



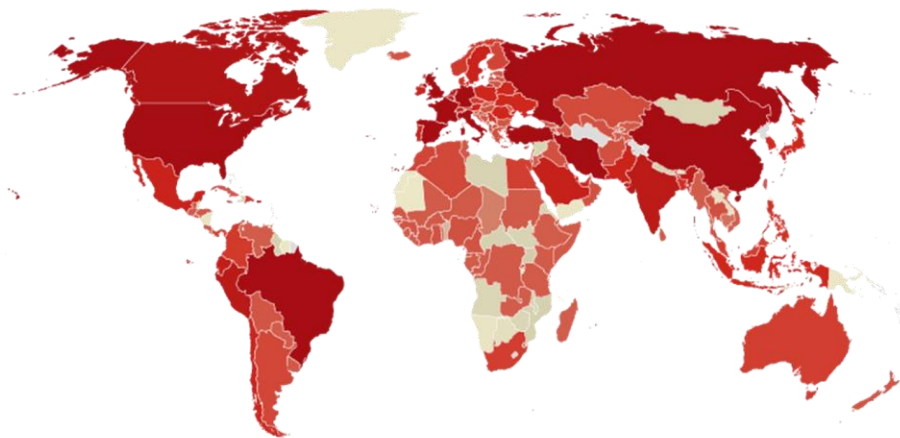
ПРОГНОЗУВАННЯ ЕПІДЕМІЧНОЇ СИТУАЦІЇ ВИКЛИКАНОЇ ЗАХВОРЮВАННЯМ COVID-19 ЗА ДОПОМОГОЮ МЕТОДУ АНАЛІЗУ ЧАСОВИХ РЯДІВ

Виконав студент 1 курсу
2 (магістерського) рівня
Киян М.Є.

Керівник курсової роботи
канд. техн. наук
Ковалюк Т.В.



Актуальність



ХРОНОЛОГІЯ ПОДІЙ

- 31 грудня 2019 року – звіт від китайської влади про новий тип вірусу у місті Ухань
- 13 січня – перший випадок зараження за межами Китаю
- 3 березня – перший зафіксований випадок захворювання в Україні
- 11 березня – Комітет Всесвітньої організації охорони здоров'я: пандемія коронавірусу – надзвичайна ситуація міжнародного значення

НАГАЛЬНІ ПИТАННЯ КОЖНОЇ КРАЇНИ

- Чи вводити/послаблювати/знімати карантин?
- Якого навантаження лікарень очікувати?
- За яких умов можна стверджувати, що ризик виникнення нового спалаху мінімальний?

ПІДРҀРУНТЯ ДЛЯ ПРИЙНЯТТЯ РІШЕНЬ

- Створення моделі та прогнозування епідемічного процесу

МЕТА

прогнозування епідемічної ситуації
викликаного захворюванням COVID-19
у різних країнах, у тому числі і
України



Часові ряди та їх аналіз. Поняття

Coronavirus Cases:

4,220,901

[view by country](#)

Deaths:

284,817

Recovered:

1,508,928

Ситуація в Україні

181552

протестовано

15648

хворих на Covid-19

3288

одужало

408

летальних випадків

ACTIVE CASES

2,427,156

Currently Infected Patients

2,380,071 (98%)

In Mild Condition

47,085 (2%)

Serious or Critical

[Show Graph](#)

CLOSED CASES

1,793,745

Cases which had an outcome:

1,508,928 (84%)

Recovered / Discharged

284,817 (16%)

Deaths

[Show Graph](#)

ЧАСОВИЙ РЯД

– зібраний у різні моменти часу статистичний матеріал про значення будь-яких параметрів досліджуваного процесу

АНАЛІЗ ЧАСОВИХ РЯДІВ

– сукупність математико-статистичних методів аналізу, призначених для виявлення структури часових рядів і для їх прогнозування

ІСНУЮЧІ МЕТОДИ

– інтерполяції, двох крайніх точок, середніх групових точок, Холта-Уінтерса, Трігга-Ліча, Чоу, локальна апроксимація, сингулярний спектральний аналіз ...

1

Сингулярний спектральний аналіз

(Singular spectrum analysis)

- ✓ метод аналізу часових рядів, заснований на перетворенні одновимірного часового ряду у багатовимірний ряд з подальшим застосуванням до отриманого багатомірного часового ряду методу головних компонент;
- ✓ області використання: кліматологія, морська наука, геофізика, машинобудування, обробка зображень, медицина, економетрика тощо.



ML.NET

2

ML.NET та принципи роботи

(Singular spectrum analysis)

- ✓ ML.NET – це кросплатформна середина машинного навчання з відкритим вихідним кодом (Windows, Linux, macOS) для розробників .NET;
- ✓ У основі лежить модель машинного навчання, яка визначає кроки, які необхідно виконати для отримання прогнозів на основі вхідних даних.



Обробка статистичних даних

- ✓ Більше ніж 19 тисяч датасетів (один з найпопулярніших порталів у сфері Data Science)
- ✓ Доступ до даних через публічний API для зареєстрованих користувачів (через Authentication Header)
- ✓ Результат – .zip архів

The screenshot shows the Kaggle website interface. The browser address bar displays the URL: `kaggle.com/imdevskp/corona-virus-report# covid_19_clean_complete.csv`. The page title is "COVID-19 Dataset" and the description is "Number of Confirmed, Death and Recovered cases every day across the globe". The dataset is updated 8 hours ago (Version 102) by Devakumar kp. The page includes a search bar, a navigation menu with options like Home, Compete, Data, Notebooks, Discuss, Courses, and More, and a sidebar with options like Data, Tasks (1), Kernels (303), Discussion (34), Activity, Metadata, Download (39 MB), and New Notebook. The page also shows a Usability score of 10.0, a License of "Other (specified in description)", and tags including "statistics, law, infectious diseases, public health, health foundations and medical research and 2 more".

РОЗРОБКА ЗАСТОСУВАННЯ

Функціональні вимоги

- ✓ Оновлення статистичних даних з платформи Kaggle за допомогою її API.
- ✓ Завантаження користувачем статистичних даних у .csv форматі.
- ✓ Відображення на карті світу даних щодо кількості хворих, одужавших та померлих за останній день з набору даних по кожній країні.
- ✓ Вибір країни для подальших обчислень та прогнозування.
- ✓ Відображення на графіку кількості хворих, одужавших та померлих за увесь час вибраної країни.
- ✓ Побудова кругової діаграми за вибраний день.
- ✓ Прогноз даних з наступними введеними параметрами: кількість днів та параметр розміру вікна для алгоритму SSA.
- ✓ Тестування прогнозу на основі порівняння з реальними даними з датасету (параметри аналогічні попереднім вимогам: кількість днів та параметр розміру вікна для алгоритму SSA).
- ✓ Відображення реальних, спрогнозованих, а також спрогнозованих для тестування даних на графіку.
- ✓ Обрахунок метрик прогнозування даних.



Аналіз отриманих результатів

y_j – фактичне значення змінної для j -го виміру,

\hat{y}_j – прогнозне значення досліджуваної змінної,

N – розмір певної вибірки вимірювань

ПОМИЛКА ПРОГНОЗУВАННЯ

$$e_y = y_j - \hat{y}_j$$

СЕРЕДНЯ ВІДСОТКОВА ПОМИЛКА ПРОГНОЗУВАННЯ

$$MAE = 100\% * \frac{\sum_{j=1}^N |y_j - \hat{y}_j|}{N}$$

СЕРЕДНЄ КВАДРАТИЧНЕ ВІДХИЛЕННЯ

$$RMSE = \sqrt{\frac{\sum_{j=1}^N (y_j - \hat{y}_j)^2}{N}}$$

СЕРЕДНЯ ВІДСОТКОВА АБСОЛЮТНА ПОМИЛКА ПРОГНОЗУВАННЯ

$$MAPE = 100\% * \frac{1}{N} \sum_{j=1}^N \frac{|y_j - \hat{y}_j|}{y_j}$$

ДЕМОНСТРАЦІЯ



ДЯКУЮ ЗА УВАГУ!

