ივანე ჯავახიშვილის სახელობის თბილისის სახელმწიფო
უნივერსიტეტის ეკონომიკისა და ბიზნესის ფაკულტეტი

ლევან გაფრინდაშვილი

საქართველოში კორონავირუსის გავრცელების პროგნოზირება (ეკონომეტრიკული და მანქანური სწავლების მეთოდები)

სამაგისტრო პროგრამა ეკონომიკა

ნაშრომი შესრულებულია ეკონომიკის მაგისტრის აკადემიური ხარისხის მოსაპოვებლად

ხელმძღვანელი: პროფესორი იური ანანიაშვილი
ეკონომიკის მეცნიერებათა დოქტორი,
ეკონომეტრის კათედრის ხელმძღვანელი

თბილისი 2020

ანოტაცია

2019 წლის დეკემბერში ჩინეთში კორონავირუსის პირველი კერის გაჩენის შემდეგ სხვადასხვა სახის მოდელი იქნა შემუშავებული და გამოყენებული ეპიდემიის პროგნოზირებისათვის. ნაშრომი ეხება კოვიდ 19-ის გავრცელების პროგნოზირებასა და სახელმწიფოს მიერ გატარებული პრევენციული ღონისძიებების შეფასებას. კვლევა განხორციელდა საქართველოს მაგალითზე. პროგნოზირებისათვის გამოყენებული იქნა ეპიდემიოლოგიური მოდელები, ეკონომეტრიკული დროითი მწკრივების კლასიკური მოდელები და ასევე მანქანური სწავლების მოდელი. გარდა ამისა, ეპიდემიის გავრცელების ხასიათისა და ტრენდის გათვალისწინებით, ჩვენს მიერ შემოთავაზებულ იქნა პოლინომიალური მოდელი მოდიფიცირებული არქიტექტურით. გარდა ამისა, ნაშრომში განხილულია ეპიდემიის რამდენიმე პარამეტრი/კოეფიციენტი. მათ შორის უმნიშვნელოვანესია რეპროდუქციის რიცხვი, რომლის მოდელირება რამდენიმე მეთოდით განვახორციელეთ. მიღებული შედეგები არაიდენტური, თუმცა გარკვეული გარემოებების გათვალისწინებით შესადარისი იყო.

კვლევის შედეგებმა აჩვენა, რომ პროგნოზირებისათვის იდეალური მოდელის აგება რთულია. კონკრეტულ მოდელს გააჩნია თავისი უპირატესობები. კერძოდ, ფენომენოლოგიური მოდელები კარგად პროგნოზირებენ ეპიდემიის ტრენდს, ხანგრძლივობასა და ინფიცირებულთა ჯამურ რაოდენობას, თუმცა დღიური შემთხვევების პროგნოზირებისას დიდ შეცდომას უშვებენ. საპირისპიროდ, მანქანური სწავლების მოდელები მოკლევადიან პერიოდში შედარებით ზუსტ დღიური ინფიცირების პროგნოზს აკეთებენ, თუმცა მონაცემების შეზღუდულობის გამო, გრძელვადიანი პროგნოზირება უჭირთ. სახელმწიფოს მიერ გატარებული შეკავების ღონისძიებების მოდელირებისათვის და ოპტიმალური შეზღუდვის დონის განსაზღვრისათვის კი საუკეთესო არჩევანს განყოფილებიანი მოდელები წარმოადგენენ.

Annotation

Since the first outbreak of the epidemic in China in December 2019, various models have been developed and used to predict coronavirus. The paper focuses on predicting the spread of COVID 19 and evaluating preventive measures taken by the government. The study was conducted by employing Georgian coronavirus data. Epidemiological models, classical econometric time series models, and also machine learning models were used for forecasting. Furthermore, considering the nature and trend of the spread of the epidemic, we have proposed a polynomial model with modified architecture. The paper also discusses several parameters/ratios of the epidemic. The most important one, the number of reproductions, was modeled with several approaches. The result was not identical but comparable to certain circumstances.

The results of the study showed that it is difficult to construct an ideal model for prediction. Each model has its advantages. In particular, phenomenological models well predict the epidemic trend, duration, and the total number of infected people, although they make a big mistake when predicting daily cases. In contrast, machine learning models predict relatively accurate daily infection in the short term, although due to data limitations, it is difficult to predict in the long run. Furthermore, compartmental models are the best choice for modeling state-controlled restrictive measures and determining the optimal level of restraint.

შინაარსი

შესავალი

2019 წლის დეკემბერს ჩინეთში, ჰუბეის პროვინციაში, გავრცელდა ახალი მწვავე რესპირატორული სინდრომის ტიპის ინფექცია, კორონავირუსი-2 (SARS-CoV-2), რომელიც კოვიდ-19-ის (COVID-19) სახელით გახდა ცნობილი და ორი თვის შემდეგ მთელი მსოფლიო მოიცვა. გაურკვევლობისა და არეულობის თვენახევრიანი პერიოდის შემდეგ, ჩინეთმა მთლიანი ქვეყნის მასშტაბით უპრეცედენტო „შეკავების ღონისძიებები“ განახორციელა, მათ შორის ჰუბეის პროვინციის ჩაკეტვა და უმეტეს პროვინციებში საკარანტინო რეჟიმის გამოცხადება. 2020 წლის მარტში, რადიკალური ღონისძიებების გატარებიდან ექვსი კვირის შემდეგ, ჩინეთმა წარმატებით შეაჩერა ვირუსის გავრცელება და დღიური ინფიცირების რაოდენობა ორციფრიან ნიშნულებამდე შემცირდა. ჩინეთის წარმატების მიუხედავად, მსოფლიოს დანარჩენ ქვეყნებში ინფექციის გავრცელება შეუქცევადად მიმდინარეობდა. ჯანდაცვის საერთაშორისო ორგანიზაციამ მას პანდემიის სტატუსი მიანიჭა. მდგომარეობას ართულებდა ვირუსის საწინააღმდეგო პრეპარატის არარსებობა და მისი გავრცელების სიმარტივე. დაავადებიდან გამოწვეული ზიანი თანაბარი სიმკაცრით აისახა ადამიანთა ჯანმრთელობასა და ეკონომიკაზე. პრეპარატის არარსებობის გამო ვირუსთან ბრძოლის ერთადერთ საშუალებად იქცა სოციალური დისტანცირება. შეჩერდა საერთაშორისო მიმოსვლა, შეიზღუდა ვაჭრობა, საზოგადოებრივი ტრანსპორტით სარგებლობა, საგანმანათლებლო დაწესებულებები და სხვ., რამაც მკვეთრად გაზარდა უმუშევრობა და დიდი ზიანი მიაყენა ეკონომიკას.

ცალკეული სამთავრობო თუ კომერციული ორგანიზაციებისათვის დიდი მნიშვნელობა შეიძინა პანდემიის გავრცელების პროგნოზირებამ ე.წ. დაბრუნების გეგმის (roll back plan) შედგენისათვის და სახელმწიფოს მიერ განხორციელებული ცალკეული ღონისძიებების ეფექტების შეფასებამ, შეზღუდვის სარგებლისა და დანაკარგების ოპტიმიზაციისათვის. დაავადების მოსალოდნელი მასშტაბის ცოდნა მნიშვნელოვანია სამედიცინო ეკიპირებისა და მედიკამენტების მარაგის ზომის განსაზღვრისათვის, საჭიროების შემთხვევაში ახალი სამედიცინო დაწესებულებების მშენებლობისათვის (როგორც ეს ჩინეთში განხორციელდა). შეკავების პოლიტიკის ალტერნატიული სცენარების მოდელირება საშუალებას იძლევა ერთმანეთს

შეუდარდეს დანაკარგები და მოხდეს შეკავების ზომის სწორად განსაზღვრა. სამეცნიერო სფეროში, აგრეთვე, დიდი ძალისხმევა იქნა მიმართული საბაზისო რეპროდუქციის რიცხვის (basic reproduction number R_0) შეფასებისათვის და ვირუსის ტრაექტორიის განსასაზღვრად, რაც გიჩვენებს დაავადების გადადებისუნარიანობას (transmissibility). საკითხის აქტუალურობიდან გამომდინარე, პროგნოზირებამ დიდი ინტერესი გამოიწვია ეკონომეტრიკოსთა და მონაცემთა მეცნიერთა რიგებში. აღნიშნულ თემასთან დაკავშირებით, მონაცემთა ანალიზისა და პროგნოზირებისათვის განკუთვნილ ვებ-გვერდზე „kaggle.com”-ზე შეიქმნა კონკურსები. სტატისტიკური პროგრამის R-ის გუნდმა სპეციალურად კორონავირუსისათვის განავითარა რამდენიმე პაკეტი. კორონავირუსის პროგნოზირებას ასევე მრავალი ნაშრომი მიეძღვნა www.medrxiv.org-ზე.

მოცემული კვლევის მიზანს წარმოადგენს მოდელის აგება, რომელიც მოახდენს კორონავირუსის გავრცელების პროგნოზირებას, შესაძლებლობის ფარგლებში, მაქსიმალური სიზუსტით.

კვლევის მიზნიდან გამოდინარე რამდენმე ამოცანაა გადასაწყვეტი. პირველ რიგში, საჭიროა ზუსტი მონაცემების მოძიება და დამუშავება. შემდეგ უნდა მოხდეს ეპიდემიის მოდელირებისათვის ლიტერატურაში გავრცელებული მოდელების შესწავლა და მათი მორგება ჩვენი ქვეყნის მონაცემებზე. მოდელების სპეციფიკაცია უნდა მოხდეს მონაცემების სტრუქტურისა და ქვეყნის ეპიდ-პოლიტიკისა თუ დემოგრაფული თავისებურებიდან გამომდინარე. ასევე უნდა შეირჩეს მეტრიკა, რომლითაც მოხდება მოდელის სიზუსტის შეფასება. საბოლოოდ, საუკეთესო მოდელი უნდა განისაზღვროს შერჩეული მეტრიკის მიხედვით.

მოდელების აგებისათვის ძირითადად გამოყენებულია სტატისტიკური პროგრამა R-ის სხვადასხვა პაკეტი. გამონაკლისია ნეირონული ქსელები, რომლის მოდელირებისათვის გამოყენებული Python-ის keras ბიბლიოთეკა.

მონაცემები აღებულია github.com-დან, რომელიც შეგროვებულია ჯონ ჰოფკინსის უნივერსიტეტის მიერ [47].

მიუხედავად იმისა, რომ განვითარებულ ქვეყნებში ეკონომეტრიკის მანქანური სწავლების მეთოდები აქტიურად გამოიყენება სამედიცინო სფეროს მონაცემების

ანალიზისათვის, საქართველოში ამ კუთხით დიდი დეფიციტია. ამაზე მეტყველებს ის ფაქტიც, რომ ქვეყანაში სამეცნიერო კვლევის მიზნებისათვის ხელმისაწვდომი არ არის მაღალი ხარისხის სამედიცინო შინაარსის მონაცემები. აქედან გამომდინარე, კვლევა ფაქტობრივად სიახლეს წარმოადგენს ქართულ სამედიცინო-ეკონომეტრიკულ (health econometrics) სივრცეში. კვლევის სიახლეს წარმოადგენს აგრეთვე პოლინომიალური მოდელი, რომლის სპეციფიკაცია მოხდა ჩემს მიერ ეპიდემიის ხასიათისა და კვლევის მიზნის გათვალისწინებით. მოცემულ ნაშრომი შედგება ექვსი ნაწილისაგან. პირველი ნაწილი ეძღვნება კორონავირუსთან და სხვა მწვავე რესპირატორული დაავადებების პროგნოზირებასთან დაკავშირებულ ლიტერატურის მიმოხილვას.

მეორე ნაწილში განვიხილავთ ე.წ. განყოფილებიან მოდელებს, რომლებიც აქტიურად გამოიყენება ეპიდემიების მოდელირებისათვის. თეორიულად აღვწერთ SIR და SEIR მოდელებს და შევაფასებთ SIR სიზუსტეს საქართველოს მონაცემების საშუალებით. ასევე მიმოვიხილავთ გაფართოებულ SIR მოდელს, სადაც გათვალისწინებულია საქართველოს მთავრობის მიერ განხორციელებული სხვადასხვა ღონისძიებების ეფექტები.

მესამე თავში განვიხლავთ ე.წ. S ფორმის (S shape) მრუდებს და მათი საშუალებით განვახორციელებთ ეპიდემიის გავრცელების პროგნოზირებას. მიმოვიხილავთ ლოგისტიკური ზრდის მოდელს, გომპერსის მოდელს, და რიჩარდის მოდელს. ვისაუბრებთ თითოეული მოდელის უპირატესობასა და ნაკლოვანებაზე და მოვახდენთ მათი სიზუსტის შედარებას. მესამე თავში, ასევე, ავღწერთ და გავიანგარიშებთ საბაზისო და ეფექტურ რეპროდუქციის რიცხვს სხვადასხვა მეთოდის საშუალებით.

მეოთხე თავი ეთმობა დროითი მწკივების პროგნოზირების კლასიკური ეკონომეტრიკული მეთოდების განხილვას, როგორიცაა ავტორეგრესიული მცურავი საშუალოს მოდელი (ARIMA). აგრეთვე მიმოვიხილავთ ჩვენს მიერ შემოთავაზებულ პოლინომიალურ მოდელს, სადაც ქვეყანაში ინფექციის დაწყებიდან T-ე დღეს ინფიცირებულთა რაოდენობის პროგნოზირება ხდება T-ის მეორე რიგის პოლინომისა და ამ დღეს ცალკეული ქვეყნების მიხედვით დათვლილი ინფიცირების ზრდის ტემპების საშუალოსა და მედიანას საშუალებით.

მეხუთე თავში მიმოვიხილავთ მანქანური სწავლების თანამედროვე მეთოდს რეკურენტული ნეირონული ქსელების ერთ-ერთ სახეობას: მოკლევადიანი მეხსიერების გრძელი (Long Short Term Memory (LSTM)) მოდელს, რომელიც გამოიყენება ხმის ამოცნობის, ხელნაწერის ამოცნობის და დროითი მწკრივების პროგნოზრებისათვის.

მეექვსე თავში, დასკვნაში, წარმოდგენელი იქნება მოდელების სიზუსტის შეფასება და შეირჩევა საუკეთესო მოდელი შესაბამისი კრიტერიუმის მიხედვით. მოდელირებისათვის გამოყენებულია ორთვიანი პერიოდის მონაცემები, ხოლო პროგნოზის სიზუსტე შეფასებულია მომდევნო 40-50 დღის მონაცემებზე. ზოგ შემთხვევაში, მოდელის მოთხოვნების გათვალისწინებით, მოდელის აგებისათვის შეირჩევა განსახვავებული დროის ჰორიზონტი. ოთხივე ტიპის მეთოდი მორგებულია ემპირიულ მონაცემებზე და შედარებულია მათი მორგების ხარისხი საშუალო კვადრატული შეცდომის საშუალებით (RMSE).

1. ლიტერატურის მიმოხილვა

მწვავე რესპირატორული დაავადებების პროგნოზრების ლიტერატურაში გავრცელებულია სხვადასხვა ტიპის სტატისტიკური მოდელები, როგორიცაა განყოფილებიანი მოდელები (compartment models), ლოგისტიკური მოდელები, ავტორეგრესიული მოდელები. უკანასკნელ პერიოდში იყო მცდელობა მანქანური სწავლების მოდელების გამოყენებისა, თუმცა ნაკლები პოპულარობით.

კე ვუმ, დიდიერ დარსეტმა და სხვ. (Wu, Darcet, & al., 2020) ზრდის მოდელებით განახორციელეს კოვიდ-19-ის პროგნოზირება ჩინეთის 29 პროვინციაში, აზიისა და ევროპის რამდენიმე ქვეყანაში. მათ გამოიყენეს ლოგისტიკური ზრდის მოდელი, განზოგადოებული ლოგისტიკური ზრდის მოდელი, განზოგადებული ზრდის მოდელი და განზოგადებული რიჩარდის მოდელი. ჩინეთის მონაცემები დაყოფილია სამ პერიოდად განხორციელებული ღონისძიებების შესაბამისად და თითოეულ პერიოდში გამოთვლილია ზრდის/კლების ტემპი. მორგების ყველაზე კარგი ხარისხი (დეტერმინაციის კოეფიციენტის მიხედვით) აჩვენა რიჩარდის მოდელმა.

ფაიროზა ამირა ჰამზარმა, ჩერ ჰან ლაუმ და სხვ. (Hamzah, Lau, & al., 2020) გამოიყენეს SEIR მოდელი მსოფლიოს მასშტაბით კორონავირუსის გავრცელების მოდელირებისათვის. მოდელის მიხედვით ინფიცირებულთა მაქსიმალური რაოდენობა, 425 066 მლნ დაფიქსირდება 2020 წლის 23 მაისს, ხოლო სექტემბრის შუა პერიოდში ეპიდემია დასრულდება.

ტომას ვილდინგმა (Wilding, 2020) SIR მოდელის გამოყენებით განახორციელა გაერთიანებულ სამეფოში კოვიდ-19-ის გავრცელების პროგნოზირება. მოდელში ეპიდემიის საწყის ეტაპზე აიგო და გათვალისწინებული არ იყო სახელმწიფოს მიერ განხორციელებული შეკავების ღონისძიებები შესაბამისად, ეპიდემიის სიმწვავე ზედმეტობით იყო პროგნოზირებული: 26 აპრილისათვის ინფიცირებულთა რიცხვს 6 მილიონისთვის უნდა მიეღწია.

სუნიტა ტივარმა და სხვ. (Tiwari, Kumar, & al., 2020) გამოიყენეს ეპიდემიის გავრცელების ჩინეთის ტრენდი ინდოეთში ინფიცირებულთა და გარდაცვლილთა რიცხვის პროგნოზირებისათვის. სიკვდილიანობის შეფასებისათვის მათ გაითვა-ლისწინეს ქვეყნის დემოგრაფიული თავისებურებები. ინდოეთის მოსახლეობაში

ასაკოვანთა წილის გათვალისწინებით, სიკვდილიანობის მოსალოდნელი ინტენსივობა სხვა ქვეყნებთან შედარებით ნაკლები იყო პროგნოზირებული.

ორიგინალური ნაშრომი შემოგვთავაზეს ჩენგ-შანგ ჩანგმა და სხვ (Chen, Lu, & al., 2020). მათ კვლევის დასაწყისში დასვეს ექვსი კითხვა:

1. არის თუ არა შესაძლებელი კოვიდ 19-ის შეჩერება და ახდენს თუ არა გავლენას შეკავების ღონისძიებები (მოძრაობის აკრძალვა, ქალაქის მასშტაბით გარკვეული ობიექტების დაკეტვა, საგანმანათლებლო პროპაგანდა ჯანდაცვის საკიტხებთან დაკავშირებით და სხვ.) ეპიდემიის შეჩერებაზე.
2. თუ შესაძლებელია კოვიდ 19-ის შეჩერება, როდის იქნება ეპიდემიის პიკი და როდის დასრულდება იგი?
3. რა გავლენას ახდენს სიმპტომების გარეშე მიმდინარე ეპიდემია დაავადების გავრცელებაზე?
4. თუ კოვიდ 19-ის შეჩერება ვერ მოხდება, მოსახლეობის რა წილი უნდა დაინფიცირდეს იმისათვის რომ ჯოგური იმუნიტეტის ეფექტი მიიღწეს?
5. რამდენად ეფექტურია სოციალური დისტანცირება?
6. თუ კოვიდ 19 ვერ შეჩერდება მოსახლეობის რა წილი დაინფიცირდება გრძელვადიან პერიოდში?

აღნიშნული კითხვების საპასუხოდ მათ განავითარეს დინამიკური SIR მოდელი დროში ცვალებადი კოეფიციენტებით, რომლის საშუალებითაც შესაძლებელია ეფექტური რეპროდუქციის კოეფიციენტის გაანგარიშება. დროში ცვალებადი ბეტა და გამა კოეფიციენტების გასაანგარიშებლად და პროგნოზირებისათვის მათ გამოიყენეს სასრული ინპულსური პასუხის (Finate Impulse Response (FIR)) ფილტრი წრფივ სისტემაში. ხოლო ფილტრის კოეფიციენტების შესაფასებლად გამოიყენეს რეგულირებული უმცირეს კვადრატთა ვარიანტი ე.წ. ქედის რეგრესია (ridge regression). SIR მოდელში ინფიცირებული მოსახლეობა დაყვეს ორ ნაწილად, სიმპტომიანი და უსიმპტომო ინდივიდები, რომელთაც შეესაბამებოდათ რეპროდუქციის განსხვავებული კოეფიციენტები (უსიმპტომო ინფიცირებული ნაკლებად ერიდება კონტაქტებს და შესაბამისად უფრო მაღალი რეპროდუქციის კოეფიციენტის მაჩვენებლით ხასიათდება). ქსელებში დაავადების გავრცელების მოდელირებისათვის გამოიყენეს დამოუკიდებელი

კასკადების მოდელი (Independent Cascade Model) და მისი საშუალებით განსაზღვრეს სოციალური დისტანცირების ეფექტი. მათ განახორციელეს ინფიცირებულთა და დაავადებულთა ერთდღიანი პროგნოზები საკმაოდ მაღალი სიზუსტით (3%-იანი შეცდომა).

SIR მოდელი გამოყენებულ იქნა ბანგლადეშში ეპიდემიის ტრენდის აღსაწერად და პროგნოზირებისათვის მუჰამედ რაჰმენისა და სხვ. (Rahman, Ahmed, & al., 2020) მიერ. გამომდინარე იქედან, რომ აღნიშნული მოდელი თავად ვერ ითვალისწინებს სოციალური დისტანცირებისა და სხვა შეკავების ღონისძიებების ეფექტს ეპიდემიის გავრცეებაზე, მათ ხელოვნურად შეამცირეს ე.წ. დაავადების რისკის ქვეშ მყოფი (Susceptible) მოსახლეობის საწყისი მოცულობა. სიმულაცია განხორციელდა რამდენიმე სცენარის მიხედვით (სოციალური დისტანცირების წესებს იცავს მოსახლეობის 0%, 50%, 60%, 70%, 80%, 90%, 99% და 99.5%). კვლევამ გარკვეულწილად პარადოქსული შედეგები აჩვენა. ინფიცირებულთა რაოდენობა 0%-იანი სოციალური დისტანციის შემთხვევაში 163 689 383-ს მიაღწევდა, დისტანცირების მასშტაბის ზრდასთან ერთად მონოტონურად შემცირდება და 99.5%-იანი სოციალური დისტანციის პირობებში 47297-ს მიაღწევდა. პირველი სცენარის მიხედვით პიკი მიიღწეოდა 2020 წლის 11 ივნისს, პიკის თარიღი კვლავ მონოტონურად მცირდება და საუკეთესო სცენარის მიხედვით ამავე წლის თორმეტ მაისს დასრულდება. ინფიცირებულთა რაოდენობის მონოტონური შემცირება სოციალური დისტანცირების ზრდის უკუპროპორციულად ლოგიკურია, თუმცა ლიტერატურაში საპირისპირო თეორიას ვხვდებით პიკის წერტილთან დაკავშირებით. სოციალური დისტანცირების ღონისძიებები მიმართულია ეპიდემიის მრუდის გაბრტყელებაზე, რაც ნიშნავს, რომ ეპიდემიის ხანგრძლივობა იზრდება, თუმცა მკვეთრად იკლებს ყოველდღიურად დაფიქსირებბული ინციდენტების რაოდენობა.

ალი აჰმადმა, იასინ ფადაეიმ და სხვ. (Ahmadi, Fadaei, & al.) გამოიყენეს შეცდომების მესამე რიგის პოლინომიალური მოდელი, გომპერსისა და ფონ ბერტალანფისა მოდელები ირანში ეპიდემიის გავცელების მოდელირებისათვის. მოდელები აიგო სხვადასხვა სცენარების მიხედვით, რომლებიც შემუშავდა ეპიდემიოლოგების, ბიოსტატისტიკოსებისა და მათემატიკოსების მიერ. ყველაზე ოპტიმისტური პროგნოზი გაკეთდა ბერტალანფის მოდელის მიხედვით (48200 ინფიცირებული ეპიდემიის

დასრულების დროს - 13 მაისს), ხოლო პესიმისტური შედეგი პოლინომიალურმა მოდელმა აჩვენა 58000 ინფიცირებული. მკვლევარების მიერ საუკეთესო მოდელად ბერტალანფის მოდელი იქნა მიჩნეული.

ჩინეთში კორონავირუსის ეპიდემიის მოდელირებისათვის ბერტალანფის, გომპერსისა და ლოგისტიკური მოდელები გამოიყენეს კევინ ლიმ, ლინ ჯიამ და სხვ. ყველაზე კარგი მორგების ხარისხი აჩვენა ლოგისტიკურმა მოდელმა, ხოლო გომპერსის მოდელმა ბერტალანფის მოდელზე უკეთესი შედეგი დააფიქსირა.

გომპერსის მოდელი გამიყენეს ჩეხმა მკვლევარებმა ჯირი მაზურეკმა და სუსანა ნენიცკოვამ (Mazurek & Nenickova, 2020) აშშ-ში კორონავირუსით ინფიცირებულთა და გარდაცვლილთა რიცხვის და პიკის თარიღის პროგნოზირებისათვის. მოდელმა მაღალი მორგების ხარისხი აჩვენა დეტერმინაციის კოეფიციენტის მიხედვით (R^2= 99.61%). პიკი პროგნოზირებული იყო 17 აპრილისათვის, ხოლო ეპიდემიის დასრულების დროს ინფიცირებულთა მთლიანი რაოდენობა მიაღწევს 2 084 000 ადამიანს. R^2= 99.44% მორგების ხარისხის მქონე მოდელის მიხედვით გარდაცვლილთა რაოდენობა ეპიდემიის დასრულების შემდეგ 110 000 იქნება.

ფენომენოლოგიური მოდელების (ექსპონენციალური ზრდის, განზოგადებული ზრდის, ლოგისტკური ზრდის, განზოგადებული ლოგისტიკური ზრდის, რიჩარდის მოდელის, განზოგადებული რიჩარდის) და SEIR მოდელის პერფორმანსის შედარება ირანში კოვიდ 19-ით ინფიცირებულთა მონაცემებზე განახორციელეს ჰ. მასჯედმა და სხვ. (Masjedi, Rabajante, & al., 2020) მკვლევრებმა განახორციელეს მომავალი თვეში ეპიდემიის გავრცელების პროგნოზირება. ფენომენოლოგიური მოდელის პროგნოზი მერყეობდა 68 486-სა და 118923-ს შორის, ხოლო SEIR მოდელმა აპრილის ბოლოდან ეპიდემიის მეორე ტალღის წამოწყება იპროგნოზა.

ფუდანის უნივერსიტეტის ჩენგის ჯგუფმა კოვიდ 19 ეპიდემიის პროგნოზირებისთვის შეიმუშავა ერთჯაჭვიანი (single-chain Fudan-CCDC model) მოდელი, რომელიც შემდეგ ნიან შაოსა და ვენბინ ჩენის (Shao, Yan, & al., Multi-chain Fudan-CCDC model for COVID-19 -- a revisit to Singapore's case, 2020) მიერ განვითარდა მრავალჯაჭვიან (Multi-chain) მოდელად. მოდელის მოდიფიკაციის მიზეზი იყო ეპიდემიის გავრცელების ტრენდის არაერთგვაროვნება. სამხრეთ კორეის მონაცემებზე

დაკვირვების შედეგად მათ შეამჩნიეს ზრდის ტემპის უეცარი მკვეთრი ცვლილება. მრავალჯაჭვიანი მოდელი საშუალებას იძლევა სხვადასხვა განტოლებით მოდელირდეს ეპიდემიის გავრცელების განსხვავებული სტრუქტურის პერიოდები. მოდელი შემდეგ გამოყენებულ იქნა ირანში და სინგაპურში ეპიდემიის ტრენდის მოდელირებისათვის (Shao, Pan, & al., Multi-chain Fudan-CCDC model for COVID-19 in Iran, 2020). მოდელმა მორგების კარგი ხარისხი აჩვენა, ამავდროულად, ორჯაჭვიანი მოდელის სიზუსტე უფრო მაღალი იყო ვიდრე სამჯაჭვიანის. მოდელის მიხედვით ირანში ეპიდემია 10 მაისს უნდა დასრულებულიყო და ინფიცირებულთა პროგნოზირებული ჯამური რაოდენობა 80 ათასს მიაღწევდა ამ დროისათვის.

საბაზისო და ეფექტური რეპროდუქციის რიცხვები შეფასდა საზოგადოებრივი ჯანმრთელობის მონაცემების ანალიზის ცენტრის მიერ ინდოეთის მონაცემების მიხედვით (Mohapatra, 2020) ეფექტური რეპროდუქციის მნიშვნელობა საწყის ეტაპზე არასტაბილური იყო და მერყეობდა 0.28-სა და 2.5-ს შორის, თუმცა ერთი თვის შემდეგ მეტნაკლებად დასტაბილურდა და 1.27-ის ფარგლებში მოძრაობდა 0.01 სტანდარტული გადახრით.

2. განყოფილებიანი მოდელები

ინფექციური დაავადებების მოდელირებისათვის ხშირად გამოიყენება ე.წ. განყოფილებიანი მოდელები (compartmental models). კლასიკური განყოფილებიანი მოდელები ეფუძვნებიან ექსპონენციალური ზრდის დაშვებას ეპიდემიის საწყის ეტაპზე. მოსახლეობა დაყოფილია ჯგუფებად. თითოეულ ჯგუფში შემავალ ყოველ ინდივიდს აქვს ერთი და იგივე მახასიათებელი. მოდელირებისთვის გამოიყენება ჩვეულებრივი (დეტერმინირებული) დიფერენციალური განტოლებები, თუმცა რეალური გრაფიკის ასახვისათვის, ზოგ შემთხვევაში, განიხილავენ უფრო კომპლექსურ სტოხასტურ ვარიანტს. განყოფილებიანი მოდელები გამოიყენება დაავადების გავცელების სხვადასხვა მახასიათებლის პროგნოზირებისათვის, როგორიცაა მაგალითად, ინფიცირებულთა მთლიანი რაოდენობა დაავადების დასრულებისას, ეპიდემიის ხანგრძლივობა, პიკის მომენტი და სხვ. მათი საშუალებით შეგვიძლია განვსაზვროთ რა გავლენას ახდენს სხვადასხვა სცენარი შედეგებზე, რამდენად ეფექტური იქნება გატარებული ღონისძიება, როდისაა ოპტიმალური ამა თუ იმ ღონისძიების (მაგალითად, კარანტინის) დაწყება, რა არის ოპტიმალური ხანგრძლივობა და ა.შ.

2.1 SIR მოდელი

განყოფლიებიანი მოდელების ერთ-ერთ სახეს წარმოადგენს SIR მოდელი. მოდელი შეიცავს სამ გაყოფილებას: S არის ინფიცირების რისკის ქვეშ მყოფთა (susceptible) რაოდენობა, I აღნიშნავს ინფიცირებულთა (Infected) რაოდენობას, ხოლო R გამოჯანმრთელებულთა და გარდაცვლილთა (Removed) რაოდენობის ჯამია. ეს მოდელი გამოიყენება გადამდები დაავადებების მოდელირებისათვის, როგორიცაა წითელა, ყბაყურა, წითურა და სხვ.

განყოფილებებში ინდივიდთა რიცხოვნობა წარმოადგენს დროის ფუნქციას, S=S(t); I=I(t); R=R(t), და იგი სქემატურად გამოისახება შემდეგი დიაგრამით:

სადაც β აღნიშნავს დროის ერთეულში ინდივიდის კონტაქტების რაოდენობის საშუალოსა და ინფიცირებულთან კონტაქტის შემთხვევაში დაავადების გადაცემის

ალბათობის ნამრავლს. S და I განყოფილებებს შორის გადაცემის კოეფიციენტი არის βI/N, სადაც N წარმოადგენს მთლიან მოსახლეობას და შესაბამისად I/N არის ინფიცირებულთან კონტაქტების წილი მთლიან კონტაქტებში. I-დან და R-ში გადასვლის კოეფიციენტი (transition rate) არის γ=1/D, სადაც D აღნიშნავს გამოჯანმრთელების ხანგრძლივობას.

SIR მოდელის აგებისათვის გამოვიყენებთ დაშვებას (Smith & Moore, 2004), რომ ინფიცირების რისკის ქვეშ მყოფი ჯგუფის რიცხოვნობა არ იზრდება, რაც ნიშნავს რომ ჩვენ არ ვითვალისწინებთ დაბადებასა და იმიგრაციას. ასევე არ ვითვალისწინებთ გარდაცვალებას (გარდა ამ ეპიდემიით გარდაცვალებისა), ე.ი. რისკის ქვეშ მყოფთა ჯგუფიდან გასვლის ერთადერთი გზა არის ინფიცირებულთა კატეგორიაში გადასვლა.

SIR სისტემა აღიწერება შემდეგი დიფერენციალური განტოლებების სისტემის საშუალებით:

$$\frac{dS}{dt} = -\frac{\beta SI}{N} \tag{2.1.1}$$

$$\frac{dI}{dt} = \frac{\beta SI}{N} - \gamma I \tag{2.1.2}$$

$$\frac{dR}{dt} = \gamma I \tag{2.1.3}$$

დაშვების მიხედვით მოსახლეობის რაოდენობა უცვლელია, რაც გულისხმობს შემდეგს:

$$\frac{dS}{dt} + \frac{dI}{dt} + \frac{dR}{dt} = 0 \tag{2.1.4}$$

პირველი დიფერენციალური განტოლება აღწერს ინფიცირების რისკის ქვეშ მყოფი პოპულაციის ცვლილებას დროთა განმავლობაში, სადაც S-ისა და I-ის ნამრავლი აღწერს რისკის ქვეშ მყოფთა და ინფიცირებულთა ყველა შესაძლო კონტაქტების რიცხვს. ცვლილება არადადებითია, რაც ნიშნავს, რომ S კატეგორიაში მყოფი მოსახლეობის რიცხოვნობა მონოტონურად მცირდება. მეორე დიფერენციალური განტოლება აღწერს ინფიცირებულთა რაოდენობის ცვლილების დროზე დამოკიდებულებას. ინფიცირებულთა რიცხოვნობის ზრდის წყაროს წარმოადგენს S კატეგორიაში მყოფი მოსახლეობა, ხოლო შემცირება ხდება გამოჯანმრთელებულთა ან გარდაცვლილთა საშუალებით. მესამე დიფერენციალური განტოლება აღწერს გარდაცვლილთა და გამოჯანმრთელებულთა რიცხვის ცვლილებას დროთა განმავლობაში, რომელიც

დამოკიდებულია ინფიცირებულთა რაოდენობასა და გამა პარამეტრის მნიშვნელობაზე დადებითად.

(2.1.2) განტოლებიდან ჩანს, რომ ინფიცირების დინამიკა დამოკიდებულია ბეტა და გამა პარამეტრების მნიშვნელობაზე. რაც უფრო დიდია გამას მნიშვნელობა, მით უფრო სწრაფად მცირდება ინფიცირებულთა რაოდენობა, ხოლო ბეტა პარამეტრის ზრდა, პირიქით, აჩქარებს ინფიცირებულთა კატეგორიის ზრდას და ეპიდემიის სიმწვავეს ზრდის. ამ ორი პარამეტრის საშუალებით შეგვიძლია გამოვითვალოთ ე.წ. საბაზისო რეპროდუქციის კოეფიციენტი, რომელიც ეპიდემიის გავრცელების სხვადასხვა მახასიათებლის კარგი საზომია და რომლითაც შეგვიძლია შევაფასოთ მისი სიმწვავე. რეპროდუქციის კოეფიციენტი შემდეგნაირად გამოისახება:

$$R_0 = \frac{\beta}{\gamma}$$

აღნიშნული კოეფიციენტი არაუარყოფითია. თუ მისი მნიშვნელობა ერთზე მაღალია ეს ნიშნავს, რომ ადგილი აქვს ეპიდემიის ექსპონენციალურ ზრდას, ხოლო ერთზე დაბალი მნიშვნელობა მიუთითებს, რომ ეპიდემია ჩაქრობის ფაზაშია შესული. საბაზისო რეპროდუქციის კოეფიციენტთან ერთად ხშირად განიხილავენ ეფექტურ (დროში ცვალებად) რეპროდუქციის კოეფიციენტს, რომელიც ახასიათებს ეპიდემიის სიმწვავის დროში ცვლილებას და უფრო ნათელ გრაფიკს გვიქმნის ეპიდემიის დინამიკაზე. მათზე საუბარს მოგვიანებით განვაგრძობ, ახლა კი დავუბრუნდეთ დიფერენციალურ განტოლებათა სისტემას.

პირველი განტოლების მესამეზე გაყოფით, ცვლადების სეპარაციისა და ინტეგრირების შემდეგ მივიღებთ:

$$S(t) = S(0)e^{-R_0\left(R(t)0R(0)\right)/N} \qquad (2.1.5)$$

სადაც S(0) და R(0) რისკის ქვეშ მყოფთა და გამოჯანმრთელებულთა (გარდაცვლილებთან ერთად) საწყისი რაოდენობაა, ხოლო S(∞) და R(∞) საბოლოო რაოდენობები. ავღნიშნოთ:

$$r_\infty = \frac{R(\infty)}{N};\ s_\infty = \frac{S(\infty)}{N};\ r_0 = \frac{R(0)}{N} \text{ და } s_o = \frac{S(0)}{N};$$

მაშინ გვექნება:

$$s_\infty = 1 - r_\infty = -R_0^{-1} W\left(-s_0 R_0 e^{-R_0(1-r_0)}\right) \qquad (2.1.6)$$

სადაც W() ლამბერტის W ფუნქციაა.

(2.1.6)-დან ცხადია, რომ თუ s_0 განსხვავდება ნულისაგან, პოპულაციის ყველა ინდივიდი არ გადადის R განყოფილებაში, არამედ ნაწილი რჩება რისკის ქვეშ მყოფთა კატეგორიაში. ეს გულისხმობს, რომ ინფექცია ჩერდება არა რისკის ქვეშ მყოფი კატეგორიის გაქრობით, არამედ ინფიცირებული ადამიანების რაოდენობის კლების შედეგად.

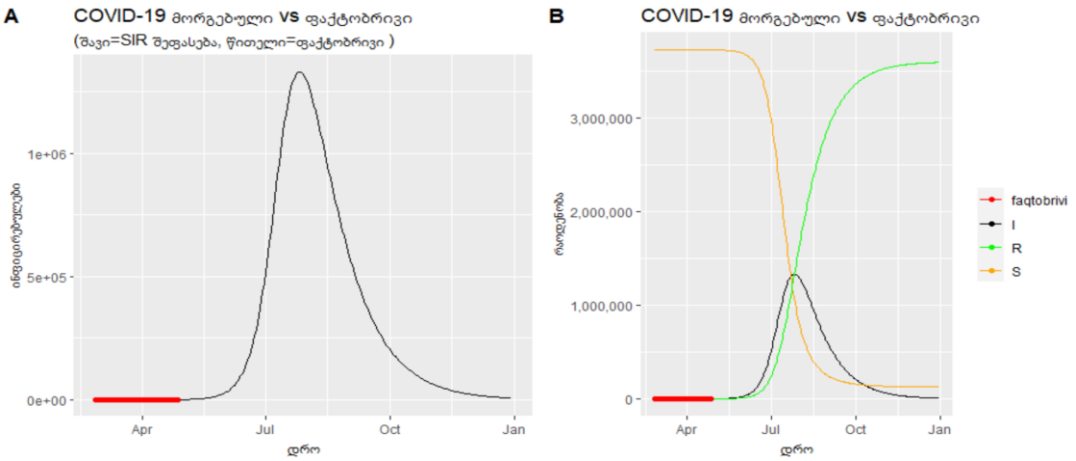
დავუბრუნდეთ საბაზისო რეპროდუქციის კოეფიციენტს და მისი საშუალებით გადავწეროთ (2.1.2) განტოლება:

$$\frac{dI}{dt} = (R_0 \frac{S}{N} - 1)\gamma I \qquad (2.1.7)$$

საიდანაც ვღებულობთ, რომ თუ: $R_0 > \frac{N}{S(0)}$, მაშინ $\frac{dI}{dt}(0) > 0$, რაც ნიშნავს, რომ ეპიდემია გავრცელდება მოსახლობის მნიშვნელოვან ნაწილში, წინააღმდეგ შემთხვევაში კი ეპიდემიის გავრცელების მასშტაბი უმნიშვნელო იქნება.

SIR მოდელის ასაგებად გამოვიყენებთ R პროგრამის deSolve პაკეტს. საწყის ეტაპზე განვსაზღვრავთ დიფერენციალური განტოლებების ფუნქციას უცნობი ბეტა და გამა პარამეტრებით. შემდეგ განვსაზღვრავთ ოპტიმიზაციის ფუნქციას (საშუალო კვადრატული შეცდომა) რომელიც არჩევს პარამეტრების იმგვარ მნიშვნელობებს, რომ ინფიცირებულთა ფაქტობრივ და მოდელით პროგნოზიებული მნიშვნელობების სხვაობების კვადრატების ჯამის მინიმიზება მოხდეს. გამას საწყის მნიშვნელობად ვუთითებთ $\frac{1}{21} = 0.04756$-ს , რადგან გამოჯანმრთელებისთვის საჭირო დრო საშუალოდ სამი კვირაა, ხოლო $\gamma = 1/D$, სადაც D აღნიშნავს საშუალოდ გამოჯანმრთელებისთვის საჭირო დროს. ბეტას საწყის მნიშვნელობად აღებულია 0.05 რეპროდუქციის საბაზისო რიცხვის შესაძლო მნიშვნელობიდან გამომდინარე. დიფერენციალური განტოლების ამოხსნის შედეგად შეფასებული კოეფიციენტების მნიშვნელობებია: $\beta = 0.14956$, $\gamma = 0.0426$; მაშასადამე რეპროდუქციის რიცხვის მნიშვნელობაა $R_0 = \frac{0.14956}{0.0426} = 3.508$, რაც ნიშნავს, რომ ერთი ინფიცირებული საშუალოდ 3.508 ადამიანს აინფიცირებს; უნდა აღინიშნოს, რომ ეს უკანასკნელი შეფასება გადაჭარბებულია და, შესაბამისად, პროგნოზის შეგედებიც რეალობისგან შორსაა. მოდელის მიხედვით პიკი 27 ივლისს

მიიღწევა, ხოლო ეპიდემიის დასრულების შემდეგ თითქმის მთელი მოსახლეობა იქნება ინფიცირებული. ეპიდემიის გავრცელების ამგვარი შეფასება განპირობებულია იმით, რომ მოდელი არ ითვალისწინებს შეკავების ღონისძიებებს. მოდელის პროგნოზის შედეგები მოცემულია გრაფიკი.1-ზე. შავი ფერით, ორივე ნახაზზე, მოცემულია ინფიცირებულთა მიმდინარე რაოდენობა(I), წითელი ფერით მოცემულია ინფიცირებულთა ფაქტობრივი მნიშვნელობა, მწვანე აღნიშნავს გამოჯანმრთელებულთა და გარდაცვლილთა ჯამურ რაოდენობას (R), ხოლო ყვითელი ფერით მოცემულია დაავადების რისკის ქვეშ მყოფი მოსახლეობა (S).



გრაფიკი.1

როგორც ავღნიშნეთ მოდელის აგებისათვის გამოყენებული დაშვებები არარელევანტურია და შესაბამისად პროგნოზის სიზუსტეც ძალიან ცუდია, კერძოდ, საშუალო კვადრატული შეცდომიდან ფესვი, RMSE=5976-ის ტოლია.

უნდა აღინიშნოს, რომ, ზოგადად, ლაბორატორიულად დადასტურებული შემთხვევები რეალურ ინფიცირებულებზე ნაკლებია. ლაბოლატორიულად დადასტურებულთა და რეალურად ინფიცირებულთა თანაფარდობას ეწოდება დადასტურების კოეფიციენტი (ascertianment rate). მოდელში შეიძლება ამ არაზუსტობის გათვალისწინება ინფიცირებულთა რაოდენობის შეწონვით. საინტერესოა, რომ თუ აღნიშნული კოეფიციენტი დროში არ იცვლება, მაშინ ადგილი აქვს საინტერესო შედეგს: დადასტურების კოეფიციენტის მიხედვით კორექტირებული მონაცემებით აგებულ SIR მოდელში ეპიდემიის დასრულების შემდეგ ინფიცირებულთა საერთო რაოდენობა იგივე იქნება, რაც კორექტირების გარეშე მოდელში, ასევე უცვლელი იქნება გამა და ბეტა კოეფიციენტი და შესაბამისად საბაზისო რეპროდუქციის რიცხვიც.

2.2 გაფართოებული SIR მოდელი

გაფართოებული SIR მოდელი ეყრდნობა დირიხლეტ-ბეტა მდგომარეობა-სივრცის მოდელს (Dirichlet-Beta state-space model (DBSSM)). ეს უკანასკნელი ჩვეულებრივი SIR მოდელისგან განსხვავებით ითვალისწინებს პარამეტრებში და დაავადების გადაცემის მექანზმში არსებულ უზუსტობებს. DBSSM მოდელი განისაზღვრება შემდეგი ფორმით:

$$y_t|\theta_t, \phi \sim Beta(\lambda\theta_t^I, \lambda(1-\theta_t^I)) \quad (2.2.1)$$

$$\theta_t|\theta_{t-1}, \phi \sim Dirichlet(kf(\theta_{t-1}, \beta, \gamma)) \quad (2.2.2)$$

სადაც y_t არის ინფიცირებულთა წილი მთლიან მოსახლეობაში, $\theta_t = (\theta_t^S, \theta_t^I, \theta_t^R)'$ აღნიშნავს ნამდვილ, მაგრამ არადაკვირვებად რისკის ქვეშ მყოფ (susceptible), ინფიცირებულ (infectious) და გამოჯანმრთელებულ და გარდაცვლილ (removed) მოსახლეობის წილს, შესაბამისად. $\phi = \{k, \theta_0, \beta, \gamma, \lambda\}$, სადაც γ>0 გამოჯანმრთელების კოეფიციენტია, β>0 ინფექციის გადაცემის კოეფიციენტია, $k > 0$, აკონტროლებს (2.2.2) განტოლების ვარიაციას, $\lambda > 0$ აკონტროლებს (2.2.1) განტოლების ვარიაციას, ხოლო $f(\theta_{t-1}, \beta, \gamma) \in R^3$-ზე დეტალურად ქვემოთ ვისაუბრებთ. DBSSM-ის დაშვებით $\theta_{0:T} = (\theta_0, \theta_1, \dots, \theta_T)$ არის პირველი რიგის მარკოვის ჯაჭვი (ე.ი. $[\theta_t|\theta_{0:(t-1)}] = [\theta_t|\theta_{t-1}]$ ყველა t-სათვის) და ყველა t≠s-თვის y_t და y_s დამოუკიდებელნი არიან მოცემული θ_t-სთვის.

$f(.)$ წარმოადგენს შემდეგი განტოლების ამონახსნს:

$$\frac{d\theta_t^S}{dt} = -\beta\theta_t^S\theta_t^I \quad (2.2.3)$$

$$\frac{d\theta_t^I}{dt} = \beta\theta_t^S\theta_t^I - \gamma\theta_t^I \quad (2.2.4)$$

$$\frac{d\theta_t^R}{dt} = \gamma\theta_t^I \quad (2.2.5)$$

(2.2.3)-(2.2.5) განტოლებათა სისტემის ამონახსნი ზუსტად განსაზღვრული არ არის, ამიტომ $f(.)$ ჩანაცვლებულია რიცხვითი მიახლოებით. ამ შემთხვევაში განვიხილავთ რანჯ-კუტას (Runge–Kutta) მეოთხე რიგის მიახლოებას (იხ. დანართი.1). არსებობს მიახლოების სხვა ვარიანტებიც, მაგალითად, ეილერის მეთოდი.

სწორედ $f(.)$ ფუნქციის საშუალებით ხდება θ_t ლატენტური მდგომარეობის გადასვლა ერთი საფეხურით წინ. $y_t|\theta_t, \phi$-ს მოდელირება ხდება ბეტა განაწილებით, რომელიც

ფართოდ გამოიყენება [0,1] ინტერვალში განსაზღვრული მონაცემებისთვის. (2.2.1)-ის პარამეტრიზაცია შემდეგნაირად მოიცემა:

$$E(y_t|\theta_t, \phi) = \theta_t^I \tag{2.2.6}$$

$$Var(y_t|\theta_t, \phi) = \frac{\theta_t^I(1-\theta_t^I)}{1+\lambda} \tag{2.2.7}$$

y_t-ს პირობითი ლოდინი გადაუადგილებელი შეფასებაა ინფიცირებული მოსახლეობის წილის ნამდვილ, მაგრამ არადაკვირვებადი მნიშვნელობისა, θ_t^I-სი. y_t-ის პირობითი ვარიაცია არის θ_t^I-ს და λ-ს ფუნქცია. λ პარამეტრი მონაწილეობს მხოლოდ ვარიაციის განსაზღვრაში და არა ლოდინის განსაზღვრაში. როდესაც მისი მნიშვნელობა უსასრულოდ იზრდება, ვარიაცია ნულისაკენ მიისწრაფვის.

$\theta_t|\theta_{t-1}, \phi$-ის მოდელირება ხდება დირიხლეტის განაწილებით, რომელიც ფართოდ გამოიყენება არაუარყოფითი მნიშვნელობის მქონე ვექტორული მონაცემებისათვის, რომელთა ჯამი ერთის ტოლია. (2.2.2)-ის პარამეტრიზაცია შემდეგი სახით მოიცემა:

$$E(\theta_t|\theta_{t-1}, \phi) = f(\theta_{t-1}, \beta, \gamma) \tag{8.2.8}$$

$$\left\{ \begin{array}{l} Var(\theta_t^S|\theta_{t-1}, \phi) \\ Var(\theta_t^S|\theta_{t-1}, \phi) \\ Var(\theta_t^S|\theta_{t-1}, \phi) \end{array} \right\} = \frac{1}{1+k}[f(\theta_{t-1}, \beta, \gamma)\mathrm{O}(\mathbb{I} - \mathrm{f}(\theta_{t-1}, \beta, \gamma))] \tag{2.2.9}$$

სადაც „0“ აღნიშნავს ჰადამარდის ნამრავლს და $\mathbb{I}$ არის 3 X 1 განზომილებიანი ერთეულოვანი ვექტორი. ლატენტური მდგომარეობის მოდელის პირობითი საშუალოს სტრუქტურა (ე.ი. θ_t-ის პირობითი ლოდინი) არის გადაუადგილებელი შეფასება ერთეული ინტერვალით დაშორებული ამონახსნისათვის (2.2.2) θ_{t-1}-ს მოცემულობით. θ_t-ს პირობითი ვარიაცია არის ფუნქცია $f(\theta_{t-1}, \beta, \gamma)$- სა და k-სი. როდესაც k პარამეტრი მიისწრაფვის უსასრულობისკენ, y_t-ს პირობითი ვარიაცია ნულისაკენ მიისწრაფვის.

უნდა აღინიშნოს, რომ არსებობს სტოხასტური SIR მოდელის სხვა მოსახერხებელი მიახლოება, როგორიცაა SIR მოდელის სტოხასტური დიფერენციალური განტოლებების ვერსია (SDE-SIR) ან დროში უწყვეტი მარკოვის ჯაჭვის ვერსია (CTMC-SIR). ეს მიახლოებები გარკვეული მიზეზების გამო არ გამოვიყენეთ. კერძოდ, SDE-SIR მოდელის გაუსიანური შეცდომები არ ითვალისწინებს განყოფილებიანი მოდელების განსაზღვრის შეზღუდულ არეალს, რის გამოც პროგნოზირებული მნიშვნელობა შეიძლება დასაშვებ მნიშვნელობათა საზღვრებს გასცდეს (მაგალითად,

ინფიცირებულთა წილის პროგნოზირებული მნიშვნელობა იყოს უარყოფითი). მეორეს მხრივ, დირიხლესა და ბეტა განაწილებები DBSSM მოდელისათვის შერჩეულია სწორედ პროგნოზირებული ლატენტური მდგომარეობიისა და მნიშვნელობების მიზანშეწონილობის უზრუნველსაყოფად. (Osthus, Hickmann, & al., 2017)

ამ მოდელზე დაყრდნობით განვითარდა გაფართოებული SIR მოდელი, სადაც შესაძლებელია გავითვალისწინოთ სახელმწიფოს მიერ გათვალისწინებული შეკავების ღონისძიებები.

2.2.1 ეპიდემიოლოგიური მოდელი დროში ცვალებადი გადაცემის კოეფიციენტით

მარტივი SIR მოდელში გადაცემის, ბეტა და გამოსვლის, გამა, კოეფიციენტები დროში უცვლელნი არიან და მაშასადამე მას არ შეუძლია ასახოს სახელმწიფოს მიერ განხორციელებული შეკავების ღონისძიებები, როგორიცაა კარანტინის რეჟიმის შემოღება, დამცავი საშუალებების გამოყენება და ინფიცირებუთა მყისიერი ჰოსპიტალიზაცია. შესაბამისად, გადაცემის ბეტა კოეფიციენტი დროთა განმავლობაში განიცდის ცვლილებას. ლილი ვანგმა და სხვ. (Wang, Zhou, & al., 2020) შემოგვთავაზეს გაფართოებული SIR მოდელი, რომელიც დროში ცვალებადი გადაცემის ბეტა კოეფიციენტის მოდელირების საშუალებას იძლევა. დავუშვათ დროის t მომენტში $q^S(t) \in [0,1]$ არის ალბათობა, რომ რისკის ქვეშ მყოფი ინდივიდი თვითიზოლაციაში იქნება, ხოლო $q^I(t) \in [0,1]$ არის ალბათობა, რომ ინფიცირებული ინდივიდი საავადმყოფოშია მოთავსებული. θ_t^S არის რისკის ქვეშ მყოფი მოსახლეობის წილი, θ_t^I ინფიცირებული მოსახლეობის წილი, ხოლო θ_t^R გამოჯანმრთელებული და გარდაცვლილი მოსახლეობის წილი დროის t მომენტში. ამის გათვალისწინებით დაავადების გადაცემის ალბათობა შემცირდება შემდეგი სახით:

$$\beta\{1 - q^S(t)\}\theta_t^S\{1 - q^I(t)\}\theta_t^I \coloneqq \beta\pi(t)\theta_t^S\theta_t^I \qquad (2.2.10)$$

სადაც $\pi(t) \coloneqq \{1 - q^S(t)\}\{1 - q^I(t)\} \in [0,1]$ გვიჩვენებს გადაცემის კოეფიციენტის შემცირების სიდიდეს. კარანტინის არ არსებობის შემთხვევაში იგი ერთის ტოლია. მისი საშუალებით SIR მოდელი შემდეგი სახით შეგვიძლია გადავწეროთ:

$$\frac{d\theta_t^S}{dt} = -\beta\pi(t)\theta_t^S\theta_t^I \quad (2.2.11)$$

$$\frac{d\theta_t^I}{dt} = \beta\pi(t)\theta_t^S\theta_t^I - \gamma\theta_t^I \quad (2.2.12)$$

$$\frac{d\theta_t^R}{dt} = \gamma\theta_t^I \quad (92.13)$$

$\pi(t)$ ფუნქციას შეიძლება ჰქონდეს როგორც დისკრეტული (საფეხუროვანი) სახე, ასევე იგი შეიძლება წარმოვადგინოთ უწყვეტი ექსპონენციალური ფუნქციის სახით: $\pi(t) = e^{-\lambda t}$, სადაც λ>0. დისკრეტული ვარიანტის შემთხვევაში ფუნქციას შემდეგი სახე აქვს:

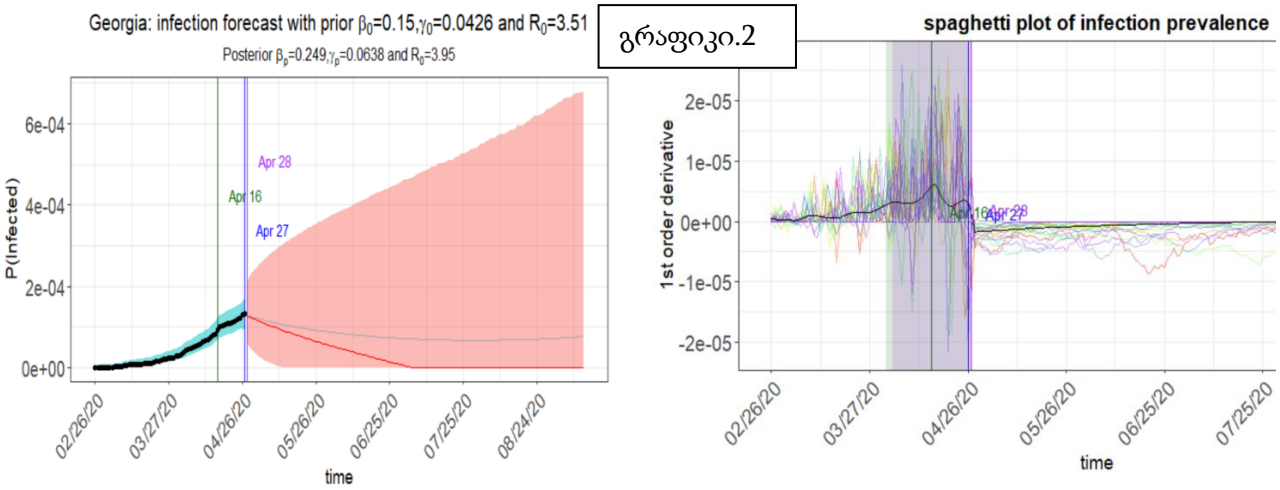
$$\pi(t) = \begin{cases} \pi_1, & \text{თუ } t < t_1; \ \textit{კარანტინის} \text{ არ არსებობის შემთხვევაში} \\ \pi_k, & \text{თუ } t < t_k; \ k - \textit{ური ღონისძიების გატარების} \text{ შემდეგ} \\ \pi_n, & \text{თუ } t > t_n; \ \textit{უკანასკნელი} \text{ ღონისძიების გატარების } \textit{შემდეგ} \end{cases}$$

აღსანიშნავია, რომ $\pi(t)$ ფუნქციის სახე ჩვენ უნდა განვსაზღვროთ, ანუ შეკავების ღონისძიებების ეფექტი უნდა შეფასდეს მკვლევარის მიერ წინასწარ და მისი მოდელით შეფასება არ ხდება.

ჩვენს შემთხვევაში დისკრეტულ ფუნქციას შემდეგი სახე აქვს:

$$\pi(t) = \begin{cases} 1, & \text{თუ } t < 21.03.2020; \ \textit{პრეზიდენტის დეკრეტის} \text{ გამოშვებამდე} \\ 0.7, & \text{თუ } 21.03.20 < t < 31.03.20; \ \textit{კარანტინის} \text{ რეჟიმის ამოქმედებამდე} \\ 0.4, & \text{თუ } t > 31.03.2020; \end{cases} \quad (10.2.14)$$

ექპონენციალური პუნქციის შემთხვევაში გადაცემის კოეფიციენტი ნელ-ნელა მცირდება. ჩვენს შემთხვევაში ექსპონენციალური ფუნქციის ლამბდა კოეფიციენტის მნიშვნელობა 0.025-ის ტოლია. რაც ნიშნავს იმას, რომ კოეფიციენტის მნიშვნელობა ეპიდემიის დაწყებიდან 36 დღეში მიაღწევს 0.4-ს (იხ. დანართი.2). გაფართოებული SIR მოდელის ასაგებად, გარდა $\pi(t)$-ის სპეციფიკაციისა, ასევე საჭიროა საპროგნოზო დროის ჰორიზონტის, მონტე კარლოს სიმულაციის რაოდენობის, ბეტა და გამა კოეფიციენტების მნიშვნელობების მითითება. ამ კოეფიციენტების მნიშვნელობებად შევარჩიე ზემოგანხილული SIR მოდელით მიღებული შეფასებები. გაფართოებული მოდელის შეფასებები მიღებულია 500000 შერჩევის მქონე მონტე კარლოს სიმულაციით. შედეგები წარმოდგენილია გრაფიკი.2-ზე.



გრაფიკი.2

მარცხენა გრაფიკი აღწერს ინფიცირების გავრცელების ტრენდს. ორდინატთა ღერძის მასშტაბი [0,1] ინტერვალშია მოთავსებული, რადგან გრაფიკზე მოცემულია ინფიცირებულთა წილი მთლიან მოსახლეობაში და არა აბსოლიტური მნიშვნელობები. წითელი ხაზით აღნიშნულია მონტე კარლოს სიმულაციის შედეგად მიღებული მედიანური მნიშვნელობა, ხოლო ნაცრისფერი ხაზი აღნიშნავს სიმუალციის საშუალო მნიშვნელობას. სტაფილოსფერი ფერით მოცემულია ინფიცირების პროგნოზის 95%-იანი ნდობის ინტერვალი. შავი ფერი გვიჩვენებს ინფიცირებულთა კუმულატიურ დაკვირვებად მნიშვნელობებს. ლურჯი, მწვანე და ნარინჯისფერი ვერტიკალური ხაზები, შესაბამისად, აღნიშნავენ დაკვირვების ბოლო დღეს, პიკის თარიღსა და მაქსიმალური წმინდა ინფიცირების თარიღს. მოდელის მიხედვით პიკი 16 აპრილს დაფიქსირდა, ხოლო წმინდა ინფიცირებულთა შემცირების თარიღი 28 აპრილს ანუ დაკვირვების ბოლო დღის მომდევნო დღეს დაფიქსირდება. ხოლო ეპიდემია დასრულდება 15 ივლისს და ინფიცირებულთა მთლიანი რაოდენობა 3189-ს მიაღწევს. მარცხენა გრაფიკზე რეპროდუქციის რიცხვის ორი მნიშვნელობაა მოცემული: საწყისი - $R_0 = 3.51$ და მოდელის შედეგად მიღებული $R_0 = 3.95$. თუმცა გასათვალისწინებელია, რომ R_0-ს მოდელით გაანგარიშებული მნიშვნელობა მის ფაქტობრივ მნიშვნელობას არ აღნიშნავს. იგი გვიჩვენებს რეპროდუქციის რიცხვის მნიშვნელობას, იმ შემთხვევაში, თუ შეკავების ღონისძიებები არ იქნებოდა გათვალისწინებული. მარჯვენა ნახაზი გვიჩვენებს მონტე კარლოს სიმულაციით მიღებული ინფიცირების ფუნქციის პირველი რიგის წარმოებულების ქცევას, ხოლო თარიღების მნიშვნელობებს ზემოთ მოყვანილი შინაარსი აქვთ. შავი ხაზი აღნიშნავს სიმულაციების საშუალო მნიშვნელობას. როდესაც

ფუნქციის პირველი რიგის წარმოებული უარყოფითი გახდება ინფიცირების ზრდის ტემპი შემცირდება და მდგომარეობა დასტაბილიზდება. უნდა აღინიშნოს, რომ პროგნოზი მგრძნობიარეა შეყვანილი პარამეტრების მიმართ. $\pi(t)$ ფუნქციისა სხვადასხვა სპეციფიკაციის შემთხვევაში პიკისა და წმინდა ინფიცირების განულების თარიღები იცვლება. ეს საშუალებას გვაძლევს ვიპოვოთ ბალანსი შეკავების ღონისძიებების სიმკაცრესა და ეპიდემიის ხანგრძლივობის ალტერნატივებს შორის. აქვე უნდა აღინიშნოს, რომ ჩვენს მიერ განსაზღვრული $\pi(t)$ ფუნქცია სინამდვილეს ზუსტად ვერ ასახავს და მაშასადამე პროგნოზირებაში გარკვეული გადაადგილება იქნება. მის ზუსტ სპეციფიკაციას ართულებს საქართველოს მთავრობის მიერ განხორციელებული მრავალმხრივი და მრავალჯერადი ღონისძიებები. ეს მოდელირებისათვის ორ პრობლემას წარმოშობს: ერთის მხრივ, რთულია მათი მასშტაბისა და ეფექტის გაზომვა, ხოლო მეორე მხრივ, მოდელისთვის რთული იქნება მცირე ინტერვალით განხორციელებული ღონისძიებების ეფექტების შეფასება.

საქართველოს მთავრობის მიერ კორონავირუსის წინააღმდეგ განხორციელებული ბრძოლის კამპანია საკმაოდ მკაცრი და შედეგიანი იყო. ხელისუფლება მყისიერად რეაგირებდა ინფიცირებულთა ყოველი ახალი კლასტერის გამოვლენაზე და ატარებდა პრევენციულ ღონისძიებებს, რათა მინიმუმამდე ყოფილიყო დაყვანილი ეპიდემიის გავრცელების საშიშროება. ამ ღონისძიებებში მოიაზრებოდა ინფიცირების გამოვლენის ტერიტორიის საზღვების ჩაკეტვა, მასობრივი კამპანია სახლში დარჩენისა და დამცავი საშუალებების გამოყენებისათვის და სხვ. რაც ნიშნავს იმას, რომ შეკავების ღონისძიებებს უწყვეტი სახე ჰქონდა. აქედან გამომდინარე, საქართველოს მონაცემებზე გაფართოებული SIR მოდელის აგებისას უმჯობესია $\pi(t)$ ფუნქციის სახედ შევარჩიოთ უწყვეტი ექსპონენციალური ფუნქცია.

უწყვეტი ექსპონენციალური ფუნქციის გამოყენებით, რომლის სახე ზემოთ განვსაზღვრეთ, არსებული შედეგები მნიშვნელოვნად განსხვავდება ზემოთ მიღებული შედეგებისაგან. მართალია, კუმულატიური წმინდა ინფიცირების კლების თარიღსა(28 აპრილი) და პიკის თარიღში (15 აპრილი) საგრძნობი სხვაობა არ შეინიშნება, თუმცა დიდი ცვლილებაა ინფიცირების დასრულების თარიღსა და მთლიან ინფიცირებულ მოსახლეობას შორის. კერძოდ, მთლიანი ინფიცირებულთა რაოდენობა 1555 ადამიანს

აღწევს, ხოლო ეპიდემიის დასრულების თარიღად 22 ივნისია პროგნოზირებული. (იხ. დანართი.3)

როგორც ჩანს შეკავების პოლიტიკის რბილი, მაგრამ მონოტონური გამკაცრება უფრო ეფექტური ღონისძიებაა ვიდრე გაცილებით მკაცრი, მაგრამ ეტაპობრივი, ერთჯერადი ღონისძიებები. სამწუხაროდ R-ის ეს პაკეტი არ იძლევა დღიურ პროგნოზსს, რის გამოც შეცდომის მეტრიკის მნიშვნელობის (RMSE) გამოთვლა არ შეგვიძლია.

2.2.2 ეპიდემიოლოგიური მოდელი კარანტინის განყოფილების დამატებით

SIR მოდელში შეკავების რეჟიმის გათვალისწინების ალტერნატიული გზა არის კარანტინის (quarantine) განყოფილების დამატება. ამ განყოფილებაში მოთავსებული ინდივიდებისათვის დაავადებულ ინდივიდებთან შეხვედრის ალბათობა ნულოვანია. კარანტინის განყოფილებაში ინდივიდები ხვდებიან დაავადების რისკის ქვეშ მყოფთა (susceptible) განყოფილებიდან და მცირდება ამ უკანასკნელთა რაოდენობა.

კარანტინის განყოფილების დამატებით SIR მოდელის განტოლებები გადაიწერება შემდეგი სახით:

$$\frac{d\theta_t^Q}{dt} = \varphi(t)\theta_t^S \tag{2.2.15}$$

$$\frac{d\theta_t^S}{dt} = -\beta\theta_t^S\theta_t^I - \varphi(t)\theta_t^S \tag{2.2.16}$$

$$\frac{d\theta_t^I}{dt} = \beta\theta_t^S\theta_t^I - \gamma\theta_t^I \tag{2.2.17}$$

$$\frac{d\theta_t^R}{dt} = \gamma\theta_t^I \tag{2.2.18}$$

$$\theta_t^S + \theta_t^Q + \theta_t^I + \theta_t^R = 1 \tag{2.2.19}$$

სადაც $\varphi(t)$ დროში ცვალებადი კოეფიციენტია, რომელიც აღნიშნავს რისკის ქვეშ მყოფთა იმ წილს, რომლებიც იცავენ სრული თვითიზოლაციის წესებს. დაშვების თანახმად $\varphi(t)$ არის დირახ დელტას საფეხუროვანი ფუნქცია, რომელიც მნიშვნელობას იცვლის ქვეყნის მასშტაბით მაკრო საკარანტინო ღონისძიებების განხორციელების თარიღის მიხედვით.

საქართველოში განხორციელებული ღონისძიებების გათვალისწინებით ფუნქცია განისაზღვრა შემდეგი სახით:

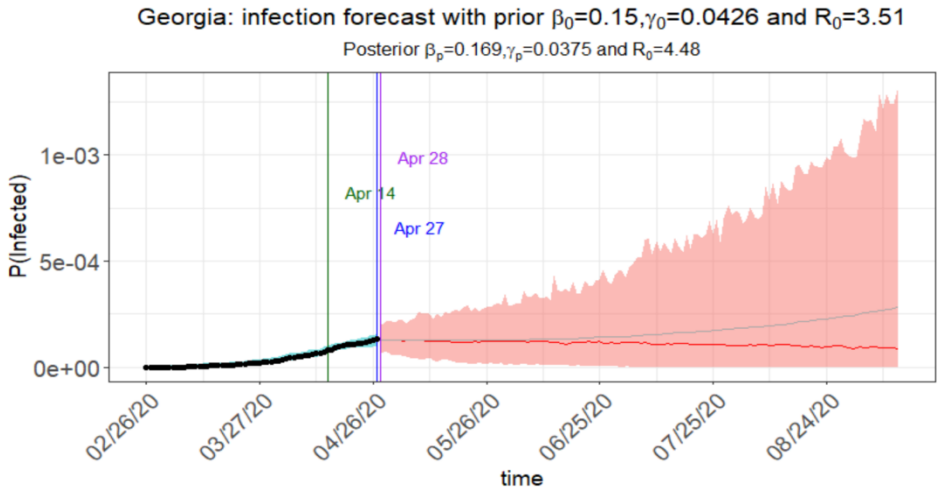
$$\varphi(t) = \begin{cases} 0.1 & თუ\ t = 21.03.2020 \quad პრეზიდენტის\ დეკრეტი \\ 0.1 & თუ\ t = 23.03.2020 \quad ბოლონისის\ ჩაკეტვა \\ 0.5 & თუ\ t = 31.03.2020 \quad კარანტინის\ შემოღება \\ 0.1 & თუ\ t = 10.04.2020 \quad ლეჩხუმის\ ჩაკეტვა \\ 0.4 & თუ\ t = 16.04.2020 \quad ოთხი\ დიდი\ ქალაქის\ ჩაკეტვა \end{cases}$$

შესაბამისად, კარანტინში (თვითიზოლაციაში) მყოფი მოსახლეობის მთლიანი წილი იქნება: $0.1\theta_{t1}^{S} + 0.1\theta_{t2}^{S} + 0.5\theta_{t3}^{S} + 0.1\theta_{t4}^{S} + 0.4\theta_{t5}^{S}$.

გარდა შეკავების ღონისძიებების გათვალისწინებისა, მოდელების უცნობი პარამეტრების და მათი ნდობის ინტერვალის შესაფასებლად გამოყენებულია მარკოვის ჯაჭვის მონტე კარლოს ალგორითმი (Markov Chain Monte Carlo (MCMC)). იგი საშუალებას იძლევა ვიპროგნოზოთ ინფიცირების ზრდის შეჩერების თარიღი და მაქსიმალური კუმულატიური წმინდა ინფიცირების (კუმულატიურ ინფიცირებას გამოკლებული კუმულატიური გამოჯანმრთელება და გარდაცვალება) თარიღი, რომლებიც მათემატიკურად წარმოადგენენ θ_t^I-ს პირველი და მეორე რიგის გადაღუნვის წერტილებს და $\dot{\theta}_t^I = 0$ და $\ddot{\theta}_t^I = 0$ ამონახსნებს, შესაბამისად. რეპროდუქციის კოეფიციენტი $R_0 = \beta/\gamma$ სადაც β და γ მოღებულია ზემოხსენებული განაწილებიდან. რადგან მოდელში გათვალისწინებულია შეკავების ღონისძიებები, R_0 შეიძლება შეიცვალოს განსხვავებული პროტოკოლის მიხედვით. დავუშვათ, t_0 არის ბოლო დღე, როდესაც შეგვიძლია დავაკვირდეთ ინფიცირებულთა და გამოჯანმრთელებულთა წილს $(Y_{0:to}^I, Y_{0:to}^R)$. $t \in [t_0 + 1, T]$ საროგნოზე პერიოდზე $Y_t^I -$ სა და Y_t^R-ს M რაოდენობის სიმულაციის პროგნოზირებისათვის განვახორციელებთ შემდეგ პროცედურებს: თითოეული m=1, 2, …, M-სათვის:

1. წინა $[\theta_t|\theta_{t-1}^{(m)}, \tau^{(m)}]$ გავრცელების პროცესიდან მივიღებთ $\theta_t^{(m)}$მნიშვნელობებს t= $t_0 + 1, \dots, T$-სთვის.
2. $[Y_t^I|\theta_{t-1}^{(m)}, \tau^{(m)}]$-სა და $[Y_t^R|\theta_{t-1}^{(m)}, \tau^{(m)}]$-დან მივიღებთ $(Y_t^{I(m)}, Y_t^{R(m)})$ პროგნოზს დაკვირვებადი მნიშვნელობების მიხედვით, t=$t_0 + 1, \dots, T$-სთვის.

ზემოთ განსაზღვრული $\varphi(t)$ ფუნქციის მიხედვით საქართველოს მონაცემებზე მორგებული მოდელის შედეგები ნაჩვენებია გრაფიკი.3-ზე.



გრაფიკი.3

მიუხედავად იმისა, რომ დაავადების რისკის ქვეშ მყოფი მოსახლეობის (S) 78%-ზე მეტი ეტაპობრივად კარანტინის განყოფილებაში გადადის, შეკავების ღონისძიების შედეგი არაეფექტურია. ეპიდემიის დასრულების სავარაუდო თარიღი 2020 წლის 6 სექტემბერია, ინფიცირებულთა საერთო რაოდენობა ამ დროისათვის 4500-ს აღწევს. ეს შედეგი კიდევ ერთხელ ადასტურებს იმ ვარაუდს, რომ ერთჯერადი, თუნდაც მკაცრი ღონისძიებები არაეფექტიანია ეპიდემიასთან საბრძოლველად. უფრო მეტიც, თუ კარანტინის რეჟიმი ზედმეტად მკაცრ პირობებში მიმდინარეობს, შესაბამისი საინფორმაციო კამპანიის გარეშე, ამან შეიძლება უკუშედეგი მოგვცეს. საზოგადოებამ შეიძლება დაკარგოს ნდობა და მოთმინება და არ დაიცვას კარანტინის რეჟიმი.

2.3 SEIR მოდელი

ინფექციური დაავადების გარკვეული ნაწილისათვის დამახასიათებელია ე. წ. ინკუბაციის პერიოდი, როდესაც ინდივიდი არის ინფექციის მატარებელი სიმპტომების გარეშე. მსგავსი ტიპის ეპიდემიის მოდელირებისათვის იყენებენ SEIR (Susceptible – Exposed – Infectious – Removed) მოდელს, რომელიც წარმოადგენს SIR მოდელის გაფართოებას დაუცველთა (exposed) ჯგუფის დამატებით. დაშვების თანახმად ინკუბაციის პერიოდი არის ექსპონენციალური განაწილების მქონე შემთხვევითი ცვლადი a პარამეტრით. (ე.ი. საშუალო ინკუბაციის პერიოდი არის a^{-1}). გარდა ამისა

მოდელი ითვალისწინებს შობადობისა (δ) და მოკვდავობის (μ) კოეფიციენტების დინამიკას. მოდელი შემდეგი დიფერეციალური განტოლებებით აღიწერება:

$$\begin{aligned} \frac{dS}{dt} &= \delta - \mu S - \beta \frac{I}{N} S \\ \frac{dE}{dt} &= \beta \frac{I}{N} S - (\mu + a)E \\ \frac{dI}{dt} &= aE - (\gamma + \mu)I \\ \frac{dR}{dt} &= \gamma I - \mu R \end{aligned} \qquad (2.3.1)$$

სადაც S+E+I+R=N მხოლოდ იმ შემთხვევაში თუ შობადობისა და მოკვდაობის კოეფიციენტი ერთმანეთის ტოლია. ამ მოდელისათვის საბაზისო რეპროდუქციის კოეფიციენტი შემდეგი სახით გამოითვლება:

$$R_0 = \frac{a}{a + \mu} * \frac{\beta}{\mu + \gamma} \qquad (2.3.2)$$

SEIR მოდელის უპირატესობა ისაა, რომ იგი უფრო რეალურად ასახავს ეპიდემიის განვითარების პროცესს, ამასთანავე ითვალისწინებს ბუნებრივ მოძრაობასაც. მისი მთავარი ნაკლი SIR მოდელთან შედარებით გაანგარიშებებისათვის საჭირო მონაცემების არ არსებობაა. მოდელირებისათვის დამატებით საჭიროა ინკუბაციის პერიოდის ემპირიული მონაცემების არსებობა ან პარამეტრების მნიშვნელობების წინასწარ ცოდნა. გამომდინარე იქედან, რომ არც მონაცემები არსებობს და არც პარამეტრების მნიშვნელობებია ცნობილი SEIR მოდელის აგება ვერ მოხერხდება.

3. ექსპონენციალური და S ფორმის მოდელები

3.1 პროგნოზი ლოგარითმულ-წრფივი მოდელის საშუალებით

გადამდები მწვავე რესპირატორული ინფექციების გავრცელებას ხშირად ექსპონენციალური ფორმა აქვს, რის გამოც მკვლევარები მათი მოდელირებისა და გავრცელების პროგნოზირებისათვის იყენებენ ექსპონენციალურ ფუნქციას. ამ ქვეთავში აღვწერთ ლოგარითმულ-წრფივ მოდელს და მისი საშუალებით გამოვითვლით რამდენიმე საინტერესო პარამეტრის მნიშვნელობას. მოდელს ქვემოთ მოცემული მარტივი სახე აქვს:

$$\log(y) = r * t + b, \qquad (3.1.1)$$

სადაც, r ზრდის ტემპია, y ინფიცირებულთა რაოდენობა (არაკუმულატიური), t აღნიშნავს ეპიდემიის დაწყებიდან გასული დღეების რაოდენობას, ხოლო b თავისუფალი წევრია. მოდელის აგებას განვახორციელებთ სტატისტიკური პროგრამა R-ის "incidence" პაკეტის საშუალებით, რომელიც შექმნილია სპეციალურად ეპიდემიების მოდელირებისათვის RECON (**R** **E**pidemic **CON**sortium) ჯგუფის მიერ. ამ პაკეტის საშუალებით შეგვიძლია განვსაზღვროთ გარდატეხის (პიკის) წერტილი, ინფიცირებულთა რაოდენობის განახევრებისათვის (ან გაორმაგებისათვის, თუ ზრდის ტენდენცია შეინიშნება) საჭირო დრო, ეპიდემიის ზრდის ტემპი (95%-იანი ნდობის ინტერვალით) და რეპროდუქციის საბაზისი რიცხვის მნიშვნელობაც თუ სერიული ინტერვალის (serial interval) განაწილება არის ცნობილი.

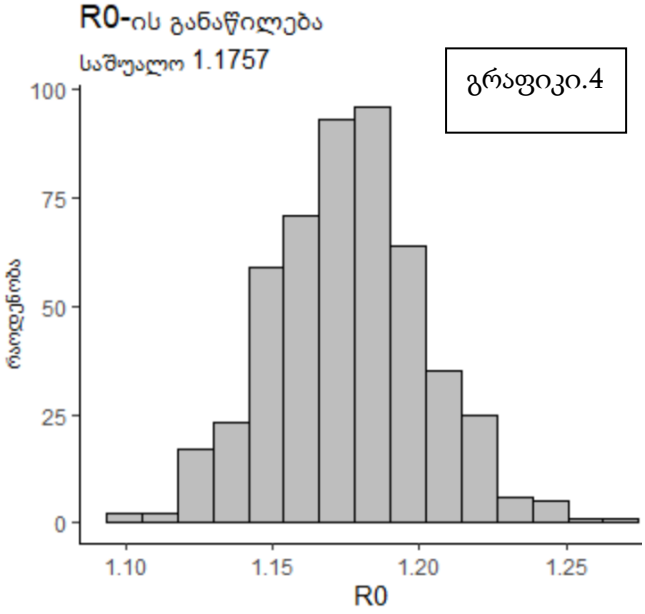
მოდელის მიხედვით პიკი 16 აპრილს დაფიქსირდა (იხ. დანართი.4). პიკამდე ინფიცირების საშუალო ზრდის ტემპი 5.32% (95%-იანი ნდობის ინტერვალი 3.686% - 6.96%) იყო, ხოლო პიკის შემდეგ კლების ტემპი -3.3% ((-15.55%) - 8.955%). ინფიცირებულთა (არაკუმულატიური) რაოდენობის განახევრებისთვის საჭირო დრო 21 დღეა. აღსანიშნავია, რომ კლების კოეფიციენტის 95%-იან ნდობის ინტერვალში ნული ექცევა, რაც ნიშნავს იმას, რომ ინფიცირების კლებისკენ წასვლაზე დამაჯერებელი საუბარი ნაადრევია.

3.2 რეპროდუქციის რიცხვი - R_0

ეპიდემიის გავრცელების პროგნოზირებისათვის ერთ-ერთი მნიშვნელოვანი კოეფიციენტია R_0, რომლის საშუალებითაც შეგვიძლია განვსაზღვროთ ეპიდემიის ტრაექტორია (ზრდა / კლება) და მისი სიმწვავე. კორონავირუსის R_0-ის შესაფასებლად მრავალი კვლევა ჩატარდა სხვადასხვა ქვეყანაში (Vaidyanathan, 2020) (Soetewey, 2020) (Ellis, 2020) (Caicedo-Ochao & al., 2020) (Zhuang, Zhao, & al., 2020). შედეგების სპექტრი საკმაოდ ფართო და არაერთგვაროვანი იყო. რაც განაპირობა, ერთი მხრივ, კვლევის მეთოდების არაერთგვაროვნებამ, მეორე მხრივ, მონაცემების სტრუქტურაში არსებულმა განსხვავებამ. მიმდინარე ქვეთავში რამდენიმე მეთოდით შევაფასებთ რეპროდუქციის რიცხვის მნიშვნელობას საქართველოსთვის, ასევე ვისაუბრებთ შედეგებში არსებულ განსხვავებების გამომწვევ მიზეზებზე.

რეპროდუქციის რიცხვის გამოსათვლელად საჭიროა სერიული ინტერვალის განაწილების პარამეტრების (საშუალო და სტანდარტული გადახრა) ცოდნა. სერიული ინტერვალი გვიჩვენებს გადაცემის ჯაჭვში შემთხვევებს შორის დროს. მისი მოდელირებისათვის ხშირად მიმართავენ გამა ან ვეიბულის განაწილებას. შესაბამისი მონაცემების არ არსებობის გამო მისი გაანგარიშება საქართველოს შემთხვევაზე არ შეგვიძლია და უნდა გამოვიყენოთ სერიული ინტერვალის პარამეტრების ლიტერატურაში გავრცელებული მნიშვნელობები. კორონავირუსის სერიული ინტერვალის მოდელირებისათვის რამდენიმე კვლევა ჩატარდა (Nishiura & al., 2020) (Du, Xu, & al., 2020) (Baum, 2020). მისი საშუალოსა და სტანდარტული გადახრის მნიშვნელობების შეფასებისას საკმაოდ არაერთგვაროვანი შედეგები იქნა მიღებული. საშუალოს მნიშვნელობა მერყეობდა 3.96-სა [ზანვეი დუ და სხვ. (Du, Xu, & al., 2020)] და 7.5-ს [ქუნ ლი და სხვ. (Li & Guan, 2020)] შორის, ხოლო სტანდარტული გადახრის მნიშვნელობები 2.2.2-სა და 4.75 (Du, Xu, & al., 2020) შორის. ჩვენ ვიხელმძღვანელებთ ჰაროში ნიშიურასა და სხვ. (Nashiura & al., 2020) მიერ მიღებული პარამეტრების შეფასებით: საშუალო - 4.7, სტანდარტული გადახრა - 2.2.2, რომელიც ახლოს არის ჯანდაცვის მსოფლიო ორგანიზაციის შეფასებასთან (Hamidouche, 2020). R-ის „discrete" და "epitrix" პაკეტების საშუალებით ავაგეთ რეპროდუქციის რიცხვის განაწილება,

რომელიც წარმოდგენილია გრაფიკი.4-ზე. საშუალო მნიშვნელობა 1.1757-ია, ხოლო მედიანური - 1.1753. R_0-ის ეს მნიშვნელობა ძლიერ განსხვავდება SIR მოდელით მიღებული მნიშვნელობისაგან. ამას რიგი ობიექტური წინაპირობები გააჩნია. პირველ



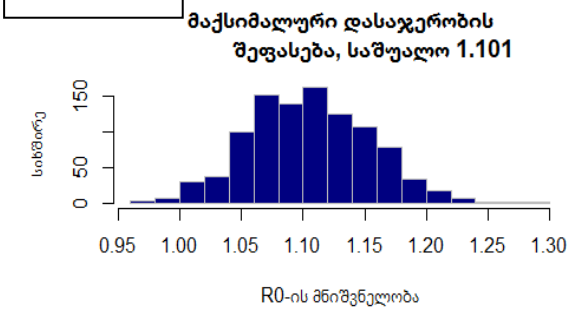
გრაფიკი.4

რიგში, საქართველოში კორონავირუსის გავრცელების ტრენდი დიდად განსხვავდება კლასიკური ეპიდემიის გრაფიკისაგან. ეს განპირობებულია მთავრობის მყისიერი და კარგად ორგანიზებული შეკავების ღონისძიებებით. ვირუსი საქართველოში შემოსვლამდე სამი თვით ადრე გავრცელდა მსოფლიოში, ამ პერიოდის განმავლობაში შესწავლილ იქნა გადაცემის გზები და მისგან თავის დაცვის საშუალებები. პრეპარატის არ არსებობის პირობებში, ეპიდემიის გავრცელების პრევენციის ყველაზე ეფექტურ გზა სოციალური დისტანცირებაა, რაც დიდად აზარალებს ეკონომიკას. ამიტომ, რიგმა სახელმწიფოებმა ეპიდემიის გამწვავევამდე უარი თქვეს სოციალური დისტანცირების ღონისძიების გატარებამდე. ჩვენთან შეკავების ღონისძიებები მყისიერად განხორციელდა, რამაც ეპიდემიას თავისუფლად გავრცელების საშუალება არ მისცა. ინფიცირების კერები იკეტებოდა და შესაბამისად იზღუდებოდა გავრცელების არეალი. ამიტომ, ეპიდემიის მახასიათებლების მოდელირებისათვის მთლიანი ქვეყნის ერთ სივრცედ განხილვა და ინფიცირების რისკის ქვეშ მყოფთა (S) კატეგორიაში მთლიანი მოსახლეობის ჩართვა გარკვეულწილად არამიზანშეწონილია. ამ გარემოების გათვალისწინებით აშკარა იყო, რომ SIR მოდელი გამა და ბეტა პარამეტრების არასწორ შეფასებას მოგვცემდა. გამა პარამეტრის საშუალო მნიშვნელობა მეტ-ნაკლები სანდოობით ცნობილი იყო და ეს მნიშვნელობა მივუთითეთ. არარელევანტური მონაცემების გამო, ერთ-ერთი პარამეტრის დაფიქსირებამ გამოიწვია რეპროდუქციის რიცხვის გადაჭარბებული შეფასება. როდესაც მოდელს მივეცით საშუალება ორივე პარამეტრის მნიშვნელობა მხოლოდ

მონაცემებზე დაყრდნობით განესაზღვრა, მიღებული R_0-ის მნიშვნელობა (R_0 =1.12) ბევრად უფრო მიუახლოვდა ზემოთ მიღებულ მნიშვნელობას.

R_0-ის შეფასების კიდევ ერთი მეთოდია მაქსიმალური დასაჯერობის მეთოდი. ეს მეთოდიც მოითხოვს სერიული ინტერვალის განაწილების პარამეტრებს და ამ შემთხვევაშიც იგივე პარამეტრები იქნება გამოყენებული მოდელირებისათვის. უფრო საიმედო შეფასების მისაღებად გამოვიყენებთ 1000 შერჩევის მქონე ბუთსთრეფ (bootstrap) სიმულაციას. სიმულაციის შედეგად მიღებული R_0-ების საშუალო მნიშვნელობა 1.101-ის ტოლია. (იხ. გრაფიკი.5).

გრაფიკი.5

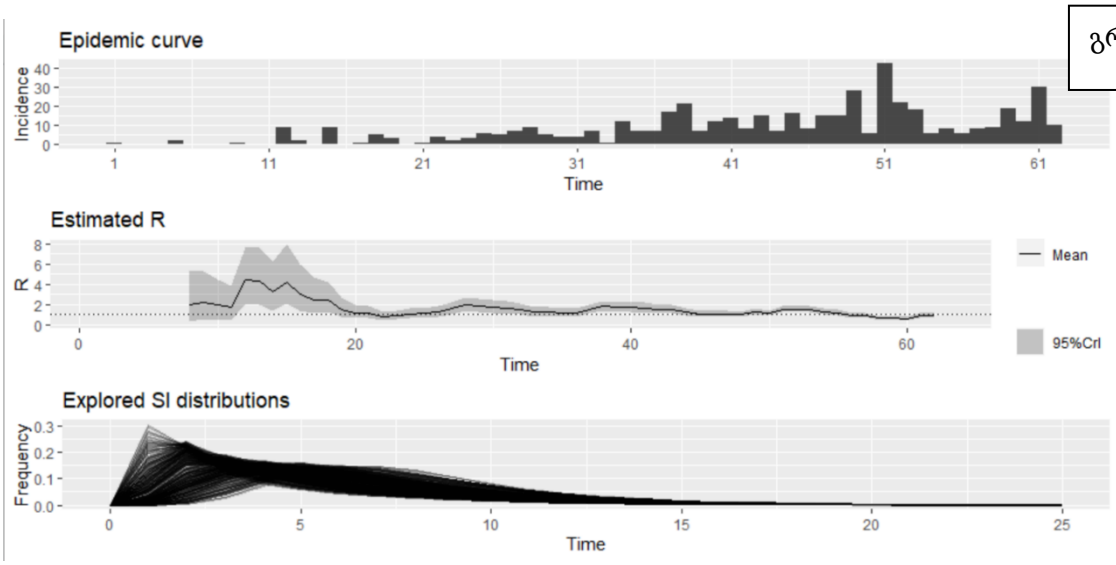


რეპროდუქციის საბაზისო კოეფიციენტს გააჩნია გარკვეული ნაკლოვანება. როგორც ავღნიშნე მისი მნიშვნელობა დამოკიდებულია დაავადების რისკის ქვეშ მყოფთა (S) წილზე მთლიან მოსახეობაში. რეალურად ამ კატეგორიის წილი მუდმივი არ არის, რაც განპირობებულია შეკავების ღონისძიებებით. ცვლილება გავლენას ახდენს კოეფიციენტის მნიშვნელობაზეც. R_0 მუდმივი რიცხვია და არ შეუძლია აღწეროს დინამიკა, ამიტომ მასზე დაყრდნობა მხოლოდ ეპიდემიის ადრეულ ეტაპზეა მიზანშეწონილი. ეპიდემიის დინამიკის კარგი მახასიათებელია ეფექტური რეპრო-დუქციის კოეფიციენტი, ან როგორც ზოგჯერ მოიხსენებენ - დროში ცვალებადი რეპროდუქციის კოეფიციენტი. საქართველოში განხორციელებული მრავალმხრივი პრევენციული ღონისძიების შედეგად R_0-ის მნიშვნელობა იცვლებოდა. ამიტომ, უმჯობესია განვიხილოთ ეფექტური რეპროდუქციის კოეფიციენტი. მის შესაფასებლად გამოვიყენებთ R-ის „EpiEstim” პაკეტს. შეფასებისთვის კვლავ დაგვჭირდება სერიული ინტერვალის განაწილება.

განსხვავებით ზემოგანხილულისა ახლა არ დავეყრდნობით სერიული ინტერვალის (SI) პარამეტრების წერტილოვან შეფასებას. ნაცვლად განვიხილავთ SI-ის საშუალოსა და სტანდარტული გადახრის ინტერვალურ შეფასებას. SI-ს საშუალოს შემდეგი განაწილება აქვს: { საშუალო 4.7; მინიმალური მნიშვნელობა 3.96; მაქსიმალური მნიშვნელობა 7.5; სტანდარტული გადახრა - 2.}. SI-ის სტანდარტულ გადახრას შემდეგი

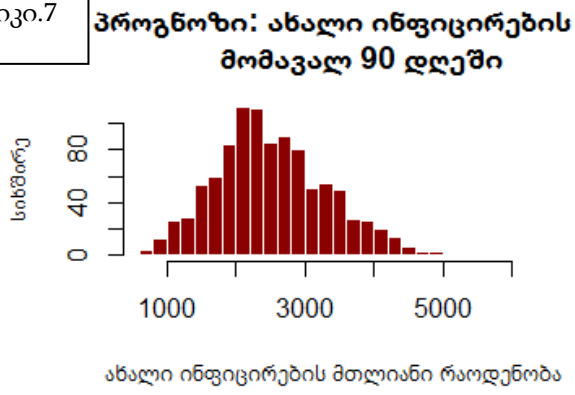
განაწილება აქვს: {საშუალო - 2.2.2, მინიმალური მნიშვნელობა - 2.2.1, მაქსიმალური მნიშვნელობა - 4.7, სტანდარტული გადახრა - 1}.

გრაფიკი.6



პირველი გრაფიკი ასახავს დღიური ინფიცირების რაოდენობებს. მეორე გრაფიკზე წარმოდგენილია ეფექტური რეპროდუქციის რიცხვი, რომელიც გამოითვლება შვიდ დღიანი ფანჯრის ინტერვალით. მესამე გრაფიკზე კი მოცემულია სერიული ინტერვალის სიმულაციის შედეგი, რომლის პარამეტრიზაცია ზემოთ ავღწერეთ.

გრაფიკი.7



სერიული ინტერვალის, ყოველდღიური ინფიცირების მონაცემებისა და რეპროდუქციის რიცხვის საშუალებით შეგვიძლია განვახორციელოთ ეპიდემიის ტრაექტორიის პროგნოზი. მოდელი ეყრდნობა განშტოების პროცესს (brancing process), სადაც დღიური ინფიცირების რაოდენობები მიჰყვებიან პუასონის პროცესს და გამოითვლებიან შემდეგი სახით:

$$\lambda = y_s w(t-s) \qquad (11.2.1)$$

სადაც w არის დისკრეტული სიმკვრივის ფუნქცია, ხოლო y_s ახალი ინფიცირების რაოდენობა დროის s მომენტში. გრაფიკი.7-ზე წარმოდგენილი შედეგები მიღებულია 1000 შერჩევის მქონე სიმულაციის შედეგად. ინფიცირებულთა ჯამური რაოდენობა 90

დღის შემდეგ 2512 იქნება. მოდელს გააჩნია ნაკლოვანება, მას არ შეუძლია ეპიდემიის დასრულების თარიღის პროგნოზირება.

3.3 S ფორმის მრუდები

გადამდები ეპიდემიის მოდელირებისას ასევე ხშირად იყენებენ S ფორმის მრუდებს. აღნიშნული ტიპის მრუდები აღწერენ პროცესებს, რომელიც საწყის ეტაპზე ექსპონენციალურად იზრდება, ხოლო გარკვეული დროის შემდეგ, როდესაც ეპიდემია მიაღწევს პიკს, ზრდის ტემპი მცირდება. პიკის წერტილში მრუდი გადაიღუნება და ეპიდემია კლებას იწყებს. S ფორმის მრუდების სახეებია: ლოგისტიკური მრუდი, ვერჰულსტ-პირლის განტოლება, პირლის მრუდი, ზრდის მრუდი, რიჩარდის ზრდის მოდელი, გომპერსის მრუდი, სიგმოიდური მრუდი, ფოსტერის მრუდი, ბასის მოდელი და ა.შ. მოცემულ ნაშრომში განვიხილავთ რიჩარდის მრუდს, გომპერსის მრუდსა და ლოგისტიკურ მრუდს. ამ უკანასკნელის ნაკლოვანებაა სიმეტრიულობის დაშვება გადაღუნვის წერტილის მიმართ, რაც აღმოფხვრილია რიჩარდის მოდელში.

ეპიდემიის საწყის ეტაპზე, შეზღუდული მონაცემების პირობებში, როდესაც მრუდის გადაღუნვის წერტილი გავლილი არ არის მრავალპარამეტრიანი მოდელის შეფასება არამიზანშეწონილია. ამიტომ, უპირველეს ყოვლისა, შევაფასებთ განზოგადო-ებული ლოგისტიკურ მოდელს (GLM), რომელიც უფრო პარსიმონული არის პარამეტრიზაციის მხრივ.

3.3.1 განზოგადოებული ლოგისტიკური მოდელი

ლოგისტიკური მოდელი აღწერს ინფიცირებულ პირთა დინამიურ ევოლუციას, მოცემული ზრდის ტემპისა და მოსახლეობის პირობით. ლოგისტიკური მრუდი აღიწერება შემდეგი სახის ფუნქციით (Tsoularis & Wallace, 2002):

$$f(x) = \frac{KP_0e^{rt}}{K + P_0(e^{rt} - 1)} \tag{3.3.1}$$

რომელიც მიიღება ჩვეულებრივი დიფერენციალური განტოლებიდან:

$$\frac{dP}{dt} = rP(1 - \frac{P}{K}) \tag{3.3.212}$$

მოდელი აღწერს ინფიცირებულთა დინამიკას, P-ს, რომელიც დამოკიდებულია r ზრდის ტემპსა და ინფიცირებულთა საწყის მნიშვნელობაზე, P_0-ზე. K წარმოადენს მრუდის მაქსიმალურ მნიშვნელობას. მოდელის შესაფასებლად გამოიყენება უმცირეს კვადრატთა მეთოდი.

განზოგადოებული ლოგისტიკური მოდელს შემდეგი სახე აქვს:

$$\frac{dC(t)}{dt} = rC^p(t)(1 - \frac{C(t)}{K}) \qquad (3.3.3)$$

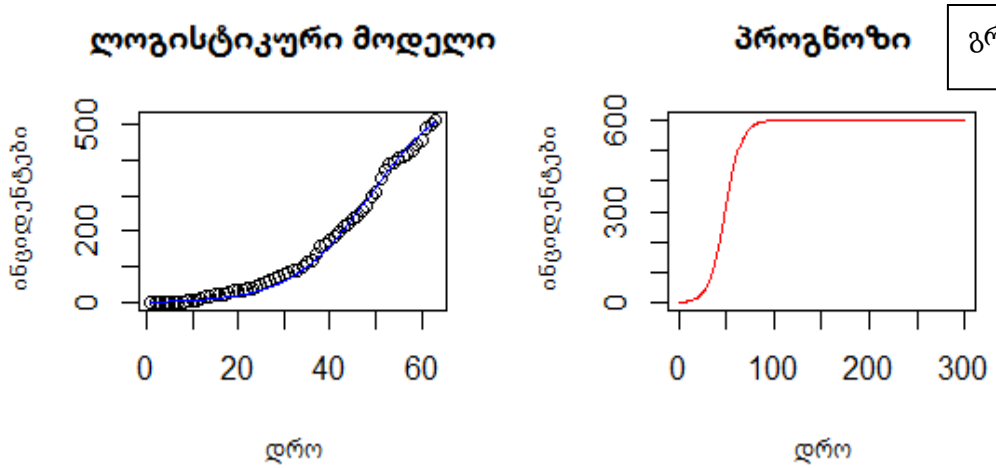
დაშვების მიხედვით (3.3.2) და (3.3.3) მოდელებში ინციდენტების ზრდის ტემპი გარკვეულ ეტაპზე შემცირედება და მრუდი გადაიღუნება. აღნიშნული დაშვება არაა სამართლიანი განზოგადოებული ზრდის მოდელის შემთხვევაში, რომელიც შემდეგი განტოლებით აღიწერება:

$$\frac{dC(t)}{dt} = rC^p(t) \qquad (3.3.4)$$

ამ მოდელით შესაძლებელია ნახევრად ექსპონენციალური ზრდის მოდელირება ($p<1$ შემთხვევაში), თუმცა მრუდის გადაღუნვის წერტილის განსაზღვრა ვერ ხერხდება. ორივე ლოგისტიკური ტიპის მოდელი მიდრეკილია ეპიდემიის მასშტაბის ნაკლებობით შეფასებისკენ, მაშასადამე ისინი შეგვიძლია გამოვიყენოთ, როგორც შეფასების ქვედა ზღვრები. ამასთანავე, განზოგადოებული ლოგისტიკური მოდელი ადრეული ნახევრად-ექსპონენციალური ზრდის მოდელირების საშუალებას იძლევა და შეუძლია უკეთესად აღწეროს შესაძლო ასიმეტირია ზრდისა და კლების დინამიკას შორის.

მოცემულ ქვეთავში განხილული მოდელები (ლოგისტიკური, რიჩარდის, გომპერსის) შეფასებულია R-ში growthrates და growthmodels პაკეტებით. სამივე მოდელის შემთხვევაში მოდელის აგება ხდება შემდეგი სახით: ხდება პარამეტრების საწყისი მნიშვნელობების ინიციალიზაცია. პარამეტრების მორგებისათვის growthrates იყენებს FME (Flexible Modelling Environment) პაკეტს. მიღებული პარამეტრებით ხდება მომავალი მნიშვნელობების პროგნოზირება წინასწარ განსაზღვრული დროითი ჰორიზონტისთვის. პროგნოზირებისთვის გამოყენებულია growthmodels პაკეტი.

ლოგისტიკური მრუდის მოდელის შედეგი მოცემულია გრაფიკი.8-ზე:



გრაფიკი.8

მოდელი მონაცემებს მაღალი ხარისხით მოერგო, კერძოდ, დეტერმინაციის კოეფიციენტი, $R^2 = 0.995$. მოდელის მიხედვით ეპიდემია დასრულდება მაისის ბოლოს. სულ დაინფიცირდება 600 ადამიანი. გადაღუნვის ანუ პიკის წერილი იქნება ეპიდემიის დაწყებიდან 49-ე დღეს. (Goshu & Koya, 2013) ცხადია, ეპიდემიის სიმკაცრე ნაკლებობითაა პროგნოზირებული და ეს შედეგი შეესაბამება ზემოთ აღნიშნულ ფაქტს, რომ ლოგისტიკური მოდელები შეიძლება გამოვიყენოთ ერთგვარ ქვედა ზღვრად. მოდელის პროგნოზირების ხარისხი შევაფასეთ მომავალი 30 დღის მონაცემებზე. მოდელმა საკმაოდ ცუდი სიზუსტე აჩვენა, კერძოდ, საშუალო კვადრატული შეცდომიდან ფესვის (RMSE) მნიშვნელობა 51.69-ის ტოლი იყო. ამ შედეგების გათვალისწინებით შეგვიძლია დავასკვნათ, რომ მოდელს აქვს ეპიდემიის ტრენდის მახასიათებლების (პიკის წერტილი, ინფიცირებულთა ჯამური რაოდენობა ეპიდემიის დასრულებისას, დასრულების თარიღი) შედარებით კარგად პროგნოზირების უნარი, თუმცა ყოველდღიურ პროგნოზს ზუსტად ვერ აკეთებს. სხვა კუთხით რომ შევხედოთ ამ საკითხს, მრუდი ეპიდემიას ახასიათებს გლუვი ტრენდით, რომლის ირგვლივ ირხევა რეალური მნიშვნელობები და ამიტომ გვაძლევს დიდ შეცდომას. თუმცა გრძელვადიან პერიოდში გადახრები ერთმანეთს აბათილებენ და ინფიცირებულთა საბოლოო რაოდენობა შედარებით ახლოსაა პროგნოზირებულ მნიშვნელობებთან. უნდა აღინიშნოს, რომ აღნიშნული პრობლემის წინაშე ვდგავართ გომპერისა და რიჩარდის მრუდების შემთხვევაშიც.

3.3.2 გომპერსის მოდელი

სიგმოიდური კლასის მრუდების ოჯახს განეკუთვნება კიდევ ერთი მოდელი - გომპერსის მრუდი, რომელიც 1825 წელს შეიქმნა ბენჟამინ გომპერსის მიერ ადამიანთა მოკვდაობის პროცესის აღსაწერად. გომპერსის მრუდი წარმატებით გამოიყენება ბიო-ლოგიაში სხვადასხვა მიმართულებით, მათ შორის, ცხოველთა და მცენარეთა ზრდის პოცესის, ბაქტერიებისა და კიბოს უჯრედების გამრავლების მოდელირებისათვის. იგი აღწერს ზრდის პროცესს, რომელიც პერიოდის საწყის და საბოლოო ეტაპზე ნელია. (Tjorve, 2017)ამასთანავე, ფუნქციის მარჯვენა (მომავლის) ასიმპტოტი უფრო დაბალი დახრილობისაა და ნელა უახლოვდება დასასრულს, ვიდრე საწყისი ასიმპტოტი, მაშასადამე, იგი არასიმეტრიულია. გომპერსის მრუდი აღიწერება შემდეგი სახით:

$$f(t) = ae^{-be^{-ct}} \tag{3.3.5}$$

სადაც

- a არის ფუნქციის ასიმპტოტი: $\lim_{t\to\infty} ae^{-be^{-ct}} = ae^0 = a$
- b განსაზღვრავს მრუდის მდებარეობას x ღერძის გასწვრივ. $b = \log(2)$ შემთხვევაში მრუდი სიმეტრიულია.
- c წარმოადგენს ზრდის ტემპს
- e ეილერის რიცხვია

ეპიდემიის შუაწერტილის შესაბამისი დროითი წერტილის საპოვნელად გამოიყენება შემდეგი ფორმულა:

$$t_{hwp} = -\frac{\ln(\frac{\ln(2)}{b})}{c} \tag{3.3.6}$$

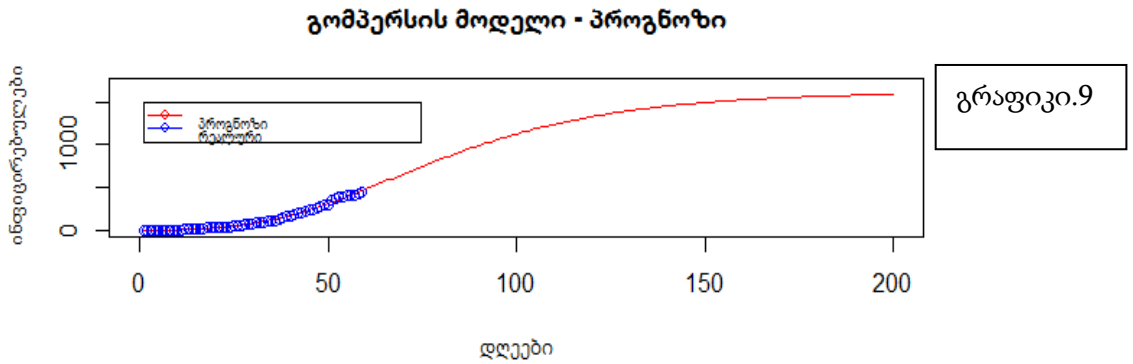
პიკის (მაქსიმალური ზრდის) წერტილი წარმოადგენს შემდეგი განტოლების ამონახსნს t-ს მიმართ:

$$\frac{d^2}{dt^2} f(t) = 0 \tag{3.3.713}$$

საიდანაც:

$$t_{max} = \ln(b)/c \tag{3.3.8}$$

გომპერსის მოდელის შედეგი მოცემულია გრაფიკ.9-ზე:



გრაფიკი.9

მოდელი მონაცემებს მაღალი ხარისხით მოერგო, $R^2 = 0.996$, რაც უმნიშვნელოდ აღემატება ლოგისტიკური მოდელის სიზუსტეს. მოდელის მიხედვით სულ დაინფიცირდება 1614 ადამიანი. ეპიდემია დასრულდება ივლისის ბოლოს, ხოლო პიკი მოხდება 66-ე დღეს, ანუ პირველ მაისს. მართალია მოდელმა ეპიდემიის ტრენდი და ტრაექტორია მეტ-ნაკლებად სწორად აღწერა თუმცა ძალიან დიდია საშუალო კვადრატული შეცდომიდან ფესვის მნიშვნელობა - RMSE=139.19, რომლის მიზეზებზეც უკვე ვისაუბრე.

3.3.3 რიჩარდის მოდელი

სამპარამეტრიანი რიჩარდის ზრდის მოდელი გამოყენებულ იქნა სხვადასხვა ლოგისტიკური ფორმის ეპიდემიოლუგიური მრუდების მოდელირებისათვის (Hsieh, 2009). განზოგადოებული რიჩარდის მოდელი განისაზღვრება შემდეგი დიფერენციალური განტოლებით:

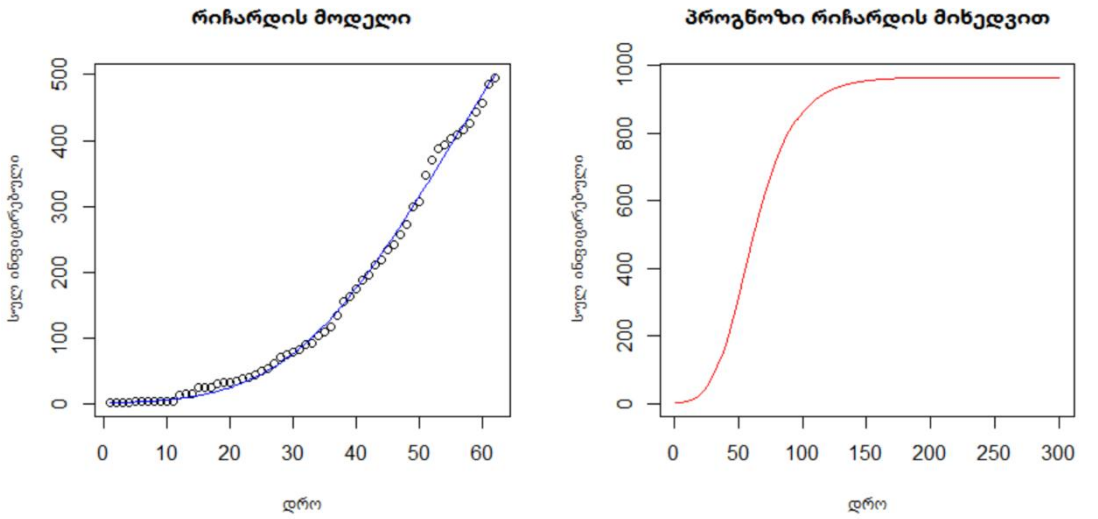
$$\frac{dC(t)}{dt} = r[C(t)]^p\left(1 - \left(\frac{C(t)}{K}\right)^a\right) \qquad (3.3.9)$$

სადაც C(t) აღნიშნავს ინფიცირებულთა კუმულატიურ რიცხვს t დროისთვის, r არის ზრდის ტემპი ადრეულ ეტაპზე და K არის ინფიცირებულთა ჯამური რაოდენობა ეპიდემიის დასრულებისას. $p \in [0,1]$ არის პარამეტრი, რომელიც ზრდის სხვადასხვა ფორმის მოდელირების შესაძლებლობას იძლევა, მათ შორის: უცვლელი ზრდა (p=0); ნახევრად-ექსპონენციალური ზრდა (0<p<1) და ექსპონენციალური ზრდა (p=1). a პარამეტრი განსაზღვრავს მარტივი ლოგისტიკური მრუდის სიმეტრიული S ფორმისგან გადახრას. $p = 1$ -სთვის მოდელი იღებს კლასიკური რიჩარდის მოდელის სახეს, ხოლო

$a = 1$ და $p = 1$ –სათვის მოდელი გარდაიქმნება განზოგადოებულ ლოგისტიკურ მოდელად. არაწრფივი უმცირეს კვადრატის ოპტიმიზაციისათვის ვიყენებთ სტანდარტულ ლევენბერგ-მარკარდის ალგორითმს. საწყის ეტაპზე ვაფიქსირებთ C პარამეტრის მნიშვნელობას, რაც ნიშნავს რომ მოდელმა უნდა შეაფასოს დარჩენილი ოთხი პარამეტრი (K, r, p, a). გომპერსის მოდელთან შედარებით უფრო მცირე პროგნოზი განახორციელა რიჩარდის მოდელმა. კერძოდ, ეპიდემიის დასრულების შემდეგ ინფიცირებულთა სრული რაოდენობა იქნება 965. იხ. გრაფიკი.10

სატესტო მონაცემებზე საშუალო კვადრატული შეცდომიდან ფესვის მნიშვნელობა იყო 46.2, რაც S ფორმის მოდელებში ყველაზე კარგი შედეგია. თუმცა, ზოგად შემთხვევაში, სიზუსტის კარგ მაჩვენებლად ვერ ჩაითვლება.

გრაფიკი.10



4. დროითი მწკრივის მოდელები

4.1 ავტორეგრესიული მცურავი საშუალოს მოდელი (ARIMA)

Y_t პროცესი მიეკუთვნება ავტორეგრესიული-მცურავი საშუალოს შერეულ პროცესს თუ ის აღიწერება p რიგის ავტორეგრესიული და q რიგის მცურავი საშუალოს შემადგენლებით. ასეთი პროცესი აღინიშნება ARMA(p,q) აბრევიატურით და აღიწერება შემდეგი განტოლებით:

$$Y_t - \mu = \sum_{j=1}^{p} \theta_j (Y_{t-j} - \mu) + \mu \sum_{j=0}^{q} \alpha_j \varepsilon_{t-j}, \alpha_0 = 1, \theta_p \neq 0, \qquad \alpha_q \neq 0 \tag{4.1.114}$$

სადაც ε_t ინოვაციაა, რომელიც წარმოქმნის თეთრ ხმაურს σ_ε^2 დისპერსიით, μ Y_t პროცესის საშუალოა, θ და α შესაბამისად მოდელის ავტორეგრესიული და მცურავი საშუალოს ნაწილების კოეფიციენტებია. (4.1) შეგვიძლია გადავწეროთ ოპერატიული ფორმით:

$$\theta(L) Y_t = \delta + \alpha(L) \varepsilon_t \tag{4.1.215}$$

სადაც $\delta = (1 - \theta_1 - \theta_2 - \cdots - \theta_p)\mu$.

ARMA მოდელის აგების მნიშვნელოვანი წინაპირობაა, დროითი მწკრივის სტაციონალურობა. დროით მკრივს ეწოდება სტაციონალური, თუ მისი საშუალო და დისპერსია, დროის ნებისმიერ მომენტში ერთმანეთის ტოლია და უსასრულობაზე ნაკლებია. ეკონომიკაში ხშირად ვხვდებით არასტაციონალურ მწკრივებს. მათი მოდელირება ხორციელდება უფრო ზოგადი ARIMA(p,d,q) მოდელით. აბრევიატურაში p-სა და q-ს იგივე განმარტება აქვთ, როგორც ზემოთ, ხოლო d მიუთითებს სხვაობის რიგზე, რომელიც საჭიროა არასტაციონალური დროითი მწკრივის სტაციონალურად გახდომისათვის. ამიტომ, მოდელის შერჩევამდე უნდა ჩატარდეს დროითი მწკრივის სტაციონალურობაზე ტესტირება. ARMA მოდელის აგებისას ერთ-ერთი გადამწყვეტი საკითხია p და q პარამეტრების სწორად შერჩევა. მათი მნიშვნელობების შერჩევისათვის სხვადასხვა მიდგომა გამოიყენება, როგორიცაა ავტოკორელოგრამისა და კერძო ავტოკორელოგრამის ანალიზი, აკაიკის, შვარცის და სხვ. კრიტერიუმების მიხედვით მოდელის სპეციფიკაცია. მოცემულ ნაშრომში ჩვენ განვიხილავთ ორივე მიდგომას, თუმცა გადაწყვეტილებას მივიღებთ აკაიკის კრიტერიუმის საშუალებით. მოდელის პარამეტრების შეფასებისათვის გამოიყენება რამდენიმე ტექნიკა:

- უმცირეს კვადრატთა მეთოდით პარამეტრების შეფასება ეფუძვნება მოდელით შეფასებულ მნიშვნელობებსა და რეალურ მნიშვნელობებს შორის სხვაობების კვადრატების ჯამის მინიმიზირებას.
- მაქსიმალური დასაჯერებლობის მეთოდის გამოყენებისას ვუშვებთ, რომ ε_t არის ნორმალურად განაწილებული შემთხვევითი სიდიდე. ასევე, შემოვიღოთ აღნიშვნები: $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots, \theta_p, \alpha_1, \alpha_2, \dots, \alpha_q)$ და $Y = (Y_1, Y_2, \dots, Y_T)$. დავუშვათ, რომ ამ მონაცემების ერთობლივი ალბათობის სიმკვრივის ფუნქცია მოიცემა შემდეგნაირად: $f(x_T, x_{T-1}, \dots, x_1; \boldsymbol{\theta})$. დასაჯერებლობის ფუნქცია არის ერთობლივი სიმკვრივე, რომელიც განისაზღვრება როგორც $\boldsymbol{\theta}$ პარამეტრების ფუნქცია მოცემული Y-ის პირობით:

$$L(\boldsymbol{\theta}|Y) = f(Y_T, Y_{T-1}, \dots, Y_1; \boldsymbol{\theta}) \qquad (4.1.3)$$

მაქსიმალური დასაჯერებლობის შეფასებას წარმოადგენს : $\widehat{\boldsymbol{\theta}_{\boldsymbol{MLE}}} = arg\max_{\theta\in\Theta} L(\boldsymbol{\theta}|Y)$, სადაც Θ არის პარამეტრების სივრცე. გამოთვლების გასამარტივებლად (4.1.3)-ს მოვდოთ ლოგარითმი: $logL(\boldsymbol{\theta}|Y) = l(\boldsymbol{\theta}|Y)$. (Triacca) დავუშვათ, რომ $l(\boldsymbol{\theta}|Y)$-ს წარმოებული $\boldsymbol{\theta}$-ს მიმართ არის უწყვეტი ყველა $\boldsymbol{\theta}$-სთვის. $l(\boldsymbol{\theta}|Y)$ -ს მაქსიმიზაციის აუცილებელი პირობა არის წარმოებულის ნულთან გატოლება.

$$\frac{\partial l(\boldsymbol{\theta}|Y)}{\partial \boldsymbol{\theta}} = 0 \qquad (4.1.4)$$

რომელსაც ეწოდება დასაჯერებლობის განტოლება, მისი ამოხსნით ვიღებთ პარამეტრების მაქსიმალური დასაჯერობლობით შეფასებას - $\widehat{\boldsymbol{\theta}_{\boldsymbol{MLE}}}$-ს.

- ბოქსმა და ჯენკინსმა შემოგვთავაზეს ოპტიმიზაციის პროცედურა - ბადეზე ძიების მეთოდი, რომელიც განიხილება უმცირეს კვადრატთა მეთოდთან კომბინირებით. ARMA პროცესი დაიყვანება მცურავი საშუალოს (MA) პროცესზე, ახალი ცვადის შემოღებით. შემდეგ გამოითვლება სხვაობები დაკვირვებულ მნიშვნელობებსა და ახალი ცვლადით აგებულ MA მოდელით პროგნოზირებულ მნიშვნელობებს შორის. ბადეზე ძიების მეთოდით ხდება იმ კოეფიციენტების შერჩევა, რომელიც მოახდენს ზემოაღნიშნული სხვაობების მინიმიზირებას.

ARMA მოდელი წარმატებით ახდენს დროითი მწკრივების მომავალი ტრაექტორიის პროგნოზირებას, რისთვისაც იყენებს მოცემული ცვლადის შესახებ არსებულ ისტორიულ მონაცემებს, რომელსაც ვუწოდებთ ინფორმაციულ სიმრავლეს.

ამგვარ ინფორმაციულ სიმრავლეზე დაფუძნებული პროგნოზის ზოგადი ჩანაწერი ასე მოიცემა:

$$Y_{T+h|T} = E\{Y_{T+h}|I_T\} \qquad (4.1.5)$$

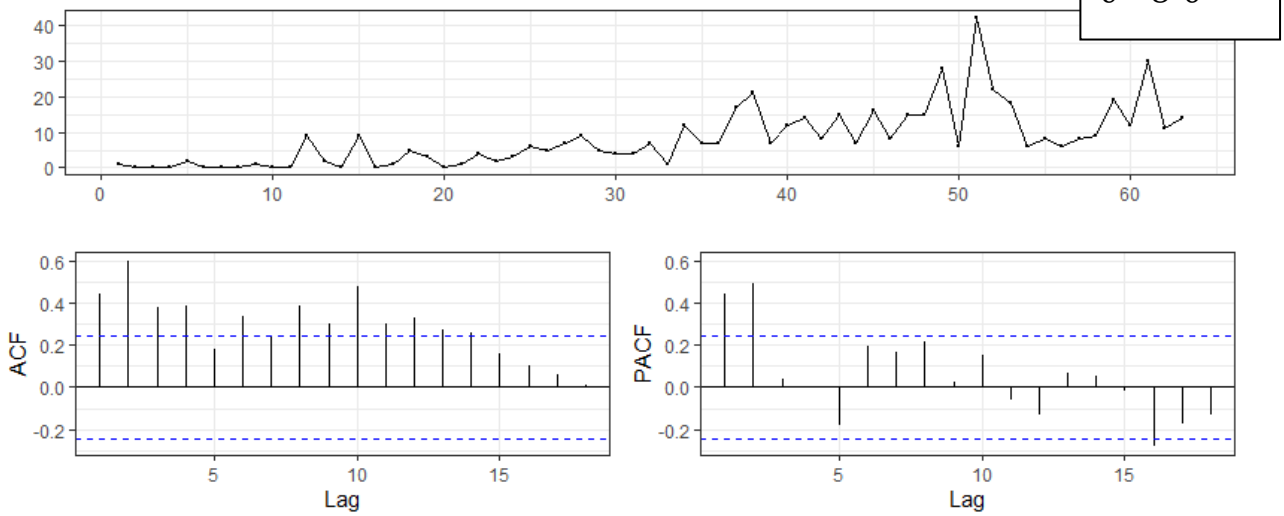
სადაც I_T აღნიშნავს მოცემულ ცვლადზე T პერიოდამდე არსებულ ინფორმაციულ სიმრავლეს. ARMA(p,q)-სთვის შეიძლება მივიღოთ ოპტიმალური პროგნოზის შედგენის რეკურსიული ფორმულა:

$$Y_{T+h|T} = \theta_1 Y_{T+h-1|T} + \cdots + \theta_P Y_{T+h-p|T} + \varepsilon_{T+h|T} + \alpha_1 \varepsilon_{T+h-1|T} + \alpha_q \varepsilon_{T+h-q|T} \qquad (4.1.616)$$

სადაც $Y_{T+h|T}$ არის Y_{T+h}-ს ოპტიმალური პროგნოზი T მომენტში h ბიჯით წინ, ხოლო $\varepsilon_{T+h|T} = 0$-ს როცა h>0, და $\varepsilon_{T+h|T} = \varepsilon_{T+h}$ სხვა შემთხვევაში. ზოგადად საპროგნოზო ინტერვალის ზრდასთან ერთად სიზუსტის ხარისხი მცირდება. ქვემოთ წარმოდგენილია კორონავირუსის გავრცელების პროგნოზი ARIMA მოდელით.

პირველ რიგში ჩავატარეთ კორელოგრამის ანალიზი. (იხ. გრაფიკი.11)

გრაფიკი.11



პირველ გრაფიკზე წარმოდგენილია დღიური ინფიცირების რაოდენობა, მეორე და მესამე გრაფიკზე შესაბამისად ACF და PACF კოლეროგრამებია წარმოდგენილი. PACF მეტყველებს მეორე რიგის AR პროცესზე, ხოლო ACF-ზე მე-14 ლაგამდე მნიშვნელოვანი კორელაციაა, რაც კიდევ ერთხელ ადასტურებს AR პროცესის არსებობას. როგორც ავღნიშნე ARMA მოდელი აიგება მხოლოდ სტაციონალურ მწკრივზე. სტაციონალურობა შემოწმებულია გაფართოებული დიკი-ფულერის ტესტით, R-ის urca პაკეტის საშუალებით. ტესტის განტოლება მოიცემა შემდეგი სახით:

$$\Delta Y(t) = a_0 + \gamma Y_{t-1} + a_2(t) + e(t) \qquad (4.1.7)$$

გამა პარამეტრის ნულთან ტოლობა ნიშნავს ერთეულოვანი ფესვის არსებობას, a_2 პარამეტრი გვიჩვენებს შეიცავს თუ არა ტრენდს მწკრივი, ხოლო a_0-ის ნულთან ტოლობა დრეიფის კომპონენტის არ არსებობაზე მიუთითებს.

შედეგი მოცემულია ცხრილ.1-ში. ცხრილი ტესტავს 3 ჰიპოთეზას:

- Tau3: $H_0: \gamma = 0$, ერთეულოვანი ფესვის არსებობა (არასტაციონალურობა)
- phi2: $H_0: \gamma = a_2 = 0$ არასტაციონალურობა და ტრენდის არარსებობა
- phi3: $H_0: \gamma = a_2 = 0$ არასტაციონალურობა, ტრენდისა და დრეიფის არარსებობა

```
Residual standard error: 6.712 on 44 degrees of freedom
Multiple R-squared:  0.5496,    Adjusted R-squared:  0.5189
F-statistic: 17.89 on 3 and 44 DF,  p-value: 9.738e-08


Value of test-statistic is: -3.4128 3.9092 5.8296

Critical values for test statistics:
      1pct  5pct 10pct
tau3 -4.04 -3.45 -3.15
phi2  6.50  4.88  4.16
phi3  8.73  6.49  5.47
```

ცხრილი.1

თუ მიღებული სტატისტიკების (value of test-statistic) აბსოლიტური მნიშვნელობები აღემატება კრიტიკულ მნიშვნელობებს (critical values for test statistics) ნულოვანი

ჰიპოთეზა უარიყოფა. მოცემული ცხრილიდან ჩანს, რომ 5%-იანი მნიშვნელოვნების დონით ვერ უარვყოფთ ერთეულოვანი ფესვის (არასტაციონალურობის) არსებობას და საჭიროა ARIMA(p,1,q) მოდელის განხილვა. მოდელის სპეციფიკაციისათვის, ანუ p-სა და q-ს მნიშვნელობების შერჩევისათვის გამოვიყენებთ R-ის forecast პაკეტს, auto.arima ბრძანებას, რომელიც თვითონ ახდენს AR და MA ნაწილის ლაგების შერჩევას ინფორმაციული კრიტერიუმების მიხედვით. ჩვენ ვიხელმძღვანელებთ აკაიკის ინფორმაციული კრიტერიუმით. შედეგი მოცემულია ცხრილ.2ში:

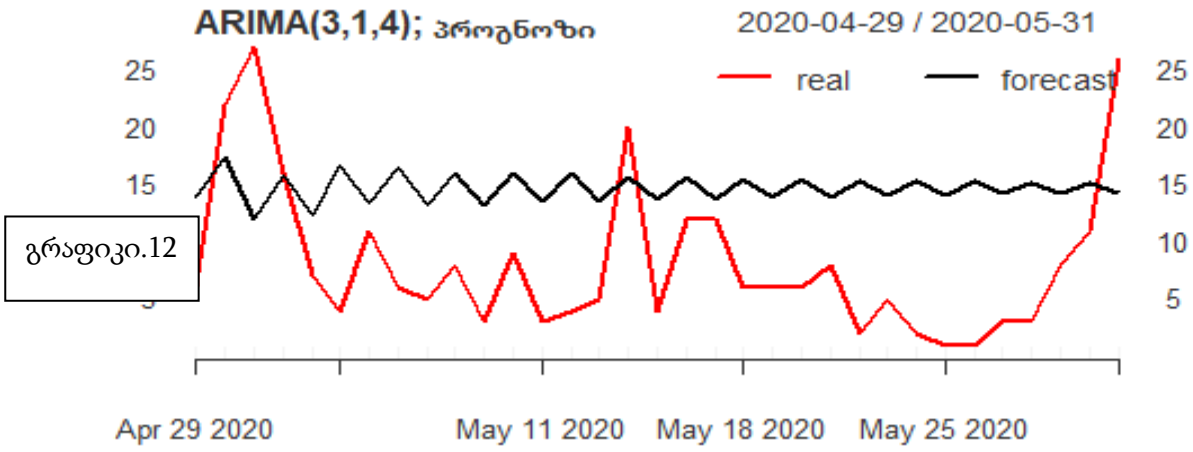
ცხრილი.2

```
Coefficients:
          ar1     ar2      ar3      ma1     ma2     ma3      ma4
      -0.0066  0.2602  -0.5960  -0.9341  0.2592  0.7079  -0.7225
s.e.   0.1779  0.1324   0.1291   0.1744  0.2673  0.2283   0.1649

sigma^2 estimated as 33.6:  log likelihood=-196.03
AIC=408.07   AICc=410.78   BIC=425.08

Training set error measures:
                    ME     RMSE      MAE MPE MAPE      MASE         ACF1
Training set 0.9631536 5.415652 3.703189 NaN  Inf 0.6578731 -0.02877879
```

როგორც ცხრილიდან ჩანს, მოდელმა ARIMA(3,1,4) სახე მიიღო, რაც ნიშნავს, დროითი მწკრივის პირველი რიგის სხვაობები წარმოადგენს მესამე რიგის მცურავი საშუალოს პროცესს, რაც ნიშნავს, რომ პროგნოზი ორი დღის შემდეგ მუდმივი გახდება. ცხრილზე ყურადღება მივაქციოთ შეცდომების ორ საზომს: RMSE-სა (საშუალო კვადრატული შეცდომიდან ფესვი) და MAE-ს (საშუალო აბსოლიტური შეცდომა), რომელთა მნიშვნელობები, შესაბამისად, 5.14 და 3.7-ია და საშუალო დონის სიზუსტეზე მეტყველებს. პროგნოზი წარმოდგენილია გრაფიკი.12-ზე:



გრაფიკი.12

როგორც ჩანს სატესტო მონაცემებზე პროგნოზი არადამაკმაყოფილებლად გამოიყურება. გაზრდილია RMSE და MAE, შესაბამისად, 9.39 და 8.6-მდე. პროგნოზის ცუდი ხარისხი აიხსნება მკაცრი რეგულაციებით, რომლის შედეგად ეპიდემიამ თავისი ბუნება ვერ გამოავლინა, ზრდის ექსპონენციალური ტემპი არ დაფიქსირდა. ინფიცირებულთა რაოდენობის მცირე რიცხვს პერიოდულად ემატება შედარებით დიდი რაოდენობა, რასაც მონაცემებში ხმაური შემოაქვს და პროგნოზირებას ართულებს.

4.2 პოლინომიალური მოდელი მსოფლიო ტრენდის გათვალისწინებით

პოლინომიალური მოდელი ეპიდემიის გავრცელების მსოფლიო ტრენდის გათვალისწინებით, შემოთავაზებული იქნა შემდეგი გარემოებების გათვალისწინებით:

I. ეპიდემიის გავრცელების პატერნი მსგავსია ქვეყნებს შორის, კერძოდ, პიკი მიიღწევა დაახლოებით 7-9 კვირაში და შემდეგ იწყება ინფიცირების ზრდის ტემპის შენელება.

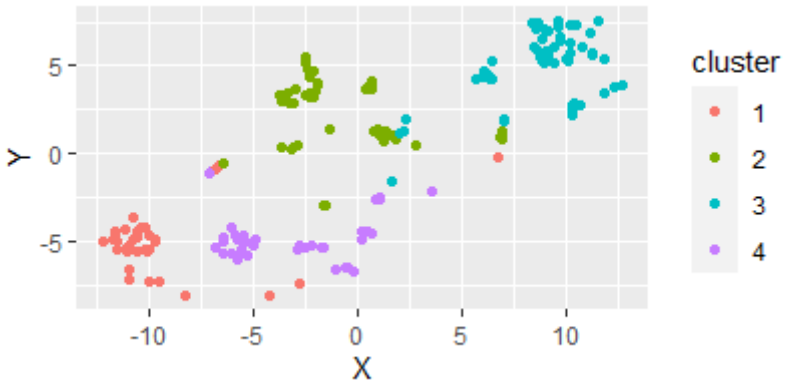
II. ეპიდემიის გავრცელება დამოკიდებული ნაკლებად არის გარემო პირობებზე (ტემპერატურა, ჰავა და ა.შ.)

III. არ არსებობს მედიკამენტი, რომელიც შეაჩერებს ეპიდემიას და შესაბამისად, მდიდარ და ღარიბ ქვეყნებს მეტ-ნაკლებად თანაბარი შესაძლებლობა აქვთ ეპიდემიასთან ბრძოლის

IV. ქვეყნების უმრავლესობამ მალევე ჩაკეტა საზღვრები. შესაბამისად, ქვეყნებს შორის მიგრაციის განსხვავება ვერ მოახდენს გავლენას ეპიდემიის გავრცელებაზე იმპორტირებული შემთხვევების საშუალებით.

I-დან გამომდინარე ინფიცირებულთა რაოდენობა და ზრდის ტემპი დროზეა (ეპიდემიის ასაკზე) დამოკიდებული. ამავდროულად კავშირი არაწრფივია (მრუდს ამოზნექილი ფუნქციის ფორმა აქვს), რის გამოც შერჩეულია დროის მეორე რიგის პოლინომი. გარდა ამისა, გამომდინარე იქედან, რომ ეპიდემიის გავრცელება ქვეყნებს შორის მეტნაკლებად მსგავსია (რადგან ჩვენს ქვეყანაში ეპიდემია უფრო გვიან დაიწყო შეგვიძლია ვისარგებლოთ სხვა ქვეყნების გამოცდილებით) მოდელში ამხსნელ ცვლადად ჩავრთავთ ქვეყნების მიხედვით ეპიდემიის დღიური ზრდის ტემპის მედიანასა და საშუალოს თითოეული დღისათვის. მიუხედავად ზემოაღწერილი მსგავსებისა, სახელმწიფოს მიერ გატარებული შეკავების ღონისძიებების, მოსახლეობის სიმჭიდროვის, ასაკობრივი შემადგენლობის და ა.შ. გათვალისწინებით, ზოგიერთ ქვეყნებს შორის არსებობს განსხვავებები. ამიტომ, პირველ ეტაპზე მედიოდების ირგვლივ დაყოფის (partitioning around medoids (PAM)) (Filaire, 2018)ალგორითმის საშუალებით ჯერ ქვეყნებს დავაჯგუფებთ 4 ჯგუფად, შემდეგ ზემოაღწერილ მოდელს

ავაგებთ იმ ქვეყნების საშუალებით, რომელ ჯგუფშიც ჩავარდება საქართველო. ქვეყნები ჯგუფდება შემდეგი პარამეტრების მიხედვით: 63 დღის განმავლობაში ეპიდემიის ზრდის ტემპის საშუალო და მედიანა, მოსახლეობის რაოდენობა, ქვეყნის ფართობი, მოსახლეობის სიმჭიდროვე, მოსახლეობის რაოდენობა უდიდეს ქალაქში, მშპ ერთ სულ მოსახლეზე, მოსალოდნელი სიცოცხლის ხანგრძლივობა, რეგიონი, ქვეყნის სტატუსი შემოსავლების მიხედვით. კლასტერინგმა საკმაოდ კარგი შედეგი აჩვენა, რაც გამოსახულია გრაფიკ.13-ზე:



გრაფიკი.13

საქართველო მოთავსებულია მეორე ჯგუფში. მოდელის აგების პრინციპი შემდეგია: მეორე კლასტერში მოთავსებული ქვეყნები დალაგდა ეპიდემიის ასაკის (პირველი ინფიცირებიდან გასული დრო) მიხედვით. შემდეგ, თითოეული ასაკისათვის გამოითვალება ქვეყნების მიხედვით ზრდის ტემპის საშუალო და მედიანა. ყოველდღიური ინფიცირების რაოდენობა დარეგრესდა ასაკის მეორე რიგის პოლინომზე, მიღებულ საშუალოზე და მედიანაზე, 63 მონაცემზე. მოდელი წარმოდგნილია ცხრილ.3-ზე:

კოეფიციენტები:

	შეფასება	სტნდ. შეცდ.	t	მნიშვნ.	Pr(>|t|)
(გადაკვეთა)	27.638457	12.162506	2.3.12	0.0268	*
ასაკი	-3.474681	0.571125	-6.084	9.91e-08	***
ასაკი^2	0.180760	0.007124	25.375	< 2e-16	***
საშუალო	-11.988343	44.507130	-0.269	0.7886	
მედიანა	-6.742421	64.319363	-0.105	0.9169	

მნიშვნელოვნის დონე:
0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

ცხრილი.3

როგორც გრაფიკიდან ჩანს, ასაკი და ასაკის მეორე ხარისხი სტატისტიკურად მნიშვნელოვანია. ხოლო მედიანა და საშუალო არამნიშვნელოვანია და კოეფიციენტებიც

არალოგიკური ნიშნისაა. მოდელს მაღალი მორგების ხარისხი აქვს დეტერმინაციის კოეფიციენის მიხედვით, დაზუსტებული $R^2 = 0.9947$. რაც შეეხება პროგნოზის სიზუსტეს, მომავალი 47 დღის საპროგნოზე ჰორიზონტზე RMSE-ისა და MAE-ის მნიშვნელობები, შესაბამისად, 11.6 და 7.9 იყო. მართალია აღნიშნული მოდელის სიზუსტე ARIMA მოდელის სიზუსტეს ჩამორჩება (RMSE-ის მიხედვით), თუმცა, მოკლევადიან პერიოდში, იგი გაცილებით უკეთეს პროგნოზს აკეთებს, ვიდრე S ფორმის მრუდები.

4.3 პროგნოზის სიზუსტის ანალიზი კავკასიის რეგიონში

საქართველოს გარდა, პროგნოზი განვახორციელეთ რუსეთში, სომხეთსა და აზერბაიჯანში. ქვეყანათა ამ ჯგუფის შერჩევა რამდენიმე მიზეზით იყო განპირობებული:

I. პირველი, აზერბაიჯანი და სომხეთი, როგორც გეოგრაფიული თვალსაზრისით, ასევე ერთ სულ მოსახლეზე მშპ-ს მიხედვით საქართველოსთან ახლოს დგანან.

II. მეორე და მნიშვნელოვანი, ამ ქვეყნების განხილვა, ფაქტობრივად, საშუალებას მოგვცემს შევაფასოთ მოდელის სიზუსტის ხარისხი ყველა ტიპის (მაღალი, საშუალო და დაბალი ინფიცირების მქონე) ქვეყნისათვის. მაღალი ინფიცირების ქვეყნების ტიპური მაგალითია რუსეთი (აბსოლუტური მნიშვნელობით) და სომხეთი (მოსახლეობასთან მიმართებით). საშუალო სიმწვავის ეპიდემიოლოგიური მდგომარეობა მოდელირებულია აზერბაიჯანის მაგალითით, ხოლო დაბალი ინფიცირების ქვეყნების მიახლოებას წარმოადგენს საქართველო.

პროგნოზი განვახორციელეთ ორი ტიპის მოდელით: ARIMA და რიჩარდის მრუდი. მათი შერჩევის მოტივი იყო ის, რომ ARIMA-მ ერთ-ერთი საუკეთესო შედეგი აჩვენა RMSE-ის მიხედვით, ხოლო რიჩარდის მრუდი კარგად აღწერს ეპიდემიის ტრენდს. პროგნოზირებისას შემდეგი მიგნებები გაკეთდა:

- მოდელების შედეგების ტრენდი არ შეცვლილა: ARIMA-ს შეცდომა საშუალოდ უფრო მცირეა ვიდრე რიჩარდის მრუდის შეცდომა.

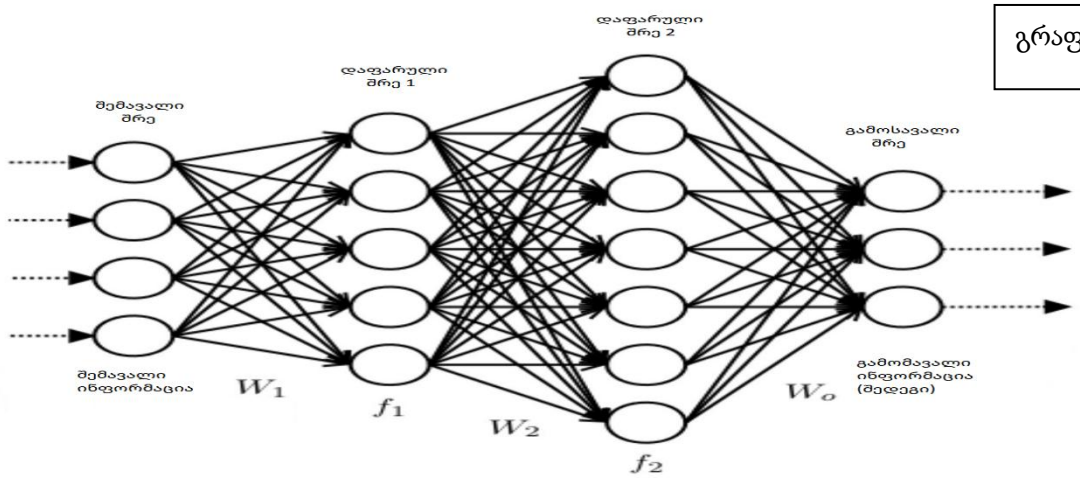
- რიჩარდის მრუდი კარგ შედეგს იძლევა მაშინ, როდესაც პიკის წერტილი უკვე გავლილია, ხოლო ადრეულ ეტაპზე პროგნოზი შესაძლოა ნაკლებობით იყოს გაკეთებული.
- დიდი ინფიცირების მქონე მონაცემების შემთხვევაში პროგნოზის სიზუსტე არ უმჯობესდება, მოდელის ხარისხი უფრო მეტად ინფიცირების ტრენდზეა დამოკიდებული (რამდენად სტაბილურად და ერთგვაროვნად იცვლება ინფიცირება), ვიდრე მონაცემების რაოდენობაზე. დეტალური შედეგები მოცემულია დანართი.7-ში.

5. ნეირონული ქსელები

5.1 ნეირონული ქსელები

უკანასკნელ პერიოდში ფართო სპექტრის ამოცანების გადაწყვეტაში აქტიურად გამოიყენება ხელოვნური ნეირონული ქსელების ალგორთმი. „ხელოვნურ ნეირონულ ქსელებს გააჩნიათ თავის ტვინის ანალოგიური თვისებები, როგორიცაა დასწავლა გამოცდილების გათვალისწინებით, რომელიც დაფუძვნებულია ადრე მიღებულ ცოდნაზე, აბსტრაქტული დასკვნების გაკეთების უნარი, შეცდომებისა და საკუთარ შეცდომებზე დასწავლის უნარი“ (კახიანი, 2004). ხელოვნური ნეირონული ქსელების სტრუქტურა წარმოადგენს ურთიერთდაკავშირებულ კვანძების (ხელოვნური ნეირონების) კავშირს. ბიოლოგიური ტვინის სინაპსისის მსგავსად, თითოეული კავშირის საშუალებით ხდება ინფორმაციის (სიგნალის) გადატანა ერთი კვანძიდან მეორეში. ხელოვნური ნეირონი მიიღებს სიგნალს, გადაამუშავებს მას და გადასცემს სხვა ნეირონს. ჩვეულებრივ, ნეირონები მოთავსებულნი არიან შრეებში. ნეირონული ქსელის სტრუქტურა მოიცავს სამი ტიპის შრეს: შემავალი შრის (input layer) საშუალებით ხდება

გრაფიკი.14



ქსელის დაკავშირება გარე სამყაროსთან (გარედან მიღებული ინფორმაციის საშუალებით), დაფარულ შრეებში (hidden layer) ხდება შემავალი შრიდან მიღებული ინფორმაციის დამუშავება, შედეგი კი აისახება გამომავალ შრეში (output layer) (იხ. გრაფიკი.14). დაფარულ შრეებში ერთი შრის გამოსავალი ამავდროულად წარმოადგენს მეორე შრის შემავალს. ყოველი შემავალი მრავლდება შესაბამის წონით კოეფიციენტზე, w_{ji}-ზე და ნამრავლი იკრიბება. წონები გვიჩვენებს თითოეული კავშირის ფარდობით

მნიშვნელოვნებას. შეწონილი ჯამი გაივლის არაწრფივ აქტივაციის ფუნქციაში და მიღებული შედეგი წარმოადგენს მოცემული კვანძის გამოსავალ მნიშვნელობას. აქტივაციის ფუნქცია განსაზღვრავს მოდელის შედეგს, სიზუსტეს, გამოთვლების ეფექტურობას რესურსების (მაგალითად, დროითი რესურსის) დანახარჯების მხრივ და კონვერგენციის სიჩქარეს.

5.1.1 აქტივაციის ფუნქციები

აქტივაციის ფუნქციის შემდეგი სახეები არსებობს: საფეხუროვანი ფუნქცია (a), წრფივი აქტივაციის ფუნქცია(b), სიგმოიდური (ლოგისტიკური) ფუნქცია(c), ჰიპერბოლური ტანგესი(d), გასწორებული წრფივი ერთეული (d) (rectified linear unit (ReLU)), პარამეტრული ReLU(e), სოფთმაქსი (f) (softmax) და სვიში (g) (Swish) [27]:

$$
\begin{aligned}
f(x) &= \begin{cases} 0, & თუ\ x < 0 \\ 1, & თუ\ x \geq 0 \end{cases} && (a)\\
f(x) &= c * x, სადაც\ c = const. && (b)\\
f(x) &= \sigma(x) = \frac{1}{1+e^{-x}} && (c)\\
f(x) &= \tanh(x) = \frac{e^{x}-e^{-x}}{e^{x}+e^{-x}} && (d)\\
f(x) &= ReLU(x) = \max(0,x) && (e)\\
f(x) &= \begin{cases} \alpha x, & თუ\ x < 0 \\ x, & თუ\ x \geq 0 \end{cases} && (f)\\
f_i(\vec{x}) &= \frac{e^{x_i}}{\sum_{j=1}^{J} e^{x_j}}\ i = 1, \dots, J - სთვის && (g)
\end{aligned}
\qquad (5.1.1)
$$

მათგან ძირითადად გამოიყენება (c), (d), (e) და (f) ფუნქციები. საფეხუროვანი ფუნქციის ნაკლია ის, რომ მისი მნიშვნელობათა არე ძალიან შეზღუდულია. წრფივი ფუნქციის წარმოებული მუდმივია და უკუგავრცელების (backpropagation) ალგორითმის გამოყენება ოპტიმიზაციის პროცესში ვერ ხერხდება, გარდა ამისა, მრავალი შრის არსებობის შემთხვევაშიც კი ბოლო შრის შედეგი წარმოადგენს საწყისი შრის წრფივ კომბინაციას, რაც რთული სტრუქტურის მოდელირების საშუალებას არ იძლევა.

სიგმოიდური ფუნქციის მნიშვნელობათა არე (0, 1) ინტერვალშია მოქცეული, რითაც იგი ერთგვარ ნორმალიზებას ახდენს მონაცემების. (Rizwan, 2018) ამასთანავე გააჩნია გლუვი გრადიენტი, რაც ახდენს საშედეგო მნიშვნელობებში ე.წ. „ნახტომების“

პრევენციას. თუმცა მისი ეს მოქნილობა ზოგჯერ პრობლემის წყარო ხდება. როდესაც შემავალი მონაცემის (X-ის) ძალიან მაღალი ან ძალიან დაბალი მნიშვნელობის პირობებში, პროგნოზის მნიშვნელობა თითქმის არ იცვლება, რაც იწვევს ე.წ. გრადიენტის გაქრობის (vanishing gradient) პრობლემას. ამგვარ პირობებში, მოდელმა შეიძლება სწავლების პროცესი შეწყვიტოს ან ძალიან ნელა დაუახლოვდეს სწორ მნიშვნელობას. სიგმოიდ ფუნქციის ნაკლოვანებაა ასევე გამოთვლებისათვის საჭირო დიდი რესურსები და ნულოცენტრულობის არ არსებობა. ეს უკანასკნელი პრობლემა გადაწყვეტილია ჰიპერბოლური ტანგესის ფუნქციაში, რომლის მნიშვნელობათა არე მოთავსებულია (-1, 1) შუალედში. შესაბამისად, მისი საშუალებით უფრო მარტივია ისეთი მონაცემების მოდელირება, რომლებიც შეიცავენ, როგორც უარყოფით, ასევე დადებით მნიშვნელობებს. რაც შეეხება ReLU-ს, მისი უპირატესობა მდგომარეობს გამოთვლების ეფექტურობაში, იგი სწრაფი კონვერგენციის საშუალებას იძლევა. ასევე, მიუხედავად იმისა, რომ ძალიან გავს წრფივ ფუნქციას, იგი არაწრფივი ფუნქციაა, გააჩნია წარმოებული და უკუგავრცელების ალგორითმის გამოყენების საშუალებას იძლევა. ძირითადი ნაკლოვანება წარმოიშვება არადადებით მნიშვნელობაზე ფუნქციის განულებიდან. ფუნქციის ნულოვან მნიშვნელობებში უკუგავრცელების ალგორითმი ვერ მუშაობს და მოდელი სწავლას წყვეტს. სოფტმაქსი ძირითადად კლასიფიკაციის პრობლემებში გამოიყენება და უმეტესწილად წარმოდგენილია გამოსავალ შრეებში. იგი შემავალ ერთეულს ანიჭებს თითოეულ კლასში მოხვედრის ალბათობას და საბოლოოდ მიაკუთვნებს იმ კლასს, რომელში მოხვედრის ყველაზე მაღალი ალბათობაც ჰქონდა.

აქტივაციის ფუნქციის განხილვის შემდეგ დავუბრუნდეთ მოდელის სწავლების (გაწვრთნის) პროცესს. უპირველეს ყოვლისა, განვმარტოთ რამდენიმე აღნიშვნა: w_{jk}^{l} – ით ავღნიშნოთ წონა რომელიც აკავშირებს $(l-1)$ შრის k-ურ ნეირონს მე- l შრის j-ურ ნეირონთან. b_j^l-ით ავღნიშნოთ მე- l შრის j-ურ ნეირონში გადახრის მნიშვნელობა. ხოლო a_j^l იყოს მე- l შრის j-ურ ნეირონის აქტივაციის ფუნქციის მნიშვნელობა. წინა შრის აქტივაციის მნიშვნელობა მომდევნო შრის აქტივაციის მნიშვნელობას უკავშირდება შემდეგნაირად:

$$a_j^l = f\left(\sum_k w_{jk}^l a_k^{l-1} + b_j^l\right) \tag{5.1.217}$$

სადაც დაჯამება ხდება $(l-1)$ შრის ყველა k-ური ნეირონის მიხედვით, ხოლო $f(.)$ ფუნქციას აქვს (5.1.1)-ში განხილული აქტივაციის ფუნქციებიდან ერთ-ერთის სახე. (5.1.2) გვიჩვენებს, რომ ყოველი შრის გამოსავალი მნიშვნელობა, ანუ აქტივაციის ფუნქციის შედეგი, ამავდროულად, წარმოადგენს მომდევნო შრის შემავალ მნიშვნელობას. უფრო მოსახერხებელი წარმოდგენისთვის (5.2) გადავწეროთ მატრიცული ფორმით:

$$a^l = f(W^l a^l + b^l) \tag{5.1.318}$$

სადაც W^l აღნიშნავს წონების მატრიცას, რომლებიც უკავშირდებიან l-ურ შრეს. ანალოგიურად b^l l-ური შრის გადახრის წევრების ვექტორია, ხოლო a^l ამავე შრის აქტივაციის ფუნქციის შედეგები. გამოთვლების გასამარტივებლად f ფუნქციის არგუმენტი ავღნიშნოთ z^l-ით: $z^l = W^l a^l + b^l$, და მას ვუწოდოთ შეწონილი შემავალი მნიშვნელობა. მაშინ გვექნება: $a^l = f(z^l)$.

5.1.2 დანაკარგების ფუნქცია

ნეირონული ქსელების მოდელის სწავლება გულისხმობს წონების იმგვარ კორექტირებას, რომ მოვახდინოთ სამიზნე ფუნქციის ოპტიმიზაცია. სამიზნე ფუნქციად განიხილება დანაკარგების (დანახარჯების) ფუნქცია (Loss (Cost) function). ოპტიმიზაცია გულისხმობს ამ ფუნქციის მინიმიზაციას ან მისი შებრუნებულის მაქსიმიზაციას. უნდა აღინიშნოს, რომ იდეალური დანაკარგების ფუნქცია არ არსებობს. მისი სახე დამოკიდებულია ამოცანის ტიპზე. კლასიფიკაციის ამოცანებში გამოიყენება ჯვარედინი ენთროპია (cross entropy), ე.წ. ღერძის ანაკარგი (hinge loss), კვადრატული ღერძის დანაკარგი (squared hinge loss) და კულბეკ ლაიბლერის განსხვავების დანაკარგი (Killback Leibler Divergence Loss), ხოლო რეგრესიული ანალიზის დროს გამოიყენება საშუალო კვადრატული შეცდომა (mean square error (MSE)); საშუალო კვარდატული ლოგარითმული შეცდომა (mean square logarithmic error) და საშუალო აბსოლუტური შეცდომა (mean absolute error (MAE)).

გამომდინარე იქედან, რომ ჩვენი ამოცანა რეგრესიული ტიპის ამოცანაა, გამოვიყენებთ საშუალო კვადრატული შეცდომის დანაკარგების ფუნქციას:

$$C = \frac{1}{2n}\sum_x \|y(x) - a^L(x)\|^2, \tag{5.1.419}$$

სადაც n საწვრთნელო მონაცემების შერჩევის ზომაა, $y(x)$ შედეგობრივი ცვლადის მნიშვნელობაა, L აღნიშნავს მოდელში შრეების საერთო რაოდენობას, ხოლო $a^L = a^L(x)$ არის აქტივაციის ფუნქციის შედეგი, x შემავალი მონაცემების პირობით.

შეცდომების ოპტიმიზაცია ხდება უკუპროპაგაციის ალგორითმით, რომელიც წონების კორექტირებისათვის იყენებს სტოხასტური გრადიენტული დაშვების ალგორითმს. უკუპროპაგაციის მიზანი არის დანაკარგების ფუნქციის ნაწილობრივი წარმოებულების გამოთვლა წონებისა და გადახრის წევრის მიმართ. ამისათვის საჭიროა დანაკარგების ფუნქციის მიმართ ორი დაშვება შემოვიღოთ. პირველი დაშვება არის ის რომ დანახარჯების ფუნქცია შეიძლება ჩაიწეროს, როგორც ინდივიდუალური დაკვირვებების დანაკარგების საშუალო ფუნქცია: $C = \frac{1}{n}\sum_x C_x$, სადაც $C_x = \frac{1}{2}\|y - a^L\|^2$ აღნიშნავს თითოეული დაკვირვების დანაკარგს. მეორე დაშვება არის ის, რომ დანაკარგების ფუნქცია შეიძლება ჩაიწეროს, როგორც მოდელის საბოლოო შედეგის ფუნქცია. ამ მოთხოვნებს აკმაყოფილებს კვარდატული შეცდომის ფუნქცია:

$$C = \frac{1}{2}\sum_x \|y - a^L\|^2 = C = \frac{1}{2}\sum_x \left(y_j - a_j^L\right)^2 \tag{5.1.5}$$

მოდელის გაწვრთა გულისხმობს დანაკარგების ფუნქციის მინიმიზაციას წონების ცვლილების საშუალებით.

5.1.3 უკუგავრცელების ალგორითმი

მოდელის გაწვრთნა მიმდინარეობს სტოხასტური გრადიენტული დაშვების ტექნიკის საშუალებით, წონები და გადახრის წევრი განახლდება შეცდომების უკუგავრცელების ალგორითმით, რომელიც ახდენს საშუალო კვადრატული შეცდომის მინიმიზაციას. გრადიენტული დაშვების მეთოდის თანახმად წონითი კოეფიციენტების ცვლილება ხდება შემდეგი წესის მიხედვით:

$$w_{j,i}(t+1) = w_{j,i}(t) - \eta\frac{\partial C}{\partial w_{j,i}(t)}, \tag{5.1.6}$$

სადაც $\eta \in (0;1)$-ს სწავლების კოეფიციენტი (learning rate) ეწოდება და იგი განსაზღვრავს სწავლების სისწრაფეს. მისი მნიშვნელობა დამოკიდებულია მონაცემების

რაოდენობაზე. როგორც წესი, მრავალი მონაცემის არსებობის შემთხვევაში $\eta = 0.01$ ან $\eta = 0.001$. როდესაც η პარამეტრი ერთთან ახლოსაა, კოეფიციენტების თითოეული დაკვირვების შეცდომა კოეფიციენტს მკვეთრად ცვლის და იწვევს ე.წ. ზედმეტად მორგების (overfitting) პრობლემას. ზედმეტად მორგების პრობლემა გულისხმობს, რომ მოდელი მონაცემების მცირე შერჩევაზე ახდენს პარამეტრების ოპტიმიზაციას, თუმცა მიღებული მოდელი დანარჩენ მონაცემებზე დიდ შეცდომას იძლევა, რადგან იგი მოერგო მონაცემების მცირე, არარეპრეზენტატულ შერჩევას. როდესაც სწავლების კოეფიციენტი ერთთან ახლოსაა, კოეფიციენტის კორექტირება ხდება სწორედ მონაცემთა მცირე შერჩევით. ოპტიმიზაციის ალგორითმი ზედმეტად მალე პოულობს დანაკარგების ფუნქციის ლოკალურ მინიმუმს და ვღებულობთ არასწორ მოდელს. ამიტომ სწავლების პროცესში მნიშვნელოვანია η პარამეტრის სწორად შერჩევა.

დავუბრუნდეთ კვლავ უკუ გავრცელების ტექნიკას. მისი მიზანია თითოეული შრისათვის (როგორც დაფარული ასევე საშედეგო (output)) მოახდინოს შეცდომის მინიმიზაცია პარამეტრების (წონების) კორექტირებით. ცხადია, დაფარული შრეებისთვის სამიზნე მნიშვნელობები ცნობილი არ არის და შესაბამისად, არ შეგვიძლია პირდაპირ დავითვალოთ დაფარული შრის დანაკარგების ფუნქციის წარმოებულები. ამ პრობლემის გადასაწყვეტად გამოიყენება ე.წ. ჯაჭვური წესი (chain rule). შეცდომების გრადიენტების დათვლა დაფარული და საშედეგო მნიშვნელობებისათვის განსხვავებულია, ამიტომ ორივე შემთხვევას განვიხილავთ ქვემოთ. ჯერ დავიწყებთ საშედეგო შრის შემთხვევის განხილვით, ხოლო შემდეგ დავუბრუნდებით დაფარული შრის შემთხვევას. გარდა ამისა, აქტივაციის ფუნქციის სახედ ავიღებთ სიგმოიდ ფუნქციას. სიგმოიდ ფუნქციას აქვს კარგი თვისება, რაც მდგომარეობს გრადიენტის მარტივად გამოთვლაში, კერძოდ სიგმოიდის გრადიენტი გამოითვლება შემდეგი სახით:

$$\frac{\partial \sigma(x)}{\partial x} = \sigma(x)(1 - \sigma(x)) \tag{5.1.7}$$

ამის გათვალისწინებით შეცდომის ფუნქციის გრადიენტი შეგვიძლია ასე გადავწეროთ:

$$\frac{\partial C}{\partial w_i} = \frac{\partial}{\partial w_i}\frac{1}{2}\sum_{d \in D}(y_d - a_d)^2$$

$$= \frac{1}{2}\sum_{d \in D} 2(y_d - a_d)\frac{\partial}{\partial w_i}(y_d - a_d)$$

$$= \sum_{d \in D} 2(y_d - a_d) \frac{\partial a}{\partial z_d} \frac{\partial z_d}{\partial w_i} \quad (5.1.8)$$

სადაც d აღნიშნავს დაკვირვების ნომერს, D მთლიანი შერჩევაა, z და a მნიშვნელობებზე კი ზემოთ ვისაუბრეთ და ამ შემთხვევაში აქტივაციის ფუნქციას სიგმოიდის სახე აქვს. (5.1.7)-ის გამოყენებით (5.1.8) შეგვიძლია შემდეგი სახით გადავწეროთ:

$$\frac{\partial C}{\partial w_i} = -\sum_{d \in D} 2(y_d - a_d)\, a_d(1 - a_d) x_{i,d} \quad (5.1.9)$$

დავუბრუნდეთ გამოსავალი შრის ელემენტებს და თითოეული j-ური ნეირონისათვის განვსაზღვროთ შეცდომა C_d. როგორც ავღნიშნე ის წარმოადგენს a_d-ის, ანუ მოდელით დათვლილი მნიშვნელობის ფუნქციას და რომელიც თავის მხრივ z-ის ფუნქციაა. იმისათვის რომ დავითვალოთ C_d-ს წარმოებული წონების მიმართ ყოველი j-სათვის, უნდა გამოვიყენოთ ჯაჭვის წესი და (5.1.9) ფორმულა. გამომდინარე იქედან, რომ ამ შემთხვევაში წარმოებული ითვლება ერთი დაკვირვებისა და j-ური ნეირონისათვის, წინა შრის i ნეირონიდან გამომავალი და j ნეირონში შემავალი წონის მნიშვნელობა კორექტირდება შემდეგი სახით:

$$w_{j,i}(t+1) = w_{j,i}(t) - \eta\big(y_j - a_d\big) a_d(1 - a_d) x_{i,d} \quad (20.1.10)$$

ახლა დავუბრუნდეთ დაფარულ შრეებს და ვნახოთ როგორ მიმდინარეობს ამ შემთხვევაში წონების კორექტირება. თავდაპირველად შემოვიღოთ ახალი ტერმინი - „j-ს ჩაშლა“ (downstream(j)). j-ს ჩაშლა ვუწოდოთ $(l-1)$ შრის j-ური ნეირონიდან l ნეირონში შემავალ მნიშვნელობებს. დავუშვათ მოდელი შეიცავს L შრეს. ვნახოთ როგორ ხდება წონების კორექტირება $L-1$ შრისათვის, რომლის ნეირონების სამიზნე მნიშვნელობა ჩვენთვის უცნობია. ამ შემთხვევაში ვიყენებთ კვლავ ჯაჭვის წესს და ვითვლით ჩვენთვის ცნობილი L –ური, ანუ საშედეგო შრეზე დათვლილი შეცდომის წარმოებულს $L-1$ შრის z_j-ის მიმართ:

$$\frac{\partial E_d}{\partial z_j} = \sum_{k \in downstream(j)} \frac{\partial C_d}{\partial z_k} \frac{\partial z_k}{\partial z_j} \quad (5.1.11)$$

ამ შემთხვევაში k წარმოადგენს ბოლო შრის ერთ-ერთ ნეირონს, ხოლო j წინა შრის ერთ-ერთი ნეირონია. მათ შორის კავშირის საილუსტრაციოდ უნდა გავიხსენოთ, რომ:

$$z_k^l = \sum_{j \in J} w_{k,j}^{l-1} f_j^{l-1}\big(z_j^{l-1}\big) \quad (5.1.12)$$

ამ შემთხვევაში f ფუნქციას სიგმოიდ ფუნქციის სახე აქვს. (5.1.12)-დან ჩანს რომ z_k^l-ს კავშირი აქვს წინა შრის წონებთან. შემოვიღოთ აღნიშვნა $\delta_k = -\frac{\partial C_d}{\partial z_k}$ და (5.1.11) შემდეგი სახით გადავწეროთ:

$$\begin{aligned}\frac{\partial C_d}{\partial z_j} &= \sum_{k\in downstream(j)} -\delta_k \frac{\partial z_k}{\partial z_j} \\ &= \sum_{k\in downstream(j)} -\delta_k \frac{\partial z_k}{\partial a_j}\frac{\partial a_j}{\partial z_j} \\ &= \sum_{k\in downstream(j)} -\delta_k w_{k,j}\frac{\partial a_j}{\partial z_j} \\ &= \sum_{k\in downstream(j)} -\delta_k w_{k,j} a_j(1-a_j)\end{aligned} \tag{5.1.13}$$

(5.1.13) გვიჩვენებს ლოგიკას, თუ როგორ ხდება დაფარული შრის პარამეტრების მიმართ დანაკარგების ფუნქციის წარმოებულის გამოთვლა ჯაჭვის წესის საშუალებით. ამ ლოგიკის განვრცობა მარტივადაა შესაძლებელი უფრო ღრმა ნეირონული ქსელის მოდელის შემთხვევაზე. უკუ გავრცელებით სწავლება მხოლოდ სტოხასტური გრადიენტული დაშვების საშუალებით არ ხდება. არსებობს სხვა უფრო პოპულარული ოპტიმაიზერები, რომლების ასევე გრადიენტის დათვლას ეფუძნებიან, როგორიცაა ადაგრადი, ადადელტა, ადამი და სხვ. (დეტალურად იხ. დანართი.5)

ოპტიმაიზერების განხილვით, პრინციპში, მოდელის პარამეტრების (წონების და გადახრის წევრის) ფუნქციის, მათი შერჩევისა და ოპტიმიზაციის გზების განხილვას მოვრჩით. მოდელის აგების პროცესში კიდევ ერთი მნიშვნელოვანი რგოლია ჰიპერპარამეტრების სწორად შერჩევა. ჰიპერპარამეტრები არიან პარამეტრები, რომელთა საშუალებითაც ხდება სწავლების პროცესის კონტროლი. ნეირონული ქსელების აგებისას ძირითადად გამოიყენება შემდეგი ჰიპერპარამეტები:დაფარული შრეების რიცხვი, სწავლების კოეფიციენტი, აქტივაციის ფუნქცია, ქვეშერჩევის ზომა, ეპოქები და ამომგდები (დეტალურად იხ. დანართი.6). პარამეტრებისგან განსხავებით, რომელთა მნიშვნელობების შერჩევა სწავლებისას ხდება, ჰიპერპარამეტრების მნიშვნელობები მოდელს უნდა მიეწოდოს გარედან. მოდელის სიზუსტე დიდადაა

დამოკიდებული მათ სწორ შერჩევაზე. ლიტერატურაში (Ippolito, 2019) გამოიყოფა ჰიპერპარამეტრების ოპტიმიზაციის ხუთი მეთოდი:

1. შემთხვევითი ძიება (Random search)
2. ბადეზე ძიება (Grid search)
3. ბაიესიანური ოპტიმიზაცია (Bayesian optimization)
4. გრადიენტზე დაფუძვნებული ოპტიმიზაცია (Gradient-based optimization)
5. ევოლუციური ოპიმიზაცია (Evolutionary optimization)

ვიდრე თითოეულ მათგანს დავახასიათებდე, უნდა ავღწერო ოპტიმიზაციის ზოგადი მიდგომა, რომელიც ფართოდ გამოიყენება პროგნოზირებაში.

ჩვეულებრივ, შეცდომის მინიმიზება საწვრთნელ მონაცემებზე მარტივია, თუმცა შერჩეული მოდელის სიზუსტე სატესტო მონაცემებზე მკვეთრად მცირდება, ანუ საქმე გვაქვს ზედმეტად მორგების (overfitting) პრობლემაზე. ამ პრობლემის გადასაწყვეტად გამოიყენება ჯვარედინი ვალიდაციის (cross validation) ტექნიკა. ზედმეტად მორგების პრობლემის თავიდან ასაცილებლად მონაცემებს ყოფენ სამ ნაწილად: საწვრთნელი შერჩევა (train set), ვალიდაციის შერჩევა (validation set) და სატესტო შერჩევა (test set). მიზანი არის სატესტო შერჩევაზე განხორციელებული პროგნოზის მინიმიზირება მოვახდინოთ. ამიტომ ამ შერჩევას მოდელის აგების ეტაპიდან გამოვრიცხავთ და დამოუკიდებლად ვინახავთ. სწავლებისა და ჰიპერპარამეტრების ოპტიმიზაციისათვის გამოვიყენებთ დანარჩენ ორ შერჩევას. კერძოდ, საწვრთნელ მონაცემებზე აიგება რამდენიმე მოდელი სხვადასხვა ჰიპერპარამეტრების საშუალებით, ხოლო თითოეული მოდელის სიზუსტე შეფასდება ვალიდაციის შერჩევაზე გაკეთებული პროგნოზით. საბოლოოდ, შეირჩევა პარამეტრების ის მნიშვნელობები, რომელთაც საუკეთესო შედეგი აჩვენეს ვალიდაციის შერჩევაზე. მიღებული მოდელი გამოიყენება სატესტო შერჩევაზე პროგნოზის გასაკეთებლადაც. ჯვარედინი ვალიდაცია წარმოადგენს აღწერილი მოდელის გაფართოებას. ამ შემთხვევაში სატესტო შერჩევის გამოყოფის შემდეგ დარჩენილი მონაცემები იყოფა 5 ან 10 ნაწილად. ყოველ ჯერზე ერთ-ერთი ნაწილი შეირჩევა ვალიდაციისათვის, ხოლო დანარჩენ მონაცემებზე აიგება მოდელი. ეს მიდგომა გვაძლევს უფრო რობასტულ პარამეტრების მიღების საშუალებას. ზოგად შემთხვევაში მთლიანი მონაცემები იყოფა შემდეგი პროპორციით: საწვრთნელი შერჩევა

80%, ვალიდაციის შერჩევა 10% და სატესტო შერჩევა 10%. იმ შემთხვევაში, თუ ვახორციელებთ ჯვარედინ ვალიდაციას მაშინ მონაცემები დაიყოფა 80%:20% პროპორციით, შესაბამისად, მოდელის განვითარებისა და ტესტირებისთვის. ცხადია ეს პროპორციები შეიძლება შეიცვალოს მონაცემთა რაოდენობისა და მკვლევარის არჩევანის მიხედვით.

დავუბრუნდეთ ოპტიმიზაციის ტექნიკის განილვას. ბადეზე ძებნა ჰიპერ-პარამეტრების ოპტიმიზაციის ტრადიციული მეთოდია. იგი გულისხმობს ჰიპერპარა-მეტრების გარკვეული ნაკრების (ბადის) მანუალურ შერჩევას. პარამეტრების მნიშვნელობები ხშირად წინასწარ განსაზღვრული ინტერვალით აიღება. საბოლოოდ შეირჩევა პარამეტრთა ის კომბინაცია, რომელიც ყველაზე ზუსტ პროგნოზს გააკეთებს ჯვარედინი ვალიდაციის მიხედვით. შემთხევევითი ძიების პროცესი ძლიერ გავს ბადეზე ძიებას. აქაც პარამეტრების მნიშვნელობათა ნაკრები შეირჩევა წინასწარ, თუმცა შემთხვევით და არა წინასწარ განსაზღვრული წესის მიხედვით, როგორც ეს ბადეზე ძებნის შემთხვევაში იყო. პარამეტრთა ოპტიმალური მნიშვნელობები მიიღება ჯვარედინი ვალიდაციის გამოყენებით. გრადიენტზე დაფუძვნებული ოპტიმიზაციის შემთხვევაში გამოითვლება დანაკარგების ფუნქციის გრადიენტი ჰიპერპარამეტრების მიმართ, რითაც ხდება ოპტიმალური მნიშვნელობებთან ეტაპობრივი დაახლოება (Larsen, Hansen, & al., 1996). რთული მოდელის შეთხვევაში სამივე ალგორითმი დავავშირებულია გამოთვლით დანახარჯებთან, ასეთ შემთხვევაში იყენებენ ბაიესიანურ ოპტიმიზაციას. ზოგადად, იგი წარმოაგდგენს გლობალურ ოპტიმიზაციის მეთოდს ხმაურიანი უცნობი სახის ფუნქციისათვის (Krasser, 2018). იგი აგებს ალბათურ მოდელს ჰიპერპარამეტრთა ურთიერთდამოკიდებულების ფუნქციის სპეციფიკაციისათვის. წინა ბიჯზე მიღებულ მოდელზე დაფუძვნებით იტერაციულად უახლოვდება უცნობი ფუნქციის ოპტიმალურ სახეს. ძიებისას გამოყენებულ მოდელს ეწოდება სუროგატი მოდელი. იგი ბაიესიანური ოპტიმიზაცია ასევე იყენებს ე.წ. შეძენის ფუნქციას, რომელიც გამოკვეთს მნიშვნელობების არეს, სადაც მოხდა გაუმჯობესება. პოპულარული სუროგატ მოდელს არის გაუსიანური პროცესი, ხოლო პოპულარული შეძენის ფუნქციების (acquisition function) გაუმჯობესების მაქსიმალური ალბათობა (maximum probability of improvement), მოსალოდნელი გაუმჯობესება (expected

improvement) და ზედა საიმედოობის ზღვარი (upper confidence bound). ევოლუციური ოპტიმიზაცია არის მეთოდოლოგია უცნობი ფუნქციის (black-box function) გლობალური ოპტიმიზაციისათვის. იგი იყენებს განვითარებად ალგორითმს ჰიპერპარამეტრების ოპტიმალურ მნიშვნელობათა სივრცის საპოვნელად. ოპტიმიზაციის პროცესი მოტივირებულია ევოლუციის ბიოლოგიური კონცეპტით:

1. გენერირდება ჰიპერპარამეტრთა საწყისი მნიშვნელობები
2. ჰიპერპარამეტრების მიხედვით შეფასდება მოდელი და შეინახება მიღებული სიზუსტის მნიშვნელობები.
3. ჰიპერპარამეტრები დალაგდება გაკეთებული პროგნოზის სიზუსტის მიხედვით
4. სიზუსტის ყველაზე ცუდ მაჩვენებელთან ასოცირებული ჰიპერპარამეტრი ჩანაცვლდება სხვა მნიშვნელობებით.
5. (2)-(4) საფეხური განმეორდება დამაკმაყოფილებელი მნიშვნელობების მიღებამდე.

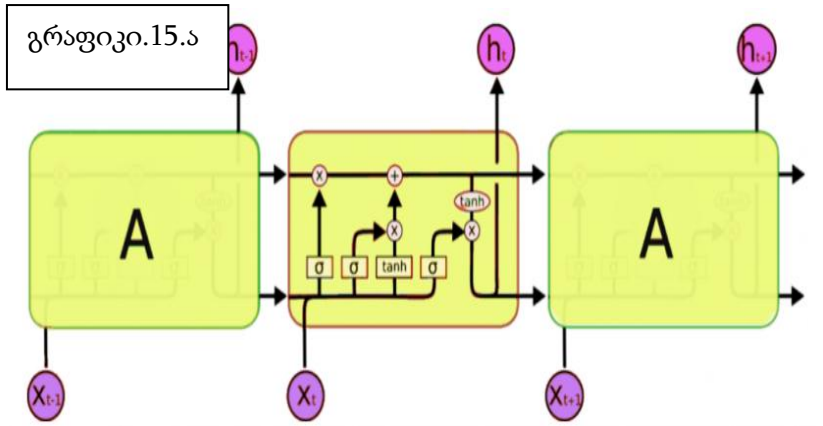
ზემოთ მიმოვიხილეთ ნეირონული მოდელის ზოგადი არქიტექტურა. კორონავირუსის პროგნოზირებისათვის გამოვიყენებთ დროითი მწკრივების ფართოდ გავცელებულ მოდელს - გრძელი მოკლე მახსოვრობის მოდელს (LSTM). ქვემოთ მიმოვიხილავთ ამ მოდელის არქიტექტურას.

5.2 მოკლევადიანი მახსოვრობის გრძელი მოდელი (Long Short Term Memory)

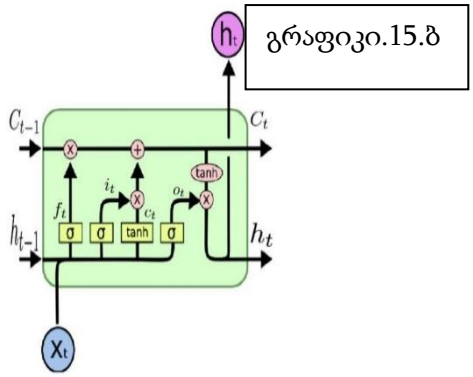
მოკლევადიანი მახსოვრობის გრძელი მოდელი (LSTM) (გრაფიკი.15.ა) წარმოადგენს ხელოვნური რეკურენტული ნეირონული ქსელების (RNN) ერთ-ერთ ნაირსახეობას, რომელიც ფართოდ გამოიყენება ღრმა სწავლებაში, როგორც ცალკეული მონაცემების დასამუშავებლად (მაგალითად გრაფიკის ამოცნობა), ასევე მიმდევრობითი მონაცემების ანალიზისათვის (საუბარი ან ვიდეო). LSTM ქსელები კარგად ერგება დროითი მწკრივის

საფუძველზე პროგნოზირებისა და კლასიფიკაციის ამოცანებს.

გრაფიკი.15.ა



გრაფიკი.15.ბ



ტიპიური LSTM ქსელი შედგება სხვადასხვა მეხსიერების ბლოკისგან, რომელსაც უჯრედები (cell) (იხ. გრაფიკი.15.ბ) ეწოდება (მართკუთხედები გრაფიკზე). მოცემული უჯრედიდან მომდევნო უჯრედში გადადის ორი ტიპის ნაკადი: უჯრედის მდგომარეობა C_t (cell state) და დაფარული მდგომარეობა h_t (hidden state). (Srivastava, 2017)

მახსოვრობის ბლოკები პასუხისმგებლები არიან ინფორმაციის დამახსოვრებაზე, ხოლო ამ ინფორმაციით მანიპულიება ხდება სამი კარის საშუალებით: დავიწყების კარის (forget gate), შესვლის კარისა (input gate) და გამოსვლის კარის (output gate). დავიწყების კარი პასუხისმგებელია უჯრედის მდგომარეობის ნაკადიდან არასაჭირო ინფორმაციის ამოღებაზე. კარში შედის ორი სახის ნაკადი: დაფარული მდგომარეობის ნაკადი, h_{t-1}, წინა უჯრედიდან და X_t, რომელიც არის შემავალი (ამხსნელი) ინფორმაცია დროის მოცემულ მომენტში. შემავალი ინფორმაცია მრავლდება წონებზე და ემატება გადახრის წევრი (bias). მიღებული მნიშვნელობა შედის სიგმოიდ ფუნქციაში და f_t-ს წონის, რომლის მნიშვნელობათა არე მოთავსებულია 0-სა და 1-ს შორის:

$$f_t = \sigma(W_f[h_{t-1}, x_t] + b_f) \qquad (5.2.121)$$

მიღებული მნიშვნელობა მრავლდება წინა უჯრედის ნაკადზე C_{t-1}-ზე. თუ f_t −ის მნიშვნელობა ნულის ტოლია, ეს ნიშნავს რომ წინა უჯრიდან მიღებული ინფორმაცია საერთოდ არ გამოიყენება, ხოლო როდესაც იგი ერთის ტოლია წინა ეტაპიდან მიღებული ინფორმაცია მთლიანად გამოიყენება შემდგომ გამოთვლებში.

შემდეგი ნაბიჯი არის, განვსაზღვროთ ახალი ინფორმაციის რა ნაწილი უნდა შევიყვანოთ უჯრედში. ამ ამოცანის გადაწყვეტაზე პასუხისმგებელია შესავალი კარი. იგი შედგება ორი ნაწილისგან. პირველი, სიგმოიდის შრე (i_t) წყვეტს რა მნიშვნელობები უნდა განახლდეს. მეორე, ჰიპერბოლური ტანგესი ქმნის ახალი მნიშვნელობების ვექტორის კანდიდატებს, $\widetilde{C_t}$. ამ ორი ვექტორის კომბინაციით იქმნება განახლებული მდგომარეობა (state).

$$i_t = \sigma(W_i[h_{t-1}, x_t] + b_i) \tag{5.2.2}$$
$$\widetilde{C_t} = tanh(W_C[h_{t-1}, x_t] + b_C).$$

ძველ მდგომარეობას ვამრავლებთ f_t-ზე და ვამატებთ ახალი კანდიდატის მნიშვნელობებს, რომელიც შეწონილია სიგმოიდის ფუნქციის მნიშვნელობით, $i_t * \widetilde{C_t}$.

$$C_t = f_t * C_{t-1} + i_t * \widetilde{C_t} \tag{5.2.3}$$

ბოლო ეტაპზე, გადაწყდება რა უნდა იყოს შედეგი/გამოსავალი (output). უჯრედის შედეგი გაივლის ჰიპერბოლური ტანგესის ფუნქციაში, რომელიც შეიწონება სიგმოიდის ფუნქციის, o_t-ს, შედეგით.

$$o_t = \sigma(W_o[h_{t-1}, x_t] + b_o) \tag{5.2.4}$$
$$h_t = o_f * \tanh(C_f)$$

შედეგად ვღებულობთ დამალული მდგომარეობის ახალ მნიშვნელობას, h_t-ს და უჯრედის ახალ მდგომარეობას, C_t.

5.2.1 LSTM მოდელის სპეციფიკაცია

მოდელი აიგო python-ის keras ბიბლიოთაკაში. გამომდინარე იქედან, რომ ნეირონული ქსელების მოდელებისათვის საჭიროა დიდი რაოდენობის მონაცემები, პროგნოზისა და მოდელის აგებისათვის გამოყენებული შერჩევის მოცულობები შევცვალეთ. კერძოდ, მოდელი აიგო პირველი 90 დღის მონაცემზე (არაკუმულატიურზე). მოდელში შედიოდა 21 დღის მოცულობის მონაცემი, რომლის მიხედვით კეთდებოდა 7 დღიანი პროგნოზი. Batch-ის ზომად შეირჩა 3, რაც ნიშნავს, რომ მოდელის წონები იცვლებოდა ყოველი სამი მონაცემის მიხედვით. მოდელის აგებისთვის გამოყენებული იქნა ორი დაფარული LSTM შრე და ერთი ჩვეულებრივი საშედეგო შრე. აქტივაციის ფუნქციებად გამოვიყენეთ რელუ (Relu), ხოლო ოპტიმაიზერად ადამი. ოპტიმიზაციის ფუნქციად შეირჩა საშუალო

კვადრატული შეცდომა, ხოლო ეპოქების რაოდენობად - 50. პროგნოზი გავაკეთეთ ბოლო 30 დღის მონაცემზე (83-ე - 110-ე დღეებზე). საწვრთნელ და სატესტო მონაცემებს რვა დღიანი კვეთა აქვს, თუმცა ამას სათანადო ახსნა მონაცემთა სიმცირით მოეძებნება. სატესტო მონაცემების პირველი სამი კვირა გამოვიყენეთ შემავალ მონაცემად, რომელმაც იპროგნოზა ბოლო ერთი კვირის განმავლობაში დღიური ინციდენტები. პროგნოზის საშუალო კვადრატული შეცდომიდან ფესვის (RMSE) მნიშვნელობა იყო 3.295, რაც საუკეთესო პოგნოზია ჩვენს მიერ განხილულ მოდელებს შორის. თუმცა ეს კარგი შედეგი ნაწილობრივ განპირობებულია საპროგნოზო ჰორიზონტის სიმცირით.

6. დასკვნა

ეპიდემიებისა და პანდემიების გავრცელების პროგნოზირება მნიშვნელოვანი ამოცანაა, როგორც კერძო, ასევე სახელმწიფო სექტორისათვის. ეს კარგად წარმოჩნდა კორონავირუსით გამოწვეული პანდემიის პირობებში. ვირუსმა, რომელიც ჩინეთში, ჰუბაის პროვინციაში 2019 წლის დეკემბერში გამოჩნდა და ორი თვის განმავლობაში თითქმის მთელი მსოფლიო მოიცვა. ვირუსის გავრცელების პრევენციის ერთადერთი გზა თვითიზოლაცია და სოციალური დისტანცირება არის. ქვეყნების უმრავლესობამ მკაცრად შეზღუდა ეკონომიკური საქმიანობა. კომერციული და სახელმწიფო სექტორისთვის სასიცოცხლოდ მნიშვნელოვანი აღმოჩნდა ეპიდემიის გავრცელების მასშტაბი და ხანგრძლივობა. კერძოდ, თუ ეპიდემია მალე არ დასრულდებოდა, საჭირო გახდებოდა დისტანციური სამუშაო სისტემის გამართვა, რაც გულისხმობს თანამშრომლებისთვის კომპიუტერებისა და სხვა საჭირო აღჭურვილობის შეძენას. ამ ხარჯების თავიდან არიდება მოხდებოდა პოზიტიური პროგნოზის შემთხვევაში. ასევე, ზოგიერთი საქმიანობის სპეციფიკიდან და ფირმის ზონიდან გამომდინარე ეპიდემიის ხანგრძლივობაზე შესაძლოა ყოფილიყო დამოკიდებული თანამშრომელთა შემცირებისა და ანაზღაურების საკითხი. მეორეს მხრივ, სახელმწიფო ზრუნავს რა მარაგების შექმნაზე (სასურსათო, მედიკამენტები და სხვ.), დიდ აქცენტს უნდა აკეთებდეს ეპიდემიის პროგნოზირებაზე. მისთვის, აგრეთვე მნიშვნელოვანი გატარებული ღონისძიებების ეფექტის შეფასებაც.

მოცემულ ნაშრომი ორ მიზანს ისახავდა: პირველი, ვირუსის გავრცელების მასშტაბის (დაავადებულთა რაოდენობა ეპიდემიის ბოლოს), ეპიდემიის დასრულების სავარაუდო თარიღის და პიკის წერტილის პროგნოზირება; და მეორე, სახელმწიფოს მიერ გატარებული ღონისძიების შეფასება და საუკეთესო მიდგომის გამოვლენა.

პირველი მიზნის ამოცანების გადასაწყვეტად განვახოციელეთ პროგნოზი ეკონომეტრიკული და მანქანური სწავლების მოდელების საშუალებით. ფენომენოლოგიურმა მოდელებმა (S ფორმის მრუდებმა), მიუხედავად იმისა, რომ დღიური ინფიცირების რაოდენობას სწორად ვერ პროგნოზირებდნენ, საკმაოდ კარგად გაართვეს თავი ზემოაღნიშნული მნიშვნელოვანი თარიღებისა და მასშტაბის

პროგნოზირებას. მიღებული ციფრები საკმაოდ ახლოს იყო რეალურ შედეგებთან. მოდელების მიხედვით ეპიდემია ზაფხულში დასრულდებოდა, ხოლო დაინფიცირებულთა ჯამური ოდენობა 1500-ის ფარგლებში მეყეობდა. დღიური ინფიცირების პროგნოზირებაში საუკეთესო შედეგი აჩვენა მანქანური სწავლების ალგორითმმა - LSTM-მა, თუმცა მოდელის სპეციფიკიდან გამომდინარე პროგნოზი მხოლოდ ერთკვირიან მონაკვეთზე იქნა განხორციელებული და შესაბამისად, მისი შედეგის შედარება სხვა მოდელებთან რელევანტური არ იქნება. დამაკმაყოფილებელი შედეგი იქნა მიღებული ARIMA მოდელისა და ჩემს მიერ შემოთავაზებული პოლინომიალური მოდელის საშუალებით.

მეორე მიზნის ამოცანებისათვის გამოყენებულ იქნა SIR მოდელი და გაფართოებული SIR მოდელი. პირველ მოდელში პროგნოზი განხორციელდა სახელმწიფოს მიერ ჩარევების გაუთვალისწინებლად, ხოლო გაფართოებულ SIR მოდელში ჩართული იყო შეკავების ღონისძიებების ეფექტი. მოდელმა აჩვენა, რომ შეკავების ღონისძიებების გარეში ინფიცირებულთა რაოდენობა ორ მილიონს გადააჭარბებდა, ხოლო გატარებული ღონისძიებების შედეგად ინფიცირებულთა რაოდენობა რამდენიმე ათასამდე შემცირდა. შესაბამისად, შეგვიძლია დავასკვნათ, რომ მთავრობის რეაქცია შედეგიანი აღმოჩნდა. გარდა ამისა, eSIR მოდელით ერთმანეთს შევადარეთ გატარებული ღონისძიებების ეფექტურობა. როგორც აღმოჩნდა, ეპიდემიასთან ბრძოლის საუკეთესო გზა არის შეკავების ღონისძიებების ნელი და უწყვეტი გამკაცრება. მყისიერი და მკაცრი ქმედებები იწვევენ საზოგადოების გაღიზიანებასა და შესაძლოა მიგვიყვანოს საპირისპირო შედეგამდე.

გამოყენებული ლიტერატურა

1. ანანიაშვილი ი.; ეკონომეტრიკა; თბილისი; გამომცემლობა მერიდიანი; 2012
2. კახიანი, გ. (2004). პროგნოზირების ადაპტური ალგორითმების შექმა ხელოვნური ნეირონული ქსელების გამოყენებით.
3. Ahmadi, A., Fadaei, Y., & al., e. (2020, 03 17). Modeling and Forecasting Trend of COVID-19 Epidemic in Iran until May 13, 2020.
4. Alto, V. (2019, 06 6). Neural Networks: parameters, hyperparameters and optimization strategies.
5. Baum, S. (2020, March 27). Serial Interval of COVID-19.
6. Brownlee, J. (2019, 08 19). A Gentle Introduction to Mini-Batch Gradient Descent and How to Configure Batch Size.
7. Caicedo-Ochao, Y., & al., e. (2020, June 1). Effective Reproductive Number estimation for initial stage of COVID-19 pandemic in Latin American Count.
8. Chen, Y.-C., Lu, P.-E., & al., e. (2020). A Time-dependent SIR model for COVID-19 with Undetectable Infected Persons.
9. Du, Z., Xu, X., & al., e. (2020, 06). serial Interval of COVID-19 among Publicly Reported Confirmed Cases.
10. Ellis, P. (2020, May 9). Test positivity rates and actual incidence and growth of diseases.
11. Filaire, T. (2018, 06 17). Clustering on mixed type data.
12. Goshu, A. T., & Koya, P. R. (2013). Derivation of inflection points of nonlinear regression curves - implications to statistics curves - implications to statistics.
13. Hamidouche, M. (2020, 03 22). COVID-19 outbreak in Algeria: A mathematical Model to predict cumulative cases.
14. Hamzah, F. A., Lau, C. H., & al., e. (2020, 3 19). CoronaTracker: World-wide COVID-19 Outbreak Data Analysis and Prediction.
15. Harko, T., & al., e. (n.d.).
16. Harko, T., & al., e. (2014, 06 1). Exact analytical solutions of the Susceptible-Infected-Recovered (SIR) epidemic model and of the SIR model with equal death and birth rates.
17. Harko, T., & al., e. (2014, 06 01). Exact analytical solutions of the Susceptible-Infected-Recovered (SIR) epidemic model and of the SIR model with equal death and birth rates.
18. Hsieh, Y.-H. (2009, 04). Richards Model: A Simple Procedure for Real-time Prediction of Outbreak Severity.
19. Ippolito, P. P. (2019, September 26). Hyperparameters Optimization.
20. Kathuria, A. (2018, 1 13). Intro to optimization in deep learning: Momentum, RMSProp and Adam.

21. Krasser, M. (2018, March 21). Bayesian optimisation.

22. Larsen, J., Hansen, L., & al., e. (1996). DESIGN AND REGULARIZATION OF NEURAL NETWORKS: THE OPTIMAL USE OF A VALIDATION SET .

23. Li, Q., & Guan, X. (2020, 03 26). Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus–Infected Pneumonia.

24. Masjedi, H., Rabajante, J., & al., e. (2020, 04 22). Nowcasting and Forecasting the Spread of COVID-19 in Iran.

25. Mazurek, J., & Nenickova, Z. (2020, 04). Predicting the number of total COVID-19 cases and deaths in the USA by the Gompertz curve.

26. Mohapatra, M. (2020, 04 22). Estimation of the reproductive number of novel Corona virus (COVID-19) in India.

27. Nashiura, H., & al., e. (2020, 02 14). Serial interval of novel coronavirus (COVID-19) infections.

28. Nishiura, H., & al., e. (2020, April). Serial interval of novel coronavirus (COVID-19) infections.

29. Osthus, D., Hickmann, K., & al., e. (2017, 04 8). Forecasting seasonal influenza with a state-space SIR model.

30. Rahman, M., Ahmed, A., & al., e. (2020, 04 19). Impact of control strategies on COVID-19 pandemic and the SIR model based forecasting in Bangladesh.

31. Rizwan, M. (2018, 05 08). How to Select Activation Function for Deep Neural Network.

32. Ruder, S. (2017, 06 15). An overview of gradient descent optimization algorithms.

33. Shao, N., Pan, H., & al., e. (2020, 04 22). Multi-chain Fudan-CCDC model for COVID-19 in Iran.

34. Shao, N., Yan, Y., & al., e. (2020, 04 13). Multi-chain Fudan-CCDC model for COVID-19 -- a revisit to Singapore's case.

35. Smith, D., & Moore, L. (2004, 12). The SIR Model for Spread of Disease - The Differential Equation Model.

36. Soetewey, A. (2020, March 31). COVID-19 in Belgium.

37. Srivastava, P. (2017, December 10). Essentials of Deep Learning : Introduction to Long Short Term Memory.

38. Tiwari, S., Kumar, S., & al., e. (2020). Outbreak Trends of Coronavirus Disease–2019 in India: A predicion.

39. Tjorve, K. (2017, 06 05). The use of Gompertz models in growth analyses, and new Gompertz-model approach: An addition to the Unified-Richards family.

40. Triacca, U. (n.d.). Estimation of the parameters of an ARMA model.

41. Tsoularis, A., & Wallace, J. (2002, 07). Analysis of Logistic Growth Models.

42. Vaidyanathan, R. (2020, April 17). Estimating COVID-19's Rt in Real-Time (Replicating in R).

43. Wang, L., Zhou, Y., & al., e. (2020, 02 29). An epidemiological forecast model and software assessing interventions on COVID-19 epidemic in China.

44. Wilding, T. (2020, 3 20). Epidemic modelling of COVID-19 in the UK using a SIR model.

45. Wu, K., Darcet, D., & al., e. (2020). Generalized logistic growth modeling of the COVID-19 outbreak in 29 provinces in China and in the rest of the world.

46. Zhuang, Z., Zhao, S., & al., e. (2020, April 22). Preliminary estimates of the reproduction number of the coronavirus disease (COVID-19) outbreak in Republic of Korea and Italy by 5 March 2020.

47. https://github.com/CSSEGISandData/COVID-19/tree/master/csse_covid_19_data/csse_covid_19_time_series
48. https://www.repidemicsconsortium.org/projects/
49. https://www.youtube.com/watch?v=9IwbALQ9kdY&list=PLUZjIBGiCHFd9uMvnx-35JJmj1XwhNXtW&index=6

დანართი

დანართი 1. რანჯ-კუტას მეოთხე რიგის მიახლოება

$f(\theta_{t-1}, \beta, \gamma)$ -ს RK4-ის მიახლოება მოიცემა შემდეგი სახით:

$$f(\theta_{t-1}, \beta, \gamma) = \begin{Bmatrix} \theta_{t-1}^{S} + \frac{1}{6}[k_{t-1}^{S1} + 2k_{t-1}^{S2} + 2k_{t-1}^{S3} + k_{t-1}^{S4}] \\ \theta_{t-1}^{I} + \frac{1}{6}[k_{t-1}^{I1} + 2k_{t-1}^{I2} + 2k_{t-1}^{I3} + k_{t-1}^{I4}] \\ \theta_{t-1}^{R} + \frac{1}{6}[k_{t-1}^{R1} + 2k_{t-1}^{R2} + 2k_{t-1}^{R3} + k_{t-1}^{R4} \end{Bmatrix} \quad (ფ1.1)$$

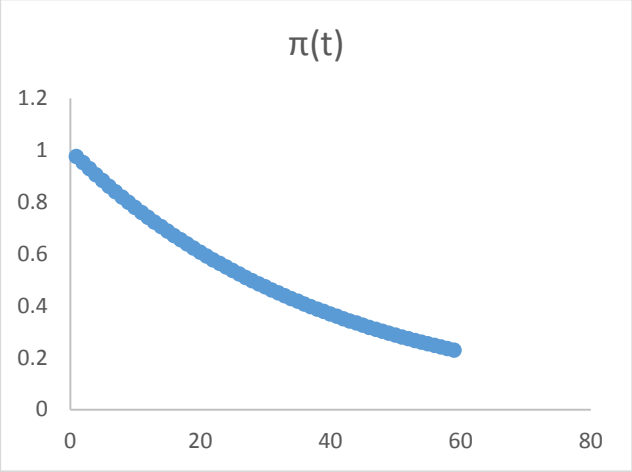
სადაც,

$$\begin{gathered} k_t^{S1} = -\beta\theta_t^S\theta_t^I \\ k_t^{S2} = -\beta[\theta_t^S + 0.5k_t^{S1}][\theta_t^I + 0.5k_t^{I1}] \\ k_t^{S3} = -\beta[\theta_t^S + 0.5k_t^{S2}][\theta_t^I + 0.5k_t^{I2}] \\ k_t^{S4} = -\beta[\theta_t^S + k_t^{S3}][\theta_t^I + k_t^{I3}] \end{gathered} \quad (ფ1.2)$$

$$\begin{gathered} k_t^{I1} = \beta\theta_t^S\theta_t^I - \gamma\theta_t^I \\ k_t^{I2} = \beta[\theta_t^S + 0.5k_t^{S1}][\theta_t^I + 0.5k_t^{I1}] - \gamma[\theta_t^I + 0.5k_t^{I1}] \\ k_t^{I3} = \beta[\theta_t^S + 0.5k_t^{S2}][\theta_t^I + 0.5k_t^{I2}] - \gamma[\theta_t^I + 0.5k_t^{I2}] \\ k_t^{I4} = \beta[\theta_t^S + k_t^{S3}][\theta_t^I + k_t^{I3}] - \gamma[\theta_t^I + k_t^{I3}] \end{gathered} \quad (ფ1.3)$$

$$\begin{gathered} k_t^{R1} = \gamma\theta_t^I \\ k_t^{R2} = \gamma[\theta_t^I + 0.5k_t^{I1}] \\ k_t^{R3} = \gamma[\theta_t^I + 0.5k_t^{I2}] \\ k_t^{R4} = \gamma[\theta_t^I + k_t^{I3}] \end{gathered} \quad (ფ1.422)$$

დანართი.2 ექსპონენციალური ფუნქცია



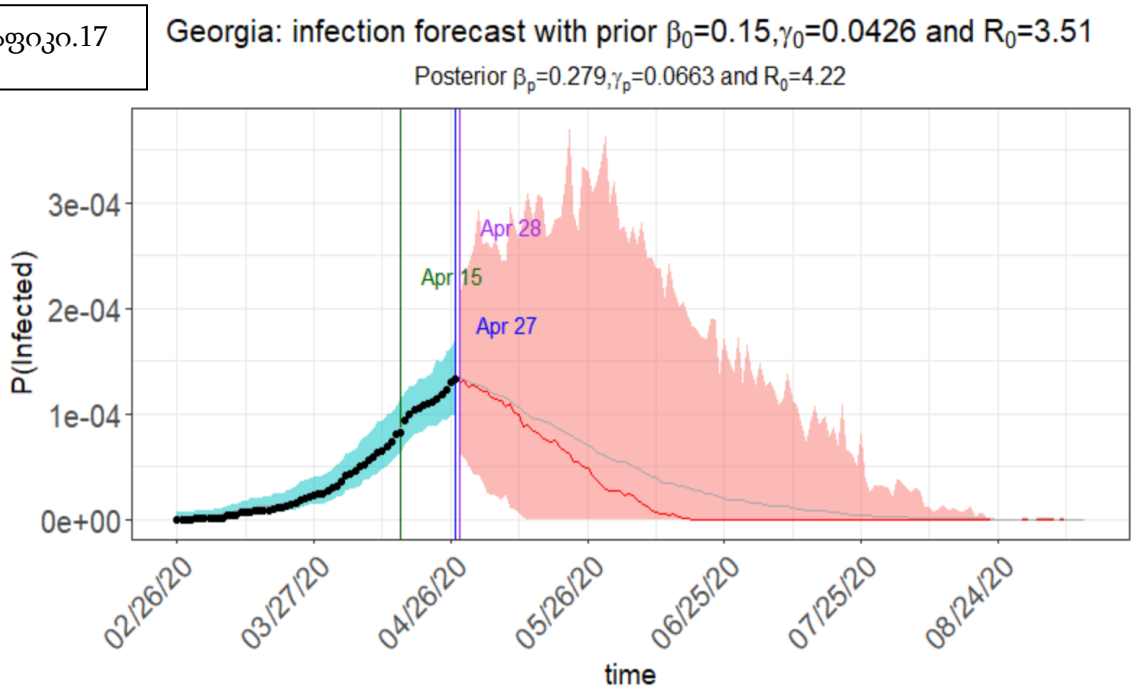
გრაფიკი.16

გრაფიკი.16-ზე მოცემულ ფუნქციას შემდეგი სახე აქვს:

$$\pi(t) = e^{-0.025t}$$

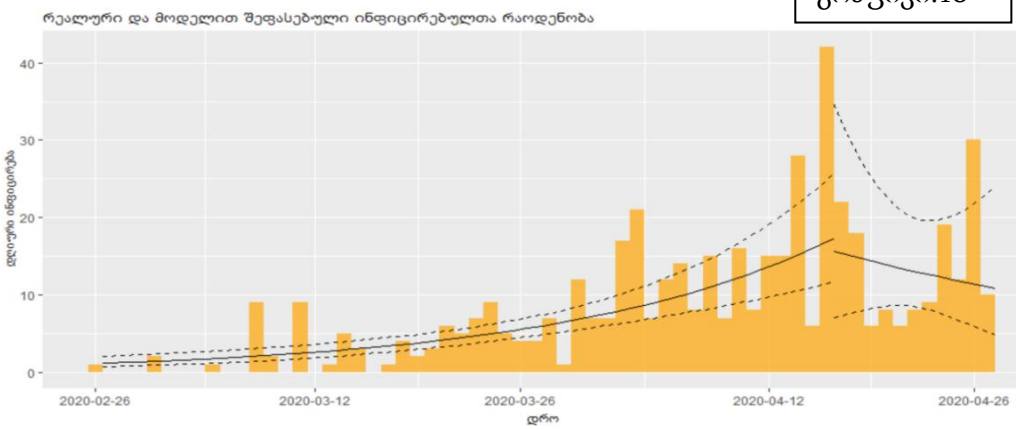
დანართი.3: ეპიდემიის პროგნოზი eSIR მოდელით, ექსპონენციალური სახის მქონე შეზღუდვების ფუნქციის გამოყენებით:

გრაფიკი.17



დანართი.4 ლოგარითმულ-წრფივი მოდელი

გრაფიკი.18



დანართი.5 ოპტიმაიზერები

ადაგრადი (Adagrad). ადაგრადი არის გრადიენტზე დაფუძვნებული ოპტიმაიზერი. ის ახდენს სწავლების კოეფიციენტის ადაპირებას პარამეტრების მიმართ, იშვიათი პარამეტრები განახლდება უფრო დიდი წონით, ანუ სწავლების კოეფიციენტის მნიშვნელობა უფრო მაღალია, ხოლო ხშირი პარამეტრები განახლდება უფრო დაბალი წონით, რაც ლოგიკურიცაა. ამ უპირატესობის საშუალებით, ეს ოპტიმაიზერი კარგად ერგება გამოხშირულ (sparse) მონაცემებს. (Ruder, 2017) ადაგრად ალგორითმის აღსაწერად შემოვიღოთ აღნიშვნები. θ-თი ავღნიშნოთ პარამეტრების ერთობლიობა, ხოლო θ_i იყოს θ ვექტორის i-ური პარამეტრი. $g_{t,i}$-ით ავღნიშნოთ სამიზნე ფუნქციის გრადიენტი θ_i-ის მიმართ:

$$g_{t,i} = \nabla_{\theta_t} C(\theta_{t,i}) \qquad (დ5.1)$$

ადაგრადი ახდენს ზოგადი სწავლების კოეფიციენტის მოდიფიკაციას თითოეულ t დროის საფეხურზე ყველა θ_i პარამეტრისათვის θ_i-ით გამოთვლილი წინა გრადიენტის საშუალებით:

$$\theta_{t+1,i} = \theta_{t,i} - \frac{\eta}{\sqrt{G_{t,ii} + \epsilon}} * g_{t,i} \qquad (დ5.223)$$

სადაც $G_t \in R^{dXd}$ არის დიაგონალური მატრიცა, რომლის თითოეული დიაგონალური ელემენტი i,i არის θ_i-ს მიმართ სამიზნე ფუნქციის გრადიენტების კვადრატების ჯამი,

ხოლო ϵ პარამეტრი გამოიყენება რათა თავიდან ავიროდოთ ნულზე გაყოფა. ადაგრადის ერთ-ერთი უპირატესობაა ის რომ მისი გამოყენების შემთხვევაში სწავლების პარამეტრის ოპტიმალური მნიშვნელობის ძიება აღარ არის საჭირო. მთავარი ნაკლი კი უკავშირდება მნიშვნელში გრადიენტის კვადრატების აკუმულირებას. სწავლების პროცესში ჯამის მნიშვნელობა იზრდება, რაც სწავლების კოეფიციენტს ძლიერ ამცირებს და ალგორითმს აღარ შეუძლია მოდელის განვითარება. ეს ხარვეზი აღმოფხვრილია ადადელტას ოპტიმაიზერში.

ადადელტა (Adadelta). ადადელტა არის ადაგრადის გაფართოება, რომელიც ჭრის სწავლების კოეფიციენტის მონოტონურად კლების პრობლემას. ყველა გრადიენტის აკუმულირების ნაცვლად, იგი ზღუდავს წინა გრადიენტების აკუმულირების სივრცეს ფიქსირებული w ზომით. ყველა წინა w რაოდენობის გრადიენტების კვადრატების შენახვის ნაცვლად, გრადიენტების ჯამი რეკურსიულად განისაზღვრება, როგორც ყველა ბოლო კვადრატული გრადიენტის კლებადი საშუალო (decaying average). საშუალო მნიშვნელობა, $E[g^2]_t$, t დროით ბიჯზე დამოკიდებულია წინა საშუალო მნიშვნელობაზე და მიმდინარე გრადიენტზე შემდეგნაირად:

$$E[g^2]_t = \gamma E[g^2]_{t-1} + (1-\gamma) g_t^2 \qquad \text{(დ5.3)}$$

γ-ს მნიშვნელობა დაახოლებით 0.9-ის ტოლია.

ადადელტას შემთხვევაში კორექტირების ნაწილი შემდეგი სახით არის წარმოდგენილი:

$$\Delta\theta_t = -\frac{\eta}{\sqrt{E[g^2]_t + \epsilon}} \qquad \text{(დ5.4)}$$

მნიშვნელი წარმოადგენს გრადიენტის საშუალო კვადრატულ (RMS) შეცდომას, ამიტომ ის შეგვიძლია ასე გადავწეროთ:

$$\Delta\theta_t = -\frac{\eta}{\sqrt{RMS[g]_t}} g_t \qquad \text{(დ5.5)}$$

რადგან განახლების ერთეული და პარამეტრის ერთეული ერთმანეთისგან განსხვავდებიან, განისაზღვრა სხვა ექსპონენციალური კლებადი საშუალო, რომლითაც ხდება პარამეტრის კვადრატების განახლება:

$$E[\Delta\theta^2]_t = \gamma E[\Delta\theta^2]_{t-1} + (1-\gamma)\Delta\theta_t^2. \qquad \text{(დ5.6)}$$

პარამეტრების განახლების საშუალო კვადრატული შეცდომა ასე ავღნიშნოთ:

$$RMS[\Delta\theta]_t = \sqrt{E[\Delta\theta^2]_T + \epsilon}. \quad (დ5.7)$$

რადგან $RMS[\Delta\theta]_t$ უცნობია, მისი მიახლოება იქნება ფესვი წინა ბიჯზე განახლებული პარამეტრის საშუალო კვადრატული შეცდომიდან. η პარამეტრს ჩავანაცვლებთ $RMS[\Delta\theta]_{t-1}$-ით, რაც მოგვცემს წონების განახლების ადადელტას წესს:

$$\Delta\theta_t = -\frac{RMS[\Delta\theta]_{t-1}}{RMS[g]_t} g_t \quad (დ5.824)$$
$$\theta_{t+1} = \theta_t + \Delta\theta_t$$

ადაპტური მომენტის შეფასება (Adam (adaptive moment estimation)). ადაპტური მომენტური შეფასება არის ოპტიმაიზერი, რომელიც ითვლის ადაპტურ სწავლების კოეფიციენტს თითოეული პარამეტრისთვის. ადადელტას მსგავსად ადამი ინახავს წინა გრადიენტის კვადრატის ექსპონენციალურ კლებად საშუალოს v_t-ს, თუმცა ამავდროულად ინახავს წინა გრადიენტის ექსპონენციალურ კლებად საშუალოს:

$$m_t = \beta_1 m_{t-1} + (1-\beta_1) g_t \quad (დ5.9)$$
$$v_t = \beta_2 v_{t-1} + (1-\beta_2) g_t^2$$

m_t და v_t წარმოადგენენ გრადიენტის პირველი და მეორე მომენტების შეფასებებს. ისინი ნულისაკენ არიან გადაადგილებულები, განსაკუთრებით საწყის ეტაპზე და როდესაც კლების კოეფიციენტები არიან მცირე (ე.ი. β_1 და β_2 არიან 1-თან ახლოს). გადაადგილების პრობლემის გადაწყვეტა ხდება კორექტირებული პირველი და მეორე მომენტების გამოთვლით:

$$\widehat{m_t} = \frac{m_t}{1-\beta_1^t} \quad (დ5.10)$$
$$\widehat{v_t} = \frac{v_t}{1-\beta_2^t}$$

კორექტირებული მომენტები გამოიყენებიან პარამეტრების განსაახლებლად:

$$\theta_{t+1} = \theta_t - \frac{\eta}{\sqrt{\widehat{v_t} + \epsilon}} \widehat{m_t} \quad (დ5.11)$$

ალგორითმის ავტორების მიერ შემოთავაზებულია შემდეგი მნიშვნელობაბი: $\beta_1 = 0.9$, $\beta_2 = 0.999$ და $\epsilon = 10^{-8}$. მათ ასევე ემპირიულად დაადასტურეს რომ ადამი პრაქტიკაშია სხვა ადაპტურ ალგორითმებზე უკეთ მუშაობს. მიუხედავად იმისა, რომ ეს არასრული სიაა ოპრიმაიზერებისა, ამ ეტაპზე დანარჩენებს არ განვიხილავთ.

დანართი.6 ჰიპერპარამეტრები

დაფარული შრეების რიცხვი: ოპტიმალური რიცხვი ან/და რიცხვის შერჩევის წესი არ არსებობს. დაფარული შრეების რაოდენობის შერჩევა დამოკიდებულია მონაცემთა რაოდენობაზე და სტრუქტურაზე. მცირე მონაცემების შემთხვევაში შესაძლოა საკმარისი იყოს 1-2 დაფარული შრე. გრაფიკების და ხელნაწერების ამოცნობის მოდელებისთვის ასევე უფრო მეტი დაფარული შრეა საჭირო, ვიდრე მარტივი დროითი მწკრივებისათვის. გასათვალისწინებელია ის ფაქტიც, რომ შრეების დამატებით გამოთვლების რაოდენობა იზრდება, რაც ზრდის სწავლების დროს. ზოგადად, ზედმეტი რაოდენობის შრეების შერჩევამ შეიძლება გამოიწვიოს მონაცემებზე ზედმეტად მორგება (overfitting) ხოლო, არასაკმარისი შრეების შემთხვევაში მოდელმა კარგად ვერ ისწავლოს მონაცემთა სტრუქტურა. ორივე შემთხვევაში მოდელის პროგნოზის სიზუსტე მცირდება.

სწავლების კოეფიციენტი (γ): ამ პარამეტრს გარკვეულწილად შევეხეთ. მისი მნიშვნელობა მოქცეულია (0;1) შუალედში. იგი გვიჩვენებს თუ რა მნიშვნელოვნებით უნდა განახლდეს წონა. გამას დაბალი მნიშვნელობისთვის ოპტიმიზაციის პროცესი ნელა მიმდინარეობს, ასევე შეიძლება წარმოიქმნას ე.წ. გრადიენტის გაქრობის პრობლემა(vanishing gradient problem). გრადიენტი ძალიან მცირდება და წონები თითქმის უცვლელნი რჩებიან. ამ დროს მოდელმა შესაძლოა ნაადრევად შეწყვიტოს სწავლება. მეორე უკიდურესობაში, როდესაც სწავლების კოეფიციენტი ზედმეტად დიდია, წონები დიდი მნიშვნელობებით იცვლებიან, რამაც შესაძლოა გამოიწვიოს დანაკარგების ფუნქციის მინიმუმის გამოტოვება. ოპტიმალური მნიშვნელობის შერჩევის კარგი მეთოდია 0.1 მნიშვნელობიდან დაწყება და შემდეგ მისი შემცირება ეტაპობრივად და პარალელურად შეცდომის მნიშვნელობაზე დაკვირვება. ამ უკანასკნელზე დეტალურად ქვემოთ ვისაუბრებ.

აქტივაციის ფუნქცია: ამ ჰიპერპარამეტრზეც უკვე ვისაუბრეთ. დამატებით ავღნიშნავ, რომ აქტივაციის ფუნქციის შერჩევისას ყოველთვის არსებობს ალტერნატივა სისწრაფესა და ეფექტურობას შორის. კერძოდ, რელუ ფუნქცია სწრაფია, მაშინ როდესაც სიგმოიდ ფუნქცია უფრო კომპლექსურია. მიიჩნევა, რომ დაფარულ შრეებისთვის

ოპტიმალურია რელუ. არგუმენტის ძალიან დიდი ან ძალიან მცირე მნიშვნელობისთვის tanh და სიგმოიდ ფუნქციის დახრა ძალიან მცირდება და სწავლის პროცესი იწელება. რაც შეეხება, გამოსავალ/საშედეგო შრეს, ამ შემთხვევაში ფუნქციის სახე შეირჩევა ამოცანის სახის (კლასიფიკაცია/რეგრესია) მიხედვით.

ქვეშერჩევის ზომა (Batch size): ეს ჰიპერპარამეტრი განსაზღვრავს ქვეშერჩევის ზომას, რომელიც გამოიყენება პარამეტრების ყოველი განახლებისათვის. საწვრთნელი მონაცემები (მთლიანი შერჩევა) შეიძლება დაიყოს ერთ ან რამდენიმე ქვეშერჩევად. მაგალითად, სტოხასტურ გრადიენტულ დაშვების ალგორიტმში გვაქვს იმდენი ქვეშერჩევა, რამდები მონაცემიცაა, რადგან, როგორც ავღნიშნე, SGD-ის შემთხვევაში პარამეტრები(წონები) ყოველი დაკვირვებისთვის ახლდება. ზოგად შემთხვევაში, მისი ზომა მთლიანი შერჩევის ზომაზეა დამოკიდებული. ე.წ. მცირე ქვეშერჩევის გრადიენტული დაშვების ალგორითმში (mini-batch gradient descent) გამოიყენება 32, 64 და 128 დაკვირვების მქონე ქვეშერჩევები. ზოგადად, მცირე ქვეშერჩევის შემთხვევაში სწავლება სწრაფად მიმდინარეობს, თუმცა გარკვეული ხმაურების (noise) ხარჯზე. შედარებით დიდი ქვეშერჩევა სწავლების ხანგრძლივობას ზრდის, თუმცა ამავდროულად იზრდება სიზუსტეც. (Brownlee, 2019)

ეპოქები (Epochs) : ჰიპერპარამეტრი განსაზღვრავს თუ რამდენჯერ უნდა გაიწვრტთნას ალგორითმი მთლიან მონაცემებზე. ეპოქების რაოდენობა ხშირად სამნიშნა რიცხვით არის წარმოდგენილი (ზოგჯერ უფრო დიდიცაა), რაც სწავლების ალგორითმს საშუალებას აძლევს მინიმუმამდე შეამციროს შეცდომა. ქვეშერჩევის განმარტებიდან ცხადია, რომ ერთი ეპოქა შეიძლება მოიცავდეს ერთ ან მეტ ქვეშერჩევას. ეპოქების ოპტიმალური რაოდენობის არჩევა ხდემა ე.წ. სწავლების მრუდის (learning curve) საშუალებით. აბცისათა ღერძზე წარმოდგენილია ეპოქების რაოდენობა, ხოლო ორდინატაზე შეცდომა. ეპოქების ოპტიმალური მნიშვნელობა არის ის, როდესაც სწავლების მრუდის დახრილობა მცირდება და დამატებითი ეპოქა უმნიშვნელოდ ამცირებს შეცდომას.

ამომგდები (dropout): ჰიჰერპარამეტრი უგულებელყოფს ზოგიერთ ნეირონს სწავლების პროცესში, რათა ნეირონული ქსელი არ იყოს ძალიან „მძიმე“. მისი გამოყენების ლოგიკა არის შემდეგი: ჩვენ არ გვინდა მოდელი ზედმეტი ინფორმაციით

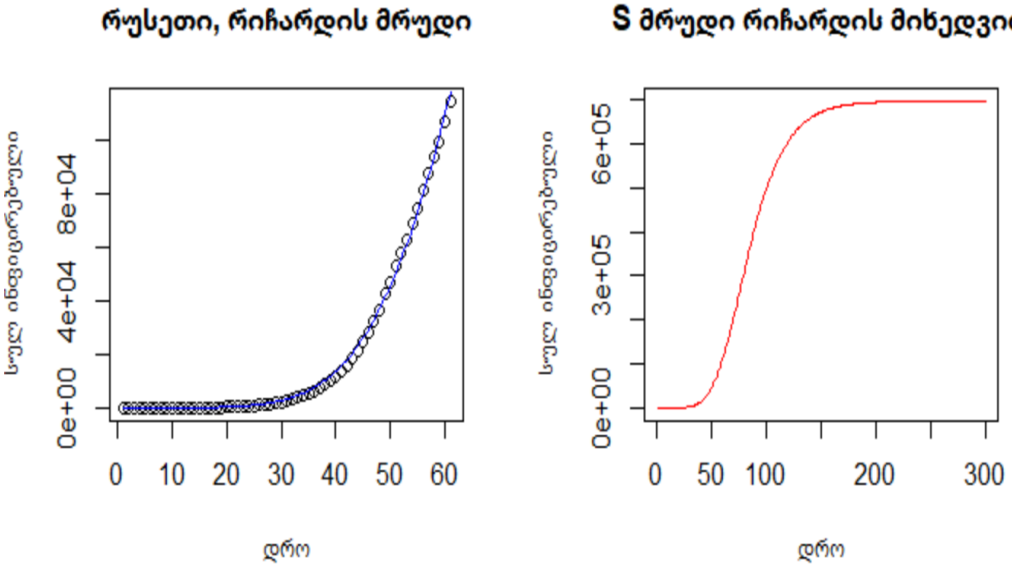
გადაიტვირთოს, გამსაკუთრებით მაშინ თუ ვვარაუდობთ რომ ზოგიერთი ნეირონი შეიძლება იყოს უსარგებლო. ალგორითმის აგების პროცესში, თითოეულ საფეხურზე, ყოველ ნეირონს ვუსადაგებთ სწავლების პროცესში დარჩენი p ალბათობას და ამოგდების (1-p) ალბათობას. (Alto, 2019)

დანართი 7. პროგნოზი კავკასიაში რიჩარდის მრუდის და ARIMA-ს საშუალებით.

რუსეთი

გრაფიკ. 19-ზე წარმოდგენილია რიჩარდის მრუდით განხორციელებული პროგნოზი რუსეთის შემთხვევაში.

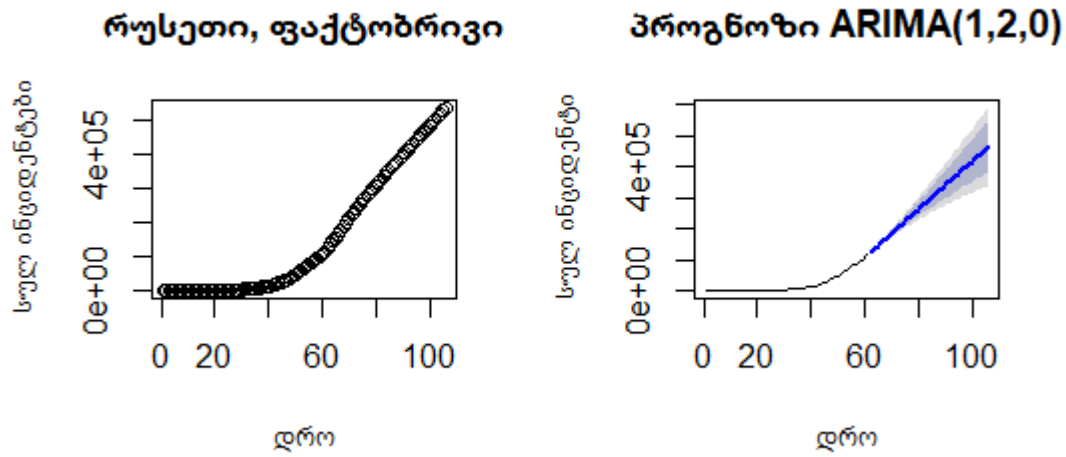
გრაფიკი.19



რიჩარდის მოდელის აგება მოხდა 61 დღის კუმულატიურ მონაცემზე. გათვალისწინებული არ იქნა პირველი ერთი თვის მონაცემები, როდესაც ინფიცირებულთა ჯამური ოდენობა არ იცვლებოდა და 3-ის ტოლი იყო. მოდელმა საშუალო კვადრატული შეცდომიდან ფესვის მიხედვით ცუდი შედეგი აჩვენა.

პროგნოზი მეტობით იყო განხოციელებული: მომდევნო 45 დღის განმავლობაში კუმულატიურმა ინფიცირებამ 536484-ს მიაღწია, ხოლო მოდელირებული მნიშვნელობა 650731-ის ტოლი იყო. საპირისპირო შედეგი მივიღეთ ARIMA(1,2,0) მოდელის შემთხვევაში. პროგნოზირება ნაკლებობით განხორციელდა (463394.5), თუმცა 80%-იან ნდობის ინტერვალში რეალური 536484 მნიშვნელობა მოექცა. გრაფიკი.20-ზე მოცემულია ARIMA-ს პროგნოზი. მიუხედავად შეცდომის სიდიდისა, პროგნოზირებული ტრენდი შეგვძლია მისაღებად მივიჩნიოთ.

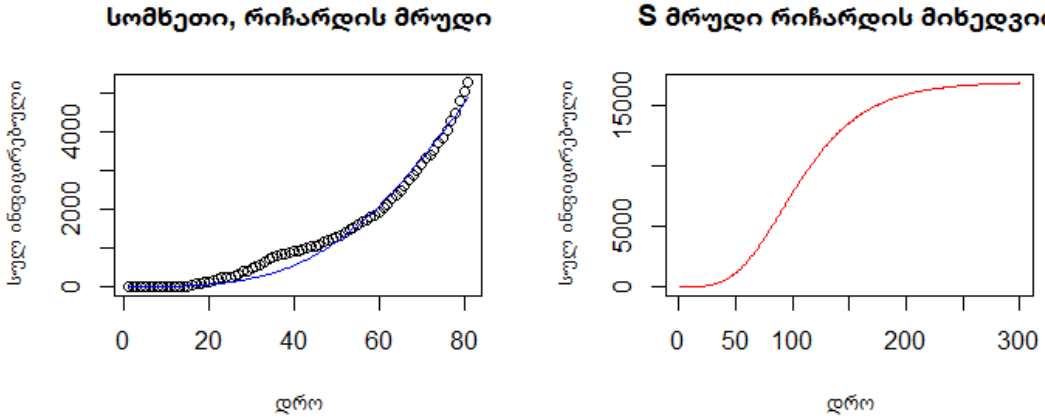
გრაფიკი.20



სომხეთი

სომხეთში ინფიცირებულთა რაოდენობა ერთ-ერთი ყველაზე მაღალია მოსახლეობასთან თანაფარდობით. პანდემიის მონაცემები ძლიერი ექსპონენციანული ტრენდით ხასიათდება. აღსანიშნავია, რომ ეპიდემიის დასაწყისში ზრდა შედარებით ნელი იყო. ამიტომ, მოდელისათვის საწვრთნელი და საპროგნოზო ჰორიზონტის კონფიგურაცია მოგვიხდა. საწვრთნელ შერჩევად ავიღეთ პირველი 81 დღის მონაცემები, ხოლო საპროგნოზოდ დანარჩენი 26 დღის მონაცემი. რიჩარდის მოდელის შედეგი მოცემულია გრაფიკი.21-ზე.

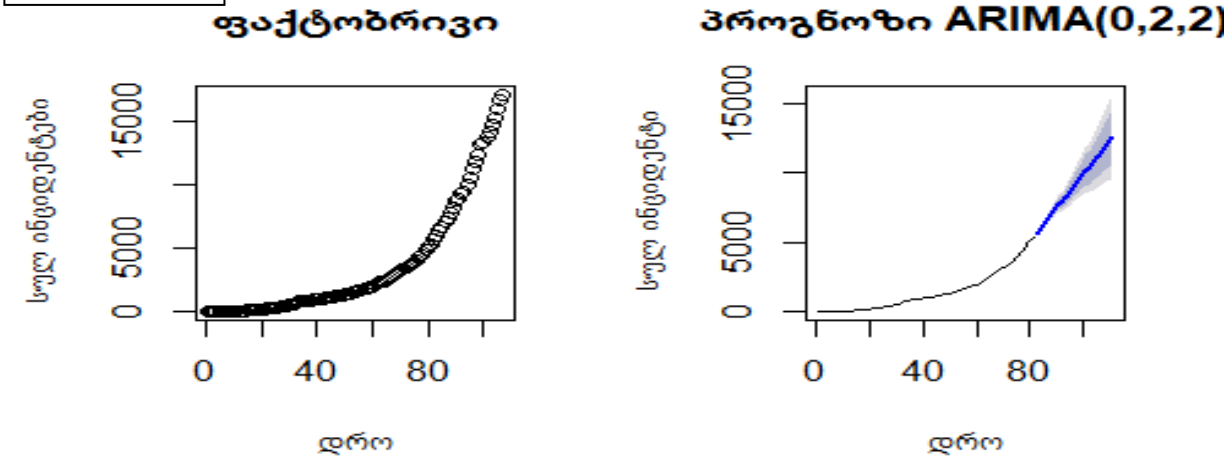
გრაფიკი.21



მოდელი ნაკლებობით აფასებს ეპიდემიის სიმკაცრეს. შეცდომა, RMSE=4858.802, ასევე, მაღალია შედარებით მცირე საპროგნოზო ინტერვალისა და მოსახლეობის რაოდენობის გათვალისწინებით. პროგნოზირებული ინფიცირება მკვეთრად ჩამორჩა ფაქტობრივ მონაცემებს. ეს შეიძლება იმით აიხსნას, რომ ეპიდემიის საწყის ეტაპზე უფრო ნაკლები ზრდა ფიქსირდებოდა ვიდრე მომდევნო პერიოდში. მოდელი სწორედ ამ საწყის მონაცემებზე გაიწვრთნა და შესაბამისად, პროგოზიც რეალურთან შედარებით მოკრძალებული გააკეთა.

სომხეთის შემთხვევაშიც ARIMA(0,2,2) პროგნოზი უფრო კარგ შედეგს იძლევა, თუმცა შეცდომა კვლავ მაღალია: RMSE=2777.323. პროგნოზირება კვლავ ნაკლებობით არის

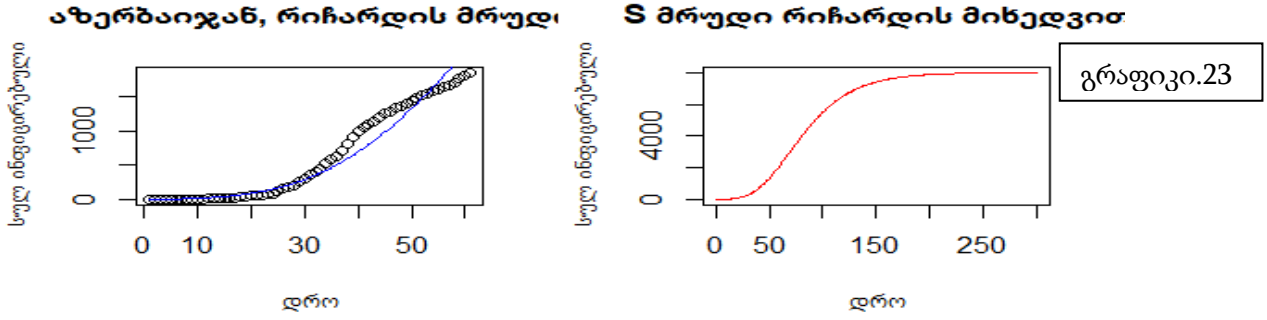
გრაფიკი.22



გაკეთებული. კერძოდ, 107-ე დღეს ინფიცირებულთა ფაქტობრივი რაოდენობა 17064-ის ტოლია, ხოლო პროგნოზირებული მხოლოდ 11635.

აზერბაიჯანი

აზერბაიჯანში ინფიცირებულთა რაოდენობა სომხეთთან შედარებით მცირეა, როგორც რაოდენობრივად, ასევე ფარდობითად. ის ეპიდემიოლოგიურად საშუალო სიმწვავის ქვეყნების ტიპიური მაგალითია.



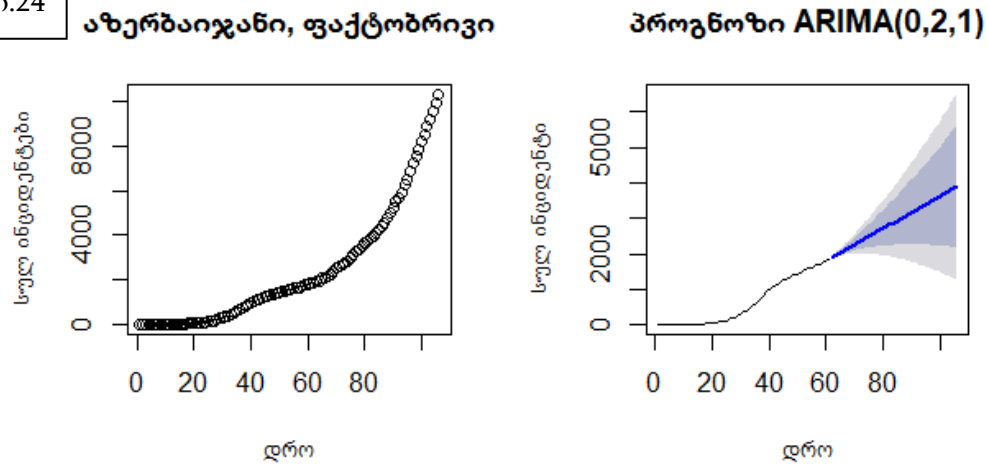
გრაფიკი.23

რიჩარდის მრუდით გაკეთებული პროგნოზის შეცდომა RMSE=2465, პროგნოზი ამ შემთხვევაშიც ნაკლებობით გაკეთდა: მოდელმა დაკვირვების ბოლო წერტილისთვის

5868 იწინასწარმეტყველა, ხოლო რეალურად ინფიცირებულთა ჯამური რაოდენობა ამ დროისათვის 10324 იყო (იხ. გრაფიკი.23).

აზერბაიჯანის მონაცემებს საუკეთესოდ მოერგო ARIMA(0,2,1) (იხ.გრაფიკი.24). თუმცა უნდა აღინიშნოს, რომ ამ გამონაკლის შემთხვევაში პროგნოზის შეცდომა რიჩარდის მრუდის შეცდომას აღემატებოდა: RMSE=2733. პროგნოზირება ამ მოდელითაც ნაკლებობით იყო განხორციელებული, რისი მიზეზიც კვლავ არასაკმარის დროით ინტერვანლში უნდა ვეძებოთ.

გრაფიკი.24



შეიძლება გაჩნდეს კითხვა, თუ რატომ არ ვზრდით ინტერვალს საკმარისი სიგრძით. პასუხი ნაშრომის მიზნიდან გამომდინარეობს: ჩვენი ერთ-ერთი მიზანი იყო პროგნოზირების განხორციელება ეპიდემიის საწყის ეტაპზე, რათა სწორად დაიგეგმოს პრევენციული ღონისძიებები. ამ შემთხვევაში 2 თვეზე უფრო დიდი ხანგრძლივობის მონაცემთა გამოყენება მიზანშეუწონელია. მართალია, ცალკეულ შემთხვევებში გამოვიყენეთ უფრო დიდი ხანგრძლივობის მონაცემები, თუმცა მოტივი ამისა იყო ის, რომ გვეჩვენებინა როგორ გაუმჯობესდება სიზუსტის ხარისხი ან/და მოდელის სპეციფიკა მოითხოვდა დიდი მოცულობის მონაცემებს.

Ivane Javakhishvili Tbilisi State University

Faculty of Economics and Business

Levan Gaprindashvili

Forecasting the Spread of Coronavirus in Georgia (Eonometric and Machine Learning Methods)

Master's Program: Economics

The work is done to obtain the academic degree of Master of Economics

Supervisor: Professor Iuri Ananiashvili,

Head of Econometrics Department,

Ivane Javakhishvili Tbilisi state university,

Faculty of Economics and Business

Tbilisi 2020