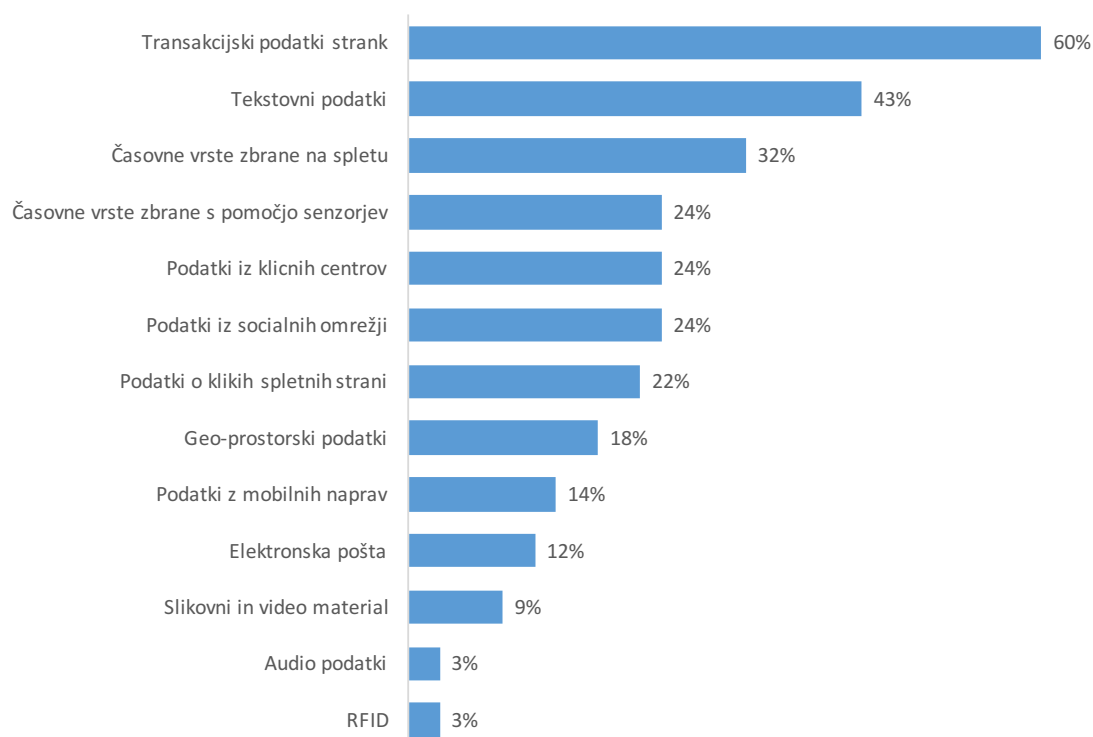
Vir: K. Rexter, 2013 Data Miner Survey Summary Report, 2013, str. 8.

Analitska hiša Gartner napoveduje, da bo do leta 2020 kar 40 % novih investicij s področja analitike in poslovne inteligence (angl. *business intelligence*) pripadlo ravno področju napovedne analitike (Gartner, 2016). Predvideva tudi, da bo do leta 2018 kar polovica velikih globalnih podjetij že uporabljala ta orodja in razvite lastne algoritme za izboljšanje konkurenčnega položaja na trgu.

Razlog za to lahko najdemo tudi v dejstvu, da se obseg podatkov s katerimi razpolaga človeštvo, iz leta v leto povečuje, poleg tega pa so v prihodu nove tehnologije, kot je internet stvari (angl. *internet of things*), ki bodo na relativno poceni način ustvarjale nove in nove podatke. Po drugi strani pa podjetja šele začenjajo razumevati pomen velikega obsega podatkov, saj samo 13 % anketiranih podatkovnih analitikov trdi, da imajo v njihovem podjetju vzpostavljen program za obdelavo velike količine podatkov (Rexer, 2013, str. 12).

Po podatkih raziskave Rexer Analytics (Rexer, 2013, str. 8) predstavljajo glavni vir podatkov, podatki zbrani pri spletnih nakupih. Te podatke namreč uporablja kar 60% anketiranih analitikov, saj so zaradi svoje globine in kakovosti najbolj verodostojni. Ostale vire podatkov, ki jih analitiki uporabljajo pri podatkovnem rudarjenju prikazuje Slika 10.

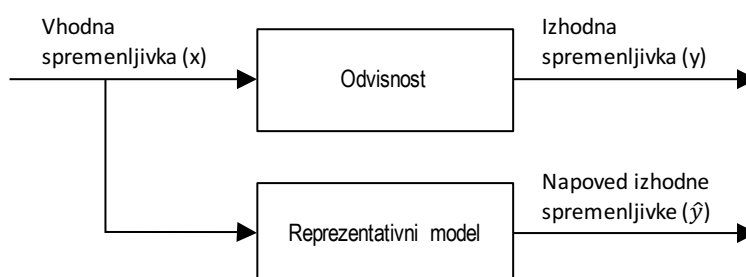
Slika 10: Viri podatkov za podatkovno rudarjenje



Vir: K. Rexter, 2013 Data Miner Survey Summary Report, 2013, str. 9.

V okviru podatkovnega rudarjenja s pomočjo napovedne analitike na osnovi podatkov iz opazovanj razvijamo reprezentativne modele, s katerimi lahko na eni strani pojasnimo razmerja med vhodnimi spremenljivkami in izhodno spremenljivko, in na drugi strani, na podlagi vrednosti vhodnih spremenljivk predvidevamo oz. napovedujemo (angl. *predict*) vrednost izhodne spremenljivke. Shematski prikaz tega procesa kaže Slika 11 (Kotu & Deshpande, 2015, str. 3).

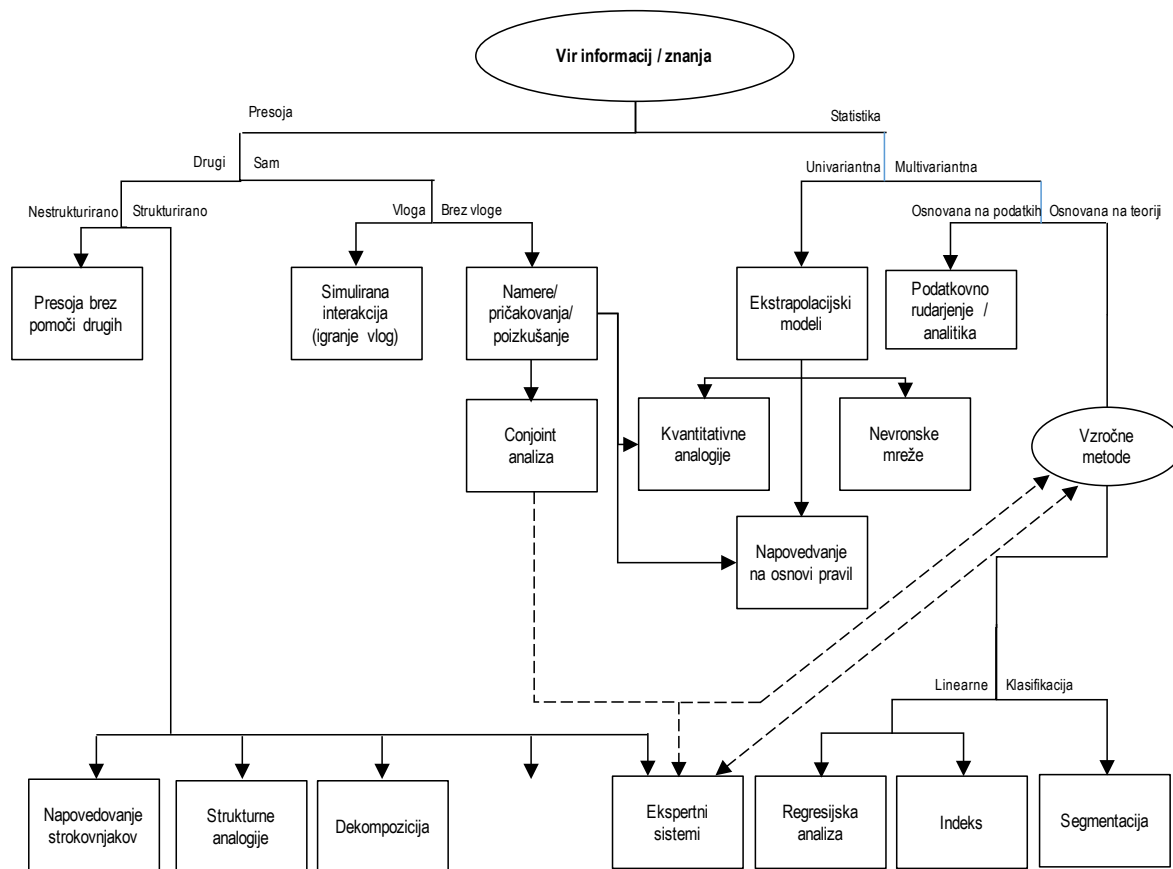
Slika 11: Napovedna analitika in modeliranje



Vir: V. Kotu & B. Deshpande, Predictive Analytics and Data Mining, 2015, str. 4.

Napovedno analitiko in podatkovno rudarjenje lahko postavimo tudi v širši okvir metod napovedovanja, ki temeljijo na uporabi podatkov in statistike, kot jih umeščata Armstrong in Green (2014) in jih shematično prikazuje Slika 12.

Slika 12: Metodološko drevo metod napovedovanja



Vir: J. S. Armstrong & K. C. Green, *Methodology Tree for Forecasting*, 2014.

Zanimiv strnjen pregled razvoja napovedovalnih metod najdemo v teh dveh intervjujih z ustanoviteljema mednarodnega inštituta napovedovalcev IIF (angl. *International Institute of Forecsasters*): Intervju z J. Scott Armstrong (Callopy, 2012) in Intervju s Spyrosom Makridakisom (Fides & Nikolopoulos, 2006).

Z namenom iskanja uporabnih vzorcev iz podatkov uporablja napovedna analitika različne metode in algoritme obdelave podatkov iz naslednjih področij (Kotu & Deshpande, 2015, str. 4):

- statistike,
- umetne inteligence (angl. *artificial intelligence*),
- strojnega učenja (angl. *machine learning*),
- teorije baz podatkov (angl. *database theory*),
- prepoznavanje vzorcev (angl. *pattern recognition*).

Vendar pa kljub napredku, ki ga je hiter razvoj računalništva dosegel v zadnjih letih, ena ključnih komponent za uporabo napovedne analitike ostaja poznavanja preučevane tematike (angl. *subject mater expertise*). Metode podatkovnega rudarjenja in napovedne analitike nam z uporabo algoritmov samo pomagajo izveči smiselne informacije iz obsežnih baz podatkov, ne morejo pa povsem nadomestiti vseh izkušenj, ki jih ima raziskovalec z obravnavano tematiko. Znani so primeri, ko pridemo do napačnih zaključkov zaradi slepega naslanjanja na rezultate algoritmov. Tako je v literaturi večkrat omenjen primer napačnega sklepanja o vzročnosti na podlagi korelacije med številom napadov morskih psov in prodajo sladoleda. Z analizo podatkov je seveda možno dokazati veliko korelacijo (angl. *correlation*) med tema dvema pojavoma, to pa še ne pomeni, da je med njima posledična povezava (angl. *causation*), saj visoke temperature na plaže privabijo veliko število ljudi, zaradi česar se poveča tako število napadov morskih psov kot prodaja sladoleda. Podoben komičen primer je opisan v članku »If TiVo Thinks You Are Gay, Here's How to Set It Straight« (Zaslow, 2002), kjer algoritem, uporabljen v sistemu za plačljivo televizijo, na podlagi obdelave podatkov o gledanju in snemanju televizijskih oddaj, sklepa na preference kupca.

Da se izognemo takim napakam Armstrong, Green in Graefe (2015) na podlagi obsežne eksperimentalne raziskave 70 člankov, predlagajo zlato pravilo napovedovanja – konservativnost. To pravilo dopolnjuje nabor 28 smernic za katere dokazujejo, da neizpolnjevanje ene od teh smernic poveča napako v povprečju za 44 %.

2.1 Proces podatkovnega rudarjenja

Metodološki postopek iskanja uporabnih povezav in vzorcev v podatkih je sestavljen iz niza iterativnih aktivnosti, združenih v proces podatkovnega rudarjenja.

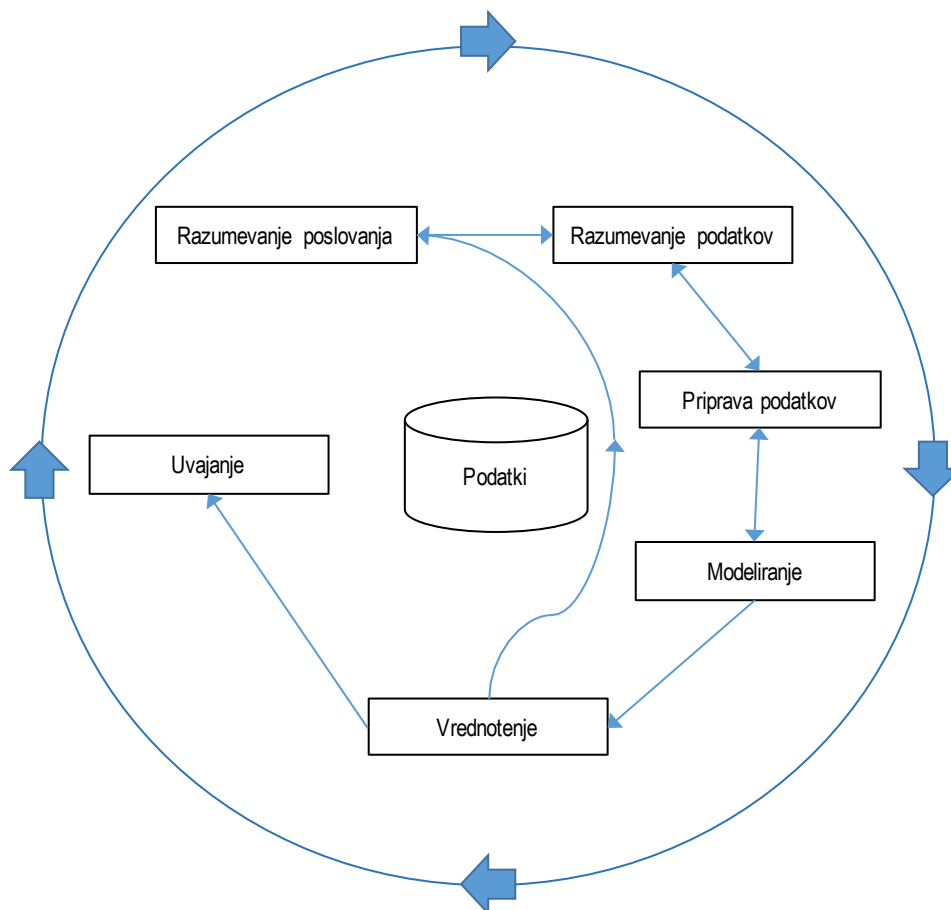
Eden najbolj pogosto uporabljenih referenčnih modelov (angl. *framework*) je CRISP-DM (akronim za angl. *Cross Industry Standard Process for Data Mining*). Ta model je razvil konzorcij več podjetij, ki se ukvarjajo s podatkovnim rudarjenjem (Chapman, Clinton, Kerber, Khabaza, Reinartz, Shearer & Wirth, 2000), in ga prikazuje Slika 13.

Model je sestavljen iz naslednjih faz:

- razumevanje poslovanja – definiranje poslovnih ciljev, ocena dejanskega stanja, določitev ciljev podatkovnega rudarjenja in izdelava projektnega načrta,
- razumevanje podatkov – zbiranje izhodiščnih podatkov, izvedba popisa podatkov, raziskava in kontrola kakovosti podatkov,
- priprave podatkov – izbira nabora podatkov, čiščenje podatkov, generiranje izvedenih podatkov, združevanje in ustrezno formatiranje podatkov,

- modeliranje – izbira tehnike modeliranja, izdelava testnega modela, izdelava modela, ocenjevanje modela in njegovih parametrov,
- vrednotenje – vrednotenje rezultatov, pregled celotnega procesa in določitev naslednjih korakov,
- uvajanje – izdelava načrta uvedbe, načrtovanje nadzora in vzdrževanja, izdelava zaključnega poročila in revizija projekta.

Slika 13: Faze CRISP-DM referenčnega modelov podatkovnega rudarjenja



Vir: P. Chapman et al., CRISP-DM 1.0 Step by Step Data Mining Guide, 2000, str. 10.

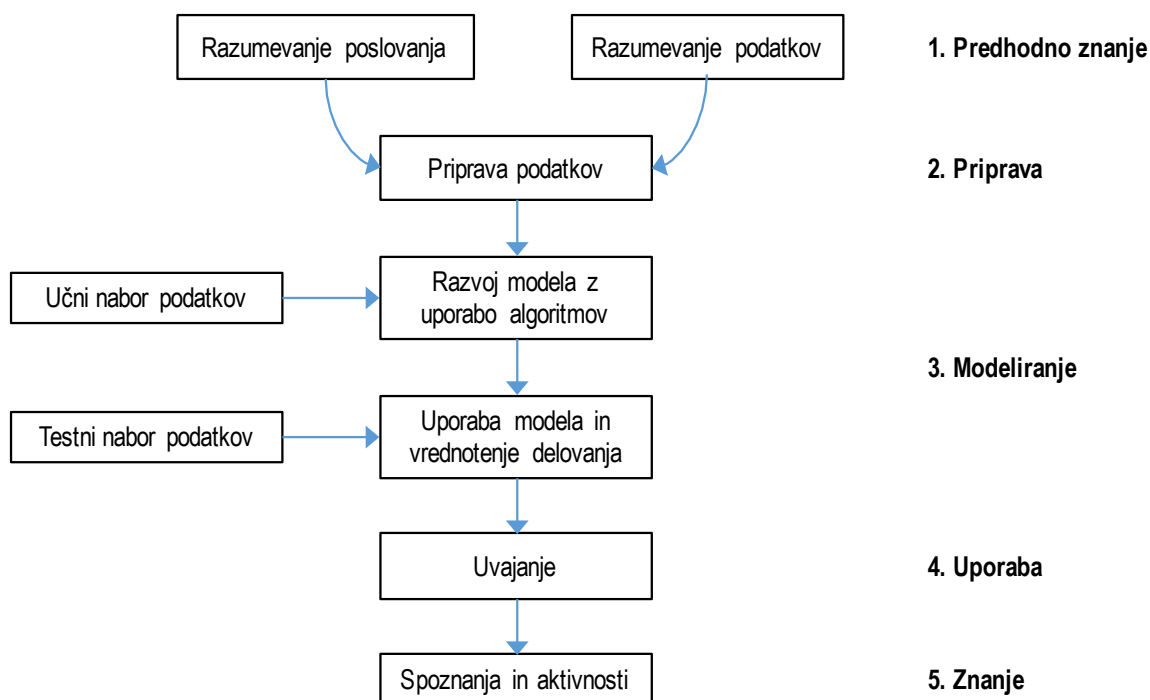
Na osnovi tega referenčnega modela Kotu in Deshpande (2015, str. 17–36) navajata proces podatkovnega rudarjenja. Ta predstavlja osnovo za izvedbo empiričnega dela tega magistrskega dela in je povzet v podpoglavjih 2.1.1 do 2.1.5. Uporabljen proces podatkovnega rudarjenja je sestavljen iz 5 korakov, pri čemer so bili za namen tega magistrskega dela dejansko uporabljeni samo prvi trije:

1. predhodno znanje,
2. priprava,
3. modeliranje,

4. uporaba,
5. znanje.

Shematični prikaz uporabljenega procesa podatkovnega rudarjenja prikazuje Slika 14, kratek opis posamičnih korakov pa sledi v nadaljevanju.

Slika 14: Proces podatkovnega rudarjenja



Vir: V. Kotu & B. Desphande, Predictive Analytics and Data Mining, 2015, str. 19.

2.1.1 Predhodno znanje

Predhodno znanje predstavlja že znane informacije v zvezi z obravnavanim problemom. Predhodno poznavanje problema nam pomaga prepoznati problem, kako je ta vključen v poslovni kontekst in katere podatke potrebujemo, da lahko problem rešimo. V konkretnem primeru so mi večletne izkušnje pri prodaji gospodarskih vozil pomagale omejiti problem (katere segmente bom preučeval), katere podatke potrebujem za rešitev problema in iz katerih virov lahko te podatke pridobim.

2.1.2 Priprava

V tem koraku podatke zberemo in jih pripravimo za uporabo. Ta korak vključuje naslednje aktivnosti:

1. Raziskovanje podatkov – z uporabo opisnih statistik in grafičnih prikazov poglobimo znanje o uporabljenih podatkih.
2. Preverjanje kakovosti podatkov – v tem koraku preverimo celovitost podatkov, ali so vrednosti posameznih polj logične, izločimo morebitne podvojene podatke ipd..
3. Obvladovanje manjkajočih vrednosti – pregledamo, ali nam kateri podatki manjkajo, saj so določeni algoritmi občutljivi na manjkajoče vrednosti posameznih vhodnih spremenljivk.
4. Določitev ustreznega tipa podatkov in njihova pretvorba – glede na vir podatkov je potrebno v nekaterih primerih posamičnim podatkom določiti ustrezen tip (npr. odstotek, celo število, kategorija ...). Nekateri algoritmi so uporabni samo za določen tip podatkov, zato je včasih potrebno podatke pretvoriti na, primer iz numeričnega tipa (celo število) v kategorični ali logični tip podatka.
5. Transformacija podatkov – nekateri algoritmi zahtevajo vhodne podatke, ki so numerični in normalizirani, saj izračunavajo medsebojno razdaljo v večdimenzionalnem prostoru. V takem primeru je potrebno podatke normalizirati.
6. Osamelci (angl. *outliers*) – to so po definiciji anomalije v naboru podatkov. Te je potrebno dovolj zgodaj identificirati in ovrednotiti, saj so lahko posledica napake ali pa predstavljajo povsem legitimen podatek. Cilj napovedne analitike je razviti reprezentativni model, ki bo odseval generalizirano odvisnost vhodnih spremenljivk od izhodne spremenljivke, zato lahko prisotnost osamelcev to odvisnost bistveno izkrivi. Na drugi strani pa je iskanje osamelcev lahko celo cilj modeliranja (npr. pri razpoznavanju zavarovalniških prevar), kar pomeni, da je njihova prisotnost v podatkih v določeni meri celo zaželena.
7. Izbor lastnosti (ang. *Feature Selection*) – v sodobnih modelih imamo pogosto opravka z velikim številom spremenljivk (več sto ali tisoč), kar je lahko ovira pri uporabi določenih algoritmov. Poleg tega so nekateri algoritmi občutljivi na medsebojno odvisnost med vhodnimi spremenljivkami. Zato je včasih potrebno nabor vhodnih spremenljivk omejiti tako, da ne pride do bistvenega poslabšanja napovednega modela.
8. Vzorčenje podatkov – z vzorčenjem del razpoložljivega nabora podatkov zaradi uporabe v analizi ali modeliranju izločimo. Za potrebe napovedne analitike pogosto razdelimo podatke na dva nabora – nabor za učenje, in nabor za testiranje. Z naborom za učenje zgradimo model, ki ga nato z naborom za testiranje preverimo na dejanskih podatkih.

2.1.3 Modeliranje

Model je abstraktna predstava podatkov in njihovih medsebojnih odnosov v danem naboru podatkov. Kotu in Deshpande (2015, str. 27) navajata, da je danes pri podatkovnem rudarjenju v uporabi več sto različnih algoritmov, ki so na voljo v različnih komercialnih kakor tudi odprtokodnih programskih orodjih. Modele, kot rezultat podatkovnega rudarjenja, lahko razdelimo v dve skupini:

1. Napovedni modeli (angl. *predictive models*) – ti modeli se na osnovi historičnih podatkov »naučijo« napovedovati izhodno spremenljivko iz vhodnih spremenljivk in spadajo v področje napovedne analitike. Primer takih modelov so regresijski modeli in modeli, ki temeljijo na analizi časovnih vrst.
2. Opisovalni modeli (angl. *descriptive models*) – ti modeli se uporabljajo za opisovanje obstoječega nabora podatkov in nimajo izhodnih spremenljivk. Primer takega modela je razvrščanje v skupine (angl. *clustering*) in modeli, temelječi na asociacijski analizi.

Glede na uporabljen algoritem in vrsto problema pa modele, ki so rezultat podatkovnega rudarjenja, lahko razdelimo v naslednje skupine (povzeto po Kotu & Deshpande, 2015, str. 9):

- klasifikacijski modeli,
- regresijski modeli,
- modeli za razvrščanje v skupine,
- modeli na osnovi asociacijske analize,
- modeli na osnovi analize časovnih vrst (angl. *time series forecasting*),
- modeli za rudarjenje teksta (angl. *text mining*),
- modeli za izbor nabora vhodnih spremenljivk (angl. *feature selection*),
- modeli za ugotavljanje anomalij (angl. *anomaly detection*).

V samem koraku modeliranja izvedemo naslednje aktivnosti:

- izbira algoritma za modeliranje – glede na vrsto problema se odločimo za ustrezen algoritem (npr. pri klasifikacijskem problemu za metodo odločitvenih dreves),
- uporaba nabora podatkov za učenje na izbranem algoritmu,
- vrednotenje modela – izvedemo ga na naboru podatkov za testiranje, saj se na ta način izognemo memoriranju podatkov oz. prevelikemu prilagajanju modela podatkom (angl. *overfitting*),
- združevanje modelov (angl. *ensemble modeling*) – gre za združevanje rezultatov več modelov, ki uporabljajo različne algoritme, združene v enoten model.

Armstrong in Green (2015, str. 1678) pri izbiri modela ugotavljata, da enostavni modeli dajejo v povprečju boljše rezultate kot kompleksni. Pri tem opredeljujeta enostavni model kot model, kjer so: metoda, prikaz akumuliranega znanja, odvisnosti v modelu in odvisnosti med modeli, napovedmi in odločitvami dovolj enostavno povezani, da jih nosilci odločitev lahko razumejo.

Ker smo priča zelo hitremu razvoju algoritmov, ki predstavljajo osnovo za modeliranje, Armstrong in Fildes (2006, str. 440) predlagata, da se izkušnje in praksa z novimi algoritmi prenašajo v obliki odprtokodne programske opreme in internetnih strani pod

pokroviteljstvom skupin SIG (angl. *Special Interest Group*), ki bi bedele nad pravilno uporabo teh algoritmov in skrbele za komunikacijo z zainteresirano javnostjo.

Na koncu koraka modeliranja dobimo zanesljiv in preverjen model, ki ga lahko uporabimo na novem (neuporabljenem) naboru podatkov.

2.1.4 Uporaba

V tem koraku razviti model predamo v uporabo. To v poslovni praksi pomeni, da razviti model vključimo v obstoječe poslovne procese in informacijsko infrastrukturo podjetja. V ta korak spadajo naslednje aktivnosti:

1. Ugotavljanje primernosti za uporabo – ali je razviti model primeren za uporabo v poslovnih procesih glede na razpoložljivost podatkov, poslovne procese, časovni okvir.
2. Tehnična integracija – napovedni model je bil po navadi razvit z enim izmed orodij za podatkovno rudarjenje (R, RapidMiner, SAS, SPSS ipd.). Ta orodja so namenjena razvoju modela, za uporabo v poslovnih procesih pa je potrebno napovedni model prenesti v poslovne aplikacije. Nekateri napovedni modeli (npr. regresijski, odločitvena drevesa) so dejansko relativno enostavne enačbe, ki jih je enostavno prenesti v poslovne aplikacije. Za lažji prenos modelov med informacijskimi sistemi je bil razvit poseben standardiziran datotečni format PMML (Predictive Model Markup Language), ki temelji na XML in za katerega razvoj in specifikacijo skrbi Data Mining Group (Data Mining Group, 2016).
3. Obvladovanje odzivnosti modela – nekateri modeli so sicer enostavni za razvoj, vendar pa so lahko računsko zelo zahtevni, zato je pri uporabi modela v praktičnih aplikacijah potrebno to ustrezno upoštevati.
4. Spremembe modela – okolje je podvrženo stalnim spremembam, zato je potrebno že pri uvajanju napovednega modela predvideti redno vzdrževanje, kjer se preveri ali so se ključni kriteriji upoštevani pri razvoju modela spremenili do te mere, da je potrebno izvesti ponovno modeliranje.
5. Asimilacija – pri nekaterih opisnih modelih vključitev v poslovne aplikacije ni glavni cilj modeliranja, temveč je lahko samo prenos spoznanj v poslovno prakso.

2.1.5 Znanje

Proces podatkovnega rudarjenja se začne s predhodnim znanjem, ki ga predstavlja znanje o podatkih in znanje o poslovanju oziroma poslovnem problemu, ki ga rešujemo. Znanje, pridobljeno s procesom podatkovnega rudarjenja, se v obliki izkušenj z uporabo algoritmov, nastavitvev algoritmov, izkušenj z obdelavo podatkov, prenese v novo znanje.

2.2 Pregled algoritmov napovedne analitike

Metode napovedne analitike lahko na kratko definiramo kot tiste metode iz nabora metod podatkovnega rudarjenja, ki so namenjene razvoju napovednih modelov. V grobem lahko probleme, s katerimi se srečujemo pri napovedni analitiki, razdelimo v dve skupini, in sicer na klasifikacijske in regresijske.

Klasifikacijski problemi so tisti, kjer želimo z vhodnimi spremenljivkami razdeliti vzorce v dva ali več različnih razredov. V tem primeru torej spremenljivka, ki jo želimo napovedati, ni zvezna, ampak diskretna (lahko numerična ali kategorična). Regresijski problemi pa se ukvarjajo z napovedovanjem numeričnih vrednosti. Izhodno spremenljivko s to vrsto metod opišemo kot funkcijo vhodnih spremenljivk. Posebno vrsto problemov predstavljajo časovne vrste, kjer je čas ena (v nekaterih primerih edina) vhodna spremenljivka.

Metode izbora vhodnih spremenljivk predstavljajo posebno vrsto problemov napovedovalne analitike, kjer zaradi optimizacije modelov ali pa manjše kompleksnosti modelov, želimo izbrati »prave« vhodne spremenljivke. V podpoglavjih 2.2.1 do 2.2.4 so na kratko predstavljeni algoritmi podatkovnega rudarjenja, ki se pretežno uporabljajo za razvoj napovednih modelov. Predstavitev je povzeta po opisih, kot jih predstavljata Kotu in Deshpande (2015, str. 63–193, str. 305–327, 347–370).

2.2.1 Klasifikacijski algoritmi

Klasifikacijski algoritmi so namenjeni uvrščanju posameznih vzorcev v dva ali več različnih razredov. S temi algoritmi, na osnovi znanih vrednosti vhodnih spremenljivk in diskretnih vrednosti izhodne spremenljivke, »naučimo« model, da zna ob znanih vhodnih spremenljivkah napovedati vrednost diskretne izhodne spremenljivke.

Glede dejstva, da v problemu, ki ga preučujem v magistrskem delu, tako na vhodu kot na izhodu, nastopajo zvezne spremenljivke, bom na tem mestu te algoritme samo naštel in na kratko opisal. Detajlni opis in primere modeliranja pa lahko najdemo v Kotu in Deshpande (2015, str. 63–163).

2.2.1.1 Odločitvena drevesa

Enostaven algoritem, kjer ima izhodna spremenljivka samo dve vrednosti (npr. »da« in »ne«). Izpeljanke tega algoritma lahko obvladajo tudi več vrednosti izhodne spremenljivke (Quinlan, 1986). Posebna izpeljanka tega algoritma so regresijska drevesa, ki omogočajo uporabo tudi numeričnih oz. zveznih izhodnih spremenljivk (Breiman, Friedman, Stone & Olshen, 1984).

2.2.1.2 Učenje pravil

Ta algoritem temelji na učenju na podlagi pravil po sistemu če-potem (angl. *if-then*) iz nabora vhodnih podatkov. Algoritem se pogosto uporablja tako za opisno kot napovedno analitiko. Primer uporabe je v preventivnem vzdrževanju strojev in naprav, kjer se na podlagi vhodnih kazalnikov z uporabo tega algoritma model odloči, kdaj je potrebno določen del preventivno zamenjati z namenom, da se prepreči okvara in s tem povezan izpad.

2.2.1.3 k-Najbližji sosed

Ta algoritem temelji na reku, da ptiči iste vrste letijo skupaj. Pri temu algoritmu nabor podatkov za učenje predstavlja vpogledno tabelo (angl. *lookup table*) kjer izhodno spremenljivko določimo na osnovi bližine vhodnih spremenljivk v n-razsežnem prostoru. Algoritem je sicer enostaven, vendar, če imamo opravka z velikim številom vhodnih spremenljivk, zahteva veliko računsko moč za uporabo v realnem času.

2.2.1.4 Naivna Bayesova metoda

Algoritem temelji na Bayesovem teoremu o pogojni verjetnosti. Algoritem uporablja predpostavko, da so vhodne spremenljivke med seboj popolnoma neodvisne, kar v praksi težko dosežemo (od tod tudi pridevnik naivna). Se pa algoritem dobro izkaže pri naboru podatkov, kjer imamo opravka z osamelci ali manjkajočimi vrednostmi, kar je velika prednost tega algoritma.

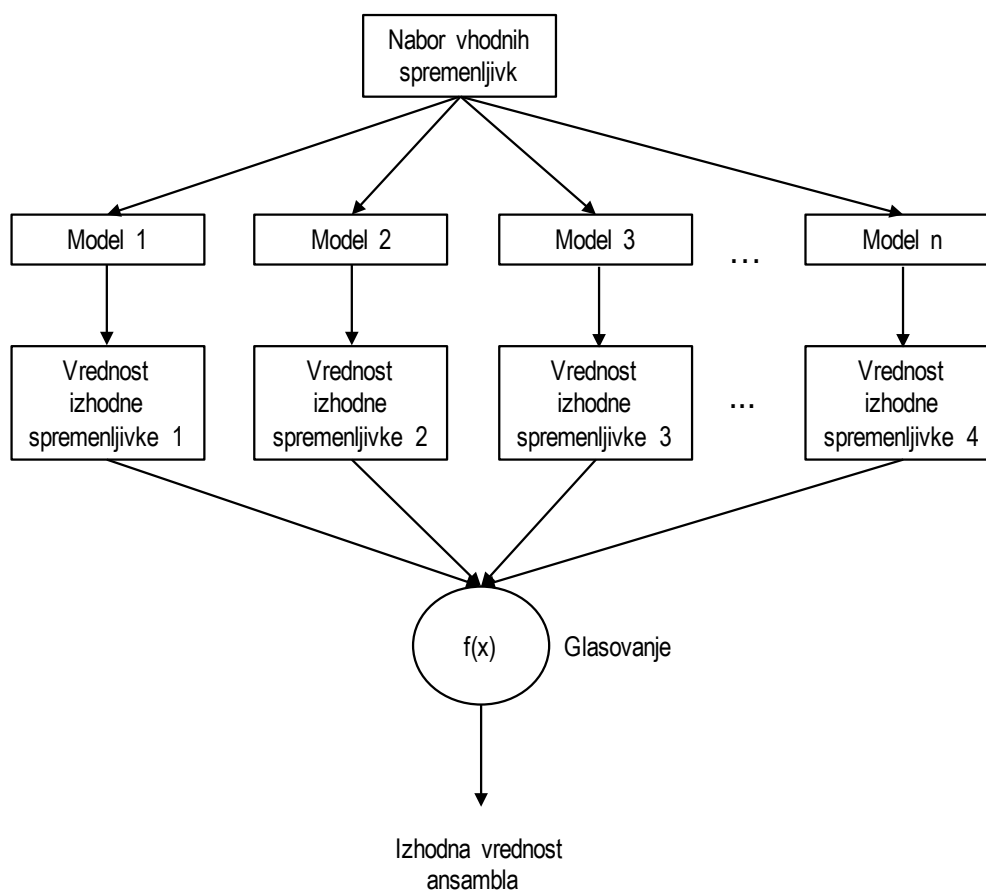
2.2.1.5 Umetne nevronske mreže

Algoritem umetnih nevronske mreže posnema delovanje nevrona v živčnem sistemu. Vhodne spremenljivke se preko več vozlišč preračunajo v izhodne. Ta algoritem se uporablja za modeliranje nelinearnih, kompleksnih odvisnosti med vhodnimi in izhodnimi spremenljivkami.

2.2.1.6 Učenje napovednih ansamblov

S tem algoritmom združimo rezultate več različnih modelov v en rezultat s pomočjo glasovanja. Način delovanja je podoben kot delo v skupini, kjer vsak član pride do svojega rezultata, nato pa skupina z glasovanjem odloči, kateri rezultat bo uporabila. Na ta način se zmanjša varianca napake, poveča napovedna moč modela in ublažijo posamične pomanjkljivosti uporabljenih algoritmov. Glede na to, da ima vsak od algoritmov napovedne analitike svoje prednosti in slabosti, je v praksi učenje napovednih ansamblov precej razširjeno. Shematski prikaz modela prikazuje Slika 15.

Slika 15: Model z uporabo učenja napovednih ansambllov



Vir: V. Kotu & B. Desphande, *Predictive Analytics and Data Mining*, 2015, str. 149.

2.2.1.7 Metoda podpornih vektorjev

Algoritem podpornih vektorjev razmeji učni nabor podatkov z mejami, ki jih definirajo podporni vektorji. Ko so z modelom te meje določene, model enostavno pogleda, v katero območje spada posamičen zapis glede na vhodne spremenljivke ter na ta način določi izhodno spremenljivko. Algoritem zahteva veliko računsko moč v fazi modeliranja, medtem ko je relativno enostaven za uvedbo v prakso.

2.2.2 Regresijski algoritmi

Regresijski algoritmi so najbolj razširjena tehnika, ki jo uporabljamo v napovedni analitiki (Rexter, 2013, str. 36). Zasnovana je tako, da vhodnim podatkom prilagodimo določeno matematično funkcijo (angl. *function fitting*). To pomeni, da lahko zapišemo, da je vrednost izhodne spremenljivke (y), funkcija vhodnih spremenljivk (X) oz. $y = f(X)$. Napovedovalna analitika za prileganje funkcije podatkom uporablja različne tehnike, od katerih sta najbolj pogosta linearna in logistična regresija.

2.2.2.1 Linearna regresija

Pri enostavnem linearnem regresijskem modelu predvidevamo, da je izhodna spremenljivka y linearno odvisna od (ene) vhodne spremenljivke x , tako kot kaže enačba (1).

$$y = b_0 + b_1x \quad (1)$$

V primeru da imamo več vhodnih spremenljivk (m), ta model imenujemo multipla linearna regresija, kjer odvisnost izhodne od vhodnih spremenljivk zapišemo, tako kot kaže enačba (2):

$$y = b_0 + b_1x_1 + b_2x_2 + \dots + b_mx_m \quad (2)$$

Algoritem multiple linearne regresije določi koeficiente b_i tako, da minimizira vsoto povprečne kvadratne napake, ki predstavlja razliko med napovedano vrednostjo ($\hat{y}_i$) in dejansko vrednostjo izhodne spremenljivke (y_i) za vsako posamično vrednost vhodnih in izhodne spremenljivke, ki so nam znane, kot prikazuje enačba (3):

$$\frac{1}{n} \sum_{i=1}^n e_i^2 = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (3)$$

Multipla linearna regresija se uporablja v napovedni analitiki tam, kjer je izhodna spremenljivka numerična vrednost. Pri uporabi v napovedni analitiki pa se je potrebno zavedati naslednjih 10 predpostavk, ki izhajajo iz klasičnega (Gaussovega) linearnega regresijskega modela (povzeto po Gujarati, 2003, str. 335):

1. regresijski model vsebuje samo linearne parametre, kot jih prikazuje enačba (2),
2. vrednosti x_i niso stohastične narave,
3. povprečna vrednost napak e_i je 0,
4. homoskedastičnost oz. da je varianca napake e_i konstantna pri dani vrednosti vhodne spremenljivke,
5. napake e_i pri danih vrednostih x_i med seboj niso korelirane oz. avtokorelirane,
6. kovarianca med napako e_i in vrednostmi vhodne spremenljivke x_i je enaka 0,
7. število opazovanj (n) je večje kot število vhodnih spremenljivk (m),
8. vrednosti vhodnih spremenljivk x_i morajo variirati,
9. regresijski model je pravilno specificiran oz. pri določitvi modela nismo naredili sistematične napake (npr. da so vhodne in izhodna spremenljivka nelinearno povezani),
10. v modelu ni prisotne popolne multikolinearnosti, kar pomeni, da vhodne spremenljivke med seboj niso popolno linearno odvisne,
11. napaka e_i je naključna in ima normalno porazdelitev.

Te predpostavke seveda v praksi ne veljajo vedno oziroma bolje rečeno vedno veljajo le delno. Zaradi teh praktičnih omejitev je potrebno, pri uporabi regresijskih modelov v praksi, upoštevati nekatera pravila, s katerimi se izognemo napačnemu tolmačenju rezultatov regresijskih modelov. Avtorji tako v literaturi navajajo različne korake, katerih cilj je čim bolj realno ovrednotiti dobljen regresijski model.

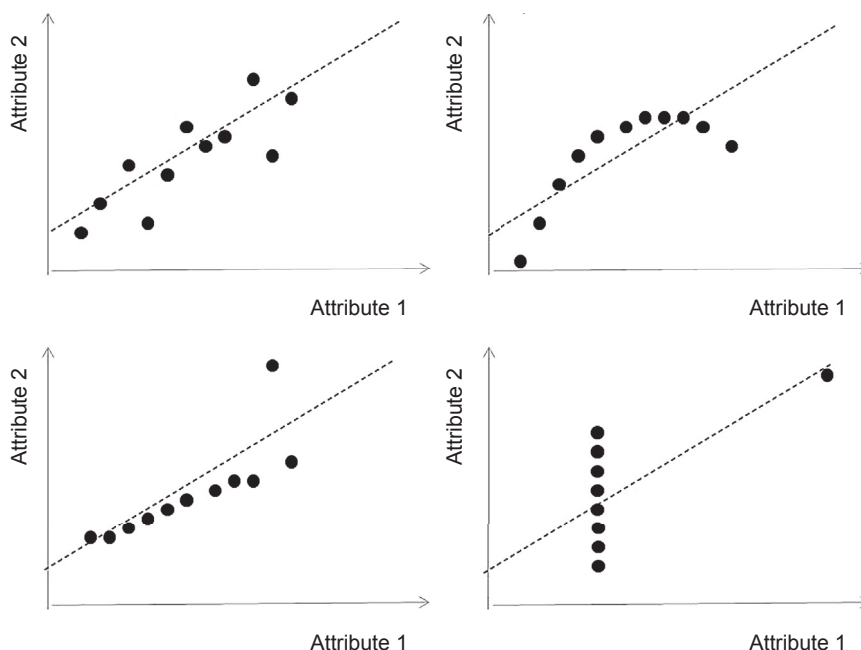
Tako Kotu in Deshpande (2015, str. 180) predlagata da se za validacijo regresijskega napovednega modela upoštevajo naslednje kontrolne točke:

1. Determinacijski koeficient (angl. *Squared Correlation* ali *Coefficient of Determination*) predstavlja razmerje med vsoto kvadratov razlike med ocenjeno vrednostjo ($\hat{y}$) in povprečno vrednostjo ($\bar{y}$) odvisne spremenljivke, ter vsoto kvadratov razlike med dejansko vrednostjo (y) in povprečno vrednostjo ($\bar{y}$) odvisne spremenljivke. Uporablja se kot glavni kriterij za oceno regresijskega modela. Model, ki ima R^2 bližje 1, pomeni boljši regresijski model. R^2 učinkovito pojasnjuje, koliko variabilnosti izhodne spremenljivke lahko pojasnimo z vhodnimi spremenljivkami. Definiran je z enačbo (4):

$$R^2 = \frac{\sum_1^n (\hat{y}_i - \bar{y})^2}{\sum_1^n (y_i - \bar{y})^2} \quad (4)$$

2. Če R^2 dobro pojasnjuje variabilnost izhodne spremenljivke (njegova vrednost blizu 1) potem morajo biti napake (angl. *residuals*), ki predstavljajo razliko med dejansko in ocenjeno vrednostjo (e), naključni. Zato je potrebno poleg vrednosti R^2 pri validaciji regresijskega modela, vedno preveriti tudi, ali so napake normalno porazdeljene. V kolikor temu ni tako, naš model vključuje sistematično napako, kar pomeni, da slabo popisuje odvisnosti izhodne spremenljivke od vhodnih.
3. Modeli z visoko nelinearnimi odvisnostmi med izhodno in vhodno spremenljivko bodo padli na zgornjih dveh kontrolnih točkah, kar pa še ne pomeni, da med njimi ne obstaja odvisnost. Najbolje to prikazuje Anscombov kvartet (Anscombe, 1973), kjer imata dve spremenljivki v vseh štirih primerih enak korelacijski koeficient, enako varianco, enako povprečno vrednost, vendar pa je njuna medsebojna odvisnost samo v enem primeru približno linearna (Slika 16). Ti primeri nakazujejo, da je potrebno pri analizi odvisnosti vhodnih in izhodne spremenljivke uporabljati tudi grafične prikaze, saj nas opiranje izključno na statistične lastnosti spremenljivk lahko zavede. Iz navedenega razloga sem se odločil, da v praktičnem delu magistrskega dela, poleg statističnih kazalnikov regresijskega modela, uporabim tudi grafične prikaze.

Slika 16: Anscombejev kvartet



Vir: F. J. Anscombe, *Graphs in Statistical Analysis*, 1973, str. 19–20.

Gujarati (2003, str. 335–490) ilustrativno prikazuje še dodatne izzive in rešitve, s katerimi se v praksi najpogosteje srečujemo pri modeliranju linearnih regresijskih modelov:

- Uporaba slamnatih vhodnih spremenljivk – ta metoda se uporablja, kadar želimo vključiti vhodne spremenljivke, ki imajo diskretne ali celo nominalne vrednosti.
- Multikolinearnost – o multikolinearnosti govorimo, kadar so vhodne spremenljivke med seboj korelirane, kar je še posebej pogosto pri določenih časovnih vrstah ekonomskih spremenljivk (npr. BDP in državna potrošnja). Posledica multikolineranosti je, da imajo izračunani regresijski koeficienti b_i izredno velike standardne odklone, kar zmanjšuje robustnost regresijskega modela.
- Majhna številčnost vzorca (angl. *micronumerosity*) – v kolikor imamo več vhodnih spremenljivk kot je velikost vzorca, je regresijske koeficiente nemogoče izračunati. Zato je potrebno zagotoviti, da je velikost vzorca v primerjavi z številom vhodnih spremenljivk dovolj velika.
- Heteroskedastičnost – o heteroskedastičnosti govorimo, kadar varianca napake e_i ni konstantna, kar se v praksi lahko zgodi (npr. kot posledica krivulje učenja pri tipkanju), vendar pa s tem kršimo predpostavko o konstantni varianci napake e_i .
- Avtokoreliranost napake e_i – do avtokoreliranosti napake e_i pogosto prihaja, kadar vhodne spremenljivke predstavljajo časovne vrste ekonomskih spremenljivk, ki so podvržene sezonskim, konjunkturalnim vplivom ali pa so lahko podvržene določenim obdelavam (npr. interpolacija, ekstrapolacija, povprečenje), ki v časovno vrsto vhodne spremenljivke vnesejo določen sistematičen vzorec.

Vse te navedene izzive moramo upoštevati pri ocenjevanju regresijskega modela oziroma se jim, kolikor je pač mogoče, poskušamo izogniti.

2.2.2.2 Logistična regresija

Logistična regresija se uporablja takrat, ko izhodna spremenljivka ni zvezna temveč diskretna. Glede na to, da imamo v našem primeru tako na vhodu kot na izhodu zvezne numerične spremenljivke, te vrste regresijskih metod ne bom podrobneje opisoval.

2.2.3 Napovedovanje časovnih vrst

Analiza časovnih vrst in napovedovanje časovnih vrst je ena izmed najstarejših tehnik podatkovnega rudarjenja.

Pri univariantni analizi časovnih vrst napovedovanje izhodne spremenljivke temelji izključno na vrednosti te iste spremenljivke v preteklosti. Predpostavljamo torej, da je vhodna spremenljivka enaka izhodni spremenljivki. Tehnike, ki se ukvarjajo s tako vrsto problemov, imenujemo metode napovedovanja osnovane na podatkih (angl. *data driven forecasting methods*).

Drugo vrsto tehnik, ki jih imenujemo metode napovedovanja, osnovane na modelih (angl. *model driven forecasting methods*), predstavljajo algoritmi, kjer predpostavljamo, da je izhodna spremenljivka funkcija časa. Te metode se razlikujejo od drugih algoritmov napovedne analitike po tem, da imajo na vhodu samo eno spremenljivko – čas.

2.2.3.1 Metode napovedovanja, osnovane na podatkih

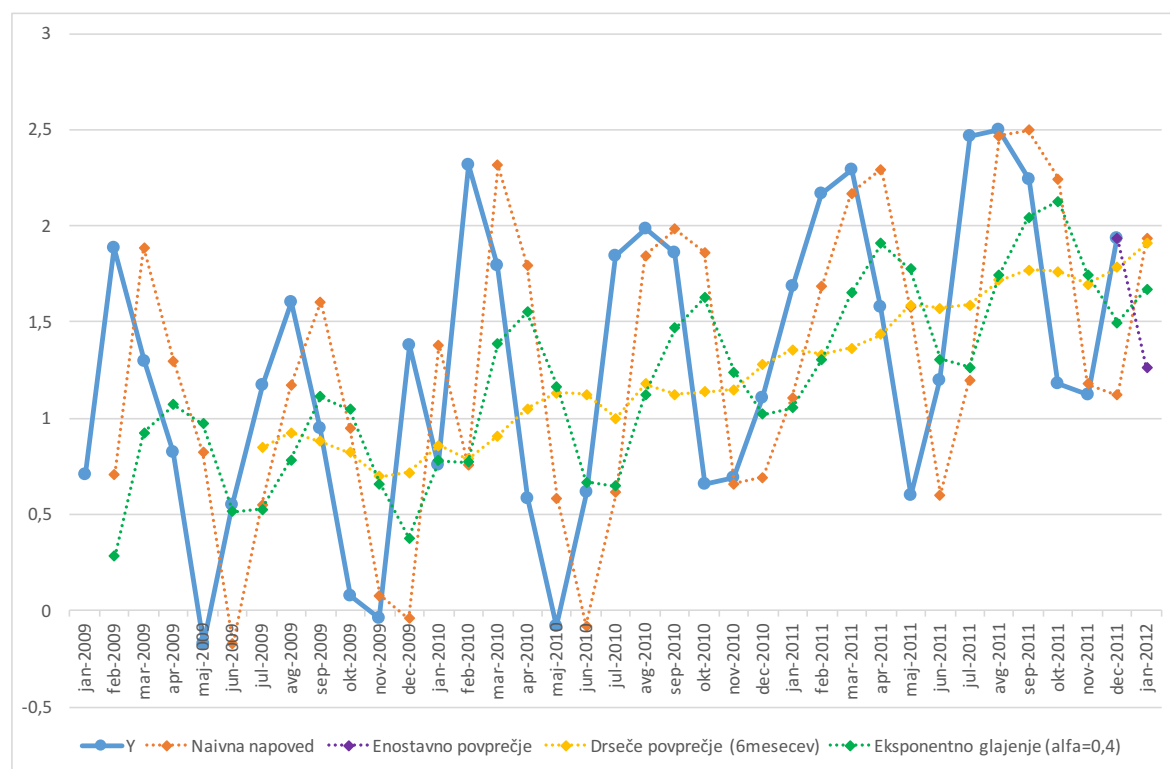
V to skupino spadajo naslednje metode:

- Naivno napovedovanje – kjer predpostavimo, da je vrednost izhodne spremenljivke enaka zadnji vrednosti izhodne spremenljivke v časovni vrsti $F_{n+1}=y_n$.
- Enostavno povprečje – tu predpostavimo, da je vrednost izhodne spremenljivke enaka povprečju vseh preteklih vrednosti te spremenljivke.
- Drseče povprečje – tu vrednost izhodne spremenljivke izračunamo kot povprečje zadnjih k vrednosti.
- Tehtano drseče povprečje – pri tej metodi predvidevamo, da imajo mlajše vrednosti v časovni vrsti večjo težo kot starejše in to upoštevamo pri izračunu tehtanega povprečja,
- Eksponentno glajenje – pri tej metodi izhodno spremenljivko (F_{n+1}) izračunamo iz tehtanega povprečja zadnje dejanske vrednosti (y_n) in napovedi zadnje dejanske vrednosti (F_n).

- Holtovo dvoparametersko eksponentno glajenje – napoved izhodne spremenljivke (F_{n+1}) temelji na dveh komponentah višine (L_n) in trenda (T_n).
- Holt-Winters triparametersko eksponentno glajenje poleg višine in trenda upošteva še sezonsko komponento vrednosti izhodne spremenljivke.

Skupna značilnost vseh metod napovedovanja, osnovanih na podatkih, je da omogočajo napoved samo za en korak v prihodnost (F_{n+1}). Primer uporabe nekaterih omenjenih metod osnovanih na podatkih kaže Slika 17.

Slika 17: Primer uporabe različnih metod napovedovanja osnovanih na podatkih



Vir: V. Kotu & B. Desphande, *Predictive Analytics and Data Mining*, 2015, str. 308; lastni izračuni.

2.2.3.2 Metode osnovane na modelih

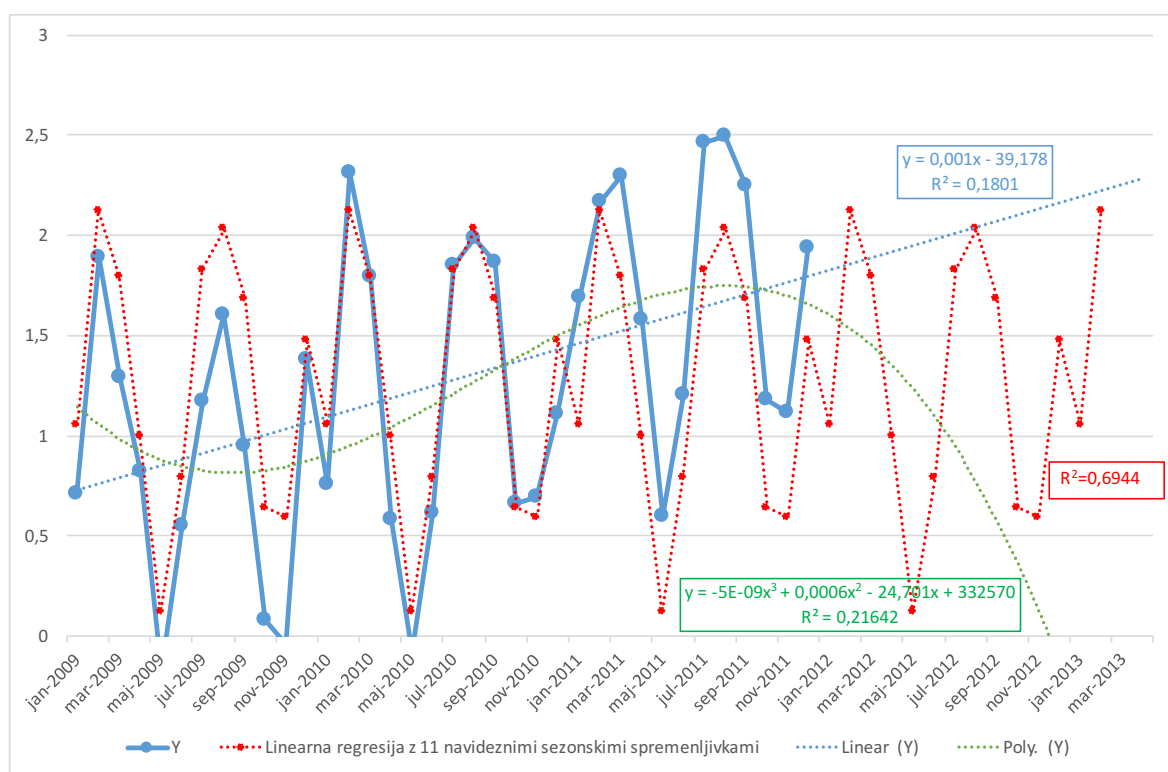
Metode, osnovane na modelih, ne poznajo omejitve napovedovanja samo za en korak v prihodnost (F_{n+1}). Skupna lastnost teh metod je, da je vrednost časovne vrste v prihodnosti funkcija preteklih vrednosti ($F_{n+h}=f(y_1, \dots, y_n)$). Med metode, osnovane na modelih prištevamo:

- Linearna regresija – pri tej metodi predvidevamo da je izhodna spremenljivka linearna funkcija časa.

- Polinomska regresija – pri tej metodi predvidevamo, da je izhodna spremenljivka polinomska funkcija časa.
- Linearna regresija z upoštevanjem sezonskosti – pri tej metodi modeliramo regresijski model z več spremenljivkami, eno predstavlja čas, ostale pa predstavljajo sezonskost (npr. četrtoletja ali mesece).
- Avtoregresijski modeli in ARIMA in njihove izvedenke predvidevajo, da so vrednosti v prihodnosti avtokorelirane z vrednostmi v preteklosti. Avtoregresijski modeli so dejansko regresijski modeli uporabljeni na preteklih vrednostih izhodne spremenljivke.

Primer uporabe linearne regresije, polinomske regresije in linearne regresije z uporabo slamnatih sezonskih spremenljivk, prikazuje Slika 18.

Slika 18: Primer uporabe metod osnovanih na modelih



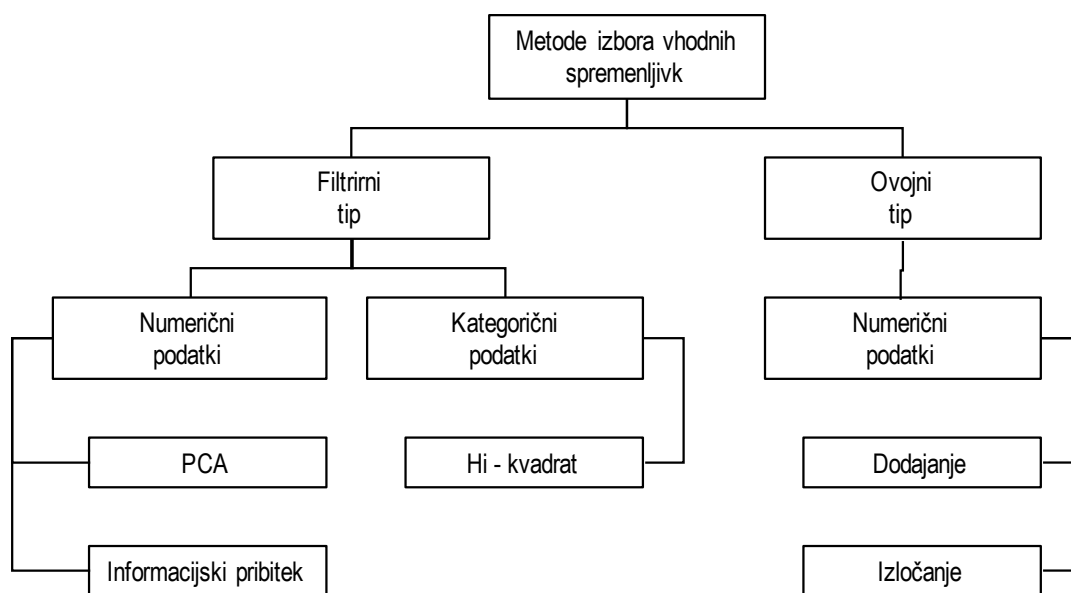
Vir: V. Kotu & B. Desphande, *Predictive Analytics and Data Mining*, 2015, str. 308; lastni izračuni.

Metode napovedovanja s pomočjo časovnih vrst bi lahko uporabil tudi pri oceni obsega posameznih segmentov gospodarskih vozil, vendar to ni predmet tega magistrskega dela, saj te metode predvidevajo, da je ocenjena vrednost velikosti segmenta odvisna izključno od časa in nobenih drugih spremenljivk. Se pa ta metoda pogosto uporablja v praksi v avtomobilski industriji za ocenjevanje obsega trga, še posebej kadar imamo opravka s sezonskimi nihanji in krajšimi časovnimi horizonti.

2.2.4 Metode izbora vhodnih spremenljivk

Kotu in Deshpande (2015, str. 348) za uporabo metod za izbor vhodnih spremenljivk navajata dva glavna tehnična razloga. Prvi je, da nabor podatkov lahko vsebuje visoko korelirane spremenljivke (npr. število prodanih enot in prihodki), kar po navadi pomeni, da z vključitvijo teh spremenljivk ne pridobimo dodatnih informacij. Poleg tega določeni algoritmi predvidevajo, da so vhodne spremenljivke med seboj popolnoma neodvisne (npr. naivni Bayesov). Pri regresijskih metodah se z ustreznim izborom nabora vhodnih spremenljivk želimo izogniti pojavi mulikolinearnosti, ko so posamične vhodne spremenljivke med seboj korelirane. V tem primeru se poveča standardna napaka pri izračunu regresijskih koeficientov, kar posledično pomeni nestabilen napovedni model. Drugi razlog je, da imamo lahko v naboru podatkov take podatke, ki nimajo nobene povezave z izhodno spremenljivko. Te povezave niso vedno znane vnaprej, še posebno, če imamo opravka z velikim številom vhodnih spremenljivk. Glede na način uporabe metode izbora vhodnih spremenljivk poznamo dve skupini metod nadzorovane in nenadzorovane. Pri nadzorovanih metodah izbora vhodnih spremenljivk, imenujemo jih tudi metode ovojnega tipa (angl. *wrapper type*), je algoritem izbora osredotočen na optimiranje določenega napovednega modela (npr. regresijskega). Pri nenadzorovanih metodah izbora vhodnih spremenljivk, imenujemo jih tudi metode filtrirnega tipa (angl. *filter type*), pa napovednega modela ne potrebujemo, algoritem sam izloči spremenljivke. Filtrirni tip se po navadi uporablja, kadar imamo v naboru podatkov izjemno veliko vhodnih spremenljivk in kadar je računska moč za izračun napovednega modela omejena. Pregled metod izbora vhodnih spremenljivk prikazuje Slika 19.

Slika 19: Splošni pregled metod izbora vhodni spremenljivk



Vir: V. Kotu & B. Desphande, *Predictive Analytics and Data Mining*, 2015, str. 351.

Za potrebe magistrskega dela sem se odločil, da uporabim metode ovojnega tipa, saj je metoda v programskem paketu RapidMiner izvedena tako, da hkrati s preprečevanjem multikolinearnosti, optimirajo tudi napovedno moč regresijskega modela.

2.2.4.1 Metode filtrirnega tipa

Med metode filtrirnega tipa prištevamo naslednje algoritme:

- PCA (angl. *Principal Component Analysis*) – ta algoritem deluje tako, da nadomesti obstoječi nabor vhodnih numeričnih spremenljivk z manjšim številom novih nadomestnih spremenljivk, ki so med seboj nekorelirane in s katerimi popišemo čim večji del variance osnovnih vhodnih spremenljivk.
- Filtriranje vhodnih spremenljivk na osnovi informacijske teorije – ta algoritem uporablja informacijski pribitek med vhodno in izhodno spremenljivko kot kriterij za izločanje, pri čemer mora biti izhodna spremenljivka nominalna.
- Filtriranje na osnovi Hi Kvadrat – ta metoda se uporablja kadar so vhodne in izhodne spremenljivke nominalne.

2.2.4.2 Metode ovojnega tipa

Metode ovojnega tipa izbora vhodnih spremenljivk delujejo tako, da optimirajo izbrani napovedni model tako, da postopoma iterativno dodajajo ali izločajo posamične spremenljivke in pri tem kontrolirajo, ali se je kakovost modela izboljšala ali ne.

Poznamo dva tipa teh algoritmov:

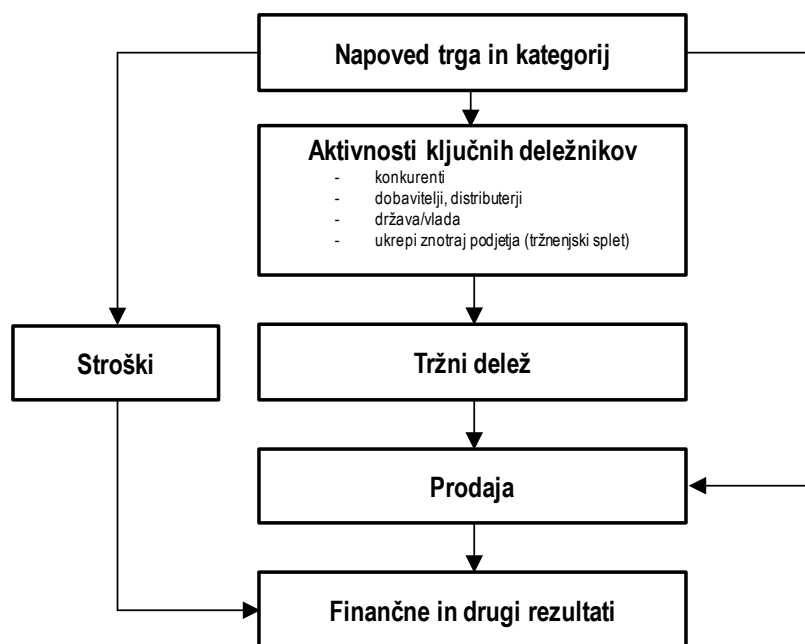
- vnaprejšnje dodajanje (angl. *forward selection*) – pri tem algoritmu z iterativnim dodajanjem vhodnih spremenljivk optimiramo napovedni model na določen kriterij in sicer tako, da pri vsaki iteraciji algoritem preveri, ali je dodana vhodna spremenljivka izboljšala napovedni model ali ne,
- povratno izločanje (angl. *backward elimination*) – pri tem algoritmu pa začnemo s polnim naborom vhodnih spremenljivk, nato pa algoritem iterativno izloča posamične vhodne spremenljivke in preverja, ali se je z izločitvijo posamične spremenljivke napovedni model poslabšal.

Pri tem je potrebno biti pozoren na dejstvo, da obe metodi delujeta iterativno in da z njuno uporabo ne dosežemo absolutnega, temveč lokalni optimum.

2.3 Uporaba napovedne analitike za napovedovanje prodaje v avtomobilski industriji

Osnovo za pravilno napovedovanje obsega prodaje predstavlja potreba po napovedovanju obsega trga oziroma segmenta trga, kot kaže Slika 1 po Armstrong in Brodie (1999, str. 2).

Slika 20: Komponente napovedi prodaje



Vir: J. S. Armstrong & R. J. Brodie, *Forecasting for Marketing*, 1999, str. 3.

Kot je omenjeno že v uvodu magistrskega dela je bilo napovedovanje obsega trga, v zadnjih 15. letih, deležno kar nekaj pozornosti različnih avtorjev. Enostaven in sistematičen pregled uporabe metod napovedovalne analitike v avtomobilski industriji sta pripravila Dudnehöffer in Dudnehöffer (2014, str. 367–377), ki ga na kratko povzemam v nadaljevanju. Glede na časovni horizont napovedi delita metode na dve glavni skupini:

- kratko- in srednjeročne napovedi – časovni horizont do 5 let,
- dolgoročne napovedi – časovni horizont od 5 do 20 let.

Pri dolgoročnih napovedih se uporabljajo predvsem analogije in scenariji na podlagi analize strukturnih podatkov, kot so: porast števila prebivalstva, gostota vozil (št. vozil na 1000 prebivalcev), rast bruto družbenega proizvoda na prebivalca in podobno. Pri kratko- in srednjeročnih napovedih pa sta v uporabi dve vrsti metod: metoda analize časovnih vrst in ekonometrični modeli na osnovi regresijskih metod.

2.3.1 Uporaba napovedi na osnovi analize časovnih vrst v avtomobilski industriji

Dudnehöffer in Dudnehöffer (2014, str. 369) zaradi specifične narave časovne vrste obsega trga delita časovno vrsto na naslednje komponente:

- trend – katere potek je determiniran z dolgoročnimi vzroki,
- cikel – katere potek je pogojen s posamičnimi konjunktturnimi cikli,
- sezona – katere potek je odvisen od posamičnih gibanj znotraj leta (npr. zima, pomlad),
- napaka – ta komponenta je naključna, niha okoli 0 in praviloma nima globljega vzroka.

Avtorja navajata, poleg teh splošnih komponent, še naslednje komponente specifične za avtomobilsko industrijo, ki jih je potrebno vključiti v pripravo napovedi:

- modelno načrtovanje avtomobilskih proizvajalcev (življenjski cikel modela),
- posebne tržne aktivnosti (promocije),
- cenovne spremembe na vozilih ali spremembe pri stroških, povezanih z vozili,
- spremembe razmer na povezanih trgih (trg rabljenih vozil, denarni trgi, finančni trgi),
- spremembe razmer v ekonomskem okolju (brezposelnost, rast proizvodnje, rast družbenega proizvoda, recesija itd.).

Vse te posebne učinke je pri časovni analizi posameznih splošnih komponent časovne vrste potrebno vključiti v napoved. Tako dobimo napovedni model, ki poleg funkcije časa vključuje tudi določene pojasnjevalne faktorje specifične za avtomobilsko industrijo.

2.3.2 Ekonometrični modeli

V nasprotju z modeli, ki temeljijo na časovnih vrstah, poskušamo z ekonometričnimi modeli na osnovi vhodnih spremenljivk, pojasniti izhodno spremenljivko (prve registracije novih vozil). Pri tem za vhodne spremenljivke praviloma uporabljamo makroekonomske spremenljivke. Po izkušnjah avtorjev (Dudnehöffer & Dudnehöffer, 2014, str. 373) na zasičenih trgih metode analize časovne vrste praviloma dajejo boljše napovedi kot ekonometrični modeli. Avtorja to pripisujeta natančnejši določitvi trenutka za zamenjavo vozila skozi zunanje kazalnike, kot so menjava modela, pogodbe o financiranju, povpraševanje na trgu rabljenih vozil in podobno.

2.3.3 Sprejemljivostni modeli

Sprejemljivostni napovedni modeli so podobni ekonometričnim modelom. Pri teh modeli nas zanima odnos med sprejemljivostjo novega proizvoda ali prodajnega kanala (izhodna spremenljivka) v odvisnosti od vhodnih spremenljivk. Vhodne spremenljivke so v teh modelih po navadi socialno-demografske lastnosti značilnosti (npr. starost, spol, izobrazba, zaposlenost ipd.) Kot pri ekonometričnih modelih se tudi tu uporabljajo regresijski modeli.

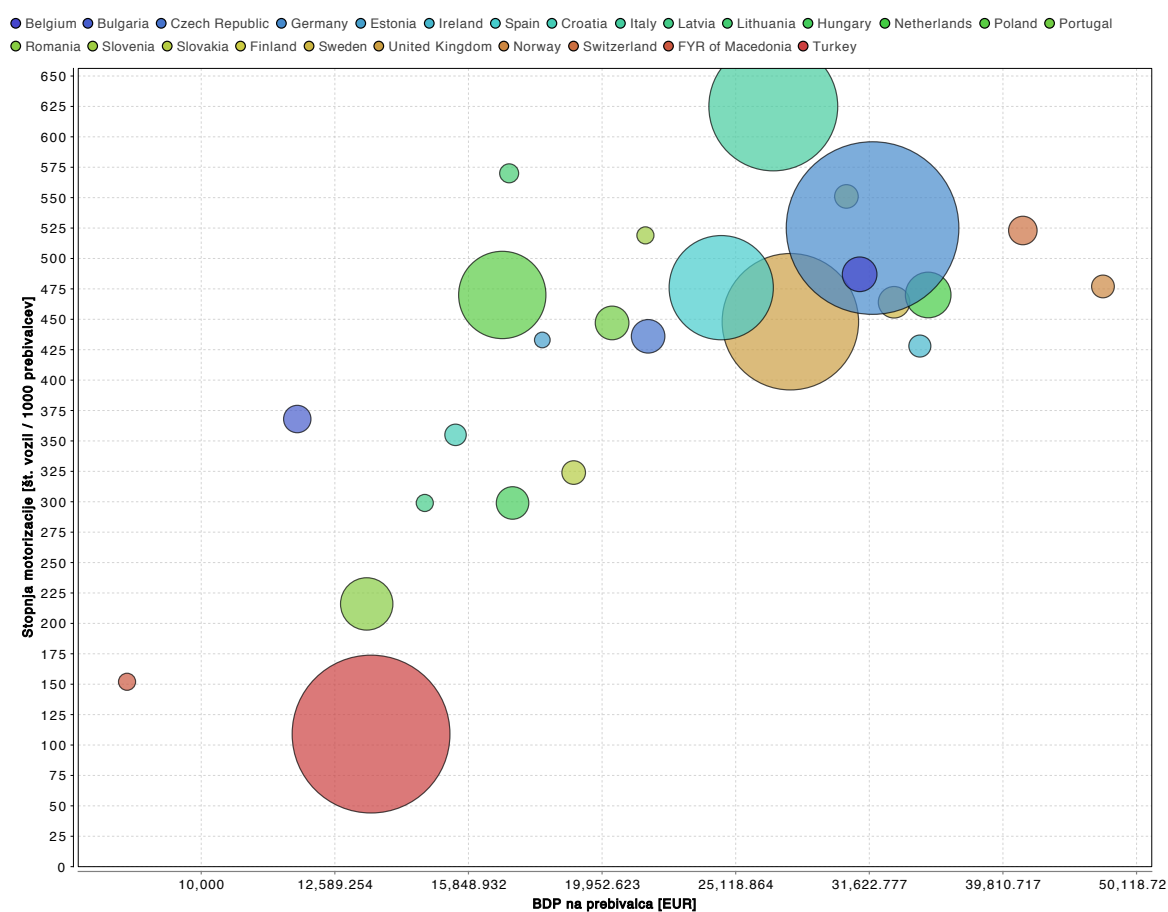
Za razliko od ekonometričnih modelov pa imamo pri sprejemljivostnih modelih med vhodnimi spremenljivkam pogosto tudi diskretne spremenljivke (spol, izobrazba, starost) zato v takem primeru uporabljamo logistične regresijske modele.

2.3.4 Dolgoročne napovedi

Ekonometrični modeli pri dolgoročnih napovedih in modeli temelječi na analizi časovnih vrst, niso več uporabni, saj je negotovost, ki jo prinašajo tehnološki razvoj, substituti, spremembe v obnašanju potrošnikov v posameznih panogah, prevelika, da bi te metode dale zadovoljive rezultate (Dudenhöffer & Dudenhöffer, 2014, str. 373).

Avtorja navajata, da je v praksi najbolj uporabna metoda stopnje motoriziranosti, ki se meri v številu vozil na 1000 prebivalcev. Pri tem se uporablja implicitna hipoteza, da se s povečanjem dohodka na prebivalca stopnja motoriziranosti povečuje (Slika 21). Iz napovedi povečanja dohodka na prebivalca in rasti števila prebivalcev nato lahko sklepamo kolikšen bo obseg števila vozil in s tem na obseg potrebe po novih vozilih.

Slika 21: Št. prebivalcev, BDP na prebivalca in stopnja motoriziranosti v letu 2011



Vir: Prirejeno po F. Dudenhöffer & K. Dudenhöffer, *Automobilmarkt-Prognosen: Modelle und Methoden*, 2014, str. 376; Eurostat; lastni izračuni.

3 PREDSTAVITEV UPORABLJENIH PODATKOV IN METODOLOGIJE

3.1 Predstavitev podatkov

3.1.1 Podatki registriranih vozilih

Podatke o registriranih gospodarskih vozilih sem dobil na portalu Nacionalnega interoperabilnostnega okvira (nio.gov.si), kjer so na voljo podatki o preseku stanja evidence vseh registriranih vozil na zadnji dan leta od leta 1992 dalje. Z uporabo poizvedbe v vsakoletni datoteki o stanju vseh registriranih vozil v Republiki Sloveniji (Ministrstvo za infrastrukturo, 2015; 2016a) sem izločil tri podatke za vsak segment vozil:

- Stanje – število vseh registriranih vozil posameznega segmenta,
- Prve_reg – število prvič registriranih vozil (prva registracija v tekočem letu L),
- Prve_reg_novi – število prvič registriranih novih vozil, pri čemer je bilo novo vozilo tisto prvič registrirano vozilo v letu L, katerega datum izdelave ni manjši od L–1.

Gospodarska vozila sem skladno s predstavljeno segmentacijo razdelil v tri segmente na podlagi naslednjih kriterijev iz baze registriranih vozil:

- lahka dostavna vozila – skupna dovoljena masa ≤ 3.500 kg,
 - 1995–2005 – polje [vrsta] = DV, TO, VV, TS,
 - 2006–2013 – polje [kategorija oznaka] = N1,
 - 2014–2015 – polje [J-Kategorija in vrsta vozila (oznaka)] = N1, N1G,
- težka tovorna vozila – skupna dovoljena masa ≥ 16.000 kg,
 - 1995–2005 – polje [vrsta] = DV, TO, VV, TS,
 - 2006–2013 – polje [kategorija oznaka] = N3,
 - 2014–2015 – polje [J-Kategorija in vrsta vozila (oznaka)] = N3, N3G,
- avtobusi,
 - 1995–2005 – polje [vrsta] = AV,
 - 2006–2013 – polje [kategorija opis] = avtobus,
 - 2014–2015 – polje [J-Kategorija in vrsta vozila (opis)] = avtobus.

3.1.2 Nabor vhodnih spremenljivk

Pregled 15. uporabljenih vhodnih spremenljivk, vir njihovih historičnih časovnih vrst in njihova frekvenca ter vir napovedi prikazuje Tabela 1, medtem ko Tabela 2 prikazuje enote za vsako posamično vhodno spremenljivko in ime spremenljivke, uporabljene v programskem paketu RapidMiner.

Tabela 1: Nabor uporabljenih vhodnih spremenljivk

Vhodna spremenljivka	Vir historičnih podatkov*	Frekvenca podatkov**	Vir napovedi
BDP Slovenija	SURS	L, K	UMAR
Št. registriranih brezposelnih	ZRSZ	L, M	UMAR
Izvoz blaga in storitev	SURS	L, K	UMAR
Uvoz blaga in storitev	SURS	L, K	UMAR
Državna potrošnja	SURS	L, K	UMAR
Zasebna potrošnja	SURS	L, K	UMAR
Investicije v osnovna sredstva	SURS	L, K	UMAR
Indeks cen življenjskih potrebščin	SURS	L, M	UMAR
Cena nafte	ECB	L, K	UMAR
Uvoz blaga in storitev Nemčija	ESTAT, SB	L, K	ifo
Izvoz blaga in storitev Nemčija	ESTAT, SB	L, K	ifo
Dodana vrednost gradbeništvo	SURS	L, K	UMAR
Dodana vrednost trgovina in popravila vozil, promet in skladiščenje, gostinstvo	SURS	L, K	UMAR
Dodana vrednost rudarstvo, predelovalne dejavnosti, oskrba z elektriko in vodo, ekologija	SURS	L, K	UMAR
Obrestna mera***	BS, ECB	L, M	EFA

Legenda:

* SURS=Statistični urad Republike Slovenije (b.l.), ZRSZ=Zavod Republike Slovenije zaposlovanje (b.l.), BS=Banka Slovenije (b.l.), ECB=Evropska centralna banka (b.l.), UMAR=Urad za makroekonomske analize in razvoj (2016), ifo=ifo Institute (2016), EFA=Economy Forecast Agency (b.l.), ESTAT=Eurostat (b.l.), SB=Statistisches Bundesamt (b.l.).

** L=letno, K=kvartalno, M=mesečno.

*** časovno vrsta podatkov za povprečno obrestno mero Banka Slovenije objavlja samo za časovno obdobje 2003–2016 zato je za obdobje 1995 do 2002 uporabljena obrestna mera sestavljena iz 12m EURIBOR in povprečne razlike med povprečno obrestno mero in 12m EURIBOR v letih 2003–2006.

Tabela 2: Pregled nabora vhodnih spremenljivk in uporabljenih enot

Vhodna spremenljivka	Ime spremenljivke v RapidMiner	Enota
BDP Slovenija	[BDP]	mio EUR stalne cene 2010
Št. registriranih brezposelnih	[Reg_brezposelnost]	število (letno povprečje)
Izvoz blaga in storitev	[Izvoz]	mio EUR stalne cene 2010
Uvoz blaga in storitev	[Uvoz]	mio EUR stalne cene 2010
Državna potrošnja	[Drzavna_potrosnja]	mio EUR stalne cene 2010
Zasebna potrošnja	[Zasebna_potrosnja]	mio EUR stalne cene 2010
Investicije v osnovna sredstva	[Investicije_v_OS]	mio EUR stalne cene 2010
Indeks cen življenjskih potrebščin	[Inflacija]	indeks povprečje obdobja/ povprečje prejšnjega obdobja
Cena nafte	[Cena_nafte]	EUR za sod
Uvoz blaga in storitev Nemčija	[Export_DE]	mrd EUR stalne cene 2010
Izvoz blaga in storitev Nemčija	[Import_DE]	mrd EUR stalne cene 2010
Dodana vrednost gradbeništvo	[DV_gradbenistvo]	mio EUR stalne cene 2010
Dodana vrednost trgovina in popravila vozil, promet in skladiščenje, gostinstvo	[DV_trgovina]	mio EUR stalne cene 2010
Dodana rudarstvo, predelovalne dejavnosti, oskrba z elektriko in vodo, ekologija	[DV_gradbenistvo]	mio EUR stalne cene 2010
Obrestna mera	[Obrestna_mera]	v %, povprečje mesečnih povpr. vrednosti

Pri izboru vhodnih spremenljivk sem upošteval naslednje kriterije:

- na voljo so javno dostopne napovedi za tekoče leto L, L+1 in praviloma L+2,
- napovedi pripravljajo priznane institucije (UMAR, ifo Institut) in so javno dostopne na spletu v prvi polovici tekočega leta (L),
- na voljo so javno dostopni podatki o časovnih vrstah od leta 1995 do 2015 (historični podatki večine ekonomskih spremenljivk za čas pred letom 1995 v večini primerov niso na voljo),
- časovne vrste spremenljivk, izraženih v EUR, so na voljo v stalnih cenah iz leta 2010,
- iz lastnih izkušenj predvidevam, da imajo izbrane spremenljivke vpliv na obseg trga gospodarskih vozil.

Historične podatke za navedenih 15. vhodnih spremenljivk za obdobje 1995–2015 sem dopolnil s podatki iz spomladanske napovedi UMAR za leto 2016, 2017 in 2018 (UMAR, 2016) ter Ifo za leto 2016 in 2017 (Ifo, 2016).

Tako sta za vsak obravnavan segment trga (težka tovorna vozila, avtobusi, lahka dostavna vozila) nastala dva nabora podatkov (letni in kvartalni nabor) s 15 neodvisnimi spremenljivkami, 1 identifikatorjem [Leto] oz. [Leto-kvartal] ter 1 odvisno spremenljivko [Prva_reg_novi] in 24 oziroma 96 zapisi.

V času priprave magistrskega dela, napoved o nemškem izvozu in uvozu ([Export_DE] in [Import_DE]) za leto 2018 še ni bila objavljena, zato sem ti dve vrednosti pustil prazni. Napovedi vhodnih spremenljivk za leto 2016, 2017 in 2018 so na voljo samo na letni ravni, ne pa za posamični kvartal, zato sem za napovedi kvartalnih vrednosti uporabil napovedano letno stopnjo rasti/padca in jo uporabil kot predvideno stopnjo rasti oziroma padca kvartal na kvartal.

V kvartalni nabor podatkov sem dodal še 3 slamnate (angl. *dummy*) sezone spremenljivke, s katerimi popišemo sezonski vpliv posamičnega kvartala. Kot navajajo Hyndman in Athanasopoulos (2016, poglavje 5.2), Gujarati (2003, str. 297–324), Kotu in Deshpande (2015, str. 314) lahko na ta enostaven način bistveno izboljšamo regresijski model, kjer je v podatkih opazen sezonski vpliv. Izsek iz podatkovne tabele prikazuje Tabela 3.

Tabela 3: Slamnate sezone spremenljivke in njihove vrednosti

[Leto-kvartal]	[Kvartal_1]	[Kvartal_2]	[Kvartal_3]
1995Q1	1	0	0
1995Q2	0	1	0
1995Q3	0	0	1
1995Q4	0	0	0
1996Q1	1	0	0
1996Q2	0	1	0
1996Q3	0	0	1
1996Q4	0	0	0
...	...	...	...
2018Q1	1	0	0
2018Q2	0	1	0
2018Q3	0	0	1
2018Q4	0	0	0

3.1.3 Napovedi vhodnih spremenljivk

Cilj magistrskega dela je razviti regresijski model, s katerim lahko na osnovi vhodnih spremenljivk ocenimo obseg posamičnega segmenta gospodarskih vozil, pri čemer je časovni horizont omejen na tekoče leto in še dve leti v naprej (v nadaljevanju L+2). Historične časovne vrste vhodnih spremenljivk sem pridobil iz javnih spletnih baz podatkov različnih institucij, ki te podatke sistematično zbirajo (Statistični Urad Republike Slovenije, Evropska Centralna Banka, Banka Slovenije). Podatki so na voljo v ustrezni obliki in formatu, tako da pri tem nisem naletel na večje težave.

Glavno omejitev pri izboru nabora vhodnih spremenljivk pa predstavljajo njihove napovedi za časovni horizont L+2. Za napoved vhodnih spremenljivk sem kot najprimernejšo izbral napoved UMAR (2016), saj je časovno na voljo med prvimi in so tako podatki na voljo še pred mesecem junijem, ko se po navadi začnejo prve aktivnosti pri pripravi poslovnih načrtov za naslednji dve leti. Pri tem je dodaten argument pri izbiri predstavljalo dejstvo, da UMAR pripravlja še jesensko napoved, ki je po analizi Banke Slovenije pričakovano bistveno bolj natančna kot spomladanska (Cimperman & Savšek, 2014).

Za napoved nemškega izvoza in uvoza sem vzel napoved priznanega nemškega inštituta ifo (ifo Institut, 2016). Ta objavlja svojo napoved skupaj z analizo gospodarskih gibanj za države EU v svoji publikaciji ifo Schnelldienst in je prav tako na voljo že konec aprila. Žal ima omenjena napoved časovni horizont samo do L+1, kar pomeni, da je potrebno v praksi napoved teh dveh vhodnih spremenljivk za L+2 napovedati oziroma oceniti.

Pri uporabi napovedi makroekonomskih spremenljivk kot vhodnih spremenljivk za potrebe napovednih modelov je pri uporabi rezultatov teh modelov potrebno upoštevati tudi dejstvo, da te napovedi niso vedno točne. Cimperman in Savšek (2014, str. 10) pri analizi natančnosti napovedi različnih tako domačih kot tujih institucij med drugim ugotavljata sledeče:

- do največjih napak napovedi je prišlo v prvih dveh letih finančne in gospodarske krize,
- natančnost jesenske napovedi je za tekoče leto praktično brez izjeme, za naslednje dve leti pa je praviloma bolj natančna od spomladanskih napovedi,
- pri napovedovanju BDP in inflacije je tako pri domačih kot pri tujih napovedih pogosto viden optimizem napovedovalcev, ki se kaže kot precenjevanje gospodarske aktivnosti in podcenjevanje inflacije,
- večje precenjevanje je zaznati pri spomladanski napovedi zasebne potrošnje in bruto investicij za tekoče leto, nekoliko manjše pa je za uvoz blaga in storitev,
- najbolj podcenjene so napovedi državne potrošnje in izvoz blaga in storitev.

3.2 Kratek opis uporabljene metodologije

3.2.1 Izbor vhodnih spremenljivk

Za potrebe optimizacije regresijskega modela sem izbral algoritem s povratnim izločanjem vhodnih spremenljivk (angl. *Backward Elimination Feature Selection*), saj je za napovedno moč regresijskega modela bolje, da ostane v modelu kakšna vhodna spremenljivka preveč kot kakšna premalo (Gujarati, 2003, str. 514).

Tako na vhodu v proces modeliranja vstopa vseh 15 vhodnih spremenljivk, ki jih prikazujeta Tabela 1 in Tabela 2. V samem procesu modeliranja sem nato operator »Backward Elimination« nastavil tako, da je z odvzemanjem posamičnih vhodnih spremenljivk algoritem optimiral regresijski model na vrednost determinacijskega koeficienta R^2 toliko časa, dokler se z odvzemom dodatne vhodne spremenljivke vrednost R^2 ne zmanjša za več kot 5 %.

3.2.2 Kratek opis uporabljenih procesov modeliranja

Programski paket RapidMiner Studio (v nadaljevanju RapidMiner) omogoča zelo enostavno in z grafičnim uporabniškim vmesnikom podprto modeliranje kompleksnih obdelav podatkov. Pri svojem delu sem uporabil RapidMiner Studio Community 7.1.000. Obdelane podatke sem za lažjo grafično prezentacijo izvozil v datoteko v formatu Microsoft Excel (.xlsx), ki mi je služila kot osnova pri izdelavi grafičnih prikazov in izvedbi dodatnih izračunov.

Zaradi avtomatizacije določenih operacij in možnosti hitrejših prilagoditev novim podatkom sem za vsak segment trga izdelal svoj nabor procesov, ki se ločijo med seboj glede na segment trga in uporabljen nabor podatkov (četrtni, letni).

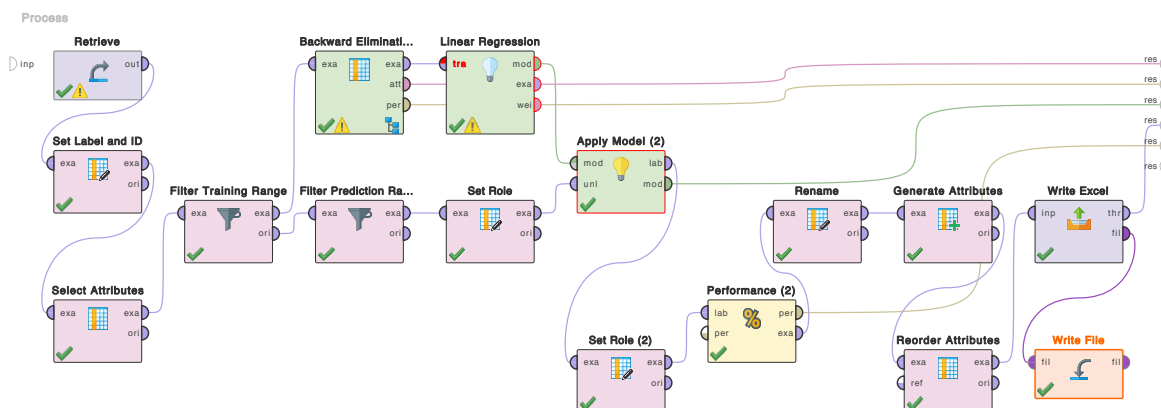
Segment trga in nabor podatkov (četrtni, letni) ne vplivata na sam proces, temveč samo na nekatere kazalnike znotraj procesa (obseg podatkov, izbor spremenljivk). Zato v nadaljevanju sledi kratek opis uporabljenega procesa v RapidMiner-ju.

Ta proces, prikazuje ga Slika 22, je sestavljen iz naslednjih korakov:

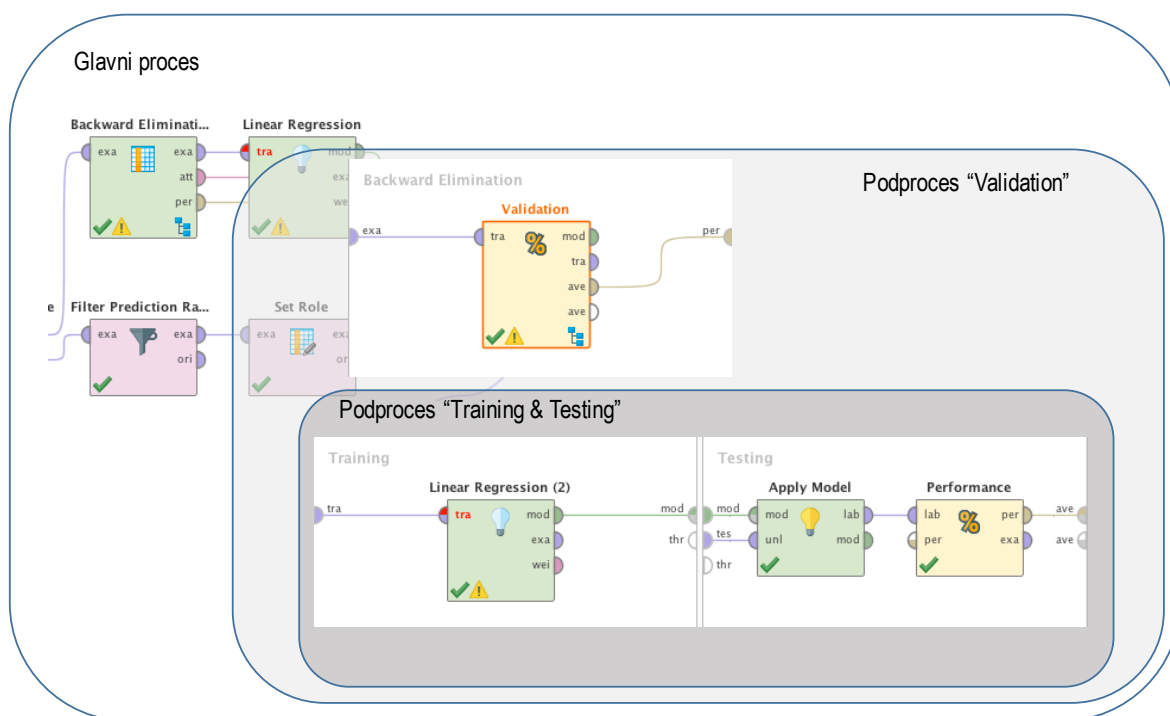
- z operatorjem »Retrieve« naložimo tabelo z podatki,
- z operatorjem »Set Label and ID« določimo odvisno spremenljivko [Prve_reg_novi] in identifikacijsko spremenljivko [Leto] oz. [Leto-kvartal],
- z operatorjem »Select Attributes« izberemo nabor vseh 15 vhodnih spremenljivk,
- z operatorjem »Filter Training Range« izberemo obseg za učenje, originalni obseg pa ponovno uporabimo v operatorju »Filter Prediction Range«, kjer izberemo vse zapise,

- v podproces »Backward Elimination« je ugnезden podproces »Validation« tako, kot kaže Slika 23,

Slika 22: Uporabljen proces v RapidMiner-ju za modeliranje regresijskega modela



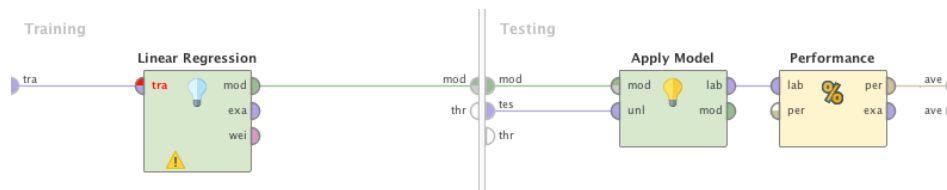
Slika 23: Shematska ponazoritev operatorja »Backward Elimination« in njegovih podprocesov



- v podprocesu »Validation« pa se skriva podproces »Training & Testing«, ki ga kaže Slika 23 in je sestavljen iz dveh delov:

- z operatorjem »Linear Regression« na delu učnega obsega podatkov izračunamo regresijski model (Training), ki ga z dobljenim modelom uporabimo na drugem delu učnega obsega podatkov (Testing),
- z operatorjem »Performance« izračunamo ocenjevalne kazalnike (RMSE, R^2), na osnovi katerih ocenjujemo dobljeni model,

Slika 24: Podproces »Training & Testing« izračun in preverjanje regresijskega modela



- podproces »Validation« je nastavljen tako, da pri letnem obsegu podatkov za izračun modela uporabi 15 zapisov, za testiranje modela pa 6 zapisov, medtem ko na četrtletnem naboru podatkov uporabi 60 zapisov za izračun modela in 25 zapisov za testiranje modela (podobsega za izračun modela in testiranje modela sta naključno izbrana),
- podproces »Validation« sem nastavil tako, da eden od izhodov podprocesa (mod) predstavlja model, ki je optimiran na čim višjo vrednost R^2 ,
- ta regresijski model uporabim v naslednjem koraku z operatorjema »Linear Regression« in »Apply Model (2)« na celotnem obsegu podatkov, da ugotovim prileganje dobljenega modela celotnemu naboru podatkov,
- operatorja »Rename« in »Generate attributes« uporabim, da generiram dodatno spremenljivko [Napaka], ki predstavlja razliko med dejansko vrednostjo prvih registracij novih vozil [Dejanski_Prve_reg_novi] in z modelom ocenjeno vrednostjo [Ocenjeni_Prvi_reg_novi],
- ostale operatorje uporabim, da v določenem vrstnem redu shranim tabelo s podatki, ki vključuje ocenjene vrednosti, dejanske vrednosti izhodne spremenljivke in njuno razliko (napako) z namenom nadaljnje obdelave v programu Microsoft Excel za izdelavo grafičnih prikazov,
- najpomembnejša izhodna rezultata predstavljata izračunan regresijski model – izhod »mod« iz operatorja »Apply Model (2)« in »Performance Vector« – izhod »exe« iz podprocesa »Validation« – kaže Slika 25,
- pri četrtletnem naboru podatkov celoten proces modeliranja zaženem dvakrat, najprej samo s celotnim naborom vhodnih spremenljivk brez slamnatih spremenljivk, nato pa na osnovi dobljenega modela še z slamnatimi spremenljivkami, vendar brez izločanja, tako da lahko primerjamo rezultate modela z in brez uporabe slamnatih spremenljivk,
- ostali izhodi so namenjeni kontroli podatkov in izračunov.

Slika 25: Izpis regresijskega modela iz programa RapidMiner

LinearRegression Napoved_Prva_reg_novi= + 0.046 * BDP + 0.162 * Uvoz - 0.003 * Reg_brezposelnost - 1.374 * Drzavna_potrosnja + 0.083 * Investicije_v_OS + 0.002 * Export_DE - 0.153 * DV_trgovina - 50.754 * Kvartal_3 + 1330.545	PerformanceVector Linear Regression root_mean_squared_error: 58.960 +/- 0.000 correlation: 0.955 squared_correlation: 0.913
--	---

Ob tem je potrebno pripomniti, da zaradi različnih velikosti obsegov posamičnih vhodnih spremenljivk dobimo pri nekaterih zelo nizke regresijske koeficiente, saj so vrednosti teh vhodnih spremenljivk zelo visoke (npr. [Export_DE]). Za potrebe napovedne sposobnosti je to sicer povsem nepomembno, saj koeficienti niso zaokroženi, kot morda na prvi pogled prikazuje Slika 25. Kljub temu sem pa zaradi tega spremenil enoto vhodnih spremenljivk [Export_DE] in [Import_DE] iz mio. EUR v mrd. EUR, tako da so koeficienti pri teh dveh spremenljivkah razvidni, tudi če je zapis omejen na tri decimalna mesta.

3.2.3 Vrednotenje dobljenih rezultatov posameznih regresijskih modelov

Za oceno rezultatov posameznih modelov sem uporabil naslednje kriterije, ki jih na kratko navajam v nadaljevanju in so povzeti po Kotu in Deshpande (2015, str. 180) in Shmueli (2011, str. 587/2267).

3.2.3.1 R^2 – determinacijski koeficient

Determinacijski koeficient (angl. *Squared Correlation* ali *Coefficient of Determination*), predstavlja razmerje med vsoto kvadratov razlike med ocenjeno vrednostjo ($\hat{y}$) in povprečno vrednostjo ($\bar{y}$) odvisne spremenljivke, ter vsoto kvadratov razlike med dejansko vrednostjo (y) in povprečno vrednostjo ($\bar{y}$) odvisne spremenljivke. Uporablja se kot glavni kriterij za oceno regresijskega modela.

Definiran je s formulo (5):

$$R^2 = \frac{\sum_1^n (\hat{y}_i - \bar{y})^2}{\sum_1^n (y_i - \bar{y})^2} \quad (5)$$

R^2 , ki je bližje 1 pomeni boljše prileganje regresijskega modela podatkom. Za primerjavo med različnimi regresijskimi modeli z različnim številom vhodnih spremenljivk se uporablja popravljen determinacijski koeficient.

Ta je definiran s formulo:

$$R_{popravljen}^2 = 1 - \frac{(1 - R^2)(n - 1)}{n - m - 1} \quad (6)$$

Pri tem (n) predstavlja število zapisov, (m) pa število vhodnih spremenljivk.

3.2.3.2 RMSE – kvadratni koren povprečne kvadratne napake

Kvadratni koren povprečne kvadratne napake (angl. *Root Mean Squared Error*) je merilo za odstopanje ocenjene vrednosti od dejanske vrednosti odvisne spremenljivke (y). Definiran je z naslednjo formulo:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (7)$$

RMSE ima enako enoto kot napaka.

3.2.3.3 Razpon napake

Razpon napake je najnižja in najvišja vrednost razlike med ocenjeno in dejansko vrednostjo odvisne spremenljivke. Model, ki ima razpon napake manjši, je boljši.

3.2.3.4 Povprečje napake

Povprečje napake je povprečje razlik med ocenjeno in dejansko vrednostjo odvisne spremenljivke.

$$\bar{e}_l = \sum_{i=1}^n y_i - \hat{y}_i \quad (8)$$

Povprečje napake bi moralo biti čim bližje 0, standardni odklon napake pa čim manjši.

3.2.3.5 Histogram porazdelitve napake e_i

Če regresijski model dobro popiše vrednost izhodne spremenljivke, potem bi morala biti porazdelitev napake e_i čim bližje normalni razporeditvi, kar je ena od osnovnih matematičnih predpostavk linearnega regresijskega modela.

3.2.3.6 MAPE – Povprečje absolutne vrednosti relativne napake

MAPE (angl. *Mean Absolute Percentage Error*) predstavlja povprečno relativno napako napovedi in je definirana z :

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \times 100\% \quad (9)$$

3.2.3.7 Matrika korelacijskih koeficientov med vhodnimi spremenljivkami

Z matriko medsebojnih korelacijskih koeficientov r_{ij}^2 bom preveril, ali so morda katere izmed vhodnih spremenljivk regresijskega modela med seboj preveč korelirane.

3.2.3.8 VIF – faktor napihnjenosti variance

VIF (angl. *Variance Inflating Factor*) je merilo za multikolinearnost in je premo sorazmeren z varianco regresijskih koeficientov modela.

$$VIF_k = \frac{1}{(1 - R_k^2)} \quad (10)$$

R_k^2 je determinacijski koeficient, če k-to vhodno spremenljivko regresiramo s preostalimi vhodnimi spremenljivkami. Višji, je, višja je varianca koeficientov linearnega regresijskega modela, in sicer zaradi visoke korelacije med dvema vhodnima spremenljivkama. Sprejemljiva vrednost VIF je do 5.

4 REZULTATI REGRESIJSKIH MODELOV ZA POSAMEZNE SEGMENTE

4.1 Lahka dostavna vozila

4.1.1 Rezultati z uporabo letnega nabora podatkov

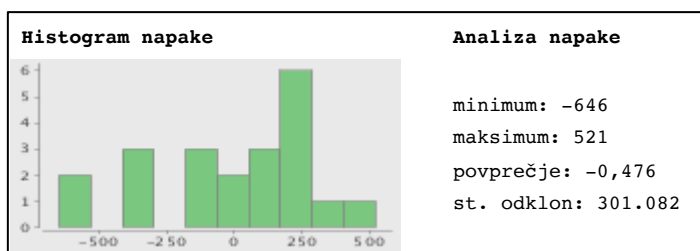
Kot prikazuje Slika 26, je proces modeliranja od 15 vhodnih spremenljivk izločil 13 spremenljivk. Obe preostali vhodni spremenljivki [DV_trgovina] in [Drzavna_potrosnja] imata regresijska koeficienta z nizko p-vrednostjo in vrednostjo t-statistike krepko nad mejno. Obe spremenljivki sta visoko korelirani z izhodno spremenljivko, saj model daje vrednost R^2 nad 0,9. Zaradi majhnega števila zapisov bi težko ocenil, da je napaka približno normalno razporejena. Iz grafičnega prikaza (Slika 27) ni opaziti, da bi imela napaka nek sistematični vzorec ali da bi bila avtokorelirana.

Slika 26: Regresijski model lahkih dostavnih vozil z uporabo letnega nabora podatkov

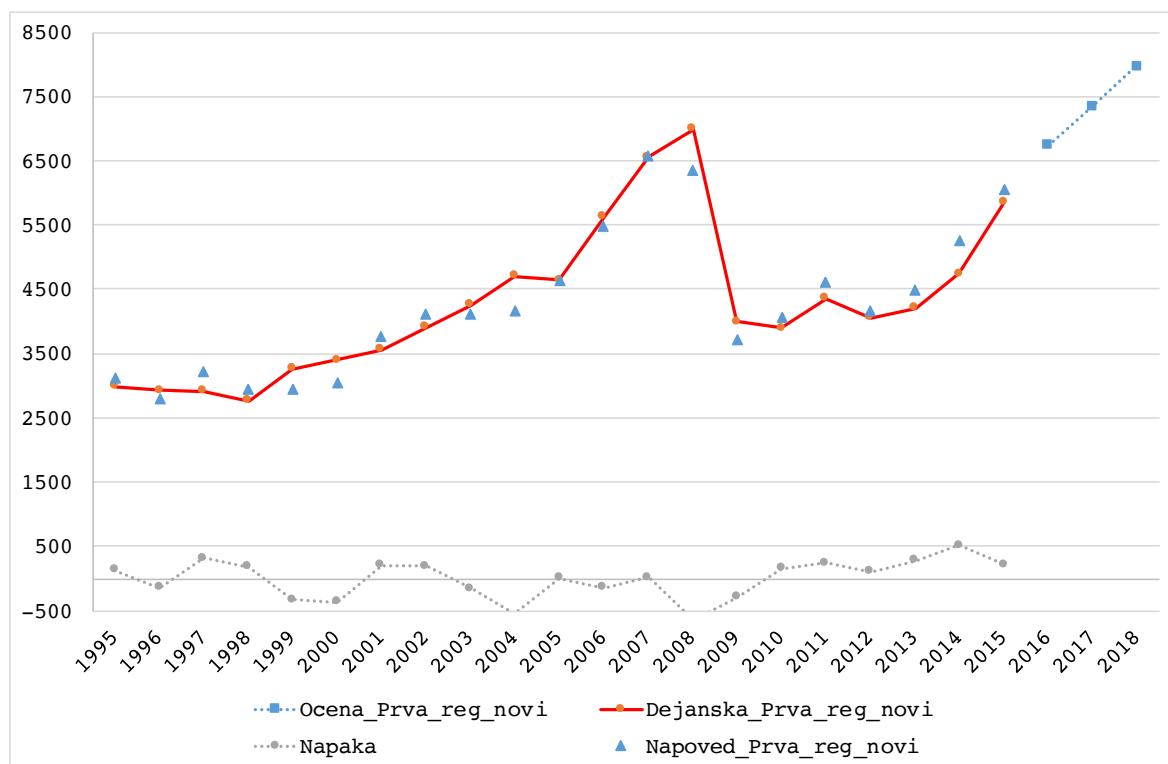
$$[\text{Napoved_Prve_Reg_novi}] = b_0 + b_1 * [\text{DV_Trgovina}] + b_2 * [\text{Drzavna_potrosnja}]$$

i	Vhodna spremenljivka x_i	Regresijski koeficijent b_i	Standardna napaka b_i	t-statistika	p-vrednost	Oznaka pomembnosti
0		3.890,253	706,121	5,509	0,000	****
1	[DV_Trgovina]	3,507	0,311	11,276	0,000	****
2	[Drzavna_potrosnja]	-2,982	0,355	-8,400	0,000	****

	Št. zapisov	R	R ²	popravljen R ²	RMSE	MAPE	stdev MAPE
Testni nabor	15	0,986	0,973		493	9,72 %	3,68 %
Celotni nabor	21	0,967	0,934	0,927	294	6,01 %	3,38 %



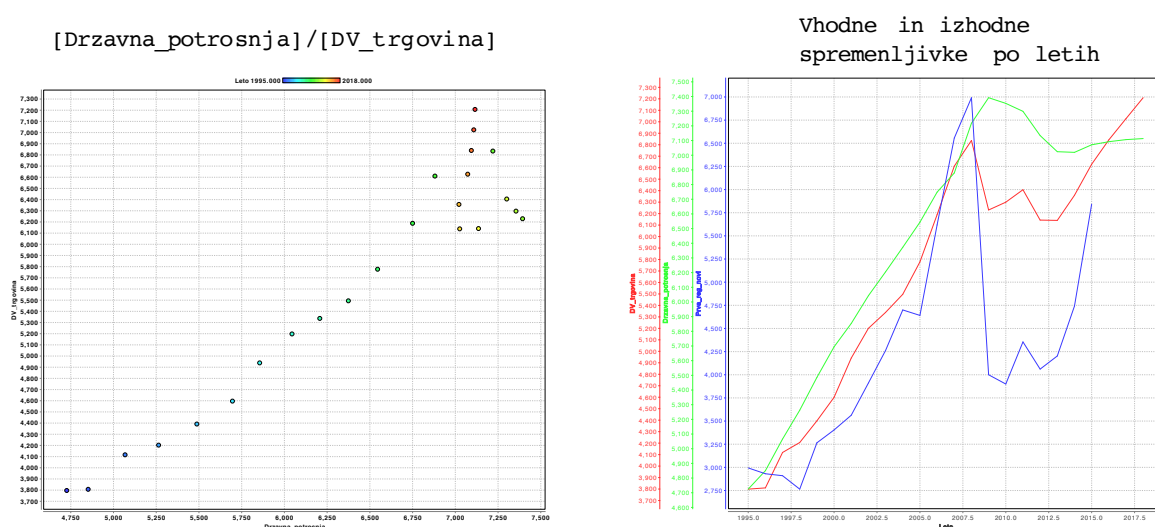
Slika 27: Grafični prikaz rezultatov regresijskega modela lahkih dostavnih vozil z uporabo letnega nabora podatkov



Sta pa med seboj visoko korelirani obe vhodni spremenljivki, kar nazorno prikazuje Slika 28. Tako vrednost medsebojnega korelacijskega koeficienta r^2 (0,95) kot vrednost VIF (19,4), nakazujeta na prisotnost pojava multikolinearnosti. To ima za posledico povečano varianco regresijskih koeficientov, kar pomeni, da dobljeni model ni robusten. Vendar zaradi visoke vrednosti R^2 ta pojav ne zmanjšuje napovedne vrednosti modela. Po drugi strani pa je zaradi majhnega nabora podatkov in dejstva, da so vhodne spremenljivke časovne vrste makroekonomskih spremenljivk, ki so v določenem časovnem obdobju med seboj močno korelirane, ta pojav do neke mere razumljiv.

Grafični prikaz gibanja obeh vhodnih in izhodne spremenljivke po letih (desni diagram Slika 28), kaže, da sta obe vhodni spremenljivki do leta 2008 rasli skupaj in sta bili zelo korelirani. V zadnjih letih pa te korelacije ni opaziti. Glede na oba predznaka regresijskih koeficientov (eden je pozitiven, drugi negativen) lahko zaključim, da dobljeni model ni stabilen in je tako omejeno primeren za napovedovanje vrednosti izhodne spremenljivke.

Slika 28: Grafični prikaz korelacij med vhodnimi in izhodnimi spremenljivkami letnega modela segmenta lahkih dostavnih vozil



	x_1	x_2	VIF_k	Ime spremenljivke
x_1	1,00		19,4	[DV_Trgovina]
x_2	0,95	1,00	19,4	[Drzavna_potrosnja]

4.1.2 Rezultati z uporabo četrtnega nabora podatkov

Proces modeliranja na naboru četrtnih podatkov izloči 12 vhodnih spremenljivk (Slika 29). Tako na osnovi četrtnih podatkov dobimo povsem druge vhodne spremenljivke v modelu kot pri letnem naboru podatkov. Glede na to, da moj namen ni modeliranje z namenom opisovanja in iskanja vzročno posledičnih relacij, temveč izključno napovedovanje vrednosti izhodne spremenljivke, je to povsem sprejemljivo.

Regresijski model daje sicer nekoliko nižji determinacijski koeficient R^2 kot pri letnem modelu, vendar so kljub temu na osnovi t-statistike in p-vrednosti vsi regresijski koeficienti pomembni. To nam nazorno kaže polje Oznaka pomembnosti, kjer so vsi regresijski koeficienti ocenjeni s štirimi zvezdicami.

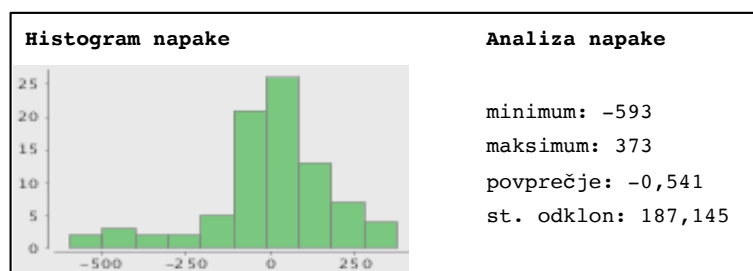
Relativna napaka je več kot dvakrat višja od letnega modela, vendar pa je porazdelitev napak bolj podobna normalni prerazdelitvi. To lahko pojasnim z dejstvom, da imam na voljo 4-krat večji nabor podatkov kot pri letnem modelu. Večji nabor podatkov na eni strani omogoči bolj naključno porazdelitev napak, na drugi strani pa s seboj prinese sezonska nihanja, ki jih z naborom vhodnih spremenljivk ni mogoče povsem pojasniti. Zato dobimo nekoliko večjo relativno napako in nižjo vrednost koeficienta R^2 .

Slika 29: Regresijski model segmenta lahkih dostavnih vozil z uporabo četrtnega nabora podatkov

$$[\text{Napoved_Prve_Reg_novi}] = b_0 + b_1 * [\text{Uvoz}] + b_2 * [\text{Drzavna_potrosnja}] + b_3 * [\text{Reg_brezposelnost}]$$

i	Vhodna spremenljivka x_i	Regresijski koeficient b_i	Standardna napaka b_i	t-statistika	p-vrednost	Oznaka pomembnosti
0		2.621,230	358,771	7,306	0,000	****
1	[Uvoz]	0,357	0,046	7,812	0,000	****
2	[Drzavna_potrosnja]	-1,532	0,305	-5,405	0,000	****
3	[Reg_brezposelnost]	-0,007	0,001	-5,405	0,000	****

	Št. zapisov	R	R^2	popravljen R^2	RMSE	MAPE	stdev MAPE
Testni nabor	60	0,941	0,885		112	9,05 %	7,62 %
Celotni nabor	85	0,846	0,715	0,704	186	12,19 %	10,46 %



Delno lahko ta učinek dodatnega nihanja v modelu na osnovi četrtnih podatkov omilim z uporabo treh slamnatih (sezonskih) spremenljivk kot prikazuje Slika 30.

Slika 30: Regresijski model segmenta lahkih dostavnih vozil z uporabo četrletnega nabora podatkov in slamnatimi spremenljivkami

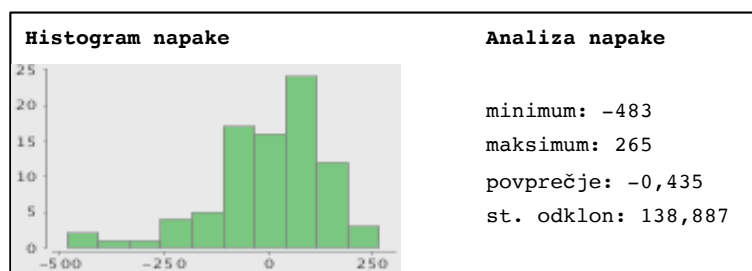
$$[\text{Napoved_Prve_Reg_novi}] = b_0 + b_1 * [\text{Uvoz}] + b_2 * [\text{Drzavna_potrosnja}] + b_3 * [\text{Reg_brezposelnost}] + b_4 * [\text{Kvartal_1}] + b_5 * [\text{Kvartal_2}] + b_6 * [\text{Kvartal_3}]$$

i	Vhodna spremenljivka x_i	Regresijski koeficient b_i	Standardna napaka b_i	t-statistika	p-vrednost	Oznaka pomembnosti
0		3.263,088	283,372	11,515	0,000	****
1	[Uvoz]	0,402	0,035	11,479	0,000	****
2	[Drzavna_potrosnja]	-1,907	0,236	-8,092	0,000	****
3	[Reg_brezposelnost]	-0,008	0,001	-7,726	0,000	****
4	[Kvartal_1]	-212,147	44,625	-4,754	0,000	****
5	[Kvartal_2]	-197,060	44,671	-4,411	0,000	****
6	[Kvartal_3]	-360,744	45,462	11,515	0,000	****

Kot prikazuje Slika 31 z uvedbo sezonskih slamnatih spremenljivk dobim še nekoliko bolj enakomerno porazdelitev napake, poveča se R^2 (iz 0,715 na 0,843), relativna napaka se zmanjša za približno dve odstotni točki, RMSE, standardni odklon napake, se zmanjša za približno četrtno.

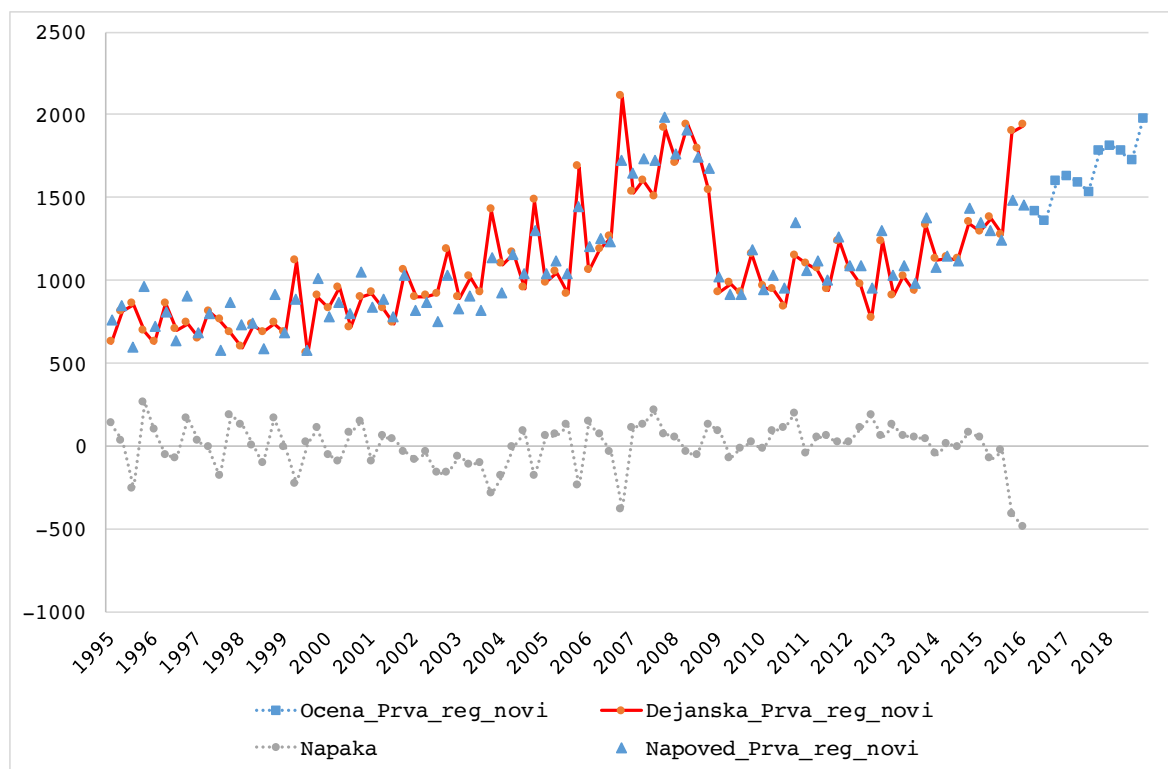
Slika 31: Kazalniki regresijskega modela segmenta lahkih dostavnih vozil z uporabo četrletnega nabora podatkov in slamnatimi spremenljivkami

	Št. zapisov	R	R^2	popravljen R^2	RMSE	MAPE	stdev MAPE
Testni nabor	60	0,969	0,940		102	8,93 %	6,67 %
Celotni nabor	85	0,918	0,843	0,831	138	10,05 %	7,84 %



Na osnovi teh rezultatov lahko vsekakor potrdim izkušnje drugih avtorjev, da pravilna uporaba slamnatih spremenljivk izboljša napovedno moč in rezultate modela. Slika 32 prikazuje rezultate napovedi še v grafični obliki.

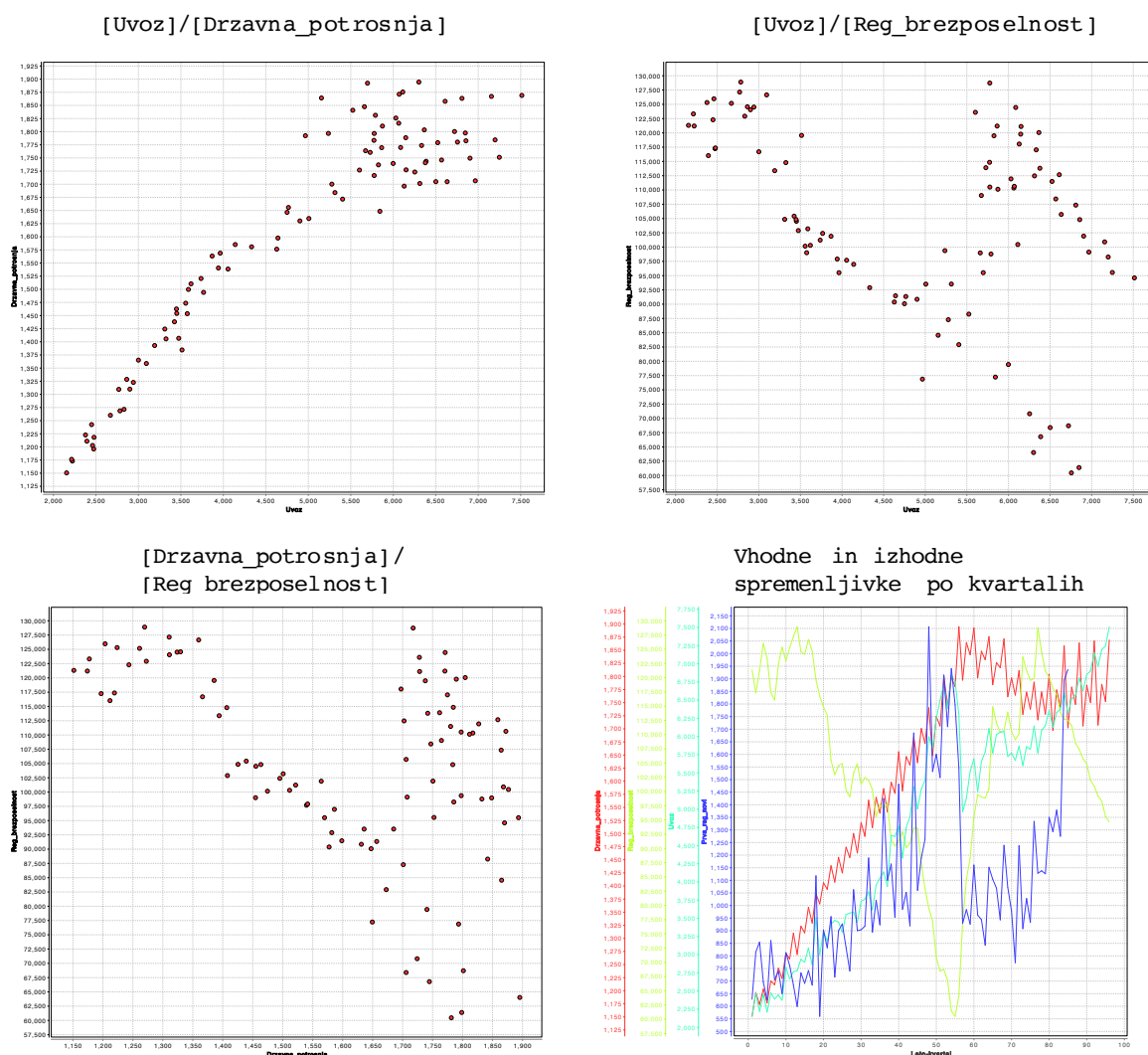
Slika 32: Grafični prikaz rezultatov regresijskega modela segmenta lahkih dostavnih vozil z uporabo četrletnega nabora podatkov in slamnatih spremenljivk



Tako kot pri letnem modelu tudi pri kvartalnem modelu ugotavljam, da sta dve od treh vhodnih spremenljivk močno korelirani, kar nakazuje na multikolinearnost. Slika 33 prikazuje odvisnost vhodnih spremenljivk med seboj in časovno vrsto vhodnih in izhodne spremenljivke. Tako imata spremenljivki [Drzavna_potrosnja] in [Uvoz] visok korelacijski koeficient ($r^2=0,9$). Vrednost VIF za obe spremenljivki je prav tako zelo visoka (nad 10), vendar pa vseeno polovico nižja kot pri letnem modelu. Podobno kot pri letnem modelu imamo tudi pri četrtnem modelu opravka z multikolinearnostjo, ki pa je glede na vrednost VIF polovico manjša kot pri letnem modelu.

Zaključim lahko, da na opazovanem segmentu lahkih dostavnih vozil četrtni model daje boljše rezultate kot letni, kljub temu da zaradi dodatnih sezonskih nihanj, ki jih lahko le delno omilimo z uporabo slamnatih spremenljivk, v napoved vnaša nekoliko povišano napako.

Slika 33: Grafični prikaz korelacij med vhodnimi spremenljivkami četrtnega modela segmenta lahkih dostavnih vozil



	x_1	x_2	x_3	VIF_k	Ime spremenljivke
x_1	1,00			10,1	[Uvoz]
x_2	0,22	1,00		1,3	[Reg_brezposelnost]
x_3	0,90	0,23	1,00	10,2	[Drzavna_potrosnja]

4.1.3 Rezultati z uporabo četrtnega nabora podatkov na letni ravni

V praksi pri ocenjevanju obsega trga zadostuje ocena na letni ravni. Nenazadnje so tudi napovedi makroekonomskih spremenljivk, ki sem jih uporabil za vhodne spremenljivke, na voljo le na letni ravni. Zato sem se odločil, da model na osnovi četrtnega nabora podatkov prevedem na letnega z enostavnim seštevanjem štirih četrtnih napovedi za vsako leto. Ta združeni model ohranja vse dobre lastnosti četrtnega modela, hkrati pa lahko, v primerjavi z letnim modelom, zaradi medsebojnega izenačevanja napak znotraj posameznega leta, opazimo manjša nihanja napak. Kot je razvidno iz rezultatov, ki jih

prikazuje Slika 34, se pričakovano s seštevanjem četrtnih napovedi relativna napaka MAPE za več kot polovico zniža, hkrati pa se poveča determinacijski koeficient R^2 , ki je celo višji kot pri letnem modelu.

Slika 34: Model segmenta lahkih dostavnih vozil z uporabo četrtnih napovedi na letni ravni

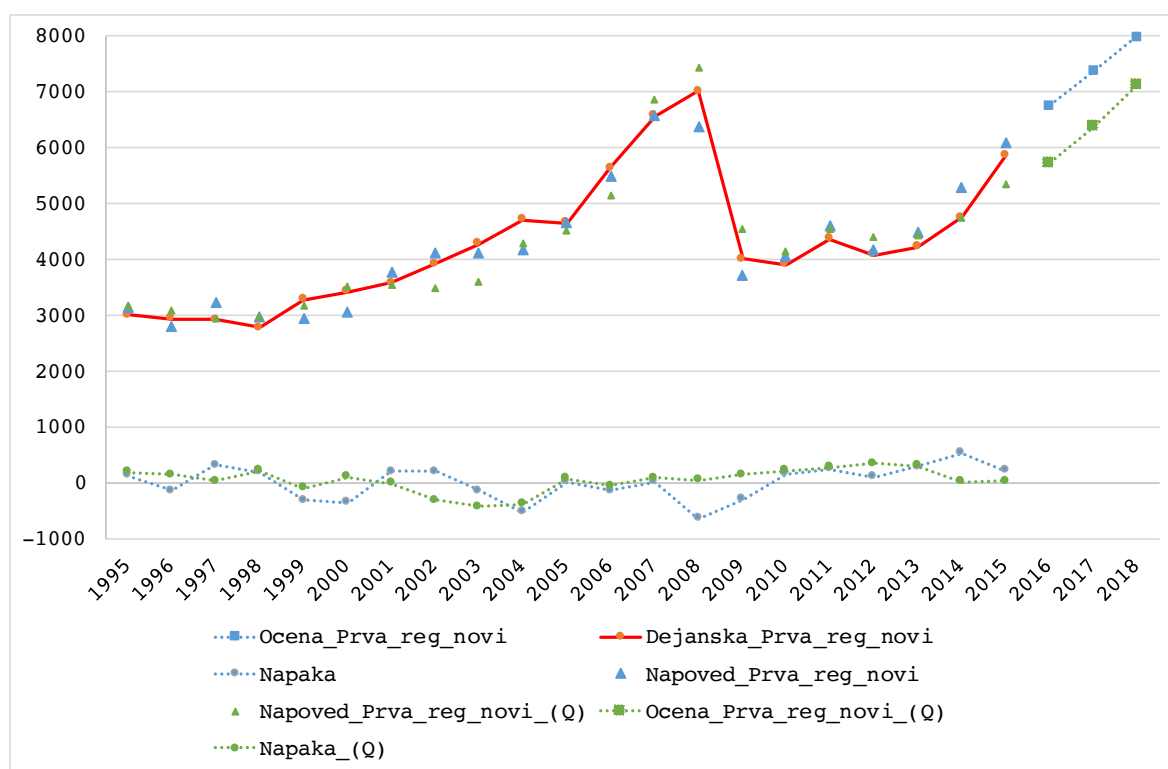
$$[\text{Napoved_Prve_Reg_novi_Q}] = [\text{Napoved_Prve_Reg_novi_Q1}] + [\text{Napoved_Prve_Reg_novi_Q2}] + [\text{Napoved_Prve_Reg_novi_Q3}] + [\text{Napoved_Prve_Reg_novi_Q4}]$$

	Št. zapisov	R	R^2	popravljen R^2	RMSE	MAPE	stdev MAPE
Celotni nabor	85	0,985	0,969	0,956	208	4,20 %	3,31 %

Analiza napake

minimum: -430
maksimum: 338
povprečje: 40,76
st. odklon: 208,72

Slika 35: Grafični prikaz rezultatov modela segmenta lahkih dostavnih vozil z uporabo četrtnega nabora na letni ravni v primerjavi z letnim naborom podatkov



Grafični prikaz rezultatov modela v primerjavi z letnim modelom prikazuje Slika 35. Opazimo lahko, da napaka (Napaka_Q) pri tem modelu bistveno manj niha kot pri letnem modelu. Nekoliko moti edino dejstvo, da se povprečje napak nekoliko oddalji od nič, kar kaže na prisotnost določene sistemske napake v modelu.

4.1.4 Povzetek ugotovitev modeliranja segmenta lahkih dostavnih vozil

Pri modeliranju segmenta lahkih dostavnih vozil lahko strnemo sledeče ugotovitve:

- obseg segmenta lahkih dostavnih vozil je mogoče dovolj dobro napovedati tako na letni kot na četrtni ravni,
- pri ocenjevanju obsega segmenta se moramo zavedati prisotnosti multikolinearnosti, ki je pogojena z naborom podatkov (letni) in vrsto vhodnih spremenljivk (časovne vrste),
- na letni ravni pa je najbolj primerno uporabiti seštevek napovedi četrtnih ocen, saj da ta model boljše rezultate kot letni in je tudi zaradi obsega nabora podatkov bolj zanesljiv.

Primerjavo glavnih kazalnikov za regresijskih modelov za segment lahkih dostavnih vozil prikazuje Tabela 4.

Tabela 4: Primerjava rezultatov regresijskih modelov za segment lahkih dostavnih vozil*

	Četrtni model	Četrtni sezonski model	Četrtni sezonski model letno	Letni model
Št. vhodnih spremenljivk	3	6	6 (24)	2
R^2 (na testnem naboru podatkov)	0,885	0,940	-	0,972
R^2 (na celotnem naboru podatkov)	0,715	0,843	0,969	0,934
popravljen R^2 (na celotnem naboru)	0,704	0,831	0,956	0,927
RMSE (na celotnem naboru podatkov)	186	138	207	294
Napaka min/max	-593/373	-483/265	-430/338	-646/521
Napaka povprečje	-0,5	-0,4	40,8	-0,5
Napaka standardni odklon	187	139	209	301
MAPE – relativna napaka +/- standardni odklon MAPE	12 % +/-10,5 %	10 % +/-7,8 %	4 % +/-3,3 %	6 % +/-3,4 %

Legenda: * Primerjava je izvedena na osnovi kazalnikov predstavljenih v poglavju 3.2.3

4.2 Težka tovorna vozila

4.2.1 Rezultati z uporabo letnega nabora podatkov

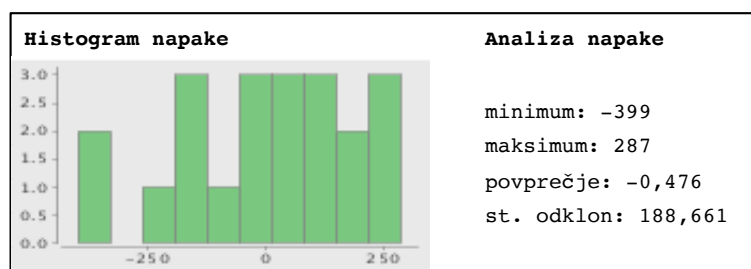
Tudi na segmentu težkih tovornih vozil proces modeliranja izloči kar 12 vhodnih spremenljivk. Kot prikazuje Slika 36 je R^2 zelo visok, koeficienti so vsi visoko ocenjeni. Napake so sicer lepo razporejene, njihova časovna vrsta daje vtis naključnosti, vendar pa zaradi majhnega obsega podatkov iz histograma težko ocenimo, ali so približno normalno razporejene.

Slika 36: Regresijski model segmenta težkih tovornih vozil z uporabo letnega nabora podatkov

$$[\text{Napoved_Prve_Reg_novi}] = b_0 + b_1 * [\text{Investicije_OS}] + b_2 * [\text{Zasebna_potrosnja}] + b_3 * [\text{Izvoz}]$$

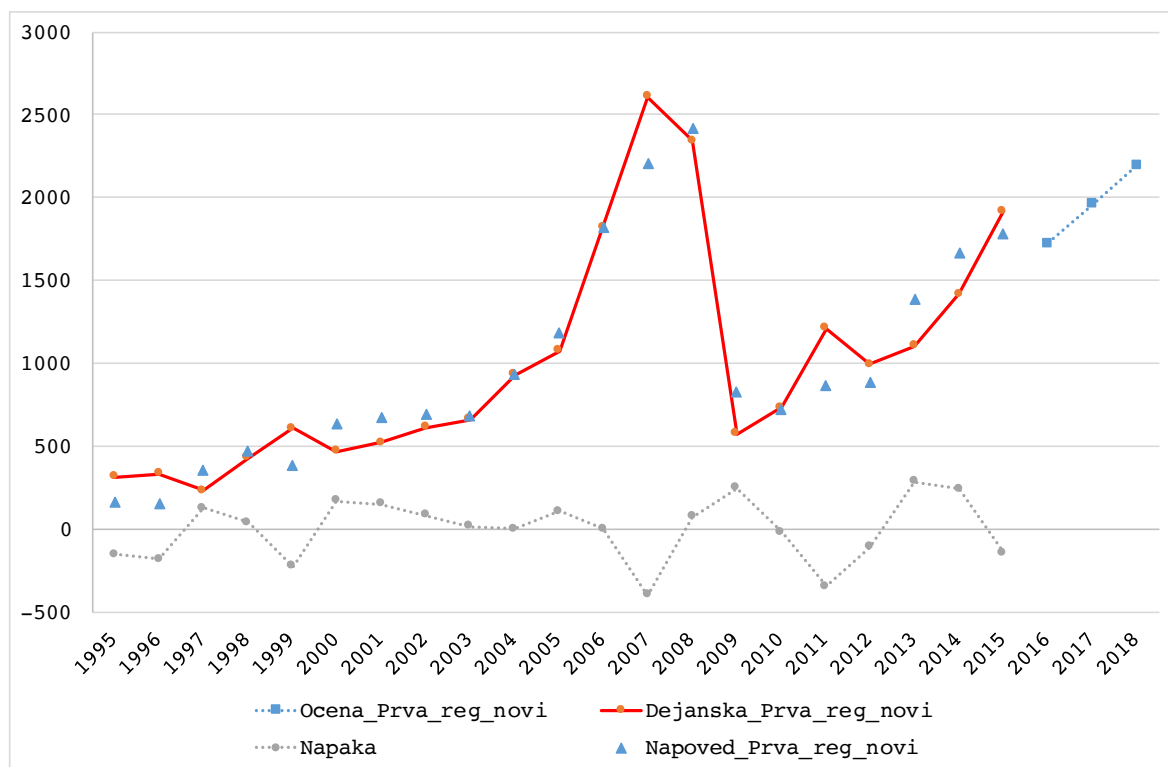
i	Vhodna spremenljivka x_i	Regresijski koeficient b_i	Standardna napaka b_i	t-statistika	p-vrednost	Oznaka pomembnosti
0		2.649,675	707,998	3,742	0,002	***
1	[Investicije_v_OS]	0,320	0,038	8,421	0,000	****
2	[Zasebna_potrosnja]	-0,418	0,068	-6,168	0,000	****
3	[Izvoz]	0,169	0,021	8,251	0,000	****

	Št. zapisov	R	R^2	popravljen R^2	RMSE	MAPE	stdev MAPE
Testni nabor	15	0,977	0,955		192	37,48 %	44,83 %
Celotni nabor	21	0,960	0,922	0,908	184	21,28 %	17,63 %



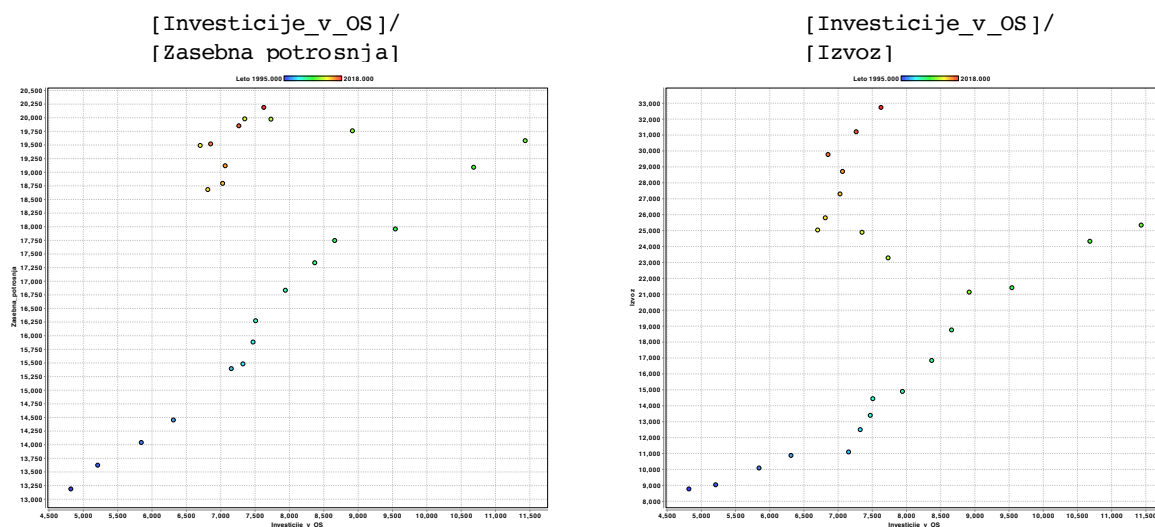
Kljub zelo visoki vrednosti R^2 moti dejstvo, da je relativna napaka MAPE dvakrat večja kot pri letnem modelu lahkih dostavnih vozil. Bistveno višji je tudi standardni odklon relativne napake MAPE, kar nakazuje na veliko odstopanje ocenjenih od dejanskih vrednosti izhodne spremenljivke. Grafični prikaz rezultatov regresijskega modela težkih tovornih vozil z uporabo letnega nabora podatkov prikazuje Slika 37.

Slika 37: Grafični prikaz rezultatov regresijskega modela težkih tovornih vozil z uporabo letnega nabora podatkov

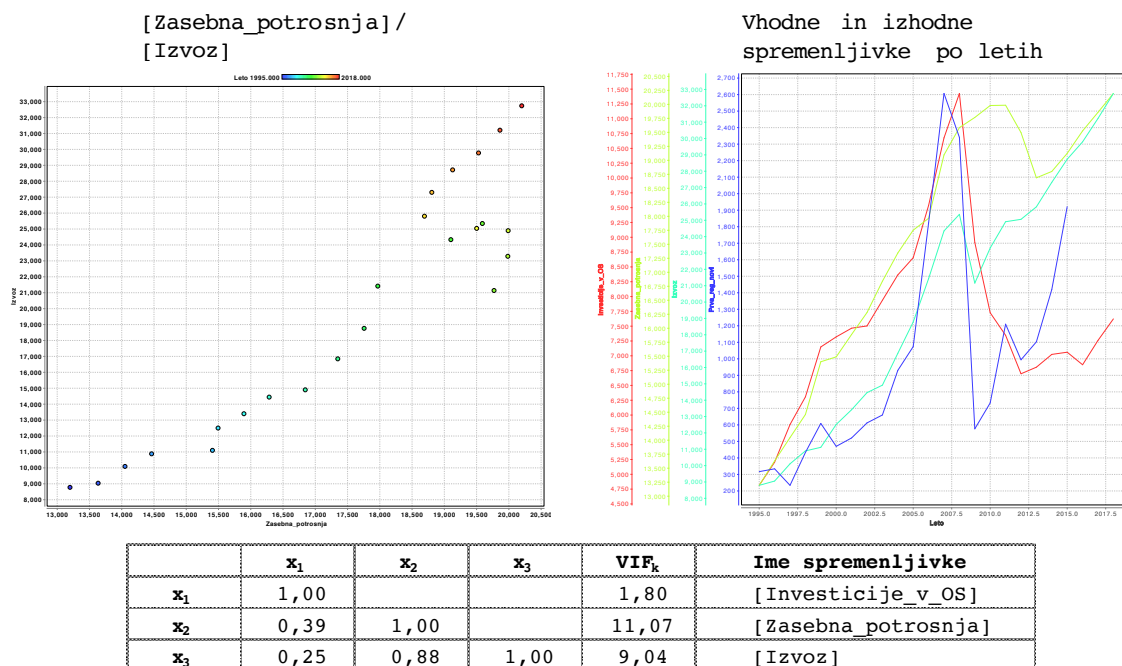


Podobno kot vsi dosednji modeli tudi ta kaže na prisotnost multikolinearnosti, čeprav so vrednosti VIF za posamične koeficiente pol nižje kot pri letnem modelu segmenta lahkih dostavnih vozil (Slika 38 in Slika 39).

Slika 38: Grafični prikaz korelacij med vhodnimi spremenljivkami letnega modela segmenta težkih tovornih vozil



Slika 39: Grafični prikaz korelacij med vhodnimi spremenljivkami letnega modela segmenta težkih tovornih vozil



4.2.2 Rezultati z uporabo četrtnetnega nabora podatkov

Rezultati modela, ki ga prikazujeta Slika 40 in Slika 41, kažejo da z uporabo četrtnetnega nabora dobimo boljšo porazdelitev napake (porazdelitev je bližje normalni), vendar pa so vsi ostali kazalniki nekoliko slabši, vključno z relativno napako MAPE, ki se pričakovano še dodatno poveča. Zanimivo je, da so uporabljene vhodne spremenljivke v modelu enake kot pri četrtnetnem modelu segmenta lahkih dostavnih vozil.

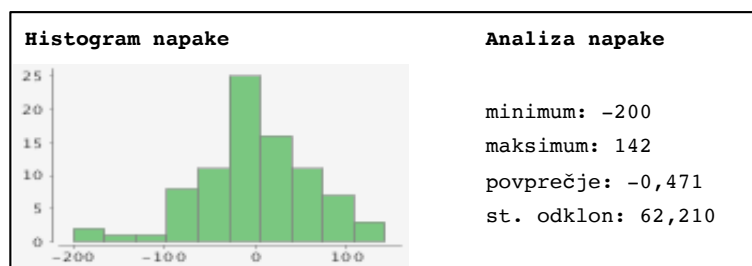
Slika 40: Regresijski model segmenta težkih tovornih vozil z uporabo četrtnetnega nabora podatkov

$$[\text{Napoved_Prve_Reg_novi}] = b_0 + b_1 * [\text{Uvoz}] + b_2 * [\text{Reg_brezposelnost}] + b_3 * [\text{Drzavna_potrosnja}]$$

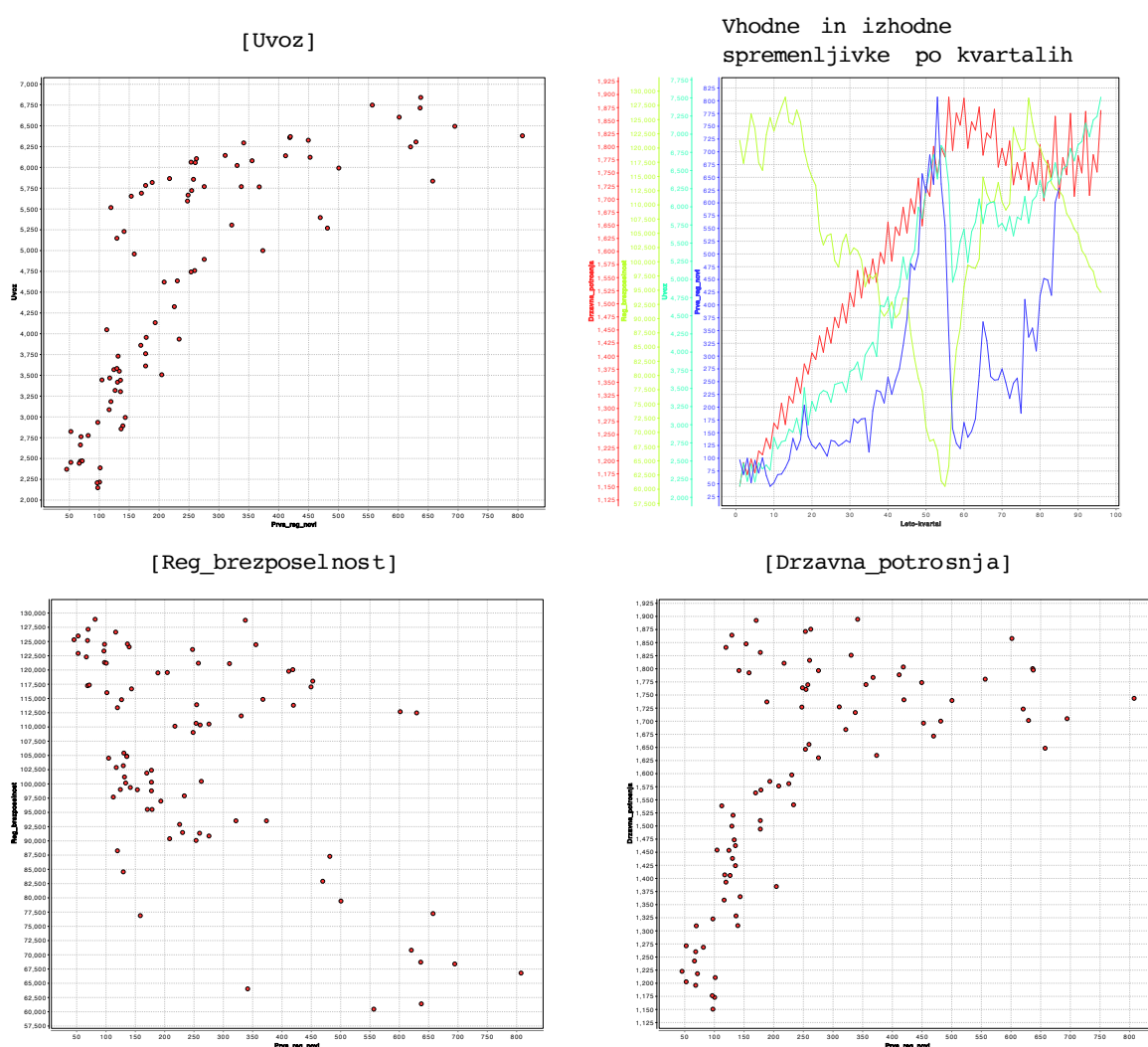
i	Vhodna spremenljivka x_i	Regresijski koeficijent b_i	Standardna napaka b_i	t-statistika	p-vrednost	Oznaka pomembnosti
0		1.347,210	119,258	11,297	0,000	****
1	[Uvoz]	0,256	0,015	16,880	0,000	****
2	[Reg_brezposelnost]	-0,003	0,000	-6,700	0,000	****
3	[Drzavna_potrosnja]	-1,238	0,101	-12,215	0,000	****

	Št. zapisov	R	R^2	popravljen R^2	RMSE	MAPE	stdev MAPE
Testni nabor	60	0,956	0,914		52	24,66 %	28,31 %
Celotni nabor	85	0,937	0,877	0,872	62	24,78 %	26,68 %

Slika 41: Analiza napake regresijskega modela segmenta težkih tovornih vozil z uporabo četrtletnega nabora podatkov



Slika 42: Grafični prikaz korelacij med vhodnimi in izhodnimi spremenljivkami četrtletnega modela segmenta težkih tovornih vozil



	x_1	x_2	x_3	VIF_k	Ime spremenljivke
x_1	1,00			10,1	[Uvoz]
x_2	0,22	1,00		1,3	[Reg_brezposelnost]
x_3	0,90	0,23	1,00	10,2	[Drzavna_potrosnja]

Dodatna grafična analiza odvisnosti vhodnih spremenljivk, uporabljenih v modelu ([Uvoz], [Drzavna_potrosnja], [Reg_brezposelnost]), od izhodne spremenljivke [Prve_reg_novi] kaže, da so dejansko vse vhodne spremenljivke korelirane z izhodno spremenljivko (Slika 42), vendar pa žal dve od njih tudi med seboj. S tem se tudi v segmentu težkih tovornih vozil pojavi enak problem, saj sta kot prikazuje Slika 42, vhodni spremenljivki [Uvoz] in [Drzavna_potrosnja] močno korelirani, kar nakazuje določeno stopnjo nestabilnosti modela. Vzroki za to so enaki kot v primeru modela za segment lahkih dostavnih vozil.

Bistveno večji relativni napaki MAPE v primerjavi s segmentom lahkih dostavnih vozil lahko pripišemo dejstvu, da je prodaja lahkih dostavnih vozil bistveno bolj razpršena (veliko manjših kupcev), kot je to pri prodaji težkih tovornih vozil. Tako zaradi bistveno manjšega obsega segmenta večji nakupi večjih kupcev, ki jih po navadi izvedejo znotraj enega kvartala, bistveno bolj vplivajo na nihanje števila prvič registriranih novih vozil med posamičnimi kvartali, kot je to pri lahkih dostavnih vozilih.

S to posebnostjo segmenta težkih tovornih vozil lahko pojasnimo tudi manjše izboljšanje modela z uporabo sezonskih slavnatih spremenljivk. Kot prikazujeta Slika 43 in Slika 44, se sicer vsi kazalniki nekoliko izboljšajo, vendar pa manj kot pri segmentu lahkih dostavnih vozil.

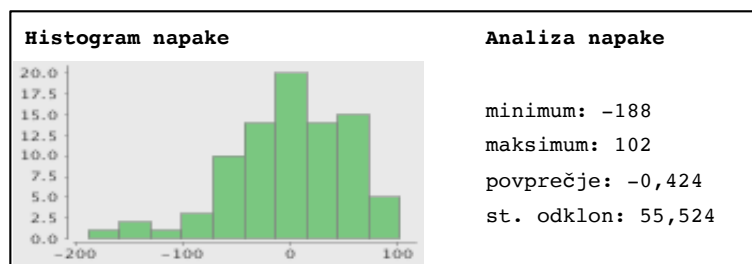
Slika 43: Regresijski model segmenta težkih tovornih vozil z uporabo četrletnega nabora podatkov in slavnatimi spremenljivkami

$$[\text{Napoved_Prve_Reg_novi}] = b_0 + b_1 * [\text{Uvoz}] + b_2 * [\text{Reg_brezposelnost}] + b_3 * [\text{Drzavna_potrosnja}] + b_4 * [\text{Kvartal_1}] + b_5 * [\text{Kvartal_2}] + b_6 * [\text{Kvartal_3}]$$

i	Vhodna spremenljivka x_i	Regresijski koeficijent b_i	Standardna napaka b_i	t-statistika	p-vrednost	Oznaka pomembnosti
0		1.421,735	113,279	12,551	0,000	****
1	[Uvoz]	0,262	0,014	18,734	0,000	****
2	[Reg_brezposelnost]	-0,003	0,000	-7,740	0,000	****
3	[Drzavna_potrosnja]	-1,283	0,094	-13,624	0,000	****
4	[Kvartal_1]	18,052	17,839	1,012	0,315	
5	[Kvartal_2]	-15,148	17,857	-0,848	0,399	
6	[Kvartal_3]	-57,969	18,173	-3,190	0,002	***

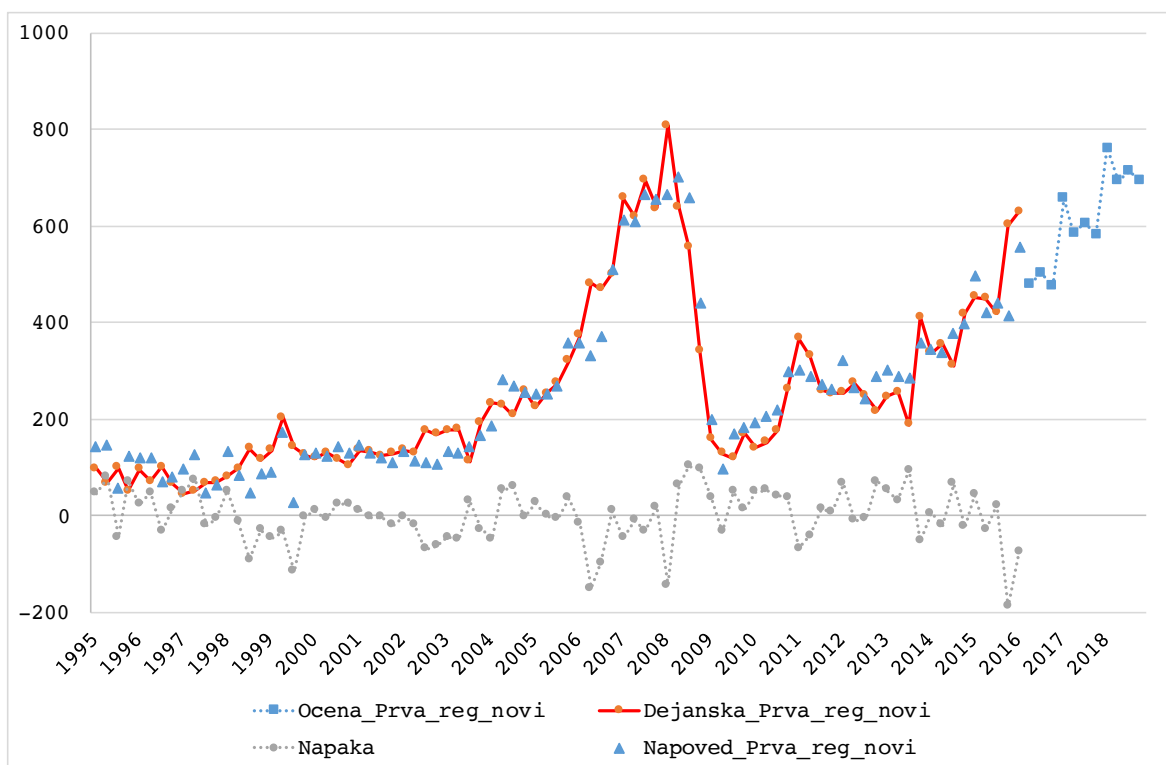
	Št. zapisov	R	R ²	popravljen R ²	RMSE	MAPE	stdev MAPE
Testni nabor	60	0,960	0,923		56	31,17 %	32,57 %
Celotni nabor	85	0,950	0,902	0,894	55	24,81 %	28,40 %

Slika 44: Analiza napake regresijskega modela segmenta težkih tovornih vozil z uporabo četrletnega nabora podatkov in slamnatih spremenljivk



Tako se R^2 iz 0,877 poveča na 0,902, RMSE se zmanjša s 63 na 55, medtem ko relativna napaka, kot merilo natančnosti modela, vztraja na visokih 25 %. Tudi pogled na regresijske koeficiente nam pokaže razliko. Če so bili regresijski koeficienti slamnatih spremenljivk pri segmentu lahkih dostavnih vozil zelo pomembni, je slika pri segmentu težkih tovornih vozil povsem drugačna. P-vrednosti in t-statistika nam kažeta, da ta koeficienta nimata bistvenega vpliva na regresijski model. Tudi če primerjamo oba histograma porazdelitve napake, lahko opazimo, da je pri modelu brez slamnatih spremenljivk, razporeditev bolj podobna normalni kot v modelu z uporabo slamnatih spremenljivk. Grafični prikaz rezultatov regresijskega modela z uporabo slamnatih spremenljivk prikazuje Slika 45.

Slika 45: Grafični prikaz rezultatov regresijskega modela segmenta težkih tovornih vozil z uporabo četrletnega nabora podatkov in slamnatih spremenljivk



Ne glede na to sem se odločil, da za uporabo četrtnega modela na letni ravni izberem model s slamnatimi spremenljivkami, saj razen pri porazdelitvi napak in standardnem odklonu relativne napake MAPE po vseh drugih kriterijih (R^2 , RMSE, standardni odklon napake) daje boljše rezultate.

4.2.3 Rezultati z uporabo četrtnega nabora podatkov na letni ravni

Podobno kot pri segmentu lahkih dostavnih vozil s tem, ko kvartalne napovedi preračunamo na letno raven, precej zmanjšamo napake. Kot prikazuje Slika 46 se še dodatno poveča determinacijski koeficient R^2 , pričakovano pa se tudi zmanjša relativna napaka MAPE (iz 25 % na 15 %).

Slika 46: Model segmenta težkih tovornih vozil z uporabo četrtnih napovedi na letni ravni

$$[\text{Napoved_Prve_Reg_novi_Q}] = [\text{Napoved_Prve_Reg_novi_Q1}] + [\text{Napoved_Prve_Reg_novi_Q2}] + [\text{Napoved_Prve_Reg_novi_Q3}] + [\text{Napoved_Prve_Reg_novi_Q4}]$$

	Št. zapisov	R	R^2	popravljen R^2	RMSE	MAPE	stdev MAPE
Celotni nabor	85	0,984	0,968	0,954	118	15,31 %	14,05 %

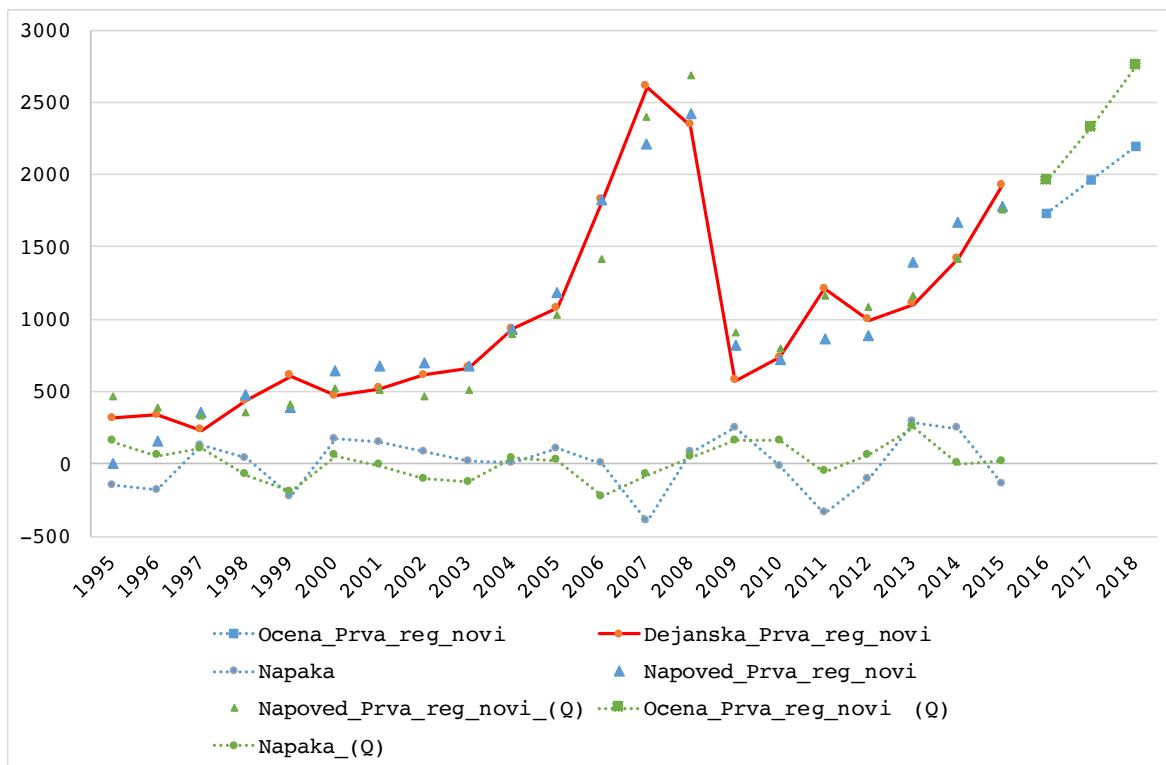
Analiza napake

minimum: -226
maksimum: 254
povprečje: 11,38
st. odklon: 120,71

Tudi v tem segmentu z uporabo četrtnih podatkov na letni ravni dobimo v vseh pogledih boljši napovedni model, kot z uporabo letnih podatkov. Iz diagrama, ki ga prikazuje Slika 47, je lepo razvidno da se nihanje napake bistveno zmanjša, kar je opaziti tudi pri modelu segmenta lahkih dostavnih vozil. Podobno pa nekoliko moti dejstvo, da se povprečna napaka nekoliko pomakne od 0 (iz 0,5 na 11,4), kar kaže na določeno mero sistematične napake.

Zanimivo je tudi dejstvo, da z uporabo četrtnega nabora podatkov na letni ravni, v nasprotju z rezultati v segmentu lahkih dostavnih vozil, pri segmentu težkih tovornih vozil letni model napoveduje nižji obseg trga kot na letni nivo preračunani četrtni model. Glede na to, da pri obeh segmentih v četrtnem modelu nastopajo iste vhodne spremenljivke, lahko to dejstvo delno pojasnimo z različnim naborom vhodnih spremenljivk pri letnem modelu.

Slika 47: Grafični prikaz regresijskega modela segmenta težkih tovornih vozil z uporabo četrtletnega nabora na letni ravni v primerjavi z letnim naborom podatkov



4.2.4 Povzetek ugotovitev modeliranja segmenta težkih tovornih vozil

Pri pregledu rezultatov modelov segmenta težkih tovornih vozil ne morem mimo dejstva, da je relativna napaka MAPE več kot 3-krat višja, kot je pri lahkih dostavnih vozilih. To je mogoče pojasniti s tem, da je segment lahkih dostavnih vozil bistveno bolj odvisen od gospodarskega okolja v Sloveniji, kot je segment težkih tovornih vozil. Ta je bistveno bolj odvisen od gospodarskih razmer v EU oziroma razmer na naših glavnih izvoznih trgih (Nemčija, Avstrija, Italija). Poleg tega je ta segment bistveno bolj reguliran. Teh razmer z izbranim naborom vhodnih spremenljivk očitno ni mogoče dovolj dobro popisati.

Na podlagi primerjave rezultatov regresijskih modelov za segment težkih tovornih vozil, ki jih prikazuje Tabela 5, lahko pridemo do sledečih ugotovitev:

- tako četrtletni kot letni model imata visoko napovedno moč, vendar moramo pri tem upoštevati omejitve, ki jih prinaša medsebojna koreliranost vhodnih spremenljivk,
- četrtletni model je tako s stališča obsega podatkov, relativne napake, kot napovedne moči, boljši od letnega modela,
- zaradi visoke relativne napake MAPE je priporočljiva uporaba četrtletnega modela.

Tabela 5: Primerjava rezultatov regresijskih modelov za segment težkih tovornih vozil*

	Četrletni model	Četrletni sezonski model	Četrletni sezonski model letno	Letni model
Št. vhodnih spremenljivk	3	6	6 (24)	3
R ² (na testnem naboru podatkov)	0,914	0,923	-	0,955
R ² (na celotnem naboru podatkov)	0,877	0,902	0,968	0,922
popravljen R ² (na celotnem naboru)	0,872	0,894	0,954	0,908
RMSE (na celotnem naboru podatkov)	63	55	118	184
Napaka min/max	-200/142	-188/102	-226/254	-399/287
Napaka povprečje	-0,5	-0,4	11,4	-0,5
Napaka standardni odklon	62	56	121	189
MAPE – relativna napaka +/- standardni odklon MAPE	25 % +/-26,7 %	25 % +/-28,4 %	15 % +/-14,1 %	21 % +/-17,6 %

Legenda: * Primerjava je izvedena na osnovi kazalnikov predstavljenih v poglavju 3.2.3

4.3 Avtobusi

4.3.1 Rezultati z uporabo letnega nabora podatkov

Rezultat je glede na ostale segmente nekoliko presenetljiv, saj, kot prikazuje Slika 48, na osnovi letnega nabora podatkov proces modeliranja izloči samo 7 vhodnih spremenljivk. Vrednost determinacijskega koeficienta R^2 je v primerjavi s preostalima segmentoma nizka, kar nam pojasni pogled na pomembnost regresijskih koeficientov.

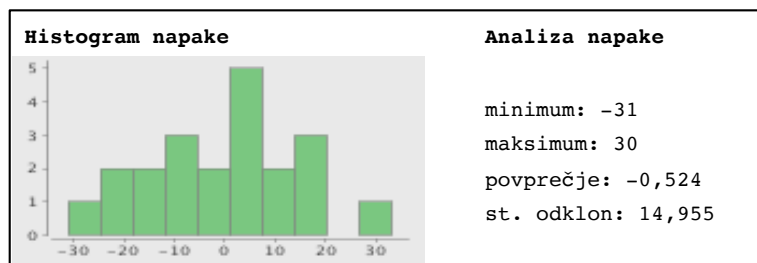
Ti na podlagi p-vrednosti in t-statistike slabo popisujejo odvisnost od med vhodnimi in izhodno spremenljivko. Napaka je glede na majhno število opazovanj relativno enakomerno porazdeljena, tudi relativna napaka MAPE je v primerjavi z ostalima dvema segmentoma majhna. Grafični prikaz rezultatov prikazuje Slika 49. Iz analize grafičnega prikaza rezultatov in gibanja dejanskega obsega trga v preteklosti, lahko vidimo, da ta segment trga gospodarskih vozil nima nekega izrazitega trenda ali cikličnega gibanja, kot je bilo relativno izrazito prisotno v segmentu lahkih dostavnih vozil in težkih tovornih vozil. To nakazuje, da morda makroekonomske spremenljivke niso najbolj primerne vhodne spremenljivke za napovedovanje obsega trga avtobusov.

Slika 48: Regresijski model segmenta avtobusov z uporabo letnega nabora podatkov

$$[\text{Napoved_Prve_Reg_novi}] = b_0 + b_1 * [\text{Export_DE}] + b_2 * [\text{DV_predelovalne}] + b_3 * [\text{BDP}] + b_4 * [\text{Uvoz}] + b_5 * [\text{Obrestna_mera}] + b_6 * [\text{Reg_brezposelnost}] + b_7 * [\text{Zasebna_potrosnja}] + b_8 * [\text{Import_DE}]$$

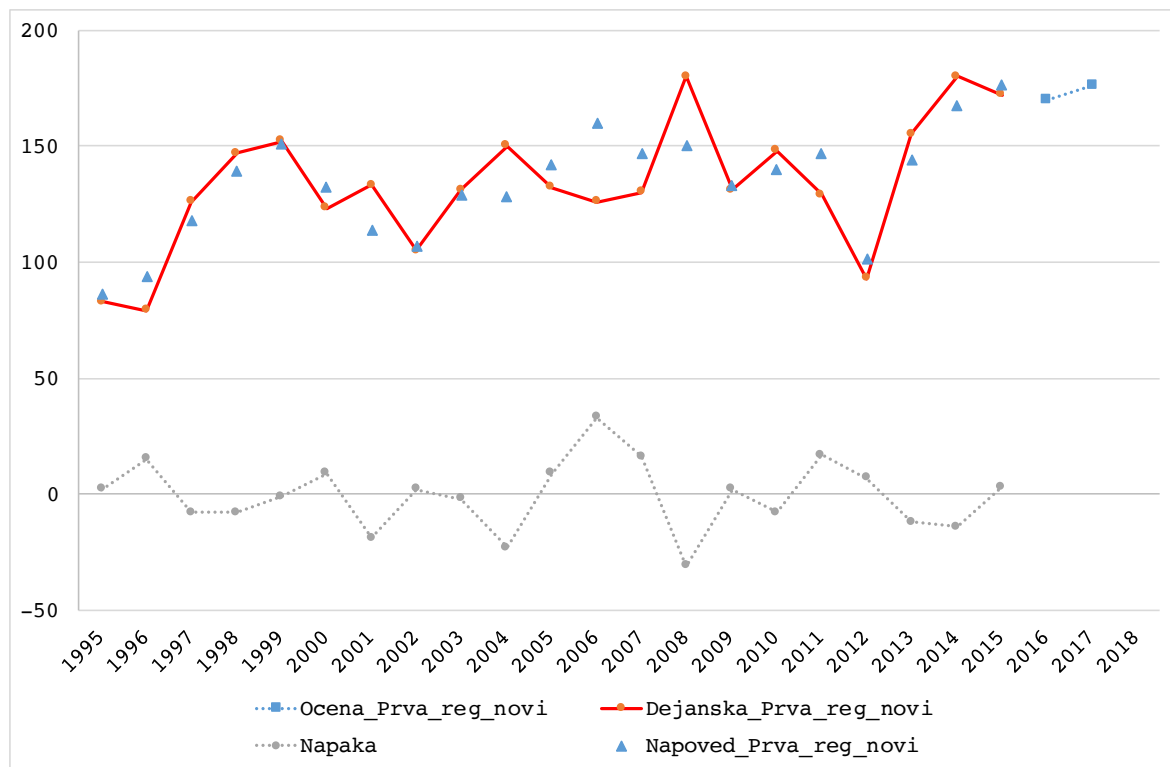
i	Vhodna spremenljivka x_i	Regresijski koeficient b_i	Standardna napaka b_i	t-statistika	p-vrednost	Oznaka pomembnosti
0		-276,120	411,866	-0,670	0,515	
1	[Export_DE]	-0,789	0,276	-2,865	0,014	**
2	[DV_predelovalne]	-0,000	0,059	-0,007	0,995	
3	[BDP]	0,026	0,040	0,653	0,526	
4	[Uvoz]	0,009	0,011	0,771	0,456	
5	[Obrestna_mera]	-5,612	6,925	-0,810	0,433	
6	[Reg_brezposelnost]	0,002	0,001	1,572	0,142	
7	[Zasebna_potrosnja]	-0,032	0,021	-1,584	0,139	
8	[Import_DE]	0,660	0,331	1,994	0,069	*

	Št. zapisov	R	R ²	popravljen R ²	RMSE	MAPE	stdev MAPE
Testni nabor	15	0,933	0,870		12	6,06 %	6,40 %
Celotni nabor	21	0,843	0,711	0,518	15	8,79 %	6,65 %



Ob velikem številu vhodnih spremenljivk je grafična analiza njihove medsebojne koreliranosti otežena. Zato Slika 50 prikazuje matriko medsebojnih korelacijskih koeficientov in VIF vrednosti za vsako posamično vhodno spremenljivko. Kot je razvidno so glede na visoke medsebojne korelacijske koeficiente r_{ij}^2 primerno zelo visoke tudi vrednosti VIF. Izstopa predvsem [BDP], ki ima izredno visoke medsebojne korelacijske koeficiente z vsemi vhodnimi spremenljivkami razen [Obrestna_mera] in [Reg_brezposelnost], kar je z vsebinskega stališča razumljivo in pričakovano. Ob dejstvu, da sem pri modeliranju uporabil tudi metodo izbora vhodnih spremenljivk, ki naj bi do neke mere preprečila, da v modelu ostanejo med seboj visoko korelirane vhodne spremenljivke, si lahko dobljene rezultate razlagam s tem, da enostavno izbrani nabor vhodnih spremenljivk ne popisuje najboljše dinamike obsega tega segmenta gospodarskih vozil.

Slika 49: Grafični prikaz rezultatov regresijskega modela segmenta avtobusov z uporabo letnega nabora podatkov



Slika 50: Prikaz korelacij med vhodnimi spremenljivkami letnega modela segmenta avtobusov

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	VIF_k
x_1	1,00								345,7
x_2	0,69	1,00							337,5
x_3	0,89	0,93	1,00						2281,0
x_4	0,94	0,88	0,98	1,00					231,9
x_5	0,05	0,04	0,06	0,04	1,00				3,6
x_6	0,10	0,58	0,36	0,27	0,01	1,00			21,4
x_7	0,86	0,87	0,96	0,92	0,05	0,28	1,00		113,7
x_8	0,99	0,67	0,87	0,92	0,05	0,09	0,09	1,00	283,8

x_1	[Export_DE]	x_5	[Obrestna_mera]
x_2	[DV_predelovalne]	x_6	[Reg_brezposelnost]
x_3	[BDP]	x_7	[Zasebna_potrosnja]
x_4	[Uvoz]	x_8	[Import_DE]

Na podlagi navedenih dejstev bi dobljeni model težko priporočil za uporabo v praksi, saj je očitno, da je kljub absolutno gledano ustrezni vrednosti determinacijskega koeficienta R^2 , ta umetno “napihnen” zaradi pojava multikolinearnosti. Če temu dodamo še majhen obseg podatkov, ki je skupen vsem letnim modelom, menim, da obstaja dovolj razlogov za dvom o napovedni vrednosti tega modela.

4.3.2 Rezultati z uporabo četrtnega nabora podatkov

V primerjavi z letnim modelom proces modeliranja na četrtnem naboru podatkov izloči 13 vhodnih spremenljivk. Rezultat je regresijski model z dvema vhodnima spremenljivkama [Investicije_v_OS] in [Reg_brezposelnost], kot ga prikazuje Slika 51. Obe spremenljivki imata regresijska koeficienta, ki na podlagi dobljenih vrednosti t-statistike in p-vrednosti dobro popisujeta odvisnost med izhodno in vhodnima spremenljivkama. Na drugi strani pa nam nizka vrednost R^2 , visoka relativna napaka MAPE in histogram porazdelitve napake ne dajejo velikega zaupanja v napovedno moč modela.

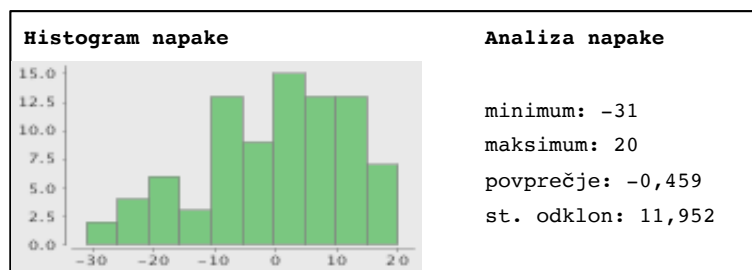
Vzrok za to gre iskati v dejstvu, da je obseg segmenta avtobusov na letni ravni zelo majhen, saj predstavlja v povprečju 3 % velikosti segmenta lahkih dostavnih vozil oziroma 13 % velikosti segmenta težkih tovornih vozil. Ko že tako majhen obseg segmenta opazujemo na nivoju četrtnega in ob dejstvu, da je v tem segmentu bolj prisotna večja koncentracija kupcev, je razumljivo, da napovedni model nima velike napovedne moči.

Slika 51: Regresijski model segmenta avtobusov z uporabo četrtnega nabora podatkov

$$[\text{Napoved_Prve_Reg_novi}] = b_0 + b_1 * [\text{Investicije_OS}] + b_2 * [\text{Reg_brezposelnost}]$$

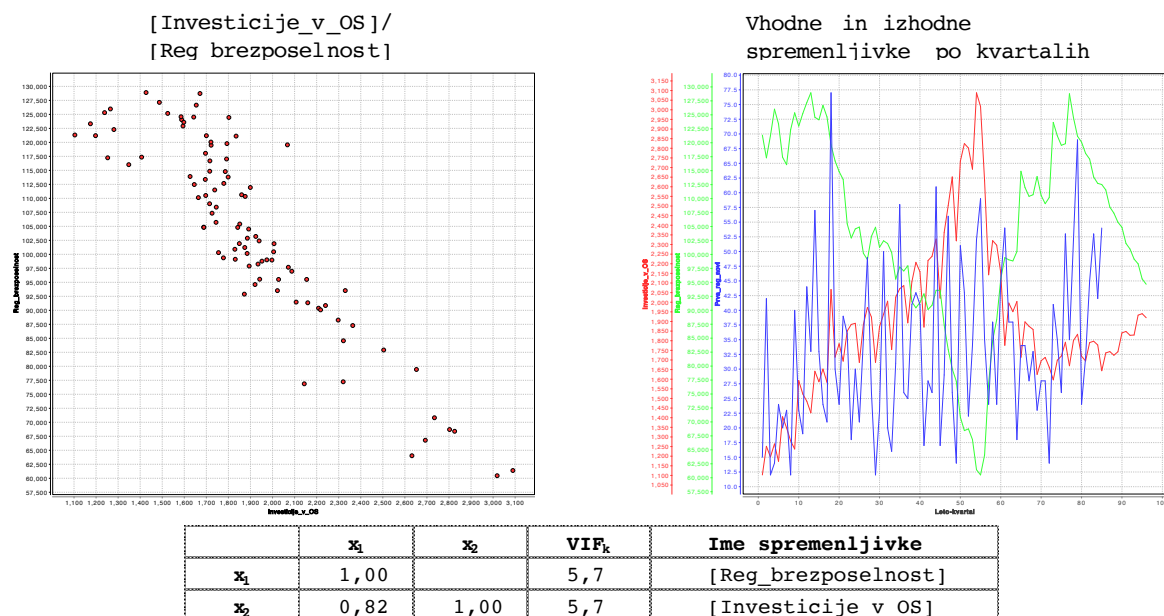
i	Vhodna spremenljivka x_i	Regresijski koeficient b_i	Standardna napaka b_i	t-statistika	p-vrednost	Oznaka pomembnosti
0		-137,657	32,407	-4,248	0,000	****
1	[Investicije_OS]	0,044	0,008	5,719	0,000	****
3	[Reg_brezposelnost]	0,001	0,000	4,742	0,000	****

	Št. zapisov	R	R^2	popravljen R^2	RMSE	MAPE	stdev MAPE
Testni nabor	60	0,707	0,499		13	38,10 %	28,48 %
Celotni nabor	85	0,541	0,292	0,275	12	34,22 %	27,02 %



Za dodatno manjše zaupanje v dobljeni regresijski model poskrbi še analiza medsebojne korelacije med vhodnima spremenljivkama, ki jo prikazuje Slika 52. Ta v primerjavi z dosedanjimi analizami kaže na relativno nizko korelacijo med vhodnima spremenljivkama ($r^2=0,82$), absolutno gledano pa po izkušnjah vrednost VIF nad 5 ni priporočljiva.

Slika 52: Grafični prikaz korelacij med vhodnimi in izhodnimi spremenljivkami četrletnega modela segmenta avtobusov



Vendar pa za napovedno moč modela to ni nujno slabo, zato je pri četrletnem modelu segmenta avtobusov bolj zaskrbljujoče dejstvo, da je determinacijski koeficient R^2 (0,292) izredno nizek. Uporaba sezonskih slammnatih spremenljivk nekoliko izboljša predvsem R^2 (iz 0,292 na 0,347), medtem ko napako (RMSE in MAPE) le minimalno zniža (Slika 53).

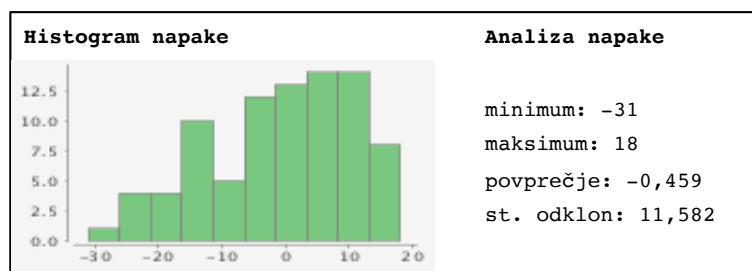
Slika 53: Regresijski model segmenta avtobusov z uporabo četrletnega nabora podatkov in slammnatimi spremenljivkami

$$[\text{Napoved_Prve_Reg_novi}] = b_0 + b_1 * [\text{Investicije_OS}] + b_2 * [\text{Reg_brezposelnost}] + b_3 * [\text{Kvartal_1}] + b_4 * [\text{Kvartal_2}] + b_5 * [\text{Kvartal_3}]$$

i	Vhodna spremenljivka x_i	Regresijski koeficijent b_i	Standardna napaka b_i	t-statistika	p-vrednost	Oznaka pomembnosti
0		-143,677	34,292	-4,190	0,000	****
1	[Investicije_OS]	0,044	0,008	5,405	0,000	****
2	[Reg_brezposelnost]	0,001	0,000	4,630	0,000	****
3	[Kvartal_1]	5,163	3,841	1,344	0,183	
4	[Kvartal_2]	8,472	3,666	2,311	0,023	**
5	[Kvartal_3]	7,813	3,670	2,129	0,036	**

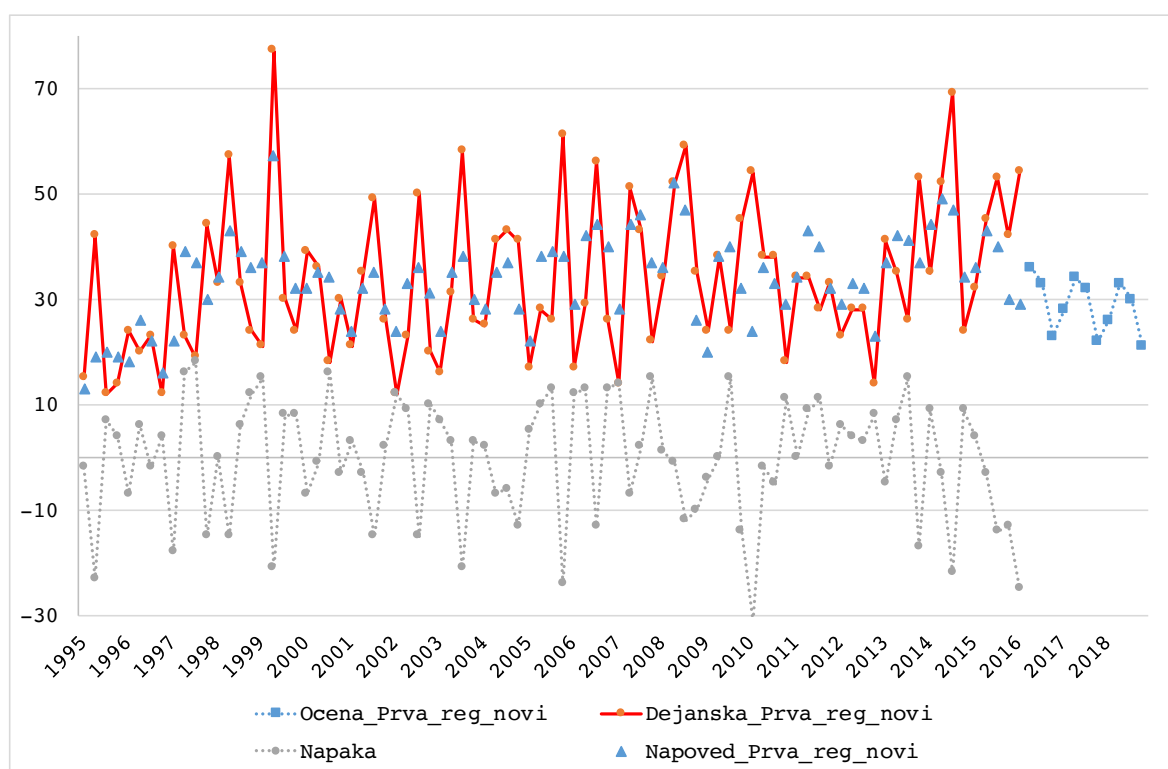
	Št. zapisov	R	R^2	popravljen R^2	RMSE	MAPE	stdev MAPE
Testni nabor	60	0,521	0,271		12	40,89 %	24,60 %
Celotni nabor	85	0,589	0,347	0,306	11	32,07 %	24,28 %

Slika 54: Analiza napake regresijskega modela segmenta avtobusov z uporabo četrletnega nabora podatkov



Visoko nihanje napake in histogram porazdelitve napake (Slika 54) nam sicer daje upanje, da model le ni tako slab, kot na prvi pogled prikazuje nizka vrednost R^2 . Dva od treh regresijskih koeficientov slamnatih sezonskih spremenljivk sta ocenjena kot nekoliko bolj pomembna, kar je ponovno drugače kot pri drugih dveh preučevanih segmentih gospodarskih vozil. Zanimivo je tudi to, da je determinacijski koeficient R^2 izračunan na testnem obsegu vzorca, manjši od R^2 , ki je izračunan na celotnem vzorcu, kar je prav tako odstopanje od rezultatov na ostalih dveh segmentih.

Slika 55: Grafični prikaz rezultatov regresijskega modela segmenta avtobusov z uporabo četrletnega nabora podatkov in slamnatih spremenljivk



Ne moremo tudi mimo dejstva, da kvartalni podatki o gibanju obsega prvih registracij avtobusov ne kažejo nekega trenda, temveč gre za relativno stacionarno nihanje, ki ga je z izbranimi vhodnimi spremenljivkami, ki popisujejo makroekonomsko okolje, očitno težko

popisati. Ne samo na četrtni ravni, tudi pri pregledu gibanja obsega prvih registracij avtobusov po letih, ki ga prikazuje Slika 49, lahko ugotovim, da obseg prvih registracij novih avtobusov nima nekega izrazitega trenda, kar lahko pripišem določeni zrelosti trga.

Po mnenju Dudnehöffer in Dudnehöffer (2014, str. 373) dajejo na zasičenih trgih napovedi temelječe na analizi časovnih vrst, bolj natančne rezultate kot ekonometrični modeli osnovani na regresijskih napovednih modelih z uporabo makroekonomskih vhodnih spremenljivk.

4.3.3 Rezultati z uporabo četrtnega nabora podatkov na letni ravni

S seštevanjem rezultatov napovedi četrtnega modela na letno raven dobimo sicer boljši rezultat kot na četrtni, vendar pa je v po vseh uporabljenih kazalnikih malenkost slabši od modela na osnovi letnega nabora podatkov. To je ponovno presenetljivo in v nasprotju s sedanjimi izkušnjami na ostalih dveh segmentih. Rezultate modela prikazuje Slika 56.

Slika 56: Model segmenta avtobusov z uporabo četrtnega nabora podatkov na letnem nivoju v primerjavi z letnim naborom podatkov

$$[\text{Napoved_Prve_Reg_novi_Q}] = [\text{Napoved_Prve_Reg_novi_Q1}] + [\text{Napoved_Prve_Reg_novi_Q2}] + [\text{Napoved_Prve_Reg_novi_Q3}] + [\text{Napoved_Prve_Reg_novi_Q4}]$$

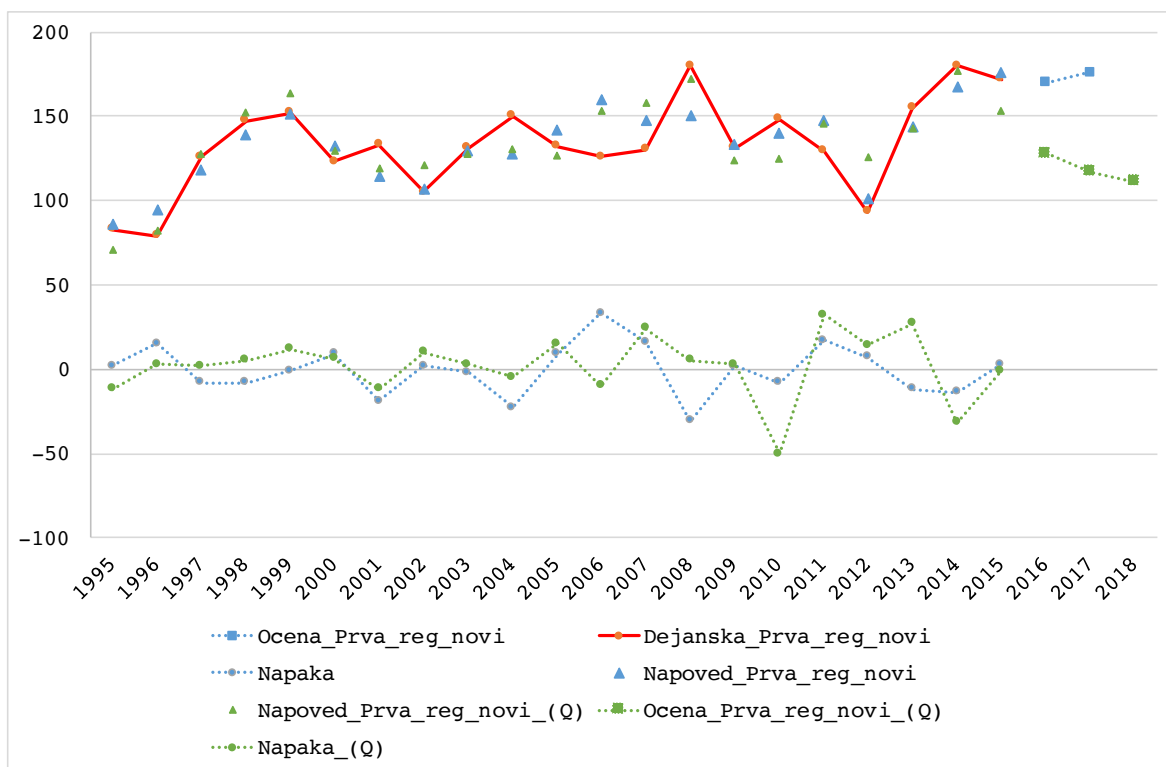
	Št. zapisov	R	R ²	popravljen R ²	RMSE	MAPE	stdev MAPE
Celotni nabor	85	0,791	0,625	0,500	18	10,08 %	8,56 %

Analiza napake

minimum: -50
maksimum: 32
povprečje: 1,86
st. odklon: 18,68

Slika 57 prikazuje rezultate modela v grafični obliki. Na diagramu lahko opazimo to, da letni in četrtni model, preračunano na letno raven nista usklajena v smeri trenda (rast, padec), kar dodatno vzbuja nezaupanje v napovedno sposobnost modela dobljenih modelov. Ne glede na to pa je potrebno upoštevati dejstvo, da je pri četrtnem regresijskem modelu algoritem za izračun regresijskih koeficientov uporabil naključno izbranih 60 zapisov za izračun 3 regresijskih koeficientov, pri letnem modelu pa samo 15 zapisov, na osnovi katerih je algoritem izračunal 9 regresijskih koeficientov. To dejstvo odseva vrednost popravljenega determinacijskega koeficienta R², ki je pri letnem modelu na osnovi četrtnega bistveno bližje R² kot pri letnem.

Slika 57: Grafični prikaz rezultatov modela avtobusov z uporabo četrtnega nabora na letni ravni v primerjavi z letnim naborom podatkov



4.3.4 Povzetek ugotovitev modeliranja segmenta avtobusov

Rezultate modeliranja segmenta avtobusov lahko strnem v naslednje ugotovitve:

- vsi modeli imajo relativno nizke koeficiente R^2 v primerjavi z ostalima dvema segmentoma,
- na podlagi uporabljenih kazalnikov (R^2 , RMSE, MAPE) da letni model boljše rezultate kot kateri koli model osnovan na četrtnem naboru podatkov,
- glede na visoko koreliranost med vhodnimi spremenljivkami letnega modela in razmerjem med obsegom podatkov (21 zapisov) in številom vhodnih spremenljivk je uporaba tega modela v praksi vprašljiva,
- glede na to, da četrtni model na letni ravni, daje skoraj tako dobre rezultate kot letni model, se kljub nekoliko slabšim kazalnikom odločim za uporabo četrtnega modela, saj temelji na večjem obsegu podatkov in ima manjše število vhodnih spremenljivk.

Primerjava med regresijskimi modeli segmenta avtobusov prikazuje Tabela 6.

Tabela 6: Primerjava rezultatov regresijskih modelov za segment avtobusov*

	Četrtni model	Četrtni sezonski model	Četrtni sezonski model letno	Letni model
Št. vhodnih spremenljivk	2	5	5 (20)	8
R^2 (na testnem naboru podatkov)	0,499	0,271	-	0,870
R^2 (na celotnem naboru podatkov)	0,292	0,347	0,625	0,711
popravljen R^2 (na celotnem naboru)	0,275	0,306	0,500	0,518
RMSE (na celotnem naboru podatkov)	12	11	18	15
Napaka min/max	-31/20	-31/18	-50/32	-31/30
Napaka povprečje	-0,5	-0,5	1,9	-0,5
Napaka standardni odklon	12	12	19	15
MAPE – relativna napaka +/- standardni odklon MAPE	34 % +/-27,0 %	32 % +/-24,3 %	10,1 % +/-8,0 %	9 % +/-6,7 %

Legenda: * Primerjava je izvedena na osnovi kazalnikov predstavljenih v poglavju 3.2.3

4.4 Povzetek rezultatov in ugotovitev

Rezultate izvedenega procesa modeliranja na vseh treh segmentih trga gospodarskih vozil lahko strnem v naslednje ugotovitve:

- vsi modeli imajo relativno visoke vrednosti R^2 , kar kaže na visoko napovedno sposobnost izračunanih regresijskih modelov,
- zaradi vrste in obsega nabora vhodnih spremenljivk (časovne vrste makroekonomskih spremenljivk) kljub uporabi algoritma selekcije s povratnim izločanjem se pri modeliranju nisem uspel popolnoma izogniti pojavu multikolinearnosti, zaradi česar so visoke vrednosti korelacijskih koeficientov »napihnjene«,
- povprečje absolutne vrednosti relativne napake MAPE je v območju, ki je sprejemljivo za uporabo napovedi v poslovni praksi na letni ravni.

Primerjava rezultatov regresijskih modelov za posamezne segmente gospodarskih vozil prikazuje Tabela 7.

Tabela 7: Primerjava rezultatov modelov za posamične segmente trga gospodarskih vozil*

	Lahka dostavna vozila	Težka tovorna vozila	Avtobusi
Št. vhodnih spremenljivk	6 (24)	6 (24)	5 (20)
R^2 (na celotnem naboru podatkov)	0,969	0,968	0,625
popravljen R^2 (na celotnem naboru)	0,956	0,954	0,500
RMSE (na celotnem naboru podatkov)	20	118	18
Napaka min/max	−430/338	−226/254	−50/32
Napaka povprečje	40,8	11,4	1,9
Napaka standardni odklon	209	121	19
MAPE – relativna napaka +/- standardni odklon MAPE	4,0 % +/-3,3 %	15,0 % +/-14,1 %	10,1 % +/-8,0 %
VIF _{x₁}	10,1	1,8	5,7
VIF _{x₂}	1,3	11,1	5,7
VIF _{x₃}	10,2	9,0	

Legenda: * Primerjava je izvedena na osnovi kazalnikov predstavljenih v poglavju 3.2.3

SKLEP

V magistrskem delu sem skladno s cilji najprej predstavil segmente trga gospodarskih vozil in njihove značilnosti. Glede na poznavanje posamičnih segmentov trga in razpoložljive podatke, sem v empiričnem delu magistrskega dela izbral tri največje in najbolj homogene segmente trga gospodarskih vozil v Sloveniji. Te segmente sem natančno tehnično opredelil glede na razpoložljive podatke o registriranih vozilih v Republiki Sloveniji. Ugotavljam, da so razpoložljivi podatki lahko dostopni in dovolj kakovostni za uporabo v podobnih raziskovalnih nalogah.

V drugem delu sem na osnovi analize strokovne literature pripravil kratek pregled metod napovedne analitike in vanj vključil spoznanja avtorjev s področja napovedovanja. Strnem lahko, da je področje napovedovanja izredno obsežno in dobro podkrepljeno tako s teoretičnimi kot tudi empiričnimi primerjalnimi študijami. Na drugi strani pa je napovedna analitika kot del podatkovnega rudarjenja še precej novo področje, ki se še uveljavlja.

V tretjemu delu magistrskega dela, sem skozi proces modeliranja razvil regresijske modele, namenjene napovedovanju obsega posamičnih segmentov gospodarskih vozil. Kot vhodne spremenljivke sem uporabil nabor izbranih makroekonomskih spremenljivk, za katere so javno dostopne tudi njihove napovedi za prihodnji dve leti. Sklenem lahko, da so razviti regresijski modeli pogojno primerni za ocenjevanje obsega trga gospodarskih vozil v Sloveniji v danem časovnem horizontu (2 leti). Zaradi uporabe časovne vrste makroekonomskih spremenljivk kot vhodnih spremenljivk je v modelih prisoten učinek multikolinearnosti, kar sicer ne zmanjšuje njihove napovedne moči, vendar pa zaradi velike variance regresijskih koeficientov povzroča nestabilnost modela. Na podlagi te ugotovitve bi bilo po mojem mnenju smiselno nadaljnje raziskave na tem področju usmeriti v razvoj modelov, ki temeljijo na združevanju napovedi več posamičnih modelov (angl. *ensemble modeling*). V te bi poleg modelov razvitih v tem magistrskem delu, vključil še modele, temelječe na drugih napovednih algoritmih (npr. časovne vrste, podporni vektorji) ali pa regresijske modele z drugačnim naborom vhodnih spremenljivk (npr. št. prevoženih ton kilometrov, število potnikov v javnem prevozu, življenjski cikel modelov, itd.).

LITERATURA IN VIRI

1. Anscombe, F.J. (1973). Graphs in Statistical Analysis. *American Statistician*, 27(1), 17–21.
2. Armstrong, J. S., Green, K. C., & Graefe A. (2015). Golden Rule of Forecasting: Be Conservative. *Journal of Business Research*, 68(8), 1717–1731.
3. Armstrong, J. S., & Brodie R. J. (1999). Forecasting for Marketing. V G. J. Hooley & M. K. Hussey (ur.), *Quantative Methods in Marketing* (2nd Ed.) (str. 92–119). London: International Thompson Business Press.
4. Armstrong, J. S., & Fildes R. (2006). Making progres in forecasting. *International Journal of Forecasting*, 2(3), 433–44.
5. Armstrong, J. S., & Green, K. C. (2014). *Methodology Tree for Forecasting*. Najdeno 20. maja 2016 na spletnem naslovu <http://www.forecastingprinciples.com/index.php/methodology-tree>
6. Armstrong, J. S., & Green, K. C. (2015). Simple versus Complex Forecasting: The Evidence. *Journal of Business Research*, 68(8), 1678–1685.
7. Armstrong, J. S., Morwitz, V.G., & Kumar, V. (2000). Sales Forecasts for Existing Consumer Products and Services: Do Purchase Intentions Contribute to Accuracy? *International Journal of Forecasting*, 16(3), 383–397.
8. Banka Slovenije. (b.l.). *Obrestne mere*. Najdeno dne 20. aprila 2016 na spletnem naslovu <http://www.bsi.si/financni-podatki.asp?MapaId=1001>
9. Breiman, L., Friedman, J., Stone, C. J., & Olshen, R. A. (1984). *Classification and regression trees*. Boca Raton: CRC Press.
10. Brühl, B., Hülsmann, M., Borscheid, D., Friedrich, C., & Reith, D., (2009). A Sales Forecast Model for the German Automobile Market Based on Time Series Analysis and Data Mining Methods. V P. Perner (ur.), *Advances in Data Mining Applications and Theoretical Aspecst* (str. 146–160). Berlin /Heidelberg: Springer.
11. Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., & Wirth, R. (2000). CRISP-DM 1.0: *Step-by-step Data Mining guide*. SPSS Inc. Najdeno 20. maja 2016 na spletnem naslovu <ftp://ftp.software.ibm.com/software/analytics/spss/support/Modeler/Documentation/14/UserManual/CRISP-DM.pdf>
12. Cimperman, F., & Savšek, S. (2014). Natančnost napovedi makroekonomskih agregatov Slovenije. *Prikazi in analize 2014, marec*. Najdeno 7. februarja 2016 na spletnem naslovu <http://www.bsi.si/library/includes/datoteka.asp?DatotekaId=5635>
13. Collopy, F. (2012). Interview of J. Scott Armstrong for the International Journal of Forecasting. *The Oracle*, 13(4), 8–14.
14. Data Mininig Group. (2016). *PMML 4.3 - General Structure*. Najdeno 20. maja 2016 na spletnem naslovu <http://dmg.org/pmml/v4-2-1/GeneralStructure.html>
15. Direktiva Sveta 85/347/EGS o spremembi Direktive 68/297/EGS o standardizaciji predpisov o brezcarinskem vnosu goriva, vsebovanega v rezervoarjih za gorivo gospodarskih motornih vozil. *Uradni list L183*, 16.07.1985. str 22–23.

16. Dudenhöffer, F., & Dudenhöffer, K. (2014). Automobilmarkt-Prognosen: Modelle und Methoden. V B. Ebel & M. B. Hofer (ur.), *Automotive Management: Strategie und Marketing in der Automobilwirtschaft* (str 367–377). Berlin: Springer Verlag.
17. Economy Forecast Agency. (b.l.). *Euribor Forecast for 2016 and 2017*. Najdeno 22. aprila na spletnem naslovu <http://longforecast.com/euribor/euribor-forecast-for-2015-2016-and-2017.html>
18. Eurostat. (b.l.). *Eurostat Data Database*. Najdeno 22. aprila na spletnem naslovu <http://ec.europa.eu/eurostat/data/database>
19. Evropska Centralna Banka. (b.l.). *Statistical Data Warehouse*. Najdeno dne 20. aprila na spletnem naslovu <http://sdw.ecb.europa.eu>
20. Fantazzini, D., & Toktamysova, Z. (2015). Forecasting German Car Sales Using Google Data and Multivariate Models. *International Journal of Production Economics*, 170(december 2015). Najdeno 26. marca 2016 na spletnem naslovu <https://mpira.ub.uni-muenchen.de/id/eprint/67110>
21. Fides, R., & Nikolopoulos, K. (2006). Spyros Makridakis: An Interview With the International Journal of Forecasting. *International Journal of Forecasting*, 22(3), 625–636.
22. Gartner. (2016, 9. februar). *Magic Quadrant for Advanced Analytics Platforms*. Najdeno 20. marca 2016 na <https://www.gartner.com/doc/reprints?id=12YC9GD6&ct=160209&st=sb>
23. Gujarati, D.N. (2003). *Basic Econometrics* (4th Ed.). New York: McGraw-Hill/Irwin.
24. Hülsmann, M., Borscheid, D., Freidrich, C. M., & Reith, D. (2012). General Sales Forecast Models for Automobile Markets and their Analysis. *Transactions on Machine Learning and Data Mining*, 5(2), 65–86.
25. Hyndman, R. J., & Athanasopoulos, G. (2016). *Forecasting: Principles and Practice*. Najdeno 8. februarja 2016 na spletnem naslovu <https://www.otexts.org/book/fpp>
26. ifo Institut. (2016). Gemeinschaftsdiagnose Frühjahr 2016: Aufschwung bleibt moderat – Wirtschaftspolitik wenig wachstumsorientiert. *ifo Schnelldienst*, 69(08), 17–58.
27. Kotu, V., & Deshpande, B. (2015). *Predictive Analytics and Data Mining*. Amsterdam:Elsevier.
28. Ministrstvo za infrastrukturo. (2015). *Evidenca registriranih vozil – preseki stanja 1992–2005*. Najdeno 22. februarja 2016 na spletnem naslovu <https://nio.gov.si/nio/data/evidenca+registriranih+vozil+preseki+stanja+19922005>
29. Ministrstvo za infrastrukturo. (2016a). *Evidenca registriranih vozil – preseki stanja 2006–2015*. Najdeno 22. februarja 2016 na spletnem naslovu <https://nio.gov.si/nio/data/evidenca+registriranih+vozil+31122012>
30. Ministrstvo za infrastrukturo. (2016b). *Prvič registrirana vozila v letu 2015, po mesecih*. Najdeno 14. maja 2016 na spletnem naslovu <https://nio.gov.si/nio/data/prvic+registrirana+vozila+v+letu+2015+po+mesecih>
31. Quinlan, J. R. (1986). Induction of Decision Trees. *Machine Learning*, 1(1), 81–106.

32. Rexer, K. (2013). *2013 Data Miner Survey Summary Report*. Prejeto 20. maja 2016 po elektronski pošti z naslova DataMinerSurvey@RexerAnalytics.com.
33. Sa-ngasoongsong, A., Bukkapatnam, S.T.S., Kim, J., Iyer, P., & Suresh, R.P. (2012). Multi-Step Sales Forecasting in Automotive Industry Based on Structural Relationship Identification. *International Journal of Production Economics*, 140(2), 875–887.
34. Shahabuddin, S. (2009). Forecasting automobile sales. *Management Research News*, 32(7), 670–682.
35. Shmueli, G. (2011). *Practical Time Series Forecasting: A Hands-On Guide* (2nd Ed.). Arlington: Statistics.com LLC.
36. Statistični urad Republike Slovenije. (2016). *Metodološka pojasnila – Registrirana cestna vozila*. Najdeno 13. maja 2016 na spletnem naslovu <http://www.stat.si/statweb/Common/PrikaziDokument.ashx?IdDatoteke=8193>
37. Statistični urad Republike Slovenije. (b.l.). *Podatkovni portal SI-STAT*. Najdeno 7. februarja 2016 na spletnem naslovu <http://pxweb.stat.si/pxweb/Dialog/statfile2.asp>
38. Statistisches Bundesamt. (b.l.). *GENESIS-Online Datenbank*. Najdeno dne 27. aprila 2016 na spletnem naslovu <https://www-genesis.destatis.de/genesis/online>
39. United Nations Economic Commission for Europe. (2016). *Consolidated Resolution on the Construction of Vehicles (R.E.3)*. Najdeno 13. maja 2016 na spletnem naslovu <http://www.unece.org/fileadmin/DAM/trans/main/wp29/wp29resolutions/ECE-TRANS-WP29-78-r4e.pdf>
40. Urad RS za makroekonomske analize in razvoj. (2016, 16. marec). *Pomladanska napoved gospodarskih gibanj 2016*. Najdeno 13. aprila 2016 na spletnem naslovu http://www.umar.gov.si/fileadmin/user_upload/publikacije/analiza/Pomladanska_napoved_2016/majska_2016_splet-18marec.pdf
41. Wang, F.-K., Chang, K.-K., & Tzeng, C.-W. (2011). Using adaptive network based fuzzy inference system to forecast automobile sales. *Expert System with Applications*, 38(8), 10587–10593.
42. Zakon o motornih vozilih. Uradni list RS št. 106/2010, 23/2015.
43. Zaslow, J. (2002). If TiVo Thinks You Are Gay, Here's How to Set It Straight. Najdeno 20. maja 2016 na spletnem naslovu <http://www.wsj.com/articles/SB1038261936872356908>
44. Zavod Republike Slovenije za zaposlovanje. (b.l.). *Registrirana brezposelnost*. Najdeno dne 23. februarja 2016 na spletnem naslovu http://www.ess.gov.si/trg_dela/trg_dela_v_stevilkah/registrirana_brezposelnost
45. Združenje evropskih avtomobilskih proizvajalcev ACEA. (2015, junij). *ACEA Pocket Guide 2015–2016*. Najdeno 20. marca 2016 na spletnem naslovu <http://www.acea.be/publications/article/acea-pocket-guide>
46. Združenje evropskih avtomobilskih proizvajalcev ACEA. (2016, 22. april). *New Commercial Vehicle Registrations European Union, March 2016*. Najdeno 12.5.2016 na spletnem naslovu http://www.acea.be/uploads/press_releases_files/20160422_PRCV-1603-FINAL.PDF

PRILOGE

KAZALO PRILOG

Priloga 1: Zbrani podatki o registriranih gospodarskih vozilih	1
Priloga 2: Uporabljene makroekonomske spremenljivke	6
Priloga 3: Procesi uporabljeni v RapidMiner Studiu	15

Priloga 1: Zbrani podatki o registriranih gospodarskih vozilih

Tabela 1: Podatki o prvih registracijah novih gospodarskih vozil po segmentih trga po letih (število vozil)

Leto	LDV Prva_reg_novi	TTV Prva_reg_novi	BUS Prva_reg_novi
1995	2.994	317	83
1996	2.931	334	79
1997	2.910	234	126
1998	2.765	433	147
1999	3.265	609	152
2000	3.404	470	123
2001	3.563	521	133
2002	3.911	612	105
2003	4.263	660	131
2004	4.700	930	150
2005	4.641	1.074	132
2006	5.618	1.823	126
2007	6.556	2.607	130
2008	6.992	2.341	180
2009	4.000	576	131
2010	3.900	733	148
2011	4.355	1.210	129
2012	4.061	994	93
2013	4.203	1.103	155
2014	4.742	1.420	180
2015	5.842	1.921	172

*Vir: Ministrstvo za infrastrukturo, Evidenca registriranih vozil – preseki stanja 1992–2005, 2015;
Ministrstvo za infrastrukturo, Evidenca registriranih vozil – preseki stanja 2006–2015, 2016a;
lastni izračuni.*

Tabela 2: Podatki o prvih registracijah gospodarskih vozil po segmentih trga po letih (število vozil)

Leto	LDV Prva_reg	TTV Prva_reg	BUS Prva_reg
1995	3.317	591	98
1996	3.120	574	91
1997	3.047	528	140
1998	2.928	720	160
1999	3.454	876	157
2000	3.511	757	133
2001	4.024	939	154
2002	4.136	1.065	131
2003	4.506	1.100	168
2004	5.052	1.626	230
2005	5.175	1.908	206
2006	6.157	3.084	203
2007	7.461	3.825	193
2008	7.863	3.267	222
2009	4.662	1.181	168
2010	4.474	1.578	199
2011	4.865	1.925	162
2012	4.506	1.710	132
2013	4.788	2.122	234
2014	5.439	2.652	279
2015	6.753	3.363	233

*Vir: Ministrstvo za infrastrukturo, Evidenca registriranih vozil – preseki stanja 1992–2005, 2015;
Ministrstvo za infrastrukturo, Evidenca registriranih vozil – preseki stanja 2006–2015, 2016a;
lastni izračuni.*

Tabela 3: Podatki o stanju registriranih gospodarskih vozil po segmentih trga po letih (število vozil na zadnji dan leta)

Leto	LDV Stanje	TTV Stanje	BUS Stanje
1995	22.528	4.968	2.478
1996	24.734	5.405	2.411
1997	26.901	5.720	2.372
1998	28.739	6.131	2.327
1999	30.653	6.763	2.319
2000	32.864	7.256	2.259
2001	40.006	11.470	2.217
2002	41.860	7.822	2.201
2003	43.788	8.501	2.195
2004	46.510	9.598	2.276
2005	48.903	10.781	2.256
2006	50.958	13.561	2.291
2007	56.125	16.532	2.349
2008	61.236	18.620	2.398
2009	62.310	18.147	2.411
2010	63.403	18.195	2.431
2011	64.527	18.038	2.442
2012	65.235	17.935	2.429
2013	66.393	18.374	2.486
2014	68.881	19.325	2.576
2015	72.653	20.726	2.645

*Vir: Ministrstvo za infrastrukturo, Evidenca registriranih vozil – preseki stanja 1992–2005, 2015;
Ministrstvo za infrastrukturo, Evidenca registriranih vozil – preseki stanja 2006–2015, 2016a;
lastni izračuni.*

Tabela 4: Podatki o prvih registracijah gospodarskih vozil po segmentih trga po četrletjih (število vozil)

Leto-četrletje	LDV Prva_reg_novi	TTV Prva_reg_novi	BUS Prva_reg_novi
1995 Q1	628	97	15
1995 Q2	815	68	42
1995 Q3	855	100	12
1995 Q4	696	52	14
1996 Q1	625	96	24
1996 Q2	861	71	20
1996 Q3	705	101	23
1996 Q4	740	66	12
1997 Q1	650	45	40
1997 Q2	813	52	23
1997 Q3	762	68	19
1997 Q4	685	69	44
1998 Q1	599	81	33
1998 Q2	733	97	57
1998 Q3	691	139	33
1998 Q4	742	116	24
1999 Q1	684	136	21
1999 Q2	1.118	204	77
1999 Q3	560	143	30
1999 Q4	903	126	24
2000 Q1	832	119	39
2000 Q2	956	130	36
2000 Q3	716	117	18
2000 Q4	900	104	30
2001 Q1	927	135	21
2001 Q2	833	133	35
2001 Q3	740	124	49
2001 Q4	1.063	129	26
2002 Q1	900	135	12
2002 Q2	904	131	23
2002 Q3	918	177	50
2002 Q4	1.189	169	20
2003 Q1	894	177	16
2003 Q2	1.021	178	31
2003 Q3	923	112	58
2003 Q4	1.425	193	26
2004 Q1	1.100	233	25
2004 Q2	1.165	230	41
2004 Q3	953	208	43
2004 Q4	1.482	259	41
2005 Q1	985	225	17
2005 Q2	1.052	253	28
2005 Q3	919	275	26
2005 Q4	1.685	321	61
2006 Q1	1.060	373	17
2006 Q2	1.188	481	29

se nadaljuje

Tabela 4: Podatki o prvih registracijah gospodarskih vozil po segmentih trga po četrtnjih (število vozil) – (nad.)

Leto-četrtnje	LDV	TTV	BUS
	Prva_reg_novi	Prva_reg_novi	Prva_reg_novi
2006 Q3	1.263	469	56
2006 Q4	2.107	500	26
2007 Q1	1.532	657	14
2007 Q2	1.601	620	51
2007 Q3	1.508	694	43
2007 Q4	1.915	636	22
2008 Q1	1.710	807	34
2008 Q2	1.941	637	52
2008 Q3	1.796	556	59
2008 Q4	1.545	341	35
2009 Q1	929	158	24
2009 Q2	984	129	38
2009 Q3	926	119	24
2009 Q4	1.161	170	45
2010 Q1	962	141	54
2010 Q2	944	153	38
2010 Q3	842	177	38
2010 Q4	1.152	262	18
2011 Q1	1.104	367	34
2011 Q2	1.069	330	34
2011 Q3	943	260	28
2011 Q4	1.239	253	33
2012 Q1	1.072	254	23
2012 Q2	978	275	28
2012 Q3	773	248	28
2012 Q4	1.238	217	14
2013 Q1	908	247	41
2013 Q2	1.028	257	35
2013 Q3	933	188	26
2013 Q4	1.334	411	53
2014 Q1	1.128	337	35
2014 Q2	1.138	355	52
2014 Q3	1.126	310	69
2014 Q4	1.350	418	24
2015 Q1	1.293	452	32
2015 Q2	1.379	449	45
2015 Q3	1.275	419	53
2015 Q4	1.895	601	42

Vir: Ministrstvo za infrastrukturo, Evidenca registriranih vozil – preseki stanja 1992–2005, 2015; Ministrstvo za infrastrukturo, Evidenca registriranih vozil – preseki stanja 2006–2015, 2016a; lastni izračuni.

Priloga 2: Uporabljene makroekonomske spremenljivke

Tabela 5: Pregled uporabljenih makroekonomskih spremenljivk, frekvenca podatkov, njihovi viri in uporabljene enote

Ime spremenljivke v RapidMiner	Ekonomska spremenljivka	Enota	Vir historičnih podatkov*	Frekvenca podatkov**	Vir napovedi
[BDP]	BDP Slovenija	mio EUR stalne cene 2010	SURS	L, K	UMAR
[Reg_brezposelnost]	Št. registriranih brezposelnih	število (letno povprečje)	ZRSZ	L, M	UMAR
[Izvoz]	Izvoz blaga in storitev	mio EUR stalne cene 2010	SURS	L, K	UMAR
[Uvoz]	Uvoz blaga in storitev	mio EUR stalne cene 2010	SURS	L, K	UMAR
[Drzavna_potrosnja]	Državna potrošnja	mio EUR stalne cene 2010	SURS	L, K	UMAR
[Zasebna_potrosnja]	Zasebna potrošnja	mio EUR stalne cene 2010	SURS	L, K	UMAR
[Investicije_v_OS]	Investicije v osnovna sredstva	mio EUR stalne cene 2010	SURS	L, K	UMAR
[Inflacija]	Indeks cen življenjskih potrebščin	indeks povprečje obdobja/ povprečje prejšnjega obdobja	SURS	L, M	UMAR
[Cena_nafte]	Cena nafte	EUR za sod	ECB	L, K	UMAR
[Export_DE]	Uvoz blaga in storitev Nemčija	mio EUR stalne cene 2010	ESTAT, SB	L, K	Ifo
[Import_DE]	Izvoz blaga in storitev Nemčija	mio EUR stalne cene 2010	ESTAT, SB	L, K	Ifo
[DV_gradbenistvo]	Dodana vrednost gradbeništvo	mio EUR stalne cene 2010	SURS	L, K	UMAR
[DV_trgovina]	Dodana vrednost trgovina in popravila vozil, promet in skladiščenje, gostinstvo	mio EUR stalne cene 2010	SURS	L, K	UMAR
[DV_gradbenistvo]	Dodana vrednost rudarstvo, predelovalne dejavnosti, oskrba z elektriko in vodo, ravnanje z odplakami, saniranje okolja	mio EUR stalne cene 2010	SURS	L, K	UMAR
[Obrestna_mera]	Obrestna mera***	v %, povprečje mesečnih povprečnih vrednosti	BS, ECB	L, M	EFA

Legenda:

* SURS=Statistični urad Republike Slovenije (b.l.), ZRSZ=Zavod Republike Slovenije zaposlovanje (b.l.), BS=Banka Slovenije (b.l.), ECB=Evropska centralna banka (b.l.), UMAR=Urad za makroekonomske analize in razvoj (2016), ifo=ifo Institute (2016), EFA=Economy Forecast Agency (b.l.), ESTAT=Eurostat (b.l.), SB=Statistisches Bundesamt (b.l.).

** L=letno, K=kvartalno, M=mesečno.

*** časovno vrsto podatkov za povprečno obrestno mero Banka Slovenije objavlja samo za časovno obdobje 2003–2016 zato je za obdobje 1995 do 2002 uporabljena obrestna mera sestavljena iz 12m EURIBOR in povprečne razlike med povprečno obrestno mero in 12m EURIBOR v letih 2003–2006.

Tabela 6: Makroekonomske spremenljivke iz letnega nabora podatkov 1/3

Leto	BDP	Izvoz	Uvoz	Reg_brezposelnost	Drzavna_potrosnja
1995	22.631	8.794	9.298	121.483	4.725
1996	23.427	9.056	9.521	119.799	4.850
1997	24.625	10.109	10.634	125.189	5.066
1998	25.434	10.900	11.704	126.080	5.262
1999	26.776	11.118	12.688	118.951	5.486
2000	27.888	12.518	13.528	106.601	5.694
2001	28.711	13.417	14.018	101.857	5.854
2002	29.812	14.462	14.807	102.635	6.043
2003	30.659	14.919	15.764	97.674	6.205
2004	31.994	16.865	17.966	92.826	6.373
2005	33.274	18.781	19.284	91.889	6.544
2006	35.156	21.433	21.673	85.836	6.748
2007	37.597	24.350	25.305	71.336	6.879
2008	38.838	25.365	26.279	63.216	7.218
2009	35.809	21.158	21.328	86.354	7.392
2010	36.252	23.306	22.786	100.504	7.353
2011	36.488	24.913	23.924	110.692	7.299
2012	35.496	25.059	23.038	110.183	7.134
2013	35.121	25.823	23.425	119.827	7.024
2014	36.191	27.321	24.366	120.109	7.020
2015	37.233	28.733	25.437	112.726	7.071
Napoved 2016	37.866	29.796	26.200	107.400	7.092
Napoved 2017	38.775	31.226	27.536	101.000	7.106
Napoved 2018	39.667	32.757	28.913	94.700	7.113

Tabela 7: Makroekonomske spremenljivke iz letnega nabora podatkov 2/3

Leto	Zasebna_potrosnja	Investicije_v_OS	Inflacija	Cena_nafte	Export_DE
1995	13.194	4.814	114	13,0	433
1996	13.630	5.205	110	16,0	458
1997	14.046	5.839	108	17,0	515
1998	14.459	6.307	108	12,0	554
1999	15.402	7.150	106	17,1	584
2000	15.489	7.319	109	31,0	665
2001	15.890	7.466	108	27,8	703
2002	16.281	7.504	108	26,5	733
2003	16.840	7.937	106	25,1	746
2004	17.344	8.365	104	30,5	832
2005	17.752	8.655	103	44,6	887
2006	17.964	9.537	103	52,9	996
2007	19.097	10.678	104	52,8	1.089
2008	19.585	11.428	106	65,9	1.110
2009	19.766	8.913	101	44,6	952
2010	19.980	7.727	102	60,7	1.090
2011	19.985	7.346	102	79,7	1.180
2012	19.497	6.697	103	86,6	1.213
2013	18.688	6.808	102	81,7	1.232
2014	18.801	7.024	100	74,5	1.282
2015	19.124	7.061	100	48,3	1.351
Napoved 2016	19.526	6.849	101	31,5	1.413
Napoved 2017	19.858	7.260	101	37,3	1.471
Napoved 2018	20.195	7.623	101	40,5	-

Tabela 8: Makroekonomske spremenljivke iz letnega nabora podatkov 3/3

Leto	Import_DE	DV_gradbenistvo	DV_trgovina	DV_predelovalne	Obrestna_mera
1995	437	1.370	3.799	5.457	8,4
1996	454	1.512	3.810	5.619	6,4
1997	496	1.579	4.119	6.014	5,7
1998	542	1.587	4.205	6.115	5,2
1999	590	1.815	4.394	6.555	4,5
2000	654	1.798	4.599	6.748	6,1
2001	660	1.792	4.941	7.114	5,4
2002	643	1.837	5.200	7.452	4,8
2003	680	1.933	5.339	7.678	3,9
2004	734	1.948	5.497	7.868	3,6
2005	776	2.013	5.779	8.237	3,6
2006	863	2.305	6.191	8.959	4,5
2007	916	2.718	6.613	9.853	5,9
2008	937	2.847	6.837	10.193	6,7
2009	847	2.464	6.232	9.175	5,9
2010	956	2.015	6.300	8.789	5,7
2011	1.023	1.812	6.408	8.700	5,8
2012	1.020	1.673	6.144	8.281	5,8
2013	1.051	1.526	6.141	8.135	5,7
2014	1.090	1.671	6.360	8.530	5,0
2015	1.154	1.616	6.630	8.380	3,5
Napoved 2016	1.217	1.430	6.842	8.656	3,4
Napoved 2017	1.276	1.488	7.027	8.959	3,4
Napoved 2018	-	1.532	7.210	9.273	3,4

Tabela 9: Makroekonomske spremenljivke iz četrtnega nabora podatkov 1/3

Leto-kvartal	BDP	Izvoz	Uvoz	Reg_brezposelnost	Drzavna_potrosnja
1995 Q1	5.346	2.201	2.151	121.370	1.151
1995 Q2	5.818	2.249	2.470	117.293	1.197
1995 Q3	5.684	2.156	2.220	121.250	1.174
1995 Q4	5.783	2.189	2.457	126.020	1.203
1996 Q1	5.472	2.111	2.211	123.374	1.177
1996 Q2	5.940	2.316	2.474	117.412	1.219
1996 Q3	5.924	2.295	2.390	116.067	1.211
1996 Q4	6.091	2.333	2.446	122.341	1.243
1997 Q1	5.679	2.242	2.374	125.365	1.223
1997 Q2	6.333	2.624	2.828	122.970	1.272
1997 Q3	6.226	2.570	2.667	125.225	1.261
1997 Q4	6.387	2.673	2.766	127.194	1.310
1998 Q1	5.955	2.569	2.780	128.948	1.269
1998 Q2	6.452	2.761	2.937	124.565	1.323
1998 Q3	6.405	2.763	2.896	124.104	1.311
1998 Q4	6.622	2.807	3.090	126.702	1.359
1999 Q1	6.202	2.591	2.859	124.622	1.329
1999 Q2	6.924	2.786	3.510	119.605	1.385
1999 Q3	6.674	2.788	2.997	116.741	1.366
1999 Q4	6.975	2.954	3.322	114.836	1.406
2000 Q1	6.595	2.934	3.188	113.428	1.394
2000 Q2	7.087	3.167	3.421	105.469	1.439
2000 Q3	7.007	3.152	3.471	102.938	1.407
2000 Q4	7.199	3.266	3.448	104.568	1.455
2001 Q1	6.852	3.272	3.308	104.904	1.425
2001 Q2	7.247	3.352	3.554	100.208	1.474
2001 Q3	7.179	3.379	3.572	99.065	1.454
2001 Q4	7.433	3.414	3.585	103.249	1.500
2002 Q1	7.076	3.426	3.445	104.871	1.463
2002 Q2	7.495	3.645	3.734	101.270	1.521
2002 Q3	7.474	3.685	3.764	102.448	1.495
2002 Q4	7.768	3.705	3.865	101.949	1.564
2003 Q1	7.287	3.535	3.616	100.357	1.511
2003 Q2	7.710	3.727	3.960	95.567	1.569
2003 Q3	7.723	3.802	4.052	97.743	1.539
2003 Q4	7.940	3.854	4.137	97.031	1.586
2004 Q1	7.556	3.783	3.939	97.957	1.541
2004 Q2	8.075	4.286	4.639	91.520	1.598
2004 Q3	8.105	4.370	4.625	90.431	1.577
2004 Q4	8.258	4.425	4.763	91.396	1.656
2005 Q1	7.783	4.223	4.329	92.943	1.581
2005 Q2	8.500	4.741	4.747	90.130	1.647
2005 Q3	8.413	4.903	4.898	90.907	1.631
2005 Q4	8.580	4.913	5.310	93.577	1.685
2006 Q1	8.200	5.010	5.003	93.566	1.635
2006 Q2	8.916	5.409	5.274	87.331	1.701

se nadaljuje

Tabela 9: Makroekonomske spremenljivke iz četrtnega nabora podatkov 1/3 (nad.)

Leto-kvartal	BDP	Izvoz	Uvoz	Reg_brezposelnost	Drzavna_potrosnja
2006 Q3	8.929	5.401	5.400	82.964	1.672
2006 Q4	9.111	5.613	5.996	79.482	1.740
2007 Q1	8.839	5.802	5.839	77.285	1.649
2007 Q2	9.536	6.141	6.250	70.858	1.724
2007 Q3	9.608	6.219	6.498	68.444	1.705
2007 Q4	9.614	6.188	6.718	68.755	1.801
2008 Q1	9.360	6.301	6.384	66.843	1.744
2008 Q2	10.112	6.732	6.845	61.438	1.798
2008 Q3	9.902	6.448	6.753	60.507	1.781
2008 Q4	9.463	5.885	6.298	64.074	1.895
2009 Q1	8.657	5.044	4.962	76.925	1.793
2009 Q2	9.123	5.223	5.151	84.611	1.865
2009 Q3	9.048	5.339	5.520	88.310	1.841
2009 Q4	8.982	5.552	5.694	95.570	1.893
2010 Q1	8.558	5.386	5.233	99.423	1.797
2010 Q2	9.283	5.855	5.657	99.029	1.848
2010 Q3	9.214	6.001	5.787	98.835	1.832
2010 Q4	9.198	6.064	6.109	100.504	1.876
2011 Q1	8.776	5.985	5.770	114.896	1.784
2011 Q2	9.434	6.394	6.027	111.994	1.827
2011 Q3	9.264	6.298	6.061	110.397	1.817
2011 Q4	9.013	6.236	6.067	110.692	1.872
2012 Q1	8.728	6.076	5.725	113.953	1.761
2012 Q2	9.090	6.376	5.773	110.562	1.797
2012 Q3	8.975	6.284	5.671	109.086	1.764
2012 Q4	8.703	6.323	5.869	110.183	1.811
2013 Q1	8.324	6.174	5.599	123.651	1.728
2013 Q2	8.968	6.513	5.860	121.244	1.770
2013 Q3	8.946	6.556	5.823	119.542	1.737
2013 Q4	8.882	6.580	6.144	119.827	1.789
2014 Q1	8.518	6.453	5.772	128.779	1.717
2014 Q2	9.268	6.797	6.084	124.499	1.771
2014 Q3	9.272	6.977	6.148	121.171	1.728
2014 Q4	9.133	7.094	6.362	120.109	1.804
2015 Q1	8.759	6.856	6.126	118.099	1.697
2015 Q2	9.522	7.221	6.330	117.081	1.774
2015 Q3	9.516	7.329	6.374	113.859	1.741
2015 Q4	9.437	7.327	6.607	112.726	1.858
Napoved 2016 Q1	8.908	7.109	6.309	112.519	1.702
Napoved 2016 Q2	9.683	7.489	6.520	111.548	1.780
Napoved 2016 Q3	9.678	7.600	6.565	108.479	1.747
Napoved 2016 Q4	9.597	7.599	6.805	107.399	1.864
Napoved 2017 Q1	9.121	7.450	6.631	105.768	1.705
Napoved 2017 Q2	9.916	7.848	6.853	104.856	1.783
Napoved 2017 Q3	9.910	7.965	6.900	101.971	1.750
Napoved 2017 Q4	9.827	7.963	7.152	100.955	1.868
Napoved 2018 Q1	9.331	7.815	6.963	99.168	1.707
Napoved 2018 Q2	10.144	8.233	7.195	98.313	1.785
Napoved 2018 Q3	10.138	8.355	7.245	95.608	1.752
Napoved 2018 Q4	10.053	8.353	7.510	94.656	1.870

Tabela 10: Makroekonomske spremenljivke iz četrtnega nabora podatkov 2/3

Leto-kvartal	Zasebna_potrosnja	Investicije_v_OS	Inflacija	Cena_nafte	Export_DE
1995 Q1	2.910	1.103	117,2	13,2	106
1995 Q2	3.379	1.251	113,4	13,5	108
1995 Q3	3.460	1.196	111,2	12,4	105
1995 Q4	3.445	1.264	109,0	12,8	114
1996 Q1	3.209	1.173	108,8	13,8	111
1996 Q2	3.399	1.405	110,7	15,1	112
1996 Q3	3.537	1.347	110,1	16,2	113
1996 Q4	3.485	1.280	109,0	18,6	123
1997 Q1	3.229	1.237	108,3	18,0	117
1997 Q2	3.583	1.593	108,6	16,2	128
1997 Q3	3.636	1.524	109,7	17,1	130
1997 Q4	3.598	1.486	109,7	16,9	139
1998 Q1	3.310	1.425	109,1	13,5	135
1998 Q2	3.596	1.642	108,6	12,8	140
1998 Q3	3.759	1.587	107,5	11,7	137
1998 Q4	3.794	1.653	106,6	10,0	143
1999 Q1	3.542	1.584	105,6	10,3	134
1999 Q2	3.984	2.066	104,4	15,0	146
1999 Q3	3.911	1.715	106,8	19,7	146
1999 Q4	3.964	1.785	107,8	23,1	158
2000 Q1	3.637	1.694	108,3	26,9	158
2000 Q2	3.868	1.850	109,3	28,8	164
2000 Q3	3.997	1.885	108,6	33,7	163
2000 Q4	3.986	1.891	109,2	34,5	180
2001 Q1	3.681	1.687	108,7	28,4	174
2001 Q2	4.068	1.884	109,4	31,7	176
2001 Q3	4.046	1.973	108,4	29,0	172
2001 Q4	4.095	1.923	107,3	22,4	181
2002 Q1	3.764	1.687	108,0	24,6	173
2002 Q2	4.143	1.873	107,6	27,8	184
2002 Q3	4.160	1.938	107,2	27,2	184
2002 Q4	4.214	2.006	107,0	26,5	192
2003 Q1	3.826	1.754	106,4	28,4	180
2003 Q2	4.320	2.026	105,6	22,7	182
2003 Q3	4.335	2.070	105,5	25,1	187
2003 Q4	4.359	2.086	104,8	24,5	197
2004 Q1	3.952	1.893	103,7	25,0	200
2004 Q2	4.453	2.106	103,7	29,3	210
2004 Q3	4.447	2.208	103,6	33,3	205
2004 Q4	4.492	2.159	103,4	34,5	216
2005 Q1	4.110	1.871	102,6	36,6	208
2005 Q2	4.628	2.217	102,3	42,2	220
2005 Q3	4.426	2.239	102,5	50,9	223
2005 Q4	4.589	2.329	102,5	48,6	236
2006 Q1	4.202	2.021	102,2	52,3	238
2006 Q2	4.649	2.363	102,9	56,2	241
2006 Q3	4.491	2.503	102,5	55,7	246
2006 Q4	4.622	2.651	102,2	47,3	271

se nadaljuje

Tabela 10: Makroekonomske spremenljivke iz četrtnega nabora podatkov 2/3 (nad.)

Leto-kvartal	Zasebna_potrosnja	Investicije_v_OS	Inflacija	Cena_nafte	Export_DE
2007 Q1	4.371	2.319	102,4	44,8	266
2007 Q2	4.897	2.733	103,0	51,0	269
2007 Q3	4.904	2.824	103,6	54,2	272
2007 Q4	4.926	2.801	105,5	61,0	282
2008 Q1	4.539	2.691	106,6	64,2	279
2008 Q2	5.083	3.088	106,6	78,5	286
2008 Q3	4.967	3.017	106,1	77,6	278
2008 Q4	4.997	2.631	103,4	43,5	268
2009 Q1	4.582	2.143	101,8	35,1	231
2009 Q2	5.039	2.320	100,7	43,8	228
2009 Q3	5.043	2.296	99,8	48,1	238
2009 Q4	5.103	2.154	101,1	51,2	254
2010 Q1	4.680	1.776	101,4	56,0	252
2010 Q2	5.095	1.996	102,1	62,6	271
2010 Q3	5.053	1.951	102,1	59,6	275
2010 Q4	5.152	2.004	101,7	64,4	291
2011 Q1	4.670	1.714	101,7	77,3	290
2011 Q2	5.065	1.898	101,7	81,3	293
2011 Q3	5.081	1.875	101,3	79,3	295
2011 Q4	5.169	1.859	102,5	80,7	302
2012 Q1	4.661	1.625	102,5	90,1	304
2012 Q2	4.940	1.696	102,4	84,6	302
2012 Q3	4.883	1.713	102,9	87,3	304
2012 Q4	5.013	1.663	102,6	84,4	303
2013 Q1	4.346	1.597	102,4	85,0	300
2013 Q2	4.771	1.699	101,5	79,0	307
2013 Q3	4.665	1.720	102,1	82,5	309
2013 Q4	4.906	1.792	101,1	80,3	316
2014 Q1	4.395	1.670	100,5	78,6	313
2014 Q2	4.812	1.801	100,6	79,9	316
2014 Q3	4.698	1.834	99,8	78,0	323
2014 Q4	4.896	1.720	100,0	61,5	330
2015 Q1	4.422	1.695	99,6	49,0	328
2015 Q2	4.865	1.790	99,4	57,4	336
2015 Q3	4.814	1.798	99,6	46,1	340
2015 Q4	5.023	1.778	99,4	40,7	347
Napoved 2016 Q1	4.515	1.644	100,7	32,5	343
Napoved 2016 Q2	4.967	1.737	100,5	38,4	351
Napoved 2016 Q3	4.915	1.744	100,7	30,9	355
Napoved 2016 Q4	5.129	1.725	100,5	27,3	363
Napoved 2017 Q1	4.592	1.743	101,3	38,4	357
Napoved 2017 Q2	5.051	1.841	101,1	45,4	366
Napoved 2017 Q3	4.999	1.848	101,3	36,5	370
Napoved 2017 Q4	5.216	1.828	101,1	32,2	378
Napoved 2018 Q1	4.670	1.830	101,5	41,7	-
Napoved 2018 Q2	5.137	1.933	101,3	49,3	-
Napoved 2018 Q3	5.084	1.941	101,5	39,6	-
Napoved 2018 Q4	5.305	1.919	101,3	35,0	-

Tabela 11: Makroekonomske spremenljivke iz četrtletnega nabora podatkov 3/3

Leto-kvartal	Import_DE	DV_gradbenistvo	DV_trgovina	DV_Predelovalne	Obrestna_mera
1995 Q1	105	301	830	1.108	9,0
1995 Q2	109	353	933	1.223	8,7
1995 Q3	111	369	1.001	1.100	8,2
1995 Q4	112	347	1.013	1.126	7,8
1996 Q1	110	336	844	1.102	6,9
1996 Q2	111	395	936	1.220	6,5
1996 Q3	115	419	1.002	1.165	6,4
1996 Q4	118	362	1.001	1.277	5,8
1997 Q1	116	352	894	1.177	5,7
1997 Q2	123	422	1.043	1.331	5,7
1997 Q3	128	428	1.090	1.247	5,7
1997 Q4	129	377	1.058	1.363	5,8
1998 Q1	128	363	924	1.255	5,5
1998 Q2	135	399	1.058	1.345	5,4
1998 Q3	138	416	1.093	1.273	5,2
1998 Q4	141	408	1.101	1.344	4,8
1999 Q1	137	405	981	1.238	4,4
1999 Q2	146	495	1.150	1.362	4,1
1999 Q3	152	467	1.100	1.317	4,5
1999 Q4	155	449	1.133	1.427	5,0
2000 Q1	152	414	1.046	1.385	5,4
2000 Q2	159	442	1.165	1.505	6,1
2000 Q3	166	488	1.178	1.422	6,5
2000 Q4	178	455	1.175	1.488	6,4
2001 Q1	163	409	1.157	1.466	5,9
2001 Q2	166	438	1.258	1.553	5,7
2001 Q3	165	488	1.255	1.459	5,4
2001 Q4	166	456	1.231	1.515	4,6
2002 Q1	150	410	1.216	1.526	4,9
2002 Q2	160	448	1.329	1.632	5,2
2002 Q3	164	502	1.334	1.551	4,8
2002 Q4	169	477	1.287	1.608	4,3
2003 Q1	163	429	1.248	1.603	4,2
2003 Q2	166	497	1.356	1.680	3,9
2003 Q3	172	525	1.371	1.626	3,7
2003 Q4	179	482	1.332	1.695	3,6
2004 Q1	173	442	1.280	1.674	3,5
2004 Q2	180	480	1.399	1.801	3,5
2004 Q3	188	542	1.435	1.738	3,6
2004 Q4	193	484	1.360	1.733	3,7
2005 Q1	176	437	1.346	1.706	3,5
2005 Q2	192	526	1.491	1.904	3,6
2005 Q3	198	547	1.507	1.799	3,6
2005 Q4	210	503	1.411	1.809	3,8
2006 Q1	206	456	1.447	1.829	4,1
2006 Q2	210	561	1.599	1.985	4,3
2006 Q3	218	649	1.597	1.928	4,7
2006 Q4	229	638	1.520	1.978	5,0

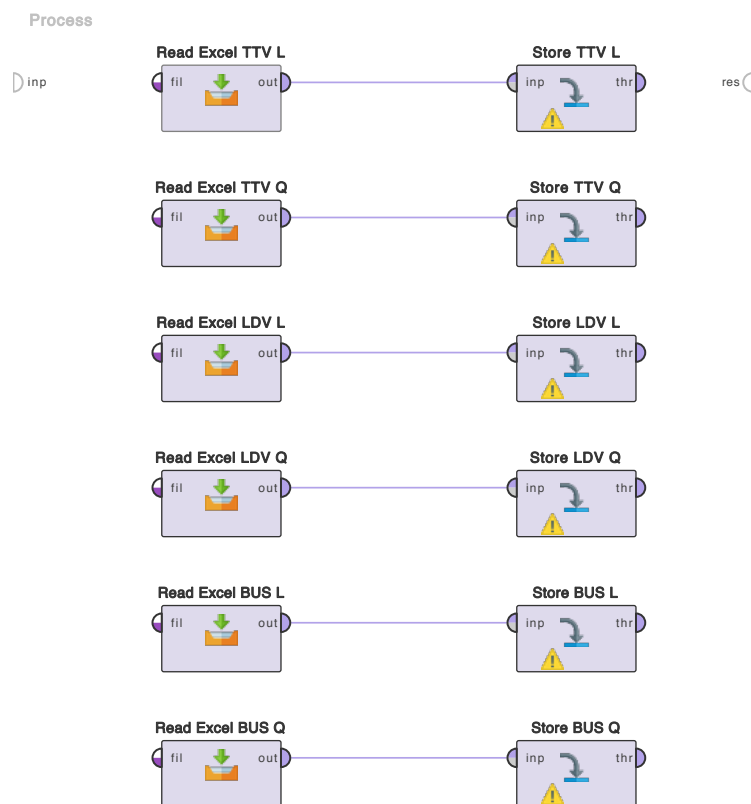
se nadaljuje

Tabela 11: Makroekonomske spremenljivke iz četrtnega nabora podatkov 3/3 (nad.)

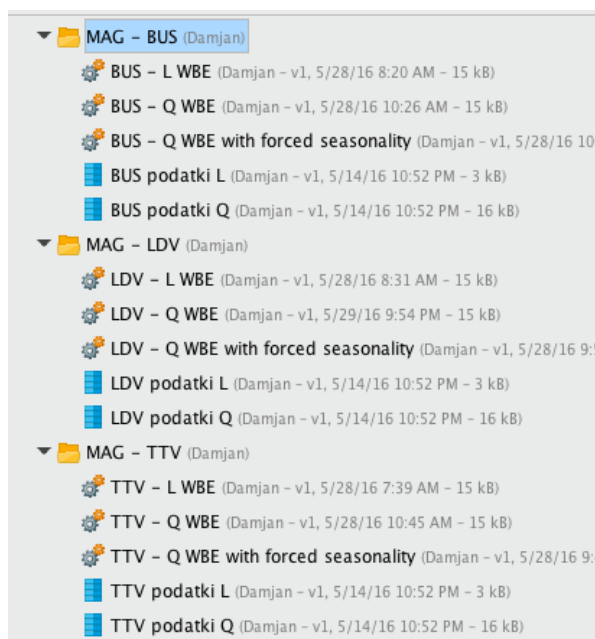
Leto-kvartal	Import_DE	DV_gradbenistvo	DV_trgovina	DV_Predelovalne	Obrestna_mera
2007 Q1	223	585	1.542	1.980	5,6
2007 Q2	224	699	1.719	2.135	5,7
2007 Q3	234	754	1.730	2.071	6,1
2007 Q4	236	680	1.607	2.112	6,3
2008 Q1	229	684	1.619	2.079	6,2
2008 Q2	233	727	1.788	2.275	6,6
2008 Q3	241	782	1.794	2.104	6,9
2008 Q4	233	654	1.626	1.924	6,9
2009 Q1	211	617	1.451	1.717	5,9
2009 Q2	202	650	1.629	1.853	6,0
2009 Q3	217	650	1.634	1.812	5,9
2009 Q4	217	547	1.509	1.801	5,9
2010 Q1	221	502	1.416	1.720	5,8
2010 Q2	239	521	1.652	2.032	5,5
2010 Q3	246	528	1.699	1.949	5,8
2010 Q4	250	464	1.533	1.951	5,7
2011 Q1	248	429	1.422	1.857	5,7
2011 Q2	254	451	1.696	2.145	5,8
2011 Q3	262	489	1.748	1.944	5,9
2011 Q4	259	444	1.542	1.887	6,0
2012 Q1	253	413	1.397	1.853	5,9
2012 Q2	251	424	1.622	2.065	5,7
2012 Q3	260	451	1.653	1.871	5,7
2012 Q4	255	384	1.473	1.838	5,7
2013 Q1	252	336	1.331	1.811	5,8
2013 Q2	259	369	1.633	2.017	5,7
2013 Q3	272	414	1.677	1.873	5,8
2013 Q4	267	407	1.501	1.916	5,6
2014 Q1	264	386	1.370	1.879	5,3
2014 Q2	268	439	1.690	2.109	5,2
2014 Q3	279	453	1.743	1.967	5,0
2014 Q4	279	394	1.556	2.023	4,6
2015 Q1	280	383	1.425	1.989	4,1
2015 Q2	282	416	1.762	2.205	3,6
2015 Q3	296	426	1.821	2.071	3,3
2015 Q4	296	392	1.622	2.116	3,2
Napoved 2016 Q1	295	339	1.470	2.054	3,1
Napoved 2016 Q2	298	368	1.818	2.278	3,1
Napoved 2016 Q3	312	377	1.880	2.139	3,1
Napoved 2016 Q4	312	347	1.674	2.186	3,1
Napoved 2017 Q1	309	352	1.510	2.126	3,1
Napoved 2017 Q2	312	383	1.867	2.357	3,1
Napoved 2017 Q3	327	392	1.930	2.214	3,1
Napoved 2017 Q4	327	361	1.719	2.262	3,1
Napoved 2018 Q1	-	363	1.549	2.201	3,1
Napoved 2018 Q2	-	394	1.916	2.440	3,1
Napoved 2018 Q3	-	403	1.981	2.291	3,1
Napoved 2018 Q4	-	372	1.764	2.341	3,1

Priloga 3: Procesi uporabljeni v RapidMiner Studiu

Slika 1: Shema procesa za zajem podatkov v repozitorij RapidMiner iz Excel datotek



Slika 2: Prikaz procesov in podatkov iz repozitorija RapidMiner



*Slika 3: Zapis uporabljenega procesa modeliranja v PMML formatu
na primeri za letni nabor podatkov za segment avtobusov*

```
<?xml version="1.0" encoding="UTF-8" standalone="no"?>
<process version="7.1.000">
  <context>
    <input/>
    <output/>
    <macros/>
  </context>
  <operator activated="true" class="process" compatibility="7.1.000" expanded="true"
name="Process">
    <parameter key="logverbosity" value="init"/>
    <parameter key="random_seed" value="2001"/>
    <parameter key="send_mail" value="never"/>
    <parameter key="notification_email" value=""/>
    <parameter key="process_duration_for_mail" value="30"/>
    <parameter key="encoding" value="SYSTEM"/>
    <process expanded="true">
      <operator activated="true" class="retrieve" compatibility="7.1.000" expanded="true"
height="68" name="Retrieve" width="90" x="45" y="34">
        <parameter key="repository_entry" value="//Local Repository/MAG - BUS/BUS podatki L"/>
      </operator>
      <operator activated="true" class="set_role" compatibility="7.1.000" expanded="true"
height="82" name="Set Label and ID" width="90" x="45" y="136">
        <parameter key="attribute_name" value="Leto"/>
        <parameter key="target_role" value="id"/>
        <list key="set_additional_roles">
          <parameter key="Prva_reg_novi" value="label"/>
        </list>
      </operator>
      <operator activated="true" class="select_attributes" compatibility="7.1.000" expanded="true"
height="82" name="Select Attributes" width="90" x="45" y="289">
        <parameter key="attribute_filter_type" value="subset"/>
        <parameter key="attribute" value=""/>
        <parameter key="attributes"
value="Prva_reg_novi|BDP|Cena_nafte|DV_gradbenistvo|DV_predelovalne|DV_trgovina|Drzavna_potrosnja|E
xport_DE|Import_DE|Inflacija|Investicije_v_OS|Izvoz|Obrestna_mera|Reg_brezposelnost|Uvoz|Zasebna_po
trosnja|DV_Predelovalne|Leto"/>
        <parameter key="use_except_expression" value="false"/>
        <parameter key="value_type" value="attribute_value"/>
        <parameter key="use_value_type_exception" value="false"/>
        <parameter key="except_value_type" value="time"/>
        <parameter key="block_type" value="attribute_block"/>
        <parameter key="use_block_type_exception" value="false"/>
        <parameter key="except_block_type" value="value_matrix_row_start"/>
        <parameter key="invert_selection" value="false"/>
        <parameter key="include_special_attributes" value="false"/>
      </operator>
      <operator activated="true" class="filter_example_range" compatibility="7.1.000"
expanded="true" height="82" name="Filter Training Range" width="90" x="246" y="187">
        <parameter key="first_example" value="1"/>
        <parameter key="last_example" value="21"/>
        <parameter key="invert_filter" value="false"/>
      </operator>
      <operator activated="true" class="filter_example_range" compatibility="7.1.000"
expanded="true" height="82" name="Filter Prediction Range" width="90" x="380" y="187">
        <parameter key="first_example" value="1"/>
      </operator>
    </process>
  </operator>
</process>
```

se nadaljuje

*Slika 3: Zapis uporabljenega procesa modeliranja v PMML formatu
na primeri za letni nabor podatkov za segment avtobusov (nad.)*

```

    <parameter key="last_example" value="24"/>
    <parameter key="invert_filter" value="false"/>
  </operator>
  <operator activated="true" class="set_role" compatibility="7.1.000" expanded="true"
height="82" name="Set Role" width="90" x="514" y="187">
    <parameter key="attribute_name" value="Prva_reg_novi"/>
    <parameter key="target_role" value="regular"/>
    <list key="set_additional_roles"/>
  </operator>
  <operator activated="true" class="optimize_selection_backward" compatibility="7.1.000"
expanded="true" height="103" name="Backward Elimination" width="90" x="380" y="34">
    <parameter key="maximal_number_of_eliminations" value="20"/>
    <parameter key="speculative_rounds" value="100"/>
    <parameter key="stopping_behavior" value="with decrease of more than"/>
    <parameter key="use_relative_decrease" value="true"/>
    <parameter key="maximal_relative_decrease" value="0.05"/>
    <parameter key="alpha" value="0.05"/>
    <process expanded="true">
      <operator activated="true" class="split_validation" compatibility="7.1.000"
expanded="true" height="124" name="Validation" width="90" x="380" y="34">
        <parameter key="create_complete_model" value="false"/>
        <parameter key="split" value="absolute"/>
        <parameter key="split_ratio" value="0.7"/>
        <parameter key="training_set_size" value="15"/>
        <parameter key="test_set_size" value="6"/>
        <parameter key="sampling_type" value="shuffled sampling"/>
        <parameter key="use_local_random_seed" value="false"/>
        <parameter key="local_random_seed" value="1992"/>
      </process expanded="true">
      <operator activated="true" class="linear_regression" compatibility="7.1.000"
expanded="true" height="103" name="Linear Regression (2)" width="90" x="112" y="34">
        <parameter key="feature_selection" value="none"/>
        <parameter key="alpha" value="0.05"/>
        <parameter key="max_iterations" value="10"/>
        <parameter key="forward_alpha" value="0.05"/>
        <parameter key="backward_alpha" value="0.05"/>
        <parameter key="eliminate_colinear_features" value="true"/>
        <parameter key="min_tolerance" value="0.05"/>
        <parameter key="use_bias" value="true"/>
        <parameter key="ridge" value="1.0E-8"/>
      </operator>
      <connect from_port="training" to_op="Linear Regression (2)" to_port="training set"/>
      <connect from_op="Linear Regression (2)" from_port="model" to_port="model"/>
      <portSpacing port="source_training" spacing="0"/>
      <portSpacing port="sink_model" spacing="0"/>
      <portSpacing port="sink_through 1" spacing="0"/>
    </process>
    <process expanded="true">
      <operator activated="true" class="apply_model" compatibility="7.1.000"
expanded="true" height="82" name="Apply Model" width="90" x="45" y="34">
        <list key="application_parameters"/>
        <parameter key="create_view" value="false"/>
      </operator>
      <operator activated="true" class="performance_regression" compatibility="7.1.000"
expanded="true" height="82" name="Performance" width="90" x="179" y="34">

```

se nadaljuje

Slika 3: Zapis uporabljenega procesa modeliranja v PMML formatu na primeri za letni nabor podatkov za segment avtobusov (nad.)

```

    <parameter key="main_criterion" value="squared_correlation"/>
    <parameter key="root_mean_squared_error" value="true"/>
    <parameter key="absolute_error" value="false"/>
    <parameter key="relative_error" value="true"/>
    <parameter key="relative_error_lenient" value="false"/>
    <parameter key="relative_error_strict" value="false"/>
    <parameter key="normalized_absolute_error" value="false"/>
    <parameter key="root_relative_squared_error" value="false"/>
    <parameter key="squared_error" value="false"/>
    <parameter key="correlation" value="true"/>
    <parameter key="squared_correlation" value="true"/>
    <parameter key="prediction_average" value="false"/>
    <parameter key="spearman_rho" value="false"/>
    <parameter key="kendall_tau" value="false"/>
    <parameter key="skip_undefined_labels" value="true"/>
    <parameter key="use_example_weights" value="true"/>
  </operator>
  <connect from_port="model" to_op="Apply Model" to_port="model"/>
  <connect from_port="test set" to_op="Apply Model" to_port="unlabelled data"/>
  <connect from_op="Apply Model" from_port="labelled data" to_op="Performance"
to_port="labelled data"/>
  <connect from_op="Performance" from_port="performance" to_port="averagable 1"/>
  <portSpacing port="source_model" spacing="0"/>
  <portSpacing port="source_test set" spacing="0"/>
  <portSpacing port="source_through 1" spacing="0"/>
  <portSpacing port="sink_averagable 1" spacing="0"/>
<portSpacing port="sink_averagable 2" spacing="0"/>
  </process>
</operator>

  <connect from_port="example set" to_op="Validation" to_port="training"/>
  <connect from_op="Validation" from_port="averagable 1" to_port="performance"/>
  <portSpacing port="source_example set" spacing="0"/>
  <portSpacing port="sink_performance" spacing="0"/>
  </process>
</operator>

  <operator activated="true" class="linear_regression" compatibility="7.1.000" expanded="true"
height="103" name="Linear Regression" width="90" x="514" y="34">
    <parameter key="feature_selection" value="none"/>
    <parameter key="alpha" value="0.05"/>
    <parameter key="max_iterations" value="10"/>
    <parameter key="forward_alpha" value="0.05"/>
    <parameter key="backward_alpha" value="0.05"/>
    <parameter key="eliminate_colinear_features" value="true"/>
    <parameter key="min_tolerance" value="0.05"/>
    <parameter key="use_bias" value="true"/>
    <parameter key="ridge" value="1.0E-8"/>
  </operator>

  <operator activated="true" class="apply_model" compatibility="7.1.000" expanded="true"
height="82" name="Apply Model (2)" width="90" x="648" y="136">
    <list key="application_parameters"/>
    <parameter key="create_view" value="false"/>
  </operator>

  <operator activated="true" class="set_role" compatibility="7.1.000" expanded="true"
height="82" name="Set Role (2)" width="90" x="648" y="340">

```

se nadaljuje

*Slika 3: Zapis uporabljenega procesa modeliranja v PMML formatu
na primeri za letni nabor podatkov za segment avtobusov (nad.)*

```

<parameter key="attribute_name" value="Prva_reg_novi"/>
<parameter key="target_role" value="label"/>
<list key="set_additional_roles"/>
</operator>
<operator activated="true" class="performance_regression" compatibility="7.1.000"
expanded="true" height="82" name="Performance (2)" width="90" x="782" y="340">
  <parameter key="main_criterion" value="first"/>
  <parameter key="root_mean_squared_error" value="true"/>
  <parameter key="absolute_error" value="false"/>
  <parameter key="relative_error" value="true"/>
  <parameter key="relative_error_lenient" value="false"/>
  <parameter key="relative_error_strict" value="false"/>
  <parameter key="normalized_absolute_error" value="false"/>
  <parameter key="root_relative_squared_error" value="false"/>
  <parameter key="squared_error" value="false"/>
  <parameter key="correlation" value="true"/>
  <parameter key="squared_correlation" value="true"/>
  <parameter key="prediction_average" value="false"/>
  <parameter key="spearman_rho" value="false"/>
  <parameter key="kendall_tau" value="false"/>
  <parameter key="skip_undefined_labels" value="true"/>
  <parameter key="use_example_weights" value="true"/>
</operator>
<operator activated="true" class="rename" compatibility="7.1.000" expanded="true" height="82"
name="Rename" width="90" x="782" y="187">
  <parameter key="old_name" value="Prva_reg_novi"/>
  <parameter key="new_name" value="Dejanska_Prva_reg_novi"/>
  <list key="rename_additional_attributes">
    <parameter key="prediction(Prva_reg_novi)" value="Ocena_Prva_reg_novi"/>
  </list>
</operator>
<operator activated="true" class="generate_attributes" compatibility="7.1.000"
expanded="true" height="82" name="Generate Attributes" width="90" x="916" y="187">
  <list key="function_descriptions">
    <parameter key="Napaka" value="[Ocena_Prva_reg_novi]-[Dejanska_Prva_reg_novi]"/>
  </list>
  <parameter key="keep_all" value="true"/>
</operator>
<operator activated="true" class="order_attributes" compatibility="7.1.000" expanded="true"
height="82" name="Reorder Attributes" width="90" x="916" y="340">
  <parameter key="sort_mode" value="user specified"/>
  <parameter key="sort_direction" value="none"/>
  <parameter key="attribute_ordering" value="Dejanska_Prva_reg_novi|Napaka"/>
  <parameter key="handle_unmatched" value="append"/>
  <parameter key="use_regular_expressions" value="false"/>
</operator>
<operator activated="true" class="write_excel" compatibility="7.1.000" expanded="true"
height="82" name="Write Excel" width="90" x="1050" y="187">
  <parameter key="file_format" value="xlsx"/>
  <parameter key="encoding" value="SYSTEM"/>
  <parameter key="sheet_name" value="RapidMiner Data"/>
  <parameter key="date_format" value="yyyy-MM-dd HH:mm:ss"/>
  <parameter key="number_format" value="#.0"/>

```

se nadaljuje

*Slika 3: Zapis uporabljenega procesa modeliranja v PMML formatu
na primeri za letni nabor podatkov za segment avtobusov (nad.)*

```

</operator>
<operator activated="true" class="write_file" compatibility="7.1.000" expanded="true"
height="68" name="Write File" width="90" x="1050" y="340">
  <parameter key="resource_type" value="file"/>
  <parameter key="filename" value="/Users/Damjan/Documents/Dropbox/M-RM Data/Export BUS L
WBE.xlsx"/>
  <parameter key="mime_type" value="application/octet-stream"/>
</operator>
<connect from_op="Retrieve" from_port="output" to_op="Set Label and ID" to_port="example set
input"/>
<connect from_op="Set Label and ID" from_port="example set output" to_op="Select Attributes"
to_port="example set input"/>
<connect from_op="Select Attributes" from_port="example set output" to_op="Filter Training
Range" to_port="example set input"/>
<connect from_op="Filter Training Range" from_port="example set output" to_op="Backward
Elimination" to_port="example set"/>
<connect from_op="Filter Training Range" from_port="original" to_op="Filter Prediction Range"
to_port="example set input"/>
<connect from_op="Filter Prediction Range" from_port="example set output" to_op="Set Role"
to_port="example set input"/>
<connect from_op="Set Role" from_port="example set output" to_op="Apply Model (2)"
to_port="unlabelled data"/>
<connect from_op="Backward Elimination" from_port="example set" to_op="Linear Regression"
to_port="training set"/>
<connect from_op="Backward Elimination" from_port="attribute weights" to_port="result 1"/>
<connect from_op="Backward Elimination" from_port="performance" to_port="result 2"/>
<connect from_op="Linear Regression" from_port="model" to_op="Apply Model (2)"
to_port="model"/>
<connect from_op="Apply Model (2)" from_port="labelled data" to_op="Set Role (2)"
to_port="example set input"/>
<connect from_op="Apply Model (2)" from_port="model" to_port="result 3"/>
<connect from_op="Set Role (2)" from_port="example set output" to_op="Performance (2)"
to_port="labelled data"/>
<connect from_op="Performance (2)" from_port="performance" to_port="result 5"/>
<connect from_op="Performance (2)" from_port="example set" to_op="Rename" to_port="example
set input"/>
<connect from_op="Rename" from_port="example set output" to_op="Generate Attributes"
to_port="example set input"/>
<connect from_op="Generate Attributes" from_port="example set output" to_op="Reorder
Attributes" to_port="example set input"/>
<connect from_op="Reorder Attributes" from_port="example set output" to_op="Write Excel"
to_port="input"/>
<connect from_op="Write Excel" from_port="through" to_port="result 4"/>
<connect from_op="Write Excel" from_port="file" to_op="Write File" to_port="file"/>
<portSpacing port="source_input 1" spacing="0"/>
<portSpacing port="sink_result 1" spacing="0"/>
<portSpacing port="sink_result 2" spacing="0"/>
<portSpacing port="sink_result 3" spacing="0"/>
<portSpacing port="sink_result 4" spacing="0"/>
<portSpacing port="sink_result 5" spacing="0"/>
<portSpacing port="sink_result 6" spacing="0"/>
</process>
</operator>
</process>

```