Міністерство освіти і науки України
Тернопільський національний технічний університет імені Івана Пулюя

Факультет комп'ютерно-інформаційних систем і програмної інженерії
(повна назва факультету)
Кафедра комп'ютерних наук
(повна назва кафедри)

КВАЛІФІКАЦІЙНА РОБОТА

на здобуття освітнього ступеня

Магістр
(назва освітнього ступеня)

на тему: Моделі та методи прогнозування часових рядів із властивістю циклічності

Виконав: студент VI курсу, групи СНнм-61
спеціальності 122 Комп'ютерні науки
(шифр і назва спеціальності)

(підпис) Воробець І.В. (прізвище та ініціали)

Керівник (підпис) Фриз М.Є. (прізвище та ініціали)

Нормоконтроль (підпис) Мацюк О.В. (прізвище та ініціали)

Завідувач кафедри (підпис) Боднарчук І.О. (прізвище та ініціали)

Рецензент (підпис) Осухівська Г.М. (прізвище та ініціали)

Тернопіль
2023

Міністерство освіти і науки України
Тернопільський національний технічний університет імені Івана Пулюя

Факультет комп'ютерно-інформаційних систем і програмної інженерії
(повна назва факультету)

Кафедра комп'ютерних наук
(повна назва кафедри)

ЗАТВЕРДЖУЮ
Завідувач кафедри
_____ Боднарчук І.О.
(підпис) (прізвище та ініціали)
«____» 2023 р.

ЗАВДАННЯ
НА КВАЛІФІКАЦІЙНУ РОБОТУ

на здобуття освітнього ступеня Магістр
(назва освітнього ступеня)

за спеціальністю 122 Комп'ютерні науки
(шифр і назва спеціальності)

Студенту Воробцю Ігорю Володимировичу
(прізвище, ім'я, по батькові)

1. Тема роботи Моделі та методи прогнозування часових рядів із властивістю циклічності

Керівник роботи Фриз Михайло Євгенович, к.т.н., доцент кафедри КН
(прізвище, ім'я, по батькові, науковий ступінь, вчене звання)

Затверджені наказом ректора від « 22 » листопада 2022 року № 4/7-950

2. Термін подання студентом завершеної роботи 23 травня 2023 р.

3. Вихідні дані до роботи наукові, літературні та інтернет-джерела щодо часових рядів, їх циклічності, моделей, методів і засобів прогнозування рядів із властивістю циклічності.

4. Зміст роботи (перелік питань, які потрібно розробити)
Вступ. 1. Аналітичний огляд літературних джерел за напрямом досліджень. 1.1 Прогнозування часових рядів та аналіз предметних областей його використання. 1.2 Складові часових рядів. 1.3 Основи прогнозування часових рядів. 2. Дослідження методів та моделей прогнозування часових рядів. 2.1 Методи та моделі експоненційного згладжування. 2.2 Регресійні моделі. 2.3 Моделі ARIMA. 2.4 Комплексні моделі прогнозування та моделі штучних нейронних мереж. 3. Розробка програмного засобу для прогнозування циклічних часових рядів та аналіз експериментальних результатів. 3.1 Постановка завдання розробки програмного засобу прогнозування циклічних часових рядів. 3.2 Обґрунтування вибору методів та моделей прогнозування циклічних часових рядів. 3.3 Розробка програмного засобу прогнозування циклічних часових рядів. 3.4 Аналіз та прогнозування експериментальних даних. 4. Охорона праці та безпека в надзвичайних ситуаціях. 4.1 Характеристика НС природного характеру. Характеристика НС соціального характеру. 4.2 Підвищення стійкості роботи об'єктів приладобудівної галузі у воєнний час. Висновки. Перелік джерел. Додатки.

5. Перелік графічного матеріалу (з точним зазначенням обов'язкових креслень, слайдів)
1. Титульний слайд. 2. Мета і завдання. 3. Об'єкт, предмет дослідження, наукова новизна і практичне значення. 4. Актуальність дослідження. 5. Патерни часових рядів. 6. Похибки прогнозів MAE і MAPE. 7. Вибір моделей. 8. Модель ARIMA. 9. Експериментальні дані. 10. Аналіз експериментальних даних. 11. Результати прогнозування. 12. Похибки отриманих прогнозів. 13. Висновки. 14. Завершальний слайд.

6. Консультанти розділів роботи

Розділ	Прізвище, ініціали та посада консультанта	Підпис, дата	
		завдання видав	завдання прийняв
Охорона праці	Мацюк О.В., доцент		
Безпека в надзвичайних ситуаціях	Клепчик В.М., ст. викладач		

7. Дата видачі завдання 14 листопада 2022 р.

КАЛЕНДАРНИЙ ПЛАН

№ з/п	Назва етапів роботи	Термін виконання етапів роботи	Примітка
1.	Ознайомлення з завданням до кваліфікаційної роботи	14.11.2022-15.11.2022	*Виконано*
2.	Підбір джерел щодо часових рядів, їх циклічності,	16.11.2022-30.11.2022	*Виконано*
	моделей, методів і засобів прогнозування рядів із		
	властивістю циклічності		
3.	Аналіз джерел та збір даних по темі роботи	01.12.2022-15.01.2023	*Виконано*
4.	Виконання дослідження та розробки програмного	16.01.2023-12.03.2023	*Виконано*
	засобу згідно з метою кваліфікаційної роботи		
5.	Оформлення розділу «Аналітичний огляд	13.03.2023-23.03.2023	*Виконано*
	літературних джерел за напрямом досліджень»		
6.	Оформлення розділу «Дослідження методів та	24.03.2023-01.04.2023	*Виконано*
	моделей прогнозування циклічних часових рядів»		
7.	Оформлення розділу «Розробка програмного засобу	02.04.2023-07.04.2023	*Виконано*
	для прогнозування циклічних часових рядів та аналіз		
	експериментальних результатів»		
8.	Виконання завдання до підрозділу «Охорона праці»	08.04.2023-13.04.2023	*Виконано*
9.	Виконання завдання до підрозділу «Безпека в	14.04.2023-22.04.2023	*Виконано*
	надзвичайних ситуаціях»		
10.	Оформлення кваліфікаційної роботи	23.04.2023-30.04.2023	*Виконано*
11.	Нормоконтроль	01.05.2023-07.05.2023	*Виконано*
12.	Перевірка на плагіат	08.05.2023	*Виконано*
13.	Попередній захист кваліфікаційної роботи	15.05.2023	*Виконано*
14.	Захист кваліфікаційної роботи	23.05.2023	

Студент _____ Воробець І.В.
(підпис) (прізвище та ініціали)

Керівник роботи _____ Фриз М.Є.
(підпис) (прізвище та ініціали)

АНОТАЦІЯ

Моделі та методи прогнозування часових рядів із властивістю циклічності // Кваліфікаційна робота освітнього рівня «Магістр» // Воробець Ігор Володимирович // Тернопільський національний технічний університет імені Івана Пулюя, факультет комп'ютерно-інформаційних систем і програмної інженерії, кафедра комп'ютерних наук, група СНнм-61 // Тернопіль, 2023 // С. 76, рис. – 15, табл. – 1, слайдів – 14, додат. – 2, бібліогр. – 62.

Ключові слова: часові ряди, шаблони часових рядів, циклічні часові ряди, прогнозування часових рядів, модель прогнозування часових рядів, модель ETS, регресійна модель, модель ARIMA.

Кваліфікаційна робота присвячена розробці програмного засобу прогнозування часових рядів, що використовує методи та моделі прогнозування з урахуванням циклічності часових рядів.

У першому розділі кваліфікаційної роботи виконано аналіз джерел щодо часових рядів та їх прогнозування. Висвітлено базові поняття щодо часових рядів і їх прогнозування та розглянуто етапи процесу прогнозування часових рядів.

У другому розділі кваліфікаційної роботи проведено дослідження методів та моделей прогнозування часових рядів. Проаналізовано можливості роботи розглянутих методів і моделей з урахуванням циклічності часових рядів.

У третьому розділі кваліфікаційної роботи здійснено вибір методів і моделей прогнозування часових рядів, які враховують їх циклічність. На основі них виконано розробку програмного засобу. Здійснено виконання розробленого програмного засобу та проведено аналіз отриманих прогнозів.

Об'єкт дослідження: процес прогнозування часових рядів із властивістю циклічності. Предмет дослідження: методи, моделі та програмні засоби прогнозування часових рядів із властивістю циклічності.

ANNOTATION

Models and Methods of Time Series Forecast with the Cyclicity Feature // Qualification work of the educational level "Master" // Vorobets Ihor Volodymyrovych // Ternopil Ivan Pulyuy National Technical University, Faculty of Computer Information Systems and Software Engineering, Department of Computer Science, SNnm-61 group // Ternopil, 2023 // P. 76, fig. – 15, tables – 1, slides – 14, annexes – 2, ref. – 62.

Key words: time series, time series patterns, cyclical time series, time series forecasting, time series forecasting model, ETS model, regression model, ARIMA model.

The qualification work is devoted to the development of a time series forecasting software that uses forecasting methods and models taking into account the cyclicality of the time series.

In the first section of the qualification work, the analysis of the sources of time series and their forecasting was performed. The basic concepts of time series and their forecasting are covered, the stages of the time series forecasting process are considered.

In the second section of the qualification work, a study of time series forecasting methods and models was conducted. The work possibilities of the considered methods and models are analyzed, taking into account the cyclicality of time series.

In the third section of the qualification work, a selection of methods and models for forecasting time series, which take into account their cyclicity, was made. Based on them, the software was developed. The developed software was implemented and the received forecasts were analyzed.

The object of research: the process of forecasting time series with the property of cyclicity. The subject of research: methods, models and software for forecasting time series with the property of cyclicity.

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ, СИМВОЛІВ, СКОРОЧЕНЬ

y_t – часовий ряд;

$\hat{y}_{T+h|T}$ – точковий прогноз часового ряду y_t на крок h на основі значень часового ряду за моменти часу $t = 1,\ldots,T$;

$\mathbb{Z}$ – множина цілих чисел;

AR (англ. autoregression – авторегресія) – авторегресійна модель;

ARIMA (англ. autoregressive integrated moving average – авторегресійне інтегроване ковзне середнє) – модель авторегресійного інтегрованого ковзного середнього;

CSV (англ. comma-separated values – значення, розділені комою) – формат файлів, у якому зберігаються табличні дані, які відокремлюються між собою символом коми;

ETS (англ. error – похибка, trend – тренд, seasonal – сезонний компонент) – модель прогнозу часових рядів, що враховує стани похибок, тренду і сезонного компонента часового ряду;

MA (англ. moving average – ковзне середнє) – модель ковзного середнього;

MAE (англ. mean absolute error) – середня абсолютна похибка;

MAPE (англ. mean absolute percentage error) – середня абсолютна відсоткова похибка;

RMSE (англ. root mean squared error) – коренева середньоквадратична похибка;

SARIMA (англ. seasonal ARIMA) – сезонна модель ARIMA;

НС – надзвичайна ситуація;

ЦЗ – цивільний захист.

ЗМІСТ

ВСТУП 9
1 АНАЛІТИЧНИЙ ОГЛЯД ЛІТЕРАТУРНИХ ДЖЕРЕЛ ЗА НАПРЯМОМ ДОСЛІДЖЕНЬ 11
1.1 Прогнозування часових рядів та аналіз предметних областей його використання 11
1.2 Складові часових рядів 12
1.2.1 Шаблони часових рядів 12
1.2.2 Декомпозиція часових рядів 15
1.3 Основи прогнозування часових рядів 16
1.3.1 Базові поняття прогнозування 16
1.3.2 Прості методи прогнозування 18
1.3.3 Оцінювання точності прогнозів 20
1.3.4 Перехресна перевірка прогнозування 23
1.4 Висновки до першого розділу 24
2 ДОСЛІДЖЕННЯ МЕТОДІВ ТА МОДЕЛЕЙ ПРОГНОЗУВАННЯ ЧАСОВИХ РЯДІВ 25
2.1 Методи та моделі експоненційного згладжування 25
2.1.1 Основні методи експоненційного згладжування 25
2.1.2 Моделі ETS 28
2.1.3 Прогнозування за допомогою моделей ETS з урахуванням циклічності часових рядів 30
2.2 Регресійні моделі 30
2.2.1 Лінійна регресія 30
2.2.2 Нелінійна регресія 31
2.2.3 Оцінка регресійних моделей 32
2.2.4 Прогнозування за допомогою регресійних моделей з урахуванням циклічності часових рядів 34

2.3 Моделі ARIMA 35
2.3.1 Авторегресійні моделі та моделі ковзного середнього 35
2.3.2 Властивості моделей ARIMA 36
2.3.3 Основні характеристики сезонних моделей ARIMA 38
2.3.4 Прогнозування за допомогою моделей ARIMA та SARIMA з урахуванням циклічності часових рядів 39
2.4 Комплексні моделі прогнозування та моделі штучних нейронних мереж 40
2.4.1 Динамічні регресійні моделі 40
2.4.2 Векторні авторегресійні моделі 41
2.4.3 Моделі штучних нейронних мереж 43
2.5 Висновки до другого розділу 44
3 РОЗРОБКА ПРОГРАМНОГО ЗАСОБУ ДЛЯ ПРОГНОЗУВАННЯ ЦИКЛІЧНИХ ЧАСОВИХ РЯДІВ ТА АНАЛІЗ ЕКСПЕРИМЕНТАЛЬНИХ РЕЗУЛЬТАТІВ 45
3.1 Постановка завдання розробки програмного засобу прогнозування циклічних часових рядів 45
3.2 Обґрунтування вибору методів та моделей прогнозування циклічних часових рядів 46
3.2.1 Модель ARIMA 46
3.2.2 Модель XGBoost 47
3.2.3 Модель LightGBM 47
3.3 Розробка програмного засобу прогнозування циклічних часових рядів .. 48
3.3.1 Вибір засобів розробки 48
3.3.2 Розробка модулів програмного засобу 49
3.4 Аналіз та прогнозування експериментальних даних 53
3.4.1 Аналіз часового ряду для прогнозування 54
3.4.2 Використання програмного засобу прогнозування циклічних часових рядів 57
3.5 Висновки до третього розділу 61

4 ОХОРОНА ПРАЦІ ТА БЕЗПЕКА В НАДЗВИЧАЙНИХ СИТУАЦІЯХ 62
4.1 Характеристика НС природного характеру. Характеристика НС соціального характеру 62
4.2 Підвищення стійкості роботи об’єктів приладобудівної галузі у воєнний час 65
4.3 Висновки до четвертого розділу..... 68
ВИСНОВКИ..... 69
ПЕРЕЛІК ДЖЕРЕЛ 70
ДОДАТКИ

ВСТУП

Актуальність теми. Прогнозування часових рядів використовується у багатьох галузях людської діяльності, де його результати часто мають важливе значення під час прийняття рішень. Розвиток обчислювальних засобів та розробка просунутих алгоритмів машинного навчання зробили можливим прогнозування на основі великих об’ємів даних та дозволили удосконалити методи та моделі прогнозування для роботи з різними часовими рядами.

Окремим типом часових рядів є часові ряди із властивістю циклічності (або циклічні часові ряди). Вони характеризуються коливаннями даних, які зумовлюються нерегулярними процесами. Приклади циклічних часових рядів, як правило, пов’язані з природними (активність сонячних плям) або економічними процесами (зменшення безробіття, пов’язане з ростом економіки).

Прогнозування таких часових рядів повинно враховувати їх циклічну поведінку. Саме тому методи та моделі прогнозування часових рядів із властивістю циклічності є актуальним напрямом дослідження.

Мета і завдання дослідження. Метою даної кваліфікаційної роботи є підвищення ефективності процесу прогнозування циклічних часових рядів шляхом розробки програмного засобу прогнозування часових рядів, що використовує методи та моделі, які враховують їх циклічну поведінку. Для досягнення поставленої мети необхідно виконати наступні завдання:

- виконати аналіз джерел щодо часових рядів, їх типів та особливостей їх прогнозування;
- здійснити огляд основних методів та моделей прогнозування часових рядів і дослідити їх можливості опрацювання циклічної поведінки ряду;
- здійснити вибір методів та моделей прогнозування часових рядів, які враховують їх циклічність, та виконати його обґрунтування;
- виконати розробку програмного засобу, який використовує методи та моделі прогнозування часових рядів, що враховують їх циклічність;

- провести аналіз точності прогнозів методів і моделей прогнозування, отриманих під час використання розробленого програмного засобу.

Об'єкт дослідження: процес прогнозування часових рядів, а саме часових рядів із властивістю циклічності.

Предмет дослідження: методи та моделі прогнозування часових рядів із властивістю циклічності, програмні засоби для прогнозування таких рядів.

Наукова новизна одержаних результатів: набув подальшого розвитку спосіб визначення методу та моделі прогнозування часового ряду, який враховує наявність циклічності, що дозволяє підвищити точність прогнозування.

Практичне значення одержаних результатів. Розроблено програмний засіб, що обчислює прогнози циклічних часових рядів за допомогою різних методів та дозволяє аналізувати ці прогнози.

Апробація результатів кваліфікаційної роботи. Основні результати проведених досліджень обговорювалися на трьох конференціях:

- X науково-технічна конференція «Інформаційні моделі, системи та технології» (м. Тернопіль, 7-8 грудня 2022 р.);
- XI Міжнародна науково-технічна конференція молодих учених та студентів «Актуальні задачі сучасних технологій» (м. Тернопіль, 7-8 грудня 2022 р.);
- VI Міжнародна студентська науково-технічна конференція ТНТУ імені Івана Пулюя «Природничі та гуманітарні науки. Актуальні питання» (м. Тернопіль, 27-28 квітня 2023 р.).

Публікації. Основні результати кваліфікаційної роботи опубліковано у трьох працях конференцій (див. додаток А).

Структура та обсяг кваліфікаційної роботи. Кваліфікаційна робота складається зі вступу, чотирьох розділів, висновків, списку літератури з 62 найменувань та 2 додатків. Загальний обсяг роботи складає 76 сторінок, з яких 58 – сторінки основного тексту, що містить 15 рисунків та 1 таблицю.

1 АНАЛІТИЧНИЙ ОГЛЯД ЛІТЕРАТУРНИХ ДЖЕРЕЛ ЗА НАПРЯМОМ ДОСЛІДЖЕНЬ

1.1 Прогнозування часових рядів та аналіз предметних областей його використання

Під часовим рядом розуміється упорядкована послідовність значень певної змінної через однакові проміжки часу [1]. Часовий ряд можна розглядати як випадковий процес y_t, $t \in \mathbb{Z}$.

Прогнозування часового ряду передбачає знаходження якомога точних значень часового ряду на основі доступної інформації: минулих значень ряду, минулих тенденцій даних і знань про майбутні події, що можуть вплинути на прогнози. Для прогнозування важливим є припущення, що майбутні тенденції даних у ряді будуть подібними до минулих тенденцій [2].

Використання прогнозування часових рядів охоплює широкий ряд предметних областей. Основні з них зображено на рисунку 1.1.

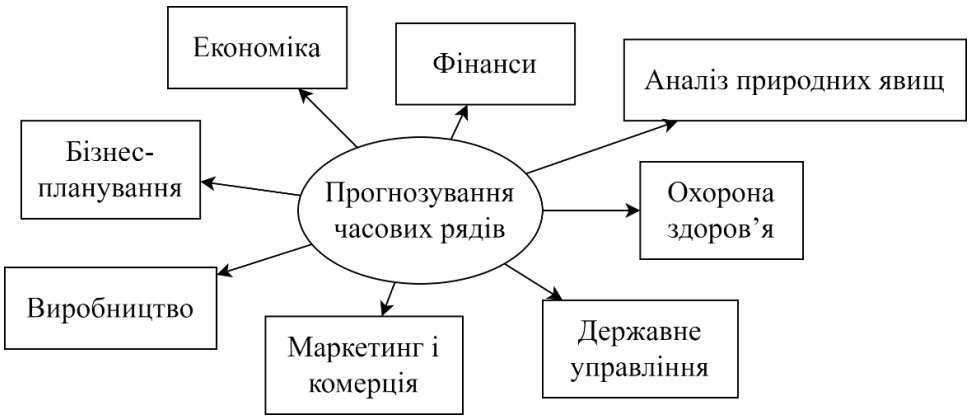

Рисунок 1.1 – Предметні області використання прогнозування часових рядів

В області фінансів прогнозування часових рядів може застосовуватися в аналізі фондового ринку або аналізі інвестиційних ризиків. Іншим прикладом є прогноз цін та курсу валют. В економіці прогнозування використовується під час аналізу економічного росту регіонів, прогнозування споживання електроенергії.

Бізнес-планування використовує прогнозування рядів для бюджетного аналізу й аналізу прибутковості [1, 4].

Прогнозування часових рядів у виробництві є необхідним для прогнозу попиту на ресурси, контролю якості, прогнозу кількості виготовленої продукції. Маркетинг і комерція користується засобами прогнозування рядів у, наприклад, прогнозі попиту й продажів різних товарів на основі відгуків користувачів чи в аналізі впливу проведення маркетингових кампаній на продажі товару [1, 6].

Прикладом прогнозування часових рядів у державному управлінні є аналіз перепису населення і прогнозування майбутньої його чисельності, планування державного або місцевого бюджету. В області охорони здоров'я прогнозування рядів використовується при прогнозі рівня зараження населення певної території з метою моделювання поширення епідемій та контролю спалахів захворювань [5, 6]. Також часові ряди та їх прогнозування використовується в діагностиці хвороб, зокрема серцево-судинних [7].

Окремою предметною областю є аналіз природних явищ. Він включає такі приклади прогнозування часових рядів: прогноз погоди, прогноз землетрусів [6].

1.2 Складові часових рядів

1.2.1 Шаблони часових рядів

Шаблон часового ряду дозволяє визначити властивості ряду та закономірності зміни значень у ньому протягом заданого часового проміжку. Загалом існує три шаблони часового ряду [2]:

- шаблон тренду;

Тренд описує тривале зменшення або збільшення даних у часовому ряді, проте воно не має бути лише лінійним. На великих часових проміжках тренд може змінюватися від зростання значень ряду до їх спадання і навпаки.

Як приклад шаблону тренду на рисунку 1.2 подано графік казначейських векселів США, укладених на ринку Чикаго протягом 100 торгових днів у 1984

році. Як можна побачити, цей графік відображає загальний тренд до спадання значень часового ряду [2].

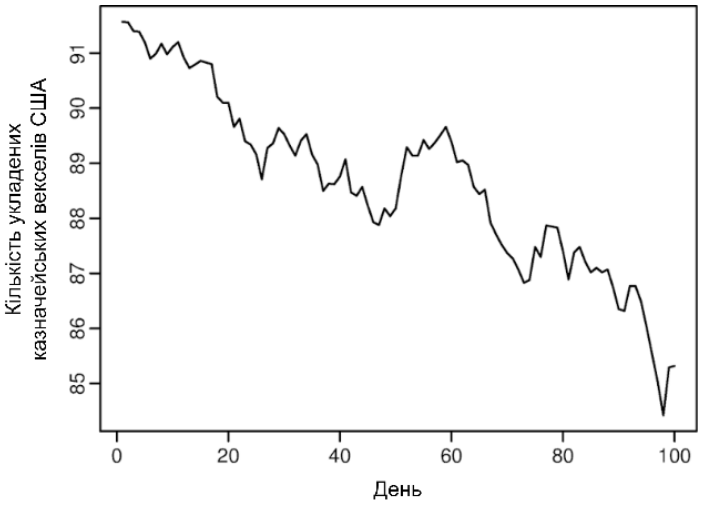

Рисунок 1.2 – Графік укладання казначейських векселів США на ринку Чикаго у 1984 році [8]

- шаблон сезонності;

У часовому ряді сезонність виникає тоді, коли на дані впливають різні сезонні фактори, такі, як день тижня, місяць, пора року. Сезонність має фіксований і визначений період.

Приклад сезонного шаблону показано на рисунку 1.3. На ньому подано графік щоквартального виробництва електроенергії в Австралії, яке проявляє помітну сезонність, а також чіткий тренд до зростання. Як можна зрозуміти, часові ряди можуть одночасно мати різні шаблони, що потрібно враховувати при виборі методу прогнозування [2, 8].

- шаблон циклічності.

Циклічність проявляється, якщо зростання й спадання у даних часового ряду не мають фіксованої частоти. Зазвичай такі коливання зумовлюються економічними умовами [8].

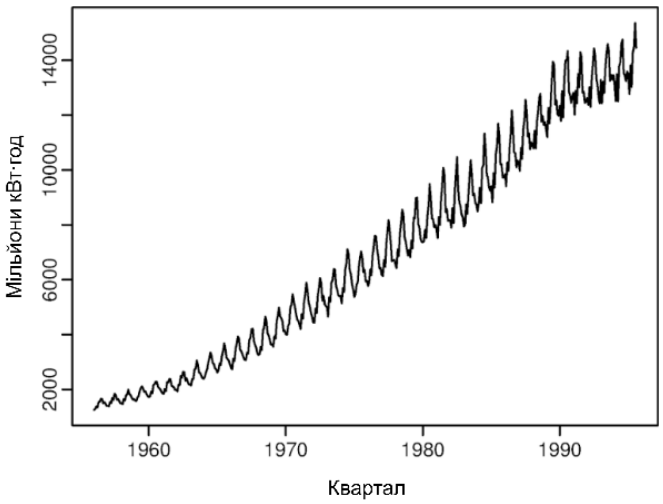

Рисунок 1.3 – Графік виробництва електроенергії в Австралії [8]

На рисунку 1.4 показано щомісячні продажі будинків в США у 1975-1995 роках. Ці дані показують чітку сезонність протягом кожного року. Також прослідковується циклічність, що проявляється спадами з подальшим підйомом кожні 6-10 років.

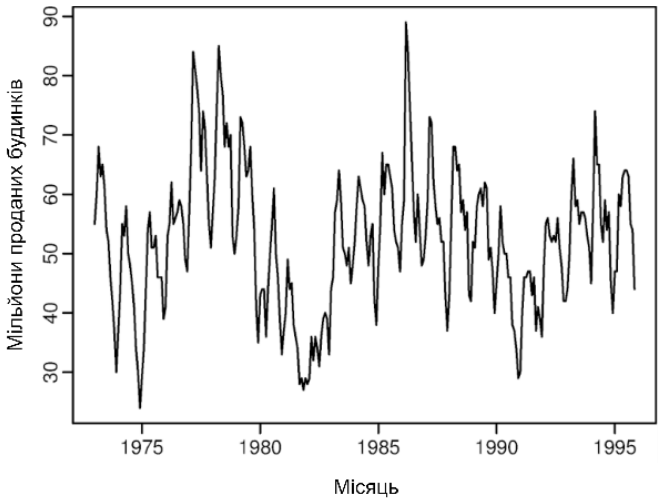

Рисунок 1.4 – Графік продажів будинків у США [8]

Крім нефіксованої частоти, циклічність від сезонності відрізняється ще й тим, що середня тривалість циклічних коливань більша від середньої тривалості сезонних, а також циклічні дані є менш сталими, ніж сезонні [2].

1.2.2 Декомпозиція часових рядів

Наявність того чи іншого шаблону в часовому ряді визначається його відповідними компонентами. Загалом часовий ряд розкладається на три компоненти [2, 9]:

- тренд-цикл (об'єднує шаблони тренду й циклічності, також може називатися просто трендом);
- сезонний компонент (деякі часові ряди можуть містити декілька сезонних компонентів, які пов'язані з різними сезонними періодами, наприклад, з тижнем, місяцем, роком);
- залишок (містить решту даних у часовому ряді, описує несистематичні й випадкові коливання) [10].

Для визначення цих компонентів використовується декомпозиція часового ряду. Вона дає можливість визначити, які шаблони більш виражені в ряді, що є корисним для аналізу часового ряду та вибору методів і моделей його прогнозування. Провести декомпозицію ряду можна двома способами: адитивним і мультиплікативним [11, 12].

Адитивна декомпозиція будує модель часового ряду на основі суми його компонентів, тобто:

$$y_t = T_t + S_t + R_t, \tag{1.1}$$

де y_t – дані часового ряду в момент часу t, T_t – тренд-цикл, S_t – сезонний компонент, а R_t – залишок.

Мультиплікативна декомпозиція, натомість, подає дані часового ряду як добуток компонентів:

$$y_t = T_t \cdot S_t \cdot R_t. \tag{1.2}$$

Вибір між адитивною й мультиплікативною декомпозицією ґрунтується на характері коливань у часовому ряді. Адитивна декомпозиція буде корисною для тих рядів, у яких коливання залишаються відносно постійними в часі. Мультиплікативна декомпозиція більше підходить, коли величина коливань ряду з часом змінюється [12].

Загальна процедура декомпозиції часового ряду описується наступними кроками [11]:

1. Знаходження оцінки тренду $\hat{T}_t$. Це можна зробити, наприклад, за допомогою згладжування ковзними середніми [2].

2. Вилучення тренду з даних ($y_t - \hat{T}_t$ – для адитивної декомпозиції, $y_t / \hat{T}_t$ – для мультиплікативної декомпозиції).

3. Визначення оцінки сезонного компонента $\hat{S}_t$. Для цього потрібно обчислити середні значення даних без тренду для кожного сезонного періоду (дня, місяця, кварталу).

4. Обчислення оцінки залишку $\hat{R}_t$ ($\hat{R}_t = y_t - \hat{T}_t - \hat{S}_t$ – для адитивної декомпозиції, $\hat{R}_t = y_t / (\hat{T}_t \cdot \hat{S}_t)$ – для мультиплікативної декомпозиції).

Розглянутий метод декомпозиції часових рядів є так званою класичною декомпозицією, яка слугує основою для інших методів декомпозиції. До таких методів належать X-11, SEATS (англ. Seasonal Extraction in ARIMA Time Series), STL (англ. Seasonal and Trend decomposition using Loess) [2].

1.3 Основи прогнозування часових рядів

1.3.1 Базові поняття прогнозування

Важливими поняттями в прогнозуванні часових рядів є тип прогнозу, який треба визначити перед прогнозуванням, а також підігнані значення й залишки, що мають значення для аналізу результатів прогнозування.

Залежно від умов використовуються два типи прогнозів, які можна отримати за допомогою використання методів і моделей прогнозування. Найбільш поширеним є точковий прогноз, що подає результат прогнозування у вигляді одного значення. Точковий прогноз часового ряду y_t на основі значень за моменти часу $t = 1,\ldots,T$ записується як $\hat{y}_{T+h|T}$, де h – горизонт прогнозування, тобто крок, на який знаходиться прогноз [2, 13].

Проте точковий прогноз ніколи не відповідає майбутнім фактичним значенням ряду. Його можна представити як середнє значення розподілу ймовірностей прогнозу. Для того, щоб показати невизначеність прогнозу, використовується інтервал прогнозування, у межах якого очікується прогнозоване значення з певною ймовірністю [13, 14]. У більшості випадків розподіл прогнозу є нормальним, тому межі інтервалу прогнозування можна обчислити наступним чином [2]:

$$\hat{y}_{T+h|T} \pm c \cdot \hat{\sigma}_h, \tag{1.3}$$

де c – множник, який залежить від ймовірності попадання прогнозу в інтервал прогнозування, $\hat{\sigma}_h$ – оцінка середнього квадратичного відхилення розподілу прогнозу на крок h.

Підігнані значення дуже схожі на точкові прогнози часового ряду. Їх відмінність полягає в тому, що підігнані значення є оцінками, отриманими з допомогою методу прогнозування на основі всіх значень у часовому ряді, у тому числі й майбутніх. Тому підігнане значення часового ряду в момент часу t позначається як $\hat{y}_t$ [2].

За допомогою підігнаних значень часового ряду можна обчислити залишки прогнозування. Залишок e_t визначається як різниця значень ряду та відповідних підігнаних значень:

$$e_t = y_t - \hat{y}_t . \tag{1.4}$$

Залишки є корисними для аналізу моделей прогнозування часових рядів. Наприклад, характеристики залишків дозволяють перевірити, чи адекватно модель врахувала інформацію з даних часового ряду [2, 15].

1.3.2 Прості методи прогнозування

Існує ряд методів прогнозування часових рядів, які є простими і досить ефективними. Основними простими методами прогнозування є метод середнього, наївний метод, сезонний наївний метод і метод дрейфу. Кожен із цих методів далі розглядається окремо.

У методі середнього прогнози всіх майбутніх значень часового ряду дорівнюють середньому значенню відомих даних у ряді. Тоді прогноз на крок h можна записати наступним чином:

$$\hat{y}_{T+h|T} = \overline{y} = (y_1 + \ldots + y_T)/T , \tag{1.5}$$

де $\overline{y}$ – середнє значення, $y_1, \ldots, y_T$ – відомі дані за моменти часу $t = 1, \ldots, T$.

Наївний метод для прогнозу використовує останнє відоме значення часового ряду. Це можна записати так [16]:

$$\hat{y}_{T+h|T} = y_T . \tag{1.6}$$

Сезонний наївний метод є вдосконаленням наївного методу, який враховує сезонність. За цим методом для прогнозу встановлюється значення часового ряду в тому самому моменті часу з минулого відомого періоду. Формально наївний метод записується як [2, 16]:

$$\hat{y}_{T+h|T} = y_{T+h-m(k+1)}, \tag{1.7}$$

де m – сезонний період, k – ціла частина від $(h-1)/m$ (кількість повних сезонних періодів у прогнозованому відрізку часу від T до $T+h$).

Іншою варіацією наївного методу є метод дрейфу, що враховує тренд часового ряду. Для врахування тренду встановлюється параметр дрейфу, який визначається як середня зміна відомих значень ряду. Тоді прогноз можна обчислити як:

$$\hat{y}_{T+h|T} = y_T + \frac{h}{T-1}\sum_{t=2}^{T}(y_t - y_{t-1}). \tag{1.8}$$

Формули (1.5)-(1.8) використовуються для обчислення точкового прогнозу. Інтервал прогнозування можна знайти за допомогою (1.3), де використовується оцінка середнього квадратичного відхилення розподілу прогнозу $\hat{\sigma}_h$. Далі подано формули обчислення $\hat{\sigma}_h$ для кожного з розглянутих простих методів прогнозування [2]:

- метод середнього: $\hat{\sigma}_h = \hat{\sigma}\sqrt{1+1/T}$,
- наївний метод: $\hat{\sigma}_h = \hat{\sigma}\sqrt{h}$,
- сезонний наївний метод: $\hat{\sigma}_h = \hat{\sigma}\sqrt{k+1}$,
- метод дрейфу: $\hat{\sigma}_h = \hat{\sigma}\sqrt{1+h/(T-1)}$,

де $\hat{\sigma}$ – оцінка середнього квадратичного відхилення залишків, T – кількість значень часового ряду, h – горизонт прогнозу, k – ціла частина від $(h-1)/m$ (кількість повних сезонних періодів у відрізку часу від T до $T+h$).

Хоч розглянуті методи прогнозування зазвичай не завжди надають найкращі результати, але вони можуть слугувати показником для розроблених методів. Тоді прості методи використовуються для порівняння з іншими методами прогнозування [2].

1.3.3 Оцінювання точності прогнозів

Для оцінювання результатів прогнозу залишки не будуть надійним показником, оскільки вони обчислюються на основі даних часового ряду, які використовувалися для побудови й підгонки моделі прогнозування. Перевірку точності прогнозу можна досягти тільки за допомогою обробки нових даних.

Саме тому для підгонки моделі прогнозування широковживаним є розділення ряду на набори навчальних і тестових даних. Набір навчальних даних використовується для навчання (підгонки) моделі та визначення її параметрів. Тестові дані використовуються для перевірки роботи моделі й оцінки її точності.

Проте при навчанні та оцінці роботи моделі варто враховувати, що модель, яка дуже добре відповідатиме навчальним даним, не зможе робити точні прогнози на основі інших даних. Надмірне навчання моделі призводить до її нездатності визначати у даних систематичний шаблон [2].

Оцінювання точності точкового прогнозу здійснюється за допомогою похибок прогнозу – різниць між значеннями часового ряду та їхніми прогнозами. Якщо навчальними даними є $y_1,\ldots,y_T$, а $y_{T+1},\ldots,y_{T+h}$ – тестові дані, то похибка прогнозу на крок h записується наступним чином:

$$e_{T+h} = y_{T+h} - \hat{y}_{T+h|T}\,. \tag{1.9}$$

Похибки прогнозу відрізняються від залишків тим, що вони обчислюються на основі тестових даних у той час, як залишки обчислюються на основі навчальних даних. Також залишки базуються на однокрокових прогнозах, похибки прогнозу, натомість, можуть отримуватися також за допомогою прогнозів на декілька кроків [2].

Для узагальнення похибки точкового прогнозу існують різні варіанти обчислення. Надалі розглядатимуться основі варіанти похибок. Узагальнені похибки прогнозу можна поділити на похибки, які залежать від масштабу, і відсоткові похибки.

Похибки, які залежать від масштабу, використовують ті самі одиниці вимірювання, що й значення часового ряду. Тому оцінки прогнозу, що базуються на цих похибках, не можна використовувати для порівняння між рядами з різними одиницями. Популярними похибками, які залежать від масштабу, є середня абсолютна похибка (англ. mean absolute error, MAE) та коренева середньоквадратична похибка (англ. root mean squared error, RMSE).

Похибка MAE обчислюється як середнє значення абсолютних значень (модулів) похибок прогнозу:

$$\mathrm{MAE} = \frac{1}{h}\sum_{i=1}^{h}\left|e_{T+i}\right|. \tag{1.10}$$

Похибка RMSE є квадратним коренем середнього значення квадратів похибок прогнозу:

$$\mathrm{RMSE} = \sqrt{\frac{1}{h}\sum_{i=1}^{h}e_{T+i}^{2}}\ . \tag{1.11}$$

При виборі однієї з цих похибок варто зазначити, що модель, яка мінімізує значення MAE, веде до медіани значень ряду, а мінімізація значень RMSE веде модель до середнього значення [17].

На противагу похибкам, які залежать від масштабу, відсоткові похибки не базуються на одиницях вимірювання, а тому їх можна використовувати для порівняння точності прогнозу між різними часовими рядами. Найчастіше використовується середня абсолютна відсоткова похибка (англ. mean absolute percentage error, MAPE), яка дорівнює середньому значенню абсолютних значень відсоткових похибок:

$$\text{MAPE} = \frac{1}{h}\sum_{i=1}^{h}\left|p_{T+i}\right|, \tag{1.12}$$

де $p_t = 100 \cdot e_t / y_t$ – відсоткова похибка прогнозу в момент часу t.

Недоліком MAPE є те, що відсоткові похибки, на яких вона, базується, невизначені для будь-яких значень y_t або мають екстремальні значення, коли y_t близькі до нуля. Іншою проблемою MAPE є упередженість оцінки, яка до відносної похибки більше враховує від'ємні значення, ніж додатні [2, 18].

Для оцінювання точності інтервалу прогнозування використовуються інші показники. Таким показником є оцінка Вінклера.

Нехай метод прогнозування дав інтервал $[u_{\alpha,t}, v_{\alpha,t}]$ і передбачається, що в момент часу t у цей інтервал потраплятиме значення ряду y_t із ймовірністю $100(1-\alpha)$ %. Оцінка Вінклера $W_{\alpha,t}$ тоді визначається за формулою:

$$W_{\alpha,t} = \begin{cases} (v_{\alpha,t} - u_{\alpha,t}) + \frac{2}{\alpha} \cdot \left(u_{\alpha,t} - y_t\right), & \text{якщо } y_t < u_{\alpha,t}, \\ (v_{\alpha,t} - u_{\alpha,t}), & \text{якщо } u_{\alpha,t} \le y_t \le v_{\alpha,t}, \\ (v_{\alpha,t} - u_{\alpha,t}) + \frac{2}{\alpha} \cdot \left(y_t - v_{\alpha,t}\right), & \text{якщо } y_t > v_{\alpha,t}. \end{cases} \tag{1.13}$$

Оцінка Вінклера буде довжиною інтервалу, якщо значення y_t входить у інтервал. Якщо ж y_t виходить за межі інтервалу, то до довжини інтервалу додається відстань, на яку y_t віддалене від інтервалу [2].

Для інтервалів прогнозування існують свої показники точності, які не залежать від масштабу й одиниць вимірювання значень часового ряду. Один з таких показників – skill score (укр. оцінка вправності прогнозу).

Показник skill score дозволяє обчислювати оцінку точності певного методу відносно іншого методу, який слугує орієнтиром. Якщо для оцінки точності використовується оцінка Вінклера, то skill score (SS) визначається як:

$$\mathrm{SS} = \frac{W_{ref} - W_f}{W_{ref}}, \tag{1.14}$$

де W_{ref} – оцінка Вінклера для методу-орієнтира, W_f – оцінка Вінклера для вибраного методу.

Значення SS можна інтерпретувати як пропорцію, що показує зростання точності вибраного методу порівняно з методом-орієнтиром [2, 19].

1.3.4 Перехресна перевірка прогнозування

Удосконаленим способом для навчання моделі прогнозування та обчислення точності прогнозу є перехресна перевірка. Ця техніка передбачає наявність кількох наборів тестових даних, кожен з яких містить лише одне значення.

Кожному такому набору тестових даних відповідає окремий набір навчальних даних, що містить значення часового ряду, які передують значенню з набору тестових даних. Для кожного набору начальних даних проводиться навчання моделі. Загальна оцінка точності обчислюється як середнє значення оцінок, обчислених на основі наборів тестових даних [2].

На рисунку 1.5 можна побачити набори значень часового ряду. Кожен набір представляє значення ряду відповідно до часу. Значення, які входять до набору навчальних даних, позначені світло-сірим кольором. Темно-сірі значення показують значення наборів тестових даних.

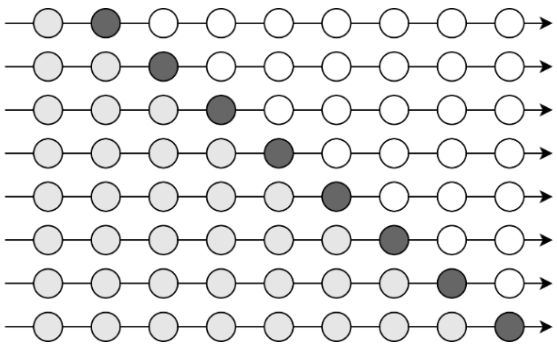

Рисунок 1.5 – Набори даних для перехресної перевірки прогнозування

Крім однокрокових прогнозів, під час проведення перехресної перевірки можна здійснювати прогнози на декілька кроків уперед [2]. Приклад такої перехресної перевірки показує рисунок 1.6.

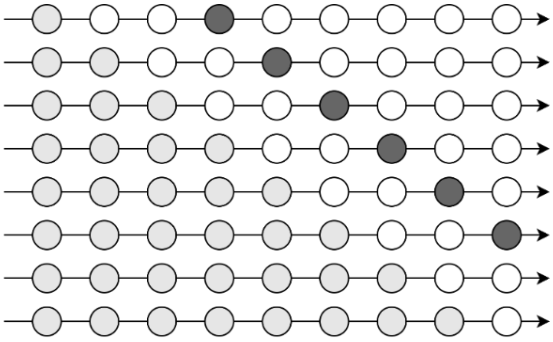

Рисунок 1.6 – Набори даних для перехресної перевірки прогнозування на 3 кроки вперед

На рисунку 1.6 видно, що набори тестових даних включають значення для оцінювання точності 3-крокових прогнозів.

1.4 Висновки до першого розділу

У першому розділі кваліфікаційної роботи освітнього рівня «Магістр» висвітлено поняття прогнозування часового ряду, проведено аналіз предметних областей, у яких використовується прогнозування часових рядів. Крім того, описано процес декомпозиції часового ряду та розглянуто шаблони часового ряду, їх характеристики. До шаблонів часового ряду належать шаблон тренду, шаблон сезонності та шаблон циклічності.

Здійснено огляд таких понять: точковий прогноз, інтервал прогнозування, підігнані значення, залишки. Висвітлено прості методи прогнозування (метод середнього, наївний метод, сезонний наївний метод, метод дрейфу), а також проаналізовано невід’ємні етапи прогнозування часових рядів: оцінювання точності прогнозів і проведення перехресної перевірки.

2 ДОСЛІДЖЕННЯ МЕТОДІВ ТА МОДЕЛЕЙ ПРОГНОЗУВАННЯ ЧАСОВИХ РЯДІВ

2.1 Методи та моделі експоненційного згладжування

2.1.1 Основні методи експоненційного згладжування

Методи експоненційного згладжування визначають прогноз на основі зваженого середнього минулих значень часового ряду так, щоб з давнішими значеннями пов'язувалися менші ваги, а з пізнішими – більші ваги. При цьому для давніших значень ваги зменшуються експоненційно [20]. Прогноз з експоненційним згладжуванням на один крок виражається записується так [2]:

$$\hat{y}_{T+1|T} = \alpha y_T + \alpha(1-\alpha)y_{T-1} + \alpha(1-\alpha)^2 y_{T-2} + \ldots, \tag{2.1}$$

де $0 \le \alpha \le 1$ – параметр згладжування, що визначає швидкість зменшення ваг.

Якщо коефіцієнт α близький до 0, то більші ваги надаються значенням ряду з минулого. Якщо ж α близький до 1, тоді більші ваги надаються останнім значенням.

Прогноз з експоненційним згладжуванням на час $T+1$ можна записати як зважене середнє останнього значення y_T та попереднього прогнозу $\hat{y}_{T|T-1}$:

$$\hat{y}_{T+1|T} = \alpha y_T + (1-\alpha)\hat{y}_{T|T-1}, \tag{2.2}$$

де $0 \le \alpha \le 1$ – параметр згладжування.

Нехай на момент часу 1 визначено перше підігнане значення l_0. Воно і буде вважатися першим прогнозом $\hat{y}_{1|0}$. Якщо на основі (2.2) виразити прогноз на час $T+1$ через усі передні значень ряду, то буде:

$$\hat{y}_{T+1|T} = \sum_{i=0}^{T-1} \alpha(1-\alpha)^i y_{T-i} + (1-\alpha)^T l_0, \tag{2.3}$$

У часовому ряді експоненційне згладжування може розглядати такі компоненти: рівень (згладжене значення), тренд і сезонний компонент. Поки що розглядається просте експоненційне згладжування, яке включає лише компонент рівня та підходить для рядів без вираженого шаблону тренду або сезонності.

Метод простого експоненційного згладжування представляється у вигляді двох рівнянь: рівняння прогнозу і рівняння згладжування для рівня (або рівняння рівня) [21]. Рівняння прогнозу записується як:

$$\hat{y}_{t+h|t} = l_t, \tag{2.4}$$

а рівняння рівня:

$$l_t = \alpha y_t + (1-\alpha) l_{t-1}, \tag{2.5}$$

де l_t – оцінка рівня на момент часу t, $0 \le \alpha \le 1$ – параметр згладжування рівня.

За рівнянням прогнозу значення прогнозу на час $t+1$ дорівнює рівню на момент часу t. Рівняння рівня визначає значення рівня на кожен момент часу t. Для використання методу простого експоненційного згладжування необхідно вибрати параметр α та початкове значення l_0.

Розширенням методу простого експоненційного згладжування, що враховує компонент тренду, є метод лінійного (адитивного) тренду Голта. Цей метод використовує одне рівняння прогнозу та два рівняння згладжування: для рівня і для тренду [2]. Рівняння методу лінійного тренду Голта:

- рівняння прогнозу: $\hat{y}_{t+h|t} = l_t + hb_t$,
- рівняння рівня: $l_t = \alpha y_t + (1-\alpha)(l_{t-1} + b_{t-1})$,

- рівняння тренду: $b_t = \beta(l_t - l_{t-1}) + (1-\beta)b_{t-1}$,

де b_t – оцінка тренду ряду на момент часу t, $0 \le \beta \le 1$ – параметр згладжування тренду, $0 \le \alpha \le 1$.

Рівняння тренду показує, що оцінка тренду дорівнює зваженому середньому оціненого тренду $l_t - l_{t-1}$ і попередньої оцінки тренду b_{t-1}. Для методу лінійного тренду Голта, крім α і l_0, потрібно встановити b_0 й параметр β.

Варіацією методу лінійного тренду Голта є метод демпфованого тренду, який використовує параметр згасання тренду для «заглушення» тренду за деякий час у майбутньому [21].

Для включення сезонного компонента розроблено метод Голта-Вінтерса, який до рівнянь згладжування компонентів додає рівняння сезонного компонента [2]. Розрізняється адитивний і мультиплікативний метод Голта-Вінтерса. Адитивний метод відображає сезонний компонент в абсолютних величинах та підходить тоді, коли коливання у часовому ряді є відносно постійними. Мультиплікативний метод відображає сезонний компонент у відносних величинах і є кращим, коли коливання змінюються з часом.

Для адитивного методу Голта-Вінтерса рівняння компонентів:

- рівняння прогнозу: $\hat{y}_{t+h|t} = l_t + hb_t + s_{t+h-m(k+1)}$,
- рівняння рівня: $l_t = \alpha(y_t - s_{t-m}) + (1-\alpha)(l_{t-1} + b_{t-1})$,
- рівняння тренду: $b_t = \beta(l_t - l_{t-1}) + (1-\beta)b_{t-1}$,
- рівняння сезонного компонента: $s_t = \gamma(y_t - l_{t-1} - b_{t-1}) + (1-\gamma)s_{t-m}$,

де s_t – оцінка сезонного компонента ряду на момент часу t, m – сезонний період, k – ціла частина від $(h-1)/m$, $0 \le \gamma \le 1-\alpha$ – параметр сезонного компонента, $0 \le \alpha \le 1$, $0 \le \beta \le 1$.

Тут рівняння сезонного компонента визначає сезонний компонент як зважене середнє поточного значення сезонного компонента (без рівня та тренду) $y_t - l_{t-1} - b_{t-1}$ і значення сезонного компонента в минулому сезоні s_{t-m}.

Для вибору методу експоненційного згладжування потрібно враховувати всі комбінації методів для тренду й сезонного компонента [2]. Кожен метод можна позначити як (t, s), де t – тренд, а s – сезонний компонент. На рисунку 2.1 показано структуру можливих варіантів компонентів.

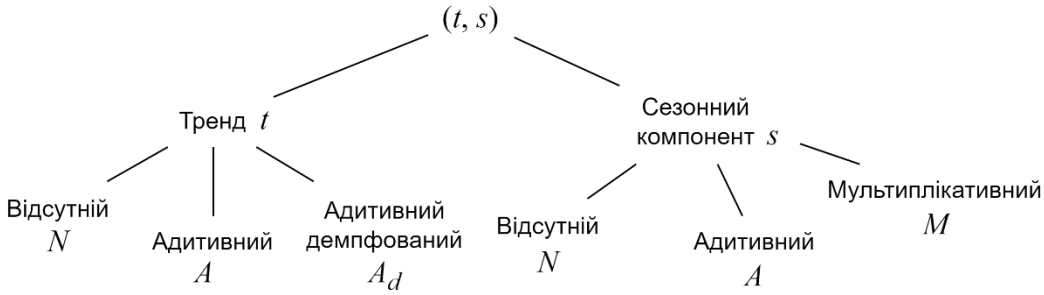

Рисунок 2.1 – Можливі варіанти компонентів
для методу експоненційного згладжування

На рисунку видно, що в методі експоненційного згладжування тренд може бути відсутнім (N), адитивним (A) або адитивним демпфованим (A_d).Варіантами сезонного компонента є: відсутній (N), адитивний (A) і мультиплікативний (A).

2.1.2 Моделі ETS

На основі методів експоненційного згладжування створені статистичні моделі, які дозволяють генерувати розподіли прогнозів, а зокрема надати інтервали прогнозування. Модель експоненційного згладжування складається з рівняння спостереження, яке описує дані часового ряду, та рівнянь стану, що описує зміну компонентів (або станів): рівня, тренду, сезонного компонента. Тому такі моделі називаються моделями простору станів [2].

Для кожного методу експоненційного згладжування існують дві моделі, які використовують або адитивні, або мультиплікативні похибки. Так, моделі простору станів називаються також моделями ETS (англ. error – похибка, trend – тренд, seasonal – сезонний компонент) і позначаються як ETS(e, t, s), де e – похибки, t – тренд, а s – сезонний компонент [22].

Можливі варіанти моделі подані на рисунку 2.2.

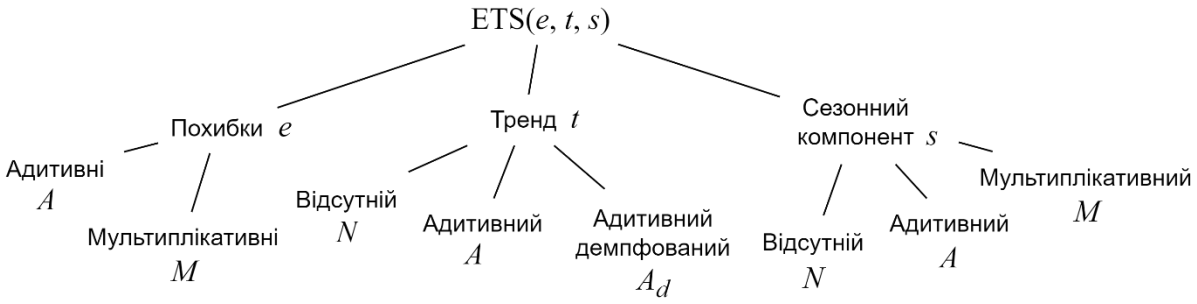

Рисунок 2.2 – Можливі варіанти моделі ETS

Модель ETS(*A*, *N*, *N*) позначає модель простого експоненційного згладжування з адитивними похибками, яка не використовує рівнянь тренду чи сезонного компонента. Цю модель можна описати рівняннями [22]:

- рівняння спостереження: $y_t = l_{t-1} + \varepsilon_t$;
- рівняння стану: $l_t = l_{t-1} + \alpha\varepsilon_t$.

де ε_t – залишок на момент часу t, який є білим шумом з нормальним розподілом, який має середнє значення 0, дисперсію σ^2.

Модель простого експоненційного згладжування з мультиплікативними похибками ETS(*M*, *N*, *N*) також можна описати цими рівняннями [22]:

- рівняння спостереження: $y_t = l_{t-1} + \varepsilon_t$;
- рівняння стану: $l_t = l_{t-1} + \alpha\varepsilon_t$.

Проте вона використовує відносні похибки:

$$\varepsilon_t = \frac{y_t - \hat{y}_{t|t-1}}{\hat{y}_{t|t-1}} \tag{2.6}$$

де ε_t – білий шум, що має нормальний розподіл із середнім значенням 0 та дисперсією σ^2.

Для опису моделі ETS(*M*, *N*, *N*) використовуються рівняння:

- рівняння спостереження: $y_t = l_{t-1} + l_{t-1}\varepsilon_t$ або $y_t = l_{t-1}(1+\varepsilon_t)$;
- рівняння стану: $l_t = l_{t-1} + \alpha l_{t-1}\varepsilon_t$ або $l_t = l_{t-1}(1+\alpha\varepsilon_t)$.

Так само можна описати й інші моделі ETS на основі всіх можливих методів експоненційного згладжування.

2.1.3 Прогнозування за допомогою моделей ETS з урахуванням циклічності часових рядів

Для прогнозування часового ряду за допомогою експоненційного згладжування використовується модель ETS. Вибір найкращої моделі та оцінка параметрів згладжування, початкових значень станів пов'язані з максимальним значенням правдоподібності, що можна назвати ймовірністю того, як добре статистична модель відповідає заданій вибірці даних.

Точкові прогнози моделей ETS отримуються з рівнянь спостереження для $t = T+1,\ldots,T+h$. Також встановлюється $\varepsilon_t = 0$ для всіх $t > T$. Оскільки для моделей ETS точкові прогнози є середніми значеннями розподілу ймовірностей прогнозу, то інтервали прогнозування обчислюються за (1.3). Знайти середнє квадратичне відхилення прогнозу σ_h можна за допомогою дисперсії прогнозу σ_h^2, яка обчислюється залежно від самої моделі ETS [2].

Моделі ETS враховують сезонний компонент часових рядів і добре визначають сезонність. Однак вони не передбачають циклічності. Жодна така модель не може відтворити аперіодичні цикли. Тому моделі ETS не підходять для прогнозування часових рядів з вираженою циклічністю [23].

2.2 Регресійні моделі

2.2.1 Лінійна регресія

Суть регресійних моделей полягає в тому, що для прогнозування часового ряду (змінної) використовується залежність від інших часових рядів (змінних). Лінійна регресія визначає залежність за лінійною моделлю і має формулу:

$$y_t = \beta_0 + \beta_1 x_{1,t} + \beta_2 x_{2,t} + \ldots + \beta_k x_{k,t} + \varepsilon_t, \tag{2.7}$$

де y_t – значення залежної змінної y на момент часу t, $x_{1,t}, x_{2,t}, \ldots, x_{k,t}$ – значення незалежних змінних $x_1, x_2, \ldots, x_k$ на момент часу t.

У (2.7) коефіцієнт β_0 відображає прогнозоване значення залежної змінної y, коли всі незалежні змінні $x_1, x_2, \ldots, x_k$ дорівнюють нулю. Коефіцієнти $\beta_1, \beta_2, \ldots, \beta_k$ визначають вплив певної незалежної змінної на залежну змінну після врахування впливу всіх інших незалежних змінних. Похибка ε_t є відхиленням значень змінної y від лінійної моделі [24].

Для отримання робочої регресійної моделі необхідно на основі даних оцінити коефіцієнти $\beta_0, \beta_1, \beta_2, \ldots, \beta_k$. Зробити це можна шляхом мінімізації суми квадратів похибок [25]. Тоді потрібно мінімізувати:

$$\sum_{t=1}^{T} \varepsilon_t^2 = \sum_{t=1}^{T} (y_t - \beta_0 - \beta_1 x_{1,t} - \beta_2 x_{2,t} - \ldots - \beta_k x_{k,t})^2, \tag{2.8}$$

Процес оцінки коефіцієнтів, при яких сума квадратів похибок буде мінімальною, називається навчанням моделі. У результаті навчання отримуються оцінені коефіцієнти, які можна відставити в рівняння регресії, щоб отримати підігнані значення [2]. Це записується так:

$$\hat{y}_t = \hat{\beta}_0 + \hat{\beta}_1 x_{1,t} + \hat{\beta}_2 x_{2,t} + \ldots + \hat{\beta}_k x_{k,t}, \tag{2.9}$$

де $\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, \ldots, \hat{\beta}_k$ – оцінені коефіцієнти.

2.2.2 Нелінійна регресія

У деяких випадках потрібно використовувати нелінійну регресію. Найпростішим способом побудови нелінійної регресії є перетворення

незалежної змінної на певну нелінійну функцію. Хоч тоді регресія буде функціонально нелінійною, проте її модель матиме лінійні параметри [2].

У загальному випадку модель нелінійної регресії записується формулою:

$$y = f(x) + \varepsilon, \tag{2.10}$$

де f – нелінійна функція змінної x.

Функція f може бути квадратичною, логарифмічною тощо. Одним з варіантів вираження f є кусково-лінійна функція. Для кусково-лінійної функції визначаються точки, які називаються вузлами. Між вузлами функція представлена у вигляді прямих, у самих вузлах нахил цих прямих змінюється.

Для вираження кусково-лінійної функції вводиться змінна $x_1 = x$. Також для кожного вузла додається змінна x_i, $i = \overline{2,k}$, так що:

$$x_i = \begin{cases} x - c_i, & \text{якщо } x \geq c_i, \\ 0, & \text{якщо } x < c_i, \end{cases} \tag{2.11}$$

де c_i – вузли.

Найважливішим у побудові кусково-лінійної функції є вибір кількості вузлів та їх розташування. Такий вибір часто є складним і загалом дещо довільним. Хоч алгоритми для вибору вузлів й існують, проте вони не часто використовуються. Зазвичай вибір вузлів проводиться інтуїтивним способом на основі аналізу наявних даних [26].

2.2.3 Оцінка регресійних моделей

Після встановлення незалежних змінних і навчання регресійної моделі необхідно визначити, чи ця модель достатньо наближена до реальних даних часового ряду, тобто здійснити оцінку регресійної моделі.

В оцінці моделі вагоме значення має аналіз залишків [25]. Для регресійної моделі залишки на момент часу t обчислюються наступним чином:

$$e_t = y_t - \hat{y}_t = y_t - \hat{\beta}_0 - \hat{\beta}_1 x_{1,t} - \hat{\beta}_2 x_{2,t} - \ldots - \hat{\beta}_k x_{k,t}. \qquad (2.12)$$

Залишки мають дві основні властивості [2, 27]:

- середнє значення залишків дорівнює нулю, тобто $\sum_{t=1}^{T} e_t = 0$;
- кореляція між залишками та значеннями незалежних змінних дорівнює нулю, таким чином, $\sum_{t=1}^{T} x_{k,t} e_t = 0$;

Виконання цих властивостей дозволяє визначити, чи навчена модель відповідає реальним даним. Для їх перевірки будуються й розбираються різні графіки на основі залишків.

Ще однією із властивостей залишків є відсутність їх автокореляції. Якщо в залишків моделі присутня певна автокореляція, тоді існує певна інформація, яку варто врахувати в моделі, щоб покращити прогнози. Для перевірки автокореляції залишків можна скористатися графіком автокореляційної функції. Екстремально великі значення на графіку вказують на сплески автокореляції, проте невелика кількість таких сплесків не вказує на значний вплив автокореляції на прогнози.

Для перевірки кореляції між залишками і значеннями незалежних змінних необхідно побудувати й дослідити графіки розсіювання між залишками та кожною з цих змінних. Наявність певної закономірності на діаграмах свідчитиме про нелінійний зв’язок.

Важливою також є відсутність кореляції між залишками й підігнаними значеннями. Якщо діаграма розсіювання показує певні закономірності, тоді в залишках може бути гетероскедастичність, якої потрібно уникати.

Також при оцінці моделі варто перевірити, чи залишки мають нормальний розподіл. Для цього потрібно скористатися діаграмою розподілу залишків [2].

2.2.4 Прогнозування за допомогою регресійних моделей з урахуванням циклічності часових рядів

Прогнози часових рядів при використанні регресійних моделей можна розділити на два типи: прогнози ex-ante і прогнози ex-post [2].

Прогнози ex-ante (укр. перед подією) – прогнози, зроблені на основі тих даних часового ряду, які є на момент прогнозування. Тому для знаходження цих прогнозів використовуються прогнози незалежних змінних, які отримуються різними способами. Так, прогнози незалежних змінних можна обчислити за допомогою різних моделей і методів прогнозування. Також ці прогнози можна отримати із зовнішніх джерел, наприклад урядових установ чи організацій, що спеціалізуються на прогнозуванні.

Прогнози ex-post (укр. після події) – прогнози, зроблені на основі пізніших даних часового ряду. Для прогнозів ex-post використовуються фактичні значення незалежних змінних, а прогнози не робляться на певний час у майбутньому.

Загалом прогнози ex-post не є справжніми прогнозами, проте вони важливі для аналізу поведінки створених регресійних моделей. Прогнози ex-ante, натомість, є справжніми, зробленими на майбутнє.

Особливістю регресійних моделей є те, що вони дозволяють визначати взаємозв’язки між прогнозованою змінною та незалежними змінними. Проте важливим для прогнозів регресійних моделей є також майбутні значення всіх незалежних змінних, які можна замінити відповідними прогнозами.

Якщо розглядати циклічні часові ряди, то регресійні моделі для прогнозування змінної (часового ряду) потрібні циклічні незалежні змінні (часові ряди), які мають зв’язок із прогнозованою змінною. Також слід зауважити, що для прогнозування значень циклічної змінної у майбутньому потрібні прогнози циклічних незалежних змінних.

Окремо варто зазначити використання кусково-лінійної регресії для прогнозування циклічних часових рядів, адже циклічна поведінка ряду добре описується прямими, які змінюють свій нахил на вузлах. Для обчислення

прогнозів за допомогою моделей кусково-лінійної регресії використовуються прогнози значень таких вузлів.

Для того, щоб визначити інтервал прогнозування за допомогою простої регресійної моделі з однією незалежною змінною x, потрібно знайти його межі за формулою [2]:

$$\hat{y} \pm c \cdot \hat{\sigma}_e \cdot \sqrt{1 + \frac{1}{T} + \frac{(x - \overline{x})^2}{(T-1) \cdot s_x^2}}, \tag{2.13}$$

де $\hat{y}$ – прогноз, c – множник, який залежить від ймовірності попадання прогнозу в інтервал прогнозування, $\hat{\sigma}_e$ – оцінка середньоквадратичного відхилення залишків, T – кількість значень ряду, $\overline{x}$ – середнє значення змінної x, s_x^2 – середньоквадратичне відхилення значень змінної x.

Оцінка середньоквадратичного відхилення залишків $\hat{\sigma}_e$ обчислюється наступним чином:

$$\hat{\sigma}_e = \sqrt{\frac{1}{T-k-1} \sum_{t=1}^{T} e_t^2}, \tag{2.14}$$

де T – кількість значень часового ряду, k – кількість незалежних змінних.

2.3 Моделі ARIMA

2.3.1 Авторегресійні моделі та моделі ковзного середнього

Авторегресійні моделі дозволяють прогнозувати значення змінної за допомогою лінійної комбінації минулих значень змінної. Залежно від кількості минулих значень, які використовуються в лінійній комбінації, авторегресійна модель має певний порядок. Так, авторегресійна модель поряду *p* записується як AR(*p*) (англ. autoregression) і обчислюється за формулою [2, 28]:

$$y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \ldots + \phi_p y_{t-p} + \varepsilon_t, \tag{2.15}$$

де c – константа, ε_t – білий шум, $\phi_1, \phi_2, \ldots, \phi_p$ – коефіцієнти, які дозволяють визначати різні шаблони часового ряду.

Головним обмеженням авторегресійних моделей є стаціонарність даних. Обмеження також стосуються параметрів моделі. Вони зазвичай залежать від її порядку.

На відміну від авторегресійних моделей, моделі ковзного середнього використовують значення минулих похибок [29]. Модель ковзного середнього порядку q записується як MA(q) (англ. moving average). Формулу, яка визначає модель ковзного середнього, подано нижче [2]:

$$y_t = c + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \ldots + \theta_q \varepsilon_{t-q}, \tag{2.16}$$

де c – константа, ε_t – білий шум, $\theta_1, \theta_2, \ldots, \theta_q$ – коефіцієнти для визначення різних шаблонів часового ряду.

Для само, як і авторегресійні моделі, моделі ковзного середнього мають обмеження щодо стаціонарності даних і значень параметрів.

2.3.2 Властивості моделей ARIMA

Якщо об’єднати авторегресійну модель (AR) та модель ковзного середнього (MA), то можна отримати модель ARIMA (англ. autoregressive integrated moving average), тобто модель авторегресійного інтегрованого ковзного середнього. «Інтегроване» означає, що для обчислення прогнозів часових рядів використовуються їх різниці для досягнення стаціонарності даних, що є обмеженням складової авторегресійної моделі та моделі ковзного середнього [30].

Формально модель ARIMA можна записати формулою:

$$\Delta^d y_t = c + \phi_1 \Delta^d y_{t-1} + \ldots + \phi_p \Delta^d y_{t-p} + \theta_1 \varepsilon_{t-1} + \ldots + \theta_q \varepsilon_{t-q} + \varepsilon_t, \qquad (2.17)$$

де $\Delta^d y_t$ – різниця значення змінної y_t порядку *d*, різниця першого порядку записується як $\Delta y_t = y_t - y_{t-1}$.

Загалом модель записується як ARIMA(*p*, *d*, *q*), де:

- *p* – порядок авторегресійної частини;
- *d* – порядок різниці;
- *q* – порядок частини ковзного середнього.

Вибір значень *p*, *d* і *q* є досить складним [2]. Параметри *p* і *q* мають таке ж значення, як в авторегресійних моделях і моделях ковзного середнього. Також важливими є значення параметра *d* і константи c. Різні їх комбінації впливають на те, як змінюватимуться значення прогнозів у майбутньому.

Якщо $c = 0$ $d = 0$, то далекі прогнози будуть наближатися до нуля, якщо $c = 0$ і $d = 1$ – далекі прогнози наближатимуться до певного ненульового сталого значення. Якщо $c = 0$ і $d = 2$ – то далекі прогнози будуть іти по прямій. Якщо розглянути $c \neq 0$, тоді при $d = 0$ далекі прогнози будуть наближатися до середнього значення даних часового ряду, при $d = 1$ далекі прогнози будуть іти по прямій, при $d = 2$ далекі прогнози будуть іти за квадратичним трендом [31].

Значення *d* також впливає на розміри інтервалів прогнозування. Чим більшим буде *d*, тим швидше зростатимуть інтервали прогнозування, а при $d = 0$ інтервали матимуть однаковий розмір.

Для аналізу значень параметрів моделі ARIMA можна використовувати графіки автокореляційної функції і часткової автокореляційної функції часового ряду [31]. Автокореляція визначає зв’язок між значеннями y_t та y_{t-k} для будь-якого k. Однак цей зв’язок може бути хибним через те, наприклад, що він випливає з пов’язаних проміжних значень між y_t і y_{t-k}. Для усунення впливу пов’язаних значень і застосовується часткова автокореляція.

Щоб отримати модель ARIMA з $p = 0$ і $q = q_0$, потрібні такі умови:

- часткова автокореляційна функція є синусоїдою або експоненційно спадає;
- автокореляційна функція має пік у точці q_0, а після q_0 таких піків немає.

Щоб для моделі ARIMA визначити $p = p_0$ і $q = 0$, необхідні умови:

- автокореляційна функція є синусоїдою або експоненційно спадає;
- часткова автокореляційна функція має пік у точці p_0, а неї піків немає

Після визначення параметрів *p*, *d* та *q* необхідно оцінити коефіцієнти $\phi_1, \phi_2, \ldots, \phi_p$ і $\theta_1, \theta_2, \ldots, \theta_q$. Це можна зробити шляхом вибору з найкращої моделі з вибраними коефіцієнтами. Вибір моделі проводить на основі знаходження максимального значення правдоподібності, так само, як і для моделей ETS.

2.3.3 Основні характеристики сезонних моделей ARIMA

Моделі ARIMA також можуть використовуватися в моделюванні та прогнозуванні сезонних даних. Для цього до моделі включається сезонна авторегресійна частина і сезонна частина ковзного середнього. Така модель називається сезонною моделлю ARIMA або SARIMA (англ. seasonal ARIMA). Сезонна модель ARIMA (SARIMA) позначається як ARIMA(p, d, q)(P, D, Q)$_m$, де додаткові параметри:

- P – порядок сезонної авторегресійної частини;
- D – порядок сезонної різниці;
- Q – порядок сезонної частини ковзного середнього;
- m – сезонний період.

Додаткові сезонні частини дозволяють моделі враховувати сезонну поведінку часового ряду при прогнозуванні. Сезонна різниця призначена для забезпечення стаціонарності даних, кратних відлікам у сезонний період, тобто однакових даних у різних сезонах [2, 32].

У формулі моделі SARIMA сезонні частини перемножуються на відповідні несезонні частини. Сезонна авторегресійна частина записується так:

$$\Phi_1 \Delta^D y_{t-m} + \Phi_2 \Delta^D y_{t-2m} + \ldots + \Phi_P \Delta^D y_{t-P \cdot m}, \quad (2.18)$$

де $\Phi_1, \Phi_2, \ldots, \Phi_P$ – коефіцієнти сезонної авторегресійної частини, $\Delta^D y_t$ – сезонна різниця значення змінної y_t.

Формула сезонної частини ковзного середнього:

$$\Theta_1 \varepsilon_{t-m} + \Theta_2 \varepsilon_{t-2m} + \ldots + \Theta_Q \varepsilon_{t-Q \cdot m}, \quad (2.19),$$

де $\Theta_1, \Theta_2, \ldots, \Theta_Q$ – коефіцієнти сезонної частини ковзного середнього, ε_t – білий шум.

2.3.4 Прогнозування за допомогою моделей ARIMA та SARIMA з урахуванням циклічності часових рядів

Для отримання точкових прогнозів на основі моделей ARIMA та SARIMA необхідно виконувати наступний порядок дій [2]:

1) розписати рівняння моделі так, щоб у лівій частині містилося значення y_t, а решта – у правій частині;

2) у рівнянні замінити t на $T + h$;

3) у правій частині рівняння замість майбутніх значень записати їх прогнози, замість майбутніх похибок – нулі, замість минулих похибок – відповідні залишки;

4) виконувати попередні дії: починати з $h = 1$, а потім повторювати кроки для $h = 2, 3, \ldots$, поки не будуть обчислені всі прогнози.

Порівняно з моделями ETS моделі ARIMA можуть працювати з часовими рядами із властивістю циклічності. При прогнозуванні циклічних рядів за допомогою моделей ARIMA важливим є параметр *p*. Основною умовою для роботи з циклічними часовими рядами є значення параметра $p \geq 2$ [2].

Інтервали прогнозування для моделей ARIMA та SARIMA обчислюються на основі точкових прогнозів за формулою (1.3). Визначення середнього квадратичного відхилення розподілу прогнозу $\hat{\sigma}_h$ є досить складним, і формули його обчислення можуть відрізнятися залежно від параметрів моделі. Саме знаходження інтервалів прогнозування базується на припущеннях, що залишки мають нормальний розподіл і не є корельованими.

Для моделей ARIMA інтервали прогнозування зростають зі збільшенням горизонту прогнозування. При $d = 0$ інтервали прогнозування будуть сходитися, так що для довгих прогнозів вони є по суті однаковими. При $d \geq 1$ інтервали прогнозування будуть продовжувати зростати для подальших прогнозів.

2.4 Комплексні моделі прогнозування та моделі штучних нейронних мереж

Раніше розглядалися найпопулярніші моделі прогнозування часових рядів: моделі експоненційного згладжування ETS, регресійні моделі та моделі ARIMA. Далі йде короткий огляд комплексних моделей, які базуються на попередніх, а саме динамічні регресійні моделі та векторні авторегресійні моделі. Крім цього, здійснюється огляд моделей штучних нейронних мереж.

2.4.1 Динамічні регресійні моделі

Динамічні регресійні моделі дозволяють при прогнозуванні враховувати інформацію з інших змінних, як і звичайні регресійні моделі. Разом із тим, вони враховують динаміку часових рядів, як моделі ARIMA. Динамічна лінійна регресійна модель, по суті, є звичайною лінійною регресійною моделлю, похибки якої визначаються відповідно до моделі ARIMA.

Динамічна лінійна регресійна модель описується на формулою [33]:

$$y_t = \beta_0 + \beta_1 x_{1,t} + \beta_2 x_{2,t} + \ldots + \beta_k x_{k,t} + \eta_t \qquad (2.20)$$

де y_t – значення залежної змінної, $x_{1,t}, x_{2,t}, \ldots, x_{k,t}$ – значення незалежних змінних, $\beta_0, \beta_1, \beta_2, \ldots, \beta_k$ – коефіцієнти регресійної моделі, η_t – значення похибок регресійної моделі.

Похибки регресійної моделі визначаються за моделлю ARIMA так:

$$\Delta^d \eta_t = c + \phi_1 \Delta^d \eta_{t-1} + \ldots + \phi_p \Delta^d \eta_{t-p} + \theta_1 \varepsilon_{t-1} + \ldots + \theta_q \varepsilon_{t-q} + \varepsilon_t \qquad (2.21)$$

де c – константа, $\phi_1, \phi_2, \ldots, \phi_p$, $\theta_1, \theta_2, \ldots, \theta_q$ – коефіцієнти моделі ARIMA, ε_t – похибки моделі ARIMA, які є білим шумом.

Важливою умовою динамічної регресійної моделі є стаціонарність змінних. Якщо якісь зі змінних нестаціонарні, то оцінені коефіцієнти будуть неправильними. Для досягнення стаціонарності змінних до них застосовуються різниці.

Результати прогнозування динамічних регресійний моделей отримуються шляхом об’єднання прогнозів регресійної моделі та моделі ARIMA. Інтервали прогнозування для динамічних регресійний моделей, як і для звичайних регресійних моделей, не враховують невизначеність у прогнозах незалежних змінних. Тому ці інтервали потрібно вважати залежними від оцінених майбутніх значень незалежних змінних [2, 33].

При прогнозуванні циклічних часових рядів динамічні регресійні моделі мають такі самі особливості, як звичайні регресійні моделі: прогнозована змінна (числовий ряд) залежить від пов’язаних циклічних незалежних змінних (рядів).

2.4.2 Векторні авторегресійні моделі

Суттєвим обмеженням регресійних моделей є те, що вони враховують лише односторонній зв’язок – на залежну змінну впливають незалежні змінні і не навпаки. Для опису двостороннього зв’язку двох або більше змінних використовуються векторні авторегресійні моделі. У цій моделі змінні визначаються так, ніби всі однаково впливають одна на одну.

Векторна авторегресійна модель поряду *p* записується як VAR(*p*) (англ. – vector autoregression). Якщо розглянути дві змінні y_1 та y_2, де y_1 впливає на y_2, а y_2 впливає на y_1, то модель VAR(*p*) описується двома рівняннями [34]:

$$\begin{aligned} y_{1,t} &= c_1 + \phi_{11,1} y_{1,t-1} + \ldots + \phi_{11,p} y_{1,t-p} + \phi_{12,1} y_{2,t-1} + \ldots + \phi_{12,p} y_{2,t-p} + \varepsilon_{1,t}, \\ y_{2,t} &= c_2 + \phi_{21,1} y_{1,t-1} + \ldots + \phi_{21,p} y_{1,t-p} + \phi_{22,1} y_{2,t-1} + \ldots + \phi_{22,p} y_{2,t-p} + \varepsilon_{2,t}, \end{aligned} \tag{2.22}$$

де c_1 і c_2 – константи, $y_{1,t}$ і $y_{2,t}$ – значення змінних y_1 і y_2, $\varepsilon_{1,t}$ і $\varepsilon_{2,t}$ – значення похибок, які є білим шумом, коефіцієнт $\phi_{ij,k}$ визначає вплив змінної y_j на змінну y_i у відліку k.

Так само, як моделі ARIMA та динамічні регресійні моделі, векторні авторегресійні моделі мають одну умову щодо даних – стаціонарність. Для усунення нестаціонарності даних у моделі можуть використовуватися різниці.

Векторні авторегресійні моделі можуть прогнозувати циклічні часові ряди, так само, як моделі ARIMA та авторегресійні моделі, які є частиною моделей ARIMA. Найбільше векторні авторегресійні моделі підходять для прогнозування кількох циклічних часових рядів, яка є взаємопов’язані між собою.

Векторна авторегресійна модель надає прогнози для всіх включених змінних. Обчислення прогнозу на крок h використовує оцінені коефіцієнти. Для моделі з двома змінними прогнози обчислюються за формулами [2]:

$$\begin{aligned} \hat{y}_{1,T+h|T} &= \hat{c}_1 + \hat{\phi}_{11,1} y_{1,T-1} + \ldots + \hat{\phi}_{11,p} y_{1,T-p} + \hat{\phi}_{12,1} y_{2,T-1} + \ldots + \hat{\phi}_{12,p} y_{2,T-p}, \\ \hat{y}_{2,T+h|T} &= \hat{c}_2 + \hat{\phi}_{21,1} y_{1,T-1} + \ldots + \hat{\phi}_{21,p} y_{1,T-p} + \hat{\phi}_{22,1} y_{2,T-1} + \ldots + \hat{\phi}_{22,p} y_{2,T-p}. \end{aligned} \tag{2.23}$$

Знаходження прогнозів відбувається по поряду для всіх значень h. Спочатку обчислюється прогноз для $h = 1$. Потім на основі нього обчислюється прогноз для $h = 2$ і так далі.

2.4.3 Моделі штучних нейронних мереж

Штучними нейронними мережами називаються обчислювальні системи, які базуються на мережах розміщення нейронів у мозку. Вони дозволяють будувати моделі прогнозування, які використовують складні нелінійні зв'язки між незалежними змінними та залежною змінною [2].

Нейронна мережа складається зі штучних нейронів, які поділені на шари: шар входу і виходу, приховані шари. Загальну структуру такої нейронної мережі подано на рисунку 2.3.

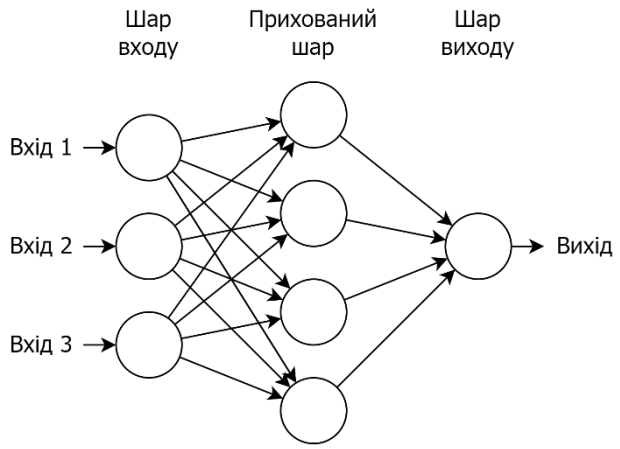

Рисунок 2.3 – Штучна нейронна мережа з трьома вхідними нейронами, одним прихованим шаром та одним вихідним нейроном

Вхідні нейрони відповідають незалежним змінним, якщо будувати модель нейронної мережі на основі регресії. Якщо ж використовується авторегресія, то вхідними нейронами є минулі відліки змінної. Вихідним нейроном є залежна змінна, значення якої й отримуються з моделі. Нейрони у прихованих шарах забезпечують нелінійність мережі.

У багатошаровій мережі нейрони кожного рівня отримують вхідні дані від нейронів попередніх рівнів. Для кожного нейрона вхідні дані об'єднуються за допомогою лінійної комбінації вихідних даних нейронів попереднього шару. Для цієї лінійної комбінації використовуються вагові коефіцієнти. На виході

нейронів прихованих шарів до даних застосовується нелінійна функція, що зменшує вплив екстремальних значень [35].

Вагові коефіцієнти вибираються шляхом навчання моделі нейронної мережі, що полягає у мінімізації функції витрат. Спочатку для вагових коефіцієнтів використовуються випадкові значення. Тоді у ході навчання ці значення оновлюються. Модель нейронної мережі навчається так кілька разів з використанням різних випадкових початкових значень вагових коефіцієнтів [2].

Використання нейронної мережі та прогнозування часового проводиться ітеративно. Спочатку на основі наявних даних обчислюється прогноз на крок уперед. Потім на основі цього прогнозу і попередніх даних обчислюється прогноз на два кроки вперед і так далі. Слід зазначити, що моделі нейронних мереж можуть добре прогнозувати циклічні часові ряди, так само, як моделі ARIMA та регресійні моделі.

2.5 Висновки до другого розділу

У другому розділі кваліфікаційної роботи проведено огляд моделей прогнозування часових рядів, які найчастіше використовуються, та аналіз їх можливостей прогнозування циклічних рядів. До цих моделей належать моделі експоненційного згладжування ETS, регресійні моделі та моделі ARIMA.

Моделі ETS не можуть визначати циклічність рядів і відтворювати аперіодичні цикли. Тому вони не підходять до прогнозування циклічних часових рядів. Регресійні моделі, натомість, дозволяють таке прогнозування. Проте для цього використовуються значення інших пов’язаних циклічних рядів. Моделі ARIMA також підтримують циклічність часових рядів, для прогнозування яких вони встановлюють обмеження значень параметра p.

Крім цього, також розглянуто комплексні моделі прогнозування часових рядів, а саме динамічні регресійні моделі, векторні авторегресійні моделі, та моделі штучних нейронних мереж.

3 РОЗРОБКА ПРОГРАМНОГО ЗАСОБУ ДЛЯ ПРОГНОЗУВАННЯ ЦИКЛІЧНИХ ЧАСОВИХ РЯДІВ ТА АНАЛІЗ ЕКСПЕРИМЕНТАЛЬНИХ РЕЗУЛЬТАТІВ

3.1 Постановка завдання розробки програмного засобу прогнозування циклічних часових рядів

На основі аналізу розглянутих методів і моделей визначається завдання щодо розробки програмного засобу прогнозування часових рядів із циклічністю.

Результатом розробки є програмний засіб, що допомагає проводити аналіз прогнозування циклічних часових рядів на основі різних методів і моделей. Програмний засіб може використовуватися для подальшого дослідження методів і моделей прогнозування часових рядів, які можуть враховувати їх циклічність.

Так, завдання полягає в розробці програмного засобу прогнозування циклічних часових рядів за допомого різних методів і моделей. Робота програмного засобу описується наступним чином: спочатку приймається файл, у якому записано часовий ряд, також приймається горизонт прогнозування, можливі значення параметрів для методів і моделей прогнозування.

Далі програмний засіб обчислює точкові прогнози отриманого часового ряду для кожного методу й моделі. Отриманий горизонт прогнозування визначається в межах часового проміжку ряду так, щоб було можна оцінювати прогнози та порівнювати їх з реальними значеннями часового ряду.

Обчислені прогнози відображаються на графіках разом зі значеннями часового ряду. Після обчислення прогнозів програмний засіб визначає похибки для оцінювання й порівняння між собою методів та моделей.

Методи та моделі прогнозування часових рядів, які будуть реалізовані у програмному засобі, вибираються серед методів і моделей, розглянутих у другому розділі. При виборі враховуються можливості прогнозування циклічних часових рядів та пов’язані з цим особливості.

3.2 Обґрунтування вибору методів та моделей прогнозування циклічних часових рядів

Для вибору методів та моделей прогнозування розглядаються наступні: моделі ETS, що використовують експоненційне згладжування, регресійні моделі, моделі ARIMA, комплексні моделі, моделі машинного навчання. Серед цих методів і моделей для реалізації в програмному засобі вибрано модель ARIMA, моделі XGBoost та LightGBM, які описуються далі.

3.2.1 Модель ARIMA

Якщо розглядати основні методи та моделі прогнозування часових рядів (моделі ETS, регресійні моделі, моделі ARIMA), то для роботи з циклічними даними найкраще підходить модель ARIMA. Як уже згадувалося, жодна модель ETS не працює із циклічними даними. Справа в тому, що цикли – коливання, які не мають однакового періоду. Моделі ETS, відповідно, не можуть визначати й відтворювати аперіодичну циклічну поведінку часового ряду [23].

Якщо взяти регресійну модель, то вона обчислює значення часового ряду на основі його зв’язку з іншими рядами. У такому випадку для прогнозування потрібно використовувати значення кількох рядів. Проте за поставленим завданням зв’язок між кількома часовими рядами не розглядається, і тому регресійна модель у цьому випадку не підходить для прогнозування циклічних часових рядів.

Модель ARIMA, натомість, якнайкраще підходить для прогнозування циклічних часових рядів. Будь-яка модель ARIMA задається за допомогою трьох параметрів: p (порядок авторегресійної частини), d (порядок різниці) та q (порядок частини ковзного середнього). Параметр p має велике значення для роботи з циклічними даними. Так, тільки тоді, коли p має значення 2 або більше, модель ARIMA може відтворювати циклічну поведінку часового ряду [2].

3.2.2 Модель XGBoost

XGBoost (англ. eXtreme Gradient Boosting – екстремальне градієнтне підсилення) – програмна бібліотека з відкритим вихідним кодом, яка надає функціонал для виконання градієнтного підсилення. Градієнтне підсилення є технікою машинного навчання, яка на основі слабких моделей визначає сильну, використовуючи для цього дерева рішень [36, 37].

До задач машинного навчання, які можуть виконувати алгоритми XGBoost, належить класифікація, регресія, ранжування. Разом із тим, XGBoost може створювати моделі для прогнозування рядів, зокрема ті, що враховують циклічність [37].

Бібліотека XGBoost створена для забезпечення високої продуктивності градієнтного підсилення та підвищення швидкості пов’язаних обчислень. До особливостей і переваг XGBoost можна зарахувати наступне [36, 38, 39]:

- підтримка різних мов програмування, зокрема R та Python, а також Java, C++, Scala, Julia;
- можливість роботи на різних операційних системах (Windows, Linux, macOS) та підтримка різних хмарної інтеграції з AWS (Amazon Web Services), Microsoft Azure, GCP (Google Cloud Platform), кластерами Yarn;
- підтримка розподілених обчислень за допомогою таких фреймворків, як Apache Spark, Apache Hadoop, Apache Flink та Dask;
- робота в умовах обмеженості ресурсів, що включає оптимізацію пам’яті кешу, використання доступних ядер процесора для розпаралелювання під час побудови дерева рішень, можливість виконання обчислень у зовнішній пам’яті.

3.2.3 Модель LightGBM

LightGBM (англ. Light Gradient-Boosting Machine – машина легкого градієнтного підсилення) також є бібліотекою градієнтного підсилення, як і XGBoost. Бібліотека LightGBM розроблена компанією Microsoft та має відкритий вихідний код. До платформ, які підтримує LightGBM, належить Linux

Windows та macOS. Сама бібліотека LightGBM доступна для мов R, Python, C++ і C#. Також LightGBM підтримує хмарні платформи Microsoft Azure, AWS, GCP та фреймворки розподілених обчислень Apache Spark і Dask [40]. Так само, як XGBoost, LightGBM за допомогою машинного навчання створює моделі класифікації, регресії, а також моделі, що можуть виконувати прогнозування рядів із циклічністю [41, 42].

У бібліотеках LightGBM та XGBoost дещо відрізняється принцип градієнтного підсилення, а саме принцип побудови дерев рішень. У XGBoost дерева рішень будуються по рівнях горизонтально, а LightGBM, натомість, здійснює побудову дерева вертикально в напрямку листа, що робить алгоритм швидшим у виконанні .

Проте ціною зменшення часових втрат є зростання ризику перенавчання моделі. Таким чином, LightGBM не здатна створювати такі надійні моделі, як XGBoost. Також, якщо порівняти з XGBoost, бібліотека LightGBM у градієнтному підсиленні використовує додаткові методики, які при збереженні високого рівня точності збільшують швидкість роботи алгоритму [38].

3.3 Розробка програмного засобу прогнозування циклічних часових рядів

3.3.1 Вибір засобів розробки

Для розробки програмного засобу прогнозування циклічних часових рядів використовується мова програмування Python, а також підключені бібліотеки XGBoost, LightGBM, Pandas, Sktime, Matplotlib.

Мова Python має ряд переваг, які впливають на її вибір для обробки даних та їх аналізу. Такими перевагами є простота її освоєння та наявність відкритого вихідного коду, що дозволяє завдяки спільноті швидко вдосконалювати Python. Не менш важливою перевагою є наявність широкого спектру доступних бібліотек, що надають функції від простої обробки й візуалізації даних та їх

статистичного аналізу та машинного навчання. Таким чином, для бібліотеки градієнтного підсилення XGBoost та LightGBM підтримують пакети для використання в Python.

Pandas – це бібліотека Python з відкритим вихідним кодом, яка надає структури даних та функції для роботи з часовими рядами й даними, організованими у вигляді таблиць. Функції Pandas реалізують читання й запису даних з підтримкою CSV-файлів (англ. comma-separated values – значення, розділені комою), текстових файлів, файлів Microsoft Excel, баз даних SQL (англ. Structured Query Language – мова структурованих запитів) тощо [43].

Бібліотека Sktime використовується для виконання завдань машинного навчання з використанням часових рядів. Sktime реалізує інтерфейс створення, навчання, застосування моделей, який уніфікований з іншими бібліотеками Python. Серед реалізованих завдань машинного навчання пропонується класифікація, кластеризація, прогнозування часових рядів [44].

Бібліотека Matplotlib надає можливості для візуалізації даних за допомогою мови Python. Функціонал Matplotlib дозволяє створювати статичні, інтерактивні, анімовані графіки. Для побудови використовуються різні типи графіків: графіки розсіювання, стовпикові графи, гістограми тощо [45].

3.3.2 Розробка модулів програмного засобу

Під час розробки програмного засобу прогнозування циклічних часових рядів самий програмний засіб розділено на чотири модулі, кожен з яких має відповідний файл. Це модулі tsforecast, forecasting, evaluating, plotting. Модуль tsforecast відповідає за саме виконання програмного засобу та взаємодію з користувачем. Модулі forecasting, evaluating, plotting містять додатковий функціонал: для обчислення прогнозів за допомогою різних моделей, обчислення похибок прогнозу, побудови графіків часових рядів.

На рисунку 3.1 показано структуру модулів програмного засобу. За цим рисунком модуль tsforecast використовує функціонал модулів forecasting,

evaluating, plotting. Доступ до самого модуля tsforecast має користувач, який через цей модуль може користуватися програмним засобом після його запуску.

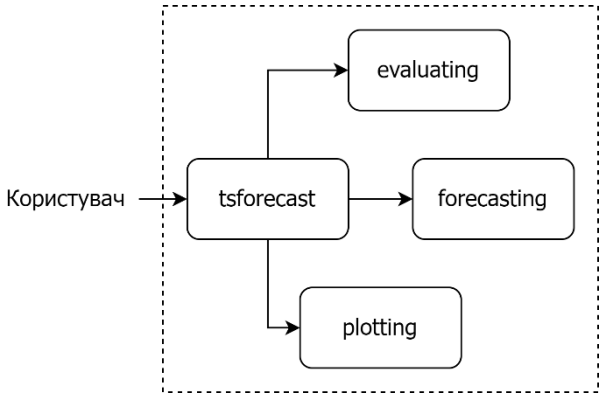

Рисунок 3.1 – Структура модулів програмного засобу

Модулю tsforecast відповідає файл «tsforecast.py», код якого міститься в додатку Б, у лістингу Б.1. Модуль tsforecast додає функціонал з модулів forecasting, evaluating та plotting. У tsforecast міститься одна функція main, яка виконується при запуску програмного засобу. Зразу після початку роботи main відбувається перевірка аргументів командного рядка.

Принцип запуску такий: у командному рядку із введеними аргументами запускається програмний засіб. Він на основі отриманих даних відкриває файл з часовим рядом, створює для нього прогнози, обчислює похибки та відображає це все для користувача. Підтримується п’ять аргументів командного рядка:

- csv_file – назва CSV-файлу, у якому зберігається часовий ряд;
- period – періодичність даних ("Y" використовується для щорічних даних, "M" – для щомісячних і "Q" – для щоквартальних);
- p_min – мінімальне значення параметра *p* у моделі ARIMA;
- p_max– максимальне значення параметра *p* у моделі ARIMA;
- test_prop – частка тестових даних, яка входить до всього набору даних часового рядку (необов’язковий аргумент, за замовчуванням дорівнює 0,1).

Спочатку перевіряється наявність введених аргументів (додаток Б). Потім визначається наявність необов’язкового аргументу test_prop. Далі йде перевірка

значення period. Якщо який аргумент не пройшов перевірку, то програмний засіб відображає відповідне повідомлення та припиняє свою роботу.

Після цього йде відкриття CSV-файлу за назвою csv_file. Для структури даних у файлі встановлюється вимога – дані часового ряду мають міститися у двох стовпцях: перший має містити часові мітки та бути індексом, а другий – містити самі дані часового ряду. Після успішного відкриття файлу та збереження даних іде перевірка значень p_min та p_max. За логікою значення p_max має бути більшим за значення p_max. Також, щоб відповідати обмеженню цифрових даних, у моделі ARIMA значення має бути не меншим, ніж 2.

Далі функція main виконує розділення всіх даних на набори навчальних та тестових даних. При розділенні розмір набору тестових даних обчислюється на основі значення test_prop. Сформовані набори даних використовуються у при обчисленні прогнозів. Для цього використовуються функції forecast_arima, forecast_xgboost та forecast_lightgbm, які містяться в модулі forecasting. Також при обчисленні прогнозу за допомогою функції forecast_arima отримуються значення параметрів p, d і q моделі ARIMA, яка використовувалася для прогнозування.

Потім у функції main іде побудова графіків отриманих прогнозів. Для побудови й відображення графіків використовується функція plot_forecast з модуля plotting. Далі відбувається обчислення похибок прогнозів MAE та MAPE для кожної моделі. Для цього використовується функція forecast_errors, яка описана в модулі evaluating. Отримані похибки відображаються користувачеві у вигляді таблиці. На цьому програмний засіб закінчує свою роботу.

Далі розглядається вміст файлів модулів forecasting, plotting та evaluating. Вони мають відповідні файли «forecasting.py», «plotting.py» і «evaluating.py». Код усіх цих файлів можна побачити також у додатку Б, а саме в лістингах Б.2, Б.3 і Б.4 відповідно.

До модуля forecasting належить файл «forecasting.py». Цей файл містить три функції: forecast_arima, forecast_xgboost та forecast_lightgbm (лістинг Б.2).

Функція forecast_arima виконує обчислення прогнозів за допомогою моделі ARIMA й повертає значення прогнозів та параметрів *p*, *d* і *q* моделі ARIMA. Як вхідні параметри функція приймає навчальні й тестові дані, а також мінімальне та максимальне значення *p*. Для створення моделі ARIMA функція forecast_arima використовує клас AutoARIMA з бібліотеки Sktime. Ціллю AutoARIMA є отримання оптимальних параметрів *p*, *d* і *q*. Алгоритм спочатку визначає необхідний порядок різниці *d*, щоб забезпечити стаціонарність даних. Далі він для діапазонів *p* та *q* навчає всі варіанти моделі, серед яких він визначає найкращий [46].

Після навчання створеної моделі ARIMA визначаються її параметри та обчислюється прогноз, що й повертається функцією forecast_arima у вигляді результату.

Функція forecast_xgboost виконує прогнозування з використанням моделі з бібліотеки XGBoost. На вході forecast_xgboost приймає навчальні й тестові дані. У середині самої функції спочатку створюється модель XGBRegressor, далі за допомогою функції make_reduction з Sktime здійснюється перетворення моделі з трансформацією ковзного вікна розміром з набір тестових даних.

Після того проводиться перехресна перевірка з рандомізованим пошуком для оптимізації визначених параметрів (window_length – розмір вікна, n_estimators – кількість дерев у XGBoost, max_depth – максимальна глибина дерева) [47]. Після навчання моделі здійснюється обчислення прогнозу даних часового ряду. Обчислений прогноз повертається через результат функції.

Схожим чином працює й функція forecast_lightgbm. Вона також приймає навчальні й тестові дані. Проте вона повертає прогнози, отримані шляхом використання моделі з бібліотеки LightGBM. Для цього функція створює модель LGBMRegressor. Далі використовує make_reduction, щоб перетворити модель з трансформацією ковзного вікна розміром з набір тестових даних.

Далі функція здійснює перехресну перевірку із сітковим пошуком для оптимізації параметрів: window_length – розмір вікна, max_depth – максимальна

глибина дерева [47]. Потім після навчання моделі функція обчислює прогноз даних, який повертає у вигляді результату.

Модуль plotting містить одну функцію plot_forecast, основним завданням якої є побудова та відображення графіка прогнозу (лістинг Б.3). На вході функція приймає навчальні та тестові дані, а також дані прогнозу та заголовок графіка. Для побудови графіка plot_forecast використовує функцію plot_series з бібліотеки Sktime.

Самий графік будується так: на одному полі відображаються графіки навчальних і тестових даних. Також на цьому полі відображається графік даних прогнозу. Для кожного з цих графіків встановлюється позначка – "Навчальні дані", "Тестові дані" та "Дані прогнозу". Над полем з графіками відображається вказаний заголовок.

Модуль evaluating також складається з однієї функції forecast_errors (лістинг Б.4). Ця функція у вигляді вхідних параметрів приймає тестові дані та дані прогнозу, а повертає обчислені значення похибок MAE та MAPE. Для обчислення похибок прогнозу використовуються функції mean_absolute_error і mean_absolute_percentage_error з бібліотеки Sktime.

3.4 Аналіз та прогнозування експериментальних даних

Для демонстрації роботи розробленого програмного засобу потрібен часовий ряд із властивістю циклічності. Циклічна поведінка може бути характерною для економічних показників країн, що розвиваються та мають нестабільну економіку. До таких країн належать країни Африки, у яких кожні декілька років при зміні уряду відбувається зміна напрямку економічної політики.

Тоді у вигляді експериментальних даних з офіційного ресурсу Світового банку [48] береться набір даних, який містить часовий ряд експорту товарів і послуг з Мавританії у відсотках від ВВП протягом 1961-2021 років.

3.4.1 Аналіз часового ряду для прогнозування

Перш, ніж запускати програмний засіб прогнозування цього часового ряду, варто виконати його аналіз, визначити його статистичні характеристики. Дослідження й аналіз часового ряду здійснюється на мові Python і за допомогою середовища Jupyter Notebook. Увесь програмний код, який використовується при дослідженні й аналізі ряду, міститься в лістингах 3.1-3.2.

Часовий ряд зберігається у файлі «export_mrt.csv», де він представлений у вигляді таблиці, що містить два стовпці. Перший стовпець містить роки, за які є дані про експорт. А другий – самі дані у відсотках від ВВП.

Якщо розглядати роботу в середовищі Jupyter Notebook, то спочатку підключаються необхідні бібліотеки (лістинг 3.1): для роботи з табличними даними (Pandas), для роботи з часовими рядами (Statsmodels) і для візуалізації даних (Matplotlib). Далі йде відкриття файлу «export_mrt.csv», збереження часового ряду у вигляді таблиці та побудова його графіка, який зображено на рисунку 3.2.

Лістинг 3.1 – Підключення бібліотек та відкриття файлу «export_mrt.csv»

```
import pandas as pd
import statsmodels as sm
import statsmodels.api as sma
import matplotlib.pyplot as plt

data = pd.read_csv("export_mrt.csv", usecols=[0, 1], index_col=0)
data.plot(figsize=(12, 5))
```

На цьому графіку видно, що часовий ряд має циклічну поведінку. Так, періоди циклів сягають близько 10-12 років. Тому цей часовий ряд чудово підходить для використання розробленого програмного засобу.

Проте для подальшої роботи з цим рядом потрібно перевірити його на стаціонарність. Визначити, чи є часовий ряд стаціонарним, чи ні, можна за допомогою розширеного тесту Дікі-Фуллера [49]. За цим тестом, щоб довести стаціонарність ряду, потрібно спростувати нульову гіпотезу, яка припускає, що

ряд не є стаціонарним. Таким чином, для доведення альтернативної гіпотези про стаціонарність ряду, потрібно, щоб обчислене *p*-значення було менше за рівень значущості 0,05 [50].

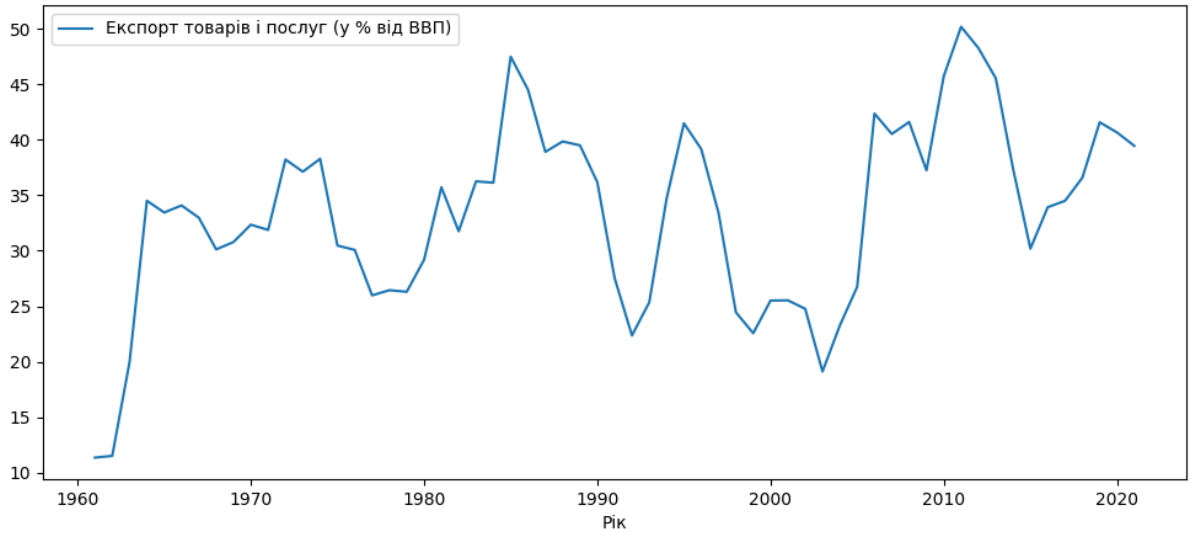

Рисунок 3.2 – Графік часового ряду щорічного експорту

У лістингу 3.2 показано код обчислення *p*-значення за розширеним тестом Дікі-Фуллера та побудови графіків автокореляції та часткової автокореляції числового ряду. На рисунку 3.3 подано результат виконання цього коду.

Лістинг 3.2 – Обчислення *p*-значення, побудова графіка часового ряду, графіків його автокореляції та часткової автокореляції

```
fig = plt.figure(figsize=(12, 7))
layout = (2, 2)
ts_ax = plt.subplot2grid(layout, (0, 0), colspan=2)
p_value = sma.tsa.stattools.adfuller(data)[1]
ts_ax.set_title(f"Розширений тест Дікі-Фуллера: p-значення =
{p_value:.5f}")
data.plot(ax=ts_ax)
acf_ax = plt.subplot2grid(layout, (1, 0))
pacf_ax = plt.subplot2grid(layout, (1, 1))
sm.graphics.tsaplots.plot_acf(data, ax=acf_ax,
title="Автокореляція")
sm.graphics.tsaplots.plot_pacf(data, ax=pacf_ax, title="Часткова
автокореляція")
plt.tight_layout()
```

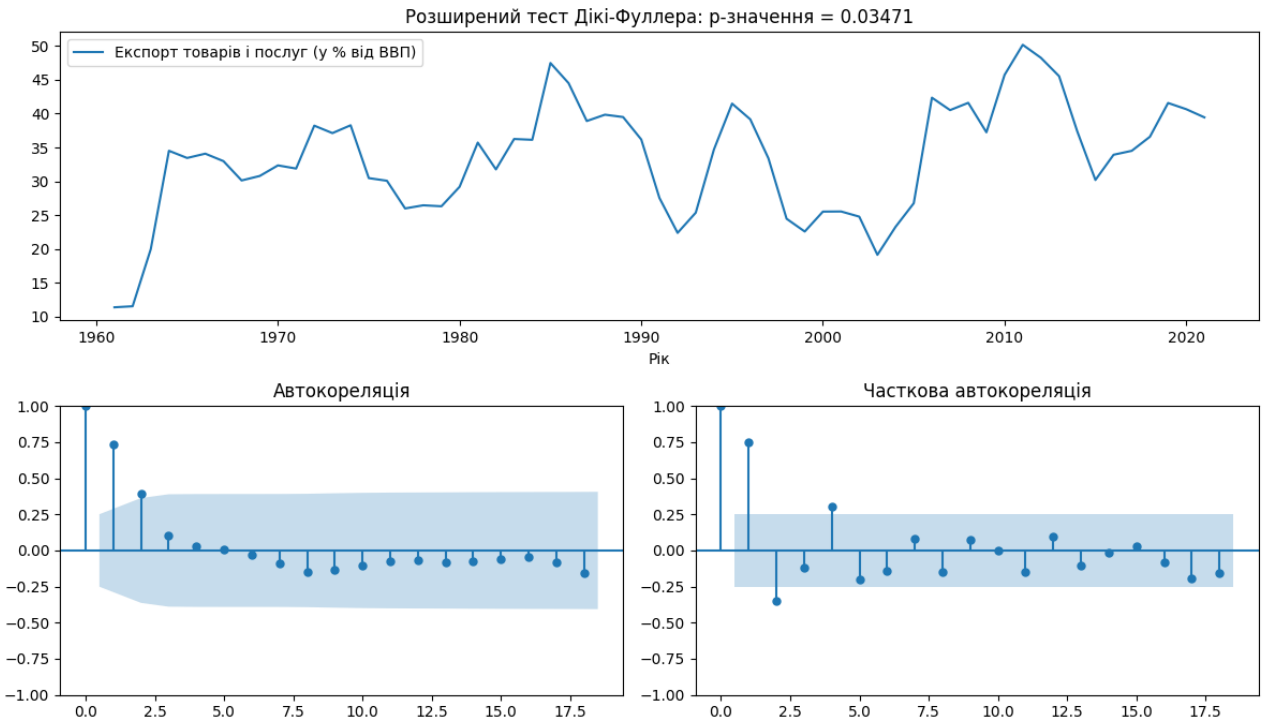

Рисунок 3.3 – Графік часового ряду і графіки його автокореляції та часткової автокореляції

За рисунком 3.3 обчислене p-значення $= 0{,}03471$, що є менше, ніж рівень значущості 0,05. Це означає, що часовий ряд є стаціонарним і надалі може використовуватися в методах і моделях прогнозування. Залишилося розглянути графіки автокореляції і часткової автокореляції.

Як уже згадувалося в другому розділі, автокореляція та часткова автокореляція грають важливу роль у визначенні параметрів моделі ARIMA, яка використовуватиметься для прогнозування під час використання програмного засобу. За цими графіками можна встановити можливе значення параметра p [2]. Так, p може мати значення p_0, якщо:

- автокореляція експоненційно спадає або має синусоїдну поведінку;
- часткова автокореляція має відчутне пікове значення у точці p_0, а після p_0 немає ніяких піків.

Якщо розглянути графік автокореляції часового ряду, то можна сказати, що вона експоненційно спадає. Якщо розглянути графік часткової автокореляції,

тут видно два піки, після яких немає ніяких різких зрушень. Це піки в точках 2 та 4. Для прогнозування часового ряду з моделлю ARIMA програмний засіб використовує алгоритми, які потребують діапазону значень параметра *p* – з мінімальним і максимальним значенням. Тоді для створення моделі ARIMA можна встановити мінімальне значення *p* як 2, а максимальне – як 4.

На цьому закінчується дослідження та аналіз вибраного часового ряду. Далі розглядається використання розробленого програмного засобу для прогнозування цього часового ряду.

3.4.2 Використання програмного засобу прогнозування циклічних часових рядів

Як зазначалося раніше, при запуску розробленого програмного засобу для нього подаються також аргументи командного рядка, більшість з яких уже визначена. Залишився лише аргумент test_prop, який означає частку тестових даних, які будуть прогнозуватися.

Під час своєї роботи розроблений програмний засіб ділить значення часового ряду на дві частини: набір навчальних даних і набір тестових даних. Навчальні дані використовуються для навчання моделі прогнозування, а тестові дані – для порівняння з даними прогнозу та його оцінювання. Тобто test_prop визначає не тільки розмір набору навчальних даних, які дозволяють навчити модель. Цей аргумент також визначає горизонт прогнозування, на якому значення прогнозів порівнюються зі значенням тестових даних та оцінюється точність моделі.

Щоб визначити test_prop, можна розглянути розмір часового ряду. Самий ряд містить значення експорту за 1961-2021 роки, що становить 61 значення. Якщо дивитися на частку тестових даних, то для такого розміру ряду підходить значення 0,15. Тоді набір тестових даних міститиме $61 \cdot 0{,}15 = 9{,}15 \approx 9$ значень ряду і горизонт прогнозування також дорівнюватиме 9.

Отже, для запуску розробленого програмного засобу тепер визначено всі необхідні аргументи командного рядка:

- назва файлу csv_file – «export_mrt.csv»;
- періодичність даних period – щорічні дані ("Y");
- мінімальне значення параметра *p* p_min – 2;
- максимальне значення параметра *p* p_max – 4;
- частка тестових даних test_prop – 0,15.

Після записку у командному рядку із заданими аргументами програмний засіб відкриває файл з даними, робить їх прогнози, а потім відображає ці прогнози у графіках. Далі йде огляд прогнозів кожної моделі.

Першим іде прогноз, отриманий за допомогою моделі ARIMA. Графік цього прогнозу подано на рисунку 3.4. Як можна побачити, при створенні й налаштуванні параметрів програмний засіб вибрав модель ARIMA(4, 0, 0), яка еквівалентна авторегресійній моделі AR(4).

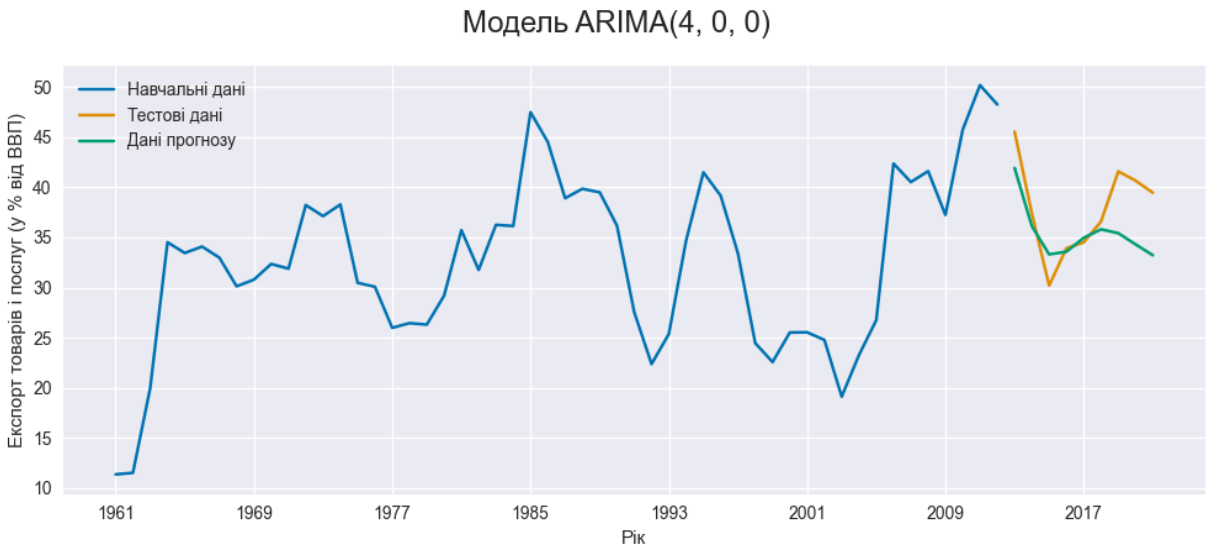

Рисунок 3.4 – Графік прогнозу щорічного експорту
за допомогою моделі ARIMA(4, 0, 0),

Що стосується порівняння даних прогнозу з тестовими даними, то можна сказати, що дані прогнозу досить точно відтворюють тестові дані, особливо на

проміжку 2013-2018 років. До недоліків прогнозу можна віднести те, що модель не змогла передбачити різке падіння експорту у 2015 році. Також вона не побачила зростання експорту з 2018 року.

Загалом, якщо аналізувати прогноз візуально, то модель ARIMA добре змогла визначити точки різкого спаду експорту у 2013-2014 роках і точки поступового росту 2016-2018 роках.

Другий прогноз – прогноз з використанням моделі з бібліотеки XGBoost, який подано на рисунку 3.5. Після аналізу його графіку можна сказати модель змогла добре зрозуміти загальний шаблон значень ряду. Проте самі прогнози не дуже добре збігаються з тестовими значеннями: на проміжку 2013-2016 років прогнози сильно зміщуються вниз, а прогнози на 2017-2021 роки спрогнозовані із випередженням.

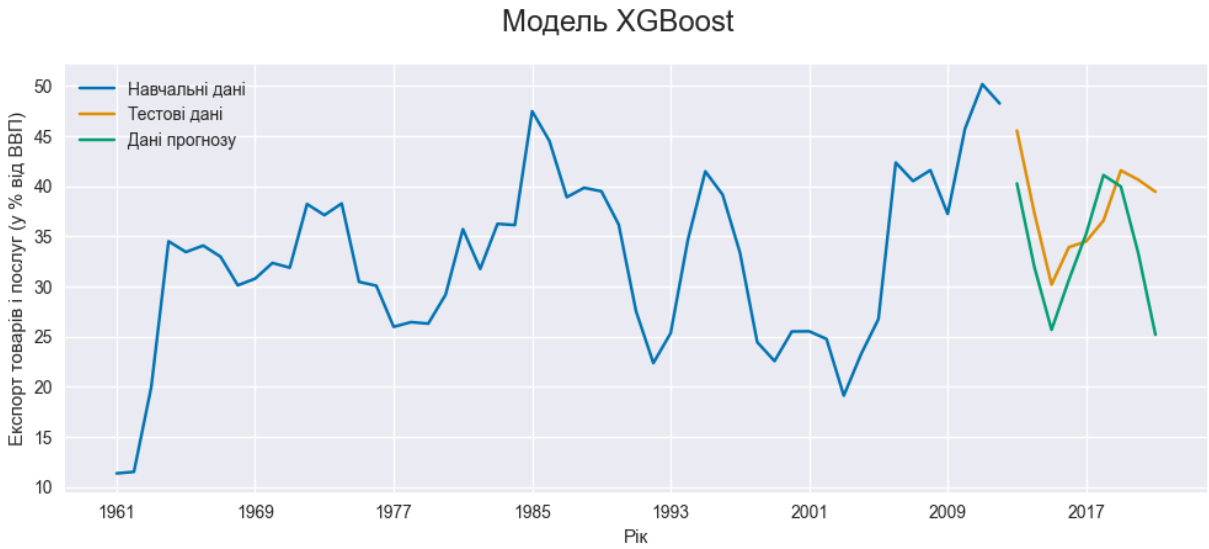

Рисунок 3.5 – Графік прогнозу щорічного експорту
за допомогою моделі з бібліотеки XGBoost

Таким чином, хоч модель XGBoost і змогла точно визначити поведінку ряду на прогнозованому проміжку, але вона не зуміла більше скоротити відстань між даними прогнозу та тестовими даними. Саме цей недолік найбільше буде враховуватися при обчислені похибок.

Третім прогнозом є прогноз, який використовує модель з бібліотеки LightGBM. Графік цього прогнозу можна побачити на рисунку 3.6. З графіку видно, що модель LightGBM просто відобразила прогноз у вигляді прямої, проігнорувавши дані ряду протягом усього прогнозованого проміжку.

Причиною такого прогнозу може бути неправильне навчання модель, а це вказує на проблему її налаштування під дані.

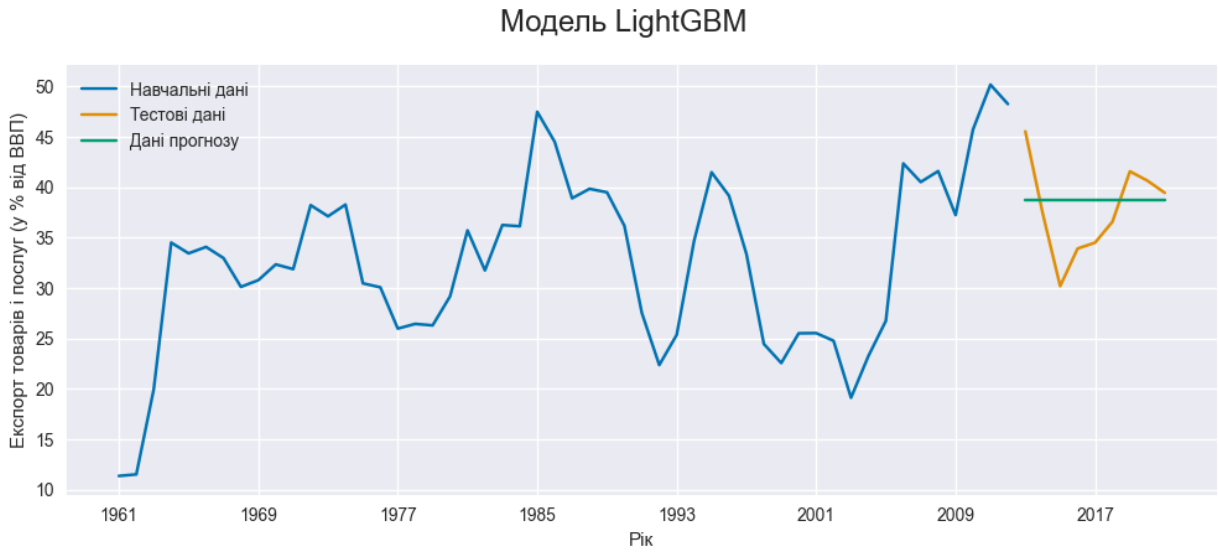

Рисунок 3.6 – Графік прогнозу щорічного експорту
за допомогою моделі з бібліотеки LightGBM

Разом із прогнозами розроблений програмний засіб обчислює й відображає їх похибки. У таблиці 3.1 подано значення похибок обчислених прогнозів: MAE (абсолютної) та MAPE (абсолютної відсоткової).

Таблиця 3.1 – Порівняння похибок прогнозів для кожної моделі

Модель	**MAE**	**MAPE**
ARIMA	3,15	8,04 %
XGBoost	5,21	13,67 %
LightGBM	3,71	10,29 %

На таблиці 3.1 можна побачити, що найменші значення похибок мають прогнози, отримані з моделі ARIMA. При аналізі рисунку 3.3 видно, що прогнози моделі ARIMA найкраще підлаштовані до реальних значень.

Якщо дивитися на похибки, то модель XGBoost має найгірші прогнози. Але попри неточне підлаштування прогнози моделі XGBoost дуже добре враховують поведінку ряду на прогнозованому проміжку.

При заданому часовому ряді найгірший прогноз надала модель LightGBM. Попри те, що її прогноз ніяк не реагує на зміну реальних значень, похибки її прогнозу є досить низькими. Це пов'язано з неправильним налаштуванням моделі, при якому вона намагається досягти мінімальних значень в обчислені похибок MAE і MAPE.

3.5 Висновки до третього розділу

У третьому розділі кваліфікаційної роботи виконано постановку завдання розробки, що полягає в розробці програмного засобу, який за даними з циклічним часовим рядом робить точкові прогнози з використанням різних методів та моделей, будує графіки цих прогнозів й обчислює їхні похибки. Також здійснено вибір методів та моделей, які використовуються у розробці. До них належить модель ARIMA, а також моделі з бібліотек XGBoost і LightGBM.

Потім виконано розробку програмного засобу прогнозування циклічних часових рядів, що включає вибір засобів розробки, проектування структури програмного засобу та розробку його модулів.

Для демонстрації роботи програмного засобу вибрано часовий ряд з набору даних експорту Мавританії протягом 1961-2021 років. Щодо цього ряду проведено дослідження його циклічності, стаціонарності та визначено параметри для прогнозування. Здійснено запуск розробленого програмного засобу для прогнозування вибраного часового ряду за допомогою моделей ARIMA, XGBoost і LightGBM та виконано аналіз цих прогнозів.

4 ОХОРОНА ПРАЦІ ТА БЕЗПЕКА В НАДЗВИЧАЙНИХ СИТУАЦІЯХ

4.1 Характеристика НС природного характеру. Характеристика НС соціального характеру

Згідно з Кодексом цивільного захисту України залежно від характеру походження надзвичайні ситуації (НС) поділяються на такі види [51, 52]: НС техногенного характеру, НС природного характеру, НС соціального характеру, НС воєнного характеру. Надалі розглядаються НС природного характеру та НС соціального характеру.

НС природного характеру – це наслідки небезпечних геофізичних, геологічних, метеорологічних, гідрологічних явищ, природних пожеж, інфекційних захворювань людей і сільськогосподарських тварин, зміни стану водних ресурсів та біосфери тощо [52].

До геофізичних явищ, які спричиняють НС, належать виверження вулканів і землетруси. Виверження вулканів, пов’язані з проникненням магми на поверхню землі. Їх наслідками вивержень є знищення будівель і загибель значної кількості людей, виникнення пожеж під дією вулканічної лави, викиди в повітря шкідливих речовин. Землетруси – сейсмічні явища, що виникають у результаті раптових зміщень і розривів у шарах землі. Землетруси часто супроводжуються великими жертвами серед населення, руйнуванням споруд, виникненням виробничих аварій, провалами цілих населених пунктів [52, 53].

Найпоширенішими небезпечними геологічними явищами є зсуви, обвали, селі. Зсуви – ковзкі зміщення ґрунтових мас униз по схилу, що спричиняють знищення сільськогосподарських угідь, руйнування будівель, небезпеку для добування корисних копалин,.

Обвали виникають при відриві мас гірських порід або снігових мас і їх скочування зі схилів, урвищ. Основною причиною їх появи є людська діяльність (неправильне ведення гірських розробок, будівельних робіт). Селями

називаються грязьові потоки ґрунту, каміння, гірських порід, що виникають унаслідок різкого підйому води в руслах гірських річок під час сильних злив, інтенсивного танення снігу, деградації ґрунтів на гірських схилах тощо [53].

Небезпечні метеорологічні явища, що спричиняють НС, – сильні вітри, сильні снігопади, сильні морози й ожеледі. Сильні вітри – це рухи повітряних мас з великою швидкістю і тривалістю, які мають руйнівний характер. Залежно від швидкості сильні вітри поділяються на бурі, урагани, смерчі. Сильні снігопади виникають під час інтенсивного випадання снігу та призводять до погіршення видимості на дорогах або навіть до припинення руху транспорту.

Сильні морози пов’язані з падінням температури повітря до -30 °С і нижче, а сильні ожеледі пов’язані з утворення тонкого шару льоду після дощу при падінні температури. Морози та ожеледі ускладнюють пересування в населених пунктах для пішоходів і водіїв [53, 54].

Гідрологічними небезпечними явищами є паводки та повені. Паводками називаються нерегулярні підвищення рівня річок, що виникають під час сильних дощів або відлиги. Повені – великі затоплення місцевості, спричинені значним підйомом рівня води в річках, озерах, водосховищах. Наслідки повеней охоплюють вимивання родючого шару ґрунту, загибель людей, пошкодження й руйнування будівель, знищення сільськогосподарських культур і загибель свійських тварин тощо [53].

Природні пожежі виникають внаслідок як природніх явищ (посуха, блискавка), так і діяльності людини (спалювання відходів). Основними видами природних пожеж є степові та лісові. Степові пожежі виникають на відкритій місцевості та залежать від пори року. Лісові пожежі поділяються на низові, верхові та підземні.

Низові пожежі розвиваються з горінням рослин і рослинних залишків, розташованих безпосередньо на ґрунті або на невеликій висоті. Верхові пожежі, натомість, пов’язані з горінням крон дерев і чагарників. Підземні пожежі розгортаються у торф’яних шарах ґрунту на глибині 50 см [53].

Інфекційні захворювання людей і тварин – це хвороби, які можуть передаватися від заражених людей чи тварин до здорових. Збудниками інфекційних захворювань в основному є мікроорганізми (віруси, бактерії, гриби) та їхні токсини. Результатами інфекційних захворювань є погіршення стану здоров'я або загибель значної кількості людей, скорочення чисельності сільськогосподарських тварин [52].

НС соціального характеру – це прояви злочинних дій, які мають антиконституційний і терористичний напрям. НС соціального характеру можуть виникати в різних регіонах світу і тому є небезпечними в багатьох країнах. Злочинними діями, пов'язаними з НС соціального характеру, є загрози або здійснення терористичних актів та інші злочини: встановлення вибухових пристроїв у громадських місцях, крадіжка зброї, викрадення, захоплення, знищення суден [54, 55].

До терористичних актів належать збройні напади, захоплення й утримання ядерних об'єктів, зв'язку й телекомунікацій, замах на життя членів екіпажу повітряного або морського судна [56]. Терористичні акти є проявами тероризму, який полягає у використанні або загрозі застосування насильства через взяття заручників, тортур, убивств, залякування населення й органів влади тощо. У новітній час тероризм досить поширений і є глобальною проблемою [54, 57].

Залежно від причин виникнення тероризм можна поділяти на політичний, релігійний і кримінальний. Політичний тероризм має на меті боротьбу проти політичного чи економічного устрою в певній країні або регіоні світу. Причинами релігійного тероризму є релігійні конфлікти, які базуються на протистоянні релігійних переконань і вірувань. Кримінальний тероризм пов'язаний із сутичками злочинних угруповань з державою та між собою з ціллю отримання власної вигоди [54].

Також джерелом НС соціального характеру є глобальна злочинність, зростання якої часто зумовлюється економічними кризами, спадом рівня життя людей, недоліками законодавства та іншими факторами [53].

До вивчення й аналізу надзвичайних ситуацій (НС) різного характеру походження входить і прогнозування НС на основі даних з минулого. Це схоже на прогнозування часових рядів, у тому числі й часових рядів із властивістю циклічності. Тоді методи та моделі прогнозування різних часових рядів є важливим у прогнозуванні й дослідженні НС.

4.2 Підвищення стійкості роботи об’єктів приладобудівної галузі у воєнний час

Стійкістю роботи об’єкта господарської діяльності (у тому числі об’єкта приладобудівної галузі) називається спроможність об’єкта виготовляти продукцію в запланованому обсязі під час НС, а у випадку порушення матеріального постачання або слабких руйнувань – спроможність відновити виробництво в найкоротший термін [58, 59].

Стійкість роботи об’єкта під час будь-якої НС (зокрема й у воєнний час) визначається факторами:

- надійність захисту працівників об’єкта впливу вибухової зброї та зброї масового ураження;
- здатність будівель і споруд об’єкта протистояти наслідкам НС мирного та воєнного часу;
- захищеність об’єкта від вторинних факторів ураження;
- надійність постачання об’єкта електроенергією, водою, сировиною, матеріалами, інструментами;
- стійкість і оперативність управління виробництвом та цивільним захистом (ЦЗ);
- готовність об’єкта до рятувальних й інших невідкладних робіт та робіт, пов’язаних з відновленням порушеного виробництва.

В умовах воєнного часу підвищення стійкості роботи об’єкта є одним з головних завдань ЦЗ. Підвищення стійкості роботи об’єкта приладобудівної

галузі у воєнний час можна досягти за допомогою виконання ряду заходів, серед яких розглядаються інженерно-технічні, технологічні, організаційні заходи [59].

Інженерно-технічні заходи полягають у забезпечення стійкості будівель і споруд об’єкта, його матеріально-технічного устаткування [59, 60].

До інженерно-технічних заходів входить:

- забезпечення захисту працівників від вибухової зброї та зброї масового ураження;
- захист виробничих споруд та обладнання;
- захист об’єкта від вторинних факторів ураження;
- забезпечення надійного постачання електроенергії, води, роботи комунальних мереж;
- підвищення матеріально-технічного постачання об’єкта.

За захист працівників об’єкта відповідають спеціальні захисні споруди (сховища, укриття), у яких вони можуть укриватися під час ракетних ударів або в разі застосування зброї масового ураження. Ці споруди плануються й будуються заздалегідь [58, 60].

Забезпечення захисту споруд і будівель має відбуватися ще на етапі проектування. Для цього при побудові використовуються міцніші матеріали і конструкції. Найважливіші споруди будуються із заглибленням або зі зменшенням площі стін для забезпечення кращого опору від ударних хвиль. Щодо споруд з різною висотою застосовуються різні заходи: низькі споруди частково обсипаються ґрунтом, а високі – закріплюються відтяжками. Трубопроводи для захисту встановлюються заглиблено у ґрунті.

Захист обладнання об’єкта полягає у мінімізації його руйнувань і пошкоджень. Для захисту обладнання потрібно використовувати захисні пристрої та розміщувати його у приміщеннях на глибині [60, 61].

Вторинними факторами ураження можуть бути пожежі або витік небезпечних речовин. Для захисту об’єкта від вторинних факторів ураження проводяться такі заходи: підвищення вогнестійкості конструкцій, побудова

водоймищ для гасіння пожеж, побудова відокремлених сховищ для легкозаймистих і отруйних речовин [60].

Для забезпечення надійного постачання електроенергії підключаються декілька зовнішніх джерел або встановлюється автономне резервне джерело. Стала робота систем водопостачання вимагає побудови резервних джерел і використання оборотного водопостачання, що дозволяє повторно використовувати воду для технічних цілей. Висока стійкість роботи комунальних мереж пов’язана зі встановленням кільцевої системи та побудовою окремих каналізаційних систем [59, 60].

Підвищення постачання запасів сировини, матеріалів, інструментів забезпечує безперебійне виробництво в умовах воєнного часу. У випадку ураження об’єкта для збереження цих запасів їх необхідно розосереджено зберігатися в захищених місцях [60].

Технологічні заходи призначені для зміни технологічного процесу роботи об’єкта з метою спрощення виробництва продукції та обмеження можливості виникнення вторинних факторів ураження.

Організаційні заходи передбачають розробку ефективних дій керівництва, підрозділів ЦЗ під час НС для захисту працівників, проведення рятувальних та інших робіт, відновлення виробництва [60]. Організаційні заходи включають:

- підвищення надійності й оперативності управління виробництвом і ЦЗ;
- забезпечення готовності об’єкта до відновлення порушеного виробництва.

Основою надійності й оперативності управління ЦЗ є його сталість на всіх етапах ведення. Так, на об’єкті повинно забезпечуватися стале управління ЦЗ в умовах нападу ворога, проведення евакуації, рятувальних робіт та ін. [62].

Підготовка об’єкта до відновлення виробництва полягає в розробці різних планів відновлення – для різних окремих варіантів можливого ураження. Такі плани відновлення повинні розроблятися з урахуванням головної мети – якомога швидшого виробництва продукції [60, 62].

Забезпечення стійкості об’єктів приладобудівної галузі у воєнний час має важливе значення у випуску електронних пристроїв, комп’ютерів та забезпеченні ними користувачів у державі в умовах воєнного часу. Це у свою чергу тісно пов’язане з використанням комп’ютерів для розробки й запуску програмного засобу, що виконує прогнозування різних часових рядів, зокрема часових рядів із властивістю циклічності.

4.3 Висновки до четвертого розділу

У четвертому розділі кваліфікаційної роботи розглянуто види надзвичайних ситуацій (НС), висвітлено поняття НС природного характеру. Також оглянуто небезпечні явища, що спричиняють НС природного характеру, здійснено аналіз причин їх виникнення і наслідків. Розглянуто саме такі явища: геофізичні, геологічні явища, метеорологічні, гідрологічні явища, пожежі, інфекційні захворювання людей і тварин.

Далі проаналізовано поняття НС соціального характеру. Висвітлено поняття терористичного акту та проаналізовано явище тероризму і його видів. Розглянуто поняття глобальної злочинності та причин її існування.

Крім того, також розглянуто поняття стійкості роботи будь-якого об’єкта господарської діяльності, зокрема об’єкта приладобудівної галузі, та описано фактори, якими вона визначається. Висвітлено заходи підвищення стійкості роботи об’єкта у воєнний час, які можна поділити на інженерно-технічні, технологічні та організаційні заходи.

До інженерно-технічних заходів входить забезпечення захисту працівників об’єкта та підтримка стійкості будівель і споруд. Технологічні заходи проявляються у запобіганні виникненню вторинних факторів ураження. Організаційні заходи полягають у забезпеченні ефективності управління виробництвом і цивільним захистом під час проведення рятувальних та робіт з відновлення виробництва.

ВИСНОВКИ

У результаті виконання кваліфікаційної роботи було здійснено розробку програмного засобу, який використовує моделі прогнозування часових рядів із властивістю циклічності.

У першому розділі кваліфікаційної роботи:

- проведено аналіз джерел щодо часових рядів та їх прогнозування;
- визначено циклічні часові ряди як ряди, які мають шаблон циклічності, висвітлено поняття точкового прогнозу й інтервалу прогнозування
- розглянуто прості методи прогнозування, описано процеси оцінювання точності прогнозів та проведення перехресної перевірки

У другому розділі роботи:

- досліджено популярні моделі прогнозування часових рядів (моделі ETS, регресійні моделі, моделі ARIMA), комплексні моделі (динамічні регресійні моделі, векторні авторегресійні моделі) та моделі штучних нейронних мереж;
- проаналізовано роботу розглянутих моделей з урахуванням циклічності часових рядів.

У третьому розділі роботи:

- у результаті аналізу розглянутих моделей вибрано модель ARIMA, також вибрано моделі з бібліотек машинного навчання XGBoost і LightGBM;
- виконано розробку програмного засобу прогнозування часових рядів та здійснено його виконання з використанням експериментальних даних;
- проведено порівняльний аналіз отриманих прогнозів, за яким модель ARIMA найвищу точність прогнозу, модель XGBoost добре визначає поведінку даних, модель LightGBM показує найгірші результати прогнозу.

У розділі «Охорона праці та безпека в надзвичайних ситуаціях» розглянуто характеристики НС природного характеру та НС природного характеру, описано їх особливості. Також висвітлено заходи, спрямовані на підвищення стійкості роботи об’єктів приладобудівної галузі у воєнний час.

ПЕРЕЛІК ДЖЕРЕЛ

1. Introduction to Time Series Analysis. Definitions, Applications and Techniques. Information Technology Laboratory | NIST. URL: https://www.itl.nist.gov/div898/handbook/pmc/section4/pmc41.htm (date of access: 30.04.2023).

2. Hyndman R.J., Athanasopoulos G. Forecasting: Principles and Practice. OTexts, 2021.

3. Nielsen A. Practical Time Series Analysis: Prediction with Statistics and Machine Learning. O'Reilly Media, Inc., 2019. 504 p.

4. Statswork. Application of Time Series Analysis in Financial Economics. Medium. URL: https://statswork.medium.com/application-of-time-series-analysis-in-financial-economics-c1a7e1b69d04 (date of access: 30.04.2023).

5. Time Series Analysis and Forecasting: Examples, Approaches, and Tools. AltexSoft. URL: https://www.altexsoft.com/blog/business/time-series-analysis-and-forecasting-novel-business-perspectives/ (date of access: 30.04.2023).

6. Time Series Forecasting Methods, Techniques & Models | InfluxData. InfluxData. URL: https://www.influxdata.com/time-series-forecasting-methods/ (date of access: 30.04.2023).

7. Time Series Forecasting for Healthcare Diagnosis and Prognostics with the Focus on Cardiovascular Diseases / C. Bui et al. 6th International Conference on the Development of Biomedical Engineering in Vietnam (BME6). Singapore, 2017. P. 809–818.

8. Time Series Patterns. Hydropedia - Pratik Solanki's Blog. URL: http://hydropedia.blogspot.com/2014/03/time-series-patterns.html (date of access: 30.04.2023).

9. Components of a time series. IHMC Sample Knowledge Models. URL: https://cmapskm.ihmc.us/rid=1052458821502_1749267941_6906/components.pdf (date of access: 30.04.2023).

10. Time Series Analysis: The Basics. Australian Bureau of Statistics. URL: https://www.abs.gov.au/websitedbs/d3310114.nsf/home/time+series+analysis:+the+basics (date of access: 30.04.2023).

11. Decomposition Models | STAT 510. PennState: Statistics Online Courses. URL: https://online.stat.psu.edu/stat510/lesson/5/5.1 (date of access: 30.04.2023).

12. How to Decompose Time Series Data into Trend and Seasonality - MachineLearningMastery.com. MachineLearningMastery.com. URL: https://machinelearningmastery.com/decompose-time-series-data-trend-seasonality/ (date of access: 30.04.2023).

13. Types of forecasts | Forecasting and Analytics with ADAM. OpenForecast. URL: https://www.openforecast.org/adam/types-of-forecasts.html (date of access: 30.04.2023).

14. Types of Forecasts. Intensity. URL: http://intensity.com/wp-content/uploads/2016/07/Forecasting-Note-No1.pdf (date of access: 30.04.2023).

15. Residuals | STAT 462. Statistics Online. URL: https://online.stat.psu.edu/stat462/node/116/ (date of access: 30.04.2023).

16. Introduction to Simple Forecasting Methods. WFM eLearning. URL: https://wfmelearning.com/introduction-to-simple-forecasting-methods/ (date of access: 30.04.2023).

17. Measuring accuracy of point forecasts | Forecasting and Analytics with ADAM. OpenForecast. URL: https://openforecast.org/adam/errorMeasures.html (date of access: 30.04.2023).

18. Evaluating Forecast Accuracy. Universität Wien. URL: https://homepage.univie.ac.at/robert.kunst/procip.pdf (date of access: 30.04.2023).

19. Wheatcroft E. Interpreting the skill score form of forecast performance metrics. International Journal of Forecasting. 2019. Vol. 35, no. 2. P. 573–579.

20. What is Exponential Smoothing? Information Technology Laboratory | NIST. URL: https://www.itl.nist.gov/div898/handbook/pmc/section4/pmc43.htm (date of access: 30.04.2023).

21. Exponential Smoothing | Time Series Analysis. Bookdown. URL: https://bookdown.org/mpfoley1973/time-series/exponential.html (date of access: 30.04.2023).

22. Introduction to ETS | Forecasting and Analytics with ADAM. OpenForecast. URL: https://openforecast.org/adam/ETSConventional.html (date of access: 30.04.2023).

23. Cyclic and seasonal time series. Rob J Hyndman. URL: https://robjhyndman.com/hyndsight/cyclicts/ (date of access: 30.04.2023).

24. Beers B. What is Regression? Definition, Calculation, and Example. Investopedia. URL: https://www.investopedia.com/terms/r/regression.asp (date of access: 30.04.2023).

25. Linear Regression. Department of Statistics and Data Science. URL: http://www.stat.yale.edu/Courses/1997-98/101/linreg.htm (date of access: 30.04.2023).

26. Hamers J. Piecewise Linear Regression. Medium. URL: https://medium.com/databulls/piecewise-linear-regression-30b63ff8a90d (date of access: 30.04.2023).

27. Thabethe A. Linear Regression with Time Series Data. Econometrics. 2005. Vol. 2.

28. Autoregressive Models. Statistics Online. URL: https://online.stat.psu.edu/stat501/book/export/html/996 (date of access: 30.04.2023).

29. Moving Average Models (MA models) | STAT 510. PennState: Statistics Online Courses. URL: https://online.stat.psu.edu/stat510/lesson/2/2.1 (date of access: 30.04.2023).

30. Hayes A. Autoregressive Integrated Moving Average (ARIMA) Prediction Model. Investopedia. URL: https://www.investopedia.com/terms/a/autoregressive-integrated-moving-average-arima.asp (date of access: 30.04.2023).

31. Introduction to ARIMA models. Duke University. URL: https://people.duke.edu/~rnau/411arim.htm (date of access: 30.04.2023).

32. Seasonal ARIMA models | STAT 510. PennState: Statistics Online Courses. URL: https://online.stat.psu.edu/stat510/lesson/4/4.1 (date of access: 30.04.2023).

33. STAT481/581: Introduction to Time Series Analysis. Department of Mathematics & Statistics | New Mexico's Flagship University. URL: https://math.unm.edu/~lil/Stat581/10-dynamic-regression.pdf (date of access: 30.04.2023).

34. Vector Autoregressions. Introduction to Econometrics with R. URL: https://www.econometrics-with-r.org/16-1-vector-autoregressions.html (date of access: 30.04.2023).

35. What are Neural Networks? IBM. URL: https://www.ibm.com/topics/neural-networks (date of access: 30.04.2023).

36. What is XGBoost?. NVIDIA Data Science Glossary. URL: https://www.nvidia.com/en-us/glossary/data-science/xgboost/ (date of access: 30.04.2023).

37. XGBoost Algorithm. Amazon Web Services. URL: https://docs.aws.amazon.com/sagemaker/latest/dg/xgboost.html (date of access: 30.04.2023).

38. Saha S. XGBoost vs LightGBM: How Are They Different. neptune.ai. URL: https://neptune.ai/blog/xgboost-vs-lightgbm (date of access: 30.04.2023).

39. XGBoost. XGBoost. URL: https://xgboost.ai/ (date of access: 30.04.2023).

40. Welcome to LightGBM’s documentation!. LightGBM 3.3.5 documentation. URL: https://lightgbm.readthedocs.io/en/v3.3.5/index.html (date of access: 30.04.2023).

41. How LightGBM works. Amazon Web Services. URL: https://docs.aws.amazon.com/sagemaker/latest/dg/lightgbm-HowItWorks.html (date of access: 30.04.2023).

42. LightGBM - Overview. Microsoft on GitHub. URL: https://microsoft.github.io/SynapseML/docs/features/lightgbm/LightGBM%20-%20Overview/ (date of access: 30.04.2023).

43. pandas - Python Data Analysis Library. pandas. URL: https://pandas.pydata.org/about/ (date of access: 30.04.2023).

44. Welcome to sktime. sktime documentation. URL: https://sktime.readthedocs.io/en/stable/ (date of access: 30.04.2023).

45. Matplotlib – Visualization with Python. Matplotlib. URL: https://matplotlib.org/ (date of access: 30.04.2023).

46. AutoARIMA. sktime documentation. URL: https://www.sktime.net/en/latest/api_reference/auto_generated/sktime.forecasting.arima.AutoARIMA.html (date of access: 30.04.2023).

47. Hyperparameter Optimization With Random Search and Grid Search. MachineLearningMastery.com. URL: https://machinelearningmastery.com/hyperparameter-optimization-with-random-search-and-grid-search/ (date of access: 30.04.2023).

48. Exports of goods and services (% of GDP) - Mauritania. World Bank Open Data. URL: https://data.worldbank.org/indicator/NE.EXP.GNFS.ZS?contextual=default&end=2021&locations=MR&name_desc=false&start=1960&view=chart (date of access: 30.04.2023).

49. Masui T. Multi-step Time Series Forecasting with ARIMA, LightGBM, and Prophet. Medium. URL: https://towardsdatascience.com/multi-step-time-series-forecasting-with-arima-lightgbm-and-prophet-cc9e3f95dfb0 (date of access: 30.04.2023).

50. Chaudhary M. Why is Augmented Dickey–Fuller test (ADF Test) so important in Time Series Analysis. Medium. URL: https://medium.com/@cmukesh8688/why-is-augmented-dickey-fuller-test-adf-test-so-important-in-time-series-analysis-6fc97c6be2f0 (date of access: 30.04.2023).

51. Кодекс цивільного захисту України : Кодекс України від 02.10.2012 р. № 5403-VI : станом на 31 берез. 2023 р. URL: https://zakon.rada.gov.ua/laws/show/5403-17#Text (дата звернення: 30.04.2023).

52. Березюк О.В., Лемешев М.С. Безпека життєдіяльності. URL: https://web.posibnyky.vntu.edu.ua/fmbt/berezyuk_bezpeka_zhittyediyalnosti/index.htm (дата звернення: 30.04.2023).

53. Толок А.О., Крюковська О.А. Безпека життєдіяльності : навч. посіб. Кам'янське : ДДТУ, 2011. 215 с.

54. Лекція 2. Надзвичайні ситуації мирного та воєнного часу, їх вплив на безпеку життєдіяльності населення України. Інформаційно-обчислювальний центр – Національний технічний університет «Харківський політехнічний інститут». URL: http://web.kpi.kharkov.ua/safetyofliving/wp-content/uploads/sites/171/2017/02/go_lekcia_2.pdf (дата звернення: 30.04.2023).

55. Приходько Р.В. Надзвичайна ситуація: поняття, зміст, характеристика. Державне будівництво. 2011. № 1. URL: http://nbuv.gov.ua/UJRN/DeBu_2011_1_47 (дата звернення: 30.04.2023).

56. Тема 4. Соціальні небезпеки, їх види та особливості. ОДУВС – Одеський державний університет внутрішніх справ. URL: https://oduvs.edu.ua/wp-content/uploads/2016/09/Lektsiya-OP-4.pdf (дата звернення: 30.04.2023).

57. Війна і тероризм. Council of Europe. URL: https://www.coe.int/uk/web/compass/war-and-terrorism (дата звернення: 30.04.2023).

58. Охорона праці в галузі та цивільний захист : навч. посіб. / Ю.А. Гасило, О.А. Крюковська. К.О. Левчук, Р.Я. Романюк. Кам'янське : ДДТУ, 2017. 369 с.

59. Журавель М.О. Конспект лекцій дисципліни «Цивільний захист і охорона праці в галузі» : конспект лекцій. Запоріжжя : НУ «Запорізька політехніка», 2020. 49 с.

60. Лекція № 4 на тему: Стійкість роботи промислових об'єктів у надзвичайних ситуаціях. Освітній сайт КНУБА. URL: https://org2.knuba.edu.ua/pluginfile.php/114575/mod_resource/content/1/Лекція%20№4.pdf (дата звернення: 30.04.2023).

61. Основи цивільного захисту : навч. посіб. / В.О. Васійчук, В.Є. Гончарук, С.І. Качан, С.М. Мохняк. Львів : НУ «Львівська політехніка», 2010. 417 с.

62. Фесенко Г.В., Д'яконов В.І. Курс лекцій з дисципліни «Цивільний захист» : курс лекцій. Харків : ХНУМГ ім. О. М. Бекет., 2015. 98 с.

ДОДАТКИ

Додаток А

Матеріали конференцій

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
ТЕРНОПІЛЬСЬКИЙ НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ
ІМЕНІ ІВАНА ПУЛЮЯ

М А Т Е Р І А Л И

Х НАУКОВО-ТЕХНІЧНОЇ КОНФЕРЕНЦІЇ

«ІНФОРМАЦІЙНІ МОДЕЛІ, СИСТЕМИ ТА ТЕХНОЛОГІЇ»

7–8 грудня 2022 року

ТЕРНОПІЛЬ
2022

УДК 004.67
І. Воробець
(Тернопільський національний технічний університет імені Івана Пулюя, Україна)

ПОРІВНЯННЯ МЕТОДІВ ПРОГНОЗУВАННЯ ЧАСОВИХ РЯДІВ

UDC 004.67
I. Vorobets

COMPARISON OF TIME SERIES FORECASTING METHODS

Прогнозування часових рядів, має велике значення у плануванні. Особливу роль воно відіграє в бізнесі та фінансах: прогнозування щоденного курсу валют, цін акції компаній, місячних/квартальних продажів товару тощо. Також прогнозування часових рядів використовується у метеорології (наприклад температури повітря, кількості опадів), у енергетиці (наприклад прогнозування споживання електроенергії, води, газу).

Методи прогнозування значною мірою залежать від доступних даних. Так, якщо немає доступних даних, то використовуються якісні методи, що надають суто припущення на основі розроблених підходів. В іншому випадку, використовуються кількісні методи [1]. Надалі розглядатимуться саме кількісні методи прогнозування.

У кількісному прогнозуванні можуть використовуватися різноманітні методи, що можуть розроблятися в межах окремих сфер застосування. Кожен з цих методів має певні властивості, точність та вартість. Найпростіші методи використовують лише дані про змінну, яку потрібно прогнозувати. Для цього вони розділяють часові ряди на компоненти: тренд, сезонність і залишки або «шум». Кожен з цих компонентів відображає певний шаблон поведінки ряду. Методи прогнозування, які найчастіше використовуються, розробляються на основі регресивних моделей, моделей експоненційного згладжування, моделей ARIMA (Autoregressive Integrated Moving Average, авторегресивне інтегроване ковзне середнє).

Регресивні моделі прогнозування спираються опис зв'язку між декількома рядами. Застосування регресивної моделі дозволяє прогнозувати числові значення часового ряду на основі даних, які на нього впливають. Ця модель надає можливість виражати у вигляді певної функції залежність цільової змінної від інших змінних [2].

У моделях експоненційного згладжування для обчислення прогнозів використовуються зважені середні значення минулих даних у часовому ряді. Варто зазначити, що ваги експоненційно зменшуються для давніших даних. Тобто, для раніших даних значення ваг буде більшим [1, 3].

Моделі ARIMA, як і моделі експоненційного згладжування, є найбільш популярними моделями прогнозування, які застосовуються до нестаціонарних часових рядів. ARIMA спирається на моделі авторегресії (AR, Autoregression) та ковзного середнього (MA, Moving Average). Для усунення нестаціонарності ця модель використовує інтегровані ряди [3].

У даній роботі розглядаються основні методи прогнозування часових рядів, концепції, на яких вони базуються. Проводиться порівняння розглянутих методів залежно від особливостей їх використання та залежно від властивостей часових рядів, до яких вони застосовуються.

Література

1. Hyndman R.J., Athanasopoulos G. Forecasting: Principles and Practice. 3rd ed. Melbourne, Australia: OTexts, 2021. 442 p.
2. Palma W. Time series analysis. Wiley, 2016. 579 p.
3. Pathak P.P. Time Series Forecasting – A Complete Guide. Medium. URL: https://medium.com/analytics-vidhya/time-series-forecasting-a-complete-guide-d963142da33f.

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
Тернопільський національний технічний університет імені Івана Пулюя (Україна)
Університет імені П'єра і Марії Кюрі (Франція)
Маріборський університет (Словенія)
Технічний університет у Кошице (Словаччина)
Вільнюський технічний університет ім. Гедимінаса (Литва)
Міжнародний університет цивільної авіації (Марокко)
Наукове товариство ім. Т.Шевченка

АКТУАЛЬНІ ЗАДАЧІ СУЧАСНИХ ТЕХНОЛОГІЙ

Збірник
тез доповідей

XI Міжнародної науково-практичної конференції молодих учених та студентів
7-8 грудня 2022 року

УКРАЇНА
ТЕРНОПІЛЬ – 2022

УДК 004.67

І.В. Воробець
Тернопільський національний технічний університет імені Івана Пулюя, Україна

ВПЛИВ КОМПОНЕНТІВ ЧАСОВИХ РЯДІВ НА ВИБІР МЕТОДУ ПРОГНОЗУВАННЯ

I.V. Vorobets

THE INFLUENCE OF TIME SERIES COMPONENTS ON THE CHOICE OF FORECASTING METHOD

Часовим рядом можна назвати послідовність певних даних, упорядкованих відповідно до часу. Тоді часовий ряд можна розглядати як сукупність двох змінних. Перша змінна – це час (точки відліку), а друга – це самі дані (чисельні значення). Найбільшу цінність аналіз часових рядів має в передбаченні майбутньої поведінки змінної на основі її значень у минулому. Передбачення часових рядів широко використовується в бізнес-плануванні, енергетиці, метеорології тощо.

Різні сили, що визначають поведінку часових рядів, описуються компонентами часових рядів. До компонентів часових рядів належать: тренд, сезонність, циклічність та випадковість або «шум» [1].

Тренд показує загальну тенденцію зростання або спадання числових значень даних протягом тривалого періоду часу. Проте не обов'язково, щоб напрямок зміни значень даних був сталим. У довгостроковому періоді тенденція зростання значень може змінюватися на спадання і навпаки. Тренд може бути лінійним або нелінійним, що можна визначити за допомогою графіку значень ряду у відповідні моменти часу.

Сезонність відображає вплив на значення часового ряду різних сезонних факторів (наприклад дня тижня, місяця, пори року). Це означає, що поведінка сезонних рядів є дуже регулярною. Для сезонності характерний фіксований та визначений період.

Часовий ряд проявляє циклічність тоді, коли зміни у даних не мають фіксованого періоду. Умовно цикл можна поділити на чотири фази: процвітання, спад, застій, підйом. Цикл може бути регулярним, але неперіодичним [1, 2].

Часто виникають проблеми розрізненні сезонності та циклічності, проте ці поняття є кардинально різними. Сезонні коливання мають незмінний період, пов'язаний з певним сезонним фактором. Натомість циклічні коливання відзначаються відсутністю фіксованого періоду. Загалом, тривалість циклів є більшою за тривалість сезонних коливань, а самі значення величин у циклах характеризуються меншою мінливістю, ніж значення у сезонних коливаннях [2].

Ще один компонент часових рядів – випадковість або «шум». За визначенням випадкові коливання є нерегулярними та не можуть використовуватися для передбачення значень у часовому ряді [1].

Компоненти тренду, сезонності та циклічності можуть у тій чи іншій мірі міститися в багатьох рядах. При виборі оптимального методу прогнозування часового ряду необхідно зважати на наявність в ньому цих компонентів. Саме особливості використання методів прогнозування для різних шаблонів компонентів часових рядів і розглядаються в даній роботі.

Література

1. Pathak P.P. Time Series Forecasting – A Complete Guide. Medium. URL: https://medium.com/analytics-vidhya/time-series-forecasting-a-complete-guide-d963142da33f.

2. Hyndman R.J., Athanasopoulos G. Forecasting: Principles and Practice. 3rd ed. Melbourne, Australia : OTexts, 2021. 442 p.

Міністерство освіти і науки України,
Тернопільський національний технічний університет
імені Івана Пулюя
Маріборський університет (Словенія)
Технічний університет в Кошице (Словаччина)
Каунаський технологічний університет (Литва)
Львівський національний університет
імені Івана Франка,
Гірничо-металургійна академія ім. Станіслава Сташиця (Польща)
Луцький національний технічний університет,
Чернівецький національний університет
імені Юрія Федьковича,
Вроцлавський економічний університет (Польща)
Університет технологій та економіки
імені Хелени Ходковської (Польща)
Донбаська державна машинобудівна академія

Студентське наукове товариство

VI МІЖНАРОДНА

студентська науково - технічна конференція

"ПРИРОДНИЧІ ТА ГУМАНІТАРНІ НАУКИ.

АКТУАЛЬНІ ПИТАННЯ"

27-28 квітня 2023 р.

(збірник тез конференції)

Тернопіль 2023

УДК 004.67

Воробець І. – ст. гр. СНнм-61
Тернопільський національний технічний університет імені Івана Пулюя

ВИКОРИСТАННЯ МОДЕЛЕЙ ARIMA ДЛЯ ПРОГНОЗУВАННЯ ЧАСОВИХ РЯДІВ ІЗ ВЛАСТИВІСТЮ ЦИКЛІЧНОСТІ

Науковий керівник: к.т.н., доц. Фриз М.Є.

Vorobets I.
Ternopil Ivan Puluj National Technical University

USING ARIMA MODELS FOR FORECASTING TIME SERIES WITH CYCLICITY FEATURE

Supervisor: Ph.D., Assoc. Prof. Fryz M.Ye.

Ключові слова: прогнозування часових рядів, циклічність, моделі ARIMA.
Keywords: time series forecasting, cyclicity, ARIMA models.

Часовий ряд є упорядкованою відповідно до часу послідовністю значень певної змінної. По суті, часовий ряд варто розглядати як випадковий процес з дискретним часом.

Прогнозування часових рядів полягає у знаходженні майбутніх значень часового ряду з якомога найбільшою точністю на основі минулих значень часового ряду та інформації про його поведінку у минулому. Сфери застосування прогнозування часових рядів обмежуються лише використанням самих рядів. Так, прогнозування рядів має широкий спектр різних сфер застосування: від аналізу фондового ринку та оцінювання росту економіки окремих регіонів до моделювання поширення епідемій і прогнозу погоди.

Шаблон поведінки часового ряду визначається трьома його властивостями: трендом, циклічністю та сезонністю [1]. Тренд показує тенденцію зростання або спадання значень ряду протягом певного періоду часу. На великому часовому проміжку тренд може описувати зміну значень від зростання до спадання або навпаки. Властивість циклічності проявляється тоді, коли відбуваються коливання значень часового ряду без фіксованого періоду. Сезонність описує вплив на поведінку часового ряду різних сезонних факторів, наприклад, день тижня, місяць, квартал року. Відмінність сезонності та циклічності полягає в тому, що сезонність має фіксований і визначений період. Циклічність, натомість, передбачає довшу середню тривалість коливань і більш мінливу їх величину.

Кожен часовий ряд має зазвичай декілька таких властивостей. Саме на основі властивостей тренду, циклічності та сезонності й базуються методи та моделі прогнозування часових рядів. Одними з найбільш часто вживаних моделей прогнозування є моделі ARIMA (англ. autoregressive integrated moving average, укр. авторегресійне інтегроване ковзне середнє). Фактично моделі ARIMA є поєднанням авторегресійних моделей (англ. autoregressive model, AR) і моделей ковзного середнього (англ. moving average model, MA).

Авторегресійні моделі прогнозують значення часового ряду на основі його минулих значень. Моделі ковзного середнього визначає прогноз часового ряду на основі похибок минулих прогнозів. Основним обмеженням авторегресійних моделей і моделей ковзного середнього є стаціонарність часового ряду. Таке саме обмеження

мають і моделі ARIMA. При роботі з нестаціонарними даними моделі ARIMA використовують різниці часових рядів для досягнення їх стаціонарності.

Загалом модель ARIMA описується за формулою:

$$\Delta^d y_t = c + \phi_1 \Delta^d y_{t-1} + \ldots + \phi_p \Delta^d y_{t-p} + \theta_1 \varepsilon_{t-1} + \ldots + \theta_q \varepsilon_{t-q} + \varepsilon_t ,$$

де y_t – значення часового ряду на момент часу t, $\Delta^d y_t$ – різниця ряду ступеня d, різниця першого ступеня записується як $\Delta y_t = y_t - y_{t-1}$, c – константа, ε_t – значення білого шуму, $\phi_1, \ldots, \phi_p$ – коефіцієнти авторегресійної моделі, $\theta_1, \ldots, \theta_q$ – коефіцієнти моделі ковзного середнього.

Визначення моделі ARIMA для певного часового ряду полягає у виборі значень параметрів p, d і q, що є досить складним завданням. Параметр p позначає порядок авторегресійної частини авторегресійної, d – ступінь різниці ряду, q – порядок частини ковзного середнього [1].

Для прогнозування часових рядів із циклічністю добре підходять моделі ARIMA. Важливе значення при цьому має параметр p. Так, для створення прогнозів циклічних рядів необхідною умовою є $p > 1$ [2].

У ході використання моделі прогнозування отримуються точкові прогнози на визначений горизонт прогнозування. Оцінити отримані прогнози, можна за допомогою двох основних похибок: середньої абсолютної похибки (англ. mean absolute error, MAE) та середньої абсолютної відсоткової похибки (англ. mean absolute percentage error, MAPE) [1]. Похибка MAE обчислюється як середнє абсолютних значень похибок прогнозу та використовує ті самі одиниці, що й значення часового ряду. Помилка MAPE обчислюється як середнє абсолютних значень відсоткових похибок, які подають значення у відсотках і не залежать від одиниць значень ряду.

Для побудови моделей прогнозування часових рядів існує багато різних інструментів. У даній роботі прогнозування здійснюється за допомогою бібліотеки Statsmodels. Statsmodels – це відкрита бібліотека мови програмування Python, яка надає функції для проведення описової статистики, побудови й оцінки статистичних моделей, виконання статистичних тестів [3]. Зокрема, Statsmodels дозволяє працювати з моделями прогнозування часових рядів, у тому числі і з моделями ARIMA.

Завданням даної роботи є використання засобів бібліотеки Statsmodels для прогнозування часового ряду із властивістю циклічності. У якості такого часового ряду вибрано набір даних активності сонячних плям Національного центру геофізичних даних США [4]. Цей набір даних включає щорічні дані кількості сонячних плям, зібрані протягом 1700-2000 років.

Література:

1. Hyndman R.J., Athanasopoulos G. Forecasting: principles and practice. 3rd ed. Melbourne, Australia : OTexts, 2021. 442 p.
2. Hyndman R.J. Cyclic and seasonal time series. URL: https://robjhyndman.com/hyndsight/cyclicts.
3. Introduction – statsmodels. URL: https://www.statsmodels.org/stable/index.html.
4. Solar Indices | NCEI. URL: https://www.ngdc.noaa.gov/stp/solar/solar-indices.html.

Додаток Б

Код вихідних файлів програмного засобу

Лістинг Б.1 – Код файлу «tsforecast.py»

```
import sys
import pandas as pd
from sktime.forecasting.model_selection import
temporal_train_test_split

import warnings
warnings.simplefilter("ignore")

from forecasting import forecast_arima, forecast_xgboost,
forecast_lightgbm
from plotting import plot_forecast
from evaluating import forecast_errors


def main(argv):

    # Перевірка, чи користувач ввів хоч якісь аргументи
    if len(argv) not in [5, 6]:
        print("Команда: tsforecast.py csv_file period p_min p_max
[test_perc]")
        print("csv_file - назва CSV-файлу, у якому зберігається
часовий ряд")
        print("period = періодичність даних у файлі (Y щорічні, M
щомісячні, Q щоквартальні)")
        print("p_min - мінімальне значення параметра p")
        print("p_max - мінімальне значення параметра p")
        print("test_prop - частка тестових даних у всьому наборі
даних часового ряду")
        quit(-1)

    # Опрацювання необов'язкового аргумента test_perc
    if len(argv) == 6:
        try:
            test_prop = float(argv[5])
        except ValueError:
            print("Помилка! Неправильно введено відношення
тестових даних!")
            quit(1)
    else:
        test_prop = 0.1

    # Перевірка аргумента period
    period = argv[2]
    if period not in ["Y", "M", "Q"]:
        print("Помилка! Неправильно введено періодичність даних!")
```

```
        quit(1)

    # Читання файлу із заданим ім'ям
    try:
        file = argv[1]
        ts_data = pd.read_csv(file, usecols=[0, 1], index_col=0)
        ts_data.index = pd.PeriodIndex(ts_data.index, freq=period)
    except FileNotFoundError:
        print(f"Помилка! Неможливо знайти файл \"{file}\"!")
        quit(2)

    # Перетворення даних з pandas.DataFrame у pandas.Series
    ts_data = ts_data.squeeze()
    if not isinstance(ts_data, pd.Series):
        print(f"Помилка! Дані з файлу \"{file}\" записані
неправильно!")
        quit(3)

    # Опрацювання аргументів p_min і p_max
    try:
        p_min = int(argv[3])
        p_max = int(argv[4])
    except ValueError:
        print("Помилка! Неправильно введено значення p_min або
p_max!")
        quit(1)
    if p_min >= p_max:
        print("Помилка! Значення p_min має бути меншим за значення
p_max!")
        quit(4)
    if p_min < 2:
        print("Помилка! Значення p_min має бути більшим або рівним
2!")
        quit(5)

    # Розділення даних на набори навчальних і тестових даних
(відповідно до test_prop)
    test_num = int(test_prop * ts_data.size)
    train_data, test_data = temporal_train_test_split(ts_data,
test_size=test_num)

    # Визначення прогнозів за допомогою моделей ARIMA, XGBoost та
LightGBM
    arima_forecast, arima_order = forecast_arima(train_data,
test_data, p_min, p_max)
    forecasts = {
        f"ARIMA{arima_order}": arima_forecast,
        "XGBoost": forecast_xgboost(train_data, test_data),
        "LightGBM": forecast_lightgbm(train_data, test_data)
    }
```

```
    # Побудова графіків прогнозів
    for model, forecast in forecasts.items():
        plot_forecast(train_data, test_data, forecast, f"Модель
{model}")

    # Обчислення похибок прогнозів MAE і MAPE
    errors = {}
    for model, forecast in forecasts.items():
        mae, mape = forecast_errors(test_data, forecast)
        errors[model] = {"MAE": mae, "MAPE": mape}

    errors = pd.DataFrame(errors).T
    print("Похибки прогнозів моделей:")
    print(errors)


if __name__ == "__main__":
    main(sys.argv)
```

Лістинг Б.2 – Код файлу «forecasting.py»

```
from sktime.forecasting.base import ForecastingHorizon
from sktime.forecasting.arima import AutoARIMA
from sktime.forecasting.model_selection import (
    ForecastingRandomizedSearchCV, SingleWindowSplitter,
    ExpandingWindowSplitter, ForecastingGridSearchCV
)
from sktime.performance_metrics.forecasting import
MeanAbsolutePercentageError
from sktime.forecasting.compose import make_reduction

import xgboost as xgb
import lightgbm as lgb


def forecast_arima(train, test, p_min, p_max):
    """
    Обчислює прогнози за допомогою моделі ARIMA. Приймає навчальні
дані `train`
    тестові дані `test`, мінімальне `p_min` та максимальне
значення `p_max` параметра p.

    Повертає кортеж зі даними прогнозу та параметрами моделі.
    """

    # Створення й навчання моделі ARIMA
    model = AutoARIMA(start_p=p_min, max_p=p_max,
suppress_warnings=True)
    model.fit(train)
```

```
    # Визначені параметри моделі ARIMA
    model_order = model.get_fitted_params(deep=False)["order"]

    # Прогнозування на основі тестових даних
    fh = ForecastingHorizon(test.index, is_relative=False)
    forecast = model.predict(fh=fh)

    return forecast, model_order


def forecast_xgboost(train, test):
    """
    Обчислює прогнози за допомогою моделі з бібліотеки XGBoost.
Приймає
    навчальні дані `train` та тестові дані `test`.


    Повертає дані прогнозу.
    """

    # Створення моделі XGBRegressor та підготовка до перехресної
перевірки
    regressor = xgb.XGBRegressor(objective="reg:squarederror")
    forecaster = make_reduction(regressor,
window_length=test.size)

    # Параметри, які визначаються під час перехресної перевірки
    params = {
        "window_length": [5, 10, 15, 20, 25, 30],
        "estimator__n_estimators": [100, 500, 1000],
        "estimator__max_depth": [3, 5, 10, 15, 20]
    }

    # Навчання моделі за допомогою перехресної перевірки
    cv = SingleWindowSplitter(window_length=train.size,
fh=test.size)
    rscv = ForecastingRandomizedSearchCV(
        forecaster, cv=cv,
        param_distributions=params, n_iter=100
    )
    rscv.fit(train)

    # Прогнозування на основі тестових даних
    fh = ForecastingHorizon(test.index, is_relative=False)
    forecast = rscv.predict(fh=fh)

    return forecast
```

```
def forecast_lightgbm(train, test):
    """
    Обчислює прогнози за допомогою моделі з бібліотеки LightGBM.
Приймає
    навчальні дані `train` та тестові дані `test`.

    Повертає дані прогнозу.
    """

    # Створення моделі LGBMRegressor і підготовка до перехресної
перевірки
    regressor = lgb.LGBMRegressor()
    forecaster = make_reduction(regressor,
window_length=test.size)

    # Параметри для перехресної перевірки
    params = {
        "window_length": [5, 10, 15, 20, 25, 30],
        "estimator__max_depth": [3, 5, 10, 15, 20]
    }

    # Навчання моделі за допомогою перехресної перевірки
    cv = ExpandingWindowSplitter(initial_window=int(train.size *
0.5))
    gscv = ForecastingGridSearchCV(
        forecaster, cv=cv, param_grid=params,
        scoring=MeanAbsolutePercentageError(symmetric=True)
    )
    gscv.fit(train)

    # Прогнозування на основі тестових даних
    fh = ForecastingHorizon(test.index, is_relative=False)
    forecast = gscv.predict(fh=fh)

    return forecast
```

Лістинг Б.3 – Код файлу «plotting.py»

```
import matplotlib.pyplot as plt
import matplotlib.style
from sktime.utils.plotting import plot_series
matplotlib.style.use("seaborn")


def plot_forecast(train, test, forecast, title):
    """
    Будує і відображає графік прогнозу. Для цього використовує
навчальні дані
    `train`, тестові дані `test`, дані прогнозу `forecast`, а
також заголовок `title`.
    """
```

```
    plot_series(train, test, forecast, title=title, markers=["", 
"", ""],
        labels=["Навчальні дані", "Тестові дані", "Дані 
прогнозу"],
        x_label=forecast.index.name, y_label=forecast.name
    )
    plt.show()
```

Лістинг Б.4 – Код файлу «evaluating.py»

```
from sklearn.metrics import mean_absolute_error, 
mean_absolute_percentage_error


def forecast_errors(test, forecast):
    """
    Обчислює похибки прогнозу MAE та MAPE на основі
    тестових даних `test` і даних прогнозу `forecast`.

    Повертає кортеж значень MAE, MAPE.
    """
    mae = mean_absolute_error(test, forecast)
    mape = mean_absolute_percentage_error(test, forecast)

    return mae, mape
```