

Analiza otvorenih podataka

Krznarić, Karlo

Undergraduate thesis / Završni rad

2019

Degree Grantor / Ustanova koja je dodijelila akademski / stručni stupanj: **University of Zagreb, Faculty of Organization and Informatics / Sveučilište u Zagrebu, Fakultet organizacije i informatike**

Permanent link / Trajna poveznica: <https://um.nsk.hr/um:nbn:hr:211:502156>

Rights / Prava: [Attribution-NoDerivs 3.0 Unported](#) / [Imenovanje-Bez prerada 3.0](#)

Download date / Datum preuzimanja: **2024-08-05**



Repository / Repozitorij:

[Faculty of Organization and Informatics - Digital Repository](#)



**SVEUČILIŠTE U ZAGREBU
FAKULTET ORGANIZACIJE I INFORMATIKE
VARAŽDIN**

Karlo Krznarić

ANALIZA OTVORENIH PODATAKA

ZAVRŠNI RAD

Varaždin, 2019.

SVEUČILIŠTE U ZAGREBU
FAKULTET ORGANIZACIJE I INFORMATIKE
V A R A Ž D I N

Karlo Krznarić

Matični broj: 45129/16 - R

Studij: Poslovni sustavi

ANALIZA OTVORENIH PODATAKA

ZAVRŠNI RAD

Mentorica:

Doc. dr. sc. Dijana Oreški

Varaždin, rujan 2019.

Karlo Krznarić

Izjava o izvornosti

Izjavljujem da je moj završni rad izvorni rezultat mojeg rada te da se u izradi istoga nisam koristio drugim izvorima osim onima koji su u njemu navedeni. Za izradu rada su korištene etički prikladne i prihvatljive metode i tehnike rada.

Autor/Autorica potvrdio/potvrdila prihvaćanjem odredbi u sustavu FOI-radovi

Sažetak

U ovom radu napravljen je pregled otvorenih podataka te je napravljena analiza skupa podataka sa sustava e-Građani. Teorijski dio rada objašnjava najvažnije pojmove vezane uz otvorene podatke, njihove prednosti i nedostatke te razvoj otvorenih podataka u Hrvatskoj. U poglavlju podatkovne analize predstavljeni su koncepti rudarenja podataka, proces rudarenja i tehnike koje se koriste za rudarenje podataka. Također, teorijski dio obuhvaća pregled prethodnih pretraživanja gdje je navedeno i opisano nekoliko različitih istraživanja vezanih uz otvorene podatke. Kod empirijskog dijela, napravljena je klaster analiza pomoću alata *BigML*. Za skup podataka odabrani su podaci o korištenju usluga sustava e-Građani. Odabrani podaci opisani su i napravljena je priprema podataka kako bi bili spremni za analizu. Analizom se željelo dobiti uvid u korištenje pojedinih usluga sustava e-Građani. U poglavlju modeliranja podataka i opisa rezultata predstavljeni su rezultati analize. Objašnjen je postupak analiziranja, a dobiveni rezultati su interpretirani te opisani. Na kraju je donesen zaključak cijele analize i dane su smjernice na temelju analize koje mogu koristiti za daljnja istraživanja ili kao primjer za izgradnju ostalih elektroničkih sustava.

Ključne riječi: podaci, otvoreni podaci, rudarenje podataka, klaster, klasteriranje, e-Građani, analiza

Sadržaj

1. Uvod.....	1
2. Karakteristike otvorenih podataka.....	2
2.1. Pojam otvorenih podataka	2
2.2. Prednosti i nedostaci otvorenih podataka	3
2.3. Otvoreni podaci u Republici Hrvatskoj	5
3. Podatkovna analiza	7
3.1. Rudarenje podataka	7
3.2. Tehnike rudarenja podataka	11
4. Pregled prethodnih istraživanja	14
5. Analiza otvorenih podataka	16
5.1. Opis odabranog skupa podataka	16
5.2. Opis problema	18
5.3. Priprema podataka za analizu	18
5.4. Modeliranje podataka	20
5.4.1. Opis i diskusija rezultata	21
6. Zaključak.....	31
Popis literature	32
Popis slika	34
Popis tablica	35

1. Uvod

Podaci su važan dio svakodnevice ljudi i predstavljaju činjenice koje imaju određeno značenje ili vrijednost. U poslovnom svijetu, kvalitetni i točni podaci mogu uvelike utjecati na uspješnost poduzeća te donijeti prednost nad sveprisutnom konkurencijom. S obzirom na brzorastući broj podataka u svijetu, dolazi do potražnje za tehnologijama koje će prikupljati i analizirati te podatke. U tu svrhu razvijene su *Big Data* tehnologije i tehnike rudarenja podataka koje omogućuju poduzećima i organizacijama prikupljanje i analizu velikih količina strukturiranih i nestrukturiranih podataka u realnom vremenu. Takve tehnologije i tehnike daju mogućnost izrade kvalitetnih analiza podataka koje pomažu prilikom donošenja važnih odluka kako u poslovnom tako i u privatnom svijetu.

Podaci nad kojima se danas uvelike koriste spomenute tehnologije i koji dobivaju sve veću važnost, nazivaju se otvoreni podaci. Oni predstavljaju golem resurs koji još nije u potpunosti iskorišten, odnosno ima veliki potencijal. Korist od otvorenih podataka možemo pronaći u područjima poput inovacija, poboljšanja učinkovitosti državnih službi, mjerenja utjecaja politike i sl. Ideja je pružiti građanima podatke koji im mogu biti korisni za svakodnevni život i koje mogu slobodno i višekratno upotrebljavati. Otvoreni podaci mogu biti korisni za donošenje boljih odluka u svakodnevnom životu, imaju velik ekonomski značaj, a važni su i za samu vladu. Danas se u Hrvatskoj stavlja sve veća važnost na razvitak i analizu otvorenih podataka po uzoru na ostatak Europe. Tome pridonosi podatak da je Hrvatska 2017. godine svrstana u skupinu zemalja predvodnica u području otvorenih podataka.

U ovom radu prikazat će se teorijski dio vezan uz otvorene podatke te će biti objašnjeno rudarenje podataka i koje metode se koriste kod istog. Također, biti će predstavljeno nekoliko prethodnih istraživanja na ovu temu. U empirijskom dijelu analizirati će se jedan skup otvorenih podataka kako bi dobili uvid u važnost i potencijal ovakve vrste podataka. Podaci nad kojima će se provesti analiza preuzeti su sa sustava e-Građani. Sustav e-Građani namijenjen je svim građanima Hrvatske i cilj mu je pomoći korisnicima da jednostavnije i uz manje vremena dođu do podataka i dokumenata koji su im potrebni. Analizom ovog skupa podataka želi se dobiti uvid u to kakva je potreba za određenim uslugama sustava e-Građani te kakvo je zadovoljstvo tim uslugama. Naposljetku će biti prikazan pregled i opis rezultata dobivenih analizom odabranog skupa podataka kako bi mogli donijeti zaključke o tome koje su usluge poželjne te ih treba razvijati, a koje su manje potrebne i ne pridonose poboljšanju života građana u velikoj mjeri.

2. Karakteristike otvorenih podataka

Otvorene podatke možemo definirati kao podatke koji se mogu slobodno i bez ograničenja koristiti nebrojeno puta te dijeliti s bilo kime uz uvjet imenovanja autora. Takvi podaci moraju biti dostupni u cijelosti, u preglednom obliku i u mogućnosti s izmjenom sadržaja, moraju imati dozvolu za korištenje izvan izvornog područja namjene, odnosno za redistribuciju te svi ljudi moraju imati pravo na njihovo korištenje. (Open data handbook, bez dat.)

2.1. Pojam otvorenih podataka

Prema Portalu otvorenih podataka Republike Hrvatske (2015) otvoreni podaci su podaci koje stvaraju tijela javne vlasti, a čijom se uporabom u komercijalne i/ili nekomercijalne svrhe može stvoriti dodana vrijednost ili ekonomska korist.

„Primjeri skupova otvorenih podataka su:

- geolokacijski podaci
- prometni podaci
- meteorološki podaci
- podaci iz područja okoliša itd.

Otvoreni podaci u pravilu ne sadrže osobne podatke, ili drugim riječima, ukoliko su podaci zaštićeni Zakonom o zaštiti osobnih podataka, iste se ne smatra otvorenim podacima jer njihova objava nije dozvoljena.“ (Portal otvorenih podataka Republike Hrvatske, 2015).

Mnoge države diljem svijeta, pa tako i Hrvatska, potiču pokret koji zagovara što veću dostupnost otvorenih podataka široj javnosti. Otvoreni podaci i njihovo pružanje javnosti pomaže u povećanju transparentnosti između javnog sektora i građana države, potiče inovacije te ekonomski razvoj. Otvaranjem podataka javnosti, ohrabruju se građani da se aktivnije uključe u politički život i provođenje zakona. Također, velik dio Europskih država iskorištava otvorene podatke kako bi olakšali život svojim građanima. Građani Danske iskoristili su otvorene podatke o lokacijama javnih zahoda kako bi napravili stranicu koja pokazuje lokacije svih javnih zahoda po gradovima kako bi ljudi mogli s lakoćom i na jednom mjestu pronaći tražene informacije kada su vani u šetnji i sl. Za građane New Yorka napravljena je aplikacija koja pokazuje sve parkove pogodne za pse na temelju otvorenih podataka koji daju podatke o tome gdje se nalaze spomenuti parkovi. Poznato je da i kompjuterska usluga poput *Google translate*-a koristi otvorene podatke iz raznih europskih dokumenata kako bi poboljšala svoje algoritme za prevođenje. (Open data handbook, bez dat.)

Prema svemu navedenom možemo vidjeti da otvoreni podaci imaju veliki potencijal, a njihovom analizom i primjenom možemo doprinijeti u raznim područjima djelovanja.

2.2. Prednosti i nedostaci otvorenih podataka

Velika prednost otvorenih podataka u tome je što se njihovom uporabom može stvoriti dodana društvena ili gospodarska vrijednost kao npr. kod izrade novih aplikacija, sustava i znanstvenih istraživanja. Visok stupanj transparentnosti između javne uprave i građana još je jedna od prednosti otvorenih podataka. Građani na raspolaganju imaju pouzdanu bazu podataka kojoj pristupaju s lakoćom i u kojoj mogu pronaći razne korisne i zanimljive informacije. Samim time, svim zainteresiranim stranama omogućeno je dobivanje novih znanja na temelju otvorenih podataka koje im mogu pomoći kod vlastitog života i poslovanja. Razvojem i povećanjem dostupnosti otvorenih podataka, gospodarski sektor dobio je lakši pristup informacijama i znanju što rezultira razvojem novih i inovativnih usluga te stvaranjem novih poslovnih modela, a u konačnici i povećanjem tržišta te profitabilnosti tog sektora. Osim što otvoreni podaci donose velike prednosti u gospodarstvu, prednosti se javljaju i u vidu otvaranja novih radnih mjesta. Dolazi do veće potražnje za radnom snagom koja ima potrebna znanja i kompetencije u radu s podacima kako bi se ti podaci što bolje iskoristili te kako bi njihovom analizom poduzeća i organizacije dobile dodane vrijednosti. Pružanjem otvorenih podataka javnosti, građane države se potiče na uključivanje u politički život te u bolje provođenje zakona s obzirom da sada sve skupine ljudi imaju jednako pravo na informacije. Također, analizom nekoliko slučajeva otvorenih podataka dokazano je da otvoreni podaci donose veliku korist kod uštede vremena i pomažu u postizanju prednosti vezanih uz očuvanje okoliša. (Europski portal podataka, 2015)



Slika 1. Prednosti otvorenih podataka (Izvor: *Europski portal podataka*, 2015)



Slika 2. Korisnici otvorenih podataka (Izvor: *Infotrend*, 2016)

S druge strane, možemo izdvojiti određene nedostatke otvorenih podataka. Ti nedostaci zapravo su prepreke u prihvaćanju otvorenih podataka. Kako je koncept otvorenih podataka relativno novi i nepoznati, građani i dalje nemaju osjećaj da je njihovo pravo tražiti javne informacije vezane uz njihov život i djelovanje njihove države. Samim time potrebno je puno ulaganja u razvitak svijesti građana kako bi se prihvatio ovaj koncept i njegova važnost. U mnogim zajednicama ljudi nemaju povjerenja u javne podatke i smatraju da se javno objavljuju podaci koji nisu od velike koristi za njih i zajednicu. Također, javljaju se mišljenja da otvoreni podaci sami po sebi nisu toliko korisni ako nisu analizirani te da zbog toga najveću korist imaju istraživači i znanstvenici koji imaju dovoljno znanja da te analize provode. Problem se javlja u informatičkoj nepismenosti građana koja je u nekim državama na zabrinjavajućoj razini zbog čega pristupanje otvorenim podacima i njihovo razumijevanje postaje problem. Vrlo je važno da svaka država potiče objavljivanje i iskorištavanje otvorenih podataka, ali da se pridaje važnost i prikazivanju tih podataka na tehnički prihvatljiv način. Potrebno je poticati marketing otvorenih podataka i informirati građane o svim mogućnostima koje oni pružaju. Osnivanje portala otvorenih podataka, kako na europskoj tako i na državnim razinama, koristan je korak u promicanju korištenja otvorenih podataka čime se dobiva komunikacijski kanal za komuniciranje s korisnicima. (Forum za javnu upravu, 2015) Otvoreni podaci imaju veliki potencijal koji se polako razvija te uz jačanje svijesti o korištenju tih podataka i razvojem informatičke pismenosti građana može se postići puno više.

2.3. Otvoreni podaci u Republici Hrvatskoj

Kao i mnoge europske države, Hrvatska je također prepoznala potencijal otvorenih podataka te je 2018. godine pokrenula politiku otvorenih podataka. Spomenuta politika predstavlja strateški smjer razvoja politike otvorenosti i transparentnosti javne uprave i provedbom te politike želi se stvoriti određeno okruženje za poticanje otvorenih podataka koji bi se opet mogli koristiti u svrhu stvaranja novih vrijednosti. (Središnji državni portal, 2018)

Glavni ciljevi politike otvorenih podataka Republike Hrvatske pružanje su maksimalne otvorenosti svih podataka uzimajući u obzir najviše standarde kvalitete i otvorenosti, stvaranje dodane društvene i gospodarske vrijednosti koristeći otvorene podatke, dobivanje visoke razine znanja na temelju otvorenih podataka i razvijanje svijesti o njihovoj ponovnoj upotrebljivosti te povezivanje javnih tijela na lokalnoj i regionalnoj razini u provođenju aktivnosti. Ovom politikom žele se promicati međunarodna načela otvorenih podataka koja zagovaraju otvorenost podataka kao pravilo čime se osigurava da podaci javnog sektora budu javno objavljeni u skladu sa svim standardima. Isto tako, načelom brojnosti i kvalitete otvorenih podataka želi se potaknuti objavljivanje što većeg broja otvorenih podataka koji su točni, sveobuhvatni te kvalitetni. Još jedna od mjera politike zagovara dostupnost otvorenih podataka svima, odnosno ovom mjerom nastoji se osigurati dostupnost podataka što većem broju korisnika. Želi se osigurati objavljivanje otvorenih podataka u formatima koji su prikladni za dohvaćanje, preuzimanje i pretraživanje u svim uobičajenim aplikacijama. Politikom otvorenih podataka potiče se zainteresirana javnost da surađuje s tijelima javnog sektora te da građani iznose svoje želje i stavove o tome koji bi podaci trebali biti javno objavljeni. (Musa, Bevandić, Herak, Jadrijević, Kovačić, Luša, Vrček, 2018, str. 8)

Razvoju koncepta otvorenih podataka u Hrvatskoj uvelike je pridonijelo pokretanje državnog portala otvorenih podataka. Počeo je s radom 2015. godine te je povezan s Europskim portalom podataka. Portal služi za prikupljanje, kategorizaciju i distribuciju otvorenih podataka javne uprave. Cilj portala je omogućiti širenje i razvitak otvorenih podataka na jednome mjestu te potaknuti ljude na korištenje dostupnih podataka s portala. Danas broji 588 skupova podataka, a neki od skupova podataka su ekološki podaci, lista lijekova, popis tijela javne vlasti, popis udruga, vremenske prognoze, adresari, registri, statistika sustava e-Građani itd. Nadalje, o korisnosti i kvaliteti podataka objavljenih na portalu svjedoči popis aplikacija, odnosno softverskih rješenja koja su razvijena na temelju otvorenih podataka s portala. Trenutno se u katalogu portala nalazi 21 aplikacija, a neke od njih su HAKmap, HZnet, LLHZZO – lista lijekova HZZO-a, Vodostaj itd. (Portal otvorenih podataka Republike Hrvatske, 2019) Uz aplikacije prikazane na portalu, postoji i niz aplikacija koje su izrađene koristeći otvorene podatke portala, a objavljene su na mrežnim trgovinama.

Promicanjem i provođenjem politike otvorenih podataka želi se u Hrvatskoj povećati transparentnost javne uprave, potaknuti sudjelovanje javnosti i uključiti društvo u odlučivanje te potaknuti suradnja na svim razinama, a ponajviše suradnja javnog s privatnim sektorom. Iako je Hrvatska svrstana u skupinu zemalja predvodnica u području otvorenih podataka, ovaj koncept još je relativno novi te je potrebno uložiti puno napora kako bi se ostvarili zadani ciljevi. Polako, ali sigurno prepoznaje se potencijal otvorenih podataka i koristi koju oni donose te se stvara određena kultura otvorenih podataka. Ovaj koncept sve se više razvija te Hrvatskoj otvara mogućnosti da se poveže s ostalim državama i time ostvari napredak, ali također omogućuje suradnju javnog i privatnog sektora te na taj način pospješuje ekonomske učinke, potiče zapošljavanje i promovira nove poslovne modele. Kontinuiranim razvojem, podizanjem svijesti, educiranjem građana, povećanjem kvalitete i količine, otvoreni podaci će imati sve veću i snažniju ulogu u ekonomskom napretku Hrvatske. Podatak o izravnoj koristi od nekoliko milijardi eura godišnje od otvorenih podataka na području Europske unije može biti samo dodatna motivacija za smjer u kojem je Hrvatska krenula i koji bi još snažnije trebala razvijati.

Tablica 1. Najaktivniji izdavači otvorenih podataka u Republici Hrvatskoj do 2019.god.

Izdavač	Broj skupova podataka
Grad Rijeka	149
Državni zavod za statistiku	79
Grad Zagreb	77
Hrvatska agencija za okoliš i prirodu	39
Agencija za mobilnost i programe Europske unije	20
Ministarstvo uprave	17
Hrvatska narodna banka	17
Ministarstvo pravosuđa	16
Grad Pula	15
Hrvatska agencija za nadzor financijskih usluga	13

(Izvor: *autorski rad, vlastita izrada prema podacima s Portala otvorenih podataka Republike Hrvatske, 2019*)

3. Podatkovna analiza

Živimo u svijetu gdje postoji veliki broj podataka koji se svakodnevno prikuplja te je od velike važnosti da se ti podaci analiziraju, a dobiveni rezultati iskoriste u obliku novih znanja. Mnoga poduzeća prikupljene podatke spremaju u svoje baze podataka analizirajući ih kako bi mogli što bolje razumjeti što njihovi klijenti žele da bi mogli svoju ponudu prilagoditi njima. Tako veliki broj podataka koji se sakuplja, doveo je do potražnje za kvalitetnim alatima, tehnologijama i procesima koje bi te podatke mogle obraditi i analizirati na način da pomoću njih steknemo nova znanja. Takav jedan proces otkrivanja zakonitosti, obrazaca i trendova u velikim skupovima podataka nazivamo rudarenje podataka.

3.1. Rudarenje podataka

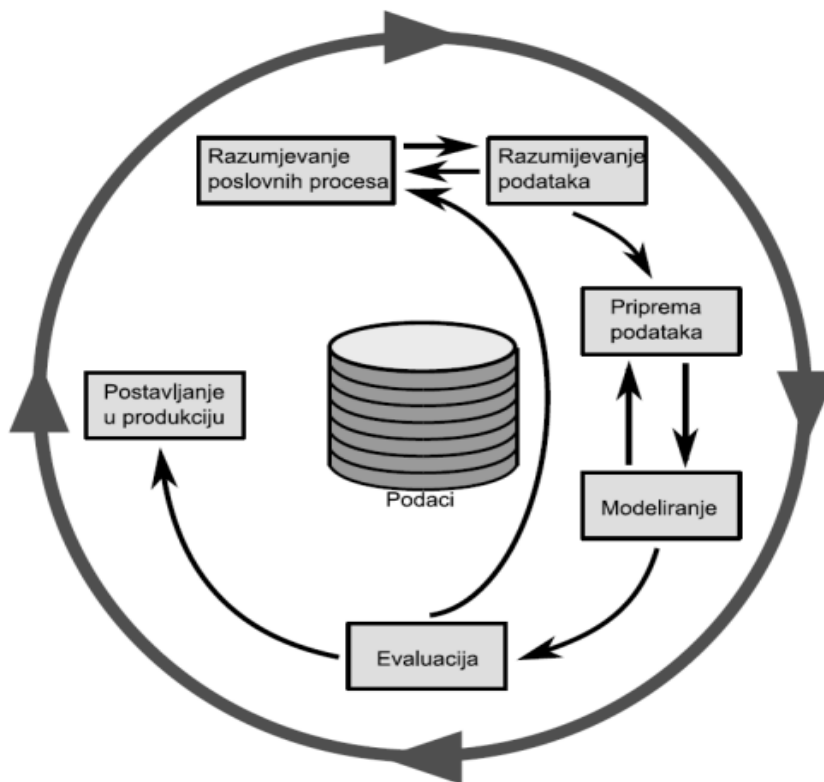
Možemo reći da je područje rudarenja podataka mlado, dinamično i puno noviteta. Rudarenjem podataka pretvara se velika količina podataka u znanje. Još poznato i pod nazivom „otkrivanje znanja u podacima“, rudarenjem podataka velike se količine podataka stavljaju u upotrebu te otkrivaju novi uzorci i obrasci koji poduzećima mogu biti korisni za daljnje napredovanje i donošenje odluka. Podaci za rudarenje mogu biti iz raznih baza podataka, skladišta podataka, interneta i drugih informacijskih repozitorija. Naravno, skupovi nad kojima se provodi rudarenje podataka moraju biti veliki skupovi kako bi algoritmi imali što veći broj podataka za kreiranje modela. Primjenom današnjih tehnika i analiza nad strukturiranim i nestrukturiranim podacima omogućeno je uočavanje novih pravila, obrazaca, kreiranje novih ideja, olakšano donošenje odluka i izrada predikcija te budućih ishoda. Sve do pojave tehnika rudarenja podataka, poduzeća i organizacije morale su donositi odluke na temelju vlastitog znanja i intuicije bez pomoći tehnologije, ali danas je za većinu organizacija nezamislivo donošenje novih odluka bez prediktivnih modela i analiza. Za primjer možemo uzeti proizvođača kompjutera *Dell* koji je želio poboljšati produktivnost svojih zaposlenika, odnosno prodavača. Okrenuli su se analizi podataka i prediktivnoj analitici te su analizirali svoju bazu podataka o potencijalnim kupcima kako bi se pronašli najvjerojatniji ispitanici. Nakon što su dobili više informacija o potencijalnim kupcima, omogućeno je kreiranje personaliziranih ponuda i letaka za njihove kupce te je time postignuto povećanje u prodaji. (T. Larose i D. Larose, 2015, str. 3) Ovo je samo jedan primjer kako rudarenje podataka može pomoći u prodajnom poslovanju, a isto tako veliku korist ima i u ostalim granama poslovanja.

Kada govorimo o podacima koji se mogu rudariti, možemo reći da podaci mogu biti bilo koje vrste sve dok imaju korist za ciljanu uporabu. Oblik podataka koji se obično koristi za rudarenje su podaci iz baza podataka, skladišta podataka i transakcijski podaci. Spomenuta

tehnika može se iskoristiti i na drugim oblicima podataka poput multimedijских podataka, grafičkih podataka, internetskih podataka itd. Naravno, razvitkom ove grane prihvaćat će se sve više oblika podataka za rudarenje.

Rudarenje podataka područje je računalne znanosti koje ima puno zadataka, a neki od najvažnijih su deskripcija, estimacija, predikcija, klasifikacija, klasteriranje i asocijacija. Pomoću deskripcije analitičari opisuju uzorke i trendove koji se nalaze iza nekog skupa podataka. Važno je da modeli dobiveni rudarenjem podataka budu što transparentniji, odnosno rezultati modela moraju opisivati jasne uzorke i pravilnosti koje svaka osoba može prepoznati i objasniti. Neke od metoda rudarenja podataka, poput stabla odlučivanja, puno su transparentnije što se tiče interpretacije dobivenih rezultata dok rezultate dobivene metodom neuronske mreže mogu objasniti samo profesionalni analitičari i ostali stručnjaci tog područja. Estimacija, koja se još naziva i regresija, zadaća je rudarenja podataka kojom se aproksimiraju vrijednosti numeričke ciljne varijable koristeći skup numeričkih ili kategoričkih prediktorskih varijabli. Koriste se kompletni zapisi koji daju vrijednost ciljne varijable uz prediktorske varijable kako bi se izradili modeli, a tada se za novo opažanje procjenjuje vrijednost ciljne varijable na temelju vrijednosti prediktora. Primjeri korištenja estimacije u svakodnevnom životu su procjena količine novaca koju će nasumična četveročlana obitelj potrošiti za školsku opremu ili procjena ocjena srednjoškolca na temelju ocjena iz osnovne škole. Kod predikcije rezultati leže u budućnosti, tj. kod predikcije ciljnu varijablu predstavlja buduća varijabla. Za primjere predikcije možemo uzeti predviđanje cijene dionica tri mjeseca unaprijed, predviđanje povećanja postotka smrti u prometu ako se poveća ograničenje brzine i predviđanje pobjednika svjetskog prvenstva na temelju usporedbe statistika timova. Sljedeći zadatak rudarenja podataka je klasifikacija kod koje se za ciljnu varijablu uzima kategorija za razliku od estimacije gdje se uzima numerička varijabla. Za primjer kategoričke varijable možemo uzeti prihode koji se kod klasifikacije mogu podijeliti na tri kategorije: visoki, srednji i niski prihodi. Svaki zapis u skupu podataka sadrži informaciju o ciljnoj varijabli te o skupu ulaznih varijabli. Primjer tih varijabli može biti zanimanje, spol ili dob. Stručnjak koji provodi klasifikaciju može na temelju postojećih zapisa iz baze odrediti prihode za osobe koje nisu zapisane u bazi. Klasifikacija ima veliku primjenu u svakodnevnom životu poput dijagnosticiranja prisutnih bolesti, procjenjivanja da li je prijava za hipoteku dobar ili loš kreditni rizik, određivanja da li je oporuku zaista napisala preminula osoba ili je krivotvorena i sl. Što se tiče klasteriranja, ono se odnosi na grupiranje zapisa, zapažanja ili slučajeva u klase sličnih objekata. Klasteriranje se ne bavi klasificiranjem, predikcijama ili estimacijama već pokušava segmentirati cijeli skup podataka u homogenu podgrupu ili klaster na način da sličnost zapisa unutar klastera bude maksimalna, a sličnosti sa zapisima izvan klastera budu minimalne. Na kraju imamo asocijaciju koja ima zadatak pronaći attribute koji idu zajedno. Poznata je u poslovnom svijetu gdje se još

naziva i analiza afiniteta ili analiza potrošačke košarice, a pomoću nje omogućeno je otkrivanje pravila za kvantificiranje veza između dviju ili više atributa. Primjere upotrebe asocijacije pronalazimo u poslovanju i istraživanjima, a neki od njih su ispitivanje razmjera djece čiji im roditelji čitaju, a i oni sami su dobri čitači ili određivanje slučajeva u kojima će novi lijek imati opasne nuspojave. (T. Larose i D. Larose, 2015, str. 10 - 17) Iz navedenog možemo vidjeti da svaki od zadataka rudarenja podataka ima primjenu u mnogim segmentima ljudskog djelovanja i donosi veliku korist u svakodnevnom poslovanju te životu ljudi.



Slika 3. Faze u životnom ciklusu rudarenja podataka (Izvor: *Banić, Štajcer, 2016*)

Mnoge tvrtke u zadnjih nekoliko godina počele su pristupati rudarenju podataka na različite načine i razvijati svoje slučajeve. Kako se ovaj koncept počeo koristiti u raznim organizacijama, pojavila se potreba za kreiranjem međustručnog standarda koji je neutralan prema industriji i alatima te koji je aplikacijski neutralan. Kreiran pod nazivom „Međustručni standardizirani postupak za rudarenje podataka“ (*CRISP - DM*), ovaj besplatno dostupan standard omogućava uklapanje postupka rudarenja podataka u opću strategiju za rješavanje problema u poslovanju ili istraživačkim jedinicama. (T. Larose i D. Larose, 2015, str. 6) Razvojem spomenutog standarda, organizacije su dobile generalne smjernice vezane uz faze i korake koje bi projekt rudarenja podataka trebao pratiti, a one ih

mogu dodatno prilagoditi svojim potrebama. Prema *CRISP - DM* standardu, svaki projekt ili postupak rudarenja podataka trebao bi se sastojati od šest faza i važno je napomenuti da svaka sljedeća faza ovisi o tome kakav će biti rezultat, odnosno ishod prethodne faze. Na slici 3 prikazane su faze u životnom ciklusu rudarenja podataka te je moguće primijetiti da je cijeli taj ciklus, odnosno proces iterativan i adaptivan.

Prva je faza kod *CRISP - DM* standarda razumijevanje poslovnih procesa. Prema toj fazi, prvotno je potrebno navesti i objasniti što određenim projektom želimo postići te koji su ciljevi tog projekta kada govorimo o poslovnoj ili istraživačkoj jedinici u cjelini. Kada smo jasno odredili ciljeve i ograničenja, tada ih je potrebno pretvoriti u definiciju problema rudarenja podataka kako bismo znali što se procesom rudarenja podataka želi postići. Zadnji je dio ove faze pripremanje početne strategije za postizanje ciljeva. Nakon faze razumijevanja poslovnih procesa dolazi faza razumijevanja podataka. Početni dio ove faze predstavlja prikupljanje potrebnih podataka. Nakon toga potrebno je napraviti istraživačku analizu prikupljenih podataka kako bi se поближе upoznali s podacima koji će se koristiti i kako bi otkrili početne pokazatelje. Sljedeći korak je procjenjivanje kvalitete podataka. Na kraju ove faze poželjno je odabrati zanimljive podskupove koji bi mogli sadržavati djelotvorne uzorke. Treća faza, koja se zove priprema podataka, radno intenzivna je faza i ona pokriva sve aspekte pripremanja završnog skupa podataka koji će se koristiti za nadolazeće faze. Potrebno je zatim odabrati slučajeve i varijable koje želimo analizirati i koje su prikladne za analizu. Završni koraci ove faze su izvršavanje transformacija nad pojedinim varijablama i čišćenje sirovih podataka kako bi bili spremni za modeliranje. Faza modeliranja, koja predstavlja četvrtu fazu u *CRISP - DM* standardu, započinje odabirom i primjenom odgovarajuće tehnike za modeliranje podataka. Kako bi dobili što optimiziranije rezultate, potrebno je namjestiti postavke modela. Također, za isti problem rudarenja podataka često je moguće koristiti više tehnika modeliranja. Peta faza je faza evaluacije. Modeli iz faze modeliranja su do tada napravljeni te je potrebno napraviti evaluaciju kvalitete i efektivnosti tih modela. Nadalje, u ovoj fazi potrebno je provjeriti da li dobiveni modeli zadovoljavaju ciljeve i ograničenja koji su bili zadani u prvoj fazi. Važno je utvrditi da li su pokriveni svi aspekti poslovnog ili istraživačkog problema i na kraju donijeti odluku o korištenju rezultata rudarenja podataka. Zadnja faza procesa rudarenja podataka, prema spomenutom standardu, faza je primjene ili postavljanja u produkciju. Iako su modeli kreirani, to ne znači potpuni završetak projekta. Potrebno je koristiti stvorene modele i primijeniti ih u postizanju ciljeva. Korisna je izrada izvještaja ili zahtjevnija primjena poput paralelne implementacije procesa rudarenja podataka u drugim odjelima. (T. Larose i D. Larose, 2015, str. 6 - 8)

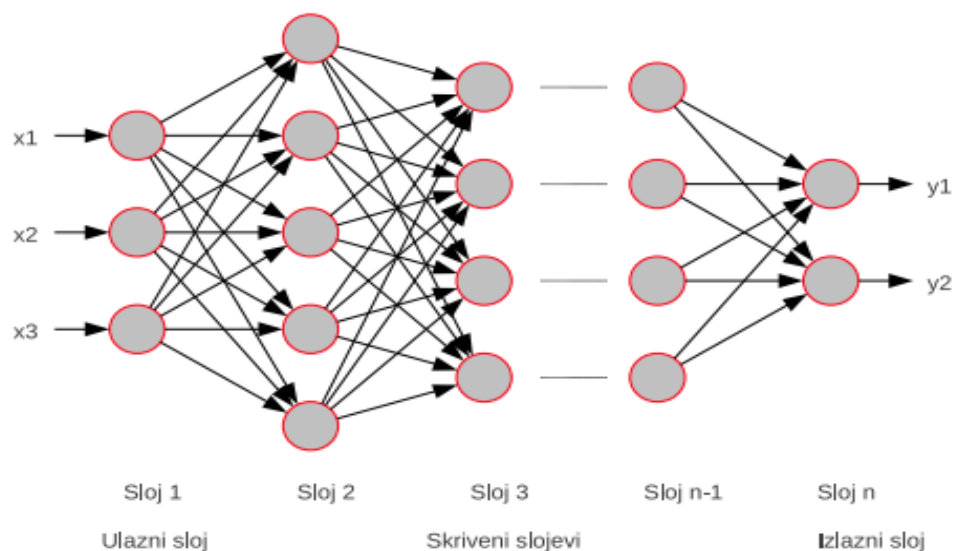
3.2. Tehnike rudarenja podataka

Postoje razne tehnike koje se danas koriste prilikom rudarenja podataka. Jedna od glavnih zadaća u ovom procesu odabir je prikladne tehnike pomoću koje ćemo dobiti kvalitetne i željene rezultate. U nastavku će biti navedene i ukratko objašnjene najvažnije tehnike današnjeg svijeta rudarenja podataka.

- **Klasteriranje** - Ova tehnika, poznata još pod nazivom grupiranje, predstavlja proces grupiranja skupa podatkovnih objekata u više grupa ili klastera na način da objekti koji se nalaze unutar istog klastera imaju velike sličnosti, a s objektima iz ostalih klastera imaju velike razlike. Spomenute sličnosti i razlike među objektima određuju se na temelju vrijednosti atributa i često uključuju mjerenja udaljenosti. Korist od ove tehnike nalazi se u njezinoj mogućnosti pronalaska nepoznatih grupa unutar nekog skupa podataka što nam može olakšati razumijevanje određenih pojava. Klasteriranje se u velikoj mjeri koristi u poslovnoj inteligenciji gdje pomaže kod organiziranja velikog broja ljudi u grupe na temelju karakternih sličnosti. Na taj način, omogućava se izrada posebnih poslovnih strategija za svaku od grupa kako bi se poboljšalo upravljanje odnosima s kupcima i dobila mogućnost izrade kvalitetnijih i personaliziranih ponuda. Također, klasteriranje ima primjenu u pretraživanju interneta. Određeni pojam koji pretražujemo na internetu može vratiti velik broj rezultata s obzirom da danas postoji puno web stranica. Pomoću klasteriranja možemo podijeliti dobivene rezultate u grupe kako bi se prikazale samo one grupe odnosno podaci koji najbolje odgovaraju traženom pojmu. Vrijedno je spomenuti metodu klasteriranja pod nazivom partijska metoda ili metoda dijeljenja. To je jednostavna i osnovna verzija klaster analize koja organizira objekte nekog skupa u nekoliko grupa ili klastera. (Han, Kamber, Pei, 2012) Kod nje se koristi algoritam k – srednjih vrijednosti koji identificira broj centroida i tada alokira svaku podatkovnu točku najbližem klasteru održavajući centroide što manjima. (J.Garbade, 2018)
- **Stablo odlučivanja** - Kod stabla odlučivanja postoji kolekcija čvorova odluke povezanih granama koji se protežu prema dolje od korijenskih čvorova sve dok ne dođu do završnih čvorova. Prema dogovoru, korijenski čvor početak je stabla odlučivanja i nalazi se na vrhu dijagrama. Atributi kojima raspolažemo testiraju se na čvorovima odluke pri čemu svaki mogući ishod rezultira granom. Nakon toga, svaka grana vodi do drugog čvora odluke ili do završnog čvora. (T. Larose i D. Larose, 2015, str. 317) Pojednostavljeno, stablo odlučivanja sastoji se od mogućnosti između kojih moramo odlučiti, posljedica koje se mogu dogoditi ukoliko odaberemo neku od ponuđenih mogućnosti i vjerojatnosti da će se neka od tih posljedica dogoditi. Kada je potrebno predstaviti i objasniti kompleksnu situaciju donošenja odluka, stablo odlučivanja pruža odlične mogućnosti. Osim navedenog, ova

tehnika pomaže osobama koje donose odluke kao potpora razumijevanju osnovnog problema izbora i obuhvaćanju neizvjesnih događaja koji utječu na ishode i posljedice inačica. Velika prednost ove tehnike je njezina jednostavnost i transparentnost.

- **Neuronske mreže** - Neuronske mreže predstavljaju osnovni pokušaj imitiranja određene vrste nelinearnog učenja koje se pojavljuje u neuronskim mrežama pronađenima u prirodi. Možemo reći da je to tehnika umjetne replike ljudskog mozga kojom se nastoji simulirati postupak učenja. Neuronske mreže sastoje se od skupa međusobno povezanih procesnih elemenata, čvorova, čija se funkcionalnost temelji na biološkom neuronu. (Bašić, Čupić, Šnajder, 2008) Nadalje, ovom tehnikom mogu se prikupiti ulazni podaci iz velikog broja ulaza koji se zatim obrađuju u što kraćem vremenu i na kraju dobivamo izlazni parametar na jednom izlazu. Prednosti neuronskih mrežama su u tome što imaju visoku toleranciju na oštećene i iskvarene podatke kao i njihova sposobnost da klasificiraju uzorke na koje još nisu naučeni. Nadalje, mogu se koristiti kada imamo malo znanja o odnosima između atributa i klasa. Uspjeh neuronskih mreža očituje se na mnoštvu podataka iz stvarnog svijeta kao npr. kod prepoznavanja ljudskog rukopisa, osposobljavanja računala za izgovaranje engleskog teksta ili unaprjeđivanja laboratorijske medicine. Kritika je ove tehnike u tome što je rezultate teško interpretirati te je prosječno potrebno puno više sati rada da bi se dobili željeni rezultati u odnosu na ostale tehnike rudarenja podataka. (Han, Kamber, Pei, 2012)



Slika 4. Primjer slojevite neuronske mreže (Izvor: Čupić, Bašić, Golub, 2013)

- Regresijska analiza** - Ova tehnika omogućava ispitivanje veza između dvije ili više varijabli koje promatramo. Regresijska analiza pokušava odrediti jačinu veze između jedne zavisne varijable i niza ostalih promjenjivih varijabli, odnosno nezavisnih varijabli. Proces izvođenja regresije omogućava nam da sa sigurnošću odredimo koji faktori su najznačajniji, koji mogu biti odbačeni i kako ti faktori utječu jedni na druge. Vrste regresijske analize koje se najčešće koriste su linearna i logistička regresija. Linearna regresija podrazumijeva pronalaženje najbolje linije koja odgovara dvama atributima na način da se jedan atribut može iskoristiti kako bi se predvidio neki drugi atribut ili varijabla. (Han, Kamber, Pei, 2012) Logistička regresija predstavlja generalizaciju linearne regresije. Većinom se koristi za predviđanje brojčanih varijabli. Kod ovakve vrste, umjesto da se predviđa hoće li se dogoditi neki događaj, izrađuje se model za predviđanje logaritma izgleda njegove pojave. (Two Crows Corporation, 2005) Regresija se može koristiti za projiciranje cijena na temelju faktora poput dostupnosti, potražnje i konkurencije. Sportski analitičari koriste regresiju za predviđanje broja golova koje će igrač zabiti u narednim utakmicama na temelju prethodnih nastupa.
- Pravila asocijacije** - Tehnika pravila asocijacije omogućava pronalazak poveznica između dvije ili više varijabli u velikim bazama podataka. Upotrebom ove tehnike možemo pronaći skrivene obrasce u skupovima podataka, a ti obrasci mogu se koristiti za identificiranje varijabli i pojavu različitih varijabli koje se pojavljuju s najvišom učestalošću. Pravila asocijacije, kao tehniku rudarenja podataka, možemo promatrati kroz proces koji se sastoji od dva koraka. Prvi korak je pronalazak najčešćih skupova predmeta. Drugi korak odnosi se na generiranje pravila pridruživanja iz pronađenih skupova predmeta. (Han, Kamber, Pei, 2012) Primjenu tehnike pravila asocijacije pronalazimo u ispitivanju i predviđanju ponašanja kupaca. Također, korisna je kod otkrivanja zlonamjernih softvera, analiziranja tržišne košarice, određivanja izgleda kataloga, rasporeda prodavaonica itd.

4. Pregled prethodnih istraživanja

Istraživanja na temu otvorenih podataka posljednjih godina zastupljena su u velikoj mjeri. Popularnosti otvorenih podataka pridonosi sve veći razvoj informacijsko - komunikacijske tehnologije. Kako se razvijaju nove tehnologije, tako se i pronalaze nove koristi i prednosti ove vrste podataka. U nastavku će biti prikazano nekoliko istraživanja vezanih uz ovu temu.

Jedno od istraživanja otvorenih podataka govori o njihovom korištenju za predviđanje odgovora turista na turističke atrakcije. Povećanjem broja korisnički generiranog sadržaja poput recenzija, komentara i prošlih iskustava, pojavila se velika količina dostupnih informacija. Turisti te informacije mogu iskoristiti kako bi lakše donijeli odluke vezane uz mjesto putovanja, smještaj, prijevoz, hranu i slično. Ovim istraživanjem željelo se doznati u kolikoj mjeri analiza ovih dostupnih odnosno otvorenih podataka, kojima svaki turist ima mogućnost pristupa, može predvidjeti njihov odgovor na određene turističke destinacije. Istraživanjem se pokušao predvidjeti stav turista prema turističkoj atrakciji transformirajući velike količine otvorenih podataka u vrijedne prijedloge. Pri analizi korišten je algoritam slučajne šume, a skup podataka za analizu preuzet je s *TripAdvisor*-a. Iako razvoj tehnologije pruža nove usluge i alate za potporu turistima kod donošenja odluka, ono također donosi nove izazove zbog velike količine dostupnih podataka koje je potrebno strukturirati i limitiranog korištenja tih podataka od strane turističkih destinacija, agencija i ugostiteljskih menadžera. Jedan od prvih pokušaja da se istraži upotreba otvorenih podataka za predviđanje reakcija turista prema određenoj destinaciji upravo je ovo istraživanje. Prema rezultatima, spomenutom analizom moguće je s velikom točnošću odrediti trend potrošačkog uvažavanja određenih destinacija. Turističkim menadžerima preporuča se usvajanje analize otvorenih podataka kako bi bolje predvidjeli atraktivnost određene turističke destinacije. Ovo istraživanje moglo bi poslužiti kao temelj za razvijanje analize otvorenih podataka u turističke svrhe i prilagođenih marketinških strategija. (Pantano, Priporas, Stylos, 2017)

Prema nedavnom istraživanju u Americi, slike i podaci dobiveni od sustava za snimanje Zemlje iz svemira uvelike su pomogli u razvitku znanosti i ekonomije. Američki svemirski projekt *Landsat*, pokrenut je u svrhu promatranja Zemlje iz svemira. Godine 2008. *Landsat* program donio je odluku o pokretanju politike otvorenih podataka kojom su svi podaci vezani uz taj projekt objavljeni na internetu. Taj događaj doveo je do znatnog povećanja u preuzimanju slika i podataka dobivenih *Landsat* sustavom. Istraživanje navodi da je analizom šesnaest ekonomskih sektora koji su koristili *Landsat* podatke utvrđeno značajno povećanje produktivnosti te da je ekonomska korist spomenutih podataka viša od milijardu dolara. Također, politika otvorenih podataka donijela je korist vladi, sveučilištima i organizacijama s

ograničenim resursima. Istraživanje navodi da je objavljivanje podataka *Landsat* projekta navelo i ostale nacionalne opservacijske programe da pokrenu politiku otvorenih podataka i objave svoje podatke svima na korištenje bez naknade. Istraživanje u svom zaključku navodi da se sve više javnih i privatnih sektora te organizacija oslanja na podatke iz *Landsat* projekta. Jedan od primjera je konvencija Ujedinjenih naroda o klimatskim promjenama koja je specificirala mehanizme smanjenja krčenja šuma i degradacije šuma na temelju slika i podataka dobivenih iz *Landsat* sustava. Otvoreni podaci iz *Landsat* sustava omogućili su Americi da učvrsti svoj položaj kao svjetski predvodnik u pružanju podataka vezanih uz promatranje Zemlje. Prikupljeni i objavljeni podaci omogućuju drugim projektima da sva dobivena znanja iskoriste kako bi unaprijedili svoje sustave te taj otvoreni pristup polako postaje norma u svijetu kako bi se maksimizirale prednosti satelitskih podataka. Koncept otvaranja i objavljivanja podataka omogućio je podacima o promatranju Zemlje da dobiju važnu ulogu u razvoju znanosti. (Zhu et al., 2019)

S obzirom da se ovaj rad bavi analizom otvorenih podataka iz sustava e-Građani, koji predstavlja jedno od inovacija na području elektroničkih usluga Hrvatske, u nastavku će biti prikazano istraživanje vezano uz inovacije na području elektroničkih usluga. Istraživanje se odnosi na pojavljivanje novih elektroničkih trendova u Aziji. U azijskoj zemlji Bangladeš, na relativno maloj površini živi veliki broj ljudi. S obzirom na tu činjenicu, postojali su veliki problemi kod javnog prijevoza i transporta. Ljudi su za kupnju karata za bus ili vlak trebali stajati u redu po nekoliko sati, a za vrijeme praznika ili blagdana, kada svi putuju svojim kućama, bilo je potrebno čekati cijelu noć da bi došli na red za kupnju karte. Kako bi se riješio taj problem, donesena je odluka o izradi elektroničke usluge pod nazivom *Shohoz* preko koje bi tamošnji stanovnici mogli kupovati karte. Također, osim elektroničke kupnje karata, stanovnici su ovom uslugom dobili mogućnost da se te karte dostave u njihove domove. Ovo istraživanje analizira operaciju, strukturu, reklamiranje i marketinšku strategiju elektroničke usluge *Shohoz*. Nadalje, istraživanjem se želi preporučiti spajanje tehnologije s postojećim prometnim sustavom prema primjeru *Shohoz* e-usluge. Napravljena je kvalitativna analiza, a podaci za analizu su prikupljeni iz sekundarnih izvora poput časopisa, web stranica itd. Dobiveni rezultati govore da je izrada ove e-usluge pokrenula novu nišu na tržištu i uspješno potaknula nove inovacije u području e-usluga. Istraživanjem se preporuča da se spomenuta usluga proširi po cijeloj zemlji, odnosno da se započne postupak decentralizacije. Inovacije koje je donijela e-usluga *Shohoz* potaknule su susjedne zemlje, poput Indije, da naprave svoje sustave kako bi svojim građanima olakšali putovanje. (Jalil, 2016)

5. Analiza otvorenih podataka

5.1. Opis odabranog skupa podataka

Kod analize koristi se skup podataka preuzet s Portala otvorenih podataka Republike Hrvatske. Podaci za taj skup dobiveni su iz sustava e-Građani, a govore o korištenju pojedinih e-usluga tog sustava. Objavljeni su kao otvoreni podaci koji su postavljeni u strojno čitljivom obliku te svaki građanin i poduzetnik ima mogućnost preuzimanja tih podataka te korištenja u određene svrhe. Objavljivanjem podataka sustava e-Građani, Ministarstvo uprave želi potaknuti pristupačnost, redistribuciju i globalnu uključenost otvorenih podataka u Hrvatskoj. Skupom podataka o korištenju usluga sustava e-Građani želi se pružiti uvid u to koje e-usluge se najviše koriste, a koje usluge možda nisu potrebne. Skup podataka koji je objavljen, uključuje podatke o korištenju e-usluga u razdoblju od 2014. - 2019. godine.

Što se tiče samih podataka, podijeljeni su u pet skupina atributa, a to su redom:

- Naziv_e_usluge - tekstualni atribut koji predstavlja naziv korištene e-usluge
- Ukupni_broj_prijava - numerički atribut koji predstavlja ukupan broj prijava korisnika na određenu uslugu sustava
- Broj_jedinstvenih_korisnika - numerički atribut koji predstavlja broj jedinstvenih korisnika koji su koristili određenu uslugu
- Textbox7 (Prosječan broj prijava po korisniku) - numerički atribut koji predstavlja prosječan broj prijava po korisniku na određenu uslugu
- TextBox13 (Suma svih prijava na usluge sustava) - numerički atribut koji predstavlja ukupan broj prijava na sve usluge sustava, vrijednost atributa u svakom je retku jednaka pa prema tome nema neku veliku važnost kod analize

Trenutno se u sustavu nalazi 85 e-usluga koje su građanima na raspolaganju. S obzirom na to, odabrani skup podataka sadrži 85 instanci za svaki atribut. U nastavku se nalazi prikaz jednog dijela originalnog skupa podataka kako bi se dobio uvid u to kako odabrani podaci izgledaju.

Naziv_e_usluge	Ukupni_broj_prijava	Broj_jedinstvenih_korisnika	Textbox7	Textbox13
Osobni korisnički pretinac	8.641.288	556.958	16	22.862.363
e-Dnevnik	2.211.461	73.089	30	22.862.363
eUsluge Porezne uprave	1.940.707	289.631	7	22.862.363
HZMO korisničke stranice	1.539.079	210.274	7	22.862.363
e-Matice	1.362.492	391.490	3	22.862.363
HZMO elektronički zapis	1.359.474	310.395	4	22.862.363
e-Usluge MUP-a	920.282	303.881	3	22.862.363
Burza rada	840.750	111.564	8	22.862.363
Elektroničke potvrde	409.971	189.194	2	22.862.363
REGOS Ena	321.983	169.160	2	22.862.363
eFondovi	316.568	8.850	36	22.862.363
ZIS OSS	307.889	84.000	4	22.862.363
Moj račun - Regos	300.010	126.145	2	22.862.363
Izabrani liječnik	283.266	129.070	2	22.862.363
mojID	262.249	145.021	2	22.862.363
Uvjerenje da se ne vodi kazneni postupak	223.534	122.993	2	22.862.363
ISSP	161.316	74.373	2	22.862.363
Portal zdravlja	152.093	34.175	4	22.862.363
Uvid u OIB	140.816	83.552	2	22.862.363
Centralna prijava na studijske programe	119.688	18.694	6	22.862.363
Realizirani recepti	98.251	42.366	2	22.862.363
Komunalne usluge i naknade	88.666	46.401	2	22.862.363
Registar birača - e-Privremeni upis	84.310	41.311	2	22.862.363
HZZO Otvorene narudžbe	80.184	30.956	3	22.862.363
EKZO	76.889	46.888	2	22.862.363
Uvjerenje iz kaznene evidencije	71.669	45.684	2	22.862.363
e-Obrt	65.307	15.082	4	22.862.363
Uvid u ePKK za strane fizičke osobe	56.756	17.174	3	22.862.363
eVisitor	51.946	8.163	6	22.862.363
eEnergetskiCertifikat	38.569	1.181	33	22.862.363
Porezna prijava za obračun i plaćanje posebnog poreza na motorna vozila	33.308	14.633	2	22.862.363
Aplikacija za e-savjetovanje	29.381	6.945	4	22.862.363
Izdavanje elektroničke isprave Grada Zagreba	25.978	5.844	4	22.862.363
Elektronička komunikacija sa sudovima	22.378	1.632	14	22.862.363
Registar birača	20.099	12.572	2	22.862.363
SISSI	16.977	726	23	22.862.363
EDIP	16.497	194	85	22.862.363
e-Zahtjev za izdavanje vozačke dozvole	16.088	7.133	2	22.862.363
eNovorođenče	15.505	5.361	3	22.862.363
e-Pomorac	13.649	2.963	5	22.862.363
e-Plovilo	12.476	5.785	2	22.862.363
SKDD e-Ulagatelj	11.477	5.779	2	22.862.363
Online tečajevi Srca	9.443	5.050	2	22.862.363
e-Razmjena studentskih ocjena	8.701	6.987	1	22.862.363
NIAS SocSkrb	8.606	5.645	2	22.862.363
e-Prijava boravišta hrvatskih državljana	8.380	4.779	2	22.862.363
e-Zahtjev za izdavanje ePutovnice	7.844	4.271	2	22.862.363

Slika 5. Prikaz odabranog skupa podataka (Izvor: *Portal otvorenih podataka Republike Hrvatske, 2019*)

5.2. Opis problema

U ovom radu, analizom skupa podataka o korištenju usluga sustava e-Građani, želi se vidjeti da li građani Hrvatske koriste taj sustav i uolikoj mjeri. Također, analizom korištenja usluga pokušat će se donijeti zaključci o tome da li je sustav e-Građani riješio dio problema papirologije, koji je u Hrvatskoj značajan, s obzirom da sada građani imaju mogućnost pregleda i preuzimanja potrebnih dokumenata kod kuće preko interneta. Nadalje, prema korištenosti usluga želi se saznati da li sustav najviše koriste akademski građani, poduzetnici ili neka treća skupina. Dobiveni rezultati i zaključci analize mogu poslužiti kao primjer koji daje uvid u to da li je isplativo uvođenje jednog ovakvog sustava i uz koje grupe ljudi i kategorije usluga bi novi sustav trebao biti vezan da bi se najviše isplatio i olakšao život građanima.

5.3. Priprema podataka za analizu

Faza pripreme podataka za analizu obuhvaća određivanje potrebnih podataka, njihovu transformaciju i vrednovanje. Preuzete podatke potrebno je prilagoditi za rad u alatu *BigML* koji će se koristiti za analizu odnosno modeliranje. Kod originalnih podataka preuzetih sa sustava e-Građani, numerički zapisi sadrže točku kao simbol za razdvajanje tisućica. S obzirom da se za analizu koristi alat *BigML*, potrebno je točku zamijeniti zarezom ili jednostavno ukloniti kako bi se podaci mogli pravilno učitati i koristiti u odabranom alatu.

Također, nazive određenih atributa potrebno je preoblikovati kako bi imali više smisla. To se odnosi na atribut pod nazivom „Textbox7“ koji se u originalnom skupu podataka nalazi u četvrtom stupcu. Nakon proučavanja podataka u tom stupcu, vidljivo je da se taj atribut odnosi na prosječan broj prijava po korisniku na određenu uslugu. Prema tome, naziv spomenutog atributa promijenjen je u „Prosjecan_broj_prijava_po_korisniku“ kako bi se lakše snalazili kod same analize.

Drugi atribut koji je potrebno promijeniti, u originalnom skupu podataka, nalazi se u petom stupcu pod nazivom „Textbox13“. Podaci u tom stupcu zapravo prikazuju ukupan broj prijava na sve usluge sustava. S obzirom da je to jedinstven broj, za bilo koju instancu u skupu podataka nalazi se isti broj. Kako je vrijednost spomenutog atributa uvijek ista, kod analize neće imati veliku važnost pa ga je poželjno izbaciti. Prednost je alata *BigML* što prepoznaje atribute, odnosno podatke koje je potrebno izbaciti te ih označava crvenim uskličnikom, a to je moguće prepoznati i iz grafičkog prikaza (histograma) koji alat nudi što će biti objašnjeno kod analize. Navedena prednost korisna je kod velikih skupova podataka s puno atributa gdje je teško prepoznati podatke koji nisu potrebni za analizu, tj. koji nemaju utjecaja na rezultate. Na slici 6 moguće je vidjeti dio odabranog skupa podataka nakon izvršene pripreme.

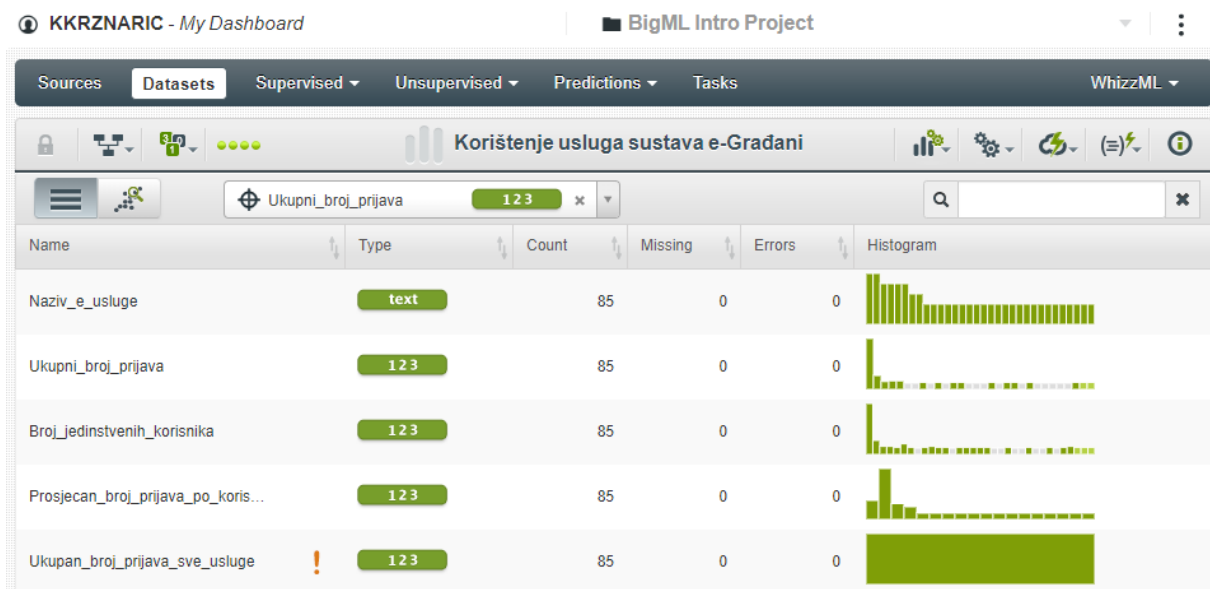
Naziv_e_usluge	Ukupni_broj_prijava	Broj_jedinstvenih_korisnika	Prosjecan_broj_prijava_po_korisniku	Ukupan_broj_prijava_sve_usluge
Osobni korisnički pretinac	8641288	556958	16	22862363
e-Dnevnik	2211461	73089	30	22862363
eUsluge Porezne uprave	1940707	289631	7	22862363
HZMO korisničke stranice	1539079	210274	7	22862363
e-Matice	1362492	391490	3	22862363
HZMO elektronički zapis	1359474	310395	4	22862363
e-Usluge MUP-a	920282	303881	3	22862363
Burza rada	840750	111564	8	22862363
Elektroničke potvrde	409971	189194	2	22862363
REGOS Ena	321983	169160	2	22862363
eFondovi	316568	8850	36	22862363
ZIS OSS	307889	84000	4	22862363
Moj račun - Regos	300010	126145	2	22862363
Izabrani liječnik	283266	129070	2	22862363
mojID	262249	145021	2	22862363
Uvjerenje da se ne vodi kazneni postupak	223534	122993	2	22862363
ISSP	161316	74373	2	22862363
Portal zdravlja	152093	34175	4	22862363
Uvid u OIB	140816	83552	2	22862363
Centralna prijava na studijske programe	119688	18694	6	22862363
Realizirani recepti	98251	42366	2	22862363
Komunalne usluge i naknade	88666	46401	2	22862363
Registar birača - e-Privremeni upis	84310	41311	2	22862363
HZZO Otvorene narudžbe	80184	30956	3	22862363
EKZO	76889	46888	2	22862363
Uvjerenje iz kaznene evidencije	71669	45684	2	22862363
e-Obrt	65307	15082	4	22862363
Uvid u ePKK za strane fizičke osobe	56756	17174	3	22862363
eVisitor	51946	8163	6	22862363
eEnergetskiCertifikat	38569	1181	33	22862363
Porezna prijava za obračun i plaćanje posebnih doprinosa	33308	14633	2	22862363
Aplikacija za e-savjetovanje	29381	6945	4	22862363
Izdavanje elektroničke isprave Grada Zagreba	25978	5844	4	22862363
Elektronička komunikacija sa sudovima	22378	1632	14	22862363
Registar birača	20099	12572	2	22862363
SISSI	16977	726	23	22862363
EDIP	16497	194	85	22862363
e-Zahtjev za izdavanje vozačke dozvole	16088	7133	2	22862363
eNovorođenče	15505	5361	3	22862363
e-Pomorac	13649	2963	5	22862363
e-Plovilo	12476	5785	2	22862363
SKDD e-Ulagatelj	11477	5779	2	22862363

Slika 6. Prikaz odabranog skupa podataka nakon pripreme (Izvor: *Portal otvorenih podataka Republike Hrvatske, 2019*)

5.4. Modeliranje podataka

Modeliranje podataka napravljeno je pomoću alata *BigML*. Alat *BigML* je platforma za strojno učenje koja olakšava rješavanje i automatizaciju klasifikacije, regresije, predviđanja vremenske serije, analize klastera, detekcije anomalija itd. Za modeliranje odabranog skupa podataka korištena je klaster analiza. Kao što je već spomenuto, klaster analiza obavlja grupiranje objekata na temelju sličnih karakteristika koje ti objekti posjeduju.

Na slici 7 možemo vidjeti prikaz odabranog skupa podataka koji je uspješno učitano u alat. Prema prikazanom, svaki atribut ima 85 instanci podataka što nam potvrđuje da su svi podaci iz pripremljenog skupa učitani i prepoznati. Također, u zadnjem stupcu možemo vidjeti grafički prikaz vrijednosti pojedinog atributa koji se naziva histogram. Histogram omogućava prepoznavanje podataka koje je potrebno izbaciti. Kao što je već spomenuto u dijelu pripreme podataka za analizu, zadnji atribut pod nazivom „Ukupan_broj_prijava_sve_usluge“ ima istu vrijednost za svaku instancu što znači da neće utjecati na rezultate analize, odnosno ne donosi nikakvu promjenu. To možemo i prepoznati u alatu tako da proučimo histogram atributa te ako je njegov prikaz ujednačen znači da nam taj atribut nije potreban. Također, alat nam to daje do znanja prikazujući crveni uskličnik pokraj atributa.

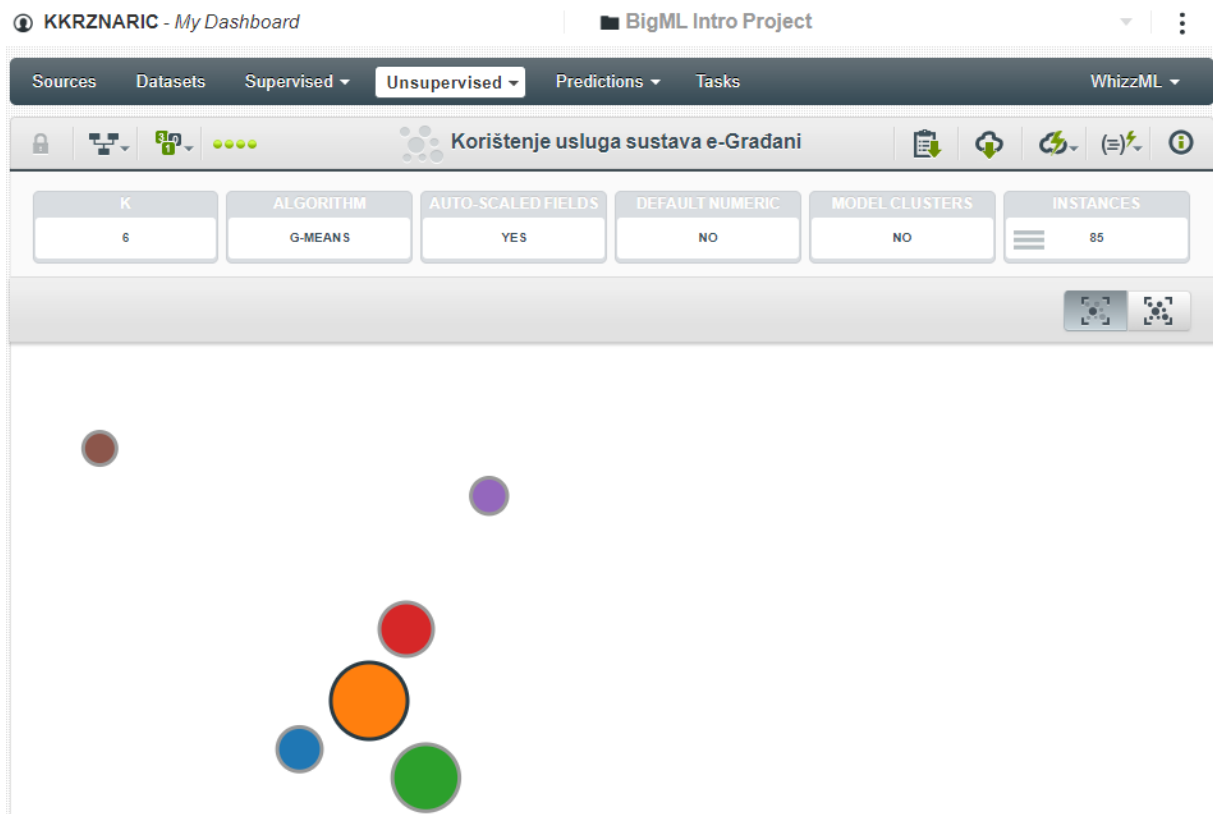


Slika 7. Prikaz skupa podataka u alatu BigML (Izvor: alat za strojno učenje BigML)

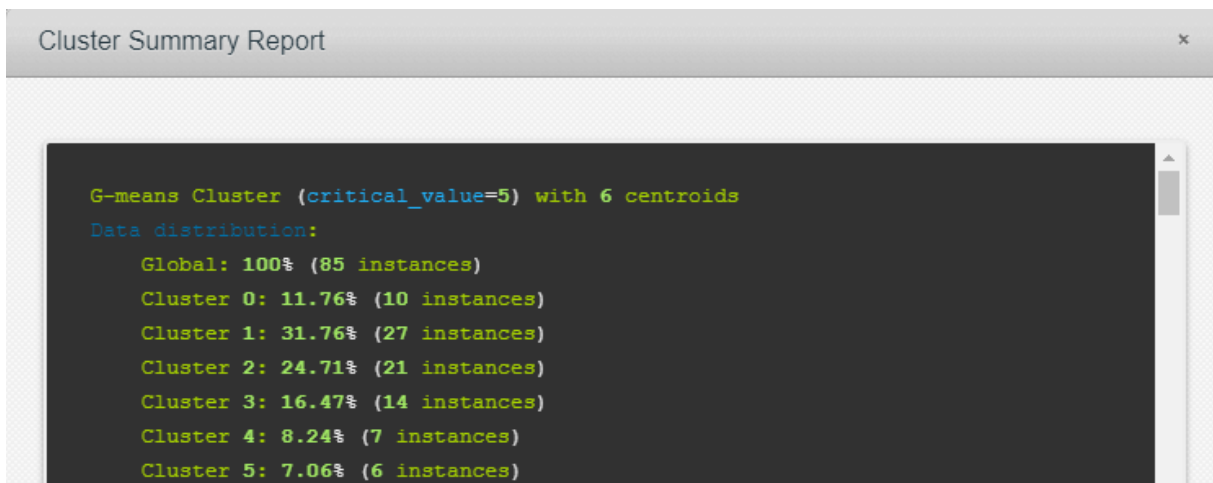
5.4.1. Opis i diskusija rezultata

Rezultati analize bit će prikazani i opisani u ovom poglavlju. Kod klaster analize prvo je korištena metoda k - srednjih vrijednosti. S obzirom da je kod te metode potrebno samostalno odrediti broj klastera koje želimo dobiti, prvo su korišteni najmanji brojevi, ali rezultati takve analize nisu bili dobri te raspored instanci po klasterima nije bio optimalan. Postupno povećanje broja klastera nije rezultiralo boljim rasporedom instanci podataka po klasterima te nikako nije bilo moguće dobiti dobru distribuciju instanci po klasterima. Kako upotrijebljena metoda nije donosila dobre rezultate, za analizu je korištena druga metoda pod nazivom g - srednjih vrijednosti. To je metoda slična prethodnoj metodi, ali kod nje alat sam određuje optimalan broj klastera na temelju svojih algoritama kako bi dobili što bolju distribuciju instanci po klasterima. Nakon pregleda dobivenih rezultata, bilo je vidljivo da je metoda g - srednjih vrijednosti bila uspješna te da je uspješno dobivena optimalna distribucija. Korištenjem te metode ukupno je dobiveno šest klastera. Na slici 8 možemo vidjeti raspored klastera u alatu nakon provedene klaster analize. Također, vidljivo je da su ljubičasti klaster (klaster 4) i smeđi klaster (klaster 5) odvojeni od ostalih klastera i jedan od drugoga dok su ti ostali klasteri bliže jedni drugima, odnosno zajedno. Kako se sličnost između dva klastera temelji na dvije najbliže točke u različitim klasterima, možemo zaključiti da dva odvojena klastera imaju manje sličnosti s ostalim klasterima koji su zajedno.

Kako bi potvrdili da su dobiveni rezultati dobri i optimalni, na slici 9 možemo pogledati izvješće klaster analize koje je alat kreirao. Točnije, prema izvješću možemo vidjeti i potvrditi da nijedan dobiveni klaster nema manje od 5% ili više od 35% instanci podataka što je jako dobra distribucija. Iako ne postoji standard za distribuciju, općeprihvaćeni i preporučeni raspon je najmanje 5%, a najviše 35% instanci podataka po klasteru što je i dobiveno u rezultatima klaster analize.

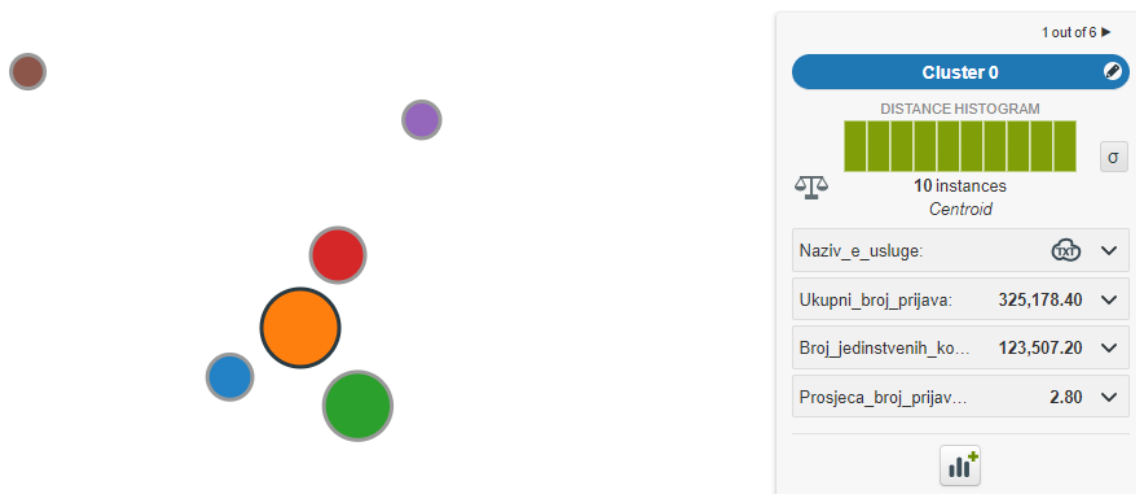


Slika 8. Prikaz rezultata klaster analize (Izvor: alat za strojno učenje BigML)



Slika 9. Sažeti prikaz izvještaja klaster analize (Izvor: alat za strojno učenje BigML)

U nastavku ovog poglavlja objašnjen je svaki klaster te su iznesene i interpretirane neke zakonitosti koje možemo donijeti prema podacima pojedinog klastera.



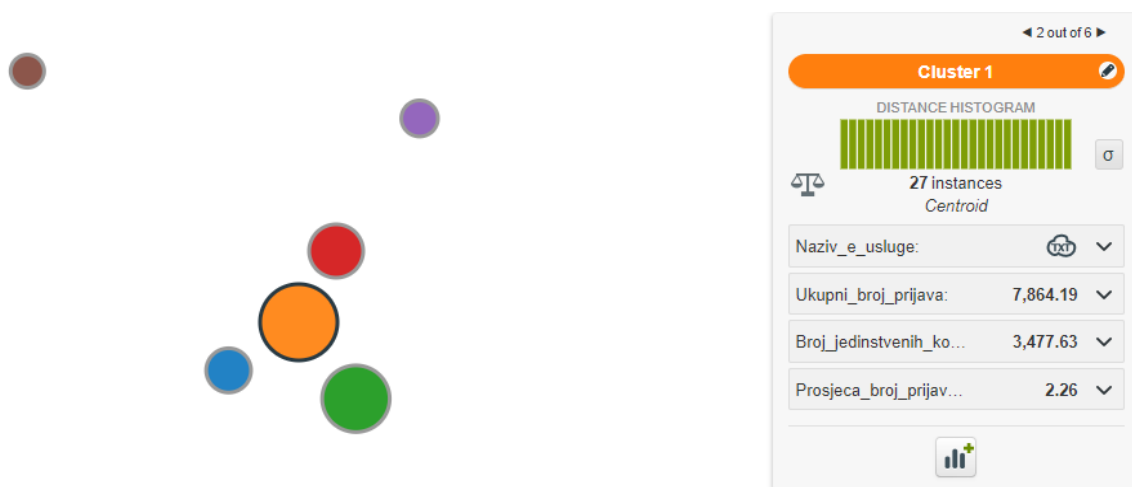
Slika 10. Prikaz podataka klastera 0 (Izvor: alat za strojno učenje BigML)

Naziv_e_usluge	Ukupni_broj_prijava	Broj_jedinstvenih_korisnika	Prosječan_broj_prijava_po_korisniku
Burza rada	840750	111564	8
Elektroničke potvrde	409971	189194	2
REGOS Ena	321983	169160	2
ZIS OSS	307889	84000	4
Moj račun - Regos	300010	126145	2
Izabrani liječnik	283266	129070	2
mojID	262249	145021	2
Uvjerenje da se ne vodi kazneni postupak	223534	122993	2
ISSP	161316	74373	2
Uvid u OIB	140816	83552	2

Slika 11. Prikaz instanci skupa podataka koje se nalaze u klasteru 0 (Izvor: alat za strojno učenje BigML)

Na slici 10 možemo vidjeti da klaster 0 ima ukupno deset instanci podataka. Što se tiče ukupnog broja prijava, koji iznosi oko 325 tisuća, možemo reći da je korištenost ovih usluga umjerena, odnosno imaju srednju korištenost. To bi značilo da se ovaj klaster odnosi na e-usluge koje nisu potrebne često nego na nekoj mjesečnoj ili čak godišnjoj bazi. Broj jedinstvenih korisnika u ovom klasteru relativno je visok što znači da se ovdje radi o uslugama sustava e-Građani kojima građani ne pristupaju više puta nego jednom kada saznaju tražene informacije ili preuzmu potreban dokument one im više nisu potrebne dulje vrijeme. To nam potvrđuje i prosječan broj prijava na pojedinu uslugu po korisniku koji je u ovom klasteru nizak, odnosno iznosi blizu tri. Iako iz ovog klastera nije moguće donijeti neke zakonitosti na temelju instanci podataka koje se nalaze u njemu, iz svega navedenog i prema popisu e-usluga koje se nalaze u ovom klasteru možemo zaključiti da bi se ovaj klaster mogao odnositi na usluge

potrebne pri zapošljavanju, odnosno za prijavu na radno mjesto. E-usluge koje na to upućuju su burza rada koja je važan dio procesa zapošljavanja, elektroničke potvrde pomoću kojih građani mogu dobiti dokumente i potvrde koji su potrebni za zapošljavanje, izabrani liječnik pomoću koje se može provjeriti lista liječnika i obaviti potrebne liječničke preglede prije zapošljavanja, uvjerenje da se ne vodi kazneni postupak kako bi poslodavac imao dokaz o nekažnjavanju itd. Sve su to usluge koje su građanima potrebne, ali jednom kada prikupe sve informacije i dokumente nemaju potrebe za ponovnom prijavom u bliskoj budućnosti. E-usluge koje se nalaze u ovom klasteru mogli bi okarakterizirati kao usluge koje se srednje koriste i predstavljaju grupu e-usluga koje svakom građaninu nude njegove osobne informacije.

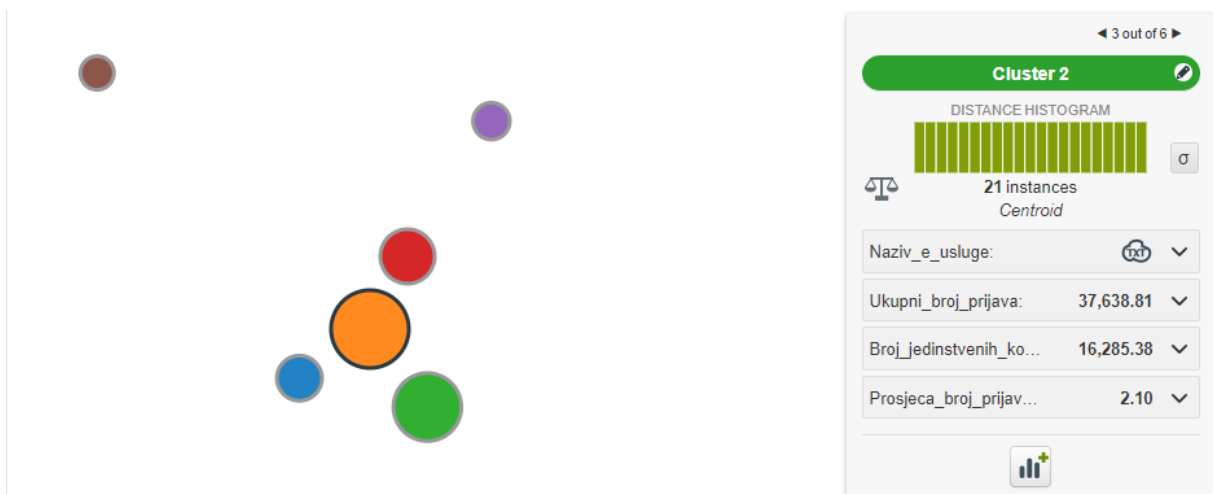


Slika 12. Prikaz podataka klastera 1 (Izvor: alat za strojno učenje BigML)

Naziv_e_usluge	Ukupni_broj_prijava	Broj_jedinstvenih_korisnika	Prosječan_broj_prijava_po_korisniku
Uvid u ePKK za strane fizičke osobe	56756	17174	3
Aplikacija za e-savjetovanje	29381	6945	4
e-Zahtjev za izdavanje vozačke dozvole	16088	7133	2
eNovorođenče	15505	5361	3
e-Plovilo	12476	5785	2
SKDD e-Ulagatelj	11477	5779	2
Online tečajevi Srca	9443	5050	2
e-Razmjena studentskih ocjena	8701	6987	1
NIAS SocSkrb	8606	5645	2
e-Zahtjev za izdavanje ePutovnice	7844	4271	2
ePodnesak	7445	5900	1
MojZagreb	6768	3624	2
ePoljoprivreda	4851	2714	2
e-Detektivi	3908	3643	1
Digitalna komora HGK	2926	2428	1
e-Osiguranje tražbina	2623	2106	1
eDozvola - predaja zahtjeva za gradnju i prostorno uređenje	2452	566	4
ePostupci u području intelektualnog vlasništva	1700	1478	1
e-Savjetovanja	762	315	2
e-Imovinske kartice	659	196	3
e-Ovlaštenja	592	198	3
Postupci vezani uz članstvo Hrvatske komore arhitekata	559	173	3
e-Ovlaštenja za kontrolu projekata	267	180	1
Metaregistar	220	53	4
eDozvola	201	160	1
CTS uvid u predmete	79	23	3
Uvid u članstvo u obveznom mirovinskom fondu	44	9	5

Slika 13. Prikaz instanci skupa podataka koje se nalaze u klasteru 1 (Izvor: alat za strojno učenje BigML)

Kada pogledamo klaster 1 možemo vidjeti da je on puno veći što se tiče broja instanci od klastera 0. Taj klaster ujedno je i najveći klaster u analizi, a sadrži dvadeset sedam instanci podataka što čini 31.76% od cijelog skupa podataka. Ukupni broj prijava u ovom klaster je vrlo nizak i odmah možemo zaključiti da se tu radi o uslugama koje se malo koriste, odnosno koje nisu od velikog značaja za građane Hrvatske. Nadalje, prosječan broj prijava po korisniku na određenu uslugu je nizak što znači da građani vrlo rijetko koriste ove usluge i da ponovno korištenje iste usluge nema veliku svrhu. Možemo pretpostaviti da se ovdje radi o uslugama za koje građani još nisu svjesni da postoje u elektroničkom obliku te da ih mogu koristiti preko interneta umjesto odlazaka u ustanove koje se bave tim uslugama. Za primjer možemo uzeti primjer iz svakodnevnog života vezan uz izdavanje ili produljivanje vozačke dozvole. Puno ljudi ne zna da postoji e-usluga preko koje se može dati zahtjev za izdavanje ili produljivanje vozačke dozvole zbog čega odlaze u policijsku postaju gdje je taj proces dugotrajniji. Kada još uzmemo u obzir dolazak do tamo i čekanje u redu, korištenje ove usluge preko interneta bi svima olakšalo život. Zaključak je da ovaj klaster predstavlja grupu e-usluga koje se ne koriste u velikoj mjeri ili nisu još prepoznate kao korisne među građanima Hrvatske.

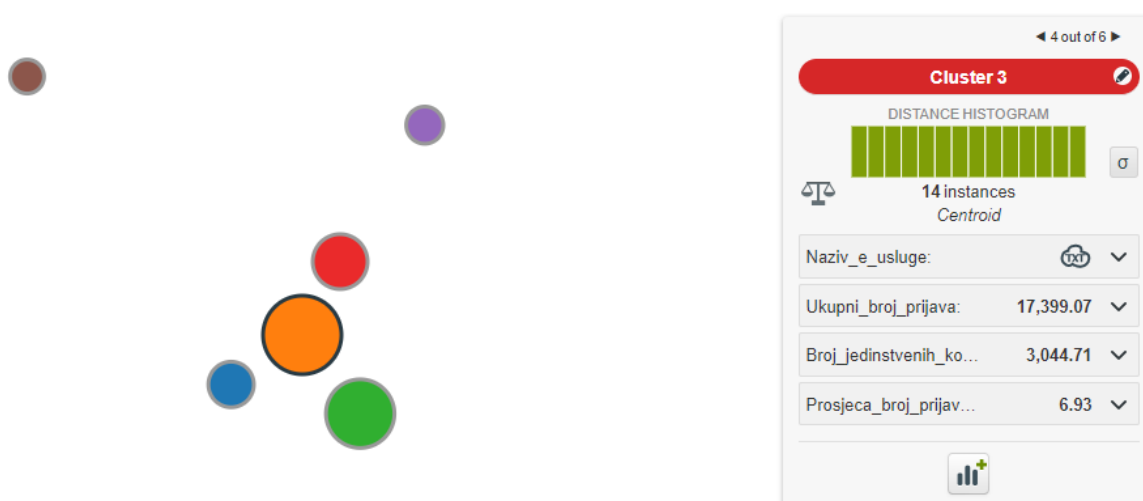


Slika 14. Prikaz podataka klastera 2 (Izvor: alat za strojno učenje BigML)

Naziv_e_usluge	Ukupni_broj_prijava	Broj_jedinstvenih_korisnika	Prosjecan_broj_prijava_po_korisniku
Portal zdravlja	152093	34175	4
Realizirani recepti	98251	42366	2
Komunalne usluge i naknade	88666	46401	2
Registar birača - e-Privremeni upis	84310	41311	2
HZZO Otvorene narudžbe	80184	30956	3
EKZO	76889	46888	2
Uvjerenje iz kaznene evidencije	71669	45684	2
e-Obrt	65307	15082	4
Porezna prijava za obračun i plaćanje posebnog poreza na motorna vozila	33308	14633	2
Registar birača	20099	12572	2
Registracija objekata koji pružaju uslugu smještaja strancima	6082	4282	1
Vodne usluge Međimurskih voda	4617	2035	2
Suglasnosti i punomoći u postupcima iz djelokruga MUP-a	4151	2605	2
Portal otvorenih podataka	3035	1629	2
Home for homeless servis sustava AAI@EduHr	949	799	1
Obvezni mirovinski fond (prijava/promjena) - pilot faza	539	415	1
eKomunalno baza komunalnih naknada i doprinosa	195	127	2
Nacionalni informacijski sustav prijava i upisa u srednje škole	35	16	2
Moj EKZO	19	9	2
eIDAS usluga	13	5	3
Registracija operatora bespilotnih zrakoplova za kategorije B2 i C1	4	3	1

Slika 15. Prikaz instanci skupa podataka koje se nalaze u klasteru 2 (Izvor: alat za strojno učenje BigML)

Sljedeći klaster je klaster 2 koji sadrži dvadeset jednu instancu. Ovo je klaster s nešto većim brojem ukupnih prijava, ali je i dalje nizak u odnosu na ukupne prijave u ostalim klasterima. Prosječan broj prijava po korisniku na određenu uslugu je također nizak što znači da korisnici ne koriste istu uslugu vrlo često. Možemo zaključiti da usluge u ovom klasteru nisu od velike važnosti za svakodnevni život građana i poslovanje ljudi pa se iz tog razloga i ne koriste u velikoj mjeri. U ovom klasteru se kao najkorištenije spominju e-usluge vezane uz zdravlje što je upućuje na to da postoje usluge sustava e-Građani koje pomažu građanima u zdravstvenim situacijama i prepoznate su kao korisne.

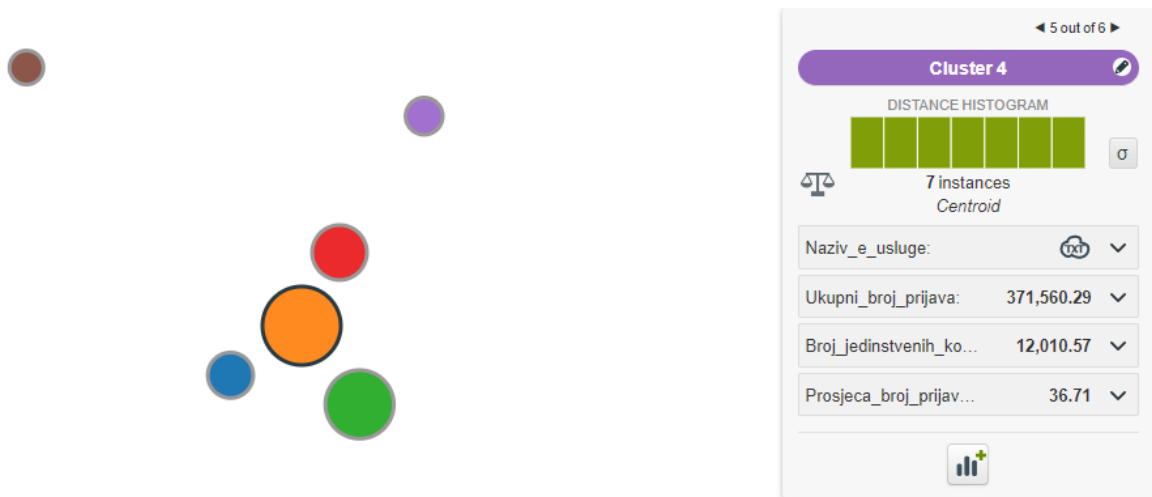


Slika 16. Prikaz podataka klastera 3 (Izvor: alat za strojno učenje BigML)

Naziv_e_usluge	Ukupni_broj_prijava	Broj_jedinstvenih_korisnika	Prosjecan_broj_prijava_po_korisniku
Centralna prijava na studijske programe	119688	18694	6
eVisitor	51946	8163	6
Izdavanje elektroničke isprave Grada Zagreba	25978	5844	4
Elektronička komunikacija sa sudovima	22378	1632	14
e-Pomorac	13649	2963	5
e-Prijava boravišta hrvatskih državljana	8380	4779	2
DGU - Sustav za polaganje stručnih ispita i izdavanje potvrda	760	439	2
CARNET Login	315	38	8
eInstalateriOIE	172	29	6
e-Pristojbe - Modul korisničkog sučelja za tijela javne vlasti	100	20	5
e-Pristojbe - Modul za upravljanje matičnim podacima	87	5	17
e-Potvrde za učenike	68	12	6
Slanje e-poruka	44	5	9
ePIOTGradjevniProizvodi	22	3	7

Slika 17. Prikaz instanci skupa podataka koje se nalaze u klasteru 3 (Izvor: *alat za strojno učenje BigML*)

Na slici 16 možemo vidjeti klaster 3 koji sadrži četrnaest instanci. Ukupan broj prijava na e-usluge u ovom je klasteru vrlo malen što upućuje na to da se neke od usluga ne koriste puno, ali njegova je specifičnost u tome što je prosječan broj prijava po korisniku na određenu e-uslugu znatno veći u odnosu na većinu klastera. Na temelju toga možemo zaključiti da se ovdje radi o uslugama na koje se građani prijavljuju više puta. Također, vidljivo je da se ovdje radi o e-uslugama koje građani koriste za poslovanje, a na neke od njih se je potrebno prijaviti više puta tjedno ili mjesečno radi prirode posla. U pitanju su usluge poput eVisitor-a preko kojeg iznajmljivači prijavljuju i odjavljuju svoje goste, e-Pristojbe preko koje građani mogu plaćati svoje upravne pristojbe i naknade te usluga Elektroničke komunikacije sa sudovima preko koje odvjetnici i javni bilježnici na tjednoj ili mjesečnoj bazi šalju podneske i priloge sudu te primaju sudska pismena. Navedene usluge imaju velik prosječan broj prijava po korisniku u odnosu na cijeli skup podataka što nam potvrđuje zašto su smještene u isti klaster. Kako se kod ovog klastera radi o e-uslugama koje imaju veći prosječan broj prijava po korisniku u odnosu na ostale klastere, možemo reći da se ovdje nalaze i usluge koje koriste i akademski građani Hrvatske. Tijekom školovanja i studiranja često je potrebno doći do raznih dokumenata poput potvrda i isprava, a to rezultira mnogim prijavama na usluge. Akademski građani su većinom mladi ljudi, informatički pismeni, te su dobro upućeni u dodatne mogućnosti koje nudi sustav e-Građani. Ovaj zaključak potvrđuju i e-usluge koje se nalaze u ovom klasteru poput CARNET login-a, e-Potvrde za učenike te Centralne prijave na studijske programe.



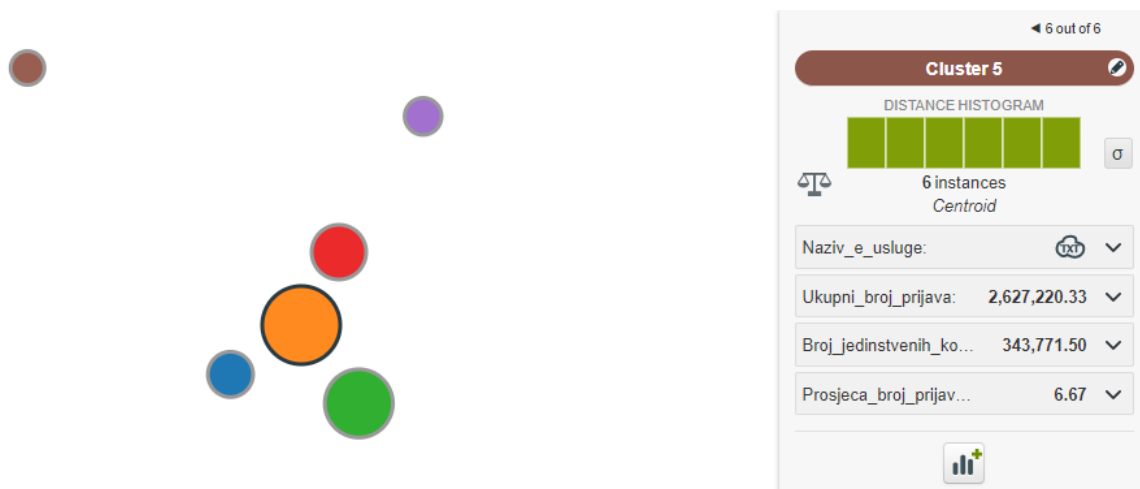
Slika 18. Prikaz podataka klastera 4 (Izvor: alat za strojno učenje BigML)

Naziv_e_usluge	Ukupni_broj_prijava	Broj_jedinstvenih_korisnika	Prosječan_broj_prijava_po_korisniku
e-Dnevnik	2211461	73089	30
eFondovi	316568	8850	36
eEnergetskiCertifikat	38569	1181	33
SISSI	16977	726	23
EDIP	16497	194	85
RegistarEUsluga	622	25	25
Predaja zahtjeva za dodjelom OIB-a	228	9	25

Slika 19. Prikaz instanci skupa podataka koje se nalaze u klasteru 4 (Izvor: alat za strojno učenje BigML)

Klaster 4 sastoji se od sedam instanci što ga čini jednim od manjih klastera dobivenih analizom. Kao što se može vidjeti na slici 18, ukupan broj prijava nije malen što nam govori da se u ovom klasteru nalaze e-usluge koje se koriste u većoj mjeri i daju građanima Hrvatske određenu korist. Najveću pažnju u ovom klasteru privlači prosječan broj prijava po korisniku na određenu uslugu koji iznosi oko trideset sedam. Možemo vidjeti da sve e-usluge iz ovog klastera imaju jako velik prosječan broj prijava po korisniku te je to razlog zašto su smještene u isti klaster. Spomenuta vrijednost je vrlo velika te predstavlja veliko odstupanje od ostalih klastera. Iz svega navedenog možemo zaključiti da se kod ovog klastera radi o e-uslugama koje imaju velik broj prijava po korisniku. To bi značilo da se u ovom klasteru nalaze e-usluge koje se koriste na dnevnoj ili tjednoj bazi s obzirom da se korisnici često prijavljuju na njih. Za pretpostaviti je da e-usluge iz ovog klastera građani koriste u svakodnevnom životu i poslovanju. Možemo reći da se tu radi o uslugama vezanima uz praćenje financija s obzirom da je svakom građaninu i zaposleniku bitno da redovito prati svoje financijsko stanje. Također, ovdje se spominju e-usluge koje se koriste u školstvu što je razumljivo s obzirom da su

obrazovne ustanove mjesta gdje se svakodnevno koriste razne usluge koje imaju velik broj prijava po korisniku.



Slika 20. Prikaz podataka klastera 5 (Izvor: alat za strojno učenje BigML)

Naziv_e_usluge	Ukupni_broj_prijava	Broj_jedinstvenih_korisnika	Prosječan_broj_prijava_po_korisniku
Osobni korisnički pretinac	8641288	556958	16
eUsluge Porezne uprave	1940707	289631	7
HZMO korisničke stranice	1539079	210274	7
e-Matice	1362492	391490	3
HZMO elektronički zapis	1359474	310395	4
e-Usluge MUP-a	920282	303881	3

Slika 21. Prikaz instanci skupa podataka koje se nalaze u klasteru 5 (Izvor: alat za strojno učenje BigML)

Zadnji dobiveni klaster ujedno je i najmanji klaster koji se sastoji od samo šest instanci. Na slici 20 možemo vidjeti da je ukupan broj prijava u ovom klaster daleko veći od ukupnog broja prijava u ostalim klasterima. Sukladno tome, očigledno je da se ovaj klaster odnosi na one e-usluge koje se najviše koriste u sustavu e-Građani i koje su najpotrebnije građanima Hrvatske. Prosječan broj prijava po korisniku na određenu uslugu je relativno visok što je još jedan od pokazatelja velike korištenosti e-usluga koje se nalaze u klasteru 5. Većina e-usluga spomenutog klastera ima milijun ili nekoliko milijuna ukupnih prijava što su vrlo velike brojke kada pogledamo koliko Hrvatska ima stanovnika. Ovim klasterom pokazano je da postoji određeni broj e-usluga sustava e-Građani koje se koriste u velikoj mjeri te građanima Hrvatske olakšavaju poslovanje i svakodnevni život. Te usluge odnose se na dobivanje osobnih informacija i preuzimanje osobnih dokumenata, olakšavanje poreznim obveznicima pregled poreznih obveza, izdavanje dokumenata MUP-a o prebivalištu i boravištu itd. Kada pogledamo e-usluge koje se nalaze u ovom klasteru možemo zaključiti da građani Hrvatske najviše koriste

one e-usluge koje im pružaju određene osobne podatke, dokumente i uvide u poslovanje odnosno informacije korisne za osobni život i vlastito poslovanje.

Nakon analize svakog pojedinog klastera mogu se izdvojiti najvažnije zakonitosti koje su dobivene, odnosno prepoznate iz podataka. Klaster analizom dobivene su grupe podataka pomoću kojih je bilo moguće donijeti kvalitetne zaključke i zakonitosti. Na kraju možemo reći da postoji velik broj e-usluga koje se često koriste i pomažu građanima Hrvatske da lakše i brže dođu do potrebnih informacija, dokumenata i sl. Također, e-usluge koje se najviše koriste vezane su uz školovanje i studiranje, što znači da su akademski građani prepoznali prednosti sustava e-Građani i koriste ga u velikoj mjeri. S druge strane, prepoznato je da pojedine e-usluge imaju veliku ulogu i u poslovanju te poslovnim ljudima olakšavaju vođenje poslovanja. To se najviše odnosi na olakšano praćenje i plaćanje naknada koje je potrebno podmiriti tijekom poslovanja. U rezultatima je primijećeno da postoje pojedine e-usluge koje imaju jako velik broj prijava po korisniku, a odnose se većinom na praćenje financija poput dohodaka i primitaka, prijave za financiranje projekata, praćenje pristojbi itd. Za razliku od e-usluga koje imaju nekoliko tisuća ili milijuna prijava, pojedine usluge imaju vrlo nizak broj prijava. Velika je mogućnost da su te usluge nedovoljno razvijene ili građanima jednostavno ne pružaju ono što je njima potrebno. Te e-usluge većinom se odnose na neke manje zastupljene i razvijene grane poslovanja u Hrvatskoj. Nadalje, prema rezultatima možemo reći da postoji znatan broj e-usluga koje su korisne, ali građani jednostavno ne prepoznaju njihovu korisnost, nisu upućeni u njihovo postojanje u elektroničkom obliku ili nisu informatički pismeni da bi se njima koristili. Korisnike sustava najviše zanimaju informacije koje su usko povezane uz njih kao pojedinca te kako mogu što brže i efikasnije doći do dokumenta, informacija o porezima, zdravstvu te o zapošljavanju, a manje su im bitne informacije vezane uz opće stanje države. Na kraju opisa rezultata možemo zaključiti da većina e-usluga sustava e-Građani donose prednosti u svakodnevnom životu građana Hrvatske te da je izrada ovakvog sustava bila korak u pravom smjeru kako bi se potaknula informatizacija javne uprave. Smatram da je bitno ulagati u razvitak ovakvih sustava, ali još bitnije je potaknuti građane na korištenje e-usluga. Važno je educirati građane o prednostima korištenja e-usluga i široj javnosti dati do znanja da postoje razne prednosti elektroničkih usluga. Korištenje spomenutog sustava, u većem broju, smanjuje opterećenje javne uprave što u konačnici rezultira boljim poslovanjem.

6. Zaključak

Otvoreni podaci danas imaju sve veću ulogu u životu ljudi i svakodnevnom poslovanju. Prepoznate su prednosti i koristi koje donose otvoreni podaci te se zbog toga sve više ulaže u njihov razvoj i analizu. Postoje razne mogućnosti koje nude ovi podaci, a na nama je da ih probamo što kvalitetnije iskoristiti s ciljem postizanja novih znanja. Živimo u svijetu neprestanog napretka na području informacijske tehnologije i važno je da iskoristimo svu tehnologiju i dostupne podatke za dobivanje novih spoznaja koje mogu olakšati živote mnogih ljudi. Otvoreni podaci imaju važnu ulogu u postizanju transparentnosti zemlje kojom se stvara povjerenje između države i njezinih građana. Također, svaka država trebala bi educirati svoje građane o korisnosti otvorenih podataka i poticati ih na korištenje u vlastite svrhe ili u poslovne svrhe. Prednosti spomenutih podataka prepoznala je i Republika Hrvatska koja zadnjih nekoliko godina sve intenzivnije radi na razvoju otvorenih podataka. Hrvatska je pokretanjem politike otvorenih podataka i razvojem Portala otvorenih podataka Republike Hrvatske započela s uvođenjem ovog koncepta te je danas svrstana u skupinu zemalja predvodnica na području otvorenih podataka. Danas postoji sve veća potražnja za radnom snagom koja ima potrebna znanja da iskoristi svu količinu podataka koja se danas nudi. Mnoge organizacije svoje poslovanje temelje na analizi podataka te su im potrebni kvalificirani radnici kojih danas nema dovoljno što je jedan od problema na ovom području. U ovom radu predstavljeno je nekoliko najpoznatijih tehnika rudarenja podataka. Bitno je napomenuti da svaka od tih tehnika zahtjeva puno znanja i uloženog vremena kako bi se napravila kvalitetna analiza i interpretacija rezultata. Također, dobiveni modeli i predikcije ne mogu uvijek dati točne i sigurne rezultate koji se očekuju zato što se često radi o velikom broju faktora koji utječu na krajnji ishod. Neovisno o tome, rudarenje podataka ima veliku primjenu i sigurno će u budućnosti donijeti nova znanja i koristi za ljudske živote.

Nad otvorenim podacima bitno je raditi razne analize kako bi otkrili neke zakonitosti u njima ili dobili nova znanja. U ovom radu napravljena je klaster analiza kojom su otkrivene pojedine zakonitosti na području korištenja e-usluga u Hrvatskoj. Važno je za Hrvatsku i njezine građane da se razvija što više ovakvih sustava zato što uvelike olakšavaju rad i poslovanje. Analizom je utvrđeno da građani Hrvatske koriste mnoge e-usluge u velikoj mjeri što je dobar korak prema informatizaciji javne uprave i njezine povezanosti s privatnim sektorom. Pojedine e-usluge imaju manju posjećenost zbog manjka informiranosti građana i nedovoljne razvijenosti tih e-usluga, ali njihovim daljnjim razvojem te poticanjem i educiranjem građana o prednostima njihovog korištenja sigurno će se i to promijeniti na bolje. Dobiveni rezultati analize korištenog skupa otvorenih podataka mogu poslužiti kao nit vodilja za daljnji razvoj sustava e-Građani ili nekih sličnih sustava.

Popis literature

- [1] Čupić, M., Bašić, B. D., Golub, M. (2013). *Neizrazito, evolucijsko i neuroračunarstvo*.
- [2] Europski portal podataka (2015). *Prednosti uporabe podataka*. Preuzeto 18.8.2019. s <https://www.europeandataportal.eu/hr/using-data/benefits-of-open-data>
- [3] Forum za javnu upravu (2015). *Uprava u digitalno doba: otvoreni podaci / 8. forum za javnu upravu*. Zagreb: Friedrich Ebert Stiftung, ured za Hrvatsku
- [4] Garbade, M. J. (2018). *Understanding K-means Clustering in Machine Learning*. Preuzeto 22.8.2019. s <https://towardsdatascience.com/understanding-k-means-clustering-in-machine-learning-6a6e67336aa1>
- [5] Han, J., Kamber, M., Pei, J. (2012). *Data Mining Concepts and Techniques* (3. izd.). Elsevier Inc.
- [6] Jalil, M. S. (2016) *E-service Innovation: A Case Study of Shahoz.com*. Preuzeto 24.8.2019. s <https://www.sciencedirect.com/science/article/pii/S187704281630516X>
- [7] Larose, D.T., Larose, C.D. (2015). *Data Mining and Predictive Analytics* (2. izd.). John Wiley & Sons, Inc., Hoboken, New Jersey
- [8] Musa, A., Bevandić, D., Herak, D., Jadrijević, L., Kovačić, M., Luša, Z., Vrčec, N. (2018). *Priručnik o ponovnoj uporabi informacija za tijela javne vlasti: Otvoreni podaci za sve*. Zagreb: Povjerenik za informiranje
- [9] Open data handbook. (bez dat.). *Čemu služe otvoreni podaci?* Preuzeto 16.8.2019. s <http://opendatahandbook.org/guide/hr/why-open-data/>
- [10] Pantano, E., Priporas, C. V., Stylos, N. (2017). „You will like it!“ *Using open data to predict tourists' response to a tourist attraction*. Preuzeto 24.8.2019. s <https://www.sciencedirect.com/science/article/pii/S0261517716302680#sec5>
- [11] Portal otvorenih podataka Republike Hrvatske (2015). *Što je Portal otvorenih podataka Republike Hrvatske?* Preuzeto 17.8.2019. s <https://data.gov.hr/o-portalu-cijeli>
- [12] Portal otvorenih podataka Republike Hrvatske. (bez dat) *e-Građani – statistika*. Preuzeto 24.8.2019. s <https://data.gov.hr/dataset/e-gradjani-statistika/resource/b0aeeb96-b97d-41a4-b0b7-95472af30ff6>
- [13] Središnji državni portal (2018). *Politika otvorenih podataka*. Preuzeto 20.8.2019. s <http://data.gov.hr/library/politika-otvorenih-podataka>
- [14] Two Crows Corporation (2005). *Introduction to Data Mining and Knowledge Discovery* (3. izd.). Preuzeto 23.8.2019. s http://www.twocrows.com/intro-dm.pdf?bcsi_scan_998937D63ACB8A9E=0&bcsi_scan_filename=intro-dm.pdf
- [15] Zhu, Z., Wulder, M. A., Roy, D. P., Woodcock, C. E., Hansen, M. C., Radeloff, V. C., Haeley, S. P., Schaaf, C., Hostert, P., Strobl, P., Pekel, J. F., Lymburner, L., Pahlevan,

N., Scambos, T. A. (2019). *Benefits of the free and open Landsat data policy*. Preuzeto 24.8.2019. s <https://www.sciencedirect.com/science/article/pii/S0034425719300719>

[16] Alat BigML. (2019). Preuzeto s <https://bigml.com/>

Popis slika

Slika 1: Prednosti otvorenih podataka.....	3
Slika 2: Korisnici otvorenih podataka.....	4
Slika 3: Faze u životnom ciklusu rudarenja podataka.....	9
Slika 4: Primjer slojevite neuronske mreže.....	12
Slika 5: Prikaz odabranog skupa podataka.....	17
Slika 6: Prikaz odabranog skupa podataka nakon pripreme.....	19
Slika 7: Prikaz skupa podataka u alatu BigML.....	20
Slika 8: Prikaz rezultata klaster analize.....	22
Slika 9: Sažeti prikaz izvještaja klaster analize.....	22
Slika 10: Prikaz podataka klastera 0.....	23
Slika 11: Prikaz instanci skupa podataka koje se nalaze u klasteru 0.....	23
Slika 12: Prikaz podataka klastera 1.....	24
Slika 13: Prikaz instanci skupa podataka koje se nalaze u klasteru 1.....	24
Slika 14: Prikaz podataka klastera 2.....	25
Slika 15: Prikaz instanci skupa podataka koje se nalaze u klasteru 2.....	26
Slika 16: Prikaz podataka klastera 3.....	26
Slika 17: Prikaz instanci skupa podataka koje se nalaze u klasteru 3.....	27
Slika 18: Prikaz podataka klastera 4.....	28
Slika 19: Prikaz instanci skupa podataka koje se nalaze u klasteru 4.....	28
Slika 20: Prikaz podataka klastera 5.....	29
Slika 21: Prikaz instanci skupa podataka koje se nalaze u klasteru 5.....	29

Popis tablica

Tablica 1: Najaktivniji izdavači otvorenih podataka u Republici Hrvatskoj do 2019.godine.....6