

**ԿՈՐՈՆԱՎԻՐՈՒՍԱՅԻՆ ՀԻՎԱՆԴՈՒԹՅԱՆ ՀԵՏԵՎԱՆՔՈՎ ՀԵՏԱԳԱ
ՄԱՀԵՐԻ ՔԱՆԱԿԻ ԿԱՆԽԱՏԵՍՈՒՄԸ ՄԵՔԵՆԱՅԱԿԱՆ ՈՒՍՈՒՑՄԱՆ
ՄԻՋՈՑՈՎ**

ՀՏԴ 004.8

DOI 10.56246/18294480-2022.12-11

ՄԱՐՏԻՐՈՍՅԱՆ ԱՐՄԱՆ

տեխնիկական գիտությունների թեկնածու,
«Թիմկյուեր» ընկերության ավագ ծրագրավորող
էլիոսյ՝ armanmartirosyan00@gmail.com

ԳԱԼՍՏՅԱՆ ԴԱՎԻԹ

ՀԱՊՀ «Սինոփսիս Արմենիա» ուսումնական դեպարտամենտի
«Ծրագրային ճարտարագիտություն» բաժնի
մագիստրատուրայի 2-րդ կուրսի ուսանող,
«Օփեն Սոֆթ Քոնսալթ» ընկերության ծրագրավորող
էլիոսյ՝ galstyandavid1998@gmail.com

ՀԱՐՈՒԹՅՈՒՆՅԱՆ ԷԴՈՒԱՐԴ

ՀԱՊՀ «Սինոփսիս Արմենիա» ուսումնական դեպարտամենտի
«Ծրագրային ճարտարագիտություն» բաժնի
մագիստրատուրայի 2-րդ կուրսի ուսանող,
«Կրիսպ» ընկերության ծրագրավորող
էլիոսյ՝ eduard.harutyunyan.1999@gmail.com

ՆԻԿՈՂՈՍՅԱՆ ԿԱՐԵՆ

Հայ-ռուսական համալսարանի «Համակարգչային ծրագրավորում»
բաժնի մագիստրատուրայի 2րդ կուրսի ուսանող,
«Սցիլա» ընկերության ծրագրավորող
էլիոսյ՝ karen.nikoghosyan.98@gmail.com

**Համաճարակի ժամանակ հիվանդների մահացության ռիսկերի վաղ
կանխարեսումը կարող է նվազեցնել մահացությունը՝ ապահովելով ռե-
սուրսների բաշխման և բուժման արդյունավետ պլանավորում: Այս հոդ-
վածի նպատակն է կորոնավիրուսային հիվանդության հետևանքով հե-
տագա մահվան դեպքերի քանակի կանխարեսման համար մշակել մեքե-
նայական ուսուցման K-մոդակա հարևանների մոդել՝ օգտագործելով**

համաճարակային իրավիճակի վերաբերյալ գոյություն ունեցող տվյալների հավաքածուն:

Ուսումնասիրվել է կորոնավիրուսային հիվանդության հետևանքով մահվան դեպքերի վրա ազդող հիմնական գործոնները. միջին տարիքը, սրտանոթային հիվանդությամբ պայմանավորված մահացության մակարդակը, շաքարախտի տարածվածությունը և այլն: Վերոնշյալ տվյալների հավաքածուն նորմալացվել է՝ առանձնացված հատկանիշները միևնույն չափողականության բերելու և ավելի ճշգրիտ արդյունքներ ստանալու համար: Կառուցված մոդելի նկատմամբ կատարված հետազոտությունների արդյունքում ստացվել է մինիմալ միջին քառակուսային սխալանքի արժեքը՝ 0.056 ($K = 3$ դեպքում), որի համար մոդելի ճշգրտության աստիճանը կազմում է 94%: Ինչպես նաև կառուցված մոդելի միջոցով վերլուծվել են առանձնացված հատկանիշները և ստացվել է դրանց ազդեցության ցուցանիշը, ըստ որի կորոնավիրուսային հիվանդության հետևանքով մահացության դեպքերի աճը հիմնականում պայմանավորված է շաքարային դիաբետի տարածվածությամբ և նոր հաստատված դեպքերի քանակով:

Այս հոդվածի սահմաններում կատարված վերլուծությունների արդյունքները կարող են որոշակի ներդրում ունենալ կորոնավիրուսային հիվանդության դեմ աշխարհի պայքարում:

Բանալի բառեր՝ կորոնավիրուս, արհեստական բանականություն, մեքենայական ուսուցում, K -մոդակա հարևան, հիպերչափական կարգավորում, տվյալների նորմալացում, էվկլիդեսյան հեռավորություն:

Ամբողջ աշխարհում կորոնավիրուսային հիվանդության հետևանքով առաջացած համաճարակը հսկայական վնաս է հասցրել տնտեսական և առողջապահական ոլորտներին՝ 2022 թվականի փետրվարի 1-ի դրությամբ թողնելով մոտ 6 միլիոն մահվան դեպք [1]: Առանց արդյունավետ բուժման և վիրուսային նոր շտամների առաջացման հնարավորության օրական միջին 7000 մահացության գլոբալ մակարդակը կարող է հանգեցնել մեկ տարվա ընթացքում մոտ 2.5 միլիոն մարդու մահվան:

Տեխնոլոգիական զարգացման դարաշրջանում արհեստական բանականությունն ու տվյալագիտությունը կարևոր դեր են խաղում առողջապահության ոլորտում: Հիվանդությունների ախտորոշման, կլինիկական կան-

խատեսումների բարելավման և առողջապահական մեծ տվյալների հավաքածուներում բարդ օրինաչափություններ հայտնաբերելու համար շատ գիտնականներ և առողջապահական կազմակերպություններ քրտնաջան աշխատում են ավտոմատացված համակարգեր ստեղծելու համար, որոնք կարող են օգտագործվել վարակի հետազա տեմպերը կանխատեսելու համար: Արհեստական բանականության գործիքներն օգտագործվում են կորոնավիրուսի դեմ պայքարելու համար տարբեր մասշտաբներով՝ համաճարակաբանական մոդելավորումից [2,3,4] մինչև անհատական ախտորոշում [5,6]:

Նյութեր և մեթոդներ

K-մոտակա հարևանների (KNN) մոդել

K-մոտակա հարևանների մոդելը լայնորեն օգտագործվում է դասակարգման և հետընթացի խնդիրների լուծման մեջ, ինչպես նաև ոչ ամբողջական արժեքների հաշվարկման և տվյալների հավաքածուների նմուշառման համար: K-մոտակա հարևանների մոդելի կառուցման ալգորիթմի մուտքին տրվում է ուսուցման տվյալների հավաքածուն, իսկ ելքին ստացվում են կանխատեսման կամ դասակարգման համար ընտրված տվյալները: Այս ալգորիթմի աշխատանքի ընթացքում որոշվում են տվյալ կետին ամենամոտ K հարևան կետերը, և այդ կետերի տվյալների հիման վրա կատարվում են հետազա կանխատեսումները: Կետերի մեջ չափվում է Էվկլիդեսյան հեռավորությունը՝ որոշելու, թե որքան մոտ է ուսուցման հավաքածուի յուրաքանչյուր անդամ կանխատեսման կամ դասակարգման համար ընտրված տվյալներին [8]:

$$d(x, y) = \sqrt{\sum_{i=1}^n (y_i - x_i)^2}$$

Քանաձև 1. Տրված x և y կետերի միջև Էվկլիդեսյան հեռավորությունը հաշվելու բանաձև:

Երբ ուսուցման տվյալների հավաքածուն պարունակում է տարբեր չափողականություն ունեցող բաղադրիչներ, անհրաժեշտ է նախքան K-մոտակա հարևանների մոդելին փոխանցելը կատարել նորմալացում: Տվյալների նորմալացումը կատարվում է 2-րդ բանաձևի միջոցով, որի արդյունքում բոլոր դաշտերը բերվում են միևնույն չափման սանդղակի՝ բարձրացնելով վերջնական արդյունքների ճշգրտության աստիճանը[9]:

$$X_i = \frac{X - X_{min}}{X_{max} - X_{min}}$$

Քանաձև 2. Տվյալների նորմալացման քանաձև:

որտեղ x -ը՝ տվյալ արժեք, X_i -ը՝ նորմալացված արժեք, X_{min} -ը՝ նվազագույն արժեք, իսկ X_{max} -ը՝ առավելագույն արժեք:

Ինչպես նաև վերջնական արդյունքների ճշգրտության աստիճանը բարձրացնելու նպատակով օգտագործվում է հիպերչափական կարգավորման **K-ական խաչաձև ստուգման** մեթոդը [11]:

Առաջարկվող մեթոդ

Կորոնավիրուսի հետևանքով մահվան հետագա ելքերի քանակը կանխատեսելու նպատակով օգտագործվել է մեքենայական ուսուցման K-մոտակա հարևանների մոդելը: Մոդելի վարժեցման համար վերցվել է տվյալների հավաքածուի հեղինակների կողմից օրեցօր թարմացվող աղբյուրից [7], որը օգտագործվում է բազում օտարերկրյա լրագրողների, քաղաքականական գործիչների և հետազոտողների կողմից (օրինակ՝ “Գուգլ” ընկերության կորոնավիրուսի տեմպերը նկարագրելու համար օգտագործվում է այդ աղբյուրը): Հետևյալ աշխատանքի համար վերոնշյալ տվյալների հավաքածուից առանձնացվել են այն բնութագրիչները, որոնք սերտորեն կապված են վարակի աճման արագության հետ: Ստորև ներկայացված են այդ բնութագրիչները՝

- ընդհանուր դեպքեր,
- նոր դեպքեր,
- նոր մահեր,
- խստության ինդեքս՝ կառավարության խստության ինդեքս,
- բնակչություն,
- միջին տարիք,
- սրտանոթային հիվանդությամբ պայմանավորված մահացության մակարդակ,
- շաքարախտի տարածվածություն,
- ձեռքի լվացման հարմարություններ:

Առանձնացված տվյալները նույն չափողականության բերելու և K-մոտակա հարևանների մոդելի ուսուցման համար ավելի հարմար դարձնելու նպա-

տակով կատարվել են հետևյալ ձևափոխությունները՝

- 1) տեսակավորվել են ըստ ժամանակացույցի,
- 2) խմբավորվել են ըստ օրերի,
- 3) նորմալիզացվել են 1-ին բանաձևի միջոցով,

Ստորև ներկայացված է տվյալների հավաքածուից հատված նորմալաց-
նելուց առաջ և հետո(աղյուսակ 1., աղյուսակ 2.):

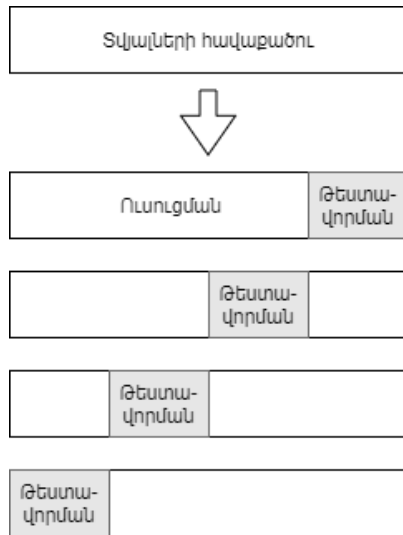
Աղյուսակ 1. Ոչ նորմալացված տվյալների հավաքածու:

date	new_deaths	cardiovasc_death_rate	people_fully_vaccinated	total_cases	median_age	new_cases	population	stringency_index	diabetes_prevalence	handwashing_facilities
9/1/2021	222.079787	262.5598207	908.2941176	286295364	30.243956	12107.99	120770850	57.98189655	7.927765957	95
9/2/2021	277.624339	262.5927166	1312911.918	341650976	30.2833333	6858.81	112313403	56.56719101	8.101502591	95
9/3/2021	183.164021	262.5927166	3477095.333	376437333	30.2833333	6802.9	112313403	55.90151685	8.101502591	95
9/4/2021	240.989529	262.5927166	6001079.224	432344612	30.4021277	11943.35	109252226	58.09268156	8.218877551	95
9/5/2021	167.217617	262.406	12194903	506660207	30.3592593	9832.295	108758562	56.6752514	8.351363636	95

Աղյուսակ 2. Նորմալացված տվյալների հավաքածու:

date	new_deaths	cardiovasc_death_rate	people_fully_vaccinated	total_cases	median_age	new_cases	population	stringency_index	diabetes_prevalence	handwashing_facilities
9/1/2021	0.72020933	0.983564919	6.86E-06	0.0018661	0.00422534	0.8307627	0.0126345	0.68699215	0.3140836	0.001755792
9/2/2021	0.90034146	0.983841275	0.009910443	0.0022269	0.00842856	0.4706019	0.0054565	0.678456701	0.467605348	0.001755792
9/3/2021	0.5940047	0.983841275	0.026246663	0.0024536	0.00842856	0.4667658	0.0054565	0.669735049	0.467605348	0.001755792
9/4/2021	0.78153402	0.983841275	0.045298816	0.002818	0.02110894	0.8194663	0.0028583	0.698443654	0.571323306	0.001755792
9/5/2021	0.54229018	0.982271686	0.092052554	0.0033024	0.01653307	0.6746209	0.0024393	0.679872507	0.688394155	0.001755792

Նորմալացված տվյալների հավաքածուն բաժանվել է ուսուցման և թես-
տավորման մասերի՝ 80/20 տոկոսային հարաբերությամբ, ինչպես նաև կա-
ռուցվող մոդելի ճշգրտությունն ու հուսալիությունը գնահատելու համար կա-
տարվել է **հիպերչափական կարգավորում**՝ K-ական խաչաձև ստուգման
մեթոդով (նկար 3):

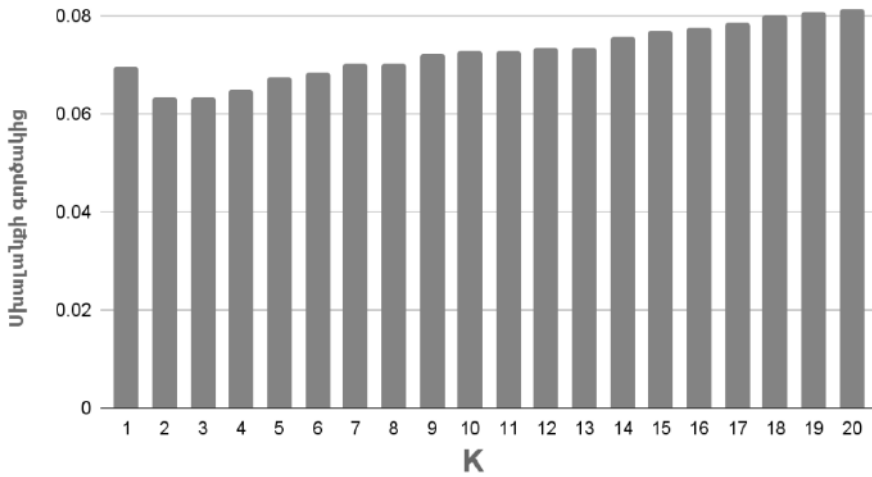


Նկար 3. Հիպերչական կարգավորում K-ական խաչաձև ստուգման մեթոդով՝ $K=5$ դեպքում:

Հետազոտական աշխատանքի ժամանակ տարբեր K արժեքների համար կառուցվել է **K-մոտակա հարևանների մոդելը**, որին փոխանցվել է ստացված ուսուցման տվյալների հավաքածուն: Ընտրվել է մինիմալ միջին քառակուսային սխալանք [10] ունեցող K արժեքը և այդ մոդելի համար կատարվել կանխատեսում:

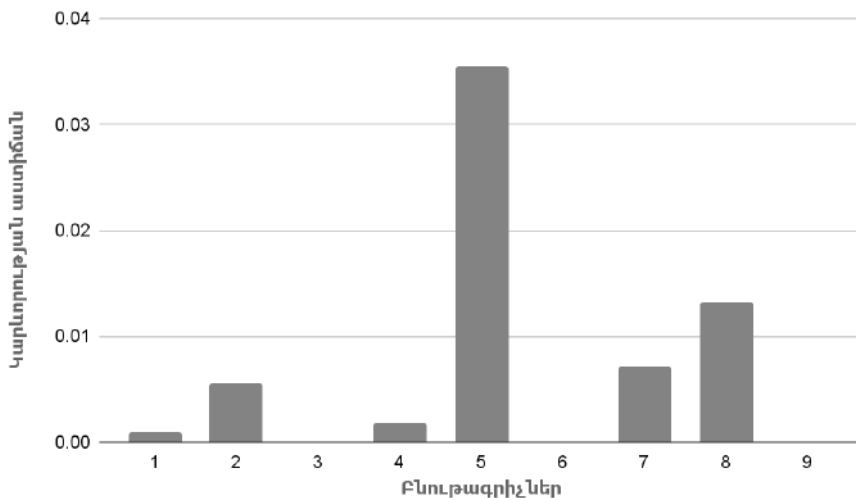
Արդյունքներ և քննարկում

Կատարված հետազոտությունները ցույց տվեցին, որ ընտրված տվյալների հավաքածուի համար **K-մոտակա հարևանների մոդելի** մինիմալ միջին քառակուսային սխալանքի արժեքը ստացվում է 0.056 ($K = 3$ դեպքում) (գրաֆիկ 1.):



Գրաֆիկ 1. Տարբեր K արժեքների համար կառուցված K-մոտակա հարևանների մոդելի միջին քառակուսային սխալանքի արժեքի կախումը մոտակա հարևանների քանակից:

Կառուցած մոդելի վրա հետազոտությունները ցույց տվեցին, որ առանձնացված բնութագրիչները հետևյալ հարաբերակցությամբ են ազդում վերջնական արդյունքի՝ նոր մահվան դեպքերի քանակի կանխատեսման ճշգրտության վրա (գրաֆիկ 2.):



Գրաֆիկ 2. Առանձնացված հատկանիշների կարևորության ցուցանիշը՝ նոր մահերի կանխատեսման համար:

Գրաֆիկ 2-ում ինդեքսները տրված են հետևյալ հաջորդականությամբ՝ սրտանոթային մահացության մակարդակը, ընդհանուր դեպքեր, միջին տարիքը, նոր դեպքեր, բնակչություն, կառավարության խստության ինդեքս, ձեռքի լվացման հարմարություններ:

Մինիմալ միջին քառակուսային սխալանք ունեցող K արժեքի համար մոդելի ճշգրտության աստիճանը կազմում է 94%:

Եզրակացություն

Հաշվի առնելով համաճարակային իրավիճակի վերաբերյալ առկա տվյալները՝ կառուցված մոդելի վրա ուսումնասիրվել է կորոնավիրուսի տարածման վրա ազդող հիմնական գործոնները, ինչպիսիք են **միջին տարիքը, սրտանոթային հիվանդությամբ պայմանավորված մահացության մակարդակը, շաքարախտի տարածվածությունը**: Արդյունքում ստացվել է, որ կորոնավիրուսային հիվանդության հետևանքով մահացության դեպքերի աճը հիմնականում պայմանավորված է շաքարային դիաբետի տարածվածությունից, ինչպես նաև նոր հաստատված դեպքերի քանակից: Օգտագործելով **K-մոտակա հարևանների մոդելը**՝ ստացվել է մինիմալ միջին քառակուսային սխալանքի արժեքը՝ 0.056, $K = 3$ դեպքում, որի համար մոդելի ճշգրտության աստիճանը կազմում է 94%:

Այս հետազոտության արդյունքները կարող են որոշակի ներդրում ունենալ կորոնավիրուսային հիվանդության դեմ աշխարհի պայքարում: Հետազոտության արդյունքները կօգտագործվեն հաջորդիվ կատարվելիք ուսումնասիրություններում, որտեղ պլանավորվում է ավելի հստակ անդրադառնալ պատվաստանյութերի դերակատարությանը համաճարակային իրադրության վրա:

Օգտագործված գրականության ցանկ

1. <https://www.worldometers.info/coronavirus/>, Accessed: 03.10.2022
2. Raju Vaishya, Mohd Javaid, Ibrahim Haleem Khan and Abid Haleem Artificial Intelligence (AI) applications for COVID-19 pandemic, Diabetes Metab Syndr. 2020 (pp. 337–339).
3. Bird J. J., Barnes C. M., Premebida C., Ekart A. & Faria D. R. Country-level pandemic risk and preparedness classification based on covid-19 data: A machine learning approach. 2020 (pp. 1-8)

- 
4. Wang P., Zheng X., Li J. & Zhu B. Prediction of epidemic trends in covid-19 with logistic model and machine learning technics. *Chaos, Solitons & Fractals* 139, 2020
 5. Mei X. et al. Artificial intelligence-enabled rapid diagnosis of patients with covid-19. *Nature Medicine* 2020. (pp. 1 - 5)
 6. Jin C. et al. Development and evaluation of an artificial intelligence system for covid-19 diagnosis. *Nature communications* 11, 2020. (pp. 1 - 14)
 7. <https://github.com/owid/covid-19-data/tree/master/public/data>, Accessed: 03.15.2022
 8. Spencer, Neil H. "5.4.5 Squared Euclidean Distances", *Essentials of Multivariate Data Analysis*, CRC Press, 2013, (pp. 95).
 9. <https://medium.com/analytics-vidhya/why-is-scaling-required-in-knn-and-k-means-8129e4d88ed7>, Accessed: 03.16.2022
 10. <https://c3.ai/glossary/data-science/root-mean-square-error-rmse/> Accessed: 03.18.2022
 11. https://scikit-learn.org/stable/modules/cross_validation.html Accessed: 03.10.2022

PREDICTION OF FURTHER DEATH FROM CORONAVIRUS DISEASE THROUGH MACHINE LEARNING

MARTIROSYAN ARMAN

*PhD in Technical Sciences,
Senior Software Engineer at "TeamViewer" Company
e-mail: armanmartirosyan00@gmail.com*

GALSTYAN DAVID

2nd year Master Student at the Educational Department "Synopsis Armenia" of
NPUA, the Faculty of "Software Engineering",
Software Developer at "Open Soft Consult" Company
e-mail: galstyandavid1998@gmail.com

HARUTYUNYAN EDUARD

2nd year Master Student at the Educational Department "Synopsis Armenia" of
NPUA, the Faculty of "Software Engineering",
Software Developer at "Krisp" Company
e-mail: eduard.harutyunyan.1999@gmail.com

NIKOGHOSYAN KAREN

2nd year Master Student at the Armenian-Russian University, the
Faculty of "System Programming",
Software Developer at "Scylla" Company
e-mail: karen.nikoghosyan.98@gmail.com

Early prediction of patient mortality risk during an epidemic can reduce mortality by ensuring effective resource allocation and treatment planning. The purpose of this article is to develop a machine learning model of k-nearest neighbors to predict the number of deaths as a result of coronavirus disease, taking into account existing data on the epidemiological situation.

The main factors influencing death from coronavirus disease were studied, such as average age, mortality from cardiovascular diseases, the prevalence of diabetes, etc. The above data set has been normalized to bring the chosen features to the same dimensionality and for more accurate results. As a result of the studies conducted with respect to the constructed model, the minimum average square error was obtained: 0.056 (in the case of $k = 3$), for which the degree of accuracy of the model is 94%. According to the built-in model, the chosen features were analyzed, and their impact index was obtained, according to which the increase in deaths due to coronavirus disease is mainly due to the prevalence of diabetes and the number of newly confirmed cases.

The results of the analysis carried out within the framework of this work may make a certain contribution to the fight against coronavirus in the world.

Keywords: *coronavirus, artificial intelligence, machine learning, k-nearest neighbors, hyperparameter optimization, data normalization, euclidean distance.*

ПРОГНОЗИРОВАНИЕ ЧИСЛА ПОСЛЕДУЮЩИХ СМЕРТЕЙ ОТ КОРОНАВИРУСНОЙ БОЛЕЗНИ ПОСРЕДСТВОМ МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ

МАРТИРОСЯН АРМАН

*Кандидат технических наук,
Старший программист компании "Тимвьюер"
электронная почта: armanmartirosyan00@gmail.com*

ГАЛСТЯН ДАВИД

Студент 2-го курса магистратуры учебного отдела "Синописис Армения"

*факультета “Программная инженерия” НПУА,
Программист компании “Опен Софт Консалт”
электронная почта: galstyandavid1998@gmail.com*

АРУТЮНЯН ЭДУАРД

Студент 2-го курса магистратуры учебного отдела

“Синописис Армения” НПУА

факультета “Программная инженерия”,

Программист компании “Крисп”

электронная почта: eduard.harutyunyan.1999@gmail.com

НИКОГОСЯН КАРЕН

Студент 2-го курса магистратуры кафедры “Системное программирование”

Армяно-Российского Университета,

программист компании “Scylla”

электронная почта: karen.nikoghosyan.98@gmail.com

Раннее прогнозирование риска смертности пациентов во время эпидемии может снизить смертность за счет эффективного распределения ресурсов и планирования лечения. Целью данной статьи является разработка модели машинного обучения k-ближайших соседей для прогнозирования количества смертей в результате коронавирусной болезни с учетом имеющихся данных об эпидемиологической ситуации.

Были изучены основные факторы, влияющие на смерть от коронавирусной болезни, такие как средний возраст, смертность от сердечно-сосудистых заболеваний, распространенность диабета и др. Приведенный выше набор данных был нормализован, чтобы привести выбранные признаки к одинаковой размерности и для получения более точных результатов. В результате исследований, проведенных по построенной модели, была получена минимальная средняя квадратичная ошибка 0,056. (в случае $k = 3$), для которых степень точности модели составляет 94%. При помощи встроеной модели были проанализированы выбранные признаки и получен их импакт-индекс, согласно которому рост смертности от коронавирусной болезни происходит в основном за счет распространенности сахарного диабета и количества впервые подтвержденных случаев.

Результаты анализов, проведенных в рамках этой работы, могут внести определенный вклад в борьбу с коронавирусом в мире.



Ключевые слова: *коронавирус, искусственный интеллект, машинное обучение, K-ближайших соседей, оптимизация гиперпараметров, нормализация данных, Евклидово расстояние*

Հոդվածը ներկայացվել է խմբագրական խորհուրդ 15.04.2022թ.:

Հոդվածը գրախոսվել է 16.04.2022թ.: