

Analiza globalne ekonomske nejednakosti korištenjem otkrivanja znanja u bazama podataka

Markoč, Zdeslav

Master's thesis / Diplomski rad

2023

Degree Grantor / Ustanova koja je dodijelila akademski / stručni stupanj: **University of Zagreb, Faculty of Economics and Business / Sveučilište u Zagrebu, Ekonomski fakultet**

Permanent link / Trajna poveznica: <https://urn.nsk.hr/urn:nbn:hr:148:995462>

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

Download date / Datum preuzimanja: **2024-07-25**



Repository / Repozitorij:

[REPEFZG - Digital Repository - Faculty of Economics & Business Zagreb](#)



Sveučilište u Zagrebu

Ekonomski fakultet

Integrirani preddiplomski i diplomski studij

Poslovna ekonomija – smjer Menadžerska informatika

**ANALIZA GLOBALNE EKONOMSKE NEJEDNAKOSTI
KORIŠTENJEM OTKRIVANJA ZNANJA U BAZAMA
PODATAKA**

Diplomski rad

Zdeslav Markoč

Zagreb, veljača 2023.

Sveučilište u Zagrebu

Ekonomski fakultet

Integrirani preddiplomski i diplomski studij

Poslovna ekonomija – smjer Menadžerska informatika

**ANALIZA GLOBALNE EKONOMSKE NEJEDNAKOSTI
KORIŠTENJEM OTKRIVANJA ZNANJA U BAZAMA
PODATAKA**

**GLOBAL ECONOMIC INEQUALITY ANALYSIS USING
KNOWLEDGE DISCOVERY IN DATABASES**

Diplomski rad

Student: Zdeslav Markoč

JMBAG studenta: 0067550760

Mentorica: Prof. dr. sc. Mirjana Pejić Bach

Zagreb, veljača 2023.

Sažetak i ključne riječi

Globalna ekonomska nejednakost prisutna je u svim sferama života te izravno pogađa stanovnike svih zemalja svijeta te igra značajnu ulogu u osnovnim procesima svakog gospodarstva. Možemo ju promatrati kao nejednakost unutar država, ali isto tako kao nejednakost između različitih svjetskih gospodarstava i time uočiti veliki jaz između životnog standarda stanovnika bogatih i razvijenih svjetskih zemalja te onih koji nisu imali tolike sreće da se rode u dobrim uvjetima, stanovnika zemalja u razvoju. Fokus ovog diplomskog rada bit će na ekonomskoj nejednakosti između zemalja, ne zanemarujući nejednakosti unutar zemlje. Cilj ovog rada je analiza globalne ekonomske nejednakosti te iskoristiti otkrivanje znanja u bazama podataka za ispitivanje nejednakosti i razumijevanje čimbenika koji tome pridonose.

Podaci nad kojima je provedena analiza te samo istraživanje ovog rada preuzeti su sa službenih stranica Svjetske banke. Korištenjem softvera otvorenog koda Weka napravljena je klaster analiza te analiza, tj. prognoziranje vremenskih serija. Analizirana su tri pokazatelja u trima odabranim godinama (1995., 2005. te 2015.): BDP per capita, Ginijev koeficijent te očekivana životna dob pri rođenju. Klaster analizom podataka BDP-a per capita prikazana je velika razlika između vrijednosti centroida najbogatijeg klastera te najsiromašnijeg klastera. Analizom je uočen i pozitivan trend puno bržeg rasta vrijednosti BDP-a per capita zemalja siromašnijih klastera od zemalja bogatijih klastera. Predviđanjem vremenskih serija napravljen je pogled u budućnost za 15 narednih godina. Analiza je napravljena nad podacima BDP-a per capita nad ukupno 16 svjetskih zemalja, a pokazala je kako zemlje koje pripadaju siromašnijim klasterima pokazuju veći potencijal za značajnim gospodarskim rastom te prelaskom u rangove bogatih zemalja.

Ključne riječi: globalna ekonomska nejednakost, otkrivanje znanja, baze podataka, klaster analiza, analiza vremenskih serija

Summary and keywords

Global economic inequality is present in all spheres of life and directly affects the inhabitants of all countries of the world and plays a significant role in the basic processes of every economy. We can see it as inequality within countries, but also as inequality between different world economies and thus notice a large gap between the standard of living of residents of rich and developed world countries and those who were not so lucky to be born in good conditions, residents of developing countries. The focus of this thesis will be on economic inequality between countries, without neglecting inequalities within the country. The aim of this paper is to analyze global economic inequality and use the discovery of knowledge in databases to examine inequalities and understand the factors that contribute to them.

The data on which the analysis was carried out and the research of this paper were taken from the official website of the World Bank. Using open-source software Wek, a cluster of analysis and analysis of time series. Three indicators were analyzed in three selected years (1995, 2005, and 2015): GDP per capita, Gini coefficient, and life expectancy at birth. Cluster analysis of GDP per capita data shows a large difference between the value of the centroid of the richest cluster and the poorest cluster. The analysis also showed a positive trend of much faster growth of GDP values per capita in countries of poorer clusters than in countries of richer clusters. Predicting the time series is a glimpse into the future for the next 15 years. The analysis was made on GDP per capita data over a total of 16 world countries and showed that countries belonging to poorer clusters show greater potential for significant economic growth and transition to the ranks of rich countries.

Keywords: global economic inequality, knowledge discovery, databases, cluster analysis, time series analysis

IZJAVA O AKADEMSKOJ ČESTITOSTI

Izjavljujem i svojim potpisom potvrđujem da je diplomski rad isključivo rezultat mog vlastitog rada koji se temelji na mojim istraživanjima i oslanja se na objavljenu literaturu, a što pokazuju korištene bilješke i bibliografija. Izjavljujem da nijedan dio rada nije napisan na nedozvoljen način, odnosno da je prepisan iz necitiranog rada, te da nijedan dio rada ne krši bilo čija autorska prava. Izjavljujem, također, da nijedan dio rada nije iskorišten za bilo koji drugi rad u bilo kojoj drugoj visokoškolskoj, znanstvenoj ili obrazovnoj ustanovi.

U Zagrebu, _____

Student: _____

(potpis)

STATEMENT ON THE ACADEMIC INTEGRITY

I hereby declare and confirm by my signature that the final thesis is the sole result of my own work based on my research and relies on the published literature, as shown in the listed notes and bibliography. I declare that no part of the thesis has been written in an unauthorized manner, i.e., it is not transcribed from the non-cited work, and that no part of the thesis infringes any of the copyrights. I also declare that no part of the thesis has been used for any other work in any other higher education, scientific or educational institution.

U Zagrebu, _____

Student: _____

(potpis)

Sadržaj

1.	Uvod.....	1
1.1.	Predmet i cilj rada.....	1
1.2.	Izvor podataka i metode prikupljanja	1
1.3.	Sadržaj i struktura rada	2
2.	Globalna ekonomska nejednakost.....	3
2.1.	Pojmovno određenje i značajke ekonomske nejednakosti	3
2.2.	Ekonomska nejednakost unutar zemlje	7
2.2.1.	Kuznetsovi valovi.....	8
2.3.	Ekonomska nejednakost između zemalja	11
2.4.	Učinci ekonomske nejednakosti u svijetu	15
3.	Otkrivanje znanja u bazama podataka	17
3.1.	Baze podataka.....	18
3.2.	Otkrivanje znanja u bazama podataka	19
3.2.1.	Definicija poslovnog problema	21
3.2.2.	Priprema podataka.....	21
3.2.3.	Modeliranje	22
3.2.4.	Implementacija	22
3.3.	Metode otkrivanja znanja u bazama podataka.....	23
3.3.1.	Klaster analiza	23
3.3.2.	Prognoziranje vremenskih serija	26
3.4.	Programi za otkrivanje znanja u bazama podataka	27
3.4.1.	Weka softver	28
4.	Analiza globalne ekonomske nejednakosti.....	30
4.1.	Metodologija istraživanja	30
4.2.	Korišteni podaci.....	31

4.3.	Rezultati istraživanja	34
4.3.1.	Klaster analiza	34
4.3.2.	Analiza vremenske serije	71
4.4.	Analiza rezultata	76
4.4.1.	Klaster analiza	76
4.4.2.	Analiza vremenske serije	83
5.	Zaključak	85
	Literatura.....	87
	Popis slika	89
	Popis tablica.....	91
	Popis grafikona	92

1. Uvod

1.1. Predmet i cilj rada

Tema globalne ekonomske nejednakosti izravno pogađa stanovnike svih zemalja svijeta, odnosno sve potrošače, poslodavce, investitore i kućanstva svih globalnih ekonomija. Globalna ekonomska nejednakost igra veliku ulogu u svakodnevnom životu stanovnika svake nacionalne ekonomije. Gledajući povijesne podatke, prije dva stoljeća svijet je bio mjesto puno veće jednakosti, doduše, jednakosti u siromaštvu. Neka su svjetska gospodarstva od tada značajno napredovala, poput Švedske, čiji su stanovnici postali 30 puta bogatiji, ali mnoge druge zemlje jedva da su napredovale. Zbog globalne nejednakosti, oni koji su rođeni u zemljama koje su postigle značajan ekonomski rast i razvoj odrastaju s mnogo boljim životnim uvjetima i više mogućnosti, dok oni koji nisu te sreće odrastaju u siromaštvu s vrlo malo, često minimalnih mogućnosti. Fokus ovog diplomskog rada bit će na ekonomskoj nejednakosti između zemalja, ne zanemarujući nejednakosti unutar zemlje. Cilj ovog rada je analiza globalne ekonomske nejednakosti te iskoristiti otkrivanje znanja u bazama podataka za ispitivanje nejednakosti i razumijevanje čimbenika koji tome pridonose.

1.2. Izvor podataka i metode prikupljanja

Podatci koji će se koristiti za provođenje istraživanja, tj. samu analizu globalne ekonomske nejednakosti korištenjem otkrivanja znanja u bazama podataka preuzeti su sa službenih stranica Svjetske banke (data.worldbank.org). Korištenjem softvera Weka napravit će se klaster analiza te analiza, tj. prognoziranje vremenskih serija. Analizirat će se tri pokazatelja, a to su BDP per capita (kako bi se analizirala ekonomska nejednakost između zemalja), Ginijev koeficijent (koji prikazuje dohodovne nejednakosti unutar država) te očekivana životna dob pri rođenju (koja ukazuje na razvijenost zemlje i kvalitetu života u istoj). Podaci su priloženi samome radu na zadnjim stranicama istoga.

1.3. Sadržaj i struktura rada

Diplomski se rad sastoji od ukupno pet poglavlja. Nakon uvoda, drugo poglavlje obrađuje temu globalne ekonomske nejednakosti. Kroz poglavlje o globalnoj ekonomskoj nejednakosti pojmovno će se odrediti te prikazati značajke ekonomske nejednakosti. Prolaskom kroz tematiku ekonomske nejednakosti unutar zemlje, ali zatim i između zemalja razjasnit će se razlika između ta dva pojma, ali isto tako i naglasiti korelacija istih. Naposljetku poglavlja, kroz rad će se prikazati učinci ekonomske nejednakosti u svijetu. Treće poglavlje donosi tematiku otkrivanja znanja u bazama podataka koja će biti pokrivena prolaskom te detaljniji uvid u svijet baza podataka, otkrivanja znanja u bazama podataka te metodama otkrivanja znanja u bazama podataka. Programi za otkrivanje znanja u bazama podataka posljednji je dio poglavlja, ali ne manje važan budući da će se kroz isti prikazati softver korišten u samom istraživanju koje slijedi u četvrto, ujedno i predzadnjem poglavlju ovoga rada. Analiza globalne ekonomske nejednakosti naslov je četvrtog poglavlja koje prikazuje metodologiju istraživanja, korištene podatke u istome te zatim rezultate samog istraživanja i njihovu analizu. Posljednje poglavlja rada je zaključak kroz koji će biti prikazan kratki osvrt na cjelokupni rad te rezultate provedene analize.

2. Globalna ekonomska nejednakost

2.1. Pojmovno određenje i značajke ekonomske nejednakosti

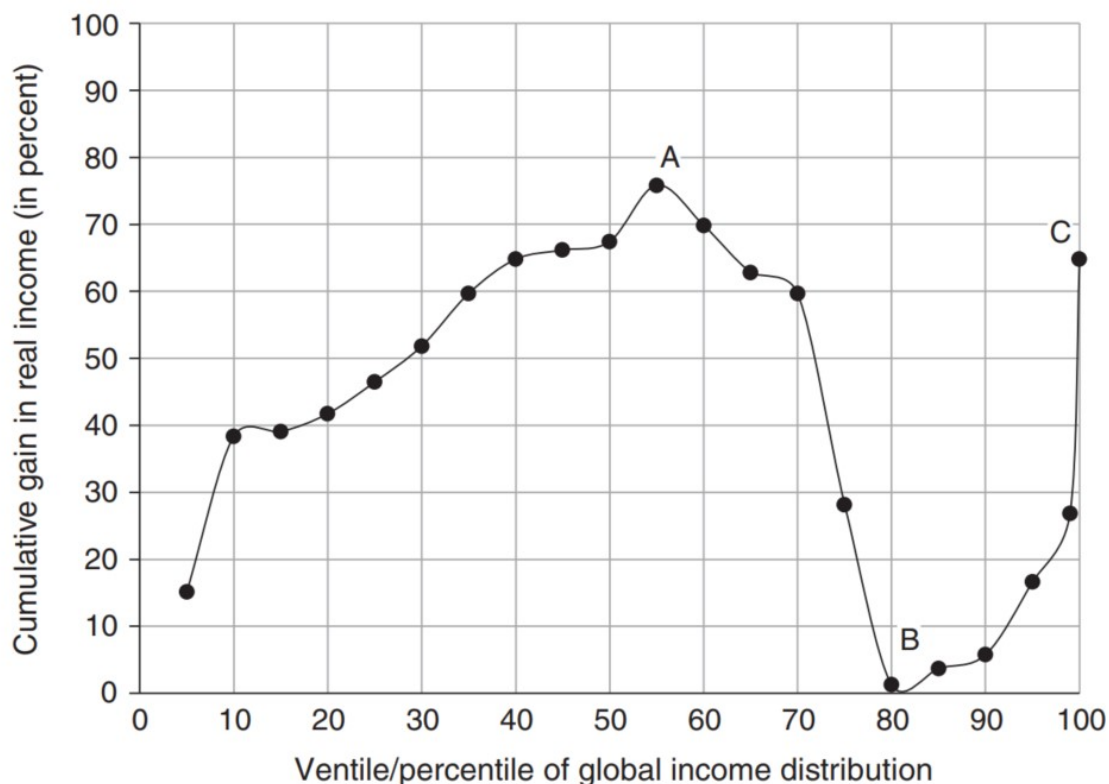
Kao mnogi drugi pojmovi u društvenim znanostima, i pojam nejednakosti u ekonomskoj znanosti ima široko značenje. Najjednostavnije predstavljeno, nejednakost podrazumijeva nedostatak jednake raspodjele predmetne veličine koja se raspoređuje po određenim jedinicama analize, bile te jedinice grupne ili pojedinačne (Dragičević, 2016). Grupna nejednakost, odnosno horizontalna nejednakost (Stewart, Brown i Mancini, 2005) pojava je koja podrazumijeva situaciju u kojoj određenoj grupi nije distribuiran ravnomjerni dio ukupnog zbroja jedinica koji se dijeli. Slično, ali donekle drugačije opisujemo pojedinačnu, tj. vertikalnu nejednakost kao situaciju u kojoj pak pojedincu nije ravnomjerno raspoređen njegov ili njen dio ukupnog zbroja dijeljenih jedinica. Kako Stewart u svome radu ističe, nejednakost je mjerljiva na različitim razinama, pa se tako ista može mjeriti na razini pojedinaca, grupa, država i na svjetskom nivou te isto tako na razini različitih tipova grupa poput grupa na temelju spola, etniciteta, rase, starosti, ideologiji, itd (Stewart, Brown i Mancini, 2005).

Globalnu nejednakost, tj. nejednakost u dohotku građana možemo promatrati kao zbroj svih nacionalnih nejednakosti kojem pridružujemo zbroj ukupnih razlika u srednjim vrijednostima dohodaka između zemalja. Zbroj svih nacionalnih nejednakosti bavi se nejednakosti u prihodima bogatih i siromašnih Amerikanaca, bogatih i siromašnih Meksikanaca, i tako dalje. Druga se komponenta bavi razlikama u dohotku između zemalja, kao na primjer razlika u dohotku između Sjedinjenih Američkih Država, Meksika, Španjolske, Maroka, i tako dalje (Milanovic, 2016).

Promatrajući ekonomski pokazatelj relativnog dobitka u realnom BDP-u per capita na razini svjetskog dohotka za razdoblje takozvane „visoke globalizacije“ koje možemo svrstati između godina 1988. te 2008., odnosno između Pada berlinskog zida te Svjetske financijske krize, dolazimo do zaključka kako takozvani dobitci nisu ravnomjerno distribuirani među svjetskom populacijom. Navedeni je vremenski raspon razdoblje komunikacijsko tehnološke revolucije koja je omogućila poduzećima da realociraju

proizvodne pogone u svjetske zemlje nižeg troška rada te potaknu globalizaciju (Milanovic, 2016).

Grafikon 1 Relativni dobitak realnog dohotka po stanovniku prema razini globalnog dohotka 1988. - 2008.



Izvor: Milanovic, B. (2016) *Global Inequality: A New Approach for the Age of Globalization*. Cambridge, MA and London, England: Harvard University Press

Na prikazanom grafikonu upravo se maloprije spomenuto može očitati, relativni dobitak realnog dohotka po stanovniku prema razini globalnog dohotka za razdoblje „visoke globalizacije“ od 1988. do 2008. godine. U točkama A i C možemo vidjeti takozvane pobjednike globalizacije čiji se dohodak značajno povećao, dok u točki B očitavamo takozvane gubitnike.

Promotrimo prvo pobjednike u točki A prikazanoj na grafu. U devet od deset slučajeva radi se o ljudima iz azijskih gospodarstava u usponu, pretežno Kine, ali i Indije, Tajlanda, Vijetnama i Indonezije. Oni nisu najbogatiji ljudi u tim zemljama, jer su bogati

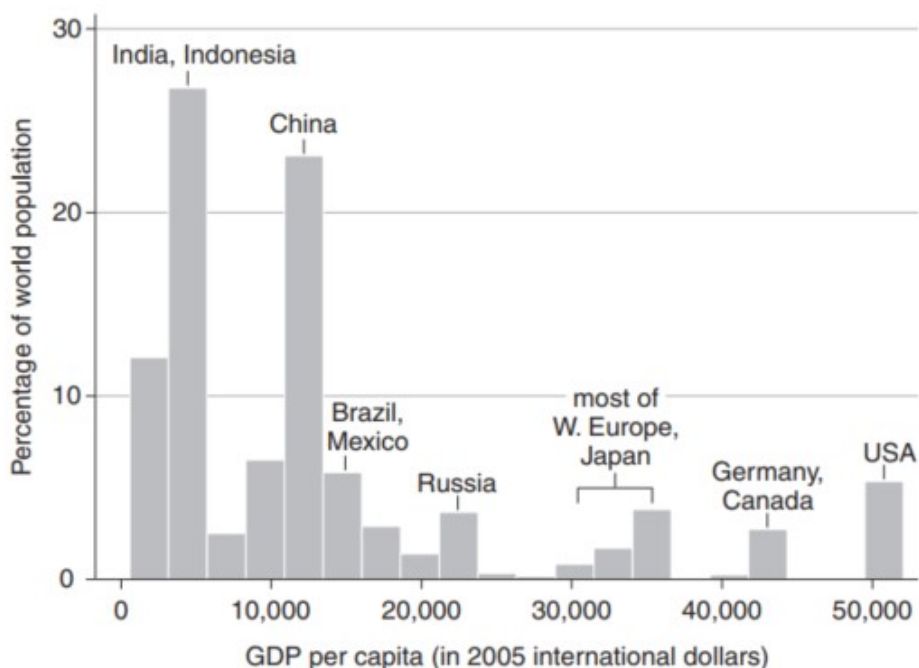
pozicionirani više u globalnoj distribuciji dohotka. To su ljudi u sredini distribucije u svojim zemljama, ali i u samome svijetu. Te su skupine bile glavni pobjednici globalizacije između 1988. i 2008., a nazivamo ih "globalna srednja klasa u nastajanju" kako bismo ih ipak mogli razlikovati od zapadne srednje klase od koje su, ipak, i dalje relativno siromašniji (Milanovic, 2016).

U točki B prikazanoj na grafu možemo uočiti takozvane gubitnike globalizacije, ali su navedeni gubitnici ipak i dalje bogatiji od ranije spomenute skupine u točki A. Također, primjećujemo da je vrijednost na okomitoj osi u točki B gotovo nula, što ukazuje na odsutnost bilo kakvog rasta realnog dohotka tijekom dvadeset godina. Gotovo svi koji se nalaze u točki B su iz bogatih gospodarstava OECD-a. Ako zanemarimo one među njima koji su relativno nedavno postale članicama OECD-a č, odnosno nekoliko zemalja istočne Europe, Čile i Meksiko, oko tri četvrtine ljudi u ovoj skupini su građani takozvane staro bogate zemlje zapadne Europe, Sjeverne Amerike te Oceanije i Japana. Na isti način kako Kina dominira u točki A, SAD, Japan i Njemačka dominiraju u točki B. Ljudi u točki B općenito pripadaju nižim polovinama raspodjele dohotka svojih zemalja. Oni su iz donjih pet decila u Njemačkoj, koja je od 1988. do 2008. uspjela ostvariti kumulativni rast od samo između 0 i 7 posto; iz donje polovice raspodjele dohotka u SAD-u, koji je doživio realni rast između 21 i 23 posto; i iz nižih decila u Japanu, gdje je zabilježen ili pad realnog dohotka ili ukupnog rasta od 3 do 4 posto. Ove ljude koji sigurno nisu pobjednici globalizacije možemo nazvati i nižom srednjom klasom bogatog svijeta (Milanovic, 2016).

U točki C, ujedno i zadnjoj točki, prikazani su globalno gledajući iznimno bogati ljudi čiji su realni dohodci značajno porasli u razdoblju od 1988. do 2008. te ih također smatramo pobjednicima globalizacije, kao i spomenutu globalnu srednju klasu u nastajanju prikazanu u točki A (Milanovic, 2016).

Usporedbom skupina B i C može se doći do još jednog vrlo važnog zaključka. Vidjeli smo da se skupina B, s nultom ili zanemarivom dobiti od globalizacije, sastoji uglavnom od niže srednje klase i siromašnijih segmenata stanovništva bogatih zemalja. Nasuprot tome, skupina C, takozvani pobjednici globalizacije, sastoji se od bogatijih klasa iz istih zemalja. Očita je implikacija da su se jazovi u dohotku između vrha i dna povećali u bogatom svijetu i da je globalizacija pogodovala onima koji su u bogatim zemljama već bili u boljem položaju. Ovo također nije posve iznenađujuće, budući da je općenito poznato da su se nejednakosti unutar nacija u bogatom svijetu povećale tijekom posljednjih 25 do 30 godina (Milanovic, 2016).

Grafikon 2 Raspodjela svjetske populacije prema realnom BDP-u po glavi stanovnika zemlje u kojoj ljudi žive (2013. godina)



Izvor: Milanovic, B. (2016) *Global Inequality: A New Approach for the Age of Globalization*. Cambridge, MA and London, England: Harvard University Press

Na prethodno prikazanom grafikonu prikazana je raspodjela svjetske populacije prema realnom BDP-u per capita zemlje u kojoj ljudi žive te je vrlo jasno i lako uočljivo kako vrlo mali dio svjetskog stanovništva živi u srednjim razinama prihoda. Zaključak pri promatranju grafikona bio bi kako takozvana "praznina u sredini" još uvijek uvelike karakterizira distribuciju svjetskog stanovništva prema prosječnom dohotku (ili BDP-u po glavi stanovnika) zemlje u kojoj ljudi žive (Milanovic, 2016).

2.2. Ekonomska nejednakost unutar zemlje

Pri analizi ekonomske nejednakosti unutar zemalja postoji pojam opisan od strane skupine autora Milanovic, Lindert i Williamson koji ne smije biti zanemaren, a to je takozvana granica mogućnosti nejednakosti (engl. Inequality possibility frontier). Koncept se temelji na ideji da postoji kompromis između ekonomskog rasta i nejednakosti dohotka, te da različita društva mogu odabrati različite točke u tom kompromisu ovisno o svojim vrijednostima i prioritetima. Granica mogućnosti nejednakosti predstavlja maksimalnu razinu dohodovne nejednakosti koja se može postići uz održavanje određene razine ekonomskog rasta (Milanovic, Lindert i Williamson, 2011).

U poznatom djelu "Zašto nacije propadaju" Acemoglua i Robinsona, autori tvrde da su ključna odrednica uspjeha nacije njene ekonomske institucije, koje definiraju kao pravila koja upravljaju načinom na koji se resursi raspoređuju i koriste. Oni tvrde da nacije propadaju kada su njihove institucije ekstraktne, što znači da koriste malobrojnoj eliti nauštrb većine građana. S druge strane, nacije uspijevaju kada su njihove institucije inkluzivne, što znači da koriste većini građana i potiču gospodarski rast. Argumenti koje autori prikazuju u svome djelu pokazuju kako su ekstraktne institucije u zemljama poput pred kolonijalne Afrike i pred moderne Europe dovele do siromaštva i stagnacije, dok su inkluzivne institucije u zemljama poput srednjovjekovne Europe i postkolonijalne Afrike dovele do prosperiteta i rasta. Oni također pokazuju kako je uspon ekstraktnih institucija u nekim zemljama, poput Rusije i Kine, doveo do povećane nejednakosti i smanjenog gospodarskog rasta. Zaključno, autori kroz svoje djelo tvrde da je nejednakost među zemljama uglavnom uzrokovana načinom na koji su dizajnirane njihove ekonomske institucije i da su uključive institucije ključne za dugoročni prosperitet i rast zemlje (Acemoglu i Robinson, 2012). Navedeno ukazuje na visok stupanj korelacije između unutarnje ekonomske nejednakosti, ekonomskih i političkih prilika i ne prilika s jedne strane te vanjske ekonomske nejednakosti, odnosno ekonomske nejednakosti između zemalja.

2.2.1. Kuznetsovi valovi

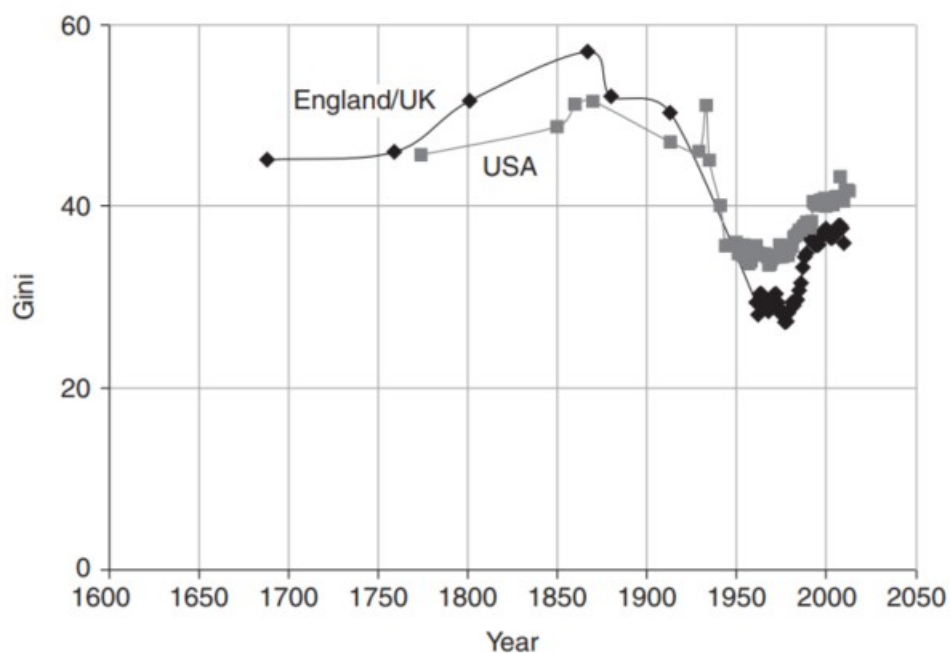
Posljednjih je desetljeća kroz nove ekonomske pokazatelje poznata Kuznetsova teorija, ideja da je nejednakost niska na vrlo niskim razinama dohotka, a zatim raste kako se gospodarstvo razvija i na kraju opet pada na visokim razinama dohotka djelomično, ako ne i u potpunosti opovrgnuta. Porast nejednakosti u visini dohotka građana bogatih zemalja upravo je glavni razlog zbog čega Kuznetsovu hipotezu ne možemo uzeti za u potpunosti ispravnu. Nedvojbena porast nejednakosti u Sjedinjenim Državama, Ujedinjenom Kraljevstvu, pa čak i u nekim prilično egalitarnim zemljama poput Švedske i Njemačke, jednostavno je nespojiv s Kuznetsovom hipotezom (Milanovic, 2016).

Nedavni rast nejednakosti u razvijenim i bogatim državama koji se kosi s Kuznetsovom hipotezom, pokušali su objasniti znanstvenici Tinbergen, kao i Goldin i Katz prikazom koncepta utrke između obrazovanja i onoga što je poznato kao tehnološki napredak usmjeren na vještine. Naime, fenomen porasta nejednakosti u razvijenim i bogatim državama ovim se konceptom pokušao objasniti kroz tvrdnju da su nadnice kvalificiranijih radnika porasle su više nego plaće manje kvalificiranih. U spomenutoj utrci ne postoji nijedan teorijski argument koji bi nam rekao pod kojim uvjetima možemo očekivati da će utrku pobijediti tehnologija, čime se povećava nejednakost, a pod kojim uvjetima obrazovanje, čime se smanjuje nejednakost. U Tinbergenovoj izvornoj formulaciji, međutim, utrku je trebalo pobijediti obrazovanje, s više visokokvalificiranih ljudi koji su postajali sve brojniji kako su zemlje postajale bogatije i s vještinama koje su tako ublažile učinke tehnološke promjene. Zbog toga je Tinbergen očekivao da će takozvana premija za vještinu pasti na nulu. Ali i ovdje se dogodilo upravo suprotno: premija za vještinu pokazala je snažan porast u većini naprednih zemalja tijekom proteklih dvadeset godina. Tinbergenova teorija, kao i Kuznetsova ukazuje na to kako će se nejednakost smanjivati s razvojem, a na temelju ekonomskih pokazatelja možemo reći kako je to svakako zaključak koji je nedvosmisleno u suprotnosti s činjenicama (Milanovic, 2016).

Osim spomenutih autora i znanstvenika, Thomas Piketty je kroz svoj rad „Kapital u 21. stoljeću“ pokušao objasniti spomenuti fenomen te u potpunosti opovrgnuti Kuznetsovu hipotezu. Piketty je tvrdio kako je do smanjenja nejednakosti došlo zbog neobičnih i posebnih uvjeta i događaja koji su bili obilježeni političkim snagama ratova, oporezivanjem za financiranje ratova, socijalističkom ideologijom i pokretima te ekonomskom konvergencijom koja je stopu rasta plaća držala iznad stope rasta prihoda od

imovine. Takozvana normalna kapitalistička konstelacija, pod kojom danas živimo, donosi, prema Pikettyjevom mišljenju, rastuću nejednakost, kao što je to bilo u razdoblju prije Prvog svjetskog rata. Ova teorija stoga objašnjava oba dijela Kuznetsove krivulje - koja je, prema Pikettyjevu mišljenju, u obliku slova U, a ne obrnutog oblika U, kako je mislio Kuznets (Milanovic, 2016).

Grafikon 3 Nejednakost u Engleskoj/UK i Sjedinjenim Državama od 17. stoljeća do 21. stoljeća



Izvor: Milanovic, B. (2016) *Global Inequality: A New Approach for the Age of Globalization*. Cambridge, MA and London, England: Harvard University Press

No, ako pogledamo grafikon broj 3 te obratimo posebnu pozornost na razdoblje 18. i 19. stoljeća, vidimo porast nejednakosti koju Pikettyjeva teorija ne objašnjava (Milanovic, 2016). Na grafikonu je prikazano kretanje razine Ginijevog koeficijenta, mjere koja se koristi za izračunavanje nejednakosti u dohotku koja postoji između građana određene države. Vrijednost Ginijevog koeficijenta kreće se između 0 i 100 (ili pak 0 i 1) te pritom 0 označava potpunu jednakost, a vrijednost od 100 označava potpunu nejednakost. U ekonomskoj se znanosti nejednakost grafički može prikazati još i Lorenzovom krivuljom.

Kroz svoj je rad, autor Branko Milanovic 2016. godine predstavio teoriju Kuznetsovih valova koja nadopunjuje Kuznetsovu hipotezu te objašnjava promjene u nejednakosti u razdoblju prije industrijske revolucije, razdoblju koje je uslijedilo do revolucije Reagan-Thatcher te najnovijem razdoblju u kojem trenutno živimo (Milanovic, 2016.).

Razdoblje prije Industrijske revolucije autor je okarakterizirao kao razdoblje u kojemu nije postajala pozitivna ili negativna korelacija između prosječnog prihoda te razine nejednakosti. Plaće i nejednakost povećavali su ili smanjivali idiosinkratski događaji kao što su epidemije, nova otkrića poput otkrića Amerike ili novih trgovačkih putova između Europe i Azije, invazije i ratovi. Kada bi se nejednakost smanjila kako bi prosječni dohodak i plaće porasle, a siromašni bi postali malo manje siromašni, populacija bi se povećala do neodrživih razina i na kraju bi bila smanjena višim stopama smrtnosti među siromašnima zbog smanjenog prosječnog dohotka. To bi gurnulo siromašne natrag na egzistencijalnu razinu i podiglo nejednakost na prethodnu, višu razinu (Milanovic, 2016.).

S dolaskom Industrijske revolucije te pratećim porastom u razini prosječnog prihoda, situacija se mijenja te plaće generalno gledajući rastu zajedno s prihodom. Autor Milanovic ističe dvije najvažnije činjenice vezano za kretanje razine nejednakosti u vrijeme Industrijske revolucije. Prvo, nejednakost se sada može povećati više nego prije jer veći ukupni dohodak omogućuje dijelu stanovništva da uživa mnogo veće prihode bez da sve ostale spusti ispod točke gladovanja. Drugo, nakon industrijske revolucije, nejednakost i prosječni dohodak ušli su u odnos koji prije nije postojao, kada je srednji dohodak bio fiksiran (Milanovic, 2016.).

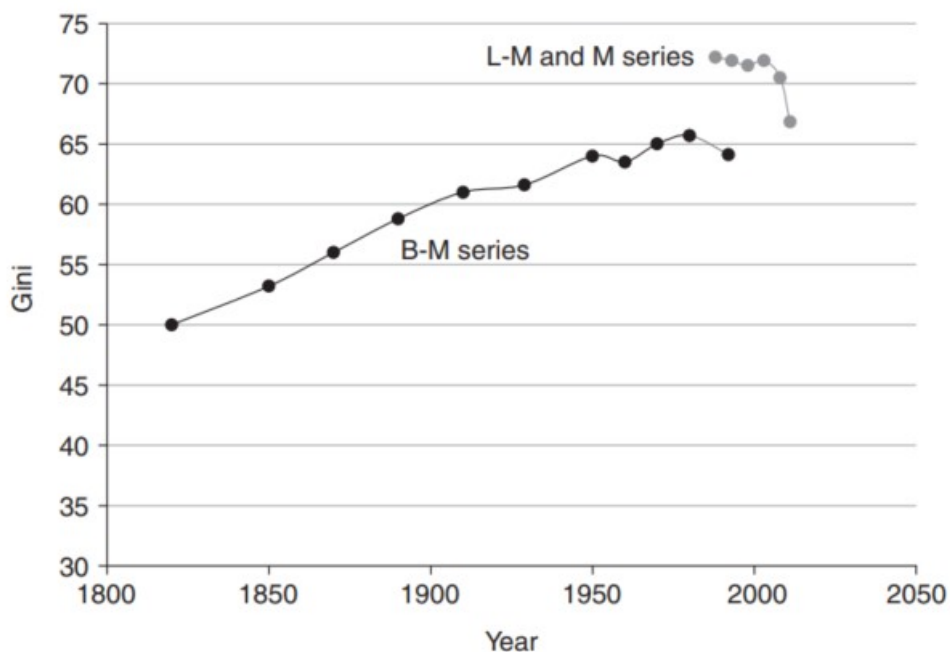
Prvi svjetski rat autor Milanovic ističe kao silu koja je uzrokovala smanjenje nejednakosti, a ta je sila kroz većinu 20. stoljeća imala svoj utjecaj zahvaljujući, nažalost, prvom i drugom svjetskom ratu, kao i brojnim sukobima za vrijeme trajanje takozvanog hladnog rata. Sile koje su smanjile nejednakost nakon Prvog svjetskog rata prestale su djelovati do kraja 1980-ih. U to vrijeme datiramo početak druge Kuznetsove krivulje za bogate zemlje. To je desetljeće otvorilo novu, odnosno drugu tehnološku revoluciju, obilježenu značajnim promjenama u informacijskoj tehnologiji, globalizaciji i sve većoj važnosti heterogenih poslova u uslužnom sektoru. Ova je revolucija, kao i industrijska revolucija s početka devetnaestog stoljeća, povećala razlike u prihodima. Nove tehnologije koje su snažno nagrađivale više kvalificiranu radnu snagu, povećanje udjela i povrata kapitala, sve otvorenija gospodarstva bogatih zemalja konkurenciji iz Kine i Indije, sve je to dovelo do djelomičnog povećanja nejednakosti. Struktura potražnje, a time i radnih mjesta,

pomaknula se prema uslužnim djelatnostima, u kojima je radila manje kvalificirana i lošije plaćena radna snaga. S druge strane, neki poslovi u uslužnom sektoru, poput financija, bili su iznimno visoko plaćeni. Ovo je proširilo raspodjelu plaća, a naposljetku i dohotka (Milanovic, 2016.).

2.3. Ekonomska nejednakost između zemalja

Kako je prethodno u radu spomenuto, globalnu nejednakost između svjetskih zemalja možemo promatrati kao zbroj svih nacionalnih nejednakosti kojem pridružujemo zbroj ukupnih razlika u srednjim vrijednostima dohodaka između zemalja. Zbroj svih nacionalnih nejednakosti bavi se nejednakosti u prihodima bogatih i siromašnih građana svijeta.

Grafikon 4 Globalna nejednakost, 1820. - 2011.



Izvor: Milanovic, B. (2016) *Global Inequality: A New Approach for the Age of Globalization*. Cambridge, MA and London, England: Harvard University Press

Grafikon broj 4 po redu prikazuje kretanje Ginijevog koeficijenta u razdoblju od 1820. do 2011. godine kojim se vrlo jasno prikazuje procijenjena nejednakost prihoda između svih građana svijeta u posljednja dva stoljeća, na temelju tri različita, ali povezana izvora. Vidljivo je iz grafikona da se nejednakost povećavala sve do kraja dvadesetog stoljeća i da je od tada u opadanju. Serija B-M koristi međunarodne dolare iz 1990., a serije L-M i M koriste međunarodne dolare iz 2005. te navedeno objašnjava prekid u grafu (Milanovic, 2016.).

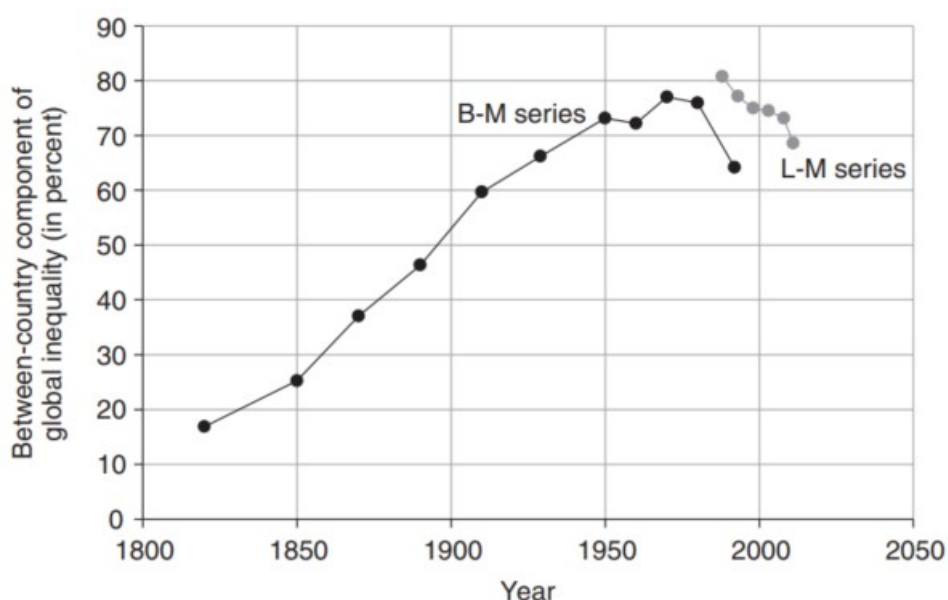
Od kasnih 1980-ih do otprilike prijelaza u dvadeset i prvo stoljeće, razina globalne nejednakosti bila je relativno konstantna, oscilirajući na nešto iznad 70 Ginijevih bodova. Detaljna analiza pokazuje da ta stabilnost ovisi o Kini: ako se Kina isključi iz izračuna, globalna Gini vrijednost raste tijekom vremena. Kina je do 2000. godine imala ulogu glavnog globalnog lidera u ujednačavanju prihoda, a nakon 2000. pridružuje joj se i Indija (Milanovic, 2016.). Stopa gospodarskog rasta Kine znatno je viša od svjetskog prosjeka od 1977. te s obzirom na nizak inicijalni prihod, vrlo se lako očekuje da djeluje kao sila izjednačavanja. U manjoj mjeri isto može vrijediti za Indiju, koja je rasla sporije od Kine, ali još uvijek brže od svjetskog prosjeka od 1980. godine (Anand i Segal, 2008.). Te su zemlje najprije držale pod kontrolom rast globalne nejednakosti, a potom su pridonijele smanjenju ukupne razine nejednakosti. Otprilike od 2000. postoje nepogrešivi znakovi smanjenja globalne nejednakosti, a gotovo svaka sljedeća godina za koju postoje podaci, pokazuje blagi pad Ginijevog koeficijenta (Milanovic, 2016.).

No, treba uzeti nekoliko dodatnih faktora u obzir. Kao prvo, spomenuti rezultati prikazani u grafikonu broj 4 prikazuju pad nejednakosti u relativno kratkom vremenskom periodu. Kao drugo, pad nejednakosti u svijetu nastao je kao posljedica napretka Azijskih zemalja uz usporavanje gospodarskog rasta razvijenih zapadnih zemalja. Za pretpostaviti je kako će se trend smanjenja pada globalne ekonomske nejednakosti smanjivati, tj. jenjavati kako gospodarski rast Azijskih zemalja u razvoju bude jenjavao. Treća i posljednja, ali ne manje bitna činjenica koju treba uzeti u obzir je ta da pri istraživanju dohotka građana vrlo lako dolazi do pogrešnog računanja ukupnog prihoda najbogatijih jedan posto građana, s obzirom na različite izvore i kompleksnost izvora prihoda. Iako podaci iz 2011. pokazuju

da je pad globalne nejednakosti prilično strm, ako bismo bili konzervativni, a u ovakvim područjima bismo trebali biti, točniji zaključak bi bio da dokazi upućuju na to da je globalna nejednakost dohotka stabilna ili u padu. Konzervativnija, tj. opreznija izjava bila bi da nema dokaza o rastućoj globalnoj nejednakosti dohotka, a razlika u dohotku između zapadne i azijske srednje klase očito se smanjuje (Milanovic, 2016.).

No, pitanje koje se postavlja pri analizi globalne ekonomske nejednakosti između zemalja i unutar zemalja je svakako – što više utječe na globalnu nejednakost, nejednakosti između zemalja ili nejednakosti unutar zemalja? Odgovor na to pitanje može dati grafikon u nastavku.

Grafikon 5 Udio međudržavne komponente u globalnoj nejednakosti, 1820. - 2011.



Izvor: Milanovic, B. (2016) *Global Inequality: A New Approach for the Age of Globalization*. Cambridge, MA and London, England: Harvard University Press

Na prethodnom grafikonu broj 5 vidljivo je za koliki je postotak globalne nejednakosti, mjereno Theilovim ili Theilovim entropijskim indeksom, zaslužna nejednakost između zemalja, odnosno jazovima između nacionalnih dohodaka po glavi stanovnika. Povećanjem navedenog udjela govori nam kako nacionalni prosječni prihod postaje važnija karika od individualnih okolnosti određene osobe u određivanju ekonomske nejednakosti. Prednost Theilovog indeksa nad Ginijevim je ta što je on u potpunosti djeljiv, tj. razgradiv

između klase i lokacije. Element lokacije bio je skoro pa zanemariv 1820. godine jer je samo 20 posto globalne nejednakosti bilo uzrokovano nejednakostima između zemalja. Velika većina globalne nejednakosti proizlazila je, u to vrijeme, samo iz činjenice u kakvu je osoba obitelj ili društvenu klasu rođena, tj. iz nejednakosti unutar zemlje. Nije bilo bitno je li osoba rođena u Engleskoj, Kini ili Rusiji, bilo je bitno u kakvim je uvjetima unutar svoje zemlje rođena. No, kako nam grafikon prikazuje, navedeno se kroz vrijeme počelo mijenjati. Tako je u sredini 20. stoljeća došlo do stanja u kojemu je 80% svjetske nejednakosti proizlazilo iz, upravo obrnuto, nejednakosti između zemalja. Situacija je u svijetu bila te još uvijek je, da je od veće važnosti činjenica u kojoj je zemlji osoba rođena nego u kojim uvjetima unutar svoje, nebitno koje, zemlje (Milanovic, 2016.).

U radu Dragičevića "Suvremeni trendovi međunarodne ekonomske nejednakosti" autor ispituje čimbenike koji su pridonijeli pozitivnim trendovima međunarodne ekonomske nejednakosti posljednjih godina. Dragičević tvrdi da postoji nekoliko ključnih pokretača ovog trenda, opisanih u nastavku.

Globalizacija, sve veća integracija svjetskog gospodarstva kroz trgovinu, ulaganja i migracije dovela je do rasta gospodarstava u nastajanju i propadanja razvijenih gospodarstava, što je dovelo do smanjenja međunarodne ekonomske nejednakosti. Tehnološki napredak, uspon novih tehnologija i automatizacije doveo je do povećane produktivnosti i gospodarskog rasta, što je pomoglo smanjenju siromaštva i nejednakosti. Ljudski kapital, poboljšanja u obrazovanju i zdravstvu dovela su do povećanja ljudskog kapitala, što je dovelo do povećane produktivnosti i gospodarskog rasta. Institucije, uspostava stabilnih i uključivih institucija, kao što su vlasnička prava i vladavina prava, pomogla je promicanju gospodarskog rasta i smanjenju nejednakosti. Politike, progresivne politike kao što su redistributivno oporezivanje, mreže socijalne sigurnosti i politike za smanjenje siromaštva, pomogle su smanjiti nejednakost u mnogim zemljama (Dragičević, 2016). Sveukupno, Dragičević tvrdi da su ovi čimbenici pridonijeli pozitivnim trendovima međunarodne ekonomske nejednakosti koje smo vidjeli posljednjih godina, te da će se ti trendovi vjerojatno nastaviti i u budućnosti.

2.4. Učinci ekonomske nejednakosti u svijetu

Velika ekonomska nejednakost samo je jedna dimenzija globalne nejednakosti. Postoje mnogi drugi aspekti do kojih je ljudima stalo. No, budući da je visok prihod toliko važan za dobre životne uvjete, ove se druge nejednakosti preslikavaju na ekonomsku nejednakost. Oni koji žive s većim primanjima imaju višestruke prednosti. Promatrajući podatke kojima je baza razina prihoda, a zatim se na bazu povezuju podaci iz različitih životnih sfera, možemo zaključiti kako u područjima u kojima je viši prosječni prihod ljudi žive dulje, djeca umiru rjeđe, majke rjeđe umiru, liječnici se mogu usredotočiti na manje pacijenata, ljudi imaju bolji pristup čistoj pitkoj vodi i struji, mogu putovati više, imaju više slobodnog vremena, imaju bolji pristup obrazovanju i boljem učenju rezultate, a ljudi su zadovoljniji svojim životima. Nejednakost uvjeta života ljudi odražava svjetsku ekonomsku nejednakost (Roser, 2021).

Kao jedan od učinaka ekonomske nejednakosti u svijetu autor Milanovic ističe pojam takozvane premije za državljanstvo. Vrlo je važno istaknuti činjenicu kako je vrijednost Ginijevog koeficijenta malo ispod 70 u čak najviše nejednakim zemljama svijeta poput Južne Afrike ili Kolumbije. Kao što smo vidjeli u prethodnom poglavlju, svijet je nejednak na vrlo poseban način. Najveći dio nejednakosti, kada je podijelimo na nejednakost unutar zemalja i nejednakost među zemljama, nastaje zbog potonje. Kada su razlike u prihodima među zemljama velike, tada prihod osobe značajno ovisi o tome gdje živi, odnosno gdje je rođena, budući da 97 posto svjetskog stanovništva živi u zemljama u kojima je rođeno. Premija za državljanstvo koju netko dobiva rođenjem u bogatijoj zemlji u biti je renta (Milanovic, 2016.).

Globalnu distribuciju dohotka možemo zamisliti kao dugu motku, sličnu jarbolu zastave na kojoj su razine dohotka označene odozdo. Pa tako od egzistencijalnog minimuma (PPP \$200-\$300) do maksimalnog dohotka po kućanstvu od gotovo PPP \$200.000. Svaka država u svijetu sukladno svojoj razvijenosti te položaju na globalnoj ljestvici ima vlastiti raspon uzduž motke niz koji se njeni stanovnici mogu kretati. Pa tako na primjer Indija ima raspon od PPP \$250 do PPP \$7.000, Južna Koreja od PPP \$1.6000 do PPP \$80.000 te SAD od PPP \$2.500 do PPP \$180.000. Svaka osoba svojim rođenjem dobiva određenu poziciju uzduž motke unutar raspona vlastite zemlje te ju navedeno ne smješta samo na određeni položaj u distribuciji nacionalnog dohotka već isto tako i u distribuciji svjetskog

dohotka. Osoba svoj položaj može poboljšati na tri načina. Prvi je način, mogli bismo reći pasivni način, taj da se osoba može nadati kako će se cjelokupni raspon države u kojoj živi pomaknuti uzduž motke te samim time pogurati cijelo svoje stanovništvo prema naprijed i povisiti položaj svojih građana u distribuciji globalnog dohotka. Drugi bi način bio da svojim trudom i radom poboljša svoj položaj unutar raspona koji država unutar koje živi ima u distribuciji svjetskog dohotka, no u tom će slučaju kad-tad dosegnuti limit. Treća, zadnja mogućnost, bila bi da se osoba preseli iz siromašnije zemlje u bogatiju zemlju koja ima veći i ili viši raspon uzduž motke koja predstavlja distribuciju globalnog dohotka (Milanovic, 2015.).

Upravo shvaćajući te na vlastitoj koži osjećajući učinke premije za državljanstvo te globalne ekonomske nejednakosti, mnogi se stanovnici siromašnijih te manje razvijenih zemalja odlučuju na migracije. Velike migracije iz ekonomskih razloga možemo pripisati upravo visokoj razini nejednakosti u svijetu. Upravo tako, sama globalizacija u svojoj osnovi sadrži jednu veliku kontradikciju. Globalizacija, u svome širem značenju, podrazumijeva lagano, tj. neotežano kretanje faktora proizvodnje, dobara, tehnologije te samih ideja diljem svijeta. No, dok se po svijetu više-manje slobodno kreće kapital, usluge te roba, navedeno se ne događa s radom. Posljednjih desetljeća pod stalnim smo dojmom o velikim migracijama te migrantskim krizama, a dobar dio istih možemo pripisati upravo činjenici kako je dio migranata počeo bolje shvaćati razlike u prosječnim dohodcima između zemalja. Promatrajući svjetsku kartu, možemo uvidjeti kako je na područjima na kojima je velika razlika između dviju susjednih zemalja ili regija u prosječnom nacionalnom dohotku, postavljena i vrlo velika barijera migracijama (Milanovic, 2016.).

Postoje dokazi da velike promjene u mnogim društvima diljem svijeta, poput internacionalizacije, financijske otvorenosti, brzog razvoja informacijske tehnologije i migracija, stvaraju specifične obrasce nejednakosti u industrijaliziranim gospodarstvima i gospodarstvima u razvoju. Izazov za budućnost bit će suočavanje s posljedicama ovih povećanja i eventualnih smanjenja nejednakosti te nastojanje da se stvori ne samo globalno uravnoteženije društvo, već i suočavanje s velikim nejednakostima unutar naših zemalja (Mills, 2009.)

3. Otkrivanje znanja u bazama podataka

Podaci se u današnje vrijeme stvaraju u ogromnim količinama nevjerojatno visokom brzinom, a najbolje o tome govori činjenica da dnevno trenutnim tempom proizvodimo preko 2.5 kvintilijuna bajtova podataka (Marr, 2018). Sve bržim razvojem Interneta stvari (engl. Internet of Things – IoT) očekuje se sve veći rast proizvedene količine podataka, a s razlogom se postavlja pitanje kako navedenu količinu podataka možemo iskoristiti za korisne te vrlo precizne analize, kreiranje raznih modela te stvaranje relevantnih predikcija u različitim sferama života i znanosti.

Jedna od sfera života u kojima je korištenje velike količine dostupnih podataka u svrhu otkrivanja znanja od velike važnosti je poslovni svijet. Kvalitetne se poslovne odluke sve češće donose uz pomoć poslovnog znanja o samome poduzeću te njegovom okruženju, otkrivenog iz baza podataka. Kako bi se određeno poslovno znanje steklo, potrebno je analizirati odgovarajuće informacije dobivene iz podataka prikupljenih iz različitih izvora dok je u isto vrijeme podataka sve više, a vremena za njihovu obradu sve manje te upravo radi toga procesi otkrivanja znanja u bazama podataka moraju biti čim kraći i učinkovitiji (Ćurko i Španić, 2016).

Kroz ovo poglavlje detaljnije će se približiti teorijski dio procesa otkrivanja znanja u bazama podataka (engl. Knowledge discovery from data), kao i sami proces skladištenja podataka. Pojam velikih podataka te metode analize, poznato kao i „rudarenje“ navedenih podataka za izvlačenje konkretnih informacija te donošenje relevantnih zaključaka također će biti obrađeni u ovome poglavlju.

3.1. Baze podataka

Iako postojanje baze podataka nije preduvjet za rudarenje podataka, u praksi, proces otkrivanja znanja u bazama podataka, posebice za velika poduzeća, mnogo je lakši zahvaljujući postojanju baze podataka te prakse skladištenja podataka u istu. Primarni cilj baze podataka je povećati takozvanu inteligenciju procesa odlučivanja i znanja ljudi uključenih u ovaj proces (Kantardžić, 2011).

Skladište podataka, tj. bazu podataka možemo promatrati kao bazu koja za potrebe dobivanja poslovnog znanja unutar sebe sadrži posebno pripremljene podatke za pokretanje zahtjevnih analiza (Ćurko i Španić, 2016).

Također, može se reći da je skladište podataka statički prikaz podataka u skladištu, a skladištenje podataka je sustav koji pruža cjelokupni dinamički pregled, dizajniran za dobivanje prikladnih podataka za pohranu podataka, te njihovu obradu u potrebne informacije i znanja za donošenje odluka i upravljanje poslovanjem (Ćurko i Španić, 2016).

3.2. Otkrivanje znanja u bazama podataka

Nastavno na prethodni dio rada u kojem je naglašena velika količina podataka koja se svakodnevno kreira bilo kroz poslovne transakcije ili pak društvene interakcije putem raznih internetskih servisa i aplikacija, navedeni se podaci, uspješno pohranjeni, tj. skladišteni u bazama podataka mogu iskoristiti za ostvarenje napretka i poslovnih ciljeva poduzeća. Za navedeno se koristi otkrivanje znanja iz baza podataka (Pejić Bach i Kerep, 2011).

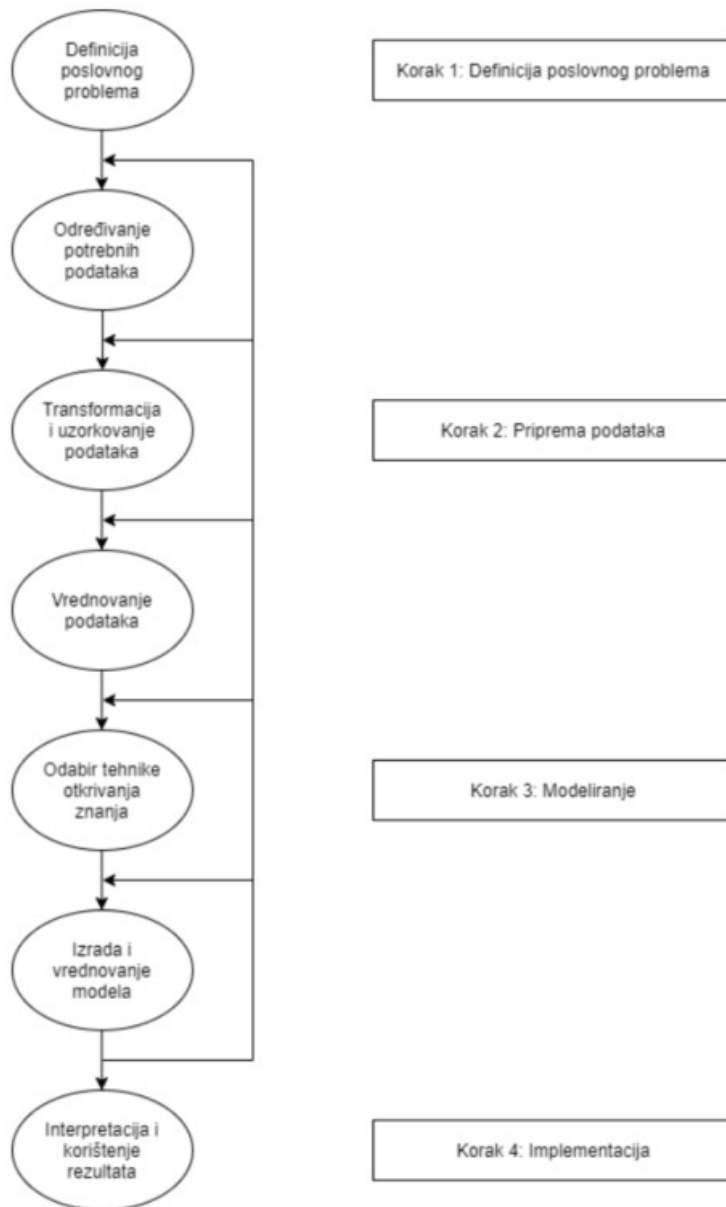
Već spomenuti neobuzdani rast količine podataka u bazama podataka posljednjih desetljeća, koje se sve češće koriste za svakodnevne aktivnosti, stavlja rudarenje podataka na čelo novih poslovnih tehnologija. Procjenjuje se da se svakih 20 mjeseci količina podataka pohranjenih u svjetskim bazama podataka udvostruči, a iako je teško kvantitativno opravdati ovu brojku, kvalitativno se svi stručnjaci informacijske tehnologije mogu složiti sa eksponencijalnom stopom rasta koju struka uviđa. Kako strojevi koji mogu vršiti rudarenje podataka postaju sve češći i procesorski jači uz konstantan priljev novih, svježih podataka, tako rastu i prilike za rudarenje podataka te ono postaje naša jedina nada u otkrivanju skrivenih obrazaca. Inteligentno analizirani podaci već dokazano kroz brojne poslovne primjere dovode do novih uvida i konkurentskih prednosti u poslovnom okruženju te samim time postaju visoko vrijedan resurs (Witten, Frank i Hall, 2012).

Otkrivanje znanja u bazama podataka možemo opisati kao proces analiziranja podataka iz različitih perspektiva te zatim sumiranje istih kako bismo dobili korisne informacije (Sharma, Bajpai i Litoriya, 2012.). Sama svrha otkrivanja znanja iz baza podataka je uspješna identifikacija informacija koje mogu poboljšati uspješnost poduzeća. Proces se vrlo često naziva i rudarenje podataka jer se u velikim količinama podataka zapravo traže i otkrivaju obrasci u podacima koji daljnjom analizom i interpretacijom daju informacije od velike vrijednosti (Pejić Bach i Kerep, 2011).

Kako bi se osigurala čim veća efikasnost i uspješnost pri procesu otkrivanja znanja iz baza podataka, važno je pridržavati se hodograma procesa otkrivanja znanja iz baza podataka koji je prikazan na slici broj 1 Proces otkrivanja znanja iz baza podataka (Pejić Bach i Kerep, 2011). Iz navedene je slike, tj. hodograma na prvu odmah uočljivo kako je sami proces iterativan, odnosno u slučaju određen pogreške ili neuspješno dovršenog koraka, uvijek je moguće vratiti se na prethodni korak. Tako su koraci u procesu prikazani na slici

sljedeći: definicija poslovnog problema, priprema podataka, modeliranje te implementacija (Pejić Bach i Kerep, 2011). Sami koraci, kao i pripadajući pod procesi bit će detaljnije opisani te objašnjeni u nastavku.

Slika 1 Proces otkrivanja znanja iz baza podataka



Izvor: Pejić Bach, M. i Kerep, I. (2011) Weka: alat za otkrivanje znanja iz baza podataka. Zagreb: Mikrorad (2011., str. 10)

3.2.1. Definicija poslovnog problema

Prvi korak u otkrivanju znanja iz baza podataka je svakako postupak definiranja poslovnog problema. Problem se izražava u obliku pitanja, a odgovor se može saznati tek nakon što je proces otkrivanja znanja iz baze podataka završen. Najbolji način za definiranje poslovnog problema je analiza područja u kojima je proces otkrivanja znanja već dokazano bio uspješan (Pejić Bach i Kerep, 2011).

3.2.2. Priprema podataka

Proces pripreme podataka drugi je u nizu koraka u ukupnom procesu otkrivanja znanja iz baza podataka, a sastoji se od određivanja potrebnih podataka, transformaciju i uzorkovanje podataka te na posljetku samo vrednovanje podataka (Pejić Bach i Kerep, 2011).

Baze podataka te skladištenje podataka unutar istih, koncepti spomenuti na početku trećeg poglavlja od velike su važnosti upravo u ovom koraku otkrivanja znanja iz baza podataka jer u slučaju kada iste postoje unutar poduzeća one postaju jedinstvena točka pristupa podacima organizacije koja nadilazi različite odjele te različite standarde, tj. metode i navike prikupljanja i pohranjivanja podataka od strane različitih odjela unutar organizacije (Witten, Frank i Hall, 2012).

Pri procesu otkrivanja znanja iz baze podataka odluku o tome koji će podaci biti korišteni za izradu modela unutar projektnog tima najčešće donose analitičar, stručnjak iz poduzeća te informatičar. Popis podataka koji će se koristiti u izradi modela konačan su rezultat ovog koraka, a zatim je potrebno podatke s popisa transformirati u oblik pogodan za otkrivanje znanja. Tako podaci moraju zadovoljiti određen kriterije po pitanju svojeg oblika. Tabelarni oblik je osnovni i glavni oblik u kojemu se podaci moraju nalaziti te se u recima moraju nalaziti opažanja, a u stupcima varijable koje najčešće definira stručnjak iz poduzeća na temelju dostupnih atributa iz baza podataka. U slučajevima kada se u procesu koriste podaci iz transakcijske baze podataka poduzeća, vrlo se često isti moraju sumirati kako bi bili korisni za otkrivanje znanja (Pejić Bach i Kerep, 2011).

Kako se u svakoj bazi podataka vrlo često nalaze takozvane netipične vrijednosti i problematični podaci, tako se i u ovome koraku procesa otkrivanja znanja iz baze podataka mora odlučiti što će biti napravljeno po tom pitanju. Zadržavanje u analizi, zamjena s nekim drugim vrijednostima ili pak potpuno izbacivanje iz analize su tri opcije koje su dostupne projektnom timu u vidu suočavanja s netipičnim vrijednostima u bazi podataka koja se koristi u procesu (Pejić Bach i Kerep, 2011).

3.2.3. Modeliranje

Nakon što su podaci odabrani te prema potrebi transformirani ili uzorkovani kako bi se proces otkrivanja znanja iz baze podatak čim kvalitetnije mogao dovršiti, ovisno o prirodi podataka te ranije definiranom poslovnom problemu odabire se jedna od tri struci poznatih skupina metoda modeliranja. Riječ je o metodama za klasifikaciju, za predviđanje vrijednosti ili pak metode za grupiranje (Pejić Bach i Kerep, 2011).

Za predviđanje događaja koriste se metode za klasifikaciju, kao što su na primjer stablo odlučivanja, logit regresija i neuralne mreže.

Za predviđanje numeričkih vrijednosti koriste se metode za predviđanje vrijednosti, kao što su na primjer linearna regresija, metode vremenskih serija te isto kao i kod metoda za klasifikaciju, neuronske mreže.

Za identifikaciju tipičnih skupina koriste se metode za grupiranje poput metoda analize tržišne košarice ili pak metoda klaster analize.

3.2.4. Implementacija

Uloga stručnjaka za poslovanje od velike je važnosti u ovom koraku implementacije rezultata unutar kojega se provodi interpretacija samih rezultata te njihova primjena u poslovanju. Menadžeri unutar poduzeća, koji posjeduju specifična poslovna znanja zahvaljujući kojima rezultate otkrivanja znanja mogu uspješno interpretirati, nažalost

najčešće nemaju dovoljno razvijene kompetencije u domeni samog procesa otkrivanja znanja iz baza podataka te je upravo radi toga od velike važnosti rezultate modela prikazati u obliku čim jednostavnijem za interpretaciju. Najčešće su ti oblici popis pravila ili pak grafikoni. Osim što uspješna implementacija uvelike ovisi o prezentaciji rezultata modela, od velike je pomoći ukoliko se rezultati implementiraju u informacijski sustav samog poduzeća (Pejić Bach i Kerep, 2011).

3.3. Metode otkrivanja znanja u bazama podataka

Prilikom otkrivanja znanja u bazama podataka korisniku na raspolaganju stoje različite metode koje je moguće koristiti u softverima za rudarenje podataka i strojno učenje, a glavna je podjela istih na metode za klasifikaciju, za predviđanje vrijednosti te metode za grupiranje (Pejić Bach i Kerep, 2011). Tako su na primjer neke od češće primjenjivih i poznatijih metoda: stablo odlučivanja, analiza netipičnih vrijednosti, analiza tržišne košarice, neuronske mreže, logit regresija, linearna regresija, metode vremenskih serija, klaster analiza, asocijativna pravila i mnoge druge metode. U nastavku rada fokus će biti stavljen na dvije metode koje će kasnije biti korištene u analizi podataka koji su tema ovog diplomskog rada, metoda prognoziranja vremenskih serija te klaster analiza.

3.3.1. Klaster analiza

Organiziranje podataka u skupine jedan je od temeljnih pristupa koji nam omogućava razumijevanje i učenje. Klaster analiza formalno je proučavanje metoda i algoritama za prirodno grupiranje, tj. klasteriranje objekata prema izmjerenim ili percipiranim intrinzičnim karakteristikama ili sličnostima (Kantardžić, 2011). Također, važno je napomenuti kako je klaster analiza (engl. Cluster Analysis) multivarijantna metoda kojom se postiže grupiranje uzoraka (Cornish, 2007). Uzorci za grupiranje predstavljeni su kao vektor mjerenja, ili formalnije, kao točka u višedimenzionalnom prostoru. Uzorci unutar

važeceg klastera sličniji su jedan drugome nego uzorcima koji pripadaju drugom klasteru. Možemo reći kako je metodologija klasteriranja posebno prikladna za istraživanje međuodnosa među uzorcima kako bi se napravila preliminarna procjena strukture uzorka (Kantardžić, 2011).

Najčešći primjer koji se za pojašnjenje ove metode otkrivanja znanja iz baza podataka koristi je primjer znanosti psihijatrije u kojemu se karakterizacija pacijenata prema skupinama određenih simptoma može koristiti za određivanje za pojedinu bolest odgovarajućih različitih oblika liječenja. Drugi, možda malo ekonomski gledano bliži struci primjer je primjer u znanosti marketinga u kojemu pak za ciljano oglašavanje vrlo često pomoću klaster analize stručnjaci identificiraju određene skupine potencijalnih kupaca (Cornish, 2007).

Pri samome izvođenju klaster analize postoji nekoliko metoda izvedbe, a svrstavamo ih u dvije glavne kategorije, u hijerarhijske i u ne-hijerarhijske metode (Cornish, 2007).

Hijerarhijske metode pak dijelimo na aglomerativne te na djeljive metode. Aglomerativne metode su metode u kojima subjekti počinju u svom zasebnom klasteru, a zatim se dva najbliža, odnosno najbližnja klastera kombiniraju i ponavljaju dok svi subjekti ne budu u jednom klasteru. Metoda završava na način da se među svim klasterima odabire optimalan broj klastera te navedeno postaje rješenje, tj. rezultat samog postupka. Djeljive metode su ustvari metode obrnute od aglomerativnih metoda, tj. postupak započinje na način da su na početku svi subjekti u istome klasteru te se prethodno opisana aglomerativna metoda ponavlja obrnutim redoslijedom sve dok svaki subjekt nije u vlastitom, zasebnom klasteru. Važno je napomenuti kako su aglomerativne metode češće korištene od djeljivih. U nastavku će ukratko biti opisane različite metode koje se koriste za određivanje koje klastera treba spojiti u svakoj fazi u sklopu hijerarhijske metode (Cornish, 2007).

U metodi najbližeg susjeda, tj. metodi jednostruke veze udaljenost između dva klastera definirana je kao udaljenost između dva najbliža člana ili susjeda. Ova je metoda relativno jednostavna, ali vrlo često kritizirana jer ne uzima u obzir strukturu klastera i može rezultirati problemom ulančavanja pri čemu klasteri na kraju ispadnu dugi i neravni. Međutim, savjetuje se njena primjena ispred ostalih metoda u slučajevima kada prirodne nakupine nisu sferičnog ili eliptičnog oblika (Cornish, 2007).

Iduća je metoda najdaljeg susjeda, tj. metoda potpunog povezivanja. U ovom slučaju udaljenost između dva klastera definirana je kao najveća udaljenost između članova, tj.

udaljenost između dva subjekta koji su najudaljeniji. Ova metoda nastoji proizvesti kompaktne grozdove slične veličine, ali kao i prethodna metoda najbližeg susjeda, ne uzima u obzir strukturu klastera (Cornish, 2007).

Metoda povezivanja prosjeka, često nazivana UPGMA, udaljenost između dva klastera izračunava se kao prosječna udaljenost između svih parova subjekata u dva klastera. Smatra se vrlo robusnom metodom (Cornish, 2007).

U metodi centroida se izračunava, kako joj i sami naziv govori, centroid, tj. srednja vrijednost za svaku varijablu svakog klastera i koristi se udaljenost između težišta. Grozdovi čija su težišta međusobno najbliža podliježu spajanju. Ova metoda je također prilično robusna (Cornish, 2007).

U Wardovoj metodi svi mogući parovi klastera se kombiniraju te se izračunava kvadrirani zbroj udaljenosti unutar svakog klastera. Navedeno se zatim zbraja za sve klasterne, a izabire se kombinacija koja daje najmanji zbroj kvadrata. Ova metoda teži stvaranju grozdova približno jednake veličine, što nije uvijek poželjno. Unatoč svojoj osjetljivosti na izuzetke, tj. ekstremne vrijednosti, to je jedna od najpopularnijih metoda prosječnog povezivanja (Cornish, 2007).

Ne-hijerarhijske su metode ujedno poznate i kao algoritam k-srednjih vrijednosti, a jedna od važnijih karakteristika im je činjenica da se željeni broj klastera unaprijed određuje te se zatim traži najbolje rješenje. Postupak započinje odabirom početnih takozvanih klaster centara kojima se zatim pripisuju, tj. dodjeljuju svi subjekti. U idućem se koraku centroidi klastera formiraju te se ponovno izračunava udaljenost svakog pojedinog subjekta od centroida klastera kojemu pripada. Postupak se ponavlja dokle god centroidi ostaju stabilni. Ova se metoda najčešće koristi u slučajevima kada su u proces otkrivanja znanja iz baza podataka uključeni veliki setovi podataka te se vrlo često preferira korištenje ove metoda zbog mogućnosti da se subjekt premješta iz jednog u drugi klaster. Glavni nedostaci ove metode bili bi ti da je vrlo često teško u početku znati koliki broj klastera postoji u procesu te se stoga postupak mora više puta ponoviti te da je sama metoda jako osjetljiva te vrlo ovisna o inicijalnom izboru klaster centara (Cornish, 2007).

Važno napomena pri korištenju metoda klaster analize je ta da klaster analiza nema mehanizam za razlikovanje relevantnih od nebitnih varijabli te samim time izbor varijabli uključenih u klaster analizu mora biti potkrijepljen konceptualnim razmatranjima. Ovo je

vrlo važno jer formirani klasteri mogu biti vrlo ovisni o uključenim varijablama (Cornish, 2007).

3.3.2. Prognoziranje vremenskih serija

Prognoziranje vremenskih serija jedna je od metoda za predviđanje vrijednosti kojima se otkriva znanje iz baza podataka, a osnovicu metode čine podaci vremenske serije te uočavanje trendova i uzoraka kretanja vrijednosti u budućnosti na temelju dosadašnjih dostupnih podataka. Drugim riječima, prognoziranje vremenskih serija možemo definirati kao pronalaženje matematičke funkcije koja će približno generirati povijesne uzorke u vremenskoj seriji i koristiti navedene funkcije za izradu dugoročnih ili kratkoročnih predviđanja budućih vrijednosti. U podacima vremenskih serija, sekvencijski podaci sastoje se od dugih nizova numeričkih podataka, zabilježenih u jednakim vremenskim intervalima, na primjer po minuti, po satu ili po danu. Podaci vremenskih serija mogu biti generirani zahvaljujući mnogim prirodnim i ekonomskim procesima i pojavama kao što su burze vrijednosnica ili robe te znanstvenim, medicinskim ili prirodnim promatranjima (Han, Kamber i Pei, 2012).

Iako je regresijska analiza vremenskih serija podataka značajno proučavana u poljima statistike i analize signala, vrlo je često potrebno na podacima vremenskih serija napraviti trend analizu (engl. Trend analysis) kako bi se istom mogao napraviti integrirani model na temelju četiri glavna stupa koji karakteriziraju podatke vremenske serije. Trendovi ili dugoročna kretanja ukazuju na generalnom smjeru kretanja u kojem se podaci vremenske serije u obliku grafa kreću kroz vrijeme. Ciklična kretanja koja su ustvari dugoročne oscilacije oko linije ili krivulje trenda. Sezonske su varijacije gotovo identični obrasci koji se pojavljuju u vremenskoj seriji te slijede tijekom odgovarajućih sezona kao što su sezone blagdanske kupovine. Za učinkovitu analizu trendova podatke je često potrebno desezonizirati na temelju sezonskog indeksa izračunatog auto korelacijom. I posljednja, nasumična kretanja koje se karakteriziraju kao sporadične promjene uslijed slučajnih događaja poput radnih sporova ili najavljenih kadrovskih promjena unutar poduzeća (Han, Kamber i Pei, 2012).

3.4. Programi za otkrivanje znanja u bazama podataka

Kako je prethodno u radu spomenuto, na tržištu postoji veliki broj neplaćenih i plaćenih specijaliziranih softverskih rješenja bez kojih bi primjena metode modeliranja u procesu otkrivanja znanja podataka bile praktički nemoguća, a razlikuju se prema tome koje metode u sebi sadržavaju, tj. sadrže li samo jednu metodu te su za navedenu specijalizirani ili pak sadrže veći broj metoda (Pejić Bach i Kerep, 2011). Tako su među poznatijim programima Statistica Dana Miner, SAD Enterprise Miner te IBM Intelli Miner. Među programima besplatnima za korištenje, tj. alatima takozvanog otvorenog koda su najpoznatiji pak Rapid Miner, Alpha Miner, Knime i Weka.

Ogromna količina podataka koja se svakodnevno proizvodi jedan je od najvećih izazova s kojim se stručnjaci informacijske tehnologije susreću posljednjih desetljeća. Takozvani veliki podaci (engl. Big Data) koje definiramo kroz takozvana četiri V obilježja (volumen, raznolikost, promjenljivost te vjerodostojnost) (Ćurko i Španić, 2016) doveli su do razvoja tehnologije velikih podataka kojom se uspješno upravlja istima, ali isto tako i do razvoja strojnog učenja (engl. Machine Learning) koje struci donosi mogućnost filtriranja neobrađenih informacija te pretvorbu istih u korisno i primjenjivo znanje koje menadžeri unutar poduzeća različitih gospodarskih grana mogu koristiti za donošenje kvalitetnih i pravovremenih poslovnih odluka. Također, strojno je učenje zaslužno za stvaranje temeljne veze između današnjeg softverskog inženjerstva te napredne tehnologije umjetne inteligencije (Pejić Bach i Kerep, 2011). Upravo se metodama strojnog učenja vodila oformljena skupina istraživača na Sveučilištu Waikato na Novom Zelandu koja je devedesetih godina prošlog stoljeća razvila prvu verziju Weka softvera, o kojem će više biti napisano u nastavku.

3.4.1. Weka softver

Weka softver jedan je od poznatijih softvera otvorenog koda, tj. softvera koji je dostupan u potpunosti besplatno za korištenje u svrhu otkrivanja znanja iz baze podataka. Softver je u svome punom nazivu Waikato Environment for Knowledge Analysis nastao na Sveučilištu Waikato na Novom Zelandu devedesetih godina prošlog stoljeća, a u biti predstavlja paket Java klasa biblioteka koje korištenjem modernih algoritama ostvaruju rudarenje podataka i/ili strojno učenje (Pejić Bach i Kerep, 2011). Weka poboljšava obradu podataka obuhvaćajući alate za diskretizaciju, transformaciju, pa čak i obuku algoritama učenja na određenom skupu podataka. Weka alati mogu biti korišteni za prethodnu obradu skupa podataka, primjenu sheme učenja, te zatim i ispitivanje izvedbe klasifikatora kako bi se mogla procijeniti njegova učinkovitost. Weka čak olakšava analizu rezultata klasifikacije skupa podataka bez programiranja (Witten, Frank i Hall, 2012).

Prilikom korištenja Weka softvera korisnik od strane istoga dobiva potpunu podršku i uslugu kroz cjelokupni proces rudarenja podataka, uključujući početne procese pripreme ulaznih podataka, statističku analizu shema učenja, vizualizaciju ulaznih podataka, ali i samih krajnjih rezultata učenja. Ovaj softver otvorenog koda sadrži razne alate, pa tako i alate za regresiju, klasifikaciju, pripremu i grupiranje podataka te grafički prikaz asocijativnih pravila. Također, kroz softver je moguće razviti nove sheme strojnog učenja. Stručnjaci različitih područja koji rade na projektima otkrivanja znanja iz baza podataka, pomoću softvera Weka mogu vrlo lako identificirati znanje iz baza podataka koje bi inače bile prevelike da bi se na njima radila klasična, ručna analiza takozvanom metodom „korak po korak“ (Pejić Bach i Kerep, 2011).

Sami softver programirana je u Java programskom jeziku te sadrži GUI (enlg. Graphic user interface), tj. grafičko korisničko sučelje putem kojeg korisnik ima mogućnost na jednostavan i brz način upravljati datotekama podataka te kreirati grafička rješenja i vizualizacije prigodne za kasniju analizu (Sharma, Bajpai i Litoriya, 2012). Weka se može koristiti za učenje o skupovima podataka primjenom metode učenja i analizom rezultata, kao i za izradu modela koji mogu predvidjeti nove podatke. Samo sučelje softvera ima izbornik metoda učenja koje se mogu mijenjati te svaka metoda ima podesive parametre koji se mogu mijenjati u uređivaču svojstava ili objekta (Witten, Frank i Hall, 2012).

Kako je ranije spomenuto, Weka softver za otkrivanje znanja iz baza podataka ne spada pod skupinu alata komercijalnog tipa te za razliku od istih postoji te je dostupan u okruženju otvorenog koda (engl. open source) i distribuiran je pod uvjetima GNU Opće Politike Licenci (engl. GNU General Policy Licence). To u praksi znači da bilo koji korisnik koji putem interneta preuzme softver ima slobodu isti koristiti za kreiranje daljnjih softverskih rješenja, prenamjenu izvornog koda ili pak korištenje samo pojedinih dijelova koda softvera u svrhe koje su korisniku potrebne (Pejić Bach i Kerep, 2011). Također, Weka softver moguće je ugraditi u bilo koju drugu aplikaciju ili skladište podataka putem generalnog API-ja (engl. Application programming interface), tj. sučelja za programiranje aplikacija (Sharma, Bajpai i Litoriya, 2012). Upravo radi toga, kroz posljednjih se dvadesetak godina razvio veći broj projekata otvorenog koda koji svoj uspjeh mogu i moraju djelomično pripisati Weka softveru jer su vlastiti razvoj temeljili na njegovom kodu. Projekti koje vrijedi istaknuti su projekti poput Bioweka (otkrivanje znanja u biokemiji, biologiji te bioinformatici), MOA (engl. Massive Online Analysis), Rapidminer (platforma za prediktivnu analizu i poslovnu inteligenciju poznata kao jedna od vodeći platformi otvorenog koda za rudarenje podacima), GridWeka, Weka-Parallel, GATE te mnogi drugi (Pejić Bach i Kerep, 2011).

4. Analiza globalne ekonomske nejednakosti

4.1. Metodologija istraživanja

Metodologija istraživanja za analizu klasteriranja i analizu vremenskih serija zemalja svijeta na temelju bruto domaćeg proizvoda (BDP) po glavi stanovnika, Gini koeficijent i očekivanog životnog vijeka za godine 1995., 2005. i 2015., kao i predviđanje vremenskih serija 10 godina unaprijed od 2015., uključivat će korištenje softvera za rudarenje podataka Weka. Istraživanje ima za cilj razumjeti gospodarski i društveni razvoj zemalja, kretanje globalne ekonomske nejednakosti i međusobne odnose zemalja između različitih pokazatelja tijekom vremena.

Proces prikupljanja podataka uključivat će dobivanje podataka o BDP-u po glavi stanovnika, Gini koeficijentu i očekivanom životnom vijeku za svaku zemlju za godine 1995., 2005. i 2015. Podaci će biti prethodno obrađeni kako bi se osiguralo da su u prikladnom formatu za analizu i rješavanje bilo kakvih nedostajućih ili netočnih vrijednosti. Podaci će se zatim učitati u Weka softver za daljnju analizu.

Klaster analiza će se provesti pomoću algoritma k-srednjih vrijednosti. Broj klastera bit će postavljen na četiri, što odgovara dohodovnoj klasifikaciji Svjetske banke u zemlje s visokim dohotkom, višim srednjim dohotkom, nižim srednjim dohotkom i niskim dohotkom. Algoritam će zatim grupirati zemlje na temelju vrijednosti triju pokazatelja, BDP-a po glavi stanovnika, Gini koeficijenta i očekivanog životnog vijeka, za godine 1995., 2005. i 2015.

Analiza vremenske serije uključivat će izradu modela vremenske serije vrijednosti BDP per capita za nasumično odabrani set zemalja koji će se sastojati od po četiri predstavnika svakog pojedinačnog klastera (uzoraka) dobivenih klasterne analize za 2015. godinu. Modeli će se koristiti za izradu predviđanja za sljedećih 15 godina i za razumijevanje trendova i obrazaca u podacima.

Rezultati klaster analize omogućit će uvid u gospodarski i društveni razvoj zemalja i odnose između različitih pokazatelja. Analiza vremenskih serija pružit će informacije o

budućem gospodarskom i društvenom razvoju zemalja, na temelju trendova i obrazaca u podacima.

4.2. Korišteni podaci

Podaci korišteni u istraživanju preuzeti su sa službene stranice Svjetske banke, tj. iz podatkovnog centra Svjetske banke poznatog pod nazivom Banka podataka. Podaci su pohranjeni pod grupom podataka zvanim pokazatelji svjetskog razvoja (engl. World Development Indicators). Pokazatelji svjetskog razvoja (WDI) kompilacija je podataka o međunarodnom razvoju iz Svjetske banke i raznih drugih izvora. Uključuje više od 1000 pokazatelja o različitim aspektima razvoja kao što su gospodarski rast, siromaštvo, zdravlje, obrazovanje i drugi. Podaci su predstavljeni u formatu vremenske serije za različite zemlje i regije, a namijenjeni su informiranju kreatora politika i javnosti o statusu globalnih razvojnih napora.

U nastavku rada prikaze su tablice atributa korištenih prilikom analize globalne ekonomske nejednakosti otkrivanjem znanja u bazama podataka. Naime, u Tablici broj 1 prikazani su atributi korišteni prilikom analize podataka BDP-a per capita svjetskih zemalja, njihovi formati te isto tako modaliteti za nominalne attribute te maksimalna i minimalna vrijednost za numeričke attribute. U Tablici broj 2 prikazani su atributi korišteni prilikom analize podataka Ginijevog koeficijenta svjetskih zemalja, njihovi formati te isto tako modaliteti za nominalne attribute te maksimalna i minimalna vrijednost za numeričke attribute. U Tablici broj 3 prikazani su atributi korišteni prilikom analize podataka očekivane životne dobi svjetskih zemalja, njihovi formati te isto tako modaliteti za nominalne attribute te maksimalna i minimalna vrijednost za numeričke attribute.

Tablica 1 Tablica atributa BDP per capita

Naziv atributa	Format atributa	Modaliteti atributa (nominalnih)	Maksimalna vrijednost (numeričkih)	Minimalna vrijednost (numeričkih)
Naziv države	Nominalni	Afganistan, Albanija, Alžir...		
BDP per capita 1995.	Numerički		101866,604	134,343
BDP per capita 2005.	Numerički		124197,275	151,682
BDP per capita 2015.	Numerički		167313,266	293,455

Izvor: Rad autora; DataBank: World Development Indicators, The World Bank [izvor podataka]. Dostupno na: <https://databank.worldbank.org/source/world-development-indicators>

Tablica 2 Tablica atributa Ginijev koeficijent

Naziv atributa	Format atributa	Modaliteti atributa (nominalnih)	Maksimalna vrijednost (numeričkih)	Minimalna vrijednost (numeričkih)
Naziv države	Nominalni	Afganistan, Albanija, Alžir...		
Ginijev koeficijent 1995.	Numerički		59,60	23
Ginijev koeficijent 2005.	Numerički		59,50	25,20
Ginijev koeficijent 2015.	Numerički		51,90	25,50

Izvor: Rad autora; DataBank: World Development Indicators, The World Bank [izvor podataka]. Dostupno na: <https://databank.worldbank.org/source/world-development-indicators>

Tablica 3 Tablica atributa Očekivana životna dob

Naziv atributa	Format atributa	Modaliteti atributa (nominalnih)	Maksimalna vrijednost (numeričkih)	Minimalna vrijednost (numeričkih)
Naziv države	Nominalni	Afganistan, Albanija, Alžir...		
Očekivana životna dob 1995.	Numerički		79,536	31,037
Očekivana životna dob 2005.	Numerički		81,925	42,518
Očekivana životna dob 2015.	Numerički		7,867	50,881

Izvor: Rad autora; DataBank: World Development Indicators, The World Bank [izvor podataka]. Dostupno na: <https://databank.worldbank.org/source/world-development-indicators>

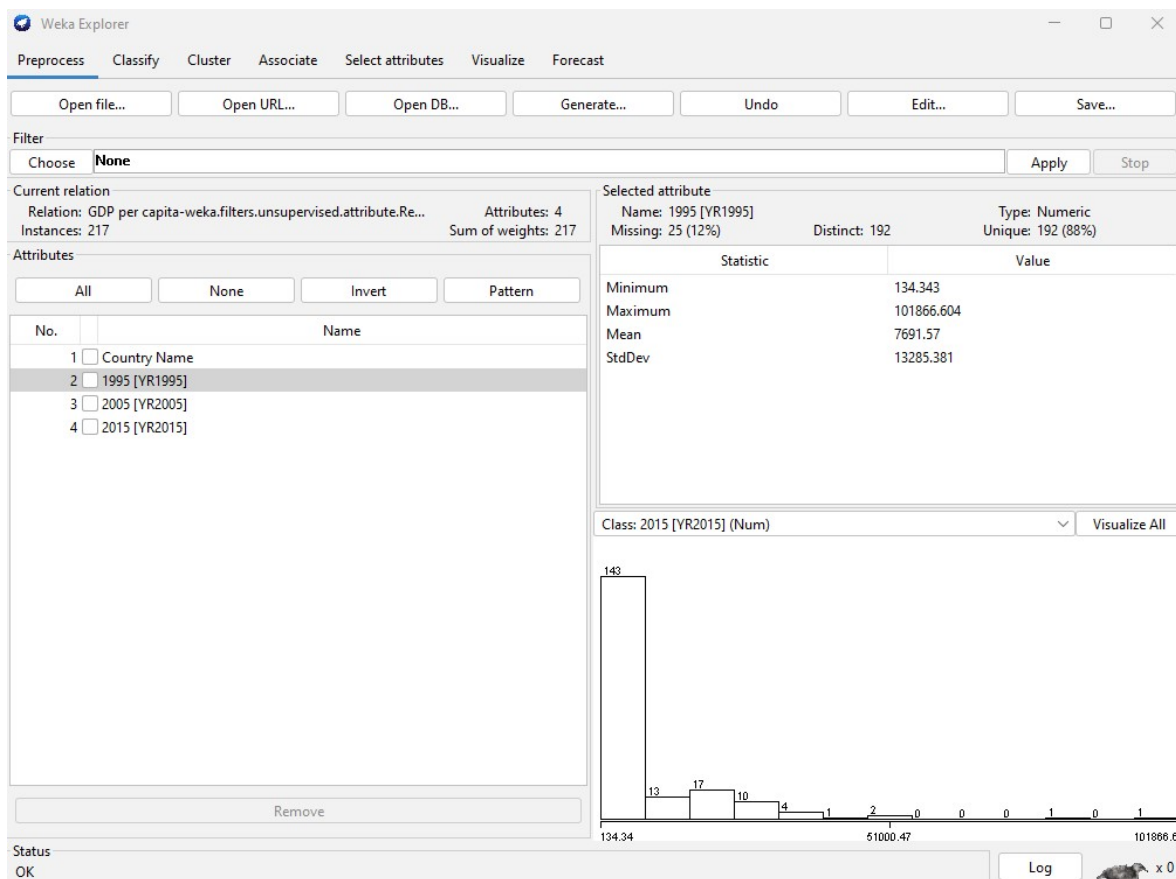
4.3. Rezultati istraživanja

4.3.1. Klaster analiza

Podatke preuzete sa službene stranice Svjetske banke, tj. baze podataka pod nazivom Banka podataka u formatu .csv prvotno pomoću Weka softvera pretvaramo u .arff format. ARFF (engl. Attribute-Relation File Format) datoteka je ASCII tekstualna datoteka koja opisuje popis instanci koje dijele određeni skup atributa. ARFF datoteke razvio je Projekt strojnog učenja na Odsjeku za računalne znanosti Sveučilišta Waikato za upotrebu sa softverom za strojno učenje Weka.

Jednom pretvorene podatke iz .csv formata u .arff format prikladan za korištenje u softveru Weka, navedeni softver pokrećemo. Ulaskom u „Explorer“ te učitavanjem datoteke u „Preprocess“ otvara nam se prozor prikazanim podacima BDP per capita prema godinama. Godine koje nisu odabrane za istraživanje jednostavnim odabirom brišemo te ostavljamo samo podatke za 1995., 2005. te 2015. godinu zajedno s nazivom zemalja, kao što je prikazano na Slici 2.

Slika 2 "Preprocess" u Weka softveru

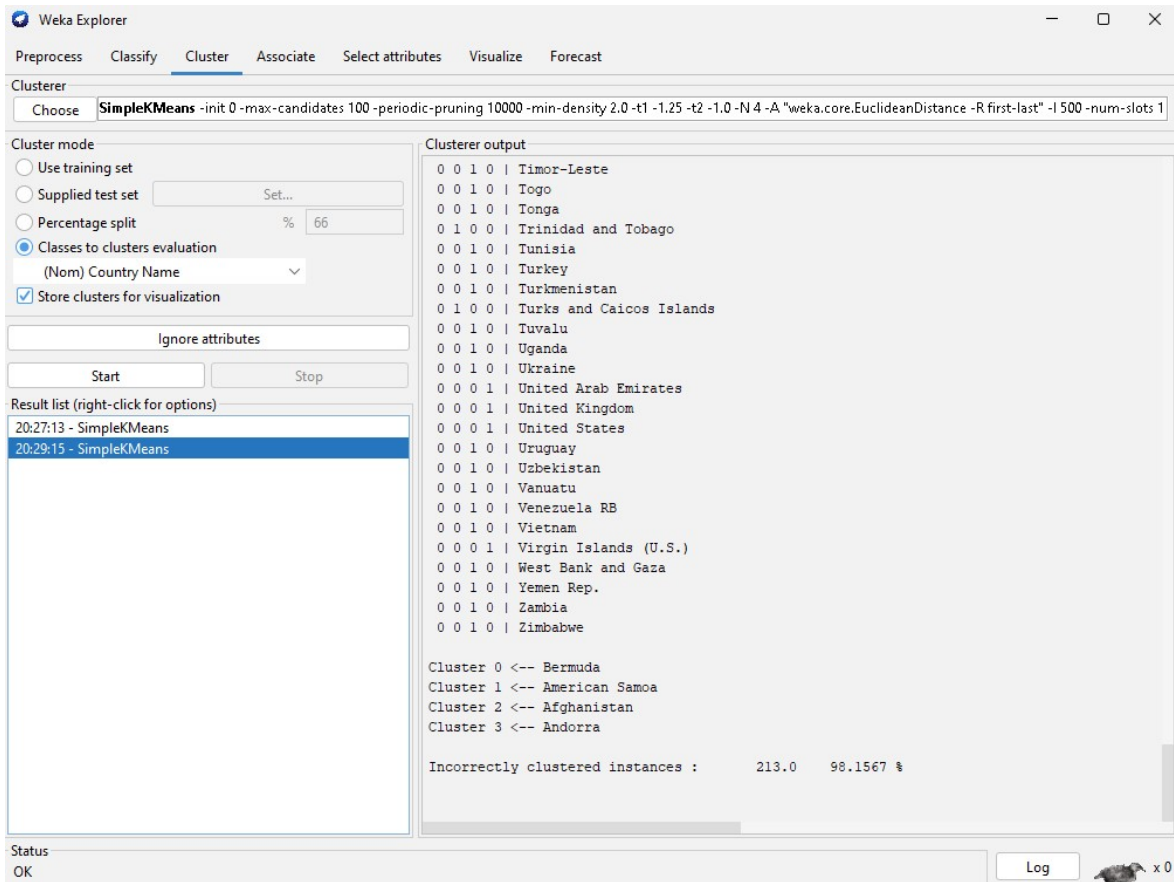


Izvor: Rad autora; DataBank: World Development Indicators, The World Bank [izvor podataka]. Dostupno na: <https://databank.worldbank.org/source/world-development-indicators>; Weka softver

Kao što možemo vidjeti na Slici broj 2, već su nam u „Preprocess“ dijelu Weka softveru za klikom određenu godinu vidljivi podaci poput minimalne i maksimalne vrijednosti, srednje vrijednosti, standardne devijacije, nedostajalih te jedinstvenih vrijednosti.

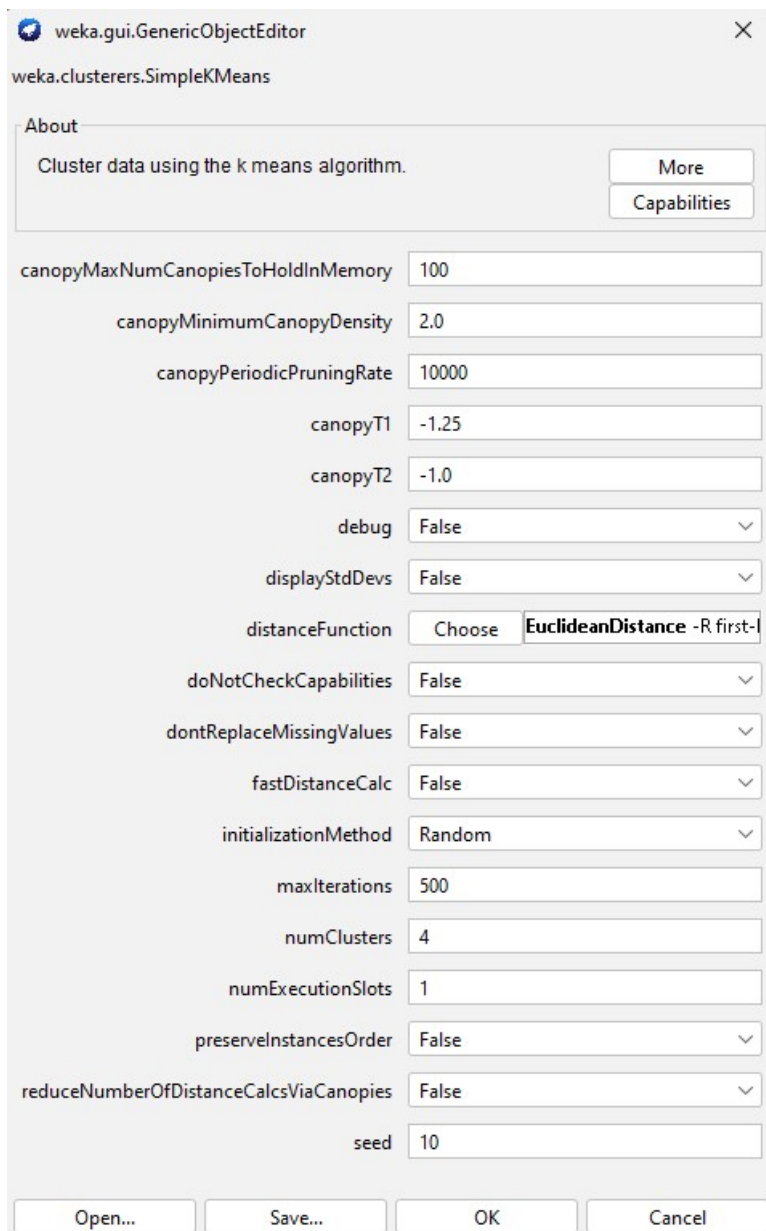
Nakon završenog „preprocess“ procesa u Weka softveru odabiremo prozor „Cluster“ u kojemu nakon odabira algoritma kojim želimo provoditi klaster analizu imamo mogućnost postaviti posebne postavke algoritma. U ovom istraživanju odabiremo algoritam SimpleKMeans te broj klastera postavljamo na četiri. Na Slikama broj 3 te 4 vidimo prvo prozor „Cluster“ s pripadajućim rezultatitima te zadane postavke algoritma SimpleKMeans kojim se provodi klaster analiza.

Slika 3 Prikaz prozora "Cluster"



Izvor: Rad autora; DataBank: World Development Indicators, The World Bank [izvor podataka]. Dostupno na: <https://databank.worldbank.org/source/world-development-indicators>; Weka softver

Slika 4 Postavke algoritma SimpleKMeans



Izvor: Rad autora; DataBank: World Development Indicators, The World Bank [izvor podataka]. Dostupno na: <https://databank.worldbank.org/source/world-development-indicators>; Weka softver

Na slici broj 5. vidljiv je rezultat klaster analize za 1995. godinu na temelju obrađenih podataka u Weka softveru. Na samoj slici vidljiv je broj atributa, vrijednost centroida klastera, broj instanci te broj država u samim klasterima.

Slika 5 Prikaz rezultata klaster analize BDP per capita za 1995. godinu

```

Clusterer output

=== Run information ===

Scheme:      weka.clusterers.SimpleKMeans -init 0 -max-candidates 100 -periodic-pruning 1
Relation:    GDP per capita-weka.filters.unsupervised.attribute.Remove-R5-19,21-24,26-29,
Instances:   217
Attributes:  2
              1995 [YR1995]

Ignored:     Country Name
Test mode:   Classes to clusters evaluation on training data

=== Clustering model (full training set) ===

kMeans
=====

Number of iterations: 15
Within cluster sum of squared errors: 0.3602894422286733

Initial starting points (random):

Cluster 0: 16439.356361
Cluster 1: 285.552548
Cluster 2: 242.620397
Cluster 3: 7690.576284

Missing values globally replaced with mean/mode

Final cluster centroids:

Attribute      Full Data      Cluster#
              (217.0)      0          1          2          3
=====
1995 [YR1995]  7691.57 70421.6567  9489.1802  1565.6015  25926.0464

Time taken to build model (full training data) : 0 seconds

=== Model and evaluation on training set ===

Clustered Instances

0          4 ( 2%)
1          50 ( 23%)
2         136 ( 63%)
3          27 ( 12%)

```

Izvor: Rad autora; DataBank: World Development Indicators, The World Bank [izvor podataka]. Dostupno na: <https://databank.worldbank.org/source/world-development-indicators>; Weka softver

U nastavku rada izlistani su rezultati klaster analize za 1995. godinu prema državama, tj. iz navedenog se popisa može vrlo lako vidjeti pripadnost pojedine države određenom klasteru čije su vrijednosti centroida određene u prethodnom koraku prikazanom na Slici broj 5. Izlistani rezultati prikazani su tako redom na Slici broj 6, na Slici broj 7 te na Slici broj 8.

Slika 6 BDP per capita 1995. godine: pripadnost države klasteru (1)

Clusterer output		Clusterer output	
Class attribute: Country Name			
Classes to Clusters:			
0 1 2 3	<-- assigned to cluster	0 0 1 0	Congo Dem. Rep.
0 1 0 0	Afghanistan	0 0 1 0	Congo Rep.
0 0 1 0	Albania	0 0 1 0	Costa Rica
0 0 1 0	Algeria	0 0 1 0	Cote d ivoire
0 1 0 0	American Samoa	0 0 1 0	Croatia
0 0 0 1	Andorra	0 0 1 0	Cuba
0 0 1 0	Angola	0 1 0 0	Curacao
0 1 0 0	Antigua and Barbuda	0 1 0 0	Cyprus
0 1 0 0	Argentina	0 1 0 0	Czech Republic
0 0 1 0	Armenia	0 0 0 1	Denmark
0 1 0 0	Aruba	0 0 1 0	Djibouti
0 0 0 1	Australia	0 0 1 0	Dominica
0 0 0 1	Austria	0 0 1 0	Dominican Republic
0 0 1 0	Azerbaijan	0 0 1 0	Ecuador
0 1 0 0	Bahamas The	0 0 1 0	Egypt Arab Rep.
0 1 0 0	Bahrain	0 0 1 0	El Salvador
0 0 1 0	Bangladesh	0 0 1 0	Equatorial Guinea
0 1 0 0	Barbados	0 0 1 0	Eritrea
0 0 1 0	Belarus	0 0 1 0	Estonia
0 0 0 1	Belgium	0 0 1 0	Eswatini
0 0 1 0	Belize	0 0 1 0	Ethiopia
0 0 1 0	Benin	0 1 0 0	Faroe Islands
0 0 0 1	Bermuda	0 0 1 0	Fiji
0 0 1 0	Bhutan	0 0 0 1	Finland
0 0 1 0	Bolivia	0 0 0 1	France
0 0 1 0	Bosnia and Herzegovina	0 0 0 1	French Polynesia
0 0 1 0	Botswana	0 0 1 0	Gabon
0 0 1 0	Brazil	0 0 1 0	Gambia The
0 1 0 0	British Virgin Islands	0 0 1 0	Georgia
0 1 0 0	Brunei Darussalam	0 0 0 1	Germany
0 0 1 0	Bulgaria	0 0 1 0	Ghana
0 0 1 0	Burkina Faso	0 1 0 0	Gibraltar
0 0 1 0	Burundi	0 1 0 0	Greece
0 0 1 0	Cabo Verde	0 0 0 1	Greenland
0 0 1 0	Cambodia	0 0 1 0	Grenada
0 0 1 0	Cameroon	0 1 0 0	Guam
0 0 0 1	Canada	0 0 1 0	Guatemala
0 1 0 0	Cayman Islands	0 0 1 0	Guinea
0 0 1 0	Central African Republic	0 0 1 0	Guinea-Bissau
0 0 1 0	Chad	0 0 1 0	Guyana
0 1 0 0	Channel Islands	0 0 1 0	Haiti
0 0 1 0	Chile	0 0 1 0	Honduras
0 0 1 0	China	0 0 0 1	Hong Kong SAR China
0 0 1 0	Colombia	0 0 1 0	Hungary
0 0 1 0	Comoros	0 0 0 1	Iceland
0 0 1 0	Congo Dem. Rep.	0 0 1 0	India
0 0 1 0	Congo Rep.	0 0 1 0	Indonesia
		0 0 1 0	Iran Islamic Rep.
		0 0 1 0	Iraq
		0 0 0 1	Ireland

Izvor: Rad autora; DataBank: World Development Indicators, The World Bank [izvor podataka]. Dostupno na: <https://databank.worldbank.org/source/world-development-indicators>; Weka softver

Slika 7 BDP per capita 1995. godine: pripadnost države klasteru (2)

Clusterer output	
0 0 1 0	Iraq
0 0 0 1	Ireland
0 1 0 0	Isle of Man
0 0 0 1	Israel
0 0 0 1	Italy
0 0 1 0	Jamaica
0 0 0 1	Japan
0 0 1 0	Jordan
0 0 1 0	Kazakhstan
0 0 1 0	Kenya
0 0 1 0	Kiribati
0 1 0 0	Korea Dem. Peoples Rep.
0 1 0 0	Korea Rep.
0 1 0 0	Kosovo
0 1 0 0	Kuwait
0 0 1 0	Kyrgyz Republic
0 0 1 0	Lao PDR
0 0 1 0	Latvia
0 0 1 0	Lebanon
0 0 1 0	Lesotho
0 1 0 0	Liberia
0 0 1 0	Libya
1 0 0 0	Liechtenstein
0 0 1 0	Lithuania
1 0 0 0	Luxembourg
0 0 0 1	Macao SAR China
0 0 1 0	Madagascar
0 0 1 0	Malawi
0 0 1 0	Malaysia
0 0 1 0	Maldives
0 0 1 0	Mali
0 1 0 0	Malta
0 0 1 0	Marshall Islands
0 0 1 0	Mauritania
0 0 1 0	Mauritius
0 0 1 0	Mexico
0 0 1 0	Micronesia Fed. Sts.
0 0 1 0	Moldova
1 0 0 0	Monaco
0 0 1 0	Mongolia
0 1 0 0	Montenegro
0 0 1 0	Morocco
0 0 1 0	Mozambique
0 0 1 0	Myanmar
0 0 1 0	Namibia
0 1 0 0	Nauru
0 0 1 0	Nepal
0 0 0 1	Netherlands
0 0 0 1	New Caledonia
0 1 0 0	New Zealand

Clusterer output	
0 0 0 1	New Caledonia
0 1 0 0	New Zealand
0 0 1 0	Nicaragua
0 0 1 0	Niger
0 0 1 0	Nigeria
0 0 1 0	North Macedonia
0 1 0 0	Northern Mariana Islands
0 0 0 1	Norway
0 1 0 0	Oman
0 0 1 0	Pakistan
0 1 0 0	Palau
0 0 1 0	Panama
0 0 1 0	Papua New Guinea
0 0 1 0	Paraguay
0 0 1 0	Peru
0 0 1 0	Philippines
0 0 1 0	Poland
0 1 0 0	Portugal
0 1 0 0	Puerto Rico
0 1 0 0	Qatar
0 0 1 0	Romania
0 0 1 0	Russian Federation
0 0 1 0	Rwanda
0 0 1 0	Samoa
0 1 0 0	San Marino
0 1 0 0	Sao Tome and Principe
0 1 0 0	Saudi Arabia
0 0 1 0	Senegal
0 0 1 0	Serbia
0 1 0 0	Seychelles
0 0 1 0	Sierra Leone
0 0 0 1	Singapore
0 1 0 0	Sint Maarten (Dutch part)
0 0 1 0	Slovak Republic
0 1 0 0	Slovenia
0 0 1 0	Solomon Islands
0 1 0 0	Somalia
0 0 1 0	South Africa
0 1 0 0	South Sudan
0 1 0 0	Spain
0 0 1 0	Sri Lanka
0 1 0 0	St. Kitts and Nevis
0 0 1 0	St. Lucia
0 1 0 0	St. Martin (French part)
0 0 1 0	St. Vincent and the Grenadines
0 0 1 0	Sudan
0 0 1 0	Suriname
0 0 0 1	Sweden
1 0 0 0	Switzerland
0 0 1 0	Syrian Arab Republic

Izvor: Rad autora; DataBank: World Development Indicators, The World Bank [izvor podataka]. Dostupno na: <https://databank.worldbank.org/source/world-development-indicators>; Weka softver

Slika 8 BDP per capita 1995. godine: pripadnost države klasteru (3)

```
Clusterer output
0 1 0 0 | Somalia
0 0 1 0 | South Africa
0 1 0 0 | South Sudan
0 1 0 0 | Spain
0 0 1 0 | Sri Lanka
0 1 0 0 | St. Kitts and Nevis
0 0 1 0 | St. Lucia
0 1 0 0 | St. Martin (French part)
0 0 1 0 | St. Vincent and the Grenadines
0 0 1 0 | Sudan
0 0 1 0 | Suriname
0 0 0 1 | Sweden
1 0 0 0 | Switzerland
0 0 1 0 | Syrian Arab Republic
0 0 1 0 | Tajikistan
0 0 1 0 | Tanzania
0 0 1 0 | Thailand
0 1 0 0 | Timor-Leste
0 0 1 0 | Togo
0 0 1 0 | Tonga
0 0 1 0 | Trinidad and Tobago
0 0 1 0 | Tunisia
0 0 1 0 | Turkey
0 0 1 0 | Turkmenistan
0 1 0 0 | Turks and Caicos Islands
0 0 1 0 | Tuvalu
0 0 1 0 | Uganda
0 0 1 0 | Ukraine
0 0 0 1 | United Arab Emirates
0 0 0 1 | United Kingdom
0 0 0 1 | United States
0 1 0 0 | Uruguay
0 0 1 0 | Uzbekistan
0 0 1 0 | Vanuatu
0 0 1 0 | Venezuela RB
0 0 1 0 | Vietnam
0 1 0 0 | Virgin Islands (U.S.)
0 0 1 0 | West Bank and Gaza
0 0 1 0 | Yemen Rep.
0 0 1 0 | Zambia
0 0 1 0 | Zimbabwe

Cluster 0 <-- Liechtenstein
Cluster 1 <-- Afghanistan
Cluster 2 <-- Albania
Cluster 3 <-- Andorra
```

Izvor: Rad autora; DataBank: World Development Indicators, The World Bank [izvor podataka]. Dostupno na: <https://databank.worldbank.org/source/world-development-indicators>; Weka softver

Klaster analiza ponovljena je prema istim postavkama za 2005. godinu te za 2015. godinu. Na Slici broj 9, Slici broj 10, Slici broj 11 i Slici broj 12 možemo vidjeti rezultate za 2005. godinu te na Slici broj 13, 14, 15 i 16 rezultate klaster analize za 2015. godinu.

Slika 9 Prikaz rezultata klaster analize BDP per capita za 2005. godinu

```

Clusterer output
=== Run information ===

Scheme:      weka.clusterers.SimpleKMeans -init 0 -max-candidates 100 -periodic
Relation:    GDP per capita-weka.filters.unsupervised.attribute.Remove-R5-19,21
Instances:   217
Attributes:  2
              2005 [YR2005]

Ignored:
  Country Name
Test mode:   Classes to clusters evaluation on training data

=== Clustering model (full training set) ===

kMeans
=====

Number of iterations: 16
Within cluster sum of squared errors: 0.4872665897820068

Initial starting points (random):

Cluster 0: 23300.738363
Cluster 1: 10963.406378
Cluster 2: 321.723661
Cluster 3: 13791.454762

Missing values globally replaced with mean/mode

Final cluster centroids:
Attribute      Full Data      Cluster#
              (217.0)      0          1          2          3
=====
2005 [YR2005] 12807.2511 76187.5725 15057.9318 2471.8799 36486.8269

Time taken to build model (full training data) : 0 seconds

=== Model and evaluation on training set ===

Clustered Instances

0          9 ( 4%)
1         39 ( 18%)
2        137 ( 63%)
3         32 ( 15%)

```

Izvor: Rad autora; DataBank: World Development Indicators, The World Bank [izvor podataka]. Dostupno na: <https://databank.worldbank.org/source/world-development-indicators>; Weka softver

Slika 10 BDP per capita 2005. godine: pripadnost države klasteru (1)

Clusterer output		Clusterer output	
Class attribute: Country Name			
Classes to Clusters:			
0 1 2 3	<-- assigned to cluster	0 0 1 0	Congo Dem. Rep.
0 0 1 0	Afghanistan	0 0 1 0	Congo Rep.
0 0 1 0	Albania	0 0 1 0	Costa Rica
0 0 1 0	Algeria	0 0 1 0	Cote d ivoire
0 0 1 0	American Samoa	0 1 0 0	Croatia
0 0 0 1	Andorra	0 0 1 0	Cuba
0 0 1 0	Angola	0 1 0 0	Curacao
0 1 0 0	Antigua and Barbuda	0 1 0 0	Cyprus
0 0 1 0	Argentina	0 1 0 0	Czech Republic
0 0 1 0	Armenia	0 0 0 1	Denmark
0 1 0 0	Aruba	0 0 1 0	Djibouti
0 0 0 1	Australia	0 0 1 0	Dominica
0 0 0 1	Austria	0 0 1 0	Dominican Republic
0 0 1 0	Azerbaijan	0 0 1 0	Ecuador
0 0 0 1	Bahamas The	0 0 1 0	Egypt Arab Rep.
0 1 0 0	Bahrain	0 0 1 0	El Salvador
0 0 1 0	Bangladesh	0 1 0 0	Equatorial Guinea
0 1 0 0	Barbados	0 0 1 0	Eritrea
0 0 1 0	Belarus	0 1 0 0	Estonia
0 0 0 1	Belgium	0 0 1 0	Eswatini
0 0 1 0	Belize	0 0 1 0	Ethiopia
0 0 1 0	Benin	0 0 0 1	Faroe Islands
1 0 0 0	Bermuda	0 0 1 0	Fiji
0 0 1 0	Bhutan	0 0 0 1	Finland
0 0 1 0	Bolivia	0 0 0 1	France
0 0 1 0	Bosnia and Herzegovina	0 1 0 0	French Polynesia
0 0 1 0	Botswana	0 0 1 0	Gabon
0 0 1 0	Brazil	0 0 1 0	Gambia The
0 1 0 0	British Virgin Islands	0 0 1 0	Georgia
0 0 0 1	Brunei Darussalam	0 0 0 1	Germany
0 0 1 0	Bulgaria	0 0 1 0	Ghana
0 0 1 0	Burkina Faso	0 1 0 0	Gibraltar
0 0 1 0	Burundi	0 1 0 0	Greece
0 0 1 0	Cabo Verde	0 0 0 1	Greenland
0 0 1 0	Cambodia	0 0 1 0	Grenada
0 0 1 0	Cameroon	0 0 0 1	Guam
0 0 0 1	Canada	0 0 1 0	Guatemala
0 1 0 0	Cayman Islands	0 0 1 0	Guinea
0 0 1 0	Central African Republic	0 0 1 0	Guinea-Bissau
0 0 1 0	Chad	0 0 1 0	Guyana
1 0 0 0	Channel Islands	0 0 1 0	Haiti
0 0 1 0	Chile	0 0 1 0	Honduras
0 0 1 0	China	0 0 0 1	Hong Kong SAR China
0 0 1 0	Colombia	0 1 0 0	Hungary
0 0 1 0	Comoros	1 0 0 0	Iceland
0 0 1 0	Congo Dem. Rep.	0 0 1 0	India
0 0 1 0	Congo Rep.	0 0 1 0	Indonesia
		0 0 1 0	Iran Islamic Rep.
		0 0 1 0	Iraq
		0 0 0 1	Ireland

Izvor: Rad autora; DataBank: World Development Indicators, The World Bank [izvor podataka]. Dostupno na: <https://databank.worldbank.org/source/world-development-indicators>; Weka softver

Slika 11 BDP per capita 2005. godine: pripadnost države klasteru (2)

Clusterer output		Clusterer output	
0 0 1 0	Iraq	0 0 0 1	New Caledonia
0 0 0 1	Ireland	0 0 0 1	New Zealand
0 0 0 1	Isle of Man	0 0 1 0	Nicaragua
0 1 0 0	Israel	0 0 1 0	Niger
0 0 0 1	Italy	0 0 1 0	Nigeria
0 0 1 0	Jamaica	0 0 1 0	North Macedonia
0 0 0 1	Japan	0 1 0 0	Northern Mariana Islands
0 0 1 0	Jordan	1 0 0 0	Norway
0 0 1 0	Kazakhstan	0 1 0 0	Oman
0 0 1 0	Kenya	0 0 1 0	Pakistan
0 0 1 0	Kiribati	0 1 0 0	Palau
0 1 0 0	Korea Dem. Peoples Rep.	0 0 1 0	Panama
0 1 0 0	Korea Rep.	0 0 1 0	Papua New Guinea
0 1 0 0	Kosovo	0 0 1 0	Paraguay
0 0 0 1	Kuwait	0 0 1 0	Peru
0 0 1 0	Kyrgyz Republic	0 0 1 0	Philippines
0 0 1 0	Lao PDR	0 0 1 0	Poland
0 0 1 0	Latvia	0 1 0 0	Portugal
0 0 1 0	Lebanon	0 1 0 0	Puerto Rico
0 0 1 0	Lesotho	0 0 0 1	Qatar
0 0 1 0	Liberia	0 0 1 0	Romania
0 0 1 0	Libya	0 0 1 0	Russian Federation
1 0 0 0	Liechtenstein	0 0 1 0	Rwanda
0 0 1 0	Lithuania	0 0 1 0	Samoa
1 0 0 0	Luxembourg	1 0 0 0	San Marino
0 1 0 0	Macao SAR China	0 0 1 0	Sao Tome and Principe
0 0 1 0	Madagascar	0 1 0 0	Saudi Arabia
0 0 1 0	Malawi	0 0 1 0	Senegal
0 0 1 0	Malaysia	0 0 1 0	Serbia
0 0 1 0	Maldives	0 1 0 0	Seychelles
0 0 1 0	Mali	0 0 1 0	Sierra Leone
0 1 0 0	Malta	0 0 0 1	Singapore
0 0 1 0	Marshall Islands	0 1 0 0	Sint Maarten (Dutch part)
0 0 1 0	Mauritania	0 1 0 0	Slovak Republic
0 0 1 0	Mauritius	0 1 0 0	Slovenia
0 0 1 0	Mexico	0 0 1 0	Solomon Islands
0 0 1 0	Micronesia Fed. Sts.	0 1 0 0	Somalia
0 0 1 0	Moldova	0 0 1 0	South Africa
1 0 0 0	Monaco	0 1 0 0	South Sudan
0 0 1 0	Mongolia	0 0 0 1	Spain
0 0 1 0	Montenegro	0 0 1 0	Sri Lanka
0 0 1 0	Morocco	0 1 0 0	St. Kitts and Nevis
0 0 1 0	Mozambique	0 0 1 0	St. Lucia
0 0 0 1	Myanmar	0 1 0 0	St. Martin (French part)
0 0 1 0	Namibia	0 0 1 0	St. Vincent and the Grenadines
0 1 0 0	Nauru	0 0 1 0	Sudan
0 0 1 0	Nepal	0 0 1 0	Suriname
0 0 0 1	Netherlands	0 0 0 1	Sweden
0 0 0 1	New Caledonia	1 0 0 0	Switzerland
0 0 0 1	New Zealand	0 0 1 0	Syrian Arab Republic

Izvor: Rad autora; DataBank: World Development Indicators, The World Bank [izvor podataka]. Dostupno na: <https://databank.worldbank.org/source/world-development-indicators>; Weka softver

Slika 12 BDP per capita 2005. godine: pripadnost države klasteru (3)

```
Cluster output
0 1 0 0 | Somalia
0 0 1 0 | South Africa
0 1 0 0 | South Sudan
0 0 0 1 | Spain
0 0 1 0 | Sri Lanka
0 1 0 0 | St. Kitts and Nevis
0 0 1 0 | St. Lucia
0 1 0 0 | St. Martin (French part)
0 0 1 0 | St. Vincent and the Grenadines
0 0 1 0 | Sudan
0 0 1 0 | Suriname
0 0 0 1 | Sweden
1 0 0 0 | Switzerland
0 0 1 0 | Syrian Arab Republic
0 0 1 0 | Tajikistan
0 0 1 0 | Tanzania
0 0 1 0 | Thailand
0 0 1 0 | Timor-Leste
0 0 1 0 | Togo
0 0 1 0 | Tonga
0 1 0 0 | Trinidad and Tobago
0 0 1 0 | Tunisia
0 0 1 0 | Turkey
0 0 1 0 | Turkmenistan
0 1 0 0 | Turks and Caicos Islands
0 0 1 0 | Tuvalu
0 0 1 0 | Uganda
0 0 1 0 | Ukraine
0 0 0 1 | United Arab Emirates
0 0 0 1 | United Kingdom
0 0 0 1 | United States
0 0 1 0 | Uruguay
0 0 1 0 | Uzbekistan
0 0 1 0 | Vanuatu
0 0 1 0 | Venezuela RB
0 0 1 0 | Vietnam
0 0 0 1 | Virgin Islands (U.S.)
0 0 1 0 | West Bank and Gaza
0 0 1 0 | Yemen Rep.
0 0 1 0 | Zambia
0 0 1 0 | Zimbabwe

Cluster 0 <-- Bermuda
Cluster 1 <-- Antigua and Barbuda
Cluster 2 <-- Afghanistan
Cluster 3 <-- Andorra
```

Izvor: Rad autora; DataBank: World Development Indicators, The World Bank [izvor podataka]. Dostupno na: <https://databank.worldbank.org/source/world-development-indicators>; Weka softver

Slika 13 Prikaz rezultata klaster analize BDP per capita za 2015. godinu

```

Clusterer output
=== Run information ===

Scheme:      weka.clusterers.SimpleKMeans -init 0 -max-candidates 100 -periodic-
Relation:    GDP per capita-weka.filters.unsupervised.attribute.Remove-R5-19,21-
Instances:   217
Attributes:  2
              2015 [YR2015]

Ignored:
Country Name
Test mode:   Classes to clusters evaluation on training data

=== Clustering model (full training set) ===

kMeans
=====

Number of iterations: 14
Within cluster sum of squared errors: 0.5702786483615832

Initial starting points (random):

Cluster 0: 28396.908423
Cluster 1: 11283.398054
Cluster 2: 484.153137
Cluster 3: 20627.926929

Missing values globally replaced with mean/mode

Final cluster centroids:

Attribute      Full Data      Cluster#
              (217.0)      0          1          2          3
=====
2015 [YR2015]  16635.7502  104009.9751  18110.492  3595.2802  43718.9823

Time taken to build model (full training data) : 0 seconds

=== Model and evaluation on training set ===

Clustered Instances

0          9 ( 4%)
1          47 ( 22%)
2         130 ( 60%)
3          31 ( 14%)

```

Izvor: Rad autora; DataBank: World Development Indicators, The World Bank [izvor podataka]. Dostupno na: <https://databank.worldbank.org/source/world-development-indicators>; Weka softver

Slika 14 BDP per capita 2015. godine: pripadnost države klasteru (1)

Clusterer output		Clusterer output	
Class attribute: Country Name			
Classes to Clusters:			
0 1 2 3	<-- assigned to cluster	0 0 1 0	Congo Dem. Rep.
0 0 1 0	Afghanistan	0 0 1 0	Congo Rep.
0 0 1 0	Albania	0 1 0 0	Costa Rica
0 0 1 0	Algeria	0 0 1 0	Cote d ivoire
0 1 0 0	American Samoa	0 1 0 0	Croatia
0 0 0 1	Andorra	0 0 1 0	Cuba
0 0 1 0	Angola	0 1 0 0	Curacao
0 1 0 0	Antigua and Barbuda	0 1 0 0	Cyprus
0 1 0 0	Argentina	0 1 0 0	Czech Republic
0 0 1 0	Armenia	0 0 0 1	Denmark
0 1 0 0	Aruba	0 0 1 0	Djibouti
0 0 0 1	Australia	0 0 1 0	Dominica
0 0 0 1	Austria	0 0 1 0	Dominican Republic
0 0 1 0	Azerbaijan	0 0 1 0	Ecuador
0 0 0 1	Bahamas The	0 0 1 0	Egypt Arab Rep.
0 1 0 0	Bahrain	0 0 1 0	El Salvador
0 0 1 0	Bangladesh	0 1 0 0	Equatorial Guinea
0 1 0 0	Barbados	0 1 0 0	Eritrea
0 0 1 0	Belarus	0 1 0 0	Estonia
0 0 0 1	Belgium	0 0 1 0	Eswatini
0 0 1 0	Belize	0 0 1 0	Ethiopia
0 0 1 0	Benin	0 0 0 1	Faroe Islands
1 0 0 0	Bermuda	0 0 1 0	Fiji
0 0 1 0	Bhutan	0 0 0 1	Finland
0 0 1 0	Bolivia	0 0 0 1	France
0 0 1 0	Bosnia and Herzegovina	0 1 0 0	French Polynesia
0 0 1 0	Botswana	0 0 1 0	Gabon
0 0 1 0	Brazil	0 0 1 0	Gambia The
0 1 0 0	British Virgin Islands	0 0 1 0	Georgia
0 0 0 1	Brunei Darussalam	0 0 0 1	Germany
0 0 1 0	Bulgaria	0 0 1 0	Ghana
0 0 1 0	Burkina Faso	0 1 0 0	Gibraltar
0 0 1 0	Burundi	0 1 0 0	Greece
0 0 1 0	Cabo Verde	0 0 0 1	Greenland
0 0 1 0	Cambodia	0 0 1 0	Grenada
0 0 1 0	Cameroon	0 0 0 1	Guam
0 0 0 1	Canada	0 0 1 0	Guatemala
1 0 0 0	Cayman Islands	0 0 1 0	Guinea
0 0 1 0	Central African Republic	0 0 1 0	Guinea-Bissau
0 0 1 0	Chad	0 0 1 0	Guyana
0 1 0 0	Channel Islands	0 0 1 0	Haiti
0 1 0 0	Chile	0 0 1 0	Honduras
0 0 1 0	China	0 0 0 1	Hong Kong SAR China
0 0 1 0	Colombia	0 1 0 0	Hungary
0 0 1 0	Comoros	0 0 0 1	Iceland
0 0 1 0	Congo Dem. Rep.	0 0 1 0	India
0 0 1 0	Congo Rep.	0 0 1 0	Indonesia
		0 0 1 0	Iran Islamic Rep.
		0 0 1 0	Iraq
		0 0 0 1	Ireland

Izvor: Rad autora; DataBank: World Development Indicators, The World Bank [izvor podataka]. Dostupno na: <https://databank.worldbank.org/source/world-development-indicators>; Weka softver

Slika 15 BDP per capita 2015. godine: pripadnost države klasteru (2)

Clusterer output		Clusterer output	
0 0 1 0	Iraq	0 0 0 1	New Caledonia
0 0 0 1	Ireland	0 0 0 1	New Zealand
1 0 0 0	Isle of Man	0 0 1 0	Nicaragua
0 0 0 1	Israel	0 0 1 0	Niger
0 1 0 0	Italy	0 0 1 0	Nigeria
0 0 1 0	Jamaica	0 0 1 0	North Macedonia
0 0 0 1	Japan	0 1 0 0	Northern Mariana Islands
0 0 1 0	Jordan	1 0 0 0	Norway
0 0 1 0	Kazakhstan	0 1 0 0	Oman
0 0 1 0	Kenya	0 0 1 0	Pakistan
0 0 1 0	Kiribati	0 1 0 0	Palau
0 1 0 0	Korea Dem. Peoples Rep.	0 1 0 0	Panama
0 1 0 0	Korea Rep.	0 0 1 0	Papua New Guinea
0 0 1 0	Kosovo	0 0 1 0	Paraguay
0 1 0 0	Kuwait	0 0 1 0	Peru
0 0 1 0	Kyrgyz Republic	0 0 1 0	Philippines
0 0 1 0	Lao PDR	0 1 0 0	Poland
0 1 0 0	Latvia	0 1 0 0	Portugal
0 0 1 0	Lebanon	0 1 0 0	Puerto Rico
0 0 1 0	Lesotho	0 0 0 1	Qatar
0 0 1 0	Liberia	0 0 1 0	Romania
0 0 1 0	Libya	0 0 1 0	Russian Federation
1 0 0 0	Liechtenstein	0 0 1 0	Rwanda
0 1 0 0	Lithuania	0 0 1 0	Samoa
1 0 0 0	Luxembourg	0 0 0 1	San Marino
1 0 0 0	Macao SAR China	0 0 1 0	Sao Tome and Principe
0 0 1 0	Madagascar	0 1 0 0	Saudi Arabia
0 0 1 0	Malawi	0 0 1 0	Senegal
0 0 1 0	Malaysia	0 0 1 0	Serbia
0 0 1 0	Maldives	0 1 0 0	Seychelles
0 0 1 0	Mali	0 0 1 0	Sierra Leone
0 1 0 0	Malta	0 0 0 1	Singapore
0 0 1 0	Marshall Islands	0 0 0 1	Sint Maarten (Dutch part)
0 0 1 0	Mauritania	0 1 0 0	Slovak Republic
0 0 1 0	Mauritius	0 1 0 0	Slovenia
0 0 1 0	Mexico	0 0 1 0	Solomon Islands
0 0 1 0	Micronesia Fed. Sts.	0 0 1 0	Somalia
0 0 1 0	Moldova	0 0 1 0	South Africa
1 0 0 0	Monaco	0 0 1 0	South Sudan
0 0 1 0	Mongolia	0 1 0 0	Spain
0 0 1 0	Montenegro	0 0 1 0	Sri Lanka
0 0 1 0	Morocco	0 1 0 0	St. Kitts and Nevis
0 0 1 0	Mozambique	0 0 1 0	St. Lucia
0 0 1 0	Myanmar	0 1 0 0	St. Martin (French part)
0 0 1 0	Namibia	0 0 1 0	St. Vincent and the Grenadines
0 0 1 0	Nauru	0 0 1 0	Sudan
0 0 1 0	Nepal	0 0 1 0	Suriname
0 0 0 1	Netherlands	0 0 0 1	Sweden
0 0 0 1	New Caledonia	1 0 0 0	Switzerland
0 0 0 1	New Zealand	0 0 1 0	Syrian Arab Republic

Izvor: Rad autora; DataBank: World Development Indicators, The World Bank [izvor podataka]. Dostupno na: <https://databank.worldbank.org/source/world-development-indicators>; Weka softver

Slika 16 BDP per capita 2015. godine: pripadnost države klasteru (3)

```
Cluster output
0 0 1 0 | Somalia
0 0 1 0 | South Africa
0 0 1 0 | South Sudan
0 1 0 0 | Spain
0 0 1 0 | Sri Lanka
0 1 0 0 | St. Kitts and Nevis
0 0 1 0 | St. Lucia
0 1 0 0 | St. Martin (French part)
0 0 1 0 | St. Vincent and the Grenadines
0 0 1 0 | Sudan
0 0 1 0 | Suriname
0 0 0 1 | Sweden
1 0 0 0 | Switzerland
0 0 1 0 | Syrian Arab Republic
0 0 1 0 | Tajikistan
0 0 1 0 | Tanzania
0 0 1 0 | Thailand
0 0 1 0 | Timor-Leste
0 0 1 0 | Togo
0 0 1 0 | Tonga
0 1 0 0 | Trinidad and Tobago
0 0 1 0 | Tunisia
0 1 0 0 | Turkey
0 0 1 0 | Turkmenistan
0 1 0 0 | Turks and Caicos Islands
0 0 1 0 | Tuvalu
0 0 1 0 | Uganda
0 0 1 0 | Ukraine
0 0 0 1 | United Arab Emirates
0 0 0 1 | United Kingdom
0 0 0 1 | United States
0 1 0 0 | Uruguay
0 0 1 0 | Uzbekistan
0 0 1 0 | Vanuatu
0 1 0 0 | Venezuela RB
0 0 1 0 | Vietnam
0 0 0 1 | Virgin Islands (U.S.)
0 0 1 0 | West Bank and Gaza
0 0 1 0 | Yemen Rep.
0 0 1 0 | Zambia
0 0 1 0 | Zimbabwe

Cluster 0 <-- Bermuda
Cluster 1 <-- American Samoa
Cluster 2 <-- Afghanistan
Cluster 3 <-- Andorra
```

Izvor: Rad autora; DataBank: World Development Indicators, The World Bank [izvor podataka]. Dostupno na: <https://databank.worldbank.org/source/world-development-indicators>; Weka softver

Prilikom klaster analize drugog po redu pokazatelja čiji su podaci za 1995., 2005. te 2015. godinu preuzeti sa službene stranice Svjetske banke, Ginijevog koeficijenta, zbog vrlo visokog stupnja oskudnosti podataka za države svijeta, potrebna je bila prethodna intervencija u .arff datoteci. Naime, zbog nepostojanja podataka o vrijednosti Ginijevog koeficijenta za velik broj svjetskih država za razdoblje prije 21. stoljeća, ali i u ranim godinama 21. stoljeća, radi relevantnosti klaster analize u Weka softveru analiza je napravljena samo nad zemljama koje imaju podatke za minimalno dvije od tri godine koje su predmetom analize. Stoga je klaster analiza za pokazatelj Ginijev koeficijent u ovome radu napravljena za ukupno trideset svjetskih zemalja koje su zadovoljile prethodno opisani kriterij. Navedene zemlje prikazane su na Slici broj 17 kroz mogućnost Weka softvera zvanu „arff viewer“. Na Slikama broj 18 i broj 19 prikazani su rezultati klaster analize za 1995. godinu, na Slikama broj 20 i broj 21 za 2005. godinu te na Slikama broj 22 i broj 23 za 2015. godinu.

Slika 17 Prikaz probраних zemalja za klaster analizu pokazatelja Ginijev koeficijent

The screenshot shows a window titled 'ARFF-Viewer' with a file named 'Gini index new.arff'. The data is presented in a table with the following structure:

No.	1: Country Name Nominal	2: 1995 [YR1995] Numeric	3: 2005 [YR2005] Numeric	4: 2015 [YR2015] Numeric
1	Argentina	48,9	47,7	
2	Austria	31,1	28,7	30,5
3	Azerbaijan	34,7	26,6	
4	Bangladesh	32,9	33,2	
5	Belgium	28,4	29,3	27,7
6	Brazil	59,6	56,3	51,9
7	Costa Rica	45,7	47,5	48,4
8	Cote d ivoire	40,6		41,5
9	Denmark	23,0	25,2	28,2
10	Egypt Arab Rep.	30,1		31,8
11	El Salvador	49,9	48,5	40,6
12	Ethiopia	44,6		35,0
13	Finland	23,5	27,6	27,1
14	Germany	28,9	32,1	31,7
15	Greece	37,0	34,6	36,0
16	Honduras	55,5	59,5	49,2
17	Ireland	37,0	33,8	31,8
18	Italy	35,2	33,8	35,4
19	Malaysia	48,5		41,1
20	Norway	26,0	30,6	27,5
21	Panama	57,8	53,8	50,8
22	Paraguay	58,2	51,4	47,6
23	Spain	36,7	32,4	36,2
24	Sweden	25,2	26,8	29,2
25	Tunisia	41,7	37,7	32,8
26	Ukraine	39,3	29,0	25,5
27	United Kingdom	36,3	34,3	33,2
28	United States	39,9	41,0	41,2
29	Uruguay	40,9	44,7	40,1
30	Venezuela RB	47,2	49,5	

Izvor: Rad autora; DataBank: World Development Indicators, The World Bank [izvor podataka]. Dostupno na: <https://databank.worldbank.org/source/world-development-indicators>; Weka softver

Slika 18 Prikaz rezultata klaster analize Ginijev koeficijent za 1995. godinu

```

Clusterer output
=== Run information ===

Scheme:      weka.clusterers.SimpleKMeans -init 0 -max-candidates 100 -periodic-
Relation:    Gini index new-weka.filters.unsupervised.attribute.Remove-R3-4
Instances:   30
Attributes:  2
              1995 [YR1995]

Ignored:     Country Name
Test mode:   Classes to clusters evaluation on training data

=== Clustering model (full training set) ===

kMeans
=====

Number of iterations: 7
Within cluster sum of squared errors: 0.2067931154366253

Initial starting points (random):

Cluster 0: 32.9
Cluster 1: 40.6
Cluster 2: 30.1
Cluster 3: 39.9

Missing values globally replaced with mean/mode

Final cluster centroids:
Attribute      Full Data      Cluster#
                (30.0)         0           1           2           3
=====
1995 [YR1995]  39.4767        35.1125     54.0571     26.4429     42.4875

Time taken to build model (full training data) : 0 seconds

=== Model and evaluation on training set ===

Clustered Instances

0      8 ( 27%)
1      7 ( 23%)
2      7 ( 23%)
3      8 ( 27%)

```

Izvor: Rad autora; DataBank: World Development Indicators, The World Bank [izvor podataka]. Dostupno na: <https://databank.worldbank.org/source/world-development-indicators>; Weka softver

Slika 19 Ginijev koeficijent 1995. godine: pripadnost države klasteru

```
Class attribute: Country Name
Classes to Clusters:

0 1 2 3 <-- assigned to cluster
0 1 0 0 | Argentina
1 0 0 0 | Austria
1 0 0 0 | Azerbaijan
1 0 0 0 | Bangladesh
0 0 1 0 | Belgium
0 1 0 0 | Brazil
0 0 0 1 | Costa Rica
0 0 0 1 | Cote d ivoire
0 0 1 0 | Denmark
0 0 1 0 | Egypt Arab Rep.
0 1 0 0 | El Salvador
0 0 0 1 | Ethiopia
0 0 1 0 | Finland
0 0 1 0 | Germany
1 0 0 0 | Greece
0 1 0 0 | Honduras
1 0 0 0 | Ireland
1 0 0 0 | Italy
0 1 0 0 | Malaysia
0 0 1 0 | Norway
0 1 0 0 | Panama
0 1 0 0 | Paraguay
1 0 0 0 | Spain
0 0 1 0 | Sweden
0 0 0 1 | Tunisia
0 0 0 1 | Ukraine
1 0 0 0 | United Kingdom
0 0 0 1 | United States
0 0 0 1 | Uruguay
0 0 0 1 | Venezuela RB

Cluster 0 <-- Austria
Cluster 1 <-- Argentina
Cluster 2 <-- Belgium
Cluster 3 <-- Costa Rica
```

Izvor: Rad autora; DataBank: World Development Indicators, The World Bank [izvor podataka]. Dostupno na: <https://databank.worldbank.org/source/world-development-indicators>; Weka softver

Slika 20 Prikaz rezultata klaster analize Ginijev koeficijent za 2005. godinu

```

Clusterer output
=== Run information ===

Scheme:      weka.clusterers.SimpleKMeans -init 0 -max-candidates 100 -periodic
Relation:    Gini index new-weka.filters.unsupervised.attribute.Remove-R2,4
Instances:   30
Attributes:  2
              2005 [YR2005]

Ignored:
              Country Name
Test mode:   Classes to clusters evaluation on training data

=== Clustering model (full training set) ===

kMeans
=====

Number of iterations: 4
Within cluster sum of squared errors: 0.17012515278282353

Initial starting points (random):

Cluster 0: 33.2
Cluster 1: 38.292308
Cluster 2: 41
Cluster 3: 29.3

Missing values globally replaced with mean/mode

Final cluster centroids:

Attribute      Full Data      Cluster#
                (30.0)         0           1           2           3
=====
2005 [YR2005]  38.2923      33.4571     39.5099     51.775     27.975

Time taken to build model (full training data) : 0 seconds

=== Model and evaluation on training set ===

Clustered Instances

0      7 ( 23%)
1      7 ( 23%)
2      8 ( 27%)
3      8 ( 27%)

```

Izvor: Rad autora; DataBank: World Development Indicators, The World Bank [izvor podataka]. Dostupno na: <https://databank.worldbank.org/source/world-development-indicators>; Weka softver

Slika 21 Ginijev koeficijent 2005. godine: pripadnost države klasteru

```
Class attribute: Country Name
Classes to Clusters:

0 1 2 3 <-- assigned to cluster
0 0 1 0 | Argentina
0 0 0 1 | Austria
0 0 0 1 | Azerbaijan
1 0 0 0 | Bangladesh
0 0 0 1 | Belgium
0 0 1 0 | Brazil
0 0 1 0 | Costa Rica
0 1 0 0 | Cote d ivoire
0 0 0 1 | Denmark
0 1 0 0 | Egypt Arab Rep.
0 0 1 0 | El Salvador
0 1 0 0 | Ethiopia
0 0 0 1 | Finland
1 0 0 0 | Germany
1 0 0 0 | Greece
0 0 1 0 | Honduras
1 0 0 0 | Ireland
1 0 0 0 | Italy
0 1 0 0 | Malaysia
0 0 0 1 | Norway
0 0 1 0 | Panama
0 0 1 0 | Paraguay
1 0 0 0 | Spain
0 0 0 1 | Sweden
0 1 0 0 | Tunisia
0 0 0 1 | Ukraine
1 0 0 0 | United Kingdom
0 1 0 0 | United States
0 1 0 0 | Uruguay
0 0 1 0 | Venezuela RB

Cluster 0 <-- Bangladesh
Cluster 1 <-- Cote d ivoire
Cluster 2 <-- Argentina
Cluster 3 <-- Austria
```

Izvor: Rad autora; DataBank: World Development Indicators, The World Bank [izvor podataka]. Dostupno na: <https://databank.worldbank.org/source/world-development-indicators>; Weka softver

Slika 22 Prikaz rezultata klaster analize Ginijev koeficijent za 2015. godinu

```

Clusterer output
=== Run information ===

Scheme:      weka.clusterers.SimpleKMeans -init 0 -max-candidates 100 -periodic-
Relation:    Gini index new-weka.filters.unsupervised.attribute.Remove-R2-3
Instances:   30
Attributes:  2
              2015 [YR2015]

Ignored:
Country Name

Test mode:   Classes to clusters evaluation on training data

=== Clustering model (full training set) ===

kMeans
=====

Number of iterations: 4
Within cluster sum of squared errors: 0.11148983107025077

Initial starting points (random):

Cluster 0: 36.615385
Cluster 1: 41.5
Cluster 2: 31.8
Cluster 3: 41.2

Missing values globally replaced with mean/mode

Final cluster centroids:

Attribute      Full Data      Cluster#
                (30.0)         0           1           2           3
=====
2015 [YR2015]  36.6154      35.5062     49.58      29.1       40.9

Time taken to build model (full training data) : 0 seconds

=== Model and evaluation on training set ===

Clustered Instances

0      10 ( 33%)
1       5 ( 17%)
2      10 ( 33%)
3       5 ( 17%)

```

Izvor: Rad autora; DataBank: World Development Indicators, The World Bank [izvor podataka]. Dostupno na: <https://databank.worldbank.org/source/world-development-indicators>; Weka softver

Slika 23 Ginijev koeficijent 2015. godine: pripadnost države klasteru

```
Class attribute: Country Name
Classes to Clusters:

0 1 2 3 <-- assigned to cluster
1 0 0 0 | Argentina
0 0 1 0 | Austria
1 0 0 0 | Azerbaijan
1 0 0 0 | Bangladesh
0 0 1 0 | Belgium
0 1 0 0 | Brazil
0 1 0 0 | Costa Rica
0 0 0 1 | Cote d ivoire
0 0 1 0 | Denmark
0 0 1 0 | Egypt Arab Rep.
0 0 0 1 | El Salvador
1 0 0 0 | Ethiopia
0 0 1 0 | Finland
0 0 1 0 | Germany
1 0 0 0 | Greece
0 1 0 0 | Honduras
0 0 1 0 | Ireland
1 0 0 0 | Italy
0 0 0 1 | Malaysia
0 0 1 0 | Norway
0 1 0 0 | Panama
0 1 0 0 | Paraguay
1 0 0 0 | Spain
0 0 1 0 | Sweden
1 0 0 0 | Tunisia
0 0 1 0 | Ukraine
1 0 0 0 | United Kingdom
0 0 0 1 | United States
0 0 0 1 | Uruguay
1 0 0 0 | Venezuela RB

Cluster 0 <-- Argentina
Cluster 1 <-- Brazil
Cluster 2 <-- Austria
Cluster 3 <-- Cote d ivoire
```

Izvor: Rad autora; DataBank: World Development Indicators, The World Bank [izvor podataka]. Dostupno na: <https://databank.worldbank.org/source/world-development-indicators>; Weka softver

Nakon uspješno provedene klaster analize pokazatelja BDP per capita te Ginijevog indeksa za godine 1995., 2005. te 2015., u nastavku slijedi analiza pokazatelja očekivane životne dobi za navedene godine. Proces klaster analize očekivane životne dobi provodimo na jednak način kao i s dosadašnjim pokazateljima te zbog dostatnih podataka za većinu svjetskih zemalja, za razliku od Ginijevog koeficijenta, nema potrebe za dodatnom intervencijom u uređivanje seta podataka prije provođenja klaster analize. Na Slikama broj 24, 25, 26 i 27 prikazani su rezultati klaster analize za 1995. godinu, na Slikama pod brojevima 28, 29, 30 i 31 za 2005. godinu te na Slikama pod brojevima 32, 33, 34 i 35 za 2015. godinu.

Slika 24 Prikaz rezultata klaster analize očekivanu životnu dob za 1995. godinu

```

Clusterer output
=== Run information ===

Scheme:      weka.clusterers.SimpleKMeans -init 0 -max-candidates 100 -perio
Relation:    Life expectancy -weka.filters.unsupervised.attribute.Remove-R5-
Instances:   217
Attributes:  2
              1995 [YR1995]

Ignored:
Country Name
Test mode:   Classes to clusters evaluation on training data

=== Clustering model (full training set) ===

kMeans
=====

Number of iterations: 7
Within cluster sum of squared errors: 0.6574656561754446

Initial starting points (random):

Cluster 0: 73.622
Cluster 1: 51.184
Cluster 2: 46.641
Cluster 3: 71.099

Missing values globally replaced with mean/mode

Final cluster centroids:
Attribute      Full Data      Cluster#
              (217.0)      0          1          2          3
=====
1995 [YR1995]  65.7668      74.9339    57.2992    46.8784    67.3423

Time taken to build model (full training data) : 0 seconds

=== Model and evaluation on training set ===

Clustered Instances

0      74 ( 34%)
1      37 ( 17%)
2      26 ( 12%)
3      80 ( 37%)

```

Izvor: Rad autora; DataBank: World Development Indicators, The World Bank [izvor podataka]. Dostupno na: <https://databank.worldbank.org/source/world-development-indicators>; Weka softver

Slika 25 Očekivana životna dob 1995. godine: pripadnost države klasteru (1)

Clusterer output		Clusterer output	
Class attribute: Country Name			
Classes to Clusters:			
0 1 2 3	<-- assigned to cluster	0 0 1 0	Congo Dem. Rep.
0 1 0 0	Afghanistan	0 1 0 0	Congo Rep.
1 0 0 0	Albania	1 0 0 0	Costa Rica
0 0 0 1	Algeria	0 0 1 0	Cote d ivoire
0 0 0 1	American Samoa	1 0 0 0	Croatia
0 0 0 1	Andorra	1 0 0 0	Cuba
0 0 1 0	Angola	0 0 0 1	Curacao
1 0 0 0	Antigua and Barbuda	1 0 0 0	Cyprus
1 0 0 0	Argentina	1 0 0 0	Czech Republic
0 0 0 1	Armenia	1 0 0 0	Denmark
1 0 0 0	Aruba	0 1 0 0	Djibouti
1 0 0 0	Australia	0 0 0 1	Dominica
1 0 0 0	Austria	0 0 0 1	Dominican Republic
0 0 0 1	Azerbaijan	0 0 0 1	Ecuador
1 0 0 0	Bahamas	0 0 0 1	Egypt Arab Rep.
1 0 0 0	Bahrain	0 0 0 1	El Salvador
0 1 0 0	Bangladesh	0 0 1 0	Equatorial Guinea
1 0 0 0	Barbados	0 1 0 0	Eritrea
0 0 0 1	Belarus	0 0 0 1	Estonia
1 0 0 0	Belgium	0 1 0 0	Eswatini
0 0 0 1	Belize	0 0 1 0	Ethiopia
0 1 0 0	Benin	1 0 0 0	Faroe Islands
0 0 0 1	Bermuda	0 0 0 1	Fiji
0 1 0 0	Bhutan	1 0 0 0	Finland
0 1 0 0	Bolivia	1 0 0 0	France
1 0 0 0	Bosnia and Herzegovina	0 0 0 1	French Polynesia
0 1 0 0	Botswana	0 1 0 0	Gabon
0 0 0 1	Brazil	0 1 0 0	Gambia The
0 0 0 1	British Virgin Islands	0 0 0 1	Georgia
1 0 0 0	Brunei Darussalam	1 0 0 0	Germany
0 0 0 1	Bulgaria	0 1 0 0	Ghana
0 0 1 0	Burkina Faso	0 0 0 1	Gibraltar
0 0 1 0	Burundi	1 0 0 0	Greece
0 0 0 1	Cabo Verde	0 0 0 1	Greenland
0 1 0 0	Cambodia	1 0 0 0	Grenada
0 0 1 0	Cameroon	1 0 0 0	Guam
1 0 0 0	Canada	0 0 0 1	Guatemala
0 0 0 1	Cayman Islands	0 0 1 0	Guinea
0 0 1 0	Central African Republic	0 0 1 0	Guinea-Bissau
0 0 1 0	Chad	0 0 0 1	Guyana
1 0 0 0	Channel Islands	0 1 0 0	Haiti
1 0 0 0	Chile	0 0 0 1	Honduras
0 0 0 1	China	1 0 0 0	Hong Kong SAR China
0 0 0 1	Colombia	0 0 0 1	Hungary
0 1 0 0	Comoros	1 0 0 0	Iceland
0 0 1 0	Congo Dem. Rep.	0 1 0 0	India
0 1 0 0	Congo Rep.	0 0 0 1	Indonesia
		0 0 0 1	Iran Islamic Rep.
		0 0 0 1	Iraq
		1 0 0 0	Ireland

Izvor: Rad autora; DataBank: World Development Indicators, The World Bank [izvor podataka]. Dostupno na: <https://databank.worldbank.org/source/world-development-indicators>; Weka softver

Slika 26 Očekivana životna dob 1995. godine: pripadnost države klasteru (2)

Clusterer output	Clusterer output
0 0 0 1 Iraq	1 0 0 0 New Caledonia
1 0 0 0 Ireland	1 0 0 0 New Zealand
0 0 0 1 Isle of Man	0 0 0 1 Nicaragua
1 0 0 0 Israel	0 0 1 0 Niger
1 0 0 0 Italy	0 0 1 0 Nigeria
1 0 0 0 Jamaica	1 0 0 0 North Macedonia
1 0 0 0 Japan	0 0 0 1 Northern Mariana Islands
0 0 0 1 Jordan	1 0 0 0 Norway
0 0 0 1 Kazakhstan	0 0 0 1 Oman
0 1 0 0 Kenya	0 1 0 0 Pakistan
0 1 0 0 Kiribati	1 0 0 0 Palau
0 0 0 1 Korea Dem. Peoples Rep.	1 0 0 0 Panama
1 0 0 0 Korea Rep.	0 1 0 0 Papua New Guinea
0 0 0 1 Kosovo	0 0 0 1 Paraguay
1 0 0 0 Kuwait	0 0 0 1 Peru
0 0 0 1 Kyrgyz Republic	0 0 0 1 Philippines
0 1 0 0 Lao PDR	1 0 0 0 Poland
0 0 0 1 Latvia	1 0 0 0 Portugal
1 0 0 0 Lebanon	1 0 0 0 Puerto Rico
0 1 0 0 Lesotho	1 0 0 0 Qatar
0 0 1 0 Liberia	0 0 0 1 Romania
0 0 0 1 Libya	0 0 0 1 Russian Federation
1 0 0 0 Liechtenstein	0 0 1 0 Rwanda
0 0 0 1 Lithuania	0 0 0 1 Samoa
1 0 0 0 Luxembourg	0 0 0 1 San Marino
1 0 0 0 Macao SAR China	0 1 0 0 Sao Tome and Principe
0 1 0 0 Madagascar	0 0 0 1 Saudi Arabia
0 0 1 0 Malawi	0 1 0 0 Senegal
1 0 0 0 Malaysia	0 0 0 1 Serbia
0 0 0 1 Maldives	1 0 0 0 Seychelles
0 0 1 0 Mali	0 0 1 0 Sierra Leone
1 0 0 0 Malta	1 0 0 0 Singapore
0 0 0 1 Marshall Islands	0 0 0 1 Sint Maarten (Dutch part)
0 1 0 0 Mauritania	1 0 0 0 Slovak Republic
0 0 0 1 Mauritius	1 0 0 0 Slovenia
1 0 0 0 Mexico	0 0 0 1 Solomon Islands
0 0 0 1 Micronesia Fed. Sts.	0 0 1 0 Somalia
0 0 0 1 Moldova	0 1 0 0 South Africa
0 0 0 1 Monaco	0 0 1 0 South Sudan
0 1 0 0 Mongolia	1 0 0 0 Spain
1 0 0 0 Montenegro	0 0 0 1 Sri Lanka
0 0 0 1 Morocco	0 0 0 1 St. Kitts and Nevis
0 0 1 0 Mozambique	1 0 0 0 St. Lucia
0 1 0 0 Myanmar	1 0 0 0 St. Martin (French part)
0 1 0 0 Namibia	0 0 0 1 St. Vincent and the Grenadines
0 0 0 1 Nauru	0 1 0 0 Sudan
0 1 0 0 Nepal	0 0 0 1 Suriname
1 0 0 0 Netherlands	1 0 0 0 Sweden
1 0 0 0 New Caledonia	1 0 0 0 Switzerland
1 0 0 0 New Zealand	1 0 0 0 Syrian Arab Republic

Izvor: Rad autora; DataBank: World Development Indicators, The World Bank [izvor podataka]. Dostupno na: <https://databank.worldbank.org/source/world-development-indicators>; Weka softver

Slika 27 Očekivana životna dob 1995. godine: pripadnost države klasteru (3)

```
Clusterer output
1 0 0 0 | Slovak Republic
1 0 0 0 | Slovenia
0 0 0 1 | Solomon Islands
0 0 1 0 | Somalia
0 1 0 0 | South Africa
0 0 1 0 | South Sudan
1 0 0 0 | Spain
0 0 0 1 | Sri Lanka
0 0 0 1 | St. Kitts and Nevis
1 0 0 0 | St. Lucia
1 0 0 0 | St. Martin (French part)
0 0 0 1 | St. Vincent and the Grenadines
0 1 0 0 | Sudan
0 0 0 1 | Suriname
1 0 0 0 | Sweden
1 0 0 0 | Switzerland
1 0 0 0 | Syrian Arab Republic
0 1 0 0 | Tajikistan
0 0 1 0 | Tanzania
0 0 0 1 | Thailand
0 1 0 0 | Timor-Leste
0 1 0 0 | Togo
0 0 0 1 | Tonga
0 0 0 1 | Trinidad and Tobago
1 0 0 0 | Tunisia
0 0 0 1 | Turkey
0 0 0 1 | Turkmenistan
0 0 0 1 | Turks and Caicos Islands
0 0 0 1 | Tuvalu
0 0 1 0 | Uganda
0 0 0 1 | Ukraine
1 0 0 0 | United Arab Emirates
1 0 0 0 | United Kingdom
1 0 0 0 | United States
1 0 0 0 | Uruguay
0 0 0 1 | Uzbekistan
0 0 0 1 | Vanuatu
1 0 0 0 | Venezuela RB
1 0 0 0 | Vietnam
1 0 0 0 | Virgin Islands (U.S.)
0 0 0 1 | West Bank and Gaza
0 1 0 0 | Yemen Rep.
0 0 1 0 | Zambia
0 0 1 0 | Zimbabwe

Cluster 0 <-- Albania
Cluster 1 <-- Afghanistan
Cluster 2 <-- Angola
Cluster 3 <-- Algeria
```

Izvor: Rad autora; DataBank: World Development Indicators, The World Bank [izvor podataka]. Dostupno na: <https://databank.worldbank.org/source/world-development-indicators>; Weka softver

Slika 28 Prikaz rezultata klaster analize očekivanu životnu dob za 2005. godinu

```

Clusterer output
=== Run information ===

Scheme:      weka.clusterers.SimpleKMeans -init 0 -max-candidates 100 -perio
Relation:    Life expectancy -weka.filters.unsupervised.attribute.Remove-R5-
Instances:   217
Attributes:  2
              2005 [YR2005]

Ignored:
  Country Name
Test mode:   Classes to clusters evaluation on training data

=== Clustering model (full training set) ===

kMeans
=====

Number of iterations: 17
Within cluster sum of squared errors: 0.9659248773344442

Initial starting points (random):

Cluster 0: 74.287
Cluster 1: 54.449
Cluster 2: 53.411
Cluster 3: 73.256

Missing values globally replaced with mean/mode

Final cluster centroids:
Attribute      Full Data      Cluster#
                (217.0)      0          1          2          3
=====
2005 [YR2005]  68.5053      77.5997    62.2005    51.1291    70.3409

Time taken to build model (full training data) : 0.01 seconds

=== Model and evaluation on training set ===

Clustered Instances

0      71 ( 33%)
1      32 ( 15%)
2      34 ( 16%)
3      80 ( 37%)

```

Izvor: Rad autora; DataBank: World Development Indicators, The World Bank [izvor podataka]. Dostupno na: <https://databank.worldbank.org/source/world-development-indicators>; Weka softver

Slika 29 Očekivana životna dob 2005. godine: pripadnost države klasteru (1)

Clusterer output					Clusterer output				
Class attribute: Country Name									
Classes to Clusters:									
0	1	2	3	<-- assigned to cluster	0	0	1	0	Congo Dem. Rep.
0	1	0	0	Afghanistan	0	0	1	0	Congo Rep.
1	0	0	0	Albania	1	0	0	0	Costa Rica
0	0	0	1	Algeria	0	0	1	0	Cote d ivoire
0	0	0	1	American Samoa	1	0	0	0	Croatia
0	0	0	1	Andorra	1	0	0	0	Cuba
0	0	1	0	Angola	0	0	0	1	Curacao
1	0	0	0	Antigua and Barbuda	1	0	0	0	Cyprus
1	0	0	0	Argentina	1	0	0	0	Czech Republic
0	0	0	1	Armenia	1	0	0	0	Denmark
1	0	0	0	Aruba	0	1	0	0	Djibouti
1	0	0	0	Australia	0	0	0	1	Dominica
1	0	0	0	Austria	0	0	0	1	Dominican Republic
0	0	0	1	Azerbaijan	1	0	0	0	Ecuador
0	0	0	1	Bahamas	0	0	0	1	Egypt Arab Rep.
1	0	0	0	Bahrain	0	0	0	1	El Salvador
0	0	0	1	Bangladesh	0	0	1	0	Equatorial Guinea
1	0	0	0	Barbados	0	1	0	0	Eritrea
0	0	0	1	Belarus	0	0	0	1	Estonia
1	0	0	0	Belgium	0	0	1	0	Eswatini
0	0	0	1	Belize	0	0	1	0	Ethiopia
0	1	0	0	Benin	1	0	0	0	Faroe Islands
1	0	0	0	Bermuda	0	0	0	1	Fiji
0	1	0	0	Bhutan	1	0	0	0	Finland
0	1	0	0	Bolivia	1	0	0	0	France
1	0	0	0	Bosnia and Herzegovina	1	0	0	0	French Polynesia
0	0	1	0	Botswana	0	1	0	0	Gabon
0	0	0	1	Brazil	0	1	0	0	Gambia The
0	0	0	1	British Virgin Islands	0	0	0	1	Georgia
1	0	0	0	Brunei Darussalam	1	0	0	0	Germany
0	0	0	1	Bulgaria	0	1	0	0	Ghana
0	0	1	0	Burkina Faso	0	0	0	1	Gibraltar
0	0	1	0	Burundi	1	0	0	0	Greece
0	0	0	1	Cabo Verde	0	0	0	1	Greenland
0	1	0	0	Cambodia	0	0	0	1	Grenada
0	0	1	0	Cameroon	1	0	0	0	Guam
1	0	0	0	Canada	0	0	0	1	Guatemala
0	0	0	1	Cayman Islands	0	0	1	0	Guinea
0	0	1	0	Central African Republic	0	0	1	0	Guinea-Bissau
0	0	1	0	Chad	0	0	0	1	Guyana
1	0	0	0	Channel Islands	0	1	0	0	Haiti
1	0	0	0	Chile	0	0	0	1	Honduras
0	0	0	1	China	1	0	0	0	Hong Kong SAR China
1	0	0	0	Colombia	0	0	0	1	Hungary
0	1	0	0	Comoros	1	0	0	0	Iceland
0	0	1	0	Congo Dem. Rep.	0	1	0	0	India
0	0	1	0	Congo Rep.	0	0	0	1	Indonesia
					0	0	0	1	Iran Islamic Rep.
					0	0	0	1	Iraq
					1	0	0	0	Ireland

Izvor: Rad autora; DataBank: World Development Indicators, The World Bank [izvor podataka]. Dostupno na: <https://databank.worldbank.org/source/world-development-indicators>; Weka softver

Slika 30 Očekivana životna dob 2005. godine: pripadnost države klasteru (2)

Clusterer output		Clusterer output	
0 0 0 1	Iraq	1 0 0 0	New Caledonia
1 0 0 0	Ireland	1 0 0 0	New Zealand
0 0 0 1	Isle of Man	0 0 0 1	Nicaragua
1 0 0 0	Israel	0 0 1 0	Niger
1 0 0 0	Italy	0 0 1 0	Nigeria
1 0 0 0	Jamaica	1 0 0 0	North Macedonia
1 0 0 0	Japan	0 0 0 1	Northern Mariana Islands
0 0 0 1	Jordan	1 0 0 0	Norway
0 1 0 0	Kazakhstan	1 0 0 0	Oman
0 0 1 0	Kenya	0 1 0 0	Pakistan
0 1 0 0	Kiribati	0 0 0 1	Palau
0 0 0 1	Korea Dem. Peoples Rep.	1 0 0 0	Panama
1 0 0 0	Korea Rep.	0 1 0 0	Papua New Guinea
0 0 0 1	Kosovo	0 0 0 1	Paraguay
0 0 0 1	Kuwait	0 0 0 1	Peru
0 0 0 1	Kyrgyz Republic	0 0 0 1	Philippines
0 1 0 0	Lao PDR	1 0 0 0	Poland
0 0 0 1	Latvia	1 0 0 0	Portugal
1 0 0 0	Lebanon	1 0 0 0	Puerto Rico
0 0 1 0	Lesotho	1 0 0 0	Qatar
0 0 1 0	Liberia	0 0 0 1	Romania
0 0 0 1	Libya	0 1 0 0	Russian Federation
1 0 0 0	Liechtenstein	0 0 1 0	Rwanda
0 0 0 1	Lithuania	0 0 0 1	Samoa
1 0 0 0	Luxembourg	0 0 0 1	San Marino
1 0 0 0	Macao SAR China	0 1 0 0	Sao Tome and Principe
0 1 0 0	Madagascar	0 0 0 1	Saudi Arabia
0 0 1 0	Malawi	0 1 0 0	Senegal
0 0 0 1	Malaysia	0 0 0 1	Serbia
0 0 0 1	Maldives	0 0 0 1	Seychelles
0 0 1 0	Mali	0 0 1 0	Sierra Leone
1 0 0 0	Malta	1 0 0 0	Singapore
0 0 0 1	Marshall Islands	0 0 0 1	Sint Maarten (Dutch part)
0 1 0 0	Mauritania	0 0 0 1	Slovak Republic
0 0 0 1	Mauritius	1 0 0 0	Slovenia
1 0 0 0	Mexico	0 0 0 1	Solomon Islands
0 1 0 0	Micronesia Fed. Sts.	0 0 1 0	Somalia
0 0 0 1	Moldova	0 0 1 0	South Africa
0 0 0 1	Monaco	0 0 1 0	South Sudan
0 1 0 0	Mongolia	1 0 0 0	Spain
0 0 0 1	Montenegro	1 0 0 0	Sri Lanka
0 0 0 1	Morocco	0 0 0 1	St. Kitts and Nevis
0 0 1 0	Mozambique	1 0 0 0	St. Lucia
0 1 0 0	Myanmar	1 0 0 0	St. Martin (French part)
0 0 1 0	Namibia	0 0 0 1	St. Vincent and the Grenadines
0 0 0 1	Nauru	0 1 0 0	Sudan
0 1 0 0	Nepal	0 0 0 1	Suriname
1 0 0 0	Netherlands	1 0 0 0	Sweden
1 0 0 0	New Caledonia	1 0 0 0	Switzerland
1 0 0 0	New Zealand	1 0 0 0	Syrian Arab Republic

Izvor: Rad autora; DataBank: World Development Indicators, The World Bank [izvor podataka]. Dostupno na: <https://databank.worldbank.org/source/world-development-indicators>; Weka softver

Slika 31 Očekivana životna dob 2005. godine: pripadnost države klasteru (3)

```
Clusterer output
0 0 1 0 | Somalia
0 0 1 0 | South Africa
0 0 1 0 | South Sudan
1 0 0 0 | Spain
1 0 0 0 | Sri Lanka
0 0 0 1 | St. Kitts and Nevis
1 0 0 0 | St. Lucia
1 0 0 0 | St. Martin (French part)
0 0 0 1 | St. Vincent and the Grenadines
0 1 0 0 | Sudan
0 0 0 1 | Suriname
1 0 0 0 | Sweden
1 0 0 0 | Switzerland
1 0 0 0 | Syrian Arab Republic
0 1 0 0 | Tajikistan
0 0 1 0 | Tanzania
0 0 0 1 | Thailand
0 1 0 0 | Timor-Leste
0 0 1 0 | Togo
0 0 0 1 | Tonga
0 0 0 1 | Trinidad and Tobago
1 0 0 0 | Tunisia
0 0 0 1 | Turkey
0 1 0 0 | Turkmenistan
0 0 0 1 | Turks and Caicos Islands
0 0 0 1 | Tuvalu
0 0 1 0 | Uganda
0 0 0 1 | Ukraine
1 0 0 0 | United Arab Emirates
1 0 0 0 | United Kingdom
1 0 0 0 | United States
1 0 0 0 | Uruguay
0 0 0 1 | Uzbekistan
0 0 0 1 | Vanuatu
0 0 0 1 | Venezuela RB
1 0 0 0 | Vietnam
1 0 0 0 | Virgin Islands (U.S.)
0 0 0 1 | West Bank and Gaza
0 1 0 0 | Yemen Rep.
0 0 1 0 | Zambia
0 0 1 0 | Zimbabwe

Cluster 0 <-- Albania
Cluster 1 <-- Afghanistan
Cluster 2 <-- Angola
Cluster 3 <-- Algeria
```

Izvor: Rad autora; DataBank: World Development Indicators, The World Bank [izvor podataka]. Dostupno na: <https://databank.worldbank.org/source/world-development-indicators>; Weka softver

Slika 32 Prikaz rezultata klaster analize očekivanu životnu dob za 2015. godinu

```

Clusterer output
=== Run information ===

Scheme:      weka.clusterers.SimpleKMeans -init 0 -max-candidates 100 -perio
Relation:    Life expectancy -weka.filters.unsupervised.attribute.Remove-R5-
Instances:   217
Attributes:  2
              2015 [YR2015]

Ignored:
              Country Name
Test mode:   Classes to clusters evaluation on training data

=== Clustering model (full training set) ===

kMeans
=====

Number of iterations: 14
Within cluster sum of squared errors: 0.9566018505733047

Initial starting points (random):

Cluster 0: 75.725
Cluster 1: 57.359
Cluster 2: 60.631
Cluster 3: 74.651

Missing values globally replaced with mean/mode

Final cluster centroids:
Attribute      Full Data      Cluster#
                (217.0)      0          1          2          3
=====
2015 [YR2015]  72.0429      80.2593    57.1594    64.9432    73.1201

Time taken to build model (full training data) : 0 seconds

=== Model and evaluation on training set ===

Clustered Instances

0      61 ( 28%)
1      24 ( 11%)
2      35 ( 16%)
3      97 ( 45%)
    
```

Izvor: Rad autora; DataBank: World Development Indicators, The World Bank [izvor podataka]. Dostupno na: <https://databank.worldbank.org/source/world-development-indicators>; Weka softver

Slika 33 Očekivana životna dob 2015. godine: pripadnost države klasteru (1)

Clusterer output		Clusterer output	
Class attribute: Country Name			
Classes to Clusters:			
0 1 2 3	<-- assigned to cluster	0 1 0 0	Congo Dem. Rep.
0 0 1 0	Afghanistan	0 0 1 0	Congo Rep.
1 0 0 0	Albania	1 0 0 0	Costa Rica
0 0 0 1	Algeria	0 1 0 0	Cote d ivoire
0 0 0 1	American Samoa	1 0 0 0	Croatia
0 0 0 1	Andorra	1 0 0 0	Cuba
0 1 0 0	Angola	1 0 0 0	Curacao
0 0 0 1	Antigua and Barbuda	1 0 0 0	Cyprus
0 0 0 1	Argentina	1 0 0 0	Czech Republic
0 0 0 1	Armenia	1 0 0 0	Denmark
0 0 0 1	Aruba	0 0 1 0	Djibouti
1 0 0 0	Australia	0 0 0 1	Dominica
1 0 0 0	Austria	0 0 0 1	Dominican Republic
0 0 0 1	Azerbaijan	0 0 0 1	Ecuador
0 0 0 1	Bahamas	0 0 0 1	Egypt Arab Rep.
1 0 0 0	Bahrain	0 0 0 1	El Salvador
0 0 0 1	Bangladesh	0 1 0 0	Equatorial Guinea
1 0 0 0	Barbados	0 0 1 0	Eritrea
0 0 0 1	Belarus	1 0 0 0	Estonia
1 0 0 0	Belgium	0 1 0 0	Eswatini
0 0 0 1	Belize	0 0 1 0	Ethiopia
0 1 0 0	Benin	1 0 0 0	Faroe Islands
1 0 0 0	Bermuda	0 0 1 0	Fiji
0 0 0 1	Bhutan	1 0 0 0	Finland
0 0 0 1	Bolivia	1 0 0 0	France
1 0 0 0	Bosnia and Herzegovina	1 0 0 0	French Polynesia
0 0 1 0	Botswana	0 0 1 0	Gabon
0 0 0 1	Brazil	0 1 0 0	Gambia The
0 0 0 1	British Virgin Islands	0 0 0 1	Georgia
0 0 0 1	Brunei Darussalam	1 0 0 0	Germany
0 0 0 1	Bulgaria	0 0 1 0	Ghana
0 1 0 0	Burkina Faso	0 0 0 1	Gibraltar
0 1 0 0	Burundi	1 0 0 0	Greece
0 0 0 1	Cabo Verde	0 0 0 1	Greenland
0 0 1 0	Cambodia	0 0 0 1	Grenada
0 1 0 0	Cameroon	1 0 0 0	Guam
1 0 0 0	Canada	0 0 0 1	Guatemala
0 0 0 1	Cayman Islands	0 1 0 0	Guinea
0 1 0 0	Central African Republic	0 1 0 0	Guinea-Bissau
0 1 0 0	Chad	0 0 0 1	Guyana
1 0 0 0	Channel Islands	0 0 1 0	Haiti
1 0 0 0	Chile	0 0 0 1	Honduras
0 0 0 1	China	1 0 0 0	Hong Kong SAR China
0 0 0 1	Colombia	0 0 0 1	Hungary
0 0 1 0	Comoros	1 0 0 0	Iceland
0 1 0 0	Congo Dem. Rep.	0 0 1 0	India
0 0 1 0	Congo Rep.	0 0 0 1	Indonesia
		0 0 0 1	Iran Islamic Rep.
		0 0 0 1	Iraq
		1 0 0 0	Ireland

Izvor: Rad autora; DataBank: World Development Indicators, The World Bank [izvor podataka]. Dostupno na: <https://databank.worldbank.org/source/world-development-indicators>; Weka softver

Slika 34 Očekivana životna dob 2015. godine: pripadnost države klasteru (2)

Clusterer output		Clusterer output	
0 0 0 1	Iraq	1 0 0 0	New Caledonia
1 0 0 0	Ireland	1 0 0 0	New Zealand
0 0 0 1	Isle of Man	0 0 0 1	Nicaragua
1 0 0 0	Israel	0 1 0 0	Niger
1 0 0 0	Italy	0 1 0 0	Nigeria
0 0 0 1	Jamaica	0 0 0 1	North Macedonia
1 0 0 0	Japan	0 0 0 1	Northern Mariana Islands
0 0 0 1	Jordan	1 0 0 0	Norway
0 0 0 1	Kazakhstan	1 0 0 0	Oman
0 0 1 0	Kenya	0 0 1 0	Pakistan
0 0 1 0	Kiribati	0 0 0 1	Palau
0 0 0 1	Korea Dem. Peoples Rep.	1 0 0 0	Panama
1 0 0 0	Korea Rep.	0 0 1 0	Papua New Guinea
0 0 0 1	Kosovo	0 0 0 1	Paraguay
0 0 0 1	Kuwait	0 0 0 1	Peru
0 0 0 1	Kyrgyz Republic	0 0 0 1	Philippines
0 0 1 0	Lao PDR	1 0 0 0	Poland
0 0 0 1	Latvia	1 0 0 0	Portugal
1 0 0 0	Lebanon	1 0 0 0	Puerto Rico
0 1 0 0	Lesotho	1 0 0 0	Qatar
0 0 1 0	Liberia	0 0 0 1	Romania
0 0 0 1	Libya	0 0 0 1	Russian Federation
1 0 0 0	Liechtenstein	0 0 1 0	Rwanda
0 0 0 1	Lithuania	0 0 0 1	Samoa
1 0 0 0	Luxembourg	0 0 0 1	San Marino
1 0 0 0	Macao SAR China	0 0 0 1	Sao Tome and Principe
0 0 1 0	Madagascar	0 0 0 1	Saudi Arabia
0 0 1 0	Malawi	0 0 1 0	Senegal
0 0 0 1	Malaysia	0 0 0 1	Serbia
1 0 0 0	Maldives	0 0 0 1	Seychelles
0 1 0 0	Mali	0 1 0 0	Sierra Leone
1 0 0 0	Malta	1 0 0 0	Singapore
0 0 0 1	Marshall Islands	1 0 0 0	Sint Maarten (Dutch part)
0 0 1 0	Mauritania	0 0 0 1	Slovak Republic
0 0 0 1	Mauritius	1 0 0 0	Slovenia
0 0 0 1	Mexico	0 0 0 1	Solomon Islands
0 0 1 0	Micronesia Fed. Sts.	0 1 0 0	Somalia
0 0 0 1	Moldova	0 0 1 0	South Africa
0 0 0 1	Monaco	0 1 0 0	South Sudan
0 0 0 1	Mongolia	1 0 0 0	Spain
0 0 0 1	Montenegro	0 0 0 1	Sri Lanka
0 0 0 1	Morocco	0 0 0 1	St. Kitts and Nevis
0 1 0 0	Mozambique	0 0 0 1	St. Lucia
0 0 1 0	Myanmar	1 0 0 0	St. Martin (French part)
0 0 1 0	Namibia	0 0 0 1	St. Vincent and the Grenadines
0 0 0 1	Nauru	0 0 1 0	Sudan
0 0 0 1	Nepal	0 0 0 1	Suriname
1 0 0 0	Netherlands	1 0 0 0	Sweden
1 0 0 0	New Caledonia	1 0 0 0	Switzerland
1 0 0 0	New Zealand	0 0 0 1	Syrian Arab Republic

Izvor: Rad autora; DataBank: World Development Indicators, The World Bank [izvor podataka]. Dostupno na: <https://databank.worldbank.org/source/world-development-indicators>; Weka softver

Slika 35 Očekivana životna dob 2015. godine: pripadnost države klasteru (3)

```
Clusterer output
0 0 0 1 | Slovak Republic
1 0 0 0 | Slovenia
0 0 0 1 | Solomon Islands
0 1 0 0 | Somalia
0 0 1 0 | South Africa
0 1 0 0 | South Sudan
1 0 0 0 | Spain
0 0 0 1 | Sri Lanka
0 0 0 1 | St. Kitts and Nevis
0 0 0 1 | St. Lucia
1 0 0 0 | St. Martin (French part)
0 0 0 1 | St. Vincent and the Grenadines
0 0 1 0 | Sudan
0 0 0 1 | Suriname
1 0 0 0 | Sweden
1 0 0 0 | Switzerland
0 0 0 1 | Syrian Arab Republic
0 0 0 1 | Tajikistan
0 0 1 0 | Tanzania
0 0 0 1 | Thailand
0 0 1 0 | Timor-Leste
0 1 0 0 | Togo
0 0 0 1 | Tonga
0 0 0 1 | Trinidad and Tobago
0 0 0 1 | Tunisia
0 0 0 1 | Turkey
0 0 1 0 | Turkmenistan
0 0 0 1 | Turks and Caicos Islands
0 0 0 1 | Tuvalu
0 0 1 0 | Uganda
0 0 0 1 | Ukraine
1 0 0 0 | United Arab Emirates
1 0 0 0 | United Kingdom
1 0 0 0 | United States
1 0 0 0 | Uruguay
0 0 0 1 | Uzbekistan
0 0 0 1 | Vanuatu
0 0 0 1 | Venezuela RB
0 0 0 1 | Vietnam
1 0 0 0 | Virgin Islands (U.S.)
0 0 0 1 | West Bank and Gaza
0 0 1 0 | Yemen Rep.
0 0 1 0 | Zambia
0 1 0 0 | Zimbabwe

Cluster 0 <-- Albania
Cluster 1 <-- Angola
Cluster 2 <-- Afghanistan
Cluster 3 <-- Algeria
```

Izvor: Rad autora; DataBank: World Development Indicators, The World Bank [izvor podataka]. Dostupno na: <https://databank.worldbank.org/source/world-development-indicators>; Weka softver

4.3.2. Analiza vremenske serije

Istraživanje predviđanja vremenskih serija u Weka softverskom programu može se provesti zahvaljujući proširenju programa zvanom „timeseriesForecasting“, odnosno proširenju za prognoziranje vremenskih serija. Proširenje je dostupno putem takozvanog „Package manager“ alata kojim se brzim i jednostavnim načinom instalira dodatno programsko proširenje. Podatke u .csv formatu potrebno je prije korištenja u svrhu predviđanja vremenskih serija transponirati. Također, uzorci svakog od klastera dobivenih klaster analizom podataka BDP per capita za 2015. godinu odabrani su nasumičnim odabirom. U svakom su uzorku klastera ukupno četiri države, tj. podaci BDP per capita za četiri države svijeta. Odabrane države prikazane su u Tablici broj 4.

Tablica 4 Uzorci pojedinačnih klastera dobivenih klaster analizom podataka BDP per capita svjetskih zemalja za 2015. godinu

Klaster 0	Klaster 1	Klaster 2	Klaster 3
Luksemburg	Argentina	Alžir	Australija
Švicarska	Estonija	Dominika	Francuska
Bermuda	Portoriko	Gruzija	Japan
Norveška	Turska	Rusija	Novi Zeland

Izvor: Rad autora; ; DataBank: World Development Indicators, The World Bank [izvor podataka]. Dostupno na: <https://databank.worldbank.org/source/world-development-indicators>; Weka softver

Podatke BDP per capita u .csv formatu zatim učitvamao u Weka softver u „Preprocess“ odjeljku. Pomoću „Forecast“ odjeljka u softveru, nakon postavljenih postavki za provođenje predikcije vremenskih serija podataka započinjemo sami proces. Prikaz Weka softvera prilikom postavljanja osnovnih postavki i parametara procesa vidljiv je na Slici broj 36, a rezultati samog prognoziranja na Slikama pod brojem 37, 38, 39 i 40.

Slika 36 Proces analize vremenskih serija u Weka softveru

The screenshot shows the Weka Explorer interface with the 'Forecast' tab selected. The 'Target Selection' section includes buttons for 'All', 'None', 'Invert', and 'Pattern'. A table lists selected target countries: Algeria, Argentina, Australia, and Bermuda. The 'Parameters' section on the right is configured with 'Number of time units to forecast' set to 15, 'Time stamp' set to '<Use an artificial time index>', 'Periodicity' set to 'Yearly', 'Confidence intervals' set to 95%, and 'Perform evaluation' checked. The 'Output/Visualization' section displays the following information:

```
Relation: GDF per capita
Instances: 41
Attributes: 17
|>Country Name
Algeria
Argentina
Australia
Bermuda
Dominica
Estonia
France
Georgia
Japan
Luxembourg
New Zealand
Norway
Puerto Rico
Russian Federation
Switzerland
Turkey
```

The status bar at the bottom indicates 'Transforming input data...' and includes a 'Log' button and a small dog icon.

Izvor: Rad autora; DataBank: World Development Indicators, The World Bank [izvor podataka]. Dostupno na: <https://databank.worldbank.org/source/world-development-indicators>; Weka softver

Slika 37 Buduća predviđanja BDP per capita (1)

=== Future predictions from end of training data ===												
inst#	Algeria	Argentina	Australia	Bermuda	Dominica	Estonia	France	Georgia	Japan	Luxembourg	New Zealand	Norway
1	2203.0555	2758.8348	10211.8372	11218.2178	966.6302	?	12713.3652	?	9463.3539	16531.1149	7467.168	15772.2409
2	2237.0868	2776.3221	11854.5885	13425.977	1090.2009	?	11104.9813	?	10360.1783	13837.13	7813.8875	15512.5067
3	2210.302	2927.8974	12788.5493	14166.1713	1193.5101	?	10496.9286	?	9575.6077	12590.9768	7656.4757	15224.8939
4	2312.6556	3553.3775	11538.3984	15902.0204	1324.6495	?	9993.4212	?	10421.2124	12374.0304	7596.312	14927.5175
5	2467.3461	2659.7082	12454.5517	17469.8256	1478.1717	?	9419.6961	?	10978.9198	12126.9392	6713.7601	14989.4658
6	2582.879	2926.1265	11459.1796	18269.5355	1632.2474	?	9763.3274	?	11576.6921	12481.9658	7600.5652	15753.5528
7	2756.955	3613.6217	11394.9382	20450.6645	1863.5061	?	13540.2464	?	17113.2623	18149.8693	9427.5539	18883.2658
8	2807.5036	3562.8761	11645.4526	22411.7957	2113.9272	?	16302.4474	?	20748.9909	22443.4315	12330.9168	22505.8977
9	2417.3762	3985.1925	14279.1594	24253.1755	2402.7718	?	17679.9967	?	25059.0074	25219.3543	13759.1556	24207.2815
10	2215.9639	2383.8675	17828.0482	25517.9217	2617.8168	?	17694.3093	?	24822.7756	24618.0696	13312.3856	24281.0961
11	2408.823	4833.4834	18243.4714	26841.5197	2860.3225	?	21793.8355	1614.6482	25371.4642	33465.4782	13663.0216	28242.9437
12	1781.6214	5735.36	18855.762	27700.3101	3122.6532	?	21765.065	?	28915.0082	35747.3378	12230.0795	28596.933
13	1776.0283	6823.5388	18604.1893	28668.6617	3171.8247	?	23813.7122	?	31414.9946	39570.8616	11793.1392	30523.9851
14	1807.2975	6969.1197	17667.1871	30900.6943	3466.9989	?	22380.1077	547.3349	35681.9639	40066.7243	13094.3455	27963.6652
15	1507.8967	7483.1403	19079.3802	31476.0637	3721.2724	?	23496.5202	519.8576	39933.5151	43933.2354	15280.3128	29315.8419
16	1452.2784	7408.7087	20358.3334	33989.7232	3860.8005	3134.3898	26890.2176	578.3446	44197.6191	51032.3496	17400.4202	34875.7043
17	1603.9403	7721.3541	21904.2941	44826.7891	4120.5812	3380.9263	26871.8313	689.0594	39150.0396	50444.3591	18794.436	37321.9742
18	1619.7977	8213.1251	23509.4235	48478.8833	4292.1137	3682.9523	24228.9464	807.0323	35638.232	46641.6409	17474.187	36629.0309
19	1586.0039	8289.5076	21345.971	51371.7408	4593.8633	4093.3925	24974.2739	851.526	32428.7556	47445.3811	14738.4456	34788.3599
20	1588.3489	7774.7362	20588.9604	54245.4597	4751.0921	4140.9366	24673.203	673.5434	36610.1683	50855.2717	15322.2238	36371.051
21	1745.0271	7708.0991	21697.7085	56284.1686	4787.8014	4070.609	22364.0294	749.9085	39169.3596	48538.7829	13641.1027	38131.4606
22	1740.6067	7208.3731	19527.3236	58893.9594	4883.003	4505.8583	22433.555	801.9904	34406.1825	48399.2293	13882.8568	38542.7151
23	1781.8289	2593.4046	20117.7889	62583.1002	4770.8519	5341.6289	24177.3356	853.5165	32820.7936	52754.2793	16974.1875	43084.4725
24	2103.3813	3349.8063	23492.4052	66111.7252	4897.3084	7203.523	29568.8888	1010.008	35387.0374	65544.592	21913.7082	50134.8908
25	2610.1854	4277.7216	30513.9417	70359.3191	5216.8724	8914.1036	33741.2619	1305.0475	38298.9802	76404.415	25420.2349	57603.836
26	3113.0949	5109.8522	34080.9999	75882.0339	5160.8891	10412.6443	34760.185	1642.7609	37812.895	80957.9413	27751.0655	66810.4785
27	3478.71	5919.0123	36117.488	95221.8589	5518.4232	12639.4001	36443.6263	1996.0571	35981.546	90712.7993	26654.5932	74148.3201
28	3950.513	7245.4469	41001.143	104287.3875	5951.8634	16744.5845	41508.4327	2635.3539	35779.0245	107338.8652	32479.9817	85139.9604
29	4923.6316	9020.8733	49654.9106	106935.4863	6468.9448	18204.9665	45334.118	3324.7359	39876.304	119932.2437	31252.9626	96944.0956
30	3883.2709	8225.1376	42783.3226	101407.764	6903.1458	14711.7353	41575.422	2822.6674	41308.9968	108987.6776	28208.3623	79977.6971
31	4480.7863	10385.9644	52087.9723	101875.2841	6967.365	14663.0446	40638.3353	3233.2959	44968.1562	110777.9079	33676.7741	87693.7901
32	5455.6794	12843.8642	62874.1457	97774.1421	7065.4604	17464.9201	43790.726	4021.7433	48760.0789	118869.2999	38387.6271	100600.5624
33	5592.2201	13082.6643	68027.8417	98431.8652	6849.4559	17404.2008	40874.7004	4421.8182	49145.2804	112591.1162	39973.8808	101524.1419
34	5499.5873	13080.2547	68156.6279	99471.6389	7016.3801	19050.5059	42592.9354	4623.7457	40898.6479	119966.0349	42976.6496	102913.4508
35	5493.0567	12334.7982	62511.6906	98467.684	7317.4861	20234.1174	43011.2631	4739.1883	38476.3952	123514.1967	44572.8988	97019.1828

Izvor: Rad autora; DataBank: World Development Indicators, The World Bank [izvor podataka]. Dostupno na: <https://databank.worldbank.org/source/world-development-indicators>; Weka softver

Izvor: Rad autora; DataBank: World Development Indicators, The World Bank [izvor podataka]. Dostupno na: <https://databank.worldbank.org/source/world-development-indicators>; Weka softver

Slika 40 Buduća predviđanja BDP per capita (4)

14304.405	1834.8618	42740.1282	4499.7375
15220.9913	1330.7572	41786.2341	4116.1706
16192.127	1771.5941	38952.0342	4337.478
18123.1987	2100.3525	39727.8467	3142.921
18731.4594	2377.5296	42578.7631	3687.9561
19557.1202	2975.1254	49470.3971	4760.104
20988.9923	4102.3648	54878.471	6101.6321
21959.3227	5323.4631	56546.7857	7456.2961
22935.9412	6920.1891	59300.6172	8101.8569
23664.8823	9101.255	65359.5192	9791.8825
24898.3346	11635.2729	74572.2324	10941.1721
25768.7259	8562.8133	72083.1677	9103.4741
26435.7488	10674.9958	77117.126	10742.775
27278.883	14311.0843	91254.0348	11420.5555
27944.7339	15420.8745	86547.6709	11795.6335
28513.1657	15974.6446	88109.4868	12614.7816
28981.4573	14095.6487	89684.7076	12157.9904
29763.4883	9313.0136	84776.1422	11006.2795
30627.1634	8704.8984	83073.2801	10894.6034
31108.7606	10720.3327	83352.0887	10589.6677
31604.7015	11287.3553	86388.405	9454.3484
32850.5486	11497.6492	85334.5195	9121.5152
32290.9211	10126.7218	87097.0365	8536.4333
33538.413	12336.2854	97179.1056	11761.293
34059.8737	13651.6318	97495.1979	11932.9094
34812.2538	13316.4708	94025.1494	13236.1889
35058.937	13802.8213	99419.6918	13667.1374
34473.6171	11059.7674	101766.6561	12574.4238
35377.1212	11530.2925	97390.7677	11794.4154
34446.8013	13351.1265	95800.2623	11709.9164
34897.5036	14604.6467	99995.5077	13943.386
36558.151	14446.1662	101420.9812	13262.3069
36653.1144	13024.4634	95038.1282	12002.7629
36780.9582	14175.08	100244.2333	12330.4315
37171.8257	13666.7657	106008.8467	12147.8576
37695.1578	13018.05	97663.6219	11239.2783
37991.238	14035.6629	97092.6682	12246.0147
38013.6148	15185.0766	104926.9515	14242.4087

Izvor: Rad autora; DataBank: World Development Indicators, The World Bank [izvor podataka]. Dostupno na: <https://databank.worldbank.org/source/world-development-indicators>; Weka softver

4.4. Analiza rezultata

4.4.1. Klaster analiza

Klaster analiza napravljena je nad podacima BDP per capita, Ginijevog koeficijenta te očekivane životne dobi za godine 1995., 2005. te 2015. godinu. Broj klastera za svaku godinu određen je na ukupno četiri. U Tablici broj 5 prikazane su vidljive vrijednosti centroida klastera dobivenih analizom pokazatelja BDP per capita kroz navedene godine. U Tablici broj 6, prikazanoj neposredno nakon Tablice broj 5, vidljive su pak postotne promjene u vrijednostima centroida klastera.

Tablica 5 Centroidi klastera BDP per capita

Godina	Klaster 0	Klaster 1	Klaster 2	Klaster 3
1995.	70.421,66	9.489,18	1.565,60	25.926,05
2005.	76.187,57	15.057,93	2.471,88	36.486,83
2015.	104.009,98	18.110,49	3.595,28	43.718,98

Izvor: Rad autora; DataBank: World Development Indicators, The World Bank [izvor podataka]. Dostupno na: <https://databank.worldbank.org/source/world-development-indicators>; Weka softver

Tablica 6 Postotne promjene u vrijednostima centroida klastera BDP per capita

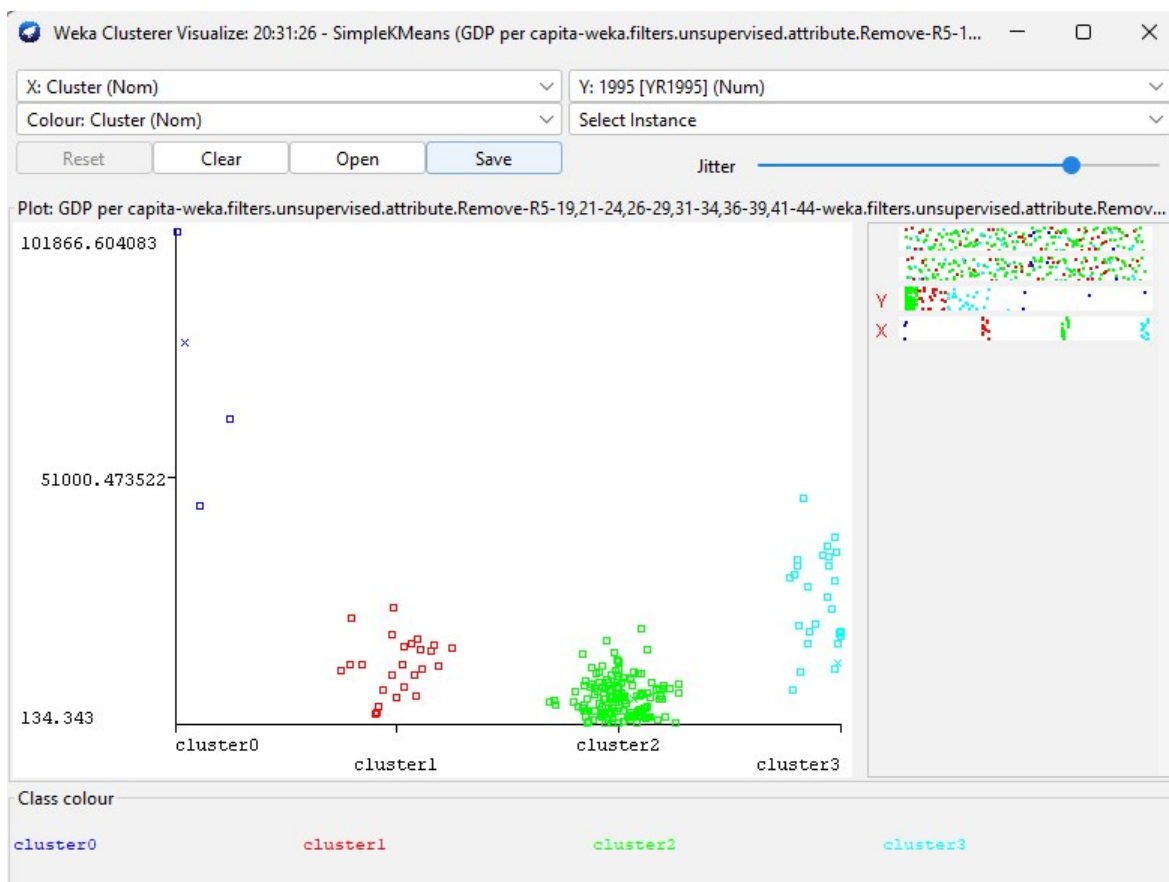
Godina	Klaster 0	Klaster 1	Klaster 2	Klaster 3
2005./1995.	8%	59%	58%	41%
2015./2005.	37%	20%	45%	20%

Izvor: Rad autora; DataBank: World Development Indicators, The World Bank [izvor podataka]. Dostupno na: <https://databank.worldbank.org/source/world-development-indicators>; Weka softver

Promatrajući navedene rezultate nameće se zaključak kako je Klaster 0 klaster najbogatijih zemalja svijeta te zatim redom prema razini bogatstva slijede Klaster 3, Klaster 2 te na posljednjem mjestu Klaster 1. Promotrimo li razliku između vrijednosti centroida najbogatijeg klastera (Klaster 0) te najsiromašnijeg (Klaster 1) uočiti ćemo veliku razliku u razini BDP per capita. Upravo ta razlika prikazuje globalnu ekonomsku nejednakost. Promotrimo li trend kojim se pokazatelj BDP per capita mijenja kroz godine, odnosno u ovom slučaju trend rasta vrijednosti centroida spomenutih klastera, uočiti ćemo kako je u razdoblju od 1995. do 2005. rast najsiromašnijih zemalja svijeta dosegao visokih 59%, usporedno s rastom najbogatijih zemalja svijeta u istom razdoblju od samo 8%. U razdoblju nakon, od 2005. do 2015. situacije se mijenja te dolazi do obrnutog slučaja u kojemu najsiromašnije zemlje usporavaju svoj rast na 20%, dok najbogatije zemlje ostvaruju rast vrijednosti centroida svoga klastera od visokih 37%. Klaster 2, drugi najsiromašniji klaster te ujedno treći najbogatiji, kroz godine prikazuje stabilan rast u vrijednostima od 58% te 45%, dok Klaster broj 3 vrijednost svog centroida u razdoblju od 1995. do 2005. podiže za visokih 41%, ali pak taj isti rast usporava u razdoblju nakon na nešto nižih 20%. Zanimljiva je i činjenica da ukupni zbroj vrijednosti centroida Klastera 3, Klastera 2 i Klastera 1 čini u 2015. godini ukupno 62,90% vrijednosti centroida najbogatijeg klastera. U 2015. godini najbogatijem je klasteru pripadalo svega 9 svjetskih zemalja, Klasteru broj 3 31, Klasteru broj 1 47 zemalja te Klasteru broj 2, ujedno i najsiromašnijem klasteru, čak 130 svjetskih zemalja.

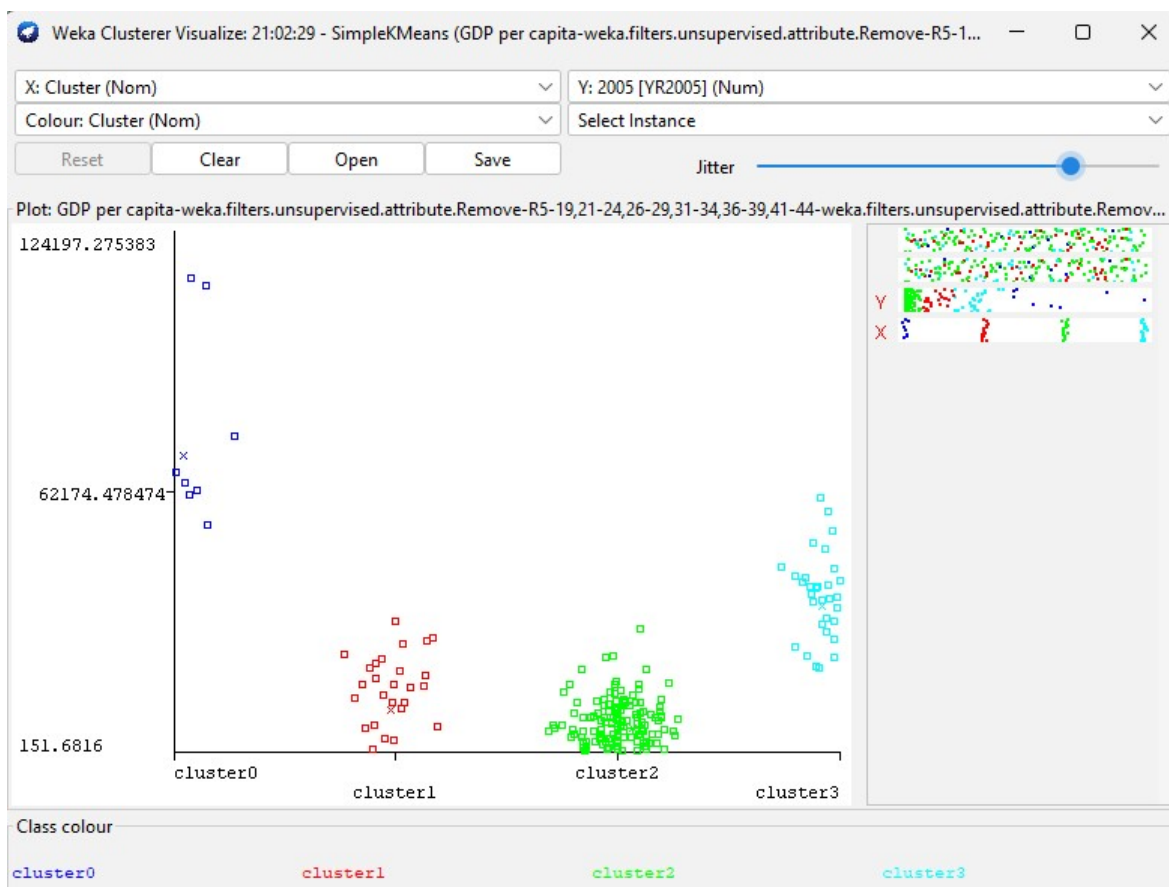
Na Slikama pod brojem 41, brojem 42 te brojem 43 grafički je prikazana podjela i rasprostranjenost svjetskih zemalja u određenim klasterima. Na priloženim je grafičkim prikazima vidljivo kako u najbogatijem klasteru postoji nekolicina zemalja koja drastično odskaka od prosjeka te centroida klastera. Isto tako, vidljivo je, upravo ono što je jasno i iz rezultata klaster analize, kako su svjetske zemlje smještene u klastere broj 1 i 2, ujedno dva najsiromašnija klaster vrlo sličnih vrijednosti BDP-a per capita. Naizgled se čini kako je koncentracija vrijednosti zemalja prema centroidu klastera najjača u klasteru broj 2, najsiromašnijem klasteru, no, treba uzeti u obzir kako se u njemu, također, nalazi i najveći broj svjetskih zemalja.

Slika 41 Grafički prikaz rasprostranjenosti zemalja u klasterima BDP per capita 1995. godine



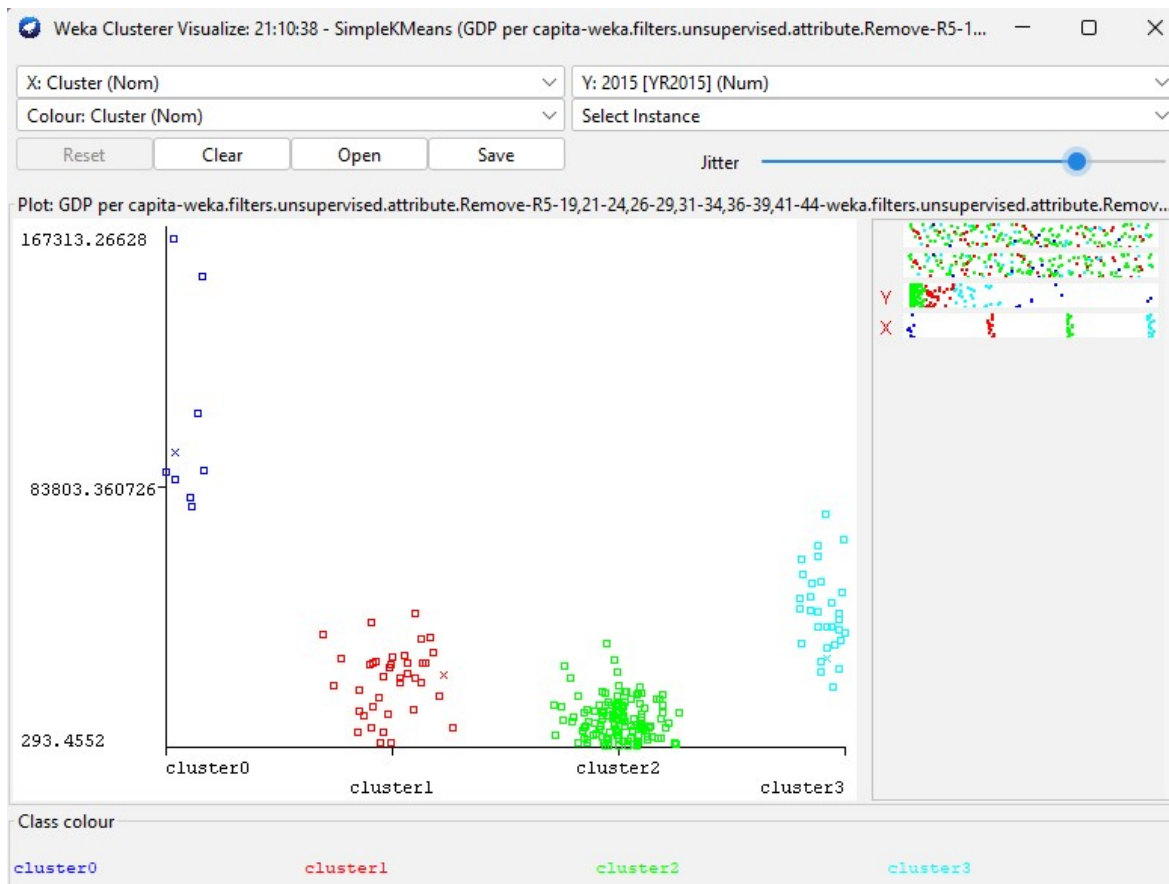
Izvor: Rad autora; DataBank: World Development Indicators, The World Bank [izvor podataka]. Dostupno na: <https://databank.worldbank.org/source/world-development-indicators>; Weka softver

Slika 42 Grafički prikaz rasprostranjenosti zemalja u klasterima BDP per capita 2005. godine



Izvor: Rad autora; DataBank: World Development Indicators, The World Bank [izvor podataka]. Dostupno na: <https://databank.worldbank.org/source/world-development-indicators>; Weka softver

Slika 43 Grafički prikaz rasprostranjenosti zemalja u klasterima BDP per capita 2015. godine



Izvor: Rad autora; DataBank: World Development Indicators, The World Bank [izvor podataka]. Dostupno na: <https://databank.worldbank.org/source/world-development-indicators>; Weka softver

U Tablici broj 7 prikazane su vidljive vrijednosti centroida klastera dobivenih analizom pokazatelja Ginijevog koeficijenta kroz odabrane godine analize.

Tablica 7 Centroidi klastera Ginijevog koeficijenta

Godina	Klaster 0	Klaster 1	Klaster 2	Klaster 3
1995.	35,11	54,06	26,44	42,49
2005.	33,46	39,51	51,78	27,98
2015.	35,51	49,58	29,10	40,90

Izvor: Rad autora; DataBank: World Development Indicators, The World Bank [izvor podataka]. Dostupno na: <https://databank.worldbank.org/source/world-development-indicators>; Weka softver

Vrijednost Ginijevog koeficijenta označava razinu nejednakosti unutar zemlje. Čim višom vrijednosti koeficijenta te bližom maksimalnoj vrijednosti od 100, zemlja se smatra dublje podijeljenom te kao zemlja veće nejednakosti. Isto tako, čim je vrijednost Ginijevog koeficijenta niža te bliža minimalnoj vrijednosti, nuli, tako se zemlja smatra više jednakom. U godini 1995. možemo zaključiti kako je Klaster 2 taj čije zemlje imaju najvišu razinu jednakosti, a zatim ga slijede Klaster 0, Klaster 3 te posljednji Klaster 1. Klaster 3 pak u 2005. godini preuzima ulogu kao klaster s najnižim centroidom i samim time najvišim stupnjem jednakosti unutar zemalja koje mu pripadaju, a slijede ga Klaster 0, Klaster 1 te posljednji Klaster 2. U zadnjoj godini provedene klaster analize pobjedničku ulogu kao klaster s najnižim centroidom vrijednosti Ginijevog koeficijenta preuzima klaster broj 2, a slijede ga redom Klaster broj 0, Klaster broj 3 te Klaster broj 1 na posljednjem mjestu.

U Tablici broj 8 prikazane su vidljive vrijednosti centroida klastera dobivenih analizom pokazatelja očekivane životne dobi kroz goine provedene analize. U Tablici broj 9, prikazanoj neposredno nakon tablice broj 8, vidljive su pak postotne promjene u vrijednostima centroida klastera.

Tablica 8 Centroidi klastera očekivane životne dobi

Godina	Klaster 0	Klaster 1	Klaster 2	Klaster 3
1995.	74,93	57,30	46,88	67,34
2005.	77,60	62,20	51,13	70,34
2015.	80,26	57,16	64,94	73,12

Izvor: Rad autora; DataBank: World Development Indicators, The World Bank [izvor podataka]. Dostupno na: <https://databank.worldbank.org/source/world-development-indicators>; Weka softver

Tablica 9 Postotne promjene u vrijednostima centroida klastera očekivane životne dobi

Godina	Klaster 0	Klaster 1	Klaster 2	Klaster 3
2005./1995.	4%	9%	9%	4%
2015./2005.	3%	-8%	27%	4%

Izvor: Rad autora; DataBank: World Development Indicators, The World Bank [izvor podataka]. Dostupno na: <https://databank.worldbank.org/source/world-development-indicators>; Weka softver

Promatrajući Tablicu broj 8 s prikazanim vrijednostima centroida pojedinih klastera pokazatelja očekivane životne dobi, dolazi se do zaključka kako su zemlje Klastera 0 vodeće kroz sve tri promatrane te analizirane godine te da stanovnici tih zemalja žive znatno dulje od na primjer stanovnika zemalja koje pripadaju Klasteru broj 1 ili Klasteru broj 2. Klaster broj 3 te stanovnici zemalja koje pripadaju navedenom klasteru imaju očekivanu životnu dob konstantno na drugom mjestu poretka klastera. Postotne promjene

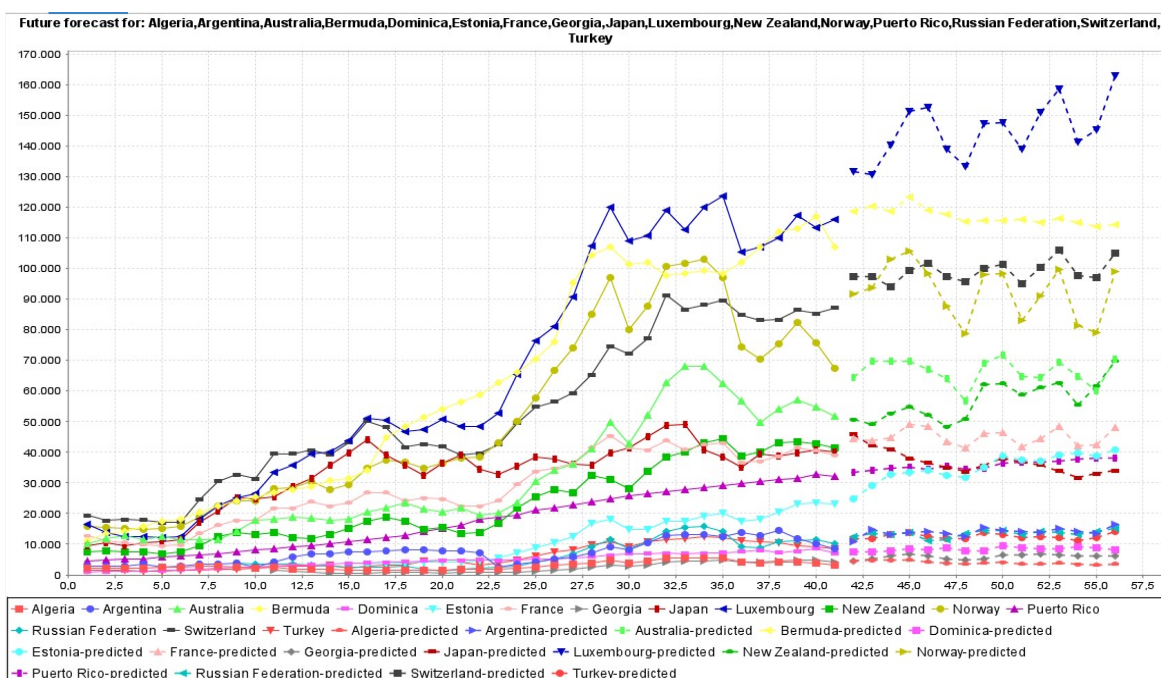
ukazuju na puno brži rast očekivane životne dobi Klastera 2 u odnosu na Klaster 0 ili Klaster 3, čije se zemlje smatraju razvijenim zemljama te u razdoblju od 2005. godine do 2015. godine doživljava rast od čak 27%. Zabrinjavajući je podatak da se u zemljama koje pripadaju Klasteru broj 1 očekivana životna dob zapravo smanjila za 8%. Za razliku od pokazatelja BDP per capita, pozitivni zaključak klaster analize pokazatelja očekivane životne dobi je taj što u Klaster 0 te Klaster 3, dva klastera zemalja s najduljom očekivanom životnom dobi, pripada čak 73% analiziranih zemalja u 2015. godini.

4.4.2. Analiza vremenske serije

Analiza vremenske serije, odnosno proces predviđanja vremenskih serija kroz softver Weka napravljen je na uzorcima zemalja iz pojedinih klastera dobivenih klaster analizom razina BDP per capita svjetskih zemalja. U svakom je uzorku nasumičnom odabirom bilo po četiri zemlje, tj. četiri zemlje pojedinih klastera. Za odabranih ukupno šesnaest zemalja svijeta napravljena je analiza vremenske serije, odnosno predviđanje za kretanje razine BDP-a per capita u narednih 15 godina. Rezultati predviđanja kretanja razine BDP-a per capita vidljivi su na Grafikonu broj 6.

Zanimljivost koja se pri Grafikonu broj 6 prikazuje je predikcija kretanja pokazatelja BDP per capita za zemlje najbogatijeg klastera dobivenog klaster analizom. Od četiri zemlje najbogatijeg klastera, opći trend rasta, uz doduše pojedina razdoblja značajnog pada, predviđeno je samo za jednu zemlju, Luksemburg. Za preostale zemlje navedenog klastera vidljiva je opća stagnacija (Švicarska) u kretanju BDP-a per capita, ili čak u nekim slučajevima (Bermuda, Norveška) blagi pad. Unutar Klastera broj 3, drugog najbogatijeg klastera, vrijedi istaknuti prognozu rasta zemalja poput Novog Zelanda i Australije, relativnu stagnaciju Francuske te blagi, ali vrlo vjerojatno osjetni, pad Japana.

Grafikon 6 Predviđanje kretanja pokazatelja BDP per capita



Izvor: Rad autora; DataBank: World Development Indicators, The World Bank [izvor podataka]. Dostupno na: <https://databank.worldbank.org/source/world-development-indicators>; Weka softver

Prebacivanjem fokusa na zemlje dvaju najsiromašnijih klastera, vrijedi istaknuti dvije zemlje koje u ovoj prognozi obećavaju najviše kretanjem pokazatelja BDP per capita. Te dvije zemlje su Estonija i Portoriko. Za Estoniju i Portoriko na temelju grafikona stvara se zaključak kako bi kroz narednih 15 godina vrlo vjerojatno dospjeli u Klaster broj 3, drugi najbogatiji klaster prema izvršenoj klaster analizi. No, te su dvije države ujedno i jedine koje ovom analizom vremenske serije pokazuju znakove potencijala za prelazak u više razine BDP-a per capita. Preostale zemlje dvaju siromašnijih klastera ovom analizom daju znakove blage stagnacije ili pak jedva zamjetnog rasta.

Usporedbom karakteristika gospodarstava zemalja poput Estonije i Portorika kao dvije zemlje koje teže ka značajnom rastu razine BDP-a per capita kroz narednih 15 godina, ali isto tako i zemalja poput Japana, Norveške i Bermude s ostalim svjetskim zemljama te karakteristikama njihovih gospodarstava, rezultati ove analize mogli bi se poopćiti. No, navedeno bi moglo biti dosta nezahvalno s obzirom na posebitosti različitih svjetskih gospodarstava te vrlo česte međuovisnosti istih na globalnoj ekonomskoj sceni.

5. Zaključak

Globalna ekonomska nejednakost prisutna je u svim sferama života te izravno pogađa stanovnike svih zemalja svijeta te igra značajnu ulogu u osnovnim procesima svakog gospodarstva. Možemo ju promatrati kao nejednakost unutar država, ali isto tako kao nejednakost između različitih svjetskih gospodarstava i time uočiti veliki jaz između životnog standarda stanovnika bogatih i razvijenih svjetskih zemalja te onih koji nisu imali tolike sreće da se rode u dobrim uvjetima, stanovnika zemalja u razvoju. Pri analizi ekonomske nejednakosti, bilo između zemalja ili unutar zemlje, najčešće se promatraju pokazatelji poput BDP-a per capita ili prosječnog realnog dohotka po stanovniku, ali ne smijemo zaboraviti pokazatelje kvalitete života poput očekivanje životne dobi ili pak stope smrtnosti novorođene djece. Upravo je to razlog zbog kojeg se u ovome radu pri analizi globalne ekonomske nejednakosti otkrivanjem znanja iz baza podataka Svjetske banke među analiziranim pokazateljima osim BDP-a per capita te Ginijevog koeficijenta pronašao i pokazatelj očekivane životne dobi pri rođenju.

Zahvaljujući razvoju tehnologije te obilju podataka kojemu danas svjedočimo, korištenje strojnog učenja, skladištenja podataka te različitih metoda otkrivanja znanja iz baza podataka stvorilo je mogućnost kreiranja brzih te pouzdanih analiza povijesnih ekonomskih podataka te stvaranje relativno pouzdanih predikcija kretanja određenih ekonomskih pokazatelja kroz budućnost. Kroz ovaj diplomski rad prikazan je proces otkrivanja znanja u bazama podataka. Organizacijom podataka u klustere, odnosno skupine pomoću provođenja klaster analize ekonomisti imaju priliku bolje razumjeti povijesne podatke te kretanja pokazatelja globalne ekonomske nejednakosti.

Cilj ovog rada bio je analiza globalne ekonomske nejednakosti te korištenje otkrivanja znanja u bazama podataka za ispitivanje nejednakosti i razumijevanje čimbenika koji tome pridonose. Podaci su prilikom izrade diplomskog rada te provođenja istraživanja preuzeti sa službenih stranica Svjetske banke. Korištenjem softvera otvorenog koda Weka napravljena je klaster analiza te analiza, tj. prognoziranje vremenskih serija. Analizirana su tri pokazatelja u trima odabranim godinama (1995., 2005. te 2015.): BDP per capita (ekonomska nejednakost između zemalja), Ginijev koeficijent (dohodovne nejednakosti unutar država) te očekivana životna dob pri rođenju (razvijenost zemlje i kvalitetu života u istoj). Klaster analizom podataka BDP-a per capita prikazana je velika razlika između

vrijednosti centroida najbogatijeg klastera te najsiromašnijeg klastera, kao i činjenica da zbroj vrijednosti centroida triju od ukupno četiri klastera čini svega 62,90% vrijednosti centroida najbogatijeg klastera. Analizom je uočen i pozitivan trend puno bržeg, tj. većeg rasta vrijednosti BDP-a per capita zemalja siromašnijih klastera od zemalja bogatijih klastera. Analizom vremenske serije, tj. predviđanjem vremenskih serija napravljen je pogled u budućnost za ukupno 15 narednih godina. Analiza je napravljena nad podacima BDP-a per capita nad ukupno 16 svjetskih zemalja. Zemlje dvaju siromašnijih klastera, određenih u prvom dijelu istraživanja klaster analizom, pokazuju veći potencijal za značajnim rastom porastom vrijednosti BDP-a per capita te prelaskom u „viši rang“ država prema navedenom pokazatelju. Odnosno, optimistično vjerovanje na temelju analize moglo bi biti da nas očekuje budućnost manje razine globalne ekonomske nejednakosti te više razine kvalitete života za veći dio ljudske populacije u svijetu.

Literatura

Knjige i znanstveni članci:

1. Cornish, R. (2007) Cluster Analysis. Mathematics Learning Support Centre.
[online] Dostupno na:
<http://www.statstutor.ac.uk/resources/uploaded/clusteranalysis.pdf>
2. Ćurko, K. i Španić M. (2016) Skladištenje podataka: put do znanja i poslovne inteligencije. Zagreb: Ekonomski fakultet
3. Pejić Bach, M. i Kerep, I. (2011) Weka: alat za otkrivanje znanja iz baza podataka. Zagreb: Mikrorad
4. Kantardžić, M. (2011) Data mining: concepts, models, methods, and algorithms. 2 izd. Hoboken: Wiley-Interscience
5. Han, J., Kamber, M. & Pei, J. (2012) Data mining: Concepts and techniques, 3rd.ed, Waltham: Morgan Kaufmann Publishers.
6. Witten, I., H., Frank, E. & Hall, M., A. (2005) Data Mining: Practical Machine Learning, 2nd ed. Amsterdam: Elsevier
7. Sharma, N., Bajpai, A., & Litoriya, R. (2012) Comparison the various clustering algorithms of weka tools, International Journal of Emerging Technology and Advanc Engineering [online], 2 (5). Dostupno na:
https://ijetae.com/files/Volume2Issue5/IJETAE_0512_13.pdf
8. Dragičević, O. (2016) Savremeni trendovi međunarodne ekonomske nejednakosti, Politička misao, god. 53, br. 2, 2016, str. 31-63. [online]. Dostupno na:
<https://hrcak.srce.hr/file/249613>
9. Milanovic, B. (2015) Global Inequality Of Opportunity: How Much Of Our Income Is Determined By Where We Live? The Review of Economics and Statistics, 97(2), str. 452–460. [online]. Dostupno na: <http://www.jstor.org/stable/43556185>
10. Milanovic, B. (2016) Global Inequality: A New Approach for the Age of Globalization. Cambridge, MA and London, England: Harvard University Press [online]. Dostupno na: <https://doi.org/10.4159/9780674969797>
11. Anand, Sudhir i Segal, Paul (2008) What Do We Know about Global Income Inequality?, Journal of Economic Literature, 46, str. 57-94. 10.1257/jel.46.1.57.
12. Mills, M. (2009) Globalization and Inequality. European Sociological Review, 25(1), str. 1–8. [online]. Dostupno na: <http://www.jstor.org/stable/25548302>

13. Stewart, Frances, Graham K. Brown i Luca Mancini (2005) Why horizontal inequalities matter: some implications for measurement, Centre for Research on Inequality, Human Security and Ethnicity. CRISE Working Paper No. 19.
14. Milanovic, B., Lindert, P.H. i Williamson, J.G. (2011) Pre-industrial Inequality, The Economic Journal, Vol. 121, No. 551, pp. 255-272, Wiley on behalf of the Royal Economic Society [online]. Dostupno na: <http://www.jstor.org/stable/41057775>
15. Acemoglu, D. i Robinson, J.A. (2012) Zašto nacije propadaju: izvorišta moći, napretka i siromaštva. Zagreb: Mate d.o.o.

Web stranice:

1. Roser, M. (2021) Global economic inequality: what matters most for your living conditions is not who you are, but where you are [online]. England and Wales: Our World In Data. Dostupno na: <https://ourworldindata.org/global-economic-inequality-introduction>
2. Marr, B. (2018) How Much Data Do We Create Every Day? The Mind-Blowing Stats Everyone Should Read [online]. New Jersey, SAD: Forbes. Dostupno na: <https://www.forbes.com/sites/bernardmarr/2018/05/21/how-much-data-do-we-create-every-day-the-mind-blowing-stats-everyone-should-read/>
3. DataBank: World Development Indicators, The World Bank [izvor podataka]. Dostupno na: <https://databank.worldbank.org/source/world-development-indicators>
4. How does the World Bank classify countries?, The World Bank. Dostupno na: <https://datahelpdesk.worldbank.org/knowledgebase/articles/378834-how-does-the-world-bank-classify-countries>
5. Attribute-Relation File Format (ARFF), Sveučilište Waikato. Dostupno na: <https://www.cs.waikato.ac.nz/ml/weka/arff.html>

Popis slika

Slika 1 Proces otkrivanja znanja iz baza podataka	20
Slika 2 "Preprocess" u Weka softveru	35
Slika 3 Prikaz prozora "Cluster"	36
Slika 4 Postavke algoritma SimpleKMeans.....	37
Slika 5 Prikaz rezultata klaster analize BDP per capita za 1995. godinu	38
Slika 6 BDP per capita 1995. godine: pripadnost države klasteru (1).....	40
Slika 7 BDP per capita 1995. godine: pripadnost države klasteru (2).....	41
Slika 8 BDP per capita 1995. godine: pripadnost države klasteru (3).....	42
Slika 9 Prikaz rezultata klaster analize BDP per capita za 2005. godinu	43
Slika 10 BDP per capita 2005. godine: pripadnost države klasteru (1).....	44
Slika 11 BDP per capita 2005. godine: pripadnost države klasteru (2).....	45
Slika 12 BDP per capita 2005. godine: pripadnost države klasteru (3).....	46
Slika 13 Prikaz rezultata klaster analize BDP per capita za 2015. godinu	47
Slika 14 BDP per capita 2015. godine: pripadnost države klasteru (1).....	48
Slika 15 BDP per capita 2015. godine: pripadnost države klasteru (2).....	49
Slika 16 BDP per capita 2015. godine: pripadnost države klasteru (3).....	50
Slika 17 Prikaz probranih zemalja za klaster analizu pokazatelja Ginijev koeficijent...	51
Slika 18 Prikaz rezultata klaster analize Ginijev koeficijent za 1995. godinu	52
Slika 19 Ginijev koeficijent 1995. godine: pripadnost države klasteru	53
Slika 20 Prikaz rezultata klaster analize Ginijev koeficijent za 2005. godinu	54
Slika 21 Ginijev koeficijent 2005. godine: pripadnost države klasteru	55
Slika 22 Prikaz rezultata klaster analize Ginijev koeficijent za 2015. godinu	56
Slika 23 Ginijev koeficijent 2015. godine: pripadnost države klasteru	57
Slika 24 Prikaz rezultata klaster analize očekivanu životnu dob za 1995. godinu	59
Slika 25 Očekivana životna dob 1995. godine: pripadnost države klasteru (1).....	60

Slika 26 Očekivana životna dob 1995. godine: pripadnost države klasteru (2).....	61
Slika 27 Očekivana životna dob 1995. godine: pripadnost države klasteru (3).....	62
Slika 28 Prikaz rezultata klaster analize očekivanu životnu dob za 2005. godinu	63
Slika 29 Očekivana životna dob 2005. godine: pripadnost države klasteru (1).....	64
Slika 30 Očekivana životna dob 2005. godine: pripadnost države klasteru (2).....	65
Slika 31 Očekivana životna dob 2005. godine: pripadnost države klasteru (3).....	66
Slika 32 Prikaz rezultata klaster analize očekivanu životnu dob za 2015. godinu	67
Slika 33 Očekivana životna dob 2015. godine: pripadnost države klasteru (1).....	68
Slika 34 Očekivana životna dob 2015. godine: pripadnost države klasteru (2).....	69
Slika 35 Očekivana životna dob 2015. godine: pripadnost države klasteru (3).....	70
Slika 36 Proces analize vremenskih serija u Weka softveru	72
Slika 37 Buduća predviđanja BDP per capita (1)	73
Slika 38 Buduća predviđanja BDP per capita (2)	74
Slika 39 Buduća predviđanja BDP per capita (3)	74
Slika 40 Buduća predviđanja BDP per capita (4)	75
Slika 41 Grafički prikaz rasprostranjenosti zemalja u klasterima BDP per capita 1995. godine	78
Slika 42 Grafički prikaz rasprostranjenosti zemalja u klasterima BDP per capita 2005. godine	79
Slika 43 Grafički prikaz rasprostranjenosti zemalja u klasterima BDP per capita 2015. godine	80

Popis tablica

Tablica 1 Tablica atributa BDP per capita.....	32
Tablica 2 Tablica atributa Ginijev koeficijent	32
Tablica 3 Tablica atributa Očekivana životna dob	33
Tablica 4 Uzorci pojedinačnih klastera dobivenih klaster analizom podataka BDP per capita svjetskih zemalja za 2015. godinu	71
Tablica 5 Centroidi klastera BDP per capita	76
Tablica 6 Postotne promjene u vrijednostima centroida klastera BDP per capita.....	76
Tablica 7 Centroidi klastera Ginijevog koeficijenta	81
Tablica 8 Centroidi klastera očekivane životne dobi.....	82
Tablica 9 Postotne promjene u vrijednostima centroida klastera očekivane životne dobi	82

Popis grafikona

Grafikon 1 Relativni dobitak realnog dohotka po stanovniku prema razini globalnog dohotka 1988. - 2008.....	4
Grafikon 2 Raspodjela svjetske populacije prema realnom BDP-u po glavi stanovnika zemlje u kojoj ljudi žive (2013. godina).....	6
Grafikon 3 Nejednakost u Engleskoj/UK i Sjedinjenim Državama od 17. stoljeća do 21. stoljeća.....	9
Grafikon 4 Globalna nejednakost, 1820. - 2011.....	11
Grafikon 5 Udio međudržavne komponente u globalnoj nejednakosti, 1820. - 2011....	13
Grafikon 6 Predviđanje kretanja pokazatelja BDP per capita	84