

**НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ
«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ
імені ІГОРЯ СІКОРСЬКОГО»**

*Факультет інформатики та обчислювальної техніки
Кафедра автоматизованих систем обробки інформації та управління*

Рівень вищої освіти – *другий (магістерський)*

Спеціальність – *126 «Інформаційні системи та технології»*

Освітньо-наукова програма *«Інформаційні управляючі системи та технології»*

В.о. завідувача кафедри

_____ Олександр ПАВЛОВ

«__» _____ 2021 р.

**ЗАВДАННЯ
на магістерську дисертацію студенту**

Мадоян Гаяне Оганесівні

1. Тема дисертації «Прогнозування поширення COVID-2019 в Україні», науковий керівник дисертації Гуляницький Леонід Федорович, доктор технічних наук, затверджені наказом по університету від «12» березня 2021 р. № 809-с

2. Строк подання студентом дисертації “ 10 ” травня 20 21 р.

3. Об’єкт дослідження перебіг та процеси поширення захворювання.

4. Перелік завдань, які потрібно розробити:

- виконати огляд відповідної літератури, щоб визначити, які основні методи використовуються для аналізу процесів епідеміології;
- виконати аналітичний огляд та провести детальний аналіз існуючих сучасних методів прогнозування, порівняти їх точність;
- розробити експертний алгоритм прогнозування поширення COVID-19;
- розробити програмну реалізацію прогнозування COVID-19 у реальному часі, з використанням поточних статистичних даних;
- виконати експериментальне дослідження розроблених методів прогнозування та їх ефективності.

5. Орієнтовний перелік графічного (ілюстративного) матеріалу

Плакат 1 Функціонально-логічна структура програмного забезпечення

Плакат 2 UML-діаграма варіантів використання

Плакат 3 Діаграма послідовності

Плакат 4 UML-діаграма класів

Плакат 5 Математична постановка задачі

Плакат 6 Розроблений алгоритм прогнозування

6. Орієнтовний перелік публікацій

Мадоян Г. О. Математичні моделі прогнозування поширення Covid-19 та їх наслідків // *Priority directions of science and technology development. Abstracts of the 4th International scientific and practical conference. SPC "Sci-conf.com.ua". Kyiv, Ukraine. 2020. Pp. 461-465.*

Мадоян Г.О. Прогнозування поширення COVID-2019 в Україні / Г.О. Мадоян, Л.Ф. Гуляницький // *Матеріали VI всеукраїнської науково-практичної конференції молодих вчених та студентів «Інформаційні системи та технології управління» (ICTY-2021) – м. Київ.: НТУУ «КПІ ім. Ігоря Сікорського», 22-23 квітня 2021 р.*

7. Консультанти розділів дисертації

Розділ	Прізвище, ініціали та посада консультанта	Підпис, дата	
		завдання видав	завдання прийняв

8. Дата видачі завдання “ 1 ” лютого 20 21 р.

Календарний план

№ з/п	Назва етапів виконання магістерської дисертації	Строк виконання етапів магістерської дисертації	Примітка
1	<i>Аналіз предметної області</i>	<i>02.02</i>	
2	<i>Аналіз існуючих методів прогнозування</i>	<i>08.02</i>	
3	<i>Постановка та формалізація математичної моделі задачі</i>	<i>15.02</i>	
4	<i>Розробка алгоритму прогнозування</i>	<i>22.02</i>	
5	<i>Розробка інформаційного та програмного забезпечення</i>	<i>8.03</i>	
7	<i>Проведення експериментальних досліджень розроблених алгоритмів</i>	<i>22.03</i>	
8	<i>Оформлення документації</i>	<i>10.04</i>	
9	<i>Подання роботи на попередній захист</i>	<i>20.04</i>	
10	<i>Подання роботи на основний захист</i>	<i>10.05</i>	

Студент

Гаяне МАДОЯН

Науковий керівник

Леонід
ГУЛЯНИЦЬКИЙ

РЕФЕРАТ

Магістерська дисертація: 86 с., 15 рис., 15 табл., 41 джерело, 1 додаток.

Актуальність. Пандемія COVID-19, яка викликана коронавірусом SARS-CoV-2, призвела до глобальних соціально-економічних наслідків в усьому світі.

Прогнози випадків захворювання, кількості смертей та щоденних показників по госпіталізації можуть допомогти у інформування міністерств охорони здоров'я. Дана інформація допоможе при прийнятті рішень, так як можна хоча б частково спрогнозувати наслідки епідемії. Насамперед є важливим прогнозування випадків госпіталізації, так як важливо розуміти ситуацію по вільним ліжкам у регіонах.

На допомогу у прогнозуванні приходять математичні та зокрема статистичні моделі, хоча вони виконують завдання частково лише і надають короткострокові прогнози.

Варто впроваджувати модифіковані методи прогнозування, тим самим допомагаючи у розробці стратегій подальших дій та у прийманні рішень на державному рівні. За допомогою цих методів ми можемо оцінювати ситуацію у минулому, а таким чином це дозволить краще прогнозувати ситуацію, яка може скластися у майбутньому.

Зв'язок роботи з науковими програмами, планами, темами. Робота виконувалась на кафедрі автоматизованих систем обробки інформації та управління Національного технічного університету України «Київський політехнічний інститут ім. Ігоря Сікорського» в рамках теми «Ефективні методи розв'язання задач теорії розкладів» (Державний реєстраційний номер 0117U000919).

Метою дослідження є аналіз поширення гострого респіраторного вірусного захворювання, викликаного вірусом SARS-CoV-2 та підвищення ефективності прогнозування його в Україні.

Для досягнення поставленої мети необхідно виконати наступні **завдання**:

- виконати огляд відповідної літератури, щоб визначити, які основні методи використовуються для аналізу процесів епідеміології;
- виконати аналітичний огляд та провести детальний аналіз існуючих сучасних методів прогнозування, порівняти їх точність;
- розробити експертний алгоритм прогнозування поширення COVID-19;
- розробити програмну реалізацію прогнозування COVID-19 у реальному часі, з використанням поточних статистичних даних;
- виконати експериментальне дослідження розроблених методів прогнозування та їх ефективності.

Об’єктом дослідження є перебіг та процеси поширення захворювання.

Предметом дослідження є методи прогнозування унікальних процесів.

Методи дослідження. Для виконання поставлених завдань у роботі було використано: статистичні методи прогнозування (Гольта-Вінтерса, ARIMA) та метод прогнозування XGBoost з машинного навчання. Дані методи були використані для аналізу та порівняння результатів прогнозування, для подальшого створення нового алгоритму прогнозування.

Наукова новизна одержаних результатів полягає у розробленому алгоритму прогнозування, який на основі базового прогнозу коригує результати у відповідності до подібних часових рядів у минулому.

Публікації. Матеріали роботи опубліковані у збірнику Четвертої міжнародної науково-практичної конференції PRIORITY DIRECTIONS OF SCIENCE AND TECHNOLOGY DEVELOPMENT, а також в матеріалах VI всеукраїнської науково-практичної конференції молодих вчених та студентів «Інформаційні системи та технології управління» (ІСТУ-2021).

COVID-19, МЕТОДИ ПРОГНОЗУВАННЯ, ЧАСОВІ РЯДИ.

ABSTRACT

Master's dissertation: 86 pp., 15 figs., 15 tables, 41 sources, 1 appendix.

Topicality. The COVID-19 pandemic, caused by the SARS-CoV-2 coronavirus, has had global socio-economic consequences worldwide.

Disease forecasts, deaths, and daily hospitalization rates can help inform health ministries. This information will help in decision-making, as you can at least partially predict the consequences of the epidemic. First of all, it is important to predict hospitalizations, as it is important to understand the situation of free beds in the regions.

Mathematical and in particular statistical models come to the aid of forecasting, although they perform tasks only partially and provide short-term forecasts.

Modified forecasting methods should be introduced, thus helping to develop follow-up strategies and decision-making at the state level. With these methods, we can assess the situation in the past, and thus it will better predict the situation that may develop in the future.

Connection of work with scientific programs, plans, themes. The work was performed at the Department of Automated Information Processing and Control Systems of the National Technical University of Ukraine "Kyiv Polytechnic Institute. Igor Sikorsky" in the framework of the topic «Effective methods for solving problems of schedule theory» (No DR 0117U000919).

The purpose of the study is an analysis of the spread of acute respiratory viral disease caused by SARS-CoV-2 virus and increase the efficiency of its prediction in Ukraine.

To achieve this goal it is necessary to perform the following **tasks**:

- review the relevant literature to determine what the main methods are used to analyze the processes of epidemiology;
- perform an analytical review and conduct a detailed analysis of existing modern forecasting methods, compare their accuracy;
- to develop an expert algorithm for predicting the spread of COVID-19;

- to develop a software implementation of COVID-19 forecasting in real time, using current statistics;
- perform an experimental study of the developed forecasting methods and their effectiveness.

The object of research – there is research is the course and processes of disease spread.

The subject of research – there are methods for forecasting unique processes.

The research methods. To perform the tasks were used in the work: statistical forecasting methods (Holt-Winters, ARIMA) and the method of forecasting machine learning XGBoost. These methods were used to analyze and compare the results of forecasting, to further create a new forecasting algorithm.

The scientific novelty of the obtained results lies in the developed forecasting algorithm, which on the basis of the basic forecast adjusts the results in accordance with similar time series in the past.

Publications. Materials of the work are published in the collection of the Fourth International Scientific and Practical Conference PRIORITY DIRECTIONS OF SCIENCE AND TECHNOLOGY DEVELOPMENT, as well as in the materials of the VI All-Ukrainian Scientific and Practical Conference of Young Scientists and Students "Management and Technology" Information.

COVID-19, METHODS OF FORECASTING, TIME SERIES.

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ, СИМВОЛІВ, ОДИНИЦЬ ВЕЛИЧИН І ТЕРМІНІВ

ІС інформаційна система

ЧР часовий ряд

МГУА методи групового урахування аргументів

ОС операційна система

ПК персональний комп'ютер

ЗМІСТ

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ, СИМВОЛІВ, ОДИНИЦЬ ВЕЛИЧИН І ТЕРМІНІВ	7
ВСТУП.....	10
1 ОГЛЯД ІСНУЮЧИХ РІШЕНЬ З ПРОГНОЗУВАННЯ ПОШИРЕННЯ COVID- 19	11
1.1 Огляд предметної області.....	11
1.2 Показники та критерії, які використовуються при аналізі пандемічних процесів	14
1.3 Існуючі рішення з прогнозування.....	16
1.3.1 Імітаційне моделювання.....	18
1.3.2 Метод групового урахування аргументів	19
1.3.3 Епідеміологічні моделі	20
1.3.4 Статистичні методи прогнозування ЧР	25
1.3.5 Методи машинного навчання	28
Висновки до розділу	29
2 МОДЕЛІ ТА МЕТОДИ	30
2.1 Поширені математичні моделі	30
2.2 Математична постановка задачі	36
2.3 Алгоритми розв’язування задачі.....	37
2.4 Результати експериментальних досліджень	38
Висновки до розділу	41
3 ОПИС ПРОГРАМНОГО ТА ТЕХНІЧНОГО ЗАБЕЗПЕЧЕННЯ.....	43
3.1 Опис програмного забезпечення.....	43

	9
3.1.1 Вимоги до програмного забезпечення	43
3.1.2 Засоби розробки	44
3.1.3 Архітектура програмного забезпечення	49
3.1.4 Користувацький інтерфейс	51
Висновки до розділу	55
4 РОЗРОБКА СТАРТАП-ПРОЕКТУ	56
4.1 Опис ідеї проекту	56
4.2 Технологічний аудит ідеї проекту	59
4.3 Аналіз ринкових можливостей запуску стартап проекту	60
Висновки до розділу	72
ВИСНОВКИ	73
ПЕРЕЛІК ПОСИЛАНЬ.....	74
ДОДАТОК А.....	79
Графічний матеріал	79
Плакат 1 Функціонально-логічна структура програмного забезпечення	80
Плакат 2 UML-діаграма варіантів використання	81
Плакат 3 Діаграма послідовності.....	82
Плакат 4 UML-діаграма класів	83
Плакат 5 Математична постановка задачі	84
Плакат 6 Схема розробленого алгоритму	85

ВСТУП

Спалах "пневмонії невідомої етіології" на початку грудня 2019 року у Китаї перетворився на світову епідемію. Враховуючи статистичні дані, які оновлюються щодня, щодо вилікуваних, заражених та померлих від COVID-19 на карті поширення вірусу від Центру системних досліджень та інжинірингу при Університеті Джона Гопкінса (Johns Hopkins University), підтверджена кількість захворювань зростає у геометричній прогресії у більшості країн світу.

В інформації від університету мова йдеться саме про діагностовані випадки, а не фактичні, тому в результаті різниця може бути суттєвою. Офіційно зареєстрованих випадків набагато менше, ніж випадків захворювання в цілому. У деяких людей захворювання перебігає без помітних симптомів або взагалі без симптомів.

Прогнозування подальшого поширення COVID-19 є складною задачею, так як ми не володіємо повними даними, а ті що є – динамічно змінюються. Керівництвам країн потрібно вирішувати медичні, фінансові та регуляційні питання, а для цього треба володіти інформацією з упередженням, щоб розуміти, які можуть бути наслідки від їх рішень і мати розробленими наступні плани дій.

У роботі висвітлені питання прогнозування поширення COVID-19 в Україні, беручи до уваги поденні дані з кількістю госпіталізованих, а також будуть порівняні існуючі методи прогнозування та розроблено експертну модифікацію методу.

1 ОГЛЯД ІСНУЮЧИХ РІШЕНЬ З ПРОГНОЗУВАННЯ ПОШИРЕННЯ COVID-19

1.1 Огляд предметної області

Коронавіруси [1] — це велика група віруси з оболонкою РНК, важливість яких була визнана завдяки виявленню коронавірусу як збудника важкого гострого респіраторного синдрому (SARS). (“Coronavirus disease (COVID-19),” n.d.) Всесвітня пандемія COVID-19 викликана останнім виявом коронавірусу SARS-CoV-2, і є «третім задокументованим поширенням коронавірусу тварин на людей всього за три десятиліття, яке призвело до великої епідемії». Спалах SARS [2] у 2003 році призвів приблизно 10% (774 випадки смерті) летальних випадків, а в період з 2012 по 2019 рік від MERS було зафіксовано 2494 випадки смерті. Хоча показник летальності від COVID-19 становить кілька відсотків, пов'язана з цим пандемія викликала набагато більше смертей у всьому світі. Міжнародне дослідницьке співтовариство в даний час бореться з цією глобальною кризою у галузі охорони здоров'я. Робляться численні спроби розробити методи прогнозування еволюції пандемії в часі з цілями розробки політики, яка буде заснована на наукових даних. Дані мають три фази:

- експоненціальне зростання на ранній стадії (A);
- проміжний етап, перехідний від збільшення до зменшення (B);
- повільно-регресивний заключний етап (C).

Беручи до уваги цю характеристику, для досліджень використовується статистика екстремальних значень (EV) для оцінки щоденних чисел в (C) і загального числа смертей за даними (A) і (B).

Пандемія коронавірусної хвороби 2019-2021 — пандемія, яка була спричинена одноланцюговим РНК-вмісним штамом SARS-CoV-2 [3]. Спалах захворювання вперше було виявлено у грудні 2019 у м.Ухань, Хубей, КНР. Спалах був визнаний ВООЗ пандемією у березні 2020 р., а саме 11 березня. Станом на 19 квітні 2021 року маємо понад 141 мільйон випадків по всьому світу,

майже у всіх країнах та територіях. Унаслідок захворювання померло понад 3 мільйони осіб, близько 80,4 мільйонів осіб одужали.

У липні 2020 року Тедросом Аданом Гебрейесусом, який є головою Всесвітньої організації охорони здоров'я, було названо основні варіанти розвитку подій, тобто сценарії поширення коронавірусу у різних країнах світу [4]:

- перший сценарій відноситься до країн, які були заздалегідь попереджені і мали уявлення про майбутній спалах хвороби — в результаті яким вдалося уникнути масштабних спалахів (деякі з країн Карибського басейну, Південно-Східної Азії, Тихоокеанського регіону та Африки);

- другий сценарій стосувався багатьох країн Європи — масштабні спалахи захворювання, які вдалося контролювати завдяки діям керівництва;

- третій сценарій розгортався у країнах, де вдалося подолати перший спалах хвороби, але були послаблені обмеження, що призвело до необхідності боротьби з подальшими хвилями захворювання;

- четвертий сценарій несе під собою фазу інтенсивної передачі інфекції (Америка, Південна Азія і кілька країн Африки).

Передача вірусу відбувається у переважній більшості випадків саме повітряно-крапельним шляхом та під час близьких контактів [5]. Можна заразитися при контакті із зараженою поверхнею, доторкнувшись потім до обличчя, наприклад, яке має слизові оболонки. Вважається, що найбільш заразною особою є людина з наявними симптомами, проте можливе зараження від людини без симптомів, бо їх може і не бути. Між моментом зараження та початком появи перших симптомів проходить від 2 до 14 днів. До поширених відомих симптомів відносяться гарячка, кашель та задишка. Також ускладнення можуть нести за собою гострий респіраторний дистрес-синдром та запалення легень.

Не існувало вакцини до недавнього часу та специфічних противірусних препаратів. Первинне лікування — суто підтримувальне та симптоматичне. До

профілактичних заходів відносяться соціальне дистанціювання, самоізоляція осіб з підозрою на зараження та дотримання правил особистої гігієни.

До заходів, які запобігають поширенню вірусу входять введення карантину та комендантської години, обмеження подорожей, перенесення або скасування різноманітних заходів, контроль ризику на робочому місці та закриття деяких об'єктів. До інших заходів можна віднести сувору ізоляцію, наприклад в Хубей; введення регіональних та загальнонаціональних карантинів у інших країнах; обмеження в'їзду пасажирів та закриття кордонів, введення перевірок на вокзалах та в аеропортах; введення заборони на виїзд пасажирів. Всі навчальні заклади були зачинені на місцевому або національному рівнях в більш ніж 124 країнах, що значно вплинуло на понад 1,2 млрд студентів.

Пандемія призвела до глобальних соціально-економічних серйозних наслідків. Усі скасування чи відтермінування культурних, спортивних, та релігійних подій, а також поширення побоювань з приводу дефіциту постачання усіляких необхідних продуктів та товарів, що спричинили у певних регіонах панічні покупки. Також інтернетом поширювалася дезінформація та різноманітні теорії змови про вірус, які стосуються природи виникнення вірусу. У низці країн спостерігалися випадки ксенофобії та расизму по відношенню до жителів Китаю. У деяких країнах Східної та Південно-Східної Азії навіть спалахнули протести.

Прогнози кількості смертей, госпіталізацій і випадків захворювання допоможуть інформувати галузь охорони здоров'я при прийнятті рішень, прогнозуючи можливі наслідки пандемії COVID-19, тому дуже важливим є прогнозування випадків, зайнятих ліжок по лікарням на кілька тижнів вперед.

Математичні моделі справляються із завданням лише частково і видають короткострокові й досить мінливі прогнози. Спиратися на досвід інших країн корисно, але, знову ж таки, не дуже надійно: виявилося, і менталітет, і дисциплінованість громадян різних країн значно впливають на перебіг і поширення коронавірусної інфекції. Щоб будувати хоч приблизні версії про

майбутнє, багато вчених шукають надійні методи в минулому. Один з таких давно існує в медицині - експертна оцінка відразу декількох досвідчених фахівців.

Можна впровадити методи прогнозування, тим самим допомагаючи розробляти кращі стратегії і приймати раціональні рішення. Ці методи дозволяють оцінювати ситуації в минулому, і таким чином це дозволяє краще прогнозувати ситуацію, яка може скластися у майбутньому. Враховуючи унікальність нинішньої пандемії, важливо також розроблювати прогнози, які враховують і можливі сценарії розвитку епідемії, чим можуть допомогти підготуватися до потенційних загроз і наслідків.

1.2 Показники та критерії, які використовуються при аналізі пандемічних процесів

Всесвітня організація охорони здоров'я розробила «COVID-19 Strategic Preparedness and Response. Monitoring and Evaluation Framework» [6]. Метою цієї системи моніторингу та оцінки є підтримка країн ЄС та європейської економічної зони та Великобританії у їхніх зусиллях щодо моніторингу впровадження та наслідків заходів реагування на COVID-19. Структура подає показники для ряду ключових опор діяльності з підготовки, запобігання та контролю COVID-19 та надає керівництво країнам щодо збору та аналізу даних для запропонованих показників [7]–[9].

Для аналізу та подальшого прогнозування враховуються різні статистичні, математичні та медичні параметри. Одними з найважливіших параметрів є кількість смертей, кількість переносників, інкубаційний період, доступність медичних закладів, швидкість зараження після контакту, вік хворих, час виявлення та інші.

Щоденно оновлюються статистичні дані, які відображають кількість підтверджених випадків, кількість інфікованих, одужавших та померлих. На

сторінці Міністерства охорони здоров'я України надається оперативна інформація щодо забезпеченості закладів у вигляді інтерактивного Power BI звіту [10]. На рисунку 1.1 наведений скріншот екрану, на якому продемонстрована кількість вільних місць по регіональним центрам України.

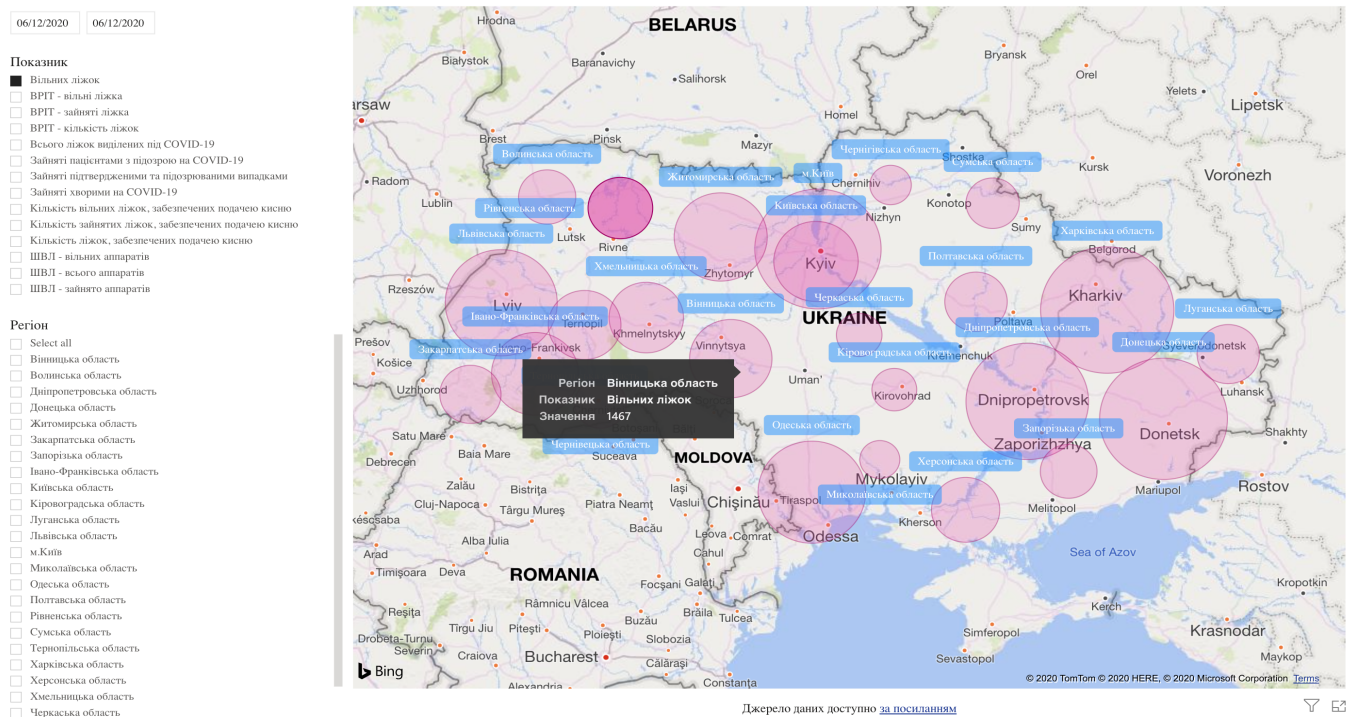


Рисунок 1.1 – Кількість вільних ліжко-місць по регіональним центрам

Одним з важливих показників є коефіцієнт поширення коронавірусу R_t . Цей коефіцієнт показує кількість зареєстрованих хворих за останні t днів, поділену на кількість підтверджених випадків інфікування за ці дні. Також існує коефіцієнт летальності, який відповідає кількості зареєстрованих смертей, яка поділена на кількість зареєстрованих випадків протягом певного періоду. Частота смертей від інфекції визначається як кількість зареєстрованих смертей поділену на кількість інфікованих. Часто установлюють $t = 7$, тобто йдеться мова про тижневе рухоме вікно. Вважається, що його значення менше 1 протягом останніх t днів і дозволяє говорити про загасання епідемії [11], [12], [13].

Проте летальність більшою мірою визначається у відповідності з рівнем та доступністю надання медичної допомоги.

1.3 Існуючі рішення з прогнозування

Підходи для вирішення проблеми прогнозування поділяються на такі категорії:

- імітаційне моделювання. З допомогою даного напрямку можливо спрогнозувати швидкість, з якою може розповсюджуватися епідемія, а також час «критичного порогу» по завантаженню основних ресурсів допомоги населенню;
- метод групового урахування аргументів (МГУА);
- епідеміологічні моделі. Для моделювання поширення інфекційних захворювань використовується класична модель SIR, а також її модифікації [14]. Математичне моделювання в епідеміологічній області дає можливість змодельовати появу та розповсюдження хвороботворних мікроорганізмів;
- статистичні методи прогнозування ЧР;
- методи машинного навчання.

У [15] описаний стан прогнозування COVID-19 станом на червень 2020 року, коли вже пройшло майже 7 місяців з початку поширення вірусу. Для точного прогнозування дуже важливим є розуміння природного прогресування захворювання. Хвороба, як правило, прогресує через вплив інфекції.

Розглядаються чотири важливі фактори, які беруть участь у формуванні епідеміологічної ситуації — фактори навколишнього середовища, носій збудника, господар заражений та патогени або інфекційні агенти. Інфекція передається господареві, коли носій збудника контактує з господарем у певному середовищі.

Авторами були визначені та формалізовані наступні проблеми, пов'язані з прогнозуванням подальшої ситуації:

- відслідковування заражених осіб та осіб, які контактували з ними;
- довгий інкубаційний період. Так як інкубаційний період складає 14 днів, то стає неможливим визначення стану особи заздалегідь;

- недостатньо належних даних. Іноді ці дані не структуровані або надходять у різний час, а точність даних є однією з запорук ефективної моделі прогнозування;
- перенавчання моделі. Може статися перенавчання моделі прогнозування, що призведе до некоректних результатів на нових даних;
- надто очищені дані. Очистка даних є важливим етапом на початку прогнозування, проте з цим потрібно бути обережними;
- велика кількість даних. Використання великої кількості показників із даних теж може негативно вплинути на точність моделі, що варто враховувати;
- неправильний вибір алгоритму та ознак на вхід;
- складність обраної моделі. Якщо модель є складною, то це може значно вплинути на загальну продуктивність моделі.

Прогнозні моделі використовуються для формування прогнозів залежно від базової моделі. Точність моделі залежить від базового алгоритму. До цього часу статистична модель використовувалась для прогнозування або визначення зв'язку між атрибутами або змінними. Що стосується COVID-19, ситуація різко змінюється з кожним днем. Як раніше бачили, дані збираються у багатьох різних форматах, а також обсяг даних величезний. У цьому сценарії моделі глибокого навчання можуть стати в нагоді. Алгоритми глибокого навчання можуть забезпечити кращі результати. Якщо в майбутньому доступні дані часових рядів, алгоритми ШІ можуть бути корисними.

У цьому дослідженні було представлено загальне комплексне дослідження з аналізу COVID-19, прогнозування його поширення, наслідків та заходів контролю. Основним внеском цього дослідження науковці бачать аналіз декількох моделей прогнозування, знаходження проблем цих моделей та надання рекомендацій щодо боротьби з цією пандемією. На основі доступних методів прогнозування вони вивчали різні статистичні, аналітичні, математичні та медичні (симптоматичні та безсимптомні) параметри. Було проведено

категоризацію різних методів прогнозування на чотири основні набори, що включають великі масиви даних з різних джерел, таких як ВООЗ, Національні джерела даних, соціальні медіа. Наведена класифікація безсумнівно може допомогти дослідникам наштовхнутися на ідеї для подальшого прогнозування

Було вирішено розглянути по кожному з підходів наукові роботи та детально ознайомитися з отриманими результатами.

1.3.1 Імітаційне моделювання

Так у роботі [16] було взято за основу індивідуальну модель для того, щоб змодельовати поширення COVID-19 в одному з міст Південної Кореї. Так як індивід може заразитися знаходячись у людних місцях, наприклад, в школах або на роботі, під час громадських або релігійних зборів. Для дослідження штучно створили синтетичну вибірку на основі даних по перепису населення. Автори хотіли отримати уявлення про поточну епідемію, для прогнозування майбутнього поширення, а також проаналізували як вплине відкриття шкіл на цей процес.

У роботі [17] показана симуляція для сценарію з 25 тисячами студентів та трьома тисячами робітників, які прибувають у кампус на семестр тривалістю 90 днів. Кінець симуляції відбувається тоді, коли досягнуто значення у 100 випадків зараження на день. Для моделювання авторами був проаналізований перелік параметрів:

- швидкість прибуття студентів;
- динаміка поширення;
- кількість тестувань;
- ймовірність зараження при контакті;
- вікові категорії та ризики для персоналу;
- оцінка ймовірності летальних випадків;
- частота контактів та дотримання дистанції.

Були отримані кілька сценаріїв, де кількість випадків починає зростати у геометричній прогресії з перших днів, що означає переведення університету на дистанційну роботу до кінця семестру. При моделюванні не були розглянуті фактори, пов'язані зі складністю надання медичних послуг, хоча спалах у містечку, яке не готове до великої кількості пацієнтів, може додатково навантажити систему охорони здоров'я, впливаючи на послуги та час відновлення. Автори дійшли до того, що можливо отримати більше користі з їх моделі, шляхом надання фактичних статистичних даних щодо застосування масок та політики соціального дистанціювання. Загалом припускається, що дана модель може бути корисною та перевіреною на більшій вибірці університетів.

Говорячи про імітаційне моделювання в цілому, можна зробити висновок, що це досить важливий підхід для розуміння поширення COVID-19. До недоліків можна віднести недостатню кількість статистичних даних. Наразі можливо тільки штучно моделювати усі ті параметри, які були виділені авторами попередньої роботи.

1.3.2 Метод групового урахування аргументів

Даний підхід є дуже корисним, оскільки на базі прогнозу, який можна отримати за допомогою будь-якої статистичної моделі, можна робити різні модифікації. У [18] подані результати моделювання процесу пандемії в Україні Covid-19 з використанням офіційних статистичних даних про підтверджені випадки. Основною метою роботи є визначення динамічних закономірностей процесу, який подаються як ЧР. У роботі використовуються чотири методи для побудови прогнозних різницевих моделей типу авторегресії. Базовим вважають прогноз, який отримується за допомогою звичайної авторегресії. Далі авторами порівнюються модифіковані прогнози, які отримані з базового прогнозу за допомогою комбінаторно-генетичного алгоритму GMDH COMBI-GA та методу Lasso. Це дослідження корисне для визначення тенденції розвитку коронавірусу

у часі та прогнозування його подальшої поширення з метою вжиття певних захисних заходів. Прогноз за допомогою комбінаторно-генетичного алгоритму виявився найбільш показовим.

1.3.3 Епідеміологічні моделі

У [19] при прогнозуванні епідемії в Індії порівнюються модель SIR та модифікована Logistic Growth модель. У випадку з COVID-19 вважається, що максимальне обмеження буде визначатися загальною чисельністю населення, і приріст обов'язково знизиться, коли більша частина населення перехворіє. Причина використання Logistic Growth для моделювання спалаху захворюваності на COVID-19 базується на доказах того, що епідемія характеризується експоненціальним зростанням на ранніх стадіях і певним спадом пізніше (втім, досвід 2020 р. показав, що на відміну від інших інфекцій, спостерігаються хвилеподібні спалахи і в подальшому).

В цій роботі Logistic Growth модель використали для короткострокового прогнозування, а модель SIR для того, щоб спрогнозувати максимальну кількість активних випадків захворювання та пік захворюваності. Також була задіяна модель Time Interrupted Regression для оцінки впливу карантину та інших втручань. Були порівняні результати, які дали перші дві обрані моделі для прогнозування сукупної кількості захворювань в Індії та 4 інших регіонах з високою захворюваністю. Після порівняння результатів вийшло, що модель Logistic Growth добре впоралася з прогнозуванням сукупної кількості захворювань. Дана модель підійшла у цьому випадку для прогнозування, тому що епідемія має експоненціальний ріст, проте потім вона має стабілізуватися, на відміну від експоненціальної моделі. Загалом модель Logistic Growth дала точні результати, з урахуванням фактичних значень захворюваності в Індії, проте на короткострокову перспективу.

У [20] розглядаються варіанти прогнозування поширення коронавірусної інфекції у кількох країнах світу: Китай, Південна Корея, Індія, Італія, Австралія, США та окремо штат Техас. Досліджується еволюція різних популяцій та розглядаються різноманітні важливі параметри поширення захворювання у різних регіонах. Для прогнозування було взято набір даних, який містив дати, загальну кількість випадків, кількість інфікованих та кількість летальних випадків відповідно.

Наведені результати попереднього аналізу показали, що застосування нелінійних перетворень даних для тренування моделі на підготованих даних при найкращих параметрах видало гірший результат, ніж застосування розроблених методів у статті. Однією з причин отриманих результатів можна рахувати те, що були представлені дані з різних країн, а це означає, що методологія їх збору була не одна й та сама.

Засоби математичного моделювання є ефективними інструментами у розумінні закономірностей спалахів хвороб. Вони дають нам корисні прогнози в контексті впливу втручання на зменшення кількості захворювань.

У цій роботі була розширена класична модель SIR з можливістю компенсувати стрибки кількості сприйнятливих осіб, доповнену записаними даними з Китаю, Південної Кореї, Індії, Австралії, США, Італії та штату Техас у США, які надають уявлення про поширення COVID-19 у громадах. В усіх випадках моделі прогнозування можуть бути застосованими до публічних даних, причому деякі з них підходять краще, ніж інші. Для Китаю фактична кількість інфекцій впала швидше, ніж передбачали моделі, що є показником успіху заходів, які були вчасно реалізовані китайським урядом. У середині квітня у Китаї спостерігався стрибок кількості смертей, що призвело до менш надійних оцінок кількості смертей, передбачених моделлю SIR. Сприйнятливий населення в Південній Кореї дуже швидко впало до нуля, що свідчить про те, що уряд швидко прийняв дії у боротьбі із розповсюдженням вірусу. На початок червня 2020 р.

пікової кількості заражень в Індії ще не було досягнуто. Отже, модельні прогнози дають лише мінімальні оцінки тривалості пандемії в країні, загальної сукупної кількості інфекцій та смертей. Тематичне дослідження вірусу в Австралії показує важливість включення сплесків, де кількість сприйнятливих особин може бути збільшена. Ці сплески може бути пов'язані із прибуттям заражених людей з-за кордону та інфікованих людей з круїзного судна Ruby Princess. Також цікавим прикладом є дані із США, оскільки існує безліч епіцентрів вірусу, що виникають у різний час. Це ускладнює вибір відповідних параметрів моделі. Результати для Техасу показують, що модель може бути застосована до громад, крім країн. Італія наводить приклад, коли існує прекрасна збіжність між опублікованими даними та модельними прогнозами.

Таким чином, застосована модель SIR у цьому випадку забезпечує теоретичну базу для дослідження поширення вірусу COVID-19 у межах спільнот. Модель може дати уявлення про еволюцію поширення вірусу в часі, чого лише дані не дають. У цьому контексті це може бути застосовано до спільнот за наявності надійних даних. Його сила також полягає в тому, що в міру додавання нових даних до моделі, легко регулювати її параметри та забезпечувати найкращі результати враховуючи реальні дані та прогнози з моделі. Результати досліджень показують, що модель SIR є більш придатною для прогнозування епідемічної тенденції через поширення хвороби, оскільки вона може стримувати сплески та коригуватися враховуючи реальні дані. Порівнюючи опубліковані дані з прогнозами, можна передбачити успіх державних втручань. Дані, що розглядалися, бралися між січнем та червнем 2020 року, що містить набори даних до та під час здійснення жорстких та контрольних заходів. Проведений аналіз також підтвердив успіх і невдачі в деяких країнах щодо вжитих заходів контролю.

Поширення коронавірусної інфекції в Україні почалося з виявлених перших випадків у Чернівецькій області. Після цього, враховуючи світову

статистику, питання прогнозування поширення коронавірусної інфекції почало стояти гостро.

Автори досліджень наголошують, що повинні бути вжиті суворі адекватні заходи для подальшого запобігання та контролю поширення COVID-19. Країни світу вжили заходів для зменшення кількості заражених громадян, наприклад, заходи блокування, програми інформування, що рекламуються за допомогою засобів масової інформації, дезінфекційні кампанії для рук тощо, для того щоб уповільнити передачу хвороби. Додаткові заходи, включаючи підходи до раннього виявлення та ізоляцію сприйнятливих осіб, щоб уникнути змішування їх з безсимптомними та самокарантинними особами, обмеження руху та медичне лікування показали, що вони можуть допомогти запобігти збільшенню кількості заражених осіб. Посильна політика блокування може застосовуватися в різних сферах, якщо це можливо. Відповідно до цього, у країнах із високим рівнем захворюваності на COVID-19 необхідно якнайшвидше впровадити необхідну політику охорони здоров'я, щоб контролювати її поширення. Модель SIR, яка використовується тут, є простою, і, отже, прогнози, які виходять, можуть бути недостатньо точними, що також залежить від опублікованих даних та їх надійності. Однак, як показують дані моделі, одне, що є певним, це те, що COVID-19 не збирається рухатись швидко чи легко.

Національна академія наук України регулярно надає на своєму сайті прогнози розвитку епідемії, які розроблені робочими групами з математичного моделювання проблем та від учених Інституту проблем ринку та економіко-екологічних досліджень НАН України. Звіт робочої групи подає аналіз первинних статистичних даних розвитку епідемії, порівняння нових даних з результатами попереднього прогнозу, а також прогноз на черговий період (тиждень чи два) [21]. Для аналізу розвитку епідемії в Україні використовуються дані, які оприлюднюються на регулярних щоденних брифінгах, дані з сайту Ради національної безпеки і оборони України, а також з інших джерел.

У цьому звіті однією з моделей для прогнозування була обрана модель Facebook Prophet. Ця модель подається у вигляді рівняння:

$$y(t) = g(t) + s(t) + h(t) + \varepsilon_t \quad (1.1)$$

Модель з рівняння (1.1) містить в собі:

- тренд, з логістичною чи лінійною функцією;
- компонент сезонності, який відповідає за моделювання періодичних змін, що пов'язані з тижневою та річною сезонністю;
- компонент, який відповідає за аномальні дні, в тому числі і за певні нерегулярні дні, наприклад, як чорна п'ятниця;
- помилку, яка містить інформацію, яка не врахована моделлю.

Робочою групою було обрано саме цю модель, бо вона враховує аномальні дати, які можуть бути святами, дати потенційного зростання людей на відпочинку, а також дати, коли послаблювалися карантинні обмеження. Враховуючи всі параметри, було побудовано модель з помилкою в 3.5%.

Другою була обрана компартментна модель SEIR-U. Дана модель входить до моделей сімейства SIR, тільки саме цей варіант був розроблений робочою групою при Президії НАН України. Модель SEIR-U враховує наявність хворих, у яких немає симптомів, а також озволяє обчислювати кількість пацієнтів, які госпіталізовані. Математичний апарат моделей SEIR подається у вигляді чисельного розв'язку систем диференціальних рівнянь.

Після порівняння результатів, які видали розглянуті моделі, було зроблено висновки, що моделі типу SEIR можуть бути застосовані для створення моделюючого інструментарію регіонального та національного рівнів.

У роботі [22] від серпня 2020 року науковець Ігор Нестерук розглядає нові хвилі пандемії COVID-19 в Україні та його динаміку. Хвилі епідемії обумовлюються багатьма факторами, такими як зміни умов карантину, активність збудника, дотримання карантинних обмежень. Для прогнозування хвиль епідемії у даній роботі була використана класична модель SIR. Було

виділено чотири хвили та проведені розрахунки для п'ятої то шостої хвили. Враховуючи період прогнозування, результати прогнозу кінцевої кількості хворих погіршилися і це є цілком логічним, за рахунок початку періоду відпусток та недотримання маскового режиму.

1.3.4 Статистичні методи прогнозування ЧР

У роботі [23] від вересня 2020 року був проаналізований ріст пандемії COVID-19 в Індії з 4 березня по 11 червня, використовуючи регресійний аналіз, модель з авторегресією і рухомим середнім, експоненційне згладжування та модель Гольта-Вінтерса. В ході дослідження були виявлені проміжки часу, де розповсюдження зростало у геометричній регресії, у квадратичній та навіть лінійно. Після аналізу кожного з методів прогнозування, було зроблено висновок, що на найбільшому проміжку часу ARIMA дає кращі результати з точністю 97,38%. Далі йде модель Гольта-Вінтерса, яка дала точність в 97,11%. У випадку з прогнозуванням кількості випадків у Індії немає сенсу порівнювати сценарії з іншими країнами, так як відрізняються демографічні показники та особливості захворювання. Далі автори сегментували ряди і будували прогнози на менші часові проміжки, тому було вирішено брати за основу ARIMA, яка надала фінально точність у 99,86%.

У наступній роботі [24] розглядається застосування логістичної моделі для прогнозування.

Підсумовуючи результати проведених досліджень, логістична модель належним чином описує зростання кількості випадків COVID-19 з часом. Це було проілюстровано у статті на графіках, де показано фактичні дані та побудовані відповідні логістичні криві. Очевидно, що для багатьох країн використання простого логістичного рівняння може бути допустим тільки при високій точності наявних даних. Використання узагальненої логістичної кривої значно покращує результати, включаючи країни, для яких логістична модель занадто груба

(наприклад, Південна Корея). Було зазначено, що прогнозне число загальної кількості випадків в узагальненій логістичній моделі трохи вище, ніж у простій логістичній моделі, і підхід до граничного постійного значення затримується у часі.

Було також проілюстровано можливості логістичної моделі для опису кількості хворих на день. В середньому теоретична модель досить добре описувала реальні дані, але розкид точок все ще не малий, і іноді відхилення досягали 50% і вище, хоча в середньому вони були менше 50%. Ці відмінності особливо очевидні біля піку епідемії, коли бажано мати більш точний прогноз для медичних установ. Наскільки цей розкид краще описаний іншими моделями (наприклад, моделями SIR), буде зрозуміло найближчим часом, коли з'являться результати відповідних досліджень. З математичної точки зору, отримана різниця у використанні логістичної моделі для опису двох характеристик: загальна кількість випадків (N) та кількість виявлених за день (K) очевидна: крива $N(t)$ дорівнює інтегралу по відношенню до $K(t)$, і, отже, він більш плавний і визначений. Для опису залежності $K(t)$, принаймні на якісному рівні, краще використовувати стохастичні рівняння логістичної моделі рівняння, де вводяться зовнішні випадкові сили або випадкові коефіцієнти. Вони допомагають зрозуміти ступінь розповсюдження даних і, що найголовніше, кількість можливих відхилень під час епідемії. Таку роботу ще потрібно зробити у майбутньому.

Автори розуміють, що для реального прогнозу розвитку епідемії необхідно мати багатофакторні моделі, які включають розподіл населення на різні групи (діти, люди похилого віку тощо), умови життя (транспортні потоки між територіями, щільність населення тощо). Такі моделі повинні включати звичайні диференціальні рівняння та диференціальні рівняння з частинними похідними високого порядку з урахуванням відстаючих аргументів та інтегральних термінів. Зараз розробляються такі складні моделі. Тим не менше, аналіз в рамках простих моделей з невеликої кількістю параметрів є важливим, оскільки дозволяє якісно

описати процес, зрозуміти роль певних факторів та виявити певні явища (стохастизація, фрактальність, нелінійність), які також є цікавими для інших галузей фізики і техніки. У цьому сенсі отримані вище результати демонструють можливості добре розробленої логістичної моделі для опису такої великої епідемії, як COVID-19.

На закінчення автори зазначають, що реальна мінливість даних щоденних тестів на віруси пов'язана з неконтрольованими умовами адміністративних дій (наприклад, змінною кількістю обстежуваних щодня) та соціальною поведінкою людей (міграція, незахищені контакти між їх тощо). Це цікава задача дослідити внесок кожного фактора, який не входить до простої математичної моделі. Найкращим наближенням для таких неконтрольованих факторів є наближення випадковими функціями. Ця випадковість не пов'язана із властивостями фізичних моделей, їх стохастичністю. Звичайно, ця мінливість не може пояснити особливості механізмів епідемічної динаміки, і основною перевагою даного аналізу є визначення можливих рівнів неконтрольованих факторів, коли математичні моделі використовуються для аналізу поширення епідемії.

Згідно зі статистичними даними, підтверджена кількість захворювань на COVID-19 зростає в геометричній прогресії у більшості країн світу. В Італії та Іспанії вже були періоди, коли чергова хвиля пандемії призвела до великого навантаження на систему охорони здоров'я. Таким чином, прогнозування поширення COVID-19 відіграє ключову роль. По-перше, важливо інформувати уряди та медичних працівників, щоб було розуміння чого очікувати і які заходи перш за все вводити, а по-друге, мотивувати населення дотримуватись заходів, які вводяться для уповільнення розповсюдження, щоб не призвести до песимістичних сценаріїв.

У роботі [25] проведено дослідження у ході якого були зроблені прогнози для Пакистану з використанням ARIMA. Було зроблено три прогнози з такими параметрами (0,2,1) (2,2,0) та (1,2,1) для підтверджених випадків, одужалих та

померлих відповідно. Дійшли до висновку, що для прогнозування є важливим мати точні дані від міністерств охорони здоров'я, а також обрати метод прогнозування. Також було згадано, що ми все ще маємо не достатню кількість історичних даних, тому це завжди нам дає певну невизначеність, неможливо виявити певні закономірності у даних.

1.3.5 Методи машинного навчання

У статті [26] були використані методи машинного навчання для створення прогнозів на 10 днів вперед по кількості заражених, кількості одужалих та кількості померлих. За основу були взяті дані від університету Хопкінса. Зупинилися на методах навчання з учителем, так як у нас є історичні дані з кількістю заражених, одужалих та померлих. Застосували такі методи:

- лінійна регресія;
- метод регресії «лассо» (Least Absolute Shrinkage and Selection Operator) [27];
- метод опорних векторів (SVM) [28];
- експоненційне згладжування.

У ході дослідження було визначено, що лінійна регресія в даному випадку дала кращі результати за інші методи. Автори зазначають, що покращення їх прогнозів дозволить урядам вчасно діяти та приймати рішення для запобігання кризових ситуацій, які пов'язані з COVID-19.

У наступній роботі [29] дані отримуються з Міністерства охорони здоров'я Ізраїлю. Набір даних містить щоденні початкові записи про всіх мешканців, які проходили тестування на COVID-19 по всій країні. Дослідниками була розроблена модель, яка передбачає результати тестування, використовуючи вісім бінарних ознак: стать, вік (від 60 років), чи був контакт із зараженою особою та наявність п'яти клінічних симптомів. Початково був 51 831 запис про перевірених осіб. Модель була побудована з використанням дерев рішень з

градієнтним бустингом. Цей підхід виявився ефективним та показав хорошу точність, проте для цього треба мати дані про тестування, симптоми людей, а не тільки один часовий ряд.

Висновки до розділу

Математичне моделювання є ефективним інструментом виявлення динаміки процесів поширення інфекцій, що дозволяє отримувати прогнози поширення COVID-19. В існуючих наукових роботах розглядаються класичні методи моделювання появи та поширення мікроорганізмів, такі як SIR та його модифікації, також імітаційне моделювання грає важливу роль. Не на останньому місці знаходяться методи машинного навчання, а особливо статистичні методи. Важливі питання все ще залишаються невизначеними, і точних відповідей для прогнозування динаміки поширення епідемії складно отримати.

У багатьох роботах дані для дослідження були представлені не лише у вигляді часового ряду. У різних країнах це відбувається по-своєму, вони надають державні ресурси, де можна знайти навіть набори даних з результатами тестувань та симптомами.

Критично важливою є якість доступних і достовірних статистичних даних. Так як ми не маємо у вільному доступі усіх даних, які могли б бути корисними, то можемо оперувати лише даними по кількості виявлених випадів у день та даним по госпіталізації.

Враховуючи результати вищенаведених досліджень, що були проведені, можна зробити висновок, що використання єдиної методики прогнозування не є вірним рішенням. Прогнози, отримані з використанням певного методу прогнозування, потребують експертних корекцій для досягнення результатів, які дійсно можуть відобразити поточну ситуацію та бути релевантними для проведених досліджень.

2 МОДЕЛІ ТА МЕТОДИ

2.1 Поширені математичні моделі

Поява та розповсюдження інфекційних захворювань представляє собою такий собі механізм факторів, які взаємодіють між собою. Такими факторами виступають навколишнє середовище, де знаходяться хвороботворні організми та мікроорганізми, а також разом з цим ще й зовнішня та внутрішня динаміка населення. У зв'язку з цим математичні методи прогнозування відіграють ключову роль для розуміння перебігу епідемії та для розробки стратегій, які дозволяють стримувати інфекційні захворювання, що швидко поширюються, за відсутності будь-яких специфічних противірусних препаратів або ефективної вакцини [41].

У 1927-1933 роках вчені Кермак та Маккендрік запропонували базову математичну модель у епідеміологічній області, яка отримала назву SIR і з допомогою якої стало можливим моделювання появи та поширення мікроорганізмів. Схематичне представлення зображено на рисунку 2.1. Суть полягає у поділенні населення на групи:

- особи сприйнятливі до захворювання (позначаються, як S - susceptible);
- особи інфіковані (позначаються, як I – infected);
- особи, які одужали (позначається, як R – recovered).

Дослідження цих моделей відбувається з використанням диференціальних рівнянь.

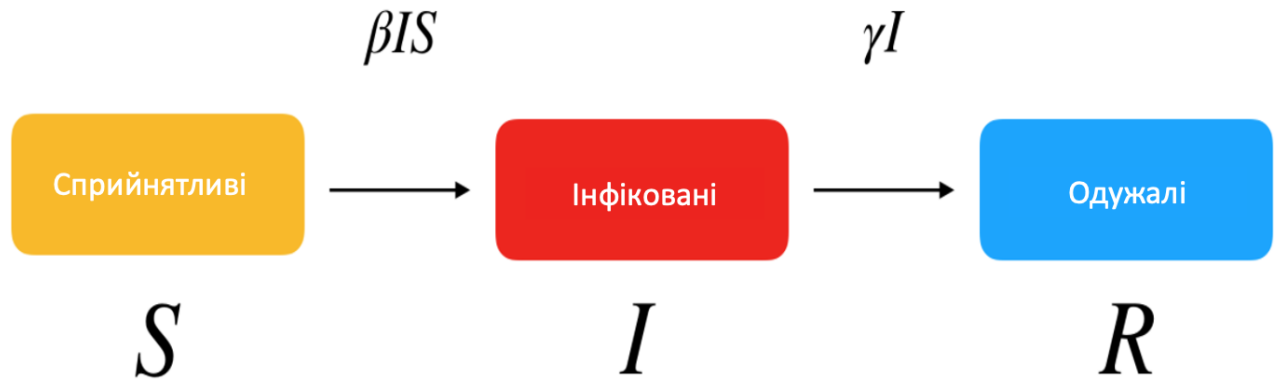


Рисунок 2.1 – Модель SIR схематично

Модель SIR представляється у вигляді диференціальних рівнянь:

$$\frac{dS}{dt} = -\beta IS \quad (2.1)$$

$$\frac{dI}{dt} = \beta IS - \gamma I \quad (2.2)$$

$$\frac{dR}{dt} = \gamma I \quad (2.3)$$

де β – параметр, який показує середню кількість людей, яких одна інфікована людина може заразити за одиницю часу; γ – параметр, який показує швидкість відновлення за певний період часу, D – середній період часу, протягом якого інфікована людина залишається заразною, зворотно-пропорційний параметр до γ ; R – середня кількість інфікованих людей від уже інфікованої людини у період зараження хворобою.

Нині існує сімейство моделей поширення інфекцій, які розроблені на основі моделі SIR:

- модель для опису динаміки захворювань з тимчасовим імунітетом (SIRS) [30];
- модель для опису поширення захворювань з інкубаційним періодом (SEIR) [31];

- модель для поширення захворювання, до якого не виробляється імунітет (SIS);
- модель, що враховує імунітет дітей, набутий внутрішньоутробно (MSEIR).

Для прогнозування COVID-19 частіше всього використовують модифікації моделі SEIR. Схематично вона подана на рисунку 2.2.

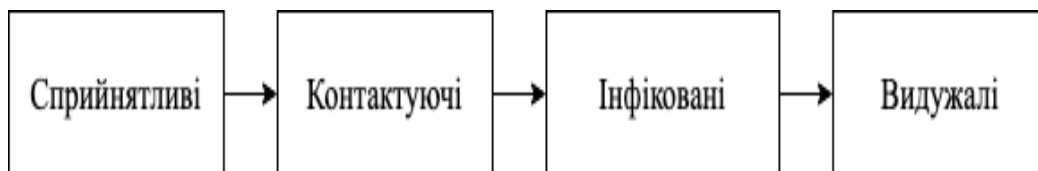


Рисунок 2.2 – Схематичне представлення моделі SEIR

До моделей, які розглядаються для прогнозування поширення коронавірусної інфекції, відноситься модель Logistic Growth [32]. Диференціальне рівняння для цієї моделі подається так:

$$\frac{dN}{dt} = rN\left(1 - \frac{N}{N_{\infty}}\right) \quad (2.4)$$

де $N(t)$ – загальна кількість людей, які постраждали від епідемії; N_{∞} – максимальні кількість заражених людей протягом всієї епідемії; r – темп росту епідемії.

У 1960 році Вінтерс розширив метод Гольта шляхом врахування сезонних коливань. Позначимо період сезонності як m .

Для опису мультиплікативної моделі Вінтерса визначають чотири рівняння:

1. Експоненційно згладжені ряди

$$L_t = \alpha \frac{y_t}{S_{t-m}} + (1 - \alpha)(L_{t-1} + T_{t-1}) \quad (2.5)$$

2. Оцінка тренду

$$T_t = \beta(L_t - L_{t-1}) + (1 - \beta)T_{t-1} \quad (2.6)$$

3. Оцінка сезонності

$$S_t = \gamma \frac{y_t}{L_t} + (1 - \gamma) S_{t-m} \quad (2.7)$$

4. Прогноз на τ періодів

$$\hat{y}_{n+\tau} = (L_n + \tau T_n) S_{n+\tau-m} \quad (2.8)$$

де

L_t – нове згладжене значення або оцінка поточного рівня;

α – постійна згладжування для цього рівня;

y_t – нове спостереження або реальне значення величини за період);

β – постійна згладжування для оцінки тренду;

T_t – оцінка тренду;

γ – постійна згладжування для оцінки сезонності;

S_t – оцінка сезонності;

τ – кількість періодів, на які будується прогноз (період упередження);

n – число наявних спостережень;

$\hat{y}_{n+\tau}$ – прогноз на τ періодів.

Рівняння (2.5) використовується для коригування згладжених рядів. За допомогою поділу y_t на S_{t-m} елімінуються сезонні ефекти, якщо такі наявні.

Вже після того, як враховані сезонність та тренд, оцінка згладжуються у рівняннях (2.6) та (2.7), а рівняння (2.8) стосується прогнозування.

Важливо на початку обрати параметри α , β , γ , а також визначити початкові значення для згладжених рядів L_t , тренду T_t та коефіцієнту, який відповідає за сезонність S_t .

Тож експоненційне згладжування є популярною методикою, яка використовується для цілей короткострокового прогнозування.

Для прогнозування також використовуються такі статистичні моделі:

– авторегресійні моделі порядку p : AR(p);

- моделі із рухомим середнім порядку q : $MA(q)$;
- комбінація моделей $AR(p)$ та $MA(q)$ — $ARMA(p,q)$, де p – порядок авторегресійної частини моделі, а q – частини РС.

Загальний вигляд такої комбінованої моделі:

$$y_t = \phi_0 + \phi_1 y_{t-1} + \dots + \phi_p y_{t-p} + \varepsilon_t - \omega_1 \varepsilon_{t-1} - \dots - q \varepsilon_{t-q}, \quad (2.9)$$

Де y_t відгук (залежна змінна) в момент часу t ; $\phi_0, \phi_1, \dots, \phi_p$ – коефіцієнти, які оцінюються $AR(p)$; $\varepsilon_t, \dots, \varepsilon_{t-q}$ – помилки у відповідні періоди; $\omega_1, \omega_2, \dots, \omega_q$ – коефіцієнти, які оцінюються $MA(q)$;

За допомогою моделі $ARMA(p,q)$ описується широкий спектр поведінки стаціонарних часових рядів. Якщо в (2.9) $q=0$, то маємо звичайну авторегресійну модель порядку p ; якщо ж $p=0$, то маємо стандартну модель РС.

Якщо ми маємо ряд, який не є стаціонарним, то в моделі $ARIMA$ здійснюється взяття різниць, щоб отримати стаціонарний процес: за допомогою першої різниці усувається лінійний тренд, за допомогою другої — квадратичний.

У загальному випадку, щоб отримати стаціонарний часовий ряд, можна знаходити різниці порядку d . k -та різниця подається у вигляді:

$$\Delta^k y_t = \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \dots + \alpha_p y_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q} \quad (2.10)$$

З моделі для $\Delta^k y_t$ можливо отримати модель для початкового ряду. Наприклад, для різниці першого порядку Δy :

$$y_t = y_{t-1} + \Delta y_t \quad (2.11)$$

Друга різниця представиться у вигляді:

$$\Delta^2 y_t = \Delta(\Delta y_t) = \Delta(y_t - y_{t-1}) = y_t - 2y_{t-1} + y_{t-2} \quad (2.12)$$

Також для прогнозування використовують моделі машинного навчання. Яскравим претендентом серед цих моделей є XGBoost. XGBoost (Extreme Gradient Boosting) відноситься до методів навчання з учителем, де подаємо дані для тренування на вхід, щоб визначити цільову змінну. Це такий метод є ансамблевим алгоритмом градієнтного бустингу, який використовує дерева рішень в основі. При цьому методі дерева будуються по одному, де кожне нове дерево допомагає виправляти помилки, допущені раніше побудованим деревом.

Основна формула для розрахунку прогнозу виглядає так:

$$\hat{y}_i = \sum_{k=1}^K f_k(x_i), f_k \in F, \quad (2.13)$$

де y_i – це прогнозоване значення; x_i – вектор ознак; $f_k(x_i)$ – значення, розраховане для кожного дерева; K – загальна кількість дерев.

Модель XGBoost є адитивною по відношенню до кожного дерева. Функція, яка відповідає за те, як дерева оцінюються та обчислюються:

$$f_t(x) = \omega_{q(x)}, \omega \in R^T, q: R^d \rightarrow \{1, 2, \dots, T\}, \quad (2.14)$$

де $q(x)$ – це функція, яка приписує ознаки x конкретному листу поточного дерева t ;

$\omega_q(x)$ – це тоді оцінка листків для поточного дерева t та поточних ознак x .

Підводячи підсумок, як тільки проходить навчання моделі, що є найскладнішою частиною проблеми, прогнозування просто зводиться до визначення правильного листа для кожного дерева на основі особливостей та підсумовування значень, доданих до кожного листа.

Проблеми із формальним застосуванням найкращих із відомих методів прогнозування ілюструє рисунок 2.3, де подані фактичні дані МОЗ України та результати прогнозування кількості ушпиталених на початок травня методом

ARIMA з вибором найкращих значень його параметрів на основі оцінки точності на навчальній вибірці.

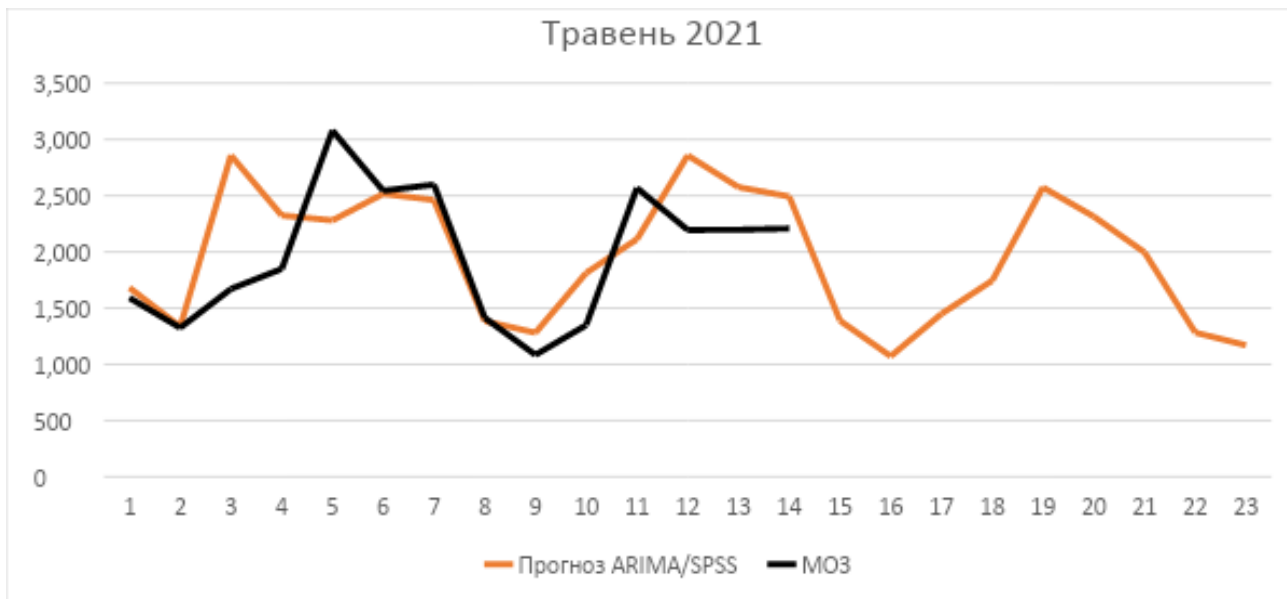


Рисунок 2.3 – Порівняння результатів ARIMA з фактичними даними

2.2 Математична постановка задачі

Загалом ми маємо задачу прогнозування ЧР у такій постановці.

Маємо дискретні моменти часу $t = 1, 2, \dots, T$ та часовий ряд $y_1, y_2, \dots, y_T$, де кожний член ЧР $\{y_t\}, t \in [1, T]$ може бути поданий так:

$$y_t = u_t + v_t + w_t + z_t + \theta_t + \varepsilon_t, \quad (2.14)$$

де u_t – довготривала складова, тренд; v_t – циклічна компонента; w_t – сезонна складова, що формується під впливом сезонних коливань показника протягом заданого періоду часу, зазвичай року (часто включається у циклічну компоненту); z_t – компонента, яка забезпечує порівнюваність елементів ЧР; θ_t – управляюча змінна, за допомогою якої впливають на члени ЧР з метою формування в майбутньому його бажаної траєкторії (це – керований прогноз, який є основою для сценарного підходу).

Потрібно створити таку модель прогнозування, для якої середньоквадратичне відхилення фактичного значення від прогнозованого

прямує до мінімального для заданого P . Тобто потрібно визначити прогнозні значення $y_{T+1}, \dots, y_{T+P}$.

За критерій точності можна взяти показник середньоквадратичного відхилення:

$$\varepsilon = \sqrt{\frac{\sum_{i=1}^n (y - \hat{y}_i)^2}{n}}, \quad (2.15)$$

По суті, (2.14) і є цільовою функцією, яку потрібно мінімізувати.

2.3 Алгоритми розв'язування задачі

Схема алгоритму:

1. Формування ряду рухомих сум за період довжиною в тиждень:

$$y_1, y_2, \dots, y_T, \quad t \in [1, T];$$

2. Прогноз за допомогою ARIMA на n значень наперед

– вибір параметрів для ARIMA, при яких маємо мінімальний показник RMSE;

– розрахунок прогнозу з підібраними параметрами;

3. Коригування значень прогнозу:

– розрахунок кореляції між останніми n значеннями нативного ряду $y_{T-n+1}, \dots, y_T$ та значеннями нативного ряду у минулому $y_i, \dots, y_{i+n-1}$, де $i = \overline{1, T-2n}$.

Нехай $x = y_{T-n+1}, \dots, y_T$, $y = y_i, \dots, y_{i+n-1}$, тоді для розрахунку коефіцієнта кореляції маємо формулу:

$$r = \frac{\sum (x - m_x)(y - m_y)}{\sqrt{\sum (x - m_x)^2 \sum (y - m_y)^2}}, \quad (2.16)$$

де m_x та m_y – середні значення векторів x та y відповідно.

– для частини ЧР з найбільшим показником кореляції розраховуємо частки кожного дня за період з n днів:

$$d_j = \frac{y_i}{\sum_{j=1}^n y_i}, \quad (2.17)$$

де $j = \overline{1, n}$.

– відповідно до днів тижня та розрахованих часток коригуємо прогнозні значення і отримуємо прогноз

$$y_{T+1}d_1, \dots, y_{T+n}d_n. \quad (2.18)$$

2.4 Результати експериментальних досліджень

Було виділено декілька етапів дослідження:

- прогнозування нативного ряду;
- прогнозування на основі тижневих рухомих сум (TPC) з накладанням профілю тижня. Для кількох тижнів без свят і аномальних викидів розраховуються усереднені частки кожного з днів тижня, і тоді прогнозна величина розподіляється по дням тижня згідно часток;
- прогнозування різницевого ряду.

Для порівняння були використані методи Гольта-Вінтерса та ARIMA, а також метод машинного навчання XGBoost. Було оцінено апріорну та апостеріорну точність даних методів на різних видах часових рядів. Апріорна точність розрахована за допомогою середньоквадратичної помилки (RMSE) і наведена у таблиці 2.1.

Таблиця 2.1 – Априорна точність розрахованих прогнозів

RMSE	Нативний ряд	Згладжений ряд (свята, вихідні)	TPC	Ряд різниць
ARIMA	1 120,89	1 294,71	6 408,29	804,83
Гольта-Вінтерса	970,29	1 021,58	11 093,09	744,28
XGBoost	1 053,79	1 054,31	4 984,65	789,39

На рисунку 2.4 показані всі прогнози отримані за допомогою порівнюваних методів.

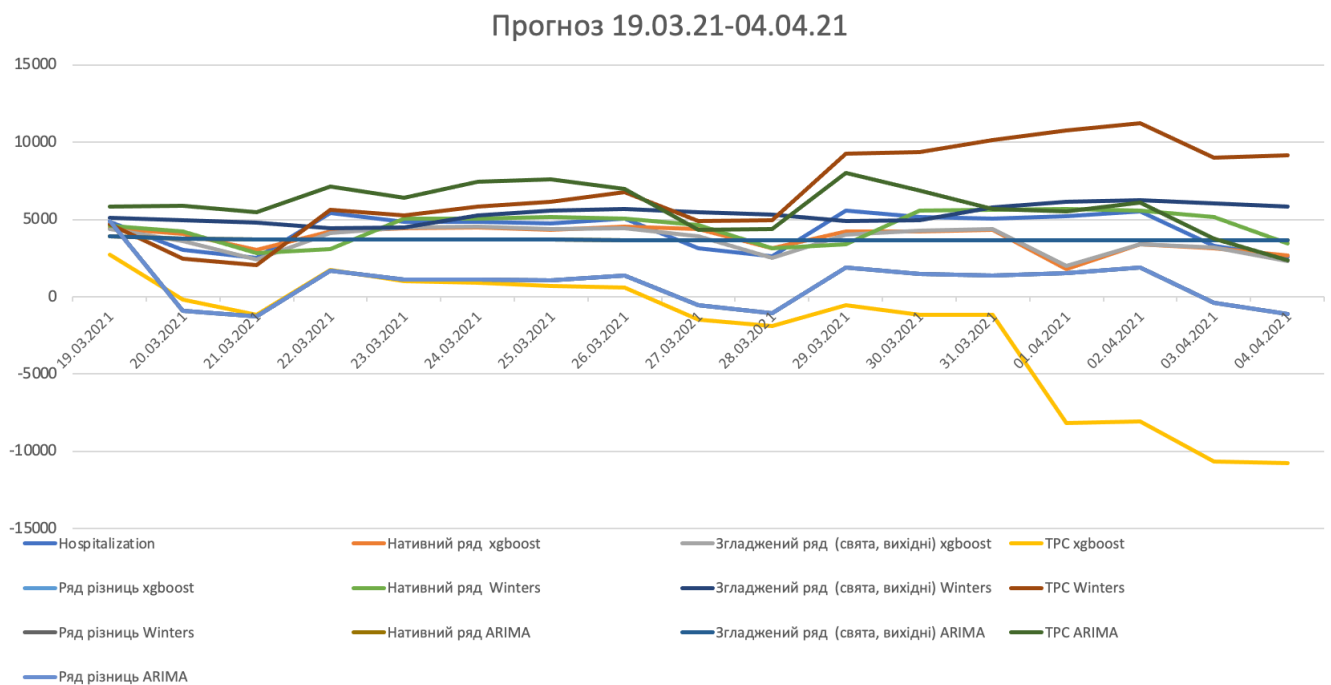


Рисунок 2.4 – Графік з розрахованими прогнозними значеннями та фактичними даними

З такого порівняння складно виділити методи, тож відберемо кілька прогнозів, які мають найменші значення середньоквадратичної помилки за априорною оцінкою.

Після того, як були відібрані методи з найменшими значеннями середньоквадратичних помилок, маємо наступну картину на рисунку 2.5.

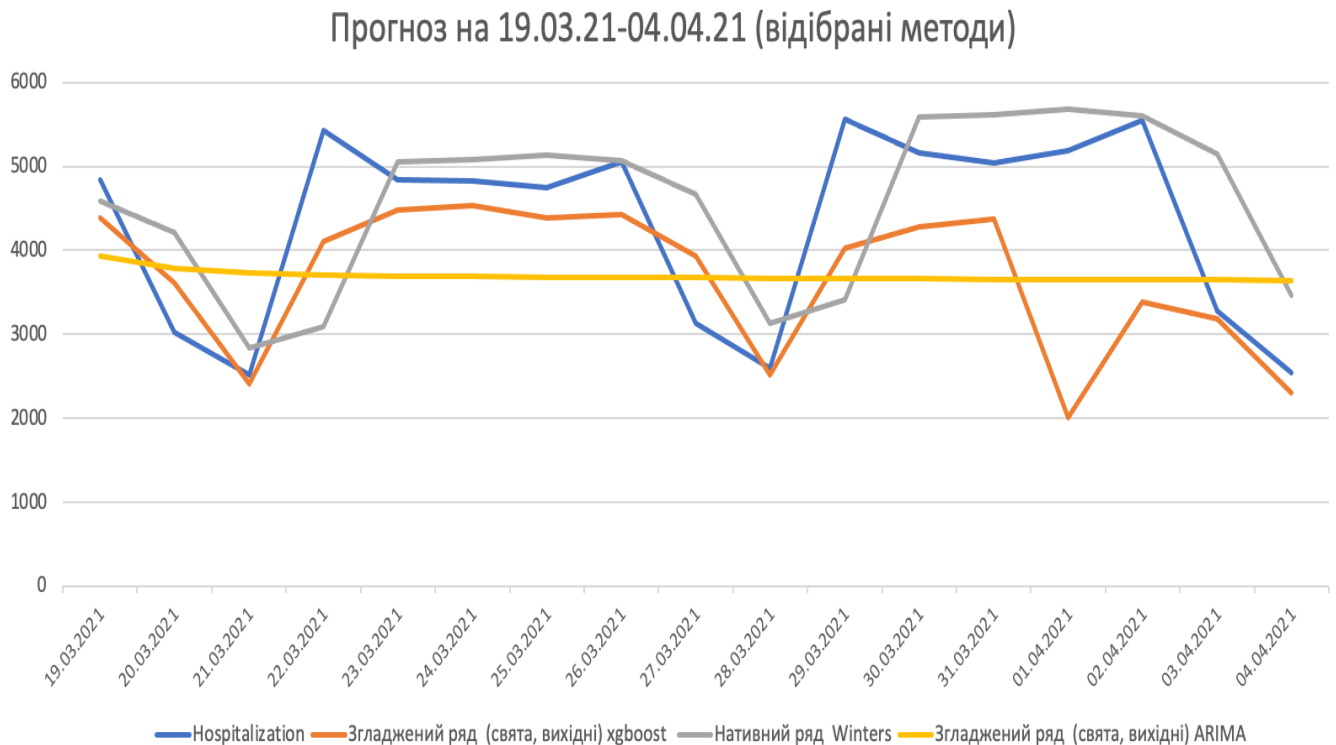


Рисунок 2.5 – Графік з розрахованими прогностичними значеннями кращими методами та фактичними даними

У таблиці 2.2 можемо бачити значення середньоквадратичної помилки за апостеріорною оцінкою.

Таблиця 2.2 – Апостеріорна точність відібраних методів

RMSE	Нативний ряд	Згладжений ряд (свята, вихідні)	TPC	Ряд різниць
ARIMA		1075,93		
Вінтерс				
XGBoost	1214,14	1138,11		

Було відібрано кращі методи для порівняння з розробленим алгоритмом прогнозування.

Отримали результати, які показані на рисунку 2.6.

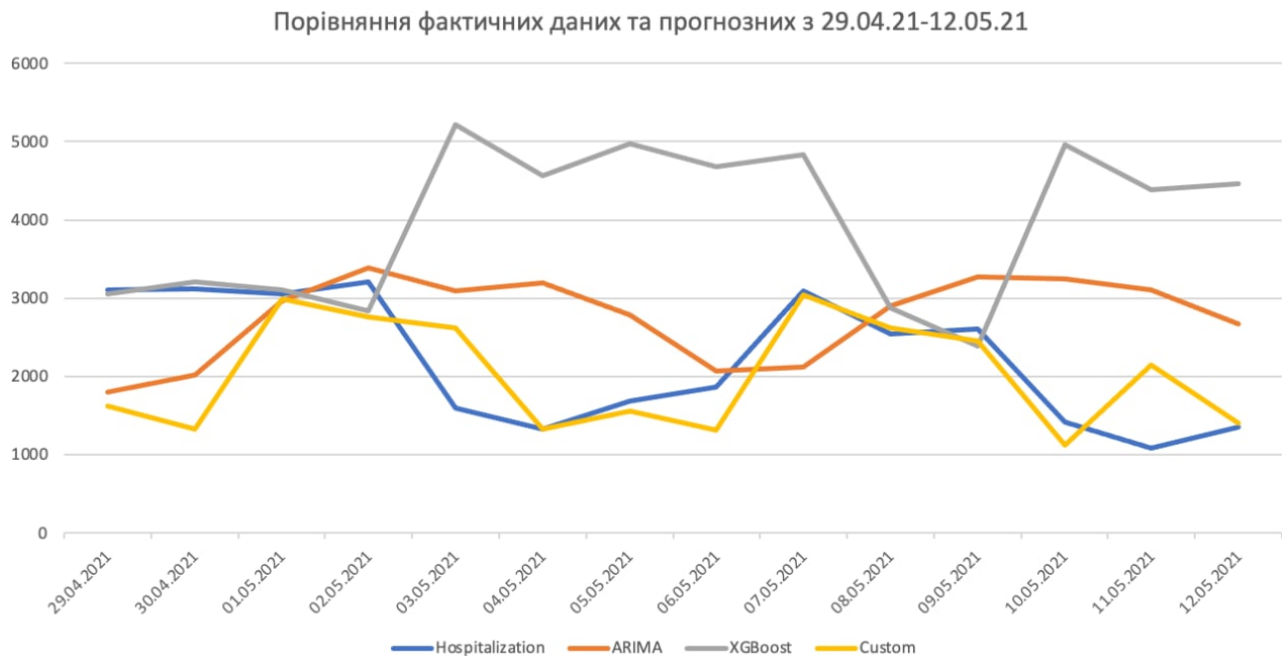


Рисунок 2.6 – Графік з розрахованими прогнозними значеннями кращими методами, фактичними даними та результатами роботи розробленого алгоритму

Для оцінки точності використаємо показник, який покаже значення у відсотках – MAPE (Mean absolute percentage error). Результати порівняння помилок показані у таблиці 2.3.

Таблиця 2.3 – Апріорна точність відібраних методів та розробленого алгоритму

	ARIMA	XGBoost	Custom
MAPE	63,0096852	55,8778868	44,1986523

Як бачимо з результатів, розроблений алгоритм показав кращі результати за відібрані методи, та його точність складає близько 56%.

Висновки до розділу

Методи математичного моделювання є ефективними інструментами виявлення динаміки процесів поширення. З їх допомогою отримуємо прогнози поширення COVID-19. В існуючих наукових роботах розглядаються класичні методи моделювання появи та поширення мікроорганізмів, такі як SIR та його

модифікації. Важливі питання все ще залишаються невизначеними, і точних відповідей для прогнозування динаміки поширення епідемії складно отримати. Це залежить від якості доступних статистичних даних, а саме від точної кількості заражених осіб, тож це може призводити до неоднозначних результатів та невідповідних прогнозів. Було розроблено алгоритм, який на основі прогнозу за допомогою статистичної моделі ARIMA здійснює подальше коригування. Для кожного прогнозу були взяті значення ряду за останній період та розрахована кореляція з такими ж інтервалами у минулому, проходжуючись по усьому часовому ряду. Це було пророблено для того, щоб знайшовши схожий інтервал можливо було зробити припущення щодо подальшого розвитку подій. Було розраховано частку кожного дня за цей тиждень, та на цю частку був скоригований прогноз який дає модель з вже підібраними параметрами. Були отримані досить гарні результати у порівняння з існуючими методами прогнозування.

3 ОПИС ПРОГРАМНОГО ТА ТЕХНІЧНОГО ЗАБЕЗПЕЧЕННЯ

3.1 Опис програмного забезпечення

У ході розробки програмного продукту, як частини магістерської дисертації, було обрано складну задачу, яка стосується прогнозування в реальному часі. Прогнозування має здійснюватися у реальному часі із можливостями вибору горизонту прогнозування та методу прогнозування. Також результати можна зберегти локально для подальшого аналізу. Для того, щоб це було можливо, потрібно отримувати і дані у реальному часі, наприклад, поденну кількість госпіталізованих або кількість ліжко-місць. Це ставить ще одну важливу задачу, яка стосується збору та збереження даних. Далі нижче наведено перелік вимог до програмного забезпечення, а також опис обраних засобів розробки та технологій. У додатку А (плакат 2) зображено UML-діаграму варіантів використання.

3.1.1 Вимоги до програмного забезпечення

У таблиці 3.1 наведемо список функціональних вимог до ПЗ.

Таблиця 3.1 – Функціональні вимоги до програмного забезпечення

№	Назва	Пріоритет
1	Система надає можливість переглядати щоденно оновлені фактичні дані	Високий
2	Система надає можливість перегляду базового тижневого прогнозу	Високий
3	Система надає можливість задавати параметр діапазону прогнозу: – 7 днів; – 14 днів – 21 день.	Високий

Продовження таблиці 3.1.

№	Назва	Пріоритет
4	Система надає можливість задавати методу прогнозування: – ARIMA; – XGBoost; – Custom.	Високий
5	Система надає можливість збереження обраного прогнозу та фактичних даних для подальшого аналізу	Високий
6	Система надає можливість відображати результати прогнозів графічно	Високий

Кожна з функцій системи організована у вигляді окремих модулів, що наведено у додатку А (плакат 1) функціонально-логічну структуру програмного продукту.

Інформаційна система з прогнозування поширення COVID-19 успішно пройшла випробування і готова до експлуатації. Розроблений програмний продукт допомагає у короткий час отримувати прогноз з урахуванням обраного діапазону та методу прогнозування. Дана функціональність може принести велику користь на державному рівні. Вищі органи управління можуть керуватися отриманою інформацією та продумувати подальші заходи, які можна ввести для покращення ситуації із захворюваністю, спираючись на отримані прогнози у реальному часі, або ж після роботи державних аналітиків над отриманими прогнозами, використовуючи опцію завантаження результатів.

3.1.2 Засоби розробки

Для розробки програмного продукту була обрана клієнт-серверна архітектура з мікросервісами [33], так як у нас є певна кількість ресурсів, для яких потрібний обмежений правами доступ великої кількості користувачів. Усі

серверні частини написані на мові програмування Python. Python є об'єктно-орієнтованою мовою програмування, яка відноситься до високорівневих, має динамічну сувору типізацію [34]. Мікросервіси представляються у якості Azure Functions.

Azure Functions – це serverless рішення, які надає хмарна платформа Azure [35], що дозволяють виконувати код, який може бути ініційований подіями, наприклад, по таймеру, за GET-запитом та без попереднього налаштування середовища [36]. До переваг обраного сервісу відноситься його вартість та можливість налаштувати виконання функції по таймеру. За замовчуванням очікується, що кожна функція буде містити за точку входу глобальний метод `main()` в `__init__.py`.

HTTP-запити використовуються, щоб вказати серверу на те, яку ми дію бажано виконати з ресурсом. Для того, щоб розділити дії з ресурсами на рівні HTTP-запитів, були придумані наступні варіанти запитів [38]:

- GET (отримання ресурсу);
- POST (створення ресурсу);
- PUT (оновлення ресурсу);
- DELETE (видалення ресурсу).

Для розробки веб-застосування був обраний JavaScript-фреймворк AngularJS [37]. За допомогою цього інструменту у поєднанні з бібліотекою Angular Material було вирішено зробити базовий інтерфейс застосування, який має містити графіки з поточними та прогнозованими даними, а також можливість вибору методу прогнозування та горизонту прогнозу.

Перш за все стояла задача збору даних. Були проаналізовані доступні ЧР, які стосуються перебігу захворюваності на COVID-19. Ряд з кількістю підтверджених випадків в день відразу не розглядався, так як ці дані не є показовими і дуже залежать від щоденної кількості тестів, дня тижню та терміну передачі цієї інформації до статистичних відділів. Після пошуку потрібних

ресурсів було знайдено сайт, який відноситься до Апарату РНБО України. На рисунку 3.1 представлено загальний вигляд початкової сторінки.

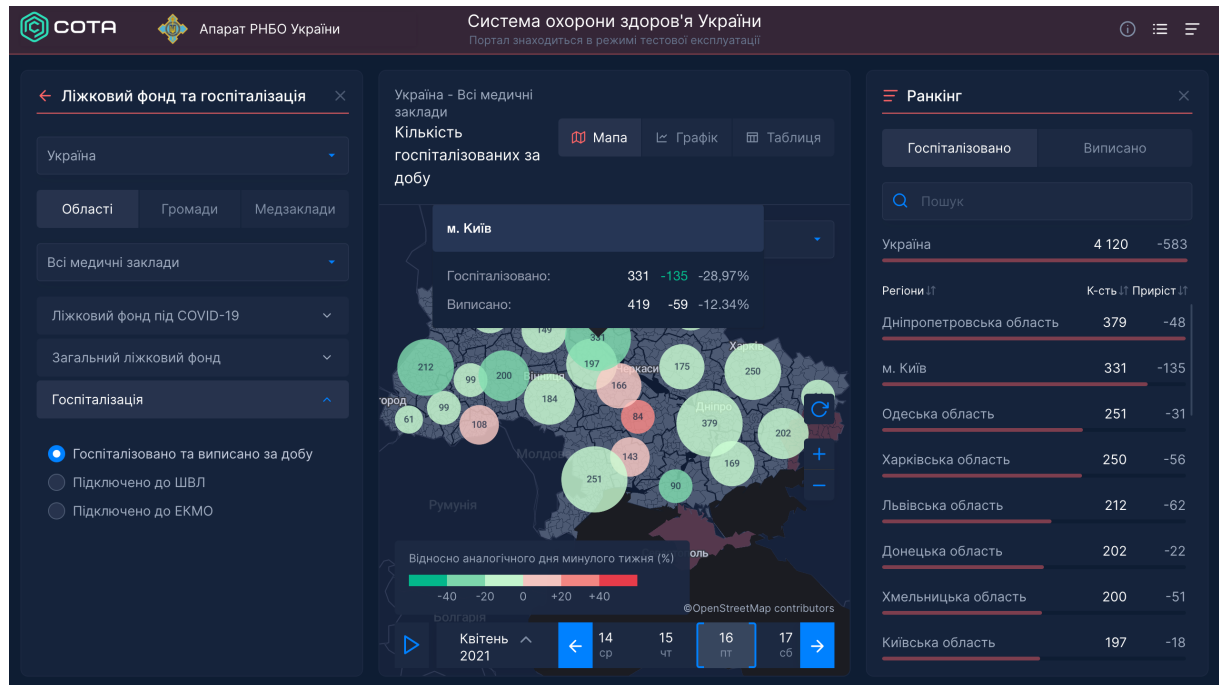


Рисунок 3.1 – Головна сторінка порталу системи охорони здоров'я України

Було прийнято рішення відібрати дані по щоденній госпіталізації та поставало питання як це краще робити та де зберігати.

Доступ до даних з сайту РНБО можливо отримати за допомогою відповідних HTTP-запитів.

Тож для отримання даних виконуються відповідний GET-запити <https://health-security.rnbo.gov.ua/api/beds/hospitalization/ranking> з параметрами.

Параметри GET-запиту:

- bedType;
- regionId;
- otgId;
- aggLevel;
- date;
- hospitalType.

Так як нас цікавлять дані по госпіталізації, то передаємо у якості параметру `bedType` значення «hospitalized». На рисунку 3.2 відображена відповідь серверу на запит даних по госпіталізації, яка отримана через Postman.

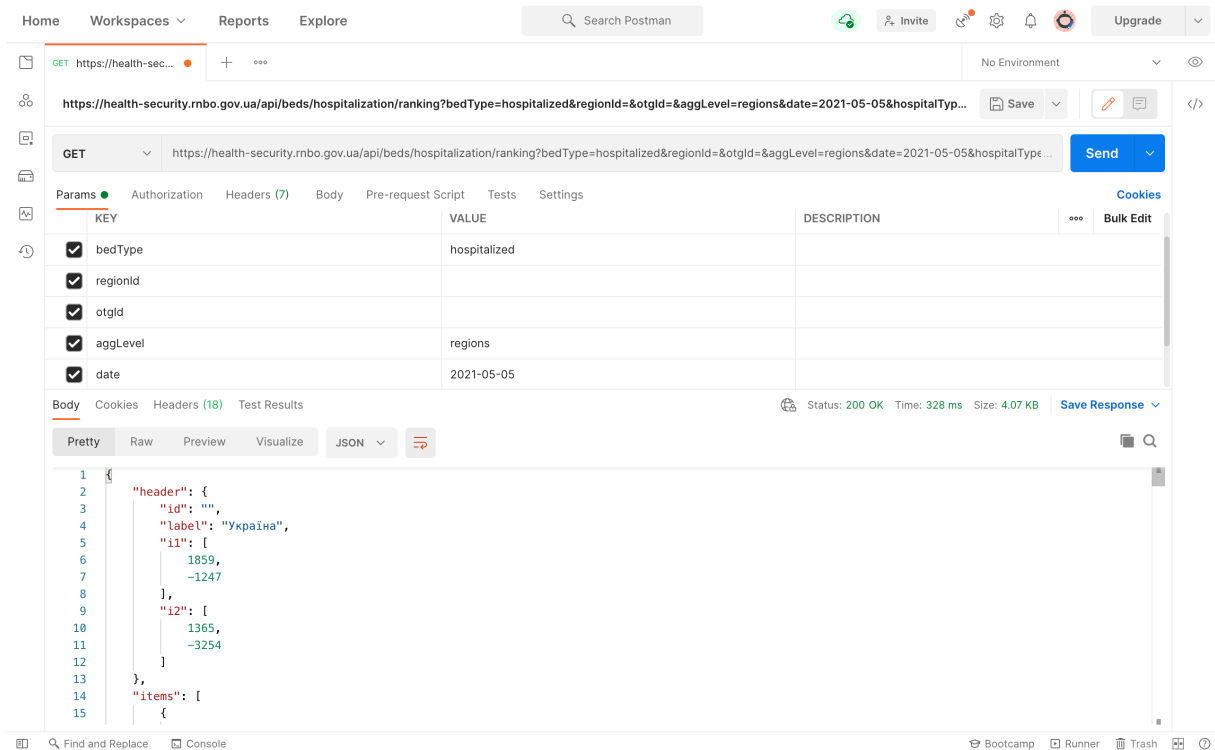


Рисунок 3.2 – Приклад отриманої відповіді від серверу за GET запитом

Вибір був зупинений на хмарній платформі від Microsoft – Azure. Серед ресурсів, які надає хмарна платформа, були обрані Azure Functions для обчислень та збору даних, а також Storage Account для збереження даних по госпіталізації після збору.

Було реалізовано timer trigger функцію, яка за налаштуваннями CRON о 7 ранку збирала доступні дані по госпіталізації [39]. Потім дані зберігаються до файлу, який знаходиться у Storage Account [40].

Повертаючись до веб-застосування, ми маємо сторінку з графічним відображеннями ситуації по госпіталізації. Шляхом вибору значень у випадючих списках, вони далі передаються за допомогою GET запиту до опублікованої Azure Functions у наступному вигляді <https://hospitalizationforecast.azurewebsites.net/api/HttpTrigger> та з параметрами:

- numdays, що відповідає горизонту прогнозу, наприклад, 7, 14 або 21 день;
- method, що відповідає за метод прогнозування, наприклад, ARIMA, XGBoost або Custom.

На рисунку 3.3 зображений приклад відповіді серверу на GET-запит прогнозу методом Custom на 7 днів у форматі JSON.

```

1  {
2    "Date": [
3      "2021-05-12",
4      "2021-05-13",
5      "2021-05-14",
6      "2021-05-15",
7      "2021-05-16",
8      "2021-05-17",
9      "2021-05-18"
10   ],
11   "Hospitalization": [
12     2823.8018277600922,
13     2642.2952825398293,
14     1210.0436348017538,
15     2302.3830251273366,
16     1509.2544244981873,
17     1201.2433174577409,
18     2820.5017087560873
19   ]
20 }
```

Рисунок 3.3 – JSON файл з відповіддю серверу на GET-запит по отриманню прогнозу

Після того, як параметри були передані та була отримана відповідь від серверу, на головній сторінці веб-застосування оновлюється графік у відповідності до обраних параметрів.

На рисунку 3.4 представлена архітектура програмного продукту, який працює у реальному часі.

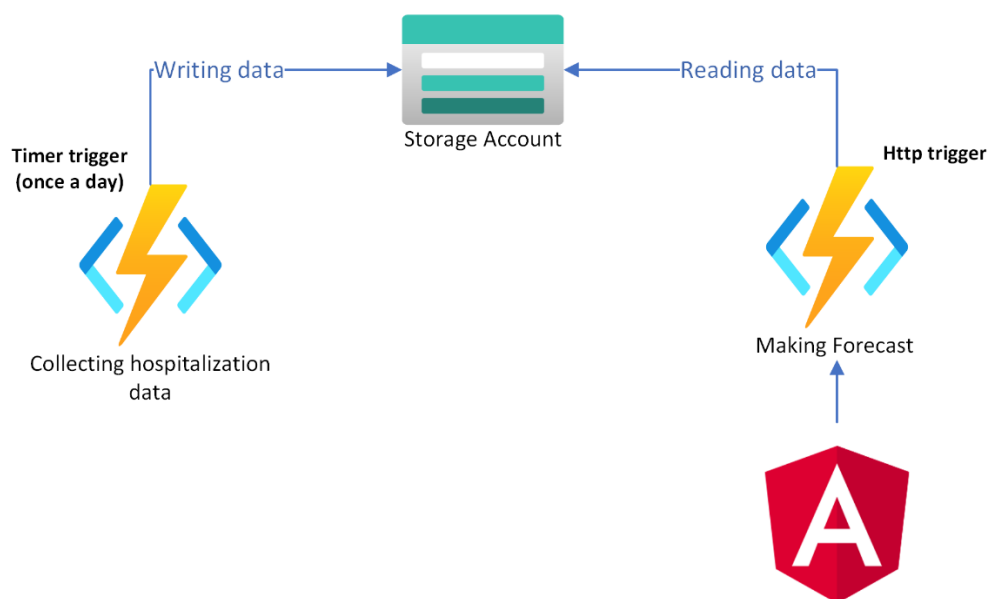


Рисунок 3.4 – Архітектура програмного продукту

3.1.3 Архітектура програмного забезпечення

Програмний продукт розроблений з ціллю прогнозування кількості госпіталізованих на найближчі тижні у реальному часі. В кінцевому результаті користувач отримує прогноз з обраним методом прогнозування та на обраний період.

Архітектура програмного забезпечення може бути представлена у вигляді діаграми класів, яка наведена у додатку А (плакат 4).

У додатку А (плакат 3) представлена діаграма послідовності, для того, щоб показати взаємодію об'єктів у часі. Тобто вона відображає задіяні об'єкти та послідовність їх взаємодії через повідомлення.

У таблиці 3.1 містяться описи класів та їх методів, які реалізовані в рамках розробки програмного продукту.

Таблиця 3.1 – Класи та їх методи

Назва	Примітка
TSForecast – клас, в якому відбувається розрахунок прогнозу обраним методом на певний горизонт	
getForecast	Метод, у якому відбувається прогнозування обраним методом
ARIMA – клас, в якому відбувається підбір параметрів, тренування та формування прогнозу з методом ARIMA	
getBestParams	Метод, у якому відбувається підбір оптимальних параметрів для прогнозування
modelTrain	Метод, у якому відбувається тренування моделі
predict	Метод, який використовує натреновану модель та повертає значення прогнозу на заданий діапазон
XGBoost_forecast – клас, в якому відбувається тренування моделі машинного навчання XGBoost та подальший прогноз	
addFeatures	Метод, у якому трансформуються вхідні дані для виділення додаткових ознак
modelTrain	Метод, у якому відбувається тренування моделі
modelPredict	Метод, у якому відбувається прогноз
Custom – клас, в якому відбувається прогнозування з урахуванням коригувань	
calcCorrelation	Метод, за допомогою якого розраховується кореляція між значеннями останнього наявного тижня та між тижнями у минулому
getARIMAForecast	Метод, у якому створюється об'єкт класу ARIMA для отримання початкового прогнозу

Продовження таблиці 3.1.

Назва	Примітка
getCorrections	Метод, у якому значення отриманого прогнозу коригуються у відповідності до кореляційних значень, часток днів тижня та тренду.

3.1.4 Користувацький інтерфейс

Користувацький інтерфейс програмного продукту повинен бути зручно організованим для користувача, а також відповідати певним стандартам дизайну інтерфейсів.

Інтерфейс містить у собі основну сторінку, яка початково відображає прогноз на 7 днів за допомогою статистичної моделі Бокса-Дженкінса. Екранна форма програмного продукту, на якій зображена початково завантажена сторінка, наведена на рисунку 3.5.

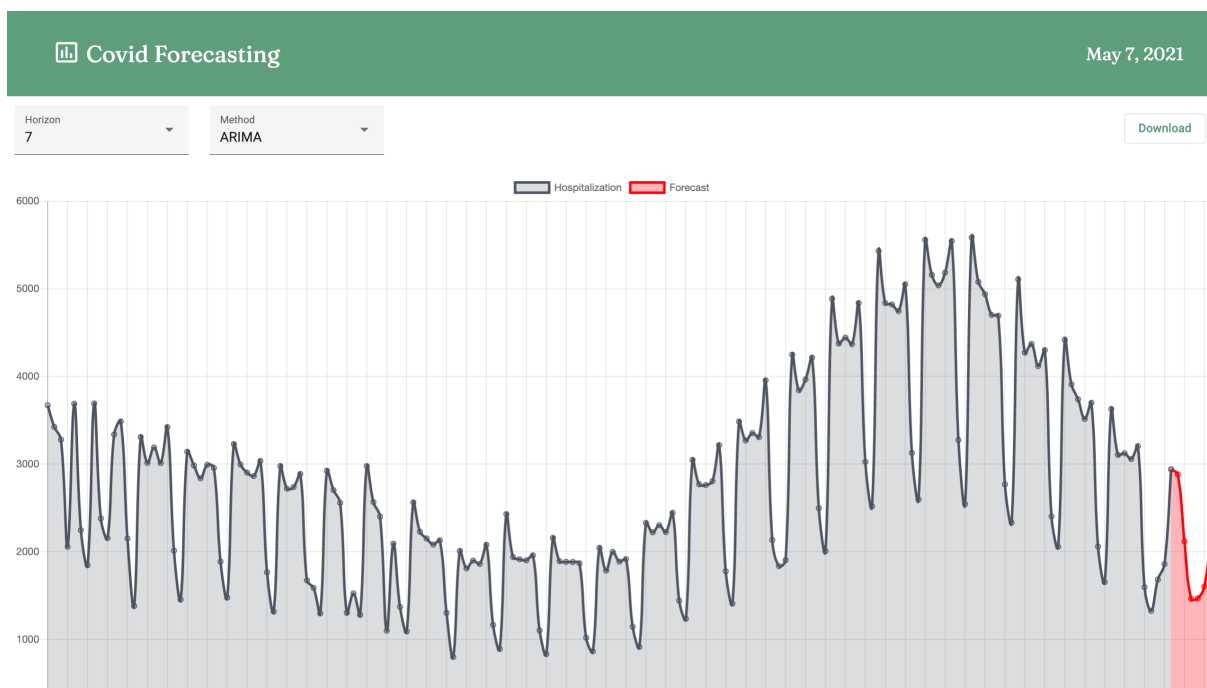


Рисунок 3.5 – Екранна форма початково завантаженої сторінки (прогноз з використанням моделі Бокса-Дженкінса)

На основній сторінці ми бачимо два випадаючих списки, де можна обрати горизонт прогнозу та метод прогнозування.

Варіанти для вибору горизонту прогнозу:

- 7 днів;
- 14 днів;
- 21 день.

Варіанти для вибору методу прогнозування:

- ARIMA (модель Бокса-Дженкінса);
- XGBoost;
- Custom (розроблений алгоритм прогнозування).

Екранні форми програмного продукту для формування прогнозів з різними горизонтами та з використанням реалізованих методів прогнозування наведені на рисунках 3.6–3.8.

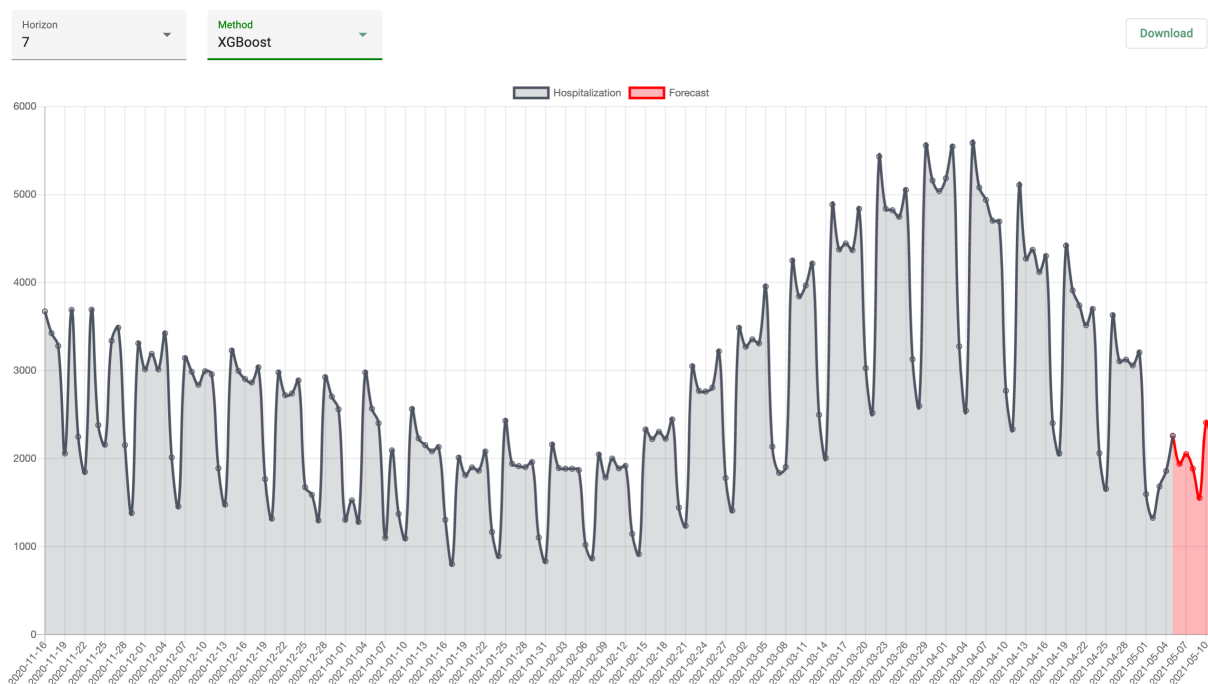


Рисунок 3.6 – Екранна форма головної сторінки з прогнозом на 7 днів за допомогою XGBoost

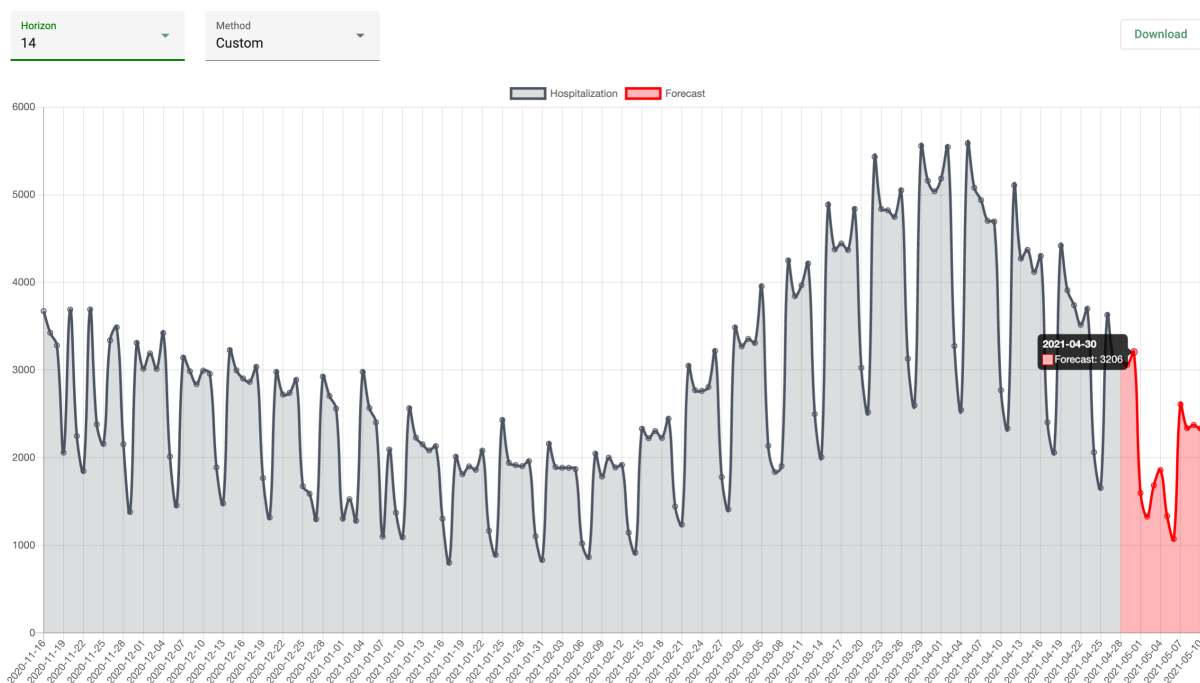


Рисунок 3.7 – Екранна форма головної сторінки з прогнозом на 14 днів за допомогою розробленого алгоритму прогнозування (Custom)

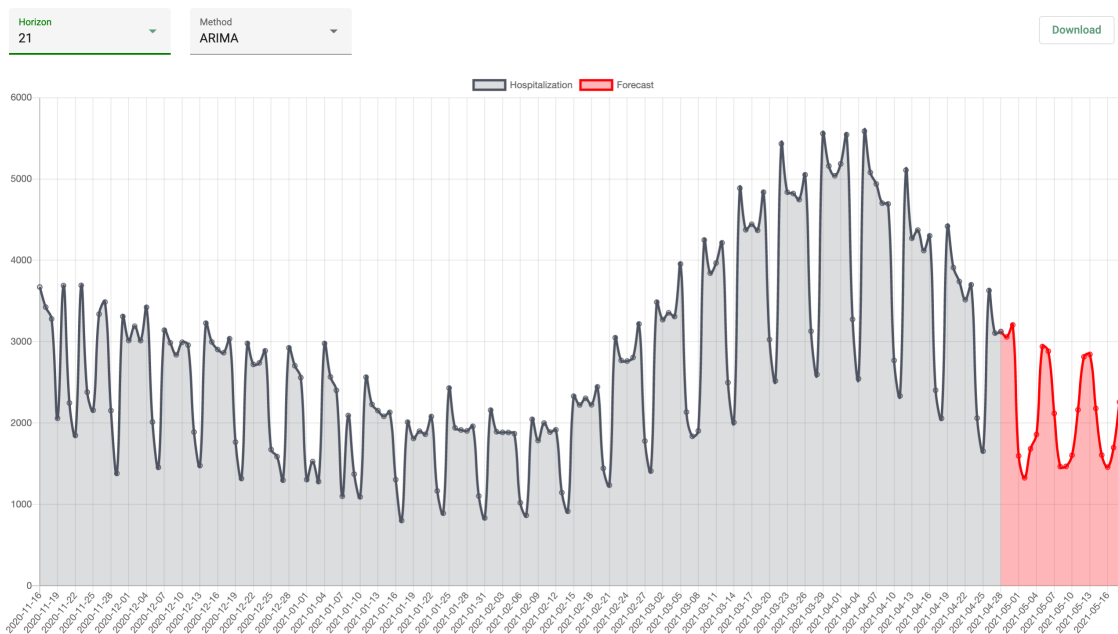


Рисунок 3.8 – Екранна форма головної сторінки з прогнозом на 21 день за допомогою ARIMA

На головній сторінці доступна опція завантаження результатів прогнозування разом з історичними даними, для подальшого аналізу, по кнопці «Download». На рисунку 3.9 наведена екранна форма, де відображено вміст завантаженого файлу після натискання на кнопку «Download».

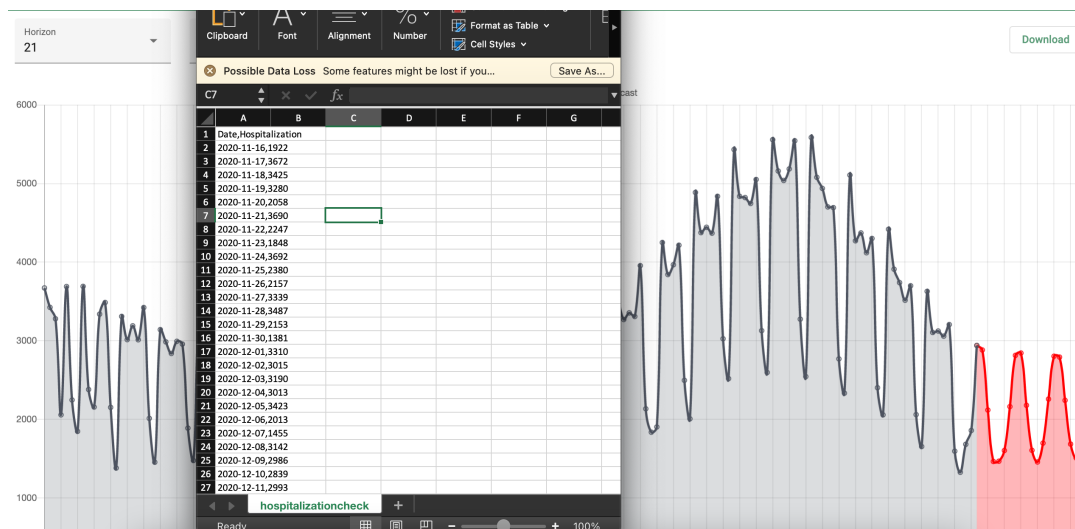


Рисунок 3.9 – Екранна форма завантаженого файлу на фоні головної сторінки програмного продукту

Надійність. Програмний продукт має достатньо високий рівень надійності та готовність до відпрацювання з помилками, яку можуть бути виконані користувачем.

Мобільність. Програмний продукт працює на ПК, із встановленим браузером, щоб було можливо переглянути веб-сторінку.

Зручність використання. Програмний продукт повинен бути зі зручним «user-friendly» інтерфейсом, який дозволить користувачам користуватися системою інтуїтивно, без попереднього ознайомлення з програмним продуктом.

Усі користувачі мають доступ до перегляду інформації на головній сторінці та до збереження потрібного їм прогнозу у вигляді файлу з розширенням .csv на локальний ПК.

Висновки до розділу

У розділі з описом програмного та технічного забезпечення був здійснений опис системи з прогнозування поширення COVID-19 в Україні. Було описано етап збору даних, які потрібні для прогнозування, а також розкрито технологічну сторону цього етапу, з використанням хмарних технологій та даних від РНБО. Також було здійснено опис процесу розробки та взаємодії компонентів програмного продукту.

4 РОЗРОБКА СТАРТАП-ПРОЕКТУ

4.1 Опис ідеї проекту

Опис ідеї стартап-проекту, напрямки застосування описані у таблиці 4.1.

Таблиця 4.1 – Опис ідеї стартап-проекту

Зміст ідеї	Напрямки застосування	Вигоди для користувача
Розробка інформаційної системи з прогнозування поширення COVID-19	1. Прогнозування випадків госпіталізації	Розуміння на державному рівні епідеміологічної ситуації
	2. Прогнозування кількості зайнятих ліжко-місць	Розуміння на регіональних рівнях потреби у коригуванні кількості ліжко-місць
	3. Уточнення прогнозу у зв'язку з коригуванням даних	Експерт з прогнозування буде здатен провести власний аналіз та зробити незалежні висновки

Інформаційна система з прогнозування поширення COVID-19 дозволяє отримати тижневий прогноз за останніми статистичними даними по госпіталізації; дозволяє отримати прогноз обираючи діапазон прогнозування та метод.

Довгостроковими перспективами є:

- додавання можливості прогнозування за агрегованими даними (тиждень, місяць);
- додавання можливості отримати прогноз за відкоригованими експертом даними;
- збільшення кількості обласних медичних установ, які будуть використовувати дану систему прогнозування.

Потреби в стартовому фінансуванні:

Стартовий капітал = 10000 грн

У таблиці 4.2 наведено техніко-економічні характеристики ідеї.

Таблиця 4.2 – Визначення сильних, слабких та нейтральних характеристик ідеї проєкту розробки інформаційної системи з прогнозування поширення COVID-19

№ з/п	Техніко-економічні характеристики ідеї	(потенційні) товари/концепції конкурентів				W (слабка сторона)	N (нейтральна сторона)	S (сильна сторона)
		Мій проєкт	Конкурент 1 BlueDot	Конкурент 2 Diagnostiс Robotics	Конкурент 3 IHME			
1.	Бюджетне фінансування	виділення власних коштів	розробка комерційна	розробка комерційна	державна розробка	повна відсутність фінансування	часткова бюджетне фінансування	виділення бюджетних коштів на розробку
2.	Використання сучасної техніки	використано найсучаснішу техніку та технічні засоби для проєктування	використано найсучаснішу техніку для проєктування	використано найсучаснішу техніку для проєктування	використано найсучаснішу техніку для проєктування	використано техніку, яка є застарілою	часткова комплектація технікою	використано найсучаснішу техніку та технічні засоби для проєктування

Продовження таблиці 4.2.

№ з/ п	Техніко- економіч ні характер истики ідеї	(потенційні) товари/концепції конкурентів				W (слабк а сторо на)	N (нейтр альна сторо на)	S (сильн а сторо на)
3.	Належна матеріаль но- технічна база	розра бка систе ми здійсн юється на власні й техніці і з викор истан ням власн их коштів	розроб ка здійсн юється на базі інформ аційног о центру	розр обка здійс нюєт ся на базі інфо рмац ійног о цент ру	розроб ка здійсн юється на базі бюдже тної устано ви	розроб ка здійсн юється на базі інфор маційн ого центру	розроб ка здійсн юється на базі бюдже тної устано ви	розроб ка систем и здійсн юється на власні й техніці з викори стання м власни х коштів
4.	Налагодж ена система реклами продукту	рекла ма відсут ня	реклам а присут ня	рекл ама прис утня	реклам а відсутн я	немає реклам и	частко ва реклам а	реклам а присут ня

Продовження таблиці 4.2.

№ з/п	Техніко-економічні характеристики ідеї	(потенційні) товари/концепції конкурентів				W (слабка сторона)	N (нейтральна сторона)	S (сильна сторона)
5.	Високий рівень розробки	запропоновані методи та алгоритми є досконалими	запропоновані методи та алгоритми є досконалими	запропоновані методи та алгоритми є досконалими	запропоновані методи та алгоритми є досконалими	на базі розробки є помилки вона потребує доробок	розробко є майже досконалою	запропоновані методи та алгоритми є досконалими
6.	Професіонали програмісти	Головний розробник - студент	Головний розробник - група професіоналів	Головний розробник - група професіоналів	Головний розробник - група професіоналів	Головний розробник - студент	Головний розробник – програм професіонал	Головний розробник - група професіоналів

4.2 Технологічний аудит ідеї проекту

Визначення технологічної здійсненності ідеї проекту передбачає аналіз таких складових у таблиці 4.3:

- за якою технологією буде виготовлено товар згідно ідеї проекту?
- чи існують такі технології, чи їх потрібно розробити/доробити?
- чи доступні такі технології авторам проекту?

Таблиця 4.3 – Технологічна здійсненність ідеї проекту розробки інформаційної системи планування ресурсів IT-проектів

№ п/п	Ідея проєкту	Технології її реалізації	Наявність технологій	Доступність технологій
1	Дослідження особливостей формування інформаційної системи з прогнозування поширення COVID-19	Технологія 1 (технологія збору даних)	реалізована	доступні
2		Технологія 2 (наявність бази досліджень)	наявні	доступні
3		Технологія 3 (база проведення досліджень (випробувань))	потрібно розробити	доступні
4		Технологія 4 (оформлення результатів дослідження)	потрібно розробити	доступні
Обрана технологія реалізації ідеї проєкту розробки інформаційної системи з прогнозування поширення COVID-19: є можливою				

4.3 Аналіз ринкових можливостей запуску стартап проєкту

Визначення ринкових можливостей, які можна використати під час ринкового впровадження проєкту, та ринкових загроз, які можуть перешкодити реалізації проєкту, дозволяє спланувати напрями розвитку проєкту із урахуванням стану ринкового середовища, потреб потенційних клієнтів та пропозицій проєктів-конкурентів. Цю інформацію представлено у таблиці 4.4.

Таблиця 4.4 – Попередня характеристика потенційного ринку стартап-проекту

№ п/п	Показники стану ринку (найменування)	Характеристика
1	Кількість головних гравців, од	4
2	Загальний обсяг продаж, грн/ум.од	55000
3	Динаміка ринку (якісна оцінка)	зростає
4	Наявність обмежень для входу (вказати характер обмежень)	немає
5	Специфічні вимоги до стандартизації та сертифікації	ДСТУ 2844-94 ДСТУ ISO 9000-2007 (ISO 9000:2005, IDT) ДСТУ ITU-T G.957:2010
6	Середня норма рентабельності в галузі (або по ринку), %	32

На основі проведеного дослідження є можливість стверджувати про привабливість проекту для входження на ринок за попереднім оцінюванням.

У таблиці 4.5 наведено перелік характеристик потенційних клієнтів даного стартап-проекту, а у таблиці 4.6 наведено відповідно фактори загроз та можливостей відповідно. У таблиці 4.7 наведено ступеневий аналіз конкуренції на ринку.

Таблиця 4.5 – Характеристика потенційних клієнтів стартап-проекту

№ п/п	Потреба, що формує ринок	Цільова аудиторія (цільові сегменти ринку)	Відмінності у поведінці різних потенційних груп клієнтів	Вимоги споживачів до товару
1	Інформація у реальному часі, яка показує можливі варіанти прогнозу поширення COVID-19	Державні структури, обласні лікарні	Органи державного управління зможуть коригувати заходи щодо запобігання поширенню інфекції, обласні лікарні зможуть правильно розраховувати кількість вільних ліжко-місць	відповідність ДСТУ 2844-9 Обов’язкова наявність сертифікатів

Таблиця 4.6 – Фактори загроз

№ п/п	Фактор	Зміст загрози	Можлива реакція компанії
1	Відсутність оновлених даних	Невчасне оновлення даних державними структурами	Інформаційна система може надати результати, які не враховують останні дані
2	Нестабільність політичної ситуації як на території України так і в світі	Варіювання курсу валют	Може порушити надійну систему постачальників

Продовження таблиці 4.6.

№ п/п	Фактор	Зміст загрози	Можлива реакція компанії
3	Висока вартість інформаційної системи	Збільшення ціни на систему	Підвищить агресивність конкурентів
4	Економічні складності при розробці	Повна відсутність фінансування або його мінімальне фінансування	Може порушити фінансове забезпечення компанії

Таблиця 4.7 – Фактори можливостей

№ п/п	Фактор	Зміст можливості	Можлива реакція компанії
1	Тривале існування інформаційної системи з прогнозування поширення COVID-19	Тривале існування інформаційної системи з прогнозування поширення COVID-19	На ринку дає можливість виходу на нові ринки у інших країнах
2	Моніторинг потреб споживачів інформаційної системи з прогнозування поширення COVID-19	Розуміючи потреби споживачів, розширювати діапазон методів та вносити поправки до існуючих, аналіз нових даних	Розширення діапазону методів та способів прогнозування

Продовження таблиці 4.7.

№ п/п	Фактор	Зміст можливості	Можлива реакція компанії
3	Висока вартість інформаційної системи з прогнозування поширення COVID-19	Встановлення високої ціни на інформаційну систему з прогнозування поширення COVID-19	Ускладнює вихід на нові ринки інформаційної системи з прогнозування поширення COVID-19

У таблиці 4.8 наведено ступеневий аналіз конкуренції на ринку, а у таблиці 4.9 описано аналіз конкуренції в галузі за М. Портером.

Таблиця 4.8 – Ступеневий аналіз конкуренції на ринку

Особливості конкурентного середовища	В чому проявляється дана характеристика	Вплив на діяльність підприємства (можливі дії компанії, щоб бути конкурентоспроможною)
1. Вказати тип конкуренції - монополія/олігополія/ монополістична/чиста	Локальний/національний бізнес. Глобальні сили є не досить вагомими по відношенню до локальних сил, які визначаються наявністю сертифікації, відповідності держ нормам і стандартам, регулюванням інформаційної галузі державою.	працює в рамках підвищення параметрів процесу планування ресурсів ІТ-проектів
2. За рівнем конкурентної боротьби- локальний/національний/...	Міжнародний	Ведучи конкуренцію на міжнародному рівні, компанії необхідно прикласти належні зусилля для охоплення всього ринку щодо реалізації інформаційної системи з прогнозування поширення COVID-19

Продовження таблиці 4.8.

Особливості конкурентного середовища	В чому проявляється дана характеристика	Вплив на діяльність підприємства (можливі дії компанії, щоб бути конкурентоспроможною)
3. За галузевою ознакою - міжгалузева/ внутрішньогалузева	Внутрішньогалузева. Конкуренція на ринку ведеться в інформаційній галузі України	Необхідно зосередити зусилля на пошуку конкурентних переваг, які дозволять компанії займати стійкі конкурентні позиції
4. Конкуренція за видами товарів: - товарно-родова - товарно-видова - між бажаннями	Товарно-родова. Конкуренція на рівні технології задоволення потреб. Існує конкуренція з іншими моделями, алгоритмами	Методи параметрів інформаційної системи з прогнозування поширення COVID-19 водночас веде конкурентну боротьбу як з товарами-субститутами

Продовження таблиці 4.8.

Особливості конкурентного середовища	В чому проявляється дана характеристика	Вплив на діяльність підприємства (можливі дії компанії, щоб бути конкурентоспроможною)
5. За характером конкурентних переваг - цінова / нецінова	Нецінова. При виборі алгоритмів та методів споживач звертає увагу на ефективність методів та рівень якості інформаційної системи з прогнозування поширення COVID-19. Цінова. Для значної частки споживачів ціна є визначальною при виборі інформаційної системи з прогнозування поширення COVID-19	Головною конкурентною перевагою є унікальність позиціонування інформаційної системи з прогнозування поширення COVID-19
6. За інтенсивністю - марочна / не марочна	Марочна	Диференціація методів та моделей за мотивом задоволення необхідних потреб споживачів

Таблиця 4.9 – Аналіз конкуренції в галузі за М. Портером

Складові аналізу	Процес	Висновки
Прямі конкуренти в галузі	Навести перелік прямих конкурентів	Прямих конкурентів на ринку України не існує, які б надавали прогноз у реальному часі з обраними діапазонами та методами.
Потенційні конкуренти	Визначити бар'єри входження в ринок	Бар'єри входу на ринок є порівняно незначними. Вартість організації бізнесу з виробництва сучасних інформаційних систем з прогнозування поширення COVID-19 сягає 10 тис. дол.
Постачальники	Визначити фактори сили постачальників	Існує чітка залежність від постачальників як якості інформаційної системи з прогнозування поширення COVID-19.

Продовження таблиці 4.9.

Складові аналізу	Процес	Висновки
Клієнти	Визначити фактори сили споживачів	Споживачі відносяться до державних установ та медичних закладів обласних рівнів. Покупка програмних додатків та алгоритмів реалізації прогнозування поширення COVID-19 несе не імпульсивний характер, це критична необхідність, щоб розроблювати подальші плани з запобігання поширенню інфекції.
Товари-замінники	Фактори загроз з боку замінників	Конкуренції наразі немає, так як було знайдено лише кілька аналогів, які працюють за кордоном.

Отже, відповідно до наведеного вище аналізу головними силами, які діють на конкуренцію в галузі є постачальники та споживачі. Також в силу розвитку ринку все більшого значення набуває інтенсивність конкуренції між існуючими конкурентами та загроза зі сторони товарів-субститутів [75-80].

Таким чином в межах структурного підходу до аналізу конкуренції тип конкуренції – монополістична конкуренція.

У таблиці 4.10 наведено пояснення та опис факторів конкурентоспроможності.

Таблиця 4.10 – Обґрунтування факторів конкурентоспроможності

№	Фактор конкурентоспроможності	Обґрунтування вибору
1	Частка ринку	Враховуючи той факт, що тип родового середовища в галузі – консолідований ринок, тобто існує група компаній, які контролюють разом понад 40% ринку, а також те, що інтенсивність суперництва між діючими конкурентами при низьких темпах зростання ринку є однією з головних сил, одним з найважливіших факторів конкурентоспроможності виступає частка ринку, яку займає виробник. В таких умовах чим більше частка ринку, тим більшими ринковими можливостями володіє виробник.
2	Ціна	Чим вигіднішою є ціна для споживача, тим вірогідніше його вибір.
3	Асортимент	В умовах збільшення інтенсивності між існуючими конкурентами завоювання споживачів відбувається за рахунок нових методів та алгоритмів.
4	Доступ до каналів розподілу	Споживач далеко не завжди проявляє прихильність до певної категорії розробників і дуже схильний до експериментів. В цьому випадку завоювати лояльність споживача дуже складно і ще складніше її утримати.
5	Торговий маркетинг	

Продовження таблиці 4.10.

№	Фактор конкурентоспроможності	Обґрунтування вибору
6	Рівень диференціації ТМ	В умовах ведення конкурентної боротьби на споживчому ринку, де попит є ірраціональним та існує велика кількість виробників і розробників при фактично відсутній різниці між товарами, що пропонуються, ключовим фактором успіху є здатність чітко диференціювати ТМ від ТМ конкурентів, надаючи споживачеві унікальну цінність.
7	Репутація виробника	Якщо компанія має бездоганну репутацію, особливо у сфері якості своєї продукції, то рівень довіри до неї зростає. Також репутація виробника важлива при виході на ринок з новими товарами, або при виході на нові сегменти, що полегшує позитивне сприйняття новинок.
8	Рівень лояльності до бренду	Чим вище рівень лояльності, тим більше компанія має прихильних, а значить постійних споживачів.
9	Унікальність позиціонування	В умовах монополістичної конкуренції, коли фактор диференціації ТМ є ключовим засобом ведення конкурентної боротьби, важливим є створення та підтримання унікального позиціонування, що створює певний захист від конкурентних зіткнень.

Кінець таблиці 4.10.

№	Фактор конкурентоспроможності	Обґрунтування вибору
10	Маркетинговий бюджет	Від розміру маркетингового бюджету залежить здатність здійснювати маркетингову стратегію підприємства. Маркетингові заходи мають забезпечувати інші конкурентні переваги такі, як рівень диференціації, лояльності, репутація виробника, дистрибуція та просування в торгових точках.

Висновки до розділу

У четвертому розділі здійснено розробку стартап проекту створення інформаційної системи з прогнозування поширення COVID-19. Здійснено аналіз конкурентного середовища, розкрито фактори впливу на систему, здійснено обґрунтування факторів конкурентоспроможності та обґрунтовано її важливість перш за все на державному рівні. В силу розвитку ринку все більшого значення набуває інтенсивність конкуренції між існуючими конкурентами на міжнародному рівні.

Проект є цілісною системою для входження на ринок за попереднім оцінюванням. Він принесе допомогу у коригуванні заходів, які мають бути прийняті на державному рівні, щоб запобігти розповсюдженню вірусу, а також зможе допомогти аналітикам на обласних рівнях у визначенні поточної ситуації з ліжко-місцями у медичних закладах.

ВИСНОВКИ

У магістерській дисертації здійснено розробку алгоритму прогнозування поширення COVID-19 в Україні за поденними даними по госпіталізації. Даний алгоритм лежить в основу веб-застосування, яке в реальному часі видає прогноз за обраний період.

У роботі були виконані такі завдання:

- був здійснений огляд предметної області, так як ситуація з COVID-19 постійно змінюється;
- був виконаний огляд відповідної літератури по прогнозуванню саме епідеміологічних процесів;
- виконано аналітичний огляд існуючих рішень по прогнозуванню, таких як імітаційне моделювання, метод групового урахування аргументів, епідеміологічні моделі, статистичні моделі та методи машинного навчання.

Даний програмний продукт може бути застосований державними установами та лікарнями, так як органи державного управління зможуть коригувати заходи щодо запобігання поширенню інфекції, а обласні лікарні зможуть правильно розраховувати кількість вільних ліжко-місць.

У результаті виконання магістерської дисертації:

- розроблено алгоритм прогнозування поширення COVID-19, який на основі базового прогнозу проводить коригування;
- розроблено програмну реалізацію алгоритму та створено веб-застосування, яке надає прогнози з обраним діапазоном та методом у реальному часі;
- виконано експериментальне дослідження розробленого алгоритму у порівнянні з іншими та порівняно їх точність.

9. Monitoring and Evaluation Framework COVID-19 Strategic Preparedness and Response (SPRP), TechNet-21. URL: [https://www.technet-21.org/en/library/main/6439-monitoring-and-evaluation-framework-covid%E2%80%9119-strategic-preparedness-and-response-\(sprp\)](https://www.technet-21.org/en/library/main/6439-monitoring-and-evaluation-framework-covid%E2%80%9119-strategic-preparedness-and-response-(sprp)).
10. Microsoft Power BI. URL: <https://app.powerbi.com/view?r=eyJrIjoiZDc1NTdjOTMtYjBjNC00ODZkLTk3MGEtNDM5ZTY1ODA3OGY5IiwidCI6ImY2MTBjMGI3LWJkMjQtNGIzOS04MTBiLTNkYzI4MGFmYjU5MCIsmiOjh9&pageName=ReportSection>.
11. Rt COVID-19. URL: <https://rt.live/>.
12. Coronavirus: What is the R number and how is it calculated? - BBC News. URL: <https://www.bbc.com/news/health-52473523>.
13. Reproduction number (R) and growth rate (r) of the COVID-19 epidemic in the UK, p. 86, 2020.
14. W. O. Kermack and A. G. McKendrick, Contributions to the mathematical theory of epidemics—I, *Bltm Mathcal Biology*, vol. 53, no. 1–2, pp. 33–55, Mar. 1991, doi: 10.1007/BF02464423.
15. G. R. Shinde, A. B. Kalamkar, P. N. Mahalle, N. Dey, J. Chaki, and A. E. Hassanien, Forecasting Models for Coronavirus Disease (COVID-19): A Survey of the State-of-the-Art, *SN Computer Science*, vol. 1, no. 4, 2020, doi: 10.1007/s42979-020-00209-9.
16. W.-S. Son and RISEWIDs Team, Individual-based simulation model for COVID-19 transmission in Daegu, Korea, *Epidemiol Health*, vol. 42, p. e2020042, 2020, doi: 10.4178/epih.e2020042.
17. N. Ghaffarzadegan, Simulation-based what-if analysis for controlling the spread of Covid-19 in universities, *PLOS ONE*, vol. 16, no. 2, p. e0246323, Feb. 2021, doi: 10.1371/journal.pone.0246323.
18. N. Shakhovska and M. O. Medykovskyy, Eds., GMDH-Based Discovering Dynamic Regularities of the Ukraine Covid-19 Pandemic Process, in

Advances in Intelligent Systems and Computing V, Cham, 2021, pp. 456–470, doi: 10.1007/978-3-030-63270-0_30.

19. B. Malavika, S. Marimuthu, M. Joy, A. Nadaraj, E. S. Asirvatham, and L. Jeyaseelan, Forecasting COVID-19 epidemic in India and high incidence states using SIR and logistic growth models, *Clinical Epidemiology and Global Health*, vol. 9, pp. 26–33, Jan. 2021, doi: 10.1016/j.cegh.2020.06.006.

20. I. Cooper, A. Mondal, and C. G. Antonopoulos, A SIR model assumption for the spread of COVID-19 in different communities, *Chaos Solitons Fractals*, vol. 139, p. 110057, Oct. 2020, doi: 10.1016/j.chaos.2020.110057.

21. Прогноз розвитку епідемії COVID-19 в Україні на 14–28 грудня 2020 року («Прогноз ПГ-32»). URL: <https://www.nas.gov.ua/UA/Messages/Pages/View.aspx?MessageID=7277>.

22. I. Nesteruk, Нові хвилі пандемії COVID-19 в Україні. New waves of COVID-19 pandemic in Ukraine. 2020.

23. V. K. Sharma and U. Nigam, Modeling and Forecasting of COVID-19 Growth Curve in India, *Transactions of the Indian National Academy of Engineering*, pp. 1–14, Sep. 2020, doi: 10.1007/s41403-020-00165-z.

24. Logistic equation and COVID-19. URL: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7444956/>.

25. M. Yousaf, S. Zahir, M. Riaz, S. M. Hussain, and K. Shah, Statistical analysis of forecasting COVID-19 for upcoming month in Pakistan, *Chaos, Solitons & Fractals*, vol. 138, p. 109926, Sep. 2020, doi: 10.1016/j.chaos.2020.109926.

26. R. K. Mojjada, A. Yadav, A. V. Prabhu, and Y. Natarajan, Machine learning models for covid-19 future forecasting, *Mater Today Proc*, Dec. 2020, doi: 10.1016/j.matpr.2020.10.962.

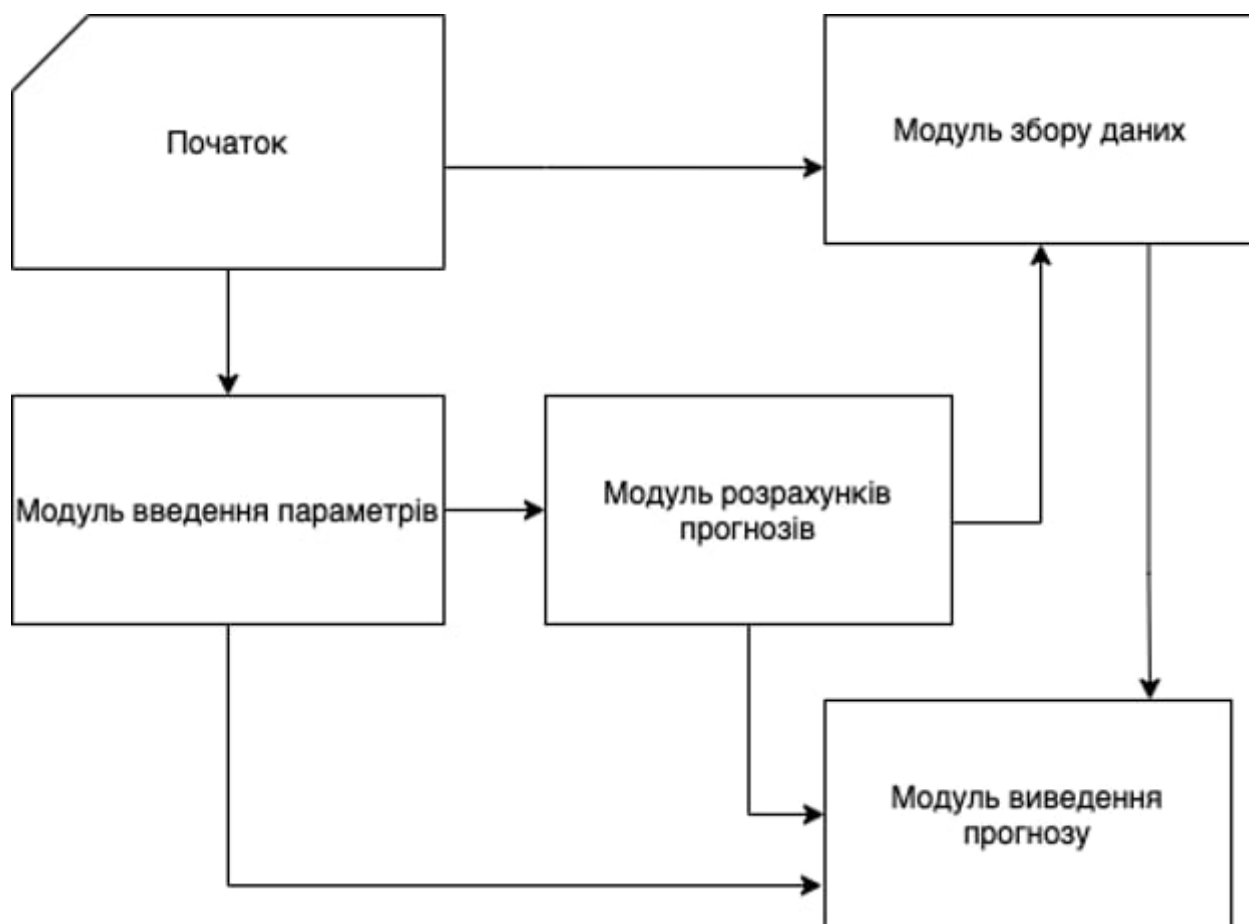
27. Stephanie, Lasso Regression: Simple Definition, *Statistics How To*, Sep. 24, 2015. URL: <https://www.statisticshowto.com/lasso-regression/>.

28. Support Vector Machines: A Simple Explanation, KDnuggets. URL: <https://www.kdnuggets.com/support-vector-machines-a-simple-explanation.html>.
29. Y. Zoabi, S. Deri-Rozov, and N. Shomron, Machine learning-based prediction of COVID-19 diagnosis based on symptoms, npj Digital Medicine, vol. 4, no. 1, Art. no. 1, Jan. 2021, doi: 10.1038/s41746-020-00372-6.
30. D. Greenhalgh and I. A. Moneim, SIRS Epidemic Model and Simulations Using Different Types of Seasonal Contact Rate, Systems Analysis Modelling Simulation, vol. 43, no. 5, pp. 573–600, May 2003, doi: 10.1080/023929021000008813.
31. S. He, Y. Peng, and K. Sun, SEIR modeling of the COVID-19 and its dynamics, Nonlinear Dyn, vol. 101, no. 3, pp. 1667–1680, Aug. 2020, doi: 10.1007/s11071-020-05743-y.
32. Exponential growth & logistic growth (article) | Khan Academy. URL: <https://www.khanacademy.org/science/ap-biology/ecology-ap/population-ecology-ap/a/exponential-logistic-growth>.
33. K. Gayantha, Software Architectural Patterns, Medium, Feb. 09, 2020. URL: <https://medium.com/@96kavindugayantha/software-architectural-patterns-17b05bfa3aef>.
34. П. Николай and Д. Владимир, Python 3. Самое необходимое, 2-е изд. БХВ-Петербург, 2019.
35. Службы облачных вычислений | Microsoft Azure. URL: <https://azure.microsoft.com/ru-ru/>.
36. Getting started with Azure Functions | Microsoft Docs. URL: <https://docs.microsoft.com/en-us/azure/azure-functions/functions-get-started?pivots=programming-language-python>.
37. AngularJS — Superheroic JavaScript MVW Framework. URL: <https://angularjs.org/>.

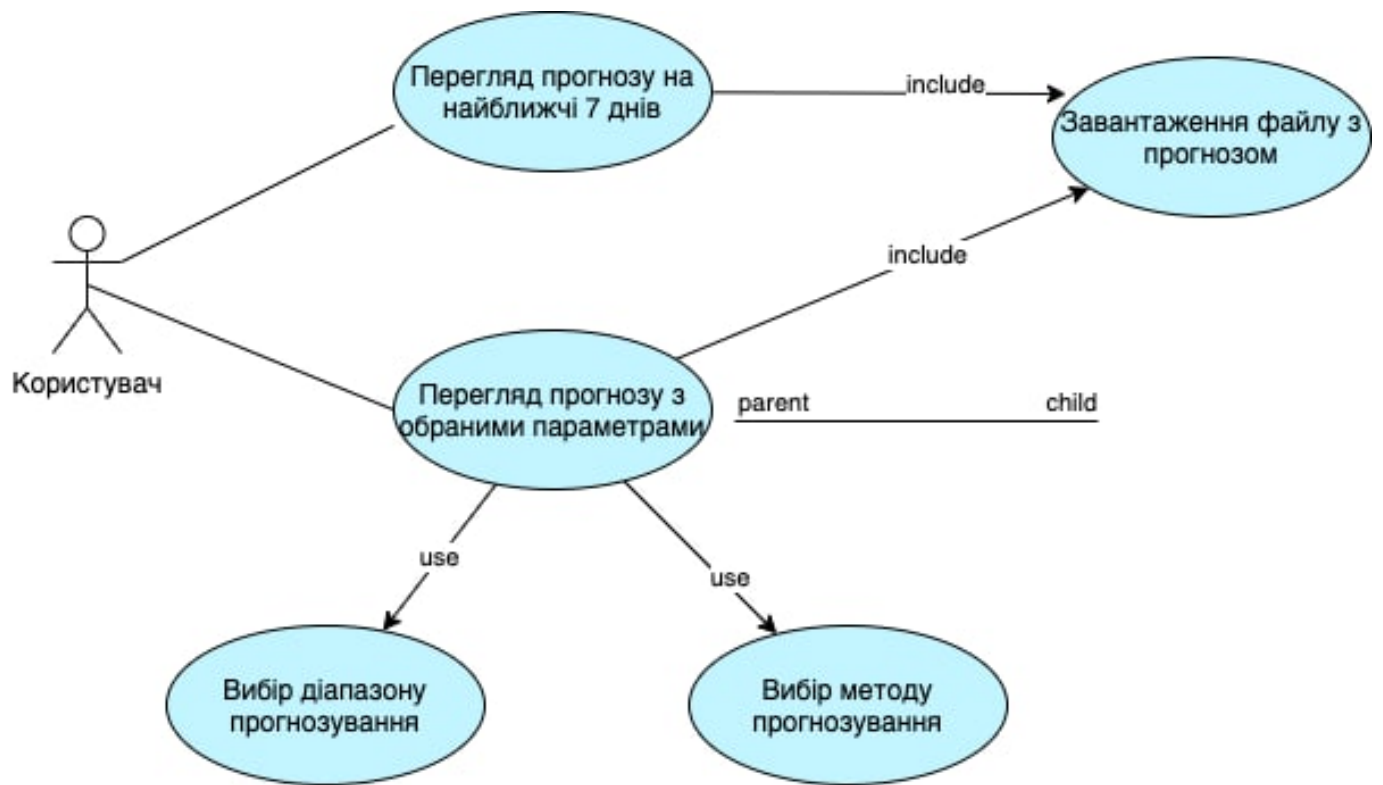
38. Методы HTTP запроса - HTTP | MDN. URL: <https://developer.mozilla.org/ru/docs/Web/HTTP/Methods>.
39. craigshoemaker, Timer trigger for Azure Functions. URL: <https://docs.microsoft.com/en-us/azure/azure-functions/functions-bindings-timer>.
40. Storage account overview - Azure Storage | Microsoft Docs. URL: <https://docs.microsoft.com/en-us/azure/storage/common/storage-account-overview>.
41. Мадоян Г. О. Математичні моделі прогнозування поширення Covid-19 та їх наслідків // Priority directions of science and technology development. Abstracts of the 4th International scientific and practical conference. SPC “Sci-conf.com.ua”. Kyiv, Ukraine. 2020. Pp. 461-465.

ДОДАТОК А

Графічний матеріал

Плакат 1 Функціонально-логічна структура програмного забезпечення

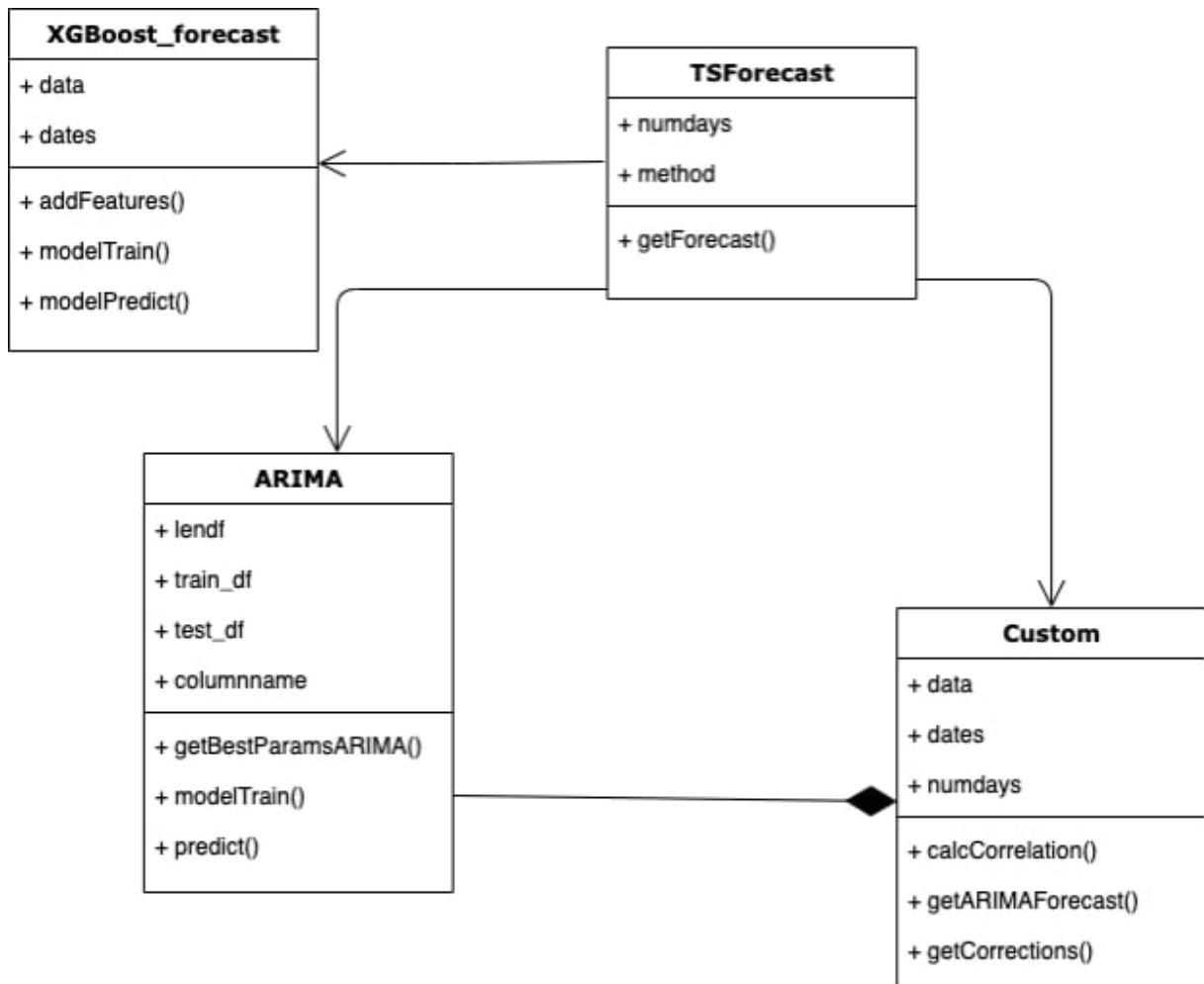
Плакат 2 UML-діаграма варіантів використання



Плакат 3 Діаграма послідовності



Плакати 4 UML-діаграма класів



Плакати 5 Математична постановка задачі

Маємо дискретні моменти часу $t = 1, 2, \dots, T$ та часовий ряд $y_1, y_2, \dots, y_T$, де кожний член ЧР $\{y_t\}, t \in [1, T]$ може бути поданий так:

$$y_t = u_t + v_t + w_t + z_t + \theta_t + \varepsilon_t, \quad (2.14)$$

де u_t – довготривала складова, тренд; v_t – циклічна компонента; w_t – сезонна складова, що формується під впливом сезонних коливань показника протягом заданого періоду часу, зазвичай року (часто включається у циклічну компоненту); z_t – компонента, яка забезпечує порівнюваність елементів ЧР; θ_t – управляюча змінна, за допомогою якої впливають на члени ЧР з метою формування в майбутньому його бажаної траєкторії (це – керований прогноз, який є основою для сценарного підходу).

Потрібно створити таку модель прогнозування, для якої середньоквадратичне відхилення фактичного значення від прогнозованого прямує до мінімального для заданого P . Тобто потрібно визначити прогнозні значення $y_{T+1}, \dots, y_{T+P}$.

За критерій точності можна взяти показник середньоквадратичного відхилення:

$$\varepsilon = \sqrt{\frac{\sum_{i=1}^n (y - \hat{y}_i)^2}{n}}, \quad (2.15)$$

По суті, (2.14) і є цільовою функцією, яку потрібно мінімізувати.

Плакати 6 Схеми розробленого алгоритму

1. Формування ряду рухомих сум за період довжиною в тиждень: $y_1, y_2, \dots, y_T, t \in [1, T]$;
2. Прогноз за допомогою ARIMA на n значень наперед
 - вибір параметрів для ARIMA, при яких маємо мінімальний показник RMSE;
 - розрахунок прогнозу з підібраними параметрами;
3. Коригування значень прогнозу:
 - розрахунок кореляції між останніми n значеннями нативного ряду $y_{T-n-1}, \dots, y_T$ та значеннями нативного ряду у минулому $y_i, \dots, y_{i+n-1}$, де $i = \overline{1, T - 2n}$.

Нехай $x = y_{T-n-1}, \dots, y_T, y = y_i, \dots, y_{i+n-1}$, тоді для розрахунку коефіцієнта кореляції маємо формулу:

$$r = \frac{\sum (x - m_x)(y - m_y)}{\sqrt{\sum (x - m_x)^2 \sum (y - m_y)^2}},$$

де m_x та m_y – середні значення векторів x та y відповідно.

- для частини ЧР з найбільшим показником кореляції розраховуємо частки кожного дня за період з n днів:

$$d_j = \frac{y_i}{n} \sum_{j=1}^n y_i$$

де $j = \overline{1, n}$.

- відповідно до днів тижня та розрахованих часток коригуємо прогнозні значення і отримуємо прогноз $y_{T+1}d_1, \dots, y_{T+n}d_n$.