

НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ
«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ імені ІГОРЯ СІКОРСЬКОГО»
(КПІ ім. Ігоря Сікорського)

ФАКУЛЬТЕТ БІОМЕДИЧНОЇ ІНЖЕНЕРІЇ
кафедра БІОМЕДИЧНОЇ КІБЕРНЕТИКИ

«До захисту допущено»

В.о. завідувач кафедри БМК

_____ **Євген НАСТЕНКО**

“ ” _____ 2023р.

Дипломна робота
на здобуття ступеня бакалавра
за освітньо-професійною програмою
«Комп'ютерні технології в біології та медицині»
спеціальності 122 «Комп'ютерні науки»

на тему: **Інтерактивний додаток до прогнозування захворюваності**
COVID-19 на основі даних про вакцинацію

Виконала: студентка IV курсу, групи БС-93

ШЛАПАК МИЛАНА ІГОРІВНА

(прізвище, ім'я, по батькові)

(підпис)

Керівник: *асистент каф. біомедичної кібернетики*
(БМК) Давидович Ілля Вікторович

(посада, науковий ступінь, вчене звання, прізвище, ім'я, по батькові)

(підпис)

Консультант з розділів дипломної роботи:

проф. Каф. БМК, д.т.н. проф. Третьякова Лариса Дмитрівна

(назва розділу) (посада, вчене звання, науковий ступінь, прізвище, ім'я, по батькові)

(підпис)

Рецензент: *доцент каф. біомедичної інженерії, к.т.н.*

Рудніцька Олена Володимирівна

(посада, науковий ступінь, вчене звання, прізвище, ім'я, по батькові)

(підпис)

Засвідчую, що у цій дипломній роботі немає
запозичень з праць інших авторів без
відповідних посилань.

Студентка _____
(підпис)

Київ – 2023 року

Національний технічний університет України
«Київський політехнічний інститут імені Ігоря Сікорського»
Факультет біомедичної інженерії
Кафедра біомедичної кібернетики

Рівень вищої освіти – перший (бакалаврський)

Спеціальність – 122 «Комп'ютерні науки»

Освітньо-професійна програма «Комп'ютерні технології в біології та медицині»

ЗАТВЕРДЖУЮ

В.о. завідувач кафедри БМК

_____ Євген НАСТЕНКО

« 30 » травня 2023 р.

ЗАВДАННЯ
на дипломну роботу студентці

ШЛАПАК МИЛАНА ІГОРІВНА

(прізвище, ім'я, по батькові)

1. Тема роботи **Інтерактивний додаток до прогнозування
захворюваності COVID-19 на основі даних про вакцинацію**

Керівник роботи

Давидович Ілля Вікторович

(прізвище, ім'я, по батькові, науковий ступінь, вчене звання)

затверджені наказом по університету від «31_» травня 2023 р. №2106-с

2. Термін подання студентом звіту **18-20 травня 2023 року**
3. Вихідні дані до роботи статистичні дані щодо захворюваності на COVID-19
4. Зміст роботи *Анотації (на двох мовах); Вступ; Огляд літературних джерел; Теоретична частина; Практична частина; Загальні висновки; Список використаних джерел.*
5. Перелік ілюстративного матеріалу 17 слайдів, 31 рисуноків.
6. Орієнтовний перелік публікацій — не заплановано
7. Консультанти розділів роботи

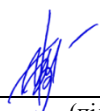
Розділ	Прізвище, ініціали та посада консультанта	Підпис, дата	
		завдання видав	завдання прийняв
Дипломної роботи	Третьякова Л. Д. д.т.н. проф. каф БМК	10.04.23	05.06.2023

7. Дата видачі завдання ***30 травня 2023 року.***

Календарний план

№ з/п	Назва етапів виконання дипломної роботи	Термін виконання етапів роботи	Примітка
1	Отримати завдання за темою ДР на практику	До 15.02.2023р.	виконано
2	Переддипломна практика	За графіком	виконано
3	Виконання розділів ДР (Вступ, аналітичний огляд літературних джерел, теоретична частина)	До кінця практики	виконано
4	Виконання розділів ДР (практична частина, загальні висновки, список джерел)	Не пізніше 1 тижня до засідання каф-ри	виконано
5	Перевірка ДР науковим керівником	Не пізніше 1 тижня до засідання каф-ри	виконано
6	Подання в електронному вигляді ДР та анотації до неї на перевірку нормоконтролера та плагіат (UNICHECK).	---- « -----	виконано
7	Надання документів на засідання кафедри	За день до засідання	виконано
8	Предзахист ДР та допуск до захисту дисертації	Згідно плану каф.	виконано
9	Подання ДР рецензенту. Отримання рецензії.	До подання пакету документів до ЕК	виконано
10	Подання пакету документів по ДР та супровідних до неї документів до захисту в ЕК ¹	За 5 днів до дати захисту ДР за графіком	виконано
11	Захист ДР в ЕК		

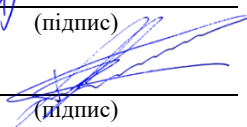
Студент


(підпис)

Милана ШЛАПАК

(ім'я, ПРІЗВИЩЕ)

Керівник ДР


(підпис)

Ілля ДАВИДОВИЧ

(ім'я, ПРІЗВИЩЕ)

Нормоконтролер

(підпис)

Галина КОРНІЄНКО

(ім'я, ПРІЗВИЩЕ)

¹ не пізніше ніж за 5 днів до затвердженої дати захисту ДР в ЕК

АНОТАЦІЯ

Дипломна робота за темою «Інтерактивний додаток до прогнозування захворюваності COVID-19 на основі даних про вакцинацію» виконана студентом кафедри біомедичної кібернетики ФБМІ Шлапак Миланою Ігорівною студентки зі спеціальності 122 «Комп'ютерні науки» за освітньо-професійною програмою «Комп'ютерні технології в біології та медицині» та складається зі: вступу; 3 розділів (аналіз літератури, матеріали та методи, практична частина), висновків до кожного з цих розділів; загальних висновків; списку використаних джерел, який налічує 23 джерела. Загальний обсяг роботи 65 сторінки.

Актуальність теми. Актуальність теми "Інтерактивний додаток для прогнозування COVID-19 на основі даних про вакцинацію" полягає в тому, що вона пропонує спосіб зробити прогнозування COVID-19 більш доступним та для користувачів, а також прогнозує з урахуванням даних щодо вакцинації.

Крім того, інтерактивний додаток можна використовувати для вивчення різних сценаріїв прогнозування і для того, щоб побачити, як зміни в даних можуть вплинути на прогнози.

Мета і завдання роботи.

Метою роботи є розробка інтерактивного додатку, який здатний прогнозувати захворюваність на COVID-19 з урахуванням даних щодо вакцинації. Основна ціль полягає у створенні інструменту, який надасть користувачам зручну та доступну інформацію про тенденції поширення хвороби.

Її досягнення передбачає вирішення наступних завдань:

1. Аналіз вітчизняних та зарубіжних джерел.
2. Розробити модель прогнозування.
3. Реалізувати прогноз захворюваності COVID-19 та створити інтерфейс.

Використані методи. Для реалізації прогнозу було використано мову програмування Python, бібліотеки, Pandas, Matplotlib, Seaborn. Користувацький інтерфейс з використанням фреймворку PyQt5.

Отримані результати. Інтерактивний додаток до прогнозування COVID-19 на основі даних про вакцинацію з використанням поліноміальної регресії.

Ключові слова. Захворюваність, COVID-19, вакцинація, моделі прогнозування, аналіз.

Бібліографічний опис ДР

Шлапак М. І. Інтерактивний додаток до прогнозування COVID-19 на основі даних про вакцинацію : дипломна роб. бакалавра : 122 Комп'ютері науки / Шлапак Милана Ігорівна. – Київ, 2023. – 65 с.

ABSTRACT

The thesis on "Interactive application for forecasting the incidence of COVID-19 based on vaccination data" was written by a student of the Department of Biomedical Cybernetics of FBMI *Shlapak Milana Ihorivna*, a student majoring in 122 "Computer Science" under the educational and professional programme "Computer Technologies in Biology and Medicine" and consists of introduction; 3 chapters (*literature review, materials and methods, practical part*), conclusions to each of these chapters; general conclusions; list of references, which includes 23 sources. The total volume of the paper is 65 pages.

Relevance of the topic. The relevance of the topic "Interactive application for COVID-19 forecasting based on vaccination data" is that it offers a way to make COVID-19 forecasting more accessible and user-friendly, and also forecasts based on vaccination data.

In addition, the interactive application can be used to explore different forecasting scenarios and see how changes in the data can affect the forecasts.

Purpose and objectives of the work.

The aim of this work is to develop an interactive application that can predict the incidence of COVID-19 based on vaccination data. The main goal is to create a tool that will provide users with convenient and accessible information about the trends in the spread of the disease.

Achieving it involves solving the following *tasks*:

1. Analysis of domestic and foreign sources.
2. Develop a forecasting model.
3. Implement a forecast of COVID-19 incidence and create an interface.

Methods used. To implement the forecast, the Python programming language, libraries, Pandas, Matplotlib, Seaborn were used. The user interface is based on the PyQt5 framework.

Results. An interactive application for forecasting COVID-19 based on vaccination data using polynomial regression.

Keywords. Morbidity, COVID-19, vaccination, forecasting models, analysis.

ЗМІСТ

ВСТУП.....	9
РОЗДІЛ 1 АНАЛІЗ ЛІТЕРАТУРИ.....	10
1.1. Пандемія COVID-19	10
1.2. Вакцинація	13
1.3. Проблематика прогнозування	14
Висновки до розділу 1	15
РОЗДІЛ 2 МАТЕРІАЛИ ТА МЕТОДИ	17
2.1 Методи прогнозування	19
2.2 Лінійна регресія	20
2.3 Поліноміальна регресія	25
Висновок до розділу 2	28
РОЗДІЛ 3 ПРАКТИЧНА ЧАСТИНА	29
3.1. Підготовка даних та аналіз датасету.....	29
3.2. Побудова моделі. Лінійна регресія	42
3.3. Побудова моделі. Поліноміальна регресія.....	43
3.4. Інтерфейс інтерактивного додатку.....	47
3.5. Безпека життєдіяльності та охорона праці	55
3.6. Розрахунок економічного ефекту за темою ДР	57
Висновки до розділу 3	58
ЗАГАЛЬНІ ВИСНОВКИ	60
СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ.....	62

ВСТУП

Завдання створити інтерактивний додаток до прогнозування захворюваності COVID-19 на основі даних про вакцинацію є дуже актуальним в сучасному світі, оскільки пандемія COVID-19 мала і має вплив на глобальне здоров'я та економіку.

Метами досягнення

Мета і завдання роботи

Метою роботи є розробка інтерактивного додатку, який здатний прогнозувати захворюваність на COVID-19 з урахуванням даних щодо вакцинації. Основна ціль полягає у створенні інструменту, який надасть користувачам зручну та доступну інформацію про тенденції поширення хвороби.

Її досягнення передбачає вирішення наступних завдань:

Аналіз вітчизняних та зарубіжних джерел.

Розробити модель прогнозування.

Реалізувати прогноз захворюваності COVID-19 та створити інтерфейс.

Використані методи. Для реалізації прогнозу було використано мову програмування Python, бібліотеки, Pandas, Matplotlib, Seaborn. Користувацький інтерфейс з використанням фреймворку PyQt5.

Отримані результати. Інтерактивний додаток до прогнозування COVID-19 на основі даних про вакцинацію з використанням поліноміальної регресії.

Структура роботи

Дипломна робота за темою «Інтерактивний додаток до прогнозування захворюваності COVID-19 на основі даних про вакцинацію» виконана студентом *Шлапак Миланою Ігорівною студенткою зі спеціальності 122 «Комп'ютерні науки» за освітньо-професійною програмою «Комп'ютерні технології в біології та медицині»*, побудована за класичним типом та викладена на 64 сторінках машинописного тексту. Вона складається з: вступу; 3 розділів (*аналіз літератури, матеріали та методи, практична частина*), висновків до кожного з цих розділів; загальних висновків; списку використаних джерел, який налічує 23 джерел та додатків (5 – на кирилиці, 18 – на латиниці). В роботі представлено 31 рисунок.

РОЗДІЛ 1

АНАЛІЗ ЛІТЕРАТУРИ

1.1. Пандемія COVID-19

На початку 2020 року весь світ сколихнула пандемія коронавірусу. Захворювання стрімко ширилось країнами світу та спричиняло паніку населення. А головним питанням було коли віднайдуть вакцину та чи буде вона ефективною.

Для початку вченим потрібно було розібратись з етимологією захворювання. Коронавірус – вірусне захворювання, яке проявляється такими симптомами як температуру тіла, кашель, проблеми з диханням, головний біль, втрата смаку та нюху, проблеми з шлунково-кишковим трактом. [5]

Першим випадком захворювання на COVID-19 в Україні є випадок, який зафіксували 3 березня 2020 року [4]. Проте перший спалах захворюваності вірусом був зафіксований в Ухані, Китай, в грудні 2019 [10]. Спочатку 2020 року ВООЗ оголосила спалах захворюваності – це надзвичайна ситуація в галузі суспільної охорони здоров'я, а вже 11 березня 2020 року оголосила пандемію. [2]

За інформацією Всесвітньої організації охорони здоров'я (WHO), станом на 10 травня 2023 було зафіксовано більш ніж 750 мільйонів випадків захворювань на COVID-19 та майже 7 мільйонів смертей від цього вірусу по всьому світу. [22]

Нижче наведені графіки захворюваності COVID-19, за даними сервісу Our World in Data, в різних країнах на різних континентах за весь період пандемії. Графіки показують, як кількість заражень змінювалася з часом, починаючи з перших випадків, що було зафіксовано в країнах світу в січні 2020 року, і до кінця пандемії (10.05.2023) [6].

На графіку нижче (рис. 1.1) зображена зареєстрована кількість випадків захворювання в США. Чітко видно, що найбільший пік приходить на 12 січня 2020 року з 3740 випадками за добу на 1 мільйон населення. Цей пік відображає надзвичайно високу інцидентність хвороби у вказаний період часу. Дані вказують

на серйозну загрозу здоров'ю та громадському здоров'ю, вимагають негайних заходів для контролю та зменшення поширення захворювання.

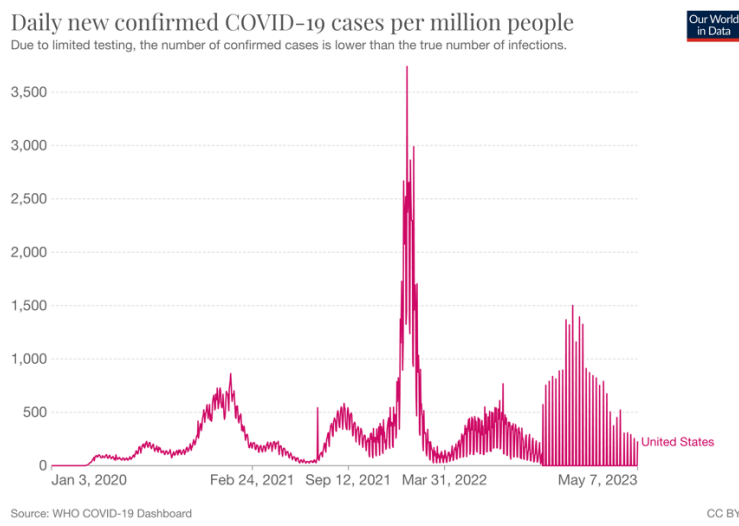


Рисунок 1.1 – Графік захворюваності COVID-19 в США [6]

Згідно з відомостями на представленому графіку (рис. 1.2), найвищий рівень захворюваності в Японії був зафіксований 28 серпня 2022 року, коли кількість випадків сягнула більше 2,5 тисяч за добу на один мільйон населення. Цей високий показник є свідченням широкого поширення хвороби та серйозної загрози для здоров'я місцевої спільноти.

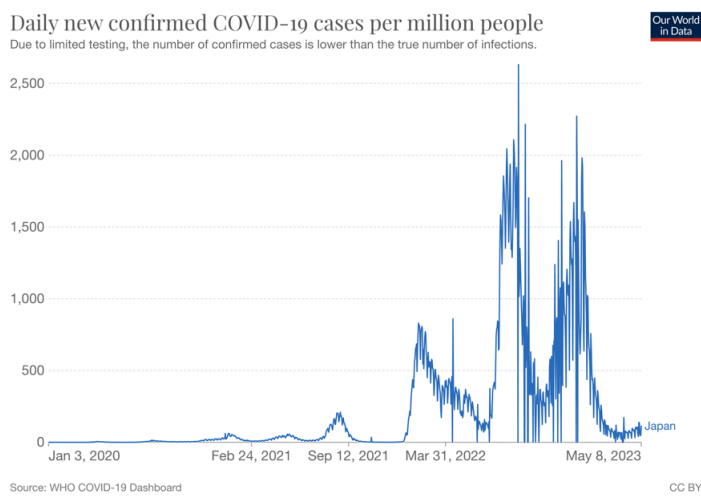


Рисунок 1.2 – Графік захворюваності COVID-19 в Японії [6]

Серед усіх дат, найвища кількість зафіксованих випадків захворювань на COVID-19 в Бразилії була виявлена 5 лютого 2022 року (рис. 1.3). У цей день, статистика показала дуже високий рівень захворюваності, який становив 1 385 випадків за добу на мільйон населення. Цей показник є дзеркалом загальної поширеності вірусу та серйозністю його впливу на здоров'я громади.

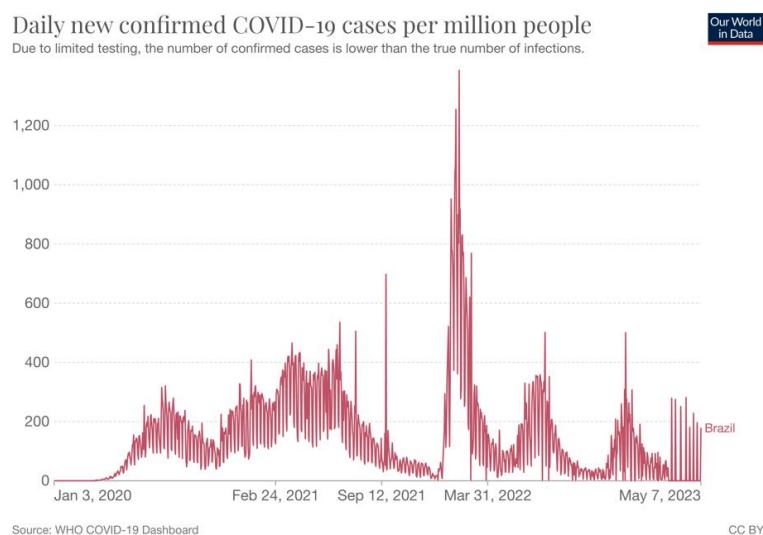


Рисунок 1.3 – Графік захворюваності COVID-19 в Бразилії [6]

Згідно з інформацією на графіку (рис. 1.4), станом на 24 березня 2022 року в Німеччині було зафіксовано вражаючу кількість випадків - 3 693 захворювання на добу на мільйон населення.

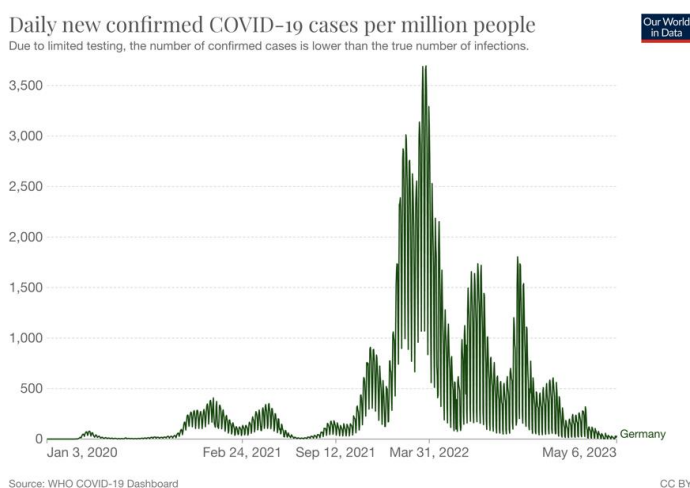


Рисунок 1.4 – Графік захворюваності COVID-19 в Німеччині [6]

Аналізуючи попередні графіки, можна зробити висновок, що найвищий рівень захворюваності на COVID-19 поступово переносився від заходу до сходу протягом року. Це свідчить про поширення хвороби по всій планеті і відображає динаміку її впливу на різні регіони.

Дивлячись на дані від Our World in Data про захворюваність на COVID-19 в різних країнах світу, можна зробити кілька висновків. Найбільша кількість випадків захворювання виявляється в країнах з великою кількістю населення, таких як США та Бразилія, але це не завжди так. Наприклад, Японія має значно меншу кількість населення, але все ж стала свідком значного піку випадків захворювання в 2022 році. Але важливішим є те, що в країнах було вжито заходів під час боротьби з поширенням хвороби (а саме вакцинація), які дали результати та призвели до зменшення кількості захворювань. Отже важливим пунктом для дослідження є саме вакцинація.

1.2. Вакцинація

Зважаючи на дані графіків (рис. 1.1-1.4), що наведено вище, однозначно людство потребувало рішень. Тож одразу вчені почали працювати над вакциною, проте основною проблемою було те, що певна кількість населення не сприймала вакцину серйозно.

Згідно з дослідженням, опублікованим в журналі Nature Communications, у 2021 році сприйняття вакцини від COVID-19 було нерівномірним у світі. Найбільше прийняття вакцинації спостерігалось в Китаї (85,6%), а найменша - у Франції (40,3%). [12]

Такі дані вказують на те, що потрібно вжити заходів, аби довести більшості населення важливість вакцини та її користь.

Інша стаття від Nature Communications наголошує на тому, що головними причинами відмовлення від вакцинації є недостатня інформація та спекуляції щодо ефективності та безпеки вакцин. [18]

Цю інформацію підтверджує інша стаття, автором якої є Зоя Гадецька. Згідно з її дослідженням, рівень вакцинації населення є важливим фактором, який впливає на рівень захворювання. Важливим пунктом є те, що розвинені країни фіксують значно більшу кількість вакцинованих, ніж країни з перехідною економікою [1]. Саме це свідчить, що країни з гіршим економічним становищем не мають змоги донести до своїх громадян важливість вакцинації та її вплив на перебіг пандемії.

Отже, дані вказують на необхідність створення інтерактивного додатку, що містить інформацію про вакцинацію, її вплив на поширення захворюваності. Тому важливою задачею додатку є поширення достовірної інформації про вакцини та допомога користувачам у прийнятті рішення про їх вакцинацію.

1.3. Проблематика прогнозування

З початком пандемії велика кількість людей почала працювати з прогнозуванням захворюваності. Зрозуміло, що вибір моделі прогнозування залежав від того, які дані є та які чинники будуть розглядатись.

Згідно зі статтею на Springer Link, автори підкреслюють важливість використання математичних моделей в прогнозуванні пандемічних хвороб таких як COVID-19. Вони допомагають зрозуміти тенденції та прогнозувати майбутні розвиток ситуації, що допомагає приймати обґрунтовані рішення та розробляти ефективні стратегії в управлінні та протидії пандемії COVID-19 [11].

Стаття The Institution of Engineering and Technology [23] наголошує на складності прогнозування тенденцій COVID-19 і на тому, що жодна модель не може забезпечити абсолютну точність у передбаченнях майбутніх подій. Однак, використання різних моделей та аналіз їх прогнозів може допомогти зробити більш точні прогнози тенденцій COVID-19 у майбутньому.

У даному дослідженні було розглянуто прогнозування тенденцій розвитку COVID-19 в Канаді, використовуючи чотири різні моделі, але зупинитись хочеться на трьох: логістичній моделі, моделі Гауса та моделі SIR.

На практиці, логістичні, гауссівські та SIR моделі, які використовуються в даному дослідженні, мають свої переваги та обмеження. Логістична та гауссівська моделі використовують математичні рівняння для прогнозування епідемічних трендів, і їхні прогнози базуються на виборі лише декількох гіпер-параметрів, що зменшує ступінь невизначеності у моделюванні. Однак, ці моделі є досить простими і не можуть адекватно описати складні епідемічні траєкторії та врахувати всі фактори, які впливають на поширення хвороби.

Стосовно моделі SIR, в статті зазначено, що вона є однією з найпростіших динамічних моделей прогнозування грипу, яка може прямо відображати спалахи епідемії. Однак, використання цієї моделі для прогнозування тенденцій є складним, оскільки результати моделі дуже чутливі до попередньо визначених гіпер-параметрів. Тобто, навіть незначні зміни в параметрах можуть призвести до двох різних траєкторій прогнозу [23].

Також, розглядаються методи прогнозування за регресії [16]. Більш детально буде розібрано методи прогнозування для даної роботи в наступному розділі.

Висновки до розділу 1

В результаті аналізу низки літературних джерел, створених відомими фахівцями в галузі епідеміології, отримано такі висновки:

- Захворювання на COVID-19 швидко ширилось в усіх країнах світу, призводячи до великої кількості смертельних випадків, тривалої хвороби серед всіх соціальних груп, економічного спаду та інших негативних наслідків.
- Вакцинація виявилася ефективним засобом обмеження поширення вірусу COVID-19, навіть при початковому супротиві з боку деякої частини населення.
- Прогнозування темпів поширення захворювання та можливих наслідків відіграє важливу роль у наданні інформації та визначенні

стратегій для обмеження подальшого поширення небезпечних інфекцій.

Завдяки цим висновкам, було отримано високоякісну інформацію, що розкриває ключові аспекти цієї глобальної проблеми.

РОЗДІЛ 2

МАТЕРІАЛИ ТА МЕТОДИ

Прогнозування стану здоров'я є новим напрямом прогнозування, що виявляється корисним інструментом для передбачення майбутніх подій та ситуацій у галузі охорони здоров'я. Це може включати передбачення попиту на медичні послуги та потреб у медичній допомозі. Прогнозування сприяє профілактичній медицині та розробці стратегій втручання в систему охорони здоров'я, надаючи постачальникам медичних послуг попередню інформацію для вжиття відповідних заходів щодо мінімізації ризиків та управління попитом.

Прогнозування в галузі охорони здоров'я потребує надійних даних, інформації та аналітичних інструментів для передбачення конкретних станів або ситуацій у сфері охорони здоров'я. Оскільки немає універсального підходу до прогнозування стану здоров'я, часто використовуються різні методи для прогнозування загальних або специфічних станів здоров'я. Крім того, не існує чітко визначених часових рамок прогнозування здоров'я, які відповідають вибору методів або підходів, що часто використовуються для прогнозування здоров'я. [21]

Також важливим для прогнозування захворюваності чи стану здоров'я важливо орієнтуватись на схему наведену нижче (рис. 2.1). Ці етапи допомагають визначити вимоги та інструменти прогнозування стану здоров'я в широкому розумінні. Крім того, вони встановлюють основні процеси, пов'язані з розробкою та вдосконаленням стратегії прогнозування змін у стані здоров'я з плином часу.

В статті ВМС [13] наводять покрокову структуру, що демонструє динамічний процес, у якому моделі прогнозування, розроблені на будь-якому етапі, постійно вдосконалюються з метою досягнення прогнозованої мети або задоволення потреб клієнта.

Крок 1: Потрібно уточнити концепції та ідеї, пов'язані із проблемними станами здоров'я, які мають великий вплив на систему охорони здоров'я. Представити чітку специфікацію прогнозованого результату стану здоров'я і чітко визначити період, на який розрахований прогноз.

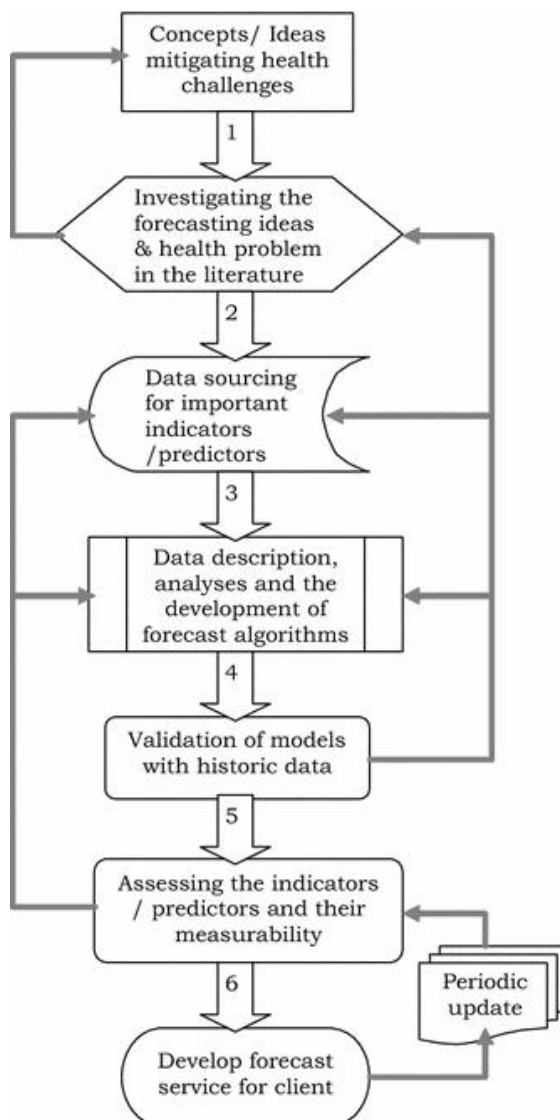


Рисунок 2.1 – Схема концептуальної основи для прогнозування [21]

Крок 2:

Варто скористатись літературою, щоб визначити причинні або сильно корельовані змінні, які пов'язані з визначеними показниками стану здоров'я в Кроці 1 (може знадобитися консультація з експертом для збагачення знань в цій галузі);

Крок 3:

Визначити джерела даних як для показників стану здоров'я (Крок 1), так і для всіх потенційних прогностичних змінних та переконатись в наявності та повноті даних (тобто перевірити наявність прогалів у часових рядах даних).

Крок 4:

Підготувати набори даних для базового статистичного аналізу, включаючи описову статистику та розробку прогностичних алгоритмів. Цей крок включає

очищення та керування даними, а також генерацію додаткових змінних для подальшого аналізу.

Крок 5:

Створити прогностичні моделі та їх перевірити на основі різних наборів подібних історичних даних.

Крок 6:

Оцінити та визначити остаточний перелік індикаторів, необхідних для якісної прогностичної моделі (моделей), на підставі практичного доступу до їхніх показників.

Крок 7:

Розробити високоспеціалізовані та індивідуальні прогностичні послуги для конкретної мети або клієнта, а також періодичне оновлення моделей.

Тому для досягнення мети прогнозування захворюваності на COVID-19 та вакцинації в різних країнах, потрібно ґрунтовно дослідити та оцінити різні методології, які призначені саме для цієї цілі.

2.1 Методи прогнозування

Існує кілька методів, які можна використовувати для прогнозування захворюваності COVID-19. Основні з них включають:

1. Статистичні моделі: Ці моделі базуються на аналізі статистичних даних про поширення вірусу та захворюваності. Можуть використовуватися різні методи, включаючи лінійну регресію, поліноміальну регресію, експоненційні моделі та ARIMA (авторегресійні інтегровані зміщені ковзні середні) моделі [17].
2. Математичні моделі передачі хвороб: Ці моделі використовують математичні рівняння для опису процесу передачі хвороби в популяції. Найбільш відомою моделлю такого типу є модель SIR (Susceptible-Infectious-Recovered), яка враховує кількість осіб, які ще можуть

захворіти (схильні), інфіковані особи та осіб, які одужали або отримали імунітет [9].

3. Машинне навчання: Машинне навчання використовує різноманітні методи для здобуття знань та навичок у комп'ютерних системах. Серед цих методів є нейронні мережі, дерева рішень, метод опорних векторів (SVM) та інші алгоритми машинного навчання. Вони можуть бути використані для прогнозування захворюваності COVID-19 на основі великої кількості вхідних даних, включаючи соціально-економічні показники, демографічні дані, показники медичної системи тощо.
4. Ансамблеві методи: Ці методи поєднують декілька моделей разом для отримання більш точних та надійних прогнозів. Наприклад, можна застосовувати ансамблеві методи, такі як градієнтний бустинг (Gradient Boosting) або випадковий ліс (Random Forest), для поєднання результатів декількох моделей лінійної регресії, поліноміальної регресії.

Для цієї роботи було обрано перший пункт – статистичні моделі, а саме лінійну та поліноміальну регресії. Адже лінійна регресія є одним з найпростіших та найпоширеніших методів прогнозування, в той час як поліноміальна регресія розширює можливості лінійної регресії, дозволяючи моделювати нелінійні залежності між змінними.

2.2 Лінійна регресія

Лінійна регресія є статистичним інструментом, який використовується для прогнозування майбутніх значень на основі попередніх даних [13]. У випадку прогнозування, лінійна регресія допомагає знайти найкращу лінійну функцію, яка найкраще відповідає вхідним даним і може бути використана для передбачення майбутніх значень.

На графіку (рис. 2.2) зображений приклад простої лінійної регресії, де є лише одна незалежна змінна.

Формально, модель лінійної регресії виражається у вигляді:

$$y = b_0 + b_1x_1 + b_2x_2 + \dots + b_nx_n, \quad (2.1)$$

де:

- y - цільова змінна, яку потрібно прогнозувати,
- $b_0, b_1, b_2, \dots, b_n$ - коефіцієнти моделі, які визначають вагу кожної вхідної змінної,
- $x_1, x_2, \dots, x_n$ - вхідні змінні, які використовуються для прогнозування [3].

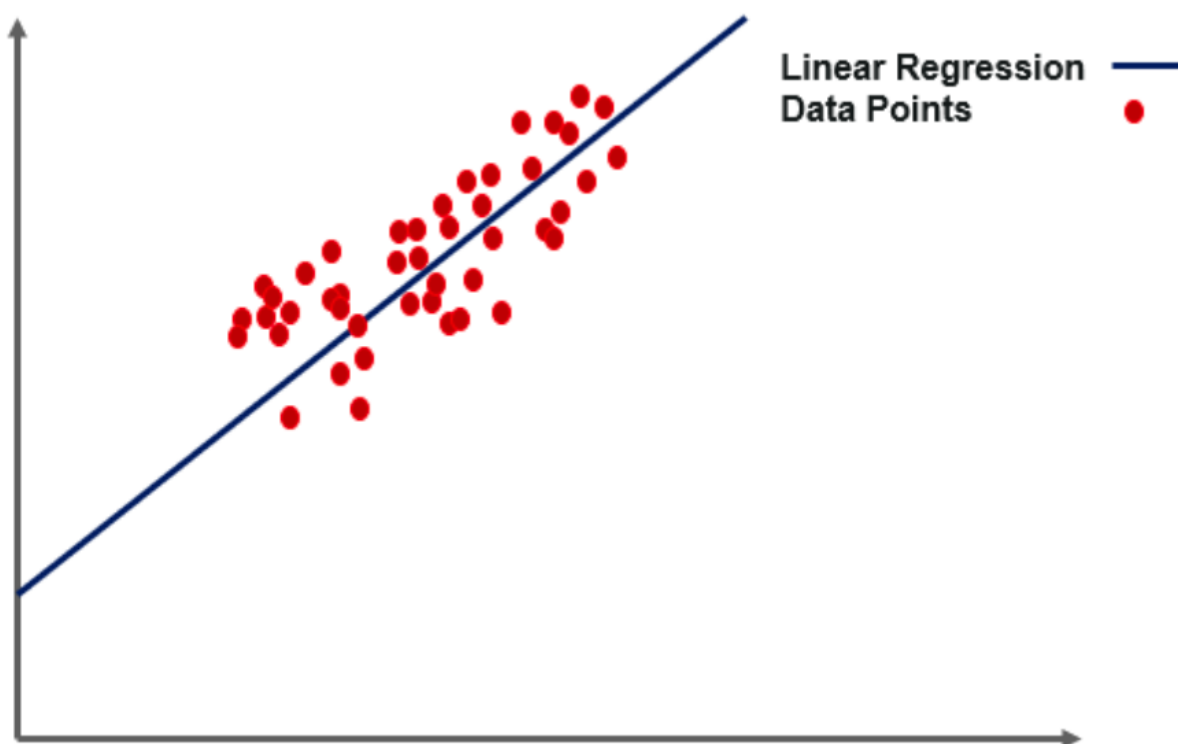


Рисунок 2.2 – Лінійна регресія [14]

Мета лінійної регресії полягає в знаходженні оптимальних значень коефіцієнтів $b_0, b_1, b_2, \dots, b_n$, що мінімізують суму квадратів відхилень між прогнозованими значеннями і фактичними значеннями цільової змінної. Цей

процес називається навчанням моделі і використовується для побудови лінійної функції, яка найкраще підходить до даних.

Тепер, що стосується прогнозування захворюваності COVID-19 за допомогою лінійної регресії, можна використовувати дані про різні фактори, які можуть впливати на поширення вірусу або його вплив на населення. Наприклад, такими факторами можуть бути:

- Демографічні дані: Розмір населення може впливати на швидкість поширення вірусу. Більше населення може сприяти швидкому поширенню COVID-19.
- Густота населення: Регіони з високою густотою населення можуть бути більш схильними до швидкого поширення вірусу через близькість контактів між людьми.
- Соціальні заходи: Введення обмежень, таких як карантин, закриття шкіл та підприємств, може впливати на поширення вірусу і його негативний вплив.
- Рівень гігієни: Санітарні умови, такі як доступ до води, миючих засобів та засобів особистої гігієни, можуть впливати на поширення вірусу.
- Медична інфраструктура: Наявність достатньої кількості лікарень, лікарів, медичного обладнання та інших ресурсів може визначати здатність системи охорони здоров'я ефективно боротися зі спалахом COVID-19.
- Вакцинація: Рівень вакцинації в популяції може впливати на поширення вірусу. Високе покриття вакцинації може знизити ризик захворювання та сприяти контролю епідемії COVID-19. Ефективна вакцинаційна кампанія може допомогти знизити швидкість поширення вірусу та його небажані наслідки, такі як важкі випадки захворювання та смертельні випадки.

Для прогнозування захворюваності COVID-19 можна зібрати дані про ці фактори для різних регіонів або країн і використовувати лінійну регресію, щоб побудувати модель, яка оцінює вплив цих факторів на поширення вірусу і прогнозує

кількість захворювань у майбутньому. Прогнози, отримані за допомогою цієї моделі лінійної регресії, можуть допомогти в оцінці тенденцій поширення COVID-19 та допомогти в прийнятті рішень щодо здійснення необхідних заходів для контролю та управління пандемією.

Важливо врахувати, що лінійна регресія має як свої плюси, так і свої недоліки у вигляді деяких обмежень.

Переваги лінійної регресії:

- лінійна регресія демонструє високу ефективність при моделюванні даних, де взаємозв'язок між змінними може бути виражений лінійно;
- цей метод є легким у впровадженні та інтерпретації, і його можна ефективно навчити на великих наборах даних;
- лінійна регресія також проявляє добрі властивості щодо управління перенавчанням, завдяки методам зменшення розмірності, регуляризації та перехресної перевірки;
- крім того, вона має можливість екстраполювати результати за межі конкретного набору даних, що є ще однією перевагою лінійної регресії.

[14]

Недоліки лінійної регресії:

- лінійна регресія передбачає лінійну залежність між вхідними змінними та цільовою змінною, що може не враховувати складніші нелінійні взаємодії між змінними;
- лінійна регресія може бути чутлива до викидів або незвичайних даних, що можуть вплинути на точність моделі, великі викиди можуть негативно впливати на оцінки коефіцієнтів та зміщати прогнози;
- лінійна регресія може стикатися з проблемою мультиколінеарності, коли дві або більше змінних взаємно залежні або мають високу кореляцію; саме це може призвести до нестабільних та недостовірних оцінок коефіцієнтів [14];
- невідповідність – стан, який виникає, коли модель машинного навчання не може адекватно відтворити надані їй дані, зазвичай це відбувається,

коли гіпотеза моделі не може належним чином відповісти на дані, в результаті чого виникає розбіжність між прогнозованими значеннями та фактичними даними [20]. Приклад такого випадку наведений на графіку нижче (рис. 2.3).

У більшості реальних сценаріїв зв'язок між змінними набору даних не є лінійним, а отже, пряма лінія не відповідає даним належним чином. У таких ситуаціях більш складна функція може більш ефективно врахувати дані. Через це більшість лінійних регресійних моделей мають низьку точність.

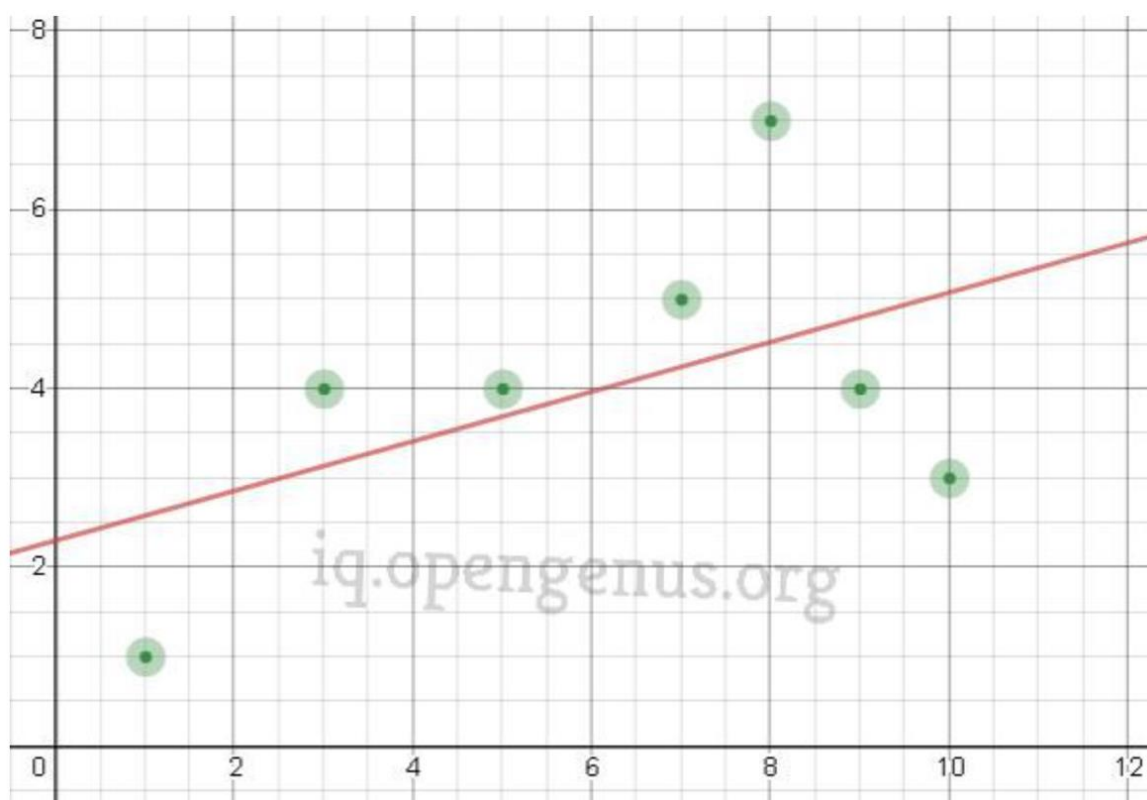


Рисунок 2.3 – Приклад невідповідності [20]

Саме тому у прогнозуванні захворюваності COVID-19, на додаток до лінійної регресії, можуть бути застосовані інші більш складні моделі, такі як поліноміальна регресія, дерева рішень, нейронні мережі або методи машинного навчання, які здатні врахувати нелінійні залежності та взаємодію між факторами.

2.3 Поліноміальна регресія

Поліноміальна регресія – це метод регресійного аналізу, який використовується для моделювання зв'язку між залежною (y) та незалежною змінною (x) у формі полінома n -го степеня. Рівняння поліноміальної регресії представлено нижче:

$$y = b_0 + b_1x_1 + b_2x_2^2 + \dots + b_nx_n^n \quad (2.2)$$

У цьому рівнянні $b_0, b_1, b_2, \dots, b_n$, позначають коефіцієнти регресії, які визначають форму та похил розподілу поліноміальної кривої. Чим вищий степінь полінома (n), тим більше гнучкість в моделюванні нелінійних залежностей між змінними [15]. На графіку (рис.2.4), що знаходиться нижче, зображений приклад зображення поліноміальної регресії.

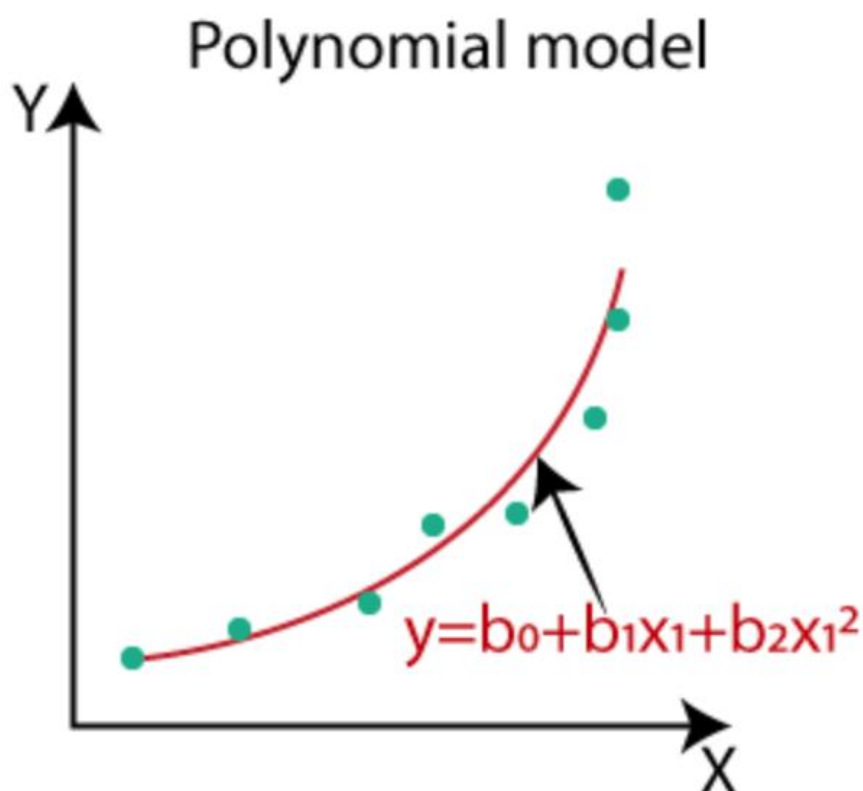


Рисунок 2.4 – Приклад зображення поліноміальної регресії [19]

Поліноміальна регресія дозволяє адаптуватися до більш складних форм функцій, що описують дані, а не обмежується простою лінійною залежністю. Цей метод дозволяє підібрати поліном, який найкраще апроксимує дані і дозволяє виконувати прогнози на основі цієї моделі.

Для побудови моделі поліноміальної регресії використовуються поліноми різних ступенів, наприклад, квадратичні (ступінь 2), кубічні (ступінь 3) і т. д. Після побудови моделі можна провести аналіз коефіцієнтів поліномів, щоб зрозуміти, які змінні мають найбільший вплив на прогнозування. Важливо врахувати, що використання поліноміальної регресії для задачі дипломної роботи прогнозування захворюваності COVID-19 має свої обмеження і потребує обережності при інтерпретації результатів.

При використанні поліноміальної регресії в даній задачі прогнозування захворюваності COVID-19, незалежними змінними можуть бути дата або час, а залежна змінна може бути кількість нових випадків COVID-19 наступного дня.

Також важливо врахувати, що використання поліноміальної регресії в задачі прогнозування захворюваності COVID-19 має свої переваги та обмеження, а також потребує обережності при інтерпретації результатів.

До переваг належить:

- гнучкість: поліноміальна регресія досить гнучка, щоб пристосовуватися до широкого діапазону кривизни. Вона може адаптуватися до нелінійних залежностей між змінними, які не можуть бути відображені за допомогою простої лінійної моделі. За допомогою поліноміальної регресії можна врахувати кривизну та вигини у залежності та точніше відтворити поведінку даних;
- вибір функцій: поліноміальна регресія може легко підходити під широкий спектр функцій. Вона дозволяє використовувати поліноми різних ступенів (наприклад, квадратичні, кубічні і т.д.) для моделювання залежності. Це дає можливість більш точно апроксимувати дані та краще відтворювати їх поведінку.

- наближення зв'язку: поліноміальна регресія пропонує найкраще наближення зв'язку між двома залежними та незалежними змінними. Оскільки поліноміальна регресія може моделювати складніші форми функцій, вона здатна точніше відтворити залежності між змінними. Це дозволяє зробити більш точні прогнози та аналізувати взаємозв'язки між змінними з більшою достовірністю. [19]

Недоліки поліноміальної регресії:

- існує ризик перенавчання при використанні поліноміальної регресії, коли модель добре адаптується до навчальних даних, але погано узагальнюється на нові дані, це може призвести до складності моделі та неправильних прогнозів;
- вибір ступеня полінома має велике значення: занадто низький ступінь може пропустити складні нелінійні залежності, тоді як занадто високий ступінь може призвести до перенавчання. Важливо збалансувати складність моделі та її здатність до узагальнення;
- важлива роль належить валідації моделі – це дозволяє оцінити, наскільки добре модель узагальнюється на нові дані і чи можна їй довіряти для прогнозування захворюваності COVID-19. Для порівняння прогнозних значень з фактичними даними можна застосовувати різноманітні метрики, такі як середня квадратична помилка або коефіцієнт детермінації. Ці метрики дозволяють оцінити точність прогнозування шляхом вимірювання різниці між прогнозними значеннями та реальними даними.

При інтерпретації результатів поліноміальної регресії в задачі прогнозування захворюваності COVID-19 важливо враховувати контекст та характер даних. Додаткові змінні, такі як вакцинація, ефективність обмежувальних заходів та медичні показники, можуть впливати на результати моделі. Крім того, прогнозування захворюваності COVID-19 є складним завданням через багатофакторний характер пандемії та її динамічний характер. Для досягнення надійних результатів застосування поліноміальної регресії в задачі прогнозування

захворюваності COVID-19, необхідно уважно аналізувати дані і враховувати взаємодію між різними факторами, такими як епідеміологічні показники, обмежувальні заходи та медичні інтервенції.

Висновок до розділу 2

У даному розділі було розглянуто методи прогнозування, зокрема лінійну регресію та поліноміальну регресію, їх переваги та недоліки. Враховуючи отримані результати, можна зробити наступні висновки:

Лінійна регресія є потужним інструментом для моделювання залежностей між змінними, особливо в тих випадках, коли залежність між ними може бути описана лінійною функцією. Вона має просту інтерпретацію та широке застосування. Проте, вона не може вловити складні нелінійні залежності, що можуть присутні у деяких даних. Це обмеження може бути подолане за допомогою поліноміальної регресії.

Поліноміальна регресія дозволяє моделювати нелінійні залежності та використовувати поліноміальні функції для опису даних. Вона є гнучким методом, який може адаптуватися до різноманітних форм залежностей. Проте, важливо врахувати ризик перенавчання, особливо при використанні великої кількості ступенів полінома. Також важливо збалансувати складність моделі та її здатність до узагальнення шляхом вибору оптимального ступеня полінома.

Для обох методів важливо проводити валідацію моделі на незалежних тестових даних для оцінки їх ефективності та здатності до узагальнення. Це допомагає переконатися, наскільки добре модель працює на нових даних і чи можна їй довіряти для прогнозування.

Завдяки цим висновкам, було отримано високоякісну інформацію, яка допоможе при практичному виконанні дипломної роботи.

РОЗДІЛ 3

ПРАКТИЧНА ЧАСТИНА

Для виконання дипломної роботи був використаний датасет, який було отримано з платформи Kaggle. Цей датасет містить значну кількість інформації про захворюваність на COVID-19, вакцинацію, дати та смертність[8]. Загалом, в датасеті міститься 67 стовпців і 166 325 рядків, що свідчить про його обсяг і розгорнутість. Датасет охоплює період часу з 24 лютого 2020 року по 5 березня 2022 року, що дає можливість аналізувати дані за проміжок понад два роки. Використання такого широкого діапазону дат дозволяє вивчати тенденції та патерни у розповсюдженні та контролі COVID-19 на тривалому періоді часу. Цей датасет є цінним ресурсом для проведення аналізу та розробки моделей, які використовуються в інтерактивному додатку для прогнозування захворюваності на COVID-19.

3.1. Підготовка даних та аналіз датасету

Перед початком аналізу та моделювання імпортуємо необхідні бібліотеки, які забезпечують нам функціональність та інструменти для роботи з даними та побудови моделей. Ось перелік бібліотек, які було використано:

- **Pandas:** Ця бібліотека надає потужні інструменти для маніпулювання та аналізу даних. Вона дозволяє ефективно завантажувати, обробляти та аналізувати дані з датасету.
- **NumPy:** Ця бібліотека є основним інструментом для наукових обчислень у мові програмування Python [24]. Вона надає нам високопродуктивні структури даних, такі як масиви, інструменти для роботи з цими масивами, а також математичні функції для виконання різноманітних операцій.

- Matplotlib: Ця бібліотека використовується для візуалізації даних. Вона дозволяє побудовувати різноманітні графіки, діаграми та інші візуальні представлення даних.
- Seaborn: Це високорівнева бібліотека для візуалізації даних, яка базується на Matplotlib. Вона надає простий та зручний інтерфейс для побудови складних графіків та статистичних візуалізацій.
- Ці бібліотеки дозволяють ефективно працювати з даними, виконувати обчислення та створювати інформативні візуалізації, що є важливими етапами аналітичної роботи та моделювання в рамках даної дипломної роботи.

Щоб розпочати аналіз та моделювання, було імпортовано датасет, який містить інформацію про захворюваність та вакцинацію COVID-19. Для цього було використано датасет під назвою "COVID-19 dataset", який був завантажений з платформи Kaggle. Однак, варто зазначити, що цей датасет також доступний на сайті "Our World in Data" [7].

Як говорилося вище, цей датасет містить обширну інформацію про кількість захворювань, вакцинацію та інші важливі показники, пов'язані з пандемією COVID-19. Завдяки цьому датасету є можливість досліджувати тенденції, виявляти зв'язки та робити прогнози, що є для написання даної дипломної роботи.

Наступним кроком після імпорту датасету було проведення первинного аналізу, що включало перевірку вмісту колонок. Під час цього етапу було оцінено значення кожної колонки в контексті конкретної задачі прогнозування захворюваності COVID-19.

Перевірка вмісту колонок була важливою частиною процесу аналізу датасету. Це означало, що кожна колонка була проглянута, щоб з'ясувати, яку саме інформацію вона містить і які дані можуть бути використані для прогнозування захворюваності COVID-19.

Колонки датасету можуть містити різноманітні дані, такі як країна, кількість захворювань, кількість смертей, тощо. Важливо було з'ясувати, які саме колонки

мають вагому значення для прогнозування захворюваності COVID-19 та які можуть бути ігноровані.

Було помічено, що в датасеті присутня велика кількість комірок, які не мали вагомого впливу на прогнозування захворюваності і, отже, не були необхідними для нашої задачі. В цих колонках були відсутні дані або ж вони містили інформацію, яка не була релевантною для даного дослідження.

Для того, аби точно виявити потрібні для дослідження колонки, було побудовано кореляційну матрицю для 67 колонок з датасету (рис. 3.1), що дозволило виявити залежності між різними змінними. На основі цієї матриці було прийнято рішення обрати колонки з найвищою кореляцією, щоб спростити аналіз та сконцентруватися на ключових аспектах.

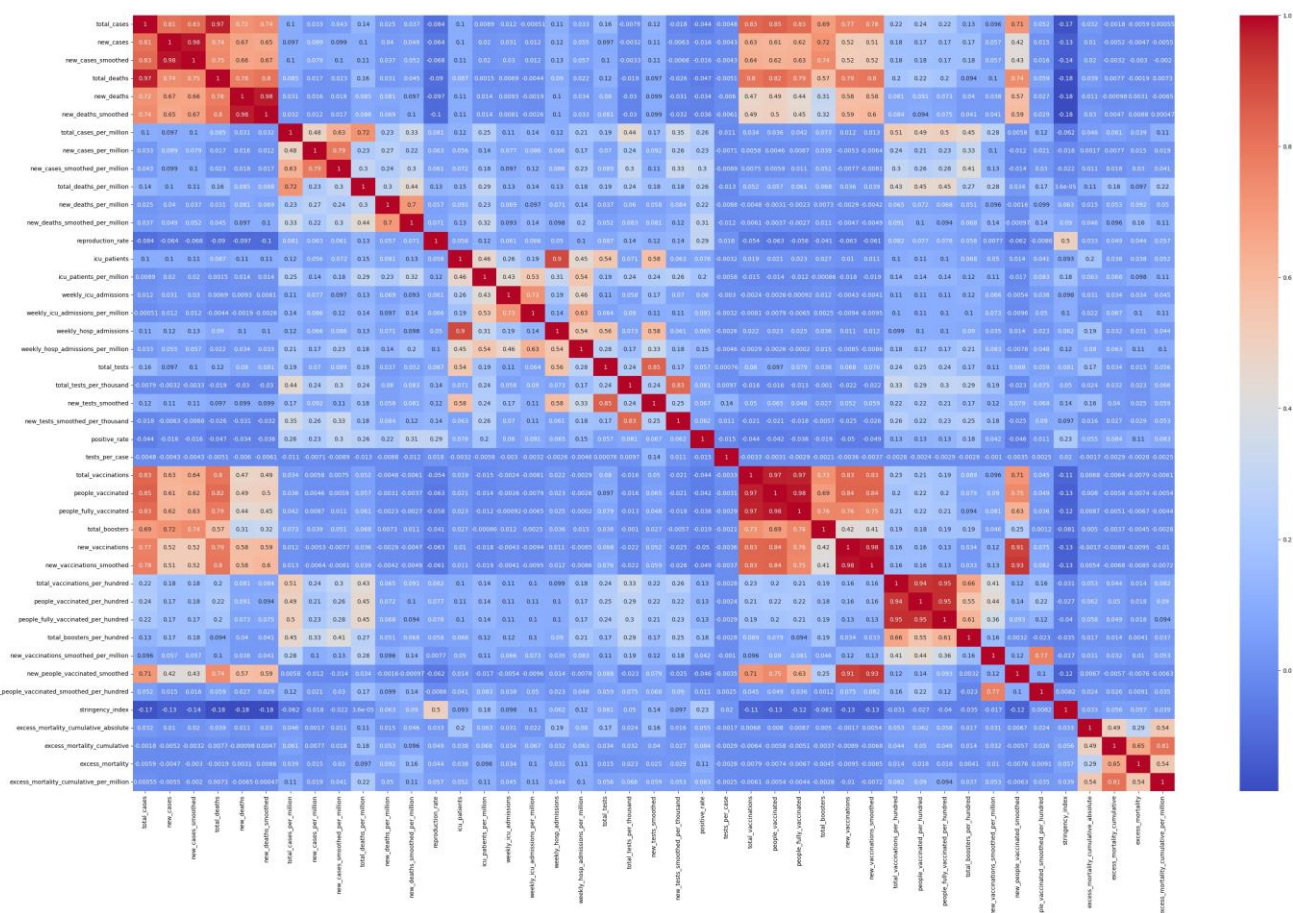


Рисунок 3.1 – Датасет після скорочення даних

Кореляційна матриця є потужним інструментом, який допомагає виявити статистичні зв'язки між різними змінними. Вона показує, наскільки сильно

залежать одна змінна від іншої, використовуючи кореляційний коефіцієнт. Коефіцієнт кореляції знаходиться в діапазоні від -1 до 1, де значення, які наближаються до 1, вказують на наявність позитивної кореляції, тоді як значення, що наближаються до -1, свідчать про наявність негативної кореляції. Значення, близькі до 0, вказують на відсутність кореляції між змінними.

Побудована кореляційна матриця була важливим кроком для визначення найважливіших колонок у датасеті. Шляхом оцінки значень кореляційних коефіцієнтів між кожною парою колонок було виявлено зв'язки, які допомогли вибрати колонки з найвищою кореляцією. Це дало змогу зосередитися на ключових аспектах та спростити подальший аналіз датасету.

Після проведення аналізу кореляційної матриці, було вибрано лише десять ключових колонок зі згаданих 67. Ці обрані колонки представляють значущі показники, такі як:

1. 'iso_code' – міжнародний унікальний ідентифікатор країни;
2. 'continent' – континент, на якому розташована країна;
3. 'location' – назва країни;
4. 'date' – дата, коли було зроблено запис;
5. 'total_cases' – кількість випадків захворювання;
6. 'new_cases' – нові випадки;
7. 'total_deaths' – кількість смертей,
8. 'new_deaths' – нові випадки смертей;
9. 'total_vaccinations' – кількість вакцинацій;
10. 'new_vaccinations' – кількість нових вакцинованих.

Вибір обмеженої кількості колонок забезпечив більш чистий та зрозумілий датасет, який був готовий до подальшого аналізу та використання для побудови моделі прогнозування захворюваності COVID-19. Цей підхід дозволив зменшити складність датасету та зосередитися на найбільш важливих показниках, спрощуючи подальшу роботу з ним.

Нижче представлений фрагмент датасету (рис. 3.2), який відображає дані після проведених маніпуляцій. У цьому фрагменті містяться лише обрані десять

ключових колонок, які були відібрані на основі їх важливості та кореляції з задачею прогнозування захворюваності COVID-19. Це дозволило зробити датасет більш структурованим та спрощеним, що полегшує подальші аналізи та розробку моделей.

	iso_code	continent	location	date	total_cases	new_cases	total_deaths	new_deaths	total_vaccinations
10	AFG	Asia	Afghanistan	2020-03-05	5.0	0.0	NaN	NaN	NaN
11	AFG	Asia	Afghanistan	2020-03-06	5.0	0.0	NaN	NaN	NaN
12	AFG	Asia	Afghanistan	2020-03-07	8.0	3.0	NaN	NaN	NaN
13	AFG	Asia	Afghanistan	2020-03-08	8.0	0.0	NaN	NaN	NaN
14	AFG	Asia	Afghanistan	2020-03-09	8.0	0.0	NaN	NaN	NaN
15	AFG	Asia	Afghanistan	2020-03-10	8.0	0.0	NaN	NaN	NaN
16	AFG	Asia	Afghanistan	2020-03-11	11.0	3.0	NaN	NaN	NaN
17	AFG	Asia	Afghanistan	2020-03-12	11.0	0.0	NaN	NaN	NaN
18	AFG	Asia	Afghanistan	2020-03-13	11.0	0.0	NaN	NaN	NaN
19	AFG	Asia	Afghanistan	2020-03-14	14.0	3.0	NaN	NaN	NaN
20	AFG	Asia	Afghanistan	2020-03-15	20.0	6.0	NaN	NaN	NaN
21	AFG	Asia	Afghanistan	2020-03-16	25.0	5.0	NaN	NaN	NaN
22	AFG	Asia	Afghanistan	2020-03-17	26.0	1.0	NaN	NaN	NaN
23	AFG	Asia	Afghanistan	2020-03-18	26.0	0.0	NaN	NaN	NaN

Рисунок 3.2 – Датасет після скорочення даних

Проаналізувавши скорочений датасет, було помічено, що деякі колонки містять комірки з пропущеними даними. Важливо зазначити, що пропущені дані не означають відсутність даних повністю, а лише вказують на те, що в певні дати були відсутні випадки, смерті або вакцинації.

У випадку цього датасету, для забезпечення коректності аналізу та моделювання, було прийнято рішення замінити пропущені значення на нулі. Такий підхід дозволяє зберегти консистентність даних та уникнути проблем з обробкою пропущених значень у подальшому аналізі.

Заміна пропущених значень на нулі може бути виправдана в даному контексті, оскільки нуль відображає відсутність певної події, такої як випадки захворювання, смерті або вакцинації. Це дозволяє зберегти структуру датасету та забезпечити правильність аналізу та моделювання наступних кроків у дослідженні COVID-19.

Вигляд датасету після заміни пустих комірок нулями показано нижче (рис. 3.3). Заміна пропущених значень на нулі дозволяє враховувати ці дані в подальших обчисленнях та розрахунках, але варто пам'ятати, що це рішення може мати вплив на кінцеві прогнози та результати моделі.

	iso_code	continent	location	date	total_cases	new_cases	total_deaths	new_deaths	total_vaccinations
10	AFG	Asia	Afghanistan	2020-03-05	5.0	0.0	0.0	0.0	0.0
11	AFG	Asia	Afghanistan	2020-03-06	5.0	0.0	0.0	0.0	0.0
12	AFG	Asia	Afghanistan	2020-03-07	8.0	3.0	0.0	0.0	0.0
13	AFG	Asia	Afghanistan	2020-03-08	8.0	0.0	0.0	0.0	0.0
14	AFG	Asia	Afghanistan	2020-03-09	8.0	0.0	0.0	0.0	0.0
15	AFG	Asia	Afghanistan	2020-03-10	8.0	0.0	0.0	0.0	0.0
16	AFG	Asia	Afghanistan	2020-03-11	11.0	3.0	0.0	0.0	0.0
17	AFG	Asia	Afghanistan	2020-03-12	11.0	0.0	0.0	0.0	0.0
18	AFG	Asia	Afghanistan	2020-03-13	11.0	0.0	0.0	0.0	0.0
19	AFG	Asia	Afghanistan	2020-03-14	14.0	3.0	0.0	0.0	0.0
20	AFG	Asia	Afghanistan	2020-03-15	20.0	6.0	0.0	0.0	0.0
21	AFG	Asia	Afghanistan	2020-03-16	25.0	5.0	0.0	0.0	0.0
22	AFG	Asia	Afghanistan	2020-03-17	26.0	1.0	0.0	0.0	0.0
23	AFG	Asia	Afghanistan	2020-03-18	26.0	0.0	0.0	0.0	0.0

Рисунок 3.3 – Датасет після заміни

Після підготовки датасету до ґрунтового аналізу, ми переходимо до вибору конкретної країни для подальшого дослідження. Адже у кожної країни свій рівень медицини і свої методи захисту, тож доцільніше проводити прогнозування для кожної країни окремо, ніж об'єднувати усі країни разом.

За приклад, обрано Німеччину. Для кращого уявлення про динаміку захворюваності в Німеччині будемо кілька графіків.

Важливо враховувати, що у Німеччині проживає значна кількість населення, і це має прямий вплив на кількість захворюваності. Тому, з метою наочності та зрозуміння масштабів проблеми, шкалу значень по вертикальній осі на графіку було зображено у форматі $1 \cdot 10^7$, тобто кожна одиниця на графіку відповідає 10 мільйонам випадків захворювання.

Цей формат допомагає візуалізувати величезний масштаб проблеми і показати, що кількість захворюваності в Німеччині дуже велика.

Починаємо з графіка, що відображає загальну кількість випадків COVID-19 в Німеччині протягом певного періоду часу (рис. 3.4). Цей графік дозволить нам спостерігати тенденцію поширення хвороби та з'ясувати, чи спостерігається зростання або спад кількості випадків.

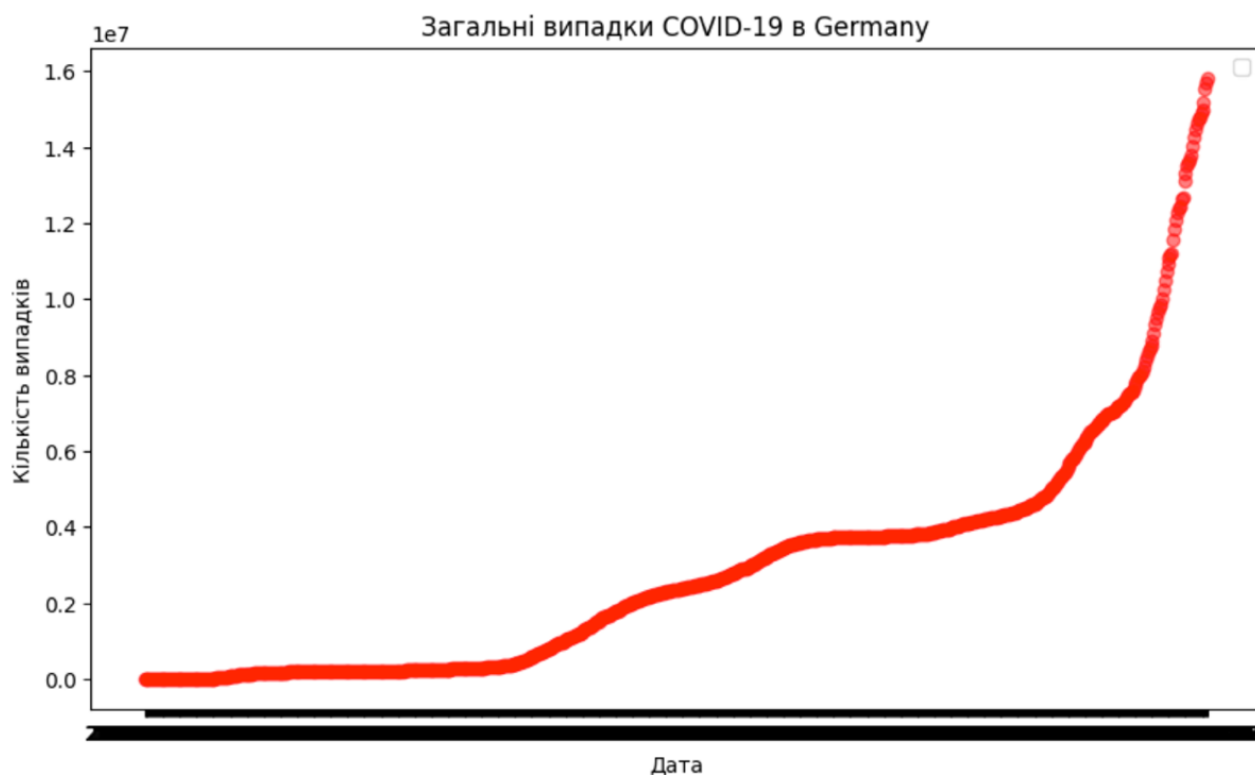


Рисунок 3.4 – Графік загальних випадків COVID-19 в Німеччині

Згідно з графіком (рис. 3.4), видно, що кількість загальних випадків COVID-19 в Німеччині зростала протягом тривалого періоду від 24 лютого 2020 року до 5 березня 2022 року. Адже цей показник визначає суму усіх випадків захворюваності на COVID-19.

Далі, будемо графік, який відображає загальну кількість смертей від COVID-19 в Німеччині протягом того ж періоду часу. Цей графік допоможе нам визначити, як впливає хвороба на смертність.

Відповідно до попереднього графіка загальних випадків захворювань COVID-19 (рис.3.4), на графіку загальної смертності від COVID-19 (рис. 3.5) кількість смертей збільшувалась, адже це загальне значення за весь період (з 24.02.2020 по 05.03.2022) в Німеччині.

Кореляція вказує на статистичний зв'язок між двома змінними - кількістю захворюваності та смертністю. У даному випадку, існує позитивна кореляція, оскільки зі збільшенням кількості захворювань COVID-19 може збільшуватись і кількість смертей пов'язаних з цим захворюванням.

Ця тенденція є логічною, оскільки важкі форми захворювання можуть призводити до серйозних ускладнень та смерті. Також, збільшена кількість хворих може перевантажити систему охорони здоров'я, що може впливати на якість та доступ до медичної допомоги, що може призводити до більшого числа смертей.

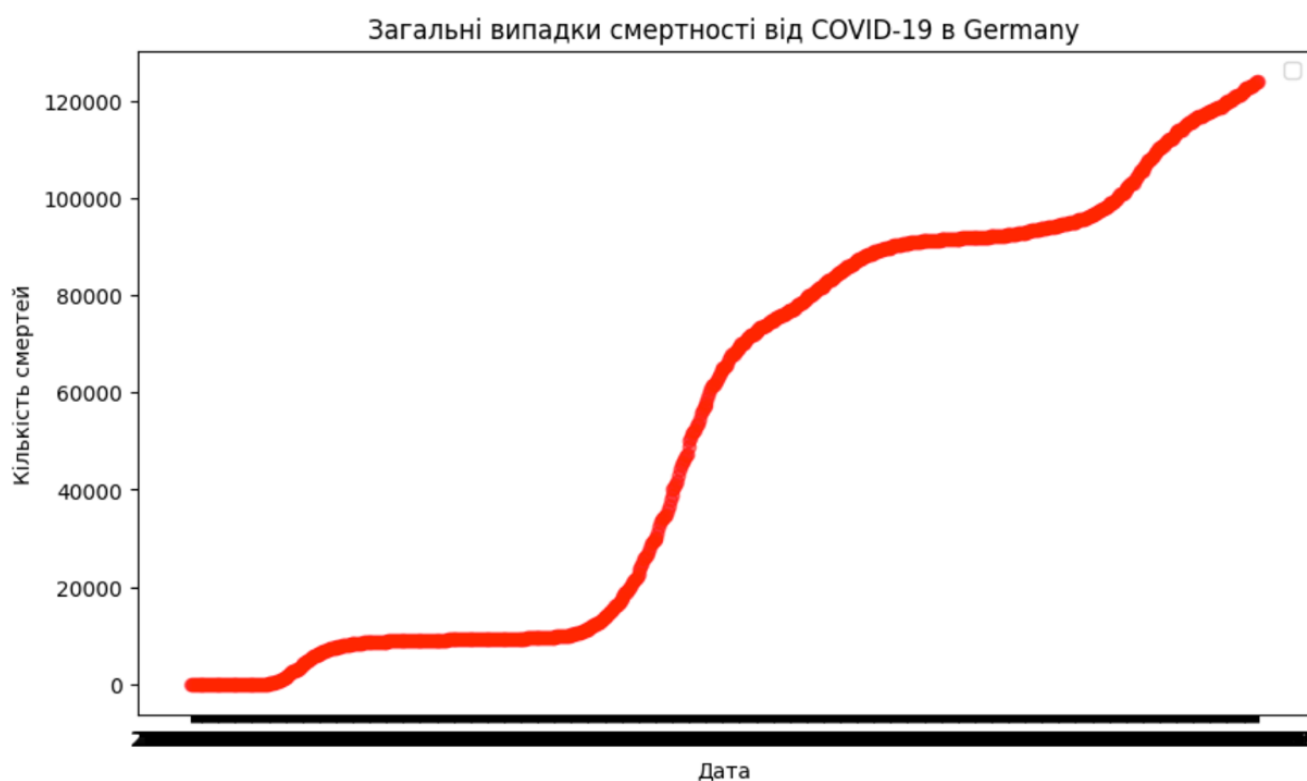


Рисунок 3.5 – Графік загальних випадків смертей від COVID-19 в Німеччині

Далі, будуюмо графік (рис. 3.6) кількості нових випадків COVID-19. Згідно з цим графіком, відзначається, що існують періоди, коли кількість випадків зростає швидко, що може бути пов'язано з появою нових штамів хвороби або іншими факторами, що сприяють поширенню захворюваності. У таких періодах може бути необхідно прийняти додаткові заходи для контролю і обмеження поширення COVID-19.

Також помітно, що після періоду зростання кількості випадків, можуть наступати періоди зменшення, що може бути результатом ефективної реалізації заходів контролю і вакцинації. Однак, варто враховувати, що залежно від ситуації, можуть бути нові хвилі захворюваності, і важливо продовжувати дотримуватись рекомендацій та заходів безпеки для запобігання поширенню COVID-19.

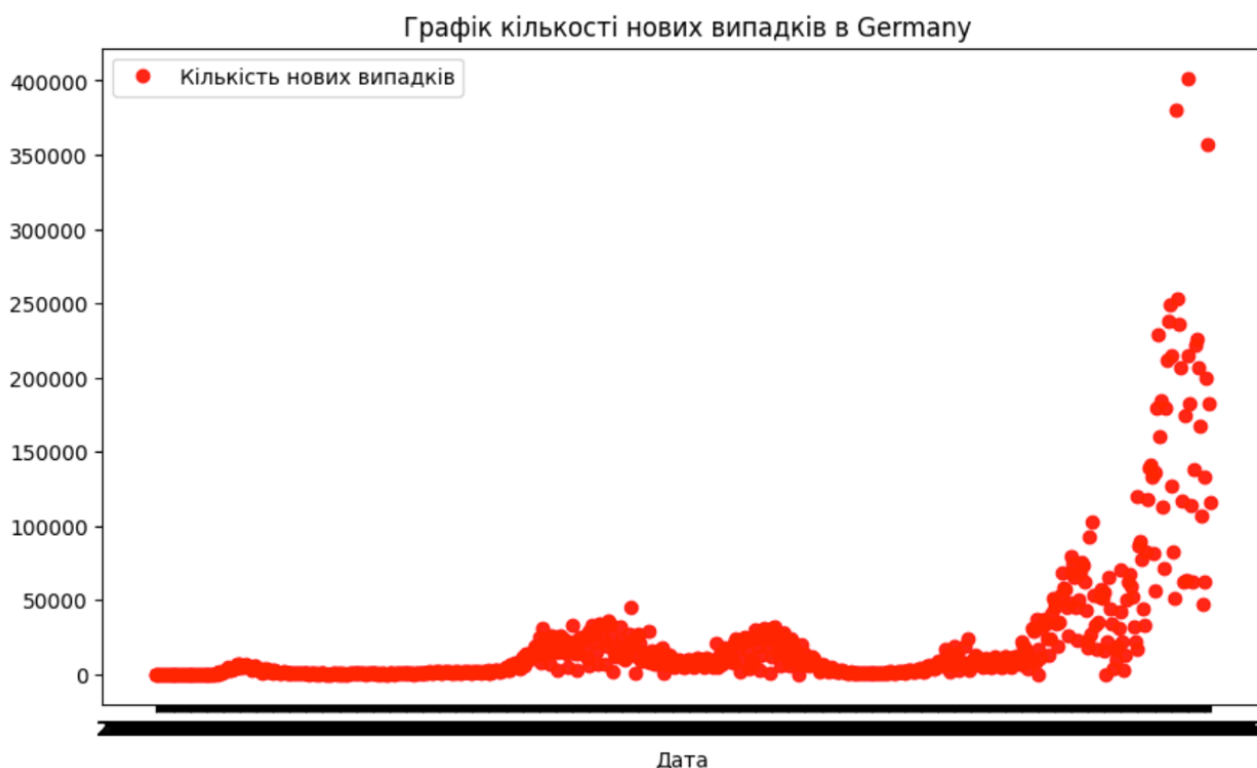


Рисунок 3.6 – Графік кількості нових випадків COVID-19 в Німеччині

З аналізу графіка кількості вакцинацій в Німеччині (рис. 3.7) стає очевидним, що на певному етапі було розпочато активну вакцинацію населення. Це свідчить про впровадження масштабних заходів щодо захисту населення від COVID-19.

На графіку можна побачити, що починаючи з певного часового моменту, кількість проведених вакцинацій почала значно зростати. Це вказує на те, що влада та організації охорони здоров'я вклали значні зусилля в організацію та проведення вакцинаційної кампанії. За допомогою цієї кампанії значна кількість людей отримала вакцину, що свідчить про зростання обізнаності та розуміння щодо важливості імунізації для здоров'я громади.

Активна вакцинація населення є показником напрацювання ефективної стратегії контролю та стримування поширення вірусу. Це свідчить про зусилля, спрямовані на забезпечення колективного захисту населення. Вакцинація великої кількості людей допомагає знизити ризик зараження та поширення вірусу в громаді, а також зменшує важкість хвороби та ризик розвитку ускладнень.

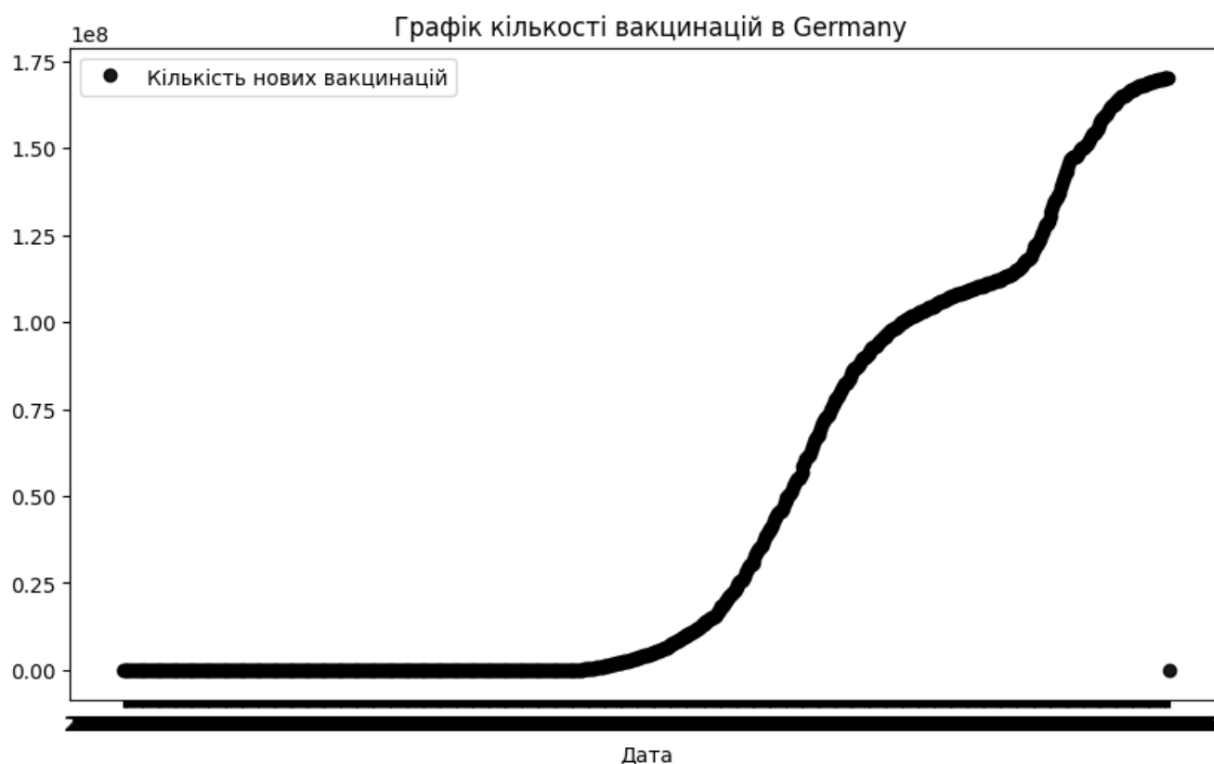


Рисунок 3.7 – Графік кількості вакцинації в Німеччині

Якщо звернути увагу на графік кількості нових випадків захворюваності (рис. 3.6), то можна побачити, що в період початку вакцинації кількість нових випадків зображувалась хвилями, тобто на ці дані мали вплив вакцинації, нових штамів, або ж більша кількість тестувань населення. Це може бути результатом ефективності вакцинації в контролюванні поширення хвороби. Зменшення кількості нових випадків свідчить про успішну стратегію протидії пандемії шляхом вакцинації населення.

Після огляду цих графіків, стало легше зрозуміти динаміку поширення COVID-19 в Німеччині, вплив заходів контролю та вакцинації на кількість випадків.

Це важливі дані, які можуть допомогти у прийнятті рішень щодо подальших заходів для прогнозування захворюваності COVID-19.

Далі побудуємо кореляційну матрицю. Це важливий крок у аналізі даних, оскільки дозволяє встановити зв'язки між різними змінними. Кореляційна матриця показує ступінь лінійної залежності між парами змінних і допомагає виявити потенційні взаємозв'язки та залежності між ними.

Коефіцієнт кореляції вимірює ступінь лінійної залежності між двома змінними і приймає значення в проміжку -1 – 1 . Значення, що ближче до одиниці показує позитивну лінійну залежність, значення близьке до -1 показує негативну лінійну залежність. Якщо значення близьке до нуля, то це означає відсутність лінійної залежності.

Отримавши значення коефіцієнтів кореляції для всіх пар змінних, було побудовано кореляційну матрицю (рис. 3.8), де кожен елемент матриці представляє собою коефіцієнт кореляції між відповідними змінними. Кореляційна матриця представлена графічно у вигляді теплової карти, де кольори відображають рівень кореляції.

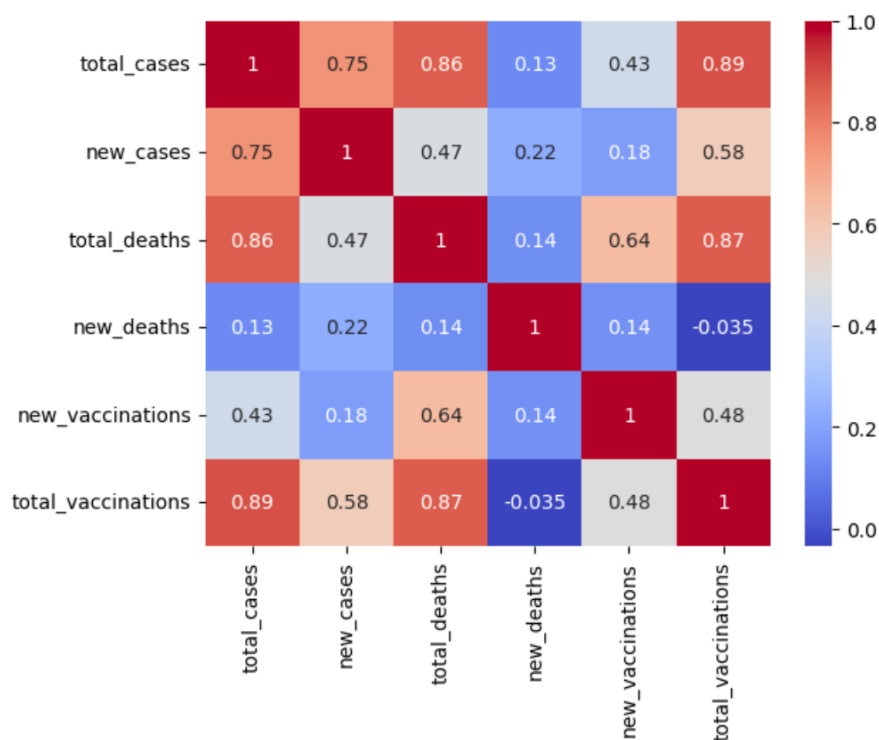


Рисунок 3.8 – Кореляційна матриця

Згідно з кореляційною матрицею (рис. 3.8), ми спостерігаємо дуже сильний зв'язок між показниками загальної кількості випадків COVID-19 і загальної кількості смертей, що є очевидним результатом. Це свідчить про те, що країни з високим рівнем захворюваності відповідно мають достатньо високий рівень смертності.

Також важливим є високий рівень кореляції між загальною кількістю випадків COVID-19 і загальною кількістю вакцинацій. Це означає, що країни з більшою кількістю вакцинацій, ймовірно, мають меншу кількість випадків захворювання.

На основі цих спостережень прийнято рішення використовувати загальну кількість вакцинацій для прогнозування кількості нових захворювань. Кореляційний коефіцієнт 0,89 свідчить про значну позитивну залежність між цими змінними.

З метою отримання більшого розуміння взаємозв'язків між стовпчиками датасету, було побудовано графік (рис. 3.9), який візуалізує ці залежності.

На головній діагоналі графіка розташовані гистограми або оцінки щільності розподілу для кожної окремої змінної. Наприклад, якщо розглянемо взаємозв'язок між "кількістю випадків", "кількістю смертей" та "кількістю вакцинацій", то на головній діагоналі будуть графіки, які відображають розподіл цих змінних.

Взаємозв'язки між парами змінних представлені на решті графіків, розташованих поза головною діагоналлю. Наприклад, для визначення залежностей між "кількістю випадків" та "кількістю смертей", або між "кількістю випадків" та "кількістю вакцинацій", можуть бути показані в вигляді точкових графіків, діаграм розсіювання або лінійних регресійних моделей.

Такий тип візуалізації дозволяє отримати загальне уявлення про взаємозв'язки між змінними у датасеті і може бути корисним для виявлення потенційних залежностей або взаємодій між змінними. Він допомагає отримати візуальне представлення даних та допомагає зрозуміти їхню взаємодію.

Отже, аналізуючи наведені дані і застосовуючи кореляційну матрицю, можна встановити зв'язки між різними показниками та використовувати їх для прогнозування та прийняття рішень у контексті COVID-19.

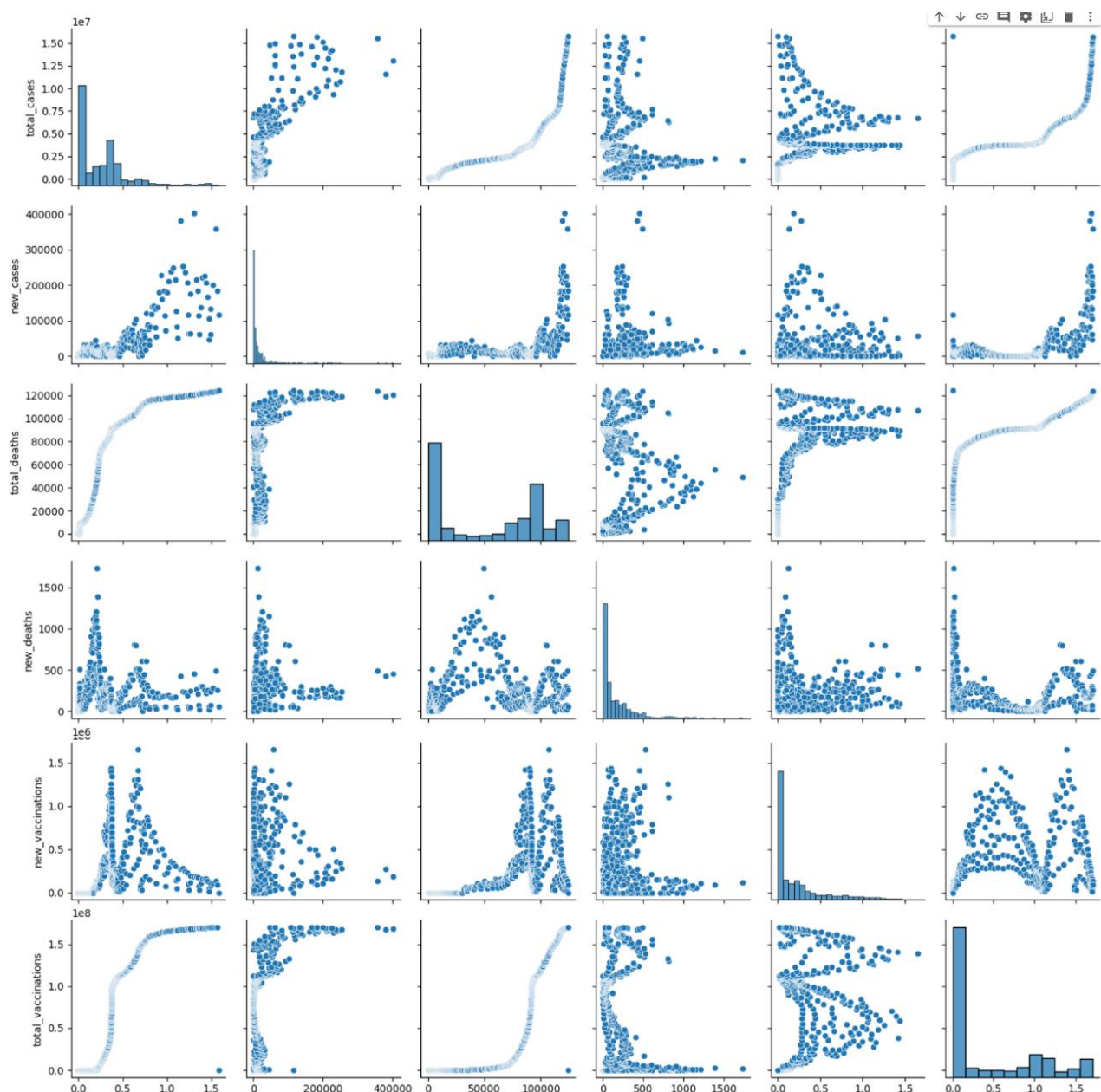


Рисунок 3.9 – Візуалізація відношень стовпчиків датасету

Кореляційна матриця надає нам важливу інформацію про ступінь залежності між змінними у датасеті. За її допомогою ми можемо виявити, як один параметр змінюється при зміні іншого, та оцінити силу цієї залежності.

Цей аналіз може допомогти зрозуміти, як різні фактори впливають на поширення хвороби, на її смертність та на ефективність вакцинації. Наприклад,

можна виявити позитивну кореляцію між кількістю випадків та кількістю смертей, що свідчить про високий рівень смертності при великій кількості випадків.

Отже, аналіз даних та використання кореляційної матриці надають нам цінні інструменти для розуміння зв'язків між показниками та використання цих знань для прогнозування та прийняття рішень у боротьбі з пандемією COVID-19.

3.2. Побудова моделі. Лінійна регресія

На цьому етапі ми переходимо до побудови моделі авторегресії для прогнозування захворюваності на COVID-19, враховуючи показник вакцинації і зважаючи на високі коефіцієнти кореляції, які були зазначені раніше.

Модель авторегресії дозволить нам урахувати тенденції та зміни в часовому ряді, а також зробити прогноз на майбутні періоди, використовуючи показник вакцинації як один із важливих факторів.

На цьому етапі ми переходимо до побудови моделі лінійної регресії для прогнозування захворюваності на COVID-19, використовуючи показник вакцинації. В якості незалежної змінної в моделі ми беремо значення загальної кількості вакцинацій. Цей показник відображає прогрес вакцинаційної кампанії в країні (в нашому випадку розглядається Німеччина).

Залежною змінною в моделі є кількість нових захворювань, яка вказує на зростання або зниження кількості випадків захворюваності.

Переходимо до оцінки ефективності даної моделі. Як видно з рис.3.10, значення середньої квадратичної помилки (Mean Squared Error, MSE) складає 17595.289, що вказує на розброс прогнозів моделі відносно фактичних значень.

Mean Absolute Error: 17595.288792005103
R-squared: 0.4135555712843667

Рисунок 3.10 – Значення середньої квадратичної помилки та коефіцієнта детермінації (лінійна регресія)

Значення коефіцієнта детермінації (R-squared) становить 0.41, що означає, що лише 41% варіації залежної змінної можна пояснити моделлю.

Хоча значення MSE та R-squared вказують на певну лінійну кореляцію між показником вакцинації та кількістю нових захворювань, вони також свідчать про значний розкид прогнозів моделі відносно реальних значень. Це може вказувати на те, що лінійна регресія може бути недостатньо гнучкою для точного прогнозування.

З огляду на ці результати, можна припустити, що при застосуванні поліноміальної регресії буде отримано кращу модель. Поліноміальна регресія дозволяє враховувати нелінійні залежності між змінними, що може покращити точність прогнозів. Подальше дослідження та налаштування моделі поліноміальної регресії можуть привести до кращих результатів у прогнозуванні захворюваності на COVID-19 на основі показника вакцинації.

3.3. Побудова моделі. Поліноміальна регресія

На даному етапі вирішено застосувати поліноміальну регресію для аналізу та моделювання даних.

Застосування поліноміальної регресії передбачає моделювання залежності між залежною змінною (y) та незалежною змінною (x) у вигляді полінома n -го степеня. При цьому коефіцієнти полінома будуть оцінюватися шляхом підгонки моделі до наших даних.

Тож застосування поліноміальної регресії може допомогти виявити складніші шаблони та зв'язки в наших даних, які не можуть бути виявлені за допомогою простої лінійної регресії. Це дає нам можливість зробити більш точні прогнози та аналізувати вплив різних факторів на нашу залежну змінну.

Після побудови поліноміальної регресії, ми виводимо значення середньої квадратичної помилки (Mean Squared Error, MSE) та коефіцієнта детермінації, щоб оцінити ефективність моделі. Згідно з рис. 3.11, видно, що отримані значення стали кращими. Зокрема, значення середньої квадратичної помилки скоротилося до

15393.23, порівняно з попередніми результатами (рис. 3.10), де значення MSE було вище. Це свідчить про покращення роботи моделі.

Mean Absolute Error: 15393.23475496345
R-squared: 0.6683179372157755

Рисунок 3.11 – Значення середньої квадратичної помилки та коефіцієнта детермінації (поліноміальна регресія)

Також це підтверджується коефіцієнтом детермінації, який став більшим за попередні результати, а саме 0.67. Коефіцієнт детермінації вказує на те, яка частка дисперсії залежної змінної пояснюється моделлю. Значення, яке було отримано після побудови поліноміальної регресії, свідчить про те, що модель стала більш адекватною і краще пояснює залежність між кількістю захворювань на COVID-19 та показником вакцинації.

Отже, результати поліноміальної регресії свідчать про покращення точності та адекватності моделі, що дозволяє зробити більш надійні прогнози щодо кількості нових захворювань на COVID-19 на основі показника вакцинації.

Проте, отримані результати після побудови поліноміальної регресії з однією змінною можуть бути значно покращені. З метою досягнення ще більшої ефективності роботи моделі, спробуємо підвищити степінь полінома.

Збільшення степені полінома дозволить моделі виразніше апроксимувати складніші залежності між змінними. Це означає, що модель зможе краще узгоджуватись з даними та точніше прогнозувати значення захворюваності на COVID-19 на основі показника вакцинації.

Варто відзначити, що збільшення степені полінома може призвести до більшої складності моделі та збільшення ризику перенавчання. Тому важливо бути обережними та знайти баланс між точністю моделі та її загальною простотою.

Після збільшення степені полінома та побудови моделі з новими показниками, буде знову оцінена її ефективність за допомогою статистичних метрик, таких як середня квадратична помилка та коефіцієнт детермінації.

Після проведених обчислень та збільшення степені полінома у моделі, ефективність моделі ще покращилась. Отримані результати підтверджують це.

Знижене значення середньої квадратичної помилки до 12031.45 є ознакою покращення точності прогнозів та зменшення розбіжностей між спостережуваними та прогнозованими значеннями.

Також добрим показником є підвищення значення коефіцієнта детермінації до 0.72. Чим ближче значення коефіцієнта детермінації до 1, тим краще модель узгоджується з даними та ефективніше прогнозує результати.

Отже, отримані результати свідчать про покращення ефективності моделі після підвищення степені полінома. Це може бути важливим відкриттям у прогнозуванні захворюваності на COVID-19 на основі показника вакцинації, що може сприяти удосконаленню стратегій контролю та прогнозування пандемічних заходів.

Для визначення оптимального степеня полінома в регресійній моделі можна побудувати графік залежності степеня полінома від похибки. Це допоможе з'ясувати, який степінь полінома найкраще підходить для даного набору даних.

На графіку буде вісь X, яка представлятиме степінь полінома, а вісь Y - значення похибки (наприклад, середня квадратична помилка). Чим нижче значення похибки, тим краще модель.

Почнемо з невеликого діапазону степенів полінома, наприклад, від 1 до 10. Для кожного степеня полінома побудуємо модель і обчислимо відповідну похибку. Отримані значення похибки записуємо на графік (рис.3.12).

Mean Absolute Error: 12031.445334484355
R-squared: 0.7179742610931695

Рисунок 3.12 – Значення середньої квадратичної помилки та коефіцієнта детермінації після підвищення степеню полінома

В процесі дослідження графіка залежності степеня полінома від похибки, було виявлено, що похибка зменшується і стабілізується після 4 степеня полінома (рис.3.13). Це означає, що подальше підвищення степеня полінома не призводить до значної зміни у похибці.

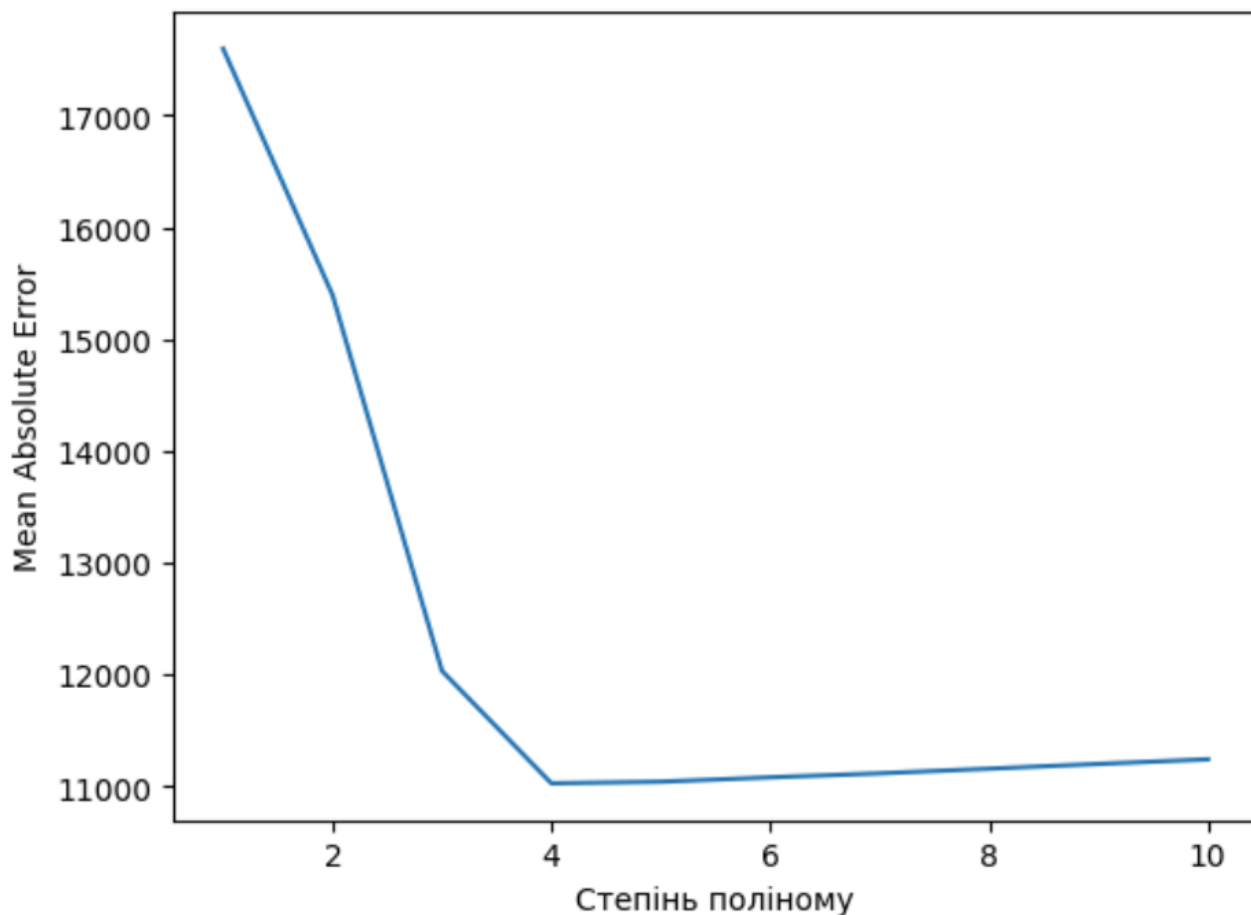


Рисунок 3.13 – Графік залежності степеню полінома від похибки

Цей результат важливий, оскільки він надає нам інформацію про оптимальний степінь полінома для наших даних. Потрібно зазначити, що використання надто високих степенів полінома може призвести до перенавчання моделі, коли вона надмірно пристосовується до навчальних даних і втрачає здатність до загальних прогнозів.

Отже, у цьому випадку, оптимальний степінь полінома для побудови моделі прогнозування захворюваності на COVID-19 для Німеччини залежить від наявних даних і складає 4. При цьому степені полінома модель є достатньо ефективною і добре апроксимує залежність між вакцинацією та кількістю нових захворювань.

Важливо пам'ятати, що вибір оптимального степеня полінома може варіюватись залежно від конкретного датасету (в цьому випадку залежить від вибору країни). Тому ретельний аналіз та експерименти з різними степенями полінома допоможуть вибрати найбільш підходящу модель для кожного конкретного випадку.

3.4. Інтерфейс інтерактивного додатку

Завершуючи опис даного інтерактивного додатку, перейдемо до розбору його інтерфейсу. При запуску додатку, користувачі бачать добре організований екран з декількома ключовими елементами.

Для реалізації налаштувань інтерфейсу була використана потужна бібліотека PyQt5, яка надає широкі можливості для створення та налаштування графічного інтерфейсу користувача. Завдяки її функціоналу та гнучкості, розробники змогли точно контролювати вигляд та поведінку кожного елементу на екрані. Використання PyQt5 дало можливість легко створити інтуїтивно зрозумілий та естетично привабливий інтерфейс, який забезпечує зручну навігацію та ефективну взаємодію з додатком для кожного користувача.

При першому запуску додатку, на екрані з'являється весь представлений функціонал.

На рис. 3.14 зображено вікно інтерактивного додатку, яке розроблено спеціально для прогнозування захворюваності на COVID-19. Це вікно пропонує користувачеві широкі можливості для отримання актуальної інформації та аналізу даних щодо поширення хвороби.

Інтерфейс додатку до прогнозування має привабливий та інтуїтивно зрозумілий вигляд, який надає користувачу легкість у користуванні. Дизайн інтерфейсу розроблений з урахуванням зручності користувача, щоб забезпечити йому зручну та ефективну роботу з додатком.

Першим є вибір країни, який є зручним та функціональним. На цьому розділі вибору країни розміщений величезний перелік країн з усього світу. Він представлений у вигляді випадаючого списку, який дозволяє користувачам зручно прокручувати та вибирати країни шляхом простого прокладання шляху курсору.

Список країн впорядкований за алфавітом, що дозволяє легко знайти потрібну країну за першою літерою. Такий підхід спрощує навігацію та дозволяє користувачам швидко знаходити бажану країну для подальшого аналізу та прогнозування захворюваності на COVID-19.

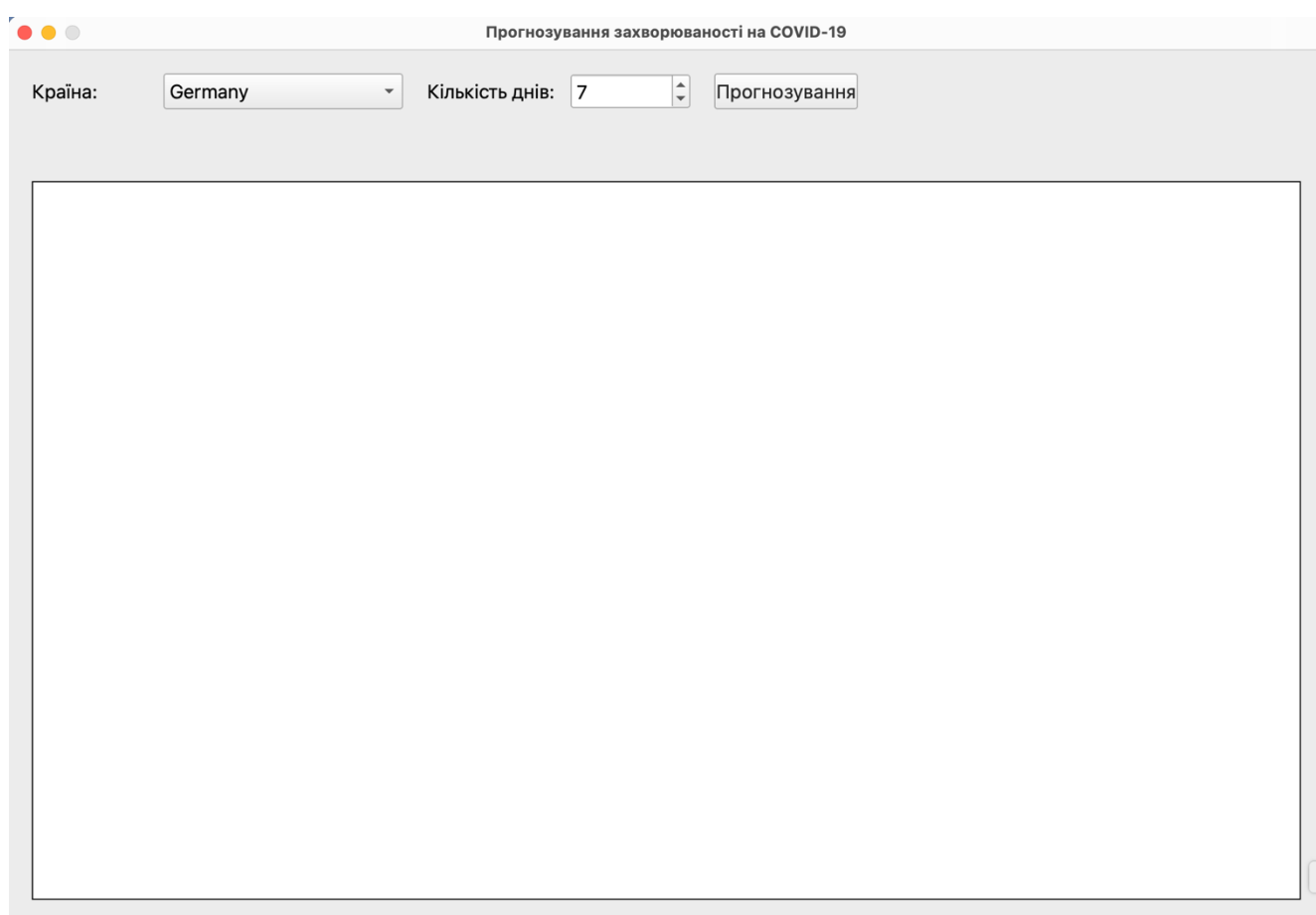


Рисунок 3.14 – Вигляд інтерфейсу.

Цей інтуїтивний та організований вибір країни дарує користувачам можливість легко встановлювати бажані налаштування, щоб глибше проаналізувати та прогнозувати захворюваність на COVID-19 у вибраній країні.

Ілюстрація списку вибору країни представлена нижче, на рис. 3.15.

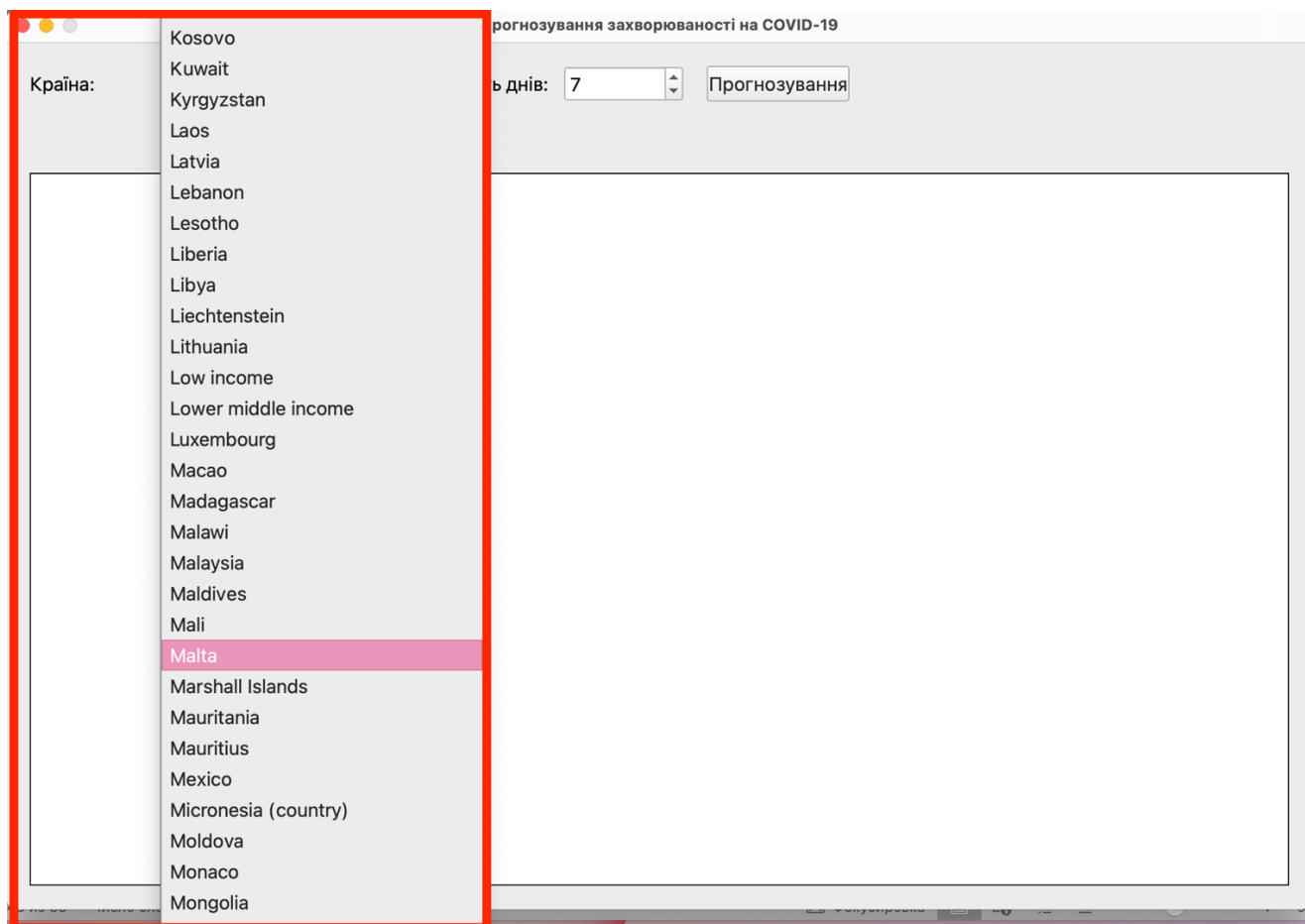


Рисунок. 3.15 – Вигляд інтерфейсу. Випадаючий список

Після того, як користувач вибрав країну, наступним кроком є вибір кількості днів для прогнозу захворюваності на COVID-19 (рис. 3.16).

В інтерактивному додатку надається можливість встановити максимальну кількість днів прогнозу, яка обмежена 10 днями.

Важливо зазначити, що при виборі більшої кількості днів для прогнозу, точність результатів може зменшуватися. Це пояснюється тим, що чим далі в майбутнє прогнозується захворюваність на COVID-19, тим складніше передбачити вплив різних факторів на захворюваність.

Наприклад, зміна епідеміологічних умов, впровадження нових заходів або зміна поведінки населення можуть вплинути на розповсюдження вірусу і змінити прогнозовані показники. Тому, при виборі кількості днів для прогнозу, варто уважно розглянути потреби та обмеження аналізу, забезпечуючи баланс між точністю та довжиною прогнозу.

Саме через ці причини, було обрано обмежити кількість днів прогнозу до 10.

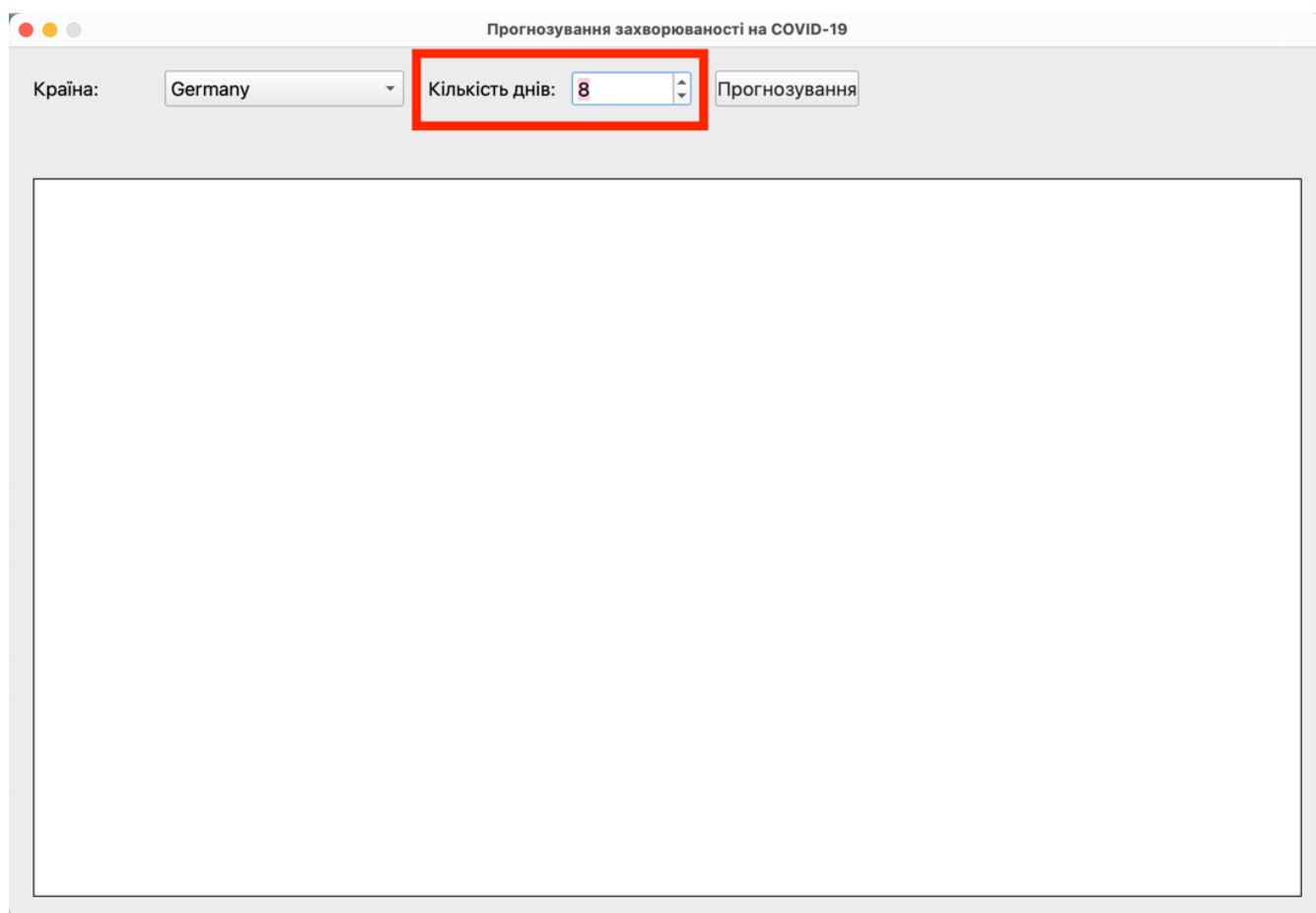


Рисунок 3.16 – Вигляд інтерфейсу. Вибір кількості днів прогнозу

Після вибору країни та кількості днів для прогнозу, на екрані є важлива кнопка «Прогнозування» (рис. 3.17). Ця кнопка має ключову роль у запуску процесу прогнозування захворюваності COVID-19.

Користувач, натискаючи на кнопку, активує алгоритми та моделі, які аналізують наявні дані, враховують тенденції та розраховують майбутні прогнози. Це важлива функція, яка надає користувачеві можливість отримати прогнозні дані щодо захворюваності на COVID-19 для обраної країни та періоду, що він вибрав.

Після натискання кнопки «Прогнозування» не лише запускається процес обробки даних та розрахунку прогнозних результатів, але також на екрані з'являється цікавий інтерактивний графік прогнозу захворюваності COVID-19. Такий приклад представлено нижче на рис. 3.18. Для візуалізації було обрано Німеччину.

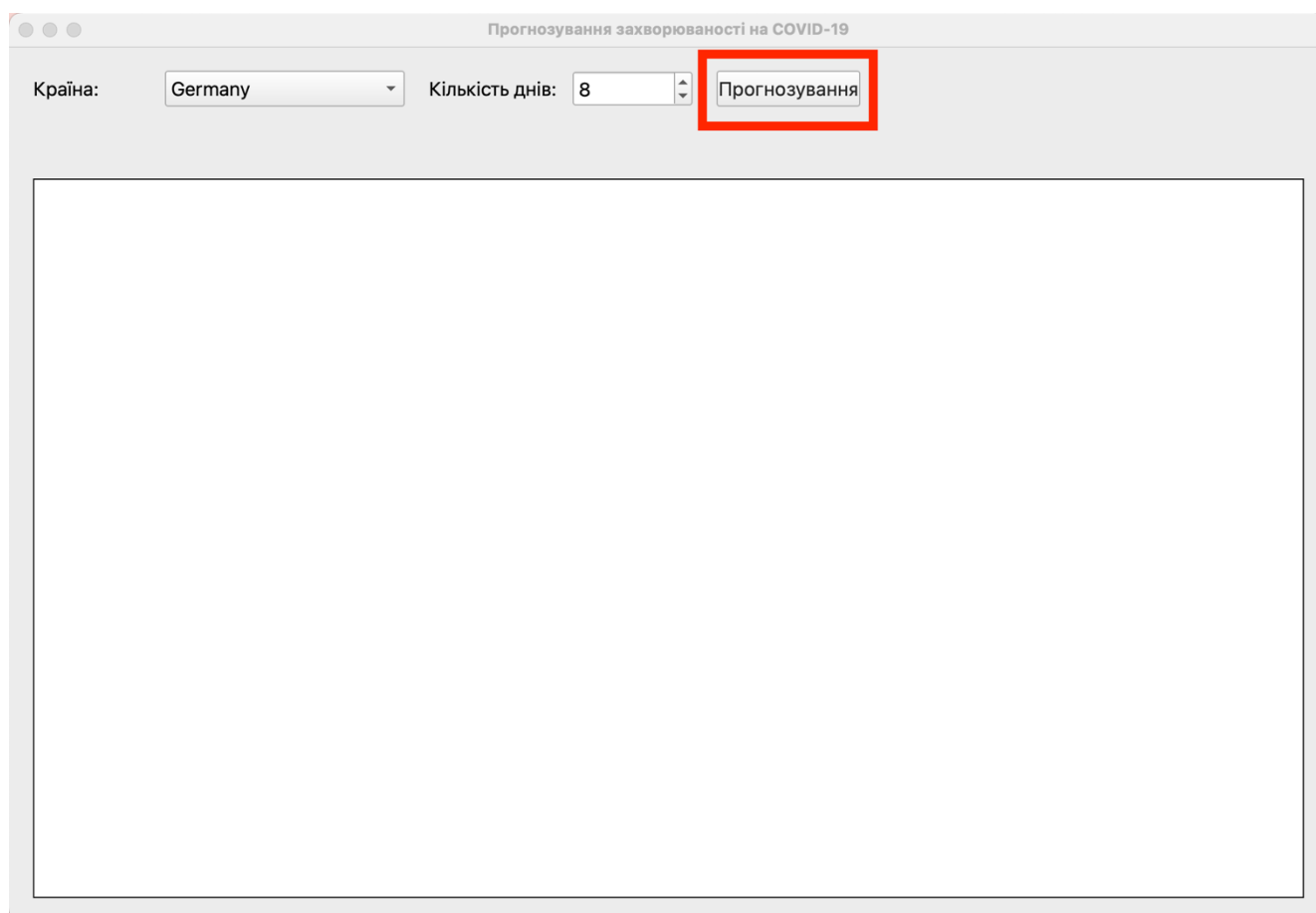


Рисунок 3.17 – Вигляд інтерфейсу. Кнопка «Прогнозування»

Графік відображає очікувану динаміку кількості випадків захворювання протягом вибраного періоду, враховуючи дані та алгоритми прогнозування. Це дозволяє користувачеві візуально спостерігати за тенденціями та змінами у захворюваності, а також отримувати уявлення про потенційний розвиток ситуації у майбутньому.

При виборі країни, наприклад, Німеччини, та встановленні кількості днів прогнозу на 7 днів, на екрані з'являється графік (рис. 3.18) з прогнозом захворюваності на COVID-19. Цей графік відображає очікувану динаміку кількості випадків захворювання протягом вибраного періоду. Крім того, над графіком розташоване текстове поле, яке містить точні цифрові значення прогнозу для кожного окремого дня. Це дозволяє користувачеві отримати конкретні числові дані щодо прогнозованої кількості випадків захворювання на кожен день прогнозу.

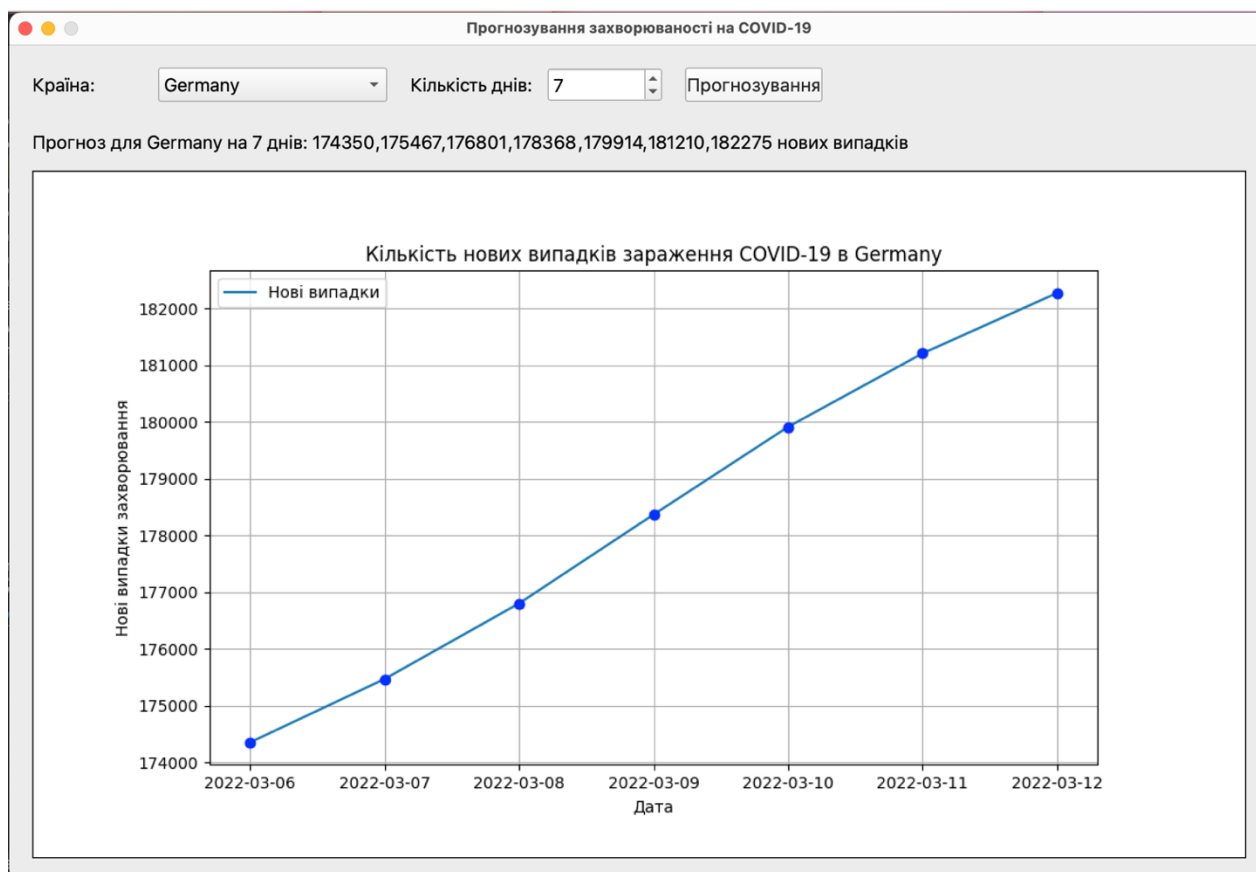


Рисунок 3.18 – Прогноз на тиждень. Німеччина

Текстове поле забезпечує точність та деталізацію прогнозу, щоб користувач міг отримати конкретну інформацію щодо очікуваних змін у захворюваності та відстежувати тенденції. Таке поєднання графічного відображення та числових значень надає користувачеві повний обсяг інформації для аналізу та зроблення відповідних висновків щодо прогнозу захворюваності на COVID-19 у вибраній країні та періоді.

Для того, щоб перевірити роботу інтерактивного додатку, було вирішено вибрати інші країни та різні періоди прогнозу. Один із таких прикладів - прогноз захворюваності на COVID-19 для Мальти протягом 10 днів. На рис. 3.19 наведено графік з прогнозом для Мальти, який відображає очікувану динаміку кількості випадків захворювання протягом обраного періоду. Цей графік надає можливість оцінити очікувані тенденції та зміни у захворюваності на COVID-19 для Мальти протягом 10 днів.

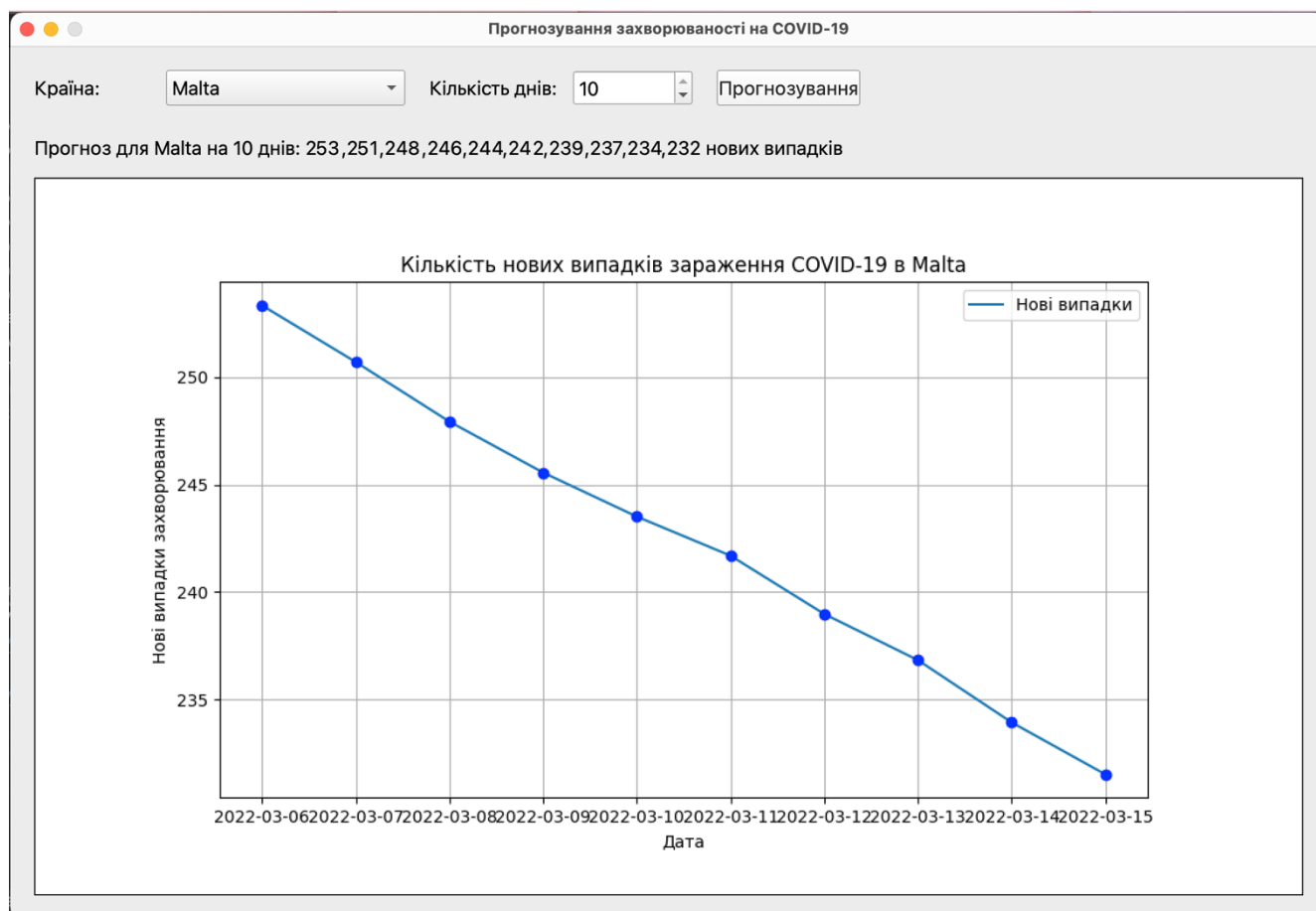


Рисунок 3.19 – Прогноз на 10 днів. Мальта

Користуючись цим графіком, можна проводити порівняння між різними країнами та періодами прогнозу, щоб отримати краще розуміння динаміки захворюваності та зробити аналітичні висновки.

Для подальшої перевірки роботи інтерактивного додатку до прогнозування захворюваності COVID-19, було обрано Україну та встановлено прогноз на 3 дні вперед. Графік прогнозу для України на рис. 3.20 нижче демонструє очікувану динаміку кількості випадків захворювання протягом обраного періоду. Зверху графіка, вказано точні числові значення кількості захворювань на наступні три дні: 23087, 23185, 23273. Ці числа надають конкретну інформацію про очікувану кількість захворювань в Україні на кожен з вибраних днів. Використовуючи цей графік та числові дані, користувач може отримати уявлення про прогнозовану тенденцію захворюваності та зробити аналітичні висновки щодо ситуації з COVID-19 в Україні на найближчі три дні.

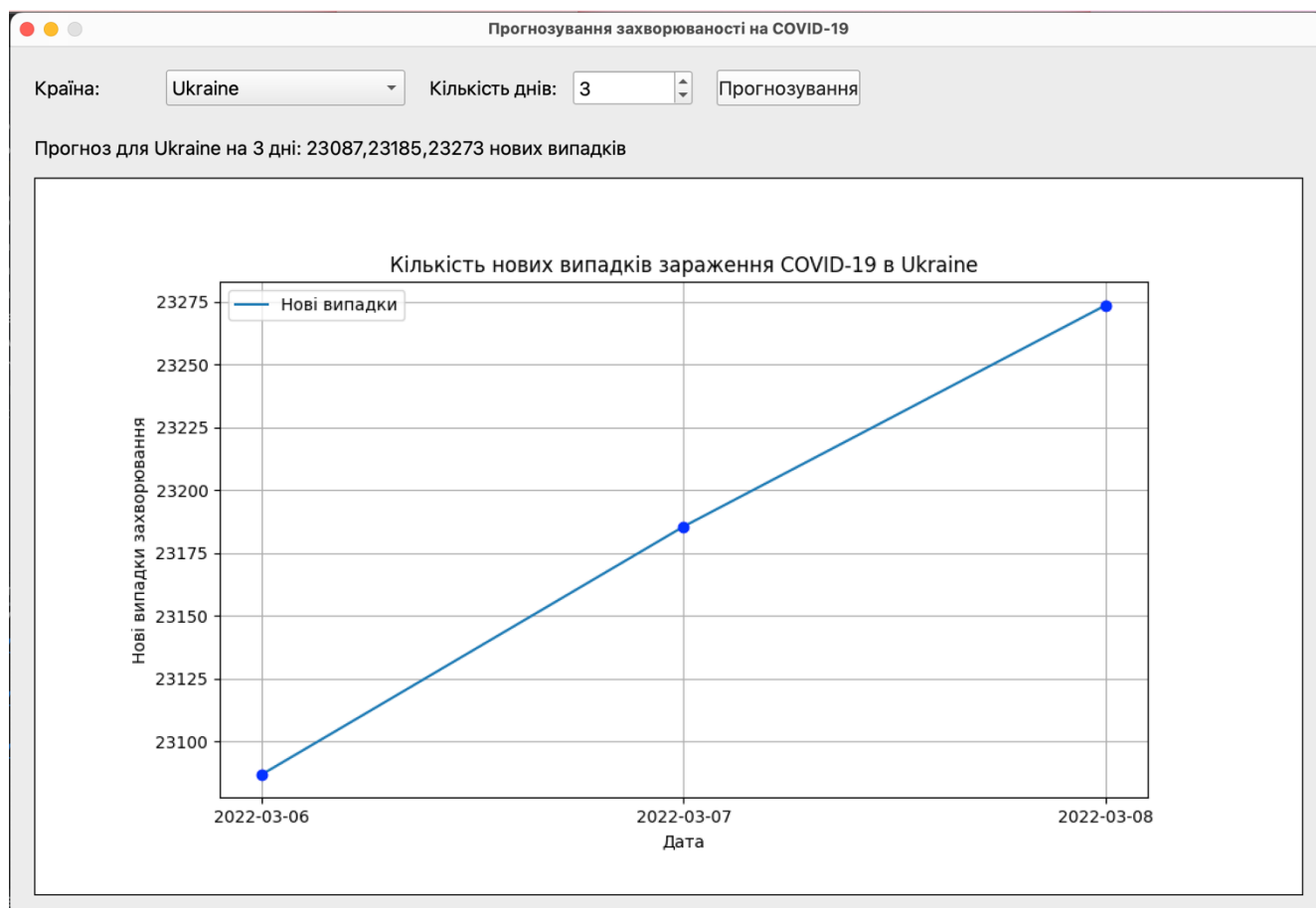


Рисунок 3.20 – Прогноз на 3 дні. Україна

Останньою країною, яку було вибрано для прогнозування, стали Сполучені Штати Америки. Прогноз на 5 днів для США представлений на графіку, позначеному як рис. 3.21.

Цей графік відображає очікувану зміну кількості захворюваності на COVID-19 протягом обраного періоду. Графік дозволяє користувачеві отримати візуальне уявлення про очікувану тенденцію у розповсюдженні хвороби в США протягом наступних п'яти днів.

Додатково, числові значення кількості захворювань на кожен день наведені над графіком, що надає точні дані для кожного прогнозованого дня. Аналізуючи цей графік та числові дані, користувач може отримати інформацію про передбачувану динаміку COVID-19 у США та зробити висновки щодо очікуваного розвитку ситуації в найближчі п'ять днів.

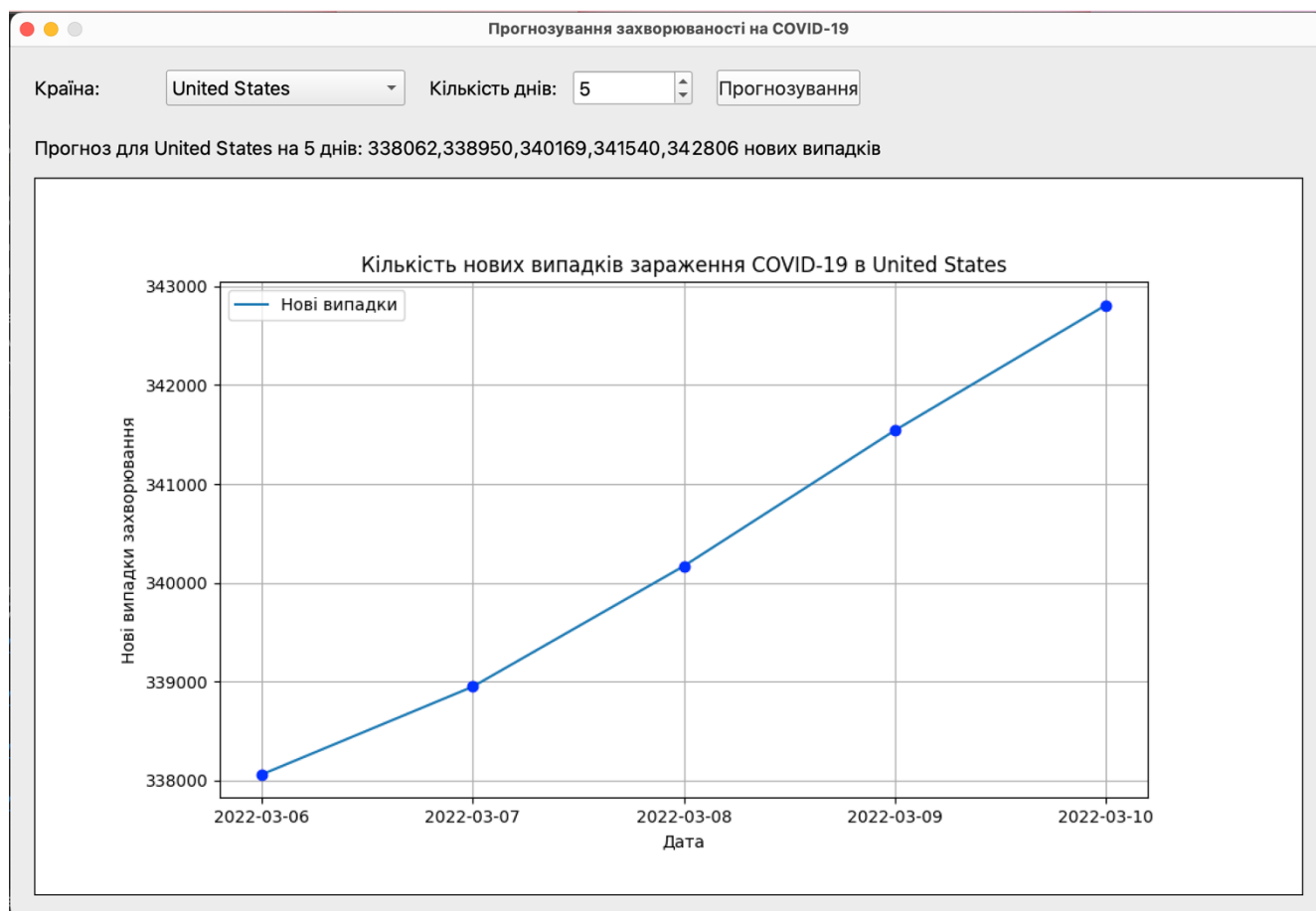


Рисунок 3.21 – Прогноз на 5 днів. США

Як видно з попередньо представлених рисунків, інтерактивний додаток до прогнозування захворюваності на COVID-19 виконує свою роботу на високому рівні. Він надає користувачам доступ до потужних інструментів та функцій для аналізу та прогнозування динаміки захворюваності.

3.5. Безпека життєдіяльності та охорона праці

Під час розробки програмного забезпечення було проведено дослідження, спрямоване на виявлення та оцінку потенційних ризиків та небезпек, пов'язаних з використанням системи інтерактивного додатку до прогнозування COVID-19, який розроблявся в ході дипломної роботи.

Виявлення та оцінка ризиків і небезпек є важливим етапом управління безпекою та захистом. В даному випадку, цей процес був проведений відповідно до

плану №3 з метою виявлення потенційних загроз, оцінки рівня ризику та впровадження необхідних заходів для вирішення зазначених питань.

У рамках дипломної роботи було розглянуто приклад робочого місця сімейного лікаря (рис. 3.22), який планує використовувати програмний додаток у своєму кабінеті. Детальний аналіз був проведений з метою ідентифікації можливих ризиків та шкідливих факторів, що можуть виникати в такому робочому середовищі, а також визначення відповідних заходів безпеки лікаря та пацієнтів.

У звіті з переддипломної практики детально описані характеристики кабінету, в якому планується використання програмного застосунку. Наведено повний перелік обладнання, що присутнє у кабінеті, зокрема: медичну кушетку, стіл, комп'ютер, стільці, шафу, раковину з тумбою, кондиціонер, вогнигасник, пожежний сповіщувач, відро для сміття, тощо.

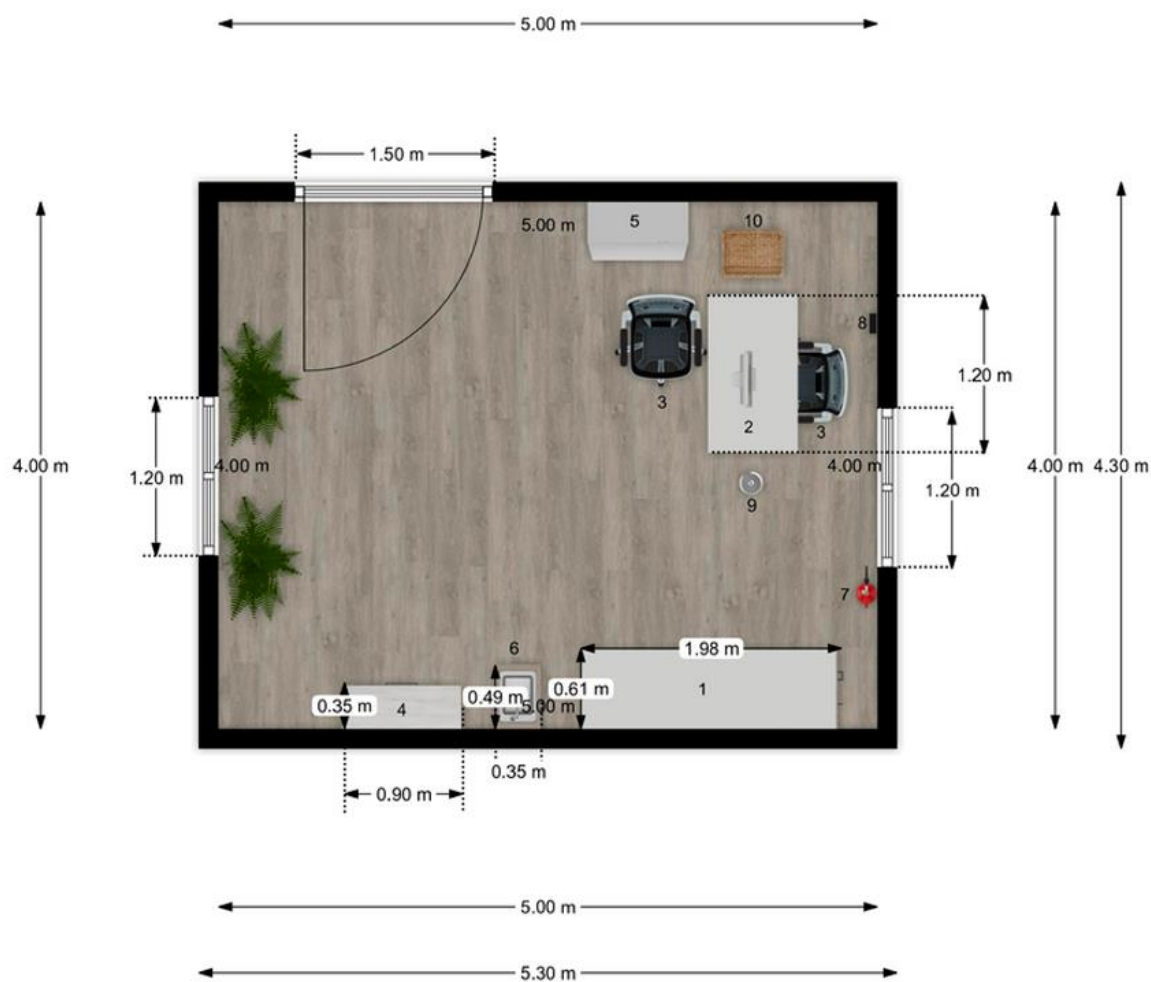


Рисунок. 3.22 – Схема кабінету лікаря

При проведенні розрахунків були використані норми охорони праці для забезпечення безпеки та здоров'я співробітників.

3.6. Розрахунок економічного ефекту за темою ДР

У звіті переддипломної практики була проведена оцінка ключових функціональних та вартісних характеристик розробленого інтерактивного додатку до прогнозування захворюваності COVID-19.

Під час аналізу було розглянуто різні підходи до досягнення визначеної мети та найкращий варіант був вибраний на основі аналізу економічних факторів та характеристик розробленого інтерактивного додатку.

З метою вирішення основних функцій було розроблено морфологічну карту, яка відображена на рис. 3.23. Ця карта включає запропоновані способи реалізації основних функцій програмного продукту та його компонентів.

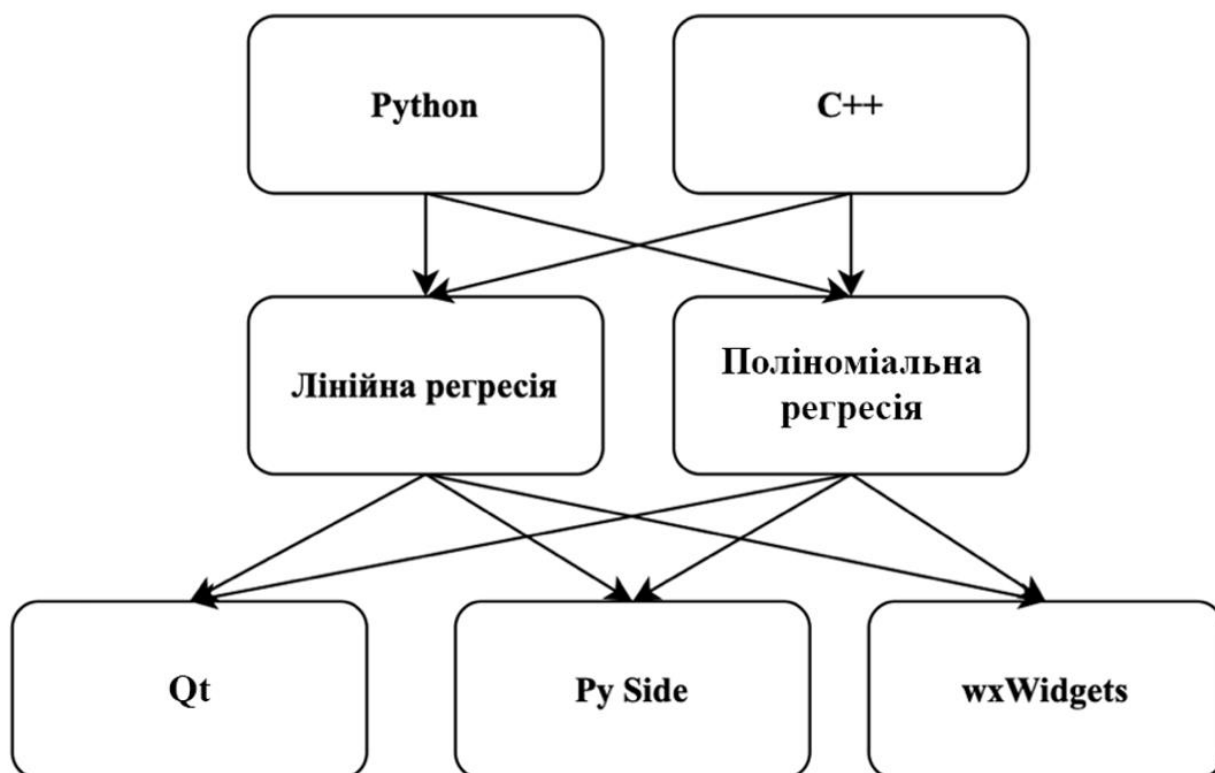


Рисунок 3.23 – Морфологічна карта

Під час проведення економічного аналізу в рамках дипломної роботи було здійснено обґрунтування розробки інтерактивного додатку для прогнозування захворюваності на COVID-19 з використанням методу функціонально-вартісного аналізу.

В результаті аналізу основних функцій було виявлено два найбільш перспективних варіанти реалізації продукту. З цих варіантів найбільш ефективним виявився той, що має найвище значення коефіцієнта техніко-економічного рівня, при цьому вартість розробки становить 308 996,72 грн. Це обґрунтовує вибір даного варіанту як оптимального з економічної точки зору.

Описаний варіант передбачає наступні характеристики:

- Використання мови програмування Python.
- Застосування поліноміальної регресії.
- Розробку інтерфейсу користувача на основі PyQt.

Висновки до розділу 3

У цьому розділі було проведено аналіз даних щодо захворюваності на COVID-19, зосередившись на залежності між кількістю вакцинацій та кількістю нових захворювань. Було використано різні методи, включаючи лінійну та поліноміальну регресію, для побудови моделей прогнозування.

За допомогою кореляційного аналізу було встановлено сильний зв'язок між кількістю вакцинацій та кількістю нових захворювань. Це дало підставу вважати, що вакцинація впливає на зниження захворюваності на COVID-19.

Починаючи зі звичайної лінійної регресії, було перейдено до поліноміальної регресії, щоб покращити точність моделі прогнозування. За допомогою оцінки середньої квадратичної помилки (MSE) та коефіцієнта детермінації (R-squared), було з'ясовано, що поліноміальна регресія з показником 4 є оптимальною для наших даних.

Після введення полінома четвертого степеня, помічено поліпшення в ефективності моделі. Значення середньої квадратичної помилки зменшилося, а

коефіцієнт детермінації збільшився, що свідчить про кращу апроксимацію та прогнозування.

Однак варто зауважити, що додаткове підвищення степеня полінома не призводило до подальшого зниження похибки. Це показує, що потрібно обережно підходити до вибору степеня полінома, оскільки надмірна складність може призвести до перенавчання моделі.

У цьому розділі було розглянуто створений інтерфейс до прогнозування та детально розказано про його роботу.

Також у цій частині було проаналізовано основні можливі небезпеки, які можуть виникнути під час роботи у вказаному кабінеті. При аналізі було виявлено, що основною небезпекою є небезпека біологічного походження. Так як лікар на пряму стикається з COVID-19. Також є загроза пожежі.

Було проведено техніко-економічне обґрунтування розробки інтерактивного додатку прогнозування захворюваності на COVID-19 за допомогою методу функціонально-вартісного аналізу.

Враховуючи це, можна зробити висновок, що вакцинація грає важливу роль у контролі та зниженні поширення COVID-19. Чим більше людей отримають вакцину, тим менша ймовірність нових захворювань.

ЗАГАЛЬНІ ВИСНОВКИ

У рамках дипломної роботи було проведено дослідження, спрямоване на аналіз літератури та використання методів прогнозування для вирішення проблематики, пов'язаної з пандемією COVID-19 та вакцинацією. Були використані методи лінійної та поліноміальної регресії для побудови моделей прогнозування. Також було проведено економічний аналіз і написано розділ БЖД.

- У розділі аналізу літератури детально розглянуто пандемію COVID-19, вакцинацію та проблематику прогнозування в цьому контексті. Зроблено висновки, що пандемія COVID-19 є актуальною проблемою, а вакцинація відіграє важливу роль у її подоланні. Також виявлено, що прогнозування може бути ефективним інструментом для стратегічного планування та контролю за поширенням захворюваності.
- У розділі матеріалів та методів було детально описано використані методи прогнозування з точки зору теорії, зокрема лінійну та поліноміальну регресію. Проаналізовано результати побудови моделей, зокрема з використанням лінійної та поліноміальної регресії.
- У практичній частині були підготовлені дані та проведений аналіз датасету. Далі були побудовані моделі лінійної та поліноміальної регресії для прогнозування. При використанні поліноміальної регресії було досягнуто таких результатів: значення середньої квадратичної помилки – 12031.45, а значення коефіцієнта детермінації – 0.72.
- Створено та описано інтерфейс інтерактивного додатку до прогнозування захворюваності COVID-19. Також було проведено характеристику приміщення, в якому будуть використовувати інтерактивний додаток, та оцінка небезпечних і шкідливих факторів, зокрема біологічної безпеки та пожежної безпеки. Було проведено розрахунок економічного ефекту за темою дипломної роботи

Загальною метою дипломної роботи було дослідження та аналіз важливих аспектів, пов'язаних з пандемією COVID-19, вакцинацією, прогнозуванням,

безпекою життєдіяльності та економічними аспектами. Завдяки проведеним дослідженням та аналізу було отримано певні висновки та рекомендації, а також створено інтерактивний додаток до прогнозування захворюваності на COVID-19, що і було темою дипломної роботи.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Гадецька З. Г., Меркотан М. Аналіз і прогнозування рівня захворюваності на covid-19 в країнах європейського регіону. *Економіка та суспільство*. 2022.
URL: <https://economyandsociety.in.ua/index.php/journal/article/view/1356>
(дата звернення: 18.05.2023).
2. Заявление по итогам второго совещания комитета по чрезвычайной ситуации в соответствии с международными медико-санитарными правилами, в связи со вспышкой заболевания, вызванного новым коронавирусом 2019 г. (ncov). *World Health Organization (WHO)*.
URL: [https://www.who.int/ru/news-room/detail/30-01-2020-statement-on-the-second-meeting-of-the-international-health-regulations-\(2005\)-emergency-committee-regarding-the-outbreak-of-novel-coronavirus-\(2019-ncov\)](https://www.who.int/ru/news-room/detail/30-01-2020-statement-on-the-second-meeting-of-the-international-health-regulations-(2005)-emergency-committee-regarding-the-outbreak-of-novel-coronavirus-(2019-ncov)) (дата звернення: 18.05.2023).
3. Карташов М. Імовірність, процеси, статистика : навч. посіб. Київ : Видавничо-поліграфіч. центр 'Київ. університет', 2008.
URL: https://probability.knu.ua/userfiles/kmv/VPS_Pv.pdf (дата звернення: 18.05.2023).
4. Українська правда. В Україні – перший випадок коронавірусу. *Українська правда*.
URL: <https://www.pravda.com.ua/news/2020/03/3/7242332/> (дата звернення: 18.05.2023).
5. Що таке коронавірус? Симптоми та профілактика. *MEDIKOM | Приватна багатопрофільна клініка в Києві*.
URL: <https://medikom.ua/koronavirus-simptomy-i-profilaktika/> (дата звернення: 18.05.2023).
6. COVID-19 data explorer. *Our World in Data*.
URL: <https://ourworldindata.org/explorers/coronavirus-data->

[explorer?facet=entity&uniformYAxis=0&country=USA~BRA~JPN~DEU&Interval=7-day+rolling+average&Relative+to+Population=true&Color+by+test+positivity=false&Metric=Confirmed+cases](https://ourworldindata.org/explorers/coronavirus-data-explorer?facet=entity&uniformYAxis=0&country=USA~BRA~JPN~DEU&Interval=7-day+rolling+average&Relative+to+Population=true&Color+by+test+positivity=false&Metric=Confirmed+cases) (date of access: 18.05.2023).

7. COVID-19 data explorer. *Our World in Data*.
URL: <https://ourworldindata.org/explorers/coronavirus-data-explorer?tab=table&zoomToSelection=true&facet=none&country=~UKR&pickerSort=asc&pickerMetric=location&Interval=7-day+rolling+average&Relative+to+Population=true&Color+by+test+positivity=false&Metric=Confirmed+cases> (date of access: 18.05.2023).
8. COVID-19 dataset. *Kaggle: Your Machine Learning and Data Science Community*.
URL: <https://www.kaggle.com/datasets/georgesaavedra/covid19-dataset> (date of access: 18.05.2023).
9. da Costa A. L. P. Mathematical modeling of the infectious diseases: key concepts and applications. *ClinMed*.
URL: <https://clinmedjournals.org/articles/jide/journal-of-infectious-diseases-and-epidemiology-jide-7-209.php?jid=jide> (date of access: 18.05.2023).
10. Davidson H. First Covid-19 case happened in November, China government records show - report. *the Guardian*.
URL: <https://www.theguardian.com/world/2020/mar/13/first-covid-19-case-happened-in-november-china-government-records-show-report> (date of access: 18.05.2023).
11. Forecasting COVID-19 outbreak progression using hybrid polynomial-Bayesian ridge regression model - Applied Intelligence. *SpringerLink*.

- URL: <https://link.springer.com/article/10.1007/s10489-020-01942-7> (date of access: 18.05.2023).
12. Lazarus J. V. Revisiting COVID-19 vaccine hesitancy around the world using data from 23 countries in 2021. *Nature communications*. 3801. URL: <https://www.nature.com/articles/s41467-022-31441-x> (date of access: 18.05.2023).
 13. Linear regression forecast (LRF) | charts help and tutorials. *Trading Technologies*. URL: <https://library.tradingtechnologies.com/trade/chrt-ti-linear-regression-forecast.html> (date of access: 18.05.2023).
 14. Linear regression for machine learning | intro to ML algorithms | edureka. *Edureka*. URL: <https://www.edureka.co/blog/linear-regression-for-machine-learning/> (date of access: 18.05.2023).
 15. Machine learning polynomial regression - javatpoint. *www.javatpoint.com*. URL: <https://www.javatpoint.com/machine-learning-polynomial-regression> (date of access: 18.05.2023).
 16. Performance evaluation of regression models for COVID-19: a statistical and predictive perspective / M. Ayoub Khan et al. *ScienceDirect*. 2021. URL: <https://www.sciencedirect.com/science/article/pii/S2090447921003385> (date of access: 18.05.2023).
 17. Preisler H. K., Westerling A. L. Statistical model for forecasting monthly large wildfire events in western united states. *Journal of applied meteorology and climatology*. 2007. URL: <https://journals.ametsoc.org/view/journals/apme/46/7/jam2513.1.xml> (date of access: 18.05.2023).
 18. Ratzan S. C. A global survey of potential acceptance of a COVID-19 vaccine. *Nature medicine*. 2021. 354. URL: <https://www.nature.com/articles/s41591-020-01226-0> (date of access: 18.05.2023).
 19. Simplilearn. What is python polynomial regression in machine learning? | simplilearn. *Simplilearn.com*. URL: <https://www.simplilearn.com/what-is->

python-polynomial-regression-in-machine-learning-article (date of access: 18.05.2023).

20. Singh N. Advantages and disadvantages of linear regression. *OpenGenus IQ: Computing Expertise & Legacy*. URL: <https://iq.opengenus.org/advantages-and-disadvantages-of-linear-regression/> (date of access: 18.05.2023).
21. Soyiri I. N., Reidpath D. D. An overview of health forecasting - environmental health and preventive medicine. *BioMed Central*. URL: <https://environhealthprevmed.biomedcentral.com/articles/10.1007/s12199-012-0294-6> (date of access: 18.05.2023).
22. WHO coronavirus (COVID-19) dashboard. *WHO Coronavirus (COVID-19) Dashboard | WHO Coronavirus (COVID-19) Dashboard With Vaccination Data*. URL: <https://covid19.who.int/> (date of access: 18.05.2023).
23. Zhang W. Predicting COVID-19 trends in Canada: a tale of four models. *The institution of engineering and technology*. 2020. URL: <https://ietresearch.onlinelibrary.wiley.com/doi/full/10.1049/ccs.2020.0017> (date of access: 18.05.2023).
24. Зарічковий О. Програмні засоби для деблюрінга зображень за допомогою генеративних змагальницьких систем : Дипломна робота. Київ, 2019. 98 с.