



Univerza v Ljubljani
Fakulteta za računalništvo
in informatiko



Univerza v Ljubljani
Fakulteta za družbene vede

Poročilo Delovnega sklopa 1 (tehnični pregled zbirk), Delovnega sklopa 2 Metodologija izgradnje analitičnih zbirk, Delovnega sklopa 4 Demonstracijski primeri in Delovnega sklopa 5 (zbirka podatkov)

Ciljni raziskovalni projekt (CRP) Zasnova informacijski rešitev za odločanje o SKP na podlagi podatkov (ITzaSKP)

Avtorji: dr. Blaž Zupan, mag. Maša Kerstein, Vesna Tanko, Andreja Kovačič, dr. Janez Štebe, Brigita Bočkaj

Uredil: dr. Marko Lovec

Ljubljana, november 2020

Kazalo vsebine

Uvod.....	3
Delovne naloge 1.3, 1.4, 2.1 in 2.2/4.2 (prof. Blaž Zupan)	4
Delovna naloga 1.3: Tehnično-informacijske značilnosti in združljivost podatkovnih baz	4
Uporabljene tehnologije	4
Struktura podatkov	4
Uporaba ključev	4
Odprtost baze, dostopnost podatkov	5
Programski vmesniki, programski dostop.....	5
Spremljajoča analitična orodja.....	5
Dokumentacija	6
Delovna naloga 1.4.: Identifikacija ključnih stvarnih pomanjkljivosti.....	6
Ključne pomankljivosti	6
Izkušnje s podatkovno analitiko podatkov MKGP v projektu CRP	7
Delovna naloga 2.1 in 2.2.: Gradnja podatkovnih zbirk, predlogi in izboljšave	10
Priložnosti.....	12
Delovne naloge 4.1 in 4.2.....	13
Eksperiment 1: Atributivna večnamenska zbirka podatkov o KMG za obdobje 2015-2019 (mag. Maša Kerstein, Vesna Tanko).....	13
Oblikovanje večletne podatkovne zbirke za namen reševanja analitičnih nalog	14
Odločitev o izbiri naloge za prikaz.....	15
Analiza podatkov za kmetije »mladi kmet«	16
Analiza vseh kmetij	22
Zaključek	29
Eksperiment 2: Analiza napovedljivosti opuščanja kmetijskih površin (Andreja Kovačič)	31
Podatkovne zbirke.....	31
Obdelava podatkov	31
Modeliranje.....	33
Vrednotenje	34
Možne izboljšave.....	36
Delovna naloga 5.1: opis in arhiviranje podatkovnih zbirk (Brigita Bočkaj, dr. Janez Štebe)	37
Metapodatkovni opis izbranih evidenc.....	37
Prezem podatkovnih zbirk ARSKTRP	40
Priloga: predlogi demonstracijskih nalog.....	44

Uvod

Pričujoče poročilo je del Ciljnega raziskovalnega projekta (CRP) Zasnova informacijskih rešitev v podporo odločanju o Skupni kmetijski politiki (SKP) na podlagi podatkov (ITzaSKP). Nanaša se na informacijsko-tehnični pregled administrativnega podatkovnega sistema (del nalog DS1), metodologijo za izgradnjo analitičnih podatkovnih zbirk (DS2) in izvedbo demonstracijskih nalog (DS4) – z vidika podatkovne znanosti.

Informacijsko-tehnični del projekta temelji na eksperimentalnem pristopu, katerega namen je bil na konkretnem primeru oceniti stanje in potrebne vložke za smiselno ciljno in splošno uporabo (naprednih) podatkovnih metod in tehnik ter podati ustrezne predloge.

V okviru DS1 je bil opravljen pregled administrativnih registrov, evidenc in zbirk MKGP in organov v sestavi na podlagi dokumentacije in podatkovnih intervjujev, ki je omogočil splošno oceno stanja z družboslovnega vidika podatkovne znanosti. Register kmetijskih gospodarstev (RKG) je bil prepoznan kot osrednji register podatkov, predvsem z vidika rabe površin (GERKi, raba), vključno z okoljsko-prostorskimi sloji, in subjektov. Prepoznana je bila možnost oblikovanja statistične zbirke v okviru podatkovnega sistema Agencije za kmetijske trge in razvoj podeželja (AKTRP), ki združuje številne atributivne podatke o KMG-zavezancih z vidika pristojnosti izvajanja ukrepov in izplačil. Kot relevantni so bili med drugim prepoznani tudi registri živali v reji v okviru Uprave za varno hrano, veterinarstvo in varovanje rastlin (UVHVVR).

V okviru DS2 je bila razvita splošna metodologija za izgradnjo zbirk za analitično delo. Temeljila je na indikativnih nalogah, ki so jih oblikovali člani projektne skupine (Priloga, Tabela 1 in 2) ter na možnostih prepoznanih v okviru DS1. Na podlagi tega je bila testno oblikovana (a) atributivna statistična zbirka mikro podatkov o 57.000 KMG za obdobje enega leta prevzetih s strani AKTRP za podatkovno-statistične analize (profiliranje, korelacije, napovedovanje-vzročnost) in prostorski prikaz na zemljevidu, v okolju Orange, ter (b) idejna zasnova ciljne izgradnje zbirke prostorskih podatkov za ugotavljanje sprememb v (kmetijski) rabi prostora (raba, GERKi) z nekaterimi atributivnimi podatki (KMG-subjekti, stalež, ukrepi) .

V okviru DS3 je potekalo posvetovanje z različnimi deležniki, kjer je bila med drugim na javnem dogodku junija 2019 predstavljena letna zbirka atributivnih podatkov o KMG prevzeta s strani AKTRP in možnosti v okolju Orange. Sprejet je bil dogovor o odprtem pozivu po nalogah za namen demonstracije v okviru DS4 na katerega se je odzvalo več predlagateljev s strani MKGP in AKTRP. Pregled zbranih pobud/potreb (Priloga, Tabela 3) je pokazal na številne in raznolike potrebe preprostih pregledov, statističnih analiz in vizualizacije v zvezi z različnimi atributivnimi podatki. Med kompleksnejšimi nalogami so izstopale prostorske spremembe kot je zaraščanje, nekateri dejavniki v zvezi s tem in predvidevanje.

Na podlagi izsledkov DS3 je bila sprejeta odločitev, da se v okviru DS4 (demonstracijske naloge) razširi atributivno zbirko podatkov o KMG v okolju Orange, ki je skladno z izraženimi potrebami omogočala raznolike statistične analize na statični zbirki s številnimi atributi ter vizualni izpis na zemljevidu, na dostopne podatke za obdobje izvajanja trenutne SKP (2015-2019), kar bi omogočilo smiselno ugotavljanje časovnih trendov v zaključenem obdobju izvajanja ukrepov. Kot drugo pa je bil sprejet dogovor o zbirki prostorskih podatkov (raba, GERKi), ter nekaterih atributivnih podatkov o KMG (subjekti, stalež, ukrepi) iz različnih registrov in evidenc za daljše obdobje za ugotavljanje sprememb v prostoru in dejavnikov na ravni različnih (prostorskih) agregacij.

Delovne naloge 1.3, 1.4, 2.1 in 2.2/4.2 (prof. Blaž Zupan)

Delovna naloga 1.3: Tehnično-informacijske značilnosti in združljivost podatkovnih baz

Cilj delovne naloge 1.3 je ocena stanja tehničnih in informacijskih rešitev na področjih, ki jih pokriva Ministrstvo za kmetijstvo, gozdarstvo in prehrano (MKGP) s stališča možnosti uvajanja tehnologij napredne podatkovne analitike in uvajanja sistemov za podporo odločanja. Stanje je bilo ocenjeno v neformalnih pogovorih z ustreznimi službami ministrstva, strukturiranim razgovorom z vodjem službe za informacijsko tehnologijo in ocene na osnovi pridobljenih izkušenj, ki smo jih tekom projekta pridobili pri poskusih črpanja in urejanja podatkov in pilotnega načrtovanja shem za podatkovno analitiko. V tem delu poročila podajamo opažanja, ki izhajajo iz teh pogovorov in predvsem iz strukturiranega razgovora, opažanja, zaključke in predloge pa potem strnemo v sintezi, o kateri poročamo v delu poročila, ki obravnava Delovno nalogo 1.4.

Uporabljene tehnologije

Informacijski sistem oziroma elementi podatkovnega jezera, v katerem MKGP shranjuje in obdeluje podatke, temelji na plačljivih in licenčnih tehnologijah, kot sta Oracle in Excel. Baze podatkov MKGP so skrbno načrtovane, dobro vzdrževane, a kot take razpršene in vsebujejo različne evidence in registre, mnogokrat kot odziv na zahteve EU po zbiranju in sistematizaciji podatkov s področja kmetijstva in v, na primer, podporo tekočim zakonskim uredbam, politikam EU in podpori črpanja namenskih sredstev. Tudi dostop do aplikacij je namenski, večinoma spleten in zgrajen z uporabo licenčnih tehnologij (npr. Oracle Application Express). Aplikacije se največkrat, z izjemo nekaj pomembnih aplikacij, kot je RKG, navezujejo na eno samo bazo podatkov, ki je bila zgrajena po naročilu s strani tretje stranke in je vzdrževana namensko za podporo aplikacije.

Struktura podatkov

Baze slonijo na skrbno načrtanih podatkovnih shemah, ki so odraz potreb po zadovoljevanju izbranih ukrepov. Pravice na podatkovnih objektih so skrbno urejene, podatki pa primerno zaščiteni in varovani. Struktura podatkov je vidna neposredno iz baz podatkov, kar vključuje seznam tabel, atributov in ključev ter njihovo povezanost. Za uporabo v podatkovni analitiko bi namesto take strukture pričakovali za vsako bazo dokumentiran vmesnik (API), ki omogoča varen programski dostop do podatkov. V to strukturo ima vpogled Služba za informacijsko tehnologijo, ki suvereno pozna ustroj sistema, uporabljeno tehnologijo in možnosti dostopa, in da sproti in po potrebi upravlja s podatki. Primer sicer dobre dokumentacije in rešitve spletnih dostopov je RKG in »Dokumentacija RKG spletne storitve«, do katere hitrost dostopa in ustreznost za črpanje podatkov v repozitorij za podatkovno analitiko bi bilo potrebno preveriti v nadaljnjih projektih.

Uporaba ključev

Ključni v obstoječih bazah podatkov so največkrat specifični aplikacijam; upoštevajo se zakonske podlage. Med aplikacijami se ključni, ali pa opisni podatki, lahko ujemajo, ni pa sistem načrtovan tako, da bi se posamezne baze izognile replikaciji podatkov. Problem povezovanja med posameznimi bazami je sicer manj v strukturi podatkov ali nekompatibilnih ključih, ampak bolj kadrovski, v preobremenjenosti in podhranjenosti obstoječe ekipe, ki skrbi za informacijsko podporo, in pomanjkanju systemskega pogleda na vrednost podatkov, ki jih sicer skrbno zbira in premalo izkorišča MKGP. Na primer, na ministrstvu ni podatkovnih analitikov, ki bi skrbeli za integracijski aspekt izkoriščanja podatkov in s tem za povezljivost. Trenutna podatkovna struktura posameznih

podatkovnih baz je odlična, primerno načrtana s stališča posameznih aplikacij, ki pa ostanejo nepovezane in služijo specifičnim aplikacijam in točno določenim vsebinskim problemom in službam, ki skrbijo za tekoča opravila. Kar je seveda pozitivno, a s stališča podatkovne analitike in možnosti, ki bi jih zbrani podatki skupaj, združeno, nudili tako v operativnih kot strateških nalogah močno omejuje. Služba za informacijsko podporo torej uspešno in kljub premajhni kadrovske zasledbi skrbi za arhitekture in infrastrukture, katere upravljalec je Ministrstvo za javno upravo, a je vsebinska koordinacija, ki bi skrbelo za integracijo, sistematizacijo, skupne registre, normalizacijo baz s stališča izogibanja redundantnosti, podhranjena. Pri integraciji podatkov glavna prepreka torej ni tehnologija ali neurejenost podatkovnih baz in manjkajoči ključni. Problem je sistemski in kadrovske.

Odprtost baze, dostopnost podatkov

Vse ključne baze MKGP so zaprte in niso javno dostopne, oziroma so dostopne samo v delu, kot ga ponujajo obstoječe aplikacije, ki so povezane z bazami, omogočajo vnos in prikazujejo podatke. Nastavki za odprte dostope so za posamezne baze že izvedeni – primer take je RKG – je pa dostop do podatko omogočen samo uporabnikom, ki imajo za to pravno podlago in ustrezen certikat. Funkcionalnost aplikacij, ki so javno dostopne, je tipično določena, aplikacije s stališča analitike izpolnjuje točno izbrane funkcije, ki so z gledišča uporabnika nespremenljive. Kar je seveda pravilno in v skladu z obstoječimi zakoni in predpisi o hranjenju osebnih podatkov. Tehnologija, na kateri je podatkovna struktura zgrajena, ponuja možnosti vzpostavitve programskih vmesnikov za dostop do podatkov, a vzpostavitev in dokumentacija teh ni sistematično urejena zaradi navidezne (oziroma dejanske) ločitve in nepovezanosti in fokusa na vertikalno funkcionalnost tipa ena-baza-ena-aplikacija. Poskusov integracije, kot smo se ga konceptualno lotili v pričujočem projektu, do sedaj na MKGP ni bilo in zato tudi ne sistematične ureditve programskih dostopov do različnih podatkovnih baz. Naj omenimo tu tudi skrb, da četudi bi se MKGP odločil za vzpostavitev takih dostopov, bi s tem dodatno obremenil že tako preobremenjeno službo za informatiko.

Problem trenutne razpršenosti rešitev, torej podatkovnih baz in aplikacij, je tudi razpršenost ljudi, ki s temi rešitvami upravljajo. Tako ni točno jasno (zapisano), kdo so ključne osebe, ki bi lahko nudile podporo in odprle podatke za nove aplikacije na področju podatkovne analitike in ki bi skrbele in vzdrževale dostop do različnih podatkovnih shem.

Z drugimi besedami: ekipe, ki bi imela pregled nad celotno zbirko podatkov in ki bi z njo upravljala, na MKGP ni. Problem je razpršenost rešitev, baz, aplikacij, podatkovnih struktur, in osebja, ki upravljajo s temi ali pa ki skrbijo za uporabniški del. Dodatni problem je tudi podvajanje in optimizacija zapisov.

Programski vmesniki, programski dostop

Do podatkov, v splošnem, je moč dostopati z ustreznimi pravicami in z uporabo povpraševalnih stavkov SQL. Kot smo zapisali zgoraj trenutno uporabljena tehnologija omogoča vzpostavitev programskega dostopa; bralne ali tudi druge pravice je moč urediti, za analitiko pa bi bil dovoljšen bralni dostop z uporabo ustreznih gesel. Ureditve programskega dostopa je torej možna, a zahteva precejšen angažma s strani preobremenjene in kadrovske podhranjene službe za informacijsko tehnologijo.

Spremljajoča analitična orodja

MKGP oziroma njegovo izbrano osebje uporablja sicer zastareli Oracle Discoverer, uvedba poslovne inteligence pa ni stekla zaradi zahtev po anonimizaciji. V načrtu je gradnja podatkovnega skladišča z anonimiziranimi podatki. Delo poteka v sodelovanju z Ministrstvom za javno upravo (MJU), kjer pa se

pojavljajo nejasnosti glede izbire in vpeljave novih tehnologij. Na primer, Oracle Business Intelligence Suite, s katerim bi morda lahko MKGP podprl poslovno analitiko, ni kupljen zaradi alternativnih aktivnosti na MJU pri projektu Skrinja.

V splošnem opažamo odvisnost rešitev od dragih komercialnih orodij in odvisnost od opreme, za katero država plačuje letne licence. Uporabe odprtokodnih, prosto-dostopnih rešitev na MKGP-ju (in na MJU-ju) nismo zasledili. Z uporabo dragih orodij ministrstva tudi zapirajo možnosti za sodelovanje širšega kroga deležnikov pri urejanju in gradnji informacijske infrastrukture.

Dokumentacija

Skupne, systemske dokumentacije, ki bi naslavljal podatkovno integracijo in sistematizacijo podatkovnih struktur in analitike, ni. Dokumentacija bi recimo bila dobrodošla novim deležnikom, na primer, razvijalcem podatkovne analitike, in bi lahko vsebovala popis podatkovnih baz, popis programskih vmesnikov za dostop do podatkov, načine avtentikacije, in seznam odgovornih za baze in njene razširitve. Obstoječa dokumentacija je namesto systemske vezana na posamezne aplikacije, posamezne rešitve in posamezne sheme in odvisna od izvajalcev del, ki so razvili baze in pripadajočo aplikacijo.

Delovna naloga 1.4.: Identifikacija ključnih stvarnih pomanjkljivosti

Ključne pomankljivosti

Kadrovska podhranjenost: trije zaposleni v službi za informacijsko tehnologijo zaradi preobremenjenosti in skrbi za druge naloge ne bodo zmogli bremena uvajanja podatkovne analitike. Na MKGP ni službe za podatkovno analitiko, oziroma skrbnika z ustreznimi znanji računalništva, statistike in računskih pristopov, ki bi urejal področje integracije podatkov, skrbel za gradnjo (neobstoječih) podatkovnih skladišč in pazil, da je v vseh prihajajočih razpisih poskrbljeno za integracijo, dostop do skupnih registrov in nepodvajanje podatkov. Načeloma za podatkovno analitiko obstaja nastavek v »Službi za analize«, kje so zaposleni vodja in dva sodelavca, a trenutno ta služba s pristojnostmi in znanji ne pokriva področja podatkovnih znanosti. Na končni delavnici projekta je bil s strani te službi izkazan velik interes pri sodelovanju na področju projektov, ki bi podatkovne znanosti pripeljale na MKGP.

Razdrobljenost rešitev: rešitve tipa ena-baza-ena-aplikacija sicer zadovoljujejo trenutne potrebe na MKGP, a zaradi ločenega tipa izvedbe ne omogočajo horizontalne integracije in s tem preprečujejo ali pa vsaj močno zavirajo uporabo podatkovne analitike. Zdi se, da je z vsakim razpisom MKGP po novih rešitvah na področju IT sicer dobro poskrbljeno za posamezne izbrane vidike delovanja ministrstva, a da se pri tem pozablja na integracijo podatkov preko vseh obstoječih baz.

Pridobivanje novih znanj: znanja podatkovne analitike, vizualizacije podatkov in možnosti podpore sprotnega odločanja na podlagi podatkov so specifična. V pogovorih smo zasledili, ne pa tudi kvantitativno ocenili, da pridobivanje teh znanj ni urejeno sistematično oziroma da področje podatkovne analitike in znanosti o podatkih navkljub njegovi aktualnosti ni prioritarno.

Odvisnost od komercialnih orodij, zaprtost: MKGP uporablja drago programsko opremo, s tem ostaja odvisno do specifičnih ponudnikov te opreme in specifičnih razvijalcev. Drugačen pristop bi zahteval načrtno uporabo odprtokodnih programskih rešitev, zahtevo po uporabi te v tekočih razpisih, zahtevo po odprtosti orodij in zahtevo po ustreznih programskih vmesnikih, s katerimi bi lahko uredili povezavo med sicer ločenimi usmerjenimi rešitvami.

Izkušnje s podatkovno analitiko podatkov MKGP v projektu CRP

V namene pridobivanja uvida v kompleksnost podatkov in možnosti povezovanja podatkov iz različnih baz smo v projektu CRP izvedli nekaj pilotnih projektov podatkovne analitike. V prvem sklopu teh projektov smo gradili analitične delokroge za študij trendov vlaganja zahtevkov EK za obdobje 2015-2019, v drugem pa gradili modele in analizirali njihovo točnost pri napovedljivosti opuščanja površin. Za vsakega od teh sklopov je bilo potrebno pridobiti surove podatke, pri čemer so nam izčrpno pomagali sodelavci iz MKGP. Zahtevnejša analitika je bila narejena na podlagi podatkov AKTRP, ki vključujejo tudi izvirne podatke MKGP in UVHVVR, prav tako so pri pripravi pomagali informatiki AKTRP. Izkaže se, da tudi ta korak ni trivialen in da zahteva precej časa, kar kaže na potrebo po standardiziranih vmesnikih za črpanje podatkov (slika 1). Surove podatke je bilo treba nato obdelati in jih prilagoditi za analitiko. Predobdelave tega tipa so bile zahtevne in vključevale pisanje programov v jeziku Python, tipično dolžine nekaj tisoč vrstic. Cilj je bil recimo gradnja tabele, v kateri so profilirane kmetije po tipu rabe GERK-ov, oddanih zahtevkih, prejetih financiranj, in podobnih. Delo, ki smo ga opravili za predobdelavo podatkov kažejo na potrebnost specializiranih baz podatkov, ki integrirajo podatke iz različnih virov in profilirajo objekte (kmetije, GERK-e in katerekoli druge geografske sklope ali pa subjekte), ki bi jih želeli vključiti v analitiko.

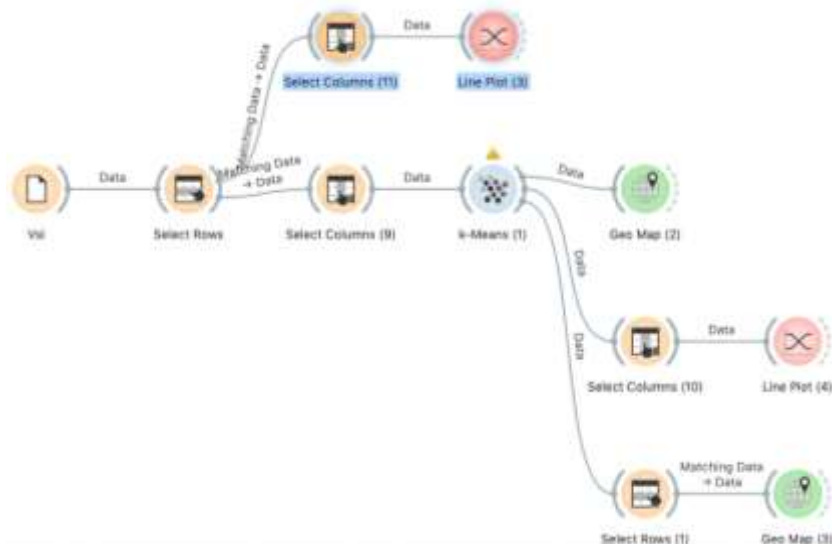
Register ekstenzivnih sadovnjakov	Šifrant raba	Evidenca EK prijava
Register hmeljišč	Šifrant avtohtonih in tradicionalnih pasem	Evidenca pridelovalcev semen
Register intenzivnih sadovnjakov	živali	Evidenca Izpolnjevanje pogojev - plačilne pravice
Register oljčnikov	Šifrant izjem gnojilnega načrta	Seznam bioplinarn in kompostarn
Register vinogradov	Šifrant podatki o živalih	Evidenca Dobrobit živali prašiči
Register CRG	Šifrant shema	Evidenca Negativna lista
Register konj	Šifrant presek slojev z rabami	Evidenca Blokirani KMG_MIDi ARSKTRP
Register CPZD	Šifrant slojev	Evidenca Izjema KMG_MIDi ARSKTRP
Register OMD	Šifrant kmetijskih zemljišč v uporabi	Evidenca obmejnih GERK-ov
Register CRPŠ DŽ stalež	Šifrant zahteve za sloj	Evidenca izobraževanje
Register CRG Živali	Šifrant rab za zahtevo	Evidenca EK vstop
Register CRG Premiki	Šifrant sort	Evidenca EK cert
Register CRG KMG MID-GMID	Šifrant KMRS	Evidenca EKSEME cert
Register CRG Napoved premikov	Šifrant Tip kolobarja	Evidenca EK travojede živali
	Šifrant napoved	Evidenca EK čebele
	Šifrant živali na planini in pašniku	Evidenca EKOP območij
	Šifrant kontrol	Evidenca AVTP
	Šifrant Razlog izločitve	Evidenca višja sila
	Šifrant rubrika	Evidenca kršitev DŽ
	Šifrant vrst živali	Evidenca dovoljenj za gojenje konoplje in vrtnega maka
	Šifrant držav	Evidenca habitati in ptice
	Šifrant opomb za šifre KMRS	Evidenca Združitev KMG
	Šifrant vrst primarnih proizvodov	Evidenca CC ugotovitev
	Šifrant predpogoj kontrole	Evidenca pooblastil za planine
	Šifrant zelenjadnic	Evidenca KOPOP in EK obveznosti
	Šifrant diverzifikacijskih skupin	Evidenca Trajno travinje
	Povezovalna tabela med šifrantoma diverzifikacijskih skupin in KMRS	Evidenca semenskih posevkov
	Šifrant samodejnega ponujanja KMRS na rabah	Evidenca mejice
	Šifrant nedovoljenih kombinacij zahtevkov	Evidenca ATGN
	Šifrant skupin EK	
	Povezovalna tabela med šifrantoma KMRS in skupin EK	
	Šifrant skupin EKSEME	
	Povezovalna tabela med šifrantoma KMRS in skupin EKSEME	
	Šifrant razlogov za blokiran prenos	
	Šifrant kontrolnih ugotovitev	
	Šifrant kategorij prašičev	
	Šifrant izvorov kršitev in zapisnikov	
	Šifrant stopenj krištev	
	Šifrant malih kršitev	
	Šifrant PZR/DKOS	

Povezovalna tabela med starimi in novimi šiframi ugotovitev
Šifrant razlog zmanjšanje KOPOP in EK
Šifrant razlog za neprijavo površine
Uradni zaznamki
Povezovalna tabela s šiframi ugotovitev

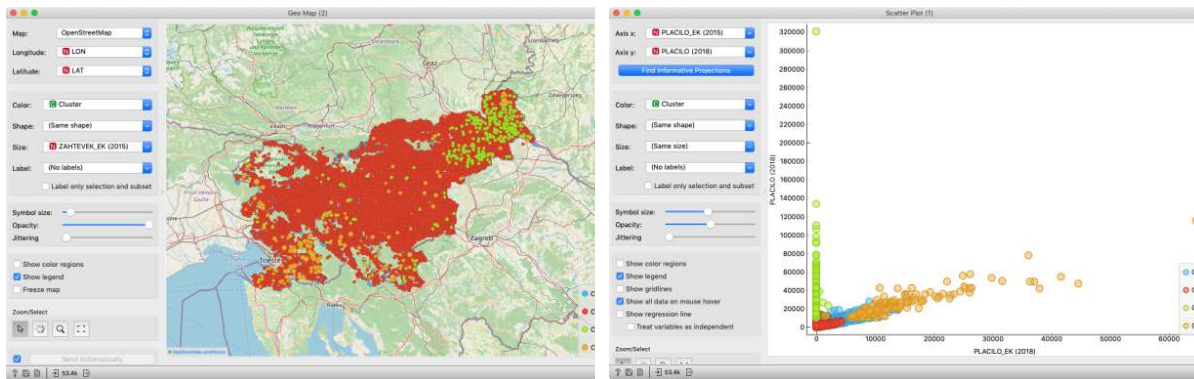
Slika 1 | Študija trendov vlaganja zahtevkov EK je zahtevala pridobivanje podatkov iz raznolikih registrov, šifrantov in evidenc, ki so prisotnih v različnih tabelah in bazah podatkov obstoječega podatkovnega jezera. Potrebno je bilo tudi sestaviti ustrezno programsko opremo za integracijo surovih podatkov.

V pilotnih projektih smo preverili uporabnost grafičnih uporabniških vmesnikov in sistemov za podatkovno analitiko, ki omogočajo vizualno specifikacijo analitičnih delokrogov. Uporabljali smo nekatere napredne analitične tehnike, kot so metode razvrščanja v skupine, metode gradnje in vrednotenja napovednih modelov in kombiniranje teh metod s podatkovnimi vizualizacijami, ki so vključevale prikaze na geografskih kartah. Primer enega od zgrajenih analitičnih delokrogov je prikazan na sliki 2, primer analize rezultatov razvrščanja v skupine pa na sliki 3.

V drugem sklopu pilotnih aplikacij smo analizirali možnost gradnje modelov, ki bi lahko napovedovali opuščanje kmetijskih površin oziroma napovedovali spremembo načina izrabe. Tudi tu je bil med časovno najbolj zamudnimi deli projekta predobdelava surovih podatkov in podatkovno profiliranje kmetij. Rezultati so obetavni, napovedi opuščanja kmetijskih površin so možne ($AUC > 0.75$, slika 4), smiselno je tudi zgrajeni model (slika 5). Modele te vrste je moč še dodatno izboljšati z obogatitvijo profilov kmetij oziroma dodajanjem značilnik, ki kmetije popišejo. Pri tem problemu se je izkazalo na primer, da zaradi različnih ključev podatkov iz prejšnjega sklopa analitičnih problemov (vlaganje zahtevkov) nismo mogli izkoristiti, čeprav bi bili ti lahko zelo informativni.



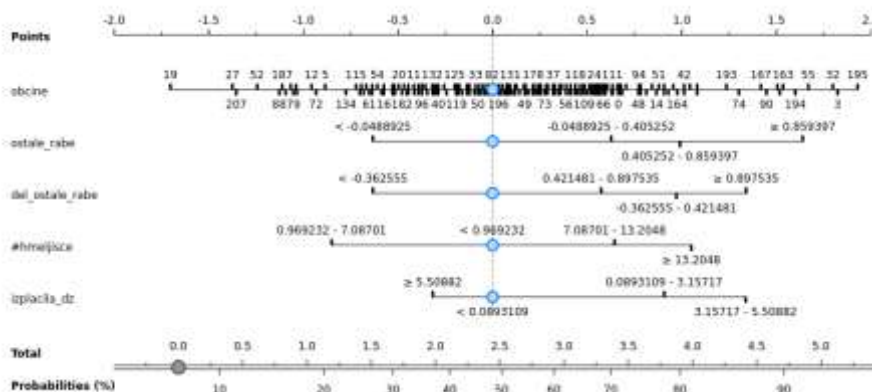
Slika 2 | Primer analitičnega delokroga z branjem podatkov, izborom kmetij, izrisom trendov (zgornja veja delokroga) in razvrščanjem v skupine (gradnik k-means za razvrščanje po metodi voditeljev) ter analizo skupin (spodnja veja delokroga).



Slika 3 | Primer grafičnega prikaza rezultatov gručenja, ki izpostavijo kmetije s Primorske in Goriške regije na eni strani in Pomurje, Dravsko in Ptusjko polje na drugi. Za slednjo se na primer v analizi izkaže, da tam kmetje vlagajo manj zahtevkov.

Model	AUC \uparrow	CA	F1	Precision	Recall
Naive Bayes	0.756	0.888	0.892	0.897	0.888
Logistic Regression	0.728	0.927	0.901	0.906	0.927
Random Forest	0.706	0.926	0.900	0.900	0.926
Constant	0.500	0.925	0.890	0.856	0.925

Slika 4 | Napovedne točnosti različnih tehnik modeliranja dobljene s testiranjem na neodvisni testni množici pri nalogi napovedovanja opuščanja kmetijskih površin. Po kriteriju površine pod AUC krivuljo se presenetljivo dobro odreže naivni Bayesov model, ki pa očitno potrebuje dodatno kalibracijo.



Slika 5 | Naivni Bayesov nomogram za izračun verjetnosti opuščanja zemljišč, iz izbranimi faktorji urejenimi po pomembnosti (občina kot najbolj informativen faktor). Možne so dodatne izboljšave modela, na primer s segmentacijo občin in dodajanjem ostalih faktorjev za karakterizacijo kmetij.

Delo na naštetih primerih s področja podatkovne analitike nas je opozorilo na sledeča »ozka grla«, ki bi jih bilo pri uvajanju teh tehnologij v bodoče dobro odpraviti:

- Pridobivanje surovih podatkov iz obstoječih podatkovnih baz je lahko zamudno, predvsem pa poteka preko angažiranja že tako prezasedenega osebja IT podpore v širšem smislu. Za vpeljevanje podatkovne analitike je potrebna standardizacija dostopa do podatkov preko ustrezno zgrajenih programskih vmesnikov.

- Za podatkovno analitiko je potrebna predobdelava podatkov in gradnja profilov zemljišč in kmetij. Gre sicer za ponavljajoča opravila, ki jih je možno standardizirati v sklopu inkrementalne izgradnje podatkovnega skladišča, ki pravilno upošteva tudi časovno dimenzijo.
- Analitiko je moč približati uporabnikom z uporabo enostavnejših grafičnih vmesnikov, ki uporabljajo delokroge. Zaželeno bi bila izgradnja vtičnikov, ki bi se lahko neposredno programsko priključili na podatkovne baze in (še neobstoječa) skladišča.
- Integracija podatkov v podatkovnem skladišču bo zahtevala usklajevanje med seboj nekompatibilnih ključev, uporabljenih v različnih bazah podatkov, a se bo to usklajevanje dalo rešiti na nivoju programskih skript in najbrž ne bo zahtevalo zahtevnih ročnih posegov ali sprememb v samih obstoječih bazah podatkov.
- Grafični uporabniški vmesniki, kot na primer ta, ki je bil uporabljen tudi v omenjenih pilotih, bi lahko služili tudi enostavnemu povpraševanju po bazi (npr. preštej, potem pa na karti prikaži vse kmetije, kjer vzgajajo trto) in ne nujno samo bolj kompleksnim opravlilom, kot so segmentacija in gradnja napovednih modelov.

Delovna naloga 2.1 in 2.2./4.2.: Gradnja podatkovnih zbirk, predlogi in izboljšave

Skladno z analizo stanja ter praktičnimi poskusi gradnje aplikacij, ki slonijo na podatkovni analitiki, ugotavljamo, da je trenutno stanje zbiranja podatkov na MKGP dobro, da pa so podatki porazdeljeni v bazah, ki primarno naslavlja posamezne delujoče aplikacije (slika 6). Odsvetujemo projekt, ki bi za vse te baze in podatke skušal zgraditi enoten repozitorij. Namesto tega svetujemo ciljnemu pristopu s serijo manjših projektov, ki bi naslavljali perečo problematiko operativne in strateške analitike in v korakih gradili vmesnike do obstoječih baz, sistem za integracijo podatkov ter uvajali programske ter spletne rešitve za podatkovno analitiko.

Konkretno, predlagamo:

1. **MKGP naj postavi oziroma imenuje koordinatorja za podatkovno analitiko.** Ta naj bo vključen v pripravo vseh razpisov na področju IT, naj tesno sodeluje tako z domenskimi eksperti kot s službo za informatiko in naj pri delu svetuje odločevalcem in pripravljalcem politik. Vloga koordinatorja je tudi priprava razpisov projektov, katerih cilj je uvajanje analitičnih tehnik in naprednih tehnik podatkovne analitike.
2. **Uvajanje tehnik podatkovne analitike naj poteka preko manjših, pilotnih projektov, s točno določenimi, praktičnimi in uporabnimi cilji.** Predlagamo, da MKGP ob tesnem sodelovanju z domenskimi strokovnjaki izbere na začetku tri, proti koncu pa do deset problemov, ki so pereči, ponavljajoči in ob obstoječi infrastrukturi težko rešljivi oziroma zahteva njihova analiza trenutno veliko ročnega dela (pisanje SQL stavkov, zahtevno obračanje podatkov v Excelu). Za projekte naj MKGP išče zunanje izvajalce, a naj zahteva odprte, dobro dokumentirane rešitve ki zahtevajo minimalno vzdrževanje. Velikost takih projektov naj bo od 40,000 do 70,000 EUR, s časom izvedbe do pol leta in s petletno vzdrževalno pogodbo z do 3,000 EUR vzdrževalnih stroškov na leto. Ocenjen strošek ne vključuje dela, ki je potrebno za spremembe na obstoječih bazah podatkov oziroma dela, kjer se implementira ustrezne programske vmesnike. Pilotni projekti naj vključujejo tudi izobraževanje uporabnikov. Poudarek naj bo na enostavni uporabi z vmesniki, ki uporabnikom dopuščajo gradnjo kompleksnejših analitičnih postopkov.

3. **Uvajanje odprtokodnih orodij.** Predlagamo, da s e vsi pilotni projekti izvedejo na odprtokodnih tehnologijah, ki finančno ne predstavljajo nikakršnega dodatnega bremena.
4. **Izobraževanje uporabnikov.** Napredna podatkovna analitika mora biti enostavno uporabna in intuitivna za domenske strokovnjake na MKGP, ki naj jih bo moč za uporabo uvedene tehnologije usposobiti v enem delovnem dnevu.
5. **MKGP naj v naslednjih treh letih zasleduje cilj porabe vsaj 10% sredstev za IT za integracijo podatkov, razvoj napredne podatkovne analitike in razvoj vmesnikov, ki bodo odločevalcem nudili možnost enostavnega dostopa do podatkov.**
6. **V razpise za IT tehnologije naj MKGP vključuje zahteve po integraciji z obstoječimi bazami, izogibanju redundantnosti, ter vzpostavitvi dobro dokumentiranih programskih vmesnikov za dostop do vseh podatkov.** Zahteve naj pripravi in pri vključitvi v razpise sodeluje koordinator za podatkovno analitiko.

Od vseh zgoraj naštetih je najpomembnejša vzpostavitev vloge koordinatorja za podatkovno analitiko in vzpostavitev mehanizmov dela na manjših pilotnih projektih, ki bi sčasoma na MKGP uvedli prakso uporabe naprednih tehnik podatkovne analitike. Gre za potencialno rizične projekte z večjim vplivom na delo ministrstva. Rizične zato, ker z njimi MKGP vzpostavlja nove prakse in ker so projekti močno odvisni od možnosti integracije z obstoječimi bazami in vzpostavitvijo novih programskih vmesnikov in okolij za dostop do podatkov. Koordinacija projektov naj skrbi za, če je le možno, enotno odprtokodno platformo in sočasen razvoj podatkovnega skladišča, ki lahko po kompleksnosti in velikosti raste z opravljenimi pilotnimi projekti.

Pomemben je tudi izbor problematike, ki jo rešujejo pilotni projekti. Ta naj bo izbrana predvsem z vidika uporabnosti rešitev in reševanja problemov, ki so kritični za MKGP. Za ustrezen izbor takih projektov MKGP predlagamo organizacijo delavnice, ki se jo udeležijo deležniki-uporabniki in domenski eksperti, odločevalci, predstavniki IT podpore in zunanja skupina strokovnjakov s področja podatkovne analitike in gradnje uporabniških vmesnikov.

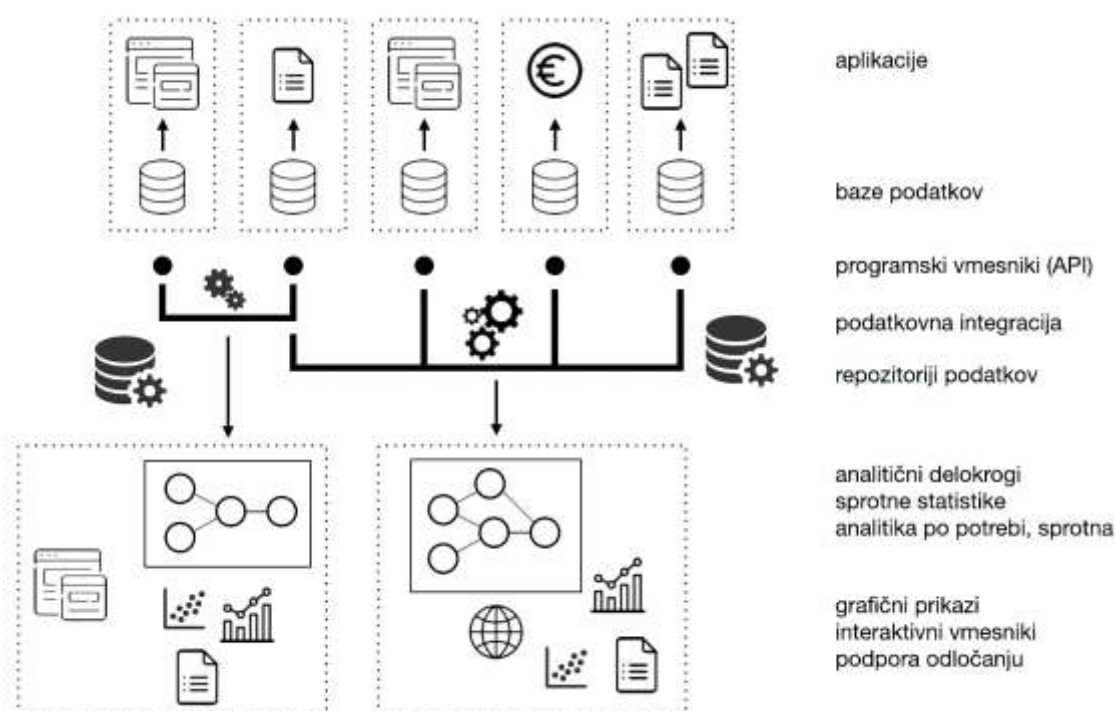
Priporočamo tudi reševanje problematike podatkovne analitike s serijo manjših projektov (slika 6). MKGP naj prične z enim in nadaljuje z dvema manjšima projektoma, in jih uporabi za učenje v pisanju projektnih zahtev in učenje uvajanja podatkovne analitike v delo ministrstva. Nadaljuje naj z manjšimi projekti, ki vsak rešuje specifične zahteve uporabnikov, a, če je le možno, temelji na tehnologijah, ki so jih uvajali prejšnji projekti. Predlagamo izvedbo prvih treh projektov v enem letu, in izvedbo ostalih projektov, skupaj do deset projektov, v naslednjih dveh letih. Projekti naj bodo kratki, z izvedbenimi roki od treh do šestih mesecev.

Pri planiranju projektov je nujna določitev skupine MKGP, ki projektu sledi in ki vključuje koordinatorja podatkovne analitike, odločevalca, ključnega uporabnika in osebo iz službe za informatiko, ki lahko zagotovi programski dostop do ustreznih baz. Sodelovanje službe za informatiko je nujno že pri zasnovi projekta, saj je prav ureditev programskega dostopa do ustreznih surovih podatkov lahko na kritični poti projekta in mora biti pri vzpostavitvi projekta jasno, da je tak dostop izvedljiv in da bodo za projekt na ta način dostopni vsi potrebni podatki. Ključna vloga uporabnika je vzpostavitev ciljnih zahtev za projekt in v sodelovanju pri oblikovanju uporabniških vmesnikov oziroma potrebnih analitičnih delokrogov in izboru podatkovnih vizualizacij.

Serija manjših projektov lahko naslovi zelo različne, a pereče teme. Spremljanje in napovedovanje opuščanja kmetijskih zemljišč, tema, ki jo naslavljamo v pilotnem projektu, je samo ena od možnih

tem, ki dobro prikaže uporabnost integracije podatkovnih virov. Drugi primeri tem so na primer analiza izrabe kmetijskih zemljišč v namene planiranja kmetijske in okoljske politike, ali pa opazovanje trendov na področju ekološkega kmetijstva, prepoznavanje tipov zemljišč in tipov kmetij.

Eden od projektov lahko na primer naslovi tudi vprašanje enostavnega dostopa do osnovnih statističnih podatkov in podatkov agregiranih po kmetijah ali geografskih enotah. Dostop do teh podatkov je v trenutnem sistemu seveda omogočen, a po nepotrebnem obremenjuje zaposlene. Čar takega projekta je v načrtovanju uporabniških vmesnikov in preverjanju njihove praktičnosti pri uporabnikih, kjer so cilji začrtati sisteme, s katerimi lahko hipno pridemo do enostavnih in zahtevnejših statistik in ključnih podatkov za odločanje.



Slika 6 | Sedanja organizaciji podatkov, s katerimi upravljata MKGP in AKTRP in ki večinsko sledijo modelu ena-aplikacija-ena-baza (zgornji del grafičnega prikaza) je za uvajanje podatkovne analitike potrebno dodati programske vmesnike in funkcionalnosti za integracijo podatkov (srednji del). Vmesnike in integracijo je moč dodajati po potrebi, nad temi vmesniki pa graditi posamezne, možno ločene aplikacije, ki naslavljaajo ključne potrebe MKGP-ja za operativno in strateško analitiko.

Priložnosti

Tako kot praktično v vseh ostalih institucijah sistema javnega upravljanja, kot tudi praktično v vseh podjetjih na področju evropske skupnosti, je tudi na MKGP uporaba podatkovne znanosti v povojih. Kar seveda glede na nedavni razvoj te stroke in njenih orodij ne preseneča. Je pa podatkovna znanost za MKGP izjemna priložnost, saj ministrstvo lahko gradi na spektru odlično postavljenih in načrtovanih podatkovnih baz, na množici podatkov, ki so tam zbrani in urejeni, in na perečih problemih s področja kmetijskih in okoljskih politik, ki kličejo po integrirani uporabi že zbranih podatkov. **Podatkovna analitika je zato za MKGP izjemna priložnost, kjer še prav nič ni zamujeno in kjer Slovenija lahko orje ledino na področju digitalizacije in uspešnih uporab naprednih računskih pristopov in umetne inteligence.**

Delovne naloge 4.1 in 4.2

Eksperiment 1: Atributivna večnamenska zbirka podatkov o KMG za obdobje 2015-2019 (mag. Maša Kerstein, Vesna Tanko)

Podatkovna analitika (s podporo pri odločanju) predstavlja kmetijskemu sektorju pri nas, kot tudi v EU, velik izziv. MKGP je v okviru prehoda sedanje skupne kmetijske politike (SKP) na politiko, temelječo na rezultatih, sprejel izziv kot priložnost, kar priča razpis zadevnega projekta razvoja IT rešitev v podporo SKP, ki vključuje področja digitalizacije in uspešnih uporab naprednih računskih pristopov in umetne inteligence.

Zato je v tem kontekstu pomemben podatek, da začetek tega projekta (november 2018) časovno sovпада s pomembnim mejnikom iz aprila 2019, ko je 25 evropskih držav, med njimi tudi Slovenija, podpisalo izjavo o sodelovanju z naslovom „Pametna in trajnostna digitalna prihodnost evropskega kmetijstva in podeželja“ z namenom, da bi sprejeli številne ukrepe za podporo uspešni digitalizaciji kmetijstva in podeželja v Evropi. Ta pomeni priznanje potenciala digitalnih tehnologij za reševanje pomembnih in nujnih gospodarskih, socialnih, podnebnih in okoljskih izzivov, s katerimi se soočata kmetijsko-živilski sektor EU in podeželje. Prepoznan je napredek v robotiki za natančno kmetovanje in sistemih za izvajanje SKP, ki temeljijo na rešitvah za upravljanje digitalnih podatkov. Sporočilo berlinskega svetovnega foruma za hrano in kmetijstvo naslavlja cilj prispevati k napredku k pametni in trajnostni prihodnosti evropskega kmetijstva, pri čemer se podpisnice izjave zavezujejo prizadevati si za krepitev podpor raziskavam, vzpostavitev inovacijske infrastrukture, ustvarjanje evropskega podatkovnega prostora za pametne kmetijsko-živilske aplikacije in povečanje učinka rešitev v podporo prehodu SKP k politiki, ki temelji na rezultatih.

Že pred izvedbo prve delavnice junija 2019 je bil sprejet dogovor o odprtem pozivu po nalogah, ki jih MKGP v širšem kontekstu naslavlja kot prioritete kot pomoč pri izvajanju svojih rednih nalog in/ali za namen strateškega načrtovanja SKP po letu 2020. Odzvalo se je več predlagateljev s strani MKGP in AKTRP, zlasti strokovnih delavcev s področij, ki pri svojem delu redno uporabljajo in upravljajo z različnimi bazami in/ali načrtujejo ukrepe oziroma delajo na področju uredbe in postopkov v zvezi z integriranim administrativnim in kontrolnim sistemom (IAKS) ter sistemom navzkrižne skladnosti.

Poudarjamo pomen vključenosti domenskih ekspertov – skrbnikov in vsebincev ter hkrati pozivamo k standardizaciji, integraciji in odpiranju podatkov ter univerzalnih orodij za krepitev učinkovitosti in transparentnosti odločanja.

Zbrane pobude in potrebe smo zbrali v zbirni preglednici, ki je pokazala na številne in raznolike potrebe preprostih pregledov, statističnih analiz in vizualizacije v zvezi z različnimi atributivnimi podatki - potrdila se je izkazana potreba administrativno-odločevalske strukture po standardizaciji podatkov in večnamenskih aplikacijah za pregled, analizo in vizualizacijo. Med kompleksnejšimi nalogami so izstopale prostorske spremembe kot je zaraščanje kmetijskih površin ter vprašanje potenciala za naslavljanje bolj zelenih, okoljsko naravnanih ukrepov, nekateri dejavniki v zvezi s tem in predvidevanje. Na podlagi izsledkov smo sprejeli odločitev, da se v okviru DS4 demonstracijske naloge razširi atributivno zbirko podatkov o kmetijskih gospodarstvih (KMG) v okolju [Orange](#), ki je skladno z izraženimi potrebami omogočala raznolike statistične analize na statični zbirki s številnimi atributi ter vizualni izpis na zemljevidu, na obdobje izvajanja trenutne SKP (2015-2020), kar bi omogočilo smiselno ugotavljanje časovnih trendov v zaključenem obdobju izvajanja ukrepov.

Poudariti velja, pa dodana vrednost naloge ni v tem, da bi z orodjem pridobili bolj natančen ali dodelan odgovor, ko gre za posamezno vprašanje - čeprav smo pri nekaterih zadevah orali ledino tako s sodelujočimi kolegi iz MKGP, kot tudi organov v sestavi.¹ Prednost našega eksperimenta je zlasti v tem, da omogoča v kratkem času odgovore na številna vprašanja, ki jih ob vzpostavljanju podatkovne zbirke morda sploh še nismo poznali. Orodje omogoča sprotne, hitre preveritve, npr. v okviru delovnega sestanka se lahko preveri na koga se določen ukrep nanaša, kje prihaja do težav pri izvedbi, pragove potencialnih ukrepov ... brez obremenjevanja skrbnikov. Pri tem pa je pomembno omejiti tveganja, ki jih uporaba novih orodij prinaša – potrebna je vključenost domenskih ekspertov in dvosmerna komunikacija z IT eksperti, kakor tudi izobraževanje in osveščanje odločevalske strukture in deležnikov. Potrebno je graditi skupnost ustvarjalcev in uporabnikov podatkov za smiselno uporabo, izboljšati zbirke in načine uporabe.²

Oblikovanje večletne podatkovne zbirke za namen reševanja analitičnih nalog

V analizo so vključena vsa kmetijska gospodarstva (KMG), ki vlagajo zahteve za kmetijske subvencije v obdobju 2015 – 2019. KMG smo, ob pomoči in usmeritvi strokovnih sodelavcev MKGP ter AKTRP - nosilci ukrepov/shem/integriranega administrativnega kontrolnega sistema in sistema navzkrižne skladnosti - za namen konkretne naloge, razvrstili v 4 TIPE KMETIJ:

- živinorejska (v stolpcu SBV2017_ImetnikiRejnihZivali 'da' in hkrati mora imeti posamezen KMG 10 GVŽ ali več);
- vrtnarska (vsaj dve zelenjadnici in hkrati v min obsegu 1 ha zelenjadnic skupaj);
- poljedelska (skladno po šifrantu KMRS min obseg 1 ha poljščin (pšenico, koruzo, ječmen, oljna ogrščica ...) + omejitev obremenitve GVŽ: 0 do 0,5 GVŽ/ha. Izračun obremenitve na interval 0-0,5 GVŽ/ha: suma GVŽ na kmetiji/ KZU =kmetijske površine v uporabi na kmetiji);
- trajni nasadi (kmetije, ki imajo vsaj en vinograd/sadovnjak, min 1 ha).

Namen izbrane tipologije je z uporabo čim več kakovostnih podatkov skozi vrsto aplikacij, zbirk, registrov v določenem časovnem obdobju oblikovati tipe kmetij, ki so prilagojene slovenskim razmeram in predpisanim omejitvam ukrepov, na podlagi katerih s pomočjo izbire različnih kriterijev za različne cilje politik lahko raziskujemo posamezne skupine in poiščemo vzorce, pomembne za naše razmere in stanje, napovemo trende in ovrednotimo potencial. Zato so v našem primeru kmetije lahko razvrščene v toliko tipov kmetij, kolikor kriterijem ustrezajo. Vendar pa je prekrivanje oz. podvojevanje v smislu nekompatibilnosti z vidika načrtovanja intervencij v okviru kmetijske politike, prilagojene nacionalnim razmeram, preprečeno s postavitvijo konkretnih mej/kriterijev znotraj posameznega tipa kmetij. Vendar gre zgolj za eno od možnosti, ki smo jo v nadaljevanju predstavili. Povsem mogoče bi bilo določiti in integrirati za namen stojnega učenja tudi drugačne tipologije, npr. usmerjene na statusno – pravne oblike kmetijskih gospodarstev, glede na oblikovane kriterije skladno s statističnim

¹ Trenutni model reševanja nalog za potrebe kmetijske politike je utirjen in omejen na uporabo IT orodij, ki ne vključujejo podatkovne analitike in reševanja nalog z uporabo umetne inteligence. Nov zakonodajni reformni paket SKP naslavlja prav vpeljavo tovrstnih orodij tako na različnih področjih, kakor tudi sistemsko. Z razmišljanjem, ki ga podatkovna analitika zahteva v zameno za hitre in kakovostne strateške rešitve, se na področju kmetijstva še nismo srečali

² Različni načini interpretacije, ki bi lahko bili sporni/problematični, bodo prišli do izraza takrat, ko bodo podatki na voljo zunanjim uporabnikom z različnimi interesi, k čimer EU, ne samo na ravni širše javnosti temveč tudi v okviru reformirane SKP že močno in vztrajno poziva. V tem smislu je boljše, da ustvarjalci podatkov (MKGP) sami proaktivno vodijo proces odpiranja in uporabe podatkov.

raziskovanjem kmetijskih gospodarstev in raziskovanjem metod kmetijske proizvodnje v zvezi s seznamom značilnosti, ki se zbirajo v raziskovanju o strukturi kmetijskih gospodarstev (kot določajo evropski in nacionalni predpisi s tega področja) ali po katerikoli drugi analizi kmetijskih gospodarstev glede na njihove osnovne značilnosti. Prednost orodja je namreč prav v tem, da omogoča poljubne tipologije glede na namen in znanje uporabnika. Ekonomist ali okoljevarstvenik bi gotovo uporabila drugačne tipologije, pri čemer bi bila velika razlika že v obsegu zajema surovih podatkov v osnovni podatkovni tabeli kot podlage za uporabo kriterijev posamezne tipologije. Tipe kmetij bi lahko razvrstili / opredelili tudi z uporabo umetne inteligence (UI).

Vsi podatki in končne spremenljivke, uporabljene v raziskavi, so zaradi predaje zbirke v arhiv družboslovnih podatkov pojasnjene in opisane, kar *omogoča ustrezno nadaljnjo uporabo podatkov s strani tretjih oseb in za različne namene.*

Enoten PROFIL KMG, ločen v zbirkah za posamezna leta od 2015 do 2019 smo oblikovali na način, da smo posamezne spremenljivke iz posameznih aplikacij obdržali v njihovi osnovni vrednosti, nekatere, z istim imenovalcem (kjer je bilo to primerno zaradi ogromne količine podatkov) pa združili v nadpomenko, in zapisali nov naziv sestavljene spremenljivke. Osnovna baza na AKTRP, ki se uporablja za namen subvencij je DATAPROD (Oracle). Glavne uporabljene zbirke/sheme/tabele v bazi, ki smo jih pri oblikovanju večletne podatkovne zbirke uporabili, so, na nivoju aplikacij:

- KSS 2015 – 2019 → *naziv aplikacije:* Vnos subvencij (KSS) (*tudi Zajem*)
- SBV 2015 – 2019 → *naziv aplikacije:* Obdelava subvencij
- KONTROLE 2015 – 2019 → *naziv aplikacije:* Kontrola subvencij
- SBV 2015 – 2019 → *naziv aplikacije:* Obravnava kontrolnih ugotovitev (PINSU)
- OBR 2015 – 2019 → *naziv aplikacije:* Obračun subvencij AKTRP

Znotraj Zajema so vključeni vsi potrebni:

- Registri (npr. registri trajnih nasadov, živali - VOLOS registri, itd...),
- Šifranti (šifrant rabe, KMRS, skupin EKSEME, vrst živali, PZR/DKOS, itd...) in
- Evidence (izobraževanj, plačilne pravice, evidenca združitvev KMG, itd...).

Pri Trend vlaganja zahtevkov EK število zahtevkov EK pomeni seštevek GERKov (ali poljin), za katere posamezen kmet vlaga/uvlavlja zahtevke za ukrep ekološkega kmetovanja.

Odločitev o izbiri naloge za prikaz

Za prikaz analitične naloge v okviru večletne zbirke smo si zastavili primer s področja ekološkega kmetovanja, ki se, zlasti zaradi zadnjih dogajanj, uvršča med kompleksne naloge strateškega načrtovanja in prioritete naloge MKGP. Že v zakonodajnem paketu SKP post 2020 so cilji ambiciozne okoljske politike ob zeleni arhitekturi vgrajeni v ambiciozno kmetijsko politiko, lani je bil sprejet še evropski Zeleni dogovor, katerega cilj je izboljšati dobrobit ljudi. Dodatno zaostritev okoljskih zahtev prinašata s strani EK nedavno sprejeti Strategija o biodiverziteti in Strategija od vil do vilic. Slednja med 4 ključnimi cilji zahteva tudi povečanje deleža površin pod ekološkim načinom kmetovanja (vsaj 25 %). Izhodišče naloge je populacija kmetov, ki vlaga zahtevke za subvencije v okviru zbirne vloge (ZV), na letnem nivoju to populacijo predstavlja cca 57.0000 kmetov. Zahtevke (skladno z definicijo predpisov s področja PRP) za ukrep ekološkega kmetovanja (EK) pa vlaga v povprečju le cca 3.200 kmetov na leto.

Pri reševanju naloge smo si pomagali z orodjem za podatkovno rudarjenje - Orange. Orodje je namenjeno vizualizaciji in analizi podatkov ter omogoča uporabo naprednejših tehnik strojnega učenja, kot so npr. klasifikacija, regresija, gručenje.

Cilj: pregledamo stanje podatkovne zbirke o subvencijah za ukrep EK. Z uporabo orodja Orange sestavimo delokrog, ki združi profile kmetij za vsa obdobja v eno tabelo. (Profil kmetije je predstavljen z eno vrstico v tabeli.) Na grafu prikažemo trend vlaganja EK zahtevkov in trend odobrenih plačil. Da bi odkrili morebitne vzorce v podatkih, se odločimo za uporabo postopkov nenadzorovanega učenja. Podatke normaliziramo in gručimo. Preverimo in ovrednotimo gručenje za različno število skupin (2-8). Na zemljevidu prikažemo rezultate gručenja na dve in štiri skupine. S pomočjo razsevnega diagrama raziskujemo posamezne skupine, ki so nam zanimive z vidika reševanja naloge. Skupino, ki najbolj odstopa, podrobneje pregledamo in preučimo s pomočjo grafov in zemljevida. Vprašamo se ali napovedan trend predvideva dosego zastavljenega cilja do določenega časovnega obdobja (leto 2030).

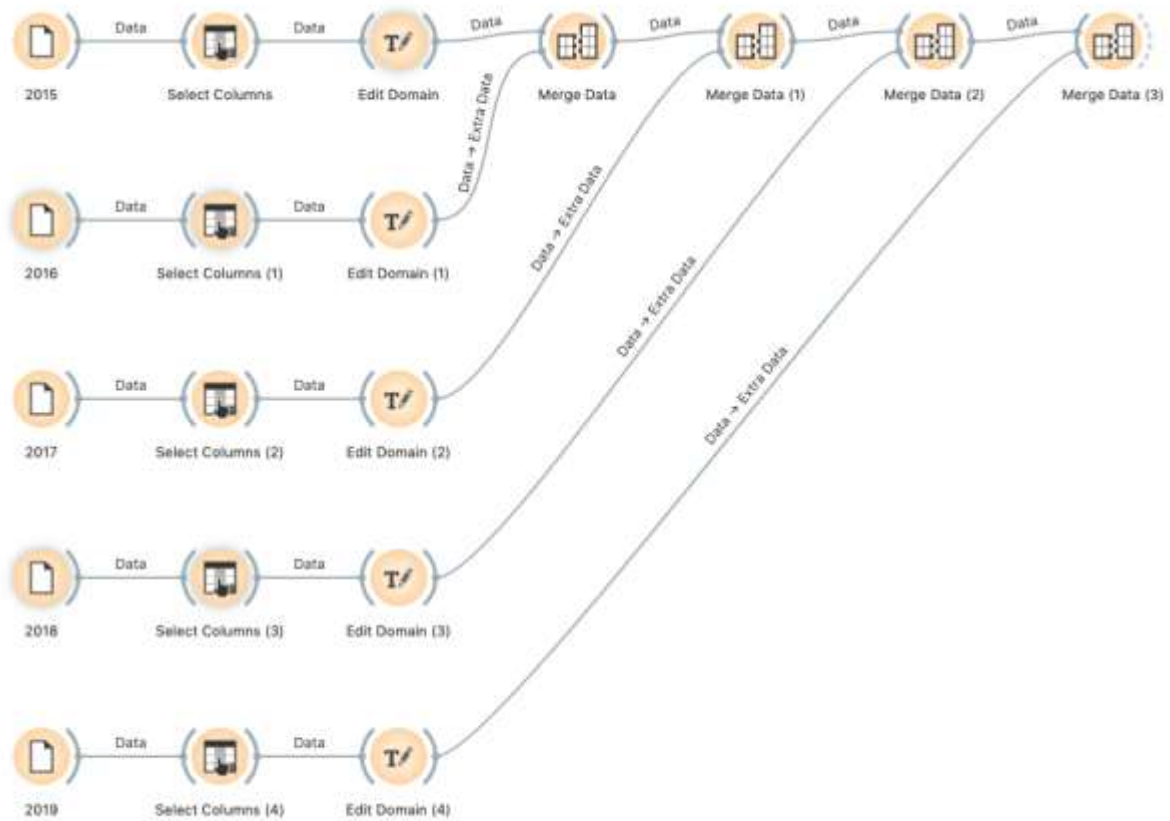
Vemo, da je za obdobje 2014-2020 za ukrep EK namenjenih cca 66 mio EUR, to je cca 10 mio na leto. Črpanje sredstev je dobro.

STANJE: cca 3000 KMG (letno) v izbranem obdobju vključuje cca 10% kmetijskih površin, kar predstavlja cca 40.000 ha. Za to pridobijo kmetje cca 10 MIO€/leto.

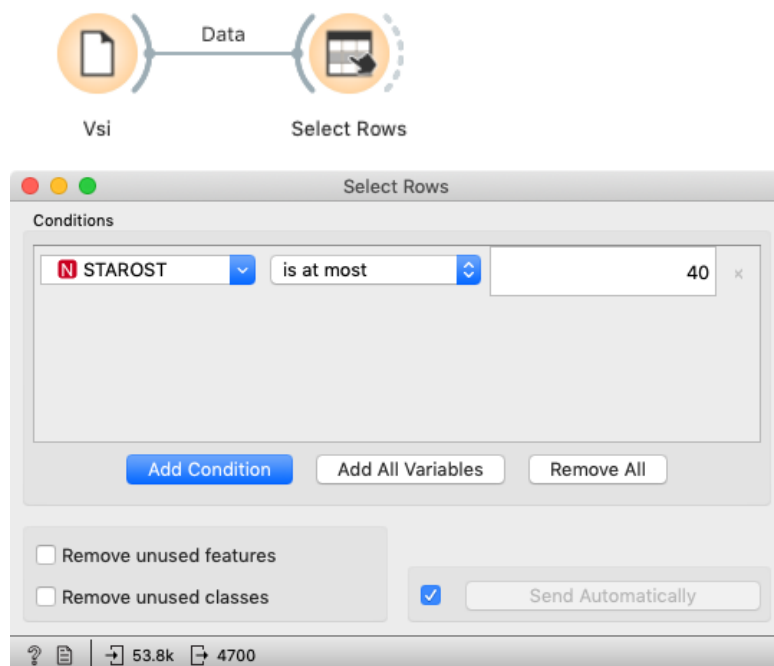
Analiza podatkov za kmetije »mladi kmet«

Ker je za dosego ciljev predlogov ukrepov/intervencij nove SKP uporaba različnih IKT orodij (od preciznega kmetovanja do uporabe različnih aplikacij) s strani kmetovalcev zaželeno, priporočena ali pa celo obvezna, smo se v prvem delu naloge koncentrirali na mlajšo populacijo kmetov, ki ji je tehnologija bližja in jo pogosteje uporablja. Starostno omejitev smo izenačili s starostnim pragom iz predpisov, ki še omogoča dodelitev dodatnih vzpodbud in prejemka subvencij iz naslova ukrepa »Mladi kmet«. Ta populacija je po predvidevanjih tudi dovezetnejša za novejša ukrepe (varnost, kakovost pridelane hrane), hkrati pa ima praviloma višji nivo izobrazbe kot starejša populacija kmetov (naraščajoč trend starostnikov, nosilcev kmetij, je prepoznan kot problematika našega kmetijstva).

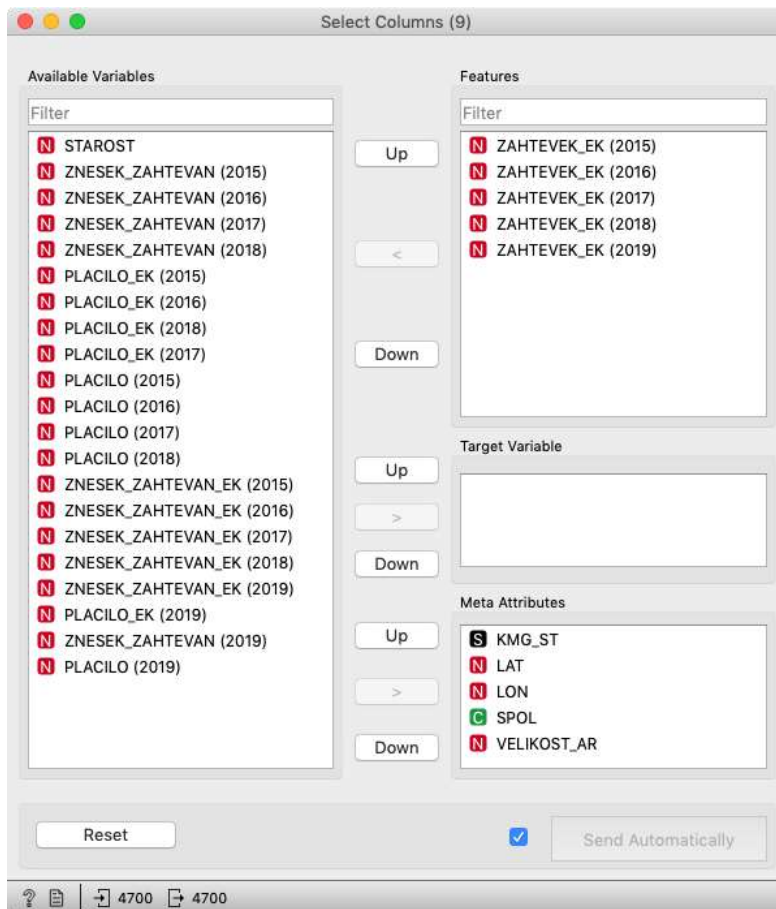
Slika 1 | Sestavimo delokrog, ki združi profile (ena vrstica v tabeli predstavlja eno kmetijo) za obdobje 2015-2019 v eno tabelo. V večini primerov reševanja problemov bi gradnike prikazanega delokroga združili v en nov gradnik, ki bi predstavljal gradnik branja podatkov iz baze. V našem primeru združene podatke shranimo v novo datoteko, ki jo uporabljamo v naslednjih korakih (1-3).



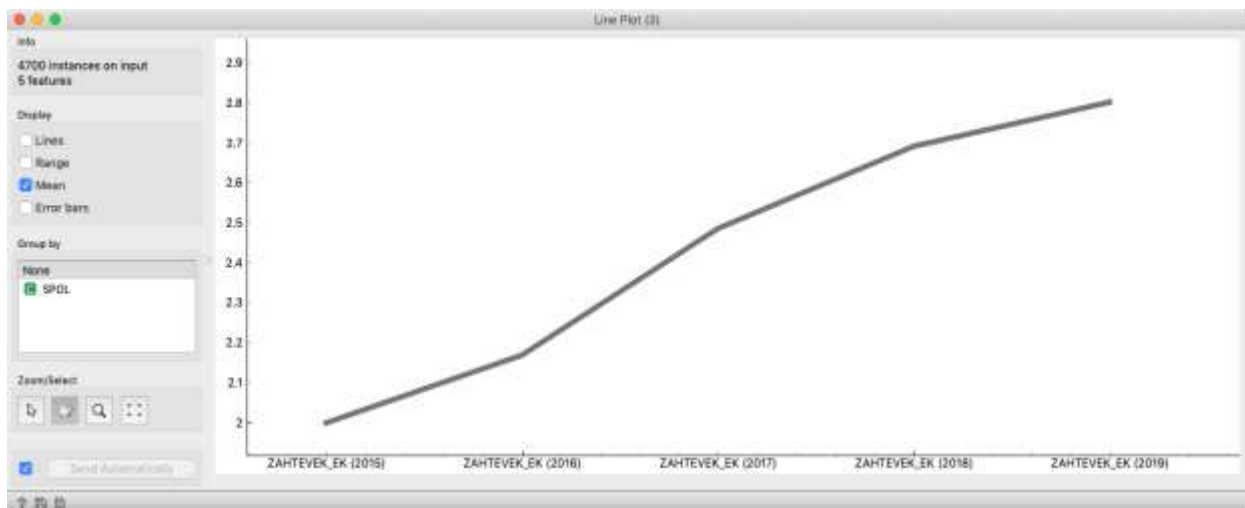
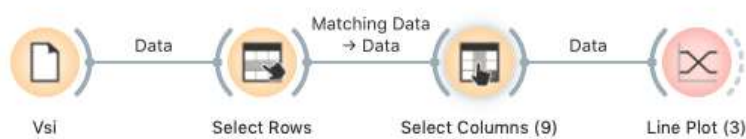
Slika 2 | (Korak 1) Preberemo združene podatke in dodamo omejevanje kmetij glede na starost nosilca (40 let).



Slika 3 | (Korak 2) Izmed vseh stolpcev izberemo tiste, ki opisujejo trend vloženih EK zahtevkov skozi leta (2015 – 2019).



Slika 4 | (Korak 3) Trend prikažemo na grafu (prikazano je le povprečje).

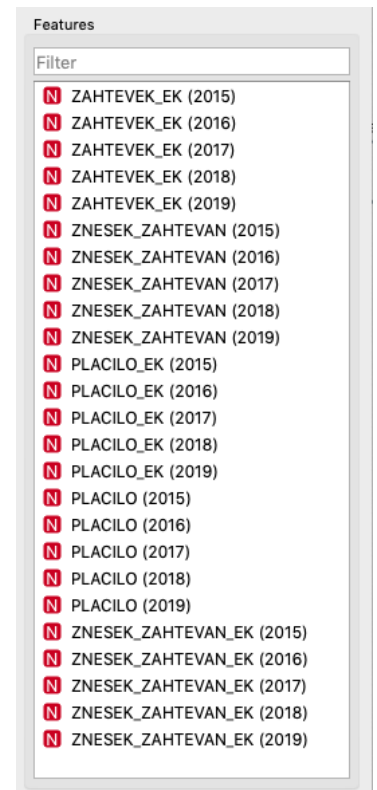


Trend vlaganja zahtevkov nekoliko raste . Na x osi je prikazan zahtevek EK po letih, na y osi pa število zahtevkov EK. V analizo so vključeni le tisti KMG, ki so skozi vsa leta analize vlagali minimalno en zahtevek EK. Ker po naši definiciji opredelitve »zahtevka EK« štejejo GERKe, za katere se uveljavlja zahtevek, je številka nekoliko višja, 4.700 enot (razlika od cca 3.200 KMG, ki vlagajo zahtevke EK skozi leta).

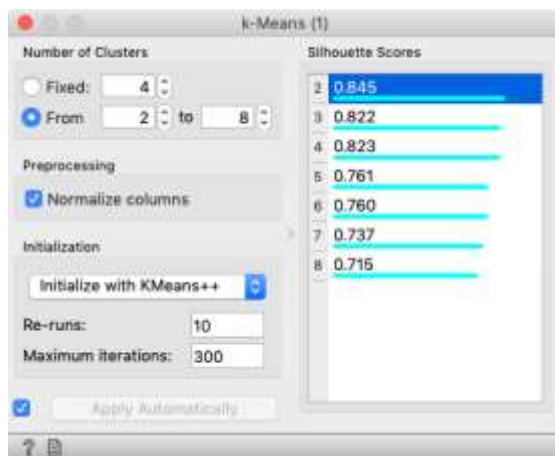
Slika 5 | Analizo nadaljujemo tako, da naredimo gručenje podatkov ([K-means](#)):



Slika 6 | Izberemo naslenje stolpce:



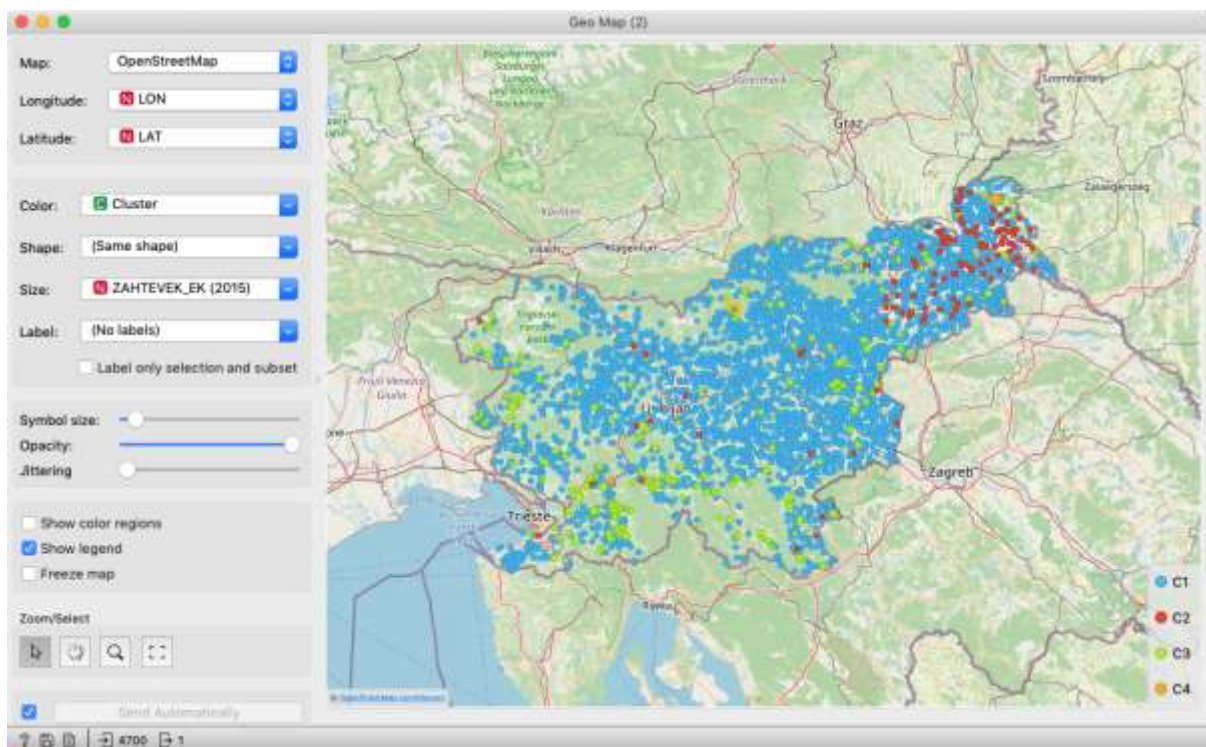
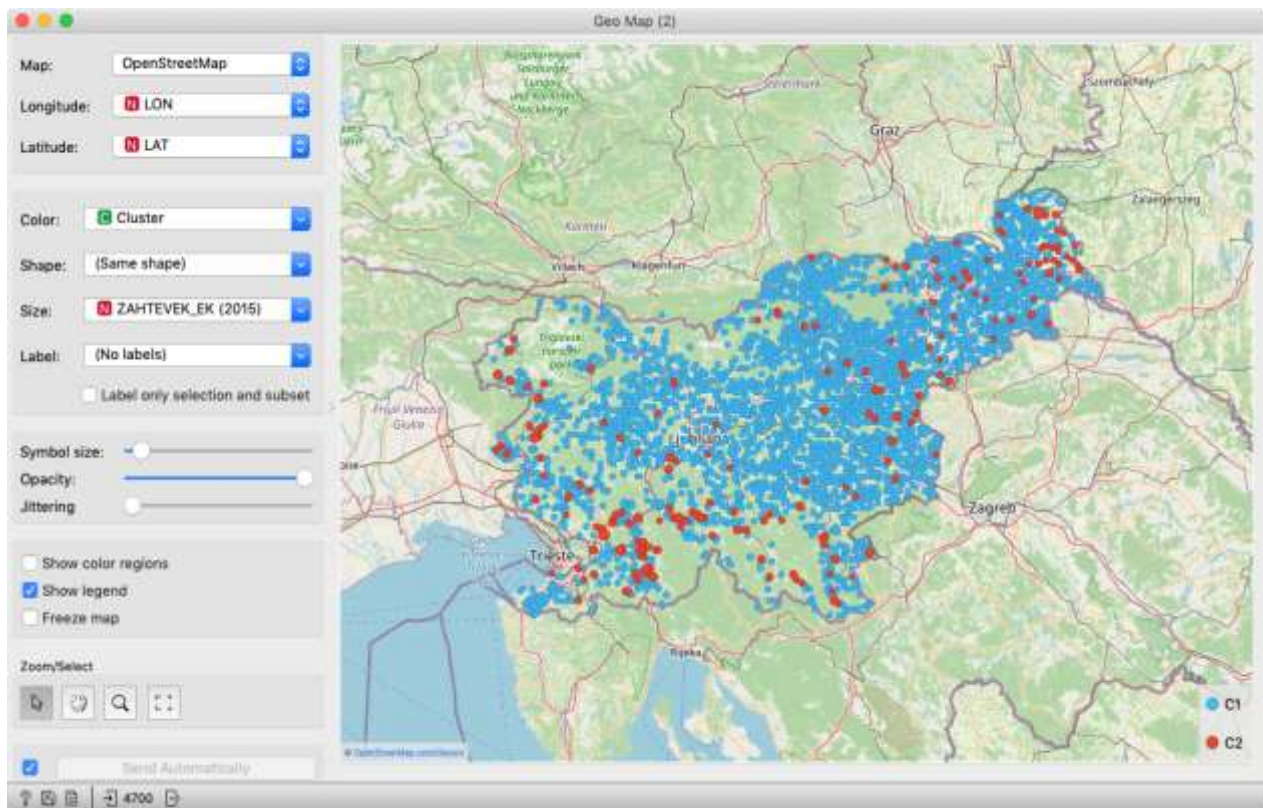
Slika 7 | Rezultat gručenja



Slika 8 | Pregledamo rezultate gručenja, kjer smo dosegli največji Silhouette Score (t.j. gručenje profilov na dve skupini):

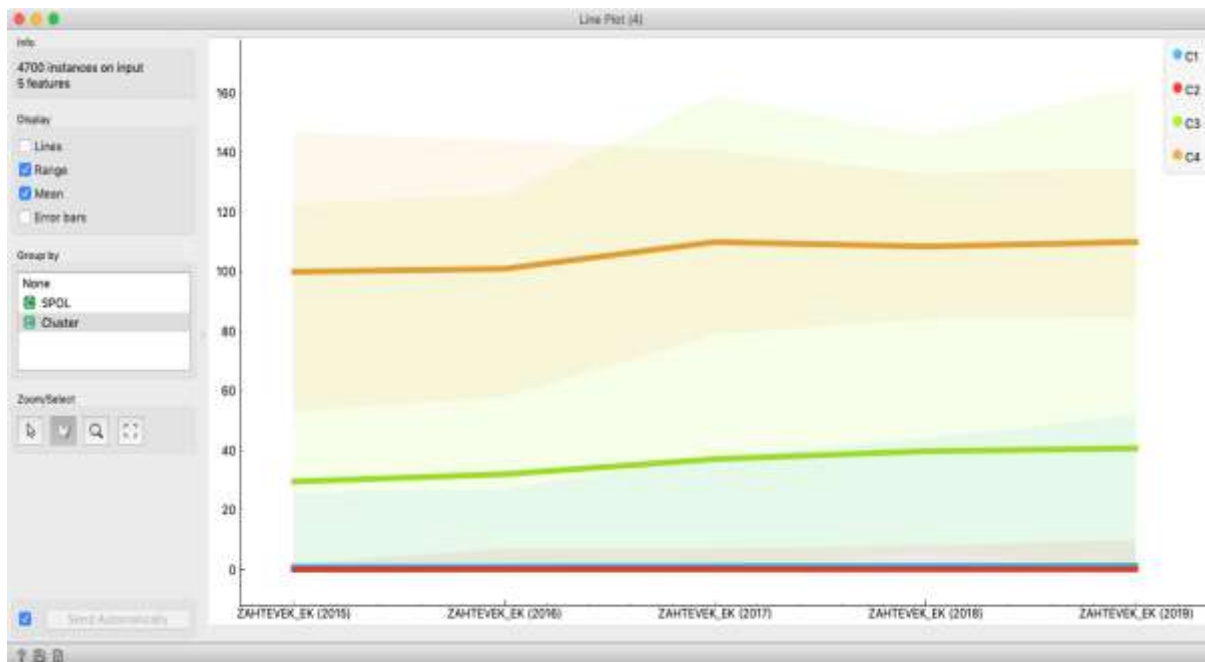


Slika 9 | Na hitro zgleda, da so v C1 gruči tisti nosilci kmetij, ki ne vlagajo EK zahtevkov, v C2 pa obratno. Ker nas zanimajo tisti, ki ne vlagajo EK zahtevkov, gledamo naprej (iščemo razloge, zakaj in kateri kmetje na določenih območjih ne vlagajo zahtevkov EK → potencial). S tega vidika je zanimivejše gručenje na 4 skupine.



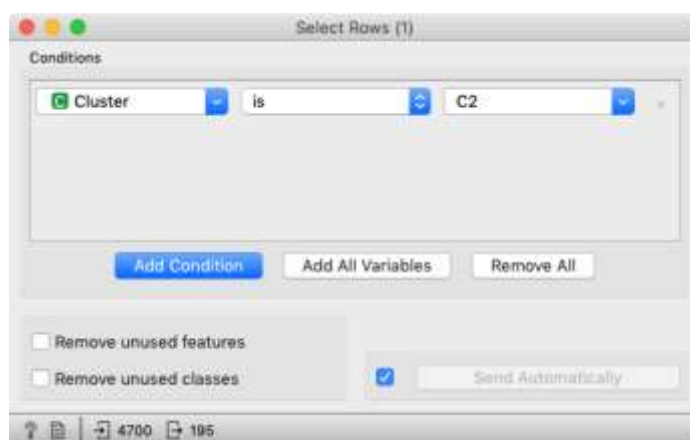
Slika 10 | Gručenje na 4 skupine.

Slika 11 | Ponovno prikažemo trend števila vloženih EK zahtevkov skozi leta, le da je tokrat grupiran po grućah (prikazana sta povprećje in domet).

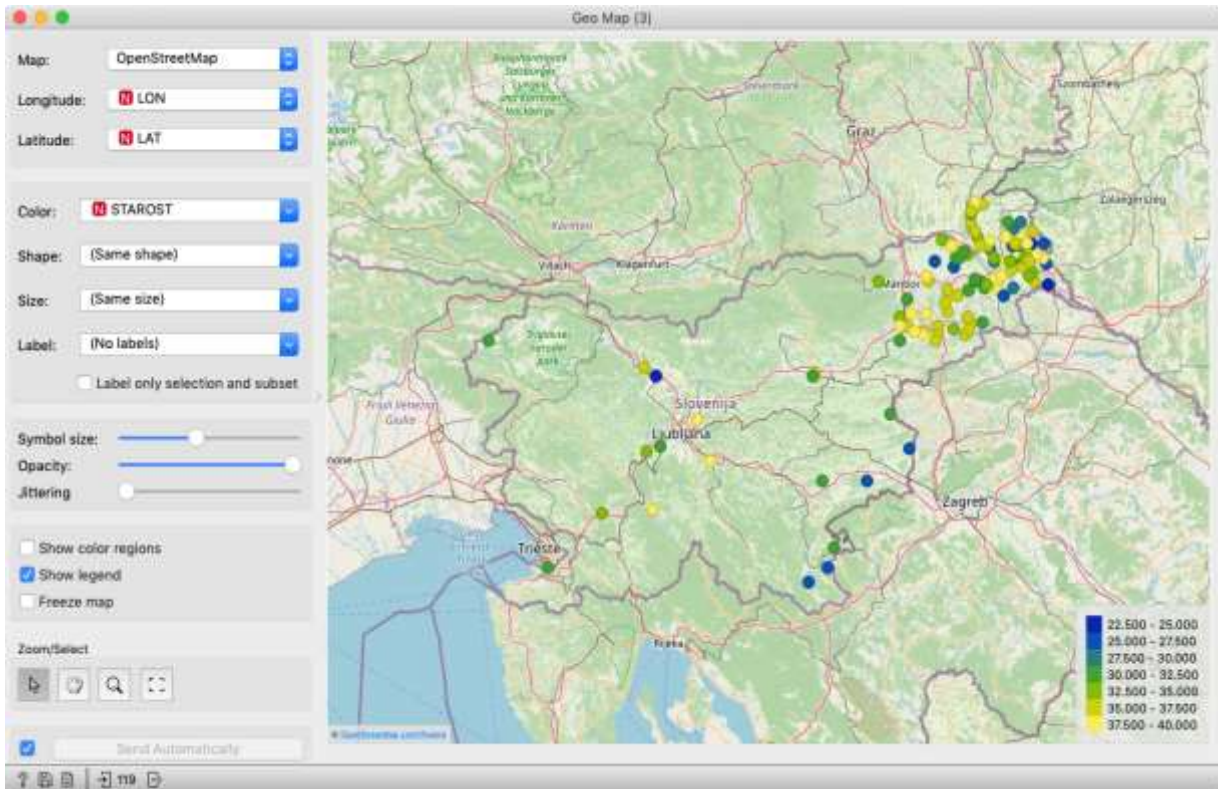


Razberemo lahko, da Oranćna skupina (C4) vlaga veliko zahtevkov EK – tu verjetno ni već veliko potenciala za preprićevanje v Ńe bolj EKO usmerjenost. Nasprotno pa Rdeća skupina (C2) ne vlaga EK zahtevkov ali pa vlaga zelo, zelo malo - zato nas za reŃitev naŃe naloge ta skupina zanima, saj predstavlja potencialne nove kmete, ki bi se lahko vkljućili v ukrep. Vendar pa na to vpliva veliko Ńtevilo faktorjev, ki jih je potrebno analizirati in poiskati vzroke za takŃno stanje, ter v nadaljevanju poiskati vzorce (interakcije med ukrepi, struktura kmetij, starostna/ izobrazbena struktura, okoljske danosti, velikost kmetij ipd).

Slika 12 | Kot izhaja iz Slike 11 v skupini C2 skoraj nihće ne vlaga zahtevkov EK, pogledjmo jih podrobneje:

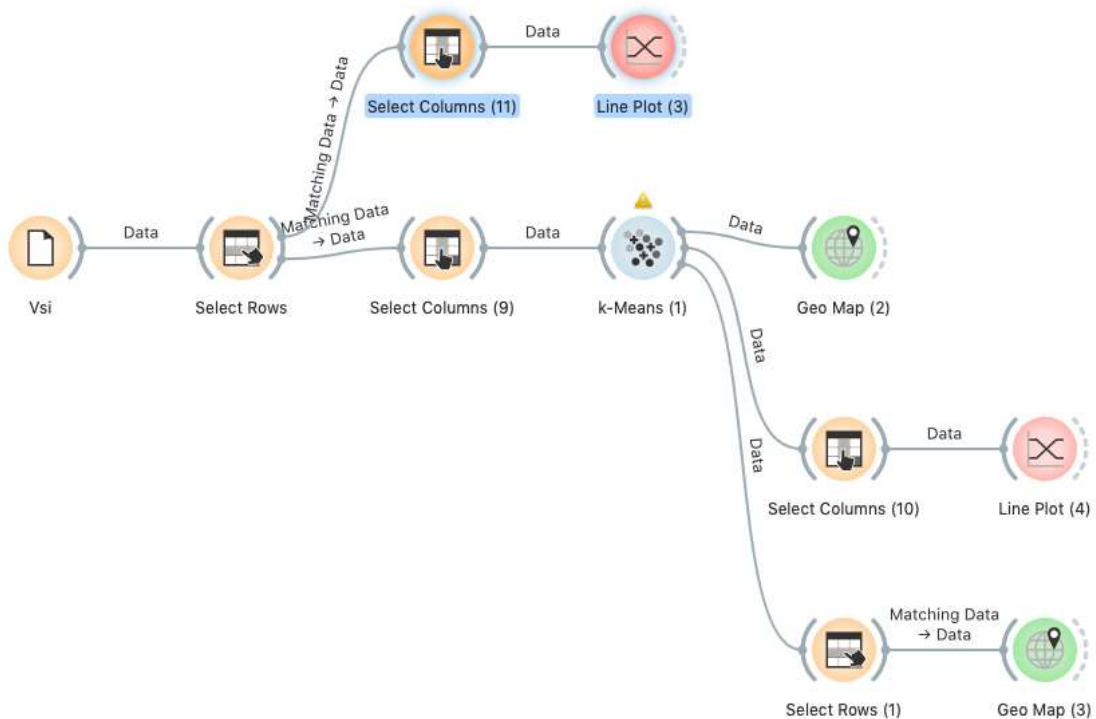


Slika 13 | Izberemo le tiste nosilce, katerih kmetije so v skupini C2. Prikažemo jih na mapi, kjer barva točke pove starost nosilca kmetije, razdeljene v 7 kategorij. Ali smo našli deficitarno območje (Hiter pogled na zemljevid Slovenije → Prekmurje, natančneje Pomurje in ptujsko polje)?

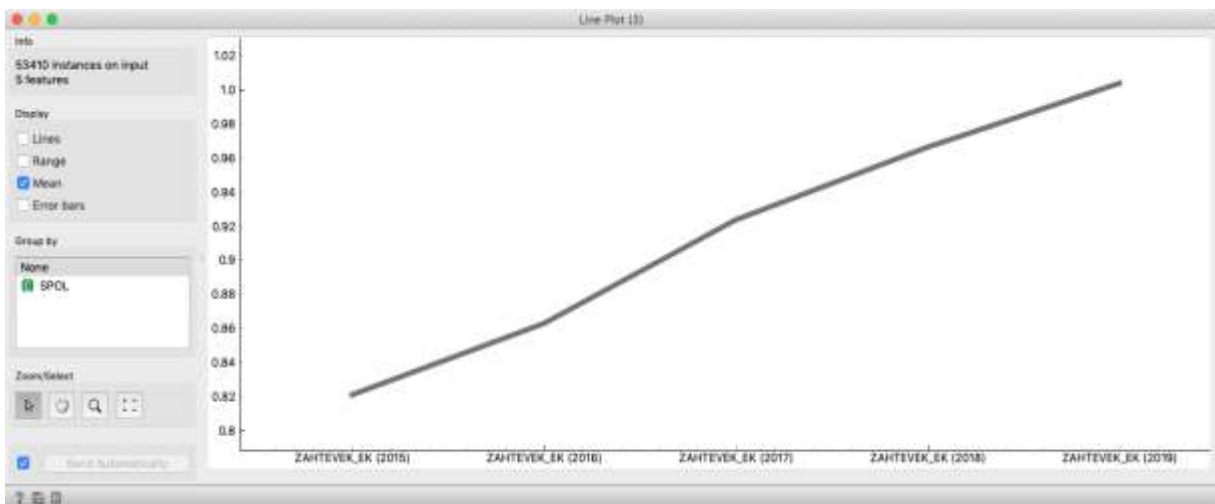


Analiza vseh kmetij

Slika 14 | Naredimo enak poskus še za vse kmetije (ne glede na starost nosilca).

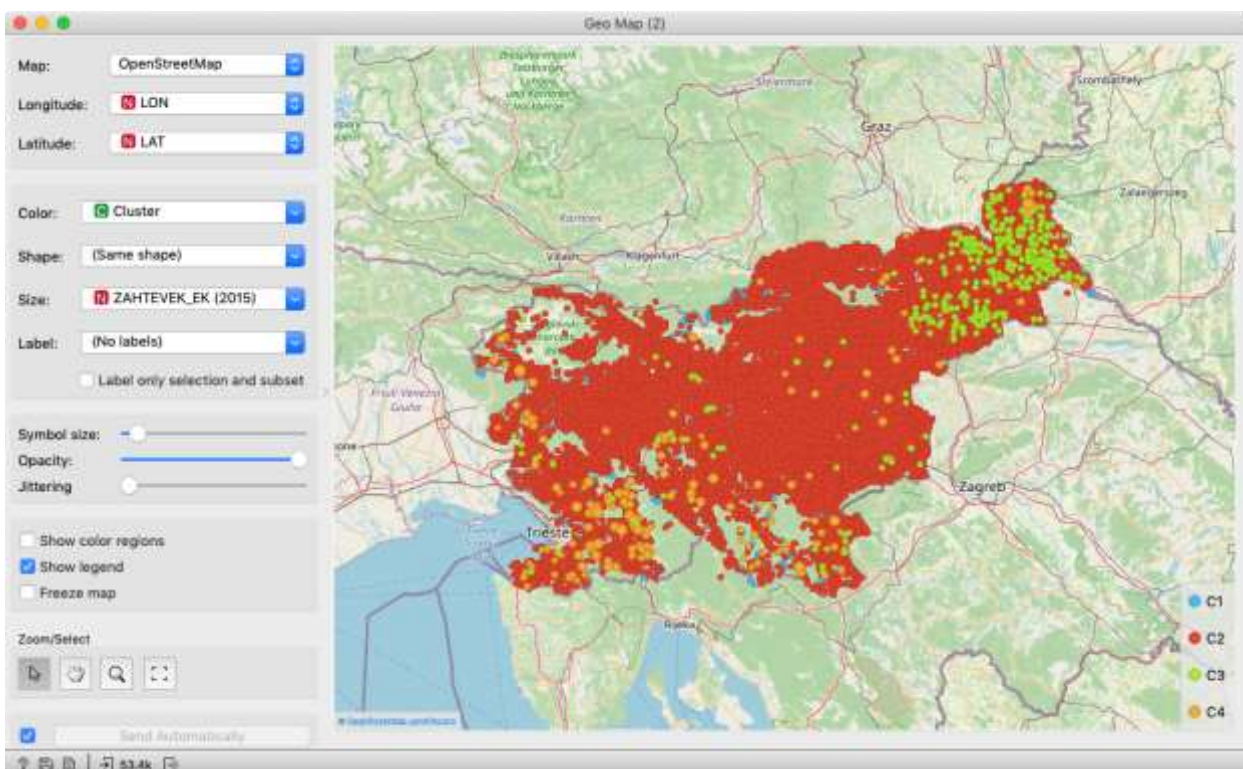


Slika 15 | Trend vlaganja EK zahtevkov je podoben, kot pri mladih kmetih (le vrednost je manjša).

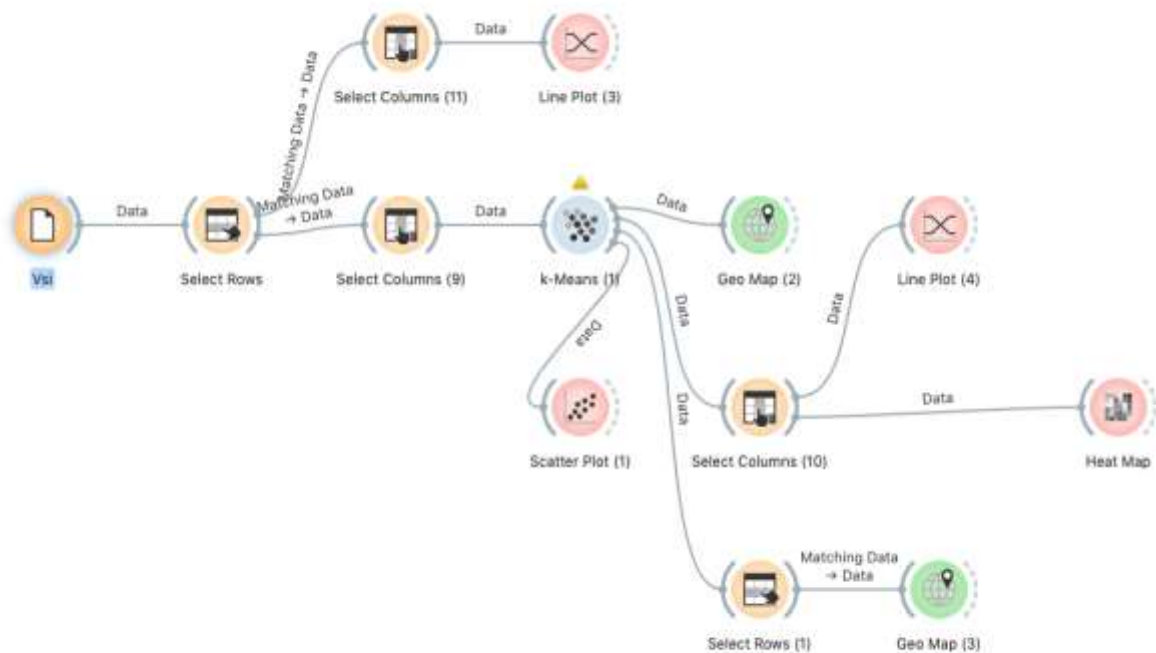


Na x osi je prikazan zahtevek EK po letih, na y osi pa število zahtevkov EK. Štejemo samo tiste KMG, ki so vlagali v vseh letih, ki so predmet analize. Trend vlaganja zahtevkov je podoben kot pri mladih kmetih, ki vlagajo EK, le vrednost je nekoliko nižja.

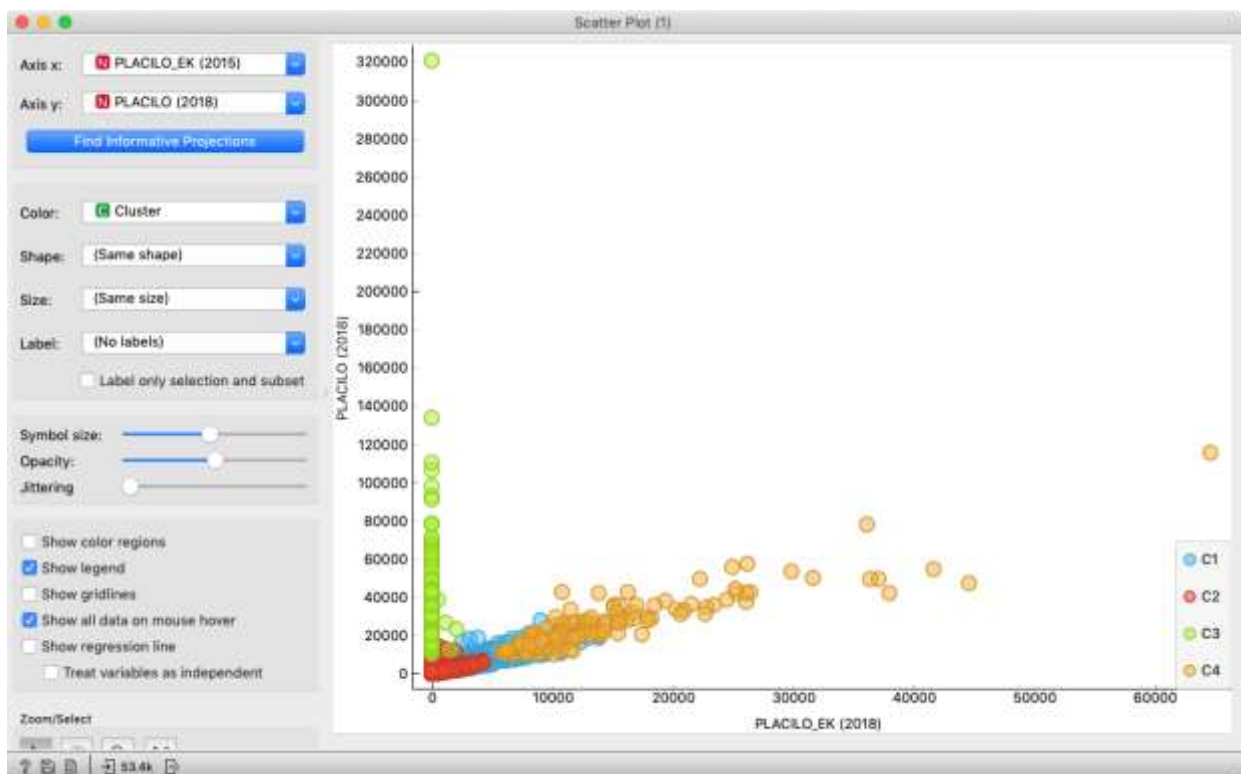
Slika 16 | Pred gručenjem podatke normaliziramo. Gručenje na 4 skupine. Pregledamo rezultate gručenja z Geo Map gradnikom :



Slika 17 | Poglejmo kaj te štiri skupine so:

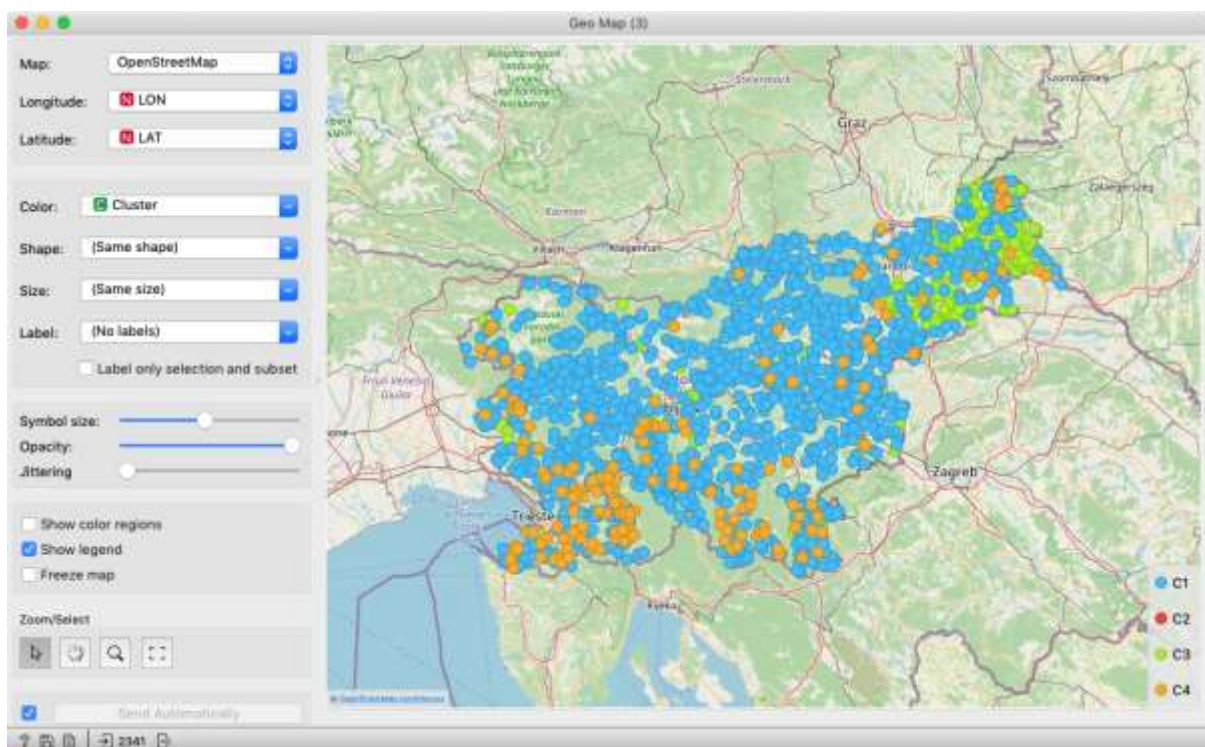
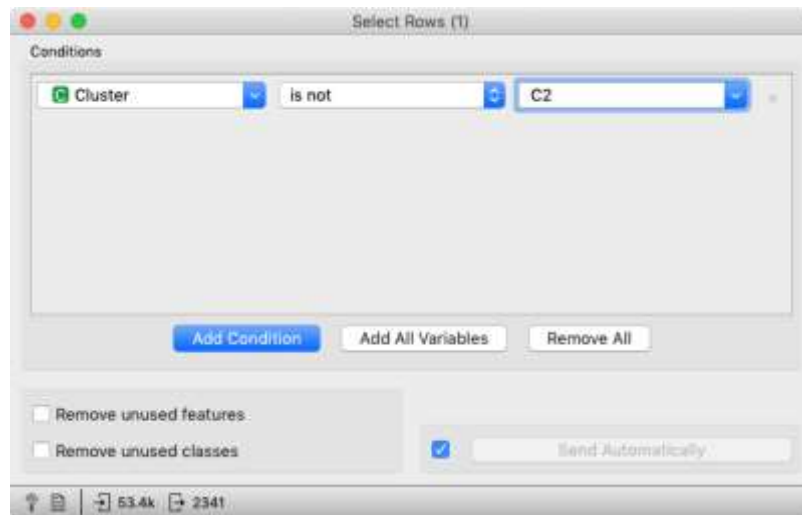


Skupin pravzaprav še ne razumemo, zato gledamo naprej. Pregledamo jih s pomočjo razsevnega diagrama



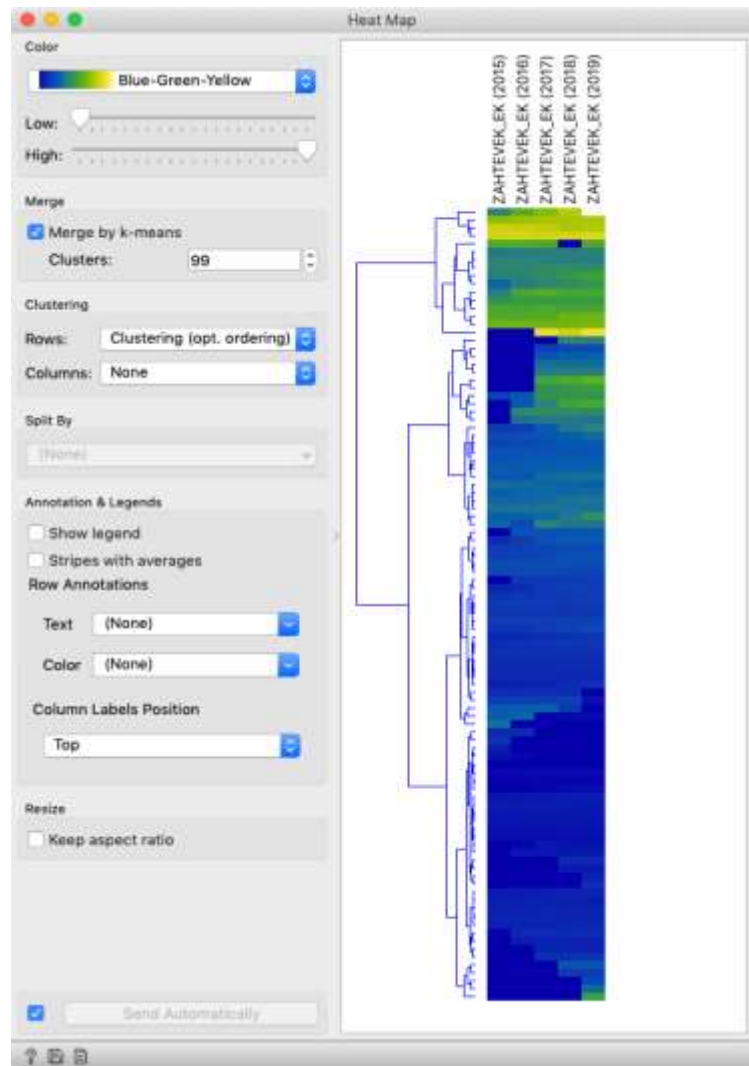
Slika 18 | Razsevni diagram opiše pomen štirih skupin. »Zelena skupina« (C3) NE vlaga zahtevkov EK, dobi pa veliko subvencij na račun ostalih vloženih zahtevkov iz naslova subvencij (drugi ukrepi).

Slika 19| Za začetek odstranimo »rdečo skupino« (C2), ki je največja in prejme najmanj subvencij in morda ni tako zanimiva, ker se take kmetije nahajajo povsod.

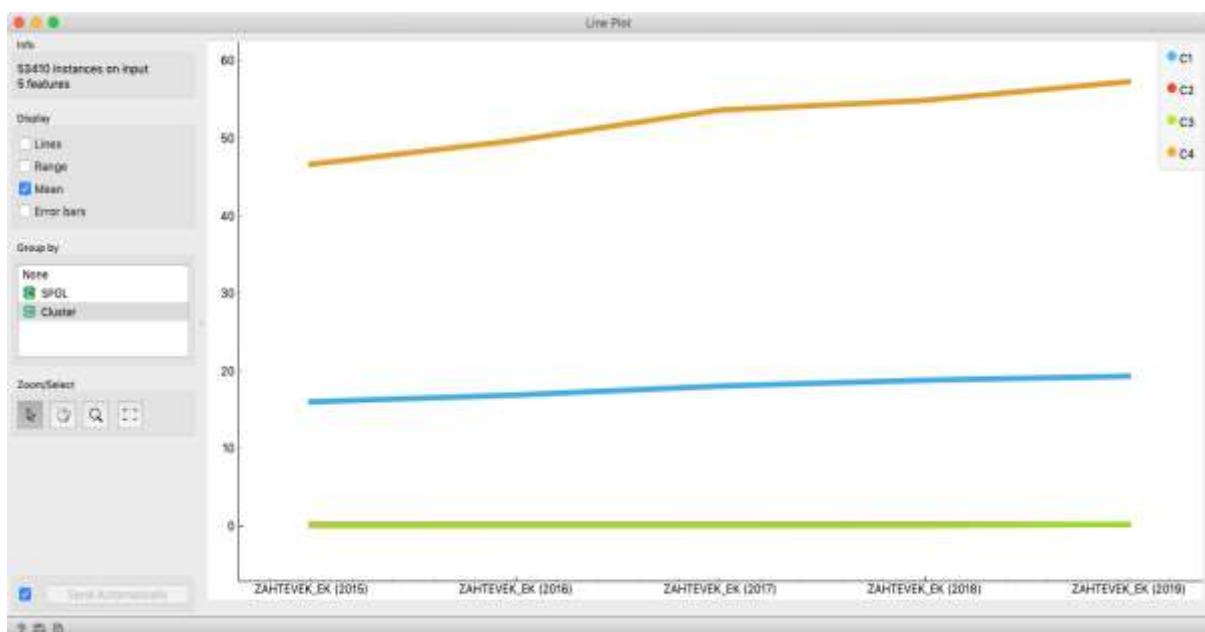


Slika 20| Še prikaz na karti.

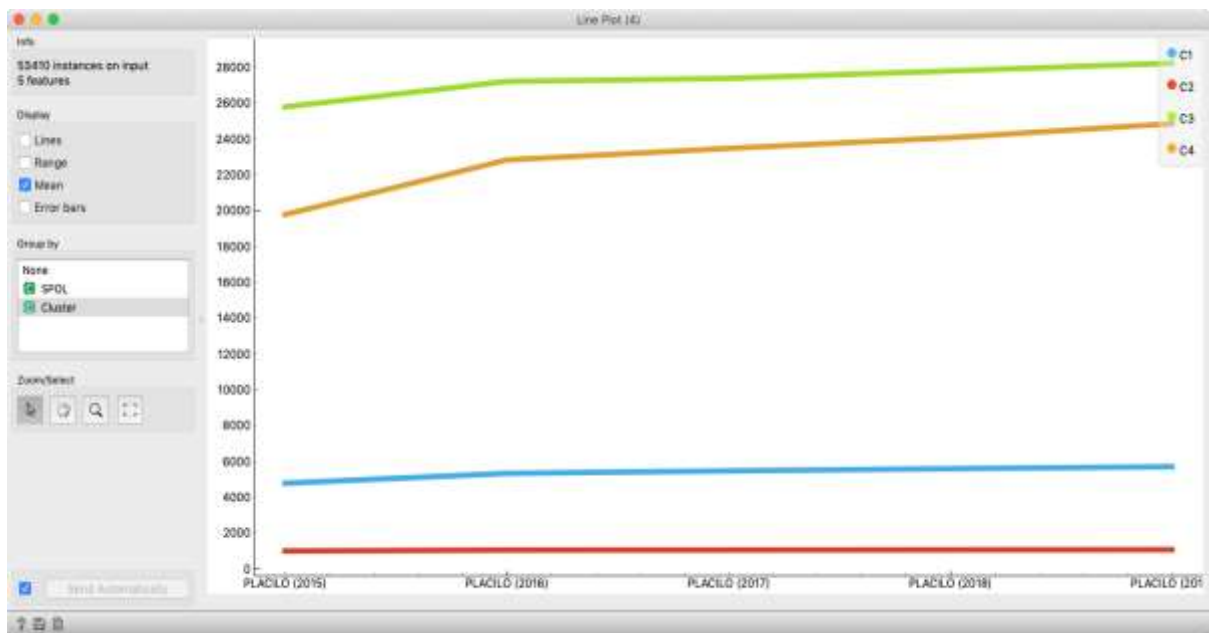
Slika 21 | Poglejmo še eno zanimivo vizualizacijo trenda vlaganja EK zahtevkov. Na sliki vsaka vrstica prikazuje eno kmetijo, barva pa pomeni število zahtevkov. Kmetije lahko označimo in pregledujemo dalje.



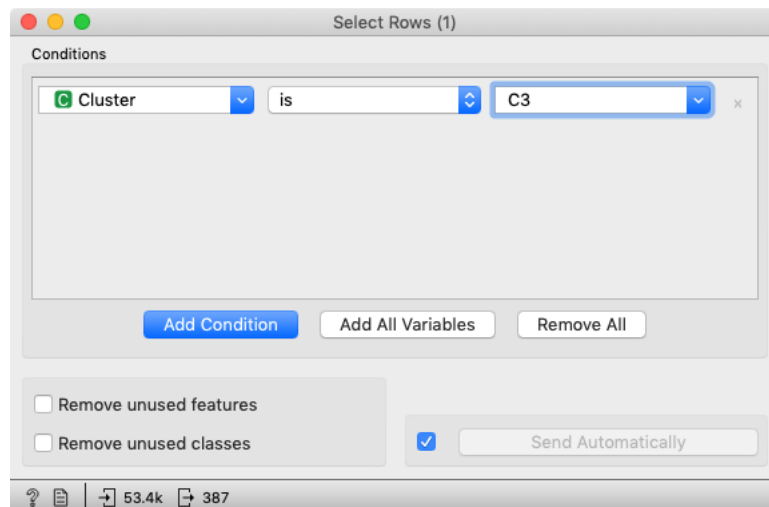
Slika 22 | Nadaljujemo razlago razsevnega diagrama. »Zelena skupina« (C3) ne vlaga EK zahtevkov, dobi pa veliko subvencij na račun ostalih (to je vidno na razsevnem diagramu, potrdimo pa s pregledom naslednjih grafov).



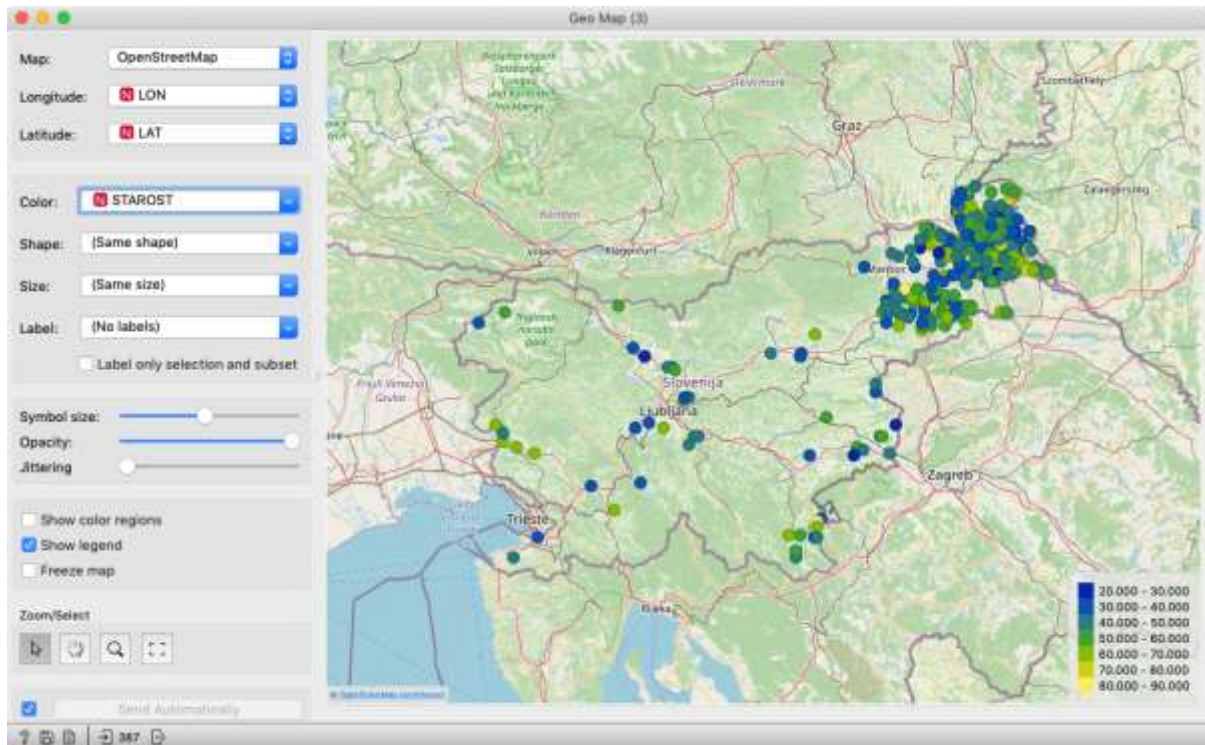
Slika 22 | Plačila skozi leta.



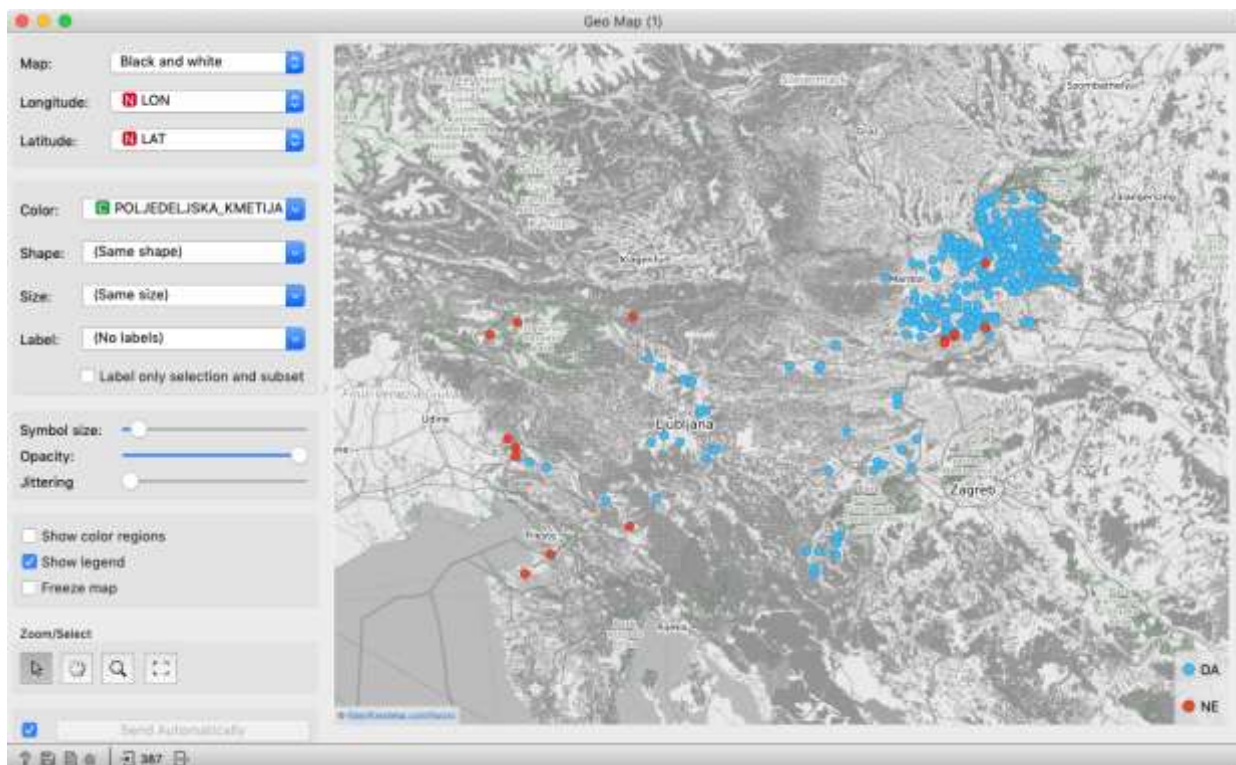
Slika 23 | Poglejmo kje se te kmetije nahajajo. Izberimo le skupino C3.



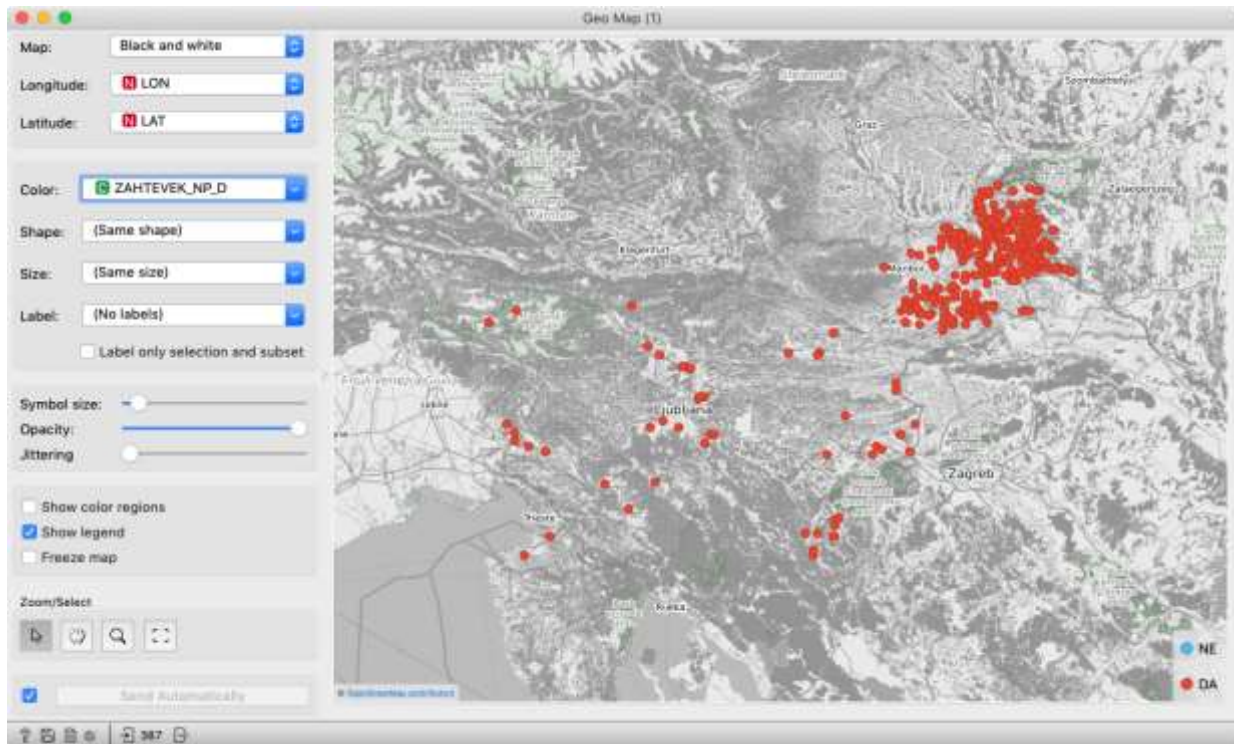
Slika 24 | Točke pobarvavimo glede na starost nosilca kmetije. Ne zdi se na bi starost kaj dosti vplivala na vlaganje zahtevkov EK (brez omejitve starosti nosilca KMG, 7-starostnih kategorij).



Slika 25 | Poglejmo tipe teh kmetij (glede na definirane osnovne 4 tipe kmetij) . Večinoma gre za tip poljedelske kmetije. /Prikaz tipov kmetij po želji/



Slika 26 | Kam (v katere ukrepe) ti kmetje vlagajo zahteve za subvencije? Vse KMG vlagajo v NP, večina jih vlaga v KOPOP, nekatere v OMD, le peščica pa v EK. (Ker smo gledali obdobje 2015 – 2019 npr, kmet, ki je samo eno leto vložil zahtevek EK, pade v populacijo, ki vlaga EK, tudi če je ta iz naslova EK pridobil min sredstva in v drugih letih ni vložil nobenega zahtevka EK. Prav zato te tudi izpostavljam. Lahko se analiza nadaljuje – zakaj).



Zaključek

Za zaključek odgovorimo na zastavljeno vprašanje: Ali napovedan trend vlaganja zahtevkov EK omogoča doseg cilja: pod ekološko pridelavo vsaj 25% kmetijskih površin do leta 2030?

Predstavimo ugotovitve, kot npr. zaskrbljujoče je, da starost nosilca KMG ne vpliva na vlaganje zahtevkov EK. Ugotavljamo, da brez ukrepov in celostnega ukrepanja ni pričakovati sprememb. Ne glede na to, da je bil *primer naloge zastavljen pred sprejetjem zaveze na ravni EU, da bodo zahteve strategij prenesene v zakonodajni okvir SKP (še niso), in s tem za DČ obligatorne, so nam rezultati in predstavljeni trendi lahko v pomoč in usmeritev pri naslavljanju intervencije iz naslova Ekološkega kmetovanja v prihodnji SKP.*

Ugotovitve so lahko berljive, rezultati pa predstavljeni, saj vsakokratni prikaz zemljevida omogoča lažjo percepcijo (vizualizacija rezultata).

Naloga lahko služi kot uvod v nadaljnje, targetirane vsebinske naloge, izdelane na podlagi preučitve stanja /izhodišče/, opredelitve potreb /prioritizacija/ in možnosti za doseg zahtevanega stanja /rezultat/.

Potrdila se je ugotovitev, da je opredelitev podatkov in njihova prioritizacija pri poskusih črpanja in urejanja podatkov ter pilotnega načrtovanja shem za podatkovno analitiko ključna. Nov izvedbeni

model temelji na principu voditeljstva. Za učinkovito uporabo orodja za podatkovno rudarjenje je razdelitev vlog, lahko tudi na nivoju oblikovanja skupine, ključna. Usklajen temeljni načrt (konsenz!) ciljnih zahtev/vrednosti ne sme biti predmet nenehnega vmesnega spreminjanja posameznih parametrov in celo temeljnih ciljnih zahtev, saj le-to povzroči resne spremembe z (ne)obvladljivimi posledicami. Te imajo nadalje vpliv in sovpliv na posledice povezanih aktivnosti/procesov, kar rezultat (časovno, finančno, vsebinsko) oddaljuje od začrtane strategije in dosege cilja (poraba resursov je neminovno nesorazmerno višja od pričakovanj naročnika/ uporabnika).

Pridobivanje surovih podatkov iz obstoječih podatkovnih baz (in posameznih aplikacij) zahteva skrbno obdelavo, pri čemer je potrebno nenehno medsebojno komuniciranje in interdisciplinarno sodelovanje najmanj strokovnjakov s področij vsebine (kmetijstvo, živinoreja, naravovarstvo, ...) in področja IT (poznavanje posameznih aplikacij, podatkovna analitika, vzdrževanje shem, uporaba ključev, dostopnost podatkov, programski vmesniki) in odločevalcev - vodstva. Potrebno je poznavanje procesov organizacij/e (AKTRP) in toka podatkov (od prejema izvirnega podatka, t.i. zajem podatkov, vgrajenega v posamezno aplikacijo – obdelave podatka za specifičen namen – do izhodnih podatkov, opredeljenega kot rezultat). Nujno je opremljanje in opolnomočenje kadrov s podatkovnimi znanji in kompetencami. Nadaljnje predobdelave podatkov za namen takega tipa naloge so bile izjemno zahtevne.

Takšna analitika omogoča še:

- iskanje »tipičnega kršitelja« glede na razlike med SUBV vlogami in prejetimi sredstvi ter rezultati kontrol,
- hiter vpogled, koliko KMG bi bilo potencialnih upravičencev za določen ukrep/intervencijo glede na kriterije,
- hiter prikaz korelacij med ukrepi in identifikacijo razlogov za takšno stanje, s prostorskim prikazom (npr SV Slovenija: KMG z njivskimi površinami raje izberejo operacije KOPOP (POZ, VOD) kot EK, saj so zahteve lažje dosegljive in znesek SUBV višji) → ukrepanje?
- usmerjanje izobraževanj, svetovanj, informiranja, ...,
- planiranje virov: znanje, lokacija, kompetence, ... realna časovnica, ... ipd;
- celostno obravnavo Slovenije za načrtovanje optimalno izbranih in naravnanih intervencij, ki bodo zadostile pogojem zakonodajnega okvirja in omogočile implementacijo s pozitivnim učinkom za kmete, potrošnike in varstvo okolja.

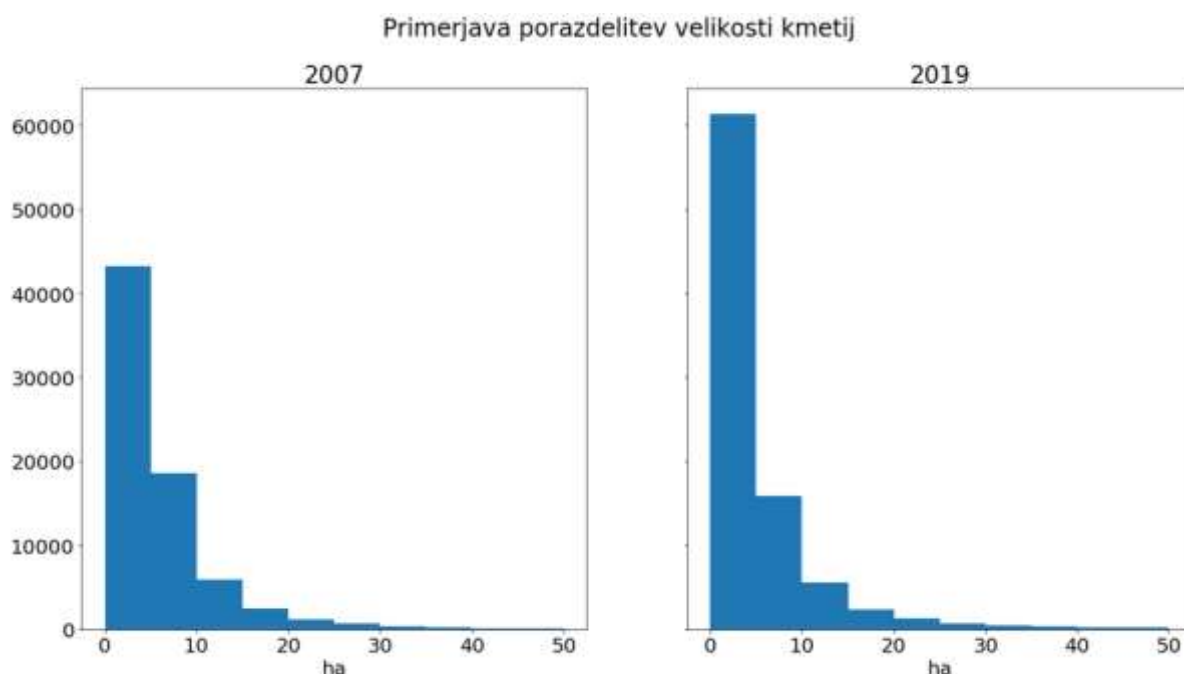
Zlasti iz razloga, ker orodja danes omogočajo tudi zelo dobro argumentacijo nenamerno ali namerno izdelanih zaključkov in njihovo interpretacijo, če jih uporabljajo uporabniki, ki ne poznajo/ne poznajo dovolj vsebine ali osnov analitike bodisi želijo najti dokaze za apriori določene trditve bodisi so v vlogi zastopnikov različnih interesnih skupin, je pri izvedbi nalog na področju podatkovnega rudarjenja/analitike in predstavitvi rezultatov, ugotovitev, sklepov in napolnil potrebna stalna vključenost in komunikacija domenskih ekspertov – skrbnikov baz in vsebinec z IT strokovnjaki ter odločevalsko strukturo.

Eksperiment 2: Analiza napovedljivosti opuščanja kmetijskih površin (Andreja Kovačič)

Cilj te naloge je analiza napovedljivosti opuščanja kmetijskih površin, torej ali se napovedovati sploh da. Uspešnost modeliranja, in s tem napovedi, je v veliki meri odvisna od podatkov, ki jih imamo. Zato smo se v tej nalogi osredotočili samo na 'pogerkanе' površine, kajti samo o njih imamo dovolj podatkov. Na voljo smo imeli podatkovne zbirke iz MKGP in ARSKTRP, za obdobje več let. Izkaže se, da je opuščanje površin napovedljivo, vendar se ga da še precej izboljšati.

Podatkovne zbirke

Uporabili smo podatke o GERKih, ki so poleg lokacije in površine vsebovali še podatke o obdelovalcu-nosilcu KMG (fizična/pravna oseba, letnica rojstva, spol), rabi, povprečnem naklonu, povprečni ekspoziciji in povprečni nadmorski višini. Sloj dejanske rabe smo uporabili za potrjevanje podatkov v GERKih in za ugotavljanje rabe površin, ki so bile izločene iz registra GERKov. Iz ARSKTRP smo pridobili podatke o staležu živine skozi leta in podatke o različnih izplačilih (OMD, DŽ, EK, EKO0, KOPOP, KOP, SKOP). Vse uporabljene zbirke so vsebovale podatke od 2007 do 2019.



Slika 1 | Primerjava porazdelitev velikosti kmetij pokaže, da se število majhnih kmetij skokovito povečuje. Dolgi rep v obeh letih kaže na majhno število velikih kmetij. Graf je lahko nekoliko zavajajoč, saj lahko narobno sklepamo, da je kmetijskih površin v Sloveniji vedno več. Vemo pa, da je v zadnjih letih ta trend negativen, rast pa pripišemo vpisovanju že obstoječih KMG v register.

Obdelava podatkov

Pri agregaciji podatkov v profile kmetij smo naleteli na nekaj težav. Velik del podatkov od leta 2007 do 2011 smo takoj izločili, saj podatki o GERKih vsebujejo preveč šuma. Izločiti smo morali tudi atribut ekspozicije GERKa, saj zaradi načina beleženja preveč podatkov manjka. Za kategorizacijo opuščene površine smo uporabili preseke z dejansko rabo. Dejanska raba se v Sloveniji zajema po kosih, približno tretjino površja na leto. Da bi se izognili primerjanju registra GERK z dejansko rabo, kjer bi bili posamezni kosi že zastareli, smo raje uporabili dejansko rabo, ki ima vse slike zajete po letu, iz katerega so podatki o GERKih. Zaradi tega lahko GERKe prehitro označimo kot zaraščene. Posledično smo iz obdelave izločili tudi leta 2017 do vključno 2019, saj za njih nimamo dovolj svežih slik dejanske rabe.

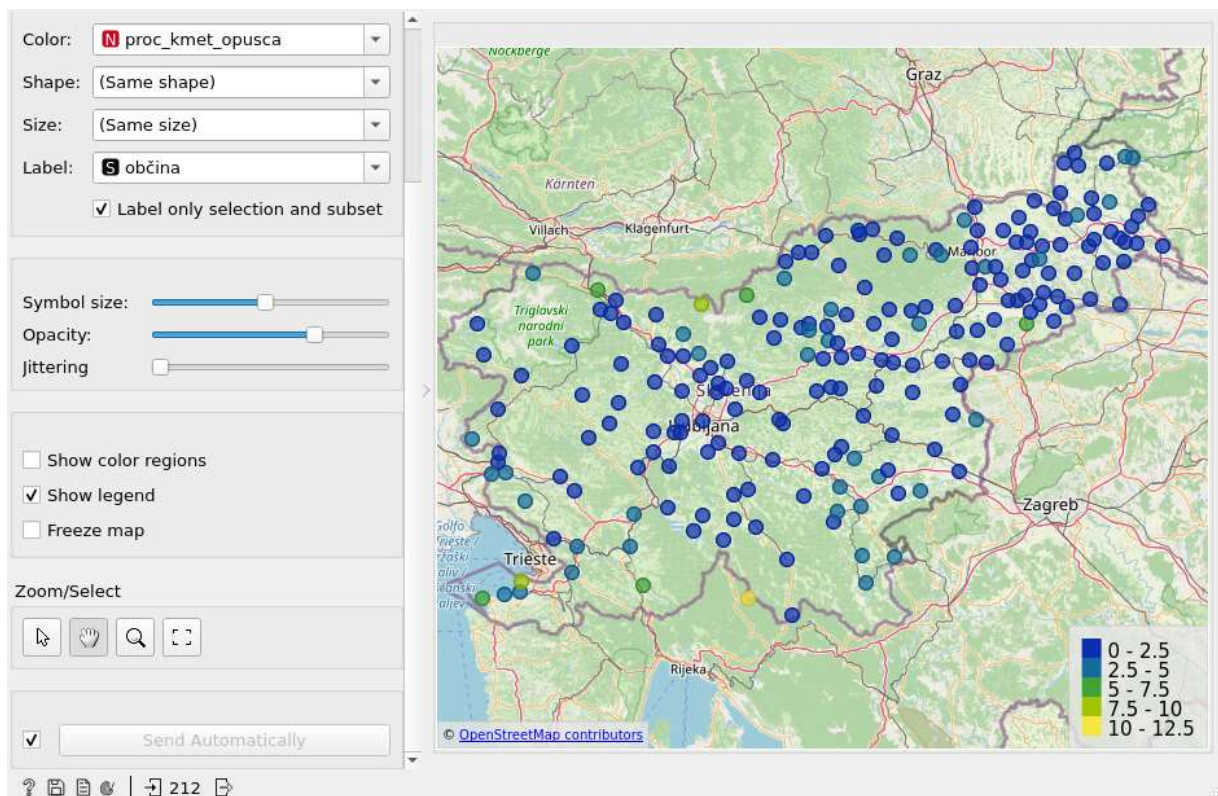
Kot zaraščene smo označili GERKe, ki imajo več kot 15 ara površine označene s šiframi dejanske rabe 1410, 1500, 1600, 2000. Preseke smo naredili z vektorskimi preseki, s pomočjo knjižice shapely. Za kmetije smo agregirali atribut, ki označuje opuščanje. Če je vsaj ena površina označena kot opuščena, velja, da kmetija opušča površine. Sam kriterij je sicer zlahka prilagodljiv, najbolje ga je oblikovati v sodelovanju z domenskimi eksperti.



Slika 2 | Prikaz postopnega zajema dejanske rabe skozi leta.

leto	# gerkov	% opuščenih gerkov	# kmetij	% kmetij, ki opušča
2011	689780	0.5	56753	4.61
2012	688555	0.43	56271	3.71
2013	687377	0.4	56116	3.77
2014	687862	0.53	56018	4.94
2015	566903	0.2	45993	1.99
2016	568236	0.2	46118	2.08

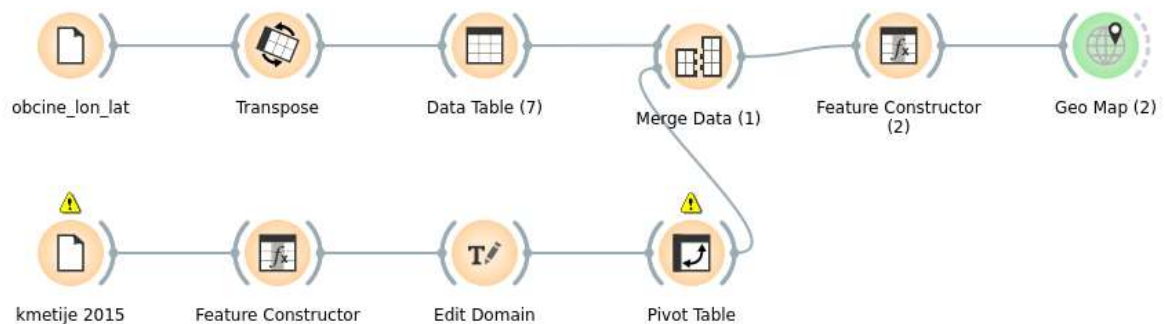
Slika 3 | Osnovne statistike združenih podatkov iz baz ARSKTRP in MKGP. Nenaden padec števila kmetij izvira iz baze izplačil kmetijam skozi leta (v letih 2014-2015 gre za prehod med dvema cikloma politike, zbirke podatkov ARSKTRP pa so vezane na posamezen cikel). Za same analize smo uporabili podmnožico kmetij, tiste ki so vsebovane tudi v bazi GVŽ, tako je število bolj konsistentno. Hkrati smo napovedovali opuščanje samo za posamezno leto, saj so časovne vrste zaradi nekonsistenc v podatkih manj zanesljive.



Slika 4 | Kmetije ki opuščajo površine, agregirane po občinah, za leto 2014. V večini občin opušča površine manj kot 2.5% kmetij. Nekaj občin sicer izstopa, npr. Solčava, Jezersko, Osilnica, Ankarana. Razlike v občinah nakazujejo na možno koristnost občine kot značilke kmetije. Na sliki so imena občin zavoljo berljivosti izpuščena.

Modeliranje

Primerjali smo nekaj osnovnih modelov: naključni gozdovi, linearna regresija in naivni Bayes. Slednja dva sta posebej zanimiva, saj omogočata enostaven vpogled v vpliv atributov na napovedi modela.



Slika 5 | Večina modeliranja smo izvedli v programu Orange. Obdelava podatkov, ki je zajemala filtriranje, zlivanje in agregacijo podatkov pa je zahtevala podatkom specifične skripte. Na sliki je primer osnovnega delokroga.

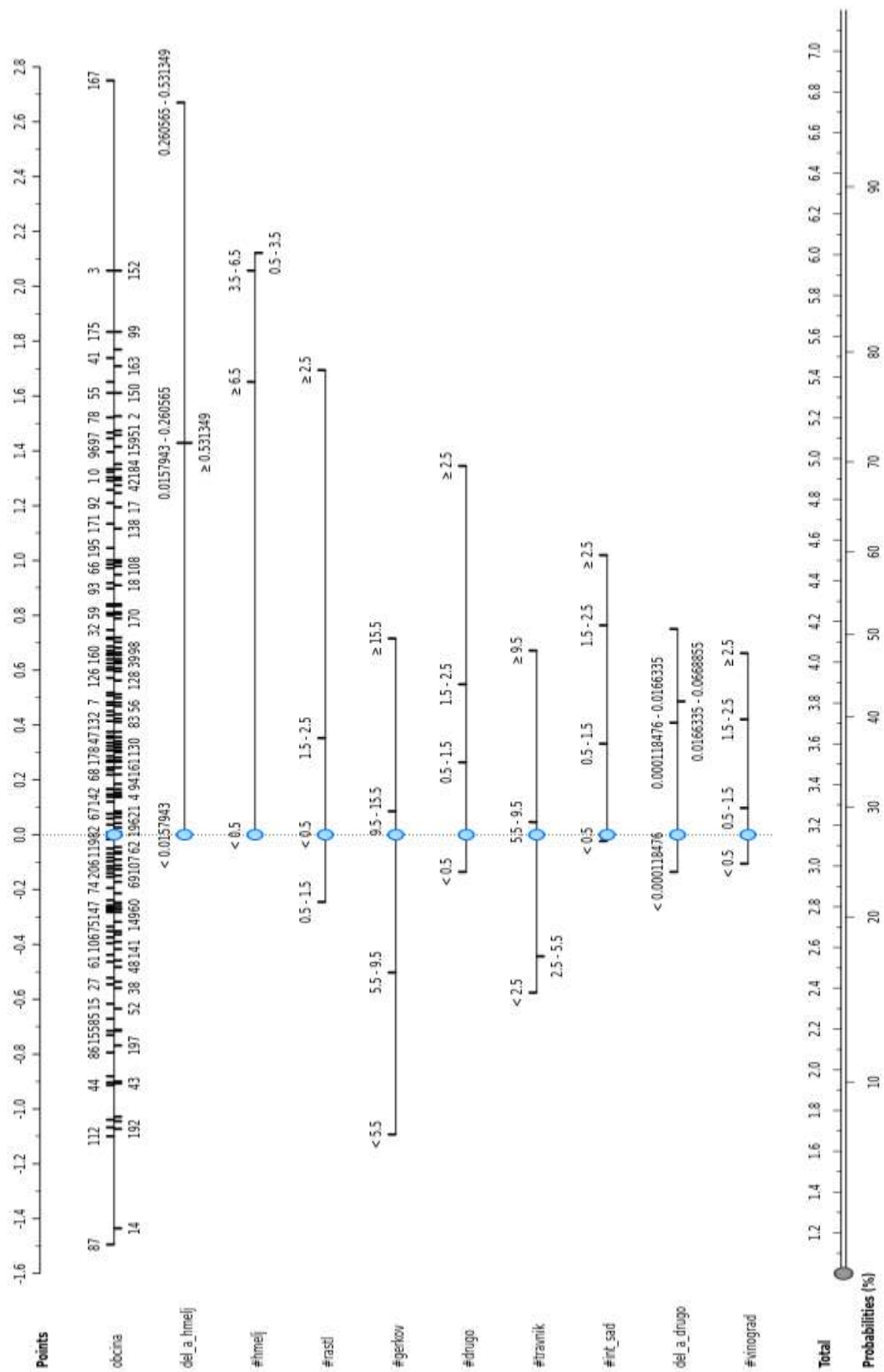
Vrednotenje

Profile za posamezno leto smo razdelili v dve neodvisni množici. Na prvi, ki je zajemala približno 70% podatkov, smo modele učili in jih nato testirali na preostanku podatkov. Ta postopek smo ponovili večkrat (10) in rezultate povprečili.

Optimizirali smo AUC, t.i. površino pod krivuljo, ki se lahko potem optimizira glede na potrebe končnega uporabnika. Model se potem lahko uravnava, glede na to, koliko kmetij, ki opušča površine, smo pripravljene zgrešiti, v korist točnosti napovedi opuščanja. AUC ima razpon od 0 do 1, naivni bayes in logistična regresija dosejata AUC okrog 0.7, kar potrjuje napovedljivost, hkrati pa je še vedno precej možnosti za izboljšanje.

Sampling		Evaluation Results			
<input type="radio"/> Cross validation		Model	AUC	CA	F1
Number of folds:	10	Random Forest	0.598	0.984	0.976
<input checked="" type="checkbox"/> Stratified		Naive Bayes	0.713	0.976	0.973
<input type="radio"/> Cross validation by feature		Logistic Regression	0.697	0.984	0.977
		Constant	0.500	0.984	0.976
<input checked="" type="radio"/> Random sampling					
Repeat train/test:	10				
Training set size:	70 %				
<input checked="" type="checkbox"/> Stratified					

Slika 6 | Napovedne točnosti različnih modelov. Najboljše se odreže naivni Bayes, vendar bi se rezultat dalo še izboljšati.



Slika 7 | Naivni Bayesov nomogram, atributi so urejeni po pomembnosti, padajoče. Največji vpliv ima občina, iz katere je kmet in sama sestava kmetije: delež hmelja, število površin z neko rabo, število površin. S pomočjo nomograma lažje razumemo napovedi za posameznega kmeta.

Možne izboljšave

Prej opisan problem postopnega zajema dejanske rabe se da omiliti z združevanjem slojev iz različnih let dejanske rabe. Tako bi se izognili delu prehitrega označevanja površin kot zapuščenih. K uspešnosti napovedovanja bi pripomogle dodatne značilke iz drugih baz, na primer podrobnejši podatki o poljščinah. Potencial za izboljšavo je tudi oblikovanju kriterijev opuščanja in pri oblikovanju profilov samih, pri čemer je nujno sodelovanje domenskih ekspertov. Zanimiv je tudi pogled na kmete in kmetijske površine z vidika analize preživetja.

Delovna naloga 5.1: opis in arhiviranje podatkovnih zbirk (Brigita Bočkaj, dr. Janez Štebe)

Arhiv družboslovnih podatkov (v nadaljevanju ADP) je v okviru delovnega sklopa 5 izvedel dvoje nalog. Prvič, organiziral je pripravo metapodatkovnega opisa izbranih evidenc, ki jih vodi Ministrstvo za kmetijstvo, gozdarstvo in prehrano (v nadaljevanju MKGP). Ter drugič, izvedel je prevzem in objavo podatkovnih zbirk, ki so jih za prikaz analitičnih pristopov v okviru projekta pripravili na Agenciji za kmetijske trge in razvoj podeželja (v nadaljevanju ARSKTRP), v Katalogu ADP.

Potrebno je omeniti, da gre v primeru obravnavanih zbirk za tematsko specifične mikro-podatke s področja kmetijstva, medtem ko je metapodatkovni opis pripravljen po strukturi Data Documentation Initiative (DDI), ki je uveljavljen pri opisovanju družboslovnih podatkov. S tem, ko so podatki pripravljeni na tovrsten način, se jih bo približalo širši množici potencialnih uporabnikov, ki lahko do njih dostopajo na pregleden način.

Podatki so na voljo za uporabo tako notranjim uporabnikom MKGP in ARSKTRP kot zunanjim uporabnikom, raziskovalcem, študentom in drugim. S to nalogo smo preverili in pokazali, da metapodatkovni opis po strukturi DDI ponudi izčrpen in informativen vpogled v podatke, tako da jih je mogoče poiskati, se seznaniti z obstojem podatkov, kako je moč do njih dostopati, oceniti njihovo uporabnost za želene analize in jih v kombinaciji z drugimi podatki uporabiti za namene presoje ukrepov politik, njihovo načrtovanje in za druge raziskovalne namene.

V nadaljevanju je predstavljen potek dela obeh nalog.

Metapodatkovni opis izbranih evidenc

Izmed številnih evidenc, ki jih vodi MKGP je bil narejen izbor 11 evidenc, za katere je ADP pripravil metapodatkovni opis. Izbor evidenc sta pripravila g. Miran Tisu, Služba za register kmetijskih gospodarstev (v nadaljevanju g. Miran) in ga. Maša Kerstein, MKGP (v nadaljevanju ga. Maša). Izbrane evidence so:

1. Evidenca dejanske rabe kmetijskih in gozdnih zemljišč
2. Evidenca kmetijskega knjigovodstva
3. Evidenca komasacij
4. Evidenca o izobraževanju in usposabljanju za potrebe kmetijstva in razvoja podeželja
5. Evidenca območij z omejenimi možnostmi za kmetijsko dejavnost
6. Evidenca organizacij proizvajalcev
7. Evidenca osuševalnih in namakalnih sistemov
8. Evidenca pridelovalcev in predelovalcev ekoloških kmetijskih pridelkov ali živil
9. Register kmetijskih gospodarstev
10. Register pridelovalcev grozdja in vina
11. Zbirka podatkov Mreže za podeželje

Metapodatkovni opis zajema več sklopov metapodatkov, pri čemer je potrebno poudariti, da je zapolnitev metapodatkovnih polj odvisna od vsebine same evidence:

- osnovni metapodatki³: naslov, avtor, kontakt, povzetek, področje, ključne besede, vsebinsko področje/tema, jezik podatkov, nosilec podatkov, datum in kraj izdelave, sodelujoči, finančna podpora, distributer, dajalec, datum izročitve, časovno pokritje podatkov, obdobje zbiranja podatkov, vrsta podatkov, podatki o seriji, povezani materiali, publikacije reference, viri podatkov
- geoprostorski metapodatki: geografsko pokritje, najmanjša geografska enota, ki je dostopna v evidenci
- metapodatki s področja metodologije izvora podatkov: enota za analizo, populacija, časovna opredelitev podatkov, zbiratelj podatkov, pogostost zbiranja, tip vzorca, metoda zbiranja podatkov, uporabljen instrument, opis situacije zbiranja podatkov, uporabljene kontrole, očiščevanje podatkov
- Pogoji: informacije o dostopu do podatkov

Metapodatkovni opis je narejen v aplikaciji Dataverse, ki je odprtokodni program za deljenje, hranjenje, citiranje, iskanje in analizo raziskovalnih podatkov (glej <https://dataverse.org/>). Dataverse uporabljajo različne institucije in organizacije po svetu. V Evropski uniji je eden največjih Dataverse repozitorijev na Nizozemskem, ki zagotavlja upravljanje s podatki 11 nizozemskih univerz (glej <https://dataverse.nl/>), podobno velja tudi za norveški Dataverse (glej <https://dataverse.no/>), medtem, ko ga Avstrijski družboslovni arhiv uporablja kot platformo za izvajanje svojih storitev (glej <https://data.aussda.at/>).

ADP aplikacijo uvaža za potrebe samoarhiviranja, kjer ustvarjalec podatkov po koncu procesiranja podatke in metapodatke samostojno preda v podatkovni repozitorij in tako deli dalje. Podobno kot pri zgoraj omenjenih primerkih nacionalnih Dataverse repozitorijev je v prihodnje predvideno uvajanje te iste enotne platforme tudi za dostop do raziskovalnih podatkov osnovnega Kataloga ADP, kjer so na voljo podatki iz sklopa trajnega digitalnega arhiviranja.

Storitev samoarhiviranja na podlagi Dataverse (glej <https://dataverse.adp.fdv.uni-lj.si/>) je ADP vzpostavil v letu 2020, pri čemer trenutno potekajo nadgradnje za prilagajanje specifičnim zahtevam in testiranja funkcionalnosti. Široka uporaba je predvidena v začetku leta 2021.

Metapodatki in ostala gradiva, povezana z raziskavo oz. projektom, v katerem so nastali podatki, so brez registracije dostopna vsem obiskovalcem spletne strani, kar pomeni, da lahko uporabnik prosto brez predhodne registracije pregleduje in išče po katalogu ter prenaša k sebi spremna gradiva. Pri slednjem so možne tudi izjeme, kjer je za spremno gradivo, zaradi njegove vsebine (tajnost, občutljivost, itd.) opredeljen drugi strožji dostop.

Dostop do raziskovalnih podatkov v večini primerov zahteva registracijo zaradi izpolnjevanja določenih pogojev, razen, kadar so podatki ponujeni v javno rabo brez vsakih omejitev. Ob prenosu podatkov se

³ Polji datum izdelave (en. production date) in datum izročitve (en. deposit date) se nanašata na podatkovno datoteko, ki pa v primeru opisa evidenc ni bila priložena, kljub temu smo polji zapolnili, pri čemer datum izdelave ponazarja datum popisa stanja evidenc, datum izročitve pa datum prvega vnosa zapisa evidence v aplikacijo Dataverse

mora uporabnik strinjati s pogoji uporabe za dotično gradivo oz. raziskovalne podatke. Uporabnik se sicer preko opisa raziskave seznanja z morebitnimi posebnimi omejitvami in izjemami glede dostopa, ki jih je določil dajalec ob predaji, dodatno se mora pri registraciji strinjati s splošnimi določili in pogoji uporabe. Pri sami registraciji se oblikujejo različni uporabniški profili, na podlagi katerih je nato omogočen dostop do podatkov, v kolikor to omogočajo določeni pogoji uporabe (npr. licenca, ki dovoljuje ali omejuje komercialno rabo).

Potrebno je poudariti, da je naveden način regulacije dostopa določen za ADP primerek Dataverse in ga je možno prilagoditi za morebitne druge primerke. Za scenarij morebitne vzpostavitve kompletne zbirke mikro-podatkov s področja kmetijstva je posebej atraktivna možnost, da uporabnik lahko tudi sam arhivira svojo raziskavo in jo ponudi v sekundarno rabo, in sicer tako, da si ustvari račun, izpolni opis raziskave ter naloži podatke in z njimi povezana gradiva. Vsakemu gradivu določi tudi pogoje dostopa.

Pri opisu evidenc smo pripravili le metapodatkovni opis, medtem ko sami podatki iz tekočih registrov zaradi varovanja zaupnosti in drugih omejitev niso priloženi. Je pa uporabnik preko podanega opisa seznanjen s poenotenimi informacijami glede pogojev za dostop do podatkov. Upravljalci evidenc lahko pripravijo mikro-podatke ob specifikaciji na zahtevo uporabnika, tako kot smo v okviru projekta določili in nato prevzeli mikro-podatke za prikaz možnih analiz.

Nekateri podatki (večinoma agregirani) in minimalni metapodatki izbranih evidenc so sicer vidni in dostopni preko različnih drugih portalov:

- Portal MKGP (<https://rkg.gov.si/vstop/>)
- Javni pregledovalnik grafičnih podatkov MKGP <http://rkg.gov.si/GERK/WebView/>)
- SiStat (<https://pxweb.stat.si/SiStat>)
- Eurostat (<https://ec.europa.eu/eurostat/home>)
- https://ec.europa.eu/agriculture/rica/database/consult_std_reports_en.cfm
- https://www.kis.si/Porocila_o_stanju_v_kmetijstvu_OEK

Za potrebe projekta CRP ITzaSKP smo na platformi Dataverse oblikovali ločen podkatalog, poimenovan *Kmetijski podatki* (glej https://dataverse.adp.fdv.uni-lj.si/dataverse/kmetijski_podatki). Podkatalog vsebuje metapodatke evidenc in sledi priporočilom evropskega podatkovnega konzorcija CESSDA z uporabo kontroliranih besedišč Data Documentation Initiative (glej <https://ddialliance.org/controlled-vocabularies>), ki se tudi sicer, kot je že omenjeno v uvodu, uporablja pri pripravi metapodatkov na področju družboslovja. Pri vpisu ključnih besed je bil, v kolikor je bilo to mogoče, uporabljen tezaver ELSST (glej <https://elsst.ukdataservice.ac.uk/>). S tem so izpolnjeni pogoji, da se bodo metapodatki prikazovali pri iskanju preko skupnega CESSDA kataloga (glej <https://datacatalogue.cessda.eu/>).

Trenutno je metapodatkovni opis evidenc narejen v slovenskem jeziku. Ob soglasju skrbnikov evidenc na MKGP bomo metapodatke evidenc prevedli, uskladili in objavili tudi v angleški različici podkataloga. Na ta način bodo evidence postale vidne in se jih bo dalo iskati tudi v širšem prostoru.

Sam metapodatkovni opis evidenc je potekal v prostorih MKGP skupaj s skrbniki posamezne evidence ob prisotnosti g. Mirana in/ali ga. Maše. Opis je bil izveden v roku dveh tednov (5 terminov - 2 oz. 3 evidence na termin), pri čemer so skrbniki pred dogovorjenim terminom prejeli seznam metapodatkovnih polj, ki so predvidena za izpolnjevanje ter povezavo do kontroliranega DDI besedišča.

Vpis za posamezno evidenco je potekal direktno v aplikacijo Dataverse, trajal pa je v povprečju od pol do ene ure, odvisno od posamezne evidence. Po opravljenem vpisu na MKGP je ADP opise pregledal in jih dopolnil (vsebinsko, spletne povezave na uradne dokumente, publikacije, itd.) na podlagi razpoložljivih vsebin na spletu.

Pripravljeni opisi so bili poslani v avtorizacijo, na način, da je bila vsebina opisa iz aplikacije pripravljena v prosto besedilni obliki z možnostjo dodajanja dopolnitev in komentarjev. Opisi so bili posredovani g. Miranu, ki je predstavljal vezni člen med skrbniki evidenc in ADP, ter predani skrbnikom v pregled. Postopek avtorizacije metapodatkovnih opisov je tudi sicer uveljavljena praksa v ADP. V nadaljevanju sta opise pregledala tudi g. Miran ter neodvisno od vseh še pregledovalec g. Matjaž Rotenhajzer, MKGP. Na sestanku v času priprave opisov, je bilo namreč dogovorjeno, da bi bilo dobrodošlo, da pripravljene opise pregleda tudi oseba, ki sicer prihaja iz stroke kmetijstva in podatke tudi vsebinsko pozna, vendar z njimi ne operira na tovrsten način in oceni, ali kot potencialni uporabnik na osnovi opisa pridobi dovolj informacij o podatkih. ADP je nato opise popravil v skladu s prejetimi komentarji ter jih ponovno posredoval v pregled, tokrat v sami aplikaciji. Ker aplikacija še ni v javni uporabi, je bilo v ta namen generirano enolično uporabniško ime in geslo, ki se ga je preko g. Mirana posredovalo vsem skrbnikom. Po preteku nekaj dni (zaradi morebitnih dodatnih popravkov) je ADP evidence objavil. V kolikor se bo v prihodnosti pojavila potreba po popravi, bo na pobudo MKGP to opravil ADP.

Samo delo je z obeh strani zahtevalo kar nekaj vložnega časa in truda: od uskladitve terminov za pripravo opisov do same izvedbe opisov, njihove dopolnitve, pregledov, poprave, objave, itd. Ker gre za kompleksne in področno specifične podatke, ADP sam ne bi mogel pripraviti ustreznega metapodatkovnega opisa, zato je bilo ključno sodelovanje skrbnikov evidenc kot poznavalcev njihove vsebine. V tem delu gre zahvala predvsem g. Miranu ter ga. Maši za angažiranost, vsem skrbnikom evidenc in g. Rotenhajzerju za dodatni pregled metapodatkovnih opisov. V samem procesu dela je bilo sicer izvedenih nekaj aktivnosti z namenom olajšanja in hitrejšega poteka dela (npr. vnaprejšnje posredovanje pripravljenih tabel zahtevanih polj, spletna povezava na besedišče, direktni vpis v aplikacijo), ki so sicer pripomogle k manjši porabi časa, vendar ne v tolikšni meri, kot je bilo pričakovano.

Prevzem podatkovnih zbirk ARSKTRP

Primarni namen generiranja podatkov je bil prikaz možne uporabe kmetijskih podatkov, za kar je bilo potrebno pridobiti čim bolj splošne podatke. Izbrani so bili podatki, ki jih vodi ARSKTRP, pri čemer gre za t. i. profile kmetijskih gospodarstev. Sam prikaz možnosti analiz so opravili kolegi iz Fakultete za računalništvo, Univerze v Ljubljani (glej ustrezno poglavje). Podatki so bili s pomočjo sql stavkov generirani (g. Matjaž Kolar, ARSKTRP ter ga. Maša) ter shranjeni v .xlsx format. Ker je bilo v pripravo podatkov vloženo kar nekaj truda in v taki obliki drugje niso na voljo, smo se odločili, da jih posredujemo v ADP, kjer bi jih ponudili v drugo rabo različnim uporabnikom. Sama naloga je namenjena prikazu možnosti širšega analitičnega izkoriščanja razpoložljivih podatkov, kadar se iz evidenc, ki so za večino uporabnikov nedostopne, pripravi izvoz verzije podatkov, ki je anonimizirana in pregledno urejena za namen neodvisne druge rabe.

Splošne informacije o zbirki:

- podatki generirani s pomočjo sql stavkov ter shranjeni v .xlsx format
- 5 podatkovnih zbirk: 2015, 2016, 2017, 2018, 2019
- posamezna zbirka vsebuje cca. 300-400 spremenljivk in cca 57.000 enot
- podatki v zbirki: nekaj splošnih podatkov o nosilcu KMG (spol, leto rojstva), občina KMG, KMG-ID (umetno generiran), plačila, kršitve, poljine, št. glav živine, kontrole, tip kmetije⁴, itd.
- populacija: posamezno KMG, ki je v tekočem letu oddalo vlogo za subvencijo
- anonimizacija: umetno generiran KMG-ID, leto rojstva nosilca (iz celotnega datuma rojstva, bo v datoteki le letnica)

Postopek prevzema podatkovnih zbirk sledi protokolu, kot je sicer v praksi v ADP (glej <https://www.adp.fdv.uni-lj.si/deli/postopek/>) in zajema:

- Izpolnitev dokumentacije:
 - izjava o Izročitvi
 - obrazec za opis raziskave, na podlagi katerega se pripravi metapodatkovni opis. Obrazec je dostopen v .doc ali .pdf formatu na spletni strani ADP ali pa preko aplikacije 1ka. V primeru slednjega se dajalcu generira enolično geslo za dostop ter posreduje povezavo do obrazca
 - izjava naročnika /financerja v kolikor je potrebno
- Ureditve podatkovne datoteke glede na [priporočila ADP](#)
- Priprava ostalih gradiv potrebnih za razumevanje podatkov (šifranti, kodirna knjiga, izpis frekvenc, podatki o izvedbi, kopije publikacij in ostalih spremnih dokumentov, ki so bili del zbiranja podatkov ali pa so pomembni za njihovo razumevanje, itd.)
- Podatkovne datoteke in gradiva se posreduje preko zaščitenega oblaka [ADP hramba](#)

Zbirke bodo objavljene na spletni strani ADP v Katalogu ADP (glej <https://www.adp.fdv.uni-lj.si/opisi/>), pri čemer opis sledi metapodatkovnemu modelu CESSDA z uporabo kontroliranega besedišča Data Description Initiative (DDI) ter ključnih besed tezavra ELSST (glej povezave pri evidencah).

Za razliko od opisa evidenc bo metapodatkovni opis zbirk takoj pripravljen tako v slovenskem kot tudi angleškem jeziku. ADP namreč vse opise raziskav ob prevzemu podatkov v sistem trajne digitalne hrambe in dostopa ponudi v slovenski in angleški različici Kataloga ADP. Ker je Katalog ADP že žet (ang. harvesting) s strani že prej omenjenega CESSDA kataloga, to pomeni, da bodo zbirke takoj po objavi vidne in najdljive tudi v širšem evropskem prostoru. Posamezne zbirke bodo objavljene in v sekundarno rabo ponujene kot samostojne objave, ki bodo povezane v serijo. Tako je omogočena možnost dodajanja novih zbirk (npr. generirani podatki za leto 2020), v kolikor bo v prihodnje izražen interes s strani ARSKTRP.

Režim dostopa do Kataloga ADP, metapodatkov, gradiv in raziskovalnih podatkov je podrobno opisan v enem izmed predhodnih odstavkov. V sklopu opisa raziskave pa uporabnik pridobi tudi informacije o dostopu in citiranju, vključno z oceno in statusom raziskave. Tu je potrebno omeniti, da so na spletni strani, uporabniku prosto dostopne tudi absolutne frekvenčne porazdelitve, kar pomeni, da lahko uporabnik na enostaven način pridobi uvid v podatke, vendar z njimi ne more operirati, medtem, ko

⁴ Tip kmetije v zbirki je umetno generiran za potrebe demonstracije uporabnosti podatkov v okviru projekta.

lahko do relativne frekvenčne porazdelitve prosto dostopa preko spletnega vmesnika Nesstar. Spletni vmesnik Nesstar (glej <https://nesstar2.adp.fdv.uni-lj.si/webview/?language=sl>) je platforma, ki ima enako vlogo kot spletna stran (iskanje po katalogu, pregled metapodatkov, ostalih gradiv, podatkov, prenos gradiv, itd.), le da ob predhodni registraciji (če ni določeno drugače) omogoča tudi prenos datotek in analizo podatkov (tabele, grafi, korelacija, regresija, ...). Uporabnik se torej lahko prosto seznanj z relativnimi frekvenčnimi porazdelitvami, samo operiranje s podatki pa zahteva registracijo. Dostop do podatkov določa dajalec podatkov. V ADP sicer omogočamo različne vrste dostopa, od tistega najbolj odprtega t. i. prosti dostop, do najbolj zaprtega t. i. dostop pod posebnimi pogoji (glej <https://www.adp.fdv.uni-lj.si/uporabi/kako/pravila/>). Režim dostopa dotičnih zbirk bo standardni dostop, ki vključuje dostop do datoteke za javno rabo oz t. i. »Public Use File« ob predhodni registraciji.

Četudi so zbirke trenutno še v fazi prevzema, je bilo s strani vseh vpletenih opravljeno že kar nekaj dela. Izvedeni so bili sestanki, kjer se je ADP seznanil s samo vsebino zbirk, dajalcu (ga. Maša in g. Miran) je bil predstavljen postopek predaje, katero dokumentacijo je potrebno izpolniti ter katero spremeno gradivo je potrebno pripraviti za lažje razumevanje in uporabo podatkov. Dajalec pa se je prav tako seznanil z možnostmi dostopa, ki jih ponuja ADP. Pojavilo se je tudi vprašanje o potrebi po anonimizaciji podatkov. ADP je zbirke sicer že prejel v pregled, pri čemer so bili podatki pregledani po predpisanem postopku, prav tako nekaj pripravljenih gradiv (pripravljeni šifranti). V procesu pregleda je bilo ugotovljeno nekaj nejasnosti oz. dilem, ki so bile v nadaljevanju skupaj z dajalcem tudi v večini razrešene. Tukaj gre zahvala tudi g. Matjažu Kolarju, ARSKTRP, za dodatno preverjanje podatkov, ki je omogočilo razjasnitev nekaterih dilem. Podatkovne zbirke kakor tudi vsa dokumentacija ter ostala gradiva so trenutno v pripravi na ARSKTRP. Po njihovem prevzemu bo ADP nadaljeval z delom.⁵

Vsa prejeta gradiva bodo po prevzemu na ADP po ustaljenem protokolu ponovno skrbno pregledana (v primeru dilem se kontaktira dajalca), opravilo se bo ovrednotenje podatkov za znanost in njihovo dolgoročno uporabnost, namreč podatki sprejeti v ADP štejejo kot znanstvena objava po merilih ARRS in jih je potrebno navajati v seznamu uporabljene literature. Sledila bo priprava metapodatkov, podatkov in gradiv v paket za dolgotrajno hrambo in v paket za dostop in nadaljnjo uporabo. Podatkovne datoteke bodo uporabnikom na voljo v različnih formatih. Pripravljen metapodatkovni opis se bo, tako kot v primeru opisa evidenc, poslal v avtorizacijo oz. pregled avtorju, se po potrebi popravil in objavil. O objavi se bo obvestilo dajalca, kakor tudi potencialne uporabnike preko ustaljenih kanalov ADP (spletna stran ADP, podstran ADP na strani FDV, družbena omrežja, e-novice, FDV napovednik).

ARSKTRP s predajo omenjenih podatkov v ADP ustvarja zgled za ostale skrbnike podatkov glede možnosti širitve kroga uporabnikov podatkov za različne namene, tako za naloge, povezane s področjem dela posamezne ustanove kot za druge naloge. Kot je bilo uvodoma že omenjeno, je bil namen tvorjenja zbirke primarno prikaz možnosti uporabe in ne predaja v arhiv. Delo na zbirki bo tudi v bodoče od vseh sodelujočih zahtevalo še kar nekaj vloženega truda in časa, namreč podatke je zaradi svoje specifičnosti in priprave potrebno ustrezno opremiti z dodatnim gradivom in pojasnili, ki bodo omogočila pravilno uporabo podatkov. Korektno sodelovanje in visoka angažiranost zagotavljajo, da bodo zbirke v čim krajšem možnem času pripravljene za objavo. Nekateri udeleženci so že na zaključnem srečanju izrazili zanimanje za uporabo podatkov, tako za učenje analitičnih pristopov orodja Orange, ki so ga predstavili kolegi s Fakultete za računalništvo, kot za druge namene. Podatki

⁵ V času priprave poročila prevzem zbirk v ADP še ni bil realiziran, so pa bile izvedene aktivnosti potrebne za predajo. Poročilo tako zajema opis aktivnosti, ki so že bile izvedene ter v nadaljevanju aktivnosti, ki bodo sledile prevzemu.

so v takšni obliki edinstveni in primerni tudi za raziskovalno delo in za vključevanje v učno okolje s študenti. Z njimi si lahko pomagajo tudi kmetijski svetovalci pri svojem delu na terenu.

Zgled priprave opisov evidenc in predaje ter deljenja podatkov za drugo rabo, ki smo ga pripravili v okviru projekta, odpira še nadaljnje možnosti sodelovanja. Preko področne zbirke, ki smo jo za zgled oblikovali na ADP, lahko podatke, ustrezno urejene in dokumentirane ter z razrešenimi pravnimi in etičnimi vidiki, delijo tudi drugi skrbniki in ustvarjalci podatkov. Zlasti ustanove znanja, univerze in inštituti, ki so tako uporabniki kot ustvarjalci podatkov, lahko prostor izkoristijo za predajo in deljenje podatkov iz raziskovalnih projektov po načelih odprte znanosti, kakor se uveljavlja v mednarodnem okolju (glej poročilo iz prvega dela projekta). Pa tudi druge javne ustanove lahko prispevajo in sodelujejo pri pripravi obogatenih podatkov, povezljivih ali združenih iz različnih virov. Denimo Statistični urad RS bi lahko pripravil in predal v nadaljnjo uporabo mikro-podatke iz vzorčnih raziskovanj Strukture kmetijskih gospodarstev in vsakih deset let iz Popisa kmetijstva. Pomembno bi bilo informirati in organizirati zainteresirane akterje s področja kmetijstva, začevši z oblikovalci politik in interesenti za njihovo načrtovanje in presojo učinkovitosti, ter vseh ostalih.

Priloga: predlogi demonstracijskih nalog

Tabela 1: Splošni predlogi raziskovalne skupine (Mag. Maša Kerstein)

Profiliranje	Korelacija	Napovedovanje
<p>1. Enotni profil KMG</p> <ul style="list-style-type: none"> - Oblikovanje enotnega profila KMG v podporo odločanju na več ravneh: kmet, svetovalec/svetovalna služba, nadzorne službe za administrativne kontrole/kontrole na kraju, strokovne službe (ZRSVN), zakonodajne službe (MKGP, MOP), raziskovalne skupine in znanstvene institucije. - Profil naj bo uporaben (a) na mikro ravni (KMG) in (b) na makro ravni (ocena stanja, načrtovanje in spremljanje ukrepov in politik). - Posamezna vprašanja: Ali je mogoče iz podatkov v okviru zbirne vloge (ZV) za pridobitev subvencij opredeliti tip kmetije? (t.j. živinorejska, vrtnarska, poljedelska, kmetije s trajnimi nasadi, kmetija brez živine). Pri tem uporabimo tehnike strojnega učenja, ki temeljijo na čim večjem številu podatkov, vendar ne tistih, ki so neposredno povezani z opredelitvijo tipov (kriteriji). - V drugi fazi za opredelitev tipov uporabimo prostorske podatke (sloji RKG, hidrografska mreža, podatki ARSO) in druge podatke (npr. FADN).* 	<p>4. Korelacija</p> <ul style="list-style-type: none"> - Ali obstaja povezava med velikostjo kmetije (npr. uporabimo razmejitev na pragu 6,5ha) in kontrolami (administrativnimi, na kraju samem, ločeno po površinskih – NP, KOPOP, EK, DŽ, OMD, CC, živali – NP). - Kakšna je razlika po velikosti in vrsti (glej nalogo 3)? - Ali je povezava značilna upošteva druge dejavnike?* 	<p>1. (-9.) Trendi in predvidevanja strukturnih sprememb</p> <ul style="list-style-type: none"> - Na podlagi sprememb na ravni posamezne KMG, vključno s spremembo rabe v nekmetijska zemljišča (zaraščanje) izpeljati trende v dolgoročnem obdobju in v povezavi z izvajanjem ukrepov v posameznih obdobjih. - Kakšna so predvidevanja po ključnih kategorijah? Kje se skriva potencial? Katere ukrepe izboljšati? Kako smiselno povezati kmetijske, okoljske in socialne ukrepe? - Primer: kako z natančnejšim definiranjem pogojev kmetovanja (naklon, višinska lega, zaraščanje, naravovarstvene omejitve) bolj ciljno usmerjati podpore?*
<p>2. Enotni profil KMG – uporaba</p> <ul style="list-style-type: none"> - Ali lahko iz profila KMG sklepamo na smiselno usmeritev za optimizacijo prihodkov, potencialno 'težavne' pofile in temu prilagajamo ukrepe? S kakšno verjetnostjo to trdimo za posamezen tip kmetije? Kateri od atributov ima največji in kateri najmanjši vpliv na izbrani tip kmetije? - Glede na tipe kmetij oblikujemo podtipove: npr. za živinorejski tip po proizvodnji glede na vrsto proizvoda (mlečne/mesne), vrtnarski tip za zeliščarske, poljubno združujemo kmetije po 	<p>5. Korelacija – družbenoekonomski</p> <ul style="list-style-type: none"> - Ali obstaja povezava med starostjo nosilca in prejemki ter med starostjo in kršitvami? Uporabimo prag do vključno 40 let starosti. Koliko je teh KMG in kje se nahajajo? Kakšni so velikostni razredi kmetije in ukrepi, ki jih uveljavljajo? - Koliko jih uveljavlja zahtevek Mladi kmet v okviru NP? Ali uveljavljajo tudi druge ukrepe (ne-IAKS) za pridobitev sredstev (znanje, 	

<p>več tipih: npr. poljedelska in živinorejska, zato, da določimo smiselne ukrepe ob ustreznih pogojih.</p> <ul style="list-style-type: none"> - Primer potencialnega vprašanja: katere kmetije imajo potencial za razvoj v živinorejski tip kmetij? Ali za ukrep ekološkega kmetovanja-čebele (EK_čebele)? Na katerih območjih (občine) se nahajajo? Ali je to področje primerno in ustreznosti za razvoj te panoge in posledično spodbude? - Kakšno je število potencialnih upravičencev glede na vlagatelje po posameznem ukrepu, ločeno za (a) ukrepe z 1-letno obveznostjo (npr. DŽ) in (b) 5-letno obveznostjo (npr. KOPOP).* 	<p>investicije)? V koliko primerih prejemnikov sredstev iz naslova ukrepov »mladih kmetov« pride čez čas do ponovne spremembe nosilca kmetije?</p> <ul style="list-style-type: none"> - Kakšne so značilnosti KMG v starostni skupini nosilca nad 80 let (ukrepi, pridobljena sredstva, uveljavljanje zahtevkov)?* 	
<p>3. Enotni profil KMG – uporaba za nadzor</p> <ul style="list-style-type: none"> - Katere kršitve so tipične za posamezno vrsto tipa kmetij (glej nalogo 1)? Kršitve smo opredelili kot razliko med zahtevanim in izplačanim zneskom po posamezni obračunski skupini glede na ukrep (OMD, NP, KOPOP, EK, DŽ). Prikaz na karti za usmerjanje izvajanja nadzora. - Kako tipi kmetij sovpadajo s posameznimi kršitvami (statistična verjetnost)? Ali obstajajo značilne kršitve za posamezna območja in za določen tip kmetij? 	<p>6. Korelacija – načrtovanje</p> <ul style="list-style-type: none"> - Ali obstaja povezava med višino subvencij (pod/nad 2.000€, pod/nad 100.000€) in ugotovitvami kontrol (glej nalogo 4)? Na katerih območjih so te kmetije in katere zahteve vlagajo? - Podobno za izbrane cilje: ali so izbrani sklepi skladno s trenutno strategijo kmetijske politike (kazalci outputa in rezultata)? Ali izbrani ukrepi pripomorejo k doseganju ciljev trenutne strategije kmetijske politike (kazalci vpliva)? Prikaz na mapi. - Ali je mogoče z uporabo različnih zbirk izboljšati kakovosti podatkov in predvideti 'napake'?* 	

*Siva barva: obravnavano na delavnici 2019 na podlagi letne zbirke podatkov AKTRP

Tabela 2: Horizontalni primeri z vidika treh deležnikov

RKG	Naravovarstvo	KGZS
<ul style="list-style-type: none"> - Koliko GERKov leži skladno s katastrom (parcele)? - Koliko GERKov obdelujejo dejansko lastniki (GERKi- uradni zakupi ZK, GERKi- ne uradni zakupi)? Ali je glede na slednje smiselno, da se prične s kampanjo uradnih zakupov? - KMG – MIDi/ GERKi / posebna območja varstvenih režimov / prikaz vplivov KMG – MID / ciljno naravnano okoljsko in investicijsko financiranje - Ali lahko na osnovi podatkov določimo potencialna območja za posamezne agrarne operacije (komasacije ipd.) Npr. ali veliki zakupniki dosegajo boljše rezultate v primerjavi z manjšimi (proizvodnji, okoljski ...)? - Ali je možno na osnovi podatkov določiti učinkovitost delovanja javnih služb, ki delujejo pod okriljem MKGP (npr. problem zakupov, zaraščanja ugotovljeno s pomočjo križanja podatkov v povezavi z izvedenimi ukrepi posamezne službe) 	<ul style="list-style-type: none"> - Koliko je trajnostno naravnanih kmetij (pogoj: max. obremenitev = 0,75GVŽ/ha) in na katerih območjih (občine) ležijo? - Koliko kmetij ima vse površine (in nadalje 10%, 50%) v območju NATURA2000? Kakšna je struktura kmetijskih zemljišč na NATURA 2000 leta 2007 in v primerjavi z letom 2017? - Kolikšna je povprečna velikost kmetije na NATURA2000? Koliko kmetij izvaja naravovarstvene ukrepe (ti so usmerjeni na NATURA2000) tipa KRA na KOPOP + OMD? Kje ležijo? - Ali se z naravovarstvenimi ukrepi (našteje taksativno po šifrantu operacij) ohranja splošna biodiverziteteta? Slikovni prikaz na karti. (Kazalnik 1: Koliko KMG s kolikšnimi površinami je vključenih v te ukrepe na NATURA2000; Kazalnik 2: Primerjava stanj 2007 : 2017). - Kakšno je gospodarjenje s trajnimi travniki (raba 1300)? Slikovni prikaz (Kazalnik 1: delež trajnih travnikov 2007: 2017 + v ar in št. KMG; Kazalnik 2: kolikšen je delež 100% prekritja? V ar in št. KMG) Ta prikazuje dejanski OOTT. - Koliko trajnih travnikov v NATURA 2000 spada v okvir OOTT? (pod pogoji nove SKP-GAEC) in primerjava z dosedanjim stanjem. 	<ul style="list-style-type: none"> - Spajanje podatkov med bazami: vloge za NP., RKG, VOLOS, GOVEDO SI; FARM MANAGER, FADN knjigovodstvo za namen izdelave svetovalnega orodja, ki je lahko del širše podatkovne baze/aplikacije/portala -> Kako omogočiti uvoz ažurnih podatkov iz administrativnih zbirk <ul style="list-style-type: none"> i. Za analize ii. Za izpolnjevanje vlog (gnojilni načrti) iii. Za uporabniška orodja (seznam dovoljenih sredstev) - Predvidevanje na katerih območjih bo opuščanje – zaraščanje? <p>Analiza: kakšne kmetije so na teh območjih:</p> <ul style="list-style-type: none"> - starost nosilca, - članstvo, - velikost , - opremljenost,... <p>-> Kakšni ukrepi bi bili glede na ugotovitve dejavnikov (zgoraj) potrebni, da bi preprečili zarast / da bi obdržali vitalno kmetijo in s tem socialno poseljenost.</p>

Tabela 3: predlogi prispeli na odprti poziv 2019

<p>1. RKG: možnost samodejnega izpisa in prikaza na karti za pogoste poizvedbe:</p> <ul style="list-style-type: none"> - osnovne statistike (število KMG, povprečna velikost) in demografija (starost nosilca) po prostorski enoti (občine, UE) - skupna površina poljin in struktura zemljišč po kategorijah rabe/GERKih (njive, travinje, gozd) po prostorskih enotah - dele območij (npr. glede na vse GERK) ki potencialno lahko vstopijo v ukrep (npr. KOPOP) - pokritost površin prostorskih enot z rabo in GERKi po območjih NATURA2000 - vinogradi po prostorskih enotah (površina, povprečna vleikost, povprečna starost trsov, povprečen naklon, sorte ipd.) - dopolnilne dejavnosti po vrsti, številu po prostorski enoti. <p>Prikaz časovne vrste in trenda.</p>	<p>3. Predlogi AKTRP</p> <p>a) Medved in volk</p> <ul style="list-style-type: none"> - Koliko KMG-MIDov in upravičenih površin v posamezni občini je vključenih v varovalne ukrepe KRA_OGRM, KRA_VARPA, KRA_VARPP, KRA_CRED in KRA_PAST po območjih iz evidence MV1(osrednje območje pojavljanja medveda) in MV2 (osrednje območje pojavljanja volka). - Koliko KMG-MIDov v posamezni občini z zahtevkom DŽ-drobnica leži v sloju MV2? (zbirke zahtevki DŽ-drobnica, KMG-MIDi, uveljavljena površin iz ZV iz AKTRP, evidenci sloja MV1 in MV2 iz AKTRP/RKG ter občine iz RKG) <p>b) Zelenjadarji in sadjarji</p> <ul style="list-style-type: none"> - Ali je mogoče ugotoviti razmerje med količino pridelane zelenjave (oz. prodajo) in prijavljenimi površinami pri zelenjadarjih, upošteva ekološko in integrirano pridelavo? Upoštevati bi bilo mogoče tudi podatke o uporabi FFS. <p>Zelenjadar je KMG-MID, ki prideluje zelenjavo oz. zelišča na tleh (GERKi 1100-njiva, 1161-hmeljišče v premeni, 1180-trajne rastline na njivskih površinah), na prostem, na vsaj 0,1 ha zemljišč oz. vsaj 0,02ha površin v rastlinjakih (GERKi 1190-rastlinjak, 1191-rastlinak, kjer pridelava ni v tleh) oz. njivskih površinah, kjer pridelava ni v tleh (kontejner - GERK 1181). V ZV je površina z zelenjavo določena s šiframi '402', '733', '734', '703' in '444'.</p> <ul style="list-style-type: none"> - Ali je mogoče podobno ugotavljati tudi za sadje (prodajalci brez proizvodnje)? <p>c) Svetovalne službe</p> <ul style="list-style-type: none"> - Katere svetovalne službe imajo v svojem okolju največ nosilcev s čezmerno prijavo površin? (kontrola s terena) <p>d) Spremembe GERK</p> <ul style="list-style-type: none"> - kateri GERKi in v katerih občinah se najbolj spreminjajo v rabi in površini? (oddaja ZV brez zahtevkov oz. le enkrat v 5 letih za KOPOP in EK ni možna zaradi preveč spremenjenih površin v RKG). - Ali so razlog zakupne pogodbe, komasacije, obnove trajnih nasadov ...? (vir: GERKi in občine iz RKG) <p>e) Gnojevka</p>
<p>4. MKGP: Zaraščanje</p> <p>Kmetijske površine, ki se ne obdelujejo, se zarastejo in v določenem trenutku 'izginejo' iz RKG. To se lahko zgodi na dva načina: (1) zaraščena površina je napačno prijavljena kot kmetijska površina upravičena za ukrepe (na ZV) kar se ugotovi pri pregledu na kraju samem, na kar RKG dotični GERK zmanjša. (2) Nosilec KMG opusti kmetijsko dejavnost in na UE v RKG izloči površino iz GERKa.</p> <p>Koliko in kje so površine, ki se med leti (med zaporednima ZV) zarastejo in 'izginejo' iz RKG, kateremu KMG-ju so pripadale in kakšna raba GERK je bila na teh površinah pred izločitvijo?</p> <p>Zbirke za pridobitev podatkov vključujejo ZV zaporednih let in RKG (KMG-MID, razlog izbrisa), Evidenco dejanske rabe.</p> <p>Upravičeni GERKi za ukrepe (1100 – njiva, 1131 – začasno travinje, 1150 – njiva za rejo polžev, 1160 – hmeljišče, 1161 – hmeljišče v premeni, 1170 – jagode na njivi, 1180 – trajne rastline na njivskih površinah, 1181 – trajne rastline na njivskih</p>	

<p>površinah, kjer pridelava ni v tleh, 1190 – rastlinjak, 1192 – rastlinjak s sadnimi rastlinami, 1211 – vinograd, 1212 – matičnjak, 1221 – intenzivni sadovnjak, 1222 – ekstenzivni sadovnjak, 1230 – oljčnik, 1240 – ostali trajni nasadi, 1300 – trajni travnik, 1320 – travinje z razpršenimi neupravičenimi značilnostmi, (v RKG je določena največja upravičena površina tega GERK-a, in navedena kot numerični atribut GERK-a), 1610 – kmetijsko zemljišče v pripravi).</p> <p>Evidenca dejanske rabe (1410 – Kmetijsko zemljišče v zaraščanju; 1500 – Drevesa in grmičevje; 1600 – Neobdelano kmetijsko zemljišče; 2000 – Gozd; 3000 – Pozidano in sorodno zemljišče). Evidenca dejanske rabe kmetijskih in gozdnih zemljišč se ažurira na 3 leta; vsako leto 1/3 Slovenije.</p>	<ul style="list-style-type: none"> - Kateri KMG-MIDI oddajajo gnoj drugemu KMG-MID, ki v RKG nima vpisanih površin in nima lastne bioplinarne? Trenutno so tretirani kot kršitelji. Ali imajo svojo kompostarno, so prodajalci gnoja? Kdo so KMG-MID in kje so območja oddanega in prejetega gnoja? Gre za vzorec in kakšni so vzroki? (KMG-MIDI oddajalcev in prejemnikov iz ZV, površine KMG-MID iz RKG, dopolnilne dejavnosti iz RKG) f) Drugo <ul style="list-style-type: none"> - Določitev prostorske razporeditve kmetij glede na starost nosilcev in raven izobrazbe nosilcev. Uporaba razpoložljivih podatkov za ugotavljanje ali so nosilci uporabniki pametnih telefonov. - Vpliv investicijskih sredstev na spremembo velikosti kmetije, število kmetij in depopulacijo v regiji. g) Dopolnitev nalog: <ul style="list-style-type: none"> - profiliranje: naravovarstvena (izvajanje HAB, MET, STE, VTR), EK kmetija, vzdrževana kmetija (brez nosilcev na sedežu KMG ali z enim samim, ki se letno obdela a brez živečih). - korelacije: regija, starost, oddaljenost od središč oz. izpostav KGZS, povezava z internetom (informacije). Povezava med zaraščanjem in obtežbo, številom članov na KMG, oddaljenostjo od središč, naklonom, številom kršitev. - Naravovarstvo: izobrazbena struktura in starost nosilcev, ki izvajajo naravovarstvene ukrepe, zemljiška struktura kmetij, vrsta kmetije (poslovni subjekt ali fizična oseba). - Podrobna analiza EK kmetij: pridelava hrane, trajno travinje na spodnji meji obtežbe (vloga reje – nekatere težko zagotavljajo večjo obtežbo), poveza s subvencijami, preoravanjem OOTT.
---	--