

Технічні науки

УДК: 311.

Якимчук Вікторія Сергіївна

кандидат технічних наук, доцент кафедри БМК

Національний технічний університет України

«Київський політехнічний інститут імені Ігоря Сікорського»

Якимчук Виктория Сергеевна

кандидат технических наук, доцент кафедры БМК

Национальный технический университет Украины

«Киевский политехнический институт имени Игоря Сикорского»

Iakymchuk Victoria

PhD, Associate Professor at the Department of Biomedical Cybernetics

National Technical University of Ukraine

"Igor Sikorsky Kyiv Polytechnic Institute"

Іванова Яна Олегівна

студентка

Національного технічного університету України

«Київський політехнічний інститут імені Ігоря Сікорського»

Иванова Яна Олеговна

студентка

Национального технического университета Украины

«Киевский политехнический институт имени Игоря Сикорского»

Ivanova Yana

Student of the

National Technical University of Ukraine

"Igor Sikorsky Kyiv Polytechnic Institute"

**ВИКОРИСТАННЯ ДАНИХ ПОШУКОВИХ ЗАПИТІВ ДЛЯ
ПРОГНОЗУВАННЯ ІНФЕКЦІЙНИХ ЗАХВОРЮВАНЬ НАСЕЛЕННЯ
УКРАЇНИ**

**ИСПОЛЬЗОВАНИЕ ДАННЫХ ПОИСКОВЫХ ЗАПРОСОВ ДЛЯ
ПРОГНОЗИРОВАНИЯ ИНФЕКЦИОННЫХ ЗАБОЛЕВАНИЙ
НАСЕЛЕНИЯ УКРАИНЫ**

**USING OF SEARCH QUERIES TO PREDICT INFECTIOUS DISEASES
IN UKRAINE**

***Анотація.** Точне прогнозування інфекційних захворювань може зменшити наслідки спалахів хвороби. У дослідженні було запропоновано використання даних про запити користувачів, для відстеження захворюваності населення. Використання Інтернет-ресурсу постійно набирає обертів, тому дані пошукового трафіку, соцмереж та інших веб-джерел можуть сприяти кращому розумінню та аналізу поведінки користувачів.*

Проведене дослідження показало, що дані пошукових запитів населення України корелюють із даними по кількості спалахів інфекційних захворювань. Було знайдено коефіцієнт кореляції та довірчий інтервал для цього значення. Проведено аналіз числового ряду. Побудовано модель та перевірено її на тестових даних.

***Ключові слова:** пошукові запити, прогнозування спалахів, Google Trends, ARIMA, тест Дікі-Фулера, аналіз часових рядів.*

***Аннотация.** Точное прогнозирование инфекционных заболеваний может уменьшить последствия вспышек болезни. В исследовании было предложено использование данных о запросах пользователей, для отслеживания заболеваемости населения. Использование Интернет-ресурса постоянно набирает обороты, поэтому данные поискового*

трафика, соцсетей и других веб-источников могут способствовать лучшему пониманию и анализу поведения пользователей.

Проведенное исследование показало, что данные поисковых запросов населения Украины коррелируют с данными о количестве вспышек инфекционных заболеваний. Было найдено коэффициент корреляции и доверительный интервал для этого значения. Проведен анализ числового ряда. Построена модель и проверено ее на тестовых данных.

Ключевые слова: поисковые запросы, прогнозирование вспышек, Google Trends, ARIMA, тест Дики-Фуллера, анализ временных рядов.

Summary. Accurate forecasting of infectious diseases can reduce the effects of outbreaks. The article is devoted to using of search data to track the morbidity. The use of the Internet resource is constantly gaining momentum, therefore search traffic, social media and other web sources can contribute to a better understanding and analysis of user behavior.

The research conducted shows that search queries of Ukraine's population correlate with dataset on a number of infectious diseases. The correlation coefficient and confidence interval for this time series were calculated. The analysis of the time series has been carried out. A model was built and tested on test data.

Key words: search queries, outbreak forecasting, Google Trends, ARIMA, Dickey-Fuller test, time series analysis.

Вступ. Сьогодні існує досить несприятлива епідеміологічна ситуація у всьому світі, враховуючи територію України [1]. Відбувається багато непередбачуваних спалахів інфекційних захворювань. Вихід з ситуації, що склалася, полягає у своєчасному і точному прогнозі захворюваності населення на обраній місцевості, що надасть можливість прийняття, більш перспективних, рішень у плані підготовки системи медико-профілактичних

і терапевтичних заходів. Таким чином, залишається актуальною проблема теоретичної розробки та практичної реалізації методів прогнозування інфекційної захворюваності для забезпечення санітарно-епідеміологічного благополуччя громадян, збереження і поліпшення їх здоров'я. На даний час математичний апарат прогнозування інфекційної захворюваності представлений досить широким спектром методів, що забезпечують ефективність прогностичних рішень, які приймалися на їх основі [2]. Сучасними напрямками наукового пошуку в цій області є як модифікація класичних методів аналізу часових рядів стосовно захворюваності населення, так і розробка принципово нових, що спираються на останні досягнення науки і техніки способів вивчення динаміки поширення інфекційних захворювань. Спільною рисою таких досліджень є вхідні дані, які формуються та надаються від медичних закладів. Вони включають і пацієнтів, які прийшли на консультацію до лікаря та за медичною допомогою. Однак з розвитком технологій, все більша кількість населення, із першими симптомами хвороби, звертається до Інтернет-ресурсів для знаходження «вірних» методів лікування.

За даними дослідження, проведеного в рамках проекту USAID (Агентство США з міжнародного розвитку) «(Без)коштовна медицина» та Благодійний Фонд «Пацієнти України», в Україні виявлено майже 69% хворих, які займаються самолікуванням [3]. Це призводить до збільшення випадків з ускладненнями та скорочення тривалості життя людей. Хворі такого типу, які схильні до самолікування, починають масово звертатися за допомогою лише тоді, коли з'являється сильне погіршення їх здоров'я. При цьому медичні заклади не готові одночасно прийняти таку кількість пацієнтів, оскільки було прогнозовано меншу кількість випадків.

З огляду на величезний соціальний, економічний та медичний тягар інфекційних захворювань, важливо вміти передбачувати спалахи хвороби [4]. Такі прогнози дозволять забезпечити своєчасну та адекватну підготовку

медичного персоналу, ресурсів медичних закладів, таких як противірусні препарати (особливо вакцини) або наявність необхідних медичних приладів. Це знизить рівень смертності та кількість пацієнтів. Тому, дані Інтернет-запитів одне із джерел інформації, яке може допомогти будувати точні прогнози.

Матеріали та методи дослідження. Дані пошукових запитів були отримані з сервісу Google Trends [5]. Дана платформа показує відносний обсяг пошуку, який являє собою частку запиту по терміну для заданого періоду часу і місця розташування. Можна отримати часовий ряд для будь-якого обраного терміну в будь-якій країні, або регіоні, за вибраний період часу. Наприклад, певний рік, кілька років, 3 тижні, 4 місяці, 30 днів, 7 днів, 4 години, 1 годину. Крім того, різні терміни в різних регіонах можна порівнювати одночасно. Дані завантажуються з Інтернету в форматі «.csv». Діапазон значень від 0 до 100 нормалізується за найбільшою часткою запиту терміну, за часовим рядом. Це значення також представляє індекс «цікавості до пошуку», де 100 позначається, як пікова популярність, а число 50 являє рівно половину популярності [6]. Була завантажена інформація за 2019 рік для терміну «кір», регіон – Україна.

Статистику захворюваності на кір було отримано із сайту Центру громадського здоров'я (ЦГЗ) Міністерства охорони здоров'я (МОЗ) України [7]. Дані представлені як кількість спалахів, станом на кожен тиждень року за 2019 рік по всій території України.

З ціллю оцінити актуальність такого підходу, спершу було визначено зв'язок пошукового трафіку з офіційними спалахами із сайту ЦГЗ [7]. Для визначення зв'язку між даними було застосовано коефіцієнт кореляції Пірсона. Оскільки, даний показник розраховано за вибіркою, яка включає лише дані за 2019 рік, то отримане значення не істинне, а є оцінкою кореляції. Так, як обрана вибірка є репрезентативною, то справжнє значення

коефіцієнта кореляції знаходиться десь відносно недалеко від оцінки. Наскільки далеко, було визначено через довірчий інтервал.

Для прогнозу спалаху інфекційного захворювання використано модель $ARMA(p,d,q)x(P,D,Q)$. Вона складається із авторегресійної моделі, моделі ковзного середнього та сезонної компоненти. Параметр p – порядок моделі AR (p), d – порядок інтегрування, q – порядок моделі MA (q), P – порядок сезонної складової AR (P), D – порядок інтеграції сезонної складової, Q – порядок сезонної складової MA (Q) [12]. Для побудови моделі взято з сервісу Google Trends набір даних, який являє собою ряд, що містить щотижневі дані про пошукову активність користувачів, які робили запит в браузері за терміном «кір» на всій території України протягом п'яти років.

Перед побудовою моделі потрібно перевірити дані на стаціонарність, а також знайти необхідний набір параметрів. Ряд є стаціонарним, якщо його статистичні властивості, такі як середня величина, дисперсія з часом залишаються постійними. Оскільки, не завжди можливо зробити такі візуальні висновки із загального графіка ряду, можна застосувати *Rolling Statistics*. Необхідно побудувати ковзне середнє значення та ковзне середнє відхилення. Під ковзним середнім і середнім відхиленням розуміють, що в будь-який момент часу « t » необхідно взяти середнє значення за минулий рік, тобто за останні 12 місяців. Однак це більш візуальна техніка, тому важливо додатково зробити тест Дікі-Фуллера (DF-тест, Dickey-Fuller test) [13]. Це один з статистичних тестів для перевірки стаціонарності альтернативної гіпотези, що варіюється у залежності від версії тесту. Тут нульова гіпотеза полягає у тому, що ряд нестаціонарний та вказує на наявність одиночного кореня в авторегресійній моделі. Результати тесту включають в себе статистику тесту і деякі критичні значення для рівнів достовірності відмінностей. Якщо «*Статистика тесту*» менша за «*Критичну цінність*», то можемо відхилити нульову гіпотезу і сказати, що

ряд є стаціонарним і приймаємо альтернативну гіпотезу [11]. У даному дослідженні під альтернативною гіпотезою

Для оптимізації цього процесу аналізу було створено веб-додаток, який містить всі необхідні інструменти для роботи з даними.

Для інструментальної бази додатку використовувалися бібліотеки мови програмування *Python*: *NumPy*, *SciPy*, *Pandas*, *Matplotlib*, *Statsmodels*. *NumPy* застосовувалася для наукових обчислень. Вона працює з N-мірним об'єктом масиву і надає базові математичні функції, такі як розмір, форма, середнє значення, стандартне відхилення, мінімум, максимум, а також деякі більш складні функції, такі як лінійні алгебраїчні функції і перетворення Фур'є. Бібліотека *Pandas* надає високоефективні і прості, у використанні, структури даних, такі як серії. Вона розширює функціональність *Python* від простого збору та підготовки даних до аналізу. *Science Python* – це бібліотека, яка використовується для наукових і технічних обчислень. Вона надає функції для оптимізації, обробки сигналів і зображень, інтеграції, інтерполяції та лінійної алгебри. *Statsmodels* широко використовується для статистичного моделювання, машинного навчання та глибокого навчання, оскільки містить різні моделі регресії, класифікації та кластеризації, які є можливість налаштовувати. Для візуалізації даних, таких як лінійний графік, гістограма, діаграми розсіювання, гістограма, використовують бібліотеку *Matplotlib*. Вона містить всі функції, пов'язані з графіком, необхідні від побудови графіка до маркування.

Для написання інтерфейсу використовували фреймворк *Dash*. Додатки *Dash* – веб-сервери, які запускають *Flask* і обмінюються даними з пакетами *JSON* через *HTTP*. Їхній інтерфейс забезпечує компоненти можливостями *React.js*. Додатки, розроблені за допомогою *Dash*, відображаються у веб-браузері та можуть бути розгорнуті на серверах.

Результати дослідження. Оцінювання доцільності запропонованого підходу відбувалось спершу через визначення зв'язку пошукового трафіку з офіційними даними. Оскільки, часовий ряд, який завантажений із *Google Trends* відображає ступінь популярності запиту, а статистичні дані із Центру громадського здоров'я – кількість пацієнтів, для подальшого аналізу отримані дані були нормовані. Нормування проводилося за методом мінімакс (1).

$$X_{norm} = \frac{x - x_{min}}{x_{max} - x_{min}}, \quad (1)$$

де x_{min} – найменше значення обраного ряду;

x_{max} – найбільше значення того ж ряду.

Розглянемо отримані нормовані дані (рис. 1) і можна побачити, що разом із «поетапним» зменшенням випадків кору у цій епідемії різко зменшилася кількість онлайн-пошуків, пов'язаних із кором, у Google-пошуку.

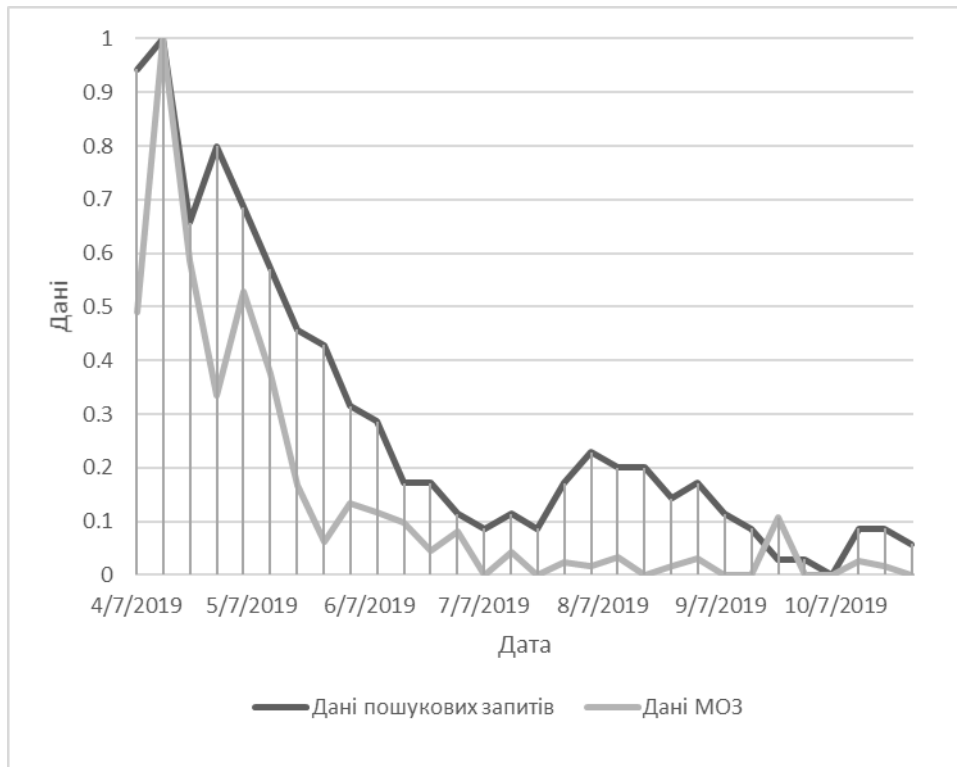


Рис. 1. Графік нормованих часових рядів тижневого обсягу *Google Trends* та зафіксованих спалахів хвороби «кір»

На графіку розсіювання можна побачити, що точки певним чином скупчені навколо лінії тренду, що показує наявність зв'язку між запитами в інтернеті про «кір» та даними МОЗ про спалахи хвороби (рис. 2).

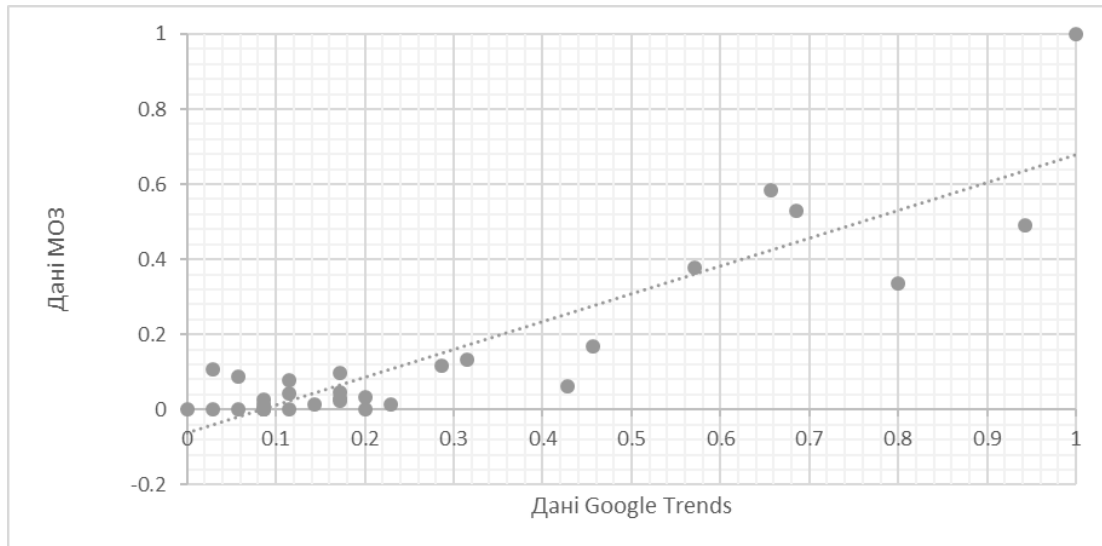


Рис. 2. Графік розсіювання між нормованими даними про пошукові запити та про спалахи хвороби

Знайдений коефіцієнт кореляції Пірсона становить 0.895. Отже, дані *Google Trends* мають статистично значущі та позитивні кореляції із кількістю підтверджених випадків захворювань [8]. Для визначення довірчого інтервалу використовувалося перетворення Фішера (2) та стандартна похибка (3).

$$z = 0,5 \ln \left(\frac{1+r}{1-r} \right), \quad (2)$$

$$se_z = \frac{1}{\sqrt{n-3}}. \quad (3)$$

Виходячи з властивостей нормального розподілу було знайдено верхню і нижню межі довірчого інтервалу для z . Для цього визначимо квантілі стандартного нормального розподілу для заданої довірчої ймовірності, тобто кількість стандартних відхилень від центру розподілу [9].

$$c_\gamma = N^{-1} = \left(\frac{1+\gamma}{2} \right), \quad (4)$$

де c_γ – квантиль стандартного нормального розподілу;

N^{-1} – функція зворотного стандартного розподілу;

γ – довірна ймовірність (часто 95%).

Межі розраховувались за допомогою:

$$z_{L(U)} = z \pm se_z \cdot c_\gamma. \quad (5)$$

де se_z – стандартна похибка.

Для обчислених меж застосовано зворотне перетворення Фішера [10].

$$r_{L(U)} = \frac{e^{2z_{L(U)}} - 1}{e^{2z_{L(U)}} + 1}. \quad (6)$$

де $z_{L(U)}$ – знайдені межі.

Вище наведені формули було застосовано та проведено всі відповідні розрахунки (табл. 1).

Таблиця 1

Знаходження довірного інтервалу

Показник	Значення
Коефіцієнт кореляції Пірсона	0.89540249
z (довірчий інтервал)	1.44853339
se (стандартна похибка)	0.18898224
$z_{.95\%}$ (квантиль стандартного нормального розподілу)	1.95996398
Нижня 95% межа z_L	1.07813501
Верхня 95% межа z_U	1.81893177
Нижня 95% межа r_L	0.79250647
Верхня 95% межа r_U	0.94873181

Знайдений інтервал [0.79250647; 0.94873181] показує, що кореляція є високою, та дає підстави вважати, що захворюваність населення пов'язана певним чином з відображенням пошукових запитів в Інтернет-ресурсах. Отже, запропоновані дані можна використовувати для прогнозування захворюваності. Для побудови більш точної моделі було взято дані за 5 років.

Основним положенням, на якому базується використання часових рядів для прогнозування, є те, що фактори, що впливають на отримані дані,

впливали певним чином на процес у минулому, і передбачається, що вони будуть діяти схожим чином і в майбутньому. Із-за наявних у часових рядах певних компонентів, перед прогнозуванням проводиться їхній аналіз. Використано *STL*-декомпозицію (рис. 3) та *Rolling Statistics* для перегляду таких складових (рис. 4).

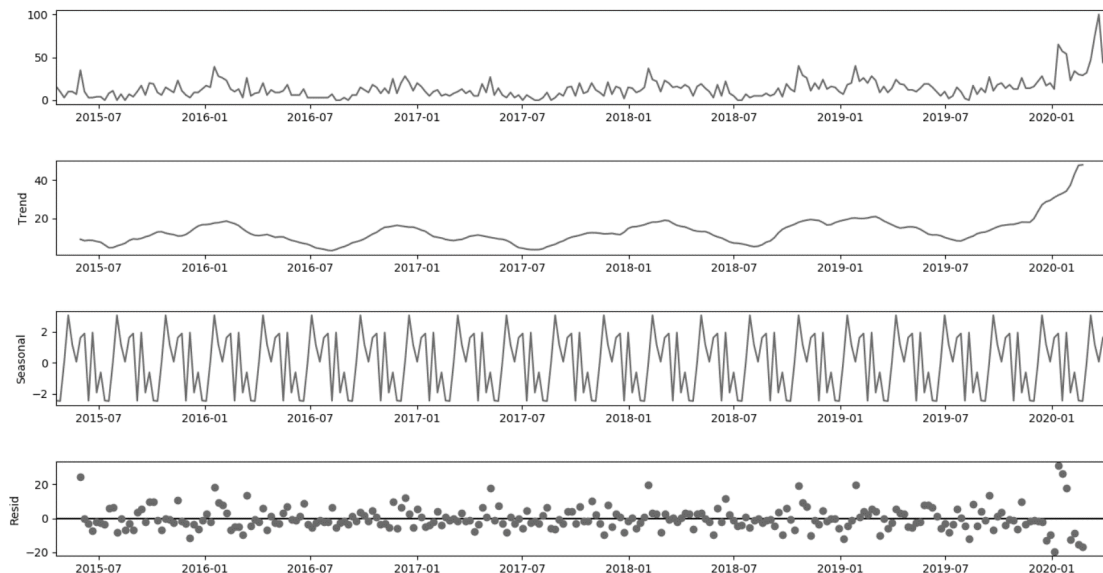


Рис. 3. Декомпозиція числового ряду

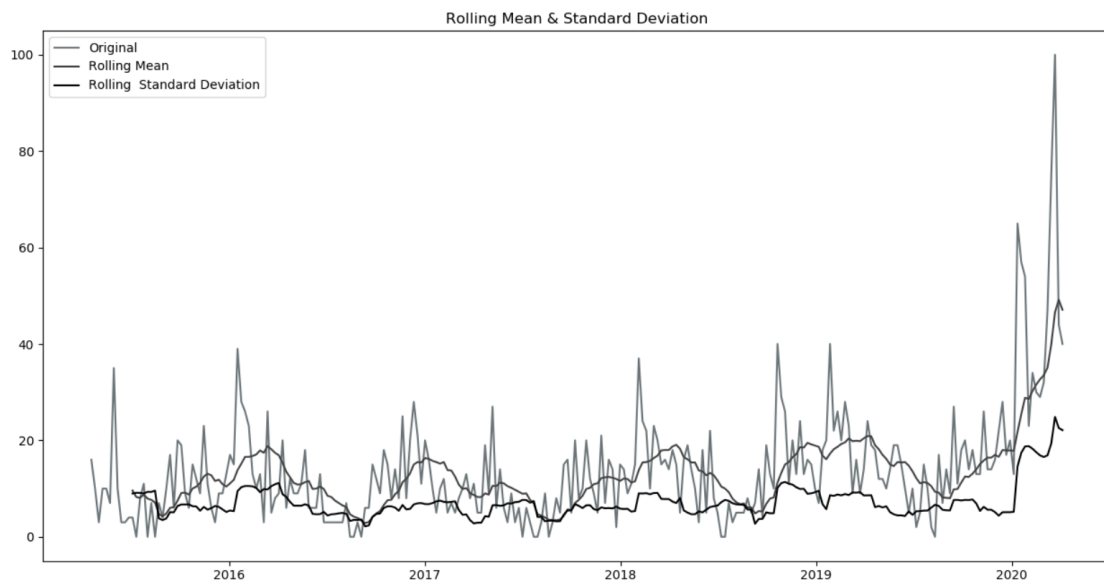


Рис. 4. Rolling Statistics

На рис. 4, де відображено ковзне середнє значення та ковзне стандартне відхилення, можна помітити, що показники мають

закономірність змінюватися із часом. Оскільки, візуально важко точно оцінити ряд, було застосовано тест Дікі-Фулера (рис. 5).

```
Results of Dickey-Fuller Test:  
Test Statistic           -1.342223  
p-value                  0.609641  
#Lags Used                10.000000  
Number of Observations Used  249.000000  
Critical Value (1%)      -3.456888  
Critical Value (5%)     -2.873219  
Critical Value (10%)    -2.572994
```

Рис. 5. Результати тесту Дікі-Фулера для початкового ряду

Згідно отриманого p -значення, видно, що гіпотеза про стаціонарність ряду відхиляється. Також *Test Statistic* більша за критичні значення, що також показує нестаціонарність ряду. На наступному кроці було проведено сезонне диференціювання ряду (рис. 6). Тест Дікі-Фелера (рис. 7), як і графік середніх значень, показав, що числовий ряд є стаціонарним.

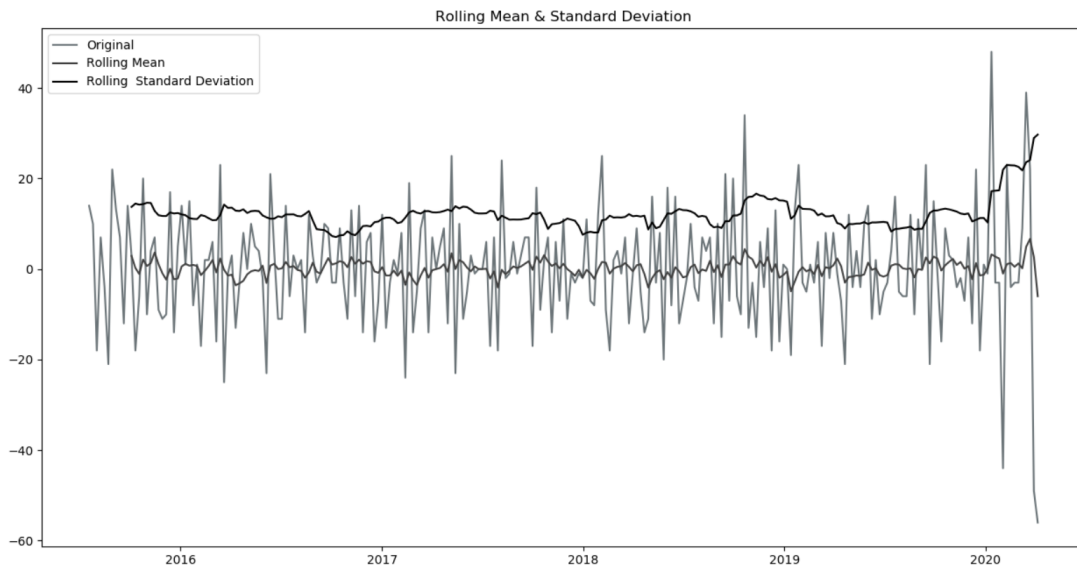


Рис. 6. Rolling Statistics після сезонного диференціювання

```
Results of Dickey-Fuller Test:  
Test Statistic           -6.928844e+00  
p-value                  1.097354e-09  
#Lags Used                1.400000e+01  
Number of Observations Used  2.320000e+02  
Critical Value (1%)       -3.458855e+00  
Critical Value (5%)       -2.874080e+00  
Critical Value (10%)      -2.573453e+00
```

Рис. 7. Результати тесту Дікі-Фулера ряду з сезонним диференціюванням

Для знаходження початкових параметрів p , P , Q та q було застосовано графіки автокореляції та часткової автокореляції (рис. 8). Вони показують значення кореляції Пірсона між початковим рядом та цим же рядом, але із затримкою на певну кількість позицій, які називають лагами. Значення Q – це номер останнього сезонного лага (по порядку) зі значущою автокореляцією, q – номер останнього несезонного лага зі значущою автокореляцією, P – номер останнього сезонного лага зі значною частковою автокореляцією, p – номер останнього несезонного лага зі значною частковою автокореляцією. Такі значення є лише початковим наближенням до справжніх параметрів моделі.

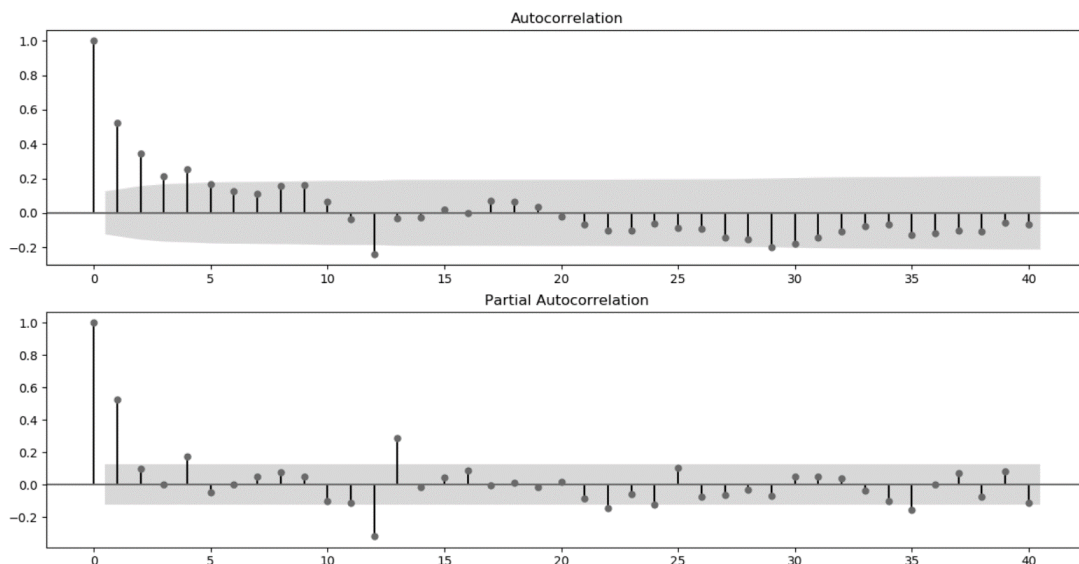


Рис. 8. Графіки автокореляції та часткової автокореляції

Далі було проаналізовано моделі із всіма можливими параметрами, які менші та дорівнюють початковим значенням q , Q , p , P . Для порівняння моделей використовувався критерій Акаїке (AIC) та Байєсівський інформаційний критерій (BIC) [12]. Кращою є та модель, для якої значення критеріїв найменше. Значення критеріїв для побудованих моделей наведено в таблиці 2. Отримані моделі відсортовані за зростанням значення критеріїв, тому в таблиці наведено лише перші 8 моделей, оскільки всі наступні будуть мати ще більше значення.

Таблиця 2

Порівняння моделей ARIMA

	Параметри моделі	Значення критерію AIC	Значення критерія BIC
1	ARIMA(3, 0, 1)x(0, 0, 2)	1815.386	1843.461
2	ARIMA (3, 0, 2)x(0, 0, 2)	1816.723	1848.307
3	ARIMA (3, 0, 2)x(1, 0, 1)	1817.959	1849.544
4	ARIMA(3, 0, 2)x(1, 0, 2)	1818.465	1853.559
5	ARIMA(3, 0, 2)x(0, 0, 1)	1818.611	1846.686
6	ARIMA (1, 0, 2)x(1, 0, 2)	1821.836	1849.911
7	ARIMA(2, 0, 2)x(0, 0, 1)	1823.425	1847.991
8	ARIMA (0, 0, 0)x(0, 0, 0)	1977.192	1984.210

Результати перебору моделей показали, що найбільш оптимальною згідно до значень AIC та BIC є модель з параметрами: $p=3$, $q=1$, $d=0$, $P=0$, $Q=2$, $D=0$.

Дані пошукових запитів розділено на тренувальну (75% від всієї вибірки) та тестову (25%) вибірки. Також було побудовано прогноз пошукових запитів (рис. 9).

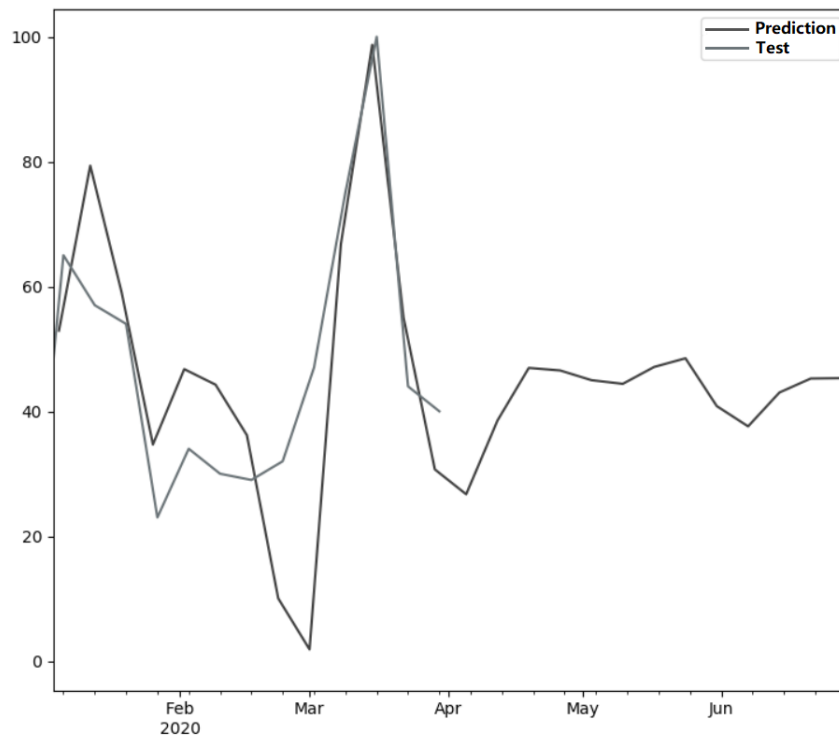


Рис. 9. Побудова прогнозу пошукових запитів

З графіку видно, що лінія побудованого прогнозу за допомогою моделі $ARIMA(3, 0, 1) \times (0, 0, 2)$ проходить досить близький до вхідного ряду. Для детальнішого аналізу знайдено різницю значень прогнозу та тестової вибірки та проведено статистичний аналіз (рис. 10).

count	1.400000e+01
mean	-4.304501e-07
std	2.398817e+01
min	-3.991954e+01
25%	-1.839888e+01
50%	-4.126339e+00
75%	1.755113e+01
max	4.496438e+01

Рис. 10. Аналіз різниць прогнозованого ряду та тестової вибірки

Середнє значення та середнє відхилення залишків малі, що свідчить про досить гарну якість прогнозу, який необхідно було провести. Також, від'ємні показники відображають наявність випередження у прогнозі, проводячи порівняння з фактичними значеннями тестової вибірки. Таку ситуацію можна легко скорегувати, додавши коефіцієнт, який дорівнює

середньому значенню всіх не співпадінь прогнозу із вхідним рядом, до значень, що обчислюються.

Висновки. Традиційні системи спостережень обмежують здатність органів охорони здоров'я ефективно реагувати на спалахи інфекційних захворювань. Оскільки, інформація *Google Trends* збирається та обробляється майже в режимі реального часу, пошукові запити в Інтернет-ресурсах створюють моніторингові дані набагато швидше ніж традиційні системи. Проведено аналіз для дослідження кореляції між даними про кір із пошукових запитів та зафіксованими випадками Центру громадського здоров'я на території України протягом п'яти років. Виявлено, що пошукові запити в *Google Trends* мають сильну кореляцію з зафіксованими випадками кору в Україні. Для побудови прогнозу пошукового трафіку для терміну «кір» вхідні дані розділено на тестову та тренувальну вибірку, яка містила дані про активність користувачів Інтернету. Проаналізовано декілька моделей, серед яких найкращий результат отримала $ARIMA(3, 0, 1)x(0, 0, 2)$. Прогнозовані дані, згідно одержаної моделі *ARIMA*, практично на 70% схожі з фактичними даними.

Результати цього дослідження показали, що новий інструмент спостереження спалаху інфекційних захворювань через *Google Trends* може відображати динамічну, своєчасну інформацію органам охорони здоров'я та надавати, майже в реальному часі, показники поширеності інфекційних захворювань. Доведено, що дані пошукових запитів можна прогнозувати. Це є корисним у ряді обставин: оптимізація планування витрат у лікарнях (кількість медичних працівників, засобів захисту, медичних приладів тощо), виробництво лікарських препаратів, включаючи вакцини та противірусні лікарські засоби, вжиття запобіжних заходів, таких як закриття шкіл, державних установ та закладів громадської діяльності.

Література

1. World health statistics 2019: monitoring health for the SDGs, sustainable development goals // World Health Organization. 2019. URL: https://www.who.int/gho/publications/world_health_statistics/2019/en/.
2. Technological Forecasting and Social Change. // Elsevier. 2017. №115.
3. (Без)коштовна медицина. 2017. URL: <https://moz.gov.ua/article/statistic/rezultati-doslidzhennja-bezkoshtovna-medicina>.
4. Глобальное бремя болезней. // «Украинский медицинский журнал». 2017.
5. Google Trends: [Веб-сайт]. Google LLC, 2020. URL: <https://trends.google.com/trends/?geo=UA>
6. Google Trends lessons : [Веб-сайт]. Google LLC, 2020. URL: <https://newsinitiative.withgoogle.com/training/lessons?tool=Google%20Trends&image=trends>
7. Центр громадського здоров'я МОЗ України: [Веб-сайт]. Державна установа «Центр громадського здоров'я Міністерства охорони здоров'я України», 2020. URL: <https://phc.org.ua/kontrol-zakhvoryuvan/inshi-infekciyni-zakhvoryuvannya/operativni-dani-pro-spalakhi> – ОПЕРАТИВНІ ДАНІ ПРО СПАЛАХИ
8. Карасев В.А., Богданов С.Н., Лёвшина Г.Д.. Теория вероятностей и математическая статистика. Раздел 2. Математическая статистика. Учебно-методическое пособие. М.: Учеба (МИСиС). 2006. 116 с.
9. Гржибовский А. М. Корреляционный анализ / А. М. Гржибовский. // Экология человека. 2008. №9. С. 50–60.
10. Анализ данных. Учебное пособие. М.: Мир науки, 2019. Сетевое издание. URL: <https://izd-mn.com/PDF/32MNNPU19.pdf>

11. Robert Nau. Statistical forecasting: notes on regression and time series analysis / Robert Nau // Fuqua School of Business Duke University. 2019. URL: <https://people.duke.edu/~rnau/411home.htm>.
12. Hyndman R. Forecasting: Principles and Practice / R. Hyndman, G. Athanasopoulos // Monash University, Australia. 2018. URL: <https://otexts.com/fpp2/>.
13. Brownlee J. Introduction to Time Series Forecasting With Python: How to Prepare Data and Develop Models to Predict the Future / Brownlee, 2018. 367 c.