УДК 004.6:616-036.22
DOI: 10.31866/2617-796x.3.2.2020.220590

Булига Костянтин,
кандидат технічних наук, доцент кафедри комп'ютерних наук,
Київський національний університет культури і мистецтв,
Київ, Україна
piton54@ukr.net
https://orcid.org/0000-0002-9537-3226

Булига Олена,
старший викладач кафедри інформаційно-аналітичної
діяльності та інформаційної безпеки,
Національний транспортний університет,
Київ, Україна
hellen.bulyga@gmail.com
https://orcid.org/0000-0003-1612-6735

ВИКОРИСТАННЯ ІНФОРМАЦІЙНИХ ТЕХНОЛОГІЙ ДЛЯ АНАЛІЗУ ЕПІДЕМІОЛОГІЧНОГО СТАНУ

Метою статті є пошук оптимальних методів побудови тренду з метою аналізу епідеміологічної ситуації в разі подальшого розвитку пандемії COVID-19. Докладно розглянуто завдання, які виникли зі зростом випадків інфікування восени 2020 року.

Методи дослідження: узагальнення методу найменших квадратів Гауса та методологія Facebook Prophet.

Новизною проведеного дослідження є порівняльний аналіз переваг і недоліків систем, що використовують для оперативної обробки статистичних даних, зокрема електронних таблиць MS Excel.

Висновки. Викладений у статті матеріал дає наочні приклади виконання статистичних розрахунків з використанням широко розповсюджених програмних продуктів. Розглянуто проблеми, що виникають у зв'язку з правдивістю статистичної інформації. Представлено приклади реальних розрахунків трендів епідеміологічної ситуації.

Ключові слова: пандемія COVID-19; інформаційна система; MS Excel; метод найменших квадратів; лінія тренду; статистичний аналіз.

© Булига К. Б.
© Булига О. А.

Вступ. Захворювання на COVID-19, яке на початку 2020 року сприймали як рідкісну й екзотичну хворобу, перетворилося на пандемію світового масштабу. Зокрема, на сайті РНБО України (https://covid19.rnbo.gov.ua/) зображено масштаби епідеміологічної ситуації у світі та нашій державі. Для оцінки перспективи росту кількості випадків захворювання використовують різноманітні методи статистичного аналізу (Аналіз поточної ситуації та моделювання сценаріїв поширення захворювання COVID-19, 2020; Прогноз розвитку епідемії COVID-19 в Україні на 23 листопада – 7 грудня 2020 року («Прогноз РГ-29»)). Наприклад, у роботі (Прогноз розвитку епідемії COVID-19 в Україні на 23 листопада – 7 грудня 2020 року («Прогноз РГ-29»)) використано найсучаснішу модель Facebook Prophet, яка демонструє високу ефективність щодо моделювання часових рядів, що містять аномальні дати, різні види сезонності та лінійну чи нелінійну динаміку впливу різних складових моделі (Быстрый старт прогнозирования временных рядов на практическом примере с использованием FB Prophet, 2019).

Метод використовує модель часових рядів з трьома основними компонентами моделі: тренд, сезонність і свята. Вони об'єднані в такому рівнянні:

$$y(t) = g(t) + s(t) + h(t) + \varepsilon t,$$

де:

$g(t)$ – кусково-лінійна або логістична крива зростання для моделювання неперіодичних змін у часових рядах;

$s(t)$ – періодичні зміни (наприклад, щотижнева/річна сезонність);

$h(t)$ – вплив свят (надається користувачем) з нерегулярними графіками;

εt – термін «похибка» враховує будь-які незвичайні зміни, що не враховуються моделлю.

Використовуючи час як регресор, Facebook Prophet намагається узгодити кілька лінійних і нелінійних функцій часу в ролі компонентів. Моделювання сезонності як адитивного компонента – це той же підхід, який використовують для експоненціального згладжування в техніці Холта-Вінтерса (Hyndman and Athanasopoulos, 2018). Facebook Prophet формулює проблему прогнозування не як залежність кожного спостереження в межах часового ряду, а як вправу з підгонки кривої. У дослідженні розроблено та застосовано алгоритм налаштування багатьох параметрів цієї моделі, який прогнозує дані на задану кількість днів вперед, але дані наявних спостережень за останні дні використовували для вибору найкращої моделі з налаштованих. Проведено дослідження для періоду прогнозування 14 днів. На жаль, прогнозовані результати не збігаються з уже відомими даними, які доступні на момент написання цієї роботи (таб. 1).

Вочевидь некерованість і непрогнозованість суспільно-політичної та економічної ситуації в Україні істотно впливають на результати статистичних досліджень. У такому разі доцільно використовувати класичні методи прогнозування, але на незначних часових відрізках. У цій роботі для побудови трендів використано класичний алгоритм методу найменших квадратів (МНК) (Hyndman and Athanasopoulos, 2018), реалізований у програмному продукті MS Excel (Булига, Булига та Волощук, 2018).

Результати дослідження. Розглянемо побудову лінії тренду щоденного зростання в Україні кількості захворювань на Covid-19 у листопаді 2020 року. Статистичні дані публікують на порталі (https://covid19.rnbo.gov.ua/). MS Excel дає змогу будувати тренди декількома способами: 1) за допомогою вбудованих функцій; 2) використовуючи апарат побудови діаграм; 3) використовуючи надбудову Solver, або російською мовою Поиск решения. На рис. 1 наведено діаграму побудови ліній тренду за даними з 1-го по 22-ге листопада трьох типів: лінійного, показникового та логарифмічного.

Таблиця 1

Дата	Нижня межа довірчого інтервалу, кількість випадків	Прогнозоване значення, кількість випадків	Верхня межа довірчого інтервалу, кількість випадків	Реальне значення, кількість випадків	Похибка, %
23.11.2020	10975	11229	11464	12287	9 %
24.11.2020	13566	13799	14033	13882	1 %
25.11.2020	14254	14501	14728	15331	5 %
26.11.2020	14585	14827	15047	16218	9 %
27.11.2020	15777	16008	16257	16294	2 %
28.11.2020	16478	16723	16967	12978	29 %
29.11.2020	13822	14073	14324	9946	41 %
30.11.2020	12263	12523	12775	12498	0 %
01.12.2020	15081	15346	15598	13141	17 %
02.12.2020	16220	16495	16779	14496	14 %
03.12.2020	16578	16879	17176	15131	12 %
04.12.2020	17465	17768	18055	13285	34 %
05.12.2020	18165	18483	18826	11590	59 %
06.12.2020	15602	15906	16230	8641	84 %

На основі цих трендів зроблено прогноз до 30-го листопада, порівняння результатів якого з новою статистикою наведено на рис. 2.

Обидві діаграми наочно демонструють, що показники суботи та неділі значно порушують загальну тенденцію зростання випадків захворювання. Відповідно до концепції Facebook Prophet можна відкинути статистику вихідних, тоді результати спостережень значно гладшають. Нові результати наведено на рис. 3.

З рисунка 3 видно, що лінійний тренд з достатньою точністю відображає тенденцію розвитку подій. Тому подальші дослідження виконувалися з лінійною функцією.

Числовий експеримент полягав у тому, що всім статистичним даним надано ваговий коефіцієнт 1, потім, змінюючи його, можна створити новий статистичний ряд з урахуванням можливих збурень у часі. Зокрема, враховуючи спадання кількості тестів на вихідні, змінювалися вагові коефіцієнти суботи і неділі, які

припадали на перші 22 дні листопада. Наявність статистики дала змогу оцінити можливості такого підходу. Результати порівняння наведено в таб. 2.

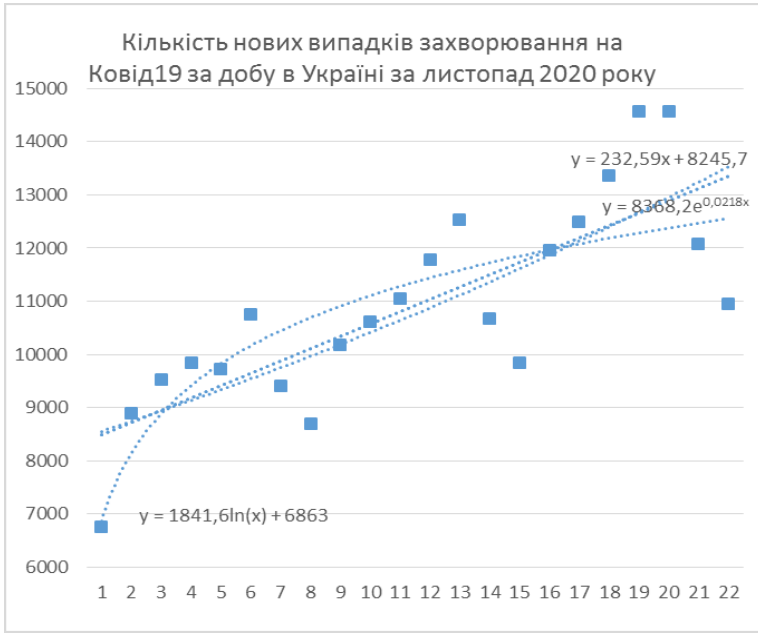

Рис. 1. Побудова ліній тренду на основі 22-х днів листопада

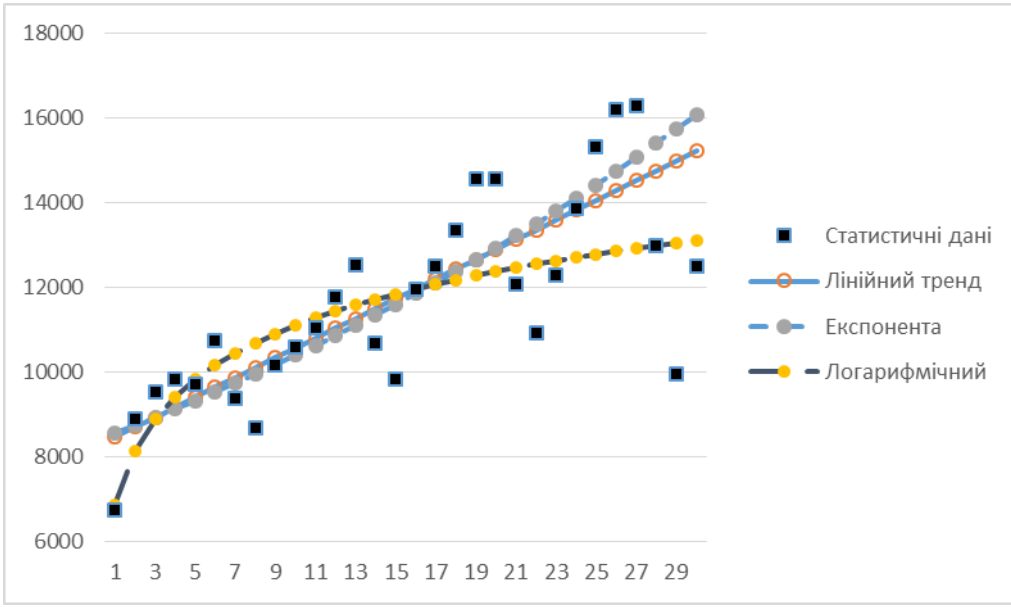

Рис. 2. Порівняння прогнозу з наявною статистикою

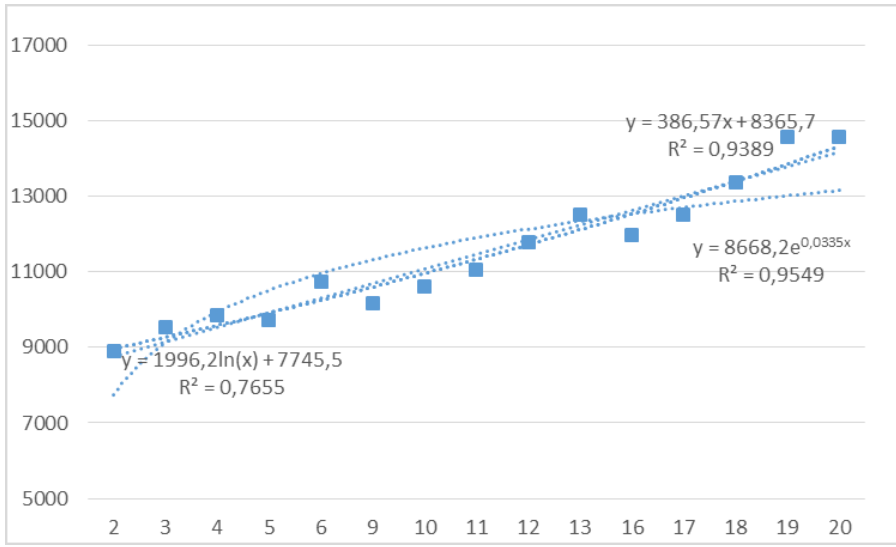

Рис. 3. Побудова ліній тренду на основі 22-х днів листопада (без вихідних)

Таблиця 2

Ваговий множник	Дата, листопад								Середня похибка
	23	24	25	26	27	28	29	30	
	Похибка								
1,00	11%	0%	8%	12%	11%	14%	51%	22%	16,04%
0,65	4%	13%	21%	24%	23%	2%	30%	4%	15,07%
0,70	2%	12%	19%	22%	21%	0%	33%	7%	14,39%
0,75	0%	10%	17%	20%	20%	2%	36%	9%	14,33%
0,80	2%	8%	15%	19%	18%	5%	39%	12%	14,67%
0,85	5%	6%	14%	17%	16%	7%	42%	14%	15,01%
0,90	7%	4%	12%	15%	14%	9%	45%	17%	15,35%
0,95	9%	2%	10%	14%	13%	11%	48%	19%	15,69%
1,05	13%	1%	7%	10%	9%	16%	54%	24%	16,75%

Найкращий результат з точки зору значення середньої похибки відповідає ваговому коефіцієнту 0,75. Зроблено прогноз на перші 6 днів грудня на основі запропонованої методики та проведено порівняння з наявними даними та результатами роботи (Прогноз розвитку епідемії COVID-19 в Україні на 23 листопада – 7 грудня 2020 року).

Висновки. Аналізуючи отримані дані, дійшли висновку, що аномальні результати вихідних днів листопада значно погіршують загальну статистику. З рис. 4 видно, що запропонований підхід дає найкращий рівень точності прогнозування, а з урахуванням трудомісткості та кількості виконавців має значну перевагу для оперативних розрахунків.

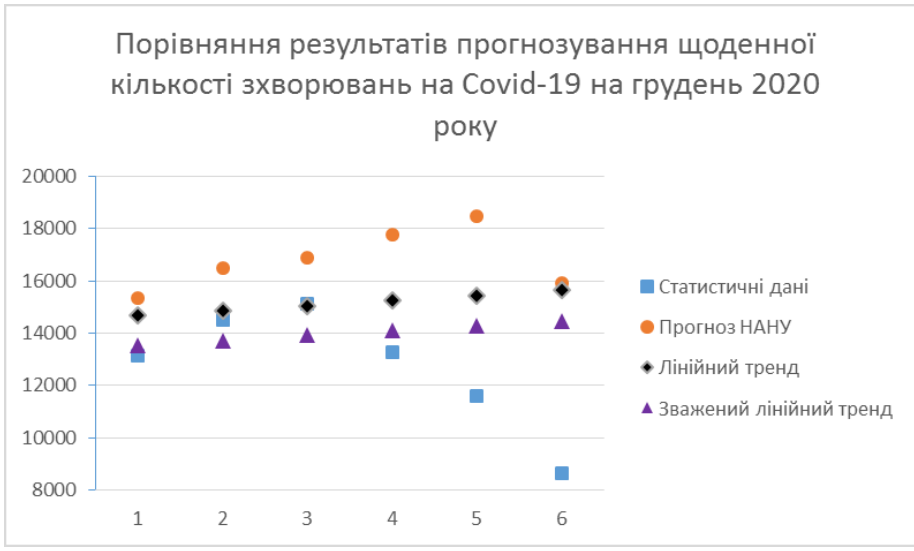

Рис. 4. Діаграма кількості нових випадків захворювання на період 1–6 грудня

СПИСОК ПОСИЛАНЬ

Аналіз поточної ситуації та моделювання сценаріїв поширення захворювання COVID-19, 2020. *Міністерство охорони здоров'я України,* [online] 18 червня. Доступно: <https://moz.gov.ua/article/news/analiz-potochnoi-situacii-ta-modeljuvannja-scenariiv-poshirennja-zahvorjuvannja-covid-19> [Дата звернення 30 листопада 2020].

Булига, К.Б., Булига, О.А. та Волощук, О.М., 2018. *Практикум з чисельних методів в інформатиці*. Київ: Київський національний університет культури і мистецтв.

Быстрый старт прогнозирования временных рядов на практическом примере с использованием FB Prophet, 2019. *Машинное обучение, нейронные сети, искусственный интеллект.* [online] Доступно: <https://www.machinelearningmastery.ru/a-quick-start-of-time-series-forecasting-with-a-practical-example-using-fb-prophet-31c4447a2274> [Дата обращения 10 ноября 2020].

Епідемічна ситуація щодо COVID-19 у регіонах україни, 2020. *Центр громадського здоров'я Міністерства охорони здоров'я України.* [online] 9 листопада. Доступно: <https://www.phc.org.ua/news/epidemichna-situaciya-schodo-covid-19-u-regionakh-ukraini> [Дата звернення 30 листопада 2020].

Прогноз розвитку епідемії COVID-19 в Україні на 23 листопада – 7 грудня 2020 року («Прогноз РГ-29»), 2020. *Національна академія наук України.* [online] 23 листопада. Доступно: <https http://www.nas.gov.ua/UA/Messages/Pages/View.aspx?MessageID=7187 [Дата звернення 30 листопада 2020].

Система моніторингу расширения эпидемии коронавірусу, 2020. *Апарат Ради національної безпеки і оборони України,* [online] Доступно: <https://covid19.rnbo.gov.ua/> [Дата звернення 30 листопада 2020].

У світі понад 62 мільйони випадків COVID-19: карта, 2020. *ZN.UA*. [online] 29 листопада. Доступно: <https://zn.ua/ukr/WORLD/chislo-hvorih-na-koronavirus-u-sviti-dosyagaye-1-5-mln-ssha-yak-i-ranishe-v-liderah-karta-343909_html> [Дата звернення 30 листопада 2020].
Hyndman, R.J. and Athanasopoulos, G., 2018. 7.3 Holt-Winters' seasonal method. In: *Forecasting: Principles and Practice.* 2nd ed. [online] Melbourne: OTexts. Available at: <https://otexts.com/fpp2/holt-winters.html> [Accessed 5 October 2020].

REFERENCES

Analiz potochnoi sytuatsii ta modeliuvannia stsenariiv poshyrennia zakhvoriuvannia COVID-19 [Analysis of the current situation and modeling of scenarios for the spread of the disease COVID-19], 2020. *Ministry of health of Ukraine*, [online] 18 June. Available at: <https://moz.gov.ua/article/news/analiz-potochnoi-situacii-ta-modeljuvannja-scenariiv-poshirennja-zahvorjuvannja-covid-19> [Accessed 30 November 2020].
Bulyha, K.B., Bulyha, O.A. and Voloshchuk, O.M., 2018. *Praktykum z chyselnykh metodiv v informatytsi* [Workshop on numerical methods in computer science]. Kyiv: Kyivskyi natsionalnyi universytet kultury i mystetstv.
Bystryj start prognozirovanija vremennyh rjadov na prakticheskom primere s ispolzovaniem FB Prophet [Quick start of time series forecasting on a practical example using FB Prophet], 2019. *Mashinnoe obuchenie, nejronnye seti, iskusstvennyj intellekt*. [online] Available at: <https://www.machinelearningmastery.ru/a-quick-start-of-time-series-forecasting-with-a-practical-example-using-fb-prophet-31c4447a2274> [Accessed 10 November 2020].
Epidemichna sytuatsiia shchodo COVID-19 u rehionakh ukrainy [Epidemic situation regarding COVID-19 in the regions of Ukraine], 2020. *Tsentr hromadskoho zdorovia Ministerstva okhorony zdorovia Ukrainy.* [online] 9 November. Available at: <https://www.phc.org.ua/news/epidemichna-situaciya-schodo-covid-19-u-regionakh-ukraini> [Accessed 30 November 2020].
Hyndman, R.J. and Athanasopoulos, G., 2018. 7.3 Holt-Winters' seasonal method. In: *Forecasting: Principles and Practice.* 2nd ed. [online] Melbourne: OTexts. Available at: <https://otexts.com/fpp2/holt-winters.html> [Accessed 5 October 2020].
Prohnoz rozvytku epidemii COVID-19 v Ukraini na 23 lystopada – 7 hrudnia 2020 roku ("Prohnoz RH-29") [Forecast of the COVID-19 epidemic in Ukraine on November 23 – December 7, 2020 ("Forecast RG-29")], 2020. *The National Academy of Sciences of Ukraine*. [online] 23 November. Available at: <http://www.nas.gov.ua/UA/Messages/Pages/View.aspx?MessageID=7187> [Accessed 30 November 2020].
Systema monitorynhu rasshyrenyia эpydemyy koronavirusu [Monitoring system for the spread of the coronavirus epidemic], 2020. *National Security and Defense Council of Ukraine*, [online] Available at: <https://covid19.rnbo.gov.ua/> [Accessed 30 November 2020].
U sviti ponad 62 miliony vypadkiv COVID-19: karta [In the world more than 62 million cases of COVID-19: a map], 2020. *ZN.UA*. [online] 29 November. Available at: <https://zn.ua/ukr/WORLD/chislo-hvorih-na-koronavirus-u-sviti-dosyagaye-1-5-mln-ssha-yak-i-ranishe-v-liderah-karta-343909_html> [Accessed 30 November 2020].

UDC 004.6:616-036.22

Bulyha Kostiantyn,
PhD in Technical Sciences, Associate Professor at the Department of Computer Science,
Kiev National University of Culture and Arts,
Kyiv, Ukraine
piton54@ukr.net
https://orcid.org/ 0000-0002-9537-3226

Bulyha Olena,
Senior Lecturer at the Department of Information Technologies and Design,
State University of Infrastructure and Technology,
Kyiv, Ukraine
hellen.bulyga@gmail.com
https://orcid.org/0000-0003-1612-6735

THE USE OF INFORMATION TECHNOLOGIES FOR ANALYSIS OF EPIDEMIOLOGICAL SITUATION

The aim of the article is to find optimal methods of trend building in order to analyze the epidemiological situation in case of further development of the COVID-19 pandemic. The tasks arisen due to the increase in infection cases in the fall of 2020 are considered in detail.

The research methodology lies in the generalization of the least Gaussian squares method and Facebook Prophet methodology.

The novelty of the study is a comparative analysis of the advantages and disadvantages of systems used for online processing of statistical data, in particular, MS Excel spreadsheets.

Conclusions. The material presented in the article gives clear examples of statistical calculations using widely used software products. The problems arising from the reliability of statistical information have been considered. Examples of real calculations of trends in the epidemiological situation have been presented.

Keywords: pandemic COVID-19; information system; MS Excel; the least squares method; trend line; statistical analysis.

УДК 004.6:616-036.22

Булыга Константин,
кандидат технических наук, доцент кафедры компьютерных наук,
Киевский национальный университет культуры и искусств,
Киев, Украина
piton54@ukr.net
https://orcid.org/0000-0002-9537-3226

Булыга Елена,
старший преподаватель кафедры информационно-аналитической деятельности и информационной безопасности,
Национальный транспортный университет,
Киев, Украина
hellen.bulyga@gmail.com
https://orcid.org/0000-0003-1612-6735

ИСПОЛЬЗОВАНИЕ ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ ДЛЯ АНАЛИЗА ЭПИДЕМИОЛОГИЧЕСКОГО СОСТОЯНИЯ

Целью статьи является поиск оптимальных методов построения тренда с целью анализа эпидемиологической ситуации в случае дальнейшего развития пандемии COVID-19. Подробно рассмотрены задачи, возникшие с ростом случаев инфицирования осенью 2020 года.

Методы исследования: обобщение метода наименьших квадратов Гаусса и методология Facebook Prophet.

Новизной проведенного исследования является сравнительный анализ преимуществ и недостатков систем, используемых для оперативной обработки статистических данных, в частности электронных таблиц MS Excel.

Выводы. Изложенный в статье материал дает наглядные примеры выполнения статистических расчетов с использованием широко распространенных программных продуктов. Рассмотрены проблемы, возникающие в связи с достоверностью статистической информации. Представлены примеры реальных расчетов трендов эпидемиологической ситуации.

Ключевые слова: пандемия COVID-19; информационная система; MS Excel; метод наименьших квадратов; линия тренда; статистический анализ.

01.12.2020