

# Napovedni modeli razvoja gozdov v Sloveniji

Marko Debeljak <sup>1\*</sup>, Bernard Ženko <sup>1</sup>, Aleš Poljanec <sup>2</sup>

<sup>1</sup> Institut Jožef Stefan, Odsek za tehnologije znanja

<sup>2</sup> Univerza v Ljubljani, Biotehniška fakulteta

\* Jamova 39, 1000 Ljubljana, tel.: 01 4773124, E-mail: [marko.debeljak@ijs.si](mailto:marko.debeljak@ijs.si)

## Uvod

Na področju analize ekoloških podatkov se vedno bolj kažejo potrebe po razvoju in uporabi metod in orodij, ki bi vključevale sodobne pristope strojnega učenja in informacijske teorije ter bi se dopolnjevale s klasičnimi statističnimi metodami. Pričakovana kompatibilnost bi omogočila izpolnitev večjega številnih zahtev glede značilnosti sodobnih orodij za analizo ekoloških podatkov in zagotovila višjo kvaliteto rezultatov analiz podatkov. Podatkovno rudarjenje tako npr. uporablja metode strojnega učenja, ki v veliki meri uspešno vključujejo tako pristope klasične statistike kot tudi informacijske teorije. Orodja strojnega učenja se uspešno uporabljajo za analizo podatkov, odkrivanja zakonitosti v podatkih ter za gradnjo kvantitativnih in kvalitativnih modelov. Z metodami strojnega učenja se lahko iz podatkov avtomatsko naučimo zakonitosti, ki v njih veljajo ali zgradimo model obravnavanega sistema. Modele lahko zapišemo v človeku razumljivi obliki (npr. v obliki pravil, odločitvenih dreves, enačb) ali pa v obliki, ki je uporabna samo za napovedovanje novih primerov (npr. nevronske mreže, modeli podpornih vektorjev, itd.). Dosedanje izkušnje so pokazale, da so za analizo ekoloških podatkov še posebej primerni modeli v obliki odločitvenih dreves, ki imajo hierarhično strukturo, in napovedujejo vrednosti odvisne spremenljivke, pri čemer lahko napovedujejo vrednost ene ali več odvisnih spremenljivk hkrati. Raziskava obravnava presojo primernosti uporabe metod podatkovnega rudarjenja za analizo in izdelavo scenarijev razvoja gozdov v Sloveniji, pri čemer smo se osredotočili na analizo časovne dinamike spreminjanja lesne zaloge, pojasnjevanje in napovedovanje skupne lesne zaloge.

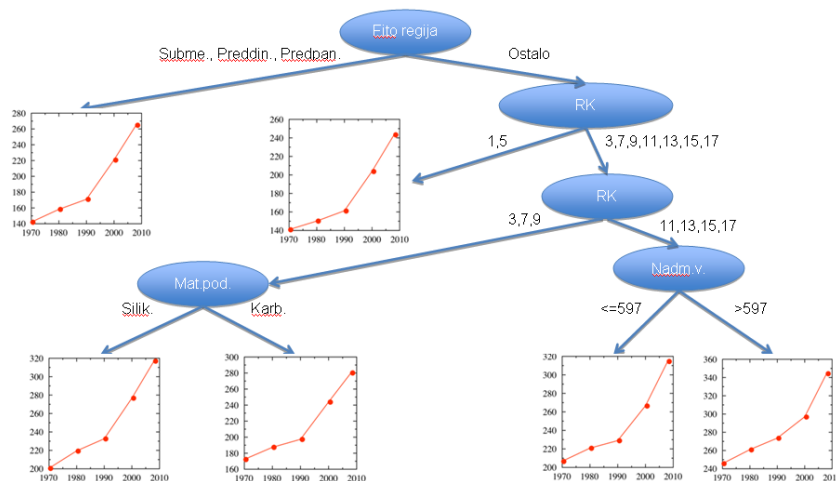
## Metode dela

Izhodišče raziskave predstavlja podatkovna zbirka *Silva-SI* (Poljanec, 2008). Zbirka vsebuje podatke za 21052 oddelkov za pet obdobji od 1970 do 2008 (1970, 1980, 1990, 2000 in 2008). Za potrebe analize smo glede na lastniški delež države v oddelku, oblikovali skupini »državnih« gozdov, z 100 % lastniškim deležem države (5237 oddelkov) in »zasebnih« gozdov, kjer je lastniški delež države nižji od 100 % (15815 oddelkov). Časovno dinamiko spreminjanja lesne zaloge oddelkov v celotnem preučevanem obdobju (1970 – 2008) smo modelirali s pomočjo odločitvenih dreves za napovedno razvrščanje (Predictive Clustering Trees - PCT), ki smo jih gradili s pomočjo orodja CLUS (Blokkeel & Struyf, 2002). Pojasnjevalni in napovedni modeli (napoved za leto 2018) za skupno lesno zalogo po oddelkih so imeli

obliko modelnih odločitvenih dreves zgrajenih z orodjem WEKA (Witten & Frank 2005). Vse tri tipe modelov smo zgradili na osnovi podatkov za celotno lesno zalogo in za posamezne razširjene debelinske stopnje A, B in C. Ker v podatkovni zbirki *Silva-SI* število oddelkov »zasebnih« gozdov močno prevladuje, v nadaljevanju podajamo rezultate modelov celotne lesne zaloge le za to skupino gozdov.

## Rezultati

Struktura modela časovne dinamike (Slika 1) pokaže, da ima na obliko trenda lesne zaloge zasebnih gozdov največji vpliv fitogeografska regija, sledi rastiščni koeficient, medtem ko sta na zadnjem, najmanj vplivnem hierarhičnem nivoju matična podlaga in nadmorska višina. Trendi lesne zaloge so vseskozi naraščajoči. Vrednosti napak se od prvega obdobja (1970) zmanjšujejo tako, da je povprečna absolutna napaka (MAE) modelirane vrednosti za leto 2008 boljša za 12 %, relativna povprečna kvadratna napaka (RMSE) pa za 22 %. Koeficient korelacije ostaja za vse napovedi členov časovne vrste približno enak.

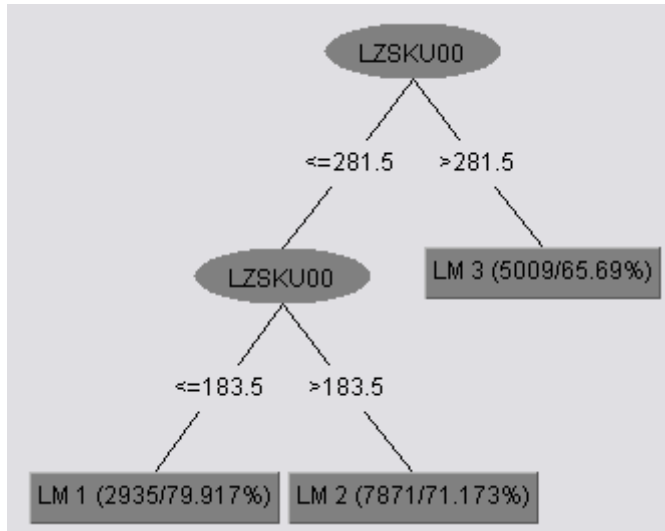


Slika 1: Model časovne dinamike lesne zaloge v »zasebnih« gozdovih.

Pojasnjevalni modeli skupne lesne zaloge po oddelkih za posamezna preučevana obdobja 1970-2008 kažejo, da na lesno zalogo preučevanega obdobja najmočneje vpliva skupna lesna zaloga preteklega obdobja. Opozoriti velja, da pri modeliranju lesne zaloge leta 1970 nismo razpolagali s podatki o lesni zalogi leta 1960, zato se v tem modelu atribut "pretekla lesna zaloga" ne pojavlja, kot najpomembnejši atribut pa nastopa fitogeografska regija. Zaradi zelo nizkega koeficienta korelacije je model za leto 1970 nereprezentativen.

Višino lesne zaloge za leto 2018 smo izračunali s pomočjo ekstrapolacije trenda lesne zaloge posameznih linearnih modelov odločitvenega drevesa, ki smo ga zgradili za leto 2008 (Slika 2). V prvi fazi izgradnje napovednih modelov smo na primeru modela za leto 2008 izvedli verifikacijo ekstrapolacijske metode. Pri tem smo za izhodišče uporabili model lesne zaloge za leto 2000, kateremu smo na osnovi podatkov za leto 1990 in 2000 izračunali ekstrapolirano vrednost trenda za leto 2008. Rezultate modelnih napovedi smo

preverili (verificirali) z obstoječimi dejanskimi vrednostmi za leto 2008. Ker so bila odstopanja med napovedanimi in dejanskimi vrednostmi zelo majhna (največje odstopanje je bilo 9,2 %), smo metodologijo prenesli v drugo fazo izgradnje napovednega modela. Za izhodišče smo uporabili model za leto 2008, kjer smo na osnovi vrednosti za leto 2000 in 2008 izračunali vrednosti ekstrapolacijske funkcije za leto 2018.



Slika 1: Model lesne zaloge »zasebnih« gozdov za leto 2008.

Vrednosti linearnih ekstrapolacijskih modelov LM1, LM2 in LM3 za leto 2018, kažejo na nadaljnje povečevanje lesne zaloge ob predpostavki, da se gospodarjenje z gozdovi v prihodnjem desetletju ne bo bistveno spremenilo. Največje povečanje lahko pričakujemo v oddelkih z nižjo lesno zalogo v letu 2008, najmanjše povečanje pa v oddelkih z visoko lesno zalogo (Preglednica 1).

Preglednica 1: Napovedane vrednosti lesne zaloge »zasebnih« gozdov za leto 2018.

	LM1	LM2	LM3
Ekstrapolacija regresijske funkcije; $y=kx+n$ (k, n)	9,48; - 18815	5,84; -11464	1,83; -3336,6
Povprečna dejanska vrednost lesne zaloge v letu 2008	224,4	280,0	351,7
Napovedna (modelna) vrednost lesne zaloge v letu 2008	224,5	279,6	351,7
Napovedna (modelna) vrednost lesne zaloge v letu 2018	319,3	338,1	370,1
Indeks rasti (leto 2008 =100)	142,3	120,8	105,2

## Zaključek

Na osnovi dobljenih rezultatov lahko zaključimo, da podatkovna zbirka *Silva-SI* omogoča zanesljivo uporabo metod podatkovnega rudarjenja za izdelavo modelov časovne dinamike lesne zaloge, modelov za razlago stanja lesne zaloge po posameznih obdobjih in za izgradnjo napovednih modelov lesne zaloge. Uporabljen pristop omogoča dinamično modeliranje, ki optimizira strukturo modela glede na značilnosti

podatkov, omogoča oblikovanje podskupin podatkov in določitev najvplivnejših dejavnikov in povezave med njimi (odkrivanje direktnih in indirektnih vplivov) in tako prispeva k pridobivanju novega znanja iz podatkov. Z uporabo kvalitativnega odločitvenega modeliranja bi rezultate podatkovnega rudarjenja lahko povezali z obstoječim ekspertnim znanjem ter gozdarstvu ponudili učinkovit in uporabniku prijazen sistem za podporo odločanju o prihodnjem ravnanju z gozdovi.

## **Literatura**

Blockeel, H., Struyf, J. 2002. Efficient algorithms for decision tree cross-validation. *Journal of Machine Learning Research*, 3:621–650.

Poljanec, A. 2008. Strukturne spremembe gozdnih sestojev v Sloveniji v obdobju 1970-2005. Doktorska disertacija, Ljubljana: 126 s.

Witten, I.H., Frank, E. 2005. *Data Mining: Practical Machine Learning Tools and Techniques*. Morgan Kaufmann, San Francisco, CA, USA, second edition.