

Analiza i rudarenje podataka na projektima namijenjenih provedbi kohezijske politike Europske unije u RH

Stegić, Matija

Master's thesis / Diplomski rad

2023

Degree Grantor / Ustanova koja je dodijelila akademski / stručni stupanj: **University of Split, Faculty of economics Split / Sveučilište u Splitu, Ekonomski fakultet**

Permanent link / Trajna poveznica: <https://um.nsk.hr/um:nbn:hr:124:873694>

Rights / Prava: [Attribution-NonCommercial-ShareAlike 3.0 Unported / Imenovanje-Nekomercijalno-Dijeli pod istim uvjetima 3.0](#)

Download date / Datum preuzimanja: **2024-12-29**

Repository / Repozitorij:

[REFST - Repository of Economics faculty in Split](#)



SVEUČILIŠTE U SPLITU

EKONOMSKI FAKULTET

DIPLOMSKI RAD

**Analiza i rudarenje podataka na projektima namijenjenih
provedbi kohezijske politike Europske unije u RH**

Mentor:

Prof. dr. sc. Nikša Alfirević

Student:

Matija Stegić

Split, rujan, 2023

IZJAVA O AKADEMSKOJ ČESTITOSTI

Ja, Matija Stegić _____ ,

(ime i prezime)

izjavljujem i svojim potpisom potvrđujem da je navedeni rad isključivo rezultat mog vlastitog rada koji se temelji na mojim istraživanjima i oslanja na objavljenu literaturu, što pokazuju korištene bilješke i bibliografija. Izjavljujem da niti jedan dio navedenog rada nije napisan na nedozvoljeni način te da nijedan dio rada ne krši autorska prava. Izjavljujem, također, da nijedan dio rada nije korišten za bilo koji drugi rad pri bilo kojoj drugoj visokoškolskoj, znanstvenoj ili obrazovnoj ustanovi.

Split, 2023. godine

Vlastoručni potpis : 

SADRŽAJ:

1. UVOD	1
1.1. Problem istraživanja	1
1.2. Predmet istraživanja.....	2
1.3. Istraživačke hipoteze.....	3
1.4. Ciljevi istraživanja	3
1.5. Metode istraživanja	4
1.6. Doprinost istraživanja.....	5
1.7. Struktura rada.....	5
2. FINANCIRANJE EU PROJEKATA NA PODRUČJU RH	6
2.1. Analiza iskorištenosti ESI fondova u državama članicama EU.....	6
2.2. Strukturni i investicijski fondovi	13
2.3. NUTS klasifikacija regija	16
2.4. Kohezijska politike Europske unije	18
2.5. Operativni program “Konkurentnost i kohezija 2014.-2020.”	19
2.6. Operativni program “Učinkoviti ljudski potencijali 2014.-2020.”	20
3. RUDARENJE PODATAKA.....	21
3.1. Pojam i važnost rudarenja podataka	21
3.2. Proces rudarenja podataka	22
3.2.1 Razumijevanje poslovnog problema.....	23
3.2.2 Razumijevanje podataka.....	25
3.2.3 Priprema podataka	26
3.2.4. Modeliranje	28
3.2.5. Evaluacija.....	29
3.2.6. Isporuka.....	30
3.3 Metode rudarenja podataka.....	30
3.3.1. Korelacija	31
3.3.2 Analiza varijance.....	32
3.3.3 Klasteriranje	33
3.3.4 Stabla odlučivanja	38
4. ANALIZA I RUDARENJE PODATAKA NA PRIMJERU JAVNE BAZE PODATAKA O EU PROJEKTIMA	41

4.1 Opis skupa podataka	41
4.1.1. <i>Sadržaj skupa podataka</i>	41
4.1.2. <i>Popis skupa podataka</i>	41
4.2 Vizualizacija podataka	44
4.3 Analiza podataka i rezultati istraživanja	65
4.4 Osvrt na hipoteze	80
5. ZAKLJUČAK	83
LITERATURA	85
POPIS SLIKA I TABLICA	92
SAŽETAK	93
SUMMARY	94

1. UVOD

1.1. Problem istraživanja

„Poslovna inteligencija (BI) je novo područje istraživanja primjene ljudskih kognitivnih sposobnosti i tehnologije umjetne inteligencije na upravljanje i podršku odlučivanju na različitim organizacijskim problemima. Uključuje softvere za ekstrakciju, transformaciju i učitavanje podataka (ETL), skladištenje podataka, upite i izvještavanje u bazama podataka, višedimenzijsku analizu podataka, rudarenje podataka i vizualizaciju podataka“ (J., Ranjan, 2005, str. 60-70). Stoga, čini široku kategoriju aplikacija i tehnologija za prikupljanje, pružanje pristupa i analizu podataka. Pojedinaac ili organizacija odluku ne donosi na osnovu podataka već na osnovu relevantnih informacija iskoristivih kao podrška odlučivanju zbog čega se primjena poslovne inteligencije u organizacijama smatra potencijalnim izvorom održive konkurentne prednosti.

Kao dio BI, rudarenje podataka predstavlja “netrivijalni proces identificiranja valjanih, novih, potencijalno korisnih i konačno razumljivih uzoraka u podacima pohranjenim u strukturiranim bazama podataka”, gdje su podaci organizirani kao zapisi strukturirani prema kategorijskim, ordinalnim i kontinuiranim varijablama (Moss L.T. & Atre S., 2003, str. 381-397). To je disciplina otkrivanja korisnih informacija iz podataka i stavljanja tih informacija u praktičnu upotrebu automatiziranom obradom velike količine podataka u svrhu identificiranja obrazaca i trendova u podacima.

S obzirom da organizacije pohranjuju sve veće količine podataka generirane tijekom svojih poslovnih aktivnosti, uzorci koji ukazuju na uspješnost poslovnih procesa obično se nalaze „zakopani“ u tim povijesnim podacima (Nassar K., 2006, str. 139-148). Stoga, pravilna analiza podataka i otkrivanje uzoraka u podacima je od iznimne strateške važnosti organizacijama jer tko posjeduje prave informacije, taj posjeduje i znanje kako donositi bolje odluke. Europska Unija, koja je regionalna organizacija europskih država i jedina organizacija ove vrste na svijetu, nije iznimka. Ona osigurava određene iznose financijskih sredstava za provođenje projekata i programa, a neki od tih programa su OP „Konkurentnost i kohezija 2014.-2020.“ i OP „Učinkoviti ljudski potencijali 2014.-2020.“ čiji setovi podataka o financiranim projektima predstavljaju podlogu za primjenu tehnika rudarenja podataka u empirijskom dijelu rada.

Ispravna, transparentna i učinkovita potrošnja proračuna EU-a provodi se kroz pažljivo planiranje (Središnji državni ured za razvojnu strategiju i koordinaciju fondova Europske unije [SDURF], 2009, str. 13). Projekti financirani iz proračuna Europske unije utječu na životni standard stanovnika Hrvatske. OPKK provodi

kohezijsku politiku kojom potiče ulaganja u infrastrukturne investicije te pruža potpore razvoju poduzetništva i istraživačkim djelatnostima dok OPULJP potiče rast zapošljavanja i poboljšanje socijalne kohezije, unutar Hrvatske, alokacijom sredstava prema najpotrebnijima područjima i sektorima. Sredstva koja EU dodjeljuje iznose do 6,88 milijardi eura, a tijekom 2020. godine, Hrvatska je bila na samom začelju po korištenju financijskih sredstava iz fondova. Bila je na 27. mjestu među 28 zemalja Europske unije. Iako je tokom godina povećala dinamiku apsorpcije financijskih sredstava te su stvorene nove poslovne mogućnosti i povećana su ulaganja, zanimljiv je uvid u raspodjelu alokacije prema određenim atributima (županija, prioritetna os, i slično) upravo zbog činjenice da, Hrvatska, spada u jednu od lošijih u EU po stopi isplaćenih sredstava, a kohezijska politika predstavlja važan instrument razvoja. Također, pretpostavlja se da su veći iznosi sredstava dodijeljeni većim projektima koji su obično kompleksniji te osim što zahtijevaju više sredstava, zahtijevaju i više vremena pa tako sama logika nalaže da je veličina projekta određena iznosom dodijeljenih sredstava što, primjerice, uključujući određene atribute pruža uvid u karakteristike velikih projekata. Pravilnim segmentiranjem podataka dolazi se do klasifikacijskih obilježja. Na internetskoj stranici strukturnih fondova nalaze se javno dostupni podaci o financiranim projektima OPKK-a i OPULJP-a, tehnike rudarenja podacima upotrijebit će se za identificiranje uzoraka odnosno istražiti će se mogućnosti tehnika rudarenja podataka u dobivanju relevantnih informacija iz odabranog skupa podataka kako bi se otkrili odnosi, logičnosti, pravilnosti ili bilo kakve strukture među podacima koje mogu biti korisne te mogu olakšati donošenje odluka.

1.2. Predmet istraživanja

Istraživanje se temelji na praktičnosti primjene metoda i tehnika rudarenja podacima na odabranom setu podataka financiranih projekata OP "Konkurentnost i kohezija 2014.-2020." i setu podataka OP "Učinkoviti ljudski potencijali 2014.-2020.". Setovi podataka su javno dostupni na stranicama europskih strukturnih i investicijskih fondova i namijenjeni su osiguranju transparentnosti i pružanju informacija javnosti o financiranim projektima te korisnicima potpora iz fondova Europske unije. Prije početka analize potrebno je urediti podatke jer kvalitetno pripremljeni podaci predstavljaju temelj za uspješno izvršavanje daljnjih analiza. Pojam, koncept i tehnike rudarenja podataka će se teoretski obraditi, a analiziranjem seta podataka donijeti će se zaključci o dodijeljenim sredstvima projektima i njihovoj alokaciji prema određenim atributima kao što su županija, prioritetna os, operativni program i slično. Metodama i tehnikama rudarenja podataka će se nastojati identificirati uzorci o projektima financiranih sredstvima iz ESI fondova izdvajanjem onih projekata sličnih obilježja odnosno ukazat će se na značaj tehnika rudarenja podataka u

prepoznavanju obilježja određenih projekata i njihovim grupiranjem će se nastojati doći do informacija koje mogu biti korisne pojedincima ili organizacijama pri donošenju odluka o poslovanju. Sredstva europskih fondova mogu koristiti sve vrste organizacija iz javnog, privatnog i civilnog sektora. Osim navedenih, sredstva europskih fondova mogu koristiti građani pojedinci i njihove inicijative.

1.3. Istraživačke hipoteze

Na temelju prethodno utvrđenih problema i predmeta istraživanja postavljaju se sljedeće hipoteze:

H1:...Postoji povezanost između dodijeljenih vrijednosti bespovratnih sredstava EU i vremenskog razdoblja trajanja projekta te EU stope sufinanciranja projekata namijenjenih provedbi kohezijske politike

H2:...Ugovoreni projekti se značajno razlikuju prema prosječnom iznosu bespovratnih sredstava EU s obzirom na prioritete financiranja projekata namijenjenih provedbi kohezijske politike.

H3:...Ugovorena vrijednost bespovratnih sredstava EU i vremensko razdoblje trajanja projekta namijenjenih provedbi kohezijske politike imaju utjecaj na tip korisnika.

H4:... H4:...Ugovorena vrijednost bespovratnih sredstava EU ima utjecaj na vrstu fonda iz kojeg se projekt financira.

Na temelju prikupljenih podataka i njihovom analizom metodama rudarenja podataka donesene hipoteze će se prihvatiti ili odbaciti.

1.4. Ciljevi istraživanja

Analiziranjem projekata, koji su financirani sredstvima iz ESI fondova, nastoji se doći do korisnih informacija identificiranjem skrivenih uzoraka. Cilj istraživanja je prikazati mogućnosti praktične primjene metoda rudarenja podataka u dobivanju relevantnih informacija iz odabranog skupa podataka kako bi se otkrili odnosi, logičnosti, pravilnosti ili bilo kakve strukture među podacima. Također, testirat će se i primjenjivost alata na odabranom setu podataka i donijeti zaključci o projektima koji se financiraju iz europskih fondova jer utječu na poboljšanje života stanovnika Hrvatske poticanjem ulaganja u infrastrukturu (promet, energetika, zaštita okoliša, ICT) i razvoj poduzetništva i istraživačkih djelatnosti.

1.5. Metode istraživanja

U istraživanju će se primijeniti sljedeće metode znanstveno-istraživačkog rada:

- *Induktivna metoda i deduktivna metoda* - Induktivnom metodom se osmišljeno i dosljedno od pojedinačnih činjenica dolazi do donošenja zaključka o općem sudu. To je metoda kojom se spoznaju nove činjenice i zakonitosti. S druge strane, primjenom *deduktivne metode* se iz općih stavova dolazi do konkretnih zaključaka odnosno konkretnih činjenica.
- *Metoda analize i sinteze* – je metoda detaljnog ispitivanja složenih pojmova i zaključaka i njihovog svodenja na jednostavnije sastavne dijelove te istraživanje svakog dijela za sebe i u odnosu na druge dijelove. Nasuprot tome, metoda sinteze objašnjava stvarnost spajanjem jednostavnih misaonih tvorevina u složene i onih složenijih u još složenije povezujući ih sve u jedinstvenu cjelinu.
- *Metoda klasifikacije* – je metoda kojom se vrši podjela općeg pojma na posebne, u okviru opsega pojma.
- *Metoda kompilacije* – je postupak preuzimanja tuđih rezultata istraživačkih radova odnosno njihovih opservacija, stavova i zaključaka. To je metoda kojom se iz više različitih djela sastavlja literarno djelo. (Zelenika R. , 2004).
- *Metoda rudarenja podataka* – je metoda kojom se otkrivaju korisne informacije iz podataka i kojom se te informacije stavljaju u praktičnu upotrebu automatiziranom obradom velike količine podataka kako bi se identificirali obrasci i trendovi u podacima.

1.6. Doprinos istraživanja

Podaci s kojima se danas raspolaže su ogromni ili neprecizni i tada se radi o složenoj strukturi podataka kod kojih statističke tehnike ne uspijevaju, ali rudarenje podataka uspijeva zbog čega i predstavlja rješenje problema za rad s velikim skupovima podataka. Dobivene informacije o dodijeljenim sredstvima i raspodjeli alokacije projektima prema određenim atributima iz ESI fondova istraživanjem varijabli odabranog seta podataka mogu koristiti poboljšanju donošenja odluka sadašnjim i budućim korisnicima potpora EU, te svima onima koji ovo istraživanje pronađu korisnim. Također, može ukazati na praktičnost primjene rudarenja podataka u dobivanju relevantnih informacija kako bi se otkrili odnosi, logičnosti, pravilnosti ili bilo kakve strukture među podacima odabranog seta podataka. S obzirom na dinamičnost današnjeg tržišta, te brzi razvoj tehnologije, organizacije trebaju biti konstantno pratiti trendove na tržištu kako bi mogle stvarati i održavati konkurentsku prednost, a poslovna inteligencija predstavlja izrazito bitan čimbenik poslovnog uspjeha.

1.7. Struktura rada

Rad se sastoji od dva dijela: teorijskog i empirijskog dijela. U prvome dijelu rada opisan je problem i predmet istraživanja na temelju kojih su postavljene hipoteze istraživanja. Također, opisan je cilj rada, a koji se odnosi na prikaz mogućnosti praktične primjene metoda rudarenja podataka u dobivanju relevantnih informacija iz odabranog skupa podataka, opisane su metode i doprinos istraživanja. Ukratko će se prikazati iskorištenost sredstava iz EU fondova te će se opisati pojam i važnost provedbe kohezijske politike Europske unije, njenih operativnih programa i NUTS klasifikacija regija u RH. Predstaviti će se pojam, proces i objasniti će se metode rudarenja podataka koje će se koristiti u istraživanju. U empirijskom dijelu rada će se detaljnije opisati skup podataka, primijeniti će se metode rudarenja podataka na odabranom setu podataka i objasniti će se dobiveni rezultati istraživanja.

2. FINANCIRANJE EU PROJEKATA NA PODRUČJU RH

Osnovni cilj osnivanja Europske unije je uspostavljanje regionalne organizacije europskih država koje ostvaruju zajedničke ciljeve radi stvaranja zajedničkog tržišta na području Europske unije čime bi se države članice bolje pozicionirale u svijetu. Ti zajednički ciljevi podrazumijevaju postizanje uravnoteženog gospodarskog i društvenog razvoja, visoke razine zaposlenosti, zaštite prava i interesa građana.

Kao jedan od alata kojima se postiže ostvarenje zajedničkih ciljeva su objavljeni natječaji kojima se financiraju projekti u svim zemljama članicama. Projekti se financiraju iz proračuna Europske unije u koji svaka zemlja članica uplaćuje određeni iznos ovisno o sredstvima kojima raspolaže. Prema tome, što je gospodarstvo neke zemlje veće, to je veći iznos sredstava koji uplaćuje, i obrnuto. Hrvatska je pristupila Europskoj Uniji u 2013. godini.

Objavljeni natječaji se formiraju prema operativnim programima koji predstavljaju detaljne planove kojima države članice predviđaju određenu potrošnju sredstava tijekom programskog razdoblja. Svakim operativnim programom je određeno koji će se ciljevi, kojima se rukovodi kohezijska politika, nastojati ostvariti sredstvima dostupnim tom operativnom programu pa stoga ono što nije određeno u operativnom programu, ne može se niti financirati.

2.1. Analiza iskorištenosti ESI fondova u državama članicama EU

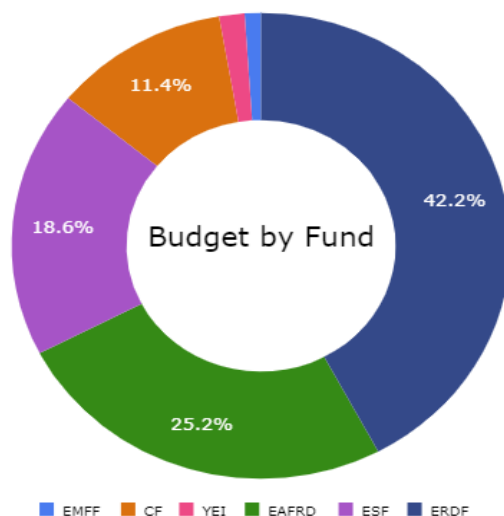
Cilj Europskih strukturnih i investicijskih fondova (ESI), u financijskoj perspektivi od 2014. do 2020., je maksimiziranje njihovog doprinosa strategiji Europa 2020. kojom su definirana tri socioekonomska cilja koja su se trebala ostvariti do 2020. godine. To su: *pametna rast* (razvoj gospodarstva temeljen na znanju i inovacijama), *održiva rast* (resursno učinkovitije, ekološki prihvatljivije i konkurentnije gospodarstvo) i *uključiva rast* (poticanje gospodarstva s visokom stopom zaposlenosti koje ostvaruje socijalnu i teritorijalnu koheziju).

S obzirom na usmjerenost kohezijske politike prema ostvarenju ciljeva strategije 2020., u nastavku je dan grafički prikaz udjela ukupnih financijskih sredstava Europske unije prema ESI fondovima (2014.-2020. godine). Iz grafičkog prikaza 1. vidi se da je najveći postotni udio sredstava EU, čak 42,2%, izdvojeno za financiranje Europskog fonda za regionalni razvoj (ERDF-a), a kojeg slijedi Europski poljoprivredni fond (EAFRD) sa 25,2%, Europski socijalni fond (ESF) sa postotnim udjelom od 18,6% te Kohezijski fond za koji je izdvojeno 11,4% ukupnog EU financiranja. Najmanji postotni udio financijskih sredstava izdvojen je za Europski fond za pomorstvo, ribarstvo i akvakulturu te za Inicijativu za zapošljavanje mladih (YEI).

Grafički prikaz 1.

Udio ukupnih financijskih sredstava EU prema ESI fondovima

ESIF 2014-2020: Total budget by Fund, EUR billion (daily update)



Izvor: Platforma otvorenih podataka za koheziju, <https://cohesiondata.ec.europa.eu/overview>

Isto tako, zanimljiv je pregled ukupnog financiranja EU iz ESI fondova prema zemlji članici (2014.-2020.) koji je vidljiv na grafičkom prikazu 2., na kojem se može uočiti da je Poljska zemlja članica koja prednjači po EU financiranju iz EFRR fonda i Kohezijskog fonda. EFRR je fond koji služi smanjenju razlika u stupnju razvijenosti europskih regija, dok se KF koristi za smanjenje gospodarskih i socijalnih razlika te za promicanje održivog razvoja.

Nakon Poljske, zemlje članice s najvećim ostvarenim financiranjem EU iz ESI fondova su: Italija (EFRR i ESF), Španjolska (EFRR i EAFRD), Francuska (EAFRD i EFRR), Rumunjska (EFRR i EAFRD), Njemačka (EAFRD-a i EFRR), Portugal (EFRR i ESF), Mađarska (EFRR i KF) i slično.

Bitno je napomenuti da provođenjem određenih javnih politika EU svakoj zemlji članici dostupan je određen iznos sredstava iz ESI fondova koji može koristiti prema NUTS klasifikaciji.

Na grafičkom prikazu 2. vidljivo je da razvijenije zemlje Europske unije (Italija, Španjolska, Francuska, Njemačka, Ujedinjeno Kraljevstvo, Austrija, Finska, Švedska, Irska, Belgija, Nizozemska, Danska i Luksemburg), ne koriste sredstva Kohezijskog fonda (KF) čije aktivnosti uključuju transeuropske mreže

prijevoza te okoliš. Transeuropskom mrežom prijevoza se financiraju veliki infrastrukturni projekti kao što je Pelješki most u Hrvatskoj, a aktivnošću vezanom uz okoliš se financiraju projekti koji doprinose okolišu u pogledu energetske učinkovitosti, razvoju željezničkog i javnog prijevoza i slično.

Među zemljama članicama, po financiranju Europske unije iz EAFRD-a redom se ističu: Francuska, Njemačka, Austrija, Finska, Švedska, Irska, Nizozemska, Danska, Luksemburg i Latvija. Tako se uočava da skoro sve navedene zemlje članice pripadaju razvijenim zemljama članicama EU. Hrvatska, kao zemlja članica, najviši iznos financiranja iz ESI fondova ostvaruje iz EFRR fonda, a kojeg slijedi EAFRD fond, KF te ESF fond.

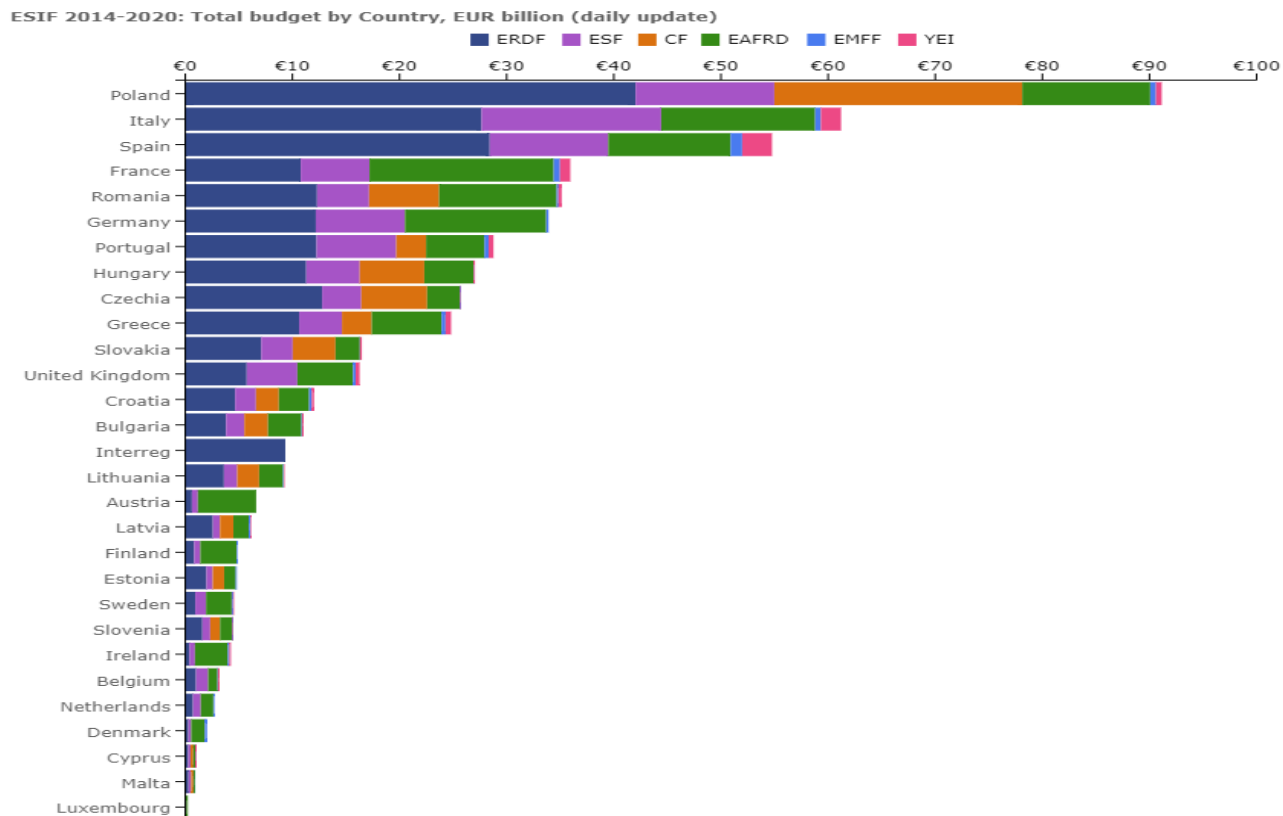
Tako se zaključuje da razvijene zemlje imaju dovoljno razvijenu transeuropsku mrežu prijevoza i okoliš te nemaju velike potrebe za financiranjem iz kohezijskog fonda u usporedbi sa ostalim zemljama članicama koje su u razvoju (npr. Hrvatska), već su njihove aktivnosti usmjerene na jačanje ruralnog razvoja financiranjem projekata kroz EAFRD fond.

S druge strane, zemlje u razvoju ili zemlje članice koje su najviše financijskih sredstava apsorbirale iz Kohezijskog fonda redom su: Poljska, Rumunjska, Češka, Mađarska, Slovačka, Portugal, Grčka, Hrvatska, Bugarska, Litvanija, Latvija, Estonija, Slovenija.

Najveći iznos financiranja iz Europskog socijalnog fonda (ESF), glavnog financijskog instrumenta Europske unije za ulaganje u ljudske resurse (zaposlene, mlade ljude i nezaposlene), ostvarile su: Italija, Poljska, Španjolska, Njemačka, Portugal, Francuska.

Grafički prikaz 2.

EU financiranje iz ESI fondova prema zemlji članici (2014.-2020.)



Izvor: Platforma otvorenih podataka za koheziju, <https://cohesiondata.ec.europa.eu/overview>

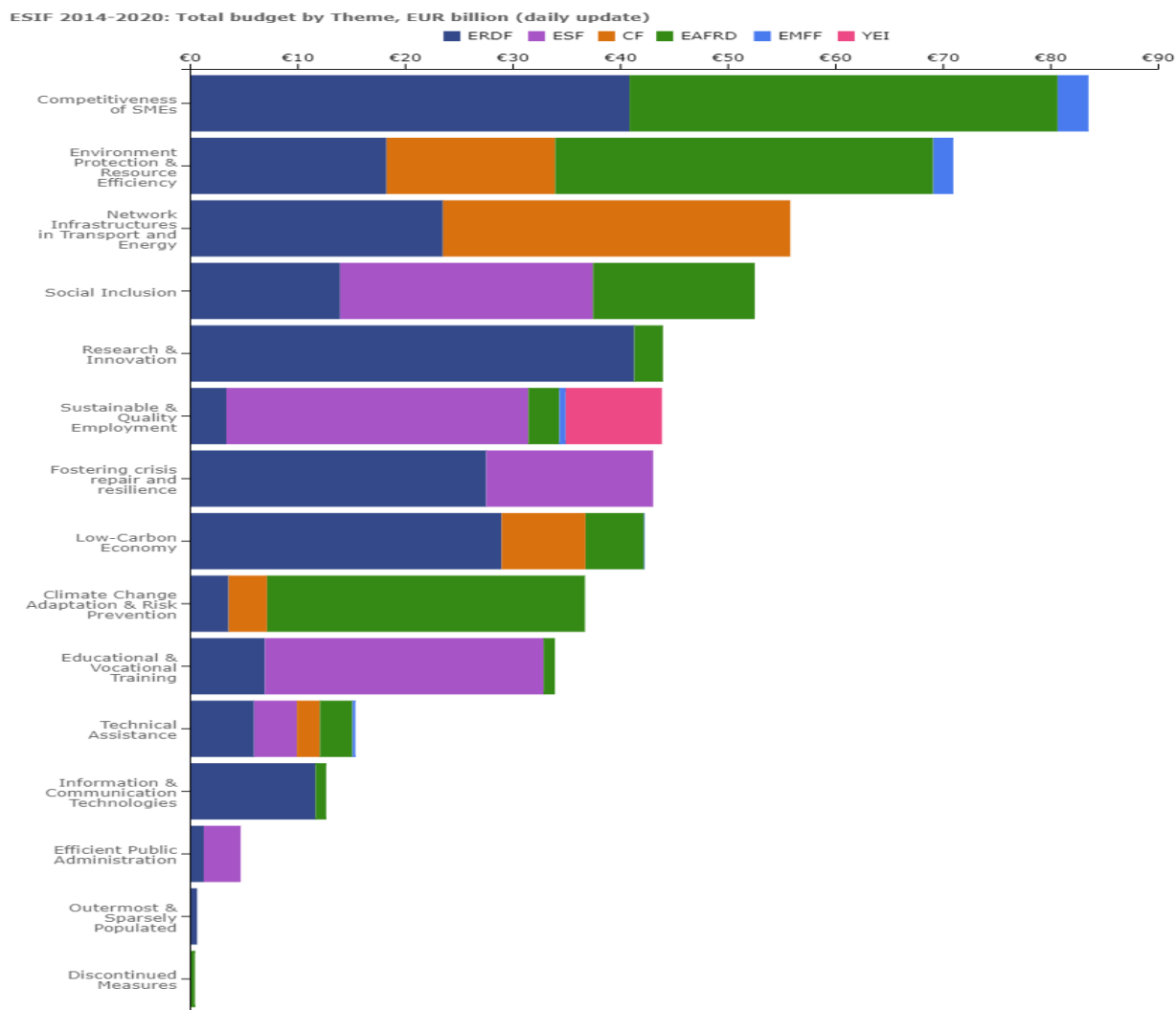
Na slici dolje (Grafički prikaz 3.), prikazani su prioritetni ciljevi financirani iz ESI fondova na području EU (2014.-2020.). Tako je uočljivo da je najviši iznos EU financiranja usmjeren doprinosu konkurentnosti malih i srednjih poduzeća, cilja koji se, većim dijelom, financira iz EFRR-a i EAFRD-a. Nakon toga, ističe se ostvarenje prioritetnog cilja zaštite okoliša i učinkovitosti resursa financiranjem iz EFRR, KF i EAFRD-a, potom ostvarenje prioritetnog cilja usmjerenog na infrastrukturnu mrežu energije i prometa financiranjem iz KF-a i EFRR-a te ostvarenje prioritetnog cilja usmjerenog na socijalnu uključenost najvećim dijelom financiranog iz ESF-a.

Tako, prioritetni cilj „Istraživanje i inovacije“ najvećim dijelom financiran je iz EFRR fonda, kao i cilj poticanje oporavka i otpornosti na krize, nisko-ugljična ekonomija te informacijske i komunikacijske

tehnologije. Ciljevi održivosti i kvalitetnog zapošljavanja, poticanja edukacije i obrazovanja te učinkovitosti javne administracije su većim dijelom financirani iz ESF-a.

Grafički prikaz 3.

Prioritetni ciljevi financirani EU sredstvima iz ESI fondova (2014.-2020.)



Izvor: Platforma otvorenih podataka za koheziju, <https://cohesiondata.ec.europa.eu/overview>

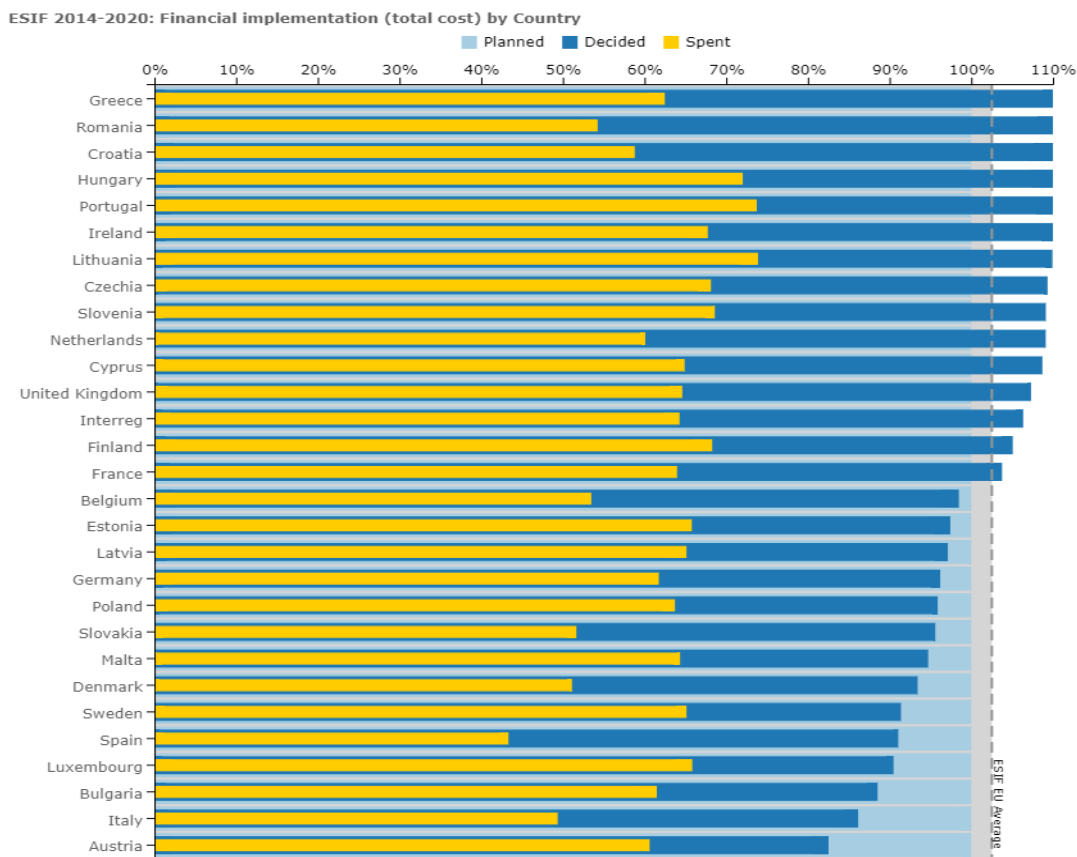
Na grafičkom prikazu 4., prikazana je iskorištenost EU sredstava prema zemlji članici (2014.-2020.) odnosno prikazana su:

- Planirana sredstva: Ukupan budžet programa
- Odlučena: Financijska sredstva dodijeljena odabranim projektima
- Potrošena: Prijavljeni troškovi odabranih projekata

Prema navedenom, zaključuje da najveći postotak ukupno potrošenih EU sredstava imaju Litvanija, Portugal i Mađarska - više od 70%, a najmanji postotak od 43% ima Španjolska. Za Hrvatsku, postotak potrošenih EU sredstava iznosi 59%, dok je odlučenih 123%.

Grafički prikaz 4.

Udio iskorištenosti EU sredstava prema zemlji članici (2014.-2020.)

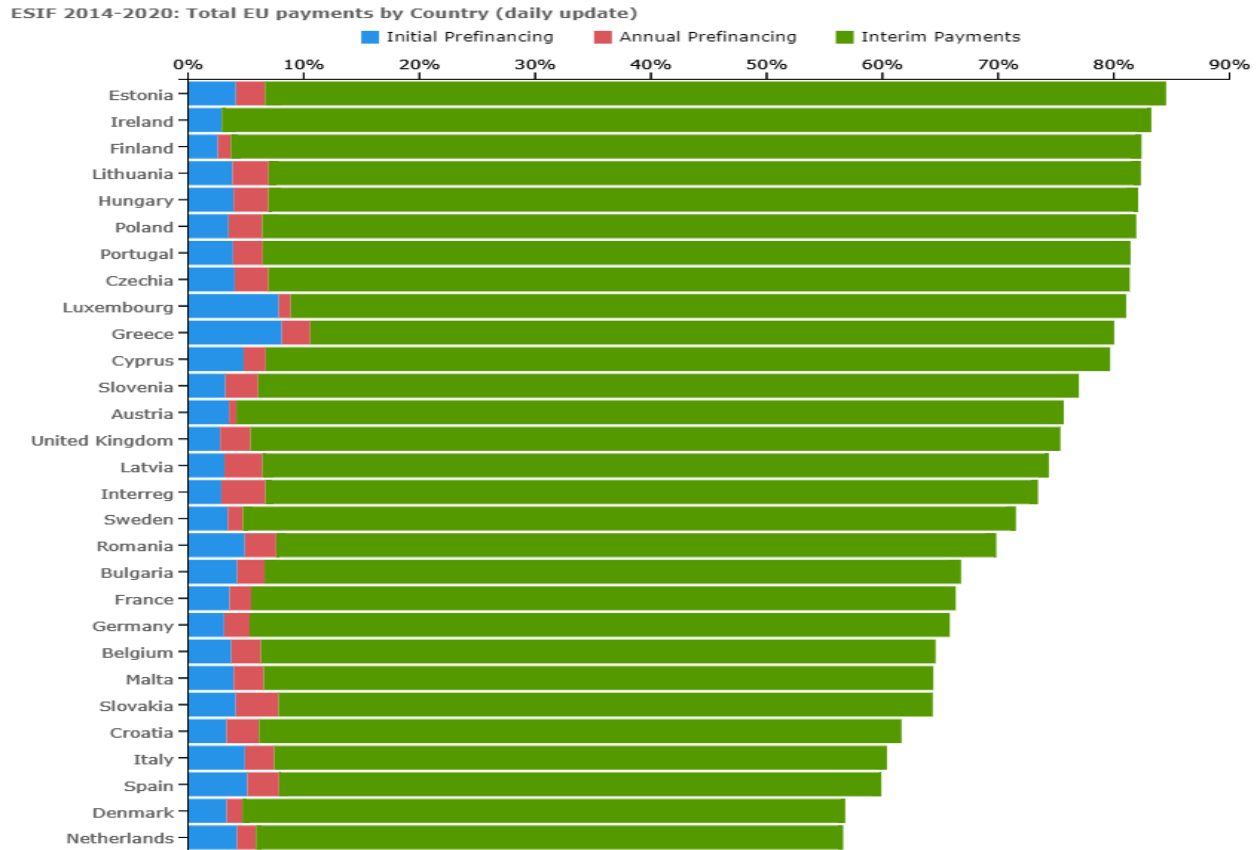


Izvor: Platforma otvorenih podataka za koheziju, <https://cohesiondata.ec.europa.eu/overview>

Po postotku ukupnih isplata Europske unije iz ESI fondova po zemljama članicama (2014.-2020.) prikazanih na grafičkom prikazu 5., izdvaja se Estonija na prvom mjestu s ostvarenih 85% isplata EU od ukupno dodijeljenih sredstava toj zemlji za ostvarenje ciljeva. Zadnje mjesto zauzela je Nizozemska s tek 57% isplata EU od ukupno dodijeljenih sredstava, dok ukupne isplate EU prema Hrvatskoj iznose 62%.

Grafički prikaz 5.

Udio ukupnih isplata EU po zemljama članicama (2014.-2020.)



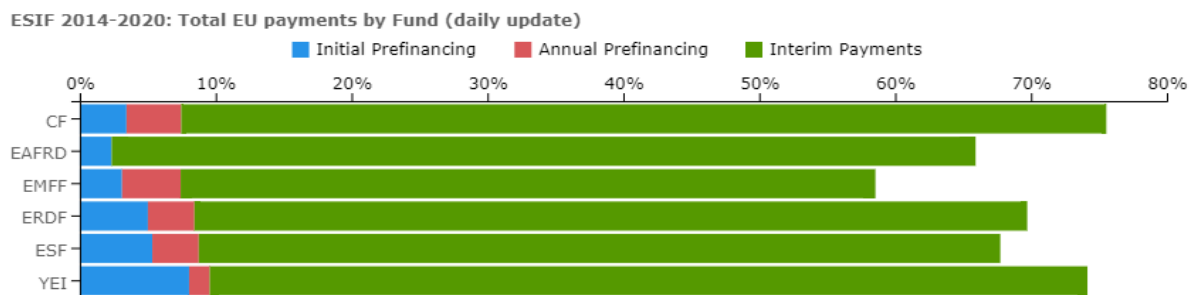
Izvor: Platforma otvorenih podataka za koheziju, <https://cohesiondata.ec.europa.eu/overview>

Postotak ukupnih isplata EU prema vrsti fonda s obzirom na ukupno dodijeljena sredstva zemlji članici prikazane su na grafičkom prikazu 6. na kojem je jasno vidljivo da je najveći postotak isplaćenih sredstava iz kohezijskog fonda (76%) i Inicijativi za zapošljavanje mladih (74%), a najmanji postotak iz Europskog fonda za pomorstvo i ribarstvo (59%).

Od ukupnih isplata EU prema fondu u Hrvatskoj, a koje su vidljive na slici 7., najveći postotak (89%) isplaćenih sredstava od ukupno dodijeljenih sredstava zemlji članici odnosi se fond „Inicijativa za zapošljavanje mladih“ što ga svrstava u jedini fond koji je iznad prosječnih EU isplata (74%). Najmanji postotak ostvarenih isplata EU je iz Europskog socijalnog fonda (ESF) u iznosu od 31% te je daleko ispod prosjeka EU (68%), kao što je i postotak ostvarenih isplata EU iz kohezijskog fonda.

Grafički prikaz 6.

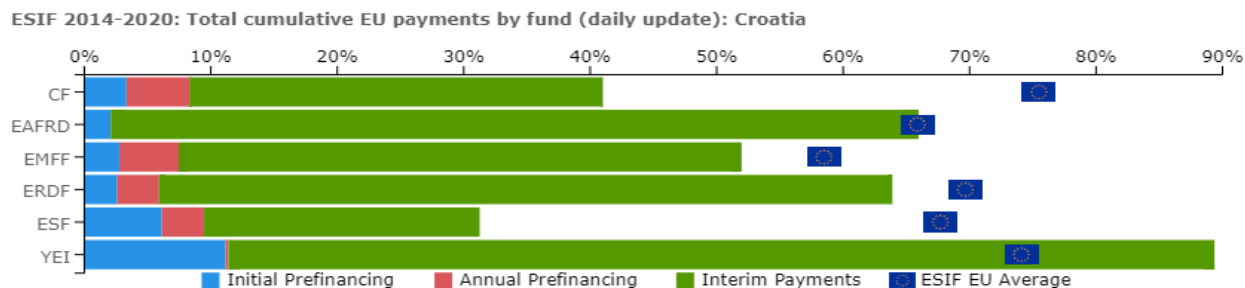
Udio ukupnih isplata EU prema fondu (2014.-2020.)



Izvor: Platforma otvorenih podataka za koheziju, <https://cohesiondata.ec.europa.eu/overview>

Grafički prikaz 7.

Udio ukupnih isplata EU prema fondu u Hrvatskoj (2014.-2020.)



Izvor: Platforma otvorenih podataka za koheziju, <https://cohesiondata.ec.europa.eu/overview>

2.2. Strukturni i investicijski fondovi

Od 2014. do 2020. godine, kohezijska politika financirana je putem Europskih strukturnih i investicijskih fondova (ESI) kojima se usmjerilo više od polovine sredstava EU-a za ostvarenje zajedničkih ciljeva.

Europski fondovi su financijski instrumenti koji služe za podupiranje provedbe pojedine politike Europske unije u zemljama članicama, a isplaćenim sredstvima iz ESI fondova doprinosi se održivom rastu, ulaganjima u stvaranje radnih mjesta, povećanju konkurentnosti te omogućava konvergencija manje razvijenih članica ka više razvijenim članicama i regijama Europske unije.

Prema zajedničkoj regulativi kojom je određena upotreba ESI fondova, Europska Komisija (EK) je identificirala jedanaest tematskih ciljeva u okviru kojih svaka država članica odabire investicijske prioritete

i definira svoje specifične ciljeve. U skladu s time, Hrvatskoj je od 2014. do 2020. godine bilo na raspolaganju pet ESI fondova.

Kohezijska politika Europske unije financirana je trima fondovima:

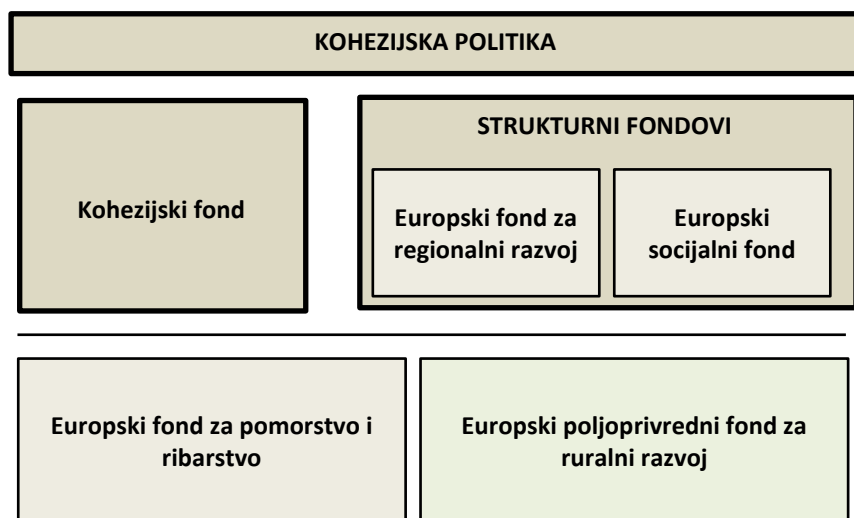
1. Kohezijskim fondom (KF-om)
2. Europskim fondom za regionalni razvoj (EFRR-om)
3. Europskim socijalnim fondom (ESF-om).

Europski fond za regionalni razvoj (EFRR) i Europski socijalni fond (ESF) spadaju u strukturne fondove. Od 2014.-2020. godine, Hrvatskoj su na raspolaganju bila još dva fonda iz ESIF-a, a to su:

4. Europski poljoprivredni fond za ruralni razvoj (EAFRD) i
5. Europski pomorski i ribarski fond (EFPR).

Grafički prikaz 8.

Europski strukturni i investicijski fondovi



Izvor: Prikaz autora

Kohezijski fond (KF) je namijenjen onim državama članicama koje su najmanje razvijene, a čija je vrijednost bruto nacionalnog proizvoda po stanovniku manja od 90% prosjeka Europske unije.

Zbog navedenog, korist ovog fonda je u smanjivanju gospodarskih i socijalnih razlika te promicanju održivog razvoja tako da se sredstva fonda upotrebljavaju za provođenje projekata koji doprinose

poboljšanje okoliša i infrastrukture koja je dio transeuropske prometne mreže (mreža cesta, željeznica, zračnih luka i vodne infrastrukture EU).

Aktivnosti kohezijskog fonda odnose se na: pružanje potpora gospodarstvu s niskom razinom ugljičnog dioksida, promicanju prilagodbe klimatskim promjenama i ublažavanju rizika, zaštiti okoliša, promicanju efektivne upotrebe resursa, promicanju održivog prometa te razvoju ključnih infrastrukturnih mreža.

Hrvatska korištenjem kohezijskog fonda, kao instrumenta kojim se provodi kohezijska politika, ima pravo na sufinanciranje u iznosu od 80% - 85% ukupne vrijednosti projekata, a iznos alokacije za svaku državu članicu određen je kriterijima površine i broja stanovnika.

Kohezijski fond karakterizira financiranje velikih infrastrukturnih projekata za koje se najmanje 25 milijuna eura izdvaja za okolišne, a 50 milijuna eura za poboljšanje prometne infrastrukture. Korisnici kohezijskog fonda su uglavnom tijela javne vlasti, dok poslovni sektor financiranje ostvaruje kroz sudjelovanje u postupcima javne nabave za isporuku dobara i usluga te za obavljanje radova.

Europski fond za regionalni razvoj (EFRR), koji spada u strukturne fondove, ima cilj poboljšanja ekonomske i socijalne kohezije EU te smanjenje razlika u razvoju između njenih regija i država članica.

Aktivnosti fonda odnose se na: ulaganja za stvaranje i očuvanje radnih mjesta putem potpora za ulaganja financijskih sredstava u razvoj malih i srednjih poduzeća (SME); infrastrukturna ulaganja; pružanje usluga stanovnicima u području energetike, okoliša, prometa te ICT-a; ulaganja u socijalnu, zdravstvenu i obrazovnu infrastrukturu; poticanje lokalnog i regionalnog razvoja i istraživanja; i tehnička pomoć.

Kroz korištenje Europskog fonda za regionalni razvoj Hrvatskoj je bilo na raspolaganju sufinanciranje od 50%-75% ukupne vrijednosti projekta, a sredstva fonda mogla su koristiti tijela državne uprave, SME, znanstveno-istraživački sektor te strane tvrtke koje se nalaze u regiji koja je pokrivena jednim od operativnih programa.

Europski socijalni fond (ESF) spada u strukturne fondove kao i EFRR te je jedan od važnijih načina financiranja organizacijama civilnoga društva. Fond je čiji se doprinos ogleda u poboljšanju gospodarske, socijalne i teritorijalne kohezije te je glavni financijski instrument EU za ulaganje u ljudske resurse (zaposlenike, mlade ljude i nezaposlene). Stoga, financijska sredstva fonda koriste se za povećanje mogućnosti zapošljavanja stanovnika EU, promicanje kvalitetnijeg obrazovanja te za poboljšanje situacija najugroženijih osoba koje su na rubu siromaštva.

Uloga fonda ističe se u ublažavanju posljedica koje uzrokuju ekonomske krize, a što se posebno odnosi na rast broja nezaposlenih i broja siromašnih. Udio sufinanciranja projekata je od 50% do 85% (95% u posebnim slučajevima) ukupnih troškova projekta. Udio sufinanciranja projekta ovisi i o bogatstvu regije u kojoj se želi sufinancirati projekt. Korisnici sredstava fonda su javne uprave, nevladine organizacije, dobrotvorne ustanove, dobrotvorne tvrtke, udruge radnika i poslodavaca, stariji radnici koji stječu nove vještine financiranim projektom, ili mladi koji trebaju dobiti posao, kao i oni pojedinci koje žele pokrenuti vlastitu djelatnost.

Cilj *Europskog poljoprivrednog fonda za ruralni razvoj (EAFRD)* je osnaživanje ruralnog razvoja Hrvatske. Tako, u neke od ciljeva fonda kojima se nastoje smanjiti postojeće gospodarske razlike među regijama EU razvojem poljoprivrednog sektora su postizanje konkurentnosti poljoprivrede, osiguranje ujednačenog prostornog razvoja ruralnih gospodarstava i ruralnih zajednica te stvaranje i očuvanje radnih mjesta. Udio sufinanciranja projekata iz ovog fonda je do 85% ukupne vrijednosti projekta. Fond je čijim se ciljevima želi podržati obnova i integracija ruralnih područja u ekonomske tokove zemalja EU.

Europskim fondom za pomorstvo i ribarstvo (EFPR) financira se ribarska industrija i zajednice obalnog područja kako bi postale što konkurentnije i održivije na tržištu ribarstva i akvakulture, ali i kako bi se ostvario ujednačen teritorijalni razvoj ribarstvenih područja. Udio sufinanciranja projekata iz ovog fonda je do 85% ukupne vrijednosti projekta. Tako, sredstva fonda se usmjeravaju na uzgoj ribe, školjkaša, podvodnog bilja odnosno na sve sektore ribarske industrije kao i na obradu i marketing proizvoda ribarstva.

2.3. NUTS klasifikacija regija

Nomenklatura prostornih jedinica za statistiku, (NUTS) sustav je za identifikaciju i klasifikaciju prostornih jedinica za potrebe službene statistike u zemljama članicama EU, koji je razvio Europski ured za statistiku, s ciljem uspoređivanja regija unutar Europe (Hrvatska gospodarska komora [HGK], 2015, str. 21)

NUTS predstavlja osnovu kojom se određuje koliko su određene prostorne jedinice prihvatljive za korištenje fondova EU u okviru kohezijske politike i na temelju podjele određenih teritorijalnih jedinica na NUTS regije, omogućava se izrada programskih dokumenata koji podržavaju kohezijsku politiku.

Kohezijska politika pokriva sve regije u EU, ali veći dio sredstava se alocira najpotrebitijim područjima. To su regije čiji je BDP ispod 75% prosjeka EU (Hrvatska gospodarska komora [HGK], 2015, str. 21).

Sredstva Kohezijske politike su, u financijskom razdoblju 2014. - 2020., usmjerena na tri skupine:

- manje razvijene regije - čiji je BDP po stanovniku manji od 75% prosjeka EU,
- regije u tranziciji - čiji je BDP po stanovniku između 75% i 90% prosjeka EU,
- razvijene regije - čiji je BDP po stanovniku viši od 90% prosjeka EU (Hrvatska gospodarska komora [HGK], 2015, str. 21).

Cilj kohezijske politike je smanjenje nejednakosti među regijama i državama članicama kako bi se na razini cijele Europske unije ostvario održiv gospodarski rast kojim bi EU ostvarila bolju konkurentnost na globalnom tržištu. To se posebno nastoji ostvariti u NUTS II regijama gdje su prisutni veliki infrastrukturni projekti, a čiji učinci su mjerljivi samo na većem području. Dobar primjer velikog infrastrukturnog projekta u Hrvatskoj je izgradnja Pelješkog mosta.

Sredstva kohezijske politike, namijenjena ostvarenju zajedničkih ciljeva EU, usmjeravaju se na NUTS regije prema broju stanovnika koji je ključan kriterij za određivanje NUTS regija u EU. Administrativna podjela i geografsko razgraničenje su još dva kriterija prema kojima se definiraju NUTS regije.

Tablica 1.

NUTS regije i broj stanovnika

<i>NUTS REGIJE</i>	<i>MINIMALAN BROJ STANOVNIKA</i>	<i>MAKSIMALAN BROJ STANOVNIKA</i>
<i>NUTS I</i>	<i>3.000.000</i>	<i>7.000.000</i>
<i>NUTS II</i>	<i>800.000</i>	<i>3.000.000</i>
<i>NUTS III</i>	<i>150.000</i>	<i>800.000</i>

Izvor: Prikaz autora

Hrvatska, kao zemlja članica, je uključena u nomenklaturu prostornih jedinica za statistiku (NUTS) te je, u financijskom razdoblju od 2014. do 2020. godine, bila podijeljena na dvije statističke regije što je ovisilo o stupnju gospodarske razvijenosti. Tako se razlikuje:

1. Kontinentalna Hrvatska i
2. Jadranska Hrvatska (Hrvatska gospodarska komora [HGK], 2015, str. 24)

Cijela Hrvatska je ostvarila pravo na korištenje sredstva iz strukturnih fondova stjecanjem članstva u EU. Tako se ovisno o stupnju razvijenosti pojedine regije (NUTS), razlikuje i ostvareni postotak sufinanciranja sredstvima namijenjenih provedbi kohezijske politike.

U Hrvatskoj su, u financijskom razdoblju od 2014. godine do 2020. godine, županije koje ne pripadaju Gradu Zagrebu bile uskraćene za korištenje ESI sredstava zbog toga utjecaja Grada Zagreba na cijelu NUTS regiju. Grad Zagreb ima visok BDP po stanovniku za razliku od ostalih županija koje imaju nisku razinu BDP-a po stanovniku. Utjecaj Grada Zagreba i podjela Hrvatske na dvije NUTS regije, Kontinentalnu i Jadransku Hrvatsku, negativno je utjecao na ostale županije jer su imale slabiju mogućnost korištenja sredstava iz ESI fondova. BDP po stanovniku cijele regije zbog Grada Zagreba, monocentričnog središta Hrvatske, prelazio je 60%, a s obzirom da su referentni pragovi kojima se određivala razina sredstava koja će se koristiti po regiji iznosila 45%, 60%, 75% i 100% prosjeka Europske unije prema BDP-u po stanovniku, ostale županije su bile uskraćene za korištenje najpovoljnije razine sredstava.

U kontinentalnoj Hrvatskoj živi 2,87 milijuna stanovnika s BDP-om 61,67% prosjeka EU 27 (za razdoblje 2008-2010), dok regija Jadranska Hrvatska ima 1,41 milijuna stanovnika s BDP-om 60,00% EU prosjeka. Prema navedenom, obje NUTS regije klasificiraju se kao manje razvijene regije, čiji je BDP po stanovniku manji od 75% prosjeka EU 27.

2.4. Kohezijska politike Europske unije

Kohezijska politika odnosno regionalna politika, instrument je financijske solidarnosti. Jedna je od najznačajnijih javnih politika Europske unije kojom se potiče ravnomjeran razvoj svih država članica i njihovih regija. Instrument je financijske solidarnosti jer usmjerava sredstva prema područjima i sektorima kojima su najpotrebnija. Njome se nastoje ostvariti napredak u pet ključnih područja:

1. Zapošljavanje – želi se postići zaposlenost 75% stanovnika čija je dob od 20 do 64
2. Istraživanje i Inovacije – planira se investiranje 3% ukupnog BDP-a EU u istraživanje i razvoj
3. Obrazovanje – želi se postići smanjenje stope napuštanja škola (manje od 10%) i trećeg stupnja obrazovanja za najmanje 40% stanovnika u dobi od 30 do 34 godine
4. Socijalna uključenost/siromaštvo – želi se postići smanjenje rizika siromaštva i socijalne isključenosti za najmanje 20 milijuna stanovnika zemalja članica EU
5. Klimatsko-energetski ciljevi – želi se postići smanjenje emisije stakleničkih plinova za 20% u odnosu na 1990. godinu i smanjenje do 30% ukoliko je moguće. Isto, želi se postići da se za 20% više energije dobiva iz obnovljivih izvora te se za isti postotak želi povećati energetska učinkovitost (Hrvatska gospodarska komora [HGK], 2015, str. 9-10).

U posljednjem sedmogodišnjem razdoblju, od 2014. do 2020., financiranje je bilo usmjereno na jedanaest tematskih ciljeva koji su bili usklađeni sa strategijom Europa 2020. čijoj je financijskoj potpori predana kohezijska politika. Strategijom Europa 2020. provode se tri prioriteta, a to su: pametan rast, održiv rast i uključiv rast.

Kao što je već prethodno navedeno (grafički prikaz 8.), kohezijska politika EU financirana je iz:

1. Kohezijskog fonda
2. Europskog fonda za regionalni razvoj te
3. Europskog socijalnog fonda

S obzirom na postojanje velikih razlika u razvitku nekih regija i zemalja članica EU, cilj regionalne politike je upravo smanjivanje ključnih ekonomskih, socijalnih i teritorijalnih razlika koje postoje u regijama EU i zemljama članicama. Stoga je gotovo trećina proračuna Europske unije usmjereno na Kohezijsku politiku - 351,8 milijardi eura u financijskom razdoblju od 2014. do 2020. godine.

U tom razdoblju, Hrvatska je upravljala sa dva operativna programa u sklopu kohezijske politike Europske unije. Jedan od njih, OP "Konkurentnost i kohezija", financira se iz EFRR-a i KF-a, a drugi OP "Učinkoviti ljudski potencijali", financira se iz ESF-a.

Operativni programima se određuju prioritetni ciljevi u koje će se usmjeravati sredstva iz fondova EU. Jednom usvojeni, predstavljaju osnovu za korištenje financijskih sredstava Kohezijske politike. U ovome slučaju, to se odnosi na financijsku perspektivu 2014. - 2020.. Područja koja nisu obuhvaćena Operativnim programima, ne mogu se sufinancirati sredstvima EU.

Na svakoj je državi članici da, u sklopu procesa programiranja, sama odredi svoje ulagačke i razvojne prioritete (u skladu s ciljevima strategije EUROPA 2020) te rasporedi alocirana sredstva za financijsko razdoblje na koje se OP odnosi prema tima investicijskim prioritetima (Hrvatska gospodarska komora [HGK], 2015, str. 33).

2.5. Operativni program "Konkurentnost i kohezija 2014.-2020."

OP "Konkurentnost i kohezija 2014.-2020." temeljni je programski dokument u kojem su određena prioritetna područja koja podržavaju provedbu kohezijske politike Europske unije. OPKK-om se nastoji ostvariti cilj Ulaganje za rast i radna mjesta infrastrukturnim ulaganjima u području energetike, prometa, zaštite okoliša i ICT-a te potporama razvoju poduzetništva i istraživačkim djelatnostima. Osim regionalne

nejednakosti koja je uzrokovana zemljopisnim i ostalim socioekonomskim čimbenicima, glavni izazovi sa kojima se Hrvatska susreće su niska konkurentnost gospodarstva, visoka nezaposlenost, raskorak između obrazovnog sustava i ekonomskih potreba, nedostatna gospodarska i javna infrastruktura i usluge (Europski strukturni i investicijski fondovi, 2021). Operativni program “Konkurentnost i kohezija” sufinanciran je iz EFRR-a i KF-a, a prioritetne osi obuhvaćaju:

1. Jačanje gospodarstva primjenom istraživanja i inovacija
2. Korištenje informacijske i komunikacijske tehnologije
3. Poslovna konkurentnost
4. Promicanje učinkovitosti i obnovljivih izvora energije
5. Klimatske promjene i upravljanje rizicima
6. Zaštita okoliša i održivost resursa
7. Povezanost i mobilnost
8. Socijalno uključivanje i zdravlje
9. Obrazovanje, vještine i cjeloživotno učenje
10. Tehnička pomoć

2.6. Operativni program “Učinkoviti ljudski potencijali 2014.-2020.”

Europski socijalni fond (ESF) kao instrument Europske unije zemljama članicama pruža potporu za ulaganje u ljudski kapital i jačanje konkurentnosti cjelokupnog područja EU. S obzirom da je pozornost OPULJP-a usmjerena na poboljšanje socioekonomskih uvjeta u Hrvatskoj, posebna usmjerenost na osobe koje su izložene riziku od siromaštva i socijalne isključenosti.

Osnovni cilj Operativnog programa “Učinkoviti ljudski potencijali” je pridonijeti rastu zapošljavanja i poboljšanju socijalne kohezije u Hrvatskoj (Hrvatska gospodarska komora [HGK], 2015). Tako razlikujemo ulaganja u prioritetna područja:

1. Visoka zapošljivost i mobilnost radne snage
2. Socijalno uključivanje
3. Obrazovanje i cjeloživotno učenje
4. Dobro upravljanje

Bespovratna sredstva EU kojima se financiraju projekti iz ESF-a pomažu ljudima u unaprjeđivanju svojih vještina i olakšavaju im integraciju na tržište rada. Usmjerenost ovog Operativnog programa je u borbi protiv siromaštva i socijalne isključenosti te poboljšanje učinkovitosti javne uprave.

3. RUDARENJE PODATAKA

3.1. Pojam i važnost rudarenja podataka

Većina uskladištenih podataka nikad nije korištena ili su podaci prikupljeni iz različitih izvora. Takva baza podataka sadrži ogromnu količinu skrivenog znanja koja organizaciji može biti od ključne važnosti za stjecanje konkurentske prednosti pa je sve veća potreba organizacija za pronalaženjem novih i učinkovitijih načina izvlačenja korisnih informacija iz skladišta podataka.

Tako rudarenje podataka predstavlja skup metoda i tehnika za istraživanje i analizu skupova podataka, najčešće velikih, na automatski ili poluautomatski način kako bi se među tim podacima pronašla određena nepoznata ili skrivena pravila i povezanost. To je umjetnost izvlačenja informacija odnosno znanja iz podataka, a razlikuju se opisne i prediktivne tehnike rudarenja podataka. Opisne (eksplorativne) tehnike su namijenjene za izvlačenje prisutnih, ali "zakopanih" informacija u masi podataka, dok su prediktivne tehnike namijenjene ekstrapolaciji novih informacija na temelju sadašnjih informacija.

Rudarenje podataka definira se i kao "netrivijalni proces identificiranja valjanih, novih, potencijalno korisnih i konačno razumljivih uzoraka u podacima pohranjenim u strukturiranim bazama podataka", gdje su podaci organizirani kao zapisi strukturirani prema kategorijskim, ordinalnim i kontinuiranim varijablama (Moss L.T. & Atre S., 2003, str. 381-397). To je disciplina koja se temelji na statistici i matematici, a svojom pojavom je pomogla u otkrivanju obradivih podataka u skladištu podataka, uklanjanju suvišnih i dobivanju što točnijih podataka na što je moguće brži način.

S obzirom na eksplozivni rast izvješća i grafikona, potreban je dodatan oprez kod interpretacije podataka jer ponekad previše informacija u podacima može značiti upravo suprotno, a to je da podaci uopće ne sadrže nikakve informacije. To može rezultirati postupnim smanjenjem znanja o korisnicima neke organizacije. Cilj rudarenja podataka je pružanje pomoći u rješavanju te kompleksnosti odnosno pružanje pomoći pri donošenju odluka jer se od tehnika rudarenja podataka očekuje da mogu pružiti više od samo stvaranja što realnijeg modela koji će pomoći u razumijevanju stvarnosti. Rudarenje podataka omogućuje ograničenje ljudske subjektivnosti u procesima donošenja odluka kao i baratanje s velikom količinom datoteka velikom brzinom (zahvaljujući rastućoj snazi računala).

Anketa provedena na portalu www.kdnuggets.com u srpnju 2005. godine dala je uvid u kojim se područjima najčešće koristi rudarenje podataka. Prema provedenoj anketi to su: bankarstvo (12%), CRM

(12%), izravni marketing (8%), otkrivanje prijevvara (7%), osiguranje (6%), maloprodaja (6%), telekomunikacije (5%), znanstvena istraživanja (4%) i zdravstvo (4%).

U bankarskom i financijskom sektoru upotrebljava se pri formiranju marketinških strategija, izradi analize rizika, određivanju kreditnog limita, i slično. U zdravstvenom sustavu za dijagnosticiranje bolesti, određivanju metoda liječenja, a u obrazovnom sektoru koristi se za određivanje uspješnosti prolaznosti studenata, odabira zanimanja prema demografskim i osobnim karakteristikama, određivanju profila i preferencija studenata koji su pristupili ispitu i slično. Učestala je upotreba i u javnom sektoru u predviđanju utjecaja poreza na proračun, predviđanju vremena i slično (Mumine K.K., 2017, str. 128-132). Javna baza podataka o EU projektima sadrži informacije o projektima kohezijske politike (EFRR, KF, ESF i IZM), o vrijednosti projekta, isplatama EU-a Hrvatskoj, programima i postignućima za razdoblje 2014-2020. koje predstavljaju predmet istraživanja ovog rada zajedno sa korištenim tehnikama rudarenja podataka.

Cilj analitičkog CRM-a (Customer Relationship Management), koji označava upravljanje odnosima s korisnicima, je ekstrakcija, pohrana, analiza i proizvodnja relevantnih informacija da bi se osigurao sveobuhvatan, integrirani pogled na korisnika u poslovanju kako bi se njegov profil i potrebe potpunije razumjeli. Zahvaljujući CRM-u organizacija može segmentirati korisnike, razviti i održavati dugoročne odnose sa stalnim korisnicima, odrediti korake postupanja prema onima koji to nisu i prilagoditi tržišne ponude i promociju. Primjerice, kod EU financiranja segmentacijom korisnika dobio bi se uvid u tip korisnika (SME, velika poduzeća, obrti, ministarstva i slično), njegovom ostvarenom iznosu sufinanciranja od strane EU, gradu, županiji, NUTS regiji financiranja projekta i slično, čime se dobije i sveukupan uvid u ostvarenje prioritetnih ciljeva EU na području Hrvatske, a time i ostvarenost operativnih programa (OPKK i OPULJP).

3.2. Proces rudarenja podataka

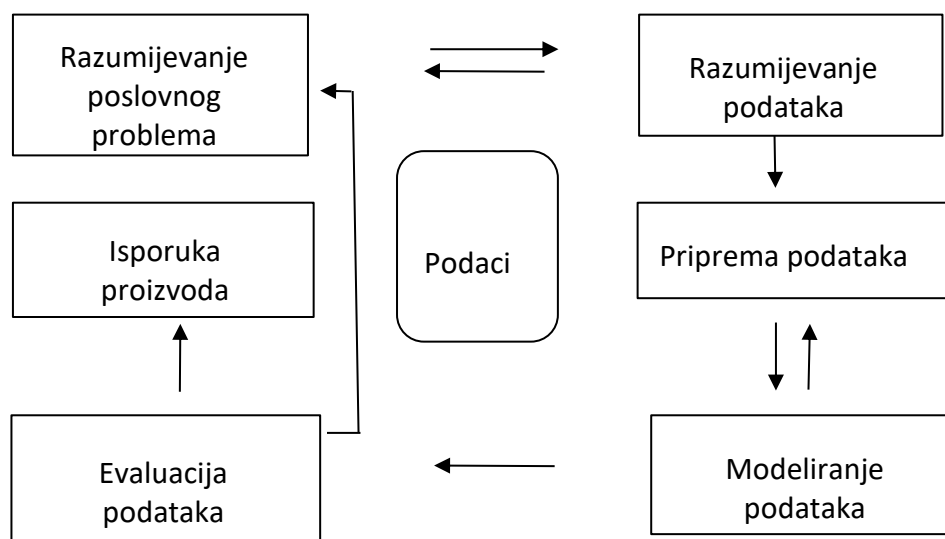
Svi prikupljeni podaci neke organizacije pohranjuju se u skladište podataka ili u podatkovno jezero otkud se započinje s njihovom analizom. To je početna faza rudarenja podataka kojom započinje primjena CRISP-DM metodologije.

CRISP-DM metodologija (*Cross-Industry Standard Process for Data Mining*) je robusna i dobro dokazana metodologija. To je procesni model koji opisuje životni ciklus projekta rudarenja podataka kroz šest faza s povezanim aktivnostima koje treba dovršiti tijekom svake faze gdje je naglasak stavljen na iterativni pristup za uspješnu provedbu projekta rudarenja podataka. Iterativni pristup podrazumijeva razdiobu projekta na iteracije radi procjene napretka projekta u odnosu na početno postavljene ciljeve projekta rudarenja

podataka čime se minimizira rizik od završavanja projekta bez ispunjenja početno postavljenih ciljeva projekta jer se izradom svakog manjeg dijela projekta, navedeni testira, provjerava i stavlja u upotrebu. Ukoliko se ustanove greške, vrše se potrebne izmjene te se iteracija ponavlja sve dok se ne utvrdi ispravnost iste. Svakom iteracijom dobije se precizniji model projekta rudarenja podataka odnosno svaka faza CRISP-DM metodologije informira sljedeću, ali i prijašnju fazu te svaki dio procesa trenutno i ponovno informira modele. Ova metodologija omogućuje repliciranje projekata (Grafički prikaz 9.).

Grafički prikaz 9.

Životni ciklus projekta rudarenja podataka CRISP-DM



Izvor: Prikaz autora

3.2.1 Razumijevanje poslovnog problema

U prvoj fazi rudarenja podataka, prije no što se pristupi podacima ili alatima, stječe se jasno razumijevanje problema koji se namjerava riješiti tako da se definira što se želi postići i da se odrede razlozi zbog kojih se žele postići određeni ciljevi. Zanemarivanjem početnog koraka može se utrošiti previše vremena kako bi se dobio pravi odgovor, ali na pogrešno pitanje. Cilj procesa je otkriti čimbenike koji bi mogli utjecati na ishod projekta. Fazu razumijevanja poslovnog problema čine četiri zadatka, a to su:

- *Određivanje poslovnih ciljeva* - Prvo što je potrebno napraviti u svakome projektu je saznati što se točno želi postići. Jedan od razloga za to je da analitičari mogu dosta vremena uložiti u analizu

podataka samo da bi se, na kraju, otkrilo da njihov menadžment nije zainteresiran za problem koji se istražuje. Stoga, potrebno je započeti s jasnim razumijevanjem: *problema* koji se želi istražiti, *poslovnih ciljeva, ograničenja* (ograničenja na ono što se može učiniti, rješenja koja se mogu koristiti, kada se posao mora dovršiti i slično), *utjecaja* (načina na koji se problem i moguća rješenja uklapaju u poslovanje). Na završetku ovog zadatka isporučuju se tri kratka izvješća: pozadina - opisivanje poslovne situacije koja pokreće projekt, poslovni ciljevi - definirati što organizacija namjerava postići projektom (npr. poslovni cilj može biti povećanje broja financiranih projekata za 10% u ovoj godini.) i kriteriji poslovnog uspjeha - definirati način mjerenja uspješnog rezultata projekta.

- *Procjena situacije* – U ovom zadatku nastoji se dobiti što više detalja o mogućim problemima s postizanjem poslovnih ciljeva. Podrobnijim istraživanjem i pronalaženjem činjenica u ovom zadatku, izgrađuje se jasnije objašnjenje problema opisanih u zadatku - *određivanje poslovnih ciljeva*. Na kraju, isporučuje se pet izvješća: *popis resursa* - sadrži popis svih sredstava dostupnih projektu (ljudi, softver, podaci); *zahtjevi, pretpostavke i ograničenja* - sadrži sve zahtjeve projekta (raspored završetka projekta, pravne i sigurnosne obveze, zahtjeve za prihvatljivo dovršen projekt), pretpostavke koje je dao projekt tijekom rudarenja podataka i ograničenja poput dostupnosti resursa ili veličine skupa podataka; *rizici i nepredviđene situacije* - sadrži identificirane uzroke koji bi mogli odgoditi dovršenje projekta te kontingencijski plan za svaki mogući uzrok koji bi mogao odgoditi dovršenje projekta; *terminologija* - sadrži poslovni pojmovnik i pojmovnik rudarenja podataka koji se odnosi na projekt; *analiza troškova i koristi* - sadrži koristi i troškove projekta. Ukoliko koristi ne premašuju troškove, potrebno je ponovno razmotriti analizu i projekt koji se želi provesti. Donositelji odluka će radije dodjeljivati resurse onim projektima koji smanjuju troškove, nego onima kojima je cilj povećanje prihoda.
- *Određivanje ciljeva rudarenja podataka* – Potrebno je definirati mali dio unutar veće slike svakog pojedinca koji sudjeluje u postizanju ciljeva. Primjerice, ukoliko je poslovni cilj smanjiti odljev kupaca, tada ciljevi mogu identificirati stope odljeva za različite segmente kupaca i razviti model koji predviđa koji kupci predstavljaju najveći rizik. Na kraju ovog zadatka pišu se dva izvješća: ciljevi rudarenja podataka - u kojem se definiraju rezultati rudarenja podataka poput modela, izvješća, prezentacija i obrađenih skupova podataka te kriteriji uspjeha rudarenja podataka koji služe kao podrška kriterijima poslovnog uspjeha, a potrebno ih je definirati u kvantitativnom smislu (npr.

točnost modela). Ukoliko kriteriji moraju biti kvalitativni, potrebno je identificirati osobu koja vrši procjenu.

- *Izrada projektnog plana* – je korak u kojem se opisuje plan za realizaciju projekta rudarenja podataka, a time i postizanja poslovnih ciljeva. Na kraju ovog zadatka pišu se dva izvješća. Prvo izvješće je *projektni plan* koji predstavlja kratak pregled plana proširen s rasporedom završetka svakog pojedinog zadatka, potrebnim resursima, ulaznim i izlaznim vrijednostima za svaki zadatak i definiranjem zavisnosti tako da naredni zadatak ne može započeti dok se trenutni zadatak ne dovrši. Drugo izvješće odnosi se na *početnu procjenu alata i tehnika* koje je važno dobro procijeniti rano u projektu jer mogu utjecati na ishod cijelog projekta.

3.2.2 Razumijevanje podataka

Faza razumijevanja podataka služi kao podrška prethodnom koraku, razumijevanju poslovnog problema. Nastoje se otkriti zanimljivi podskupovi na kojima će se formirati hipoteze koje odgovaraju poslovnom problemu uz pomoć kojih će se doći do informacija skrivenih u podacima. Također, identificiraju se mogući problemi koji postoje u prikupljenim podacima što može zahtijevati povratak na prethodnu fazu i revidiranje plana. Ključni zadaci ove faze su:

- *Prikupljanje podataka* koje se ostvaruje iz različitih izvora nakon što su postavljeni poslovni ciljevi i definiran plan rudarenja podataka. Svaki zadatak u planu ovisi o tome raspolaže li se s pravim podacima. Tako se razlikuju: postojeći podaci (oni kojima organizacija već raspolaže), kupljeni podaci te dodatni podaci (intervju, anketa i slično). Na kraju ovog zadatka isporučuje se samo jedno izvješće, a to je izvješće o početno prikupljenim podacima. U tom izvješću mora se verificirati da su podaci prikupljeni ili da je dobiven pristup podacima, da je testiran proces pristupa podacima te je potrebno verificirati postojanje podataka. Isto, potrebno je učitati podatke u sve alate koji će se koristiti za rudarenje podataka kako bi se provjerila kompatibilnost alata s podacima.
- *Opisivanje podataka* započinje s pripremom općeg opisa podataka s kojima se raspolaže. Rezultat/ispоруka ovog zadatka je izvješće o opisu podataka u kojem se opisuju izvori, formati podataka, broj slučajeva, brojevi i opisi polja podataka i sve druge informacije koji bi mogle biti od važnosti. Isto, izvršava se kratka procjena prikladnosti prikupljenih podataka ciljevima rudarenja

podataka. Zbog stjecanja dojma o informativnosti podataka, često se koriste jednostavne statističke tehnike.

- *Istraživanje podataka* uključuje detaljnije istraživanje svake varijable u podacima. Predstavlja fazu u kojoj se podaci istražuju postavljanjem upita, vizualizacijom podataka i tehnikama izvještavanja. Kod istraživanja podataka važno je upoznati podatke, uočiti probleme s njihovom kvalitetom i urediti ih tako da budu spremni za sljedeću fazu. Na kraju ovog zadatka sastavlja se izvješće o istraživanju podataka. Ono sadrži detaljniji opis podataka od izvješća o opisu podataka uključujući distribucije, sažetke i sve što upućuje na to da postoji problem sa kvalitetom podataka. Upravo u izvješću o istraživanju podataka mogu se pronaći rezultati istraživanja, uključujući i pronalazak prve ili početne hipoteze i njihov utjecaj na projekt (Smart Vision, 2016, par.2).
- *Verificiranje kvalitete podataka* izvršava se ispitivanjem o tome jesu li podaci potpuni, imaju li ikakvih pogrešaka ili nedostajućih vrijednosti te koji je postotak nedostajućih vrijednosti naspram ukupnim sakupljenim podacima. Sve uočene probleme sa kvalitetom podataka odnosno nedostatke podataka i način njihovog uklanjanja kao i korištenje alternativnog izvora podataka potrebno je iskazati u izvještaju o kvaliteti podataka. Ukoliko postoje ozbiljniji problemi s kvalitetom podataka te nije moguće pronaći odgovarajuće rješenje, potrebno je ponovno razmotriti poslovne ciljeve ili plan rudarenja podataka.

3.2.3 Priprema podataka

Faza pripreme podataka za analizu oduzima najveći dio vremena u cjelokupnom procesu rudarenja podataka. Veći dio prikupljenih i sačuvanih podataka zahtijeva dodatnu obradu prije nego što je spremna za modeliranje. Zadaci ove faze su:

- *Odabir podataka* je korak u kojem se odlučuje koji će se podaci koji su na raspolaganju zapravo koristiti kao osnova za daljnje analize, a koji neće. Isto, dat će se valjani razlozi za korištenje određenih podataka i obrnuto prema tome koliko su bitni za ostvarenje postavljenih ciljeva, vodit će se računa o kvaliteti podataka kao i o tehničkim problemima poput ograničenja broja redaka i stupaca koje alati mogu obraditi ili prikladnost formata podataka. Postoje dva načina odabiranja podataka. Jedan je odabiranje stavki (redaka) što uključuje donošenje odluka o tome koji će se

proizvodi, korisnici, projekti uključiti i slično, a drugi je odabiranje atributa ili obilježja (stupaca) što uključuje donošenje odluka o karakteristikama poput iznosa transakcije ili prihoda kućanstva ili veličine projekta (Smart Vision, 2016, par.2).

- *Pročišćavanje podataka* predstavlja bitan korak pripreme seta podataka za daljnju analizu. S obzirom da odabrani podaci sigurno nisu bez nekih pogrešaka ili nedostajućih vrijednosti, potrebne su izmjene istih: pronalaskom izvora podataka kako bi se izvršili potrebni ispravci, izuzimanjem određenih slučajeva ili stavki, zamjenom nekih stavki podataka zadanim vrijednostima i slično. Potreban je valjan razlog za ispravljanje ili unošenje novih podataka. Isto, ne mora se koristiti cijeli set podataka, već se mogu koristiti podskupovi podataka za dio ili za cijeli projekt rudarenja podataka.
- *Konstrukcija podataka* znači da je ponekad potrebno izvesti nove atribute iz već postojećih atributa, primjerice stopu financiranja projekta i ukupnu vrijednost projekta kako bi se dobio iznos dodijeljenih EU sredstava, ili je potrebno sumirati podatke. Tako se razlikuju izvješća za izvedene atribute (opisuju novonastala polja i objašnjavaju razloge njihova nastanka) i generirani zapisi (opisuju novonastale redove i razloge njihova nastanka).
- *Integracija podataka* je postupak spajanja nekih ili svih podataka koji se nalaze u različitim setovima podataka i provođenja operacija sažimanja podataka kako bi ih se pripremilo za fazu modeliranja.
- *Formatiranje podataka* je korak koji se vrši jer su podaci često u formatima koji su neprikladni za modeliranje (Brown, S.M., 2014). Primjerice, transformiranje kategorijskih varijabli u numeričke varijable i obrnuto.

Numerički podaci mogu biti kontinuirani ili diskretni. Kontinuirani podaci su oni koji mogu imati bilo koju vrijednost i koji su izraženi kao decimalni brojevi, dok diskretni podaci imaju cjelobrojne vrijednosti (Žuvić-Butorac M., 2006, str. 63-79). Primjeri kontinuiranih podataka su ukupna vrijednost projekta, vrijednost EU financiranja, vrijednost koncentracije šećera u krvi, indeks tjelesne mase i slično, dok se kao primjer diskretnih podataka može navesti broj djece, broj projekata, godine trajanje projekta, mjeseci trajanja projekta i slično.

S druge strane, kategorijski podaci mogu se podijeliti na ordinalne i nominalne podatke. Ordinalni podaci su raspoređeni na ljestvici vrijednosti i među njima se može uspostaviti odnos. Ono što je važno naglasiti je da se vrijednost na ljestvici ne može kvantificirati (Žuvić-Butorac M., 2006, str. 63-79). Nominalni podaci

su oni koji se opisuju riječima kao što je spol, rasa, krvna grupa, NUTS regija, tip fonda, prioriteta os i slično. Kategorijski podaci se još mogu kvantificirati kodiranjem. Na primjer, označavanje NUTS regija sa brojevima od 0 i 1 jer postoje dvije NUTS regije, Jadranska Hrvatska i Panonska

Važno je istaknuti da informacijska vrijednost podataka opada od numeričkih ka kategorijskim varijablama pa se tako numerički podaci mogu prevesti u kategorijske varijable kao što je ukupna vrijednost projekta u ordinalne varijable definiranjem kategorija: malen projekt, srednji i veliki projekt, dok obrnut slijed nije moguće izvršiti. Zbog toga numerički podaci imaju veću informacijsku vrijednost, iako samo snižavanje njihove vrijednosti kategorizacijom može biti korisno kod sumarnih analiza podataka (Žuvić-Butorac M., 2006, str. 63-79).

Rezultat faze pripreme podataka u procesu rudarenja podataka je set podataka koji je spreman za modeliranje i izvješće koje opisuje skup podataka. Ukoliko se napravi pogreška u ovoj fazi procesa rudarenja podataka, sljedeća faza neće dati pouzdane rezultate.

3.2.4. Modeliranje

Modeliranje podataka je ključna faza procesa rudarenja podataka jer se u ovoj fazi obavlja analiza podataka i pronalaze informacije koje postoje u podacima. Sve prethodne faze smatraju se pripremom za ovu fazu procesa. Kod modeliranja podataka mogu se primjenjivati različite tehnike rudarenja podataka na isti problem, a čest je slučaj da se tijekom modeliranja primijete neki problemi s podacima ili se pak dobiju nove ideje za kreiranje novih podataka. Ova faza uključuje dosta isprobavanja, pogrešaka i novih pokušaja te predstavlja najzujbudljiviju fazu CRISP-DM procesa vremenski manje zahtjevnju u odnosu na fazu pripreme podataka koja je vremenski najzahtjevnija kako bi podaci bili ispravni za uvođenje u model. Faza modeliranja podataka se provodi kroz tri međusobno ovisna koraka:

- *Odabir tehnike modeliranja* razmatra koliko su odgovarajuće tehnike modeliranja koje su predložene u fazi razumijevanja problema. Nova saznanja o svojstvima i karakteru podataka koji se analiziraju, proizašla iz faza razumijevanja i pripreme podataka, mogu dovesti do promjene inicijalnog odabira u korist neke prikladnije tehnike (Ujević F., str. 3). S obzirom da rezultat cjelokupnog procesa ovisi o tome koja je tehnika modeliranja korištena, potreban je dodatan oprez kada se vrši pregled odabranih tehnika modeliranja.

- U koraku *odabira procedure testiranja*, prije konstrukcije modela potrebno je definirati proceduru za testiranje kvalitete generiranih modela (Ujević F., str. 3). Tako, potrebno je razdvojiti set podataka na dva skupa gdje se jedan skup podataka koristi za konstrukciju modela, a drugi skup podataka, testni uzorak, za testiranje modela. Skup podataka za konstrukciju modela (training data) koristi se za uklapanje matematičkih obrazaca u model podataka, dok se drugi skup podataka, testni uzorak (test data), koristi tokom procesa izrade modela training setom podataka kako bi se izbjegla pojava *overfittinga* (stvaranja modela koji je savršen za samo jedan set podataka, ali ni za jedan drugi). Za reprezentativnost testnog uzorka koriste se statističke tehnike kako bi se osigurala kvaliteta modela. Nakon postizanja željenog stanja skupa podataka, testnog uzorka, odabranom tehnikom modeliranja se konstruiraju modeli kako bi se izdvojili oni koji pokazuju bolje rezultate na testnom uzorku odnosno oni koji su kvalitetniji.
- U postupku *konstrukcije modela*, variranjem različitih parametara, koji su određeni tehnikom modeliranja, traži se njihova optimalna kombinacija kako bi se dobio kvalitetni model (Ujević F., str. 3). Većina alata pruža mogućnost prilagodbe raznih postavki parametara koje utječu na strukturu konačnog modela. Isto, potrebno je opisati, navesti vrstu dobivenog modela (linearna regresija, stablo odlučivanja) i korištene varijable. U ovom koraku, potrebno je obratiti pozornost na preostale raspoložive resurse jer razvoj modela može potrajati duže no što je planirano.

3.2.5. Evaluacija

Ova faza usmjerena je na evaluaciju izrađenih modela i na donošenje odluka o tome kako upotrijebiti dobivene rezultate (Shafique, U. & Quaiser H., 2014, str. 217-222). Izrađeni model trebao bi biti kvalitetan. Prije prelaska na fazu isporuke, potrebno je još jednom, ali temeljitije izvršiti procjenu modela i pregledati poduzete korake kako bi se ustanovilo da li će dobiveni model pomoći u ostvarivanju osnovnih ciljeva projekta. Isto, potrebno je utvrditi postoji li neki poslovni problem koji je preskočen odnosno koji prilikom izrade modela nije uzet u obzir (Wirth, R., & Hipp, J., 2000, str. 29-39). Ukoliko se ustanove razlozi za dodatnom korekcijom modela, određuje se nova kombinacija parametara tehnike modeliranja s kojom se inicira nova iteracija konstruiranja modela (Ujević F., str. 3). Fazu evaluacije može se podijeliti se na tri glavna zadatka, a to su: ocjena rezultata, pregledavanje procesa i određivanje sljedećeg koraka. Ovisno o rezultatima ocjene rezultata i pregledavanju procesa, slijedi donošenje odluka o narednim koracima. Primjerice, odluka o tome da li će se preći na posljednju, fazu isporuke, ili će se donijeti odluka o novom

projektu rudarenja podataka gdje uzimaju u obzir preostali resursi i proračun koji mogu utjecati na donošenje odluka (Smart Vision, 2016, parag. 4).

3.2.6. Isporka

Isporka je posljednja faza CRISP-DM metodologije. Prije primjene konstruiranih modela potrebno je izraditi plan primjene modela u kojem se definira strategija i konkretizira način praktične upotrebe modela (Ujević F., str. 3). Ova faza je usmjerena na određivanje načina upotrebe stečenog znanja (tako da ga korisnik može koristiti) i dobivenih rezultata u poslovanju. U većini slučajeva će korisnik, a ne analitičar podataka biti taj koji će implementirati rezultate.

Kako bi se pratila isporuka rezultata rudarenja podataka, projektu je potreban detaljni plan monitoringa. Tako, na kraju svakog projekta piše se završni izvještaj kako bi se utvrdilo što se ispravno napravilo, gdje su nastale greške, što se može poboljšati, stečeno iskustvo tijekom projekta i slično (Smart Vision, 2016, parag. 4). Taj izvještaj sažima cijeli projekt sklapanjem svih do tada napravljenih izvještaja i dodavanjem sažetog pregleda cjelokupnog projekta i njegovih rezultata. Izvještaj prikazuje bitna iskustva stečena u procesu analize podataka, te opisuje i objašnjava postignute rezultate (Ujević F., str. 4).

3.3 Metode rudarenja podataka

Ključni korak projekta rudarenja podataka je odabrati one metode koje najviše odgovaraju postojećem poslovnom problemu. Metodama rudarenja podataka izvlače se korisne informacije iz ogromnih količina podataka pa je, s obzirom na veliku količinu podataka koja se dnevno pohranjuje, interes organizacija za rudarenjem podataka u konstantnom porastu radi pružanja pomoći kod donošenja odluka. Rudarenjem podataka izrađuje se model koji predstavlja algoritam (povezuje ulaze i ciljne varijable ili izlaze). Uspješnost provođenja rudarenja podataka ovisi o izabranoj metodi i o kvaliteti podataka. Isto, preporučuje se upotreba više različitih metoda na istom skupu podataka, kako bi se njihovi rezultati mogli međusobno usporediti. Tim postupkom postiže se raznovrsnost te se pruža više mogućih rješenja koja se mogu primijeniti za određeni slučaj. U nastavku ovoga rada detaljnije se obrađuju neke od metoda rudarenja podataka, a to su: korelacija, anova, klasteriranje i stabla odlučivanja.

3.3.1. Korelacija

Pod pojmom korelacije podrazumijeva se međuzavisnost, odnosno povezanost slučajnih varijabli (Rozga A., 2006, str. 179-185). Ako korelacija postoji moguće je odrediti njezin smjer i intenzitet, ali i testirati njezinu značajnost uobičajenim t-testom. Kod smjera korelacije razlikuje se pozitivna i negativna korelacija. Pozitivna korelacija je prisutna onda kada rast jedne varijable prati rast druge varijable, odnosno pad jedne varijable prati pad druge varijable. Smjer i intenzitet povezanosti između dviju varijabli ovisi o njihovoj kovarijaciji, što znači da varijacije jedne varijable ovise o varijacijama druge varijable. Mjerenjem intenziteta povezanosti između varijabli dobiva se koeficijent linearne korelacije (r) koji može poprimiti vrijednost iz segmenta $-1 \leq r \leq +1$ pa tako prema intenzitetu korelacija može biti:

$r > 0.80$ jaka pozitivna korelacija,

$0.50 < r \leq 0.80$, srednje jaka pozitivna korelacija,

$0 < r \leq 0.50$, slaba pozitivna korelacija.

Na isti način zaključuje se ako je korelacija negativna. (Rozga A., 2006, str. 179-185)

Ukoliko je korelacija potpuna, govori se o funkcionalnoj povezanosti varijabli. U tome slučaju vrijednost jedne slučajne varijable može se s potpunom sigurnošću odrediti pomoću vrijednosti druge varijable te vrijednost koeficijenta linearne korelacije iznosi ± 1 . Ukoliko je korelacija djelomična, radi se o stohastičkoj ili statističkoj vezi (Rozga A., 2006, str. 179-185).

Postoji nekoliko različitih koeficijenta korelacije, od kojih je svaki prikladan za određeni tip podataka. Najčešći je Pearsonov koeficijent linearne korelacije (r) koji se koristi na omjerne i intervalne skale mjerenja (numeričke varijable). Ako je korelacija linearna, tada se ista može analitički opisati linearnom regresijskom jednadžbom kojom se najbolje opisuje statistička povezanost jedne numeričke varijable s drugom ili više numeričkih varijabli (Rozga A., 2006, str. 179-185). Ukoliko se nakon izračuna koeficijenta korelacije pokaže da vrijednost koeficijenta nije značajna, tada je predviđanje vrijednosti zavisnih parametara regresijskom jednadžbom beskorisno. Korelacija je, zbog jednostavnosti primjene i sposobnosti identificiranja temeljnih zavisnosti između varijabli, jedna od metoda rudarenja podataka koja je najviše u upotrebi. Zbog sposobnosti identificiranja temeljnih zavisnosti između varijabli može generirati korisne uvide koji mogu biti polazišna točka za daljnje istraživanje i stjecanje uvida. Također, ukoliko su u setu podataka prisutni

outlieri, rezultati kovarijance i korelacijske matrice nisu pouzdani. Samo mali dio outliera u setu podataka, a u nekim slučajevima i samo jedan outlier, može iskriviti konačne rezultate što tada korelacijsku metodu čini u potpunosti beskorisnom.

3.3.2 Analiza varijance

Djelovanje različitih promjenjivih faktora na numeričku vrijednost slučajne varijable X može se istraživati analizom varijance (Rozga A., 2006, str. 167-172). Točnije, kod testiranja postoji li značajna razlika između aritmetičkih sredina tri ili više osnovnih skupova, odnosno da li promatrani uzorci pripadaju istoj populaciji.

Analizom varijance ukupna varijanca raščlanjuje se na dio koji se pripisuje djelovanju nekoga (ili nekih) promjenjivih faktora i dio koji ostaje neobjašnjen (rezidualni dio) (Rozga A., 2006, str. 167-172). Najčešće je u primjeni analiza varijance s jednim promjenjivim faktorom (One-Way ANOVA) i analiza varijance s dva promjenjiva faktora (Two-Way ANOVA).

Kod analize varijance s jednim promjenjivim faktorom (ANOVA) cilj je ispitati odnos varijacija između uzoraka s varijacijama unutar uzoraka što se naziva empirijskim F-omjerom. Ako je F-omjer statistički značajan, odbacuje se nulta hipoteza da se aritmetičke sredine značajno ne razlikuju, a ako F-omjer nije statistički značajan, odnosno ako je signifikantnost testa manja od 5%, prihvaća se alternativna hipoteza da postoji statistički značajna razlika između aritmetičkih sredina.

Drugim riječima, nultom hipotezom se pretpostavlja da određeni faktor značajno ne djeluje na promatranu slučajnu varijablu, dok se alternativnom hipotezom pretpostavlja da određeni faktor statistički značajno djeluje na promatranu slučajnu varijablu. Primjerice, zbog toga što je za OPKK predviđen određeni iznos sredstava iz EU, a unutar OPKK postavljeni su prioritetni ciljevi koji se nastoje ostvariti te kojima se dodjeljuju različiti iznosi sredstava, postavljamo nultu hipotezu (H_0) da ugovoreni projekti značajno ne djeluju na raspodjelu sredstava prema prosječnom iznosu dodijeljenih ESI sredstava s obzirom na prioritetnu os, dok alternativnom hipotezom pretpostavljamo da ugovoreni projekti značajno djeluju na raspodjelu sredstava prema prosječnom iznosu dodijeljenih ESI sredstava s obzirom na prioritetnu os. Kod analize varijance s dva promjenjiva faktora postojala bi dva faktora koja statistički značajno djeluju na promatranu slučajnu varijablu, primjerice županija i prioritetna os.

U osnovi, vrši se testiranje grupa kako bi se vidjelo postoji li razlika među njima. Tako financirane projekte možemo razlikovati prema prioritetnim osima OPKK-a na čak desetak grupa kao što su: Jačanje gospodarstva primjenom istraživanja i inovacija; Korištenje informacijske i komunikacijske tehnologije; Poslovna konkurentnost; Promicanje učinkovitosti i obnovljivih izvora energije; Klimatske promjene i upravljanje rizicima, itd.. Navedenim primjerom želi se vidjeti za koji prioritetni cilj, koji se nastoji ostvariti, je dodijeljen veći iznos sredstava.

U regresijskoj analizi, ANOVA se upotrebljava kako bi se odredio utjecaj nezavisne varijable na zavisnu varijablu. ANOVA je važna i svestrana analiza koja se koristi u analizi podataka i u analiziranju veza koje postoje među podacima. Kod ANOVE bitno je razlikovati:

- zavisnu varijablu - koja je kontinuirana varijabla (Y) - dodijeljena ESI sredstva
- nezavisnu varijablu - koja je kategorijska varijabla (X) - prioritetna os. Ove varijable se nazivaju i faktori.
- Nultu hipotezu - aritmetičke sredine se značajno ne razlikuju.
- Razinu faktora - Svaki faktor može imati više razina poput veliki, srednji, mali projekt. Ili prema prioritetnoj osi, to su: Korištenje informacijske i komunikacijske tehnologije, Poslovna konkurentnost, Promicanje učinkovitosti i obnovljivih izvora energije, Klimatske promjene i upravljanje rizicima, itd.

3.3.3 Klasteriranje

Klasteriranje, odnosno grupiranje podataka, jedna je od najčešćih tehnika analize podataka. Osnovni cilj klasteriranja je grupiranje podataka prema zadanom kriteriju sličnosti uz postizanje velike sličnosti između podataka unutar iste grupe, a male sličnosti između podataka koji pripadaju različitim grupama (Pratljačić S., 2019, str. 1-8). Primjerice, grupiranje EU projekata prema iznosu financiranja čime će se grupe međusobno razlikovati prema veličini projekata (veliki, srednji, mali projekt), a unutar grupe će biti slični po iznosu financiranja odnosno projekti sličnih iznosa bit će svrstani u istu grupu pa će tako projekt od tristo milijuna kuna biti svrstan pod velike projekte, a projekt od pedeset tisuća kuna pod male projekte i slično. Unutar grupe, projekti će biti slični po svojim karakteristikama, ali ne i isti po nekim drugim karakteristikama poput prioritetne osi, fonda iz kojeg su financirani ili županije u kojoj će se projekt odvijati i slično.

S obzirom na to da postoji veći broj algoritama za klasteriranje, postavlja se pitanje koji odabrati za određeni poslovni problem jer ne znači da će isti algoritam biti učinkovit na različitim skupovima podataka. Onaj koji je dobar, neće biti osjetljiv ni na količinu podataka ni na broj varijabli te će se moći nositi sa outlierima.

Klasteriranje je najraširenija deskriptivna metoda analize podataka koja se koristi kod velikih skupova podataka, a cilj joj je pronaći homogene podskupove koji se mogu analizirati. Dakle, to je statistička metoda grupiranja varijabli u određeni broj klastera (segmenata) koji nisu unaprijed definirani nego se otkrivaju tokom postupka te nije poznat klaster kojem će pripasti svaka varijabla. Razlog tome je što nema zavisne varijable jer klasteriranje spada u deskriptivne, a ne u prediktivne metode. Kao rezultat uspješno provedenog klasteriranja javlja se unutarnja homogenost i vanjska heterogenost grupiranih varijabli. Najčešće područje upotrebe metode klasteriranja je u segmentaciji tržišta prema korisnicima sličnih karakteristika radi što boljeg pozicioniranja proizvoda. Primjerice, u marketingu, klasteriranje je korisno za identificiranje različitih profila korisnika gdje donositelj odluka nakon formiranja klastera može formirati strategiju pristupa svakom identificiranom segmentu korisnika.

Većina prediktivnih algoritama nije dobra s pretjerano velikim brojem varijabli jer korelacije između varijabli mogu utjecati na njihovu sposobnost predikcije. S obzirom da je teže opisati heterogenu grupu s malim brojem varijabli, grupe formirane klasteriranjem su korisne jer se radi o homogenim grupama koje se mogu opisati malim brojem varijabli. Općenito, klasteriranje se koristi kao preliminarna analiza podataka te se očekuje da cjelokupan proces klasteriranja bude razumljiv svima. Isto, ako se podaci prirodno strukturiraju u dva klastera, ali su tražena tri klastera, postojat će proizvoljni element u podjeli na broj klastera. Stoga, veoma je važno otkriti koji je potreban broj klastera u određenoj bazi podataka. Dobar postupak klasteriranja će:

- otkriti strukture prisutne u podacima
- omogućiti određivanje optimalnog broja klastera
- dati jasno diferencirane klastere
- dati klastere koji su stabilni kod pojave malih promjena u podacima
- učinkovito obrađivati velike količine podataka

- se nositi s različitim vrstama varijabli (kvantitativne i kvalitativne) ukoliko je potrebno (Tuffery, S., 2011, str. 242).

Razlikujemo hijerarhijsko i nehijerarhijsko klasteriranje. Kod nehijerarhijskog klasteriranja svaki objekt se smješta u jedan od k disjunktnih klastera. Broj klastera mora biti unaprijed određen. Najpoznatiji nehijerarhijski algoritam je algoritam k -srednjih vrijednosti tj. k -means (Ungaro T., 2016, str. 2-12).

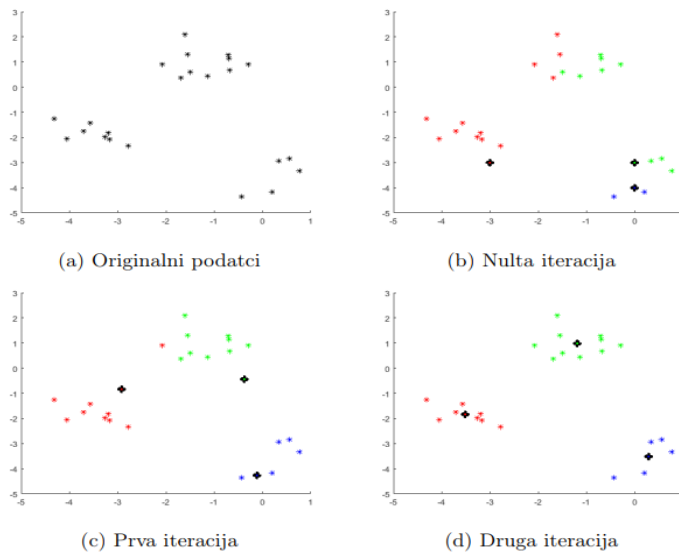
K-srednjih vrijednosti ili k-means klasteriranje koristi se kod obrade podataka koji nemaju definirane kategorije. Glavni cilj algoritma je podjela podataka u k grupe, gdje k predstavlja broj klastera koje je potrebno unaprijed odrediti. Algoritam radi iterativno, particionira skup u manje klastere, računa njihove centre pa razmješta elemente skupa po klasterima ovisno o njihovoj udaljenosti od centara klastera. Na grafičkom prikazu 10., prikazan je iterativni rad algoritma čiji je cilj particionirati skup na k -klastera odnosno pronaći idealnu particiju da funkcija cilja ima minimalnu vrijednost varijance, tj. traže se najhomogeniji klasteri. Najprikladniji je za primjenu zbog svoje učinkovitosti u klasteriranju velikih setova podataka, no algoritam je moguće koristiti samo na numeričkim vrijednostima te se loše nosi sa netipičnim vrijednostima, tj. outlierima. Algoritam k -srednjih vrijednosti (engl. k -means) sastoji se od nekoliko koraka:

1. Odabira broja klastera k .
2. Inicijaliziranja k centara klastera (slučajnim odabirom).
3. Pridruživanja svakog od n objekata najbližem centru klastera.
4. Promjena centara klastera pretpostavljajući da su objekti stavljeni u točne klastere.
5. Ponavljanje trećeg i četvrtog koraka sve dok niti jedan od n objekata ne promijeni svoj klaster.

Centri klastera su obično srednja vrijednost objekata unutar klastera. Prednost ovog algoritma je što je jednostavan za implementaciju, te intuitivan jer optimizira sličnost unutar klastera (Ungaro T., 2016, str. 2-12).

Grafički prikaz 10.

K-means algoritam



Izvor: Nemčić J. (2017): Spektralne metode grupiranja podataka, Sveučilište J.J. Strossmayera u Osijeku, Osijek [Internet], raspoloživo na: <https://repositorij.mathos.hr/islandora/object/mathos%3A197/datastream/PDF/view>

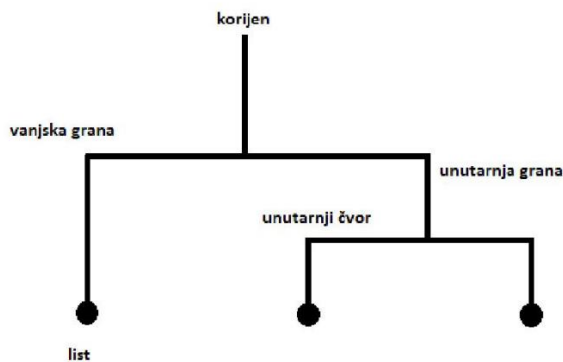
K-mod klasteriranje omogućava klasteriranje velikih setova podataka kategorijskih varijabli bez potrebe za njihovom transformacijom. Karakteristika nominalnih varijabli je da se među njima ne može uspostaviti prirodni poredak jer su neuređene, a njihova transformacija u numeričke varijable ne daje dobre rezultate. Algoritam *k-modova* je sličan algoritmu *k-srednjih vrijednosti*, te se struktura algoritma ne mijenja. Jedina razlika je što se srednje vrijednosti klastera zamjenjuju s modovima odnosno s najčešćom vrijednošću unutar svakog klastera kako bi se odredila njegova centralna vrijednost. Isto, ovaj algoritam je robusniji na outlieri od *k-means* algoritma i njime se vrši klasteriranje kategorijskih podataka poput demografskih podataka korisnika, tržišnih segmenata ili anketa. Koristan je kod analize podataka za dobivanje uvida u podatke i donošenje informiranih odluka.

Hijerarhijsko klasteriranje se može podijeliti na aglomerativnu i divizijsku metodu klasteriranja. Aglomerativno klasteriranje, polazi od pojedinih objekata koji se povezuju u sve veće klastere, dok se kod divizijskog klasteriranja polazi od svih objekata udruženih u jedan klaster koji se zatim dijele do pojedinih objekata. Klasteri se formiraju pomoću mjere sličnosti između svakog novog elementa i svih ostalih već

prije određenih klastera. Na grafičkom prikazu 12. nalaze se obje metode hijerarhijskog klasteriranja. Stablo kojim se prikazuje raspored klastera nastalih hijerarhijskim klasteriranjem naziva se dendrogram i grafičkim prikazom 11. dan je kratki prikaz grananja stabla. Sličnost između dva objekta u dendrogramu je reprezentirana visinom odnosno duljinom najnižeg unutarnjeg čvora kojeg dijele. Ako je dendrogram u koordinatnom sustavu, na osi ordinata vidi se visina na kojoj se pojedini klasteri spajaju, dok su na osi apscisa raspoređeni objekti. U dendrogramu se netipične vrijednosti (engl. outlier) tj. podaci jako različiti od ostalih, uočavaju kao izolirana grana (Ungaro T., 2016, str. 2-12). Na grafičkom prikazu 11. outlier je predstavljen listom koji je s korijenom povezan vanjskom granom. Outlier predstavlja zaseban klaster, dok unutarnji čvor povezan s unutarnjom granom koja je povezana s korijenom sačinjava drugi klaster koji se sastoji od dva objekta prikazana crnim točkama.

Grafički prikaz 11.

Dendrogram

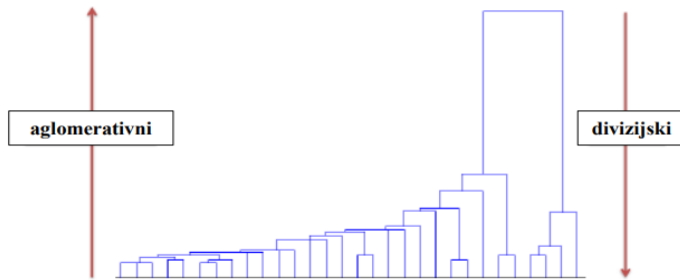


Izvor: Ungaro T. (2016): Klasterska analiza [Internet], raspoloživo na: <https://core.ac.uk/download/pdf/197898517.pdf>

Za razliku od divizijskog klasteriranja, karakteristika aglomerativnih hijerarhijskih algoritama je što gledaju svaki objekt posebno te između svaka dva traže najbolji par koji će spojiti u klaster. To ponavljaju dok svi klasteri nisu spojeni. Divizijski hijerarhijski pristup se manje primjenjuje od aglomerativnog pristupa, a karakterizira ga da se polazi od svih objekata udruženih u jedan klaster koji se zatim dijele do pojedinih objekata. Divizijski hijerarhijski pristup se manje primjenjuje od aglomerativnog pristupa, a karakterizira ga da se polazi od svih objekata udruženih u jedan klaster koji se zatim dijele do pojedinih objekata.

Grafički prikaz 12.

Hijerarhijski pristup



Izvor: Terzić D. (2020): Metode grupiranja podataka, Sveučilište u Zagrebu, Fakultet organizacije i informatike, Varaždin [Internet], raspoloživo na: <https://repositorij.foi.unizg.hr/islandora/object/foi%3A6298/datastream/PDF/view>

Opći oblik hijerarhijskog algoritma čine sljedeći koraci:

1. Inicijalni klasteri su opservacije
2. Izračunavaju se udaljenosti između klastera
3. Spajaju se dva najbliža klastera koja se zamjenjuju s jednim klasterom
4. Ponovno se započinje s drugim korakom dok god se svi klasteri ne svedu na samo jedan klaster koji sadrži sve opservacije.

Stablo se može smanjiti ili povećati kako bi se dobio manji ili veći broj klastera. Broj klastera može odabrati statističar kako bi optimizirao određene kriterije statističke kvalitete (Tuffery, S.,2011, str. 253-254).

3.3.4 Stabla odlučivanja

Stablo odlučivanja je metoda rudarenja podataka koja se zasniva na pretpostavci da određena situacija može rezultirati određenim brojem ishoda sa određenom vjerojatnošću za svaki od mogućih ishoda. Kao što stablo polazi od korijena, pa se grana na pojedine grane te završava listovima tako i stablo odlučivanja pomoću grana predstavlja pravce odlučivanja, a krajnji rezultat predstavlja listove (Perošević M., 2020, str.12) . To je grafički prikaz procesa odlučivanja koji se sastoji od niza slijedno povezanih odluka (Šučur J., 2021, str.4).

Jedna od prednosti stabla odlučivanja je mogućnost rada s kategorijskim i kontinuiranim ulaznim vrijednostima te mogućnost kombinacije sa ostalim metodama rudarenja podataka. Ova metoda spada u

prediktivnu metodu kojom se dobiva grafički prikaz (u obliku stabla) utjecaja ulaznih varijabli na izlaznu varijablu, u obliku klase ili kategorija.

Svaki čvor u stablu predstavlja jednu ulaznu varijablu, na čijim su rubovima označena „djeca-čvorovi“ za svaku moguću vrijednost neke ulazne varijable. Svaki list u stablu predstavlja vrijednost ciljne (izlazne) varijable ako su dane vrijednosti ulaznih varijabli predstavljene putom od korijena stabla do tog lista. Stablo se dobiva „učanjem“ na podacima, na način da se vrši grananje izvornog skupa podataka u podskupove na temelju testiranja vrijednosti varijabli. Proces se ponavlja na svakom izvedenom podskupu na rekurzivni način. Rekurzija je završena kada podskup određenog čvora ima sve iste vrijednosti izlazne varijable, ili kada daljnje grananje više ne pridonosi poboljšanju rezultata (Zekić-Sušac M.&Frajman-Jakšić A.&Drvenkar N.,2009, str. 317-318). Stablo odlučivanja se u pravilu sastoji od niza povezanih odluka, i svaka je ovisna o svojoj prethodnoj.

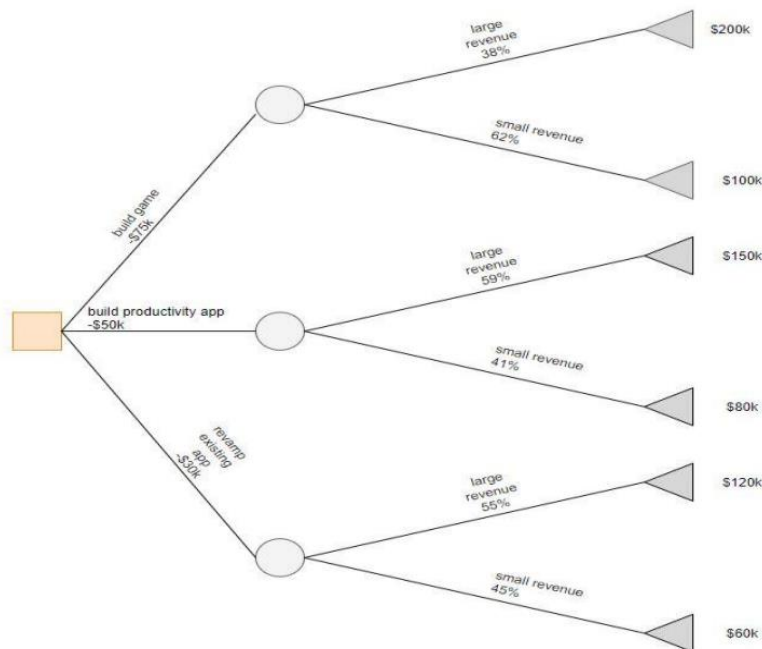
Grafičkim prikazom 13. prikazano je stablo odlučivanja. Svako stablo odlučivanja sastoji se od tri vrste čvorova i dvije vrste grana pa je tako na grafičkom prikazu 13. kvadratom prikazan čvor odluke, krugom - čvorovi mogućih posljedica, a trokutom - završni čvorovi. Svi dijelovi stabla su povezani granama, pri čemu se razlikuju grane alternativnih aktivnosti i grane mogućih posljedičnih stanja. Grane alternativnih aktivnosti izlaze iz čvora odluke (kvadrata), dok grane posljedica izlaze iz čvora posljedica (krugova). Tako, elemente stabla odlučivanja čine:

- *Čvor odluke*, koji je prikazan u obliku kvadrata, i u tom čvoru donositelj odluke odabire za koju će se od opcija odlučiti. Kvadrat prikazan na slici grana se na nekoliko mogućnosti između kojih se može odabrati. Isto, u stablu odlučivanja može biti više čvorova odluke - kvadrata, a ne samo jedan. U tome slučaju, prvi čvor se naziva početni čvor i predstavlja glavnu odluku. Iz ovog čvora izlaze grane alternativnih aktivnosti.
- *Čvorovi mogućih posljedica*, koji su prikazani u obliku krugova, predstavljaju moguće ishode svake od mogućnosti. U čvorove mogućih posljedica (krugovi) ulaze grane alternativnih aktivnosti, a iz njih izlaze grane mogućih posljedičnih stanja.
- *Završni čvor*, koji je prikazan u obliku trokuta i u njima završava svaka od mogućih alternativa. To je obavezan element stabla odlučivanja za svaku navedenu alternativu. U njih ulaze grane mogućih posljedičnih stanja.

- *Grane alternativnih aktivnosti* su grane koje izlaze iz čvora odluke (kvadrata). Iz svakog čvora odluke moraju izlaziti barem dvije grane alternativnih aktivnosti koje imaju određenu vrijednost. Grane alternativnih aktivnosti rezultiraju čvorom mogućih posljedica (krugovima), ali u složenijim stablima odlučivanja mogu rezultirati čvorom mogućih posljedica (krugovima), čvorom odluke (kvadratima), ili pak završnim čvorom (trokutima).
- *Grane mogućih posljedičnih stanja* su grane koje izlaze iz čvora mogućih posljedica (krugova) te prikazuju posljedice koje se mogu dogoditi. Na grani mogućih posljedičnih stanja nalazi se vjerojatnost da će se svaka pripadna posljedica dogoditi. Isto, mogu završavati u bilo kojem od čvorova.

Grafički prikaz 13.

Stablo odlučivanja



Izvor: Kiđemet M.(2018): Primjena metode stablo odlučivanja, Sveučilište u Zagrebu, Fakultet organizacije i informatike, Varaždin [Internet], raspoloživo na: <https://core.ac.uk/download/pdf/197829835.pdf>

4. ANALIZA I RUDARENJE PODATAKA NA PRIMJERU JAVNE BAZE PODATAKA O EU PROJEKTIMA

4.1 Opis skupa podataka

Kao što je već spomenuto, setovi podataka financiranih projekata OP “Konkurentnost i kohezija 2014.-2020.” i OP “Učinkoviti ljudski potencijali 2014.-2020.” spadaju u sekundarne podatke, prethodno prikupljene koji su javno dostupni na stranici europskih strukturnih i investicijskih fondova (<https://strukturnifondovi.hr/>) te su namijenjeni osiguranju transparentnosti i pružanju informacija javnosti o projektima i korisnicima potpore iz fondova EU. Navedeni setovi podataka sadrže podatke o financiranim projektima od 2014. godine do 2020. godine. Države članice odnosno upravljačka tijela dužna su objaviti popis projekata po operativnim programima sa svim informacijama koje su propisane regulativom Europske unije te ga redovito održavati (Strukturni fondovi, 2023, parag. 1). S obzirom da popis projekata sadrži određeni broj kategorijskih varijabli, a tehnike rudarenje podataka su ograničene u rukovođenju s kvalitativnim varijablama, potrebne varijable se kodiraju te time svode na numeričke varijable kako bi bile uključene u analizu podataka. Iz seta podataka uklanjaju se irelevantne varijable, odnosno set podataka se uređuje kako bi bio spreman za daljnju obradu.

4.1.1. Sadržaj skupa podataka

Odabrani setovi podataka su spojene Excel datoteke seta OP “Konkurentnost i kohezija 2014.-2020.” odnosno datoteka “OPKK.xls” i seta OP “Učinkoviti ljudski potencijali 2014.-2020.” odnosno datoteka “OPULJP.xls” . Sastoji se od 29 stupaca i 9615 redaka. Isto, set podataka sadrži izvedene varijable poput godine trajanja projekta, mjeseci trajanja projekta, bespovratnih sredstava EU i slično te uvedenih varijabli poput prioritetne osi, NUTS regije, fonda iz kojeg se financiraju projekti i slično. Navedeni set podataka sadrži popis projekata financiranih od strane EU u razdoblju od 2014.-2020. godine. Dakle, svaki stupac sadrži informaciju o projektu prema nazivu korisnika, nazivu projekta, ukupnoj vrijednosti projekta, EU stopi financiranja, prioritetnoj osi, lokaciji provedbe projekta, NUTS regiji kojoj projekt pripada, bespovratnim sredstva EU i slično.

4.1.2. Popis skupa podataka

Kao što je prethodno navedeno, skup podataka se sastoji od 9615 obilježja o EU projektima provedenih u Hrvatskoj u razdoblju od 2014. do 2020. godine koji su raspoređeni u 29 varijabli. Detaljan popis skupa podataka nalazi se u sljedećoj tablici.

Tablica 2. Popis skupa podataka

<i>Redni broj</i>	<i>Naziv varijable</i>	<i>Opis varijable</i>	<i>Tip varijable</i>
1.	<i>ID projekta</i>	<i>Jedinstveni identifikator svakog pojedinog projekta</i>	<i>Nominalna</i>
2.	<i>Prioritetna os</i>	<i>Tematski ciljevi strategije EU</i>	<i>Nominalna</i>
3.	<i>Naziv operacije</i>	<i>Naziv svakog projekta</i>	<i>Nominalna</i>
4.	<i>Naziv korisnika</i>	<i>Naziv svakog korisnika financiranog projekta</i>	<i>Nominalna</i>
5.	<i>Tip korisnika</i>	<i>Korisnici mogu biti mala i srednja poduzeća, neprofitne organizacije, javna tijela, istraživači i ostali.</i>	<i>Nominalna</i>
6.	<i>Datum početka operacije</i>	<i>Datum kada je započeo projekt</i>	<i>Vremenska</i>
7.	<i>Datum završetka operacije</i>	<i>Datum kada je projekt završen</i>	<i>Vremenska</i>
8.	<i>Status provedbe projektnih aktivnosti</i>	<i>Trenutni status projekta koji može biti u tijeku ili završen.</i>	<i>Nominalna</i>
9.	<i>Lokacija provedbe projekta</i>	<i>Lokacija provedbe projekta - županija</i>	<i>Nominalna</i>
10.	<i>Eu stopa sufinanciranja u %</i>	<i>Udio EU sufinanciranja od ukupne ugovorene vrijednosti bespovratnih sredstava</i>	<i>Numerička</i>
11.	<i>Stopa nacionalnog, javnog i privatnog sufinanciranja u %</i>	<i>Udio nacionalnog financiranja projekata od ukupne ugovorene vrijednosti bespovratnih sredstava.</i>	<i>Numerička</i>
12.	<i>Bespovratna sredstva - Nacionalni dio + javni i privatni doprinos korisnika - HRK</i>	<i>Količina nacionalnih novčanih sredstava dodijeljenih projektu</i>	<i>Kontinuirana</i>
13.	<i>Bespovratna sredstva - EU dio - HRK</i>	<i>Količina novčanih sredstava EU dodijeljenih pojedinom projektu</i>	<i>Numerička</i>

14.	<i>UKUPNI PRIHVATLJIVI IZDACI = Ukupna ugovorena vrijednost projekta HRK</i>	<i>Bespovratna sredstva ukupno + doprinos korisnika = Ukupni prihvatljivi troškovi = Ukupna ugovorena vrijednost projekta HRK</i>	<i>Numerička</i>
15.	<i>Posredničko tijelo razine 1</i>	<i>Nacionalno tijelo koje obavlja delegirane funkcije vezane za odabir projekata za financiranje. Primjerice, u suradnji s Posredničkim tijelom razine 2 priprema smjernice za prijavitelje za ocjenjivanje i odgovorno je za odabir projekata za financiranje.</i>	<i>Nominalna</i>
16.	<i>Posredničko tijelo razine 2</i>	<i>Nacionalno tijelo koje obavlja delegirane funkcije koje se odnose na provjeru jesu li financirani proizvodi i usluge isporučeni, jesu li prikazani izdaci korisnika za projekt stvarno nastali te udovoljavaju li nacionalnim i pravilima EU tijekom cijelog razdoblja provedbe projekta.</i>	<i>Nominalna</i>
17.	<i>Fond</i>	<i>EU fondovi iz kojih se financiraju projekti. Razlikuju se: KF, EFRR, ESF i slično.</i>	<i>Nominalna</i>
18.	<i>Operativni program</i>	<i>Operativni programi kohezijske politike kojima pripadaju projekti mogu biti OPKK i OPULIP.</i>	<i>Nominalna</i>
19.	<i>NUTS regija</i>	<i>NUTS regija kojoj pripada projekt koja može biti Jadranska Hrvatska ili Kontinentalna Hrvatska.</i>	<i>Nominalna</i>
20.	<i>Sažetak operacije</i>	<i>Detaljan tekstualni opis svakog pojedinog projekta.</i>	<i>Nominalna</i>
21.	<i>Kategorija intervencije</i>	<i>Prema kategoriji intervencije projekti su svrstani u ostvarenje određenog prioritetnog cilja, npr. kategorija intervencije 024-Željeznica (osnovne transeuropske prometne mreže/TEN-T)</i>	<i>Nominalna</i>
22.	<i>Mjesec početka operacije</i>	<i>Mjesec u kojem je započeo projekt.</i>	<i>Nominalna</i>
23.	<i>Mjesec završetka operacije</i>	<i>Mjesec u kojem završava odvijanje projekta.</i>	<i>Nominalna</i>

24.	<i>Godina početka operacije</i>	<i>Godina u kojoj je započeo projekt.</i>	<i>Nominalna</i>
25.	<i>Godina završetka operacije</i>	<i>Godina u kojoj završava projekt.</i>	<i>Nominalna</i>
26.	<i>Mjeseci trajanja operacije</i>	<i>Dužina trajanja projekta izražena u mjesecima</i>	<i>Numerička</i>
27.	<i>Godine trajanja operacije</i>	<i>Dužina trajanja projekta izražena u godinama</i>	<i>Numerička</i>
28.	<i>Veličina projekta</i>	<i>Kategoriziranje projekta prema ukupnoj ugovorenoj vrijednosti projekta razlikuje se: veliki, mali i slično</i>	<i>Ordinalna</i>
29.	<i>Interv.vrijednost projekta</i>	<i>Kategoriziranje ukupne vrijednosti projekta u intervale od 0-50 milijuna kuna, 50-100 milijuna kuna, i slično</i>	<i>Ordinalna</i>

Izvor: Prikaz autora

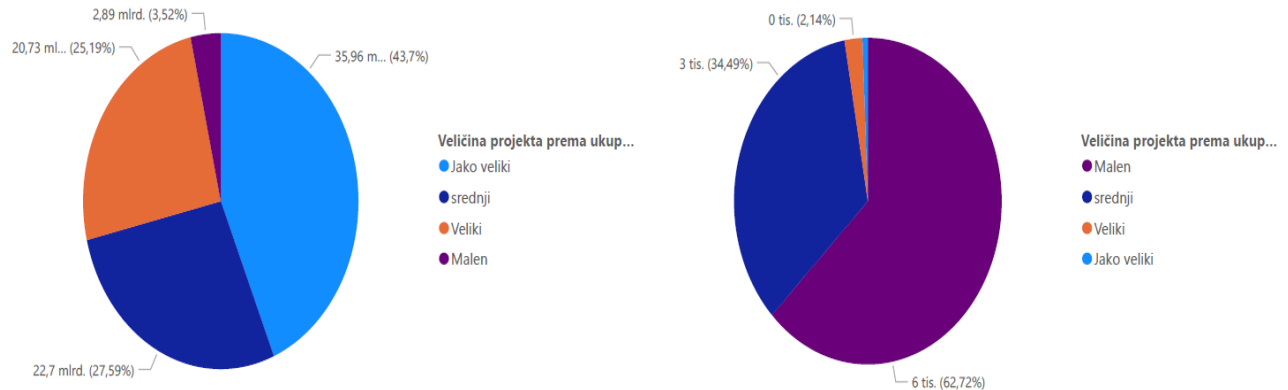
4.2 Vizualizacija podataka

Kohezijskom politikom promiče se i podržava ukupan ravnomjeran razvoj država članica i njihovih regija. To je instrument financijske solidarnosti jer koncentrira sredstva prema područjima i sektorima kojima su najpotrebnija.

Zbog postojanja velikih teritorijalnih i demografskih razlika koje mogu predstavljati prepreku integraciji i razvoju EU, uspostavljeni su mehanizmi solidarnosti u obliku fondova. Tako, postoje tri kohezijska aspekta EU-a koji zajedno primaju potporu u okviru kohezijske politike i strukturnih fondova, a to su: socijalna kohezija, ekonomska kohezija i teritorijalna kohezija. Sva tri aspekta su obuhvaćena dvama setovima podataka OPKK i OPULJP u kojima se nalaze podaci o financiranim projektima kojima se ostvaruju ciljevi kohezijske politike EU, a koji su prikazani pomoću Microsoft Power BI alata za vizualizaciju podataka. Tako, na grafičkom prikazu 14. prikazan je odnos između broja projekata i ukupnog iznosa bespovratnih sredstava EU prema veličini projekta.

Grafički prikaz 14.

Udio iznosa bespovratnih sredstava EU i broja projekata prema veličini projekata



Veličina projekta prema ukupno ugovorenoj vrijednosti projekta	Zbroj: Bespovratna sredstva EU - HRK
Jako veliki	35.962.202.733,70
Malen	2.894.936.208,42
srednji	22.700.646.258,60
Veliki	20.732.811.769,45
Ukupno	82.290.596.970,17

Veličina projekta prema ukupno ugovorenoj vrijednosti projekta	Zbroj: Broj projekata
Jako veliki	62
Malen	6031
srednji	3316
Veliki	206
Ukupno	9615

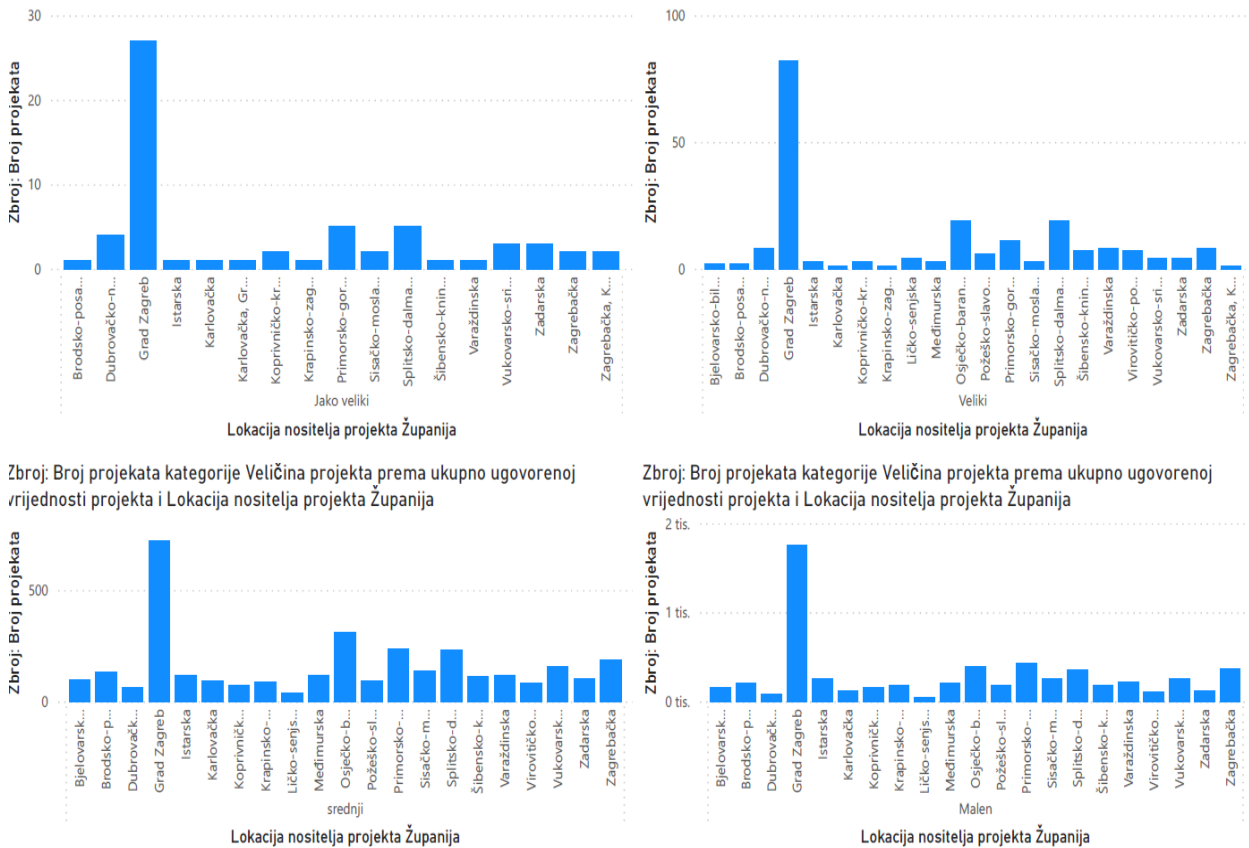
Izvor: Prikaz autora

Na grafičkom prikazu 14. najveći udio ukupnih bespovratnih sredstava EU, čak 43,7 posto, izdvojeno je za financiranje jako velikih projekata što se odnosi na samo 62 projekta od ukupno 9615 projekata. Najmanji udio ukupnih bespovratnih sredstava EU izdvojen je za financiranje malih projekata (3,52 posto) odnosno na financiranje oko šest tisuća projekata. U jako velike projekte spadaju projekti čija je ukupna ugovorena vrijednost projekta veća od 300 milijuna kuna, a u male projekte spadaju projekti do 1,5 milijuna kuna.

Ukupna ugovorena vrijednost srednjih projekata kreće se od 1,5 milijuna kuna do 50 milijuna kuna, dok se za velike projekte ukupno ugovorena vrijednost projekta kreće od 50 do 300 milijuna kuna. Udio bespovratnih sredstava EU za srednje projekte iznosi 27,59 posto, a za velike projekte 25,19 posto od ukupnog postotka bespovratnih sredstava EU dodijeljenih projektima. Prema navedenim informacijama ostvareno je više od tri tisuće projekata srednje veličine i oko dvjesto velikih projekata u vremenskom periodu od 2014.godine do 2020.godine koji su namijenjeni provedbi kohezijske politike.

Grafički prikaz 15.

Raspodjela projekata prema veličini i broju projekata po županiji



Izvor: Prikaz autora

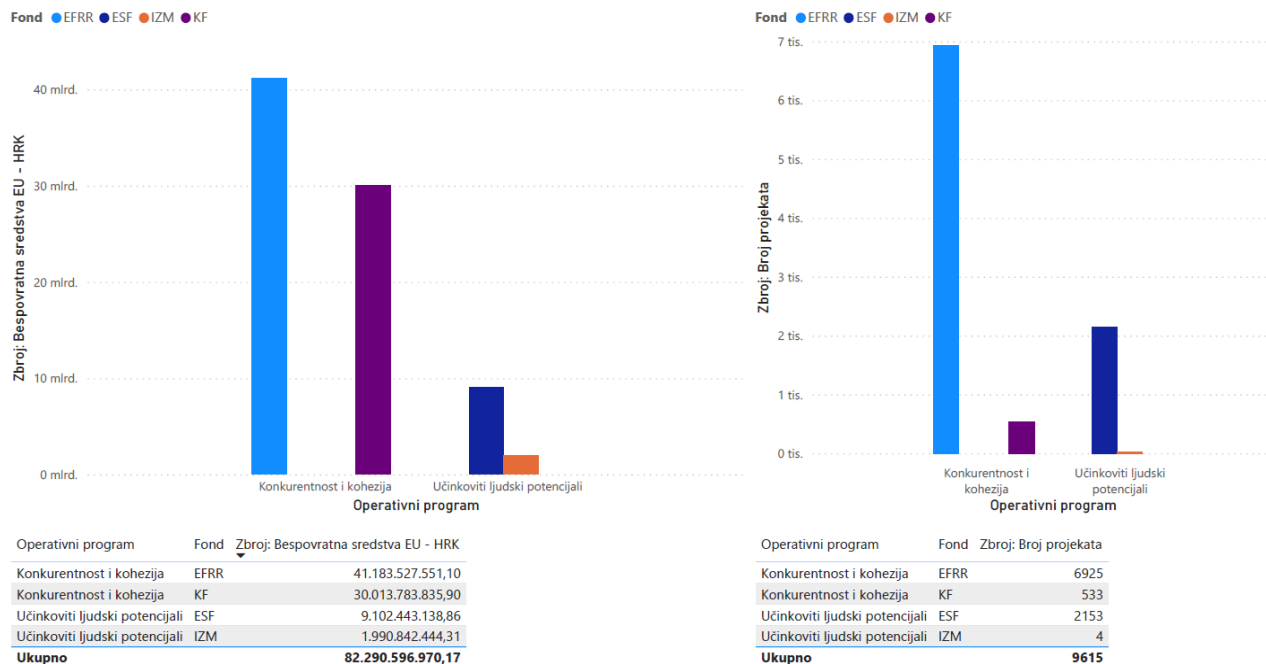
Tako, na slici 15. nalazi se raspodjela projekata prema veličini i broju projekata po županiji. Na slici se može vidjeti da Grad Zagreb prednjači u apsorpciji bespovratnih EU sredstava i po broju projekata. Najveći broj jako velikih projekata nalazi se u Gradu Zagrebu, njih 27. Može se naglasiti da su se još 4 projekta odvijala na lokacijama dviju županija od kojih se jedna odnosi na Grad Zagreb pa se može reći da se radi o 28 jako velika projekta.

Niti približno blizu po broju projekata i iznosu financijskih sredstava, Grad Zagreb slijede županije Primorsko-goranska i Splitsko-dalmatinska sa 5 jako velikih projekata te Dubrovačko-neretvanska sa 4 jako velika projekta. Isto, 82 velika projekta odvijaju se na području Grada Zagreba kojeg slijede Osječko-baranjska i Splitsko-dalmatinska sa 19 velikih projekata te Primorsko-goranska sa 11 velikih projekata. Najveći broj

projekata srednje veličine, onih čija se ukupna ugovorena vrijednost kreće od 1,5 milijuna kuna do 50 milijuna kuna, nalazi se u Gradu Zagrebu (724 projekta), a slijede ga: Osječko-baranjska županija sa 309 projekata, Primorsko-goranska sa 237 projekata, Splitsko-dalmatinska županija sa 233 projekta i ostali. Najveći broj malih projekata, onih čije je financiranje do 1,5 milijuna kuna, isto se odnosi na Grad Zagreb sa 1757 malih projekata što je razumljivo s obzirom da se samo u Gradu Zagrebu nalazi oko 800 tisuća stanovnika i da je grad Zagreb monocentrično središte Hrvatske. Po broju projekata i iznosu bespovratnih sredstava slijede ga Primorsko-goranska županija (429 projekata), Osječko-baranjska županija (392 projekta) te Splitsko-dalmatinska (358 projekata). S obzirom da se bespovratna sredstva EU isplaćuju iz fondova EU, zanimljiv je prikaz sveukupno isplaćenih bespovratnih sredstava iz fondova, a time i operativnih programa.

Grafički prikaz 16.

Operativni program i vrsta fonda prema broju projekata i vrijednosti bespovratnih sredstava EU



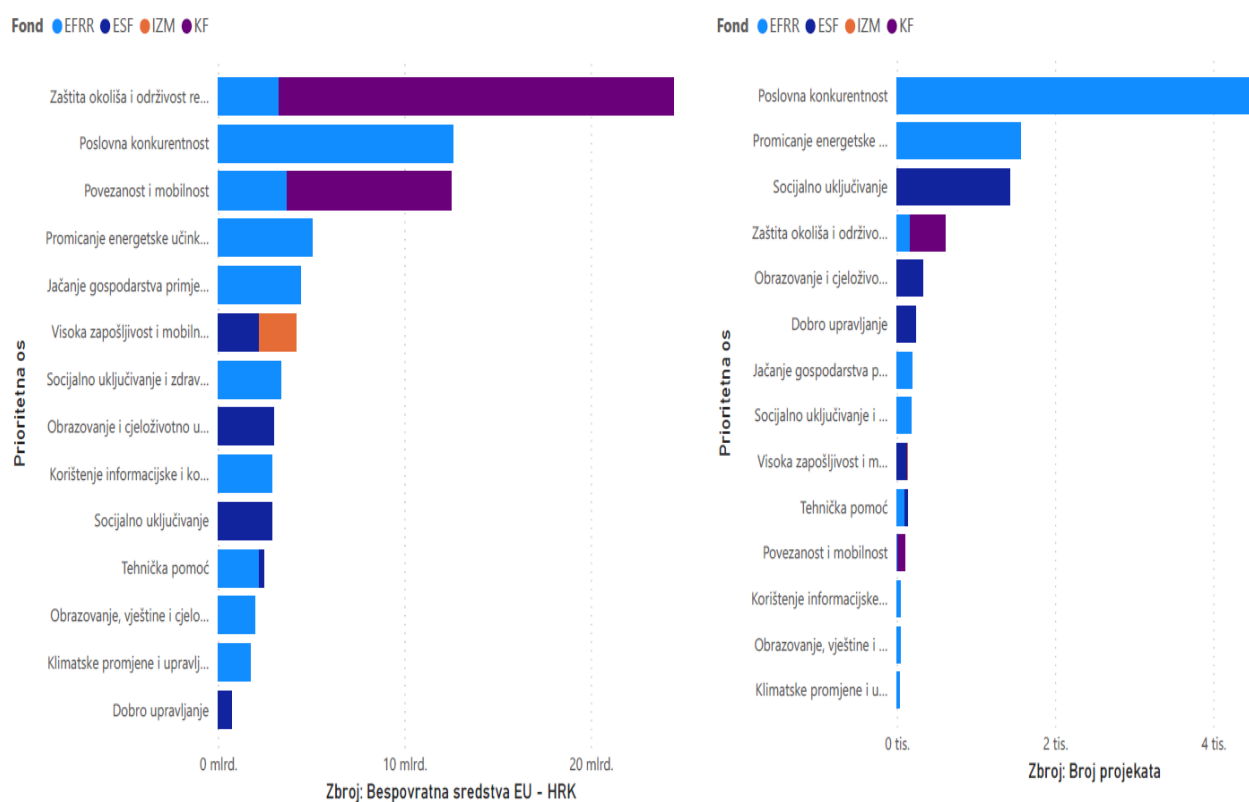
Izvor: Prikaz autora

Na slici 16. vidljivo je da je iz OP „Konkurentnost i kohezija“ isplaćeno najviše bespovratnih sredstava EU i to iz fondova EFRR i KF što je razumljivo s obzirom da obuhvaćaju prioritetne osi poput poslovne konkurentnosti, povezanosti i mobilnosti te zaštite okoliša i održivosti resursa. Tako, iz EFRR fonda

financirano je 6925 projekata, dok su iz KF financirana 533 projekta. Financiranje projekata iz OP „Učinkoviti ljudski potencijali“ odvija se putem ESF i IZM fonda pa su iz ESF-a financirana 2153 projekta, a iz IZM-a 4 projekta. S obzirom da se iz OP financiraju projekti preko dostupnih fondova, zanimljivo je vidjeti koliko projekata i za ostvarenje kojih prioritetnih ciljeva im je dodijeljeno najviše financijskih sredstava.

Grafički prikaz 17.

Vrijednost bespovratnih EU sredstava i broj projekata prema fondu i prioritetnoj osi



Izvor: Prikaz autora

Tako, prema podacima prikazanim na slici 17. najveći iznos bespovratnih EU sredstava je iz Kohezijskog fonda dodijeljeno projektima za ostvarenje prioritetnih ciljeva Zaštita okoliša i održivost resursa te za prioritetni cilj Povezanost i mobilnost. U prioritetni cilj Zaštita okoliša i održivost resursa spadaju ulaganja u unaprjeđenje procjene kakvoće zraka nadogradnjom i modernizacijom nacionalnih i lokalnih mreža za trajno praćenje kakvoće zraka kako bi se uspostavio odgovarajući sustav za nadzor i upravljanje onečišćenjem zraka odnosno žele se uvesti klimatski osjetljive mjere protiv onečišćenja zraka. S druge

strane, prioritetni cilj Povezanost i mobilnost karakterizira ulaganje u prometnu infrastrukturu potrebnu za suvremenu i dobro povezanu europsku ekonomiju koja podrazumijeva razvoj projekata koji olakšavaju kretanje robe i ljudi odnosno cilj je kojim se jača teritorijalna kohezija između gradova i izoliranih područja.

Iz EFRR fonda financirani su projekti kojima se nastoje ostvariti svi preostali prioritetni ciljevi osim prioritetnih ciljeva OP „Učinkoviti ljudski potencijali“. Prema tome, iz EFRR-a se u najvećem iznosu bespovratnih sredstava EU nastoje ostvariti prioritetni ciljevi Poslovna konkurentnost, Promicanje energetske učinkovitosti i obnovljivih izvora energije, Jačanje gospodarstva primjenom istraživanja i inovacija i ostali. Prioritetnim ciljem Poslovna konkurentnost promiče se poticanje poduzetništva putem poslovnih inkubatora, osigurava se bolji pristup financiranju za MSP-ove te se stvara povoljno okruženje za osnivanje i razvoj poduzeća i slično. Glavni prioriteti financiranja cilja Promicanje energetske učinkovitosti i obnovljivih izvora energija je usmjerenost na promicanje mjera obnovljivih izvora energije i energetske učinkovitosti te smanjenju emisija stakleničkih plinova.

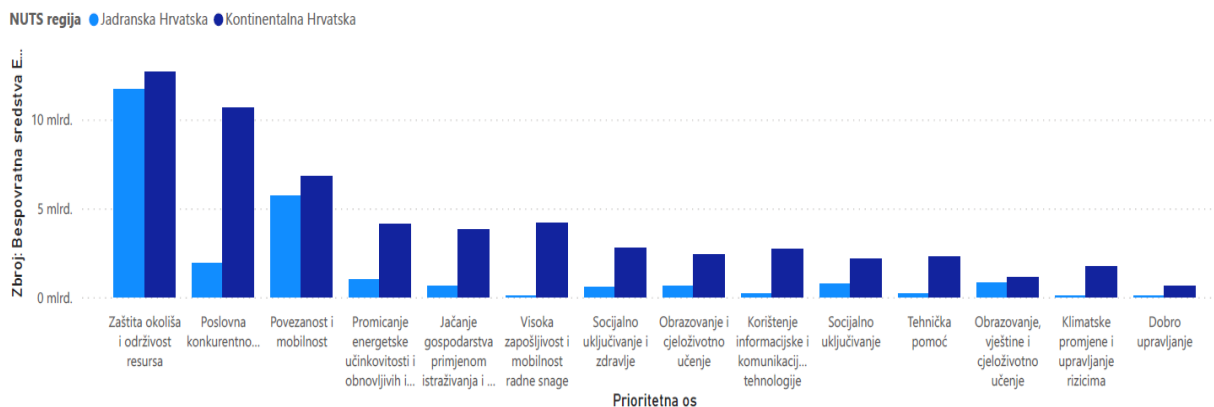
Najmanji iznos bespovratnih sredstava EU dodijeljen je financiranju projekata iz OPULJP-a za ostvarenje prioritetnog cilja Dobro upravljanje. Cilj tog prioriteta je ulaganje u institucionalne kapacitete i učinkovitost javnih uprava i usluga, povećanje djelotvornosti javne uprave i unapređenje kapaciteta te funkcioniranja pravosuđa kroz ulaganja u ljudske resurse.

Također, najveći broj projekata, njih 4580, financirani su iz EFRR-a za ostvarenje prioritetnog cilja Poslovna konkurentnost. Po broju ostvarenih projekata slijede ga prioritetni ciljevi Promicanje energetske učinkovitosti i obnovljivih izvora energije (EFRR), Socijalno uključivanje (ESF), Zaštita okoliša i održivost resursa (EFRR i KF) i ostali. Najmanji broj projekata, njih 38, je financirano iz prioritetne osi Klimatske promjene i upravljanje rizicima (EFRR). Glavni prioriteti cilja Socijalno uključivanje odnose se na povećanje zapošljivosti, borbi protiv siromaštva i socijalne isključenosti, poboljšanju pristupa visokokvalitetnim uslugama zdravstvene skrbi i socijalnim uslugama, promicanju društvenog poduzetništva i slično.

Kao što se vidi na grafičkom prikazu 18., zanimljiv je i uvid u isplaćena bespovratna sredstva EU prema prioritetnoj osi i NUTS regiji jer se jasno vidi da NUTS regija Kontinentalna Hrvatska prednjači naspram Jadranske Hrvatske po iznosu bespovratnih EU sredstava po svakoj prioritetnoj osi.

Grafički prikaz 18.

Vrijednost bespovratnih sredstava EU prema NUTS regiji i prioritetnoj osi



Prioritetna os	Jadranska Hrvatska	Kontinentalna Hrvatska	Ukupno
Zaštita okoliša i održivost resursa	11.718.388.770,29	12.706.572.997,28	24.424.961.767,57
Poslovna konkurentnost	1.931.587.706,16	10.667.213.430,02	12.598.801.136,18
Povezanost i mobilnost	5.734.447.936,86	6.795.520.336,22	12.529.968.273,08
Promicanje energetske učinkovitosti i obnovljivih izvora energije	967.125.798,84	4.084.560.811,14	5.051.686.609,98
Jačanje gospodarstva primjenom istraživanja i inovacija	616.414.551,68	3.804.261.017,49	4.420.675.569,17
Visoka zapošljivost i mobilnost radne snage	30.876.924,11	4.177.405.773,55	4.208.282.697,66
Socijalno uključivanje i zdravlje	594.133.034,82	2.756.831.246,06	3.350.964.280,89
Obrazovanje i cjeloživotno učenje	623.826.020,22	2.374.805.201,99	2.998.631.222,21
Korištenje informacijske i komunikacijske tehnologije	188.687.155,53	2.720.082.757,11	2.908.769.912,64
Socijalno uključivanje	735.815.853,48	2.148.815.630,39	2.884.631.483,88
Tehnička pomoć	188.227.746,92	2.287.868.685,12	2.476.096.432,04
Obrazovanje, vještine i cjeloživotno učenje	837.931.447,45	1.122.013.728,34	1.959.945.175,79
Klimatske promjene i upravljanje rizicima	18.974.964,21	1.732.374.848,21	1.751.349.812,43
Dobro upravljanje	68.018.718,16	657.813.878,49	725.832.596,66
Ukupno	24.254.456.628,74	58.036.140.341,43	82.290.596.970,17

Izvor: Prikaz autora

Naime, kod obje NUTS regije najveći iznos bespovratnih EU sredstava odnosi se na ostvarenje prioritetne osi Zaštita okoliša i održivost resursa kojom se žele uvesti klimatski osjetljive mjere protiv onečišćenja zraka, a ostale redom navedene prioritetne osi su: Poslovna konkurentnost, Povezanost i mobilnost, Promicanja energetske učinkovitosti i obnovljivih izvora energije, Jačanje gospodarstva primjenom istraživanja i inovacija i ostali. Za razliku od prioritetne osi Zaštita okoliša i održivost resursa po kojoj se NUTS regije prate po dobivenom bespovratnom financiranju (obje NUTS regije su financirane sa više od deset milijardi kuna), za prioritetnu os Poslovna konkurentnost, koja bi trebala poticati MSP-ove te rast i razvoj inovativnosti i poduzetništva, razlika je i više nego uočljiva.

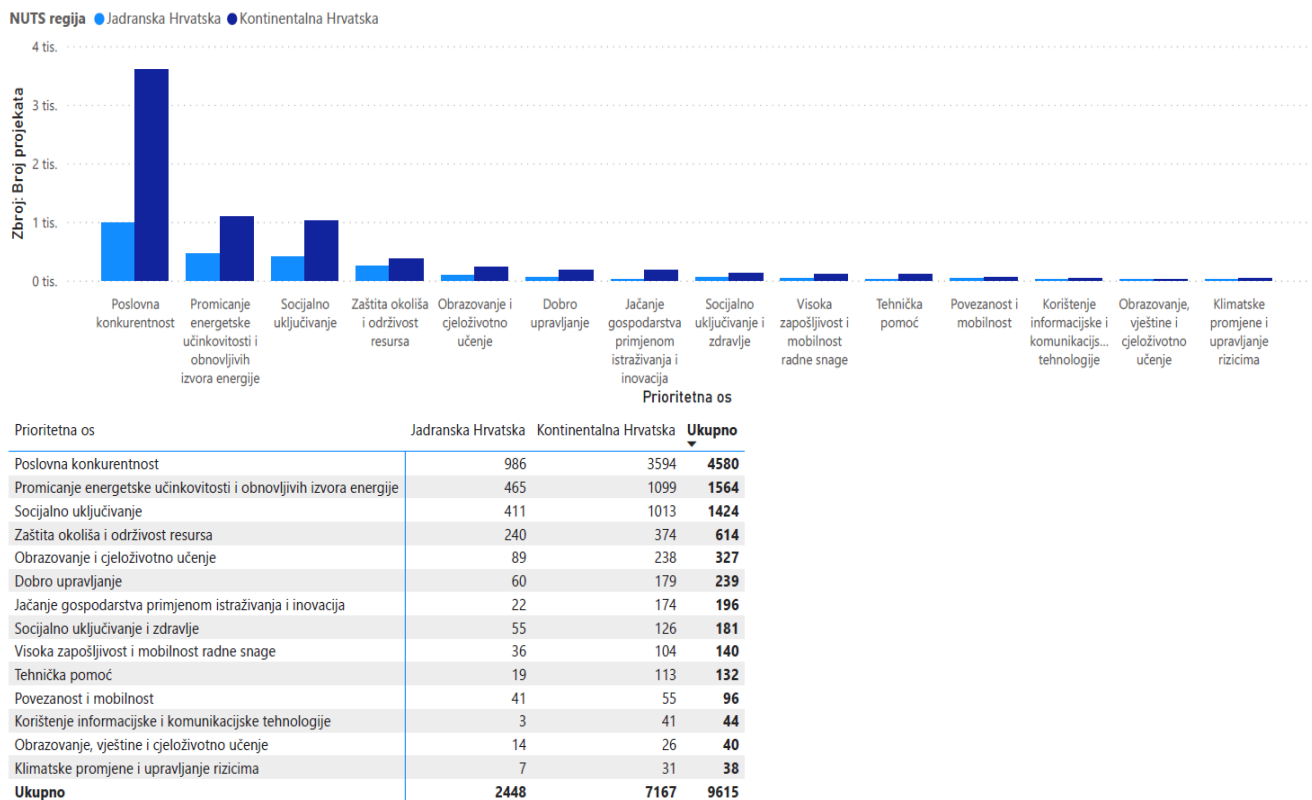
Tako, za Kontinentalnu Hrvatsku taj iznos se kreće oko jedanaest milijardi kuna, dok za Jadransku Hrvatsku iznos bespovratnih EU sredstava je manji od dvije milijarde kuna. Isto, za Prioritetnu os Povezanost i mobilnost NUTS regije se prate po iznosu bespovratnog financiranja (od šest do sedam milijardi kuna), dok za preostale prioritetne osi postoji jako velika razlika u iznosu bespovratnog financiranja između NUTS

regija. Primjerice, za cilj Promicanje energetske učinkovitosti i obnovljivih izvora energije iznos financiranja kreće se oko četiri milijarde kuna za Kontinentalnu Hrvatsku, a manje od milijarde kuna za Jadransku Hrvatsku. Za cilj Jačanje gospodarstva primjenom istraživanja i inovacija iznos financiranja projekata je manje od četiri milijarde kuna za Kontinentalnu Hrvatsku, a više od šesto milijuna kuna za Jadransku Hrvatsku. Za OPULJP-ov prioritetni cilj Visoku zapošljivost i mobilnost radne snage iznos financiranja projekata je viši od četiri milijarde kuna za Kontinentalnu Hrvatsku i nešto manji od šesto milijuna kuna za Jadransku Hrvatsku.

Isto, zanimljiv je grafički prikaz 19. na kojem je prikazan broj projekata po prioritetnoj osi i NUTS regiji gdje je jasno vidljivo da NUTS regija, Kontinentalna Hrvatska, prednjači i po broju projekata naspram Jadranske Hrvatske, što je razumljivo s obzirom da prednjači i po dobivenom iznosu bespovratnih sredstava EU.

Grafički prikaz 19.

Ukupan broj projekata prema prioritetnoj osi i NUTS regiji



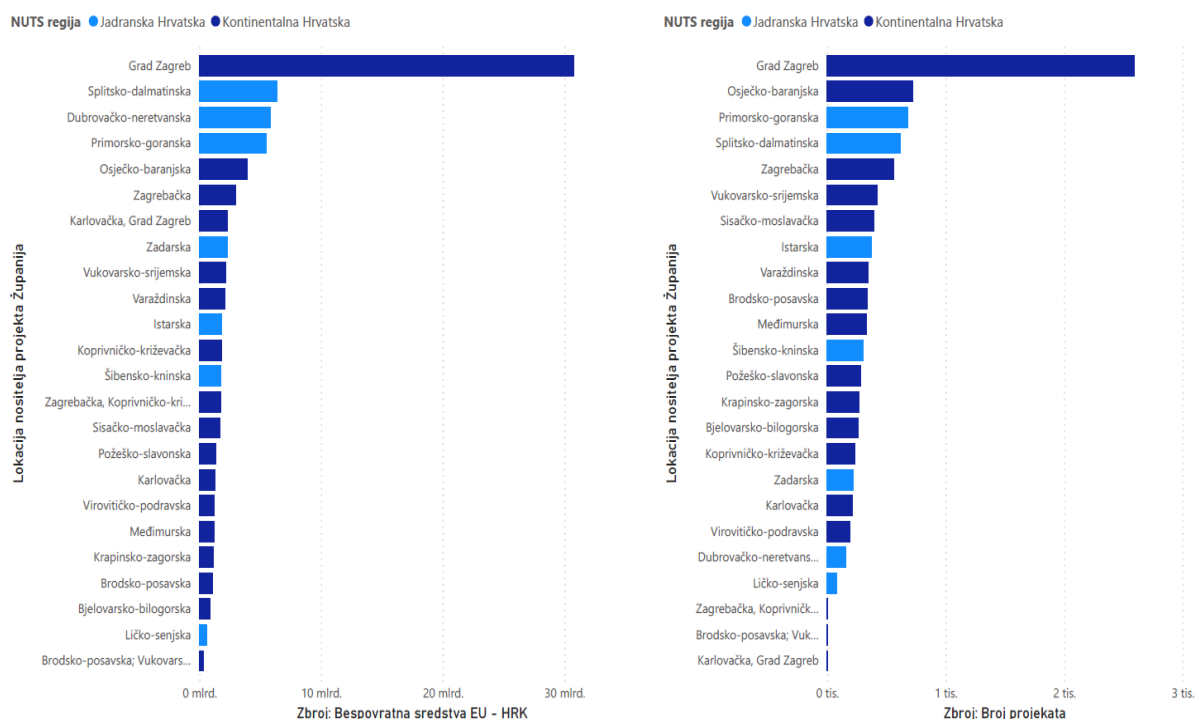
Izvor: Prikaz autora

Tako, za NUTS regiju Kontinentalna Hrvatska najveći broj projekata ostvaren je iz prioritarnog cilja Poslovna konkurentnost (manje od četiri tisuće projekata). U NUTS regiji Jadranska Hrvatska je slična situacija, najviše projekata je ostvareno iz cilja Poslovne konkurentnosti, ali je taj broj projekata dosta manji te je u financirano manje od tisuću projekata. Drugi prioritetni cilj, po broju projekata je Promicanje energetske učinkovitosti i obnovljivih izvora energije sa više od tisuću projekata za Kontinentalnu Hrvatsku i oko petsto projekata za Jadransku Hrvatsku. Najmanji broj projekata se ostvaruje iz Prioritarnog cilja Klimatske promjene i upravljanje rizicima za Jadransku Hrvatsku sa sedam projekata te Kontinentalnu Hrvatsku sa 31 projektom. Po broju projekata prate ga ciljevi Obrazovanje, vještine i cjeloživotno učenje te cilj Korištenje informacijske i komunikacijske tehnologije gdje Kontinentalna Hrvatska ostvaruje 41 projekt, a Jadranska Hrvatska tek tri projekta.

Ukupan iznos bespovratnih sredstava EU i broja projekata prema županiji i NUTS regiji prikazan je grafičkim prikazom 20., na kojem NUTS regija Kontinentalna Hrvatska prednjači po broju projekata i iznosu bespovratnih sredstava posebno u Gradu Zagrebu.

Grafički prikaz 20.

Ukupan iznos bespovratnih sredstava EU i broja projekata prema županiji i NUTS regiji



Izvor: Prikaz autora

Županije Kontinentalne Hrvatske koje su apsorbirale dosta manji iznos bespovratnih sredstava i u kojima je financiran dosta manji broj projekata nego u Gradu Zagrebu, a koje se ističu po korištenju bespovratnih sredstava su redom navedene: Osječko-baranjska, Zagrebačka, Vukovarsko-srijemska, Varaždinska, Koprivničko-križevačka, Sisačko-moslavačka i druge. Županija Kontinentalne Hrvatske sa najmanjim iznosom ostvarenih bespovratnih sredstava je Bjelovarsko-bilogorska. Županije koje se nakon Grada Zagreba izdvajaju po broju projekata su redom navedene: Osječko-baranjska, Zagrebačka, Vukovarsko-srijemska, Sisačko-moslavačka, Varaždinska, Brodsko-Posavska, Međimurska, Krapinsko-zagorska, Bjelovarsko-bilogorska i druge. Županija Kontinentalne Hrvatske koja je ostvarila najmanji broj projekata je Virovitička-podravska.

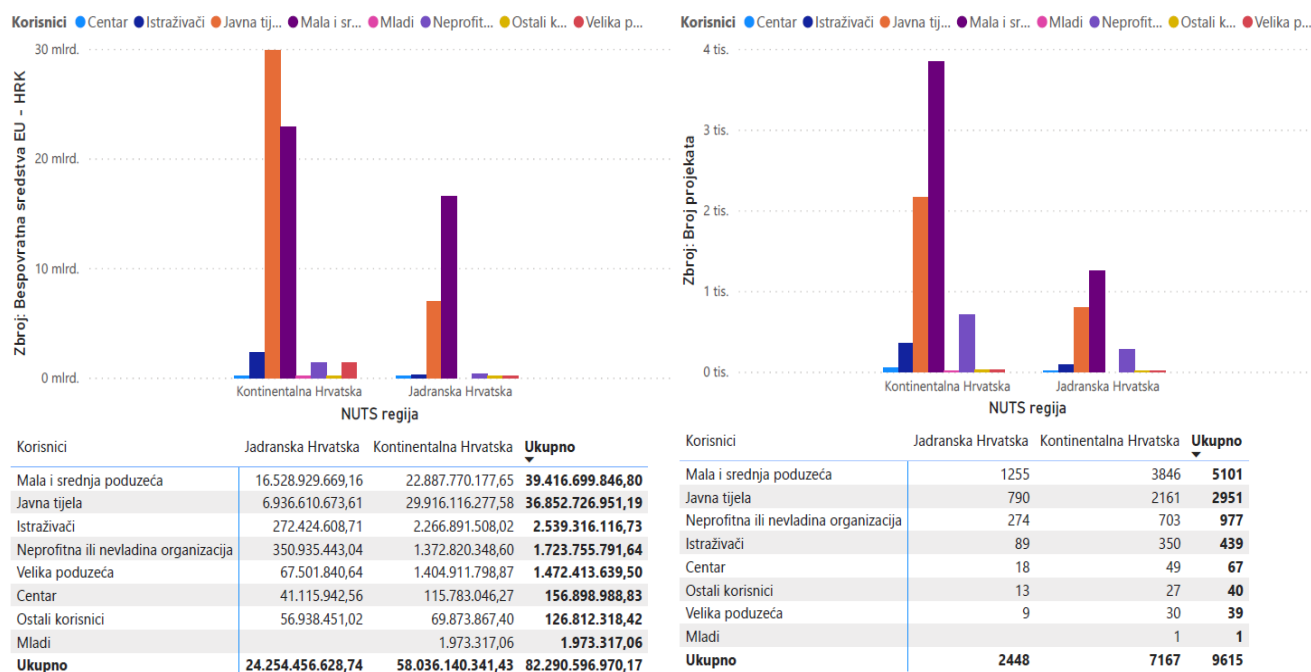
U NUTS regiji Jadranskoj Hrvatskoj najveći iznos bespovratnih sredstava EU je ostvarila Splitsko-dalmatinska županija. Nešto manje bespovratnih sredstava EU ostvarile su Dubrovačko-neretvanska i Primorsko-goranska županija. Nakon njih slijede Zadarska, Istarska pa Šibensko-kninska županija, a najmanje bespovratnih sredstava EU apsorbirala je Ličko-senjska županija. Za Jadransku Hrvatsku, najveći broj projekata kojima se ostvaruju ciljevi kohezijske politike ostvarila je Primorsko-goranska županija sa 682 projekta, a iza koje je sa nešto manjim brojem projekata Splitsko-dalmatinska županija, njih 615. Slijedeće u nizu po broju projekata su: Istarska, Šibensko-kninska, Zadarska, Dubrovačko-neretvanska pa s najmanje ostvarenih projekata Ličko-senjska županija. Uočava se da je Dubrovačko-neretvanska županija druga županija Jadranske Hrvatske po iznosu financiranja, a s dosta manjim brojem projekata, točnije njih 160, što ukazuje na prisutnost većih projekata u toj županiji.

Isto, ističe se pet projekata koji se istovremeno odvijaju u dvije županije, a od kojih se tri odvijaju u Zagrebačkoj i Koprivničko-križevačkoj županiji i odnose se na projekte financirane iz prioritetne osi Povezanosti i mobilnosti iz OP „Konkurentnost i kohezija“. Projekti su financirani iz Kohezijskog fonda, a korisnik je javno tijelo pod nazivom HŽ infrastruktura d.o.o., a projektima se želi rekonstruirati, modernizirati i izgraditi nove željezničke pruge. Projektima se radi na povezivanju osnovne mreže prometne infrastrukture s transeuropskim mrežama i koridorima što je jedan od osnovnih uvjeta za ravnomjeran razvoj zemalja Europske unije. Ovim projektima omogućava se integracija nacionalnih mreža prometnica u jedinstvenu Transeuropsku mrežu prometnica (TEN-T). Četvrti projekt odvija se na području Karlovačke županije i Grada Zagreba koji je TEN-T projekt čiji je korisnik HŽ Infrastruktura d.o.o. koji se financira iz Kohezijskog fonda. Peti projekt je projekt Hrvatskih voda financiran iz EFRR fonda na području Brodsko-posavske i Vukovarsko-srijemske županije kojim se žele modernizirati savski nasipi što spada pod financiranje prioriteta Klimatske promjene i upravljanje rizicima.

S obzirom da je određeni iznos bespovratnih sredstava EU dodijeljen određenom broju projekata po prioritetnim osima iz operativnih programa prema NUTS regijama, na grafičkom prikazu 21. iščitava se i na koji tip korisnika se to odnosi. Tako, za NUTS regije najveći iznos bespovratnih EU sredstava isplaćen je malim i srednjim poduzećima i to za Kontinentalnu Hrvatsku u vrijednosti od 23 milijarde kuna sa manje od četiri tisuće projekata od sveukupnih 9615 projekata, a za Jadransku Hrvatsku u vrijednosti od više od 16 milijardi kuna sa više od tisuću projekata.

Grafički prikaz 21.

Vrijednost bespovratnih sredstava EU i broj projekata prema NUTS regiji i korisnicima



Izvor: Prikaz autora

U Kontinentalnoj Hrvatskoj Javna tijela su dobila financiranje za više od dvije tisuće projekata sa trideset milijardi kuna, a u Jadranskoj Hrvatskoj za osamsto projekata u sveukupnom iznosu od šest milijardi kuna. Za kontinentalnu Hrvatsku po dobivenim bespovratnim sredstvima EU za projekte izdvajaju se sljedeći korisnici: istraživači (više od dvije milijarde kuna sa više od tristo projekata), neprofitne ili nevladine organizacije (više od milijardu kuna sa sedamsto projekata), velika poduzeća (više od milijardu kuna sa trideset projekata), centri (više od sto milijuna kuna sa pedeset projekata), ostali korisnici (oko sedamdeset milijuna kuna sa trideset projekata) i mladi (dva milijuna kuna sa jednim projektom).

Za Jadransku Hrvatsku, nakon malih i srednjih poduzeća te javnih tijela po dobivenim bespovratnim sredstvima EU za projekte izdvajaju se sljedeći korisnici: istraživači (manje od tristo milijuna kuna sa devedeset projekata), neprofitna ili nevladina organizacija (oko tristo pedeset milijuna kuna sa manje od tristo projekata), velika poduzeća (manje od sedamdeset milijuna kuna sa devet projekata), centri (više od četrdeset milijuna kuna sa dvadesetak projekata), ostali korisnici (manje od šezdeset milijuna kuna sa više od deset projekata) i mladi.

Jako veliki projekti financiraju se iz ESF-a, IZM-a, EFRR i KF-a. Tako, izdvajaju se dva jako velika projekta iz OPULJP-a na području Grada Zagreba i to su projekti HZZ-a. Jednim projektom se podupire samozapošljavanje, a drugim se provodi mjera aktivne politike zapošljavanja mladih. Za financiranje projekata iz OPKK-a po iznosu financiranja izdvajaju se projekti HBOR grupe preko kojih se omogućava odobravanje ESIF kredita za rast i razvoj čiji su korisnici mala i srednja poduzeća te se izdvajaju krediti za energetska učinkovitost javne zgrade čiji su korisnici javna tijela. Isto, izdvajaju se i krediti koje odobrava Hrvatska agencija za malo gospodarstvo, inovacije i investicije čiji su korisnici mala i srednja poduzeća i slično. U ostalim županijama jako veliki projekti odnose se na ulaganja u poboljšanje vodno-komunalne infrastrukture ili na sustav vodoopskrbe, odvodnje i pročišćavanja otpadnih voda. Ulaže se i u gospodarenje otpadom, izgradnje željeznica, razvoj zračne luke Dubrovnik, cestovnu povezanost s Južnom Dalmacijom što se odnosi na Dubrovačko-neretvansku županiju te na izgradnju potrebnih prometnica (TEN-T). U Dubrovačko-neretvanskoj županiji izdvajaju se još dva jako velika projekta to je ulaganje u izgradnju centra za gospodarenje otpadom i razvoj vodnokomunalne infrastrukture. U jako velike projekte spadaju projekti čija je ukupna ugovorena vrijednost projekta veća od 300 milijuna kuna, a u male projekte spadaju projekti do 1,5 milijuna kuna kao što je već prije navedeno.

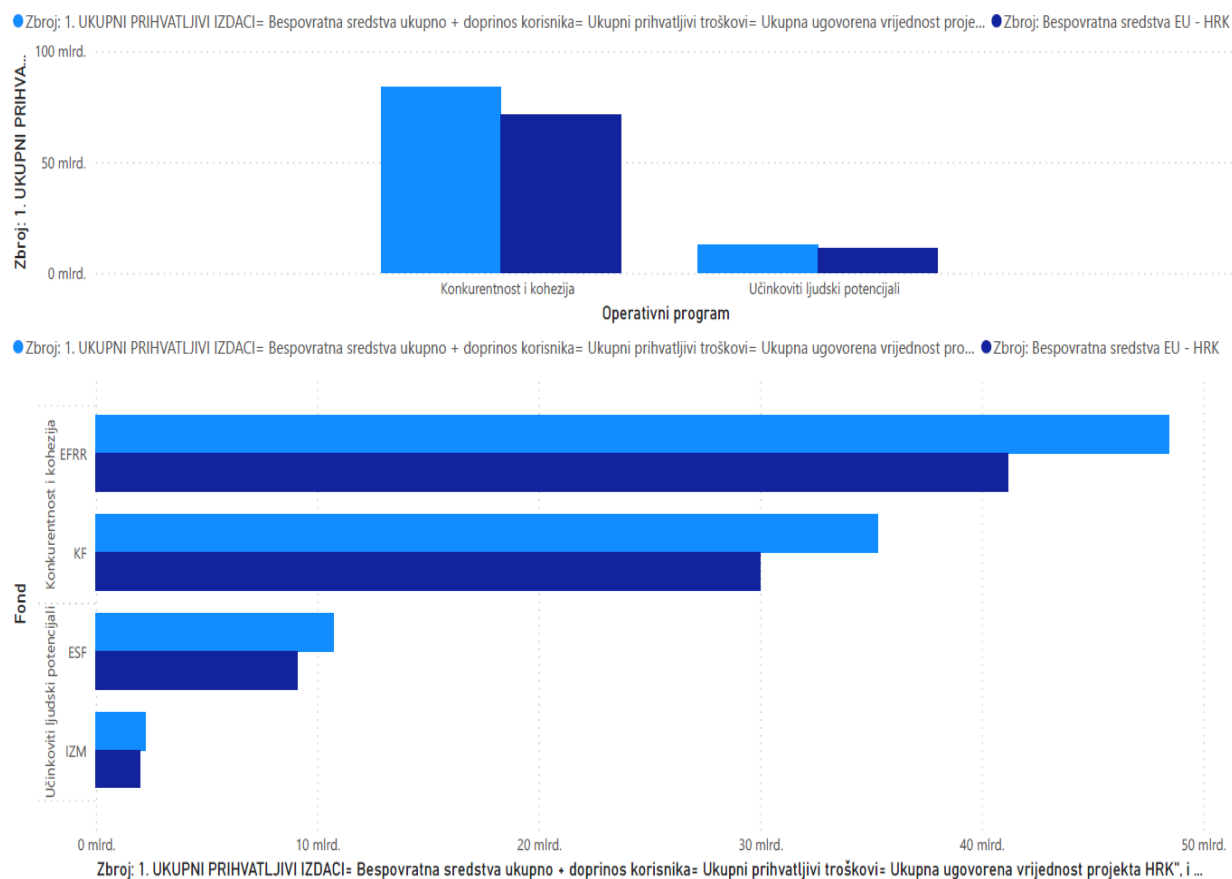
Preko grafičkog prikaza 22. dobiva se uvid u ukupna bespovratna sredstva EU uložena u projekte iz OPKK-a i OPULJP-a, kao i ukupna ugovorena vrijednost svih projekata koji se odnose na provedbu kohezijske politike u Hrvatskoj. Iz Operativnog programa „Konkurentnost i kohezija“ isplaćeno je više od sedamdeset milijardi kuna, dok je iz Operativnog programa „Učinkoviti ljudski potencijali“ isplaćeno nešto više od jedanaest milijardi kuna za financiranje projekata. Iz Operativnog programa „Konkurentnost i kohezija“ je isplaćeno više od 41 milijarde kuna iz EFRR fonda, a iz Kohezijskog fonda je isplaćeno više od 30 milijarde kuna bespovratnih sredstava EU.

S druge strane, iz Operativnog programa „Učinkoviti ljudski potencijali“ isplaćeno je više od 9 milijardi kuna, a iz IZM fonda su isplaćene oko 2 milijarde kuna. Isto, usporedbom ukupne ugovorene vrijednosti projekata i bespovratnih sredstava može se vidjeti koliki je iznos financiran bespovratnim sredstvima EU i

iz kojeg fonda. Tako, većina projekata financirana je 85 posto bespovratnim sredstvima EU, dok se preostali iznos, od 15 posto, financira nacionalnim doprinosom ili javnim i privatnim financiranjem korisnika. Za OPKK iznos nacionalnog financiranja za sve projekte iznosi 15%. Kod OPULJP-a iznos nacionalnog doprinosa za većinu projekata je 15 posto, ali neki projekti su dijelom financirani javnim i privatnim financiranjem. Razlika između ukupnog bespovratnog financiranja EU i ukupne ugovorene vrijednosti projekta je vidljiva na slici.

Grafički prikaz 22.

Ukupna ugovorena vrijednost projekata i bespovratnih sredstava EU prema Operativnim programima i fondovima namijenjenih provedbi kohezijske politike



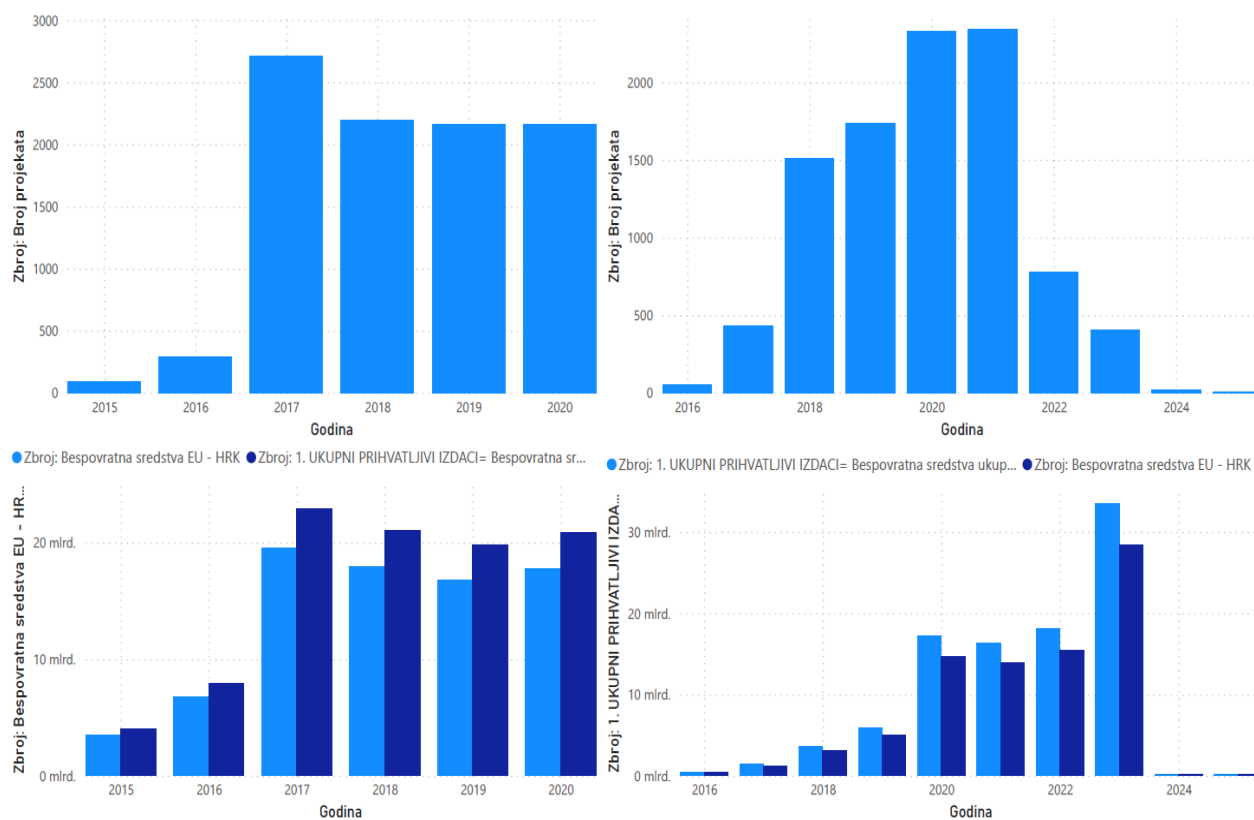
Izvor: Prikaz autora

Financiranje bespovratnim sredstvima EU započelo je u 2014. godini. Prema grafičkom prikazu 23., vidi se da su na početku financijske perspektive jako mali iznosi bespovratnih sredstava kao i što je financiran mali broj projekata. Tako, započelo je financiranje više od 350 projekata čija je vrijednost bespovratnih

sredstava iznosila približno deset milijardi kuna u 2015. i 2016. godini. S druge strane, samo u 2017. godini je započelo financiranje oko tri tisuće projekata čija je vrijednost bespovratnih sredstava iznosila približno dvadeset milijardi kuna. Od 2018. do 2020. godine broj projekata je ostajao približno isti. Tako je za svaku godinu, od 2018. godine, započeto financiranje više od dvije tisuće projekata čija je ukupna vrijednost bespovratnih sredstava EU svake godine iznosila oko 17 milijardi kuna.

Grafički prikaz 23.

Ukupan broj projekata i bespovratnih sredstava EU prema godini početka i godini završetka projekta



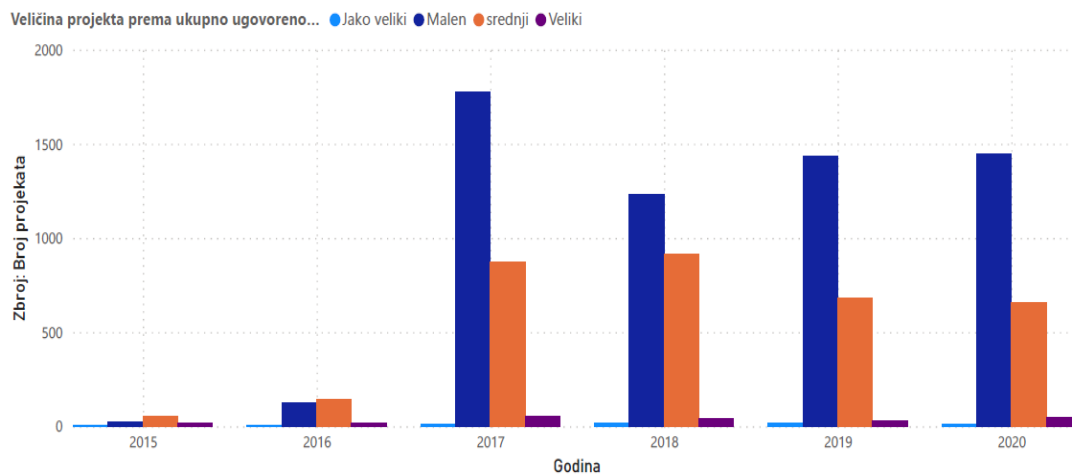
Izvor: Prikaz autora

Isto, na grafičkom prikazu 23. vidi se da je više projekata započelo u 2016. i 2017. godini nego što ih je završeno, ali je zato više projekata završeno u 2020. i 2021. godini nego što ih je započeto. Završetak nekih projekata proteže se i do 2025.godine, a zanimljivo je da je u 2023. godini završeno oko 400 projekata čija je ukupna vrijednost iznosa bespovratnih sredstava EU bila manja od 29 milijarde kuna što je neusporedivo visoka vrijednost iznosa bespovratnih sredstava EU u usporedbi s ostalim godinama.

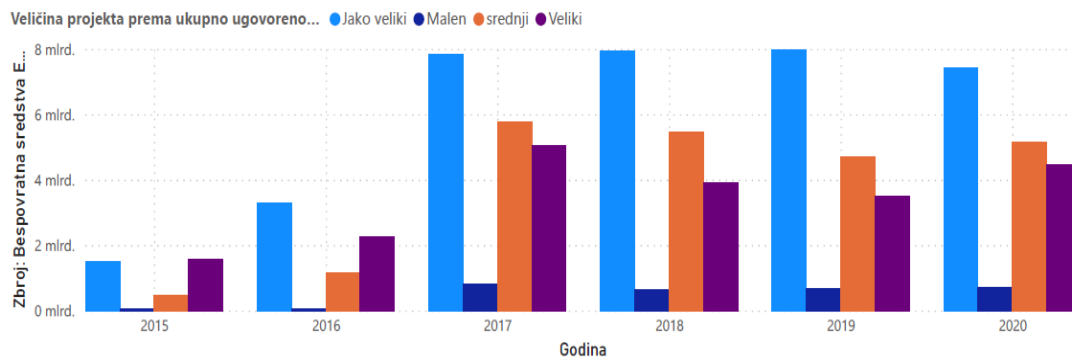
Kao što je prikazano na grafičkom prikazu 24., u 2015. godini i 2016. godini je jako malo projekata financirano bespovratnim sredstvima EU, ali po iznosu bespovratnih sredstava isticali su se jako veliki i veliki projekti. Primjerice, iako su u 2015. godini započela dva jako velika projekta, njihova vrijednost bespovratnih sredstava iznosila je oko dvije milijarde kuna, a za 16 velikih projekata vrijednost bespovratnih sredstava je isto iznosila oko dvije milijarde kuna. Vrijednost malih i srednjih projekata nije prešla milijardu kuna te je iznosila oko 500 milijuna kuna za niti 100 projekata koji su započeli te godine. Razlog zbog kojega je financirano jako malo projekata može biti do natječaja koji su se provodili i do nedovoljne informiranosti javnosti o financiranju iz EU fondova što se promijenilo u 2017. godini kada je porastao broj projekata koji su započinjali, a sa brojem projekata povećala se i ukupna vrijednost bespovratnih sredstava EU.

Grafički prikaz 24.

Ukupan broj projekata prema godini početka operacije i veličini projekt



Zbroj: Bespovratna sredstva EU - HRK kategorije Godina i Veličina projekta prema ukupno ugovorenoj vrijednosti projekta



Izvor: Prikaz autora

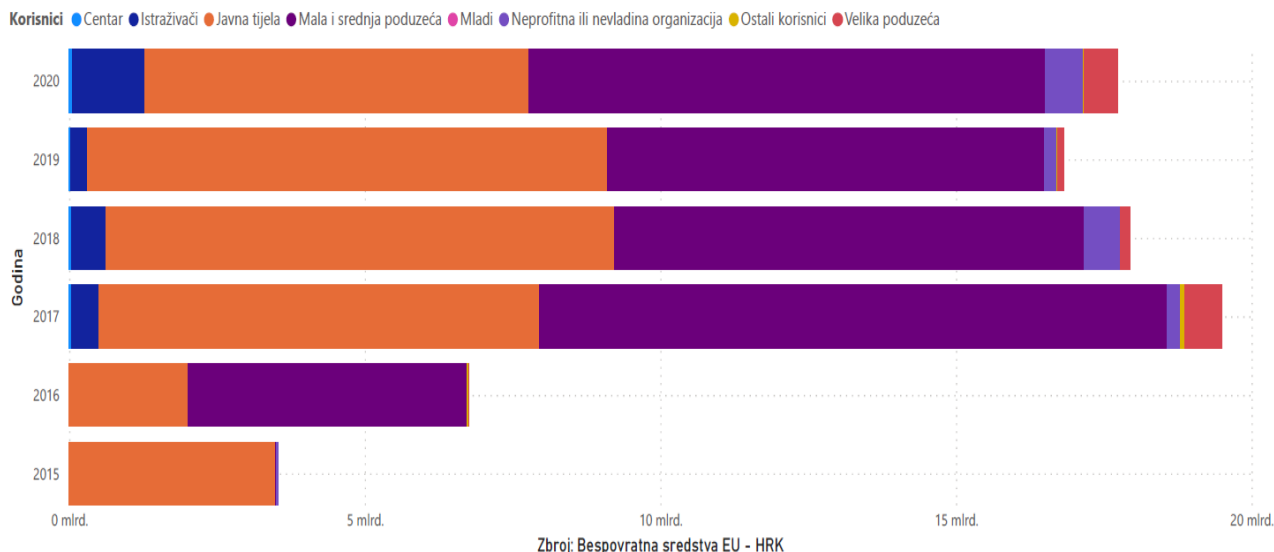
Tako, u 2017. godini broj jako velikih projekata koji su započinjali povećao se na 11 projekata čija je ukupna vrijednost bespovratnog financiranja EU iznosila skoro osam milijardi kuna, dok je za 50-ak velikih projekata ukupna vrijednost bespovratnog financiranja iznosila oko pet milijardi kuna. Za nešto više od osamsto projekata srednje veličine ukupna vrijednost bespovratnog financiranja iznosila je oko šest milijardi kuna, dok je za skoro dvije tisuće malih projekata ukupna vrijednost bespovratnog financiranja EU iznosila oko osamsto milijuna kuna što znači da se bespovratno financiranje povećalo sa petnaestak milijuna kuna u 2015. godini na osamsto milijuna kuna u bespovratnim sredstvima EU u 2017. godini. U zadnjoj godini financiranja, 2020. godini, broj projekata koji su započinjali i iznos bespovratnog financiranja bio je približno isti. Primjerice, ukupna vrijednost bespovratnog financiranja za jako velike projekte iznosila je oko osam milijardi kuna, za velike projekte ta vrijednost se kretala između 3,5-4,5 milijarde kuna, za srednje projekte oko 4,6-5,1 milijarde kuna, a za male projekte ukupna vrijednost bespovratnog financiranja kretala se oko 650-700 milijuna kuna.

Na grafičkom prikazu 25., vidi se da je u 2015. godini, godini započinjanja projekta, ukupna vrijednost bespovratnih sredstava iznosila oko 4 milijarde kuna za Javna tijela odnosno više od 98% ukupnog bespovratnog financiranja ostvarenog za 2015. godinu, dok su jedan manji dio (0,41%) činila financiranja projekata malih i srednjih poduzeća u iznosu od četrnaest milijuna kuna te neprofitnih ili nevladinih organizacija u iznosu od 30 milijuna kuna (0,88%).

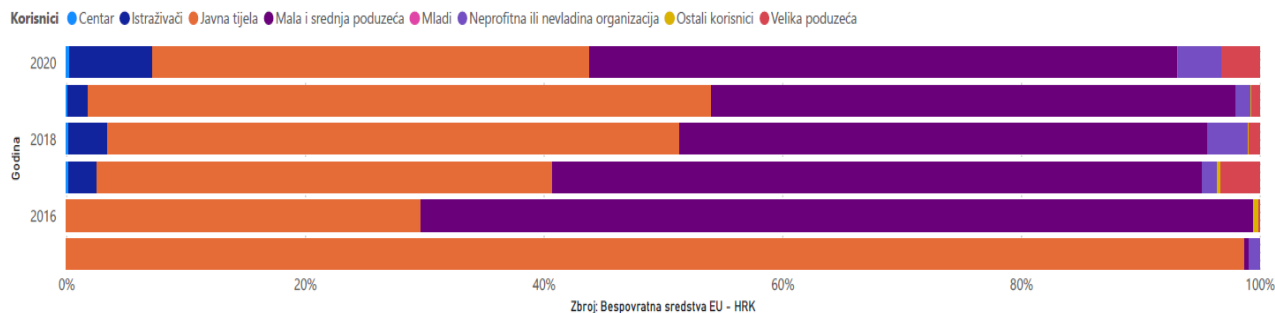
U 2016. godini, na projekte Javnih tijela odnosilo se 29,75% ukupnog iznosa bespovratnih sredstava EU ostvarenih za tu godinu odnosno oko dvije milijarde kuna, na projekte malih i srednjih poduzeća 69,71% odnosno približno 5 milijardi kuna, a preostali iznos odnosi se na velika poduzeća (0,08%) i na ostale korisnike (0,46%). U 2017. godini najveći udio financiranja odnosio se na mala i srednja poduzeća negdje oko 54,45% odnosno oko 10,6 milijardi kuna te na javna tijela negdje oko 38,23% odnosno oko 7,4 milijarde kuna, a ostali iznosi bespovratnog financiranja EU odnose se na velika poduzeća, ostale korisnike, centre, istraživače i neprofitne ili nevladine organizacije. Vrijednost bespovratnih sredstava EU za projekte koji su započeli u 2017. godini, a koji se odnose na velika poduzeća iznose oko 630 milijuna kuna, za istraživače oko 452 milijuna kuna te za neprofitne organizacije oko 233 milijuna kuna. U 2018. godini i 2019. godini najveća vrijednost bespovratnih sredstava EU započelih projekata kretala se od 8,6 do 8,8 milijarde kuna za Javna tijela te za mala i srednja poduzeća od 7,3 do 7,9 milijarde kuna.

Grafički prikaz 25.

Ukupan iznos bespovratnih sredstava EU prema godini početka projekta i vrsti korisnika



Zbroj: Bespovratna sredstva EU - HRK kategorije Godina i Korisnici



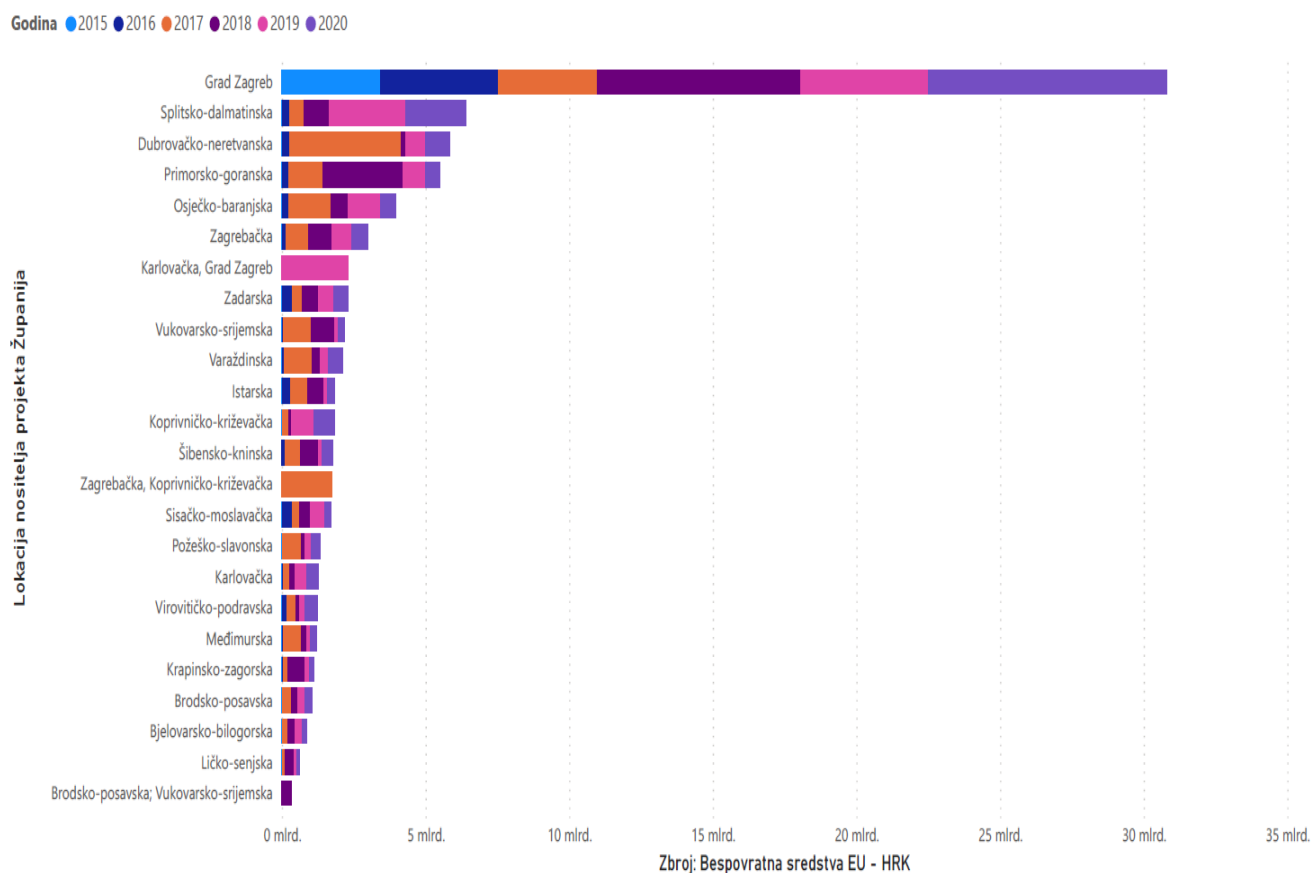
Izvor: Prikaz autora

U 2020. godini, zadnjoj godini financijske perspektive 2014.-2020. godine, najveća vrijednost bespovratnih sredstava EU započetih projekata iznosila je 8,7 milijardi kuna za najveće korisnike, a to su mala i srednja poduzeća te za javna tijela je iznosila oko 6,4 milijarde kuna. Za istraživače vrijednost bespovratnih sredstava EU započetih projekata iznosila je oko 1,2 milijarde kuna, za neprofitne organizacije 641 milijuna kuna, za velika poduzeća 561 milijuna kuna te za centre 54 milijuna kuna. Pod korisnike kao centre financirani su projekti Kliničkog bolničkog centra Split, Razvojni centar poduzetništva, Centar za odgoj i obrazovanje, Volonterski centri, Centar za ruralni razvoj, Centar za socijalno poduzetništvo i slično. Pod Javna tijela podrazumijevaju se općine, gradovi, županije, Domovi zdravlja, poliklinike, sveučilišta, domovi za starije i slično, dječji vrtići, Državna geodetska uprava, Hrvatski zavod za mirovinsko osiguranje, Ministarstva i slično.

Za koju županiju je apsorbirano najviše bespovratnih sredstava EU za projekte po godini započinjanja projekta, prema grafičkom prikazu 26. vidi se da ukupna vrijednost bespovratnih sredstava EU bila najveća za projekte koji su započeli u Gradu Zagrebu od 2014. godine do 2020. godine. U 2015. godini ukupna vrijednost bespovratnih sredstava EU započetih projekata iznosila je oko 3,4 milijarde kuna, za 2016. godinu oko 4 milijarde kuna, za 2017. godinu oko 3,4 milijarde kuna, za 2018. godinu oko 7 milijardi kuna, 2019. godinu oko 4,4 milijarde kuna i za 2020. godinu oko 8,3 milijarde kuna. Tako da je Grad Zagreb kroz fondove EU financiran sa nešto više od 30 milijardi kuna.

Grafički prikaz 26.

Ukupna vrijednost bespovratnih sredstava EU prema županiji i godini početka projekta



Izvor: Prikaz autora

Sljedeća županija koja se ističe po započetim financiranim projektima bespovratnim sredstvima EU je Splitsko-dalmatinska županija čiji su projekti korisnika financirani u sveukupnom iznosu bespovratnih sredstava EU sa nešto više od 6 milijardi kuna od 2014. godine do 2020. godine. S obzirom da nije ostvareno financiranje projekata u 2015. godini, ukupna vrijednost bespovratnih sredstava EU započetih projekata

za 2016. godinu iznosila je nešto manje od 270 milijuna kuna, za 2017. godinu oko 500 milijuna kuna, za 2018. godinu oko 866 milijuna kuna, za 2019. godinu 2,6 milijardi kuna, a u 2020. godini vrijednost bespovratnih sredstava EU započelih projekata Splitsko-dalmatinske županije iznosila je oko 2,1 milijarde kuna. U Dubrovačko-neretvanskoj županiji, najveći iznos bespovratnih sredstava EU ostvaren je u 2017. godini čija je vrijednost bespovratnih sredstava EU započelih projekata iznosila oko 2,6 milijarde kuna, u Primorsko-goranskoj najveći iznos bespovratnih sredstava EU projekata koji su započeli u 2018. godini iznosio je približno oko 2,8 milijardi kuna, u Osječkoj-baranjskoj najveća vrijednost bespovratnih sredstava EU ostvarena je u 2017. godini i iznosila je oko 1,4 milijarde kuna. Ostale županije redom navedene su: Zagrebačka županija u 2018. godini sa oko 800 milijuna kuna vrijednosti bespovratnih sredstava EU započelih projekata, Zadarska županija u 2018. godini sa započetim projektima vrijednosti bespovratnih sredstava EU od 561 milijuna kuna, Vukovarsko-srijemska županija u 2017. godini sa 964 milijuna kuna, Varaždinska u 2017. godini sa 968 milijuna kuna, Istarska županija u 2017. godini sa 600 milijuna kuna, Koprivničko-križevačka u 2019. godini sa oko 800 milijuna kuna, Šibensko-kninska u 2018. godini sa više od 620 milijuna kuna, Sisačko-moslavačka najviše je bespovratnih sredstava EU apsorbirala za financiranje projekata koji su započeli u 2019. godini negdje oko 500 milijuna kuna. Ličko-senjska županija koja je u svakom pogledu imala najmanji sveukupan iznos financiranja bespovratnim sredstvima EU u 2018. godini je ostvarila najveći iznos financiranja projekata bespovratnim sredstvima EU odnosno započeli su projekti čija je vrijednost bespovratnih sredstava EU iznosila nešto manje od 330 milijuna kuna.

Ukupna vrijednost bespovratnih sredstava EU prema prioritetnoj osi i godini kada su započeli projekti nalazi se na grafičkom prikazu 27. prema kojem je najveći ukupan iznos bespovratnih sredstava EU ostvaren za realiziranje Prioritetne osi Zaštita okoliša i održivost resursa negdje oko 24 milijarde kuna, a najmanji ukupan iznos bespovratnih sredstava EU ostvaren je za realiziranje cilja Dobro upravljanje OP „Učinkoviti ljudski resursi“.

Za ostvarenje Prioritetnog cilja Zaštita okoliša i održivost resursa najveća vrijednost bespovratnih sredstava EU je za projekte koji su započeli u 2018. godini te iznosi oko 6,7 milijardi kuna. Vrijednost bespovratnih sredstava EU projekata koji su započeli u 2019. g iznosi oko 6,6 milijarde kuna, u 2017. godini iznosi oko 5,1 milijarde kuna, u 2020. godini iznosi 4,7 milijardi kuna, dok je najmanja vrijednost bespovratnih sredstava EU apsorbirana za projekte koji su započeli u 2016. godini i iznosila je oko 1,2 milijarde kuna, u 2015. godini nema zabilježenih podataka o financiranim projektima.

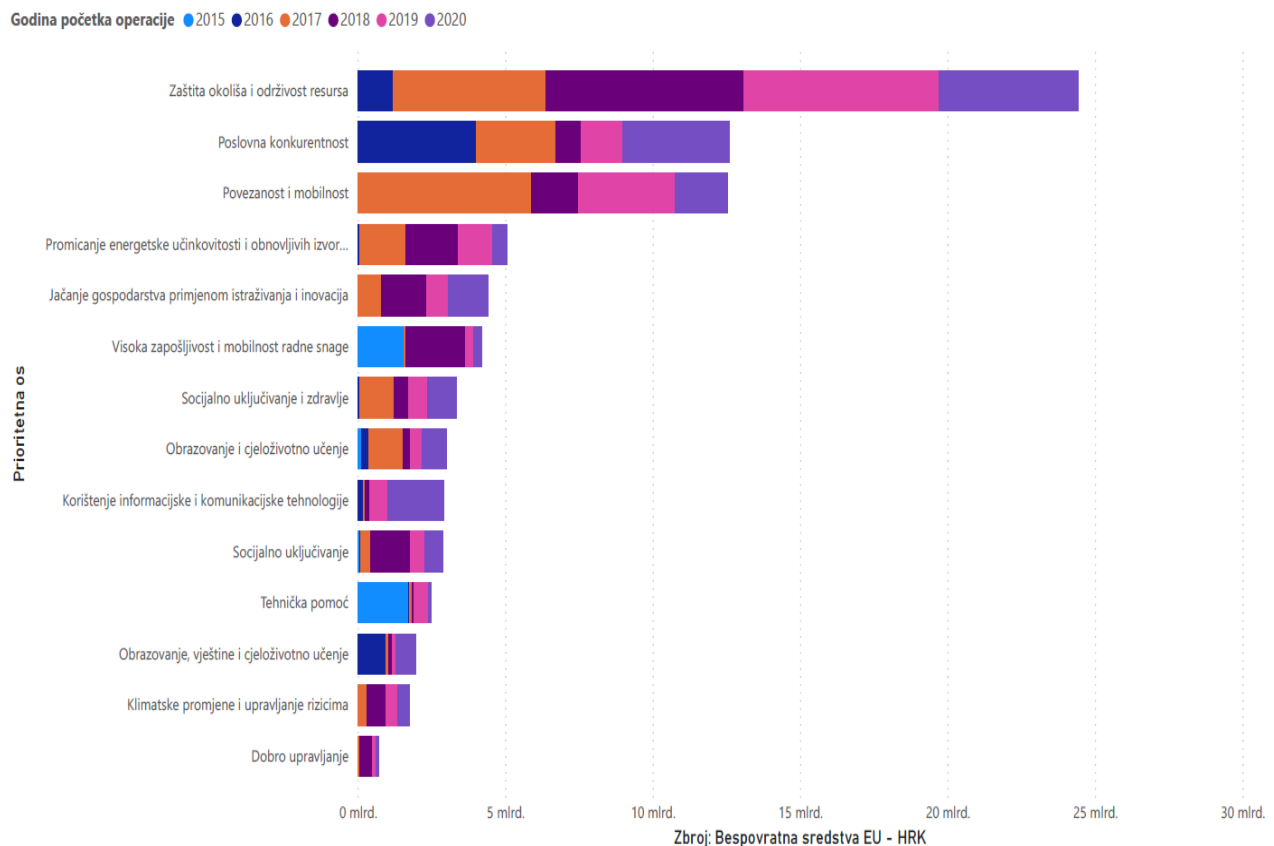
Prema vrijednosti bespovratnih sredstava EU, drugi prioritetni cilj koji se ističe je cilj Poslovna konkurentnost za čije je ostvarenje financirano projekata u vrijednosti bespovratnih sredstava EU oko 13

milijardi kuna. Za projekte koji su započeli u 2016. godini vrijednost bespovratnih sredstava EU iznosi oko 4 milijarde kuna, za projekte koji su započeli u 2017. godini vrijednost bespovratnih sredstava EU iznosi oko 2,6 milijardi kuna, za započete projekte 2018. godine vrijednost bespovratnih sredstava EU iznosi oko 856 milijuna kuna što je najmanji iznos ostvaren u nekoj godini za ostvarenje cilja Poslovna konkurentnost. U 2019. godini vrijednost bespovratnih sredstava EU započetih projekata iznosi 1,4 milijarde kuna, a u 2020. godini taj iznos se kreće oko 3,6 milijardi kuna.

Prema vrijednosti bespovratnih sredstava EU, ostali ciljevi koji se nastoje ostvariti financiranjem projekata redom navedeni su: Povezanost i mobilnost, Promicanje energetske učinkovitosti i obnovljivih izvora energije, Jačanje gospodarstva primjenom istraživanja i inovacija, Visoka zapošljivost i mobilnost radne snage, Socijalno uključivanje i zdravlje, Obrazovanje i cjeloživotno učenje i ostali.

Grafički prikaz 27.

Ukupna vrijednost bespovratnih sredstava EU prema prioritetnoj osi i godini početka projekta



Izvor: Prikaz autora

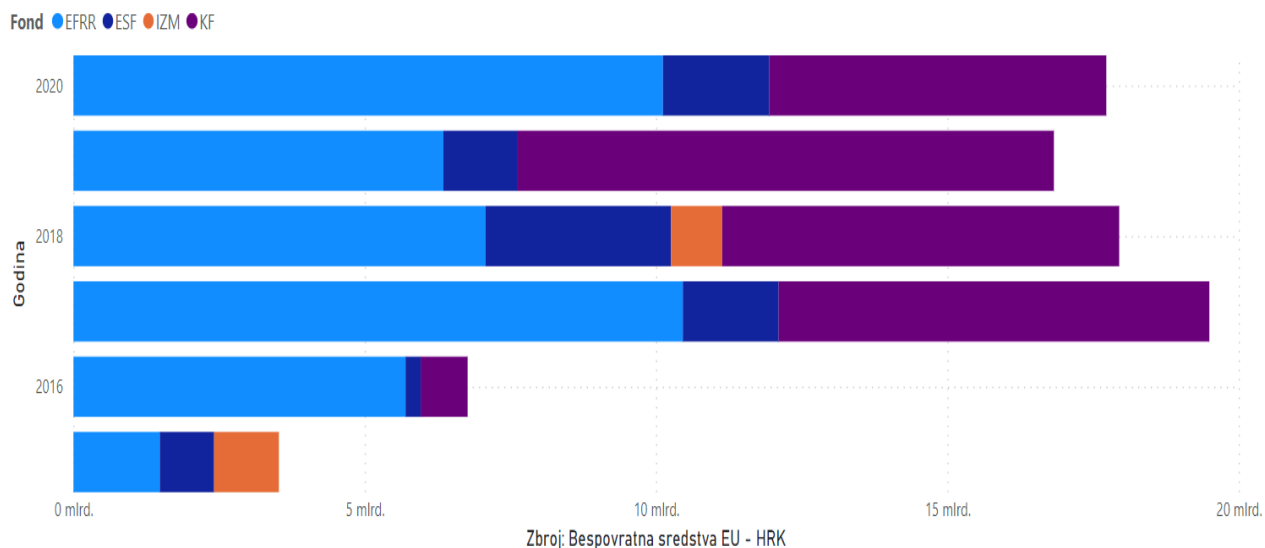
U 2017. godini su započeli projekti za koje je najveća vrijednost bespovratnih sredstava EU iznosila oko 5,8 milijardi kuna za ostvarenje cilja Povezanost i mobilnost. Za ostvarenje cilja Promicanje energetske učinkovitosti i obnovljivih izvora najveća vrijednost bespovratnih sredstava EU dodijeljenih projektima koji su započeli u 2018. godini i iznosila je oko 1,8 milijardi kuna. Za ostvarenje cilja Jačanje gospodarstva primjenom istraživanja i inovacija najveća vrijednost bespovratnih sredstava EU dodijeljenih projektima koji su započeli u 2018. godini i iznosila je oko 1,5 milijardi kuna. Sljedeći cilj je Visoka zapošljivost i mobilnost radne snage za čije je ostvarenje najveća vrijednost bespovratnih sredstava EU dodijeljenih projektima koji su započeli 2018. godine i iznosila je oko 2 milijarde kuna.

Najmanji ukupan iznos bespovratnih sredstava EU dodijeljenih projektima ostvaren je za Prioritetni cilj Dobro upravljanje i iznosi više od 700 milijuna kuna. Najveća vrijednost bespovratnih sredstava EU za ovaj prioritetni cilj ostvarena je za projekte započete u 2018. godini i iznosi oko 430 milijuna kuna.

Na grafičkom prikazu 28., vidljivo je kolika je ukupna vrijednost bespovratnih sredstava EU dodijeljenih projektima prema godini u kojoj su započeli projekti i prema fondu iz kojeg su isplaćivana bespovratna sredstva EU. Tako, najmanja ukupna vrijednost bespovratnih sredstava EU projekata koji su započeli u 2015. godini iznosi oko 3,5 milijarde kuna, dok je najveća ukupna vrijednost bespovratnih sredstava EU projekata koji su započeli u 2017. godini i iznosi oko 19,5 milijardi kuna.

Grafički prikaz 28.

Ukupna vrijednost bespovratnih sredstava EU prema fondu i godini početka operacije



Izvor: Prikaz autora

U 2015. godini najveća vrijednost bespovratnih sredstava EU za projekte financirana je iz EFRR fonda i iznosila je oko 1,4 milijarde kuna, a slijedi ga IZM fond sa 1,1 milijardu kuna te Europski socijalni fond sa oko 930 milijuna kuna. S obzirom da je najveća ukupna vrijednost bespovratnih sredstava EU projekata koji su započeli u 2017. godini, najveći iznos bespovratnih sredstava EU financirano je iz EFRR fonda u iznosu od 10,4 milijarde kuna, a kojeg slijedi financiranje projekata iz Kohezijskog fonda čija je vrijednost bespovratnih sredstava EU iznosila približno 7,4 milijarde kuna te ESF fonda sa približno 1,6 milijardi kuna. U zadnjoj godini financijske perspektive, za projekte započete 2020. godine, vrijednost bespovratnih sredstava EU projektima isplaćenih prema fondovima iznosila je oko 10,1 milijardu kuna za EFRR, oko 5,7 milijardi kuna za KF te oko 1,7 milijardi kuna za ESF.

4.3 Analiza podataka i rezultati istraživanja

Prebacivanjem svih podataka u Excel tablicu omogućeno je provođenje analize i procesa rudarenja podataka na javnoj bazi projekata EU. No, prije nego što se provede analiza rudarenja podataka, sirove podatke u tablici je potrebno formatirati, urediti nedostajuće podatke njihovom eliminacijom ili pronalaženjem ispravnih vrijednosti ukoliko je to moguće te ih pretvoriti u oblik prilagođen za analizu podataka. U istraživanju na zadanoj bazi podataka koristi se „RapidMiner“ i „Orange“ koji su open-source alati namijenjen početnicima i stručnjacima za provođenje analize i rudarenja podataka.

S obzirom da u setu podataka prevladavaju kategorijske varijable, gdje je potrebno, provest će se kodiranje kategorijskih varijabli u numeričke varijable, u obliku „dummy“ varijabli koje daju bolje rezultate ili će se provesti kodiranje varijabli „label“ metodom koje je karakteristično za ordinalne varijable. Primjerice, ordinalna varijabla je veličina projekta jer razlikujemo male, srednje, velike i jako velike projekte koji će se kodirati brojevima 0,1,2,3. S obzirom da su NUTS regija i Operativni program dihotomne varijable, bit će dovoljno kodiranje varijabli u binarni oblik 0 i 1.

Započinje se s klasteriranjem kao jednom od primijenjenih metoda podataka na odabranom setu podataka. U obzir su uzete samo numeričke varijable: vrijednost bespovratnih sredstava EU, vrijednost bespovratnih sredstava nacionalnog, privatnog i javnog doprinosa, mjeseci trajanja projekta i EU stopa sufinanciranja projekta. Pomoću k-means operatora, opažanja su se raspodijelila u četiri klastera. Ukupno 9615 opažanja raspoređeno je u klaster na sljedeći način: klaster nula sadrži 9540 opažanja, klaster jedan sadrži 5 opažanja dok klaster broj dva ima 70 opažanja. Prethodno navedeno možemo vidjeti na sljedećoj slici.

Grafički prikaz 29.

Model klastera

```
Cluster 0: 9540 items
Cluster 1: 5 items
Cluster 2: 70 items
Total number of items: 9615
```

Izvor: Prikaz autora u alatu RapidMiner Studio

Cilj klasteriranja je pronalaženje prirodne veze među svim opažanjima unutar uzroka, koji se prema određenim sličnostima razvrstavaju u klaster, te se srednje vrijednosti tih klastera međusobno raspoređuju.

Grafički prikaz 30.

Prikaz srednjih vrijednosti za svaki atribut prema 3 zadana klastera

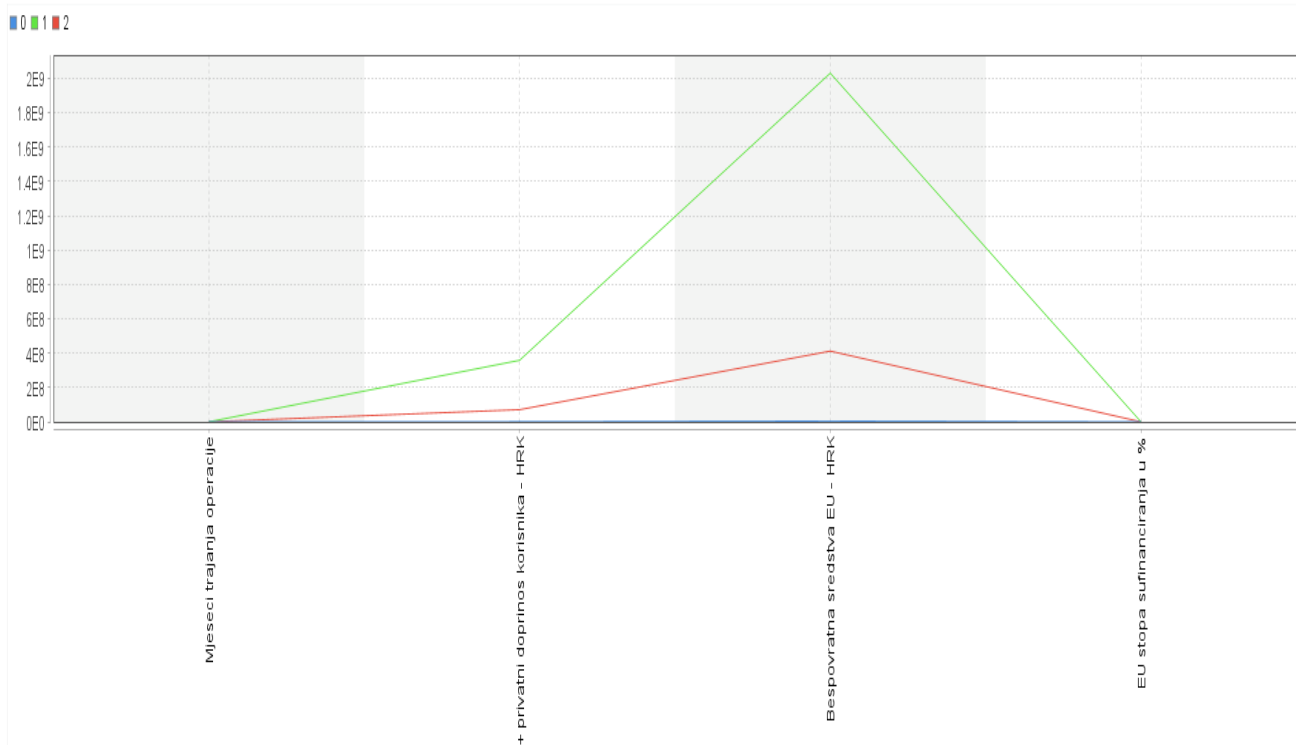
Attribute	cluster_0	cluster_1	cluster_2
Mjeseci trajanja operacije	17.902	47.200	48.357
Bespovratna sredstva - Nacionalni dio + Javni + privatni doprinos	802046.594	358169858.892	70765906.472
Bespovratna sredstva EU - HRK	4541054.833	2029629200.386	411725540.880
EU stopa sufinanciranja u %	0.850	0.850	0.852

Izvor: Prikaz autora u alatu RapidMiner Studio

Iz prikaza srednjih vrijednosti može se vidjeti da klaster 2 ima najveću prosječnu vrijednost vremenskog razdoblja trajanja projekta i EU stope sufinanciranja. S druge strane, klaster 0 ima najmanju prosječnu vrijednost vremenskog razdoblja trajanja projekta (17,9 mjeseci u prosjeku), vrijednosti bespovratnih sredstava EU (prosječno 4,5 milijuna kuna) i sredstava nacionalnog, javnog i privatnog doprinosa (prosječno 800 tisuća kuna). Također, iz grafičkog prikaza 31. može se uočiti koji klaster ima najveću srednju vrijednost za određeni atribut pa se tako sa najvećom srednjom vrijednošću bespovratnih sredstava EU i nacionalnog, javnog i privatnog doprinosa izdvaja klaster 1, a potom i klaster 2.

Grafički prikaz 31.

Grafički prikaz srednjih vrijednosti za svaki atribut prema 3 zadana klastera

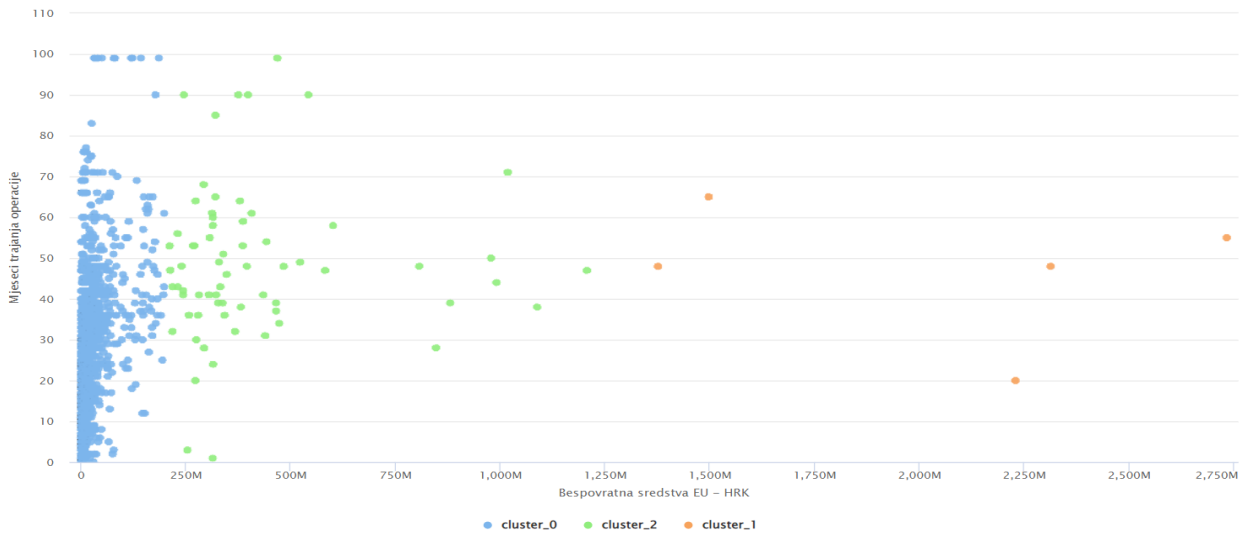


Izvor: Prikaz autora u alatu RapidMiner Studio

Set podataka sadrži veliku količinu outliera. No, s obzirom da se radi o istinitim outlierima, opažanja su zadržana u setu podataka jer za uklanjanje outliera iz seta podataka mora postojati jako dobro obrazloženje, a koje u ovom slučaju ne postoji. Tako, iz rezultata klasteriranja može se uočiti da klaster 1 sadrži opažanja čija se vrijednost bespovratnih sredstava EU varira od 1,3 milijarde kuna i više, a u klasteru 2 vrijednost bespovratnih sredstava varira od 220 milijuna kuna do 1,25 milijardi kuna. Sve vrijednosti bespovratnih sredstava EU koje su manje od 220 milijuna kuna pripadaju klasteru 0 što se može lako uočiti i na grafičkom prikazu 46. na kojem su prikazani klasteri s obzirom na vrijednosti bespovratnih sredstava EU i vremensko razdoblje trajanja projekta.

Grafički prikaz 32.

Klasteri s obzirom na bespovratna sredstva EU i vremensko razdoblje trajanja projekta



Izvor: Prikaz autora u alatu RapidMiner Studio

Klasteri dobiveni k-means metodom mogu se upotrijebiti kako bi se s obzirom na istražene varijable stekli uvidi u set podataka. Tako na grafičkom prikazu 33. poznavajući broj opažanja koji se nalazi u svakome klasteru, može se zaključiti da se u klasteru 1 nalaze jako veliki projekti s obzirom da ukupna vrijednost bespovratnih sredstava EU za samo 5 opažanja iznosi oko 10 milijardi kuna. Isto, ističe se klaster 0 jer sadrži najveći broj opažanja, njih 9540, a ukupna vrijednost bespovratnih sredstava EU iznosi nešto više od 40 milijardi kuna.

Grafički prikaz 33.

Klasteriranje prema ukupnoj vrijednosti bespovratnih sredstava EU

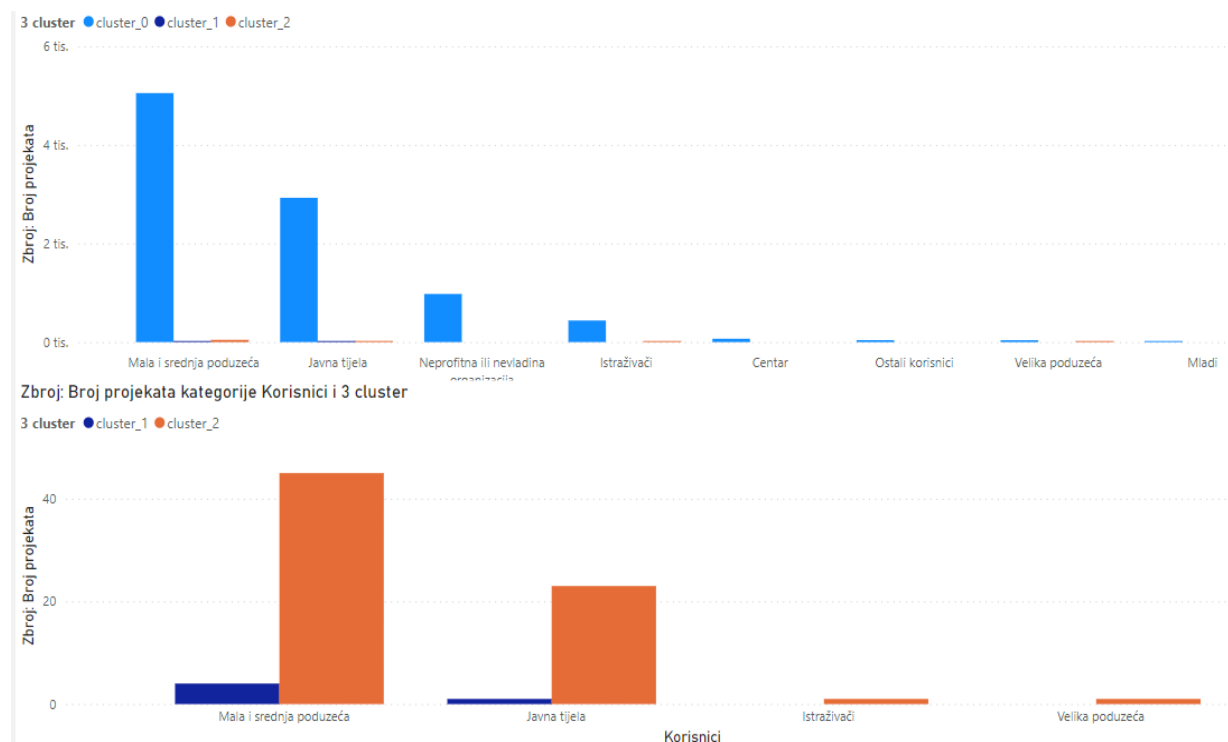


Izvor: Prikaz autora u Power BI alatu

Tako, na grafičkom prikazu 34. može se vidjeti da za klaster 0, sa najvećim brojem opažanja, prevladavaju mala i srednja poduzeća, Javna tijela, neprofitne ili nevladine organizacije te istraživači. U klasteru 1, najveći broj korisnika odnosi se na mala i srednja poduzeća i Javna tijela, a u klasteru 2 najveći broj korisnika odnosi se na mala i srednja poduzeća i Javna tijela, a najmanji broj korisnika odnosi se na velika poduzeća i istraživače.

Grafički prikaz 34.

Klasteriranje prema tipu korisnika

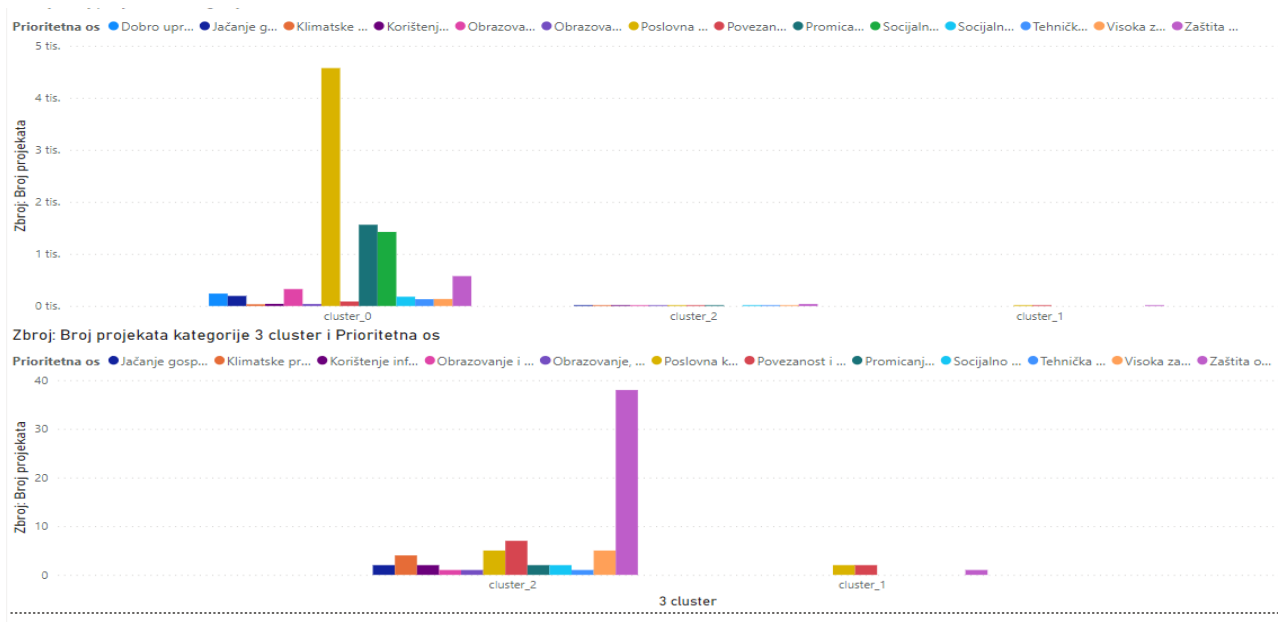


Izvor: Prikaz autora u Power BI alatu

Na grafičkom prikazu 35. vidi se da je u klasteru 0, daleko najzastupljeniji prioritet „Poslovna konkurentnost“ kojeg slijede prioriteti „Promicanje učinkovitosti i obnovljivih izvora energije“ i „Socijalna uključenost i zdravlje“. U klasteru 2 najzastupljeniji prioritet je „Zaštita okoliša i održivost resursa“, a u klasteru 1 ističe se „Poslovna konkurentnost“ i „Povezanost i mobilnost“.

Grafički prikaz 35.

Klasteriranje prema vrsti prioritetne osi

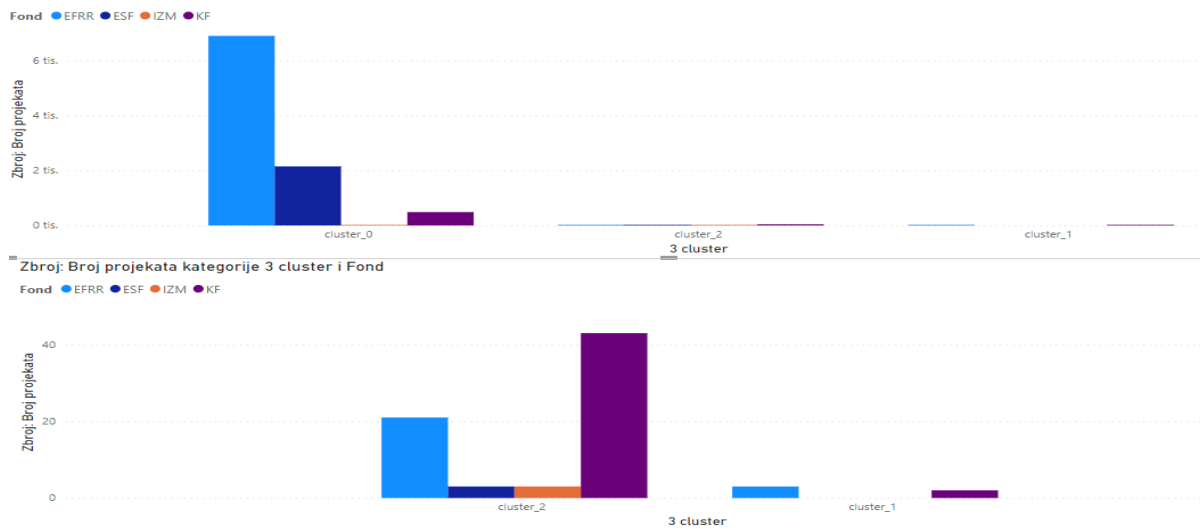


Izvor: Prikaz autora u Power BI alatu

Svrstavajući projekte prema sličnim karakteristikama metodom klasteriranja za varijablu „fond“, dobivena tri klastera ukazuju na to da se u klasteru 0 projekti najviše svrstavaju prema EFRR fondu, potom ESF fondu i KF fondu. U klasteru 1 projekti su svrstani prema EFRR i KF fondu, a u klasteru 2 najzastupljeniji su KF fond i EFRR fond.

Grafički prikaz 36.

Klasteriranje prema vrsti fonda

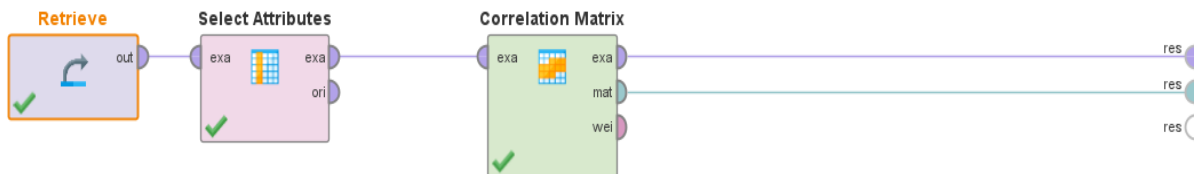


Izvor: Prikaz autora u Power BI alatu

Druga metoda rudarenja podataka koja će se primijeniti i kojom će se nastojati dokazati postojanje povezanosti između varijabli je korelacija. Tako, na slici 37. dan je grafički prikaz izrađenog korelacijskog modela, a na slici 38. prikazana je korelacijska matrica koja sadrži procijenjene vrijednosti korelacijskog koeficijenta za svaki par varijabli. Osim procijenjenih vrijednosti korelacijskog koeficijenta može se i dokazati značajnost pojedinih koeficijenta korelacije.

Grafički prikaz 37.

Korelacijski model



Izvor: Prikaz autora u alatu Rapid Miner Studio

Grafički prikaz 38.

Korelacijska matrica OP „Konkurentnost i kohezija“

Attributes	Mjeseci trajanja operacije	Bespovratna sredstva EU - HRK	EU stopa sufinanciranja u %
Mjeseci trajanja operacije	1	0.292	-0.017
Bespovratna sredstva EU - HRK	0.292	1	0.003
EU stopa sufinanciranja u %	-0.017	0.003	1

Izvor: Prikaz autora u alatu RapidMiner Studio

Pearsonov koeficijent korelacije između varijabli iznosa bespovratnih sredstava EU i vremenskog razdoblja trajanja projekta unutar OP „Konkurentnost i kohezija“ iznosi $r=0,292$ što znači da postoji pozitivna i relativno slaba povezanost. Postojanje povezanosti upućuje na to da se s povećanjem iznosa bespovratnih sredstava EU projekta, pretpostavlja i duže vremensko razdoblje trajanja projekta i obrnuto s dužim vremenskim razdobljem trajanja projekta, očekuje se i veći iznos bespovratnih sredstava EU nekog projekta. Isto, ako je kraće vremensko razdoblje projekta, očekuju se i manji iznosi bespovratnih sredstava EU projekta i obrnuto.

Nakon izračuna koeficijenta korelacije, izračunava se njegova signifikantnost te se zaključuje da empirijska signifikantnost koeficijenta korelacije iznosi $\alpha^* \approx 0\%$ što je manje od 1% i znači da je koeficijent korelacije

statistički značajan. Koeficijent korelacije ukazuje na postojanje smjera i intenziteta povezanosti između dviju varijabli, ali ne ukazuje na uzročno-posljedičnu povezanost.

Grafički prikaz 39.

Korelacijska matrica OP „Učinkoviti ljudski potencijali“

Attributes	Mjeseci trajanja operacije	Bespovratna sredstva EU - HRK	EU stopa sufinanciranja u %
Mjeseci trajanja operacije	1	0.242	0.087
Bespovratna sredstva EU - HRK	0.242	1	0.665
EU stopa sufinanciranja u %	0.087	0.665	1

Izvor: Prikaz autora u alatu RapidMiner Studio

Na grafičkom prikazu 39. prikazana je korelacijska matrica OP „Učinkoviti ljudski potencijali“. Pearsonov koeficijent korelacije između varijabli iznosa bespovratnih sredstava EU i vremenskog razdoblja projekta iznosi $r=0,242$ što znači da postoji pozitivna i relativno slaba povezanost. Postojanje povezanosti upućuje na to da se s povećanjem iznosa bespovratnih sredstava EU projekta, očekuje i duže vremensko razdoblje trajanja projekta i obrnuto. Nakon izračuna koeficijenta korelacije, izračunava se njegova signifikantnost te se zaključuje da empirijska signifikantnost koeficijenta korelacije iznosi $\alpha^* \approx 0\%$ što je manje od 1% i znači da je koeficijent korelacije statistički značajan.

Pearsonov koeficijent korelacije iznosi $r=0,665$ za varijable EU stope sufinanciranja u % i bespovratnih sredstava EU što znači da postoji pozitivna i srednje jaka povezanost. Postojanje povezanosti upućuje na to da se s povećanjem iznosa bespovratnih sredstava EU projekta, očekuje i povećanje EU stope sufinanciranja projekta i obrnuto. Empirijska signifikantnost koeficijenta korelacije iznosi $\alpha^* \approx 0\%$ što je manje od 1% i znači da je koeficijent korelacije statistički značajan.

S obzirom da svi projekti OP „Konkurentnost i kohezija“ imaju EU stopu sufinanciranja u iznosu od 85%, nema smisla istraživati korelaciju između iznosa bespovratnih sredstava EU i stope sufinanciranja jer će svaka promjena varijable iznosa bespovratnih sredstava EU, njeno smanjenje ili povećanje, imati stopu sufinanciranja u iznosu od 85% odnosno i kod projekata s većim iznosom bespovratnih sredstava EU ne bi došlo do povećanja stope sufinanciranja, čime se može pretpostaviti da će se većem projektu, čija je veličina određena ukupnom vrijednošću projekta, dodijeliti i veći iznos bespovratnih sredstava EU. Tako, na grafičkom prikazu 32., dan je prikaz korelacijske matrice sa varijablama iznosa bespovratnih sredstava EU i najmanjih i najvećih ukupnih vrijednosti projekata kako bi se istražilo postojanje povezanosti. Set

podataka sadrži 9615 opažanja, a projekti veličine od 0-50 milijuna kuna čine 9347 opažanja odnosno 97% seta podataka. Broj velikih projekata je jako mali, ali se izdvajaju sa svojom ukupnom vrijednošću projekta i iznosima bespovratnih sredstava EU projekta pa su u obzir uzeti samo oni najveći projekti čija je ukupna vrijednost od jedne milijarde kuna i više. U OPKK-u ima 7223 projekata čija se vrijednost kreće od 0-50 milijuna kuna, a u OPULJP-u se nalaze 2123 projekta čija se vrijednost kreće od 0-50 milijuna kuna.

Grafički prikaz 40.

Korelacijska matrica OP „Konkurentnost i kohezija“ i OP „Učinkoviti ljudski potencijali (*dummy varijable*)

Attributes	Interv. Vrije. Projekt = 0-50	Interv. Vrije. Projekt = 1000-...	Bespovratna sredstva EU - HRK
Interv. Vrije. Projekt = 0-50	1	-0.200	-0.539
Interv. Vrije. Projekt = 1000-...	-0.200	1	0.783
Bespovratna sredstva EU - ...	-0.539	0.783	1

Izvor: Prikaz autora u alatu RapidMiner Studio

Pearsonov koeficijent korelacije između varijabli iznosa bespovratnih sredstava EU i ukupne vrijednosti projekata od 0-50 milijuna kuna iznosi $r=-0,539$ što znači da postoji negativna i srednje jaka povezanost. Postojanje povezanosti upućuje na to da se s većim iznosom bespovratnih sredstava EU projekta, očekuje i manja ukupna vrijednost projekata od 0-50 milijuna i obrnuto da se s manjim iznosom bespovratnih sredstava EU, očekuje i povećanje ukupne vrijednosti projekata od 0-50 milijuna kuna. Dobiveni koeficijent korelacije može se protumačiti da su manji projekti financirani manjim iznosom bespovratnih sredstava EU te da će s povećanjem iznosa bespovratnih sredstava EU, biti manje financiranih projekata čija je ukupna vrijednost od 0-50 milijuna kuna i obrnuto. Tako, sa smanjenjem iznosa bespovratnih sredstava EU, očekuje se više financiranih projekata čija je ukupna vrijednost od 0-50 milijuna kuna. Nakon izračuna koeficijenta korelacije, izračunava se njegova signifikantnost te se zaključuje da empirijska signifikantnost koeficijenta korelacije iznosi $\alpha^*\approx 0\%$ što je manje od 1% i znači da je koeficijent korelacije statistički značajan.

Pearsonov koeficijent korelacije za varijable bespovratna sredstva EU i ukupne vrijednosti projekata od milijardu kuna i više iznosi $r=0,783$ što znači da postoji srednje jaka pozitivna korelacija. Postojanje povezanosti upućuje na to da se s većim iznosom bespovratnih sredstava EU projekta, očekuje i povećanje ukupne vrijednosti projekata čija je vrijednost veća od milijardu kuna i obrnuto. Tako, sa manjim iznosom bespovratnih sredstava EU, pretpostavlja se i smanjenje ukupne vrijednosti projekata čija je vrijednost veća od milijardu kuna. Nakon izračuna koeficijenta korelacije, izračunava se njegova signifikantnost te se

zaključuje da empirijska signifikantnost koeficijenta korelacije iznosi $\alpha^* \approx 0\%$ što je manje od 1% i znači da je koeficijent korelacije statistički značajan.

U setu podataka prisutni su projekti koji se izdvajaju od drugih ili po vremenskom razdoblju trajanja projekta ili po vrijednosti bespovratnih sredstava EU, primjera radi mogu se izdvojiti projekti čija vrijednost bespovratnih sredstava EU iznosi više od 2,5 milijardi kuna, a vremensko razdoblje trajanja projekta je godinu dana, a čiji je korisnik Hrvatska agencija za malo gospodarstvo, inovacije i investicije. Operacija projekta se odnosi na COVID-19 zajam. Takav prisutan outlier može utjecati na rezultat analize, ali s obzirom da se radi o istinitom outlieru, ne postoji opravdan razlog za njegovo odbacivanje iz seta podataka.

Sljedeća tehnika koja će se primijeniti u analizi podataka je ANOVA koja kombinira nominalne varijable s numeričkim varijablama te se smatra prikladnom za dani set podataka u kojem dominiraju nominalne varijable.

Na slici 41. prikazana je ANOVA matrica numeričkih i nominalnih varijabli seta podataka iz koje je vidljiva njihova signifikantnost. Tako, može se zaključiti da postoji statistički značajna razlika za sve numeričke varijable s obzirom na odgovarajuće faktore osim za EU stopu sufinanciranja projekata koja se može tumačiti na način da ne postoji statistički značajna razlika projekata prema prosječnoj EU stopi sufinanciranja s obzirom na županiju projekta, NUTS regiju, tip korisnika, prioritetnu os i slično odnosno da se ne može odbaciti nulta hipoteza.

Grafički prikaz 41.

ANOVA matrica OP „Konkurentnost i kohezija“

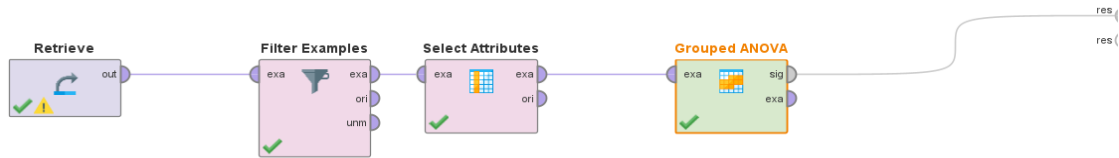
ANOVA Attribute	group Županija projekta	group NUTS regija	group Korisnici	group Prioritetna os	group Fond
Mjeseci trajanja operacije	0.000	0.002	0.000	0.000	0.000
Bespovratna sredstva EU - HRK	0.000	0.023	0.000	0.000	0.000
EU stopa sufinanciranja u %	0.929	0.219	0.805	0.979	0.273

Izvor: Prikaz autora u alatu RapidMiner Studio

Isto, uz pomoć F-testa može se ispitati pretpostavka da vrsta prioriteta financiranja ne djeluje značajno na prosječna bespovratna sredstva EU projekta.

Grafički prikaz 42.

ANOVA model



Izvor: Prikaz autora u alatu RapidMiner Studio

Grafički prikaz 43.

ANOVA rezultati OP „Konkurentnost i kohezija“ (iznos bespovratnih sredstava EU i prioriteta financiranja)

Source	Square Sums	DF	Mean Squares	F	Prob
Between	94491796724006865000.000	9	1049908852488962690.000	230.445	0.000
Residuals	33933155435736070000.000	7448	4556009054207313.000		
Total	43382335108136730000.000	7457			

Probability for random values with the same result: 0.000
 Difference between actual mean values is probably significant, since 0.000 < alpha = 0.050

Izvor: Prikaz autora u alatu RapidMiner Studio

Testiranje značajnosti djelovanja faktora može se prikazati F-testom. S obzirom da je empirijska vrijednost $F^*=230,445$ veća od tablične vrijednosti $F_{(9;7448)}^{0,05}=1,831$ pri razini signifikantnosti od 5%, prihvaća se pretpostavka da je djelovanje vrste prioriteta financiranja na bespovratna sredstva EU statistički značajno odnosno da postoji statistički značajna razlika projekata prema prosječnim bespovratnim sredstvima EU s obzirom na prioritete financiranja.

Na slici 44. prikazana je ANOVA matrica OP „Učinkoviti ljudski potencijali“ iz koje je vidljiva signifikantnost varijabli. Tako, može se zaključiti da ne postoji statistički značajna razlika projekata prema prosječnoj EU stopi sufinanciranja s obzirom na županiju projekta, tip korisnika, NUTS regiju te da ne postoji statistički značajna razlika projekata prema prosječnom vremenskom razdoblju trajanja projekta s obzirom na NUTS regiju odnosno ne može se odbaciti nulta hipoteza.

Grafički prikaz 44.

ANOVA matrica OP „Učinkoviti ljudski potencijali“

ANOVA Attribute	group Županija projekta	group NUTS regija	group Korisnici	group Prioritetna os	group Fond
Mjeseci trajanja operacije	0.028	0.263	0.000	0.000	0.000
Bespovratna sredstva EU - HRK	0.001	0.011	0.000	0.000	0.000
EU stopa sufinanciranja u %	0.997	0.329	0.814	0.000	0.000

Izvor: Prikaz autora u alatu RapidMiner Studio

Kao i u prethodnom primjeru, uz pomoć F-testa može se ispitati pretpostavka da vrsta prioriteta financiranja ne djeluje značajno na prosječna bespovratna sredstva EU projekta.

Grafički prikaz 45.

ANOVA rezultati OP „Učinkoviti ljudski potencijali“ (iznos bespovratnih sredstava EU i prioriteta financiranja)

Source	Square Sums	DF	Mean Squares	F	Prob
Between	17935555731121632.000	4	44838888932780408.000	40.110	0.000
Residuals	2405717951995568100.000	2152	1117898676577866.200		
Total	2585073507726688800.000	2156			

Probability for random values with the same result: 0.000

Difference between actual mean values is probably significant, since 0.000 < alpha = 0.050

Izvor: Prikaz autora u alatu RapidMiner Studio

Testiranje značajnosti djelovanja faktora može se prikazati F-testom. S obzirom da je empirijska vrijednost $F^*=40,110$ veća od tablične vrijednosti $F_{(4;2152)}^{0,05}=2,372$ pri razini signifikantnosti od 5%, prihvaća se pretpostavka da je djelovanje vrste prioriteta financiranja na bespovratna sredstva EU statistički značajno odnosno da postoji statistički značajna razlika projekata prema prosječnim bespovratnim sredstvima EU s obzirom na prioritete financiranja.

Naime, prioriteta financiranja određuju se unutar nekog Operativnog programa te se sa svakim ugovorenim projektom ostvaruju ciljevi kohezijske politike za određeno financijsko razdoblje. S obzirom na pripadnost prioriteta financiranja određenom Operativnom programu može se, primjera radi, ispitati ovisnost prioriteta financiranja o operativnom programu Hi-kvadrat testom s obzirom da se radi o dvije nominalne varijable. Rezultati Hi-kvadrat testa prikazani su u IBM SPSS Statistics alatu jer RapidMiner u korištenoj verziji ne nudi tu mogućnost odnosno potrebna je posebna licenca.

Grafički prikaz 46.

Rezultat Hi-kvadrat testa ovisnosti prioriteta financiranja o operativnom programu

Chi-Square Tests			
	Value	df	Asymptotic Significance (2-sided)
Pearson Chi-Square	9491,574 ^a	13	,000
Likelihood Ratio	10103,195	13	,000
N of Valid Cases	9615		

a. 0 cells (0,0%) have expected count less than 5. The minimum expected count is 8,52.

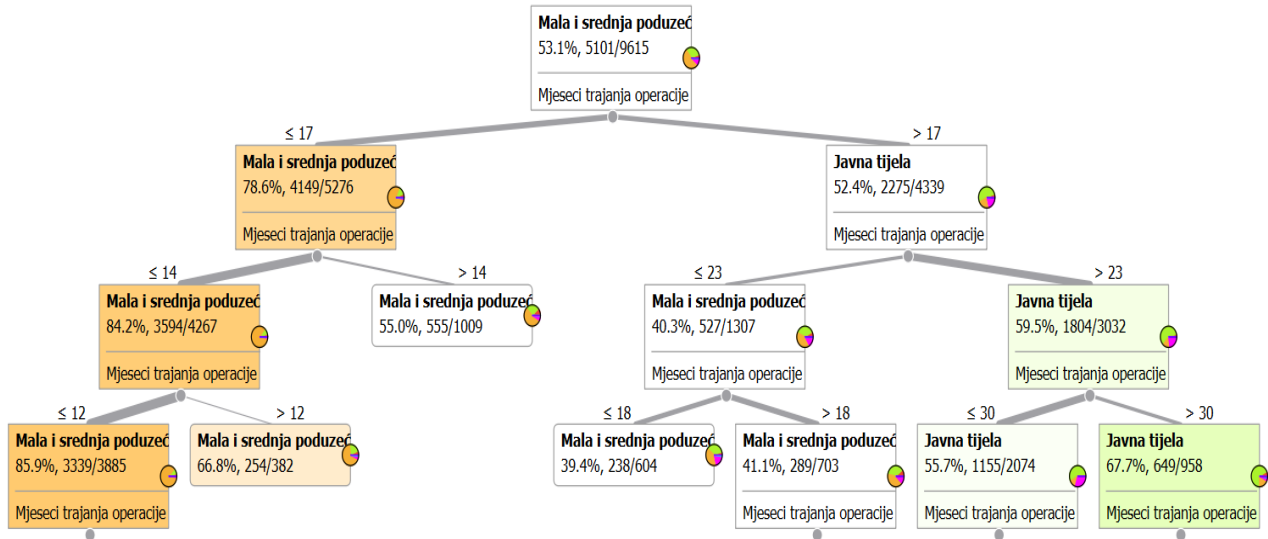
Izvor: Prikaz autora u IBM SPSS Statistics alatu

S obzirom na dobivene rezultate, na razini signifikantnosti od 5%, prihvaća se pretpostavka da postoji statistički značajna ovisnost između prioriteta financiranja i Operativnih programa te se nul hipoteza odbacuje ($\chi^2 = 9491,574, df = 13, p = 0,00$).

Sljedeća tehnika koja će se primijeniti na odabranom setu podataka je „stablo odlučivanja“ koje kategorizira varijable prema sličnim karakteristikama i odlukama.

Grafički prikaz 47.

Stablo odlučivanja prema vremenskom razdoblju trajanja projekta ovisno o tipu korisnika

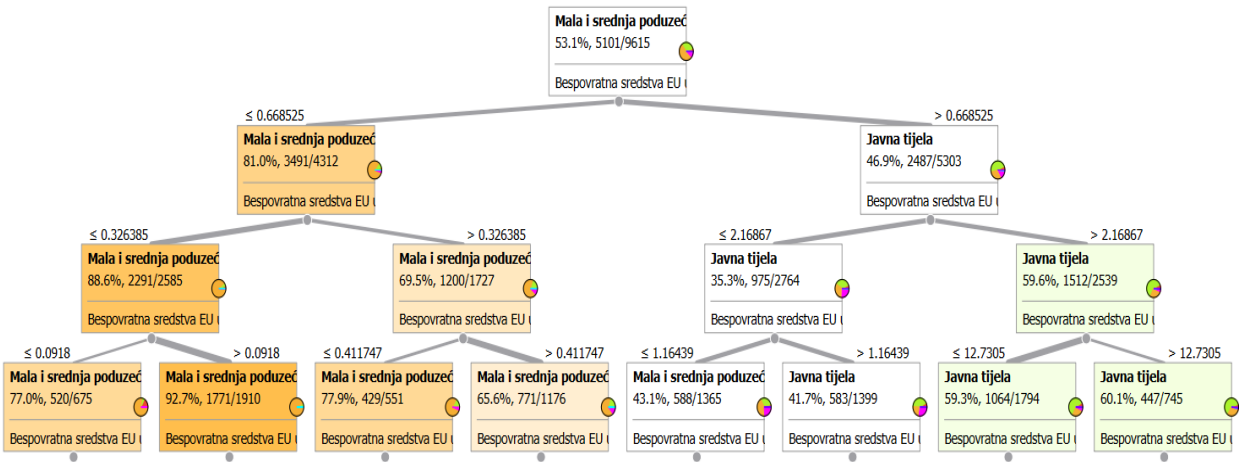


Izvor: Prikaz autora u Orange alatu

Korištenjem metode stabla odlučivanja, uzimajući u obzir varijable „vremensko razdoblje trajanja projekta“ i „korisnik“ iz dobivenih rezultata može se zaključiti da su se projekti podijelili na projekte čije je vremensko razdoblje trajanja kraće od 17 mjeseci i na projekte čije je vremensko razdoblje trajanja duže od 17 mjeseci. Kod svih projekata čije je vremensko razdoblje trajanja projekta kraće od 17 mjeseci prevladavaju mala i srednja poduzeća (78,6%), dok kod svih projekata čije je vremensko razdoblje trajanja duže od 17 i duže od 23 mjeseci prevladavaju javna tijela (59,5%). Tako i za projekte čije je vremensko razdoblje trajanja duže od 23 mjeseci i duže od 30 mjeseci prevladavaju projekti Javnih tijela (67,7%), dok kod projekata čije je vremensko razdoblje trajanja duže od 17 mjeseci, a kraće od 23 mjeseci prevladavaju mala i srednja poduzeća (40,3%).

Grafički prikaz 48.

Stablo odlučivanja prema vrijednosti bespovratnih sredstava EU ovisno o tipu korisnika

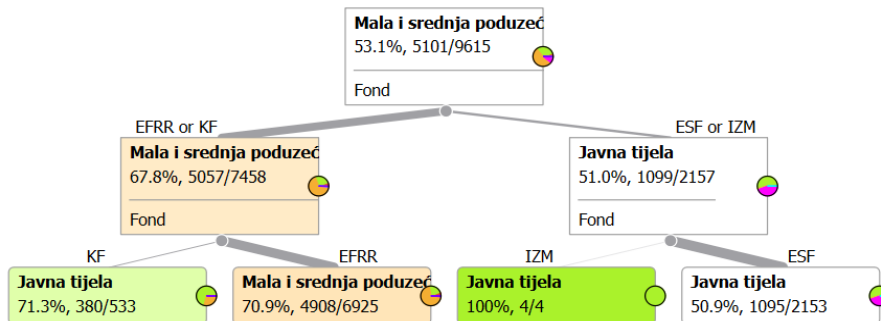


Izvor: Prikaz autora u Orange alatu

Korištenjem metode stabla odlučivanja, uzimajući u obzir varijable „iznos bespovratnih sredstava EU“ i „korisnik“ iz dobivenih rezultata može se zaključiti da su se projekti podijelili na projekte čiji je iznos bespovratnih sredstava manji ili je jednak 668 tisuća kuna i na projekte čiji je iznos bespovratnih sredstava veći od 668 tisuća kuna. Kod svih projekata čija je vrijednost bespovratnih sredstava EU manja ili jednaka 668 tisuća kuna prevladavaju mala i srednja poduzeća, dok kod svih projekata čija je vrijednost bespovratnih sredstava veća od 668 tisuća kuna prevladavaju javna tijela.

Grafički prikaz 49.

Stablo odlučivanja prema vrsti fonda ovisno o tipu korisnika

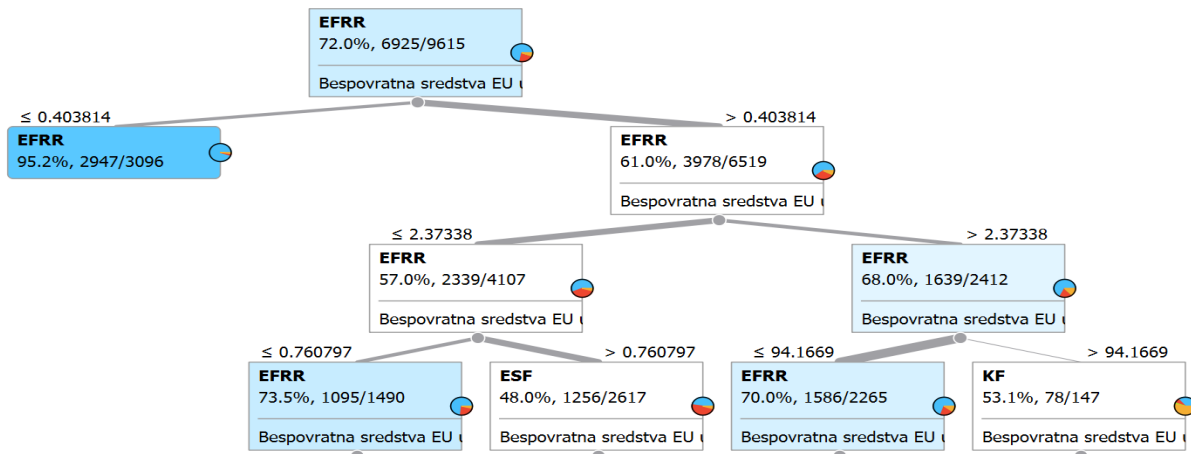


Izvor: Prikaz autora u Orange alatu

Korištenjem metode stabla odlučivanja, uzimajući u obzir varijable „fond“ i „korisnik“ iz dobivenih rezultata može se zaključiti da su se projekti podijelili na one koji se financiraju iz EFRR ili KF fonda (OPKK) i na one koji se financiraju iz ESF ili IZM fonda (OPULJP). Projekti koji se dijele na one koji se financiraju iz EFRR ili KF fonda prevladavaju mala i srednja poduzeća (67,8%), dok kod projekata koji se financiraju iz ESF ili IZM fonda prevladavaju Javna tijela (51%). Tako kod projekata koji se financiraju iz KF fonda prevladavaju Javna tijela (71,3%), dok za projekte koji se financiraju iz EFRR fonda prevladavaju mala i srednja poduzeća(70,9%).

Grafički prikaz 50.

Stablo odlučivanja prema vrijednosti bespovratnih sredstava EU ovisno o vrsti fonda



Izvor: Prikaz autora u Orange alatu

Korištenjem metode stabla odlučivanja, uzimajući u obzir varijable „vrsta fonda“ i „vrijednost bespovratnih sredstava EU“ iz dobivenih rezultata može se zaključiti kako su se projekti podijelili na one čija je vrijednost bespovratnih sredstava EU manja ili jednaka 404 tisuće kuna i na projekte čija je vrijednost bespovratnih sredstava EU veća od 404 tisuće kuna. Kod svih projekata čija je vrijednost bespovratnih sredstava EU manja ili jednaka 404 tisuće kuna prevladavaju projekti financirani iz EFRR fonda (95,2%). Tako, i kod svih projekata čija je vrijednost bespovratnih sredstava EU veća od 404 tisuće kuna prevladavaju projekti financirani iz EFRR fonda (61%). Ukoliko je vrijednost bespovratnih sredstava veća od 2,37 milijuna kuna, a manja od 94,17 milijuna kuna, prevladavaju projekti financirani iz EFRR fonda (70%), a ukoliko je vrijednost bespovratnih sredstava EU veća od 2,37 milijuna kuna i veća od 94,17 milijuna kuna projekti su financirani iz KF fonda (53,1%).

4.4 Osvrt na hipoteze

H1:...Postoji povezanost između dodijeljenih vrijednosti bespovratnih sredstava EU i vremenskog razdoblja trajanja projekta te EU stope sufinanciranja projekata namijenjenih provedbi kohezijske politike

H_{1a}:...Postoji povezanost iznosa dodijeljenih iznosa bespovratnih sredstava EU i vremenskog razdoblja trajanja projekta namijenjenih provedbi kohezijske politike.

Korelacijskom matricom „OPKK“ i „OPULJP“ između varijabli vrijednost bespovratnih sredstava EU i vremenskog razdoblja trajanja projekta izraženog u mjesecima istražuje se postojanje povezanosti Pearsonovim koeficijentom korelacije. Na temelju vrijednosti koeficijenta korelacije koji za OPKK iznosi $r=0,292$, a za OPULJP iznosi $r=0,242$ može se utvrditi postojanje slabe, pozitivne i statistički značajne povezanosti ($p=0,00$; $< 0,01$) između varijabli vrijednost bespovratnih sredstava EU i vremenskog razdoblja trajanja projekta. Postojanje povezanosti upućuje na to da se s povećanjem iznosa bespovratnih sredstava EU projekta, pretpostavlja i duže vremensko razdoblje trajanja projekta i obrnuto s dužim vremenskim razdobljem trajanja projekta, očekuje se i veća vrijednost bespovratnih sredstava EU nekog projekta.

Na temelju rezultata provedene korelacijske analize može se donijeti zaključak da se hipoteza H_{1a} prihvaća kao istinita.

H_{1b}:...Postoji povezanost vrijednosti dodijeljenih bespovratnih sredstava EU i stope EU sufinanciranja projekata namijenjenih provedbi kohezijske politike.

Korelaciju OP „Konkurentnost i kohezija“ između vrijednosti bespovratnih sredstava EU i stope sufinanciranja projekata nema smisla istraživati jer skoro svi projekti OPKK-a imaju istu stopu sufinanciranja što znači da će uz smanjenje ili povećanje vrijednosti bespovratnih sredstava, stopa sufinanciranja projekata iznositi 85%, a prema korelacijskoj matrici Pearsonov koeficijent korelacije iznosi $r=-0,017$ što znači da postoji negativna, slaba korelacija koja nije statistički značajna ($p=0,132$; $> 0,01$) pa će u obzir biti uzet Pearsonov koeficijent korelacije OP „Učinkoviti ljudski potencijali“ koji iznosi $r=0,665$ za varijable EU stope sufinanciranja u % i bespovratnih sredstava EU što znači da postoji pozitivna i srednje jaka povezanost koja je statistički značajna ($p=0,00$; $< 0,01$). Postojanje povezanosti upućuje na to da se s povećanjem vrijednosti bespovratnih sredstava EU projekta, očekuje i povećanje EU stope sufinanciranja projekta i obrnuto.

Na temelju rezultata provedene korelacijske analize prihvatit će se hipoteza H1b kao istinita jer OPULJP program je namijenjen provedbi kohezijske politike čija je povezanost EU stope sufinanciranja i vrijednosti bespovratnih sredstava ipak statistički značajna.

H2:...Ugovoreni projekti se značajno razlikuju prema prosječnom iznosu bespovratnih sredstava EU s obzirom na prioritete financiranja projekata namijenjenih provedbi kohezijske politike.

H_{1a}:...Ugovoreni projekti se značajno razlikuju prema prosječnom iznosu bespovratnih sredstava EU s obzirom na prioritete financiranja OP „Konkurentnost i kohezija“

Testiranje značajnosti djelovanja faktora prikazano je F-testom (str.75). S obzirom da je empirijska vrijednost $F^*=230,445$ veća od tablične vrijednosti $F_{(9;7448)}^{0,05}=1,831$ pri razini signifikantnosti od 5%, prihvaća se pretpostavka da je djelovanje vrste prioriteta financiranja na bespovratna sredstva EU statistički značajno odnosno da postoji statistički značajna razlika projekata prema prosječnim bespovratnim sredstvima EU s obzirom na prioritete financiranja te se hipoteza H_{1a} prihvaća kao istinita.

H_{1b}:...Ugovoreni projekti se značajno razlikuju prema prosječnom iznosu bespovratnih sredstava EU s obzirom na prioritete financiranja OP „Učinkoviti ljudski potencijali“.

Testiranje značajnosti djelovanja faktora može se prikazati F-testom (str.76). S obzirom da je empirijska vrijednost $F^*=40,110$ veća od tablične vrijednosti $F_{(4;2152)}^{0,05}=2,372$ pri razini signifikantnosti od 5%, prihvaća se pretpostavka da je djelovanje vrste prioriteta financiranja na bespovratna sredstva EU statistički značajno odnosno da postoji statistički značajna razlika projekata prema prosječnim bespovratnim sredstvima EU s obzirom na prioritete financiranja te se hipoteza H_{1b} prihvaća kao istinita.

H3:...Ugovorena vrijednost bespovratnih sredstava EU i vremensko razdoblje trajanja projekta namijenjenih provedbi kohezijske politike imaju utjecaj na tip korisnika.

H_{1a}:... Ugovorena vrijednost bespovratnih sredstava EU namijenjenih provedbi kohezijske politike ima utjecaj na tip korisnika (str.78)

S obzirom na rezultate metode stabla odlučivanja, može se zaključiti da vrijednost bespovratnih sredstava utječe na tip korisnika. Svi projekti čije je vrijednost bespovratnih sredstava EU:

- manje ili jednako 668 tisuća kuna - mala i srednja poduzeća (81,00%)
- veće od 668 tisuća kuna – Javna tijela (46,9%)
- veće od 668 tisuća kuna i veće od 2,1 milijuna kuna – Javna tijela (59,6%)

- manje od 2,1 milijuna kuna i manje ili jednako 1,16 milijuna kuna – mala i srednja poduzeća (43,1%)

Prihvaća se H_{1a} kao istinita.

H_{1b}:... Vremensko razdoblje trajanja projekata namijenjenih provedbi kohezijske politike ima utjecaj na tip korisnika (str.77)

S obzirom na rezultate metode stabla odlučivanja, može se zaključiti da vremensko razdoblje trajanja projekta utječe na tip korisnika. Svi projekti čije je vremensko razdoblje:

- Kraće od 17 mjeseci – mala i srednja poduzeća (78,6%)
- Duže od 17 mjeseci, a manje od 23 mjeseca – Javna tijela (59,5%)
- Duže od 23 mjeseca i duže od 30 mjeseci – Javna tijela (67,7%)
- Duže od 17 mjeseci, a kraće od 23 mjeseca – mala i srednja poduzeća (40,3%).

Prihvaća se H_{1b} kao istinita.

H₄:...Ugovorena vrijednost bespovratnih sredstava EU ima utjecaj na vrstu fonda iz kojeg se projekt financira.

S obzirom na rezultate metode stabla odlučivanja, može se zaključiti da ugovorena vrijednost bespovratnih sredstava EU utječe na vrstu fonda. Svi projekti čija je vrijednost bespovratnih sredstava:

- Manja ili jednaka 404 tisuće kuna – EFRR fond (95,2%)
- Veća od 404 tisuće kuna – EFRR fond (61%)
- Veća od 2,3 milijuna kuna, a manja od 94,17 milijuna kuna – EFRR fond (70%)
- Veće od 2,37 milijuna kuna i veće od 94,17 milijuna kuna – KF fond (53,1%).

Prihvaća se H₄ hipoteza kao istinita.

5. ZAKLJUČAK

Analizom i rudarenjem podataka projekata namijenjenih provedbi kohezijske politike dolazi se do korisnih informacija identificiranjem skrivenih uzoraka. Analizirani set podataka sadrži outliere, koji se otkrivaju metodom klasteriranja, a time i se otkriva i struktura podataka. Jedan od klastera sadrži samo velike projekte čija se ukupna vrijednost projekta iznosi od 1,4 milijarde kuna do skoro 4 milijarde kuna. Projekti tog klastera odnose se na poticanje rasta i razvoja ESIF kreditima, COVID-19 zajam namijenjen subjektima malog gospodarstva za suzbijanje štetnih posljedica nastalih korona virusom te projekti koji se odnose na poboljšanje cestovne, željezničke i vodno-komunalne infrastrukture. U drugom klasteru nalaze se projekti čija je vrijednost od 220 milijuna kuna do 1,4 milijarde kuna, dok se u klasteru koji ima najveći broj projekata manje od 220 milijuna ističu projekti namijenjeni Poslovnoj konkurentnosti i da je najviše korisnika koji se svrstavaju pod „mala i srednja poduzeća“.

Isto, iz analize podataka se uočava da se najveći broj projekata i s najvećim iznosom sufinanciranja nalazi u Gradu Zagrebu, a da odstupanje koje se pojavljuje u iznosima sufinanciranja i broju projekata u Jadranskoj regiji se odnosi na velike infrastrukturne projekte poboljšanja vodno-komunalne infrastrukture, cestovne povezanosti, izgradnju zračne luke, centra za gospodarenje otpadom i slično odnosno projekti kojima se poboljšava kohezija zemlje. Nije neobično što Grad Zagreb dominira u apsorpciji bespovratnih sredstava iz ESI fondova jer je Zagreb monocentrično središte Hrvatske.

Korelacijom, kao metodom rudarenja podataka, stekao se uvid u povezanost varijabli u setu podataka i kako se s promjenom jedne varijable, očekuje i promjena druge varijable, npr. povećanjem vrijednosti bespovratnih sredstava EU, može se očekivati i duže vremensko razdoblje trajanja projekta. Treća primijenjena metoda je ANOVA kojom je stečen uvid o značajnoj razlici projekata prema prosječnom iznosu bespovratnih sredstava EU s obzirom na prioritete financiranja. S obzirom da svaki prioritet financiranja odgovara određenom Operativnom programu, može se očekivati da će se prema nekim prioritetima financiranja kao što je „Povezanost i mobilnost“ financirati veći projekti koji zahtijevaju veće iznose bespovratnih sredstava. Posljednja korištena metoda je stabla odlučivanja kojom se stekao uvid o utjecaju bespovratnih sredstava EU i vremenskog razdoblja trajanja projekta na vrstu fonda i tip korisnika pa tako ukoliko se radi o projektima čija je vrijednost bespovratnih sredstava manja od 404 tisuće kuna, bit će financirani iz EFRR fonda, a ukoliko se radi o projektima većim od 2,37 milijuna kuna i većim od 94,17 milijuna kuna bit će financirani iz Kohezijskog fonda. Ukoliko se radi o projektima koji su kraći od 17 mjeseci i duži od 17 mjeseci, ali kraći od 23 mjeseca korisnik će biti mala i srednja poduzeća. Ukoliko je trajanje

projekta duže od 17 mjeseci, a kraćim od 23 mjeseca i dužim od 30 mjeseci korisnik će biti Javna tijela. S obzirom na dobivene informacije može se zaključiti da je potrebno informirati i educirati javnost o natječajima projekata kako bi se povećao broj sufinanciranih projekata, alocirati sredstva ESI fondova prema manje razvijenim područjima od Grada Zagreba, prilagoditi prioritete financiranja i podjelu na NUTS regije kako bi sredstva postala dostupnija i drugim regijama te i dalje poticati rast i razvoj malih i srednjih poduzeća kroz prioritet „Poslovna konkurentnost“ jer mala i srednja poduzeća stvaraju rast i razvoj gospodarstva. Isto, i dalje ulagati infrastrukturni razvoj Hrvatske kroz Kohezijski fond te ulagati u obrazovanje i cjeloživotno učenje čime se otvaraju mogućnosti za rast i razvoj vještina i sposobnosti stanovništva.

Vizualizacija podataka omogućava slikovit i jednostavan način prikazivanja rezultata istraživanja koji je prihvatljiv korisniku. Može se reći i da uspješnost provođenja rudarenja podataka ovisi o području primjene jer nije svaka metoda rudarenja podataka primjenjiva na svakom setu podataka. Pravilno provedena analiza može biti od strateške važnosti organizaciji jer otkrivanjem pravih informacija iz podataka, donose se informirane odluke te omogućuje ostvarivanje konkurentske prednosti na tržištu. S obzirom na dinamičnost današnjeg tržišta, te brzi razvoj tehnologije, organizacije trebaju biti konstantno pratiti trendove na tržištu kako bi mogle stvarati i održavati konkurentsku prednost, a poslovna inteligencija predstavlja izrazito bitan čimbenik poslovnog uspjeha.

LITERATURA

A.J. Chamatkar, P.K. Butey (2014); Importance of Data Mining with Different Types of Data Applications and Challenging Areas, Int. Journal of Engineering Research and Applications [Internet], 4 (5), 38-41, raspoloživo na: [https://www.academia.edu/7675788/Importance of Data Mining with Different Types of Data Applications and Challenging Areas](https://www.academia.edu/7675788/Importance_of_Data_Mining_with_Different_Types_of_Data_Applications_and_Challenging_Areas)

Adrians P. Zantinge D. (1996): Data mining, Addison-Weasley Longman, England, Harlow

Aggarwal., C., C. (2015): Data Mining: The textbook. New York; Springer International Publishing [Internet], raspoloživo na: https://doc.lagout.org/Others/Data%20Mining/Data%20Mining_%20The%20Textbook%20%5BAggarwal%202015-04-14%5D.pdf

Allemar, J.P.D. (2019): Predicting Scholarship Grants Using Data Mining techniques, International Journal of Machine Learning and Computing, Vol. 9, No.4 [Internet], raspoloživo na: <https://pdfs.semanticscholar.org/fc56/0233126e175716605d722eb70b8508ee9f5d.pdf>

Alqallaf F., Konis K., Martin R., Zamar R. (2002): Scalable Robust Covariance and Correlation Estimates for Data Mining, Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, [Internet], raspoloživo na: https://www.researchgate.net/publication/2565827_Scalable_Robust_Covariance_and_Correlation_Estimates_for_Data_Mining

Apolonio P.(2018): Klasteriranje k-sredinama, Sveučilište u Zagrebu, Prirodoslovno-matematički fakultet, Zagreb [Internet], raspoloživo na: <https://repositorij.unizg.hr/islandora/object/pmf%3A4641/datastream/PDF/view>

Arnerić J. (2012): Vježbe iz kolegija: Statistička analiza, Ekonomski fakultet, Split

Azevedo A., Santos M. (2008): KDD, SEMMA AND CRISP-DM: A PARALLEL OVERVIEW [Internet], 182-185., raspoloživo na: <https://recipp.ipp.pt/bitstream/10400.22/136/3/KDD-CRISP-SEMMA.pdf>

Brown, M. S. (2014): Business Understanding, [Internet], raspoloživo na: <http://www.dummies.com/programming/big-data/phase-1-of-the-crisp-dm-process-model-business-understanding/>

Brown., S. M. (2014): Data Mining For Dummies. New Jersey; John Wiley & Sons, Inc.

Chamatkar., J., A., Butey., K., P. (2014): Importance of Data Mining with Different Types of Data Applications and Challenging Areas , Int. Journal of Engineering Research and Applications [Internet], 5 (3), 38-41, raspoloživo na: <https://core.ac.uk/download/pdf/27193111.pdf> i www.ijera.com

Coenen F. (2004); Data Mining: Past, Present and Future, The Knowledge Engineering Review, Cambridge University Press, Ujedinjeno Kraljevstvo [Internet], Vol. 00:0, 1–24, raspoloživo na: <https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.185.1202&rep=rep1&type=pdf>

Cohesion open data platform: 2014-2020 Cohesion Policy Overview, [Internet], raspoloživo na: <https://cohesiondata.ec.europa.eu/overview>

Đulabić V. (2014): Kohezijska politika Europske unije kao instrument razvoja Hrvatske, Političke analize, 5(17), str. 17-22. [Internet], raspoloživo na: [210286 \(srce.hr\)](http://210286.srce.hr)

Europski strukturni i investicijski fondovi (2014): Operativni program Učinkoviti ljudski potencijali 2014.-2020., Zagreb [Internet], raspoloživo na: [OPULJP_hr-1.pdf \(strukturnifondovi.hr\)](http://OPULJP_hr-1.pdf)

Europski strukturni i investicijski fondovi (2021): Operativni program Konkurentnost i kohezija 2014.-2020., Zagreb [Internet], raspoloživo na: https://strukturnifondovi.hr/wp-content/uploads/2021/12/Izmjene_OPKK_lipanj-2021.pdf

Hrvatska gospodarska komora (2013): Ususret fondovima kohezijske politike u Hrvatskoj: Kohezijska politika, Strukturni fondovi, Kohezijski fond i primjeri [Internet], raspoloživo na: https://www.safu.hr/datastore/filestore/10/Ususret_fondovima_Kohezijske_politike_HGK.pdf

Hrvatska gospodarska komora (2015): Kohezijska politike Europske unije i Hrvatska 2014.-2020.: Vodič kroz strateški okvir i pregled mogućnosti financiranja [Internet], raspoloživo na: <https://izvoz.gov.hr/UserDocsImages/dokumenti/Kohezijska%20politika%20EU%20i%20Hrvatska%202014.-2020..pdf>

Huang, J.Z. (1997): A Fast Clustering Algorithm to Cluster Very Large Categorical Data Sets in Data Mining. Workshop on Research Issues on Data Mining and Knowledge Discovery, [Internet], raspoloživo na: http://www.diag.uniroma1.it/~sassano/STAGE/Fast_Clustering.pdf

Jourdan, Z., Rainer R. K., Marshall T. E.(), Business Intelligence: An Analysis of the Literature, Taylor & Francis LTD [Internet], raspoloživo na: https://www.researchgate.net/profile/Zack-Jourdan/publication/220630374_Business_Intelligence_An_Analysis_of_the_Literature_1/links/5547bf8f0cf2e2031b384867/Business-Intelligence-An-Analysis-of-the-Literature-1.pdf

K. P. Sinaga, M.-S. Yang (2020): "Unsupervised K-Means Clustering Algorithm," in IEEE Access, vol. 8, [Internet], raspoloživo na: <https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=9072123>

Kantardzic, M. (2020): Data Mining: Concepts, Models, Methods and Algorithms. New Jersey; John Wiley & Sons, Inc.

Kapetanović L. (2015): Primjena cluster analize u projektiranju proizvodnih sustava, Sveučilište u Zagrebu, Fakultet strojarstva i brodogradnje, Zagreb, [Internet], raspoloživo na: http://repositorij.fsb.hr/4772/1/Kapetanovic_2015_zavrzni_preddiplomski.pdf

Kashwan K.R., Velu C. (2013): Customer Segmentation Using Clustering and Data Mining Techniques, International Journal of Computer Theory and Engineering.[Internet], raspoloživo na: <https://www.researchgate.net/publication/271302240>

Khan, R.A. (2012): Business Intelligence: An integrated approach, Business Intelligence Journal [Internet], 5 (1), 64-70, raspoloživo na: [https://www.researchgate.net/profile/Rafi-Khan-4/publication/266489043 BUSINESS INTELLIGENCE AN INTEGRATED APPROACH/links/5b3c70424585150d23f6940e/BUSINESS-INTELLIGENCE-AN-INTEGRATED-APPROACH.pdf](https://www.researchgate.net/profile/Rafi-Khan-4/publication/266489043_BUSINESS_INTELLIGENCE_AN_INTEGRATED_APPROACH/links/5b3c70424585150d23f6940e/BUSINESS-INTELLIGENCE-AN-INTEGRATED-APPROACH.pdf)

Kidemet M.(2018): Primjena metode stablo odlučivanja, Sveučilište u Zagrebu, Fakultet organizacije i informatike, Varaždin [Internet], raspoloživo na: <https://core.ac.uk/download/pdf/197829835.pdf>

Kraljević, S., Staničić, O. (2020): PRIMJENA ALGORITAMA DUBINSKE ANALIZE PODATAKA I STROJNOG UČENJA ZA KLASIFIKACIJU I PREDIKCIJU U DRUŠTVENOM PODRUČJU. Polytechnic and design, 8 (1), 38-46. [Internet], raspoloživo na: <https://hrcak.srce.hr/242766>

Kumar D., Bhardwaj D. (2011); Rise of Data Mining: Current and Future Application Areas, International Journal of Computer Science Issues [Internet], 8 (5), 256-260, raspoloživo na: <https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.402.6760&rep=rep1&type=pdf#page=276>

Mirošević I. (2016): Algoritam k-sredina, *KoG*, [Internet] 20 (20), 91-98, raspoloživo na: <https://hrcak.srce.hr/file/256969>

Moss., L.T. i Atre, S. (2003) Business Intelligence Roadmap: The Complete Project Lifecycle for Decision-Support Applications. SAD; Addison Wesley

Mumine, K.K. (2017): An overview: The impact of data mining applications on various sectors, Technical Journal [Internet], 11 (3), 128-132, raspoloživo na: <https://hrcak.srce.hr/file/275281>

Nassar K. (2006): Application of data-mining to state transportation agencies' projects databases, IT con Vol 12. [Internet], str 139-148, raspoloživo na: https://itcon.org/papers/2007_8.content.06891.pdf

Nemčić J. (2017): Spektralne metode grupiranja podataka, Sveučilište J.J. Strossmayera u Osijeku, Osijek [Internet], raspoloživo na: <https://repozitorij.mathos.hr/islandora/object/mathos%3A197/datastream/PDF/view>

Perošević M. (2020): Stablo odlučivanja te primjena u ekonomiji; [Internet], raspoloživo na: <https://zir.nsk.hr/islandora/object/efos%3A4074/datastream/PDF/view>

Praadep R., Shubha S. (2010): A survey of clustering techniques, International Journal of Computer Applications Volume 7–No.12 [Internet], raspoloživo na: https://www.researchgate.net/profile/Pradeep-Rai-2/publication/49586251_A_Survey_of_Clustering_Techniques/links/58088c6f08ae63c48fec841d/A-Survey-of-Clustering-Techniques.pdf

Pratljačić S. (2019): Spektralna analiza grafa i klasteriranje., Sveučilište u Zagrebu, Fakultet elektrotehnike i računarstva, Zagreb, [Internet], raspoloživo na: https://www.fer.unizg.hr/download/repository/Spektralna_analiza_grafa_i_klasteriranje_Pratljacic.pdf

Pyle, D. (1999): Data Preparation for Data Mining, San Francisco, Morgan Kaufmann Publishers, Inc. [Internet], raspoloživo na: https://temida.si/~bojan/MPS/materials/Data_preparation_for_data_mining.pdf

Ranjan, J. (2005): Business Intelligence: Concepts, components, techniques and benefits, Journal of Theoretical and Applied Information Technology [Internet], 9 (1), 60-70, raspoloživo na: <http://www.jatit.org/volumes/research-papers/Vol9No1/9Vol9No1.pdf>

Rokach L., Maimon O., (2005): Clustering Methods. [Internet], raspoloživo na: <https://www.researchgate.net/publication/226748490>

Rozga A. (2006): Statistika za ekonomiste, Ekonomski fakultet, Split

Scheps., S. (2008): Business Intelligence For Dummies. Indianapolis; Wiley Publishing Inc.

Schollo., A., Kautz., K. (2010): Towards an understanding of Business Intelligence. ACIS 2010 Proceedings [Internet], Paper 86., raspoloživo na: <https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.862.99&rep=rep1&type=pdf>

Shafique, U. i Quaiser H. (2014): A Comparative Study of Data Mining Process Models (KDD, CRISP-DM and SEMMA), International Journal of Innovation and Scientific Research [Internet], Vol.12, 217-222,

raspoloživo

na: <https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.686.840&rep=rep1&type=pdf>

SmartVision (2016): What is the CRISP-DM methodology?, [Internet], raspoloživo na: <https://www.sv-europe.com/crisp-dm-methodology/#businessunderstanding>

Solarte. J. (2002): A Proposed Data Mining Methodology and its Application to Industrial Engineering, Master's Thesis, University of Tennessee, [Internet], raspoloživo na: https://trace.tennessee.edu/cgi/viewcontent.cgi?article=3549&context=utk_gradthes

Središnji državni ured za razvojnu strategiju i koordinaciju fondova Europske unije (2009): Europski fondovi za hrvatske projekte, potpora EU, PHARE 2005-0707-010101 [Internet], raspoloživo na: https://www.safu.hr/datastore/filestore/10/Europski_fondovi_za_hrvatske_projekte.pdf

Šućur J. (2021): Stabla odlučivanja; [Internet], raspoloživo na: <https://repositorij.mathos.hr/islandora/object/mathos%3A583/datastream/PDF/view>

Tuffery, S. (2011): Data Mining and Statistics for Decision Making; United Kingdom; John Wiley & Sons Ltd.

Ujević F.: Postupci i tehnike dubinske analize podataka, Kvalifikacijski rad [Internet], raspoloživo na: https://www.inf.uniri.hr/images/studiji/poslijediplomski/kvalifikacijski/Filip_Ujevic_-_Postupci_i_tehnike_dubinske_analize_podataka.pdf

Ungaro T. (2016): Klasterska analiza, Sveučilište u Zagrebu, Prirodoslovno-matematički fakultet, Zagreb [Internet], raspoloživo na: <https://core.ac.uk/download/pdf/197898517.pdf>

Uremović K. (2016): Statističke metode grupiranja u analizi podataka, Sveučilište u Zagrebu, Prirodoslovno-matematički fakultet, Zagreb [Internet], raspoloživo na <https://repositorij.pmf.unizg.hr/islandora/object/pmf%3A5556/datastream/PDF/view>

Wijaya C. Y. (2021): CRISP-DM Methodology For Your First Data Science Project [Internet], raspoloživo na: <https://towardsdatascience.com/crisp-dm-methodology-for-your-first-data-science-project-769f35e0346c>

Wirth R., Hipp J. (2000). CRISP-DM: Towards a standard process model for data mining. Proceedings of the 4th International Conference on the Practical Applications of Knowledge Discovery and Data Mining [Internet], raspoloživo na: <http://www.cs.unibo.it/~danilo.montesi/CBD/Beatriz/10.1.1.198.5133.pdf>

Youguo L., Haiyan W. (2012): A Clustering Method Based on K-Means Algorithm, Physics Procedia, [Internet] (25) 1104-1109, raspoloživo na: <https://core.ac.uk/download/pdf/82600940.pdf>

Zaidi M., Saeed Bushra A., Nukman I., Nukman. (2012): Experimental Data Mining Techniques (Using Multiple Statistical Methods), International Journal of Computer Science issues. 9. [Internet], raspoloživo na:
https://www.researchgate.net/publication/267801593_Experimental_Data_Mining_Techniques_Using_Multiple_Statistical_Methods

Zekić-Sušac M., Frajman-Jakšić A., Drvenkar N.(2009):Neuronske mreže i stabla odlučivanja za predviđanje uspješnosti studiranja [Internet], raspoloživo na: <https://hrcak.srce.hr/file/73924>

Zelenika, R. (2004) : Znanost o znanosti, Ekonomski fakultet u Rijeci

Žapčević S., Butala P. (2015): Otkrivanje znanja i metode rudarenja podataka u proizvodnim sistemima, 10th International Scientific Conference on Production Engineering DEVELOPMENT AND MODERNIZATION OF PRODUCTION [Internet], raspoloživo na:
<https://tfb.ba/repozitorij/2/eZbornik/eZbornik2015/Seid%20%C5%BDap%C4%8Devi%C4%87,%20Peter%20Butala%20-%20OTKRIVANJE%20ZNANJA%20I%20METODE%20RUDARENJA%20PODATAKA%20U%20PROIZVODNIM%20SISTEMIMA.pdf>

Žuvić-Butorac M. (2006): Usporedba nebrojčanih podataka, Acta medica Croatica [Internet], 60 (Suppl.1), 63-79, raspoloživo na:
https://www.researchgate.net/publication/7252666_Characteristics_of_categorical_data_analysis/link/58380e4d08aef00f3bf9e460/download

Terzić D. (2020): Metode grupiranja podataka, Sveučilište u Zagrebu, Fakultet organizacije i informatike, Varaždin [Internet], raspoloživo na: <https://repositorij.foi.unizg.hr/islandora/object/foi%3A6298/datastream/PDF/view>

POPIS SLIKA I TABLICA

Grafički prikaz 1.	7
Grafički prikaz 2.	9
Grafički prikaz 3.	10
Grafički prikaz 4.	11
Grafički prikaz 5.	12
Grafički prikaz 6.	13
Grafički prikaz 7.	13
Grafički prikaz 8.	14
Grafički prikaz 9.	23
Grafički prikaz 10.	36
Grafički prikaz 11.	37
Grafički prikaz 12.	38
Grafički prikaz 13.	40
Grafički prikaz 14.	45
Grafički prikaz 15.	46
Grafički prikaz 16.	47
Grafički prikaz 17.	48
Grafički prikaz 18.	50
Grafički prikaz 19.	51
Grafički prikaz 20.	52
Grafički prikaz 21.	54
Grafički prikaz 22.	56
Grafički prikaz 23.	57
Grafički prikaz 24.	58
Grafički prikaz 25.	60
Grafički prikaz 26.	61
Grafički prikaz 27.	63
Grafički prikaz 28.	64
Grafički prikaz 29.	66
Grafički prikaz 30.	66
Grafički prikaz 31.	67
Grafički prikaz 32.	68
Grafički prikaz 33.	68
Grafički prikaz 34.	69
Grafički prikaz 35.	69
Grafički prikaz 36.	70
Grafički prikaz 37.	71

Grafički prikaz 38.....	71
Grafički prikaz 39.....	72
Grafički prikaz 40.....	73
Grafički prikaz 41.....	74
Grafički prikaz 42.....	74
Grafički prikaz 43.....	75
Grafički prikaz 44.....	75
Grafički prikaz 45.....	76
Grafički prikaz 46.....	76
Grafički prikaz 47.....	77
Grafički prikaz 48.....	78
Grafički prikaz 49.....	78
Grafički prikaz 50.....	79

Tablica 1. NUTS regije i broj stanovnika.....	17
---	----

Tablica 2. Popis skupa podataka.....	42
--------------------------------------	----

SAŽETAK

Sastavni dio BI je rudarenje podataka. To je proces identificiranja korisnih i razumljivih uzoraka u podacima odnosno otkrivanja korisnih informacija uz pomoć alata za rudarenje podataka kojima se dobivene informacije stavljaju u praktičnu upotrebu. Pravilno provedena analiza može biti od strateške važnosti organizaciji jer otkrivanjem pravih informacija iz podataka, donose se informirane odluke te omogućuje ostvarivanje konkurentne prednosti na tržištu. Europska Unija, regionalna organizacija europskih država i jedina organizacija ove vrste na svijetu, nije iznimka. Ona osigurava financijska sredstva za niz projekata i programa s ostvarenjem ciljeva kohezijske politike na području zemalja članica. Financijska sredstva isplaćuju se iz ESI fondova prema određenim programima i njihovim prioritetnim ciljevima. U Hrvatskoj su to OP "Konkurentnost i kohezija 2014.-2020." i OP "Učinkoviti ljudski potencijali 2014.-2020." čiji setovi podataka o financiranim projektima namijenjenih provedbi kohezijske politike, provođeni na lokalnoj i regionalnoj razini, predstavljaju podlogu za provedbu tehnika rudarenja podataka u ovome radu. Istraživanje se temelji na primjenjivosti metoda i tehnika rudarenja podacima na odabranim setovima podataka koji su javno dostupnih na stranicama europskih strukturnih i investicijskih fondova. Tako, u istraživanju je teoretski obrađen pojam, koncept i tehnike rudarenja podataka te je dana analiza isplaćenih ESIF sredstava zemljama članicama Europske unije u financijskoj perspektivi 2014.-2020. Nadalje, opisana je kohezijska politika u Hrvatskoj pod koju spadaju i fondovi iz kojih se projekti financiraju, Operativni programi, NUTS raspodjela Hrvatske na Jadransku i Kontinentalnu Hrvatsku gdje je prema određenom

broju stanovnika za razvoj svake regije na raspolaganje stavljen određeni iznos financijskih sredstava i slično. U empirijskom dijelu rada, Microsoft Power BI alatom izvršena je vizualizacija podataka, a daljnjim analiziranjem seta podataka odabranim tehnikama rudarenja podataka za rudarenjem podataka omogućava se stjecanje uvida u odnose, logičnosti, pravilnosti ili bilo kakve strukture među podacima. Osim toga, prikazana je i primjenjivost alata rudarenja podataka na odabranom setu podataka te su doneseni zaključci o projektima namijenjenih provedbi kohezijske politike financiranih iz ESI fondova. Dobivene informacije o dodijeljenim sredstvima i raspodjeli alokacije projektima prema određenim atributima iz ESI fondova istraživanjem varijabli odabranog seta podataka mogu koristiti poboljšanju donošenja odluka sadašnjim i budućim korisnicima potpora EU, te svima onima koji ovo istraživanje pronađu korisnim.

SUMMARY

Data mining is an integral part of BI. It is the process of identifying useful and understandable patterns in data, or discovering useful information through data mining tools that put the obtained information into practical use. A well-conducted analysis can be of strategic importance to an organisation because by uncovering the right information from data, informed decisions are made and competitive advantage can be achieved in the market. The European Union, a regional organisation of European states and the only organisation of its kind in the world, is no exception. It provides funding for a range of projects and programmes to achieve the objectives of cohesion policy in the territory of the Member States. Financial resources shall be disbursed from the ESI Funds in accordance with the specific programmes and their priority objectives. In Croatia, these are the OP 'Competitiveness and Cohesion 2014.-2020' and the OP 'Effective Human Resources 2014.-2020.' whose datasets on funded projects for the implementation of cohesion policy, implemented at local and regional level, provide the basis for the implementation of data mining techniques in this research. The research shall be based on the applicability of data mining methods and techniques on selected datasets which are publicly available on the websites of the European Structural and Investment Funds. Therefore, in research a theoretical data mining term, concepts and techniques are elaborated and an analysis of disbursed ESIF funds to the member states of the European Union in the financial perspective of 2014.-2020. was given. Furthermore, it describes the cohesion policy in Croatia which includes the funds from which projects are financed, the Operational Programmes, the NUTS distribution of regions in Croatia into Adriatic and Continental Croatia where according to a certain

number of inhabitants a certain amount of financial resources is made available for the development of each region. In the empirical part of the research, the Microsoft Power BI tool performs a data visualization and further analysis of the data set by selected data mining techniques allows to gain insight into the relationships, logic, regularity or any structure between the data. In addition, the applicability of the data mining tool on the selected data set was presented and conclusions were drawn on the projects intended for the implementation of the cohesion policy financed from the ESI funds. The obtained information about the allocated funds and the distribution of the allocation to projects according to certain attributes from the ESI funds by researching the variables of the selected data set can be used to improve decision-making to current and future beneficiaries of EU grants and to all those who find this research useful.