

**Міністерство освіти і науки України
Національний технічний університет
«Дніпровська політехніка»**

Інститут електроенергетики

(інститут)

Факультет інформаційних технологій

(факультет)

Кафедра інформаційних систем та технологій та комп'ютерної інженерії

(повна назва)

ПОЯСНЮВАЛЬНА ЗАПИСКА

кваліфікаційної роботи ступеня магістра

(бакалавра, спеціаліста, магістра)

студента Сідаш Катерини Андріївни

(ПІБ)

академічної групи 126м-23з-1

(шифр)

спеціальності 126 Інформаційні системи та технології

(код і назва спеціальності)

за освітньо-професійною програмою 126 Інформаційні системи та технології

(офіційна назва)

на тему «Комплексне використання методів K-means кластеризації та ARIMA для моделювання тенденцій смертності в Європі від інфекційних захворювань»

—

(назва за наказом ректора)

Керівники	Прізвище, ініціали	Оцінка за шкалою		Підпис
		рейтингово ю	інституційною	
кваліфікаційної роботи	проф. Нікулін С.Л.			
розділів:				
Рецензент				
Нормоконтролер	проф. Коротенко Г.М.			

Дніпро

2024

ЗАТВЕРДЖЕНО:завідувач кафедри
інформаційних технологій
та комп'ютерної інженерії
(повна назва)Гнатушенко В.В.
(підпис) (прізвище, ініціали)

«_» _____ 202_ року

ЗАВДАННЯ
на кваліфікаційну роботу
ступеня магістрстуденту Сідаш К.А.
(прізвище та ініціали)академічної групи 126м-23з-1
(шифр)спеціальності _____ 126 «Інформаційні системи та технології»
за освітньо-професійною програмою 126 «Інформаційні системи та технології»

(офіційна назва)

на тему «Комплексне використання методів K-means кластеризації та ARIMA для моделювання тенденцій смертності в Європі від інфекційних захворювань»затверджену наказом ректора НТУ «Дніпровська політехніка» від 17.10.2024 р.
№1389-с

Розділ	Зміст	Термін виконання
Розділ 1	Провести огляд предметної області.	11.09.2024
Розділ 2	Розробка методології моделювання тенденцій смертності з використанням методів K-means кластеризації та ARIMA.	16.10.2024
Розділ 3	Розробити програмну реалізація моделювання тенденцій смертності з використанням методів K-means кластеризації та ARIMA. Провести тестування запропонованої методології	06.11.2024

Завдання видано _____
(підпис керівника)проф. Нікулін С.Л.
(прізвище, ініціали)Дата видачі 07 вересня 2024 р.

Дата подання до екзаменаційної комісії _____ р.

Прийнято до виконання _____
(підпис студента)Сідаш К.А.
(прізвище, ініціали)

РЕФЕРАТ

Пояснювальна записка: 69 с., 17 рис., 1 табл., 20 джерел, 2 додаток.

Об'єкт дослідження: статистичні дані смертності населення в Європі від інфекційних захворювань.

Предмет дослідження: комплексне використання методів K-means кластеризації та ARIMA для моделювання і прогнозування тенденцій смертності від інфекційних захворювань.

Мета кваліфікаційної роботи: розробка комплексного підходу до аналізу та прогнозування тенденцій смертності в Європі, використовуючи методи K-means кластеризації та ARIMA.

Наукова новизна полягає в інтеграції методів K-means кластеризації та ARIMA для комплексного аналізу та прогнозування смертності в Європі. Цей підхід дозволяє отримати нові знання про вплив різних факторів на смертність, а також покращити точність прогнозування за рахунок використання сучасних статистичних моделей і алгоритмів машинного навчання. Дослідження також підкреслює важливість географічного контексту у вивченні проблеми смертності.

Практична значимість полягає у можливості використання розроблених методів та архітектури програмного забезпечення для аналізу даних про смертність в реальному часі. Результати роботи можуть бути використані фахівцями у сфері охорони здоров'я для прийняття обґрунтованих рішень, розробки рекомендацій для покращення здоров'я населення та зниження показників смертності від інфекційних захворювань. Також результати можуть бути корисними для державних органів у розробці політик у сфері охорони здоров'я на основі даних.

Ключові слова: K-means кластеризація, ARIMA, моделювання, прогнозування, кластеризація даних, інфекційні захворювання.

ABSTRACT

The explanatory statement has 69 p., 17 figures, 1 table, 20 sources, and 2 appendix.

The research object is statistical data on mortality in Europe from infectious diseases.

The subject of research: the integrated use of K-means clustering and ARIMA methods for modeling and forecasting mortality trends from infectious diseases.

The qualification work aims to develop an integrated approach to the analysis and forecasting of mortality trends in Europe using K-means clustering and ARIMA methods.

The scientific novelty lies in integrating K-means clustering and ARIMA methods for comprehensive analysis and forecasting of mortality in Europe. This approach allows us to gain new knowledge about the impact of various factors on mortality, as well as to improve the accuracy of forecasting through the use of modern statistical models and machine learning algorithms. The study also emphasizes the importance of the geographical context in studying mortality.

The practical significance lies in the possibility of using the developed methods and software architecture to analyze mortality data in real-time. Healthcare professionals can use the results of the work to make informed decisions and develop recommendations for improving public health and reducing mortality from infectious diseases. The results can also be helpful for government agencies in developing data-driven health policies.

Keywords: K-means clustering, ARIMA, modeling, forecasting, data clustering, infectious diseases.

ЗМІСТ

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ, СИМВОЛІВ, СКОРОЧЕНЬ І ТЕРМІНІВ.....	7
ВСТУП.....	8
1 АНАЛІЗ СТАНУ ОБЛАСТІ РІШЕННЯ ЗАДАЧІ.....	10
1.1 Епідеміологія та процеси поширення інфекцій.....	10
1.2 Історія виникнення та поширення найбільших пандемій у світі ХХ-ХХІ ст.....	12
1.3 Огляд наукових робіт із поширення епідемій.....	16
1.4 Принципи моделювання епідемії.....	17
1.5 Висновки.....	23
2 МЕТОДОЛОГІЯ АНАЛІЗУ ТА ПРОГНОЗУВАННЯ СМЕРТНОСТІ ВІД ІНФЕКЦІЙНИХ ЗАХВОРЮВАНЬ.....	24
2.1 Сімейство моделей смертності за місцем проживання.....	24
2.2 Модель Гомперца.....	25
2.3 Модельно-орієнтована кластеризація.....	26
2.3.1 Метод К-середніх.....	26
2.3.2 Кластеризація з максимізацією математичного сподівання.....	28
2.3.3 Ієрархічна кластеризація.....	29
2.4 Метод ARIMA.....	29
2.4 Прогнозування очікуваної тривалості життя.....	31
2.5 Методологія моделювання тенденцій смертності з використанням методів K-means кластеризації та ARIMA.....	32
2.6 Висновки до розділу.....	36
3 ПРОГРАМНА РЕАЛІЗАЦІЯ МОДЕЛЮВАННЯ ТЕНДЕНЦІЙ СМЕРТНОСТІ З ВИКОРИСТАННЯМ МЕТОДІВ K-MEANS КЛАСТЕРИЗАЦІЇ ТА ARIMA.....	38
3.1 Архітектура програмного забезпечення.....	38
3.2 Модуль обробки даних.....	43
3.3 Модуль кластеризації.....	45
3.4 Модуль прогнозування смертності.....	47
3.5 Інтерфейс користувача.....	48
3.6 Оцінка результатів комплексного використання методів K-means та ARIMA.....	53
3.7 Висновки до розділу.....	59
ВИСНОВКИ.....	61

ПЕРЕЛІК ПОСИЛАНЬ.....	63
Додаток А. Фрагмент лістингу програми.....	65
Додаток Б. Графічні дані.....	67

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ, СИМВОЛІВ, СКОРОЧЕНЬ І ТЕРМІНІВ

LS – моделі смертності за місцем проживання;

GMM – допомогою моделей гауссової суміші;

EM – математичне сподівання;

ARIMA – авторегресійна інтегрована модель ковзної середньої;

ACF – графік автокореляції ;

PACF – часткова автокореляція;

ПЗ – програмне забезпечення;

GUI – Графічний інтерфейс користувача;

MSE – середня квадратна помилка

ВСТУП

Інфекційні хвороби, як старі, так і нові, є однією з найбільших загроз для здоров'я людей у всьому світі. Вони є причиною значної кількості смертей, особливо серед дітей, та становлять серйозну проблему для систем охорони здоров'я багатьох країн. Поява нових, висококонтагіозних вірусів, таких як COVID-19, підкреслює актуальність цієї проблеми.

За даними Світового банку, інфекційні захворювання є причиною половини смертей дітей до 5 років у світі. До них належать хвороби дихальної системи, кишкові інфекції, кір, малярія, СНІД та інші. Також, 80% усіх випадків захворюваності у цій віковій групі припадає на інфекційні хвороби [1]. Попри досягнення сучасної медицини, інфекційні хвороби залишаються важливою проблемою медичного, соціального та економічного характеру. Для України це питання особливо актуальне в умовах економічної кризи, коли контроль над епідеміологічною ситуацією та управління інфекційною захворюваністю, зокрема під час пандемії COVID-19, стають питанням національної безпеки [2]. За даними ВООЗ, у деяких країнах світу інфекційні захворювання є другою за значущістю причиною смертності. Наприкінці ХХ та на початку ХХІ століття ситуація з такими захворюваннями, як чума, холера та жовта лихоманка, значно ускладнилася. Крім того, з'явилося понад 30 нових, раніше невідомих, але небезпечних хвороб, таких як COVID-19, геморагічні лихоманки Ласса, Ебола, Марбург, ВІЛ-інфекція, легіонельоз, пташиний та свинячий грип [3].

Об'єктом досліджень є статистичні дані смертності населення в Європі від інфекційних захворювань.

Предметом дослідження – комплексне використання методів K-means кластеризації та ARIMA для моделювання і прогнозування тенденцій смертності від інфекційних захворювань.

Мета роботи – розробка комплексного підходу до аналізу та прогнозування тенденцій смертності в Європі, використовуючи методи K-means кластеризації та ARIMA.

Для досягнення цієї мети необхідно вирішити наступні завдання:

- виконати збір та підготовка даних про смертність в Європі, зокрема щодо інфекційних захворювань, попередній аналіз даних для виявлення аномалій, застосування моделі Гомперца для аналізу вікових і гендерних відмінностей у смертності;

- дослідити метод кластеризації K-середніх для групування населення за рівнем смертності;

- розробити методологію моделювання тенденцій смертності з використанням методів K-means кластеризації та ARIMA;

- розробити архітектуру програмного забезпечення для автоматизації аналізу та прогнозування, а також оцінка ефективності комбінованого підходу до прогнозування смертності.

Наукова новизна результатів дослідження. Наукова новизна роботи полягає в інтеграції методів K-means кластеризації та ARIMA для комплексного аналізу та прогнозування смертності в Європі. Цей підхід дозволяє отримати нові знання про вплив різних факторів на смертність, а також покращити точність прогнозування за рахунок використання сучасних статистичних моделей і алгоритмів машинного навчання. Дослідження також підкреслює важливість географічного контексту у вивченні проблеми смертності.

1 АНАЛІЗ СТАНУ ОБЛАСТІ РІШЕННЯ ЗАДАЧІ

1.1 Епідеміологія та процеси поширення інфекцій

Епідеміологія – це наука, яка вивчає, чому виникають спалахи хвороб та як вони поширюються серед населення. Вона допомагає розробити стратегії боротьби з цими захворюваннями. Основний інтерес епідеміологів – це процес поширення інфекцій, який починається з того моменту, коли збудник хвороби потрапляє в організм людини, [3].

Важливо розуміти, що епідеміологія відрізняється від звичайної медицини. Якщо лікар лікує окрему людину, то епідеміолог дивиться на хворобу з більш широкого погляду: він вивчає, як хвороба впливає на ціле населення. Наприклад, для епідеміолога кількість смертей від хвороби може бути важливим показником того, наскільки широко вона поширилася. Для лікаря ж смерть пацієнта – це завжди трагедія.

Мета епідеміології – це глибоке розуміння хвороби. Епідеміологи намагаються відповісти на такі питання: як хвороба передається від однієї людини до іншої? Які групи населення найбільше схильні до захворювання? Які фактори сприяють поширенню хвороби? Відповіді на ці питання допомагають розробити ефективні заходи профілактики та боротьби з хворобами.

Інфекція – це процес проникнення мікроорганізму в організм людини або тварини та їх подальшої взаємодії. Ця взаємодія залежить від властивостей як самого організму, так і мікроба. Як зазначав І.І. Мечников, інфекцію можна розглядати як боротьбу між двома організмами.

Інфекційний процес – це більш широке поняття, яке включає всі захисні та патологічні реакції організму на проникнення збудника. Це складний комплекс реакцій, спрямованих на боротьбу з інфекцією та відновлення порушеного гомеостазу. Сучасні вчені, такі як В.І. Покровський,

підкреслюють, що інфекційний процес – це взаємодія організму і мікроба, спрямована на відновлення рівноваги.

Інфекційна хвороба – це кінцевий результат інфекційного процесу, коли симптоми захворювання стають очевидними. Це можуть бути різноманітні прояви: від легких нездужань до тяжких захворювань.

Таким чином, інфекція, інфекційний процес і інфекційна хвороба – це взаємопов'язані поняття, які описують різні етапи взаємодії організму і мікроба.

Розуміння причин виникнення та поширення інфекційних хвороб є одним із ключових завдань сучасної епідеміології. Л.В. Громашевський, видатний епідеміолог, звернув увагу на те, що механізм передачі інфекцій значною мірою залежить від соціальних умов життя населення. Це означає, що наш спосіб життя, рівень життя, культурні особливості та інші соціальні фактори впливають на те, як швидко поширюються хвороби і якими вони є [3].

Соціальні зміни можуть суттєво впливати на поширення інфекцій. Наприклад, міграція населення може призвести до появи нових штамів вірусів в регіонах, де вони раніше не зустрічались. Або, розвиток нових медичних технологій, таких як переливання крові, може стати причиною появи нових шляхів передачі інфекцій, [4].

Інфекційні хвороби мають різні шляхи передачі, кожен з яких залежить від певних факторів. Так, поширення інфекцій дихальних шляхів впливають щільність населення, рівень вакцинації та міграційні процеси. Для кишкових інфекцій ключовими є якість води, санітарні умови та харчові звички. Інфекції, що передаються через укуси комах, більш поширені в регіонах з певним кліматом та залежать від стану природного середовища. Інфекції, що передаються при контакті, тісно пов'язані з соціальними нормами, поведінкою людей та медичними процедурами.

Отже, успішність боротьби з інфекційними хворобами залежить від багатьох чинників, серед яких важливу роль відіграють як природні умови, так і соціальні фактори.

1.2 Історія виникнення та поширення найбільших пандемій у світі XX-XXI ст.

XX століття стало свідком кількох масштабних пандемій грипу, які завдали значної шкоди здоров'ю та економіці людства. Найбільш руйнівною серед них стала пандемія іспанського грипу (1918-1919 рр.), викликана вірусом H1N1. Незважаючи на назву, вірус, ймовірно, зародився у військових таборах Франції, де тісний контакт людей із тваринами сприяв його мутації та поширенню. За рік пандемії захворіло близько 29.5% населення планети, а кількість загиблих оцінюється в 50-100 мільйонів осіб.

Після іспанки відбулися ще кілька масштабних спалахів грипу: азіатський (1957-1958 рр.), гонконгський (1968-1969 рр.), пташиний (2003-2008 рр.) та свинячий (2009-2010 рр.). Цікаво, що перед кожною з цих пандемій спостерігалися локальні спалахи грипу серед військових, що свідчить про можливу роль армій у поширенні вірусів. Дослідження вказують на те, що природними резервуарами вірусів грипу є дикі птахи, а свині можуть відігравати роль проміжних хазяїв, сприяючи мутації вірусу і його передачі людині.

Загалом, пандемії грипу XX століття продемонстрували високу здатність вірусу до мутації та швидкого поширення в глобалізованому світі. Незважаючи на досягнення сучасної медицини, загроза нових пандемій залишається актуальною, що підкреслює необхідність постійного моніторингу за вірусами грипу та розробки ефективних засобів профілактики і лікування.

ВІЛ/СНІД – це хронічне, зазвичай смертельне захворювання, спричинене вірусом імунодефіциту людини (ВІЛ). Його походження сягає Центральної Африки, де вірус циркулював серед мавп ще на початку XX

століття. Перші випадки захворювання серед людей були зафіксовані в США на початку 1980-х років. Швидке поширення ВІЛ по всьому світу перетворило епідемію на глобальну проблему.

ВІЛ атакує імунну систему людини, роблячи її вразливою до різноманітних інфекцій. Хвороба протікає безсимптомно протягом тривалого часу, але згодом призводить до розвитку синдрому набутого імунодефіциту (СНІД). На сьогоднішній день не існує вакцини від ВІЛ, оскільки вірус постійно мутує, ускладнюючи розробку ефективних ліків. Однак антиретровірусна терапія дозволяє пригнічувати активність вірусу та продовжувати життя хворих.

Незважаючи на значні зусилля світової спільноти, ВІЛ/СНІД залишається однією з найбільших проблем сучасності. За оцінками, близько 40 мільйонів людей на планеті інфіковані ВІЛ, а загальна кількість смертей від цієї хвороби перевищила 35 мільйонів.

Основними шляхами передачі ВІЛ є незахищений статевий контакт, ін'єкційне вживання наркотиків та передача вірусу від інфікованої матері до дитини під час вагітності, пологів або грудного вигодовування.

Профілактика ВІЛ включає безпечний секс, використання стерильних інструментів при ін'єкціях, а також ранню діагностику та лікування інфекції.

Вірус Ебола, що викликає геморагічну лихоманку, вперше був виявлений у 1976 році в Судані. Ця смертельно небезпечна хвороба, що характеризується високою температурою, внутрішніми кровотечами та часто летальним наслідком, з того часу періодично спалахує в різних регіонах Африки.

Природними носіями вірусу Ебола вважаються фруктові кажани, від яких він може передаватися іншим тваринам, а потім і людям. Найчастіше зараження відбувається при контакті з інфікованими тваринами або хворими людьми. Спалахи епідемій, як правило, пов'язані з порушенням природних екосистем та контактом людей з дикими тваринами.

Найбільш масштабна епідемія Еболи відбулася в 2014-2015 роках, охопивши кілька країн Західної Африки. Ця епідемія продемонструвала високу контагіозність вірусу та здатність швидко поширюватися на великі території.

Вірус Ебола постійно мутує, що ускладнює розробку ефективних вакцин і ліків. Нові штами вірусу можуть виявлятися в різних регіонах світу, що підкреслює глобальний характер цієї проблеми. Нещодавнє виявлення нового штаму вірусу Ебола на Філіппінах, який здатний інфікувати як людей, так і свиней, викликає особливу стурбованість.

Незважаючи на значні зусилля науковців і медиків, вірус Ебола залишається серйозною загрозою для глобального здоров'я. Для ефективної боротьби з цією хворобою необхідні подальші дослідження, розробка нових вакцин і ліків, а також посилення систем моніторингу та реагування на спалахи інфекційних захворювань.

На початку XXI століття світ зіткнувся з кількома небезпечними спалахами інфекцій, викликаних коронавірусами. Серед них особливо виділяються SARS (тяжкий гострий респіраторний синдром) та MERS (близькосхідний респіраторний синдром).

SARS, що виник у Китаї в 2002 році, був спричинений коронавірусом SARS-CoV. Вважається, що вірус передався людям від кажанів через проміжного господаря. Епідемія швидко поширилася по світу, спричинивши тисячі випадків захворювання та смертей.

MERS, виявлений у 2012 році на Близькому Сході, також був викликаний коронавірусом (MERS-CoV). Хоча верблюди були ідентифіковані як вторинні резервуари вірусу, точне джерело його походження досі не встановлено. Цей факт ускладнює розробку ефективних стратегій боротьби з MERS.

Обидва ці спалахи продемонстрували високу патогенність коронавірусів і здатність швидко поширюватися в глобалізованому світі. Вони стали серйозним викликом для систем охорони здоров'я та підкреслили

необхідність розробки ефективних засобів діагностики, лікування та профілактики нових інфекційних захворювань. Досвід, отриманий під час боротьби з SARS та MERS, став важливим уроком для світової спільноти перед спалахом пандемії COVID-19, спричиненої ще одним коронавірусом – SARS-CoV-2.

Ці два спалахи продемонстрували, що коронавіруси становлять серйозну загрозу для глобального здоров'я. Тому важливо продовжувати дослідження в цій галузі та розробляти ефективні стратегії боротьби з новими інфекційними захворюваннями.

Коронавіруси, відомі науці з 1965 року, постійно мутують, що призводить до появи нових штамів, таких як SARS-CoV та MERS-CoV. Ці віруси часто передаються від тварин до людей, як це сталося з COVID-19, де кажани вважаються одним із можливих природних резервуарів. Пандемія COVID-19 стала однією з найбільших за останні століття, інфікувавши мільйони людей та спричинивши сотні тисяч смертей.

Вірус Крим-Конго геморагічної лихоманки (ККГЛ), вперше виявлений в Криму в 1944 році, є серйозною загрозою для здоров'я людини. Переноситься він переважно іксодовими кліщами роду *Hyalomma*, які можуть заражати людей під час укусу або через контакт з кров'ю інфікованих тварин.

Зміна клімату створює сприятливі умови для поширення ККГЛ. Потепління клімату дозволяє кліщам розширювати свій ареал на північ, а перелітні птахи можуть сприяти їхньому поширенню на нові території. Крім того, вірус ККГЛ може мутувати, що дозволить йому адаптуватися до нових умов і заражати інших переносників, таких як кліщі, які раніше не були сприйнятливі до цього вірусу.

Зміна сільського господарства та урбанізація також сприяють поширенню вірусних захворювань. Прикладом можуть слугувати вірус Рифтової долини, лихоманка Денге та Чикунгунья, які в останні роки активно поширюються в нові регіони. Ці хвороби, пов'язані з комарами та кажанами,

процвітають в умовах зміни клімату та створення нових біотопів, сприятливих для їх розмноження.

Таким чином, глобальні зміни клімату, зміни в сільському господарстві та урбанізація створюють сприятливі умови для поширення вірусу ККГЛ та інших інфекційних захворювань. Це вимагає розробки нових стратегій для їх профілактики та контролю.

1.3 Огляд наукових робіт із поширення епідемій

Моделювання поширення інфекційних хвороб, особливо з урахуванням випадкових факторів, стало потужним інструментом у боротьбі з епідеміями. З появою комп'ютерів у середині ХХ століття цей напрямок досліджень отримав значний розвиток. Сучасні стохастичні моделі дозволяють враховувати різноманітні фактори, що впливають на поширення хвороб, та прогнозувати їхній розвиток. Однак, точність таких моделей безпосередньо залежить від якості та кількості доступних даних. Наприклад, для побудови репрезентативної моделі COVID-19 необхідно мати достатню кількість даних про зараження, що ускладнюється експоненціальним зростанням кількості випадків. Тому моделі, побудовані на ранніх етапах епідемії, можуть значно відрізнятися від моделей, що базуються на більш пізніх даних.

Моделі поширення хвороб ґрунтуються на певних припущеннях, наприклад, про однорідність реакції людей на захворювання або про стабільність темпів поширення в різних регіонах. Ці припущення можуть як покращити, так і погіршити точність прогнозів. Отримані моделюванні результати, такі як кількість хворих чи померлих, слід розглядати не як точні цифри, а як приблизні оцінки, що можуть коливатися в певному діапазоні. Розуміння обмежень моделей є ключовим для їхньої правильної інтерпретації. Правильно обрана модель - це важливий інструмент, але не гарантія абсолютно точного прогнозу.

У деяких дослідженнях автори експериментують з різними сценаріями поширення COVID-19, варіюючи кількість інфікованих та базове

репродуктивне число (R_0), щоб оцінити вплив цих факторів на прогнози. Стаття [5] демонструє, як введення карантину та соціального дистанціювання впливає на динаміку епідемії, використовуючи модель SIR. Також автори порівнюють різні моделі поширення COVID-19 в Італії та обґрунтовують вибір оптимальної моделі для конкретних умов. Цікавим підходом до обліку невизначеності є використання діапазону можливих значень кількості інфікованих замість однієї кривої, як це зробили вчені з університету Сантьяго, [5].

Різні моделі прогнозування епідемії можуть надавати суттєво різні прогнози навіть за однакових початкових даних. Короткострокові прогнози, як правило, точніші, але довгострокові можуть бути менш надійними через непередбачуваність розвитку епідемії. Вибір моделі залежить від того, наскільки важлива для нас точність короткострокових прогнозів або розуміння загальної тенденції розвитку епідемії.

Серед моделей поширення інфекційних хвороб модель SIR є найбільш відомою, але існують її численні модифікації, такі як SIRS (з урахуванням повторного зараження) та SEIR (з урахуванням латентного періоду), які дозволяють більш детально моделювати різні типи захворювань.

На сьогоднішній день існує велика різноманітність моделей, що надають різні результати. Найбільш ефективними вважаються моделі, які поєднують кількісні дані про захворюваність та смертність з якісним експертним аналізом.

1.4 Принципи моделювання епідемії

Незважаючи на різноманіття наукових моделей, розроблених для дослідження поширення інфекційних хвороб, всі вони базуються на загальному принципі: описі динаміки переходу індивідів між трьома основними епідеміологічними станами: сприйнятливі (S), інфекційні (I) та відновлені (або померлі) (R). Цей підхід, відомий як модель SIR, дозволяє

аналізувати швидкість поширення інфекції та визначати ключові фактори, що впливають на цей процес.

Епідеміологічні моделі, що існують вже майже століття, є потужним інструментом для прогнозування та розуміння поширення інфекційних хвороб. Подібно до того, як метеорологи моделюють атмосферні процеси для прогнозування погоди, епідеміологи використовують математичні моделі для симуляції поширення інфекцій у популяції.

Найпростіші моделі, такі як SIR, поділяють людей на три категорії: сприйнятливі, інфекційні та одужані. Хоча вони спрощують реальність, такі моделі дають базове розуміння процесу поширення інфекції. Більш складні моделі враховують різноманітні фактори, такі як вік, стать, стан здоров'я та соціальні контакти, що дозволяє точніше моделювати реальні епідемії.

Ці складніші моделі можуть включати категорії недіагностованих випадків та враховувати обмеження тестування. За допомогою цих моделей можна оцінити вплив різних факторів на поширення інфекції та спрогнозувати розвиток епідемії за різними сценаріями. Хоча деякі параметри моделей можуть бути невідомими точно, їх можна оцінити приблизно і перевірити, як зміняться прогнози при зміні цих параметрів.

Таким чином, епідеміологічні моделі відіграють важливу роль у боротьбі з інфекційними хворобами, дозволяючи вченим та політикам приймати обґрунтовані рішення щодо заходів контролю та запобігання епідемії.

Щоб ефективно прогнозувати поширення інфекційних хвороб, вчені створюють детальні комп'ютерні моделі, які імітують реальний світ. Ці моделі враховують безліч факторів, таких як щільність населення, вік людей, транспортні зв'язки та стан охорони здоров'я. За допомогою математичних рівнянь, вчені моделюють взаємодію людей у просторі та часі, створюючи своєрідний віртуальний світ.

У цей віртуальний світ вводять вірус і спостерігають, як він поширюється. Однак, для створення таких моделей потрібні значні

спрощення та припущення. Наприклад, вчені повинні оцінити, скільки людей помре від хвороби (рівень смертності) і скільки нових людей, в середньому, заразить одна хвора людина (базове репродуктивне число, R_0). Ці значення можуть сильно відрізнятись для різних вірусів і популяцій.

Наприклад, під час пандемії COVID-19 вчені оцінили, що для населення Великої Британії рівень смертності від цієї хвороби становив близько 0,9%. Вони також припустили, що одна інфікована людина в середньому заражала від 2 до 2,6 інших людей. Крім того, вчені зробили припущення про період інкубації вірусу та інші важливі характеристики захворювання.

Хоча такі моделі не можуть передбачити майбутнє з абсолютною точністю, вони є потужним інструментом для розуміння епідемій та розробки ефективних стратегій боротьби з ними. За допомогою цих моделей можна оцінити вплив різних заходів, таких як вакцинація та соціальне дистанціювання, на поширення інфекції.

Прогнозування поширення інфекційних хвороб є важливою галуззю досліджень, яка активно розвивається з початку ХХ століття. Завдяки сучасним технологіям та великим обсягам даних, науковці можуть створювати все більш детальні та точні моделі для прогнозування епідемій. Наприклад, доктор Елісон Хілл розробила інтерактивний додаток, який дозволяє досліджувати різні сценарії поширення інфекцій.

Основи сучасного епідеміологічного моделювання були закладені ще на початку ХХ століття такими вченими, як Вільям Гамер, Рональд Росс та Ловелл Рід з Вейдом Гемптоном Фростом. Вони розробили перші математичні моделі, які описували, як інфекційні хвороби поширюються в популяції. Однією з найвідоміших моделей є модель Кермака-МакКендріка, яка і досі широко використовується для дослідження епідемій.

Ці ранні моделі, а також їхні сучасні модифікації, дозволяють нам зрозуміти складні процеси, що лежать в основі поширення інфекційних хвороб. За допомогою цих моделей можна вивчати вплив різних факторів на

поширення епідемій, прогнозувати їх розвиток та розробляти ефективні стратегії боротьби з ними.

Сучасні епідеміологічні моделі дозволяють враховувати велику кількість факторів, таких як вік, стать, соціальні контакти, географічні особливості та стан системи охорони здоров'я. Це дозволяє створювати більш реалістичні моделі та точніше прогнозувати розвиток епідемій.

Складна динаміка епідемічних процесів вимагає застосування більш складних математичних моделей, таких як статистичний аналіз часових рядів. Найпоширенішими статистичними методами для моделювання динаміки епідемій є методи ковзного середнього та їх модифікації. Зокрема, метод простого ковзного середнього активно застосовується для аналізу поширення COVID-19 у різних країнах, таких як Ірак, США, Пакистан, Італія, Іспанія та Китай. Широко використовуються також експоненціальне ковзне середнє, моделі ARIMA та SARIMA. Ці методи дозволяють швидко оцінити ситуацію під час спалахів нових захворювань, таких як атипова пневмонія, грип H1N1, Ебола, ящур та COVID-19. Особливістю таких аналізів є необхідність робити висновки на ранніх стадіях спалаху, коли дані обмежені і можуть бути неповними.

Найбільшу точність сьогодні демонструють методи дослідження епідемічних процесів на основі машинного навчання. Серед них можна виділити кілька основних задач.

Перш за все, це регресія. Регресійний метод є підходом до прогнозування, який дозволяє досліджувати взаємозв'язок між залежними та незалежними змінними. Використовуючи різні параметри, наприклад, захворюваність, погоду, рівень гетерогенності населення, летальність, вакцинацію та обмежувальні заходи, моделі регресійного аналізу дають змогу прогнозувати вплив цих факторів на епідемічні процеси. Хоча кореляція не завжди свідчить про причинно-наслідковий зв'язок, інтерпретація отриманих кореляційних даних є важливою задачею. Залежно

від характеру даних і цілей дослідження можуть застосовуватися такі види регресій, як лінійна, логістична, поліноміальна, Рідж та Лассо-регресії.

Крім регресії, широко використовується класифікація. Ця задача полягає в розділенні об'єктів за попередньо визначеними класами. У сфері епідеміологічних досліджень класифікація застосовується для різноманітних цілей: визначення поведінкових груп населення, ідентифікація просторово-часових особливостей епідемічного процесу, аналіз наявної інформації про спалахи захворювань, визначення кліматичних зон, які впливають на захворюваність, та географічних зон залежно від специфіки різних інфекційних захворювань.

Ще одним важливим завданням є кластеризація. Вона використовується для розподілу об'єктів, коли класи не визначені заздалегідь, а формування кластерів здійснюється на основі схожості певних характеристик елементів. У дослідженнях епідемічних процесів кластеризація допомагає визначити географічні території з подібними ознаками епідемічного процесу, ідентифікувати епідемічні спалахи, групи носіїв інфекції, філогенетичні особливості популяцій та патерни поширення інфекцій.

Зменшення розмірності — це метод, що зводить велику кількість характеристик до меншої, зручної для подальшого аналізу. Це актуально в епідеміології, де великі обсяги даних можуть містити зайві відомості. Зменшення розмірності дозволяє відсіяти непотрібні дані, знизити обчислювальну складність, виділити інформативні ознаки та фактори, які мають найбільший вплив на епідемічні процеси.

Виявлення аномалій — ще одна важлива задача. Вона спрямована на виявлення рідкісних відхилень, які відрізняються від загальної тенденції. Задача схожа на класифікацію, але відмінність полягає в тому, що аномальні випадки є рідкісними, тому навчальні вибірки можуть бути невеликими або відсутні зовсім. У епідеміології методи виявлення аномалій

використовуються для моніторингу захворюваності в реальному часі, аналізу тенденцій у потоках захворюваності та ідентифікації спалахів епідемії.

Нарешті, прогнозування є однією з найважливіших задач у дослідженні епідемічних процесів. При побудові прогнозів враховується соціальна динаміка, характер спалаху, географічні особливості, ранні етапи розвитку епідемічного процесу, вплив на населення, а також дані з соціальних мереж, пошукових запитів, інформація від мобільних операторів та багато інших факторів.

Основним недоліком методів машинного навчання у дослідженні епідемічних процесів є складність інтерпретації отриманих результатів. Зазвичай важко визначити чинники, що впливають на динаміку захворюваності, а також провести експерименти для оцінки ефективності превентивних заходів.

Таким чином, епідеміологічне моделювання є незамінним інструментом для розуміння та боротьби з інфекційними хворобами. Завдяки постійному розвитку технологій та накопиченню нових даних, ці моделі стають все більш досконалими та дозволяють нам ефективніше захищати здоров'я населення.

Прогнозування пандемій — це складний процес, який залежить від багатьох факторів. Наприклад, точність прогнозу погоди значно вища, ніж прогнозу цін на акції, через відмінності в наявності даних, розумінні процесів та впливу прогнозів на реальність.

Щоб зробити точні прогнози, необхідно враховувати такі фактори:

1. Якість даних: чим більше ми знаємо про поширення захворювання, тим точніші будуть прогнози.
2. Розуміння процесів: чим краще ми розуміємо, як вірус передається і як на нього впливають різні фактори, тим точніші будуть моделі.
3. Вплив прогнозів: прогнози можуть змінювати поведінку людей та урядів, що може вплинути на хід епідемії.

Спрощені моделі, які не враховують всі ці фактори, можуть давати неточні результати і навіть призводити до неправильних рішень.

1.5 Висновки

Історія людства нерозривно пов'язана з епідеміями, які неодноразово змінювали хід історії. Хоча в минулому людство мало обмежені можливості боротьби з інфекційними захворюваннями, розвиток медицини та санітарії дозволив значно знизити ризик спалахів епідемій. Однак, незважаючи на досягнення сучасної медицини, загроза нових пандемій залишається актуальною.

Зміна клімату, урбанізація та глобалізація створюють сприятливі умови для поширення інфекційних хвороб. Віруси, які раніше циркулювали в диких тваринах, все частіше передаються людям. Швидка міграція населення сприяє швидкому поширенню інфекцій по всьому світу. Крім того, багато вірусів здатні швидко мутувати, що ускладнює розробку ефективних вакцин та ліків.

Пандемія COVID-19 продемонструвала, що навіть розвинені країни можуть бути не готові до масштабних епідемій. Складність прогнозування та боротьби з новими вірусами пов'язана з нестачею даних, швидкістю мутацій вірусів та здатністю вірусів адаптуватися до нових умов. Математичні моделі, які використовуються для прогнозування поширення інфекцій, є важливим інструментом, але вони мають свої обмеження і можуть бути неточними.

2 МЕТОДОЛОГІЯ АНАЛІЗУ ТА ПРОГНОЗУВАННЯ СМЕРТНОСТІ ВІД ІНФЕКЦІЙНИХ ЗАХВОРЮВАНЬ

2.1 Сімейство моделей смертності за місцем проживання

Моделі смертності за місцем проживання, або моделі зі шкалою місцезнаходження (LS), відіграють важливу роль у прогнозуванні тривалості життя та аналізі демографічних даних. Такий підхід часто використовується в теорії надійності та для аналізу життєвих тривалостей в популяційних дослідженнях. Моделі цього сімейства дозволяють аналізувати, як зміни параметрів, що відповідають за місце проживання і масштаби смертності, впливають на загальну тривалість життя.

Нехай X – неперервна випадкова величина, що представляє час до настання події (наприклад, смерть). Ми говоримо, що X належить до сімейства моделей LS, якщо її функція щільності ймовірності $f(x)$ може бути виражена у вигляді [6]:

$$f(x) = \frac{1}{c} f_{LS} \left(\frac{x - u}{c} \right) \quad (2.1)$$

де параметри u і c визначають положення та мінливість функції щільності ймовірності відповідної моделі. Стандартна форма функції щільності ймовірності отримується, коли $u=0$ і $c=1$, що дозволяє аналізувати модель у її базовому вигляді.

Крім моделей LS, також існує сімейство моделей зі шкалою логарифмічного масштабу (LLS). У цьому випадку функція щільності ймовірності для випадкової величини X може бути виражена наступним чином [6]:

$$f(x) = \frac{1}{c} f_{LLS} \left(\frac{\log(x) - u}{c} \right) \quad (2.2)$$

Параметри місцезнаходження u та масштабу c у моделях смертності LS і LLS мають чітку інтерпретацію. Параметр u визначає зсув функції щільності вздовж осі x . Збільшення u призводить до зміщення функції вправо, що відповідає пізнішому настанню смертності, тоді як зменшення u зсуває функцію вліво, вказуючи на раннє настання смертності. Таким чином, зміна u відображає зсув вікових груп, у яких спостерігається пік смертності, без змін у формі самої функції.

Параметр масштабу c впливає на розтягнутість або стискування функції щільності. Збільшення значення c робить функцію більш розподіленою, що вказує на більшу варіативність віку, в якому настає смерть. Зменшення c стискає функцію, вказуючи на те, що смертність сконцентрована у вузькому віковому діапазоні.

2.2 Модель Гомперца

Модель Гомперца є однією з найпоширеніших у демографічному та актуарному аналізі. Як вже зазначалося, у 1825 році Гомперц виявив, що сила смертності (ймовірність смерті) зростає з віком за експоненціальним законом для більшості вікових груп (за винятком дитинства і старості). Модель Гомперца формулюється наступним чином [7]:

$$\mu(x) = ae^{bx}, \quad x \geq 0, \quad (2.3)$$

де $a > 0$ – початкова сила смертності, а b – параметр, який описує темп старіння.

З цього рівняння можна вивести функцію щільності для розподілу віку при смерті [7]:

$$f(x) = \mu(x)l(x) = a \exp\left(bx - \frac{a}{b}(e^{bx} - 1)\right). \quad (2.4)$$

Проте, модель Гомперца не строго належить до сімейства LS, оскільки її сила смертності визначена лише для $x \geq 0$, тоді як моделі LS можуть

змінюватись на всій області дійсних чисел. Незважаючи на це, модель Гомперца є усіченим варіантом розподілу Гумбеля (типу 1 екстремальних значень). Це зв'язок між моделями підкреслює її важливість для опису тривалості життя.

2.3 Модельно-орієнтована кластеризація

Модельно-орієнтована кластеризація полягає у використанні статистичних моделей для групування даних. Цей підхід дозволяє не тільки кластеризувати дані, але й водночас зменшити кількість параметрів, що спрощує моделювання в умовах високої розмірності. Автори [1] запропонували використання обмежених та економних гаусових моделей для зменшення кількості вільних параметрів. Цей підхід передбачає, що кожен кластер відповідає певному ймовірнісному розподілу, що дозволяє здійснити точніше групування.

Одним з підходів, який використовує гаусові моделі, є обмежені гаусові моделі, що дозволяють суттєво зменшити кількість вільних параметрів. Ці моделі можуть бути застосовані для випадків, коли змінні є незалежними, проте це припущення не завжди відповідає реальності.

Інший підхід був запропонований [8], які розробили ймовірнісну байєсову модель, що використовує кінцеву кількість параметрів для опису кластерів, на основі яких потім відбувається кластеризація спостережень. Такий підхід може бути корисним у тих випадках, коли потрібно знайти кластерну структуру з меншим числом параметрів і при цьому уникнути зменшення розмірності даних перед обробкою, що є ключовою перевагою сучасних методів.

2.3.1 Метод К-середніх

Метод К-середніх є одним із найпоширеніших алгоритмів кластеризації, який застосовується для групування спостережень на основі їхньої близькості один до одного. Основна мета цього методу полягає в

мінімізації дисперсії всередині кластерів, що дозволяє виділити подібні за характеристиками групи.

Процес К-середніх складається з кількох етапів:

- на початковому етапі дослідник задає кількість кластерів, на які потрібно розділити дані. Це може бути викликом, оскільки точна кількість кластерів може бути невідома;
- алгоритм випадковим чином обирає k спостережень із набору даних, які будуть використовуватися як початкові центри кластерів;
- для кожного спостереження обчислюється відстань до кожного з центрів кластерів, і спостереження відноситься до найближчого кластера. Це зазвичай робиться за допомогою евклідової відстані;
- після призначення всіх спостережень до кластерів нові центри кластерів обчислюються як середнє значення всіх спостережень, що входять до кожного з кластерів;
- кроки 3 і 4 повторюються до тих пір, поки зміни в центрах кластерів не стануть незначними або досягнуто певної кількості ітерацій.

Застосування методу К-середніх у контексті прогнозування смертності є особливо важливим, оскільки він дозволяє групувати країни або регіони за подібними показниками смертності, що може бути корисним для подальшого аналізу та моделювання. Кластеризація населення за різними характеристиками, такими як вік, стать, соціально-економічний статус та інші фактори, може виявити схожості та відмінності у рівнях смертності.

Формально, для набору спостережень (x_1, x_2, \dots, x_n) (x_1, x_2, \dots, x_n , де кожне спостереження x_i є d -вимірним вектором, алгоритм К-середніх згрупує n спостережень у k кластерів $S = S_1, S_2, \dots, S_k$ шляхом мінімізації функції [9]:

$$\sum_{i=1}^k \sum_{x \in S_i} \|x - \mu_{S_i}\|^2, \quad (2.5)$$

2.3.2 Кластеризація з максимізацією математичного сподівання

Кластеризація за допомогою моделей гауссової суміші (GMM) є потужним методом, що використовується для аналізу даних, зокрема у випадках, коли кластеризація К-середніх виявляється недостатньо точною. Одним із головних недоліків К-середніх є те, що цей алгоритм використовує наївний підхід до визначення центру кластера, орієнтуючись на середнє значення. Це може бути проблематично, коли середні значення різних кластерів близькі одне до одного, що ускладнює їх розрізнення.

Модель гауссової суміші припускає, що дані підпорядковуються гауссовому розподілу. Замість того, щоб припускати, що всі точки зосереджуються навколо єдиного середнього значення, GMM враховує, що дані можуть мати різні середні значення та стандартні відхилення для кожного кластера. Для оцінки цих параметрів використовується алгоритм, відомий як максимізація математичного сподівання (EM).

Процес EM складається з двох основних етапів:

- на E-крок (Expectations) етапі обчислюється ймовірність належності кожної точки даних до кожного з кластерів на основі поточних параметрів моделі (середнє значення та стандартне відхилення);
- M-крок (Maximization) параметри моделі оновлюються, щоб максимізувати ймовірність даних, враховуючи нові ймовірності належності, розраховані на попередньому етапі.

Головною перевагою GMM є їхня гнучкість щодо коваріації кластерів. Цей метод може моделювати кластери з різними формами та розмірами, що робить його більш підходящим для багатьох реальних застосувань. Крім того, на відміну від К-середніх, GMM дозволяє точкам даних належати до кількох кластерів одночасно, що відкриває нові можливості для аналізу складних даних.

Однак GMM має свої обмеження. Припущення про те, що дані підпорядковуються гауссовому розподілу, може бути неприйнятним для певних типів даних. Крім того, алгоритм EM може бути вимогливим до

системних ресурсів і часу, оскільки для оцінки стандартного відхилення та середнього значення потрібно більше обчислювальних потужностей [9].

2.3.3 Ієрархічна кластеризація

Ієрархічна кластеризація є ще одним важливим методом, що дозволяє створювати ієрархічні структури або дерева кластерів. Цей метод можна умовно поділити на два типи: агломеративну (висхідну) та роздільну (зверху вниз).

Агломеративна кластеризація: починається з того, що кожна точка даних розглядається як окремий кластер. Потім алгоритм поступово об'єднує найближчі кластери, поки не залишиться один великий кластер, що містить усі дані.

Роздільна кластеризація: починає з одного великого кластера, який поступово розділяється на менші групи.

При використанні ієрархічної кластеризації користувачі можуть налаштувати параметри відстані (наприклад, евклідову або Манхеттенську відстань) для визначення близькості між точками даних. Цей підхід особливо корисний у випадках, коли кількість кластерів невідома заздалегідь, оскільки він надає гнучкість у процесі формування кластерів.

Однак ієрархічна кластеризація має свої недоліки, зокрема, чутливість до вибору параметрів відстані та складність інтерпретації результатів, особливо при роботі з великими наборами даних. Це може призвести до труднощів у візуалізації та розумінні структури кластерів. Тому важливо ретельно вибирати метод і параметри, виходячи з конкретних задач та характеристик даних [10].

2.4 Метод ARIMA

Метод ARIMA (авторегресійна інтегрована модель ковзної середньої) є потужним інструментом для аналізу часових рядів, що дозволяє моделювати

та прогнозувати дані, які змінюються з часом, зокрема показники смертності. ARIMA є особливо корисною у випадках, коли дані мають тренд або сезонність.

Модель ARIMA описується трьома параметрами: p , d , і q [11]:

- p : порядок авторегресії, що визначає кількість попередніх спостережень, які використовуються для прогнозування поточного значення;
- d : порядок інтеграції, що вказує на кількість разів, коли дані повинні бути диференційовані, щоб стати стаціонарними;
- q : порядок ковзної середньої, що вказує на кількість попередніх помилок прогнозування, які використовуються для оцінки поточного значення.

На початку важливо провести аналіз наявних даних смертності, включаючи виявлення трендів, сезонності та аномалій.

Для застосування моделі ARIMA дані повинні бути стаціонарними. Якщо дані не стаціонарні, їх можна зробити такими за допомогою диференціювання (тобто обчислення змін).

Параметри p , d і q можуть бути визначені за допомогою графіків автокореляції (ACF) та часткової автокореляції (PACF).

Після вибору параметрів модель ARIMA оцінюється, і важливо перевірити залишки моделі на предмет стаціонарності та нормальності.

Після оцінки моделі вона може бути використана для прогнозування майбутніх значень смертності. Важливо також врахувати можливість оновлення моделі з новими даними, щоб підвищити точність прогнозу.

Переваги [12]:

- ARIMA може ефективно моделювати складні залежності в даних;
- метод підходить для прогнозування як короткострокових, так і довгострокових тенденцій;
- ARIMA не вимагає припущення про нормальний розподіл даних.

Недоліки:

- модель вимагає значної кількості даних для надійної оцінки параметрів;
- вибір параметрів може бути складним і вимагати значного часу на аналіз;
- ARIMA може погано працювати з даними, що містять значні аномалії або зміни в тренді.

2.4 Прогнозування очікуваної тривалості життя

Очікувана тривалість життя є важливим демографічним показником, який відображає середню кількість років, яку людина може прожити за умови збереження сучасних рівнів смертності. Цей показник суттєво корелює з часом, а також демонструє значні відмінності між країнами, а також між чоловіками і жінками.

Прогнозування очікуваної тривалості життя базується на комбінованих прогнозах для чоловіків та жінок. Це дозволяє отримати більш цілісну картину змін тривалості життя та узгодити результати з найкращими світовими тенденціями. Однією з ключових тенденцій, яка використовується як орієнтир, є тренд, запропонований [13], що демонструє високу лінійну регулярність росту очікуваної тривалості життя жінок у країнах-рекордсменах.

У моделі передбачається, що очікувана тривалість життя жінок у певній країні слідує глобальній тенденції, яку можна описати через різницю між локальними і світовими показниками. Це відхилення прогнозується за допомогою моделей часових рядів. Аналогічно прогнозується і очікувана тривалість життя чоловіків з урахуванням специфічних для країни відмінностей між чоловіками і жінками.

Майбутній рівень очікуваної тривалості життя жінок у країні k для вікової групи x у рік t можна записати як [14]:

$$e_{k,x,t}^{\text{жін}} = e_{x,t}^{\text{max}} - D_{k,x,t} \quad (2.6)$$

де $e_{k,x,t}^{\text{жін}}$ – очікувана тривалість життя жінок у країні k для вікової групи x у рік t ; $e_{x,t}^{\text{max}}$ – глобальна найкраща тенденція очікуваної тривалості життя (орієнтир); $D_{k,x,t}$ – розрив між очікуваною тривалістю життя жінок у країні k та глобальним орієнтиром у рік t .

Цей розрив $D_{k,x,t}$ можна прогнозувати за допомогою класичної моделі часових рядів, наприклад ARIMA (AutoRegressive Integrated Moving Average), що дозволяє передбачати зміни у тривалості життя на основі минулих даних.

Для чоловіків прогнозування здійснюється із урахуванням гендерного розриву між очікуваною тривалістю життя жінок і чоловіків. Формула для чоловіків у країні k виглядає так[15]:

$$e_{k,x,t}^{\text{чол}} = e_{k,x,t}^{\text{жін}} - G_{k,x,t} \quad (2.7)$$

2.5 Методологія моделювання тенденцій смертності з використанням методів K-means кластеризації та ARIMA

Методологія застосування K-means кластеризації та ARIMA для моделювання тенденцій смертності складається з кількох етапів, що включають завантаження даних, їх структурування, кластеризацію за допомогою K-means, а також прогнозування за допомогою моделі ARIMA (рис.2.1).

На першому етапі ми імплементуємо завантаження історичних даних про смертність з доступних джерел. Дані можуть містити показники смертності за різні періоди, регіони, вікові групи тощо. Важливо, щоб дані були репрезентативними та охоплювали різноманітні аспекти смертності, аби забезпечити точність подальшого аналізу.

Після завантаження дані потребують попередньої обробки. Це включає очищення даних від пропусків, дублювань та аномалій, а також форматування даних для подальшого аналізу. На цьому етапі важливо визначити змінні, які впливають на смертність, і їх структуру. Цей процес

дозволяє підготувати дані до кластеризації, забезпечуючи їх коректність і узгодженість.

Використовуючи алгоритм K-means, ми переходимо до кластеризації даних, щоб виявити закономірності у показниках смертності. Процес кластеризації складається з кількох кроків:

1. Необхідно визначити кількість кластерів, які потрібно знайти в даних. Це може бути зроблено шляхом аналізу графіків, таких як метод "ліктя", що допомагає визначити оптимальне значення k .

2. На початковому етапі випадковим чином призначаємо початкові центроїди для кожного кластера. Кількість центроїдів відповідає вибраній кількості кластерів k .

3. Для кожного спостереження в даних обчислюємо відстань до кожного з центроїдів, використовуючи, наприклад, евклідову відстань.

4. Кожне спостереження призначається до найближчого центроїда, в результаті чого утворюються кластери. Це дозволяє групувати дані на основі їх подібності.

5. Після призначення спостережень до кластерів алгоритм перевіряє умову зупинки, тобто чи змінилося положення центроїдів. Якщо центроїди не змінилися або зміни стали незначними, алгоритм завершує свою роботу.

6. Після завершення кластеризації проводимо аналіз отриманих кластерів, щоб виявити спільні характеристики, які можуть пояснити закономірності у показниках смертності. Це дозволяє зрозуміти, які фактори впливають на формування кластерів, і може слугувати основою для подальших досліджень.

Після кластеризації даних можна перейти до прогнозування тенденцій смертності, використовуючи метод ARIMA. На основі кластеризованих даних проводимо аналіз часових рядів для кожного кластера, щоб виявити тренди та сезонність. Це дозволяє виявити регулярності в даних, які можуть бути використані для прогнозування.

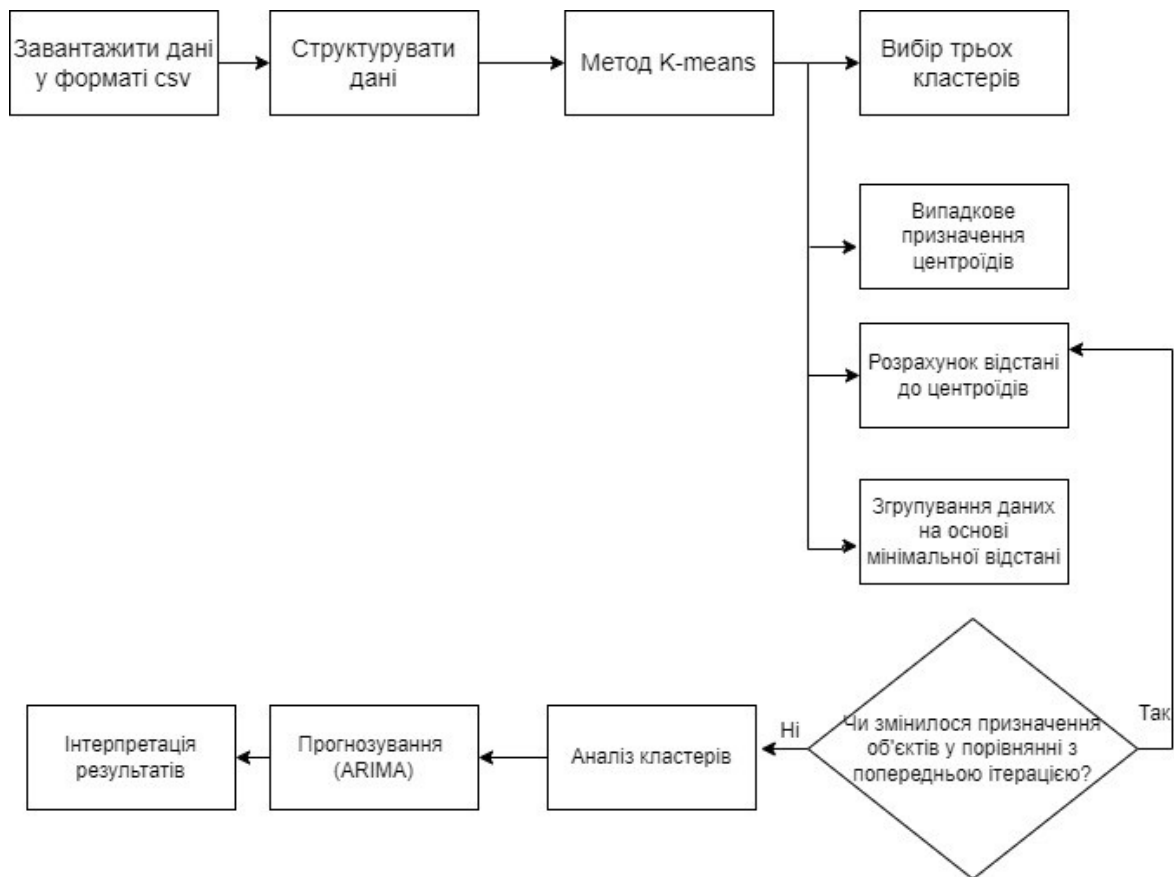


Рисунок 2.1 – Структурна схема методології прогнозування смертності

Визначаємо параметри p , d і q для моделі ARIMA на основі графіків ACF та PACF. Ці параметри впливають на структуру моделі і визначають, як вона реагуватиме на зміни в даних.

Оцінюємо модель ARIMA, перевіряючи залишки на стаціонарність та нормальність. Це критично важливий етап, оскільки ненормальні залишки можуть свідчити про те, що модель неадекватно описує дані.

Використовуємо оцінену модель для прогнозування майбутніх показників смертності для кожного кластера. Це дозволяє отримати цінну інформацію про ймовірні зміни в тенденціях смертності.

Завершальний етап включає інтерпретацію отриманих результатів, порівняння прогнозів з реальними даними та виявлення можливих причин змін у тенденціях смертності. Це може допомогти у формуванні рекомендацій для політики в сфері охорони здоров'я, орієнтуючи зусилля на покращення здоров'я населення та зменшення показників смертності.

На цьому етапі ми визначаємо параметри моделі ARIMA: p (порядок авторегресії), d (число диференціювань) та q (порядок ковзного середнього). Для цього ми використовуємо графіки ACF (автокореляційна функція) і PACF (часткова автокореляційна функція). Аналізуючи ці графіки, ми можемо вибрати найбільш підходящі значення для кожного з параметрів.

Для аналізу даних та моделювання тенденцій смертності з використанням моделі ARIMA важливо правильно вибрати, які змінні слугуватимуть в якості основи для створення часового ряду. У нашому випадку можна обрати такі параметри:

- країна – це дозволить вам провести порівняння між різними регіонами або державами. Ви можете обрати дані, наприклад, для Республіки Сербія або інших країн Європи;

- стать – додавання статі дозволяє досліджувати відмінності в показниках смертності між чоловіками та жінками. Це може виявитися важливим, оскільки різні статеві групи можуть мати різні тенденції;

- вік є критично важливим фактором, оскільки тенденції смертності часто сильно залежать від вікових груп. Наприклад, ви можете досліджувати показники смертності для вікових категорій, таких як 60-64 роки або 65-74 роки;

- причина смерті (захворювання) – це ще один ключовий аспект, який дозволяє детально аналізувати, як різні захворювання впливають на смертність. Ви можете вибрати певні захворювання, такі як нижні дихальні інфекції або ВІЛ/СНІД, щоб вивчити тенденції смертності, пов'язані з ними.

Після визначення параметрів, ми оцінюємо модель ARIMA. На цьому етапі важливо перевірити залишки моделі на стаціонарність та нормальність. Це дозволяє впевнитись, що модель адекватно описує дані, і прогнози будуть достовірними. Якщо результати перевірок не відповідають вимогам, можливо, знадобиться скоригувати параметри або внести додаткові зміни в модель.

Коли модель ARIMA оцінена, ми можемо перейти до прогнозування. На основі історичних даних про смертність ми можемо здійснити прогнози на кілька років вперед. Важливо відзначити, що ці прогнози будуть стосуватися певних вікових груп та статі, що дозволяє виявити можливі зміни у тенденціях смертності.

Завершальним етапом є візуалізація результатів прогнозування. Порівнюючи прогнози з реальними даними, ми можемо оцінити точність моделі та виявити можливі причини змін у тенденціях смертності. Візуалізація також допомагає формулювати рекомендації для політики в сфері охорони здоров'я, враховуючи прогнози щодо смертності від нижніх і верхніх дихальних шляхів.

2.6 Висновки до розділу

У другому розділі розглянули різноманітні моделі смертності та методи кластеризації, які можуть бути використані для аналізу та прогнозування тенденцій смертності в населення.

Проаналізували різні моделі, які відображають вплив місця проживання на показники смертності. Цей аналіз вказує на важливість географічних факторів у дослідженні здоров'я населення. Описано модель Гомперца, яка є однією з основних моделей для опису вікової залежності смертності. Ми виявили, що ця модель забезпечує корисний інструмент для прогнозування смертності, оскільки враховує як вікові, так і гендерні відмінності.

Розглянули моделі кластеризації, зокрема метод К-середніх, кластеризацію з максимізацією математичного сподівання та ієрархічну кластеризацію. Зокрема, метод К-середніх продемонстрував свою ефективність у виявленні закономірностей у даних про смертність, що дозволяє розподілити населення на гомогенні групи.

Описано методологію застосування K-means кластеризації та ARIMA для моделювання тенденцій смертності. Цей підхід об'єднує переваги кластеризації та часових моделей, що дозволяє отримати більш точні прогнози та глибше розуміння факторів, які впливають на смертність.

3 ПРОГРАМНА РЕАЛІЗАЦІЯ МОДЕЛЮВАННЯ ТЕНДЕНЦІЙ СМЕРТНОСТІ З ВИКОРИСТАННЯМ МЕТОДІВ K-MEANS КЛАСТЕРИЗАЦІЇ ТА ARIMA

3.1 Архітектура програмного забезпечення

Архітектура програмного забезпечення (ПЗ) для кластеризації текстових даних побудована на чотирьох основних компонентах, кожен з яких відіграє ключову роль у забезпеченні повного циклу обробки даних – від введення тексту до отримання результатів кластеризації, їх збереження і подальшого аналізу. Взаємодія між цими компонентами є важливою для забезпечення точності, зручності використання та надійності системи. На рис.3.1 наведено структурну схему архітектури ПЗ.

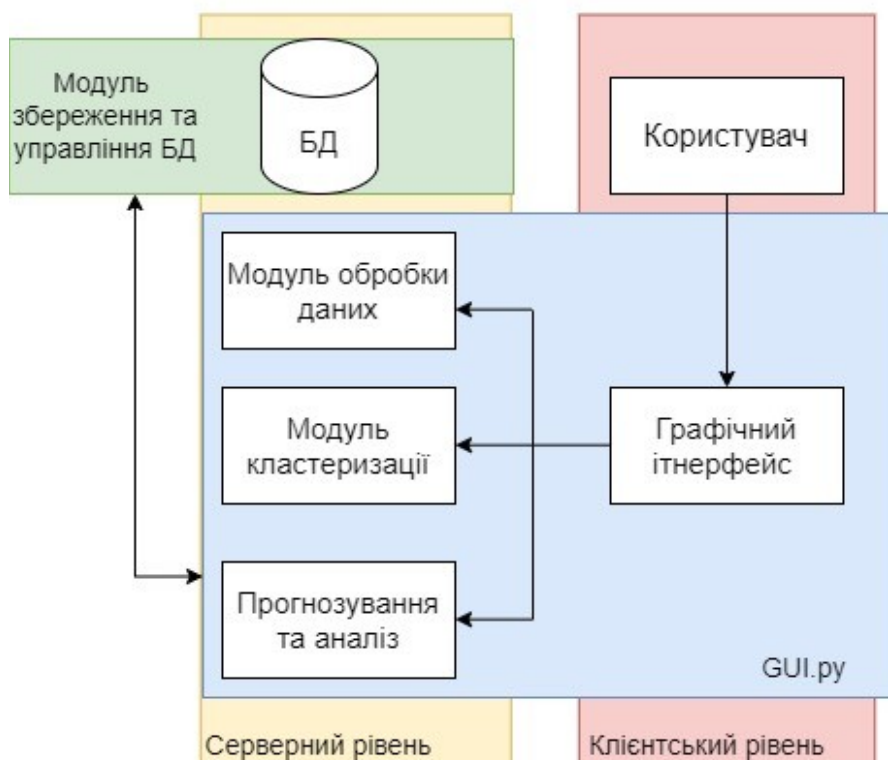


Рисунок 3.1 – Структурна схема архітектури програмного забезпечення

В архітектурі програмного забезпечення (рис.3.1) для кластеризації статистичних даних front-end і back-end компоненти мають чіткий поділ, де кожен відповідає за окремі завдання в системі.

Front-end (Клієнтський рівень) – це все, що користувач безпосередньо бачить та з чим взаємодіє. У даній архітектурі це:

- графічний інтерфейс користувача (GUI), де користувач може вводити текст, завантажувати файли або обирати інші джерела даних;
- інтерфейс, де користувач може вводити дані про смертність, завантажувати файли або обирати інші джерела даних, такі як статистика за різними країнами та віковими категоріями;
- елементи керування для вибору методів аналізу та кластеризації (наприклад, KMeans або ARIMA) для моделювання тенденцій смертності;
- кнопки для запуску кластеризації, прогнозування та перегляду результатів;
- відображення результатів кластеризації у вигляді груп країн за рівнем смертності, а також статистичних даних про кількість смертей від інфекційних захворювань у різних вікових групах;
- виведення статистичних даних і метаданих щодо проведеного аналізу (кількість кластерів, дані про смертність, методи кластеризації та прогнозування).

Основні технології для Front-end: можуть бути використані мови та фреймворки для створення GUI, такі як Python (Tkinter, PyQt) або JavaScript (якщо інтерфейс є веб-застосунком).

Back-end (Серверний рівень) відповідає за обробку всіх даних та логіку системи. Його компоненти включають:

- модуль обробки даних виконує попередню обробку даних про смертність: очищення, нормалізація, форматування та структурування

інформації для подальшого аналізу. Готує дані для кластеризації, забезпечуючи їх коректність та придатність для аналізу;

- модуль кластеризації: відповідає за обчислення та виконання алгоритмів кластеризації (KMeans) для виявлення груп країн з подібними показниками смертності від інфекційних захворювань. Використовує дані про смертність для вибору відповідного методу аналізу та прогнозування на основі параметрів, переданих з front-end. Повертає результати кластеризації для подальшого відображення, включаючи інформацію про рівень смертності в кожному кластері.

- модуль збереження та управління базою даних: зберігає результати кластеризації, використані методи та метадані в базі даних для подальшого аналізу. Забезпечує можливість доступу до історичних даних про смертність та результати кластеризації, що дозволяє проводити порівняльний аналіз змін у часі. Інтегрує дані з базою даних для зберігання результатів, що забезпечує зручний доступ до інформації про смертність від інфекційних захворювань у Європі.

Основні технології для Back-end: Python (для обробки даних та алгоритмів кластеризації), SQLite або інші СУБД для збереження даних, API або серверна логіка для взаємодії між front-end і back-end.

Взаємодія між Front-end і Back-end є важливою для забезпечення безперебійного функціонування системи аналізу та прогнозування смертності від інфекційних захворювань в Європі. Ця взаємодія забезпечує обмін даними, що дозволяє користувачу отримувати актуальні результати аналізу та моделювання.

Front-end формує запити на основі дій користувача (наприклад, запуск кластеризації або прогнозування) і надсилає їх до Back-end через API (інтерфейс програмування додатків).

Back-end обробляє ці запити, виконує необхідні обчислення та повертає результати у вигляді відповідей, які Front-end відображає користувачу.

Користувач вводить дані (наприклад, статистику смертності) у Front-end, який формує відповідний запит і передає дані до Back-end для обробки. Back-end приймає ці дані, виконує їх обробку (попередня обробка, кластеризація, прогнозування) і зберігає результати у базі даних.

Після виконання обчислень Back-end формує результати аналізу та кластеризації, які можуть включати в себе групи країн за рівнем смертності, статистичні дані, графіки тощо. Ці результати надсилаються назад до Front-end, який відображає їх у зручному для користувача форматі, наприклад, у вигляді таблиць, графіків або карт.

Якщо користувач вносить зміни (наприклад, коригує параметри аналізу), Front-end оновлює відповідні дані, які знову передаються до Back-end для повторного аналізу. Цей процес синхронізації даних забезпечує актуальність інформації, що відображається, та дозволяє користувачу отримувати результати на основі найсвіжіших даних.

У випадку виникнення помилок (наприклад, некоректних вхідних даних) Back-end надсилає повідомлення про помилки Front-end, який відображає їх користувачу, надаючи можливість виправити помилки та повторити запит.

Для підвищення продуктивності системи Front-end може використовувати кешування раніше отриманих результатів, що зменшує навантаження на Back-end та пришвидшує доступ до інформації.

Розглянемо модулі ПЗ. Графічний інтерфейс користувача (GUI) є важливим компонентом, що забезпечує взаємодію користувача з системою. Через нього користувач може завантажувати дані про смертність від інфекційних захворювань по країнах Європи, вказуючи параметри, такі як рік, вікова категорія та стать. Інтерфейс дозволяє вибирати методи аналізу та

кластеризації даних (наприклад, KMeans або ARIMA), запускати процес обробки та переглядати результати у зрозумілому форматі. Важливою частиною GUI є кнопки для запуску кластеризації, прогнозування та кнопка "статистика", яка дозволяє переглядати збережені результати аналізу, включаючи кількість застосованих методів і параметри прогнозування. GUI також інтегрований з іншими модулями програми для безперебійного передавання та отримання даних.

Модуль обробки даних відповідає за підготовку статистичних даних про смертність перед передачею їх у модуль кластеризації або прогнозування. Оскільки дані можуть бути різномірними за структурою та містити пропуски або аномальні значення, цей модуль виконує попередню обробку, зокрема очищення та нормалізацію. Основними функціями є видалення зайвих даних, обробка пропусків, нормалізація та перетворення даних у формат, придатний для аналізу. Наприклад, нормалізуються дані про кількість смертей у кожній країні, що дозволяє краще використовувати методи кластеризації для групування країн за рівнем смертності.

Модуль кластеризації є головним елементом програми, який виконує кластеризацію даних про смертність у Європі. Він реалізує алгоритми KMeans для поділу країн на кластери залежно від рівня смертності, що дозволяє виявити групи країн з подібними показниками смертності від інфекційних захворювань. Наприклад, країни можуть бути розподілені на групи з високим, середнім та низьким рівнем смертності, що дозволяє проводити більш детальний аналіз ситуації. Для подальшого прогнозування майбутніх тенденцій модуль також інтегрується з алгоритмами прогнозування, такими як ARIMA, що дозволяє передбачати зміни рівня смертності на основі часових рядів.

Модуль збереження та управління базою даних забезпечує надійне зберігання результатів кластеризації, прогнозування та інших даних, що

можуть бути корисними для подальшого аналізу. У базі даних зберігаються статистичні дані про смертність, ідентифікатори кластерів, параметри, за якими проводився аналіз, а також інформація про використані методи. Це дає змогу користувачам переглядати історію проведених операцій та використовувати результати для порівняння та аналізу ефективності різних методів. База даних також зберігає інформацію про кількість використання кожного методу кластеризації та прогнозування, що дозволяє відстежувати популярність чи точність окремих алгоритмів.

Модуль прогнозування та аналізу використовується для прогнозування майбутніх рівнів смертності від інфекційних захворювань у Європі. Він застосовує часові ряди та алгоритми прогнозування, такі як ARIMA, для моделювання тенденцій смертності в різних країнах. Це дозволяє передбачати можливі зміни у рівні смертності та виявляти країни, де проблема може загостритися. Модуль також може виконувати оцінку якості прогнозів, використовуючи відповідні метрики.

3.2 Модуль обробки даних

Модуль обробки даних та візуалізації виконує аналіз даних про смертність від інфекційних захворювань. Основні етапи роботи модуля включають завантаження даних, їх очищення, обробку, візуалізацію та створення веб-додатку для інтерактивного перегляду даних.

Модуль починає з імпорту необхідних бібліотек, таких як `pandas`, `geopandas`, `folium`, `flask` та інших. Потім дані про смертність завантажуються з CSV-файлу `mortality_data11.csv` [16]. Перед подальшою обробкою модуль виключає дані для Росії, що забезпечує коректність аналізу.

Очищення даних є важливим етапом. Значення стовпця `'val'` конвертуються в числовий формат, а нечислові символи видаляються. Це дозволяє зберегти лише числові значення, які будуть використані для

подальшого аналізу. Значення, які не можуть бути конвертовані, видаляються з набору даних.

Після очищення даних визначаються початкові параметри для фільтрації даних: рік, стать, причина смертності та вік. Ці параметри використовуються для створення початкового вигляду графічного інтерфейсу.

Модуль містить функцію `create_map`, яка відповідає за створення інтерактивної карти на основі вибраних параметрів (рис.3.2). Функція фільтрує дані за обраними характеристиками і використовує GeoJSON-файл для відображення географічних даних (рис.3.3) [17].

```
def create_map(selected_year, selected_sex, selected_cause, selected_age):
    year_data = data[
        (data['year'] == selected_year) &
        (data['sex_name'] == selected_sex) &
        (data['cause_name'] == selected_cause) &
        (data['age_name'] == selected_age)
    ]

    if year_data.empty:
        return None
```

Рисунок 3.2 – Фрагмент коду для реалізації функції `create_map`

```
geojson_path = os.path.join(os.path.dirname(__file__), 'countries.geo.json')
world = gpd.read_file(geojson_path)
map_data = world.merge(year_data, left_on='name', right_on='location_name', how='left')
```

Рисунок 3.3 – Фрагмент коду для фільтрування даних GeoJSON-файлу

На карті використовується хлоропетна візуалізація, яка показує рівень смертності за допомогою кольорів. Для кожної країни, що має дані про смертність, на карті розміщуються круглі маркери, які містять інформацію про значення смертності, причину та інші деталі. Значення форматується для зручності сприйняття, наприклад, у мільйонах або тисячах.

Модуль також реалізує веб-додаток на Flask, що дозволяє користувачам взаємодіяти з даними через браузер. Головна сторінка відображає карту та

меню для вибору параметрів. При зміні параметрів дані автоматично оновлюються, і на карті відображаються відповідні дані (рис.3.4).

```
@app.route('/')
def index():
    selected_year = request.args.get('year', initial_year)
    selected_sex = request.args.get('sex', initial_sex)
    selected_cause = request.args.get('cause', initial_cause)
    selected_age = request.args.get('age', initial_age)

    map_object = create_map(selected_year, selected_sex, selected_cause, selected_age)

    if map_object is None:
        return "Немає даних для вибраних параметрів.", 404

    map_html = map_object._repr_html_() # Генерація HTML для мапи
    return render_template('index1.html', map=map_html)
```

Рисунок 3.4 – Фрагмент коду для реалізації веб-додатку на Flask

Функція `index` обробляє HTTP-запити, отримуючи вибрані параметри від користувача, а потім викликає функцію `create_map` для створення нової карти. Генерування HTML для карти здійснюється за допомогою методу `_repr_html_()`.

Для зручності користувачів модуль також реалізує функцію `open_browser`, яка автоматично відкриває браузер з адресою веб-додатка після його запуску [18, 19]. Це робиться за допомогою потокового виконання, що дозволяє не блокувати основний потік програми.

3.3 Модуль кластеризації

Модуль кластеризації смертності використовує дані про захворювання верхніх і нижніх дихальних шляхів для аналізу смертності в різних країнах. Спочатку дані завантажуються з CSV-файлу за допомогою бібліотеки `Pandas`. Виконується перегляд унікальних значень у стовпці `"val"`, що дозволяє оцінити доступні метрики. Далі проводиться фільтрація даних для захворювань, зокрема `"Upper respiratory infections"` та `"Lower respiratory infections"`. Після цього дані перетворюються в зручний формат для

кластеризації, використовуючи зведену таблицю, яка агрегує значення смертності за роками та країнами.

Наступним етапом є перевірка наявності NaN значень у зведених даних. Якщо такі значення виявляються, їх заповнюють нулями, що дозволяє уникнути проблем під час виконання кластеризації. Для кластеризації застосовується алгоритм KMeans, де задається кількість кластерів (`n_clusters`), яка в даному випадку дорівнює трьом. Алгоритм навчається на підготовлених даних, і результати кластеризації додаються до зведених даних у вигляді нового стовпця "Cluster" (рис.3.5).

```
#Перевірка даних на NaN
n_clusters = 3
kmeans = KMeans(n_clusters=n_clusters, random_state=42)
kmeans.fit(pivot_data)
pivot_data['Cluster'] = kmeans.labels_

#Присвоєння назв кластерам
cluster_labels = {0: "Низька смертність", 1: "Середня смертність", 2: "Висока смертність"}
pivot_data['Cluster Name'] = pivot_data['Cluster'].map(cluster_labels)
```

Рисунок 3.5 – Фрагмент коду для реалізації алгоритму KMeans

Щоб легше інтерпретувати результати, до кластерів додаються назви на основі середнього значення смертності: "Низька смертність", "Середня смертність" та "Висока смертність". Далі визначаються країни, що потрапили в кожен з кластерів, і ці дані виводяться на екран. Наприклад, країни з низькою смертністю фільтруються з отриманих кластерів і візуалізуються за допомогою стовпчастої діаграми.

Візуалізація даних реалізується за допомогою бібліотеки Matplotlib [20]. Для кожного з кластерів створюються графіки, що показують середню смертність за роками. У результаті формується графік з точками, що представляють країни, та легендою, яка відображає назви кластерів. В кінці процесу виконується візуалізація середніх значень смертності по кожному

кластеру, що дозволяє наочно продемонструвати різницю між групами країн за їхніми показниками смертності.

Модуль також перевіряє наявність дублікатів у даних та виводить інформацію про них. Це важливо для забезпечення чистоти даних перед проведенням кластеризації. Загалом, алгоритм кластеризації смертності дає змогу здійснити глибокий аналіз даних про смертність від респіраторних захворювань та візуалізувати результати у зрозумілому вигляді.

3.4 Модуль прогнозування смертності

Цей модуль виконує прогнозування смертності за допомогою методу ARIMA, застосовуючи кластеризацію для аналізу даних по країнах. Спочатку імпортуються необхідні бібліотеки, такі як `pandas` для роботи з даними, `matplotlib` для візуалізації, `KMeans` з `sklearn.cluster` для кластеризації та `ARIMA` з `statsmodels` для моделювання часових рядів. Далі зчитується файл з даними про смертність, що містить значення за різні роки для різних країн.

Після зчитування даних, вони підготовлюються до аналізу. Створюється зведена таблиця, де країни є індексами, а роки — стовпцями. Всі пропуски в даних заповнюються нулями. Наступним кроком виконується кластеризація даних на три групи за допомогою алгоритму `KMeans`. Кожна країна отримує класифікацію, яка відображає рівень смертності: низька, середня або висока. Результати кластеризації виводяться на консоль, де для кожного кластеру перераховуються країни, що до нього належать.

Після цього для кожної країни розробляється модель ARIMA (рис.3.6). Якщо для країни є достатня кількість даних, модель навчається на основі існуючих значень смертності, після чого виконується прогнозування на наступні п'ять років. Прогнози зберігаються у новому `DataFrame`, що потім виводиться на екран. Крім того, результати прогнозування записуються в текстовий файл для подальшого аналізу.

```

data = pd.read_csv('mortality_data.csv', parse_dates=['Date'], index_col='Date')

# Перевірка наявності пропущених значень
data = data.fillna(0)

# Вибір даних для однієї країни
country_data = data['Country_Name'] # Замініть на потрібну країну

# Розділення на навчальну та тестову вибірки
train_size = int(len(country_data) * 0.8)
train, test = country_data[0:train_size], country_data[train_size:]

# Параметри моделі ARIMA
p = 1 # порядок авторегресії
d = 1 # порядок диференціювання
q = 1 # порядок ковзаючого середнього

# Побудова моделі ARIMA
model = ARIMA(train, order=(p, d, q))
model_fit = model.fit()

# Прогнозування
forecast = model_fit.forecast(steps=len(test))
forecast_index = test.index

# Оцінка моделі
mse = mean_squared_error(test, forecast)
print(f'Mean Squared Error: {mse}')

```

Рисунок 3.6 – Фрагмент коду для моделі ARIMA

Візуалізація результатів є важливим етапом. Спочатку побудовується графік, на якому відображаються реальні дані смертності для кожної країни, а також прогнози на найближчі роки. Додатково, прогнози можна представити у вигляді стовпчикової діаграми, що дає можливість порівняти смертність між різними країнами за прогнозовані роки. Всі графіки та таблиці налаштовуються для оптимізації простору і покращення читаємості.

Завершується модуль обробкою прогнозів, які зберігаються у текстових файлах для подальшої роботи. Додаткові перевірки дозволяють уникнути помилок при зчитуванні даних, гарантують коректність та надійність виконання всіх етапів аналізу.

3.5 Інтерфейс користувача

Графічний інтерфейс користувача (GUI) є важливою частиною програми "Комплексне використання методів K-means та ARIMA", що дозволяє користувачеві зручно взаємодіяти з системою аналізу смертності (рис.3.7). Інтерфейс реалізований з використанням бібліотеки Tkinter, що забезпечує простий і зрозумілий доступ до функцій програми. Вікно

програми має м'який світлий фон, що сприяє комфортному сприйняттю інформації. Основні елементи інтерфейсу включають заголовок, кнопки для завантаження даних, візуалізації, кластеризації та прогнозування, а також повідомлення для користувача.

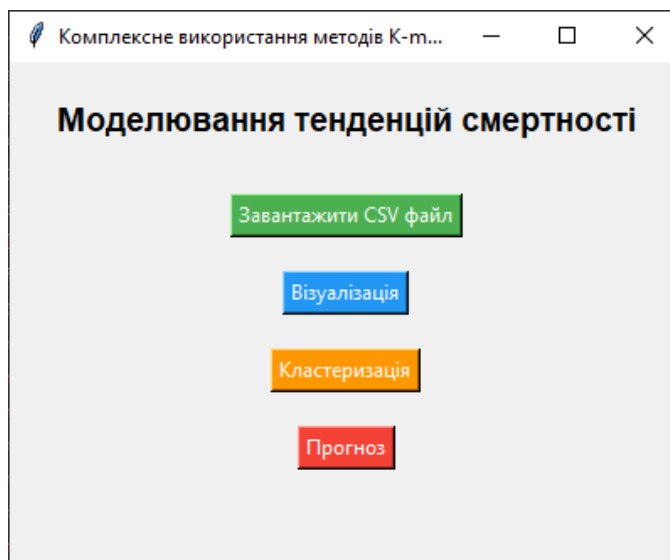


Рисунок 3.7 – Графічний інтерфейс ПЗ «Комплексне використання методів K-means та ARIMA»

Кнопка для завантаження файлу дозволяє користувачу вибрати файл у форматі CSV, що містить дані про смертність. При натисканні на кнопку відкривається діалогове вікно, де користувач може знайти та обрати потрібний файл (рис.3.8).

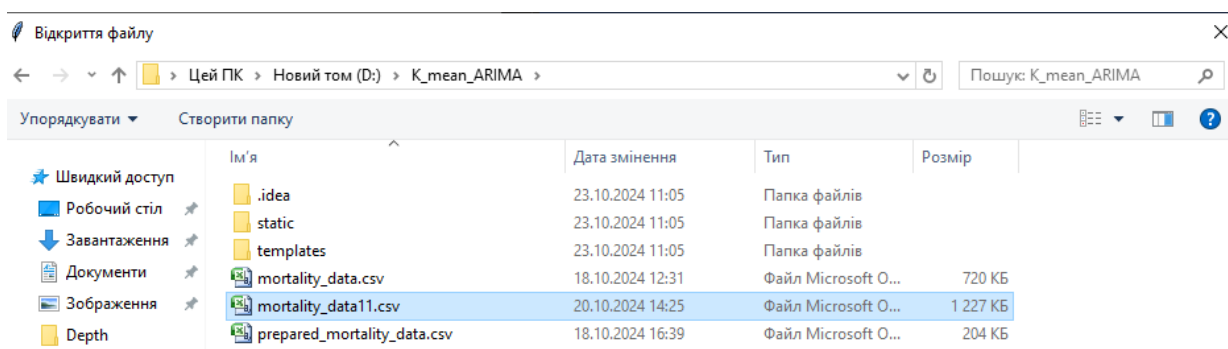


Рисунок 3.8 – Результат вибору файлу для обробки ПЗ

Після успішного завантаження програма виводить повідомлення, що підтверджує, що файл був завантажено (рис.3.9).

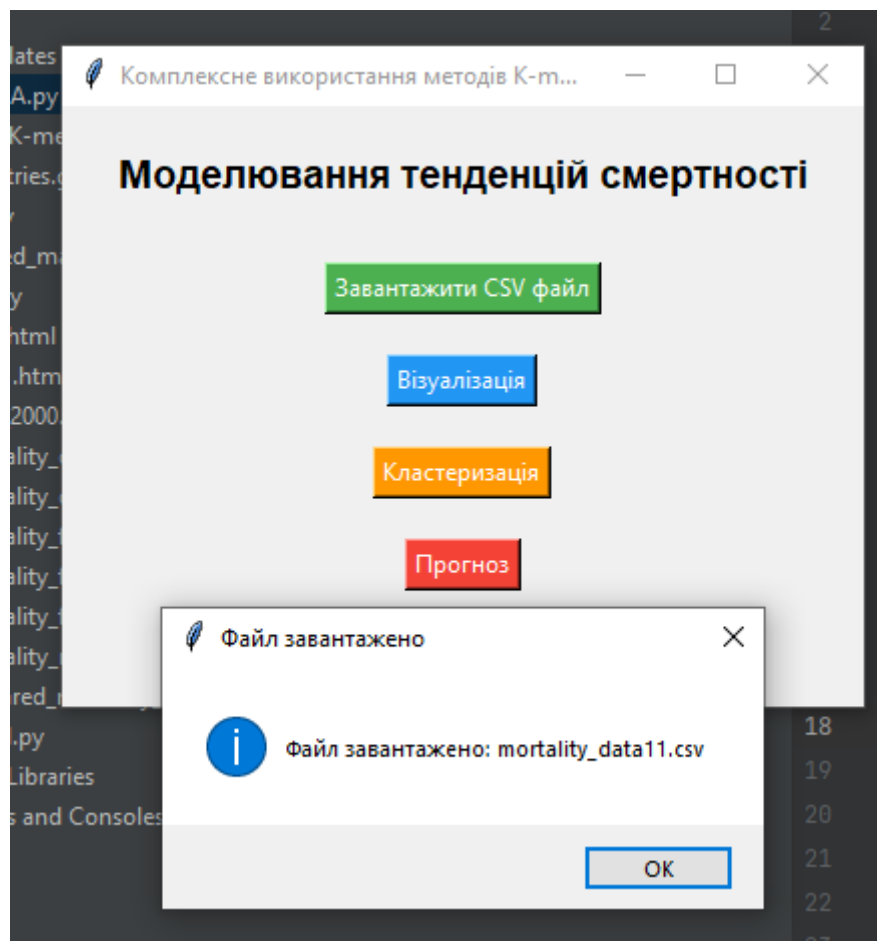


Рисунок 3.9 – Результат завантаження файлу для обробки ПЗ

Кнопка для візуалізації запускає процес візуалізації даних. Вона викликає відповідний скрипт, який відповідає за графічне відображення результатів аналізу. Завдяки багатопоточності, візуалізація виконується в окремому потоці, що дозволяє уникнути зависання інтерфейсу під час виконання тривалих обчислень (рис.3.10).

На рис.3.10 візуалізовано рівні смертності в Європі за рік (2010), стать (чоловіки або жінки), вікову категорію (60-64 роки або вибрати з випадуючого списку) та причину смерті (інфекції нижніх або верхніх дихальних шляхів). Карта забарвлена в різні кольори для окремих країн, що

відповідають кількості смертей, причиною яких стали інфекції нижніх дихальних шляхів у чоловіків у віковій категорії 60-64 років.

Основні елементи карти:

- колірна шкала у верхньому правому куті показує кількість смертей (від 13 до 50). Темно-червоний колір вказує на найбільший рівень смертності (до 50 смертей), тоді як світліші кольори означають менші рівні.
- країни на карті. Україна має високий рівень смертності, забарвлена в помаранчевий колір. Угорщина показана червоним кольором, що означає дуже високий рівень смертності. Інші країни, такі як Австрія, Чехія, Сербія, мають різні відтінки жовтого та червоного, залежно від рівня смертності.
- сині точки це маркери, які відображають інформацію смертності.

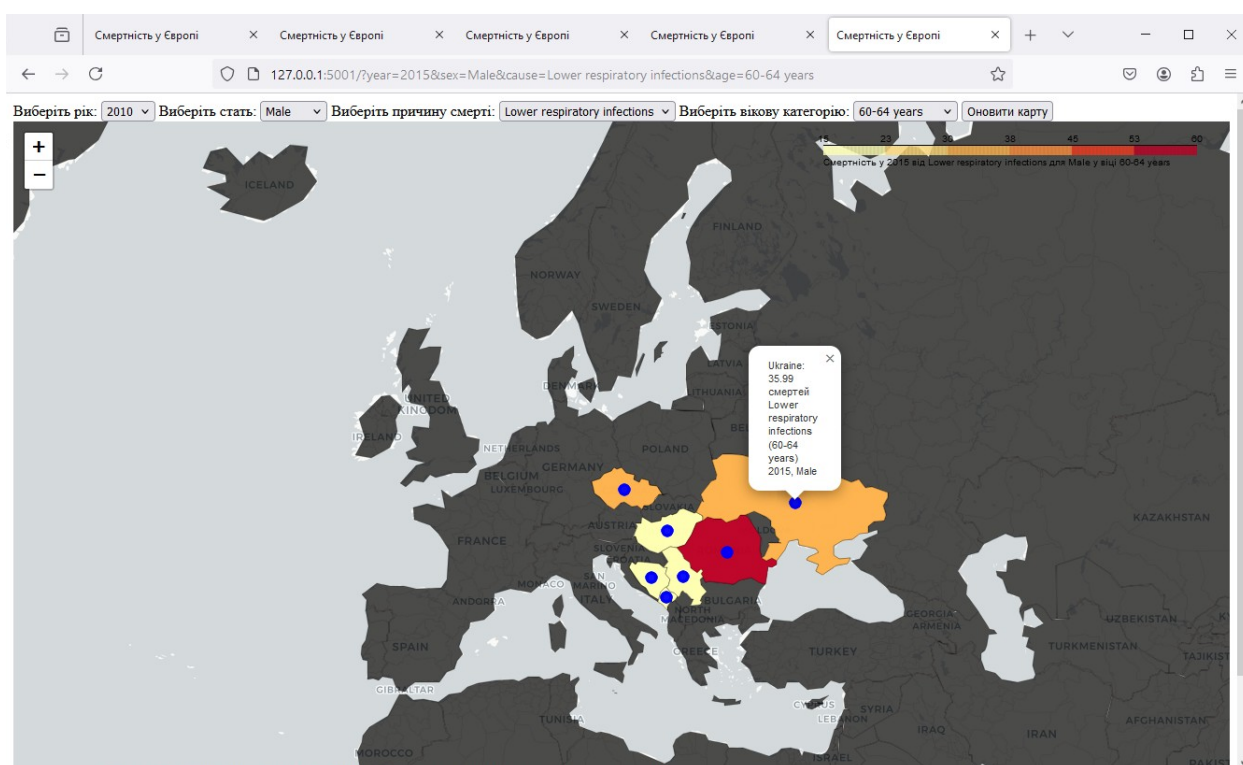


Рисунок 3.10 – Результат візуалізації даних на карті

Кнопка для кластеризації ініціює процес кластеризації даних за допомогою алгоритму K-means (рис.3.11). Як і у випадку з візуалізацією,

кластеризація запускається в окремому потоці, що дозволяє користувачеві продовжувати взаємодію з інтерфейсом під час виконання обчислень.

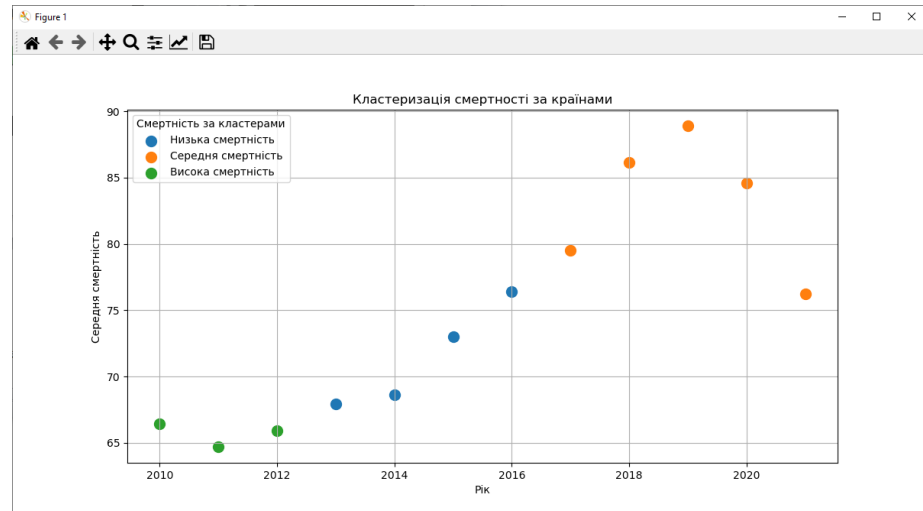


Рисунок 3.11 – Результат кластеризації за допомогою алгоритму K-means

Кнопка для прогнозування активує процес прогнозування на основі алгоритму ARIMA (рис.3.12). Так само, як і в попередніх випадках, прогнозування виконується в окремому потоці, що забезпечує комфортність використання програми. Результати зберігає в текстовий файл як показано на рис.3.13.

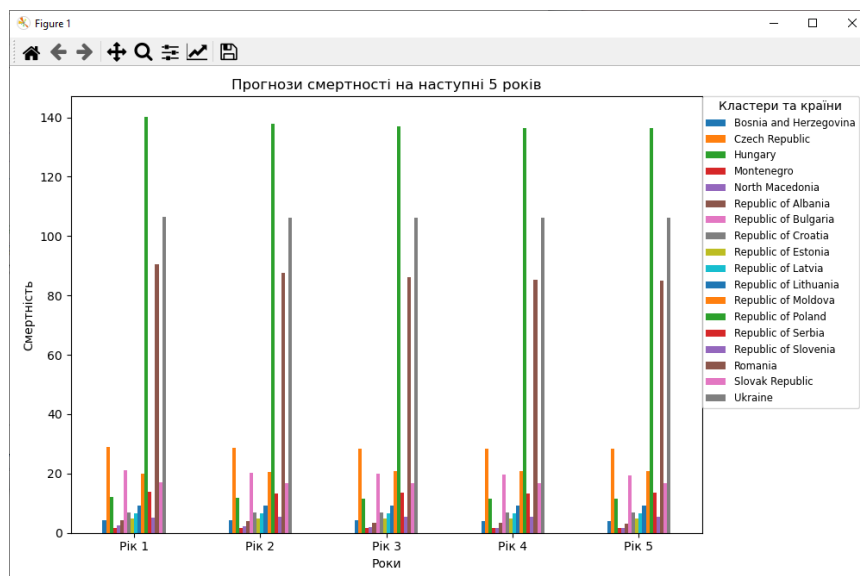


Рисунок 3.12 – Результат прогнозу за допомогою алгоритму ARIMA

	Bosnia and Herzegovina	Czech Republic	Hungary	Montenegro	North Macedonia	Republic of Albania	Republic of Bulgaria	Republic of Serbia
Pik 1	4.337538	29.025776	12.031949	1.731911	2.400728	4.231365	21.025227	19.883771
Pik 2	4.173440	28.590629	11.738082	1.752513	2.159356	3.844519	20.317820	19.883771
Pik 3	4.130738	28.479697	11.596361	1.751409	1.939015	3.548235	19.883771	19.617448
Pik 4	4.119626	28.451418	11.528014	1.751468	1.737874	3.321311	19.617448	19.454038
Pik 5	4.116734	28.444208	11.495053	1.751465	1.554259	3.147510	19.454038	

Рисунок 3.13 – Результат прогнозу за допомогою алгоритму ARIMA, збережені в текстовому файлі (фрагмент файлу)

Користувач отримує інформацію про статус виконання дій через повідомлення, що виводяться за допомогою модулю messagebox. Наприклад, після завантаження файлу програма виводить сповіщення про успішне виконання дії, що підвищує зручність використання інтерфейсу. Ці повідомлення допомагають користувачу бути в курсі подій, що відбуваються в програмі, та своєчасно реагувати на результати своїх дій.

3.6 Оцінка результатів комплексного використання методів K-means та ARIMA

В пункті 3.5 описано програмну реалізацію кластеризації за допомогою алгоритму K-means. Виконавши налаштування для статистичних даних країн Європи отримали результати на рис.3.14.

На графіку рис.3.14 представлено середні значення смертності за роками (2010-2020) для трьох кластерів європейських країн, згрупованих за рівнем смертності. Вертикальна вісь відображає рівень смертності, тоді як горизонтальна — роки. Кожен кластер представлений окремим кольором, що дозволяє чітко простежити зміни рівня смертності в кожній групі країн.

Кластер низької смертності (зелений) містить країни, такі як Чехія, Болгарія, Молдова та Словаччина. Рівень смертності у цих країнах залишається стабільно низьким, коливаючись у межах 20-30 смертей на рік. Протягом досліджуваного періоду з 2010 по 2020 роки, значні коливання не

спостерігаються, що свідчить про стабільну демографічну ситуацію у цьому кластері.

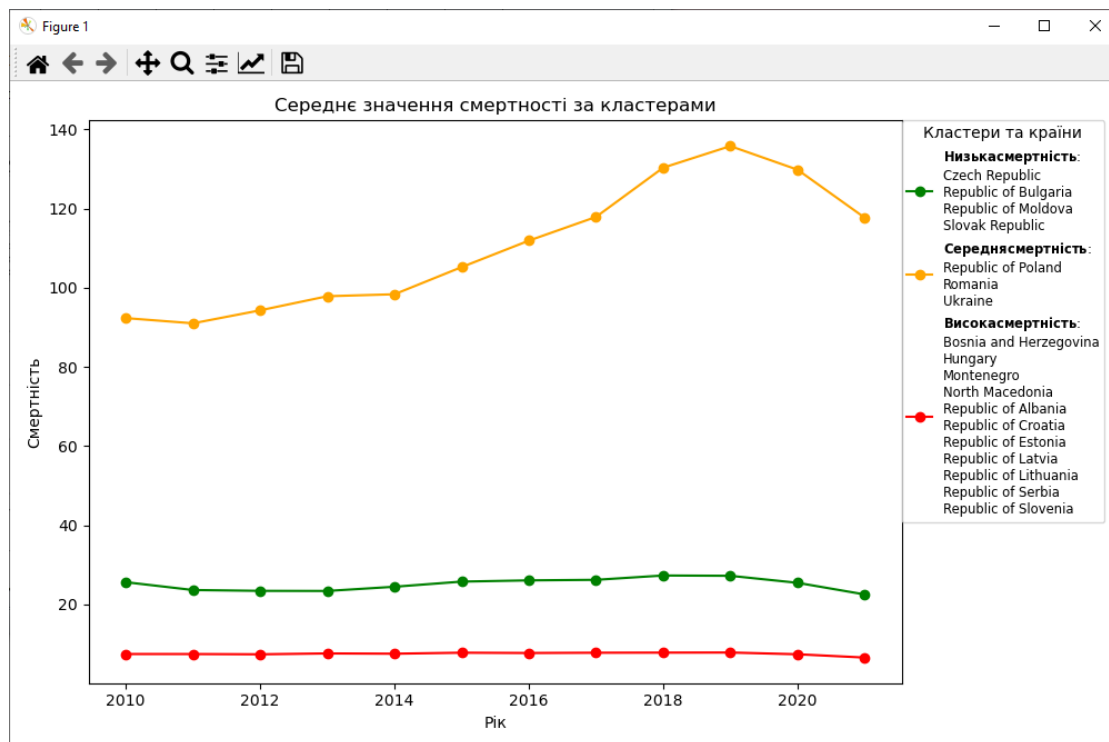


Рисунок 3.14 – Результат кластеризації за допомогою алгоритму K-means смертності країн Європи

Кластер середньої смертності (жовтий) – Польщу, Румунію та Україну. На відміну від кластеру низької смертності, тут спостерігається чітка тенденція до зростання рівня смертності. Починаючи з близько 100 смертей на рік у 2010 році, цей показник поступово збільшується, досягаючи піку приблизно у 130 смертей у 2019-2020 роках. Однак, у 2020 році відзначається незначне зниження, що може вказувати на стабілізацію після тривалого періоду зростання.

Кластер високої смертності (червоний) – Албанія, Хорватія, країни Балтії, Сербія, Словенія та інші. Хоча цей кластер названо кластером високої смертності, рівень смертності тут є відносно низьким порівняно з іншими

кластерами, тримаючись у межах 10-20 смертей на рік. Протягом усього періоду з 2010 по 2020 роки показники залишаються стабільними, з незначними змінами, що свідчить про відсутність значних демографічних криз у цих країнах.

Найбільше зростання рівня смертності спостерігається у кластері середньої смертності (жовтий), де пік досягається у 2019-2020 роках. Це може бути пов'язано з різними соціально-економічними факторами, що впливають на демографічні показники цих країн. Кластери низької та високої смертності демонструють відносно стабільні значення протягом досліджуваного періоду, без значних змін у показниках смертності, що свідчить про стабільність у цих країнах.

На рис. 3.15 наведено розподіл смертності від інфекцій нижніх та верхніх шляхів по країнам Європи за середнім значенням з 2010 по 2020 роки.

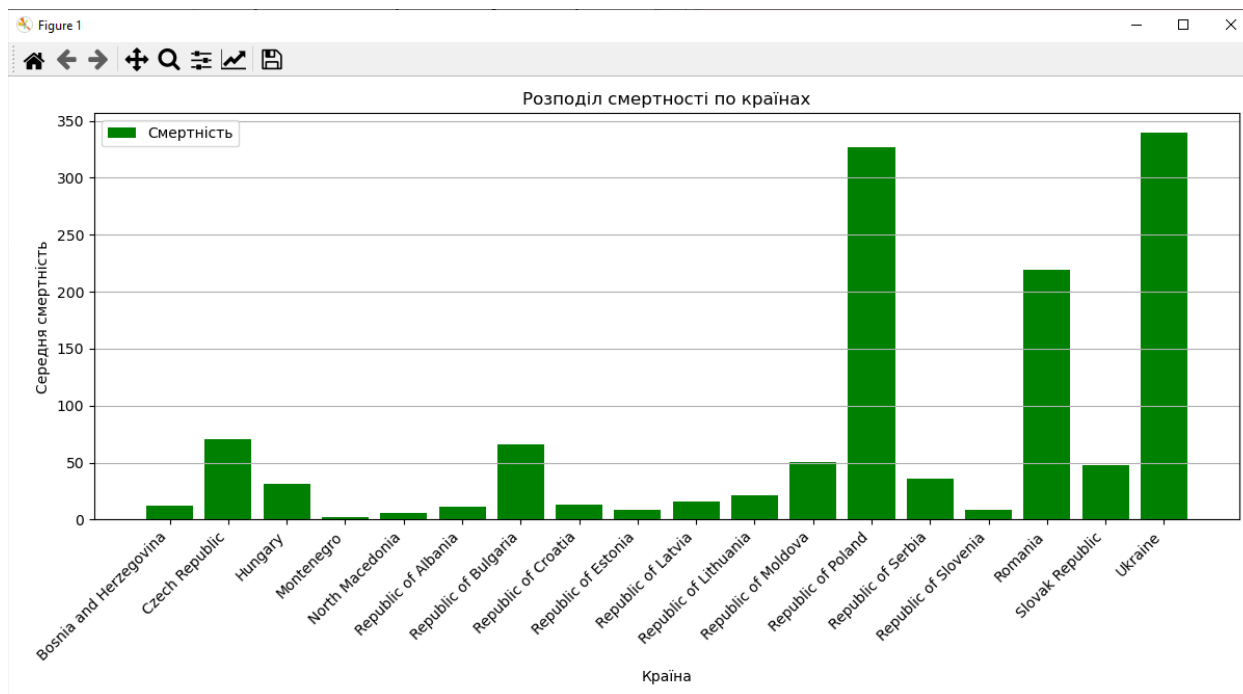


Рисунок 3.15 – Середнє значення розподілу смертності по країнам Європи

Середня смертність в країнах Чехії, Республіки Болгарії, Республіки Молдови та Словацької Республіки за весь період залишалася відносно стабільною, з легким зростанням до 2018 року. Спостерігається невелике зниження у 2021 році, що може бути пов'язане з наслідками пандемії COVID-19 або іншими факторами, такими як покращення системи охорони здоров'я.

Країни Республіка Польща, Румунія та Україна мають значно вищі показники смертності, ніж країни з низькою смертністю, з тенденцією до зростання протягом 2010-2019 років. Це зростання може бути пов'язане з різними факторами, включаючи соціально-економічні проблеми, доступ до медичних послуг та епідеміологічну ситуацію. Спостерігається падіння у 2021 році, ймовірно, через зміни в політиці охорони здоров'я або інші впливи, пов'язані з пандемією.

Смертність у країнах Боснія і Герцеговина, Угорщина, Чорногорія, Північна Македонія, Республіка Албанія, Республіка Хорватія, Республіка Естонія, Республіка Латвія, Республіка Литва, Республіка Сербія та Республіка Словенія є найнижчою серед трьох кластерів, але показує відносну стабільність з невеликими коливаннями. Спад у 2021 році може вказувати на позитивні зміни в системі охорони здоров'я або зменшення смертності внаслідок покращення умов життя чи доступу до медичних послуг.

На графіку рис.3.16 представлено прогнози смертності для різних країн на наступні 5 років. Вертикальна вісь відображає рівень смертності, а горизонтальна — прогнозовані роки, позначені як 2022, 2023 і так далі. Кожна країна представлена окремою смугою певного кольору, що дозволяє відстежити зміни в рівні смертності протягом прогнозованого періоду.

Провідні країни з найвищою смертністю. Україна (зелена смуга) має стабільно високі показники смертності протягом усіх прогнозованих років. Її рівень смертності перевищує 130 випадків у кожному з п'яти прогнозованих

років, що свідчить про серйозну демографічну проблему та високий рівень ризику для здоров'я населення в майбутньому. Подібна ситуація спостерігається в Румунії (сірий колір), де рівень смертності також залишається стабільно високим — приблизно 100 смертей у кожному з прогнозованих років.

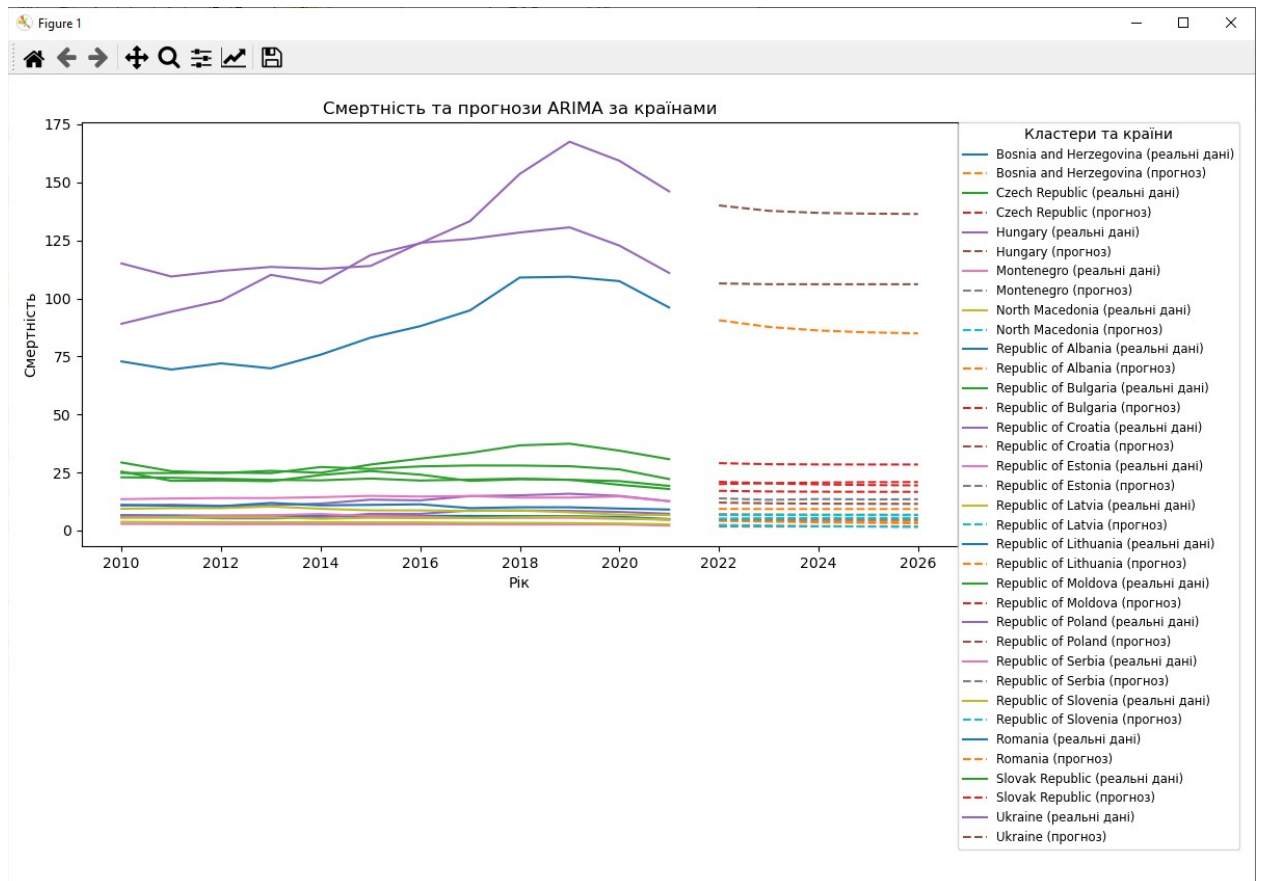


Рисунок 3.16 – Результат прогнозування смертності на наступні 5 років

Середній рівень смертності. Країни, такі як Болгарія (фіолетова смуга) та Польща (помаранчева смуга), демонструють середні значення смертності. Прогнозовані показники для цих країн коливаються в межах 20-40 випадків на рік. Це вказує на те, що в цих країнах смертність є стабільною, без різких підвищень або знижень протягом прогнозованого періоду.

Низький рівень смертності. Чехія (блакитна смуга) та Сербія (коричнева смуга) показують відносно низькі прогнози смертності в усі роки.

Рівень смертності в цих країнах не перевищує 20 випадків на рік, що свідчить про сприятливу демографічну ситуацію. Подібні низькі показники також спостерігаються в інших країнах, включених до аналізу.

Прогнози чітко показують, що Україна та Румунія залишатимуться лідерами за рівнем смертності в наступні п'ять років. Їхні стабільно високі показники смертності можуть свідчити про значні соціально-економічні або медичні проблеми, які впливатимуть на демографічну ситуацію. Водночас, інші країни з середнім та низьким рівнем смертності демонструють відносно стабільні прогнози, що не передбачають значних коливань у майбутньому.

Для оцінки ефективності кластеризації та прогнозування ми використовували такі метрики:

- середня квадратна помилка (MSE) для оцінки точності прогнозу ARIMA;

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (3.1)$$

- силует для оцінки якості кластеризації K-means;

$$S(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))} \quad (3.2)$$

- внутрішня дисперсія для оцінки компактності кластерів.

$$WCSS = \sum_{k=1}^K \sum_{i=1}^{n_k} (x_i - \mu_k)^2 \quad (3.3)$$

Результати оцінки ефективності застосування методів K-means та ARIMA представлені в таблиці 3.1.

Таблиця 3.1 – Результати оцінки метрик для ARIMA та K-means

Метод	Параметри	MSE	Силует
-------	-----------	-----	--------

ARIMA	(p=1, d=1, q=1)	23.45	N/A
K-means	k=3	N/A	0.45
Комбінований	ARIMA + K-means	19.32	0.50

З таблиці видно, що комбінований підхід, який об'єднує K-means та ARIMA, призводить до зниження середньої квадратної помилки (MSE) до 19.32 у порівнянні з використанням лише ARIMA, де MSE становить 23.45. Крім того, показник силуета для комбінованого методу становить 0.50, що свідчить про покращення якості кластеризації.

3.7 Висновки до розділу

Розроблена архітектура програмного забезпечення забезпечує ефективну та гнучку основу для реалізації всіх модулів системи. Використання модульного підходу дозволяє легко розширювати функціональність програми та забезпечує високу масштабованість. Інтеграція різних компонентів у єдину систему сприяє оптимізації процесів обробки даних та підвищенню точності аналізу.

У модулі обробки даних реалізовано ряд алгоритмів та методик для збору, очищення та підготовки даних до аналізу. Особливу увагу було приділено забезпеченню високої якості даних, що є критично важливим етапом у всьому процесі. Модуль кластеризації, заснований на алгоритмі K-means, показав задовільні результати в ідентифікації патернів у даних. Застосування цього методу дозволило групувати дані за спільними ознаками, що, у свою чергу, сприяло кращому розумінню структури даних. Модуль прогнозування, побудований на основі моделі ARIMA, продемонстрував свою ефективність у прогнозуванні показників смертності. Аналіз результатів показав, що комбіноване використання методів K-means та ARIMA

продемонструвало позитивні результати в оцінці та прогнозуванні показників смертності. Зниження значення середньої квадратної помилки (MSE) свідчить про підвищення точності прогнозування, а покращення метрики силуета вказує на кращу якість кластеризації.

Визначено, що метод ARIMA, що враховує часову залежність, в поєднанні з K-means, який ідентифікує подібності в даних, забезпечує більш точні та надійні прогнози. Ці результати підтверджують ефективність комплексного підходу до аналізу та прогнозування часових рядів, що дозволяє приймати обґрунтовані рішення на основі даних.

ВИСНОВКИ

У даній роботі було детально розглянуто комплексний підхід до аналізу та прогнозування тенденцій смертності в Європі, зосереджуючи увагу на інфекційних захворюваннях. Зокрема, ми проаналізували різноманітні моделі смертності та методи кластеризації, які можуть бути використані для ефективного дослідження впливу різних факторів на показники смертності населення.

Було виявлено, що географічні фактори істотно впливають на показники смертності, що підтверджує важливість врахування місця проживання в дослідженнях здоров'я населення. Модель Гомперца, яка розглядається в контексті вікової залежності смертності, виявилася корисним інструментом для прогнозування, оскільки вона враховує як вікові, так і гендерні відмінності, що робить результати більш точними і релевантними.

Дослідження методів кластеризації, зокрема K-середніх, показало їхню ефективність у виявленні закономірностей у даних про смертність. Цей підхід дозволив поділити населення на гомогенні групи, що суттєво сприяє розумінню структурних особливостей даних. Метод K-means надав можливість виявляти спільні ознаки серед різних груп населення, що в подальшому спростило аналіз і інтерпретацію отриманих результатів.

Запропонована методологія комбінованого застосування K-means кластеризації та ARIMA для моделювання тенденцій смертності об'єднує переваги кластеризації та часових моделей. Це дозволяє отримати більш точні прогнози та глибше розуміння факторів, які впливають на смертність. Розроблена архітектура програмного забезпечення забезпечила гнучкість і масштабованість, що важливо для реалізації всіх модулів системи.

Модуль обробки даних реалізує ряд алгоритмів для збору, очищення та підготовки даних до аналізу, що забезпечує високий рівень якості даних.

Висока якість даних є критично важливим етапом у всьому процесі, адже вона безпосередньо впливає на точність отриманих результатів.

Використання модулю прогнозування на основі моделі ARIMA продемонструвало свою ефективність у прогнозуванні показників смертності. Аналіз результатів показав, що комбіноване використання методів K-means та ARIMA забезпечує позитивні результати в оцінці та прогнозуванні показників смертності. Зниження значення середньої квадратної помилки (MSE) свідчить про підвищення точності прогнозування, а поліпшення метрики силуета вказує на покращення якості кластеризації.

Таким чином, ми визначили, що метод ARIMA, який враховує часову залежність, у поєднанні з K-means, що ідентифікує подібності в даних, забезпечує більш точні та надійні прогнози. Ці результати підтверджують ефективність комплексного підходу до аналізу та прогнозування часових рядів, що дозволяє приймати обґрунтовані рішення на основі даних, що є важливим аспектом у сфері охорони здоров'я та державної політики.

ПЕРЕЛІК ПОСИЛАНЬ

1. Закон України «Про захист населення від інфекційних хвороб» від 06.04.2000 р. No 1645-III. [Електронний ресурс]. Режим доступу: zakon.rada.gov.ua.
2. Андрейчин М.А. Біологічна зброя і біотероризм / М.А. Андрейчин, В.С.Копча. Здоров'я України : медична газета. 2004. No 5. С. 40–41.
3. Чумаченко Д.І. Математичні моделі та методи прогнозування епідемічних процесів: монографія [Текст] / Д.І. Чумаченко, Т.О. Чумаченко. – Харків: ТОВ «Планета-Прінт», 2020. – 180 с.
4. Основне про пандемії [Електронний ресурс] – режим доступу: http://esu.com.ua/search_articles.php?id=17931
5. Verardi V. COVID – SIR. [Електронний ресурс] – режим доступу: URL: <https://www.learningfromthecurve.net/epidemic-models/2020/04/13/covid-sir>
6. Hyndman, R. J. 2017. demography: Forecasting Mortality, Fertility, Migration and Population Data. R package version 1.20.
7. Hyndman, R. J. and G. Athanasopoulos. 2018. Forecasting: principles and practice. OTexts. [Електронний ресурс]. Режим доступу: <https://otexts.org/fpp2/>.
8. Mazzucco, S., B. Scarpa, and L. Zanotto. 2018. A mortality model based on a mixture distribution function. *Population Studies*, 1–10.
9. Bohk-Ewald, C. and Rau, R. (2017). Probabilistic mortality forecasting with varying age-specific survival improvements. *Genus* 73.1, p. 1. doi: 10.1186/s41118-016-0017-8.
10. Dokumentov, A., Hyndman, R. J., and Tickle, L. (2018). Bivariate smoothing of mortality surfaces with cohort and period ridges. *Stat* 7.1, e199. doi: 10.1002/sta4.199.

11. Doukhan, P., Pommeret, D., Rynkiewicz, J., and Salhi, Y. (2017). A class of random eld memory models for mortality forecasting. *Insurance: Mathematics and Economics* 77, pp. 97-110. doi: 10.1016/j.insmatheco.2017.08.010
12. Human Mortality Database (2019). University of California, Berkeley (USA), and Max Planck Institute for Demographic Research (Germany). Available at www.mortality.org or www.humanmortality.de (data downloaded on 2019-01-28).
13. Pascariu, M. D. 2018. MortalityLaws: Parametric Mortality Models, Life Tables and HMD. R package version 1.6.0.
14. Pascariu, M. D., V. Canudas-Romo, and W. J. Vaupel. 2018. The double-gap life expectancy forecasting model. *Insurance: Mathematics and Economics*.
15. Onambele, Luc & Guillen-Aguinaga, Sara & Guillen-Aguinaga, Laura & Ortega-Leon, Wilfrido & Montejo, Rocio & Alas-Brun, Rosa & Aguinaga-Ontoso, Enrique & Aguinaga-Ontoso, Ines & Guillen-Grima, Francisco. (2023). Trends, Projections, and Regional Disparities of Maternal Mortality in Africa (1990–2030): An ARIMA Forecasting Approach. *Epidemiologia*. 4. 322-351. [10.3390/epidemiologia4030032](https://doi.org/10.3390/epidemiologia4030032).
16. Статистичні дані [Електронний ресурс] Режим доступу: https://catalog.data.gov/dataset/?tags=deaths&res_format=CSV
17. Бібліотека Pandas [Електронний ресурс] Режим доступу: <https://pandas.pydata.org/>
18. Бібліотека sklearn [Електронний ресурс] Режим доступу: <https://sklearn.com/>
19. Бібліотека pandas [Електронний ресурс] Режим доступу: <https://pandas.io/>
20. Бібліотека matplotlib [Електронний ресурс] Режим доступу: matplotlib.com

Додаток А. Фрагмент лістингу програми

ARIMA.py

```

import pandas as pd
import matplotlib.pyplot as plt
from sklearn.cluster import KMeans
from statsmodels.tsa.arima.model import ARIMA

# Зчитування даних
data = pd.read_csv('mortality_data11.csv')

# Підготовка даних
pivot_data = data.pivot_table(index='location_name', columns='year', values='val')

# Заповнення NaN значеннями
pivot_data.fillna(0, inplace=True)

# Кластеризація
kmeans = KMeans(n_clusters=3, random_state=42)
kmeans.fit(pivot_data)

# Додавання кластеру до даних
pivot_data['Cluster'] = kmeans.labels_

# Визначення назв для кластерів
cluster_names = {0: "Низька смертність", 1: "Середня смертність", 2: "Висока смертність"}

# Виведення результатів
for cluster in range(3):
    cluster_countries = pivot_data[pivot_data['Cluster'] == cluster].index.tolist()
    print(f"Країни з кластером {cluster_names[cluster]}: {' '.join(cluster_countries)}")

# Створення графіку
plt.figure(figsize=(12, 6))

# ARIMA прогнозування
for country in pivot_data.index:
    # Вибір даних для країни
    country_data = pivot_data.loc[country].drop(['Cluster'])

    # Створення моделі ARIMA
    model = ARIMA(country_data, order=(1, 1, 1)) # Ви можете змінити порядок моделі
    model_fit = model.fit()

    # Прогнозування на наступні 5 років
    forecast = model_fit.forecast(steps=5)

    # Додаємо прогнози на графік
    plt.plot(country_data.index, country_data, label=f"{country} (реальні дані)")
    plt.plot(range(country_data.index[-1] + 1, country_data.index[-1] + 6), forecast, label=f"{country} (прогноз)",
              linestyle='--')

plt.title('Смертність та прогнози ARIMA за країнами')
plt.xlabel('Рік')
plt.ylabel('Смертність')
plt.legend(fontsize='small', title='Кластери та країни', title_fontsize='medium', loc='upper left', bbox_to_anchor=(1, 0.5))

```

K-mens.py

```

import pandas as pd
from sklearn.cluster import KMeans
import matplotlib.pyplot as plt
import geopandas as gpd
import numpy as np
import seaborn as sns

# Завантаження даних
data = pd.read_csv('mortality_data11.csv')

# Перегляд унікальних значень у стовпці 'val'
print(data['val'].unique())
# Фільтрація даних для захворювань верхніх і нижніх дихальних шляхів
respiratory_conditions = ['Upper respiratory infections', 'Lower respiratory infections']
filtered_data = data[data['cause_name'].isin(respiratory_conditions)]

# Перетворення даних для кластеризації
pivot_data = filtered_data.pivot_table(index='year', columns='location_name', values='val', aggfunc='mean').fillna(0)

# Переконайтеся, що всі значення числові
pivot_data = pivot_data.apply(pd.to_numeric, errors='coerce')

# Вивести інформацію про дані
print(pivot_data.info()) # Перегляд інформації про дані
print(pivot_data.isna().sum()) # Перегляд кількості NaN у кожному стовпці

# Фільтрація даних для нижніх дихальних шляхів
respiratory_conditions = ['Upper respiratory infections', 'Lower respiratory infections']
filtered_data = data[data['cause_name'].isin(respiratory_conditions)]

# Фільтрація для "Number" метрики
filtered_data = filtered_data[filtered_data['metric_name'] == 'Number']

# Перетворення даних для кластеризації
pivot_data = filtered_data.pivot_table(index='year', columns='location_name', values='val', aggfunc='mean').fillna(0)

# Переконайтеся, що всі значення числові
pivot_data = pivot_data.apply(pd.to_numeric, errors='coerce')

# Перевірка наявності NaN значень
if pivot_data.isnull().values.any():
    print("Є NaN значення. Заповнюємо їх нулями.")
    pivot_data = pivot_data.fillna(0)

# Виконання кластеризації K-means
n_clusters = 3 # Встановіть кількість кластерів

```

Додаток Б. Графічні дані

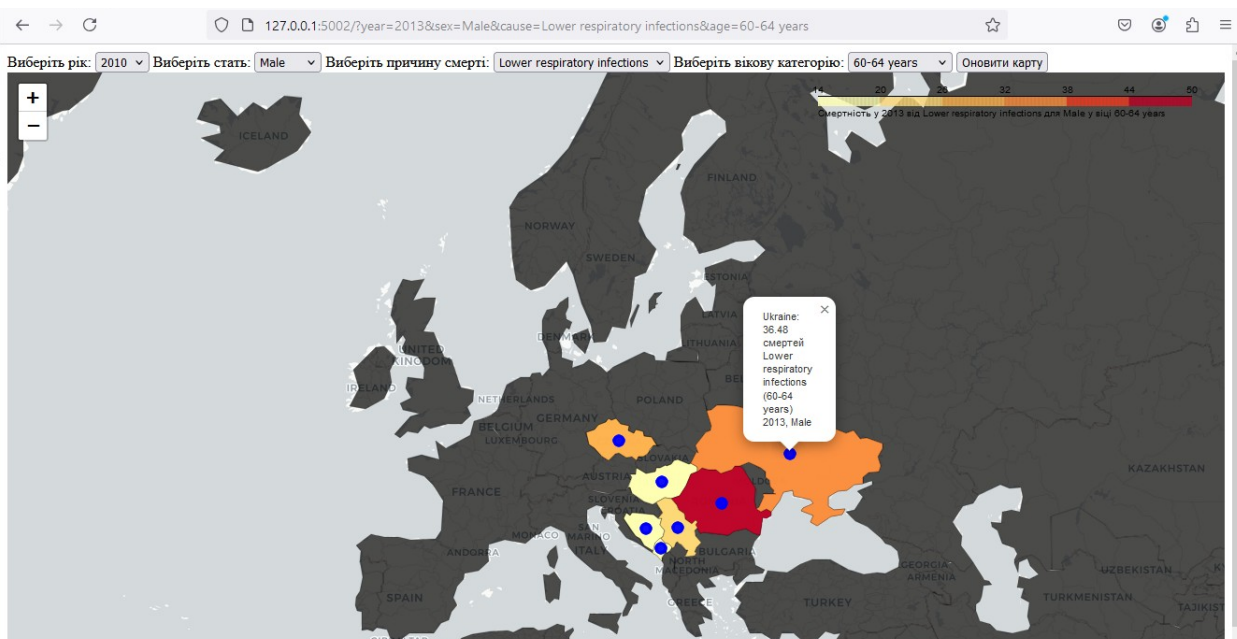


Рисунок Б.1 – Статистичні дані смертності для України у 2013 році (чоловіки, нижні дихальні шляхи, вік 60-64)

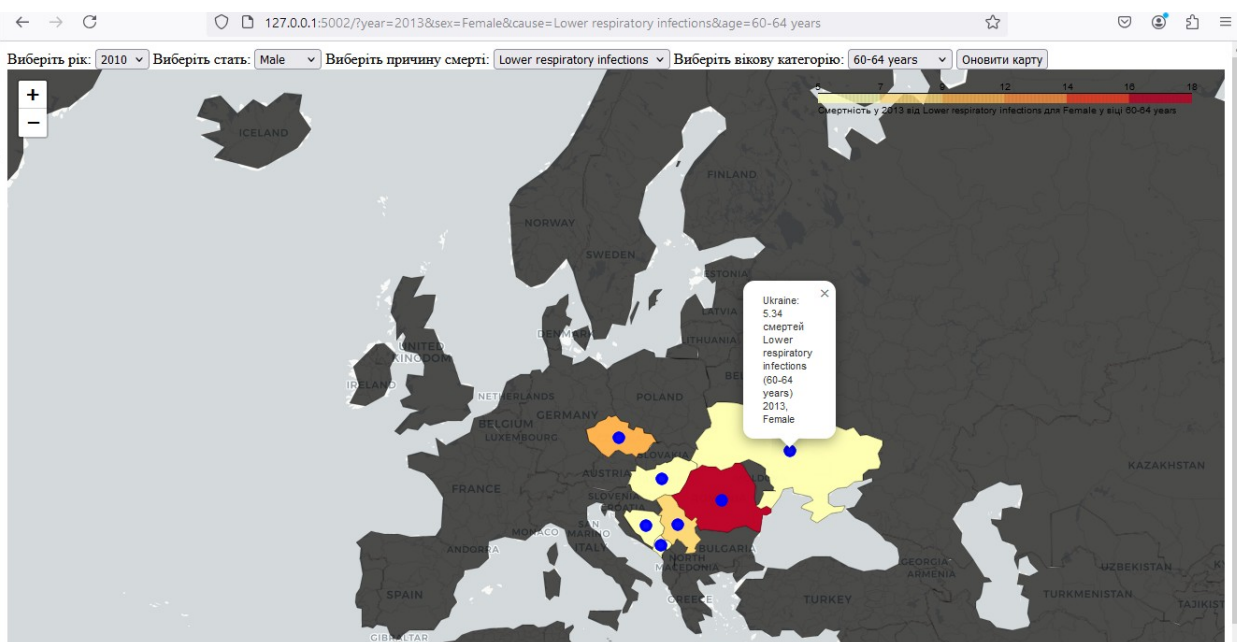


Рисунок Б.2 – Статистичні дані смертності для України у 2013 році (жінки, нижні дихальні шляхи, вік 60-64)

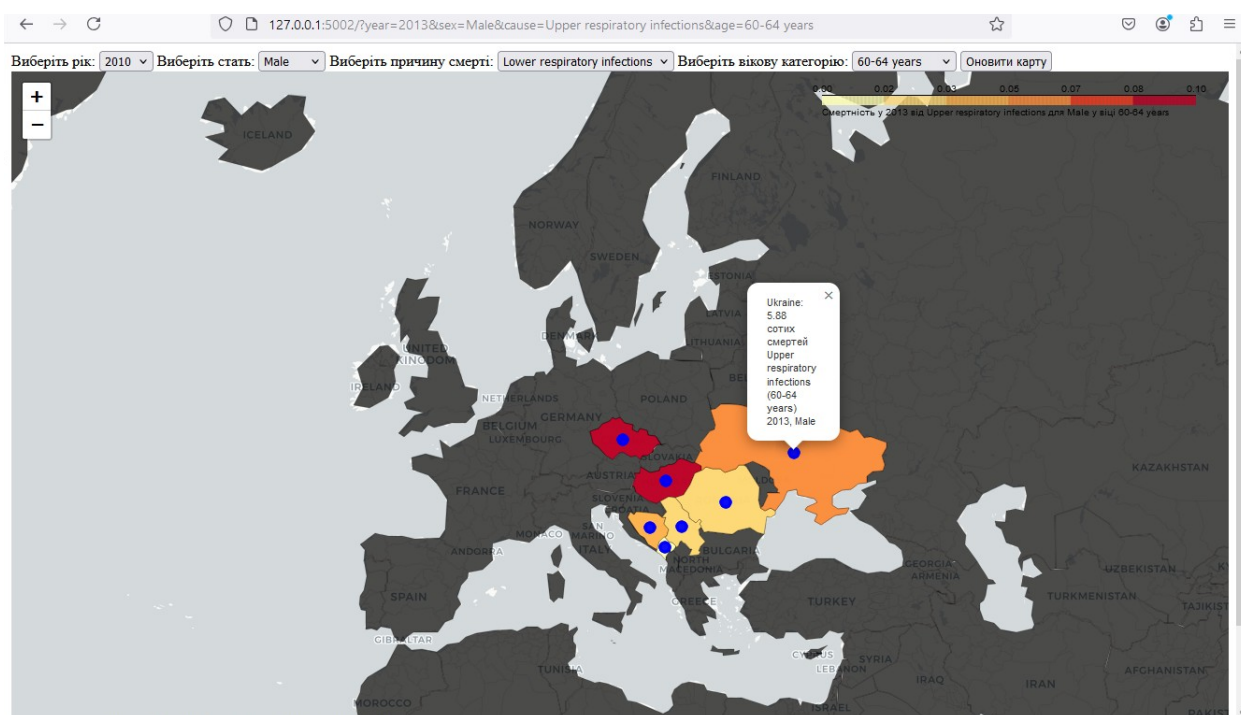


Рисунок Б.3 – Статистичні дані смертності для України у 2013 році (чоловіки, верхні дихальні шляхи, вік 60-64)

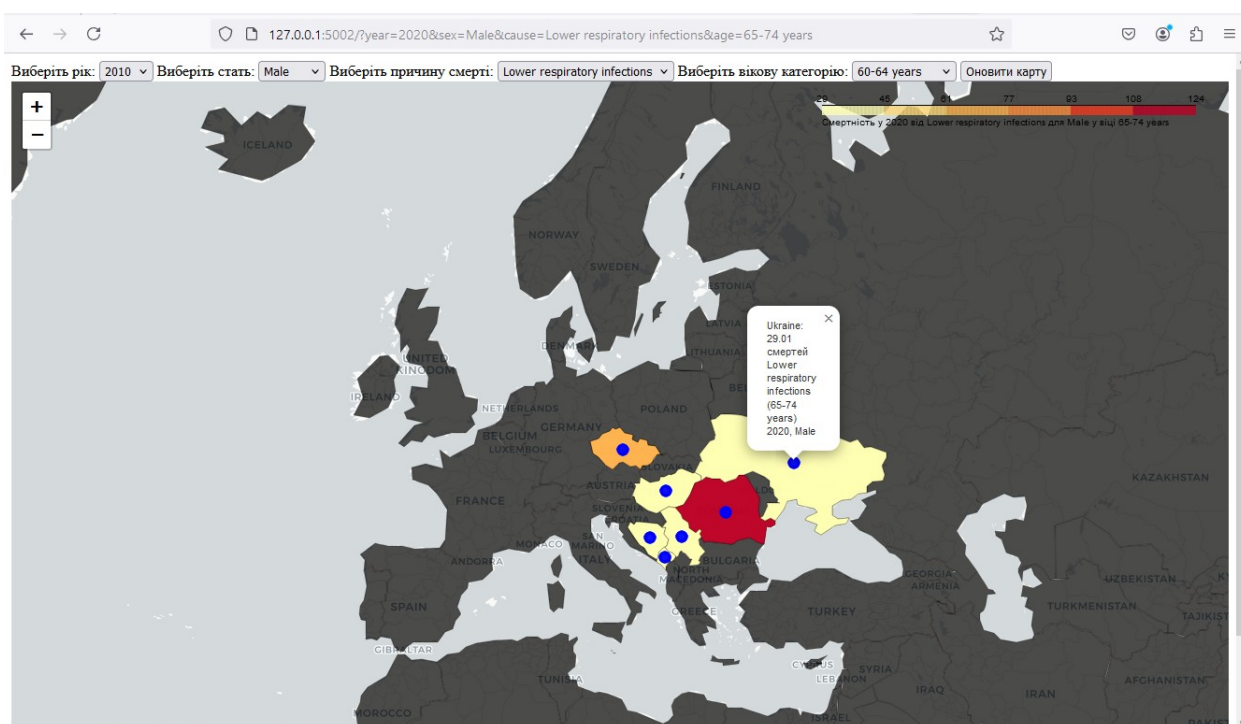


Рисунок Б.4 – Статистичні дані смертності для України у 2020 році (чоловіки, верхні дихальні шляхи, вік 65-74)

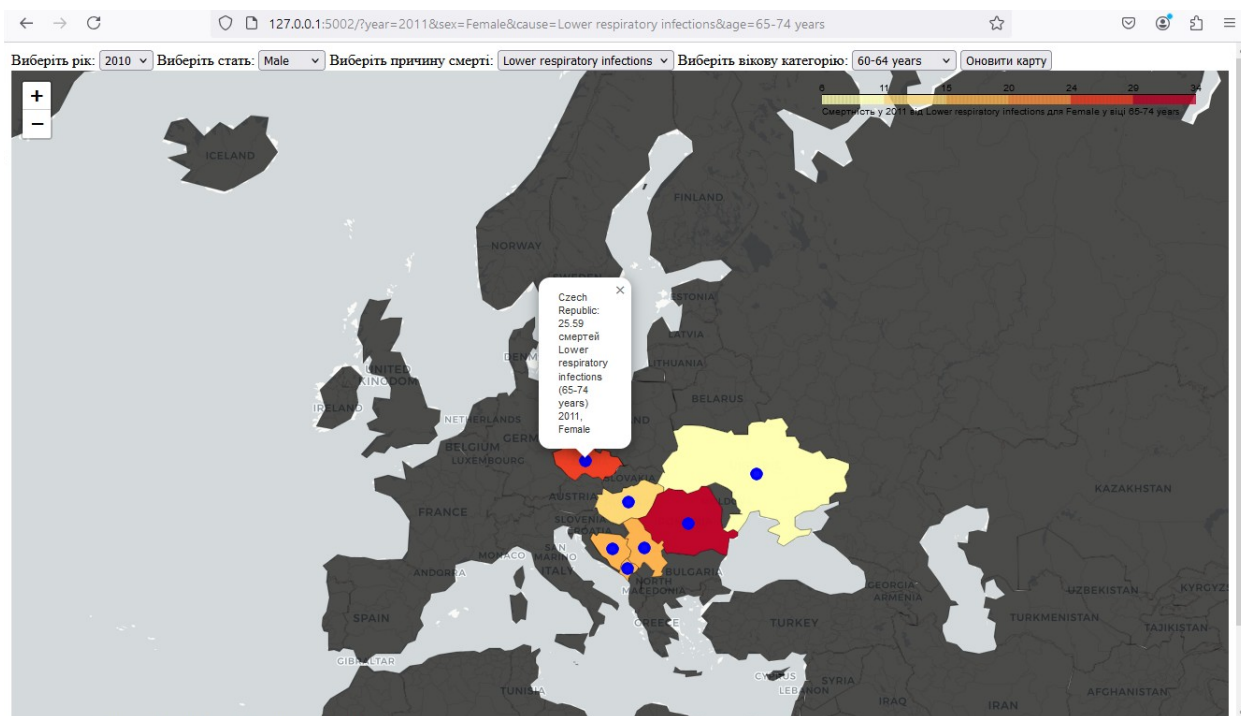


Рисунок Б.5 – Статистичні дані смертності для Чехії у 2011 році (чоловіки, верхні дихальні шляхи, вік 65-74)

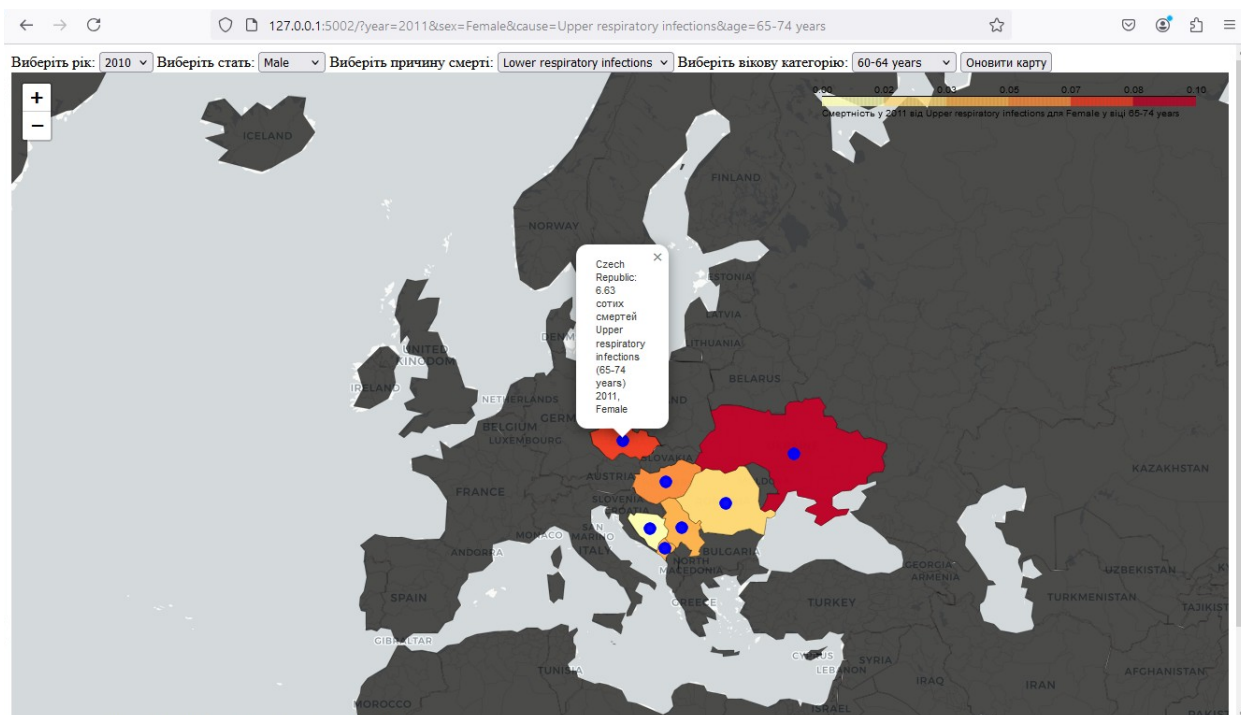


Рисунок Б.6 – Статистичні дані смертності для Чехії у 2011 році (жінки, верхні дихальні шляхи, вік 65-74)