

# Predicția evoluției cazurilor de infecție cu virusul SARS-CoV2 în România

**Autori:** Prof. univ. dr. Daniel Traian Pele (coordonator), Prof. univ. dr. Miruna Mazurencu-Marinescu-Pele, Conf. univ. dr. Vasile Alecsandru Strat, Prof. univ. dr. Ileana Gabriela Niculescu-Aron, Asist. univ.dr. Raluca Dana Căplescu, Student Ionuț Alexandru Popescu, Masterand Raul Cristian Bâg, Masterand Marius-Valentin Cristian, Masterand Ionela Argentina Andrei.

Citare recomandată: Pele, D.T. et al. 2020. Predicția evoluției cazurilor de infecție cu virusul SARS-CoV2 în România, White paper, Bucharest 2020, <a href="https://github.com/danpele/Covid-19-Romania/covid_paper.pdf">https://github.com/danpele/Covid-19-Romania/covid_paper.pdf</a>
---

**Notă:** Acest material nu reprezintă în mod necesar punctul de vedere oficial al Academiei de Studii Economice din București, ci este rezultatul cercetării nefinanțate a echipei care a realizat studiul. Echipa care a realizat acest studiu este formată dintr-un grup de profesori de la Departamentul de Statistică și Econometrie, Facultatea de Cibernetică, Statistică și Informatică Economică, Academia de Studii Economice din București, precum și studenți și masteranzi ai Academiei de Studii Economice din București.

**București, aprilie 2020**

# Predicția evoluției cazurilor de infecție cu virusul SARS-CoV2 în România

- Rezultate preliminare -

## Cuprins

1. Introducere.....	2
1.1. Informații de context.....	2
1.2. Scopul studiului .....	4
2. Bazele de date.....	4
3. Modele de predicție.....	6
4. Rezultate .....	7
4.1. Time series K-means clustering .....	7
4.2. Modelul SEIR - HCD .....	11
4.3. Modelul de rețea neuronală LSTM .....	13
4.4. Modele ARIMA .....	14
4.5. Modelul NLS .....	15
5. Concluzii.....	18
6. Limitări ale studiului.....	20
7. Observații .....	20
8. Bibliografie.....	21

## 1. Introducere

### 1.1. Informații de context

#### *Debut*

Coronavirusul este o specie virală de tip ARN cu capsulă, ce cuprinde mai multe tulpini, majoritatea afectând în mod exclusiv animalele (cămile, lilieci, câini, pisici). Uneori, pot apărea mutații care facilitează transmiterea virusului și la om (SARS-CoV, MERS-CoV, SARS-CoV2). Genomul SARS-CoV2 este într-o proporție de 86.5%-96% identic cu coronavirusul întâlnit la lilieci. Primele cazuri suspecte au fost notificate către OMS pe 31 decembrie 2019<sup>1</sup>, la doar peste trei săptămâni de la apariția primelor cazuri simptomatice pe 8 decembrie 2019<sup>2</sup>. Singurul focar cunoscut a fost detectat pentru prima oară în Wuhan, China, spre sfârșitul lunii decembrie 2019.

#### *Scurte repere temporale*

Pe 26 februarie 2020 s-a înregistrat oficial primul caz de infectare cu noul coronavirus pe teritoriul României la un tânăr din județul Gorj. Deja, din 8 martie 2020, se interzic manifestările publice sau private, în spații deschise sau închise, cu un număr de participanți mai mare de 1.000 de persoane. Pe 9 martie 2020, ora 12:00, intră în vigoare deciziile Comitetului Național pentru Situații Speciale de Urgență privind impunerea unor restricții de trafic dinspre și înspre statele afectate de infecția cu noul coronavirus. Astfel, sînt sistate cursele aeriene către și dinspre Italia.

Pe 16 martie 2020 se instituie starea de urgență pe teritoriul României, iar pe 18 martie apare Ordonanța militară nr. 1/2020 privind unele măsuri de primă urgență care privesc aglomerările de persoane și circulația transfrontalieră a unor bunuri.

#### *Cele mai relevante inițiative de studiere a COVID-19, atât din punct de vedere administrativ, cât și științific*

Oficiul pentru Politica Științei și Tehnologiei de la Casa Albă (OSTP) a reunit grupuri și companii de cercetare într-o coaliție (inclusiv Kaggle) pentru a pregăti datele de cercetare în regim deschis a COVID-19 (CORD-19) pentru a încerca să abordeze întrebările științifice cheie cu privire la COVID-19. Aceste întrebări sînt formulate de către Academii Naționale de Științe, Inginerie și Medicină (NASEM) și Organizația Mondială a Sănătății (OMS).

În cele ce urmează prezentăm o selecție a celor mai utilizate modele pe plan mondial de predicție a noilor cazuri de infectare cu noul coronavirus și a deceselor.

<sup>1</sup> „Pneumonia of unknown cause – China. Disease outbreak news”. [World Health Organization](#). 5 ianuarie 2020. Arhivat din [original](#) la 7 ianuarie 2020. Accesat în 6 ianuarie 2020.

<sup>2</sup> Schnirring, Lisa (14 ianuarie 2020). „Report: Thailand’s coronavirus patient didn’t visit outbreak market”. CIDRAP (în engleză). Arhivat din [original](#) la 14 ianuarie 2020. Accesat în 15 ianuarie 2020.

## Modele statistice

- Modele autoregresive: Auto Regressive Integrated Moving Average (ARIMA) aplicat datelor epidemiologice de la Johns Hopkins pentru a determina tendința prevalenței și incidenței COVID-2019<sup>3</sup>;
- Poisson Autoregressive Model<sup>4</sup>;
- Modele de tip panel cu erori autoregresive de ordinul 1-AR(1)<sup>5</sup>.

## Modele epidemiologice

- Susceptible-Infected-Recovered-Dead (SIRD) model<sup>6</sup>;
- Stan time-varying latent variable model of COVID-19 spread<sup>7</sup>.

## În România

Dr. Mihnea Muraru (Facultatea de Automatica si Calculatoare, Universitatea Politehnica din București) a elaborat un model matematic de predicție a apariției de noi cazuri de COVID-19 în Romania, folosind Multivariate Adaptive Regression Spline - MARS.<sup>8</sup>

Bogdan Micu a făcut câteva analize asupra nivelului de infecții care va afecta România bazându-se pe Epidemic Calculator<sup>9</sup>.

Institutul pentru Măsurători și Evaluare în domeniul Sănătății, IHME funcționează în cadrul Universității din Washington și a rulat mai multe modele matematice pentru a prezice prima zi în care nu se vor mai înregistra decese în România din cauza COVID-19. Astfel, s-a ajuns la concluzia că, dacă măsurile de distanțare socială vor fi respectate, de pe 8 mai nu vor mai exista decese în România din cauza noului coronavirus. Însă, până la acel moment, arată IHME,

<sup>3</sup> Domenico Benvenuto, Marta Giovanetti, Lazzaro Vassallo, Silvia Angeletti, Massimo Ciccozzi, Application of the ARIMA model on the COVID-2019 epidemic dataset, Data in Brief, Volume 29, 2020, 105340, ISSN 2352-3409, <https://doi.org/10.1016/j.dib.2020.105340>.

<sup>4</sup> Agosto, Arianna and Giudici, Paolo, A Poisson Autoregressive Model to Understand COVID-19 Contagion Dynamics (March 9, 2020). Available at SSRN: <https://ssrn.com/abstract=3551626> or <http://dx.doi.org/10.2139/ssrn.3551626>

<sup>5</sup> Sarkodie SA, Owusu PA. Investigating the cases of novel coronavirus disease (COVID-19) in China using dynamic statistical techniques [published online ahead of print, 2020 Apr 4]. *Heliyon*. 2020;e03747. doi:10.1016/j.heliyon.2020.e03747

<sup>6</sup> a. Cleo Anastassopoulou, Lucia Russo, Athanasios Tsakris, Constantinos Siettos, Data-Based Analysis, Modelling and Forecasting of the COVID-19 outbreak, doi: <https://doi.org/10.1101/2020.02.11.20022186>. Now published in *PLOS ONE* doi: [10.1371/journal.pone.0230405](https://doi.org/10.1371/journal.pone.0230405)

b. Binti Hamzah FA, Lau C, Nazri H, Ligt DV, Lee G, Tan CL, et al. CoronaTracker: Worldwide COVID-19 Outbreak Data Analysis and Prediction. [Submitted]. Bull World Health Organ. E-pub: 19 March 2020. doi: <http://dx.doi.org/10.2471/BLT.20.255695>

<sup>7</sup> Kubinec, R., & Carvalho, L. M. (2020, April 1). A Retrospective Bayesian Model for Measuring Covariate Effects on Observed COVID-19 Test and Case Counts. <https://doi.org/10.31235/osf.io/jp4wk>

<sup>8</sup> <https://appliedpsychology.ro/covid19.html>

<sup>9</sup> [https://gabgoh.github.io/COVID/index.html?CFR=0.009&D\\_hospital\\_lag=5&D\\_incubation=5.2&D\\_infectious=2.9&D\\_recovery\\_mild=11.1&D\\_recovery\\_severe=10.75&I0=2025&InterventionAmt=0.32999999999999996&InterventionTime=2.012933333333322&P\\_SEVERE=0.1424&R0=2.42&Time\\_to\\_death=21.41&logN=16.82&fbclid=IwAR1ZMatoELgeAZ\\_baWeEs-LG\\_TbPjFBJH\\_wfDvZljirK6iuNG78L2mN5VzM](https://gabgoh.github.io/COVID/index.html?CFR=0.009&D_hospital_lag=5&D_incubation=5.2&D_infectious=2.9&D_recovery_mild=11.1&D_recovery_severe=10.75&I0=2025&InterventionAmt=0.32999999999999996&InterventionTime=2.012933333333322&P_SEVERE=0.1424&R0=2.42&Time_to_death=21.41&logN=16.82&fbclid=IwAR1ZMatoELgeAZ_baWeEs-LG_TbPjFBJH_wfDvZljirK6iuNG78L2mN5VzM)

numărul deceselor va ajunge la 804, cifră care se va menține cel puțin până la data de 4 august 2020, conform calcului matematic.<sup>10</sup>

## 1.2. Scopul studiului

Situația actuală în care virusul SARS-CoV2 prezintă o răspândire accelerată indică o necesitate de luare a unor măsuri imediate. Viteza cu care COVID-19 poate provoca incapacitate națională ilustrează faptul că într-o perioadă scurtă de timp este posibil ca situații similare cu cele văzute în China, Italia și mai noi în Spania să fie observate în alte țări ale Uniunii Europene, țările din sud-estul Europei sau Marea Britanie.

Riscul de depășire a capacității sistemului de asistență medicală în următoarea perioadă este considerat ridicat. Evaluarea impactului, înțelegerea modului în care progresează pandemia și a riscurilor asupra capacității sistemului de sănătate poate determina aplicarea unor măsuri eficiente de prevenire și control. Prin urmare, studierea dinamicii epidemiei recent apărută și în creștere rapidă este importantă, dar în același timp provocatoare din cauza cantității limitate de date disponibile.

Predicția evoluției cazurilor de infecție cu virusul SARS-CoV2 în România este crucială pentru a înțelege cum progresează epidemia. Astfel, se poate răspunde în mod adecvat la amenințare, luând măsuri eficiente sanitare, economice, sociale împotriva acesteia.

În acest studiu preliminar ne propunem să realizăm predicții privind evoluția cazurilor de infectări cu virusul SARS-CoV2 și numărului de decese cauzate de acest virus în România, folosind datele disponibile la nivel național și internațional. Plecând de la premisa că nu există model perfect, utilizăm o gamă variată de modele predictive (detaliat mai jos), fiecare model având avantajele și dezavantajele sale din punctul de vedere al descrierii realității. Predicțiile oferite de aceste modele vor fi actualizate săptămânal și vor fi disponibile publicului larg.

Aceste predicții se pot constitui într-un instrument ce poate fi utilizat de către autorități pentru o mai bună fundamentare a deciziilor în lupta cu epidemia.

De asemenea, acest studiu se dorește a fi baza pentru realizarea unei serii de lucrări științifice ce vor fi diseminate în mediul academic național și internațional.

Toate rezultatele prezentate în acest studiu sînt complet reproductibile, bazele de date și codurile sursă fiind disponibile la adresa <https://github.com/danpele/Covid-19-Romania>.

## 2. Bazele de date

Datele folosite pentru această cercetare provin din surse multiple:

- Baza de date de la Johns Hopkins University, care agregă zilnic la nivel mondial informațiile referitoare la cazurile confirmate și decesele cauzate de Covid-19 (<https://coronavirus.jhu.edu/map.html>), în perioada 22 ianuarie 2020-7 aprilie 2020.
- Baza de date realizată de echipa de cercetare pe baza comunicatelor zilnice oferite de către Grupul de Comunicare Strategică, disponibile pe site-ul Ministerului Afacerilor Interne ([www.mai.gov.ro](http://www.mai.gov.ro)). Baza de date cuprinde următoarele variabile:

<sup>10</sup> [https://www.dcnews.ro/covid-19-previziune-ihme-cand-nu-vor-mai-muri-oameni-in-romania\\_741250.html](https://www.dcnews.ro/covid-19-previziune-ihme-cand-nu-vor-mai-muri-oameni-in-romania_741250.html)

- Numărul de cazuri confirmate;
- Numărul de decese;
- Numărul de persoane în carantină;
- Numărul de persoane în izolare la domiciliu;
- Numărul de persoane internate la ATI;
- Numărul de teste efectuate.

Pentru realizarea acestei baze de date am utilizat următoarea procedură:

- Pentru perioada 26 februarie 2020-19 martie 2020 am extras informațiile din ultimul comunicat al zilei respective.
- Pentru perioada 20 martie 2020 – 7 aprilie am extras informațiile din comunicatul Grupului de Comunicare Strategică de la ora 13.00 al zilei respective.

Figura 1 prezintă evoluția numărului de cazuri confirmate, respectiv evoluția numărului de persoane aflate în izolare, carantinate și a numărului de teste efectuate.

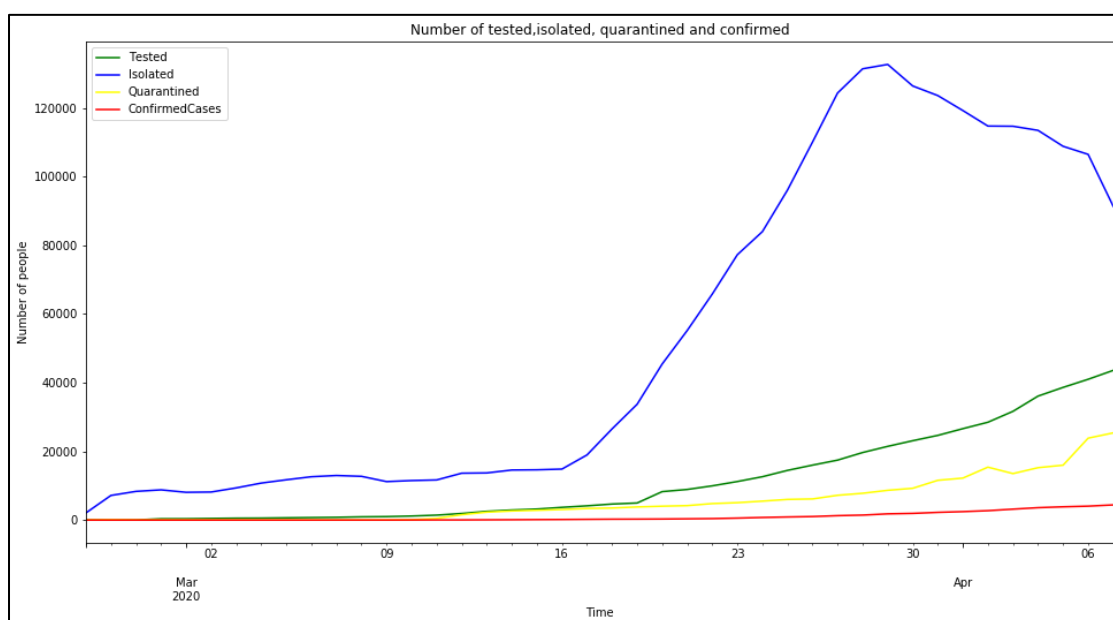


Figura 1. Evoluția numărului de cazuri confirmate, respectiv evoluția numărului de persoane aflate în izolare, carantinate și a numărului de teste efectuate în România, în perioada 22 feb 2020 - 7 apr 2020.



[Analysis](#)

Evoluția numărului de decese, cazuri vindecate și cazuri internate la terapie intensivă este prezentată în Figura 2.

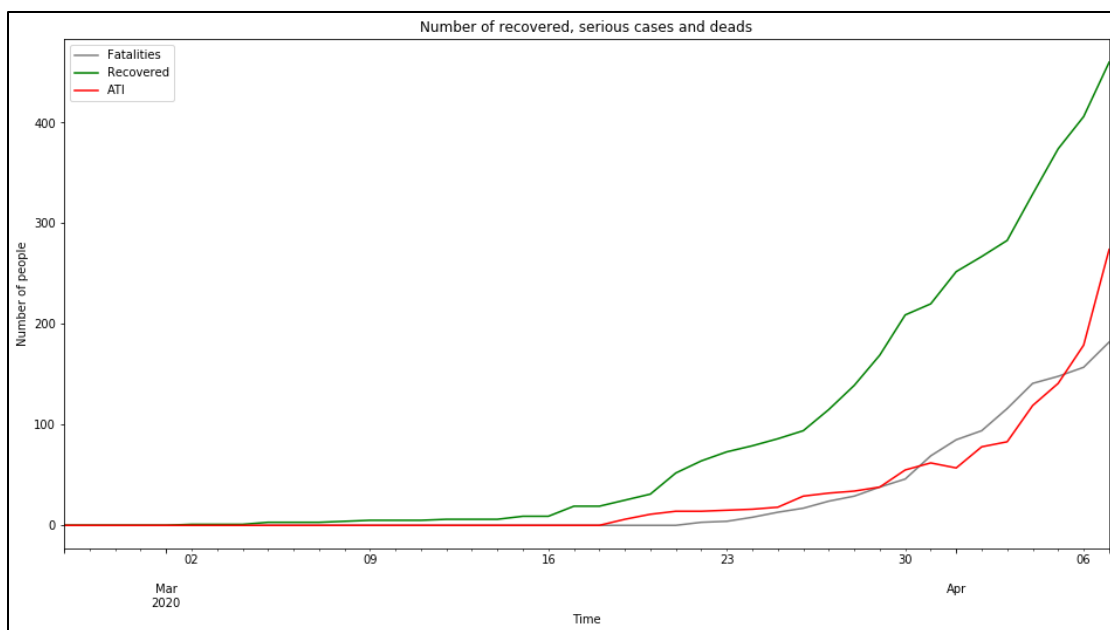


Figura 2. Evoluția numărului de decese, cazuri vindecate și cazuri internate la terapie intensivă în România, în perioada 22 feb 2020 - 7 apr 2020.



[Analysis](#)

### 3. Modele de predicție

Modelele de predicție folosite pot fi clasificate în două categorii:

- Modele epidemiologice (de exemplu SIR, SEIR);
- Modele statistice, folosind machine learning și inteligență artificială (rețele neuronale).

În cazul ambelor clase de modele am utilizat următoarea divizare a bazei de date:

- Antrenarea modelului pe datele disponibile cu privire la România în perioada 26 februarie 2020 – 31 martie 2020.
- Testarea și validarea modelului pe datele disponibile cu privire la România în perioada 1 aprilie 2020 - 7 aprilie 2020.
- Predicția pe baza modelului pentru perioada 8 aprilie 2020 - 8 mai 2020.

Drept indicator de măsurare a performanței modelului a fost ales *MSLE - Mean squared logarithmic error*:

$$MSLE = L(y, \hat{y}) = \frac{1}{N} \sum_{i=0}^N (\log(y_i + 1) - \log(\hat{y}_i + 1))^2, \text{ unde } y \text{ este valoarea reală a variabilei}$$

de interes, iar  $\hat{y}$  este valoarea estimată de model. Spre deosebire de indicatorul clasic de eroare (*MSE - Mean Squared Error*), MSLE este o măsură care corectează pentru magnitudinea variabilei de interes, i.e. ia în calcul doar diferența procentuală dintre valoarea reală și valoarea prezisă de model.

În această versiune a studiului sînt prezentate modelele utilizate, la modul general, precum și rezultatele finale pentru numărul de infectări și numărul de decese; informații mai detaliate se vor regăsi într-un articol științific, ce va fi redactat și publicat ulterior.

## 4. Rezultate

#### 4.1. Time series K-means clustering

Pentru a putea încadra tiparul de evoluție al epidemiei din România în tiparul de evoluție global, am folosit o metodă de clusterizare a evoluției<sup>11</sup> seriilor de timp (*time series K-means clustering*, vezi de exemplu Aghabozorgi et al., 2015), aplicată pentru numărul de cazuri confirmate și numărul de decese, folosind informațiile din baza de date a Johns Hopkins University. Pentru analiză, am folosit pachetul *tslearn* din Python<sup>12</sup>.

Conform acestei metode, există tipare diferite de evoluție a numărului de infectări și a numărului de decese, la nivel global.

Astfel, din punctul de vedere al evoluției numărului de infectări, România se află în același cluster cu zone precum Spania, Germania, Polonia, New York, Belgia, Suedia etc. (Figura 3).

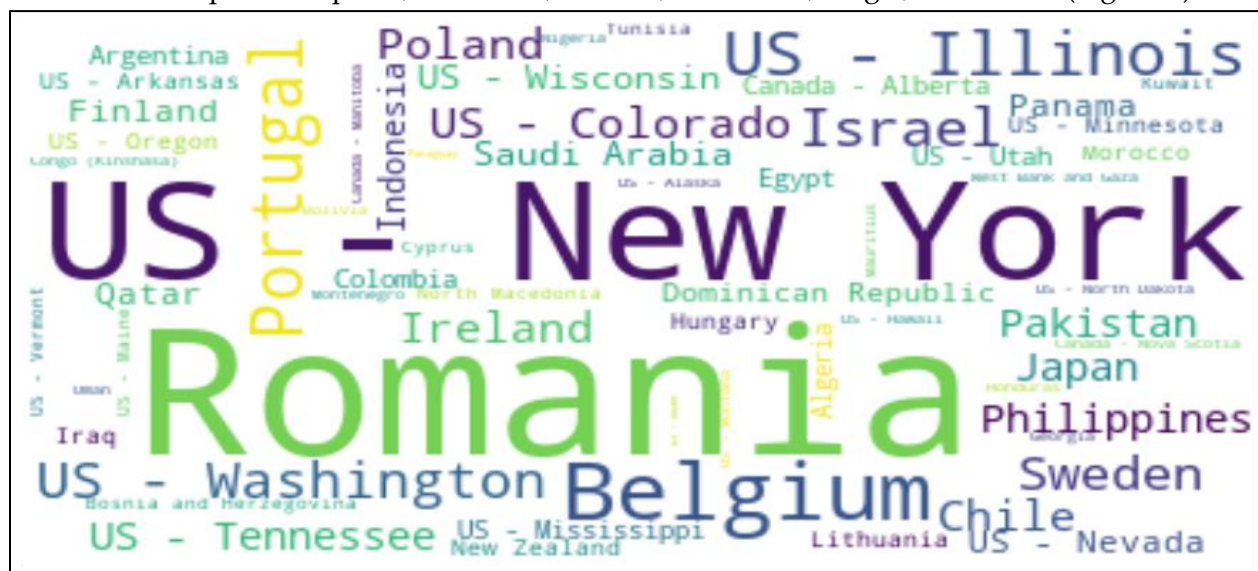


Figura 3. Componentele clusterului în care se află România din punctul de vedere al evoluției numărului de cazuri confirmate<sup>13</sup>.



## Clustering RO

<sup>11</sup> Această metodă realizează clusterizarea țărilor din punctul de vedere al traiectoriilor cazurilor confirmate și a deceselor, nu a valorilor absolute. Deci, practic, în același cluster se regăsesc țări cu traiectorii similare ale infectărilor și deceselor, după cât de lent sau abrupt este fenomenul.

<sup>12</sup> <https://buildmedia.readthedocs.org/media/pdf/tstlearn/latest/tstlearn.pdf>

<sup>13</sup> Mărimea numelor regiunilor este proporțională cu numărul de cazuri confirmate, cu excepția României, unde am folosit un factor de mărire de 33.



Conform modelului de evoluție din acest cluster, numărul de cazuri este încă pe un trend ascendent, fără a se întrezări o aplatizare a curbei (Figura 4).

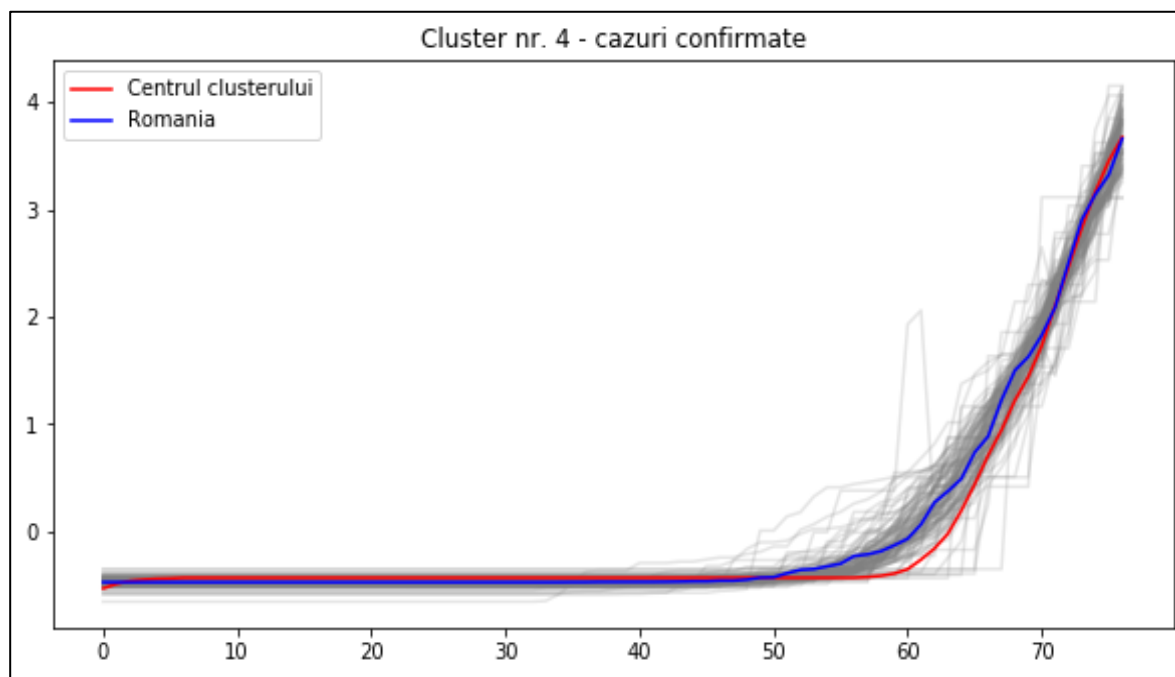


Figura 4. Clusterul în care se află România din punctul de vedere al evoluției numărului de cazuri confirmate<sup>14</sup>.



[Clustering RO](#)

Figura 5 prezintă evoluția numărului de cazuri pentru clusterul care cuprinde țări precum China, Cipru, Iran, Italia, Japonia, Coreea de Sud; aceste țări au intrat deja într-un regim de recesiune a epidemiei, curba infectărilor aflându-se într-o fază de platou.

Din punctul de vedere al numărului de decese, evoluția numărului de cazuri plasează România în același cluster cu Spania, New York, Belgia, Elveția, Germania etc. (Figura 6).

<sup>14</sup> Axa Ox: numărul de zile înregistrate de la 22 ianuarie 2020; Axa Oy: numărul de cazuri standardizat.

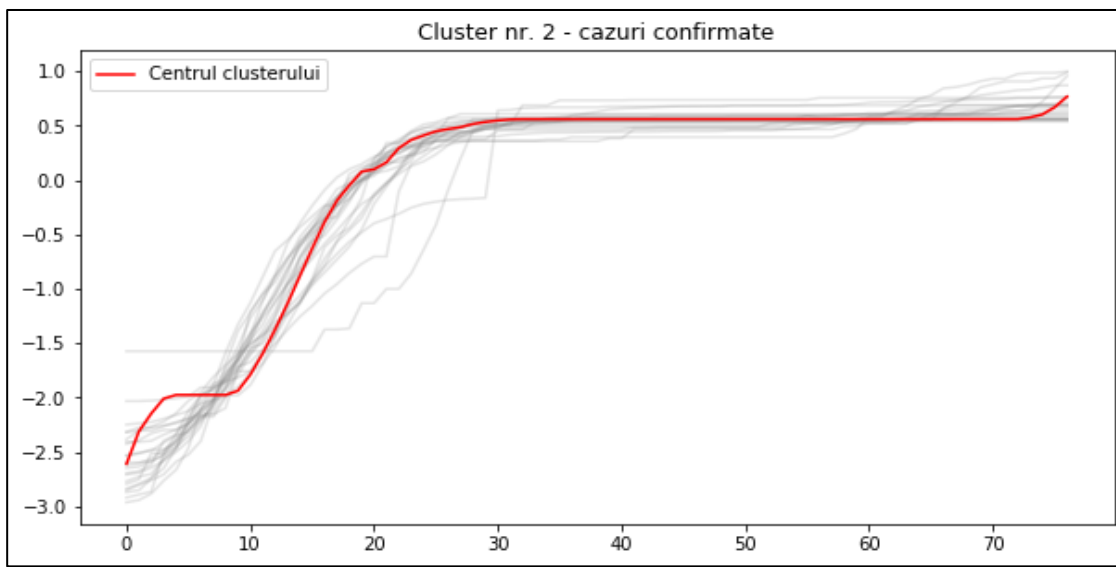


Figura 5. Clusterul în care se află China, Coreea de Sud, Italia și Iran din punctul de vedere al evoluției numărului de cazuri confirmate<sup>15</sup>.



## Clustering RO

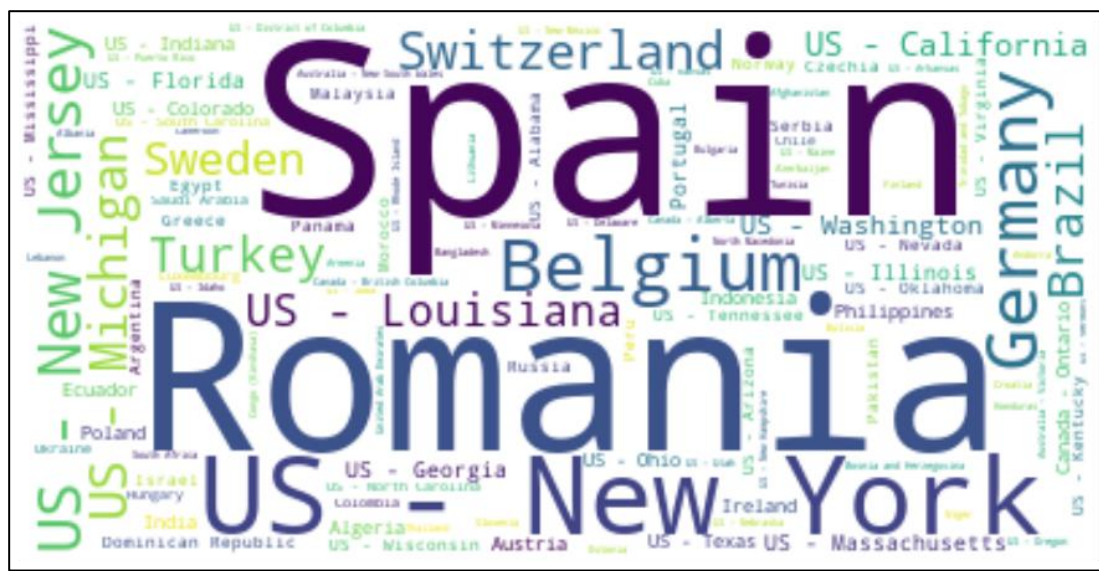


Figura 6. Componentele clusterului în care se află România din punctul de vedere al evoluției numărului de decese<sup>16</sup>.



## Clustering RO

<sup>15</sup> Axa Ox: numărul de zile înregistrate de la 22 ianuarie 2020; Axa Oy: numărul de cazuri standardizat.

<sup>16</sup> Mărimea numelor regiunilor este proporțională cu numărul de cazuri confirmate, cu excepția României, unde am folosit un factor de mărire de 33.

Conform modelului de evoluție din acest cluster, numărul de decese este încă pe un trend ascendent, fără a se întrezări o aplatizare a curbei (Figura 7).

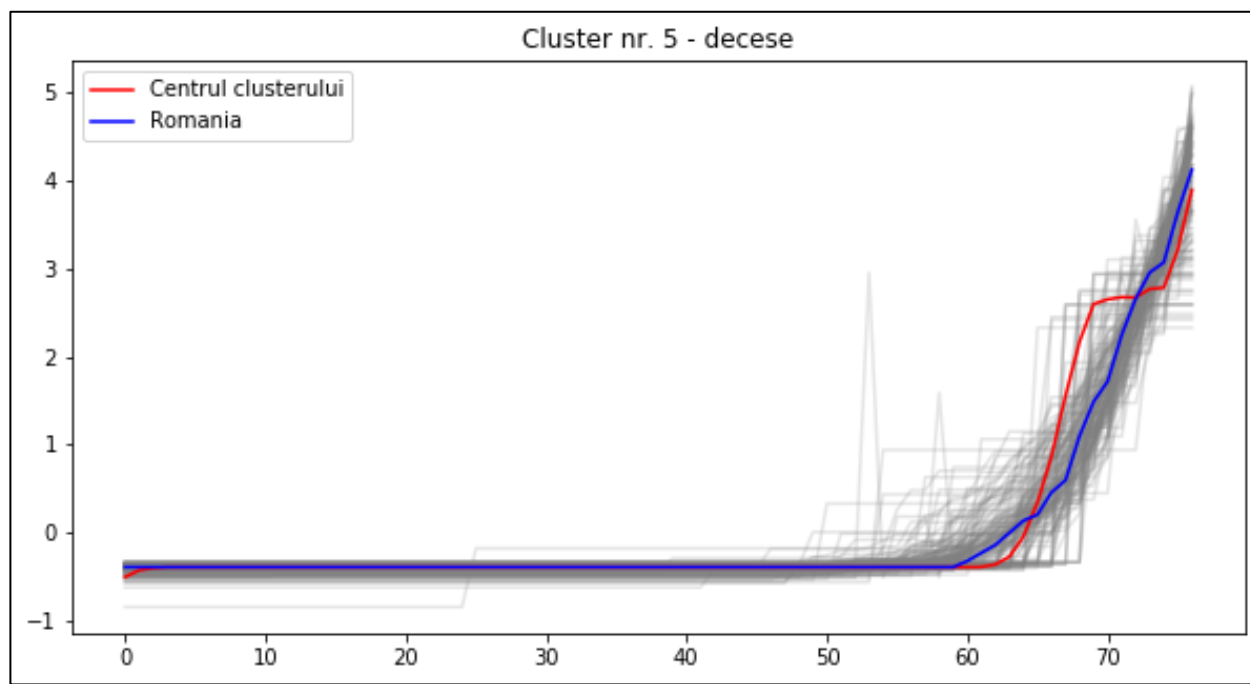


Figura 7. Clusterul în care se află România din punctul de vedere al evoluției numărului de decese<sup>17</sup>.



[Clustering RO](#)

Figura 8 prezintă evoluția numărului de decese pentru clusterul care cuprinde țări precum China, Cipru, Iran, Italia, Japonia, Coreea de Sud, țări care au intrat deja într-un regim de recesiune a epidemiei, curba deceselor aflându-se într-o fază de platou.

<sup>17</sup> Axa Ox: numărul de zile înregistrate de la 22 ianuarie 2020; Axa Oy: numărul de cazuri standardizat.

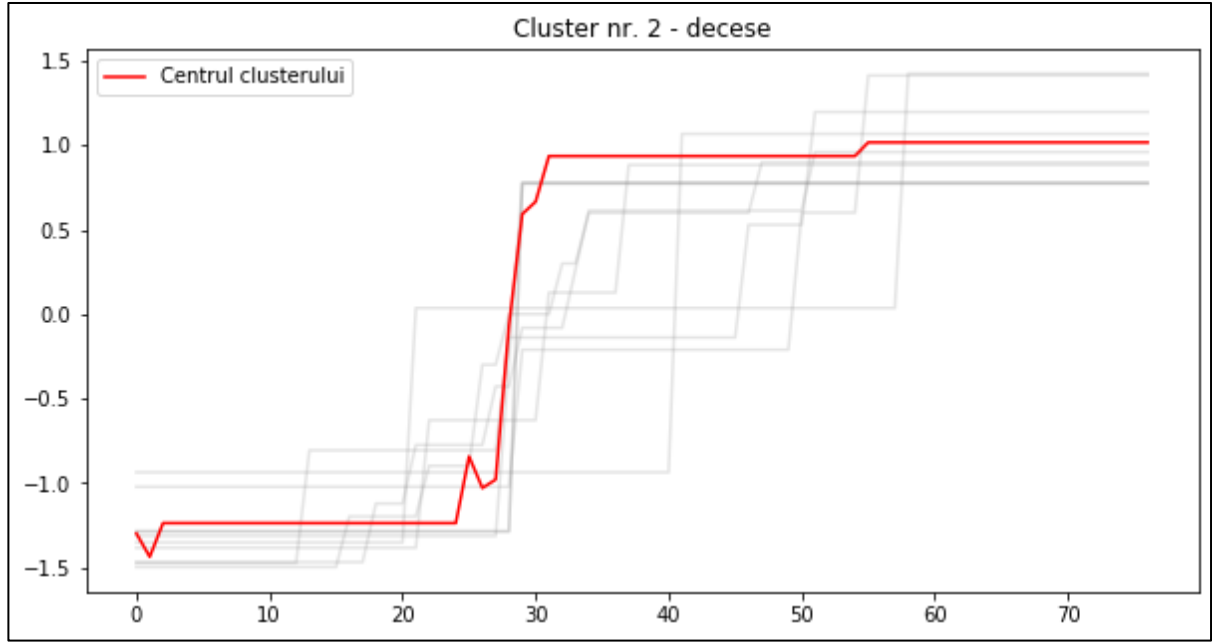


Figura 8. Clusterul în care se află China, Coreea de Sud, Italia, Iran, Japonia din punctul de vedere al evoluției numărului de decese<sup>18</sup>.



[Clustering\\_RO](#)

#### 4.2. Modelul SEIR - HCD

Modelul SEIR (*Susceptible, Exposed, Infected, Recovered – Hospitalized, Critical, Death*) este un model epidemiologic compartimental care modelează evoluția bolii folosind un set de ecuații diferențiale. În acest studiu am folosit o variantă modificată a modelului SEIR - HCD (Wu et al., 2020), care poate fi descris prin ecuațiile următoare:

- $\frac{dS}{dt} = -\frac{R_t}{T_{inf}} \cdot IS$  .
- $\frac{dE}{dt} = \frac{R_t}{T_{inf}} \cdot IS - \frac{E}{T_{inc}}$  .
- $\frac{dI}{dt} = \frac{E}{T_{inc}} - \frac{I}{T_{inf}}$  .
- $\frac{dR}{dt} = \frac{m_a I}{T_{inf}} + (1 - c_a) \frac{H}{T_{hosp}}$  .
- $\frac{dH}{dt} = \frac{1 - m_a}{T_{inf}} \cdot IS + \left[ (1 - f_a) \frac{C}{T_{crit}} \right] - \frac{H}{T_{hosp}}$  .
- $\frac{dC}{dt} = \frac{c_a H}{T_{hosp}} - \frac{C}{T_{crit}}$  .

<sup>18</sup> Axa Ox: numărul de zile înregistrate de la 22 ianuarie 2020; Axa Oy: numărul de cazuri standardizat.

$$\bullet \quad \frac{dD}{dt} = \frac{f_a C}{T_{crit}}.$$

În ecuațiile de mai sus, avem următoarele notații:

- $R_t$  este numărul de reproducere la momentul  $t$ ; dacă  $R_t < 1$ , atunci epidemia încetează.
- $S$  este numărul de indivizi care sînt susceptibili de a fi infectați.
- $E$  este numărul indivizilor care sînt expuși virusului, dar care încă nu sînt contagioși.
- $I$  este numărul indivizilor care sînt infectați și contagioși.
- $R$  este numărul indivizilor care au fost infectați, dar care s-au vindecat.
- $H$  este numărul indivizilor spitalizați, fără a fi în stare critică.
- $C$  este numărul indivizilor spitalizați în stare critică (ATI – Anestezie Terapie Intensivă).
- $D$  este numărul indivizilor decedați ca urmare a infecției.
- $T_{inc}$  - perioada de incubație (zile).
- $T_{inf}$  - perioada de contagiune (zile).
- $T_{hosp}$  - perioada în care pacientul este spitalizat înainte de a se recupera sau de a intra în stare critică (zile).
- $T_{crit}$  - perioada în care pacientul este în stare critică (fie se recuperează, fie moare) (zile).
- $m_a$  - ponderea infectărilor care sînt asimptomatice sau ușoare.
- $c_a$  - ponderea cazurilor severe care devin critice.
- $f_a$  - ponderea cazurilor critice care sînt fatale.

Pentru modelul nostru am folosit datele Grupului de Comunicare Strategică pentru estimarea parametrilor modelului SEIR-HCD, inițializînd modelul cu următoarele valori:

- $N = 19237691$  (volumul populației).
- $R_0 = 2.79$  (ECDC, 2020).
- $T_{inc} = 5$  (Chang et al., 2020).
- $T_{inf} = 8$  (Chang et al., 2020).
- $T_{hosp} = 4$  (Chang et al., 2020).
- $T_{crit} = 14$  (Chang et al., 2020).
- $m_a = 68\%$  (ECDC, 2020).
- $c_a = 0.77\%$  (ECDC, 2020).
- $f_a = 11\%$  (ECDC, 2020).

Predicțiile obținute pe baza modelului SEIR-HCD sînt prezentate în Figura 9.

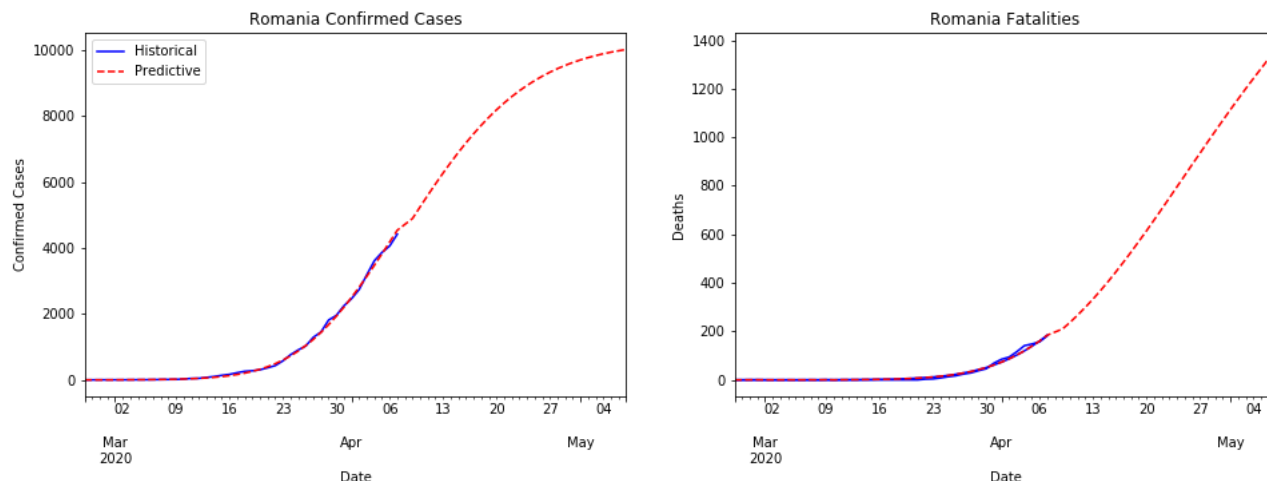


Figura 9. Predicția numărului de cazuri de infectări cu Covid-19 folosind modelul SEIR-HCD.



[SEIR\\_HCD\\_RO](http://SEIR_HCD_RO)

Potrivit acestui model, numărul de cazuri de infectare va urma un trend exponențial în următoarea lună, depășind 10000 de cazuri în prima săptămână din mai, întrevăzându-se o aplatizare a curbei. În ceea ce privește numărul de decese, modelul SEIR-HCD estimează depășirea numărului de 1200 de decese în primele zile ale lunii mai.

#### 4.3. Modelul de rețea neuronală LSTM

Modelul LSTM – *Long Short Term Memory* (Hochreiter, 1997) este un model de rețea neuronală recurentă (*RNN – Recurrent Neural Network*) care are conexiuni de tip feedback. Arhitectura standard a unei rețele LSTM este prezentată în Figura 10.

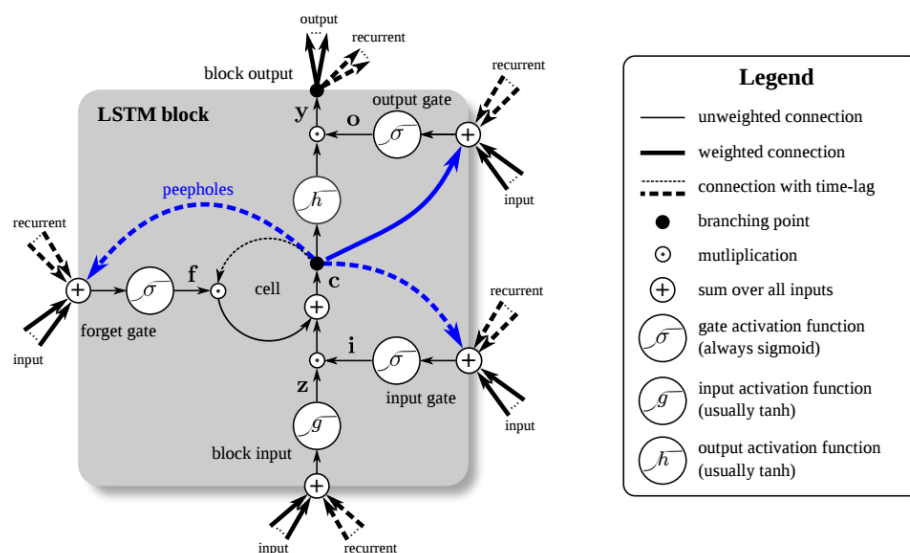


Figura 10. Arhitectura rețelei LSTM. Sursa: Greff et al. (2017)

Rețeaua LSTM a fost antrenată pe întreaga bază de date a Johns Hopkins University (306 de țări/regiuni și 23256 observații zilnice), în perioada 26 februarie 2020 – 31 martie 2020. Rețeaua are un strat vizibil cu intrare multidimensională (21113 observații zilnice, 294 țări/regiuni, 7 zile în trecut), două straturi ascunse cu 64 și 32 neuroni și un strat de ieșire care face predicție pentru numărul de cazuri confirmate și numărul de decese. Funcția implicită de activare este RELU, rețeaua este instruită pentru 200 de *epochs*, se folosește un *batch size* de 64, iar rata de învățare este 0.001.

Predicțiile obținute pe baza modelului LSTM sînt prezentate în Figura 11.

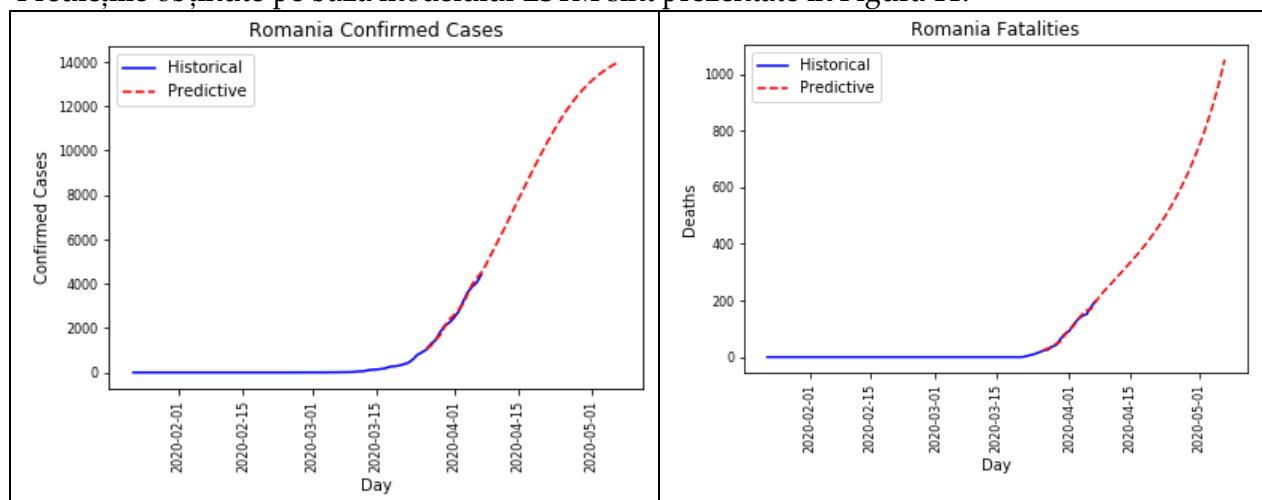


Figura 11. Predicția numărului de cazuri de infectări cu Covid-19 folosind LSTM.



[LSTM RO](#)

Potrivit acestui model, numărul de cazuri de infectare va urma un trend exponențial în următoarea lună, depășind 12000 de cazuri în ultima decadă din aprilie, întrevăzându-se o aplatizare a curbei începînd cu prima săptămîină din mai.

În ceea ce privește numărul de decese, modelul LSTM estimează depășirea numărului de 1000 de decese în primele zile ale lunii mai.

#### 4.4. Modele ARIMA

Pe baza datelor disponibile, au fost estimate modele ARIMA ( $p, d, q$ ) pentru evoluția numărului de cazuri confirmate și a numărului de decese:

$$X_t \sim ARIMA(p, d, q):$$

$$(1 - \phi_1 L - \dots - \phi_p L^p)(1 - L)^d X_t = \mu + (1 + \theta_1 L + \dots + \theta_q L^q) \varepsilon_t, \quad \varepsilon_t \sim WN(0, \sigma_\varepsilon^2)$$

Figura 12 prezintă rezultatele unui prim model ARIMA estimat pentru numărul de cazuri noi confirmate în fiecare zi și numărul de decese noi confirmate zilnic.

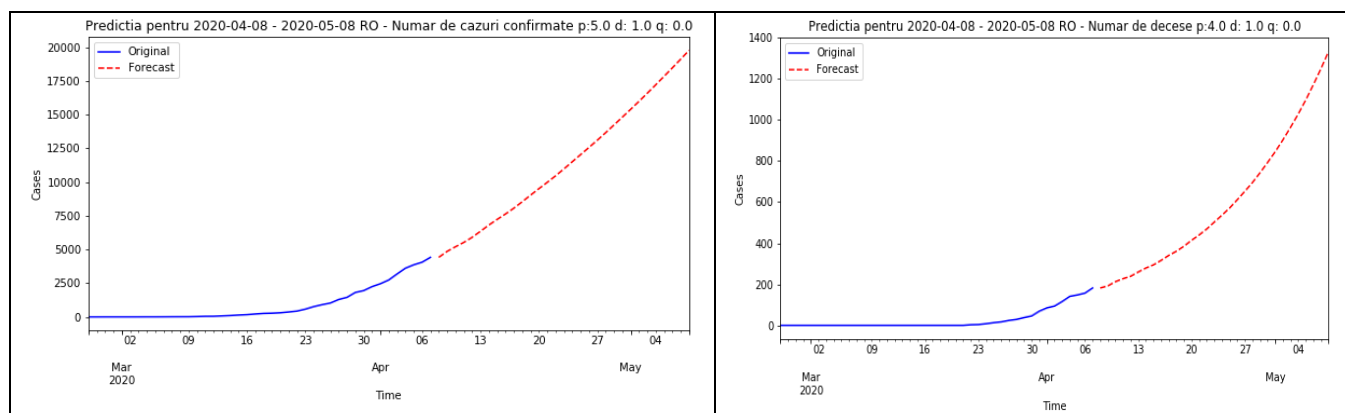


Figura 12. Predicția pe baza modelelor ARIMA (5,1,0) estimat pentru numărul de cazuri și ARIMA (4,1,0) estimat pentru logaritmul numărului de decese.



#### [ARIMA RO 1](#)

Conform acestor modele, în prima săptămână din mai numărul de infectări ar putea ajunge la 20000 de cazuri, iar numărul de decese ar putea ajunge la 1200 de cazuri.

Figura 13 prezintă predicția pe baza unui set de modele ARIMA pentru numărul de cazuri și numărul de decese, modele estimate pe numărul zilnic cumulat de cazuri. Conform acestor modele, în prima săptămână din mai se pot depăși 12500 de infectări cu SARS-CoV2 în România și se poate trece de pragul de 600 de decese cauzate de acest virus.

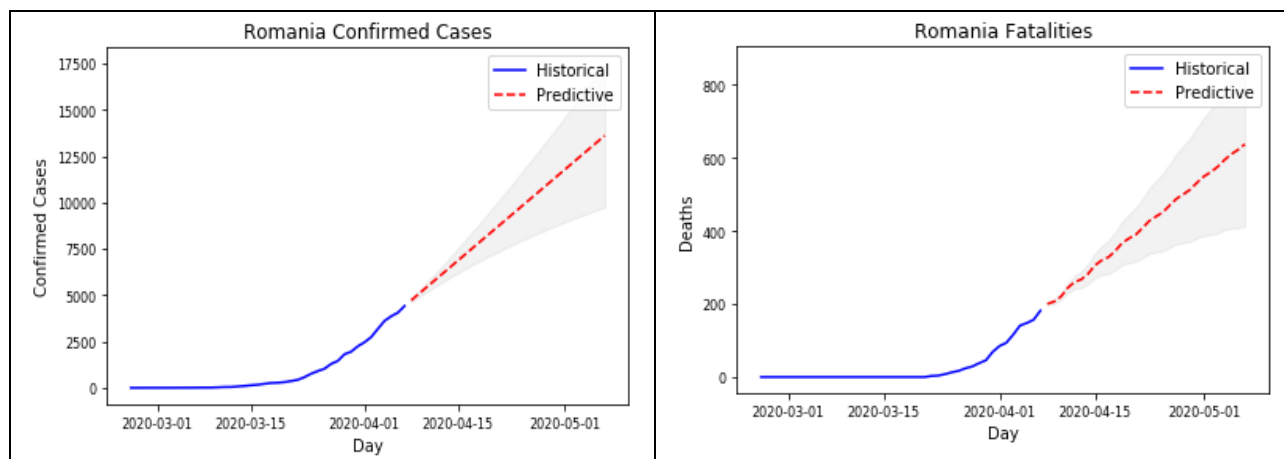


Figura 13. Predicția pe baza modelelor ARIMA (0,2,1) – numărul de cazuri și ARIMA (2,2,2) – numărul de decese.



#### [ARIMA RO 2](#)

### 4.5. Modelul NLS

Modelul NLS (*Nonlinear least squares*) realizează o estimare a metodei celor mai mici pătrate pentru un model neliniar.

Modelul neliniar ales pentru estimarea trendului infecțiilor SARS-Cov2, este o regresie logistică de forma:



$$y(t) = \frac{K}{1 + e^{\frac{-(t-b)}{a}}}$$

unde:

- $y(t)$  – Numărul de infectări la momentul  $t$ ;
- $K$  – Numărul total de infectări posibil;
- $t$  – Timpul (numărul de zile de la înregistrarea primului caz de infectare);
- $b$  – Ziua în care s-a atins cel mai mare număr de infectări;
- $a$  – Rata de infectare.

Testarea modelului s-a realizat pe datele din China de la începutul pandemiei până în ziua cu cel mai mare număr de infectări, obținând rezultatele din Figura 14.

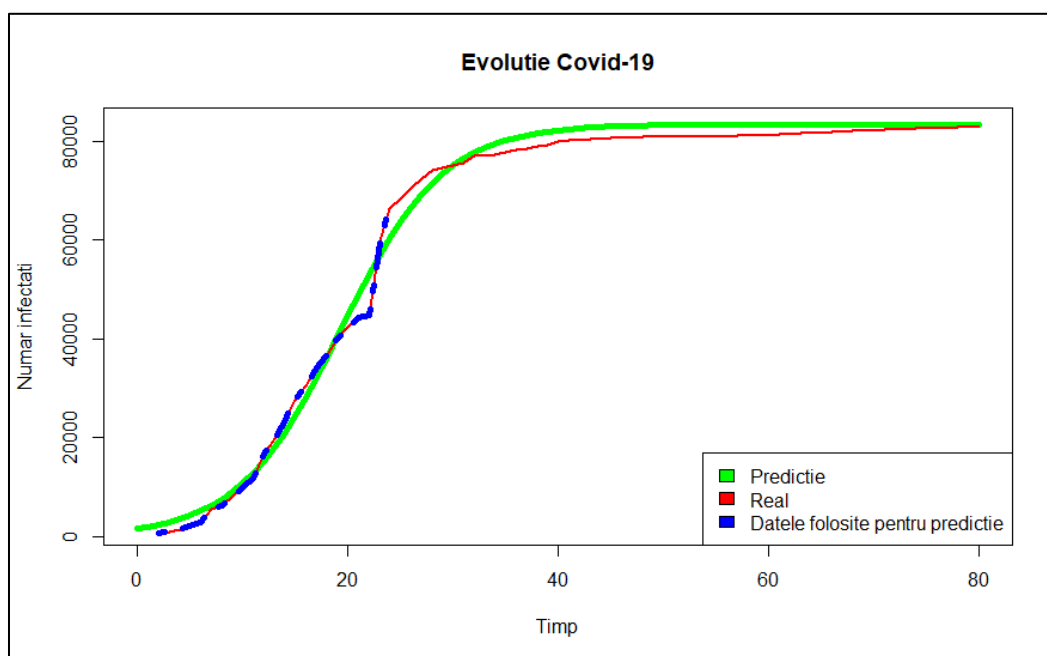


Figura 14. Testarea capacității de predicție a modelului pentru China



[NLS](#)

Modelul reușește să estimeze destul de bine creșterea și punctul maxim pe care îl atinge virusul SARS-Cov2 în China. Predicția obținută în România pe baza acestui model este descrisă în Figura 15.

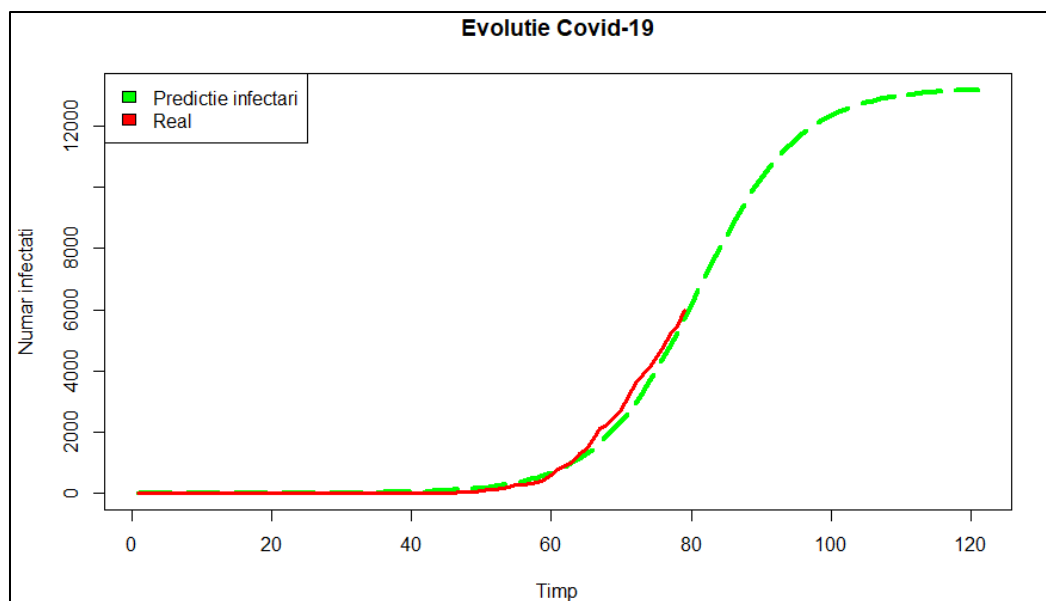


Figura 15. Estimarea modelului logistic pentru infecțiile cu Covid-19 în România



[NLS](#)

Modelul prezice un număr total de infectați de 13240, maximum fiind atins la începutul lunii mai. Ziua cu cele mai multe cazuri de infectări (variabila „b”) a fost 11 aprilie, sau în cazul nostru ziua cu numărul 80 de la începerea înregistrării datelor.

Pentru a obține o prezicere a numărului de vindecări am folosit același model, decalând „b” cu valoarea declarată de OMS (Organizația Mondială a Sănătății) pentru timpul de vindecare, și anume până la 2 săptămâni pentru persoanele infectate cu o versiune ușoară sau moderată a virusului și 3 spre 6 săptămâni pentru persoanele cu o versiune agresivă a bolii (*Report of the WHO-China Joint Mission on Coronavirus Disease 2019*), rezultatele fiind prezentate în Figura 16.

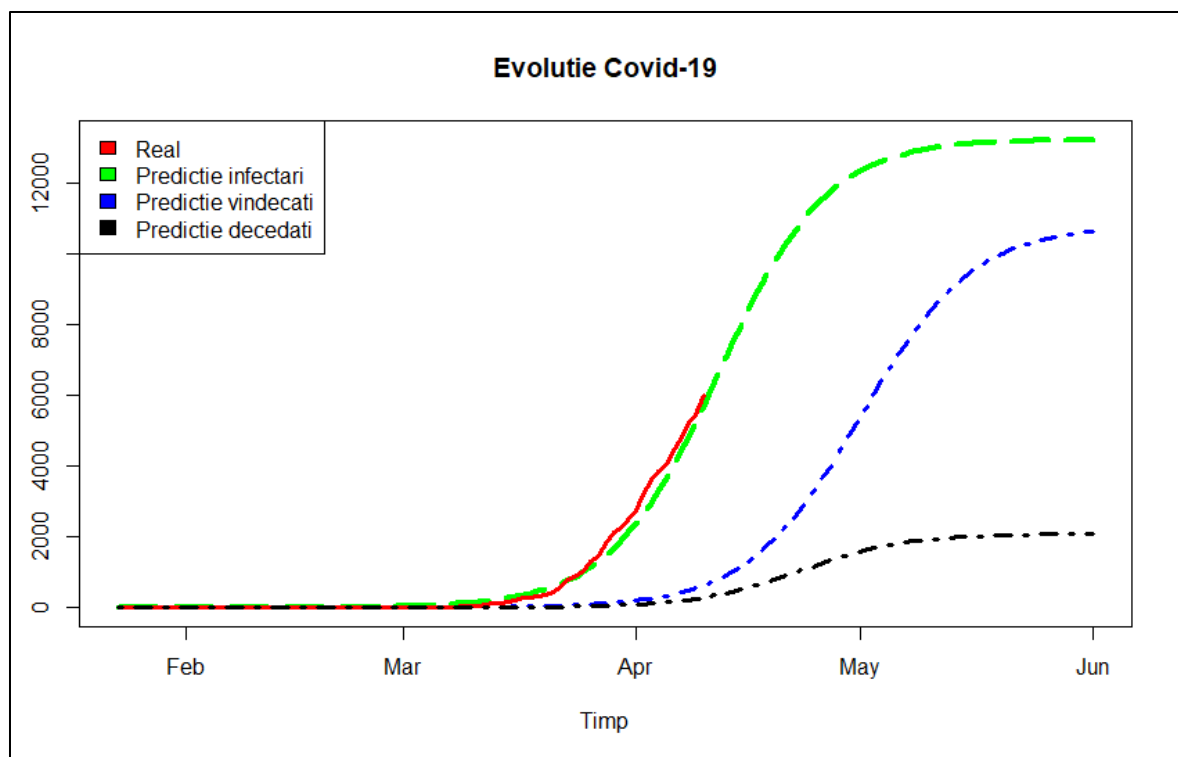


Figura 16. Predicția pe baza modelului logistic



[NLS](#)

## 5. Concluzii

Pe baza modelelor estimate se conturează trei scenarii privind evoluția numărului de cazuri confirmate pentru următoarea lună (Figura 17):

- Scenariul optimist – conform căruia în prima săptămână din mai, numărul de cazuri confirmate ar putea ajunge la 10000-12000, iar curba ar începe să se aplatizeze (conform modelelor SEIR-CHD, NLS).
- Scenariul median – conform căruia, în prima săptămână din mai, numărul de cazuri confirmate ar putea ajunge la 15000, fără aplatizarea curbei (modelele LSTM\_RO, ARIMA\_RO\_2).
- Scenariul pesimist - conform căruia, în prima săptămână din mai, numărul de cazuri confirmate ar putea ajunge la 20000, fără aplatizarea curbei (modelul ARIMA\_RO\_1).

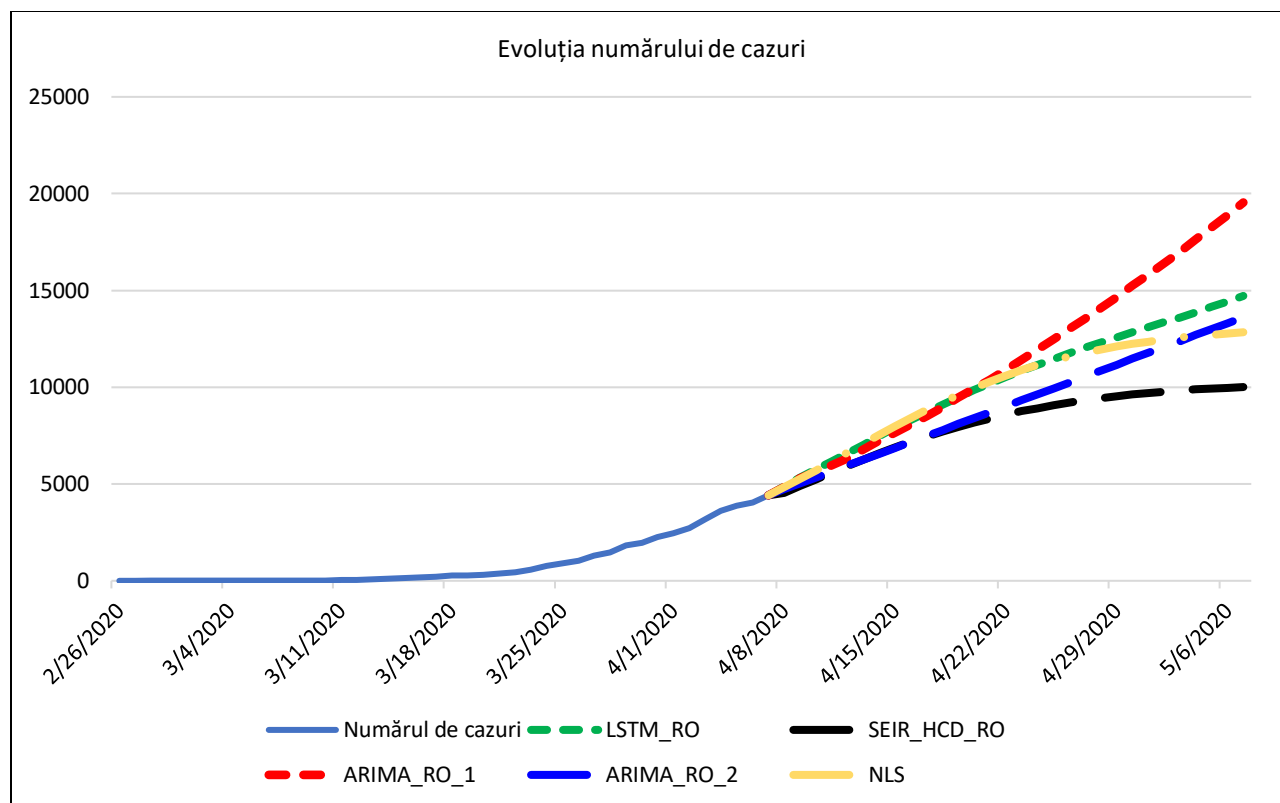


Figura 17. Evoluția numărului de cazuri, pe baza predicției

În ceea ce privește numărul de decese, se conturează de asemenea trei scenarii (Figura 18):

- Scenariul optimist – conform căruia, în prima săptămână din mai, numărul de decese ar putea ajunge la 600 iar curba ar începe să se aplatizeze (conform modelului ARIMA\_RO\_2).
- Scenariul median – conform căruia, în prima săptămână din mai, numărul de decese ar putea ajunge la 1400 (modelele LSTM\_RO, SEIR-HCD).
- Scenariul pesimist - conform căruia, în prima săptămână din mai, numărul de decese ar putea ajunge la 1800-2000 (modelele ARIMA\_RO\_1, NLS).

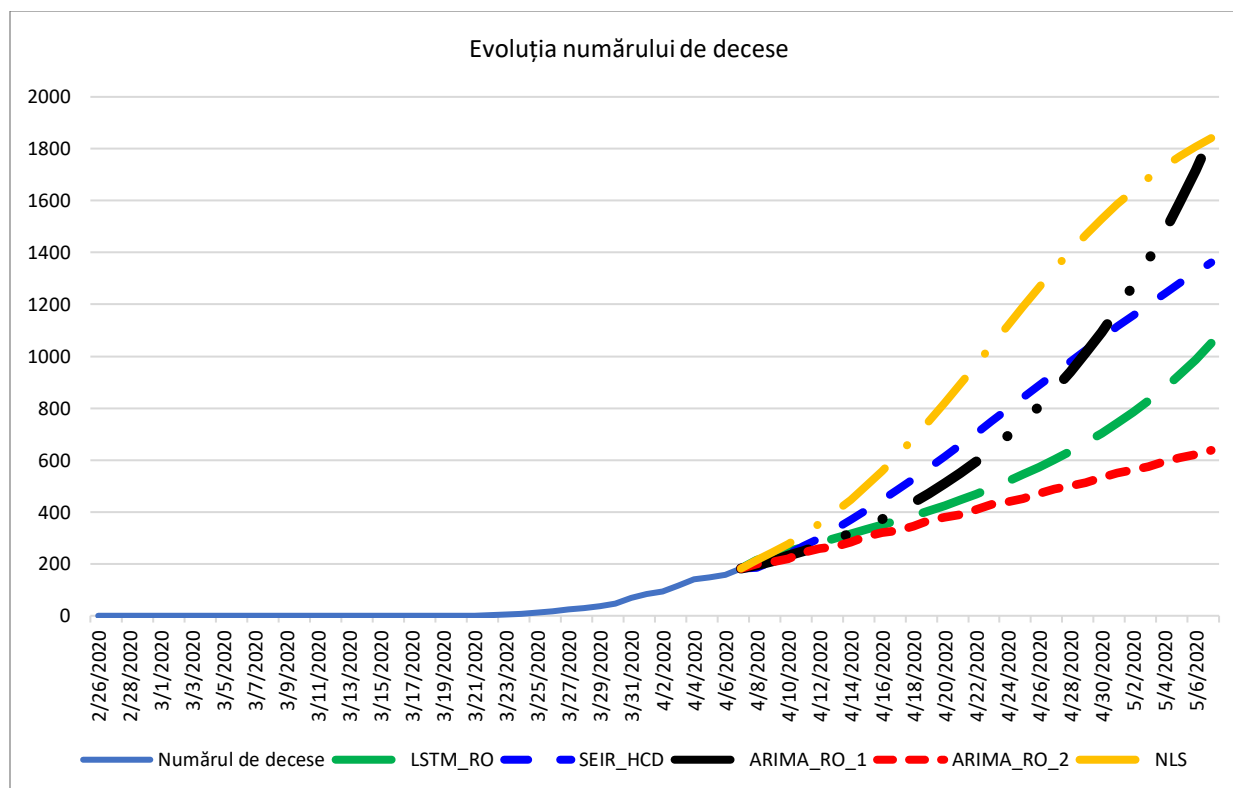


Figura 18. Evoluția numărului de decese, pe baza predicției

Predicția pe baza acestor modele va fi actualizată folosind informațiile ce vor fi disponibile pînă la data de 15 aprilie 2020.

## 6. Limitări ale studiului

Studiul nostru are cîteva limitări evidente, printre care amintim:

- Serii de date disponibile pentru România au lungime temporală relativ redusă, ceea ce poate duce la predicții eronate.
- Predicțiile obținute pe baza modelelor prezentate aici se bazează pe ipoteza implicită că evoluțiile din trecut se vor replica în viitor.
- Incidența virusului SARS-Cov2 este diferită pe grupe de vîrstă și pentru diferite patologii; existența acestor informații suplimentare ar aduce o îmbunătățire a gradului de acuratețe a predicțiilor.
- Nu se poate estima cu acuratețe gradul de conformare a cetățenilor la restricțiile impuse de autorități, prin ordonanțe militare.

## 7. Observații

- Toate programele folosite pentru estimarea acestor modele, precum și bazele de date, sînt disponibile la această adresă: <https://github.com/danpele/Covid-19-Romania>.

- Începînd cu 17 aprilie 2020, predicțiile vor fi actualizate săptămînal, pe baza datelor disponibile.
- Pentru comunicare, vă rugăm să folosiți adresa de email [danpele@ase.ro](mailto:danpele@ase.ro).
- Acest studiu nu a beneficiat de finanțare publică sau privată, autorii contribuind în mod voluntar la realizarea cercetării.

## 8. Bibliografie

1. Hochreiter, S., Schmidhuber, J. 1997. Long short-term memory. *Neural Computation*. 9 (8): 1735–1780. <https://doi.org/10.1162/neco.1997.9.8.1735>.
2. K. Greff, R. K. Srivastava, J. Koutník, B. R. Steunebrink and J. Schmidhuber, 2017. LSTM: A Search Space Odyssey, *J. IEEE Transactions on Neural Networks and Learning Systems*, vol. 28, no. 10, pp. 2222–2232, Oct. 2017. <https://doi.org/10.1109/TNNLS.2016.2582924>.
3. E. Cheynet, 2020. Generalized SEIR Epidemic Model (fitting and computation), <https://www.github.com/ECheyne/SEIR> (Accessed 9 April 2020).
4. Wu JT, Leung K, Leung GM, 2020. Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study. [https://doi.org/10.1016/S0140-6736\(20\)30260-9](https://doi.org/10.1016/S0140-6736(20)30260-9).
5. ECDC, 2020. Coronavirus disease 2019 (COVID-19) in the EU/EEA and the UK – eighth update - 8 April 2020, <https://www.ecdc.europa.eu/sites/default/files/documents/covid-19-rapid-risk-assessment-coronavirus-disease-2019-eighth-update-8-april-2020.pdf> (Accessed 9 April 2020).
6. Chang, D., Mo, Guoxin, Yuan, Xin, Tao, Yi, Peng, Xiaohua, Wang, Fusheng, 2020. Time Kinetics of Viral Clearance and Resolution of Symptoms in Novel Coronavirus Infection. *J. American journal of respiratory and critical care medicine*. <https://doi.org/10.1164/rccm.202003-0524LE>.
7. Domenico Benvenuto, Marta Giovanetti, Lazzaro Vassallo, Silvia Angeletti, Massimo Ciccozzi, 2020. Application of the ARIMA model on the COVID-2019 epidemic dataset. <https://doi.org/10.1016/j.dib.2020.105340>.
8. Agosto, Arianna and Giudici, Paolo, 2020. A Poisson Autoregressive Model to Understand COVID-19 Contagion Dynamics. <http://dx.doi.org/10.2139/ssrn.3551626>.
9. Sarkodie SA, Owusu PA, 2020. Investigating the cases of novel coronavirus disease (COVID-19) in China using dynamic statistical techniques, *J. Heliyon*. <https://www.sciencedirect.com/science/article/pii/S2405844020305922>.
10. Cleo Anastassopoulou, Lucia Russo, Athanasios Tsakris, Constantinos Siettos, 2020. Data-Based Analysis, Modelling and Forecasting of the COVID-19 outbreak. <https://doi.org/10.1101/2020.02.11.20022186>.
11. Binti Hamzah FA, Lau C, Nazri H, Ligot DV, Lee G, Tan CL, et al., 2020. CoronaTracker: Worldwide COVID-19 Outbreak Data Analysis and Prediction. E-pub: 19 March 2020. <http://dx.doi.org/10.2471/BLT.20.255695>.
12. Kubinec, R., & Carvalho, L. M., 2020. A Retrospective Bayesian Model for Measuring Covariate Effects on Observed COVID-19 Test and Case Counts. *J. SocArXiv*, <https://doi.org/10.31235/osf.io/jp4wk>.

13. Report Coronavirus Disease, 2020. Report of the WHO-China Joint Mission on Coronavirus Disease 2019 (COVID-19). <https://www.who.int/docs/default-source/coronaviruse/who-china-joint-mission-on-covid-19-final-report.pdf> (Accessed 9 April 2020).
14. Coronavirus Resource Center, 2020. COVID-19 Dashboard by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University (JHU). <https://coronavirus.jhu.edu/map.html> (Accessed 9 April 2020).
15. Aghabozorgi, S., A. S. Shirkhorshidi, and T. Y. Wah. 2015. Time-series clustering – a decade review. Information Systems 53, 16 – 38.
16. Ministerul Afacerilor Interne, 2020. Informare COVID -19. <https://www.mai.gov.ro/> (Accessed 10 April 2020).
17. tslearn package documentation. 2020. <https://buildmedia.readthedocs.org/media/pdf/tslearn/latest/tslearn.pdf>, Accessed on 13 April 2020.