

Analiza podataka o COVID-19 pandemiji

Vratović, Chiara

Master's thesis / Diplomski rad

2023

Degree Grantor / Ustanova koja je dodijelila akademski / stručni stupanj: **University of Rijeka, Faculty of Engineering / Sveučilište u Rijeci, Tehnički fakultet**

Permanent link / Trajna poveznica: <https://um.nsk.hr/um:nbn:hr:190:703415>

Rights / Prava: [Attribution 4.0 International](#)/[Imenovanje 4.0 međunarodna](#)

Download date / Datum preuzimanja: **2025-03-11**



Repository / Repozitorij:

[Repository of the University of Rijeka, Faculty of Engineering](#)



SVEUČILIŠTE U RIJECI
TEHNIČKI FAKULTET
Diplomski studij računarstva

Diplomski rad

Analiza podataka o COVID-19 pandemiji

Rijeka, srpanj 2023.

Chiara Vratović
0069082770

SVEUČILIŠTE U RIJECI
TEHNIČKI FAKULTET
Diplomski studij računarstva

Diplomski rad

Analiza podataka o COVID-19 pandemiji

Mentor: prof. dr. sc. Zoran Čarija

Rijeka, srpanj 2023.

Chiara Vratović
0069082770

Umjesto ove stranice umetnuti zadatak
za završni ili diplomski rad

Izjava o samostalnoj izradi rada

Izjavljujem da sam samostalno izradila ovaj rad.

Rijeka, srpanj 2023.

Chiara Vratović

Zahvala

Zahvaljujem v. asist. dr. sc. Ivani Lučin na podršci tijekom pisanja ovoga rada i korisnim raspravama i savjetima. Zahvaljujem obitelji i prijateljima na podršci tijekom studiranja. Bogu hvala, gotovo je!

Sadržaj

| | |
|--|----------|
| Popis slika | ix |
| Popis tablica | xi |
| 1 Uvod | 1 |
| 2 O pandemiji | 3 |
| 2.1 Općenito o COVID-19 | 3 |
| 2.2 Pozadina i značaj pandemije | 4 |
| 2.3 Trenutno stanje pandemije (2023.) | 4 |
| 2.4 Modeli predviđanja COVID-19 pandemije | 5 |
| 3 Metodologija | 6 |
| 3.1 Izvori podataka | 6 |
| 3.2 Metode analize podataka | 8 |
| 3.2.1 Deskriptivna analiza | 8 |
| 3.2.2 Korelacijska analiza | 8 |
| 3.2.3 Analiza trenda korištenjem regresije | 9 |
| 3.2.4 Vizualizacija podataka | 11 |
| 3.3 Korišteni softver | 12 |
| 3.4 Varijable i mjere | 13 |

Sadržaj

| | | |
|----------|---|-----------|
| 3.4.1 | COVID-19 | 13 |
| 3.4.2 | Modeli predviđanja | 14 |
| 3.5 | Ograničenja i pretpostavke | 15 |
| 4 | Usporedba COVID-19 podataka kroz 2020., 2021. i 2022. | 17 |
| 4.1 | Rezultati deskriptivne analize | 17 |
| 4.2 | Rezultati korelacijske analize | 20 |
| 4.2.1 | Korelacija procijepljenosti i smrtnosti | 20 |
| 4.2.2 | Korelacija gustoće naseljenosti i smrtnosti | 21 |
| 4.2.3 | Korelacija srednje starosti stanovništva i smrtnosti | 22 |
| 4.2.4 | Korelacija stanovništva starijeg od 70 godina i smrtnosti | 23 |
| 4.2.5 | Korelacija udjela stanovništva koje živi u ekstremnom siromaštvu i smrtnosti | 24 |
| 4.2.6 | Korelacija BDP-a po glavi stanovnika i smrtnosti | 25 |
| 4.2.7 | Korelacija indeksa strogosti odgovora vlade i slučajeva zaraze | 26 |
| 4.3 | Rezultati analize trenda | 27 |
| 4.3.1 | Značajnost trenda | 33 |
| 5 | Trenutno stanje pandemije (2023.) | 38 |
| 5.1 | Stanje u svijetu | 38 |
| 5.2 | Usporedba s prijašnjim godinama | 41 |
| 5.3 | Stanje u Hrvatskoj | 45 |
| 6 | Analiza modela predviđanja tijekom COVID-19 pandemije | 52 |
| 6.1 | Analiza prosječne točnosti modela za predviđanja COVID-19 u Europi | 52 |
| 6.2 | Analiza točnosti i pouzdanosti pojedinih modela za predviđanje širenja virusa | 56 |
| 6.3 | Opis najuspješnijih modela | 63 |

Sadržaj

| | | |
|----------|--|-----------|
| 6.3.1 | Epiforecasts tim | 64 |
| 6.3.2 | MUNI tim | 65 |
| 6.3.3 | Swiss Data Science Center tim | 66 |
| 6.4 | Analiza točnosti i pouzdanosti pojedinih modela po polugodištu | 67 |
| 7 | Zaključak | 70 |
| | Bibliografija | 73 |
| | Pojmovnik | 75 |
| | Sažetak | 76 |

Popis slika

| | | |
|------|---|----|
| 3.1 | Vizualno objašnjenje kutijastog dijagrama. Oznake s lijeve strane daju nazive za grafičke elemente, oznake s desne strane daju odgovarajuću sažetu statistiku.[1] | 12 |
| 4.1 | Kutijasti dijagrami novih slučajeva zaraze po kontinentima | 18 |
| 4.2 | Kutijasti dijagrami novih smrtnih slučajeva po kontinentima | 18 |
| 4.3 | Linearni trendovi novih slučajeva zaraze po kontinentima | 28 |
| 4.4 | Linearni trendovi novih slučajeva zaraze - svi kontinenti | 28 |
| 4.5 | Linearni trendovi novih slučajeva smrti po kontinentima | 29 |
| 4.6 | Linearni trendovi novih slučajeva smrti - svi kontinenti | 30 |
| 4.7 | Kvadratni trendovi novih slučajeva zaraza i smrti - Europa, 2020. godina | 31 |
| 4.8 | Kvadratni trendovi novih slučajeva zaraza i smrti - Europa, 2021. godina | 31 |
| 4.9 | Kvadratni trendovi novih slučajeva zaraza i smrti - Europa, 2022. godina | 32 |
| 4.10 | Kvadratni trendovi novih slučajeva zaraza i smrti - Europa, 2023. godina | 32 |
| 5.1 | Apsolutan broj slučajeva zaraze i smrti u 2023. po kontinentima | 39 |
| 5.2 | Relativan broj slučajeva zaraze i smrti u 2023. po kontinentima, na milijun stanovnika | 40 |

Popis slika

| | | |
|------|--|----|
| 5.3 | Stupčasti dijagram ukupnih slučajeva zaraze i smrtnih slučajeva po svim kontinentima od 2020. do 2023. | 42 |
| 5.4 | Stupčasti dijagram ukupnih slučajeva zaraze i smrtnih slučajeva po svim kontinentima od 2020. do 2023. u logaritamskoj skali | 42 |
| 5.5 | Stupčasti dijagram relativnih slučajeva zaraze i smrtnih slučajeva po svim kontinentima od 2020. do 2023., normalizirani na milijun stanovnika | 44 |
| 5.6 | Stupčasti dijagram relativnih slučajeva zaraze i smrtnih slučajeva po svim kontinentima od 2020. do 2023., normalizirani na milijun stanovnika u logaritamskoj skali | 44 |
| 5.7 | Godišnji broj slučajeva zaraze i smrtnih slučajeva u Republici Hrvatskoj | 46 |
| 5.8 | Godišnji broj slučajeva zaraze i smrtnih slučajeva u Republici Hrvatskoj u logaritamskoj skali | 46 |
| 5.9 | Relativni godišnji broj slučajeva zaraze i smrtnih slučajeva u Republici Hrvatskoj na milijun njenih stanovnika | 47 |
| 5.10 | Relativni godišnji broj slučajeva zaraze i smrtnih slučajeva u Republici Hrvatskoj na milijun njenih stanovnika, u logaritamskoj skali . | 48 |
| 5.11 | Trendovi novih slučajeva zaraze u Republici Hrvatskoj | 49 |
| 5.12 | Trendovi novih smrtnih slučajeva u Republici Hrvatskoj | 49 |

Popis tablica

| | | |
|-----|---|----|
| 3.1 | Tumačenje rezultata | 9 |
| 4.1 | Rezultati korelacijske analize između procijepljenosti i smrtnosti korištenjem Spearmanovog koeficijenta korelacije ranga | 20 |
| 4.2 | Rezultati korelacijske analize gustoće naseljenosti stanovništva i smrtnosti korištenjem Spearmanovog koeficijenta korelacije ranga | 21 |
| 4.3 | Rezultati korelacijske analize srednje starosti stanovništva i smrtnosti korištenjem Spearmanovog koeficijenta korelacije ranga | 22 |
| 4.4 | Rezultati korelacijske analize broja stanovništva starijeg od 70 godina i smrtnosti korištenjem Spearmanovog koeficijenta korelacije ranga | 23 |
| 4.5 | Rezultati korelacijske analize udjela stanovništva koje živi u ekstremnom siromaštvu i smrtnosti korištenjem Spearmanovog koeficijenta korelacije ranga | 25 |
| 4.6 | Rezultati korelacijske analize BDP-a po stanovniku i smrtnosti korištenjem Spearmanovog koeficijenta korelacije ranga | 26 |
| 4.7 | Rezultati korelacijske analize indeksa strogosti i slučajeva zaraze korištenjem Spearmanovog koeficijenta korelacije ranga | 27 |
| 4.8 | Značajnost linearnog trenda novih slučajeva zaraze i R-kvadrat vrijednosti | 33 |
| 4.9 | Značajnost linearnog trenda novih smrtnih slučajeva i R-kvadrat vrijednosti | 33 |

Popis tablica

| | | |
|------|---|----|
| 4.10 | Značajnost kvadratnog trenda novih slučajeva zaraze i R-kvadrat vrijednosti po godinama | 35 |
| 4.11 | Značajnost kvadratnog trenda novih slučajeva smrti i R-kvadrat vrijednosti po godinama | 36 |
| 5.1 | Slučajevi zaraze i smrtni slučajevi po kontinentima u 2023. (do 20.5.2023.) | 39 |
| 5.2 | Godišnji slučajevi zaraze i smrtni slučajevi u Hrvatskoj | 47 |
| 6.1 | Prosječne vrijednosti metrika točnosti različitih modela predviđanja COVID-19 u Europi | 54 |
| 6.2 | Mean Absolute Error (MAE) za predviđanje porasta slučajeva zaraze i smrtnih slučajeva | 57 |
| 6.3 | Weighted Interval Score (WIS) za predviđanje porasta slučajeva zaraze i smrtnih slučajeva | 58 |
| 6.4 | Prosječne 50%-tne vjerojatnosti pokrivanja za slučajeve zaraze i smrtnih slučajeva | 59 |
| 6.5 | Prosječne 95%-tne vjerojatnosti pokrivanja za slučajeve zaraze i smrtnih slučajeva | 60 |
| 6.6 | Prosječna pristranost za predviđanje porasta slučajeva zaraze i smrtnih slučajeva | 61 |
| 6.7 | Najbolji modeli predviđanja COVID-19 u Europi prema pojedinoj metrici | 63 |
| 6.8 | WIS za predviđanje porasta slučajeva zaraze i smrtnih slučajeva po periodima | 68 |

Poglavlje 1

Uvod

Primarni cilj ovog rada je analizirati i usporediti podatke o COronaVirus Disease of 2019 (COVID-19) iz 2020., 2021. i 2022. godine, s ciljem istraživanja trendova, uzoraka i promjena pandemije tijekom vremena, prikazati trenutnu sliku COVID-19 pandemije te analizirati modele predviđanja pandemije kako bi se utvrdila njihova točnost i pouzdanost. Istraživanje će se usredotočiti na nekoliko ključnih pokazatelja, uključujući broj potvrđenih slučajeva zaraze, smrtnih slučajeva i cijepljenja, kao i distribuciju i težinu virusa u različitim regijama i populacijama.

Da bi se postigao ovaj cilj, rad će se baviti sljedećim istraživačkim pitanjima:

1. Kako su se trendovi i uzorci podataka o COVID-19 mijenjali od 2020. do 2022. godine?
2. Koje su sličnosti i razlike u podacima o COVID-19 između različitih regija i populacija?
3. Koji su čimbenici mogli pridonijeti promjenama u podacima o COVID-19 tijekom vremena?
4. Postoje li primjetne razlike u demografskoj ili geografskoj distribuciji COVID-19 slučajeva i smrtnih slučajeva tijekom godina?
5. Koji je trenutni (2023.) obim pandemije u Hrvatskoj i svijetu?
6. Koliko su točni i pouzdani bili modeli koji su predviđali broj slučajeva zaraze i smrtnih slučajeva?

Poglavlje 1. Uvod

7. Kako saznanja iz ove analize mogu informirati buduće politike i intervencije vezane za COVID-19?

Odgovaranjem na ova istraživačka pitanja, ovo istraživanje ima za cilj pružiti sveobuhvatnu i komparativnu analizu podataka o COVID-19, osvjetljavajući prirodu i dinamiku pandemije i informirajući buduće istraživačke i političke inicijative.

Poglavlje 2

O pandemiji

U protekle tri godine, pandemija COVID-19 virusa imala je značajan utjecaj na svijet, s milijunima slučajeva zaraze i smrtnih slučajeva prijavljenih diljem svijeta. Odgovor javnosti na pandemiju bio je raznolik, s nekima koji su se pridržavali strogih smjernica, dok su drugi bili skeptični prema ozbiljnosti virusa. Kako pandemija ulazi u svoju četvrtu godinu, važno je razmotriti jesu li odgovori javnosti, vlada i politika na pandemiju u 2020. i 2021., kada su bili najrigorozniji, bili opravdani.

2.1 Općenito o COVID-19

COVID-19, također poznat kao bolest koronavirusa, uzrokuje novi koronavirus SARS-CoV-2. Virus je prvi put identificiran u Wuhanu, Kina u prosincu 2019. godine i brzo se proširio diljem svijeta, što je dovelo do pandemije. Simptomi COVID-19 mogu varirati od blagih do teških i mogu uključivati povišenu temperaturu, kašalj i kratkoću daha. Virus se uglavnom širi respiratornim kapljicama kada zaražena osoba govori, kašlje ili kihne. [2] Osim fizičkih problema koje virus može uzrokovati, COVID-19 također ima značajan ekonomski i društveni utjecaj diljem svijeta. Vlade i medicinsko osoblje provodili su niz mjera za kontrolu širenja virusa, uključujući zatvaranja (*engl.* lockdown), obvezu nošenja maski i kampanje cijepljenja. Razumijevanje prirode virusa i njegovih utjecaja ključno je za razvoj učinkovitih medicinskih strategija i odgovora na trajajuću pandemiju.

2.2 Pozadina i značaj pandemije

Pandemija COVID-19 imala je dubok utjecaj na globalno zdravlje, ekonomiju i društva. Od prijave prvih slučajeva u prosincu 2019. godine, virus se brzo proširio diljem svijeta, inficirajući stotine milijuna ljudi i uzrokujući milijune smrtnih slučajeva. Pandemija je također dovela do karantena, ograničenja putovanja i ekonomskih recesija, kao i značajnih promjena u zdravstvenim sustavima, društvenim ponašanjima i mentalnom zdravlju.

S obzirom na razmjere i složenost pandemije, postoji kritična potreba za razumijevanjem prirode i dinamike COVID-19 podataka, uključujući njihove trendove, uzorke i korelacije. Analizom i tumačenjem podataka o COVID-19, istraživači, donositelji odluka i zdravstveni stručnjaci mogu dobiti uvid u učinkovitost mjera javnog zdravlja, težinu i smrtnost virusa, utjecaj pandemije na ranjive populacije, kao i potencijalna rješenja za ublažavanje širenja i težine virusa.

Stoga ovaj rad ima za cilj pružiti sveobuhvatnu analizu podataka o COVID-19 iz više izvora i vremenskih razdoblja, uspoređujući različite pokazatelje i mjere pandemije. Na taj način, ova studija može doprinijeti nastojanjima da se razumije i kontrolira pandemija, kao i informirati buduće istraživačke i političke inicijative vezane za globalno zdravlje i hitne intervencije.

2.3 Trenutno stanje pandemije (2023.)

Godina 2023. označava kritičnu točku u kontinuiranoj borbi protiv pandemije COVID-19. Svjetska zdravstvena organizacija (*engl.* World Health Organization (WHO)), 5. svibnja 2023. objavila je da više ne smatra COVID-19 javnozdravstvenom opasnošću koja izaziva međunarodnu zabrinutost. [3] Nastavno na to, Vlada Republike Hrvatske proglasila je 11. svibnja 2023. godine kraj epidemije bolesti COVID-19 uzrokovane virusom SARS-CoV-2 u cijeloj Hrvatskoj. Unatoč tome, pojašnjeno je da je COVID-19 još uvijek globalna prijetnja i upozoreno je da bi se još mogle pojaviti nove varijante. [4] Ukratko, COVID-19 se i dalje smatra pandemijom na globalnoj razini, no ne i javnozdravstvenom opasnošću koja izaziva međunarodnu zabrinutost,

dok se u Hrvatskoj COVID-19 više ne smatra epidemijom.

2.4 Modeli predviđanja COVID-19 pandemije

COVID-19 pandemija potaknula je razvoj i upotrebu različitih modela za predviđanje širenja virusa, prognoziranje broja slučajeva i procjenu utjecaja na javno zdravlje.

Predviđanja modela igraju ključnu ulogu u donošenju odluka, oblikovanju strategija javnog zdravstva i usmjeravanju raspoloživih resursa tijekom kriznih vremena. Ona pružaju vrijedne uvide u mogući tijek pandemije, omogućavajući donositeljima odluka i zdravstvenim stručnjacima da predvide i pripreme se za različite scenarije. Međutim, važno je retrospektivno procijeniti točnost i pouzdanost ovih predviđanja kako bi se evaluirala njihova učinkovitost i unaprijedili budući napori u modeliranju.

Usporedbom različitih prediktivnih modela s podacima iz stvarnog svijeta, može se analizirati razina točnosti i identificirati obrasce ili trendove u njihovim sposobnostima predviđanja. Ova analiza pomaže razumjeti snage i ograničenja različitih pristupa modeliranju, ali također pruža vrijedne lekcije za unapređenje budućih napora u modeliranju i procesa donošenja odluka.

Važno je pristupiti analizi realno, prepoznajući izazove povezane s modeliranjem složenog i brzo razvijajućeg virusa. Cilj nije kritizirati ili umanjiti napore modelara, već stjecanje boljeg razumijevanja snaga, slabosti i ukupnih performansi različitih pristupa modeliranju tijekom globalne zdravstvene krize.

Analiza modela predviđanja prikazana je u Poglavlju 6.

Poglavlje 3

Metodologija

Ovo poglavlje pruža detaljan opis istraživačkog postupka, izvora podataka, varijabli i mjera, metoda analize podataka i ograničenja i pretpostavki studije. Objasnit će kako će biti odgovoreno na istraživačka pitanja i kako će podaci biti prikupljeni, obrađeni i analizirani. Prezentacijom jasne i sistematične metodologije, poglavlje ima za cilj osigurati valjanost, pouzdanost i generaliziranost nalaza istraživanja.

3.1 Izvori podataka

Analiza COVID-19 podataka oslanja se na više izvora, uključujući znanstvene baze podataka i javne repozitorije COVID-19 podataka. Izvori podataka odabrani su na temelju pouzdanosti, opsega i dostupnosti, te uključuju i nacionalne i međunarodne skupove podataka.

Kako bi se osigurala točnost i dosljednost podataka, u prikupljanju i obradi podataka provedeni su sljedeći koraci:

- **Prikupljanje podataka:** Podaci su prikupljeni sa službene web stranice Our World In Data (OWID) [5], koristeći ručna preuzimanja. Podaci su pohranjeni u strukturirani format, CSV datoteke, te uključuju informacije o ukupnim i novim slučajevima zaraze, ukupnim i novim slučajevima smrti, broju cijepljenih osoba, indeksima strogosti odgovora vlade, gustoće naseljenosti stanovništva,

Poglavlje 3. Metodologija

srednje starosti stanovništva, broja stanovnika starijih od 70 godina, udjela stanovništva koje živi u ekstremnom siromaštvu, BDP-a po glavi stanovnika te drugim relevantnim pokazateljima. Podaci vezani za modele predviđanja prikupljeni su sa stranica European COVID-19 Forecast Hub [6] i European Centre for Disease Prevention and Control (ECDC) [7]. European COVID-19 Forecast Hub [6] je inicijativa koja kreira predviđanja novih slučajeva zaraze i smrtnih slučajeva povezanih s COVID-19 za sljedeća četiri tjedna u zemljama diljem Europe i Ujedinjenom Kraljevstvu. U svrhu poboljšanja svojih modela, odlučili su na tjednoj bazi objavljivati evaluacije, odnosno metrike koje pokazuju koliko je predikcija pojedinog modela bila udaljena od stvarne vrijednosti. Stvarne vrijednosti podataka preuzimaju sa stranice ECDC [7].

- **Čišćenje i validacija podataka:** Prikupljeni podaci očišćeni su i validirani kako bi se uklonile sve pogreške, duplikati ili neusklađenosti. Nedostajući ili nepotpuni podaci popunjeni su uz pomoć interpolacije. Stupci tablice koji su bili nepotrebni, eliminirani su. Podaci za analizu modela uključuju tjedne evaluacije za različite modele i varijable koje su predviđane, različite vremenske okvire na temelju kojih je predikcija izvršavana, različite lokacije za koje je sprovedena, kao i različite vremenske horizonte. Tjedne evaluacije su preuzete, obrađene, analizirane, iz njih su izvučeni podaci od interesa te je na temelju njih napravljena prosječna analiza točnosti modela za predviđanja COVID-19 u Europi, kao i prosječne analize svakog modela zasebno.
- **Analiza podataka:** Očišćeni i validirani podaci analizirani su korištenjem statističkih i vizualnih metoda, kao što su deskriptivna statistika, korelacijska analiza, regresijska analiza i vizualizacija podataka. Analiza će imati za cilj identificirati trendove, obrasce i odnose u COVID-19 podacima, te ih usporediti u različitim regijama, populacijama i vremenskim razdobljima.

3.2 Metode analize podataka

3.2.1 Deskriptivna analiza

Deskriptivna analiza predstavlja skup metoda za izračunavanje, prikazivanje i opisivanje podataka kako bi se omogućilo jasno razumijevanje njihovih osnovnih karakteristika. Usredotočena je na grupiranje, sažimanje i predstavljanje podataka na smislen način, koristeći statističke mjere i grafičke prikaze. Deskriptivna analiza ima za cilj otkriti obrasce, trendove i distribucije unutar podataka, omogućujući zainteresiranima da steknu uvid u centralne tendencije, varijabilnosti i odnose među varijablama. Ova analiza pruža temelj za razumijevanje podataka prije nastavka s naprednijim statističkim analizama.[8]

3.2.2 Korelacijska analiza

Korelacijska analiza je metoda koja se koristi za izračunavanje povezanosti dviju ili više varijabli. Često se koristi u istraživanjima kako bi se utvrdilo u kojoj su mjeri promjene vrijednosti jedne varijable povezane s promjenama vrijednosti druge varijable.

Rezultat korelacijske analize je koeficijent korelacije, koji može varirati od -1 do 1. Koeficijent korelacije -1 označava savršenu negativnu korelaciju, što znači da kada jedna varijabla raste, druga varijabla pada. Koeficijent korelacije 0 označava da nema korelacije između varijabli, dok koeficijent korelacije 1 označava savršenu pozitivnu korelaciju, što znači da kada jedna varijabla raste, druga varijabla također raste. [9]

Korelacijska analiza može se koristiti za istraživanje odnosa između različitih varijabli u različitim područjima, uključujući financijske, ekonomske, zdravstvene i društvene znanosti. U kontekstu COVID-19, korelacijska analiza može se koristiti za istraživanje odnosa između slučaja ili smrtnosti od COVID-19 i drugih varijabli kao što su indeks strogoće odgovora vlade, stopa cijepljenja ili uvjeti ekstremnog siromaštva.

Postoje različite tehnike korelacijske analize, uključujući Pearsonov koeficijent korelacije te Spearmanov koeficijent korelacije ranga. Pearsonov koeficijent korelacije

Poglavlje 3. Metodologija

se često koristi za kontinuirane varijable s normalnom distribucijom. Spearmanov koeficijent korelacije ranga, neparametarski ekvivalent Pearsonovom koeficijentu korelacije, računa se ako je jedna ili više sljedećih stavki točna[10]:

- barem jedna od varijabli, x ili y , mjeri se na ordinalnoj ljestvici
- niti x niti y nisu normalno raspodijeljeni
- veličina uzorka je mala
- traži se mjera povezanosti između dviju varijabli kada je njihov odnos nelinearan.

Podaci korišteni u izračunavanjima korelacije zadovoljavat će 2. i 4. stavku, što znači da će korišteni koeficijent korelacije biti Spearmanov.

Rezultati će biti tumačeni sukladno tablici 3.1, kreiranoj prema navodima iz [9].

Tablica 3.1 Tumačenje rezultata

| Koeficijent | Tumačenje |
|-------------|----------------------------------|
| 0.00 - 0.25 | nema povezanosti |
| 0.26 - 0.50 | slaba povezanost |
| 0.51 - 0.75 | umjerena do dobra povezanost |
| 0.76 - 0.99 | vrlo dobra do izvrsna povezanost |
| 1 | matematička povezanost |

Važno je napomenuti da korelacija ne implicira uzročnost, što znači da jaka korelacija između dvije varijable ne znači nužno da jedna varijabla uzrokuje drugu. Korelacijsku analizu treba tumačiti oprezno, a daljnja istraživanja trebaju se provoditi kako bi se utvrdili uzročni odnosi između varijabli.

3.2.3 Analiza trenda korištenjem regresije

Za ispitivanje trenda novih COVID-19 slučajeva i smrtnosti tijekom vremena, korišteni su linearna regresija i kvadratni trend. Broj novih slučajeva zaraze i novih slučajeva smrti prikazan je na grafu vremena za svaku godinu, a za svaki skup podataka prilagođena je linearna regresijska linija. Nagib linije predstavlja stopu promjene

Poglavlje 3. Metodologija

broja novih slučajeva zaraze ili smrti po jedinici vremena.

Kvadratni trend vizualno je predstavljen za COVID-19 podatke za Europu. On opisuje podatke polinomnom funkcijom 2. stupnja, poznatom kao kvadratna funkcija. Kvadratni trend je zakrivljenog oblika, tvori oblik slova U ili obrnutog slova U. Smjer i strmina trenda ovise o vrijednostima koeficijenata. Zbog karakteristike COVID-19 podataka da nisu linearni, pretpostavka je da će im se kvadratni trend bolje prilagođavati.

Da bi se procijenila statistička značajnost trendova, izračunate su p-vrijednosti za svaku regresijsku analizu. P-vrijednost manja od 0.05 ukazuje da je trend statistički značajan, što znači da postoji niska vjerojatnost da se trend dogodio slučajno. Također je izračunat koeficijent determinacije (R^2), koji mjeri udio varijance u broju novih slučajeva ili smrti u linearnom regresijskom modelu.

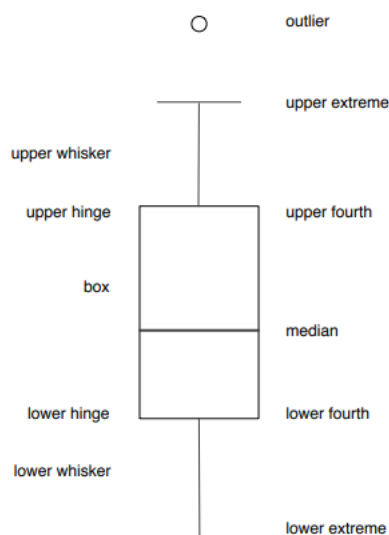
Koeficijent determinacije, označen kao R-kvadrat (R^2), je statistička mjera koja se koristi za procjenu koliko dobro se linearni regresijski model prilagođava podacima. Ovaj koeficijent pruža informaciju o tome koliki postotak varijabilnosti ovisne varijable može biti objašnjen ili predviđen pomoću nezavisnih varijabli u modelu. Drugim riječima, R-kvadrat je specifičan pokazatelj reprezentativnosti regresije. Vrijednost R-kvadrata kreće se od 0 do 1, gdje 0 označava da model ne objašnjava nikakvu varijabilnost u ovisnoj varijabli, a 1 označava da model objašnjava svu varijabilnost u ovisnoj varijabli. Što je veća vrijednost R-kvadrata, to model bolje odgovara podacima. Međutim, visoka vrijednost R-kvadrata ne znači nužno da je model dobar prediktor ovisne varijable, već treba uzeti u obzir i druge mjere kao što su p-vrijednost i analiza ostatka.

Analiza trenda omogućila je ispitivanje rasta ili pada broja novih slučajeva i smrti tijekom vremena, te kojom brzinom. Također je pomogla u identificiranju potencijalnih točaka preokreta u trendu, što može ukazivati na promjene u širenju ili utjecaju virusa.

3.2.4 Vizualizacija podataka

Vizualizacija podataka ima važnu ulogu u ovom diplomskom radu jer pruža moćno sredstvo za prikazivanje i razumijevanje kompleksnih i mnogobrojnih informacija povezanih s COVID-19 pandemijom. Kroz različite vizualne prikaze poput stupčastih dijagrama, kutijastih dijagrama, linijskih dijagrama te tablica, obrasci i trendovi u podacima su vizualno istaknuti, olakšavajući tumačenje i komunikaciju rezultata. Vizualizacije su korištene kako bi se prikazala raspodjela COVID-19 slučajeva i smrtnih slučajeva u različitim regijama, ilustrirale promjene tijekom vremena i usporedio utjecaj pandemije na kontinentima.

Budući da su kutijasti dijagrami možda manje poznati široj javnosti, bit će kratko predstavljene. Box-plot, poznat i kao kutijasti dijagram, je jednostavan graf koji prikazuje ključne statističke podatke. Izgled box-plota prikazan je i objašnjen na slici 3.1. Sastoji se od pravokutnika koji predstavlja donji i gornji kvartil, crte u središnjem dijelu koja označava median, te T-linija koje se nazivaju "whisker". Duljine whisker linija određene su parametrom, obično $1.5 \times \text{IQR}$ (interkvartilni raspon). Sve vrijednosti izvan tih granica smatraju se outlierima i prikazuju se kružićima. [1]



Slika 3.1 Vizualno objašnjenje kutijastog dijagrama. Oznake s lijeve strane daju nazive za grafičke elemente, oznake s desne strane daju odgovarajuću sažetu statistiku.[1]

3.3 Korišteni softver

Prikupljeni podaci analizirani su, obrađeni i prikazani pomoću Python programskog jezika (verzija 3.8.10) koji je korišten unutar Spyder IDE (verzija 5.4.3), besplatnog i open source okruženja za Python.

Za učinkovitu manipulaciju, obradu, čišćenje, analizu i vizualizaciju podataka u Pythonu, korištene su sljedeće knjižnice:

- **pandas** - Korištena za učitavanje i pohranu COVID-19 podataka, kreiranje vremenskih nizova, grupiranje podataka, te manipulaciju i analizu podataka općenito.
- **matplotlib.pyplot** - Korištena za kreiranje različitih vrsta grafova i vizualizacija za prikaz COVID-19 podataka.
- **numpy** - Korištena za manipulaciju s nizovima i matricama pri analizi COVID-

19 podataka.

- **scipy.stats (linregress, spearmanr)** - Korištena za provođenje statističkih izračuna poput linearne regresije i Spearmanovog koeficijenta korelacije nad COVID-19 podacima.
- **sklearn.linear_model (LinearRegression)** - Korištena za implementaciju modela linearne regresije radi analize i predviđanja trendova u COVID-19 podacima.
- **matplotlib.ticker** - Korištena za prilagodbu i formatiranje oznaka na grafikonima stvorenim pomoću matplotlib.pyplot za bolji prikaz COVID-19 podataka.
- **seaborn** - Korištena za kreiranje kutijastih dijagrama.
- **glob** - Korištena za dobivanje putanja i naziva datoteka u direktoriju radi obrade više datoteka COVID-19 podataka.
- **statistics** - Korištena za provođenje osnovnih statističkih izračuna nad COVID-19 podacima, poput srednje vrijednosti, medijana i standardne devijacije.

3.4 Varijable i mjere

3.4.1 COVID-19

Za proučavanje podataka o COVID-19 virusu, važno je definirati varijable i mjere koje će se koristiti. Varijable od interesa u ovom istraživanju uglavnom uključuju potvrđene slučajeve zaraze, smrti te cijepljenja, očitavane u različitim regijama, vremenskim razdobljima i populacijama.

Potvrđeni slučajevi zaraze odnose se na broj osoba koje su dijagnosticirane s COVID-19 na temelju laboratorijskih testova ili kliničke dijagnoze, dok se potvrđeni slučajevi smrti odnose na broj osoba koje su umrle od COVID-19. Cijepljenja se odnose na broj podijeljenih doza cjepiva protiv COVID-19.

Ove varijable će se mjeriti u različitim regijama, populacijama i vremenskim razdobljima. Mjere koje će se koristiti za proučavanje ovih varijabli uključuju deskriptivnu statistiku, poput srednje vrijednosti, medijana i standardne devijacije, re-

gresijsku i korelacijsku statistiku, poput p-vrijednosti, koeficijenta determinacije i koeficijenta korelacije, kao i vizualizacije, poput linijskih grafikona, stupčastih i kutijastih dijagrama. Mjere će pomoći u pružanju uvida u širenje i utjecaj COVID-19, te omogućiti usporedbe između različitih regija, populacija i vremenskih razdoblja.

3.4.2 Modeli predviđanja

U suvremenoj epidemiologiji zaraznih bolesti sve više se smatra da bi prognoze epidemija trebale biti probabilističke, što znači da ne bi trebale samo predviđati jedan mogući ishod, već i izračunati vlastitu nesigurnost. Takve prognoze mogu se pohraniti u obliku kvantila ili intervala. Na primjer, umjesto da se samo navede jedna predviđena vrijednost, mogu se navesti različiti kvantili ili intervali koji obuhvaćaju moguće vrijednosti. Takav pristup korišten je i u [6], gdje su prognoze pohranjivane u obliku prediktivnih intervala. Timovi koji sudjeluju trebaju izvještavati o prediktivnom medijanu i centralnim prediktivnim intervalima s nominalnim razinama 1%, 2.5%, 5%, 10%, 15%, ..., 95%, 97.5%, 99%.

Ocjenjivanje takvih intervala provodi se kroz tri koraka:

1. Širina intervala, što opisuje preciznost prognoze.
2. Kazna za promatranja koja padaju ispod donje granice intervala. Kazna je proporcionalna udaljenosti između promatranog podatka i donje granice intervala.
3. Kazna za promatranja koja padaju iznad gornje granice intervala.

Te tri komponente čine skor intervala predikcije. Da bi se pružile detaljnije informacije o prediktivnoj distribuciji, uobičajeno je prijaviti ne samo jedan, već nekoliko intervala predikcije, zajedno s prediktivnim medijanom. Tada se vrednovanje prognoza obavlja pomoću ponderiranog intervalnog skora (*engl.* WIS). On se računa na način da se skor svakog intervala predikcije pomnoži sa svojom težinom (*engl.* weight), nakon čega se to sve zbroji. WIS je koristan kod određivanja oštine i kalibracije prognoza. [11]

Metodologija analize modela predviđanja uključuje usporedbu predviđanja razli-

Poglavlje 3. Metodologija

čitih modela sa stvarnim podacima kako bi se procijenila njihova točnost i pouzdanost. Metrike točnosti korištene za analizu bit će:

- srednja apsolutna pogreška (*engl.* MAE), tjedni prosjek apsolutnih pogrešaka (*engl.* Absolute Error (AE)) koje mjere razliku između prediktivnih medijana i stvarne vrijednosti veličine
- prosječni ponderirani intervalni skor (WIS), tjedni prosjek ponderiranog intervalnog skora, koji je sveobuhvatna mjera točnosti predviđanja, uzimajući u obzir ponderirani doprinos različitih prediktivnih intervala
- prosječan udio stvarnih vrijednosti koje se nalaze unutar 50%-tnih i 95%-tnih intervala predikcije - vrijednosti pokrivenosti veće od ovih nominalnih vrijednosti ukazuju na to da su prognoze nedovoljno pouzdane, odnosno intervali predviđanja imaju tendenciju da budu preširoki, dok vrijednosti pokrivenosti manje od ovih nominalnih vrijednosti ukazuju na to da su prognoze previše pouzdane, odnosno intervali predviđanja imaju tendenciju da budu preuski
- prosječna pristranost, koja mjeri sustavno odstupanje predviđanja od stvarnih vrijednosti - pozitivna vrijednost ukazuje na precijenjenost, dok negativna vrijednost ukazuje na podcijenjenost

Navedene metrike bit će analizirane za različite vremenske horizonte, modele i ciljne varijable kako bi se utvrdila njihova uspješnost i performanse.

3.5 Ograničenja i pretpostavke

Neka od mogućih ograničenja i pretpostavki, koje bi se trebale uzeti u obzir prilikom tumačenja rezultata, mogu uključivati:

- Točnost i pouzdanost izvora podataka: Iako su izvori podataka pažljivo odabrani na temelju njihove pouzdanosti, opsega i dostupnosti, još uvijek mogu postojati neke pogreške, propusti ili neusklađenosti u podacima. Ti problemi mogu utjecati na valjanost i generaliziranost rezultata.
- Reprezentativnost uzoraka: Podaci o COVID-19-u možda nisu u potpunosti

Poglavlje 3. Metodologija

reprezentativni za cijelu populaciju, jer neke skupine mogu biti nedovoljno ili prekomjerno zastupljene u podacima. Na primjer, podaci možda neće u potpunosti obuhvatiti iskustva marginaliziranih skupina, poput siromašnih ili ruralnih populacija. To može ograničiti generaliziranost rezultata.

- Potencijalni zbunjujući čimbenici: Mogu postojati i drugi čimbenici koji utječu na ishode COVID-19, poput kvalitete zdravstvenih sustava, dostupnosti resursa, razine javne svijesti i političkog i društvenog konteksta. Ti čimbenici mogu zakomplicirati odnos između COVID-19 varijabli i objašnjavajućih čimbenika te mogu utjecati na valjanost i pouzdanost nalaza.

Poznavanje ovih ograničenja i pretpostavki omogućuje istraživanju da pruži nijansiraniju i transparentniju sliku o COVID-19 podacima i njihovim implikacijama.

Poglavlje 4

Usporedba COVID-19 podataka kroz 2020., 2021. i 2022.

4.1 Rezultati deskriptivne analize

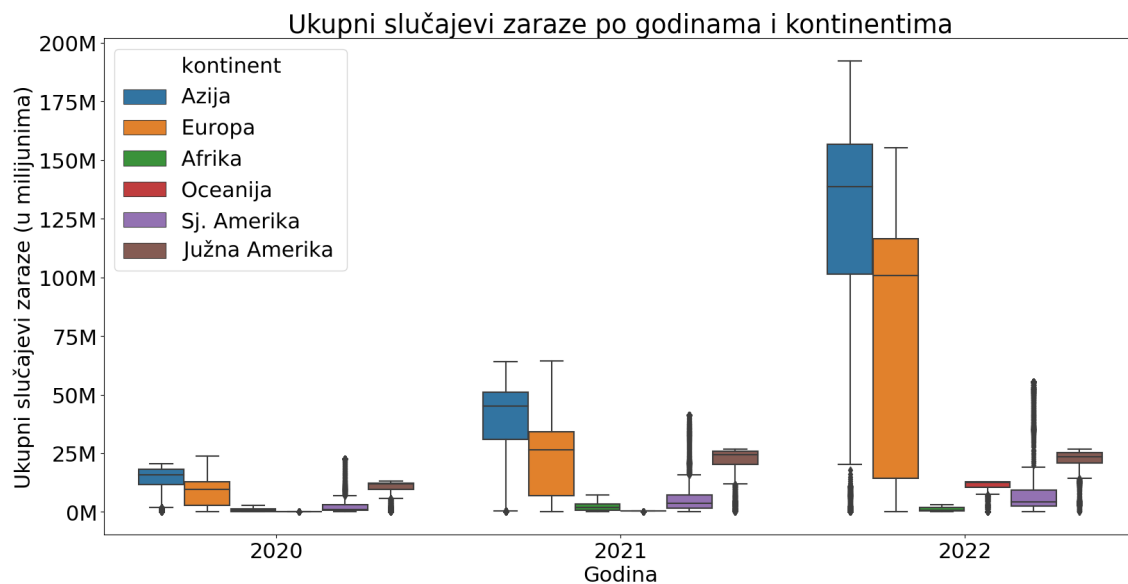
Deskriptivna analiza provedena je na skupu podataka koji sadrži podatke o COVID-19 za godine 2020., 2021. i 2022. Skup podataka je filtriran tako da uključuje podatke sa šest kontinenata: Europe, Azije, Afrike, Južne Amerike, Sjeverne Amerike i Oceanije. Ključne statistike, uključujući minimum, 0.25-kvartil, medijan, 0.75-kvartil i maksimum, izračunate su za broj ukupnih slučajeva zaraze i ukupnih smrtnih slučajeva unutar svakog kontinenta i godine. Rezultati su prikazani kutijastim dijagramima.

Na slici 4.1 prikazani su kutijasti dijagrami novih slučajeva zaraze po kontinentima, dok su na slici 4.2 prikazani kutijasti dijagrami novih smrtnih slučajeva po kontinentima.

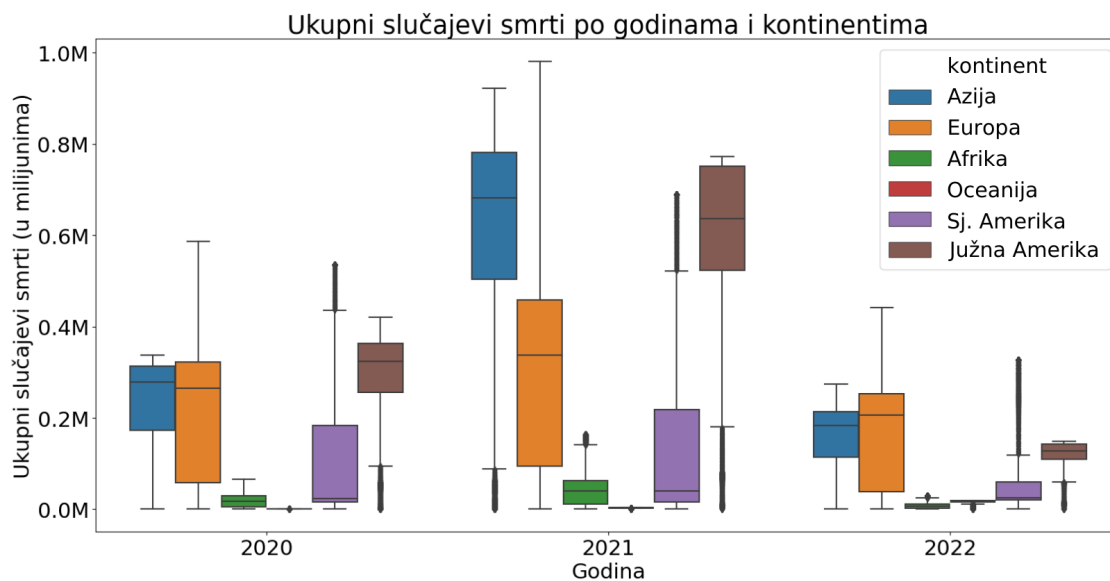
Na temelju provedene analize, iz kutijastih dijagrama mogu se izvući sljedeći opći zaključci:

Postoji značajna disperzija u broju slučajeva zaraze i smrtnih slučajeva na različitim kontinentima. Svaki kontinent ima svoje jedinstvene obrasce i trendove, što sugerira da je utjecaj pandemije geografski varirao.

Poglavlje 4. Usporedba COVID-19 podataka kroz 2020., 2021. i 2022.



Slika 4.1 Kutijasti dijagrami novih slučajeva zaraze po kontinentima



Slika 4.2 Kutijasti dijagrami novih smrtnih slučajeva po kontinentima

Poglavlje 4. Usporedba COVID-19 podataka kroz 2020., 2021. i 2022.

Općenito, došlo je do povećanja broja slučajeva zaraze od 2020. do 2021., s nekim varijacijama u 2022. Broj slučajeva smrti je rastao od 2020. do 2021., da bi kod većine kontinenata pao u 2022. Ove milijunske brojke ukazuju na to da je pandemija imala značajan utjecaj na globalno zdravlje tijekom ovih godina, pri čemu su neke regije doživjele više teških proboja virusa od drugih regija.

Postoje značajne razlike u magnitudi utjecaja pandemije među kontinentima. Azija, Europa i Južna Amerika općenito su prijavile veći broj slučajeva i smrtnih slučajeva u usporedbi s Afrikom, Oceanijom i Sjevernom Amerikom. Ovo naglašava regionalne razlike u zdravstvenim sustavima, mogućnostima testiranja i preventivnim mjerama.

Analiza pokazuje da je značajan dio podataka za slučajeve zaraze i smrtnu slučajevu bio jako malen u usporedbi s ostalim kontinentima, posebno u Africi i Oceaniji. Ovo upućuje na to da je bilo razdoblja bez prijavljenih novih slučajeva ili smrtnih slučajeva, možda zbog učinkovitih mjera suzbijanja ili ograničenih kapaciteta testiranja.

Prisutnost gornjih outliera (oni koji su iznad najviše vrijednosti) u Sjevernoj Americi, ukazuje na pojavu posebno ozbiljnih epidemija ili lokaliziranih skokova u broju slučajeva i smrti. Ti ekstremi mogu se pripisati čimbenicima kao što su gustoća naseljenosti, javnozdravstvene mjere ili pojava novih varijanti. Također, ne treba isključiti da se prisutnost određenih outliera može objasniti i greškama u podacima. Što se tiče podataka za Sjevernu Ameriku, u njih su uključeni i podaci za Srednju Ameriku, što joj, moguće, spušta statistiku za određeni iznos. Da se u obzir uzimala samo Sjeverna Amerika, bez Srednje, potencijalno bi brojke bile više.

Deskriptivna analiza pruža općeniti pregled podataka, ali istovremeno naglašava potrebu za daljnjim istraživanjem i analizom. Faktori poput kapaciteta testiranja, praksi izvještavanja i demografskih karakteristika mogu značajno utjecati na prijavljene brojke, a dublje istraživanje tih faktora doprinijelo bi sveobuhvatnijem razumijevanju utjecaja pandemije. Ovi opći zaključci pružaju uvid u globalnu i regionalnu dinamiku COVID-19 tijekom navedenih godina. Također, naglašavaju važnost uzimanja u obzir lokalnog konteksta pri tumačenju podataka.

4.2 Rezultati korelacijske analize

Korelacijska analiza provedena je računanjem Spearmanovog koeficijenta korelacije ranga.

4.2.1 Korelacija procijepljenosti i smrtnosti

Analiziranjem korelacije između broja cijepljenja i broja novih smrtnih slučajeva htjelo se utvrditi jesu li procijepljenost i smrtnost povezane. Rezultati korelacijske analize za ove dvije varijable korištenjem Spearmanovog koeficijenta korelacije ranga vidljivi su u tablici 4.1.

Tablica 4.1 Rezultati korelacijske analize između procijepljenosti i smrtnosti korištenjem Spearmanovog koeficijenta korelacije ranga

| Regija | Koeficijent | P-vrijednost |
|------------------|-------------|--------------|
| Afrika | -0.02 | 0.500 |
| Azija | 0.48 | 0.000 |
| Australija | 0.41 | 0.000 |
| Europa | 0.45 | 0.000 |
| Južna Amerika | 0.13 | 0.000 |
| Sjeverna Amerika | 0.13 | 0.000 |
| Svijet | 0.39 | 0.000 |

Varijable korištene u ovoj korelacijskoj analizi su broj novih smrtnih slučajeva i broj novih cijepljenja. Koeficijenti korelacije se kreću od -0,02 do 0,48, pri čemu je najniža vrijednost u Africi, a najviša u Aziji. Koeficijenti korelacije u svim regijama osim Afrike su pozitivni. Azija, Australija i Europa pokazuju slabu povezanost između navedenih varijabli, dok za Amerike nema značajne povezanosti. Svijet globalno pokazuje slabu, ali postojeću povezanost između dviju varijabli.

Gledajući p-vrijednosti u rezultatima, može se vidjeti da sve regije osim Afrike imaju p-vrijednost manju od 0,05, što znači da je koeficijent korelacije značajan i

Poglavlje 4. Usporedba COVID-19 podataka kroz 2020., 2021. i 2022.

da se smije tumačiti. Budući da je za Afriku p-vrijednost > 0.05 , njen koeficijent korelacije nije značajan i ne smije se tumačiti, bez obzira na njegovu vrijednost.

Sažeto, može se zaključiti da ne postoji korelacija između novih cijepljenja i novih slučajeva smrti u Amerikama, dok postoji pozitivna korelacija između ovih varijabli u svim ostalim regijama osim Afrike, gdje korelacija nije statistički značajna te se korelacija ne smije tumačiti.

4.2.2 Korelacija gustoće naseljenosti i smrtnosti

Analizom korelacije između gustoće naseljenosti stanovništva i broja smrtnih slučajeva htjelo se utvrditi jesu li gustoća naseljenosti i smrtnost povezane, budući da visoka gustoća naseljenosti može potencijalno dovesti do većih stopa prijenosa zaraze, odnosno veće smrtnosti. Varijable korištene u ovoj korelacijskoj analizi su gustoća naseljenosti i broj novih smrtnih slučajeva. Rezultati analize vidljivi su u tablici 4.2.

Tablica 4.2 Rezultati korelacijske analize gustoće naseljenosti stanovništva i smrtnosti korištenjem Spearmanovog koeficijenta korelacije ranga

| Regija | Koeficijent | P-vrijednost |
|------------------|-------------|--------------|
| Afrika | 0.03 | 0.000 |
| Azija | 0.17 | 0.000 |
| Australija | -0.06 | 0.000 |
| Europa | 0.15 | 0.000 |
| Južna Amerika | 0.47 | 0.000 |
| Sjeverna Amerika | -0.02 | 0.000 |
| Svijet | 0.11 | 0.000 |

Budući da sve regije imaju p-vrijednost < 0.05 , koeficijenti korelacije za sve regije su statistički značajni i mogu se tumačiti.

Na svim kontinentima (osim Južne Amerike) i svijetu općenito nema povezanosti između gustoće stanovništva i novih smrtnih slučajeva.

U Južnoj Americi postoji slaba do umjerena pozitivna korelacija između gus-

Poglavlje 4. Usporedba COVID-19 podataka kroz 2020., 2021. i 2022.

toće stanovništva i novih smrtnih slučajeva. To ukazuje na to da je veća gustoća stanovništva povezana s većim brojem novih smrtnih slučajeva na ovom kontinentu.

Zaključno, veća gustoća stanovništva u većini regija nije povezana s većim brojem novih smrtnih slučajeva.

4.2.3 Korelacija srednje starosti stanovništva i smrtnosti

Analizom korelacije između srednje starosti stanovništva i broja smrtnih slučajeva htjelo se utvrditi jesu li srednja starost i smrtnost povezane, budući da je logično zaključiti da što je stanovništvo starije, to su stope smrtnosti veće. Varijable korištene u ovoj korelacijskoj analizi su srednja starost stanovništva (engl. *median age*, dobiveno projekcijom UN-a za 2020.) i broj novih smrtnih slučajeva. Rezultati analize vidljivi su u tablici 4.3.

Tablica 4.3 Rezultati korelacijske analize srednje starosti stanovništva i smrtnosti korištenjem Spearmanovog koeficijenta korelacije ranga

| Regija | Koeficijent | P-vrijednost |
|------------------|-------------|--------------|
| Afrika | 0.16 | 0.000 |
| Azija | 0.03 | 0.000 |
| Australija | 0.31 | 0.000 |
| Europa | 0.57 | 0.000 |
| Južna Amerika | 0.41 | 0.000 |
| Sjeverna Amerika | 0.27 | 0.000 |
| Svijet | 0.44 | 0.000 |

Na temelju dobivenih koeficijenata korelacije, čini se da postoji pozitivna korelacija između srednje starosti stanovništva i broja novih smrtnih slučajeva u toj regiji. Sve p-vrijednosti su manje od 0.05, što znači da se koeficijenti korelacije mogu slobodno tumačiti.

U Africi i Aziji nema povezanosti između dviju navedeni varijabli. U Australiji, Sjevernoj Americi, Južnoj Americi i svijetu općenito korelacija je slaba. Najveći

Poglavlje 4. Usporedba COVID-19 podataka kroz 2020., 2021. i 2022.

koeficijent korelacije zabilježen je u Europi, gdje je korelacija između srednje starosti stanovništva i smrtnosti umjerena do dobra.

Važno je napomenuti da korelacija ne znači nužno uzročnost, a drugi čimbenici također mogu utjecati na broj novih smrtnih slučajeva u svakoj zemlji. Međutim, ovi rezultati sugeriraju da bi mogao postojati uzročno-posljedični odnos između srednje starosti stanovništva i smrtnosti.

4.2.4 Korelacija stanovništva starijeg od 70 godina i smrtnosti

Analizom korelacije između broja stanovnika starijih od 70 godina i broja smrtnih slučajeva htjelo se utvrditi jesu li te dvije varijable povezane, budući da u toj dobi organizam i imunološki sustav počinju slabiti, što može dovesti do lošije obrane od virusa, odnosno većih stopa smrtnosti. Varijable korištene u ovoj korelacijskoj analizi su broj stanovništva starijeg od 70 godina i broj novih smrtnih slučajeva. Rezultati analize vidljivi su u tablici 4.4.

Tablica 4.4 Rezultati korelacijske analize broja stanovništva starijeg od 70 godina i smrtnosti korištenjem Spearmanovog koeficijenta korelacije ranga

| Regija | Koeficijent | P-vrijednost |
|------------------|-------------|--------------|
| Afrika | 0.13 | 0.000 |
| Azija | 0.17 | 0.000 |
| Australija | 0.31 | 0.000 |
| Europa | 0.55 | 0.000 |
| Južna Amerika | 0.45 | 0.000 |
| Sjeverna Amerika | 0.36 | 0.000 |
| Svijet | 0.44 | 0.000 |

Na temelju dobivenih koeficijenata korelacije, čini se da postoji pozitivna korelacija između broja stanovništva starijeg od 70 godina u zemlji i broja novih smrtnih slučajeva u toj regiji.

Koeficijenti korelacije su statistički značajni za sve kontinente, što ukazuje da ova veza vjerojatno nije posljedica slučajnosti. Europa ima najjaču korelaciju koja

ukazuje na umjerenu do dobru povezanost dviju varijabli, slijede je Južna Amerika, svijet općenito, Sjeverna Amerika i Australija sa slabom povezanošću varijabli, te Azija i Afrika za koje varijable nisu povezane.

Važno je napomenuti da korelacija ne znači uzročnost, a drugi čimbenici također mogu utjecati na broj novih smrtnih slučajeva u svakoj zemlji. Međutim, ovi rezultati, pogotovo kada ih se promatra zajedno s rezultatima iz prošlog odjeljka (*Korelacija srednje starosti stanovništva i smrtnosti*) ukazuju da postotak stanovništva starijeg od 70 godina može biti relevantan čimbenik za razmatranje prilikom procjene utjecaja pandemije COVID-19 na različite regije.

4.2.5 Korelacija udjela stanovništva koje živi u ekstremnom siromaštvu i smrtnosti

Analizom korelacije između udjela stanovništva koje živi u ekstremnom siromaštvu i broja smrtnih slučajeva htjelo se utvrditi jesu li te dvije varijable povezane, budući da zemlje s višom razinom siromaštva mogu imati veći broj smrtnih slučajeva od COVID-19 zbog čimbenika kao što su neadekvatan pristup zdravstvenoj skrbi, loši životni uvjeti i veća prevalencija popratnih bolesti. Varijable korištene u ovoj korelacijskoj analizi su udio stanovništva koje živi u ekstremnom siromaštvu i broj novih smrtnih slučajeva. Rezultati analize vidljivi su u tablici 4.5.

Afrika, Azija, Australija i svijet općenito nemaju značajnih povezanosti između udjela stanovništva koje živi u ekstremnom siromaštvu i broja novih smrtnih slučajeva u toj regiji.

Europa, Sjeverna Amerika i Južna Amerika očituju slabu povezanost navedenih varijabli. Koeficijenti korelacije su statistički značajni za sve regije.

Za dio regija, povezanost između udjela stanovništva koje živi u ekstremnom siromaštvu i broja novih smrtnih slučajeva postoji.

Tablica 4.5 Rezultati korelacijske analize udjela stanovništva koje živi u ekstremnom siromaštvu i smrtnosti korištenjem Spearmanovog koeficijenta korelacije ranga

| Regija | Koeficijent | P-vrijednost |
|------------------|-------------|--------------|
| Afrika | -0.04 | 0.000 |
| Azija | 0.16 | 0.000 |
| Australija | 0.11 | 0.000 |
| Europa | 0.31 | 0.000 |
| Južna Amerika | 0.49 | 0.000 |
| Sjeverna Amerika | 0.49 | 0.000 |
| Svijet | 0.17 | 0.000 |

4.2.6 Korelacija BDP-a po glavi stanovnika i smrtnosti

Analizom korelacije između BDP-a po stanovniku i broja smrtnih slučajeva htjelo se utvrditi jesu li te dvije varijable povezane, kako bi se moglo odgovoriti na pitanje postoji li veza između ekonomskog prosperiteta zemlje i njezine sposobnosti da odgovori na pandemiju. Konkretno, može se usporediti jesu li zemlje s višim BDP-om po glavi stanovnika bolje sposobne provoditi učinkovite javnozdravstvene intervencije i osigurati medicinske resurse svom stanovništvu, što rezultira nižim stopama smrti od COVID-19.

Varijable korištene u ovoj korelacijskoj analizi su BDP po glavi stanovnika i broj novih smrtnih slučajeva. Rezultati analize vidljivi su u tablici 4.6.

Na temelju dobivenih koeficijenata korelacije, čini se da postoji pozitivna korelacija između BDP-a po glavi stanovnika i broja novih smrtnih slučajeva.

Koeficijenti korelacije su statistički značajni za sve kontinente. Europa, Južna Amerika, Sjeverna Amerika i svijet općenito imaju slabu povezanost navedenih dviju varijabli. Azija, Australija i Afrika nisu detektirale povezanost između dviju varijabli.

Ovi rezultati sugeriraju da postoji pozitivna povezanost između dvije varijable u nekim regijama, što znači da kako BDP po glavi stanovnika raste, smrtnost također ima tendenciju povećanja.

Tablica 4.6 Rezultati korelacijske analize BDP-a po stanovniku i smrtnosti korištenjem Spearmanovog koeficijenta korelacije ranga

| Regija | Koeficijent | P-vrijednost |
|------------------|-------------|--------------|
| Afrika | 0.22 | 0.000 |
| Azija | 0.06 | 0.000 |
| Australija | 0.18 | 0.000 |
| Europa | 0.36 | 0.000 |
| Južna Amerika | 0.34 | 0.000 |
| Sjeverna Amerika | 0.30 | 0.000 |
| Svijet | 0.40 | 0.000 |

4.2.7 Korelacija indeksa strogosti odgovora vlade i slučajeva zaraze

Analizom korelacije između indeksa strogosti (engl. *Government Response Stringency Index*, složena mjera temeljena na 9 pokazatelja odgovora vlade uključujući zatvaranje škola, zatvaranje radnih mjesta i zabrane putovanja, preračunato na vrijednost od 0 do 100 (100 = najstroži odgovor)) i broja slučajeva zaraze, htjelo se utvrditi jesu li te dvije varijable povezane, kako bi se moglo odgovoriti na pitanje u kojoj su mjeri vladine mjere odgovora povezane s brojem novih pozitivnih slučajeva uzrokovanih COVID-19.

Varijable korištene u ovoj korelacijskoj analizi su indeks strogosti odgovora vlade i broj novih pozitivnih slučajeva zaraze. Rezultati analize vidljivi su u tablici 4.7.

Na temelju dobivenih koeficijenata korelacije, čini se da postoji pozitivna korelacija između strogosti odgovora vlade neke zemlje na pandemiju COVID-19, mjereno Indeksom strogosti odgovora vlade, i broja novih slučajeva zaraze u toj regiji. Koeficijenti korelacije su statistički značajni za sve kontinente.

Svijet općenito te obje Amerike, očitovali su umjerenu do dobru povezanost navedenih dviju varijabli, dok su Afrika, Azija, Australija i Europa detektirale slabu povezanost.

Tablica 4.7 Rezultati korelacijske analize indeksa strogosti i slučajeva zaraze korištenjem Spearmanovog koeficijenta korelacije ranga

| Regija | Koeficijent | P-vrijednost |
|------------------|-------------|--------------|
| Afrika | 0.47 | 0.000 |
| Azija | 0.44 | 0.000 |
| Australija | 0.28 | 0.000 |
| Europa | 0.49 | 0.000 |
| Južna Amerika | 0.52 | 0.000 |
| Sjeverna Amerika | 0.51 | 0.000 |
| Svijet | 0.53 | 0.000 |

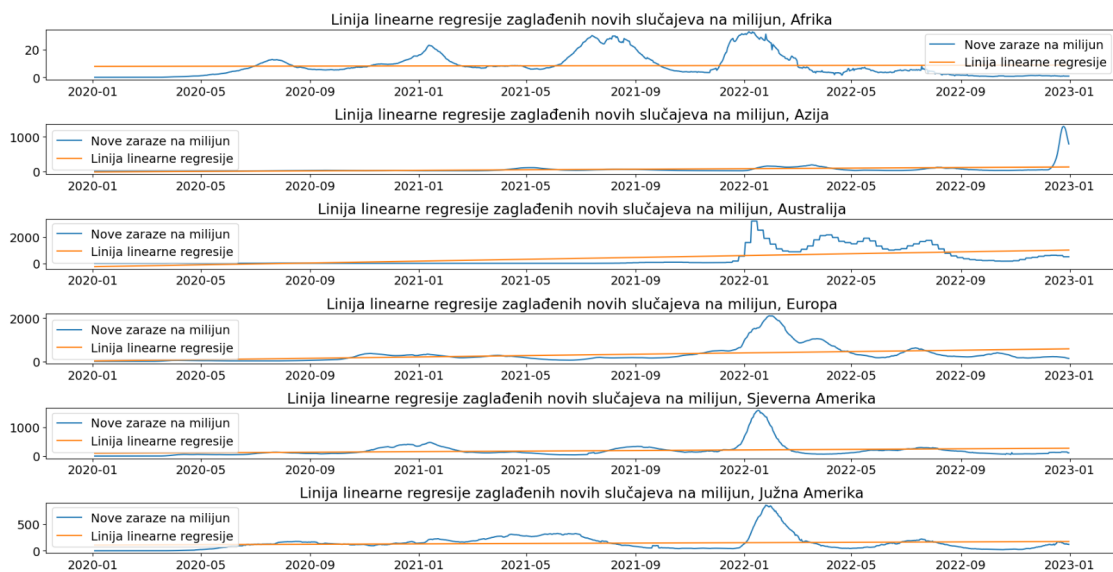
Međutim, važno je napomenuti da korelacija ne mora nužno implicirati uzročnost, s time da i drugi čimbenici također mogu utjecati na broj novih slučajeva zaraze u svakoj zemlji. Viši indeks strogosti znači da je vlada provela strože mjere za kontrolu širenja COVID-19, kao što su zatvaranje škola i radnih mjesta, zabrane putovanja i ograničenja javnih okupljanja. Očekuje se da će ove mjere dovesti do smanjenja širenja virusa i posljedično do smanjenja broja novih smrtnih slučajeva. Međutim, važno je napomenuti da koeficijent korelacije mjeri samo snagu odnosa između dvije varijable, a ne smjer odnosa. Stoga pozitivan koeficijent korelacije između indeksa strogosti i novih slučajeva zaraze ne znači nužno da strože mjere dovode do većeg broja slučajeva zaraze. To bi također moglo značiti da zemlje s višim indeksima strogosti obično imaju veći broj slučajeva i smrtnih slučajeva u početku, te stoga provode strože mjere kao odgovor.

4.3 Rezultati analize trenda

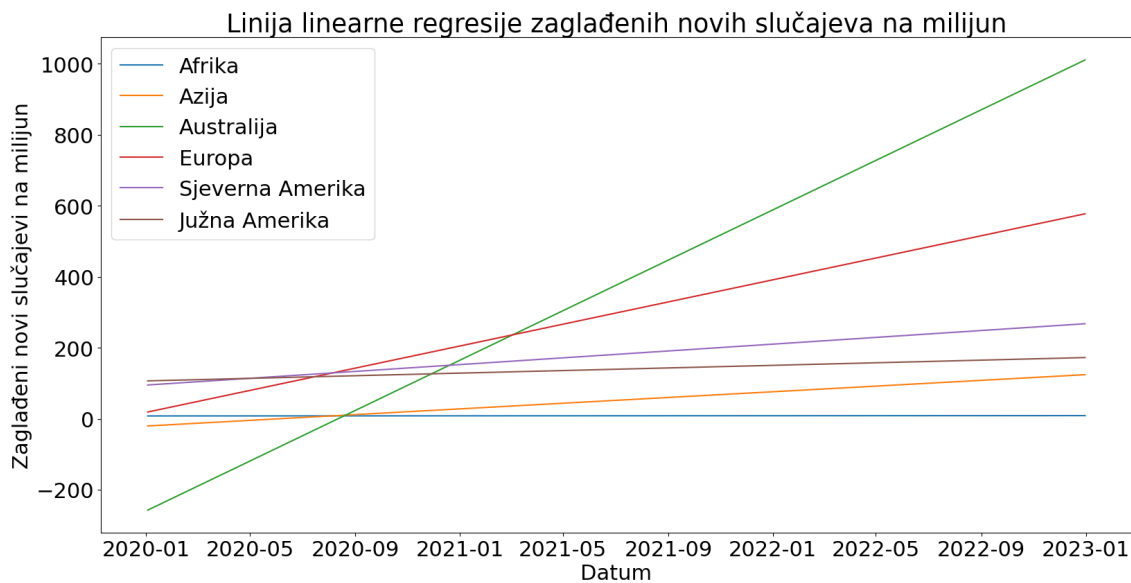
Linearni trend

Slike 4.3 i 4.4 prikazuju linije linearne regresije za nove slučajeve zaraze za svaki kontinent od 2020. do 2022. Linija linearne regresije predstavlja linearni trend u podacima i pokazuje smjer i nagib odnosa između dviju varijabli koje se analiziraju

Poglavlje 4. Usporedba COVID-19 podataka kroz 2020., 2021. i 2022.



Slika 4.3 Linearni trendovi novih slučajeva zaraze po kontinentima

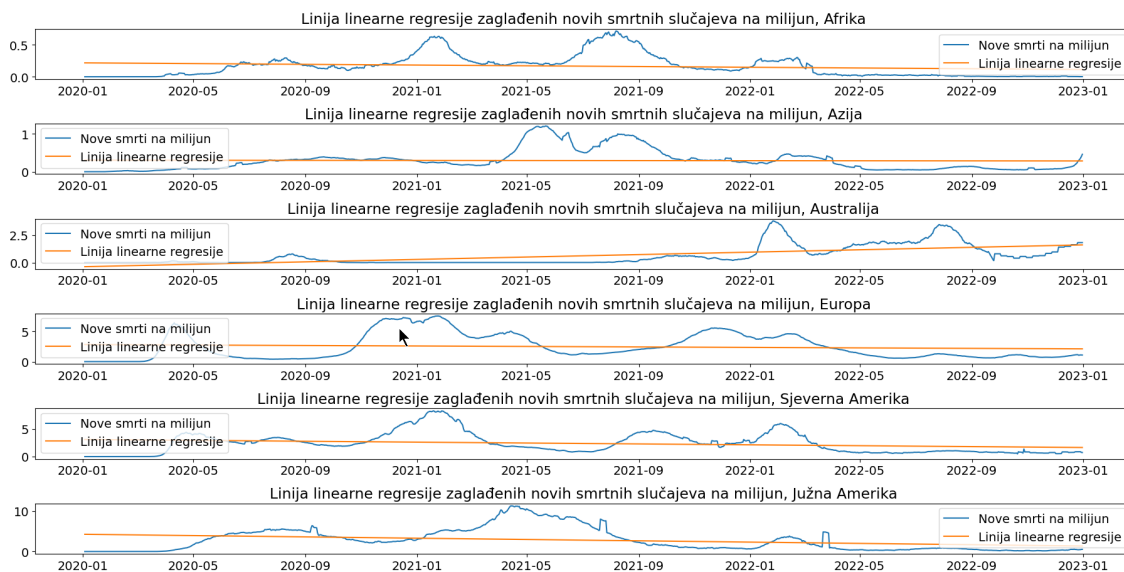


Slika 4.4 Linearni trendovi novih slučajeva zaraze - svi kontinenti

Poglavlje 4. Usporedba COVID-19 podataka kroz 2020., 2021. i 2022.

(u ovom slučaju, vrijeme i novi slučajevi zaraze). Na slici 4.3 je uz trend prikazana i zaglađena krivulja novih slučajeva zaraze za svaki kontinent, dok su na slici 4.3 prikazani samo linearni trendovi, kako bi ih se bolje moglo međusobno usporediti.

Iz slika je vidljivo da se za sve kontinente trend novih slučajeva zaraze povećava tijekom vremena. To znači da se u prosjeku broj novih slučajeva dnevno povećava s vremenom na tim kontinentima.

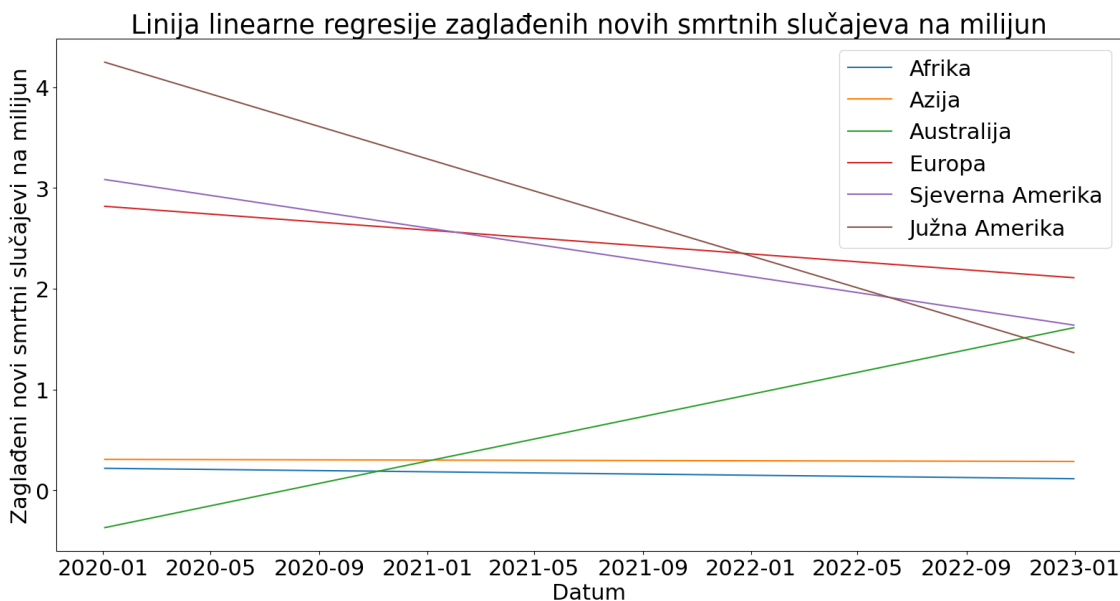


Slika 4.5 Linearni trendovi novih slučajeva smrti po kontinentima

Slike 4.5 i 4.6 prikazuju linije linearne regresije za nove smrtno slučajev za svaki kontinent od 2020. do 2022. Slično prethodnim slikama, linija linearne regresije predstavlja trend u podacima i pokazuje smjer i nagib odnosa između dviju varijabli (u ovom slučaju, vrijeme i nove smrtni slučajevi). Na slici 4.5 je uz trend prikazana i zaglađena krivulja novih smrtnih slučajeva za svaki kontinent, dok su na slici 4.5 prikazani samo trendovi, kako bi ih se bolje moglo međusobno usporediti.

Iz slika je vidljivo da se za sve kontinente osim za Australiju trend novih smrtnih slučajeva smanjuje tijekom vremena. To znači da se u prosjeku broj novih smrtnih slučajeva dnevno na tim kontinentima s vremenom smanjuje. Iznenađujuće je da trend novih smrtnih slučajeva u Australiji ne samo da raste, nego raste naglo, za

Poglavlje 4. Usporedba COVID-19 podataka kroz 2020., 2021. i 2022.



Slika 4.6 Linearni trendovi novih slučajeva smrti - svi kontinenti

razliku od svih ostalih kontinenata.

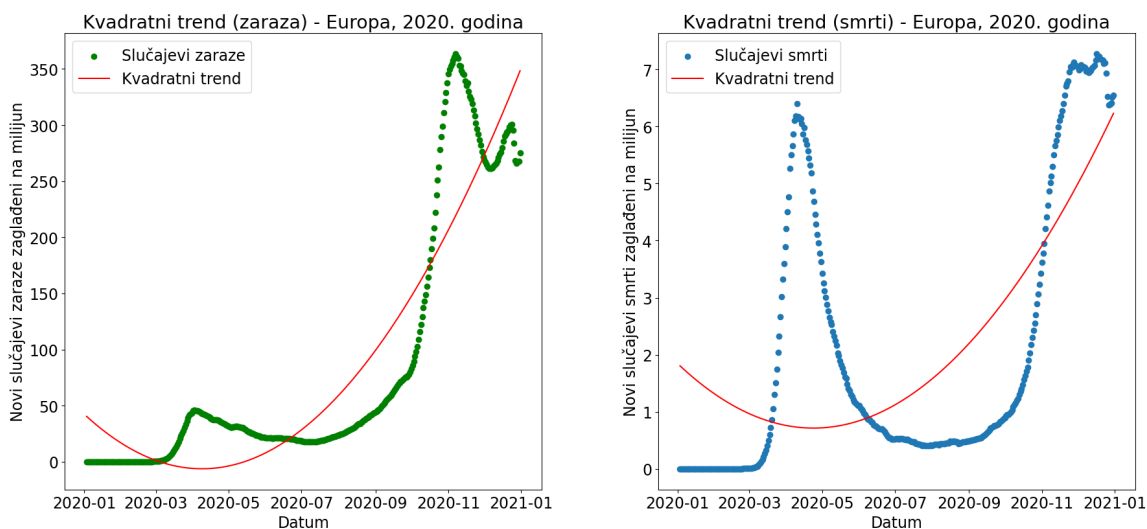
Važno je napomenuti da iako ove linije linearne regresije daju informacije o trendu u podacima, one ne obuhvaćaju nužno svu složenost i varijabilnost podataka. Također je važno te trendove tumačiti u kontekstu drugih čimbenika, kao što su stope procijepljenosti i vladine politike.

Kvadratni trend

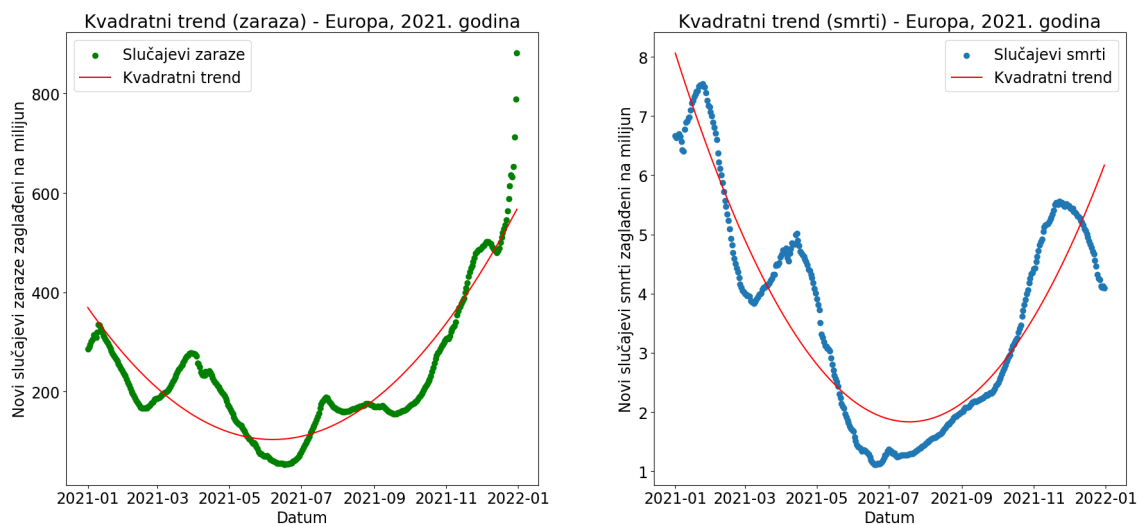
Pretpostavka je da će se, zbog svojstava COVID-19 podataka, kvadratni trend bolje prilagođavati podacima od linearnog trenda. Na slikama 4.7, 4.8, 4.9 i 4.10, prikazan je kvadratni trend COVID-19 zaraza i smrti za Europu kroz 4 godine pandemije (2020. - 2023.). Kvadratni trend prikazan je samo za Europu, budući da su svi ostali kontinenti pokazali slično ponašanje.

Već na prvi pogled, vidljivo je da se kvadratni trend bolje prilagođava podacima od linearnog. Što se zaraza tiče, može se zaključiti da je trend zaraza u Europi rastao u 2020., opadao do sredine 2021., da bi imao veliki skok s prijelazom u 2022., nakon

Poglavlje 4. Usporedba COVID-19 podataka kroz 2020., 2021. i 2022.



Slika 4.7 Kvadratni trendovi novih slučajeva zaraza i smrti - Europa, **2020.** godina

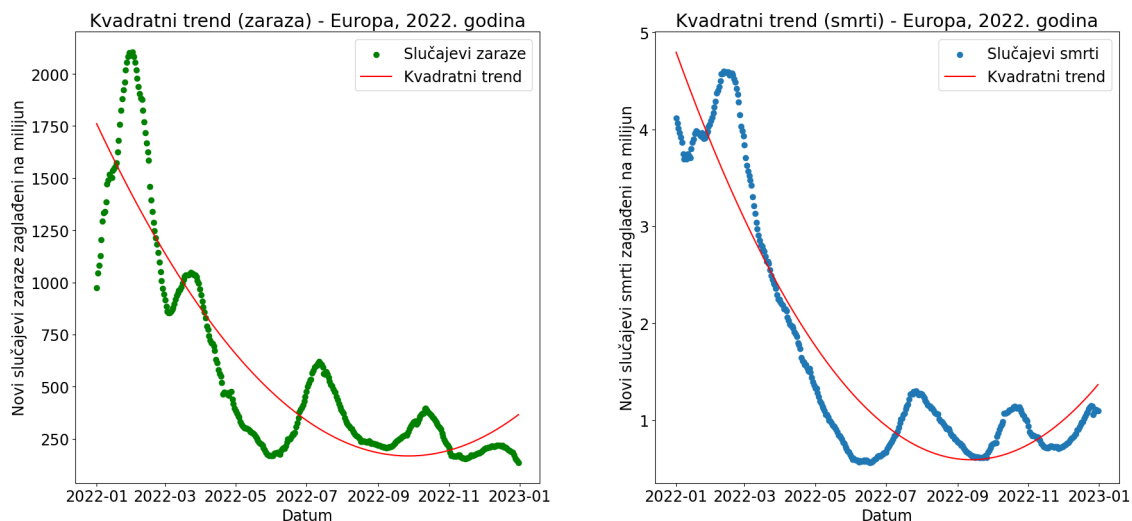


Slika 4.8 Kvadratni trendovi novih slučajeva zaraza i smrti - Europa, **2021.** godina

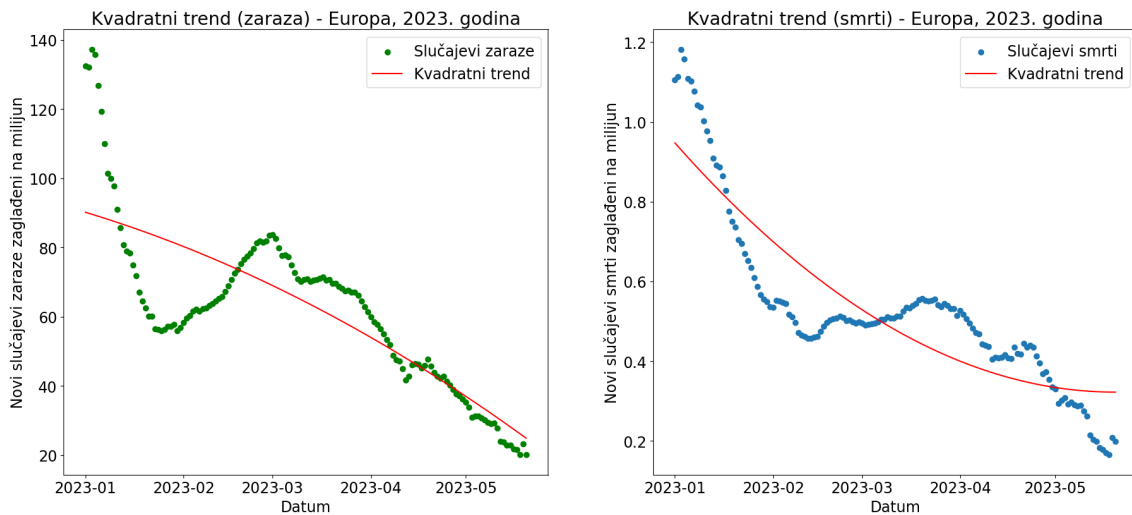
čega je uglavnom opadao do 20.5.2023.

Što se smrtnosti tiče, njen trend je značajno manji od trenda zaraza. On raste u 2020., vrhunac mu je na prijelazu 2020. u 2021., nakon čega kreće opadati da bi

Poglavlje 4. Usporedba COVID-19 podataka kroz 2020., 2021. i 2022.



Slika 4.9 Kvadratni trendovi novih slučajeva zaraza i smrti - Europa, **2022.** godina



Slika 4.10 Kvadratni trendovi novih slučajeva zaraza i smrti - Europa, **2023.** godina

opet krenuo rasti sredinom 2021. Ulaskom u 2022. opada sve do rujna, gdje se blago penje da bi opet počeo padati s prijelazom u 2023., u kojoj nastavlja svoj pad.

Iz provedene vizualne analize kvadratnog trenda da se zaključiti da se konstantno variranje COVID-19 podataka može puno bolje opisati kvadratnim trendom, nego

linearnim. U nastavku bit će određene značajnosti prikazanih trendova putem p i R-kvadrat vrijednosti.

4.3.1 Značajnost trenda

Linearni trend

Tablica 4.8 Značajnost linearnog trenda novih slučajeva zaraze i R-kvadrat vrijednosti

| Regija | P-vrijednost | R^2 |
|------------------|---------------|-------|
| Afrika | nije značajan | 0.001 |
| Azija | značajan | 0.124 |
| Australija | značajan | 0.321 |
| Europa | značajan | 0.184 |
| Južna Amerika | značajan | 0.019 |
| Sjeverna Amerika | značajan | 0.051 |

Tablica 4.9 Značajnost linearnog trenda novih smrtnih slučajeva i R-kvadrat vrijednosti

| Regija | P-vrijednost | R^2 |
|------------------|---------------|-------|
| Afrika | značajan | 0.031 |
| Azija | nije značajan | 0.001 |
| Australija | značajan | 0.426 |
| Europa | značajan | 0.010 |
| Južna Amerika | značajan | 0.088 |
| Sjeverna Amerika | značajan | 0.051 |

Tablica 4.8 prikazuje p i R-kvadrat vrijednosti kako bi se odredila značajnost linearnog trenda novih slučajeva zaraze. Tablica 4.9 prikazuje te iste vrijednosti kako bi se odredila značajnost linearnog trenda novih smrtnih slučajeva. Novodobivenim

Poglavlje 4. Usporedba COVID-19 podataka kroz 2020., 2021. i 2022.

rezultatima iz tablica 4.8 i 4.9 trebaju se nadopuniti zaključci dobiveni iz deskriptivne analize iz prethodnog poglavlja.

Što se tiče trendova novih slučajeva zaraze, tablica 4.8 potvrđuje da je trend povećanja novih slučajeva zaraze značajan za sve kontinente osim Afrike, što bi značilo da postoji mala vjerojatnost da je trend proizašao iz slučajnosti. R-kvadrat vrijednosti pokazuju da regresijski modeli za obje Amerike nisu reprezentativni, za Aziju i Europu su slabo reprezentativni, dok su za Australiju umjereno do dobro reprezentativni.

Što se tiče trendova novih smrtnih slučajeva, tablica 4.9 potvrđuje da je trend smanjenja novih smrtnih slučajeva značajan za sve kontinente osim Azije, što bi značilo da postoji mala vjerojatnost da je trend proizašao iz slučajnosti. Također, trend povećanja novih smrtnih slučajeva u Australiji je također značajan. R-kvadrat vrijednosti pokazuju da je regresijski model za Australiju umjereno do dobro reprezentativan, za Južnu Ameriku slabo reprezentativan, dok se regresijski modeli za sve ostale regije ne prilagođavaju podacima dovoljno dobro.

Kvadratni trend

Tablice 4.10 i 4.11 prikazuju koliko je kvadratni trend značajan za pojedinu regiju u određenoj godini, te, ukoliko je značajan, koliko se dobro prilagođava podacima.

Iz tablice 4.10 vidljivo je da kvadratni trend slučajeva zaraze za Europu u 2023. godini nije značajan, što znači da možemo odbaciti lijevi graf na slici 4.10. Svi ostali grafovi koji prikazuju kvadratni trend (bilo slučajeva zaraze, bilo slučajeva smrti) za Europu kroz godine su značajni te pokazuju vrlo dobru do izvrsnu povezanost podataka i kvadratnog modela, s iznimkom kvadratnog trenda smrtnosti za 2020. koji pokazuje umjerenu do dobru povezanost.

Kvadratni trendovi slučajeva zaraze za ostale kontinente su uglavnom značajni, s iznimkama Afrike i Azije u 2020., Afrike u 2021. te Australije u 2023. godini. Kvadratni trend koji najbolje opisuje podatke je onaj za Sjevernu Ameriku u 2023. godini. Podatke najlošije opisuje kvadratni trend za Australiju u 2020. godini. Ostali trendovi većinom vrlo dobro do izvrsno opisuju podatke, dok ih neki opisuju umjereno dobro do dobro.

Poglavlje 4. Usporedba COVID-19 podataka kroz 2020., 2021. i 2022.

Tablica 4.10 Značajnost kvadratnog trenda novih slučajeva zaraze i R-kvadrat vrijednosti po godinama

| Godina | Regija | P-vrijednost | R^2 |
|--------|------------------|---------------|-------|
| 2020. | Afrika | nije značajan | 0.716 |
| | Azija | nije značajan | 0.859 |
| | Australija | značajan | 0.195 |
| | Europa | značajan | 0.822 |
| | Južna Amerika | značajan | 0.802 |
| | Sjeverna Amerika | značajan | 0.840 |
| 2021. | Afrika | nije značajan | 0.010 |
| | Azija | značajan | 0.427 |
| | Australija | značajan | 0.460 |
| | Europa | značajan | 0.822 |
| | Južna Amerika | značajan | 0.686 |
| | Sjeverna Amerika | značajan | 0.335 |
| 2022. | Afrika | značajan | 0.791 |
| | Azija | značajan | 0.253 |
| | Australija | nije značajan | 0.559 |
| | Europa | značajan | 0.799 |
| | Južna Amerika | značajan | 0.556 |
| | Sjeverna Amerika | značajan | 0.500 |
| 2023. | Afrika | značajan | 0.731 |
| | Azija | značajan | 0.580 |
| | Australija | značajan | 0.620 |
| | Europa | nije značajan | 0.650 |
| | Južna Amerika | značajan | 0.807 |
| | Sjeverna Amerika | značajan | 0.914 |

Kvadratni trendovi slučajeva smrti za ostale kontinente osim Europe su većinom

Poglavlje 4. Usporedba COVID-19 podataka kroz 2020., 2021. i 2022.

Tablica 4.11 Značajnost kvadratnog trenda novih slučajeva smrti i R-kvadrat vrijednosti po godinama

| Godina | Regija | P-vrijednost | R^2 |
|--------|------------------|---------------|-------|
| 2020. | Afrika | značajan | 0.696 |
| | Azija | značajan | 0.895 |
| | Australija | značajan | 0.202 |
| | Europa | značajan | 0.402 |
| | Južna Amerika | značajan | 0.737 |
| | Sjeverna Amerika | značajan | 0.511 |
| 2021. | Afrika | značajan | 0.155 |
| | Azija | značajan | 0.563 |
| | Australija | značajan | 0.648 |
| | Europa | značajan | 0.828 |
| | Južna Amerika | značajan | 0.691 |
| | Sjeverna Amerika | značajan | 0.529 |
| 2022. | Afrika | značajan | 0.784 |
| | Azija | značajan | 0.554 |
| | Australija | značajan | 0.119 |
| | Europa | značajan | 0.874 |
| | Južna Amerika | značajan | 0.460 |
| | Sjeverna Amerika | značajan | 0.698 |
| 2023. | Afrika | nije značajan | 0.236 |
| | Azija | značajan | 0.852 |
| | Australija | značajan | 0.653 |
| | Europa | značajan | 0.758 |
| | Južna Amerika | značajan | 0.775 |
| | Sjeverna Amerika | nije značajan | 0.729 |

značajni, s iznimkom Afrike i Sjeverne Amerike u 2023. godini. Podatke najbolje

Poglavlje 4. Usporedba COVID-19 podataka kroz 2020., 2021. i 2022.

opisuje kvadratni trend za Aziju u 2020., dok ih najlošije opisuje onaj izračunat za Australiju u 2022. godini. Ostali trendovi također većinom vrlo dobro do izvrsno opisuju podatke, dok ih neki opisuju umjereno dobro do dobro.

Poglavlje 5

Trenutno stanje pandemije (2023.)

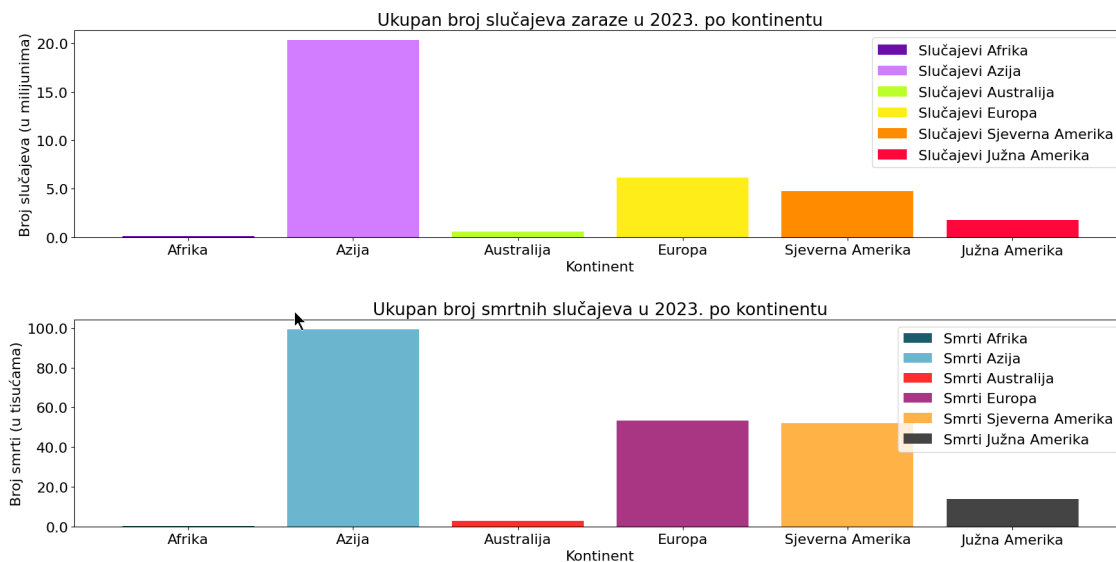
Ovaj odjeljak ima za cilj pružiti ažuriranu analizu COVID-19 situacije u 2023. godini. Stvaranje zasebnog odjeljka za 2023. godinu, umjesto stapanja s prethodnim odjeljkom u kojem su obrađeni podaci kroz 2020., 2021. i 2022., omogućuje usmjeravanje pozornosti na trenutnu situaciju, ističući najnovije podatke, trendove i razvoj. Cilj ovog pristupa je omogućiti jasnu razliku između povijesne analize i sadašnjeg stanja pandemije, kako bi se omogućilo čitateljima da se posebno usredotoče na najnovija događanja i njihove implikacije bez da ih ometa povijesna analiza.

5.1 Stanje u svijetu

Najbolji pregled trenutnog stanja COVID-19 u svijetu (2023. godine) može se odrediti analizom distribucije frekvencija, što prikazuje slika 5.1. Na njoj je prvim podgrafikonom prikazan ukupan broj slučajeva COVID-19 po kontinentu u 2023. (do 20.5.2023.), pružajući pregled distribucije slučajeva u različitim regijama. Drugi podgrafikon prikazuje ukupan broj smrtnih slučajeva od COVID-19 po kontinentu, nudeći uvid u utjecaj virusa na globalnoj razini. Zajedno, ovi dijagrami pružaju vizualni prikaz COVID-19 situacije na različitim kontinentima, ističući razlike u stopama zaraze i smrtnosti po geografskim područjima.

Iz grafikona je vidljivo da Azija i dalje ima najveći broj slučajeva zaraze i smrtnih slučajeva, značajno nadmašujući ostale kontinente. To se može pripisati čimbenicima

Poglavlje 5. Trenutno stanje pandemije (2023.)



Slika 5.1 Apsolutan broj slučajeva zaraze i smrti u 2023. po kontinentima

kao što su gustoća naseljenosti, međunarodna putovanja i različite razine zdravstvene infrastrukture.

Europa i Sjeverna Amerika također još uvijek imaju značajan broj novih slučajeva zaraze i smrtnih slučajeva. Ostali kontinenti, sudeći prema nižim brojevima novih slučajeva zaraze te smrtnih slučajeva, se bolje nose s pandemijom. Detaljnije brojke mogu se vidjeti u tablici 5.1.

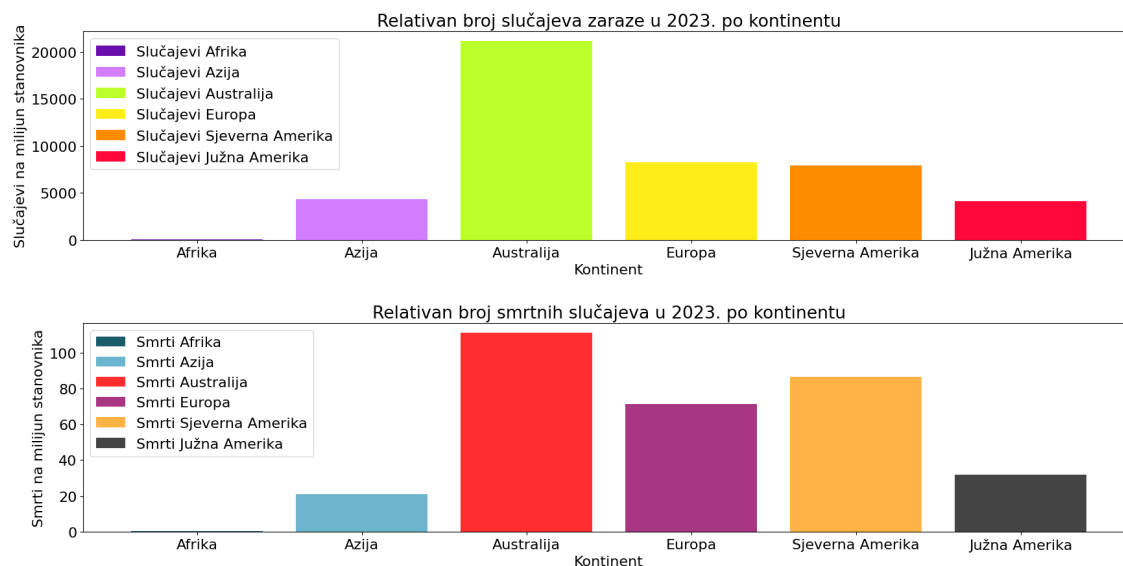
Tablica 5.1 Slučajevi zaraze i smrtni slučajevi po kontinentima u 2023. (do 20.5.2023.)

| Kontinent | Slučajevi zaraze | Smrtni slučajevi |
|------------------|------------------|------------------|
| Afrika | 88,171 | 436 |
| Azija | 20,321,600 | 99,185 |
| Australija | 553,637 | 2,905 |
| Europa | 6,174,742 | 53,233 |
| Sjeverna Amerika | 4,759,715 | 51,993 |
| Južna Amerika | 1,788,579 | 13,910 |

Kako se kontinenti uvelike razlikuju po broju stanovnika, milijuni zaraza u Aziji i

Poglavlje 5. Trenutno stanje pandemije (2023.)

milijuni zaraza u Australiji ne znače istu stvar. Stoga se prethodno prikazane brojke treba staviti u kontekst broja stanovništva koje živi na području nekog kontinenta.



Slika 5.2 Relativan broj slučajeva zaraze i smrti u 2023. po kontinentima, na milijun stanovnika

Stupčasti dijagrami na slici 5.2 prikazuju normalizirane vrijednosti COVID-19 brojke kao broj slučajeva, odnosno broj smrti, na milijun stanovnika koji žive na nekom kontinentu. Kada se relativne vrijednosti usporede s apsolutnima sa slike 5.1, vidljivo je da se slike značajno razlikuju.

Vrijednosti za Afriku su niske u oba slučaja. Azija i Australija imaju najveću razliku između apsolutnih i relativnih vrijednosti. Dok je Azija imala najviše apsolutne vrijednosti zaraza i smrtnosti (logično, budući da je kontinent s najviše stanovnika), kada se te vrijednosti normaliziraju, vidljivo je da situacija i nije toliko strašna kao što se čini. Štoviše, COVID-19 slika Azije je druga najbolja (nakon Afrike), prema relativnim vrijednostima.

Ista diskrepancija može se primijetiti i kod Australije. Dok su njene apsolutne vrijednosti druge najniže (što je, opet, logično, budući da je kontinent s najmanje stanovnika), njene relativne vrijednosti uvelike nadmašuju sve ostale kontinente. To odgovara linearnom trendu sa slika 4.4 i 4.6, gdje je vidljivo da je trend Austra-

Poglavlje 5. Trenutno stanje pandemije (2023.)

lije najbrže rastući, što se tiče slučajeva zaraze, odnosno jedini rastući, što se tiče slučajeva smrti.

Europa, Sjeverna i Južna Amerika su uglavnom zadržale iste odnose, s iznimkom broja smrtnih slučajeva na milijun stanovnika u Sjevernoj Americi, koji je nadmašio onaj u Europi, dok bi se po apsolutnim vrijednostima dalo zaključiti da je situacija sa smrtnosti u Europi gora od one u Sjevernoj Americi.

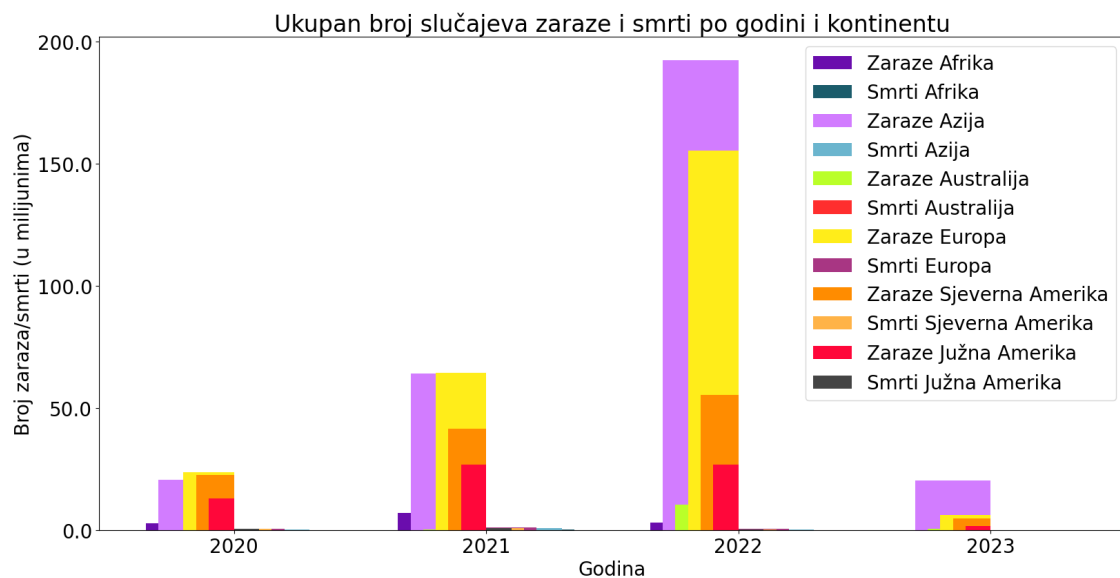
5.2 Usporedba s prijašnjim godinama

Kako bi se dobio sveobuhvatan kontekst za prikazane brojeve, ukupan broj COVID-19 slučajeva i smrtnih slučajeva po kontinentima u 2023. godini bit će uspoređen s podacima iz prethodnih godina, odnosno 2020., 2021. i 2022. godinom. Razumijevanjem tijeka pandemije kroz više godina može se poboljšati razumijevanje trenutne situacije i olakšati donošenje procjene o učinkovitosti intervencija, točnije, strategije WHO-a da više ne smatra COVID-19 javnozdravstvenom opasnošću koja izaziva međunarodnu zabrinutost te odluke Vlade RH kojom je proglasila kraj epidemije u cijeloj Hrvatskoj.

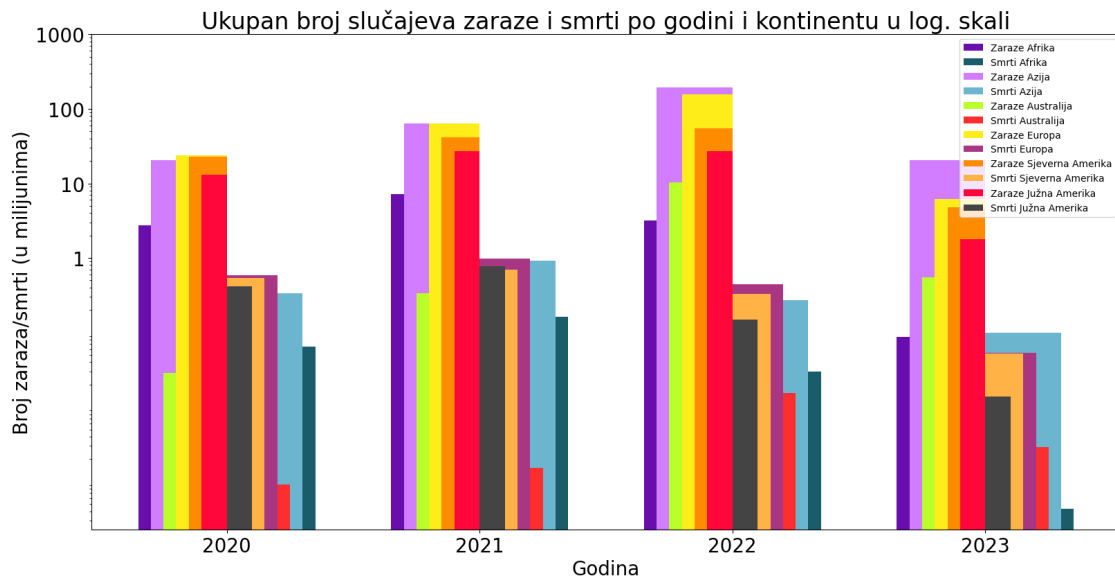
Kada se prouče grafikoni na slikama 5.3 i 5.4, koji pružaju povijesni kontekst za brojke predstavljene u prošloj sekciji, može se dobiti drugačija perspektiva. Iako se brojevi u početku mogu činiti velikima i alarmantnima, uspoređujući ih s prethodnim godinama otkriva se da nisu tako zastrašujući kao što se čine. Treba imati na umu da se podaci za 2023. prikazani na grafovima odnose samo na prvih 5. mjeseci 2023., točnije do 20.5.2023.

Grafikon na slici 5.3 prikazan je u linearnoj skali, dok je grafikon na slici 5.4 prikazan u logaritamskoj. Logaritamska skala je vrsta skale koja omogućuje učinkovitije predstavljanje podataka koji obuhvaćaju širok raspon vrijednosti. Za razliku od linearne skale, koja ravnomjerno raspoređuje vrijednosti duž osi, logaritamska skala eksponencijalno komprimira vrijednosti. To znači da se udaljenost između dviju točaka na logaritamskoj skali povećava kako vrijednosti postaju veće.

Poglavlje 5. Trenutno stanje pandemije (2023.)



Slika 5.3 Stupčasti dijagram ukupnih slučajeva zaraze i smrtnih slučajeva po svim kontinentima od 2020. do 2023.



Slika 5.4 Stupčasti dijagram ukupnih slučajeva zaraze i smrtnih slučajeva po svim kontinentima od 2020. do 2023. u logaritamskoj skali

Poglavlje 5. Trenutno stanje pandemije (2023.)

Također, logaritamska skala otkriva obrasce koji nisu odmah vidljivi na linearnoj skali (primjerice broj smrtnih slučajeva kroz godine), te je upravo zbog toga i korištena. [12] Proučavajući podatke prikazane uz pomoć logaritamske skale, treba pripaziti na percepciju podataka, budući da je ljudski mozak uglavnom navikao na linearnu skalu.

Zaključci koji se mogu izvesti promatrajući dane grafove su:

Broj slučajeva zaraze općenito je rastao za većinu kontinenata do 2022. godine, da bi zatim došlo do značajnog pada u 2023. godini. To sugerira da su možda provedene uspješne mjere suzbijanja ili da je došlo do smanjenja ukupne transmisije virusa na globalnoj razini.

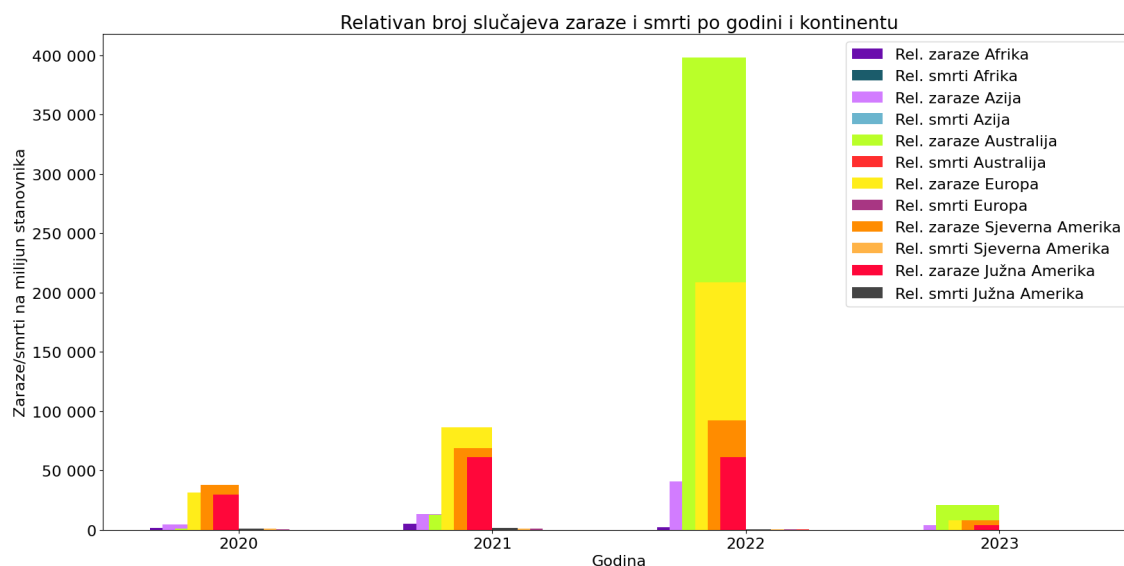
Broj smrtnih slučajeva također je rastao do kraja 2021. godine za većinu kontinenata. Međutim, u 2022. godini došlo je do smanjenja broja smrtnih slučajeva za sve kontinente osim Australije. To može ukazivati na poboljšane medicinske tretmane, bolju procijepljenost ili učinkovite mjere javnog zdravstva u kontroliranju stope smrtnosti.

Broj slučajeva zaraze je značajno veći od broja smrtnih slučajeva. To ukazuje da je ozbiljnost i stopa smrtnosti virusa bila relativno niska ili da su medicinski postupci i zdravstveni sustavi uspješno upravljali slučajevima, smanjujući stopu smrtnosti.

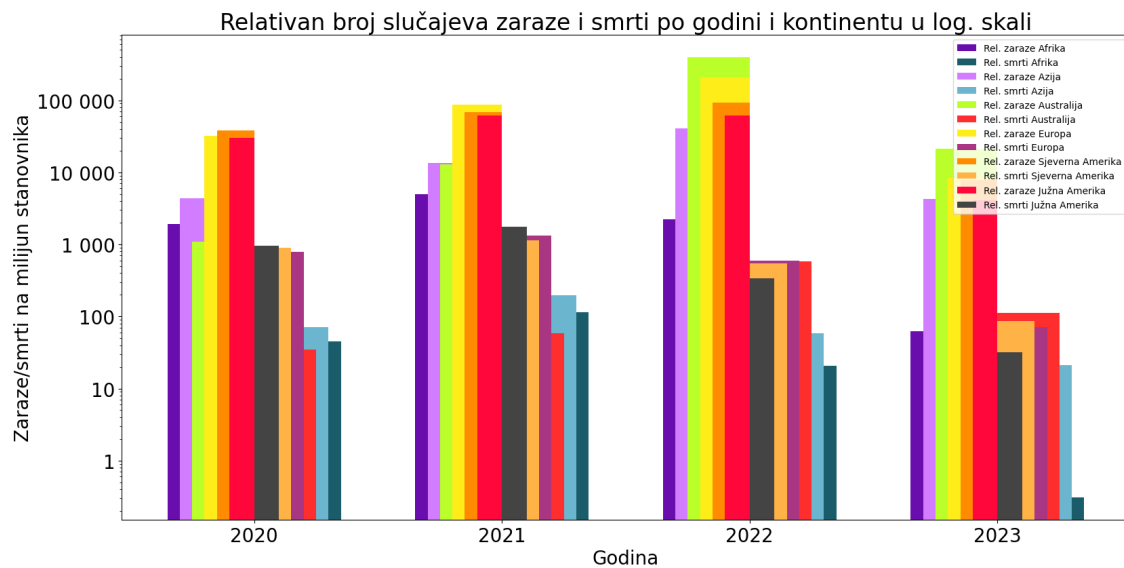
Kako bi se analizirane brojke normalizirale, stavljene su u kontekst broja stanovništva koje živi na području nekog kontinenta. Stupčasti dijagrami na slikama 5.5 i 5.6 prikazuju normalizirane vrijednosti COVID-19 brojki kao broj slučajeva, odnosno broj smrti, na milijun stanovnika nekog kontinenta, od 2020. do 2023. godine, u linearnoj i logaritamskoj skali.

Usporedbom sa slikama 5.3 i 5.4 vidljivo je da su odnosi slučajeva zaraze i slučajeva smrti uglavnom u istim omjerima. Također, uglavnom je ista i fluktuacija brojki zaraze i smrti na razini kontinenata. Ono što se razlikuje su visine stupaca za pojedini kontinent. Najveću razliku između apsolutnih i relativnih vrijednosti ima Australija. Što se tiče apsolutnih vrijednosti, na grafu je gotovo nezamjetljiva, ali kad se uspoređuju relativne vrijednosti, ona tada prednjači po broju slučajeva zaraza i smrti, posebice u zadnje dvije godine.

Poglavlje 5. Trenutno stanje pandemije (2023.)



Slika 5.5 Stupčasti dijagram relativnih slučajeva zaraze i smrtnih slučajeva po svim kontinentima od 2020. do 2023., normalizirani na milijun stanovnika



Slika 5.6 Stupčasti dijagram relativnih slučajeva zaraze i smrtnih slučajeva po svim kontinentima od 2020. do 2023., normalizirani na milijun stanovnika u logaritamskoj skali

Poglavlje 5. Trenutno stanje pandemije (2023.)

Slično ponašanje može se primijetiti i kod Azije. Sudeći prema apsolutnim vrijednostima ona ima najgoru COVID-19 situaciju, no ako se pogledaju relativne vrijednosti, može se uočiti da je njena COVID-19 slika zapravo druga najbolja, nakon Afrike.

Kada se promotri broj slučajeva zaraze na milijun ljudi za Australiju u 2022. godini, primjetno je da je on gotovo dvostruko veći od drugorangirane Europe. Doda li se tomu navod iz [13], gdje ravnateljica Instituta Doherty i liječnica za zarazne bolesti, prof. Sharon Lewin, iznosi da je *"78% osoba u dobi od 16 i više godina primilo dvije doze cjepiva do sredine prosinca 2021."*, postavlja se pitanje efikasnosti cjepiva u smanjivanju broja zaraza.

Sudeći po podacima, pitanje je je li opravdana odluka WHO-a da više ne smatra COVID-19 javnozdravstvenom opasnošću koja izaziva međunarodnu zabrinutost? To je pitanje na koje moraju odgovoriti stručnjaci na temelju najboljih dostupnih znanstvenih dokaza, stručnih mišljenja i temeljitog razumijevanja globalnog zdravstvenog konteksta, no na temelju analiziranih podataka, može se dati barem nestručno mišljenje.

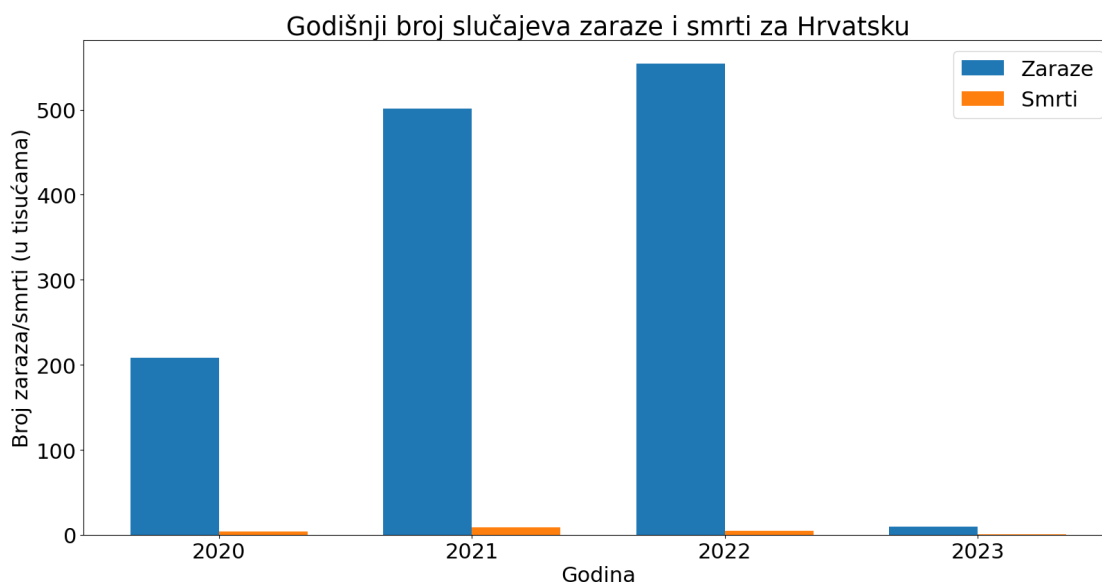
Smanjenje broja oboljelih i umrlih osoba, visoka pokrivenost cijepljenjem, poboljšana pripravnost zemalja i zdravstvenih sustava, ozbiljne socijalnoekonomske posljedice pandemije te posljedice na mentalno zdravlje pojedinca, svi ti čimbenici čine povoljne uvjete za više ne smatrati COVID-19 javnozdravstvenom opasnošću koja izaziva međunarodnu zabrinutost.

S druge strane, treba ostati na oprezu zbog novih varijanti virusa te nerazumijevanja dugoročnih učinaka virusa. Zaključak bi bio da je opravdano više ne smatrati COVID-19 javnozdravstvenom opasnošću koja izaziva međunarodnu zabrinutost, uz ostanak na oprezu i pripravnost na buduće prijetnje zdravlju.

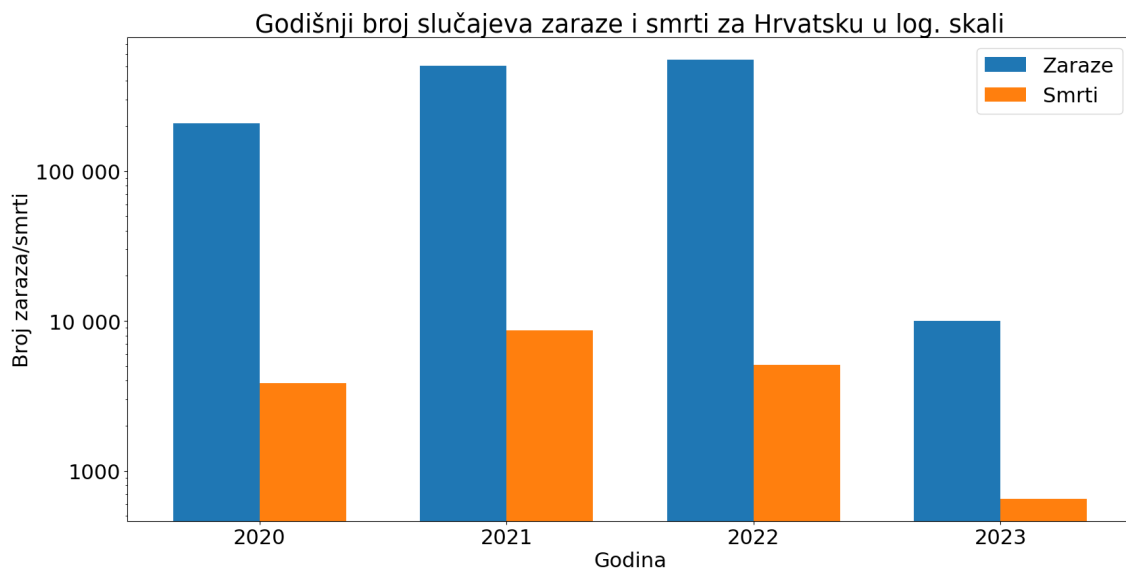
5.3 Stanje u Hrvatskoj

Slike 5.7 i 5.8 prikazuju ukupan broj COVID-19 slučajeva i smrtnih slučajeva u Hrvatskoj za godine 2020. - 2023. Dijagrami prikazuju napredak pandemije tijekom vremena, omogućujući vizualnu usporedbu trendova. Logaritamska skala na slici 5.8

Poglavlje 5. Trenutno stanje pandemije (2023.)



Slika 5.7 Godišnji broj slučajeva zaraze i smrtnih slučajeva u Republici Hrvatskoj



Slika 5.8 Godišnji broj slučajeva zaraze i smrtnih slučajeva u Republici Hrvatskoj u logaritamskoj skali

Poglavlje 5. Trenutno stanje pandemije (2023.)

pruža jasniji prikaz relativnih stopa rasta, dok linearna skala na slici 5.7 pruža realan prikaz omjera veličina. Iz dijagrama je vidljivo da je trenutno stanje pandemije u Hrvatskoj na zavidno niskoj razini, bar što se usporedbe s prijašnjim godinama tiče, a pogotovo po pitanju smrtnosti. Tablica s egzaktnim godišnjim brojkama za Hrvatsku jest tablica 5.2.

Tablica 5.2 Godišnji slučajevi zaraze i smrtni slučajevi u Hrvatskoj

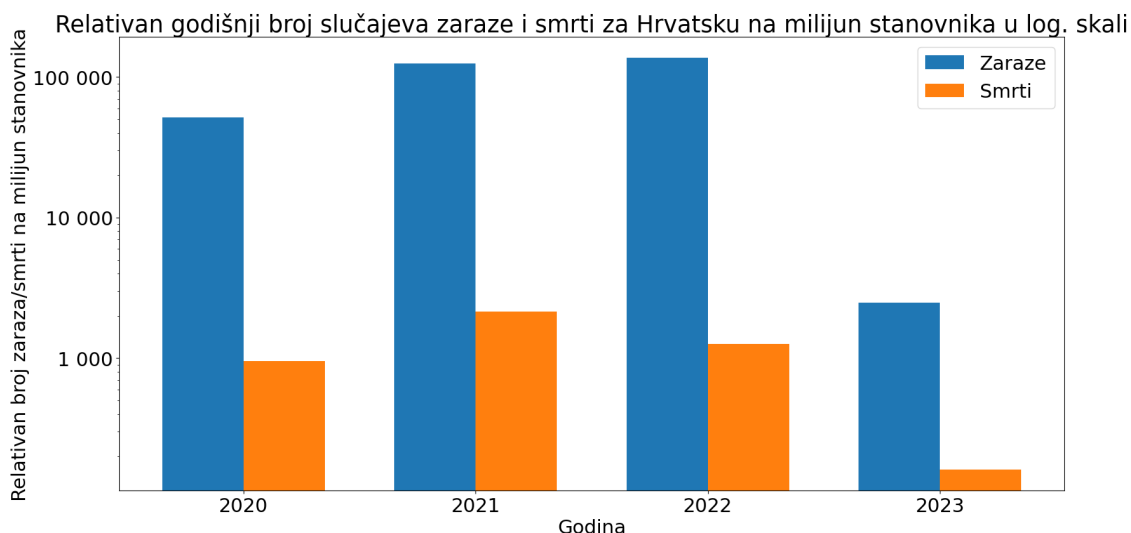
| Godina | Slučajevi zaraze | Smrtni slučajevi |
|--------|------------------|------------------|
| 2020 | 208,446.0 | 3,860.0 |
| 2021 | 501,232.0 | 8,633.0 |
| 2022 | 553,839.0 | 5,090.0 |
| 2023 | 9991.0 | 647.0 |

Kada se godišnje brojke normaliziraju, dobiju se relativne vrijednosti zaraza i smrti na milijun stanovnika u Hrvatskoj. To prikazuju slike 5.9 i 5.10 u linearnoj, odnosno logaritamskoj skali.



Slika 5.9 Relativni godišnji broj slučajeva zaraze i smrtnih slučajeva u Republici Hrvatskoj na milijun njenih stanovnika

Poglavlje 5. Trenutno stanje pandemije (2023.)



Slika 5.10 Relativni godišnji broj slučajeva zaraze i smrtnih slučajeva u Republici Hrvatskoj na milijun njenih stanovnika, u logaritamskoj skali

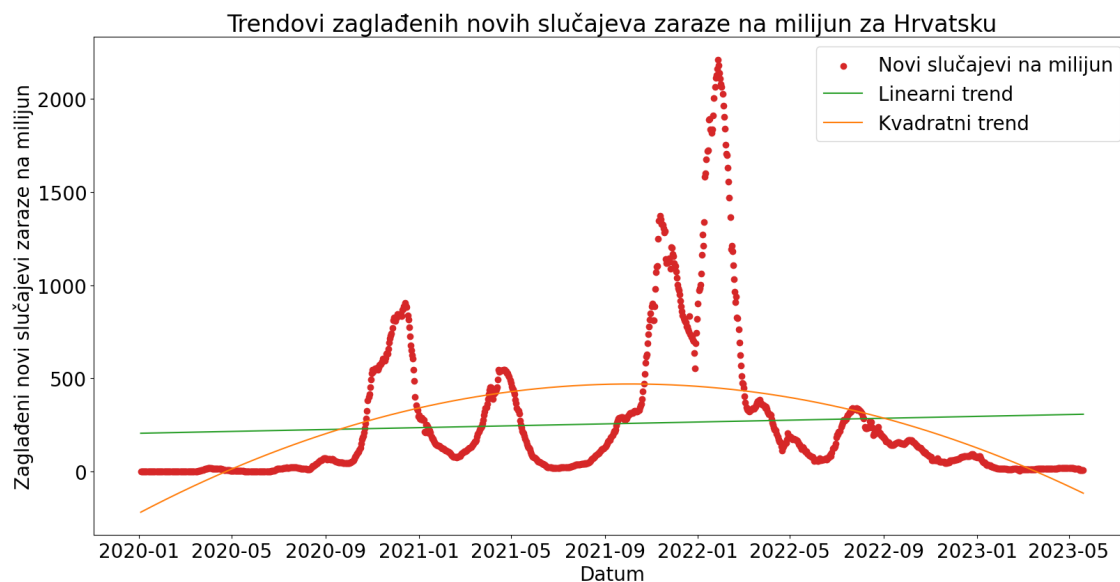
Kada se takvu situaciju usporedi sa situacijom u Europi (slike 5.5 i 5.6), vidljivo je da je situacija u Hrvatskoj bila gora nego u Europi u 2020. i 2021. godini, nakon čega se počela normalizirati te je 2022. i 2023. godine imala bolju COVID-19 sliku nego ostatak Europe, što se zaraza tiče.

Što se tiče slučajeva smrti, Hrvatska je bila ispod europskog prosjeka samo u 2023. godini, dok je sve ostale godine bila iznad europskog prosjeka. Ako se u obzir uzme da je Europa uglavnom u vrhu prema broju zaraza i smrti, može se reći da je i Hrvatska uglavnom u samom svjetskom vrhu prema broju zaraza i smrti, što je poražavajuća statistika.

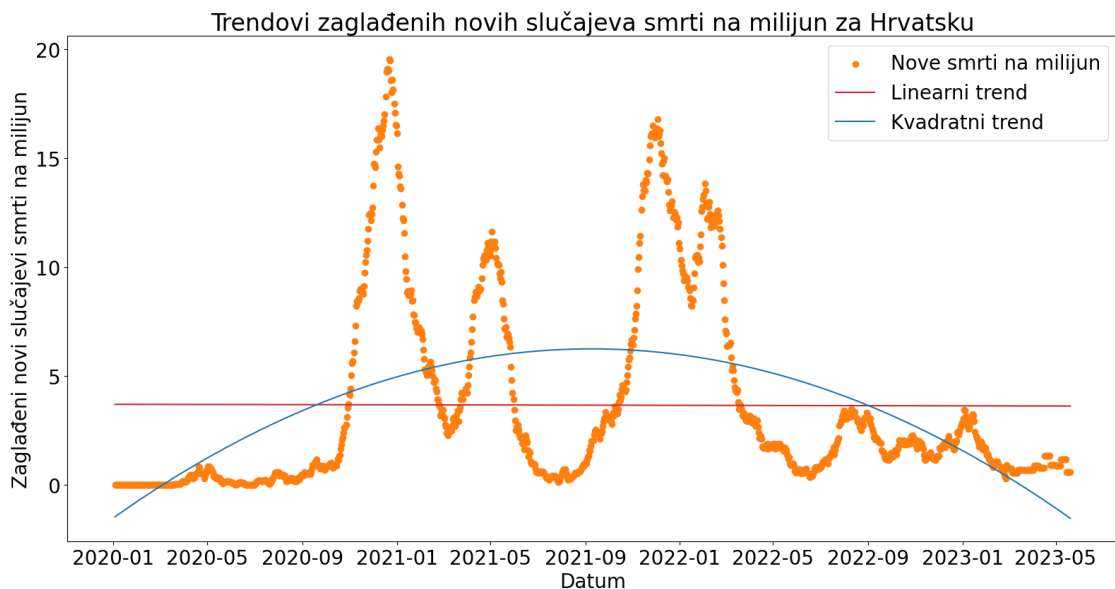
Kako bi se hrvatski trendovi lakše usporedili s grafovima trendova na razini svakog kontinenta zasebno (prikazano u prijašnjem poglavlju), analizirani su zaglađeni novi slučajevi zaraze i novi smrtni slučajevi po milijunu stanovnika u Republici Hrvatskoj. Slike 5.11 i 5.12 prikazuju ove vrijednosti, zajedno s njihovim linearnim i kvadratnim trendovima. Tehnika zaglađivanja korištena je jer pomaže smanjiti dnevne fluktuacije i pruža precizniji prikaz temeljnih trendova.

Iz grafova na slikama vidljivo je da je linearni trend novih slučajeva zaraze u

Poglavlje 5. Trenutno stanje pandemije (2023.)



Slika 5.11 Trendovi novih slučajeva zaraze u Republici Hrvatskoj



Slika 5.12 Trendovi novih smrtnih slučajeva u Republici Hrvatskoj

Poglavlje 5. Trenutno stanje pandemije (2023.)

blagom porastu, iako su brojke za 2023. vrlo niske. Razlog tomu je sporost linearnog trenda da se prilagodi promjenama, budući da je trend do 2023. bio značajno rastući. Linearni trend novih smrtnih slučajeva je u blagom, jedva primjetnom padu.

Što se tiče kvadratnih trendova, oni lijepo opisuju i zaokružuju cijeli zaplet, vrhunac i rasplet pandemije. U početku je trend zaraza rastao, sve do sredine 2021. godine. Od tada uglavnom stagnira, da bi sredinom 2022. krenuo padati. Trend smrtnih slučajeva je nešto oštriji pa nema stagnacijski plato. On raste do kraja ljeta 2021. nakon čega kreće padati. P i R-kvadrat vrijednosti pokazuju da su kvadratni trendovi značajni i umjereno dobri.

Sveukupno, dijagrami, trend linije i tablični podaci prikazani u ovom odjeljku pružaju sveobuhvatnu analizu COVID-19 situacije u Hrvatskoj u 2023. godini. Ove vizualizacije i statistike omogućuju dublje razumijevanje napretka pandemije, omogućujući usporedbu između različitih godina i identifikaciju značajnih trendova ili uzoraka.

Sudeći po navedenim podacima, pitanje je je li opravdana odluka Vlade Republike Hrvatske da proglasi kraj epidemije bolesti COVID-19? Kao i pitanje vezano za globalnu razinu, i to je pitanje na koje moraju odgovoriti stručnjaci na temelju najboljih dostupnih znanstvenih dokaza, stručnih mišljenja i temeljitog razumijevanja globalnog zdravstvenog konteksta, no na temelju analiziranih podataka, može se dati barem nestručno mišljenje.

Drastično smanjenje broja oboljelih i umrlih osoba u RH, visoka pokrivenost cijepljenjem, poboljšana pripravnost hrvatskog zdravstvenog sustava, ozbiljne socijalnoekonomske posljedice epidemije te posljedice na mentalno zdravlje građana, svi ti čimbenici čine povoljne uvjete za proglašenjem kraja epidemije COVID-19 u Hrvatskoj.

Dodatno, iz same definicije epidemije koja kaže da se epidemija odnosi na pojavu značajno većeg broja slučajeva određene bolesti unutar određene populacije, zajednice ili geografskog područja, nego što se uobičajeno očekuje, te da je karakterizira brzo širenje bolesti koje zahvaća velik broj pojedinaca u relativno kratkom vremenskom razdoblju [14], vidljivo je da temelji za proglašenjem prestanka epidemije u RH postoje, budući da trenutne brojke ne odgovaraju definiciji epidemije.

Poglavlje 5. Trenutno stanje pandemije (2023.)

S druge strane, treba ostati na oprezu zbog novih varijanti virusa te nerazumijevanja dugoročnih učinaka virusa. Zaključak bi bio da je odluka Vlade RH o proglašenju kraja epidemije COVID-19 u Hrvatskoj opravdana, uz ostanak na oprezu i pripravnost na buduće prijetnje zdravlju.

Poglavlje 6

Analiza modela predviđanja tijekom COVID-19 pandemije

6.1 Analiza prosječne točnosti modela za predviđanja COVID-19 u Europi

U ovom odjeljku bit će predstavljena analiza prosječne točnosti različitih modela predviđanja korištenih za predviđanje broja slučajeva i smrtnih slučajeva uzrokovanih COVID-19 u Europi. Analiza obuhvaća sveobuhvatnu evaluaciju tjednih predviđanja u razdoblju od 5. travnja 2021. do 6. ožujka 2023., pružajući uvide u performanse različitih modela u različitim vremenskim rasponima, predviđajućim horizontima i lokacijama.

Cilj je izračunati prosječnu točnost i pouzdanost ovih modela u predviđanju povećanja slučajeva zaraze i smrtnih slučajeva za različite vremenske horizonte.

Evaluacija modela predviđanja zaraza i smrtnih slučajeva zajedno rezultirala je sljedećim prosječnim metrikama:

- **Prosječna apsolutna pogreška (*engl.* MAE):** Prosječna apsolutna pogreška modela iznosi 3.76. To ukazuje na dobru razinu točnosti, jer su predviđanja, u prosjeku, odstupala od stvarnih vrijednosti za otprilike 3.76.
- **Ponderirani intervalni skor (WIS):** Prosječni ponderirani intervalni rezul-

Poglavlje 6. Analiza modela predviđanja tijekom COVID-19 pandemije

tat iznosi 5.39. Ova metrika mjeri kalibraciju i preciznost intervala predviđanja. Dobivena vrijednost sugerira umjerenu kalibraciju i preciznost predviđenih intervala.

- **Udio vrijednosti unutar 50%-tnog intervala predikcije (cov_50):** Udio stvarnih vrijednosti unutar 50%-tnog intervala predviđanja iznosi 0.42. To ukazuje da je, u prosjeku, otprilike 42% stvarnih vrijednosti bilo obuhvaćeno predviđenim intervalom. Idealno, ova vrijednost bi trebala biti oko 0.50 za točnu pokrivenost.
- **Udio vrijednosti unutar 95%-tnog intervala predikcije (cov_95):** Prosječna pokrivenost za 95%-tni interval predviđanja iznosi 0.78. To sugerira da je, u prosjeku, otprilike 78% stvarnih vrijednosti bilo obuhvaćeno predviđenim intervalima. Vrijednost bliska 0.95 je poželjna za točnu pokrivenost.
- **Pristranost (bias):** Modeli su pokazali malu prosječnu pristranost od 0.042. To ukazuje da su predviđanja, u prosjeku, bila nepristrana, bez sistematskog precjenjivanja ili podcjenjivanja stvarnih vrijednosti.
- **Veličina uzorka (n):** Prosječna veličina uzorka korištena u analizi iznosi otprilike 229. To sugerira relativno veliki skup podataka koji pridonosi pouzdanosti i preciznosti evaluacije.
- **Broj lokacija (n_loc):** Prosječan broj uključenih lokacija u analizu iznosi oko 30. To ukazuje na široku geografsku pokrivenost, omogućavajući sveobuhvatnu procjenu modela predviđanja.

Tablica 6.1 prikazuje prosječne vrijednosti metrika točnosti modela predviđanja COVID-19 u Europi. U drugom stupcu, "Zaraze i smrti", su prikazane prosječne vrijednosti metrika izračunate nad svim modelima zajedno, i modelima koji su predviđali zaraze, i modelima koji su predviđali smrti. U trećem stupcu, "Samo zaraze", su prikazane prosječne vrijednosti metrika izračunate nad samo onim modelima koji su predviđali zaraze. Zadnji stupac u tablici predstavlja prosječne vrijednosti metrika koje su bile izračunate nad samo onim modelima koji su predviđali smrti.

Takva analiza provedena je kako bi se utvrdilo koliko su generalno modeli predviđanja bili uspješni, te postoje li razlike u performansama modela kada predviđaju

Poglavlje 6. Analiza modela predviđanja tijekom COVID-19 pandemije

Tablica 6.1 Prosječne vrijednosti metrika točnosti različitih modela predviđanja COVID-19 u Europi

| Metrika | Zaraze i smrti | Samo zaraze | Samo smrti |
|---------|----------------|-------------|------------|
| MAE | 3.76 | 1.13 | 6.27 |
| WIS | 5.39 | 1.09 | 9.49 |
| cov_50 | 42% | 39% | 44% |
| cov_95 | 78% | 77% | 80% |
| bias | 0.042 | 0.071 | 0.014 |
| n | 229 | 229 | 230 |
| n_loc | 30 | 30 | 31 |

zaraze i modela kada predviđaju smrti.

Iz tablice je vidljivo da modeli, kada predviđaju zaraze, postižu dosta dobru razinu točnosti, s prosječnom apsolutnom pogreškom od 1.13. To sugerira da predviđanja, u prosjeku, odstupaju od stvarnih vrijednosti za otprilike 1.13. Također, prosječni ponderirani intervalni rezultat iznosio je 1.09, što ukazuje na dobro kalibrirane i precizne intervale predviđanja.

Nadalje, prosječna pokrivenost vjerojatnosti za 50% intervala predviđanja iznosila je 0.39, što ukazuje da je otprilike 39% stvarnih vrijednosti bilo obuhvaćeno predviđenim intervalom. Slično tome, prosječna pokrivenost vjerojatnosti za 95% intervala predviđanja iznosila je 0.77, što sugerira da je otprilike 77% stvarnih vrijednosti bilo obuhvaćeno predviđenim intervalom. Idealno, ove vrijednosti trebaju biti bliže 0.50 i 0.95, respektivno, za točnu pokrivenost. Budući da su vrijednosti pokrivenosti dosta manje od nominalnih vrijednosti, prognoze su previše pouzdane, odnosno intervali predviđanja imaju tendenciju da budu preuski.

Modeli, kada su predviđali zaraze, su pokazali i malu prosječnu pristranost od 0.071, što ukazuje da su predviđanja bila neutralna, u prosjeku. Analiza je uključivala prosječnu veličinu uzorka od otprilike 229 i obuhvatila oko 30 lokacija, što doprinosi pouzdanosti i sveobuhvatnoj procjeni modela.

Što se tiče performansi modela kada su predviđali smrtne slučajeve, vidljivo je

Poglavlje 6. Analiza modela predviđanja tijekom COVID-19 pandemije

da su pokazali veću prosječnu apsolutnu pogrešku od 6.27, što ukazuje na dosta veće odstupanje od stvarnih vrijednosti u usporedbi na to kada su predviđali slučajeve zaraze.

Prosječni ponderirani intervalni rezultat iznosio je 9.49, što sugerira manju kalibraciju i preciznost u intervalima predviđanja u usporedbi s predviđanjem slučajeve zaraze.

Prosječna pokrivenost vjerojatnosti za 50% intervala predviđanja iznosila je 0.44, što ukazuje da je otprilike 44% stvarnih vrijednosti bilo obuhvaćeno predviđenim intervalom. Prosječna pokrivenost vjerojatnosti za 95% intervala predviđanja iznosila je 0.80, što sugerira da je otprilike 80% stvarnih vrijednosti bilo obuhvaćeno predviđenim intervalom. Ove vrijednosti idealno bi trebale biti bliže 0.50 i 0.95 za točnu pokrivenost, no svejedno su bolje u usporedbi s predviđanjem slučajeve zaraze.

Modeli su pokazali nešto veću prosječnu pristranost od 0.014, što ukazuje na neznatno sistematsko precjenjivanje stvarnih vrijednosti. Analiza je uključivala prosječnu veličinu uzorka od otprilike 230 i obuhvatila oko 31 lokaciju, osiguravajući pouzdanu evaluaciju u različitim regijama.

Analizom je utvrđeno da modeli predviđanja ostvaruju različite rezultate kada predviđaju zaraze i kada predviđaju smrti. Modeli kada predviđaju zaraze, predviđaju ih s dobrom razinom točnosti i prilično dobro kalibriranim intervalima predviđanja. Međutim, kod predviđanja smrtnih slučajeva pokazuju veće pogreške i manje točne kalibracije u intervalima predviđanja.

Ako je glavni fokus predviđanje slučajeva zaraze, ovi modeli se mogu smatrati prihvatljivima jer pružaju relativno točna predviđanja s dobro kalibriranim intervalima. Međutim, ako je cilj predviđanje smrtnih slučajeva, može biti potrebno istražiti alternativne modele ili poboljšati postojeće modele kako bi se poboljšala njihova točnost i kalibracija.

Važno je uzeti u obzir specifične zahtjeve, kontekst i prihvatljive razine pogreške za predviđeni zadatak. Modeli pokazuju umjereni stupanj točnosti, malu pristranost i koriste relativno veliki skup podataka koji obuhvaća široko geografsko područje, što doprinosi njihovoj pouzdanosti. Međutim, niže vjerojatnosti pokrivenosti sugeriraju moguću nedovoljnu pouzdanost u predviđanjima i šire intervale predviđanja nego što

je potrebno.

Prije donošenja odluke, preporučuje se temeljito procijeniti modele u skladu s posebnim potrebama i ograničenjima predviđenog zadatka. Potrebno je razmotriti ravnotežu između točnosti i kalibracije intervala te utvrditi jesu li nešto niže vjerojatnosti pokrivenosti prihvatljive za primjenu. Ako su potrebne više razine točnosti i užji intervali predviđanja, može biti potrebno daljnje istraživanje ili poboljšanje modela.

6.2 Analiza točnosti i pouzdanosti pojedinih modela za predviđanje širenja virusa

U ovoj analizi procjenjuje se točnost i pouzdanost različitih prediktivnih modela analizirajući njihove metrike izvedbe.

Pregled modela po metrikama prikazan je u tablicama 6.2, 6.3, 6.4, 6.5 i 6.6 gdje su prikazane, za svaki model pojedinačno, prosječne vrijednosti apsolutnih pogrešaka, prosječni ponderirani intervalni skorovi, pokrivenost 50%-tnog i 95%-tnog intervala predikcije i pristranosti, redom.

Najbolji modeli u pojedinoj kategoriji su označeni crvenom bojom, dok su sljedeći najbolji označeni plavom bojom. Neki modeli predviđaju parametre za zaraze, no ne i smrtnost, i obratno, pa su takve nedostajuće vrijednosti u tablicama označene znakom '?'.¹

Iz tablice 6.2 vidljivo je da je, za predviđanje slučajeva zaraze, najbolji model po parametru MAE epiforecasts-EpiExpert_direct, dok su drugi najbolji epiforecasts-EpiExpert_Rt i BIOCOSM-Gompertz. Za predviđanje smrtnih slučajeva to je epiforecasts-EpiExpert_Rt kao najbolji te LANL-GrowthRate kao drugi najbolji.

Poglavlje 6. Analiza modela predviđanja tijekom COVID-19 pandemije

Tablica 6.2 MAE za predviđanje porasta slučajeva zaraze i smrtnih slučajeva

| Model | MAE (predviđanje zaraze) | MAE (predviđanje smrti) |
|-------------------------------|--------------------------|-------------------------|
| AMM-EpiInvert | 0.91 | - |
| BIOCOMSC-Gompertz | 0.64 | 1.06 |
| bisop-seirfilterlite | 1.40 | 1.22 |
| DSMPG-bayes | 1.33 | - |
| epiforecasts-EpiExpert | 0.89 | 0.79 |
| epiforecasts-EpiExpert_direct | 0.60 | - |
| epiforecasts-EpiExpert_Rt | 0.67 | 0.59 |
| epiforecasts-EpiNow2 | 1.26 | 1.16 |
| epiforecasts-weeklygrowth | 1.35 | - |
| EuroCOVIDhub-baseline | 1.00 | 1.00 |
| IEM_Health-CovidProject | 1.22 | 1.11 |
| ILM-EKF | 1.82 | 127.90 |
| Imperial-DeCa | - | 0.97 |
| Imperial-RtI0 | - | 1.11 |
| Imperial-sbcp | - | 1.05 |
| JBUD-HMXK | 1.20 | 2.85 |
| Karlen-pypm | 2.38 | 0.91 |
| LANL-GrowthRate | 0.83 | 0.63 |
| MUNI-ARIMA | 0.74 | 0.77 |
| MUNI-LaggedRegARIMA | - | 0.89 |
| MUNI-VAR | 1.35 | 0.98 |
| RobertWalraven-ESG | 0.99 | 0.85 |
| SDSC_ISG-TrendModel | 0.84 | 0.91 |
| UMass-MechBayes | - | 0.80 |
| UMass-SemiMech | 1.03 | 0.71 |
| USC-SIkJalpha | 1.20 | 0.84 |
| USyd-OneModelMan | - | 0.81 |
| UVA-Ensemble | 1.50 | - |

Poglavlje 6. Analiza modela predviđanja tijekom COVID-19 pandemije

Najbolji model za predviđanje slučajeva zaraze po parametru WIS, sudeći po tablici 6.3, je epiforecasts-EpiExpert_direct, dok je drugi najbolji epiforecasts-EpiExpert_Rt. Za predviđanje smrtnih slučajeva to je LANL-GrowthRate kao najbolji te epiforecasts-EpiExpert_Rt kao drugi najbolji.

Tablica 6.3 WIS za predviđanje porasta slučajeva zaraze i smrtnih slučajeva

| Model | WIS (predviđanje zaraze) | WIS (predviđanje smrti) |
|-------------------------------|--------------------------|-------------------------|
| AMM-EpiInvert | 0.85 | - |
| bisop-seirfilterlite | 1.18 | 1.08 |
| DSMPG-bayes | 1.12 | - |
| epiforecasts-EpiExpert | 0.87 | 0.74 |
| epiforecasts-EpiExpert_direct | 0.75 | - |
| epiforecasts-EpiExpert_Rt | 0.76 | 0.67 |
| epiforecasts-EpiNow2 | 1.26 | 1.18 |
| epiforecasts-weeklygrowth | 1.16 | - |
| EuroCOVIDhub-baseline | 1.00 | 1.00 |
| IEM_Health-CovidProject | 1.16 | 1.03 |
| ILM-EKF | 1.87 | 189.59 |
| Imperial-DeCa | - | 1.14 |
| Imperial-RtI0 | - | 1.19 |
| Imperial-sbcp | - | 1.18 |
| JBUD-HMXK | 1.11 | 1.45 |
| Karlen-pypm | 1.96 | 0.75 |
| LANL-GrowthRate | 0.83 | 0.61 |
| MUNI-ARIMA | 0.83 | 0.79 |
| MUNI-LaggedRegARIMA | - | 0.74 |
| MUNI-VAR | 1.03 | 0.80 |
| RobertWalraven-ESG | 1.05 | 0.86 |
| SDSC_ISG-TrendModel | 0.86 | 1.04 |
| UMass-MechBayes | - | 0.75 |
| UMass-SemiMech | 1.17 | 0.76 |
| USC-SIkJalpha | 1.32 | 0.90 |

Poglavlje 6. Analiza modela predviđanja tijekom COVID-19 pandemije

Tablica 6.4 pokazuje da su modeli koji dobro predviđaju broj zaraženih MUNI-ARIMA, kao najbolji, te JBUD-HMXK i epiforecasts-weeklygrowth. Za predviđanje smrtnih slučajeva to su UMass-MechBayes i epiforecasts-EpiNow2 kao najbolji te RobertWalraven-ESG kao drugi najbolji.

Tablica 6.4 Prosječne 50%-tne vjerojatnosti pokrivanja za slučajeve zaraze i smrtnu slučajevu

| Model | Za predviđanje zaraze | Za predviđanje smrti |
|-------------------------------|-----------------------|----------------------|
| AMM-EpiInvert | 0.39 | - |
| BIOCOMSC-Gompertz | 0.69 | 0.55 |
| bisop-seirfilterlite | 0.33 | 0.44 |
| DSMPG-bayes | 0.47 | - |
| epiforecasts-EpiExpert | 0.27 | 0.33 |
| epiforecasts-EpiExpert_direct | 0.24 | - |
| epiforecasts-EpiExpert_Rt | 0.37 | 0.42 |
| epiforecasts-EpiNow2 | 0.43 | 0.50 |
| epiforecasts-weeklygrowth | 0.52 | - |
| EuroCOVIDhub-baseline | 0.30 | 0.55 |
| IEM_Health-CovidProject | 0.31 | 0.45 |
| ILM-EKF | 0.38 | 0.75 |
| JBUD-HMXK | 0.48 | 0.37 |
| Karlen-pypm | 0.34 | 0.44 |
| LANL-GrowthRate | 0.40 | 0.36 |
| MUNI-ARIMA | 0.49 | 0.55 |
| MUNI-LaggedRegARIMA | - | 0.36 |
| MUNI-VAR | 0.21 | 0.29 |
| RobertWalraven-ESG | 0.29 | 0.49 |
| SDSC_ISG-TrendModel | 0.61 | 0.54 |
| UMass-MechBayes | - | 0.50 |
| UMass-SemiMech | 0.42 | 0.55 |
| USC-SIkJalpha | 0.25 | 0.35 |
| UVA-Ensemble | 0.28 | - |

Poglavlje 6. Analiza modela predviđanja tijekom COVID-19 pandemije

Iz tablice 6.5 je vidljivo da su modeli koji dobro predviđaju zaraze u vidu 95%-tne pokrivenosti epiforecasts-weeklygrowth i DSMPG-bayes kao najbolji te AMM-EpiInvert. Za predviđanje smrtnih slučajeva to su MUNI-ARIMA kao najbolji te ILM-EKF.

Tablica 6.5 Prosječne 95%-tne vjerojatnosti pokrivanja za slučajeve zaraze i smrtnu slučajeve

| Model | Za predviđanje zaraze | Za predviđanje smrti |
|-------------------------------|-----------------------|----------------------|
| AMM-EpiInvert | 0.93 | - |
| BIOCOMSC-Gompertz | 0.98 | 0.91 |
| bisop-seirfilterlite | 0.79 | 0.89 |
| DSMPG-bayes | 0.94 | - |
| epiforecasts-EpiExpert | 0.70 | 0.76 |
| epiforecasts-EpiExpert_direct | 0.52 | - |
| epiforecasts-EpiExpert_Rt | 0.77 | 0.80 |
| epiforecasts-EpiNow2 | 0.80 | 0.91 |
| epiforecasts-weeklygrowth | 0.94 | - |
| EuroCOVIDhub-baseline | 0.79 | 0.91 |
| IEM_Health-CovidProject | 0.70 | 0.86 |
| ILM-EKF | 0.81 | 0.97 |
| JBUD-HMXK | 0.85 | 0.74 |
| Karlen-pypm | 0.72 | 0.89 |
| LANL-GrowthRate | 0.81 | 0.76 |
| MUNI-ARIMA | 0.86 | 0.94 |
| MUNI-LaggedRegARIMA | - | 0.76 |
| MUNI-VAR | 0.52 | 0.68 |
| RobertWalraven-ESG | 0.61 | 0.77 |
| SDSC_ISG-TrendModel | 0.91 | 0.90 |
| UMass-MechBayes | - | 0.90 |
| UMass-SemiMech | 0.85 | 0.90 |
| USC-SIkJalpha | 0.43 | 0.51 |
| UVA-Ensemble | 0.56 | - |

Poglavlje 6. Analiza modela predviđanja tijekom COVID-19 pandemije

Tablica 6.6 Prosječna pristranost za predviđanje porasta slučajeva zaraze i smrtnih slučajeva

| Model | Pristranost (zaraza) | Pristranost (smrtnost) |
|-------------------------------|----------------------|------------------------|
| AMM-EpiInvert | 0.048 | - |
| bisop-seirfilterlite | 0.039 | -0.123 |
| DSMPG-bayes | 0.290 | - |
| epiforecasts-EpiExpert | 0.261 | 0.256 |
| epiforecasts-EpiExpert_direct | 0.439 | - |
| epiforecasts-EpiExpert_Rt | 0.277 | 0.387 |
| epiforecasts-EpiNow2 | -0.091 | 0.130 |
| epiforecasts-weeklygrowth | 0.153 | - |
| EuroCOVIDhub-baseline | 0.009 | 0.037 |
| IEM_Health-CovidProject | 0.104 | 0.052 |
| ILM-EKF | 0.036 | 0.129 |
| Imperial-DeCa | - | -0.232 |
| Imperial-RtI0 | - | -0.140 |
| Imperial-sbcp | - | -0.078 |
| JBUD-HMXK | 0.362 | -0.403 |
| Karlen-pypm | 0.066 | -0.049 |
| LANL-GrowthRate | -0.125 | -0.087 |
| MUNI-ARIMA | 0.006 | 0.059 |
| MUNI-LaggedRegARIMA | - | -0.032 |
| MUNI-VAR | -0.306 | 0.132 |
| RobertWalraven-ESG | -0.197 | -0.163 |
| SDSC_ISG-TrendModel | 0.034 | -0.001 |
| UMass-MechBayes | - | 0.037 |
| UMass-SemiMech | 0.110 | 0.244 |
| USC-SikJalpha | -0.052 | 0.089 |

Tablica 6.6 pokazuje da su modeli koji dobro predviđaju u pogledu pristranosti, modeli s najnižim apsolutnim vrijednostima pristranosti (najbliže nuli).

Poglavlje 6. Analiza modela predviđanja tijekom COVID-19 pandemije

Za predviđanje slučajeva zaraze to su MUNI-ARIMA kao najbolji i EuroCOVIDhub-baseline. Za predviđanje smrtnih slučajeva to su SDSC_ISG-TrendModel kao najbolji te MUNI-LaggedRegARIMA kao drugi najbolji.

Zaključno, u ovom poglavlju provedena je evaluacija različitih modela za predviđanje slučajeva zaraze i smrtnih slučajeva. Analizirane su performanse ovih modela temeljem više metrika evaluacije. Rezultati su otkrili jasne obrasce i istaknuli prednosti različitih modela za određene zadatke predviđanja.

S obzirom na MAE i WIS, model *epiforecasts-EpiExpert_direct* se istaknuo kao najbolji izvođač u predviđanju slučajeva zaraze, pokazujući izvanrednu točnost. Za predviđanje smrtnih slučajeva, model *epiforecasts-EpiExpert_Rt* se istaknuo s niskim MAE i niskim WIS, što ukazuje na njegove snažne prediktivne sposobnosti. Usto, pokazao je zapaženu izvedbu u obje kategorije. Također, kod predviđanja smrtnih slučajeva istaknuo se i *LANL-GrowthRate*.

Evaluacija udjela stvarnih vrijednosti unutar 50%-tnog intervala predikcije pokazala je da model *MUNI-ARIMA* ima najbolju kalibraciju intervala. Za predviđanje smrtnih slučajeva, modeli *UMass-MechBayes* i *epiforecasts-EpiNow2* pokazali su da dosljedno postižu veću vjerojatnost pokrivenosti, što sugerira da su njihovi intervali predviđanja pouzdani.

Što se tiče udjela stvarnih vrijednosti unutar 95%-tnog intervala predikcije, modeli *epiforecasts-weeklygrowth* i *DSMPG-bayes* dokazali su sposobnost pružanja točnih predviđanja unutar višeg intervala predikcije. Slično tome, model *MUNI-ARIMA* je pokazao izuzetnu kalibraciju intervala u predviđanju smrtnih slučajeva.

Osim toga, procjena pristranosti je pokazala da je model *MUNI-ARIMA* pokazao nisku pristranost u predviđanju slučajeva zaraze. Za predviđanje smrtnih slučajeva, model *SDSC_ISG-TrendModel* je pokazao najnižu pristranost.

Sve navedeno sažeto je u tablici 6.7 koja prikazuje pregled najboljih modela predviđanja COVID-19 u Europi za pojedinu metriku. Radi preglednosti i lakše usporedbe najboljih modela, isti model koji se pojavljuje kroz nekoliko kategorija označen je bojom.

Poglavlje 6. Analiza modela predviđanja tijekom COVID-19 pandemije

Tablica 6.7 Najbolji modeli predviđanja COVID-19 u Europi prema pojedinoj metrici

| Metrika | Predviđanje zaraza | Predviđanje smrti |
|--------------|-----------------------------|----------------------|
| MAE | EpiExpert_direct | EpiExpert_Rt |
| WIS | EpiExpert_direct | LANL-GrowthRate |
| 50%-tna v.p. | MUNI-ARIMA | EpiNow2 MechBayes |
| 95%-tna v.p. | DSMPG-bayes weeklygrowth | MUNI-ARIMA |
| Pristranost | MUNI-ARIMA | TrendModel |

*v.p. - vjerojatnost pokrivanja

Zaključno, provedena evaluacija ističe prednosti i nedostatke različitih modela za predviđanje slučajeva zaraze i smrtnih slučajeva. Može se zaključiti da su modeli tima *epiforecasts*, posebno *EpiExpert_direct*, *EpiExpert_Rt*, *weeklygrowth* i *EpiNow2*, kao i modeli tima *MUNI*, posebno *ARIMA*, bili najbolji odabir prema više evaluacijskih metrika. Ovi modeli su pokazali snažnu izvedbu u smislu točnosti, pouzdanosti i kalibracije intervala predikcije pri predviđanju broja slučajeva zaraze i smrti. Stoga, uzimajući u obzir njihovu ukupnu izvedbu, *epiforecasts* modeli i *MUNI* modeli mogu se smatrati među najboljim opcijama za svrhu predviđanja, a posebice predviđanja zaraze.

Što se tiče predviđanja smrtnih slučajeva, tu se kao najbolji ne ističe jedan specifičan model, nego nekoliko njih.

Bitno je imati na umu da odabir najboljeg modela ovisi o specifičnoj evaluacijskoj metrici i prioritetima zadatka predviđanja.

6.3 Opis najuspješnijih modela

Nazivi modela u tablicama pisani su u formatu *tim-model*, gdje tim predstavlja tim koji je taj model kreirao, a model je naziv modela koji su kreirali. Timovi koji su

kreirali ukupno najuspješnije modele su Epiforecasts / London School of Hygiene and Tropical Medicine (u tablicama epiforecasts) i Masaryk University (MUNI). Uz njih će kratko biti opisan i Swiss Data Science Center (SDSC) tim sa Sveučilišta u Ženevi te njihov TrendModel.

6.3.1 Epiforecasts tim

Epiforecasts je istraživačka grupa smještena u Centru za matematičko modeliranje infektivnih bolesti pri London School of Hygiene & Tropical Medicine. Razvijaju metodologiju i alate za *real-time* modeliranje i prognoziranje epidemija infektivnih bolesti. Sustavno procjenjuju svoje metode prema njihovoj prediktivnoj sposobnosti i primjenjuju ih kako bi unaprijedili proces donošenja odluka. Konačni cilj im je razviti i koristiti pouzdanu metodologiju koja maksimalno koristi dostupne resurse kako bi pružila korisne uvide za odgovor na epidemiju, kontrolu i prevenciju te ih učiniti dostupnim drugima u obliku alata. [15]

1. Model **EpiExpert Rt** je model za predviđanje slučajeva zaraze i smrti koji kombinira ljudska predviđanja reprodukcijskog broja (Reproduction number (R_t)) s jednadžbom obnove (*engl.* renewal equation).

R_t je prosječan broj novih infekcija koje uzrokuje jedna zaražena osoba u vremenu t u djelomično osjetljivoj populaciji. EpiExpert Rt model koristi srednju vrijednost kombinirane prognoze ljudskih predviđanja R_t . To znači da se prognoze reprodukcijskog broja prikupljaju od više pojedinaca, uključujući stručnjake i laike.

Prognoze R_t -a se prikupljaju putem ShinyApp, interaktivne web aplikacije u programskom jeziku R. Sudionici se pozivaju da odaberu distribuciju za R_t , pri čemu je zadana vrijednost normalna distribucija. Zatim specificiraju medijan i širinu (nesigurnost) prediktivne distribucije za svaki horizont prognoze. To omogućava pojedincima da izraze svoja predviđanja R_t -a kao vjerojatnosne distribucije, uzimajući u obzir varijabilnost i nesigurnost u njihovim prognozama.

Na temelju pojedinačnih prognoza, generiraju se uzorci iz distribucija speci-

ficiranih od strane svakog sudionika. Ti se uzorci R_t -a koriste za simulaciju slučajeva zaraze i smrti na temelju jednadžbe obnove koja se implementira koristeći R paket EpiNow2. [16]

2. Model **EpiExpert Direct** od prethodnog se razlikuje u tome da se kod njega ne koristi jednadžba obnove, te da pojedinci predviđaju broj slučajeva zaraze i smrti, a ne R_t . Sav ostali postupak je jednak, prognoze se prikupljaju putem ShinyApp-a, pojedinci izrađuju prognoze odabirom distribucije i specificiranjem medijana i širine odabrane prediktivne distribucije za svaki horizont prognoze. Pojedinačne prognoze se prikupljaju, nakon čega se formira srednja vrijednost ljudske predikcije. [16]
3. Model **EpiNow2**, napravljen je za procjenu broja *real-time* slučajeva i epidemioloških parametara koji variraju u vremenu. Izrađen je koristeći statistički programski jezik R. EpiNow2 model kombinira matematičko modeliranje, statističko zaključivanje i računalne tehnike, procjene početnih infekcija, reprodukcijskog broja (*engl.* R_t), inkubacijskih razdoblja i kašnjenja u prijavljivanju slučajeva za procjenu vremenski promjenjivog reprodukcijskog broja (R_t) i izradu predviđanja o slučajevima COVID-19. Uzima u obzir prethodno znanje, ažurira ga promatranim podacima i generira procjene zajedno s nesigurnostima kako bi pružio uvid u širenje virusa. [16]

6.3.2 MUNI tim

Tim MUNI sačinjavaju Andrea Kraus i David Kraus, oboje profesori Zavoda za matematiku i statistiku pri Masaryk University-ju u češkom gradu Brno. Njihov model koristi ARIMA model s detekcijom outliera te se prilagođava transformiranim tjedno agregiranim vremenskim serijama.

Autoregressive integrated moving average (ARIMA) je model koji se često koristi za analizu vremenskih serija i prognoziranje. "AR" dio u kratici "ARIMA", označava da se promjenjiva varijabla od interesa regresira na vlastite prošle vrijednosti. "MA" označava da je pogreška regresije zapravo linearna kombinacija pogrešaka čije su vrijednosti zabilježene trenutno i u različitim vremenima u prošlosti. "I" označava da su vrijednosti podataka zamijenjene razlikom između njihovih vrijed-

nosti i prethodnih vrijednosti (i taj proces diferenciranja može biti proveden više puta). Svrha svakog od ovih svojstava je da model što bolje odgovara podacima. [17]

Osim standardnog ARIMA modeliranja, model uključuje tehnike detekcije outliera. Outlieri su podaci koji značajno odstupaju od općeg obrasca podataka. Detekcija outliera pomaže identificirati i riješiti te ekstremne vrijednosti, što može poboljšati točnost i pouzdanost predviđanja modela. Također, model se prilagođava transformiranim tjedno agregiranim vremenskim serijama. To znači da se izvorni dnevni podaci transformiraju ili agregiraju u tjedne vremenske intervale prije primjene ARIMA modela. [16]

6.3.3 Swiss Data Science Center tim

Swiss Data Science Center tim je tim sa Sveučilišta u Ženevi. Sastavljen je od različitih struka, poput biomatematičara, statističara, podatkovnih znanstvenika (*engl. data scientist*), programera, elektrotehničara itd. Njihov cilj bio je razviti globalno primjenjivu metodu, integriranu u dvaput dnevno ažurirani dashboard, koja pruža procjenu trenda evolucije broja slučajeva zaraze i smrtnih slučajeva na temelju prijavljenih podataka iz više od 200 zemalja i teritorija, kao i sedmodnevno prognoziranje.

Oni smatraju da je jedan od značajnih izazova u upravljanju brzo širećom epidemijom taj da detalji dinamike potrebni za predviđanje njenog razvoja bivaju zasjenjeni kašnjenjem u identifikaciji slučajeva i smrtnosti te neregularnim prijavama. Kako bi tome doskočili, kreirali su **TrendModel** model, čija se metodologija prognoziranja uvelike oslanja na procjenu underlying trenda u promatranom vremenskom nizu koristeći robusne tehnike dekompozicije sezonalnosti trenda, što omogućava dobivanje prognoza pomoću jednostavnih, ali učinkovitih metoda ekstrapolacije na linearnoj ili logaritamskoj skali. [18]

Predviđanja TrendModela temelje se na prijavljenim brojevima slučajeva i smrtnosti na nacionalnoj ili regionalnoj razini. Modeliranje u velikoj mjeri ovisi o procjeni underlying trenda pomoću robusnog LOESS modela dekompozicije sezonalnosti trenda. Takav model omogućava uzimanje u obzir outliere, podatke koji nedostaju i kašnjenja u prijavama. Za daljnje predviđanje dnevnih slučajeva i smrtnosti, model

koristi linearnu ekstrapolaciju procijenjenog zaglađenog trenda, bilo na izvornoj ili logaritamskoj skali.

6.4 Analiza točnosti i pouzdanosti pojedinih modela po polugodištu

Cilj ovakve analize je proučiti ponašanja modela po polugodištima te vidjeti ističu li se pojedini modeli u pojedinim razdobljima, ili su najuspješniji modeli iz prethodnog potpoglavlja također najuspješniji i po razdobljima. Vrijeme je razdijeljeno u 4 razdoblja (približno polugodišta):

1. 5.4.2021. (datum prve dostupne evaluacije) do kraja 2021. (radi jednostavnosti, ovaj period bit će zvan 2021. godina)
2. siječanj do lipanj 2022. (ovaj period bit će zvan 1. dio 2022. godine)
3. srpanj do prosinac 2022. (2. dio 2022. godine)
4. siječanj do 6.3.2023. (zadnji datum prihvatljivo cjelovitih evaluacija, nakon tog datuma neki modeli su prestali s prijavljivanjem evaluacija; radi jednostavnosti, ovaj period bit će zvan 2023. godina)

Modeli će biti uspoređivani po WIS parametru, budući da on najbolje opisuje točnost modela.

U tablici 6.8 prikazane su prosječne vrijednosti parametra WIS za predviđanje porasta slučajeva zaraze i smrtnih slučajeva kroz sva 4 perioda. Iz prvog dijela tablice je vidljivo da je u prvom periodu (2021. godina) najbolji model za predviđanje slučajeva zaraze bio MUNI-ARIMA, dok je sljedeći najbolji bio LANL-GrowthRate, koji je ujedno bio i najbolji za predviđanje smrtnih slučajeva. MUNI-VAR bio je drugi najbolji.

Što se tiče prvog dijela 2022. godine, najbolji model u predviđanju zaraza bio je SDSC_ISG-TrendModel, a pratio ga je MUNI-ARIMA. Što se predviđanja smrtnih slučajeva tiče, najbolji je bio MUNI-LaggedRegARIMA, dok je drugi najbolji bio epiforecasts-EpiExpert.

Poglavlje 6. Analiza modela predviđanja tijekom COVID-19 pandemije

Tablica 6.8 WIS za predviđanje porasta slučajeva zaraze i smrtnih slučajeva po periodima

| Period | Model | WIS (zaraze) | WIS (smrti) |
|------------------|------------------------|--------------|-------------|
| 2021. | MUNI-ARIMA | 0.81 | 0.76 |
| | LANL-GrowthRate | 0.82 | 0.59 |
| | MUNI-VAR | 1.07 | 0.67 |
| 2022. (1.dio) | SDSC_ISG-TrendModel | 0.82 | 0.92 |
| | MUNI-ARIMA | 0.89 | 0.80 |
| | MUNI-LaggedRegARIMA | - | 0.70 |
| | epiforecasts-EpiExpert | 0.94 | 0.77 |
| 2022. (2. dio) | AMM-EpiInvert | 0.79 | - |
| | MUNI-ARIMA | 0.87 | 0.86 |
| | RobertWalraven-ESG | 1.01 | 0.89 |
| 2023. | SDSC_ISG-TrendModel | 0.86 | 0.97 |
| | AMM-EpiInvert | 0.88 | - |
| | MUNI-ARIMA | 0.90 | 0.86 |

U drugom dijelu te iste godine (2022.), najbolji u predviđanju zaraza bio je AMM-EpiInvert, drugi najbolji MUNI-ARIMA koji je ujedno bio i najbolji kod predviđanja smrtnih slučajeva. Drugi najbolji bio je RobertWalraven-ESG.

U 4. periodu, 2023. godini, model SDSC_ISG-TrendModel je najbolji u predviđanju zaraza i drugi najbolji u predviđanju smrti, dok je MUNI-ARIMA najbolja u predviđanju smrti i treća najbolja u predviđanju slučajeva. Tu je također značajan i AMM-EpiInvert koji je drugi u predviđanju slučajeva.

Usporedbom sva 4 perioda dolazi se do zaključka da je MUNI-ARIMA uvijek u samom vrhu kod predviđanja, bilo zaraza, bilo smrti. Zanimljivo je što takvi rezultati ne odgovaraju i rezultatima analize WIS-a iz prethodnog potpoglavlja, vidljivim u tablici 6.3. Iako MUNI-ARIMA ne stoji loše u ukupnom poretku po kriteriju WIS, nije u samom vrhu, no u samom vrhu je po drugim metrikama.

U analizi po periodima, uz MUNI-ARIMA model, još su se istaknuli i SDSC_ISG-

Poglavlje 6. Analiza modela predviđanja tijekom COVID-19 pandemije

TrendModel te AMM-EpiInvert modeli. SDSC_ISG-TrendModel dobar je na oba područja predviđanja, dok je AMM-EpiInvert specijaliziran za predviđanja zaraze. Oni su u prethodnim analizama također bili u vrhu prema nekim metrikama, no ne koliko i MUNI-ARIMA model.

Nakon provedene analize točnosti modela tijekom različitih vremenskih perioda, zaključeno je da modeli koji su najbolji u pojedinom periodu, nisu i najbolji generalno. Generalno najboljim modelima mogli bi se smatrati modeli tima Epiforecasts, no u ovoj analizi su oni gotovo nezastupljeni. Jedan mogući razlog za to može biti nekonzistentno prijavljivanje prognoza od strane timova. Neki timovi su redovito prijavljivali prognoze tijekom cijelog vremenskog perioda (2021. - 2023.), dok su drugi to radili povremeno, neki su to radili redovito ali u kraćem vremenskom periodu, a neki su čak prestali s prijavljivanjem prognoza. Također, neki su timovi tek počeli s radom nakon određenog perioda. Modeli koji su se istaknuli u određenom vremenskom periodu vjerojatno su sustavno prijavljivali kvalitetne prognoze tijekom tog razdoblja. S druge strane, modeli koji su općenito bolji, ali nisu dominirali u analizi kroz periode, vjerojatno su bili konzistentni u prijavljivanju prognoza kroz duže vrijeme sa zadovoljavajućim rezultatima.

Poglavlje 7

Zaključak

U ovom diplomskom radu provedena je detaljna analiza podataka o pandemiji COVID-19 virusa s ciljem razumijevanja njegove prirode, dinamike i utjecaja na globalno zdravlje i društvo. Kroz usporedbu podataka iz 2020., 2021. i 2022. godine, analizu trenutnog stanja pandemije u Hrvatskoj i svijetu te evaluaciju modela predviđanja, dobiveni su važni uvidi i spoznaje o trajajućoj pandemiji.

Na temelju analize podataka o broju potvrđenih slučajeva zaraze, smrtnih slučajeva i cijepljenja, primijećene su značajne promjene i varirajući trendovi tijekom protekle četiri godine. Broj potvrđenih slučajeva i smrtnih slučajeva u početku je bio u porastu diljem svijeta, što ukazuje na širenje i ozbiljnost virusa. Situacija se stabilizirala tek u 2023. godini. Također je primijećeno da su neke regije, poput Azije, Europe i Južne Amerike, doživjele veći utjecaj pandemije u usporedbi s drugim kontinentima poput Afrike, Sjeverne Amerike i Oceanije.

Tijekom pandemije, može se nagađati da su primijenjene mjere javnog zdravstva, uključujući zatvaranja, nošenje maski i kampanje cijepljenja, imale utjecaj na suzbijanje širenja virusa. Međutim, uočene su i razlike u pristupima i učinkovitosti tih mjera među različitim regijama i populacijama. To ukazuje na potrebu za daljnjim istraživanjem i evaluacijom strategija i politika javnog zdravstva. Navedeni rezultati su dobiveni samo pomoću korelacije, no ne i ispitivanjem uzročnosti, pa se ne može reći da između neke dvije varijable postoji uzročno-posljedična veza. Za donošenje čvrstih zaključaka o uzročno-posljedičnoj vezi između dviju varijabli, trebali bi

Poglavlje 7. Zaključak

se provesti eksperimenti koji bi omogućili kontrolu potencijalnih faktora koji mogu utjecati na rezultate.

Ova analiza također je obuhvatila modeliranje i predviđanje COVID-19 brojki. Evaluacija tih modela pokazala je njihovu korisnost u predviđanju širenja virusa, ali i potrebu za kontinuiranim praćenjem i ažuriranjem kako bi se osigurala njihova točnost i pouzdanost. Daljnje poboljšanje modela predviđanja bit će ključno za bolje razumijevanje pandemije i donošenje informiranih odluka.

U zaključku, ova studija pruža sveobuhvatnu analizu podataka o pandemiji COVID-19 iz više izvora i vremenskih razdoblja. Rezultati istraživanja ukazuju na ozbiljnost i globalni utjecaj pandemije, ali isto tako naglašavaju potrebu za daljnjim praćenjem, istraživanjem i intervencijama.

Odgovarajući na postavljena istraživačka pitanja, ovo istraživanje doprinosi razumijevanju trendova, uzoraka i promjena vezanih uz pandemiju COVID-19. Informacije iz ovog rada mogu biti korisne donositeljima politika, zdravstvenim stručnjacima i istraživačima kako bi se informirale i oblikovale buduće strategije za borbu protiv pandemije, sprječavanje širenja virusa i zaštita javnog zdravlja.

Važno je napomenuti da su rezultati ove analize zasnovani na dostupnim podacima do svibnja 2023. godine, te da se situacija vezana uz pandemiju COVID-19 može dalje mijenjati. Stoga je potrebno kontinuirano praćenje i ažuriranje podataka kako bi se osigurala relevantnost i primjenjivost dobivenih rezultata.

U budućnosti, preporučuje se daljnje istraživanje koje će proširiti i produbiti razumijevanje COVID-19 pandemije. Tema je izuzeto opsežna, i razne stvari se mogu analizirati, primjerice analiza COVID-19 prema varijantama virusa, odgovor na pitanja kako je virus utjecao na ekonomska stanja država i mentalna stanja građana, analiza količine lažnih vijesti (*engl.* fake news) proizvedenih tijekom pandemije, analiza porasta nasilja u obitelji za vrijeme lockdown-a, analiza školskih uspjeha djece koja su određeno vrijeme nastavu pratila od kuće, analiza porasta mentalnih bolesti među populacijom, pogotovo mlađom itd., može pružiti sveobuhvatniju sliku pandemije i njenih implikacija. Također je važno pratiti i evaluirati različite strategije javnog zdravstva kako bi se identificirale najučinkovitije intervencije za kontrolu i suzbijanje pandemije.

Poglavlje 7. Zaključak

U konačnici, samo multidisciplinarni pristup, koji uključuje znanstvenu zajednicu, donositelje politika i javnost, može osigurati cjelovit i učinkovit odgovor na pandemiju COVID-19 i slične buduće izazove.

Bibliografija

- [1] H. Wickham and L. Stryjewski, “40 years of boxplots,” Nov. 2011.
- [2] Wikipedia. (2023, 5) Covid-19. Wikimedia Foundation. , s Interneta, <https://hr.wikipedia.org/wiki/COVID-19> , lipanj 2023.
- [3] U. News. (2023, May) Who chief declares end to covid-19 as a global health emergency. United Nations. , s Interneta, <https://news.un.org/en/story/2023/05/1136367> , svibanj 2023.
- [4] V. RH. (2023, May) Plenković: Danas proglašavamo kraj epidemije koronavirusa u hrvatskoj. Vlada Republike Hrvatske. , s Interneta, <https://vlada.gov.hr/vijesti/plenkovic-danas-proglasavamo-kraj-epidemije-koronavirusa-u-hrvatskoj/38294> , svibanj 2023.
- [5] E. Mathieu, H. Ritchie, L. Rodés-Guirao, C. Appel, C. Giattino, J. Hassell, B. Macdonald, S. Dattani, D. Beltekian, E. Ortiz-Ospina, and M. Roser, “Coronavirus pandemic (covid-19),” *Our World in Data*, 2020, <https://ourworldindata.org/coronavirus>.
- [6] Reports. European COVID-19 Forecast Hub. , s Interneta, <https://covid19forecasthub.eu/reports.html> , lipanj 2023.
- [7] (2023) European Centre for Disease Prevention and Control. , s Interneta, <https://www.ecdc.europa.eu/en> , lipanj 2023.
- [8] D. C. Howell, *Statistical methods for psychology*, 8th ed. Cengage Learning, 2012.
- [9] M. Udovičić, K. Baždarić, L. Bilić-Zulle, and M. Petrovečki, “Što treba znati kada izračunavamo koeficijent korelacije?” *Biochemia Medica*, vol. 17, no. 1, pp. 10–15, 2007.

Bibliografija

- [10] A. Petrie and C. Sabin, *Medical Statistics at a Glance*, 2nd ed. Blackwell Publishing Ltd, 2005.
- [11] J. Bracher, E. L. Ray, T. Gneiting, and N. G. Reich, “Evaluating epidemic forecasts in an interval format,” *PLOS Computational Biology*, vol. 17, no. 2, p. e1008618, feb 2021.
- [12] D. Beroš, M. Čulav Markičević, Z. Lobar, and I. Martinić. Logaritamska skala. , s Interneta, <https://hrcak.srce.hr/file/424978> , lipanj 2023.
- [13] M. Davey. (2022, Dec.) Was 2022 the year that australia came to terms with covid? and what does 2023 hold? The Guardian. , s Interneta, <https://www.theguardian.com/culture/2022/dec/19/was-2022-the-year-that-australia-came-to-terms-with-covid-and-what-does-2023-hold> , lipanj 2023.
- [14] H. platforma za smanjenje rizika od katastrofe, “Epidemije i pandemije,” lipanj 2023. , s Interneta, https://civilna-zastita.gov.hr/UserDocsImages/CIVILNAZAŠTITA/PDF_ZAWEB/Epidemijeipandemije_brošuraA5-web.pdf
- [15] Epiforecasts. Centre for Mathematical Modelling of Infectious Diseases. , s Interneta, <https://epiforecasts.io/> , lipanj 2023.
- [16] S. Abbott, J. Hickson, S. Funk, Hamada S. Badr, P. Monticone, P. Ellis, J. Munday, J. Allen, C. A. B. Pearson, L. Chapman, M. DeWitt, N. Bosse, and S. Meakin, “Epinow2: Estimate real-time case counts and time-varying epidemiological parameters,” 2023. , s Interneta, <https://github.com/epiforecasts/EpiNow2/tree/main>
- [17] Wikipedia. Autoregressive integrated moving average. Wikimedia Foundation. , s Interneta, https://en.wikipedia.org/wiki/Autoregressive_integrated_moving_average , lipanj 2023.
- [18] E. Krymova, B. Béjar, D. Thanou, T. Sun, E. Manetti, G. Lee, K. Namigai, C. Choirat, A. Flahault, and G. Obozinski, “Trend estimation and short-term forecasting of covid-19 cases and deaths worldwide,” *Proceedings of the National Academy of Sciences*, vol. 119, no. 32, aug 2021.

Pojmovnik

R_t Reproduction number. 64, 65

AE Absolute Error. 15

ARIMA Autoregressive integrated moving average. 65, 66

COVID-19 COronaVIRus Disease of 2019. 1–17, 19, 24–27, 30, 32, 38, 40, 41, 43, 45, 48, 50–53, 62, 65, 70–72, 76

ECDC European Centre for Disease Prevention and Control. 7

MAE Mean Absolute Error. xii, 15, 52, 54, 56, 57, 62, 63

OWID Our World In Data. 6

WHO World Health Organization. 4, 41, 45

WIS Weighted Interval Score. xii, 14, 15, 58, 62, 63, 67, 68

Sažetak

Ovaj diplomski rad pruža analizu podataka o COVID-19 s ciljem razumijevanja pandemije. Istraživanje je obuhvatilo usporedbu podataka iz 2020., 2021. i 2022. godine te analizu trenutnog stanja pandemije u Hrvatskoj i svijetu u 2023. godini. Rezultati su otkrili varijacije u broju slučajeva, smrtnosti i stopama cijepljenja tijekom vremena. Evaluacija modela predviđanja pružila je uvide u njihovu točnost i pouzdanost. Zaključci ističu potrebu za kontinuiranim praćenjem podataka i informiranim odlukama temeljenim na njima.

Ključne riječi — COVID-19, analiza podataka, trendovi, točnost modela, pandemija

Abstract

This master's thesis provides an analysis of COVID-19 data with the aim of understanding the pandemic. The research included a comparison of data from 2020, 2021, and 2022, as well as an analysis of the current state of the pandemic in Croatia and the world in 2023. The results revealed variations in the number of cases, mortality rates, and vaccination rates over time. The evaluation of prediction models provided insights into their accuracy and reliability. The conclusions highlight the need for ongoing data monitoring and informed decision-making based on them.

Keywords — COVID-19, data analysis, trends, model accuracy, pandemic