

4. Кружкова Г. В., Костюхин Ю. Ю. Теория игр и стратегия ценообразования на вторичное сырье. Цветные металлы, 2012, №8
5. Кузьмін О. Є., Жежуха В. Й., Городиська Н. А. Інструментарій теорії ігор у ціноутворенні на інжинірингові продукти. Проблеми економіки № 2, 2014
6. Хью Кортни. Игры для менеджеров The McKinsey Quarterly, 2000, № 3
7. Aust Gerhard Game-theoretic analysis of pricing, product quality, and retail service in a three-echelon supply chain // Journal of Applied Operational Research. 2015. Vol. 7, No. 1, 2–12.
8. Maklakov S. V. Modelirovanie biznes-processov s AllFusion PM. Moscow: Dialog-MIFI, 2008. 224 p.
9. Metody i modeli analiza dannyh: OLAP i Data Mining: ucheb. pos./ Barsegyan A. A., Kupriyanov M. S., Stepanenko V. V., Holod I. Sankt-Peterburg: BHV-Peterburg, 2004. 336 p.
10. Мушник Э., Мюллер П. Методы принятия технических решений. М.: Мир, 1990. 208 с.
11. Тоценко В.Г. Методы и системы поддержки принятия решений. Алгоритмический аспект. Киев: Наукова думка, 2002. 382 с.

Статтю подано до редакції 19.10.2020

УДК: 519.252

DOI 10.33111/mise.100.6

Камінський О. Є., д.е.н.,
доцент кафедри комп'ютерної математики та інформаційної безпеки,
Політ Д. Г.,
магістрант спеціальності «Системний Аналіз»,
ДВНЗ «КНЕУ імені Вадима Гетьмана»

Kaminsky O. E., Doctor of Economic Sciences,
Associate Professor of the Department of
Computer Mathematics and Information Security,
Poliyt D. G.,
Master Student specialty «System Analysis»,
SHEI KNEU named after V. Hetman

**АНАЛІЗ ДОСТОВІРНОСТІ ІНФОРМАЦІЇ
ШОДО ПАНДЕМІЇ COVID-19 В УКРАЇНІ
(НА ПРИКЛАДІ СВІТОВИХ АГРЕГАТОРІВ ДАНИХ)**

**ANALYSIS OF THE RELIABILITY OF INFORMATION
ABOUT THE COVID-19 PANDEMIC IN UKRAINE
(ON THE EXAMPLE OF WORLD DATA AGGREGATORS)**

Анотація. Вплив пандемії COVID-19 на економічну діяльність в основному зумовлений зменшенням витрат для осіб з вищими доходами через проблеми зі здоров'ям, що, в свою чергу, впливає на підприємства, які обслуговують багатих, і в кінцевому рахунку зменшує рівень доходів і витрат низкооплачуваних працівників цих підприємств. Модель прогнозування підтверджених випадків захворювання на COVID-19 усе ще перебуває на стадії досліджень. Розуміючи тенденцію розвитку підтверджених випадків у регіоні, уряди країн намагаються контролювати пандемію, використовуючи відповідну політику. Однак загальноприйняті традиційні математичні методи та моделі прогнозування пандемій мають обмеження і не завжди дають точні оцінки. Слід зауважити, що аналіз достовірності даних щодо смертності від пандемії є досить нетривіальною задачею і складно впевнено визначити саму кількість жертв. Проте вірогідність цих показників може бути компенсовано, якщо проводити аналіз, порівнюючи дані «схожих» країн. Метою статті є представлення результатів аналізу вірогідності інформації щодо епідемії Covid-19 на основі інформації світових онлайн-агрегаторів даних, його зв'язку з географічним положенням країн світу та кількістю їх населення. У даній роботі авторами запропоновано вдосконалений метод такого аналізу та відповідне програмне забезпечення. Результати аналізу показали, що найімовірніша кількість смертей від COVID-19 в Україні має бути на 25 % вище заявлених даних. Цей метод має кращий ефект пристосування для регіонів з великою базою населення, а ефект прогнозування є точнішим. Аналіз загальної смертності під час пандемії й зіставлення цих даних з середньою смертністю в інших країнах може служити точнішою, нехай і децю попередньою, оцінкою реальної кількості жертв пандемії, та її впливу на економіку України.

Ключові слова: аналітика; пандемія; COVID-19; прогнозний аналіз; соціо-безпека

Abstract. The impact of the COVID-19 pandemic on economic activity is mainly due to lower costs for higher-income people due to health problems, which in turn affects businesses that serve the rich and ultimately reduces the income and expenditure of low-paid workers. these enterprises. The model for predicting confirmed cases of COVID-19 is still under investigation. Recognizing the trend in confirmed cases in the region, governments are trying to control the pandemic using appropriate policies. However, generally accepted traditional mathematical methods and models for predicting pandemics are limited and do not always give accurate estimates. It should be noted that the analysis of the reliability of pandemic mortality data is a rather non-trivial task and it is difficult to determine the number of victims with confidence. However, the probability of these indicators can be offset by analyzing by comparing data from «similar» countries. The aim of the article is to present the results of the analysis of the reliability of information on the David-19 epidemic based on information from global online data aggregators, its relationship with the geographical location of countries and their population. In this paper, the authors propose an improved method of such analysis and appropriate software. The results of the analysis showed that the most probable number of deaths from COVID-19 in Ukraine should be 25% higher than the stated data. This method has a better adaptation effect for regions with a large population base, and the forecasting effect is more accurate. Analysis of overall mortality during a pandemic and comparison of these data with average mortality in other countries can serve as a more accurate, albeit somewhat preliminary, estimate of the actual number of pandemic victims and its impact on Ukraine's economy.

Keywords: analytics; pandemic; COVID-19; forecast analysis; socio-security

Вступ. Для розробки ефективних заходів, що дозволять максимально зменшити вплив пандемії COVID-19 на безпеку громадян України, необхідно проводити глибокий аналіз і моделювання впливу карантинних заходів. Кожного дня кількість заражених COVID-19 збільшується. Станом на 17 листопада 2020 року в Україні було підтверджено 557 657 випадків захворювання на COVID-19, у тому числі 9 856 смертей і 250 983 повністю відновлених [1]. Через пандемію COVID-19 Україна може зіткнутися з найгіршою ре-цесією за останні десятиліття, в результаті чого понад 9 мільйонів людей опиняться у злиднях. UN ОСНА в Україні повідомляє, що з початку пандемії понад 80 % домогосподарств втратили дохід, а у більш 40 % сімей що найменше один член сім'ї втратив роботу [2]. Хоча немає даних про Східну Україну, соціально-економічні наслідки пандемії в регіоні, що постраждав від конфлікту, будуть мати руйнівний характер. Саме тому аналіз та прогнозування смертності внаслідок пандемії COVID-19 є актуальною та важливою передумовою для формулювання стратегій профілактики і боротьби з інфекційними захворюваннями, а також прийняття рішень щодо подолання їх наслідків у сфері економіки.

Аналіз останніх досліджень і публікацій.

Вплив пандемії COVID-19 на економічну діяльність в основному зумовлений зменшенням витрат для осіб з вищими доходами через проблеми зі здоров'ям, що, своєю чергою, впливає на підприємства, які обслуговують багатих, і зрештою зменшують рівень доходів і витрат низькооплачуваних працівників цих підприємств. Моделі прогнозування підтверджених випадків захворювання на COVID-19 усе ще перебувають на стадії досліджень. Розуміючи тенденцію розвитку підтверджених випадків у регіоні, уряди країн намагаються контролювати пандемію, використовуючи відповідну політику. Однак загальноприйняті традиційні математичні методи та моделі прогнозування пандемій мають обмеження і не завжди дають точні оцінки.

Наприклад, у дослідженні [1] було використано статистичний метод для оцінки CFR на ранній фазі спалаху пандемії COVID-19. Дані щодо підтверджених лабораторією випадків COVID-19 та смертей автори збирали з 10 січня по 3 лютого 2020 року та розділили їх на три групи: місто Вухань, інші міста провінції Хубей та інші провінції материкового Китаю. Проста модель лінійної регресії була застосована для оцінки CFR з кожного кластера.

У своїй роботі вчені Mizumoto K, Chowell G. [2] оцінювали ризик смертності від COVID-19 у місті Ухані, а також інших частинах Ки-

таю, крім провінції Уханя, з урахуванням затримки часу, щоб визначити ступінь тяжкості епідемії в країні. Оцінки ризику смерті в Ухані досягли значень до 12% в епіцентрі епідемії та $\approx 1\%$ в інших, більш постраждалих районах. Підвищені оцінки ризику смерті, визначено, як пов'язані з руйнуванням системи охорони здоров'я, та підкреслено дієву роль посилених заходів з охорони здоров'я, включаючи соціальне дистанціювання та обмеження руху.

У роботі Hoang P. [3] було обговорено чітку модельну функцію, яка дозволяє оцінити загальну кількість смертей серед населення, і зокрема, оцінити сукупну кількість смертей у Сполучених Штатах через поточний вірус Covid-19. Автором було проведено порівняльний аналіз результатів моделювання з двома існуючими моделями на основі нових критеріїв. Результати дослідження показали, що запропонована модель є точнішою, ніж існуючі моделі, які були засновані на американських даних про смертність від Covid-19.

У роботі Yan B., Wang J. [4] для розв'язання проблеми прогнозування смертності від пандемії запропоновано вдосконалений метод прогнозування підтверджених випадків смерті на основі нейронної мережі LSTM (Long-Short Term Memory). У роботі порівнюються відхилення між експериментальними результатами вдосконаленої моделі прогнозування LSTM і моделями прогнозування (такими як логістичні рівняння та рівняння Хілла) з реальними даними в якості еталонних. Експерименти показали, що запропонований підхід має менші відхилення від прогнозу та кращий ефект пристосування.

Постановка проблеми. Мета статті полягає в тому, щоб представити результати аналізу вірогідності інформації щодо епідемії коронавірусу (COVID-19) на основі інформації світових онлайн-агрегаторів даних, його зв'язку з географічним положенням країн світу та кількістю їх населення. У даній роботі автори пропонують вдосконалений метод такого аналізу.

Виклад основного матеріалу. За дослідженнями відомого вченого С. Кузнеця [2], рішення щодо макроекономічної політики приймаються на основі загальнодоступних статистичних даних, побудованих на основі офіційних звітів, які надають уряди. Така статистика має велике значення для розуміння загальної економічної діяльності держави, але вона має два обмеження.

По-перше, офіційні дані, як правило, можуть не бути точними, через політичні мотиви, відносно невеликі обсяги вибірки, та неповне врахування критеріїв аналізу. По-друге, така статистика, як правило, доступна лише зі значним часовим запізненням. Через

дані обмеження чинних загальнодоступних макроекономічних статистичних даних недостатньо для вивчення джерел економічних коливань і причинних наслідків глобальних подій.

Аналіз медичних даних показує, що певна кількість смертей від вірусу Covid-19 не реєструється державою, через очевидне бажання не збільшувати панічні настрої серед населення та через можливі фальсифікації збоку лікарів для отримання дозволу на поховання померлих. Слід зауважити, що аналіз достовірності даних щодо смертності від пандемії є досить нетривіальною задачею і складно впевнено визначити саму кількість жертв. Проте вірогідність цих показників може бути компенсовано, якщо проводити аналіз, порівнюючи дані «схожих» країн.

Ми можемо визначити країни схожими, якщо вони: по перше, географічно знаходяться поряд (тобто мають спільний кордон). Така умова дасть змогу ще до проведення аналізу вилучити ряд чинників, таких як: історія хвороби (яка має бути подібною у двох сусідніх країнах), клімат, туристичні шляхи, менталітет (наприклад, відношення населення до пандемії). Дана вимога реалізується нами на рівні скрипту, де залежно від обраної країни беруться дані тих країн, що знаходяться поруч.

Для отримання кількості населення кожної з країн, кількості заражених і померлих, а також медичні показники, були обрані такі інформаційні онлайн-ресурси, які наведено в табл. 1.

Таблиця 1

ОНЛАЙН — АГРЕГАТОРИ ДАНИХ ЩОДО ПАНДЕМІЇ COVID-19

Агрегатор даних	Адреса
World Health Organization (WHO)	https://www.who.int/
European Centre for Disease Prevention and Control (ECDC)	https://www.ecdc.europa.eu/en/geographical-distribution-2019-ncov-cases
DXY.cn. Pneumonia. 2020.	http://3g.dxy.cn/newh5/view/pneumonia
US CDC	https://www.cdc.gov/coronavirus/2019-ncov/index.html
BNO News	https://bnonews.com/index.php/2020/02/the-latest-coronavirus-cases/
WorldoMeters	https://www.worldometers.info/coronavirus/
lPoint3Arces	https://coronavirus.lpoint3acres.com/en
COVID Tracking Project: (US testing and Hospitali-zation Data.)	https://covidtracking.com/data .
Los Angeles Times	https://www.latimes.com/projects/california-coronavirus-cases-tracking-outbreak/
The Mercury News	https://www.mercurynews.com/tag/coronavirus/

По друге, не мають між собою великої різниці у кількості населення. Дана умова є другорядною, оскільки найголовніші коефіцієнти, які неможливо підрахувати вручну, ми скорочуємо згідно з першою умовою, але дана умова враховує, що хвороба може розповсюджуватися швидше через більшу кількість населення. Після отримання списку «схожих» країн, розраховується коефіцієнт за формулою (1), щоб врахувати той факт, що сусідня країна може мати значно кращу медичну систему.

$$K = \frac{C_1}{C_2 * M}, \quad (1)$$

де C_1 — кількість підтверджених смертей;

C_2 — кількість підтверджених заражень;

M — показники стану медичної системи країни.

Відповідно, ми можемо визначити такий алгоритм аналізу:

- 1) програмно визначаються країни-сусіди;
- 2) виконується парсинг csv-файлу з даними по населенню кожної з країн-сусідів, дані додаються до загального датасету;
- 3) з датасету виключаються країни, де кількість населення вище у 3,5 разу;
- 4) виконується парсинг csv-файлів з кількістю заражень та смертей з медичними показниками усіх країн і додається до загального датасету;
- 5) вираховується коефіцієнт відношення смертей до підтверджених заражень.

Медичні показники беруться з csv файлу.

Розрахунок здійснюється програмним модулем (прог. 1)

Прог. 1. Розрахунок медичних показників

```
medical_rating = pd.read_csv('data/medical_system.csv')
medical_rating = medical_rating.copy().drop(['Health
Distribution','Overall goal attainment', 'Health expenditure
per capita in international dollars'],axis =1)
medical_rating.index = medical_rating['Country']
medical_rating = medical_rating.drop(['Country'], axis=1)
close_countries_data = pd.merge(close_countries_data,
medical_rating, left_index=True, right_index=True)
```

Отримання даних про кількість населення здійснюється аналогічним чином, парсяться й форматуються дані з іншого csv файлу, як представлено в прог.2.

Прог. 2. Отримання даних про кількість населення

```
population = pd.read_csv(«data/population_by_country.csv»)
population.copy().drop(['Yearly Change', 'Net Change',
'Density (P/Km²)', 'Land Area (Km²)', 'Migrants (net)',
'Fert. Rate', 'Med. Age', 'Urban Pop %', 'World
Share'],axis=1)
close_countries_data = pd.merge(close_countries_data,
population, left_index=True, right_index=True)
```

Після розрахунку кожного з коефіцієнтів обираються максимальне, мінімальне та середнє значення з країн з достатньо високим рівнем медицини, оскільки, чим вище рівень медицини, тим вище вірогідність даних. За допомогою 3-ох коефіцієнтів ми отримаємо значення прогнозованої мінімальної кількості померлих, прогнозованої (найімовірнішої) кількості померлих і прогнозованої максимальної кількості померлих. Для знаходження коефіцієнта, слід знайти відношення кількості зареєстрованих смертей до заражень і загального перформансу медичної системи країни, як це відображено в прог.3.

Прог.3. Отримання даних про кількість населення

```
close_countries_data[«coef»] =
close_countries_data[«Deaths»] / (
close_countries_data[«Confirmed»] *
close_countries_data[«Overall health system performance»])

top_2_countries_by_health_level =
close_countries_data.sort_values(«Health Level»).head(2)

average_coef_by_most_healthy_counties =
sum(top_2_countries_by_health_level[«coef»]) /
len(top_2_countries_by_health_level[«coef»])

max_coef = max(close_countries_data[«coef»])
min_coef = min(close_countries_data[«coef»])

input_country_data[«Most likely deaths»] =
average_coef_by_most_healthy_counties *
input_country_data[«Confirmed»] *
input_country_data[«Overall health system performance»]
input_country_data[«Min deaths (best-case scenario)»] =
min_coef * input_country_data[«Confirmed»] *
input_country_data[«Overall health system performance»]
input_country_data[«Max deaths (Worst-case scenario)»] =
max_coef * input_country_data[«Confirmed»] *
input_country_data[«Overall health system performance»]
```

Програмне забезпечення для аналізу було реалізовано на базі Jupiter Notebook та мови програмування Python [2].

Прог. 1. Приклад подання коду комп'ютерної програми
Аналіз українських показників проводився на основі даних Польщі та Угорщини, які представлено на рис.1.

	Confirmed	Deaths	Delta_Deaths	Health Level	On level of health	Overall health system performance	Population	coef
Poland	665547	9499	419.0	45	89	50	37854825	0.000285
Hungary	131887	2883	99.0	62	105	66	9665192	0.000331

Рис. 1. Результати аналізу аналізу даних
щодо смертності для країн-сусідів України

Дані щодо кількості зараження у країн-сусідів України наведено на рис. 2.

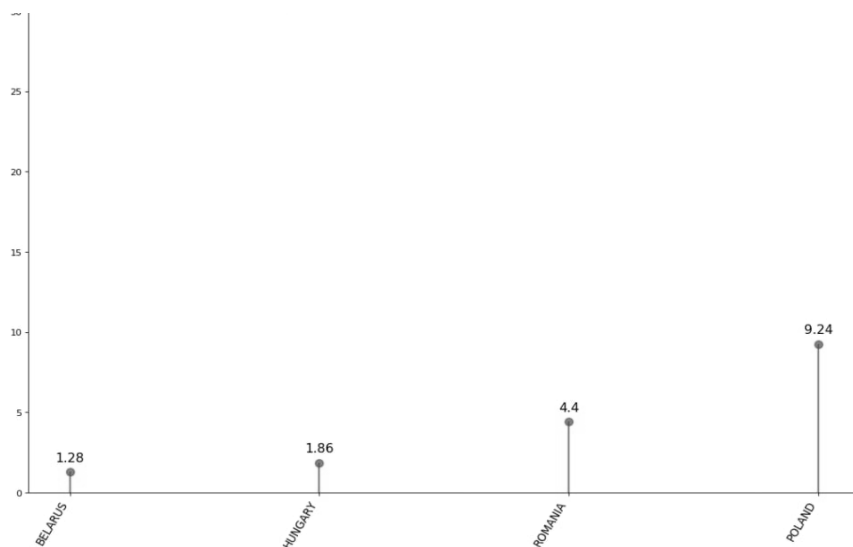


Рис. 2. Кількість заражень у країн — сусідів (x 100 000)

На основі отриманих даних було проведено аналіз ситуації в Україні, який дав такі результати, представлені на рис. 3.

	Confirmed	Deaths	Delta_Deaths	Health Level	On level of health	Overall health system performance	Most likely deaths	Min deaths (best-case scenario)	Max deaths (Worst-case scenario)	Falsification %	best-case scenario falsification %	worst-case scenario falsification %
Ukraine	527808	9604	182	70	101	79	12856	5359	13810	25	0	30

Рис. 3. Результати аналізу даних щодо смертності для України

Дані щодо кількості смертей у країн-сусідів України наведено на рис.4.

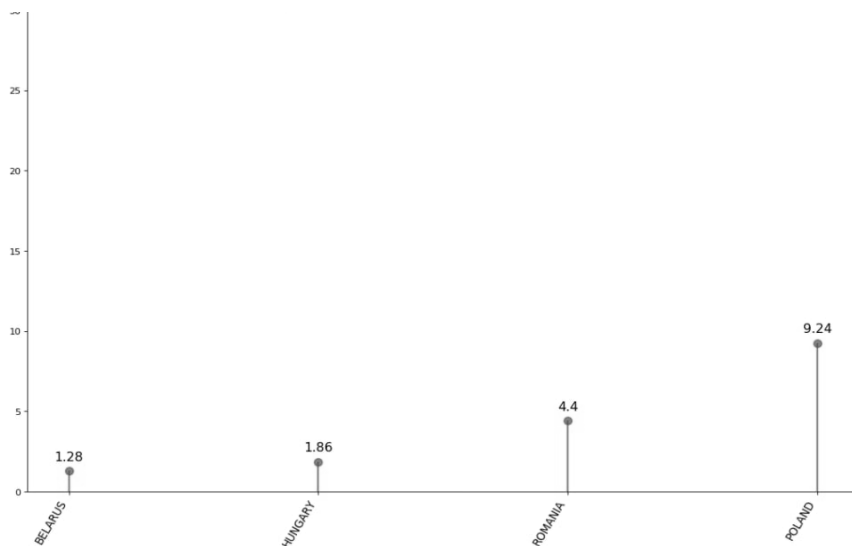


Рис.4. Кількість смертей у країн — сусідів (x 100 000)

Висновки. Результати аналізу свідчать, що найімовірніша кількість смертей від COVID-19 в Україні має бути на 25 % вище заявлених даних. Даний метод має вищу достовірність для регіонів з великою базою населення, а ефект від прогнозу відповідно, також стає вірогіднішим.

У даній роботі авторами пропонується розв’язувати такі проблеми, шляхом (1) створення публічної бази даних, яка вимірює показники смертності та зараження, стан медичної системи та інші показники на високочастотному, детальному рівні, використовуючи анонімі дані приватних компаній та світових онлайн-ресурсів,

та (2) таку нову базу даних можна використовувати для порівняльного аналізу наслідків пандемії коронавірусу (COVID-19) з іншими країнами, та використання даних аналізу для розробки рекомендацій щодо подальшої політики боротьби з наслідками пандемії майже в режимі реального часу, побудова такої бази даних є темою наших подальших досліджень.

Аналіз загальної смертності під час пандемії й зіставлення цих даних з показниками смертності в інших країнах може служити точнішою, нехай і дещо попередньою, оцінкою реальної кількості жертв пандемії, та її впливу на економіку України.

Бібліографічні посилання

1. Оперативна інформація про поширення коронавірусної інфекції 2019-nCoV. (19.11.2020) URL: <https://moz.gov.ua/article/news/operativna-informacija-pro-poshirennja-koronavirusnoi-infekcii-2019-ncov>.

2. Огляд гуманітарної ситуації — жовтень 2020 р. (19.11.2020) URL: https://www.humanitarianresponse.info/sites/www.humanitarianresponse.info/files/documents/files/ukraine_humanitarian_snapshot_20201110-ua.pdf

3. Kuznets, Simon. National Income and Its Composition, New York: National Bureau of Economic Research. 1941. <https://www.nber.org/books-and-chapters/national-income-and-its-composition-1919-1938-volume-i>

4. Yang, S., Cao, P., Du, P., Wu, Z., Zhuang, Z., Yang, L., Yu, X., Zhou, Q., Feng, X., Wang, X., Li, W., Liu, E., Chen, J., Chen, Y., & He, D. (2020). Early estimation of the case fatality rate of COVID-19 in mainland China: a data-driven analysis. *Annals of translational medicine*, 8(4), 128. <https://doi.org/10.21037/atm.2020.02.66>

5. Mizumoto K, Chowell G. Estimating Risk for Death from Coronavirus Disease, China, January-February 2020. *Emerg Infect Dis.* 2020 Jun;26(6): P.1251-1256. <https://10.3201/eid2606.200233>.

6. Hoang Pham. On Estimating the Number of Deaths Related to Covid-19. Rutgers University, Piscataway, *Mathematics* 2020, 8(5), 655; <https://doi.org/10.3390/math8050655>

7. B. Yan, J. Wang, Z. Zhang, X. Tang, Y. Zhou et al., «An improved method for the fitting and prediction of the number of covid-19 confirmed cases based on lstm,» *Computers, Materials & Continua*, vol. 64, no.3, pp. 1473–1490, 2020.

8. Jupyter Notebook. (19.11.2020) URL: <https://realpython.com/jupyter-notebook-introduction/>

Статтю подано до редакції 30.10.2020